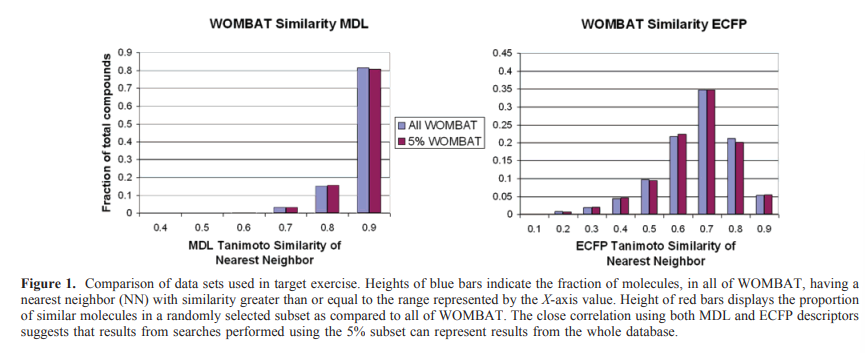
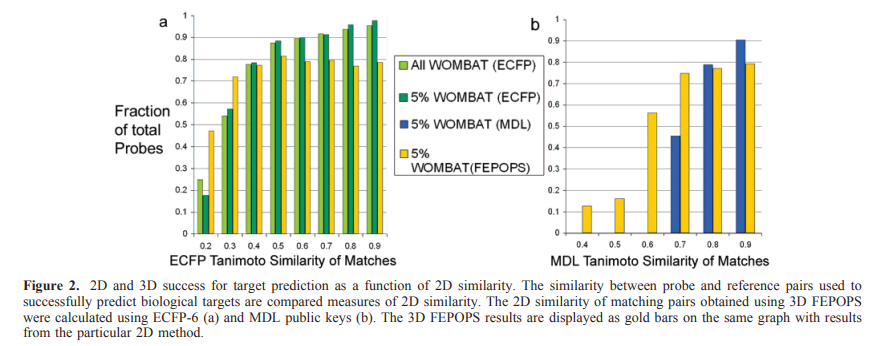
# Bridging Chemical and Biological Space: “Target Fishing” Using 2D and 3D Molecular Descriptors

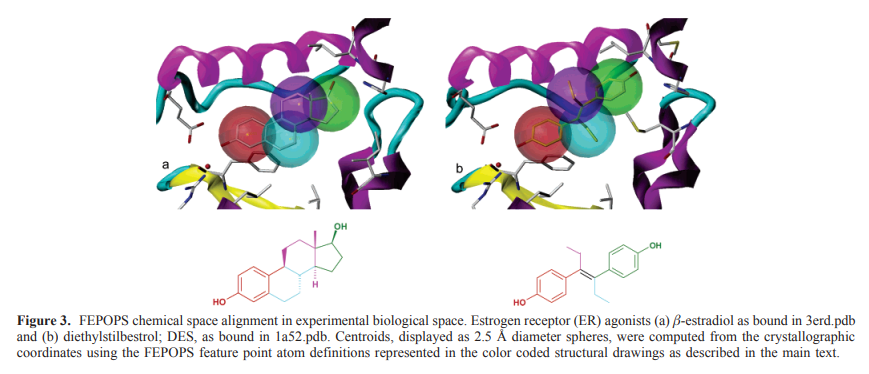
* The 2D methods employed outperform the 3D (88% vs 67% success) in correct target prediction.
* The 3D method (FEPOPS) shows promise for providing pharmacophoric alignment of the small molecules’ chemical features consistent with those seen in experiment ligand/receptor complexes.
* Current methods of cellular screening and pharmacogenetic profiling can rapidly reveal phenotypic responses to drugs but do not immediately pinpoint their molecular target.
* The researchers’ computational study is designed toward a goal: that is identifying the molecular target for a single chemical entity, or “target fishing”, based on similarity of a new compound to structures where activities against a broad panel of target is already known.
* The effective use of databases depends on having reliable methods of relating the chemical structure of the query compound to the reference compounds in the database.
* The 2D methods are very fast even the “problem of exploring the conformational space is ignored. It is expected that such methods would be useful for clustering similar compounds, or selecting diverse subsets from large libraries, but these methods are also effective for “virtual screening” of actives from large compound sets using the structures of small number of known actives as probes.
* The 3D methods would include greater information since the binding between a ligand and receptor is a 3D event.
* The 2D methods such as Daylight or Unity fingerprints were initially designed to find “more of the same” while 3D methods can enrich scaffold diversity.
* The goals of the research paper are: - to examine methods for predicting a chemical probe’s biological target based upon similarity to a reference compound. The second goal is to examine the ligand-based 3D alignments to actual 3D receptor binding.
* One hypothesis, supported by the results, is that 2D methods are favored in case of close analogues, but that 3D methods often advantages below a certain similarity threshold.
* Methods: - (A) Database, Bioactivity Data, and Database Preparation. – it has over 100,000 unique structures described as SMILES and over 240,000 biological activities in separate ISIS databases. The final groomed database contained 47 505 unique chemical structures associated with 544 biological targets.
* (B) 2D Target Fishing: MDL Structural Keys and ECFP\_6 Fingerprints: - All WOMBAT: the 2D similarity protocols, using each descriptor type, were run against the groomed database using the complete set of 47,505 chemical structures as probes.
* 5% WOMBAT: A smaller probe set of 2351 molecules reflecting a randomly chosen 5% of the entire chemical library was generated using the Random Percent Filter in Pipeline Pilot. Second runs against the full database were performed with the 2351 molecules set using both 2D descriptor types.
* 2D descriptors for all compounds were computed using both MDL public keys and SciTegic’s Extended Connectivity Fingerprints (ECFP\_6) in Pipeline Pilot.
* 3D Target Fishing: FEPOPS. All 47 505 unique compounds were input as SMILES strings. FEature POint PharmacorphoreS (FEPOPS) were calculated and stored as text in a 3D descriptor database (3DDD) yielding 815 676 records with compound IDs and associated feature point information.
* All WOMBAT: 3D analysis was not run on the full data due to computational expense.
* 5% WOMBAT: the same 2351 chemical structures, used for the 2D comparison. It is expanded with FEPOPS conformer/tautomers, and compared to the residual set of 774, 824 descriptors.
* The FEPOPS descriptors were computed and using this method, compounds are preprocessed to generate 3D structures, assign protonation states, enumerate tautomers, and calculate partial charges and atomic log *P* values.
* Ligand atoms are partitioned into four *k*-mean clusters based upon their spatial coordinates.
* Centroids are defined from the atoms of each cluster. Partial charges, log *P*, and hydrogen bond donors and acceptors of the atoms belonging to each cluster are summed and encoded into the centroids to create “feature points”.
* *K*-medoids clustering of feature points is performed to find a smaller number of representative conformers.
* (D) Fishing for Chemical Diversity with 3D Descriptors: - I Nearest Neighbors Misses (NNmiss) - first we want to access 3D performance in cases where 2D fails. Second, the smaller number of probes reduced computing times and allowed for multiple conditions to be examined. Third, using a randomly selected subset of the total nearest neighbor misses as probes left the remainder of the 2D misses in the 3DDD.
* Because some target classes have representatives with low 2D similarity in the parent database, this criterion was important for reasonable evaluation of 3D performance.
* II Similarity Filters – it explores the added value of 3D FEPOPS method for finding correct probe/reference pairings with low structure similarity. It runs multiple times against the 3DDD with different filters.
* III Probe for related chemistry/biology – ATP WOMBAT. A prospective exercise for “target fishing” would be to use a single molecule with observed activity as a probe to identify possible binding partners.
* (E) Modeling of 3D Ligand Alignments in Biological Space. the goal of this exercise was to examine the binding mode for probe/reference drug pairs having dissimilar chemical scaffolds, but similar biological activity.
* II. Mapping of FEPOPS Centroids: - Atom IDs associated with features points used for alignments were output from FEPOPS as mol2 files.
* Modeling of Estrogen Receptor Ligands. Crystallographic coordinates of the estrogen receptor (ER) complexed with both the molecular probe 17--estradiol, la52.pdb, and the reference compound diethylstilbestrol, 3erd.pdb, were downloaded from the PDB.
* Modeling of Retinoid Receptors Ligands. A 3D model of the molecular probe Targretin was generated by modifying the cyclopropyl moiety of compound LG268 as found in complex with the receptor in 1h9u.pdb to the alkene of **5**.
* FEPOPS centroids were mapped to each conformer, and each was aligned to the centroids of **5** and evaluated as described in Results and Discussion.
* Results and Discussion: - 2D and 3D Target Fishing. The 2D Nearest Neighbor (2DNN) analysis was performed using MDL and ECFP descriptor keys as described in Methods and was run on both the full and 5% data sets.



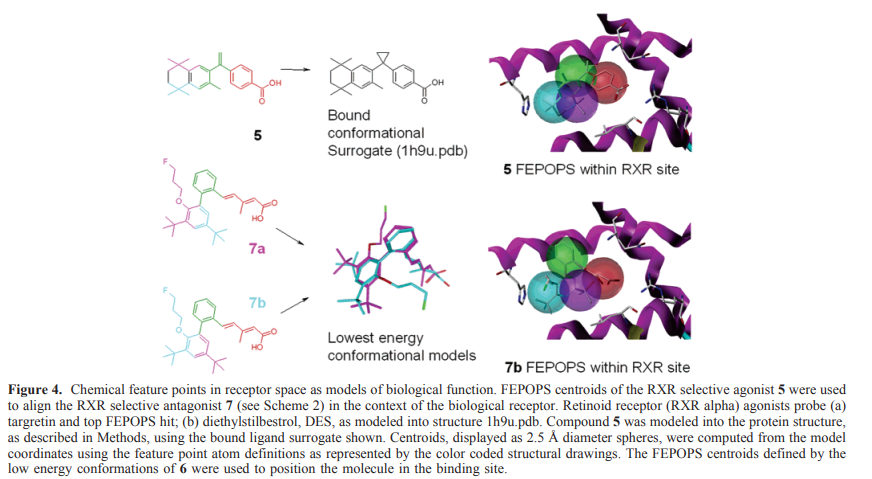
* From the figure we can see that the MDL public key method was one of the first and is still a widely used tool in cheminformatics. The “public” keys describe a chemical entity based upon the presence or absence of 166 substructural fragments from a predefined library.
* It is important to note that the lower Tanimoto similarity values for ECFP relative to MDL is a function of different scales.
* The 3D FEPOPS analysis was performed using the same 5% probe set as was used for the 2D analysis.
* The FEPOPS finds approximately 75% correct matches across the most populated chemistry space, 0.4-1.0 ECFP.
* For the NNmiss set with only a “self” filter, 3D resulted in 32.1% success compared to 0% for 2D.
* The researchers find that the results of filtering above the 0.85 level using either 2D descriptor did not return appreciably better structural diversity.



* The full results set with structural drawings and 2D Tanimoto indices for the 3D similar pairs in the 0.80 and 0.60 analysis are shown in the Supporting Information (SI).
* Pharmacophore and 3D-QSAR-based methods have also shown promise for both identification of potential nontarget interactions and scaffold hopping within a given target but have a dependence upon the external knowledge used to train the models.
* FEPOPS are an inherently fuzzy description of a molecules potential shape(s) and chemical space.
* The researchers hypothesized that the maximally overlapped feature points derived from exploring the chemistry space of molecules that effect the same target should correlate with the biological space of the actual target.
* Estrogen Receptor: - endogenous estrogens such as estradiol (E2) exert their physiological effects by binding to estrogen receptors (ER), inducing nuclear translocation, and increasing transcription.
* The synthetic nonsteroidal compound diethylstilbestrol (DES) found by their 3D target fishing protocol also binds to ER with high affinity and similarly increases transcriptional events.
* The 2D Tanimoto similarities between these molecules are 0.42-MDL and 0.13-ECFP while the 3D Pearson correlation of the feature points is 0.91.
* The 3D-FEPOPS alignments of these molecules with Pearson coefficients of 0.89 and 0.83 reverses the orientation of the fused ring system 180 along the axis between the diols, suggesting that these molecules would not have the same binding mode.
* The large effect of small changes at the 2-position upon target specificity underscores the 3D nature of the binding event and an inherent limitation of 2D similarity descriptions alone.
* RXR Receptor: - the retinoid receptors, characterized by the subfamilies of retinoic acid receptors (RAR) and retinoid X receptors (RXR), serve as obligate binding partners in many cells signaling pathways affecting, cell differentiation, proliferation, and tissue homeostasis.
* FEPOPS correctly assigned RXR binding to **5** through identification with compound **7**, also known to be RXR selective, as the closest 3D match.
* The Tanimoto scores of 2D similarity are very low, the Pearson coefficient of distances between the aligned FEPOPS of **5** and **7** is > 0.92.



* The 3D method appears to be locating some significant chemical/biological relationships that are not obvious from the underlying 2D structure.



* In the above it is illustrated that the FEPOPS defined overlap in the context of the RXR receptor.
* In this the FP one, two (red and green) and three, four (blue and violet) align in a way consistent with the bent cis geometry of its endogenous ligand.
* The probe **5** was modeled using the complex of structurally similar LG268 (8) with RXR, however, the unusual scaffold of **7** was not available in the PDB.
* Without an external reference, the FEPOPS results would best suggest the alignment of FP 1 and 2 but allow two options for FP3 and 4. If the external knowledge gain from the ER systems, the researchers can present a hypothesis.
* Compound **5** is a selective agonist for RXR while **7** is an antagonist, a situation analogous to DES, **2**, and OHT, **4**.
* The modeled binding mode of **7b** illustrated the 3-fluoroproply moiety out of the binding groove in the same receptor space relative to **5** as the side chain of **4** is to **2** in experimental solutions of the ER complexes.
* The tissue specific agonism/antagonism of various drugs is likely due to nonconserved protein side chain residues, outside the binding groove, that alter the position of the antagonizing feature relative to the ligand’s core.
* PKA Receptor-ATP/Balanol: - the researchers cast a larger net using ATP as a probe and pulled back the top 30 targets predicted by FEPOPS.
* The analysis of scaffold diversity associated with correct matches revealed that three of the pairings PKA, PKC-beta-1, and PKC-eta, were a variation of the same scaffold containing no phosphates, adenine, or ribose moieties, balanol.
* The negatively charged atoms of ATP’s FP1 (red) accepts hydrogen bonds from donating side chains in the bottom of the pocket, while balanols’ FP1 atoms accept hydrogen bonds from backbone amides of the flexible glycine rich loop above the ligands.
* For searching flexible chemical space for 3-dimensional alignments, the method has found nonobvious ways of bridging the biological space of the target molecule.
* The “fuzzy” nature of the feature point alignment is a strength that allows it to accurately describe the general receptor environment of even flexible targets such as kinases.
* Conclusions: - the 2D chemical similarity to a reference structure is a very good predictor of similar biological targets; however, there are limits to that effectiveness.
* The ECFP-6 descriptors sharply decline in predictive efficacy for neighbors with Taninoto similarity indices less than 0.40 while MDL public keys decline at indices <0.80.
* The 3D method outperforms the 2D method in the below thresholds, escaping clear correlation of performance to the underlying similarity of the chemical graph.
* The 3D method also demonstrates an ability to align chemical features in space similarly to those found in biological complexes.
* A caution is raised by the RXR example that the whole molecule overlap of chemical space is not necessarily the best descriptor of biological function.
* The analysis points out that the high performance was in a large part due to the congeneric nature of the data used.
* The FEPOPS 3D method was able to correctly predict targets for a subset of those failing 2D comparison, but the overall lower fidelity indicates that its use should be reserved for second phase analysis in the case of very low neighborhood similarity.