Laboratory of Computational Biophysical Chemistry

From Computational Physics to Structural Bioinformatics

Part 0

BIOINFORMÁTICA USP

STAMiNA Global Network

Fernando Luís BARROSO da Silva

Department of Biomolecular Sciences, FCFRP/USP, Brazil

flbarroso@usp.br

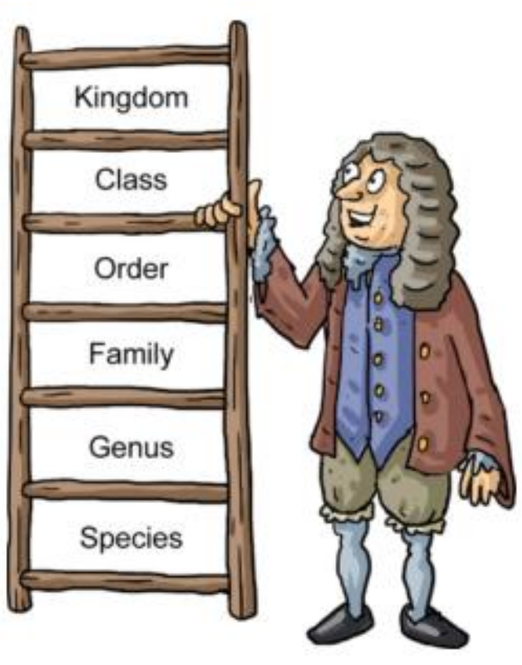
March 09, 2020

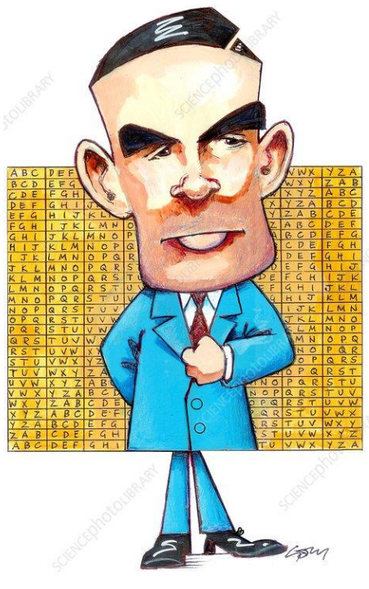
or

Computational physics applied to biological science

or

where Physics, Biology and computer science meet

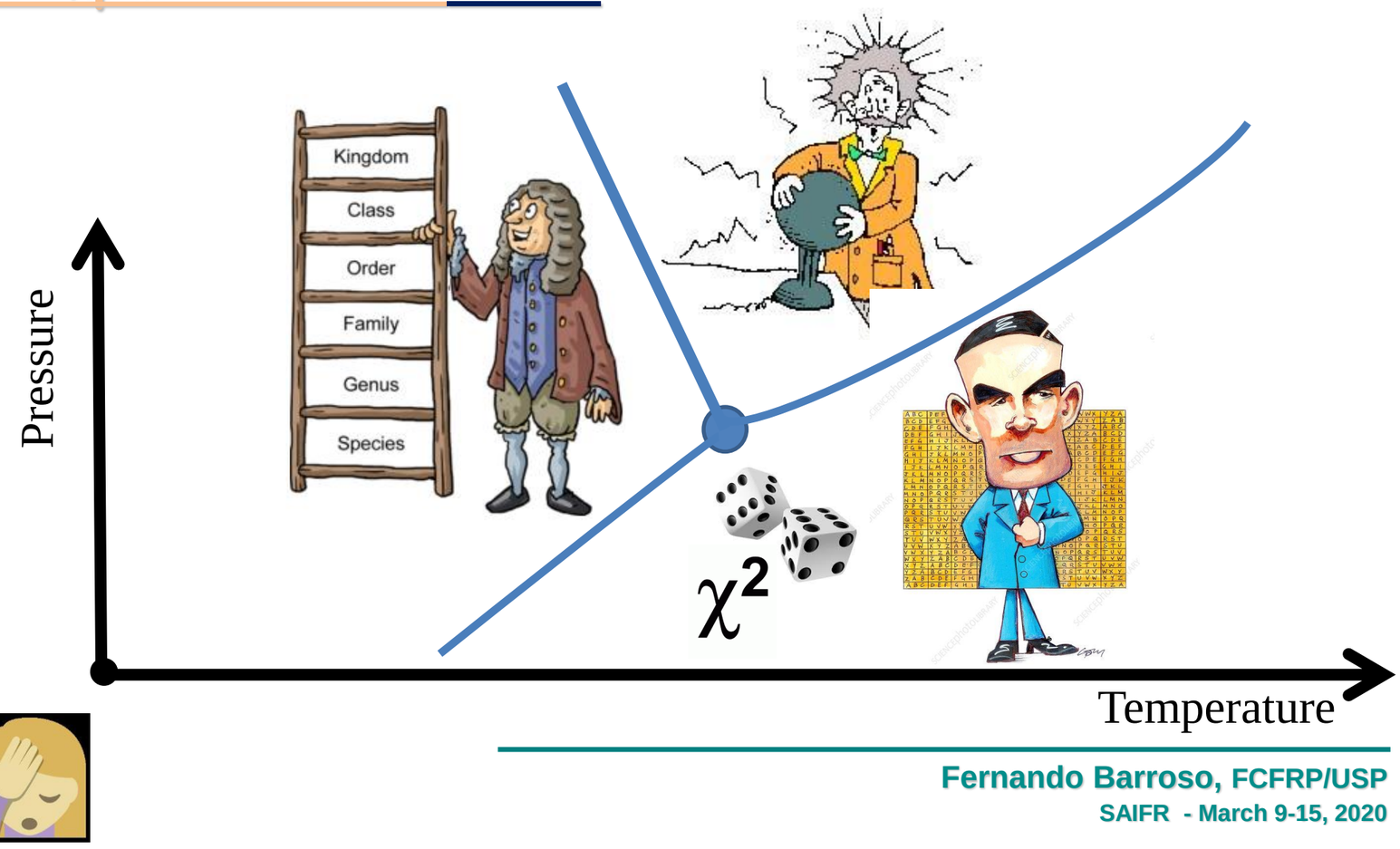




[pictures from internet, unknown author]

where Physics, Biology and computer science meet

Laboratory of Computational Biophysical Chemistry

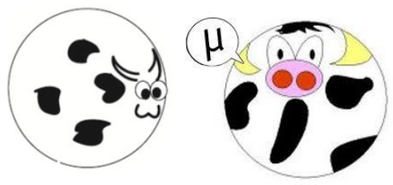


[pictures from internet, unknown author]

Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

3

What we do in our lab



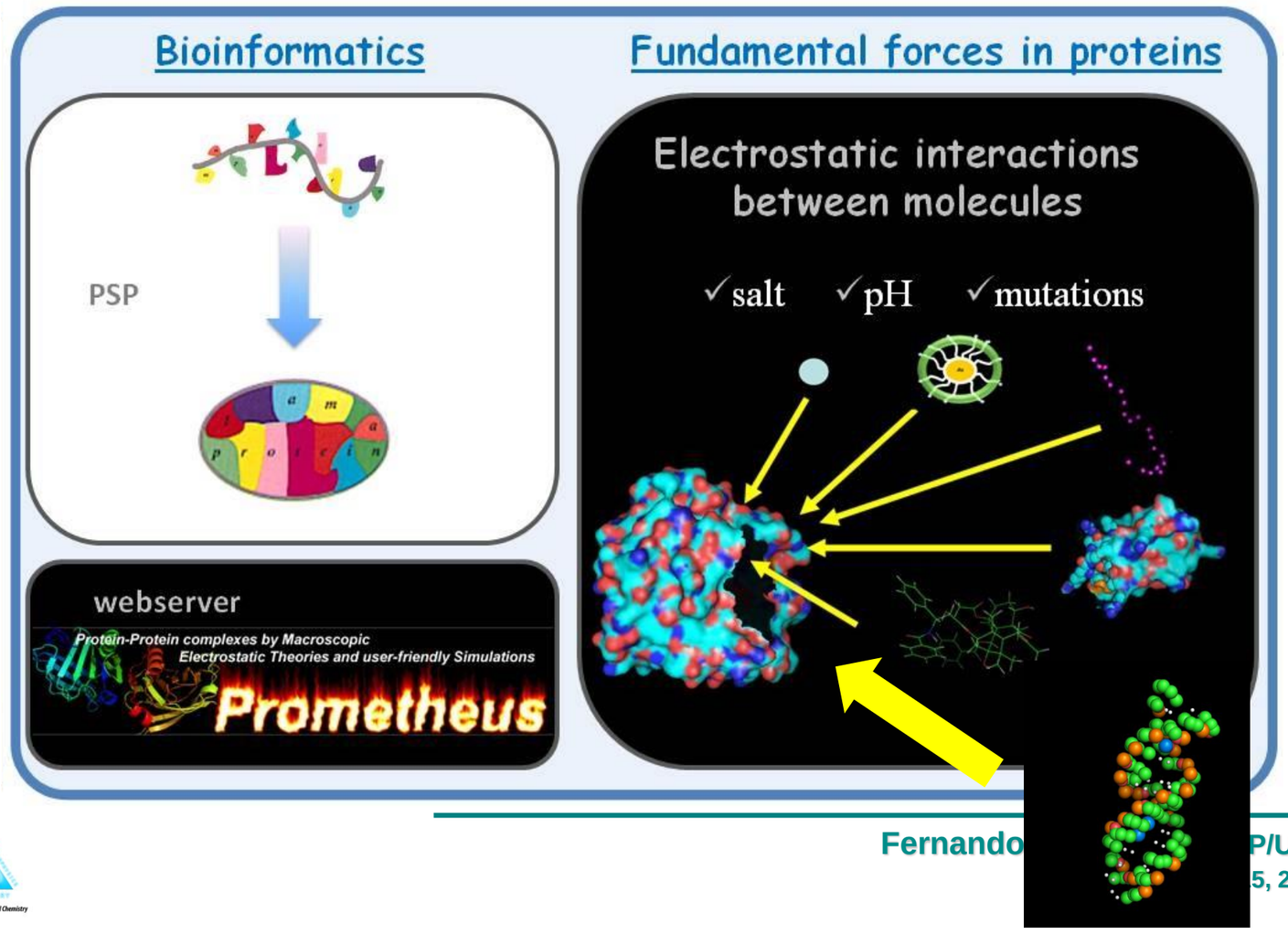
Multi-scale & multi-approach modeling

Laboratory of Computational Biophysical Chemistry

School of Pharmaceutical Sciences at Ribeirão Preto

4

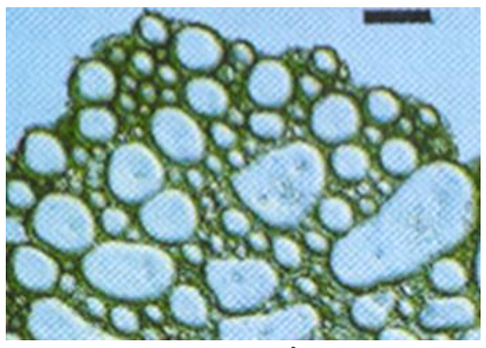
Molecular base to understand diseases, pharmaceuticals, bioseparation processes, new functionalized materials, ...



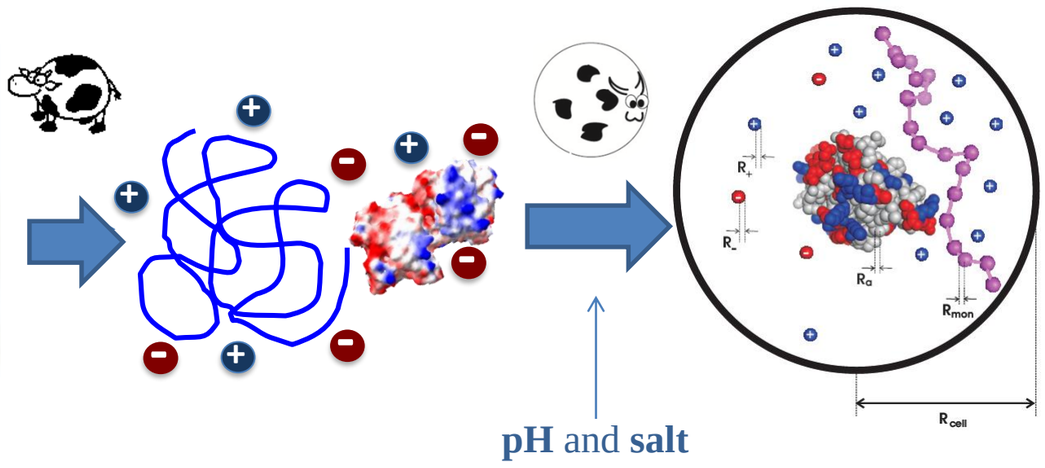
General approach: to rationalize key applied systems

From the **REAL system** to **CG models**

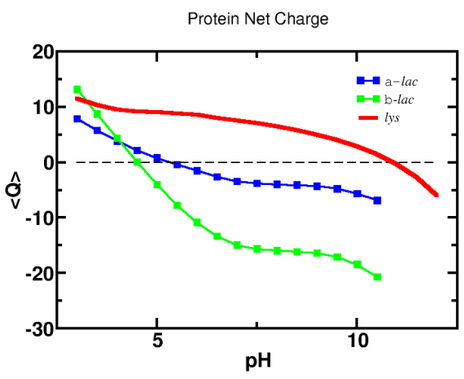
Electrostatic interactions

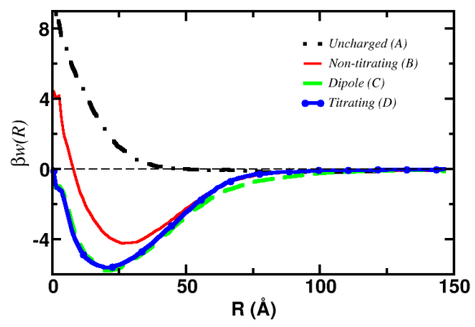


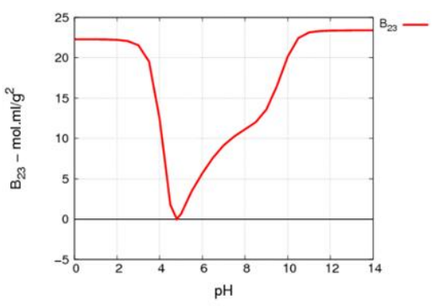
microencapsulation

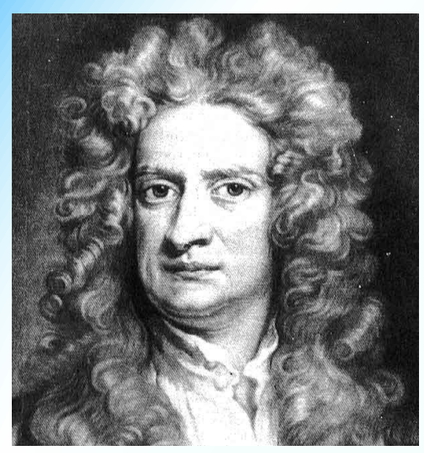


Example of measurements









From Computational Physics to Structural Bioinformatics

1. M.P.Allen e D.J. Tildesley, *Computer Simulation of Liquids*, (Oxford University Press, 1989).
2. D. Frenkel e B. Smid, *Understanding Molecular Simulation*, (Academic Press, 2001).
3. T. Schlick, *Molecular Modeling and Simulation*, (Springer, 2010).
4. J.M. Haille, *Molecular Dynamics Simulation: Elementary Methods*, (Wiley-Interscience, 1997).
5. D.C. Rappaport, *The Art of Molecular Dynamics Simulation*, (Cambridge University Press, 1997).

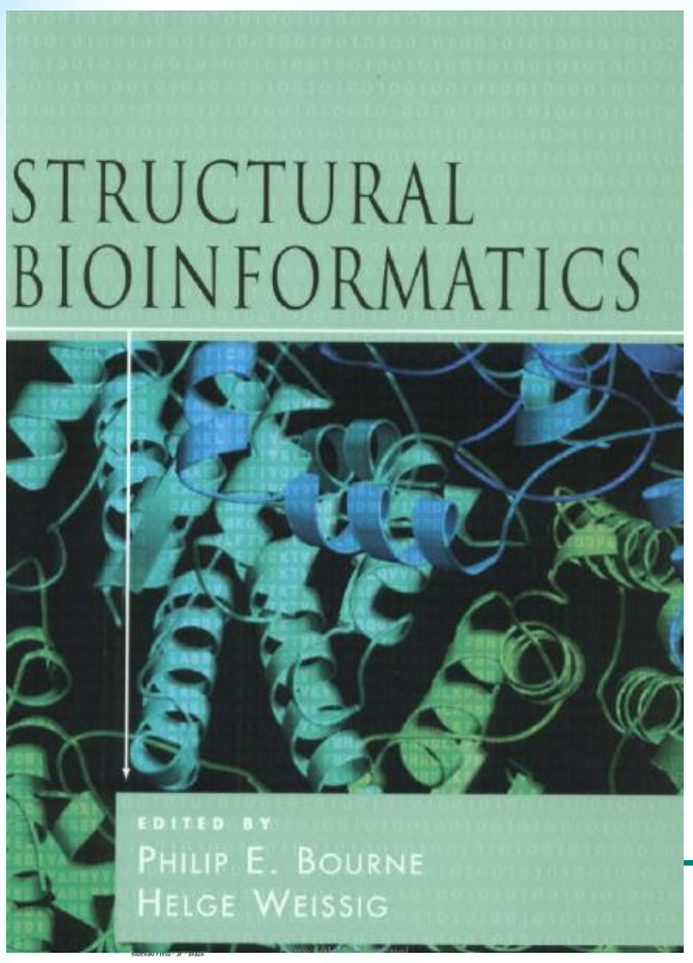
+ specific scientific papers

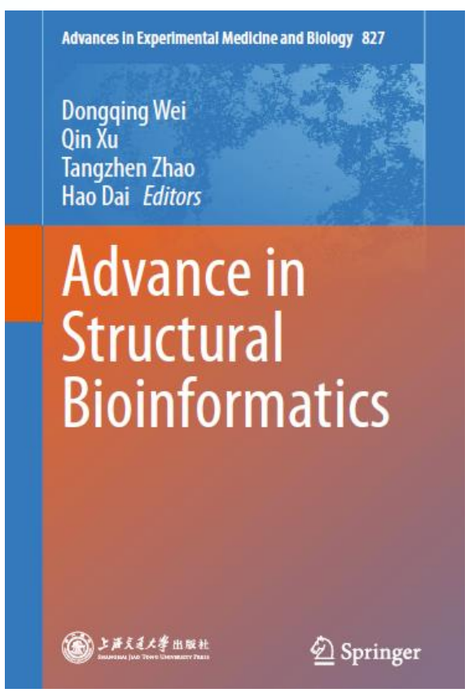
Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

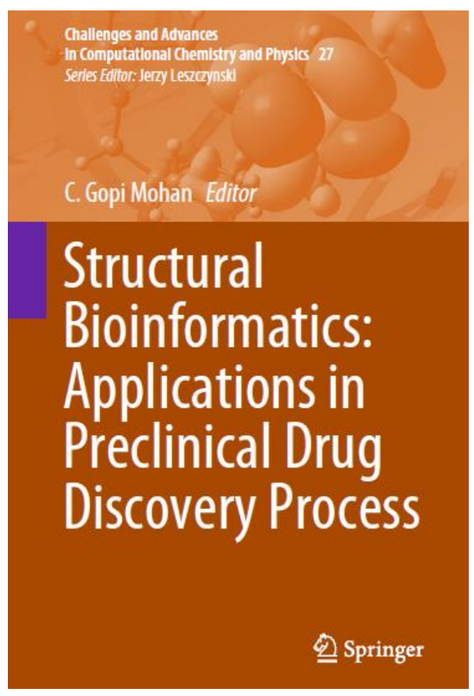
7

Laboratory of Computational Biophysical Chemistry

....Structural Bioinformatics





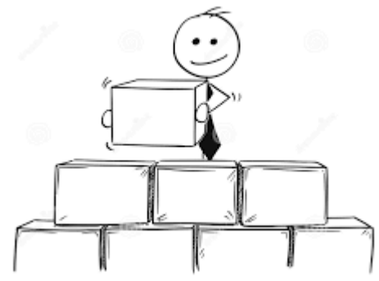


Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

8

Laboratory of Computational Biophysical Chemistry

Syllabus

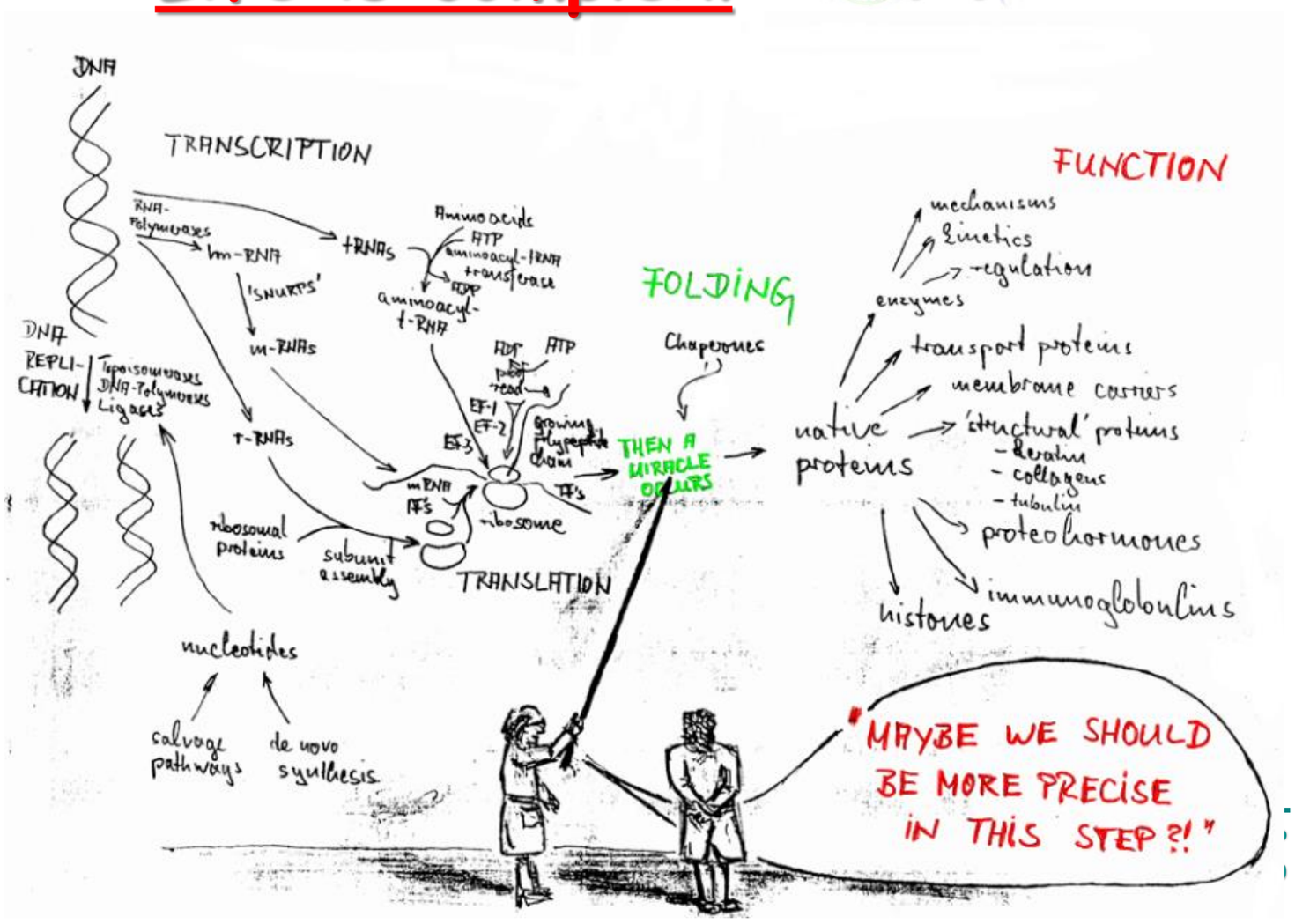


- ✓ **About us**
- ✓ **Bibliography**
- ✓ **From complex problems to key answers**
- ✓ **Early computer experiments**
- ✓ **What is Structural Bioinformatics?**
- ✓ **Molecular modeling**
- ✓ **Solving the model**
- ✓ **Electrostatic interactions & constant-pH simulation methods**

Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

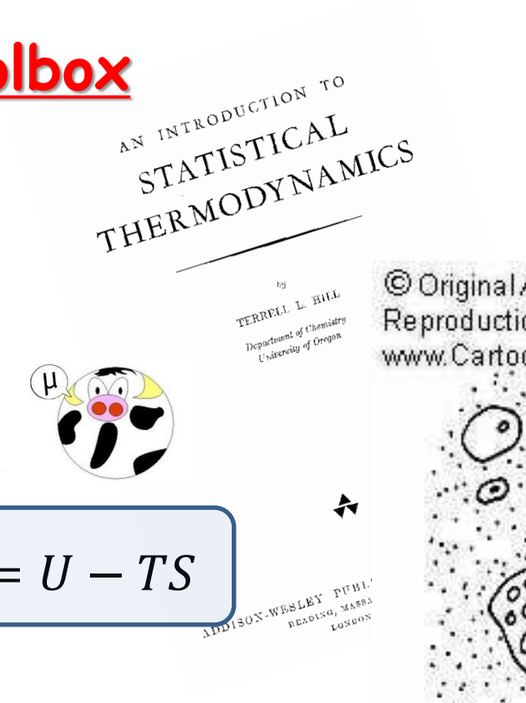
Life is complex!

Laboratory of Computational Biophysical Chemistry

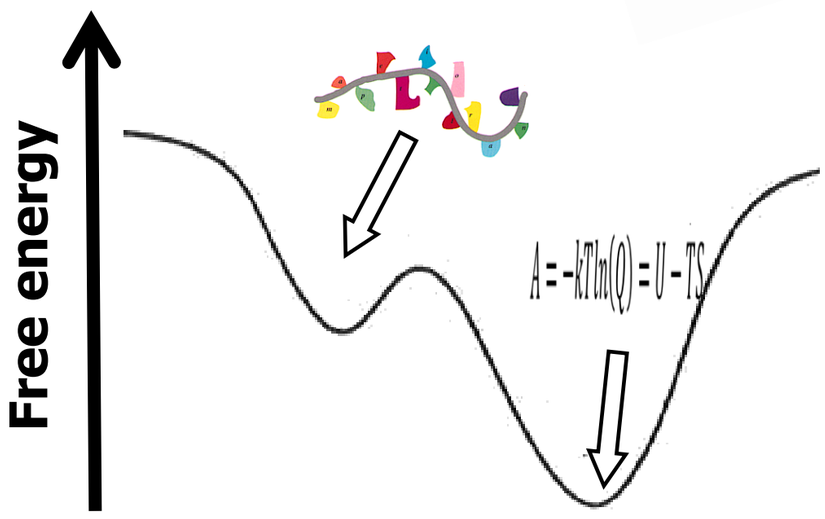


[unknown author]

Physicist´s toolbox

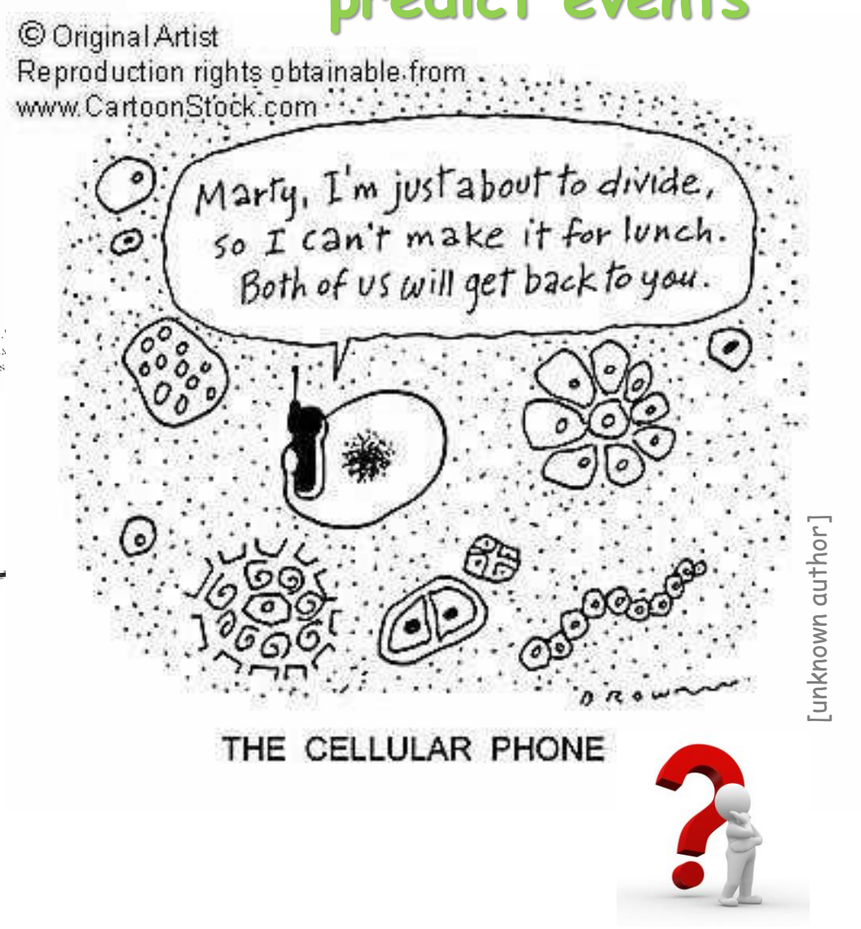


$$A = -kTln(Q) = U - TS$$



11

predict events

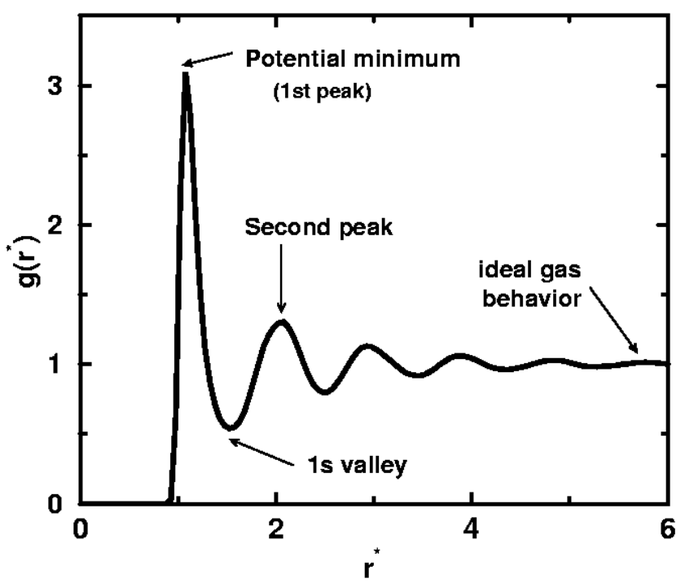


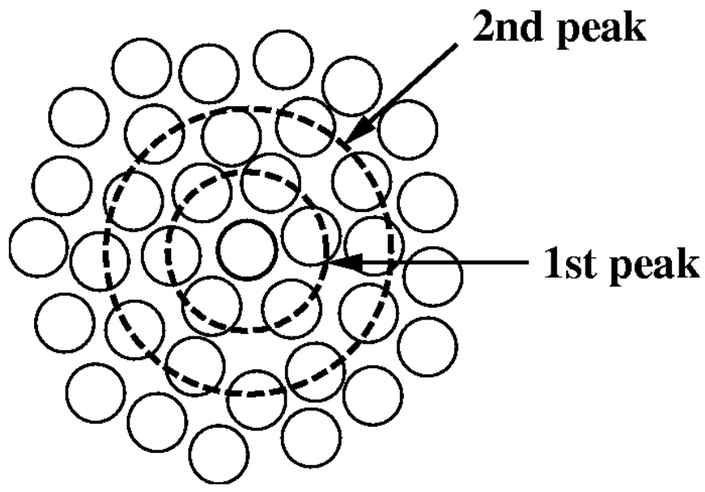
[unknown author]

$$w(r) = -k_B T \ln[\, g(r)]$$

The radial distribution function

$$g(r) = \frac{hist(r)}{\rho\, n_{obs} \Delta V}$$



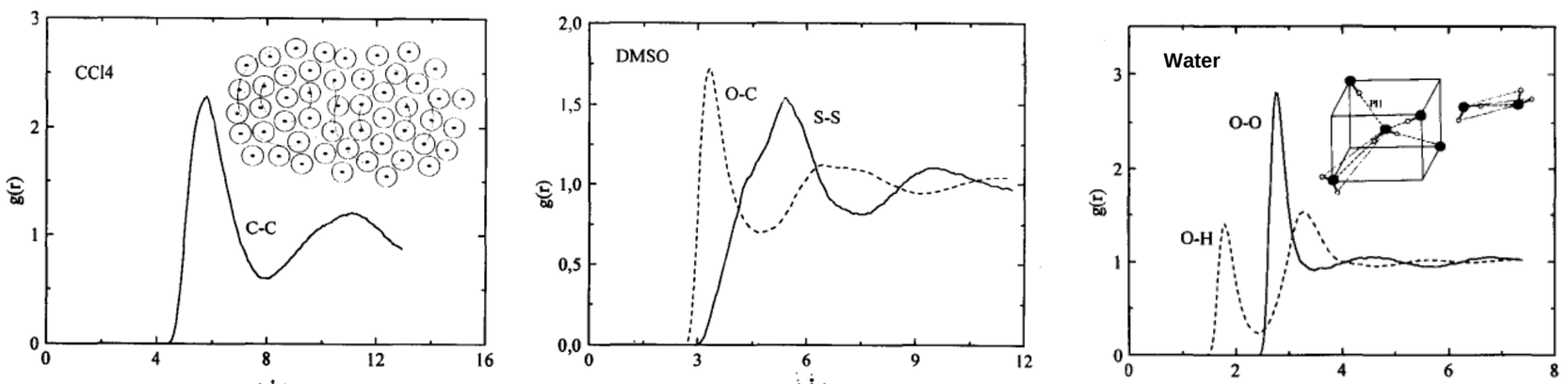


[Barroso da Silva, 2000]

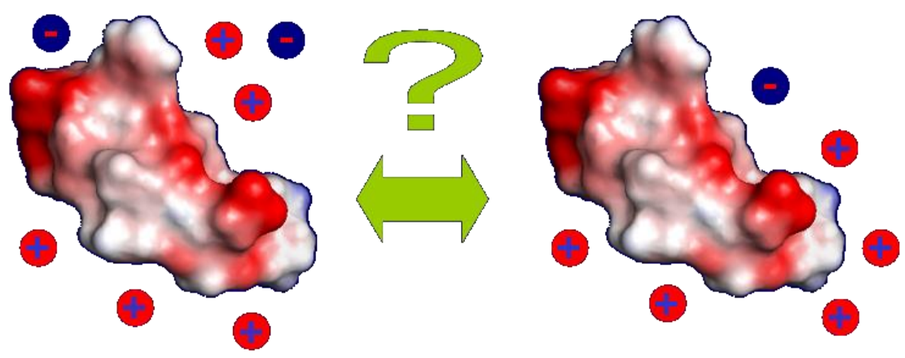
12

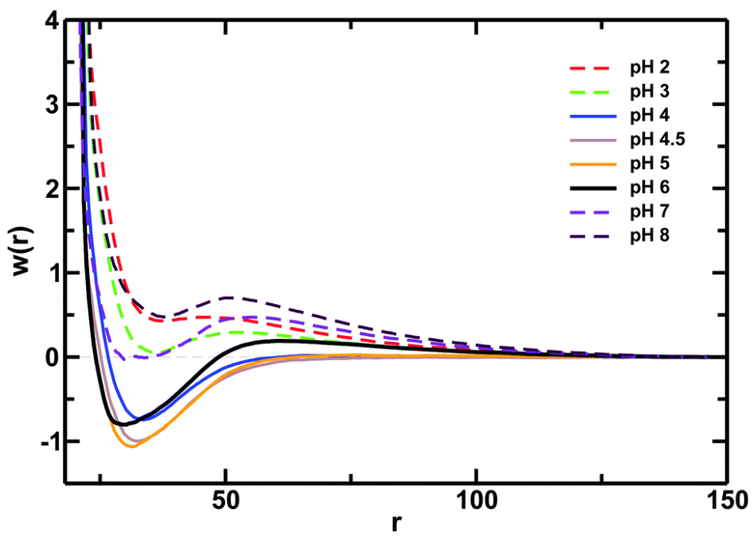
Intermolecular interactions and the radial distribution function

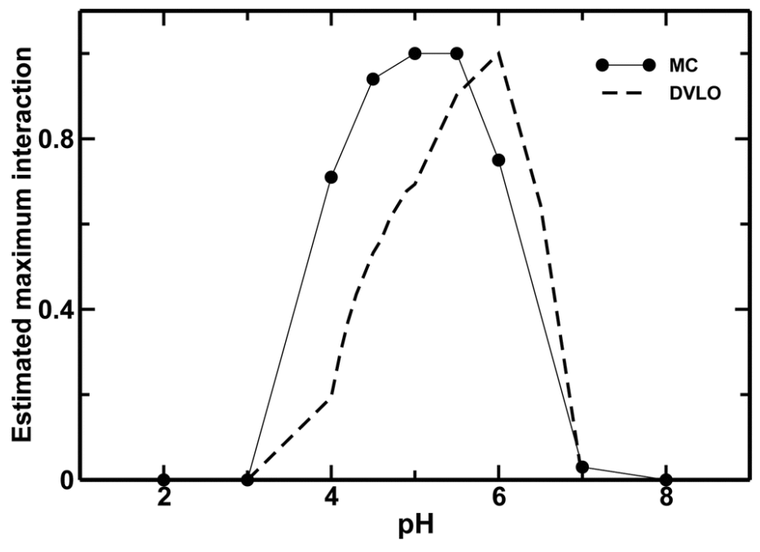
[Skaf & Barroso da Silva, 1994]



intermolecular interactions





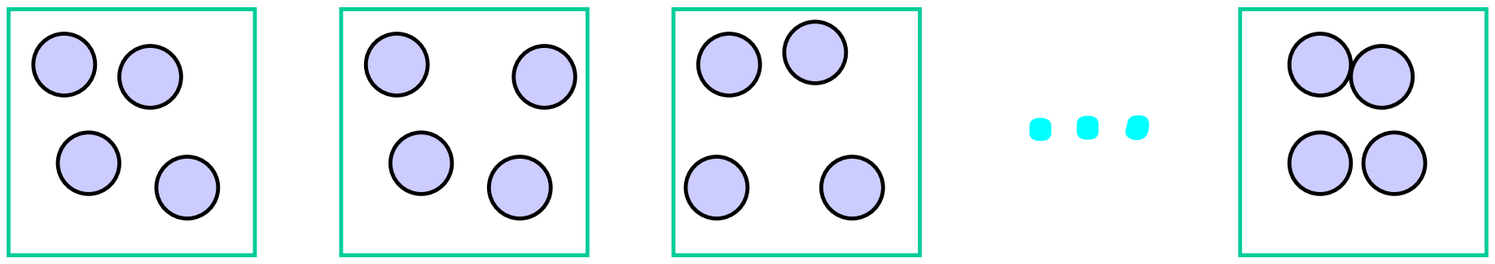


[FB, LGD, SP & PD, 2016]

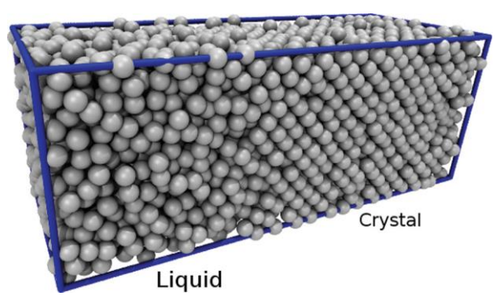
Laboratory of Computational Biophysical Chemistry

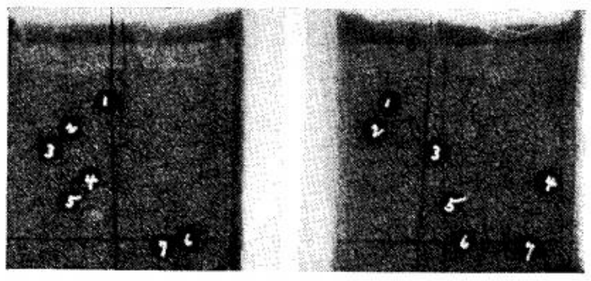
How do I generate the configurations?

$$g(r) = \frac{hist(r)}{\rho\, n_{obs} \Delta V}$$



Laboratory of Computational Biophysical Chemistry



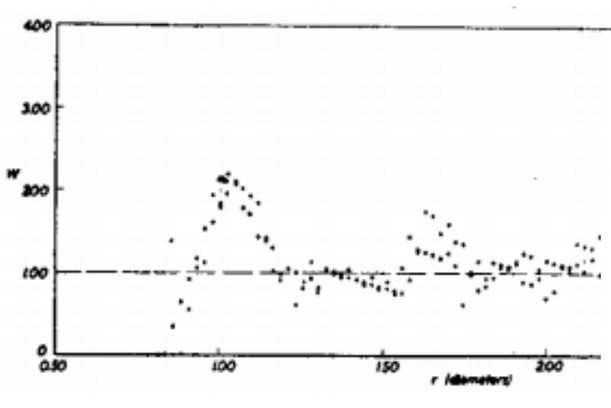


MARCH, 1936 JOURNAL OF CHEMICAL PHYSICS VOLUME 4

The Distribution of Molecules in a Model Liquid

W. E. Morrell and J. H. Hildebrand, *Department of Chemistry, University of California*

(Received January 3, 1936)



by merely pouring steel spheres onto a flat surface and measuring each time the distance between two black ones, then tabulating these distances, obtained a curve for two dimensions which had the characteristics of his *W* curve for mercury. Very similarly, Prins[5] poured seeds

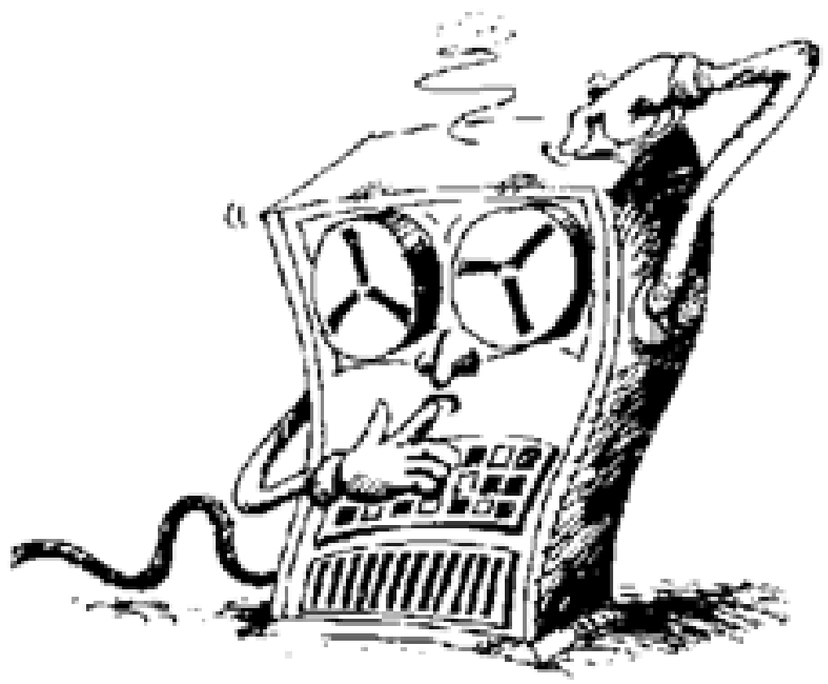
Laboratory of Computational Biophysical Chemistry

How does computer simulation enter in this picture?

Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

17

Laboratory of Computational Biophysical Chemistry



[unknown author]

Easier to be done in a computer experiment!

18

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 31, NUMBER 2 AUGUST, 1959

Studies in Molecular Dynamics. I. General Method*

B. J. Alder and T. E. Wainwright

Lawrence Radiation Laboratory, University of California, Livermore, California

(Received February 19, 1959)

A method is outlined by which it is possible to calculate exactly the behavior of several hundred interacting classical particles. The study of this many-body problem is carried out by an electronic computer which solves numerically the simultaneous equations of motion. The limitations of this numerical scheme are enumerated and the important steps in making the program efficient on the computers are indicated. The applicability of this method to the solution of many problems in both equilibrium and nonequilibrium statistical mechanics is discussed.

PHYSICAL REVIEW VOLUME 159, NUMBER 1 5 JULY 1967

Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*

Loup Verlet†

Belfer Graduate School of Science, Yeshiva University, New York, New York

(Received 30 January 1967)

The equation of motion of a system of 864 particles interacting through a Lennard-Jones potential has been integrated for various values of the temperature and density, relative, generally, to a fluid state. The equilibrium properties have been calculated and are shown to agree very well with the corresponding properties of argon. It is concluded that, to a good approximation, the equilibrium state of argon can be described through a two-body potential.

19

Table 9: History and extrapolated future of computer simulations of molecular dynamics. The future is deduced from extrapolation based on an observed increase of computing speed of a factor 10 every 5 years over the past decades (see Figure 31).

Year	Molecular system (type, size)	Length of the simulation [s]
1957	first molecular dynamics simulation (hard discs)	
1964	atomic liquid (argon)	10^{-11}
1971	molecular liquid (water)	5×10^{-12}
1977	protein in a vacuum	2×10^{-11}
1983	protein in water	2×10^{-11}
1989	protein–DNA complex in water	10^{-10}
1997	polypeptide folding in solvent	10^{-7}
2001	micelle formation	10^{-7}
200x	folding of a small protein	10^{-3}
	And the future ...	
2001	biomolecules in water (ca. 10^4 atoms)	10^{-8}
2029	biomolecules in water (folding sooner?)	10^{-3}
2034	*E. coli* bacteria (ca. 10^{11} atoms)	10^{-9}
2056	mammalian cell (ca. 10^{15} atoms)	10^{-9}
2080	biomolecules in water (as fast as nature)	10^{6}
2172	human body (ca. 10^{27} atoms)	1

Angew. Chem. Int. Ed. **2006**, *45*, 4064 – 4092


articles

Dynamics of folded proteins

J. Andrew McCammon, Bruce R. Gelin & Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

20

y of Computational
hemistry

rdf

Knowledge-based potentials for proteins

Manfred J Sippl

Current Opinion in Structural Biology 1995, **5**:229–235

J. Mol. Biol. (1997) **267**, 207–222

JMB

Novel Knowledge-based Mean Force Potential at Atomic Level

Francisco Melo and Ernest Feytmans*

An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state

Chi Zhang, Song Liu, Hongyi Zhou and Yaoqi Zhou

Protein Sci. 2004 13: 400-411

21

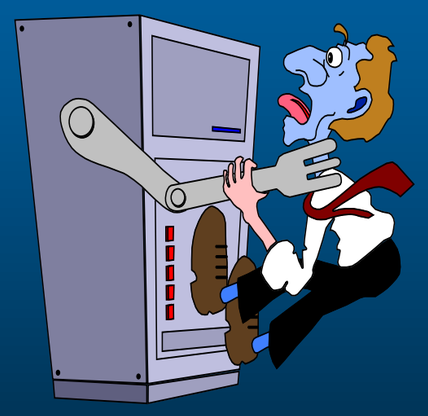
Laboratory of Computational
Biophysical Chemistry

Structural Bioinformatics

Fernando Barroso, FCFRP/USP
SAIFR - March 9-15, 2020

22

We need to create a common language!



Computer simulation
Fernando Barroso (fernando@fcfrp.usp.br)

B.P.C. – DCBM/FCFRP - USP
SAIFR, March 9-15, 2020

23

346

© 2001 Schattauer GmbH

Method Inform Med 2001; 40: 346–58

Laboratory of Computational physical Chemistry

What is Bioinformatics? A Proposed Definition and Overview of the Field

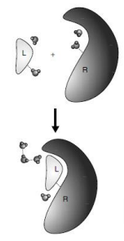
N. M. Luscombe, D. Greenbaum, M. Gerstein
Department of Molecular Biophysics and Biochemistry
Yale University, New Haven, USA

"…is conceptualising **biology** in terms of **molecules** (in the sense of **Physical chemistry**) and applying **"informatics techniques"** (derived from disciplinessuch as applied maths, computer science and statistics) to **understand** and **organise** the **information** associated with these molecules, **on a large scale**."

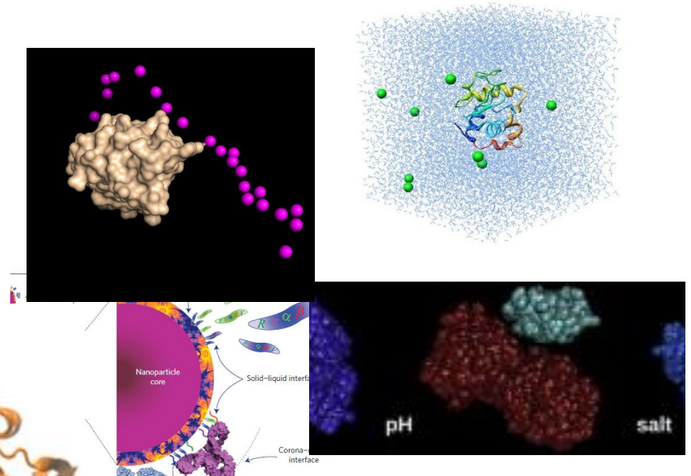
24

Structural Bioinformatics

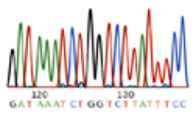
Laboratory of Computational Biophysical Chemistry



Homology modeling



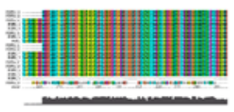
Sequencing



Template → Target

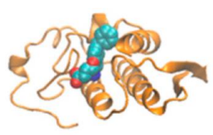
"Omics"

proteomics
data mining
alignments
genomics



Molecular modeling

structure prediction
docking



biomolecular simulation

Binding affinities
molecular dynamics
protein interactions

Statistics → Physics

25

PSP

Laboratory of Computational Biophysical Chemistry

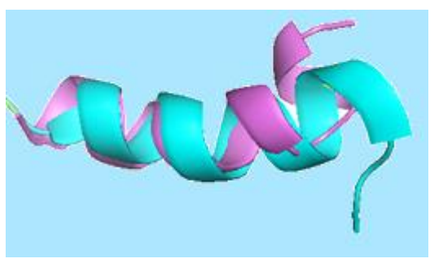
Biological approach

1) Find a known structure with a similar sequence

2) Align the sequences

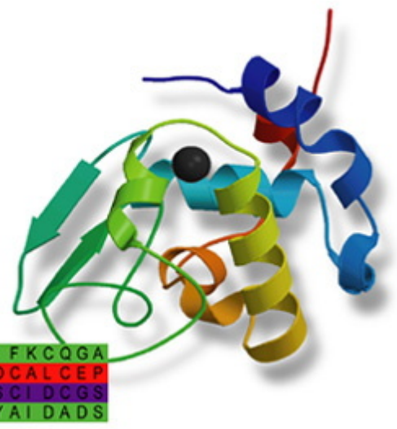
3) Model the unknown structure on the known structure using the alignment

```
TKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGK
|:|||||||::||:||||||||||.||||||||||||.|
ATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTG'
```



Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



```
AILVGSMPRRDGMERKDLLKANVKIFKCQGA
VEVCPVDCFYEGPNFLVIHPDECIDCALCEP
GACKPECPVNIIQGS--IYAIDADSCIDCGS
C--IACGACKPECPVNIIQGS--IYAIDADS
```

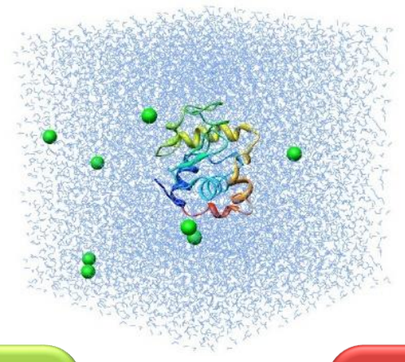
26

Physical approach

Laboratory of Computational Biophysical Chemistry

1) Choose or make a model

2) Simulate



$f(x)$

Modelling → **Solution of the model**

What Force field?

simplification
reduction in # sites

accuracy
practical relevance

Molecular dynamics?

Langevin dynamics?

Genetic algorithm?

Monte Carlo?

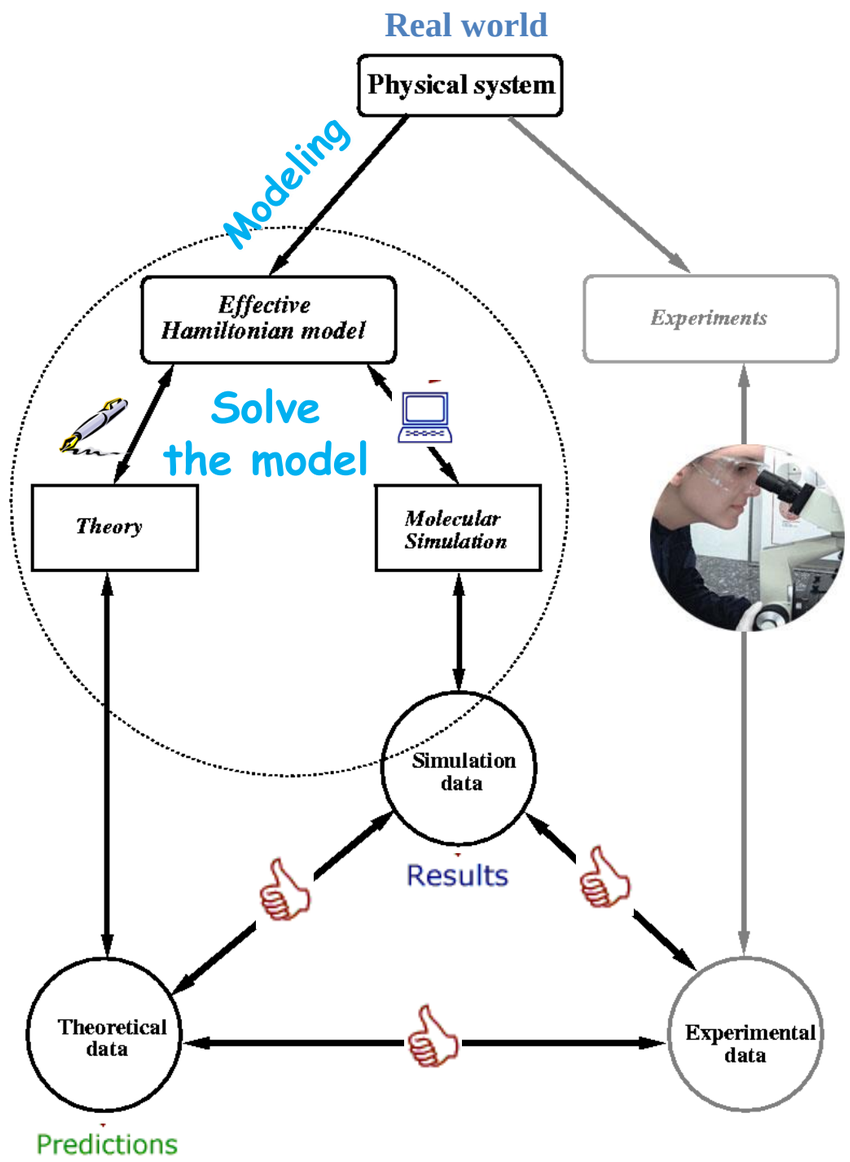
Poisson-Boltzmann?

Physical approach

Computer Experiments

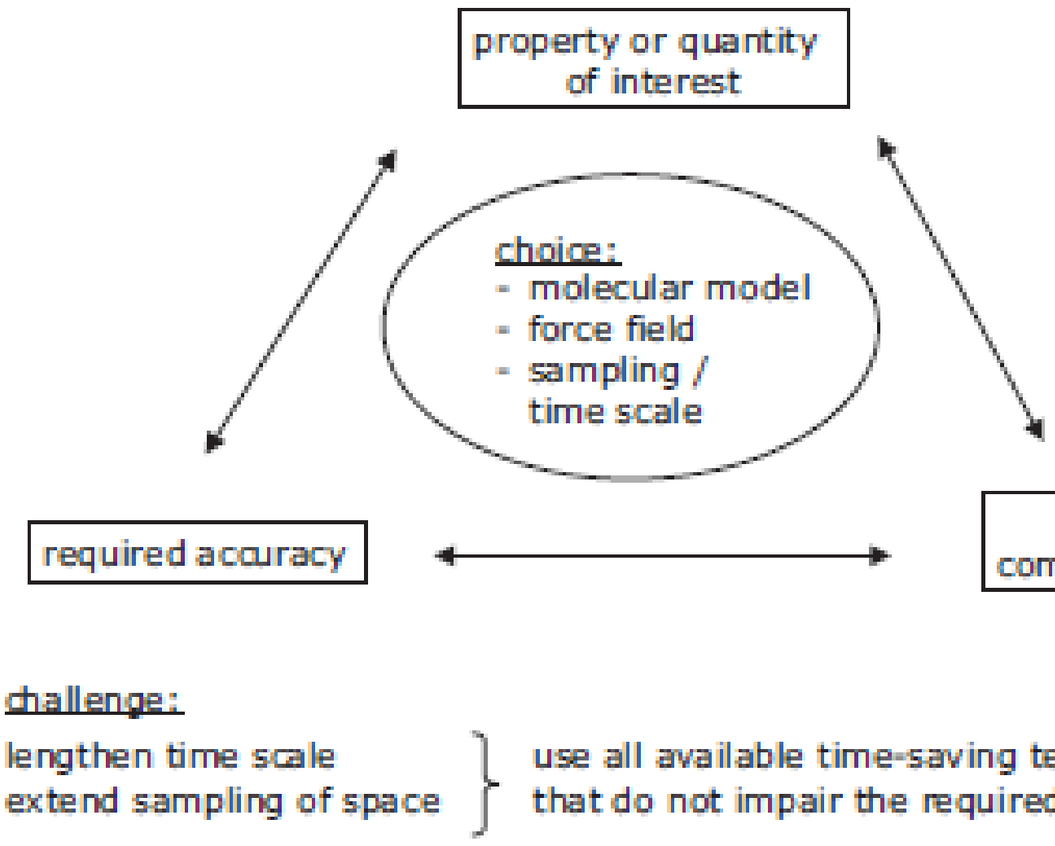
— Our main tools

- Modeling
- Implementation
- Setup
- Runs (simulation)
- Analyses



Laboratory of Computational Biophysical Chemistry

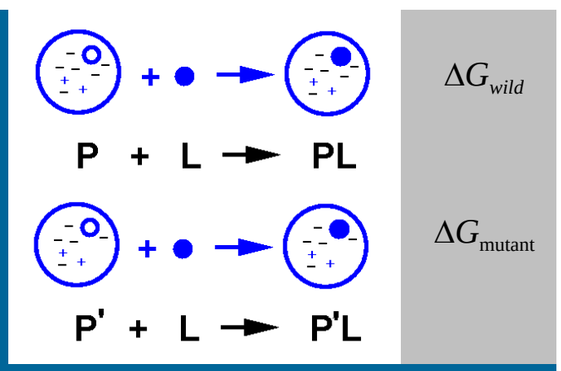
Molecular modeling

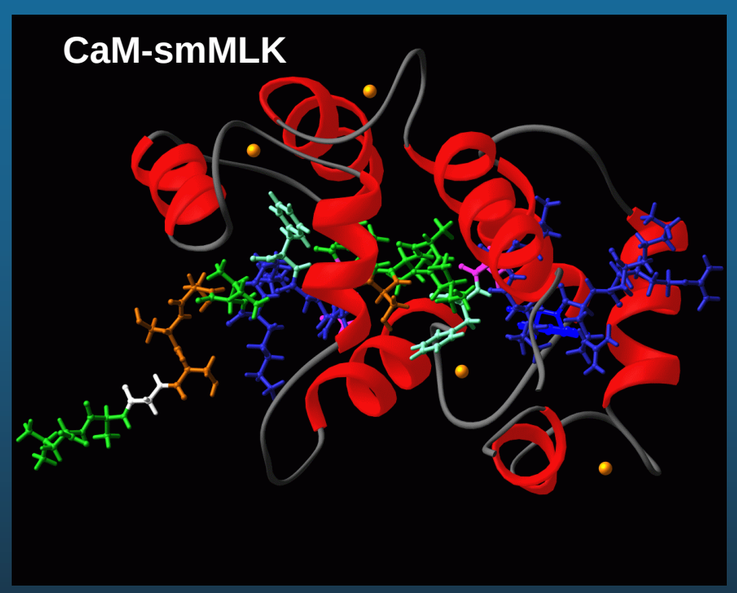


29

General questions

- ✓ Fundamental interactions
- ✓ Each aa contribution
- ✓ Molecular mechanisms





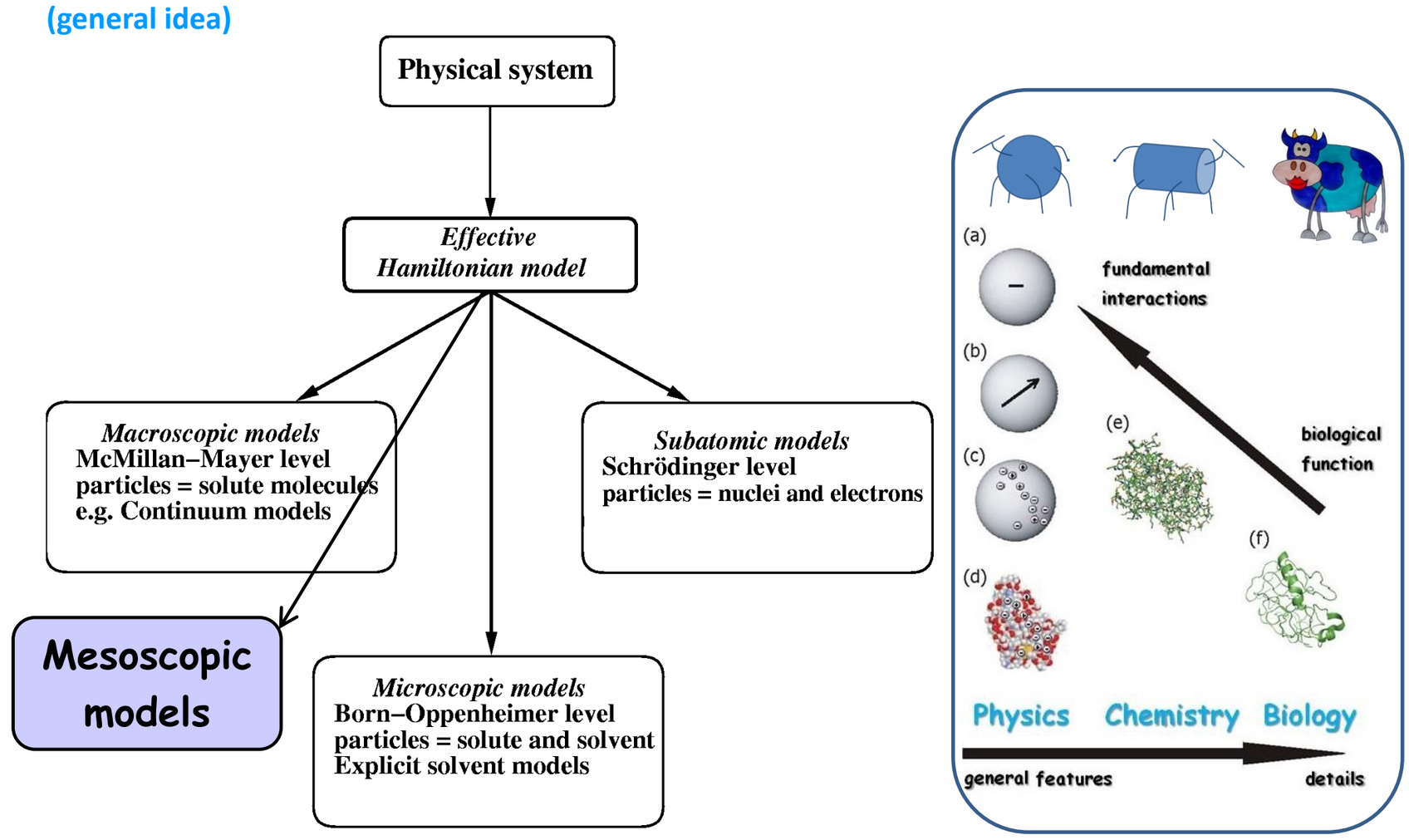
More specific questions

- Complex? In what condition? Is it stable?
- Contact residues?
- Mean smMLK conformation?
- What sites have changes on pKas? Where are they localized?
- Effect of Ca^{2+} on the CaM-smMLK?
- Effect of smMLK on the Ca^{2+}-CaM?
- Critical mutation at contact
- Other critical mutations

30

Models

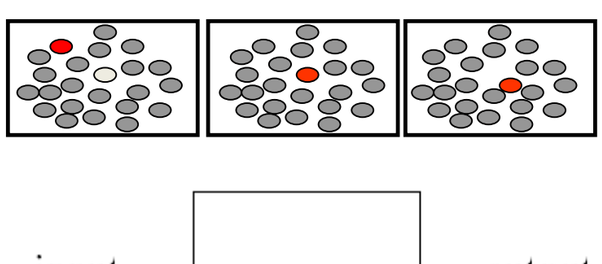
(general idea)

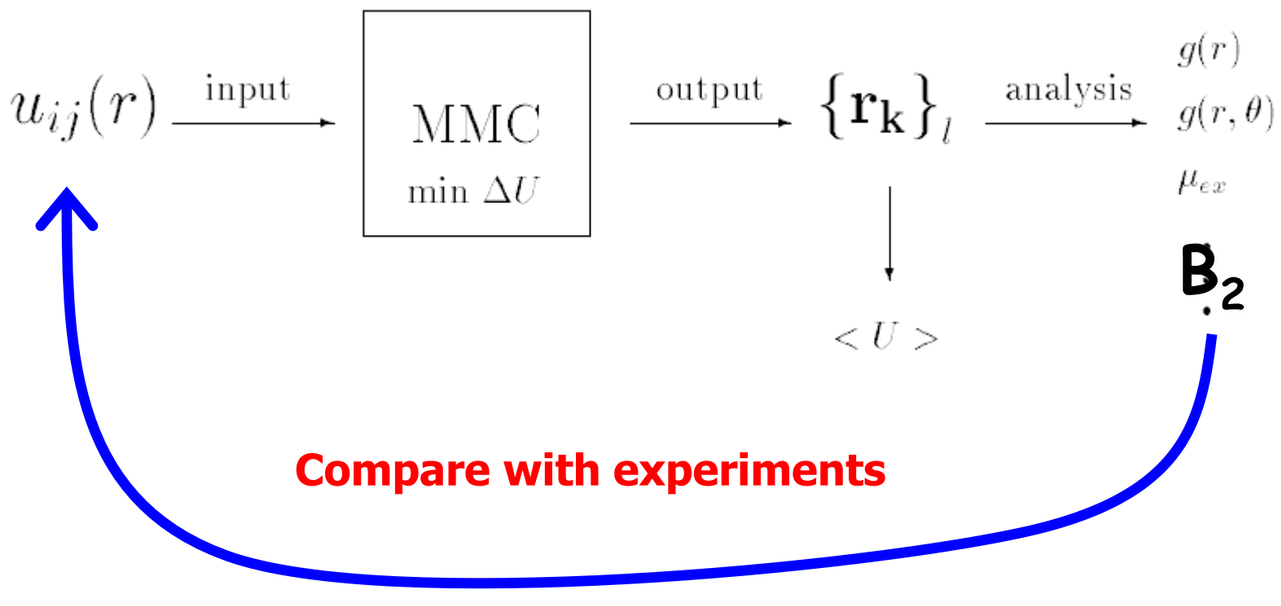


31

$$U = \frac{1}{2}\sum_{\text{bonds}} k_b(b-b_0)^2 + \frac{1}{2}\sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \frac{1}{2}\sum_{\text{dihedrals}} k_\varphi(1+\cos(n\varphi-\delta)) + \sum_{\substack{\text{non-bonded}\\ \text{pairs}}} \left[\frac{A}{r^{12}} - \frac{B}{r^6}\right] + \frac{q_1 q_2}{\varepsilon\varepsilon_0 r}$$

Direct approach





32

Angew. Chem. Int. Ed. 2006, 45, 4064 – 4092

Table 4: Choice of calibration sets of data, systems, properties, and thermodynamic phase for the derivation of the GROMOS biomolecular force-field parameter values.[7]

Type of data	Type of system	Phase	Type of properties	Force-field parameter
structural data (exptl)	small molecules	crystalline solid phase	molecular geometry: bond lengths, bond angles	b_0, θ_0, ξ_0
spectroscopic data (exptl)	small molecules	gas phase	molecular vibrations: force constants	K_b, K_θ, K_ξ
thermodynamic data (exptl)	small molecules, mixtures, solutions	condensed phase	heat of vaporization, density, partition coefficient, free energy of solvation	van der Waals: $C_{12}(i,j)$, $C_6(i,j)$, q_i(final)
dielectric data (exptl)	small molecules	condensed phase	dielectric permittivity, relaxation	charges q_i
transport data (exptl)	small molecules	condensed phase	diffusion and viscosity coefficients	$C_{12}(i,j)$, $C_6(i,j)$, q_i
electron densities (theor.)	small molecules	gas phase	quantum-chemical calculation of atom charges	charges q_i(initial)
energy profiles (theor.)	small molecules	gas phase	quantum-chemical calculation of torsional-angle rotational profiles	K_ϕ, δ, m

33

Laboratory of Computational Biophysical Chemistry

Examples of force fields

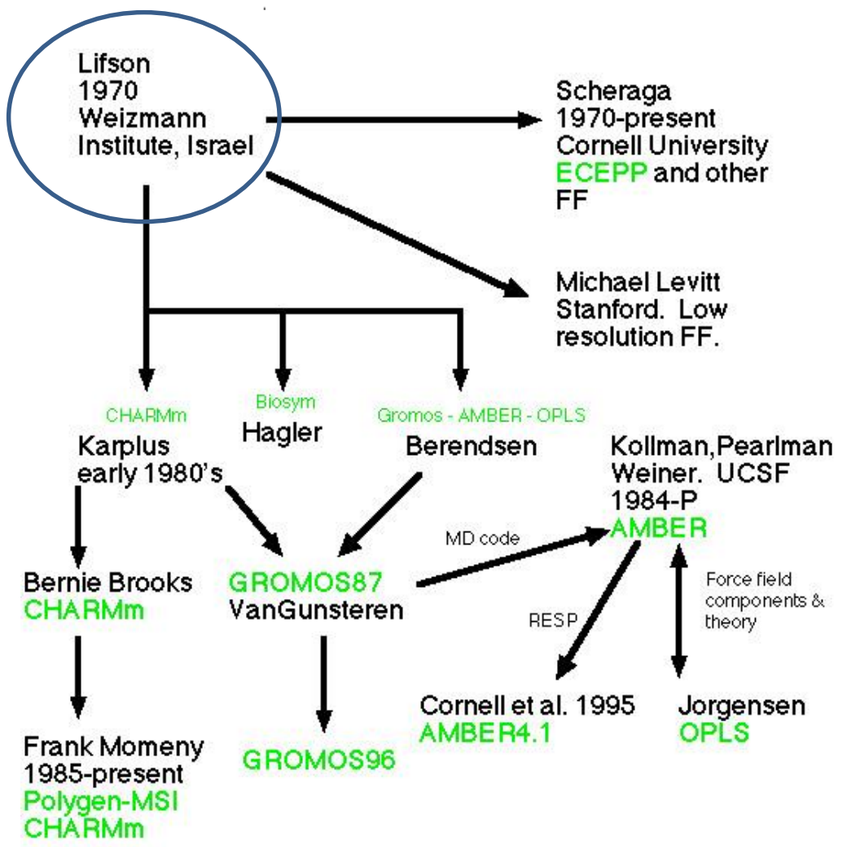
Fernando Barroso, FCFRP/USP

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 49, NUMBER 11 1 DECEMBER 1968

Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and *n*-Alkane Molecules

S. LIFSON AND A. WARSHEL
Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel
(Received 13 May 1968)

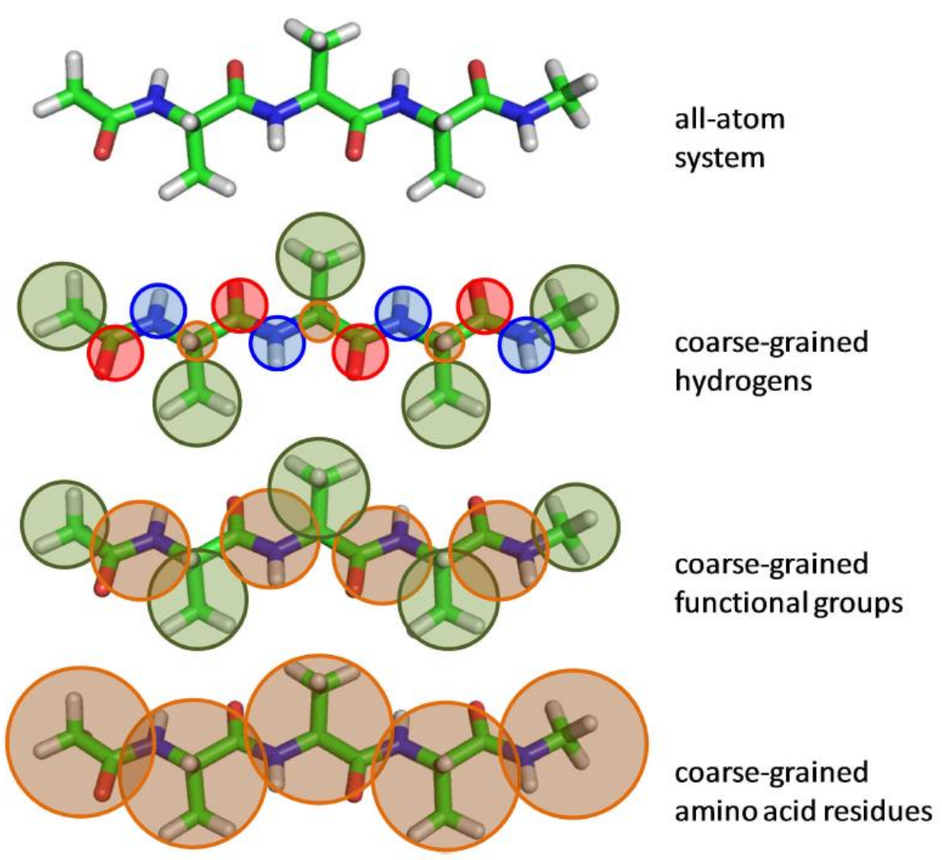
$$U = \frac{1}{2}\sum_{\text{bonds}} k_b(b-b_0)^2 + \frac{1}{2}\sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \frac{1}{2}\sum_{\text{dihedrals}} k_\varphi(1+\cos(n\varphi-\delta)) + \sum_{\substack{\text{non-bonded}\\ \text{pairs}}} \left[\frac{A}{r^{12}} - \frac{B}{r^6}\right] + \frac{q_1 q_2}{\varepsilon\varepsilon_0 r}$$



35

Laboratory of Computational Biophysical Chemistry

There are many ways to develop pseudoatoms of varying resolution. Consider polyalanine:



36

J. Chem. Theory and Comput. **2008,** *4,* 819–834

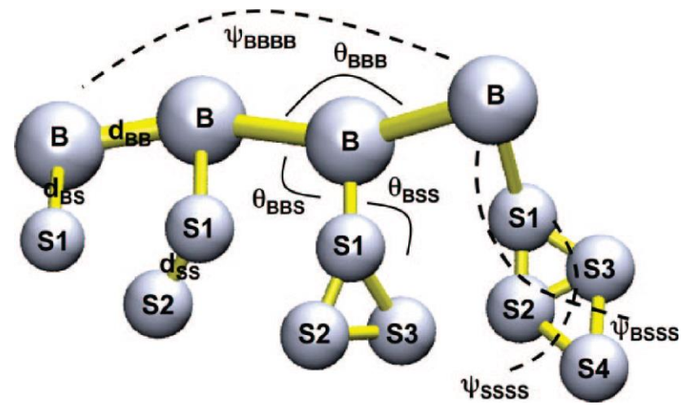
JCTC Journal of Chemical Theory and Computation

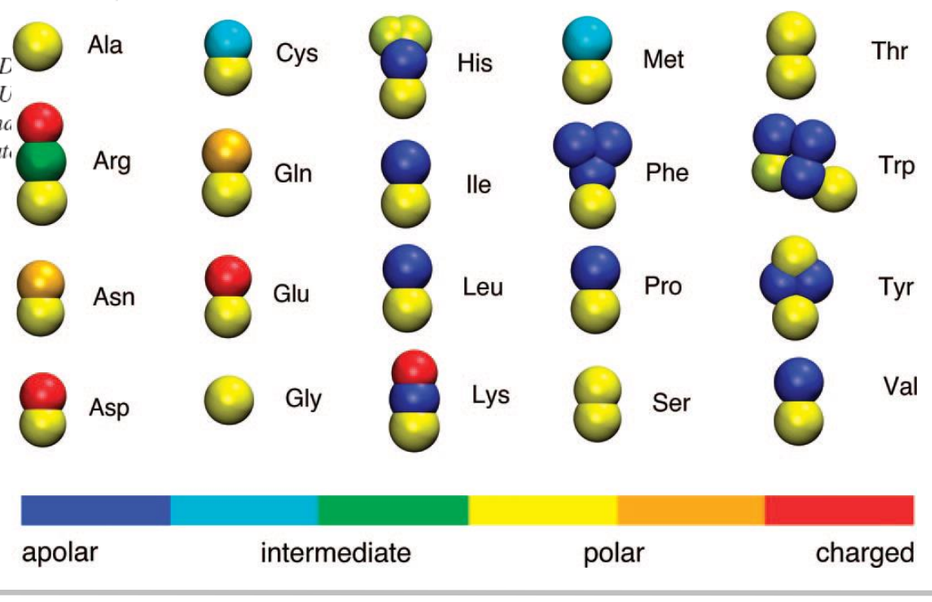
The MARTINI Coarse-Grained Force Field: Extension to Proteins

Luca Monticelli,[†] Senthil K. Kandasamy,[‡] Xavier Periole,[§] Ronald G. Larson,[‡] D. Peter Tieleman,[†] and Siewert-Jan Marrink*[,§]

Dept of Biological Sciences, University of Calgary, 2500 University D AB, T2N 1N4, Canada, Chemical Engineering Department, The U Michigan, 2300 Hayward Street, Ann Arbor, Michigan 48109, and Biomolecular Sciences and Biotechnology Institute & Zernike Institut Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received November 27, 2007





Flory–Huggins Theory

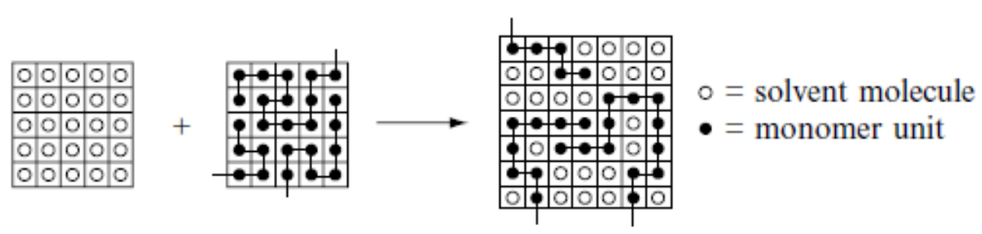
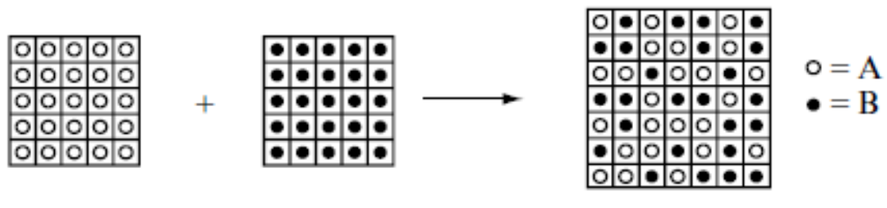
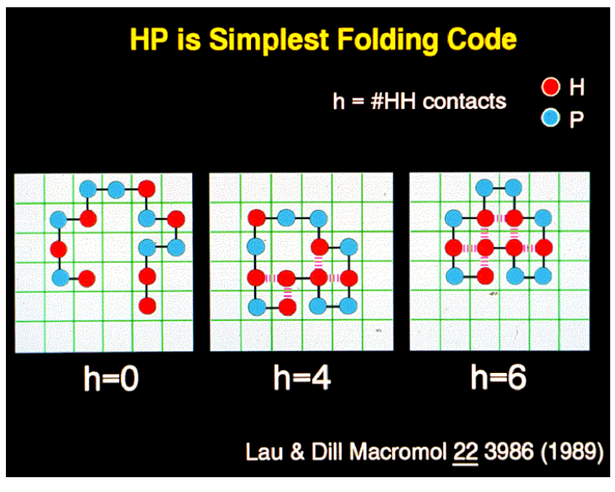


Figure 10.5 Lattice model for the random mixing of a polymer (filled circles) and a liquid (open circles)

Bragg–Williams Theory

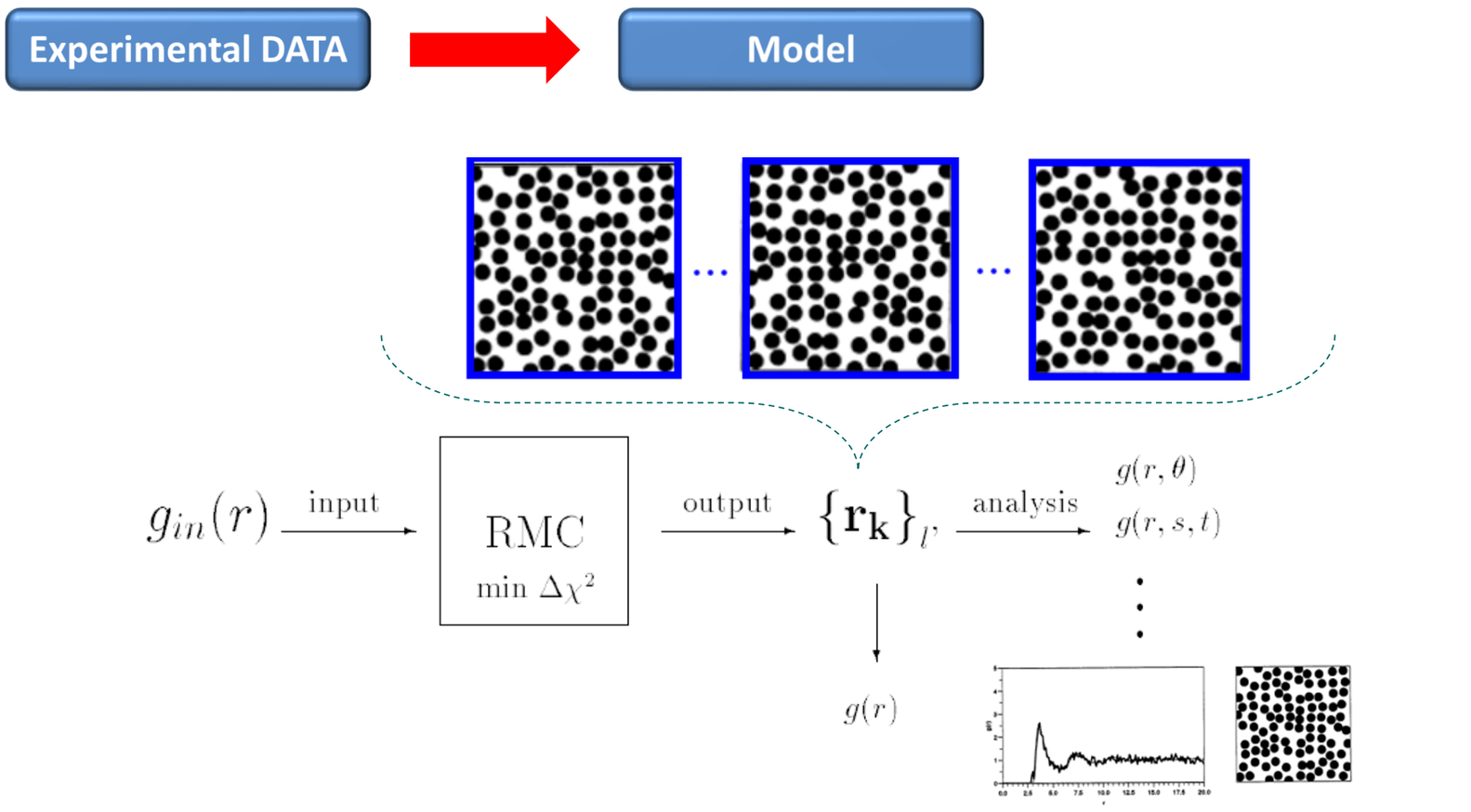
1. The mixture is random.
2. The number of nearest neighbours is constant.
3. The interaction is limited to nearest neighbours.





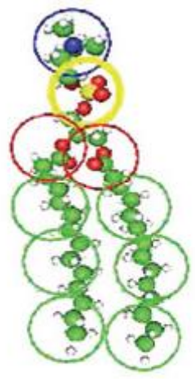
Laboratory of Computational Biophysical Chemistry

RMC aim



39

Laboratory of Computational Biophysical Chemistry

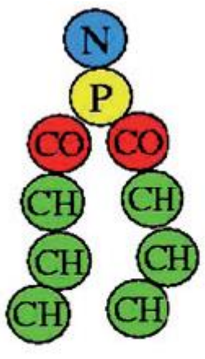


All-atom model
118 atoms

Deriving Effective Mesoscale Potentials from Atomistic Simulations

J Comput Chem 24: 1624–1636, 2003

DIRK REITH, MATHIAS PÜTZ, FLORIAN MÜLLER-PLATHE*
Max-Planck-Institut für Polymerforschung, D-55128 Mainz, Germany



Coarse-grained model
10 sites

Systematic coarse-graining of molecular models by the Newton inversion method

Alexander Lyubartsev,* Alexander Mirzoev, LiJun Chen and Aatto Laaksonen*

Faraday Discuss., 2010, **144**, 43–56

40

Solving the model

$f(x)$

Modelling

Solution of the model

41

Molecular Dynamics

Protein Structure

Protein Thermodynamics
Protein Dynamics
Protein Function

compute Forces

$$v(t) = \frac{dr}{dt} = \frac{F}{m}t + v_0$$

$$r(t) = \frac{F}{2m}t^2 + v_0 t + r_0$$

$\mathbf{F} = m\mathbf{a}$

Update positions and velocities

Update Trajectory

Analysis

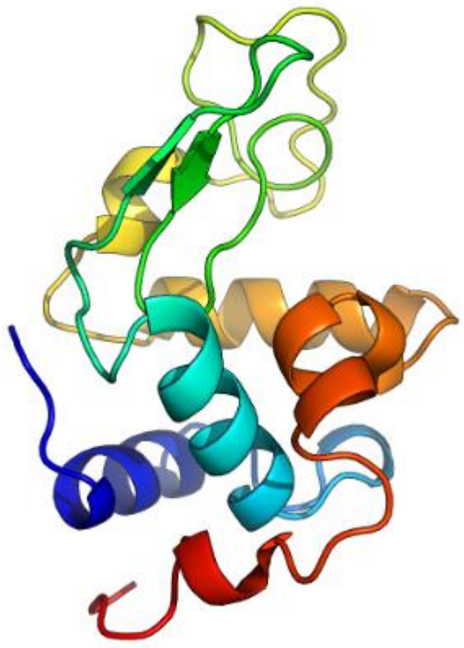
Time averages

42

GROMACS Tutorial

Lysozyme in Water

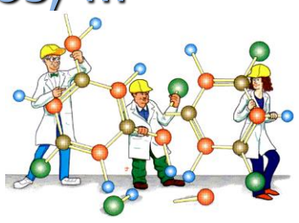
Justin A. Lemkul, Ph.D.
Virginia Tech Department of Biochemistry



Laboratory of Computational Biophysical Chemistry

- ✓ Gromacs
- ✓ Charmm
- ✓ Amber
- ✓ Discover, Insight
- ✓ Sigma, Tripos, …
- ✓ NAMD
- ✓ Tinker
- ✓ Lammps…

GROMACS
FAST. FLEXIBLE. FREE.

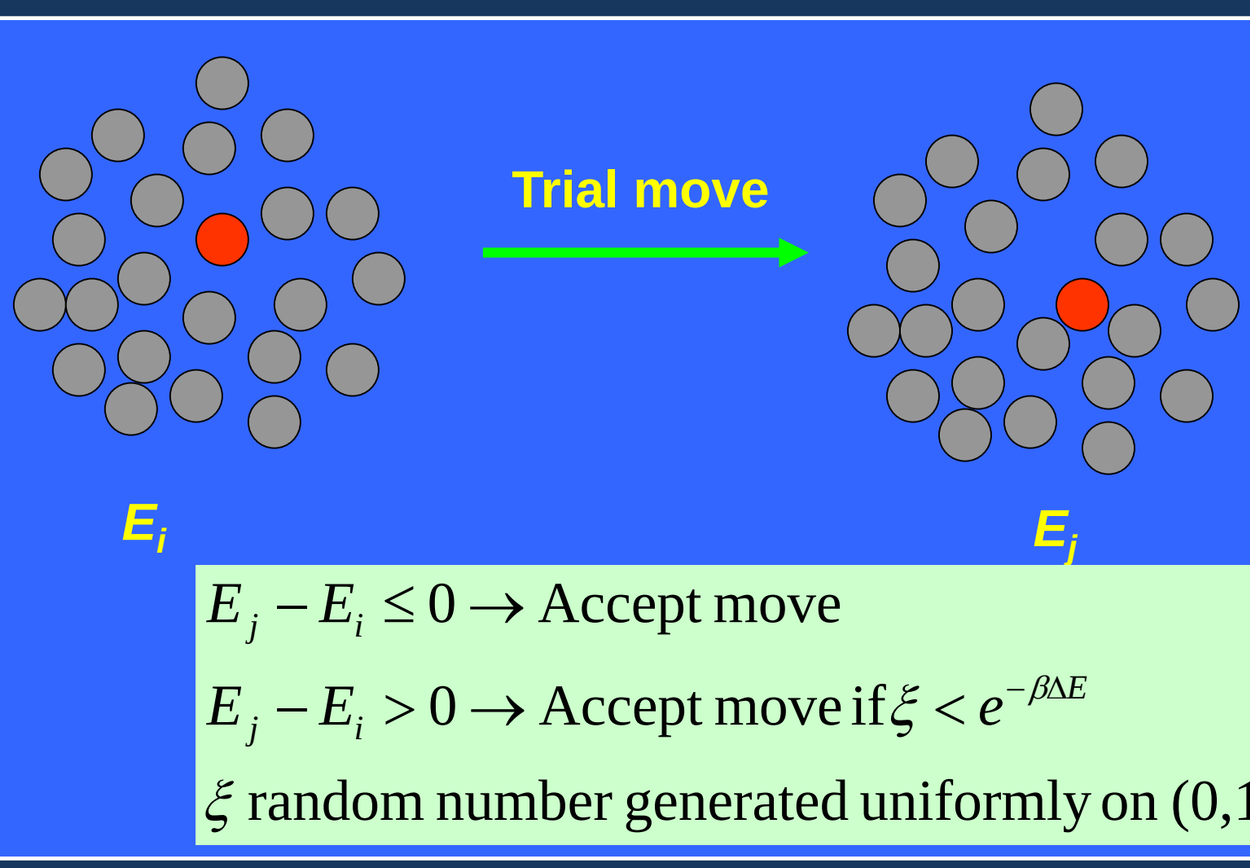


LAMMPS

http://www.mdtutorials.com/gmx/lysozyme/index.html

43

Monte Carlo



$$E_j - E_i \le 0 \rightarrow \text{Accept move}$$

$$E_j - E_i > 0 \rightarrow \text{Accept move if } \xi < e^{-\beta \Delta E}$$

ξ random number generated uniformly on (0,1)

44

Monte Carlo

Protein Structure → Protein Thermodynamics, Protein Function

Random displacement

$\propto 10^6$ steps

Evaluate Energies

$P_j \propto e^{-\beta E_j}$

Update ensemble → Analysis

Accept or reject

Ensemble averages

45

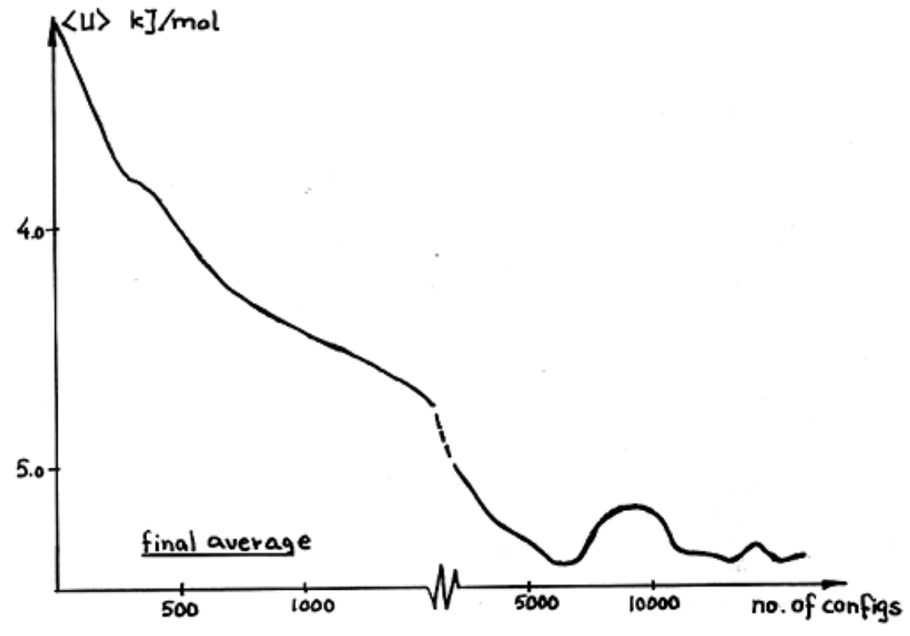
Monte Carlo

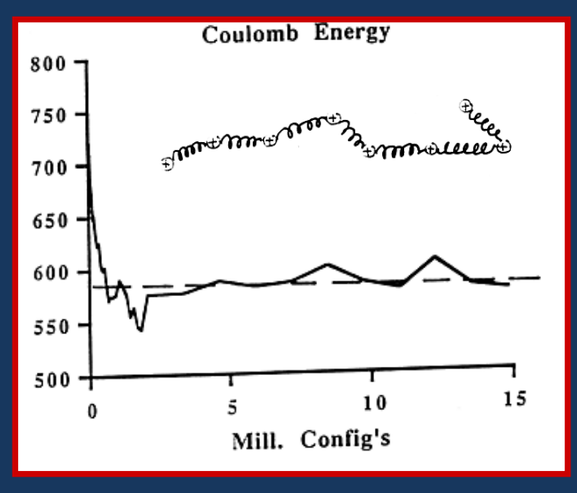
$\langle U \rangle_{final} = 5.36 \pm 0.01$ kJ/mol (after $1.5 \cdot 10^6$ configs)

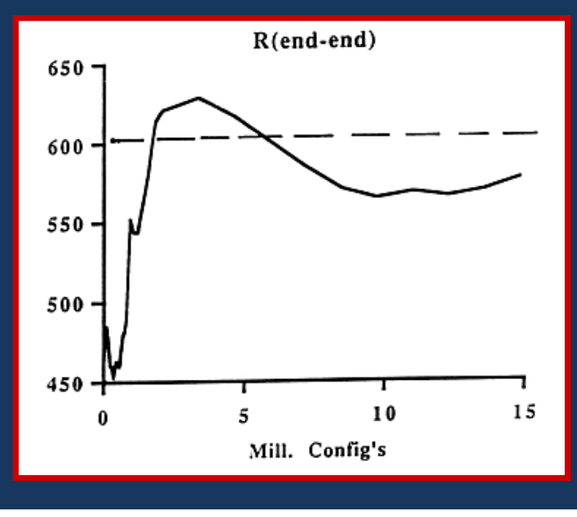
Example: 256 colloidal particles interacting via a screened Coulomb potential,

$u(r_{ij}) = f(Z_0) \frac{e^{-\kappa r_{ij}}}{r_{ij}}$; $Z_0 = 20$ $\kappa = 0.052$ Å^{-1}

the start configuration is a fcc lattice.







46

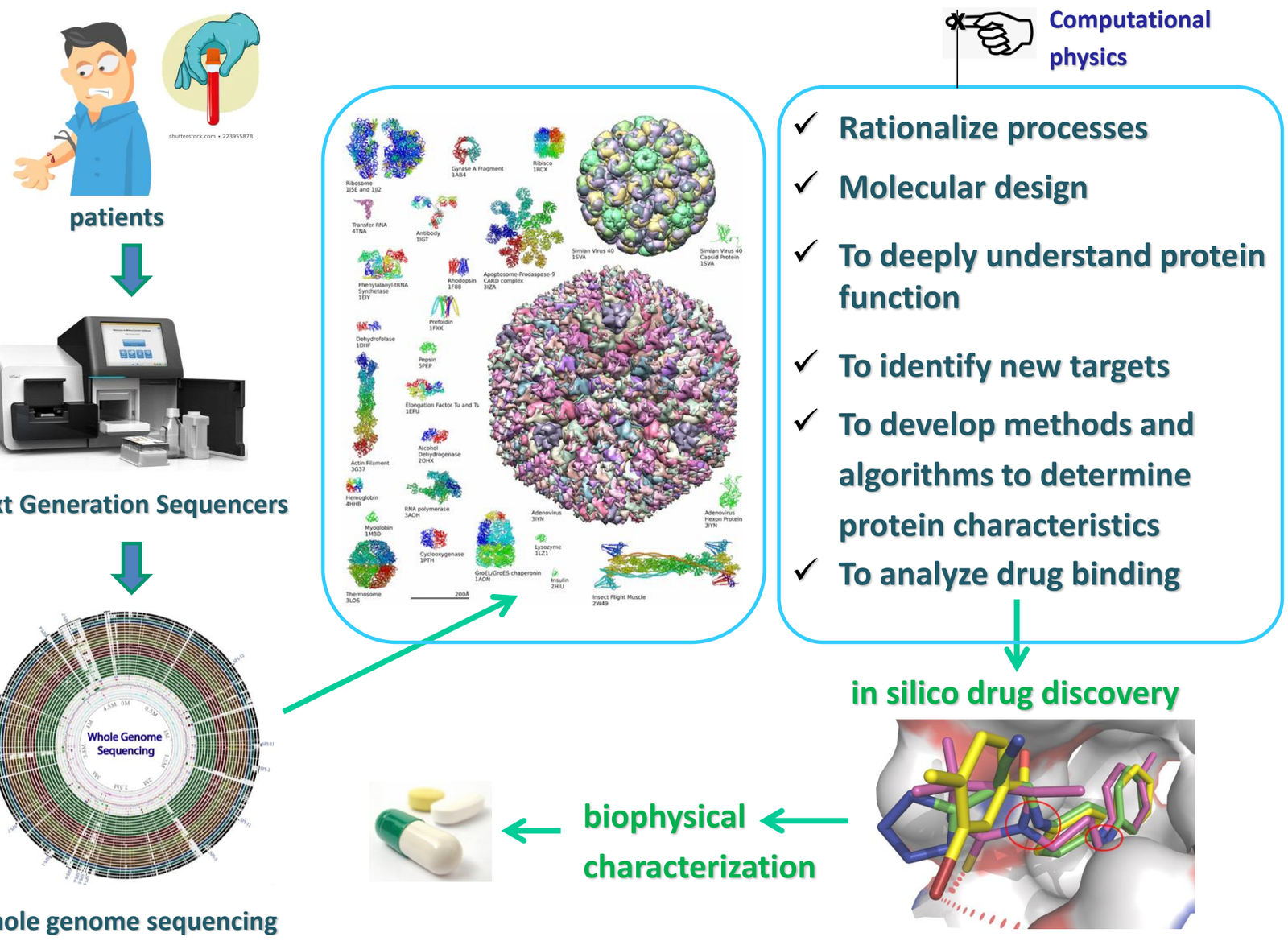
Simulation Methods for Protein Structure Fluctuations

Scott H. Northrup and J. Andrew McCammon

Biopolymers, Vol. 19, Issue 5, pp. 1001-1016 (1980)

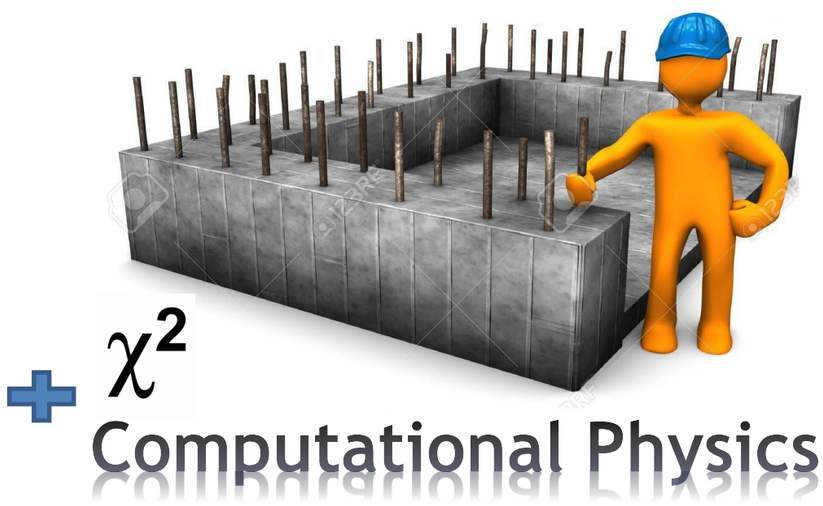
Three numerical techniques for generating thermally accessible configurations of globular proteins are considered; these techniques are the molecular dynamics method, the Metropolis Monte Carlo method, and a modified Monte Carlo method which takes account of the forces acting on the protein atoms. The molecular dynamics method is shown to be more efficient than either of the Monte Carlo methods. Because it may be necessary to use Monte Carlo methods in certain important types of sampling problems, the behavior of these methods is examined in some detail. It is found that an acceptance ratio close to 1/6 yields optimum **efficiency for the Metropolis method**, in contrast to what is often assumed. This result, together with the overall inefficiency of the Monte Carlo methods, appears to arise from the **anisotropic forces** acting on the protein atoms due to their covalent bonding. Possible ways of improving the Monte Carlo methods are suggested.

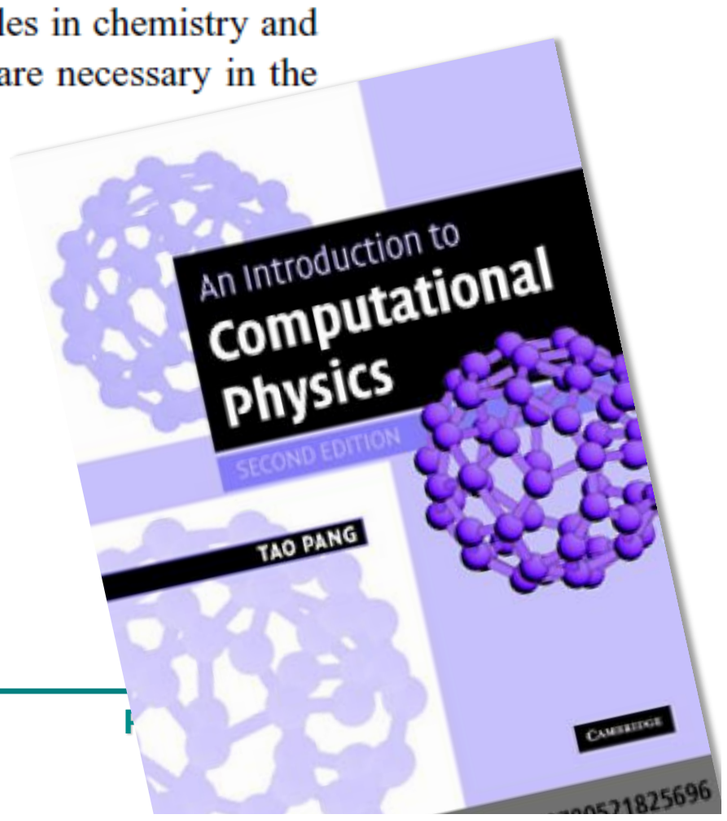
47



48

Computational physics, in my view, is the foundation of computational science. It deals with basic computational problems in physics, which are closely related to the equations and computational problems in other scientific and engineering fields. For example, numerical schemes for Newton's equation can be implemented in the study of the dynamics of large molecules in chemistry and biology; algorithms for solving the Schrödinger equation are necessary in the

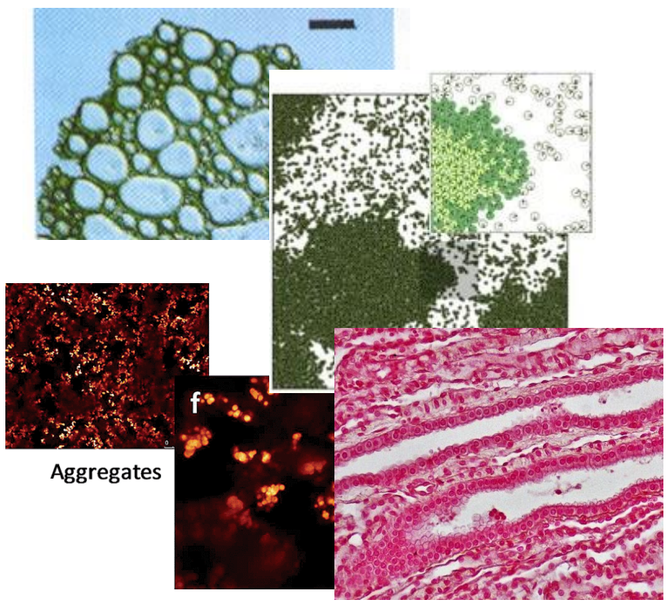


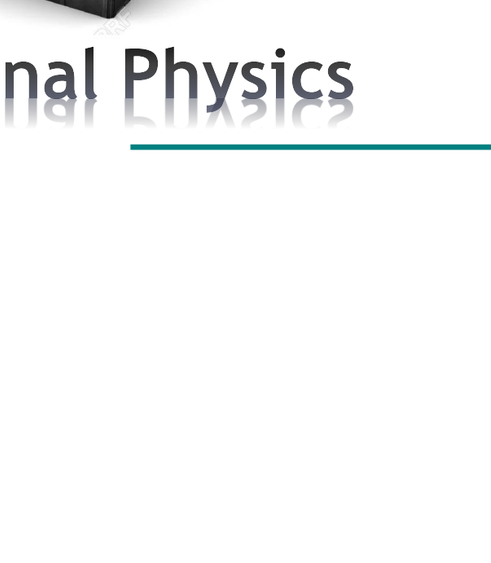


49

From Molecular to Macroscopic Scale (Bio2020)

Experiments, theory and computer simulation





Macroscopic scale

Biomolecular interactions

Single molecule

Electrolyte solutions

50

From Molecular to Macroscopic Scale (Bio2020)

Experiments, theory and computer simulation