UNDERSTANDING TAUTOMERISM (IN 2½D)

Roger Sayle

NextMove Software, Cambridge, UK

11th RDKit User Group Meeting, Berlin, Germany, Thursday 13th October 2022

UNDERSTANDING TAUTOMERISM (IN 1D, 2D, 2½D AND 3D)

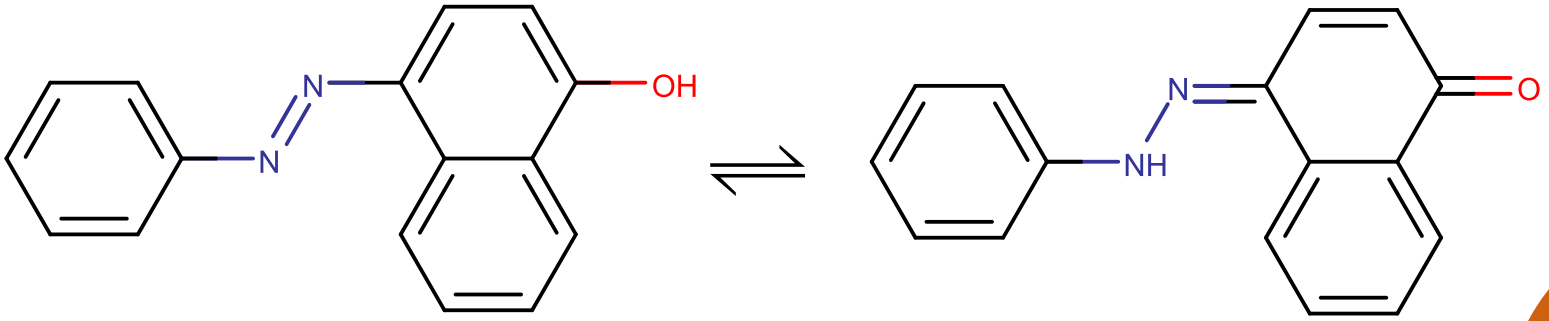
Roger Sayle

NextMove Software, Cambridge, UK

11th RDKit User Group Meeting, Berlin, Germany, Thursday 13th October 2022

TAUTOMERISM 101

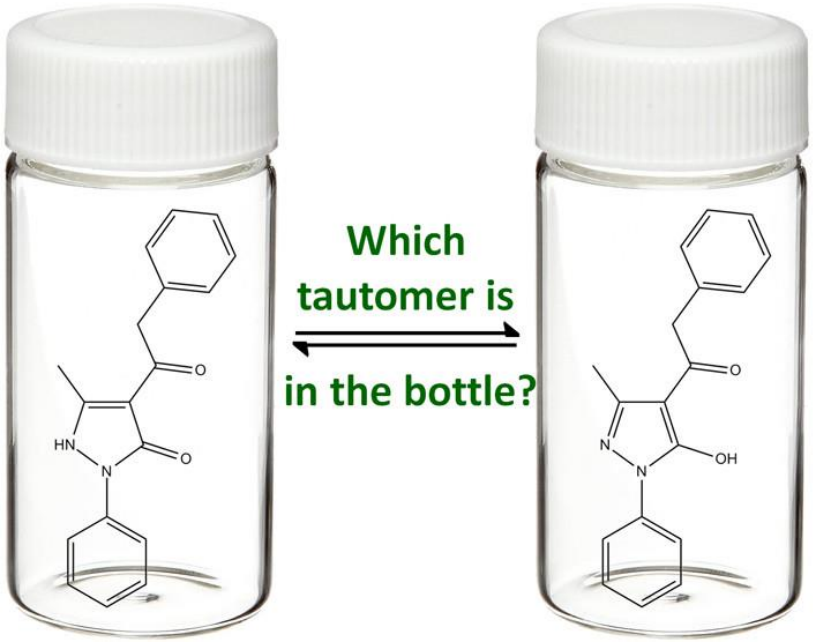
- Tautomers are isomers that (easily) interconvert by migration of hydrogen atoms[1].
- The term “tautomer” was introduced by Conrad Laar in 1885 in a paper describing the interconversion of C_6H_5-N_2-$C_{10}H_6$-OH ⇔ C_6H_5-N_2H=$C_{10}H_6$=O.



InChI=1S/C16H12N20/c19-16-11-10-15(13-8-4-5-9-14(13)16)18-17-12-6-2-1-3-7-12/h1-11,19H
InChI=1S/C16H12N20/c19-16-11-10-15(13-8-4-5-9-14(13)16)18-17-12-6-2-1-3-7-12/h1-11,17H

MOLECULES EXIST IN MULTIPLE FORMS

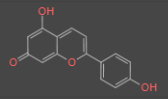
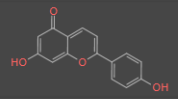
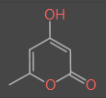
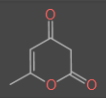
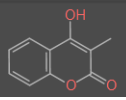
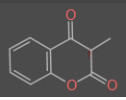
- A common misunderstanding is that there exists one “true” isomer [to be registered in a database].



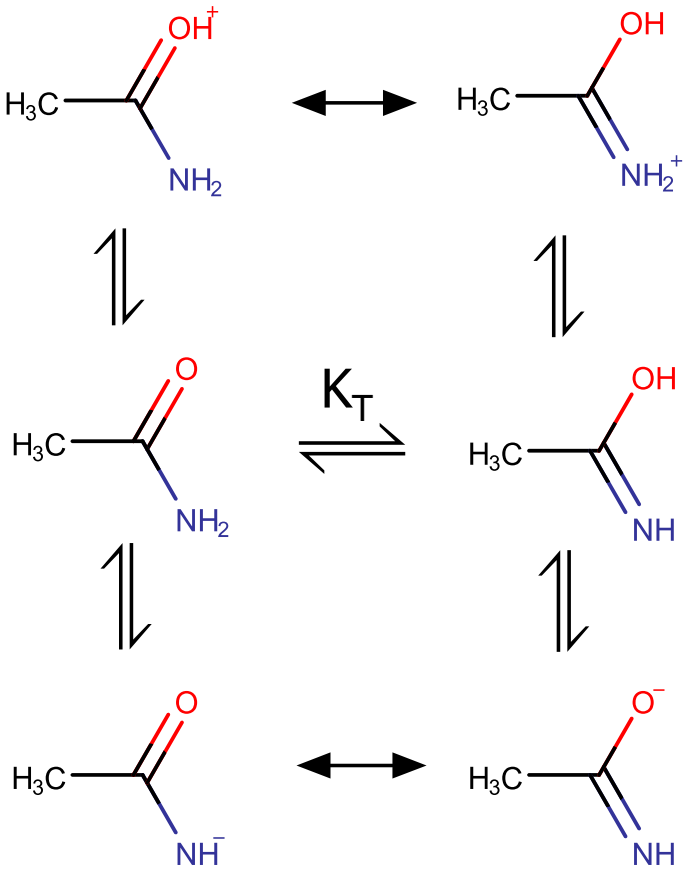
- In reality, these compounds exist as mixtures.

WAHL AND SANDER'S TAUTOBASE (2020) IN DATAWARRIOR

- An excellent resource (derived from Peter Taylor's Treatise and other sources) is Tautobase.

	Tautomer1	Tautomer2	log K	% (Tautomer1)	Preferred	State/Solvent
313			-2.35			Water
314			1.28			Water
315			3.0			Water
316					1	
317					1	Water
318				100		Water
319				100		Water

TAUTOMERS AND IONIZATION, OH MY!



common cation
pKa ~ -1.5

common anion
pKa ~ 15

TAUTOMER RATIOS BY PK_A PREDICTION

$$K_{a1} = \frac{[AB-][H+]}{[HAB]} \text{ and } pK_{a1} = -\log_{10} K_{a1}$$

$$K_{a2} = \frac{[AB-][H+]}{[ABH]} \text{ and } pK_{a2} = -\log_{10} K_{a2}$$

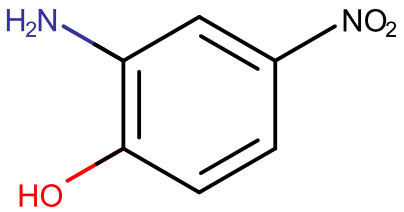
$$\frac{K_{a1}}{K_{a2}} = \frac{[AB-]/[HAB]}{[AB-]/[ABH]} = \frac{[ABH]}{[HAB]} = K_T$$

$$K_T = 10^{(pK_{a1} - pK_{a2})}$$

- Hence with an accurate way to predict acid disassociation constants (pKa), we can predict tautomer ratios.
- Even with approximate pKa predictions, we can determine the most probable/preferred tautomer.

HAMMETT AND TAFT EQUATIONS

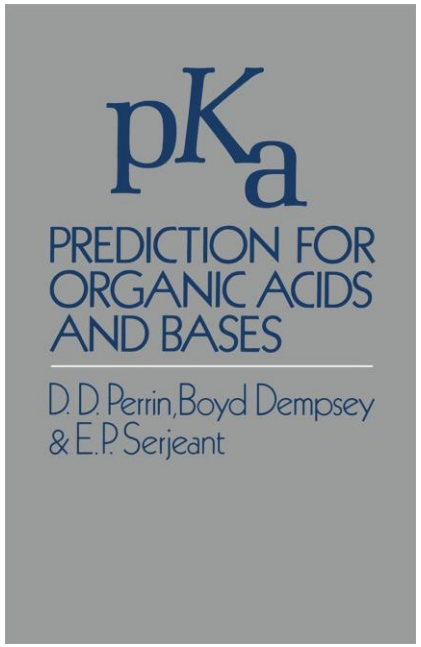
- In 1937, Louis Hammett observed the additive effects of electron withdrawing substituents on pKa.
- The pKa of phenols is given by 9.92 - 2.23 Σσ.

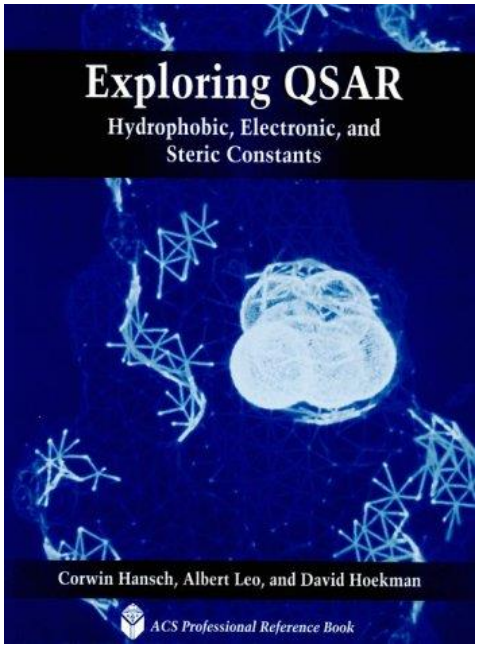


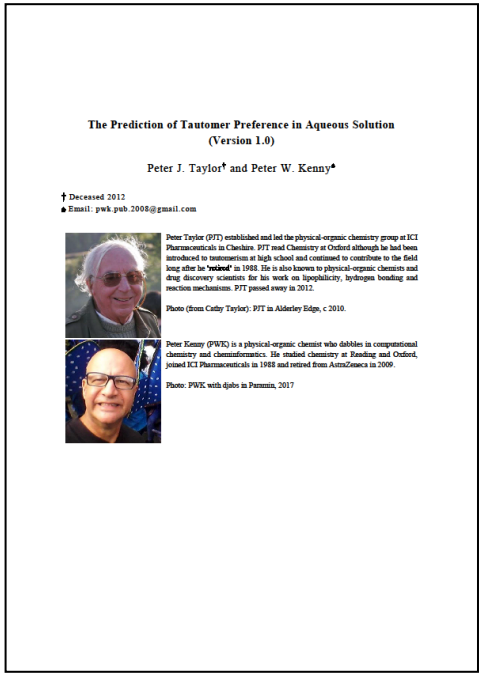
- σ_{para} for nitro is 1.24 and σ_{ortho} for amine is 0.03, so the pKa of CHEMBL 316992 (above) is

  9.92 - 2.23*(1.24+0.03) = 7.09.
- Hence is 67% deprotonated physiological pH (7.4).

HIGHLY RECOMMENDED READING







The Prediction of Tautomer Preference in Aqueous Solution
(Version 1.0)

Peter J. Taylor† and Peter W. Kenny♣

† Deceased 2012
♣ Email: pwk.pub.2008@gmail.com

Peter Taylor (PJT) established and led the physical-organic chemistry group at ICI Pharmaceuticals in Cheshire. PJT read Chemistry at Oxford although he had been introduced to tautomerism at high school and continued to contribute to the field long after he "retired" in 1988. He is also known to physical-organic chemists and drug discovery scientists for his work on lipophilicity, hydrogen bonding and reaction mechanisms. PJT passed away in 2012.

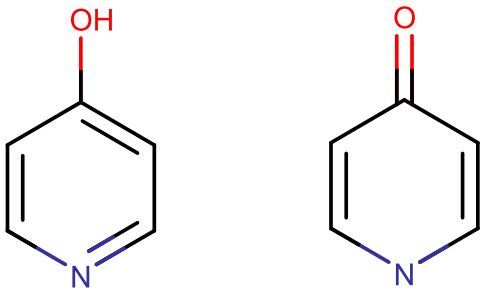
Photo (from Cathy Taylor): PJT in Alderley Edge, c 2010.

Peter Kenny (PWK) is a physical-organic chemist who dabbles in computational chemistry and cheminformatics. He studied chemistry at Reading and Oxford, joined ICI Pharmaceuticals in 1988 and retired from AstraZeneca in 2009.

Photo: PWK with djabs in Parmain, 2017

PROOF OF CONCEPT VALIDATION

- Taking the classic example of 4-pyridone, and using Hammett equations for pyridones vs. phenols, we indeed confirm that the 4-pyridone is significantly preferred.



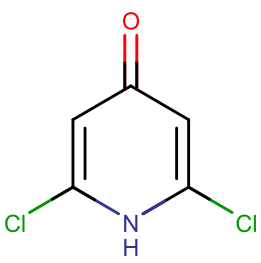
phenol = 9.92 - 2.23Σσ
= 9.92-2.23*(0.83)
= 8.07

4-pyridone = 11.12 - 4.28Σσ
= 11.12

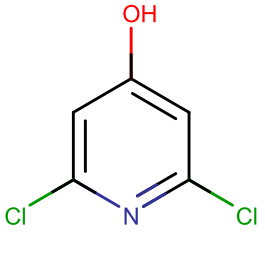
- The higher pKa indicates the preferred form.

YVONNE MARTIN'S EXAMPLE

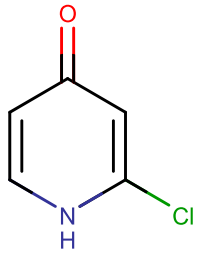
- In "Let's Not Forget Tautomers", Yvonne gave a more challenging example, that breaks the usual trend.



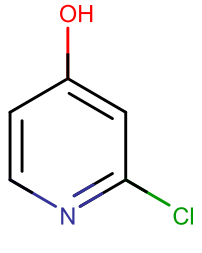
= 11.12 - 4.28*(2*0.79)
= 4.36



= 9.92 - 2.23*(0.83+2*0.37)
= 6.42



= 11.12 - 4.28*0.79
= 7.74

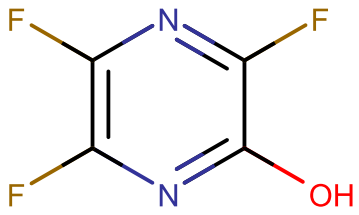


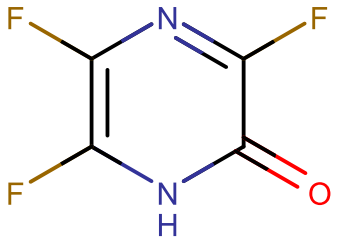
= 9.92 - 2.23*(0.83+0.79)
= 6.31

- Here the 2,6-dichloropyridinol is shown to be preferred, but 2-chloro prefer the pyridone form.
- It also shows 2,3-diCl is predominantly ionized.

TAUTOBASE #1007

- Another tricky example from Tautobase, where the pyrazinol is preferred over the pyrazinone.



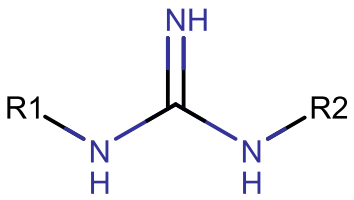


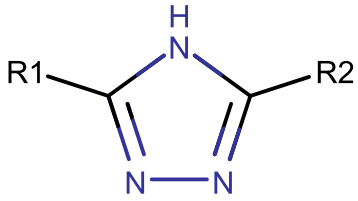
phenol = 9.92 - 2.23Σσ
= 9.92 - 2.23*(0.56+0.73+0.54+0.34+0.06)
= 4.95

2-pyridone = 11.65 - 4.28Σσ
= 11.65 - 4.28*(0.83+0.5+2*0.34)
= 3.05

CONSISTENCY VS. REALITY

- In database registration and AI/ML, it is often convenient to standardize resonant systems based upon the asymmetry of the local atom environment.

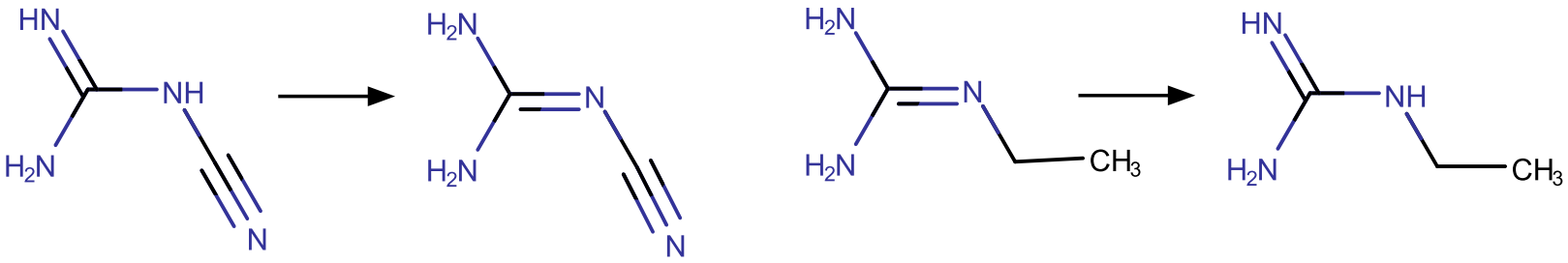




- Unfortunately, these registration heuristics often don’t match the physiological tautomeric preference.

GUANIDINE NORMALIZATION

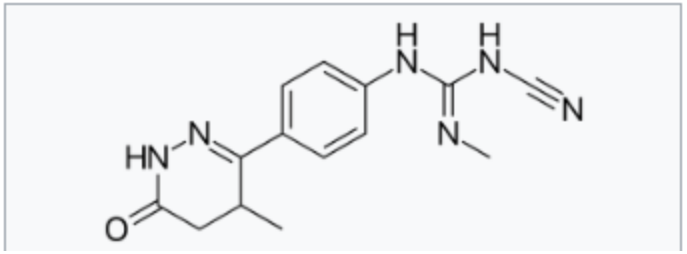
- The pKa of Guanidine is given by 14.0 - 3.60 Σσ*.
- So the double bond/deprotonation “prefers” to be towards the most electron withdrawing group.



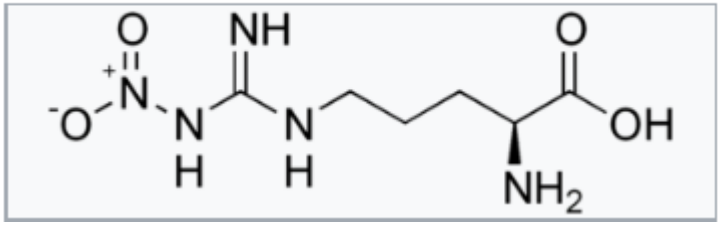
- Note that a proton is more electron withdrawing than a alkyl chain (for example in arginine).

WIKIPEDIA ROGUE'S GALLERY

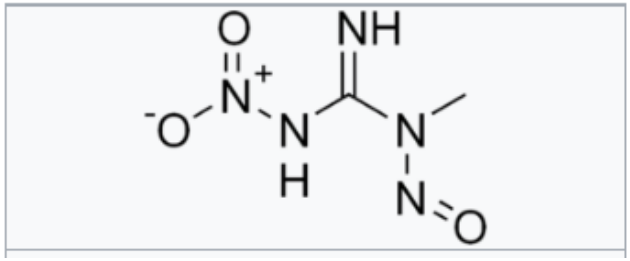
Siguazodan



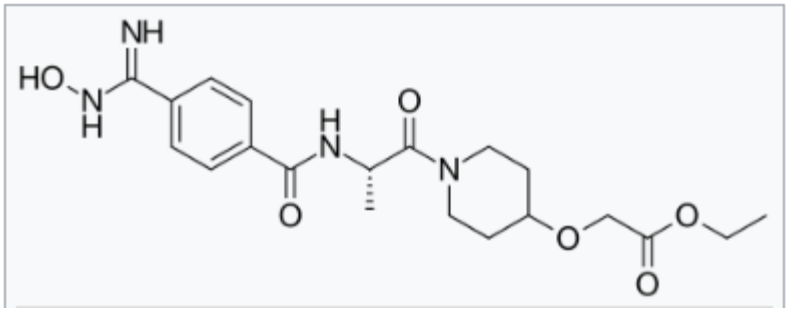
Nω-Nitro-L-arginine



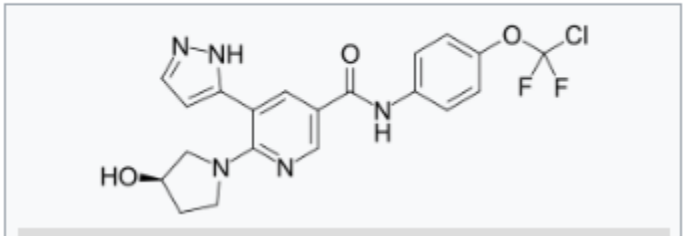
Methylnitronitrosoguanidine[1]



Sibrafiban



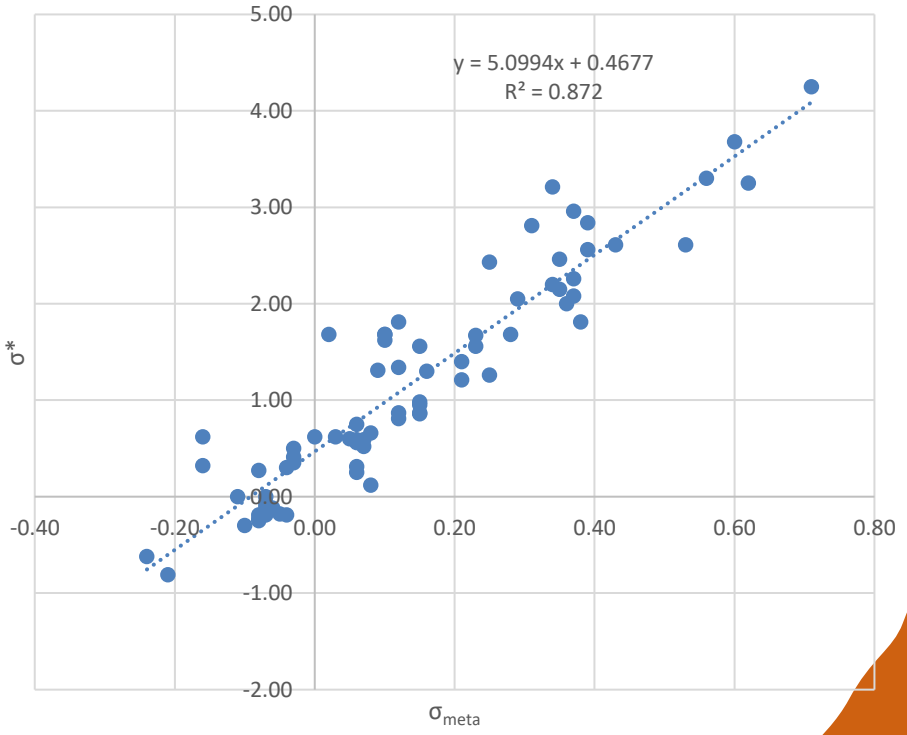
Asciminib



TAFT'S σ*

- Inductive effects are modelled by σ*, σ_{meta} or σ_I.

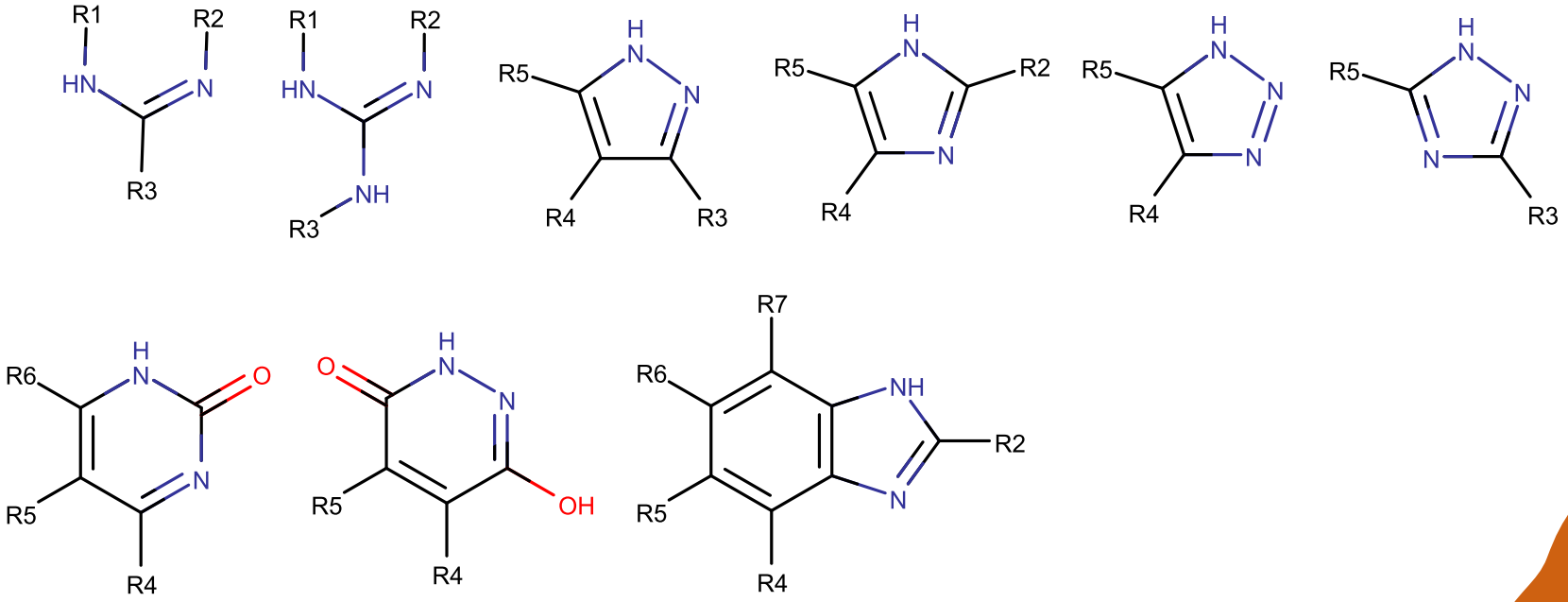
R-group	σ*	σ_{meta}
*OCCC	1.68	-0.25
*N(C)C	0.32	-0.15
*CC	0.10	-0.07
*C	0.00	-0.07
*[H]	0.49	0.00
*c1ccccc1	0.75	0.06
*O	1.34	0.12
*Cl	2.96	0.37
*C#N	3.30	0.56
*[N+](=O)[O-]	4.25	0.71



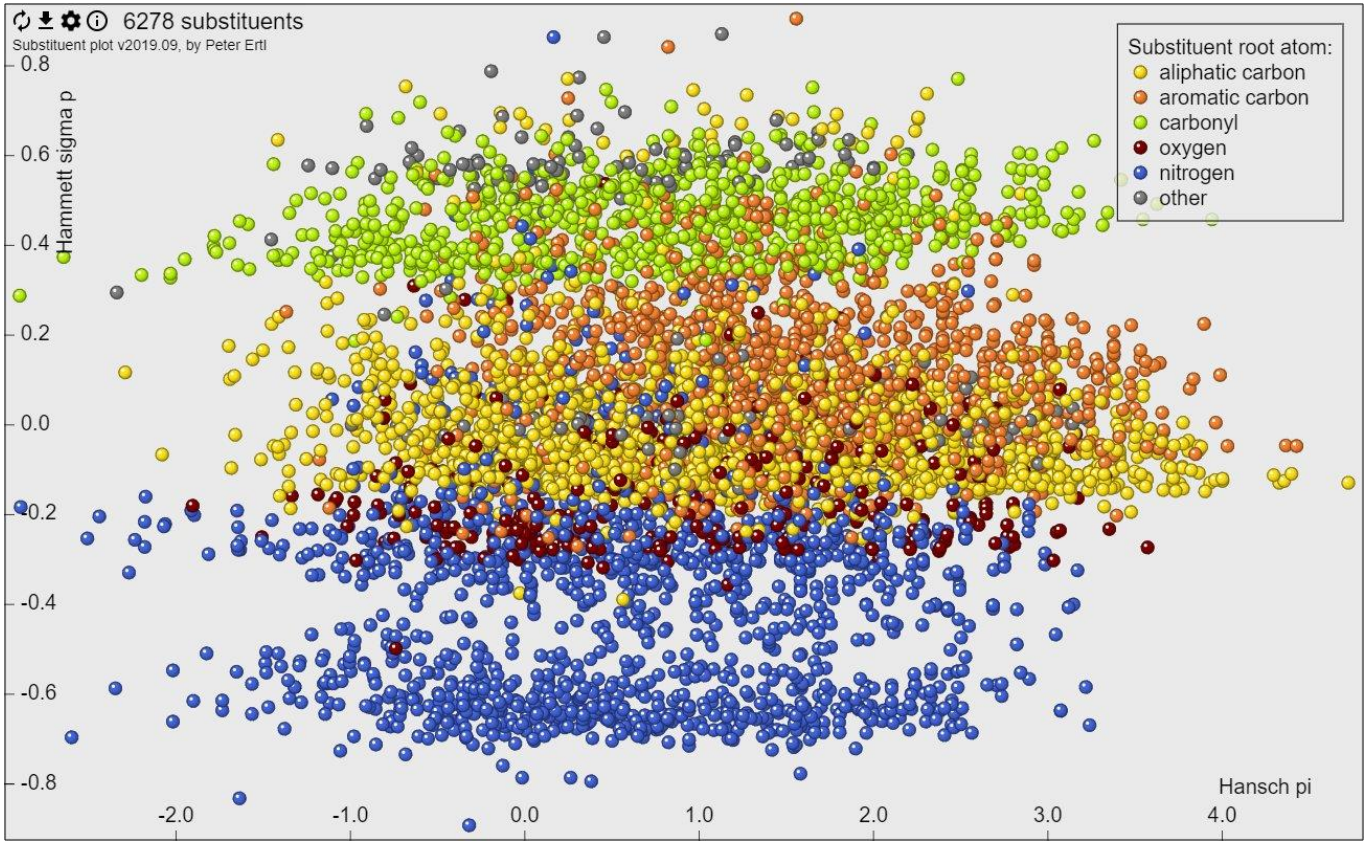
- Stocks, Alcaraz and Griffen, "On Medicinal Chemistry", 2007.

TAUTOMERIC NORMALIZATIONS

- Substructures influenced by EWG/EDG substituents.



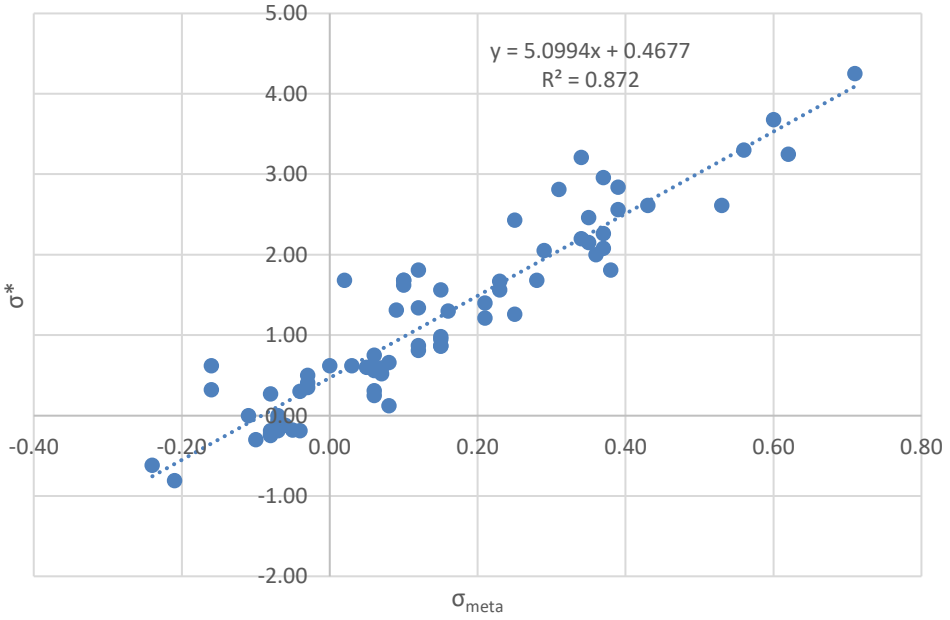
PETER ERTL'S FANTASTIC CRAIG 2.0



- Note the horizontal banding; π is a global property but σ is dominated by the attachment atom.

SIGMA, SIGMA, EVERYWHERE

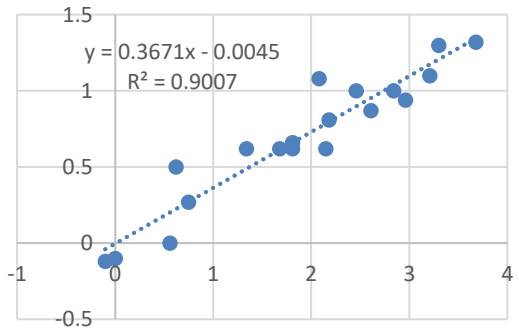
- Peter Ertl's uses of QM calculations to estimate σ values focused on Hammett's σ_{para} and σ_{meta}, but alas not Taft's σ*. Fortunately, $\sigma^* \sim 5.1\sigma_{meta} + 0.47$.



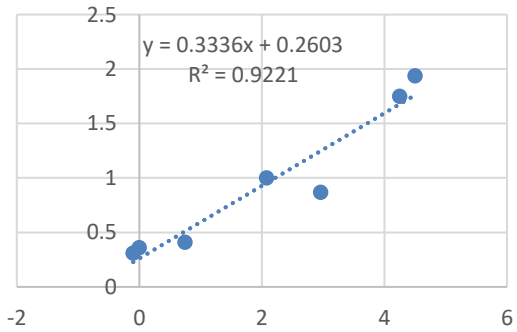
PROPAGATION COEFFICIENTS

- Likewise for propagation through common linkers…

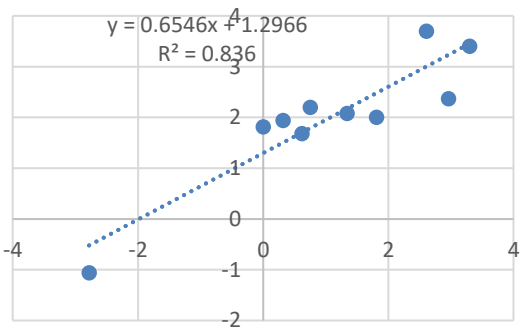
*C[R]



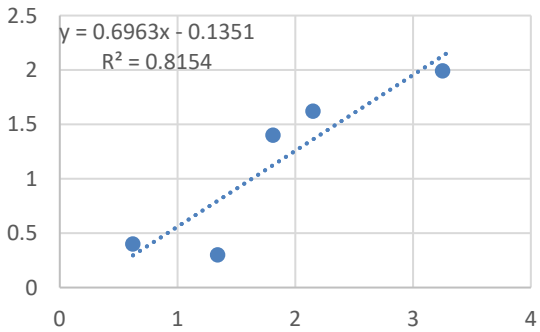
*C=C[R]



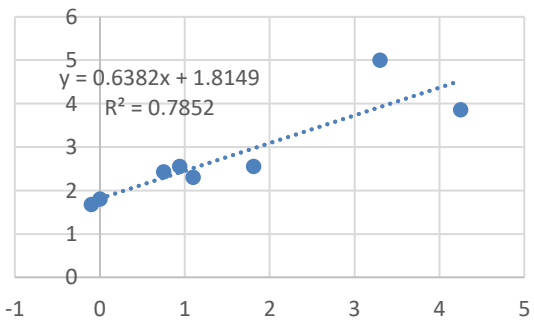
*C(=O)[R]



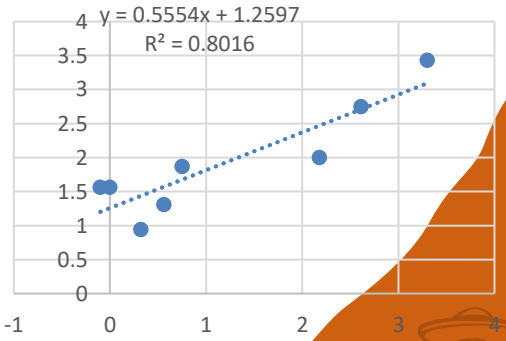
*N[R]



*O[R]



*S[R]

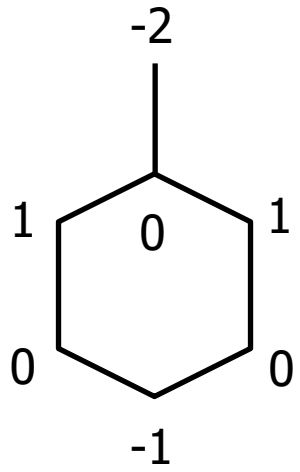


PROPAGATION COEFFICIENTS

- $\sigma^{*}(^{*}C[R]) = 0.37\sigma^{*}(^{*}[R])$
- $\sigma^{*}(^{*}C{=}C[R]) = 0.33\sigma^{*}(^{*}[R]) + 0.26$
- $\sigma^{*}(^{*}C{\#}C[R]) = 0.29\sigma^{*}(^{*}[R]) + 1.17$
- $\sigma^{*}(^{*}C(=O)[R]) = 0.65\sigma^{*}(^{*}[R]) + 1.30$
- $\sigma^{*}(^{*}N[R]) = 0.70\sigma^{*}(^{*}[R]) - 0.14$
- $\sigma^{*}(^{*}O[R]) = 0.64\sigma^{*}(^{*}[R]) + 1.81$
- $\sigma^{*}(^{*}S[R]) = 0.55\sigma^{*}(^{*}[R]) + 1.26$
- $\sigma^{*}(^{*}S(=O)[R]) = 0.50\sigma^{*}(^{*}[R]) + 2.88$
- $\sigma^{*}(^{*}S(=O)(=O)[R]) = 0.54\sigma^{*}(^{*}[R]) + 3.08$

THE ORIGINS OF σ_{META} AND σ_{PARA}

- Dewar and Grisdale (1962) decomposed σ_{meta} and σ_{para} in terms of field and mesomeric effects using Hückel theory.



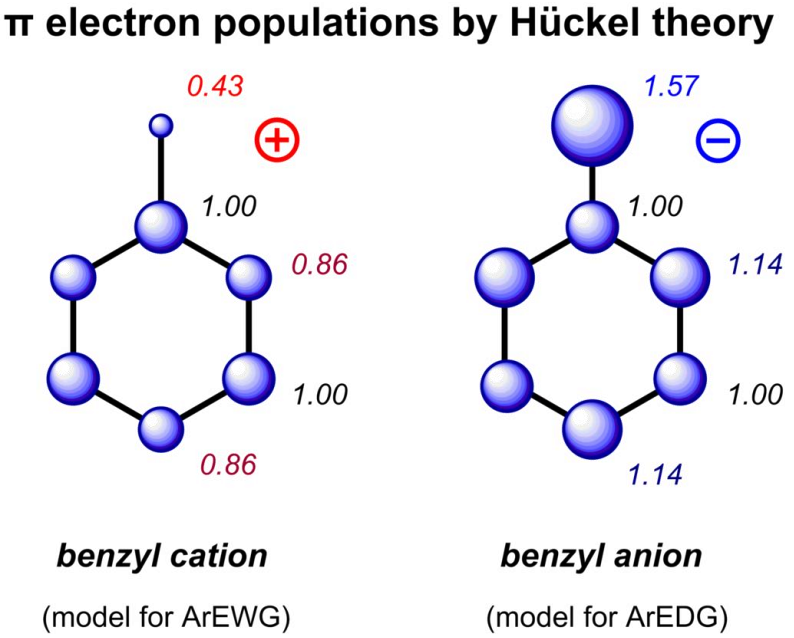
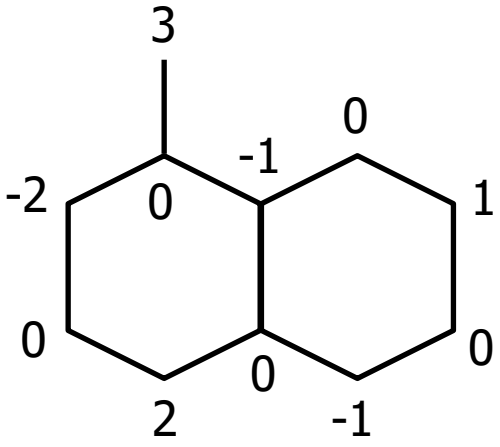
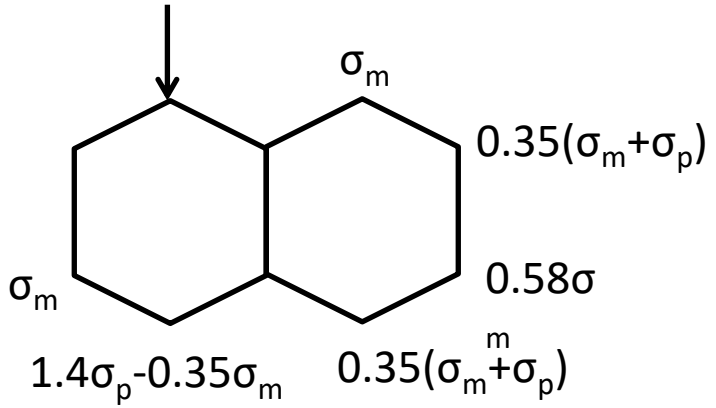


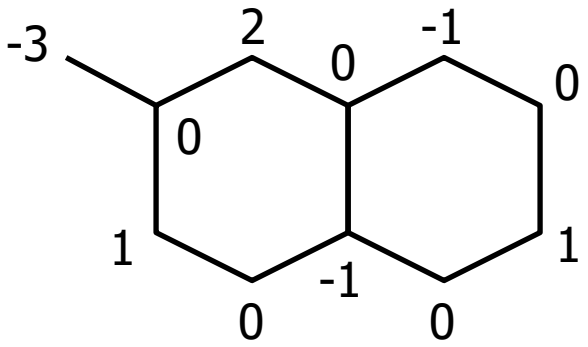
image credit:wikipedia

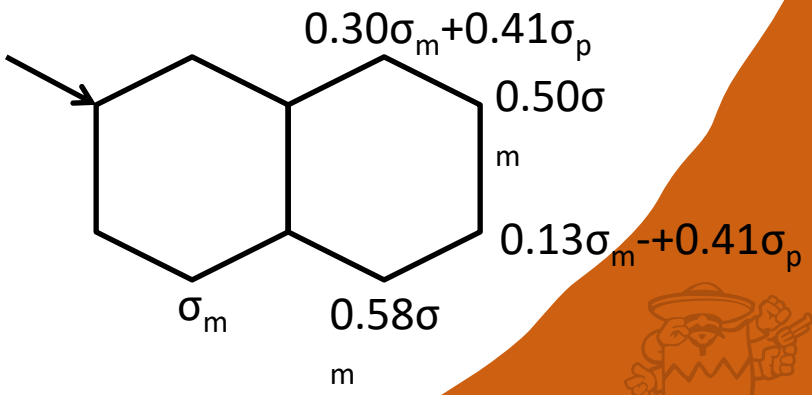
EXTENSION TO POLYAROMATIC SYSTEMS

- Longuet-Higgins (1950) and Perrin (1965) extended Hammett equation to polyaromatic rings.







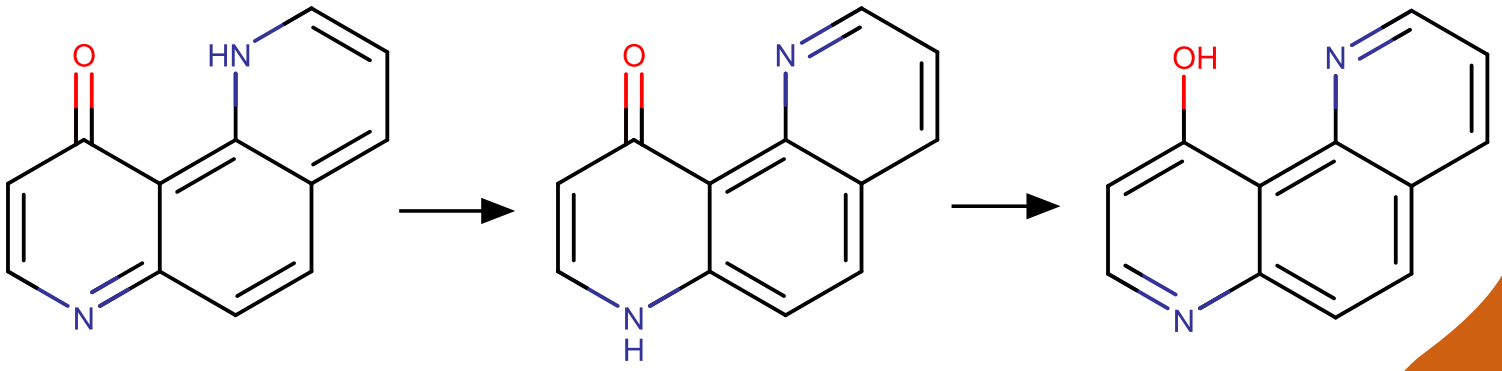


ALGORITHMIC SUMMARY

- First pass apply SMIRKS transforms to normalize to a reasonable set of tautomers (and ionization states).
- Second pass use patterns and Hammett and Taft sigma values to selected a (physiological) preference.
  - Common functional groups can be looked up in tables.
  - Propagation equations can handle common linkers.
  - Approximation algorithms can be used in missing groups have a related experimental value.
  - The connection atom type can provide a “default” value.
  - Ties can arbitrarily be split using canonical SMILES.

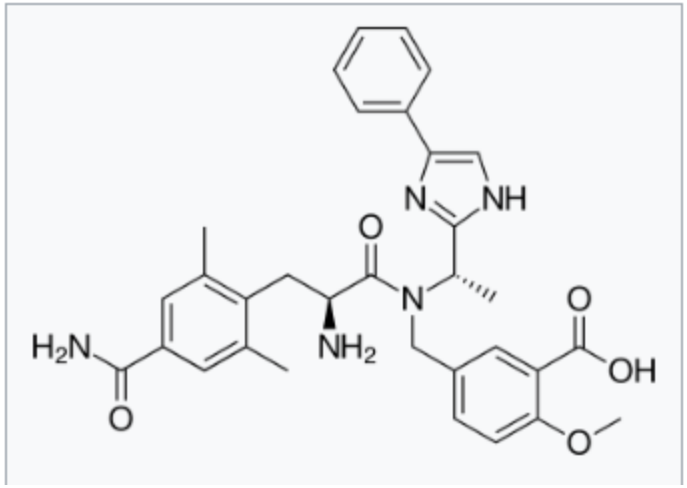
TODO

- Hydrogen bonding and lone pair repulsion effects can be handled by classic electrostatics, $\sum_{i>j} q_i q_j / r_{ij}$, where partial charges q can be determined using MMFF94 (atom types) and distances r can be determined from tables, 2D or 3D distance geometry.



IMPROVED DEPICTIONS

Eluxadoline



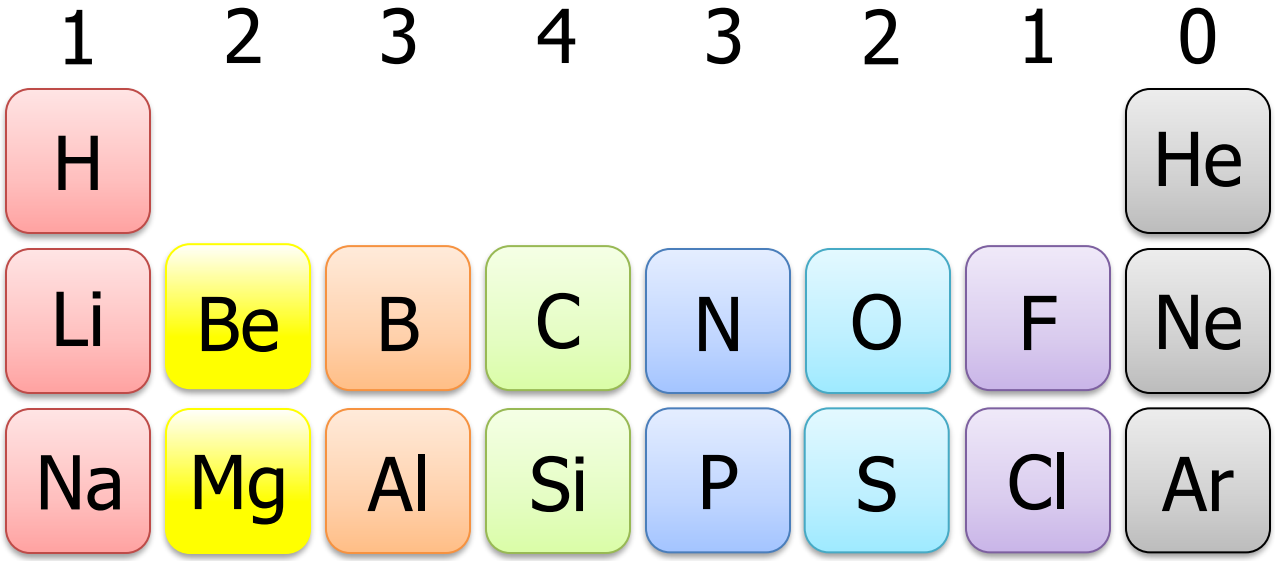
ACKNOWLEDGEMENTS

- Peter Taylor
- Peter Kenny
- Peter Ertl
- Ingvar Lagerstedt
- Thomas Sander
- John Bradshaw
- Darren Green

The NextMove Software Team

PERIODIC TABLE MADE EASY

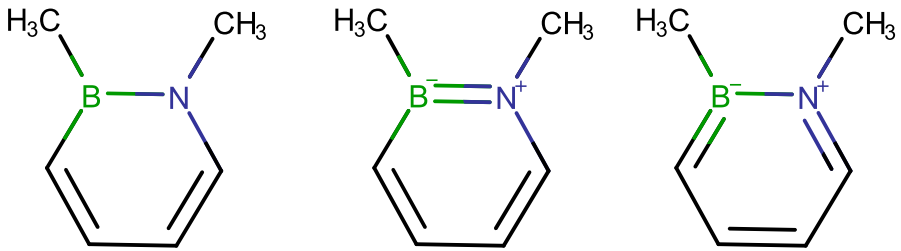
- Dmitri Mendeleev, the father of cheminformatics, introduced a visualization of chemical behaviour.



- N^+ behaves like C, C^- behaves like N, C^+ behaves like B and B^- behaves like C. Likewise, Li^+ AlH_4^-

AROMATIC BORON CHEMISTRY

- OpenSMILES considers $B(III)^{+0}$ to be aromatic (0 πe).
- I propose $B(IV)^{-1}$ to also be aromatic (1 π electron).



- RDKit considers all three structures different.
- Aromatization would make #2 and #3 equivalent.
- Normalization should probably equivalence all 3.