# Flexible Docking of 3-Phenylpropylamine into Trypsin - an Affinity Validation Study

**Industry Sector**

**Pharmaceuticals**

**Organization**

**Accelrys**

**Key Products**

**Insight11**
**Affinity**
**Discover-3**
**Binding Site Analysis**
**Ludi**
**DeCipher**

Dr Teresa Lyons, Accelrys, has used Affinity and supporting tools within InsightII to reproduce the crystal structure of trypsin in complex with 3-phenylpropylamine (PDB ID 1TNK).

This complex is one of several being used to validate Affinity as a docking application and to develop a protocol to successfully differentiate the correct complex structure from the set of structures Affinity produces.

See also:

- Flexible Docking of Benzylsuccinate Inhibitor into Carboxypeptidase A - an Affinity Validation Study (www.accelrys.com/cases/affinity/affinity2.pdf)
- Flexible Docking of Biotin into Streptavidin - an Affinity Validation Study (www.accelrys.com/cases/affinity/affinity3.pdf)

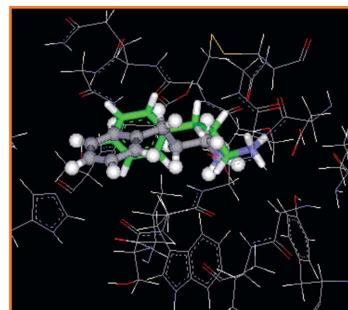
The computational challenge of docking ligands into receptors involves two distinct steps:

- Finding the correct conformation and orientation of the ligand in the binding pocket
- Correctly identifying that particular ligand pose from a set of energetically reasonable conformations and orientations.

This is widely known as the docking and scoring problem.

Affinity distinguishes itself from other docking programs in that it allows both the ligand and the binding pocket to be flexible during the calculation. Affinity starts by rotating, translating, and doing a conformational search of the ligand in the context of the binding pocket using a Monte Carlo algorithm. This is optionally followed by a Discover-based minimization or a simulated annealing phase that allows both the ligand and the binding pocket to optimize the interactions. An Affinity run results in a set of structures that fall within the energetic and geometric criteria set by the user. Assuming that conformational space has been searched adequately during the Affinity run, the challenge then becomes how to effectively pick out 'the real structure' from the result set.

This particular complex turned out to be difficult because the residues lining the binding pocket moved quite a lot during the calculations, trapping the ligand into conformations that were quite far from the original crystal structure. Remember that the goal of this project as a whole was to reproduce something close to the crystal structure conformation and location of a ligand in a protein and see what tools were needed to evaluate which structure was the closest to the original crystal structure. The first step, reproducing the crystal structure of the complex (within a couple angstroms RMS), was not trivial, as detailed below.



▲ *Figure 1: The top conformer (frame 1, in gray) overlaid with the original minimized crystal structure ligand (in green). The heavy atom RMS of the two ligands is 2.39 Angstroms.*

## Methods

- Preparing the protein: After the hydrogen atoms were added to the protein and the ligand, a heavy-atom-fixed minimization was performed to give the hydrogen atoms favorable positions. The 3-phenylpropylamine had a lot of bond and angle strain on it as modeled by the crystallographer so that it couldn't possibly be faithfully reproduced using molecular mechanics. Consequently, the hydrogen-only minimization was followed with a minimization of just the ligand in the binding pocket - this resulted in a structure of the ligand that was unstrained and 1.22 angstroms RMSD for the heavy atoms from the crystal structure position. This was the structure used for defining the binding pocket and for RMSD comparison of the Affinity-generated structures. The ligand was reoriented in the binding pocket for the start of the Affinity calculations to avoid bias towards the crystal structure position.
- Defining the binding pocket: In order to be as unbiased as possible, the binding pocket was defined based on an 'Active Site Search', a cavity-finding algorithm that is part of the Binding Site Analysis module of InsightII. Whole residues within 3 Angstroms of the cavity surface were collected to make up the binding site subset. For trypsin, the largest cavity found coincided well with the binding pocket where the inhibitor sits in the crystal structure. The binding site subset consisted of 17 residues and 3 crystal waters.
- Setting up the Affinity job: The Affinity job was set up with parameters similar to the Affinity pilot tutorial for HIV protease. No hydrogen bonding or tethered subsets were defined. The ligand was confined to stay within 5 Angstroms radius of its starting position (center of mass confinement). The docking calculation was run in two steps:
  - Monte Carlo search: 50 structures with no acceptance filter. Non-bonded method: Quartic VDW with Coulombic interactions off and VDW scaled down to 0.1. Flexible ligand with an energy range of 200 and Tolerance of 1e+6 and 100 steps minimization. All other parameters were their default value.

- Simulated annealing: The MC structures were put through a separate simulated annealing search using Cell Multipole non-bonded interactions with no distance-dependent dielectric. The VDW and Coulombic scales were originally 0.1 and their final values were brought up to 1. Avoiding a final minimization step, the initial and final temperatures of the sim lated annealing step were 480 and 280 instead of the default 500 and 300 degrees.
- Analysis of the structures: The resulting set of structures was evaluated for how close each was to the crystal structure 1TNK based on a heavy-atom RMSD of the ligand only (a function defined in Decipher). Based on this criteria, the structure most similar to the crystal structure was frame 41 of the trajectory file. All frames were analyzed using four methods:
  - Total energy as output by Affinity
  - Interaction energy between the ligand and protein as calculated with the Evaluate/Intermolecular command (Docking module)
  - Ludi Score 3 of the interaction (Ludi module)
  - Number of hydrogen bonds between protein and ligand.

A BCL script facilitated the batch analysis of the trajectory file.

**Results**

Initially, the binding pocket subset of residues was much larger (a 4 Angstrom radius was used) and the confinement radius for the ligand was 8 Angstroms. However, this trypsin-inhibitor complex showed a significant amount of motion during the Affinity run, with a small loop (residues 215-218) essentially swapping positions with the ligand and the tryptophan side-chain of 215 swinging into a completely opposite orientation relative to the backbone.

An Affinity run of 100 structures yielded nothing close to the crystal structure. So the binding pocket was redefined to include only residues within 3 Angstroms of the cavity surface and the center of mass of the ligand was confined to within 5 Angstroms of the starting position. In an uninhibited form of trypsin (5PTP), the binding site residues are within 0.2 Angstroms of those in the inhibited 1TNK complex, so restricting the motion of the protein by including fewer residues in the flexible subset is reasonable for this system.

The RMS analysis of the structures showed that none are quite as close to the crystal structure as hoped, with the most accurate structure being 2.39 Angstroms RMSD from the crystal structure. This may be due to the binding pocket residues still shifting too much. The fact that the binding pocket residues of the apoprotein (5PTP) are essentially in the same positions as with inhibited trypsin suggests that the binding pocket is rather rigid. More accurate complexes may be calculated by fixing the backbone atoms of the binding pocket residues.

| Frame | RMSD | InterE | TotalE | Ludi3 | H-bonds |
| --- | --- | --- | --- | --- | --- |
| 1 | 2.39 | -70.07 | 314.05 | 290 | 2 |
| 2 | 4.30 | -34.50 | 344.90 | 240 | 0 |
| 3 | 2.76 | -74.73 | 312.73 | 309 | 1 |
| 4 | *3.01* | *-59.30* | *305.33* | *263* | 1 |
| 5 | 4.99 | -31.04 | 334.69 | 166 | 0 |
| 6 | 7.64 | -30.63 | 345.41 | 76 | 1 |
| 7 | 9.51 | -27.12 | 356.24 | 146 | 1 |
| 8 | 3.74 | -55.67 | 321.70 | 171 | 0 |
| 9 | 8.56 | -22.83 | 329.39 | 146 | 0 |
| 10 | 6.00 | -30.90 | 352.09 | 98 | 1 |
| 11 | 7.58 | -35.27 | 352.47 | 150 | 1 |
| 12 | 6.85 | -40.70 | 356.20 | 143 | 0 |
| 13 | 6.20 | -39.22 | 340.24 | 146 | 1 |
| 14 | 9.08 | -33.22 | 343.26 | 83 | 0 |
| 15 | 6.86 | -30.39 | 357.99 | 119 | 0 |
| 16 | 9.17 | -27.18 | 365.60 | 131 | 0 |
| 17 | 5.92 | -28.37 | 338.91 | 137 | 0 |
| 18 | 9.01 | -19.65 | 351.71 | 141 | 0 |
| 19 | 8.32 | -43.13 | 367.33 | 159 | 2 |
| 20 | 6.38 | -18.55 | 368.96 | 93 | 0 |
| 21 | 5.73 | -32.15 | 345.95 | 87 | 0 |
| 22 | 4.07 | -43.34 | 349.57 | 244 | 0 |
| 23 | 4.47 | -46.05 | 347.16 | 187 | 0 |
| 24 | 5.08 | -53.79 | 338.36 | 205 | 0 |
| 25 | *3.40* | *-80.21* | *340.21* | *268* | *2* |
| 26 | 4.93 | -59.32 | 345.44 | 278 | 1 |
| 27 | 5.22 | -61.47 | 350.48 | 211 | 1 |
| 28 | 4.92 | -65.04 | 359.49 | 275 | 0 |
| 29 | 4.75 | -55.41 | 355.69 | 294 | 0 |
| 30 | 3.03 | -74.64 | 358.53 | 276 | 2 |
| 31 | 7.11 | -23.97 | 363.24 | 145 | 0 |
| 32 | 6.17 | -22.75 | 368.86 | 200 | 0 |
| 33 | 6.62 | -49.19 | 365.41 | 149 | 1 |
| 34 | 7.00 | -43.29 | 356.58 | 235 | 1 |
| 35 | 6.77 | -29.80 | 352.49 | 163 | 0 |
| 36 | 7.92 | -24.28 | 361.01 | 137 | 0 |
| 37 | 8.61 | -49.97 | 360.32 | 162 | 0 |
| 38 | 6.44 | -28.22 | 369.00 | 154 | 1 |
| 39 | 7.99 | -23.35 | 379.76 | 72 | 0 |
| 40 | 7.33 | -36.91 | 374.53 | 159 | 1 |
| 41 | 7.20 | -35.18 | 358.63 | 139 | 0 |
| 42 | 6.49 | -34.10 | 368.98 | 128 | 1 |
| 43 | 7.19 | -26.83 | 358.33 | 86 | 0 |
| 44 | 7.08 | -42.08 | 364.16 | 133 | 2 |
| 45 | 7.50 | -57.02 | 346.96 | 166 | 0 |
| 46 | 6.47 | -36.70 | 347.64 | 159 | 1 |
| 47 | 7.52 | -23.40 | 335.55 | 61 | 1 |
| 48 | 5.81 | -43.98 | 349.89 | 176 | 1 |
| 49 | 8.10 | -29.24 | 369.84 | 133 | 0 |
| 50 | 7.46 | -51.60 | 382.37 | 185 | 0 |

*Table 1: The results of the RMS analysis. Dark yellow shows the best 10% (5 frames) and light yellow shows the next 10% (frames ranked 6-10) for each of the scoring methods. The two frames closest to the 1TNK structure, by RMSD, are also highlighted in yellow.*

The best frames are frames 1 (2.39), 3 (2.76), 4 (3.01) and 30 (3.03) as evaluated by heavy atom RMSD. Ignoring the number of hydrogen bonds, the results shows that 4 different frames lie in the top 20% for all evaluation methods: 25, 3, 1, and 4. Of these, frame 1 and 25 also form two intermolecular hydrogen bonds (the maximum number from these structures). Because these two structures (frames 1 and 25) are fairly different from each other, the only way to resolve which one is the closest to the real structure (in the absence of knowledge of the real structure) is to look at which one falls into the top 10% in each category or to average the ranking (1 to 10) of each and see which one has a lower average. Although frame 1 does not rank number 1 in any one category, it does fall into the top 10% of all 'scores' whereas frame 25 does not.

**Conclusion**

This study showed that the structure of the trypsin-3-phenylpropylamine complex can be reproduced using Affinity and the most accurate structure can be distinguished from the result set using a consensus-type scoring method, taking advantage of the tools available within the InsightII software suite.