Chemical literacy for the ages
Essential skills in 2D representation

American Chemical Society
Boston, MA
August 19, 2015

Leah McEwen
Cornell University
lrm1@cornell.edu

Evan Hepler-Smith
Princeton University
ehepler@princeton.edu
molecule interpretation representation
molecule

interpretation

representation
interpretation

stereochemistry
implicit H
protonation
resonance

representation

(R)-5-((S)-1,2-dihydroxyethyl)-3,4-dihydroxyfuran-2(5H)-one

L-ascorbic acid
50-81-7
Two essential skills

Critical literacy

Historical thinking
Critical literacy

Literacy:
Read and understand formulas and identifiers

Critical literacy:
What do formulas and identifiers say & not say?
What might they imply?
Are they appropriate?

Historical thinking

(L)-5-((S)-1,2-dihydroxyethyl)-3,4-dihydroxyfuran-2(5H)-one
L-ascorbic acid
50-81-7
Two essential skills

Critical literacy

Historical thinking

In general:
history of conventions

In particular:
provenance of identifiers

L-ascorbic acid

50-81-7
When we wish to probe something which is so ingrained in us that it is second nature (as the structural formula is), it is useful to put oneself outside.

(Roald Hoffmann & Pierre Laszlo, 1991)
Geneva 3-methyl-1,3$^1$-hexanedioic acid 3-methyl-1,3$^1$-hexanedral
Formula Indexes
33) Acetat d. $\varepsilon$-Oxy-$\beta$-Ketopentan (Acetylpropylacetat).
   C 52,5 — H 7,5 — O 40,0 — M. G. 160.

1) Methyäther d. Isomannid. Sm. 44—45°; Sd. 17°, I, 317.

2) Dimethylenäther d. Pentaerythrit. Sm. 50° (B. 9).

3) 1,2-Dioxyhexahydrobenzol-1-Carbonsäure. Ca — II, 1730.

4) $\beta$-Oxy-$\delta$-Keto-$\beta$-Methykpentan-$\alpha$-Carbonsäure
   (valeriansäure). Fl. Ba$_2$, Ag$_4$ + H$_2$O (A. 266, 351).

5) Pentan-$\alpha\alpha$-Dicarbonsäure (norm. Butylmalonsäure)
   Pb, Cu + H$_2$O, Ag$_2$ (B. 17, 2218; Ph. Ch. 8, 449, I, 676).

6) Pentan-$\alpha\beta$-Dicarbonsäure (Propylbernsteinsäure)
   24, 87, 2036; 26, 1927; A. 214, 59; 304, 189; A.
   23, 439; Ph. Ch. 8, 457). — I, 677.
Diagrammatic Outline of Arrangement of Organic Compounds

in Beilstein's "Handbuch der Organischen Chemie"
(Fourth Edition)

### Acyclic Compounds

<table>
<thead>
<tr>
<th>Hydro-Carbons</th>
<th>Hydroxy Comp'ds</th>
<th>Carboxylic Compounds</th>
<th>Carboxylic Acids</th>
<th>Sulphur Acids</th>
<th>Amines</th>
<th>Nitrogen Comp'ds</th>
<th>Metallic Comp'ds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simple Carboxylic Compounds</th>
<th>Hydroxoy Carboxylic Acids</th>
<th>Carboxylic Acids</th>
<th>Sulphonic &amp; Sulphonic Acids</th>
<th>Simple Amino Compounds</th>
<th>Hydroxyl &amp; Amino Derivatives</th>
<th>Amino Derivatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>2—15</td>
<td>16—70</td>
<td>71—151</td>
<td>152—322</td>
<td>323—331</td>
<td>332—379</td>
<td>380—400</td>
</tr>
</tbody>
</table>

### Isocyclic Compounds

<table>
<thead>
<tr>
<th>Cyclic Compounds</th>
<th>Heterocyclic Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>450—498</td>
<td>1505—1504</td>
</tr>
<tr>
<td>499—606</td>
<td>1592—1591</td>
</tr>
<tr>
<td>609—890</td>
<td>1929—2351</td>
</tr>
<tr>
<td>891—1504</td>
<td>2252—2358</td>
</tr>
</tbody>
</table>

### Heterocyclic Compounds

- Heterocompounds with one oxygen atom: 2360—2865
- " two atoms: 2666—2850
- " three: 2951—3006
- " four or more: 3007—3031
- " one nitrogen atom: 3092—3457
- " two atoms: 3458—3793

- Heterocompounds with three nitrogen atoms: 3794—4088
- " four or more: 4099—4187
- " oxygen and one: 4188—4483
- " two: 4484—4669
- " three or more: 4679—4719
Diagrammatic Outline of Arrangement of Organic Compounds

in
Beilstein's "Handbuch der Organischen Chemie"
(Fourth Edition)

Acyclic Compounds

Carboxylic Acids

Simple Carboxylic Acids

Hydroxy Carboxylic Acids

Carbonyl Carboxylic Acids

Carboxylic Acids

Simple Carboxylic Acids

Hydroxy Carboxylic Acids

Carbonyl Carboxylic Acids

Sulphur Acids

Functional Derivatives

Simple Amino Compounds

Hydroxyl & Amine Derivs.

Amines

Nitrogen Comp'ds

Metallic Comp'ds

5. 3-Methylsäure-hexansäure-(I). Pentan-α,β-dicarbonsäure, Propylbernsteinsäure \( C_{10}H_{18}O_4 = CH_3 \cdot CH_2 \cdot CH_2 \cdot CH(CO_2H) \cdot CH_2 \cdot CO_2H \). B. Beim Erhitzen von 3,3-Dimethylsäure-hexansäure-(I) (WALTZ, A. 214, 59). Bei der Reduktion von Propylmaleinsäure mit Natriumamalgam (FTTTCO, GLASER, A. 304, 188; vgl. SSEMENOW, K. 23, 439). Bei der Reduktion von Propylfumarsäure mit Natriumamalgam (F. G.; vgl. DEMARÇAY, A. ch. 5) 20, 491). Bei der Reduktion von γ-Aethyl-itaconsäure \( C_9H_8 \cdot CH_3 \cdot CO_2H \cdot CH_2 \cdot CO_2H \).

S: st. No. 176.

PROPYLBERNSTEINSÄURE.

152—322

Heterocyclic Compounds

<table>
<thead>
<tr>
<th>one oxygen atom</th>
<th>2369—2965</th>
</tr>
</thead>
<tbody>
<tr>
<td>two</td>
<td>2556—3050</td>
</tr>
<tr>
<td>three</td>
<td>2591—3006</td>
</tr>
<tr>
<td>four or more</td>
<td>3007—3031</td>
</tr>
<tr>
<td>one nitrogen atom</td>
<td>3033—3457</td>
</tr>
<tr>
<td>two</td>
<td>3458—3793</td>
</tr>
</tbody>
</table>

Heterocompounds with three nitrogen atoms 3794—4088

four or more oxygen and one two three or more 4188—4483

four or more two three or more 4454—4669

four or more two three or more 4670—4719
**Geneva** 3-methyl-1,3\(^1\)-hexanedioic acid

**Richter** \( C_7H_{12}O_4 \) (+ various names)

**Beilstein** Syst. No. 176 (+ various names)

**Chem Abs** succinic acid, propyl-

3-methyl-1,3\(^1\)-hexanedial

\( C_7H_{12}O_2 \) (+ various names)

Syst. No. 95 (+ various names)

succinaldehyde, propyl-
“It is essential that all chemists who speak the French language – and with them, I hope, all of the Latin-minded – prepare themselves for battle against Anglo-Saxon illogic.”
**IUPAC (1930)**

- 1,2-pentanedicarboxylic acid
- propylsuccinic acid
- 2-propylbutanedioic acid
- propylbutanedial
- propylsuccinaldehyde
- 2-propylbutanedial
IUPAC (1930)

1,2-pentanedicarboxylic acid
propylbutanediol
2-methylpropylbutanedioic acid
propylbutanediol

"This report is not intended to interfere with the editing of Beilstein or of Chemical Abstracts"
<table>
<thead>
<tr>
<th>Source</th>
<th>Structure 1</th>
<th>Structure 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geneva</td>
<td>3-methyl-1,3(^1)-hexanedioic acid</td>
<td>3-methyl-1,3(^1)-hexanodial</td>
</tr>
<tr>
<td>IUPAC (1930)</td>
<td>1,2-pentanedicarboxylic acid</td>
<td>propylbutanedial</td>
</tr>
<tr>
<td></td>
<td>propylsuccinic acid</td>
<td>propylsuccinaldehyde</td>
</tr>
<tr>
<td></td>
<td>2-propylbutanedioic acid</td>
<td>2-propylbutanedial</td>
</tr>
<tr>
<td>Dyson</td>
<td>C6 . 3C . 1, 7X</td>
<td>C6 . 3C . 1, 7EQ</td>
</tr>
<tr>
<td>WLN</td>
<td>QVY3.1VQ</td>
<td>O:2Y3.1:O</td>
</tr>
<tr>
<td>Connect’n tables</td>
<td><img src="image1.png" alt="Connect’n tables" /></td>
<td><img src="image2.png" alt="Connect’n tables" /></td>
</tr>
<tr>
<td>CAS RN</td>
<td>618-57-5</td>
<td>105409-75-4</td>
</tr>
</tbody>
</table>

IUPAC Blue & Red books, CAS 9CI, SMILES, InChI, IUPAC PIN, etc.
Nomenclature & notation

Reference media

Systematic names
Journals
Connection Tables
Many-volume references
Registry numbers

Indexes
Card files
Molecular formulas
Magnetic tape
Punched cards
Line formulas and ciphers
Nomenclature & notation

Reference media
- Indexes
- Card files
- Magnetic tape
- Punched cards

Chemical institutions
- IUPAC
- Line formulas and ciphers
- National societies
- Chemical firms

Publishers
- Journals

Connection Tables
- Many-volume references

Systematic names
- Molecular formulas

Registry numbers

Government agencies
Indexing & registration shaped notation...
but not always in the ways you’d expect.
Indexing & registration shaped notation...
but not always in the ways you’d expect.

Logical consistency and precision vs.
flexibility and intelligibility...

a perpetual oscillation.
Indexing & registration shaped notation... but not always in the ways you’d expect.

Logical consistency and precision vs. flexibility and intelligibility... a perpetual oscillation.

Creators of notation systems usually had particular users in mind... and those users sometimes had other ideas.
Cheminformatics OLCC: A blended learning approach to teaching undergraduate chemistry students critical informatics skills

Fall 2015 Course Page: http://olcc.ccce.divched.org/Fall2015OLCC
How do they work?

Why do they work that way?

Where are they most often used?

What questions should I ask about them?
How do they work?

Why do they work that way?

Where are they most often used?

What questions should I ask about them?
How do they work?

Why do they work that way?

Where are they most often used?

What questions should I ask about them?
2.4.2. Exercise 2

Which of above form(s) of notation is/are preferable for:

<table>
<thead>
<tr>
<th>Knows little chemistry</th>
<th>Knows lots of chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Computer</td>
</tr>
<tr>
<td>Consumer</td>
<td>Google</td>
</tr>
<tr>
<td>Venture capitalist</td>
<td>MS Word</td>
</tr>
<tr>
<td>Readers of popular blog</td>
<td>Keynote</td>
</tr>
<tr>
<td>Your PI</td>
<td>SciFinder</td>
</tr>
<tr>
<td>Journal readers</td>
<td>PubChem</td>
</tr>
<tr>
<td>Cheminformaticians</td>
<td>ChemDraw</td>
</tr>
</tbody>
</table>
### Communicating chemical structure

<table>
<thead>
<tr>
<th>Molecular formula</th>
<th>Structural formula</th>
<th>Systematic name</th>
<th>SMILES</th>
<th>InChI</th>
<th>CAS RN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₅H₈</td>
<td><img src="structure1.png" alt="Structure" /></td>
<td>2-methylbuta-1,3-diene</td>
<td>CC(=C)C=C</td>
<td>InChI=1S/C5H8/c1-4-5(2)3/h4H,1-2H2,3H3</td>
<td>78-79-5</td>
</tr>
<tr>
<td>C₂H₇NO₂</td>
<td><img src="structure2.png" alt="Structure" /></td>
<td>1-aminoethane-1,2-diol</td>
<td>C(C(N)O)O</td>
<td>InChI=1S/C2H7NO2/c3-2(5)1-4/h2.4-5H,1.3H2</td>
<td>13053-46-8 (this one is kind of hidden in PubChem)</td>
</tr>
<tr>
<td>C₂H₇NO₂</td>
<td><img src="structure3.png" alt="Structure" /></td>
<td>Ammonium acetate</td>
<td>CC(=O)[O-].[NH4+]</td>
<td>InChI=1S/C2H4O2.H3N/c1-2(3)4;/h1H3,(H,3,4):1H3</td>
<td>631-61-8</td>
</tr>
<tr>
<td>C₆H₆O</td>
<td><img src="structure4.png" alt="Structure" /></td>
<td>Trick</td>
<td>Question</td>
<td>Lots</td>
<td>Of Compounds</td>
</tr>
<tr>
<td>C₇H₁₂O₄</td>
<td><img src="structure5.png" alt="Structure" /></td>
<td>Diethyl propanedioate</td>
<td>C(C(=O)C)C(=O)OCC</td>
<td>InChI=1S/C7H12O4/c1-3-10-6(8)5-7(9)11-4-2/h3-5H2,1-2H3</td>
<td>105-53-3</td>
</tr>
</tbody>
</table>

Note: the aminoethanediol example is kind of instructive. The R enantiomer first got into the CAS system in a 2014 patent and the S enantiomer isn't registered at all. Only the racemic version is accessible in PubChem, and none of them show up in ChemSpider.
Geneva

1. side chains
2. unsaturation
3. funct’l groups

IUPAC

1. funct’l groups
2. unsaturation
3. side chains
### Cxn table A

<table>
<thead>
<tr>
<th>#</th>
<th>Elem.</th>
<th>From</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

### Cxn table B

<table>
<thead>
<tr>
<th>#</th>
<th>Elem.</th>
<th>From</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>O</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

![Chemical structure](image)
**Atom table**

<table>
<thead>
<tr>
<th>Atom #</th>
<th>Elem.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
</tr>
</tbody>
</table>

**Bond table**

<table>
<thead>
<tr>
<th>Bond #</th>
<th>From</th>
<th>To</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

![Chemical structure diagram](image)
Two essential skills

Critical literacy
Read and understand formulas and identifiers
What do formulas and identifiers say & not say?
What might they imply?
Are they appropriate?

Historical thinking
In general:
history of conventions
In particular:
provenance of identifiers
For more....

evanheplersmith.com

olcc.ccce.divched.org
Acknowledgements

Cheminformatics OLCC: A blended learning approach to teaching undergraduate chemistry students critical informatics skills

Fall 2015 Course Page: http://olcc.ccce.divched.org/Fall2015OLCC
Chemical literacy for the ages
Essential skills in 2D representation

American Chemical Society
Boston, MA
August 19, 2015

Leah McEwen
Cornell University
lrm1@cornell.edu

Evan Hepler-Smith
Princeton University
ehepler@princeton.edu