Modelling Principles and Coarse Graining Methods in Computational Drug Design

Montreal, May 23 - June 2, 2006

Susanna Kube

Zuse Institute Berlin (ZIB)

Project Group

ZIB Scientific Computing:

Dept. Numerical Analysis and Modelling:

Peter Deuflhard, Frank Cordes, Marcus Weber, Ulrich Nowak, Susanna Kube, Lionel Walter

Dept. Scientific Visualization:

Hans-Christian Hege, Daniel Baum, Johannes Schmidt-Ehrenberg

Cooperation:

FU Biocomputing: **Christof Schütte**, Evelyn Dittmer, ...

FU Junior Research Group "Computational Physiology": Wilhelm Huisinga

DFG Research Center **"Matheon"**

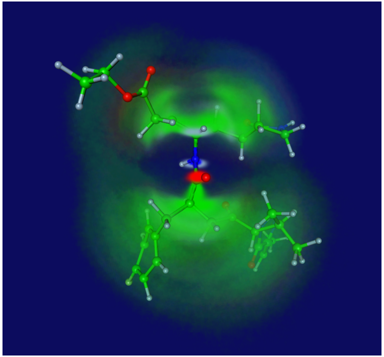
(Berlin Center for Genome-based Bioinformatics (BCB))

Virtual Lab

numerical ODEs:
molecular dynamics

nonlinear dynamics:
almost invariant sets

biochemistry:
RNA molecules,
prions,
viruses



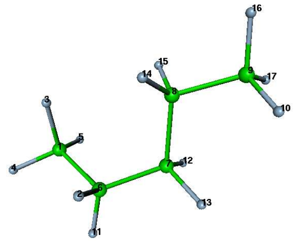
statistics:
nearly uncoupled
Markov chains

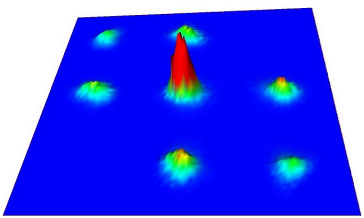
computer science:
visualization

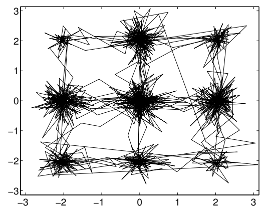
numerical linear algebra,
nonlinear optimization:
Perron cluster analysis

Basic Idea

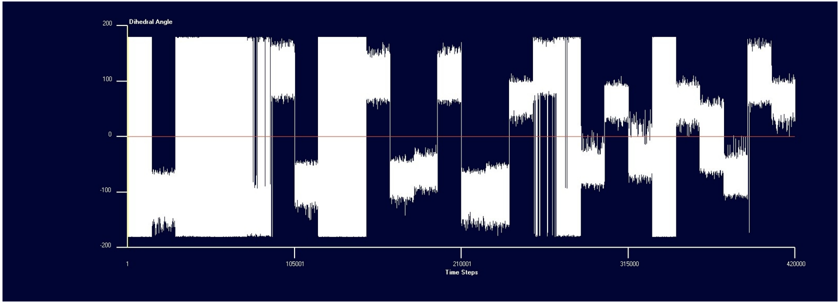
- molecular dynamics simulation
- trajectories in the space of essential dihedral angles
- identification of conformations
- average life times, transition rates/probabilities ?







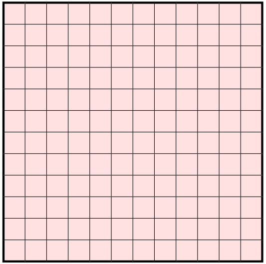
Timeseries Analysis

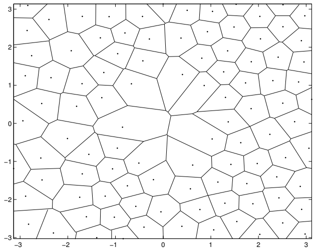


- timeseries as a Markov chain $\{X(t_n)\}_{n\in\mathbb{N}}$ in a finite state space $E = \{1, \ldots, N\}$
- discretization of the dihedral space by a set of basis functions $\{\varphi_i(\theta)\}_{i=1}^N$

$$\sum_{i=1}^{N} \varphi_i(\theta) = 1, \qquad 0 \leq \varphi_i(\theta) \leq 1$$

Discretization





Eigenproblem

A function $u(q)$ is called invariant under the Transfer Operator T^τ if

$$T^\tau u(q) \approx u(q)$$

Eigenvalue problem:

$$T^\tau u = \lambda u, \qquad \lambda \approx 1$$

Galerkin approach:

$$u(q) = \sum_{i=1}^{N} \alpha_i \varphi_i(q)$$

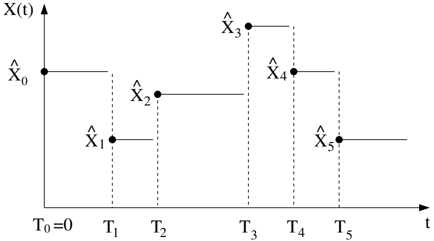
$$\sum_{i=1}^{N} \alpha_i \langle T^\tau \varphi_i, \varphi_j \rangle_{\mathcal{Q}} = \lambda \sum_{i=1}^{N} \alpha_i \langle \varphi_i, \varphi_j \rangle_{\mathcal{Q}}, \quad \forall\, j$$

$$\boxed{\overline{P}\alpha = \overline{S}\alpha\lambda}$$

T^τ selfadjoint in $L^2_{\mathcal{Q}}$ $\quad\rightarrow\quad$ $\overline{P}$ symmetric (as well as $\overline{S}$)

Alternative Approach

- modeling assumption: $\{X(t)\}$ is a regular jump homogeneous Markov chain



- then $P(t)(i,j) = \mathbb{P}(X(t+s)=j|X(s)=i)$ is a transition semigroup
  - $P(t)$ is a stochastic matrix
  - $P(0) = id$
  - $P(t+s) = P(t)P(s)$ (Chapman-Kolmogorow)

- Then there exists the infinitesimal generator

$$Q = \lim_{t \mapsto 0+} \frac{P(t) - id}{t}$$

- finite state space: $P(t) = \exp(tQ)$
- representation in terms of the embedded Markov chain

$$Q = H(K - id)$$

  - $h_i = -q_{ii}$: inverse average life time
  - $k_{ij} = q_{ij}/h_i$: probability of jumping to state j when leaving state i

- diagonalizable Q (detailed balance)

$$Q = X \Lambda X^{-1}$$

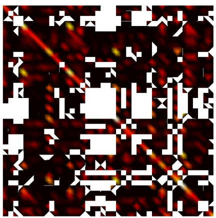
$$P(t) = \exp(tQ) = X \exp(t\Lambda) X^{-1} = X \text{diag}(\exp(t\lambda_1), \ldots, \exp(t\lambda_n)) X^{-1}$$

Eigenvalue cluster of P at 1 corresponds to an eigenvalue cluster of Q at 0.

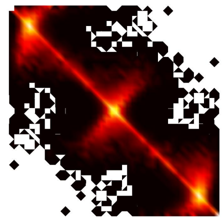
PCCA+

input:

- $P(i,j) =$ transition probability $i \to j$
- $Q(i,j) =$ transition rate $i \to j$
- real-valued eigenvectors!



$\longrightarrow$



PCCA+

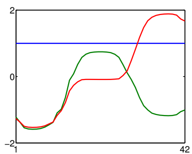
T_1	0	0
0	T_2	0
0	0	T_3

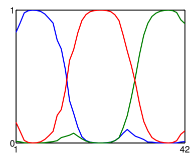
$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2.02 & -0.55 \\ 1 & -2.02 & -0.55 \\ 1 & 0.48 & -0.91 \\ 1 & 0.48 & -0.91 \\ 1 & 0.50 & 1.24 \\ 1 & 0.50 & 1.24 \end{pmatrix} \cdot \begin{pmatrix} 0.20 & 0.41 & 0.39 \\ -0.40 & 0.33 & 0.07 \\ 0.00 & -0.47 & 0.47 \end{pmatrix}$$

$\tilde{T}_1$	E_{12}	E_{13}
E_{32}	$\tilde{T}_2$	E_{23}
E_{31}	E_{32}	$\tilde{T}_3$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2.02 & -0.55 \\ 1 & -1.52 & -0.62 \\ 1 & 0.23 & -0.66 \\ 1 & 0.48 & -0.91 \\ 1 & 0.50 & 1.03 \\ 1 & 0.50 & 1.24 \end{pmatrix} \cdot \begin{pmatrix} 0.20 & 0.41 & 0.39 \\ -0.40 & 0.33 & 0.07 \\ 0.00 & -0.47 & 0.47 \end{pmatrix}$$

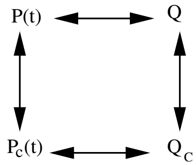
$$\boxed{\chi = X\mathcal{A}}$$





Coarse Graining

Can the dynamic process be described by a Markov chain in the conformational state space $E = \{1, \ldots, N_c\}$?



- (N_c, N_c)-coupling matrix:

$$\begin{pmatrix} p_{11} & \cdots & p_{1N_c} \\ \vdots & \ddots & \vdots \\ p_{N_c1} & \cdots & p_{N_cN_c} \end{pmatrix}$$

p_{ij} : probability to move from i to j

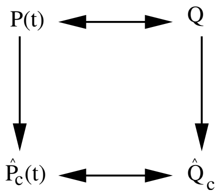
$p_{ii} \approx 1$: probability to stay within i

$p_{ii} > 0.5$

$$\begin{pmatrix} q_{11} & \cdots & q_{1N_c} \\ \vdots & \ddots & \vdots \\ q_{N_c1} & \cdots & q_{N_cN_c} \end{pmatrix}$$

q_{ij} : rate for transition from i to j

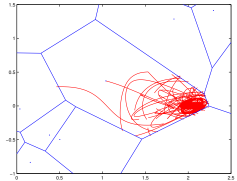
$-1/q_{ii}$: average life time of conformation i



- correct dynamic behavior: $\dot{\mathbf{x}}(t) = Q^\top \mathbf{x}(t), \quad \dot{\mathbf{y}}(t) = \hat{Q}_c^\top \mathbf{y}(t), \quad ||R\mathbf{x}(t) - \mathbf{y}(t)|| = 0, \, \forall\, t \in [0, \infty)$
- correct stationary distribution: $\tilde{\pi}^\top \hat{P}_c(t) = \tilde{\pi}^\top, \quad \tilde{\pi}^\top \hat{Q}_c = 0, \quad \tilde{\pi} = R\pi$
- correct row sum: $\sum_j \hat{P}_c(i,j) = 1, \quad \sum_j \hat{Q}_c(i,j) = 0$
- consistency: $\hat{P}_c(t) = \exp(t\hat{Q}_c)$
- negative off-diagonal entries!

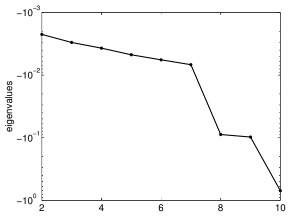
Example

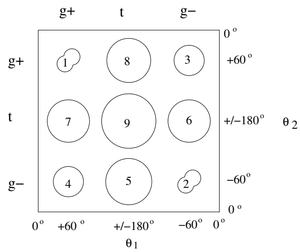
1. Discretization
2. Molecular Simulation
   - Merck Molecular Force Field
   - generation of 3000 points per cell according to the Boltzmann distribution by hybrid Monte-Carlo sampling with umbrella strategies and Gelman-Rubin convergence indicator
   - propagation by MD until the trajectories leave their starting cell

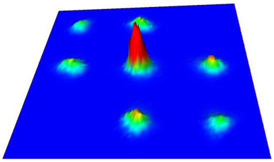


Example

3. Cluster Analysis

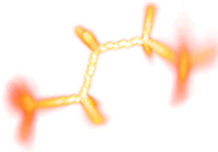




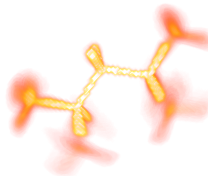


$$\text{weights} = \{0.0036, 0.0032, 0.0640, 0.0680, 0.1248, 0.1232, 0.1140, 0.1567, 0.3426\}$$

$$h^{-1} = \{10.39, 11.38, 361.68, 334.35, 198.91, 227.31, 172.40, 235.74, 283.64\} ps$$



(g+/t)



(t/t)

(t/g+)

Conclusion

- many challenging mathematical topics
- balance between modelling assumptions and numerical errors
- Visit our homepage: http://www.zib.de/MDGroup

Thank you for your attention!!!