

SUSANNA KUBE, MARCUS WEBER

# **Conformation Kinetics as a Reduced Model for Transition Pathways**

# Conformation Kinetics as a Reduced Model for Transition Pathways

Susanna Kube and Marcus Weber\*

## Abstract

The complexity of molecular kinetics can be reduced significantly by a restriction to metastable conformations which are almost invariant sets of molecular dynamical systems. With the Robust Perron Cluster Analysis PCCA+, developed by Weber and Deuffhard [6], we have a tool available which can be used to identify these conformations from a transition probability matrix. This method can also be applied to the corresponding transition rate matrix which provides important information concerning transition pathways of single molecules. In the present paper, we explain the relationship between these two concepts and the extraction of conformation kinetics from transition rates. Moreover, we show how transition rates can be approximated and conclude with an example.

## 1 Introduction

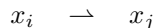
The understanding of transition pathways between different conformations of a molecule is an important issue in structural biology. Although the restriction of degrees of freedom to a few dihedral angles significantly reduces the complexity of the problem, this is still very difficult. Often, scientists are interested in single pathways, for example those over lowest energy barriers [2]. On the other hand, it is well known that molecular kinetics is not purely deterministic. All kinds of trajectories could appear, some with higher probability than others. Therefore it seems natural to consider population densities. Starting with a given probability density in position space, we are interested in the evolution of the density to figure out intermediate states.

A description of molecular dynamics based on single positions in phase space is unfeasible for large molecules. Therefore we work with a set concept based on metastable conformations as introduced in [15]. First, we reduce the position space to a number of  $N$  states represented by basis functions [18] or boxes [16]. Then we construct a transition probability matrix  $P \in \mathbb{R}^{N \times N}$  or a transition rate matrix  $Q \in \mathbb{R}^{N \times N}$  and cluster states into metastable conformations by applying PCCA+. In contrast to transition probabilities, the transition rate matrix  $Q$  is sparse because there are only transitions between neighboring states.

---

\*Zuse Institute Berlin (ZIB), Takustraße 7, D-14195 Berlin, Germany

Therefore, at the same discretization level, it is more efficient to apply numerical routines to rate matrices instead of probability matrices. Furthermore, transition rates provide important chemical information concerning transition pathways of single molecules. Given an initial weighting  $x_A$  of the states, one can compute the corresponding weights and the spatial configuration density at a next time step. This is the desired dynamic in configuration space, which is not based upon single molecules but upon ensembles. The entry  $q_{ij}$ ,  $i \neq j$ , can be considered as the reaction rate of the monomolecular reaction



where  $x_i$  stands representatively for the weight or “concentration” of state  $i$ .

The rate matrix  $Q$  can be obtained directly from the transition probability matrix  $P$ , but it also offers an alternative approach to identify metastable conformations if the transition probability matrix is not available or difficult to compute. Furthermore, we are able to reduce our model not only to a set of basis functions whose number can be very large, but also to the few metastable sets which contain all important information about the system.

## 2 From Transition Probabilities to Transition Rates

### 2.1 Markov Jump Processes

Throughout the paper, we consider an homogenous Markov process  $\{X(t) : t \geq 0\}$  on a finite state space  $S = \{0, 1, 2, \dots, N\}$  with transition probability function

$$p(t, i, j) = \mathbf{P}[X(t) = j | X(0) = i], \quad i, j \in S.$$

Such processes are also called continuous-time homogenous finite Markov chains. The Markov process is assumed to satisfy the following regularity conditions:

$$p(0, i, j) = \delta_{ij}, \tag{1}$$

$$\lim_{t \rightarrow 0^+} p(t, i, j) = \delta_{ij} \quad \forall i, j \in S. \tag{2}$$

This guarantees that the realizations are right continuous functions. Due to the fact that the state space is discrete, they are step functions. Therefore such processes are called Markov jump processes.

### 2.2 Infinitesimal Generator

Under the above conditions, with  $P(t) = (p(t, i, j))_{i, j \in S}$ , the limit

$$Q = \lim_{t \rightarrow 0^+} \frac{P(t) - Id}{t} \tag{3}$$

exists and defines the infinitesimal generator  $Q = (q(i, j))_{i, j \in S}$  with  $-\infty \leq q(i, i) \leq 0 \leq q(i, j) < \infty$  [12][3].

Markov chains on a finite state space meet

$$-q(i, i) < \infty \quad (\text{stability})$$

and

$$-q(i, i) = \sum_{j \neq i \in S} q(i, j) \quad (\text{conservation}).$$

**Definition 2.1** *The random variable*

$$T_i(t) = \inf(s \geq 0 : X(t + s) \neq i, X(t) = i)$$

*is called holding time in state  $i$ .*

If  $X(0) = i$ ,  $X(T_i)$  represents the state the Markov chain visits just after leaving state  $i$ . It can be shown [12] that  $T_i$  decays exponentially in  $t$ , i.e.

$$\mathbf{P}[T_i > s] = \exp(-\lambda(i)s), \quad \forall s > 0.$$

$\lambda(i)$  is called the jump rate associated with the state  $i \in S$ . The average life time of the state  $i$  is given by

$$\mathbf{E}[T_i] = \frac{1}{\lambda(i)}.$$

**Proposition 2.2** [12] *For a finite Markov chain  $X(t)$  in continuous time with infinitesimal generator  $Q$ , the following equations are satisfied:*

1.

$$-\frac{q(i, j)}{q(i, i)} = \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[X(t) = j | X(0) = i]}{\mathbf{P}[X(t) \neq i | X(0) = i]} = \mathbf{P}[X(T_i) = j | X(0) = i],$$

2.

$$q(i, i) = -\lambda(i).$$

In words:  $-\frac{q(i, j)}{q(i, i)}$  is the conditional probability of a transition from state  $i$  to state  $j$  given that the process starts in  $i$ , and the diagonal of  $Q$  contains the negative rates.

**Proof:** From (3) we obtain

$$\begin{aligned} q(i, i) &= \lim_{t \rightarrow 0^+} \frac{p(t, i, i) - 1}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[T_i > t] - 1}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{\exp(-\lambda(i)t) - 1}{t} \\ &= \frac{d}{dt} \exp(-\lambda(i)t)|_{t=0} \\ &= -\lambda(i). \end{aligned}$$

Furthermore,

$$\begin{aligned} q(i, i) &= \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[X(s+t) = i | X(s) = i] - 1}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[X(s+t) \neq i | X(s) = i]}{t} \end{aligned}$$

and

$$\begin{aligned} q(i, j) &= \lim_{t \rightarrow 0^+} \frac{p(t, i, j)}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[X(s+t) = j | X(s) = i]}{t}. \end{aligned}$$

Consequently

$$\begin{aligned} -\frac{q(i, j)}{q(i, i)} &= \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[X(s+t) = j | X(s) = i]}{\mathbf{P}[X(s+t) \neq i | X(s) = i]} \\ &= \lim_{t \rightarrow 0^+} \frac{\mathbf{P}[X(t) = j | X(0) = i]}{\mathbf{P}[X(t) \neq i | X(0) = i]} \end{aligned}$$

because of the homogeneity. Moreover,

$$\begin{aligned} -\frac{q(i, j)}{q(i, i)} &= \lim_{t \rightarrow 0^+} \mathbf{P}[X(s+t) = j | X(s) = i, X(s+t) \neq i] \\ &= \mathbf{P}[X(T_i) = j | X(0) = i] \end{aligned}$$

□

$P(t)$  and  $Q$  are related by the forward and backward Kolmogorov equation

$$\frac{dP(t)}{dt} = Q P(t) = P(t) Q.$$

Together with the initial condition  $P(0) = Id$ , this gives the formal solution

$$P(t) = \exp(tQ) = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}, \quad t \geq 0. \quad (4)$$

### 2.3 Stationary Distribution and Almost Invariant Sets

One can also deduce the evolution equation for an arbitrary initial distribution  $(\mu_0(i))_{i \in S}$  of the Markov jump process,

$$\mu^\top(t) = \mu_0^\top P(t),$$

with  $\mu(t) = (\mu_i(t))_{i \in S} = \{\mathbf{P}[X(t) = i]\}$ . Multiplying Kolmogorov's equation with  $\mu_0$ , we obtain

$$\frac{d\mu^\top(t)}{dt} = \mu^\top(t)Q.$$

Consequently, the infinitesimal generator  $Q$ , together with the initial distribution  $\mu_0$ , completely characterize the Markov jump process.

A stationary distribution is any probability distribution  $\pi$  with

$$\pi^\top = \pi^\top P(t), \quad \forall t \geq 0.$$

Under certain conditions (irreducibility, positive recurrence and regularity of the Markov jump process), the following statements are equivalent [12],[3]:

- There exists a stationary distribution  $\pi$  of  $P$ .
- $\pi$  meets  $\pi^\top Q = 0$ .

This is especially true for an homogenous Markov process on a finite state space [3]. A Markov process which possesses such a stationary distribution  $\pi$  is called ergodic.

Besides the stationary distribution, we are interested in almost invariant sets, i.e. unions of subsets for which the probability to stay within is close to one. The characteristic functions of such invariant sets are closely connected to the structure of eigenvectors  $x_i$  of  $P$  corresponding to eigenvalues  $\lambda_i \approx 1$  [5]

$$P(t)x = \lambda x, \quad \lambda \approx 1.$$

The Perron Cluster Algorithm [6] constructs the almost characteristic functions as linear combination of these eigenvectors. It can also be applied to  $Q$  because the eigenvectors of  $P$  and  $Q$  are the same [9]. Suppose that the infinitesimal generator  $Q$  is diagonalizable by some nonsingular matrix  $X$ .

$$Q = X\Theta X^{-1} = X \text{diag}(\theta_1, \dots, \theta_p) X^{-1}.$$

Then

$$P(t) = \exp(tQ) = X \exp(t\Theta) X^{-1} = X \text{diag}(\exp(\tau\theta_1), \dots, \exp(\tau\theta_p)) X^{-1}.$$

Since  $\exp(0) = 1$ , an eigenvalue cluster of  $P(t)$  at 1 corresponds to an eigenvalue cluster of  $Q$  at 0.

## 2.4 Reversibility

Let  $\{X(t)\}$  be an ergodic Markov jump process with state space  $S$ , transition probability matrix  $P$  and stationary distribution  $\pi$ . The Markov process is said to be reversible in time, if the transition rate out of an arbitrary state  $j$  equals the transition rate into state  $j$ , i.e.

$$\sum_i \pi_i p(t, i, j) = \sum_i \pi_j p(t, j, i), \quad \forall j \in S.$$

Obviously, a sufficient condition is given by

$$\pi_i p(t, i, j) = \pi_j p(t, j, i), \quad \forall i, j \in S \tag{5}$$

which is known as detailed balance condition.

Reversibility of the Markov process implies the existence of a stationary distribution and vice versa,

$$\pi^\top = \pi^\top P(t) \leftrightarrow \pi_j = \sum_i \pi_i p(t, i, j) \leftrightarrow \pi_j \sum_i p(t, j, i) = \sum_i \pi_i p(t, i, j).$$

The following propositions indicate a relationship between reversibility of the transition probability matrix  $P$  and reversibility of the infinitesimal generator  $Q$ .

**Proposition 2.3** [3] *If  $Q$  satisfies the detailed balance condition, the finite Markov jump process  $\{X(t)\}$  with transition probability matrix  $P(t)$  is reversible.*

**Proof:** Let  $D$  be the diagonal matrix whose elements are the stationary probabilities  $\pi_i$ ,  $i \in S$ . In matrix form, (5) reads

$$DP = P^\top D.$$

Hence,  $DP$  is symmetric. For a finite state space, pre- and post-multiplication with  $D^{-1/2}$  preserves symmetry. Consequently, a finite Markov chain is reversible, if  $D^{1/2}PD^{-1/2}$  is symmetric.

From (4) it follows that

$$D^{1/2}P(t)D^{-1/2} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \left( D^{1/2}QD^{-1/2} \right)^n.$$

Since the right hand side is symmetric ( $Q$  meets detailed balance), the left hand side is symmetric, too. Hence, reversibility of  $Q$  implies reversibility of  $P$ .  $\square$

**Proposition 2.4** *The infinitesimal generator  $Q$  of a finite Markov jump process satisfies the balance condition iff the stationary distribution  $\pi$  satisfies the stationary equation  $\pi^\top Q = 0$ .*

**Proof:**

$$\begin{aligned} (\pi^\top Q)_j &= \sum_{i \neq j} \pi_i q_{ij} + \pi_j q_{jj} = \sum_{i \neq j} \pi_i q_{ij} - \pi_j \sum_{i \neq j} q_{ji} \\ &= \sum_i \pi_i q_{ij} - \sum_i \pi_j q_{ji} = 0. \end{aligned}$$

$\square$

### 3 Conformation Kinetics

Each vector  $\mathbf{x} \in \mathbb{R}^N$  can be considered as a vector of coefficients which occur in the representation of a function defined on  $\Omega$  as a linear combination of certain basis functions  $\{\varphi_i\}_{i=1}^N$  which form a partition of unity. For example, if  $\Omega$  is divided into non-overlapping subsets  $\{A_i\}$ , the basis functions could be the characteristic functions of these sets and the vector  $\mathbf{x}$  contains the values of the piecewise constant function on these sets.

The application of PCCA+ [6] delivers membership functions  $\chi_i = \sum_{j=1}^N c_{ji} \varphi_j$  in terms of vectors  $\{\mathbf{c}_i\}_{i=1}^{N_C}$  with  $0 \leq c_{ij} \leq 1$  where  $N_C$  corresponds to the number of clusters. These vectors are obtained by a linear transformation of eigenvectors  $\{\mathbf{x}_i\}_{i=1}^{N_C}$  of  $Q$  according to eigenvalues close to zero. Set  $C = (\mathbf{c}_1, \dots, \mathbf{c}_{N_C})$ .

$$C = XA, \quad QX = \Lambda X, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_{N_C}), \quad \lambda_i \approx 0.$$

As an important property, the rows of  $C$  sum to 1. By this transformation, each basis function  $\varphi_i$  is assigned to a cluster  $j \in \{1, \dots, N_C\}$  with the weight  $c_{ij}$ .

In the following, we consider the discrete equation

$$\dot{\mathbf{x}} = Q^\top \mathbf{x}. \quad (6)$$

Densities with respect to the basis functions  $\{\varphi_i\}_{i=1}^N$  are given as vectors  $\mathbf{x}^f \in \mathbb{R}^N$  with  $\sum_i x_i^f = 1$ . In the following they are called full densities. As a reduced density, we consider a vector  $\mathbf{x}^r \in \mathbb{R}^{N_C}$ ,  $\sum_i x_i^r = 1$ . These densities represent the weights of the clusters. They are obtained by a projection of a full density onto the clusters, i.e.

$$\mathbf{x}^r = C^\top \mathbf{x}^f. \quad (7)$$

Indeed, multiplication with  $C^\top$  preserves the sum of the vector elements,

$$\sum_i x_i^r = \sum_i \sum_j c_{ij}^\top x_j^f = \sum_j (\sum_i c_{ji}) x_j^f = \sum_j x_j^f.$$

Equation (6) is not very interesting because the kinetics simply converges towards the equilibrium distribution  $\pi$ . If one is interested in a simulation of a transition from metastable conformation  $A$  to a metastable conformation  $B$  and the corresponding transition behavior, then (6) has to be solved as an initial value problem with initial distribution  $\mathbf{x}_A$  and an absorbing end state given by the distribution  $\mathbf{x}_B$ . Chemically, one would permanently eliminate conformation  $B$  out of the ensemble in order to push the reaction into the direction of this product. Mathematically this can be done by projection of  $\mathbf{x}$  onto the orthogonal complement of the desired end point  $\mathbf{x}_B$  before applying  $Q$ . Thus, the absorbing kinetics equation is:

$$\dot{\mathbf{x}}(t) = Q^\top \left( \mathbf{x} - \frac{\langle \mathbf{x}, \mathbf{x}_B \rangle}{\langle \mathbf{x}_B, \mathbf{x}_B \rangle} \mathbf{x}_B \right), \quad \mathbf{x}(0) = \mathbf{x}_A. \quad (8)$$



So far, the construction of the initial state and the end state is not yet clear. In the reduced setting, both distributions are given by  $\mathbf{x}_A = \mathbf{e}_i$  and  $\mathbf{x}_B = \mathbf{e}_j$  in  $\mathbb{R}^{N_C}$ . In general, a reduced density does not contain any information about the full density within a certain cluster. However, we know the stationary distribution in equilibrium. It is the eigenvector of  $Q$  according to the eigenvalue  $\lambda = 0$ . We assume that the initial state and the end state are characterized by this distribution restricted to the corresponding cluster, i.e.

$$\mathbf{x}_i = DC(:, i)\tilde{D}^{-1}(i, i),$$

or, in general

$$\mathbf{x}^f = DC\tilde{D}^{-1}\mathbf{x}^r, \quad (9)$$

where  $\tilde{D}$  is the matrix with  $\pi^r$  on its diagonal which is obtained from  $\pi^f$  by (7). To make this transformation feasible, note that  $\pi^r$  is transformed to  $\pi^f$ ,

$$(DC\tilde{D}^{-1}\pi^r)_i = \sum_{j=1}^{N_C} \pi_i^f c_{ij} \frac{1}{\pi_j^r} \pi_j^r = \pi_i^f \sum_{j=1}^{N_C} c_{ij} = \pi_i^f.$$

Furthermore, the transformation (9) preserves the sum of the vector elements. However, the transformations (7) and (9) do not match uniquely because in general  $\mathbf{x} \neq C^\top DC\tilde{D}^{-1}\mathbf{x}$  and  $\mathbf{x} \neq DC\tilde{D}^{-1}C^\top \mathbf{x}$ .

Nevertheless we can construct a reduced kinetics equation which operates on reduced densities. It is obtained by the following transformation:

$$\begin{aligned} \dot{\mathbf{x}}^f &= Q^\top \mathbf{x}^f \rightarrow C^\top \dot{\mathbf{x}}^f = C^\top Q^\top \mathbf{x}^f \\ &\rightarrow \boxed{\dot{\mathbf{x}}^r = C^\top Q^\top DC\tilde{D}^{-1}\mathbf{x}^r} \end{aligned} \quad (10)$$

Since  $N_C \ll N$ , we can save a considerably amount of work by using the reduced model. A similar approach was previously investigated by the group of I.L. Hofacker at the University of Vienna [20]. However, they set up the reduced rate matrix in advance. They identify macro states as basins of attraction of local minima of the energy function (see next section) by extensive search strategies. In contrast, in our method the macro states are identified automatically by clustering the full matrix. The number of macro states in our algorithm will be smaller than the number of local minima because we also take into account entropic effects.

**Remark 3.1** *We also considered transformations alternative to (7) and (9).*

- *Suppose the transformations are given by*

$$\mathbf{x}^r = Z\mathbf{x}^f \quad \text{and} \quad \mathbf{x}^f = Z^+\mathbf{x}^r \quad \text{with} \quad Z = C^\top,$$

*where  $Z^+$  denotes the pseudo inverse of  $Z$ . Then the reduced model reads*

$$\dot{\mathbf{x}}^r = ZQ^\top Z^+\mathbf{x}^r.$$

However, the row sum is not conserved. This can be circumvented by a rescaling of the result or by constructing a matrix  $\tilde{Z}^+$  which preserves this property. The kinetics simulation turned out to be a little bit slower than (10), but the general behavior was reflected in a sufficient way.

- Suppose the transformations are given by

$$\mathbf{x}^r = Z^+ \mathbf{x}^f \quad \text{and} \quad \mathbf{x}^f = Z \mathbf{x}^r \quad \text{with} \quad Z = DC\tilde{D}^{-1}.$$

The reduced model becomes

$$\dot{\mathbf{x}}^r = Z^+ Q^\top Z \mathbf{x}^r.$$

It turned out that there appear negative entries in the resulting distribution vectors. Therefore, we did not further investigate this example.

## 4 Approximation of Transition Rates

There are several possibilities to obtain the transition rate matrix  $Q$ . For example, it can be calculated directly from the transition probability matrix  $P$ . On the other hand, an approximation of  $Q$  can be used to identify clusters if the transition probability matrix is not available or difficult to compute. This can either be done by an extensive sampling or by a careful examination of the potential energy surface.

### 4.1 The Potential Energy Surface

Our goal is to construct a transition rate matrix  $Q$  which meets the detailed balance condition (5). With  $R_{ij} = \frac{\pi_j}{\pi_i}$ , the reversibility constraint becomes

$$q_{ij} = R_{ij} q_{ji}.$$

Assume,  $Q$  is scaled such that the off-diagonal entries satisfy  $0 \leq Q \leq 1$ . Then  $Q$  must satisfy

$$0 \leq q_{ij} \leq \min(1, R_{ij}).$$

According to [1],  $q_{ij}$  may be chosen arbitrarily for some fixed orientation of pairs and then  $q_{ji}$  is forced to meet the reversibility constraint.

Let the domain of interest,  $\Omega$ , be partitioned into disjoint subsets  $A_i \subset \Omega$ . In molecular dynamics, the equilibrium weights of the subsets are given by

$$\pi_i = \frac{1}{\mathcal{Z}} \int_{A_i} \exp(-\beta V(x)) dx,$$

where  $V(x)$  is the potential energy given by the underlying force field,  $\beta = 1/(k_B T)$  is the inverse temperature depending on the Boltzmann constant  $k_B$ , and  $\mathcal{Z} = \int_{\Omega} \exp(-\beta V(x)) dx$  is the scaling factor such that  $\sum_i \pi_i = 1$ . In general,  $\mathcal{Z}$  is unknown but it disappears in the fraction  $\frac{\pi_i}{\pi_j}$ . This fact is taken into

account in the following examples. The components  $\pi_i$  of  $\pi$  could be approximated by some quadrature rule,

$$\pi_i \approx \sum_{\substack{j=1 \\ x_j \in A_i}}^k \exp(-\beta V(x_j)) w_j.$$

#### 4.1.1 Examples

Every expression  $q_{ij} = \frac{1}{\pi_i} S(i, j)$  with a function  $S$  which is symmetric w.r.t.  $i$  and  $j$  meets the detailed balance condition.

**Metropolis Dynamics** The classic Metropolis choice is

$$q_{ij} = \min(1, R_{ij}) = \min(1, \frac{\pi_j}{\pi_i}) = \frac{1}{\pi_i} \min(\pi_i, \pi_j).$$

**Barker Dynamics** Chemists often use Barker's dynamic where  $q_{ij}$  is given by

$$q_{ij} = \frac{R}{1+R} = \frac{1}{\frac{\pi_i}{\pi_j} + 1} = \frac{\pi_j}{\pi_i + \pi_j} = \frac{1}{\pi_i} \frac{1}{\frac{1}{\pi_i} + \frac{1}{\pi_j}}.$$

The second factor can be interpreted as the total resistance of two parallel resistors or the harmonic average, respectively.

**Geometric Average**  $Q$  may also be chosen as

$$q_{ij} = \sqrt{\frac{\pi_j}{\pi_i}} = \frac{1}{\pi_i} \sqrt{\pi_i \pi_j}.$$

Here, the transition rate depends on the geometric average.

## 4.2 Sampling Approach

As we have explained earlier,  $-\frac{q_{ij}}{q_{ii}}$  is the probability that the Markov chain visits state  $j$  just after leaving state  $i$ . Furthermore,  $-q_{ii} = \lambda(i)$  is the inverse of the average life time of state  $i$ . The following algorithm determines the entries in line  $i$  of  $Q$  by a simulation of transition frequencies to the neighbors  $n(i)$ .

```
t_i=0;          /*total time of all trajectories*/
notr_i=0;       /*number of trajectories*/
stop=0;
for s=1:n
    for j=1:m
        choose starting position x_s(0) and momentum p_j(0);
        k=1;
        while stop==0
```

```

    {x_s(k),p_i(k)}=MDstep(x_s(0),p_j(0),dt);
    if {x_s(k),p_j(k)} belongs to neighbour n(i)
        Q(i,n(i))=Q(i,n(i))+1;
        stop=1;
        notr_i=notr_i+1;
        t_i=t_i+k*dt;
    end
    if {x_s(k),p_j(k)} not in any neighbour &
        {x_s(k),p_j(k)} not in A_i
        stop=1; /*reject trajectory*/
    end
    k=k+1;
end
end
end
Q(i,:)=Q(i,+)/notr_i;
Q(i,i)=-1;
Q(i,:)=Q(i,)*t_i/notr_i;

```

The time step  $\Delta t$  of the molecular dynamics step must be chosen carefully. If it is too small, the MD-steps are too expensive. If it is too large, many trajectories will be rejected. Furthermore, one must assure that the starting points represent the correct Boltzmann distribution within the set  $A_i$ . Therefore, this approach is time-consuming, especially for larger systems.

### 4.3 Direct Computation

Assume there already exists a transition probability matrix, for example from a dynamics simulation with hybrid Monte Carlo methods [7]. By applying the inverse operation to (4), the rate matrix can be calculated as

$$Q = \frac{1}{\tau} \log P(\tau).$$

For numerical details in computing the matrix logarithm, see [4].

## 5 Numerical Illustrations

We present the application to the n-pentane molecule  $CH_3(CH_2)_3CH_3$  which was modelled with Merck Molecular Force Field [10][11] at a temperature of 300K. The rate matrix  $Q$  was calculated directly from the transition probability matrix  $P(t)$ .  $P(t)$  itself resulted from a conformation dynamics simulation with ZIBgridfree, a program package based on meshfree methods which was developed at Zuse-Institute Berlin [19],[13].

We found 9 eigenvalues of  $Q$  close to 0,

$$\lambda = \{-3.3e-17, -3.3e-4, -4.8e-4, -5.7e-4, -7.8e-4, -9.9e-4, -1.2e-3, -4.6e-3, -5.1e-3\},$$

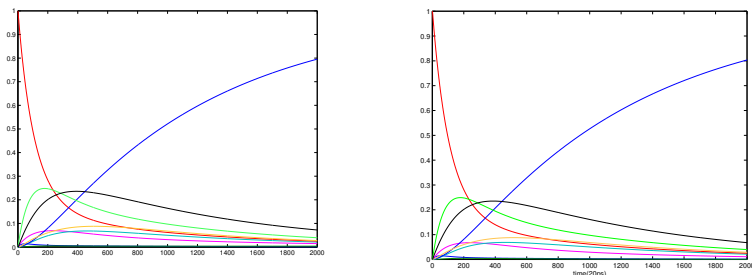


Figure 1: Matlab [8] plot of a conformation kinetics simulation. The 9 lines correspond to the weights of the 9 clusters. *Left*: From (g+/t)-conformation of pentane to the (t/g+)-conformation with the full process. Plot of the vector  $C^T \mathbf{x}^f$  every 0.01ps. *Right*: From (g+/t)-conformation of pentane to the (t/g+)-conformation with the reduced model. Plot of the vector  $\mathbf{x}^r$  every 0.01ps.

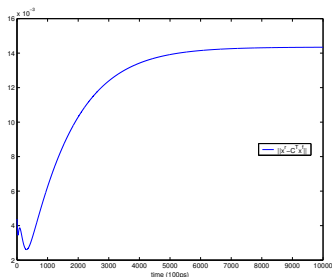


Figure 2: 2- norm of the error vector  $e = \mathbf{x}^r - C^T \mathbf{x}^f$  every 0.01ps.

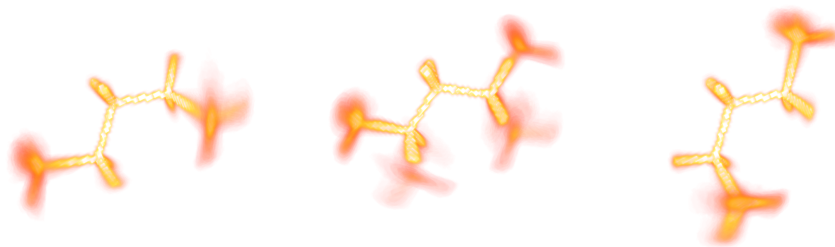


Figure 3: Volume rendering of the (g+/t)-conformations of pentane (left), the (t/g+)-conformation (right), and the corresponding transition macrostate (middle) in *amira/amiraMol* [17],[14].

followed by a gap to the 10th eigenvalue  $\lambda_{10} = -1.5e - 2$ . This corresponds to 9 metastable conformations which can be distinguished according to the orientation of one of the two dihedral angles ( $\pm g$  and  $t$  denote the  $\pm$  gauche and trans orientations):

$$\text{conformations} = \{-g/t, t/+g, -g/-g, t/t, t/-g, +g/t, +g/+g, -g/+g, +g/-g\}$$

The results for a  $(g+/t) \rightarrow (t/g+)$  transition of pentane are shown in figure 1. We performed the reduced kinetics according to (10) with  $Q \in \mathbb{R}^{9 \times 9}$  as well as the full kinetics (6) with  $Q \in \mathbb{R}^{60 \times 60}$  where we projected the results to the reduced space by (7). As one can see, the reduced model provides nearly the same results. Figure 2 shows that the error is of order  $10^{-2}$  and that it converges to a fixed value if the system reaches the equilibrium. This is due to the fact that the final reduced state with full kinetics,  $C^T D C \tilde{D}^{-1} \mathbf{x}_B$ , is not equal to the predefined reduced end state  $\mathbf{x}_B$ . Figure 1 can be interpreted as follows. The conformational change from  $(g+/t)$ -pentane to  $(t/g+)$  crosses the  $(t/t)$ -conformation which can be seen as transition state. The transition is visualised in Figure 3. The left picture shows the start conformation  $(g+/t)$ , the right one the end conformation  $(t/g+)$ . At each step of the 20ps kinetics simulation, a similar density plot can be computed. The picture in the middle shows the transition state at 3.5ps simulation length. It can be considered as the intermediate distribution of states at this particular time.

## 6 Conclusion

We have shown that the clustering method PCCA+, originally developed for transition probability matrices, can also be applied to the corresponding transition rates because the eigenvectors, which form the basis of the membership functions, remain the same. Furthermore, we have developed suitable density transformations between the unreduced space and the space spanned by the clusters. This allows us to operate on low-dimensional spaces, for example to perform kinetics simulations, such that the macro state approximations are in good agreement with the full process.

**Acknowledgments** The authors would like to thank Peter Deuffhard for motivation and useful comments on the topic as well as for initializing the contact to the theoretical chemistry group at the University of Vienna. Thanks to I.L. Hofacker, M.T. Wolfinger and W.A. Svrcek-Seiler for their explanations about reaction kinetics. We also thank W. Huisinga for his illustrative lecture on Markov processes.

## References

- [1] L. J. Billera and P. Diaconis. A geometric interpretation of the Metropolis-Hastings algorithm. *Statistical Science*, 16(4):335–339, 2001.

- [2] P. G. Bolhuis, C. Dellago, P. L. Geissler, and D. Chandler. Transition path sampling: throwing ropes over mountains in the dark. *Journal of Physics: Condensed Matter*, 12:A147–A152, 2000.
- [3] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Number Texts in Applied Mathematics in 31. Springer-Verlag New York, 1999.
- [4] P. I. Davies and N. J. Higham. A Schur-Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.
- [5] P. Deuffhard, W. Huisinga, A. Fischer, and Ch. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.
- [6] P. Deuffhard and M. Weber. Robust Perron Cluster Analysis in Conformation Dynamics. In M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte, editors, *Lin. Alg. App. – Special Issue on Matrices and Mathematical Biology*, volume 398C, pages 161–184. Elsevier Journals, 2005.
- [7] D. Frenkel and B. Smit. *Understanding Molecular Simulation – From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Press, A Division of Harcourt, Inc., www.academicpress.com/computationalscience, 2nd edition, 2002.
- [8] TheMathWorks Inc. Germany. Matlab(R) 6.5.0, 1994–2005.
- [9] G.H. Golub and C.F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996.
- [10] T.A. Halgren. *J. Am. Chem. Soc.*, 114:7827–7843, 1992.
- [11] T.A. Halgren. Merck molecular force field. *J. Comp. Chem.*, 17(I-V):490–641, 1996.
- [12] M. Kijima. *Markov Processes for Stochastic Modeling*. Stochastic Modeling Series. Chapman and Hall, 1997.
- [13] H. Meyer. Die Implementierung und Analyse von HuMfree – einer gitterfreien Methode zur Konformationsanalyse von Wirkstoffmolekülen. Master’s thesis, Free University Berlin, February 2005.
- [14] J. Schmidt-Ehrenberg, D. Baum, and H.-Ch. Hege. Visualizing dynamic molecular conformations. In *IEEE Visualization 2002*, pages 235–242. IEEE Computer Society Press, 2002.
- [15] Ch. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1999.
- [16] Ch. Schütte, A. Fischer, W. Huisinga, and P. Deuffhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys., Special Issue on Computational Biophysics*, 151:146–168, 1999.
- [17] D. Stalling, M. Westerhoff, and H.-Ch. Hege. Amira - a highly interactive system for visual data analysis. In Christopher R. Johnson and Charles D. Hansen, editors, *Visualization Handbook*. Academic Press, November 2004.
- [18] M. Weber. *Meshless Methods in Conformation Dynamics*. PhD thesis, 2005. In preparation.
- [19] M. Weber and H. Meyer. Zibgridfree - adaptive conformation analysis with qualified support of transition states and thermodynamic weights. Technical Report ZIB 05-17, Zuse Institute Berlin, 2005.
- [20] M. T. Wolfinger, W. A. Svrcek-Seiler, Ch. Flamm, I. L. Hofacker, and P. F. Stadler. Efficient computation of RNA folding dynamics. *J. Phys. A: Math. Gen.*, 37:4731–4741, 2004.