

Nature Chemistry improves publishing chemistry: a detailed analysis

Egon Willighagen 

Published March 19, 2009

Citation

Willighagen, E. (2009). Nature Chemistry improves publishing chemistry: a detailed analysis. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/40377-hz881>

Keywords

Inchi, Chemistry, Jmol

Abstract

Nature Chemistry just released the first issue with a few free papers, like Asymmetric total syntheses of (+)- and (-)-versicolamide B and biosynthetic implications by Miller et al. (DOI:10.1038/nchem.110).

Copyright

Copyright © Egon Willighagen 2009. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

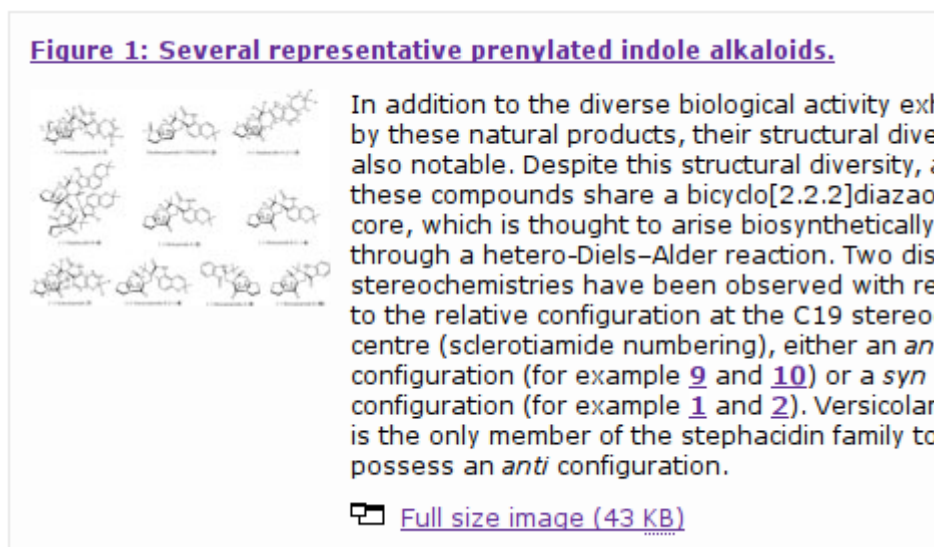
Nature Chemistry just released the first issue with a few free papers, like *Asymmetric total syntheses of (+)- and (-)-versicolamide B and biosynthetic implications* by Miller et al. (DOI:10.1038/nchem.110).

Now, we've seen the Royal Society of Chemistry's [Project Prospect](#) (see [RSC: the first publisher to go semantic!](#)) and ChemSpiders recent [ChemMantis](#) system which enriches the papers with machine readable representations of the molecules discussed in those papers. The new Nature publication has been in the works for a while, and they [asked](#) the community before what a Nature Chemistry paper should like like, and I replied in [Re: What should a Nature Chemistry paper look like?](#) .

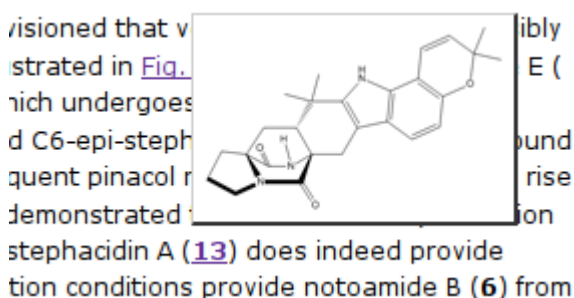
The verdict

So, have the been listening? Is the HTML they produce semantic? Is it data rich? Or is it just another hamburger? Well, I am very happy to see some of the suggestions I made picked up (though I do not fool myself in believing I am the only one that suggested those features). A tour of good things, and points for improvement.

The first impression is not shocking; it looks like any other interface, with molecules drawn as images in the paper:



All structures that are numbered and linked (as in *C6-epi-stephacidin A (Compound 13)*) have a hover-over function to popup a drawing of the structure:



chem-bla-ics

The popup image is a nice gimmick, but not really semantically useful. The link, however, is! It points to a separate supplementary page with further information which include a image of the 2D structure and, following a link, the 3D structure in [Jmol](#). Moreover, it comes with the machine readable representations:

Compound (+)-8

(+)-Versicolamide B

[View in PubChem](#)

[View in 3D \(6 KB\)](#) | [Download CML file \(5 KB\)](#) | [Download ChemDraw file of structure \(5 KB\)](#)

Chemical Formula: C₂₆H₂₉N₃O₄

Molecular Weight: 447.53

Elemental Analysis: C, 69.78; H, 6.53; N, 9.39; O, 14.30

InChI=1/C26H29N3O4/c1-22(2)10-8-14-16(33-22)7-6-15-18(14)27-20(31)25(15)13-
/h6-8,10,17H,5,9,11-13H2,1-4H3,(H,27,31)(H,28,30)/t17-,24+,25+,26+/m1/s1

InChIKey: RNWRZMCJFWSZOX-XRNKBYASBG

Standard InChI=1S/C26H29N3O4

/c1-22(2)10-8-14-16(33-22)7-6-15-18(14)27-20(31)25(15)13-26-17(23(25,3)4)12-24
/h6-8,10,17H,5,9,11-13H2,1-4H3,(H,27,31)(H,28,30)/t17-,24+,25+,26+/m1/s1

Standard InChIKey: RNWRZMCJFWSZOX-XRNKBYASSA-N

SMILES: O=C1[C@@]23N(CCC3)C(=O)[C@@]3(C[C@@]4(C([C@H]3C2)(C)C)C(=O)Nc

This is indeed interesting, and a big step forward, though please do note my comments later. For convenience, all molecules with such supplementary information is available from the special Chemical Compounds section of the paper:

- + Methods
- + Author contributions
- + Figures & tables
- + Chemical compounds
- + Acknowledgements
- + Author contributions
- + Supplementary info

Excellent! This really is a step forward towards a data-rich paper! Indeed, I will shortly write up a [Bioclipse](#) plugin for Nature Chemistry, which will download all molecular structures based on the DOI! Anyway, more on that later... For this article, that table looks like:

Compound Data Index

From the following article

[Asymmetric total syntheses of \(+\)- and \(-\)-versicolamide B and biosynthetic implications](#)

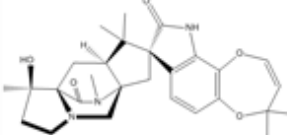
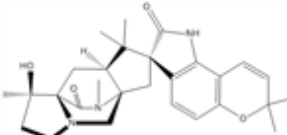
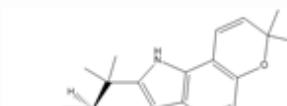
Kenneth A. Miller, Sachiko Tsukamoto & Robert M. Williams

Nature Chemistry **1**, 63-68 (2009) Published online: 19 March 2009

doi:10.1038/nchem.110

[Supplementary information](#)


[Download all compounds in a single file \(150 KB\)](#)

	<p>Compound 1 (-)-Paraherquamide A</p> <p>View in PubChem</p> <p>View compound page (6 KB) View in 3D (7 KB)</p> <p>Download ChemDraw file of structure (5 KB) Download CML file (5 KB)</p>
	<p>Compound 2 Paraherquamide F</p> <p>View in PubChem</p> <p>View compound page (6 KB) View in 3D (6 KB)</p> <p>Download ChemDraw file of structure (5 KB) Download CML file (5 KB)</p>
	<p>Compound (+)-3 (+)-Stephacidin A</p> <p>View in PubChem</p>

By now, you likely also noted the links to [PubChem](#), and indeed, upon publication of a paper, all structures are deposited in the public domain:

 **Data Source:** 

Depositor: Nature Chemistry
External ID: [nchem.110-comp13](#)

 **Depositor-Supplied Synonyms:** (Total: 2) 

C6-Epi-stephacidin A
[nchem.110-comp13](#)

At last but not least, each molecule is available in the [Chemical Markup Language](#) (with 2D coordinates)! And you know I am a very happy CML user for a long time (see e.g. Peter's recent blog [Egon Willighagen and CML](#)). BTW, one comment on the CML: the namespace used is the outdated namespace, **not** the current one (see [There can be only one \(namespace\)](#)). (But the CDK and Bioclipse will read it anyway.)

Details matter

So, while the first impression was not shocking, it was a bit deceptive. *Nature Chemistry* really changes publishing of chemistry. But I have bad news too. They need to improve the HTML they produce.

But before pointing out some missed chances, let me reply *inter alia* to Peter's recent work on the Open Source plugin for including semantic chemistry in MS-Word documents (see [How can we publish semantic chemical documents? </i>(<https://blogs.ch.cam.ac.uk/pmr/2009/03/16/how-can-we-publish-semantic-chemical-documents/>)): *Nature Chemistry seems to have done a great job with existing tools. Nevertheless, I fully back up Peters comment that while the plugin is useless without Word, the results produced with the plugin are extremely Open Standard, and enormously reusable! Indeed, while the Word file format is only formally an true Open Standard, the file format is plain XML, and extracting content bearing the CML namespace is trivial.*

Which reminds me, if someone from the Nature Chemistry team is reading this, please point me to a blog what tools actually *are* involved in publishing a Nature Chemistry paper! I think we all like to know.

Now, the HTML has room for improvement. First of all, a look at the metadata defined for the web page of the article shows a *description* and *keywords* about the journal, not the article, and the same goes for the web pages for the molecules:

```
<meta name="keywords" content="nature, chemistry, nat
<meta name="description" content="Nature Chemistry of
<link rel="shortcut icon" href="http://www.nature.com
```

Additionally, the compound details web page has no special markup for the machine readable information:

```
<p><span class="b">InChIKey:</span> YCWBTRJWYADFLQ-LROUJFHJBN</p><
```

Or, if it does, it's still mixed with markup for visual pleasing output:

```
<p class="inchi"><span class="b">Standard InChI=</span>1S/C26H29N3O3
```

Still, the HTML is clean enough to have some regular expressions extract a good deal of information, and there is also still the PubChem deposition.

Beyond connection tables

Like many other chemistry journals, Nature Chemistry does not consider properties of the molecule interesting, and NMR spectra are hidden in the Supplementary Information. This paper in particular, disregards a lot of machine readable facts by putting all experimental section bits in a PDF document. So, the next challenge for Nature Chemistry will be to get the

chem-bla-ics

authors of papers contribute the original spectra (JCAMP-DX, CMLspect, etc) in the supplementary information section. Better, have the raw data or even the NMR peak-atom annotations deposited in public repositories such (see [Open NMR data: raw curves and annotated peak lists](#)).

All in all, I am rather positive about the first Nature Chemistry issue, and like to thank the editors and paper authors for there efforts on improving publishing chemistry!