

Chemical Applications of the World-Wide-Web System

Henry S. Rzepa,*^a Benjamin J. Whitaker*^b and Mark J. Winter*^c

^a Department of Chemistry, Imperial College of Science Technology and Medicine, London, UK SW7 2AY

^b School of Chemistry, University of Leeds, Leeds, UK LS2 9JT

^c Department of Chemistry, The University of Sheffield, Sheffield, UK S3 7HF

Application of the Internet based World-Wide-Web system for efficient and intuitive delivery of chemical information in a wide variety of formats is illustrated *via* models for electronic publishing and interaction with molecular data.

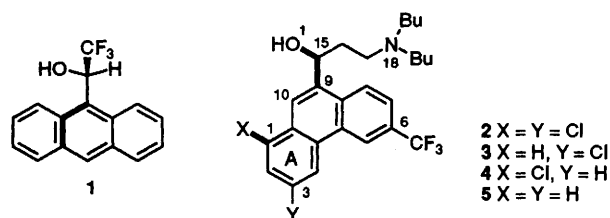
The increasing trend towards the visual and numerical complexity of molecular information derived from spectroscopic, structural or computational sources has hitherto not been matched by associated enhancements in the way the scientific community receives the information. The chemical literature has of necessity remained rooted in largely monochrome and two-dimensional presentation technology (Paper), while comparatively instant forms of communication such as electronic mail do not lend themselves to peer review and can suffer from difficulties in exchanging anything other than text-based documents. During the last three years, a new global communication medium based on 'hypermedia' concepts and termed World-Wide-Web (WWW)^{1,2} has emerged from research projects in computing science, and is now being adopted by many scientific disciplines. Only recently has it become clear that the molecular sciences have particular attributes which are well suited for this new medium.² We present here a number of novel chemical examples of the use of the WWW system which illustrate how this mechanism enables access to types of information impossible to provide on paper, and we indicate areas where rapid future development can be expected.

The WWW is founded on a peer-to-peer client-server model of information exchange³ and is based on a document type definition termed 'html' or hypertext mark-up language.³ This model has a number of attractive features for information dissemination. Each paper⁴, or more generally document fragment, is identified currently using a mechanism known as a 'URL' (uniform resource locator).⁴ The URL points to the physical location of the information and can be cited from within other documents, although the perception to the user is of a single document. Such electronic chemical communication can in principal reduce the cycle time to publication by at least an order of magnitude compared to a conventional printed journal. It is of course essential to retain the element of peer review, which could be achieved anonymously by the same electronic mechanism. This communication, for example, was mounted on WWW servers in London, Leeds and Sheffield, and was potentially available for refereeing within minutes of being submitted,⁵ *via* distribution of the URL of the paper as cited in ref. 5. We do however recognise that issues such as the integrity and publication date of the material and the nature of its accessibility and long term archival function will have to be addressed before this becomes accepted as an exclusive format for presenting and citing original work, as evidenced by the current appearance of this paper in the more conventional printed format.

We believe that this emerging technology fundamentally alters the way in which chemical data can be presented, in enabling three-dimensional structural features and even hyper-dimensional relationships such as potential energy surfaces to be easily stored and visualised on a computer. By embedding interactive 2 or 3D representations into documents, a novel forum for information dissemination and exchange is opened up. These and other themes are illustrated below.

(a) Molecular diagrams can be presented as two-dimensional mapped figures, where further analysis of a scientific point can be made to associate closely with particular molecular features, without increasing the visual complexity. Tradition-

ally this would have to be achieved by lengthy figure captions. In molecules 1–5, critical aspects of their structure⁶ (indicated here with bold bonds) can be elaborated, in our case with an audio comment rather than with written text.



(b) In hypertext format, traditional literature citations can be removed from the end of a paper and embedded as 'hyperlinks' into either the text or the 2D mapped diagrams referred to above. In principle, other sources of information such as databases or numerical data can be readily accessed from within the document and most importantly, used in the context of the discussion. Such links are activated by clicking on underlined text or diagrams.⁷ In the future, one might expect that hyperlinks could be mapped to the three-dimensional content of a structure, such as might be required to illustrate the active site of an enzyme or catalyst.

(c) Structural information can also be associated with a so called MIME (Multipurpose Internet Mail Extension) type^{7,8} and delivered as a 'hyperactive molecule' suitable for user-controlled processing such as rotation, manipulation and analysis. Traditionally this has been achieved by depositing, *e.g.* molecular coordinate files in a database archive to which general access may only be gained many months later.⁹ We have used this to link to a protein structure file,¹⁰ which when activated on the screen can be freely rotated by the user and enhanced with, *e.g.* a ribbon display or a bond length. The coordinate information is stored with a filename suffix such as .pdb, and MIME associations of the type chemical/pdb⁸ enable these files to be displayed using suitable programs on the local computer. A major advantage is the considerable saving in transmission time of the compressed atomic coordinate data, compared with 'bitmapped' 2D or 3D representations.¹⁰

Other applications of the MIME concept that we have demonstrated include: (i) linking a structure diagram automatically to generate derived molecular properties such as wave functions from a MOPAC or GAUSSIAN 92 calculation,¹¹ (ii) creating a link to a script file on a remote computer which can perform, *e.g.* an isotope distribution¹² or element percentage¹³ calculations on a specified molecular formula, (iii) using links to establish a connection to a remote database service *via* the WWW¹⁴ or a separate terminal session,¹⁵ (iv) using links and the chemical MIME concept to initiate electronic mail and video conferencing sessions to communicate with authors and collaborators rapidly, including enclosing structural diagrams with the message to appear directly on the screens of the recipient *via* the so called 'whiteboard' technique.¹⁶

(d) Remotely stored spectroscopic or analytical data could be linked to a published document, which would enable local reprocessing in order to clarify or enhance particular aspects.



Fig. 1 The protein triose phosphate isomerase visualised using the EYECHEM package via a hyperlink to a 'thumbnail' diagram

File Options Navigate Annotate Help

Document Title: [REDACTED]

Document URL: [REDACTED]

WebElements Isotope pattern

This is the result of your isotope pattern calculation for:

C6H10Br2O6

Empirical formula: C6H10Br2O6

mass	%
336	51.0
337	3.6
338	100.0
339	7.1
340	49.7
341	3.5
342	0.7
343	0.0

Back Forward Home Reload Open Save As

File Options Navigate Annotate Help

Document Title: [REDACTED]

Document URL: [REDACTED]

This is a searchable index. Enter search keywords: [REDACTED]

WebElements

Isotope pattern calculator

Please enter a chemical formula as search string. Correctly nested brackets and some pseudoelements such as Me or Cp are OK. So formulae such as C2H4Br2, EtOH, Mo(CO)6, and [Mo(CO)2(P(OMe)3)Cp]2 will work. Select here for a list of allowed pseudoelements. **The formula is case sensitive.**

This service is provided by the University of Sheffield Chemistry Department

How does it work?

This service works by the MacHTTP server software requesting HyperCard to execute the calculation. Communication between the programs is achieved with AppleScript. It's all very experimental.

Back Forward Home Reload Open Save As Clone New Window Close Window

Fig. 2 An example of using a remote script to calculate the isotopic distribution pattern defined by a molecular formula

In theory, the remote data could reside directly on the spectrometer system itself, thus establishing a direct link between the published information and its original source. The conventional security mechanisms of password protection would of course remain.

(e) Time-dependent phenomena, such as a moving image of the development of a Flame front in a rapid compression machine¹⁷ or a chemical visualiser for the simultaneous time development of the concentrations of nine species in a flow reactor,¹⁸ can be recorded as a video animation in MPEG or Quicktime format and displayed in the context of the scientific discussion.

(f) A sense of the three-dimensional perception of a complex structure, or of a derived wave function or other property can be imparted by adding a time-dependent rotation. We have used this feature to illustrate the π -facial characteristics of a calculated molecular electrostatic potential, having failed to find an entirely suitable viewing angle for printing.¹⁹ The imminent introduction of cheap 'active' liquid crystal shuttered glasses for games consoles will soon enable such animations to be viewed with a true 3D perception.

(g) At present, the text-based content of a paper can be indexed into a key-word searchable form, which in principle could be extended to a global scale, thus enhancing the browsability of the information, and encouraging serendipity. It would not be necessary to wait for, *e.g.*, an annual keyword index to perform such a search. In the future, we should expect that similar indexing might extend to the molecular structure and sub-structure content of the paper, or to databases of compound information.²⁰

(h) A feedback mechanism can be implanted into the electronic paper *via* so-called 'forms' descriptors which allow the user to fill in electronic forms on-line. Such forms have already been used to register participants at conferences, and to initiate electronic mail contact and MOPAC calculations.¹¹ Direct entry of molecular or structural information is a natural development of this concept.

Whilst we have focused on how the traditional scientific paper can be enhanced, other forms of communication can be envisaged. A scientific talk in the form of a slide show can be presented on a global scale using the WWW format.²¹ It is possible to produce reference works or textbooks in this format with live links to the internet, which might offer a constantly updatable source of current information.²² Provided initial resistance to these novel forms of communication is overcome,²³ we expect that such techniques will rapidly evolve into a fundamentally new tool for facilitating chemical communication.

We thank the JISC for an equipment grant (to H. S. R. and B. J. W.), and Dr K. Brodlied and Mr G. Stead of Leeds University and Dr P. Murray-Rust and Mr M. Hargreaves of Glaxo Group Research (Greenford) for stimulating discussions and financial support.

Received, 18th May 1994; Com. 4/02963A

Footnotes

† **Editor's note.** This communication is being published in the hope that it will stimulate discussion and encourage the use of wide-area electronic networks in chemical research.

* e-mail rzepa@ic.ac.uk, <http://www.ch.ic.ac.uk/rzepa.html>; e-mail: benw@chem.leeds.ac.uk, <http://chem.leeds.ac.uk/People/Whitaker.html>; e-mail M. Winter@sheffield.ac.uk; <http://www2.shef.ac.uk/chemistry/Chem-staff/mjw/Mark-Winter.html>

‡ Single underlining is used to indicate a hypertext link embedded in the original document.

§ Viewing documents on the WWW requires a 'browser' or 'client' computer program such as NCSA MOSAIC, which is currently freely available for X-Windows, Microsoft Windows and Apple Macintosh

systems. Further details can be obtained by connecting to the NCSA 'home' page using the URL <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html>. Publishing documents on the WWW requires 'server' software, also freely available from the same source. The local computer system must be connected to the Internet using the TCP/IP mechanism, which can be achieved either *via* a permanent institutional link, or *via* a modem to connect to an Internet service.

¶ Document preparation in html form can be achieved using several suitable word or text processor programs, including a number of public domain programs. Programs also exist which can convert (filter) existing documents, including processing a document saved in Rich Text Format or LaTeX format into html format. For further details, see <http://cbl.leeds.ac.uk/nikos/tex2html/doc/latex/latex2html/latex2html.html>, <http://www.ch.ic.ac.uk/mac/mosaic.html> or http://info.cern.ch/hypertext/WWW/Tools/Word_proc_filters.html

|| A WWW document will normally contain one or more 'hotspots' or 'buttons'. When the cursor is placed on such a spot and activated, actions such as the display of another document or a graphical image, the playing of sound or movie clips or the creation of other application windows can be invoked.

References

- 1 T. J. Berners-Lee, R. Cailliau, J.-F. Groff and B. Pollermann, CERN, World-Wide Web: The Information Universe, in *Electronic Networking: Research, Applications and Policy*, Meckler Publishing, Westport, CT, USA, 1992, vol. 2, pp. 52-58.
- 2 H. S. Rzepa, *Chemical Design Automation News*, 1994, 9(2), 1; P. Murray-Rust, Use of WWW in Biology, First International Conference on the World-Wide Web, CERN, Geneva, 1994; H. S. Rzepa, Use of WWW in Chemistry, *ibid.* The discussion papers are available as <http://cui.www.unige.ch/WWW94/Workshops/workshop.list.html> For a list of chemistry WWW sites, see <http://www2.shef.ac.uk/chemistry/chemistry-www-sites.html> or <http://www.chem.ucla.edu/chempointers.html>
- 3 T. J. Berners-Lee, *Hypertext Markup Language*, CERN, Geneva, 1993. See <http://info.cern.ch/hypertext/WWW/MarkUp/HTML.html>; <http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>; http://info.cern.ch/hypertext/WWW/MarkUp/HTMLPlus/htmlplus_1.html
- 4 T. J. Berners-Lee, *WWW Names and Addresses, URIs, URLs, URNs*, CERN, Geneva, 1993. See <http://info.cern.ch/hypertext/WWW/Addressing/Addressing.html>. The finite lifetime of a URL implicit in its apparent dependency on machine and discrete file, *viz.* 'action://server-name/file-path' is likely to be addressed by the use of a unique URN, or Uniform Resource Name, for documents in the near future.
- 5 H. S. Rzepa, Imperial College, 1994; http://www.ch.ic.ac.uk/rzepa/RSC/CC/4_02963A.html, B. J. Whitaker, University of Leeds, 1994; http://chem.leeds.ac.uk/papers/html/chem-com/4_02963A.html; M. J. Winter, University of Sheffield, 1994; http://www2.shef.ac.uk/chemistry/www-publications/4_02963A.html. We suggest that in the future, the validation, storage and archiving of such files should be on a server specifically allocated to the journal and under its control.
- 6 P. Camilleri, D. Eggleston, H. S. Rzepa and M. L. Webb, *J. Chem. Soc., Chem. Commun.*, 1994, 1135. See also http://www.ch.ic.ac.uk/rzepa/RSC/CC/3_03989G.html
- 7 N. Borenstein and N. Freed, Internet Request for Comment (RFC) No. 1521, 1993. See also <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/mailcap.html>
- 8 H. S. Rzepa and P. Murray-Rust, Internet Draft: Chemical Mime Types, May-November 1994. See <ftp://cnri.reston.va.us/internet-draft-rzepa-chemical-mime-type-00.txt>
- 9 F. H. Allen, J. E. Daview, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith and D. G. Watson, *J. Chem. Inf. Comp. Sci.*, 1991, 31, 187.
- 10 M. C. Peitsh, SWISS-Image, A Collection of MacroMolecule Images, 1994. See <http://expasy.hcuge.ch/pub/Graphics/IMAGES/GIF/>
- 11 B. J. Whitaker and H. S. Rzepa, University of Leeds and Imperial College. See <http://chem.leeds.ac.uk/Project/QCC.html> or http://www.ch.ic.ac.uk/chemical_mime.html
- 12 M. J. Winter, University of Sheffield, 1993. See <http://www2.shef.ac.uk/chemistry/web-elements/isotope.script>

- 13 M. J. Winter, University of Sheffield, 1993. See <http://www2.shef.ac.uk/chemistry/web-elements/percent.script>
- 14 M. J. Winter, University of Sheffield, 1993. See <http://www2.shef.ac.uk/chemistry/web-elements/web-elements-home.html>
- 15 H. S. Rzepa, Imperial College. See <http://www.ch.ic.ac.uk/facilities.html>
- 16 M. J. Pilling, H. S. Rzepa and B. J. Whitaker, Collaborative Molecular Modelling: An ATM-SuperJanet Pilot Project, University of Leeds and Imperial College, demonstrated on July 13th, 1994.
- 17 J. Smith, personal communication, University of Leeds. See <http://chem.leeds.ac.uk/chaos/1gn.html>
- 18 B. J. Whitaker and J. Smith, University of Leeds. See <http://chem.leeds.ac.uk/chaos/vis.html>
- 19 J. Plater, H. S. Rzepa and F. Stoppa, *J. Chem. Soc., Perkin Trans 2*, 1994, 399. See also http://www.ch.ic.ac.uk/rzepa/RSC/P2/3_07186C.html
- 20 For example the Cambridge Crystallographic Data Centre (England) at <http://csd.vx2.ccdc.cam.ac.uk/> (see also ref 9).
- 21 H. Chaves, A. M. Lobo, S. Prabhakar and H. S. Rzepa, 1st European Computational Chemistry Conference, Nancy, May 1994. See <http://www.ch.ic.ac.uk/talks/>
- 22 H. S. Rzepa, *Interactive Guide to Computational Organic Chemistry*, to be published. See also <http://www.ch.ic.ac.uk/igco/>
- 23 B. Barber, *Scientific Manpower* (NSF 61-34), 1960, pp. 36-47.