

# Protein Modeling with Discovery Studio

**Kimmo Mattila**  
**[kimmo.mattila@csc.fi](mailto:kimmo.mattila@csc.fi)**



# Discovery Studio

- Discovery Studio (DS) is a commercial molecular modeling program for biological macromolecules (proteins, nucleic acid).
- World is full of good molecular modeling programs like:
  - Sybyl
  - Maestro
  - gOpenMol
  - VMD
  - PyMol
  - Discovery Studio

Is Discovery Studio any better?



# Discovery Studio at CSC

- CSC has a national academic license and installation files for DS
- The program can be installed in users own Windows or Linux pc (requires fixed IP-address)
- You can use DS in hippu1.csc.fi using X-term, Nomachine or DS-client connection
- The token license system does not limit the amount of installations but the amount of simultaneous users. (Close your DS if you are not using it)



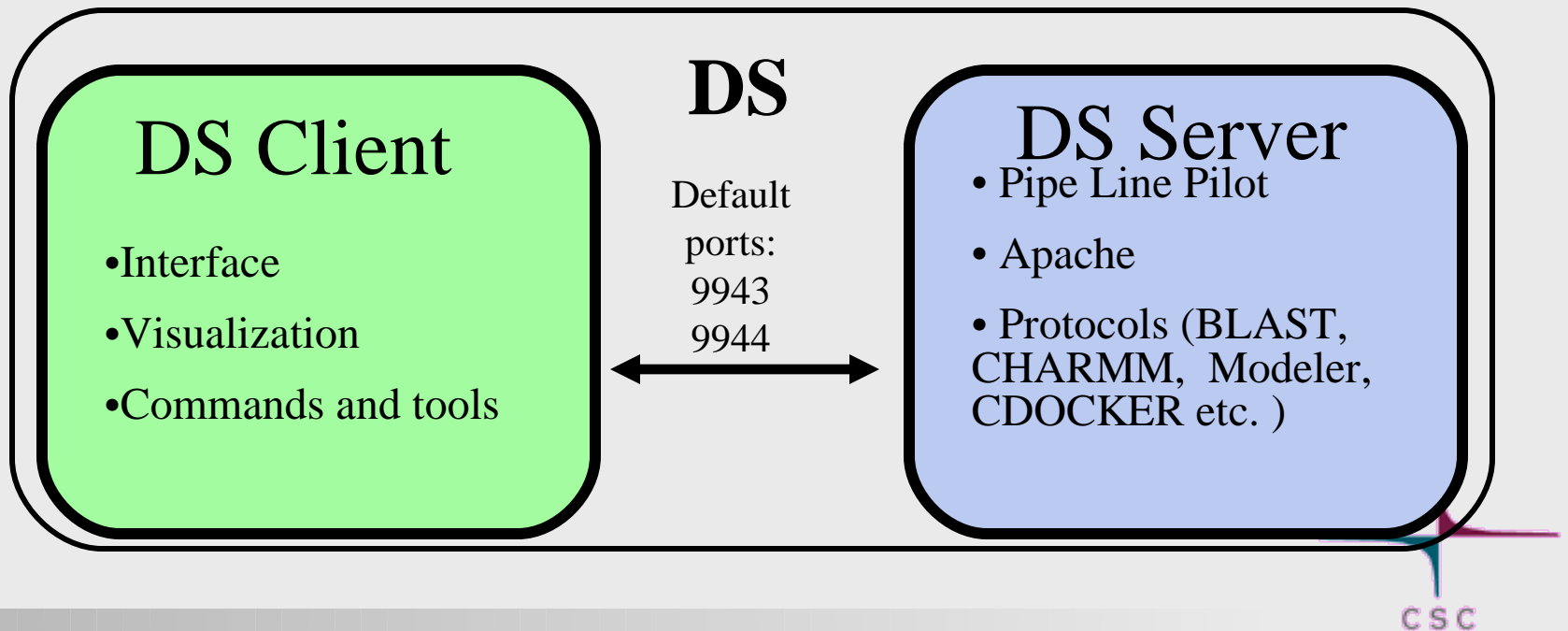
# Installing Discovery Studio

- Academic university researchers can instal the software to a local computer
- Installation instuctions can be found form:  
[http://www.csc.fi/english/research/sciences/bioscience/programs/ds/ds\\_install](http://www.csc.fi/english/research/sciences/bioscience/programs/ds/ds_install)
- A group vise license contract is needed
- A fixed IP address is needed
- Installation files are downloaded from the scientist's user interface (CSC user account is needed)  
<https://sui.csc.fi/group/sui/downloads>
- DS client only: 300 MB (windows), 400 MB (Linux)
- DS complete: 3,2 GB (windows), 3,7 GB (linux)



# Structure of Discovery Studio

- Both client and server are normally installed in the same machine
- separate server machines can be used too
- In most cases you do not need to worry about this
- If you use Hippu1 as your DS server you do not need to install the whole package but just the client



# Discovery Studio interface

Command menus

Toolbars

Protocols

Tools

Hierarchy view

Accelrys Discovery Studio Client 2.5

File Edit View Chemistry Structure Sequence Chart Scripts Window Help

Tools Protocols Files

Welcome - Html Window 1aql - Molecule Window 1aql - 2D Window

3D Window (graphics)

Protocol Name	Status	Details	Elapsed Time	Start Date
Calculate Binding Ener	Error	100% step 3/3:	21:21:39	Thu May 6 11
Dock Ligands (CDOCK)	Success	10 poses: indor	0:35:41	Thu May 6 10
Minimization	Error		0:35:04	Thu Apr 29 0
Minimization	Success	-1196994.4107	1:39:19	Thu Apr 29 0
Solvation	Success	21436 solvent r	15:35:44	Wed Apr 28 1
BLAST Search (DS Ser	Success	250 hits: P0053;	0:01:04	Mon Apr 26 1

Parameter Name	Parameter Value
Input Sequence	
Input Database	PDB_nr95
Scoring Matrix	BLOSUM62
Gap Penalties	Existence: 11 Extension: 1
Gapped Alignment	True
E-value Cutoff	10

Server: localhost:9943

# Discovery Studio menu commands

- **File menu:** Contains commands for tasks such as opening molecular data files, saving files to disk, printing, and accessing windows.
- **Edit menu:** Contains commands for tasks such as copying and pasting, selecting, finding, and setting preferences.
- **View menu:** Contains commands for tasks such as changing the way objects appear in the various views and for choosing which views should be shown or hidden.
- **Chemistry menu:** Contains commands for tasks that modify the chemical makeup of the molecules.
- **Structure menu:** Contains commands for tasks such as adding or removing labels, adding or removing structure monitors, calculating the solvent accessibility, cleaning up geometry, and superimposing multiple molecules.
- **Sequence menu:** Contains submenus and commands to manage protein sequences and protein sequence alignments.
- **Window menu:** Contains commands that allow you to control the display of open windows in the current Discovery Studio session.
- **Help menu:** Contains commands to access the Discovery Studio Help system and the Accelrys website.



# DS toolbars

- “Toolbars” are mostly shortcuts to menu commands
- However, some functions are used only through Toolbars
- Not all toolbars are normally visible. You can add or hide toolbars from:

**View | Toolbars**



# DS tools

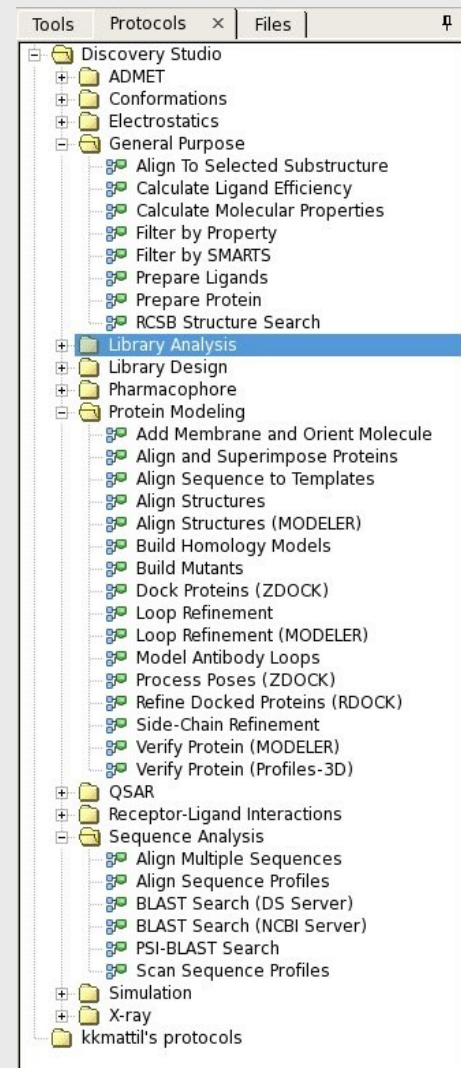
- “Tools” contain methods to analyze and modify your molecular model
- Tools panel can be made visible from:  
**View | Explorers | Tools**
- The CSC license covers most but not all the tools
- Most of the tools are run within the client but some require connection to the DS server



All		
Analyze and Edit Transmembrane Protei...	?	+ x
Analyze Binding Site	?	+ x
Analyze Docked Proteins	?	+ x
Analyze Sequences	?	+ x
Analyze Small Molecules	?	+ x
Analyze Trajectory	?	+ x
Assign Electrostatic Parameters	?	+ x
Build and Edit Nucleic Acid	?	+ x
Build and Edit Protein	?	+ x
Build and Refine X-ray Structure	?	+ x
Build Fragment	?	+ x
Customize Pharmacophore Features	?	+ x
Define and Edit Binding Site	?	+ x
Edit and Cluster Pharmacophores	?	+ x
Edit X-ray Structure	?	+ x
Navigate and Label 3D Structure	?	+ x
Place and Refine X-ray Ligand Structure	?	+ x
Protein Reports and Utilities	?	+ x
Search Small Molecule Conformations	?	+ x
Search Side-Chain Rotamers	?	+ x
Simulate Structures	?	+ x
Validate Protein Structure	?	+ x
Visualize Receptor-Ligand Interactions	?	+ x
Guide - Create and Analyze Conformati...	?	+ x
Guide - Create and Test Pharmacophore	?	+ x
Guide - Create Homology Models	?	+ x
Guide - Create Protein Sequence Align...	?	+ x
Analyze Protein Conservation Pattern	?	+ x
Superimpose Proteins	?	+ x
Guide - Dock Ligands	?	+ x
Guide - Sketch and Align Molecules	?	+ x

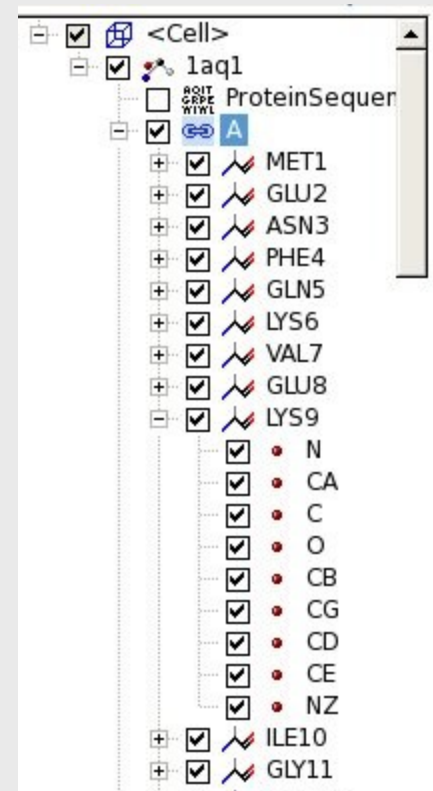
# DS Protocols

- Protocols are more advanced modeling and analysis tasks that are computed by the DS server
- Tools panel can be made visible from:  
**View | Explorers | Protocols**
- Note that the license of CSC does not cover all the protocols and tools



# Hierarchy panel

- The hierarchy panel is opened from:  
**View | Hierarchy**
- You can use hierarchy to select, atoms, amino acids or molecules
- Selections can be made for the 3D-view, Data table and the sequence window too.
- The logic in Discovery studio is:
  1. First select the target
  2. Then select the command



# Data table

- The Data table is opened from: **View | Data table**
- Data table shows data and values associated to your molecular model
- The data can be viewed, modified and sorted in the data table
- If value is in gray background it can't be modified
- Try right clicking the data table window (you can find a hidden command menus, that open by right-click, all over Discovery Studio)

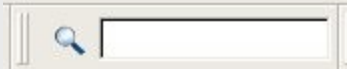
	Name	ID	Visible	Visibility Locked	Color	Parent	Type	PDB Name	1-Letter	1-Lett
1	MET1	1	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	Methionine	MET	M	M1
2	GLU2	2	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	GlutamicAcid	GLU	E	E2
3	ASN3	3	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	Asparagine	ASN	N	N3
4	PHE4	4	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	Phenylalanine	PHE	F	F4
5	GLN5	5	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	Glutamine	GLN	Q	Q5
6	LYS6	6	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	Lysine	LYS	K	K6
7	VAL7	7	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No		A	Valine	VAL	V	V7

Molecule | Cell | AminoAcidChain | ProteinSequence | AminoAcid | Atom | Chain | Residue | Group | Bond | PiCationMonitor | PiPiMonitor

# Help

- Discovery studio contains a large help system. No WWW or printed manual is available DS interface
- Open help form **Help | Topics** (if search tools are not visible in the Help window, press CONTROL + s)

Other tricks to find the missing functionality or parameter:

- Use the Feature search toolbar: 
- Try right-click to the window or object
- Check, if your parameter is located in preferences  
**Edit | preferences**



# Proteins and PDB



# Using PDB data in DS

- <http://www.rcsb.org/>
- experimentally determined protein structures are stored into PDB (Protein Data Bank) database
- Sources: X-ray diffraction (about 80%), NMR (15 %), others (5 %)
- over 65 000 structures (many of them related and nearly similar however)
- this is much less than the amount of known protein sequences (UniprotKB contains over 10 million sequences)



# PDBe Database

- “Processed” version of pdb
- Several search approaches e.g. ligand search, interface analysis,
- PISA to study assemblies interfaces and monomers
  
- <http://www.ebi.ac.uk/pdbe/>



# Using PDB data in Discovery Studio

- Protein structures can be automatically retrieved from the PDB database server using the four letter PDB-code.
  - **File | Open URL | PDB ID**
- You can do sequence based similarity searches to PDB with **BLAST protocols**.
- Protocol: **RCSB Structure search** enables metadata and sequence motif based searches.
- Local PDB formatted files can be imported too (e.g. files retrieved from PISA database).



# Using PDB files

- **Note that there can be several things that may need editing in your PDB file before you can start to use modeling tools**
- **Things that often require editing in PDB files:**
  - hydrogens are missing**
  - some side chains are missing**
  - part of main chain is missing**
  - ligand structure is not recognized**
  - multiple conformations for some side chains**
  - several structures in the PDB file**
- **Other things to check in a PDB-file:**
  - main chain omega, phi and psi angles**
  - side chain rotamers**



# Checking and fixing a protein

- **Protein Reports and Utilities** tool can be used to get an overall view to the selected PDB entry
- **Validate protein structure** tool can be used to check the PDB structure and report the problematic sites
- **Clean** function in the **Protein Reports and Utilities** tool can be used to fix some of the errors automatically
- Hydrogens can be added by **Chemistry/H**
- Protonation states can be fix with **Build and Edit Protein Tool**.
- The protonation states of neutral histidines are selected using **Edit | Preferences | Protein Utilities**



# Checking and fixing a protein

- Protocol: **Protocols | General Purpose | Prepare Protein**
  - Standardize atom names, insert missing atoms in residues and remove alternate conformations.
  - Remove water and ligand molecules depending on the settings.
  - Insert missing loop regions based on either SEQRES data or user specified loop definitions.
  - Optimize short and medium size loop regions with the LOOPER algorithm (optional).
  - Minimize the remaining loop regions (optional).
  - Calculate the pK and protonate the structure (optional).



# PDB file format

- PDB-file contains only information about atom locations (X,Y,Z).
- Data about bonds, partial charges or force field types is not included
- The fourth column contains normally R-factor, but can be something else too
- Common format for molecular data
- X-ray structures lack hydrogens and may contain several copies of the protein.
- NMR-structures contain several overlapping structures



# PDB file format

```
HEADER      RETINOL TRANSPORT                      27-JUL-92   1BRP      1BRP   2
COMPND      RETINOL BINDING PROTEIN (HOLO FORM)    1BRP      1BRP   3
SOURCE      HUMAN (HOMO SAPIENS) PLASMA
...
HELIX       1 1  VAL      6 SER      8 4 ONE SHORT TURN      1BRP   71
HELIX       2 2  PRO     146 GLU     158 1      1BRP   72
SHEET       1 S1  9 GLY     22 LYS     30 0      1BRP   73
...
SEQRES      1 A  182  GLU ARG ASP CYS ARG VAL SER SER PHE ARG VAL LYS GLU
SEQRES      2 A  182  ASN PHE ASP LYS ALA ARG PHE SER GLY THR TRP TYR ALA
SEQRES      3 A  182  MET ALA LYS LYS ASP PRO GLU GLY LEU PHE LEU GLN ASP
...
ATOM        1  N  GLU      1      22.826  21.377 -30.151  1.00100.00  1  1BRP  99
ATOM        2  CA GLU      1      23.744  21.686 -29.074  1.00100.00  1  1BRP 100
ATOM        3  C  GLU      1      23.395  23.023 -28.464  1.00100.00  1  1BRP 101
ATOM        4  O  GLU      1      22.798  23.102 -27.389  1.00100.00  1  1BRP 102
ATOM        5  CB GLU      1      25.225  21.681 -29.508  1.00100.00  1  1BRP 103
ATOM        6  CG GLU      1      26.155  20.992 -28.489  1.00100.00  1  1BRP 104
ATOM        7  CD GLU      1      27.285  21.840 -27.971  1.00100.00  1  1BRP 105
ATOM        8  OE1 GLU     1      28.301  22.075 -28.603  1.00100.00  1  1BRP 106
ATOM        9  OE2 GLU     1      27.087  22.244 -26.741  1.00100.00  1  1BRP 107
ATOM       10  N  ARG      2      23.771  24.073 -29.182  1.00100.00  1  1BRP 108
ATOM       11  CA ARG      2      23.485  25.397 -28.690  1.00 86.16  1  1BRP 109
ATOM       12  C  ARG      2      22.026  25.784 -28.629  1.00100.00  1  1BRP 110
...
HETATM 1483  O  HOH      229      2.848  65.969 -30.833  1.00 53.03  1BRP1581
HETATM 1484  O  HOH      230      38.756  38.831 -49.928  1.00 75.69  1BRP1582
```



# Forcefield methods



# Force field methods

## ➤ Quantum mechanics

- electronic structure calculations
- *Ab initio* methods
- Semi-empirical methods
- time-consuming
- high level accuracy
- < 500 heavy atoms

## ➤ Molecular mechanics

- atomic level
- simple models of the interactions to calculate the energy of molecule as a function of nuclear positions only
- Not so accurate as QM methods
- >10000 atoms



# Force Filed methods....

- Force field methods are usually used for biomolecules.
- to study complex system (i.e. the binding site in protein), quantum mechanics and molecular mechanics methods can be combined (hybrid QM/MM).



# Force Filed methods....

## ➤ Advantage

- large systems in reasonable calculation time.
- in some cases FF can provide results as accurate as the highest QM in a fraction of the CPU-time.

## ➤ Limitations

- no atom level electrostatics.
- dependent on the quality and availability of parameters.
- The calculated energy is relative.



# General form for Forcefield

- Energy consists of sum of terms each describing the energy required for distorting a molecule

$$\begin{aligned} E_{\text{pot}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{oop}} && \text{(internal terms)} \\ &+ E_{\text{elec}} + E_{\text{vdW}} && \text{(external terms)} \\ &+ E_{\text{constraint}} + E_{\text{user}} && \text{(special)} \end{aligned}$$

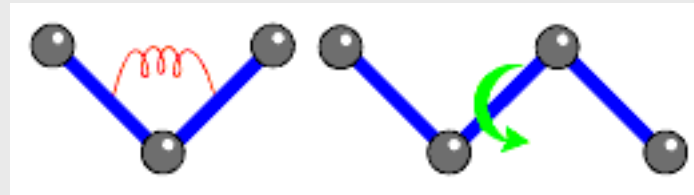
- $E_{\text{bond}}$  energy function for stretching a bond between two atoms
- $E_{\text{angle}}$  energy required for bending an angle
- $E_{\text{torsion}}$  torsional energy for rotation around a bond
- $E_{\text{elec}}$  electrostatic energy ( non-bonded interactions due to distribution of the electrons)
- $E_{\text{vdw}}$  Van der Waals energy (repulsion or attraction between non-bonded atoms)



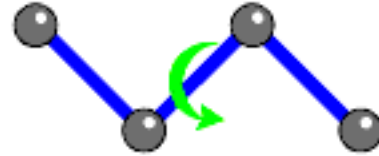
# Example. Functional form of CHARMM force field:



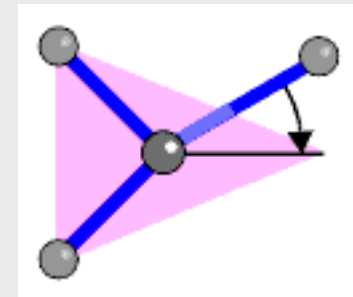
bond length



angle



dihedral angle



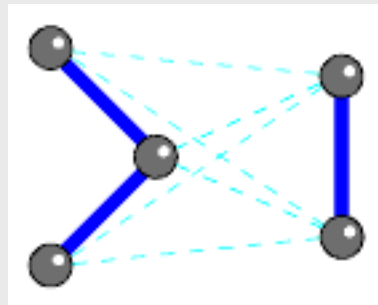
improper torsion

$$E_{\text{pot}} = \sum k_b (r - r_0)^2 + \sum K_\theta (\theta - \theta_0)^2 + \sum |k_\phi| - k_\phi \cos(n\phi) + \sum k_\chi (\chi - \chi_0)^2$$

$$+ \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \text{sw}(r_{ij}^2, r_{\text{on}}^2, r_{\text{off}}^2) + E_{\text{constraint}} + E_{\text{user}}$$

electrostatic

van der Waals



# Components of forcefield

- The forcefield contains the necessary building blocks for the calculations of energy and force:
  - A list of atom types.
  - A list of atomic charges (if not included in the atom-type information).
  - Functional forms for the components of the energy expression.
  - Parameters for the function terms.



# Forcefield....

- Torsion and non-bonded energy terms are most important for biomolecules.
- A force field is transferable (set of parameters developed on a small number of cases can be applied to much wider range of systems).
- Force fields are used in molecular mechanics and molecular dynamics calculations.



# Forcefield functions and parameters

Goal: a simple function for reproducing structural properties.

- 1) Empirical fitted force field: a functional form and parameters is designed to satisfy experimental results. ( cvff)
- 2) *Ab initio* fitted Force field: a functional form and parameters are specified using theoretical models and calculations. (cff)



# Parameter assignment

- the forcefield has the same functional form for all atoms but different parameters for each atom types.
- the atom type and its parameter depends on how that atom is bonded (example CHARMM has 38 different carbon types).
- atom type  $\neq$  atom name



# Parameter assignment...

- molecule can be neutral but the charge distribution is not equal=> partial atomic charges are determined.
- Partial charges are important!
  - hydrogen bonds
  - ionic bonds
  - dipole moment





# How to use forcefield ?

- several forcefields are available commercially.
- the validation of the force field depends on for which purpose it is designed and what properties are studied.
- the quality of force field parameters is essential.
- the complexity of functional form
- computational power (The computational time for calculating the force field energy grows as the square of the number of atoms).



- The ability to perform a calculation is no guarantee that results can be trusted !
  - unsuitable forcefield gives wrong results.
- A common problem: a lack of (good) parameters.
- different forcefields cannot usually be merged but the results can be compared.
- forcefield methods are good for predicting properties for classes of molecules where a lot of information exists.

# Forcefields in DS

Discovery Studio can use CHARMM forcefields:

CHARMm, CHARMm polar H, CHARMm19, CHARMm22,  
CHARMm27, XPLOLIG, MMFF, cff

Location of forcefield files in Discovery Studio

DiscoveryStudio17/share/forcefield

The InsightII manual chapter “**Forcefield based simulations**” gives a good introduction to force fields and their applications:

[https://extras.csc.fi/msimannual/doc/insight2005/ffbs/FF\\_SimulTOC.html](https://extras.csc.fi/msimannual/doc/insight2005/ffbs/FF_SimulTOC.html)



# Applications of forcefield methods

- Applications of forcefield methods
  - Molecular mechanics= Minimisation= Optimisation
  - Molecular dynamics=Simulation
- Basic assumption for using forcefield methods:  
**A REAL MOLECULE IS IN A STATE WHICH  
CORRESPONDS A MODEL NEAR  
POTENTIAL ENERGY MINIMUM**



# Search strategies

- Several different strategies in use.
- all methods are not using plain atomistic models
  - Molecular dynamics
  - Monte Carlo
  - Genetic algorithm
  - Fragment based method
  - Point complementarity methods

# Energy minimization

- “A REAL MOLECULE IS IN A STATE WHICH CORRESPONDS A MODEL NEAR POTENTIAL ENERGY MINIMUM” ,
- a search for the minimum of the potential energy surface defined by energy function, is done.
- Minimum energy arrangements of the atoms corresponds to stable state of the molecule.



# Three major protocols for minimization

- Steepest Descent
- Conjugate Gradient
- Adopted Bases Newton-Rhapson
- Powell



# The steepest descents method

- the gradient of potential energy determines the direction which leads to a largest reduction in energy. A step will be taken to that direction.
- Robust and simple method,
- useful when starting far from minimum
- Convergence is slow near minimum



# Conjugate gradient method

- the gradient of the previous step is included
- more efficient convergence to minimum → number of iterations smaller than for steepest descents
- works quickly when the molecule's structure is far from an energy minimum.
- the best choice for general use.
- more complex → time per iteration is longer than for steepest descents.



# Newton-Rhapson method

- use not only the first derivatives (i.e.the gradients) but also the second derivatives ( the curvature of the function) to locate a minimum.
- initial guess of structure need to be close to minimum.
- Convergence is fast near a minimum
- Computationally demanding for systems with many atoms (suitable for less than 100 atoms).



# Comparison of minimization methods

- **what determines which method is best?**
  - 1) size of system
  - 2) current state of optimization
- **robustness (i.e. ability to reach minimum regardless of initial conditions)**

**steepest descents > conjugate gradient and Newton-Raphson**
- **number of iterations**

**steepest descents > conjugate gradient > Newton-Raphson**



# Strategy in minimization

- **use steepest descents for first 10-100 steps to remove bad contacts.**
- **then use Newton-Raphson or conjugate gradients to complete minimization to convergence**
- **DS Smart minimizer:**
  1. steepest descents (max 1000 iterations)
  2. conjugate gradient (max 1000 iterations)



# Strategy in minimization

- **as a minimum approaches, the rate of convergence slows down and minimisation method crawls toward minimum at an ever decreasing speed.**
  - Convergence criteria (either for energy or conformation changes) and amount of minimization cycles are used to end minimisation.
- **The step size in minimisation algorithm is defined either by energy change in pervious step or by line search method.**



# Local or global energy minimum?

- **optimization methods can only locate the “close by” minimum, which is normally a local minimum, not a the global minimum from a given set of coordinates.**
- **To check if the minimum is the local or the global, all conformations need to be searched and the number of the minima grows typically exponentially with the number of variables.**
- **Systematic search is not possible for complete proteins and thus global minimum can not be defined.**



# Local or global energy minimum?

- **Systematic search can be used to check possible conformations, but only for small systems.**

N	Number of possible conformations	Time
1	3	3 s
5	243	4 min
15	14348907	166d

- **table<sup>1</sup>: possible conformation for linear alkanes CH<sub>3</sub>(CH<sub>2</sub>)<sub>n+1</sub>CH<sub>3</sub>)**

<sup>1</sup> F. Jensen: Introduction to Computational Chemistry

# Applications of energy minimization

- **relieve any unfavourable interactions in the initial configuration.**
- **calculations of the energy of molecular structure**
- **conformational search procedures**
- **normal mode analysis**



# Molecular dynamics

- Movements of molecule are simulated in certain temperature
- Temperature = thermal movement
- Each atom has position, mass and velocity
- Force field affects to the model according to the Newtons' law

$$F=ma$$

or

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -\nabla_i [E(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)] \quad i = 1, \dots, N$$



# Molecular dynamics

- **Total energy of the system is**
  - $E_{\text{tot}} = E_{\text{pot}} + E_{\text{kin}}$
- **Normal dynamics (Verlet/leap frog)**
- **Langevin dynamics**



# Using molecular dynamics

- **Proteins are not rigid but flexible objects**
  - Dynamical models are sometimes needed
  - fluctuations
  - Conformational search and changes
  - ligand binding
  - estimating thermodynamical parameters
  
- **Accuracy, size and timescales of the simulations are quite limited (tens of thousands of atoms, hundreds of nanoseconds)**



# Molecular dynamics parameters

## ➤ Handling of nonbonded interactions

- All interactions can't be explicitly included
- Cutoff
- Ewald summation
- Cell multipole

## ➤ Time step

- is limited by the highest frequency in the model
- 0,5-5 fs => millions of simulation steps are needed to reach nano second time scale.



# Molecular dynamics parameters

## ➤ **Temperature**

- normally 300 K
- Thermal equilibrium requires temperature control or careful heating
- Simulated annealing utilizes higher temperatures

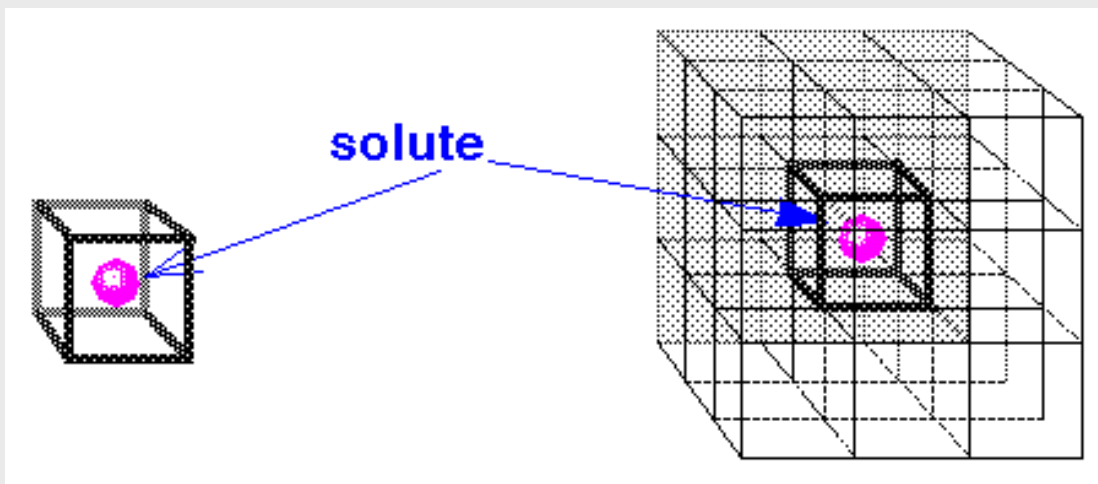
## ➤ **Other parameters**

- Force field scaling
- dielectricity



# Solvent environment

- Solvent (water) environment requires more computing but makes model more realistic
- Periodic boundary conditions are used to create continuous solvent environment
- Discovery Studio: Protocol Simulation/Solvation



# Analysis of molecular dynamics

- **The course of the simulation is recorded (trajectory).**
- **Later on several properties can be analyzed**
  - Geometry (angles distances)
  - Changes and fluctuations
  - Energies
  - Interactions, hydrogen bonds
  - Distributions and correlations
- **Discovery Studio: Animation toolbar , Analyze trajectory tool and Analysis Protocols modules.**



# Docking

- **many protein related biological processes are regulated or enabled by specific binding of small organic molecules (ligands) to the proteins**
  - signal transduction
  - enzyme activity /inhibition
- **many drugs are known or taught to work by binding a target protein**
- **what molecules could bind to the active site or where and how do the active molecules bind?**

# Docking

➤ **Two basic components:**

1. scoring function

2. search strategy

**initial position**

**optimizing search**

➤ **Systematic search is normally not possible**



# Docking

- **computer based docking can be used to predict binding geometries for large libraries of candidate molecules, if the protein structure is available**
- **speed is an issue (maximum duration few minutes/ligand)**
- **What do we want to know?**
  - does this molecule bind or not?
  - which of these molecules are most potential ligands?
  - what is the binding geometry or site of this molecule?
  - what is the binding affinity of this molecule?
- **docking does not try to simulate the binding process!**



# Scoring function

- **scoring function should distinguish the real binding modes from other binding modes**
  - force fields
  - empirical free energy functions
  - knowledge based functions
- **scoring can be used together with the conformation search method or only for ranking the search results**
- **scoring functions are the most critical issue of docking**



# Ligand Preparation in DS

## A. Manually

## B. Ligand preparation protocol

- Charges are standardized for common groups
- The largest fragment is kept
- Hydrogens are added
- The molecule is represented in Kekule form
- The ionization states may be enumerated
- Tautomers may be generated
- Isomers may be generated
- Duplicates may be removed
- *3D coordinates may be calculated (not at CSC)*



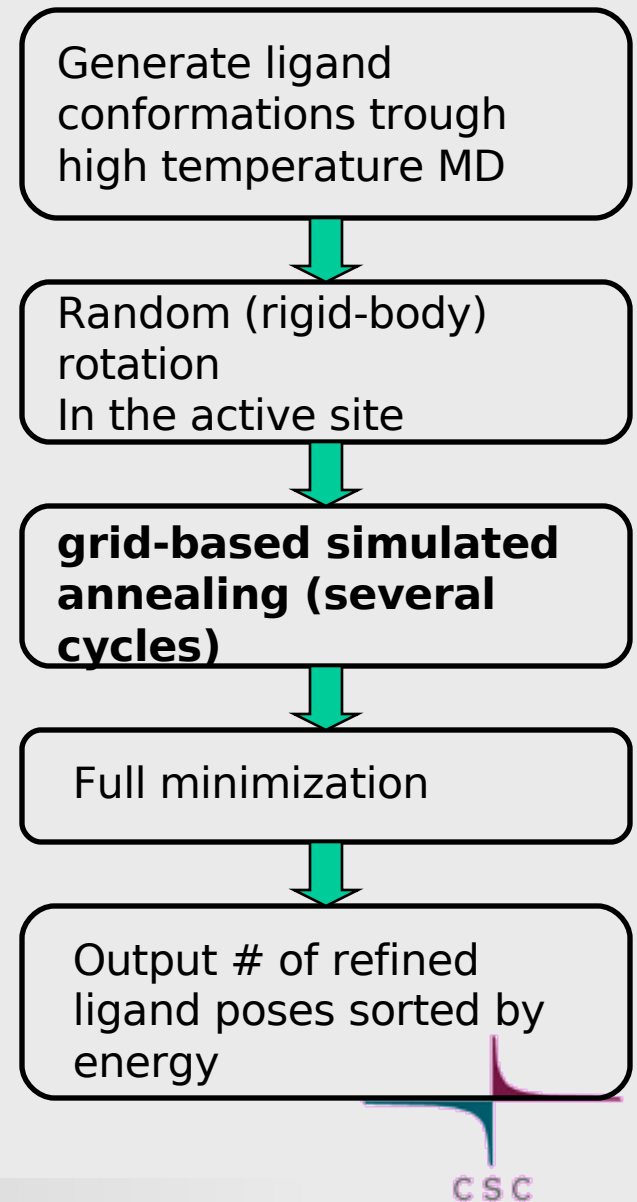
# Protein Preparation in DS

- **Protein health tool**
  - check your structure
- **Protein report and Build and edit protein**
  - fix your force field
- **Force Field tool**
  - set up the force field
  - Check the automation level from the protocols from Edit/Preferences!
- **Define and Edit Binding Site tool**
  - look for cavities
  - Define the binding site sphere using a cavity or specific site



# CDOCKER

- **CHARMm force field based docking tool**
- **Uses soft-core potentials**
- **Conformation search using simulated annealing**
- **Grid based energy evaluation**
- **Force field based scoring**



# After CDOCKER

- **You can**
- **Rescore the structures using protocol:**
  - Calculate Binding Energies
- **Optimize binding site with the ligand using protocol:**
  - Ligand Minimization
- **Study the results with protocol:**
  - Analyze Ligand Poses



# CDOCKER: Papers to read

Wu G, Robertson DH, Brooks CL 3rd, Vieth M.

**Detailed analysis of grid-based molecular docking:  
A case study of CDOCKER-A CHARMM-based MD docking algorithm.**

J Comput Chem. 2003 Oct;24(13):1549-62.

Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M.

**Lessons in molecular recognition:  
the effects of ligand and protein flexibility on molecular docking accuracy.**

J Med Chem. 2004 Jan 1;47(1):45-55.

Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL 3rd.

**Assessing scoring functions for protein-ligand interactions.**

J Med Chem. 2004 Jun 3;47(12):3032-47.

