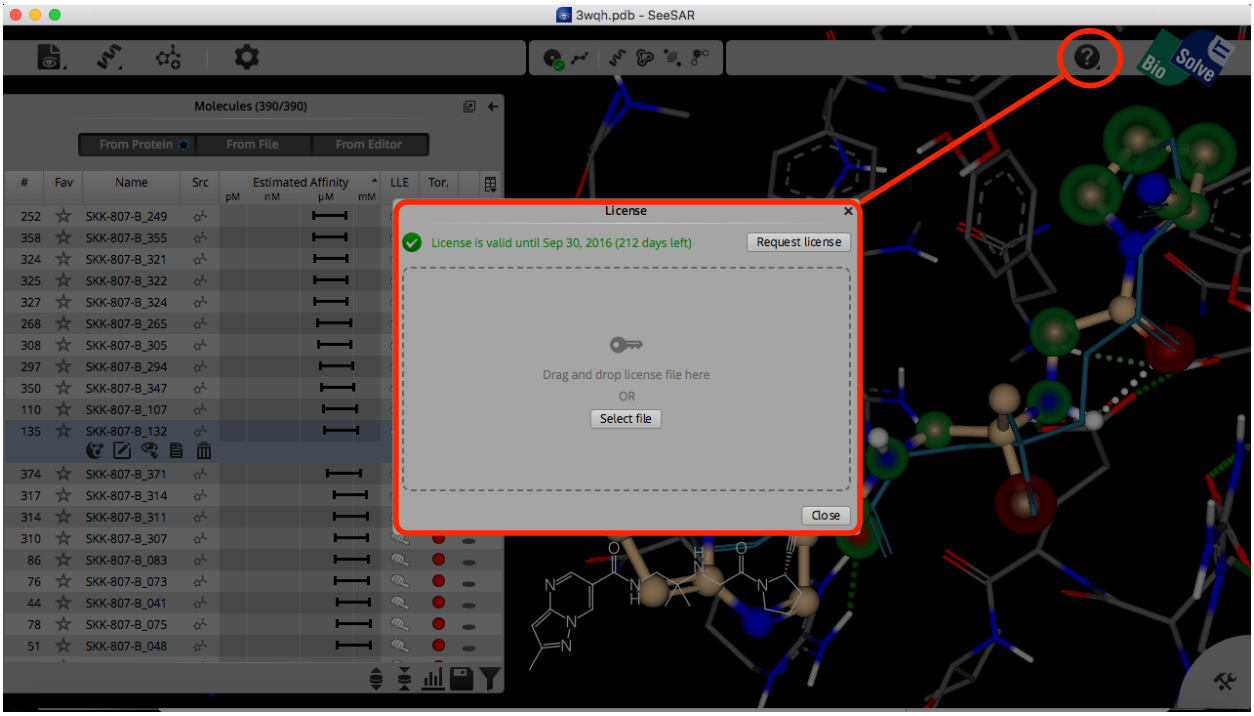
Seamless integration of 2D and 3D SAR to guide multi-parameter optimization

Workshop organized by BioSolveIT and Optibrium Ltd.

Preparation

You should have installed SeeSAR on your laptop. The license file should be installed. You can check whether you have a license when you open the information:



Check if you have all data files needed for the workshop:

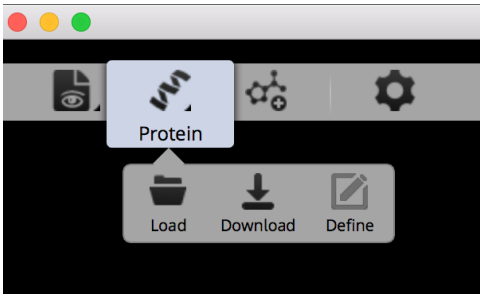
1. PDB files 3IPH, 4L6B, 3PSY, 2XJX, 2ZC9
2. SeeSAR project files 3IPH.seesar, 4L6B.seesar, 2XJX.seesar and 2ZC9.seesar
3. Molecule files 3PSY_ligand_sdf, thrombin_aligned_ref_ligs_with_properties.sdf

Example 1 - Analyzing and editing a ligand with a magic methyl

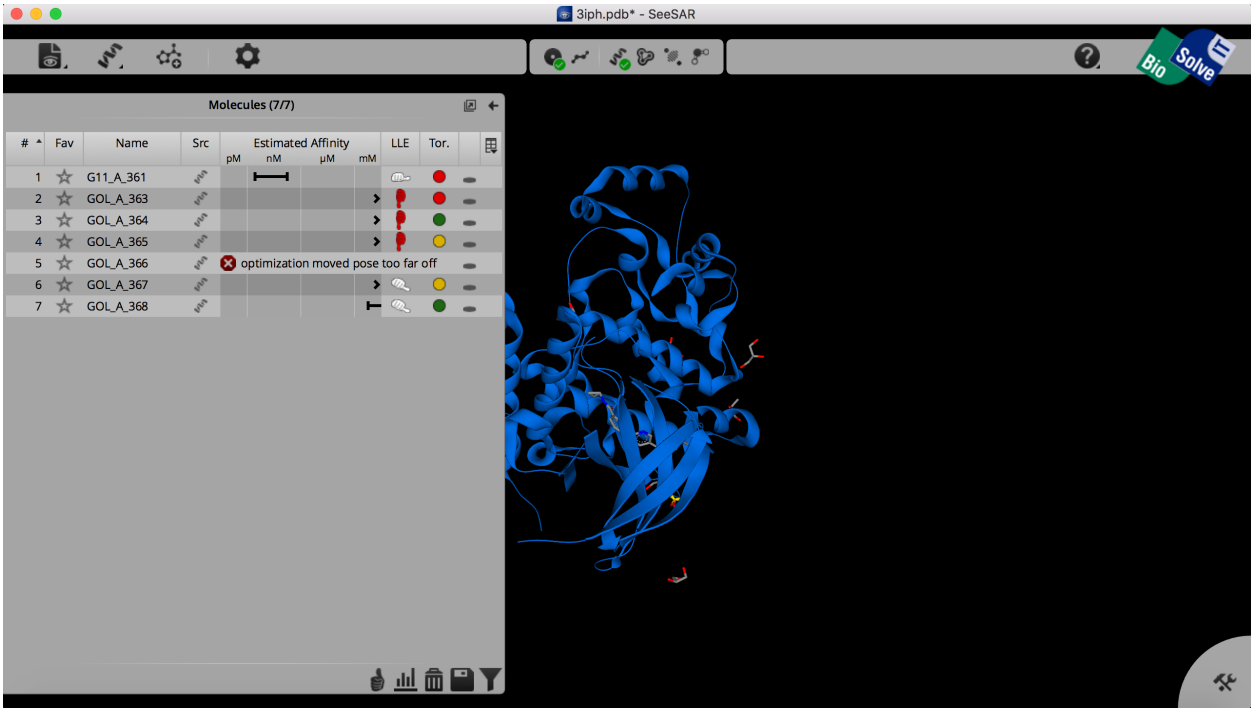
The example is taken from a publication[1] in which the importance of methyl groups on the affinity in some examples is discussed. It is a p38-MAP kinase in complex with a biphenylamide inhibitor.

We will use this example to show the editing functions in SeeSAR, and how the information about affinity, torsions, and other properties can be interpreted.

When you start up SeeSAR, enter the PDB code 3IPH into the box on the left ("Download a PDB"), or, from the menu, choose to download a structure:

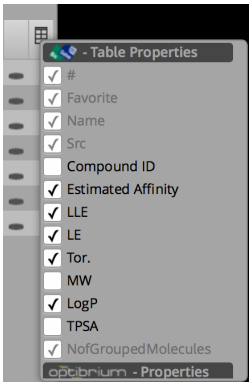


SeeSAR will pull the PDB file directly from the Protein Data Bank.
When it has successfully downloaded, it will score the co-crystallized ligands and display them in the table to the left.



Insert

The table contains useful information about the ligands such as the name, where the ligands come from (Src: Protein, File, or Editor), its estimated affinity. Other properties can be added (or left out) by clicking on the table symbol in the upper right corner of the table:



The LLE (ligand lipophilic efficiency), LE (ligand efficiency) and torsions are displayed qualitatively through symbols: the LLE and LE though thumbs up-down, the torsions through a traffic light (red, yellow, green)

The LLE is calculated as follows: ***LLE = pK_i-(c)logP***

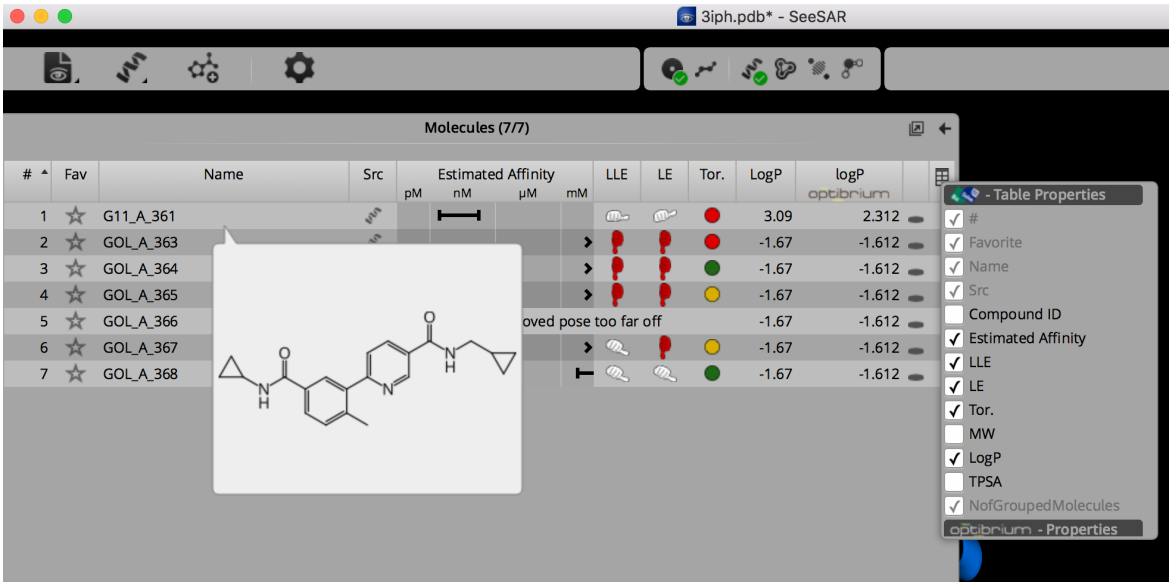
LE is calculated as follows: ***LE = ΔG/#HA***

The torsions are colored based on a statistical evaluation of the CSD. Please consult the corresponding paper[2], should you like to know more about how exactly the torsion information has been determined.

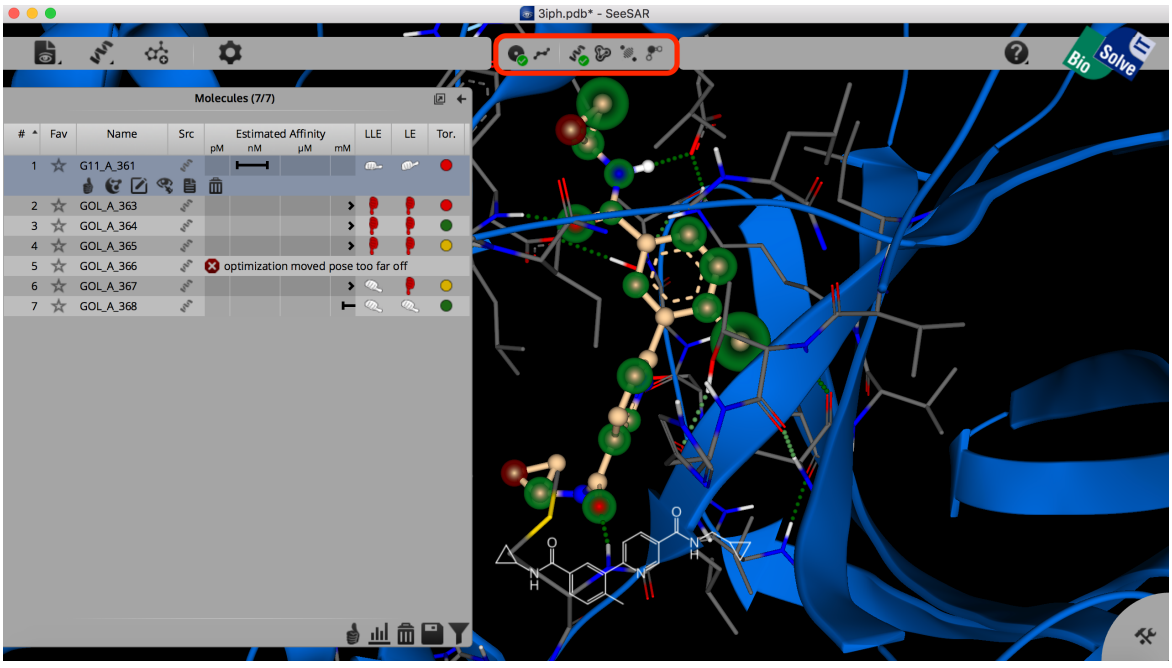
Insert End

[2] *J. Med. Chem.*, 2013, 56 (5), pp 2016-2028

When you hover over the name in the table, the 2D structure will be shown.



Click on the entry “G11_A_361”. This will show a more detailed picture of the binding affinity, broken down on an atom level. You can zoom in by pressing space.



The visualization can now be adapted according to what you want (see red circled buttons in the figure above, from left to right: Hyde affinity coronas, torsion information, protein backbone cartoon, surface, fog [more on that later], waters).

You now see the ligand with its corresponding Hyde affinity coronas. Each atom gets a score, whether it contributes favorably (green) or unfavorably (red) to the overall affinity.

Right-click on an atom and then moving your mouse uses this atom as center of rotation.

Insert

*This is how Hyde (for **hy**drogen bonding and **de**solvation) calculates affinity:*

$$\Delta G_{Hyde} = \sum \Delta G^{i}_{dehydration} + \Delta G^{i}_{H\text{-}Bond}$$

where

$$\Delta G^{i}_{dehydration} = -2.3\ RT \cdot \text{plogP}^{i} \cdot \Delta acc^{i}$$

and

$$\Delta G^{i}_{H\text{-}Bond} = \frac{2.3RT}{F_{sat}} \cdot \text{plogP}^{i} \cdot \Delta sat^{i}$$

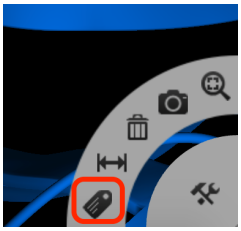
Both the dehydration and the hydrogen bond energy are very consistently described through the plogP atomic increment. Δacc^{i} ist the change in Connolly surface, Δsat^{i} the number of hydrogen bonds formed with the receptor. F_{sat} is the saturation factor, which describes the completeness of the hydrogen bond network. This factor is temperature dependent.

For more details about the derivation and theory, please consult the Hyde publication.[3]

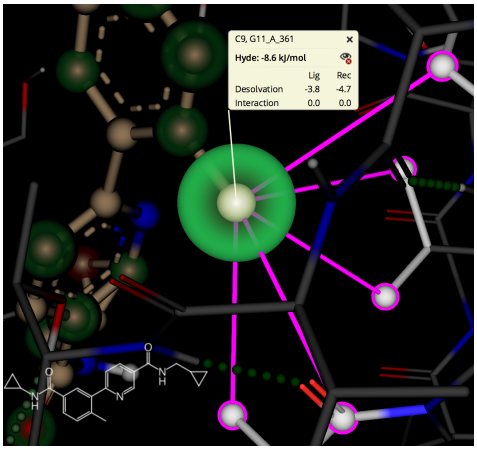
Insert end

[3] J Comput Aided Mol Des. 2013;27(1):15-29

In the lower left corner, you can choose to put labels on each atom to get a more quantitative view on each atom's binding situation. Other symbols are, from 9 to 12 o'clock: measure distance, delete all labels, screenshot (without the table), focus view (same as space)



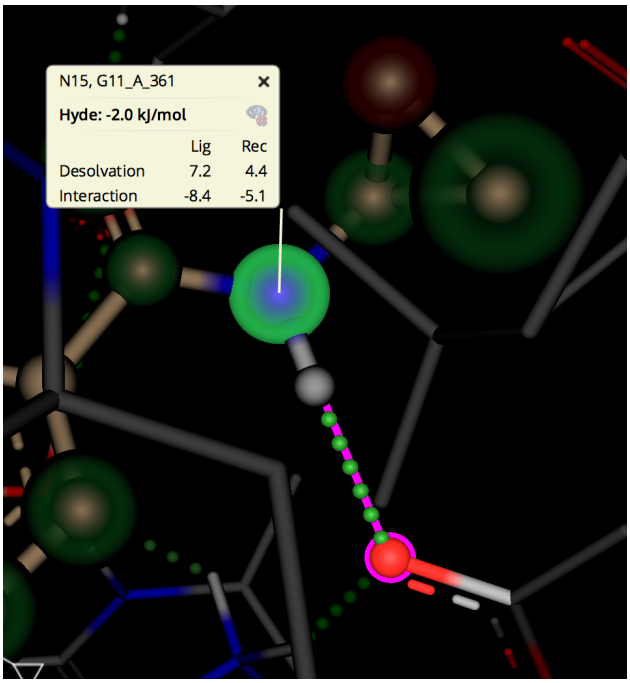
Once you press the label (circled in red), the mouse pointer will change, and when clicking on an atom, you will see a label:



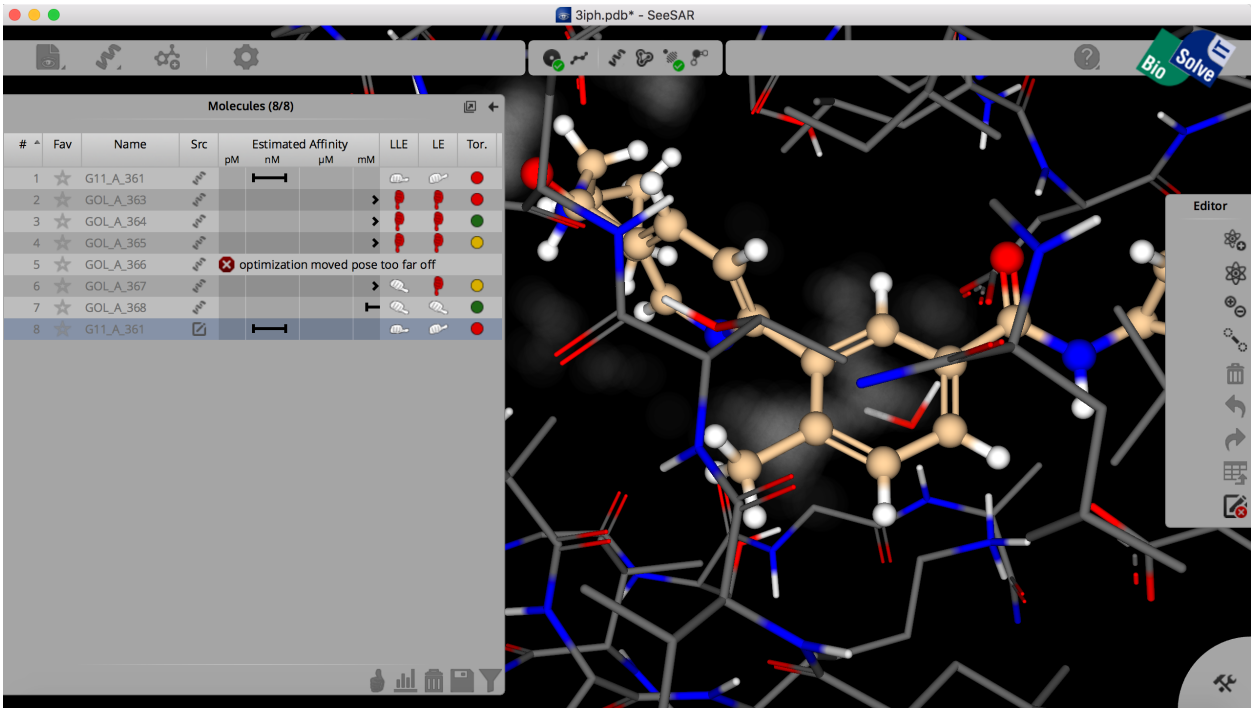
The label tell you which atom you are looking at (C9 of the ligand in focus), the overall Hyde score (-8.6 kJ/mol), and the score broken down onto ligand and receptor Desolvation and Interaction (h-bond) energy.

In the case we are looking at above, it is exclusively desolvation energy that is contributing, from the ligand (-3.8 kJ/mol) and from the receptor (-4.7 kJ/mol). By clicking on the eye symbol, interaction with the receptor is displayed. The methyl group interacts with five protein residues.

For a hydrophilic atom, the situation looks as follows:



The N15 of our ligand in focus has to pay a desolvation penalty when brought from water into the protein (7.2 kJ/mol), and the protein backbone oxygen also pays a penalty when dehydrated (4.4 kJ/mol). If there was no interaction (h-bond) energy, the Hyde affinity corona would be large and red. Due to the hydrogen bonding between the two atoms though, the desolvation loss in energy is *overcompensated* (-8.4 kJ/mol on the ligand and -5.1 kJ/mol on the protein side). The geometry of the h-bond is good, because the dashed line has a dark green color (bad geometries are displayed through lighter green or even white dashed lines).



Insert

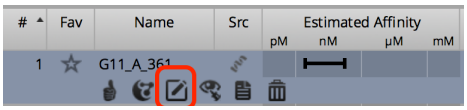
Hyde is very picky on the geometry. If we are dealing with bad PDB structures, and the hydrogen bond geometries are compromised, we will likely not get the full energetic contribution that we need to overcompensate the loss due to desolvation.
This seems like a weakness, but is what we observe in "real" hydrogen bonds as well.

Insert end

Based on the information the affinity coronas give us, we can see directly, without consulting any numbers, that the methyl group is the largest contributing group on the ligand with respect to desolvation.
This is the concept of **visual feedback**: without much interpretive work, it can be readily assessed which parts of the ligand contribute well, and which don't.

First Editing Exercise

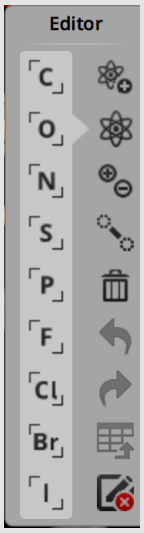
Let's now move to the fun part, the editing process. In order to edit a ligand, move your mouse pointer to the table and click the ligand you would like to edit (in this case, we have done so already). You will see a bunch of symbols underneath the table entry:



These symbols are, from left to right:
Calculating the Hyde affinity (from release 4.2 on; ligands imported are not automatically calculated anymore); FlexX docking (only the first 10 poses are kept, no parameter settings); **edit a ligand;** keep ligand as reference (will be colored blue); annotate a ligand; delete ligand.

Clicking on the edit button will bring our ligand into the edit mode:

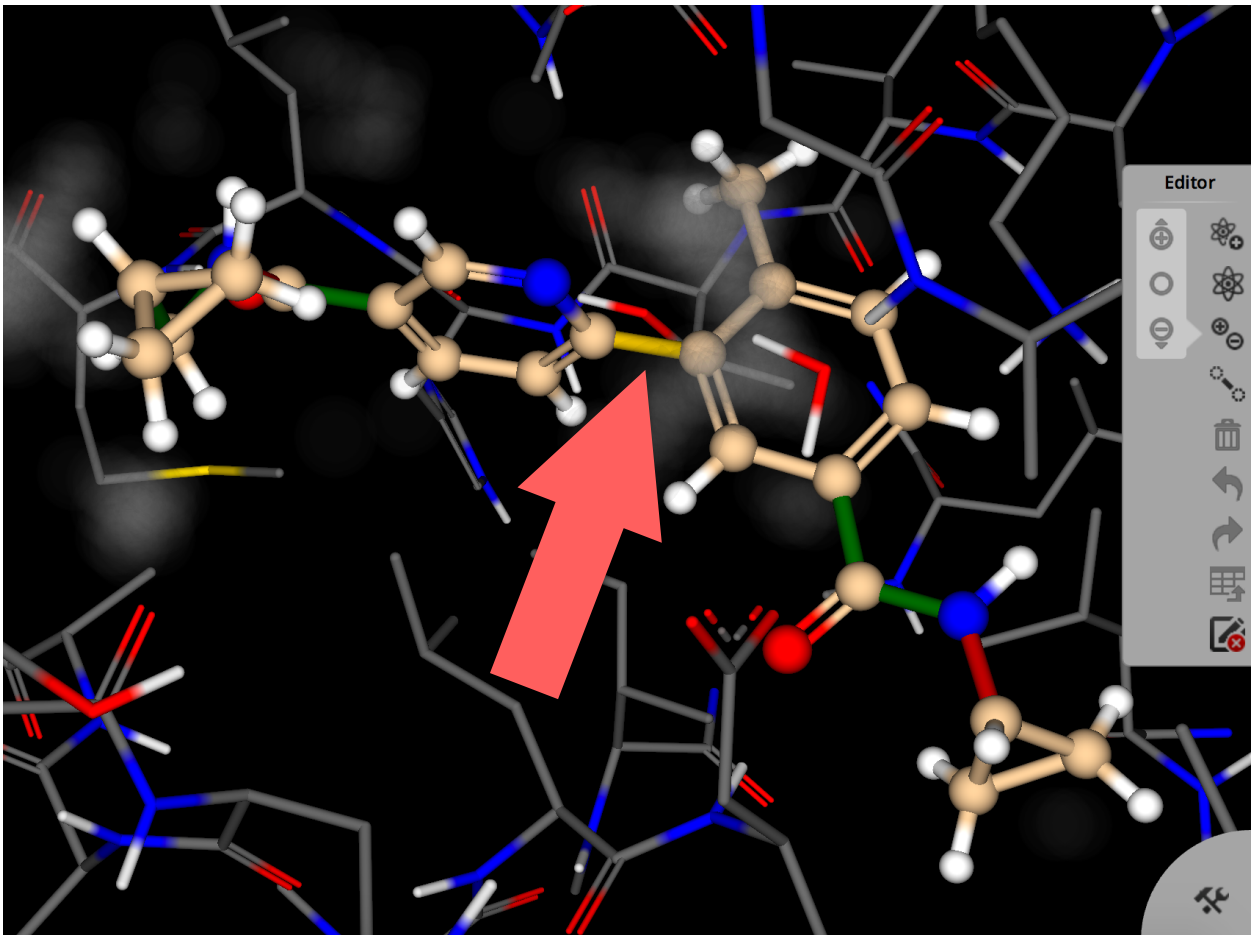
The editor will open on the right hand side of the window, and you can activate the “fog”, which shows unoccupied space in the vicinity of the ligand, where at least on fluorine atom will fit.

	
	Extending by -C, =O, or #N
	Change element to…
	Change charge
	Change bond type
	Delete (only selected)
	Undo
	Redo
	Save to table (editor remains open)
	Leave editor

Before we edit, turn on the torsions from the buttons above:

All rotatable bonds will now have a color red, yellow or green:
Green torsions are very well represented in the CSD (Cambridge Structural Database), yellow torsion occur occasionally, and red (almost) never.

Note that the torsion between the phenyl and pyridine ring is yellow.



Let us start to modify the ligand by removing the magic methyl and see what happens.

- Select the carbon C9 of the methyl group. It will turn pink
- Click the trash can in the editor.
- You will notice three things:
  - More fog will appear in the spot where the methyl group was
  - The confirmation of the ligand will change
  - The torsion between the two rings will change from yellow to red

Insert

In order to score the ligand most effectively, we need to optimize the ligand. This is done in two steps:

1. *Optimize the hydrogen bond network around the ligand and with the protein*

This is done by a method called ProToss.[4]

[4] Bietz et al. Journal of Cheminformatics 2014, **6**:12

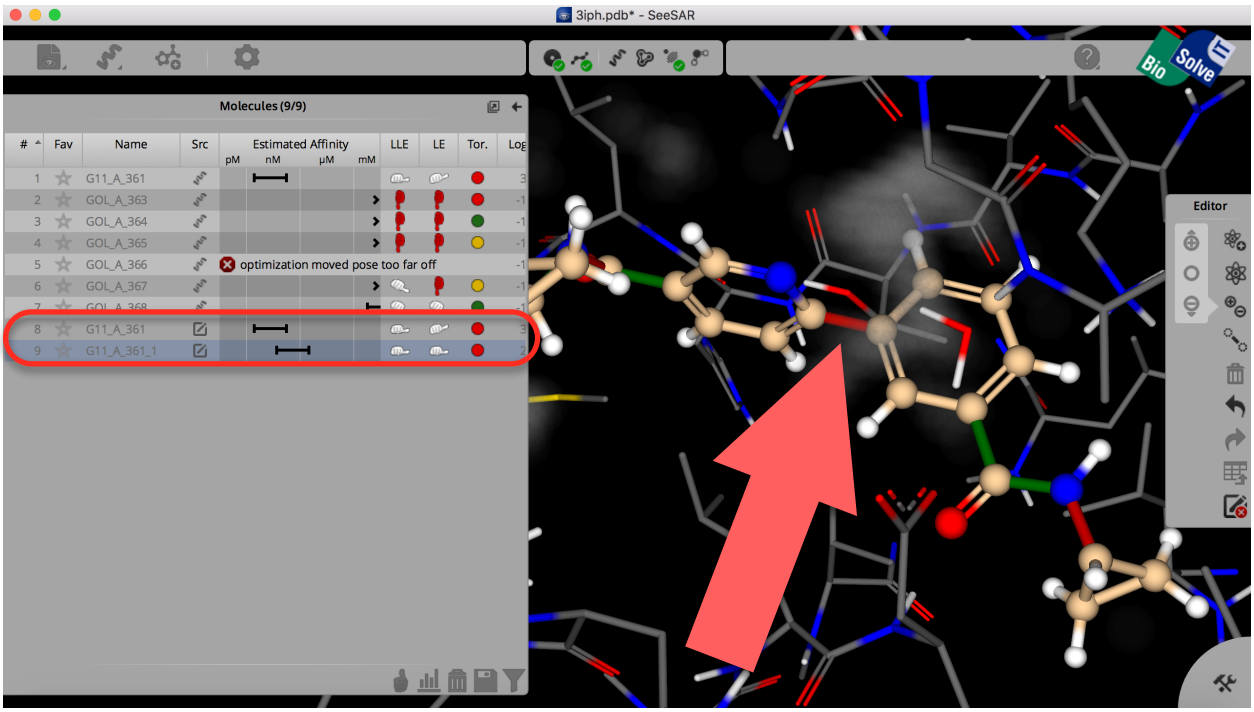
2. *Optimize the ligand's conformation and position to get the best possible hydrogen bond geometries*

$$\begin{aligned} HYDE_{opt} = & w_{LJ} E_{\text{LJ-inter}} \\ & + w_{LJ} E_{\text{LJ-intra}} \\ & + w_{HB} \Delta G_{\text{H-Bond}} \\ & + w_{HP} \Delta G_{Dehydration} \end{aligned}$$

The underlying force field consists of a 6-12 LJ potential for inter and intramolecular interactions, a weighted hydrogen bond function and a dehydration function.

Insert end

Now click the "Save to table" button in the editor (see table above for reference).
The newly edited ligand will be optimized and scored based on Hyde's scoring function. You will notice that the affinity has changed for the worse, something that was of course to be expected.

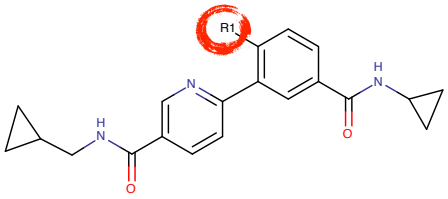


You can also see that the LLE remained the same, whereas the LE changed slightly for the worse. Since we already had one red torsion before, the change did have no effect on the overall torsional quality.
Leave the editor by clicking on the corresponding button (refer to table above)

Please note that the change in affinity in SeeSAR comes only from the removed methyl group, because its desolvation is gone. The change in torsion has no effect on the affinity in SeeSAR (it does have, however, have a dramatic effect in reality, as the two rings are locked into a very favorable conformation, which mainly drives the affinity in reality. Desolvation adds too, but much less so). The change in torsional quality is mainly a statistical one.

AP (Advanced Placement) Course:

Modify the magic methyl (C9) to a F, Cl and O-Me. What happens?



#	Fav	Name	Estimated Affinity (pM nM µM mM)	LLE	Ki (nM)
1	☆	Me			12
2	☆	H			2500
3	☆	F			25
4	☆	Cl			460
5	☆	OMe			520

W can see that the calculated pretty much follows the experimental affinity (with exception of F).
What we are after, are qualitative changes to a reference state (in this case it would be the #2). Everything better than the reference state is interesting!

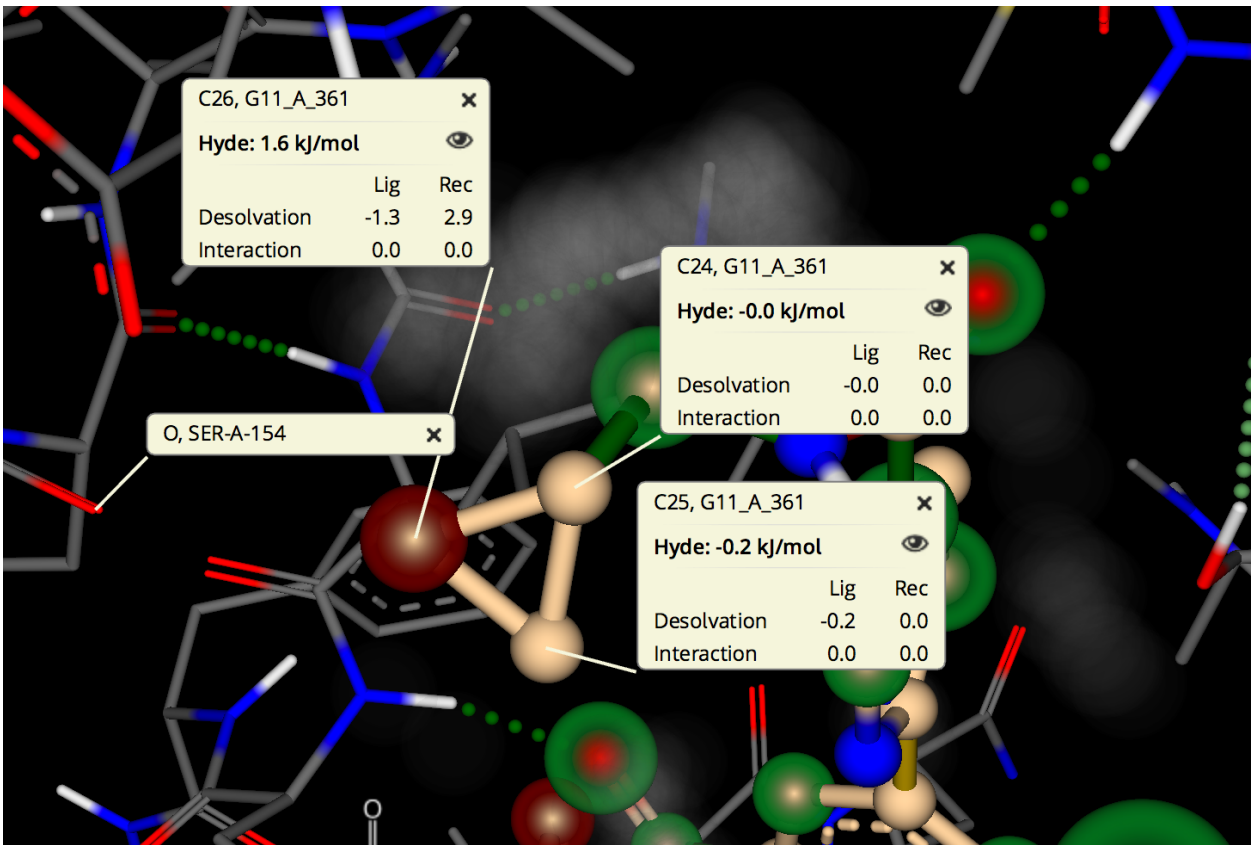
Congratulations! You have successfully edited your first ligand in SeeSAR!

Second Editing Exercise

Of course, in the real world, we do not want to destroy affinity, but improve it. Let us therefore now attempt to find a modification that improves affinity.

Go back to the initial, co-crystallized ligand (G11_A_361) and click on it.

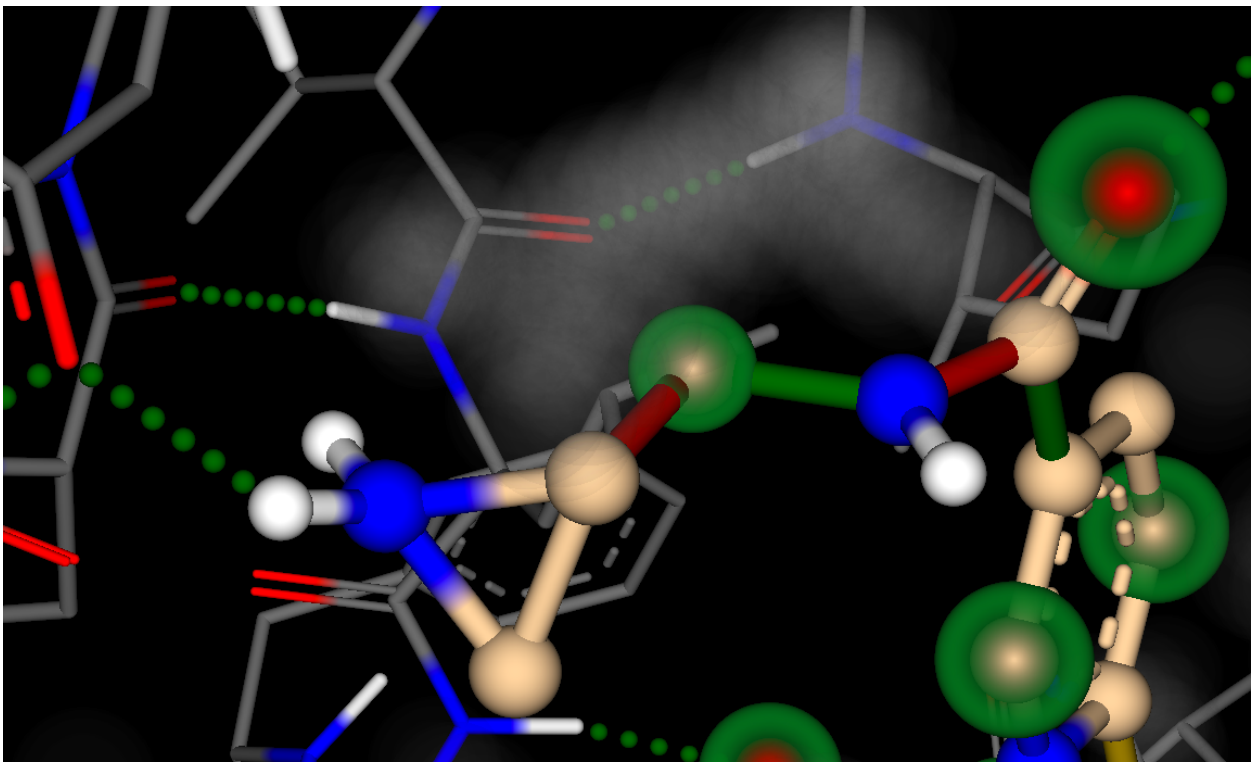
Take a look at the Hyde affinity coronas: the cyclopropyl ring (C24-C26) on the one end doesn't really contribute well, infact, atom C26 is penalyized, because it is blocking off the backbone carbonyl oxygen from entering into a hydrogen bond:



If we simply attempt to replace the carbon with a nitrogen, the bond is too short, and there is not much improvement:

- Enter the editor
- Select C26
- Type “n” on your keyboard, or select nitrogen from the list of atoms in the editor (refer to table above).
- Save to table

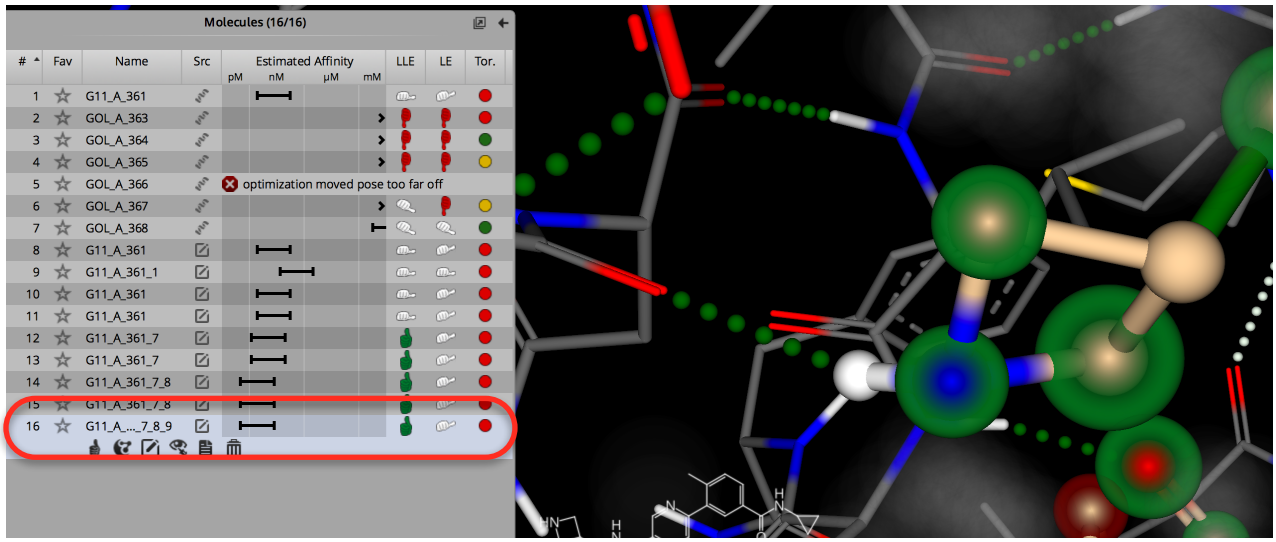
If we left the editor, we would see that although a hydrogen bond is made between the ring and the backbone oxygen, there is still almost no contribution from the ring:



Let us therefore remove the bond between the nitrogen and C24. Then do the following:

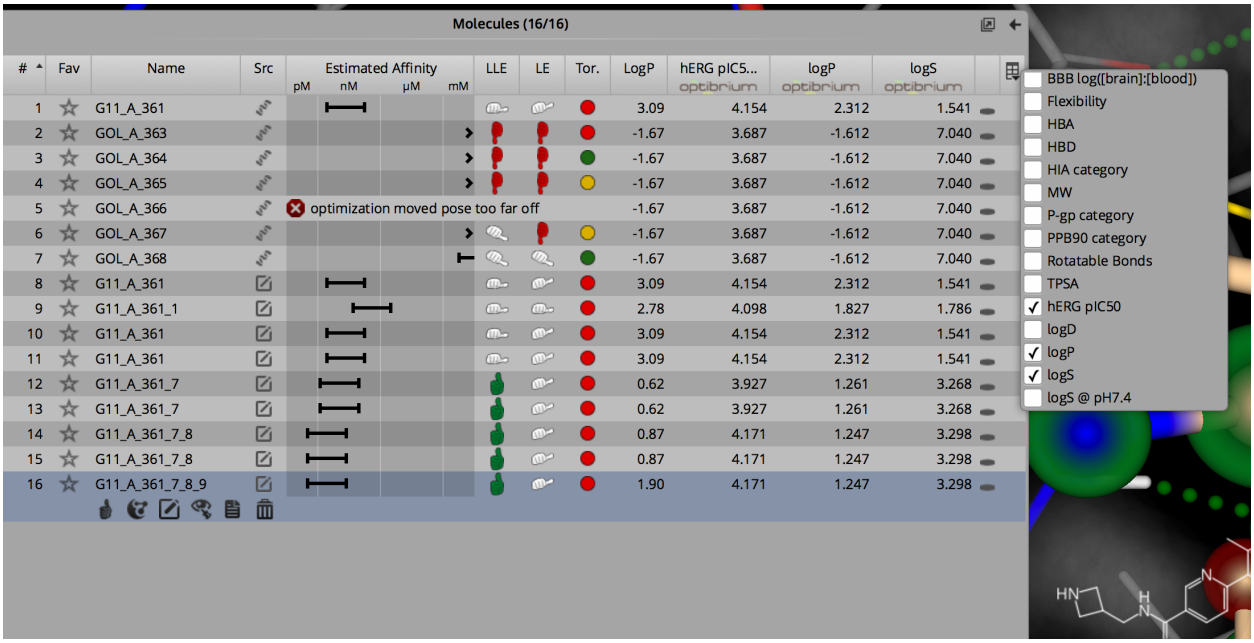
- Right click on the nitrogen and extend by a carbon
- Adjust the charge of the nitrogen to neutral (ni case it is not neutral yet)
- Make a bond between the new carbon C47 and C24
- Save to table and leave the editor

Now the picture is a different one:



Not only does every atom of the ring contribute favorably, the nitrogen atom make a hydrogen bond with the oxygen backbone atom. Moreover, we have also improved LLE.

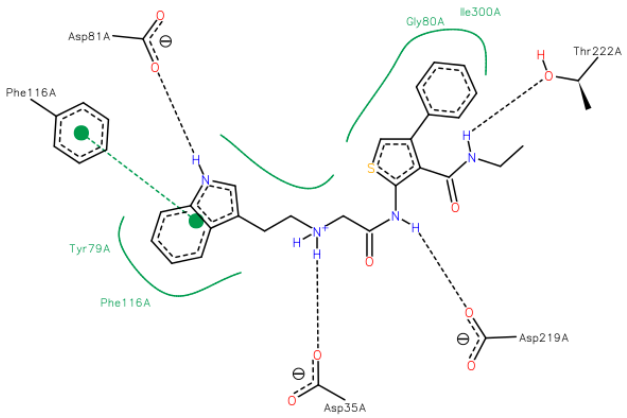
If we want to get information about other important properties of the molecule, we need to open the table properties button again:



A selection of Optibrium's QSAR models is available through the check boxes. The table can also be sorted according to these, just as it can be sorted by affinity, LLE, LE, or any other table entry.

Example 2 - A "real improvement" on Endothiapepsin

While the example above was somewhat artificial, we will look at something more real this time.
The PDB structure 4L6B shows Endothiapepsin in complex with thiophen-based inhibitor SAP128:[5]



PoseView image of 4L6B

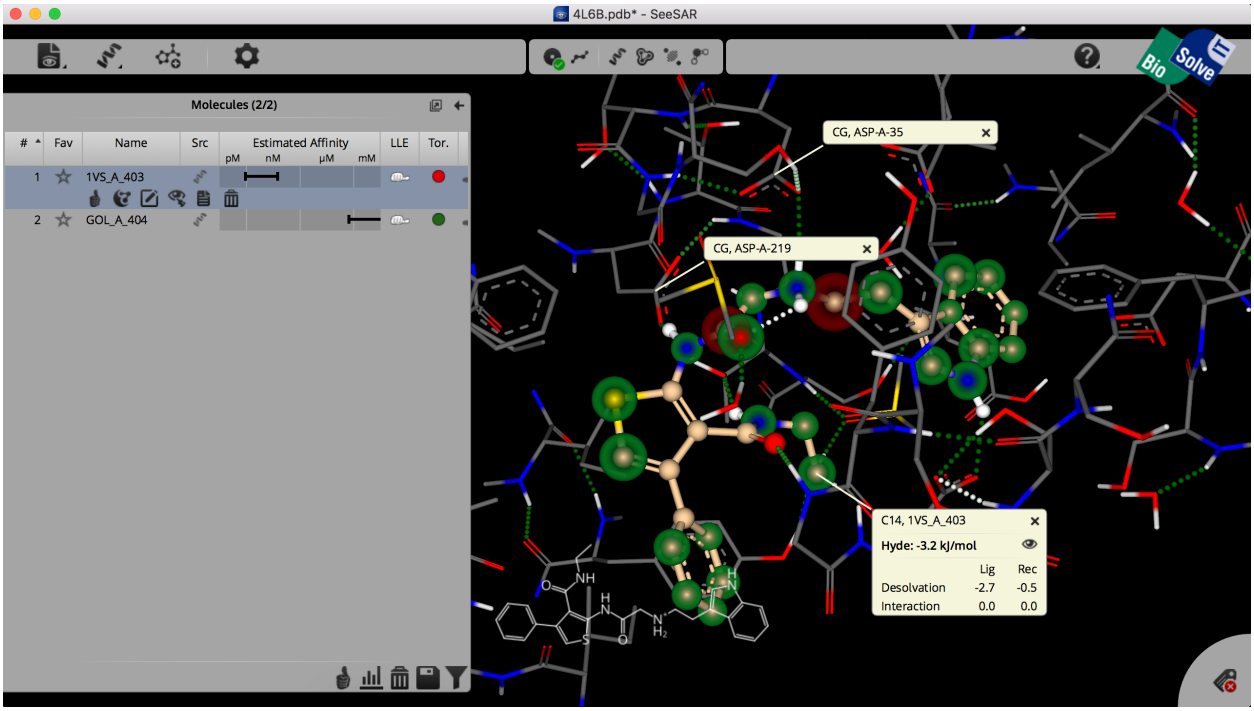
The indole ring sits in the S1-pocket, the amine and amide in the mid section bind Asp-35 and Asp-219, respectively.
A modified version, in which the ethyl amide binding to Hr-222 is replaced by a benzylamide, binds 200 fold stronger to Endothiapepsin.

Here is what you have to do:

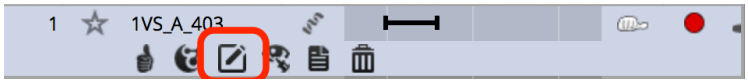
[5] The only reference is a PhD thesis in German:https://archiv.ub.uni-marburg.de/diss/z2012/0926/pdf/dhk.pdf

- Load 4L6B into SeeSAR
- Select (click) on 1VS_A_403
- Turn off the protein cartoon backbone representation
- Identify the two Asp-35 and Asp-219 by labeling them
- Identify the ethyl amide on the ligand

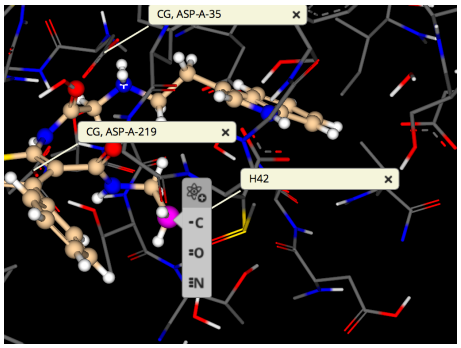
Your screen should look somewhat like this now:

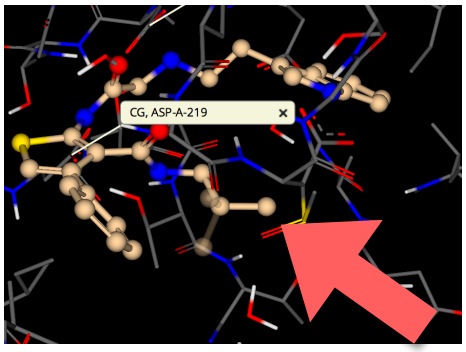


- Enter the editor
- Select C14



- Right click C14 and select “Extend by carbon”. You will have three choices. Select the one

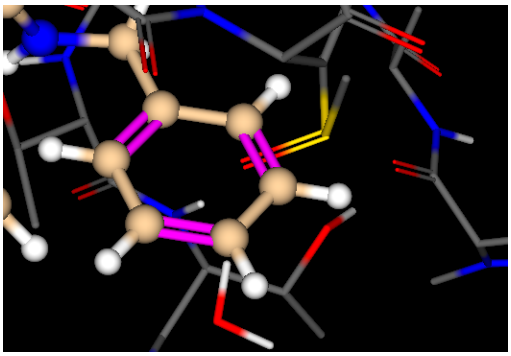




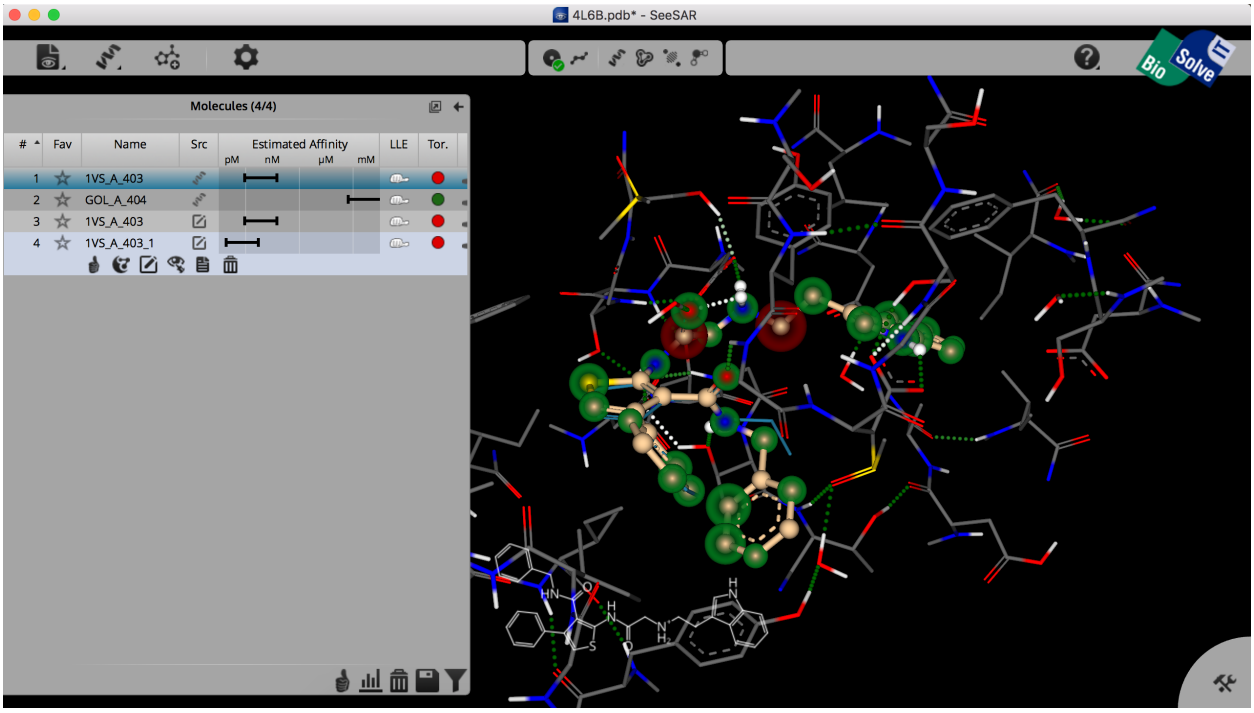
that would start to make a 6-ring.

- *Alternatively: select the hydrogen H42 an make it into a carbon by typing “c” on your keyboard.*

- Repeat this process (either one from the two above) until you have a 6-ring (make a bond between the last C-atom and the first).
- Select three alternating bonds and type “2” on your keyboard to make it a benzene ring:

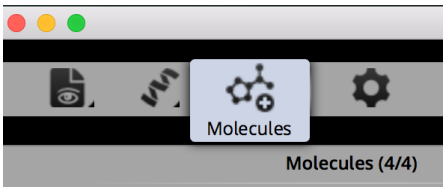


Leave the editor. This edited compound should have an improved affinity, which results in part by the benzylamide sticking in the S2 pocket:

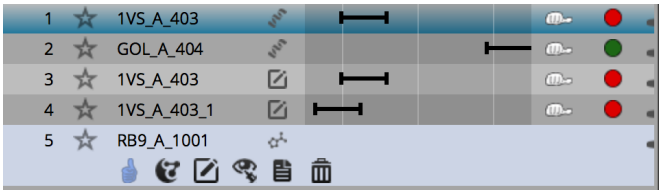


In blue is the reference compound, 4L6B.

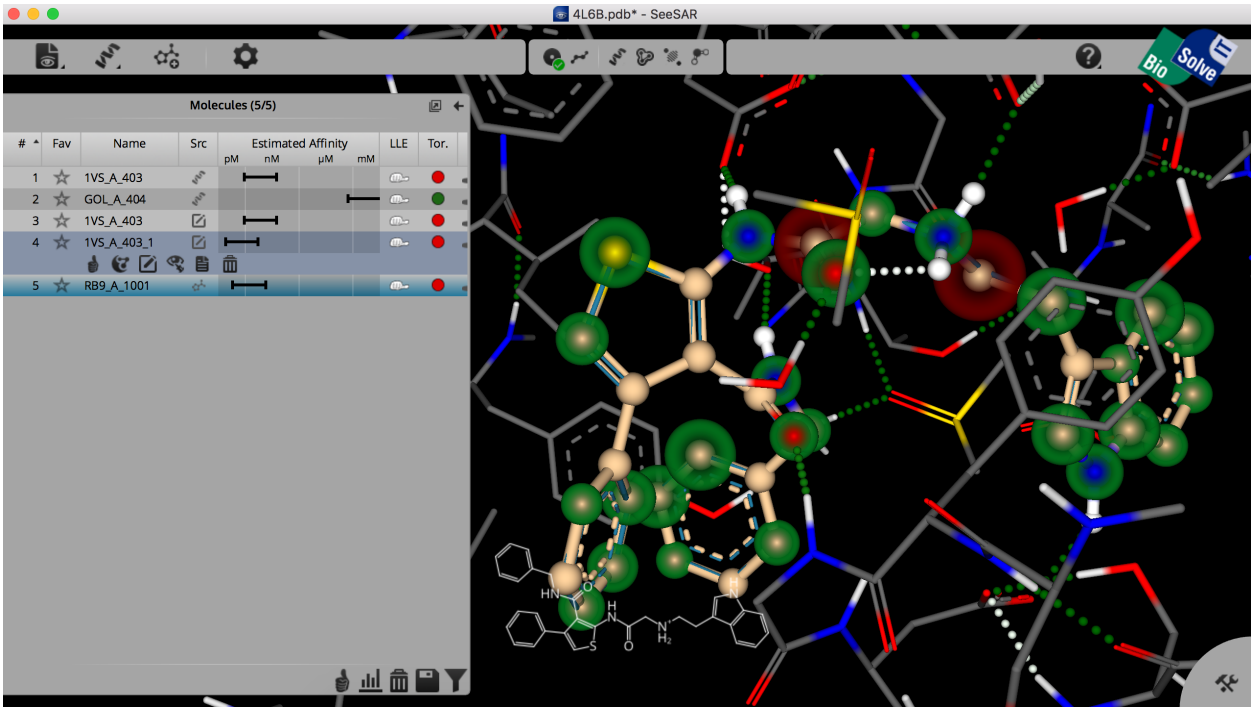
Now load the same compound from the crystal structure 3PSY into SeeSAR (the file should reside in your material you have downloaded: 3PSY_ligand_pdb.sdf):



Calculate its estimated affinity by clicking on the thumb underneath RB9_A_1001:



Make this compound the reference and click 1VS_A_403_1 (the last number may differ, based on how many variants you have saved):

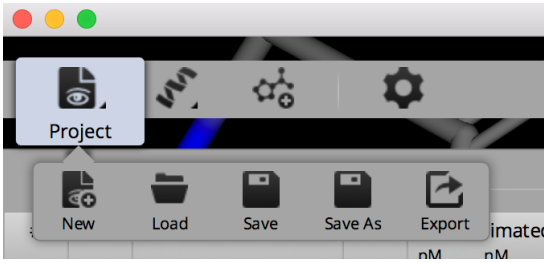


The molecules should assume the same position, with possibly just a slight difference where the benzylamide resides!

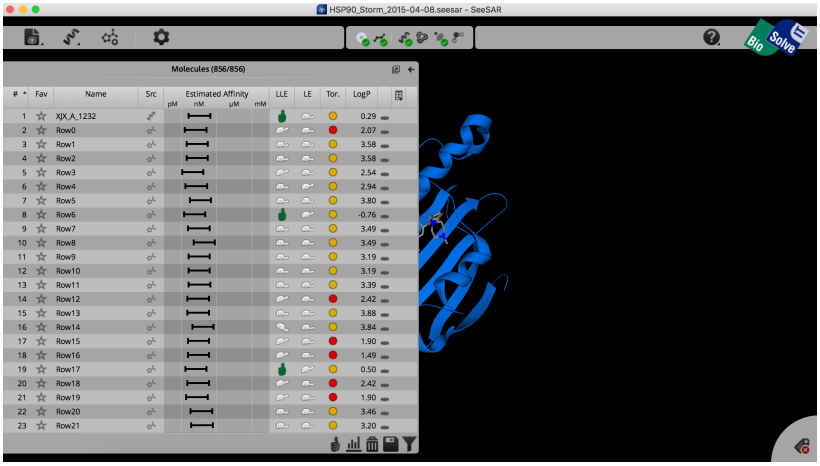
Example 3 - Improving a ligand for HSP90

This example is based on PDB 2XJX, an HSP90 with a small ligand inhibitor bound. Problem with this ligand is that a large portion protrudes into the solvent without interacting with any protein residues, and therefore doesn't contribute to the affinity. We would like to see if and how we can change that.

Open the SeeSAR project file that you have downloaded previously:



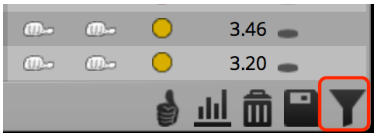
The file contains the PDB structure 2XJX, as well as 856 molecules originating from a *de novo* run with our STORM workflow in KNIME.[6]



In this example, we will showcase SeeSAR's powerful filter functions, as well as the property calculations we have built in with the help of our partners at Optibrium.

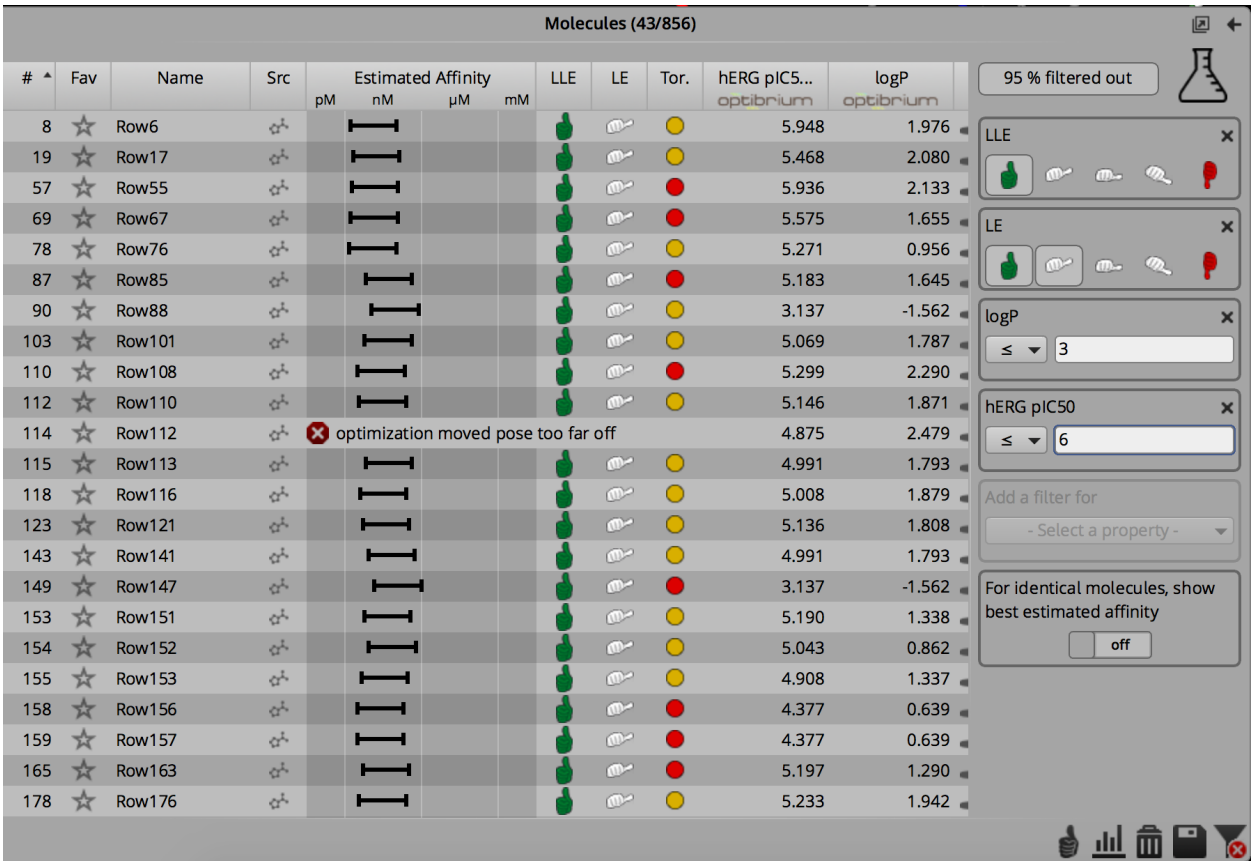
[6] For more information on this workflow, how to obtain and install it: http://www.biosolveit.de/KNIME/

Working with a few ligands is not much of a challenge, but when it comes to screening data sets or a virtual library, we need ways to make the most of such vast amounts of ligands, and extract the information we need.
Therefore, SeeSAR has built in powerful filter functions. They can be found underneath the table to the right:



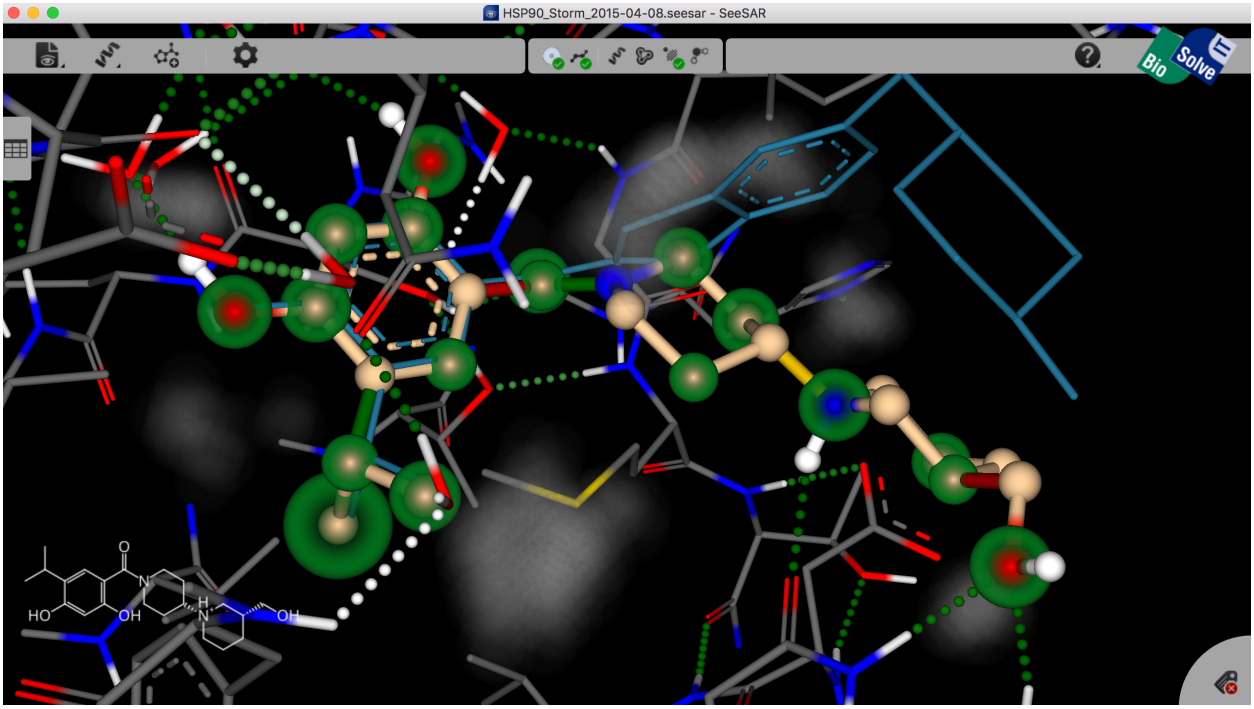
When clicked, a panel opens to the right of the table and for all properties in the table, filters can now be applied:

E.g. I would like to display only ligands with e.g. beneficial hERG values and logP, as well as LLE and LE:



#	Fav	Name	Src	Estimated Affinity (pM nM µM mM)	LLE	LE	Tor.	hERG pIC5... optibrium	logP optibrium
8	☆	Row6						5.948	1.976
19	☆	Row17						5.468	2.080
57	☆	Row55						5.936	2.133
69	☆	Row67						5.575	1.655
78	☆	Row76						5.271	0.956
87	☆	Row85						5.183	1.645
90	☆	Row88						3.137	-1.562
103	☆	Row101						5.069	1.787
110	☆	Row108						5.299	2.290
112	☆	Row110						5.146	1.871
114	☆	Row112		optimization moved pose too far off				4.875	2.479
115	☆	Row113						4.991	1.793
118	☆	Row116						5.008	1.879
123	☆	Row121						5.136	1.808
143	☆	Row141						4.991	1.793
149	☆	Row147						3.137	-1.562
153	☆	Row151						5.190	1.338
154	☆	Row152						5.043	0.862
155	☆	Row153						4.908	1.337
158	☆	Row156						4.377	0.639
159	☆	Row157						4.377	0.639
165	☆	Row163						5.197	1.290
178	☆	Row176						5.233	1.942

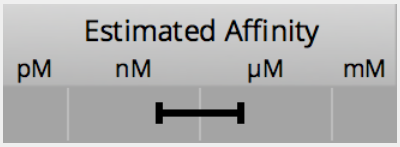
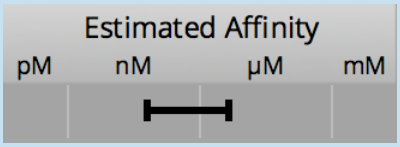
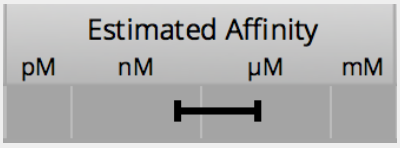
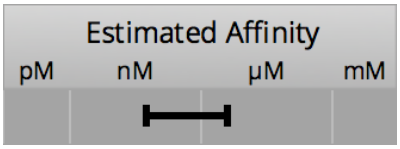
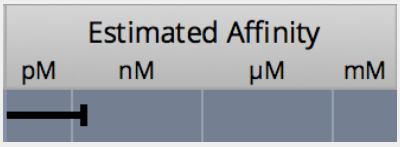
You can see that 95% of the molecules are already filtered out, and we are left with 43, a way more manageable number.
If we now sort by affinity (click on “estimated affinity” header in the table, compound “Row 731” slides to the top. We can see why this compound is so much better than the crystal structure, because it actually interacted with the protein side chain on the rim of the pocket:
The crystal structure is in blue because we selected it as reference:



Other compounds in the remaining set also have good properties.

Example 4 - Compound prioritization of a congeneric series for a thrombin example

The last example will deal with a thrombin structure.

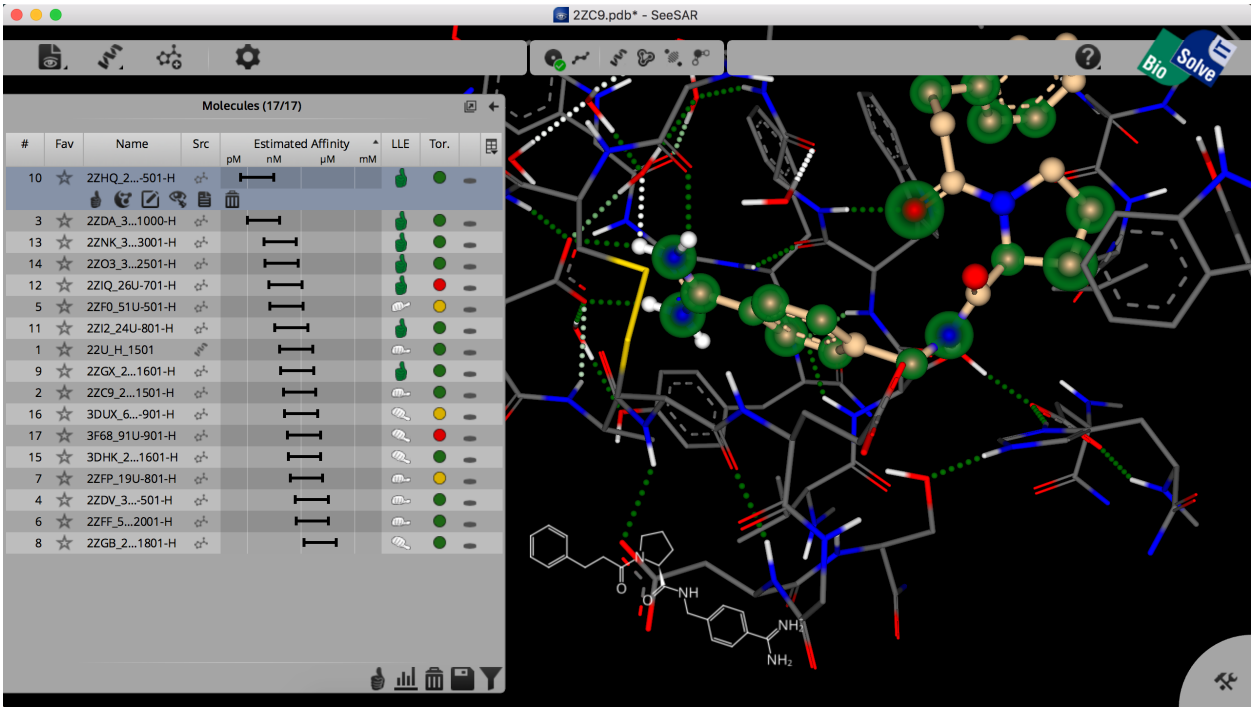
PDB	ΔG experimental	Estimated affinity Hyde
2ZDV	-30.8	
2ZC9	-38.48	
2ZFF	-28.24	
2ZF0	-33.78	
2ZDA	-47.91	

Load PDB 2ZC9 (our reference in blue in the table above) into SeeSAR. Load the set of ligands “thrombin_aligned_ref_ligs_with_properties.sdf”

The molecules in the table above are represented with their experimental and calculated binding affinity, each in their corresponding crystal structure.
We can see that the calculated affinity pretty closely represents the experimental one.

The ligands in the example file you have just loaded are these and other thrombin binding ligands *aligned* in the pocket of 2ZC9.

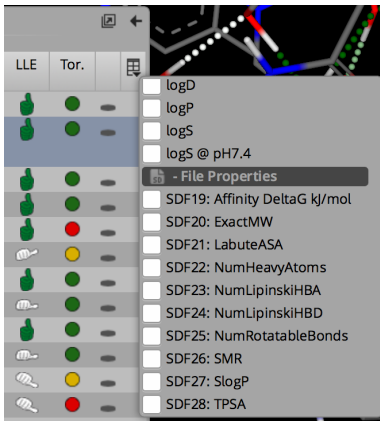
Sorting by affinity reveals the best of the 16 ligands, 2ZDA, at the top of the list (2ZHQ is not much different from 2ZDA; it has one more stereo center):



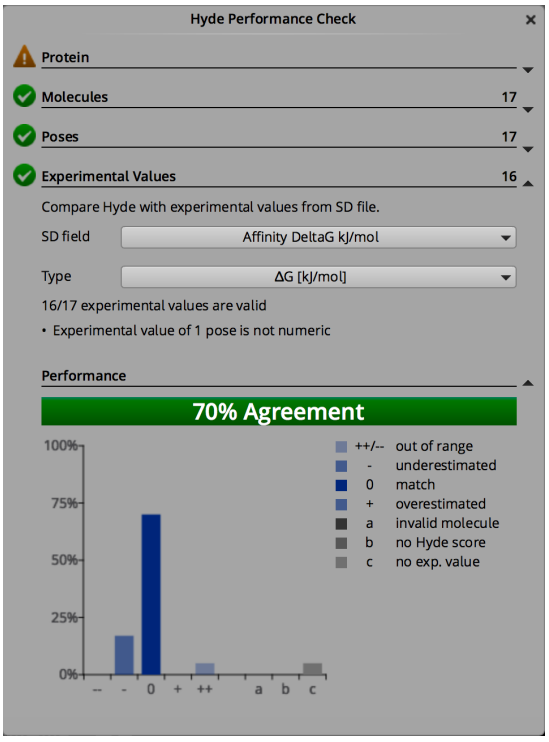
If you would like to perform any of the function below (Hyde calculation [thumb]; delete [trash can]; save [diskette]) on just a subset, select the molecules you would like to have as subset by clicking the stars in the column “Fav”.

As a last exercise we would like to show you the performance check in SeeSAR.
While we do not endorse correlation analysis per se, as their outcome is questionable (data from different sources, IC50 vs Ki, neglected errors, unsatisfactorily filled electron density in crystal structures), sometimes it is useful to gauge the deviation of calculated from experimental affinity. Therefore, we have built in the performance check:

If your molecules have an experimental affinity section in the SD file, this information will be read (and also available of the table, cp section underneath the Optibrium properties, labelled SDF properties):



If you now click the performance check, you can display how well the experimental affinity coincides with the calculated one:



Please note: a 100% agreement is not expected. A *quantitative* calculation of affinity is not possible with today's methods.

What we are after is a *qualitative* assessment of whether a change on a molecule is moving the bar to better affinity or worse. By how much is not necessarily relevant.