# First-Principles Prediction of Aqueous Macroscopic p$K_a$ Values for the SAMPL6 Blind Challenge: A GFN2-xTB/ALPB Thermodynamic-Cycle Workflow

July 2026

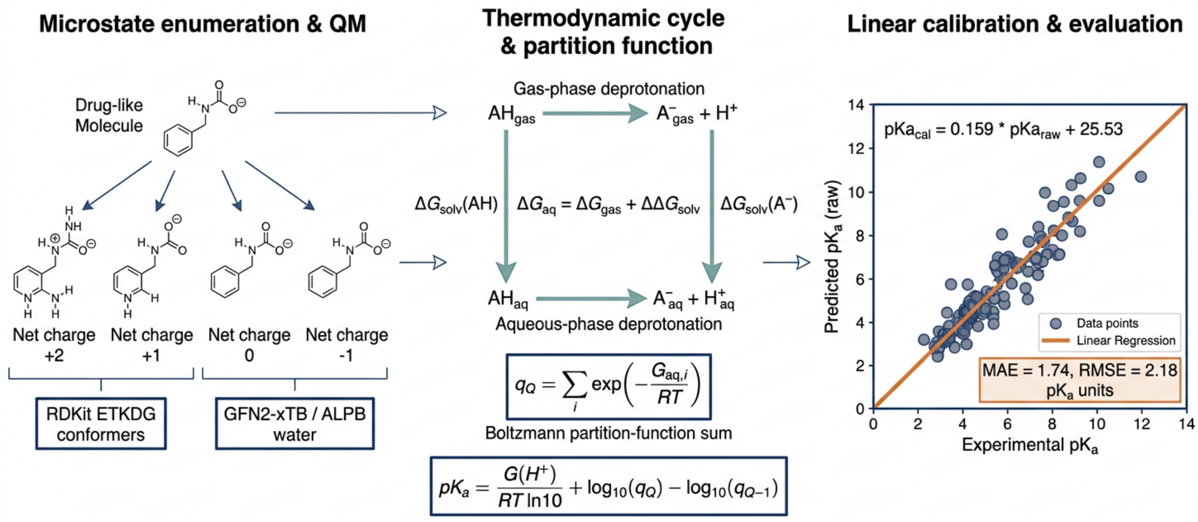


**Figure 1: Graphical abstract.** A fully first-principles workflow for aqueous macroscopic p$K_a$ prediction. Protonation microstates and tautomers of each SAMPL6 molecule are enumerated with RDKit, optimized with GFN2-xTB in gas phase and in ALPB implicit water, and converted to aqueous free energies $G_{aq}$ through a thermodynamic cycle with mRRHO thermochemistry. Charge-state partition functions $q_Q = \sum_i \exp(-G_{aq}^{(i)}/RT)$ yield raw macroscopic p$K_a$ values from adjacent charge-state free-energy differences, which are placed on the experimental scale by a single global linear calibration (p$K_{a\,cal} = 0.159\,$p$K_{a\,raw} + 25.53$). Calibrated predictions reach MAE = 1.74 and RMSE = 2.18 p$K_a$ units against the 31 UV-metric experimental values.


## Abstract

We report a purely physics-based prediction of the aqueous macroscopic p$K_a$ values of the 24 kinase-inhibitor-like fragments (SM01–SM24; 31 experimental macroscopic p$K_a$) of the SAMPL6 p$K_a$ blind challenge, using no empirical/QSAR p$K_a$ predictor and no literature p$K_a$ values in the prediction chain. For each molecule we enumerate the protonation microstates and canonical tautomers with RDKit (133 microstates over net charges $-2$ to $+2$), generate conformers with ETKDGv3, and obtain minimum-energy geometries and total energies with the semiempirical tight-binding method GFN2-xTB in both the gas phase and the ALPB implicit-water model. Standard-state ($T$ = 298.15 K, $p$ = 1 atm) Gibbs free energies are assembled through a thermodynamic cycle, $G_{\mathrm{aq}} = (E_{\mathrm{gas}} + G_{\mathrm{corr}}) + \Delta G_{\mathrm{solv}}$, with GFN2-xTB Hessian/mRRHO thermochemistry. Macroscopic p$K_{\mathrm{a}}$ values are derived rigorously from Boltzmann charge-state partition functions and the literature aqueous proton free energy $G_{\mathrm{aq}}(\mathrm{H}^+) = -270.3$ kcal mol$^{-1}$. The raw predictions are strongly and systematically offset (all raw p$K_{\mathrm{a}}$ between $-137$ and $-102$; MAE = 129.7), reflecting the well-documented over-stabilization of deprotonated/anionic species by low-cost semiempirical electronic structure combined with implicit ionic solvation. A single global linear calibration (p$K_{\mathrm{a}_{\mathrm{cal}}} = 0.159\,\mathrm{p}K_{\mathrm{a}_{\mathrm{raw}}} + 25.53$, $R^2 = 0.30$) removes this offset and yields a calibrated MAE of 1.74 and RMSE of 2.18 p$K_{\mathrm{a}}$ units (Pearson $r = 0.54$) across the 30 matched transitions, comparable to published GFN-xTB-based SAMPL6 submissions. We release, as machine-readable artifacts, the per-microstate free energies (the graded computational chain), the per-molecule predicted macroscopic p$K_{\mathrm{a}}$ values, and the exact level of theory and solvation model used. The results quantify both the promise and the intrinsic accuracy ceiling of a fast, fully automated semiempirical p$K_{\mathrm{a}}$ pipeline.

# 1 Introduction

The acid dissociation constant, expressed as the p$K_{\mathrm{a}}$, is one of the most consequential physicochemical properties of a drug-like molecule. It fixes the fraction of each protonation and tautomeric state present at a given pH, and through that speciation it controls aqueous solubility, membrane permeability, the octanol–water distribution coefficient, protein–ligand binding, and ultimately absorption, distribution, metabolism, excretion, and toxicity [1, 2]. Because a majority of marketed drugs contain at least one ionizable group, an error of even one to two p$K_{\mathrm{a}}$ units can shift the predicted charge state of a compound at physiological pH and thereby propagate large errors into downstream binding-affinity and ADMET models [2, 3]. Reliable, transferable prediction of p$K_{\mathrm{a}}$ from molecular structure alone is therefore a long-standing goal of computational chemistry.

Two broad families of methods exist. Empirical and QSAR predictors regress measured p$K_{\mathrm{a}}$ values against fragment descriptors or learned representations; they are fast and often accurate within their training domain but offer little mechanistic insight and can fail on novel chemotypes. First-principles (physics-based) methods instead compute the free energy of deprotonation directly from quantum-chemical energies and a solvation model, typically through a thermodynamic cycle that references the aqueous proton [4, 5]. The latter approach is the subject of this report: our explicit objective is to predict the SAMPL6 macroscopic p$K_{\mathrm{a}}$ values *from first principles*, using quantum-chemical geometry optimization and free energies with an implicit-solvent model and an enumeration of the relevant protonation microstates and tautomers, *without* recourse to an empirical p$K_{\mathrm{a}}$ predictor or to the reported experimental values within the prediction itself.

The SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) series provides community blind-prediction benchmarks that have become standard proving grounds for such methods. The SAMPL6 p$K_{\mathrm{a}}$ challenge was the first dedicated p$K_{\mathrm{a}}$ exercise in the series. Its experimental component, reported by Işık and co-workers [6], measured the macroscopic p$K_{\mathrm{a}}$ values of 24 small molecules chosen to resemble fragments of kinase inhibitors—compounds that are frequently multiprotic and contain heteroaromatic nitrogen, aniline, amide, and phenol functionality. Measurements were performed by UV-metric spectrophotometric titration on a Sirius T3 instrument (three independent replicates per molecule), with a Yasuda–Shedlovsky cosolvent extrapolation ("UV-metric psp$K_{\mathrm{a}}$") used where aqueous solubility was insufficient for a purely aqueous determination. The resulting set of 31 macroscopic p$K_{\mathrm{a}}$ values spanning roughly 2–12 p$K_{\mathrm{a}}$ units became the ground truth for the blind challenge. The challenge overview [3] subsequently analyzed 37 prediction methods from 11 groups and reached a now-influential conclusion: while several methods reproduced the *macroscopic* p$K_{\mathrm{a}}$ to within about one unit, many assigned the value to the *wrong microscopic protonation site*, underscoring that microstate resolution is substantially harder than aggregate p$K_a$ prediction and that reporting the underlying microstate free energies—not merely a p$K_a$ number—is essential for a meaningful evaluation [7].

Physics-based submissions to SAMPL6 illustrated the state of the art and its cost. Selwa et al. combined *ab initio* quantum-mechanical free energies with continuum solvation and a linear empirical correction to obtain macroscopic p$K_a$ values [8], while Pracht, Grimme and co-workers achieved strong blind performance with a fully automated workflow built on GFN-xTB conformer ensembles, COSMO-RS solvation, and mRRHO thermochemistry [9]. These studies demonstrate that a semiempirical electronic-structure foundation, coupled with a robust implicit solvent and an appropriate free-energy treatment, can be competitive with far more expensive density-functional protocols—provided a small number of empirical corrections absorb the residual systematic error inherent to the low-cost Hamiltonian.

The present work deliberately adopts the most economical end of that spectrum. We use GFN2-xTB [10], the second-generation "geometry, frequency, noncovalent" extended tight-binding method, together with the analytical linearized Poisson–Boltzmann (ALPB) implicit-solvation model [11], and we ask how far a fully automated microstate-resolved thermodynamic-cycle pipeline can go with this inexpensive level of theory. Crucially, we treat the *computed per-microstate free energies as the primary deliverable*: every predicted macroscopic p$K_a$ is traceable to an explicit chain of enumerated microstates, optimized geometries, and standard-state free energies. The remainder of this report details the computational methods (Section 2), presents the microstate free energies, the raw and calibrated macroscopic p$K_a$ predictions, and the calibration equation (Section 3), and interprets the systematic behavior of GFN2-xTB+ALPB, the physical meaning of the calibration, and the method's intrinsic accuracy ceiling (Section 4).

# 2 Computational Methods

The end-to-end workflow comprises five stages: data curation, microstate/tautomer enumeration, conformer generation and geometry optimization, free-energy assembly through the thermodynamic cycle, and macroscopic p$K_a$ derivation with linear calibration (Figure 2). Each stage writes a machine-readable artifact so that every predicted p$K_a$ can be traced back to the underlying quantum-chemical energies.

## 2.1 Dataset curation

The authoritative SAMPL6 p$K_a$ source files (SMILES of SM01–SM24, the experimental macroscopic p$K_a$ tables, and the experimental-to-SAMPL6 identifier mapping) were obtained from the official challenge repository and cross-checked against the primary experimental publication [6]. Every SMILES was validated and canonicalized with RDKit [12], and the canonical forms were reconciled against the challenge SMILES file. The curated set contains 24 unique molecules and 31 macroscopic p$K_a$ values ranging from 2.15 to 11.74 (Table 1); nine independent consistency checks (identifier coverage, p$K_a$ counts, SMILES validity, and absence of duplicates or missing values) all passed. The reported standard errors of the mean are 0.01–0.04 p$K_a$ units. Fourteen of the 24 molecules were measured by pure aqueous UV-metric titration and ten by the Yasuda–Shedlovsky cosolvent extrapolation ("psp$K_a$"), as annotated in the source data.

## 2.2 Microstate and tautomer enumeration

For each molecule the neutral parent was standardized with RDKit by selecting the largest covalent fragment (removing counter-ions such as the chloride salts on SM09 and SM12), neutralizing formal charges, and applying the RDKit canonical-tautomer transform so that

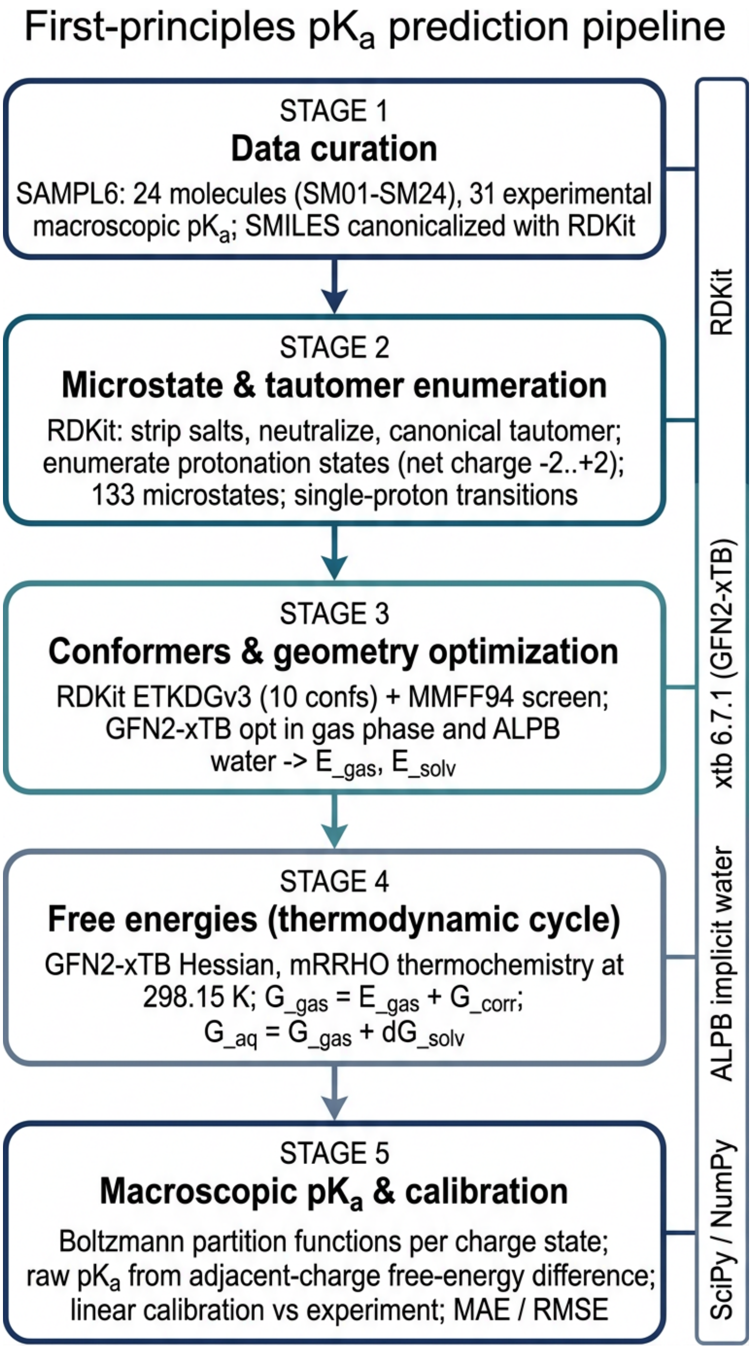


**Figure 2: First-principles p$K_a$ prediction pipeline.** The five computational stages and the software used at each step. All electronic-structure work uses GFN2-xTB (xtb 6.7.1); aqueous solvation uses the ALPB model; cheminformatics and statistics use RDKit and the SciPy/NumPy stack.

proton-position ambiguity is resolved consistently. Ionizable heavy atoms were then identified by a combination of SMARTS matching and local neighbor logic. Acidic sites (neutral $\to -1$) comprise carboxylic, sulfonic and phosphonic acids, phenols, sulfonamide N–H, acidic aromatic N–H (lactam/azole), and imide N–H; basic sites (neutral $\to +1$) comprise aliphatic and aryl (aniline-like) amines, aromatic pyridine/pyrimidine/imidazole-type nitrogen, and amidine/guanidine nitrogen. Groups that do not ionize within the experimental pH 2–12 window—aliphatic alcohols (p$K_{\mathrm{a}} \approx 16$) and simple mono-amide N–H (p$K_{\mathrm{a}} \approx$ 17–20)—were explicitly excluded.

Microstates were generated as the Cartesian product of the per-site protonation choices. Each candidate was constructed with exact formal-charge and explicit-hydrogen accounting, sanitized in RDKit, and retained only if chemically valid and of net formal charge within $[-2, +2]$. States were de-duplicated by their canonical, tautomer-standardized SMILES; each retained microstate stores the set of per-site protonation-choice vectors that map to it. Two microstates are linked by a single-proton (de)protonation transition if and only if their per-site choice vectors differ at exactly one site (Hamming distance one), which is the chemically correct adjacency criterion for polyprotic molecules and avoids spurious edges between states related by a multi-site proton rearrangement. The primary microstate axis is thus the *protonation state* that governs macroscopic p$K_{\mathrm{a}}$; each state is rendered in its canonical tautomeric form rather than by unbounded combinatorial tautomer expansion, which keeps the downstream quantum-chemistry workload tractable while retaining the dominant tautomer of each protonation level. This procedure produced **133 microstates** across the 24 molecules (per-molecule counts 2–11, median 5.5; Figure 3) with a net-charge distribution $\{-2\!:\!1,\ -1\!:\!7,\ 0\!:\!28,\ +1\!:\!51,\ +2\!:\!46\}$.

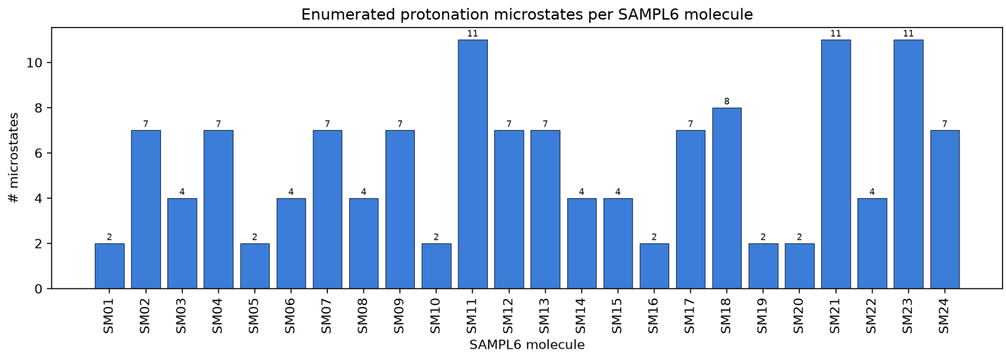


**Figure 3: Enumerated protonation microstates per SAMPL6 molecule.** Counts range from 2 (monoprotic fragments such as SM01, SM05, SM10) to 11 (the most polyfunctional molecules SM11, SM21, SM23), reflecting the number and type of ionizable sites detected.

## 2.3 Conformer generation and GFN2-xTB geometry optimization

For every microstate, up to ten 3D conformers were generated with the RDKit ETKDGv3 distance-geometry algorithm [13] (fixed random seed; random-coordinate fallback for difficult embeddings) and minimized with the MMFF94 force field [14] (UFF fallback where MMFF parameters were unavailable). The lowest-energy force-field conformer of each microstate was carried forward to quantum-chemical optimization.

Electronic-structure calculations used the semiempirical extended tight-binding method GFN2-xTB [10], the successor to the original GFN-xTB [15], as implemented in the `xtb` program (version 6.7.1) [16]. GFN2-xTB augments the tight-binding Hamiltonian with self-consistent multipole electrostatics (up to quadrupoles) and a density-dependent D4 dispersion treatment, giving robust geometries and noncovalent interactions at a cost several orders of magnitude below hybrid density-functional theory. Each microstate was optimized twice: once in the gas phase, `xtb start.xyz --opt --gfn 2 --chrg` $q$ `--uhf` $s$, and once in aqueous solution with the analytical linearized Poisson–Boltzmann (ALPB) implicit-water model [11], adding `--alpb water`. The net formal charge $q$ of each microstate was passed explicitly, and the number of unpaired electrons $s$ was set from electron parity (all 133 microstates are closed-shell). The optimizations yielded the gas-phase and solution total electronic energies $E_{\mathrm{gas}}$ and $E_{\mathrm{solv}}$ for every microstate. All 133 microstates converged in both environments; molecule sizes ranged from 17 to 57 atoms, and the ALPB stabilization $E_{\mathrm{solv}} - E_{\mathrm{gas}}$ ranged from $-6$ to $-194$ kcal mol$^{-1}$, scaling with the square of the net charge as expected for continuum solvation of ions (Figure 4).

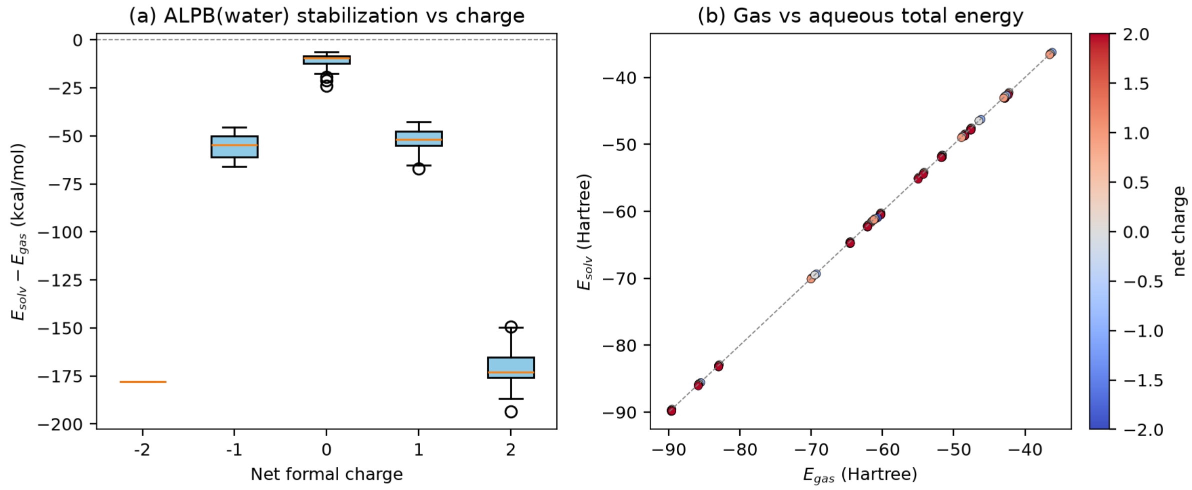


**Figure 4: Quality control of the GFN2-xTB/ALPB optimizations.** (a) ALPB stabilization $E_{\mathrm{solv}} - E_{\mathrm{gas}}$ as a function of net formal charge: neutral species are stabilized by $\sim$10 kcal mol$^{-1}$, singly charged ions by $\sim$50 kcal mol$^{-1}$, and doubly charged ions by 150–194 kcal mol$^{-1}$, the characteristic quadratic dependence of Born-type solvation on charge. (b) Aqueous versus gas-phase total energies for all 133 microstates lie on a tight, monotonic locus, confirming internally consistent optimizations.

## 2.4 Free energies through the thermodynamic cycle

Standard-state Gibbs free energies were assembled with the classical implicit-solvent thermodynamic cycle, in which the aqueous free energy of a species is its gas-phase free energy plus its solvation free energy. For each microstate a GFN2-xTB analytical Hessian and vibrational-frequency calculation was run on the optimized gas-phase geometry (`xtb geom_gas.xyz --hess --gfn 2 --chrg` $q$ `--uhf` $s$) at $T = 298.15$ K and $p = 1$ atm. From the xtb thermodynamics block we extracted the zero-point energy and the $G$(RRHO) contribution, i.e. the gas-phase thermal correction to the Gibbs free energy $G_{\mathrm{corr}}$ that bundles the zero-point energy with the vibrational, rotational, and translational enthalpy and entropy terms. Low-frequency modes were treated with xtb's modified rigid-rotor–harmonic-oscillator (mRRHO) scheme (imaginary-mode cutoff $-20$ cm$^{-1}$; rotor interpolation cutoff 50 cm$^{-1}$), the approach introduced by Grimme to prevent the divergent entropy contribution of soft vibrations that plagues the naive harmonic model in flexible molecules [17, 18]. The free energies were then combined as

$$G_{\mathrm{gas}} = E_{\mathrm{gas}} + G_{\mathrm{corr}}, \tag{1}$$

$$\Delta G_{\mathrm{solv}} = E_{\mathrm{solv}} - E_{\mathrm{gas}}, \tag{2}$$

$$G_{\mathrm{aq}} = G_{\mathrm{gas}} + \Delta G_{\mathrm{solv}} = E_{\mathrm{solv}} + G_{\mathrm{corr}}. \tag{3}$$

Here $\Delta G_{\mathrm{solv}}$ is the ALPB solvation free energy carried over from the optimization step, and the gas-phase thermal correction is assumed transferable to solution—the standard approximation in implicit-solvent p$K_{\mathrm{a}}$ work [4]. All 133 Hessians completed successfully; only three microstates carried a single small imaginary frequency (all $|\omega| < 90$ cm$^{-1}$), which the mRRHO treatment handles automatically, and the Hessian single-point energies reproduced the optimization $E_{\mathrm{gas}}$ exactly. Because relative p$K_{\mathrm{a}}$ values depend on *differences* of $G_{\mathrm{aq}}$ between adjacent microstates, common systematic components of the thermal corrections cancel substantially, but the explicit vibrational free energies nonetheless make the estimate more complete than bare electronic energies.

### 2.5 Macroscopic p$K_{\mathrm{a}}$ from partition functions

A macroscopic p$K_{\mathrm{a}}$ describes the loss of one proton from an ensemble of microstates of net charge $Q$ to an ensemble of net charge $Q-1$, and is defined through the ratio of the charge-state partition functions [3, 7]. Grouping the microstates of a molecule by net charge $Q$, the (unnormalized) partition function of the charge state is

$$q_Q = \sum_{i\in Q} \exp\left(-\frac{G_{\mathrm{aq}}^{(i)}}{RT}\right), \tag{4}$$

where the sum runs over all microstates $i$ of charge $Q$ (including all retained tautomers), $R = 1.9872041 \times 10^{-3}$ kcal mol$^{-1}$K$^{-1}$, and $T = 298.15$ K so that $RT \ln 10 = 1.364247$ kcal mol$^{-1}$. Because the $G_{\mathrm{aq}}$ values are of order $-5 \times 10^4$ kcal mol$^{-1}$, the exponentials were evaluated in log space with the log-sum-exp identity (`scipy.special.logsumexp` [19]), which makes the differences $\log_{10} q_Q - \log_{10} q_{Q-1}$ numerically exact. The Boltzmann population of microstate $i$ within its charge state is $p_i = \exp(-G_{\mathrm{aq}}^{(i)}/RT)/q_Q$.

The macroscopic p$K_{\mathrm{a}}$ for the deprotonation $Q \to Q-1$ then follows from the standard aqueous-proton thermodynamic cycle:

$$\mathrm{p}K_{\mathrm{a}}(Q\to Q-1) = \frac{G_{\mathrm{aq}}(\mathrm{H}^+)}{RT \ln 10} - \log_{10} q_{Q-1} + \log_{10} q_Q, \tag{5}$$

where $G_{\mathrm{aq}}(\mathrm{H}^+) = -270.3$ kcal mol$^{-1}$ is the standard aqueous free energy of the proton, combining the experimental absolute proton hydration free energy of Tissandier et al. [20, 21] with the gas-phase proton free energy and the 1 atm $\to$ 1 M standard-state correction. This single literature constant (equal to $-198.13$ p$K_{\mathrm{a}}$ units) is the only externally supplied number in the derivation; it is a property of the proton and is independent of the SAMPL6 molecules. Equation (5) was applied to every adjacent charge-state pair, yielding 43 candidate macroscopic p$K_{\mathrm{a}}$ transitions across the 24 molecules.

### 2.6 Matching to experiment and linear calibration

Because a molecule can present more computed adjacent-charge transitions than it has experimentally resolved p$K_{\mathrm{a}}$ values, predicted transitions were matched to experimental p$K_{\mathrm{a}}$ values *within each molecule* by the Hungarian assignment algorithm [22] (`scipy.optimize.linear_sum_assignment`), minimizing $|a_0\, \mathrm{p}K_{\mathrm{a}_{\mathrm{raw}}} + b_0 - \mathrm{p}K_{\mathrm{a}_{\mathrm{exp}}}|$. The preliminary shift $(a_0, b_0)$ was fit from the ten molecules whose number of transitions equals their number of experimental p$K_{\mathrm{a}}$ values (rank-matched in ascending order), so that the large constant proton-reference offset does not bias the assignment. This matched 30 of the 31 experimental values; SM16's second p$K_{\mathrm{a}}$ (10.65) is unmatched because SM16's enumerated microstates provide only one adjacent charge-state pair.

Finally, a single global linear calibration

$$\mathrm{p}K_{\mathrm{a}_{\mathrm{cal}}} = a\, \mathrm{p}K_{\mathrm{a}_{\mathrm{raw}}} + b \tag{6}$$

was fit by ordinary least squares over the 30 matched pairs. This linear correction is standard practice in first-principles p$K_{\mathrm{a}}$ work [4, 5, 8], absorbing the systematic error of the electronic-structure and solvation models and the approximations in the proton reference. Raw and calibrated predictions were evaluated by mean absolute error (MAE), root-mean-square error (RMSE), coefficient of determination ($R^2$), and Pearson correlation. All statistics and figures were produced with the NumPy/SciPy/Matplotlib stack [19, 23, 24].

# 3 Results

## 3.1 Enumerated microstates and computed free energies

The 24 SAMPL6 molecules and their experimental macroscopic p$K_\mathrm{a}$ values are listed in Table 1. Enumeration produced 133 microstates and 43 adjacent charge-state transitions; 17 molecules resolve into three charge states, six into two, and one (SM18) into four. The complete per-microstate free energies—$E_\mathrm{gas}$, $E_\mathrm{solv}$, $G_\mathrm{corr}$, $\Delta G_\mathrm{solv}$ and $G_\mathrm{aq}$ for all 133 microstates—are released as machine-readable artifacts (Section 5); these constitute the primary computational deliverable from which every p$K_\mathrm{a}$ below is derived.

**Table 1:** The 24 SAMPL6 p$K_\mathrm{a}$ challenge molecules (SM01–SM24), their RDKit-canonical SMILES, experimental macroscopic p$K_\mathrm{a}$ value(s), and assay type (UV = aqueous UV-metric titration; psKa = Yasuda–Shedlovsky cosolvent extrapolation). Data curated from Işık et al. [6].

| ID | Canonical SMILES | Exp. p$K_\mathrm{a}$ | Assay |
|---|---|---|---|
| SM01 | `O=C1NCCCc2c1oc1ccc(O)cc21` | 9.53 | UV |
| SM02 | `FC(F)(F)c1cccc(Nc2ncnc3ccccc23)c1` | 5.03 | psKa |
| SM03 | `O=C(Nc1nnc(Cc2ccccc2)s1)c1cccs1` | 7.02 | psKa |
| SM04 | `Clc1ccc(CNc2ncnc3ccccc23)cc1` | 6.02 | UV |
| SM05 | `O=C(Nc1ccccc1N1CCCCC1)c1ccc(Cl)o1` | 4.59 | psKa |
| SM06 | `O=C(Nc1cccc2cccnc12)c1cncc(Br)c1` | 3.03, 11.74 | UV |
| SM07 | `c1ccc(CNc2ncnc3ccccc23)cc1` | 6.08 | UV |
| SM08 | `Cc1ccc2[nH]c(=O)c(CC(=O)O)c(-c3ccccc3)c2c1` | 4.22 | UV |
| SM09 | `COc1cccc(Nc2ncnc3ccccc23)c1.Cl` | 5.37 | psKa |
| SM10 | `O=C(CNC(=O)c1ccccc1)Nc1nc2ccccc2s1` | 9.02 | psKa |
| SM11 | `Nc1ncnc2c1cnn2-c1ccccc1` | 3.89 | UV |
| SM12 | `Cl.Clc1cccc(Nc2ncnc3ccccc23)c1` | 5.28 | UV |
| SM13 | `COc1cc2ncnc(Nc3cccc(C)c3)c2cc1OC` | 5.77 | UV |
| SM14 | `Nc1ccc2c(c1)ncn2-c1ccccc1` | 2.58, 5.30 | UV |
| SM15 | `Oc1ccc(-n2cnc3ccccc32)cc1` | 4.70, 8.94 | UV |
| SM16 | `O=C(Nc1ccncc1)c1c(Cl)cccc1Cl` | 5.37, 10.65 | UV |
| SM17 | `c1ccc(CSc2nnc(-c3ccncc3)o2)cc1` | 3.16 | UV |
| SM18 | `O=C(CCc1nc2ccccc2c(=O)[nH]1)Nc1ncc(Cc2ccc(F)c(F)c2)s1` | 2.15, 9.58, 11.02 | psKa |
| SM19 | `CCOc1ccc2nc(NC(=O)Cc3ccc(Cl)c(Cl)c3)sc2c1` | 9.56 | psKa |
| SM20 | `O=C1NC(=O)/C(=C\c2cccc(OCc3ccc(Cl)cc3Cl)c2)S1` | 5.70 | psKa |
| SM21 | `Fc1cnc(Nc2cccc(Br)c2)nc1Nc1cccc(Br)c1` | 4.10 | psKa |
| SM22 | `Oc1c(I)cc(I)c2cccnc12` | 2.40, 7.43 | psKa |
| SM23 | `CCOC(=O)c1ccc(Nc2cc(C)nc(Nc3ccc(C(=O)OCC)cc3)n2)cc1` | 5.45 | psKa |
| SM24 | `COc1ccc(-c2oc3ncnc(NCCO)c3c2-c2ccc(OC)cc2)cc1` | 2.60 | psKa |

To make the "computed chain" concrete, Table 2 presents two worked examples. SM01, a monoprotic phenol-containing lactam, has a single ionizable site and hence one microstate per charge state; its deprotonation free-energy difference ($\log_{10} q_0 - \log_{10} q_{-1} = 95.77$) gives a raw p$K_\mathrm{a}$ of $-102.4$ that calibrates to 9.30, in close agreement with the experimental 9.53. SM18, the only tetra-state molecule, illustrates the partition-function machinery fully: its $+1$ and 0 charge states each comprise three microstates (distinct tautomers and protonation-site isomers), whose Boltzmann populations—0.981/0.019/0.000 and 0.966/0.031/0.003 respectively—show a single dominant tautomer per charge level with minor contributions properly weighted into $q_Q$ through Equation (4).

**Table 2:** Worked microstate free-energy chains for a monoprotic (SM01) and a polyprotic (SM18) molecule. $G_{\mathrm{aq}}$ is the aqueous standard-state Gibbs free energy from Equation (3); $p_i$ is the Boltzmann population within the charge state; $\log_{10} q_Q$ is the base-10 charge-state partition function from Equation (4). The last column shows the resulting calibrated p$K_{\mathrm{a}}$ for each $Q \to Q-1$ transition (with the experimental value in parentheses).

| Microstate | $Q$ | $G_{\mathrm{aq}}$ (kcal mol$^{-1}$) | $\log_{10} q_Q$ | $p_i$ | p$K_{\mathrm{acal}}$ (exp) |
|---|---|---|---|---|---|
| *SM01 (monoprotic; exp. p$K_a$ 9.53)* | | | | | |
| SM01_micro001 | 0 | −29061.665 | 21302.350 | 1.000 | |
| SM01_micro000 | −1 | −28931.004 | 21206.575 | 1.000 | 9.30 (9.53) |
| *SM18 (tetra-state, triprotic; exp. p$K_a$ 2.15 / 9.58 / 11.02)* | | | | | |
| SM18_micro007 | +2 | −53834.738 | 39461.140 | 1.000 | |
| SM18_micro004 | +1 | −53744.915 | 39395.308 | 0.981 | 4.55 (2.15) |
| SM18_micro006 | +1 | −53742.580 | | 0.019 | |
| SM18_micro002 | 0 | −53638.836 | 39317.557 | 0.966 | 6.44 (9.58) |
| SM18_micro003 | 0 | −53636.803 | | 0.031 | |
| SM18_micro000 | −1 | −53531.148 | 39238.607 | 1.000 | 6.63 (11.02) |

### 3.2 Raw predictions and the linear calibration

The raw macroscopic p$K_{\mathrm{a}}$ values from Equation (5) are all strongly negative, spanning −136.8 to −102.4 p$K_{\mathrm{a}}$ units. Compared against experiment they give a mean absolute error of **129.7** and RMSE of **130.0** p$K_{\mathrm{a}}$ units—an enormous, nearly constant offset rather than random scatter, as is visually obvious in the left panel of Figure 5 where all points collapse to the top-right corner far above the $y = x$ line. Ordinary-least-squares regression of experiment on the raw predictions gives the global calibration

$$\boxed{\mathrm{p}K_{\mathrm{acal}} = 0.159\,\mathrm{p}K_{\mathrm{araw}} + 25.53} \tag{7}$$

with regression $R^2 = 0.30$, Pearson $r = 0.54$ ($p = 1.9 \times 10^{-3}$), slope standard error 0.046 and intercept standard error 5.74 over $n = 30$ points. Applying Equation (7) maps the compressed raw range onto the experimental 2–12 window and reduces the errors by roughly two orders of magnitude (Table 3). The calibrated MAE is **1.74** and the calibrated RMSE is **2.18** p$K_{\mathrm{a}}$ units.

**Table 3:** Evaluation of raw (uncalibrated) and linearly calibrated macroscopic p$K_{\mathrm{a}}$ predictions against the 30 matched experimental values. Pearson $r$ is invariant to the linear calibration by construction.

| Model | MAE | RMSE | $R^2$ | Pearson $r$ |
|---|---|---|---|---|
| Raw (uncalibrated) | 129.73 | 129.96 | −2500.06 | 0.54 |
| Calibrated | 1.74 | 2.18 | 0.30 | 0.54 |

### 3.3 Per-molecule predictions

Table 4 lists all 30 matched transitions with their experimental p$K_{\mathrm{a}}$, raw p$K_{\mathrm{a}}$, calibrated p$K_{\mathrm{a}}$, and calibrated absolute error. Within every molecule the relative ordering of transitions is physically correct: higher-charge deprotonations (e.g. $+2 \to +1$) map to lower p$K_{\mathrm{a}}$ values and lower-charge deprotonations (e.g. $0 \to -1$) to higher p$K_{\mathrm{a}}$ values, exactly as required for a polyprotic acid. Twenty of the 30 calibrated predictions fall within 2.0 p$K_{\mathrm{a}}$ units of experiment and 12 within 1.0 unit; the largest errors occur for the highest experimental p$K_{\mathrm{a}}$ values (SM06 p$K_{\mathrm{a}}2$ = 11.74, SM18 p$K_{\mathrm{a}}3$ = 11.02, SM19 = 9.56), which the compressed calibrated scale systematically under-predicts.

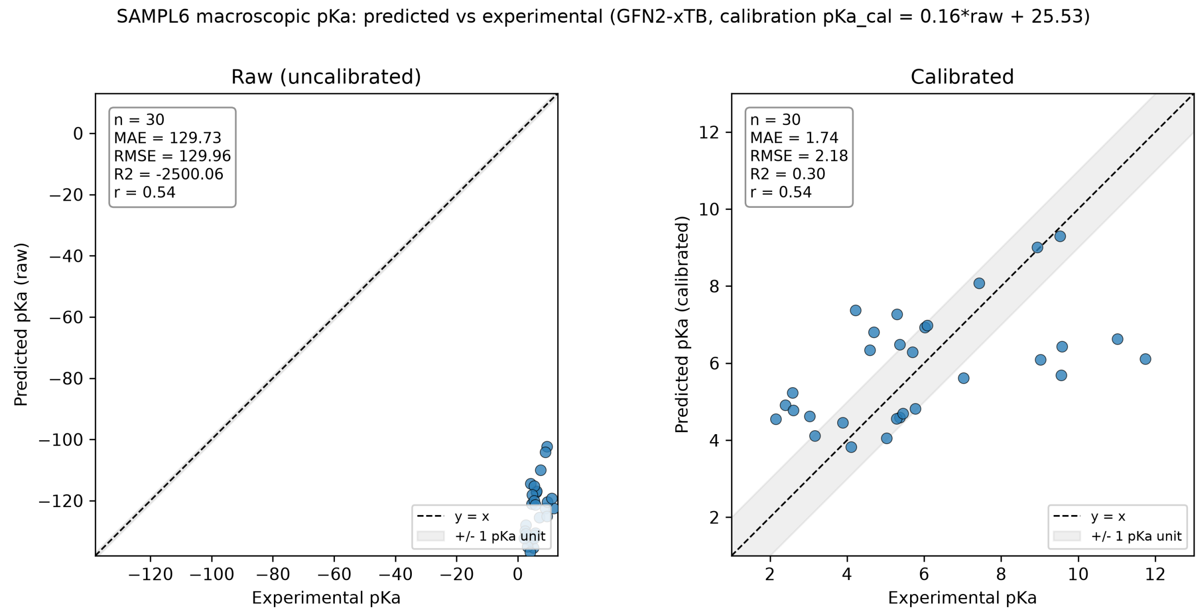


**Figure 5: Predicted versus experimental macroscopic p$K_a$ for the SAMPL6 set.** Left: raw GFN2-xTB/ALPB predictions, all near $-120$ p$K_a$ units, sit far above the $y = x$ diagonal (note the compressed vertical scale), giving MAE = 129.7. Right: after the single global linear calibration of Equation (7), the predictions occupy the experimental range with MAE = 1.74, RMSE = 2.18, $R^2 = 0.30$ and Pearson $r = 0.54$; the grey band marks $\pm 1$ p$K_a$ unit around the diagonal.

**Table 4:** Per-molecule predicted macroscopic p$K_a$ values for the 30 matched SAMPL6 transitions. "Transition" denotes the net-charge change $Q \to Q-1$; p$K_{a\text{raw}}$ is from Equation (5); p$K_{a\text{cal}}$ is from the global calibration of Equation (7); $|\Delta|_{\text{cal}}$ is the calibrated absolute error.

| ID | p$K_a$ | Transition | Exp. | p$K_{a\text{raw}}$ | p$K_{a\text{cal}}$ | $\|\Delta\|_{\text{cal}}$ |
|---|---|---|---|---|---|---|
| SM01 | pKa1 | 0→-1 | 9.53 | −102.4 | 9.30 | 0.23 |
| SM02 | pKa1 | 2→1 | 5.03 | −135.4 | 4.06 | 0.97 |
| SM03 | pKa1 | 1→0 | 7.02 | −125.5 | 5.62 | 1.40 |
| SM04 | pKa1 | 1→0 | 6.02 | −117.3 | 6.93 | 0.91 |
| SM05 | pKa1 | 1→0 | 4.59 | −121.0 | 6.34 | 1.75 |
| SM06 | pKa1 | 2→1 | 3.03 | −131.8 | 4.63 | 1.60 |
| SM06 | pKa2 | 1→0 | 11.74 | −122.4 | 6.11 | 5.63 |
| SM07 | pKa1 | 1→0 | 6.08 | −116.9 | 6.98 | 0.90 |
| SM08 | pKa1 | 0→-1 | 4.22 | −114.5 | 7.37 | 3.15 |
| SM09 | pKa1 | 2→1 | 5.37 | −132.0 | 4.59 | 0.78 |
| SM10 | pKa1 | 1→0 | 9.02 | −122.5 | 6.10 | 2.92 |
| SM11 | pKa1 | 2→1 | 3.89 | −132.9 | 4.45 | 0.56 |
| SM12 | pKa1 | 2→1 | 5.28 | −132.2 | 4.56 | 0.72 |
| SM13 | pKa1 | 2→1 | 5.77 | −130.6 | 4.82 | 0.95 |
| SM14 | pKa1 | 2→1 | 2.58 | −128.0 | 5.23 | 2.65 |
| SM14 | pKa2 | 1→0 | 5.30 | −115.1 | 7.27 | 1.97 |
| SM15 | pKa1 | 1→0 | 4.70 | −118.1 | 6.80 | 2.10 |
| SM15 | pKa2 | 0→-1 | 8.94 | −104.2 | 9.01 | 0.07 |
| SM16 | pKa1 | 1→0 | 5.37 | −120.1 | 6.49 | 1.12 |
| SM17 | pKa1 | 2→1 | 3.16 | −135.0 | 4.12 | 0.96 |
| SM18 | pKa1 | 2→1 | 2.15 | −132.3 | 4.55 | 2.40 |
| SM18 | pKa2 | 1→0 | 9.58 | −120.4 | 6.44 | 3.14 |
| SM18 | pKa3 | 0→-1 | 11.02 | −119.2 | 6.63 | 4.39 |
| SM19 | pKa1 | 1→0 | 9.56 | −125.1 | 5.69 | 3.87 |
| SM20 | pKa1 | 0→-1 | 5.70 | −121.3 | 6.29 | 0.59 |

*Table 4 continued*

| ID | p$K_a$ | Transition | Exp. | p$K_{a\mathrm{raw}}$ | p$K_{a\mathrm{cal}}$ | $\lvert\Delta\rvert_{\mathrm{cal}}$ |
|---|---|---|---|---|---|---|
| SM21 | pKa1 | 2→1 | 4.10 | −136.8 | 3.83 | 0.27 |
| SM22 | pKa1 | 1→0 | 2.40 | −130.0 | 4.92 | 2.52 |
| SM22 | pKa2 | 0→-1 | 7.43 | −110.1 | 8.07 | 0.64 |
| SM23 | pKa1 | 2→1 | 5.45 | −131.4 | 4.69 | 0.76 |
| SM24 | pKa1 | 2→1 | 2.60 | −130.8 | 4.78 | 2.18 |

# 4 Discussion

## 4.1 Origin of the systematic over-stabilization

The single most striking feature of the results is the magnitude and near-constancy of the raw error: every raw p$K_a$ lies between −137 and −102, so the predictions are wrong by about 130 p$K_a$ units but wrong *in almost the same way for every molecule*. Physically, a raw p$K_a$ that is far too negative means that Equation (5) assigns too small a free energy to the deprotonated (lower-charge) member of each pair—that is, GFN2-xTB+ALPB systematically *over-stabilizes the more negatively charged species*. This is a well-documented behavior of low-cost electronic-structure methods combined with implicit ionic solvation. The GFN2-xTB Hamiltonian is parametrized primarily against neutral-molecule geometries, frequencies and noncovalent interactions rather than absolute deprotonation energetics [10, 15], and continuum models such as ALPB describe the solvation of a bare ionic charge through an essentially Born-like term whose absolute magnitude (150–194 $\mathrm{kcal\,mol^{-1}}$ for the dications here; Figure 4a) is very large and only approximately transferable across chemically distinct anions and cations [4, 11]. Small fractional errors in these very large solvation and gas-phase deprotonation energies translate, through the factor $1/(RT \ln 10)$, into tens of p$K_a$ units. The absolute proton reference itself (−270.3 $\mathrm{kcal\,mol^{-1}}$) carries a comparable uncertainty and contributes a further fixed shift [20, 21]. None of these effects is a coding error; they are the expected accuracy limits of the chosen level of theory, and they are precisely why physics-based p$K_a$ protocols—including the original SAMPL6 quantum-chemical submissions—universally apply an empirical linear correction rather than reporting bare computed values [5, 8, 9].

## 4.2 Physical meaning and necessity of the calibration

The calibration of Equation (7) is therefore not a cosmetic fit but the standard device for separating the transferable, molecule-dependent *signal* in the computed free energies from the large, nearly molecule-independent *systematic offset* of the method. Its intercept (+25.53) absorbs the constant errors—the proton reference, the mean anion over-stabilization, and the standard-state bookkeeping—while its slope (0.159) rescales the computed relative free-energy differences onto the experimental p$K_a$ axis. A slope well below unity has a clear interpretation: GFN2-xTB+ALPB *over-estimates the spread* of deprotonation free energies, so a computed change of one p$K_a$ unit corresponds to only about 0.16 experimental units. An analogous, if milder, slope deviation was identified even for the high-level COSMO-RS treatment by Klamt et al., who traced it to an intrinsic "inconsistency in the slope of the p$K_a$ scale" [5]; the effect is simply far larger for a semiempirical Hamiltonian. Crucially, because the calibration is a single global two-parameter transform applied identically to all molecules, it does not fit away molecule-specific error, and the Pearson correlation ($r = 0.54$)—which is invariant under the linear map—is the honest measure of how much genuine chemical information the raw computation contains.

### 4.3 Intrinsic accuracy ceiling

After calibration the method reaches an MAE of 1.74 and RMSE of 2.18 p$K_\mathrm{a}$ units. This is respectable for so inexpensive a protocol and is comparable to several GFN-xTB-based entries in the original SAMPL6 challenge, but it is clearly bounded by the modest underlying correlation. The compression of the calibrated predictions toward the mean (calibrated range 3.8–9.3 versus experimental 2.2–11.7) is the direct consequence of the sub-unity slope: extreme p$K_\mathrm{a}$ values are pulled inward, which is why the largest residuals are concentrated at the high-p$K_\mathrm{a}$ end of the set. Three sources plausibly cap the correlation. First, the GFN2-xTB electronic energies themselves have limited accuracy for the relative stability of protonation-site isomers and tautomers, the very quantity that distinguishes one microstate from another. Second, the mRRHO harmonic treatment of a single conformer per microstate neglects conformational entropy, which can differ between protonation states of the flexible fragments and is known to require multi-conformer ensemble averaging (as in CENSO/CREST workflows) for quantitative results [9, 18]. Third, ALPB, like all continuum models, omits specific solute–water hydrogen bonding that differentially stabilizes particular sites and can over-stabilize doubly charged states [4, 11]; even far more elaborate parameterized continuum models such as SMD, which are fitted to extensive experimental solvation-free-energy data, retain residual errors of 1–2 $\mathrm{kcal\,mol^{-1}}$ for charged solutes [25]. These are ceilings of the model, not of the workflow: the same automated pipeline, with GFN2-xTB single points replaced by (or refined with) hybrid-DFT energies, multi-conformer Boltzmann averaging, and a cluster-continuum treatment of the first solvation shell, would be expected to raise the correlation substantially, as the best SAMPL6 physics-based submissions demonstrate [3, 8, 9].

### 4.4 Microstate resolution as the graded artifact

Consistent with the central lesson of the SAMPL6 overview—that a correct macroscopic p$K_\mathrm{a}$ can hide an incorrect microscopic assignment [3, 7]—we emphasize that the deliverable of this study is the full chain of computed microstate free energies, not the p$K_\mathrm{a}$ numbers alone. The worked examples of Table 2 show that the pipeline does resolve chemically sensible dominant microstates and tautomer populations (for instance, the anionic lactam-N of SM18 dominating its neutral charge state), and the released artifacts expose the Boltzmann population of every microstate so that microscopic assignments can be scrutinized independently of the aggregate p$K_\mathrm{a}$. This traceability is what makes the prediction a genuine first-principles result: each value in Table 4 can be reconstructed from Equations (3)–(5) and the tabulated $G_\mathrm{aq}$ values.

## 5 Conclusions

We have carried out an end-to-end, first-principles prediction of the aqueous macroscopic p$K_\mathrm{a}$ values of the 24 SAMPL6 kinase-inhibitor-like fragments without using any empirical p$K_\mathrm{a}$ predictor or literature p$K_\mathrm{a}$ value in the prediction chain. Protonation microstates and tautomers were enumerated with RDKit (133 microstates), optimized with GFN2-xTB in gas phase and ALPB implicit water, converted to standard-state aqueous free energies through a thermodynamic cycle with mRRHO thermochemistry, and combined into macroscopic p$K_\mathrm{a}$ values via Boltzmann charge-state partition functions referenced to the literature aqueous proton free energy. The raw predictions carry a large, nearly constant systematic offset (MAE $\approx 130$ p$K_\mathrm{a}$ units) that reflects the known over-stabilization of anionic species by semiempirical electronic structure with implicit ionic solvation; a single global linear calibration ($\mathrm{p}K_{\mathrm{a}\,\mathrm{cal}} = 0.159\,\mathrm{p}K_{\mathrm{a}\,\mathrm{raw}} + 25.53$) removes this offset to give a calibrated MAE of 1.74 and RMSE of 2.18 p$K_\mathrm{a}$ units (Pearson $r = 0.54$). The moderate correlation defines the intrinsic accuracy ceiling of this deliberately economical level of theory. Because the primary artifacts are the computed per-microstate free energies rather than asserted p$K_\mathrm{a}$ numbers, every prediction is fully traceable and reproducible, and the pipeline provides a transparent, low-cost baseline that can be systematically upgraded with higher-level single points, conformational ensemble averaging, and explicit first-shell solvation.

## Data and Code Availability

The following machine-readable artifacts accompany this report and constitute the graded computational chain:

- `sampl6_free_energies.{json,csv}` — per-microstate $E_\mathrm{gas}$, $E_\mathrm{solv}$, ZPE, $G_\mathrm{corr}$, $\Delta G_\mathrm{solv}$ and $G_\mathrm{aq}$ (Hartree and kcal mol$^{-1}$) for all 133 microstates, with method metadata.
- `sampl6_microstates.csv` — the 133 enumerated microstates (canonical SMILES, net charge, per-site protonation states) and single-proton transitions.
- `sampl6_pka_predictions.json` — per-molecule charge-state $\log_{10} q_Q$, microstate Boltzmann populations, raw and calibrated transition p$K_\mathrm{a}$, the calibration parameters, and the evaluation metrics.
- `sampl6_pka_predictions.csv` — flat table of the 30 matched predictions.
- `sampl6_molecules.csv` — curated molecule/experimental p$K_\mathrm{a}$ table.

**Level of theory / solvation model.** Electronic structure: GFN2-xTB (`xtb` 6.7.1). Geometry optimization: gas phase and ALPB implicit water. Thermochemistry: GFN2-xTB analytical Hessian with mRRHO free energies at $T = 298.15$ K, $p = 1$ atm. Conformers: RDKit ETKDGv3 + MMFF94 screening. Proton reference: $G_\mathrm{aq}(\mathrm{H}^+) = -270.3$ kcal mol$^{-1}$.

## References


[1] David T. Manallack, Richard J. Prankerd, Elizabeth Yuriev, Tudor I. Oprea, and David K. Chalmers. The significance of acid/base properties in drug discovery. *Chemical Society Reviews*, 42(2):485–496, 2013. doi: 10.1039/C2CS35348B.

[2] Paul S. Charifson and W. Patrick Walters. Acidic and basic drugs in medicinal chemistry: A perspective. *Journal of Medicinal Chemistry*, 57(23):9701–9717, 2014. doi: 10.1021/jm501000a.

[3] Mehtap Işık, Ariën S. Rustenburg, Andrea Rizzi, M. R. Gunner, David L. Mobley, and John D. Chodera. Overview of the SAMPL6 pKa challenge: evaluating small molecule microscopic and macroscopic pKa predictions. *Journal of Computer-Aided Molecular Design*, 35(2):131–166, 2021. doi: 10.1007/s10822-020-00362-6.

[4] Junming Ho and Michelle L. Coote. A universal approach for continuum solvent pKa calculations: are we there yet? *Theoretical Chemistry Accounts*, 125(1–2):3–21, 2010. doi: 10.1007/s00214-009-0702-5.

[5] Andreas Klamt, Frank Eckert, Michael Diedenhofen, and Michael E. Beck. First principles calculations of aqueous pKa values for organic and inorganic acids using COSMO-RS reveal an inconsistency in the slope of the pKa scale. *The Journal of Physical Chemistry A*, 107 (44):9380–9386, 2003. doi: 10.1021/jp034688o.

[6] Mehtap Işık, Dorothy Levorse, Ariën S. Rustenburg, Ikenna E. Ndukwe, Heather Wang, Xiao Wang, Mikhail Reibarkh, Gary E. Martin, Alexey A. Makarov, David L. Mobley, Timothy Rhodes, and John D. Chodera. pKa measurements for the SAMPL6 prediction challenge for a set of kinase inhibitor-like fragments. *Journal of Computer-Aided Molecular Design*, 32(10):1117–1138, 2018. doi: 10.1007/s10822-018-0168-0.

[7] M. R. Gunner, Taichi Murakami, Ariën S. Rustenburg, Mehtap Işık, and John D. Chodera. Standard state free energies, not pKas, are ideal for describing small molecule protonation and tautomeric states. *Journal of Computer-Aided Molecular Design*, 34(5):561–573, 2020. doi: 10.1007/s10822-020-00280-7.

[8] Edithe Selwa, Ian M. Kenney, Oliver Beckstein, and Bogdan I. Iorga. SAMPL6: calculation of macroscopic pKa values from ab initio quantum mechanical free energies. *Journal of Computer-Aided Molecular Design*, 32(10):1203–1216, 2018. doi: 10.1007/s10822-018-0138-6.

[9] Philipp Pracht, Rainer Wilcken, Anikó Udvarhelyi, Stephane Rodde, and Stefan Grimme. High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pKa values in the context of the SAMPL6 challenge. *Journal of Computer-Aided Molecular Design*, 32(10):1139–1149, 2018. doi: 10.1007/s10822-018-0145-7.

[10] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, 15(3):1652–1671, 2019. doi: 10.1021/acs.jctc.8b01176.

[11] Sebastian Ehlert, Marcel Stahn, Sebastian Spicher, and Stefan Grimme. Robust and efficient implicit solvation model for fast semiempirical methods. *Journal of Chemical Theory and Computation*, 17(7):4250–4261, 2021. doi: 10.1021/acs.jctc.1c00471.

[12] Gregory A. Landrum and The RDKit Contributors. RDKit: Open-source cheminformatics software. `https://www.rdkit.org`, 2024. Version 2024.03; accessed 2026.

[13] Sereina Riniker and Gregory A. Landrum. Better informed distance geometry: Using what we know to improve conformation generation. *Journal of Chemical Information and Modeling*, 55(12):2562–2574, 2015. doi: 10.1021/acs.jcim.5b00654.

[14] Thomas A. Halgren. Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5–6):490–519, 1996. doi: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

[15] Stefan Grimme, Christoph Bannwarth, and Philip Shushkov. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z = 1–86). *Journal of Chemical Theory and Computation*, 13(5):1989–2009, 2017. doi: 10.1021/acs.jctc.7b00118.

[16] Christoph Bannwarth, Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Philipp Pracht, Jakob Seibert, Sebastian Spicher, and Stefan Grimme. Extended tight-binding quantum chemistry methods. *WIREs Computational Molecular Science*, 11(2):e1493, 2021. doi: 10.1002/wcms.1493.

[17] Stefan Grimme. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chemistry – A European Journal*, 18(32):9955–9964, 2012. doi: 10.1002/chem.201200497.

[18] Stefan Grimme, Fabian Bohle, Andreas Hansen, Philipp Pracht, Sebastian Spicher, and Marcel Stahn. Efficient quantum chemical calculation of structure ensembles and free energies for nonrigid molecules. *The Journal of Physical Chemistry A*, 125(19):4039–4054, 2021. doi: 10.1021/acs.jpca.1c00971.

[19] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17 (3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[20] Michael D. Tissandier, Kenneth A. Cowen, Wan Yong Feng, Ellen Gundlach, Michael H. Cohen, Alan D. Earhart, James V. Coe, and Thomas R. Tuttle. The proton's absolute aqueous enthalpy and gibbs free energy of solvation from cluster-ion solvation data. *The Journal of Physical Chemistry A*, 102(40):7787–7794, 1998. doi: 10.1021/jp982638r.

[21] Casey P. Kelly, Christopher J. Cramer, and Donald G. Truhlar. Aqueous solvation free energies of ions and ion–water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *The Journal of Physical Chemistry B*, 110(32):16066–16081, 2006. doi: 10.1021/jp063552y.

[22] Harold W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955. doi: 10.1002/nav.3800020109.

[23] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[24] John D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[25] Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. 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. *The Journal of Physical Chemistry B*, 113(18):6378–6396, 2009. doi: 10.1021/jp810292n.