H-H model and Drug pK models

Rohan Biswas

1 Henderson-Hasselbalch Equation

Expression relating the (a) K_a of a weak acid, HA, and the pH of a solution of a weak acid or (b) K_b of a weak base, and the pOH of a solution of a weak base.

1. For Weak Acids :

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{1}$$

After taking logarithm and rearranging we get the final expression:

$$\boxed{\text{pH} = \text{p}K_a + \log\left(\frac{[A^-]}{[HA]}\right)} \tag{2}$$

2. For Weak Base (M-OH case):

$$K_b = \frac{[M^+][OH^-]}{[MOH]} \tag{3}$$

We get:

$$\boxed{\text{pOH} = \text{p}K_b + \log\left(\frac{[M^+]}{[MOH]}\right)} \tag{4}$$

3. For Weak Base (R-NH_2):

$$K_b = \frac{[R-NH_3^+][OH^-]}{[R-NH_2]} \tag{5}$$

We get:

$$\boxed{\text{pOH} = \text{p}K_b + \log\left(\frac{[R-NH_3^+]}{[R-NH_2]}\right)} \tag{6}$$

2 Polyprotic Speciation

We study fractional compositions (α) and the foundational math for calculating fractional compositions which is the proportion of each species (e.g., H_2A, HA^-, A^{2-}) at any pH. These fractions are crucial for ionization profiles, logD (distribution coefficient) in Shore partitioning, net charge calculations, etc.

2.1 Monoprotic Acid($HA \rightleftharpoons H^+ + A^-$

Equilibrium arises when we get K_a that is governed by equation (1).

The fractional compositions are:

$$\begin{aligned} \alpha_{HA} &= \frac{[H^+]}{[H^+] + K_a} \\ \alpha_{A^-} &= \frac{K_a}{[H^+] + K_a} \end{aligned} \tag{7}$$

After using mass balance and rearranging, we get:

$$\alpha_{HA} = \frac{1}{1 + 10^{pH - pK_a}} \ , \ \alpha_{A^-} = \frac{1}{1 + 10^{pK_a - pH}} \tag{8}$$

Connection to Henderson-Hasselbalch equation:

$$\frac{[A^-]}{[HA]} = 10^{pH - pK_a} \tag{9}$$

For weak bases, use pK_a of conjugate acid BH^+.

2.2 Diprotic Acid ($H_2A \rightleftharpoons HA^- + H^+ \rightleftharpoons A^{2-} + 2H^+$))

The equilibria :

$$K_{a1} = \frac{[H^+][HA^-]}{[H_2A]} \ , \ K_{a2} = \frac{[H^+][A^{2-}]}{[HA^-]} \tag{10}$$

The fractional compositions are :

$$\begin{aligned} \alpha_{H_2A} &= \frac{[H^+]^2}{D} \\ \alpha_{HA^-} &= \frac{K_{a1}[H^+]}{D} \\ \alpha_{A^{2-}} &= \frac{K_{a1}K_{a2}}{D} \end{aligned} \tag{11}$$

where $D = [H^+]^2 + K_{a1}[H^+] + K_{a1}K_{a2}$

In terms of $pH,\ pK_{a1},\ and\ pK_{a2}$.

$$\begin{aligned} \alpha_{H_2A} &= \frac{10^{-2\ pH}}{pD} \\ \alpha_{HA^-} &= \frac{10^{-(pH + pK_{a1})}}{pD} \\ \alpha_{A^{2-}} &= \frac{10^{-(pK_{a1} + pK_{a2})}}{pD} \end{aligned} \tag{12}$$

where : $pD = 10^{-2\ pH} + 10^{-(pH + pK_{a1})} + 10^{-(pK_{a1} + pK_{a2})}$

2.3 Higher Polyprotic

Say for instance H_3A:

$$pD = 10^{-3pH} + 10^{-(2pH + pK_{a1})} + 10^{-(pH + pK_{a1} + pK_{a2})} + 10^{-(pK_{a1} + pK_{a2} + pK_{a3})} \tag{13}$$

$$\alpha_{H_3A} = \frac{10^{-3pH}}{pD} \ , \ \alpha_{H_2A^-} = \frac{10^{-(2pH + pK_{a1})}}{pD} \ , \cdots \tag{14}$$

We derived these fractions by solving the equilibrium condition - setting K_a expressions equal to concentration ratios. This tells us **where** the system ends up. It does not tell us **how** it gets there, or **why** it is stable. Those questions belong to dynamics. The same fractions emerge naturally as the fixed points of the kinetic rate equations.

3 Stability of the reactions

We take a look at these reactions from a dynamical system Point of View.

3.1 Monoprotic Case (1D ODE)

If we consider the elementary reaction :

$$HA \underset{k_a}{\overset{k_d}{\rightleftharpoons}} H^+ + A^- \tag{15}$$

- Forward (dissociation) rate: $k_d[HA]$
- Reverse (association) rate : $k_a[H^+][A^-]$

Now we define $\alpha = \frac{[A^-]}{C}$ where $C = [HA] + [A^-]$ is total concentration. Then the differential equation becomes:

$$\frac{d\alpha}{dt} = \frac{1}{C}\frac{d[A^-]}{dt} = \frac{1}{C}(k_d[HA] - k_a[H^+][A^-]) \tag{16}$$

Substituting $[HA] = C(1-\alpha)$ and $[A^-] = C\alpha$:

$$\boxed{\frac{d\alpha}{dt} = k_d(1-\alpha) - k_a[H^+]\alpha} \tag{17}$$

This is linear in α because $[H^+] = 10^{-pH}$ is treated as a constant (buffered pH) which is standard pseudo-first order approximation.

Now to find fixed point, we set $\dot{\alpha} = 0$, and we get the fixed points as:

$$\alpha^* = \frac{k_d}{k_d + k_a[H^+]} = \frac{K_a}{K_a + [H^+]} = \frac{1}{1 + 10^{(pK_a - pH)}} \tag{18}$$

This is same as equation 8. H-H is the fixed-point condition, not a separate result.
Now, if we take a look at the stability, we gotta linearise:

$$f'(\alpha) = -(k_d + k_a[H^+]) < 0 \quad \forall [H^+] > 0 \tag{19}$$

which means : globally stable. Relaxation timescale $\tau = 1/(k_d + k_a[H^+]) \sim$ microseconds $\rightarrow$ equilibrium is always what you observe.

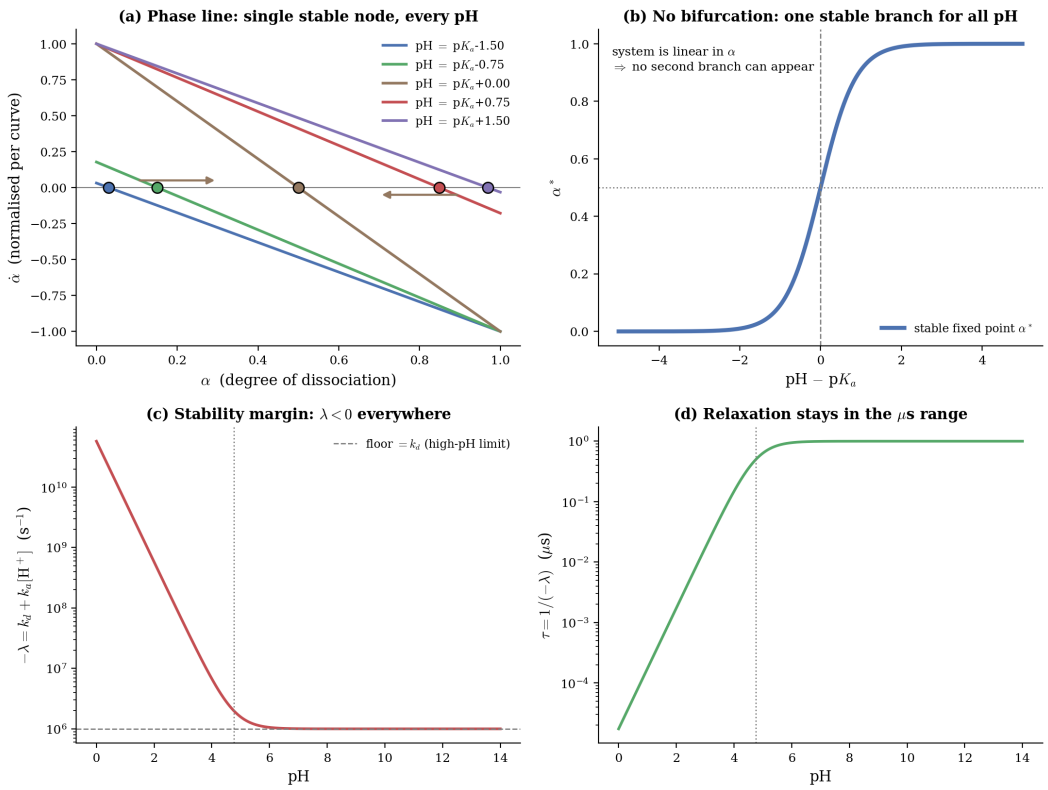

Figure 1: Global stability of the monoprotic dissociation $HA \rightleftharpoons H^+ + A^-$. (a) Phase line $\dot{\alpha}$ vs. α (normalised per curve) for pH = p$K_a + \{-1.5, -0.75, 0, 0.75, 1.5\}$: each curve is linear with a single root (filled circle) and negative slope, so every pH yields exactly one, globally stable fixed point. Arrows on the pK_a curve show the flow converging onto α^* from both sides. (b) The fixed point α^* as a function of pH−pK_a is the Henderson-Hasselbalch sigmoid; it forms a single stable branch for all pH, confirming that no bifurcation can occur in a 1D linear system. (c) Stability margin $-\lambda = k_d + k_a[\mathrm{H}^+]$, strictly positive for all pH and bounded below by k_d even as $[\mathrm{H}^+] \to 0$. (d) Corresponding relaxation time $\tau = 1/(-\lambda)$, remaining in the microsecond range across the full pH axis. Parameters: p$K_a = 4.76$, $k_d = 10^6\,\mathrm{s}^{-1}$, $k_a = k_d/K_a$.

3.2 Diprotic : 2D Linear System

For $H_2A \rightleftharpoons HA^- \rightleftharpoons A^{2-}$, we have four rate constants $(k_{d1}, k_{a1}, k_{d2}, k_{a2})$ and three species. Conservation fixes $\alpha_0 = 1 - \alpha_1 - \alpha_2$, leaving two independent variables. Writing $h = [H^+]$:

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} &= \begin{pmatrix}-(k_{d1} + k_{a1}h + k_{d2}) & k_{a2}h - k_{d1}\\ k_{d2} & -k_{a2}h\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} + \begin{pmatrix}k_{d1}\\ 0\end{pmatrix} \\ &= J\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} + \begin{pmatrix}k_{d1}\\ 0\end{pmatrix} \end{aligned} \tag{20}$$

Note : J is just a substitution for simplicity.

Now we find trace and determinant .

$$\begin{aligned} \mathrm{tr}(J) &= -(k_{d1} + k_{a1}h + k_{d2} + k_{a2}h) < 0 \\ \det(J) &= k_{d1}(k_{a2}h + k_{d2}) + k_{a1}k_{a2}h^2 > 0 \end{aligned} \tag{21}$$

Now, we know the characteristic equation of eigenvalues : $\lambda^2 - \tau\lambda + \Delta$, where λ is eigenvalue, τ is trace and Δ is determinant, we can have the equation:

$$\lambda = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2} \tag{22}$$

The discriminant value $\tau^2 - 4\Delta$ in our current case is also positive. Which means we have a globally stable "Node".

Normal thing is to invert J, we set $\dot{\alpha_0} = \dot{\alpha_1} = \dot{\alpha_2} = 0$ simultaneously. Each equation reduces to one step being individually in balance:

$$k_{d1}\alpha_0^* = k_{a1}\alpha_1^* \quad , \quad k_{d2}\alpha_1^* = k_{a2}\alpha_2^* \tag{23}$$

These are just equilibrium conditions for each step. Rearranging:

$$\frac{\alpha_1^*}{\alpha_0^*} = \frac{K_{a1}}{h} \quad , \quad \frac{\alpha_2^*}{\alpha_1^*} = \frac{K_{a2}}{h} \tag{24}$$

So:

$$\alpha_1^* = \frac{K_{a1}}{h}\alpha_0^* \quad , \quad \alpha_2^* = \frac{K_{a1}K_{a2}}{h^2}\alpha_0^* \tag{25}$$

Applying conservation $\alpha_0^* + \alpha_1^* + \alpha_2^* = 1$:

$$\alpha_0^* = \frac{h^2}{h^2 + K_{a1}h + K_{a1}K_{a2}} \tag{26}$$

This is the diprotic α_{H_2A} from equation 11. The other two follow immediately.

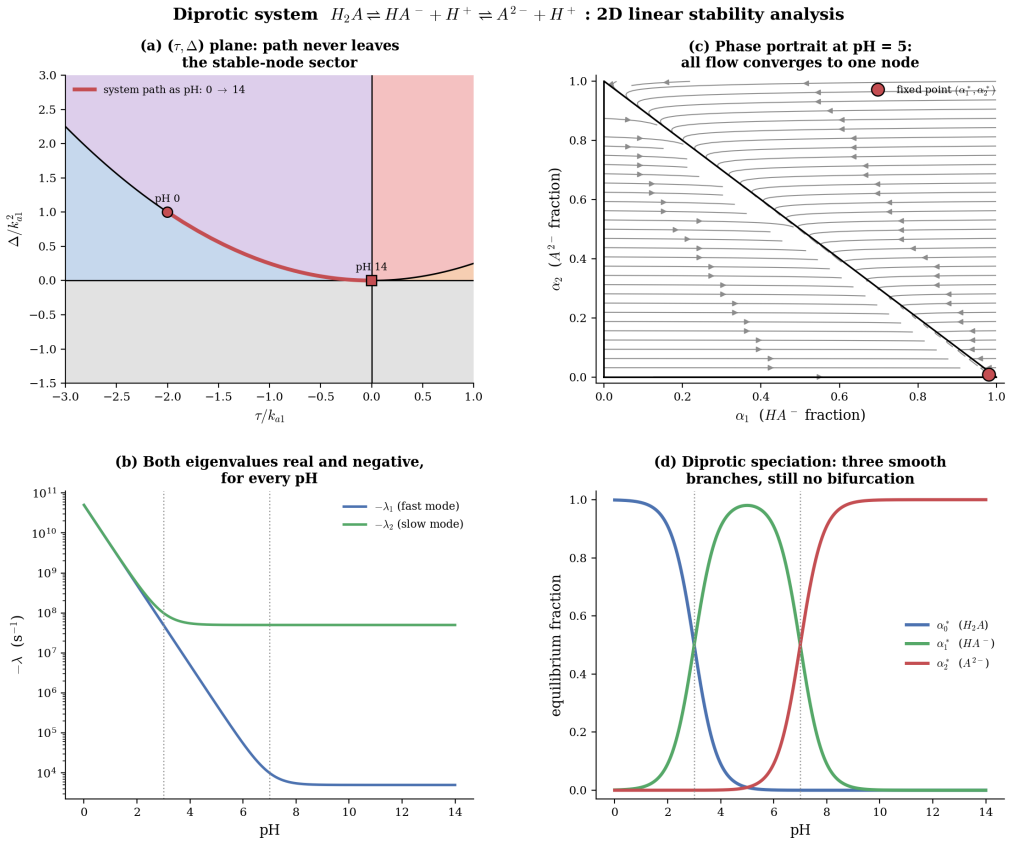

Figure 2: Stability of the diprotic system $H_2A \rightleftharpoons HA^- + H^+ \rightleftharpoons A^{2-} + H^+$. (a) Trace-determinant (τ, Δ) stability plane: the parabola $\tau^2 = 4\Delta$ separates real-eigenvalue (node) from complex-eigenvalue (spiral) behaviour, and $\tau = 0$ separates stable from unstable. The system's path as pH sweeps from 0 to 14 (red curve) remains entirely within the stable-node sector for every pH, confirming the discriminant $\tau^2 - 4\Delta > 0$ established analytically. (b) Both eigenvalues $-\lambda_{1,2}$ of the Jacobian J are real and positive (i.e. $\lambda_{1,2} < 0$) across the full pH range, with no crossing into complex behaviour, consistent with (a). (c) Phase portrait of the flow in the physical simplex (α_1, α_2) at pH = 5: all trajectories converge to the single fixed point (filled circle), the defining signature of a stable node. (d) Equilibrium speciation $\alpha_0^*, \alpha_1^*, \alpha_2^*$ vs. pH, generalising the Henderson-Hasselbalch curve to two ionisable protons; three smooth, single-valued branches with no bifurcation structure. Parameters: $\mathrm{p}K_{a1} = 3$, $\mathrm{p}K_{a2} = 7$, $k_{a1} = k_{a2} = 5 \times 10^{10}\,\mathrm{M^{-1}s^{-1}}$.

3.3 General n-Protic

The detailed-balance argument generalizes directly. At the fixed-point of the n-dimensional linear system, each of the n-steps satisfied:

$$\frac{\alpha_k^*}{a_{k-1}^*} = \frac{K_{ak}}{h} \implies \alpha_k^* = \frac{K_{a1}K_{a2}\cdots K_{ak}}{h^k}\alpha_0^* \tag{27}$$

Applying conservation $\sum_{k=0}^{n} \alpha_k^* = 1$:

$$\alpha_0^* = \frac{1}{D}\ , \quad D = \sum_{k=0}^{n} \frac{\prod_{j=1}^{k} K_{aj}}{h^k} \tag{28}$$

which in pH notation is $D = \sum_{k=0}^{n} 10^{k \cdot pH - \sum_{j=1}^{k} pK_{aj}}$ - exactly the denominator from equation 13.

The stability statement for the general case: J has a tridiagonal structure (each species communicates only with adjacent species in the ionization ladder). So, a detailed balance guarantees all eigenvalues are real and negative for any $h > 0$.

4 Shore pH-Partition Theory

4.1 The physical context

A biological membrane is a lipid bilayer which is a 3–5 nm slab of non-polar hydrocarbon chains. A charged ion cannot dissolve in a non-polar medium (Born energy barrier $\sim$40 $k_B T$ for a monovalent ion). Only the uncharged form of a drug crosses by passive diffusion. This is the single assumption underlying the entire Shore model.

4.2 Two partition coefficients: P vs D

Firstly we define the **intrinsic partition coefficient** P as the equilibrium ratio of drug concentrations across an octanol/water interface, measured "only" for the neutral species:

$$P = \frac{[\text{drug}]_{\text{octanol, neutral}}}{[\text{drug}]_{\text{water, neutral}}} \qquad (\log P \text{ is pH-independent}) \tag{29}$$

P is a property of the molecule alone. It doesn't change with pH.
The **apparent distribution ratio** D is what you actually measure experimentally: total octanol concentration over total aqueous concentration (neutral + all ionised forms):

$$D = \frac{C_m}{C_w} = \frac{[\text{neutral in octanol}]}{[\text{neutral in water}] + [\text{ionised in water}]} \tag{30}$$

Because only the neutral species partitions into octanol, the numerator equals $P \times [\text{neutral}]_w$.
Substituting $\alpha_{\text{neutral}} = [\text{neutral}]_w / C_w$:

$$D = P \cdot \alpha_{\text{neutral}} \tag{31}$$

Taking $\log_{10}$:

$$\boxed{\log\left(\frac{C_m}{C_w}\right) = \log P + \log \alpha_{\text{neutral}}} \tag{32}$$

This is Shore's equation in its most general form.

4.3 Explicit forms for drug classes

Substituting your equation (8) into (32):

Monoprotic acid (neutral = HA):

$$\log\left(\frac{C_m}{C_w}\right) = \log P - \log\left(1 + 10^{\,\text{pH} - \text{p}K_a}\right) \tag{33}$$

Monoprotic base (neutral = B, pKa of conjugate acid BH^+):

$$\log\left(\frac{C_m}{C_w}\right) = \log P - \log\left(1 + 10^{\,\text{p}K_a - \text{pH}}\right) \tag{34}$$

Diprotic acid (neutral = H_2A; $\alpha_{\text{neutral}} = \alpha_0^*$):

$$\log\left(\frac{C_m}{C_w}\right) = \log P - \log\left(1 + 10^{\,\text{pH} - \text{p}K_{a1}} + 10^{\,2\,\text{pH} - \text{p}K_{a1} - \text{p}K_{a2}}\right) \tag{35}$$

Validation: TFV at pH 7.4 ($\log P = -1.6$, $\mathrm{p}K_{a1} = 3.8$, $\mathrm{p}K_{a2} = 6.7$):

$$\begin{aligned}\log(C_m/C_w) &= -1.6 - \log\bigl(1 + 10^{3.6} + 10^{4.3}\bigr) \\ &= -1.6 - \log(1 + 3981.07 + 19952.63) = -1.6 - 4.38 = \mathbf{-5.98}\end{aligned}$$

This means $C_m/C_w \approx 10^{-6}$ which is essentially zero membrane penetration by passive diffusion. TDF raises $\log P$ from -1.6 to $+1.25$ (disoproxil esters mask the charge), bringing $\log(C_m/C_w)$ to -2.4 at plasma pH, which is way more permeable.

4.4 Where Shore fails - the TAF case

TAF has $\log P = 1.6$ and $\mathrm{p}K_a = 3.96$. Shore predicts $\log(C_m/C_w) = -1.84$ at pH 7.4 - very poor passive absorption. Clinical oral bioavailability: **80 %**. The discrepancy is because OATP1B1/B3 hepatic uptake transporters dominate TAF's disposition; they are protein-mediated, saturable, and pH-independent. Shore's model only describes *passive* diffusion. When an active transporter is present, Shore is silent about it.

5 INSTI Three-Layer Ionisation Model

5.1 Why INSTIs need a three-layer model

HIV integrase catalyses two reactions: 3′-processing (cleaves two nucleotides from each viral DNA end) and strand transfer (inserts viral DNA into host chromosome). Both reactions use the conserved D,D-35-E motif — three acidic residues that coordinate **two Mg^{2+} ions** in the active site. INSTIs work by displacing the host DNA and chelating those two metal ions directly.

The pharmacologically active species is therefore *not* just any ionised form of the drug — it is specifically the form that can donate two oxygens to the two Mg^{2+} ions simultaneously. Standard H-H gives the total ionised fraction, but not all of that fraction is chelation-competent. Hence three layers:

5.2 Layer 1 - Henderson-Hasselbalch (standard)

From 3.1:

$$\alpha_1 = \frac{1}{1 + 10^{(\mathrm{p}K_a - \mathrm{pH})}} \qquad \text{(fraction deprotonated at intracellular pH 7.1)} \tag{36}$$

Example at pH 7.1:

Drug	$\mathrm{p}K_a$	α_1
DTG	8.2	0.074
RAL	6.7	0.715
BIC	9.81	0.002

Only 7.4 % of DTG and 0.2 % of BIC are even deprotonated inside the cell. Layer 1 is the strictest bottleneck.

5.3 Layer 2 - Mg^{2+} chelation (Langmuir binding)

Of the deprotonated fraction α_1, only the portion that actually binds Mg^{2+} contributes pharmacological effect. This is a binding equilibrium — at a given free Mg^{2+} concentration $[Mg^{2+}]$, the fraction of deprotonated drug that is chelation-bound follows the **Langmuir adsorption isotherm**:

$$\theta = \frac{[Mg^{2+}]_{\text{free}}}{K_d + [Mg^{2+}]_{\text{free}}} \tag{37}$$

where K_d is the dissociation constant of the drug–Mg^{2+} complex. The Langmuir form arises from a simple 1:1 binding equilibrium: $\text{Drug}^- + Mg^{2+} \rightleftharpoons \text{Drug–}Mg^{2+}$, $K_d = [\text{Drug}^-][Mg^{2+}]/[\text{Drug–}Mg^{2+}]$.
Intracellular free Mg^{2+} is ≈ 0.8 mM (physiological range 0.5–1.0 mM; Romani 2007). The K_d values are drug-specific:

Drug	K_d (μM)	θ at 0.8 mM Mg^{2+}
DTG	5.20	0.9935
CAB	5.50	0.9932
BIC	4.80	0.9940
RAL	7.76	0.9904
EVG	8.10	0.9899

Because $[Mg^{2+}]_{free} \gg K_d$ for all five INSTIs, $\theta \approx 1$ for all of them - Layer 2 barely changes the effective fraction. The real discriminator between the drugs is Layer 1 (different pK_a values) and Layer 3.

5.4 Layer 3 - Keto-enol tautomerism (RAL, EVG only)

Raltegravir and elvitegravir have a **diketo acid (DKA)** pharmacophore which is a β-diketone system with both a keto (C=O) and an enol (C-OH) tautomeric form. The enol tautomer is the only form that can deprotonate to give the oxygen donors needed for Mg^{2+} chelation. In aqueous solution the keto form predominates (Keq < 1); only the fraction in the enol form is chelation-accessible.

$$f_{\text{enol}} = \frac{K_{\text{eq}}}{1 + K_{\text{eq}}} \quad \text{where } K_{\text{eq}} = \frac{[\text{enol}]}{[\text{keto}]} \tag{38}$$

For RAL ($\log K_{\text{eq}} \approx -0.5$): $K_{\text{eq}} \approx 0.316$, so $f_{\text{enol}} \approx 0.24$. Only $\sim$24 % of RAL is even in the right tautomeric form to chelate. DTG, CAB, BIC use a **hydroxypyrimidinone** scaffold instead. Thereby, directly presenting two oxygen donors without any tautomeric requirement ($f_{\text{enol}} = 1.0$).

5.5 Combined effective fraction

$$\boxed{\alpha_{\text{eff}} = \alpha_1 \times \theta \times f_{\text{enol}}} \tag{39}$$

Fold-change vs standard H-H:

$$\text{fold} = \frac{\alpha_{\text{eff}}}{\alpha_1} = \theta \times f_{\text{enol}} \tag{40}$$

For RAL: fold $\approx 0.994 \times 0.24 \approx 0.24$. Standard H-H overestimates RAL's binding-competent fraction by $\sim$4×. For DTG: fold ≈ 1.0. H-H is accurate because no keto-enol correction is needed.
Model caveat: θ and f_{enol} are *pH-independent* constants in this treatment (Mg^{2+} and tautomer equilibria don't shift with pH at physiological range). So the fold-change is flat vs pH which is an algebraic guarantee of the model structure, not an empirical finding.

6 Polyprotic Peptide: Net Charge Z(pH) and pI

6.1 Why peptides are not ladders

The diprotic/polyprotic model assumed a **shared ladder**: both ionisable groups are on the same chemical system, and the species are connected in a single chain $H_nA \to H_{n-1}A^- \to \cdots \to A^{n-}$. The shared denominator D works because the protons are removed *sequentially from the same scaffold*.
A peptide like T-20 has 7 classes of ionisable residues (His, Lys, Asp, Glu, Tyr, N-terminal amino, C-terminal carboxyl) distributed at separate, electronically independent sites along a 36-residue chain. Each site reaches its own equilibrium independently. There is no single sequential ladder — there are independent equilibria happening simultaneously.

6.2 Net charge formula

For each residue type i, let n_i = count, p$K_{a,i}$ = pKa:

- **Basic residue** (His, Lys, Arg, N-terminus): carries +1 when *protonated*, neutral when not. Fraction protonated $= 1/(1 + 10^{\text{pH}-\text{p}K_a})$.
- **Acidic residue** (Asp, Glu, Tyr, Cys, C-terminus): carries −1 when *deprotonated*, neutral when not. Fraction deprotonated $= 1/(1 + 10^{\text{p}K_a-\text{pH}})$.

The net charge is just the sum of all contributions:

$$\boxed{Z(\text{pH}) = \sum_{\text{basic}} n_i \cdot \frac{+1}{1 + 10^{\,\text{pH}-\text{p}K_{a,i}}} + \sum_{\text{acidic}} n_j \cdot \frac{-1}{1 + 10^{\,\text{p}K_{a,j}-\text{pH}}}} \tag{41}$$

This is the same monoprotic H-H formula (equation 8) applied once per residue class and summed — not a new equation, just many copies of equation (8) added together.

6.3 Application to T-20 (Enfuvirtide)

Sequence: `Ac-YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF-NH2` (36 residues, N-term acetylated, C-term amidated).

Residue	pK_a	Count	Type	Contribution at pH 7.4
His	6.0	1	basic	+0.038
Lys	10.53	2	basic	+2.000
Asp	3.7	1	acidic	−1.000
Glu	4.2	6	acidic	−6.000
Tyr	10.07	1	acidic	−0.002
Net charge $Z(7.4)$				**−4.96**

T-20 carries a net charge of ≈ -5.0 at physiological pH (plasma, pH 7.4). Because it is highly charged and hydrophilic, it cannot cross lipid bilayers passively (Born energy barrier ∼40–50 k_BT per charge). This is why T-20 is administered by subcutaneous injection rather than orally.

6.4 Isoelectric Point (pI)

The isoelectric point (pI) is the pH at which the net charge $Z(\text{pH}) = 0$.
For a multi-residue peptide or protein there is no simple closed-form solution (it requires solving a high-degree polynomial). Instead, note that $Z(\text{pH})$ is a continuous, strictly monotonically decreasing function of pH (each titratable group contributes a monotone sigmoid). Therefore there is exactly one root on any interval where Z changes sign.
The standard, robust numerical method is **Brent's method** (`scipy.optimize.brentq`), which combines bisection with inverse quadratic interpolation and reliably finds the root once properly bracketed.

Example: T-20 (Enfuvirtide) Computed pI ≈ 4.22 (verified: $Z(4.22) \approx +0.01 \approx 0$).

Physical Interpretation At pH = pI the molecule has zero net charge and therefore experiences minimum electrostatic repulsion between identical molecules. This usually corresponds to minimum solubility (maximum tendency to aggregate or precipitate).
Formulation scientists therefore avoid buffering a protein or peptide drug exactly at its pI. Clinical relevance: Ibalizumab (IBA) has a pI in the range $\approx 6.5 - 7.0$, close to physiological pH (7.4). This requires careful formulation pH selection (typically slightly away from pI) to maintain solubility and prevent aggregation in the vial or after injection.

7 Monoclonal Antibodies: Why Shore Fails, and What We Use Instead

7.1 Three reasons Shore is inapplicable to IBA

Shore's equation rests on one assumption: the drug crosses the membrane by **passive diffusion** driven by the concentration gradient of the neutral form. For ibalizumab (IBA), a humanised IgG4 monoclonal antibody (∼150 kDa), this assumption fails on three independent grounds:

1. **Size exclusion.** The Born energy penalty for moving a charged species into a non-polar bilayer scales as q^2/r (where r is the ion radius). For a small molecule ($r \approx 0.5$ nm) this is already prohibitive; for a 150 kDa IgG4 with hydrodynamic radius ≈ 5.5 nm, passive bilayer crossing is physically impossible regardless of ionisation state.

2. **Active transport mechanism.** IgG antibodies are distributed via the **neonatal Fc receptor (FcRn)**-mediated transcytosis. FcRn is expressed on endothelial and epithelial cells; it binds IgG at mildly acidic pH (endosome, pH ≈ 6.0) and releases it at neutral pH (cell surface / blood, pH 7.4). This is an active, saturable, receptor-mediated process and not passive diffusion. It is pH-dependent in the *opposite* sense to Shore: IgG binds its transporter more tightly at *low* pH, which is when Shore would predict the molecule to be more neutral/permeable. And that is a completely different physical mechanism.

3. **Route of administration.** IBA is given intravenously, and no GI absorption step exists, so the question Shore answers (will this drug cross the intestinal epithelium?) does not arise.

7.2 What the relevant physical chemistry property is for a mAb

The pI of an IgG determines:

- **FcRn binding affinity** (which controls plasma half-life: IgG with pI closer to 6–7 has better FcRn binding at endosomal pH; very high pI antibodies have faster clearance).
- **Non-specific tissue uptake:** Positively charged IgG (pI $>$ 7.4) binds negatively-charged heparan sulphate proteoglycans on cell surfaces, increasing non-specific uptake and shortening half-life.
- **Formulation stability:** At pH = pI the molecule carries zero net charge, minimum electrostatic repulsion between antibody molecules and minimum solubility and maximum aggregation risk. Formulation buffers are designed to stay away from pI.

7.3 The model we use

Same equation as eqn 41, applied at the amino acid composition level of the full antibody (∼1330 residues across two heavy + two light chains):

$$Z(\mathrm{pH}) = \sum_{\text{basic AA}} n_i \cdot \frac{+1}{1 + 10^{\,\mathrm{pH} - \mathrm{p}K_{a,i}}} + \sum_{\text{acidic AA}} n_j \cdot \frac{-1}{1 + 10^{\,\mathrm{p}K_{a,j} - \mathrm{pH}}} \tag{42}$$

pI found by `scipy.optimize.brentq` on $Z(\mathrm{pH}) = 0$.
Caveat on IBA specifically: The amino acid composition used here is a structural estimate based on a typical humanised IgG4 framework sequence, not ibalizumab's proprietary heavy/light chain sequence. The computed pI should be treated as an order-of-magnitude estimate.