

## COSMOfrag/COSMOtherm for drug research

### ...validated applications

- drug solubility in water
- solubility in any solvent / solvent screening
- general partition coefficients between any solvents
- special partition properties: logBB, intestinal absorption, Human Serum Albumin Binding, ...

### ...further potential applications

- unified interaction model of electrostatics, H-bonds, and lipophilicity
- valuable for molecular field analysis (MFA)
- extension to membrane modelling
- localisation of interaction sites
- solvent influence on morphology
- surely many other applications

## COSMOlogic Products

- COSMOtherm: Software for Life Science and general Fluid Phase Thermodynamics
- COSMObase: Database of 3200 DFT/ COSMOfiles for common solvents and compounds
- TURBOMOLE: Fast Quantum Chemical Program (Prof. R. Ahlrichs, Univ. of Karlsruhe, Redistribution and Support by COSMOlogic)
- COSMOfrag: Software for rapid screening and HTS applications with COSMOtherm, including various options for structure and conformational analysis and tautomer generation
- COSMOsim: Similarity Screening based on  $\sigma$ -Profiles

## COSMOlogic Services

- Consulting in Computational Chemistry and Fluid Phase Thermodynamics
- Contract Calculations in the Context of COSMOtherm
- Contract Research, Development, and Programming in Computational Chemistry related Areas

# COSMOlogic

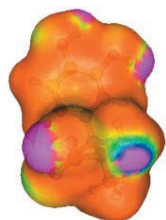
GmbH & Co. KG

Your Competent Partner for  
Computational Chemistry and Fluid Thermodynamics

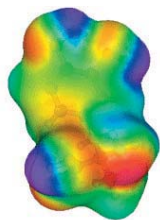
Burscheider Str. 515  
D-51381 Leverkusen  
Tel.: 49 (0) 2171 73168 0  
Fax: 49 (0) 2171 73168 9  
E-mail: [info@cosmologic.de](mailto:info@cosmologic.de)  
Web: <http://www.cosmologic.de>  
CEO: Dr. Andreas Klamt

# COSMOfrag

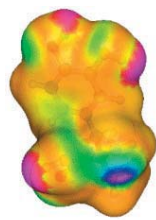
High Throughput ADME Property  
Prediction with COSMO-RS



Methyldopamin colored  
by intestinal Absorption



Methyldopamin colored  
by  $\sigma$



Methyldopamin colored  
by Blood-Brain  
Partition Coefficient

**COSMOfrag** is a tool which makes ADME property prediction with *COSMOtherm*, and *COSMOSim* similarity screenings usable in HTS projects. For that purpose, approximate  $\sigma$ -profiles of larger molecules are constructed using the  $\sigma$ -profiles of most similar fragments in a large database of precalculated molecules. Due to the fact that  $\sigma$ -profiles are distribution functions characterizing a certain molecule, they can be built up by a sum of

partial  $\sigma$ -profiles of suitable fragments. Since thermodynamic calculations with *COSMOtherm* only require fractions of a second the overall speed of the COSMO-RS method is mainly determined by the time demand of the underlying quantum chemical calculations for the molecules. On our standard highest quality level (BP-TZVP) such calculations take about 4 hours for a molecule with up to 40 heavy atoms. In order to speed up such *COSMOtherm* drug applications we have introduced a "drug-level" for the DFT/COSMO calculations that replaces the DFT geometry optimization by a semiempirical AM1 geometry optimization, followed by a single point DFT/COSMO calculation with a smaller basis set (BP-SVP). This reduces the quantum chemical time requirements by almost a factor of 30, i.e. roughly to 8 minutes per drug, at a moderate loss of accuracy which is normally acceptable in drug design applications. Nonetheless still on this level a pre-screening of up to hundreds of thousands or millions of potential drug candidates is unfeasible even on large parallel computer clusters.

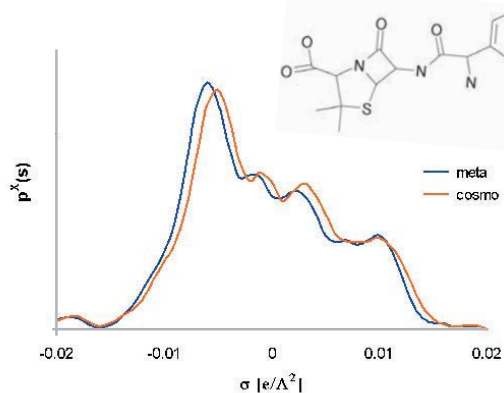


Fig. 1.  $\sigma$ -profiles of COSMOfrag meta file vs. cosmo file

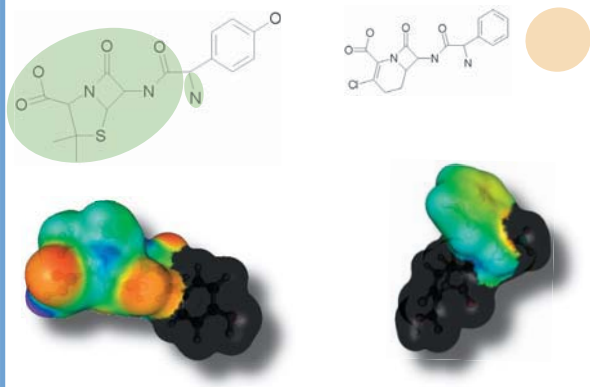


Fig. 2.  $\sigma$ -profiles of COSMOfrag meta file vs. cosmo file

### Methodology

The *COSMOfrag* program is a tool that efficiently performs the fragmentation of a new molecule by molecules of the *COSMOfrag* database (CFDB), which is composed of presently 55.000 highly diverse molecules obtained by screening procedures of different databases and libraries of drug-like compounds such as Physprop\* or NCI\*\*.

The result of the fragmentation is written to a *COSMOtherm* meta-file which then can be used as substitute for the full COSMO file in the *COSMOtherm* input. In case your compounds should not be sufficiently well represented by the CFDB you can easily run DFT/COSMO calculations for a

small number of representative compounds, add them to the database and consecutively improve the fragmentations.

The basic concept of *COSMOfrag* consists of a careful molecular perception of the input molecule which can be given in different molecular formats, including SMILES notation. Bonds and rings are analysed with respect to conjugation, E-Z-substitution, and aromaticity. Finally for each atom hash coefficients are calculated taking into account all the local information about the atom itself and about bonds and neighbour atoms. Step by step higher order hash codes are calculated which include the information about an increasing number of neighbour spheres of the atom. Thus, each two atoms that have an equivalent molecular environment up to the  $n^{\text{th}}$  neighbouring sphere have identical hash codes up to the  $n^{\text{th}}$  order. After that all similarity coefficients are converted into 5-character ASCII strings and combined to a 50 character atom code. Thus the search for most similar atoms can be done by a simple search for most similar atom codes. Based on our present database of about 55.000 drug-like compounds most molecules get fragmented into 2 - 4 fragments.

### Application

The *COSMOfrag* program has been applied to a variety of life science property predictions (partition coefficients, water-solubility, intestinal-absorption and soil-sorption) and similarity screenings with *COSMOSim*. The loss of accuracy was always below 0.05 log-units compared to direct DFT/COSMO calculations, at a throughput of 1 - 2 molecules per second on a single PC-processor.

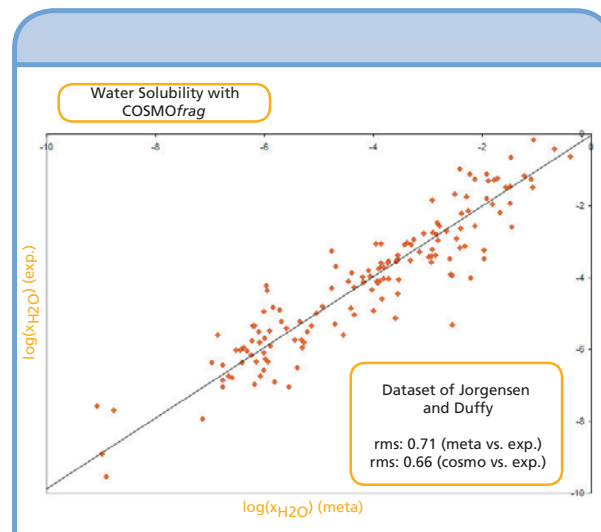


Fig. 3. *COSMOfrag* validation on BOSS-dataset

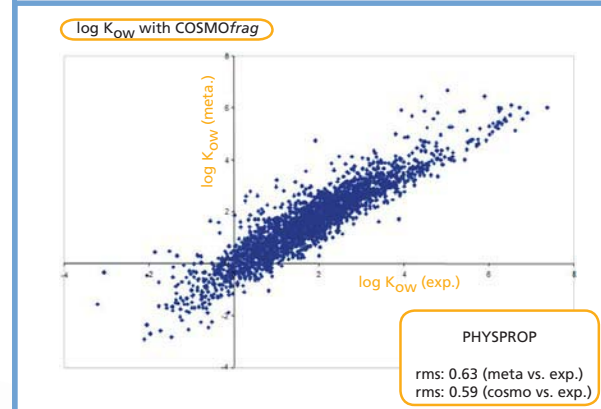


Fig. 4. 2570 exp. data from PHYSPROP

### References

- M. Hornig and A. Klamt  
J. Chem. Inf. Model, 45, 1169 (2005)
- A. Klamt, F. Eckert and M. Hornig  
J. Comp.-Aid. Mol. Design, 15, 355 (2001)

\* Physical Properties Database by Syracuse Research Corporation  
\*\* National Cancer Institute Database