Conformational Analysis

3D Structures, Conformations and Molecular Surfaces

Christof H. Schwab

Molecular Networks GmbH
Henkestr. 91
91052 Erlangen, Germany

mn-am.com

Altamira LLC
1455 Candlewood Drive
Columbus, Ohio 43235, USA

Molecular Networks and Altamira MN-AM

Erlangen, Germany
Friedrich-Alexander-Universität
1997

Columbus, Ohio, USA
The Ohio State University
2008

- Chemoinformatics
  - *3D structure generation*
  - *Physicochemical and reaction properties*
  - *Metabolic reaction knowledge*
- Computational toxicology and risk assessment
  - *Database and knowledgebase*
  - *Predictive models*
  - *Consulting services*

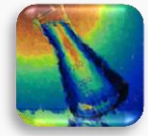
MN AM

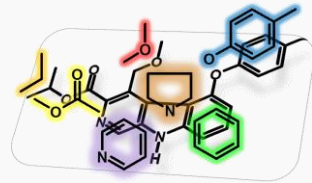
Product Lines – Chemoinformatics

- CORINA Classic
  - *Industry-standard 3D structure generation*
- CORINA Symphony
  - *Profiling and managing of chemical datasets*
  - *Workflows*
    - Structure cleaning/processing
    - Descriptor generation (properties and fragments)
- SYLVIA
  - *Estimation of synthetic accessibility of compounds*
- Public tools
  - *ChemoTyper (with ToxPrint Chemotypes)*
  - *https://chemotyper.org, https://toxprint.org*

CORINA Classic

CORINA Symphony

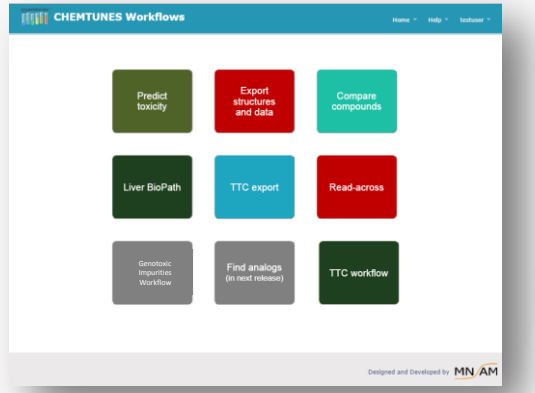




MN-AM

Product Lines – Computational Toxicology and Risk Assessment

- ChemTunes
  - *Platform to support decision making in human health and regulatory critical endpoints*
    - Toxicity database (all endpoints)
    - "Inventory" concept for compound location
- ChemTunes ToxGPS
  - *Prediction models*
    - All human health related endpoints
  - *Workflows*
    - TTC (thresholds of toxicological concern)
    - Read-Across (in development)
    - ICH M7 GTI (in development)



3D Structures, Conformations and Molecular Surfaces – Overview

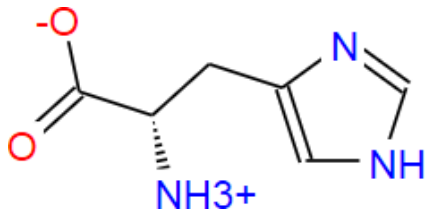
- 3D structures – why, where needed and how many
- 3D structure generation
  - *Single 3D structures and conformational ensembles*
  - *Methods, approaches, applications*
- Storage of 3D structures
  - *Examples of file formats*
- Molecular surfaces
  - *Approaches, application*

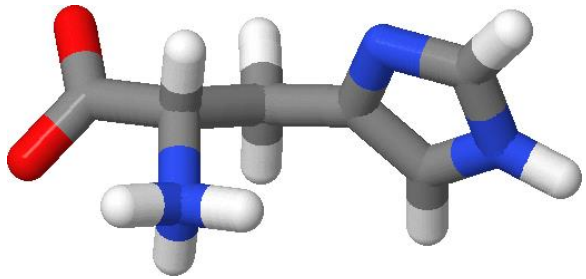
- References

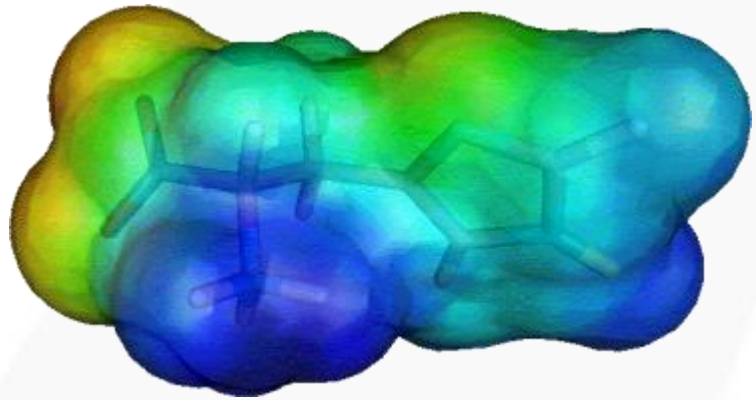
MN/AM

3D Structures – For What?

- Chemists' language
  - *2D structure diagrams*
  - *Stereochemistry*
- Molecules are 3-dimensional objects
  - *Spatial interactions of molecules cause effects*







MN AM

3D Structures – Where needed?

- 3D database generation
- Lead discovery and lead optimization
  - *Ligand- and structure-based virtual screening*
    - Docking studies, pharmacophore and similarity searching
- Prediction of chemical, physicochemical and biological properties
  - *ADMET properties*
  - *QSAR and QSPR studies*
- Structure elucidation
- Prediction of chemical reactivity
- Input to quantum-mechanical and force field calculations

MN AM

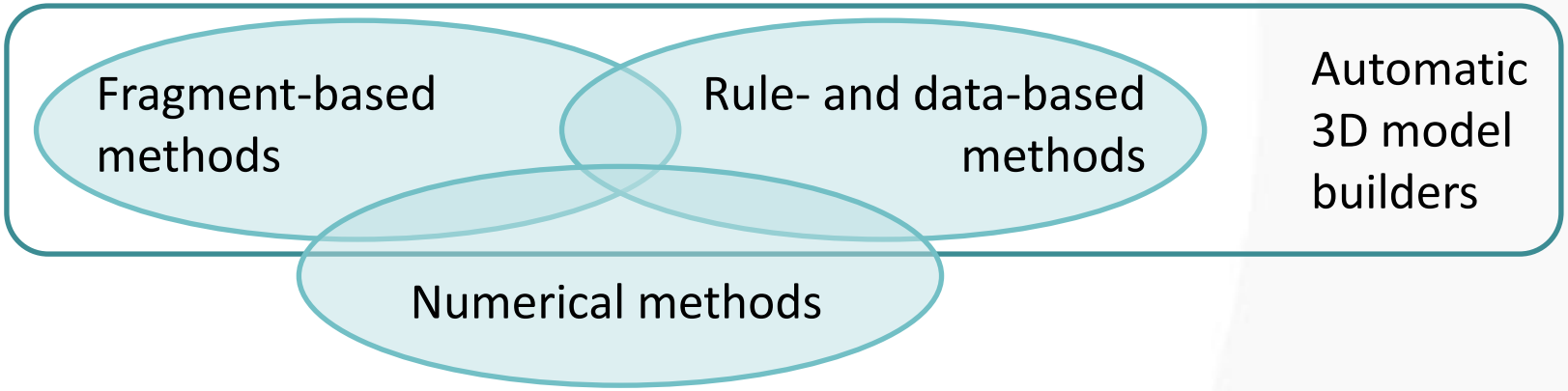
Chemical Structures – How many?

- The Cambridge Crystallographic Database (CSD)
  - *Over 850,000 experimentally-determined crystal structures*
- PubChem
  - *82 Million characterized chemical compounds*
- Chemical Abstract Service, CAS Registry
  - *121 Million unique organic and inorganic chemical substances*
- GDB, University of Berne
  - *GDB-13, 1 Billion compounds (13 atoms, C, O, N, S, and Cl)*
  - *GDB-17, 164 Billion compounds (17 atoms, C, O, N, S, and halogens)*
- RCSB Protein Data Bank, PDB
  - *123,000 biological macromolecular structures*
  - *21,000 ligand structures*

MN AM

Automatic 3D Structure Generators

- Categories of 3D molecular model building approaches



- No sharp borders between approaches
  - *Fragment-based methods require rules*
  - *Rule- and data-based methods use 3D templates*
  - *Numerical methods require starting geometries*

MN AM

3D Structure Generators

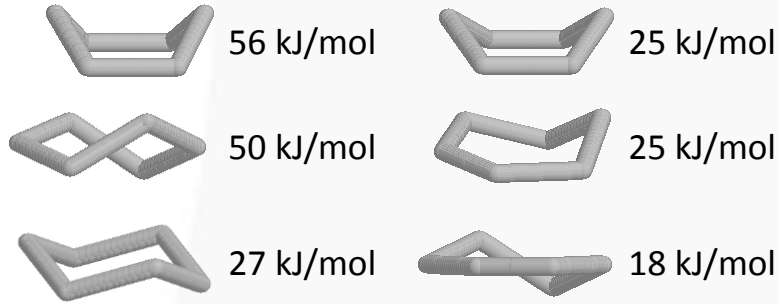
- Fragment-based methods
  - *Fragment, select appropriate 3D template, link*
  - *Minimum set of rules for fragmentation, analog searching and linking*
- Rule- and data-based methods
  - *Knowledge base from theoretical investigations and experimental structures*
  - *Explicit rules and implicit data*
- Numerical methods
  - *Molecular force field calculations*
  - *Quantum-mechanical methods*
    - Semi-empirical, DFT and *ab initio*
  - *Distance geometry*

MN AM

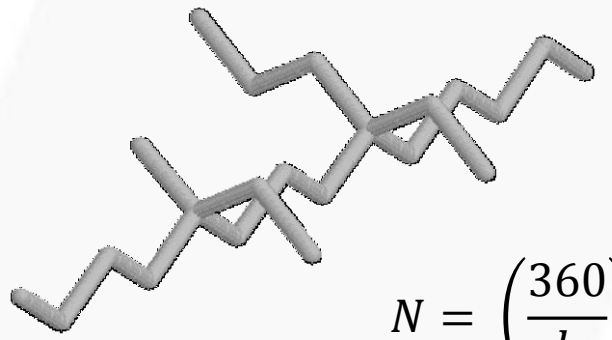
3D Structure Generators – General Issues (1/3)

- Different conformational behavior of cyclic and acyclic portions

- Ring systems with limited degrees of freedom
  - *Use of "allowed" low-energy ring templates*



- Open-chain structures with increasing number of degrees of freedom with increasing number of rotatable bonds
  - *Selection of low-energy conformation*
  - *Principle of longest pathways*

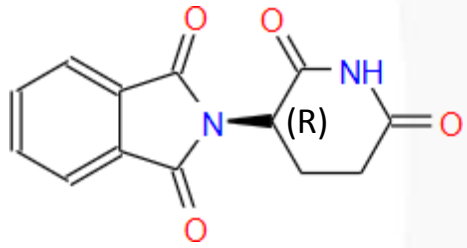


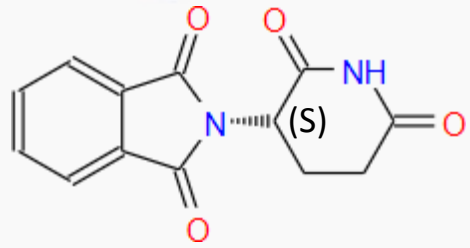
$$N = \left(\frac{360}{k}\right)^{n}$$

MN AM

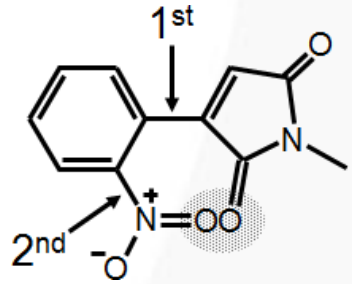
3D Structure Generators – General Issues (2/3)

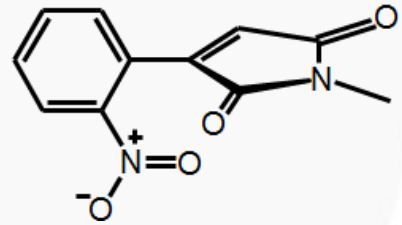
- Stereochemistry awareness
  - *Tetrahedral chiral centers and cis/trans double bonds*
  - *Coded in input structure*





- Close contacts and steric crowding
  - *Mechanisms to eliminate such situations*





- Conformational analysis
  - *Several conformations to identify a low-energy one*

MN AM

3D Structure Generators – General Issues (3/3)

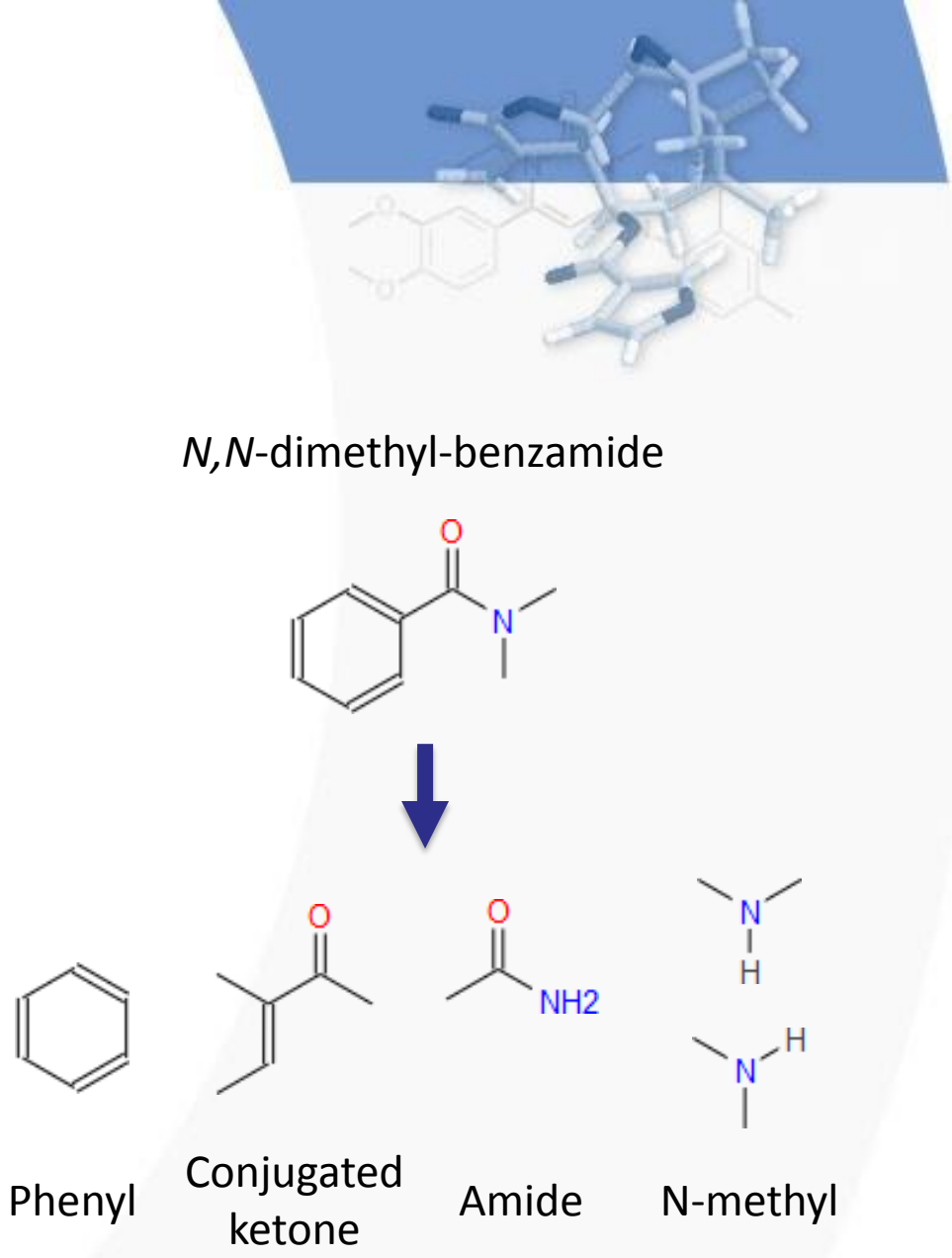
- Processing of large amounts of data
  - *Millions of structures in (company) databases*
  - *Robustness*
  - *CPU times*
  - *High conversion rates*
  - *Handling of broad range of chemical spaces*

- Quality of 3D models
  - *Close contacts and steric crowding*
  - *Reproduction of experimentally determined geometries*

Fragment-based Methods

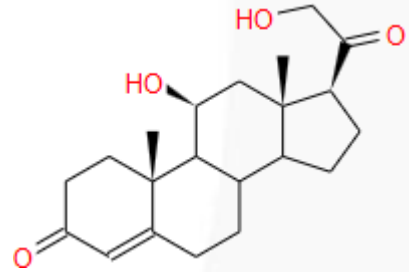
- Fragment the structure
  - *Rules for fragmentation*
- Select appropriate 3D template
  - *Library of 3D templates*
- Link 3D templates
  - *Rules for linking*

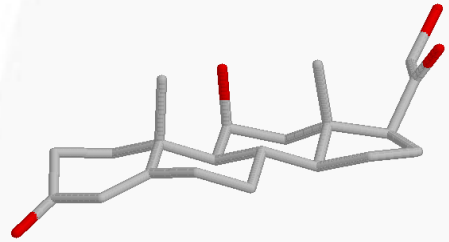
- Rules and data required

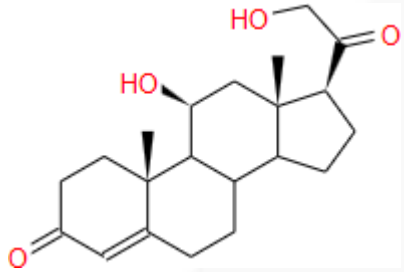


Rule and Data-Based Methods

- Knowledge base from theoretical investigations and experimental structures
- Explicit and implicit rules and data
  - *Standard bond lengths and angles*
  - *Preferred torsion angles*
  - *Ring templates*
  - *Empirical (fast) energy estimation*
    - Increments for ring templates



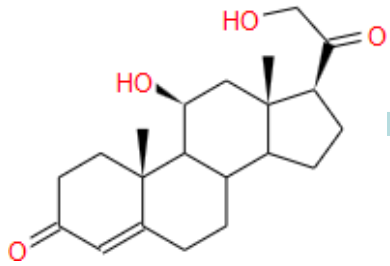




CORINA Classic

MN-AM

CORINA Classic – General Principles



Connection table
Stereo descriptors

Internal coordinates stereo information

Tables

Stereo information

Fragmentation

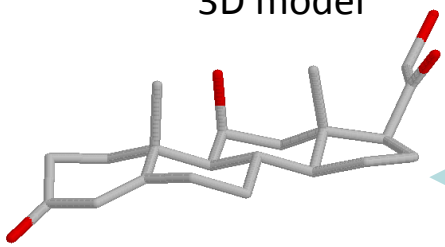
Small to medium-sized rings

Acyclic portions

Ring systems

Large/macrocyclic rings

3D model



Re-assembling conformational analysis (close contacts)

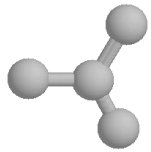
Geometry optimization

MN AM

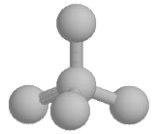
CORINA Classic

- Atom types and bond angles

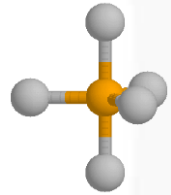
linear: 180°



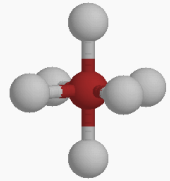
planar: 120°



tetrahedral
109.47°

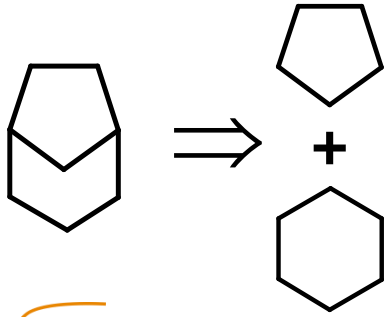


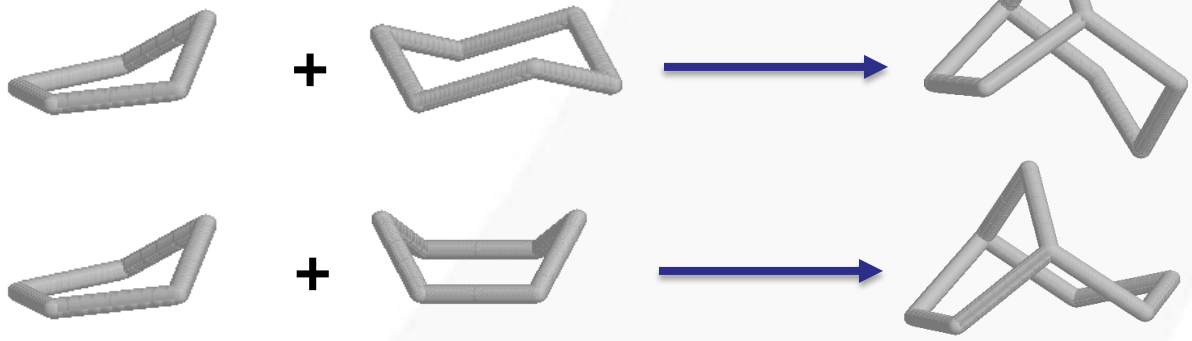
trigonal bipyramidal
90°, 120°, 180°



octahedral
90°, 180°

- Ring templates
  - *Smallest set of smallest rings*

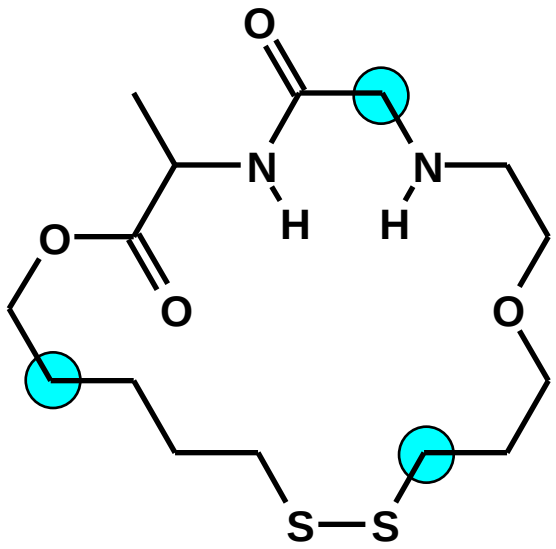




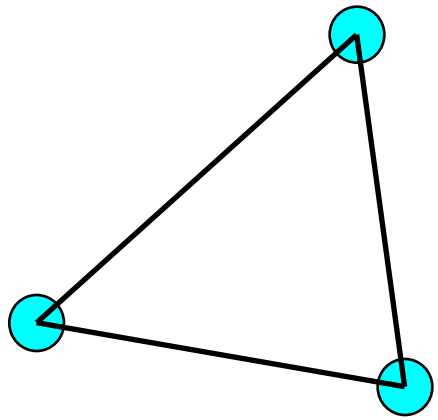
MN AM

CORINA Classic

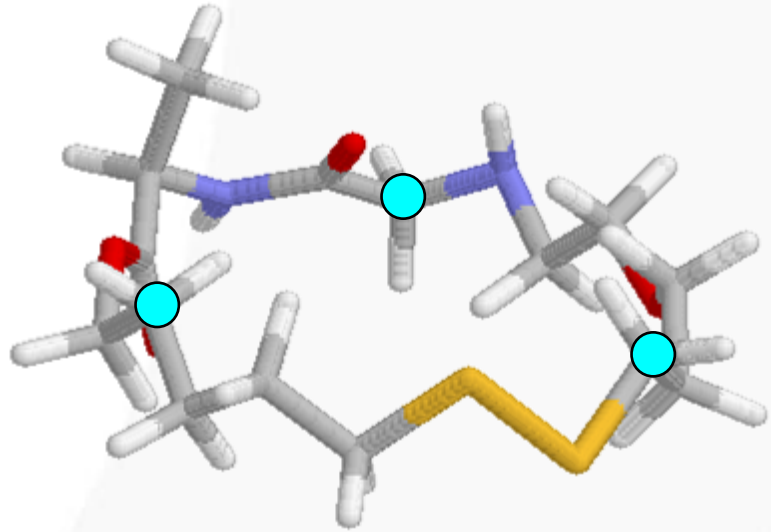
- Macrocyclic structures
  - *Principle of superstructure*



2D structure



Superstructure



3D model

CORINA Classic – Application

- Conversion of National Cancer Institute (NCI) database
  - *Number of structures* — *265,242*
  - *Structures converted* — *263,184*
  - *Structure coding errors in DB* — *543*
  - *Conversion rate* — *99.43%*
  - *CPU time* — *0.5 h*
    (x86 Linux, 1.5 GHz) — *(1,843 s, 0.007 s/molecule)*

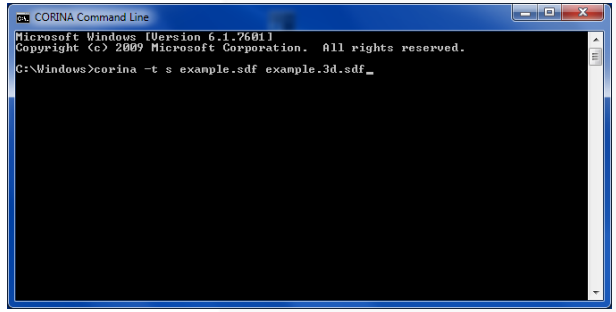
MN/AM

CORINA Classic – Key Features

- Support of a variety of chemical file formats
  - *SDF, RDF, SMILES, PDB, SYBYL MOL and MOL2, MacroModel, Maestro, CIF,...*
- Generation of multiple ring conformations
  - *Data-based for rings consisting of up to nine atoms*
  - *Interface to docking program FlexX*
- Generation of stereo isomers
  - *Automatic detection of stereo centers (tetrahedral and cis/trans)*
  - *Full/partial enumeration, preserve defined centers, duplicate detection*
- Structure "clean-up" features
  - *Adding H atoms, neutralizing formal charges, atom and bond type assignment, stereo information, removal of counter ions in salts, ...*

CORINA Classic Interfaces

- Command line interface and library version
  - *Linux shared object and Windows DLL*



- Pipeline Pilot component

- KNIME node

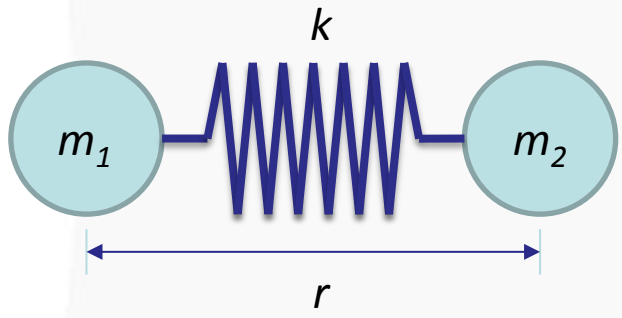
MN-AM

Numerical Methods

- Molecular force field calculations
- Quantum-mechanical methods
  - *Semi-empirical*
  - *DFT*
  - *ab initio*
- (Distance geometry)

Molecular Force Field Calculations

- Classical mechanical treatment of molecules
  - *Composed of masses: atoms*
  - *Connected by springs: bonds*
  - *Potential energy by Hooke's law*



$$V(r) = \frac{k}{2}(r - r_0)$$

- Born-Oppenheimer approximation
  - *Separation of movement of electrons from much slower movement of nuclei*
  - *Potential energy of molecule as function of atomic coordinates*

MN AM

Molecular Force Field Calculations

- Two major parts
  - *Functional form (mathematics)*
  - *Parametrization and atom typing*

- Not a single, "true" mathematical expression
  - *Different functional forms can be applied*
- Different sets of parameters and atom types
  - *Derived from experimental results*
  - *Represent thermodynamic average (rather than a particular geometry)*

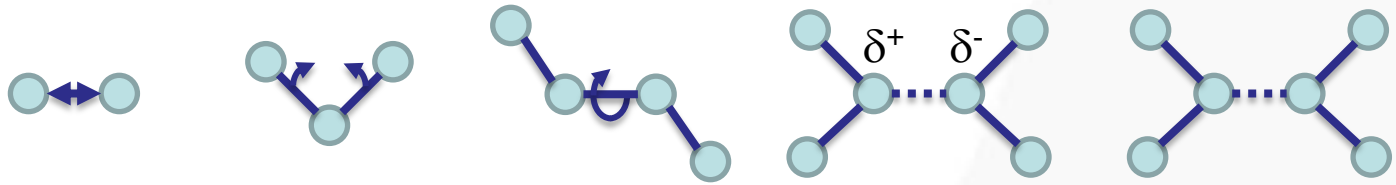
⟹ Empirical method

MN AM

Molecular Force Field Calculations

- Functional form
  - *Calculation of potential energy*
  - *Optimization of energy depending on atom coordinates*
  - *Bonded and non-bonded contributions*

$$V = \sum V_{bonds} + \sum V_{angles} + \sum V_{torsions} + \sum V_{electrostatic} + \sum V_{vanderWaals}$$



Bond stretch | Angle bend | Torsion | Electrostatic (non-bonded) | van der Waals (non-bonded)

MN AM

Molecular Force Field Calculations

- Atom typing
  - *Differentiation between hybridization state, local environment, special conditions (e.g., strained ring)*
  - *Depending on application*

- Parametrization
  - *The more atom types, the more parameters for contributions/terms in energy function required*

$$V = \sum V_{bonds} + \sum V_{angles} + \sum V_{torsions} + \sum V_{electrostatic} + \sum V_{vanderWaals}$$

MN AM

Force Fields – Applications

- Conformational analysis
  - *Studying different conformations of a molecule*
  - *Structure refinement using experimental data*
- Molecular properties
  - *Heats of formation*
- Used in molecular dynamics simulations

MN AM

Some Force Fields and Programs (1/2)

- MM2/MM3/MM4 (Allinger, University of Georgia)
  - *Small molecules, peptides, proteins*
- UFF (Rappé, Goddard III, Colorado State University)
  - *Mainly small molecules, generic force field for "whole periodic table"*
- AMBER (Kollman, University of California San Francisco)
  - *Biomolecules including solvation effects (water models TIP3/4P, SPC, POL3,…)*
- CHARMM Karplus, Harvard University)
  - *Macromolecular simulations, molecular dynamics*
- GROMOS (van Gunsteren, Berendsen, University of Groningen)
  - *Molecular dynamics, aqueous or apolar solutions, lipid systems*

Some Force Fields and Programs (2/2)

- MMFF (Halgren, Merck and Co, Inc)
  - *Optimized to handle functional groups relevant in pharma*
- Vendors
  - *BIOVIA (Accelrys)*
  - *Chemical Computing Group*
  - *Schrödinger*
- Free/open source
  - *Open Babel*
  - *RDKit*

MN AM

Quantum Mechanical Approaches

- Solving the molecular Schrödinger equation
  - $\hat{H}\Psi = E\Psi$
  - *Energy as a function of the electronic wave function*

- Three main categories according to level of theory
  - *Semi-empirical molecular orbital (MO) theory*
  - *Density functional theory*
  - *Ab initio MO theory*

- Computationally expensive
  - *Semi-empirical < DFT < ab initio*

MN AM

Quantum Mechanical Approaches

- General principles
  - *Electronic structure of molecule (built by atoms) as basis*
  - *Linear combination of atomic orbitals (LCAO) approximation to derive molecular orbital*
  - *Hartree-Fock (HF) and self-consistent field (SCF) approximation*
    - Each electron interacts with mean field of all other electrons

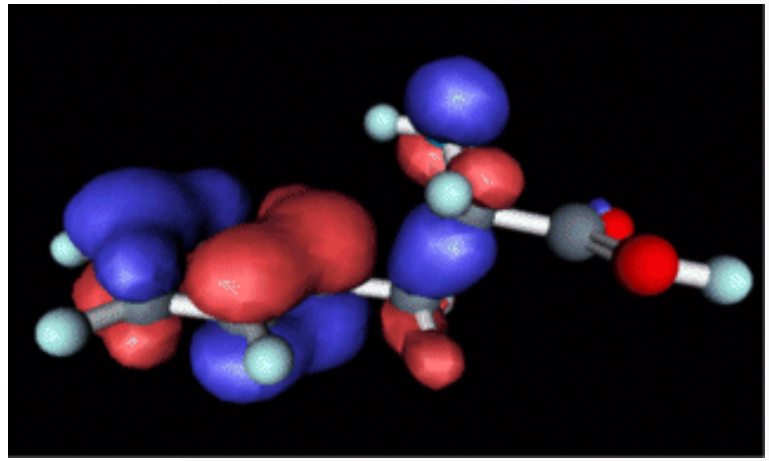
- Semi-empirical methods
  - *Parametrization of (some) terms using data from experimental data*
  - *Neglecting/approximation of (some) interaction terms*

Output of Quantum Mechanical Calculations

- Net atomic charges
- Dipole and higher multipole moments
- Polarizabilities
- Orbital energies
  - *HOMO/LUMO*
- Heat of formation

Valuable for modelling of chemical reactivity and toxicity endpoints

- Surface descriptors
- Local ionization potential
- Atomic coordinates



Molecule orbitals of phenylalanine (HOMO/LUMO)

ChemTunes

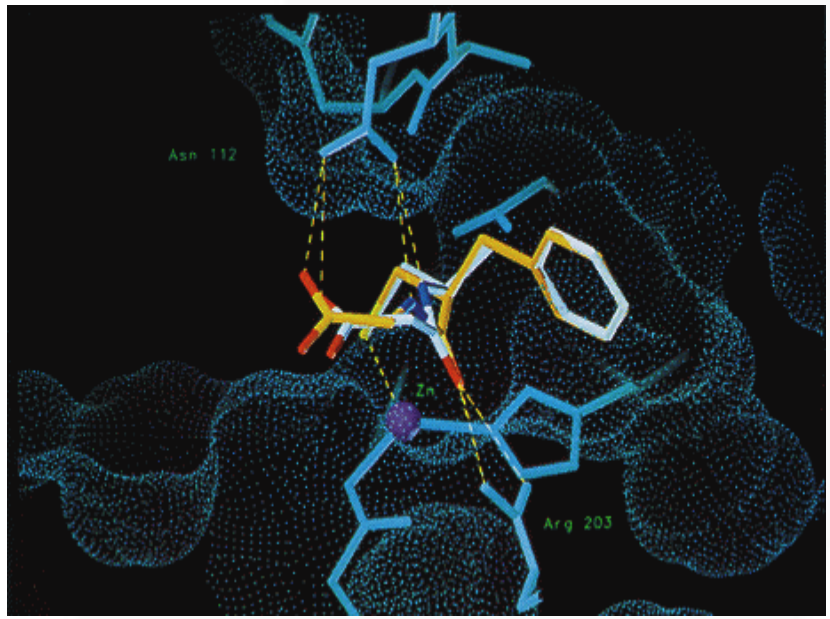
MN AM

Quantum Mechanical Program Packages

- GAUSSIAN
  - *Molecular mechanics, semi-empirical methods (AM1, PM3, …), ab initio*
- MOPAC
  - *Semi-empirical methods, AM1, PM3, …*
- Spartan
  - *Molecular mechanics, semi-empirical methods, ab initio models, DFT, …*
- Hyperchem
  - *Molecular mechanics, molecular dynamics, semi-empirical and ab initio methods*
- EMPIRE
  - *Semi-empirical methods*

Conformational Sampling

- Generate set of diverse conformations
- Biologically relevant conformations in drug discovery
  - *Bioactive conformation*
  - *Conformation in receptor-bound state*
- General case of 3D structure generation
- Applications, *e.g.*,
  - *Docking*
  - *Pharmacophore searching*

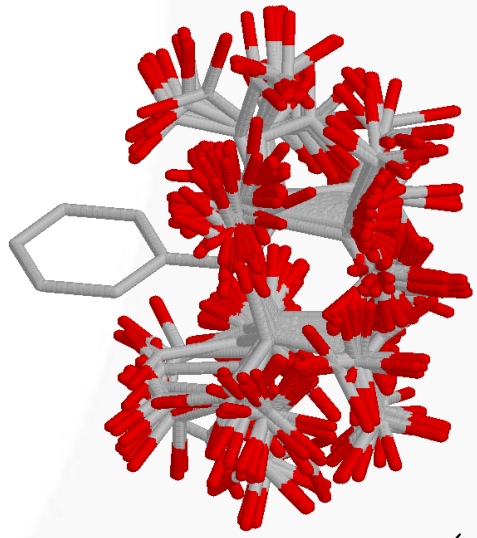


MN AM

Conformational Sampling – General Workflow

- Identification of rotatable bonds
  - *Flexible ring systems*
  - *Open-chain portions*

- Generation of conformations
  - *Application of implemented algorithm*
  - *Avoidance of "un-favorable" conformations*
  - *Combinatorial explosion*



$$N = \left(\frac{360}{k}\right)^{n}$$

MN AM

Conformational Sampling – General Workflow

- Checking for duplicates and very similar conformations
  - *Inter-conformational distances in Cartesian space*
  - *Root mean square deviation ($RMSD_{XYZ}$)*

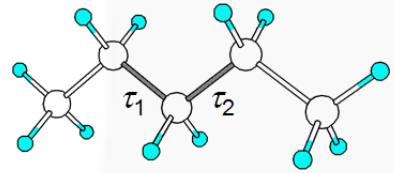
$$RMSD = \sqrt{\frac{\sum_i^N (\Delta d)^2}{N}}$$

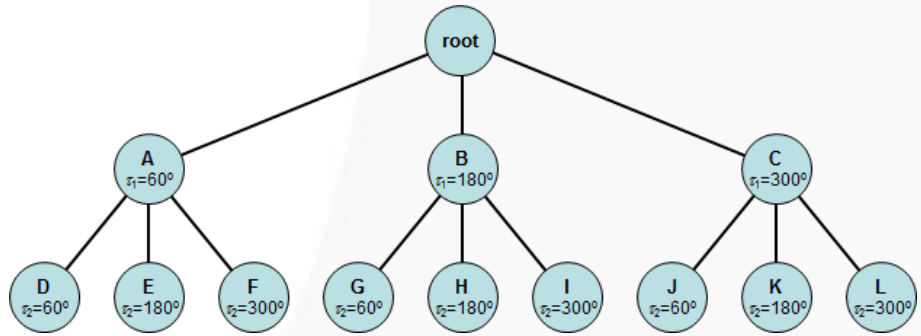
- Selection of a set of representative conformations
  - *Subsampling*
    - E.g., clustering in Cartesian space
  - *Ideally done in one of previous steps to bias search towards area of application*

MN AM

Conformational Sampling – Methods

- Rule- and data-based, fragment-based methods
  - *Empirical and fast*
- Systematic searches
  - *Exhaustive, e.g., grid searches*
  - *Combinatorial explosion*
- Random methods
  - *Random change of Cartesian or internal coordinates*
  - *Subsequent optimization*

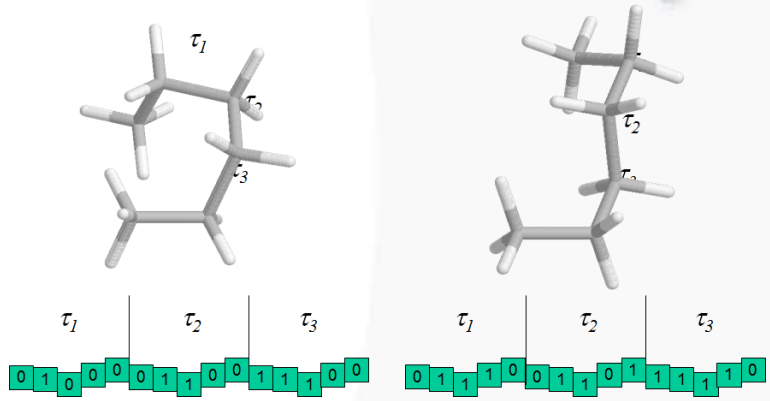




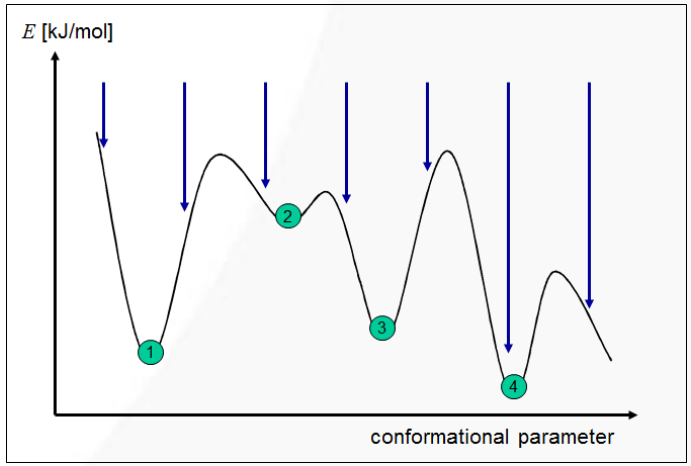
MN AM

Conformational Sampling – Methods

- Genetic algorithms
  - *Robust optimizers*
  - *Torsion angles as genes*
  - *Mutation/crossover operators*



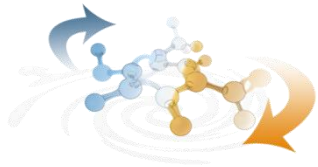
- Simulation methods
  - *Molecular dynamics*
  - *Simulated annealing*

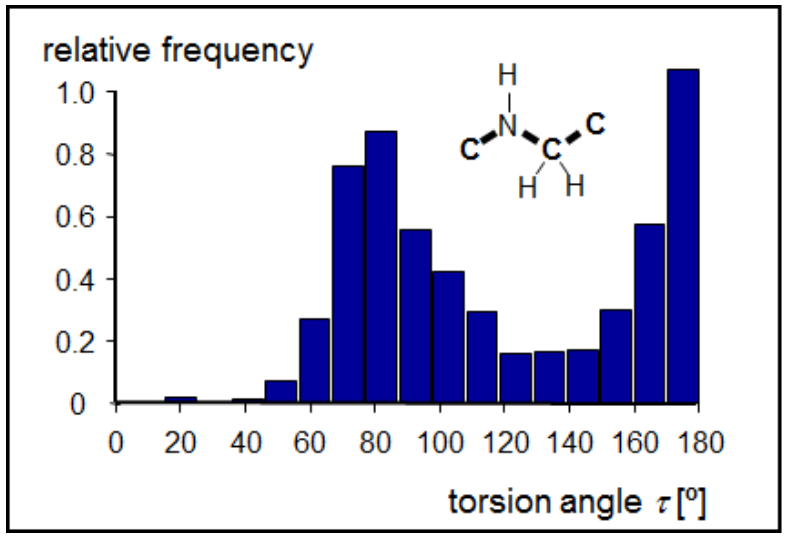


MN AM

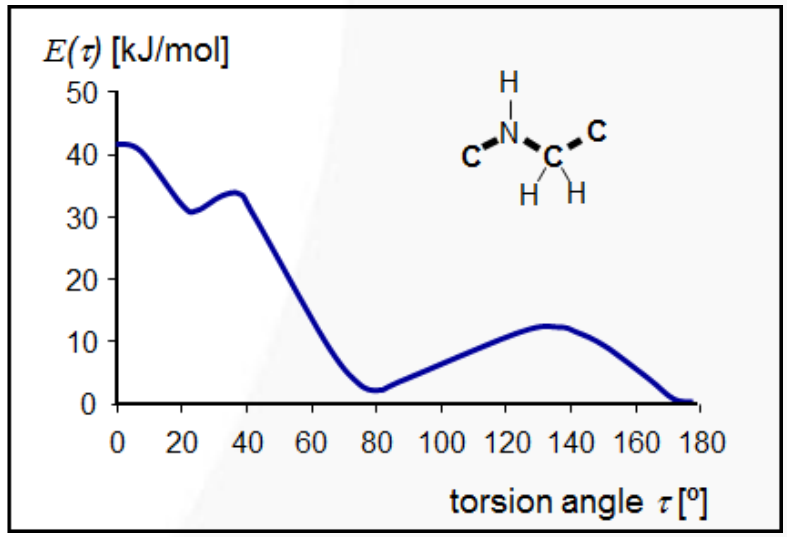
Conformer Generator ROTATE Classic

- Hybrid approach
  - *Systematic search*
  - *Rule- and data-based system*





Torsion angle distributions in small molecule crystal structures

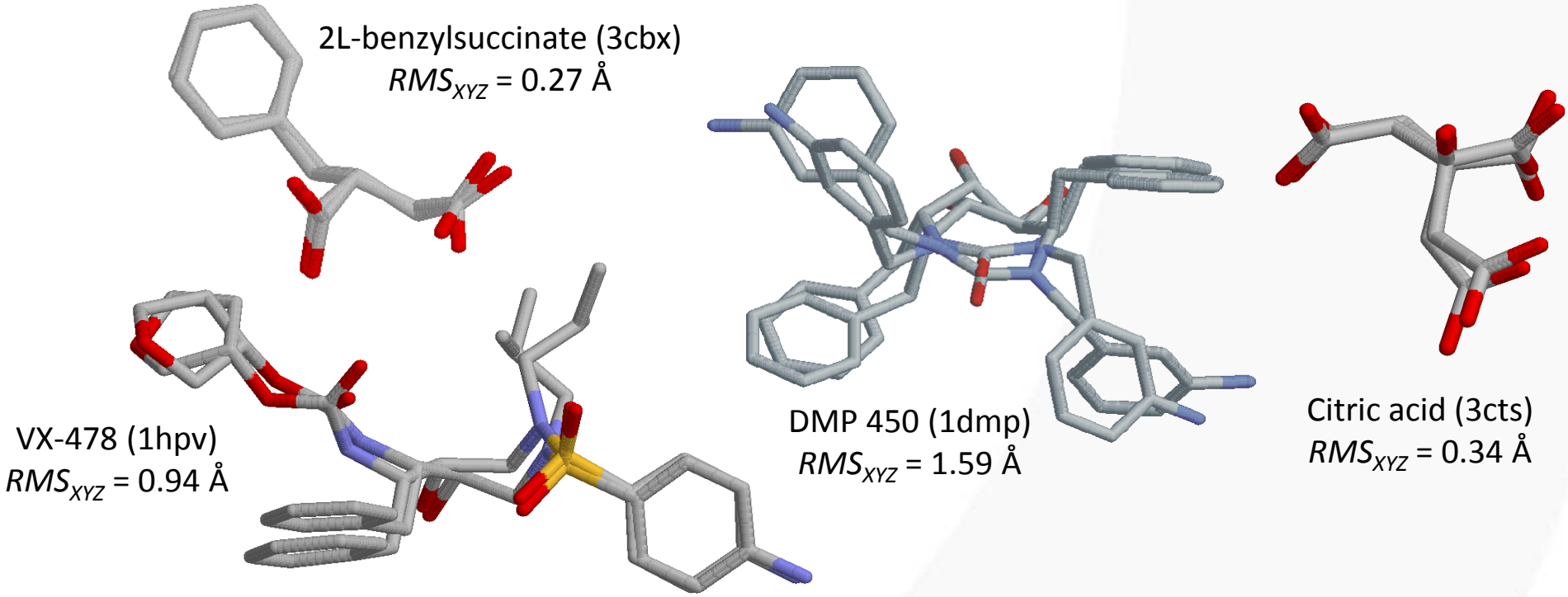


Empirical energy function

$$E(\tau) = -A \ln f(\tau)$$

Conformer Generator ROTATE Classic

- Superimpositions of experimentally determined receptor-bound geometries and ROTATE-generated models with smallest RMS_{XYZ}



MN AM

Available Methods

- CatConf/ConFirm
  - *Accelrys/Biovia*
- CEASAR
  - *Accelrys/Biovia*
- CORINA Classic and ROTATE Classic
  - *Molecular Networks*
- OMEGA
  - *OpenEye*
- CONCORD/CONFORT
  - *Tripos/Certara*
- MOE
  - *Chemical Computing Group*
- MacroModel/ConfGen
  - *Schrödinger*
- Free/open source
  - *Open Babel*
  - *RDKit*

MN AM

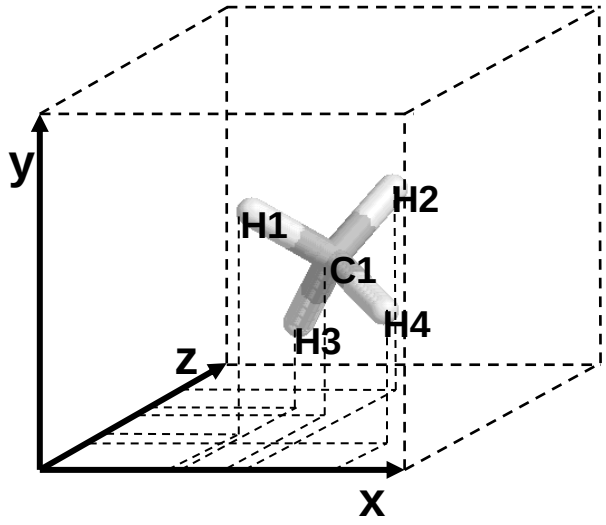
Storage of 3D Structures

- Various standard chemical file formats
- Cartesian (x,y,z) coordinates
  - *Molfile, SD file*
  - *SYBYL MOL/MOL2*
  - *PDB*
- Internal coordinates
  - *Crystallographic file formats, e.g., CIF*
  - *(Z matrix)*

MN AM

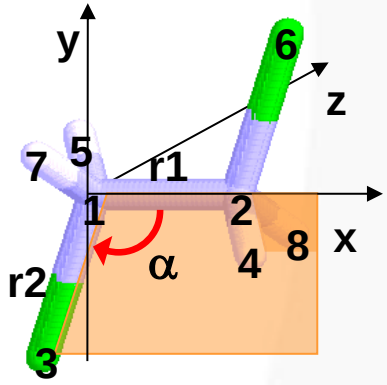
Storage of 3D Structures

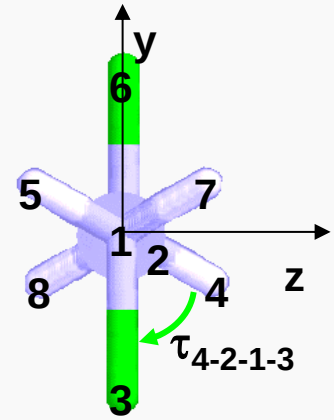
- Cartesian coordinates



	x	y	z
C1	-0.0127	1.0858	0.0080
H1	0.0021	-0.0041	0.0020
H2	1.0099	1.4631	0.0003
H3	-0.5399	1.4469	-0.8751
H4	-0.5229	1.4373	0.9048

- Internal coordinates





C1						
C2	1.5	1				
Cl3	1.7	1	109	2		
H4	1.1	2	109	1	-60	3
H5	1.1	1	109	2	180	4
Cl6	1.7	2	109	1	60	5
H7	1.1	1	109	2	-60	6
H8	1.1	2	109	1	180	7

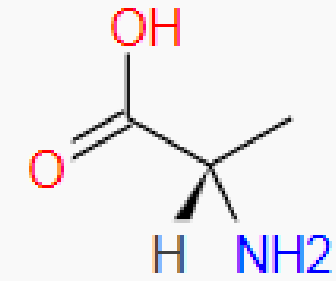
MN AM

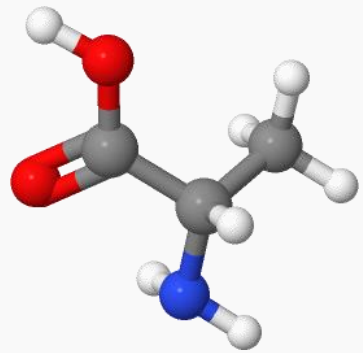
Molfile, Structure Data (SD) File

- SD V2000

```
L-alanine
          10191614583D 1   1.00000     0.00000

  7  6  0  0  0  0  0  0  0  0999 V2000
   -0.0184    1.5028    0.0103 C   0  0  0
    0.0021   -0.0041    0.0020 C   0  0  0
   -0.7002   -0.5305    1.2552 C   0  0  0
    1.0197    2.1211    0.0037 O   0  0  0
   -1.1898    2.1580    0.0194 O   0  0  0
    1.3935   -0.4748   -0.0138 N   0  0  0
   -0.5153   -0.3696   -0.8850 H   0  0  0
  1  2  1  0  0  0  0
  2  3  1  0  0  0  0
  1  4  2  0  0  0  0
  1  5  1  0  0  0  0
  2  6  1  0  0  0  0
  2  7  1  1  0  0  0
M  END
```



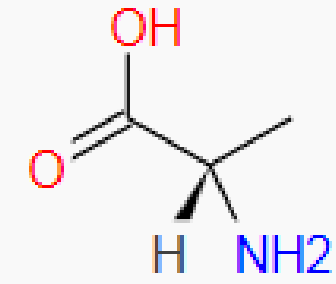


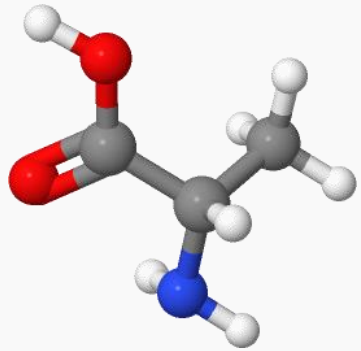
MN AM

SYBYL MOL/MOL2 File

- SYBYL MOL2 file with atom typing

```
@<TRIPOS>MOLECULE
L-alanine
   7     6     0     0     0
SMALL
NO_CHARGES
@<TRIPOS>ATOM
   1 C1             -0.0184     1.5028     0.0103 C.2
   2 C2              0.0021    -0.0041     0.0020 C.3
   3 C3             -0.7002    -0.5305     1.2552 C.3
   4 O4              1.0197     2.1211     0.0037 O.2
   5 O5             -1.1898     2.1580     0.0194 O.3
   6 N6              1.3935    -0.4748    -0.0138 N.3
   7 H7             -0.5153    -0.3696    -0.8850 H
@<TRIPOS>BOND
   1    1    2 1
   2    1    4 2
   3    1    5 1
   4    2    3 1
   5    2    6 1
   6    2    7 1
#          End of record
```



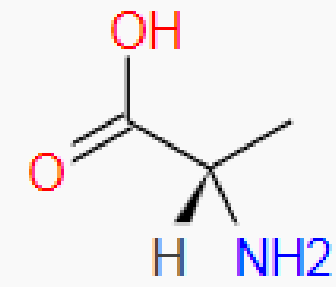


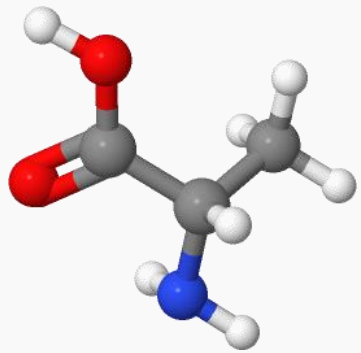
MN AM

(Brookhaven) Protein Data Bank File

- Example

```
HEADER    UNK                                      16-10-19    1UNK
COMPND    L-alanine
REMARK
HETATM    1  C1  UNK  1     -0.018   1.503   0.010  1.00 20.00
HETATM    2  C2  UNK  1      0.002  -0.004   0.002  1.00 20.00
HETATM    3  C3  UNK  1     -0.700  -0.531   1.255  1.00 20.00
HETATM    4  O4  UNK  1      1.020   2.121   0.004  1.00 20.00
HETATM    5  O5  UNK  1     -1.190   2.158   0.019  1.00 20.00
HETATM    6  N6  UNK  1      1.394  -0.475  -0.014  1.00 20.00
HETATM    7  H7  UNK  1     -0.515  -0.370  -0.885  1.00 20.00
CONECT    1    2    4    5
CONECT    2    1    3    6    7
CONECT    3    2
CONECT    4    1
CONECT    5    1
CONECT    6    2
CONECT    7    2
END
```





MN AM

Crystallographic Information File

- CIF file with internal coordinates
  - *Parameters for refinement of X-ray structures*

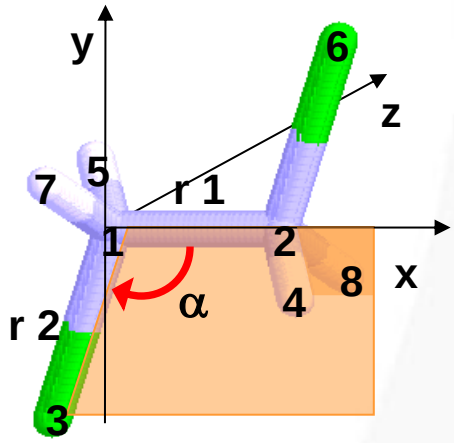
```
<keywords for connectivity>
UNK   'C1'    n/a   'C2'   START
UNK   'C2'   'C1'   'C3'    .
…
UNK   'H7'   'C2'     .     END
<keywords for bonds>
UNK   'C1'   'C2'   single   1.507   0.020
UNK   'C1'   'O4'   double   1.208   0.020
…
<keywords for bond angles>
UNK   'C2'   'C1'   'O4'   120.000   3.000
UNK   'C2'   'C1'   'O5'   120.000   3.000
…
<keywords for torsion angles>
UNK   var_000     'O4'   'C1'   'C2'   'C3'    120.000  20.0000   6
```

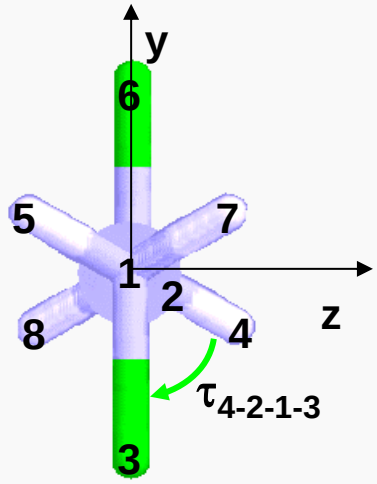
MN AM

Z Matrix

- Input to quantum mechanical programs

```
C1
C2    1.5   1
Cl3   1.7   1   109   2
H4    1.1   2   109   1   -60   3
H5    1.1   1   109   2   180   4
Cl6   1.7   2   109   1    60   5
H7    1.1   1   109   2   -60   6
H8    1.1   2   109   1   180   7
```

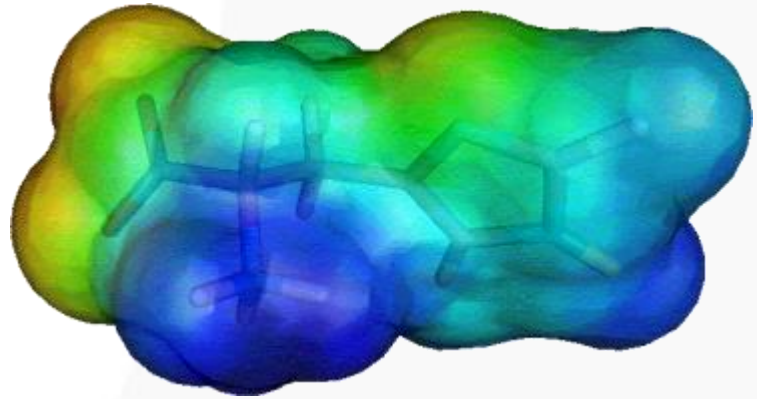




Molecular Surfaces

- Interaction between molecules mediated through surface (properties)
  - *Electrostatic, hydro/lipophilicity, H-bond donor/acceptor potential*

- Common 3D surface models
  - *van der Waals surface*
  - *Conolly surface*
  - *Solvent-accessible surface (SAS)*

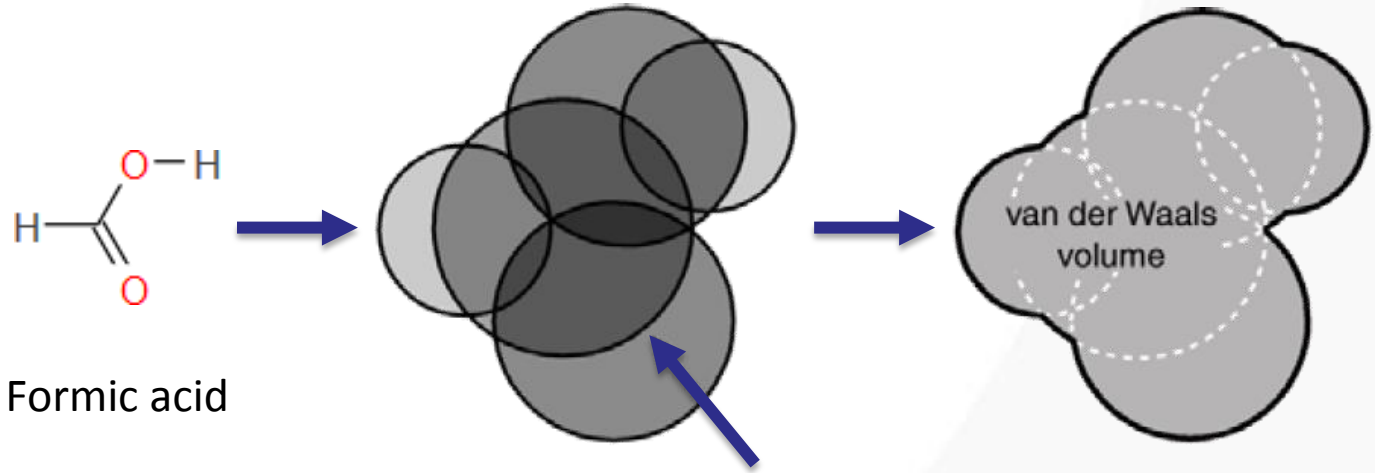


- 2D approaches
  - *Topological polar surface area*
  - *Approximate surface area*

MN AM

van der Waals Surface

- Simplest representation of molecular surface
  - *Hard-sphere model*
  - *Determined from van der Waals radii of each atom*
    - Correction by energetically-favorable distance on non-bonded atoms



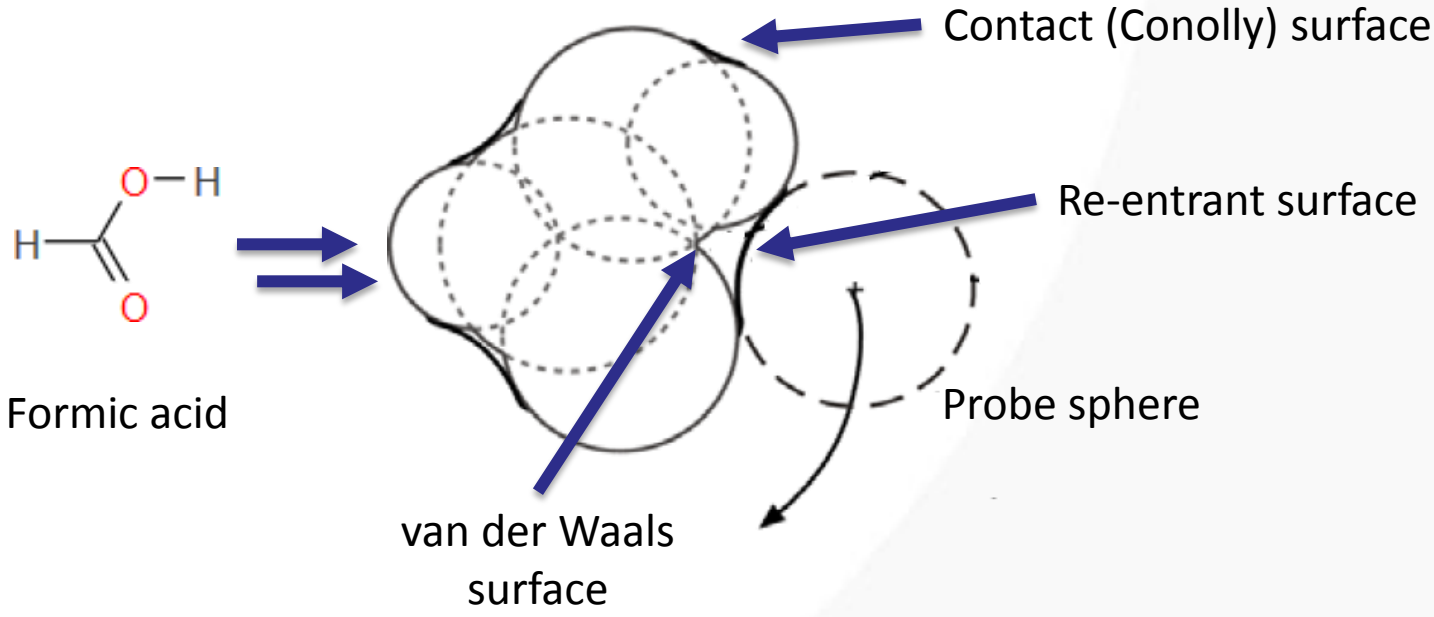
Formic acid

van der Waals radius of carboxyl carbon atom

van der Waals surface

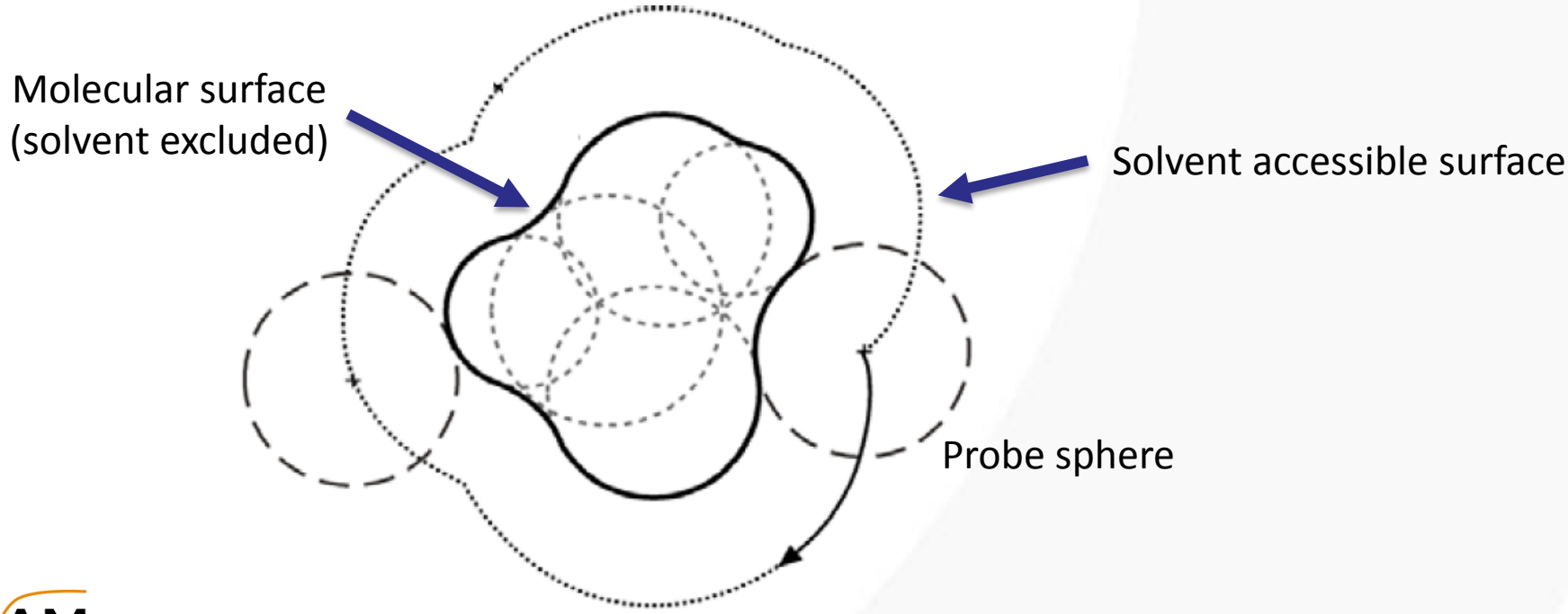
Conolly Surface

- Smoother surface (than van der Waals)
  - *"Rolling" a probe sphere (e.g., solvent molecule) over van der Waals surface*
    - Water molecule with effective radius of 1.4 Å
  - *Convex contact surface*
  - *Convex re-entrant surface where probe contacts two or more atom spheres*



Solvent Accessible Surface

- Similar to Conolly surface, but center of probe sphere defines surface
  - *Molecular surface that a solvent molecule can access*
  - *Conolly molecular surface extended by effective radius of solvent molecule*

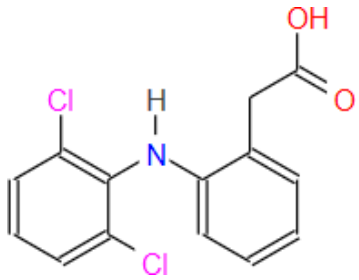


MN AM

2D Approaches for Surfaces

- Topological polar surface area (Ertl *et al.*, 2000)
  - *Estimation of polar surface area by sum of contributions of 2D polar fragments*
  - *Contribution derived by fitting of fragment-based TPSA to 3D polar surface area*
    - Based on 39,000 drug molecules
    - 43 fragments
    - r^2 of 0.98 and average error of 5.6 Å

CORINA Symphony



Query compound

Fragment	PSA	Frequency
NR_3	3.24	0
NHR_2	12.03	1
NH_2R	26.02	0
R-O-R	9.23	0
R-O-H	20.23	1
C=O	17.07	1
…	…	…

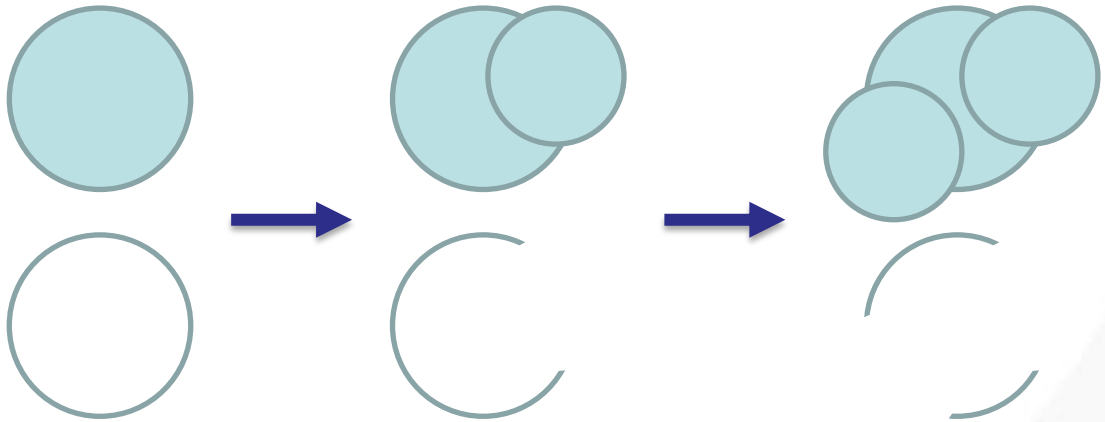
TPSA: 49.3 Å

MN AM

2D Approaches for Surfaces

- Approximate surface area (Labute, 2000)
  - *van der Waals radii and ideal bond length*
  - *Geometrical considerations to approximate overlap of spheres and remaining "free" surface*
  - *1,947 molecules*
    - r^2 of 0.96 and average error <10%

CORINA Symphony

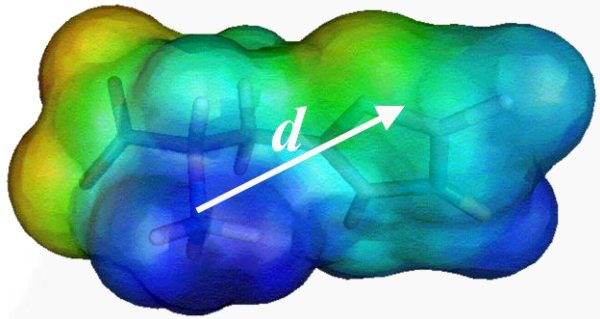


MN/AM

Molecular Surface Descriptors

- QSAR experiments
  - *Molecular descriptors independent of size and number of atoms of molecules*
  - *Invariant against translation and rotation of molecules*
  - *Fixed numbers of descriptors for each molecule in data set*

- Autocorrelation of surface properties
  - *Calculation of e.g. electrostatic potential for points on molecular surface*
  - *Application of autocorrelation formalism*
    - Sampling of pairwise surface property points in equidistant intervals

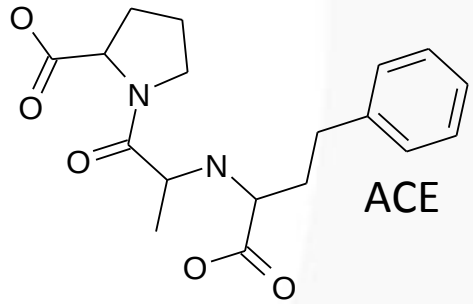


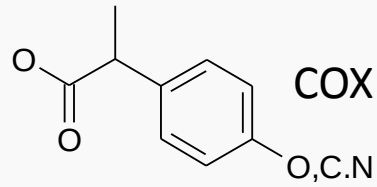
$$A(d_n) = \frac{1}{2L_n} \sum_{\substack{i,j \\ i \neq j}} p_i p_j$$

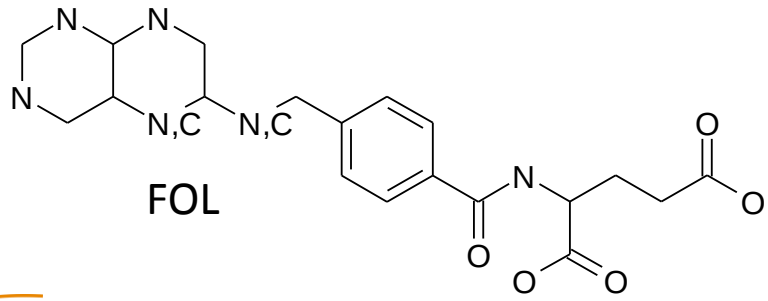
MN AM

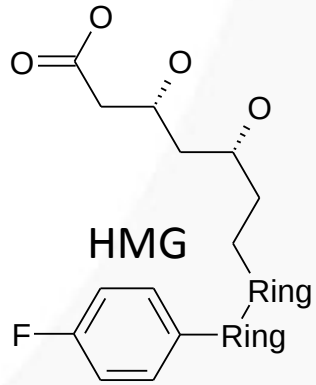
Application – Differentiating Drug Space

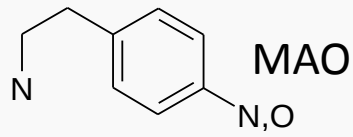
- Dataset of 207 drugs from five different pharmacological actions
  - *56 ACE inhibitors (ACE)*
  - *49 COX inhibitors (COX)*
  - *38 folic acid antagonists (FOL)*
  - *11 HMG-coA reductase inhibitors (HMG)*
  - *53 monoamine oxidase inhibitors (MAO)*
- Most similar COX and MOA
  - *Common structural features*







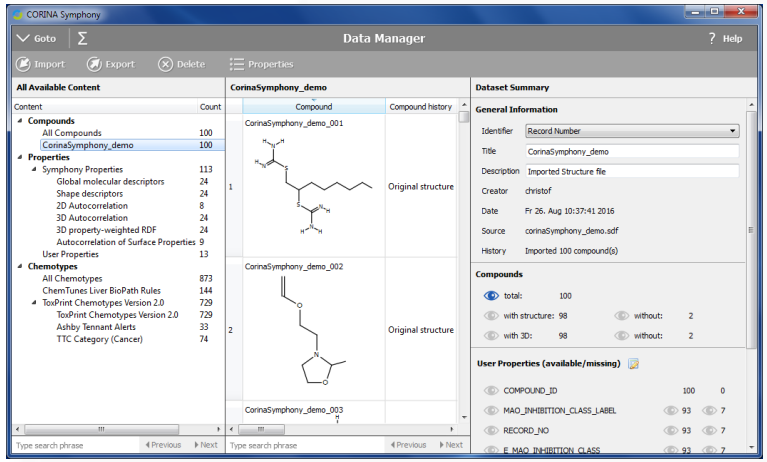




MN AM

Application – Differentiating Drug Space

- 116 CORINA Symphony descriptors
  - *Global molecular properties*
    - # H bond acceptors and H bond donors, TPSA, molecular weight, dipole moment, molecular polarizability, log*P* and log*S*
  - *3D autocorrelation vectors using partial charges, electronegativities and polarizabilities as atom pair properties*
  - *Autocorrelation of surface properties using molecular electrostatic, hydrogen bonding and hydrophobicity potential*

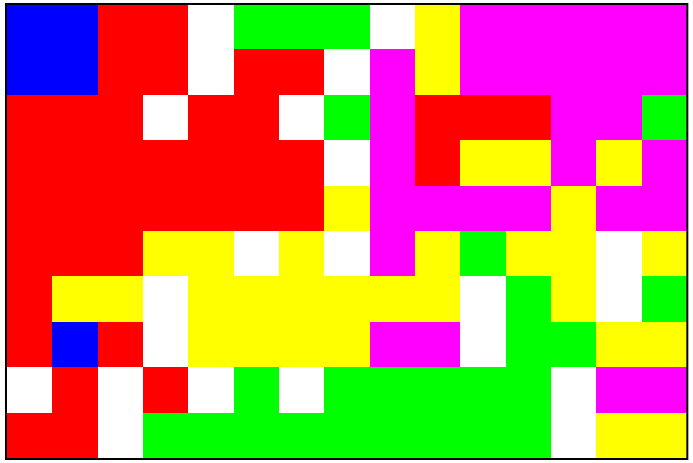


CORINA Symphony

MN AM

Unsupervised Classification

- Kohonen map trained with CORINA Symphony descriptors



207 compounds

116 descriptors

15x10 neurons

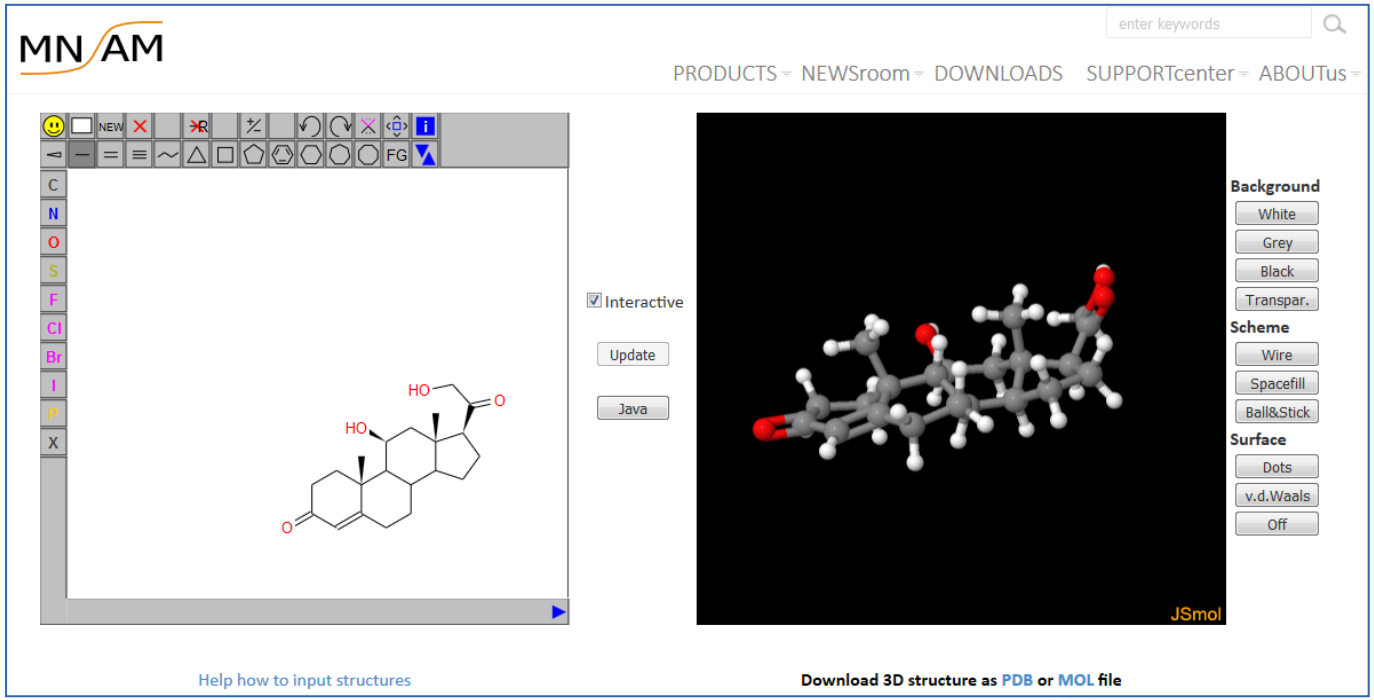
rectangular

- Acceptable separation of classes
  - *Only 5 conflict neurons*
  - *Only 4 of 10 possible class collisions*

MN AM

CORINA Classic – Online Service

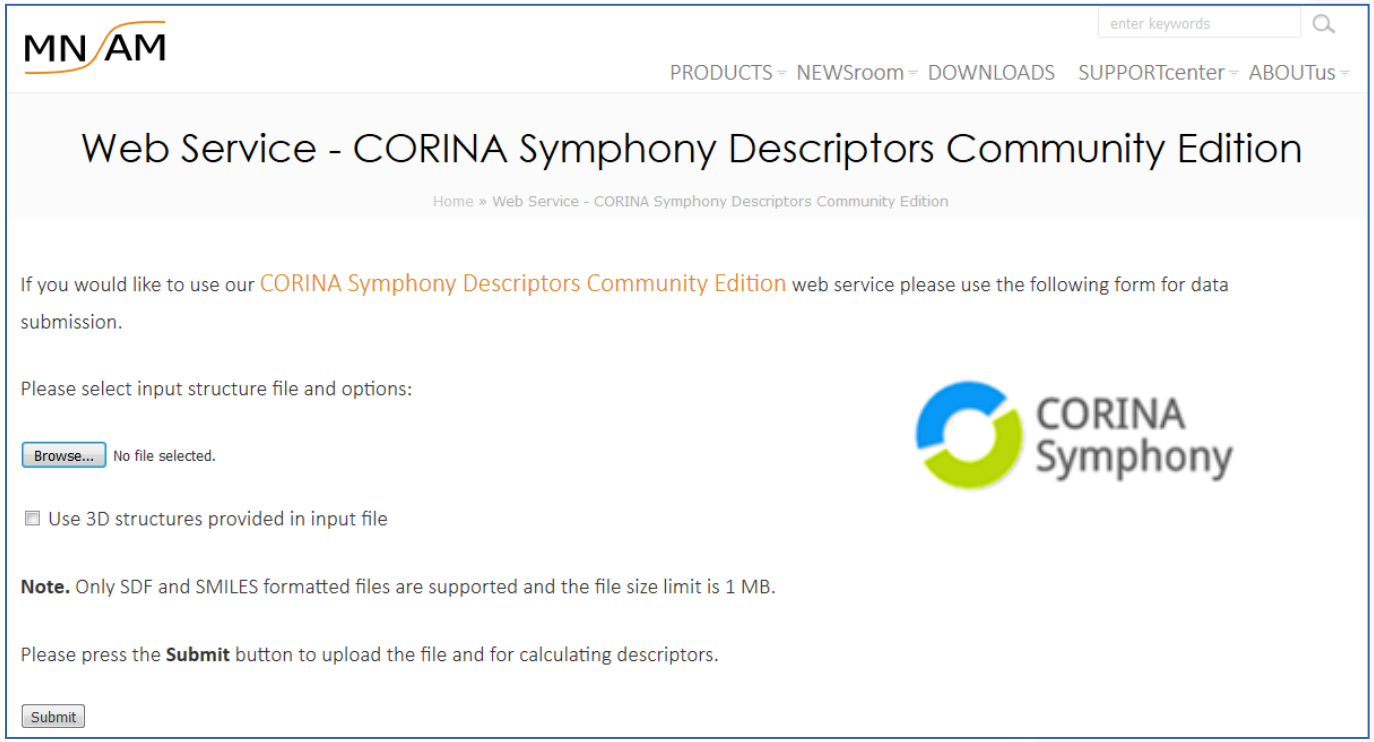
- Calculation of 3D structures
  - *https://www.mn-am.com/online_demos/corina_demo_interactive*



CORINA Classic

CORINA Symphony CE – Online Service

- Calculation of molecular properties and descriptors
  - *https://www.mn-am.com/services/corinasymphonydescriptors*



Acknowledgements

- BigChem project
  - *Barbara Gasset, Igor Tetko*

- My colleagues at MN-AM
  - *Aleksandra Mostrag-Szlichtyng, Chihae Yang, Vessela Vitcheva, Aleksey Tarkhov, Bruno Bienfait, James Rathman, Jörg Marucszyk, Oliver Sacher, Thomas Kleinöder, Tomasz Magdziarz, Vijay Gombar*

- Thank you for your attention!

- www.mn-am.com

MN AM

Selected References

- Databases
  - *Cambridge Crysatllographic Database*
    - Allen, F. H.; Kennard, O. Acc. Chem. Res., 1983, 16, 146. c) Allen, F.H.; Hoy, V.J. Cambridge Structural Database. In Encyclopedia of Computational Chemistry, Schleyer, P.v.R.; Allinger, N.L.; Clark, T.; Gasteiger, J.; Kollman, P.A.; Schaefer, III, H.F.; Schreiner, P.R. (Eds.), John Wiley & Sons, Inc., Chichester, UK, **1998**; pp. 155-167.
    - http://www.ccdc.cam.ac.uk/
  - *The PubChem project*
    - https://pubchem.ncbi.nlm.nih.gov
  - *RCSB Protein Data Bank*
    - http://www.rcsb.org/pdb/home/home.do

MN AM

Selected References (cont.)

- 3D structure and conformation generators
  - *Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. J. Chem. Inf. Comput. Sci.* ***1994****, 34, 1000-1008.*
  - *Schwab, C.H., Representation of 3D Structures. In Chemoinformatics - A Textbook, J. Gasteiger and T. Engel, Eds., Wiley-VCH, Weinheim,* ***2003****, 91-124.*
  - *Sadowski, J. Representation of 3D Structures. In Handbook of Chemoinformatics - From Data to Knowledge, J. Gasteiger, Ed., Vol. 1, Wiley-VCH, Weinheim,* ***2003****, 231-261.*
  - *Schwab, C.H. Conformational Analysis and Searching. In Handbook of Chemoinformatics - From Data to Knowledge, J. Gasteiger, Ed., Vol. 1, Wiley-VCH, Weinheim,* ***2003****, 262-301.*

Selected References (cont.)

- 3D structure and conformation generators (cont.)
  - *Renner, S.; Schwab, C.H.; Schneider, G.; Gasteiger, J. Impact of conformational flexibility on three-dimensional similarity searching using correlation vectors. J. Comp. Inf. Model.* ***2006****, 46, 2324-2332.*
  - *Schwab, C.H. Conformations and 3D pharmacophore searching. Drug Discovery Today: Technologies, Volume 7, Issue 4, Winter* ***2010****, e245-e253.*

Selected References (cont.)

- Force field calculations
  - *Leach A.R. Molecular Modelling: Principles and Applications. Prentice Hall, 2nd Edition, 2001.*
  - *Lanig, H. Molecular Mechanics. In Chemoinformatics - A Textbook, J. Gasteiger and T. Engel, Eds., Wiley-VCH, Weinheim, **2003**, 338-358.*
  - *Lanig, H. Molecular Mechanics. in Handbook of Chemoinformatics - From Data to Knowledge, J. Gasteiger, Ed., Vol. 3, Wiley-VCH, Weinheim, **2003**, 920-946.*

Selected References (cont.)

- Quantum mechanical calculations
  - *Leach A.R. Molecular Modelling: Principles and Applications. Prentice Hall, 2nd Edition, 2001.*
  - *Clark, T. Quantum Mechanics. In Chemoinformatics - A Textbook, J. Gasteiger and T. Engel, Eds., Wiley-VCH, Weinheim, **2003**, 376-400.*
  - *Clark, T. Quantum Mechanics. In Handbook of Chemoinformatics - From Data to Knowledge, J. Gasteiger, Ed., Vol. 3, Wiley-VCH, Weinheim, **2003**, 947-975.*

- Structure representation and chemical file formats
  - *Engel, T. Representation of Chemical Compounds. In Chemoinformatics - A Textbook, J. Gasteiger and T. Engel, Eds., Wiley-VCH, Weinheim, **2003**, 15-90.*

Selected References (cont.)

- Structure representation and chemical file formats (cont.)
  - *SD File format*
    - Dalby, A. et al. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. J. Chem. Inf. Comput. Sci. **1992**, 32, 244-255.
    - http://accelrys.com/products/informatics/cheminformatics/ctfile-formats/no-fee.php
  - *PDB file format*
    - Bernstein, F. C. et al. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. J. Mol. Biol. **1977**, 112, 535-542.
    - http://www.wwpdb.org/documentation/file-format
  - *CIF format*
    - Hall, S.R. et al. The Crystallographic Information File (CIF): a New Standard Archive File. Acta Cryst. **1991**, A47, 655-685.

Selected References (cont.)

- Molecular surfaces
  - *Engel, T. Representation of Chemical Compounds. In Chemoinformatics - A Textbook, J. Gasteiger and T. Engel, Eds., Wiley-VCH, Weinheim,* ***2003****, 124-129.*
  - *Connolly M.L. Analytical molecular surface calculation. J. Applied Crystallogr. 1983, 16, 548-558.*
  - *Zhao, Y.H.; Abraham; M.H.; Zissimos, A.M. Determination of McGowan Volumes for Ions and Correlation with van der Waals Volumes. J. Chem. Inf. Comp. Sci.* ***2003****, 43, 1848-1854.*
  - *Ertl, P; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Tansport Properties. J. Med. Chem.* ***2000****, 43, 3714-3717.*
  - *Labute, P. A widely applicable set of descriptors. J. Mol. Graph. Mod.* ***2000****, 18, 464-477.*

Selected References (cont.)

- Neural networks
  - *Zupan, J.; Gasteiger, J. Neural Networks in Chemistry and Drug Design. Second Edition, Wiley-VCH, Weinheim, **1999**, 380 pages.*
  - *Zupan, J.; Gasteiger, J. Neural Networks in Chemistry. Angew. Chem. Int. Ed. Engl. **1993**, 32, 503-527.*
  - *Terfloth, L.; Gasteiger, J. Self-organizing Neural Networks in Drug Design. Screening - Trends in Drug Discovery 2001, 2(4), 49-51.*