UNIVERSITY OF CAMBRIDGE

Challenging Chemistry: Solving Molecular Problems

Jonathan Goodman

University of Cambridge

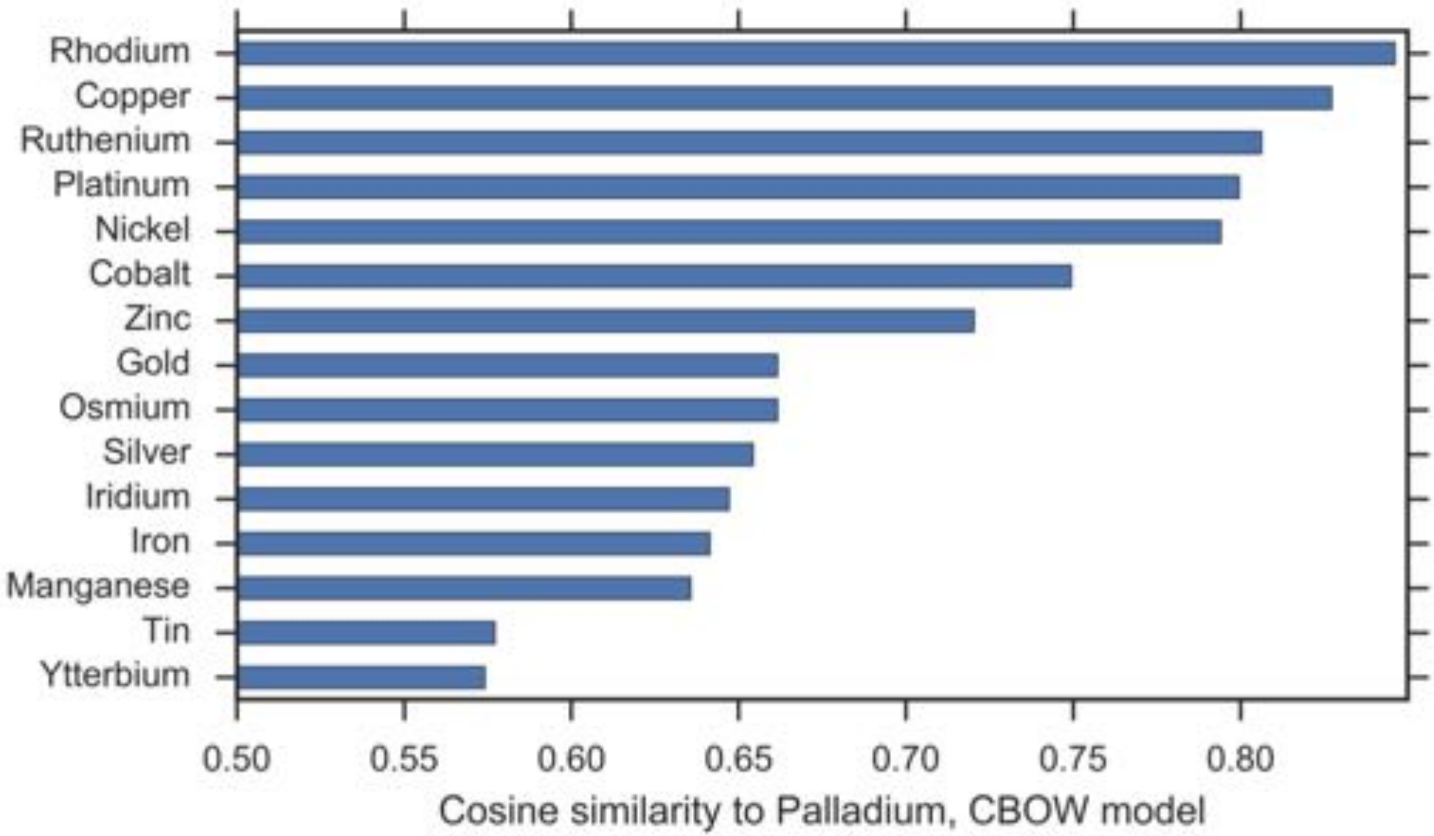
jmg11@cam.ac.uk

Challenging Chemistry: Solving Molecular Problems

- The power of AI is changing the way that chemistry is done, but chemistry is far from being solved.
- What is holding us back?
- Creating a new molecule or a necessary, but unknown, transformation is a demanding problem.
- Successes and surprises in predicting how molecules should behave can be used to try to discover the limits of our knowledge and the best ways to solve molecular problems.

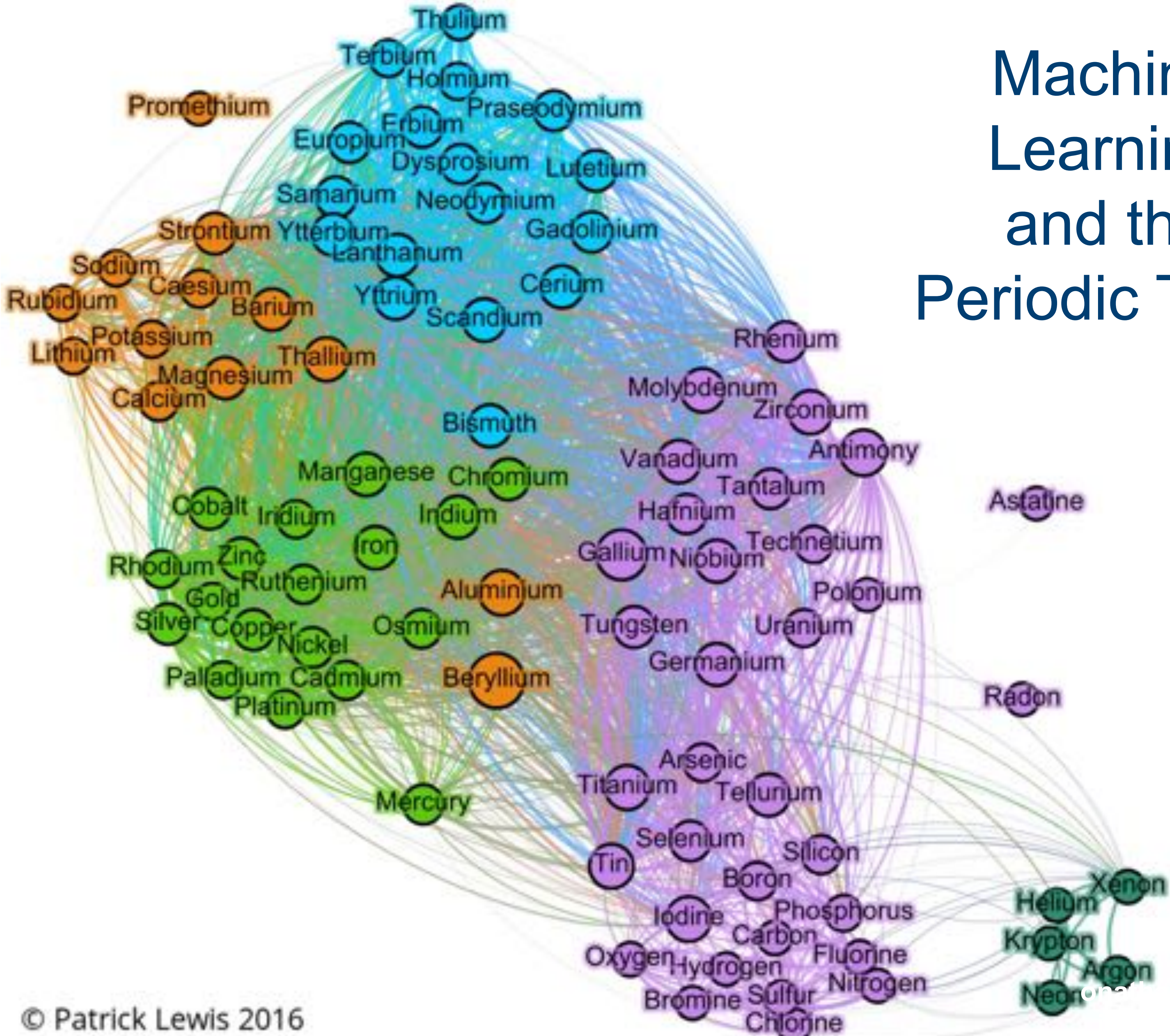
Challenging Chemistry: Solving Molecular Problems

Challenging Chemistry: Solving Molecular Problems



Patrick Lewis

Machine Learning and the Periodic Table



The Lenox Globe: *here be dragons*

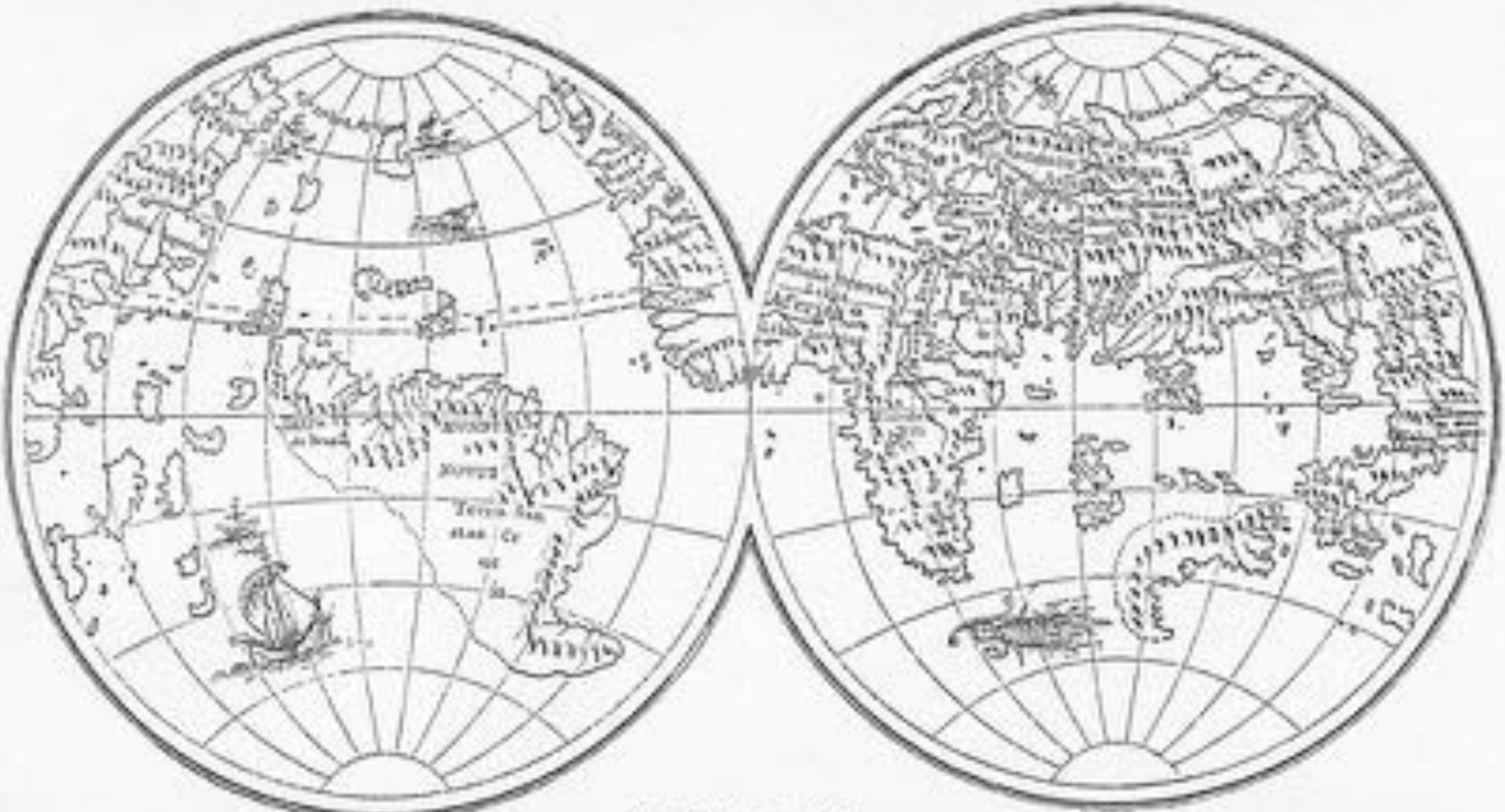
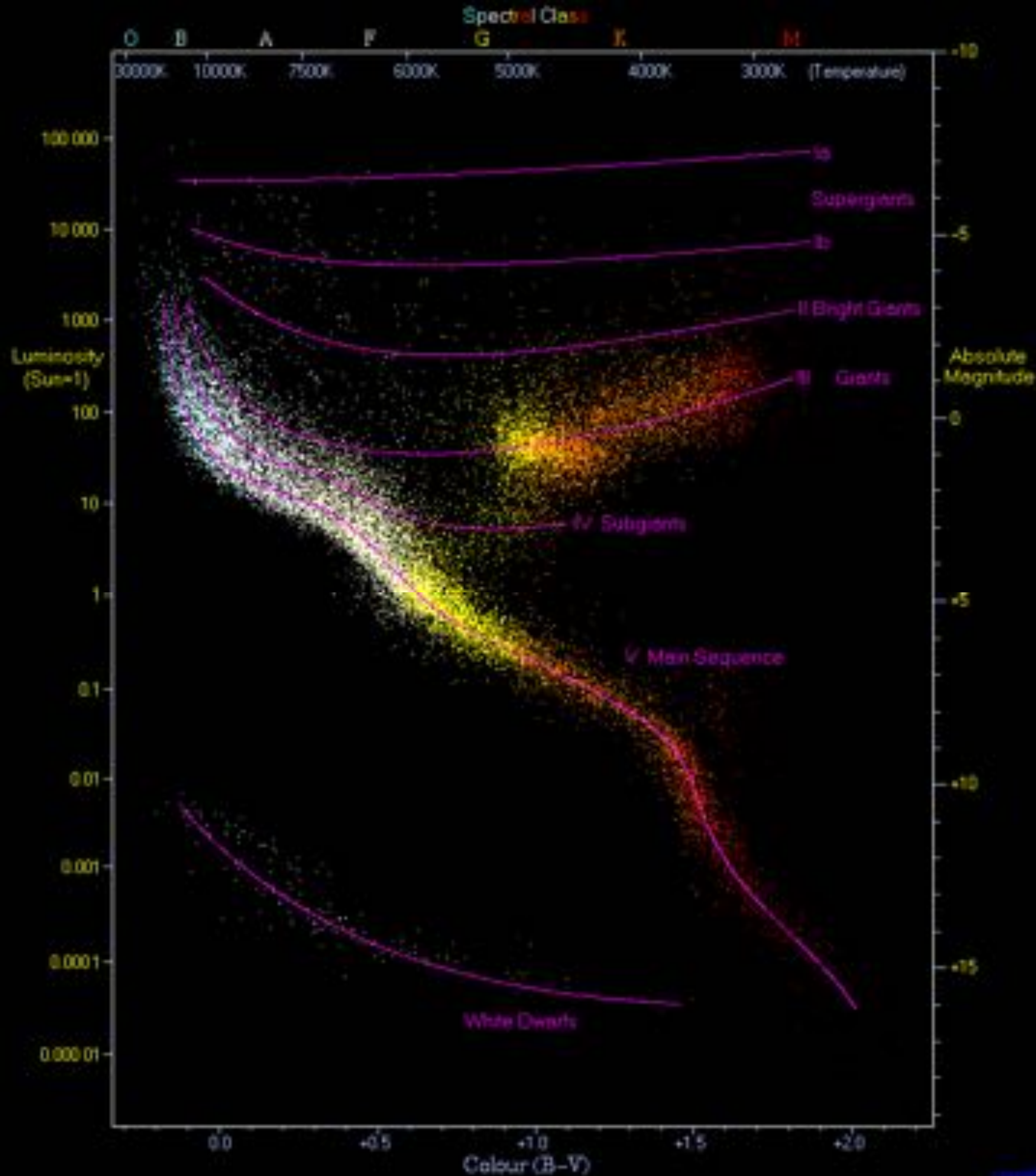


Fig. 2—Lenox Globe.



Hertzsprung-Russell diagram

22 000 stars

(from Wikipedia)

About 10^{21} stars

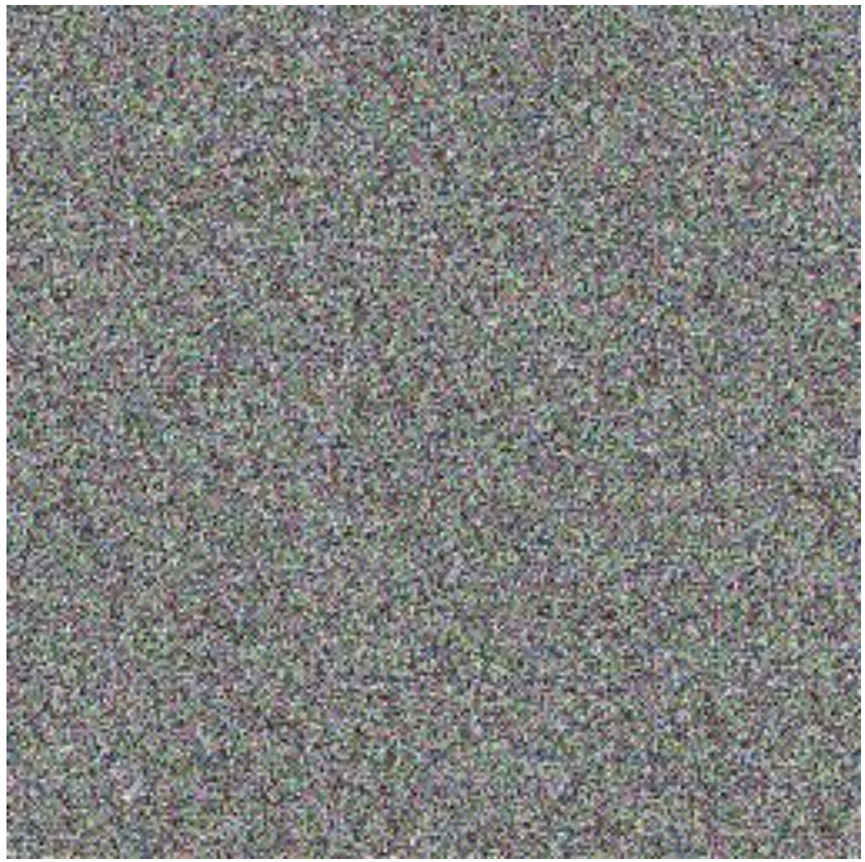
Challenging Chemistry: Solving Molecular Problems

- We know about roughly 10^8 molecules
- 10^200 small molecules possible

- ***Does every molecule we know about help us to explain 10^192 others?***

- Reactions: How many do we know? 10^7? How many possible?

How hard is chemistry?



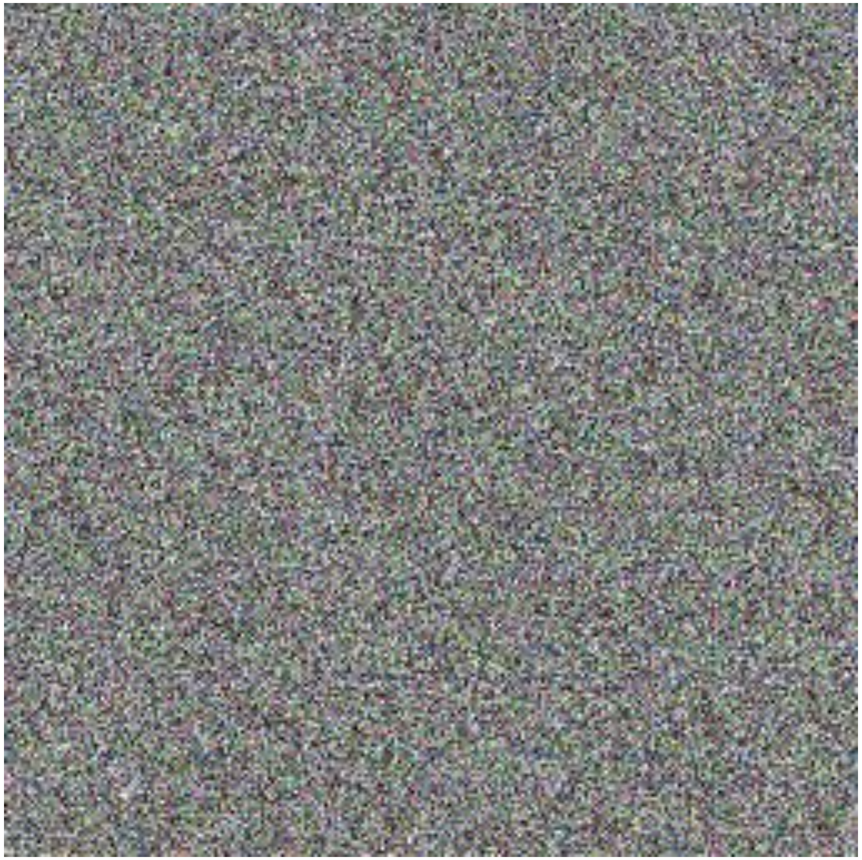
How hard is chemistry?

512 x 512
pixels

24 bits
per pixel

2^6291456

ten^two million



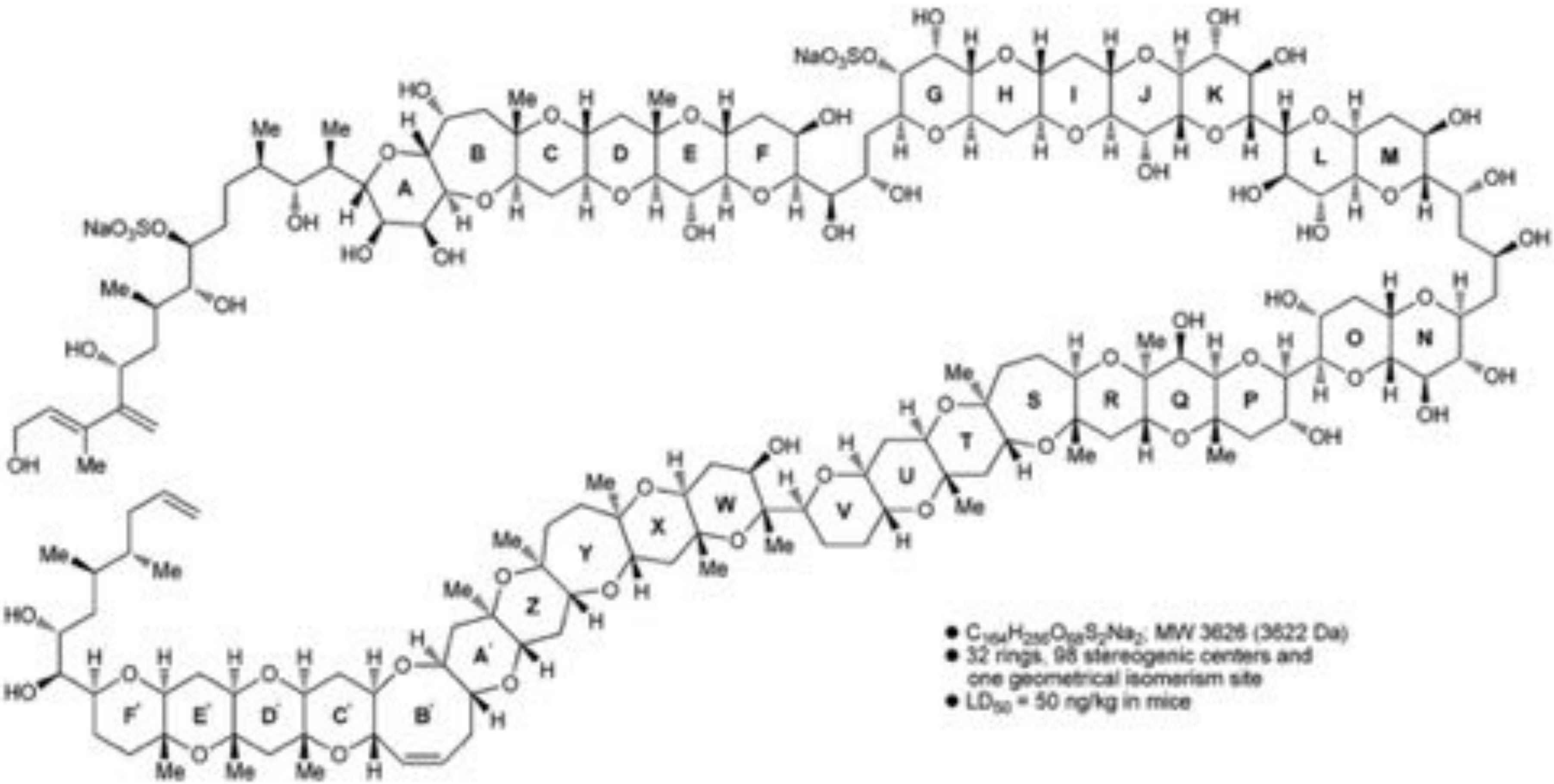
Chemistry

About 10^200
small molecules

Lots of reactions

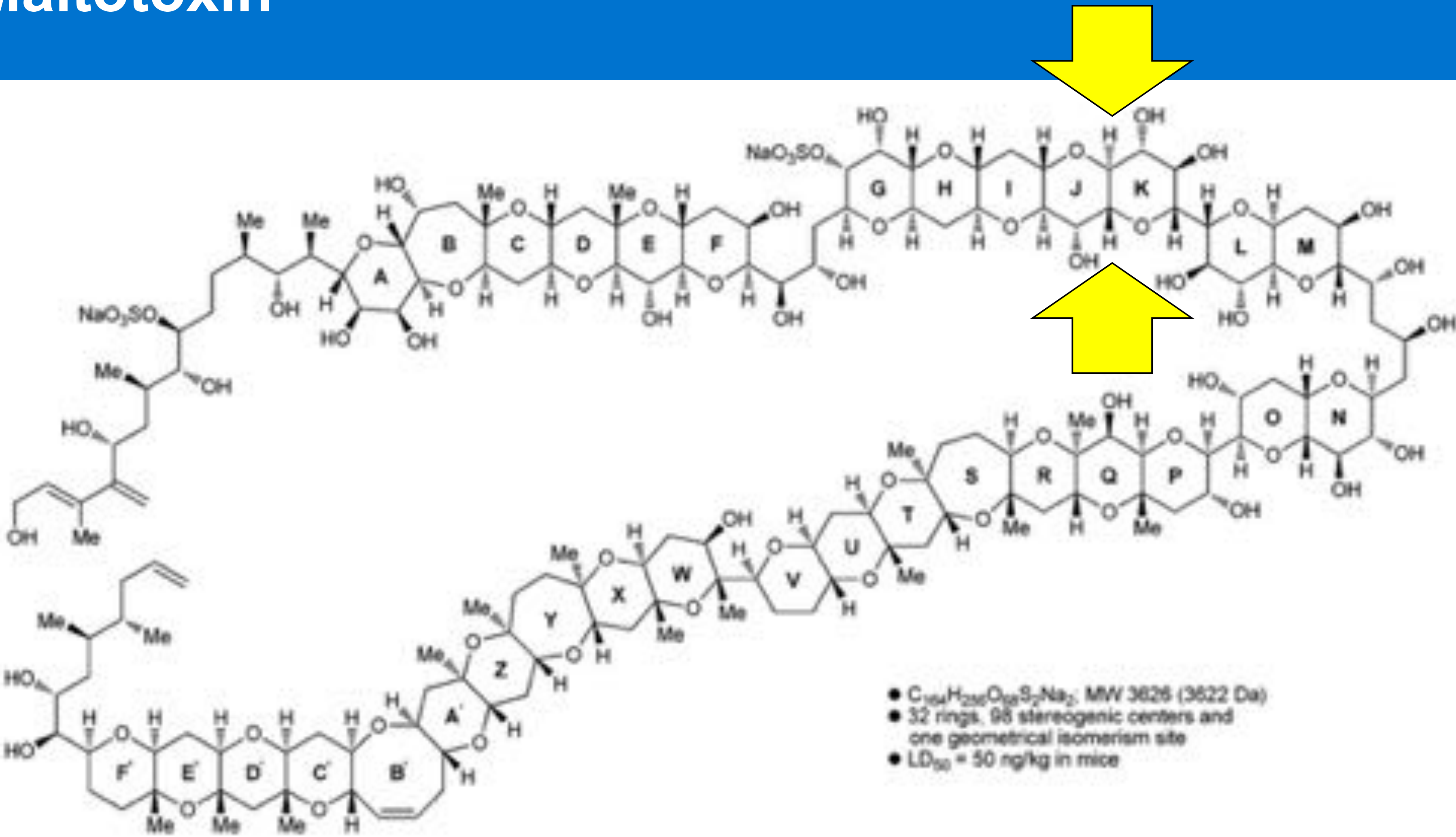
Every molecule
is interesting

Maitotoxin



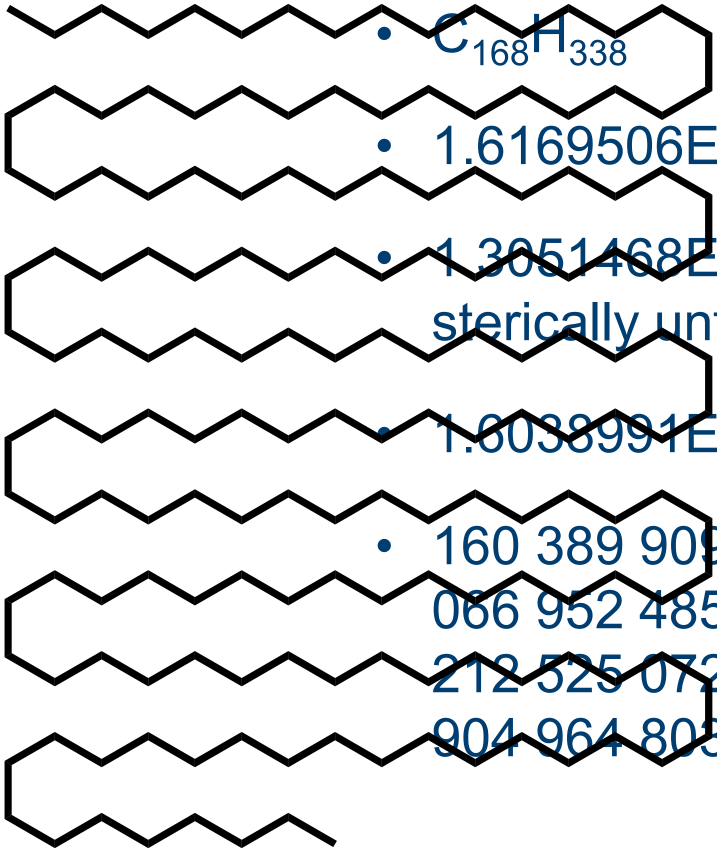
- $C_{164}H_{256}O_{68}S_2Na_2$; MW 3426 (3422 Da)
- 32 rings, 98 stereogenic centers and one geometrical isomerism site
- LD_{50} = 50 ng/kg in mice

Maitotoxin



- $C_{164}H_{256}O_{68}S_2Na_2$; MW 3626 (3622 Da)
- 32 rings, 98 stereogenic centers and one geometrical isomerism site
- LD_{50} = 50 ng/kg in mice

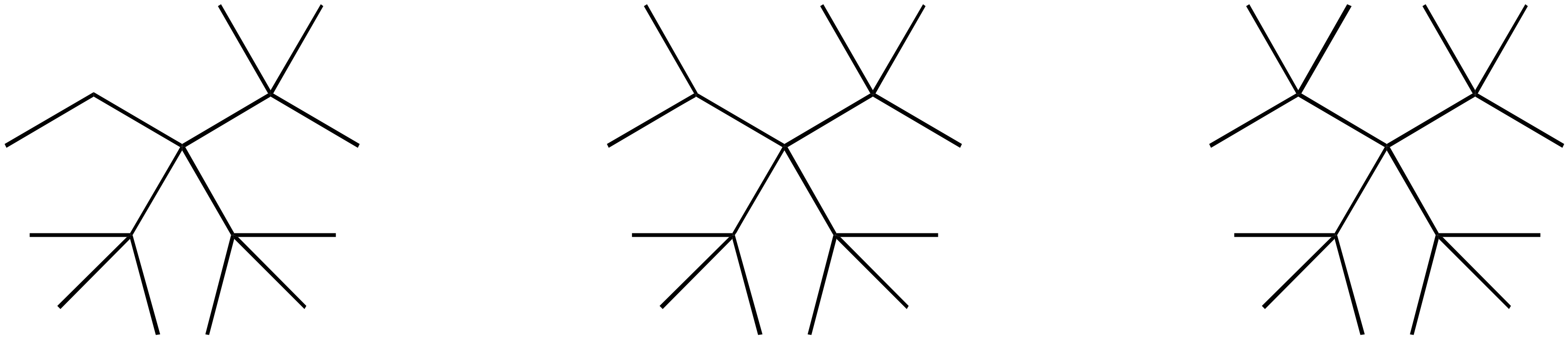
How hard is chemistry?

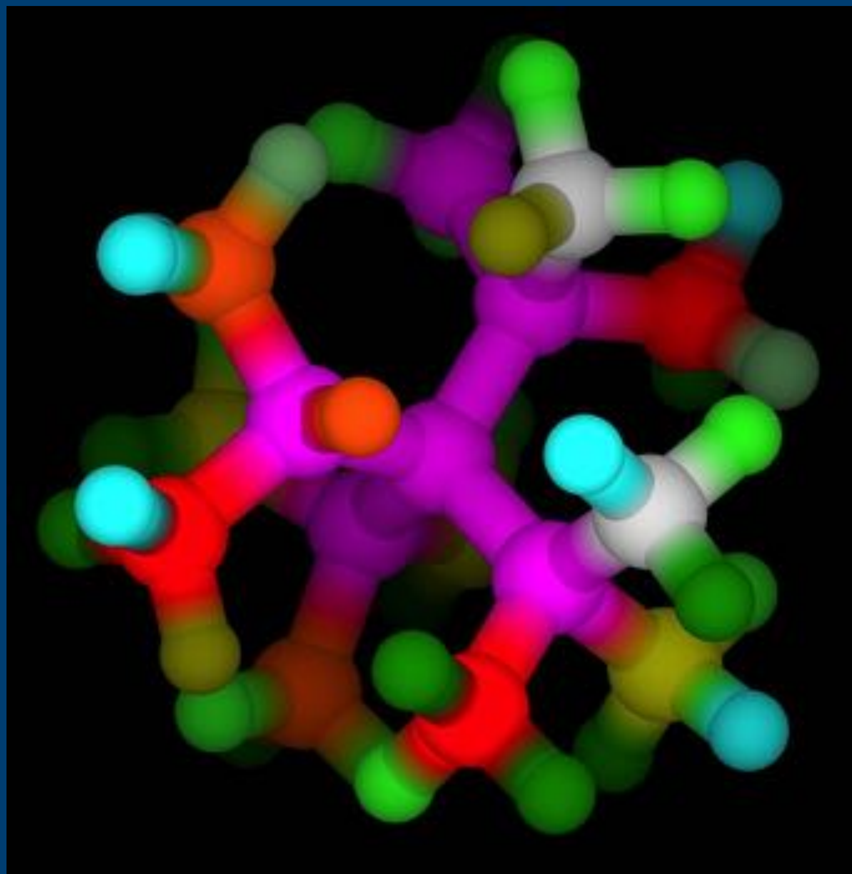


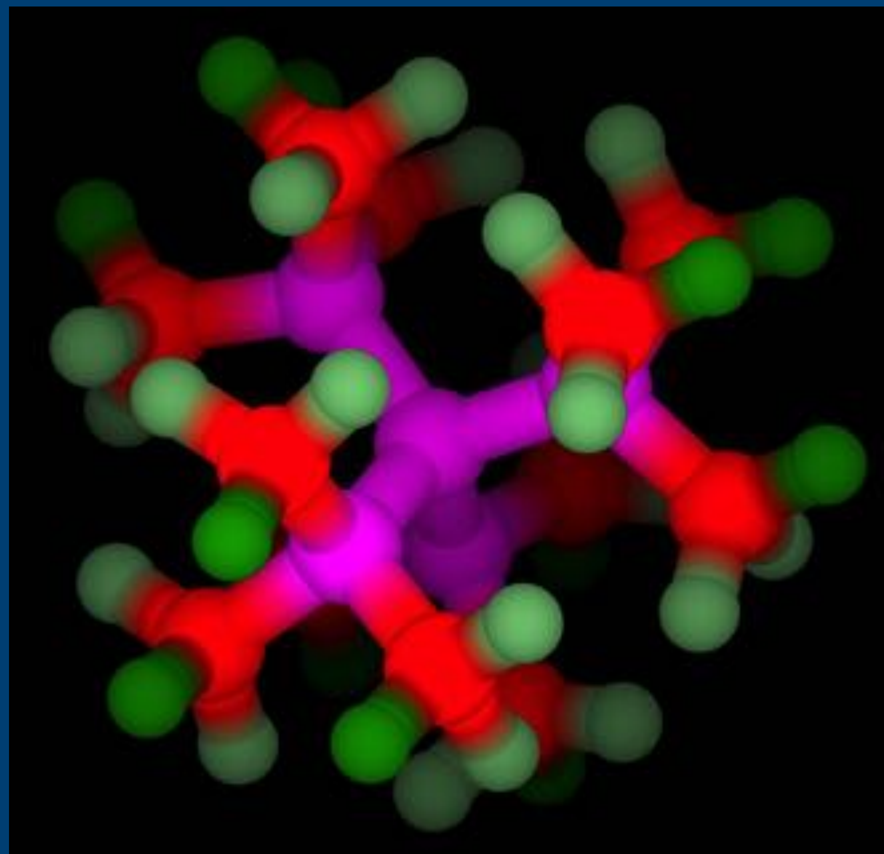
- $C_{168}H_{338}$
- 1.6169506E80 possible structures
- 1.3051468E78 of these are sterically unfavourable
- 1.6038991E80 allowed
- 160 389 909 826 333 051 035 066 952 485 856 171 917 377 212 525 072 646 363 560 301 904 964 803 211 461 164

What is the simplest molecule that cannot be made?

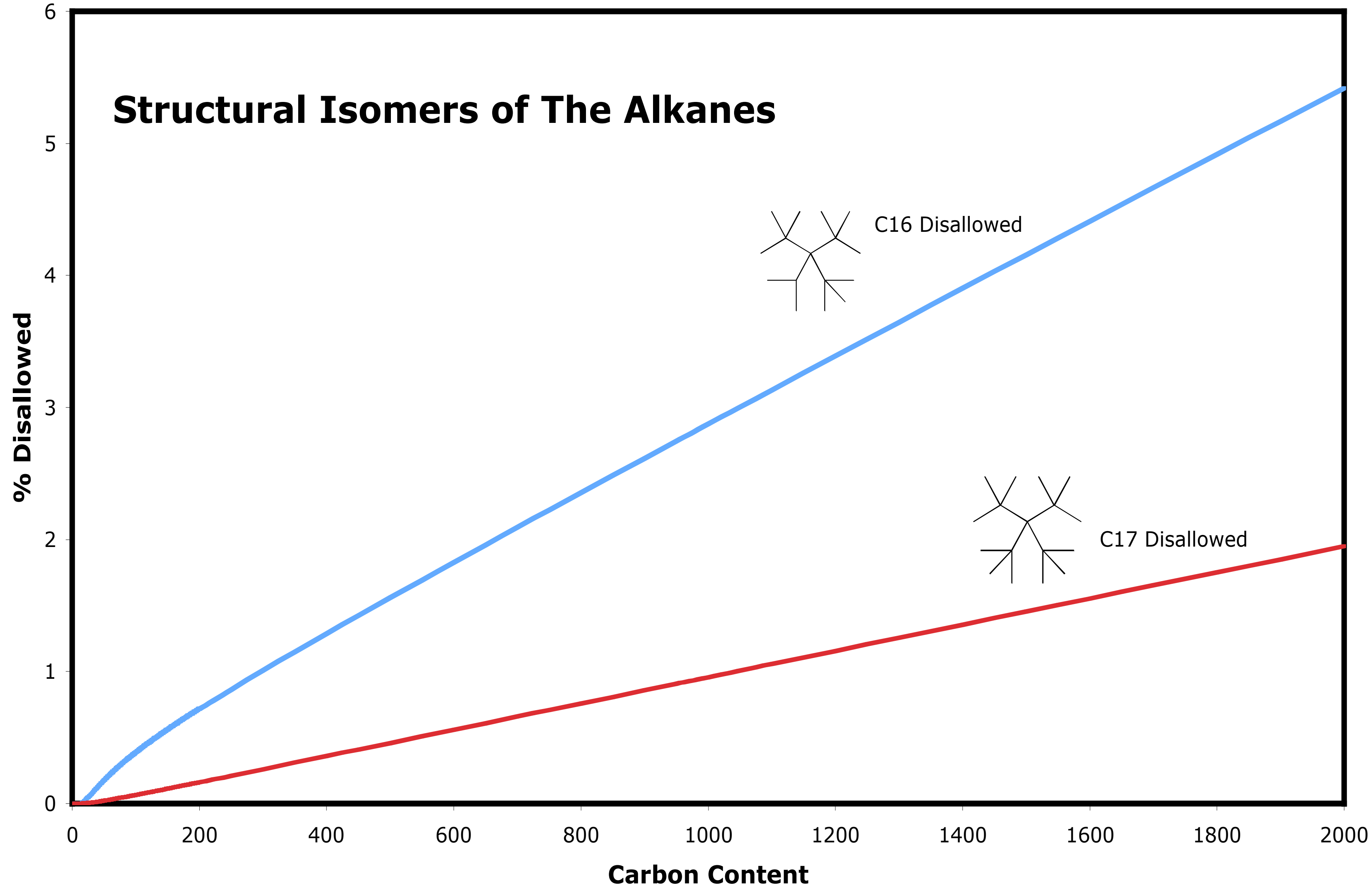
- Hydrocarbons only
- All valencies satisfied
- No rings
- Must exist under normal conditions



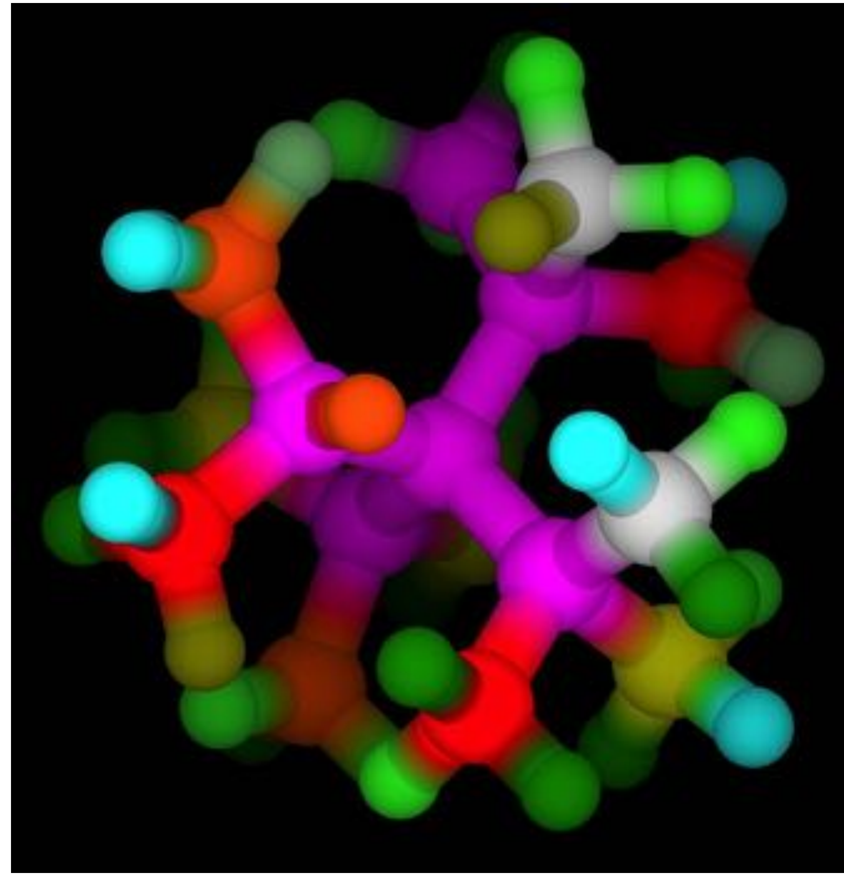


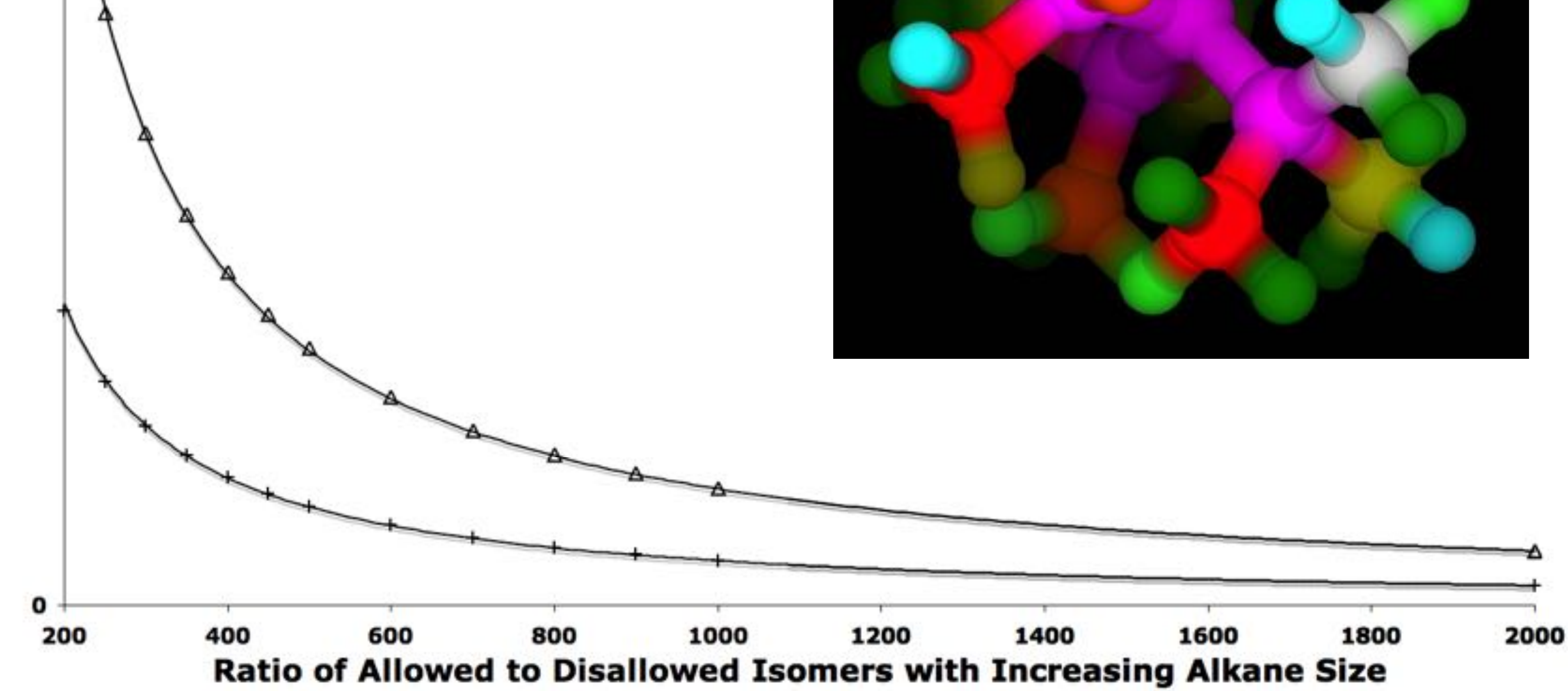


K. M. de Silva and J. M. Goodman *J. Chem. Inf. Model.* 2005, **45**, 81-87.

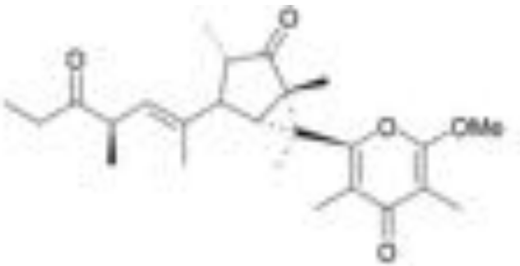

Structural Isomers of The Alkanes
C16 Disallowed
C17 Disallowed
% Disallowed
Carbon Content
0
1
2
3
4
5
6
0
200
400
600
800
1000
1200
1400
1600
1800
2000

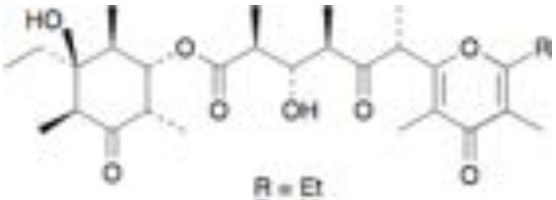
UNIVERSITY OF CAMBRIDGE



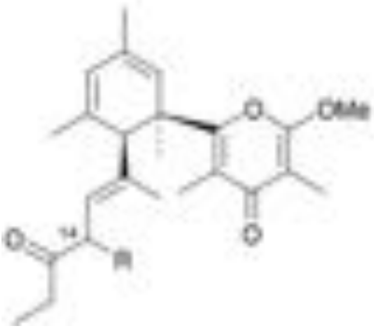


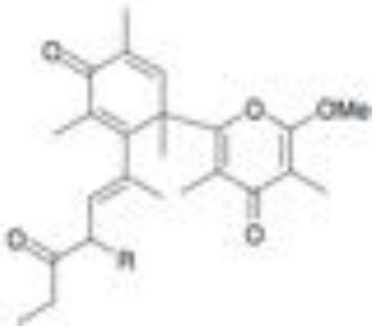
Challenging Chemistry: Solving Molecular Problems

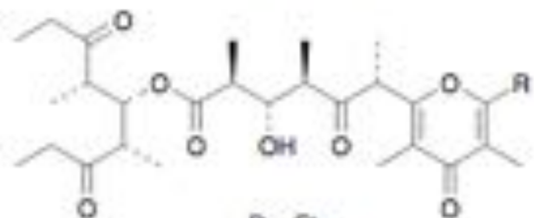




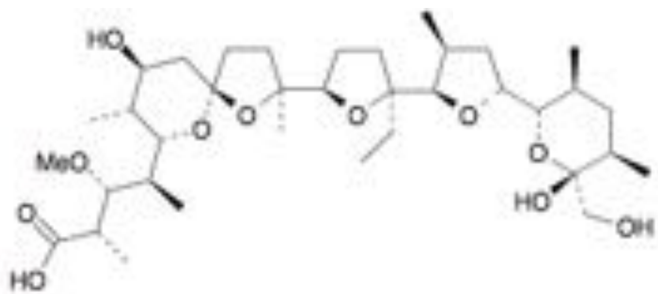
R = Et
R = Me

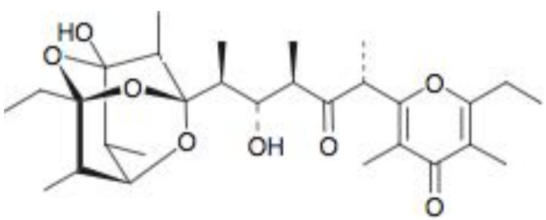


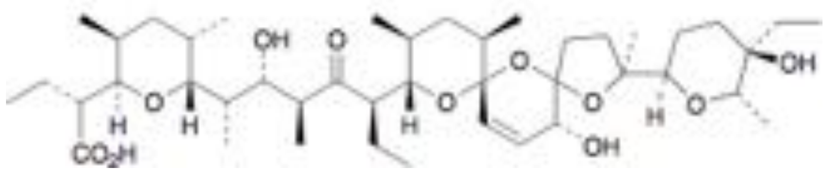




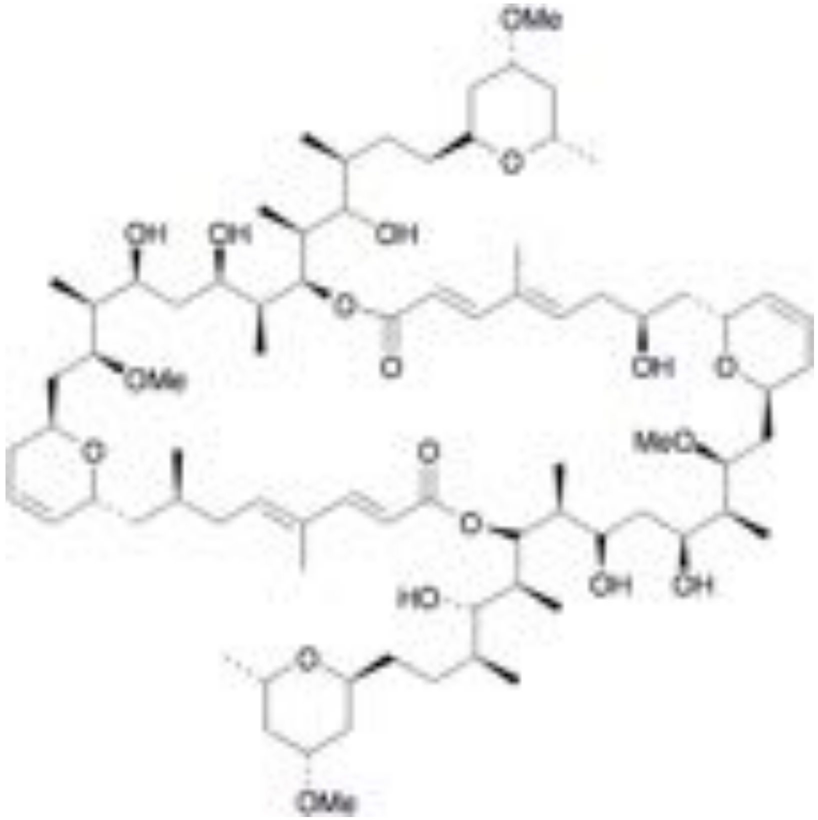
R = Et
R = Me

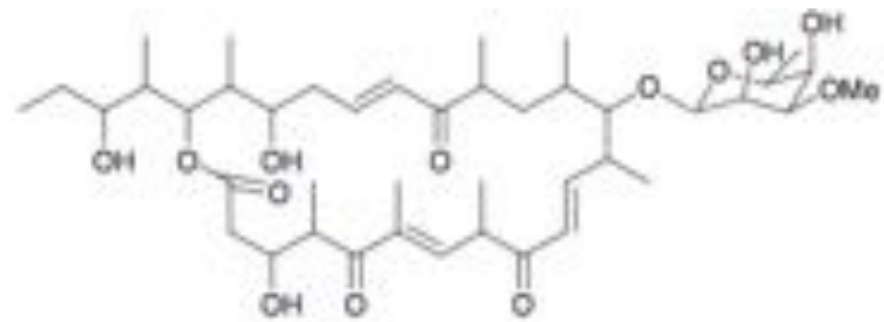


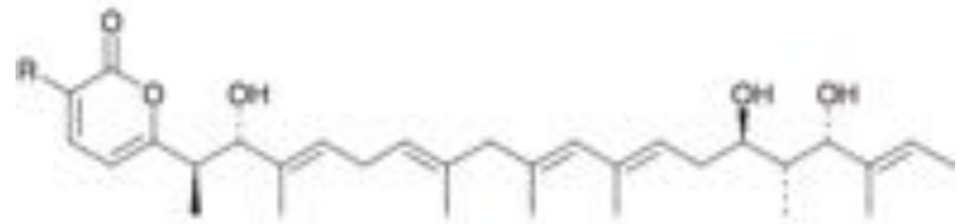




Challenging Chemistry: Solving Molecular Problems

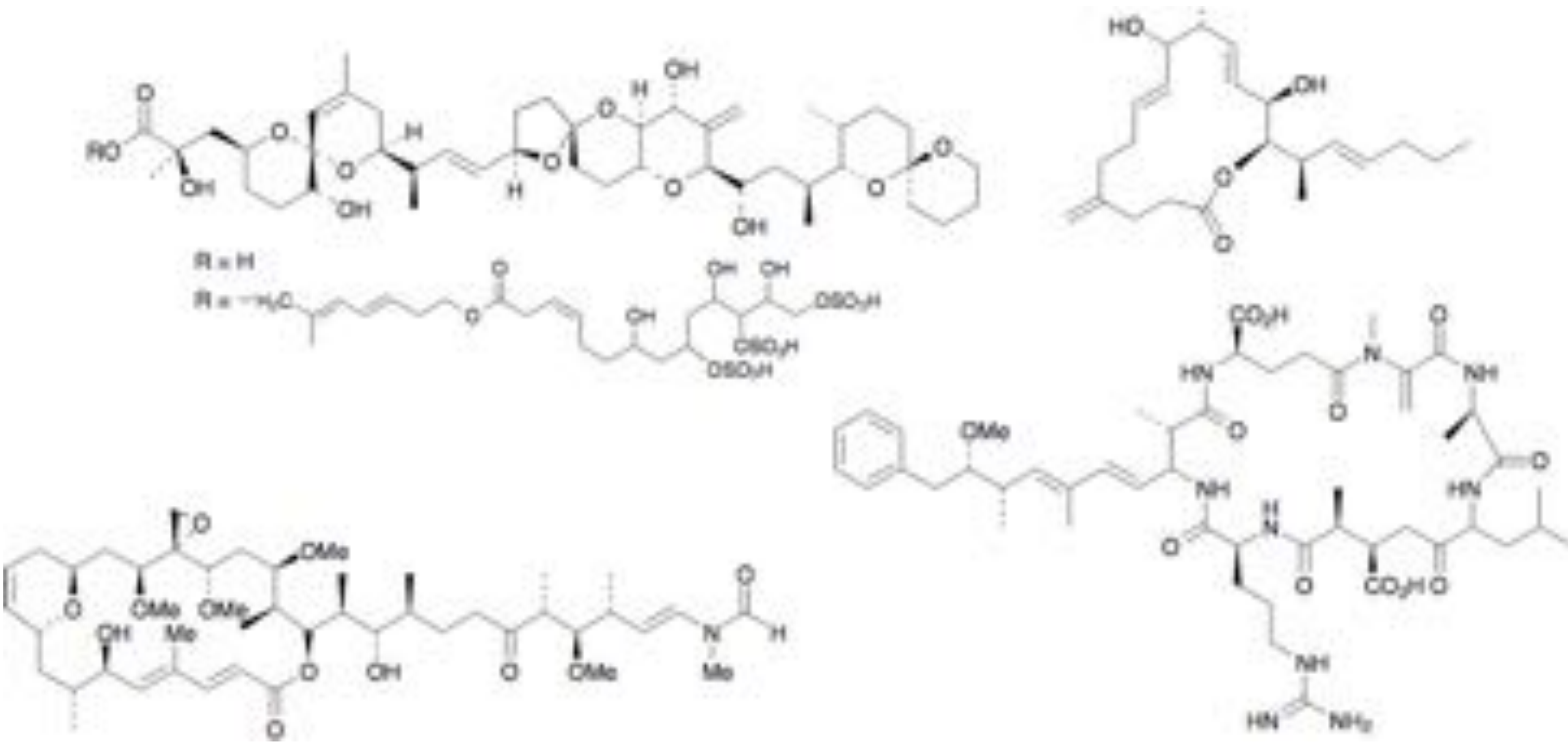




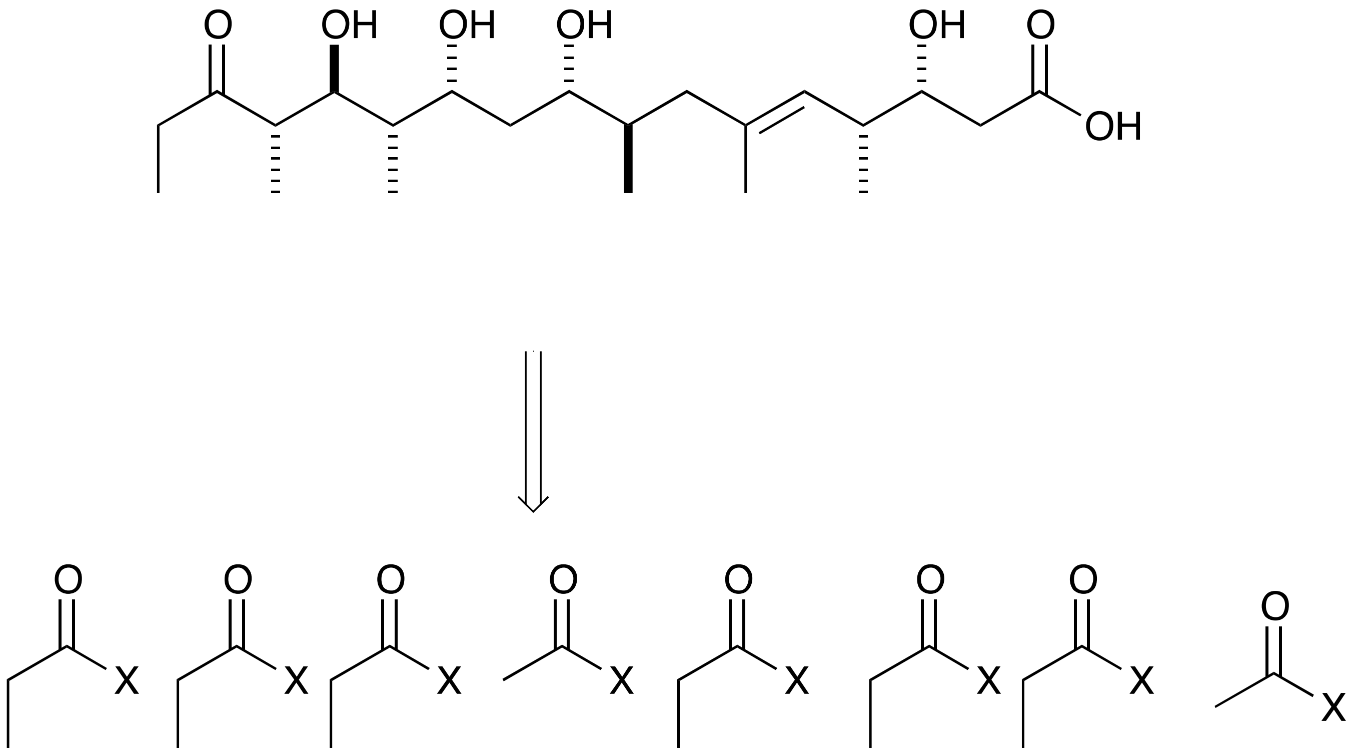


R = Me

Challenging Chemistry: Solving Molecular Problems



polypropionates



Open Access Publishing / Plan S

Some Chemical Data Providers

IUPAC

Effectively contributes to the worldwide understanding and application of the chemical sciences, to the betterment of humankind

ACS

To advance the broader chemistry enterprise and its practitioners for the benefit of Earth and its people

RSC

We're working to shape the future of the chemical sciences – for the benefit of science and humanity

Organic Syntheses, Coll. Vol. 1, p.314 (1941); Vol. 7, p.50 (1927).

HYDROGEN CYANIDE (ANHYDROUS)

[Hydrocyanic acid]

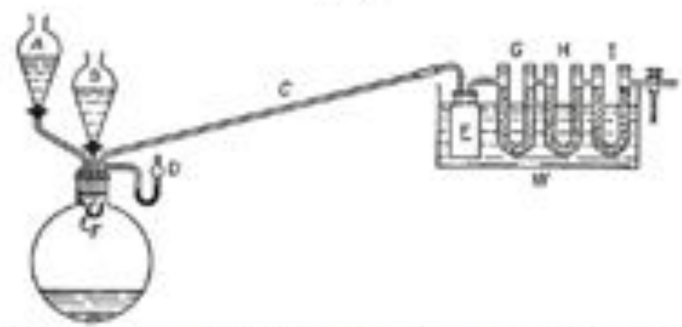
$$NaCN + H_2SO_4 \xrightarrow{H_2O} HCN + NaHSO_4$$

Submitted by K. Ziegler
Checked by Henry Gilman and L. C. Heckert.

1. Procedure

A 5-l. round-bottomed flask, set up in a good hood (Note 1), is fitted with a three-holed rubber stopper that holds two 250-cc. separatory funnels, A and B (Fig. 18). A small funnel F (about 3–4 cm. in diameter) is suspended directly under the outlets of the two separatory funnels and is attached to the rubber stopper by a loop of stiff copper wire. The discharge tube of the funnel is bent in the shape of a U so that its end is about 1 cm. below the top of the funnel. In the third hole of the rubber stopper is inserted an inclined glass tube, C, of about 10-mm. internal bore and approximately 50 cm. long. A mercury safety vent, D, is sealed in the side of this tube. C acts as an air condenser and leads to an empty gas bottle, E, of about 250-cc. capacity and then to three large U-tubes, G, H, and I, filled with anhydrous calcium chloride. The gas bottle and the U-tubes are contained in a water bath, W, warmed to 30–40°. The last calcium chloride tube is fitted with a three-way glass stopcock so that the gaseous anhydrous hydrogen cyanide may be used directly or diverted to an efficient condenser for liquefaction. The condenser is a glass coil of 4–5 mm. bore and about 50 cm. long that is surrounded by ice (Note 2) contained in a percolator arrangement like that described on p. 117.

Fig. 18.



One of the separatory funnels is filled with dilute sulfuric acid prepared by the careful addition of 392 g. (213 cc., 4 moles) of concentrated sulfuric acid to 830 cc. of water. The other separatory funnel is filled with a solution of 203 g. (4 moles) of commercial sodium cyanide (about 96 per cent) dissolved in sufficient water to make 500 cc. of solution. Evolution of hydrogen cyanide takes place on the simultaneous addition of the two solutions. Practically all the reaction occurs in the funnel, F, and the sodium bisulfate solution continuously drains into the flask so that fresh solutions are always present. The solution in the funnel remains clear as long as sufficient sulfuric acid is present. An excess of sodium cyanide colors the solution yellow and leads to the formation of a muddy brown precipitate (Note 3). By adjusting the flow of solutions the rate of evolution is easily controlled, and the preparation requires no attention beyond that involved in the occasional replenishment of the solutions in the separatory funnels. The last part of the hydrogen cyanide can be driven from the apparatus by boiling the bisulfate solution for a few minutes. The yield of acid melting at −15° to −14.5° is 100–105 g. (93–97 per cent of the theoretical amount) (Note 4) and (Note 5).

2. Notes

1. Gattermann[1] recommends that the operator smoke during the preparation, for he found that a trace of hydrogen cyanide is sufficient to give the tobacco smoke a highly characteristic flavor. This preliminary warning is useful in case of leaky apparatus or a faulty hood.
2. It is essential that the coil be cooled with ice only. A freezing mixture causes solidification of the hydrogen cyanide with consequent clogging of the apparatus.
3. If the reaction is slowed up considerably or interrupted so that the solution becomes cool, sodium bisulfate crystallizes both in the funnel and in the generating flask.
4. The hydrogen cyanide is best kept over anhydrous calcium chloride. In this way it remains clear and water-white for months; otherwise, it soon becomes yellow, owing to the formation of azulmic acid. Concentrated hydrochloric acid (two drops per 500 g. of hydrogen cyanide) has also been recommended as a stabilizer.[2]
5. If larger quantities of hydrogen cyanide are desired, the apparatus may be modified as suggested by Lindemann[3] by using a four-holed rubber stopper in the generating flask and fitting it with a siphon tube by means of which the sodium bisulfate solution can be removed from time to time. With this modification the generating flask need be of only 2- to 3-l. capacity.

Hydrogen cyanide may be prepared conveniently, probably in lower yields than in the procedure described, by adding the saturated sodium cyanide solution one-half inch below the surface of 50 per cent sulfuric acid contained in a flask. The residual amount is expelled by warming the flask on a steam bath. The apparatus is the same as that used for generating hydrogen chloride (p. 293) by passing aqueous hydrochloric acid into sulfuric acid (W. W. Hartman, private communication).

3. Discussion

Hydrogen cyanide can be prepared by the action of sulfuric acid on potassium ferrocyanide,[1] potassium cyanide[4] or sodium cyanide.[2] The procedure described is based on the methods of Ziegler[5] and Lindemann.[3]

This preparation is referenced from:

- Org. Syn. Coll. Vol. 6, 14
- Org. Syn. Coll. Vol. 6, 334

References and Notes

1. Gattermann, Ann. **357**, 318 (1907).
2. Slotta, Ber. **67**, 1030 (1934).
3. Lindemann, Ann. **431**, 291 (1923).
4. Villars, J. Am. Chem. Soc. **52**, 64 (1930).
5. Ziegler, Ber. **54**, 110 (1921); Lindemann, Ann. **431**, 290 (1923); Pauer, Monatsh. **58**, 1 (1931).

Appendix
Chemical Abstracts Nomenclature (Collective Index Number); (Registry Number)

azulmic acid

calcium chloride (10043-52-4)

Organic Syntheses, Coll. Vol. 1, p.314 (1941); Vol. 7, p.50 (1927).

HYDROGEN CYANIDE (ANHYDROUS)

[Hydrocyanic acid]

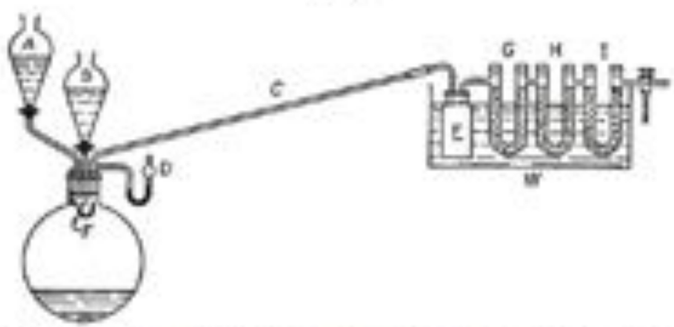
$$NaCN + H_2SO_4 \xrightarrow{H_2O} HCN + NaHSO_4$$

Submitted by K. Ziegler
Checked by Henry Gilman and L. C. Heckert.

1. Procedure

A 5-l. round-bottomed flask, set up in a good hood (Note 1), is fitted with a … stopper that holds two 250-cc. separatory funnels, A and B (Fig. 18). A small funnel F (… in diameter) is suspended directly under the outlets of the two separatory funnels and is a… the …

Fig. 18.



One of the separatory funnels is filled with dilute sulfuric acid prepared by the careful addition of 392 g. (213 cc., 4 moles) of concentrated sulfuric acid to 830 cc. of water. The other separatory funnel is filled with a solution of 203 g. (4 moles) of commercial sodium cyanide (about 96 per cent) dissolved in sufficient water to make 500 cc. of solution. Evolution of hydrogen cyanide takes place on the simultaneous addition of the two solutions. Practically all the reaction occurs in the funnel, F, and the sodium bisulfate solution continuously drains into the flask so that fresh solutions are always present. The solution in the funnel remains clear as long as sufficient sulfuric acid is present. An excess of sodium cyanide colors the solution yellow and leads to the formation of a muddy brown precipitate (Note 3). By adjusting the flow of solutions the rate of evolution is easily controlled, and the preparation requires no attention beyond that involved in the occasional replenishment of the solutions in the separatory funnels. The last part of the hydrogen cyanide can be driven from the apparatus by boiling the bisulfate solution for a few minutes. The yield of acid melting at −15° to −14.5° is 100–105 g. (93–97 per cent of the theoretical amount) (Note 4) and (Note 5).

2. Notes

1. Gattermann[1] recommends that the operator smoke during the preparation, for he found that a trace of hydrogen cyanide is sufficient to give the tobacco smoke a highly characteristic flavor. This preliminary warning is useful in case of leaky apparatus or a faulty hood.
2. It is essential that the coil be cooled with ice only. A freezing mixture causes solidification of the hydrogen cyanide with consequent clogging of the apparatus.
3. If the reaction is slowed up considerably or interrupted so that the solution becomes cool, sodium bisulfate crystallizes both in the funnel and in the generating flask.
4. The hydrogen cyanide is best kept over anhydrous calcium chloride. In this way it remains clear and water-white for months; otherwise, it soon becomes yellow, owing to the formation of azulmic acid. Concentrated hydrochloric acid (two drops per 500 g. of hydrogen cyanide) has also been recommended as a stabilizer.[2]
5. If larger quantities of hydrogen cyanide are desired, the apparatus may be modified as suggested by Lindemann[3] by using a four-holed rubber stopper in the generating flask and fitting it with a siphon tube …

1. Gattermann[1] recommends that the operator smoke during the preparation, for he found that a trace of hydrogen cyanide is sufficient to give the tobacco smoke a highly characteristic flavor. This preliminary warning is useful in case of leaky apparatus or a faulty hood.

Hydrogen cyanide can be prepared by the action of sulfuric acid on potassium ferrocyanide,[1] potassium cyanide[4] or sodium cyanide.[5] The procedure described is based on the methods of Ziegler[5] and Lindemann.[3]

This preparation is referenced from:

- Org. Syn. Coll. Vol. 6, 14
- Org. Syn. Coll. Vol. 6, 334

References and Notes

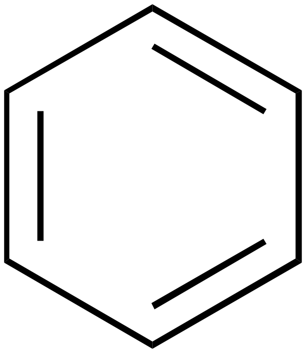
1. Gattermann, Ann. **357**, 318 (1907).
2. Slotta, Ber. **67**, 1030 (1934).
3. Lindemann, Ann. **431**, 291 (1923).
4. Villars, J. Am. Chem. Soc. **52**, 64 (1930).
5. Ziegler, Ber. **54**, 110 (1921); Lindemann, Ann. **431**, 290 (1923); Pauer, Monatsh. **58**, 1 (1931).

Appendix
Chemical Abstracts Nomenclature (Collective Index Number); (Registry Number)

azulmic acid

calcium chloride (10043-52-4)

Labelling Molecules



Benzene

SMILES

C1=CC=CC=C1

InChI

InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H

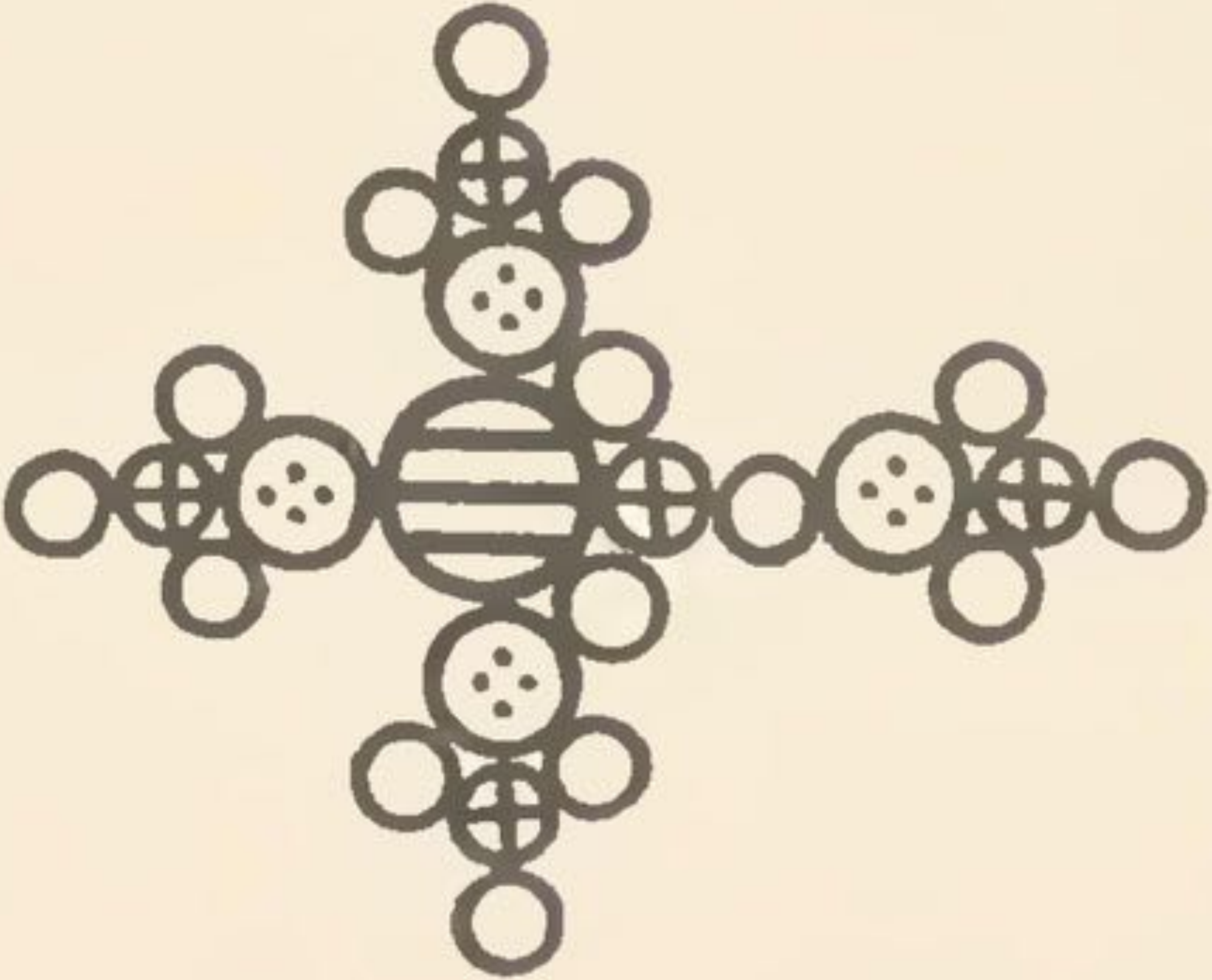
UHOVQNZJYSORNB-UHFFFAOYSA-N

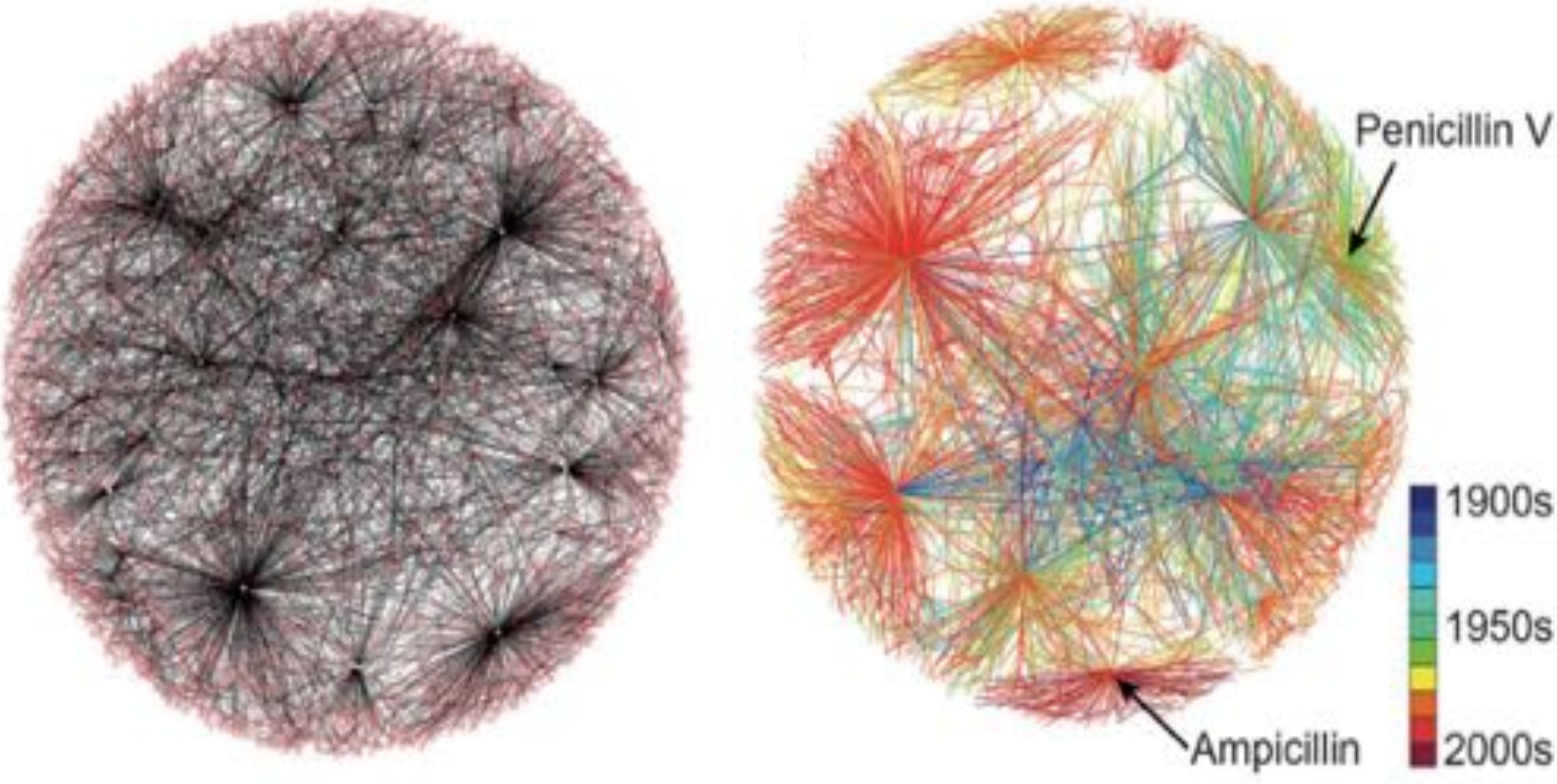
Labelling Molecules

Both SMILES and InChI provide text-based descriptions of molecular structures

Just one name for a structure and just one structure for a name *(usually)*

InChI is better, because the algorithm is open





0.1 % of the Beilstein Database
Bartosz A. Grzybowski *et al.*
Angew. Chem. Int. Ed. 2012, **51**, 7922–7927

InChI are successful because people use them

Can we label reactions too?

RInChI: example

RInChI=1.00.1S/C14H10O4/ c15-13(11-7-3-1-4-8-11)17-18-14 (16)12-9-5-2-6-10-12/h1-10H! C4H4BrNO2/c5-6-3(7)1-2-4(6)8/ h1-2H2!C7H10O2/ c1-3-4-5-6-7(8)9-2/h3-6H,1-2H3/ b4-3+,6-5+<>C7H9BrO2/ c1-10-7(9)5-3-2-4-6-8/h2-5H, 6H2,1H3/b4-2+,5-3+<>C6H6/ c1-2-4-6-5-3-1/h1-6H/d+

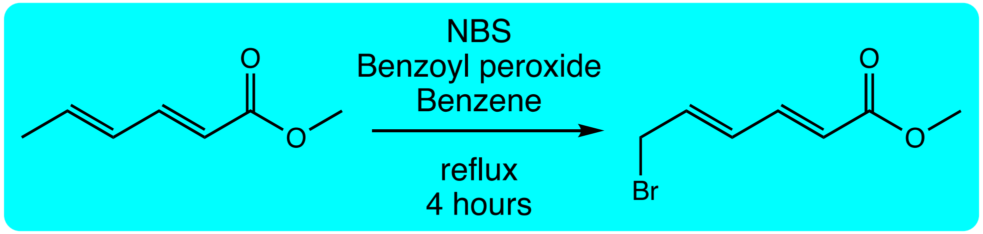
Long-RInChIKey=SA-FUHFF-OMPJBNCRMGITSC-**UHFFFAOYSA**-N-PCLIMKBDDGJMGD-**UHFFFAOYSA**-N- KWKVAGQCDSHWFK-VNKDHWASSA-N--YVJYHTBQRJXDJT-ZUVMSYQZSA-N--UHOVQNZJYSORNB- **UHFFFAOYSA**-N

Short-RInChIKey=SA-FUHFF-IOGKQBZNWJ-YVJYHTBQRJ-UHOVQNZJYS-NJUKM-NMADX-**NUHFF**-ZZZ

Web-RInChIKey=YOKVIUNDKVUECXLWI-NJXWAPQKHXRMKSA

RInChI: example

RInChI=1.00.1S/C14H10O4/ c15-13(11-7-3-1-4-8-11)17-18-14(16)12-9-5-2-6-10-12/h1-10H! C4H4BrNO2/c5-6-3(7)1-2-4(6)8/h1-2H2!C7H10O2/ c1-3-4-5-6-7(8)9-2/h3-6H,1-2H3/ b4-3+,6-5+<>C7H9BrO2/ c1-10-7(9)5-3-2-4-6-8/h2-5H, 6H2,1H3/b4-2+,5-3+<>C6H6/ c1-2-4-6-5-3-1/h1-6H/d+



Han BY, Lam NYS, MacGregor CI, Goodman JM, Paterson I (2018) *Chem Commun* **54**:3247

RInChI

- InChI – an identifier for molecules
- RInChI – an identifier for reactions

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

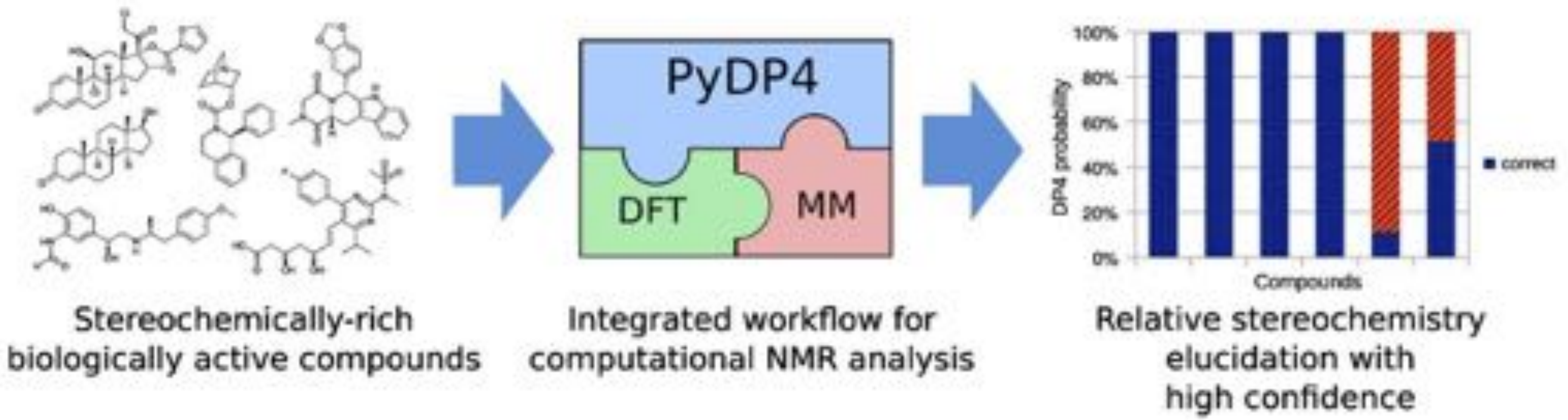
International chemical identifier for reactions (RInChI)
G. Grethe, J. M. Goodman and C. H. G. Allen
Journal of Cheminformatics 2013, **5**, 45.
DOI: 10.1186/1758-2946-5-45

http://www-rinchi.ch.cam.ac.uk/
http://www.inchi-trust.org/

RInChI next steps

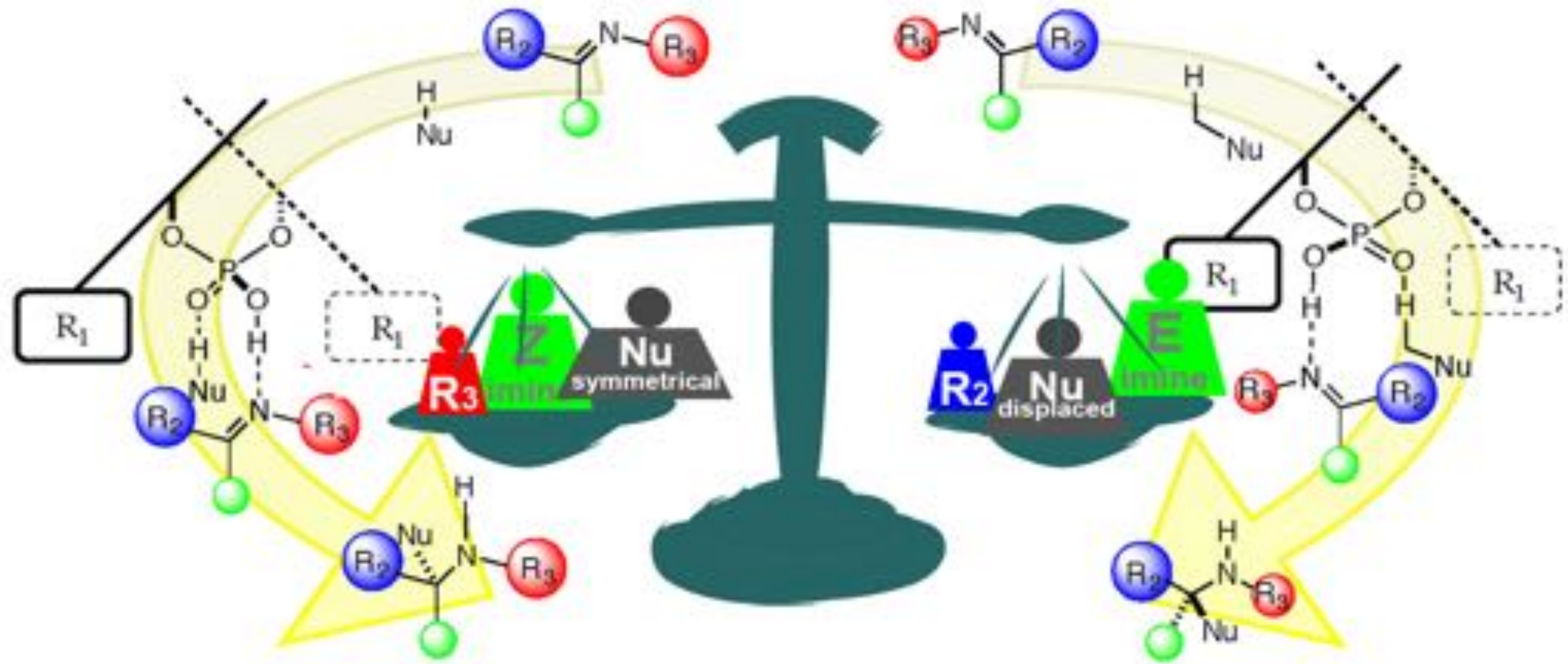
- Generic reactions
- Atom mapping
- Links to reaction classes
- *Open data, open research*
- **FAIR Molecular Data**
- **(Findable, Accessible, Interoperable, Reusable)**
- RInChI are:
  - *Easy to use; based on InChI*
  - *Canonical*
  - *Based only on the reaction, not on a central authority*

DP4 and PyDP4



Doubling the Power of DP4 for Computational Structure Elucidation
K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman
Org. Biomol. Chem. 2017, **15**, 8998-9007
DOI: 10.1039/C7OB01379E

Challenging Chemistry: Solving Molecular Problems

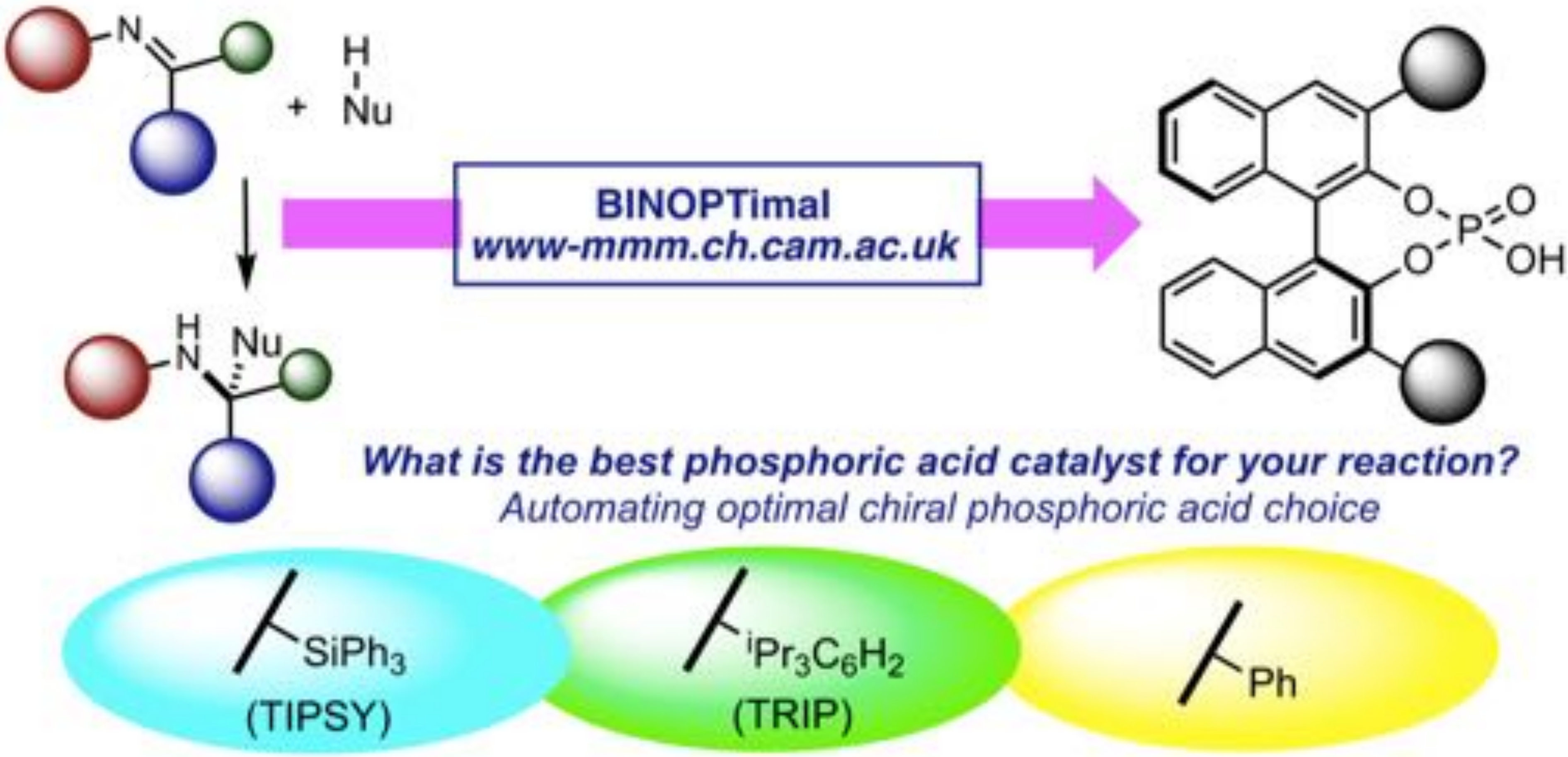


A Practical Guide for Predicting the Stereochemistry of Bifunctional Phosphoric Acid Catalyzed Reactions of Imines

J. P. Reid, L. Simon and J. M. Goodman *Acc. Chem. Res.* 2016, **49**, 1029-1041.

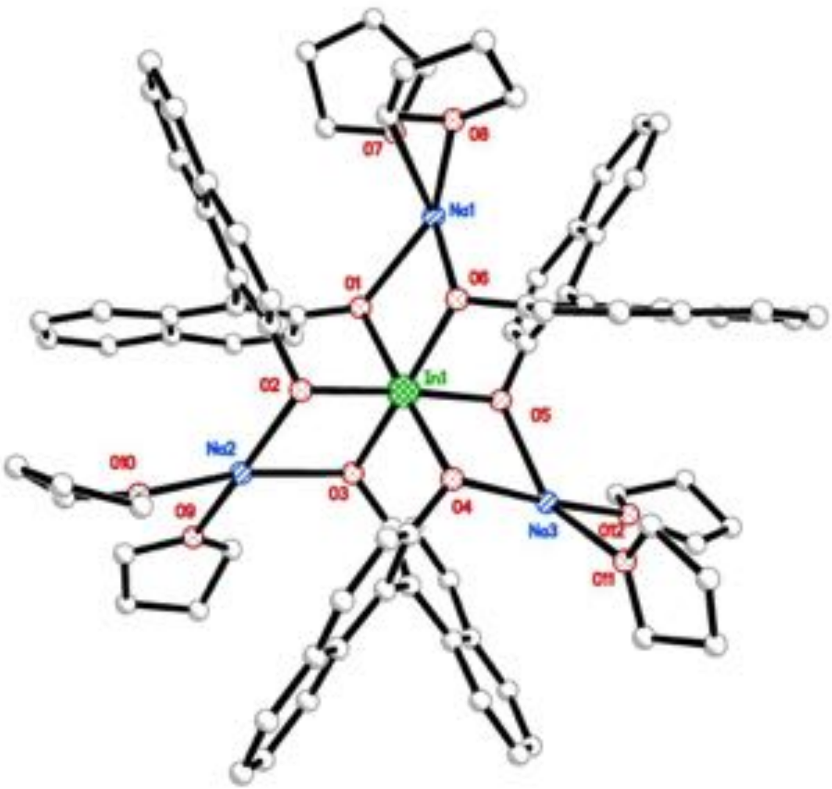
DOI: 10.1021/acs.accounts.6b00052

Challenging Chemistry: Solving Molecular Problems



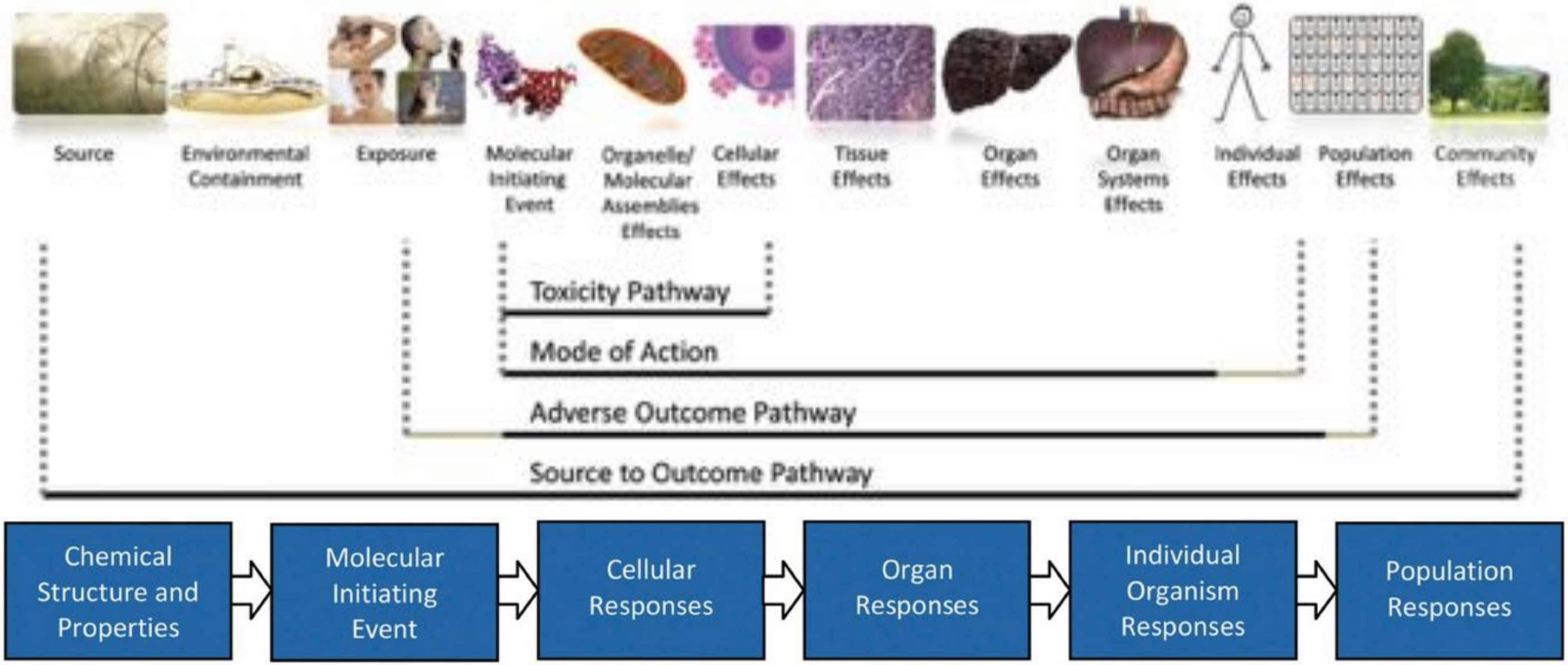
BINOPtimal: A Web Tool for Optimal Chiral Phosphoric Acid Catalyst Selection
J. P. Reid, K. Ermanis and J. M. Goodman
Chemical Communications, 2019, DOI: 10.1039/C8CC09344J

Challenging Chemistry: Solving Molecular Problems



Machine Learning
needs data

Molecular Initiating Events (MIE)



The role of chemistry in developing understanding of adverse outcome pathways and their application in risk assessment
Steve Gutsell and Paul Russell *Toxicol. Res.*, 2013, **2**, 299

Challenging Chemistry: Solving Molecular Problems

helping data escape from the laboratory
...without getting lost or captured

UNIVERSITY OF CAMBRIDGE

Challenging Chemistry: Solving Molecular Problems

Jonathan Goodman

University of Cambridge

jmg11@cam.ac.uk