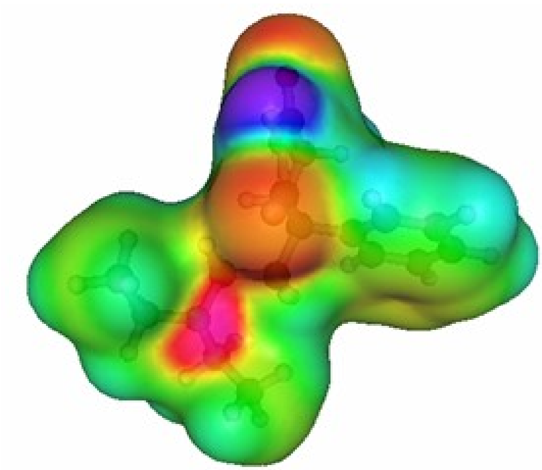
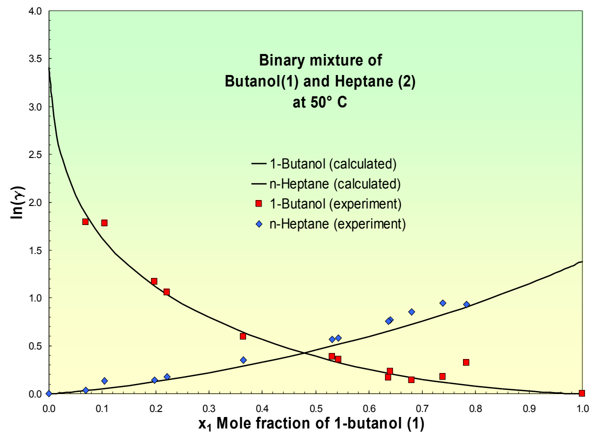
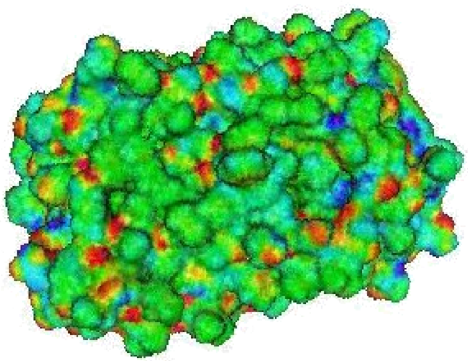
COSMO-RS:
From quantum chemistry to Cheminformatics





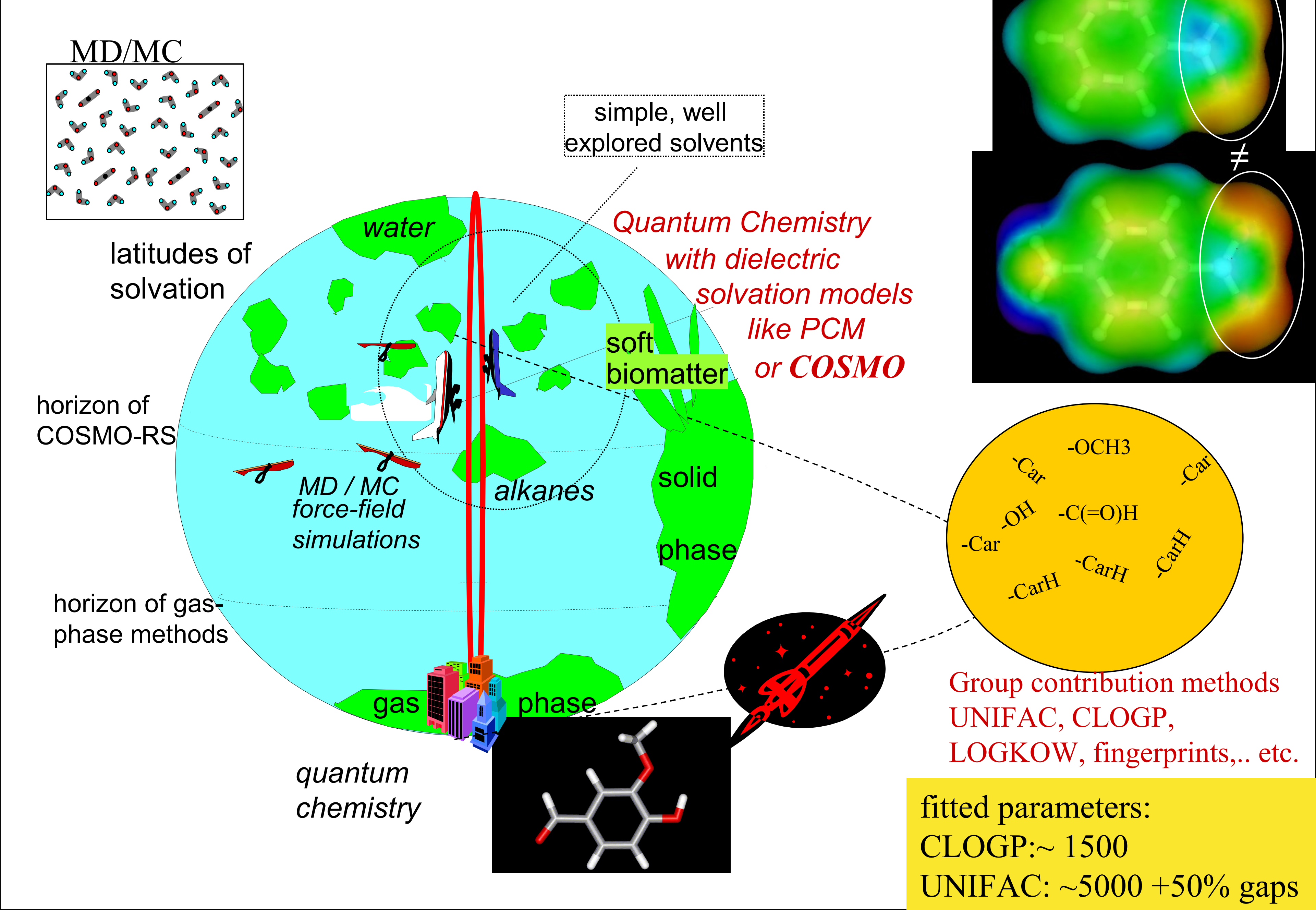


Andreas Klamt

COSMO*logic* GmbH&Co.KG, Leverkusen, Germany

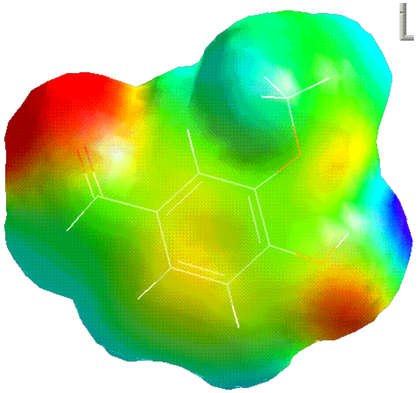
and Inst. of Physical Chemistry, University of Regensburg, Germany

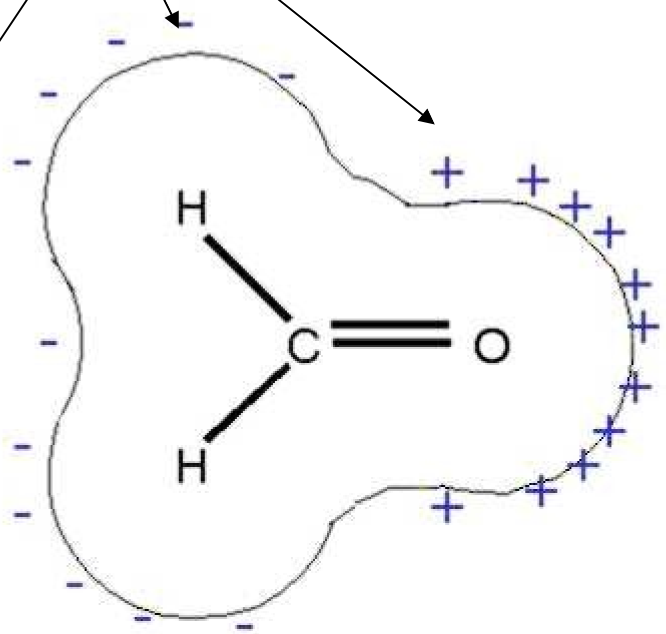
Thermophysical data prediction methods



Dielectric Continuum Solvation Models (CSM)

solute molecule embedded in a dielectric continuum,
self-consistent inclusion of solvent polarisation
(screening charges) into MO-calculation (SCRF)





- Born 1920, Kirkwood 1934, Onsager1936
- Rivail, Rinaldi et al.
- Katritzky, Zerner et al.
- Cramer, Truhlar et al. (AMSOL)
- Tomasi et al. (PCM) - Orozco et al.
- Klamt, Schüürmann (COSMO) e.g. $DMol^3$/COSMO and others

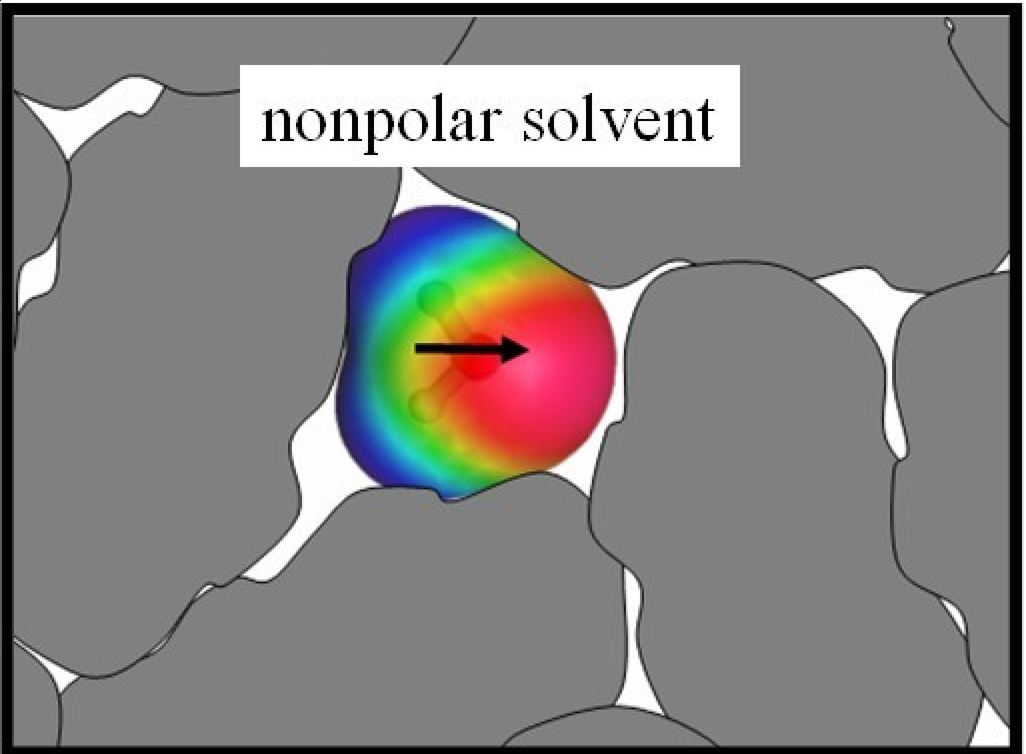
COSMO =
COnductor-like Screening Model,
just a (clever) variant of dielectric CSMs

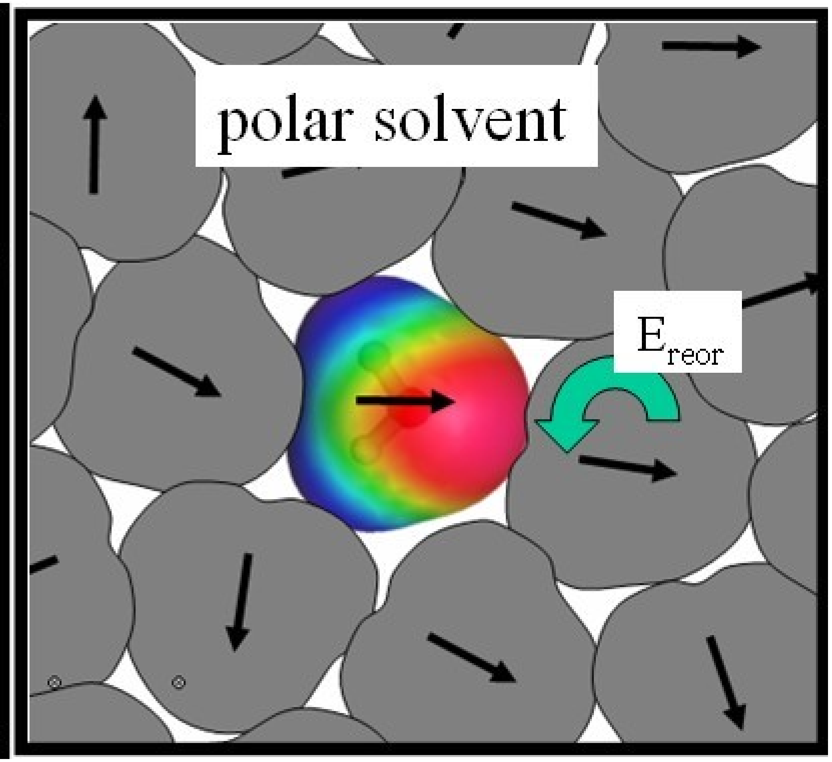
Density Functional Theory (DFT)
is appropriate level of QC!
COSMO almost as fast as gasphase!
programs: $DMol^3$, Turbomole,
Gaussian98_release2001
up to 25 atom:< 24 h on LINUX PC

- **empirical finding: cavity radii should be about 1.2 vdW-radii**
- **promising results for solvents water, alkanes, and a few other solvents**

But CSMs are basically wrong and give a poor, macroscopic description of the solvent !

Why are Continuum Solvation Models wrong for polar molecules in polar solvents?





-only electronic polarizibility

-homogeneously distributed

-linear response up to very high fields

➔dielectric continuum theory should be reasonably applicable

-discrete permanant dipoles

-mainly reorientational polarizibility

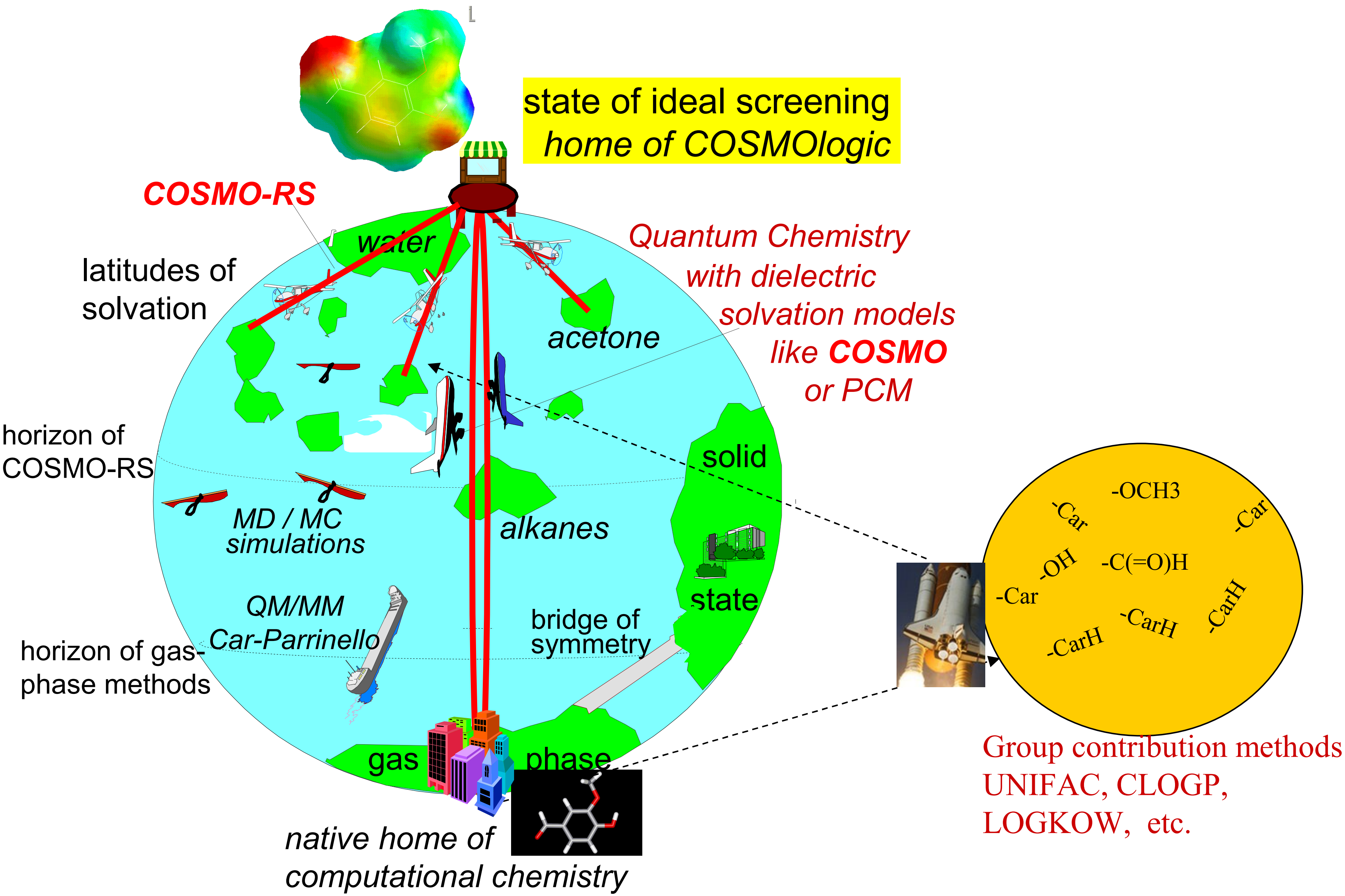
-linear response requires $E_{reor} << kT$

- typically $E_{reor} \sim 8$ kcal/mol !!!

➔no linear response, no homogenity

➔no similarity with dielectric theory

How to come to the latitudes of solvation?

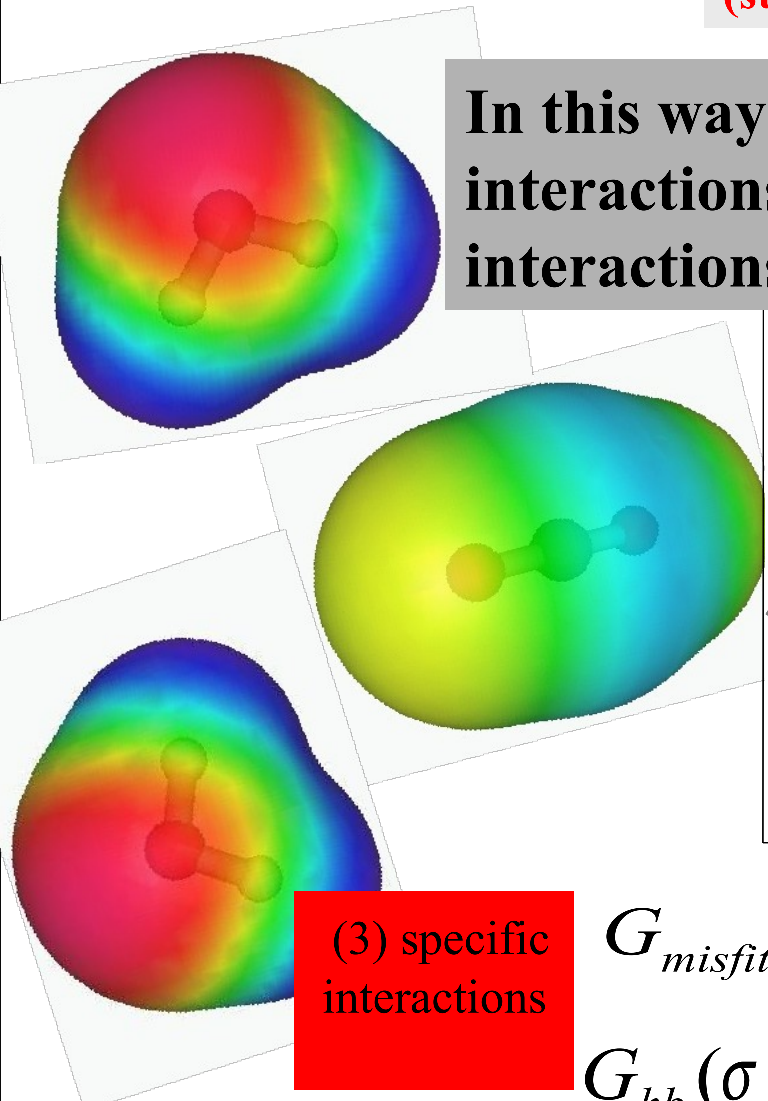


COSMO-RS:

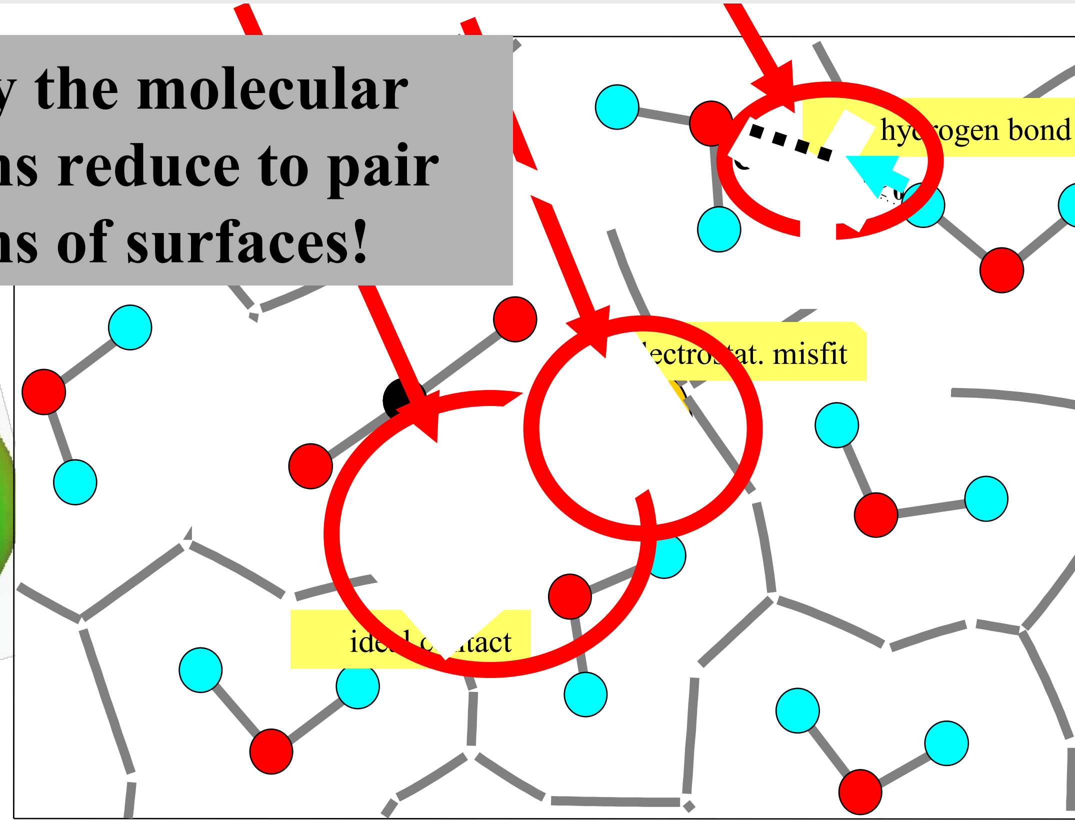
1) Put molecules into ‚virtual‘ conductor (DFT/COSMO)

2) Compress the ensemble to approximately right density

3) Remove the conductor on molecular contact areas (stepwise) and ask for the energetic costs of each step.



In this way the molecular interactions reduce to pair interactions of surfaces!



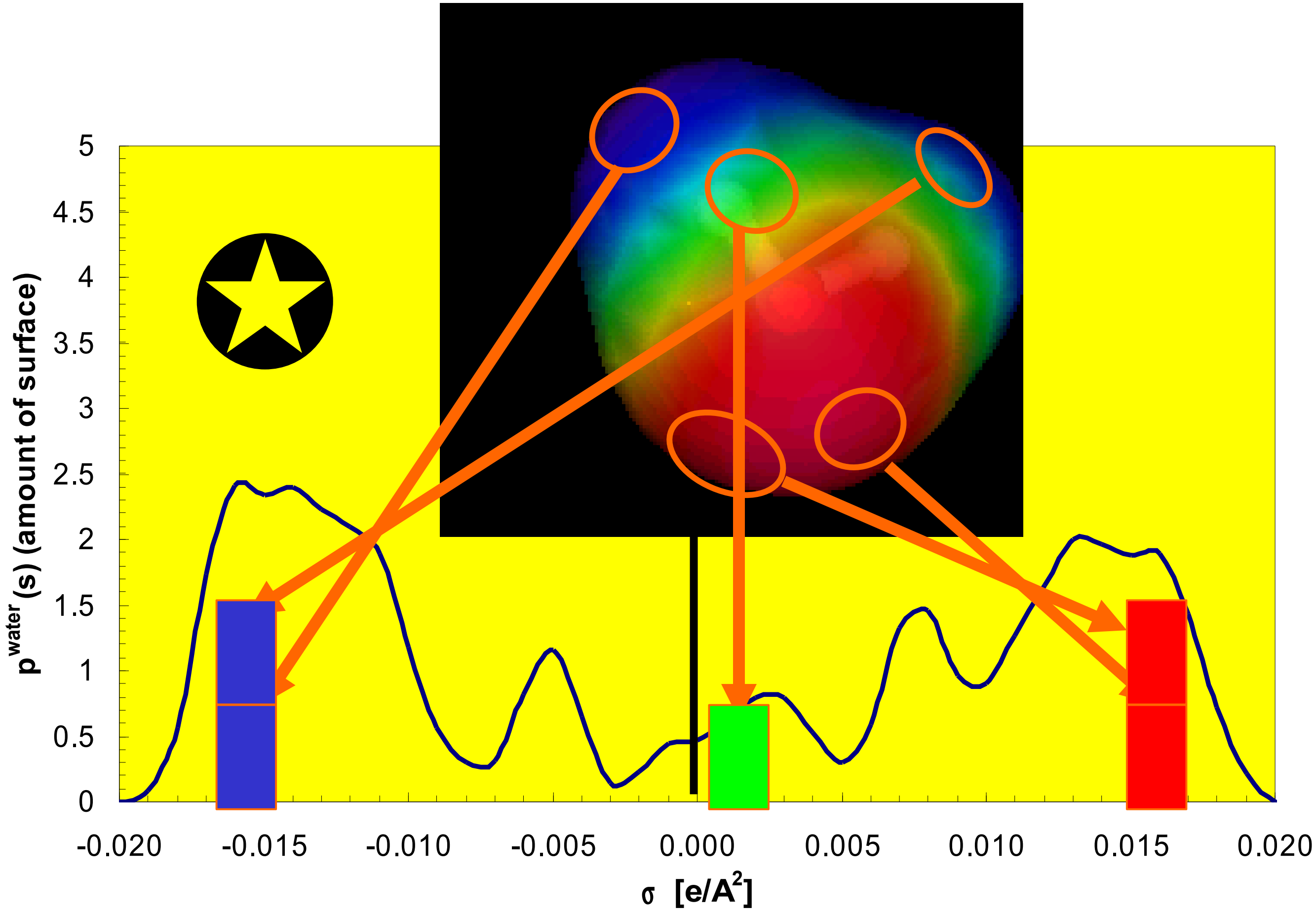
(3) specific interactions

$$G_{misfit}(\sigma,\sigma') = a_{eff}\frac{\alpha'}{2}(\sigma+\sigma')^2$$

$$G_{hb}(\sigma,\sigma') = a_{eff}\, c_{hb}(T)\min\{0;\sigma\sigma'+\sigma_{hb}^2\}$$

COSMO-RS

For an efficient statistical thermodynamics reduce the ensemble of molecules to an ensemble of pair-wise interacting surface segments !



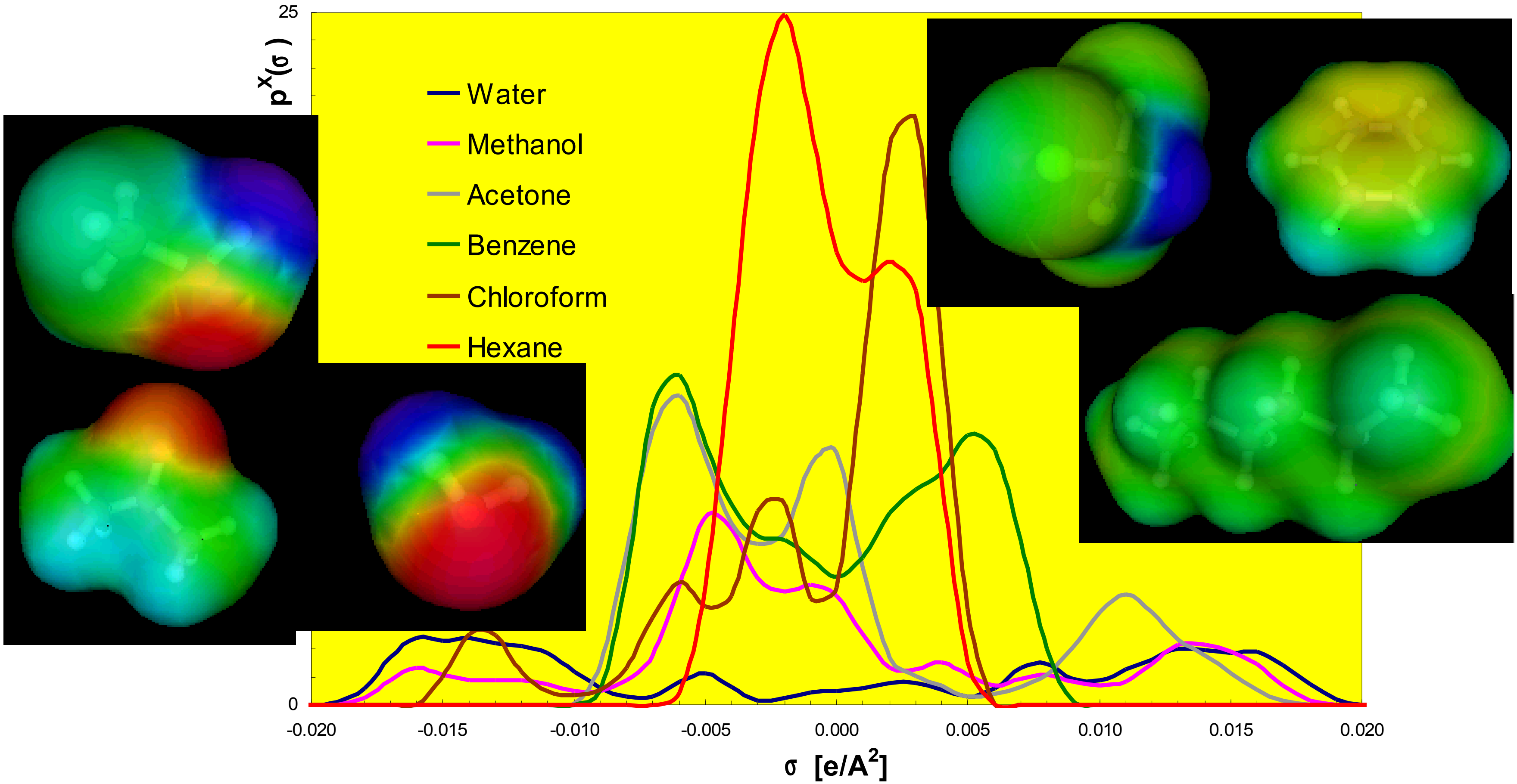
Screening charge distribution on molecular surface reduces to "σ-profile"

COSMO-RS

A. Klamt, *J. Phys. Chem.*, 99 (1995) 2224

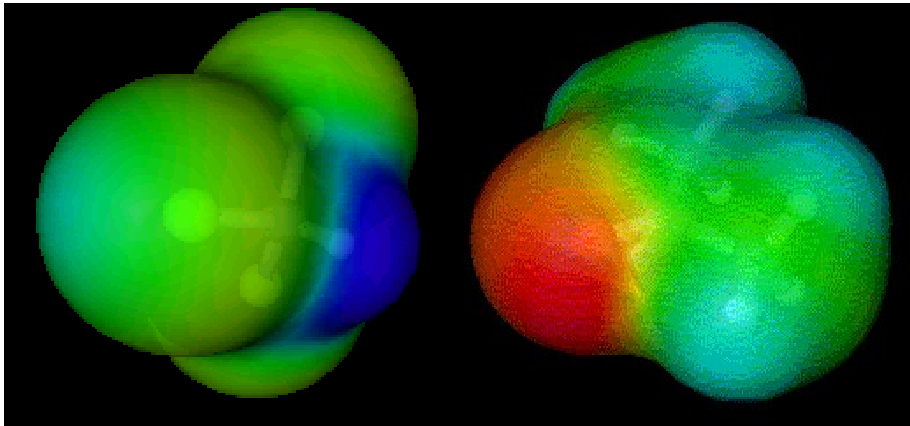
For an efficient statistical thermodynamics reduce the ensemble of molecules to an ensemble of pair-wise interacting surface segments !

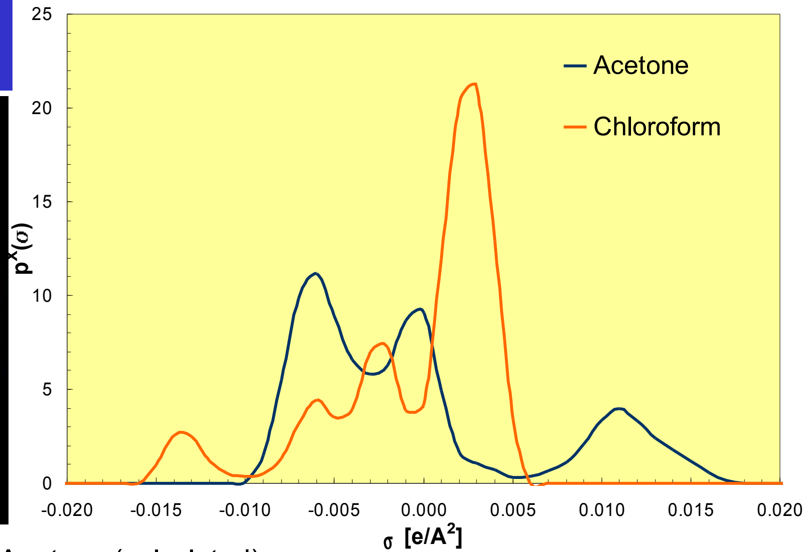
(same approximation as is UNIFAC)

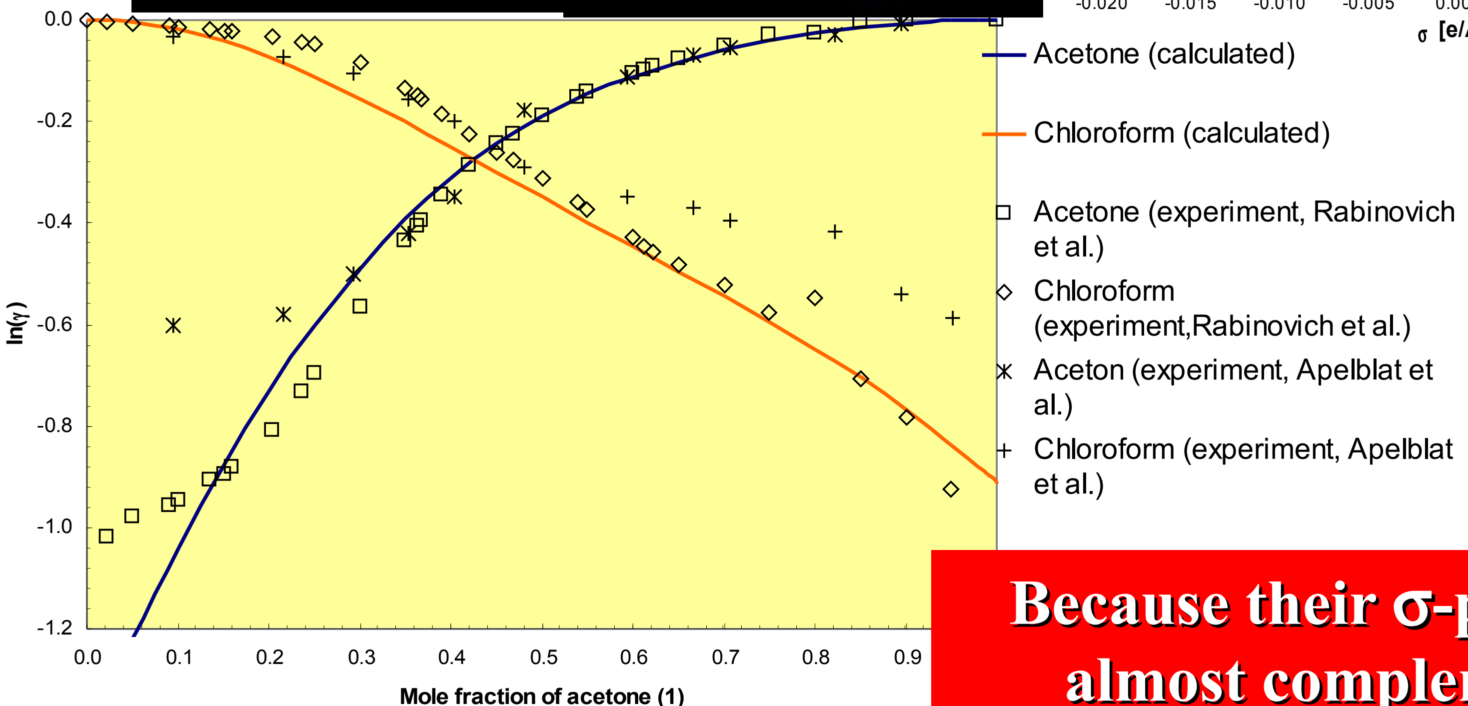


Screening charge distribution on molecular surface reduces to "σ-profile"

Why do acetone and chloroform like each other so much?







Because their σ-profiles are almost complementary!

Statistical Thermodynamics

- **Replace ensemble of interacting molecules by an ensemble S of interacting pairs of surface segments**
- **Ensemble S is fully characterized by its σ-profile $p_S(\sigma)$**

  ($p_S(\sigma)$ of mixtures is additive! -> no problem with mixtures!)
- **Chemical potential of a surface segment with charge density σ is exactly(!) described by:**

$$\mu_S(\sigma) = -kT \ln \int d\sigma' p_S(\sigma') \exp\left\{-\frac{E_{int}(\sigma,\sigma') - \mu_S(\sigma')}{kT}\right\}$$

σ-potential:
affinity of solvent for
specific polarity σ

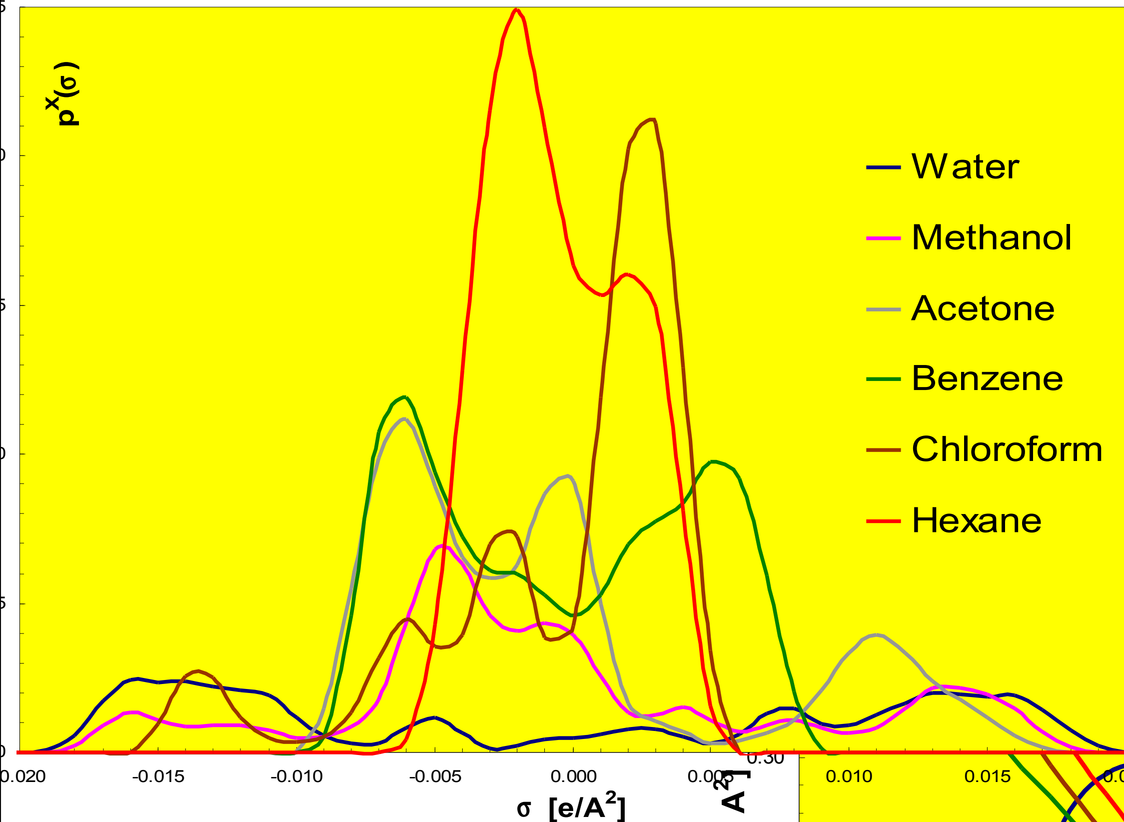
chemical potential of solute X in S:

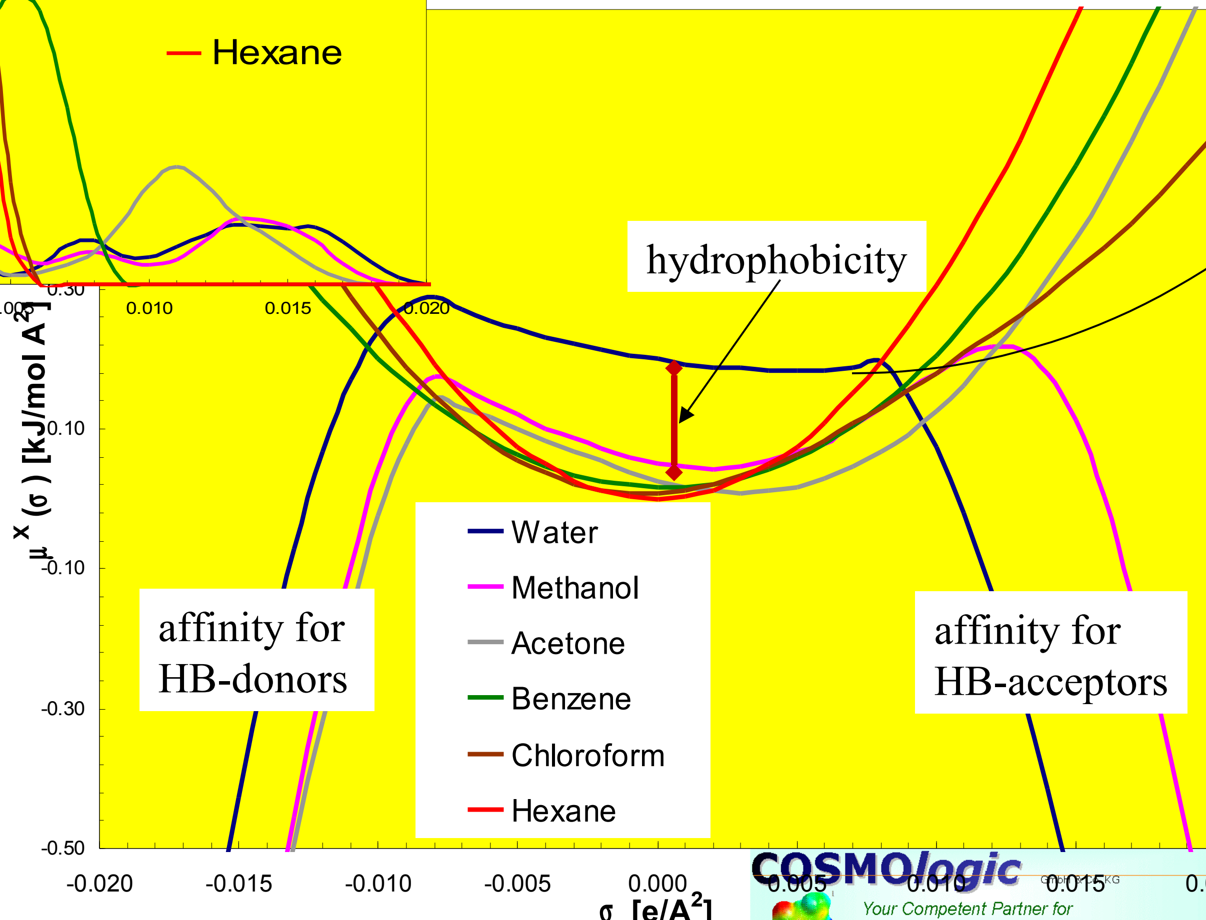
$$\mu_S^X = \int d\sigma\, p^X(\sigma)\mu_S(\sigma) - \lambda\, kT \ln \gamma_S^{X,comb}$$

combinatorial contribution:
solvent size effects

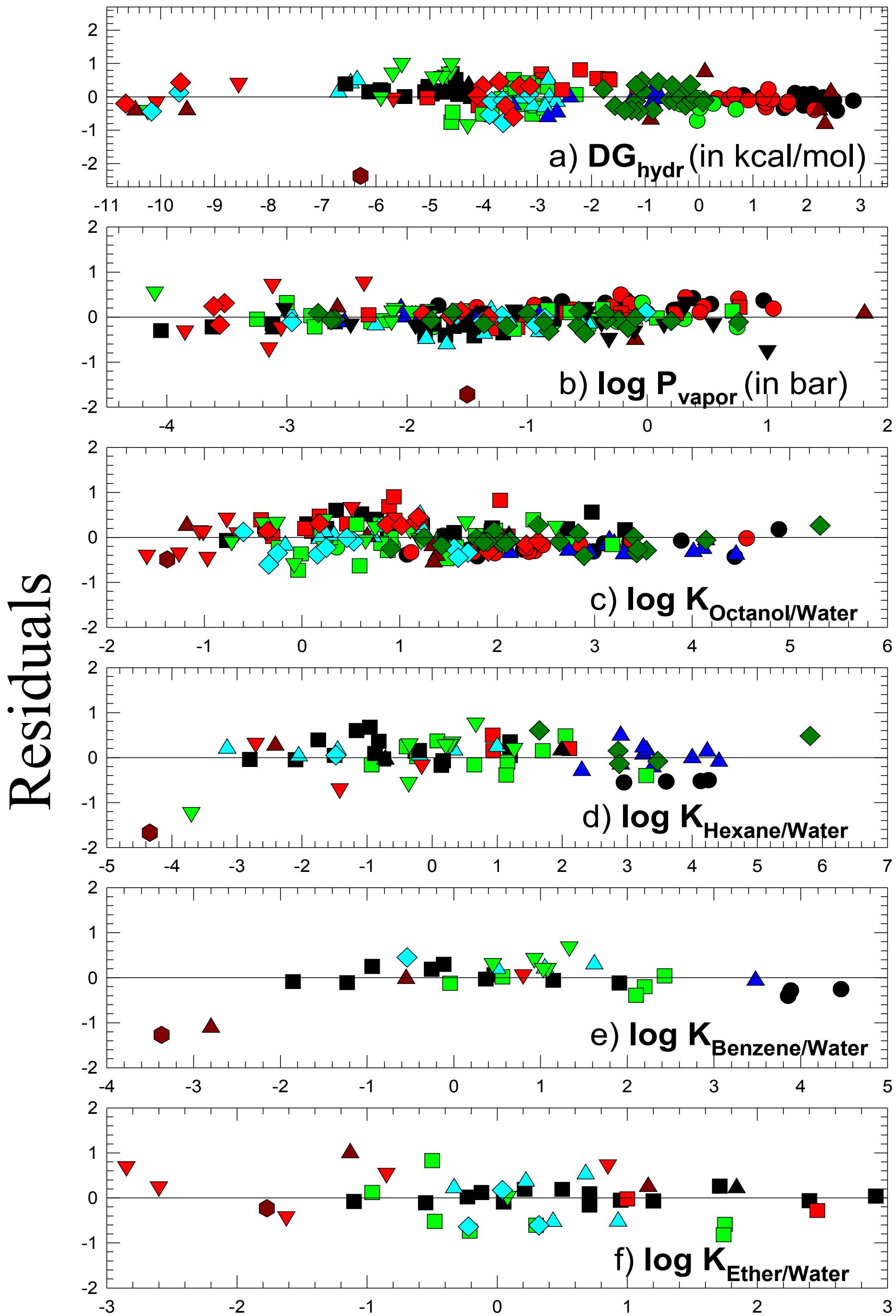
activity coefficients → arbitrary liquid-liquid equilibria

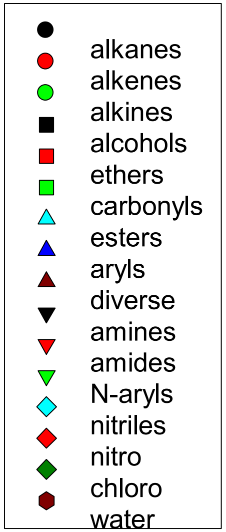
σ-profiles and σ-potentials of representative liquids





COSMO*logic*
Your Competent Partner for
Computational Chemistry and Solvation





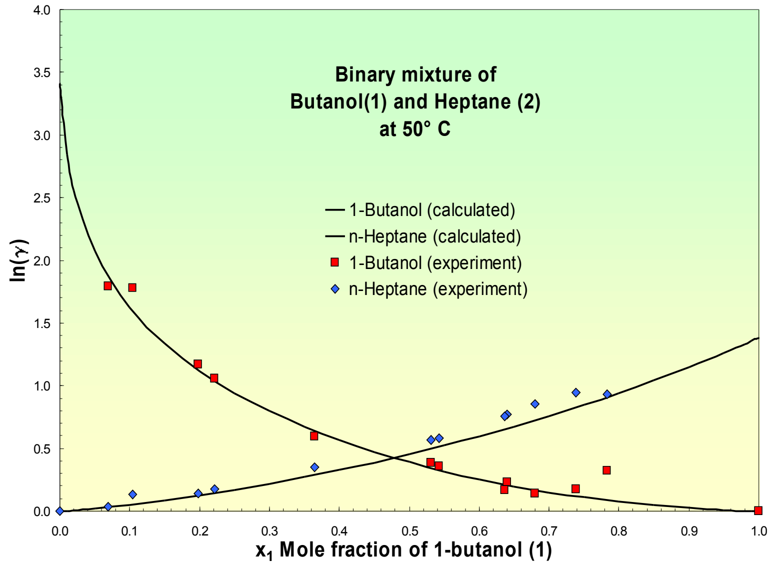
Results of parametrization based on DFT (DMol³: BP91, DNP-basis

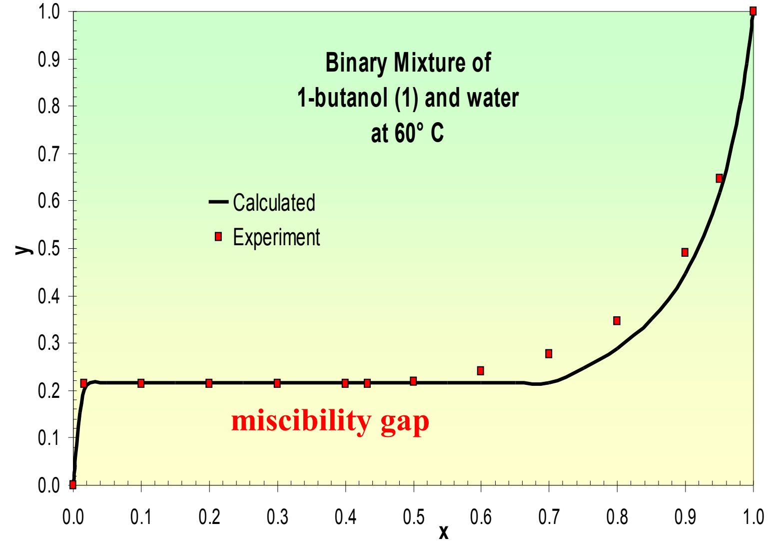
650 data
17 parameters
rms = 0.41 kcal/mol

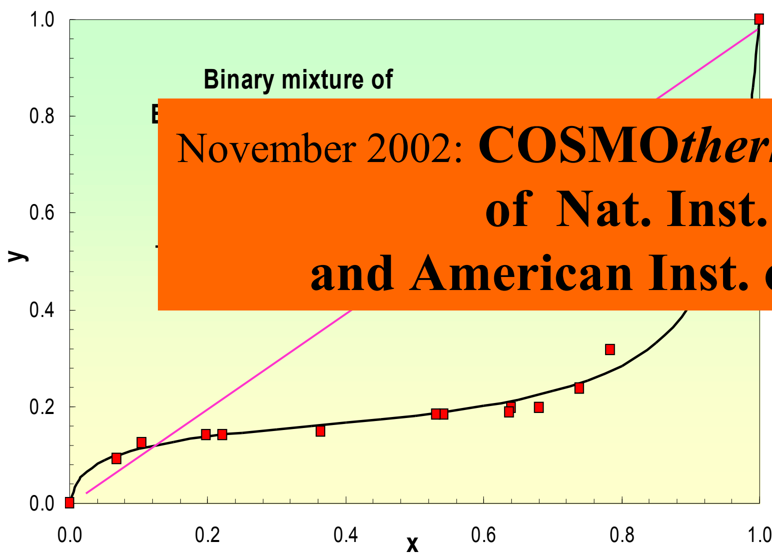
A. Klamt, V. Jonas, J. Lohrenz, T. Bürger, J. Phys. Chem. A, 102, 5074 (1998)

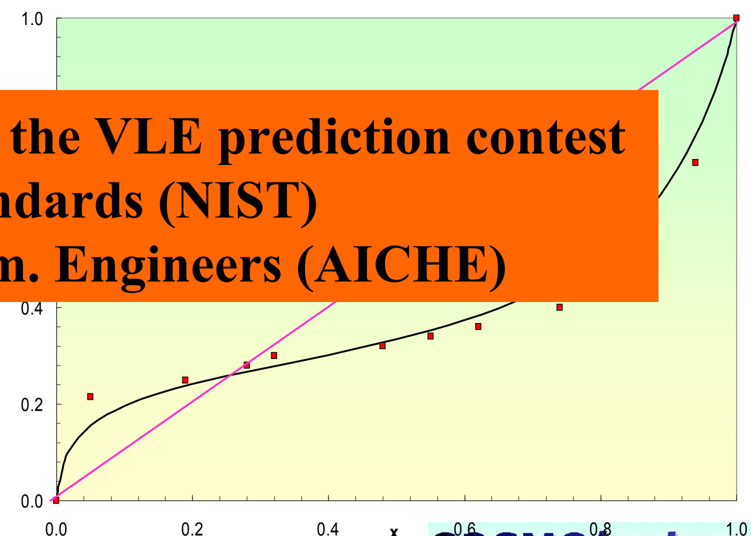
meanwhile:
COSMO*therm5.0* with Turbomole BP91/TZVP
rms = 0.36 kcal/mol

Applications to Phase Diagrams and Azeotropes





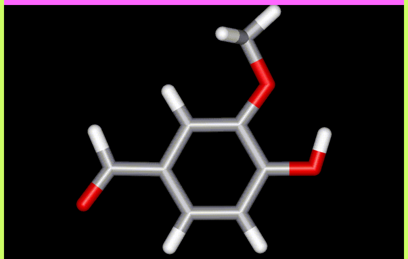




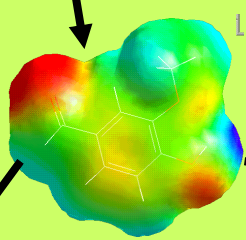
November 2002: **COSMO*therm* wins the VLE prediction contest of Nat. Inst. of Standards (NIST) and American Inst. of Chem. Engineers (AICHE)**

Flow Chart of COSMO-*RS*

Chemical Structure



Quantum Chemical Calculation with COSMO (full optimization)

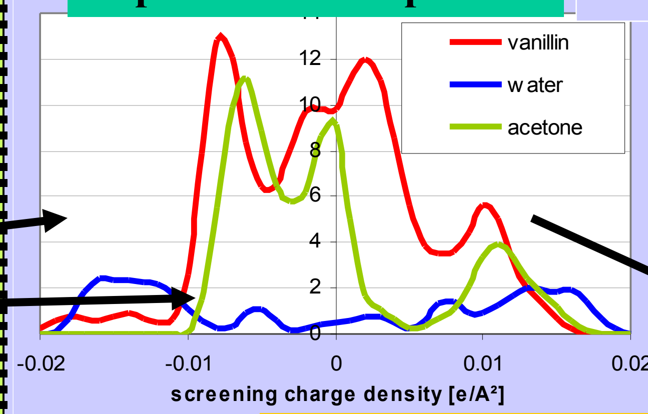


ideally screened molecule energy + screening charge distribution on surface

Database of COSMO-files (incl. all common solvents)

other compounds

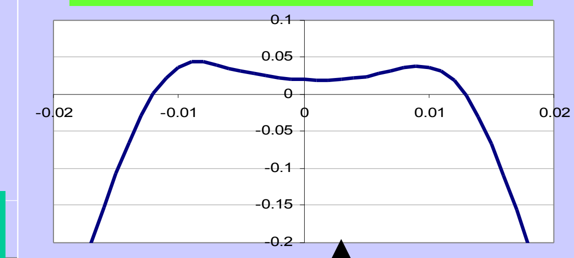
σ-profiles of compounds



σ-profile of mixture

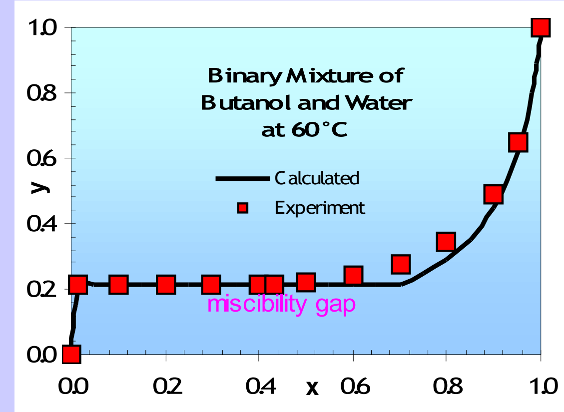
Fast Statistical Thermodynamics

σ-potential of mixture



Equilibrium data: activity coefficients vapor pressure, solubility, partition coefficients

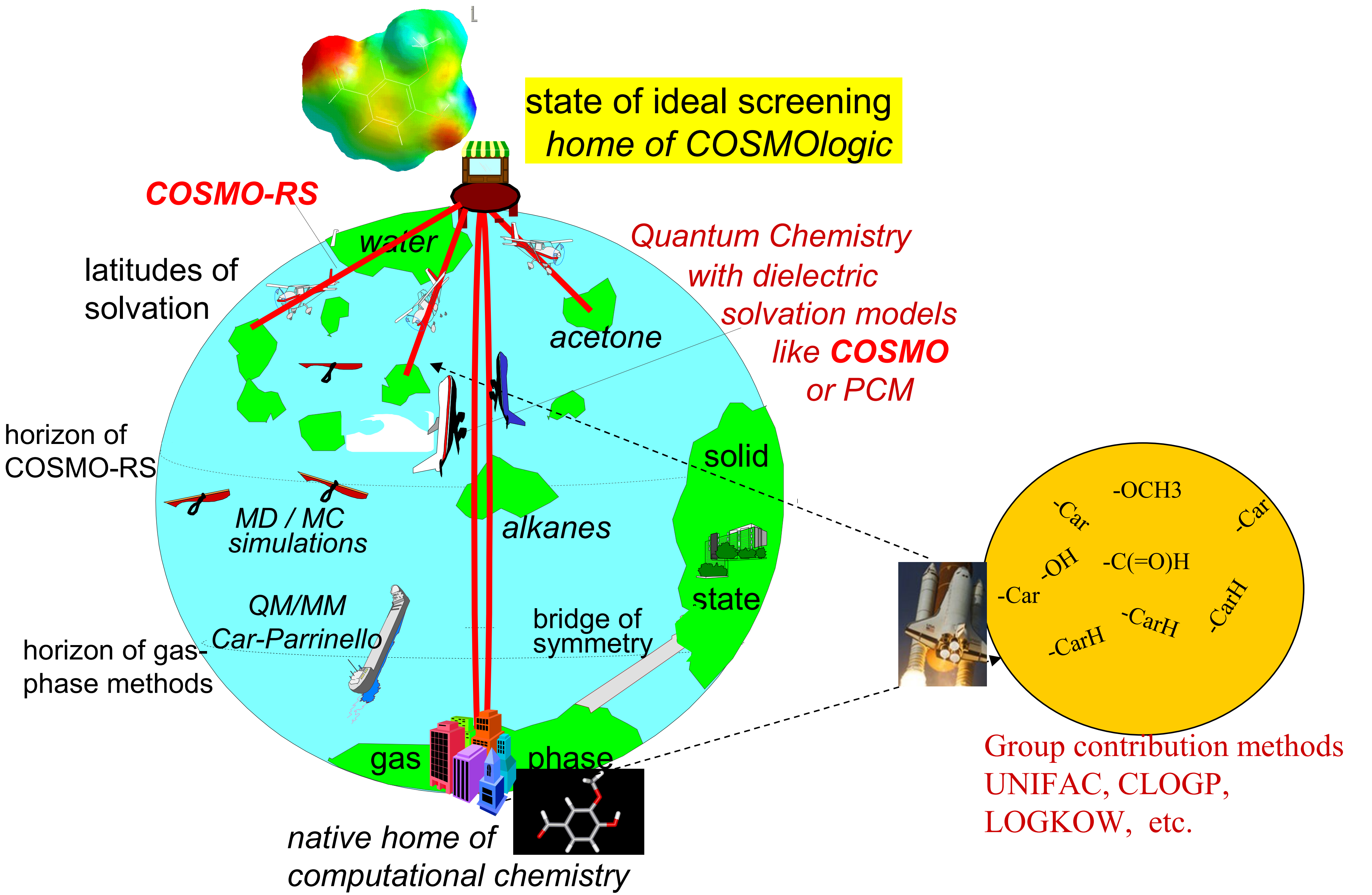
Phase Diagrams



DFT/COSMO

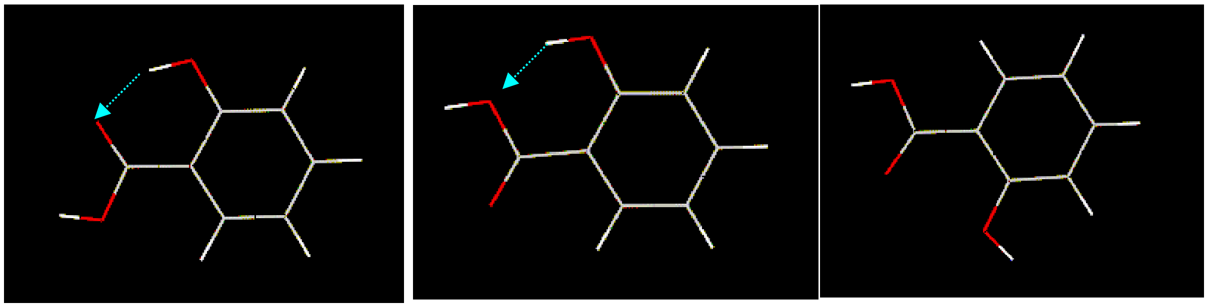
COSMO*therm*

How to come to the latitudes of solvation?



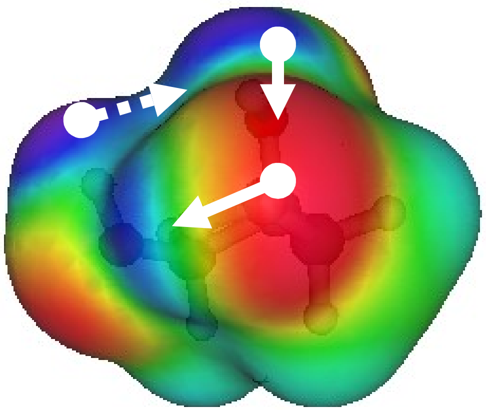
Extension of COSMO*therm* to multi-conformations

Unfortunately, many molecules have more than one relevant conformation



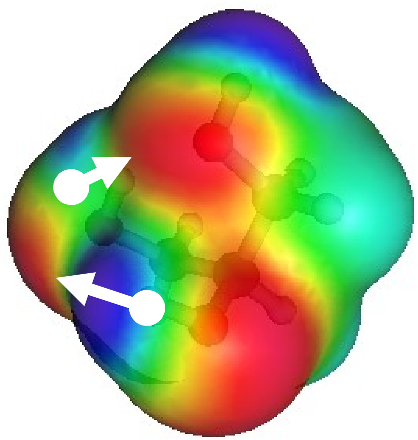
COSMO*therm* can treat a compound as a set of several conformers
- each conformer needs a COSMO calculation
- conformational population is treated consistently according to total free energy of conformers (by external self-consistency loop)

Conformational effects for glycerol



lowest COSMO conformer
all 3 donors are bound in
one 6-ring and two 5-rings,
also least polar conformer

partition coefficient between



Conclusions:

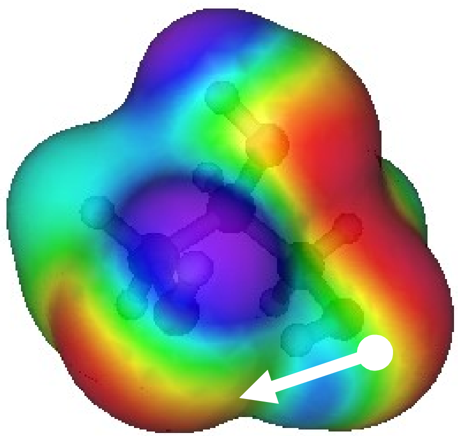
- Conformational effects can be important for the detailed understanding of phase equilibria

- In most cases one conformation dominates in all phases

- Effects are especially large for molecules with sub-optimal intramolecular HBs in solvents having strong HB acceptors, but a deficit of HB-donors.

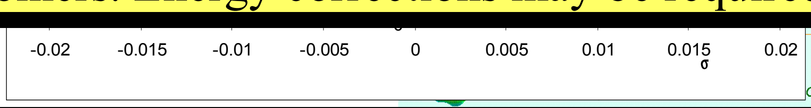
-Tautomers can be considered as a kind of conformers.

-Unfortunately the DFT level of QC is not always reliable regarding the energy differences between conformers and even more between tautomers. Energy corrections may be required.



2% in octane

41% in acetone

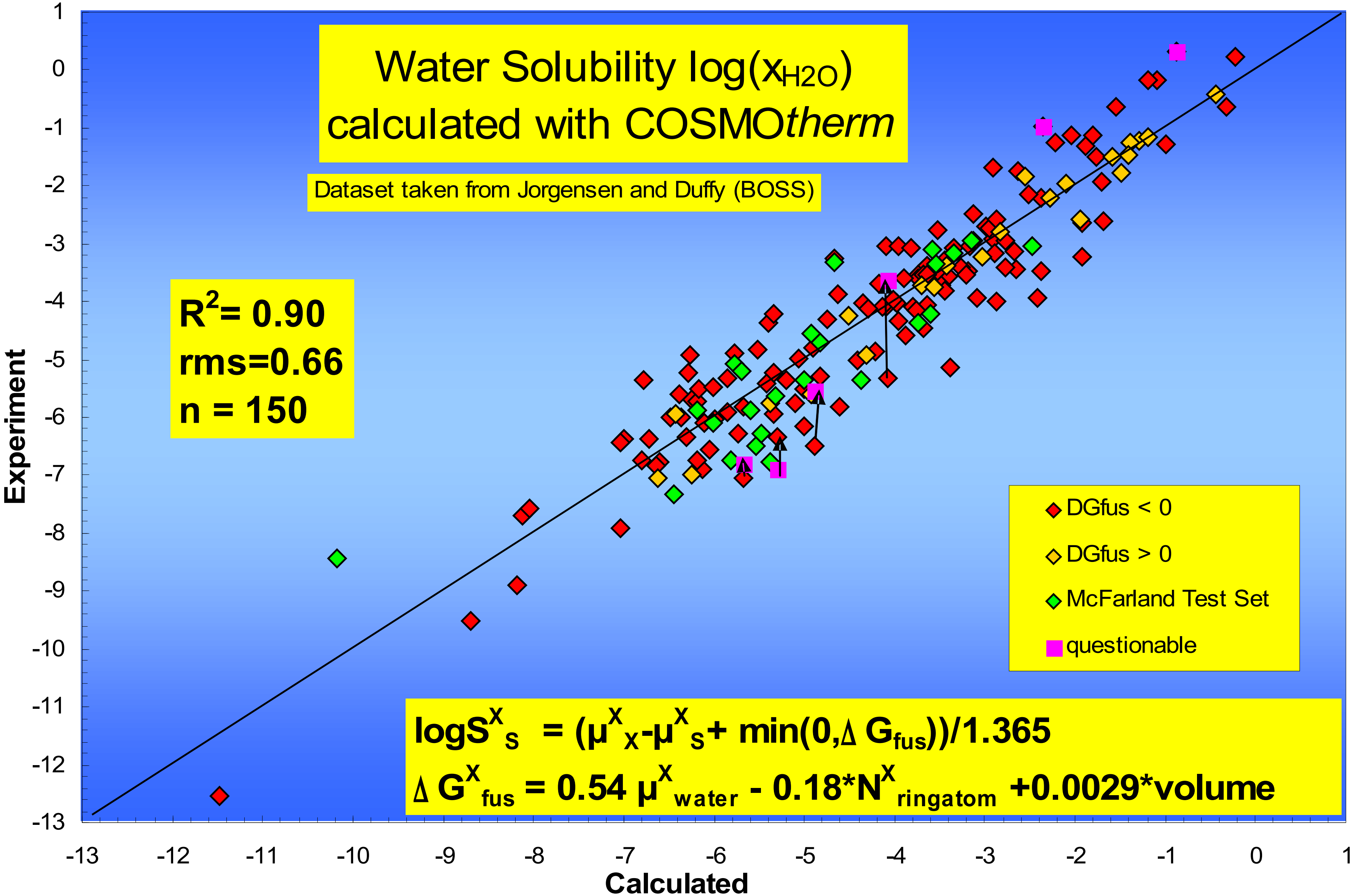


„Conformational analysis of cyclic acidic α-amino acids in aqueous solution - an evaluation of different continuum hydration models."

by Peter Aadal Nielsen, Per-Ola Norrby, Jerzy W. Jaroszewski, and Tommy Liljefors
(private comm., Ph.D. thesis)

Method	*Solvent Model*	*rms (kJ/mol)*	*rms (4 points) (kJ/mol)*	*Max Dev (kJ/mol)*
AM1	SM5.4A	4.6	5.6	9.2
PM3	SM5.4P	13.6	16.2	20.5
AM1	SM2.1	7.4	9.0	16.7
HF/6-31+G*	C-PCM	3.1	3.8	5.9
HF/6-31+G*	PB-SCRF	4.7	5.8	8.8
AMBER*	GB/SA	13.2	16.2	24.3
MMFF	GB/SA	18.5	19.9	31.4
BP-DFT/TZVP	COSMO-RS	2.2	2.6	4.8

COSMO-RS was evaluated as a blind test !!!



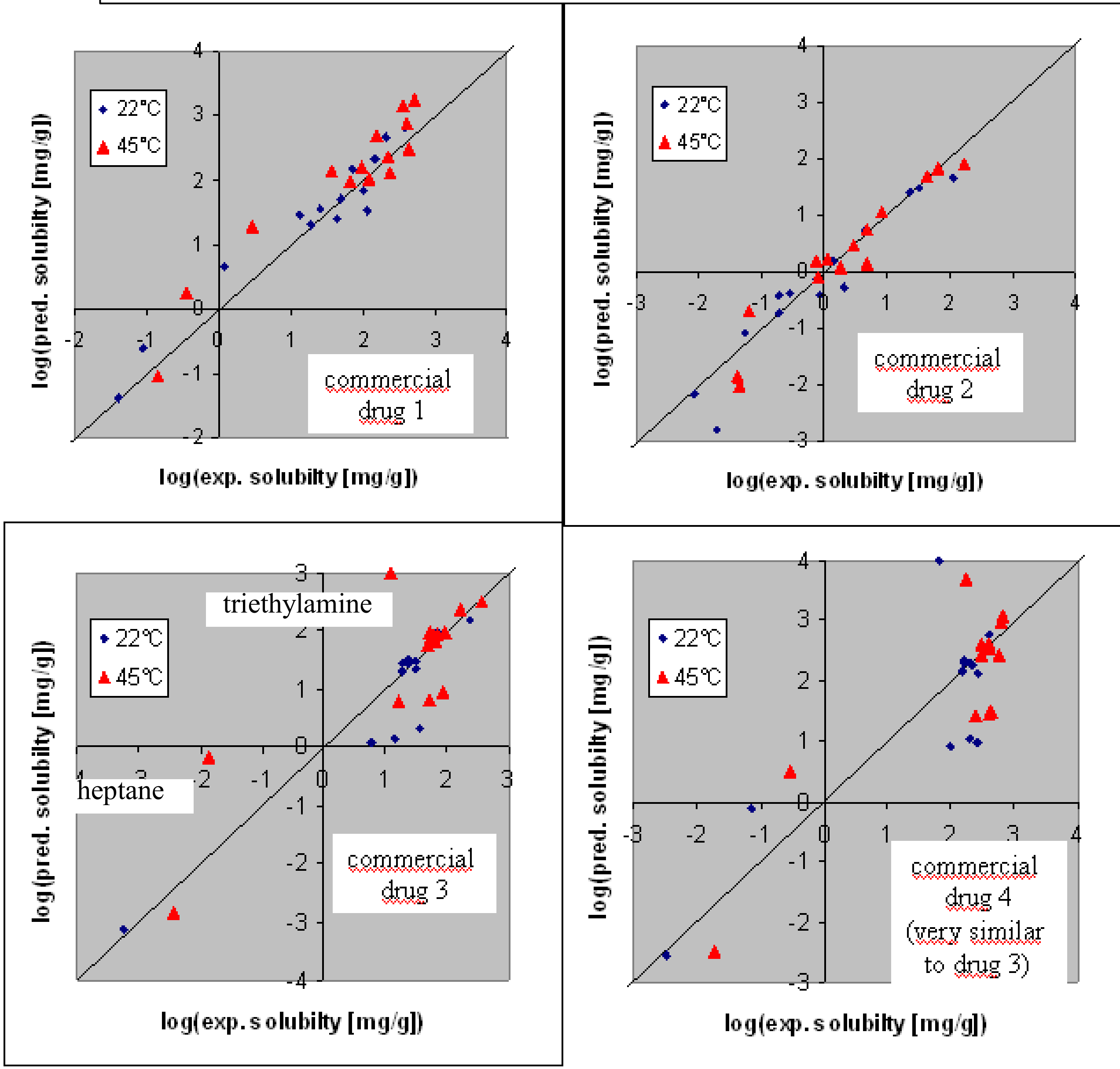
Stable model: No changes required for pesticides!

A.Klamt, F. Eckert, M. Hornig, M. Beck, and T. Bürger:
J. Comp. Chem. 23, 275-281 (2002)

COSMO*logic* GmbH & Co. KG
Your Competent Partner for Computational Chemistry and Solvation

COSMO*therm* prediction of drug solubility in diverse solvents

(blind test performed with Merck&Co., Inc., Rahway, NJ, USA)



all predictions are relative to ethanol

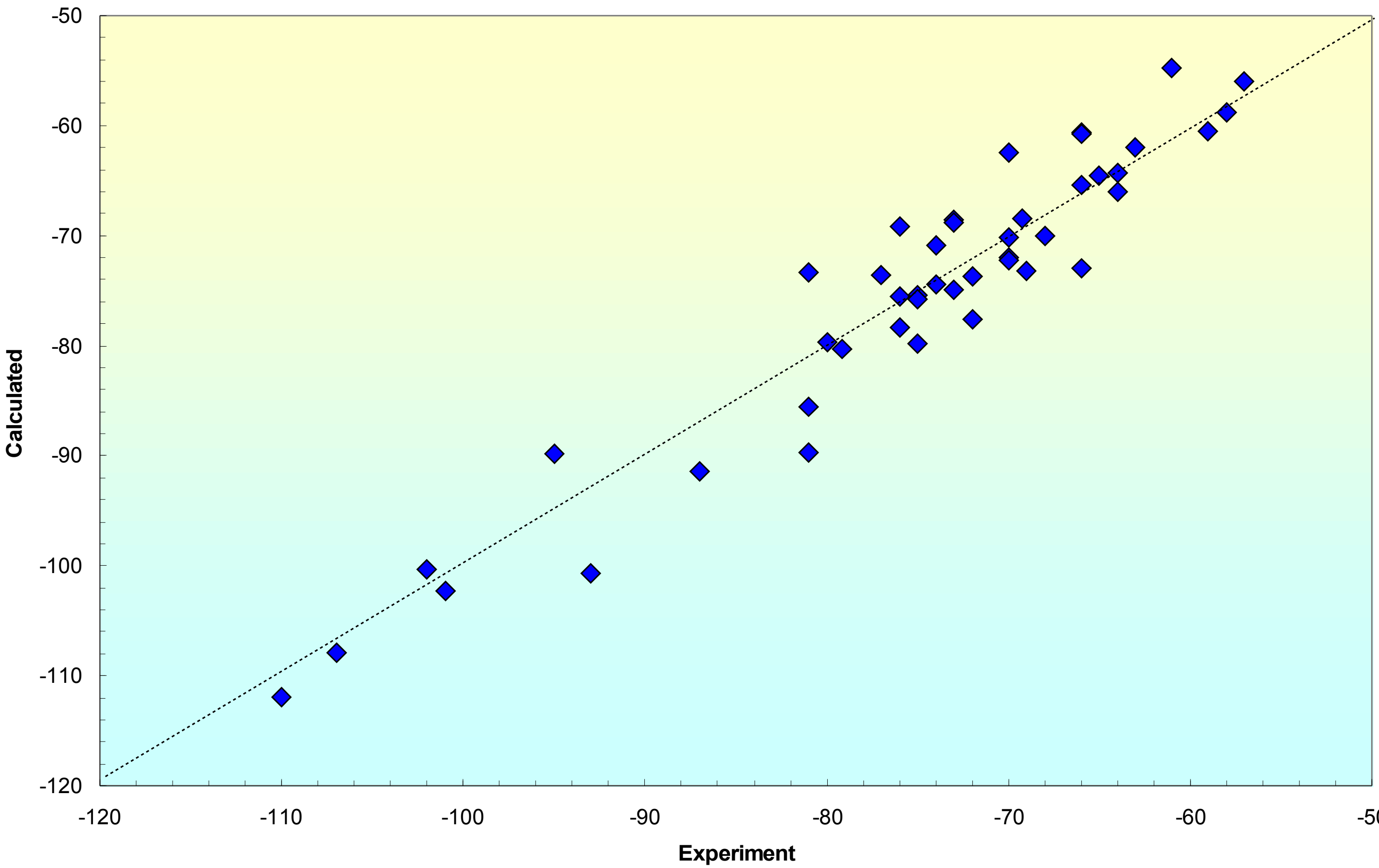
solvents:

- Water
- 1-Propanol
- 2-Propanol
- DMF
- Ethyl Acetate
- Methanol
- Heptane
- Toluene
- Chlorobenzene
- Acetone
- Ethanol
- Acetonitrile
- Triethylamine
- Butanol

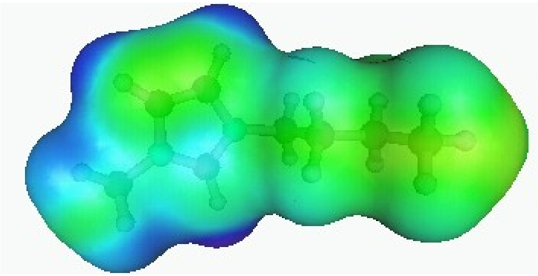
COSMO*logic* GmbH & Co. KG
Your Competent Partner for Computational Chemistry and Solvation

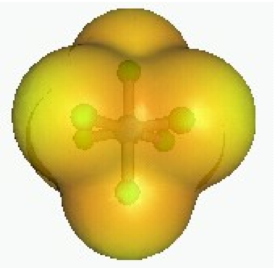
Ionic Free Energies of Hydration by COSMOtherm-Ion-Extension

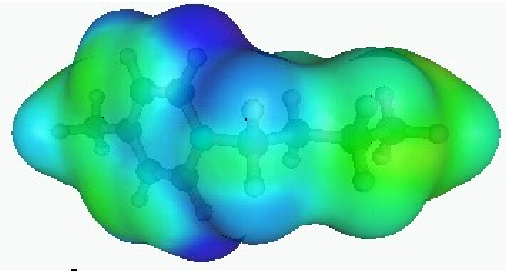
Free energy of Hydration [kcal/mol] for Ions

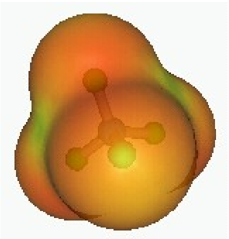


Applications of COSMO*therm* to Ionic Liquids



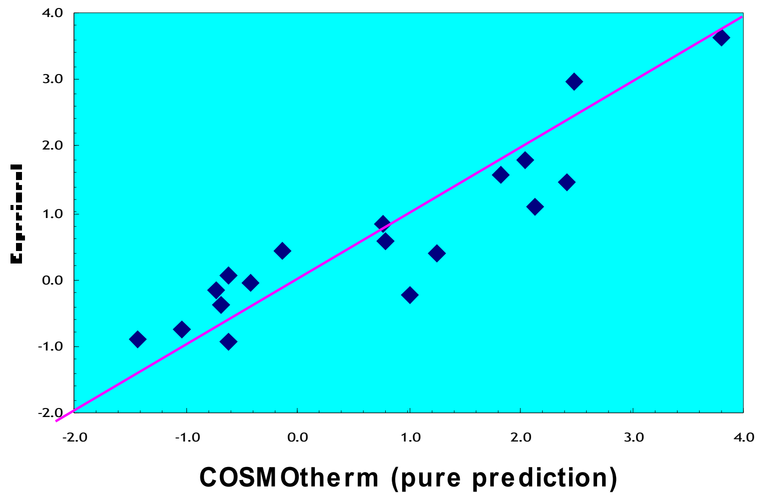




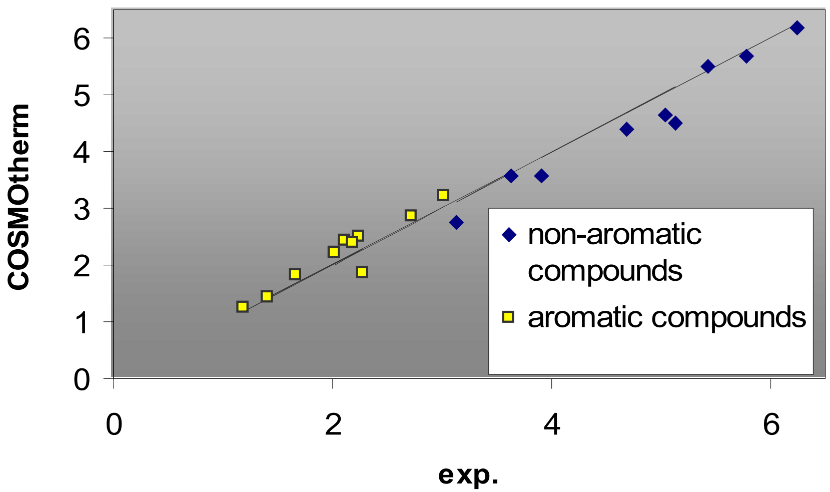


log(Partition) for H_2O / 1-butyl-3-methyl-imidazolium(+) - PF_6(-)

exp.: J.G. Huddleston,University of Alabama



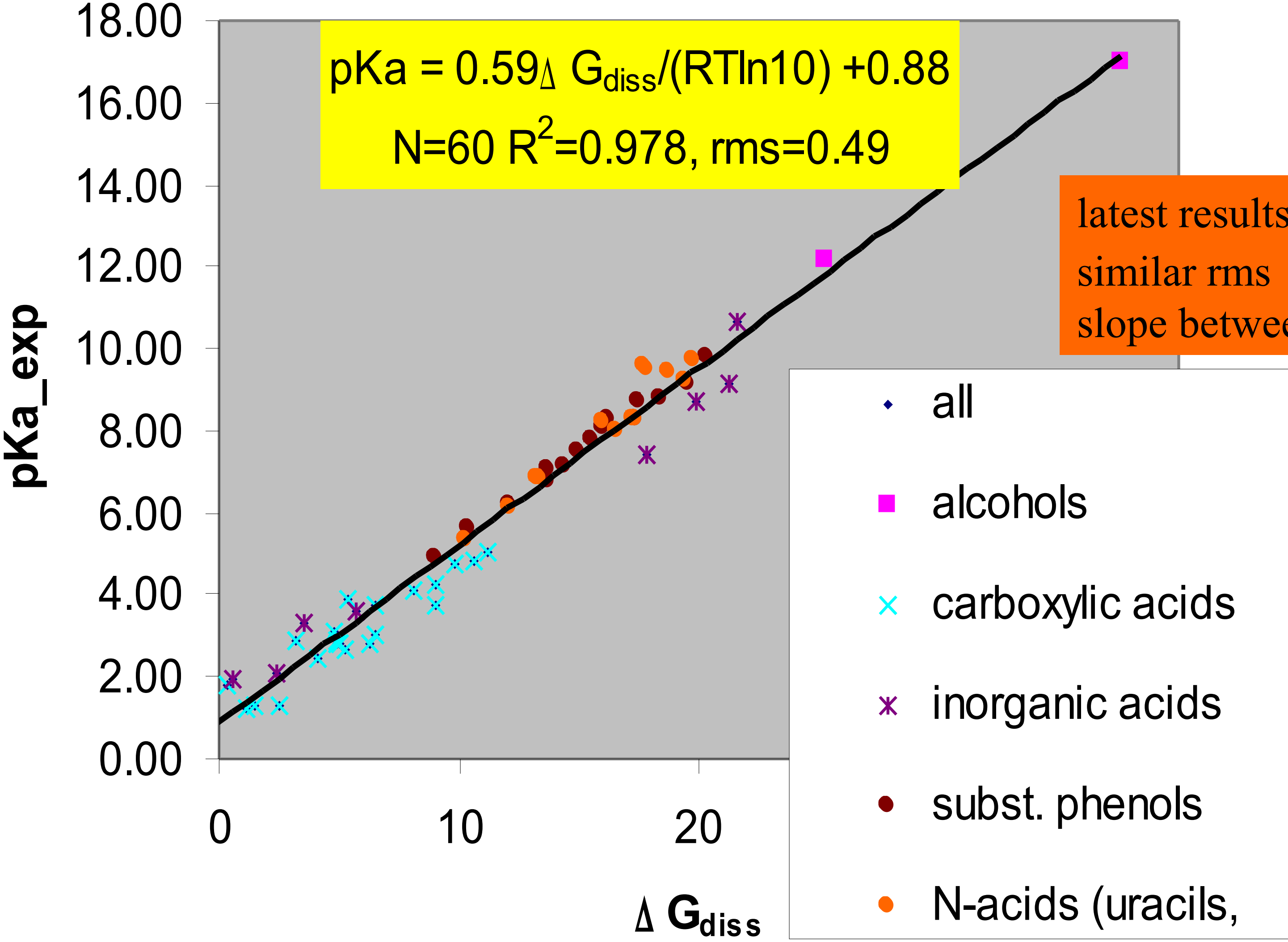
ln(gamma_inf) calc. / exp. (T=314/333K)
in 4-methyl-n-butylpyridinium BF4
Lit: Andreas Heintz, Dmitry V. Kulikov, Sergey P. Verevkin, J. Chem. Eng. Data 2001, 46, 1526-1529



COSMO*therm* appears to work well for Ionic Liquids

COSMO*therm* first principle pKa prediction

(A. Klamt, et. al. J. Phys. Chem. A, Nov. 2003)

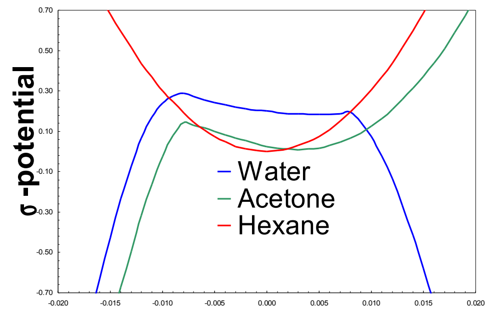


latest results for bases (pK_b):
similar rms
slope between 0.59 and 0.71

- formicacid
- aceticacid
- chloroaceticacid
- dichloroaceticacid0
- trichloroaceticacid
- n-pentanoicacid
- dimethylpropanoicacid
- benzoicacid
- oxalicacid0
- maleicacid3
- fumaricacid
- 2,2,2-trichloroethanol
- hypochlorousacid
- hypobromousacid
- hypoiodousacid
- nitrousacid
- sulfurousacid
- phosphoricacid2
- boricacid
- 5-fluorouracil
- 5-nitrouracil
- cis-5-formyluracil
- thymine
- trans-5-formyluracil
- Uracil
- and others

σ-Moment Approach

$$\mu_S(\sigma) \cong \sum_{i=-2}^{m} c_S^i f_i(\sigma) \quad with \quad f_i(\sigma) = \sigma^i \quad for\ i \geq 0\ and$$



$$f_{-2/-1}(\sigma) = f_{acc/don}(\sigma) \cong \begin{cases} 0 & if \quad \pm\sigma < \sigma_{hb} \\ \mp\sigma + \sigma_{hb} & if \quad \pm\sigma > \sigma_{hb} \end{cases}$$

Now the chemical potential of a solute X in this matrix S is:

$$\mu_S^X = \int p^X(\sigma)\mu_S(\sigma)\,d\sigma \cong \int p^X(\sigma)\sum_{i=-2}^{m} c_S^i f_i(\sigma)\,d\sigma \cong \sum_{i=-2}^{m} c_S^i M_i^X$$

$$with \quad M_i^X = \int p^X(\sigma) f_i(\sigma)\,d\sigma \quad = \sigma - moments\ of\ solute\ X$$

The coefficients can now be derived from experimental (log.) partition data by linear regression. => σ-moments are excellent QSAR-descriptors for general partition behaviour of molecules.

"The solvent space is approximately 5-dimensional!"

Zissimos, et al.: 'A comparison between the two general sets of linear free energy descriptors of Abraham and Klamt', J. Chem. Inf. Comput. Sci., **42**, 1320-1331 (2002)

→ σ -moment models for ADME proprties as

logBB, intestinal absorption, logHSA, …

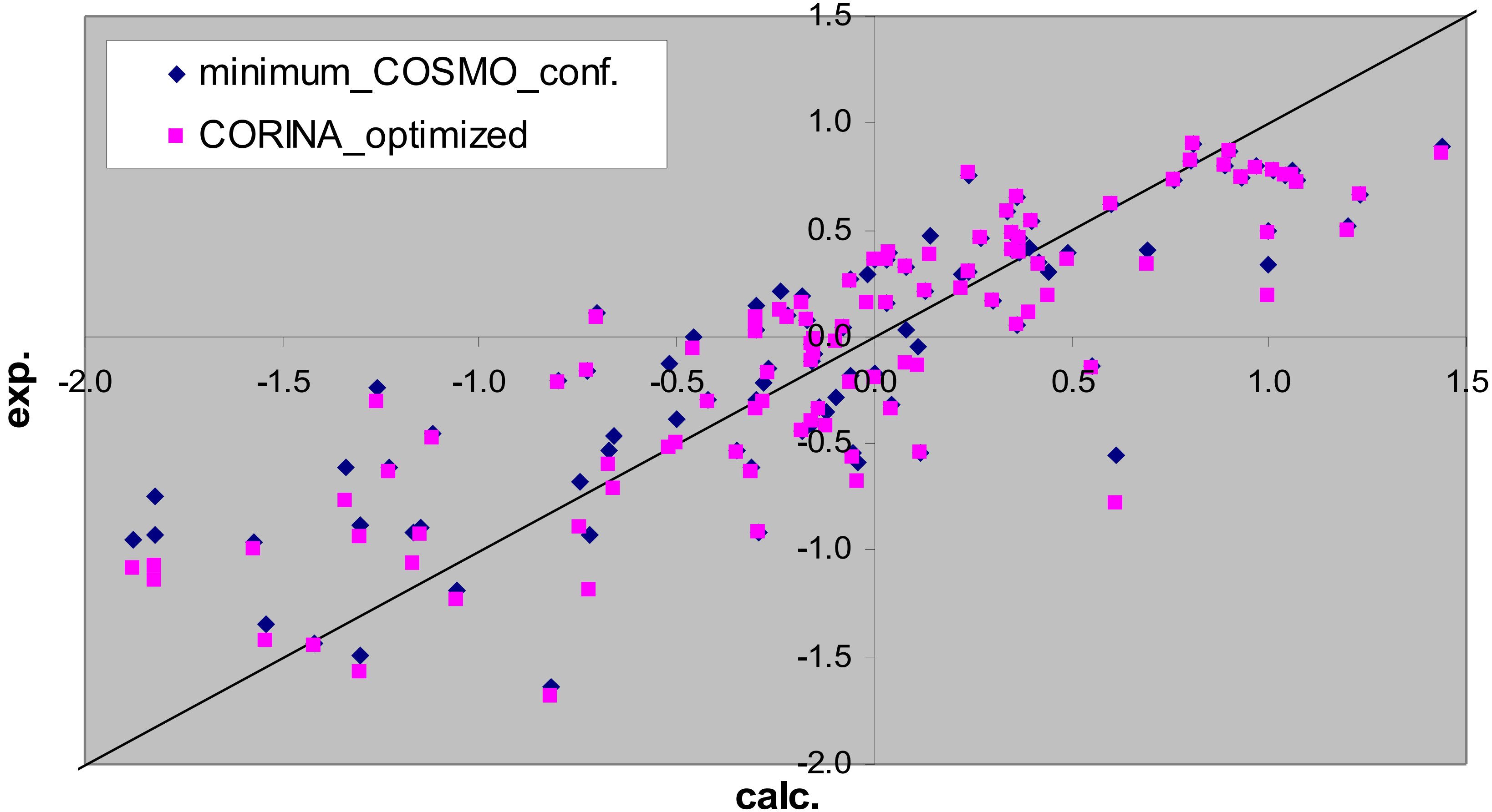
COSMO*logic* GmbH & Co. KG
Your Competent Partner for Computational Chemistry and Solvation

σ -moment logBB regression

logBB = 0.0046 area -0.017 sig2 -0.0029 sig3 +0.19

$n = 103, r^2 = 0.71, rms = 0.40$

data from: "Modeling Blood-Brain Barrier Partitioning Using Topological Structure Descriptors", Rose, Hall, Hall, and Kier, MDL-Whitepaper, 2003

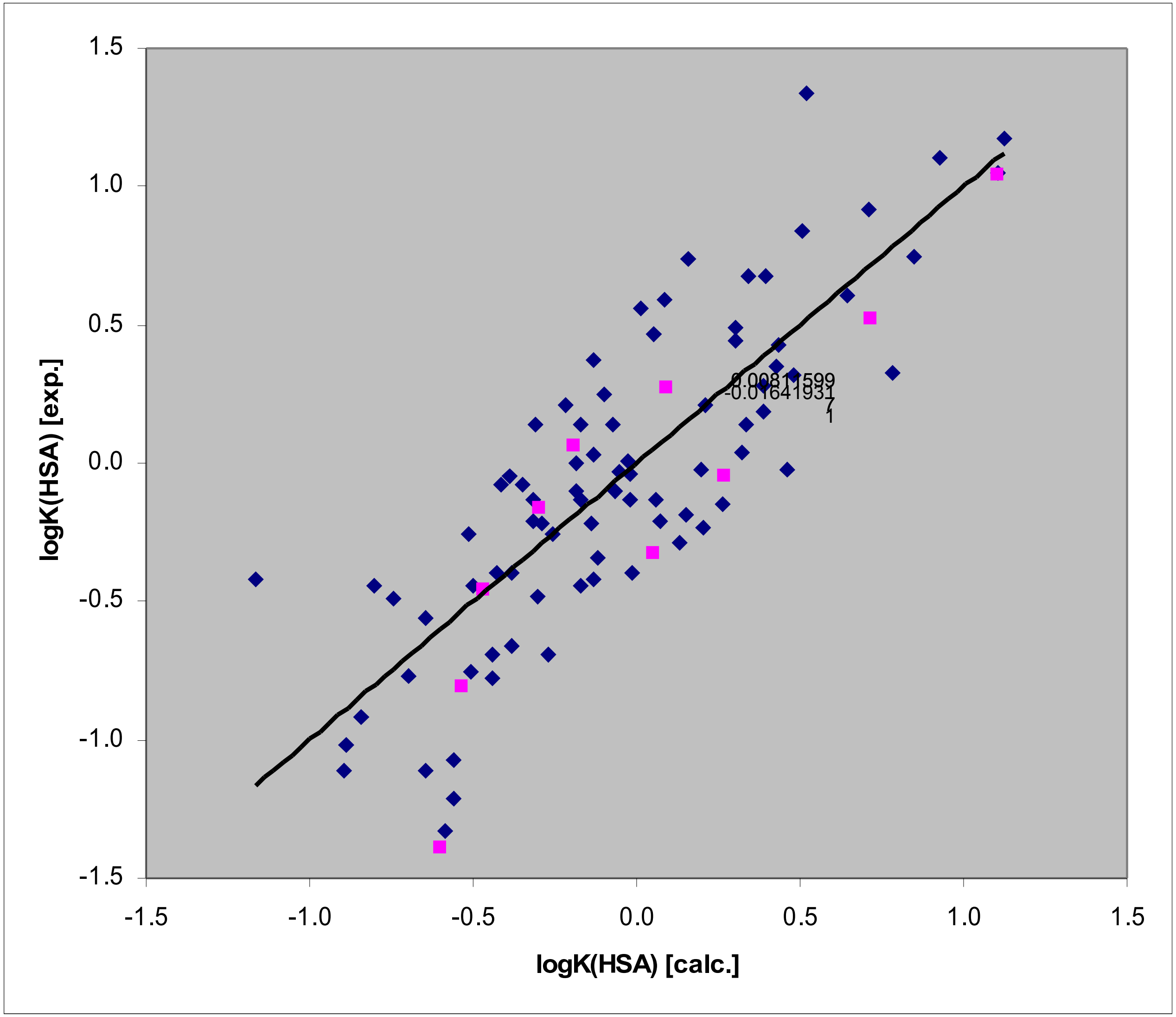


σ-moment logKHSA regression

logKHSA = 0.0081 area -0.016 sig2 -0.013 sig3 +0.145 sigHacc+0.88

n = 82, r^2 = 0.69, rms = 0.33

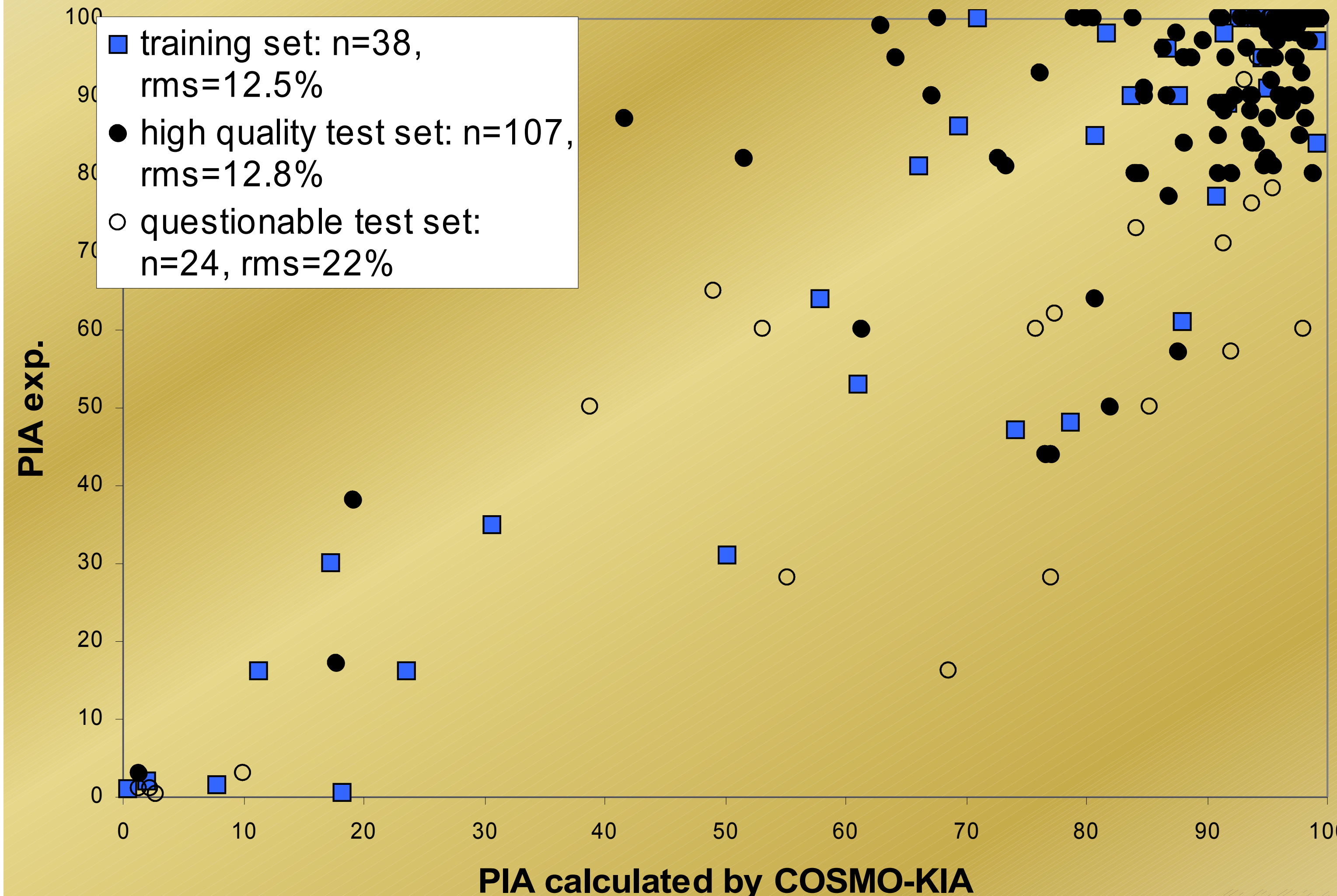
data from: Kier, Hall, Hall, MDL-Whitepaper, 2002



COSMO-RS for Percentage Intestinal Absorption (PIA)

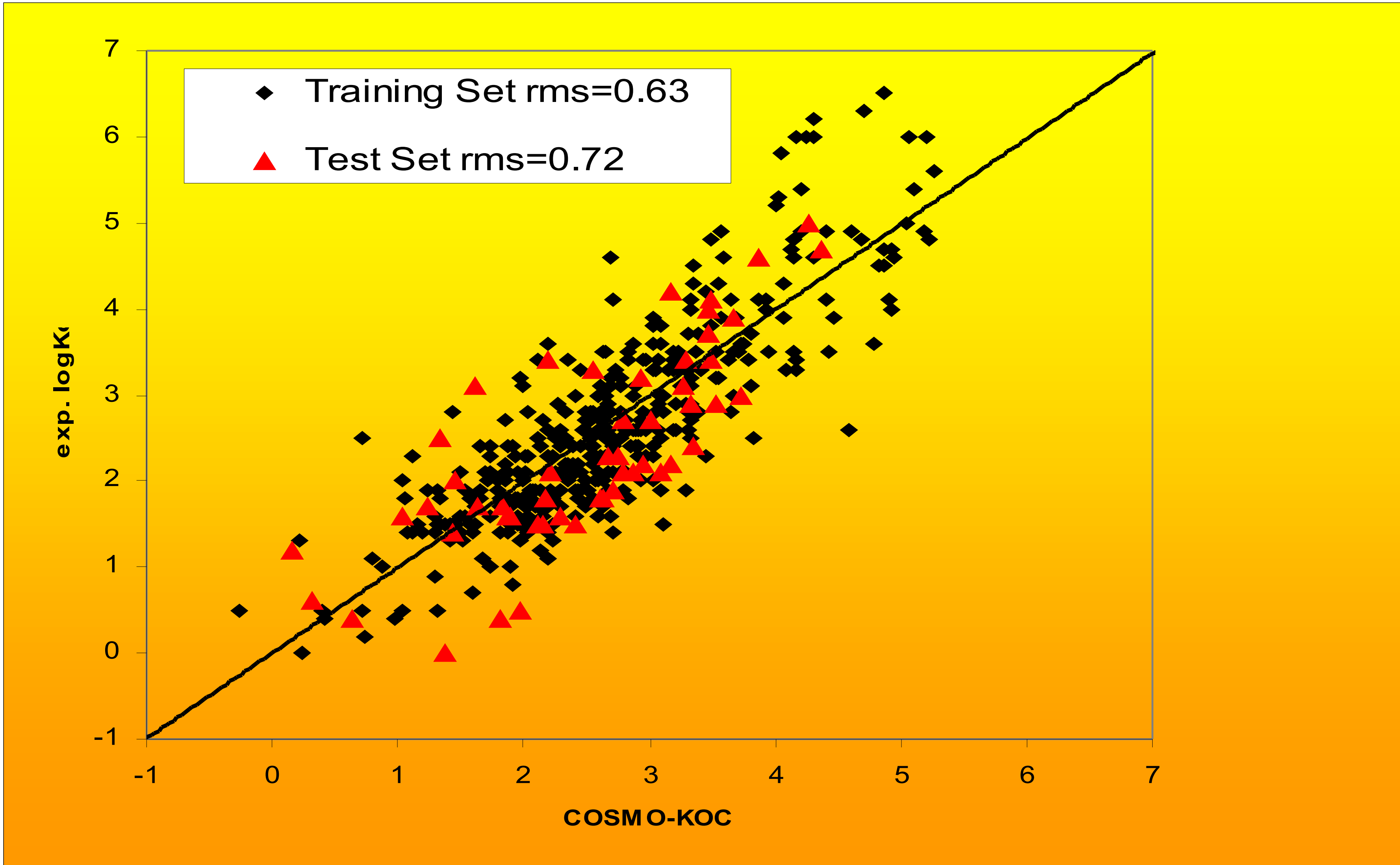
Klamt, Diedenhofen, Connolly*, Jones* (submitted) *) GlaxoSmithKline

$$\log KIA = 0.0040M_0 - 0.0053M_2 - 0.0024M_3 - 0.113M_{acc} - 0.117M_{don} + 1.37$$

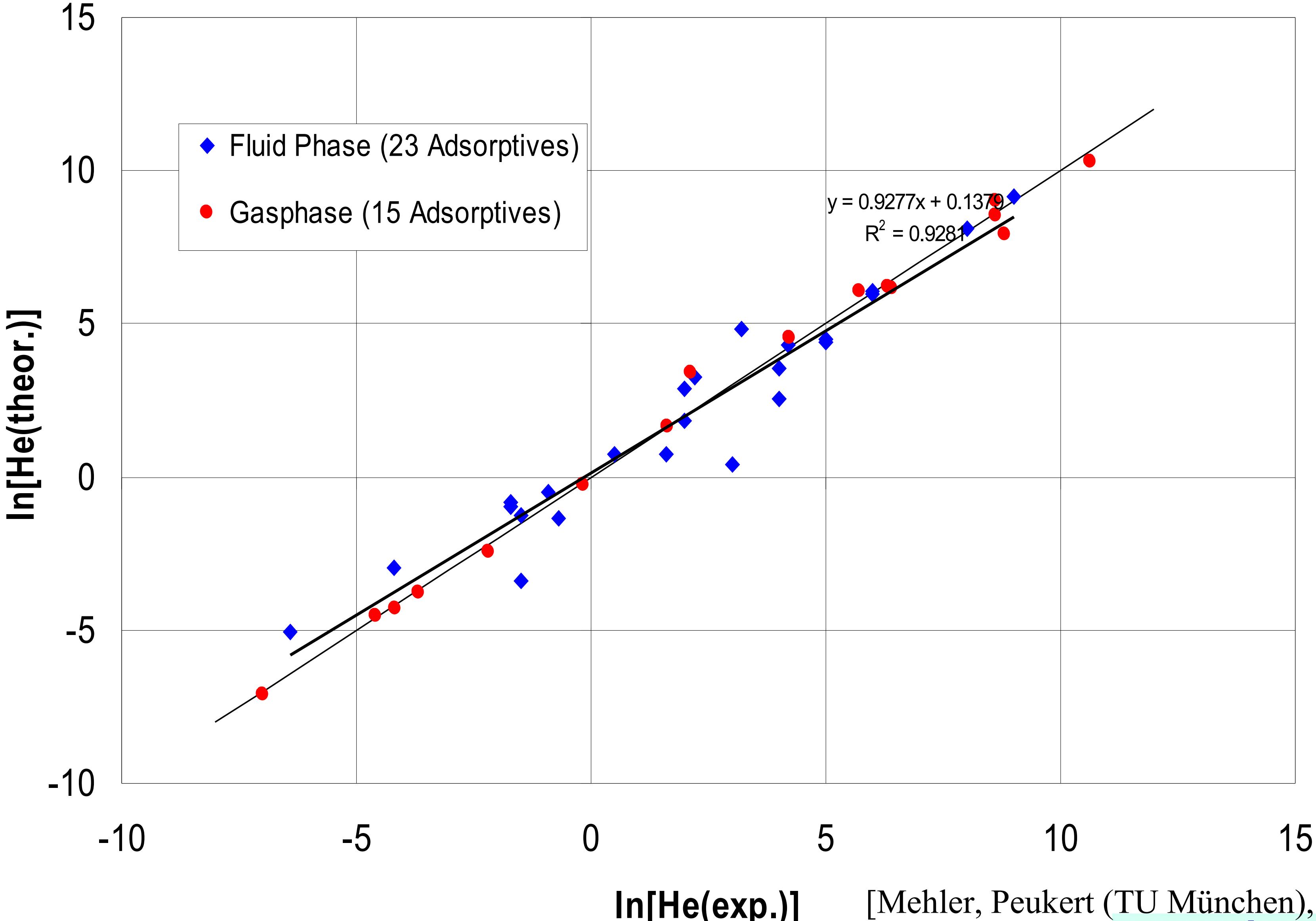


Prediction of Soil Sorption

Journal of Environmental Toxicology and Chemistry, in print

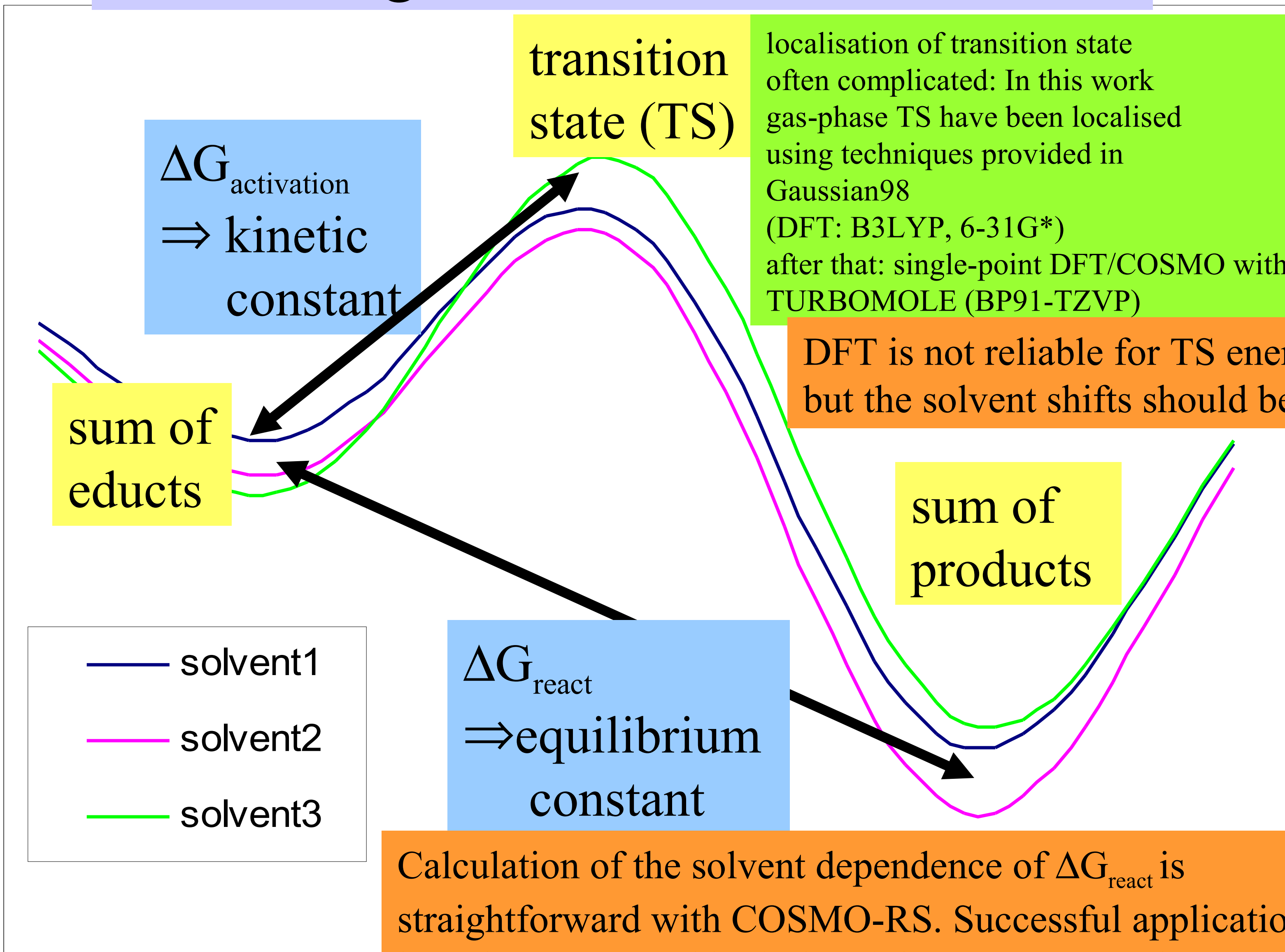


Adsorption to Activated Carbon



[Mehler, Peukert (TU München), Klamt; to be published]

Free Energies relevant for Reactions

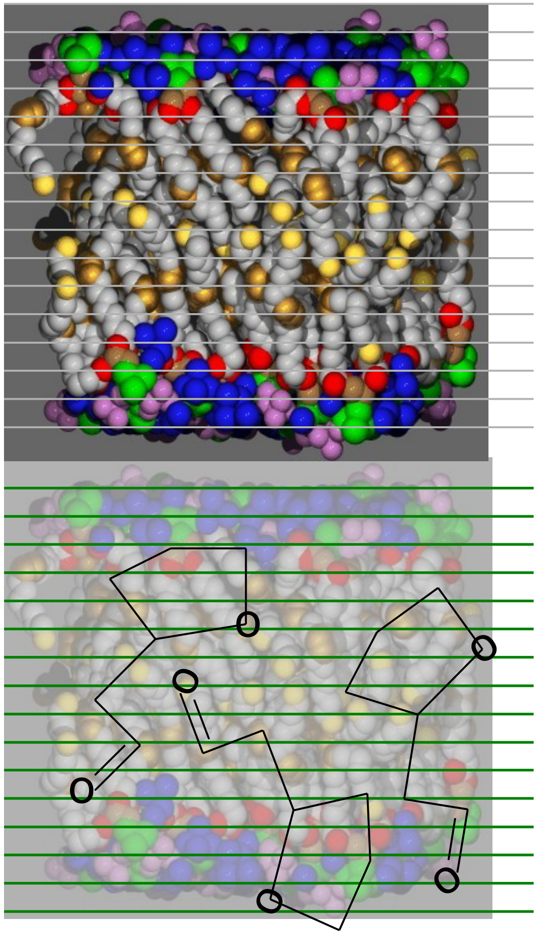


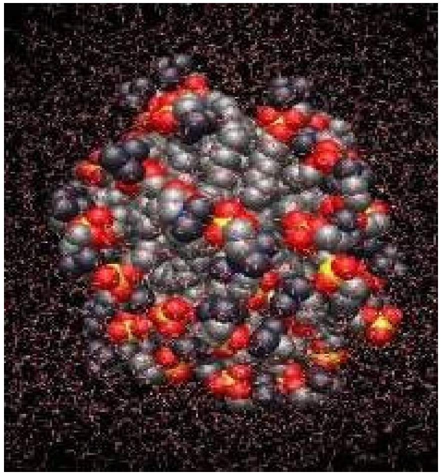
localisation of transition state often complicated: In this work gas-phase TS have been localised using techniques provided in Gaussian98 (DFT: B3LYP, 6-31G*) after that: single-point DFT/COSMO with TURBOMOLE (BP91-TZVP)

DFT is not reliable for TS energies but the solvent shifts should be reliable.

Calculation of the solvent dependence of ΔG_{react} is straightforward with COSMO-RS. Successful applications have been reported by industrial users (Dr. Franke, Degussa AG; Dr. Lohrenz, Bayer AG)

COSMO*mic*: Simulation of molecules in micelles and membranes





Concept:

-define layers of membrane (shells of micelle)

-get probability to find a certain atom of surfactant in each layer (e.g. from MD)

-convert this into a σ-profile p(σ,r) for each layer r using the COSMO-file of the surfactant

-use COSMOtherm to calculate μ(σ,r) considering each layer as a liquid mixture

-now calculate the chemical potential of a solute X in a certain postion and orientation by summing the chemical potentials of its segments in the respective layer.

-sample the chemical potentials all positions and orientations of X

-construct a total partition sum and get the probability to find the solute in a certain depth and orientation.

-also get the average volume expansion in each layer

- get a kind of micelle or membrane-water partition coefficient

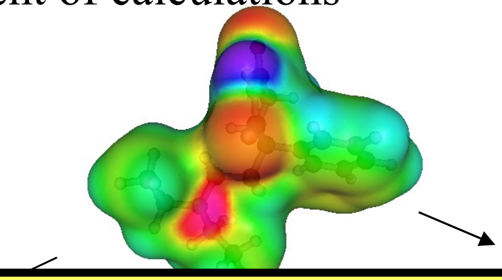
The tool COSMO*mic* facilitates all the previous steps together with COSMO*therm*

Perspective: self-consistent treatment of new surfactants; CMC prediction

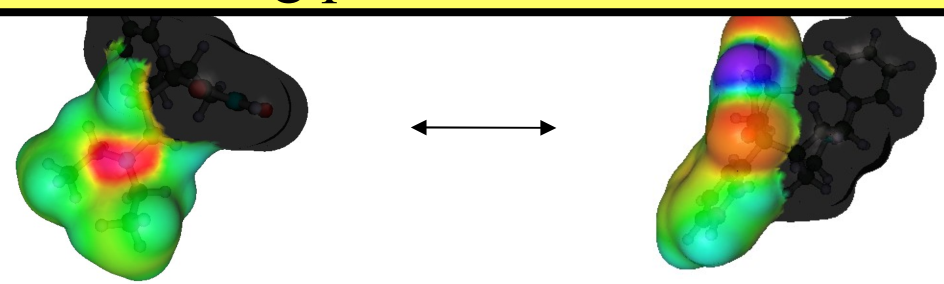
COSMO*frag*: A fast shortcut of COSMOtherm suited for HTS-ADME prediction

1) large database of precalculated drug-like compounds (about 45000)
2) for new compound find most similar fragments in database
3) compose COSMO surface from surface fragments (write a meta-file)
4) do usual COSMOtherm: solubility, partition properties

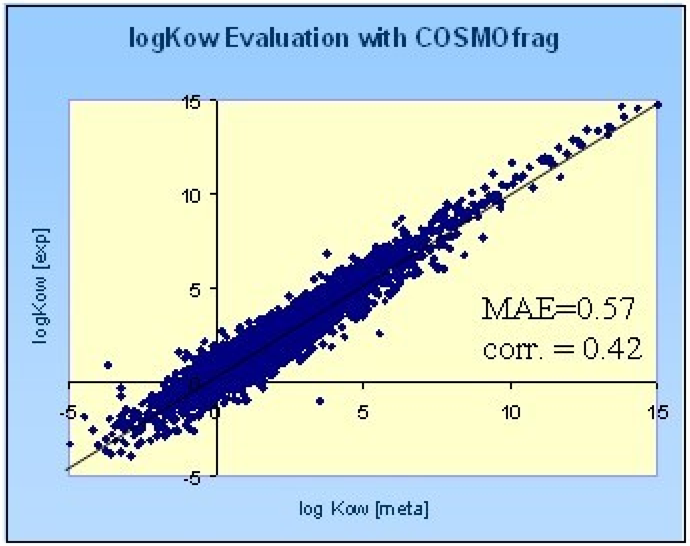
advantages:
-about 1 sec. per compound!
-you can add your typical inhouse structures to database
-simple refinement of calculations

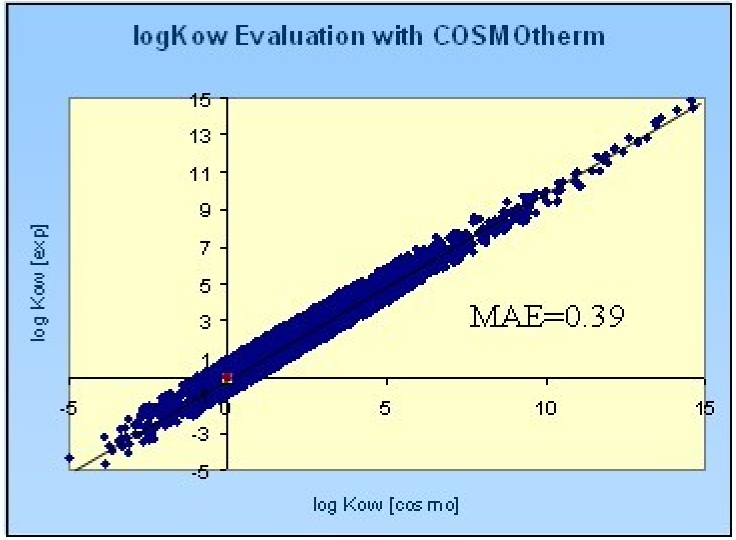


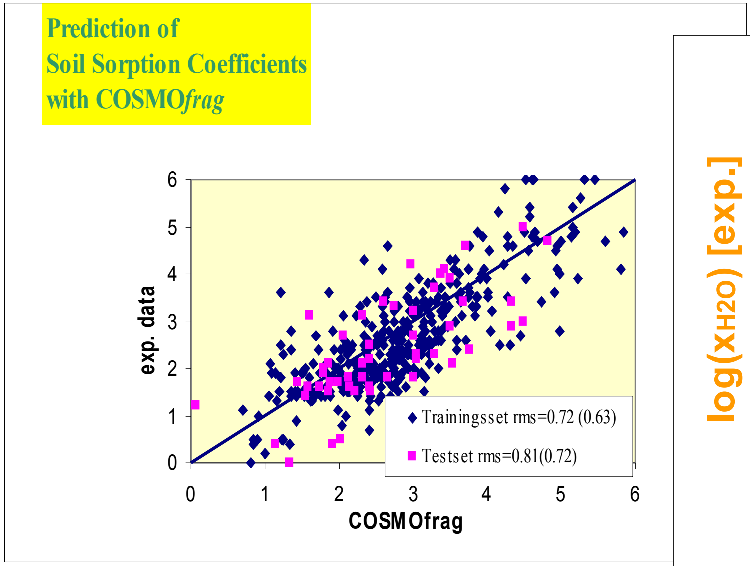
COSMOfrag ports COSMO-RS to Cheminformatics!

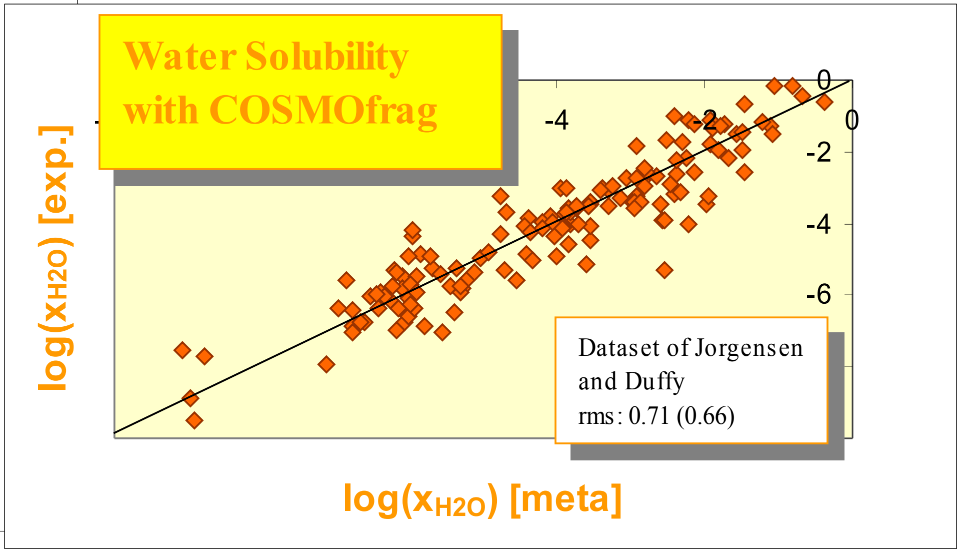


COSMO*frag:* statistics and examples



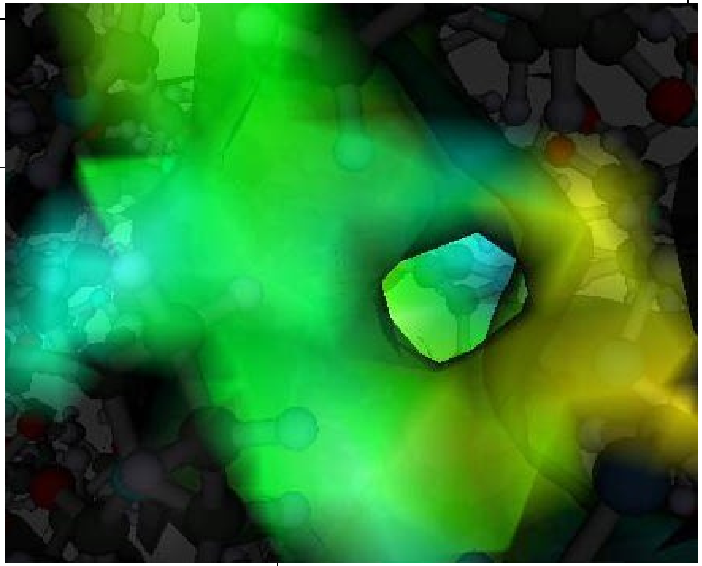


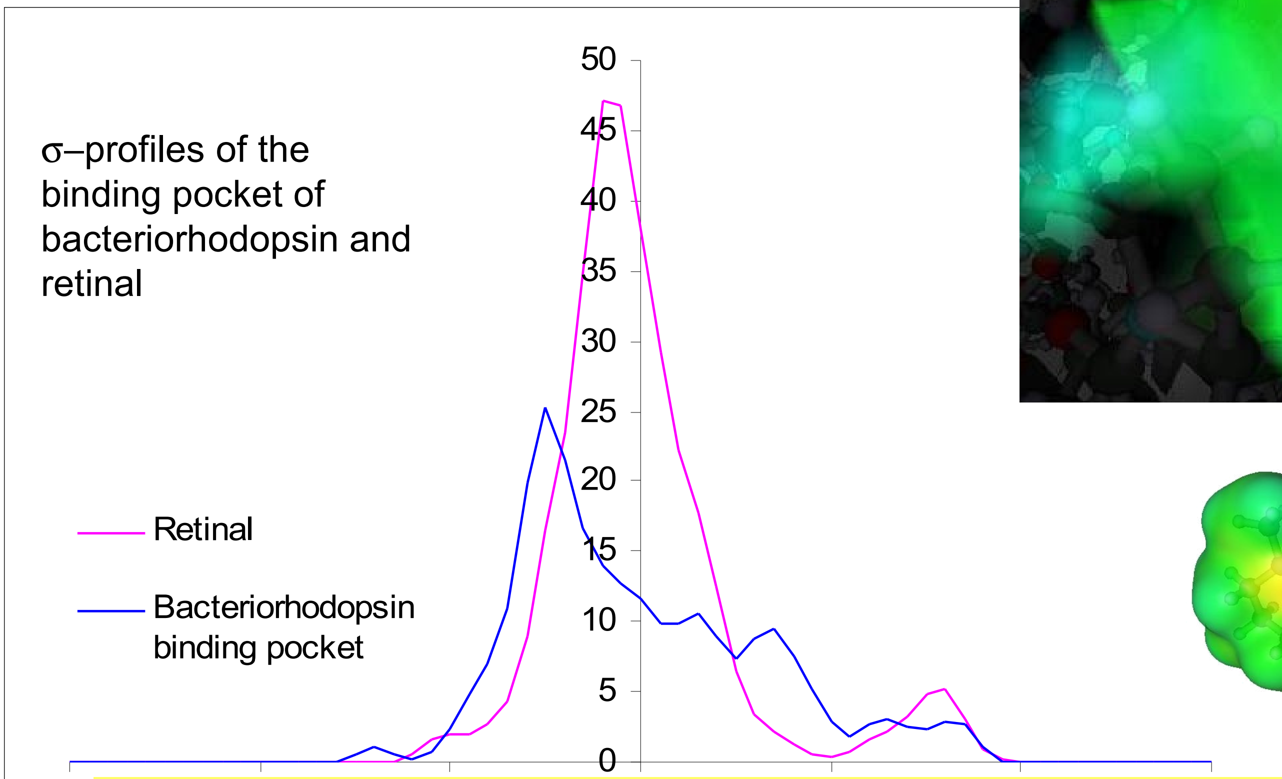


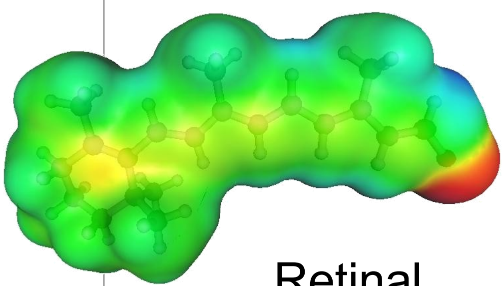


Ligand – Recptor Binding

Mouth of the Retinal binding pocket







Retinal

Meanwhile we can approximately treat enzymes and receptor pockets. The goal is to describe ligand receptor binding (incl. desolvation) based COSMO polarization cahrge densities σ.

COSMO*sim*

bio-isoster search based on s-profiles

examples by Dr. M. Thormann, Morphochem AG

If the physiological distribution and the drug-receptor binding are governed by the COSMO σ-profiles, it is reasonable to use these for drug-similarity searching:

- search for molecules with maximum similarity of σ-profiles
in order to find molecules with similar interactions, but different chemistry

-search is only based on surface polarity (σ) and not on structure

→ scaffold hopping

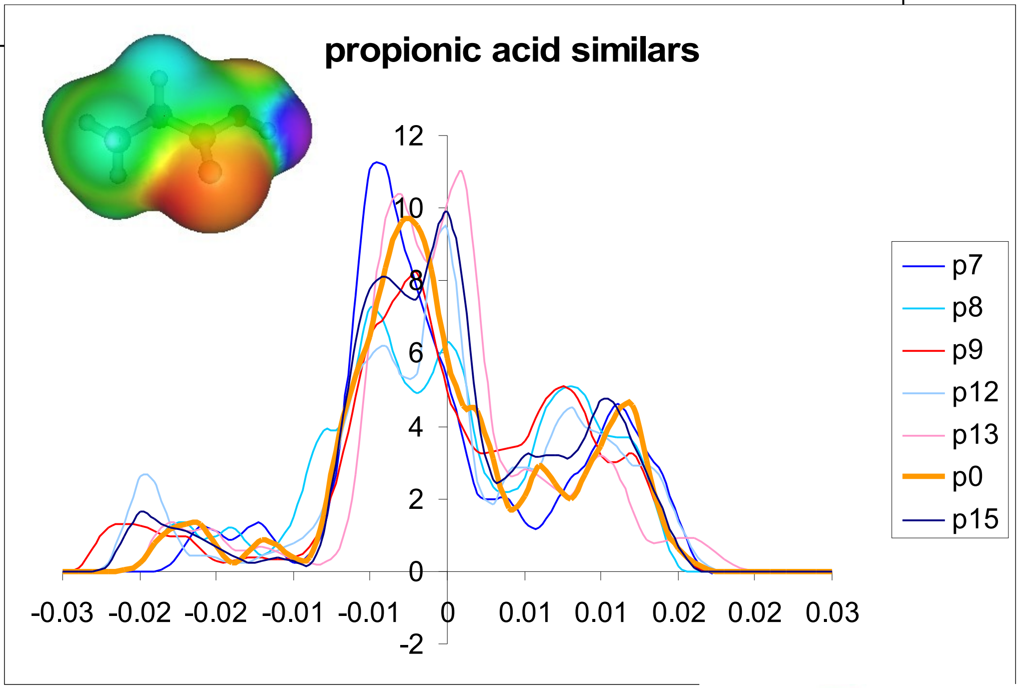
- either search over full COSMO-files of COSMOfrag-DB (48000 compounds)
-screen millions of candidate compounds using the COSMOfrag method
-Refine your search by explicit COSMO calculations on the most similar ~500 compds.

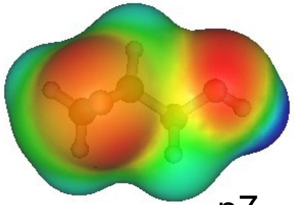
Lit: M. Thormann, A. Klamt, M. Hornig and M. Almstetter, "COSMOsim: Bioisosteric Similarity Based on COSMO-RS σ-Profiles", *J. Chem. Inf. Model.* **46**, (2006).

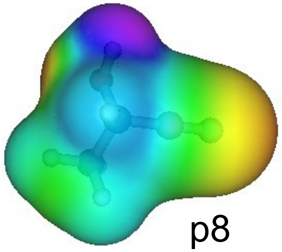
A.Bender, A. Klamt, K. Wichmann, M. Thormann, and R.C. Glen, „Molecular Similarity Searching Using COSMO Screening Charges (COSMO/3PP)“, in M.R. Berthold et al. (Eds.): CompLife 2005, LNBI 3695, pp. 175–185, 2005.Springer, Berlin Heidelberg 2005

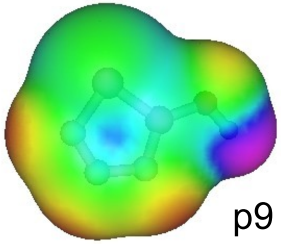
Example 1: propionic acid

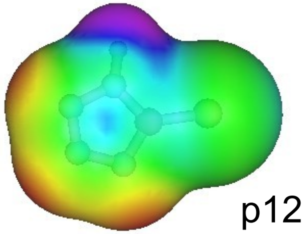
CCC(=O)O	ZFQCMUCKI	0	1
OC(=O)C=C	ITPZMBCLI	1	0.8169
CCCC(=O)O	IAVMXKDKI	2	0.7996
CC=CC(=O)O	RGQGEAHMI	3	0.791
CC(=C)C(=O)O	WCMTTAFLI	4	0.765
CC=CC(=O)O	VGZSDPDLI	5	0.7584
CC(C)C(=O)O	DGWQYNDKI	6	0.7487
OCC1CO1	SDLNNSMIA	7	0.7269
CC(O)C#N	HTYYARCJZ	8	0.7233
Oc1nnns1	NBAKLRQLI	9	0.7171
CC(O)C(=O)O	WOJBMNDKV	10	0.7109
CC(=O)O	CZWYICCKI	11	0.7052
Clc1nnn[nH]1	JMAKWZALI	12	0.7041
CC(=NO)C	EZHYEWAJI	13	0.6983
OCCC(=O)O	FFBMJKDKI	14	0.6978
CC(=O)C=NO	HOMSZUGLI	15	0.6919
Oc1csnn1	UMBRJEKLI	16	0.6885
OC(=O)C1CCC1	CUOCJIGKI	17	0.6817
OCCS	HLKLSJLHI	18	0.6804
CC1CC1C(=O)O	GXSEIQGKP	19	0.6767

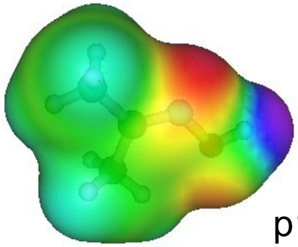


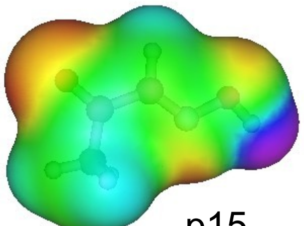

p7


p8


p9

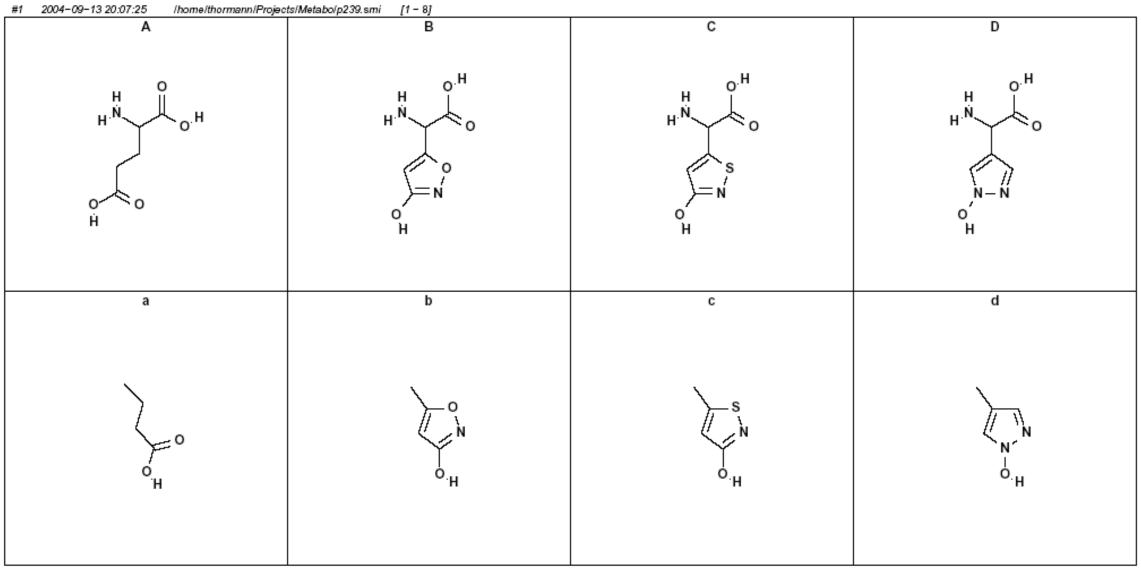

p12

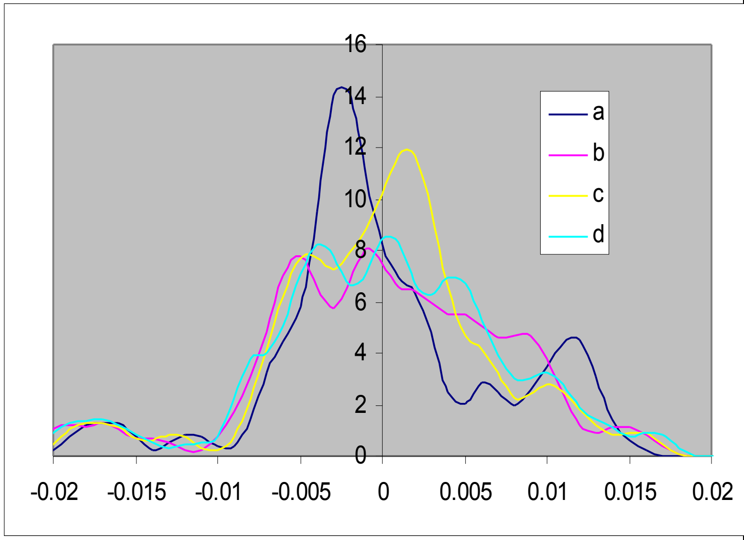

p13


p15

Example 2: Metabotropic Glutamate Receptor Ligands

Synthesis and Pharmacology of Metabotropic Glutamate Receptor Ligands
Grube-Jörgensen et al., ISMC 2004P239
Drugs of the Future 2004 (29) Suppl. A: XVIIIth Symposium on MEDICINAL CHEMISTRY





	A	B	C	D	a	b	c	d
A	1.000	0.711	0.666	0.697	0.396	0.440	0.459	0.488
B	0.711	1.000	0.852	0.835	0.406	0.487	0.459	0.530
C	0.666	0.852	1.000	0.857	0.378	0.461	0.455	0.507
D	**0.697**	**0.835**	**0.857**	**1.000**	**0.357**	**0.437**	**0.403**	**0.492**
a	0.396	0.406	0.378	0.357	1.000	0.665	0.679	0.642
b	0.440	0.487	0.461	0.437	0.665	1.000	0.742	0.792
c	0.459	0.459	0.455	0.403	0.679	0.742	1.000	0.700
d	**0.488**	**0.530**	**0.507**	**0.492**	**0.642**	**0.792**	**0.700**	**1.000**

Tanimotoprime coefficients for COSMOsim matrix
Glu (A), ibotenic acid (B), and thioibotenic acid (C) are known mGluR agonists.D is novel and does also show mGluR agonist activity with mGluR subtype specificity most similar to that of C. The querie of d to our inhouse database containing > 2.000.000 sigma profiles employing the Tanimotoprime coefficient retrieves b at rank 3 with a similarity of 0.792
(M. Thormann, Morphochem, 2004)

COSMO-RS: From Quantum Chemistry to Cheminformatics

• The quantum-chemically derived surface polarization charge densities σ provide a novel and very rich description of molecular interactions in liquids and pseudo-liquids phases, combing electrostatics, hydrogen bonding and “hydrophobic interactions“ in one picture.

• COSMO-RS provides a novel, extremely fast and efficient way to do thermodynamics based on σ-profiles.

• drug solubility and many important ADME properties can be calculated with COSMO-RS

• Quantum chemical DFT/COSMO calculations are reasonably feasible for a few hundred or thousand drug-like molecules.

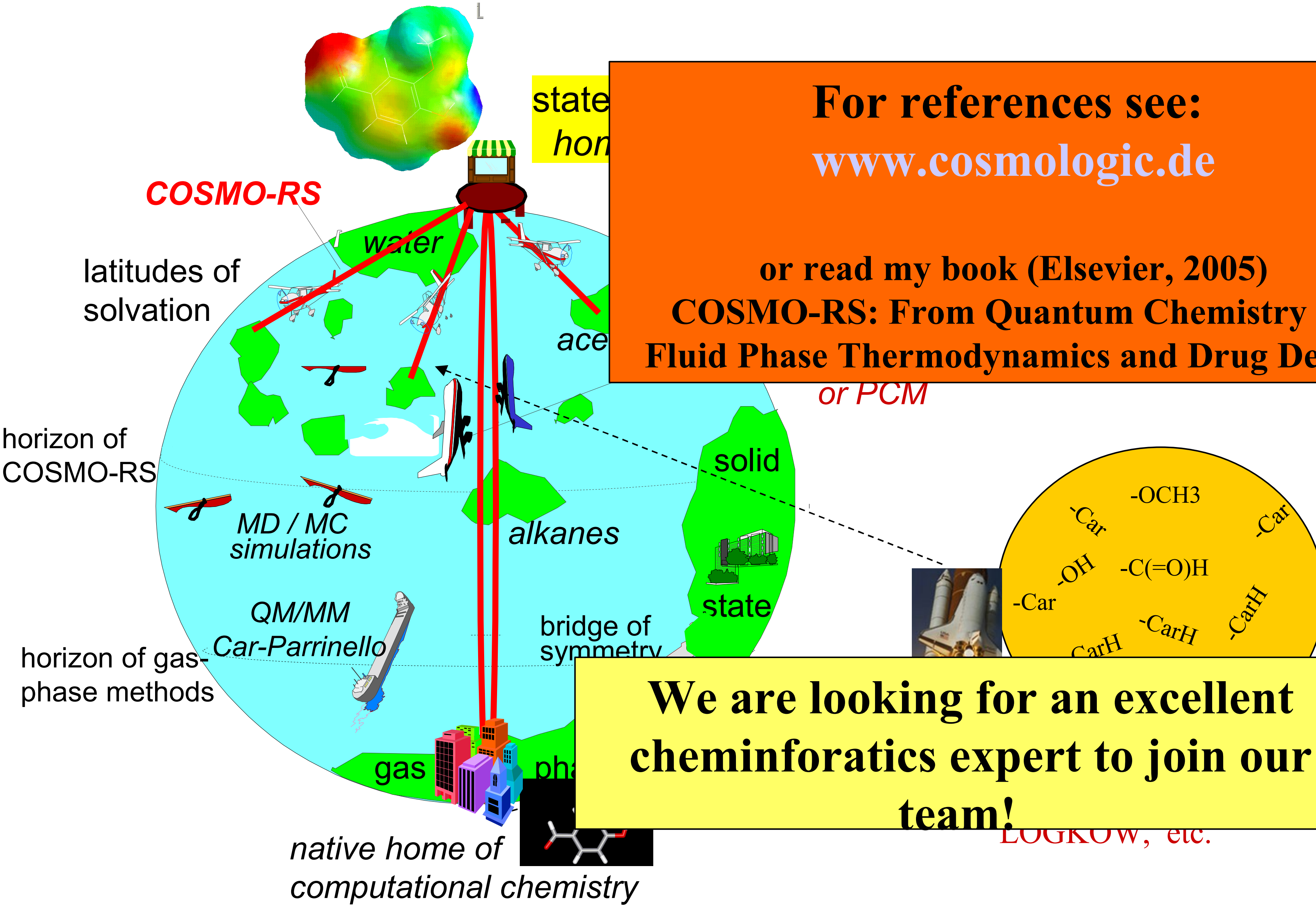
• COSMO*frag* derives approximate s-profiles for druglike compounds in a second.

• COSMO*sim* enables drug-similaity screening based on σ-profiles

----

Outlook: Ligand recepor binding based on σ-profiles

Hope you enjoyed the trip to the latitudes of solvation!



For references see:
www.cosmologic.de

or read my book (Elsevier, 2005)
COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design

We are looking for an excellent cheminforatics expert to join our team!

COSMO*logic* GmbH & Co. KG
Your Competent Partner for Computational Chemistry and Solvation

COSMO-RS for Drug-Design and -Development

• water solubility of drugs,

• Solvent Screening: relative solubilities of drugs in various solvents and mixtures

• partition behaviour between almost arbitrary phases (blood-brain, intestinal absorption, BCF, ...)

•. pK_a prediction

• visualization of partition coefficients and solubility as surface properties

• one descriptor (σ) for entire interactions
  - electrostatics
  - hydrogen bonding
  - lipophilicity/hydrophobicity

=> useful property for MFA

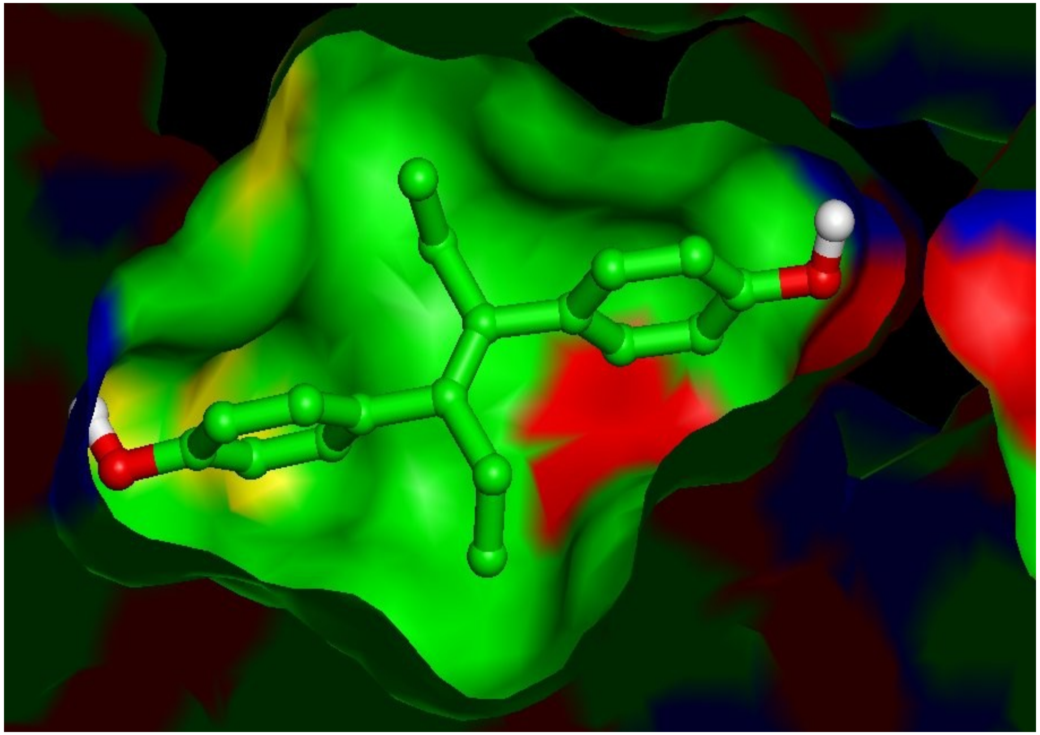
- chemical potential of crystal surfaces in solution (morphology of drugs)

•.identification of binding sites from σ-hotspots

• surface integral scoring function for docking, including desolvation

- extension to membrane and micelle partitioning

COSMO*frag:* σ-profiles built from similar fragments out of 30000 compound database brings COSMO-RS in to the range of 5 sec./compound => applicable to HTS



Ideas for drug drug-receptor binding with COSMOtherm

-we need the σ-profile
of the receptor once
(QM/MM? not yet solved)

- we simply have the σ-profile
of the ligands
(even from COSMOfrag)

Idea 1: generate scoring function
from COSMO-RS surface
interaction model

Idea 2: consider receptor pocket
as a kind of pseudo-liquid
(overestimated receptor
flexibility,
but may be interesting)

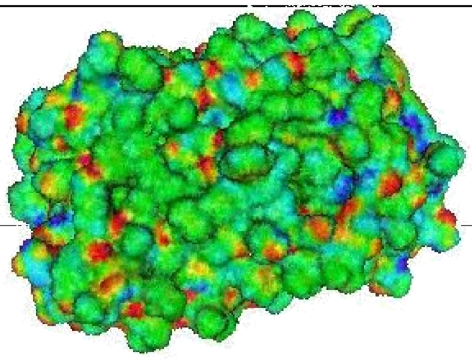
Both simply include desolvation

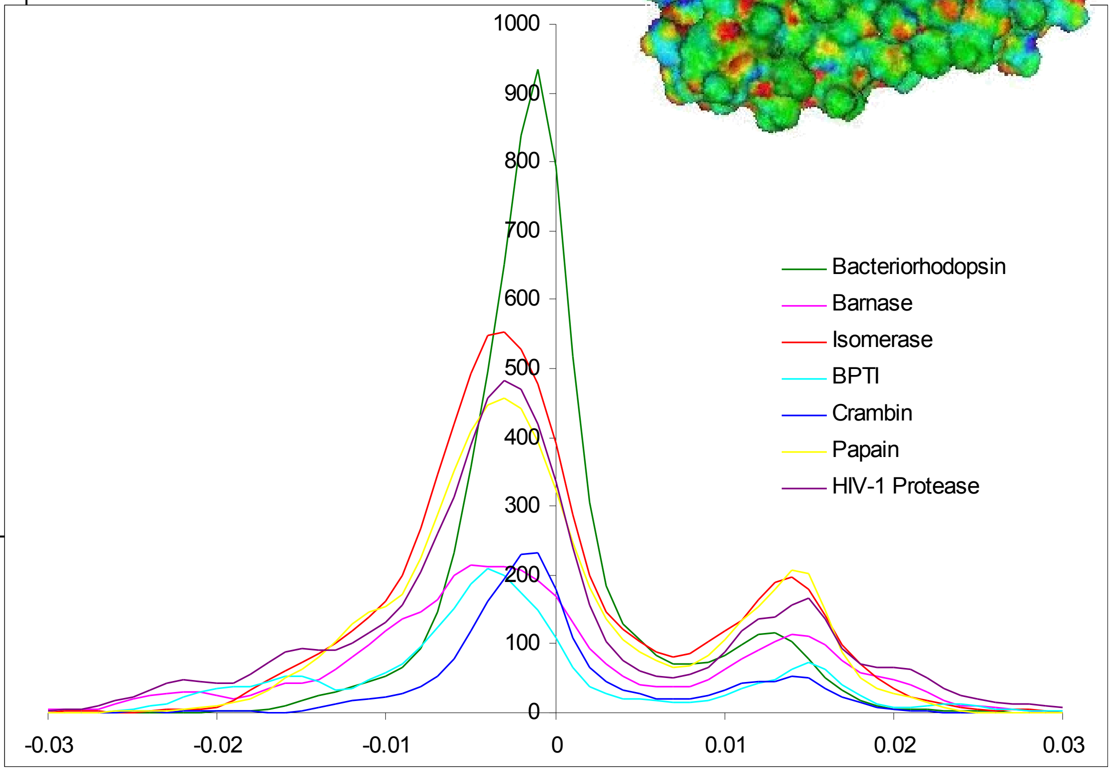
Sigma profiles of Enzymes

calculated with linear-scaling AM1/COSMO
(MOZYME in MOPAC2002)

Some common features:

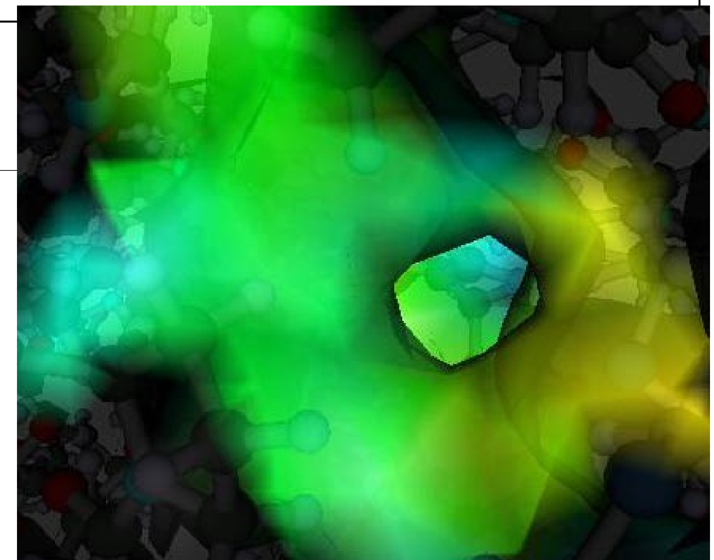
- Large charge distribution in the region around $\sigma = 0$.
- Carbonyl oxygen between 0.01 and 0.02.
- Charged side chains in the outer regions ($\sigma < -0.02$ and $\sigma > 0.02$)

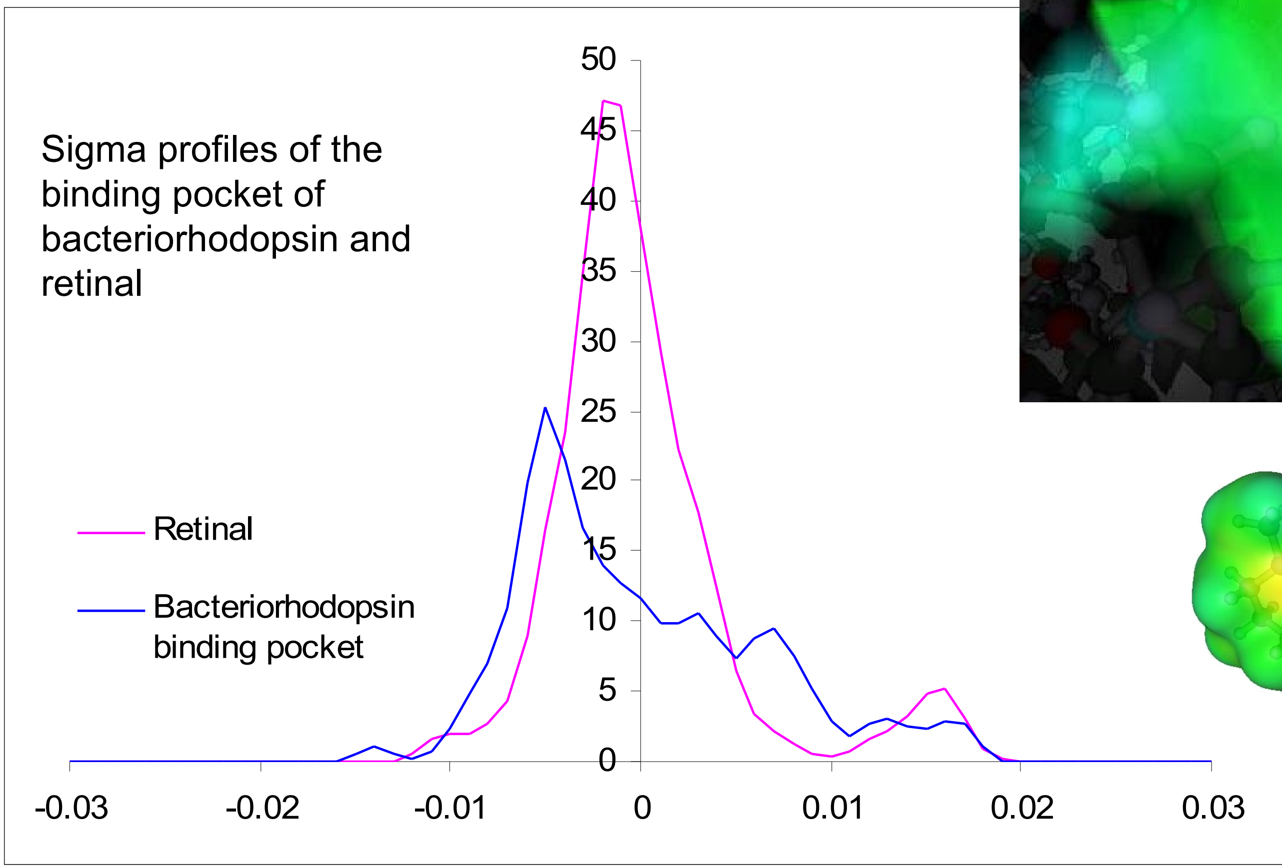


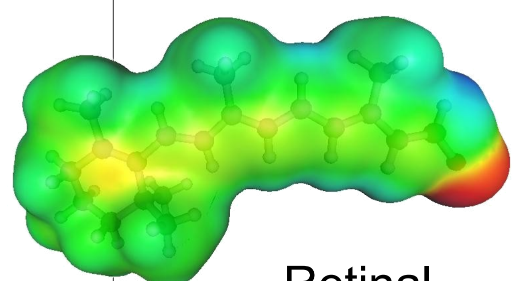


Bacteriorhodopsin and Retinal

Mouth of the Retinal binding pocket

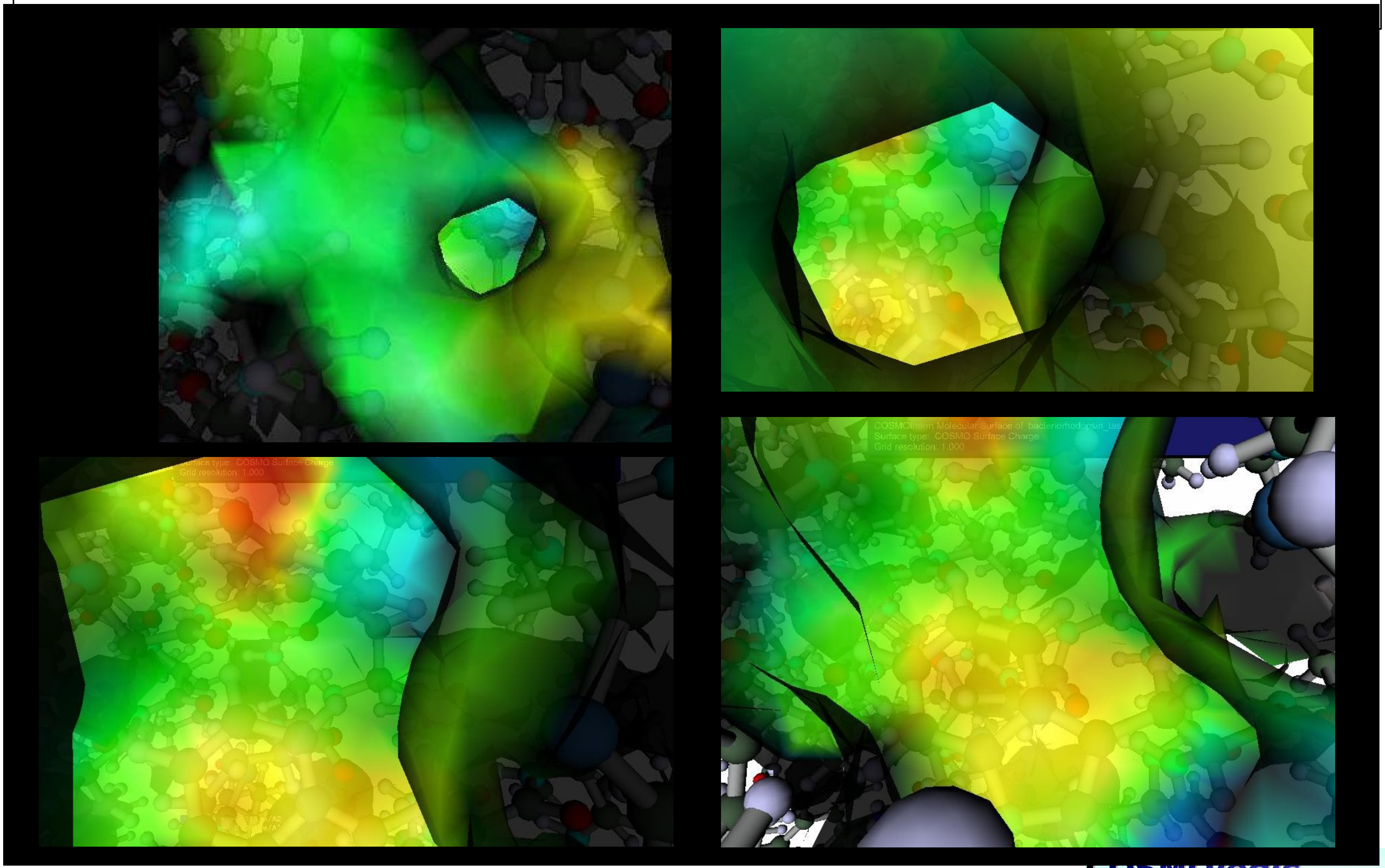






Retinal

A few shots of the binding pocket



Amino Acids: Sigma profiles on two computational levels