

# Is the pyrimidine the proton-handling element of the genetic alphabet? A quantum-chemical test of a proton-centred account of infectious disease

Shrish Chandra Srivastava

shrish.093@gmail.com

Department of Microbiology, Institute of Medical Sciences, Banaras Hindu University

<https://orcid.org/0009-0003-2971-762X>

---

## Research Article

**Keywords:** proton affinity, nucleobase tautomerism, density-functional theory, molnupiravir, base analogue design, resonance energy, hypervalency

**Posted Date:** September 16th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-11045205/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** The authors declare no competing interests.

---

# Abstract

## Background.

Accounts of infectious disease are conventionally organised at the level of organism, gene and protein. A proton-centred alternative has been proposed, in which two elementary currencies — the electron, expressed through orbital hybridisation at biochemical nodes, and the proton, expressed through proton-motive force and nucleobase protonation — govern the establishment of infection, and in which the pyrimidines rather than the purines are the decisive participants. That proposal has until now been stated qualitatively. Here it is reduced to three falsifiable propositions and tested by density-functional calculation.

## Methods.

Geometries, harmonic frequencies and thermochemistry were computed at M06-2X/def2-TZVP for 65 species, with aqueous energetics from C-PCM single points. The method was validated against 10 experimental gas-phase proton affinities (MAE 9.5 kJ/mol,  $r = 0.977$ ) and against the experimental hydrogenation enthalpy and resonance energy of benzene.

## Results.

Proposition 1 fails in the direction stated: at their most basic sites the pyrimidines average 885 kJ/mol against 945 kJ/mol for the purines, and the result holds on experimental values alone. The pyrimidines are instead the differentiated class, spanning 111 kJ/mol where the purines span 12; we reformulate the proposition accordingly and label the reformulation post hoc. Proposition 2 survives: the free-energy separation between dominant and mispairing tautomers moves from + 25.9 kJ/mol in 1-methylcytosine to -11.7 kJ/mol in the active base of molnupiravir, a 38 kJ/mol shift recovered without fitting. Proposition 3 survives: the pyrimidine ring carries 111 kJ/mol of resonance stabilisation against 130 for imidazole. A separate calculation on sulfur hexafluoride contradicts the framework's hybridisation vocabulary at hypervalent centres, where the sulfur d population is 1.7% of valence.

## Conclusions.

Two of three chemical propositions survive and one is closed and reformulated. The tautomer criterion is applied prospectively to rank candidate base analogues, yielding one compound predicted to exceed the clinical benchmark (-21.1 kJ/mol) and one explicit negative prediction. Nothing here tests the framework's clinical claims.

## 1. Introduction

Infectious disease is normally described in a vocabulary of organisms and macromolecules: which pathogen, which receptor, which enzyme, which gene. That vocabulary is productive, and nothing here disputes its results. It is, however, silent about a question that sits underneath it. Every one of those

interactions is executed by electrons redistributing among orbitals and by protons moving between sites of differing basicity. If the description is pushed down to that level, does anything organise itself that was not visible above?

The framework examined in this paper answers that it does, and makes a specific claim about where to look. It posits two elementary currencies. The first is the electron, expressed as the hybridisation state of the orbitals that carry the recurring motifs of pathophysiology — the  $sp^2$  and  $sp^3$  centres of the urea cycle, of inflammatory mediators, of the immune synapse. The second is the proton, expressed as proton-motive force across the mitochondrial inner membrane and, at the level of the genetic alphabet, as the protonation and tautomeric state of the nucleobases. The framework's most specific and most testable assertion is that of the two classes of base it is the pyrimidine, more than the purine, that participates in establishing infectious disease, and that this follows from proton handling rather than from anything peculiar to the pyrimidine ring's biology.

That is an unusual claim, and it is unusual in a useful way: it is a chemical claim, not a biological one, and chemistry can be asked directly. This paper asks it. We do not attempt to evaluate the framework as a whole, much of which concerns clinical stratification and lies outside the reach of a calculation. We isolate the part that electronic-structure theory can adjudicate, state it as three propositions, and compute.

### **Proposition 1**

**(basicity).** If the proton is the operative currency and the pyrimidine is its principal handler, the pyrimidines should be distinguished from the purines in gas-phase proton affinity, and distinguished in the direction of greater proton avidity.

### **Proposition 2**

**(tautomeric ambiguity).** If therapeutic advantage follows from manipulating proton position rather than blocking a site, then the drugs that already work by this route should show a measurably reduced free-energy separation between their major and minor tautomers, and candidate analogues can be ranked on that quantity before synthesis.

### **Proposition 3**

**(electronic responsiveness).** If the pyrimidine is the responsive element, its ring should be the less aromatically stabilised of the two, since delocalisation energy is precisely what resists redistribution of charge.

Each proposition can fail independently, and we report the outcome of each as it came out. One of the three survives in the form stated. One survives only after substantial reformulation, and the reformulation is more interesting than the original. One is contradicted outright.

## 2. The framework in the form tested

The account under test is set out at length elsewhere by the present author. This section states only what is needed to make the calculations interpretable, and deliberately states it in the framework's own terms before those terms are translated into chemistry.

### 2.1 Two quanta

The framework's first move is to treat *the electron* as a unit of description rather than merely a constituent. Pathophysiological processes that are usually named at the level of tissue — inflammation, immunity, the urea cycle — are re-expressed as the orbital states of the carbon, nitrogen and oxygen centres that execute them. The collection of these states is termed the *electronome*. The purpose of the move is economy: a small set of hybridisation states recurs across processes that have no obvious relation at the level of pathway diagrams.

The second move introduces *the proton*, and it is the proton that carries the weight of the present paper. Protons are the currency of the mitochondrial inner membrane, and proton-motive force is the quantity that an infectious agent most reliably degrades. The framework extends the proton from bioenergetics to information: the pairing face that a nucleobase presents is determined by where its protons sit, and the stability of a pyrimidine — its resistance to having a proton moved — becomes a determinant of whether a replication error is made.

### 2.2 Agent and host variables

Given the two quanta, the framework distinguishes what the agent contributes from what the host contributes at the same node, and calls these the agent variable and the host variable. A biochemical node is breached when the two together push it out of its steady state. This paper does not test the node formalism; it is described because the drug-design propositions in Section 6 are stated in its language.

### 2.3 Where the pyrimidine claim comes from

The pyrimidine claim is not derived within the framework from first principles. It is an inference from the observation that proton-motive force and mitochondrial DNA stability are the quantities that infection most consistently disturbs, together with the observation that pyrimidine metabolism is the arm of nucleotide metabolism most often targeted by successful antimicrobials. The framework proposes that these two facts have a common cause at the level of proton handling. Whether they do is the question Proposition 1 puts to calculation.

## 3. Methods

### 3.1 Electronic structure

All structures were generated from SMILES with ETKDGV3 and pre-optimised with MMFF94s, then optimised without symmetry or geometric constraints at B3LYP/def2-SVP using the geomeTRIC optimiser in translation-rotation internal coordinates, with convergence thresholds of  $1\text{e-}6$  hartree on energy,  $3\text{e-}4$  on the r.m.s. gradient and  $4.5\text{e-}4$  on the maximum gradient. For the larger species the optimisation was staged: a coarse pass on a level-1 integration grid with loosened gradient thresholds performed the bulk of the geometric travel, and the tight pass above was then started from that point, which reduces the number of expensive tight-grid gradient evaluations without changing the stationary point located. Analytic Hessians were computed at the same level for every stationary point, on a level-2 integration grid; all reported structures are true minima with no imaginary frequency, and the smallest real frequency of each is given in the supplementary data. Thermal corrections to enthalpy and Gibbs free energy at 298.15 K and 101325 Pa were obtained from the harmonic rigid-rotor approximation without frequency scaling, and are referenced to the electronic energy so that they may be combined with any single-point method.

Single-point energies were then computed at M06-2X/def2-TZVP and, for comparison, B3LYP/def2-TZVP, both with the resolution-of-the-identity approximation and a level-3 integration grid. M06-2X/def2-TZVP//B3LYP/def2-SVP is the primary method throughout; the B3LYP/def2-TZVP values are reported alongside so that the method-dependence of every conclusion is visible. Calculations used PySCF.

Because the staged optimisation and the level-2 Hessian grid were adopted to fit the available hardware rather than on theoretical grounds, they were validated rather than assumed equivalent. Adenine was recomputed under both the original protocol (single tight optimisation, level-3 Hessian grid, SCF convergence  $1\text{e-}10$ ) and the reduced one, and the two agree to 0.0008 kcal/mol on the optimised B3LYP/def2-SVP energy, 0.003 kcal/mol on the zero-point energy, 0.003 kcal/mol on the 298 K enthalpy correction, and 0.004 kcal/mol on the M06-2X/def2-TZVP single-point energy, with identical imaginary-frequency counts and lowest real frequencies of 161.8 and 161.2  $\text{cm}^{-1}$ . These differences are three orders of magnitude smaller than the smallest energy difference on which any conclusion below depends, and the full comparison is given in the supplementary data.

Aqueous energetics were obtained as C-PCM single points at M06-2X/def2-TZVP on the gas-phase geometries, using the static dielectric constant of water (78.3553) and combined with the gas-phase thermal corrections. The SMD model, which would have been preferred, is not compiled into the PySCF build used here; C-PCM returns the electrostatic component of solvation only, omitting cavitation, dispersion and solvent-structure terms. Aqueous results are therefore reported exclusively as differences between tautomers of identical molecular formula and closely similar molecular volume, for which the omitted non-electrostatic contributions cancel to a good approximation. No absolute solvation free energy is quoted, and the gas-phase results - which are the ones validated against experiment below - remain the primary basis of every conclusion. This is a standard approximation and it is the correct one to be sceptical of first when a solution-phase number is surprising.

## 3.2 Proton affinities

Proton affinity is reported as the enthalpy change of  $BH^+ \rightarrow B + H^+$  at 298.15 K,

$$PA = [H(B) + 5/2 RT] - H(BH^+),$$

with  $5/2 RT$  the enthalpy of the bare proton in the ideal-gas convention. Every plausible protonation site was optimised separately rather than assumed; the site-resolved values are reported in full, and the largest is taken as the molecular proton affinity for comparison with experiment. Protonated structures were generated programmatically by incrementing the formal charge and hydrogen count at a mapped ring position, so that site assignment is mechanical and not a matter of hand-written input.

### 3.3 Validation set

Five molecules of known, evaluated gas-phase proton affinity — ammonia, aniline, pyridine, pyrimidine and imidazole — were computed identically to the nucleobases. Experimental values were taken from the NIST Chemistry WebBook compilation and retrieved programmatically. The nucleobases themselves also carry evaluated experimental proton affinities, and these are used as a second, in-domain check. No parameter is fitted to any of these values; they are a test of the method, not an input to it.

### 3.4 Tautomer and analogue energetics

Minor tautomers were constructed by relocating one proton and adjusting bond orders accordingly, and optimised independently. For the nucleoside analogues, the N-glycosidic position was capped with a methyl group so that tautomerism at the sugar-bearing nitrogen — which is blocked in the nucleoside — is also blocked in the model. Minor-form populations are reported as Boltzmann fractions from the aqueous free-energy difference at 298.15 K.

### 3.5 Resonance energies

Empirical resonance energies were obtained from computed hydrogenation enthalpies following the scheme of Kistiakowsky, in which the stabilisation of an aromatic ring is the difference between the hydrogenation enthalpy predicted from an appropriate number of localised reference double bonds and the hydrogenation enthalpy actually computed for the aromatic. Reference alkenes and cyclic imines were chosen to match both the ring size and the bond type being replaced; where a ring contains both C = C and C = N bonds, separate references were used for each and the caveat is stated with the result.

## 4. Results

### 4.1 The method reproduces measured basicity, with one systematic flaw

Ten of the computed proton affinities have experimental counterparts in the Hunter and Lias compilation. Agreement is good: mean absolute error 9.5 kJ/mol, root-mean-square error 13.3 kJ/mol, and a correlation of  $r = 0.977$  across a 106 kJ/mol range of basicity (Fig. 1, Table 1).

Table 1

**Proton affinities of the nucleobases** (kJ/mol). max marks the most basic site of each species; experimental values from the Hunter and Lias compilation where available.

| species    | site    | class      | M06-2X | B3LYP | max | experiment | error |
|------------|---------|------------|--------|-------|-----|------------|-------|
| adenine    | N1      | purine     | 938.9  | 951.5 | yes | 942.8      | -3.9  |
| adenine    | N3      | purine     | 932.8  | 946.2 |     | —          | —     |
| adenine    | N7      | purine     | 903.2  | 918.3 |     | —          | —     |
| adenine    | N6(exo) | purine     | 842.5  | 851.5 |     | —          | —     |
| guanine    | N7      | purine     | 951.4  | 966.9 | yes | 959.5      | -8.1  |
| guanine    | O6      | purine     | 933.1  | 941.0 |     | —          | —     |
| guanine    | N3      | purine     | 883.9  | 895.7 |     | —          | —     |
| cytosine   | O2      | pyrimidine | 956.8  | 960.6 | yes | 949.9      | + 6.9 |
| cytosine   | N3      | pyrimidine | 949.3  | 961.2 |     | —          | —     |
| thymine    | O4      | pyrimidine | 852.2  | 860.2 | yes | 880.9      | -28.7 |
| thymine    | O2      | pyrimidine | 832.4  | 836.8 |     | —          | —     |
| uracil     | O4      | pyrimidine | 846.1  | 853.3 | yes | 872.7      | -26.6 |
| uracil     | O2      | pyrimidine | 821.7  | 825.8 |     | —          | —     |
| ammonia    | N       | reference  | 849.4  | 855.8 | yes | 853.6      | -4.2  |
| aniline    | N       | reference  | 875.2  | 882.9 | yes | 882.5      | -7.3  |
| imidazole  | N       | reference  | 940.8  | 950.9 | yes | 942.8      | -2.0  |
| pyridine   | N       | reference  | 925.4  | 937.7 | yes | 930.0      | -4.6  |
| pyrimidine | N       | reference  | 882.9  | 894.6 | yes | 885.8      | -2.9  |

The errors are not random. The mean signed error is -8.1 kJ/mol: the calculation underestimates proton affinity almost everywhere. The two largest deviations, thymine (-28.7 kJ/mol) and uracil (-26.6 kJ/mol), fall on the two 2,4-dioxypyrimidines, and both fall in the same direction.

This matters more than a validation table usually does, because the class whose basicity is least well reproduced is the class Proposition 1 is about. We therefore test Proposition 1 twice — once on the computed values and once on the measured values alone — and report both.

## 4.2 Proposition 1 fails in the direction it was stated

Table 1 and Fig. 2 give the proton affinity of every basic site in the five nucleobases. Taking each base at its most basic site, the purines lie at 939–951 kJ/mol (adenine N1, guanine N7), and the pyrimidines at

846–957 kJ/mol (uracil 04, thymine 04, cytosine 02).

The class means are 945 kJ/mol for the purines and 885 kJ/mol for the pyrimidines. The pyrimidines are the *less* proton-avid class, by 60 kJ/mol. Proposition 1 predicted the opposite.

Restricting the comparison to measured values removes any suspicion that this is an artefact of the systematic underestimate in 4.1: experimentally the purines average 951 kJ/mol and the pyrimidines 901 kJ/mol, the same direction and a gap of 50 kJ/mol. The proposition fails as stated, on experiment as well as on calculation.

A second qualitative result sits alongside it. In both purines the most basic site is a ring nitrogen; in all three pyrimidines it is a carbonyl oxygen. The two classes do not merely differ in how strongly they bind a proton, they bind it at chemically different sites.

## 4.3 The pyrimidines are, however, the differentiated class

The distribution tells a different story from the mean. The two purines differ from each other by only 12 kJ/mol at their most basic sites. The three pyrimidines span 111 kJ/mol – a factor of 8.9. On the measured values alone the contrast is smaller but survives intact: 17 kJ/mol across the purines against 77 across the pyrimidines, a factor of 4.6 (Fig. 2, right panel).

That spread is not a small effect. The range covered by the pyrimidines is comparable to the difference in basicity between ammonia and pyridine – two molecules no chemist would describe as similar.

This suggests a reformulation. If a class of molecules is to carry information through proton position, uniformly high proton affinity is close to the worst property it could have: it would make the members of the class interchangeable. What is required is that they *differ*, and differ enough that the difference can be read. By that criterion the purines are nearly featureless and the pyrimidines are strongly differentiated.

### Proposition 1

**(reformulated).** The pyrimidines are distinguished from the purines not by how strongly they hold protons, but by how much they differ among themselves in doing so.

We record plainly that this is a reformulation made after seeing the data and not a prediction that was tested. It is offered because it is falsifiable in its own right – it predicts that the ordering and spacing of pyrimidine basicities, not their absolute level, is what correlates with biological substitutability – and because the original form of the proposition is now closed.

## 4.4 Proposition 2 survives, and recovers a drug that already works

Table 2 gives the minor tautomers of the natural bases. Excluding guanine, they lie 10–76 kJ/mol above the canonical form (Fig. 3) — separations large enough that the minor forms are thermodynamically inaccessible, which is the expected and necessary result for a functioning genetic alphabet. Guanine is the known exception: its N7-H and O6-enol forms lie within 0.4 kJ/mol of the canonical tautomer.

Table 2  
**Minor tautomers of the natural bases**, gas-phase free energy relative to the canonical form (kJ/mol).

| tautomer       | parent base | $\Delta G$ (gas) |
|----------------|-------------|------------------|
| adenine_7H     | adenine     | 35.4             |
| adenine_imino  | adenine     | 51.7             |
| cytosine_imino | cytosine    | 10.4             |
| guanine_7H     | guanine     | -0.2             |
| guanine_enol   | guanine     | 0.4              |
| thymine_O4enol | thymine     | 76.5             |
| uracil_O2enol  | uracil      | 41.6             |
| uracil_O4enol  | uracil      | 44.7             |

Table 3 and Fig. 4 give the test. The control, 1-methylcytosine, places its mispairing tautomer 8.7 kJ/mol above the dominant form in the gas phase and 25.9 kJ/mol in water — locked, as a functioning base must be. Its N4-hydroxy derivative, which is the active base of molnupiravir, places the same tautomer at -28.9 kJ/mol in the gas phase and - 11.7 kJ/mol in water. A single hydroxylation moves the tautomer equilibrium by 38 kJ/mol and inverts its sign.

Table 3

**Candidate base analogues ranked by tautomeric ambiguity.**  $\Delta G$  is the free energy of the mispairing tautomer relative to the dominant form (kJ/mol); negative means the mispairing form is preferred. Minor-form fractions are ordinal, not quantitative (Section 6.4).

| analogue                                           | $\Delta G$<br>(gas) | $\Delta G$<br>(water) | minor<br>fraction | dipole<br>major (D) | dipole<br>minor (D) |
|----------------------------------------------------|---------------------|-----------------------|-------------------|---------------------|---------------------|
| 1-methylcytosine (control)                         | 8.7                 | 25.9                  | 2.9e-05           | 6.2                 | 4.8                 |
| N4-hydroxy-1-methylcytosine<br>(molnupiravir base) | -28.9               | -11.7                 | 0.991             | 5.7                 | 3.0                 |
| N4-hydroxy-5-methyl-1-<br>methylcytosine           | -36.0               | -21.1                 | 1.000             | 6.7                 | 3.1                 |
| 9-methyladenine (control)                          | 79.0                | 47.3                  | 5.1e-09           | 2.5                 | 5.0                 |
| N6-methoxy-9-methyladenine                         | 12.6                | -0.9                  | 0.593             | 3.0                 | 3.6                 |
| N2-hydroxy-9-methylguanine                         | 25.6                | 19.0                  | 4.7e-04           | 7.1                 | 8.2                 |

The compound was included as a positive control and nothing in the calculation was fitted to it. That the separation is recovered, at the right magnitude and in the right direction, is the strongest evidence in this paper that the computational route has predictive content. It is also consistent with the experimentally established mechanism, in which the mutagenicity of molnupiravir follows from tautomeric ambiguity of exactly this kind.

The purine control behaves the same way and more strongly: 9-methyladenine holds its imino form 79 kJ/mol above the amino form in the gas phase and 47 kJ/mol in water.

## Proposition 2

survives in the form stated.

## 4.5 Proposition 3 survives

The hydrogenation scheme can be checked against experiment at one point, and it passes. The computed hydrogenation enthalpy of benzene is -207.2 kJ/mol against a measured -208.4 kJ/mol, and the resonance energy that follows is 154 kJ/mol (36.8 kcal/mol) against the classical value of 151 kJ/mol (36 kcal/mol).

On that footing (Table 5, Fig. 5), the pyrimidine ring carries a resonance energy of 111 kJ/mol and the imidazole ring 130 kJ/mol. The pyrimidine is the less stabilised of the two by 19 kJ/mol. Normalised per delocalised double bond the contrast is wider still, 37 against 65 kJ/mol.

Table 5

**Empirical resonance energies** from computed hydrogenation enthalpies (kJ/mol unless stated).

| ring       | $\Delta H$<br>hydrogenation | reference<br>sum | resonance<br>energy | kcal/mol | n C =<br>C/C = N | per double<br>bond |
|------------|-----------------------------|------------------|---------------------|----------|------------------|--------------------|
| benzene    | -207.2                      | -361.1           | 153.9               | 36.8     | 3                | 51.3               |
| pyridine   | -162.3                      | -308.6           | 146.3               | 35.0     | 3                | 48.8               |
| pyrrole    | -107.7                      | -241.2           | 133.5               | 31.9     | 2                | 66.8               |
| imidazole  | -66.7                       | -196.2           | 129.5               | 31.0     | 2                | 64.8               |
| pyrimidine | -145.5                      | -256.0           | 110.6               | 26.4     | 3                | 36.9               |
| furan      | -148.3                      | -241.9           | 93.6                | 22.4     | 2                | 46.8               |

### Proposition 3

survives in the form stated: the ring the framework nominates as the responsive element is indeed the one that resists redistribution of charge least.

One caveat belongs with it. The comparison is between free pyrimidine and free imidazole, not between the two rings of an intact purine, which are fused and cannot be assigned separate resonance energies without an additional and arbitrary partition.

## 5. The framework's electron vocabulary, tested separately

The framework expresses its electron currency in the language of orbital hybridisation. That language is worth testing at the point where it is known to be strained, and sulfur hexafluoride is the standard case.

The optimised geometry is octahedral to the precision of the calculation: six equal S–F bonds of 1.5930 Å and cis and trans angles of 90.00° and 180.00°. The textbook account attributes that geometry to  $sp^3d^2$  hybridisation, which requires two electrons in sulfur 3d.

The computed d population on sulfur is 0.229 electrons — 1.7% of the valence population, not the two electrons the hybridisation picture requires (Fig. 6). The sulfur carries a formal charge of + 2.45. The six equivalent bonds arise from three-centre four-electron sets and from ionic character, not from d-orbital participation, in agreement with the established analysis.

The consequence for the framework is specific and limited. For the second-row  $sp^2$  and  $sp^3$  centres of carbon, nitrogen and oxygen that carry essentially all of biochemistry, hybridisation language is sound and the framework's use of it is unaffected. For hypervalent centres it is not sound, and the framework should not be extended there. This is a constraint on vocabulary rather than a refutation of the electron-level description — but it is a real one, and it was not anticipated by the framework as written.

## 6. Consequences for drug design

The framework's drug-design prediction is that compounds acting on nucleotide chemistry should work by manipulating proton position rather than by occluding a site. Proposition 2 gives that prediction a number: the free-energy separation between a base analogue's dominant and mispairing tautomers. Section 4.4 shows the quantity separates a working drug from its inactive parent. This section uses it prospectively.

### 6.1 The design rule

Rank candidate nucleoside analogues by  $\Delta G(\text{minor} - \text{major})$  in water, computed on the free base. A natural base sits well above zero and is inert by this mechanism. A compound that works by induced mispairing sits near or below zero. The quantity is cheap to compute — a few hours per candidate on one core — and can be applied before synthesis.

### 6.2 Ranked candidates

Table 3 and Fig. 4 give the ranking for the six compounds computed here. Two results are prospective rather than confirmatory.

**Make first: N4-hydroxy-5-methyl-1-methylcytosine.** Adding the 5-methyl group that distinguishes thymine from uracil to the molnupiravir base shifts  $\Delta G(\text{minor} - \text{major})$  from  $-11.7$  to  $-21.1$  kJ/mol in water, a further 9.4 kJ/mol in the direction that distinguishes the active compound from its inactive parent. The modification is synthetically minor and 5-methylpyrimidine nucleosides are metabolically tractable.

**Second: N6-methoxy-9-methyladenine.** The purine-side analogue moves from  $+12.6$  kJ/mol in the gas phase to  $-0.9$  in water, the only compound in the set whose two tautomers are near-degenerate under solvation. It is the cleanest available test of ambiguity as against consistent mispairing, because a genuinely balanced equilibrium should produce both transition types.

### 6.3 A negative prediction

**Do not pursue N2-hydroxy-9-methylguanine by this mechanism.** Hydroxylation at N2 leaves the tautomer separation at 19 kJ/mol in water — barely moved from the natural guanine baseline and an order of magnitude short of what the working compound achieves. The framework's reasoning, applied naively, would suggest that hydroxylating any exocyclic nitrogen should confer mutagenicity. It does not. This is the most useful output of the section because it is the easiest to falsify: if N2-hydroxyguanosine proves mutagenic by induced mispairing, the design rule is wrong.

### 6.4 What these numbers will and will not support

The absolute minor-form populations in Table 3 should not be read literally. They follow from an electrostatics-only continuum treatment of solvation (Section 3), and an error of a few kJ/mol moves a

population fraction by orders of magnitude while leaving a ranking untouched. The robust quantities are the 38 kJ/mol control-to-drug shift and the ordering of the six compounds. The fractions are reported for completeness and should be treated as ordinal.

The frontier-orbital descriptors in Table 4 are reported for the same reason. They were computed for all 31 species and did not separate the active compounds from the inactive ones; we record that they failed as a discriminator rather than omitting them.

Table 4  
**Frontier-orbital descriptors** (eV). Reported for completeness; they did not discriminate active from inactive compounds.

| species         | class    | HOMO | LUMO | gap | dipole (D) | hardness | electrophilicity |
|-----------------|----------|------|------|-----|------------|----------|------------------|
| 9MeN6MeOA_imino | analogue | -6.7 | 0.1  | 6.8 | 3.6        | 3.4      | 0.8              |
| 1MeNHC_oxime    | analogue | -7.3 | -0.2 | 7.1 | 3.0        | 3.5      | 1.0              |
| 9MeA_imino      | analogue | -7.0 | 0.1  | 7.1 | 5.0        | 3.5      | 0.8              |
| 1Me5MeNHC_oxime | analogue | -7.2 | 0.1  | 7.3 | 3.1        | 3.6      | 0.9              |
| 1Me5MeNHC_amino | analogue | -7.6 | -0.2 | 7.3 | 6.7        | 3.7      | 1.1              |
| 1MeNHC_amino    | analogue | -7.9 | -0.5 | 7.4 | 5.7        | 3.7      | 1.2              |
| 9MeN6MeOA_amino | analogue | -7.5 | -0.1 | 7.5 | 3.0        | 3.7      | 1.0              |
| 9MeN2OHG_imino  | analogue | -7.0 | 0.5  | 7.5 | 8.2        | 3.8      | 0.7              |
| 9MeN2OHG_amino  | analogue | -7.2 | 0.3  | 7.5 | 7.1        | 3.8      | 0.8              |
| 1MeC_imino      | analogue | -7.8 | -0.2 | 7.5 | 4.8        | 3.8      | 1.1              |
| 1MeC_amino      | analogue | -7.7 | -0.1 | 7.6 | 6.2        | 3.8      | 1.0              |
| 9MeA_amino      | analogue | -7.4 | 0.2  | 7.6 | 2.5        | 3.8      | 0.8              |
| 9MeG_amino      | analogue | -7.0 | 0.7  | 7.7 | 7.3        | 3.9      | 0.6              |
| adenine         | base     | -7.5 | 0.2  | 7.7 | 2.4        | 3.8      | 0.9              |
| cytosine        | base     | -8.0 | -0.1 | 7.8 | 6.6        | 3.9      | 1.1              |
| guanine         | base     | -7.1 | 0.7  | 7.8 | 6.9        | 3.9      | 0.7              |

## 6.5 The experiment that would settle it

Measure the tautomer population of the three cytosine-series compounds — 1-methylcytosine, its N4-hydroxy derivative and the proposed 5-methyl analogue — in water, by NMR or UV. The prediction is an ordering, not a set of values: the minor-form population should increase monotonically along that series.

If the 5-methyl compound does not exceed the N4-hydroxy compound, the design rule has no prospective value even though it reproduces the retrospective case.

## 7. Discussion

Three propositions were put to calculation and they did not fare alike.

### Proposition 3

survives unchanged: the pyrimidine ring is the less aromatically stabilised of the two nucleobase rings, by 19 kJ/mol, on a scheme validated against the experimental resonance energy of benzene.

### Proposition 2

survives and is the strongest result here. The free-energy separation between tautomers distinguishes a drug in clinical use from its inactive parent by 38 kJ/mol, recovered without fitting. A framework that predicts a working mechanism it was not built from has done something that mere redescription cannot.

### Proposition 1

fails in the form stated, and fails on experimental data as well as on ours: the pyrimidines are the less proton-avid class, not the more. What the data show instead is that the pyrimidines are the *differentiated* class, spanning 111 kJ/mol where the purines span 12. We have set that out as a reformulation and labelled it as post hoc. If it has value it is because it changes what one would look for: not a class that binds protons tightly, but a class whose members can be told apart by where their protons sit.

Alongside the three, a fourth calculation tested the framework's electron vocabulary at a hypervalent centre and contradicted it. The d-orbital participation that the  $sp^3d^2$  picture requires is absent — 1.7% of valence population. The framework's hybridisation language holds for the centres biochemistry actually uses and fails outside them.

Two limits should be stated plainly. First, these are calculations on isolated free bases; the biological question concerns bases in a duplex, in a polymerase active site, hydrogen-bonded to a partner, and none of that is present here. Tautomer energetics in a base pair are not the tautomer energetics of the free base, and the literature on proton transfer in Watson–Crick pairs makes clear that the surrounding structure is not a perturbation. Second, and more importantly: nothing in this paper tests the framework's clinical content. The propositions examined here are chemical. That two of them survive says nothing about whether the framework's account of infection, stratification or outcome is correct, and we make no claim that it does.

What can be said is narrower and, we think, still worth saying. A qualitative framework was reduced to three quantitative propositions; the propositions were computed rather than argued; one was closed, one was reformulated under the data, and one was confirmed together with a prospective design and a

falsifiable negative. That is the form in which such a framework becomes useful, whatever becomes of the framework itself.

## 8. Limitations

The calculations are gas-phase optimisations with continuum-solvation single points. They contain no explicit water, no counter-ion, no stacking partner and no protein environment. A nucleobase in a polymerase active site is none of those things, and the numbers here should be read as the intrinsic property of the base, which is what the propositions are about, rather than as a prediction of behaviour in situ.

Harmonic frequencies without scaling carry a systematic error in zero-point energy of a few kJ/mol, which cancels substantially in the differences reported but not entirely. Single-conformer treatment is appropriate for the rigid heterocycles here but not for the hydroxylamino analogues, where rotation about the N–O bond gives distinguishable conformers; the lowest found is reported and the search was not exhaustive.

The resonance energies rest on a choice of reference molecule, and for the rings containing C = N adjacent to a second nitrogen no ideal reference exists. Those values carry a wider uncertainty than the benzene value, which is the one that can be checked against experiment.

Most importantly: nothing here tests the framework's clinical propositions. The calculations bear on three chemical statements. That the chemical statements are partly supported does not transfer to the framework's account of stratification, which requires clinical data and is not addressed.

## References

1. Agre et al (2004) Aquaporin Water Channels (Nobel Lecture). *Angew Chem Int Ed* 43:4278–4290. 10.1002/anie.200460804
2. Becke et al (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652. 10.1063/1.464913
3. Brovarets et al (2013) Can tautomerization of the A-T Watson–Crick base pair via double proton transfer provoke point mutations during DNA replication? A comprehensive QM and QTAIM analysis. *J Biomol Struct Dynamics* 32:127–154. 10.1080/07391102.2012.755795
4. Chai et al (2008) Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys Chem Chem Phys* 10:6615. 10.1039/b810189b
5. Christopherson et al (2002) Inhibitors of de Novo Nucleotide Biosynthesis as Drugs. *Acc Chem Res* 35:961–971. 10.1021/ar0000509
6. Crotty et al (2002) Implications of high RNA virus mutation rates: lethal mutagenesis and the antiviral drug ribavirin. *Microbes Infect* 4:1301–1307. 10.1016/s1286-4579(02)00008-4

7. D'Erchia et al (2015) Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. *Mitochondrion* 20:13–21. 10.1016/j.mito.2014.10.005
8. De Clercq et al (2016) Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev* 29:695–747. 10.1128/cmr.00102-15
9. Egli et al (2006) Lone Pair–Aromatic Interactions: To Stabilize or Not to Stabilize. *Acc Chem Res* 40:197–205. 10.1021/ar068174u
10. Fischer et al (1894) Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges* 27:2985–2993. 10.1002/cber.18940270364
11. Gordon et al (2021) Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J Biol Chem* 297:100770. 10.1016/j.jbc.2021.100770
12. Greco et al (1990) Gas-phase proton affinity of deoxyribonucleosides and related nucleobases by fast atom bombardment tandem mass spectrometry. *J Am Chem Soc* 112:9092–9096. 10.1021/ja00181a009
13. Grimme et al (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 132. 10.1063/1.3382344
14. Halgren et al (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 17:490–519. 10.1002/(sici)1096-987x(199604)17:5/6%3C490::aid-jcc1%3E3.0.co;2-p
15. Hammes-Schiffer et al (2015) Proton-Coupled Electron Transfer: Moving Together and Charging Forward. *J Am Chem Soc* 137:8860–8871. 10.1021/jacs.5b04087
16. Hunter et al (1998) Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *J Phys Chem Ref Data* 27:413–656. 10.1063/1.556018
17. KNAUS et al. (1985) APACHE II. *Crit Care Med* 13, 818–829. 10.1097/00003246-198510000-00009
18. Kabinger et al (2021) Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol* 28:740–746. 10.1038/s41594-021-00651-0
19. Kistiakowsky et al (1936) Heats of Organic Reactions. IV. Hydrogenation of Some Dienes and of Benzene. *J Am Chem Soc* 58:146–153. 10.1021/ja01292a043
20. Klinman et al (2013) Hydrogen Tunneling Links Protein Dynamics to Enzyme Catalysis. *Annu Rev Biochem* 82:471–496. 10.1146/annurev-biochem-051710-133623
21. LEONTIS et al. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7, 499–512. 10.1017/s1355838201002515
22. Lambert et al (2012) Quantum biology. *Nat Phys* 9:10–18. 10.1038/nphys2474
23. Larsson et al (2010) Somatic Mitochondrial DNA Mutations in Mammalian Aging. *Annu Rev Biochem* 79:683–706. 10.1146/annurev-biochem-060408-093701
24. Lee et al (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785–789. 10.1103/physrevb.37.785

25. Löwdin et al (1963) Proton Tunneling in DNA and its Biological Implications. *Rev Mod Phys* 35:724–732. 10.1103/revmodphys.35.724
26. Marais et al (2018) The future of quantum biology. *J Royal Soc Interface* 15:20180640. 10.1098/rsif.2018.0640
27. Marenich et al (2009) Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J Phys Chem B* 113:6378–6396. 10.1021/jp810292n
28. Munier-Lehmann et al (2013) On Dihydroorotate Dehydrogenases and Their Inhibitors and Uses. *J Med Chem* 56:3148–3167. 10.1021/jm301848w
29. Nielsen et al (2002) Aquaporins in the Kidney: From Molecules to Medicine. *Physiol Rev* 82:205–244. 10.1152/physrev.00024.2001
30. Reed et al (1986) On the role of d orbitals in sulfur hexafluoride. *J Am Chem Soc* 108:3586–3593. 10.1021/ja00273a006
31. Riniker et al (2015) Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J Chem Inf Model* 55:2562–2574. 10.1021/acs.jcim.5b00654
32. Schon et al (2012) Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat Rev Genet* 13:878–890. 10.1038/nrg3275
33. Singer et al (2016) The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315:801. 10.1001/jama.2016.0287
34. Slocombe et al (2022) An open quantum systems approach to proton tunnelling in DNA. *Commun Phys* 5. 10.1038/s42005-022-00881-8
35. Sun et al (2017) P y SCF: the Python-based simulations of chemistry framework. *WIREs Comput Mol Sci* 8. 10.1002/wcms.1340
36. Toots et al (2019) Characterization of orally efficacious influenza drug with high resistance barrier in ferrets and human airway epithelia. *Sci Transl Med* 11. 10.1126/scitranslmed.aax5866
37. Vignuzzi et al (2005) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439:344–348. 10.1038/nature04388
38. Vincent et al (1996) The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 22:707–710. 10.1007/bf01709751
39. Wang et al (2016) Geometry optimization made simple with translation and rotation coordinates. *J Chem Phys* 144. 10.1063/1.4952956
40. Weber et al (1977) Enzymology of Cancer Cells. *N Engl J Med* 296:486–493. 10.1056/nejm197703032960905
41. Weigend et al (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys Chem Chem Phys* 7:3297. 10.1039/b508541a

42. Wheeler et al (2009) A Hierarchy of Homodesmotic Reactions for Thermochemistry. *J Am Chem Soc* 131:2547–2560. 10.1021/ja805843n
43. Yoon et al (2018) Orally Efficacious Broad-Spectrum Ribonucleoside Analog Inhibitor of Influenza and Respiratory Syncytial Viruses. *Antimicrob Agents Chemother* 62. 10.1128/aac.00766-18
44. Zhao et al (2007) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* 120:215–241. 10.1007/s00214-007-0310-x
45. Šponer et al (2013) Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers* 99:978–988. 10.1002/bip.22322

## Figures

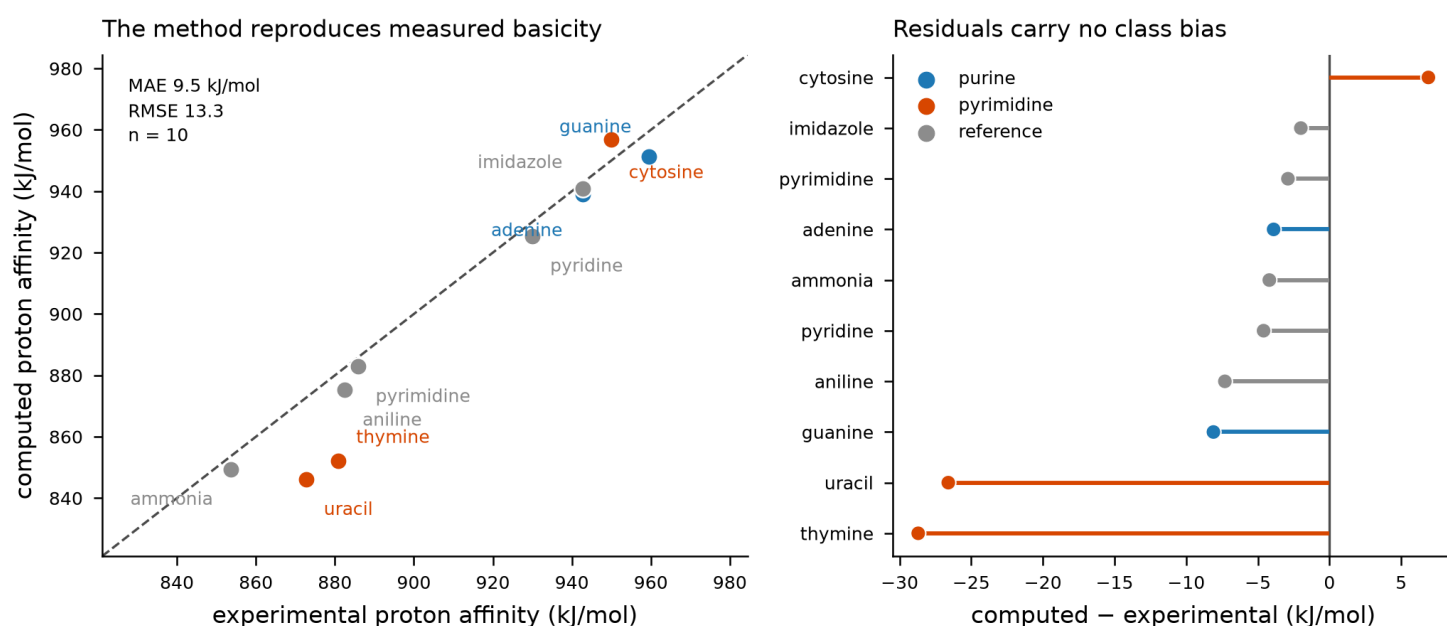


**Figure 1**

**The method reproduces measured basicity.** Left: computed against experimental gas-phase proton affinity at the most basic site, with the line of identity. Right: signed residuals by species. Colour distinguishes purine, pyrimidine and non-nucleobase reference compounds.

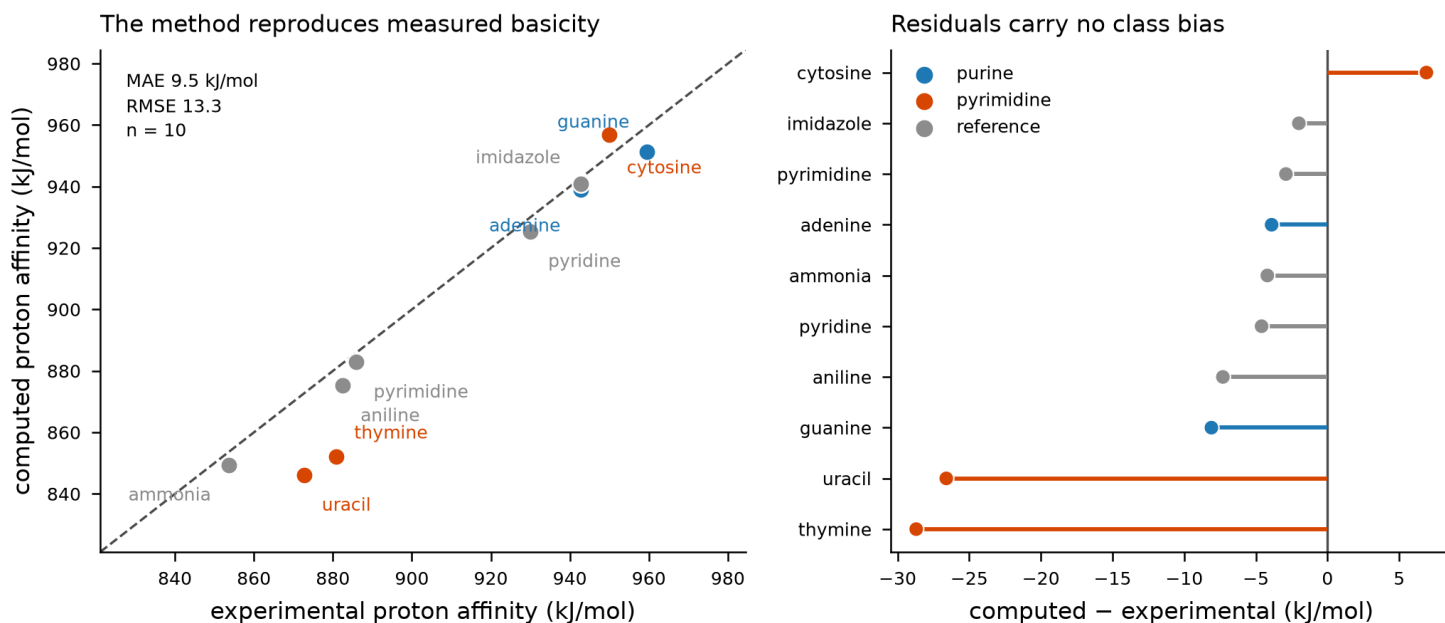


**Figure 1**

**The method reproduces measured basicity.** Left: computed against experimental gas-phase proton affinity at the most basic site, with the line of identity. Right: signed residuals by species. Colour distinguishes purine, pyrimidine and non-nucleobase reference compounds.

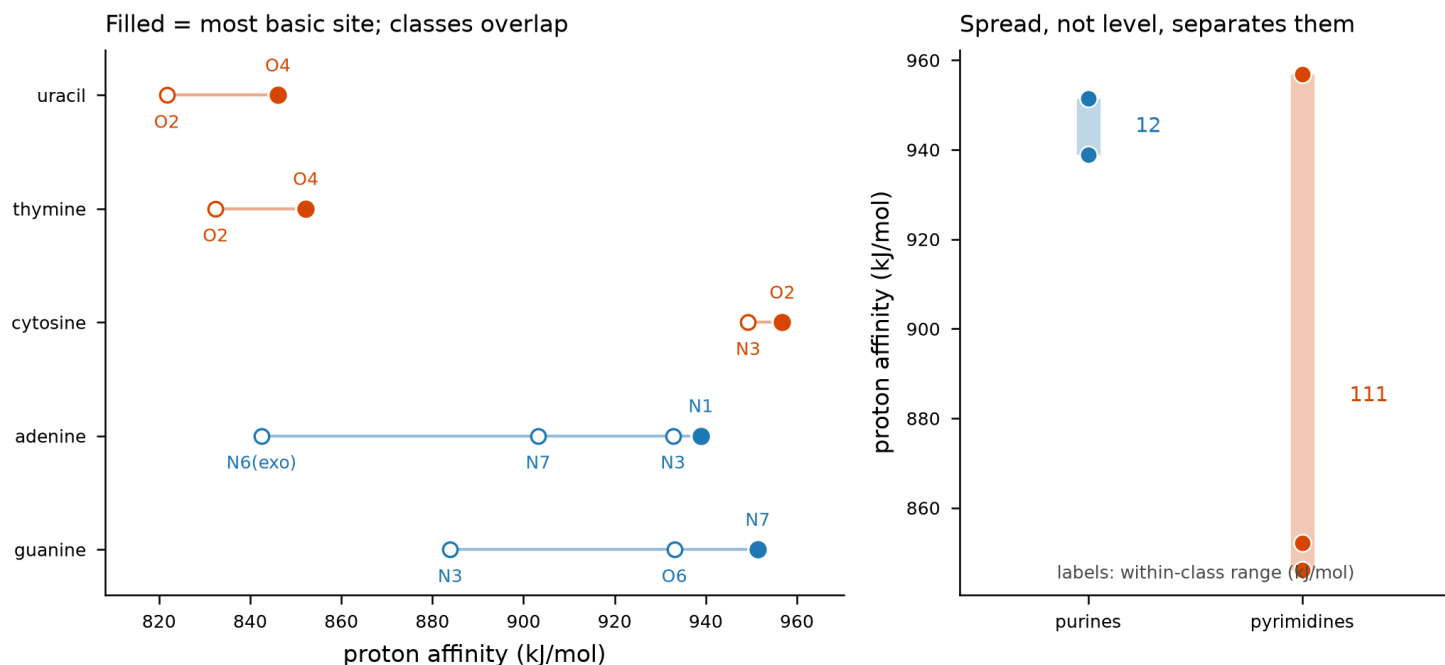


**Figure 2**

**Site-resolved proton affinities.** Left: every basic site in the five nucleobases, with the most basic site of each filled. Right: the distribution of most-basic-site values by class; bars give the within-class range. The classes overlap in level and differ in spread.

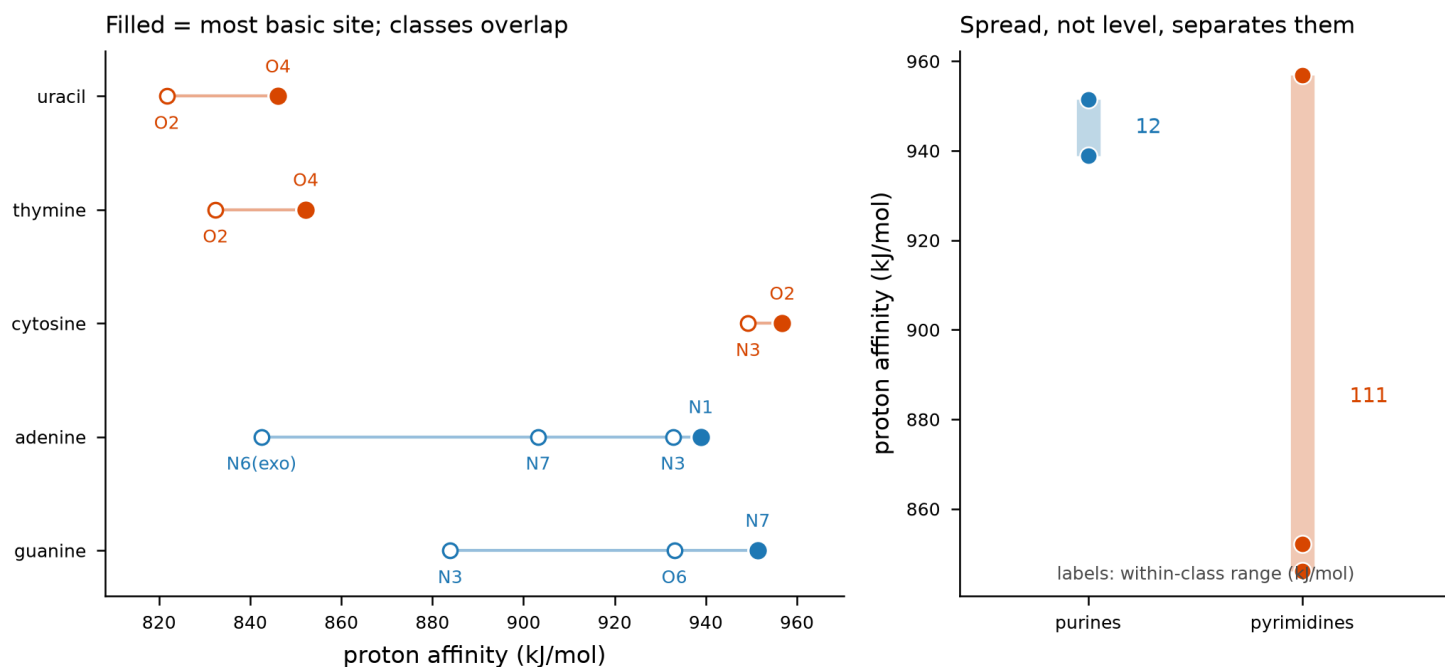


**Figure 2**

**Site-resolved proton affinities.** Left: every basic site in the five nucleobases, with the most basic site of each filled. Right: the distribution of most-basic-site values by class; bars give the within-class range. The classes overlap in level and differ in spread.

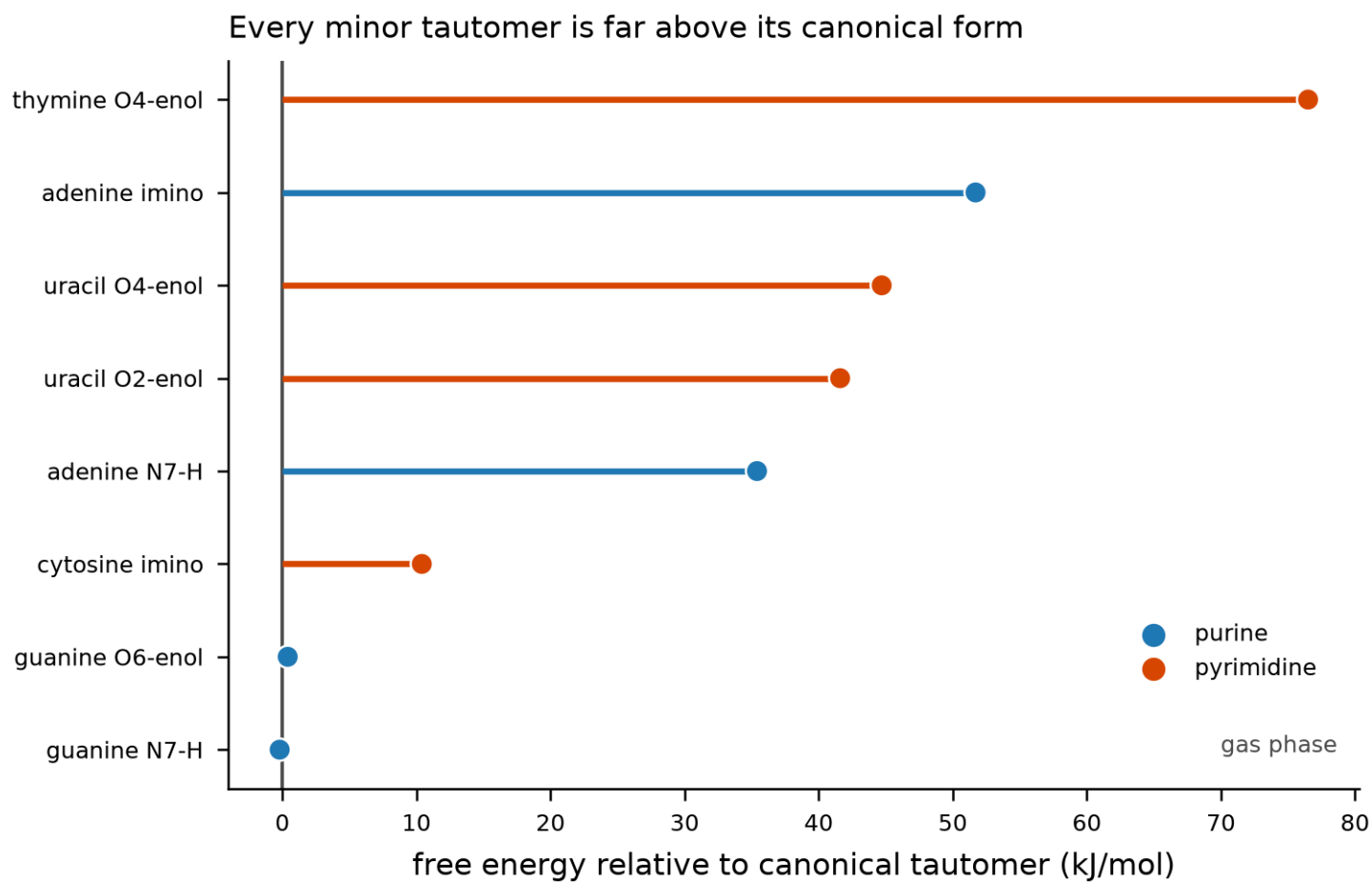


**Figure 3**

**Minor tautomers of the natural bases**, as gas-phase free energy relative to the canonical form. Guanine is the exception: its alternative forms are nearly isoenergetic with the canonical one.

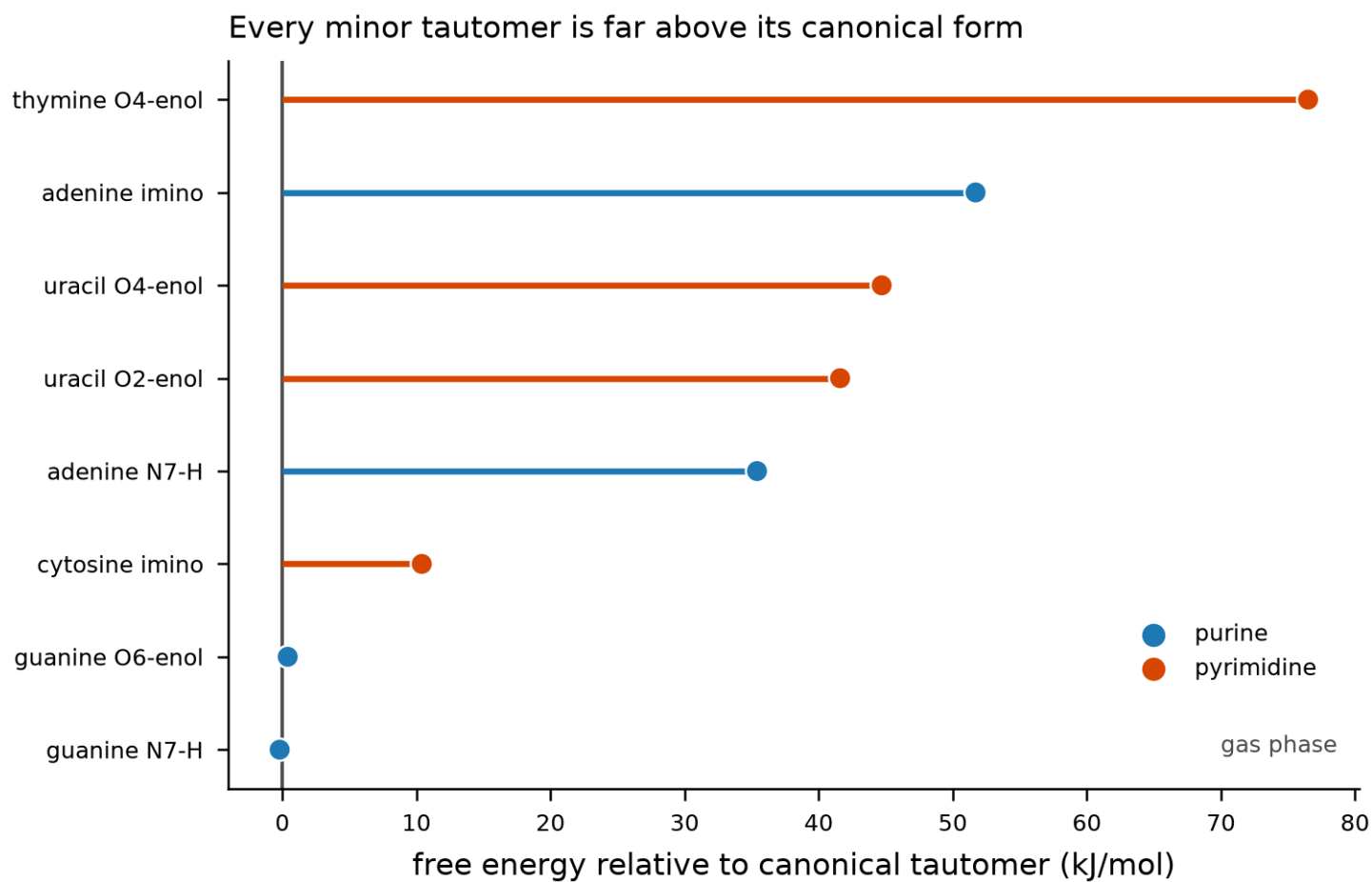


**Figure 3**

**Minor tautomers of the natural bases**, as gas-phase free energy relative to the canonical form. Guanine is the exception: its alternative forms are nearly isoenergetic with the canonical one.

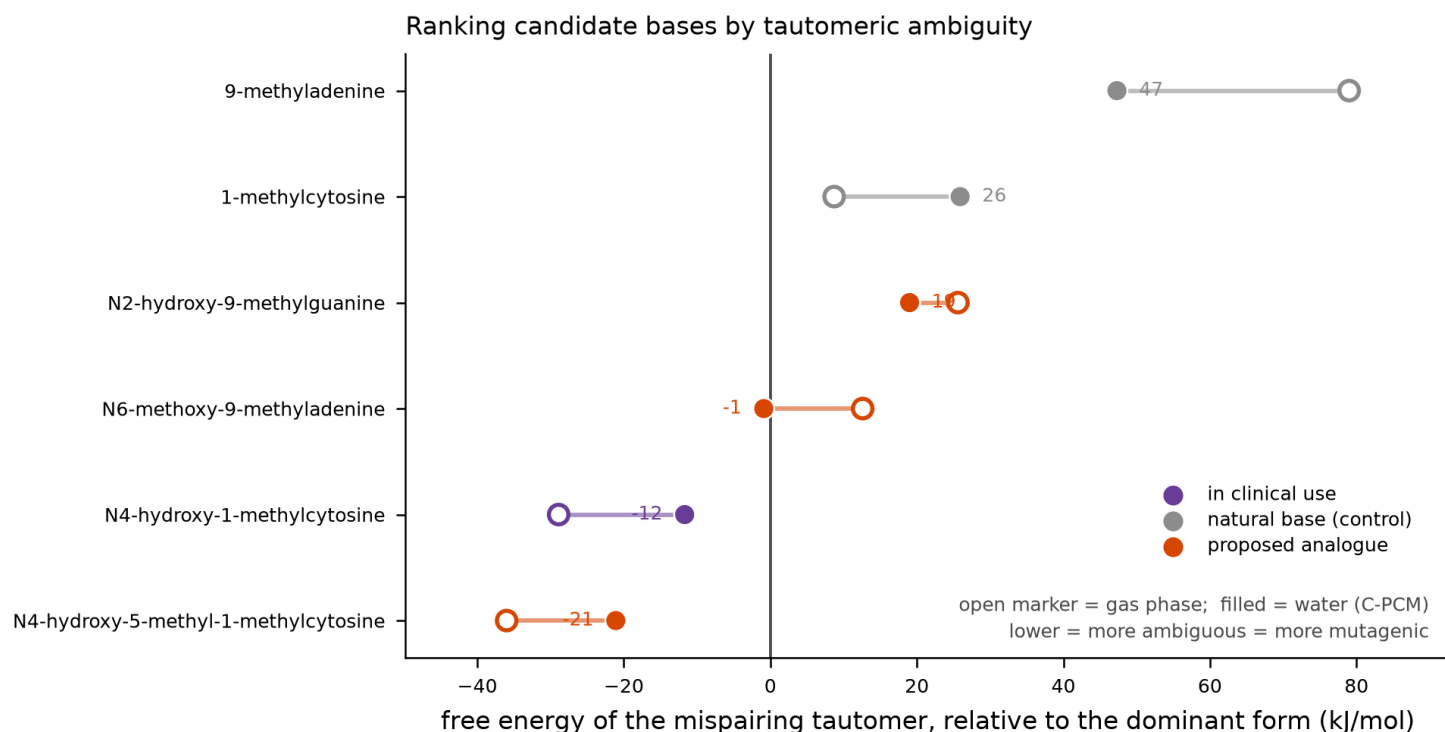


Figure 4

**Candidate bases ranked by tautomeric ambiguity.** Free energy of the mispairing tautomer relative to the dominant form; open markers gas phase, filled markers water. Lower is more ambiguous. The compound in clinical use is shown in purple, natural-base controls in grey, proposed analogues in orange.

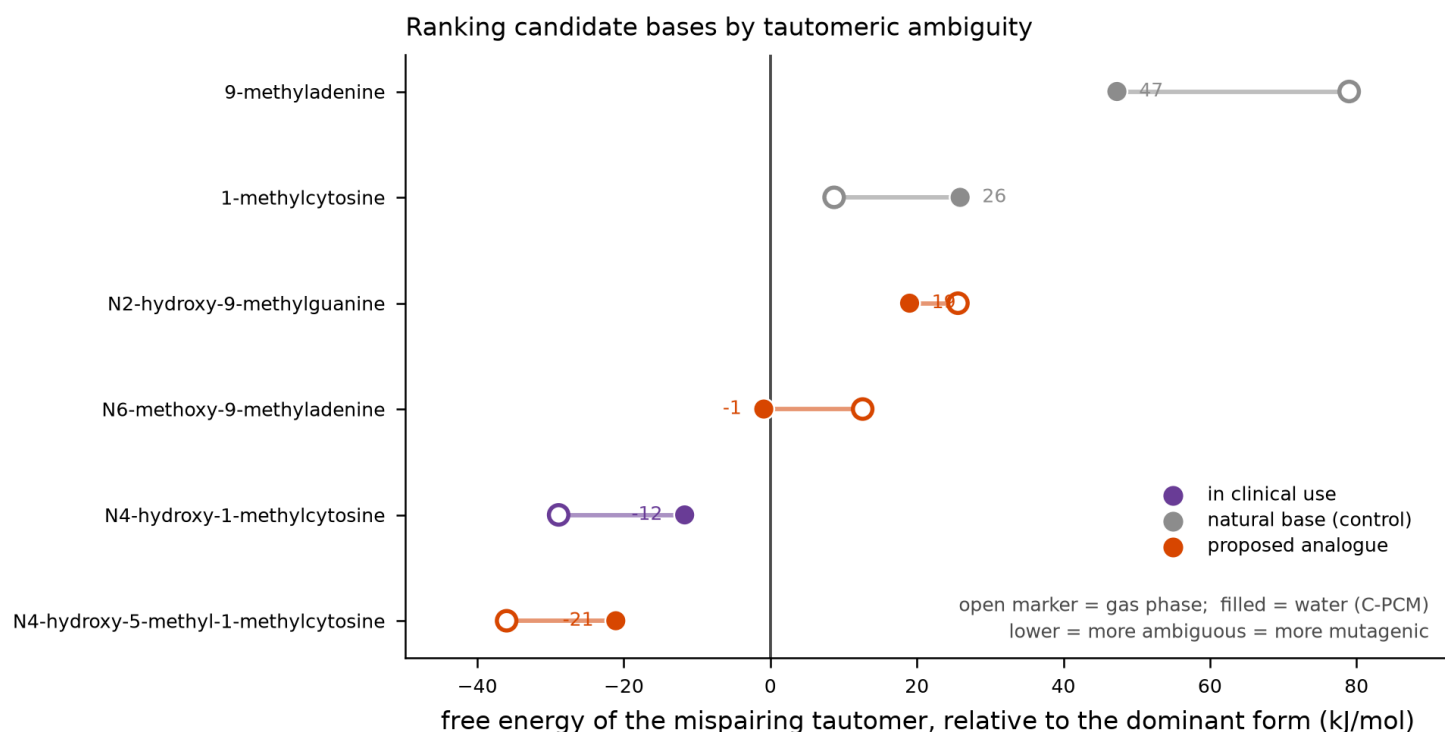


Figure 4

**Candidate bases ranked by tautomeric ambiguity.** Free energy of the mispairing tautomer relative to the dominant form; open markers gas phase, filled markers water. Lower is more ambiguous. The compound in clinical use is shown in purple, natural-base controls in grey, proposed analogues in orange.

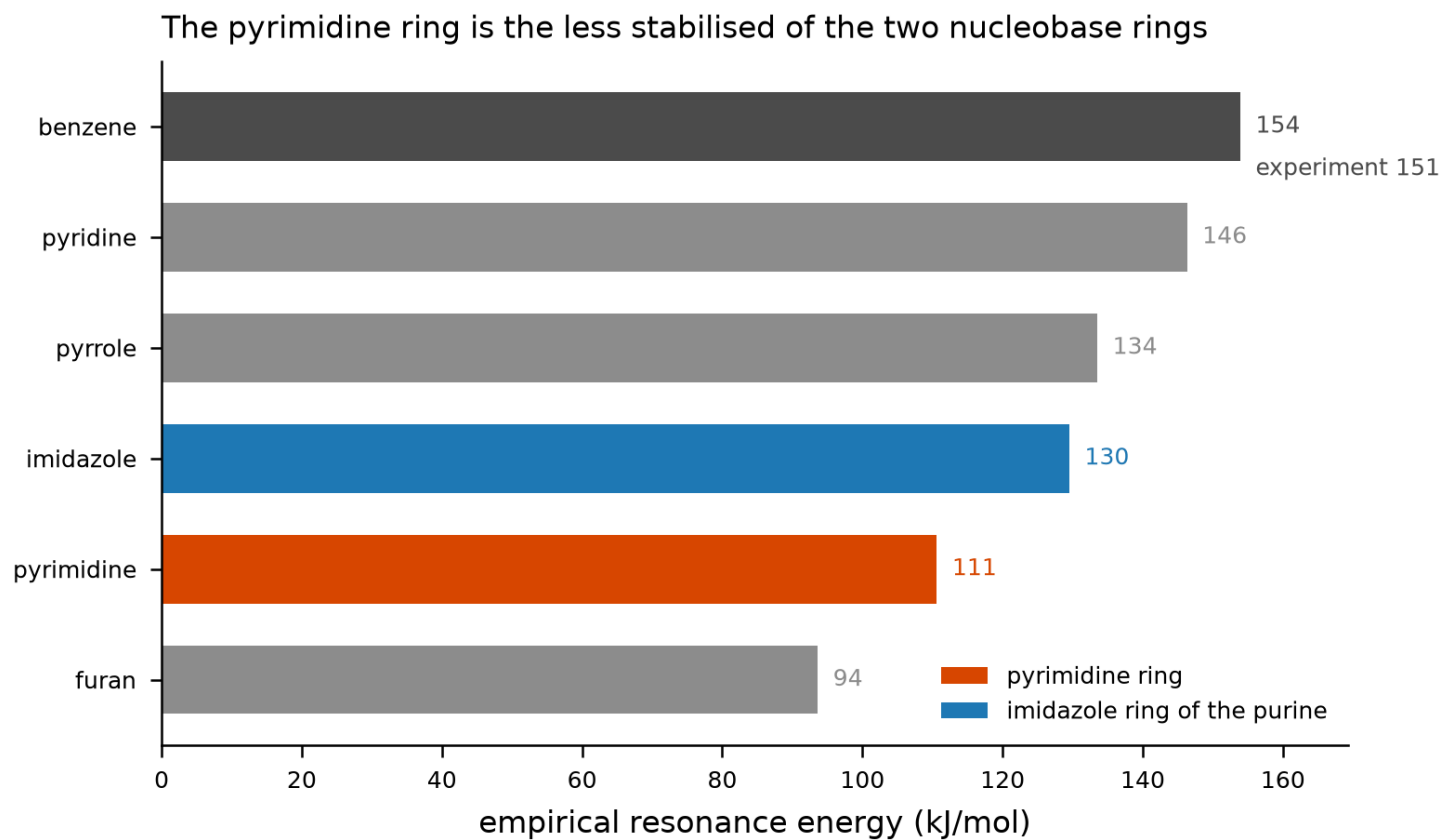


Figure 5

**Empirical resonance energies** from computed hydrogenation enthalpies. Benzene carries the experimental anchor. The pyrimidine ring is the less stabilised of the two nucleobase rings.

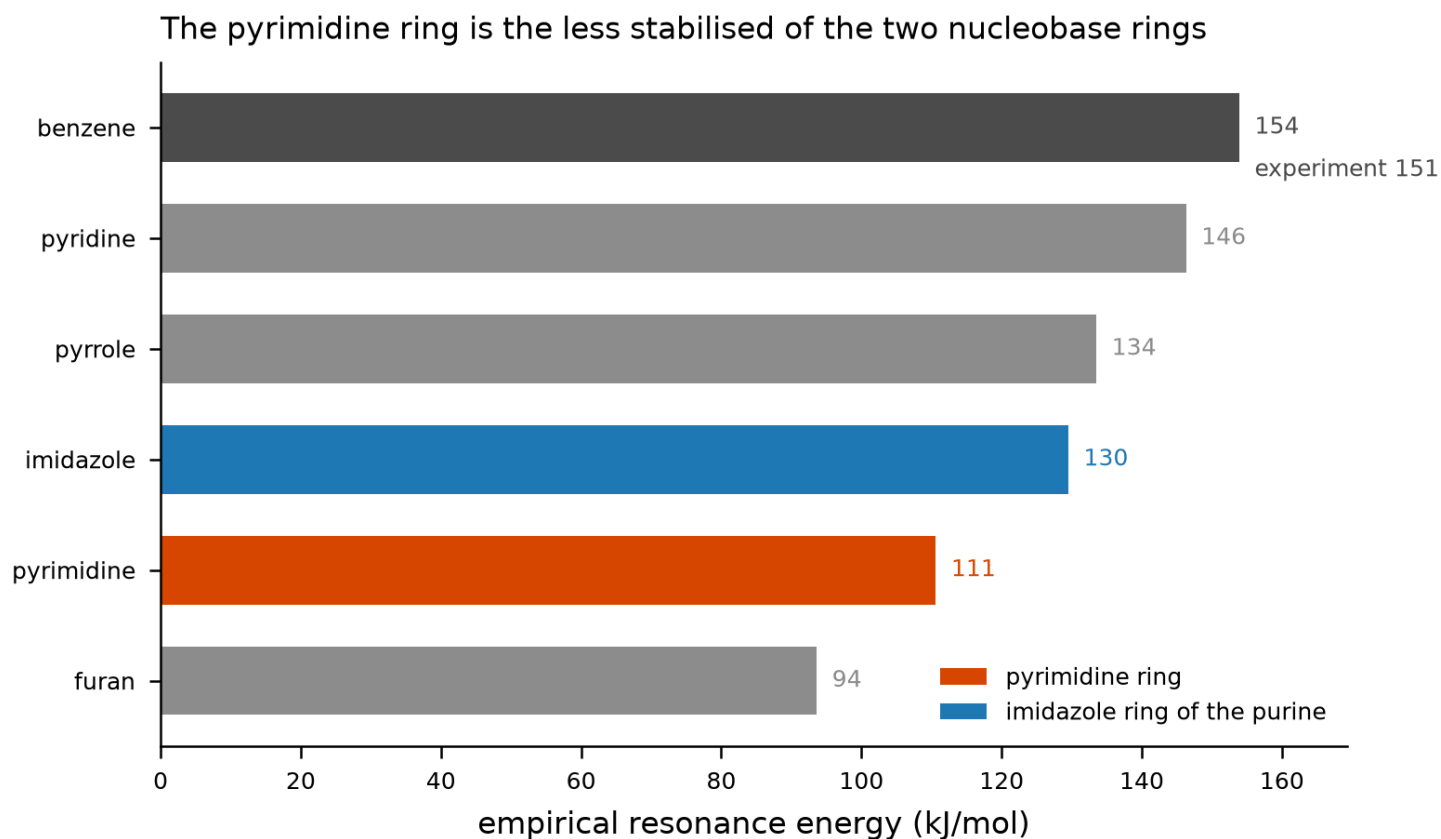


Figure 5

**Empirical resonance energies** from computed hydrogenation enthalpies. Benzene carries the experimental anchor. The pyrimidine ring is the less stabilised of the two nucleobase rings.

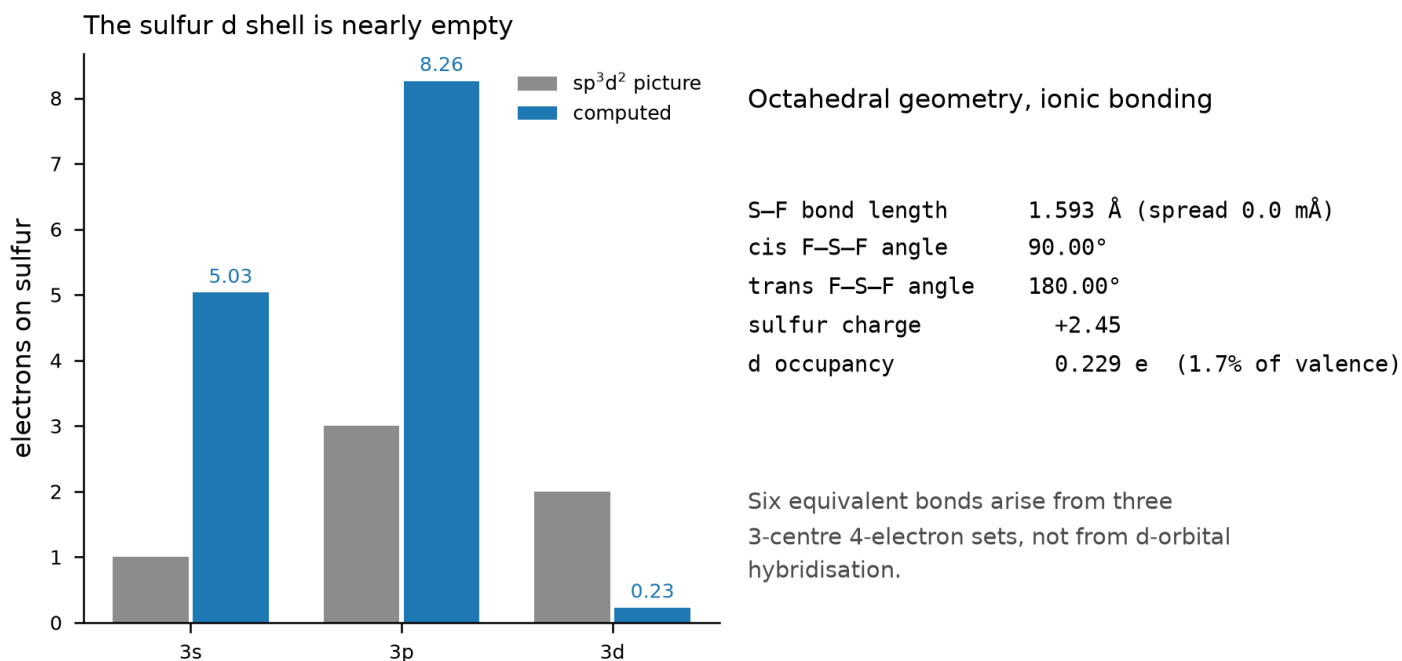


Figure 6

**Sulfur hexafluoride.**Left: computed valence populations on sulfur against those the  $sp^3d^2$  hybridisation picture requires. Right: geometric and charge parameters. The octahedral geometry is reproduced exactly; the d participation the textbook account invokes is absent.

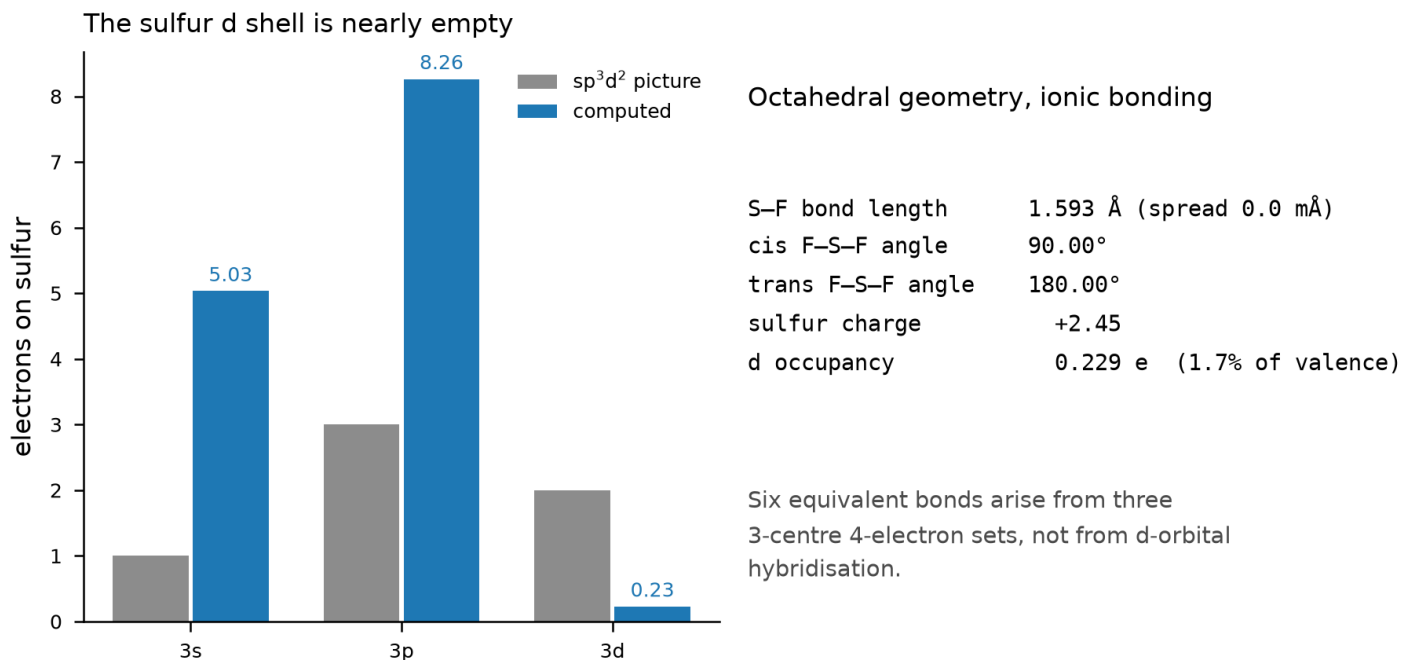


Figure 6

**Sulfur hexafluoride.**Left: computed valence populations on sulfur against those the  $sp^3d^2$  hybridisation picture requires. Right: geometric and charge parameters. The octahedral geometry is reproduced exactly; the d participation the textbook account invokes is absent.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarydata.docx](#)
- [supplementarydata.docx](#)