4. Scoring of nine ligands in p450Cam

Martin Almlöf, Martin Andér, Sinisa Bjelic, Jens Carlsson,
Hugo Gutiérrez de Terán, Martin Nervall, Stefan Trobro and Fredrik Österberg

18th August 2005

Contents

1 Scoring Functions

1.1 Theoretical background

Simple functions for estimating free energies of binding have become popular tools in the area of computer aided drug design. The functions are characterized by high speed compared to the more accurate methods of FEP and LIE. The high speed enables utilization in docking applications where large conformational searches require repeated estimations of the free energy of binding to find the most probable conformation. If the docking procedure is fast enough large virtual libraries can be screened for binders towards a target receptor.

Scoring functions can in general be divided into several categories depending on their origin.

- Empirical scoring functions, derived from parameterizations on experimental binding data.
- Knowledge based scoring functions, derived from distance to distance information in available 3D protein-ligand complexes.
- Other scoring functions such as force field based and quantum chemically based.

In this practical we will focus on the empirically based functions.

1.1.1 Empirical functions

Empirical functions rely on a (usually) linear combination of several contributing terms, e.g.

$$\Delta\Delta G_{\mathrm{bind}} = \sum_i \alpha_i \Delta G_i. \tag{1}$$

The ΔG_i's denote the binding contribution from different types of interactions and are determined for a large number of complexes for which ΔG_{bind} is known experimentally. From this, the α_i's are determined by regression so as to minimize the prediction error.

The assumption is that the total binding energy can be expressed as a linear combination of terms, each corresponding to one type of interaction, and that the contribution from each term can be determined computationally. Some common types of interactions considered by empirical scoring functions are:

- van der Waals interactions
- hydrogen bonding

- ionic and metallic interactions
- lipophilic interactions
- entropic contributions

1.2 ChemScore

The ChemScore function was published by Eldridge *et al* in 1997 and was calibrated on a set of 82 protein-ligand complexes. The goal was to develop an empirical function using simple and physically interpretable terms:

$$\begin{aligned}\Delta G_{binding} = \Delta G_{H-bond} + \Delta G_{metal} + \Delta G_{lipophilic} + \\ + \Delta G_{deformation} + \Delta G_0.\end{aligned} \tag{2}$$

where ΔG_{H-bond} considers the contributions from H-bonds, ΔG_{metal} is a metal interaction term, $\Delta G_{lipophilic}$ handles the vdW contribution, $\Delta G_{deformation}$ is a penalty term for entropic loss and ΔG_0 is a constant offset.

1.3 X-Score

X-Score is an empirical scoring function published by Wang *et al* in 2002. It was calibrated on a set of 200 protein-ligand complexes in the Protein Data Bank (PDB).

The binding contributions considered by X-Score are van der Waals interactions, lipophilic interactions, hydrogen bonding and deformation effects. Thus, X-Score takes the following form:

$$\begin{aligned}\Delta G_{binding} = \Delta G_{vdw} + \Delta G_{H-bond} + \Delta G_{lipophilic} + \\ + \Delta G_{deformation} + \Delta G_0,\end{aligned} \tag{3}$$

where ΔG_0 is a regression constant including an average of the rotational and translational entropic contributions.

X-Score is a consensus scoring function; three different algorithms for computing the lipophilic effect are used — each generating a separate scoring function. The final score is computed as the average of these three outputs.

X-score and ChemScore are very similar when you consider the terms involved, but there are substantial differences in the details of calculating the corresponding terms.

2 Practicals

The aim of this session is to get acquainted with two of the scoring functions implemented in the Q5 package, ChemScore and X-Score. We will use a set of nine inhibitors of the enzyme p450Cam. To make the practical more interesting we will investigate the difference between scoring a single snapshot and scoring a short molecular dynamics trajectory of 50 ps.

Log on to your account if you are not already logged on. Move to the "scoring" directory:
`cd scoring`
The first ligand we will score interactively, the rest is already scripted and fast to run.

2.1 ChemScore

Move to the first ligand directory called "adm" and start `Qcalc5` by typing
`cd adm`
`Qcalc5`
The first thing the program asks for is the topology file, which should be a familiar concept by now. In this case the file is called `adm.top`
Now you have come to the stage where you should choose what calculation to make. Write `chemscore` and press enter.
The program will prompt you for a mask. When you want to score a trajectory the mask specified here should be the same as the trajectory was created with. In our case all trajectories were created with the atom mask `solute`. End the mask definition with `end`.
The next question you are faced with is whether you want to score the initial topology or if you want to score a trajectory. Write `yes` to score the topology first.
The fep-file is called `lig.fep` for all the ligands. Note that the files are different for the different ligands even though they have the same name.
Finally you should define the parameter file containing parameters for `ChemScore` with the force field Oplsaa. The file is in:
`/ibg/courses/1MB280/workshop2005/Q5/chemscore_oplsaa.prm`

Now that you have defined the calculation you can start it by typing `go`. The score of the initial conformation, i.e. the crystal structure, should appear within a few seconds.

As you can see the program is still running. It is waiting for you to define a trajectory file to continue the calculation on. Type `dc1.dcd` and watch the com-

puter work. When you have scored the whole trajectory of 50 ps with snapshots every 2 ps you have 25 scoring values. Calculate the mean value by typing `mean` and finally exit the program by giving the `end` command.

2.2 X-Score

Next we will take a look at `XScore`. Start `Qcalc5` again and define the same topology as before. Now choose the option `xscore`. The procedure is just like in the `ChemScore` case until you are prompted for `Cofactor`. p450Cam has a co-factor, a heme group, which we must specify. Define it by typing `restype=HEM`. End the definition with a blank line.
The program will now ask for a code corresponding to the force field the system is represented in. We are using Oplsaa in this molecular system and the code to give `XScore` is `qoplsaa`.
Finally you should define the path to a file with input parameters for `XScore`. Enter the file name
`/ibg/courses/1MB280/workshop2005/Q5/xscore_default.input`

As before, start the calculation with `go` and wait until the score for the initial conformation appears. Then type `dc1.dcd` to score the trajectory, `mean` to get the average and `end` to stop the program.

2.3 Scoring nine ligands

In order to get some statistics for the comparison of scoring a snapshot versus a molecular dynamics trajectory we will now score a total of nine ligands in p450Cam. Start the scoring by going back to the "scoring" directory and run the shell script `score_all.sh`:

```
cd ..
csh score_all.sh
```

The script will perform the same scoring you just did with the `adm` ligand for all nine ligands. Thus it will run both `ChemScore` and `XScore` and write the result to `chemScore.log` and `XScore.log` for the respective scoring function.
When the script has finished you can view the result in `gnuplot` by running the script `plot_all.sh` which will extract data from the logfiles. The script will take the average of the `ChemScore` and the `XScore` value for the plot.
How do the points corresponding to the scored trajectories fit the experimental values? Compare them to the values from the scored crystal structures. Which method gives the best ranking? You can also compare the numbers if you open the file `plot.txt` in `emacs`. The first column is the experimental values, the second is the scored crystal structures and the third is the average from the trajectories.