# PharmacoLearn

**MPC203T Computer Aided Drug Design, M.Pharm Semester 2**

## Unit 3: Molecular Modeling and Docking

*Free notes at pharmacolearn.fyi*

### 3.1 Molecular mechanics: the force field idea

Molecular modeling begins with a bargain. Electrons are complicated, so **molecular mechanics (MM)** throws them away and treats a molecule as balls and springs: atoms are spheres with assigned sizes and charges, bonds are springs with natural lengths and stiffnesses, and the **force field** is the rule book of parameters that prices every distortion. The total energy is a sum of simple terms: bond stretching, angle bending, torsional rotation about bonds, and the non bonded terms, van der Waals attraction and repulsion and electrostatic interaction between partial charges. Different force fields, named sets such as MM2, AMBER, CHARMM and OPLS, differ in their parameters and in the atom types they recognise; protein force fields and small molecule force fields are parameterised on different evidence. The payoff is speed: MM handles thousands of atoms and lets us minimise, search conformations and run dynamics on whole protein ligand complexes. The price is absolute: bonds cannot break or form, protonation is fixed when you build the model, and any atom type the force field lacks is a guess. MM answers questions of shape and fit, not of reaction.

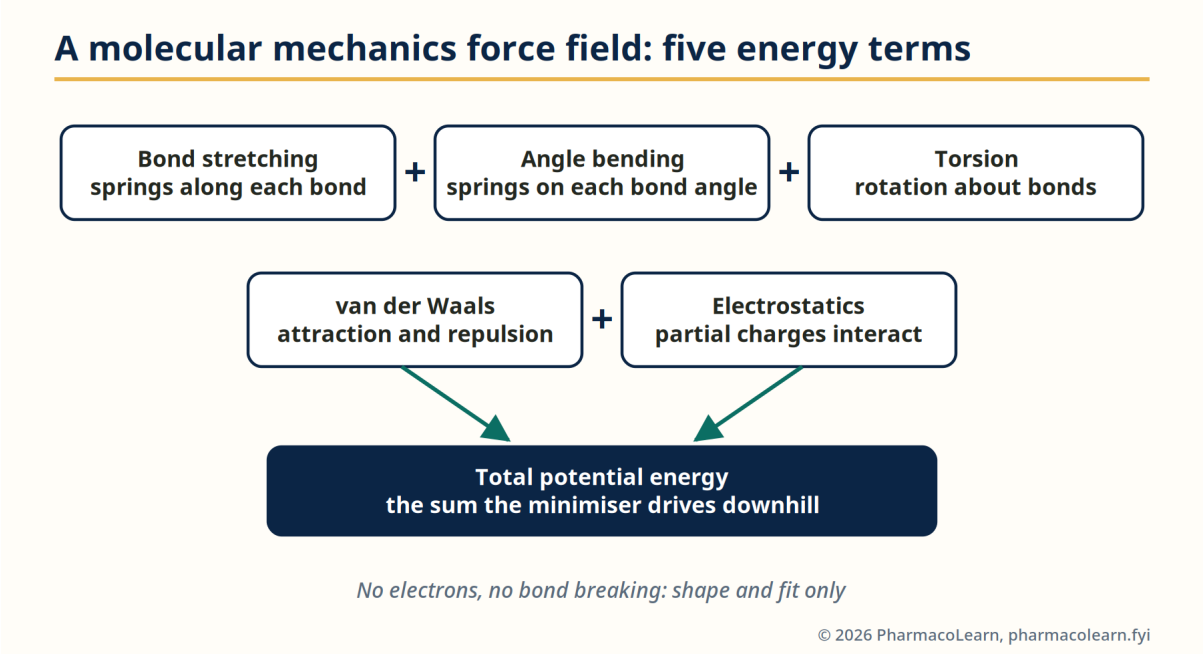


*Figure: Molecular mechanics prices a structure as a sum of bond, angle, torsion, van der Waals and electrostatic terms, with no electrons anywhere in the sum.*

### 3.2 Quantum mechanics in drug design

**Quantum mechanics (QM)** keeps the electrons, and with them the chemistry MM cannot see: bond breaking and making, charge transfer, polarisation, tautomer energies, reaction barriers, and the electrostatic potential surfaces from which accurate atomic charges are derived. The QM ladder is a trade of accuracy against cost. **Semi empirical methods** (the AM1 and PM3 family) simplify the integrals with fitted parameters and run fast enough for routine geometry and charge work on drug sized molecules. **Ab initio Hartree Fock** computes from first principles without fitted parameters, and **density functional**

**theory (DFT)** has become the working standard for accurate geometries, energies and properties of small molecules. In drug design QM earns its cost in specific places: deriving partial charges for a force field, checking which tautomer or protonation state actually binds, modelling covalent inhibitors whose warhead reacts with the target, understanding enzyme reaction mechanisms, and computing descriptors such as HOMO and LUMO energies for electronic QSAR terms. MM and QM are also combined, QM/MM, with the reaction centre treated quantum mechanically inside a mechanically treated protein: the active site gets the physics, the rest of the protein gets the springs.

| Aspect | Molecular mechanics | Quantum mechanics |
|---|---|---|
| Treats | Atoms as balls and springs, no explicit electrons | Electrons explicitly, by approximation scheme |
| Speed | Very fast, thousands of atoms | Slow, tens to a few hundred atoms for careful work |
| Can break bonds | No | Yes |
| Needs | A force field with parameters for every atom type | A method and basis set choice |
| Typical drug design jobs | Minimisation, conformations, dynamics, docking support | Charges, tautomers, covalent warheads, reaction mechanisms |

## 3.3 Energy minimization methods

A built structure arrives strained: bonds stretched by the drawing, atoms overlapping. **Energy minimization** walks the structure downhill on the potential energy surface to the nearest comfortable geometry. The methods differ in what they know about the hill. **Steepest descent** follows the gradient, the direction of steepest fall, in steps: robust far from a minimum when strains are gross, but slow and zigzagging near the bottom. **Conjugate gradient** remembers the previous direction and combines it with the new gradient, curing most of the zigzag and becoming the usual workhorse for large systems. **Newton Raphson** uses the second derivative (the curvature, the Hessian matrix) as well and converges in few steps near a minimum, but the Hessian is expensive to compute and store for big systems, and the method can misbehave far from a minimum; quasi Newton variants approximate the Hessian to keep the benefit. The **simplex method** uses no derivatives at all, moving a geometric figure across the surface, slow but forgiving. The practical protocol is therefore staged: steepest descent first to relieve the worst strain, then conjugate gradient to finish. Every student must hold one caveat fast: minimization finds the **nearest** local minimum, not the lowest one on the surface; that search problem is the next section.

## 3.4 Conformational search: finding the shapes a molecule can take

Flexible molecules live as populations of conformations, and modeling must enumerate the plausible ones. **Systematic search** rotates each rotatable bond in steps (say every 30 or 60 degrees) over the full circle and keeps the combinations that survive an energy cut off: complete for small molecules, but the count explodes combinatorially, roughly the step count raised to the number of rotatable bonds, so it fails beyond a handful of rotors. **Monte Carlo methods** change one feature at random, accept or reject the result by an energy criterion, and sample rather than enumerate, reaching larger and more flexible molecules at the price of no guarantee of completeness. **Molecular dynamics** gives the molecule kinetic energy and integrates Newton's equations, so the structure wanders the surface as it would at temperature, crossing small barriers and revealing flexible regions, hinge motions and water behaviour over time; snapshots along the trajectory are the conformational sample. **Simulated annealing**, heating and slow cooling in silico, is the hybrid used to shake a system out of poor minima. For pharmacophore work (Unit 5) and docking, the working product is a **conformational ensemble**: a stored set of low energy shapes per molecule, from which the binding one is chosen.

## 3.5 Global minimum versus bioactive conformation

The syllabus pairs two ideas that students routinely confuse. The **global minimum conformation** is the single lowest energy shape of the isolated molecule. The **bioactive conformation** is the shape the molecule actually adopts in the binding site. They are often different, and the difference costs energy: binding spends part of its budget forcing the ligand out of its favourite shape into the bound one, a conformational strain the complex must repay with interactions. Two practical consequences follow. First, a modeling study that minimises a ligand and docks only that shape has silently assumed the two conformations coincide; flexible docking and conformational ensembles exist to avoid that assumption. Second, **conformational restriction** is a design strategy: if the bioactive shape is known, locking the molecule into it with a ring or a double bond pre pays the strain and often buys potency and selectivity, at the risk that the locked shape is the wrong guess. Comparing the global minimum with a proposed bioactive conformation, and counting the energy gap between them, is a standard sanity check in structure based work; gaps beyond a few kilocalories per mole make the proposed pose suspect.

## 3.6 Molecular docking: what it is, and how a run proceeds

**Molecular docking** predicts how a small molecule sits in a target's binding site, its pose, and estimates how strongly it binds, its score. It is the workhorse of structure based design and of docking based virtual screening.

```
1. Prepare the protein .... add hydrogens, assign protonation, repair gaps
        |
        v
2. Define the site ........ binding pocket from the bound ligand or geometry
        |
        v
3. Prepare the ligands .... 3D structures, tautomers, ionisation states
        |
        v
4. Sample poses ........... place and rotate the ligand in the pocket
        |
        v
5. Score each pose ........ estimated binding energy or fit score
        |
        v
6. Inspect and test ....... eye, interactions, then the bench decides
```

Preparation is where docking studies succeed or fail quietly: a histidine protonated the wrong way, a missing structural water that actually mediates binding, a ligand drawn in the wrong tautomer, each silently poisons the result. The validation every docking protocol owes is **redocking**: remove the ligand from an experimental complex, dock it back, and demand that the top pose reproduces the experimental one, conventionally within a root mean square deviation of about 2 Angstrom. A program that cannot redock its own answer has not earned trust on unknown ligands.

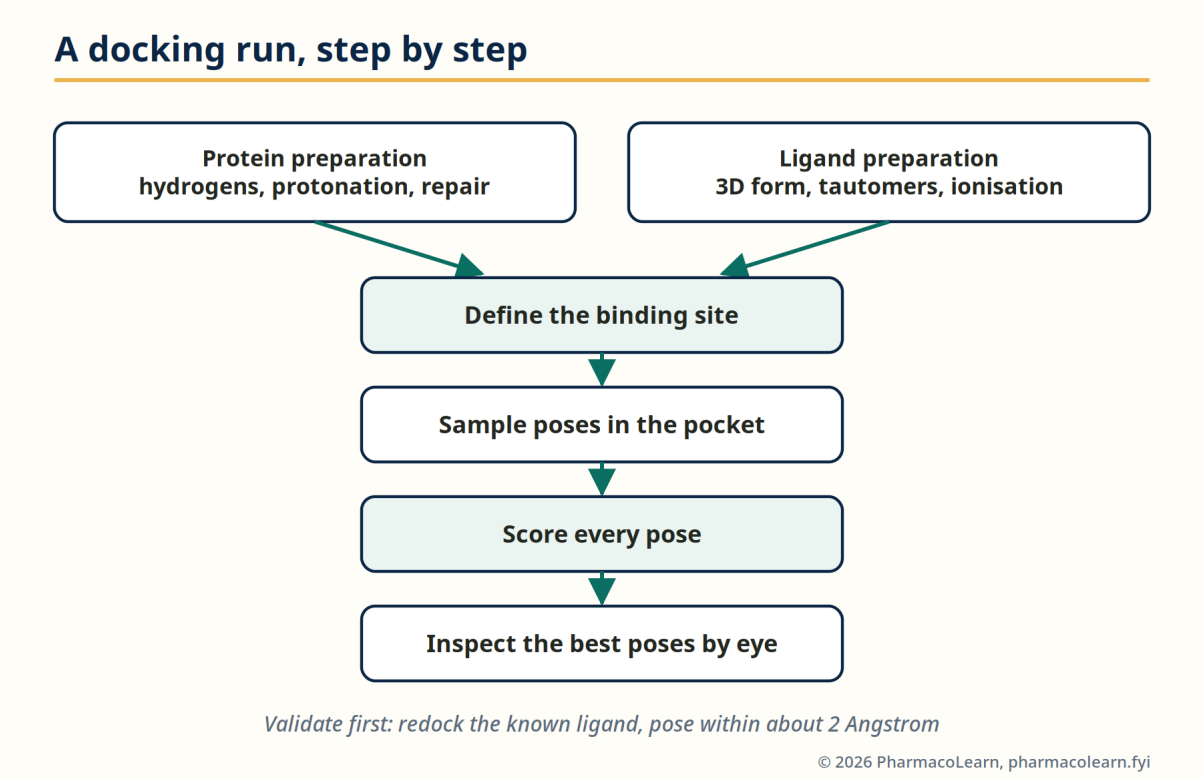


*Figure: A docking run proceeds from careful preparation through sampling and scoring to inspection, and earns trust only by redocking a known ligand correctly.*

## 3.7 Search algorithms and scoring functions

Docking divides labour between a **search algorithm**, which explores placements, and a **scoring function**, which prices them. Search strategies include systematic incremental construction, building the ligand fragment by fragment inside the pocket (the strategy of FlexX and of DOCK's anchor and grow), genetic algorithms that evolve a population of poses under selection for score (the strategy of GOLD and of AutoDock's Lamarckian variant), Monte Carlo and simulated annealing moves, and exhaustive shape matching for rigid fragments. Scoring functions come in three families. **Force field based** scores extend the MM energy of section 3.1 to the complex, counting van der Waals and electrostatic terms between ligand and protein. **Empirical** scores add up fitted contributions, a hydrogen bond term, a hydrophobic contact term, a penalty per rotatable bond, calibrated on complexes with known affinities. **Knowledge based** scores derive potentials from statistics of atom contacts in the structural databases. All three share known weaknesses: they rank poses within one ligand reasonably, compare different ligands shakily, treat water and protein motion crudely, and largely ignore entropy. Docking scores are therefore used to **enrich** a list with plausible binders and to generate binding hypotheses, not as measured affinities; the free energy methods that do better exist, at far greater cost.

## 3.8 Rigid, flexible and extra precision docking

| Mode | Protein | Ligand | Typical use |
|---|---|---|---|
| Rigid docking | Fixed | Fixed, one stored conformation | Very fast first pass over huge libraries |
| Flexible docking | Fixed (or partly) | Flexible, conformations generated during the run | The standard working mode for lead work |
| Extra precision (XP) docking | Fixed | Flexible, with a stricter, more physical scoring stage | Final ranking of shortlists, punishing false positives |

The syllabus's phrase **extra precision docking** comes from the Glide program's XP mode: after a standard precision pass, poses are rescored with a more demanding function that rewards specific motifs, such as hydrogen bonds in hydrophobic enclosures, and penalises the solvent exposed, charge mismatched poses that looser scoring lets through. The general pattern matters more than the brand: docking is run as a **cascade**, cheap and lenient first over millions, strict and expensive last over hundreds, a pattern Unit 5 formalises as the screening protocol.

### 3.9 Drug receptor interactions, in the language of docking

Docking stands or falls on the same interaction catalogue as medicinal chemistry, now with geometry attached. **Hydrogen bonds** are directional and short, roughly 2.5 to 3.2 Angstrom between heavy atoms, and a pose that satisfies the site's donors and acceptors usually beats one that merely fills space. **Hydrophobic contacts** reward burying nonpolar surface against nonpolar pocket walls, the in silico face of the log P story of Unit 1. **Electrostatic interactions** range from salt bridges between full charges to the weak pull of dipoles. **Pi stacking and cation pi interactions** organise aromatic ligands against aromatic or charged side chains. **Metal coordination** dominates in metalloenzyme sites, as with the zinc ion bound by inhibitors of carbonic anhydrase and of matrix metalloproteinases. And **water** is an honorary residue: ordered waters in a pocket may be displaced by a ligand, gaining entropy, or bridged by it, gaining enthalpy, and good protocols decide water by water which treatment applies.

### 3.10 Case target: DHFR and methotrexate

**Dihydrofolate reductase (DHFR)** reduces dihydrofolate to tetrahydrofolate, the carrier that feeds one carbon units into thymidylate and purine synthesis; block it and DNA synthesis starves. Its inhibitor family is the classical docking teaching case. **Methotrexate**, a folate analogue, binds DHFR extremely tightly in place of the substrate, and species selective relatives such as **trimethoprim**, selective for bacterial DHFR, show how small pocket differences between species are read, by eye and by docking, into selective drugs. A docking exercise on DHFR teaches the canonical lessons in one system: the pteridine ring anchors by hydrogen bonds deep in the pocket, the glutamate tail reaches toward solvent and charged residues, and the score separates substrate like poses from nonsense ones. DHFR is also a resistance lesson: point mutations in the pocket that weaken inhibitor binding while sparing substrate turnover are exactly the poses a docking program can be asked to explain.

### 3.11 Case targets: HMG CoA reductase, HIV protease and the cholinesterases

**HMG CoA reductase** catalyses the rate limiting step of cholesterol synthesis, the conversion of HMG CoA to mevalonate, and is the target of the **statins**. The statin pharmacophore carries an HMG like portion that occupies the HMG binding region and a rigid hydrophobic group that fits an adjacent pocket created when the enzyme closes over the inhibitor; docking into this site turns on reproducing the flap closure, a reminder that some pockets only exist in the bound state. **HIV protease**, a symmetric aspartic protease dimer that cleaves the viral polyprotein, was the first great triumph of structure based design: its inhibitors, beginning with **saquinavir**, approved in 1995 as the first protease inhibitor, were designed against the enzyme's known structure, peptidomimetic in ancestry, and anchored on interactions with the catalytic aspartates and the conserved water that bridges inhibitor to the flaps. The **cholinesterases** complete the syllabus set: acetylcholinesterase (AChE), which clears acetylcholine in synapses, has a deep narrow gorge with a catalytic site at its base and a peripheral site at its rim, the geography behind **donepezil**, whose long shape touches both sites, while butyrylcholinesterase (BChE), its plasma and brain relative, differs in gorge lining and selectivity patterns. Docking across these four target families, an oxidoreductase, a reductase of lipid metabolism, a viral protease and two hydrolases, is the syllabus telling the student that the method is general, and that each target still has to be learned on its own terms.