CATS Short Manual

G. Schneider, January 2003

Note for CATS*light* users
CATS*light* has limited functionality compared to the full version of the software. Its main limitation is the restriction to 50 molecules in the Library File to be searched and the unavailability of some options.

What is CATS?

The CATS (Chemically Advanced Template Search) software can be used to perform similarity searching in a collection of small molecules [1].

Molecules are described by a two-dimensional pharmacophore descriptor ("topological pharmacophores"), and CATS-similarity is defined by the degree as to which the topological pharmacophore descriptors of two molecules match (pair-wise Euklidian distance between the pharmacophore fingerprints).

Typically, one wishes to rank all compounds of a library by their pair-wise similarity to a reference or "seed compound" (the query structure). The result of a CATS run is a ranked list of all library members (Figure 1).

This kind of similarity searching has been successfully applied in a number of drug discovery projects with an emphasis on "lead- or scaffold-hopping" tasks [2,3]. It must be stressed, however, that CATS will *always* produce a ranked list of molecules. This does *not* mean that the compounds which are most similar to your query structure will necessarily exhibit similar properties (e.g., binding affinity to a receptor) as the query.
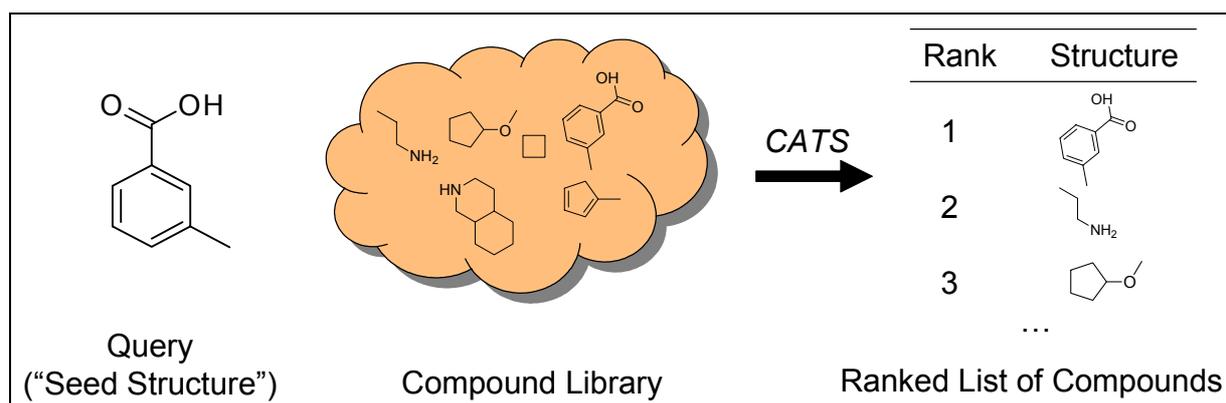

Figure 1. CATS can be used to sort the members of a compound library by similarity to a query molecule.

How can I use the CATS www interface?

The CATS web form contains three main parts: the **Query** field, the **Library** field, and the **Options** field. All three fields must be filled appropriately before pushing the "**Submit Data**" button.

Your query ("seed") structure is specified in the **Query** field. Two formats are possible, SMILES notation or SDF. You may also upload a file containing your query molecule (Figure 2). In the *light* version of the software only a single query structure may be specified.

Figure 2. The Query field of the CATS www interface.
Here the SMILES input was selected.

Your compound library to be searched is specified in the **Library** field. Two formats are possible in the full software version: use a pre-existing library on the MODLAB® server or give the location of a local library in SDF format (Figure 3). In the *light* version only a local SDF containing no more than 50 compounds may be specified. Note that the first line of each SDF entry requires a compound identifier, e.g.

```
Structure1
csChFnd70/11040211072D

 31 35  0  0  0  0  0  0  0999 V2000
    1.2124    6.3676    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.2124    4.9676    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
. . .
$$$$
Structure2
csChFnd70/11040211072D

 23 25  0  0  0  0  0  0  0999 V2000
    1.2124    7.7325    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.2124    6.3325    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
etc.
```

Figure 3. The Library field of the CATS www interface.
Here the SDF-upload input was selected (only option in CATS*light*).

Options specifying details of your CATS run can be given in the Options field. It is possible to define the contribution of the five potential pharmacophore points (lipophilic, hydrogen donor, hydrogen acceptor, positively charged, negatively charged) in terms of weighting factors (Figure 3). Setting all weights to a value of 1.0 means that no single interaction type will be preferred. You have the possibility to up- or down-weight the influence of pharmacophore points. A recommendation for CATS beginners: Use the balanced weighting scheme (all values 1.0; default setting) first, then repeat your CATS run with a weight of 0.5 or 0.0 for lipophilic interaction and compare the results.

Figure 3. The Options field of the CATS www interface.
Here balanced weights for potential pharmacophore points were selected.

How do I interpret CATS results?

A ranked list of molecules is produced by CATS. Molecules at the top of the list (low rank numbers) are more similar to the query than molecules at the bottom of the list (high rank numbers) in terms of the CATS topological pharmacophore similarity measure. For some cases it was demonstrated that molecules on the top ranks exhibit isofunctional biological activity compared to the query structure but have different backbone architecture [1-3]. One possible usage of the ranked list is the design of a focused library, i.e. the top-ranking compounds can be pooled and tested for a desired biological activity or property [4-6].

It is important to stress the fact that several similarity searching methods should be applied and compared to CATS before entering a biochemical testing campaign. Different projects demand for different similarity principles, and a method which was shown to work in one project might not be suited for a different one.

References

[1] Schneider G, Neidhart W, Giller T and Schmid G. (1999) „Scaffold-Hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew. Chemie* 111, 3068-3070; *Angew. Chemie Int. Ed.* **38**, 2894-2896.

[2] Schneider G, Chomienne-Clement O, Hilfiger L, Kirsch S, Böhm HJ, Schneider P and Neidhart W. (2000) Virtual screening for bioactive molecules by evolutionary de novo design. *Angew. Chemie Int. Ed.* **39**, 4130-4133.

[3] Schneider G, Neidhart W and Adam G (2001) Integrating virtual screening to the quest for novel membrane protein ligands. *Curr. Med. Chem. CNSA* **1**, 99-112.

[4] Schneider G (2002) Trends in virtual combinatorial library design. *Curr. Med. Chem.* **9**, 2095-2101.

[5] Schneider G and Böhm HJ (2002) Virtual screening and fast automated docking methods. *Drug Discovery Today* **7**, 64-70.

[6] Schneider G and So SS (2001) *Adaptive Systems in Drug Design*. Landes Bioscience, Georgetown.