

4.1 Molecular docking: what the computation attempts

Molecular docking predicts the bound geometry (the **pose**) of a small molecule in a target's binding site, and estimates how well it binds (the **score**). It is the engine inside structure based design and docking based screening. Every docking method solves two coupled problems. **Sampling** searches the space of possible poses: the ligand can sit in the site in many positions, rotations and conformations, and the protein may adjust around it. **Scoring** evaluates each sampled pose with a **scoring function** that estimates binding affinity from the interactions the pose makes. Docking fails in the two corresponding ways: the correct pose was never sampled, or it was sampled but scored below wrong poses. The three syllabus methods are points on the sampling spectrum.

4.2 Rigid docking

In **rigid docking**, both ligand and protein are treated as rigid bodies: the ligand's precomputed conformation is placed into the site in every orientation, and complementarity of shape and chemistry is scored. It is the fastest possible docking, able to process very large libraries, and historically the first form (the lock and key picture made computational). Its flaw is the premise: real ligands flex and real sites breathe, so a rigid search misses poses the flexible molecule would find and rejects compounds whose bound conformation differs from the one supplied. Rigid docking survives as a first pass filter in large screens, where its speed removes the obviously impossible before flexible methods spend time on the plausible.

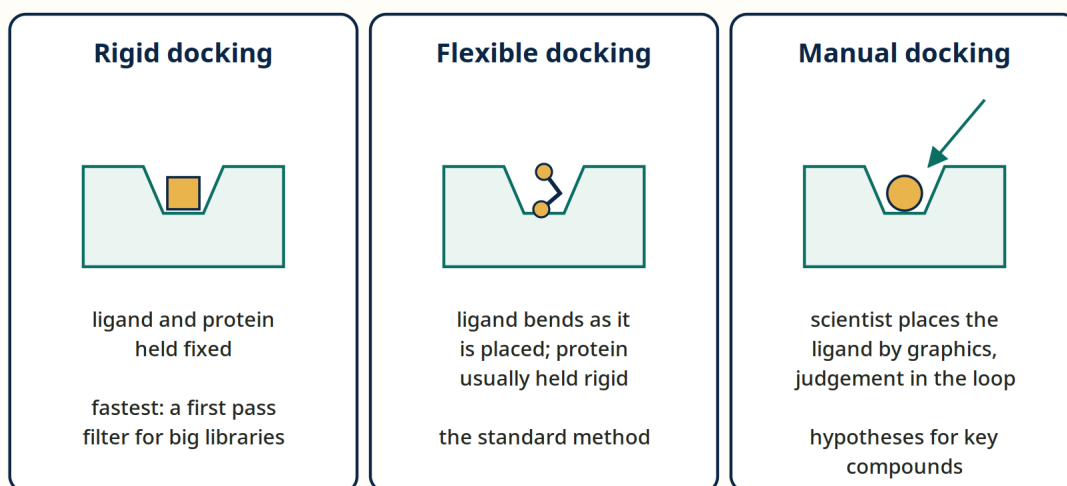
4.3 Flexible docking

Flexible docking allows the ligand to change conformation during the search, exploring its rotatable bonds while it is placed, so the pose found is a genuine bound conformation rather than a precomputed guess. The protein is usually held rigid, or granted limited flexibility in the side chains lining the site, because full protein flexibility multiplies the search space beyond routine computation. Allowing the protein to adjust to the ligand is the **induced fit** problem, and the most capable protocols treat selected side chains as movable or dock against several protein conformations. Flexible docking is the standard working method: slow relative to rigid docking, honest relative to everything the ligand actually does, and accurate enough on redocking tests, where a ligand removed from its crystal structure is usually returned within about 2 angstroms of its true pose, to be trusted for ranking within a programme.

4.4 Manual docking

Manual docking places the ligand in the site interactively, by a scientist using molecular graphics: rotating and adjusting the molecule by eye and chemical judgement, inspecting clashes and unsatisfied hydrogen bonds in real time. It is the oldest form of docking, and it survives because it does what algorithms do badly: incorporate everything the scientist knows, the assay quirks, the analogue data, the suspicion that a residue moves, into one placement decision. Its costs are throughput and reproducibility, so its modern role is hypothesis building and the careful hand examination of the few compounds that matter most.

Three Modes of Molecular Docking



Speed falls and realism rises from rigid to flexible; manual docking trades both for judgement.

© 2026 PharmacoLearn, pharmacolearn.fyi

Figure: Rigid, flexible and manual docking compared: realism rises and speed falls from the first to the second, and judgement enters with the third.

4.5 Scoring functions and their limits

A **scoring function** converts a pose into a number meant to track binding affinity. The families in standard use estimate affinity from force field terms (van der Waals and electrostatic energies of the pose), from empirical terms fitted to measured affinities of known complexes (hydrogen bonds, hydrophobic burial, entropy penalties for frozen rotatable bonds), or from knowledge based potentials distilled from the geometry statistics of many crystal structures. All three share the same honest limitation: they rank far better than they measure. A score can order a series so the best poses rise to the top, which is all screening asks, but it does not deliver a trustworthy absolute affinity, because solvent effects, protein motion and entropy are approximated crudely. Docking therefore concludes with inspection and experiment, never with the score alone, and consensus scoring, requiring several functions to agree, is a common defence against any one function's blind spots.

4.6 Docking based screening

Docking based screening is virtual screening (section 3.7) with docking as its engine: a library is docked into the target's site and ranked, and the top fraction is selected for assay. Performance is measured by enrichment of true actives in that top fraction over random selection, and the method is validated before use by redocking the co-crystallized ligand (which must return close to its experimental pose) and by checking that known actives outrank property matched decoys. Used well, it finds chemotypes similarity methods cannot, because it judges fit to the site rather than resemblance to known ligands. Used carelessly, against a poor structure or with artefacts unfiltered, it fills the assay plate with implausible molecules that score well in the computer and nowhere else, which is why drug likeness filtering and human inspection flank it in the cascade.

4.7 De novo drug design

De novo design builds the ligand inside the site from scratch. Fragments or atoms are placed at the site's interaction hot spots, the places where hydrogen bond partners, hydrophobic contacts or charges are offered, and then grown or linked into a complete molecule that satisfies them all, under the discipline of

valence and synthetic sense. Where docking selects from what exists, de novo design proposes what should exist, and its proposals routinely include chemotypes no library contains.

Its acknowledged weakness is the last step: a structure elegant in the site may be a nightmare to synthesize, or unstable, or insoluble. Modern practice therefore constrains the builder with synthetic accessibility rules and fragment libraries of purchasable pieces, and treats de novo output as inspired starting points for medicinal chemists rather than finished candidates. As a syllabus method it completes the structure based trio: docking tests the known, pharmacophores abstract the known, de novo design invents the new.

4.8 QSAR: history and development

A **quantitative structure activity relationship (QSAR)** expresses biological activity as a mathematical function of a molecule's physicochemical or structural properties, so that activity can be predicted for unsynthesized analogues. Its development, as the syllabus requires, runs in clear steps. In the nineteenth century Crum Brown and Fraser proposed that physiological activity is a function of chemical constitution, the founding idea. Around 1900 Meyer and Overton independently found that narcotic potency tracks oil water partitioning, the first quantitative property relationship and the ancestor of the lipophilicity parameter. Hammett, in the 1930s, gave physical organic chemistry its electronic parameter, sigma, and a linear free energy equation for reactivity. The modern subject begins in the early 1960s, when **Hansch** combined these strands, lipophilicity, electronic and steric terms in one regression equation for biological activity, and **Free and Wilson** published their additive substituent model in the same era. From those two foundations grew the statistical apparatus of Unit 5 and, in the 1980s, the three dimensional extension CoMFA.

4.9 SAR versus QSAR

The distinction the syllabus draws is between reading a pattern and fitting an equation:

Feature	SAR (structure activity relationship)	QSAR (quantitative SAR)
Form	Qualitative: statements about what structural changes raise or lower activity	Quantitative: an equation relating measured activity to numerical parameters
Typical statement	A para halogen increases potency in this series	$\log (1/C) = 0.9 \log P - 0.4 \sigma + 2.1$, with stated statistics
Data needed	A handful of analogues and their activities	A congeneric series with reliable activities and computed or measured parameters, enough compounds to support statistics
What it gives	Design intuition and hypotheses	Predictions for untested analogues, and an optimum (for example of $\log P$) to design toward
Risk	Over reading single pairs of compounds	Over fitting: an equation that describes the data but predicts nothing, controlled by the statistics of Unit 5

SAR is the older and universal habit of medicinal chemistry; QSAR is SAR made numerical and therefore testable, at the price of assumptions the next sections state.

4.10 Physicochemical parameters

QSAR equations are built from parameters, numbers that capture one property of a substituent or molecule. The three classical families are:

Family	Parameter	What it measures
Hydrophobic	$\log P$ (partition coefficient of the whole molecule), and π , the contribution of one substituent to $\log P$ relative to hydrogen	Tendency to leave water for a lipid like environment; governs membrane crossing and hydrophobic binding

Family	Parameter	What it measures
Electronic	Hammett sigma: sigma para and sigma meta describe a substituent's electron withdrawing (positive) or donating power; related constants describe field and resonance components	The electron density a substituent creates at a reaction or binding centre, influencing ionization and interaction strength
Steric	Taft Es (from rates of acid catalysed ester hydrolysis), molar refractivity (MR, size combined with polarizability), and simple size descriptors	The space a substituent occupies and the crowding it causes near the binding interaction

Parameters come from experiment on model systems or from computation, and the parameter tables compiled for common substituents are what make Hansch analysis practical: a new analogue's pi, sigma and Es are looked up or computed, never measured from scratch.

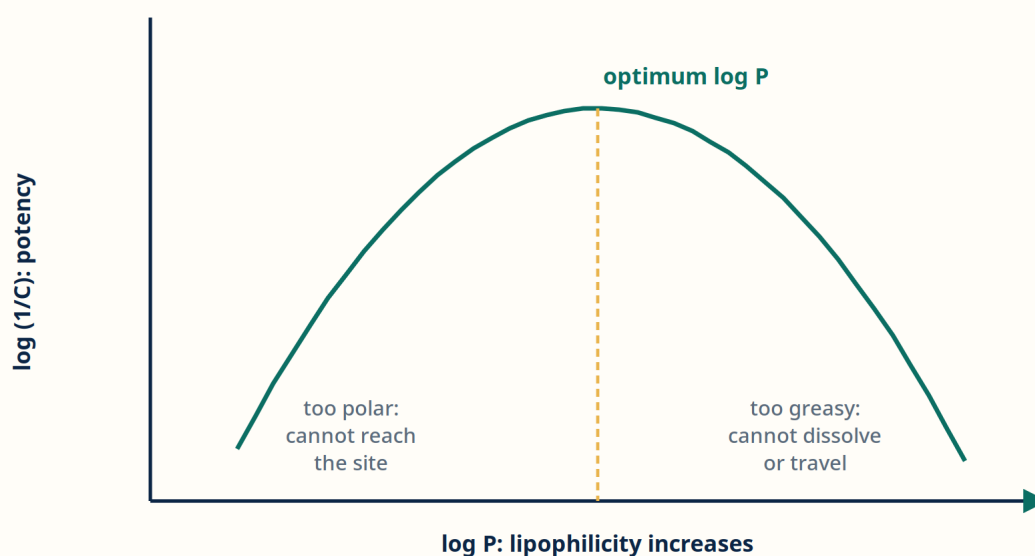
4.11 Hansch analysis

Hansch analysis fits, for a congeneric series acting by one mechanism, an equation of the form:

$$\log (1/C) = a (\log P) - b (\log P)^2 + c (\text{sigma}) + d (\text{Es}) + \text{constant}$$

where C is the molar concentration producing the defined effect, so that $\log (1/C)$ rises with potency. Each term carries a physical story. The $\log P$ terms are characteristically **parabolic**: potency first rises with lipophilicity, as the molecule partitions more readily toward its site, then falls as it becomes too greasy to dissolve, travel and leave depot phases, so the equation contains an optimum $\log P$ for the series, one of the most useful single numbers classical QSAR produces. The sigma term reports whether electron donation or withdrawal at a position helps binding; the Es term penalizes or rewards bulk near the pharmacophore. The coefficients are found by regression (Unit 5) and are believed only when the statistics, the number of compounds per parameter, and the mechanism constancy of the series support them. The method's lasting legacy is conceptual as much as numerical: it taught medicinal chemistry to think of potency as a balance of transport and binding properties that can be optimized separately and deliberately.

The Hansch Parabola: Potency Against log P



The Hansch parabola: potency rises, peaks and falls as lipophilicity grows.

© 2026 PharmacoLearn, pharmacolearn.fyi

Figure: The Hansch parabola. Potency peaks at an optimum $\log P$ and falls on both sides, too polar on the left and too greasy on the right.

4.12 Free Wilson analysis

Free Wilson analysis takes the opposite route to the same quantity. It assumes a series built on a common scaffold with substituents varied at defined positions, and it models activity as the sum of a contribution from each substituent at each position plus the activity of the unsubstituted parent:

$$\log (1/C) \text{ of an analogue} = \text{parent } (\mu) + \text{sum of substituent contributions}$$

A teaching miniature fixes the arithmetic exactly. Suppose the parent compound has $\log (1/C) = 4.00$, a para chloro substituent contributes +0.42 wherever the analysis has established it, and a meta methyl contributes -0.18. The analogue carrying both is predicted at $4.00 + 0.42 - 0.18 = 4.24$, without any measurement. The contributions are extracted from the series itself by regression on indicator variables (present or absent), so the method needs no physicochemical tables and no property assumptions. Its demands are combinatorial instead: the series must vary substituents at each position sufficiently independently for their contributions to be separable, and the model is mute about any substituent never made. Additivity is its wager: where substituents interact, for example one changing the ionization another depends on, the sum errs, and the method says so through its residuals.

4.13 The relationship between Hansch and Free Wilson analysis

The two methods are relatives, and the syllabus asks for the relationship explicitly. Formally, a Free Wilson indicator variable for a substituent behaves like a Hansch parameter that is switched on by that substituent's presence; when a substituent's Hansch parameters (its π , σ , E_s) are themselves used to explain its Free Wilson contribution across many series, the two descriptions meet. Fujita's **mixed approach** operationalizes the union: an equation containing physicochemical parameters for the positions where property logic is trusted, and indicator variables for the positions where it is not. Practically, the choice is pragmatic. Hansch analysis generalizes: from the equation, the activity of a substituent never synthesized can be estimated from its parameter values. Free Wilson analysis is self contained and assumption light, but it can only recombine substituents already present in the data. Modern QSAR descends from both: the descriptors changed, the statistics deepened (Unit 5), and the additivity and property balance ideas remained.

*More free M.Pharm notes and unit wise PDFs at pharmacolearn.fyi
Free video lectures on the PharmacoLearn YouTube channel*