THE UNIVERSITY of TEXAS
HEALTH SCIENCE CENTER AT HOUSTON
SCHOOL of HEALTH INFORMATION SCIENCES

Modeling Methods: An Overview

For students of HI 6327 “Biomolecular Modeling”

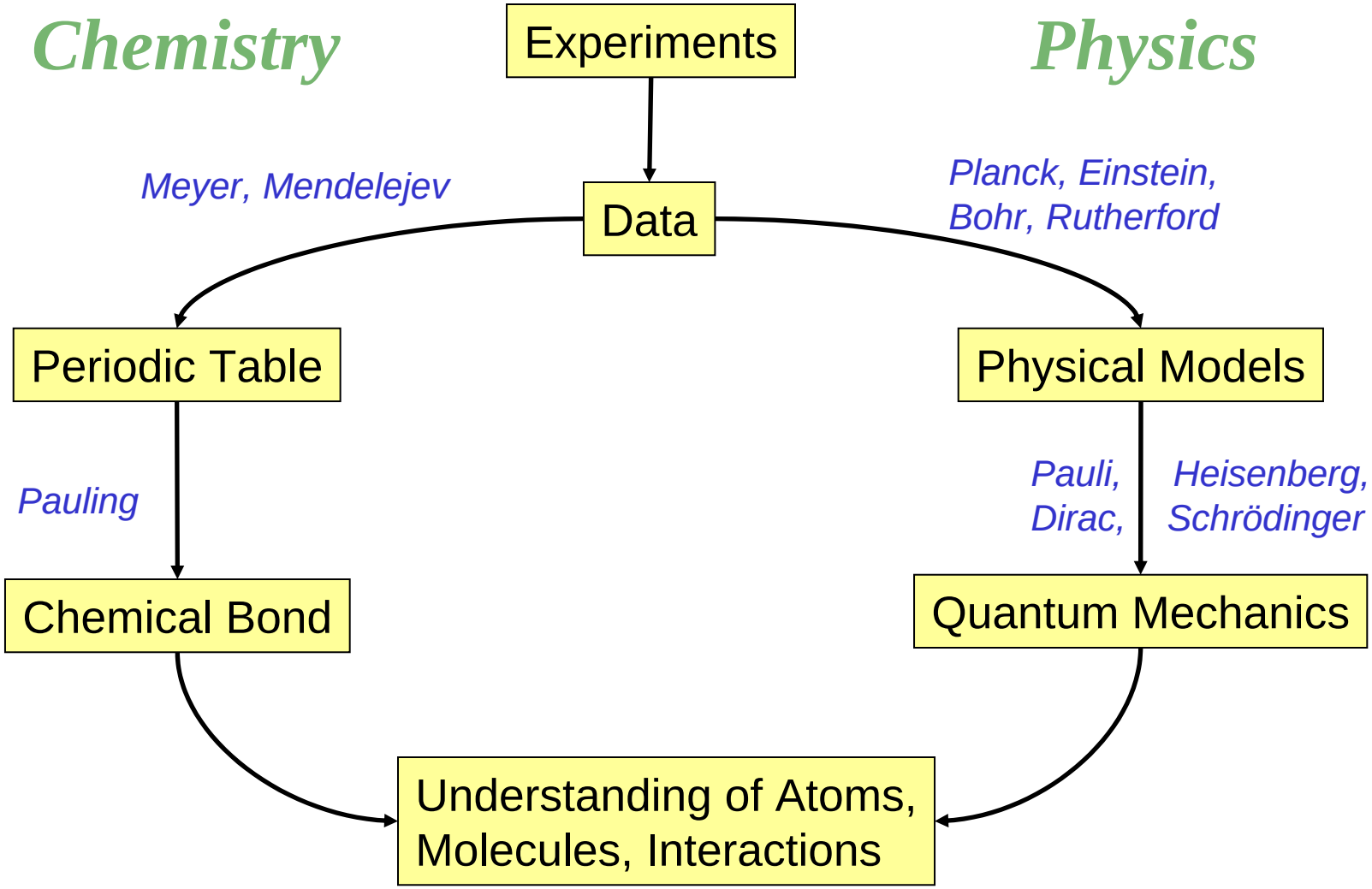
Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/modeling/01.html

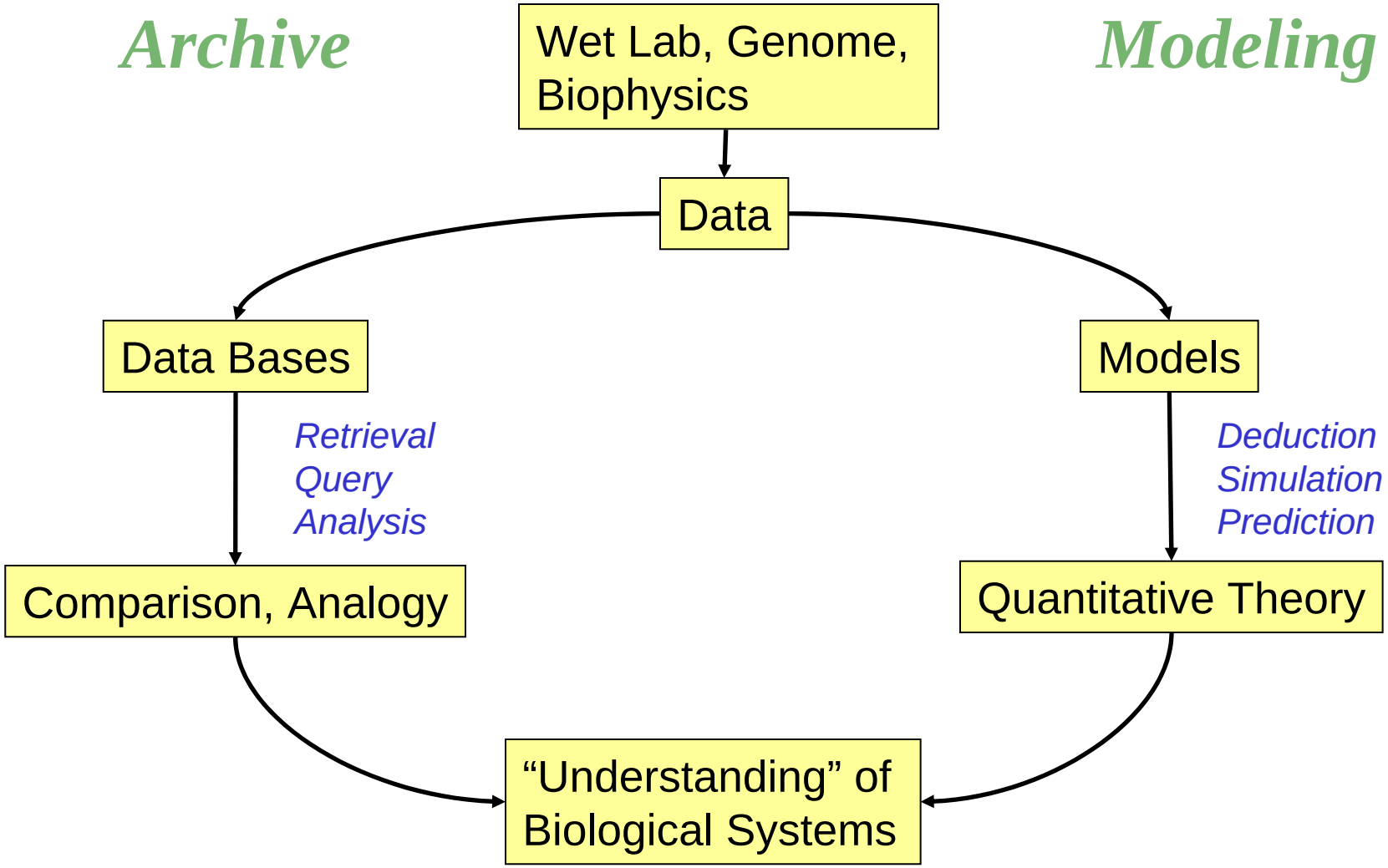
Objectives

- Overview of existing theories, models, and methods in biomolecular modeling
- Encourage further reading
- Create familiarity with methods and discuss pros and cons for homework paper

Structure of Matter: History



Structural Biology Today



Predictive Power: 70-90%

Fundamental Physical Theories

- Quantum Mechanics
- Classical Mechanics
- Electrodynamics / Theory of Relativity
- Thermodynamics / Statistical Mechanics

Levels of Detail

- Atomic/Bond: Catalytic groups of biomolecules that are involved in a chemical reaction. Goal: Understanding the chemical reaction
- Molecular: Dynamics of particular biomolecules as a whole. Goal: Understanding the dynamical behavior
- Supermolecular: Interaction between particular biomolecules (docking, assembly): Goal: Understanding structure/function
- Molecular Networks: Behavior of signaling and reaction networks in the cell and between cells (metabolism, systems biology).

What is a Model?

•a generalized, hypothetical description used in analyzing or explaining a system

• a simplified representation of a system intended to enhance our ability to understand, predict, and control the behavior of the system

• a construct by which situations, real or hypothetical, can be represented

• models should promote understanding of the phenomenon being modeled

Model Building

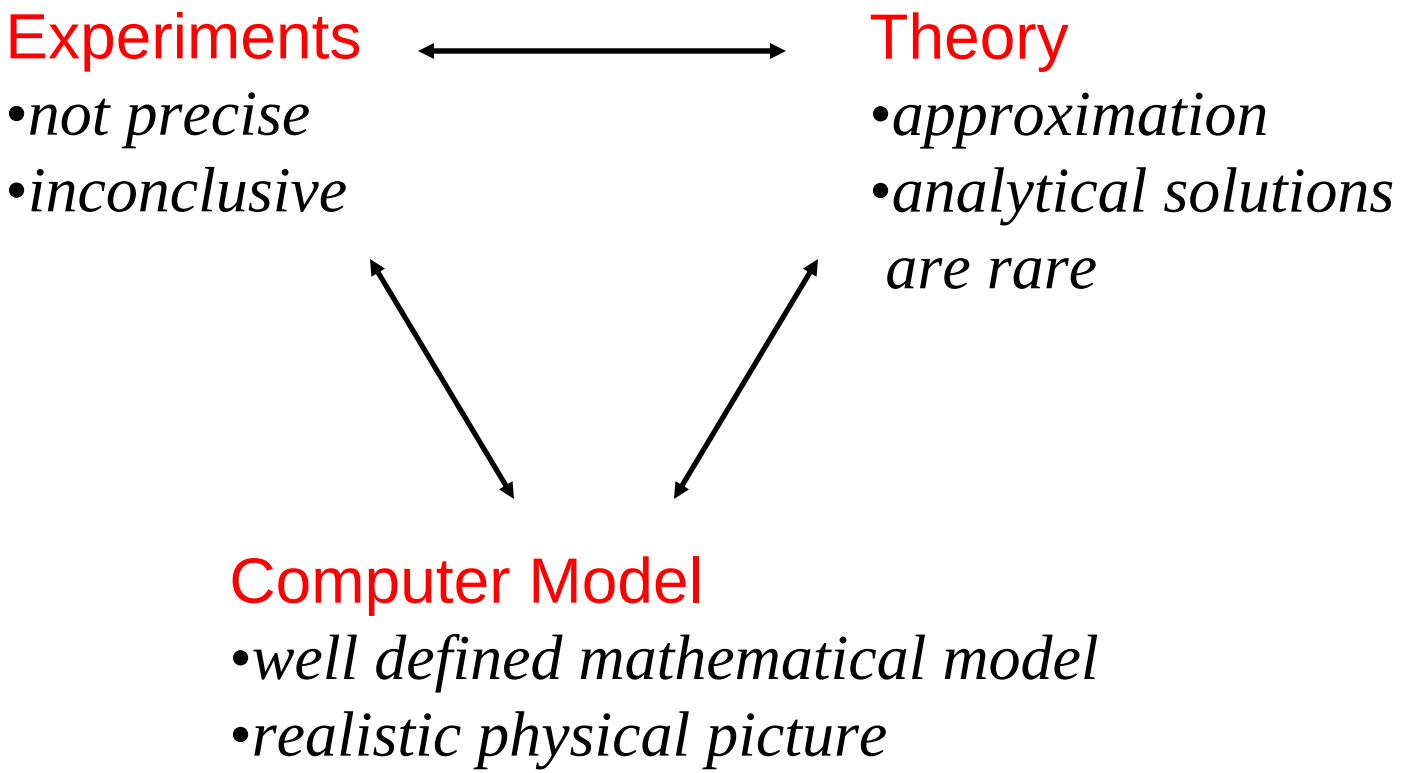
How:

- Reproduction - reproduces some aspects of the original
- Idealization - does not represent all aspects of the original
- Subjectivity - was made for a special purpose

Why:

- can explain the experimental data
- can make predictions
- be as simple as possible, not simpler
- be as complicated as necessary, not more complicated

Why Computer Models?



A Computer Model…

- captures the essence of your system in a simulation model
- enables to test hypotheses at a fraction of the cost of actually undertaking the activities which the models simulate
- begin with a simple approximation of a process and gradually refine the model as your understanding of the process improves (step-wise refinement)
- enables to achieve good approximations of very complex problems
- predict the course and results of certain actions
- understand why observed events occur
- explore the effects of modifications
- confirm that all relevant variables are known
- evaluate ideas and identify problems
- communicate the integrity and feasibility of your plans

Short History of Molecular Simulations

- Metropolis, Rosenbluth, Teller (1953) Monte Carlo Simulation of hard disks.
- Fermi, Pasta Ulam (1954) experiment on ergodicity
- Alder & Wainwright (1958) liquid-solid transition in hard spheres. “long time tails” (1970)
- Vineyard (1960) Radiation damage using MD
- Rahman (1964) liquid argon, water(1971)
- Verlet (1967) Correlation functions, ...
- Andersen, Rahman, Parrinello (1980) constant pressure MD
- Nose, Hoover, (1983) constant temperature thermostats.
- Car, Parrinello (1985) ab initio MD.

"The general theory of quantum mechanics is now almost complete. The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

Dirac, 1929

Static Models

- describe a system mathematically, in terms of equations, where the potential effect of each alternative is ascertained by a single computation of the equation.
- ignore time-based variances.

Dynamic Models

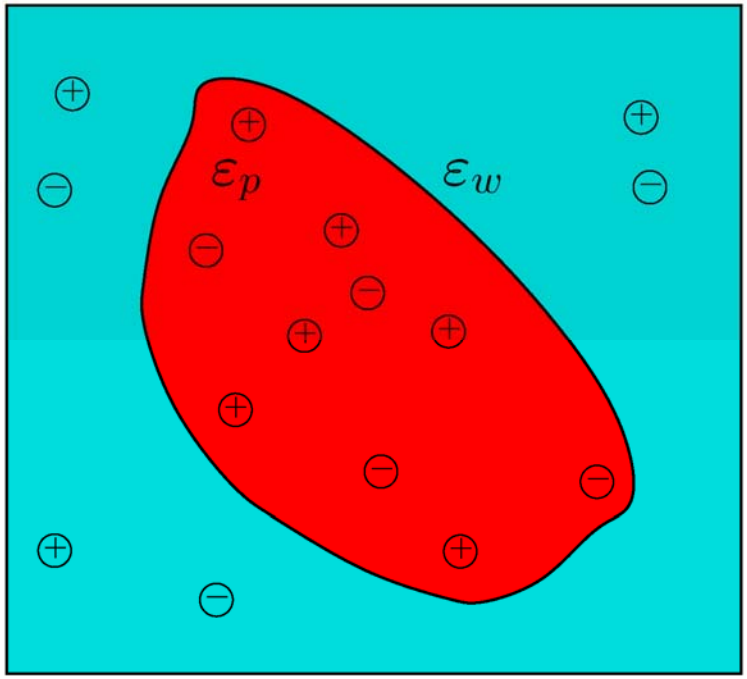
- a.k.a. “simulations”
- representation of the dynamic or time-based behavior of a system (iterative)
- can predict the outcomes of possible courses of action and can account for the effects of variances or randomness
- can predict effects of random events

Example (I): Continuum Electrostatics

Conceptual Model:

Protein: Low dielectric region
With fixed partial charges

Solvent: High dielectric region
with unlocalized (mobile) charges



A continuum electrostatic model describes molecules at atomic detail using a macroscopic description.

Example (I): Continuum Electrostatics

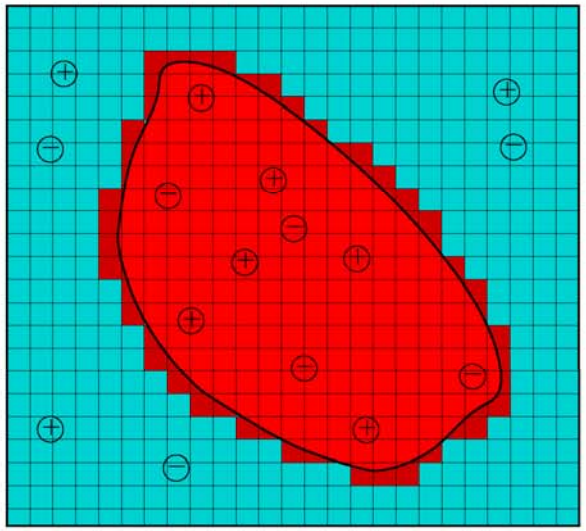
Mathematical Model: Poisson-Boltzmann Equation

$$\nabla\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}) = -4\pi\left(\rho(\mathbf{r}) + \sum_{i=1}^{K} c_i Z_i e_0 \exp\left(\frac{-Z_i e_0 \phi(\mathbf{r})}{RT}\right)\right)$$

Numerical Solution: $\longrightarrow \phi(\mathbf{r})$

(Finite Difference)

Conceptual Model:



Electrostatic Energy : $G^{elec} = \sum_{i=1}^{N} q_i \phi(\mathbf{r})$

- relative binding energies
- solvation energies
- transfer energies

Pros/Cons: Continuum Electrostatics

Pros:

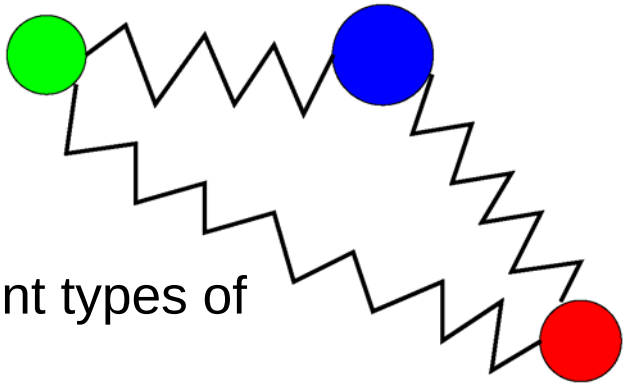
- simple model that describes certain aspects of biomolecules very well
- computationally fast
- analytical solutions for special cases

Cons:

- limited conformational flexibility (modification of model)
- static, not dynamic, i.e., no time dependence

Example (II): Molecular Mechanics

•A molecule is described by interacting (soft) spheres.



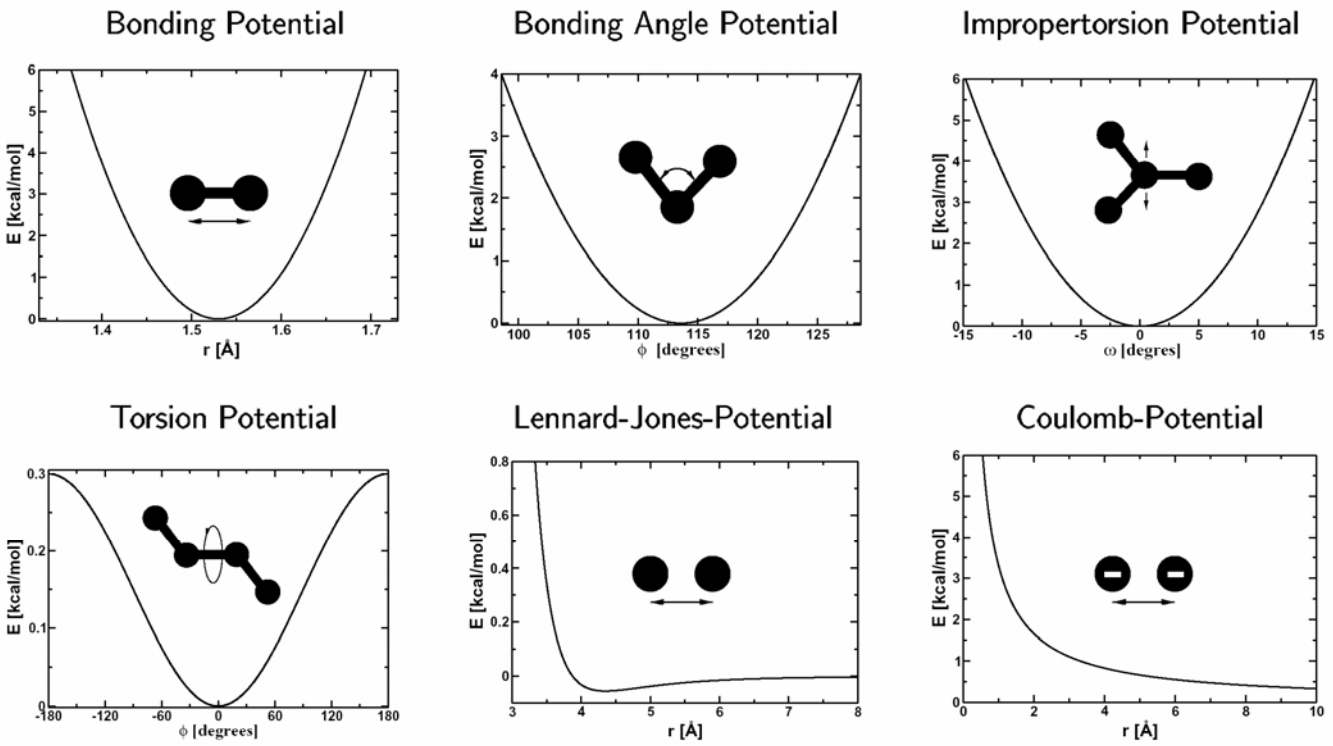
•Different types of spheres describe different types of atoms.

•The interaction between chemically bound atoms is described by special bonding interaction terms.

•The interaction of not chemically bound atoms is described by non-bonding interaction terms.

•The motion of all the atoms in the molecule is described by Newtonian classical mechanics.

Molecular Mechanics Force Field

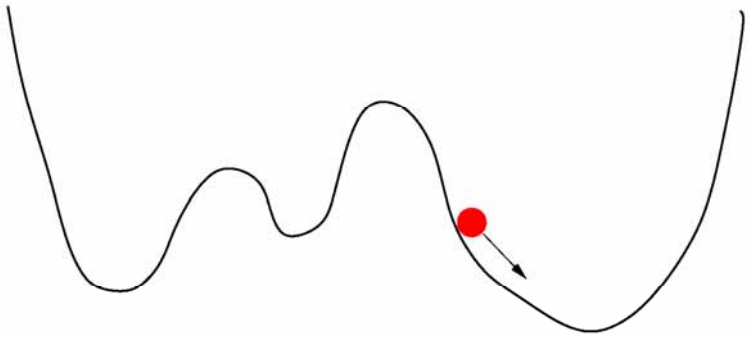


$$E = \sum_b k_b(r - r_0)^2 + \sum_\theta k_\theta(\theta - \theta_0)^2 + \sum_\omega k_\omega(\omega - \omega_0)^2$$

$$+ \sum_\phi k_\phi(1 - \cos(n\phi - \delta)) + \sum_{i<j} 4\,\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] + \sum_{i<j}\frac{q_i q_j}{r}$$

Optimization by Energy Minimization

Goal: finding low energy conformations, i.e., most probable conformations

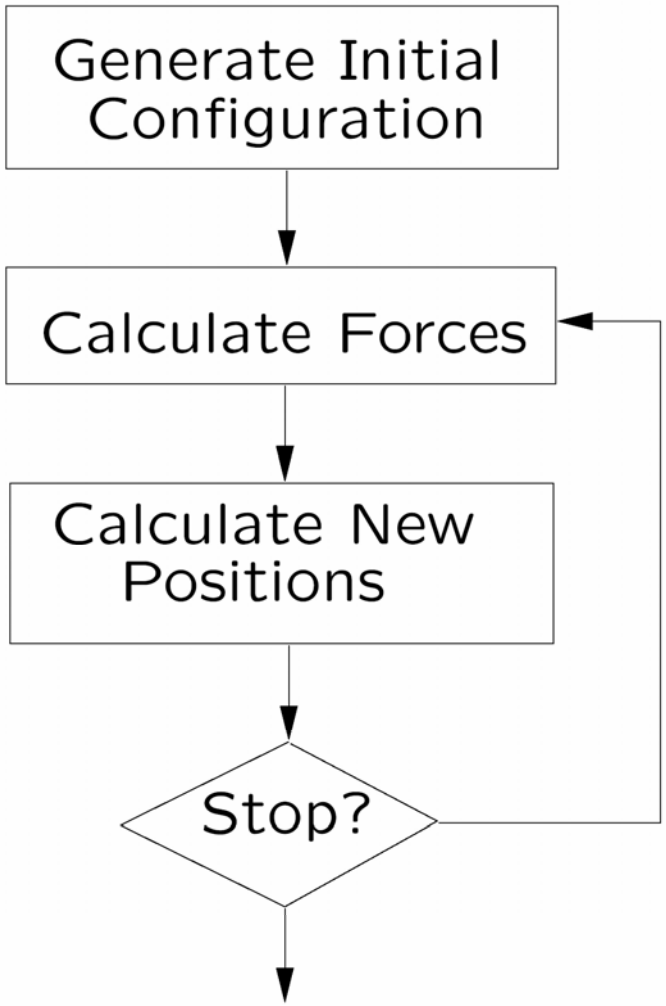


Various numerical optimization procedures are applied to find these minima.

But:

- only local minima are found
- only minimum potential energy, not minimum free energy

Molecular Dynamics



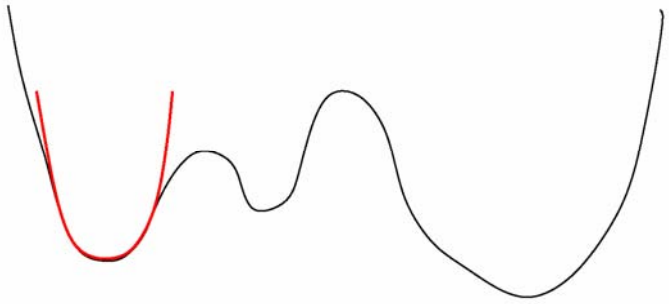
Newtons Equation of Motion

$$\frac{d^2\vec{r}_i(t)}{dt^2} = m_i^{-1}\vec{F}_i(\vec{r}_i) = -m_i^{-1}\nabla E(\vec{r}_i)$$

- Several algorithms exist to compute Molecular Dynamics trajectories.
- Trajectories describe the motion of a particular molecule with a defined starting configuration in time.
- It is assumed that the motion can be described purely classically and quantum effects do not play a role.

Normal Mode Analysis

Approximate the complex energy landscape by harmonic potentials.



Pros:

- good local approximation
- yields large scale, correlated motions
- thermodynamic theory (entropy, enthalpy, free energy)

Cons:

- no transitions between states
- forced orthogonalization of modes
- what do modes mean (are all functionally relevant)?

Brownian Dynamics

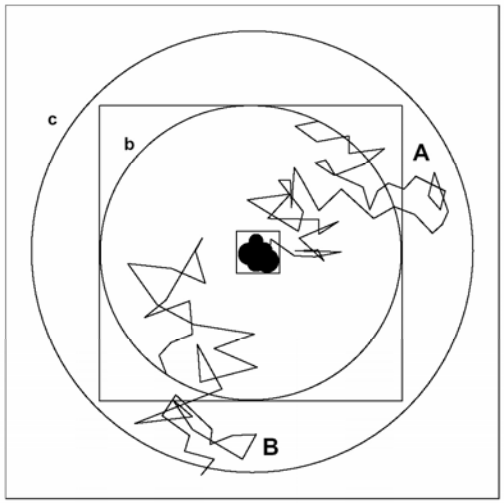
Brownian Dynamics = Newtonian Dynamics + Random Terms

$$\frac{d^2\vec{r}_i(t)}{dt^2} = m_i^{-1}\vec{F}_i + m_i^{-1}\vec{R}_i - \beta_i\frac{d\vec{r}(t)}{dt}$$

In Biomolecular Simulations:

- Diffusion of Macromolecules
- Simulation of Association Processes

Molecules are treated as rigid or only semirigid macroscopic objects.



Monte Carlo

Simulation involves random changes of the system (Name: Monte Carlo in Monaco)

Pros:

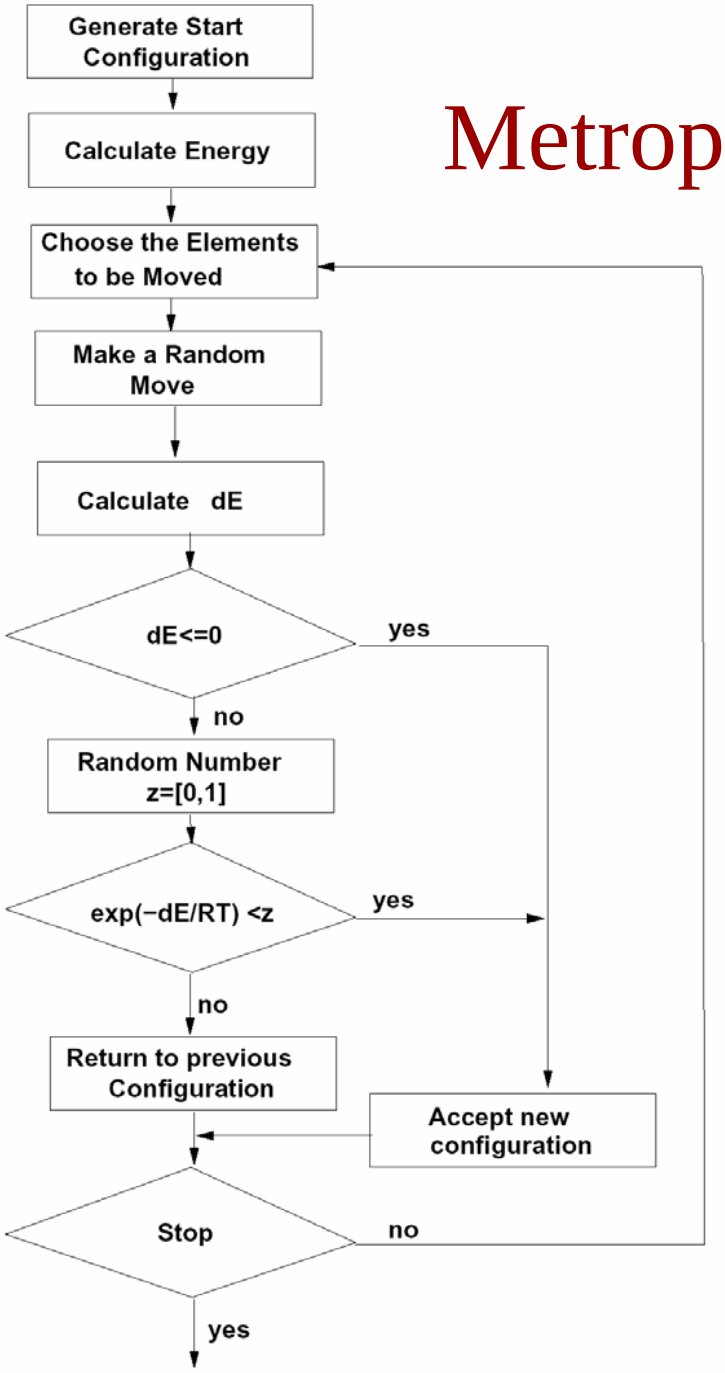
– does not require a continuous energy function (as in MD)

– number of particles can easily vary (very hard in MD)

Cons:

– highly correlated movements are hard to simulate, leads to a poor sampling of large-scale changes

Metropolis Monte Carlo



- generate well-defined thermodynamic ensembles
- distribution of states according to Boltzmann:

exp(-dE/RT)

- information about equilibrium states
- easy to implement

Pros and Cons: Molecular Mechanics

Pros:

- detailed stereochemical model that describes certain aspects of biomolecules very well
- conformational flexibility
- dynamic model (time dependence) is possible

Cons:

- computationally demanding
- large scale conformational changes are hard to model
- no electronic (quantum) desciption, no chemical reaction (bond breaking/forming), no excited states, …

Example (III): Quantum Chemistry

Describes molecules based on (approximate) quantum mechanical methods. Particle – wave duality.

- physical model: electron waves in the field of the nuclei
- results of quantum chemical calculations:
  - – geometries, structures, transition states
  - – energies: reaction energies, equilibrium constants, activation energies
  - – electronic structures: spectra, reactivity, excites states
- quantum effects (like tunneling) are included in this picture

Example (III): Quantum Chemistry

- Ab initio methods: treat molecules from first physical principles (only very few assumptions) for instance Hartree Fock, MP2, Configuration Interaction
- Semiempirical methods: use empirical parameter to compensate for neglected terms. for instance AM1, Extended Hückel theory
- Density Functional Theory (DFT): The electron density of any electronic system uniquely determines the system. The real electron density is the lowest energy density possible. Minimize “density functional” and obtain electronic structure and the correct equilibrum structures.

But: Real density functional not known: therefore all DFT methods use and approximate (guessed) functional.

Pros and Cons: Quantum Chemistry

Pros:

- detailed picture of a molecule (electrons etc.)
- electronic description, chemical reactions, excited states
- limited conformational flexibility (relaxation)

Cons:

- computationally very demanding (ab initio > DFT > semiempirical)
- static model (no time)
- only small systems: 10-1000 atoms

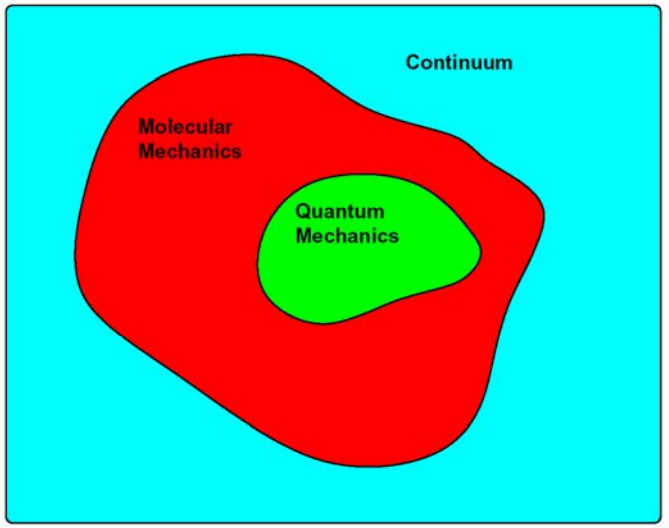
Combined QM/MM

Different parts of the system are treated by different approximations.

QM - Catalytic Site

MM - Protein Environment

CE - Solvent



Also only QM/MM, MM/CE, QM/CE or QM (DFT/semiempiric)

Pro: Relevant parts of the system can be treated more exact.

Contra: Boundaries are hard to treat.

Example (IV): Chemical Reactions

- living cells turn substances over
- not only "structural" biochemistry, but also dynamical biochemistry, i.e. biochemical reactions!!
- the kinetic of chemical reactions can be described by differential equations: Unimolecular Reaction

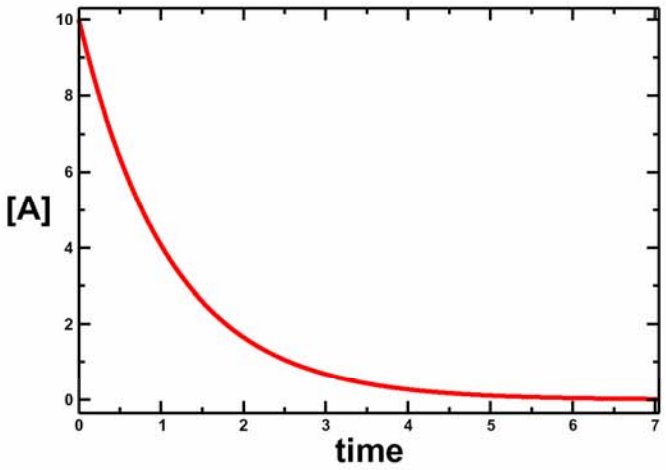
$$\frac{d[\mathrm{A}]}{dt} = -k[\mathrm{A}] \quad \longrightarrow \quad [\mathrm{A}](t) = [\mathrm{A}]_0 \; e^{-kt}$$

- more complex reaction are described by more complex differential equation

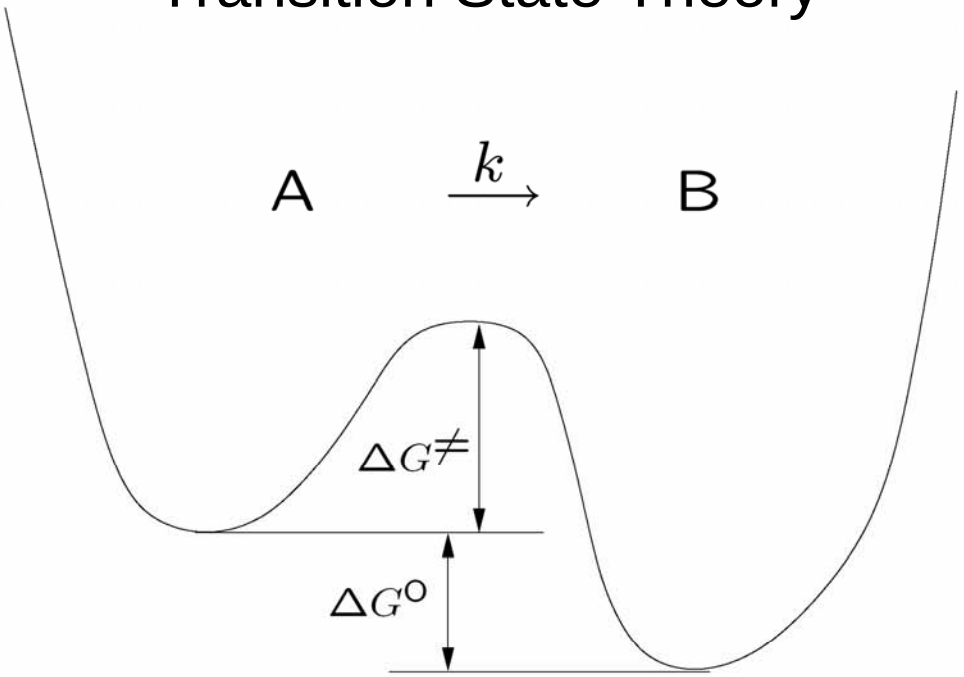
Example (IV): Chemical Reactions

$$\frac{d[\mathrm{A}]}{dt} = -k[\mathrm{A}]$$

$$[\mathrm{A}](t) = [\mathrm{A}]_0 \, e^{-kt}$$



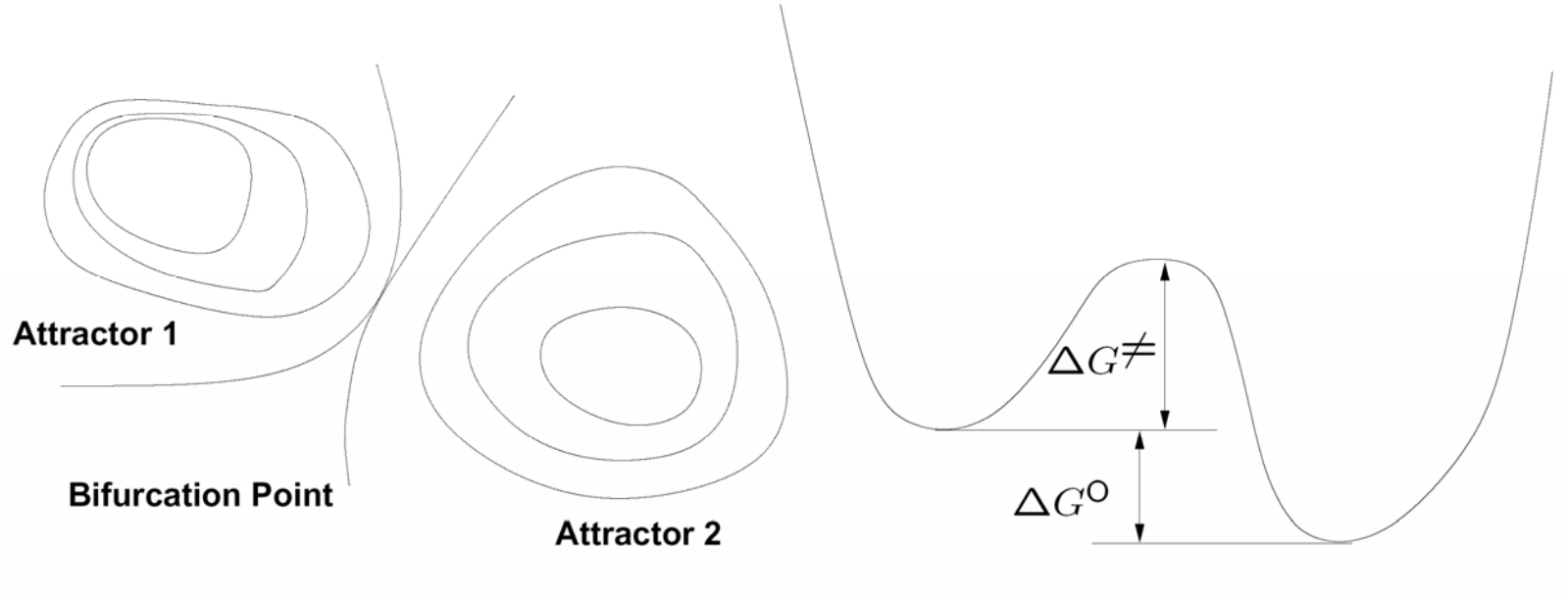
Transition State Theory



$$k \propto e^{-\frac{\Delta G^{\neq}}{RT}}$$

Example (IV): Chemical Reactions

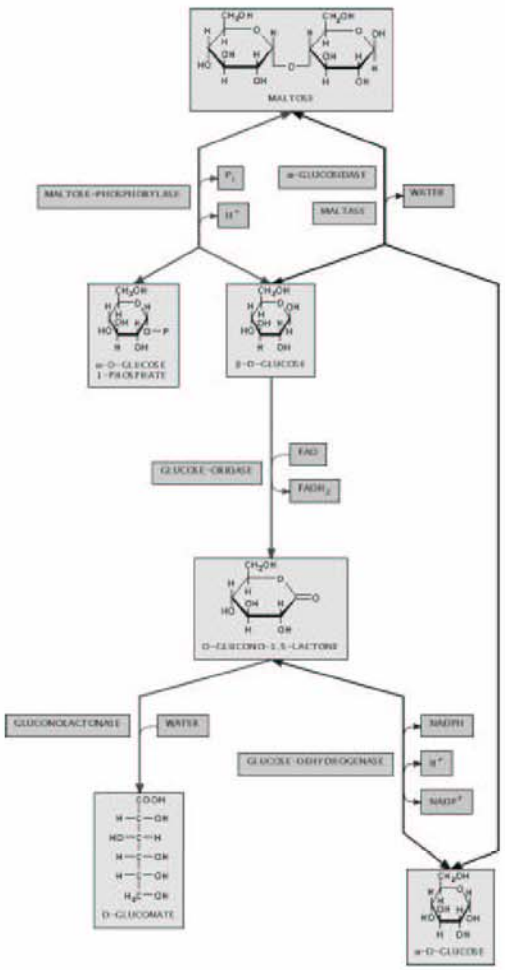
Transition State Search



Several optimization methods have been adapted to find the lowest energy path.

Other methods use Molecular Dynamics techniques.

Example (IV): Chemical Reactions



- biochemical reactions span a complicated network
- can be described as a system of coupled differential equations

$$\begin{pmatrix} \frac{d[A_1]}{dt} \\ \frac{d[A_2]}{dt} \\ \vdots \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & \cdots \\ k_{21} & k_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} [A_1] \\ [A_2] \\ \vdots \end{pmatrix}$$

- parameter experimentally measured or also calculated

Systems Biology

Conclusions

- Models allow to describe certain aspect of a system, not all. They can describe phenomena and make predictions.
- Computer models are useful to test hypotheses and to check if the result of the model agrees with reality.
- Different problems required different levels of theory.
- Only a combination of different techniques allow to describe the complex phenomena in biological molecules and biochemical networks.

Resources and Further Reading

WWW:
http://cmm.info.nih.gov/intro_simulation

Textbooks:
Schlick, Chapters 1, 7, 8, 11, 12, 13.5
Bourne & Weissig, Chapter 21

Acknowledgement:
G. Matthias Ullmann, Uni Heidelberg