

# Modeling and Interpretation of Complex Systems and Processes

Bogdan Lesyng

ICM and Faculty of Physics, Warsaw University  
(<http://www.icm.edu.pl/~lesyng/>)

and

European Centre of Excellence for  
Multiscale Biomolecular Modelling,  
Bioinformatics and Applications  
(<http://www.icm.edu.pl/mamba>)



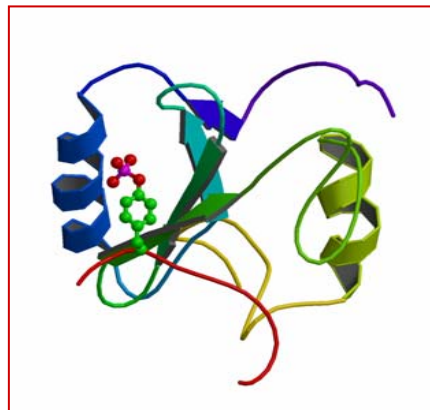
Juelich, 15-18 January, 2006

**Sequences  
at the protein &  
nucleic acids levels**

```
1  RPDFCLEPPY  10
11 TGPCCKARIIR 20
21 YFYNAKAGLC 30
31 QTFVYGGCRA 40
41 KRNNFKSAED 50
51 CMRTCAGGA  58
```



**3D & electronic  
structure**



**Function**

Dynamics,  
classical and/or  
quantum one in  
the real  
molecular  
environment

**Metabolic  
pathways &  
signalling**



**Sub-cellular  
structures & processes**



**Cell(s),  
structure(s)  
&  
functions**

In our organisms we have  $\sim 10^3$  protein kinases and phosphatases which phosphorylate/dephosphorylate other proteins activating or disactivating them. These are controllers of most of metabolic pathways.

## Science

This phylogenetic tree depicts the relationships between members of the complete superfamily of human protein kinases. Protein kinases constitute one of the largest human gene families and are key regulators of cell function. The 578 human protein kinases control protein activity by catalyzing the addition of a negatively charged phosphate group to other proteins. Protein kinases modulate a wide variety of biological processes, especially those that vary signals from the cell membrane to intracellular targets and coordinate complex biological functions.

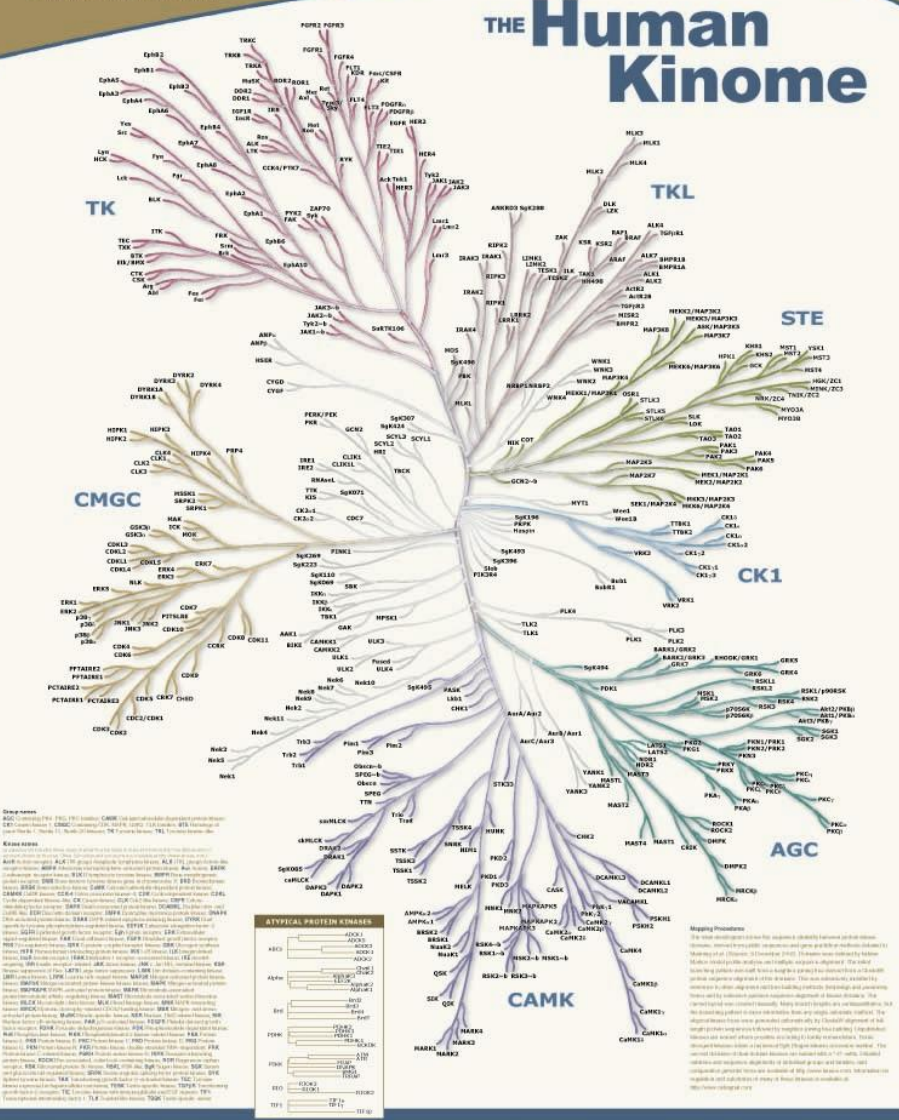
Most protein kinases belong to a single superfamily of enzymes whose catalytic domains are related in sequence and structure. The main diagram illustrates the similarity between the protein sequences of these catalytic domains. Each kinase is at the tip of a branch, and the similarity between various kinases is inversely related to the distance between their positions on the tree diagram. Most kinases fall into small families of highly related sequences, and most

families are part of larger groups. The seven major groups are labeled and colored distinctly. Other kinases are shown in the center of the tree, colored gray. The relationships shown on the tree can be used to predict protein substrates and biological function for many of the over 100 human protein kinases previously uncharacterized.

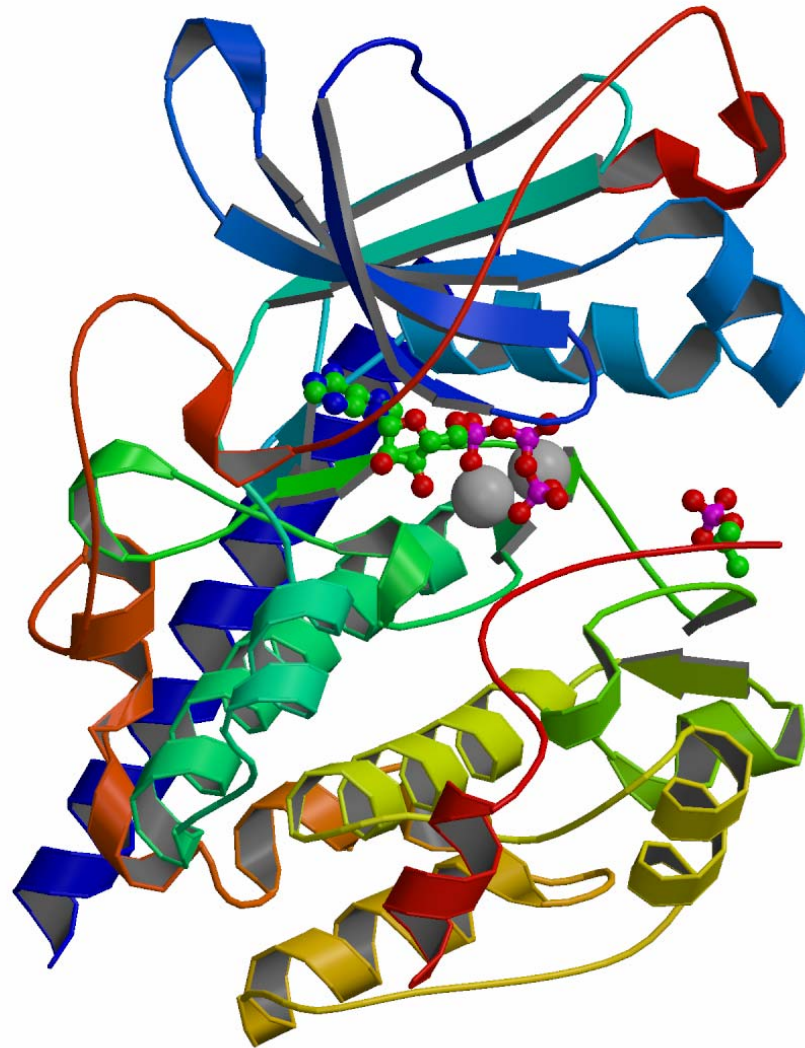
The inset diagram shown below for seven atypical protein kinase families. These proteins have a verified or strongly predicted kinase activity, but have little or no sequence similarity to members of the protein kinase superfamily. A further eight atypical protein kinases in small families of one or two genes are not shown.



# THE Human Kinome



A Protein Kinase Molecule with ATP  
(catalytic domain)



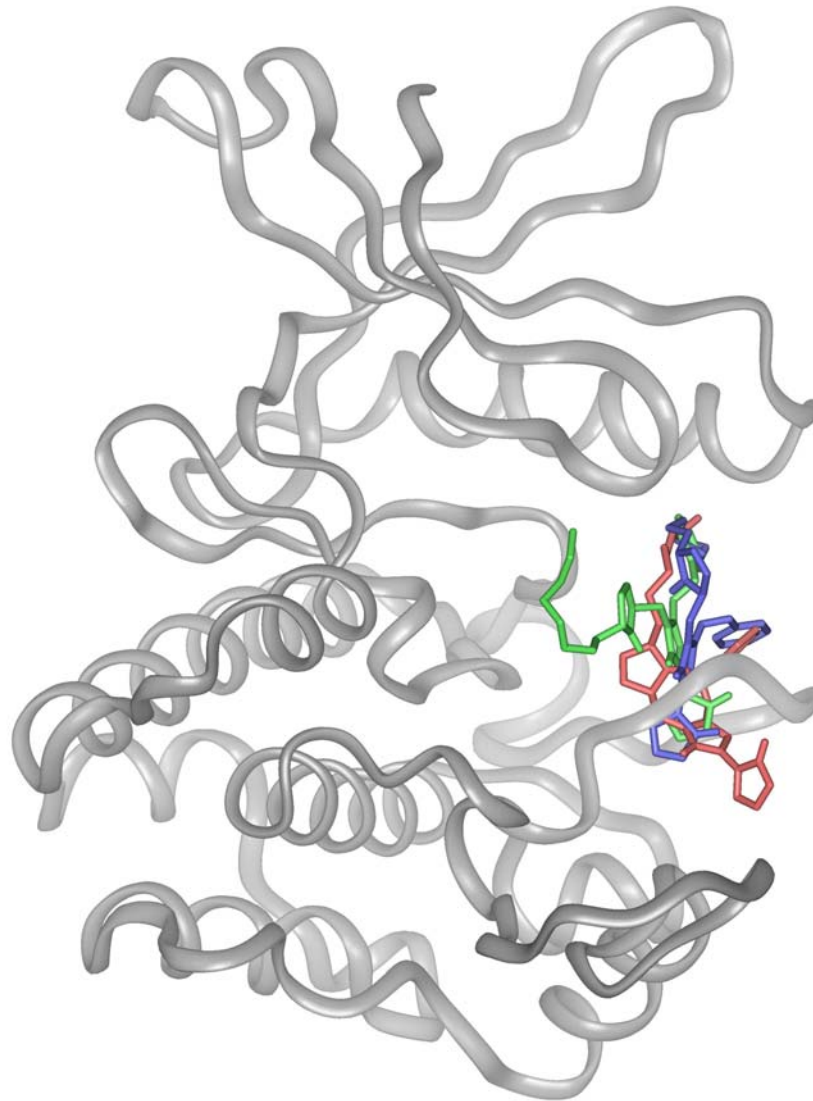
## Designing inhibitors

Information,  
conference on  
"Inhibitors of  
Protein Kinases",  
and workshops on  
"Molecular  
Recognition  
Processes"

June 26-30, 2005  
Warsaw

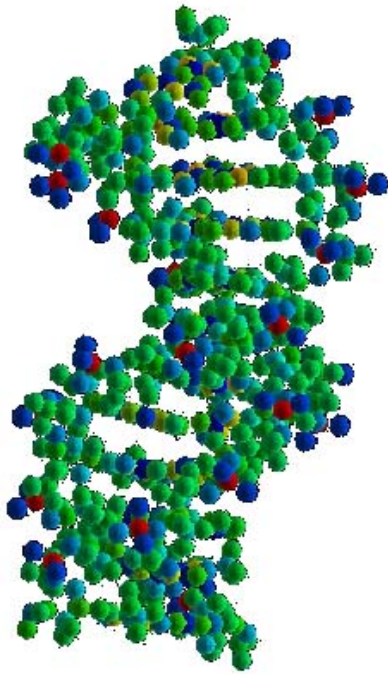
[http://www.icm.edu.pl/  
ipk2005/](http://www.icm.edu.pl/ipk2005/)

Next one  
June, 2007

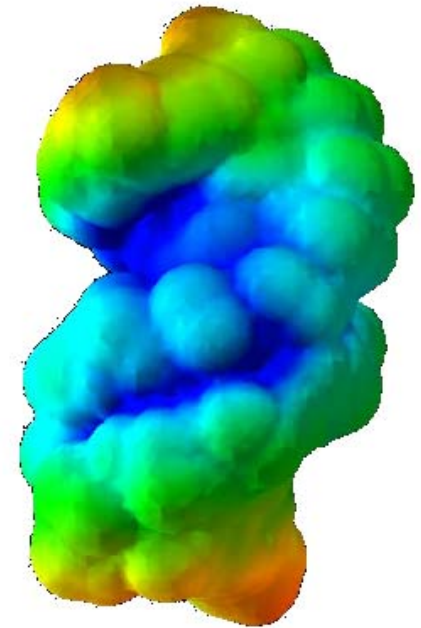


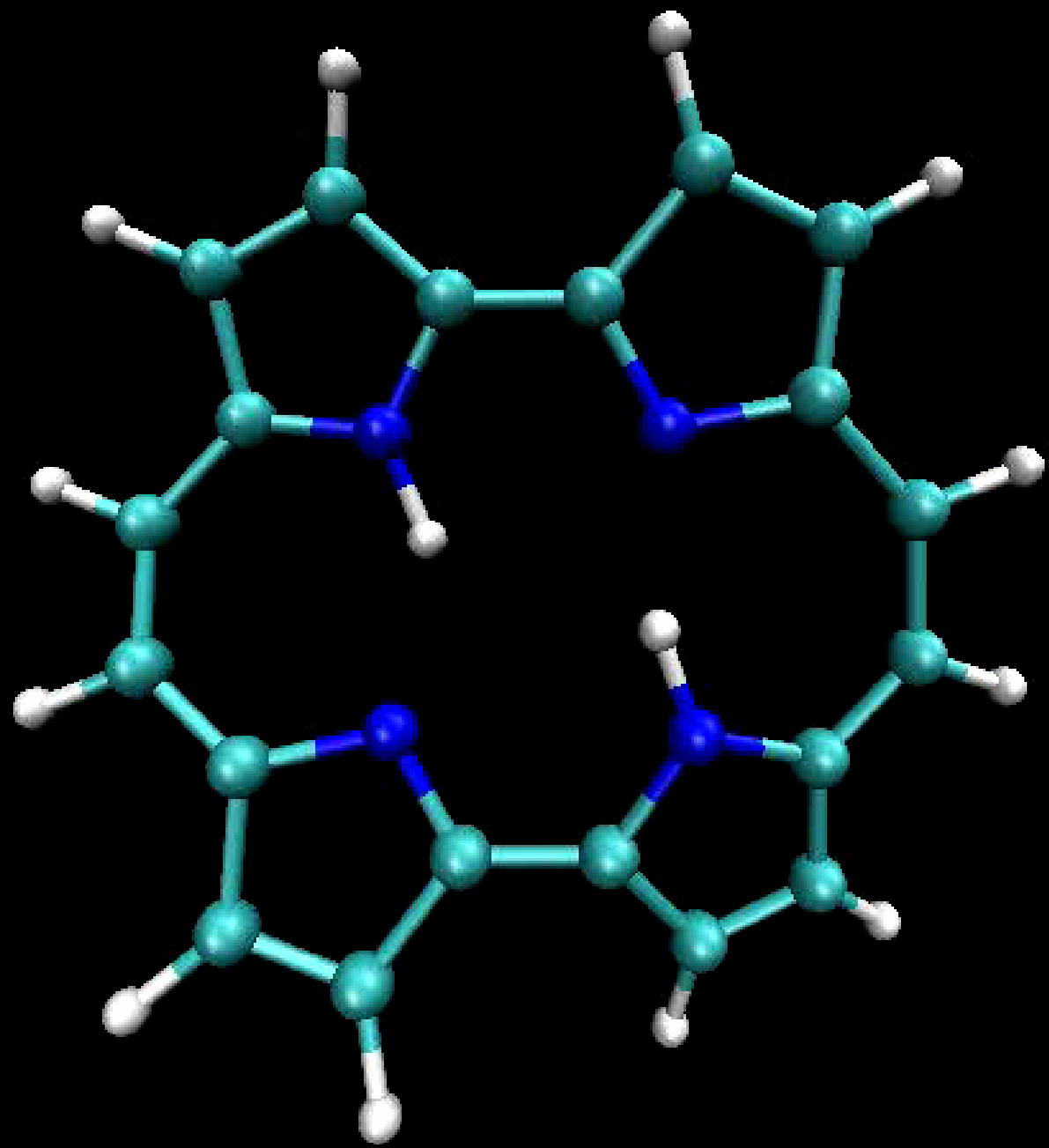
# Multi-scale modeling. Classes of models

**Microscopic models**



**Mesoscopic models**





## Microscopic generators of the potential energy function

- AVB - (quantum)
- AVB/GROMOS - (quantum-classical)
- SCC-DFTB - (quantum)
- SCC-DFTB/GROMOS - (quantum-classical)
- SCC-DFTB/CHARMM - (quantum -classical)
- ....

## Dynamics

- MD (classical)
- QD (quantum)
- QCMD (quantum-classical)
- ....

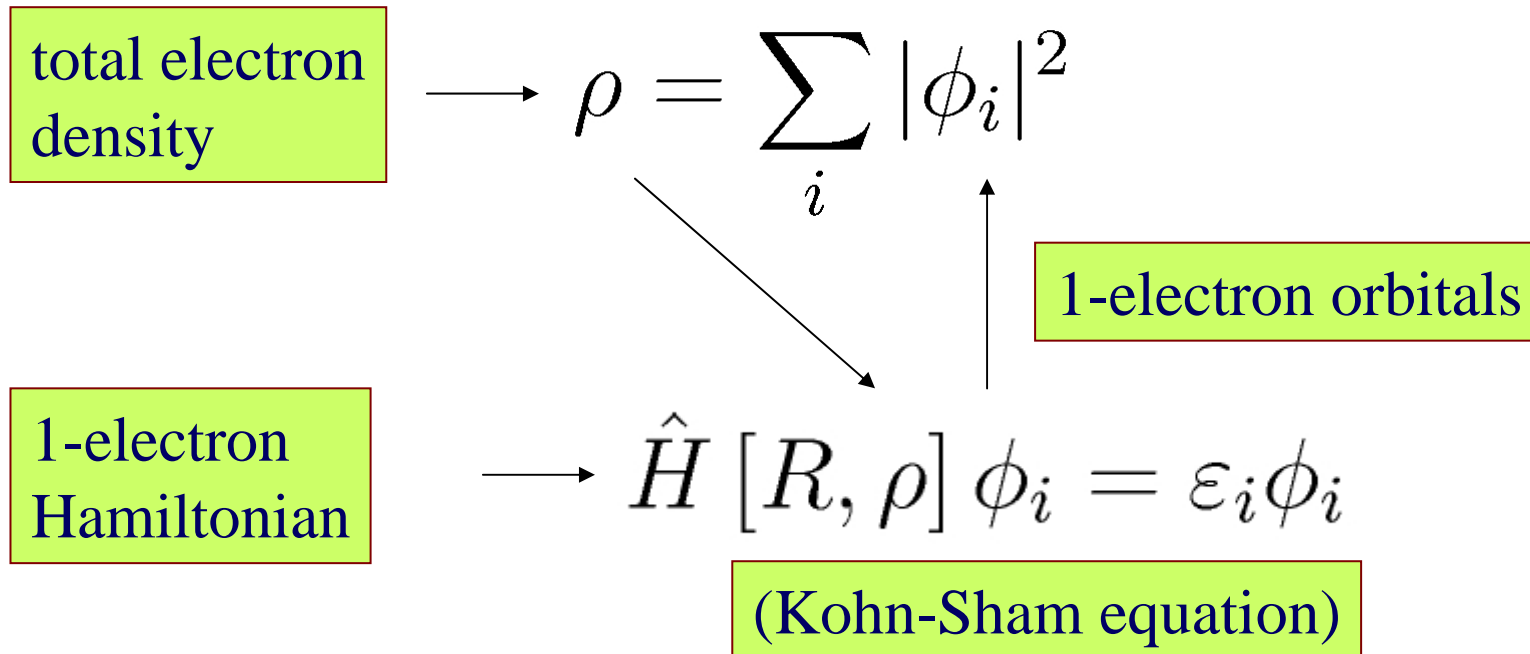
## Mesoscopic potential energy functions

- Poisson-Boltzmann (PB)
- Generalized Born (GB)
- ....

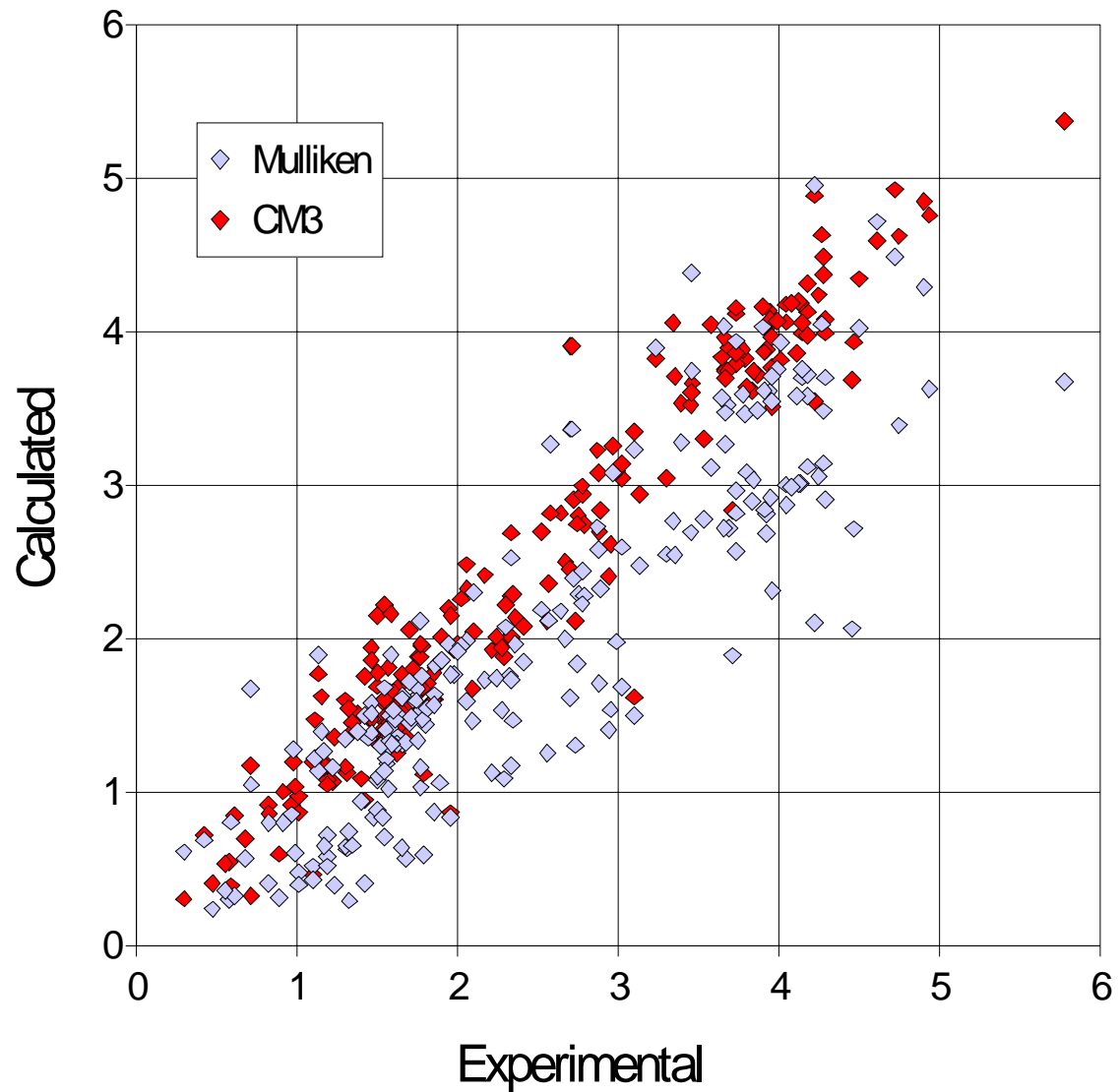
# SCC-DFTB Method

(Self Consistent Charge Density Functional Based Tight Binding Method, SCC DFTB, Frauenheim et al. Phys Stat. Sol. 217, 41, 2000)

basic DFT concepts:



# Dipole moments in Debyes



New generation of charges capable reproducing electrostatic properties, in particular molecular dipole moments.

J.Li, T.Zhu, C.Cramer, D.Truhlar, J. Phys. Chem. A, **102**, 1821(1998)

**CM3/SCC-DFTB charges**

J.A. Kalinowski, B.Lesyng, J.D. Thompson, Ch.J. Cramer, D.G. Truhlar, *Class IV Charge Model for the Self-Consistent Charge Density-Functional Tight-Binding Method*, J. Phys. Chem. A, **108**, 2545-2549 (2004)

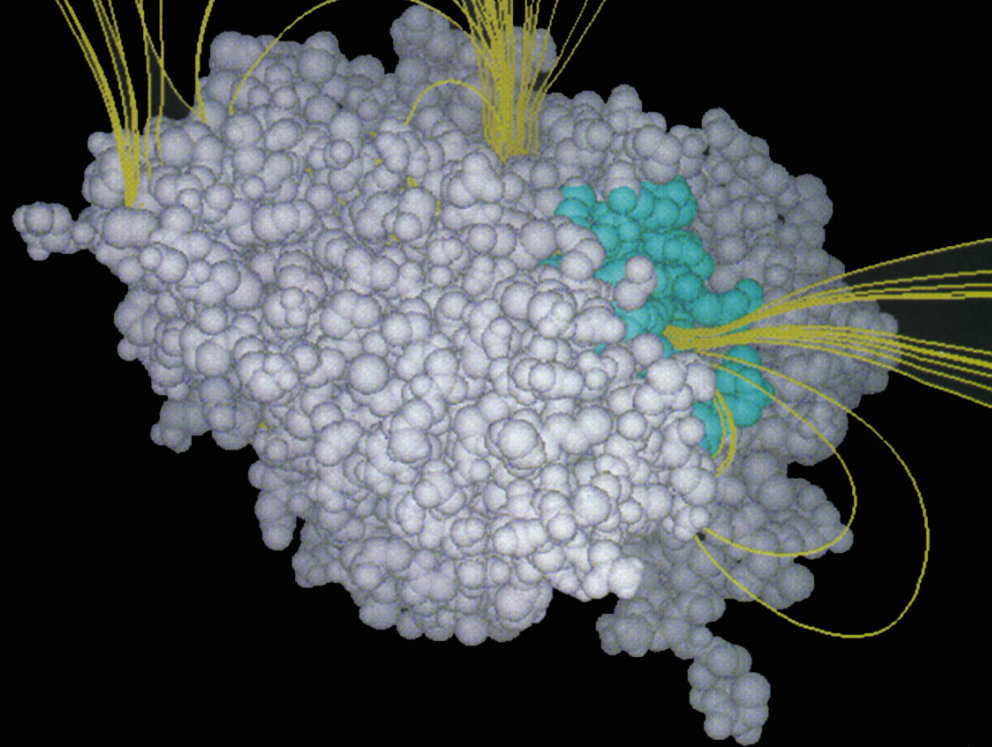
Mezoscopic models  
of molecular  
electrostatic fields  
and the energy

Studies of mezoscopic  
molecular fields  
and their relations to  
functions  
are equally important  
as structures !

AMERICAN  
ASSOCIATION FOR THE  
ADVANCEMENT OF  
SCIENCE

# SCIENCE

4 MARCH 1994 \$6.00  
VOL. 263 • PAGES 1193-1344



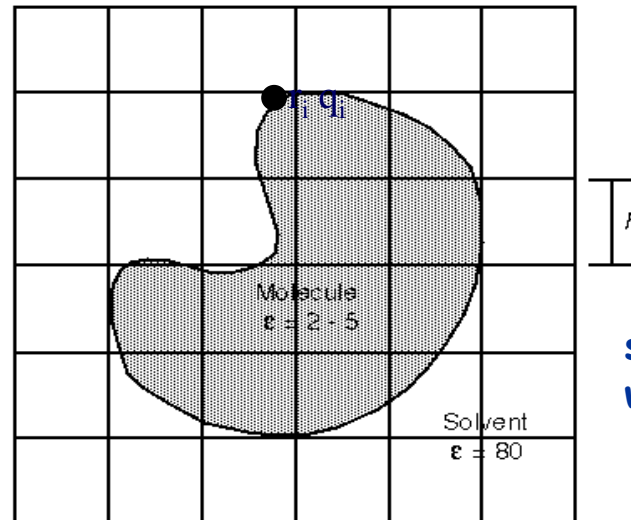
McCammon  
and coworkers

## Poisson-Boltzmann (PB) method

$$\nabla(\epsilon(\vec{r})\nabla\phi(\vec{r})) = -\sum_k q_k \delta(\vec{r}-\vec{r}_k) + \lambda(\vec{r})\epsilon(\vec{r})\kappa^2\phi(\vec{r})$$

$$\kappa^2 = \frac{2e^2 I}{\epsilon kT}$$

Debye-Huckel screening parameter, I-ionic strength



solving on a grid, or with finite elements

$$\rho_{ext}(\vec{r}) = \sum_i q_i n_i(\vec{r})$$

external ionic density

$$n_{\pm}(\vec{r}) = n_0 \exp\left\{-\frac{\pm q\phi(\vec{r})}{kT}\right\}$$

in thermodynamic equilibrium

Looking for very fast algorithms to compute  
the mean-field (mesoscopic) electrostatic energy

### Born models:

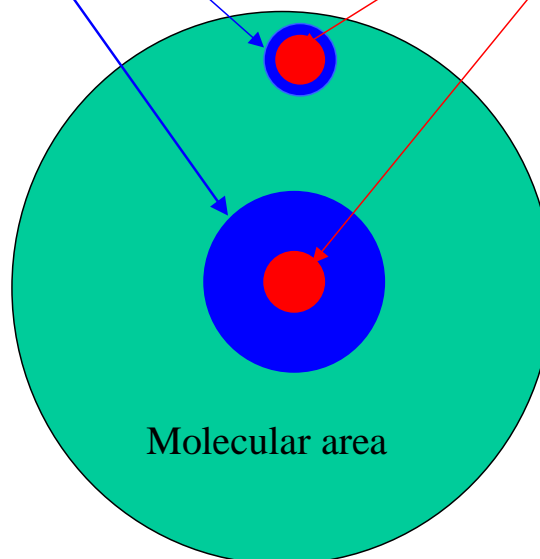
- M. Born, Z. Phys., 1, 45 (1920)
- R. Constanciel and R. Contreas, Theor. Chim. Acta, 65, 111 (1984)
- W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrikson, J. Am. Chem. Soc., 112, 6127 (1990)
- D. Bashford, D. Case, Annu. Rev. Phys. Chem., 51, 129 (2000)

If we know, so called Born-radii of atoms, we can very quickly compute the electrostatic energy. A Born radius is a geometrical property !

Born radii

and

Van der Waals radii



*The same type of atoms are characterized by different Born radii. Their values depend on geometry of the molecular system, and on localization of the atoms in the system (geometrical property). The Born radii are large inside, and are close to VdW radii on the surface.*

## Expressions for Born radii

$$\frac{1}{R_i} = \frac{1}{4\pi} \int_{\text{solvent}} \frac{1}{r^4} d^3\vec{r} \leftarrow \text{Coulomb Field appr.} \quad (\text{I})$$

A.Onufriev, D.Bashford, D.Case, *J.Phys.Chem.B*, 104,3712-3720(2000)

$$\frac{1}{R_i} = \left( \frac{3}{4\pi} \int_{\text{solvent}} \frac{1}{r^6} d^3\vec{r} \right)^{\frac{1}{3}} \leftarrow \text{Kirkwood Model} \quad (\text{II})$$

T.Grycuk, *J.Chem.Phys*, 119, 4817-4826(2003)

$$R_i = - \frac{1}{C_o A_4 + C_1 \left( \frac{3\epsilon_{ex}}{3\epsilon_{ex} + 2\epsilon_{in}} \right) A_7} + D + \frac{E}{\epsilon_{ex} + 1} \quad (\text{III})$$

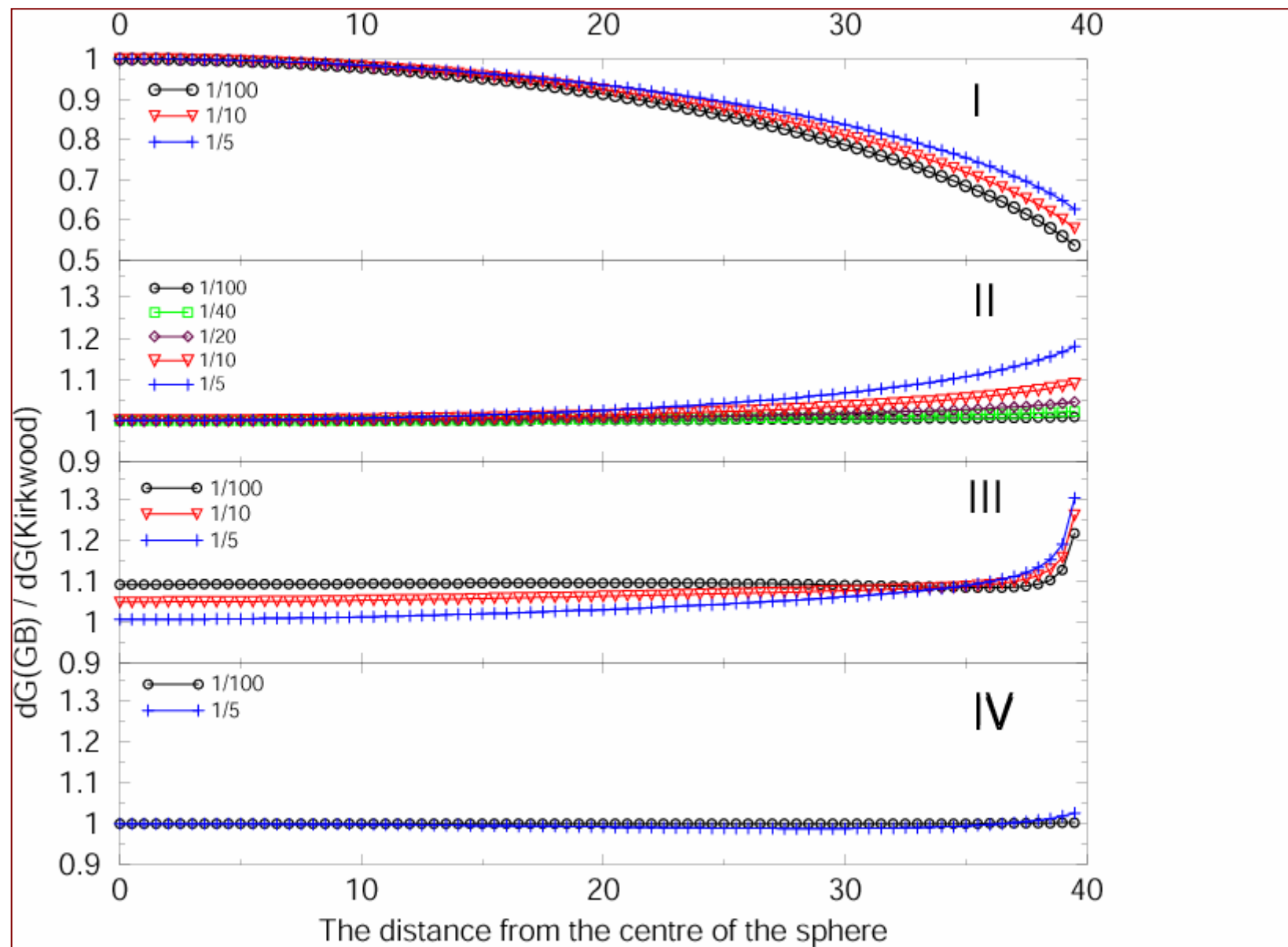
$$A_7 = \left( \frac{1}{4R_{vdW}^4} - \frac{1}{4\pi} \int_{in} \frac{1}{r^7} d^3\vec{r} \right)^{\frac{1}{4}}$$

M.Feig, W.Im, C.L.Brooks, *J.Chem.Phys.*, 120,903-911(2004)

$$\frac{1}{R_i} = \left( \frac{n-3}{4\pi} \int_{\text{solvent}} \frac{1}{r^n} d^3\vec{r} \right)^{\frac{1}{n-3}} \quad n = \frac{4.32}{\left( \frac{\epsilon_{solute}}{\epsilon_{solv}} + 0.33 \right)^{0.3}}$$

M.Wojciechowski, B.Lesyng, *J.Phys Chem, B*, 108, 18368-18376(2004) (IV)

# Ratio of the GB solvation energy to the Kirkwood solvation energy



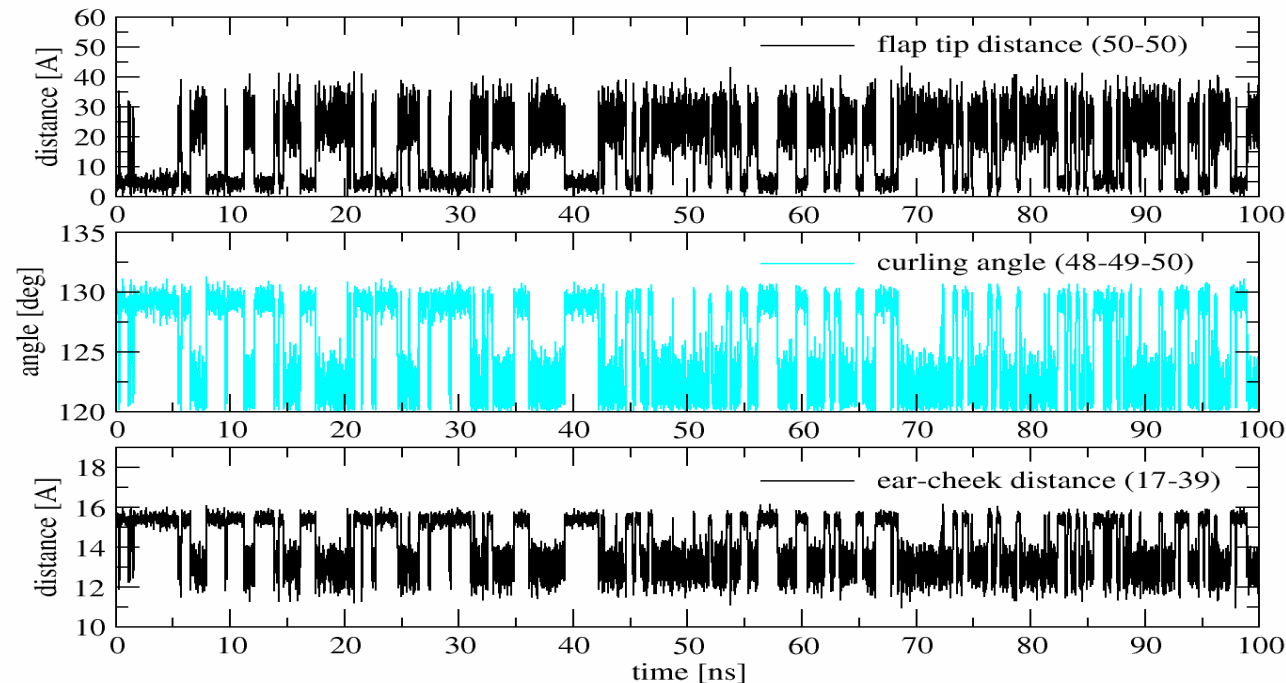
In 2003, C.W.J.Granger got  
the Nobel Prize  
for the causality analysis of  
complex economical systems.

Quite recently we have applied this strategy  
for the description  
of the dynamics (functioning)  
of complex  
molecular and nano-systems.

# Input data

- Input data: signal channels, here "observables" (snapshots) from MD or QCMD, at consecutive time-instants. These can also be projections on PCA directions

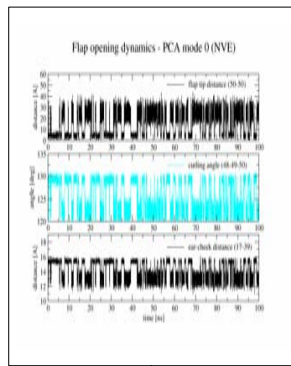
Flap opening dynamics - PCA mode 0 (NVE)



$$\mathbf{X}(t) = \{X_0(t), \dots, X_{k-1}(t)\}, \quad t = t(i) = t_0 + i \cdot \Delta t,$$

$$\mathbf{X}(t) = \sum_{i=1}^p \mathbf{A}(i) \mathbf{X}(t-i) + \mathbf{E}(t),$$

# Granger causality strategy using a multivariate autoregressive model



$$X(t) = \sum_{i=1}^p A(i)X(t-i) + E(t)$$

$\langle x_i \rangle = 0, \sigma(x_i) = 1$

$A(i) = ?$

renormalization fitting the model, MVAR

analysis in Fourier representation

$$A(f) = - \sum_{j=0}^p A(j) e^{-i2\pi f j},$$

$$H(f) = A^{-1}(f)$$

data

k channels (MD, QCMD-data),

N time - steps;

p - order of the method,

p = 20, default

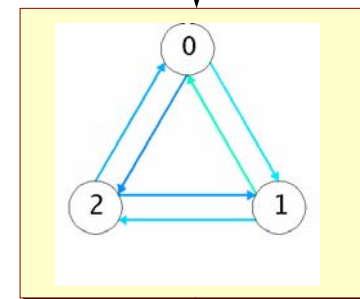
(DTF)

Directed Transfer

Functions

$$J_{ij} = \frac{1}{f_{max}} \int_0^{f_{max}} \gamma_{ij}(f) df.$$

$$\gamma_{ij}(f) = \frac{|H_{ij}|^2(f)}{\sum_{m=1}^k |H_{im}|^2(f)}$$



visualization- HTML document

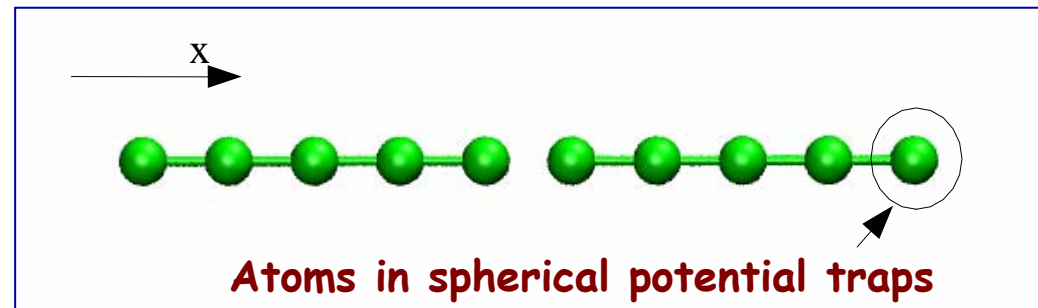
## More detailed information on test data

- Microscopic MD trajectories of a model linear set of atoms.
- Microscopic **SCC-DFTB** MD trajectories of a small, model molecule - malonaldehyde.
- MD trajectories of the HIV-1 protease  
-> signals, amplitudes from PCA.
- Coarse-grained MD trajectories of ribosome  
-> signals, amplitudes from PCA.

## Linear set of atoms - model polimer

- Dynamics using the NAMD library

$$V(\vec{r}_i, \vec{r}_j) = K_1(|\vec{r}_i - \vec{r}_j| - l_0)^2$$

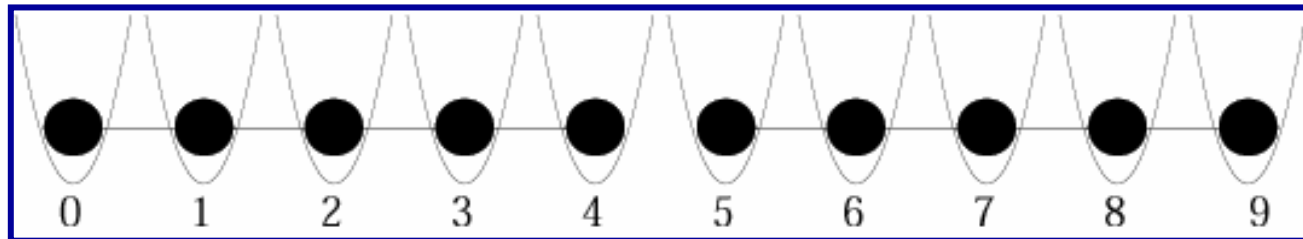


**A number of simulations carried out:**

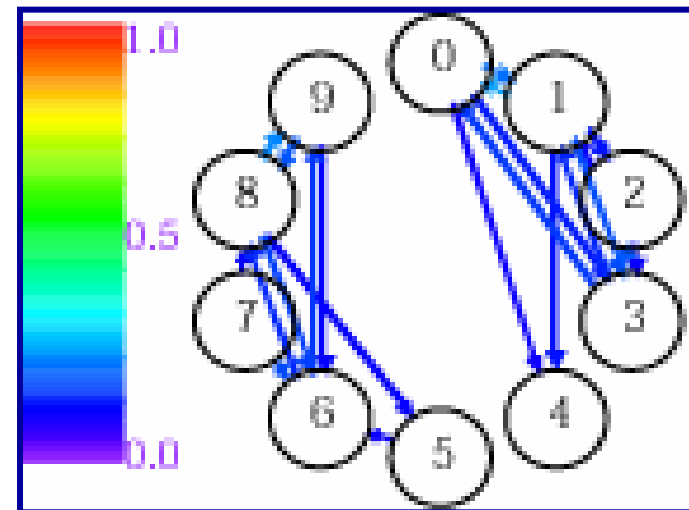
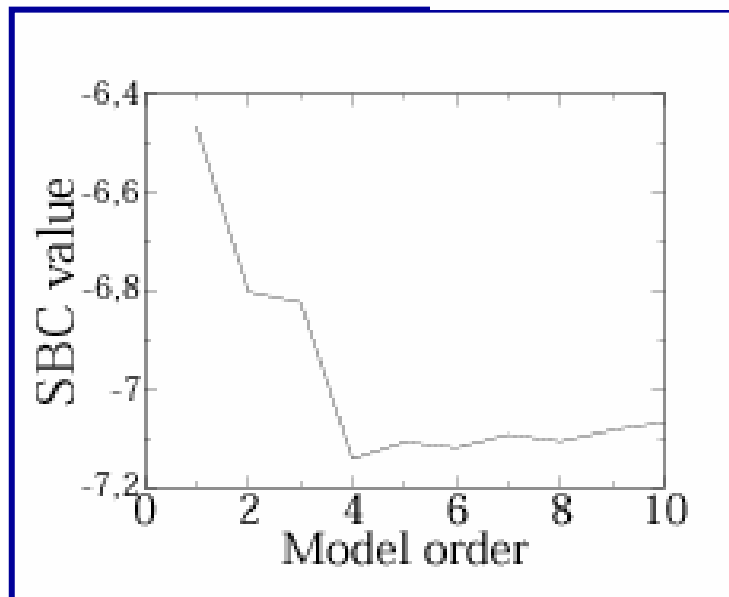
- Berendsen thermostat, harmonic traps
- Berendsen thermostat,  $V(x) = Ax^4$  traps
- Langevin dynamics

Signal for MVAR, x positions of all 10 atoms

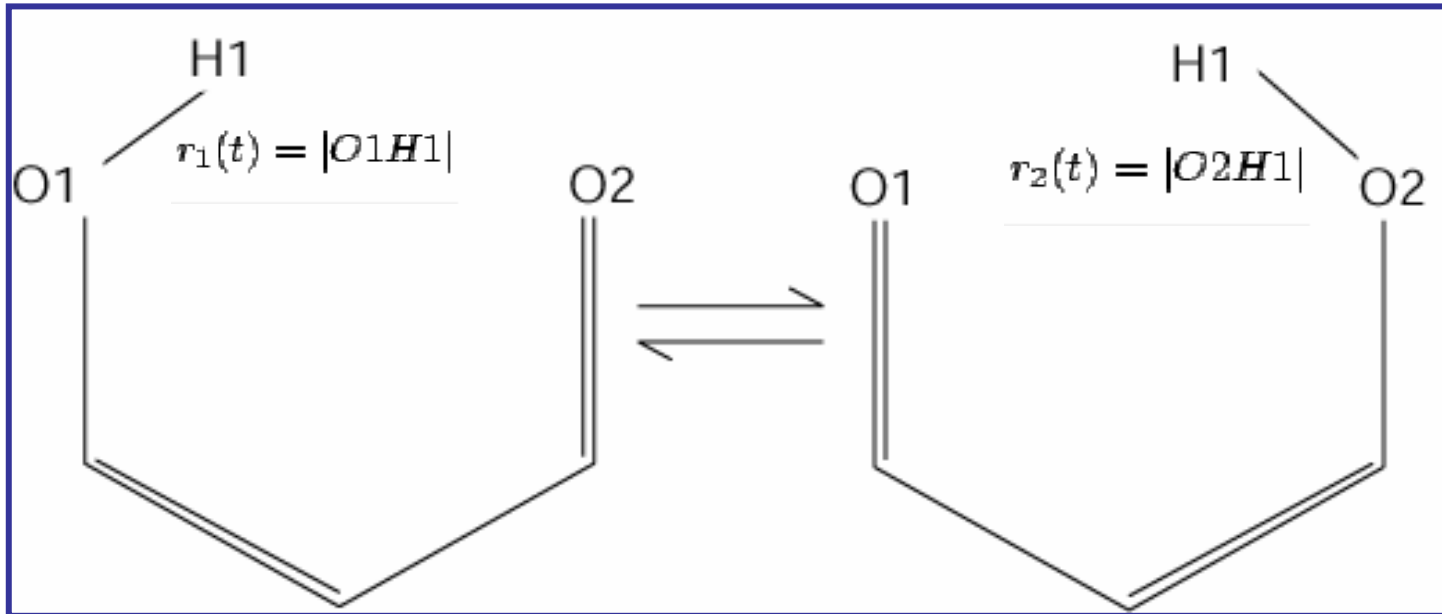
# Model polimer



$$V(\vec{r}_i, \vec{r}_j) = K_1(|\vec{r}_i - \vec{r}_j| - l_0)^2$$



# Malonaldehyde



$$X_0(t) = |O1O2|$$

$$X_1(t) = |O1H1|$$

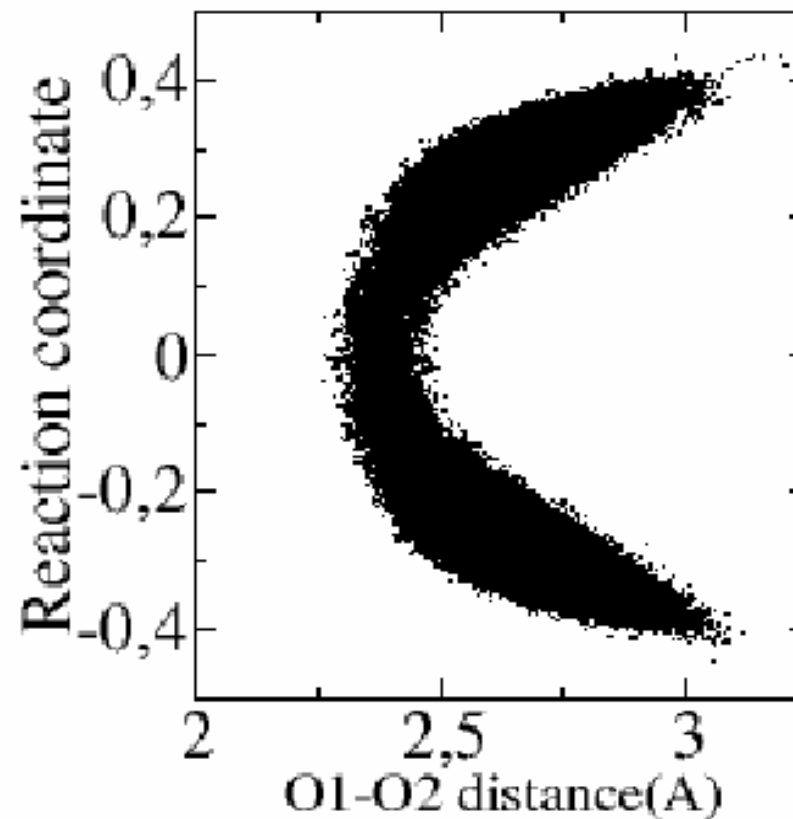
$$X_2(t) = |O2H1|$$

$$X_0(t) = |O1O2|$$

$$Y_1(t) = \frac{|r_1(t) - r_2(t)|}{r_1(t) + r_2(t)}$$

## Malonaldehyde - reaction path

$$\eta = (r_1(t) - r_2(t)) / (r_1(t) + r_2(t))$$



$$X_0(t) = |O1O2|$$

# Malonaldehyde - causality relations

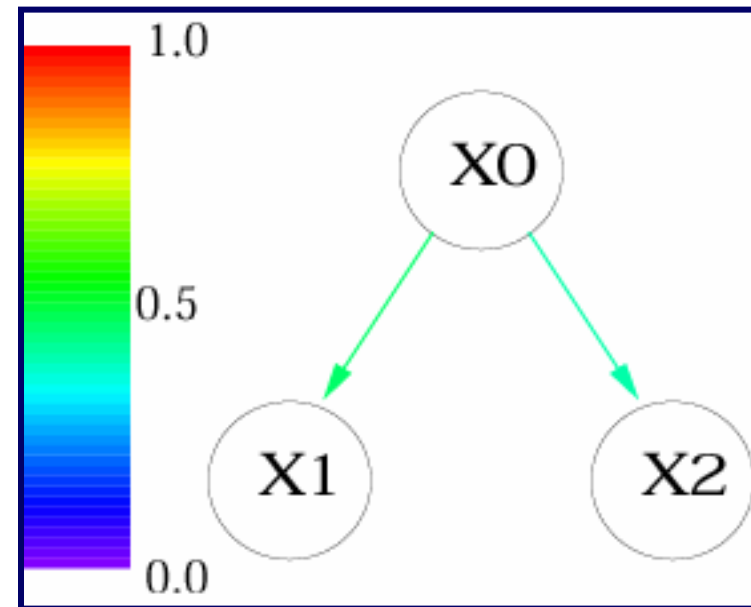
$$X_0(t) = |O1O2|$$

$$X_1(t) = |O1H1|$$

$$X_2(t) = |O2H1|$$

$$J_{ij} \in [0; 1]$$

i \ j	X0	X1	X2
X0		0.07	0.07
X1	0.46		0.11
X2	0.4	0.09	



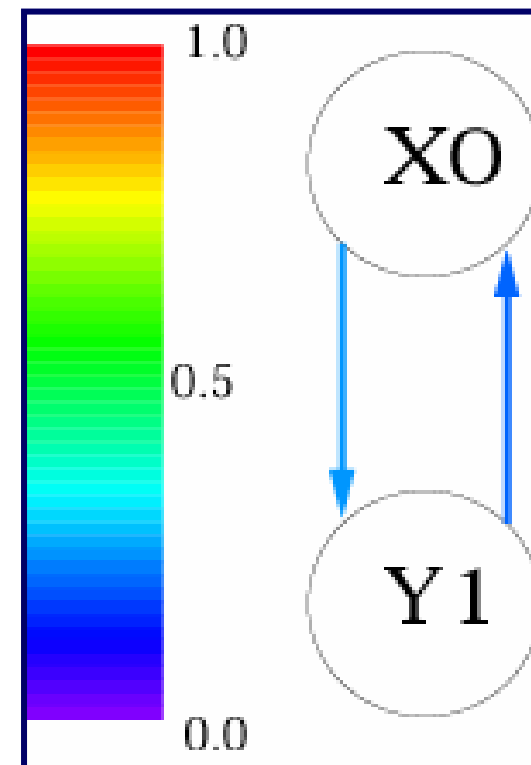
# Malonaldehyde - causality relations

$$X_0(t) = |O1O2|,$$

$$Y_1(t) = \frac{|r_1(t) - r_2(t)|}{r_1(t) + r_2(t)},$$

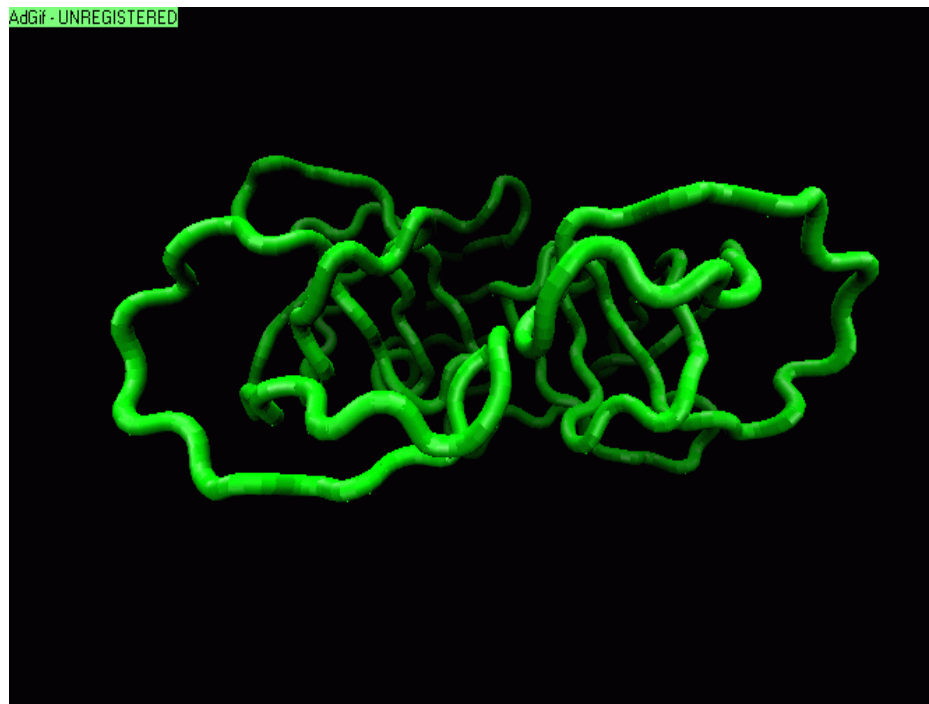
$$J_{ij} \in [0; 1]$$

i \ j	X0	Y1
X0		0.19
Y1	0.24	



# MD data for HIV-1protease: PCA data

- One of the PCA modes of HIV-protease: opening/closing of flaps (MD data from dr V.Tozzini)

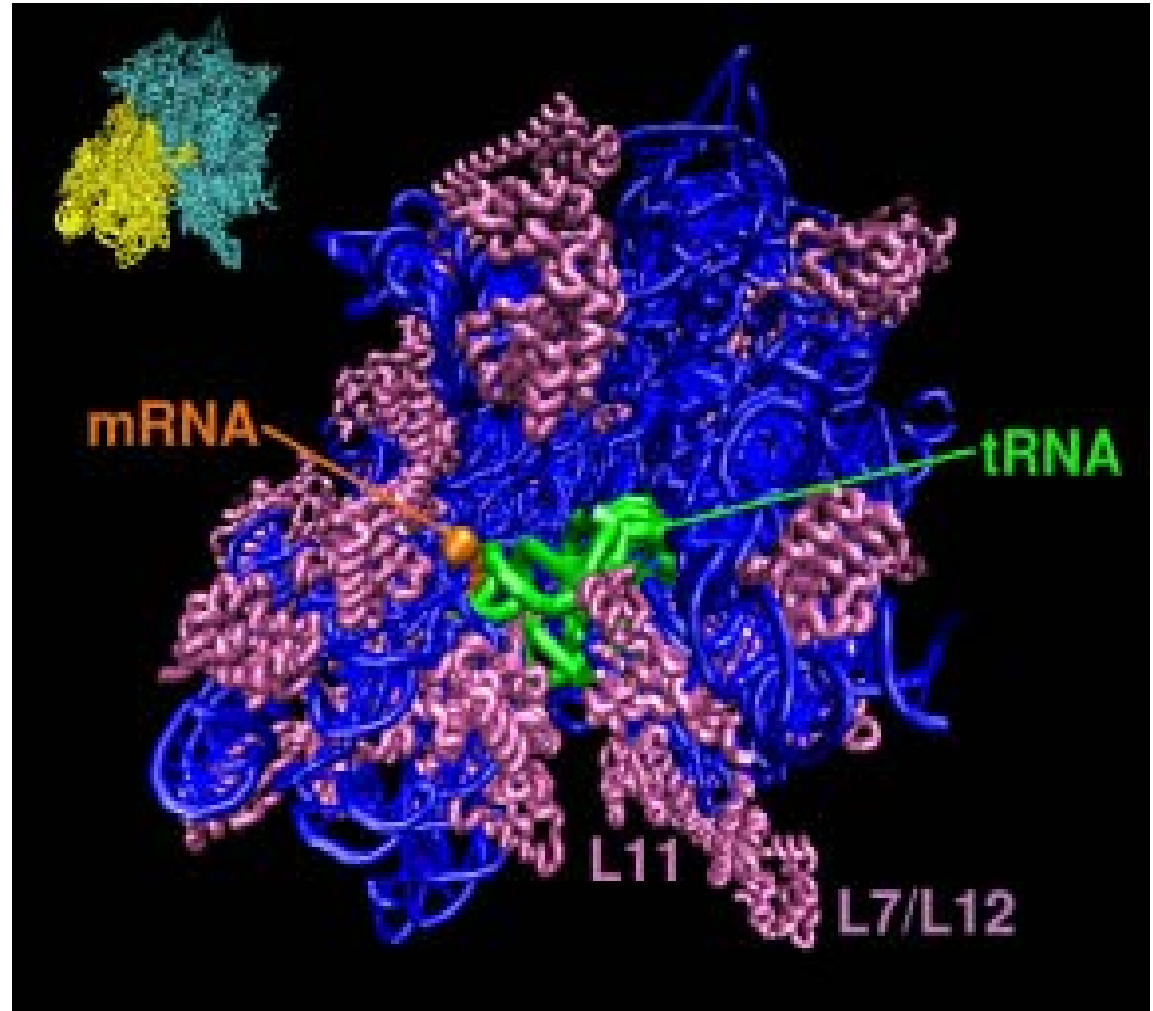


- Signals: amplitudes of projections of the trajectories on PC directions

Mode no. 0

# Dynamics of ribosome

- Coarse grain dynamics:  
objects:
  - P atoms from RNA,  
C-alpha from proteins.
- PCA projections as signals



## Conclusions & Vision

- An integrated approach, using applied informatics and multiscale modelling methods, allows to much better understand complex biomolecular and material structures and processes (important element of the e-science strategy).
- This influences positively development of advanced, practical applications, such as structure prediction, materials design molecular design including drug design, or biotechnological solutions and strategies.
- Integration of informatics, chemical and physical approaches require, however, new interdisciplinary educational programs. In particular, coupling of higher education and interdisciplinary research is highly required for high-quality science, and emerging technologies.
- E-science platforms should be based on novel WEB and GRID technologies.

## Acknowledgements:

### Coworkers

Dr. Joanna Trylska  
Dr. Maciej Geller  
Dr. Pawel Grochowski  
Dr. hab. Piotr Bala

### PhD students:

Jarek Kalinowski  
Piotr Kmiec  
Magda Gruzziel  
Michał Wojciechowski  
Adam Gorecki

### Collaboration:

Prof. T. Frauenheim  
Dr. M. Elstner

SCC-DFTB, University of Paderborn, Germany

Prof. D. Truhlar  
Dr. J. Thompson  
Dr. C. Cramer

CM3-charges, Minnesota Solvation Data Base  
University of Minnesota, USA

Prof. J.A.McCammon

Titration of proteins  
University of California at San Diego, USA

- *Simulations of adiabatic QD simulations for porphicine carried out at NIC, Juelich.*
- *Studies supported in part by "European CoE for Multiscale Biomolecular Modelling, Bioinformatics and Applications", ICM, Warsaw University.*

## Causality studies:

### *Inspirations:*

Prof. K. Blinowska

Dr P. Durka

Dr R. Kuś

Dr M. Kamiński

### *Test MD data from:*

Dr V. Tozzini (SNS, Pisa) - HIV-1 protease, ribosome

Dr J. Trylska (ICM) - HIV-1 protease, ribosome

L. Walewski, Prof. P. Bała (ICM) - malonoaldehyde