DOCKING TUTORIAL

A. The docking Workflow

1. Ligand preparation

It consists in the standardization of the molecular description (aromatization, ionisation of titratable groups, choice of the main tautomeric state) and the generation of a low energy conformer.

2. Protein preparation

It consists in:
- the definition of the receptor: selection of protein chains, co-factors, ions
- the definition of the binding site: selection of the residues flanking the binding pocket
- the check of the structure
    - ionisation state of GLU and ASP residues ((-) or neutral) and of HIS, LYS and ARG residues ((+) or neutral)
    - tautomeric state of neutral HIS residue (protonation of N-delta or N-epsilon)
    - position of polar hydrogen atoms
        - in the hydroxyl group of SER, TYR and THR residues
        - in the amine group of LYS residue
        - in the side chain amide group of ASN and GLN residues
    - crystal water molecules
    - metal coordination type
- the addition of hydrogen atoms to fill out the remaining open valences of the receptor.

3. Docking and scoring

The docking algorithm predicts the active poses of ligand. Both position and conformation are chosen based on the geometric and physico-chemical complementarity between the ligand and the receptor's binding site.

The ligand poses are ranked according to a docking score. The score is a rough measure of the fit of a ligand into the binding site.

The docking results are:

- the structure file of the best ligand poses
- the score of each pose

B. Goal of the tutorial: understanding the docking paradigm

The quality of ligand and protein preparation directly impacts the outcome of the pose prediction. The tutorial contains *re-docking** exercises to investigate the effect on docking of the ligand ionisation state and the water in binding site (E.1 section).

The docking accuracy obviously depends on the algorithm and its parameters, which are generally well documented by developers and vendors. Special care has to be given to the conformational sampling, which must suit the ligand flexibility. This issue is not examined in the tutorial.

Whatever the algorithm/program, the current limitations of the docking approach are due to:

- an unusual binding mode (e.g., poor geometrical complementarity).

  The tutorial deals with a tricky target, namely *acetylcholinesterase*. The *re-docking** exercises in section E.1 exemplify the need of expert intervention to successfully predict the true binding mode using automated docking.

- the protein binding site flexibility

  The tutorial contains a *cross-docking** exercise with induced fit by variation of rotameric state (E.2 section).

- the low accuracy of scoring functions, especially in compound ranking

  The tutorial contains a *screening** exercise to discriminate protein binders from decoys (E.3 section).

*glossary	
re-docking	docking of the ligand extracted from the X-ray structure of a ligand/protein complex back into its co-crystal protein binding site.
cross-docking	docking of the ligand extracted from the X-ray structure of a ligand/protein complex into the protein binding site extracted from another X-ray structure (apo- or holo-form of the same protein)
screening	a library of ligands is docked into a protein binding site and the compounds are ranked according to the docking score.

C. Instructions: leadIT quick start

The program leadIT 2.1.3 uses FlexX technology and has a user-friendly graphical interface to prepare, launch and analyse the docking run.

Protein preparation	`Molecules >> Prepare Receptor...`
	Select the protein PDB file and follow the instructions
Ligand preparation	`Molecules >> Choose Library...`
	Load the MOL2 file. In the tutorial, we will use ready-to-dock ligands. Un-tick the box `Protonate as in aqueous solution`.
Docking	`Docking >> Define FlexX Docking...`

Note about leadIT :

- **Protein preparation**

  The necessary steps in the protein preparation flowchart are automatically set by LeadIT after the analysis of the uploaded protein file.

  The binding site is made of all residues close enough from a reference ligand (this approach assumes that the 3D structure of a ligand/protein complex is known) or of all residues within a sphere defined by a centre atom and a radius. Other programs allow the direct detection of cavities and pockets in the structure of apo-proteins.

  Three types of crystallographic water molecules can be included in receptor:

- « *oriented water* »: hydrogen atoms are fully defined (icon water molecule)

- « *freely rotatable water* »: freely spinning molecule (icon light blue sphere)

- or « *freely rotatable, displaceable water* »: a ligand can displace the molecule (icon grey blue sphere).

The *'confirm receptor'* step automatically adds hydrogen atoms to fill out open valences.

- **Ligand preparation**

The ligand preparation tool of LeatIT is able to deal with a reference ligand or to load an external file in format MOL, MOL2 or SDF. A reference ligand is used when the solution of the docking is know in order to assess its success. An attempt is made to propose a reliable version of the protonation state of the ligand. The graphical user interface is limited to 1000 compounds. To deal with larger libraries, it is necessary to use the command line interface, which is out of the scope of this tutorial.

- **Docking**

The docking popup menu allows the user to choose the base placement method: *Entropy (Single Interaction Scan)* for hydrophobic pockets or pockets with only a few interaction sites, *Enthalpy (Triangle Matching)* for pockets containing a number of interaction sites, *Enthalpy and Entropy* a consensus of both used by default.

The docking popup menu allows the user to downscale the score of ligand poses docked at the rim of the binding pocket (*Access scaling scoring* option, which is switch on by default).

D. Course material

The **input** directory contains:

- the **pdb** directory

  pdb1acj.ent   1acj PDB entry

- the **receptor** directory

  1acj_WAT.mol2   1acj prepared receptor, including hydrogen atoms, with two water molecules

  1eve_ali_WAT.mol2   1eve prepared receptor, including hydrogen atoms, with one water molecule.

                   The protein coordinates have been 3D-aligned to 1acj protein.

- the **ligand** directory (the ligand structures are given in Figure 1)

  TAH_1acj.mol2   neutral tacrine, extracted from 1acj PDB entry

  TAH_1acj+.mol2   (+) charged tacrine, extracted from 1acj PDB entry

  A2E_1zgc.mol2   tacrine-hupyridone inhibitor, extracted from 1zgc PDB entry, whose

                   coordinates have been 3D-aligned to 1acj.

  E20_1eve.mol2   aricept, an anti-alzheimer drug, extracted from 1eve PDB entry, whose

                   coordinates have been 3D-aligned to 1acj.

  DUD.mol2   the D.U.D AchE dataset (http://dud.docking.org/)

The **flexx** directory contains:

- results of exercise E.1.

mol2 / sdf / csv / fxx files:

1acj_TAHsite65_TAHredock   re-docking of TAH into TAH 6.5 Å site in 1acj

1acj_TAHsite65_TAH+redock   re-docking of TAH+ into TAH 6.5 Å site in 1acj

1acj_TAHsite65WAT_TAH+redock   re-docking of TAH+ into TAH 6.5 Å site in 1acj, with water

- results of exercise E.2.

mol2 / sdf / csv / fxx files:

`1acj_A2Esite65WAT-A2Ecrossdock` cross-docking of A2E into AE2 6.5 Å site in 1acj, with water

`1acj_E20site65WAT-E20crossdock` cross-docking of E20 into E20 6.5 Å site in 1acj, with water

`1eve_E20site65WAT_E20redock` re-docking of E20 into E20 6.5 Å site in 1eve, with water

- underline results of exercise E.3,

mol2 / sdf / csv files:

`1acj_A2Esite65WAT_DUDscreen` score of all DUD ligands docked into A2E 6.5 Å site in1acj, with water
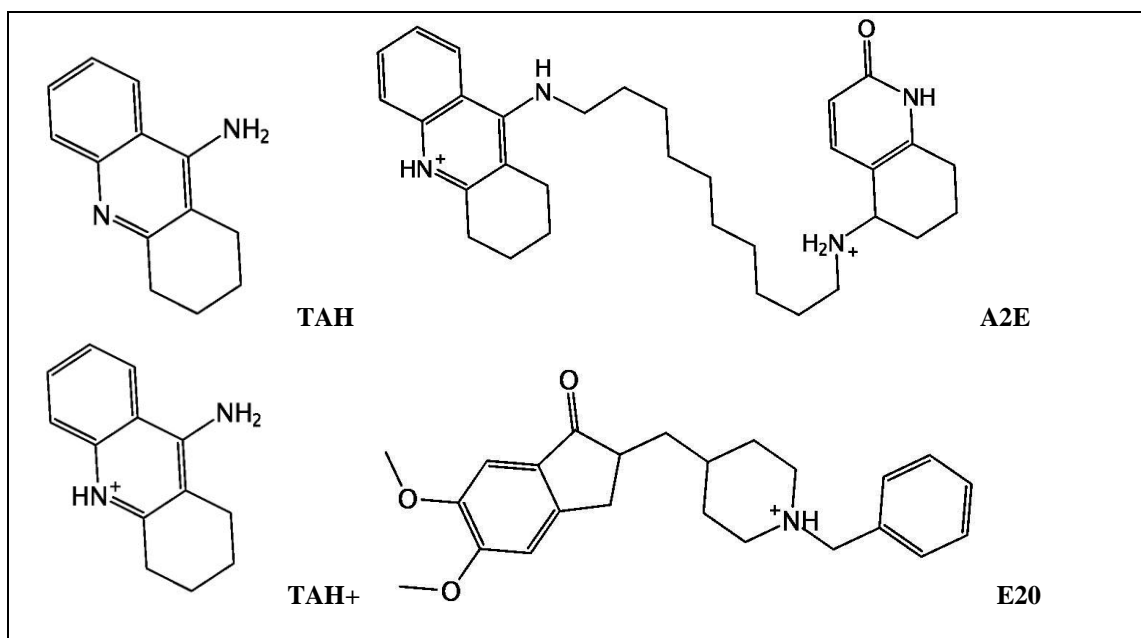
Figure 1: Structures of neutral tacrine (TAH, top left), positively charged tacrine (TAH+, bottom left), tacrine-hupyridone inhibitor (A2E, top right) and aricept (E20, bottom right)

E. Exercises: docking ligands into acetylcholinesterase

1. Re-docking tacrine (TAH) back into the acetylcholinesterase binding site

1.1. Prepare the receptor and define a 6.5A site around tacrine

Load protein	`input/pdb/pdb1acj.ent`
Choose receptor component	☒ `chain A`
Define binding site	reference ligand `THA-999A`
	include amino acids within 6.5 Å radius
resolve chemical ambiguities	In the `Residues` menu, none of the automatically selected residues establishes H-bond or ionic bond with tacrine. The orientation of polar hydrogen is thus defined in order to avoid polar interaction with ligand during docking. See values in table 1. In the `Waters` menu, un-tick all water molecules so that none of them are included into the receptor definition. Do not edit the reference ligand that can be found in the `Small Molecules` menu.
Confirm receptor	name the receptor `1acj_TAHsite65`

Amino acid	Protonation state	torsion	Comment
ASP72A	asp1	0	Side chain not in the pocket
SER81A	ser	180	default
ASN85A			default
TYR121A	Tyr	0	default
SER122A	ser	330	H-bond to a crystal water molecule
TYR130A	Tyr	300	H-bond to a crystal water molecule
GLU199A	Glu-		default
SER200A	Ser	120	H-bond to HIS440A N-epsilon nitrogen
TYR334A	Tyr	150	H-bond to ASP72A carboxylic acid
HIS440A	His1		H-bond to SER200 hydroxyl group
TYR442A	Tyr	180	Polar hydrogen does not point towards the cavity

Table 1: Proposition to resolve chemical ambiguities in the TAH site of 1acj receptor.

1.2. Load and dock the neutral tacrine

Load ligand file `input/ligand/TAH_1acj.mol2`. Do not tick the box `Protonate as in aqueous solution.`
Dock using default parameters, except in the `General Docking information` menu chose `Number of pose to keep    top10`
Analyse results to fill table 2.

1.3. Load and dock the positively charged tacrine

Load ligand file `input/ligand/TAH_1acj+.mol2`
Dock using default parameters, except in the `General Docking information` menu chose `Number of pose to keep    top10`
Analyse results to fill table 2.

1.4. Include water in the receptor, load and dock the positively charged tacrine

Go back to the `'Resolve chemical ambiguities'` step of the receptor preparation

Water `HOH-634-A` and `HOH-643-A` marked as freely rotatable water
Confirm receptor name the receptor `1acj_TAHsite65_WAT`
Load ligand file `input/ligand/TAH_1acj+.mol2`
Dock using default parameters, except `Number of pose to keep  top10`
Analyse results to fill table 2.

1.5. Results and conclusion

| site	ligand	Top pose (best score)		Best pose (best RMSD)	
		score	RMSD (Å)	Score	RMSD (Å)
1acj_TAH_site65	TAH_1acj				
1acj_TAH_site65	TAH_1acj+				
1acj_TAH_site65_WAT	TAH_1acj+				

Table 2: Re-docking of tacrine into acetylcholinesterase

No correct docking solutions are found for the neutral form of tacrine. The pka of tacrine is 9.95, hence the ligand has most probably a formal charge of +1 at physiological pH. The experimental binding of the positively charged tacrine is well predicted by the docking program, but several poses with very similar scores are far from the native one. The presence of two water molecules in the binding site yields the systematic correct placement of the positively charged tacrine.

The tacrine/ acetylcholinesterase binding mode is difficult to predict, because:
- the size of the binding pocket largely exceeds the volume occupied by the ligand
- there is only one directional polar interaction between the ligand and the protein: the H-bond between the protonated pyridine of the positively charged tacrine and the backbone carbonyl group of His440 (Figure 2).
- two water molecules H-bond bridge the ligand amine group to the receptor (Figure 2)

The docking of the neutral form of tacrine into acetylcholinesterase failed because of the fair geometrical complementarity and the lack of directional polar interactions between the two molecules. The docking of the positively charged form of tacrine was facilitated by the anchoring of the ligand in the tip of the protein pocket by the single direct H-bond. Last, the two water molecules restrict the available space during the ligand placement and provide two additional H-bond donors at key positions.
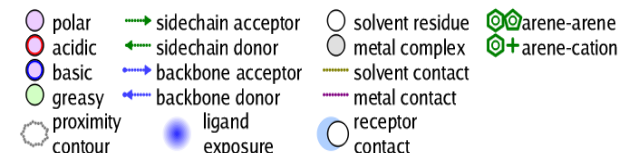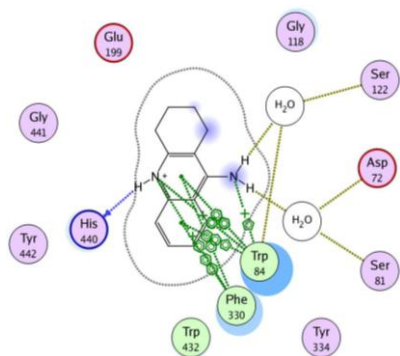
Figure 2: Experimental binding mode between tacrine and acetylcholinesterase (from 1acj PDB entry). The picture was prepared using MOEv2009.10.

2. Cross-docking of A2E and E20 inhibitors into the acetylcholinesterase

2.1. Load the 1acj receptor and define a 6.5A site around A2E

Load receptor
`input/receptor/1acj_WAT.mol2`
All chemical ambiguities of 1acj PDB file have been resolved (see Table 1). The two water molecules have been included in the file.

Define binding site
using reference ligand from `input/ligand/A2E_1zgc.mol2`
include amino acids within 6.5 Å radius

Confirm receptor
name the receptor `1acj_A2Esite65WAT`

2.2. Load and dock A2E

load ligand file `input/ligand/A2E_1zgc.mol2`. Do not tick the box `Protonate as in aqueous solution.`
Dock using default parameters, except `Number of pose to keep    top10`
Analyse results to fill table 3.

2.3. Define a 6.5A site around E20 and dock E20

Define binding site
using reference ligand from `input/ligand/E20_1eve.mol2`
include amino acids within 6.5 Å radius
The newly binding site perfectly matches the binding site defined around A2E in section 2.1, but the reference ligand was changed to allow the RMSD calculation between X-ray coordinates and docked poses.

Confirm receptor
name the receptor `1acj_E20site65WAT`

Load ligand file
`input/ligand/E20_1eve.mol2`. Do not tick the box `Protonate as in aqueous solution.`

Dock using default parameters, except `Number of pose to keep    top10`
Analyse results to fill table 3.

2.4. Load the 1eve receptor and define a 6.5A site around E20 and dock E20

Load receptor	`input/receptor/1eve_ali_WAT.mol2`
Define binding site	using reference ligand from `input/ligand/E20_1eve.mol2`
	include amino acids within 6.5 Å radius
Confirm receptor	name the receptor `1eve_E20site65WAT`
Load ligand file	`input/ligand/E20_1eve.mol2`

Dock using default parameters, except `Number of pose to keep    top10`

Analyse results to fill table 3.

2.5. Results and conclusion

| site	ligand	Top pose		Best pose	
		score	RMSD (Å)	Score	RMSD(Å)
1acj_A2E_site65WAT	A2E_1zgv				
1acj_E20_site65WAT	E20_1eve				
1eve_E20_site65WAT	E20_1eve				

Table 3: Re-docking and cross-docking of inhibitors into acetylcholinesterase

The good pose of A2E inhibitor is ranked first by the scoring function. The tacrine substructure of the A2E ligand is well docked into the acetylcholine esterase whereas the pyridone part could not be correctly placed in the protein rim because of the unsuitable rotamer of Trp279.

The experimental binding mode of E20 inhibitor into acetylcholinesterase can not be reproduced using 1acj PDB entry as receptor file because of the unsuitable rotamer of Phe330, whose side chain actually fills the space required for ligand piperidine group.

The re-docking of E20 back into the acetylcholinesterase binding site extracted from 1eve PDB entry is not satisfactory since an acceptable pose is not ranked first. The experimental binding mode is indeed very difficult to predict because the ligand does not establish any H-bonds or ionic bonds with the protein, although the ligand is positively charged and its binding site contains negatively charged residues (Figure 3).
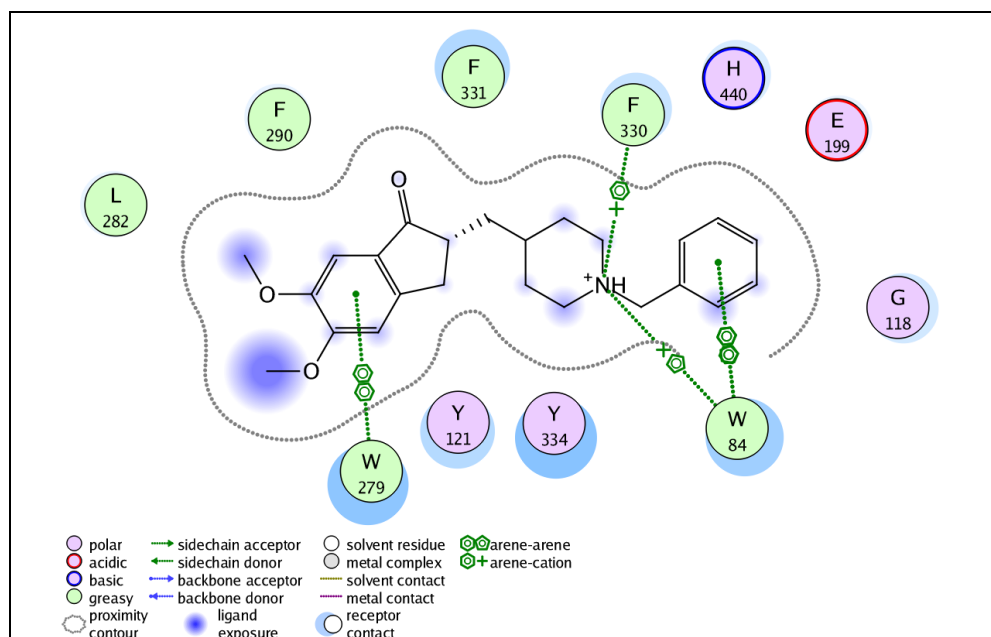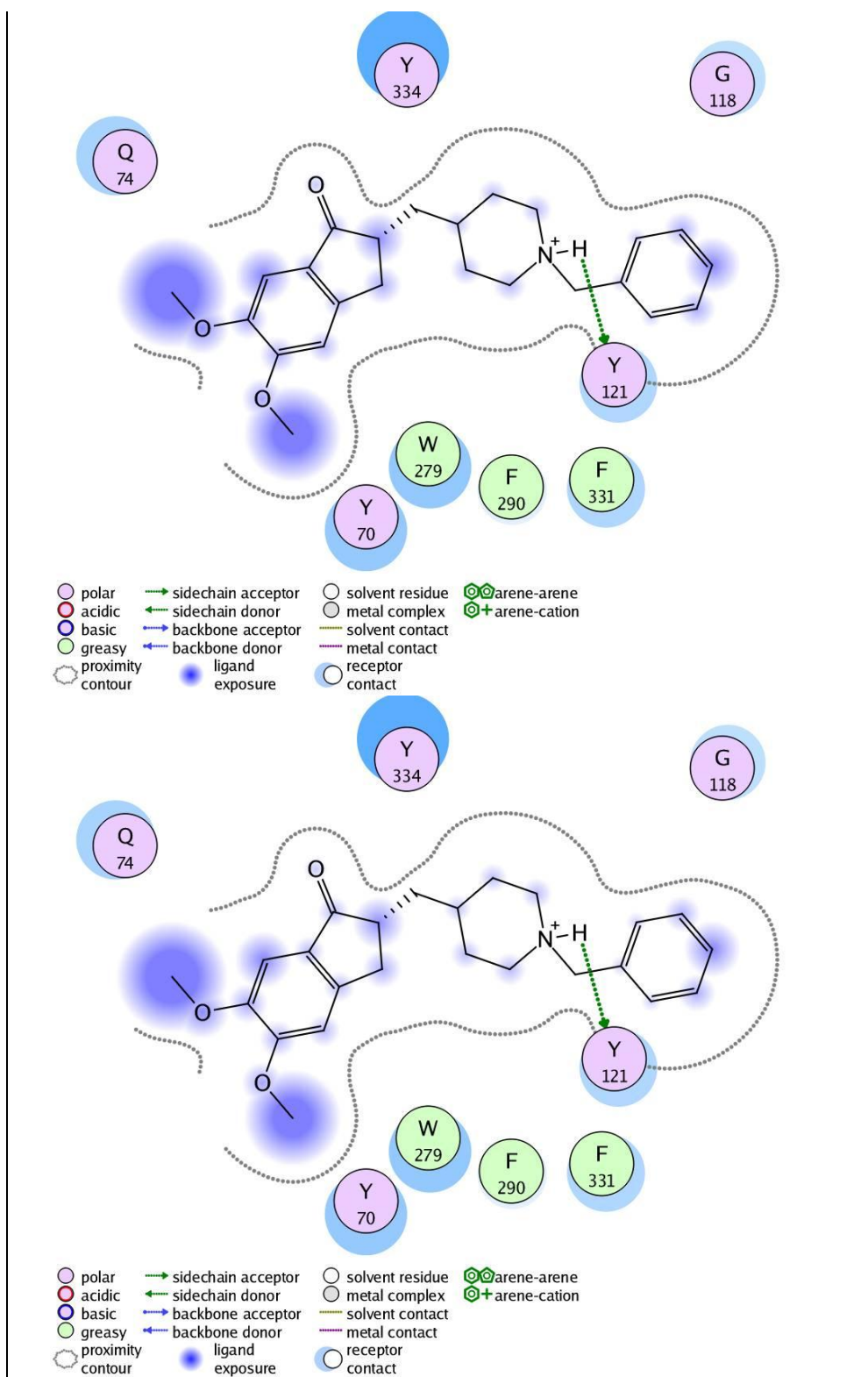
Figure 3: Experimental (1eve PDB file, top panel) and predicted (FlexX re-docking, bottom panel) binding mode between E20 and acetylcholinesterase. The pictures were prepared using MOEv2009.10.

3. Screening the DUD dataset

3.1. The DUD dataset

The DUD AchE dataset contains 107 true binders and 3892 decoys. About half of the active compounds are similar to E20, according to tanimoto coefficient computed using MACCS keys (19, 28 and 50 compounds yield values larger than 0.8, 0.7 and 0.6, respectively). Only two active compounds are similar to tacrine, namely tacrine itself and a second compound with a similarity of 0.62 (tanimoto coefficient computed using MACCS keys).

The screening of the DUD database by docking into acetylcholinesterase was published by Huang, Shoichet and Irwin in 2006 (DOI 10.1021/jm0608356). The docking was performed using DOCK3.5.54. The acetylcholinesterase was prepared without expert intervention from the 1eve PDB file. The binding site was defined around the co-crystal ligand E20. Noteworthy is the Phe330 conformation not compatible with tacrine binding.

The enrichment factors obtained by ranking compounds using the total docking energy are reported in Table 4.

3.2. Load the 1acj receptor and define a 6.5 Å site around A2E and dock the DUD database

Do not perform this task during the tutorial. It may take over 5 hours to be done. You shall find all results in the file `flexx/1acj_A2Esite65WAT-DUDscreen.csv.`

Load receptor	`input/receptor/1acj_WAT.mol2`
Define binding site	using reference ligand from `input/ligand/A2E_1zgc.mol2` include amino acids within 6.5 Å radius
Confirm receptor	name the receptor `1acj_A2Esite65WAT`
Load ligand file	`input/ligand/DUD.mol2`
Dock using default parameters, except	`Number of pose to keep      top1`
Analyse results to fill table 4.	

3.3. Results and conclusion

	Top 1%	Top20%
True positive (ACTIVE) rate, TPrate	…. / 107 =	…. / 107 =
False positive (DECOYS) rate, FPrate	…. / 3892 =	…. / 3892 =
Enrichment factor	(TPnumber / 40) ------------------ = (107 / 3999)	(TPnumber / 800) ------------------ = (107 / 3999)
Enrichment factor from DOI 10.1021/jm0608356	1.9	2.0

Table 4: Screening DUD dataset by docking into acetylcholinesterase

From exercise E.2 results, we can assume that the poor performance in screening of Huang and co-workers is due to the failure of DOCK to predict the unusual binding mode of acetylcholinesterase to E20 and similar compounds, and also to tacrine and similar compounds.

The present screening exercise, which was carried out using a receptor prepared with expert intervention (choice of 1acj entry, careful protonation of the binding site residues and inclusion of key water molecules), has slightly more success in the retrieval of true active among the top rankers.