Supplementary Information

How well do the substrates KISS the enzyme? Molecular docking program selection for feruloyl esterases

D.B.R.K. Gupta Udatha, Nobuyoshi Sugaya, Lisbeth Olsson, Gianni Panagiotou

Contents

Section A

Supplementary Table S1. Experimental substrate specificity spectra available for three feruloyl esterase enzymes[1-3]. (a) Kinetic data of Feruloyl esterase A from *Asperigillus niger* (AnFAEA). (b) Kinetic data of Feruloyl esterase B from *Asperigillus niger* (AnFAEB). (c) Kinetic data of Feruloyl esterase C from *Talaromyces stipitatus* (TsFAEC). K_m is expressed as mM and k_{cat} as kat/mol enzyme.

(a)

	Substrate	K_m	k_{cat}	k_{cat} / K_m
1	Methyl cinnamate	Inactive	Inactive	Inactive
2	Methyl 2-hydroxy cinnamate	Inactive	Inactive	Inactive
3	Methyl 3-hydroxy cinnamate	Inactive	Inactive	Inactive
4	Methyl 4-hydroxy cinnamate (Methyl p-coumarate)	Inactive	Inactive	Inactive
5	Methyl 3,4-dihydroxy cinnamate (Methyl caffeate)	Inactive	Inactive	Inactive
6	Methyl 2-methoxy cinnamate	Inactive	Inactive	Inactive
7	Methyl 3-methoxy cinnamate	1.99	12	6
8	Methyl 4-methoxy cinnamate	Inactive	Inactive	Inactive
9	Methyl 3,4-dimethoxy cinnamate	1.36	74	54
10	Methyl 3,5-dimethoxy cinnamate	0.92	15	16
11	Methyl 3,4,5-trimethoxy cinnamate	1.63	1063	652
12	Methyl 4-hydroxy-3-methoxy cinnamate (Methyl ferulate)	0.72	105	146
13	Methyl 3-hydroxy-4-methoxy cinnamate	Inactive	Inactive	Inactive
14	Methyl 4-hydroxy-3,5-dimethoxy cinnamate (Methyl sinapate)	0.45	172	382
15	Methyl 4-hydroxy-3-methoxy phenyl propionate	2.08	738	355

(b)

	Substrate	K_m	k_{cat}	k_{cat} / K_m
1	Methyl cinnamate	0.79	267	336
2	Methyl 2-hydroxy cinnamate	Inactive	Inactive	Inactive
3	Methyl 3-hydroxy cinnamate	0.55	75	138
4	Methyl 4-hydroxy cinnamate (Methyl p-coumarate)	0.014	263	18764
5	Methyl 3,4-dihydroxy cinnamate (Methyl caffeate)	0.22	411	1867
6	Methyl 2-methoxy cinnamate	0.72	41	57
7	Methyl 3-methoxy cinnamate	Inactive	Inactive	Inactive
8	Methyl 4-methoxy cinnamate	0.31	482	1555
9	Methyl 3,4-dimethoxy cinnamate	Inactive	Inactive	Inactive
10	Methyl 3,5-dimethoxy cinnamate	Inactive	Inactive	Inactive
11	Methyl 3,4,5-trimethoxy cinnamate	Inactive	Inactive	Inactive
12	Methyl 4-hydroxy-3-methoxy cinnamate (Methyl ferulate)	1.32	60	46
13	Methyl 3-hydroxy-4-methoxy cinnamate	0.85	316	372
14	Methyl 4-hydroxy-3,5-dimethoxy cinnamate (Methyl sinapate)	Inactive	Inactive	Inactive
15	Methyl 4-hydroxy-3-methoxy phenyl propionate	3.17	1410	445

(c)

	Substrate	K_m	k_{cat}	k_{cat} / K_m
1	Methyl cinnamate	0.118	1216	10351
2	Methyl 2-hydroxy cinnamate	0.015	130	8604
3	Methyl 3-hydroxy cinnamate	0.025	1444	57540
4	Methyl 4-hydroxy cinnamate (Methyl p-coumarate)	0.01	573	57300
5	Methyl 3,4-dihydroxy cinnamate (Methyl caffeate)	0.005	n.d	n.d
6	Methyl 2-methoxy cinnamate	Inactive	Inactive	Inactive
7	Methyl 3-methoxy cinnamate	0.301	1637	5442
8	Methyl 4-methoxy cinnamate	Inactive	Inactive	Inactive
9	Methyl 3,4-dimethoxy cinnamate	0.626	15	24
10	Methyl 3,5-dimethoxy cinnamate	0.075	5	71
11	Methyl 3,4,5-trimethoxy cinnamate	Inactive	Inactive	Inactive
12	Methyl 4-hydroxy-3-methoxy cinnamate (Methyl ferulate)	0.04	579	14475
13	Methyl 3-hydroxy-4-methoxy cinnamate	0.533	42	78
14	Methyl 4-hydroxy-3,5-dimethoxy cinnamate (Methyl sinapate)	0.37	556	1502
15	Methyl 4-hydroxy-3-methoxy phenyl propionate	0.016	7554	481128

Section B

***Ab-initio* structure predictions of AnFAEB and TsFAEC enzymes**

Availability of three-dimensional (3D) structures of both receptor and substrate is a prerequisite to perform molecular docking studies. In the absence of experimentally determined structure, modeling of the protein using structural bioinformatics tools provides the 3D models. Although the homology modeling approach provides reliable models, it can be applied only if the 3D structure of a similar sequence is already known (which is not the case for AnFAEB and TsFAEC). The LOMETS threading method that has been proposed[4,5] provides a platform, in which, given an amino acid sequence and a set of structures/structural patterns, a structure will be computed into which the sequence is most likely to fold. In our study, we used *ab-initio multiple threading alignment* approach[6,7] based on I-TASSER predictor that makes an alignment computation between the amino acids of the sequence and spatial positions of the 3D structure using scoring functions followed by LOMETS threading. The top five model structures were derived from I-TASSER simulations each having a *C*-score. *C*-score is a confidence score for estimating the quality of predicted models and it is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations; a high *C*-score signifies a model with a high confidence and vice-versa. Additional information of the high *C*-score I-TASSER models were given in the excel file as Supplementary Information.

Structure refinement and Validation: The coordinates of the model structures with high *C*-scores were submitted to Protein Model DataBase (PMBD)[8]; the solvent surface rendering (probe radius of 1.4) of the modeled structures colored according to the secondary structure elements and respective PMDB accession codes is given in Figure S1. Since the structure predictions were done using *ab-initio* approaches, the coordinates of the model structures with high *C*-Scores were further chosen for refinement followed by structure validation. Structure refinement of modeled structures was carried out using Discovery Studio software suite version 3.0 (Accelrys Inc, USA). The Prepare Protein protocol package in Discover studio suite was used for inserting missing atoms in incomplete residues, modeling missing loop regions[9], deleting alternate conformations (disorder), standardizing atom names, and protonating titratable residues using predicted pKs[10]. The Side-Chain Refinement protocol was used for each structure to optimize the protein side-chain conformation based on systematic searching of side-chain conformation and CHARMm Polar H energy minimization[11] using ChiRotor algorithm[12]. Energy minimization is performed on structures prior to dynamics to relax the conformation and remove steric overlap that produces bad contacts. The Minimization protocol minimizes the energy of a structure through geometry optimization and performs an energy minimization on a molecule using CHARMm[13]. Smart Minimizer algorithm was used for the minimization process which performs 1000 steps of Steepest Descent with a RMS gradient tolerance of 3, followed by Conjugate Gradient minimization for faster convergence towards a local minimum[14]. Structure evaluations were done using DOPE, which is an atomic based statistical potential in MODELER package for model evaluation and structure prediction[15]. Structure verifications were carried out using VerifyProtein-Profiles-3D that allows evaluating the fitness of a protein sequence in its current 3D environment[16]. Finally, the CHARMm based Energy minimization was done to remove steric overlaps that produces bad contacts; the initial potential energies of starting structures and the potential energy of respective minimized structures were given in Table S2. The verification scores viz., DOPE score, DOPE-HR score, Verify score and the potential energies of the modeled structures given in Table S2 reveals high quality of the models.

Supplementary Table S2. Structure evaluation scores for modeled FAEs

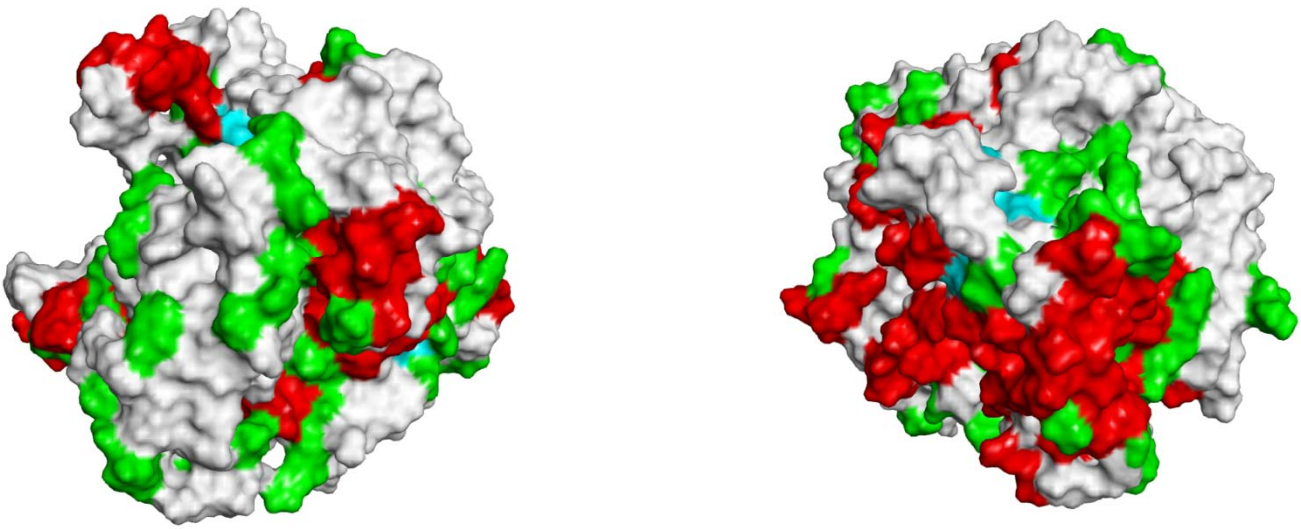
Protein	DOPE Score [a]	DOPE-HR Score	Verify Score [b]	Verify Expected High Score	Verify Expected Low Score	Initial Potential Energy (kcal/mol)	Potential Energy (kcal/mol)
AnFAEB	-46067	-28700	192	230	104	-9820	-28991
TsFAEC	-46784	-30979	194	234	105	-17562	-28884

[a] DOPE is an atomic based statistical potential in MODELER package for model evaluation and structure prediction[15]. The DOPE score of a protein can be viewed as a conformational energy which measures the relative stability of a conformation with respect to other conformations of the same protein. It can assist in choosing the best model out of a set of predicted model structures of a protein sequence. Verify Protein (MODELER) allows you to select the best structure from a collection of protein molecules with the same protein sequence. This protocol uses the MODELER DOPE (Discrete Optimized Protein Energy) method to calculate a DOPE score for each structure. A lower score indicates a statistically better model. DOPE scores are reported for each input protein structure, which can be used to compare different conformations of a protein. A lower score indicates smaller model errors which is the case in all the modeled FAEs in the present study. The DOPE score is not normalized on protein size, so DOPE scores of proteins with different sequences should not be compared directly.

[b] The Verify Protein protocol is generally employed in the final phase of a homology modeling project or for testing a preliminary protein structure based on experimental data. It depends on the principle that a protein's structure must be compatible with its own sequence. Verify Protein score (Profiles-3D) allows you to evaluate the fitness of a protein sequence in its current 3D environment[17]. The Verify scores for the modeled FAEs were shown in Table 2 along with the expected high and low Verify scores for proteins of similar size. The expected high scores are based on a statistical analysis of high-resolution structures in the Protein Data Bank. The expected low score is 45 percent of the high score and is typical of grossly misfolded structures having this sequence length. If the overall quality is lower than the expected low score, then the structure is almost certainly misfolded. Having Verify scores close to the High number is a positive indicator showing the quality of modeled FAE structures.

Supplementary Figure S1. Solvent surface rendered structures of the modeled FAEs with high *C*-Score colored according to the secondary structure (Red represent alpha helix regions; Blue represent beta strand regions; Green represent coil regions). (A) Structure of AnFAEB model submitted to PMDB with accession code PM0077390. (B) Structure of TsFAEC model submitted to PMDB with accession code PM0077391.

A) **B)**



Section C

i) Docking Algorithms

LibDock: The LibDock algorithm, an interface to the Discovery studio docking program was developed by Diller and his colleagues[18-20]. LibDock uses protein site features referred to as HotSpots. HotSpots consist of two types: polar and apolar. A polar Hotspot is preferred by a polar ligand atom (for example a hydrogen bond donor or acceptor) and an apolar HotSpot is preferred by an apolar atom (for example a carbon atom). The receptor HotSpot file is calculated prior to the docking procedure. The poses are pruned and a final optimization step is performed before the poses are scored. Ligand hydrogens, which are removed during the docking process, are added to the ligand poses.

C-DOCKER: CDOCKER algorithm is an implementation of a CHARMm based docking tool using a rigid receptor[21]. It is a grid-based molecular docking method and involves the following steps:

- A set of ligand conformations are generated using high-temperature molecular dynamics with different random seeds. This step can be skipped to dock the input conformation(s) only.
- Random orientations of the conformations are produced by translating the center of the ligand to a specified location within the receptor active site, and performing a series of random rotations. A softened energy is calculated and the orientation is kept if the energy is less than a specified threshold. This process continues until either the desired number of low-energy orientations is found, or the maximum number of bad orientations has been tried. This step can be skipped to use the input orientation only.
- Each orientation is subjected to simulated annealing molecular dynamics. The temperature is heated up to a high temperature then cooled to the target temperature.
- A final minimization of the ligand in the rigid receptor using non-softened potential is performed.
- For each final pose, the CHARMm energy (interaction energy plus ligand strain) and the interaction energy alone are calculated. The poses are sorted by CHARMm energy and the top scoring (most negative, thus favorable to binding) poses are retained.

Flexible Docking: The flexible docking algorithm in Discovery studio suite allows for receptor flexibility during docking of flexible ligands[22]. The side-chains of specified amino acids are allowed to move during docking. This allows the receptor to adapt to different ligands in an induced-fit model. The protocol uses a combination of components from other algorithms to perform the docking, and is based on methods within CHARMm to sample side-chain and ligand conformations. The Flexible Docking protocol performs the following steps:

- Calculate receptor side-chain conformations: Initially, the protocol creates protein side-chain conformations using the ChiFlex algorithm[12].
- Create ligand conformations: Low energy ligand conformations are created for the docking process using Generate Conformations.
- Perform initial placement of the ligand conformations: The ligand conformations are docked into the active site of each receptor side-chain conformation using LibDock[18].
- Clustering to remove similar ligand poses: Poses are clustered regardless of the protein conformation since the protein conformations are rebuilt during the next step.
- Refine side-chains: In the presence of the ligand, the specified receptor side-chain residues are refined using the ChiRotor algorithm[12].
- Refine docking: A final simulated annealing and energy minimization of each ligand pose is performed using CDOCKER[21].

Grid-based Ligand Docking with Energetics (Glide) v5.7: Glide[23-26] searches for favorable interactions between one or more ligand molecules and a receptor molecule, usually a protein. The ligand poses that Glide generates pass

through a series of hierarchical filters that evaluate the ligand's interaction with the receptor. The initial filters test the spatial fit of the ligand to the defined active site, and examine the complementarity of ligand-receptor interactions using a grid-based method patterned after the empirical ChemScore function[27]. Poses that pass these initial screens enter the final stage of the algorithm, which involves evaluation and minimization of a grid approximation to the OPLS-AA non-bonded ligand-receptor interaction energy. Final scoring is then carried out on the energy-minimized poses. By default, Schrödinger's proprietary GlideScore multi-ligand scoring function is used to score the poses.

Glide Extra-Precision: The extra-precision (XP) mode of Glide[23-26] combines a powerful sampling protocol with the use of a custom scoring function designed to identify ligand poses that would be expected to have unfavorable energies. The more extensive XP docking method and specialized XP scoring method are strongly coupled.

Induced Fit Docking (IFD): Schrödinger Inc (USA) has developed and validated an Induced Fit Docking algorithm[28-30] based on Glide[23-26] and the Refinement module in Prime[31] that accurately predicts ligand binding modes and concomitant structural changes in the receptor. In standard virtual docking studies, ligands are docked into the binding site of a receptor where the receptor is held rigid and the ligand is free to move. However, the assumption of a rigid receptor can give misleading results, since in reality many proteins undergo side-chain or backbone movements, or both, upon ligand binding. These changes allow the receptor to alter its binding site so that it more closely conforms to the shape and binding mode of the ligand. This is often referred to as "induced fit" and is one of the main complicating factors in structure based drug design.

The ability to model induced-fit docking has two main applications:
i) Generation of an accurate complex structure for a ligand known to be active but that cannot be docked in an existing (rigid) structure of the receptor.
ii) Rescue of false negatives (poorly scored true binders) in virtual screening experiments, where instead of screening against a single conformation of the receptor, additional conformations obtained with the induced fit protocol are used.

DOCK: MOE's DOCK application searches for favorable binding configurations between small- to medium-sized ligands and a not-too-flexible macromolecular target, usually a protein. For each ligand, a number of configurations called *poses* are generated and scored. The score can be calculated as either a free energy of binding, which takes into account solvation and entropy, or the enthalpic term of the free energy of binding, which includes a metal ligation term, or a qualitative shaped-based numerical measure. The active site is permitted to move: side chains can be tethered; backbone atoms are always fixed. The Placement phase of Dock generates poses from ligand conformations[32,33]. The available methods are given below:

Alpha Triangle: Poses are generated by superposition of ligand atom triplets and triplets of receptor site points. The receptor site points are *alpha sphere* centers, which represent locations of tight packing. At each of the iteration a random conformation is selected, a random triplet of ligand atoms and a random triplet of alpha sphere centers are used to determine the pose[32,33].

Alpha PMI: Poses are generated by aligning ligand conformations' principal moments of inertia to a randomly generated subset of alpha spheres in the receptor site. This method is most suited to tight pockets, is quite fast and the search space is greatly truncated[32,33].

Proxy Triangle: This method was developed for tackling medium to large sized ligands, which have a very large number of conformers. For small ligands, the Triangle Matcher method will be used[32]. For medium and large ligands, conformers are pre-superposed prior to being placed into the binding site, thus saving computational time.

For large ligands, the score for internal ranking will use atom representatives instead of all the ligand atoms. More time is allowed for larger ligands.

Triangle Matcher: Poses are generated by aligning ligand triplets of atoms on triplets of alpha spheres in a more systematic way than in the Alpha Triangle method[32,33].

FlexX v3.1: FlexX is a computer program for predicting protein-ligand interactions. For a given protein and a ligand, FlexX predicts the geometry of the complex as well as an estimate for the strength of binding. In this version of FlexX, the protein is assumed to be rigid and the docking algorithm in FlexX works without manual intervention. FlexX is ideal for interactive work on protein-ligand complexes as well as for screening a larger set of ligands in order to find new leads for drug design. FlexX provides the choice between the standard ***triangle algorithm*** and the Single Interaction Scan (SIS) placement; the new ***SIS algorithm*** is particularly suited for more hydrophobic pockets and pockets with only a few interaction sites[34-40].

MolDock Optimizer: The MolDock Optimizer algorithm[41] used in MVD is based on an evolutionary algorithm[42,43]. Evolutionary algorithms (EAs) are iterative optimization techniques inspired by Darwinian evolution theory. The guided differential evolution algorithm used for MolDock Optimizer is based on an EA variant called differential evolution (DE). The DE algorithm was introduced by Storn and Price in 1995[44]. Compared to more widely known EA-based techniques (e.g. genetic algorithms, evolutionary programming, and evolution strategies), DE uses a different approach to select and modify candidate solutions (individuals). The main innovative idea in DE is to create offspring from a weighted difference of parent solutions.

Simplex Evolution: Simplex Evolution (SE) algorithm is known to perform better on some complexes where the standard MolDock algorithm fails[41]. It is an alternative search heuristic, which can be used together with either the MolDock or MolDock Grid scoring functions. While other algorithms based on parallel simplex search exist, MolDock SE implementation has been modified to be suitable for docking (by the inclusion of the pose generation step, and the way the initial simplices are created). The simplex evolution algorithm performs a combined local / global search on the poses generated by the pose generator. The local search is performed using the Nelder-Mead local search algorithm[45], but unlike Nelder-Mead's original scheme, the algorithm has been extended to take the position of the other individuals in the population into account.

Iterated Simplex: The Iterated Simplex algorithm[41] is generally more robust (w.r.t. reproducing docking results with similar scores) than the SE and MolDock Optimizer. The algorithm works as follows: First an initial population of poses is created (initial number of poses is determined by the population size parameter). Afterwards, the following process will be executed until max iterations have occurred: Each individual in the population will be refined using the Simplex local search algorithm (also called Nelder-Mead). The Simplex algorithm will run for maximum steps or until the fractional difference between the best and worst vertices in the Simplex (w.r.t. the docking scoring function used) is below a given tolerance value. When all individuals have been refined, the best individual (named: iteration best solution) will be further refined using the same Simplex algorithm again but with a lower tolerance value. When max iterations have occurred the algorithm terminates and returns the best solution(s) found.

Surflex-Dock: Surflex-Dock uses an empirical scoring function and a patented search engine to dock ligands into a protein's binding site. It is particularly successful at eliminating false positive results and can, therefore, be used to narrow down the screening pool significantly, while still retaining a large number of active compounds[46]. Surflex-Dock was developed by Prof. Ajay N. Jain, University of California San Francisco (UCSF) and BioPharmics LLC. *Surflex-Dock (SFXC)* uses Surflex with the default parameter set to dock the ligands. *Surflex-Dock Screen (SFXC)* uses Surflex with the screening parameter set to dock the ligands. *Surflex-Dock Geom (SFXC)* uses Surflex with the

docking accuracy parameter set to dock the ligands. *Surflex-Dock GeomX (SFXC)* uses Surflex with the more exhaustive accuracy parameter set to dock the ligands[47].

Surflex-Dock Protein Flexibility (PF): Surflex-Dock PF allow flexibility of protein atoms whose van der Waals surface distances from ligand atoms are < 4 Å and adapts the active site conformation to the docked ligand[48]. Protein movement takes place in a second Surflex-Dock run, which produces Protein Flexibility Score set.

ii) Scoring Schemes

Ligand scoring is a method to rapidly estimate the binding affinity of a ligand, based on a candidate ligand pose geometry docked into a target receptor structure. Scoring methods typically use empirical functions developed by fitting various functional forms, which characterize various aspects of the receptor-ligand interactions against binding affinity data. An alternative approach makes use of statistical analysis of known ligand-receptor structures and the frequency of occurrence of specific receptor-ligand interactions without requiring any information about binding affinities. This approach is generally referred to as a knowledge-based approach.

LigScore1 & LigScore2 scoring functions: *LigScore1*& *LigScore2* are fast, simple, scoring functions for predicting receptor-ligand binding affinities[49,50]. Three descriptors are used to calculate *LigScore1*, which is computed in units of pK_i (-log K_i). These descriptors are:
vdW: This is a softened Lennard-Jones 6-9 potential. The vdW descriptor is expressed in units of kcal/mol.
C+pol: This is a count of the buried polar surface area between a receptor and ligand, which involves attractive receptor-ligand interactions. The C+_pol descriptor is expressed in units of $Å^2$.
TotPol^2: This is the squared sum of the total polar surface area of the receptor and the ligand molecule expressed in $Å^2$.
Three descriptors are used to calculate *LigScore2,* which is computed in units of pK_i (-log K_i). These descriptors are: *vdW, C+pol* described above &
BuryPol^2: This is the squared sum of the buried polar surface area of the receptor and the ligand molecule expressed in $Å^2$. This term is meant to account for the desolvation penalty incurred by desolvating water molecules from the ligand and receptor cavity so that the ligand can form binding interactions with the receptor.

Piecewise Linear Potential (PLP): Piecewise Linear Potential is a fast, simple, docking function that has been shown to correlate well with protein-ligand binding affinities. PLP scores are measured in arbitrary units, with negative PLP scores reported in order to make them suitable for subsequent use in consensus score calculations. Higher PLP scores indicate stronger receptor-ligand binding (larger pK_i values). Two versions of the PLP function are available: PLP1[51] and PLP2[52].

Jain scoring function*:* A.N. Jain developed an empirical scoring function through an evaluation of the structures and binding affinities of a series of protein-ligand complexes. The Jain score is a sum of five interaction terms[53]. These terms describe:

- Lipophilic interactions
- Polar attractive interactions
- Polar repulsive interactions
- Solvation of the protein and ligand
- An entropy term for the ligand

Only proximate protein-ligand atoms are considered for the pairwise interaction terms. The lipophilic and polar interaction terms are each represented by a weighted sum of a Gaussian and a sigmoidal function. This functional form is short-ranged with a pronounced maximum that occurs at close surface contacts. It also incurs a significant penalty for short contacts between protein and ligand atoms.

Potential of Mean Force (PMF): The PMF[54] and PMF04[55] scoring functions were developed based on statistical analysis of the 3D structures of protein-ligand complexes. They were found to correlate well with protein-ligand binding free energies while being fast and simple to calculate. The scores are calculated by summing pairwise interaction terms over all interatomic pairs of the receptor-ligand complex. The PMF04 score is an updated version of the original PMF score. A larger data set was used for parameterization and, additionally, metal ion and halogen potentials are included. The PMF scores are reported in arbitrary units with the sign reversed to allow for subsequent use in consensus score calculations. A higher score indicates a stronger receptor-ligand binding affinity.

Glide Score: GlideScore[23-26] is based on ChemScore, but includes a steric-clash term, adds buried polar terms devised by Schrödinger to penalize electrostatic mismatches. Glide also computes a specially constructed Coulomb-van der Waals interaction-energy score that is formulated to avoid overly rewarding charge-charge interactions at the expense of charge-dipole and dipole-dipole interactions. This score is intended to be more suitable for comparing the binding affinities of different ligands than is the "raw" Coulomb-van der Waals interaction energy.

GlideScore XP: GlideScore XP[23-26] includes additional terms over the SP scoring function. GlideScore XP specifically rewards occupancy of well-defined hydrophobic pockets by hydrophobic ligand groups. Hydrophobic reward terms are employed in empirical scoring functions such as ChemScore and the SP version of GlideScore in the form of lipophilic-lipophilic pair terms, while other empirical scoring functions use lipophilic surface-area contact terms for this purpose.

ASE Scoring*:* This score is proportional to the sum of the Gaussians $R_1R_2\exp(-0.5d^2)$ over all ligand atom - receptor atom pairs and ligand atom - alpha sphere pairs. R_1 and R_2 are the radii of the atoms in Å, or is -1.85 for alpha spheres. d is the distance between the pair in Å[32,33].

Affinity dG Scoring*:* This function estimates the enthalpic contribution to the free energy of binding using a linear function:

$G = C_{hb} f_{hb} + C_{ion} f_{ion} + C_{mlig} f_{mlig} + C_{hh} f_{hh} + C_{hp} f_{hp} + C_{aa} f_{aa}$

where the *f* terms fractionally count atomic contacts of specific types and the *C*'s are coefficients that weight the term contributions to the affinity estimate[32,33]. The individual terms are:

hb = Interactions between hydrogen bond donor-acceptor pairs. An optimistic view is taken; for example, two hydroxyl groups are assumed to interact in the most favorable way.
ion =Ionic interactions. A Coulomb-like term is used to evaluate the interactions between charged groups. This can contribute to or detract from binding affinity.
mlig =Metal ligation. Interactions between Nitrogens/Sulfurs and transition metals are assumed to be metal ligation interactions.
hh =Hydrophobic interactions, for example, between alkane carbons; these interactions are generally favorable.
hp =Interactions between hydrophobic and polar atoms; these interactions are generally unfavorable.
aa =An interaction between any two atoms. This interaction is weak and generally favorable.

Alpha HB Scoring: This score is a linear combination of two terms. The first term measures the geometric fit of the ligand to the binding site. The second term measures hydrogen bonding effects. Both terms are summed over all ligand atoms[32,33].

The first term has an attractive part and a repulsive part. The attractive part is summed over atoms in the ligand. Each ligand atom that is within 3 Å of an alpha sphere center contributes $A\exp(-0.5d^2)$ to this term, where d is the

distance from the ligand atom to the nearest alpha sphere center, and A assumes the value of -0.6845. The repulsive part is summed over all pairs of atomic overlap between the ligand and the receptor. For each pair of overlap, the contribution is between 0 and 1 depending on the severity of the overlap.

The second term measures hydrogen bond effects. For non-*sp3* donors and acceptors, hydrogen bonding sites are projected from the atom. If the site is occupied by a favorable atom, there is a score of -2 (negative means favorable). Otherwise if it is occupied by some other ligand atom, there is a score of +1. For *sp3* donors and acceptors, all favorable atoms within 3.5 Å contribute a score of -1 while all other atoms contribute +1

London dG Scoring: The London dG scoring function estimates the free energy of binding of the ligand from a given pose[32,33]. The functional form is a sum of terms:

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{HB} f_{HB} + \sum_{m-lig} c_M f_M + \sum_{atoms\ i} \Delta D_i$$

where c represents the average gain/loss of rotational and translational entropy; E_{flex} is the energy due to the loss of flexibility of the ligand (calculated from ligand topology only); f_{HB} measures geometric imperfections of hydrogen bonds and takes a value in [0,1]; c_{HB} is the energy of an ideal hydrogen bond; f_M measures geometric imperfections of metal ligations and takes a value in [0,1]; c_M is the energy of an ideal metal ligation; and D_i is the desolvation energy of atom i. The difference in desolvation energies is calculated according to the formula

$$\Delta D_i = c_i R_i^3 \left\{ \iiint_{u \notin A \cup B} |u|^{-6}\, du - \iiint_{u \notin B} |u|^{-6}\, du \right\}$$

where A and B are the protein and/or ligand volumes with atom i belonging to volume B; R_i is the solvation radius of atom i (taken as the OPLS-AA van der Waals sigma parameter plus 0.5 Å); and c_i is the desolvation coefficient of atom i. The coefficients (c, c_{HB}, c_M, c_i) were fitted from ~400 x-ray crystal structures of protein-ligand complexes with available experimental pKi data. Atoms are categorized into about a dozen atom types for the assignment of the c_i coefficients. The triple integrals are approximated using Generalized Born integral formulas.

HYDE Scoring: HYDE scoring function is based on pure physical principles: the hydrogen bond energy and the hydrophobic effect are both implemented in a consistent way via atomic *logP* increments. The new approach in *Hyde* is the heavy penalization of unmet interactions: A hydrogen bond taken out of the solvent - and not having an ideal partner in the protein; or the phenyl ring that is dehydrated and put into a hydrophilic active site region. Thus, not by rewarding H-bonds but by penalizing unfavorable situations, false positives are effectively ruled out in an *in silico* screen[56].

MolDock scoring: MolDock Score is derived from the PLP scoring functions originally proposed by Gehlhaar *et al.*[51] and later extended by Yang *et al.*[57] The MolDock scoring function further improves these scoring functions with a new hydrogen bonding term and new charge schemes. The ***MolDock Grid Scoring*** is a grid approximation using the same energy terms as the MolDock Score except that hydrogen bond directionality is not taken into account.

PLANTS scoring: The PLANTS scoring function is derived from the scoring function originally proposed by Korb *et al.*[58] MolDock further improves these scoring functions with a new hydrogen bonding term and new charge schemes[41]. ***PLANTS Grid Scoring*** is a grid approximation using the same energy terms as the PLANTS Score. The grid-based scoring functions provide a 4-5 times speed up by precalculating potential-energy values on an evenly spaced cubic grid.

The Surflex-Dock Scoring Function (Surflex Score): Surflex-Dock uses an empirically derived scoring function that is based on the binding affinities of protein-ligand complexes and on their X-ray structures. The Surflex-Dock scoring function is a weighted sum of non-linear functions involving van der Waals surface distances between the appropriate pairs of exposed protein and ligand atoms[53]. The scoring function includes the following terms:

- *Hydrophobic*--A weighted sum over all atom pairs, in which at least one atom is non-polar, of functions capturing the positive atomic contacts and the atomic interpenetration.
- *Polar*--A sum over all pairs of complementary polar atoms of a function capturing the effects of hydrogen bonds and salt bridges. This function includes a directionality term that favors hydrogen bonding geometries observed in crystal structures and a term that accounts for favorable interactions between formally charged atoms (if present).
- *Repulsive*--A sum over all pairs of polar atoms that are of the same sign. This function captures unfavorable polar contacts.
- *Entropic*--A function that models the loss of translational and rotational entropy of the ligand once it is docked. This function takes into account the number of rotatable ligand bonds and the ligand's molecular weight.
- *Solvation*--A function that captures the difference between the potential and actual numbers of hydrogen bond equivalents.
- *Crash*--The degree of inappropriate penetration into the protein by the ligand as well as the degree of internal self-clashing that the ligand is experiencing. Crash scores that are close to 0.0 are favorable.

G Score: This scoring function is drawn from the work of Willett's group[59], using the hydrogen bonding, complex (ligand-protein), and internal (ligand-ligand) energies.

PMF Score: This scoring function is drawn from the work of Muegge and Martin[54], who analyzed a large set of complexes drawn from the Protein Data Bank and developed a set of Helmholtz free energies of interactions for protein-ligand atom pairs (Potential of Mean Force, PMF).

D Score: This scoring function is drawn from the work of Kuntz's group[60], using only the charge and van der Waals interactions between the protein and the ligand.

ChemScore: This scoring function is based on the work of Eldridge, Murray, Auton, Paolini, and Mee[27]. It includes terms for hydrogen bonding, metal-ligand interaction, lipophilic contact, and rotational entropy, along with an intercept term.

Section D

Preprocessing of protein structures

One major issue with macromolecular X-ray crystal structures is that of missing or poorly resolved atomic data. Areas that cannot be well-resolved may result in either multiple models, alternate locations (atoms are assigned "partial occupancies" based on how often they are found in different locations) or data being absent altogether. The severity of missing data ranges from occasional missing atoms through to entire sections of the structure being absent. In many cases the missing data needs to be modeled and fixed before subsequent computational analyses can proceed. The purpose of protein data preparation or preprocessing is to correct for errors in the structure and to prepare macromolecular data for further computational analysis.

Preprocessing of protein structure in Accelrys Discovery Studio: The Prepare Protein protocol in Accelrys Discovery studio suite was used for preprocessing of protein structures, performing tasks such as inserting missing atoms in incomplete residues, modeling missing loop regions, deleting alternate conformations (disorder), removing waters, standardizing atom names, and protonating titratable residues using predicted pKs. For the residues with missing atoms, the atoms were inserted and the conformation of the side-chain atoms of those residues was optimized using the ChiRotor algorithm[12]. The patched residue segments were treated as loop regions and optimized using the MODELER loop refinement algorithm[61], which optimizes the added loops simultaneously and produces a single model structure. It is not intended to provide the best conformation of all loops at this stage, but only to produce a structure having reasonable loop conformations that do not clash with the rest of the structure. Then, each loop was further optimized sequentially using the Looper algorithm[9]. By default, water molecules were removed from the structure. The ligands in the protein structure were retained in the calculation by default. Finally, prediction of pK of the titratable residues and optimization of the hydrogen positions were done[62,63]. Accelrys Discovery Studio pre-processing of protein structures is a part of structure refinement and validation of modeled protein structures (AnFAEB & TsFAEC).

Preparation of protein structures in Schrödinger suite: The *Protein Preparation Wizard* in Schrödinger suite allows taking a protein from its raw state, (which may be missing hydrogen atoms and have incorrect bond order assignments, charge states, or orientations of various groups) to a state in which it is properly prepared for calculations using Schrödinger products. A typical PDB structure file consists only of heavy atoms and may include a cocrystallized ligand, water molecules, metal ions, and cofactors. Some structures are multimeric, and may need to be reduced to a single unit. Because of the limited resolution of X-ray experiments, it can be difficult to distinguish between NH and O, and the placement of these groups must be checked. PDB structures may be missing information on connectivity, which must be assigned, along with bond orders and formal charges. A set of tools assembled by Schrödinger Inc. has therefore used to prepare proteins in a form that is suitable for modeling calculations[64]. Prior to docking studies, the pre-processing of all protein structures (AnFAEA, AnFAEB & TsFAEC) was done without the native ligand using Schrödinger Suite.

Supplementary Figure S2. Comparison of ligand-receptor interactions in unprocessed and processed 1UWC crystal structures.

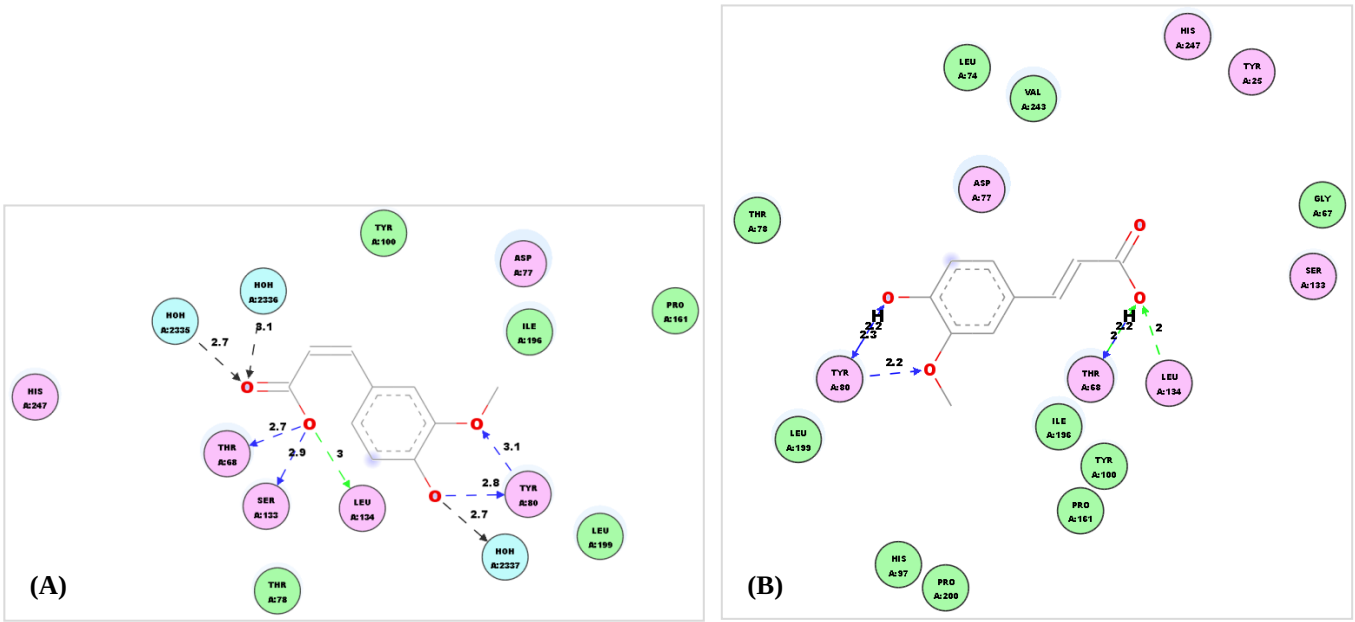
(A) and **(B)** are 2D interaction diagrams of unprocessed and processed crystal structures, respectively. Residues involved in hydrogen-bond, charge or polar interactions are represented by magenta-colored circles. Residues involved in van der Waals interactions are represented by green circles. Water molecules are represented by aquamarine circles. Hydrogen-bond interactions with non-amino acid residues are represented by a black dashed line with an arrow head directed towards the electron donor. Hydrogen-bond interactions with amino acid main chains are represented by a green dashed line with an arrow head directed towards the electron donor. Hydrogen-bond interactions with amino acid side-chains are represented by a blue dashed line with an arrow head directed towards the electron donor. The interaction distances were given in Å units. **(C)** and **(D)** are Ligand binding pattern diagrams of unprocessed and processed crystal structures respectively. The ligand is shown in stick model and the amino acid residues as line model. The binding patterns of hydrogen bond donors and acceptors were depicted in green dashed lines. **(E)** and **(F)** are Ligand binding pattern diagrams of unprocessed and processed crystal structures respectively highlighting polar and nonpolar contacts depicted as magenta lines.

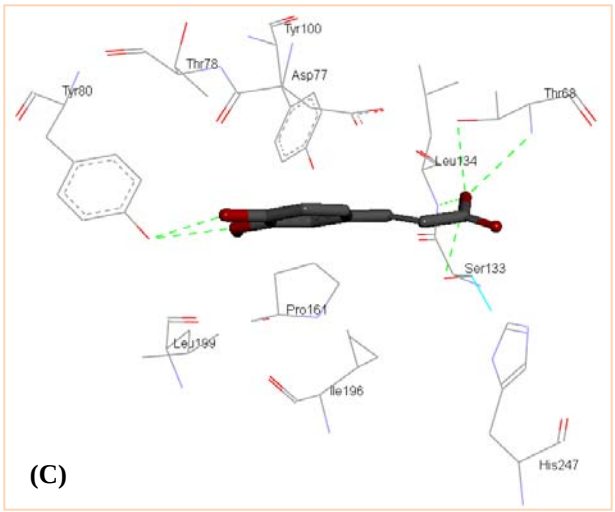
Residues involved in ligand-receptor interactions of unprocessed 1UWC structure were grouped as follows:

- Main-chain HB donors: Thr68, Leu134
- Side-chain HB donors: Thr68, Tyr80, Ser133
- Side-chain HB acceptors: Thr68, Tyr80, Ser133
- Main-chain polar contacts: Asp77, Thr68, Leu134
- Side-chain polar contacts: Thr68, Tyr80, Ser133, His247
- Side-chain nonpolar contacts: Thr68, Asp77, Tyr80, Ser133, Pro161, Leu199, Ile196, His247

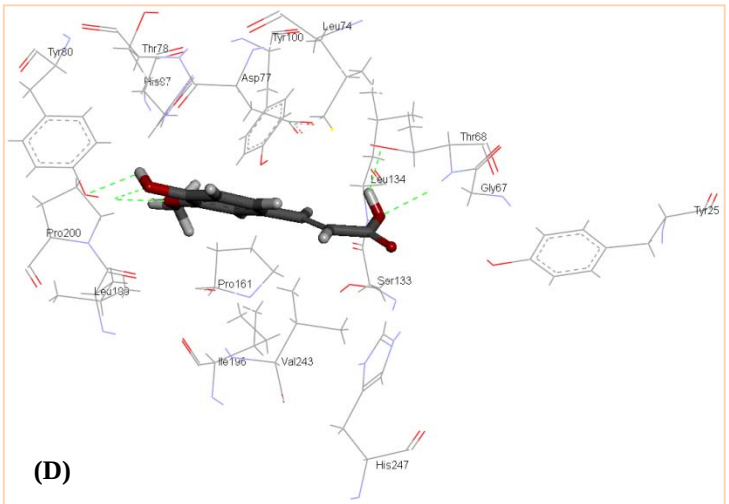
Residues involved in ligand-receptor interactions of processed 1UWC structure were grouped as follows:

- Main-chain HB donors: Thr68, Leu134
- Side-chain HB donors: Tyr80
- Side-chain HB acceptors: Thr68, Tyr80
- Main-chain polar contacts: Asp77, Thr68, Leu134
- Side-chain polar contacts: Tyr25, Thr68, Asp77, Tyr80
- Main-chain nonpolar contacts: Asp77, Leu134, Pro161, Leu199
- Side-chain nonpolar contacts: Thr68, Leu74, Asp77, Thr78, Tyr80, His97, Tyr100, Ser133, Leu134, Pro161, Leu199, Ile196, Pro200, Val243, His247

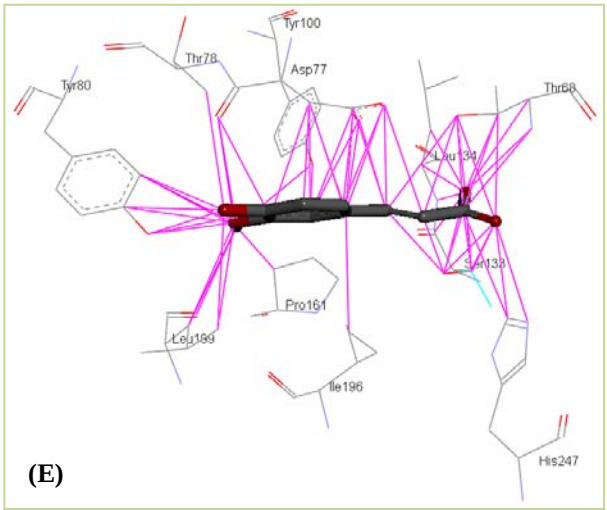



Tyr100
Thr78
Asp77
Tyr80
Thr68
Leu134
Ser133
Pro161
Leu199
Ile196
His247

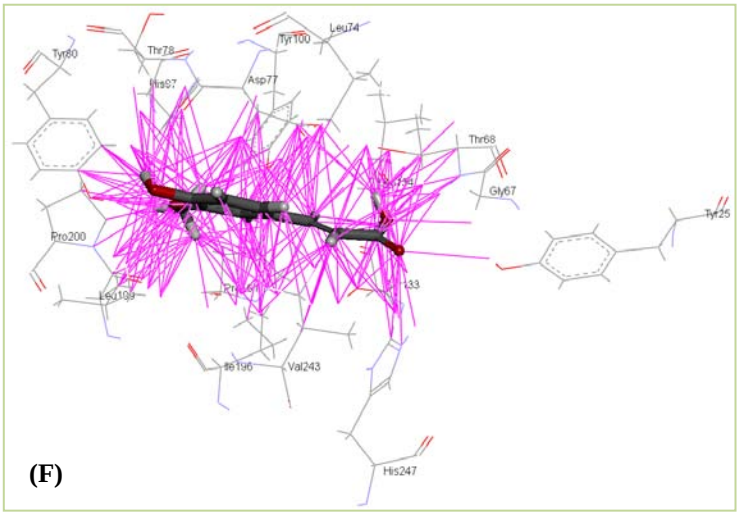
(C)


Leu74
Tyr100
Tyr80
Thr78
His87
Asp77
Thr68
Leu134
Gly67
Tyr25
Pro200
Pro161
Ser133
Leu199
Ile196
Val243
His247

(D)


Tyr100
Thr78
Asp77
Tyr80
Thr68
Leu134
Ser133
Pro161
Leu199
Ile196
His247

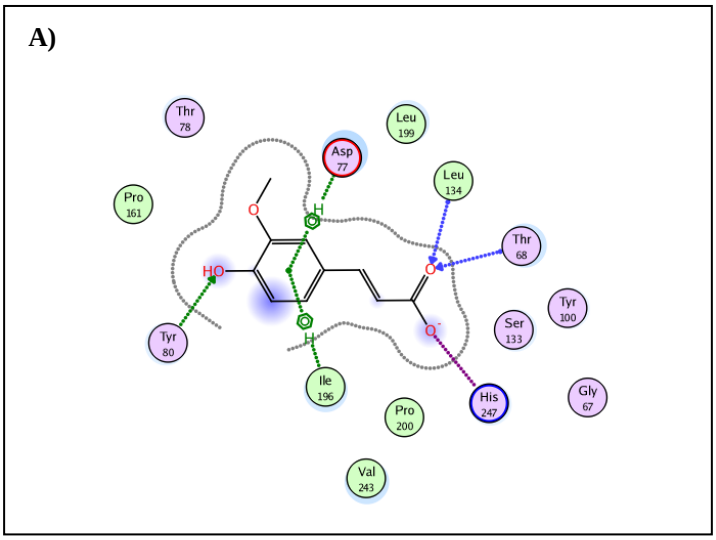
(E)


Leu74
Tyr100
Tyr80
Thr78
His87
Asp77
Thr68
Gly67
Tyr25
Pro200
Leu199
Ile196
Val243
His247

(F)

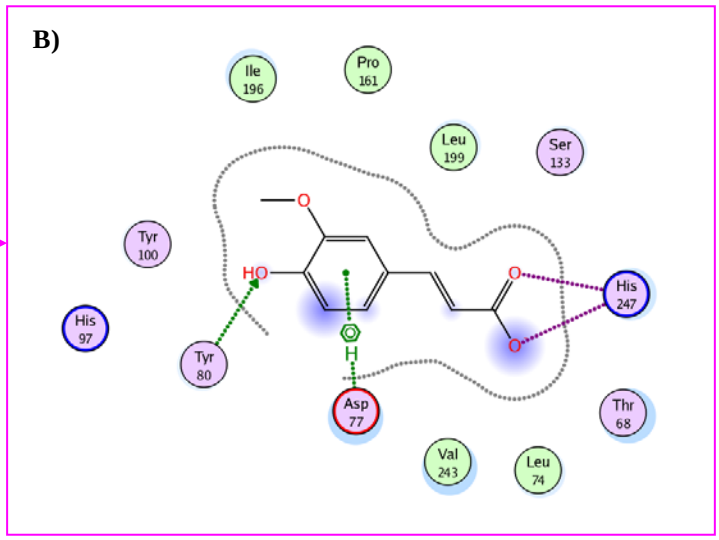
Section E

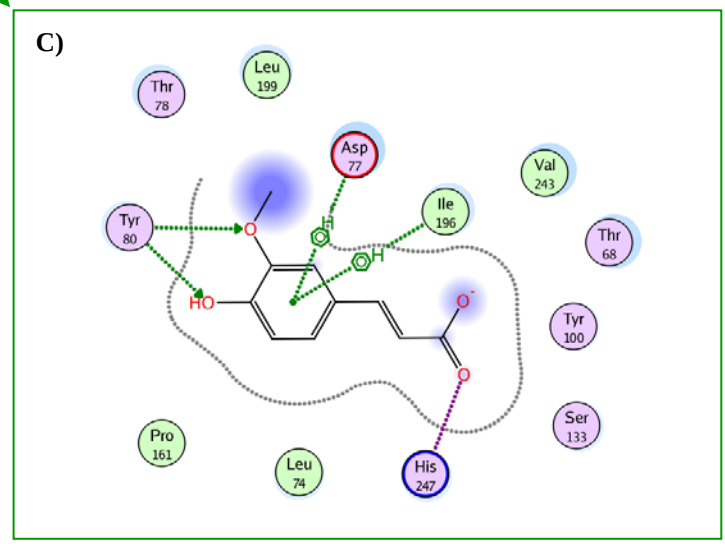
Supplementary Figure S3. Ligand interaction diagrams generated using MOE v2010.10. (A) Preprocessed 1UWC cognate ligand structure. (B) Pose 1 resulted from docking using Alpha Triangle. (C) Pose 3 resulted from docking using Alpha Triangle.

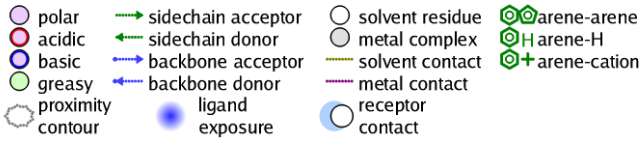


Supplementary Table S3. Resulting poses and respective KISS score values of cognate-ligand docking (1UWC) using Alpha Triangle.

Pose	RMSD (Å)	KISS score
1	**1.39**	**0.5**
2	6.40	0
3	**2.50**	**0.66**
4	5.63	0
5	6.49	0
6	8.23	0
7	8.66	0.16
8	8.58	0







Calculation of KISS score

The function for calculating the KISS score is given below:

where, I_r = Number of reproduced hydrogen bond interactions by the docked pose. I_c = Total hydrogen bond interactions present in the binding pose of processed cognate ligand crystal structure.

In the case of pose selection studies with AnFAEA (see Supplementary Table S3 and Supplementary Fig. S3 above), even though a high RMSD of 2.5 Å was observed from the binding mode seen in the crystal structure, the docked pose 3 generated by the Alpha Triangle docking algorithm reached a KISS score of 0.66. Whereas, the best pose (pose rank 1) according to the low RMSD consideration (1.39 Å) generated by the same Alpha Triangle docking algorithm was considered to be less accurate as it showed a KISS score of 0.5.

- As shown in Supplementary Fig. S3A, the ligand was able to form six hydrogen bond interactions with the receptor.
- In pose 1 (Supplementary Fig. S3B), the ligand was able to form four hydrogen bond interactions with the receptor. Only three hydrogen bond interactions were same as observed in cognate ligand structure that results a KISS score of 0.5 for this pose (Supplementary Table S3).
- In pose 3 (Supplementary Fig. S3C), the ligand was able to form five hydrogen bond interactions with the receptor. Only four hydrogen bond interactions were same as observed in cognate ligand structure that results a KISS score of 0.66 for this pose.

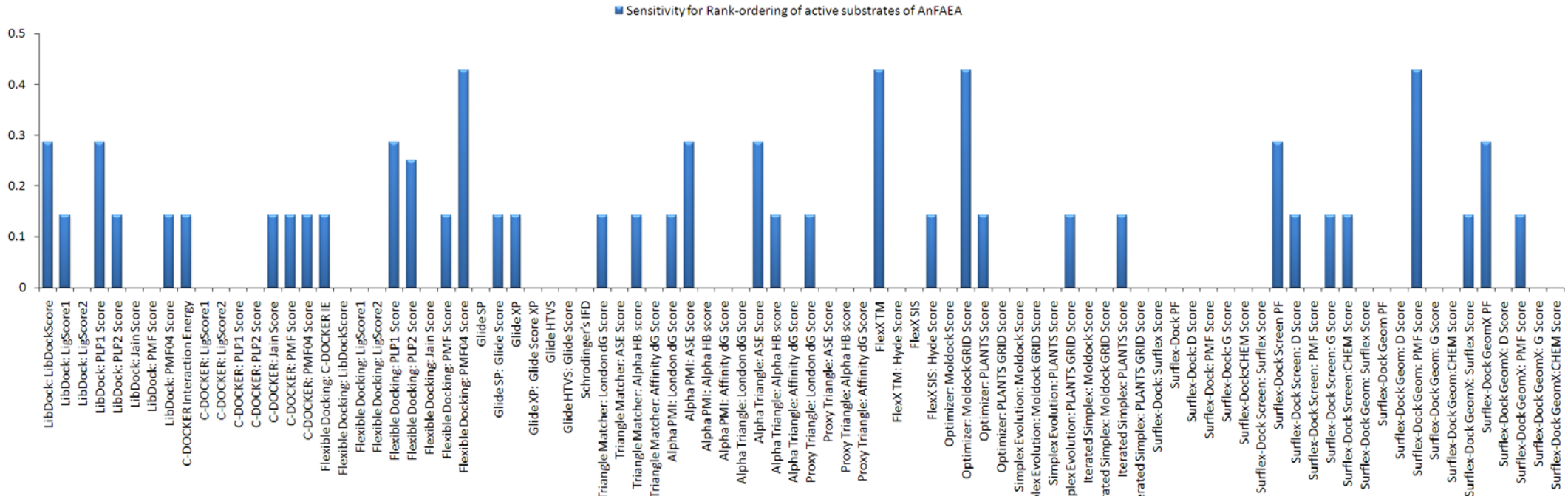
Section F. Supplementary Table S4. Ranking for docking programs with respect to their enrichment performace for the three feruloyl esterase enzymes. MCC was used as the assessment method for evaluating the docking program-scoring function sets.

	0.9 - 1.0		0.7 - 0.8		0.5 - 0.6		> 0.4
	0.8 - 0.9		0.6 - 0.7		0.4 - 0.5		

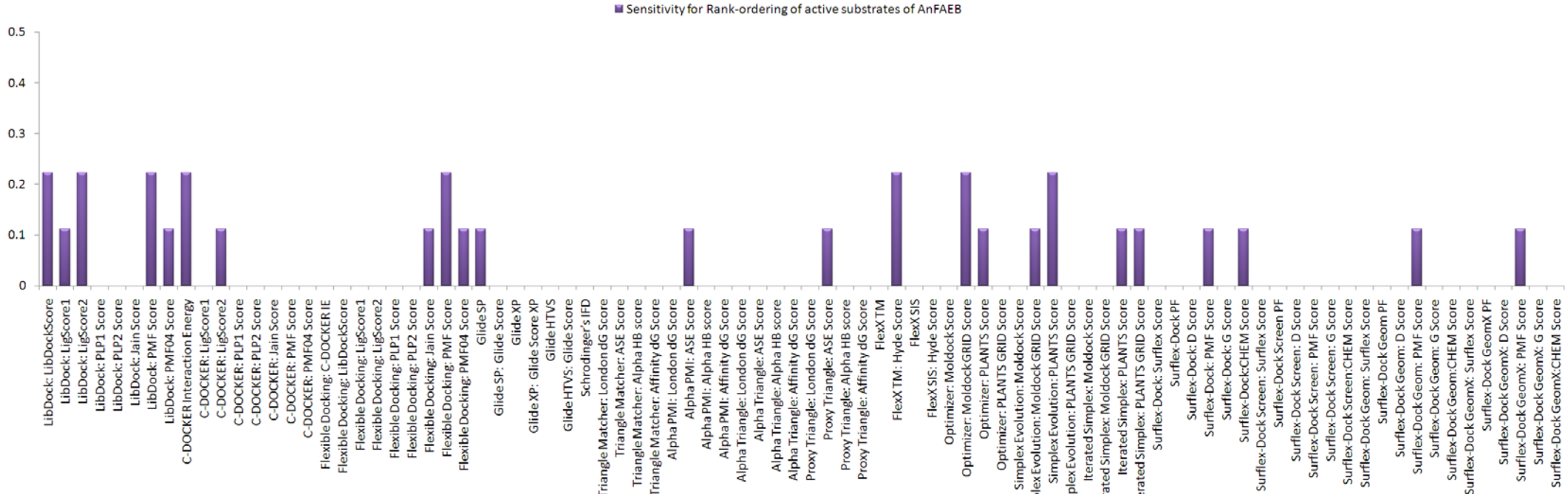
AnFAEA	MCC	AnFAEB	MCC	TsFAEC	MCC
Schrodinger's IFD	1.00	LibDock		C-DOCKER: PLP1 Score	
Surflex-Dock Screen: Surflex Score		LibDock: PLP2 Score		C-DOCKER: PMF Score	
Surflex-Dock: PMF Score		LibDock: Jain Score	0.60	C-DOCKER: PMF04 Score	
Surflex-Dock Screen:CHEM Score		Alpha Triangle: London dG Score		Glide XP: Glide Score XP	0.84
Surflex-Dock Screen: PMF Score		FlexX SIS: Hyde Score		FlexX TM	
Surflex-Dock Screen: PF Score		Optimizer: Moldock GRID Score		Iterated Simplex: PLANTS Score	
Surflex-Dock Screen: D Score		LibDock: LigScore1		Iterated Simplex: PLANTS GRID Score	
Surflex-Dock GeomX: Surflex Score		LibDock: LigScore2		C-DOCKER: LigScore1	
Surflex-Dock GeomX: PMF Score		LibDock: PLP1 Score		C-DOCKER: PLP2 Score	
Surflex-Dock GeomX: PF Score		LibDock: PMF Score		Glide SP	
Surflex-Dock GeomX: D Score		LibDock: PMF04 Score		Glide SP: Glide Score	
Surflex-Dock Geom: Surflex Score		C-DOCKER: PLP2 Score		Glide XP	
Surflex-Dock Geom: PF Score		Flexible Docking: LibDock		Schrodinger's IFD	
Surflex-Dock Geom: G Score		Flexible Docking: Jain Score		Triangle Matcher: London dG Score	
Surflex-Dock Geom: D Score		Flexible Docking: PMF Score		Triangle Matcher: ASE Score	
Simplex Evolution: Moldock Score		Flexible Docking: PMF04 Score	0.42	Triangle Matcher: Alpha HB score	
Simplex Evolution: Moldock GRID Score		Glide SP: Glide Score		Triangle Matcher: Affinity dG Score	
Optimizer: PLANTS GRID Score	0.73	Alpha PMI: Affinity dG Score		Alpha PMI: London dG Score	
Optimizer: Moldock GRID Score		FlexX TM		Alpha PMI: ASE Score	
LibDock: PMF04 Score		FlexX TM: Hyde Score		Alpha PMI: Affinity dG Score	
LibDock: PMF Score		FlexX SIS		Alpha Triangle: London dG Score	
LibDock: PLP2 Score		Optimizer: Moldock Score		Alpha Triangle: Alpha HB score	0.69
LibDock: PLP1 Score		Iterated Simplex: Moldock Score		Proxy Triangle: London dG Score	
LibDock		Iterated Simplex: PLANTS Score		Proxy Triangle: Alpha HB score	
Glide XP		Iterated Simplex: PLANTS GRID Score		Proxy Triangle: Affinity dG Score	
Glide SP: Glide Score		C-DOCKER		FlexX SIS	
FlexX TM: Hyde Score		C-DOCKER: LigScore1		FlexX SIS: Hyde Score	
FlexX TM		C-DOCKER: LigScore2		Optimizer: Moldock Score	
Flexible Docking: PMF04 Score		C-DOCKER: PLP1 Score		Optimizer: PLANTS Score	
Flexible Docking: PMF Score		C-DOCKER: Jain Score		Optimizer: PLANTS GRID Score	
Flexible Docking: PLP1 Score		C-DOCKER: PMF Score		Simplex Evolution: Moldock Score	
Alpha PMI: ASE Score		C-DOCKER: PMF04 Score		Simplex Evolution: PLANTS Score	
Alpha PMI: Alpha HB score		Flexible Docking: LigScore1		Simplex Evolution: PLANTS GRID Score	
LibDock: LigScore1		Flexible Docking: PLP1 Score		Iterated Simplex: Moldock Score	
LibDock: Jain Score		Flexible Docking: PLP2 Score		Iterated Simplex: Moldock GRID Score	
C-DOCKER: PMF04 Score		Glide SP		C-DOCKER	
Flexible Docking: C-DOCKER		Glide XP		C-DOCKER: LigScore2	
Flexible Docking: LibDock		Glide HTVS		C-DOCKER: Jain Score	
Flexible Docking: LigScore2		Glide HTVS: Glide Score		Glide HTVS	
Flexible Docking: PLP2 Score		Triangle Matcher: London dG Score		Glide HTVS: Glide Score	
Glide SP		Triangle Matcher: Alpha HB score		Alpha PMI: Alpha HB score	0.55
Glide HTVS: Glide Score		Alpha PMI: London dG Score		Alpha Triangle: ASE Score	
Triangle Matcher: London dG Score		Alpha PMI: Alpha HB score		Alpha Triangle: Affinity dG Score	
Triangle Matcher: ASE Score		Alpha Triangle: Affinity dG Score		Proxy Triangle: ASE Score	
Triangle Matcher: Alpha HB score		Proxy Triangle: London dG Score		Optimizer: Moldock GRID Score	
Triangle Matcher: Affinity dG Score		Proxy Triangle: Alpha HB score		Simplex Evolution: Moldock GRID Score	
Alpha PMI: Affinity dG Score		Proxy Triangle: Affinity dG Score		Surflex-Dock Geom: Surflex Score	
Alpha Triangle: London dG Score		Optimizer: PLANTS Score		Surflex-Dock Geom: D Score	
Alpha Triangle: ASE Score		Simplex Evolution: Moldock Score		Surflex-Dock Geom: PMF Score	
Alpha Triangle: Alpha HB score		Simplex Evolution: Moldock GRID Score		Surflex-Dock Geom: G Score	
Proxy Triangle: London dG Score	0.49	Simplex Evolution: PLANTS Score		Surflex-Dock Geom:CHEM Score	0.42
Proxy Triangle: ASE Score		Simplex Evolution: PLANTS GRID Score		Surflex-Dock GeomX: Surflex Score	
Proxy Triangle: Alpha HB score		Iterated Simplex: Moldock GRID Score		Surflex-Dock GeomX: D Score	
Proxy Triangle: Affinity dG Score		Flexible Docking: C-DOCKER		Surflex-Dock GeomX: PMF Score	
FlexX SIS: Hyde Score		Flexible Docking: LigScore2		Surflex-Dock GeomX: G Score	
Optimizer: Moldock Score		Glide XP: Glide Score XP		Surflex-Dock GeomX:CHEM Score	
Optimizer: PLANTS Score		Schrodinger's IFD	> 0.4	Surflex-Dock: Surflex Score	
Simplex Evolution: PLANTS Score		Triangle Matcher: ASE Score		Surflex-Dock: PF Score	
Simplex Evolution: PLANTS GRID Score		Triangle Matcher: Affinity dG Score		Surflex-Dock: D Score	
Iterated Simplex: Moldock GRID Score		Alpha PMI: ASE Score		Surflex-Dock: PMF Score	
Surflex-Dock: Surflex Score		Alpha Triangle: ASE Score		Surflex-Dock: G Score	
Surflex-Dock: D Score		Alpha Triangle: Alpha HB score		Surflex-Dock:CHEM Score	
Surflex-Dock: G Score		Proxy Triangle: ASE Score		Surflex-Dock Screen: Surflex Score	
Surflex-Dock:CHEM Score		Optimizer: PLANTS GRID Score		Surflex-Dock Screen: PF Score	
Surflex-Dock Screen: G Score		Surflex-Dock: Surflex Score		Surflex-Dock Screen: D Score	
Surflex-Dock Geom: PMF Score		Surflex-Dock: PF Score		Surflex-Dock Screen: PMF Score	
Surflex-Dock Geom:CHEM Score		Surflex-Dock: D Score		Surflex-Dock Screen: G Score	
Surflex-Dock GeomX: G Score		Surflex-Dock: PMF Score		Surflex-Dock Screen:CHEM Score	
Surflex-Dock GeomX:CHEM Score		Surflex-Dock: G Score		Surflex-Dock Geom: PF Score	
LibDock: LigScore2		Surflex-Dock:CHEM Score		FlexX TM: Hyde Score	
C-DOCKER		Surflex-Dock Screen: Surflex Score		Surflex-Dock GeomX: PF Score	
C-DOCKER: LigScore2		Surflex-Dock Screen: PF Score		Flexible Docking: Jain Score	> 0.4
C-DOCKER: PLP1 Score		Surflex-Dock Screen: D Score		Flexible Docking: PMF04 Score	
C-DOCKER: PLP2 Score		Surflex-Dock Screen: PMF Score		Flexible Docking: C-DOCKER	
C-DOCKER: PMF Score		Surflex-Dock Screen: G Score		Flexible Docking: LibDock	
Flexible Docking: LigScore1		Surflex-Dock Screen:CHEM Score		Flexible Docking: LigScore1	
Glide XP: Glide Score XP		Surflex-Dock Geom: Surflex Score		Flexible Docking: LigScore2	
Glide HTVS		Surflex-Dock Geom: PF Score		Flexible Docking: PLP1 Score	
Iterated Simplex: Moldock Score	> 0.4	Surflex-Dock Geom: D Score		Flexible Docking: PLP2 Score	
Surflex-Dock: PF Score		Surflex-Dock Geom: PMF Score		Flexible Docking: PMF Score	
C-DOCKER: LigScore1		Surflex-Dock Geom: G Score		LibDock	
Flexible Docking: Jain Score		Surflex-Dock Geom:CHEM Score		LibDock: LigScore1	
Alpha PMI: London dG Score		Surflex-Dock GeomX: Surflex Score		LibDock: LigScore2	
Alpha Triangle: Affinity dG Score		Surflex-Dock GeomX: PF Score		LibDock: PLP1 Score	
FlexX SIS		Surflex-Dock GeomX: D Score		LibDock: PLP2 Score	
Iterated Simplex: PLANTS Score		Surflex-Dock GeomX: PMF Score		LibDock: Jain Score	
C-DOCKER: Jain Score		Surflex-Dock GeomX: G Score		LibDock: PMF Score	
Iterated Simplex: PLANTS GRID Score		Surflex-Dock GeomX:CHEM Score		LibDock: PMF04 Score	

Section G

Supplementary Figure S4. Sensitivity of docking algorithm-scoring functions for rank-ordering active substrates of AnFAEA



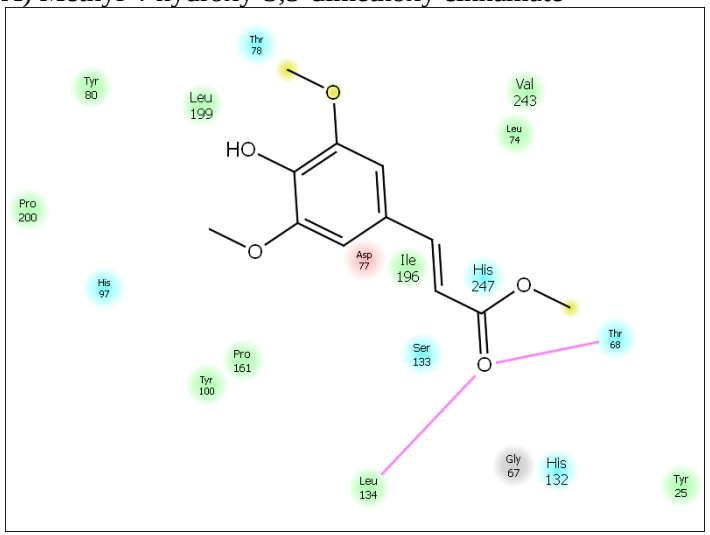
Supplementary Figure S5. Sensitivity of docking algorithm-scoring functions for rank-ordering active substrates of AnFAEB



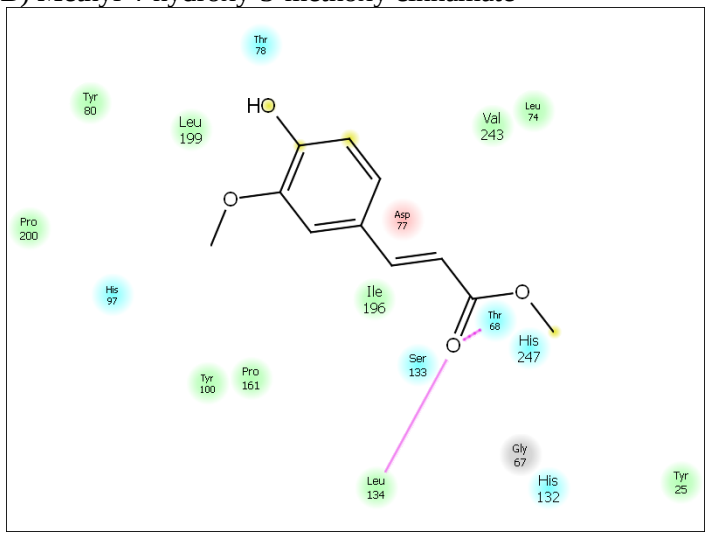
Supplementary Figure S6. Sensitivity of docking algorithm-scoring functions for rank-ordering active substrates of TsFAEC

Supplementary Figure S7. Ligand interaction diagrams for the top scoring poses of respective substrates obtained during enrichment studies for AnFAEA by Schrödinger's Glide SP algorithm. It is evident from the interaction diagrams that hydrogen bonding with Thr 68 and Leu 134 residues plays a crucial role in the binding mechanism. A) to G): Ligand interaction diagrams for the active substrates of AnFAEA. H) to O): Ligand interaction diagrams for the inactive substrates of AnFAEA.

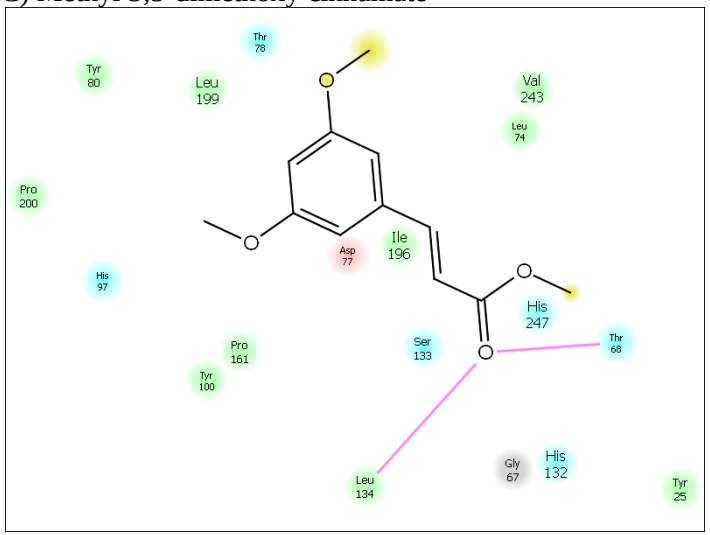
A) Methyl 4-hydroxy-3,5-dimethoxy cinnamate



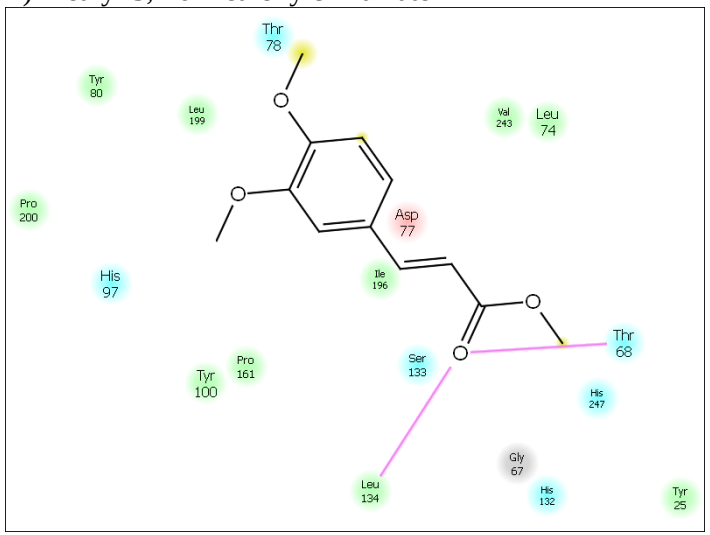
B) Methyl 4-hydroxy-3-methoxy cinnamate



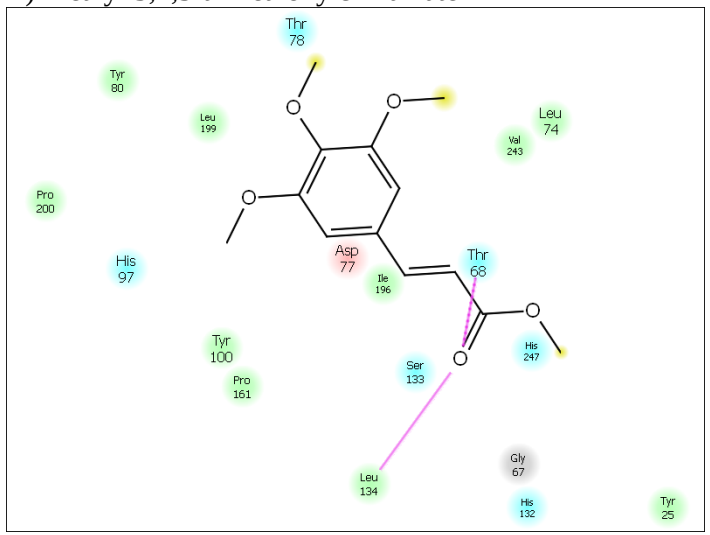
C) Methyl 3,5-dimethoxy cinnamate



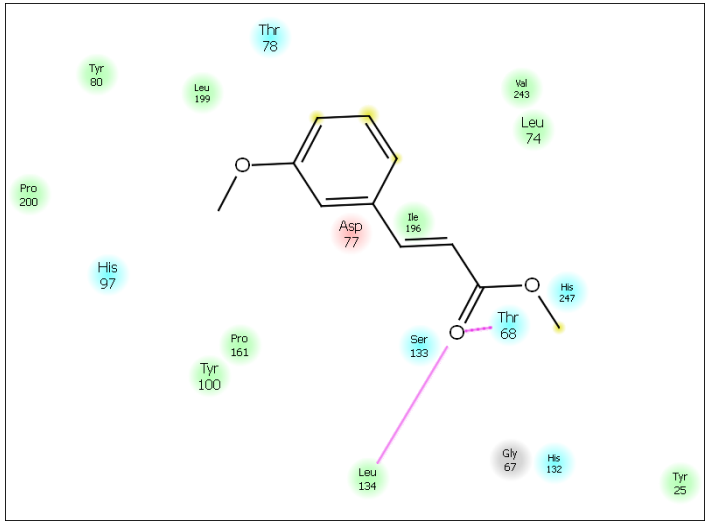
D) Methyl 3,4-dimethoxy cinnamate



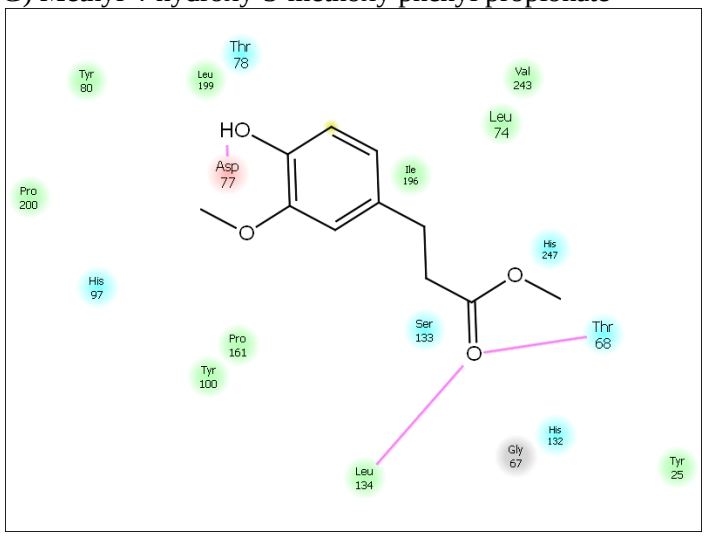
E) Methyl 3,4,5-trimethoxy cinnamate



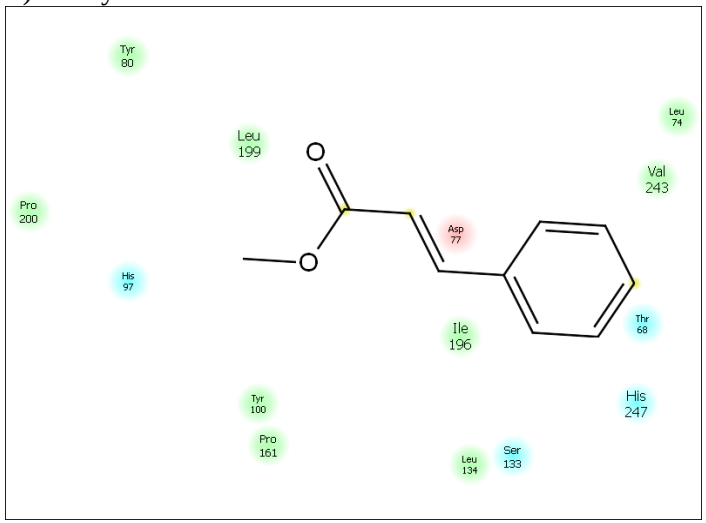
F) Methyl 3-methoxy cinnamate



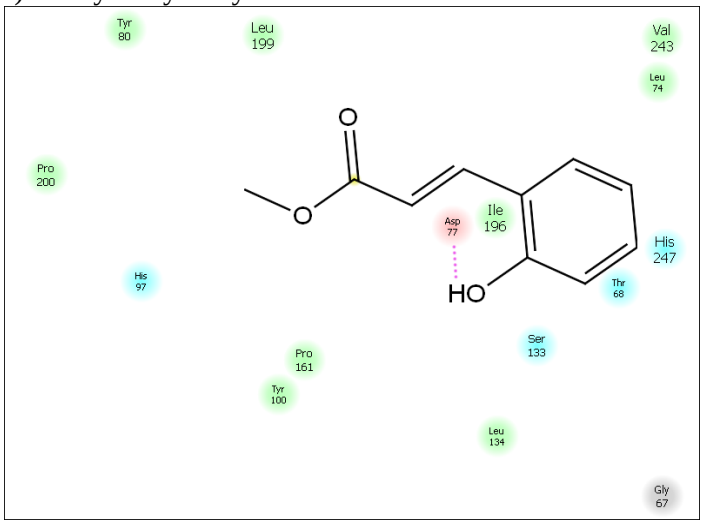
G) Methyl 4-hydroxy-3-methoxy phenyl propionate



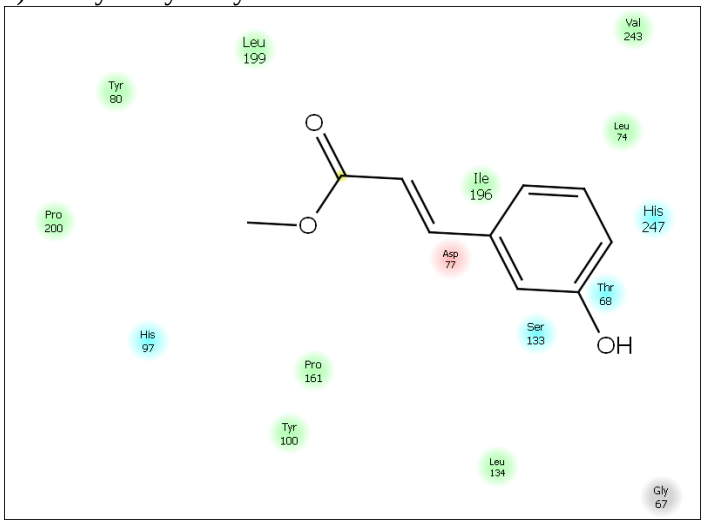
H) Methyl cinnamate



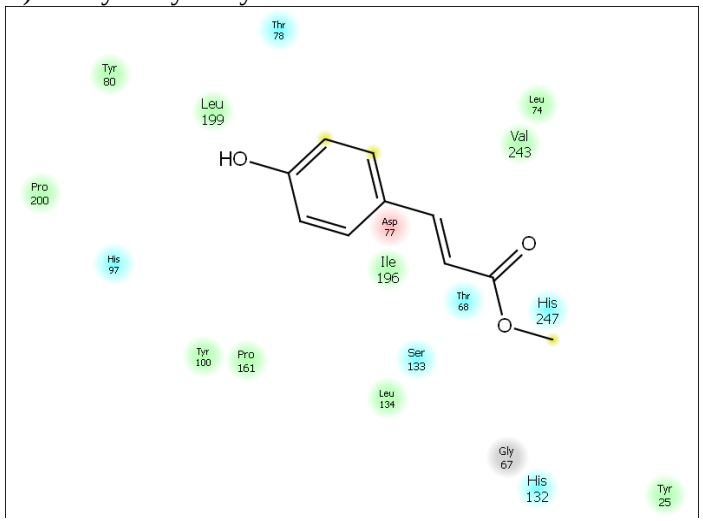
I) Methyl 2-hydroxy cinnamate



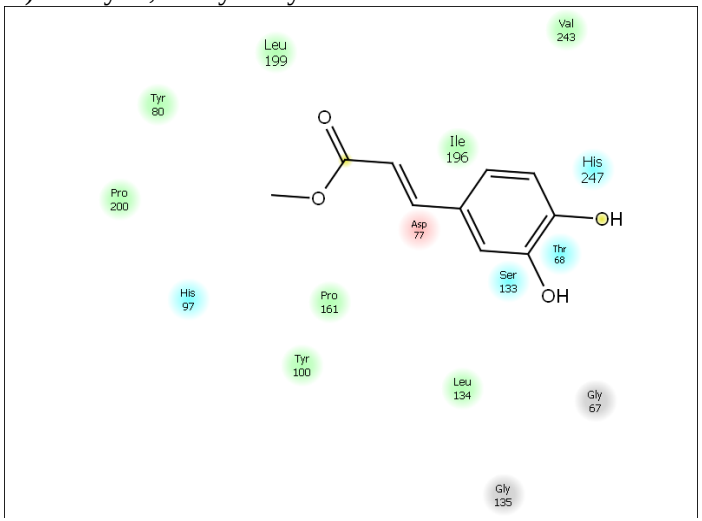
J) Methyl 3-hydroxy cinnamate



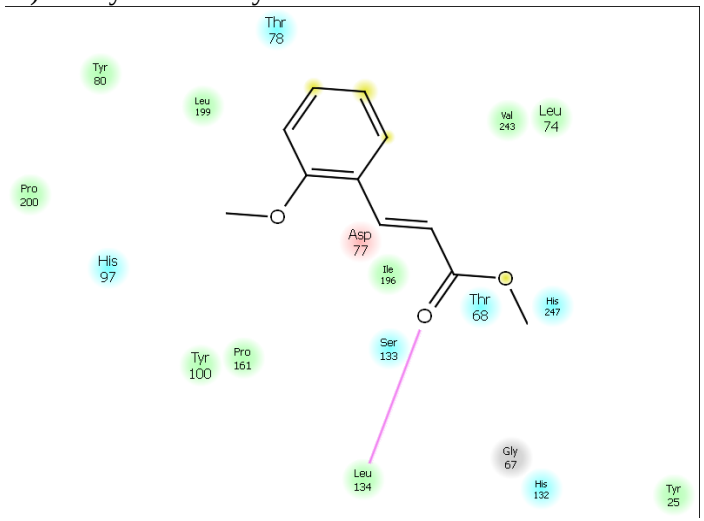
K) Methyl 4-hydroxy cinnamate



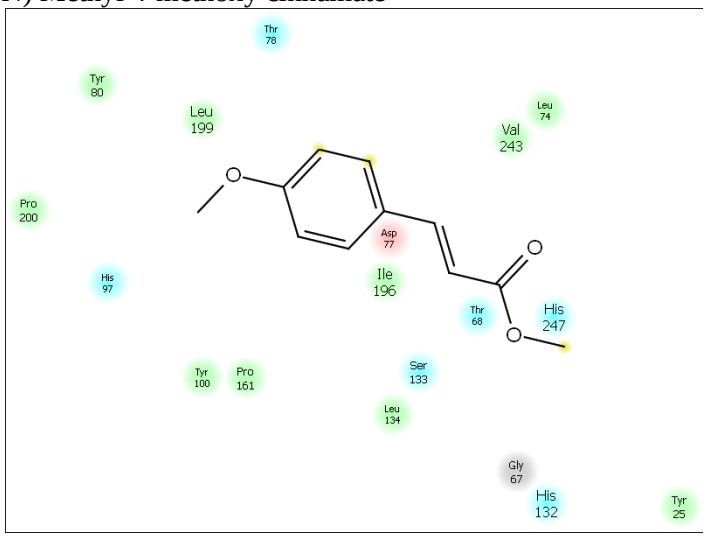
L) Methyl 3,4-dihydroxy cinnamate



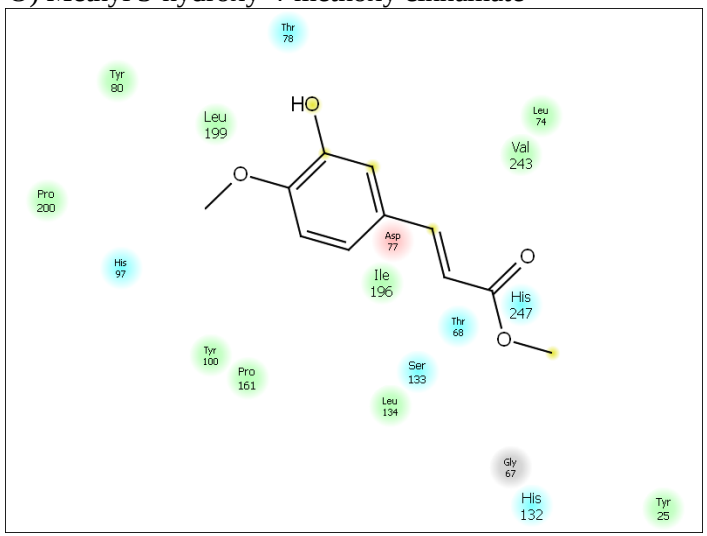
M) Methyl 2-methoxy cinnamate



N) Methyl 4-methoxy cinnamate



O) Methyl 3-hydroxy-4-methoxy cinnamate



Section H

Supplementary Table S5.

Rank-ordering of the substrates based on Glide SP score. The actives and inactives were colored in green and red respectively.

		Glide SP score (kcal/mol)	K_m (mM)
1	Methyl 4-hydroxy-3,5-dimethoxy cinnamate (Methyl sinapate)	-6.15	0.45
2	Methyl 4-hydroxy-3-methoxy cinnamate (Methyl ferulate)	-6.11	0.72
3	Methyl 3,4-dimethoxy cinnamate	-6.05	1.36
4	Methyl 3-hydroxy cinnamate	-5.94	Inactive
5	Methyl 3,5-dimethoxy cinnamate	-5.89	0.92
6	Methyl 3,4,5-trimethoxy cinnamate	-5.79	1.63
7	Methyl 3-methoxy cinnamate	-5.74	1.99
8	Methyl 3-hydroxy-4-methoxy cinnamate	-5.63	Inactive
9	Methyl 4-methoxy cinnamate	-5.50	Inactive
10	Methyl 2-hydroxy cinnamate	-5.39	Inactive
11	Methyl 4-hydroxy cinnamate (Methyl p-coumarate)	-5.33	Inactive
12	Methyl 3,4-dihydroxy cinnamate (Methyl caffeate)	-5.33	Inactive
13	Methyl 4-hydroxy-3-methoxy phenyl propionate	-5.29	2.08
14	Methyl 2-methoxy cinnamate	-4.73	Inactive
15	Methyl cinnamate	-4.68	Inactive

Supplementary Table S6.

Rank-ordering of the substrates based on Glide energy. The actives and inactives were colored in green and red respectively.

		Glide energy (kcal/mol)	K_m (mM)
1	Methyl 3,4-dimethoxy cinnamate	-36.90	1.36
2	Methyl 3,5-dimethoxy cinnamate	-34.55	0.92
3	Methyl 4-hydroxy-3,5-dimethoxy cinnamate (Methyl sinapate)	-34.53	0.45
4	Methyl 3-methoxy cinnamate	-33.98	1.99
5	Methyl 4-hydroxy-3-methoxy cinnamate (Methyl ferulate)	-33.90	0.72
6	Methyl 3-hydroxy-4-methoxy cinnamate	-32.80	Inactive
7	Methyl 2-hydroxy cinnamate	-32.60	Inactive
8	Methyl 3,4,5-trimethoxy cinnamate	-31.93	1.63
9	Methyl 4-hydroxy-3-methoxy phenyl propionate	-31.03	2.08
10	Methyl 3,4-dihydroxy cinnamate (Methyl caffeate)	-30.91	Inactive
11	Methyl 4-methoxy cinnamate	-30.13	Inactive
12	Methyl 3-hydroxy cinnamate	-28.92	Inactive
13	Methyl 2-methoxy cinnamate	-28.50	Inactive
14	Methyl 4-hydroxy cinnamate (Methyl p-coumarate)	-28.19	Inactive
15	Methyl cinnamate	-26.32	Inactive

Supplementary Table S7.

Rank-ordering of the substrates based on the combination of Key Ineraction System and Glide SP score. The actives and inactives were colored in green and red respectively.

		[a] HBI with Thr 68	[a] HBI with Leu 134	Glide SP score (kcal/mol)	K_m (mM)
1	Methyl 4-hydroxy-3,5-dimethoxy cinnamate (Methyl sinapate)	Yes	Yes	-6.15	0.45
2	Methyl 4-hydroxy-3-methoxy cinnamate (Methyl ferulate)	Yes	Yes	-6.11	0.72
3	Methyl 3,4-dimethoxy cinnamate	Yes	Yes	-6.05	1.36
4	Methyl 3,5-dimethoxy cinnamate	Yes	Yes	-5.89	0.92
5	Methyl 3,4,5-trimethoxy cinnamate	Yes	Yes	-5.79	1.63
6	Methyl 3-methoxy cinnamate	Yes	Yes	-5.74	1.99
7	Methyl 4-hydroxy-3-methoxy phenyl propionate	Yes	Yes	-5.29	2.08
8	Methyl 3-hydroxy-4-methoxy cinnamate	No	No	-5.63	Inactive
9	Methyl 2-hydroxy cinnamate	No	No	-5.39	Inactive
10	Methyl 3,4-dihydroxy cinnamate (Methyl caffeate)	No	No	-5.33	Inactive
11	Methyl 4-methoxy cinnamate	No	No	-5.50	Inactive
12	Methyl 3-hydroxy cinnamate	No	No	-5.94	Inactive
13	Methyl 2-methoxy cinnamate	No	Yes	-4.73	Inactive
14	Methyl 4-hydroxy cinnamate (Methyl p-coumarate)	No	No	-5.33	Inactive
15	Methyl cinnamate	No	No	-4.68	Inactive

[a] HBI = Hydrogen Bond Interaction

References

1 Kroon, P. A., Faulds, C. B., Brezillon, C. & Williamson, G. Methyl phenylalkanoates as substrates to probe the active sites of esterases. *European Journal of Biochemistry* **248**, 245-251 (1997).

2 Crepin, V. F., Faulds, C. B. & Connerton, I. F. Production and characterization of the Talaromyces stipitatus feruloyl esterase FAEC in Pichia pastoris: identification of the nucleophilic serine. *Protein Expres Purif* **29**, 176-184 (2003).

3 Vafiadi, C., Topakas, E., Christakopoulos, P. & Faulds, C. B. The feruloyl esterase system of Talaromyces stipitatus: determining the hydrolytic and synthetic specificity of TsFaeC. *J Biotechnol* **125**, 210-221 (2006).

4 Bowie, J. U., Luthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164-170 (1991).

5 Wu, S. & Zhang, Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **35**, 3375-3382 (2007).

6 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738 (2010).

7 Zhang, Y. & Skolnick, J. SPICKER: a clustering approach to identify near-native protein folds. *Journal of computational chemistry* **25**, 865-871 (2004).

8 Castrignano, T., De Meo, P. D., Cozzetto, D., Talamo, I. G. & Tramontano, A. The PMDB Protein Model Database. *Nucleic Acids Res* **34**, D306-309 (2006).

9 Spassov, V. Z., Flook, P. K. & Yan, L. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. *Protein Eng Des Sel* **21**, 91-100 (2008).

10 Spassov, V. Z. & Yan, L. A fast and accurate computational approach to protein ionization. *Protein science : a publication of the Protein Society* **17**, 1955-1970 (2008).

11 Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *J Chem Phys* **105**, 1902-1921 (1996).

12 Spassov, V. Z., Yan, L. & Flook, P. K. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: A side-chain prediction algorithm based on side-chain backbone interactions. *Protein Science* **16**, 494-506 (2007).

13 Momany, F. A. & Rone, R. Validation of the General-Purpose Quanta(R)3.2/Charmm(R) Force-Field. *Journal of Computational Chemistry* **13**, 888-900 (1992).

14 Fletcher, R. & Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput J* **7**, 149-154 (1964).

15 Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507-2524 (2006).

16 Eisenberg, D., Luthy, R. & Bowie, J. U. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in enzymology* **277**, 396-404 (1997).

17 Luthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-85 (1992).

18 Diller, D. J. & Merz, K. M., Jr. High throughput docking for library design and library prioritization. *Proteins* **43**, 113-124 (2001).

19 Diller, D. J. & Li, R. Kinases, homology models, and high throughput docking. *Journal of medicinal chemistry* **46**, 4638-4647 (2003).

20 Rao, S. N., Head, M. S., Kulkarni, A. & LaLonde, J. M. Validation studies of the site-directed docking program LibDock. *J Chem Inf Model* **47**, 2159-2171 (2007).

21 Wu, G., Robertson, D. H., Brooks, C. L., 3rd & Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. *Journal of computational chemistry* **24**, 1549-1562 (2003).

22 Koska, J. *et al.* Fully automated molecular mechanics based induced fit protein-ligand docking method. *J Chem Inf Model* **48**, 1965-1973 (2008).

23 Halgren, T. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry* **47**, 1750-1759 (2004).

24 Friesner, R. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **47**, 1739-1749 (2004).

25 Friesner, R. A. *et al.* Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry* **49**, 6177-6196, doi:Doi 10.1021/Jm051256o (2006).

26 Glide, v., Schrödinger, LLC, New York, NY, 2011.

27 Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aid Mol Des* **11**, 425-445 (1997).

28 Sherman, W., Beard, H. S. & Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des* **67**, 83-84 (2006).

29 Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry* **49**, 534-553 (2006).

30 Schrödinger Suite 2011 Induced Fit Docking protocol; Glide version 5.7, S., LLC, & New York, N., 2011; Prime version 3.0, Schrödinger, LLC, New York, NY, 2011.
31 Prime, v., Schrödinger, LLC, New York, NY, 2011.
32 Molecular Operating Environment (MOE), C. C. G. I., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2010.
33 Edelsbrunner, H. W. A. S., Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois 61810.
34 Bohm, H. J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *J Comput Aid Mol Des* **8**, 243-256 (1994).
35 Bohm, H. J. Ludi - Rule-Based Automatic Design of New Substituents for Enzyme-Inhibitor Leads. *J Comput Aid Mol Des* **6**, 593-606 (1992).
36 Kramer, B., Rarey, M. & Lengauer, T. CASP2 experiences with docking flexible ligands using FLEXX. *Proteins-Structure Function and Genetics*, 221-225 (1997).
37 Kramer, B., Rarey, M. & Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins-Structure Function and Genetics* **37**, 228-241 (1999).
38 Rarey, M., Kramer, B. & Lengauer, T. Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. *J Comput Aid Mol Des* **11**, 369-384 (1997).
39 Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology* **261**, 470-489 (1996).
40 Rarey, M., Wefing, S. & Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J Comput Aid Mol Des* **10**, 41-54 (1996).
41 Thomsen, R. & Christensen, M. H. MolDock: a new technique for high-accuracy molecular docking. *Journal of medicinal chemistry* **49**, 3315-3321 (2006).
42 Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs; Springer-Verlag: Berlin* (1992).
43 Michalewicz, Z. F., D. B. *How to Solve It: Modern Heuristics; Springer-Verlag: Berlin* (2000).
44 Storn, R. P., K. *Differential Evolution - A Simple And Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. Tech-report, International Computer Science Institute, Berkley* (1995).
45 Nelder, J. A. & Mead, R. A Simplex-Method for Function Minimization. *Comput J* **7**, 308-313 (1965).
46 Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry* **46**, 499-511 (2003).
47 SYBYL-X. *BioPharmics' Surflex-Dock Manual, Docking Suite Manual, Tripos - A certara company, USA* (2011).
48 Jain, A. N. Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *J Comput Aid Mol Des* **23**, 355-374 (2009).
49 Mayo, S. L., Olafson, B. D. & Goddard, W. A. Dreiding - a Generic Force-Field for Molecular Simulations. *J Phys Chem-Us* **94**, 8897-8909 (1990).
50 Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M. & Waldman, M. LigScore: a novel scoring function for predicting binding affinities. *Journal of Molecular Graphics & Modelling* **23**, 395-407 (2005).
51 Gehlhaar, D. K. *et al*. Molecular Recognition of the Inhibitor Ag-1343 by Hiv-1 Protease - Conformationally Flexible Docking by Evolutionary Programming. *Chem Biol* **2**, 317-324 (1995).
52 Gehlhaar, D. K., Bouzida, D. & Rejto, P. Reduced Dimensionality in Ligand—Protein Structure Prediction: Covalent Inhibitors of Serine Proteases and Design of Site-Directed Combinatorial Libraries. *Rational Drug Design* **ACS Symposium Series, Vol. 719**, 292–311 (1999).
53 Jain, A. N. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J Comput Aid Mol Des* **10**, 427-440 (1996).
54 Muegge, I. & Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *Journal of medicinal chemistry* **42**, 791-804 (1999).
55 Muegge, I. PMF scoring revisited. *Journal of medicinal chemistry* **49**, 5895-5902, doi:10.1021/jm050038s (2006).
56 Reulecke, I., Lange, G., Albrecht, J., Klein, R. & Rarey, M. Towards an integrated description of hydrogen bonding and dehydration: Decreasing false positives in virtual screening with the HYDE scoring function. *Chemmedchem* **3**, 885-897 (2008).
57 Yang, J. M. & Chen, C. C. GEMDOCK: A generic evolutionary method for molecular docking. *Proteins-Structure Function and Bioinformatics* **55**, 288-304 (2004).
58 Korb, O., Stutzle, T. & Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J Chem Inf Model* **49**, 84-96 (2009).
59 Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267**, 727-748 (1997).
60 Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *Journal of Molecular Biology* **161**, 269-288 (1982).
61 Fiser, A., Do, R. K. G. & Sali, A. Modeling of loops in protein structures. *Protein Science* **9**, 1753-1773 (2000).
62 Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M. & Pace, C. N. pK values of the ionizable groups of proteins. *Protein science : a publication of the Protein Society* **15**, 1214-1218 (2006).

63 Bashford, D. & Karplus, M. Multiple-Site Titration Curves of Proteins - an Analysis of Exact and Approximate Methods for Their Calculation. *J Phys Chem-Us* **95**, 9556-9561 (1991).
64 Schrödinger Suite 2011 Schrödinger Suite; Epik version 2.2, S., LLC, New York,, NY, I. v., Schrödinger, LLC, New York, NY, 2011; Prime version 2.3, & Schrödinger, L., New York, NY, 2011.