Protein-Ligand ANT System

user manual for version 1.1

O. Korb

Oliver.Korb@uni-konstanz.de

1 Configuration File Settings

PLANTS needs a configuration file that contains all data needed for docking / virtual screening and rescoring. The following sections describe the settings that can be adjusted by the user. For the case that no parameters are specified, recommended standard settings are used. PLANTS can then be executed in **virtual screening** mode by typing *PLANTS --mode screen yourconfigfile*, where *yourconfigfile* is the name of the configuration file. A **rescoring** of existing ligand poses can be carried out by typing *PLANTS --mode rescore yourconfigfile*

1.1 Search Algorithm

Since version 1.1, PLANTS offers predefined search speed settings which are set according to the scoring function chosen.

- **search_speed** *value*: search speed setting, *value* can be set to *speed1* (highest reliability, slowest setting), *speed2* (good reliability, twice as fast as speed1) or *speed4* (modest reliability, four times as fast as speed1) (standard: speed1)

The parameters can be altered by adding one or more of the following keywords to the configuration file:

- **aco_ants** *value*: number of ants (standard: 20)
- **aco_evap** *value*: evaporation factor $\rho \in [0; 1]$
- **aco_sigma** *value*: iteration scaling factor σ
- **flip_amide_bonds** *value*: activate (1) or deactivate flipping of amide bonds (standard: 0)
- **flip_planar_n** *value*: activate (1) or deactivate flipping of bonds next to planar nitrogens (standard: 1)

- **force_flipped_bonds_planarity** *value*: activate (1) or deactivate (0) automatic planarity correction for flippable bonds (standard: 0)
- **force_planar_bond_rotation** *value*: activate (1) or deactivate (0) free rotation of planar bonds (standard: 1)
- **rescore_mode** *value*: perform simplex optimization during rescoring (*value*=simplex) or only direct input conformation scoring (*value*=no_simplex) (standard: simplex)

1.2 Bindingsite

- **bindingsite_center** *value*1 *value*2 *value*3: center coordinates of the binding-site sphere (no standard value)
- **bindingsite_radius** *value*: radius of the binding-site sphere (no standard value)

1.3 Cluster Algorithm

- **cluster_rmsd** *value*: RMSD similarity threshold for cluster algorithm (standard: 2.0)
- **cluster_structures** *value*: number of structures generated by the cluster algorithm (standard: 10)

1.4 Scoring functions

Intermolecular (protein-ligand interaction scoring):

- **scoring_function** *string*: *plp*, *plp95* or *chemplp* (standard: chemplp)
- **outside_binding_site_penalty** *value*: scoring functions using precalculated grids use *value* to fill grid points outside the binding site definition (standard: 50.0)
- **enable_sulphur_acceptors** *value*: activate (1) or deactivate (0) scoring of sulphur acceptors (standard: 0)

Intramolecular ligand scoring:

- **ligand_intra_score** *string*: *clash*, *clash2* (simple heavy-atom clash terms) or *lj* (all-atom Lennard-Jones term) (standard: clash2)
- **chemplp_clash_include_14** *value*: activate (1) or deactivate (0) scoring of 1-4 interactions (standard: 0.25)
- **chemplp_clash_include_HH** *value*: activate (1) or deactivate (0) scoring of hydrogen-hydrogen interactions (standard: 0)

Keywords affecting scoring function *plp* and *plp95*:

- **plp_steric_e** *value*: well-depth for setric PLP interactions (standard: -0.4)
- **plp_burpolar_e** *value*: well-depth for occluded polar PLP interactions (standard: -0.05)
- **plp_hbond_e** *value*: well-depth for polar PLP interactions (standard: -2.0)
- **plp_metal_e** *value*: well-depth for acceptor–metal PLP interactions (standard: -4.0)
- **plp_repulsive_weight** *value*: weight for repulsive PLP interactions (standard: 0.5)
- **plp_tors_weight** *value*: weight for the ligand torsional potential (standard: 1.0)

Keywords affecting scoring function *chemplp*:

- **chemplp_weak_cho** *value*: activate (1) or deactivate (0) weak CH-O scoring (standard: 1)
- **chemplp_charged_hb_weight** *value*: weighting factor (multiplier) for charged hydrogen bonds (standard: 2.0)
- **chemplp_charged_metal_weight** *value*: weighting factor (multiplier) for charged acceptor - metal interactions (standard: 2.0)
- **chemplp_hbond_weight** *value*: weighting factor neutral - neutral and neutral - charged hydrogen bonds (standard: -3.0)
- **chemplp_hbond_cho_weight** *value*: weighting factor for CH-O interactions (standard: -3.0)
- **chemplp_metal_weight** *value*: weighting factor for neutral acceptor - metal interactions (standard: -6.0)
- **chemplp_plp_weight** *value*: weighting factor for PLP interactions (standard: 1.0)
- **chemplp_plp_steric_e** *value*: well-depth for setric PLP interactions (standard: -0.4)
- **chemplp_plp_burpolar_e** *value*: well-depth for occluded polar PLP interactions (standard: -0.1)
- **chemplp_plp_hbond_e** *value*: well-depth for polar PLP interactions (standard: -1.0)
- **chemplp_plp_metal_e** *value*: well-depth for acceptor–metal PLP interactions (standard: -1.0)

- **chemplp_plp_repulsive_weight** *value*: weight for repulsive PLP interactions (standard: 1.0)
- **chemplp_tors_weight** *value*: weight for the ligand torsional potential (standard: 2.0)
- **chemplp_lipo_weight** *value*: weighting factor for lipophilic interactions (standard: 0.0)
- **chemplp_intercept_weight** *value*: intercept value (standard: -20.0)

1.5 Input

1.5.1 Atom Typing

PLANTS exclusively uses the MOL2 file format, thus MOL2-files (including bond connectivity) must be provided for all input files. PLANTS expects correct MOL2-atom- and bond-types. This is needed for the correct identification of rotatable bonds and charged functional groups and may influence docking and virtual screening performance. Especially take care for the following cases:

- **COO^-**: oxygens should get **O.co2** and not a combination of O.3 and O.2; please **deprotonate COOH**
- **guadinium**: carbocation should get **C.cat**, nitrogens **N.pl3**
- **positively charged sp3 nitrogen**: should get **N.4**

1.5.2 Input Options

- **protein_file** *string*: protein filename
- **ligand_file** *string*: ligand filename
- **ligand_list** *string*: text file containing ligand filenames

The keywords *ligand_file* and *ligand_list* may be used several times in a configuration file to specify for example a ligand database. Also **multi-mol2** files may be used as input for docking and rescoring.

- **use_gold_input** *string*: PLANTS reads in certain options from a GOLD input file; **only** the following entries are considered:
  - **protein_datafile**
  - **ligand_data_file**
  - **directory**
  - **flip_planar_n**

- **flip_amide_bonds**
- **radius**
- **origin**

1.5.3 Ligand-Specific Data

Ligand-specific data is stored at the end of each ligand file inside the @<TRIPOS>COMMENT block.

Protein Torsions

Angles of rotatable bonds in the active site of the protein are stored in the following format. This data is read for example in the rescoring mode and the protein's conformation is restored. Format:

PLANTSPROTEINTORSIONS
NUMBER_OF_INSTANCES
PROTEIN_BOND_NUMBER ANGLE // instance 1
PROTEIN_BOND_NUMBER ANGLE // instance 2 etc.

1.6 Output

1.6.1 Output Options

- **output_dir** *string*: name of output directory; PLANTS automatically tries to create a subdirectory with the specified name and **exits** if the directory already exists to prevent overwriting existing data
- **write_protein_conformations** *value*: activate (1) or deactivate (0) output of protein conformations for scoring functions *chemplp*. (standard: 0)
- **write_protein_bindingsite** *value*: write protein binding site only (activate (1) or deactivate (0)). (standard: 0)
- **write_protein_splitted** *value*: write fixed and dynamic parts of the protein in separate files (activate (1) or deactivate (0)). (standard: 0)
- **write_rescored_structures** *value*: activate (1) or deactivate (0) output of rescored structures. (standard: 0)
- **write_multi_mol2** *value*: activate (1) or deactivate (0) output of multi-mol2 files. (standard: 1)
- **write_ranking_links** *value*: activate (1) or deactivate (0) output of ranked soft links (requires *write_multi_mol2 0*). The links are stored in subdirectory *ranking/*. Unavailable when using explicit water molecules. (standard: 0)

- **write_ranking_multi_mol2** *value*: activate (1) or deactivate (0) output of a sorted multi-mol2 file (requires *write_multi_mol2 1*). Unavailable when using explicit water molecules. Warning: this option makes heavy use of the hard disk and the sorting process may take some time depending on the size of the virtual screening. The output file is stored in subdirectory *ranking/*. (standard: 0)
- **write_per_atom_scores** *value*: activate (1) or deactivate (0) output of per molecule atom scoring values; partial atom charges are replaced. (standard: 0)
- **write_merged_ligand** *value*: activate (1) or deactivate (0) output of merged ligand files (written to *mergedStates.mol2*). (standard: 0)
- **write_merged_protein** *value*: activate (1) or deactivate (0) output of merged protein files (written to *mergedStates.mol2*). (standard: 0)
- **write_merged_water** *value*: activate (1) or deactivate (0) output of merged water files (written to *mergedStates.mol2*). (standard: 0)
- **keep_original_mol2_description** *value*: keep (1) original mol2 description field as prefix for ligand naming. (standard: 1)
- **merge_multi_conf_output** *value*: activate (1) or deactivate (0) merge of multiconformer output. This is only carried out for ranked databases (see above). (standard: 0)
- **merge_multi_conf_character** *char*: character used for merging structure names. (standard: _)
- **merge_multi_conf_after_characters** *value*: if merge character occured exactly *value* times (from right to left) the resulting prefix of the name is used for merging the structures keeping only the best scoring one for each ligand. (standard: 1)

1.6.2 Description of Output Files

The following files are generated during a virtual screening or a rescoring run:

- **ligand files**: the names are dependent on the input ligand filenames. For each conformation generated by the cluster algorithm, a suffix (format _entry_XXXXX_conf_YY.mol2) is added, where XXXXX is the number of the molecule in a (multi-) mol2 file and YY is the ligand conformation according to the cluster algorithm (YY=01 is the best scoring pose).
- **protein files**: ligandfilename_entry_XXXXX_conf_YY_protein.mol2
- **water molecule files**: ligandfilename_entry_XXXXX_conf_YY_water_ZZ.mol2, where ZZ is the water molecule identifier
- **bestranking.csv**: contains the score values, the number of scoring function evaluation and the docking time for each best-ranking ligand pose

- **ranking.csv**: same as *bestranking.csv*, but for all ligand poses
- **features.csv**: information about partial scoring function terms for all ligand poses
- **constraints.csv**: information about constraints for all ligand poses
- **protein.log**: information about the protein setup
- **ligand.log**: information about the ligand setup and the docking run
- **score.log**: scoring function parameters
- **optimizer.log**: information regarding the search algorithm settings

Explanation of the header entries for **ranking.csv** and **bestranking.csv**:

- TOTAL_SCORE: scoring function value obtained during docking
- SCORE_RB_PEN: TOTAL_SCORE plus penalty value for each ligand rotatable bond
- SCORE_NORM_HEVATOMS: TOTAL_SCORE divided by number of ligand heavy atoms
- SCORE_NORM_CRT_HEVATOMS: TOTAL_SCORE divided by cubic root of number of ligand heavy atoms
- SCORE_NORM_WEIGHT: TOTAL_SCORE divided by molecular weight of ligand
- SCORE_NORM_CRT_WEIGHT: divided by cubic root of molecular weight of ligand
- SCORE_RB_PEN_NORM_CRT_HEVATOMS: SCORE_RB_PEN divided by cubic root of number of ligand heavy atoms
- SCORE_NORM_CONTACT: TOTAL_SCORE divided by number of protein-ligand contacts
- EVAL: number of scoring function evaluations
- TIME: docking time

Explanation of the header entries for **features.csv**:

- TOTAL_SCORE: see above
- SCORE_RB_PEN: see above
- SCORE_NORM_HEVATOMS: see above

- SCORE_NORM_CRT_HEVATOMS: see above
- SCORE_NORM_WEIGHT: see above
- SCORE_NORM_CRT_WEIGHT: see above
- SCORE_RB_PEN_NORM_CRT_HEVATOMS: see above
- SCORE_NORM_CONTACT: see above
- PLPPARTHBOND: PLP hbond score
- PLPPARTSTERIC: PLP steric contact score
- PLPPARTMETAL: PLP metal interaction score
- PLPPARTREPULSIVE: PLP donor/donor and acceptor/acceptor repulsion score
- PLPPARTBURPOLAR: PLP buried polar atoms score (polar atoms occluded by nonpolar ones)
- LIG_NUM_CLASH: number of ligand atoms with PLP score greater zero
- LIG_NUM_CONTACT: number of ligand atoms with attractive PLP score
- LIG_NUM_NO_CONTACT: number of ligand atoms with zero PLP score
- CHEMPARTMETAL: CHEMSCORE metal interaction score
- CHEMPARTHBOND: CHEMSCORE hbond score
- DON: number of ligand donor atoms
- ACC: number of ligand acceptor atoms
- UNUSED_DON: number of unpaired ligand donors
- UNUSED_ACC: number of unpaired ligand acceptors
- CHEMPLP_CLASH: intra-ligand clash score
- TRIPOS_TORS: intra-ligand torsion score
- INTRAPROT_CHEMPLP_PLP: intra-protein score (only calculated for flex. side-chains)
- ATOMS_OUTSIDE_BINDINGSITE: number of ligand atoms outside binding site

1.7 Constraints

Multiple instances of the following constraints can be specified:

- **chemplp_protein_hb_constraint** *value*1 *value*2: protein hbond constraint; *value1* defines the atom number (according to the MOL2 file) and *value2* the weight. If a ligand forms a hydrogen bond with the specified protein atom, the resulting score is multiplied by *value2* (only works with scoring function *chemplp*, information is written to *features.csv*).

- **shape_constraint** $mol2-file\ weight$: shape constraint; $mol2-file$ specifies the molecule that is used for the volume overlap calculation (the more ligand atoms overlap, the better). For an optimal atom-overlap of a ligand and a shape atom *value* is added to the score (*value* should be negative, for example -3.0).

- **surface_distance_constraint** $from\ to\ weight\ (atomID)^{+}$: defines a range (between $from$ and to in Å from the protein surface) in which the specified ligand atoms (defined by $atomID$; you can specify multiple IDs separated by blanks) should be placed. If a ligand atom is placed inside the specified range, *weight* is added to the score (*weight* should be negative, for example -3.0).

- **ligand_intra_distance_constraint** $from\ to\ weight\ LatomID1\ LatomID2$: defines a distance range (between $from$ and to) for two ligand atoms ($LatomID1$ and $LatomID2$). If the distance is inside the specified range, *weight* is added to the score (*weight* should be negative, for example -3.0).

- **protein_ligand_distance_constraint** $from\ to\ weight\ PatomID\ LatomID$: defines a distance range (between $from$ and to) for a protein ($PatomID$) and a ligand atom ($LatomID$). If the distance is inside the specified range, *weight* is added to the score (*weight* should be negative, for example -3.0).

1.8 Flexible side-chains

PLANTS allows protein side-chains to be treated flexible. PLANTS expects the protein file to include the backbone information according to the mol2 specification (backbone atoms need to be marked with the *BACKBONE* tag). If this information is not available PLANTS may behave in an unexpected way.

- **flexible_protein_side_chain_string** *string*: residue label of flexible side-chain (e.g. *VAL123*)

- **flexible_protein_side_chain_number** *value*: residue number of flexible side-chain (e.g. *123*)

- **intra_protein_score_weight** *value*: weighting factor for the intramolecular protein score (standard: 0.3)

- **fix_protein_bond** *value*: keep protein bond with bond-number *value* fixed.

The use of *flexible_protein_side_chain_number* is recommended. The file *protein.log* can be used to check whether the side-chains have been identified correctly. In addition to the ligand conformations also the corresponding protein conformations can be written to files called *protein_ligandfilename* (see section *Output* for details). For rescoring purposes each ligand conformation file contains the torsion angles of the receptor side-chains and rotatable donor groups. Using such a conformation file as input will restore the bound receptor conformation.

1.9 Multiconformer Docking

PLANTS ic also capable of performing rigid-body docking, allowing for multiconformer docking of externally generated ligand conformations. Depending on the scoring function (e.g. *chemplp*), only rotatable donor groups in the ligand will be treated flexible. It is also possible to use rigid-body docking in conjunction with flexible protein side-chains (not available in *rigid_all* mode).

- **rigid_ligand** *value*: activate (1) or deactivate (0) rigid ligand docking (standard: 0)
- **rigid_all** *value*: activate (1) or deactivate (0) rigid protein and rigid ligand docking (standard: 0)

1.10 Water

Explicit water molecules can be used during docking by adding one or several of the following keywords to the configuration file:

- **water_molecule** X Y Z r [t s]: X, Y and Z specify the center of the sphere inside which the water molecule is allowed to move and r the radius of the sphere. This is the default mode. For a detailed description see below. It is also possible to fix the water molecule's translation by setting $t = 0$ and adding a switching degree of freedom with $s = 1$. In this case the water molecule can rotate and it is switched on or off by the search algorithm.
- **water_molecule_definition** $filename$: MOL2-file that contains a single water molecule (the position and orientation is arbitrary).
- **water_protein_hb_weight** *value*: weight of water-protein hydrogen bonds (standard: 1.0).
- **water_ligand_hb_weight** *value*: weight of water-ligand hydrogen bonds (standard: 1.0).
- **water_water_hb_weight** *value*: weight of water-water hydrogen bonds (standard: 1.0).

- **no_water_ligand_hb_penalty** *value*: penalty value that is added if no water-ligand hydrogen bond is formed (standard: 0.0).
- **water_enable_penalty** *value*: penalty value that is added if a water molecule is active (standard: 8.0).

If the water molecule is displaced by a ligand and moved outside the sphere, the water molecule has no score contribution. The weighting factors are only used for CHEMSCORE hydrogen bonding part of scoring function *chemplp*, which is recommended for docking with explicit water molecules. Scoring function *plp* may also be used, but in this case only a water molecule's translational degrees are optimized and the above parameters have no effect. For each complex conformation also a water molecule conformation is written if it has not been displaced by a ligand.

2 Special Modes

Besides virtual screening and rescoring PLANTS also offers some special execution modes.

2.1 Molecule Splitting

The molecule splitting mode allows to separate several molecules available in a single mol2-file. This may be especially of interest when a complete PDB-file is processed with *SPORES* and the protein, the ligands and the water molecules are expected to be available in separate files. To start PLANTS in splitting mode, type:
PLANTS --mode split molecule.mol2

The output then consists of the following files:

- **protein.mol2**: The molecule that is expected to be the protein (plus metal ions).
- **ligand_XYZ_N.mol2**: Ligand with residue label XYZ. N increases for each instance of the ligand available in the protein.
- **water_near_ligand_XYZ_N.mol2**: Water molecules in a maximum distance of 3Å to any ligand atom.
- **water.mol2**: All available water molecules.

2.2 Bindingsite Definition by Ligand

This mode calculates a sphere conatining the ligand *molecule.mol2*. The user can specify an additional distance x that is added to the sphere radius:
PLANTS --mode bind molecule.mol2 x [*protein.mol*2]

Output:

- **bindingsite.def**: Spherical bindingsite definition.
- **PLANTSactiveSite.mol2**: Active site atoms as used inside PLANTS, i.e. all atoms that lie inside the sphere (only written if protein file specified).
- **PLANTSactiveSiteResidues.mol2**: Complete active site residues (only written if protein file specified).