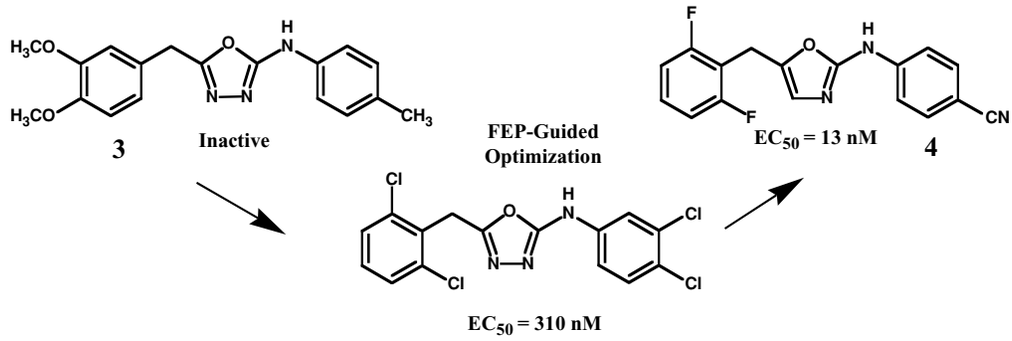


**Figure 1.4.** Progression of a false positive from docking to potent anti-HIV agents.

hydrogen-bonding groups, for example, the OH of tyrosine or serine and ammonium group of lysine. The current scoring function has been trained to reproduce experimental activity data for more than 300 complexes of HIV-RT, COX-2, FK506 binding protein, and p38 kinase.[5] It yields a correlation coefficient $r^2$ of 0.58 for the computed versus observed log(activities). The scoring function contains only five descriptors that were obtained by linear regression, including an estimate of the analog's octanol/water partition coefficient from QikProp (QPlog*P*),[15] the amount of hydrophobic surface area for the protein that is buried on complex formation, and an index recording mismatched protein/analog contacts, such as a hydroxyl group in contact with a methyl group. Interestingly, the most significant descriptor is QPlog*P*, which alone yields a fit with an $r^2$ of 0.47. Thus, the adage that increased hydrophobicity leads to increased binding is well supported, though it requires refinement for quality of fit using the host/ligand interaction energy or an index of mismatched contacts. Overdone, it also leads to ADME problems, especially poor aqueous solubility and high serum protein binding.

The results from a BOMB run include the structure for each protein/analog complex as a Protein Data Bank (PDB) file or BOSS/MCPRO Z-matrix (internal coordinate representation)[16] and a spreadsheet with one row for each analog summarizing computed quantities from the BOMB calculations, including host–analog energy components and surface area changes as well as predicted properties for the analog, including log $P_{o/w}$, aqueous solubility, and Caco-2 cell permeability from QikProp, which is called as a subroutine. The processing time for Het-NH-Ph-U using ammonia as the core is approximately 15 s per analog on a 3-GHz Pentium IV processor. The required time increases roughly linearly with the number of conformers that need to be constructed. For large libraries, multiple processors are used.

## Virtual screening

The common alternative is to perform virtual screening on available compound collections using docking software. Many reviews and comparisons for alternative software and scoring functions are available.[6,17–20] There is no question that there have been many successes with docking such that, given a target structure, it is expected to be competitive with and far more cost effective than HTS and is now an important component of lead discovery programs in the pharmaceutical industry. New success stories are reported regularly in the literature and at conferences. However, it is generally accepted that correct rank-ordering of compounds for activity is beyond the current capabilities. This is not surprising in view of the thermodynamic complexity of host/ligand binding, including potential structural changes for the host on binding, which have usually been ignored, and the need for careful consideration of changes in conformational free energetics between the bound and unbound states.[21]

In our experience, docking has been a valuable complement to de novo design (Figure 1.1). When large compound collections are docked, interesting structural motifs often emerge as potential cores that may have been overlooked otherwise. Our earliest docking effort started out well, was formally a failure, and then recovered to provide an interesting lead series that yielded potent anti-HIV agents.[5,22] Leads were sought by processing a collection of approximately 70,000 compounds from the Maybridge catalog, which was supplemented with twenty known NNRTIs. The screening protocol began with a similarity filter that retrieves 60% of the known actives in the top 5% of the screened library. The approximately 2,000 library compounds that were most similar to the known actives were then docked into the 1rt4 structure of wild-type HIV-RT, using Glide 3.5 with standard precision.[6] The top 500 compounds were then redocked and scored in Glide extraprecision (XP) mode.[23] The top 100 of these were postscored with a molecular mechanics/generalized Born/surface area (MM-GB/SA) method that was shown to provide high correlation between predicted and observed activities for NNRTIs.[22] Though known NNRTIs were retrieved well (ten were ranked in the top twenty), purchase and assaying of approximately twenty high-scoring compounds from the library failed to yield any active anti-HIV agents. Persisting, the highest-ranked library compound, the inactive oxadiazoles **3** in Figure 1.4, was pursued computationally to seek