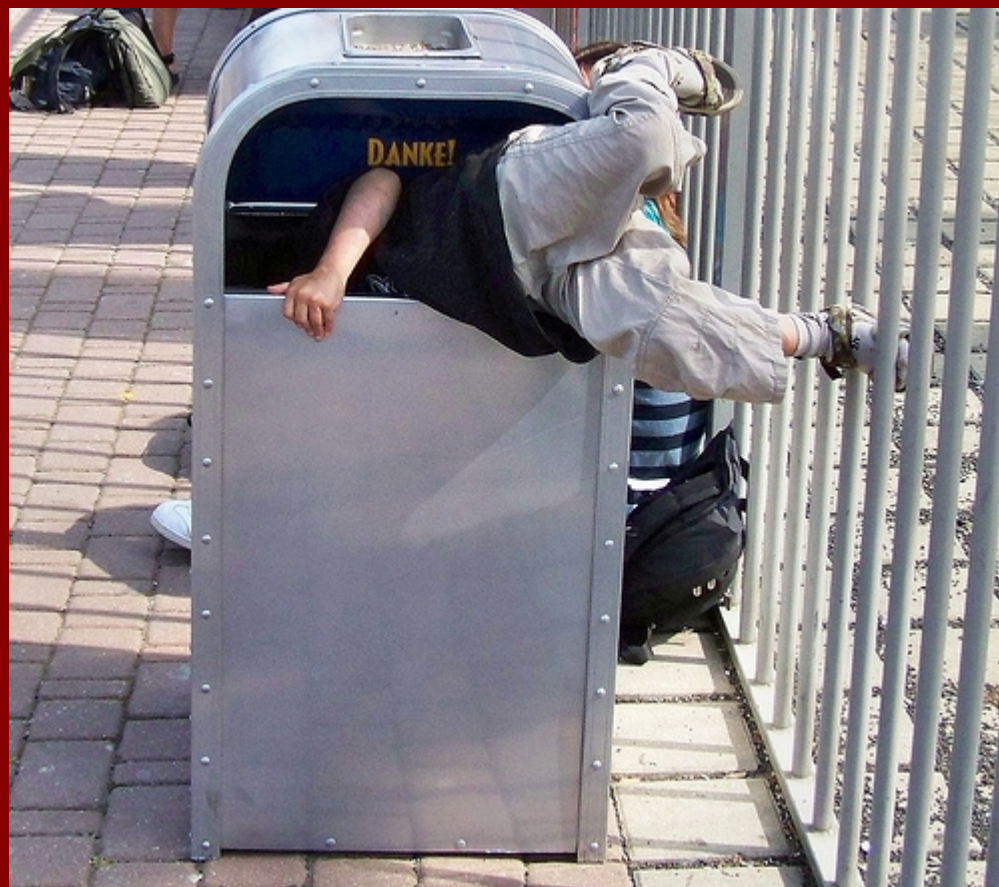


Prospecting for chemistry in publishing

Richard Kidd
ICIC 2008
kiddr@rsc.org



What are we marking up?

- Chemical compounds (InChI, ChEBI)
- Chemical classes and parts (ChEBI)
- Chemical terms from the IUPAC Gold Book
- Reactions and techniques (RXNO)

- Gene products: function, process, location (GO)
- Nucleotide and polypeptide sequence terms (SO)
- Cell types (CL)

Compound information 'BINOL'

Synonyms:

- (S)-BINOL
- 1,1'-Binaphthyl-2,2'-diol
- 2,2'-binaphthol
- BINOL
- 1,1'-binaphthol
- 1,1'-binaphthyl-2,2'-diol

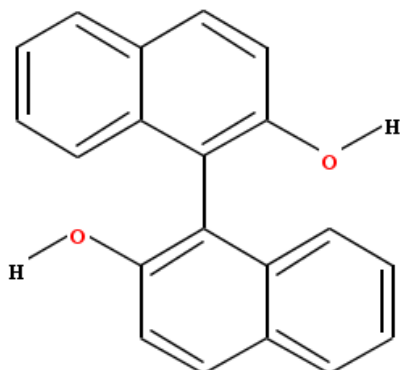
SMILES: [H]Oc2ccc1ccccc1c2c4c(ccc3ccccc34)O[H]

InChI: InChI=1/C20H14O2/c21-17-11-9-13-5-1-3-7-15(13)19(17)20-16-8-4-2-6-14(16)10-12,21-22H

InChIKey: InChIKey=PPTXVXKCQZKFBN-UHFFFAOYAX

CML (Chemical Markup Language) Representation: Download File

2-D Representation:

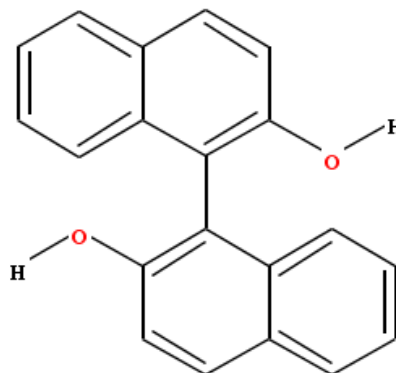


12,21-22H

InChIKey: InChIKey=PPTXVXKCQZKFBN-UHFFFAOYAX

CML (Chemical Markup Language) Representation: Download File

2-D Representation:



Other resources:

- Search for this compound in PubChem
- Search for this compound in SureChem patents

Other articles referencing this compound:

- **Ligand accelerated indium(III)-catalyzed asymmetric alkylation of aldehydes with 2-methyl-3-butyn-2-ol as an ethyne equivalent donor**
DOI: 10.1039/b614958h

Contact us | Close this window

catalytic system, which led to the identification of a new

RSS for human readers

RssReader 1.0.88.0

File Edit View Go Tools Help

Get Add Group Edit Clear Delete

Type: All Date filter: All Keyword:

My feeds

- Nature (157)
- Others (392)
- RSC (191)
- RSC Journals (1102)
- RSC - Analyst latest articles (19)
- RSC - Annu. Rep. Prog. Chem., Sect. A: Inorg. Ch...
- RSC - Annu. Rep. Prog. Chem., Sect. B: Org. Chem...
- RSC - Annu. Rep. Prog. Chem., Sect. C: Phys. Chem...
- RSC - Chem. Commun. latest articles (218)
- RSC - Chem. Soc. Rev. latest articles (30)
- RSC - Chemical Biology Virtual Journal latest article
- RSC - Dalton Trans. latest articles (127)
- RSC - Green Chem. latest articles (30)
- RSC - J. Anal. At. Spectrom. latest articles (23)
- RSC - J. Environ. Monit. latest articles (36)
- RSC - J. Mater. Chem. latest articles (109)
- RSC - Lab Chip latest articles (49)
- RSC - Mol. BioSyst. latest articles (21)
- RSC - Nat. Prod. Rep. latest articles (23)
- RSC - New J. Chem. latest articles (38)
- RSC - Org. Biomol. Chem. latest articles (112)
- RSC - Photochem. Photobiol. Sci. latest articles (2)
- RSC - Phys. Chem. Chem. Phys. latest articles (11)
- RSC - Soft Matter latest articles (22)
- Software (970)
- Technology (229)
- test (434)


RSC - Photochem. Photobiol. Sci. latest articles (25)

R	!	Headlines	Received dates	Published dates
		Picosecond time-resolved infrared study of 2...	24/05/2007 15:58	24/05/2007 00:00
		Novel emission properties of melem caused b...	23/05/2007 16:16	23/05/2007 00:00

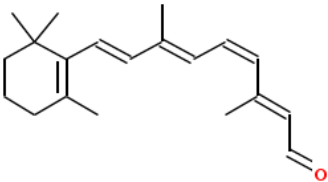
Address:

Read more | Open in browser

Chlorophyll derivatives as visual pigments for super vision in the red



Ilyas Washington, Jilin Zhou, Steffen Jockusch, Nicholas J. Turro, Koji Nakanishi, Janet R. Sparrow
(Paper from Photochem. Photobiol. Sci.)
Ilyas Washington, Photochem. Photobiol. Sci., 2007, DOI: 10.1039/b618104j
To cite this article before page numbers are assigned, use the DOI form of citation above.
Ontology Terms: photoreceptor cell; crystallin accumulating cell; visual perception; response to blue light; photoreceptor activity; response to red light
Primary Compounds:
11-cis-retinal:



Done

RSS for computers

```
<item rdf:about=http://xlink.rsc.org/?DOI=b716356h&RSS=1>
<title> [... title] </title>
<link>http://xlink.rsc.org/?DOI=b716356h&RSS=1</link>
<description> [... blah] </description>
<content:encoded> [... human-readable stuff</content:encoded>
[... dublin core stuff ...]
<content:items>
  <rdf:Bag>
    <rdf:li>
      <content:item rdf:about="info:inchi/InChI=1/C22H22NO4/c1-13-16-11-21(26-
4)20(25-3)10-15(16)8-18-17-12-22(27-5)19(24-2)9-14(17)6-7-23(13)18/h6-
12H,1-5H3/q+1"/>
    </rdf:li>
    <rdf:li>
      <content:item rdf:about="http://purl.org/obo/owl/SO#SO:0000028"/>
    </rdf:li>
  </rdf:Bag>
</content:items>
</item>
```

Text search

- “not in the Top 5” ;-(
 - Maik Annies, Syngenta, ICIC 2008
- Improved simple/advanced search under test
- Also need web service to return subject & compound results

Chemical structure search

Prospect Structure Search Results

The results show any enhanced HTML articles containing an exact match for the structure you are searching for. Similar structures are also shown.

The structure searching is at present restricted to the RSC journal articles which have been enhanced as part of RSC Project Prospect. Please email the Project Prospect team with any queries, feedback or suggestions for improvements.

Results

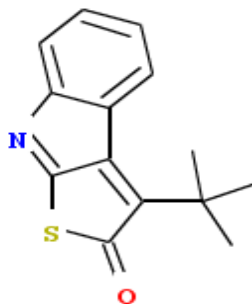
Searching for SMILES:

CC(C)(C)C=3C(=O)SC2=NC1=CC=CC=C1C2=3

InChI:

InChI=1/C14H13NOS/c1-14(2,3)11-10-8-6-4-5-7-9(8)15-12(10)17-13(11)16/h4-7H,1-3H3

2-D Representation:



Exact Matches

Thiocarbonyl induced heterocumulenic Pauson–Khand type reaction: expedient synthetic method for thieno[2,3-*b*]indol-2-ones

Takao Saito, Hiroshi Nihei, Takashi Otani, Toshiyuki Suyama, Naoki Furukawa, Masatoshi Saito, *Chem. Commun., Chem.*, 2008, (2), 172-174

DOI: 10.1039/b712739a

- [Show close matches](#)
- [Back to Search form](#)

The ALPSP/Charlensworth Award for
PUBLISHING INNOVATION
WINNER 2007



PROJECT PROSPECT
Developed by the
Royal Society of Chemistry

ALPSP
www.alpisp.org

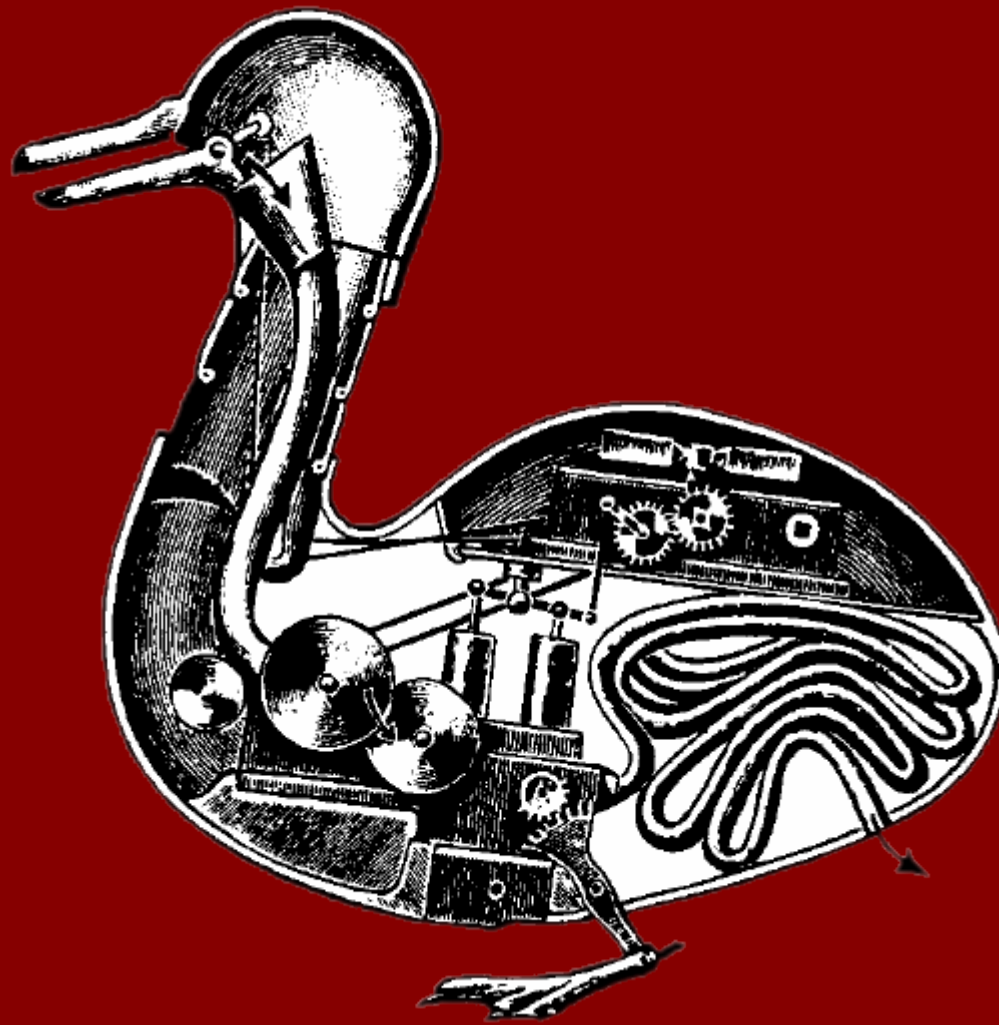
The Association of Learned and Professional Society Publishers
Shaping the Future of Learned and Professional Publishing

B. T. Donovan

Bernard Donovan President
The Association of Learned and Professional Society Publishers



How does this really work?





L^AT_EX



Data capture

XML



XML



Editing and proof-reading



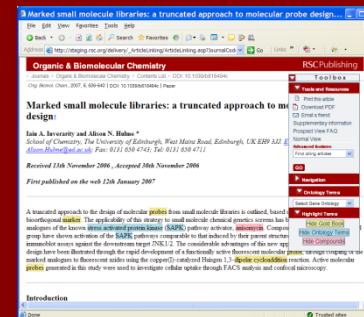
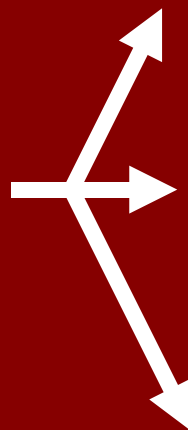
XML



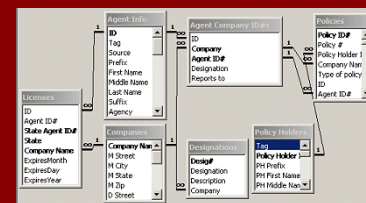
Text mining (Oscar)



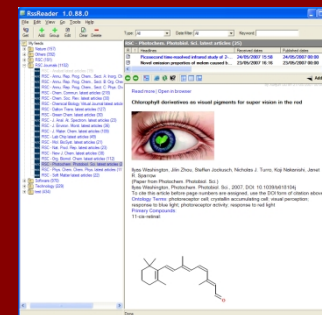
Manual QA



Enhanced HTML



Database



Enhanced RSS

Oscar



Centre For Molecular Science Informatics

Tokenizer

Morphological analyser

Name to structure converter

Workflow component

Annotation environment

<http://oscar3-chem.sourceforge.net/>

Why is this hard?



How many numbered compounds actually are named in a given paper?

iloprost (1)
tributyl-1-hexynylstannane (2)
the desired 2-heptyne (3)
methyl-Pd(II) iodide 4 or 4'
alkynylstannane 5
the hypervalent stannate 6
(alkynyl)(methyl)Pd(II) complex 7
the desired methylalkyne 8
compounds 9–14

the stannyl precursors 15 and 16
methylated compounds 17 and 18
stannyl precursor 19
iloprost methyl ester 20

“iloprost methyl ester” is the real name, but you need to know that iloprost is a monocarboxylic acid!

Annotate this...

A series of mono and di-N-2,3-epoxypropyl N-phenylhydrazones have been prepared on a large scale by reaction of the corresponding N-phenylhydrazones of 9-ethyl-3-carbazolecarbaldehyde, 9-ethyl-3,6-carbazoledicarbaldehyde, 4-dimethyl-amino-, 4-diethylamino-, 4-benzylethylamino-, 4-(diphenylamino)-, 4-(4,4'-dimethyl-diphenylamino)-, 4-(4-formyldiphenylamino)- and 4-(4-formyl-4'-methyldiphenyl-amino)benzaldehyde with epichlorohydrin in the presence of KOH and anhydrous Na₂SO₄.

From *Molecules*, via the BioNLP list

...and it gets worse

Part of speech ambiguity: tosylates; noun or verb?

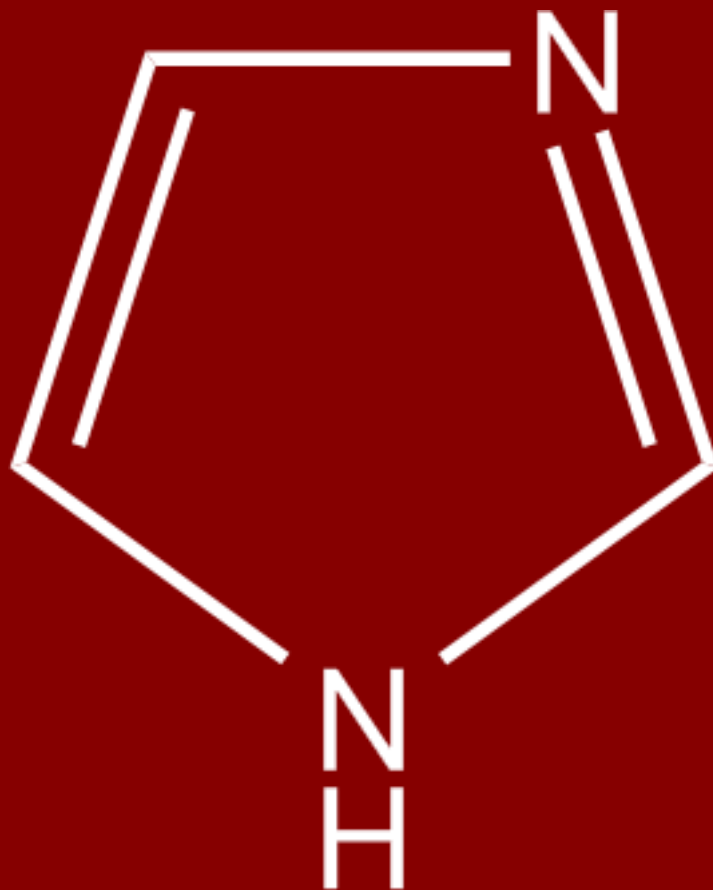
$E1cB_{\text{irrev}}$ mechanism and the $E2$ mechanism in which the transition state is $E1cB$ -like.⁴ Perhaps the surest sign that 11 and 12 undergo elimination by the $E2$ mechanism is the relative percentages of their (*E*)- and (*Z*)-alkene products.

Base-catalyzed elimination of TsOH from simple secondary alkyl tosylates under non-ion pairing conditions normally produces 15–35% (*Z*)-alkene by the $E2$ pathway, with the range reflecting steric effects as well as product stability.^{17,18}

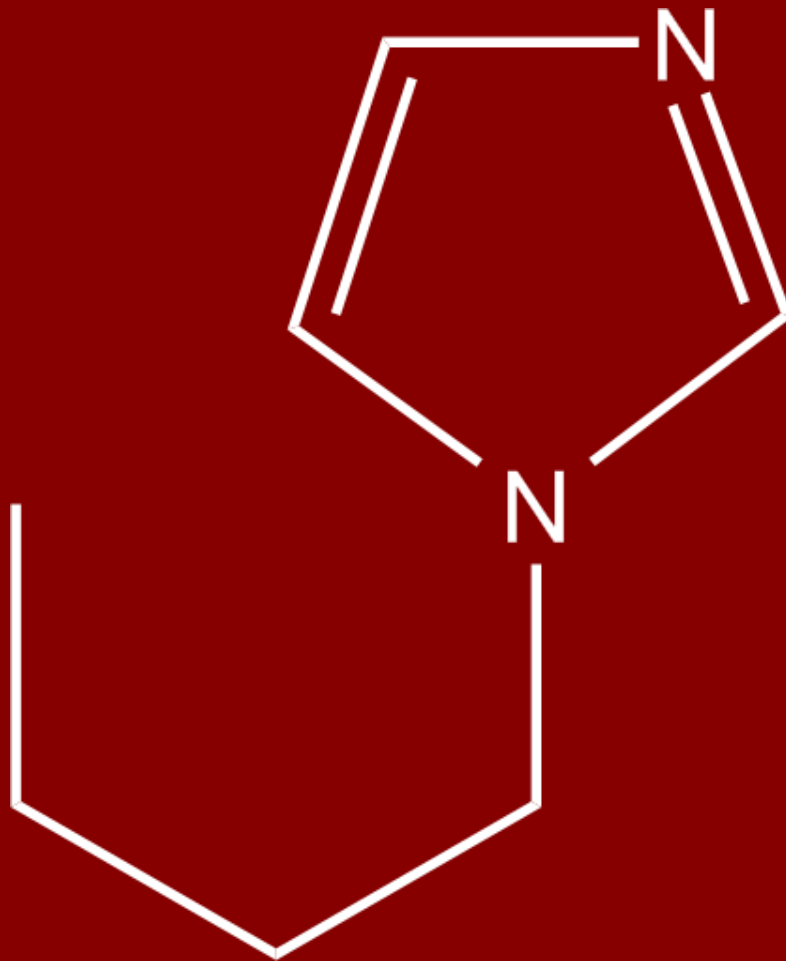
Steric effects have been implicated because other excellent leaving groups in $E2$ reactions, in particular iodide and bromide, produce only 67% as much of the (*Z*)-alkene as found with tosylate substrates.¹⁷ This differential is less

Even a simple chemical name can mean more than one thing...

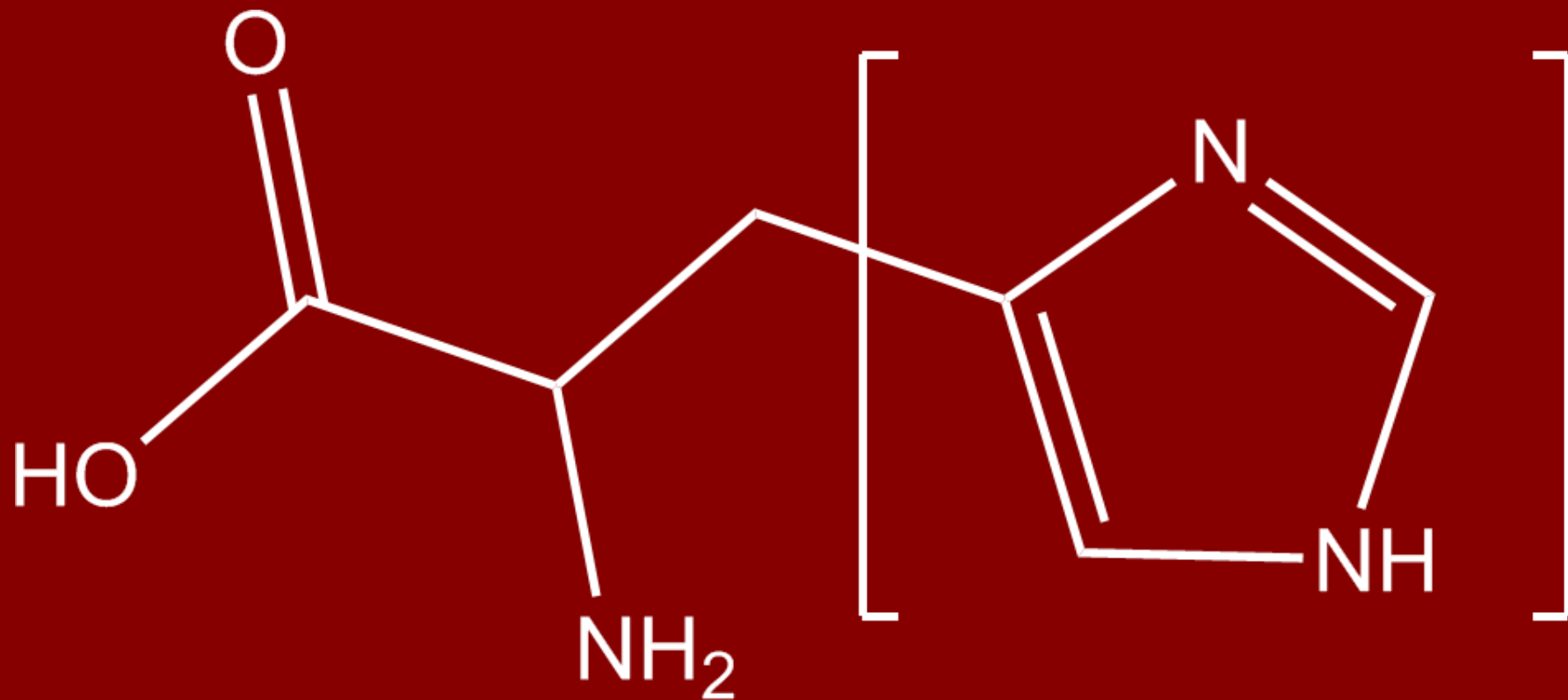
Imidazole



An imidazole



The imidazole side-chain/ group/ring/*etc.*



Can ChEBI handle this?

★ Imidazoles (CHEBI:24780)

★ Imidazole (CHEBI:16069)

★ Imidazole ring not yet

★ Imidazolyl group not yet (but methyl, benzyl,
etc.)

... and there are no disambiguation cues

Where do we get the structures?



For **compound names**:

~60% text mining from OSCAR (Corbett and Murray-Rust 2006, Batchelor and Corbett 2007)

~20% PubChem

~20% ChemDraw

For **compound numbers** (7, *cis*-8, 23b)

~70% author ChemDraw

~30% technical editors

InChI: cans and can'ts

Can tell us:

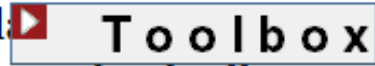
- What atoms are in a system.
- What the ligands are.
- How the non-metals are connected.
- What the geometry is around C, P, As, *etc.*

Can't tell us:

- What the geometry is around a metal.
- How the metals are connected (at least, not easily).
- Anything about polymers.

But that's enough to be going on with...

fertilisation (hpf), and come to lie in the midline. By 24 hpf their vacuoles are expanding and straightening of the trunk and tail is underway. The multiple functions of the notochord are reflected in the abundance and diversity of relevant mutants identified in large scale mutagenesis screens.^{13,14} Thus, mutations affecting specification of early notochord cells have profound defects on body patterning.¹⁵⁻¹⁷ Another larger class of mutants affect notochord differentiation and structural roles, thus *bashful*, *grumpy*, *sneezy* and *happy* are all very shortened and have unvacuolated notochord cells. Molecular characterisation of these and related loci reveal that they are defective in either secretion or notochord sheath components.¹⁸⁻²⁰ While these mutants have been characterised molecularly, another class affecting notochord morphogenesis remains uncloned. Amongst these are *leviathan*, *crash test dummy*, *gulliver* and *quasimodo*^{13,14} which, in contrast to *sleepy* and related mutants, display well-differentiated notochord cells, but with a violently distorted arrangement in both dorsomedial and mediolateral planes. Importantly, whilst *gulliver* and *quasimodo* mutants show additional characteristics, *leviathan* and *crash test dummy* show only a notochord defect. These latter mutants were isolated in separate screens and may well be allelic.



Animal classification

- those that belong to the Emperor,
- embalmed ones,
- those that are trained,
- suckling pigs,
- mermaids,
- fabulous ones,
- stray dogs,
- those included in the present classification,
- those that tremble as if they were mad,
- innumerable ones,
- those drawn with a very fine camelhair brush,
- others,
- those that have just broken a flower vase,
- those that from a long way off look like flies.

Classification problems

2,4,6-trinitrotoluene biosynthesis

“This term is obsolete as 2,4,6-trinitrotoluene is not synthesized by living organisms“

From the Gene Ontology



GoPubMed - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.gopubmed.com/

what

Top categories

- Biological Sciences [280]
 - Proteomics [31]
 - Genes [63]
 - Genomics [57]
 - Systems Biology [11]
 - Biology [90]
 - more
- Chemicals and Drugs [264]
 - Proteome [30]
 - Proteins [187]
 - DNA [58]
 - more
- biological_process [201]
 - cellularization [49]
 - signal transduction [50]
 - more
- Natural Sciences [245]
- Technology, Industry, Agriculture
- cellular_component [100]
- Techniques and Equipment [192]
- Anatomy [98]
- Organisms [118]

Search: "mol biosyst"[TA] [find it!](#)

294 articles

7: Small molecule enhancers of autophagy for neurodegenerative diseases.
 PMID: [18704227](#) [Related Articles](#)
 Sarkar S et al., *Mol Biosyst*, 4 (9): 895-901, 2008
 Neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, prion diseases and polyglutamine disorders, including Huntington's disease and various spinocerebellar ataxias, are associated with the formation of protein aggregates.

8: Protein kinase CK2 as a druggable target.
 PMID: [18704226](#) [Related Articles](#)
 Sarno S et al., *Mol Biosyst*, 4 (9): 889-94, 2008
 CK2 is probably the most pleiotropic Ser/Thr protein kinase with hundreds of endogenous substrates already known, which are implicated in a variety of cellular functions.

Transferring dat... Open Notebook 11 Now: Partly Sunny, 61° F Tue: 54° F Wed: 51° F

Making ontologies

“T.O.A.S.T.”

- Tiny ontologies all strung together
- RXNO
- Molecular processes



RXNO

the name reaction ontology

- Every chemist knows about famous chemists like Wittig, Cannizzaro, Diels, Alder, benzoin
- They're pretty unambiguous and well-suited to logical definitions
- But what organizing principle do we use?

RXNO

the name reaction ontology

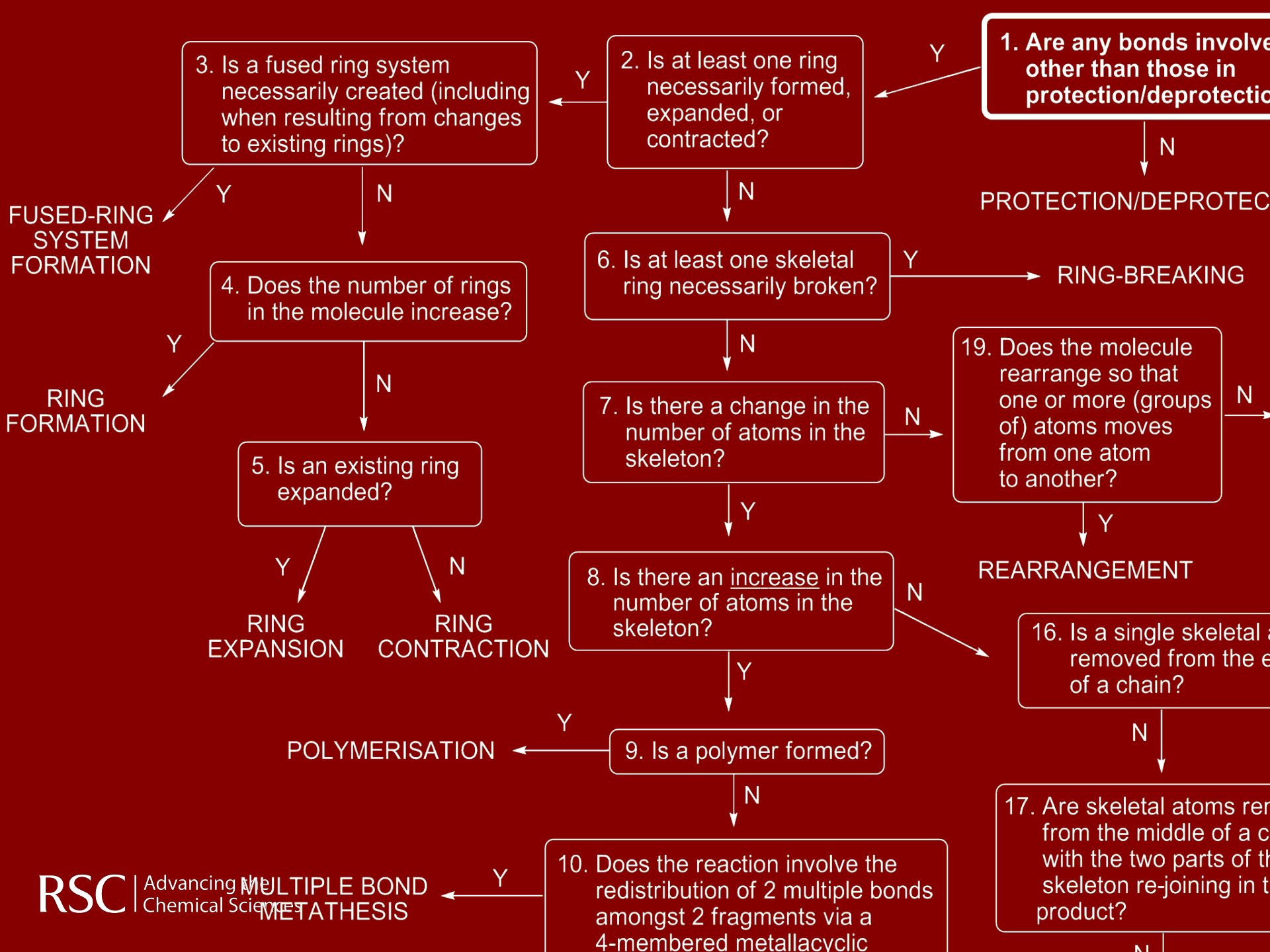
- Sort reactions by what they do to the ‘skeleton’ of the molecule.
- Skeleton-changing reactions:
 - Joinings, cleavings, rearrangements, ring formation, ring expansion
- Skeleton-preserving reactions:
 - Additions, eliminations, substitutions, protections, deprotections

RXNO

the name reaction ontology

- Quality? Subjectivity?
- Get our curators to assign reactions to categories check percentage agreement, discuss disagreements, improve guidelines, iterate to convergence.

Will be open ontology, curated by RSC. Soon.



- Classes
 - reaction
 - cleaving reaction
 - functional modification
 - addition reaction
 - deprotection reaction
 - elimination reaction
 - functional group modification
 - protection reaction
 - substitution reaction
 - allylic rearrangement
 - Arbuzov reaction
 - aromatic substitution
 - Balz-Schiemann reaction
 - Boylard-Sims oxidation
 - Buchwald-Hartwig amination
 - Elbs persulfate oxidation, phenol
 - Gattermann halogenation**
 - Goldberg reaction
 - Koerner-Contardi reaction
 - Sandmeyer reaction
 - Ullmann condensation
 - Fischer-Speier esterification
 - Fleming-Tamao oxidation
 - Gabriel synthesis
 - Ing-Manske procedure
 - Katritzky pyrylium-pyridinium method
 - Kharasch-Sosnovsky reaction
 - Michaelis-Arbuzov reaction
 - Mitsunobu reaction
 - Nicholas reaction, nucleophilic heteroa
 - Petasis reaction
 - Schotten-Baumann reaction
 - Tamao oxidation
 - Williamson ether synthesis
 - Wohl-Ziegler bromination
 - joining reaction
 - rearrangement
 - ring breaking
 - ring contraction
 - ring expansion
 - ring formation reaction

Search & Filter Parent Plugin

Term filter Advanced Options 5 results

ID	Name
RXNO:0000022	Gattermann halogenation
RXNO:0000082	Goldberg reaction
RXNO:0000178	Kharasch-Sosnovsky reaction
RXNO:0000081	Ullmann condensation
RXNO:0000040	Ullmann reaction

Autoselect Select terms Results label 5 results

Ignore obsoletes Search all Search children of selection Search ontology of selection

ID RXNO:0000022

Namespace RXNO

Name Gattermann halogenation

Definition * Comment * Cross Products

Definition

An aromatic substitution reaction where an aryl diazonium salt reacts with copper powder and hydrochloric acid or hydrobromic acid to form an aromatic chloride or bromide.

Dbxrefs

RSC:cc

Edit

Dbxrefs Synonyms * Categories

Gattermann reaction

Select a dbxref from the list to edit it, or press add to create a new dbxref

Add Del



Molecular process ontology

- But what about methylations, dihydroxylations and so forth that don't have special reagents and aren't named after a 19th century notable?
- Create a “cross-product” ontology by fitting content from an already existing ontology into a template.

Molecular process ontology

Template for <group>ation:

“The replacement of a hydrogen atom in a molecule with a <group>.”

Template for de<group>ation:

“The replacement of a <group> substituent in a molecule with a hydrogen atom.”

Template for *N*-<group>ation:

“The replacement of a hydrogen atom bound to a nitrogen atom with a <group> substituent.”

...and so forth.

Molecular process ontology

Group terms in ChEBI:

531 -yl groups

143 -o groups

45 -ene groups

32 -oxy groups

11 -yne groups

Total: 762

Patterns:

$((N|O|P|S)-)?$

$(de?)<group>ation$

Ten kinds of molecular process per group.

Equals...

Molecular process ontology

A more-or-less free ontology, inheriting the internal links from ChEBI, with

7620 terms.

Will be open ontology, curated by RSC. Soon.

More creative criticism...

Against Idiosyncrasy in Ontology

L
Univers

Abstract. The world of ontology development is full of mysteries. Recently, ISO Standard 15926 (“Lifecycle Integration of Process Plant Data Including Oil and Gas Production Facilities”), a data model initially designed to support the integration and handover of large engineering artefacts, has been proposed by its principal custodian for general use as an upper level ontology. As we shall discover, ISO 15926 is, when examined in light of this proposal, marked by a series of quite astonishing defects, which may however provide general lessons for the developers of ontologies in the future.

3. Bennett and C. Fellbaum (Eds.), *Formal Ontology and Information Systems*, (FOIS 2006), Baltimore November 9-11 2006

Data

- Experimental data checker

- Validation
- Visualisation

2-Methyl-2-hydroxymethyl-2,5-dihydrofuran **9** (50 mg, 0.51 mmol) was oxidised using OsO₄, TMEDA to yield the crude products [6 : 1 (*syn* : *anti*) by ¹H NMR]. The resulting oil was purified by flash chromatography (petrol-EtOAc, 4 : 1) to afford the pure title compound as a colourless oil (116 mg, 83%); ν_{\max} (film)/cm⁻¹ 2977, 2929, 1746, 1374, 1230, 1046^[ref], δ_{H} (300 MHz, CDCl₃) 5.43 (1H, q, *J* 5), 5.17 (1H, d, *J* 5), 4.26-4.09 (3H, m), 3.89 (1H, dd, *J* 10 and 5), 2.10 (6H, s), 2.08 (3H, s), 1.36 (3H, s)^[ref]; δ_{C} (75 MHz, CDCl₃) 170.6, 169.8, 169.5, 80.8, 76.2, 71.2, 68.9, 65.3, 22.3, 20.6, 20.4, 18.3^[ref]; Found (CI) 292.1392, C₁₂H₂₂NO₇ + NH₄ requires 292.1396.

anti-2-Methyl-2-(acetoxymethyl)tetrahydrofuran *anti*-15

2-Methyl-2-hydroxymethyl-2,5-dihydrofuran **9** (50 mg, 0.51 mmol) was oxidised using UpJohn conditions to yield the crude products [2 : 1 (*anti* : *syn*) by ¹H NMR]. The resulting oil was purified by flash chromatography (petrol-EtOAc, 4 : 1) to afford the product mixture as a colourless oil (110 mg, 79%) as an inseparable mixture of isomers (*anti* major compound); ν_{\max} (film)/cm⁻¹ 2989, 2944, 1746, 1379, 1232, 1046^[ref], δ_{H} (300 MHz, CDCl₃) 5.46-5.39 (1H, m), 5.26 (1H, d, *J* 6), 4.26-3.85 (4H, m), 2.11 (3H, s), 2.10 (6H, s), 1.26 (3H, s)^[ref]; δ_{C} (75 MHz, CDCl₃) 170.4, 169.8, 169.6, 81.2, 72.7, 71.9, 69.3, 67.9, 20.7, 20.5, 20.4, 18.3^[ref]; Found (CI) 292.1400, C₁₂H₂₂NO₇ requires 292.1396.

syn-1,2,4-Triacetoxycyclohexane *syn*-16

Cyclohex-3-en-1-ol **11** (50 mg, 0.51 mmol) was oxidised using OsO₄, TMEDA to yield the crude mixture of products [3 : 1 (*syn* : *anti*) by ¹H NMR]. The resulting oil was purified by flash chromatography (EtOAc) to afford a colourless oil (121 mg, 92%) as an inseparable mixture of isomers; mixture: ν_{\max} (film)/cm⁻¹ 2953, 1745, 1370, 1235, 1028^[ref], *syn*-**16** δ_{H} (300 MHz, CDCl₃) 5.17-5.11 (1H, m), 4.84-4.72 (2H, m), 2.04 (3H, s), 1.98 (3H, s), 1.86 (3H, s), 2.05-1.45 (6H, m); δ_{C} (75 MHz, CDCl₃) 170.1, 170.0, 169.9, 69.5, 69.3, 67.9, 31.4, 25.3, 24.7, 21.1, 20.9, 20.8^[ref]; *anti*-**16**; δ_{H} (300 MHz, CDCl₃) 5.08-5.01 (1H, m), 4.81-4.67 (2H, m), 1.99 (9H, s), 2.05-1.45 (6H, m)^[ref]; δ_{C} (75 MHz, CDCl₃) 170.3, 170.2, 170.1, 69.7, 68.8, 68.6, 31.8, 25.9, 25.3, 23.6, 21.1, 20.9^[ref]; mixture Found (CI) 276.1437, C₁₂H₂₂NO₆ requires 276.1447.

anti-1,2,4-Triacetoxycyclohexane *anti*-16

Cyclohex-3-en-1-ol **11** (50 mg, 0.51 mmol) was oxidised using UpJohn conditions to yield the crude mixture of products [1 : 1 (*syn* : *anti*) by ¹H NMR]. The resulting oil was purified by flash chromatography (EtOAc) to afford the product mixture as a colourless oil (115 mg, 88%) as an inseparable mixture of isomers; All data are consistent with *syn*- and *anti*-**16** prepared previously.

syn-1,2,4,5-Tetraacetoxycyclohexane *syn*-17

Future development

- www.sciborg.org.uk
- RSC
- Researchers
- Publishers



SciBorg - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Mail Print Wordpad

Address <http://www.sciborg.org.uk/> Go

Extracting the Science from Scientific Publications

Overview People Publications

Overview

SciBorg is a four-year project, starting October 1 2005, funded by the [EPSRC](#) under the programme for [Computer Science for e-Science](#). The project is a collaboration between three groups at the University of Cambridge:

- [The Natural Language and Information Processing group in the Computer Laboratory](#)
- [The Unilever Centre for Molecular Informatics in the Chemistry Department](#)
- [Cambridge eScience Centre](#)

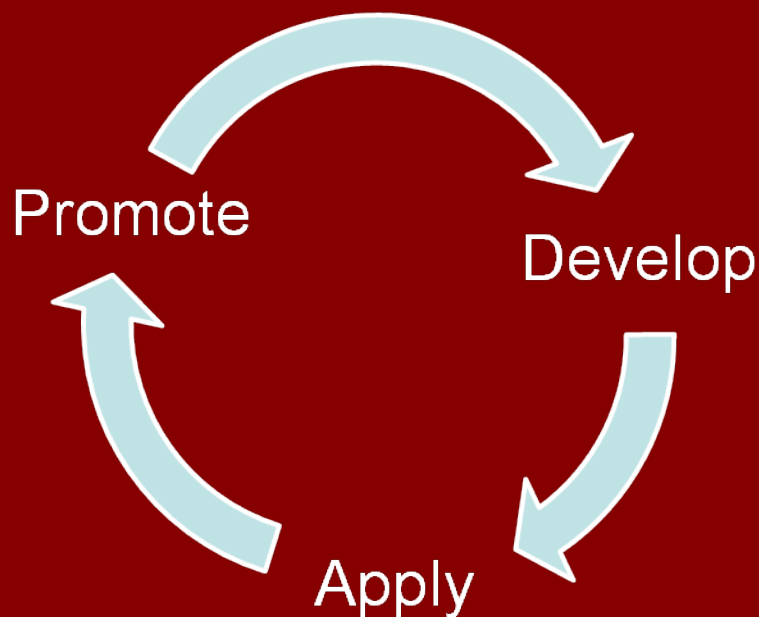
We are cooperating with three major publishers:

- [Royal Society of Chemistry](#)
- [Nature Publishing Group](#)
- [International Union of Crystallography](#)

The project summary and objectives are below. For further information, please see the [detailed project description](#), which was based on the project proposal, and the developing [SciBorg project wiki pages](#).

Internet

What we will do



- InChI implementation
- Ontology development
- Follow the science

Chemical Semantic Web



RSC Publishing



Richard Kidd
kiddr@rsc.org

www.projectprospect.org

Chemical Semantic Web



RSC Publishing



metabolite profiles indicating that epigenetic modifiers impacted fungi in a manner that was functionally distinct from that of a general cytotoxic response. Interestingly, a combination treatment, composed of a DNA methyltransferase inhibitor and histone deacetylase inhibitor, was tested and determined to be only modestly effective due to significant growth restriction and/or generation of metabolite profiles dominated by the effects of a single component in the mixture (e.g. effects of individual compound treatments were not additive as a mixture).

In order to probe the nature of metabolic remodeling induced by epigenetic modifier treatment, two fungi were selected for scale-up studies. The first isolate was obtained from a tidal pool along the coastline of Casco Bay, Portland, Maine, USA, and identified as *Cladosporium cladosporioides* by analysis of a 300 base pair sequence of the D2 region of the 26S large ribosomal subunit rRNA gene and morphological considerations.† This culture exhibited divergent responses to the small-molecule epigenetic modifiers 5-azacytidine and suberoylanilide hydroxamic acid which led to dramatic restructuring of its secondary metabolome with both treatments (Fig. 1A, C). Treatment of *C. cladosporioides* with 5-azacytidine elicited the *de novo* production of several oxylipins, three of which were characterized by NMR and MS analysis as 9Z,12Z)-11-hydroxyoctadeca-9,12-dienoic acid (1), its methyl ester (2), and glycerol conjugate (3) in substantial yields (Fig. 1B). The production of these compounds is of considerable interest since these types of metabolites are widely recognized for their important roles as intra- and inter-species cell signaling molecules. In contrast, suberoylanilide hydroxamic acid induced the production of a complex series of perylenequinones, two of which were characterized as new metabolites, cladochromes F (4) and G (5), along with four known cladochromes A (6), B (7), D (8), and E (9) and calphostin B (10) (Fig. 1D, and see ESI†). This is remarkable since this is the first reported co-occurrence of such an extensive range of cladochromes-calphostin metabolites from a single source. Moreover, the identification of 6 and 7 under epigenetic stimulation is significant since these compounds were first unique products of *Cladosporium* infection of seedlings, yet could not be obtained from an axenic mono-culture fermentation.¹⁰ While it was unclear if 6 and 7 might be the products of mixed biosynthesis, their selective production in suberoylanilide hydroxamic acid treated *Cladosporium* suggests that their biogenesis is normally tightly regulated and is a yet undefined host-specific signaling event.

The second fungal isolate was obtained from the foregut of a fifth instar luna moth (*Actias luna*; Saturniidae) larva that was cultured on an exclusive diet of sweet gum (*Liquidambar styraciflua* L.; Hamamelidaceae) leaves. Initial characterization of the fungus by analysis of the 26S rRNA gene gave a homology to *Diatype disciformis*. While control cultures of this *Diatype* sp. were relatively void of any secondary metabolites, addition of 5-azacytidine triggered a significant change in the organism's metabolic profile (Fig. 2A), resulting in the production of two new polyketides, lunalinides A (11) and B (12) (Fig. 2B). It is interesting to note that one other non-epigenetic culture treatment, elicitation with *E. coli*,† resulted in the biosynthesis of 11 and 12 which were otherwise repressed under axenic culture conditions. The results suggest that their production is under specific control of a unique environmental cue.

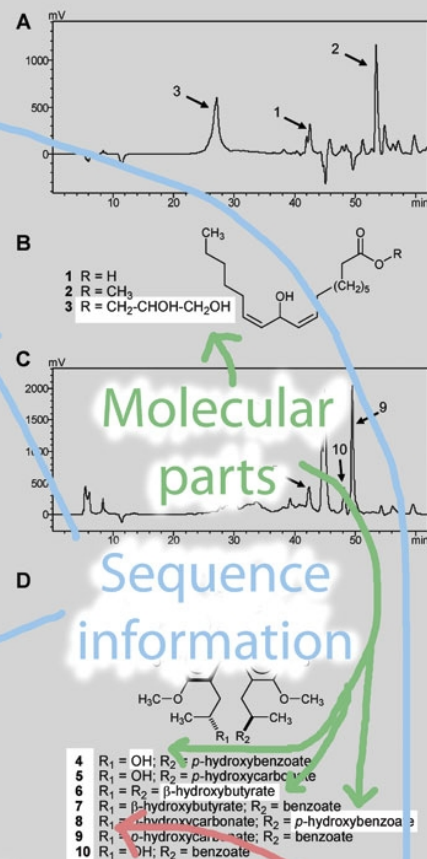


Fig. 1 Different chromatograms and structures of *C. cladosporioides* metabolites. HRESIMS chromatograms of total organic extract (TOE) from untreated and epigenetically treated (*de novo* at 210 nm) *C. cladosporioides* cultures. Difference chromatograms were generated by subtracting the chromatogram of the untreated control from the treatment groups. Peaks phasing upward represent metabolites expressed only upon epigenetic treatment or production at elevated concentrations. (A) Difference chromatogram of untreated and 5-azacytidine treated cultures yielding oxylipins 1–3 (B). (C) Difference chromatogram of untreated and suberoylanilide hydroxamic acid treated cultures yielding perylenequinones 4–10. (D) Chemical structures of compounds were characterized by HRESIMS. R₁ and R₂ are defined as in the legend. For comparison to values published in the literature, see ref. 10.

The success in identifying new natural products from fungi using epigenetic modifiers indicates that this technique is a very promising and rational approach for the native expression of silent biosynthetic pathways.

