

Triage for Tautomers: from Casualties to Computation

P. J. Taylor

(AstraZeneca, retired)

Triage: the First World War

Casualties are sorted into three categories:

(1) Those who will die anyway and for whom only palliative care is possible;

(2) The “walking wounded” who will survive with minimal attention;

(3) Those for whom proper care will make all the difference to their chances of survival.

It is therefore sensible to put most effort into the third, intermediate, category.

Computational Chemistry: the Triage System for Tautomers

(1) It is *pointless* to plunge into elaborate calculations on compounds of supposed biological or other commercial importance just because they are there, in ignorance of the chemistry and frequently of the literature. The result is to waste time and resources that could be better used elsewhere.

(2) Where it would be easy to establish tautomer ratio by experiment, that experiment should be carried out. While computation here is a legitimate exercise, it would surely be better to add to the presently slim body of accurate data that computer chemists need to calibrate their methodologies.

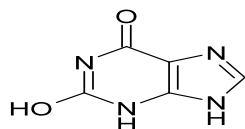
(3) Most effort needs to be concentrated on compounds, sometimes quite simple, which contain structural units that may possess the potential to provide 'building blocks' of wide applicability. These will rarely be glamorous but can be pivotal to our understanding of the phenomena, often poorly documented, that help to determine tautomeric ratio.

XANTHINE: A Cautionary Tale

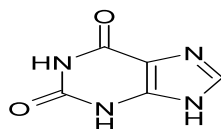
*Tautomer generation. **pKa based** dominance conditions for generating dominant tautomers -*

Josze Szegedi and Ferenc Csizmadia, ChemAxon Ltd., Budapest;
poster presented at ACS Fall National Meeting, August 2007

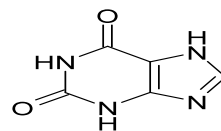
Among other analyses the authors calculated the % population, in aqueous solution, of the 15 possible tautomers of xanthine. Three of these appear below, the first, second and sixth in order of importance:



17A
60%

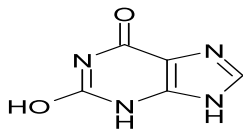


17B
33%

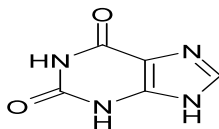


17F
0.55%

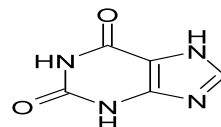
The relation between 17A and 17B requires that an iminol, normally greatly disfavoured, should be twice as important as its parent amide...



17A



17B

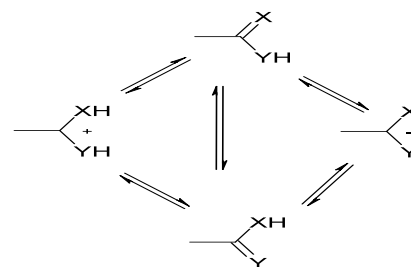
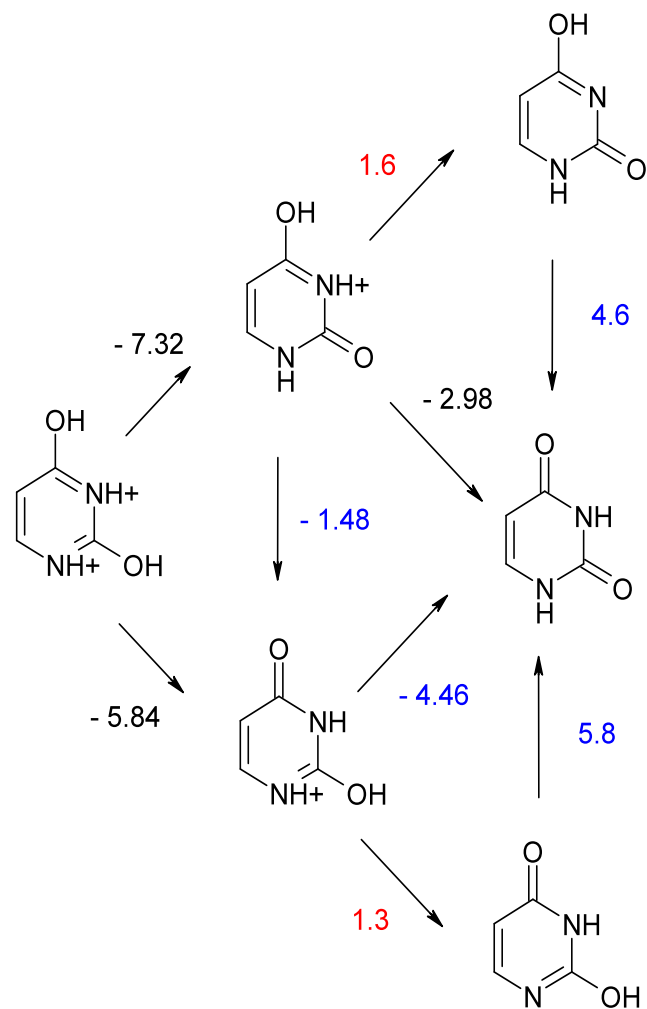


17F

The problem splits neatly into two sections:

(a) 17A / 17B: iminol formation at one amide group or both of the uracil unit, which may or may not be influenced by the imidazole moiety;

(b) 17B / 17F: the effect of a fused uracil unit on the otherwise degenerate tautomerism of the imidazole moiety.

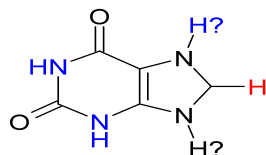


based on: C. D. Poulter and G. D. Frederick, *Tetrahedron Lett.*, 1975, **16**, 2171, with correction factors.

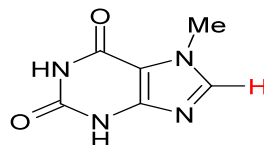
D. Lichtenberg, F. Bergmann and Z. Neimann, *J. Chem. Soc. (C)*, 1971, 1676:

(1) By proton NMR work on all the *N*-methyl derivatives of xanthine it was established that its mobile protons (in aqueous 90% DMSO) occupy those positions shown in **BLUE** below.

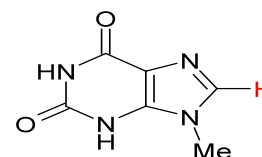
(2) δ Values (ppm) for C-H :



7.85
Xanthine



7.82
7-Me

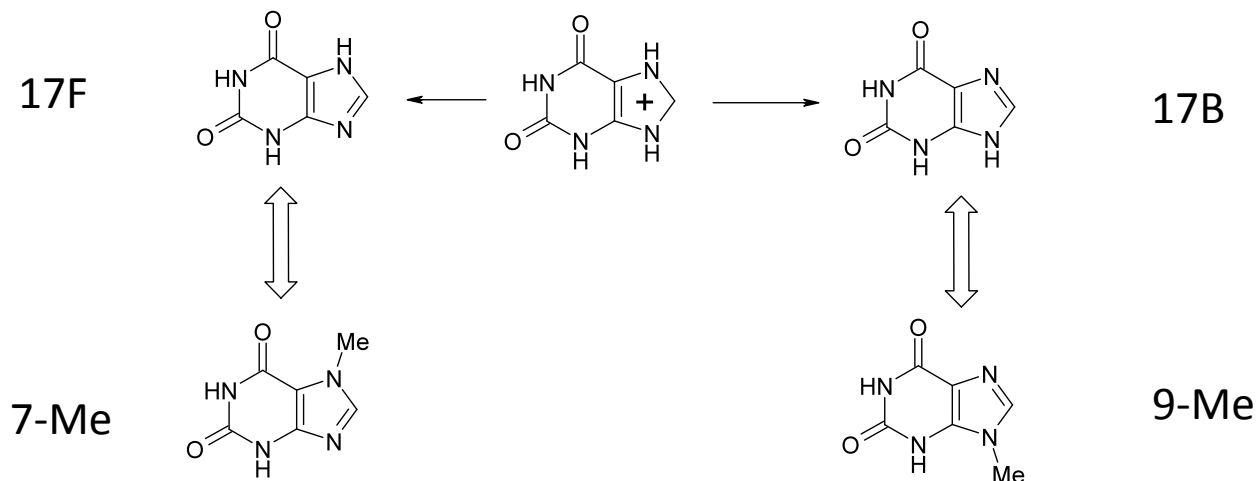


7.55
9-Me

These values unequivocally establish xanthine as the 7-H not the 9-H tautomer. On other solvent evidence DMSO strongly solvates proton donor solutes so, if it has any influence on tautomeric form here, this is likely to be in the direction of exaggerating the importance of the 9-H tautomer with its NH,NH *peri*-interaction.

THESE FACTS HAVE BEEN KNOWN FOR ALMOST 40 YEARS

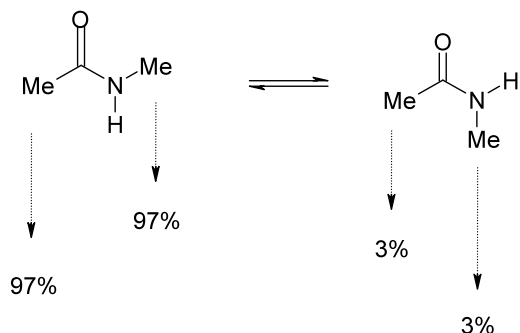
Nevertheless, a trick has been missed:



pK_a 0.8...(by NMR in 90% DMSO!)... pK_a 2.0

The result of the CUP 2009 challenge was to suggest that 17B, by far its closest rival, is disfavoured by most calculations relative to 17F by ΔE *ca.* 1.3 +/- 0.6 kcal mol⁻¹ ($n = 14$), *i.e.* about a factor of 10, two giving much larger margins. This compares with my estimated factor of *ca.* 300. It would be nice to know the truth *in water!* - and not difficult to determine it. Incidentally 17A was disfavoured by all calculations, mostly (13 out of 16) by ΔE in the range 13 - 17 kcal. mol⁻¹.

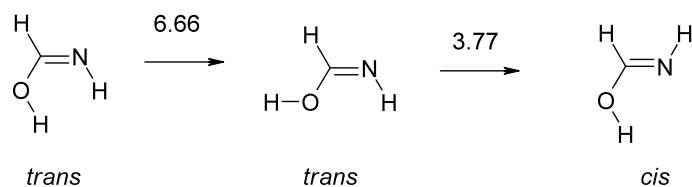
We have slid imperceptibly into the second category. More examples follow:



For *N*-methylacetamide in water at 35 ° :

C-Me singlets at δ 2.05 (97%) and 1.98 (3%)
N-Me doublets at δ 2.75 (97%) and 2.92 (3%)

(R.H. Barker and G. J. Boudreaux, *Spectrochim. Acta Part A*, 1967, **73**, 727)

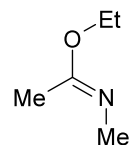


L. Radom, W. J. Hehre and J. A. Pople, *J. Am. Chem. Soc.*, 1971, **93**, 289, at the 4-31G level; results in kcal. mol⁻¹

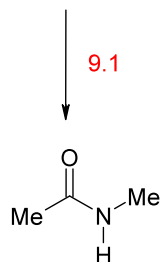
The same *cis*-configuration for *iminol ethers* is that exclusively found by dipole moment measurements in dioxane: O. Exner, in *The Chemistry of Functional Groups, Supplement A : The Chemistry of Double-Bonded Functional Groups, Part 1*, ed. S. Patai, Wiley, New York, 1977, p.1. Hence:

OPEN-CHAIN CIS- AND TRANS-AMIDES SHARE A COMMON IMINOL

pK_a 7.5 (8.5 on correction for NAlk to NH)



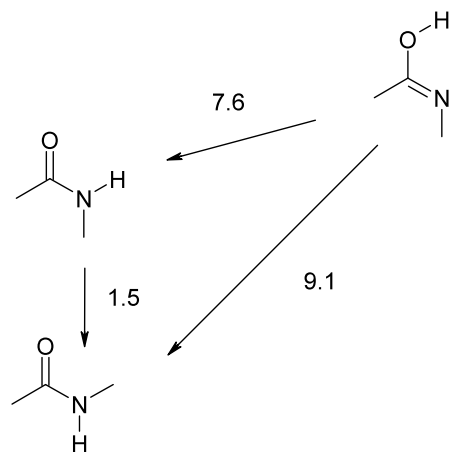
(T. C. Pletcher, S. Koehler and E. H. Cordes, *J. Am. Chem. Soc.*, 1968, **90**, 7072).



9.1

pK_a -0.56

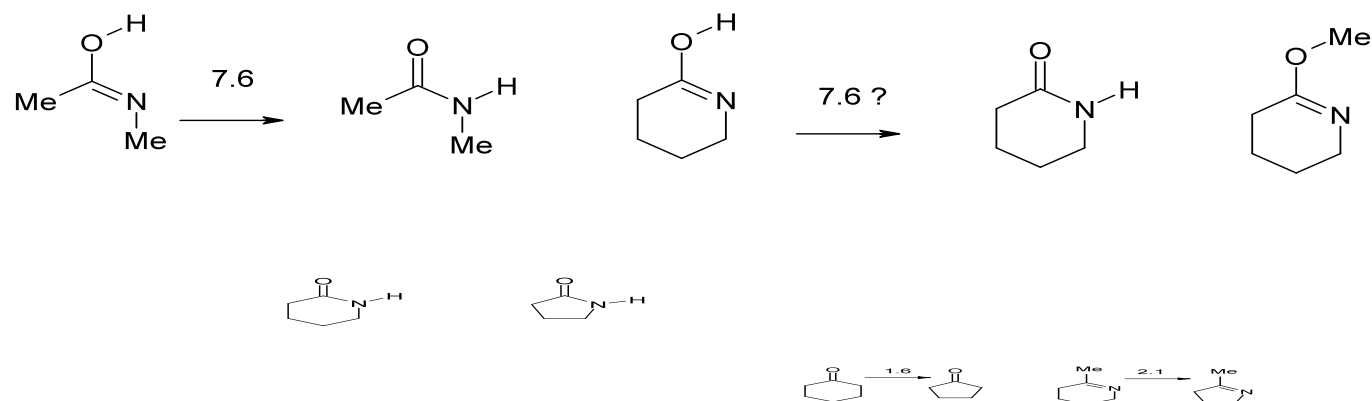
(A. Bagno, G. Lovato and G. Scorrano, *J. Chem. Soc., Perkin Trans. 2*, 1993, 1091).



Do we have here a way of estimating K_T for *cis*-amides?

And can we use it for calculating this for lactams?

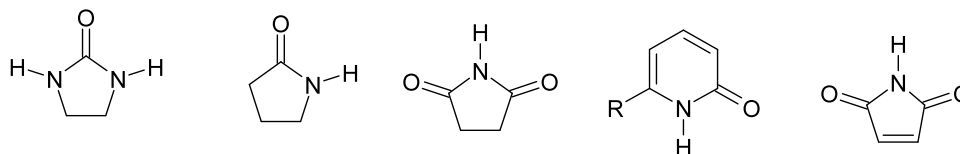
Can we extrapolate $\log K_T$ 7.6 for a *cis*-amide into the same value for a strain-free lactam? And can we push the argument still further?



PJT Guesstimate:	7.6	7.0
SAMPL 2009	8 - 11	10 - 12
(range for n)	(11/16)	(11/16)

...but really, what's the point of calculation when the experiment would be so easy?

Effects on $\log K_T$ of increasing or decreasing electronegativity:



Known value:

3.5 (R = H)

1.3 (R = Cl)

PJT Guesstimate:

(> 7)

7.0

(< 7)

(imide + 4?)

SAMPL 2009:

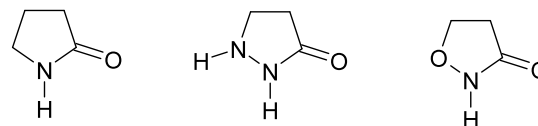
10 - 12

10 - 12

13 - 15

(11/16) (10/16)

(10/16)

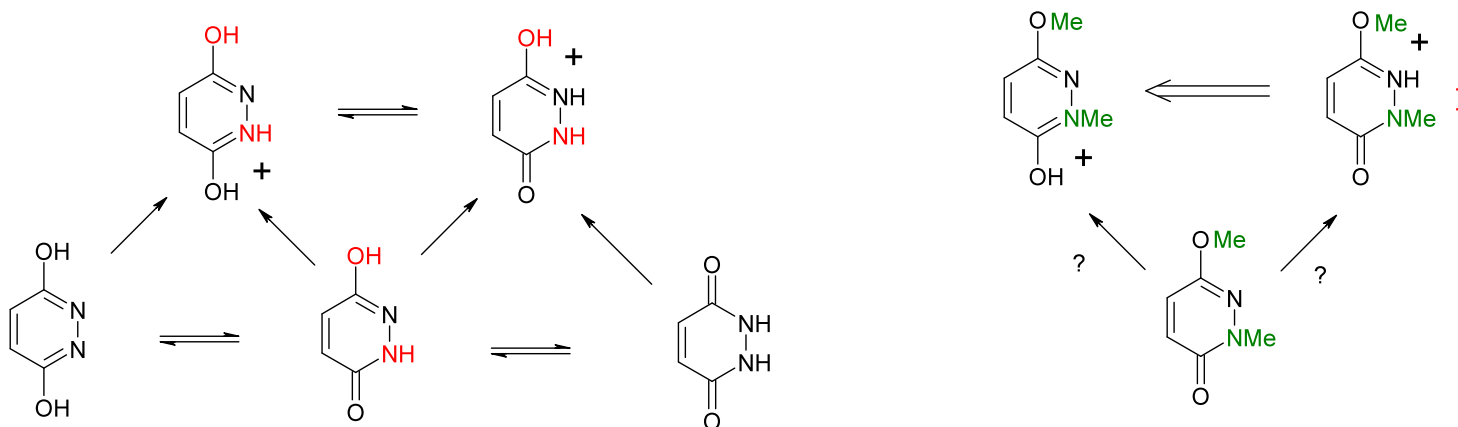


Predicted effect on $\log K_T$ of dipolar repulsion:

- 2.5

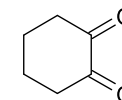
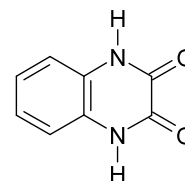
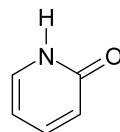
- 5

In all preceding examples, tautomeric form is ambiguous for the neutral species but unambiguous for the common cation. When *both* are ambiguous, real problems arise with the experimental methodology because *it is impossible to design a model compound* which will dictate the site of protonation. At this point, computation becomes a much more valuable option and may be the preferred one. The problem is at its worst in *symmetrical* molecules:



There is *no way* of controlling the site of protonation in the OMe, NMe model compound and the resultant cation appears to be a mixture in which protonation on C=O probably predominates (G. B. Barlin, *J. Chem. Soc., Perkin Trans. 2*, 1974, 1199).

Another possible example of the previous type. Here there are no quantitative data, but on UV evidence oxazine-2,3-dione exists in water as the dioxo-tautomer:



Known or expected $\log K_T$:

3.5

(< 3.5)

$\Delta \log K_T$ for benzofusion:

1.0

$\Delta \log K_T$ for lone pair repulsion:

- PJT guesstimate

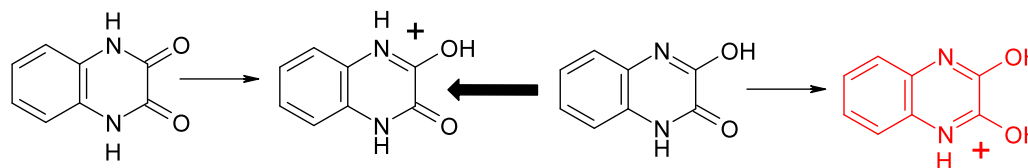
~ -5

- SAMPL 2009 challenge

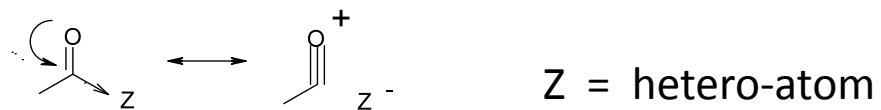
- 4.6 +/- 0.9
(n = 16)

Overall expected $\log K_T$:

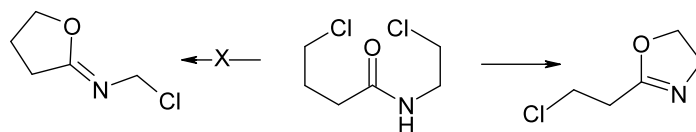
< 0 ?



This time there is no dipolar repulsion and if the **dihydroxy-cation** is disfavoured, this becomes a simple “common cation” type of tautomeric problem.



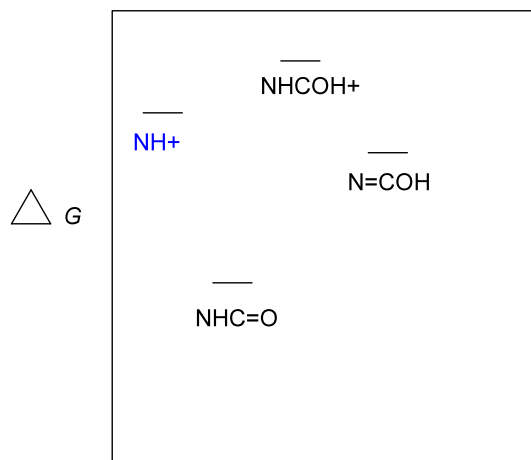
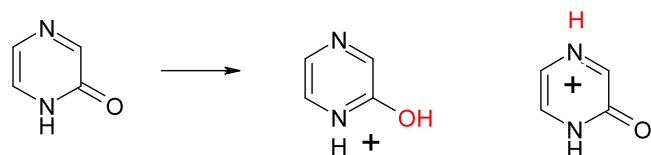
(P. L. Huyskens, H. Marshal and Th. Zeegers-Huyskens, *J. Mol. Struct.*, 1987, **158**, 379).



(from a paper presented by Jack Baldwin at a conference in or about 1980 and explicitly intended to determine which amide lone pair is the more reactive)

The 'basicity method' is valid only if compounds protonate where they are expected to. Suppose they do not?

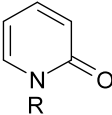
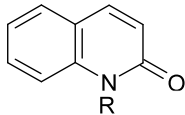
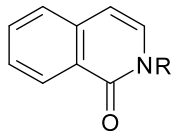
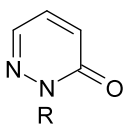
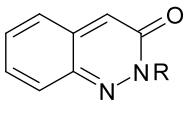
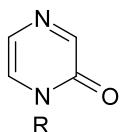
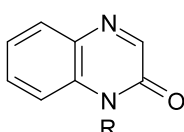
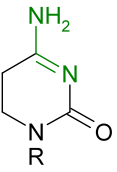
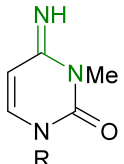
One such possible example is explored below. It provides a prime candidate for computation:



A second site for protonation of the amide is important only if more basic than the first, as shown in the free energy diagram. If that happens its effect is to *reduce* the apparent gap in pK_a between tautomers and result in a spuriously low value for K_T . While UV spectra will *quantitatively* follow the transition from neutral species to cation (or anion), they can fail to provide *qualitative* information on its nature.

Two reasons for suspicion attach to K_T for 2-pyrazinone and 2-oxazinone:

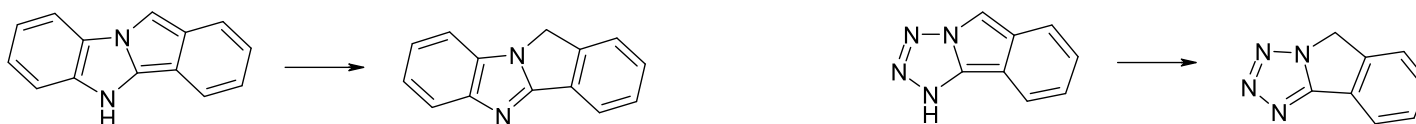
- (1) The drop in K_T relative to 2-pyridone and 2-quinolone is surprisingly large;
- (2) The 'correction factor' Δ (NR) is *negative*; that is, its NMe derivative is a stronger base than the parent compound. This is unique behaviour in the present context but is commonly found for compounds in which alkylation takes place on a nitrogen atom *which does not share the positive charge*:

					
Δ (NR)	0.43	0.40	0.60	0.30	0.82
$\log K_T$	3.5	4.5			
					
Δ (NR)	- 0.06	- 0.22	- 0.30	- 0.30 *	
$\log K_T$	1.9	2.7			

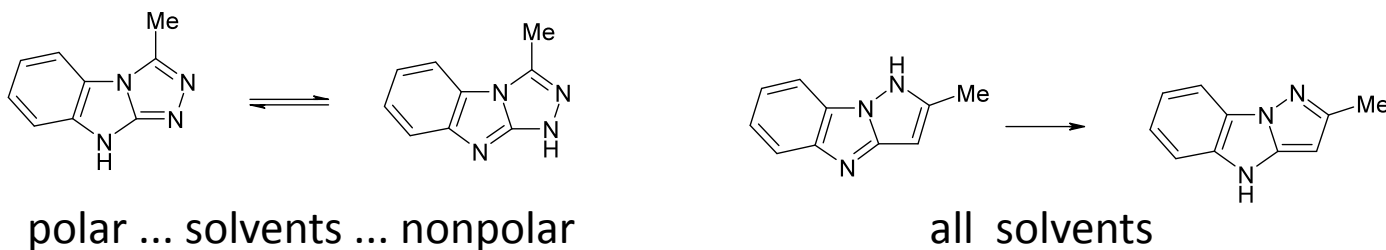
(* The species R = H is a *disfavoured* tautomer whose pK_a was determined by the 'T-jump' method: Dubois *et al.*, *J. Am. Chem. Soc.*, 1976, **98**, 6338).

Tautomer preference in (usually 5,5) ringfused heterocycles, particularly those containing bridgehead nitrogen atoms, forms a useful test bed for theories of aromaticity. “Elguero’s Rules” (J. Elguero, R. M. Claramunt and A. J. H. Summers, *Adv. Het. Chem.*, 1978, **22**, 183) attempt to rationalise what is found:

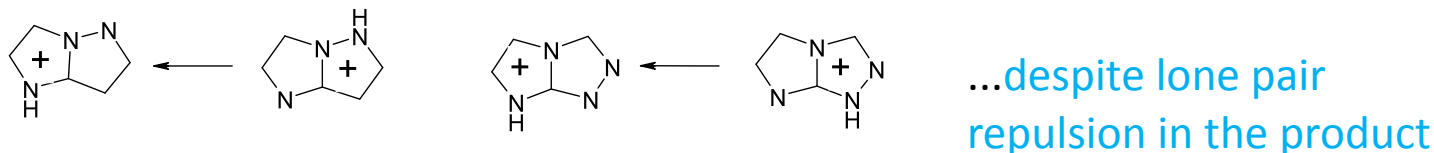
(3) Quinonoid forms are to be avoided:



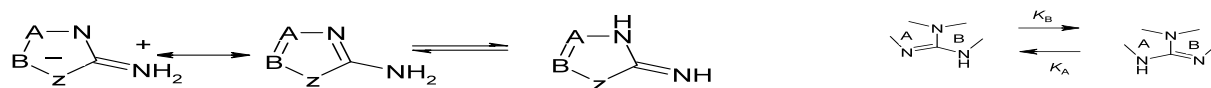
(1) Lone pair and (2) dipolar repulsion are to be avoided:



(4) The *more basic ring* will carry the mobile proton:



Leaving aside the prohibition on quinonoid forms, *all the remaining rules* can be combined, in principle, in a *single rule* that works in the following way:



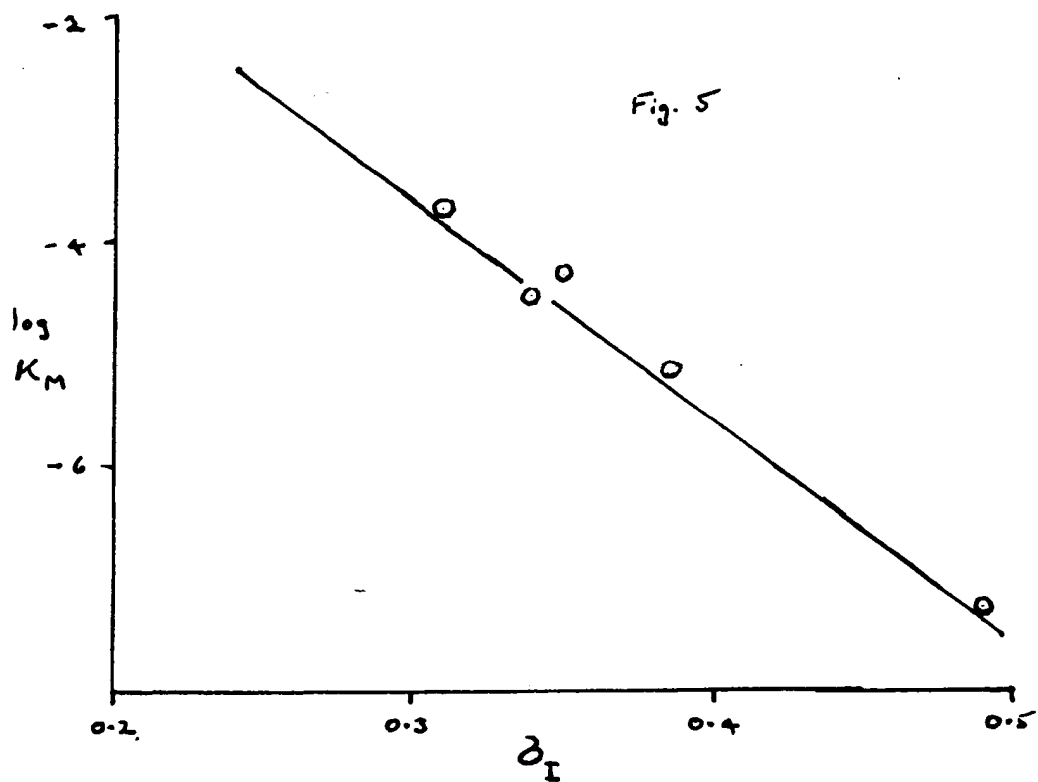
Amine Z

Amine

Imine

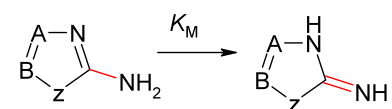
The more electronegative ring will *stabilise the amino-tautomer* so that, on ring fusion, NH will lie in the *less electronegative* ring, as predicted by Elguero. In principle *this can be quantified if the position of equilibrium for both amino-heterocycles is known*. Furthermore, *lone pair and dipolar repulsion will be AUTOMATICALLY taken into consideration*, since these help to determine the pK_a values hence $\log K_T$ of the parent amino- and imino-heterocycles.

In principle, the overall $\log K_T$ for these bicyclic heterocycles with bridgehead nitrogen could be determined by the ‘basicity method’ applied to fixed forms. In practice, it is likely that experimental determination will be difficult or impossible, since (a) many pK_a values will be too low to be readily measurable, (b) their UV spectra are likely to be poorly characterised so unsuitable for quantitative use. Hence *computation is likely to provide the method of choice*.

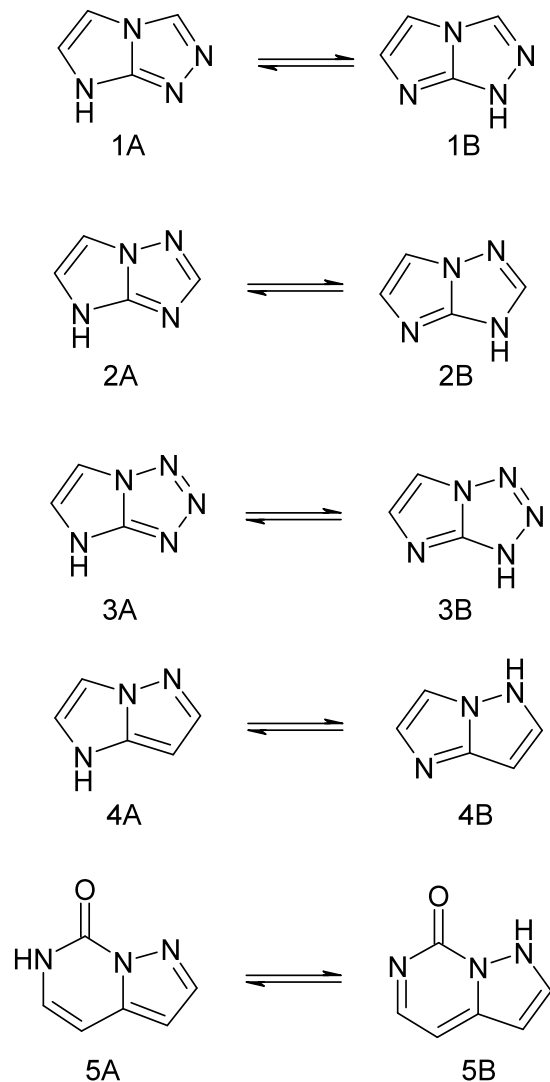


$$\log K_T = 2.38(0.42) - 19.83(1.16) \sigma_I$$

$$(n = 5 \quad r^2 = 0.990 \quad s = 0.19 \quad F = 566)$$

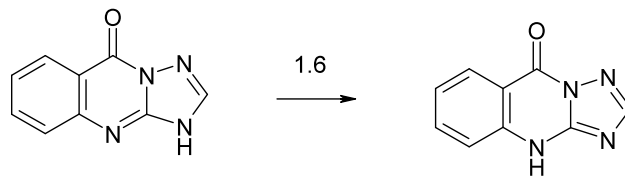


σ_I is measured at the point of attachment to the heterocycle
(P. J. Taylor and A. R. Wait, *J. Chem. Soc., Perkin Trans. 2*, 1986, 1765).



In this potential test series, A is in all cases likely to be the preferred tautomer. For 1 - 3 this is because the right hand ring is the more electronegative, and despite the lone pair clash present in 1A. Nothing is presently known about pyrrole electronegativity but the dipolar clash in 4B is likely to give the same result. It is also known qualitatively that 5A is the preferred tautomer of 5, though perhaps not only for this reason.

SAMPL 2009 presented a related but more complex series of compounds of which that below, where the tautomeric balance is known, is one such example. The present series would provide more basic information.



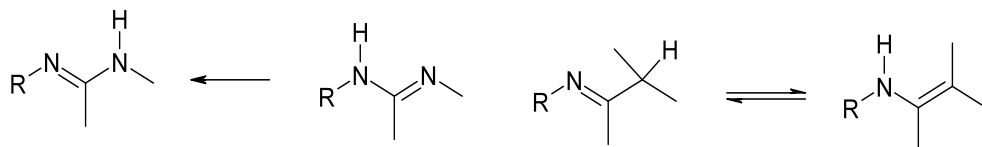
pK_a	0.47	...(NMe)...	- 1.10
λ_{max}	229, 321	(232, 334)	231, 313

(P. J. Taylor, unpublished determination).

And now for something completely different...

All examples so far have been dominated by the *field effect* ('N-type' tautomerism). The remainder are dominated by *resonance* ('C-type' tautomerism).

Contrast *amidine* with *imine-enamine* tautomerism:

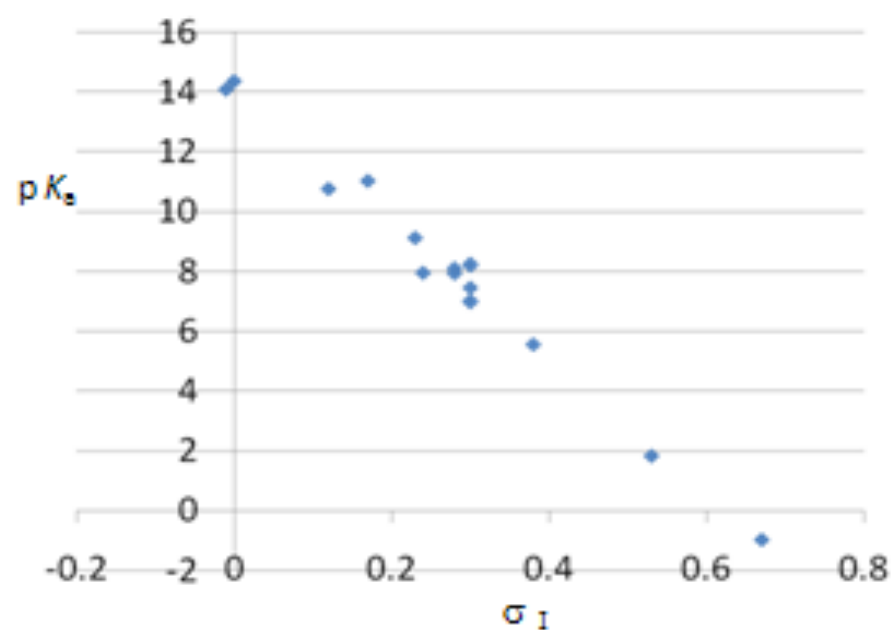
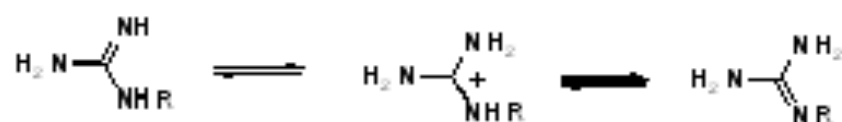


Because virtually all substituents are electron acceptors, *any* group R will cause the RN=C tautomer to predominate (σ_I almost always positive)

Substituents may be resonance *acceptors* or *donors*, with σ_R of either sign; it is found that
Donors stabilise imines;
Acceptors stabilise enamines.

Amidines are well characterised but imine-enamine tautomerism is not. It could do with computational assistance.

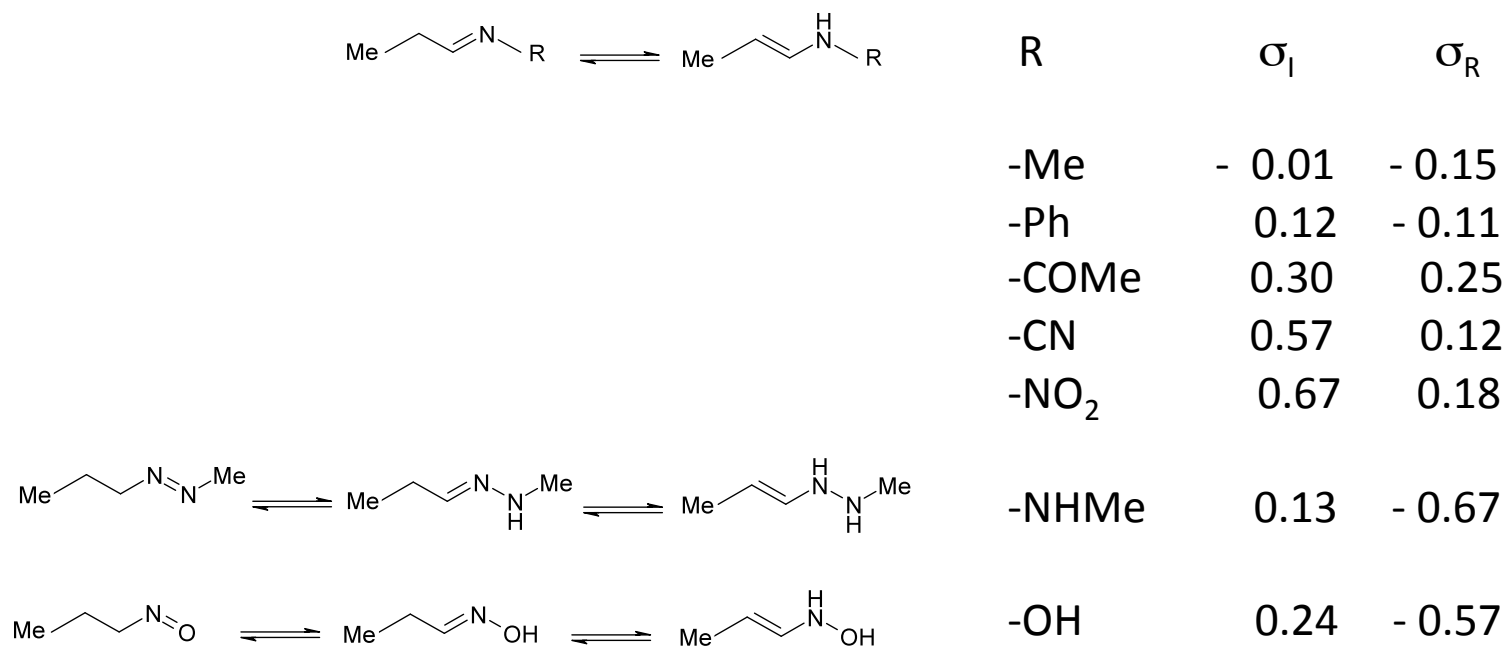
Amidines and Guanidines



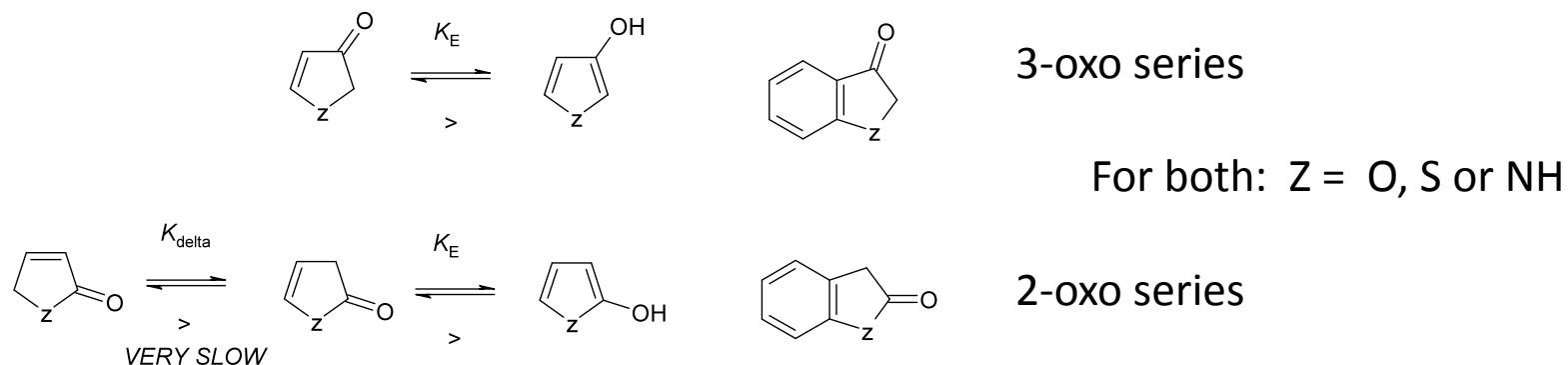
$$\begin{aligned}
 pK_a &= 14.18(0.25) - 22.58(0.78) \sigma_1 \\
 (n &= 16 \quad r^2 = 0.984 \quad s = 0.51 \quad F = 835)
 \end{aligned}$$

P. J. Taylor and A. R. Wait, *J. Chem. Soc., Perkin Trans. II*, 1986, 1765.

A suitable series for starters might be the following:



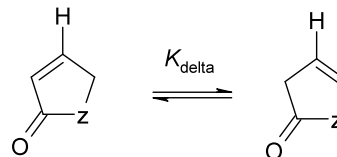
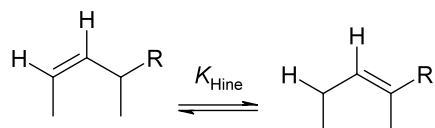
On present evidence, -Me, -Ph, -NHMe and -OH all favour the imine tautomer whereas -COMe converts this to the enamine, but the evidence is very scattered and none is accurately quantitative. The equilibria involving azo- and nitroso-compounds are known to give the imine and oxime tautomers respectively but, again, none of this information is quantitative.



The 3-oxo series has been fully investigated, for both the monocyclic and benzofused sets: B. Capon and F. C. Kwok, *J. Am. Chem. Soc.*, 1989, **111**, 5436.

No quantitative or even systematic work has been reported on the 2-oxo series. This is understandable, for several reasons:

- (1) Both steps are disfavoured, and K_E by probably a much greater margin than in the 3-oxo series;
- (2) The first, isomerisation step is very slow to attain equilibrium;
- (3) For this reason, it is quite likely that measurement of K_E would be complicated by an unpredictable degree of contamination with the overall major tautomer.

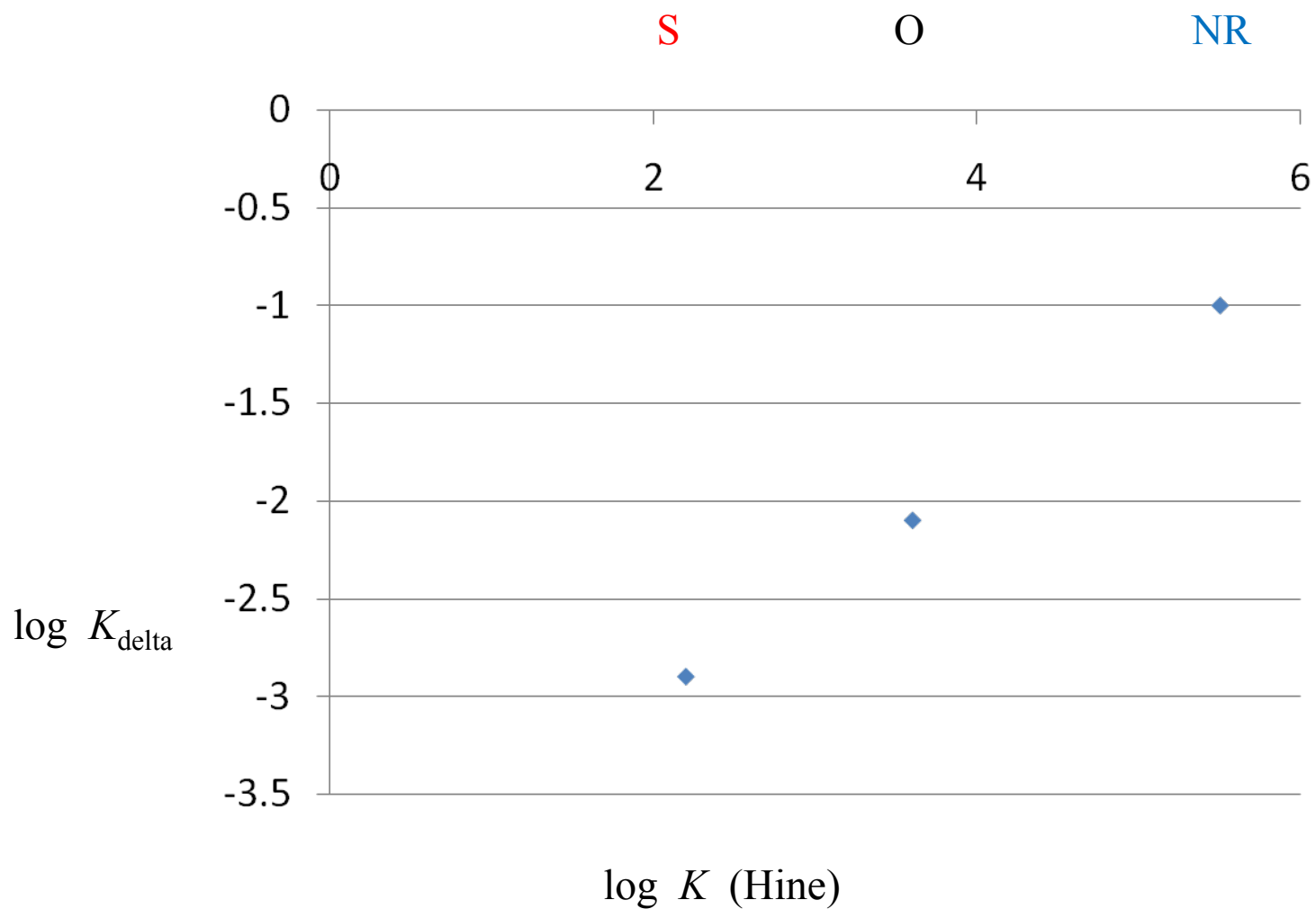


	$\log K_{\text{Hine}}$		$\log K_{\text{delta}}$
-OMe	3.6	Z = O	- 2.1 (estimate) *
-SMe	2.2	Z = S	- 2.9 " **
-NMe ₂	5.5	Z = NH/NMe	- 1.0 (J. T. Baker and S.
-COMe	2.5		Sifniades, <i>J. Org. Chem.</i> ,
-CO ₂ R	2.3		1979, 44 , 2798)
-Me	2.2		
-Ph	3.3		

(J. Hine and M. J. Skoglund, *J. Org. Chem.*, 1982, **47**, 4766; calculated for 25°C and incorporating cross-interaction terms)

For Z = O: W. Friedrichsen, calculated (B3LYP/6-31G) for isolated molecule; quoted by Katritzky *et al.*, *Adv. Het. Chem.*, 2000, **76**, 85.

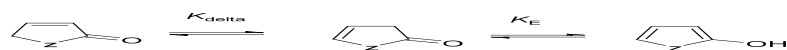
**For Z = S: P. J. Taylor, from K_{delta} values for two 5-substituted 2-oxothiophenes (-2.8, -3.0) whose known K_{Hine} values are assumed to be transferable.



$$\log K_{\text{delta}} = 0.576 \log K (\text{Hine}) - 4.17 \quad (r^2 > 0.999)$$



	$\log K_{\text{delta}}$	$\log K_E$	$\log K_{\text{delta}} K_E$
PJT Guesstimates:			
Z = NH	- 1.0	- 5.5	- 6.5
Z = O	- 2.1	- 6	- 8
Z = S	- 2.9	- 2	- 5
SAMPL 2009:			
Z = NH	- 0.65 +/- 0.15 (n = 12)	- 2.8 +/- 0.8 (n = 7)	- 3.5
		- 6.6 +/- 0.3 (n = 4)	- 7.3
Z = O	- 3.3 +/- 0.3 (n = 12)	- 4.85 +/- 0.6 (n = 7)	- 8.2
Z = S	- 3.4 +/- 0.3 (n = 11)	- 0.8 +/- 0.35 (n = 5)	- 4.2
		- 1.2 +/- 3.2 (n = 16)	- 4.6



Tentative summary - *not* conclusions:

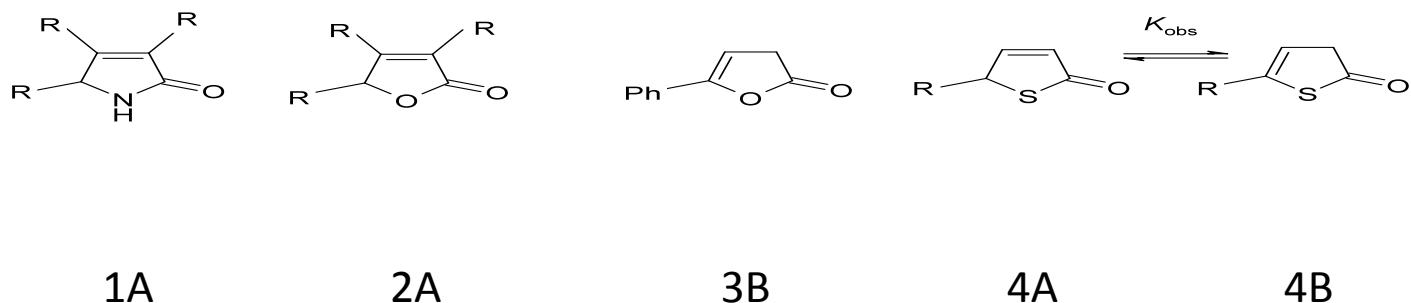
K_{delta} :

(1) Methodologies 288, 289, 320, 330 and 331 may be eliminated *a priori* since all predict $K_{\text{delta}} > 1$ for at least one parent oxoheterocycle.

(2) An approach based on results derived in a variety of ways has given an order in Z of NH > O > S with roughly equidistant spacings which is *quantitatively* matched by Hine's results for simple allylic systems. The SAMPL 2009 results give the same order but with K_{delta} for Z = O only slightly greater, in any calculation, than for Z = S. The problem with this is that the overall process leading to the enol (aromatised tautomer) for Z = O becomes disfavoured to a probably unrealistic extent.

K_E :

In general these calculations look much more reasonable, with Z = O a little more resistant to enolisation than Z = NH and Z = S much more susceptible than either. (Note that some of the empirical evidence is based on their benzofused derivatives, from which the K_{delta} step is absent). However combining K_E with K_{delta} invariably leads to problems, particularly with Z = O.



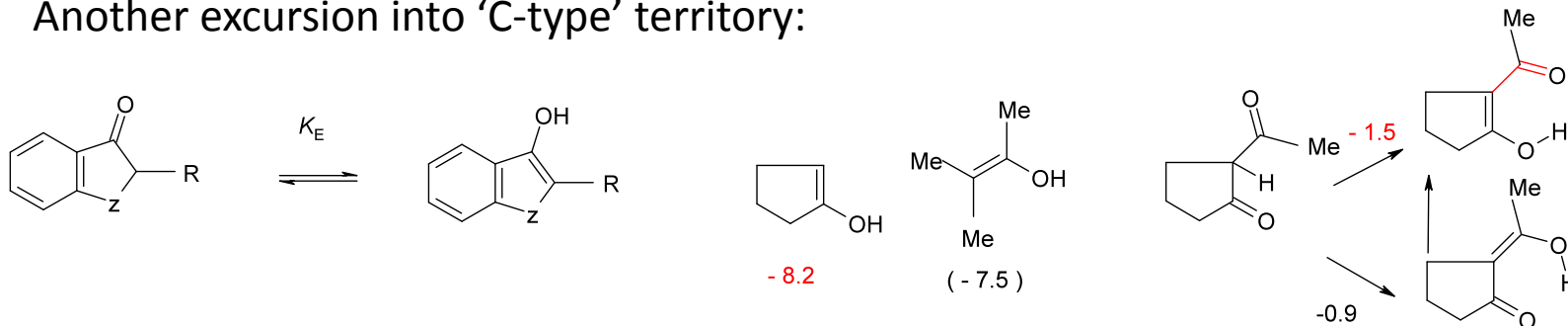
(a) For R = H or Ph, 1A and 1B are the dominant tautomers, hence $\log K_{\text{delta}} - 1$ is about the upper limit for those parent compounds in which Z = O or NH. Opposed effects of a substituent at the 3- and 5-positions should cancel.

(b) A lower limit of $\log K_{\text{delta}} \text{ ca. } 0.8$ for 3 is required for 3B to be its dominant tautomer. Given $\log K_{\text{Hine}} 3.3$ for Ph, this in turn implies a lower limit of $\log K_{\text{delta}} \text{ ca. } -2.5$ for the parent 2-oxofuranone, whose value must therefore lie between -1 and -2.5.

(c) For 4, 4B is again the dominant tautomer. Some 5-substituted derivatives have been examined quantitatively in CCl_4 , of which Me and SMe possess K_{Hine} values. These allow estimates of K_{delta} for the parent 2-oxothiophenone:

	$\log K_{\text{obs}}$	$\log K_{\text{Hine}}$	$\log K_{\text{delta}}$
R = Me	-0.6	2.2	-2.8
R = SMe	-0.8	2.2	-3.0

Another excursion into 'C-type' territory:



R = H

$\log K_E - 4.1$

(B. Capon and F. C. Kwok, *J. Am. Chem. Soc.*, 1989, **111**, 5436)

R = Me

$\Delta \log K_E$ 2.2

- 1.9 ?

R = Ph

3.3

- 0.8 ?

(J. Hine and M. Skoglund, *J. Org. Chem.*, 1982, **47**, 4766; values for simple allylic systems and likely to be lower here through cross-conjugation with OH)

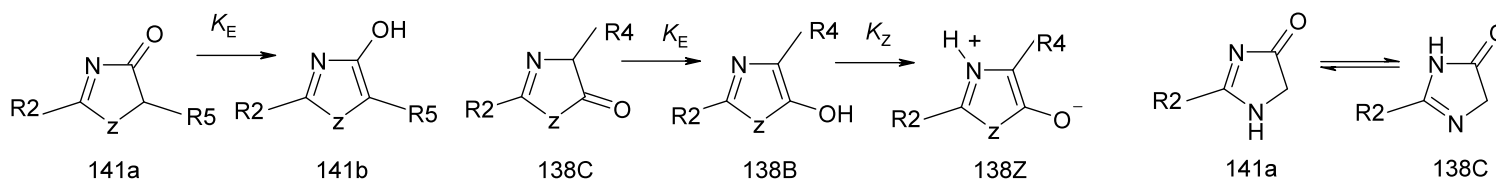
R = COMe

$\Delta \log K_E$ 6.7

2.6 ?

(based on result shown above for 2-acylation of cyclopentanone, the geometry being quite similar)

For calibration purposes it would be necessary to compute K_E for the parent compound as well as its 2-substituted derivatives. More of these could be added...



These two series of 5-membered ring oxoheterocycles present a series of puzzles which are far from being understood:

(1) For 141, $R^2 = R^5 = H$, 141a predominates. However:

(a) Alkyl and, even more, Ph at R^5 generates increasing amounts of 141b in the order $S > N > O$ (as in oxo-derivatives of thiophene, pyrrole and furan).

(b) A π -donor at R^2 will prevent detectable enolisation even for $Z = S$, presumably because of the through-conjugation now present.

(2) The predominant tautomer of 138, $R^2 = R^4 = H$ varies as follows with Z:

(a) For NMe it is mostly 138C; for S it is mostly 138B; for O it is mostly 138Z.

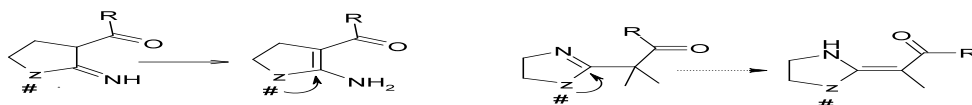
(b) Uniquely, for $Z = NH$ there is an equilibrium between 141a and 138C (shown).

For $R^2 = H$, 138C predominates, but if R^2 is a π -donor, it is locked as 141a. If R^2 is a substituent of type CH_2R where R can generate an enamine from ring $C=N$, such as $NHCH=CHR$, this will happen, as *e.g.* for $R = COX$ and (less certainly) for $R = Ph$.

(c) Substitution at R^4 in 138 has much the same effect as that at R^5 in 141.

There is plenty of scope for computational exploration here!

Lone pair directionality and its influence on tautomeric balance (Z = O or S):



5A

5B

6A

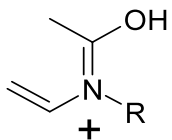
6B

Formation of the conjugated B-tautomer goes to completion to very different extents in these two series. While 5B is formed exclusively, in MeOH, CHCl₃ and the solid state,¹ 6B is formed to a much lesser extent, with 50% in CHCl₃ and 80% in DMSO for Z = S.² Intramolecular hydrogen bonding is weak in both chelates^{1,3} so its energetics do not provide an explanation.

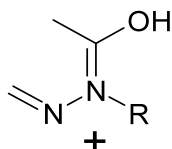
I postulate that *lone pair directionality*, providing a n,π^* interaction that will stabilise 5B but 6A, offers a likely explanation. Such contrasts in behaviour are widespread in 'C-type' tautomerism but have received no comment, let alone investigation. Can computation throw any light on them?

(¹ H. Wamhoff, H. W. Durbeck and P. Sohar, *Tetrahedron*, 1971, **27**, 5873; ² H. Wamhoff and C. Materne, *Liebigs Ann. Chem.*, 1973, 573; ³ P. J. Taylor, unpublished observations).

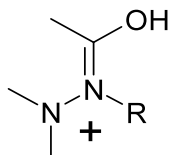
Some correction factors :



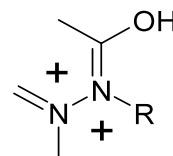
0.45



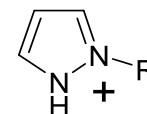
0.55



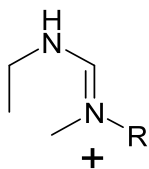
0.6



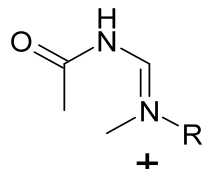
0.6



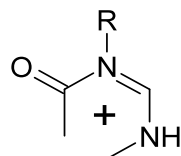
0.75



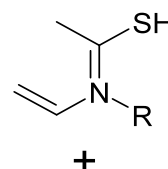
0.3



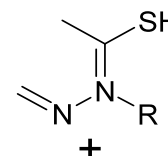
0.3



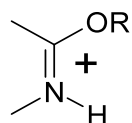
0.0



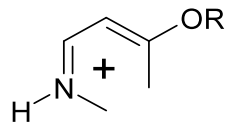
0.25



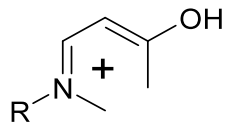
0.4



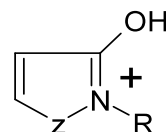
1.0



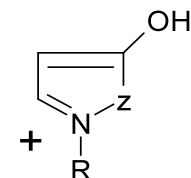
0.5



0.0

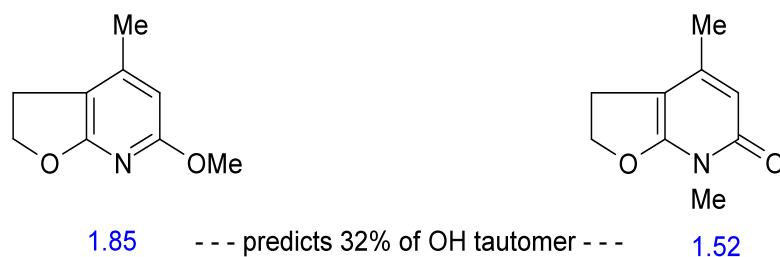
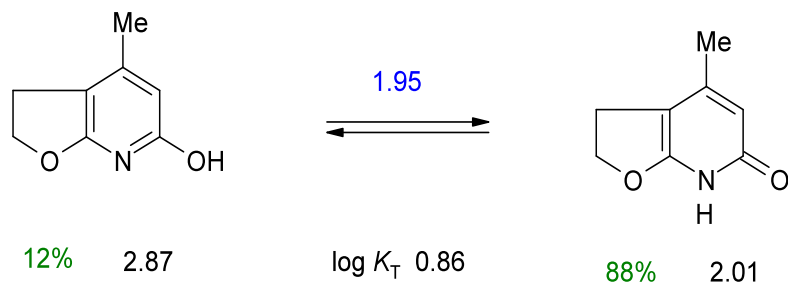


0.45

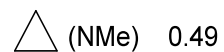


0.15

Quantitative Consequences :



(0.99 +/- 0.15 n = 6)



(0.453 +/- 0.154 n = 19)

(source of data: E. Spinner and G. B. Yeoh, *J. Chem. Soc. (B)*, 1971, 279)

SOLVATOCHROMIC PARAMETERS

$$\log Q = c + s \pi^* + a \alpha + b \beta$$

α : a measure of solvent proton donor ability;

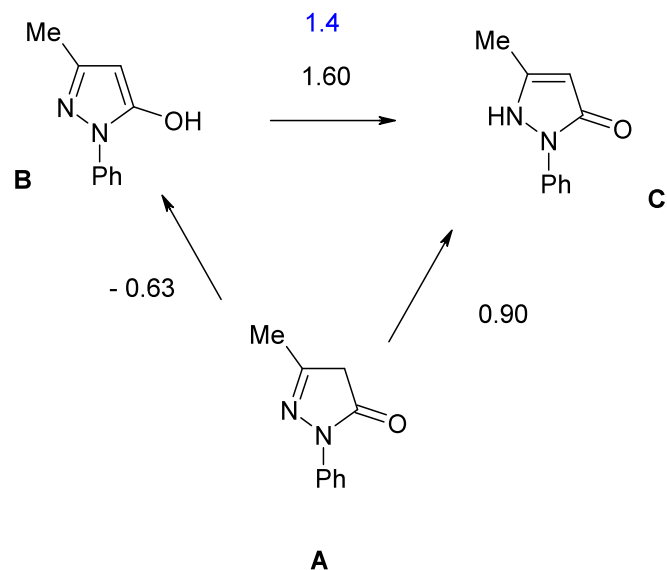
β : a measure of solvent proton acceptor ability;

π^* : some (indefinite) blend of solvent polarity with polarisability.

	Water	MeOH	EtOH	DMSO	CHCl ₃	Dioxane	Benzene	CCl ₄	Alkane	Gas phase
π^*	1.09	0.60	0.54	1.00	0.58	0.55	0.59	0.28	-0.08	-1.23
α	1.17	0.93	0.83	0	0.44	0	0	0	0	0
β	0.40	0.62	0.77	0.76	0	0.37	0.10	0	0	0

NB:

- (1) Water as a proton acceptor is barely stronger than dioxane;
- (2) The gas phase is as far removed from alkane in its ability to solvate polar solutes as the latter is from water.



Data from:

W. Freyer, H. Koppel, R. Radeglia and G. Malewski, *J. prakt. Chem.*, 1983, **325**, 238.

Analysis (PJT) used three independent sets of 9 solvents for the three tautomer ratios.

Requirement :

$$K_E K_T = K_M$$

Basicity Method: A. R. Katritzky and F. W. Maine, *Tetrahedron*, 1964, **20**, 299, 315; with correction factors: O. G. Parchment, D. V. S. Green, P. J. Taylor and I. H. Hillier, *J. Am. Chem. Soc.*, 1993, **115**, 2352.