what happens when you downscale it.” He described the change in properties from bulk, to thin coating to nanoscale as discontinuous. “There is a lot for us to do in this nanospace to develop new materials.”

The final question of the event went to Steven Freeman, who teaches innovation at the University of Pennsylvania. He asked for an example of radical change versus the incremental changes described in much of the discussion.

Connelly answered first. “I think there is a basic change in the way biology is going to impact the materials sciences. We engineer biological process and produce very specific products at room temperature, atmospheric pressure and in aqueous media with no exotic metals.”

Pappas said the incremental advances in materials over the past 20 to 30 years add up a major change. “Breakthroughs are the sum of incremental changes over time,” he said.


Neil Gussman <NeilG@chemheritage.org> works at the Chemical Heritage Foundation. CHF shares common goals with, and is an Associated Organization of, IUPAC.

How Many Miles Have We Gone, InChI by InChI?*

by Alex Tropsha and Antony Williams

How simple the communications between people of different nations would have been had we all shared the same language! The attempts to develop such a universal language adopted by the entire human population have not been tremendously successful (e.g., Esperanto), but chemists, on the other hand, have succeeded in developing a language they all understand, that of chemical structure. Chemists of all nations could easily communicate their thoughts using traditional chemical structure drawings, but this way of information exchange becomes increasingly prohibitive with the growth of chemical databases. This is when special means for compact encoding of chemical structures such as SMILES and InChI become critical, making it feasible to store and transmit huge amounts of chemical information (many current chemical databases such as PubChem or ChemSpider include many tens of millions of records).

At the ACS meeting in San Diego in March 2012, we had an opportunity to celebrate the impact of the InChI (the IUPAC International Chemical Identifier) on the ability to communicate, link, and enhance the integration of chemistry across databases, across the Internet and in publications. Unlike the proprietary SMILES format, InChI was designed to be “a freely available, nonproprietary identifier for chemical substances that can be used in printed and electronic data sources, thus enabling easier linking of diverse data compilations and unambiguous identification of chemical substances.” Over the past decade, InChI has become widely recognized and used by the chemical community to search, connect, and exchange chemical structures. The first InChI symposium to be held at an ACS meeting was organized to highlight the history, recent developments, interesting applications, and current trends in InChI. The unique gist of this full-day symposium in the San Diego 2012 program could probably be best summarized by the presentation title of one of the speakers, who had the title “InChIs here, InChIs there, InChIs everywhere!” (see page 34).

Alan McNaught (left) chaired the original IUPAC activity that lead to the development of InChI in 2000 and has followed the project since. Jason Wilde is from the Nature Publishing group, represented on the InChi-Trust.

The morning session opened with Steve Heller presenting a brief history on how the InChI project started and an update on the present status. He emphasized that InChI was meant primarily to provide a way to link information, to be an addition to what is available today, not a replacement, and is an algorithm to produce a unique label. He commented on the success of InChI, with one measure being the uncoerced adoption that has been shown by publishers, database providers, and software developers.
Antony Williams from the Royal Society of Chemistry Cheminformatics team, and member of the ChemSpider project, discussed the “Great promise of navigating the internet using InChIs” and reflected on what the original purpose of InChI was and how far we had come in a relatively short period of time in terms of meeting the goals. In contrast to his usual focus on data quality issues, he started by commenting that the talk was about quantity—how much data was now being linked together and could be discovered using InChI as the linker. He asked “What would be the situation if InChI never existed?” and answered with the fact that ChemSpider would never have been built without it.

Keith Taylor from Accelrys provided an overview of why Accelrys supports InChIs in its cheminformatics toolkit despite having its own unique identifiers already. He emphasized it was because the community asked for InChIs and found them to be valuable. He provided examples of how different online systems used InChI and what the results were when Accelrys tools were used to perform searches against the online databases.

Martin Walker from the State University of New York at Potsdam discussed the use of InChIs in Wikipedia and their use in ChemBoxes and DrugBoxes. He also discussed how InChIs became an important part of the RSC Learn Chemistry wiki, an open website hosting educational content and encouraging educators to contribute their own data. The site hosts chemical data for over 2100 common chemicals and uses InChI and InChIKeys to facilitate structure searching across the site and as the basis of interactive quizzes.

Andrey Yerin from ACD/Labs gave a fascinating presentation regarding InChIKey collisions and how to experimentally estimate the rate of their occurrence using algorithmically-generated structure libraries. His conclusion was that InChIKeys do have a very low rate of collision as the degree of randomness, as expected by the design of the InChIKey, is very high. He did provide some very amusing examples of InChIKeys that contained “hidden words” and examples are shown in the figure below.

As a final talk in the morning session Juergen Swienty-Busch talked about where InChIs are (i.e., “here, there and everywhere”). He showed how InChIs continue to proliferate in popularity across chemistry databases on the web and how Elsevier now includes them in its Reaxys database as well as in its publications associated with data. Reed-Elsevier looks forward to further advances in InChI. Elsevier Properties SA is a board member on the InChI Trust and Elsevier is involved as members of three of the working groups. This indicates their belief in the standard.

The discussion of InChI continued in the afternoon. Yulia Borodina (U.S. FDA) discussed the challenges facing the FDA Substance Registration System when dealing with complex compounds such as biopolymers or synthetic polymers. Standard InChI is not designed to handle such structures, but it could be used to describe monomers and other building blocks of these complex structures.

Marcus Sitzmann from NCI discussed how InChI/InChIKeys are used within the NCI Chemical Identifier Resolver to annotate more than 80M unique chemical structures.
Conference Call

Daniel Lowe from NextMove Software addressed features in IUPAC nomenclature that cannot be addressed by standard InChI such as accurate representation of tautomers and mixtures of stereoisomers. These cases require special treatment.

In a flash talk, Richard Kidd discussed the use of InChI within the Royal Society of Chemistry and RSC’s contribution to the development of InChI standards.

Feranc Szalai from Mcule (Budapest, Hungary) highlighted the importance of InChI for unified annotation of chemical structures in aggregated databases that is independent of drawing conventions, tautomeric states, and other features such that each unique chemical structure is annotated only once. He stated that InChI provides the best solution for this problem.

Jon Chambers from EBI (Cambridge) made similar observations based on their group experience in assembling data in the ChEMBL database from different sources. He described their UniChem system that uses the standard InChI as a means of normalizing between different sources of chemical structures.

Bill Armstrong from Louisiana State University shared his experience in creating a teaching methodology to help researchers understand InChI as a unique, non-proprietary tool for identifying chemical structures.

Jason Wilde from the Nature Publishing Group (London) described the efforts of the InChI Trust to support the development of a nonproprietary InChI standard and discussed the ongoing efforts to provide technical solutions to develop standard InChI for difficult cases of complex structures such as polymers and mixtures.

Steve Boyer from IBM (San Jose, CA) described recent efforts of their group to employ InChI for annotating chemical structures in patents as part of chemical name to structure conversion. Their approach replaces chemical names in patent documents by InChI, enabling the search for chemical structures within certain textual content in the patents and patent literature.

In the final talk of the symposium, Laura Croft from Nature Publishing Group (London) described the launch of Nature Chemistry as part of nature.com and emphasized the utility of InChI to increase the discoverability of chemical structures and related information on nature.com and on the web in general.

This brief summary of presentations at the InChI symposium underscores the importance of InChI as universal identifier of chemical structures. InChI continues to proliferate in various areas of chemical research and to serve as a critical means of chemical information dissemination and exchange. The anticipated new developments will likely warrant a new InChI symposium within the next couple of years.

Alex Tropsha and Antony Williams were the symposium organizers. Alex Tropsha <alex_tropsha@unc.edu> is an adjunct professor of biomedical engineering and of computer science at the University of North Carolina in Chapel Hill, North Carolina, USA. Antony Williams <WilliamsA@rsc.org> is with the Cheminformatics Department of the Royal Society of Chemistry and is a member of the ChemSpider project.

* Reproduced with permission from the Chemical Information Bulletin of the ACS Division of Chemical Information Vol. 64, No. 2: Summer 2012; http://bulletin.acscinf.org/node/325.

Advanced Polymeric Materials

by Vera Kovacevic and Michael Hess

POLYCHAR, The World Forum on Advanced Polymeric Materials, belongs to a series of annual conferences taking place in different countries starting in USA (from 1992 until 2003), then moving on to Portugal (2004), Singapore (2005), Japan (2006), Brazil (2007), India (2008), France (2009), Germany (2010) and Nepal (2011). The conference enjoys a high international reputation and has been sponsored by IUPAC for more than 10 years.

The 20th World Forum on Advanced Polymeric Materials—POLYCHAR 20—was held from 26–30 March 2012 in Dubrovnik (Croatia), a medieval city, included in UNESCO’s World Heritage List. The POLYCHAR Scientific Committee, with 53 countries represented, nominated the University of Zagreb, Faculty of Chemical Engineering and Technology as the organizer of POLYCHAR 20 conference, with Vera Kovacevic as the conference convener and Stanislav Kurajica as the conference chair.

The name POLYCHAR is derived from polymer characterization but goes beyond the initial goal and now addresses the whole field of polymeric materials.