Chapter 4

Virtual Ligand Screening of Chemical Libraries for Mouse MrgC11 Receptor: Combination of QSPR and Docking Methods[1]

4.1 Introduction

High-throughput screening (HTS) of chemical libraries is the widely adopted method for finding novel lead compounds in drug discovery. It enables a large number of compounds to be screened using highly automated, robotic techniques. Although HTS makes it possible in principle to test all available compounds, it is not necessarily feasible for a number of practical reasons. One of reasons is the cost of such screenings: even though the robotics and miniaturization have significantly reduced the unit cost, the huge number of compounds now available from many companies means that the overall expense can be significant. Moreover, as the available databases get larger and larger, the hit rates in HTS dramatically decrease. A possibility to avoid these problems is not to screen the whole compound set in the library experimentally, but only a small subset, which is likely to bind to the target protein receptor. This pre-selection can be performed by virtual screening (VS), which uses computer-based methods to select most promising compounds from the ligand databases for experimental assays. Virtual screening can be carried out by searching databases for molecules fitting either a known pharmacophore (ligand-based) or a three-dimensional structure of macromolecular target (structure-based). In the case of GPCRs, the limited availability of the structural data has forced the computational design of ligands to heavily rely on ligand-based drug design techniques.

[1] This work was carried out in collaboration with the Tropsha group of the University of North Carolina.

Indeed, the natural ligands can provide a good starting point, leading to useful pharmacophore models that can be used for virtual screening to identify lead structures with novel scaffolds[1]. The application of this method has been successfully demonstrated in the discovery of subtype selective agonists to the somatostatin receptor[2] and non-peptide antagonists to the urotensin II receptor[3]. Structure-based screening should be potentially more powerful than the ligand-based method since by exploiting structural information taken directly from the active site, it is possible to discover ligands with both diverse chemotypes and binding modes. However, it still suffers from docking/scoring inaccuracy, and in addition it requires the knowledge of the 3D structure of the target protein. Therefore, it has mostly been applied to targets for which a high resolution X-ray crystal structure is known. However, along with the deciphering of human genome, computational chemists are facing an overwhelming number of potential targets for which very little experimental 3D information is available. Therefore it will be very important in the near future to be able to use not only X-ray or NMR structures, but also protein models for structure-based virtual screening of chemical libraries.

The structure-based virtual screening mainly relies on a fast and accurate docking/scoring function that can be used to identify the correct binding mode. Theoretically, the most accurate estimate of the binding affinity can be obtained using force-field based methods. Examples include free energy perturbation (FEP)[4] or linear interaction energy (LIE) approaches[5]. However, the computational cost of such methods is too high to afford calculation in a high-throughput fashion. Therefore the huge chemical libraries should be filtered through a rapid pre-screening tool to identify the most promising compounds prior to engaging more computationally intensive docking approaches. The ligand-based similarity searching technique could be used for this purpose. In this approach, the ligand structures are typically represented by multiple chemical descriptors and the statistical data modeling techniques are used to establish quantitative correlation between descriptors and target properties of interest, such as binding constants or

specific biological activities[6]. Recently the Tropsha group in the University of North Carolina had developed a novel structure-based chemoinformatics approach to search for complimentary ligands based on receptor information (CoLiBRI)[7]. CoLiBRI is based on a representation to characterize both receptor active sites and their corresponding ligands in the same universal, multidimensional, chemical descriptor space. Mapping of both binding pockets and corresponding ligands onto the same multidimensional chemistry space would preserve the complementarity relationships between the binding sites and their respective ligands.

In this study, we carried out virtual screening for the mouse MrgC11 receptor, one of orphan GPCR receptors as an effort to identify small molecule ligands that behave as selective agonists or antagonists. Despite of the success of orphan GPCR-natural ligand pairing through reverse pharmacology many scientists focused on discovering new drugs appear to be bypassing the conventional deorphanizing step due to the difficulty in developing peptide libraries to look for the ligand. They perform initial high-throughput assays to find synthetic small-molecule agonists, which then can be used to explore the physiological aspects of the receptor. Here we first pre-screened compounds in the chemical database using the CoLiBRI and the resulting candidates were subsequently docked using the MSCDock method. The ‘hit’ compounds from docking were experimentally tested with the intracellular calcium release assay. In the following sections, we describe the computational methods in details and discuss the screening results.

4.2 Materials and methods

4.2.1 Pre-screening of compounds in chemical libraries

Pre-screening the compounds of the chemical libraries was carried out using the CoLiBRI program. CoLiBRI is based on the quantitative structure-property relationship (QSPR) method. It generates the molecular descriptors that capture key properties of the molecules, using the transferable atom equivalent (TAE)/RECON method. The TAE/RECON method that was developed by Breneman and co-workers[8] rapidly generates molecular electron density

Table 4.1 Electron-density-derived TAE descriptors; $\rho(r)$ represents the electron density distribution[9]

Integral Electronic Properties		
Energy		
Electronic population		
Volume		
Surface area		
Surface electronic properties (extrema, surface integral averages and histogram bins are available for each property)		
SIEP	Surface integral of electrostatic potential	
EP	Electrostatic potential	$EP(r) = \sum_{\alpha} \frac{Z_{\alpha}}{\lvert r - R_{\alpha} \rvert} - \int \frac{\rho(r')dr'}{\lvert r - r' \rvert}$
DRN	Electron density gradient normal to 0.002 e/au^3 electron-density isosurface	$\nabla\rho \cdot n$
G	Electronic kinetic energy density	$G(r) = -(1/2)(\nabla\psi^* \cdot \nabla\psi)$
K	Electronic kinetic energy density	$K = -(1/2)(\psi^*\nabla^2\psi + \psi\nabla^2\psi^*)$
DKN	Gradient of the *K* electronic kinetic energy density normal to surface	$\nabla K \cdot \boldsymbol{n}$
DGN	Gradient of the *G* electronic kinetic energy density normal to surface	$\nabla G \cdot \boldsymbol{n}$
F	Fukui F^+ function scalar value	$F^+(r) = \rho_{HOMO}(r)$
L	Laplacian of the electron density	$L(r) = -\nabla^2\rho(r) = K(r) - G(r)$
BNP	Bare nuclear potential	$BNP(r) = \sum_{\alpha} \frac{Z_{\alpha}}{\lvert r - R_{\alpha} \rvert}$
PIP	Local average ionization potential	$PIP(r) = \sum_{i} \frac{\rho_i(r)\lvert\varepsilon_i\rvert}{\rho(r)}$

distributions and evaluates the electronic surface properties, which are used for generating descriptors. It contains a library of the atomic types in a form which can transfer electron density properties. The RECON program reconstructs the electronic density properties of a molecule by assigning the closest match from a library of atom types for each atom in the molecule. The additivity principle is applied to calculate molecular descriptors by summing up the individual descriptor type values for all atoms in the molecule, using the RECON method. Therefore it is possible to derive pseudo-molecular descriptors for any group of atoms, e.g., active site fragment, making the TAE descriptors well suited for our approach. Table 4.1 shows a complete list of TAE descriptors. The local average ionization potential (PIP) of the molecule, one example of the electronic surface properties is shown onto its 0.002 e/au^3 (electrons per cubic Bohr) electron-

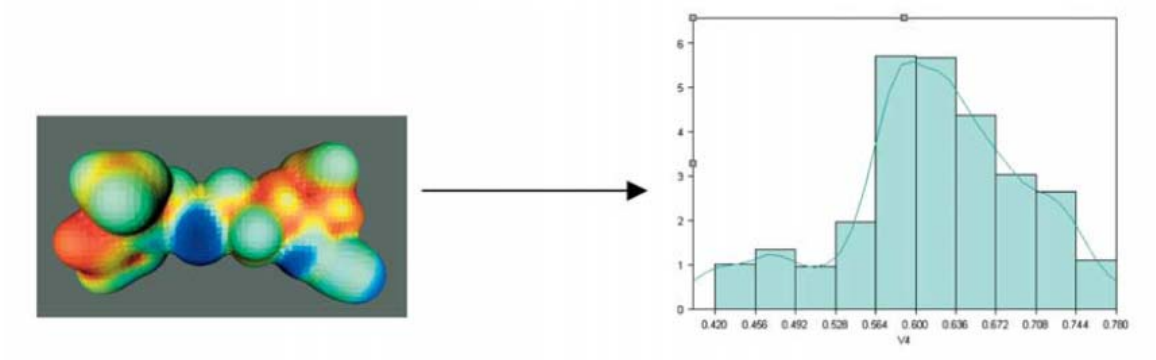


Figure 4.1 TAE local average ionization potential (PIP) surface property and its histogram distribution[9].

density surface in Figure 4.1. The distribution of this property is then presented as a histogram such as that shown on the right side of the figure. Each bin of the histogram is used as a descriptor, as well as statistical information such as maximum, minimum, and average of each surface property.

A computational geometry technique known as Delaunay tessellation is utilized to isolate receptor atoms that make contacts with bound ligands. Let us consider a collection of randomly distributed points in 2D (Fig. 4.2). By analogy, the red and blue dots represent the ligand atoms and the receptor atoms in the binding site, respectively. Delaunay tessellation partitions the space occupied by these points into a set of space filling, irregular triangles (tetrahedrons in 3D) with the original points as vertices. Therefore this method identifies all nearest neighbor triplets of vertices, including two types of interfacial triplets as shown in Figure 4.2: one ligand atom point and two receptor atom points; two ligand atom points and one receptor atom point. Applied to the 3D receptor-ligand complex case, it will generate three types of interfacial quadruplets: one ligand atom and three receptor atoms; two ligand atoms and two receptor atoms; three ligand atoms and one receptor atom. Therefore it provides a way of detecting all receptor atoms that are nearest neighbors of ligand atom. The TAE descriptors are then generated for a pseudo-molecule composed of these receptor atoms.

Using the TAE/RECON method, multiple descriptors as listed on Table 4.1 are generated for the receptor binding sites and their corresponding ligands so that each chemical entity is

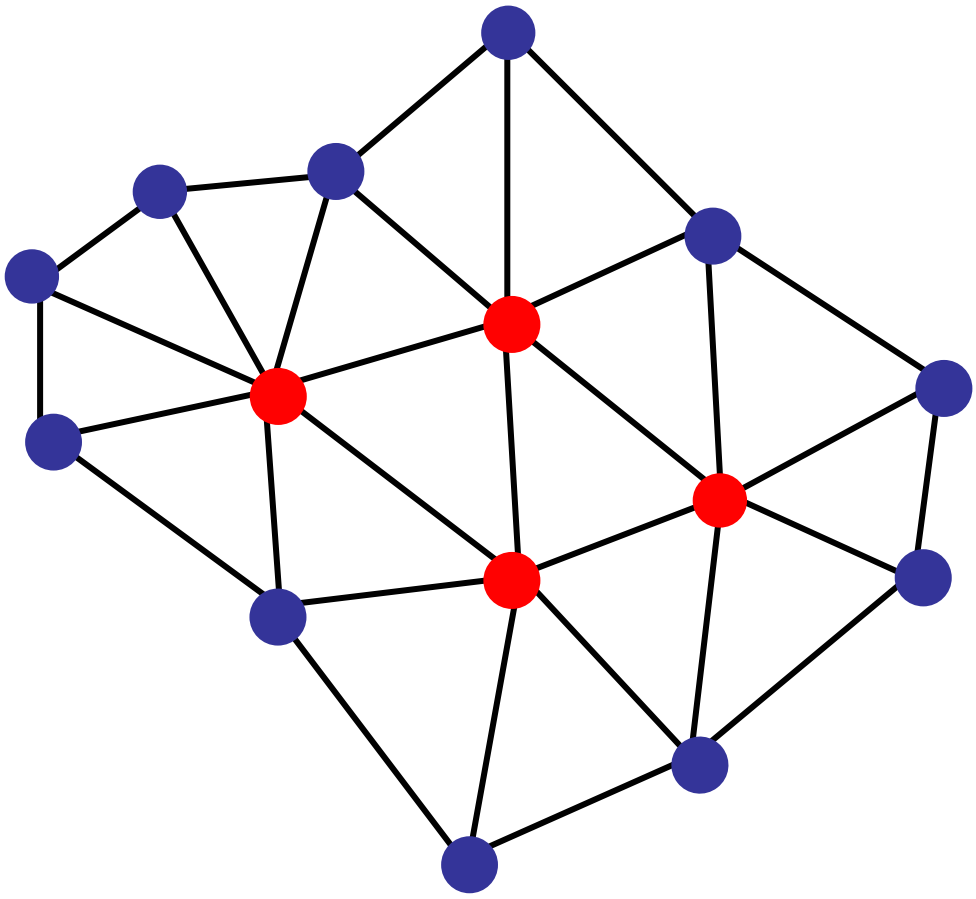


Figure 4.2 Delaunay tessllation of a collection of random points in 2D (modified from reference7)

represented as a vector in a multidimensional TAE/RECON chemical space. Since every descriptor may not be important for determining receptor-ligand complementarity, the subset of descriptors that best reflect this complementarity is determined, using a leave-one-out (LOO) cross-validation approach, in which each data value is left out in turn and a model derived using the remainder of the data. The overall procedure for selecting an optimal subset is as follows:

(1) A subset of n_{Var} descriptors (n_{Var} is a predefined number between 1 and the total number of available descriptor) is randomly selected.

(2) One of the receptors is chosen in the training set and the k nearest neighboring (kNN) receptors are selected in the n_{Var}-dimensional descriptor

space of the binding site. The coordinates of the chosen receptor's virtual ligand in the ligand space are predicted based on the relative orientation of ligands known to bind with the kNN receptors. This step is repeated until every receptor in the training set is eliminated once and all the receptor's virtual ligands are predicted. This resulting set of virtual ligands is called a CoLiBRI model.

(3) The predictive mean rank (PMR) for the model is calculated. It is related to the chemical similarity of the virtual ligands to the known ligands. The similarities are evaluated as Euclidean distances in the n_{Var}-dimensional descriptor space:

$$Dist_{i,j} = \sqrt{\sum_{d=1}^{n_{Var}} (X_{id} - X_{jd})^2}\ , \qquad \text{(Eq. 4.1)}$$

where X_{id} and X_{jd} are the dth selected descriptor for ligand i and j. The higher rank means the larger deviation of the model.

(4) Step 2 and 3 are repeated for all possible k values ($2 \le k \le$ total number of ligand - receptor pairs). The k values that leads to the lowest PMR value is chosen as optimal.

(5) The selection of n_{Var} descriptors is optimized based on simulated annealing. For a model built using randomly-sampled n_{Var} descriptors, the value of the fitness function, the inverse of its PMR value is calculated. By changing a fraction of the currently used descriptors to other randomly selected of n_{Var} descriptors, a new CoLiBRI model is generated for the new trial set (repeat steps 1 to 4) and the new corresponding fitness function is calculated. The new trial set is accepted or rejected based on the Metropolis criterion. This

Monte Carlo approach is continued as the temperature is lowered until the termination condition is satisfied.

At the end, both an optimum *k* value and an optimal subset of n_{Var} descriptors are determined and produce a model with the best predictive ability. More detailed mathematical expression is described in reference 7.

Now the CoLiBRI model is ready to be used for the ligand screening. First the target receptor is positioned in the selected descriptor subspace and its *k* nearest neighboring receptors from the training set are found. The known ligands of these *k* nearest neighboring receptors are then used to estimate the location of the target receptor's virtual ligand in the descriptor space in the same way as step 2 above. All ligands in the chemical library are ranked based on their distance to this predicted virtual ligand point (using Eq. 4.1), and the ligands with the smallest distance are considered as the most probable hit.

In our study the CoLiBRI models were generated for the dipeptide binding site using the same training set (670 complex structures from PDBbind[10]) used in reference 7 plus the predicted mMrgC11/R-F-OH complex structure.

4.2.2 Chemical libraries

Three sets of chemical libraries were screened in this study;

(1) The first set: An older version of the database from ChemDiv with 451,345 compounds was pre-screened using the CoLiBRI method. The multiple CoLiBRI models that predict complementarity were generated, varying the n_{Var} value, a number of selected descriptors used in generating a CoLiBRI model as described in section 2.1. The compound within the top 1,000 by at least one model was selected and total 3,900 compounds were collected for the next docking step.

(2) The second set: It was taken from a newer version (fall, 2004) of the database from ChemDiv with 513,000 compounds. We selected compounds that were consistently predicted to be within the top 1,000 by all models. This resulted in 442 hits.

(3) The third set: The 23 drug compounds known for producing pain relief were docked without any pre-screening. It includes some opiates (e.g. Demerol), local anesthetics (e.g. Lidocaine) and capsaicin (an agonist of vanilloid receptors in dorsal root ganglion (DRG)). All possible protonation states were considered, leading to a total of 43 ligand structures for docking.

For the pre-screened compounds from the first and second set, hydrogen atoms were added and Gasteiger charges[11] were assigned using Concord program. No further optimization was carried out before docking. For the third ligand set, Gasteiger charges were assigned and the structures were optimized in gas phase using conjugate gradient minimization using the DREIDING force field (FF)[12] on Cerius2[13].

The pre-screening of ChemDiv database for the di-peptide binding site was performed in collaboration with the Tropsha group of the University of North Carolina.

4.2.3 Molecular docking

MSC-Dock program was used for docking the pre-screened ligands. We used the Dock-Diversity Completeness protocol (DDCP). As described in chapter 2, DDCP attempts to generate a complete set of ligand configuration families with a fixed coordinate diversity. In this study the diversity was set to 0.6 Å. The rejection ratio (defined as the fraction of new configuration that belongs to previously generated families to the fraction that leads to a new family) was set to 2.2. The 50 families were selected with the best energies (by DOCK4.0 energy score) in the first phase and an average of six members in each family was generated in the second enrichment

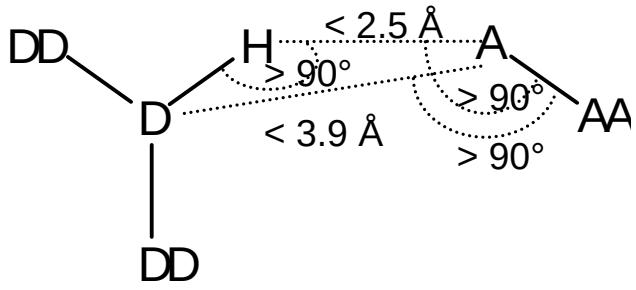


Figure 4.3 Geometric criteria for the hydrogen bonds. D is the donor heavy atom, H the hydrogen, A the acceptor, DD donor antecedent (i.e. an atom two covalent bonds away from the hydrogen) and AA acceptor antecedent.

phase. The final ~300 configurations were ordered by DOCK4.0 energy score and re-clustered with 0.6 Å of diversity to generate a new set of families. The top 5 family heads (a member with the best energy in each family) were conjugate gradient minimized (100 steps or 0.1 kcal/mol/Å of RMS force) with the ligand atoms movable and the receptor atoms fixed. Then the binding energies were then calculated for these 5 optimized ligand-receptor complex configurations. The calculated binding energy (BE) is defined by

BE = E (ligand in fixed protein) – E (ligand in water),

where the E (ligand in fixed protein) is the potential energy of the ligand calculated in the ligand-receptor complex with the coordinates of the receptor fixed. This potential energy includes the internal energy of the ligand and the interaction energy of the ligand with the receptor. E (ligand in water) is the potential energy of the free ligand in its docked conformation (snap bind energy) and its solvation energy calculated using the analytical volume generalized born (AVGB) continuum solvation method (cavity_params_1.3)[14]. The final best ligand-receptor structure was selected as the one with the most negative binding energy.

4.2.4 Selection of final hits

The ligands in the final docked conformation were sorted by three criteria; the binding energy, the van der Waals interaction energy and the energy of hydrogen bond between the

receptor and the ligand. The intermolecular hydrogen bond was determined by the geometric criteria shown in Figure 4.3[15] and its energy was evaluated using the DREIDING FF. For the first ligand set, the top 100 ligand compounds were chosen by each sorting criterion. Then we selected the compounds that were consistently within the top 100 by at least two criteria. This led to total 52 compounds. These selected compound structures were further optimized in the protein-ligand complex. The side chain conformation of receptor residues within 5 Å from the ligand was optimized using the SCREAM program and then the entire receptor-ligand complex structure was conjugate gradient minimized with 0.1 kcal/mol/Å of RMS force. This receptor-ligand complex was further refined using one cycle of annealing MD heating from 50 K to 600 K and cooling down back to 50 K in 50 K steps, with 1 ps of equilibration between temperature jumps. Here only the ligand and the receptor side chains within 5 Å of the binding pocket were allowed to move during the annealing cycle. At the end of the annealing cycle, the system was minimized to an RMS force of 0.3 (kcal/mol)/Å. The binding energy was then re-calculated for the final complex structure in the same way as described above.

The compounds in the second set were also sorted by three same criteria and the common compounds within the top 40 were selected. The 40th best binding energy is the halfway between the highest one and zero. This resulted in 21 compounds, which were optimized further as in the first set.

The pain-related compounds in the third set were sorted by their binding energy and the top 10 compounds were chosen, then the same post-optimization was carried out.

Both the protein and the ligand were described using the DREIDING FF and the protein charges were from CHARMM22[16]. All calculations used the MPSIM program[17], with nonbond interactions evaluated using the cell multipole method[18]. All simulations were performed in gas phase with the dielectric constant of 2.5.

After post-optimization, the residues of the receptor having either an intermolecular hydrogen bond or good van der Waals contact with the ligand were identified. By putting the priority on compounds having good contacts with the key residues—Tyr110 (TM3), Asp161 (TM4) and Asp179 (TM5)—26 compounds were finally chosen for experimental test. They included four outliers in the docking step to expand diversity and two pain-related compounds, capsaicin and ibuprofen.

4.2.5 Intracellular calcium release assay

The intracellular calcium release assay experiment was carried out to test activity for 26 compounds (the details are described in chapter 2). One of the known peptide agonists, F-M-R-F-NH_2 (EC50 = 168nM) was used as a control compound. To test agonistic activity, cells expressing stably mMrgC11 receptor proteins were treated with compounds in two different concentrations, 100 μM and 10 μM. To check antagonistic activity, cell sample was pre-incubated for >5min with a compound in 100 μM and 10 μM concentration and then were treated with 1 μM of F-M-R-F-NH_2. The inhibitory constant 50% (IC50), the concentration reducing the activity of 400 nM F-M-R-F-NH_2 by half was measured for the compounds showing the antagonistic effect in two ways. First, cells were pre-incubated with a compound in various concentrations and F-M-R-F-NH_2 was added later. Secondly, the compound was added to the cell sample together with F-M-R-F-NH_2 at the same time and the intracellular calcium release was measured.

4.2.6 Virtual screening of tetra-peptide binding site

The virtual screening for the tetra-peptide binding site was independently carried out in a similar way. Since the loops were in the ensemble of conformations as shown in chapter 3, the extracellular loops in the mMrgC11 receptor were not included in screening. The dataset of 800 ligand-receptor complexes from the PDBbind Database (PDB entry codes are listed in the

supporting information of reference 7) was divided into the training (used for model building; 525 structures) and the test (used for model validation; 275 structures) sets using the sphere exclusion method[19]. In building CoLiBRI models, six predicted Mrg complex structures were included in the training set; mMrgC11/(D)F-M-R-F-NH_2, mMrgC11/F-M-R-F-NH_2, mMrgC11/F-(D)M-R-F-NH_2, mMrgC11/R-F-NH_2, mMrgC11/R-F-OH and rat MrgA/adenine complex. The CoLiBRI models differ depending on the number of descriptors (4 to 40) and the content of a given number (10 content variations). Among these 370 models the top 100 models were chosen based on the PMR values for the test set of 275 receptors.

The first set of chemical library used in the previous dipeptide case was screened for the mMrgC11 receptor optimized with the bound F-(D)M-R-F-NH_2, which is the best known tetra-peptide agonist. Five F-M-R-F-NH_2 peptides (three agonists and two non-agonists), R-F-NH_2 and R-F-OH were included into the ChemDiv database, leading to total 451,352 compounds. The top 1,000 compounds were selected for each model. The models having (D)F-M-R-F-NH_2, F-M-R-F-NH_2 and F-(D)M-R-F-NH_2 as a hit after screening were identified, resulting in 92 out of 100 models. The 4,735 compound hits from the ChemDiv database were predicted by at least one of 92 models and the 16 compound hits were consistently predicted by all 92 models. However F-M-(D)R-F-NH_2 was also consistently recognized as a hit for all 92 models (false positive), indicating that the CoLiBRI model is not sensitive enough to completely distinguish between the chirally modified tetrapeptide agonists and non-agonists. Nevertheless identification of three agonists as hits provides some validation of the CoLiBRI models used in this study.

The 774 compound hits which were consistently predicted by at least 50 models were chosen for the next docking step. We also used MSC-Dock with the same parameters except for the diversity of 1.0 Å since the size (number of atoms) of hit compounds in the tetra-peptide binding site is larger than those in the di-peptide binding site.

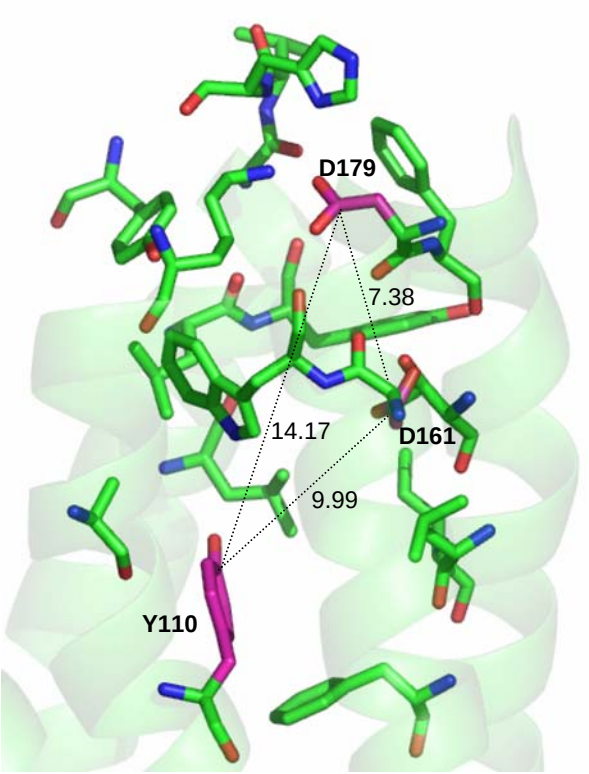


Figure 4.4 5 Å binding pocket of mMrgC11 receptor optimized with the di-peptide agonist, R-F-OH. Three key residues (Y110, D161 and D179) are identified and inter-residue distances are specified in Å for those residues.

Following the same scoring method and selection criteria (the top100 were selected for each criterion – binding energy, van der Waals interaction and hydrogen bond energy (the calculated binding energy = -41.77 to 431.47 kcal/mol; the 100th is approximately halfway between -41.77 to 0)), final 55 compounds were identified out of 774. Then these 55 complex structures were optimized in the same way as described section 2.4.

We docked F-(D)M-R-F-NH_2 with the same docking parameters and scoring method. The RMSD of the best configuration was 0.29 Å with respect to the previously predicted "true" bound configuration, validating our docking procedure.

4.3 Results and discussion

4.3.1 Hit compounds from virtual screening

Figure 4.4 shows the 5 Å binding site of the mMrgC11 receptor complexed with one of dipeptide agonists, R-F-OH. This dipeptide optimized structure was used for both pre-screening and docking. Three key residues were previously identified in the R-F dipeptide binding. Tyr110 had a good π-π interaction with F of the dipeptide, and two Asp residues, Asp161 and Asp179 interacted favorably with the sidechain of R and the N-terminus. The final hit compounds after virtual screening were listed in Figure S4.1 for the first ligand set and in Figure S4.2 for the second one. The ligand atoms forming hydrogen bonds with the receptor were specified. The contribution of each receptor residue to the van der Waals interaction was evaluated and the residues for which the absolute value of the interaction energy was larger than 3 kcal/mol were identified. Most of ligands had at least one aromatic ring, which replaced the phenyl ring of R-F dipeptide and interacted with nonpolar residues present inside the pocket such as Tyr110, Phe190 and Leu186. Some of ligands formed a hydrogen bond with Asp161 or/and Asp179, but none of the hydrogen bond partners were similar to the arginine sidechain.

By comparing the hit compounds from the first set with those from the second set, we could see that selection of the compounds consistently predicted by all CoLiBRI models provided a ligand with the higher binding energy showing better chemical contacts (i.e. contacts with all key residues) although the hit compounds showed less diversity. MOL282 (the ligand with the best binding energy in the second set) showed better binding by 6 kcal/mol than Mol2190 (the best one in the first set) and made contacts with Tyr110, Asp161 and Asp179.

Among the pain-related compounds, capsaicin and ibuprofen showed the best binding energy in docking. The binding energies were -45.11 and -43.14 kcal/mol respectively. The van der Waals interaction mainly contributed to the binding energy. Capsaicin formed a single hydrogen bond with Asp161 and ibuprofen does not have any contact with three key residues.

4.3.2 Experimental activity test

Table 4.2 Inhibitory constant 50% (IC50) of hit compounds (unit: μM)

	A	B
MOL282	46.5 ± 2.2[a]	74.6 ± 0.1[b]
capsaicin	26.0 ± 2.7[a]	N.A.
capsazepine	19.2 ± 5.9[b]	N.A.
dihydrocapsaicin	46.6	N.A.
N-vanillylnonamide	69.7 ± 17.7[b]	N.A.

A – pre-incubate a compound and then add 400 nM of F-M-R-F-NH_2, B – add a compound and 400 nM of F-M-R-F-NH_2 at the same time.
[a] mean ± SEM from triplicate independent measurements, [b] duplicate measurements
N.A.: no significant decrease in activity of F-M-R-F-NH_2 agonist is observed in >200 μM concentration.

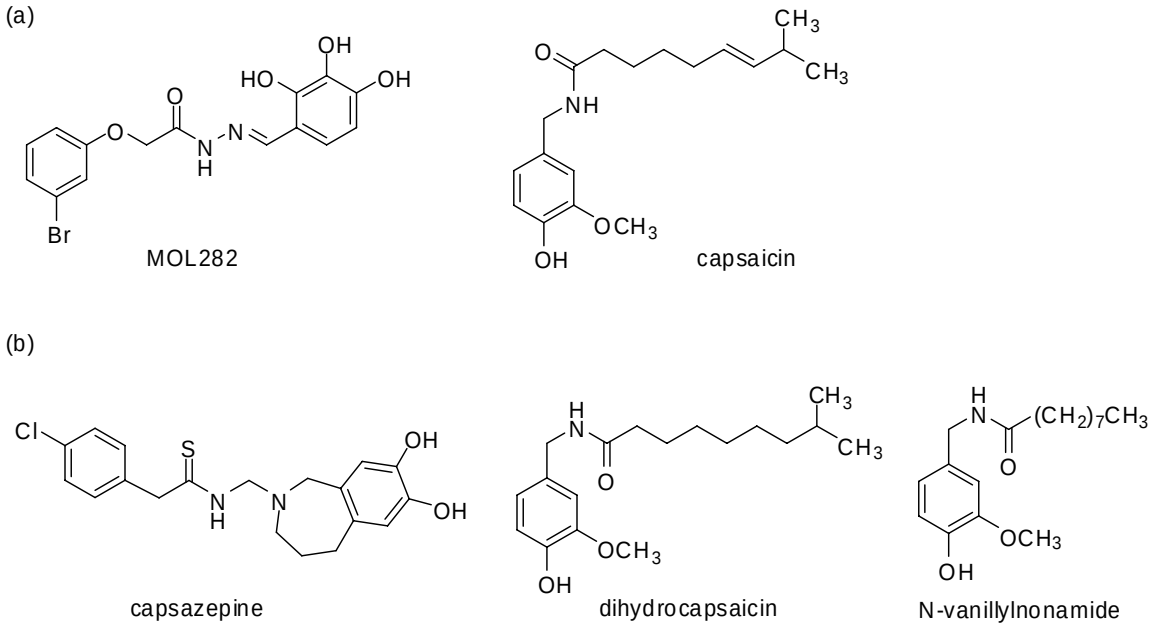


Figure 4.5 Compounds showing the inhibitory effect (a) from the hit compound set of VLS and (b) among the tested capsaicin analogs.

The agonistic activity for total 26 compounds (24 from the virtual screening plus capsaicin and ibuprofen) were tested using the intracellular calcium assay. The mMrgC11 receptor was activated by none of them up to 100 μM concentration. However some of them showed the inhibitory effect – blocking the activity of the known agonist, F-M-R-F-NH_2. Two compounds, MOL282 and capsaicin shown in Figure 4.5(a) blocked the activity of F-M-R-F-NH_2. The measured IC50s of MOL282 are 47 μM for pre-incubation case and 75 μM for simultaneous addition (Table 4.2). It means that MOL282 binds to the mMrgC11 receptor kinetically at the rate

comparable to F-M-R-F-NH_2. However capsaicin could not block the activity of the agonist when it was added together with the agonist at the same time, indicating that it is a slow binder than F-M-R-F-NH_2.

MOL282 was predicted to have the best binding energy from our virtual screening, and this experimental result provides the strong evidence that our predicted mMrgC11 structure is accurate enough to screen chemical libraries for potential ligands. Capsaicin is a well-known agonist of vanilloid receptor type 1 (VR1), which functions as a molecular integrator of painful chemical and physical stimuli[20]. Although Dong *et al.* claimed that mMrgAs and mMrgD were expressed in the $VR1^-$ sensory neurons[21], we could observe that capsaicin was able to inhibit the activity of a known agonist in the mMrgC11 receptor. Next we extended the experiment to capsaicin analogs, and five commercially available analog compounds were tested (capsazepine, dihydrocapsaicin, olvanil, N-vanillylnonamide and eugenol). Among five, three compounds showed antagonistic effect at the tens micromolar concentration. Their chemical structures are shown in Figure 4.5(b).

4.3.3 Refined docking of MOL282 and design of its derivatives

We docked the lead compound, MOL282 again into the mMrgC11 receptor in a more refined docking scheme. The conformations of MOL282 were extensively explored using the grid sampling method. Five torsion degrees of freedom were sampled by 60° steps from the initial optimized structure, leading to total 7,776 conformations. These conformations were ranked by the force field energy in gas phase and clustered with 1.0 Å of diversity. This resulted in the set of final 87 conformations. Each conformation was docked independently into the same binding region without further optimization.

The MSC-Dock with DDCP was used for docking as described in section 2.3. Here the top 25 families (instead of 5) were chosen and optimized with the receptor coordinates fixed. They were ranked by binding energy and then the top 10 configurations were determined. These 10

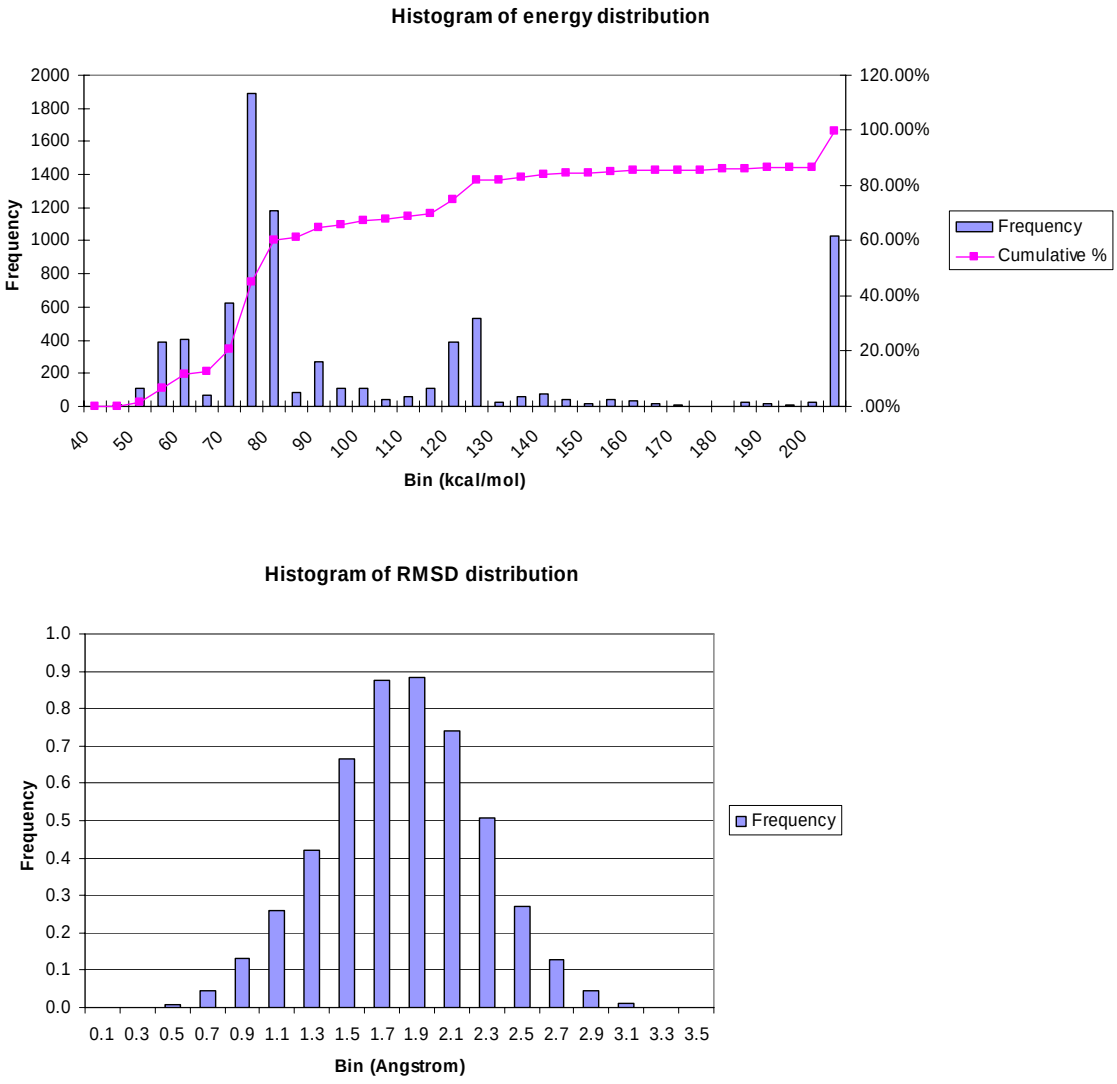


Figure 4.6 Histograms of energy and RMSD distribution for 7,776 conformations of MOL282 in grid search. The pair-wise RMSD is calculated with heavy atoms only.

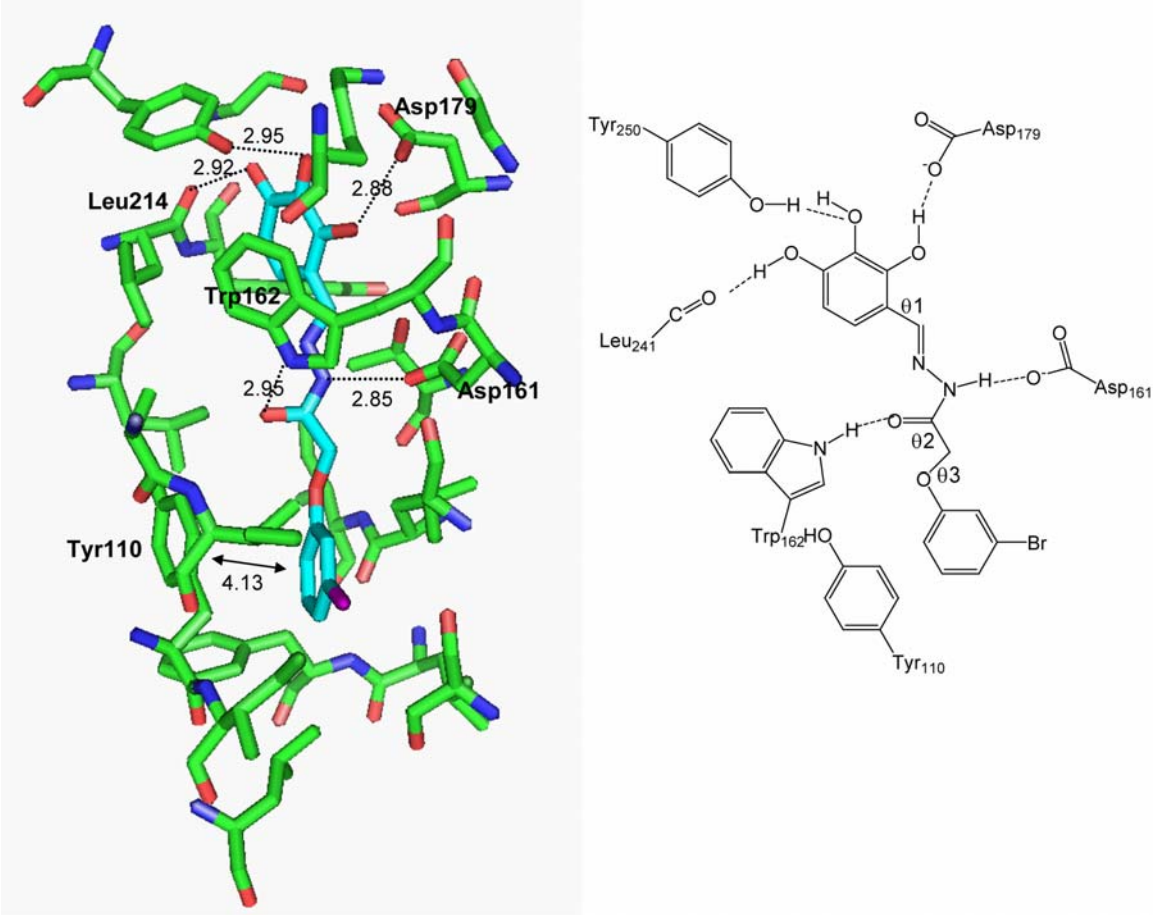
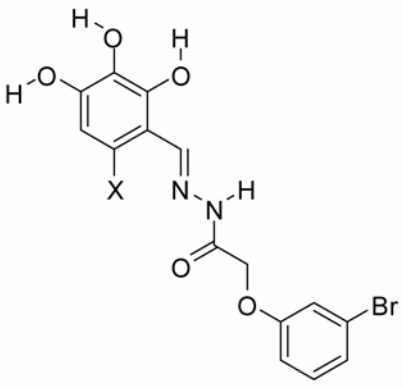


Figure 4.7 The 5 Å binding pocket of MOL282 in mMrgC11 receptor. The hydrogen bond and inter-aromatic ring distance are specified in Å.

receptor/ligand complex structures were further optimized with the conjugate gradient minimization while all atoms were movable. The final structure was then chosen with the best binding energy. Therefore we ended up with 87 optimized complex structures. No further optimization such as the sidechain replacement and annealing MD was carried out.

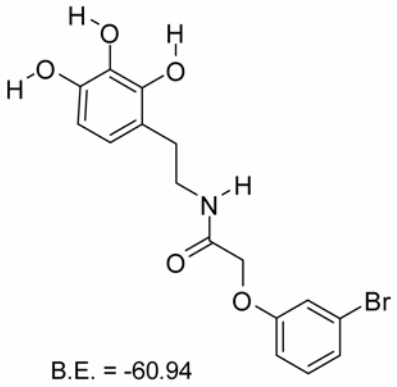
The best binding configuration across the 87 optimized structures is shown in Fig. 4.7. All three key residues interact with the ligand; Asp161 and Asp179 form hydrogen bonds with the ligand and Tyr110 participates in the π-π interaction with one of aromatic rings. Trp162, Leu241 and Tyr250 form the hydrogen bonds with the carbonyl group and the other hydroxyl groups of the ligand. However the ligand had the strain energy of ~15 kcal/mol (energy in gas phase with

(a)



X	B.E.
H	-58.81
CF3	-62.22
CH3	-63.32
ethyl	-63.93
isopropyl1	-63.33
isopropyl2	-61.70
t-butyl	-59.12

(b)



B.E. = -60.94

Figure 4.8 Suggested better binders derived from MOL282. The binding energy is in kcal/mol.

the dielectric constant of 2.5) in the docked conformation. Most strain resulted from the twist in θ1 torsion of Figure 4.7 (θ1=180° in the global minimum). To stabilize this twisted configuration, the substitution of a bulky group for the ortho hydrogen was suggested as shown in Figure 4.8(a). This bulky group also enhanced the van der Waals interaction with the receptor, leading to the increase of the binding energy. However as it became too bulky to occupy the void space in the binding pocket, it interfered binding (see the table in Fig. 4.8(a)).

Since nitrogen in C=N bond of MOL282 does not play a role in binding, C=N double bond was replaced by C-C single bond to reduce the strain seen in the docked configuration of MOL282 (Fig. 4.8(b)). This derivative of MOL282 binds to the mMrgC11 receptor similarly to MOL282, except that one of hydrogen bond partners was switched from Tyr250 to Lys99. The

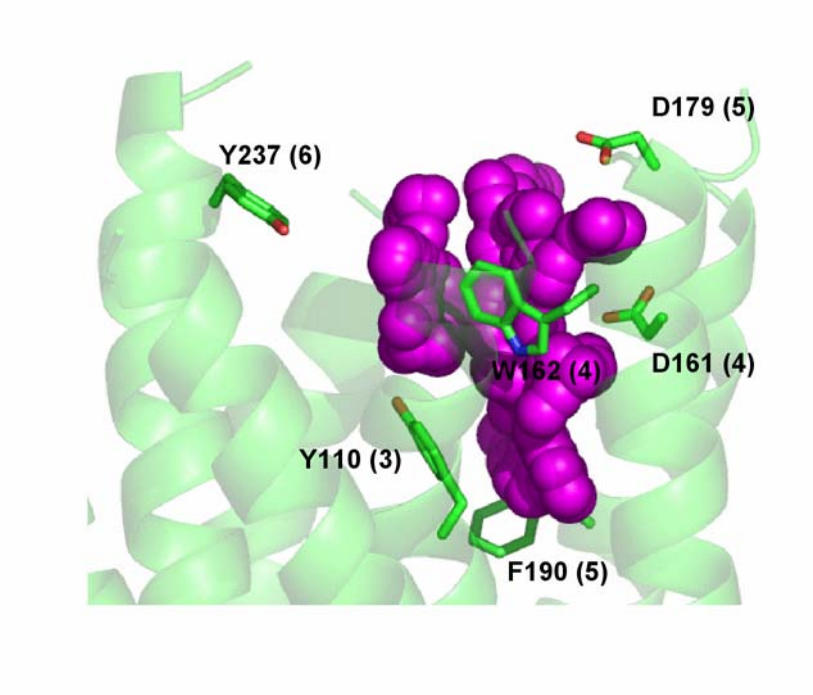


Figure 4.9 The 5 Å binding pocket of mMrgC11 receptor optimized with the tetra-peptide agonist, F-(D)M-R-F-NH_2. Six key residues identified in the previous prediction are shown in stick. The spheres representing the binding site of F-(D)M-R-F-NH_2 are colored by magenta.

strain energy of the ligand in the docked configuration decreased by ~7 kcal/mol and the snap binding energy slightly increased by ~5 kcal/mol, leading to the similar relaxed binding energy where the strain penalty was taken into account.

4.3.4 Virtual screening for F-(D)M-R-F-NH_2 bound site

The 5 Å binding site of F-(D)M-R-F-NH_2 is shown in Figure 4.9. Compared with the di-peptide binding site in Figure 4.4, the site is obviously wider. The buried surface was calculated using the Connolly MS program from Quantum Chemistry Program Exchange (QCPE) with a probe radius of 1.4 Å and a surface density of 5 dots/$Å^2$. The area for the buried part of F-(D)M-R-F-HN_2 was 466 $Å^2$, which was larger than 263 $Å^2$ for R-F-OH. The N-terminal F-(D)M part was extended towards TM6 and TM7, covering the additional TM regions. Tyr237 (TM6) is one of the key residues newly identified in the tetra-peptide binding site.

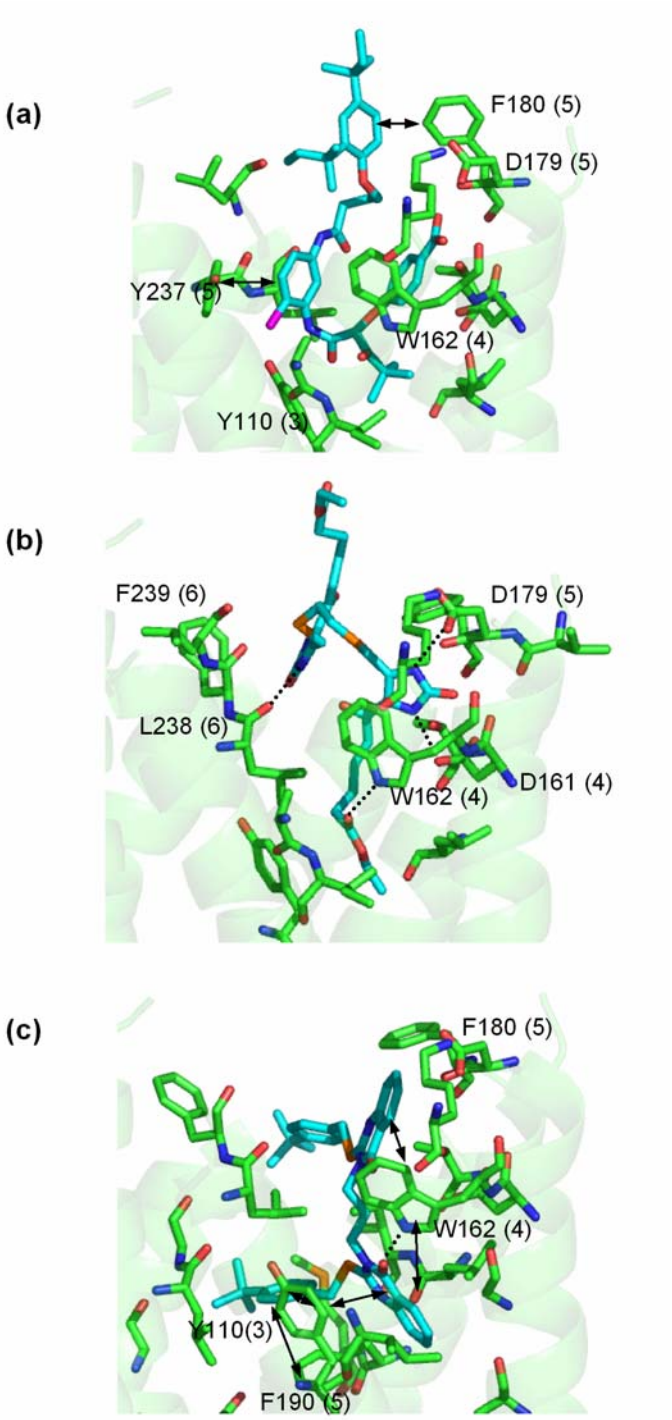


Figure 4.10 The 5 Å binding site of the best three hit compounds (a) comp242755 (b) comp241282 (c) comp391008. The intermolecular hydrogen bond is indicated by the dotted line and the aromatic interaction by the two-sided arrow.

The chemical structures of the final 55 hit compounds are shown in Figure S4.3, where the residues making a hydrogen bond or having a good van der Waals interaction (interaction energy with a ligand is greater than 3 kcal/mol) are identified together. Most are bulky since the surface area is considered as one of the descriptors, and relatively nonpolar compounds. They belong to the different class of compounds compared with those screened previously in the di-peptide case.

The detailed binding modes of the compounds with the best (comp242755), the second best (comp241282) and the third best (comp391008) binding energy are described in Figure 4.10. In comp242755, three aromatic rings interact with Trp162 (TM3), Phe180 (TM5) and Tyr237 (TM6). The *t*-butyl group has a favorable hydrophobic interaction with Tyr110 (TM3). The side chains of Trp162 (TM4) and Asp179 (TM5) are involved in the formation of hydrogen bond. However the hydrogen bond with Asp179 is unlikely if the carboxylate group in the benzoic acid part of comp242755 is deprotonated (pKa of benzoic acid = 4.20 for water at 25 °C). Since the buried receptor site might provide the different dielectric medium, the neutral form of comp242755 could be taken into account.

In comp24282, two key residues, Asp161 and Asp179 form hydrogen bonds with the ligand. Only two residues are shown to have good van der Waals interaction, but the ligand form two more hydrogen bonds with Trp162 (TM4) and Leu238 (TM6).

The comp391008 interacts with the receptor mainly through the hydrophobic interactions. The aromatic groups are well stacked with Phe190 (TM5), Tyr110 (TM3), Trp162 (TM4) and Phe180 (TM5). Asp161 and Asp179 do not interact with the ligand and are stabilized through the hydrogen bond or electrostatic interaction with Thr183 and Lys99 respectively as shown in the apo protein.

Although the hit compounds do not form as many hydrogen bonds as F-(D)M-R-F-NH_2, the nonpolar character would relieve the desolvation penalty in aqueous solution to help binding to the buried pocket of the receptor.

4.4 Summary and conclusions

The virtual screening with the combination of QSPR and docking method was carried out for the predicted mMrgC11 receptor. The antagonist ligand, MOL282 (IC50 = 46.5 μM) that had the best calculated binding energy was identified by mining ChemDiv database for the di-peptide binding site. The interactions with Asp161, Asp179 and Tyr110 shown in the agonist binding were also observed in MOL282. The novel ligands were derived from MOL282 in getting rid of the strain energy in its docked conformation. The identification of MOL282 as a hit provides the strong validation of our predicted binding site and low trial and error in the experiment (only 24 compounds were tested) demonstrates efficiency of our virtual screening method.

The different class of compounds was identified in virtual screening for the tetra-peptide binding site, having a large contribution of van der Waals interaction to the binding affinity. The experimental test of some of the top compounds would be needed to provide further validation.

The hit compounds identified in this study are certainly good staring points in designing new agonists or antagonists for the mMrgC11 receptor, and variation on the functional group in the series of ligands could be used to characterize the binding pocket. Moreover chemical characteristics of the hit compounds could provide some clues in deorphanizing Mrg receptors.

References

1. Klabunde, T. and G. Hessler, *Drug design strategies for targeting G-protein-coupled receptors.* Chembiochem, 2002. **3**(10): p. 929-944.
2. Rohrer, S.P., et al., *Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry.* Science, 1998. **282**(5389): p. 737-740.
3. Flohr, S., et al., *Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure-activity relationships and nuclear magnetic resonance studies on urotensin II.* Journal of Medicinal Chemistry, 2002. **45**(9): p. 1799-1805.
4. Kollman, P., *Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena.* Chemical Reviews, 1993. **93**(7): p. 2395-2417.
5. Aqvist, J., V.B. Luzhkov, and B.O. Brandsdal, *Ligand binding affinities from MD simulations.* Accounts of Chemical Research, 2002. **35**(6): p. 358-365.
6. Tropsha, A., *Recent Trends in Quantitative Structure-Activity Relationships*, in *Burger's medicinal chemistry and drug discovery*, D.J. Abraham, Editor. 2003, John Wiley & Sons: New York. p. 49-77.
7. Oloff, S., et al., *Chemometric analysis of ligand receptor complementarity: Identifying complementary ligands based on receptor information (CoLiBRI).* Journal of Chemical Information and Modeling, 2006. **46**(2): p. 844-851.
8. Breneman, C.M., et al., *Electron-Density Modeling of Large Systems Using the Transferable Atom Equivalent Method.* Computers & Chemistry, 1995. **19**(3): p. 161-&.
9. Breneman, C.M., et al., *New developments in PEST shape/property hybrid descriptors.* Journal of Computer-Aided Molecular Design, 2003. **17**(2): p. 231-240.

10. Wang, R.X., et al., *The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures.* Journal of Medicinal Chemistry, 2004. **47**(12): p. 2977-2980.

11. Gasteiger, J. and M. Marsili, *Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges.* Tetrahedron, 1980. **36**(22): p. 3219-3228.

12. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field for Molecular Simulations.* Journal of Physical Chemistry, 1990. **94**(26): p. 8897-8909.

13. *Cerius2 Modeling Environment, Release 4.0*, Accelrys Inc.: San Diego.

14. Zamanakos, G., *A fast and accurate analytical method for the computation of solvent effects in molecular simulations*, in *Division of Physics, Mathematics and Astronomy*. 2002, California Institute of Technology: Pasadena.

15. McDonald, I.K. and J.M. Thornton, *Satisfying Hydrogen-Bonding Potential in Proteins.* Journal of Molecular Biology, 1994. **238**(5): p. 777-793.

16. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.

17. Lim, K.T., et al., *Molecular dynamics for very large systems on massively parallel computers: The MPSim program.* Journal of Computational Chemistry, 1997. **18**(4): p. 501-521.

18. Ding, H.Q., N. Karasawa, and W.A. Goddard, *Atomic Level Simulations on a Million Particles - the Cell Multipole Method for Coulomb and London Nonbond Interactions.* Journal of Chemical Physics, 1992. **97**(6): p. 4309-4315.

19. Golbraikh, A., et al., *Rational selection of training and test sets for the development of validated QSAR models.* Journal of Computer-Aided Molecular Design, 2003. **17**(2): p. 241-253.

20. Szallasi, A. and P.M. Blumberg, *Vanilloid (capsaicin) receptors and mechanisms.* Pharmacological Reviews, 1999. **51**(2): p. 159-211.

21. Dong, X.Z., et al., *A diverse family of GPCRs expressed in specific subsets of nociceptive sensory neurons.* Cell, 2001. **106**(5): p. 619-632.

Supporting Figures

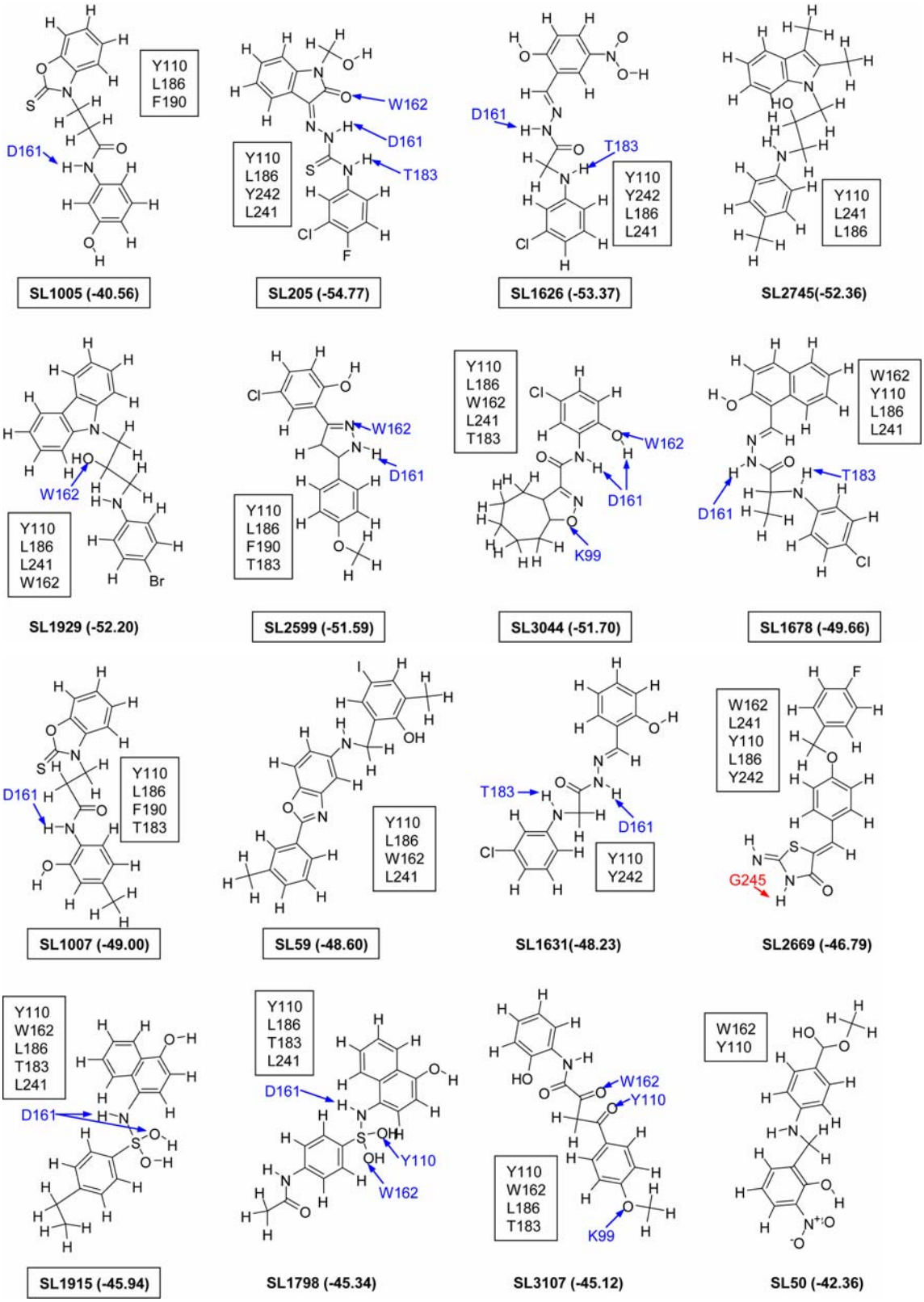


Figure S4.1 Hit compounds from the first ligand set after docking.

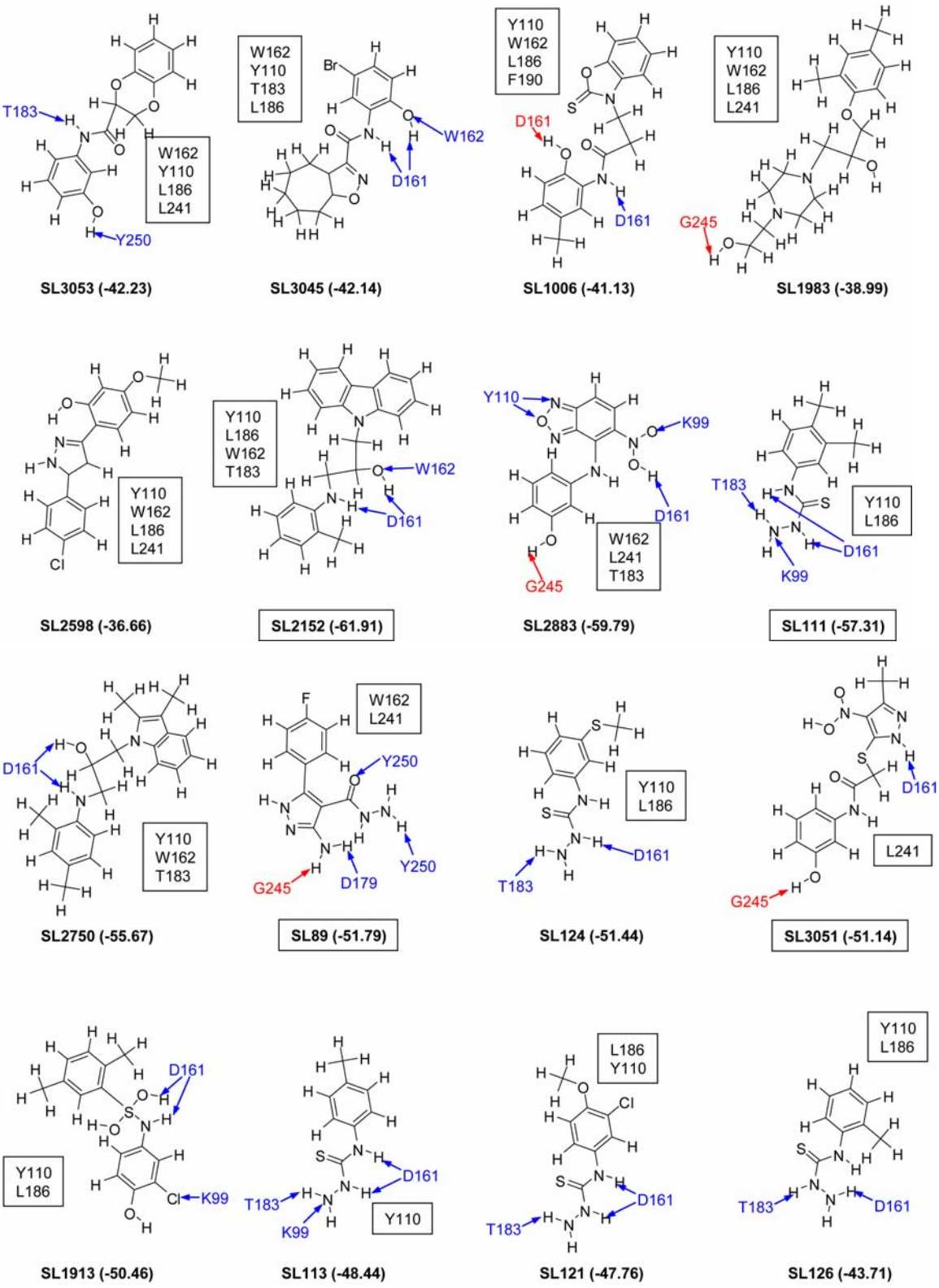


Figure S4.1 (*continued*) Hit compounds from the first ligand set after docking.

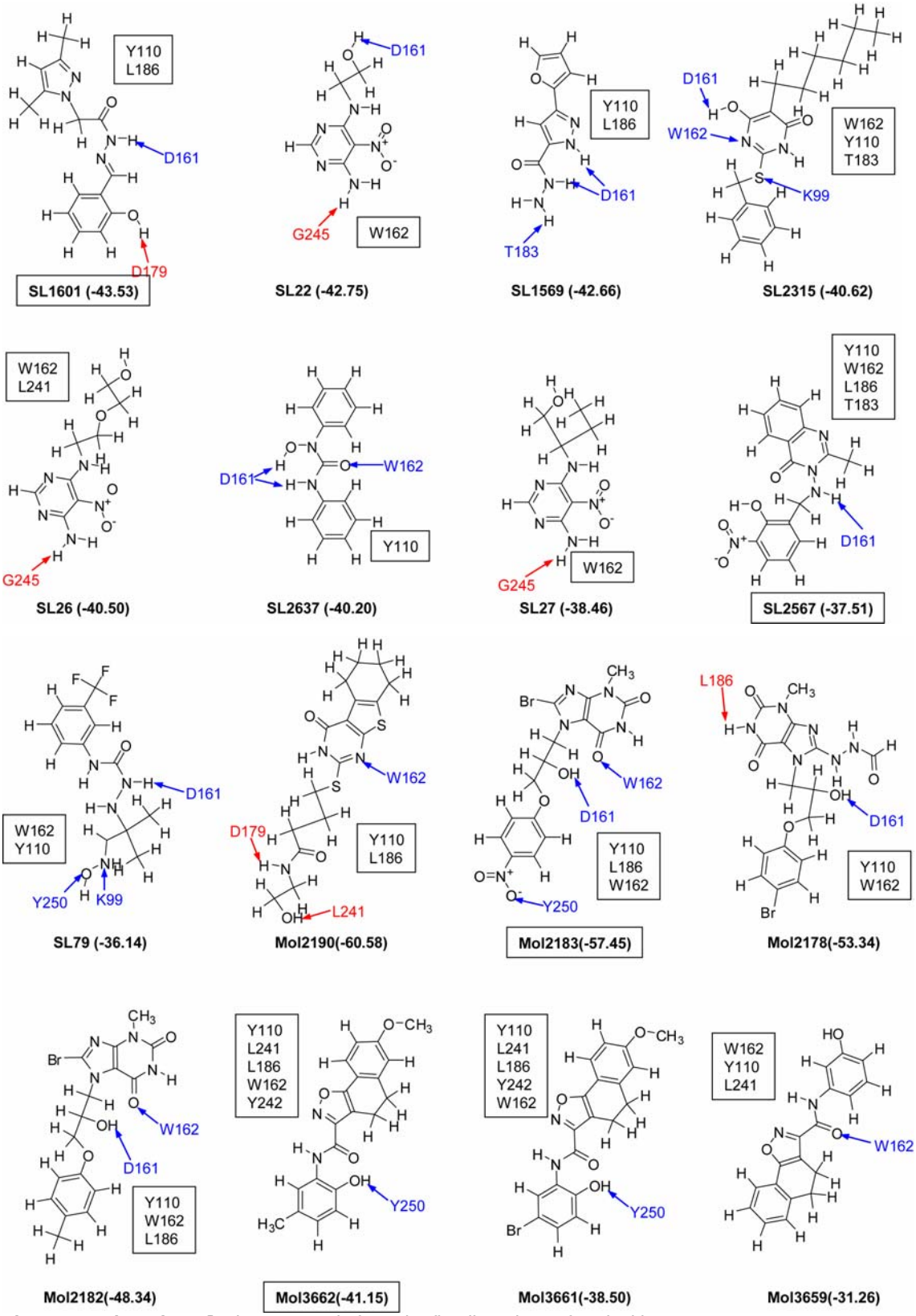


Figure S4.1 (*continued*) Hit compounds from the first ligand set after docking.

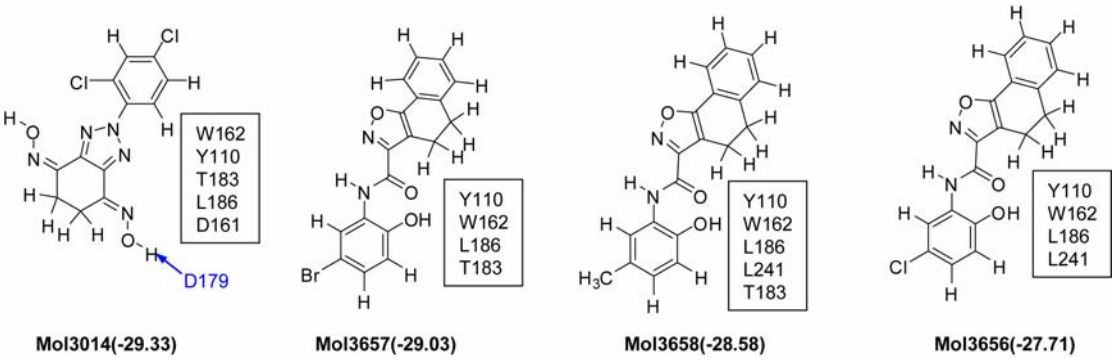


Figure S4.1 (*continued*) Hit compounds from the first ligand set after docking. The ligands whose names are enclosed by rectangular box were tested in experiment. The number in parenthesis corresponds to the calculated binding energy in kcal/mol. Residue in blue makes a hydrogen bond through its side chain with the atom indicted by the blue arrow. The residue in red has backbone atoms involved in the hydrogen bond. The residues in box have good van der Waals interactions with a ligand (E > 3 kcal/mol).

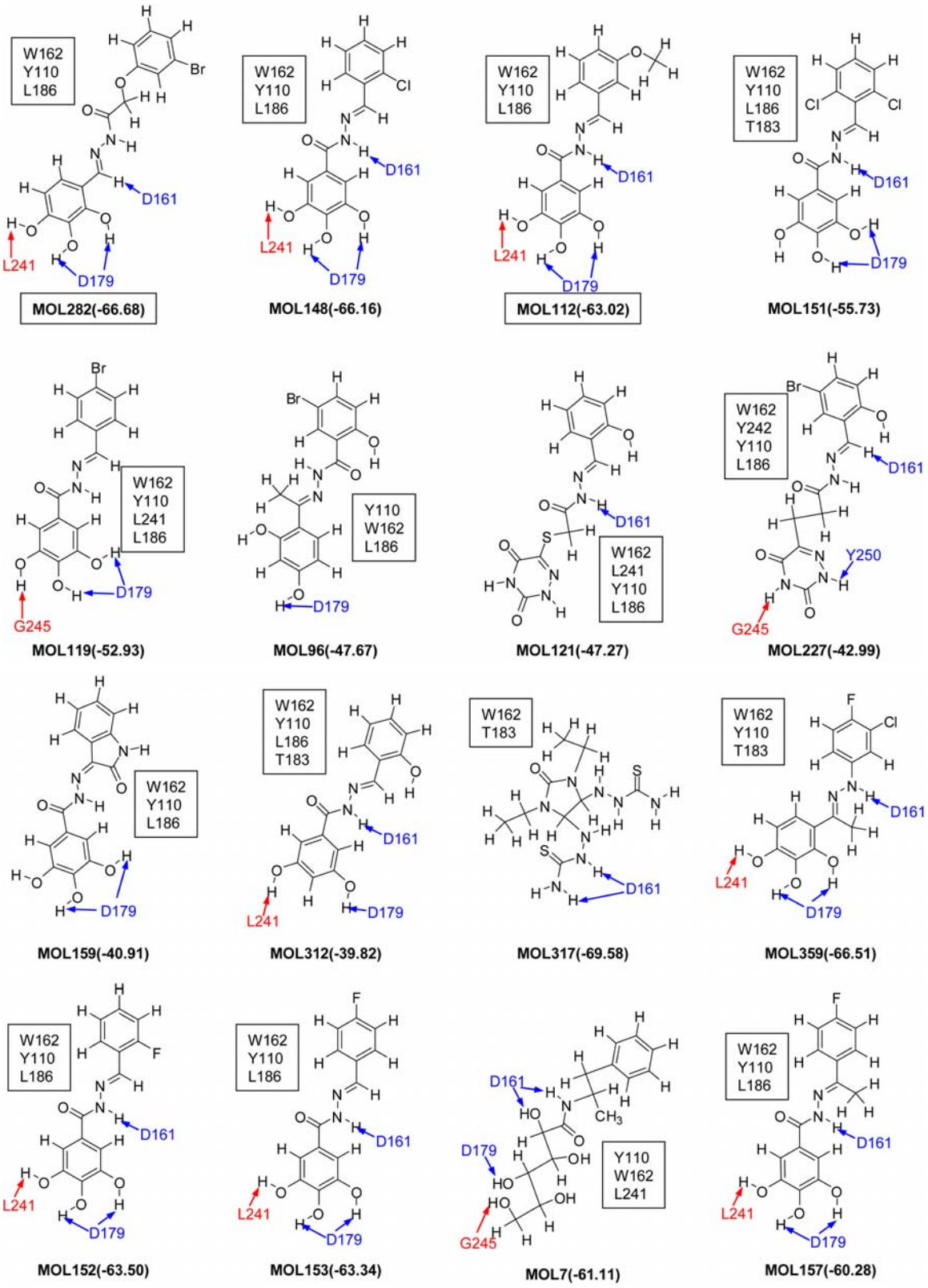


Figure S4.2 Hit compounds from the second set after docking.

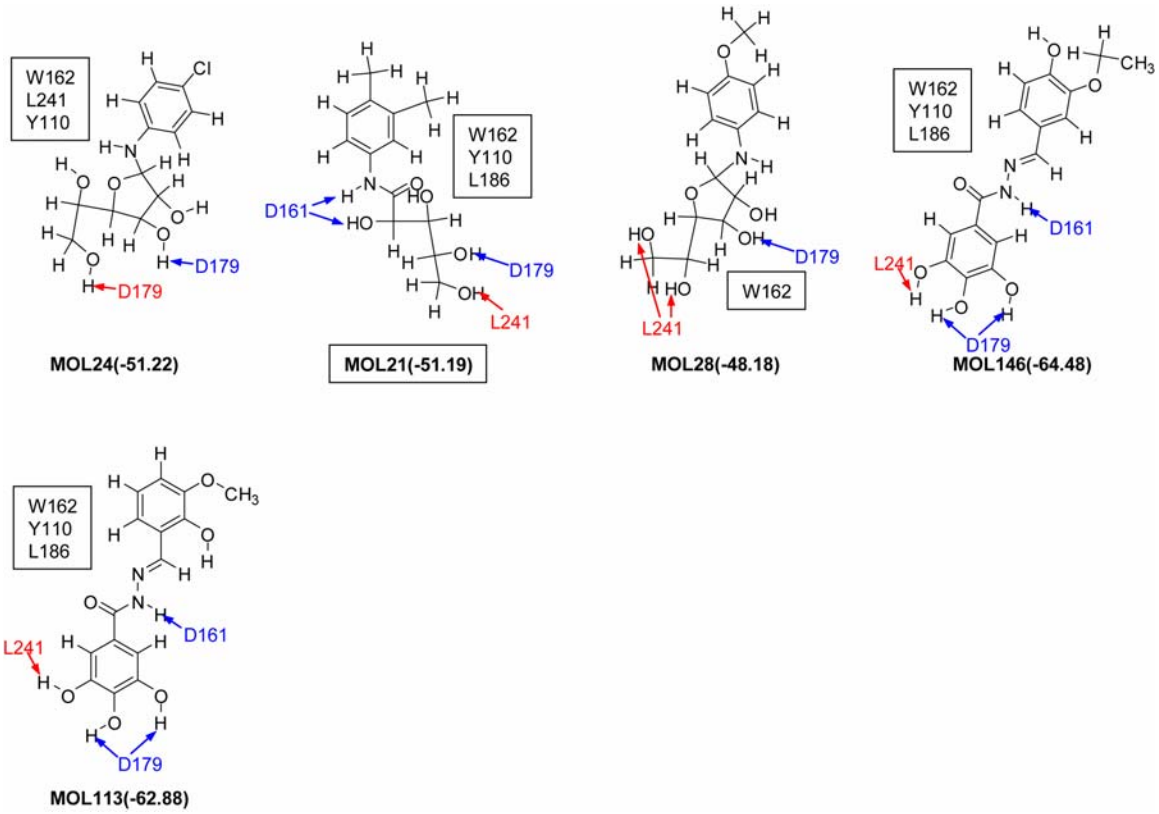


Figure S4.2 (*continued*) Hit compounds from the second set after docking.

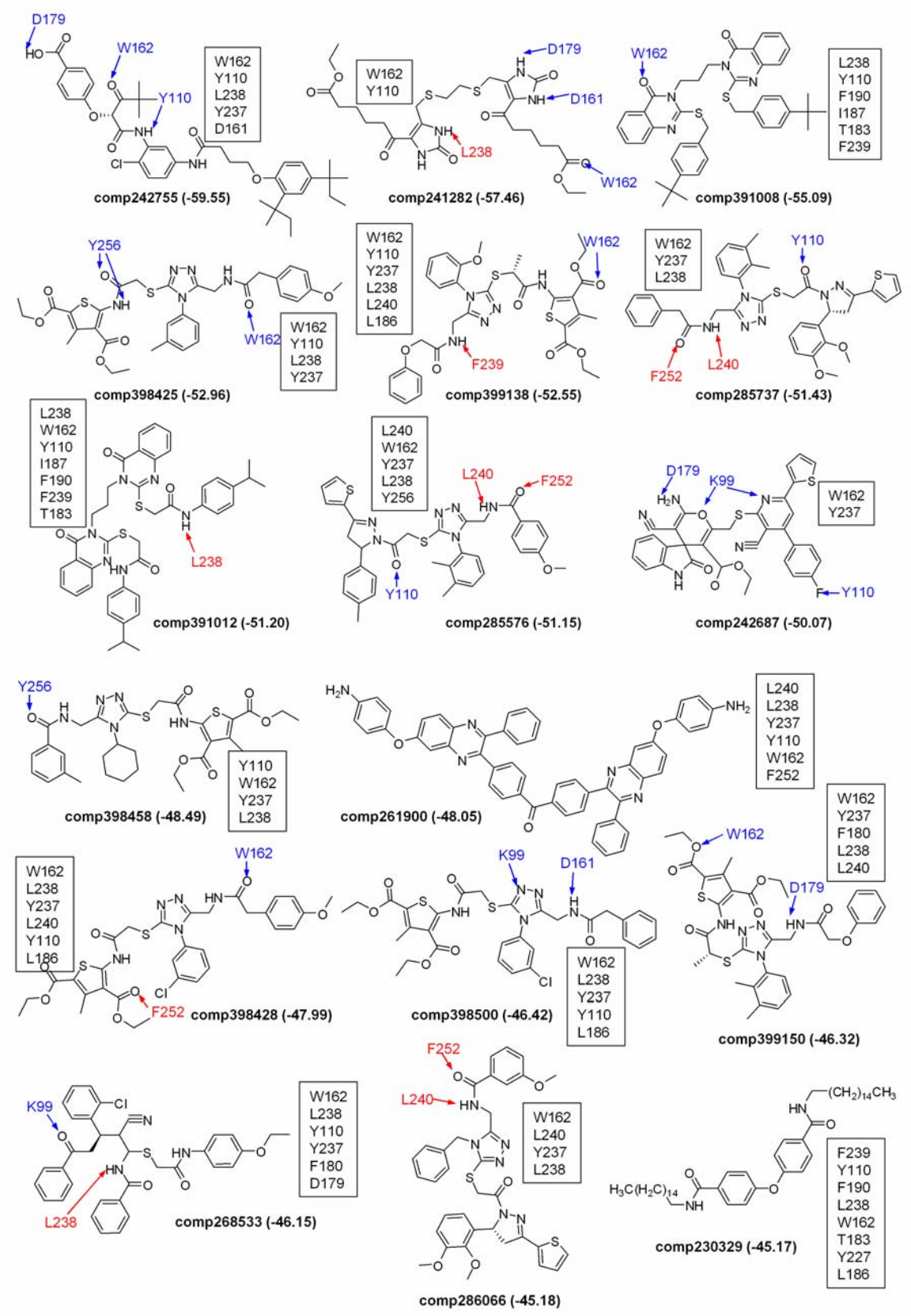


Figure S4.3 Hit compounds after virtual screening for the tetra-peptide binding site.

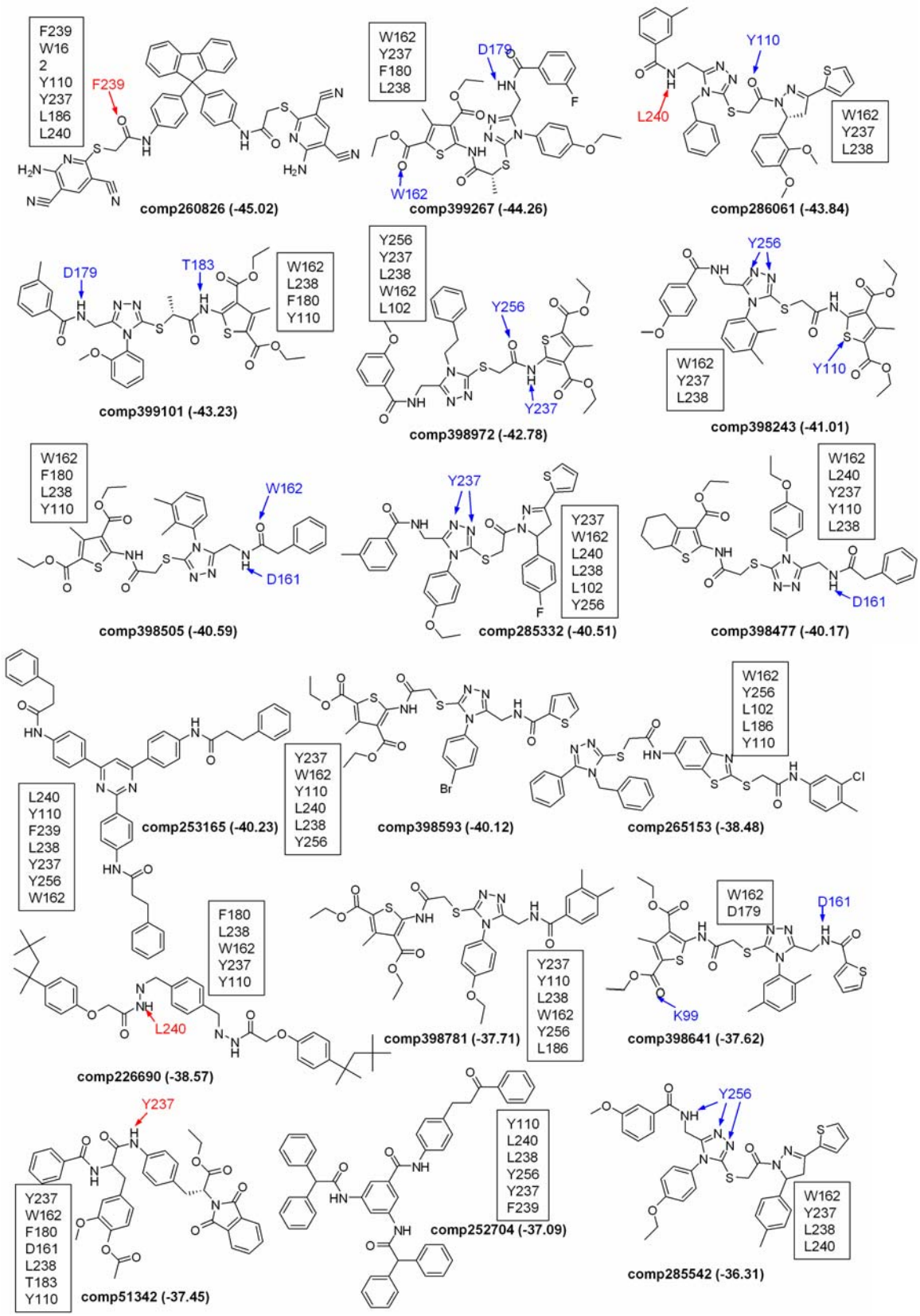


Figure S4.3 (*continued*) Hit compounds after virtual screening for the tetra-peptide binding site.

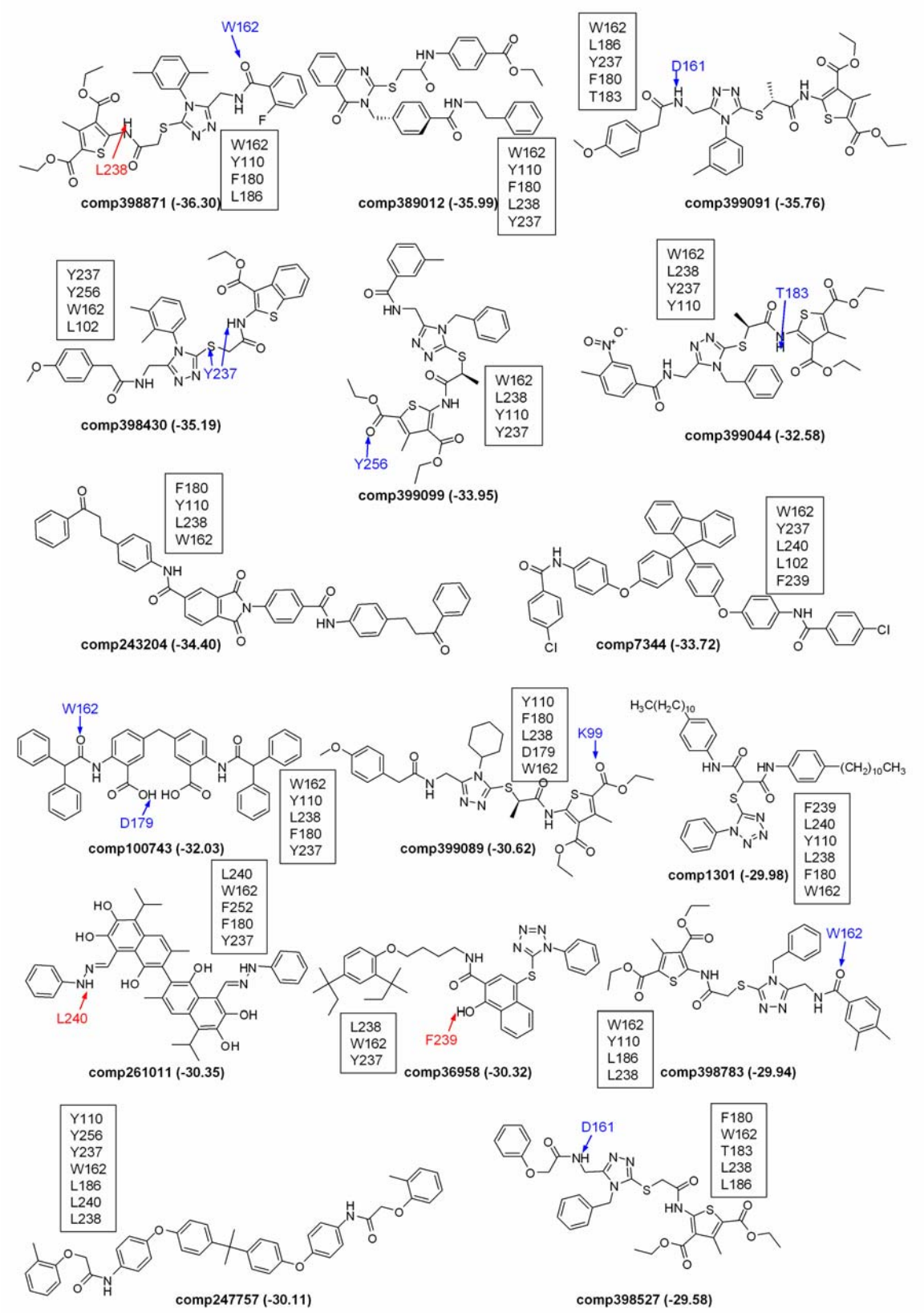


Figure S4.3 (*continued*) Hit compounds after virtual screening for the tetra-peptide binding site.

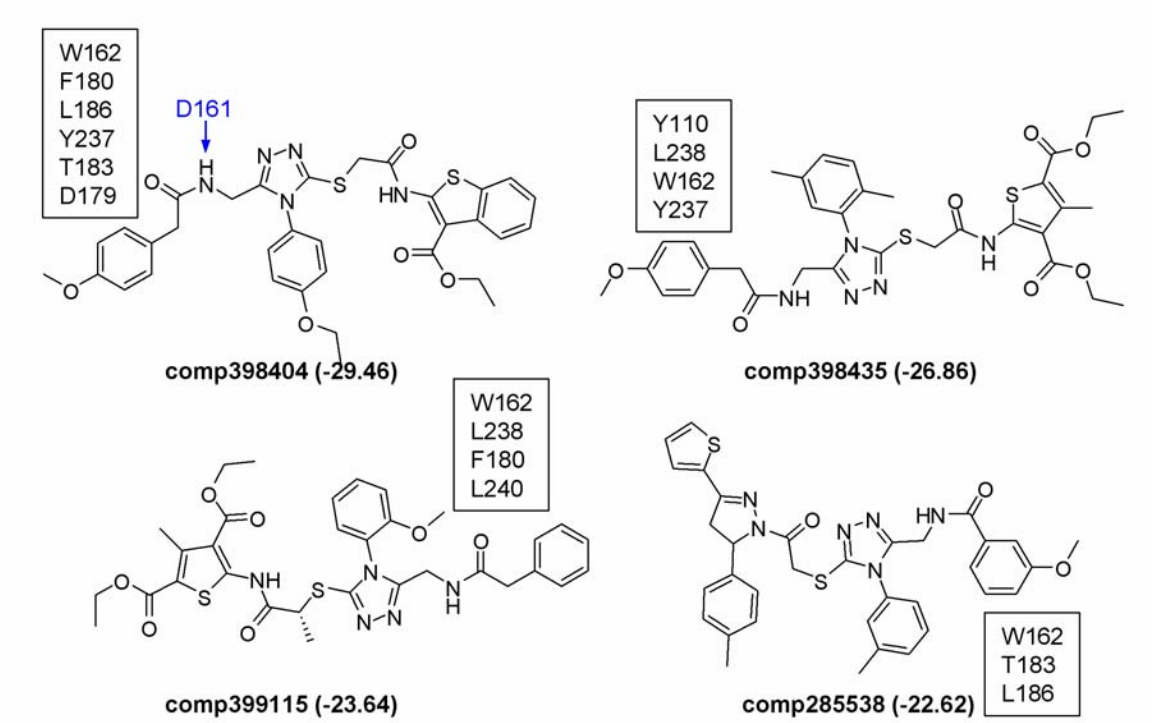


Figure S4.3 (*continued*) Hit compounds after virtual screening for the tetra-peptide binding site. The residues involved in the intermolecular hydrogen bonds or the van der Waals interactions are indicated in the same way as Figure S4.1.