

ADVANCED CHEMISTRY DEVELOPMENT, INC.

Improved Algorithm for $\log D$ Calculation Within the Percepta[®] Platform

Percepta v2024 as a New Step in PhysChem Prediction

Introduction

The algorithm offered by ACD/Labs for calculating pH-dependent octanol/water distribution coefficient ($\log D$) is a highly accurate and robust method to evaluate lipophilicity of drug-like molecules at a variety of physiologically relevant conditions.

The unique feature of the underlying pK_a predictor is its ability to account for all individual ionic forms present in the solution at given pH. This applies not only to the ionization profile itself, but also to all pK_a -dependent property predictions, such as $\log D$ or $\log S$.

However, exhaustive evaluation of all relevant microspecies comes with several tradeoffs:

- 1 **Calculation speed.** Molecules with multiple ionizable centers may have very complex ionization graphs, and their complete traversal consumes significant computational resources, resulting in calculations being perceived as slow.
- 2 **Normalization error.** Aggregating predictions from a large number of ionic forms necessitates using a special procedure for normalization of pK_a micro-constants, which can lead to significant discrepancies between the shape of the simulated $\log D$ vs. pH curve (Fig. 1), and the plot expected from the compound's apparent pK_a profile. This can cause confusion during analysis of prediction results, raising questions about the accuracy of the $\log D$ prediction itself.

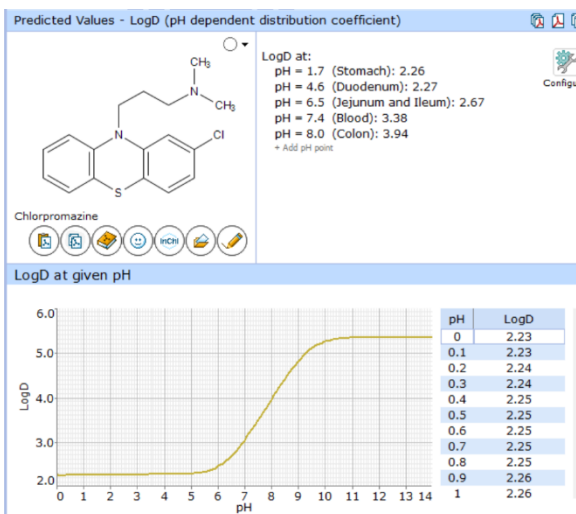
LogD Algorithm

The $\log D$ value of a molecule at given pH depends on the $\log P$ of the neutral form, and the fractions of all ionic species present at that pH. Distribution of ionic species directly depends on all pK_a micro-constants of the molecule. To calculate the distribution of ionic forms from pH, it is necessary to solve a non-linear system of equations for concentration values where pK_a micro-constants are used as coefficients.

The new generation ACD/ pK_a Classic algorithm released with Percepta v2023 improved the accuracy of pK_a predictions by the expansion of the training dataset by ~20%, and also addressed the calculation speed issue, resulting in up to 5–10-fold faster raw pK_a predictions for complex multi-protic molecules.¹ Dedicated optimizations helped to translate the speed improvement to other pK_a -dependent properties, such as $\log D$ or $\log S$.

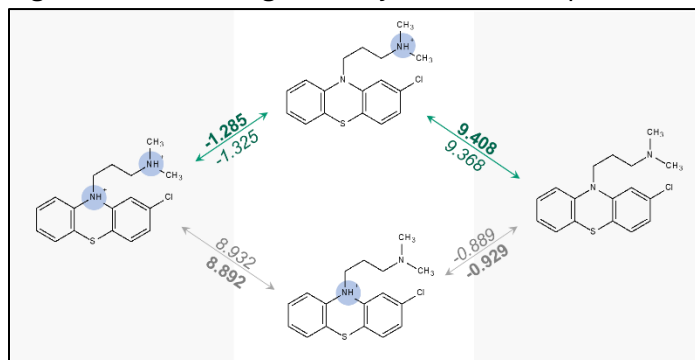
The primary goal of the $\log D$ algorithm update in Percepta v2024 was minimizing the discrepancies in the shapes of the simulated $\log D$ vs. pH and charge vs. pH curves.

FIGURE 1. Simulated $\log D$ curve in Percepta interface.



Problem: Normalization Error

Due to the excessive number of pK_a micro-constants, in a multi-protic molecule, the system of equations is "overprovisioned" and does not have a single exact solution. Instead, the final "normalized" solution must be determined by least squares optimization. pK_a micro-constants are also not exact but predicted with some error. Accumulating these prediction and normalization errors quite often leads to a situation where the distribution of ionic species, and subsequently, $\log D$, etc., deviate significantly from the compound's ionization profile as defined by predicted



values of its "major" pK_a micro-constants (see Fig. 2).

FIGURE 2. Chlorpromazine dissociation scheme with predicted pK_a micro-constants in Percepta v2022.

Bold – before normalization, *italic* – after normalization

As shown in Table 1, $\log D$ values estimated using the normalized pK_a micro-constants do not exactly match $\log D$ values back-calculated from $\log P$ and pK_a . Even larger discrepancies are observed in the distribution of ionic forms, where the midpoint in the transition between positively charged and neutral form is somewhat shifted from the predicted pK_a value of chlorpromazine.

TABLE 1. Deviations in $\log D$ and ionic form predictions introduced by normalization.

pH	logD from logP & pK_a	logD v. 2022	logD Diff.	pH	Positive Form Fraction (%)		
					From pK_a	v. 2022	Diff.
1	2.100	2.102	0.002	8.8	80.2	78.7	1.5
2	2.100	2.103	0.003	8.9	76.3	74.6	1.7
3	2.100	2.104	0.004	9.0	71.9	70.0	1.9
4	2.102	2.106	0.004	9.1	67.0	64.9	2.1
5	2.121	2.126	0.005	9.2	61.7	59.5	2.2
6	2.274	2.291	0.017	9.3	56.2	53.9	2.3
7	2.871	2.908	0.037	9.4	50.5	48.1	2.4
8	3.784	3.825	0.041	$pK_a = 9.408$	50.0	47.7	2.3
9	4.650	4.682	0.032	9.5	44.7	42.4	2.3
10	5.101	5.112	0.011	9.6	39.1	36.9	2.2
11	5.189	5.193	0.004	9.7	33.8	31.8	2.0
12	5.199	5.202	0.003	9.8	28.9	27.0	1.9
13	5.200	5.203	0.003	9.9	24.4	22.7	1.7
14	5.200	5.203	0.003	10.0	20.4	18.9	1.5

Deviations are evident even for this relatively simple molecule that has only two ionizable centers, but they can be significantly more pronounced for more complex molecules with a large number of ionizable centers. This is particularly undesirable given that micro-constants of the major dissociation reactions, and consequently the apparent pK_a values, are predicted very accurately.

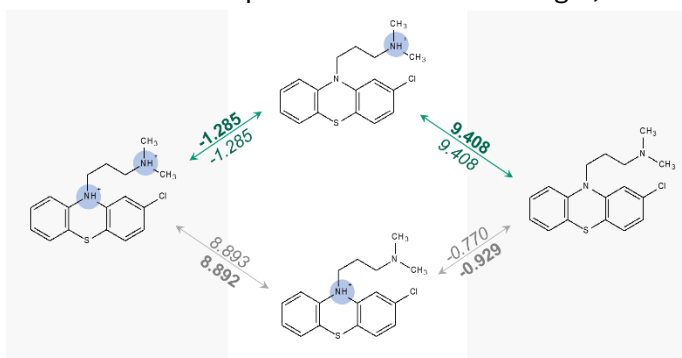
In principle, it would be possible to discard minor microstages altogether to avoid slowdown and this additional source of errors. However, there are several reasons to retain them in the calculations:

- To identify which microstages are important, and which are not, all must be calculated anyway.
- The importance of microstages is relative, as "major" and "minor" micro- pK_a for the same stage can be quite similar. Ignoring some of them may lead to loss of ionic forms with a significant presence in solution.
- For molecules with multiple ionizable centers, estimating the relevance of individual microstages is non-trivial, as any of them may turn out to be critical for the final solution.

Solution: Smart Normalization

The most viable approach to the problem was improving the normalization procedure itself, which led to the development of "Smart normalization" in Percepta v2024. The solution to the overprovisioned system of equations is still determined by minimizing the total error of all equations, but the new approach also takes into account the relative "importance" of each microstage, ensuring that the most important micro-constants deviate less from the original values before normalization. The change in calculation results for Chlorpromazine is shown in Fig.3, where the numbers in italics now represent smart-normalized values.

FIGURE 3. Chlorpromazine dissociation scheme with predicted pK_a micro-constants in Percepta v2024.



Smart normalization brings pK_a micro-constants for the most relevant dissociation reactions much closer to their initial values, leading to a significant reduction in discrepancies between $\log D$ curves estimated by different methods. The same applies to ionic form percentages, and most importantly, the midpoint (50.0%) is exactly at $pH = pK_a$ (Table 2).

TABLE 2. Improvement in accuracy of $\log D$ and ionic form predictions after smart normalization.

pH	$\log D$			pH	Positive Form Fraction (%)		
	from $\log P$ & pK_a	v. 2023	Diff.		From pK_a	v. 2023	Diff.
1	2.100	2.101	0.001	8.8	80.2	80.2	0.0
2	2.100	2.103	0.003	8.9	76.3	76.3	0.0
3	2.100	2.104	0.004	9.0	71.9	71.9	0.0
4	2.102	2.106	0.004	9.1	67.0	67.0	0.0
5	2.121	2.124	0.003	9.2	61.7	61.7	0.0
6	2.274	2.277	0.003	9.3	56.2	56.2	0.0
7	2.871	2.874	0.003	9.4	50.5	50.4	0.1
8	3.784	3.788	0.004	$pK_a=9.408$	50.0	50.0	0.0
9	4.650	4.653	0.003	9.5	44.7	44.7	0.0
10	5.101	5.105	0.004	9.6	39.1	39.1	0.0
11	5.189	5.192	0.003	9.7	33.8	33.8	0.0
12	5.199	5.202	0.003	9.8	28.9	28.8	0.1
13	5.200	5.203	0.003	9.9	24.4	24.3	0.1
14	5.200	5.203	0.003	10.0	20.4	20.4	0.0

Conclusion

- 1 The updated $\log D$ algorithm in Percepta v2024 almost completely resolves the inaccuracies in $\log D$ curve calculations caused by the redundancy of the system of equations for pK_a equilibrium.
- 2 The same applies to ionic forms distribution at different pH values, as well as pK_a or $\log D$ -dependent properties (Solubility-pH profile, absorption & bioavailability, pharmacokinetic parameters, etc.). The achieved improvements are also useful for chromatography applications.
- 3 The changes in $\log D$ /ionic forms calculations complement significant improvements in pK_a calculations available since Percepta v2023.

References

- 1 Lanevskij K, Sazonovas A, Proskura A, Kolovanov E. Ionization (pK_a) Prediction In Percepta v2023: Improvements and Evaluation. PhysChem Forum 2023, Gothenburg, Sweden, Oct. 3-4, 2023.