Fusion Protein Docking Simulation – Protocol

Testing our fusion protein on whether enzyme/tag order affects binding affinities and if occupying one site changes the other.

Step	What it does	Tools	Output
0 **Active-site identification**	Find the catalytic and substrate-binding residues of both enzymes from crystal structures — every residue within 4.5 Å of a bound ligand.	RCSB PDB, Biopython	Residue sets mapped onto the fusion sequence
1 **Structure prediction**	Predict each construct several times with different seeds – replicates.	AlphaFold 3	20 structures (4 constructs × 5 seeds)
2 **Local relax — the apo (unbound) reference**	Relax each prediction around both active sites plus a shell, everything else held rigid.	PyRosetta FastRelax (ref2015)	20 docking receptors, and the apo baseline for every later comparison
3 **Ligand preparation**	Get both substrates into a docking-ready form: TACT for the chitinase site, DACB for the deacetylase site.	PubChem, RDKit, Meeko	Two ligand PDBQT files
4 **Docking**	Dock each substrate into its own site, on every receptor. Both dockings run against the same step-2 structures.	AutoDock Vina (rigid receptor)	40 docked poses
5 **Rosetta ligand topologies**	Vina writes PDBQT: merged hydrogens, no bond orders. Rebuild each pose as a full-hydrogen molecule with its own topology, so atom names always correspond.	Meeko, RDKit, rdkit-to-params	A .params file per docked pose
6 **Induced-fit relax**	Vina held the receptor rigid. Relax again with the ligand present, against its own apo reference — plus an apo' null control.	PyRosetta FastRelax, ligand present	Holo (bound) structures, and how far each site moved
7 **Dual occupancy**	Not a second docking run. Paste both poses into the step-2 structure and relax once with both present.	File merge + one PyRosetta relax	The both-substrates-bound structure
8 **Cross-scoring**	Does occupying one site change the other's binding energy? Rescore the same pose in four receptor conformations, against each replicate's own apo baseline.	Vina --score_only	The interference effect, in kcal/mol
9 **Analysis and figures**	Outputs defined before the compute starts, so every stage has a stated purpose.	numpy, matplotlib, PyMOL	Figures and statistics

How the steps depend on each other

Main chain 0 → 1 → 2 → 4 → 4b → 5 → 6 → 8 → 9; step 3 feeds the docking at step 4 (See Figure 1).

Step 7 is the exception — it needs only steps 2 and 4, so the both-bound structure can be built as soon as docking finishes.

Every stage records its parameters, seeds, and which residue set produced which number.



Figure 1. The ten stages of the docking pipeline; *Each box gives the stage number, what it does and the software used. The main sequence runs from structure prediction through to analysis; two preparation stages feed into it from the side. Stage 0 defines the active sites by measuring them from substrate-bound crystal structures, independently of any model built later, and stage 3 builds and verifies the two substrates from their PubChem records. Both are complete before the work they feed into begins.*