# Theoretical investigation of reaction mechanisms – from methodological development to applications

PhD dissertation

**János Daru**

Supervisors:
**András Stirling**
Institute of Organic Chemistry,
Research Centre for Natural Sciences
and
**Gergely Tóth**
Department of Physical Chemistry,
Eötvös Loránd University

2015

# Contents

i

# Abstract

With the remarkable progress in computer technology, theoretical chemistry has become an essential tool in all fields of chemistry. In particular, theoretical mechanistic studies have been proved to be very important in understanding chemical reactions and to develop new, more efficient processes.My thesis discusses our efforts in this field. Part of my work has been experimentally motivated, we have explored reaction pathways of reactions which are very promising from synthetic point of view. The other part of my thesis discusses our results in method developments.

First a detailed introductory section is presented where all the relevant theoretical background and approaches are reviewed. Then first the mechanistic studies are discussed. For a palladium catalyzed C–H activation reaction our calculations identified a path where the rate-determining C–H activation step is followed by a C–C coupling via a bimetallic complex and a Pd(III)-Pd(I) reductive elimination. For a very recent silver-mediated oxidative C–H/C–H functionalization the DFT calculations have pointed out the dual role of silver: it is a reactant to initiate the reaction in a radical route and a catalyst in the later stages of the path to drive the reaction to the furan-ring formation. In a joint experimental-theoretical study of direct trifluoroethylation of indoles the computations have identified the reaction channel leading to C3 trifluorethylation of indoles, and also explained the side reactions and the role of the bases. On the basis of the results the efficiency of a given substrate–base combination could be assessed by comparing the theoretical activation barriers of the competing N- and C-alkylations. We have also studied the cooperative hydrogen splitting by Frustrated Lewis Pairs. New, remarkable aspects of the reaction paths have been identified from the reactive trajectories. The asynchronicity of the activation could be traced as the origin of the strong localization of the reaction enthalpy on the proton in the product state, whereas the reactant state temperature distribution pointed out the role of $H_2$ translation energy in the activation.

We have developed a new and efficient method to calculate rate constants. The method requires the free energy profile along a suitable reaction coordinate and dynamical information from short unbiased trajectories. We have also developed a new approach to obtain an efficient CV where we represent the committor of the system by property map. The method is based on an efficient initial geometry generation algorithm and an iterative optimization. Combination of the method with enhanced sampling techniques has been shown to provide very promising results.

# Összefoglaló

A számítástechnika rohamos fejlődésével a kémia számos területén nélkülözhetetlen eszközzé váltak az elméleti számítások. A kémiai reakciók megértésében és az új, hatékonyabb eljárások kifejlesztésében különösen fontos szerepet játszik a reakciómechanizmusok elméleti vizsgálata. Doktori értekezésem az említett területen végzett kutatásainkat összegzi. Vizsgálatainkat részben kísérleti szempontok motiválták, ugyanis ígéretes szintetikus átalakítások mechanizmusát térképeztük fel. értekezésem második része módszerfejlesztéssel kapcsolatos eredményeinket tárgyalja. A dolgozat első fejezete részletes elméleti bevezetőt nyújt és ismerteti a mechanizmuskutatás fontos módszereit. Ezt követően mechanisztikus vizsgálatainkat mutatja be.

Egy palládium–katalizált C–H aktiválási reakció vizsgálata során olyan reakcióutat azonosítottunk, amelynek sebességmeghatározó lépése a C–H aktiválás, amit egy bimetallikus palládiumkomplexen végbemenő C–C kapcsolási reakció követ. Az új C–C kötés egy Pd(III)-Pd(I) reduktív elimináció során alakul ki. Egy új ezüst-katalizált oxidatív C–H/C–H funkcionalizáció sűrűségfunkciónál-elmélet keretei között végzett vizsgálata során kimutattuk az ezüst ionok kettős szerepét. A reakció kezdeti szakaszában az ezüst a reaktáns szerepét tölt be, míg a második szakaszban katalizátorként segíti a furángyűrű kialakulását. Egy közös kísérleti-elméleti kutatás során azonosítottuk indolszármazékok C3-as helyzetű trifluoretilezésének és az azt kísérő mellékreakcióknak a mechanizmusát, ami megmagyarázta a bázis szerepét az átalakításban.Eredményeink értelmében egy adott szubsztrát és bázis kombinációjának hatékonysága megbecsülhető a versengő N- és C- alkilezések aktiválási energiájából. Vizsgáltuk Frusztrált Lewis párok kooperatív $H_2$-hasítási reakcióját. A mechanizmus egy új, fontos oldalát azonosítottuk. Számításaink alapján az aktiválás aszinkronitásának egy fontos következménye, hogy a reakcióentalpia nagy része a termékben a savas hidrogénen lokalizálódik. Ugyanakkor a reaktáns oldali hőmérséklet-eloszlás arra enged következtetni, hogy a hidrogénmolekula transzlációs energiájának meghatározó szerepe van az aktiválásban.

Kifejlesztettünk egy új, hatékony eljárást reakciósebességi állandók számítására. A módszer egy megfelelő reakciókoordináta menti szabadenergia profil számításból és rövid dinamikai szimulációkból épül fel. Munkánk során emellett kifejlesztettünk egy hatékony kollektív változót,amely a dinamikai elkötelezettséget a tulajdonságtérképpel reprezentálja. Az eljárás alapját egy hatékony szerkezetgeneráló és egy iteratív optimalizációs algoritmus képezi. A módszer ritka-esemény mintavételezési technikákkal való kombinációja ígéretes eredményekre vezetett.

# Chapter 1

# Introduction

# 1.1 Reaction mechanisms – networks of elementary reactions

The research discussed in this thesis belongs to the field of elementary reaction level computational description of reaction mechanisms. Elementary reactions are molecular transformations that proceed in one step, without any finite lifetime intermediates between their reactants and products. Exploration of the network of relevant elementary reactions provides us with the sequence of transformations, the structural properties of intermediates and with the possible branching points for side reactions. The reaction mechanism can be used for the rationalization of the ratio of different products, i.e. regio- and stereoselectivities. The understanding of the activity of catalysts and prediction of the performance of potential new catalyst candidates also require the detailed knowledge of reaction mechanisms.

The mechanistic considerations are usually based on the estimation of the relative rates of the different possible reaction pathways. The thesis discusses computational exploration of three interesting, complex reaction mechanisms with multiple reaction pathways and with possible side reactions. In these studies the mechanistic interpretations tacitly assumed the validity of the transitional state theory (TST) and the molecules have been described within a quasi-zero temperature framework. A more elaborate approach has been used for the description of the hydrogen activation reactions of three bifunctional organic compounds. The thesis also discusses two methodological developments. The first is a new rate calculation method for elementary reactions. The second is a new reaction coordinate optimizing method for elementary reactions.

## 1.1.1 Central quantities: rate constant of elementary reactions and equilibrium constant

In the followings a hypothetical

$$R_1 \ (+R_2) \longrightarrow P_1 \ (+P_2 + ...) \tag{1.1}$$
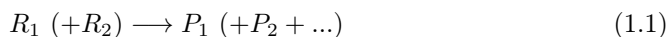
homogeneous *elementary reaction* is discussed. $R_2$ appears only if the reaction is bimolecular. For the sake of simplicity we assume that a specific reactant or product appear only once in the chemical equation (e.g. $R_1 \neq R_2$). The forward rate of the elementary step is defined by the concentration change of the reactant:

$$r_f = -\frac{d[R_1]}{dt}, \tag{1.2}$$

where the square brackets denote the actual concentration. According to the experiential observations, the rate of an elementary homogeneous reaction can be described as:

$$r_f = k_f \cdot [R_1] \ (\cdot [R_2]) \tag{1.3}$$

Here $k_f$ is the forward rate constant. If the reaction is reversible we can assume that it leads to a dynamical equilibrium:

$$R_1 \ (+R_2) \rightleftharpoons P_1 \ (+P_2 + ...) \tag{1.4}$$
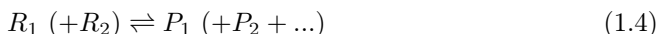
when the forward and reverse rates are equal, and the concentrations are constant. This leads to the definition of the equilibrium constant:

$$K_c = \frac{k_f}{k_r} = \frac{\Pi[P_i]}{\Pi[R_i]} \tag{1.5}$$

We note that equilibrium constant can be also defined by other quantities. For example standard equilibrium constant is defined by relative activities.

Definitions in chemistry, such as molecules, reactants or products are arbitrary. For example conformers, optical isomers, isotopomers might be considered or measured to be the same chemical species, but they may also be distinguished. On atomic level a reactant molecule (complex) turns into product through a continuous structural evolution. For example by a continuous change of the interatomic distances, and continuous redistribution of the electron density during the formation or cleavage of bonds. Therefore the border between reactant or product is also arbitrary to some extent. In spite of this arbitrariness, different definitions might give similar qualitative description. The mathematical construction which defines reactant and product is called reaction coordinate.

## 1.2 Elementary reaction – definition of reaction coordinate

The progress of the reaction can be characterized by one coordinate, which has distinct values for the reactants and products. The set of molecules considered as reactant (product) on the basis of the reaction coordinate is called reactant (product) state. The reaction coordinate is required to be able to describe the degree of transformation as well, i.e. it should change its value continuously from reactant to product state. In my thesis I will exclusively consider ground state reactions therefore I assume that the molecule with a given configuration can be classified as reactant or product. Sometimes it is not obvious how to define such a one dimensional reaction coordinate and therefore we apply a vector i.e. a set of coordinates for this purpose.

### 1.2.1 Reaction coordinate as a function of atomic coordinates

The following terms are invoked frequently in the literature of molecular simulations. However, their definition is not widely accepted. This thesis employs the following definitions [7, 8]:

- *Collective variables*(CVs) are continuous or discontinuous surjections of atomic coordinates to a lower dimension space.

- *An order parameter* is a CV that is capable to distinguish between two or more states, such as reactant or product states. Order parameters are not necessarily continuous.

- *Reaction coordinates* are continuous order parameters, that are able to describe the progress of the reactive events.

As the definitions indicate, CVs are more general than reaction coordinates. Order parameters and reaction coordinates are special sets of CVs. Usually they are expressed as functions of other CVs. The complexity of these functions varies from simple linear combinations to neural networks.

## 1.2.2  Frequently used collective variables

The simplest CVs are the actual coordinates of specific atoms with respect to some reference frames. The next level of complexity is the distance between atoms, the angle or torsion angle defined by three or four atoms. These quantities in combination with atomic masses or charges can result in wide variety of CVs. These primitive functions are frequently modified by switching functions in order to obtain better distinction between states. Such switching functions are listed below.
The rational function:

$$s(x) = \frac{1 - \left(\frac{x-d_0}{x_0}\right)^a}{1 - \left(\frac{x-d_0}{x_0}\right)^b} \tag{1.6}$$

the exponential function:

$$s(x) = \exp\left(-\frac{x-d_0}{x_0}\right) \tag{1.7}$$

the Gaussian function:

$$s(x) = \exp\left(-\frac{(x-d_0)^2}{2 \cdot x_0^2}\right) \tag{1.8}$$

the sigmoid function of sketch-map [9]:

$$s(x) = \left[1 + (2^{a/b} - 1)\left(\frac{x-d_0}{x_0}\right)\right]^{-b/a} \tag{1.9}$$

or the Fermi-like function:

$$s(x) = \frac{1}{1 + \exp(a \cdot (x - x_0))}. \tag{1.10}$$

In these equations $x$ denotes any CV, while $x_0$, $d_0$, $a$ and $b$ are constants and their actual value is selected according to the properties described by the CV (we present the parameterization as it is implemented in plumed [10]). These switching functions

can be useful in summations, as they can dramatically change the contribution of different states to the sum. Switching functions are used in coordination numbers where we count the total number of bonding pairs, or the total number of connections between two subsets of atoms. A special case of coordination numbers is called contact map [11]. A contact map sums the switching function-transformed distance of each atom from the corresponding atom in a reference structure.

Another use of the switching functions is to mimic logical gates. For example "x AND y" can be achieved by simple multiplication,

$$s(x) \cdot s(y) \tag{1.11}$$

"x XOR y" can be realized by the

$$s(x) \cdot (1 - s(y)) \tag{1.12}$$

expression. It is important to note, that these expressions provide the exact logical gate-behaviour only in their limits, where $s(x)$ approaches the Heaviside function. The logical gate like usage is included for example in the hydronium CV [12].

The Gaussian type switching function is exclusively used in path-like CVs, such as reaction-path CV [13], or property-map [14]. These complex CVs perform dimensionality reduction and define $\mathbb{R}^n \to \mathbb{R}$ surjections. The property-map utilize the following functional form:

$$X(\boldsymbol{q'}) = \frac{\sum_i^{N_r} X(q_i) \exp\left(-\gamma D(\boldsymbol{q'}, \boldsymbol{q_i})\right)}{\sum_i^{N_r} \exp\left(-\gamma D(\boldsymbol{q}, \boldsymbol{q_i})\right)} \tag{1.13}$$

$$Z(\boldsymbol{q'}) = -\frac{1}{\gamma} \log\left(\sum_i^{N_r} \exp\left(-\gamma D(\boldsymbol{q'}, \boldsymbol{q_i})\right)\right) \tag{1.14}$$

where $D(\boldsymbol{q'}, \boldsymbol{q_i})$ is the distance between the $i^{\text{th}}$ reference geometry and the actual configuration $\boldsymbol{q'}$, $X(q_i)$ is the $X$ property of the $i^{\text{th}}$ image, while $N_r$ is the total number of reference structures. $\gamma$ is a positive parameter for controlling the smoothing of the function. If $\gamma$ is zero, the mapping reduces to a constant function (the average of $X$ for the $N_r$ reference structure). If $\gamma$ is a large value, then the exponentials select the $X$ quantity of the closest structure. $Z(\boldsymbol{q})$ calculates an weighted average distance from the reference structures. The reaction path-CV [13] can be considered as a special case of property map, where $X_i$ is defined as follows:

$$X_i = \frac{i-1}{N_r} \tag{1.15}$$

Path-like CVs have raised a new and still open question, namely the proper definition of distance in the context of properties. The distance between two points of the CV space $\boldsymbol{x}(\boldsymbol{q_1})$ and $\boldsymbol{x}(\boldsymbol{q_2})$ is usually defined by the application of a metric tensor $\mathbf{M}$ in a quadratic expression:

$$D(\boldsymbol{q}_1, \boldsymbol{q}_2)^2 = [\boldsymbol{x}(\boldsymbol{q}_1) - \boldsymbol{x}(\boldsymbol{q}_2)]^T \mathbf{M} [\boldsymbol{x}(\boldsymbol{q}_1) - \boldsymbol{x}(\boldsymbol{q}_2)] \tag{1.16}$$

Various definitions are in use for the metric tensor, such as diagonal i.e. (scaled) Euclidean, or the Mahalanobis. In the latter the metric tensor is the inverse of the covariance matrix. Another widely applied distance is the root mean square deviation (RMSD). This quantity is minimized by the alignment of the molecules. The distance is defined beween the configurations $\boldsymbol{q}$ and $\boldsymbol{q}'$ as:

$$d(\boldsymbol{q}, \boldsymbol{q}') = \sqrt{\sum_i \sum_\alpha w_i [q_{i,\alpha} - c_\alpha(\boldsymbol{q}) - \sum_\beta M(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{w}')_{\alpha,\beta} (q'_{i,\beta} - c_\beta(\boldsymbol{q}'))]^2} \tag{1.17}$$

where $w_i$ is the normalised weight of the i$^{\text{th}}$ atom. $\alpha, \beta \in \{x, y, z\}$. $\mathbf{M}(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{w}')$ is the optimal alignment transformation matrix and $c_\alpha(\boldsymbol{q})$ is defined as:

$$c_\alpha(\boldsymbol{q}) = \sum_i w'_i q_{i,\alpha}. \tag{1.18}$$

$c_\alpha(\boldsymbol{q})$ equals to the center of mass if the weights are derived from the atomic masses $\{m_i\}$ as:

$$w_i = \frac{m_i}{\sum_i m_i}. \tag{1.19}$$

Physical quantities such as functions of atomic coordinates can be considered as CVs as well. For example such functions are the total energy or the dipole moment of the system.

## 1.3 Probability density with respect to collective variable(s): Free energy surface

The equilibrium constant in canonical ensemble is linked to the free energy difference between the product and the reactant states [15]

$$\Delta_{PS}^{RS} F = -k_B T \ln K, \tag{1.20}$$

where $k_B$ is the Boltzmann constant and $T$ is the thermodynamic temperature. The definition of equilibrium constant in Eq. 1.5 is based on concentrations. However, as the probability of finding the system in a given state and the concentration are proportional, the equilibrium constant between reactant state (RS) and product state (PS) can be defined on the basis of probabilities as well:

$$K = \frac{c_{PS}}{c_{RS}} = \frac{P_{PS}}{P_{RS}} \tag{1.21}$$

The connection between free energy differences and relative probabilities is therefore:

$$\Delta_{PS}^{RS} F = -k_B T \ln(P_{RS}/P_{PS}). \tag{1.22}$$

If the progress of the elementary reaction can be followed by a single reaction coordinate ($\lambda$), equilibrium constant and the free energy difference can be derived from the probability density with respect to the reaction coordinate:

$$
K \;=\; \frac{P_{PS}}{P_{RS}} = \frac{\underset{\lambda \in PS}{\int} f(\lambda)d\lambda}{\underset{\lambda \in RS}{\int} f(\lambda)d\lambda} \tag{1.23}
$$

$$
\Delta_{PS}^{RS}F \;=\; -k_B T \ln\left( \frac{\underset{\lambda \in PS}{\int} f(\lambda)d\lambda}{\underset{\lambda \in RS}{\int} f(\lambda)d\lambda} \right), \tag{1.24}
$$

where $f(\lambda)$ is the probability density with respect to the reaction coordinate, $\lambda$. The free energy difference between two infinitesimal state intervals at $PS = \lambda$ and $RS = \lambda'$ can be written:

$$
F(\lambda) - F(\lambda') = -k_B T \ln\left( f(\lambda)/f(\lambda') \right). \tag{1.25}
$$

This expression leads to the definition of an important concept, the free energy surface $F(\lambda)$:

$$
F(\lambda) = -k_B T \ln\left( f(\lambda) \right). \tag{1.26}
$$

It is important to emphasize, that these relations are valid for canonical ensembles. The probability density and therefore the free energy surface is not restricted to reaction coordinates. One can define probability densities with respect to any of CVs. The dimension of the probability density can be reduced by integration:

$$
f(\lambda_1) = \int f(\lambda_1, \lambda_2)d\lambda_2 \tag{1.27}
$$

The connection between free energy surface and probability density (Eq. 1.26) is analogous for higher dimensions:

$$
F(\lambda_1, \lambda_2) = -k_B T \ln\left( f(\lambda_1, \lambda_2) \right) \tag{1.28}
$$

$$
f(\lambda_1, \lambda_2) = \exp(-\beta F(\lambda_1, \lambda_2)), \tag{1.29}
$$

where $\beta$ is $1/k_B T$. Substituting $f(\lambda_1)$ in Eq. 1.26 from Eq. 1.27 and employing the definition of $f(\lambda_1, \lambda_2)$ from Eq. 1.29 yield the expression for dimension reduction of free energy surfaces:

$$
\begin{aligned}
F(\lambda_1) \;&=\; -k_B T \ln\left( f(\lambda_1) \right) & (1.30)\\
&=\; -k_B T \ln\left( \int f(\lambda_1, \lambda_2)d\lambda_2 \right) & (1.31)\\
&=\; -k_B T \ln\left( \int \exp(-\beta F(\lambda_1, \lambda_2))d\lambda_2 \right), & (1.32)
\end{aligned}
$$

The previous quantities were based on statistical descriptors of of any CV, such as probability density of inter-atomic distances, angles or other structural quantities.

Many distributions are readily accessible from experiments. For example in single molecule experiments, the distribution of the end to end distance of a molecule is directly measured by atomic force microscopy. The corresponding free energy surface can be calculated through Eq. 1.26. It is important to note that free energy and free energy surfaces can be connected to non-measurable microscopic physical quantities as well. Statistical mechanics connects free energy to energy. Exclusively the knowledge of Boltzmann relation is necessary for obtaining the connection between the probability of a micro state and its energy:

$$P_i = \frac{\exp\left(-\beta E_i\right)}{\sum\limits_i \exp\left(-\beta E_i\right)} \tag{1.33}$$

In continuous treatment this expressions changes by replacing the $i$ index by $\boldsymbol{p}, \boldsymbol{q}$ phase-space points:

$$P(\boldsymbol{p}, \boldsymbol{q}) = \frac{\exp\left(-\beta H(\boldsymbol{p}, \boldsymbol{q})\right) \boldsymbol{dpdq}}{\int \int \exp\left(-\beta H(\boldsymbol{p}, \boldsymbol{q})\right) \boldsymbol{dpdq}}, \tag{1.34}$$

Ensemble average of an $A(\boldsymbol{p}, \boldsymbol{q})$ quantity is:

$$\langle A \rangle = \frac{\int A(\boldsymbol{p}, \boldsymbol{q}) \exp\left(-\beta H(\boldsymbol{p}, \boldsymbol{q})\right) \boldsymbol{dpdq}}{\int \int \exp\left(-\beta H(\boldsymbol{p}, \boldsymbol{q})\right) \boldsymbol{dpdq}} \tag{1.35}$$

Eq. 1.34 can be factorized by introducing potential energy $V(\boldsymbol{q})$ and kinetic energy $E_{kin}(\boldsymbol{p})$:

$$P(\boldsymbol{p}, \boldsymbol{q}) = \underbrace{\frac{\exp\left(-\beta V(\boldsymbol{q})\right)}{\int \exp\left(-\beta V(\boldsymbol{q})\right) \boldsymbol{dq}}}_{f(\boldsymbol{q})} \boldsymbol{dq} \cdot \underbrace{\frac{\exp\left(-\beta E_{kin}(\boldsymbol{p})\right)}{\int \exp\left(-\beta E_{kin}(\boldsymbol{p})\right)}}_{f(\boldsymbol{p})} \boldsymbol{dp} \tag{1.36}$$

In case of an $A$ quantity that depends on spatial coordinates exclusively, the momentum dependence cancels in the ensemble average:

$$\langle A \rangle = \frac{\int A(\boldsymbol{q}) \exp\left(-\beta V(\boldsymbol{q})\right) \boldsymbol{dq}}{\int \exp\left(-\beta V(\boldsymbol{q})\right) \boldsymbol{dq}} \tag{1.37}$$

The spatial integral is related to the excess free energy:

$$F = -k_B T \ln\left(\int \exp\left(-\beta V(\boldsymbol{q})\right) \boldsymbol{dq}\right) \tag{1.38}$$

$$= -k_B T \ln Z, \tag{1.39}$$

where $Z$ denotes the canonical configuration integral.

The probability density with respect to any coordinate can be derived from Eq. 1.34 by the following expression:

$$f(\lambda_a) = \frac{\int \int \exp\left(-\beta H(\boldsymbol{p}, \boldsymbol{q})\right) \left(\delta(\lambda(\boldsymbol{p}, \boldsymbol{q}) - \lambda_a) \boldsymbol{dpdq}\right)}{\int \int \exp\left(-\beta H(\boldsymbol{p}, \boldsymbol{q})\right) \boldsymbol{dpdq}}, \tag{1.40}$$

which simplifies to the following equation if $\lambda$ depends only on spatial coordinates:

$$f(\lambda_a) = \frac{\int \exp\left(-\beta V(\boldsymbol{q})\right) \delta(\lambda(\boldsymbol{q}) - \lambda_a)\boldsymbol{dq}}{\int \exp\left(-\beta V(\boldsymbol{q})\right) \boldsymbol{dq}}. \tag{1.41}$$

Relative probabilities can be expressed easily:

$$\frac{f(\lambda_a)}{f(\lambda_b)} = \frac{\int \exp\left(-\beta V(\boldsymbol{q})\right) \delta(\lambda(\boldsymbol{q}) - \lambda_a)\boldsymbol{dq}}{\int \exp\left(-\beta V(\boldsymbol{q})\right) \delta(\lambda(\boldsymbol{q}) - \lambda_b)\boldsymbol{dq}}. \tag{1.42}$$

These expressions have led to numerous theoretical considerations and computational techniques, which provide excellent extensions to experimental methods. The next section focuses on some of these techniques, especially on methodologies of calculating free energy differences or free energy surfaces by computer simulations.

## 1.4 Molecular dynamics simulations

Unless indicated otherwise, we employ classical treatment for the nuclear degrees of freedom. This section summarizes the simulation techniques for a classical dynamical system obeying the forces arising from the potential energy surface (PES) (see section 1.7 for details of PES). The time evolution of such system can be described with the classical propagator [16]:

$$a(\boldsymbol{q}_t) = \exp(i\hat{L}t) \cdot a(\boldsymbol{q}_0) \tag{1.43}$$

where $L$ is the Liouville operator:

$$i\hat{L} = \sum_{\alpha=1}^{3N} \left[\frac{\partial H}{\partial q_\alpha}\frac{\partial}{\partial p_\alpha} - \frac{\partial H}{\partial p_\alpha}\frac{\partial}{\partial q_\alpha}\right] \tag{1.44}$$

The propagator can be factorized by the Trotter theorem, which leads to an exact equation in the limit of infinitesimal time-steps [17]. For a single degree of freedom with $m$ mass we can derive:

$$\exp(i\hat{L}_q t) = \lim_{\Delta t \to 0}\left[\exp\left(\frac{\Delta t}{2}F(q)\frac{\partial}{\partial p}\right)\exp\left(\Delta t\frac{p}{m}\frac{\partial}{\partial q}\right)\exp\left(\frac{\Delta t}{2}F(q)\frac{\partial}{\partial p}\right)\right]^{\frac{t}{\Delta t}} \tag{1.45}$$

The application of the propagator with a finite time step leads to the velocity Verlet [18] algorithm:

$$q(\Delta t) = q(0) + \frac{p(0)}{m}\Delta t + \frac{F(q(0))}{2m}\Delta t^2 \tag{1.46}$$

$$\frac{p(\Delta t)}{m} = \frac{p(0)}{m} + \frac{F(q(0)) + F(x(\Delta t))}{2m}\Delta t \tag{1.47}$$

The propagation of equations of the motion by the velocity Verlet algorithm is a symplectic (i.e. conserves the are in the phase space delimited by an ensemble of the

system), time-reversible scheme with $\mathcal{O}(\Delta t^3)$ error. As a symplectic integrator, this scheme results in constant energy molecular dynamics (i.e. microcanonical ensemble). In order ensure NVT conditions it is necessary to include the effect of the thermostats. The influence of the heat bath can be imposed by rescaling the velocities in each time-step or by modifying the Hamiltonian of the system. In the latter case the energy of the extended system will be conserved, while the physical system will behave according to the canonical ensemble. The change in the Hamiltonian can be applied by introducing a half time-step propagator before and after the classical Hamiltonian propagator according to the thermostat's Liouville-operator. Several different thermostat schemes are readily available in different molecular dynamics codes. Two widely applied thermostats are presented here. The Hamiltonian of the Nosé-Hoover chain [19–21] features an extended Hamiltonian with an additional term for accounting for the kinetic energy of the additional degrees of freedom and with temperature dependent potential energy terms. The expression for the Hamiltonian is the following:

$$H'(\boldsymbol{p}, \boldsymbol{q}) = H(\boldsymbol{p}, \boldsymbol{q}) + \sum_{j=1}^{M} \frac{p_{\eta_j}^2}{2Q_j} + gk_BT\eta_1 + k_BT\sum_{j=2}^{M} \eta_j \qquad (1.48)$$

This thermostat introduces chaos into the equations by coupling several extra degrees of freedom. The Nosé-Hoover method for systems with low dimensionality and with short chains might lead to ergodicity problems [21, 22].

Another efficient thermostat applies random velocity scaling in order to obtain the proper ensemble. It is called canonical sampling through velocity rescaling (CSVR) [22]. The scaling is defined by the following equation:

$$dE_{kin} = (\bar{E}_{kin} - E_{kin})\frac{dt}{\tau} + 2 + \sqrt{\frac{\bar{E}_{kin}E_{kin}}{N_f}}\frac{dW}{\sqrt{\tau}}, \qquad (1.49)$$

where $\tau$ is the time constant parameter, $N_f$ is the number of degrees of freedom and $dW$ is a random number with Gaussian distribution. $\bar{E}_{kin}$ is the average of the previous instanious kinetic energy values ($E_{kin}$). This scheme has some attractive features as in the $\tau \to \infty$ limit it gives back the Hamiltonian dynamics, while in the $\tau \to 0$ limit, the system is instantly thermalised. In contrast with Nosé-Hoover, CSVR introduces chaos by random numbers.

### 1.4.1 Free energy methods

In this section a collection of different molecular dynamics based methods is presented, that are designed to calculate free energy surfaces or free energy differences. The methods presented here can be classified into the following groups:

- methods based on distributions
- adaptive biasing potential methods

- methods based on energy difference accumulation

- methods based on force accumulation

- adaptive biasing force methods

Calculating a free energy surface or differences on the basis of histogram accumulation is called Boltzmann inversion and it is the basis of numerous others methods [23]. The method is straightforward, since Eq. 1.26 defines the connection between the probability density and the free energy surface. However there are some technical issues to consider in practice. In the case of a sufficiently long molecular dynamics simulation, the $\lambda(t)$ function can be transformed to a $f(\lambda)$ function. If $\lambda(t)$ were continuous this transformation could be done analytically. However molecular dynamics simulations apply finite time-steps. If the values for the reaction coordinate or CV are saved frequently, intermediate values can be interpolated. With the final data sets in hand, the definition of bin widths is still necessary. Larger bins are advantageous as the error of the histogram estimate is inversely proportional to the square-root of the counts per the actual bin. However larger bins result in lower resolution of the surface. This sort of compromise is even more frequent for two or three dimensional distributions where the inversion cannot be done for continuous trajectories either. One usual work-around for this issue is to build up the histogram from Gaussian kernel functions in order to avoid bins [24, 25]. However this sort of smoothing might hide important details of the probability density or result in artificially over-structured probability density in case of too small Gaussian width.

Another issue is the varying error along the reaction coordinate. If there is a low probability region along the profile it is sampled with much worse statistics. In the vast majority of the cases this probability region becomes prohibitive for the sampling. In other words the trajectory cannot escape from one of the high probability regions, or equivalently one of the free energy minima [26].

This phenomenon introduces the so-called *rare event* methods. Rare events are meant to be rare on molecular or simulation time scales, since their characteristic times are much longer than the time ranges accessible by present simulations. It is important to note that although rare events are infrequent, they are often fast, even on molecular time-scale (femto-pico second) when they actually occur. One possible solution for this problem is to apply static potentials to enhance the sampling of the low probability regions. The effect of a static potential can be derived from Eq. 1.42. If the potential $V(\boldsymbol{q})$ is substituted with a potential $V'(\boldsymbol{q}) = V(\boldsymbol{q}) + U(\lambda(\boldsymbol{q}))$, then the ratio of the perturbed probabilities will remain:

$$\frac{P'(\lambda_a)}{P'(\lambda_b)} = \frac{\exp(-\beta U(\lambda_a))}{\exp(-\beta U(\lambda_b))} \frac{P(\lambda_a)}{P(\lambda_b)} = \exp(-\beta \Delta_b^a U) \frac{P(\lambda_a)}{P(\lambda_b)} \qquad (1.50)$$

Applying Eq. 1.26 for this expression, we obtain the resultant free energy change ($\Delta_b^a F'$) between $\lambda_a$ and $\lambda_b$:

$$\Delta_b^a F' = \Delta_b^a F + \Delta_b^a U. \qquad (1.51)$$

In conclusion, static potential alters the probability density by a multiplicative exponential factor, while contributes with an additive term to the free energy surface.

A convenient way to obtain higher population in the rarely visited regions is to add a harmonic restrain to the Hamiltonian [27]:

$$U(\lambda) = \frac{k}{2}(\lambda - \lambda_0)^2 \tag{1.52}$$

This specific biasing technique is called umbrella sampling [27]. The distorted histograms can be debiased by a $\exp(+\beta U(\lambda))$ factor and the free energy information of the neighbouring umbrellas can be aligned by numerous methods. The alignment is not straightforward, as free energy surfaces are defined up to a constant only. Therefore the free energy profiles from the umbrellas have to be adjusted by a level shift. The most commonly used method for this approach is the weighted histogram analysis method (WHAM) [28, 29]. Its zero width limit is the multistate Bennett acceptance ratio (M-BAR) [30]. This method is attractive as it has no bin-width parameter. A recent promising approach is variational free energy profile (vFEP) method [31, 32]. The vFEP method is based on the maximum likelihood fitting procedure of an analytical surface (one or two dimensional spline) to the dataset. The derivatives of the resulting free energy surface can be analytically estimated, therefore its critical points and the minimum free energy path can be calculated. vFEP can also efficiently overcome one of the shortcomings of umbrella sampling based methods. Namely, in both WHAM and M-BAR type analysis of the data, a significant overlap of the neighbouring simulations is required, whereas vFEP is more robust in this aspect. It is important to note that the convergence of umbrella sampling simulations can be enhanced by switching the bias between umbrella sampling simulations. The random switches are performed by a Monte Carlo algorithm [33]. This approach is called Hamiltonian replica exchange umbrella sampling.

If the free energy profile were already known, we could apply its negative as a biasing potential to achieve flat free energy surface and uniform probability distribution. This idea leads to the next group of free energy methods, i.e. the adaptive biasing methods.

A simple application of this idea is called iterative Boltzmann inversion (IBI) [34]. The approach was originally used to derive simple pairwise potentials to mimic experimental $g(r)$ functions. However, as $g(r)$ is also related to the free energy surface, it can be considered as an inverse free energy problem. The procedure updates the empirical potential in each iteration by the following equation:

$$U_{i+1} = U_i - \alpha k_B T \ln\left(\frac{g_i(r)}{g_t(r)}\right) \tag{1.53}$$

where $\alpha$ is a scaling factor, $g_i(r)$ is the radial distribution function in the $i^{th}$ iteration, while $g_t(r)$ is the target radial distribution function. This method calculates the interatomic potential by adding extra potential terms depending on the actual probability density, in order to reproduce the target probability density.

Adaptive biasing methods apply and update an extra potential, in order to reach a uniform distribution, when the negative of the biasing potential equals to the free energy surface. The external potential can be updated iteratively, with sampling and potential update cycles, or simultaneously. The Gauss-mixture umbrella sampling (GAMUS) [35] method belongs to the former type. It samples CV values, and fits their distribution with a mixture of multidimensional Gaussians. The resulting probability density is used to construct a biasing potential:

$$H_{i+1} = H_i - k_B T \ln(P_i(\lambda)), \tag{1.54}$$

where $P_i$ denotes the Gaussian fitted probability density from the ith iteration. GAMUS shares similarities with other adaptive biasing methods, such as local elevation [36] and metadynamics [37]. Both methods utilize Gaussian or truncated Gaussian functions as GAMUS. Local elevation uses potentials centered on a grid, while metadynamics deposits the potentials wherever the trajectory explores the CV space. In contrast GAMUS accumulates histograms and applies the corresponding potentials in separate simulations. These methods can be considered as the application of the sum of Gaussian kernel functions for histogram evaluation as explained earlier. Another essential difference between GAMUS and metadynamics (or local elevation) is that the former increases the potential by the logarithmic probability, i.e. the true free energy contribution, while the latter benefits from the self-healing nature of the large number of hills. Due to its flexibility metadynamics is very popular and has become a dominant method for exploration of low (below ten) dimensional CV spaces. Several flavours of metadynamics have been developed and implemented to overcome the disadvantages of the method. One of the possible artefacts of the frequent hill deposition is that it distorts the canonical distribution by pumping energy in the system, and heats up a part of the system [38]. This effect is more pronounced if high hills are applied or if the rate of the deposition is high. It can be treated for some extent with massive thermostating i.e. coupling a thermostat to each degrees of freedom. Another solution can be an extended Lagrangian formulation [39], where the physically meaningful CVs are coupled to fictitious CVs, with fictitious masses. As the coupling is performed with harmonic springs, each CV requires two parameters: the mass and the coupling constant. In case of appropriately selected parameters, the physical and the fictitious CVs oscillate around each other, assisting in the efficient dissipation of the extra energy arising from the deposited hill potentials.

For faster exploration of the CVs multiple walker version of metadynamics can be applied. In this method multiple metadynamics simulations are performed in parallel. The simulations (walkers) build the same time-dependent potential by sharing the information of the deposited hills.

Free energy differences can also be calculated from other sources of information. One evident choice is the utilization of the potential energy. The free energy change of imposing a small perturbation on the system can be calculated in the framework of *free energy perturbation*(FEP) [40]. The free energy differences between two systems with

slightly different Hamiltonian can be calculated by estimating the ratio of probabilities as follows:

$$\frac{P_a}{P_b} = \frac{\int \exp\left(-\beta V_a(\boldsymbol{q})\right) d\boldsymbol{q}}{\int \exp\left(-\beta V_b(\boldsymbol{q})\right) d\boldsymbol{q}} \tag{1.55}$$

This can be accomplished by performing a molecular dynamics simulation and calculating the effect of the perturbation in each time step. The expression above can be further transformed:

$$\frac{P_a}{P_b} = \frac{\int \exp\left(-\beta V_a(\boldsymbol{q}) + \beta V_b(\boldsymbol{q}) - \beta V_b(\boldsymbol{q})\right) d\boldsymbol{q}}{\int \exp\left(-\beta V_b(\boldsymbol{q})\right) d\boldsymbol{q}} \tag{1.56}$$

$$= \frac{\int \exp\left(-\beta \Delta_b^a V(\boldsymbol{q})\right) \exp\left(-\beta V_b(\boldsymbol{q})\right) d\boldsymbol{q}}{\int \exp\left(-\beta V_b(\boldsymbol{q})\right) d\boldsymbol{q}} \tag{1.57}$$

The latter expression can be interpreted as an ensemble average:

$$\frac{P_a}{P_b} = \int \exp\left(-\beta \Delta_b^a V(\boldsymbol{q})\right) \cdot p_b(\boldsymbol{q}) d\boldsymbol{q} \tag{1.58}$$

$$= \langle \exp\left(-\beta \Delta_b^a V(\boldsymbol{q})\right)\rangle_b \tag{1.59}$$

$$\Delta_b^a F = -k_B T \ln\left(\langle \exp\left(-\beta \Delta_b^a V(\boldsymbol{q})\right)\rangle_b\right) \tag{1.60}$$

Free energy perturbation has the great advantage that it can utilize different Hamiltonians. Therefore it can be used for annihilation or creation of particles. Changing the number of particles makes usually broad energy distributions and give large variance for the average above. This can be handled by assuming intermediate states for example by an interpolative Hamiltonian:

$$H(\mu, \boldsymbol{p}, \boldsymbol{q}) = E_{kin}(\boldsymbol{p}) + \mu^k \cdot V_a(\boldsymbol{q}) + (1-\mu)^k \cdot V_b(\boldsymbol{q}), \tag{1.61}$$

where $k$ is an exponent and $\mu$ is a parameter for the interpolation. The change from a to b can be performed by simulation for several lambda steps.

The free energy response for infinitesimal change in $\mu$ can be derived from Eq. 1.39 by applying explicit derivation.

$$\left(\frac{\partial F}{\partial \mu}\right)_{\mu=\mu*} = -\frac{\partial k_b T \ln(Z)}{\partial \mu} \tag{1.62}$$

$$= -\frac{k_b T}{Z}\frac{\partial Z}{\partial \mu} \tag{1.63}$$

$$= \frac{\int \frac{\partial V(\boldsymbol{q})}{\partial \mu} \exp(-\beta V(\boldsymbol{q})) d\boldsymbol{q}}{\int \exp(-\beta V(\boldsymbol{q})) d\boldsymbol{q}} \tag{1.64}$$

$$= \left\langle \frac{\partial V(\boldsymbol{q})}{\partial \mu}\right\rangle_{\mu=\mu*} \tag{1.65}$$

The overall free energy change can be obtained by integrating the derivative along $\mu$. This approach is called thermodynamic integration [41]:

$$\Delta F = \int\limits_0^1 \left\langle \frac{\partial V(q)}{\partial \mu} \right\rangle d\mu \tag{1.66}$$

The integration can be performed through quadrature integration [42], which requires the calculation of the mean force at limited number of points. We note, that for particle annihilation or creation $\mu = 0$ yields singular integrand for Eq. 1.66 if $k = 1$ is used in Eq. 1.61 [42]. For practical applications $k \geq 4$ is necessary [42].

The method also has conceptual significance. Eq. 1.66 shows that the free enegy is the *potential of the mean force* (PMF) [43]. Besides the PMF notation, thermodynamics integration is a starting point to understand numerous other methods. For example blue moon sampling [44], which employs constrained $\mu$ values, or umbrella integration, which accumulates force during umbrella sampling simulation and adaptive biasing force methods [45]. The latter is similar to metadynamics, but instead of the underlying free energy, the mean force is compensated by the negative of the averaged forces. This method collects more information about the system, than adaptive biasing potential methods, as we save also the direction of the force. Nevertheless its implementation requires higher derivatives of the CVs with respect to atomic coordinates. In principle any of these methods can be applied to any CV.

Free energy methods are based on statistical averages, or iterative procedures to construct a biasing potential. As any statistical descriptor, or iterative procedure, free energy methods also have an error, that is needed to be estimated or converged until a predefined threshold. In free energy calculations the quantity of interest, such as barrier height, equilibrium constant can be converged. The simulations can be terminated upon achieving the predefined convergence criteria or the regime of small oscillations. The simulations can be considered as stochastic random processes. Therefore performing the same simulation with different initial conditions leads to independent samples of the same calculated quantities. The error corresponding any quantity can be calculated from the standard deviation of the independent simulations, and can be decreased by larger number of simulations.

## 1.4.2 Validating collective variables

Choosing a proper reaction coordinate is a challenging process. Usually several tests are necessary in order to explore the relevant degrees of freedom. Therefore it is important to validate the chosen CV set a posteriori. One of the requirements to have a good reaction coordinate is to describe the progress of the reaction. It implies that it can discriminate between reactant and product states. This can be easily tested by performing unbiased molecular dynamics simulations, starting from each state. The trajectories in the reaction coordinate space should have no overlap. In the case when

both states are known, this test should be done before starting expensive free energy calculations.

A proper reaction coordinate fulfils additional requirements. The proper set of the sufficient requirements is still not known. However there is a quantity, which is widely accepted as a test tool and a proper coordinate [46]. This quantity is the committor ($\phi(\boldsymbol{q})$), i.e. the probability that a configuration with random velocities reaches the stable product state before reaching the stable reactant state. Any reaction coordinate is expected to map the committor:

$$\phi(\boldsymbol{q}) \equiv \phi(\lambda(\boldsymbol{q})). \tag{1.67}$$

The committor of a configuration can be estimated from molecular dynamics simulations. In practice short trajectories are initialized with random velocities drawn from Boltzmann distribution. The trajectories are followed until they reach one of the stable states. If the simulation reaches the reactant its commitment is unity. If it reaches the product the commitment is zero. The committor of the configuration is the expectation value of the commitments.

There is an ensemble of geometries in the focus of our interest, defined by their commitment probability of 0.5. These configurations have the same probability to reach the reactant and product states. This set of such configurations is referred to as the *transition state ensemble*. In case of a double-well free energy surface the transition state ensemble is localized in the vicinity of the saddle of the free energy surface where the mean force is zero. The numerical procedure for examining transition state ensemble can be carried out in various ways. The typical procedure is the following. At the saddle of the approximate free energy surface a constrained or restrained molecular dynamical simulation is performed to sample the configurations representing the saddle region. Then trajectories are initiated from each configuration with random velocity. In the simple approach the average of single trajectory commitments is calculated. It is expected to be approximately 0.5. In a more elaborate, but much more computationally demanding approach, distribution of commitment averages is plotted and studied. Each commitment value is calculated as an average of multiple trajectories from a single configuration. As the theoretical distribution for commitment should be binomial with 0.5 as expectation value, the distribution should be narrow and single peaked around 0.5.

### 1.4.3 Optimization of reaction coordinates

Previously, two important physically meaningful scalar fields have been identified: free energy surface and the committor. In the followings the most important methods of dimensional reduction techniques are presented, which map the space of multiple CVs into a low dimensional set of CVs.

One of the first concepts in the field is to identify the minimum free energy path. The minimum free energy path is an object connecting the corresponding minima

of the reactant and product states. This one dimensional object is a path in the $n$ dimensional space of the $n$ CVs. Orthogonally to the path the mean force should be zero. This path can be constructed for example by exploring the full $n$ dimensional CV space and by optimizing the path on the high dimensional free energy surface. An important study of dimensional reduction techniques utilizes two CVs for this purpose: the progress along the path and the distance from the path [13]. The definition of the path is changed after each exploration of the free energy surface, in order to minimize the average of the distance from path CV (Eq.1.14) in each iteration. The path can be defined either by reference structures or by a set of $n$ dimensional CV vectors. For higher dimensions it is more economical to calculate the potential of mean force on-the-fly. This idea is employed by the *finite temperature string method* [47]. This method discretizes the path into equal length segments. The points at the end of each segment are following the direction of the mean force. The mean force can be calculated from constrained or restrained simulations. After this relaxation step, each CV is fitted with a spline by assigning row numbers for each point, and the points are then evenly redistributed along the path. The procedure is repeated until the convergence criteria are met. It is worth noting, that the component of the mean force in the path direction can be used to obtain the free energy profile along the path.

This optimization can be done in higher dimensions if it is necessary. An extreme example of that is when the applied CVs are the Cartesian coordinates of the atoms. In such case the free energy along the path can be partitioned to the atoms involved in the given chemical transformation [48].

Although the minimum free energy path is well-defined in a given CV space, projection to the path is not obvious, and to a certain extent it is ambiguous. Therefore progress along the path cannot be considered as proper reaction coordinate even if it is close to that [46]. It implies that even an excellent reaction coordinate for free energy calculations can be suboptimal for rate constant calculations or might provide unsatisfactory committor distributions.

An interesting idea to overcome this difficulty is to use the proper reaction coordinate, i.e the committor as practical reaction coordinate. In their groundbreaking paper, Ma and Dinner applied an estimate of the committor based on carefully trained and tested artificial neuron network [49]. They collected commitment data uniformly from the CV space, and trained the neural network with a least-squares cost function. Their study provided an important conceptual asset to the molecular simulation community. However their work did not prove to be practically feasible for realistic systems. Only 2 from ca. 150 one hundred and fifty citations are actual application of the method. Besides the complex training and testing procedure, its applicability is limited by the enormous amount of trajectories that are necessary for convergence.

Dynamical optimization of the parameters in contact map CV has been done by Best and Hummer [50]. They have optimized their reaction coordinate to obtain peaked average reactivity around the transition state. The average reactivity equals to the probability that a structure belongs to the ensemble of trajectories that connects the stable reactant and stable product states. This requirement can be achieved by the

maximization of the product of forward and backward committors, i.e. the product of commitment with a random velocity and with the inverted velocity.

Another important and widely used method is the likelihood maximization approach of Peters and Trout [51]. This method intends to find the optimal linear combination of CVs by maximizing the likelihood of a model committor function. The model function for the committor $\phi(\lambda_t)$ is defined as

$$\phi(\lambda_t(\boldsymbol{\alpha}, \boldsymbol{q})) = \frac{1 + \tanh(\lambda_t(\boldsymbol{\alpha}, \boldsymbol{q}))}{2}, \tag{1.68}$$

The free parameters of the model function ($\boldsymbol{\alpha}$) are optimized by maximizing the likelihood of the simulated commitment results through the following likelihood function:

$$L(\boldsymbol{\alpha}) = \prod_{i=1}^{PS} \phi(\lambda_t(\boldsymbol{\alpha}, \boldsymbol{q})) \prod_{i=1}^{\neq PS} (1 - \phi(\lambda_t(\boldsymbol{\alpha}, \boldsymbol{q}))) \tag{1.69}$$

The authors have also suggested an alternative formulation, where the central quantity is not the committor, but the average reactivity. In their latter work they concluded that the approach based on commitments is more reliable [52].

The commitment values are calculated through a path sampling algorithm which is a special variant of the *transition path sampling*(TPS) [53]. Detailed introduction to this family of methods is out of the scope of this thesis, but the main idea is presented here. The TPS algorithms start with the generation of an initial path, which can be accomplished in various ways. Then the trajectory is perturbed by one of the many available stochastic algorithms. The perturbation results in a new trajectory which is accepted or rejected based on a Monte Carlo acceptance criterion. The result of the TPS is an ensemble of reactive trajectories. Sampling of various physical quantities is possible for the path ensemble. In this particular case the *aimless shooting algorithm* is applied for path perturbation [51]. This approach assigns completely new velocities for the randomly chosen configurations at $-t$ or $+t$ from the previous shooting point. The completely new velocities ensure the fast decorrelation of the trajectories, while the biased sampling in the vicinity of the previous shooting point ensures high average reactivity and enhanced sampling around the dividing surface. Note that the reactant and product states can be connected by several reaction channels separated by high barriers. The application of the original TPS method is not appropriate for such systems. However, parallel tempering TPS technique is available to overcome this problem [54].

The combination of string method and maximum likelihood approach has been presented by Lechner, Bolhuis and co-workers [55]. The approach is similar to the original maximum likelihood approach, but the optimal reaction coordinate is assumed to be a string instead of a simple linear combination. This functional form is also able to describe complicated pathways. Its optimization is accomplished through Monte Carlo moves of the string images. The moves are rejected or accepted, depending on the variation of the logarithmic likelihood function.

## 1.5   Zero temperature limit

In this section the zero temperature limit of the presented expressions is discussed. The zero temperature limit is useful to derive approximate equations. These equations have limited validity but they provide us with a cost effective procedure to calculate fundamental physical quantities.

As Eq. 1.37 suggests in the zero temperature limit the average of any quantity is replaced by the actual value of the quantity at the lowest energy configuration. Depending on the statistical weights of other configurations this approximation can be applied at higher temperatures as well. The effect of temperature can be approximated by including contributions from the truncated series of the PES. [56]

One elegant example for zero temperature limit is deriving *zero temperature string method* [57] from finite temperature string method. Molecular dynamics in zero temperature is equivalent with small steps in the direction of the force, with instantaneous dissipation of the thermal energy. The method essentially results in repetitive geometrical optimization and image redistribution steps. The procedure converges to the *minimum energy path* (MEP) instead of the minimum free energy path.

### 1.5.1   Static calculations

The Helmholtz or the Gibbs free energies can be approximated from the local characteristics of the PES and structural properties of a molecular system. This treatment is a good approximation for a gaseous system where the molecules do not interact with each other. In such a case the partition function of the system ($Q$) can be treated as a product of the molecular partition functions. However if the molecules are identical, the equivalence of the different permutations should be taken into account by a $1/N_m!$ factor, where $N_m$ is the number of molecules. The partition function can be expressed as:

$$Q = \frac{Q_m^{N_m}}{N_m!}, \tag{1.70}$$

where $Q_m$ is the molecular partition function. The relevant thermodynamically quantities can be expressed as function of the molecular partition function and its derivatives:

$$S = R + R\ln(Q_m)) + RT\ln\left(\frac{\partial \ln Q_m}{\partial T}\right)_V \tag{1.71}$$

$$U = RT^2\left(\frac{\partial \ln Q_m}{\partial T}\right)_V \tag{1.72}$$

$$H = pV + U. \tag{1.73}$$

In the previous sections exclusively classical equations have been presented. However in the present approximation an efficient quantized formalism is available. The

molecular partition function can be expressed as:

$$Q_m = \sum_i g_i \, e^{-\beta \epsilon_i}, \tag{1.74}$$

where $g_i$ is the degeneracy and $\epsilon_i$ is the energy of the i$^{\text{th}}$ level. A further approximation can be employed for the different degrees of freedoms. The molecular energy levels of translation $(t)$, rotation $(r)$, vibration $(v)$ and the electronic $(e)$ energy are assumed to be independent. In the followings the physical quantities corresponding these degrees of freedom are denoted with the initials in brackets. If the energy contributions of these degrees of freedom are assumed to be independent, the total energy equals to the following sum:

$$\varepsilon = \varepsilon_t + \varepsilon_r + \varepsilon_v + \varepsilon_e. \tag{1.75}$$

It can be shown, that in this case the molecular partition function can be calculated as a product of the corresponding factors:

$$Q_m = Q_{m,t} \cdot Q_{m,r} \cdot Q_{m,v} \cdot Q_{m,e}. \tag{1.76}$$

The factorization leads to additive contributions to the entropy and to the internal energy:

$$S = R + S_t + S_r + S_v + S_e \tag{1.77}$$
$$U = U_t + U_r + U_v + U_e. \tag{1.78}$$

$R$ is the universal gas constant. If the relevant energy levels are known and the partition function is calculated, the entropy and internal energy contribution of the specific degrees of freedom can be calculated as:

$$S_x = R \ln (Q_{mx}) + RT \ln \left( \frac{\partial \ln Q_{mx}}{\partial T} \right)_V \tag{1.79}$$

$$U_x = RT^2 \left( \frac{\partial \ln Q_{mx}}{\partial T} \right)_V, \tag{1.80}$$

where $x \in \{t, r, v, e\}$ . The translational contribution can be exactly decoupled from the other degrees of freedom. Its contribution can be calculated from the particle in the box model:

$$Q_{m,t} = \left( \frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \frac{k_B T}{P}, \tag{1.81}$$

which leads to following entropy and internal energy contributions:

$$S_t = R \ln Q_{m,t} + \frac{3}{2} R \tag{1.82}$$

$$U_t = \frac{3}{2} RT \tag{1.83}$$

The rotational movement can be approximated as a rigid rotor. Rigidity assumption removes coupling with vibrations and facilitates simple analytic solution. The rotational

molecular partition function can be calculated from the molecule's principal moments of inertia $\Theta_r$:

$$q_r = \frac{1}{\sigma}\sqrt{\pi \frac{T^3}{\Theta_{r,A}\Theta_{r,B}\Theta_{r,C}}}. \tag{1.84}$$

Its contribution to the entropy and internal energy can be expressed as:

$$S_r = R\ln\left(Q_{m,r} + \frac{3}{2}\right) \tag{1.85}$$

$$U_r = \frac{3}{2}RT. \tag{1.86}$$

The vibrational contribution can be calculated in the harmonic approximation. If the vibrations in a system with $N_{atoms}$ number of atoms are harmonic, the $3N_{atoms}$-6 vibrations decouple into independent one dimensional oscillators. As the solution for the quantum oscillator is known, its partition function can be calculated:

$$Q_{m,v} = \prod_K \frac{e^{-\frac{\beta h\nu_K}{2}}}{1 - e^{-\beta h\nu_K}}, \tag{1.87}$$

which requires the calculation of harmonic frequencies $\{\nu_K\}$. Harmonic frequencies are readily available from the Hessian of the potential energy surface and the atomic masses. However their calculation might result in serious errors if internal rotors are not identified and calculated as vibrations. The relevant thermodinamic contributions are:

$$S_v = R\sum_K\left(\frac{\beta h\nu_K}{e^{\beta h\nu_K} - 1} - \ln\left(1 - e^{\beta h\nu_K}\right)\right) \tag{1.88}$$

$$U_v = RT\sum_K \beta h\nu_K\left(\frac{1}{2} + \frac{1}{e^{\beta h\nu_K} - 1}\right) \tag{1.89}$$

### 1.5.2   Geometry optimization

The special importance of the optimized structures have been emphasized in the previous section. There are efficient algorithms for finding the critical points of the PES. One basic approach is to consider the surface to be quasi-quadratic around the critical point. In other words, approximating the surface by its truncated Taylor-series results in a one-step convergence if this approximation is valid. This concept is referred to as Newton or Newton-Raphson method. In the quasi harmonic approximation the best assumption for the minimal energy structure can be calculated as:

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + \Delta\boldsymbol{q} \tag{1.90}$$

$$= \boldsymbol{q}_n - \boldsymbol{g}_n\mathbf{H}_n^{-1}, \tag{1.91}$$

where $\boldsymbol{g}_n$ is the gradient and $\mathbf{H}_n$ is the Hessian of the PES at the $\mathrm{n}^{\mathrm{th}}$ configuration $\boldsymbol{q}_n$. This method is the basis of many current algorithms. However, the calculation

of the Hessian in each step is computationally demanding, therefore the Hessian is usually updated after its initial exact or approximate calculation. Different algorithms are available for the Hessian update such as BFGS [58–61] or SR1 [62]. Convergence of the method could be enhanced by introducing a level shift in the Hessian matrix. This can be accomplished for example by a suitable diagonal matrix $\xi$ [63].

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n - \boldsymbol{g}_n(\mathbf{H}_n - \xi)^{-1} \tag{1.92}$$

This approach scales the steps in order to avoid too large steps. The approach is also suitable for modifying the direction of the search and facilitates search for saddle points of the potential energy surface. The *direct inversion of the iterative subspace* (DIIS) technique from Pulay [64], originally developed for enhancing the convergence of electronic structure methods, The method was adapted to geometry optimization as GDIIS [65], which have been further modified (GEDIIS [66]) by improvements of Scuseria and co-workers for the SCF problem [67]. The underlying idea is that the optimal geometry can be derived as an interpolation of the previous trial geometries. The procedure expresses the current geometry as a linear combination of the previous geometries. The coefficients are determined to minimize the norm of an approximate error vector, under the constraint that the square sum of the coefficients is equal to one. $\boldsymbol{q}_n$ and $\boldsymbol{g}_n$ are also expressed as a linear combination of the previous vectors with the optimal coefficients:

$$\boldsymbol{q}_{n+1} = \sum_i^n c_i \boldsymbol{q}_i - \sum_i^n c_i \boldsymbol{g}_i(\mathbf{H}_n - \xi)^{-1} \tag{1.93}$$

In practice the algorithms are switched depending on the RMSD threshold and also the Hessian update is varied during efficient algorithms.

## 1.6 Microscopic definition of rate constant

The rate of an elementary reaction and the rate constant have been defined in section 1.1.1 in a phenomenological manner. It was based on a macroscopic quantity: the concentration. This section focuses on the various microscopic definitions, and the possible numerical techniques to calculate rate constants from molecular level simulations. As these simulations are extremely expensive some approximate approaches are also presented.

### 1.6.1 Concepts and definitions in rate constant calculations

The equations for the description of reactions can be generally applied for any rare events. Different chemical and physical transformations, such as conformational changes, bond forming or breaking reactions or collective molecular rearrangements such as crystallisation are all rare events. They can be rationalized on the basis of a free energy profile along a reaction coordinate as a double well problem. They

are also typically characterized by separation of time-scales. Although the reactions occur rarely on molecular time scales, the actual reactive events proceed usually much faster. In other words the reactive events cover negligible time frame comparing to the typical times between two consecutive events. This relation leads to formally different microscopic definitions of rate constants with almost no practical discrepancy.

The definitions are based on an ideal equilibrium situation between two states: the reactant $(RS)$ and the product state $(PS)$. In terms of reaction coordinate the two states are adjacent. One can imagine an infinitely long canonical molecular dynamics simulation with infinite number of transitions from $RS$ to $PS$ and vice versa. The rate constant can be formulated through the *mean residence time* in the state $RS$:

$$
\begin{aligned}
k_f &= \frac{1}{\langle t_{RS}^r \rangle} & (1.94) \\
&= \frac{N^t}{t_{RS}^{tr}}, & (1.95)
\end{aligned}
$$

where $t_{RS}^r$ is the residence time in the reactant state, while $t_{RS}^{tr}$ is the total residence time in the reactant state for $N^t$ number of forward reactions. An alternative definition is based on the *mean first passage time* from $RS$ to $PS$ ($\langle t_{RS}^p \rangle$) [8]. The passage time is defined as the time necessary to reach $PS$ if the system started from $RS$. Note, that this definition is equivalent to the previous one with adjacent states. However in some formulations the non-adjacent stable states are applied, which leads to some discrepancy. The third definition of the rate constant is the transition probability from $RS$ to $PS$ per unit time [8].

## 1.6.2 Methods for rate constant calculations

The mean first passage time can be calculated by the Bennett-Chandler [68] [69] procedure. The usual derivation of the method is based on state indicator functions (order parameters): $h_{RS}(\boldsymbol{q})$ and $h_{PS}(\boldsymbol{q})$ are defined to be zero if $\boldsymbol{q}$ is out of the specific state, and unity if $q$ is within the state. The state correlation functions are used to construct a correlation function:

$$
C(t) = \frac{\langle h_{RS}(\boldsymbol{q}(0)) h_{PS}(\boldsymbol{q}(t)) \rangle}{\langle h_{RS}(\boldsymbol{q}) \rangle} \qquad (1.96)
$$

It can be shown that if there is a separation between the time scales, the $C(t)$ is approximately proportional to the forward rate constant in between the molecular and the relaxation time range. In other words for longer times than a reactive event, but shorter times than $1/k$, $C(t)$ can be approximated as:

$$
C(t) \approx k_f(t) \cdot t. \qquad (1.97)
$$

This result can be used by calculating the time derivative of the correlation function. In the followings $\lambda(q(t))$ is denoted as $\lambda_t$ and the value of the reaction coordinate

corresponding to the dividing surface is $\lambda^*$ The time derivative of the correlation function is:

$$\dot{C}(t) = \frac{\left\langle \dot{\lambda}_0 \Theta(\lambda_t - \lambda^*) \delta(\lambda^* - \lambda_0) \right\rangle}{\langle \Theta(\lambda^* - \lambda_0) \rangle}. \tag{1.98}$$

where $\Theta$ and $\delta$ are the Heaviside and the Dirac delta functions.

This expression can be simplified by the concept of the free energy surface, and its definition [8]:

$$k(t) = \dot{C}(t) \tag{1.99}$$

$$= \left\langle \dot{\lambda}_0 \Theta(\lambda_t - \lambda^*) \right\rangle_{\lambda_0=\lambda^*} \frac{\exp(-\beta F(\lambda^*))}{\int\limits_{RS} \exp(-\beta F(\lambda)) d\lambda} \tag{1.100}$$

$$= R(t) \frac{\exp(-\beta F(\lambda^*))}{\int\limits_{RS} \exp(-\beta F(\lambda)) d\lambda}. \tag{1.101}$$

These expressions are not only fundamental, but also very interesting. First of all we note that the rate constant is time dependent in this picture. It has a plateau region, where its value is independent from the actual choice of $t$. The time dependence is exclusively condensed into the factor $R(t)$, which is called the *unnormalized transmission coefficient* [8]. This factor accounts for the dynamical properties of the system, such as tendency for recrossing during a reactive event. The rest of the expression is purely static quantity. Interestingly it does not depend on the masses of the atoms. Therefore it could also be calculated from Monte Carlo simulations.

In spite of its historical role, the BC procedure is based on error cancellation. The error cancellation is exact in the case of infinite sampling. Eq. 1.100 expresses a simple physical picture. The static part calculates the probability to find the system at the top of the barrier. $R(t)$ calculates the reactive flux through the border of $RS$ and $PS$, through the *dividing surface*$(\lambda = \lambda^*)$. The $\Theta(\lambda_t - \lambda^*)$ factor simply filters trajectories which are initiated at the dividing surface $(\lambda^*)$, and reach the $PS$ after time $t$. The true reactive flux involves only trajectories that are initiated from the reactant state and reach in the product state. The time elapsed since the trajectory previously entered from $PS$ to $RS$ must be much longer than the time length of the actual reactive event. Otherwise the event is nothing else but a recrossing of an unsuccessful backward reaction attempt. Therefore this expression calculates the right average. However non-reactive trajectories will have positive and negative contributions to the average. The expression have been improved in the BC2 expression [68] [69]:

$$R(t)^{BC2} = \left\langle \dot{\lambda}_0 \Theta(\lambda_t - \lambda^*) \Theta(\lambda^* - \lambda_{-t}) \right\rangle_{\lambda_0=\lambda^*} \tag{1.102}$$

Eq. 1.102 calculates the flux through the dividing surface (at $\lambda = \lambda^*$) for trajectories which started in $RS$ at time $-t$ and arrived at $PS$ at time $t$. In practice, the BC2 procedure follows the simple approach:

- an optimal reaction coordinate ($\lambda$) is defined

- the free energy along this putative reaction coordinate is calculated

- a constrained simulation is carried out at the top of the free energy barrier

- configurations are sampled during the constrained simulation

- the configurations are assigned with random velocities and propagated until time $t$

- for each configuration, the assigned initial velocity is inverted, and the trajectory is back-propagated until $-t$

- Eq. 1.102 is calculated by averaging the $\dot{\lambda}_0$ for each configuration, but switching the contribution to zero if the configuration did not correspond to a true reactive event.

The BC2 approach raised interesting numerical issues. Namely that it averages both negative and positive contributions from reactive trajectories. In fact, the same trajectory with multiple recrossings is sampled at its different dividing surface intersections. At each crossing in the positive direction the trajectory will contribute with a positive $\dot{\lambda}_0$, while each crossing in the opposite direction will result in a negative contribution. These terms cancel out each other, and a trajectory with multiple recrossings will contribute just once, as if it had only one recrossing. The cancelling of the forward and backward recrossings is guaranteed by the canonical sampling. Although a fast recrossing has larger contribution, the probability to sample that crossing point is smaller than that of a slow crossing point. This argument explains that the true contribution of a single trajectory is obtained only in the limit of an infinite canonical sampling of its crossing points.

Another solution for accounting the recrossings is the application of the positive flux expression [70]. The algorithm is exactly the same as that of BC2, but the rate constant is calculated through an expression that gives non-zero contribution only for positive crossings. This feature results in better convergence.

$$R(t)^{PF} = \left\langle \dot{\lambda}_0 \Theta(\lambda_t - \lambda^*)\Theta(\dot{\lambda}_0) \right\rangle_{\lambda_0=\lambda^*} - \left\langle \dot{\lambda}_0 \Theta(\lambda_{-t} - \lambda^*)\Theta(\dot{\lambda}_0) \right\rangle_{\lambda_0=\lambda^*} \qquad (1.103)$$

This expression is much more efficient than BC2. However it still relays on error cancellation. A forward reactive trajectory with $n$ crossings might contribute at each crossing point if the specific crossing point is sampled. This could result in an over-counting of the reactive flux. However, the second term of the expression gives $n-1$ negative contributions if the expression is evaluated on the trajectory with inverted time.

As it is explained previously, accounting for multiple recrossings is the most problematic part of the rate constant calculations. This problem has been successfully treated by the effective positive flux (EPF) approach of van Erp and Bolhuis [71]. This

method gives a very elegant solution for the multiple crossings problem by introducing the concept of stable $RS$ and stable $PS$, and a new indicator function for a trial trajectory, denoted with $\chi_f^{EPF}(\boldsymbol{q(0)})$. The indicator function is unity if the backward propagated trajectory leads to $RS$ without recrossing $\lambda^*$ and if the forward propagated trajectory reaches the $PS$ before reaching the $RS$. This complex indicator function is used in the followingle expression:

$$R(t)^{EPF} = \left\langle \dot{\lambda}_0 \Theta(\dot{\lambda}_0) \chi_f^{EPF}(\boldsymbol{q(0)}) \right\rangle_{\lambda_0 = \lambda^*} \tag{1.104}$$

Similar expression was applied in the method of Anderson [72], which follows trajectories in time until the rate constant reaches a plateau. It calculates the velocity of the last crossing only.

An interesting expression was been derived by Hummer [73]. The formula calculates rate constants from transition path ensemble. The method defines the $N_i$ reactivity index for the i$^{th}$ trajectory. It is unity for reactive paths, and zero for nonreactive pathways. The method calculates the relative probability to find the system at the dividing surface ($f_{eq}(\lambda^*)$), the ensemble average of the ratio of reactivity ($N_i$) and the time spent at all crossing points for forward and backward propagated trajectories initiated at the dividing surface. Formally:

$$\frac{1}{k_f^{-1} + k_b^{-1}} = c_r k_f = c_p k_b = \left\langle \frac{N_i}{\sum_i |v_i|^{-1}} \right\rangle_{ds} f_{eq}(\lambda_{ds}) \tag{1.105}$$

This method has an interesting feature. It samples the transition path ensemble and gives important quantities such as committor as well. Moreover, each reactive trajectory in the sample gives its true contribution to the rate constant irrespective of which dividing surface crossing point was used as its starting point. On the other hand, the algorithm follows each trajectory till the stable reactant or stable product states. Therefore the EPF method might be more effective for systems with frequent, long, non-reactive trajectories. In contrast, other quantities, such as average reactivity or committors can be obtained from the method of Hummer.

All the methods presented above are based on the sampling of a single dividing surface. This is an obvious possibility if the dividing surface is known a priori. Methods designed to sample wider range of the reaction coordinate are presented in the followings.

An alternative formulation of the BC rate constant leads to an expression which utilizes starting points from the full reaction coordinate range between the stable reactant and product states. This expression has been derived by Ruiz-Montero, Frenkel and Brey [74]. Their derivation starts from similar indicator functions as the BC procedure, however they apply smooth functions instead of Heaviside function.

The proposed indicator function is:

$$h_{PS}(\lambda) = \frac{\int\limits_{sRS}^{\lambda} \exp(\beta F(\lambda'))d\lambda'}{\int\limits_{sRS}^{sPS} \exp(\beta F(\lambda'))d\lambda'}. \qquad (1.106)$$

Here integration limits $sRS$ and $sPS$ denote the vale of $\lambda$ corresponding to the border of the stable reactant and stable product states. The ensemble of starting configurations are sampled by applying a $-2F(\lambda)$ biasing potential. Finally the time-dependent expression for rate constant is calculated:

$$k_f(t) = \frac{1}{\langle h_A \rangle} \frac{\left\langle \int\limits_{0}^{t} \dot{\lambda}_0 \dot{\lambda}_{t'} \exp(\beta F(\lambda_{t'}) - \beta F(\lambda_0))dt' \right\rangle}{\int\limits_{A}^{B} \exp(\beta F(\lambda))d\lambda \int\limits_{A}^{B} \exp(-\beta F(\lambda))d\lambda}. \qquad (1.107)$$

The expression is similar to the Green-Kubo relation for diffusion constant [75, 76]. It is important to note that this method can be efficient for diffusive cases as it accumulates velocity information not only at single dividing surface crossing points but also in the neighbourhood of the barrier. It can also utilize calculated trajectories if the exact position of the dividing surface is not known. This feature of the method is a remarkable advantage as the calculation of the dividing surface is computationally demanding.

Transition path sampling is an alternative option for calculating rate constants [53]. This approach requires to calculate the function $P_{RS}(\lambda, t)$, the probability that a path reaches $\lambda$ at time $t$ provided that it started at the stable $RS$. As the probability to reach values of $\lambda$, that belonging to the $PS$ is very low, it requires advanced path sampling techniques, namely the path sampling analogue of umbrella sampling. The time correlation function $C(t)$ can be calculated from an integral of $P_{RS}(\lambda, t)$ with respect to $\lambda$.

The formulation of EPF has been utilized in the method called transition interface sampling [71]. The method employs multiple interfaces, and calculates the crossing probabilities. $P(\lambda_{i+1}|\lambda_i)$ is the probability that a trajectory which started at $\lambda_0$ and crosses $\lambda_i$ reaches $\lambda_{i+1}$ before reaching again $\lambda_0$. The rate constant can be calculated as:

$$k_f = \frac{N_c^+}{t} \prod_{i=1}^{n-1} P(\lambda_{i+1}|\lambda_i) \qquad (1.108)$$

where $N_c^+$ is the number of positive crossings calculated at the first interface in time $t$ in long unbiased simulation at the reactant state.

### 1.6.3 Transition State Theory

Various methods have been introduced for calculating rate constant for elementary reactions as shown in the previous section. These methods require extra computations beyond the calculation of the free energy surface along the reaction coordinate. However, the kinetics of complex reaction mechanisms can be simplified if there are parallel reaction channels with rate constants that differ in orders of magnitude. Therefore approximate calculation of the rate constants with significantly less computational effort has great practical importance. In the followings the *transition state theory* (TST) is presented, which facilitates the approximate calculation of rate constants [77].

TST assumes that if the system crosses the dividing surface then it leads to a reaction. This assumption is considered valid for the backward reactions as well. This implies that the theory neglects the recrossings. Furthermore it also relies on the assumption of quasi equilibrium, which means that the probability to find the system in the dividing surface can be derived from the equilibrium distribution. The latter assumption has been also implicitly used in all Bennett-Chandler type methods. The unnormalized transmission coefficient of TST can be easily derived from the analogous quantity appearing in the positive flux formalism (Eq. 1.103). Since in TST we assume that there are no recrossing trajectories, all trajectory with positive $\dot\lambda$ will be forward reactive. In Eq. 1.103 the $\Theta(\dot\lambda)$ will filter out all backward trajectories, and $\Theta(\lambda_t - \lambda^*)$ will be unity for all reactive trajectories. Therefore the first term simplifies to $\left\langle \dot\lambda\Theta(\dot\lambda) \right\rangle_{\lambda=\lambda^*}$. The second term of the positive flux expression provides only zero contributions within this assumption, as no trajectory can fulfil the two conditions of positive $\dot\lambda$ and product side origin if no recrossing occurs. In this way we obtain the rate expression for TST as:

$$k_f^{TST} = \left\langle \dot\lambda\Theta(\dot\lambda) \right\rangle_{\lambda=\lambda^*} \frac{\exp(-\beta F(\lambda^*))}{\int\limits_{RS} \exp(-\beta F(\lambda))d\lambda} \qquad (1.109)$$

The $\left\langle \dot\lambda\Theta(\dot\lambda) \right\rangle_{\lambda=\lambda^*}$ factor simplifies to $\frac{1}{2}\left\langle |\dot\lambda| \right\rangle_{\lambda=\lambda^*}$ in case of a symmetric distribution for $\dot\lambda$. This factor can be calculated from canonical constrained simulation at $\lambda = \lambda^*$. In the particular case, when the reaction coordinate is a Cartesian coordinate, the average can be evaluated analytically [8]

$$\frac{1}{2}\left\langle |\dot\lambda| \right\rangle_{\lambda=\lambda^*} = \frac{1}{\sqrt{2\pi\beta m}} \qquad (1.110)$$

As it has been described earlier, free energy differences can be approximately calculated from the local properties of critical points of the PES. TST can be formulated in this language as well. The description leads to expressions similar to Eq. 1.109, with minor differences. Since in this picture, states are defined as optimized structures and their vicinity, the *RS* will be replaced with a single molecular structure, and the dividing surface will be replaced by the configuration corresponding to the saddle point, the *transition state*. Probability to find the system at the transition state can

be calculated from the ratio of the canonical partition functions:

$$K^{\ddagger} = \frac{Q_{TS}}{Q_{RS}} = \exp\left(-\beta \Delta G^{\ddagger}\right), \qquad (1.111)$$

where

$$\Delta G^{\ddagger} = \Delta G_{TS} - \Delta G_R. \qquad (1.112)$$

Here $\Delta G_{TS}$ and $\Delta G_R$ are the Gibbs free energies corresponding to the TS and the reactant(s), respectively. It is important to note, that the Gibbs free energy corresponding to a $TS$ is calculated identically to the stable structures (section 1.5.1 with one difference: contribution from the normal coordinate describing the motion along the reaction coordinate is not included. The contribution from this degree of freedom is included in the pre-exponential factor. The pre-exponential factor can be calculated from different models, such as particle in the box or zero frequency limit of harmonic oscillator:

$$k^{\ddagger} = \frac{1}{\beta h} \qquad (1.113)$$

Finally the rate constant is derived as a product of the $k^{\ddagger}$ and $K^{\ddagger}$:

$$k_0^{TST} = \frac{1}{\beta h} \exp\left(-\beta \Delta G^{\ddagger}\right) \qquad (1.114)$$

It is important to note the differences between Eq. 1.114 and Eq. 1.109. Although both expressions are based on the calculation of the probability of finding the system at the dividing surface or at the transition state from free energy information, this information is calculated from different principles. It is very tempting to calculate the barrier height from a free energy surface and substituting it into Eq. 1.114, but it is both formally and practically incorrect. The height of the barrier depends on the CV describing the reactive event. Therefore the barrier height could be modified by a mathematical transformation of the reaction coordinate. Note that this dependence is cancelled by the reaction coordinate dependence of the pre-factor in exact rate calculation methods.

## 1.7  Calculation of approximate PESs

The equations for the calculation of the probability density or rate constant implicitly relied on the assumption that a given configuration can be assigned to a potential energy value. This is not trivially true, but it is one of the fundamental assumptions of theoretical chemistry, which is called the Born-Oppenheimer approximation. The non-relativistic potential energy surface can be exactly derived through wave function theory, which is briefly introduced in the followings. Then the density functional theory is presented. The section introduces some more approximate methods, and finally summarize the applied force fields.

## 1.7.1 Wave function theory (WFT)

The ground state exact potential energy can be calculated from the following functional:

$$E[\Psi] = \frac{\left\langle \Psi | \hat{H} | \Psi \right\rangle}{\langle \Psi | \Psi \rangle} \qquad (1.115)$$

$\hat{H}$ is the Hamiltonian operator of the electronic subsystem within the Born-Oppenheimer approximation and $\Psi$ is its eigenvector corresponding to the lowest eigenvalue. As the energy spectrum for bounded states has a lowest eigenvalue, the exact eigenvector (wave function) can be found by the variation of the functional with respect to $\Psi$. In practice this can be accomplished by the expansion of the wave function as a linear combination of an orthogonal set of functions, and by the variation of the coefficients. The full configuration interaction method uses the set of Slater-determinants as a basis for this expansion.

Although this method is exact, it utilizes an infinite expansion. Moreover, the method has a factorial scaling, which prohibits practical applications. Approximations can be done by truncation of this series. However it results in size consistency error. The relevant coefficients and the energy can be obtained through alternative, non-variational approaches as well, such as coupled cluster or perturbation theory. The former and specific forms of the latter have the advantage over configuration interaction, namely that these are size consistent methods.

Even if the $N!$ scaling can be considerably reduced to $N^6$ or $N^7$ without sacrificing the applicability for well behaving systems (where N is the number of electrons); large systems, high throughput calculations, and systems with strongly correlated electronic structure motivated the development of alternative formalisms.

## 1.7.2 Density functional theory (DFT)

The DFT expresses the energy as a functional of the total ground state electron density:

$$E = F[\rho]. \qquad (1.116)$$

This description could be highly beneficial as the electron density is a three dimensional function, while the wave-function is a much more complicated $3N$ dimensional object. Although Hohenberg and Kohn proved an existence theorem and also a variational theorem for this functional fifty years ago [78], its form is still not known. The main difficulty is that the kinetic energy of the electronic system could not be sufficiently modelled as a functional of the electron density. As this term has large contribution to the energy, any approximation causes severe error.

The practical importance of the density functional theory can be dated to the introduction of the Kohn-Sham formalism [79]. In this picture some more complex variables have been introduced into the functional, such as the wave-function of a non-interacting electronic system, with the same density as the interacting system

of interest. The Schrödinger-equation can be solved exactly for a non-interacting electronic system and its energy is defined by the density. Therefore the exact electronic energy can be calculated from the density. The only problem to be solved is to find the proper potential, which turns the non-interacting electron density to be equivalent with the interacting electron density.

$$E_0 = E[\rho_0] = \left\langle \Psi[\rho_0] \left| \hat{T} + \hat{V}_s \right| \Psi[\rho_0] \right\rangle \tag{1.117}$$

There is no exact external potential or equivalently exact Kohn-Sham functional either. However as the kinetic energy and the electron-electron, nuclear-nuclear and electron-nuclear classical Coulomb terms are known, only a small portion of the total energy needs to be modelled. This remaining part of the electronic energy describes the exchange interaction between parallel spin electrons, and the correlation interaction. This part of the functional is referred to as exchange-correlation functional. The functional derivative of the exchange-correlation functional provides the last term ($V_{XC}[\rho(\boldsymbol{r})]$) to the potential:

$$V_s(\boldsymbol{r}) = \sum_{i \neq j} \frac{Z_i Z_j}{|\boldsymbol{q}_i - \boldsymbol{q}_j|} - \sum_i \frac{Z_i}{|\boldsymbol{r} - \boldsymbol{q}_i|} + \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \mathrm{d}^3\boldsymbol{r}' + V_{XC}[\rho(\boldsymbol{r})] \tag{1.118}$$

where $\boldsymbol{r}$ is the electronic degrees of freedom, and $Z_i$ is the charge of the $i^{\text{th}}$ nucleus. The exchange-correlation functional has been modelled in various ways. The approximate functionals might depend on other functions besides the electron density [80](as in local density approximation), such as spin-densities (local spin-density approximation) or the gradients of the spin densities (generalized gradient approximation, GGA) or the second derivatives of the spin densities (meta-generalized gradient approximation, meta-GGA) [81]. As excluding the self-interaction error is challenging in density functional theory [82], the Kohn-Sham orbitals and Hartree-Fock exchange expression are also often used in approximate exchange-correlation functionals (hybrid functionals) [83]. Distance based mixing of Hartree-Fock and DFT exchange is used in range separated functionals [84]. Recently second order perturbation theory on the Kohn-Sham orbitals has been utilized in order to improve precision [85].

The description of dispersion is not sufficient for the most density functionals. Therefore often semiempirical dispersion is added to the energy. A frequently used formulation from Grimme and co-workers [86] gives two and three atom contributions depending exclusively on atomic positions.

The two atom contribution is calculated in the following expression:

$$E^{(2)} = \sum_{AB} \sum_{n=6,8,10...} s_n \frac{C_n^{AB}}{d_{AB}^n} D_n(d_{AB}), \tag{1.119}$$

where $s_n$ is a functional dependent scaling factor, while $C_6^{AB}$ is calculated from the Casimir-Polder [87] formula through time-dependent density functional theory from

hydrides. Higher order coefficients are calculated from $C_6^{AB}$ with recursive expressions. The coefficients can be calculated in a system dependent manner. The numbers of neighbours are calculated for each atom through a Fermi coordination number formula such as Eq. 1.10 and for intermediate numbers of coordination the property map formula (Eq. 1.13 is used for interpolation). Finally $D_n(d_{AB})$ is the so-called dumping function which eliminates the singularity at zero distance ($d_{AB} = 0$). Various forms of Fermi functions (Eq. 1.10) are in use for this purpose. Three-body terms are calculated through the Axilrod–Teller–Muto formula [88, 89]:

$$E^{ABC} = \frac{C_9^{ABC}(3\cos\alpha\cos\beta\cos\gamma + 1)}{(d_{AB}d_{BC}d_{AC})^3} \tag{1.120}$$

where the $C_9^{ABC}$ coefficient is calculated from the pairwise $6^{\text{th}}$ order coefficients. Angels and distances are denoted according to usual notion of an ABC triangle.

### 1.7.3 Summary of the applied DFT forcefields

The BLYP exchange correlation functional is a cost efficient GGA functional. It contains exchange functional of Becke [90], and the correlation term of Lee, Yang and Parr [91]. The exchange part contains only a simple correction to the LSDA expression, with a single parameter [90]. The correlation term includes two fitted parameters. As the functional neglects long-range dispersion, it is advantageous to apply together with Grimme-corrections [86].

The M06 functional of Thrular and co-workers [92] is a hybrid-meta-GGA exchange-correlation functional. Although the functional is strongly semiemperical (the optimization procedure involved 35 parameters), it fulfils exact important criteria and designed with deep physical understanding. The functional development involved a versatile strategy with the following elements: constraint satisfaction (3 constraints to obtain correct uniform electron gas limit), empirical fitting, mixing of Hartree–Fock and approximate DFT exchange and the modelling of the exchange-correlation hole. The correlation energy is calculated differently for parallel spin and antiparallel spin interactions and the former vanishes for any one-electron system. The exchange term is a linear combination of the Hartree–Fock and DFT exchange. The latter is a mixture of the PBE [93] and the LSDA description.

In comparison with 12 other functionals, M06 proved to be outstanding in main-group thermochemistry, barrier heights, non-covalent interactions, and organometallic thermochemistry. A recent validation of the method has shown that it is reliable also for organometallic kinetics [94]. It can also handle difficult cases, such as ligand dissociation. As this fuctional describe the middle-range correlation as well, it is usually applied without additional dispersion correction.

The $\omega$B97XD is a range separated hybrid functional from Head-Gordon and co-workers [95]. The functional contains pairwise additive dispersion term to $\omega$B97X [96], which itself relies on 17 fitted parameters. Unlike many long range corrected hybrid functionals, it contains a small fraction of exact-exchange at short distances as well.

The optimization of the functional has also involved constrains to fulfil the uniform electron gas limit.

## 1.7.4 Molecular mechanics (MM)

As it was shown in the previous section, weaknesses of a given approximate forcefield can be compensated with physically motivated additive $N$-atom contributions. This is not only true for dispersion, but for other quantities as well. In general the potential energy surface can be approximated as a sum of $N$-atom potentials:

$$E(R) = \sum_{n=1} U_n(R). \tag{1.121}$$

Since the particular form of $U_n(R)$ can be derived from different principles, various formulations are in use in molecular modelling. The potential energy in the Amber forcefield [97] is the following expression:

$$
\begin{aligned}
E(\boldsymbol{q}) \;=\; & \sum_{\text{bonds}} k_b(d - d_0)^2 \;+ \\
& \sum_{\text{angles}} k_a(\theta - \theta_0)^2 \;+ \\
& \sum_{\text{torsions}} \sum_n \frac{1}{2} V_n[1 + \cos(n\omega - \omega_0)] \;+ \\
& \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij}\epsilon_{ij}\left[\left(\frac{d_{0ij}}{d_{ij}}\right)^{12} - 2\left(\frac{d_{0ij}}{d_{ij}}\right)^{6}\right] \;+ \\
& \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij}\frac{Q_i Q_j}{4\pi\epsilon_0 d_{ij}}.
\end{aligned}
$$

The first two summations are the applications of the harmonic approximation for the potential energy in internal coordinates (interatomic distances, and angles). The third summation describes the potential energy surface along torsional angles. The potential energy is extended into a few Fourier components. The fourth line of the expression contains the Lennard-Jones model potential for van der Waals interactions. Comparison with the Grimme potential shows, that it neglects important three-body contributions, and also some higher order two-body terms. The fifth line is a simple Coulomb term with static atomic point charges.

# Chapter 2

# Applications

# 2.1 Palladium catalysed C–H activation reaction [1]

## 2.1.1 Introduction

2-amino-benzophenon is an important starting material of the synthesis of various heterocycles, such as cinnolines, acridones, indazoles, indoles, quinolines [98]. Recently the synthesis of 2-amino benzophenones by palladium-catalyzed cross-dehydrogenative coupling between anilides and aromatic aldehydes has been developed applying by different methodologies. The methods developed by Zhou [99] and Kwong [100] could provide the desired products at 120 and 90 ° C respectively. These procedures applied toluene as solvent and required palladium(II) trifluoroacetate ($Pd(TFA)_2$) catalyst.

Recently Yu and co-workers [98] applied pivalanilide in a similar reaction scheme. Their procedure requires milder conditions (40 °C) and a more economical catalyst combination. The C–H activation was carried out by palladium(II) acetate ($Pd(OAc)_2$) catalyst in combination with trifluoroacetic acid (TFA). They have also proposed a reaction mechanism based on their experimental findings, where the carbopalladation is the rate determining step. This assumption is also validated by their kinetic isotope effect measurements. According to their proposal the carbopalladation step is followed by the oxidation of the metallic center of the intermediate by the attachment of the benzoyl radical. Finally reductive elimination takes place. Their mechanistic proposal is based on the assumption of the in situ generation of $Pd(TFA)_2$ catalyst.
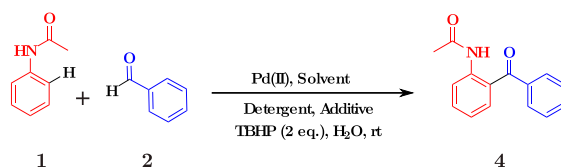


**Figure 2.1:** Procedure for oxidative coupling of acetanilide and benzaldehyde through C–H activation.

Novak and co-workers [1] (our experimental partners) have found that the reaction can be performed in water with $Pd(OAc)_2$ catalyst and with various acids at room temperature. In order to increase the reaction efficiency they have applied detergents (Fig. 2.1). In spite of the experimental findings by Yu and co-workers [98], important details of the mechanism remained unravelled. The formation of a more active catalyst $Pd(TFA)_2$ could not be convincingly supported, as non-coordinating acids, such as sulphuric or perchloric acid provided the same conversion as TFA. As the carbopalladation is the rate-determining step, the study of the subsequent steps is difficult by experimental techniques.

On the other hand for analogous C-H activation reactions important observations could be done regarding the steps after carbopalladation. For example Dong and co-workers isolated and characterised a bimetallic Pd-complex from a coupling reaction of aryl esters and aldehydes [101]. Bimetallic palladium complex formation from 2-phenylpyridine [102] and benzo[*h*]quinoline [103] was also reported by the group

of Ritte. We also intended to study the role of monometallic or bimetallic species in the C–C coupling reaction between anilides and aromatic aldehydes. We aimed to determine the evolution of the oxidation state of the palladium center during the reaction. In particular, we have focused on the possible sequences of two important steps, the reductive elimination and the single electron transfer from the palladium to the $^{t}$BuO· radical. Depending on this squence the reductive elimination can feature a Pd(IV) → Pd(II) or a Pd(III) → Pd(I) change [1].

### 2.1.2 Methodology

Our mechanistic study have been carried out by employing static calculations on the M06 [92] potenetial energy surface. The Kohn-Sham orbitals were expanded in the 6-31+G* [104] basis set. For palladium the LANL2DZ [105] basis set was applied with additional diffuse and polarization functions. The effect of the core electrons and the relativistic effects were modelled through the LANL2 [105] effective core potential. The interaction with the solvent have been taken into account by employing the SMD [106] solvent model.

### 2.1.3 Mechanistic study

Based on the experimental evidences we have examined the mechanism of the reaction by assuming that the first step is the C–H activation. The corresponding structures and the free energy profile for this transformation is depicted on Fig. 2.2. The association of the Pd(OAc)$_2$ and the acetanilide proved to be slightly endergonic. The association features 2.9 kcal/mol free energy change. The C–H activation step from **1.1** to **3.1** is only slightly endergonic and includes a 17.2 kcal/mol barrier (blue profile on Fig. 2.2). The effect of strong acids can be rationalized by noticing that the dissociation of one of the acetate ligands from Pd(II) acetate ion is enhanced by its protonation (green profile). This step results in remarkable free energy release (-22.2 kcal/mol). The proton abstraction by the acetate ligand proceeds through a slightly more favourable elementary step in comparison with the two-acetate-ligand analogue. This step involves only 15.8 kcal/mol activation free energy. It is important to note that concerted metalation-deprotonation is a frequently applied concept for the rationalization of C–H activation of arenes. This concept was applied for example by Macgregor [107] and Fagnou [108, 109] for the description of carbopalladation mechanisms in similar systems.

On the basis of our mechanism the effect of the acid is the following. By the spontaneous protonation of one of the acetate ligands the acidic media shifts the equilibrium in the direction of the active species (palladium(II)-acetanilide-acetate complex) formation. In this scheme a further advantage is that the palladium complex with a single acetate ligand undergoes intermolecular proton abstraction slightly easier.
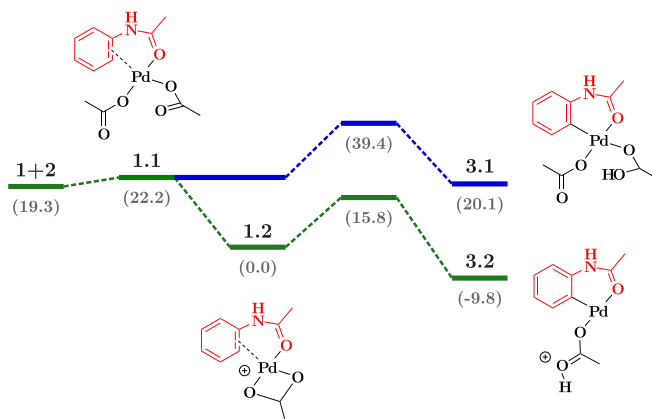
**Figure 2.2:** Gibbs free energy profiles for the two different C–H activation mechanisms. Values in parentheses are in kcal/mol. Green: one acetate ligand. Blue: two acetate ligands.

The C–C coupling of **1** and **2** was first investigated with a monometallic model system. Fig. 2.3 (green profile) summarizes the results in terms of free energy changes and a schematic mechanism. The transformation from **3.2** to **3.3** involves the deprotonation of the complex. This step is unfavorable (15.7 kcal/mol). It is important to note that the present continuum solvation model might overestimate this energy change. The hydrogen abstraction from the benzaldehyde molecule by $^{t}$BuOOH:

$$^{t}BuOOH + HCOPh \longrightarrow {}^{t}BuO \cdot + \cdot COPh + H_2O \qquad (2.1)$$
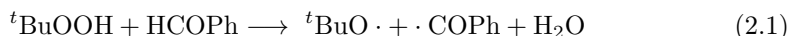
requires an additional 4.4 kcal/mol (**3.3** → **3.4**) investment. Association of the benzoyl radical and complex **3.3** leads to intermediate **3.5**. The barrier for the reductive elimination step is only 1.7 kcal/mol, which is followed by an instantaneous free energy release of -10.4 kcal/mol. Additional -42.8 kcal/mol is released during the closing steps. These involve the redox reaction between the Pd(I) species and the $^{t}$BuO· radical as well as the protonation of the resulting $^{t}$BuO$^{-}$ anion.
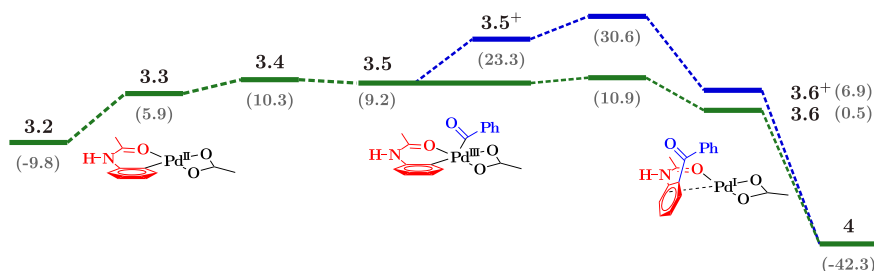


**Figure 2.3:** Gibbs free energy profiles for the mono-metallic mechanism of the C–C coupling reaction. Energies are in kcal/mol. Green: Pd(III) to Pd(I) pathway. Blue: Pd(IV) to Pd(II) pathway.

As the reductive elimination in palladium catalysed reactions often proceeds through IV to II states, we have investigated the oxidation of complex **3.5** (Fig. 2.3,

blue diagram). The oxidation from **3.5** to **3.5$^+$** by the $^t$BuO· radical is exergonic (-14.1 kcal/mol). The reductive elimination step requires an additional 7.3 kcal/mol activation. The overall barrier with this route is 40.4 kcal/mol. Therefore the monometallic C–C coupling pathway is more favourable along the III to I reductive elimination pathway than along the IV to II route. It is important to emphasize that the overall activation barrier for the C-C coupling step exceeds 20 kcal/mol for the more favourable pathway as well. These activation free energy values are significantly larger than 15.8 kcal/mol for the C–H activation. Therefore this pathway is not compatible with the experimental proposal, i.e. the carbopalladation is not the rate-determining step.
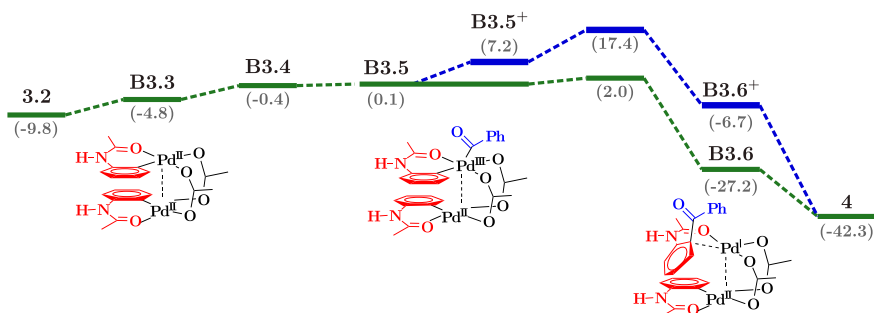


**Figure 2.4:** Gibbs free energy profiles for the bi-metallic mechanism of the C–C coupling reaction. Green: Pd(III) to Pd(I) pathway. Blue: Pd(IV) to Pd(II) pathway.

As the monometallic C–C coupling is incompatible with the experimental findings we have sought other possibilities. We have assumed the deprotonation and dimerization of **3.2** to **B3.3** (See Fig. 2.4, green profile). This transformation is spontaneous (-10.7 kcal/mol). Going from **B3.3** to **B3.4** requires the same energy investment for the H abstraction from benzaldehyde as in the monometallic route. The association of the bimetallic complex and the benzoyl radical requires only negligible activation. Likewise the reductive elimination is an almost spontaneous reaction with 1.0 kcal/mol free energy barrier. The Gibbs free energy balance of this step is also remarkable, -29.2 kcal/mol. The free energy level of the products is reached via analogous steps to those in the monometallic route. The overall activation free energy for the C–C coupling is 11.8 kcal/mol. Assuming this path as the mechanism for the C–C coupling, the C–H activation is the rate-determining step.

Although the reductive elimination proved to be remarkably efficient for the bimetallic pathway, we have also investigated the possibility of an early oxidation. This pathway is depicted in blue on Fig. 2.4. The effect of oxidation on this path is very similar to the effect on the mono-metallic path. The redox process is also endergonic and the reductive elminiation requires again larger activation energy in comparison with the Pd(III)→ Pd(I) path. The overall barrier for this pathway is 27.2 kcal/mol, which is again incompatible with the experiments.

In summary we have given a mechanism for the palladium catalysed C–H activation of acetanilide and its coupling with benzaldehyde in the presence of $^t$BuOOH. Our mechanism features the rate-determining C–H activation step, which is catalysed by a palladium complex with one acetate ligand. According to our proposal the C–H activation is followed by a C–C coupling via a bimetallic complex and a Pd(III)$\rightarrow$ Pd(I) reductive elimination.

## 2.2  Silver-mediated furan formation [2]

### 2.2.1  Introduction

Methods for selective synthesis of furan and pyridine rings from simple building blocks are important transformations. These rings are important intermediates of organic syntheses and they appear in numerous natural products [110, 111]. The synthesis of various substituted furans has been reported in a recent work of Lei at al [112]. Their synthetic strategy was based on an oxidative coupling of $\beta$-ketoesters with terminal-acetylenes via a silver mediated transformation (Figure 2.5). The same procedure led to pyrrol derivatives, if enamines were used instead of $\beta$-ketoesters. The final protocol for the reaction has been obtained by systematic variation of the applied Ag sources, bases and solvents. Remarkably, the reaction did not provide the expected product if toluene or THF solvents were applied. In contrast, the yield varied from 40 to 100% if polar solvents such as DMF or DMSO were chosen as media. The combination of $Ag_2CO_3$ and carbonate bases were moderately successful, and AgOAc in combination with KOAc bases did not yield any product. The application of silver-carbonate with silver-acetate proved to be very efficient with full conversion.
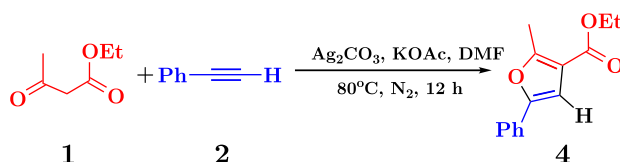


**Figure 2.5:** Silver-mediated oxidative C-H/C-H functionalization of **1** and **2**.

The authors carried out mechanistic investigations as well. They have postulated silver phenylacetylide as an intermediate. Their concept is also supported by the production of the same product in the reaction of silver phenylacetylide and ethyl acetoacetate. They also postulated coupling between the silver acetylide salt and the $\alpha$ carbon of the $\beta$-keto ester. However this idea has not been justified experimentally. The assumption that silver phenylacetylide formation from the reaction of $Ag_2CO_3$ and phenylacetylene in polar solvents can occur, has been supported by vibrational spectroscopy results. In two subsequent works of Lei and his co-workers silver-phenylacetylide has been detected by Raman spectroscopy [113] and in situ IR [114] spectroscopy. In both studies DMSO solvent has been employed.

The aim of our study was to explore several possible pathways for this interesting silver-mediated furan formation. On the basis of the corresponding free energy profiles we have selected the most likely routes. The mechanisms emerged from these pathways have been supported by additional experiments [2].

## 2.2.2 Methodology

For this study we have applied the M06 functional and the SMD solvent model. The geometry optimizations and frequency calculations employed a double zeta basis set: 6-31+G* [104] for all main group elements, and the LANL2DZ for silver. The effect of core electrons was modelled by the LANL2 effective core potential [105]. The solvation free energy contribution has been taken into account by single point SMD [106] calculations employing triple zeta basis set. For the main group elements 6-311++G(3df,3pd) [115] basis set were employed, while the LANL2TZ basis [116] with additional set of polarization and diffuse functions was used for silver. The additional polarization and diffuse functions have been taken from the aug-ccpVTZ-PP basis set.

## 2.2.3 Mechanistic study

We have selected the reaction of phenylacetylene and ethyl-acetoacetate for the mechanistic study, which has been used in the development of the synthetic procedure of Lei and co-workers [112]. The investigation of the reaction is particularly challenging due to the solid reactants (KOAc and $Ag_2CO_3$) and solid product (elementary silver). As the precise value for the solvation free energy change cannot be calculated within the applied theoretical framework, we have assumed that a certain fraction of the base and the silver source is already in the solvent phase. [2] In this approach we have an endergonic process with a free energy change which would only shift each Gibbs free energy diagrams with the same, constant value. The free energy change of silver precipitation has been calculated from the experimental values of atomisation enthalpy and entropy of silver [117], and the calculated solvation free energy for the solvation of a single silver atom.
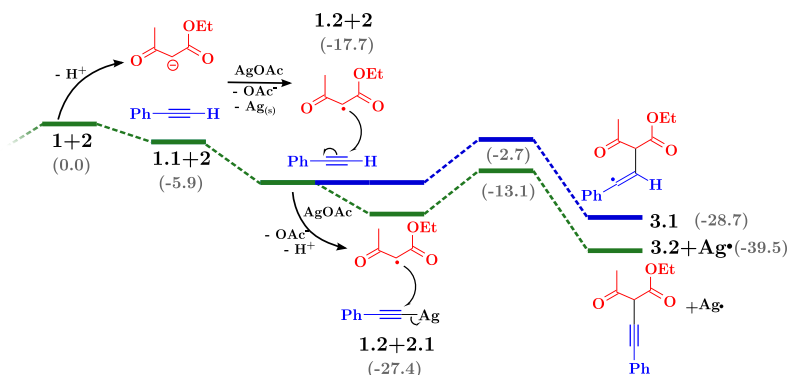


**Figure 2.6:** Gibbs free energy profile and mechanism for the C-C coupling part. Energies are in kcal/mol. Green: pathway featuring **2.1**. Blue: pathway featuring **2**.

The initial step of our mechanism is the deprotonation of the ethyl acetoacetate at the $\alpha$ carbon atom. This is followed by the single electron transfer oxidation of the anion to a radical by the silver cation. These two steps are exergonic and

release in total -17.7 kcal/mol Gibbs free energy. The radical mechanism of the reaction has been verified by our experimental partners. The application of radical scavengers such as TEMPO or BHT could inhibit the product formation. We have investigated two possibilities for the C–C coupling reaction mechanism. The first is the direct coupling of the phenylacetylene and the radical, while the second possibility features the formation of silver acetylide and its coupling with the radical. The two mechanisms and the corresponding free energy diagrams have been presented on Fig. 2.6. The pathways feature similar barriers (15.0 and 14.3 kcal/mol) and similar reaction Gibbs free energy changes (-11.0 kcal/mol and -12.1 kcal/mol). In contrast, the two reactions result in chemically significantly different species. In the first case the reaction yields an organic radical intermediate, while the silver mediated path provides a closed-shell intermediate and a silver atom. In the followings the latter reaction is described.
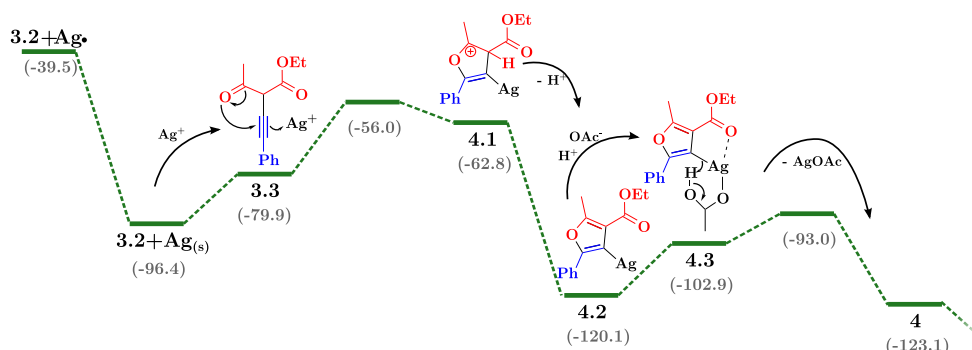


**Figure 2.7:** Gibbs free energy profile and mechanism for path **A**. Energies are in kcal/mol.

Fig. 2.7 summarizes the reaction path following an ionic cyclisation mechanism. We have assumed that the atomic silver precipitates from the reaction mixture after the C–C coupling step. It is important to note that the formation of silver mirror on the reaction chamber is indeed observed in the experiments. This path also features the silver catalysis of the cyclisation reaction. The silver ion has catalytic effect by polarizing the triple-bond and enhancing the nucleophilic attack on it. Despite the silver catalysis the path involves prohibitively high barrier for the cyclisation (40.4 kcal/mol). The instantaneous stabilization of the system after the high barrier is only 6.8 kcal/mol. However surprisingly large amount of free energy (-57.3 kcal/mol) is released during the deprotonation of intermediate **4.1**. This large energy suggests that the deprotonation step should be considered during an earlier phase of the mechanism in order to find more favourable pathways. The deprotonation reaction results in an interesting silver-organic intermediate (**4.2**) with considerable stability. As the experimental procedure includes acidic work up, one might assume that the reaction stops at this stage, and the silver organic compound is decomposed by the reaction with the HCl-solution. Our experimental partners therefore carried out experiments to test this hypothesis with isotope labelling. To this end the reaction mixture has been exposed to $D_2O/DCl$ during the workup. The deutered product has not been

identified irrespective of **2** or **2.1** have been chosen as reactant for the reaction. These experiments also support the presence of **2.1** as an important intermediate of the reaction.

The exchange of the silver ion with a proton is slightly favoured under the conditions (by -3.0 kcal/mol). Nevertheless this step requires a barrier that is rather high for a proton exchange reaction (27.1 kcal/mol).

As pathway **A** suggested the importance of the early deprotonation reaction, we have focused on the effect of an earlier proton abstraction on the cyclisation reaction. The corresponding free energy profile and the structure of the relevant intermediates are depicted on Fig. 2.8. The deprotonation of **3.2** leads to the enolate ion **3.4** with an exergonicity of -19.0 kcal/mol. The enolate moiety is a stronger nucleophile comparing to the original ketone-group. This is also reflected in the reduced barrier height (29.5 kcal/mol) for the cyclisation reaction. This amount of activation free energy is compatible with the temperature of the reaction.
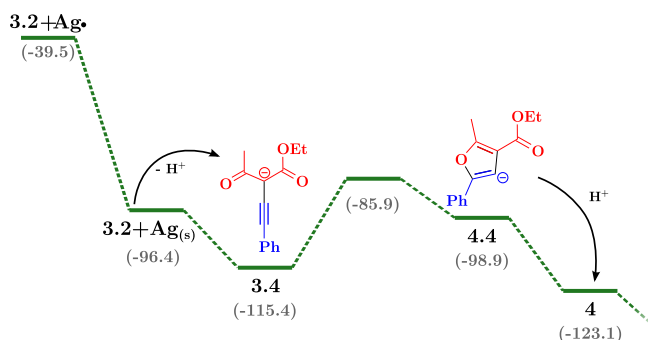


**Figure 2.8:** Gibbs free energy profile and mechanism for path **B**. Energies are in kcal/mol.

In path **C** (Fig. 2.9) the synergistic effect of $Ag^+$ and the early stage deprotonation are studied. This idea is based on the assumption that the entropy loss of the association of the organic species and the silver cation is compensated to a larger extent by the Coulomb interaction in this case. As expected, the overall barrier from state **3.4** is modest (21.4 kcal/mol), especially at the temperature of the reaction. The cyclisation reaction results in a silver-furanyl compound. The last section of this path is identical to the closing steps of pathway **A**. Interestingly in this case the barrier of the proton exchange reaction is higher than the barrier of cyclisation reaction.
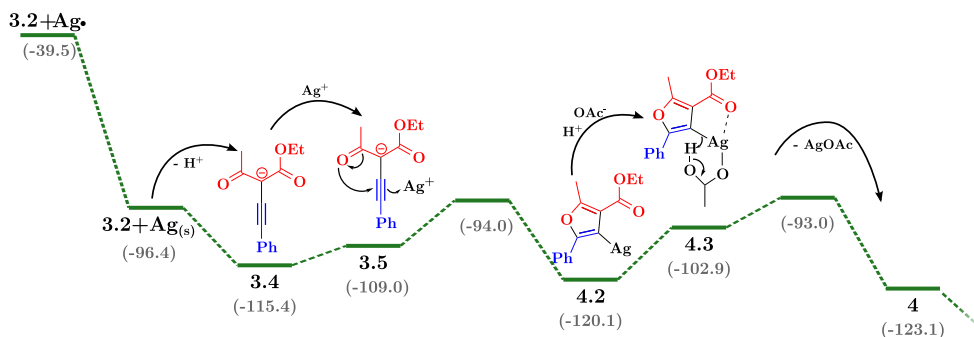
**Figure 2.9:** Gibbs free energy profile and mechanism for path **C**. Energies are in kcal/mol.

Another interesting possibility can be derived from the assumption that the cyclisation reaction precedes the silver precipitation (route **D**, Fig. 2.10). In this case the reaction might proceed through a radical pathway without the reduction of the organic moiety. However the direct cyclisation step in this case features a barrier of 33.8 kcal/mol. Therefore this pathway is inferior to the previous ionic pathways. As a summary of the presented mechanisms we can conclude that path **C** is our best mechanism if the deprotonation of the alkyne is assumed.

Although the reaction typically applied for terminal alkynes, some examples for internal alkynes have been presented in the supportin information of the original synthetic work [112]. The desired product has been identified in only traces for diphenylacetylene and in only 10% for the methyl- and phenyl-substituated analogues. As our previous pathways have not involved the reaction of these substrates, we have followed the reaction starting from the radical intermediate **3.1** as well.
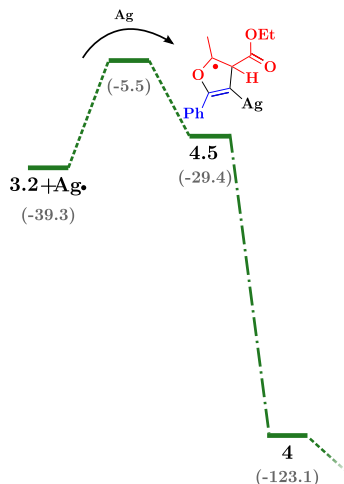


**Figure 2.10:** Gibbs free energy profile and mechanism for path **D**. Energies are in kcal/mol. The dash-dot line indicates that intermediates between **4.5** and **4** along the segment have not been identified.

On the basis of our previous experience about the cyclisation reaction we have assumed that the intermediate is promptly deprotonated and that the resulting radical ion associates with a silver ion. The corresponding pathway and its free energy profile are depicted on Fig. 2.11. The deprotonation reaction requires 18.2 kcal/mol activation, while the association reaction is slightly favoured in terms of free energy (-3.0 kcal/mol). The cyclisation reaction requires only 16.3 kcal/mol of activation and provides the desired product instantly.
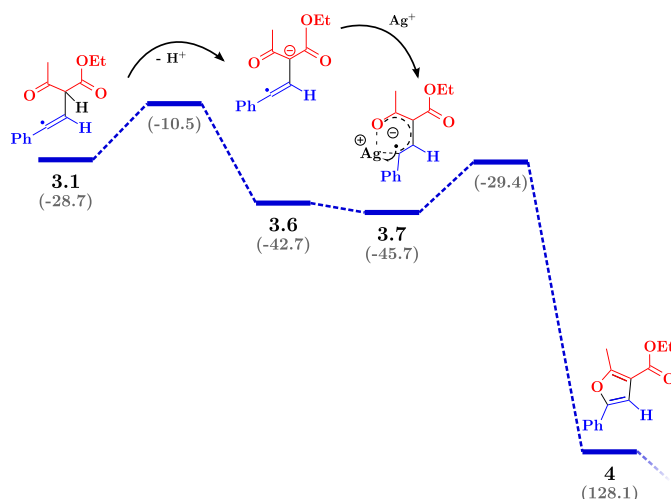


**Figure 2.11:** Gibbs free energy profile and mechanism for path **E**. Energies are in kcal/mol.

In order to test our hypothesis on this alternative mechanism we have calculated the first barrier of pathway **E** for the methyl and phenyl substituted compounds as well (Fig. 2.12). The results are in agreement with the experimental order of reactivities. It is important to note that the height of the barrier alone does not explain the reduced reactivity. However the higher activation barrier for the reaction of the unstable radical compound **1.2** might open the possibility for side reactions.
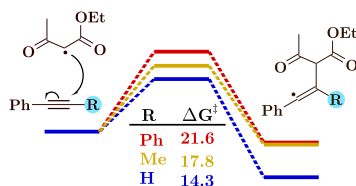


**Figure 2.12:** Gibbs free energy profiles for the C-C coupling reaction for terminal and two different internal alkynes. Energies are in kcal/mol.

In summary we have identified the relevant reaction pathways for the silver mediated-furan synthesis reaction of Lei and co-workers [112]. Our mechanistic proposal for furan formation from terminal acetylenes and ethyl acetoacetate features a C–C coupling between silver acetylide and the radical intermediate from ethyl

acetoacetate. The cyclisation reaction is assumed to be silver ion assisted ionic process. For internal acetylides we have proposed the direct bond formation between the ethyl-acetoacetate radical and the acetylene along with a cyclisation reaction with atomic silver formation.

## 2.3 Direct 2,2,2-trifluoroethylation of indoles via C–H functionalization [3]

### 2.3.1 Introduction

The synthetic procedures for the installation of trifluormethyl group onto aromatic and heteroaromatic systems have remarkable significance. These approaches are uncommon, but yield products with high pharmacological significance. Trifluoroethyl groups can increase the lypophilicity and the metabolic stability of the molecules. Our experimental partners have successfully developed a procedure for selective C3 modification of indole derivatives. The transformation introduces a 2,2,2-trifluoroethyl group to the C3 position through C-H activation. The procedure is depicted on Fig. 2.13. It is important to note that the reaction requires the presence of 2,6-di-tert-butylpyridine (DTBPY) in order to achieve maximum yield.
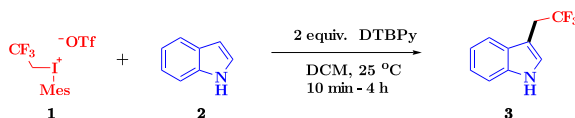


**Figure 2.13:** The reaction for trifluorethylation of indoles in our mechanistic study.

### 2.3.2 Methodology

In order to understand this valuable synthetic transformation we have performed mechanistic calculations. Our theoretical approach used the $\omega$B97XD functional [95] in combination with the 6-31+G* [104] basis set for geometry optimization and frequency calculations while the 6-311++G(3df,3pd) basis set [115] for single point SMD calculations [106] (with $CH_2Cl_2$ solvent). For iodine atoms we have applied LANL2 effective core potentials. LANL2DZ [105] and LANL2TZ basis sets [116] with additional diffuse and polarization functions from aug-cc-pVDZ-PP and aug-cc-pVTZ-PP bases sets respectively, have been used as basis sets for iodine.

### 2.3.3 Mechanistic study

We have started our study with the calculation of the free energy change for the dissociation of the iodonium salt. Our calculations have shown that the product state is slightly more favoured (by 0.9 kcal/mol). We therefore consider the dissociated form as the resting state. The direct electrophilic attack of the iodonium salt to the indole features a 25 kcal/mol high barrier. The step is highly exergonic (by 34.2 kcal/mol) therefore irreversible. The deprotonation of the sigma complex by DTBPy features 18.3 kcal/mol activation free energy. The large activation free energy can be rationalized by the steric hindrance of the base.
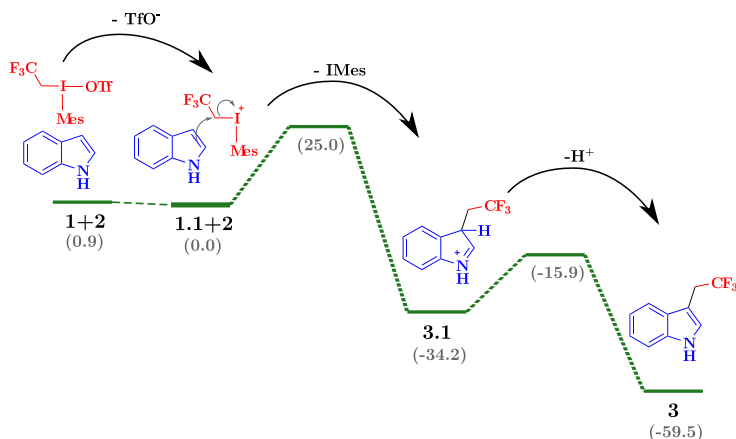
**Figure 2.14:** Free energy profile and mechanism for the trifluoroethylation reaction. Energies are in kcal/mol.

In the absence of DTBPy the reaction yielded a side product. Since the separation of the side product was not successful, DTBPy is essential for a successful reaction protocol. We aimed to understand the side reaction as well. Understanding this process is necessary in order to rationalize the further developments of this reaction. Fig. 2.15 summarizes our proposed mechanism for the side reaction and the corresponding Gibbs free energy profile. We have started our mechanistic considerations by analysing the role of the base in the reaction. Earlier we have postulated that the role of the DTBPy is to deprotonate **3.1.** Therefore in the absence of the base the proton will be abstracted by some other species. As the triflate ions are not appropriate candidates for proton abstraction, the only possibility to be protonated is the indole molecule. This assumption seems reasonable as the protonation of indole known from experiments [118] and from theoretical calculations as well [119]. The proton exchange between **3.1** and the indole molecule proved to be exergonic (6 kcal/mol). We have assumed that the reaction proceeds via association of the protonated and the neutral indole molecules. This reaction is endergonic (10.0 kcal/mol) and requires an activation free energy of 16.1 kcal/mol. This protonated species can exchange proton with another indole molecule in a slightly endergonic process. Intermediate **5.2** is a neutral secondary amine. Such molecules readily undergo N-alkylation reactions with iodonium salts. The corresponding barrier height for this transformation is 20.0 kcal/mol. This step provides the protonated side product. The overall free energy barrier for this process is remarkably large (30.3 kcal/mol). However the deprotonation and activation (according to the mechanism described in Fig. 2.14 ) of protonated indole requires slightly larger overall activation free energy (31.2 kcal/mol). Therefore the side reaction is slightly more favourable.
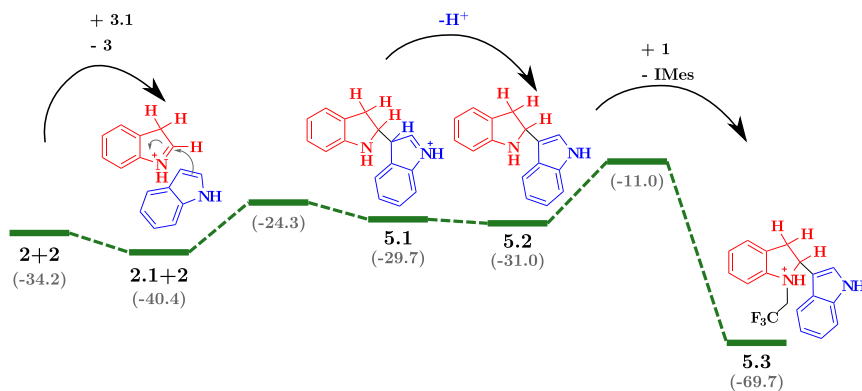
**Figure 2.15:** Free energy profile and mechanism for the dimerization side reaction. Energies are in kcal/mol.

The qualitative picture presented above suggests that the side reaction can be supressed by employing moderate strength bases. However the application of organic bases might result in the trifluorethylation of these substances. Figure 2.16 summarizes the conversion of indole to the desired product in the presence of various organic amines and the Gibbs free energy profile for the trifluorethylation side reaction of the bases. The correlation between the activation free energies for the N-alkylation reaction and the conversions to the desired product is obvious. If the activation Gibbs free energy for this side reaction is smaller than the activation necessary for the indole functionalization reaction, only traces of the product is detected. If the activation free energy barrier for the side reaction is in the range of the barrier for the desired reaction the useful conversion remains small to moderate. The DTBPy proved to be the only base in our scope which has significantly higher barrier against alkylation reaction than that of the indole molecule.
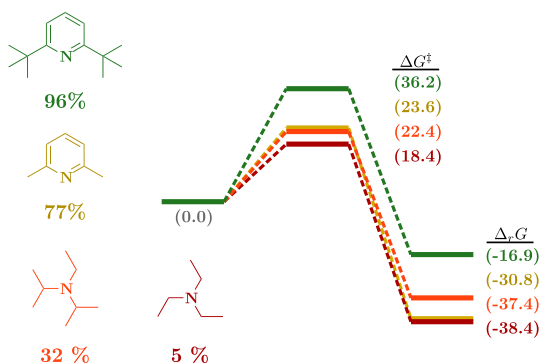


**Figure 2.16:** Conversion of the trifluorethylation of indole in the presence of amines in comparison with the Gibbs free energy profiles of the trifluorethylation of the amines. Values in parentheses are in kcal/mol.

In summary, our research on the mechanism of C3 trifluorethylation of indoles examined the target reaction and two experimentally observed side reactions. The desired product is formed through a simple electrophyl substitution reaction on

the indole ring and the deprotonation of the $\sigma$-complex. The followed side reactions are related to with the protonation of indole in the absence of suitable bases and the trifluorethylation of the bases. Our calculations on the latter reaction also revealed the role of the sterical hinderence of DTBPy.

# 2.4  Frustrated Lewis-Pairs [4]

## 2.4.1  Introduction

Frustrated Lewis-Pair (FLP) systems are based on the concept of cooperative bifunctional catalysis. FLP systems consist of a Lewis aceptor (A) and Lewis donor (D). The direct reaction (quenching) of the donor and acceptor molecules or fragments (in the case of intramolecular FLPs) is unfavoured. The quenching is prevented by the high degree of steric hindrance or the high ring strain in the product state. FLPs can facilitate the activation of otherwise inert reagents, even molecular hydrogen. The activated hydrogen molecule can be utilized in various reductive processes that can yield valuable organic substances. The catalytic hydrogenation of imines [120–122], enamines [121], enols ethers [123], N-heterocycles [124], alkenes [125], alkynes [126] and carbonyl compounds [127] have been successfully accomplished by the utilisation of different FLP catalysts.

The mechanism of the hydrogen activation has been described and rationalized by static DFT calculations [128]. In general, the process features a pre-equilibrium between the active and inactive form of the FLP. The active form of the FLPs is able to cleave heterolytically molecular hydrogen by stabilizing the proton and the hydride by the Lewis base and acid sites, respectively. According to the static calculations, the hydrogen dissociation proceeds in a single elementary step [128]. Interpretation of the hydrogen activation step is based on molecular orbital approach [129]. Ideal orbital interaction between the bonding orbital of the hydrogen and the empty orbital of the Lewis acceptor requires a side-on orientation. In contrast an end-on orientation is necessary to ensure sufficient overlap between the lone pair of the Lewis donor and the antibonding orbital of the hydrogen molecule. The calculated structural parameters for the TS-s of $H_2$ activation by FLPs demonstrated the validity of this picture. In addition, the cooperative nature of the $H_2$ activation has been revealed and discussed [129, 130].

Recently Pu and Privalov have published theoretical studies which combine PES exploration with ab initio molecular dynamics simulations [131–135]. Their results suggest the existence of a quasi-stable bounded state in hydrogen activation reactions [135]. This assumption implies that hydrogen activation is a two-step process and that the static calculations for these systems provide qualitatively false mechanistic picture. However, identification of intermediate states in finite temperature picture requires free energy calculations or the estimations of the lifetimes from canonical molecular dynamics simulations. Without such calculations the relevance of the assumed intermediate state can not be convincingly shown.

In order to asses the validity of the single-step hydrogen activation model in finite temperature, we have performed free energy calculations in combination with ab initio MD. This assessment also provides an opportunity to compare the static picture emerged from earlier studies with this new dynamical mechanistic view and discuss the validity of the approximation used in static approaches. We have studied intramolecular

FLPs, where the dissociation of the Lewis donor and acceptor counterparts is prevented by the linking group.

### 2.4.2 Methodology

All simulations have been performed employing the CP2K code [136] and the BLYP-D3 [86,90,91] method. The effect of core electrons has been described by the Goedecker-Teter-Hutter [137] pseudopotentials, while the valence orbitals have been extended in a double-$\zeta$ basis set. An auxiliary plane-wave basis set was applied for the electron density with a kinetic energy cut-off of 300 Ry. The molecular dynamics simulations were carried using a time-step of 0.5 fs and a global Nosé-Hover thermostat-chain. The dynamics of the system has been characterized by two CVs: a coordination number for the description of the bonding between the hydrogen atoms and the relevant atoms of the linked FLPs ($CV_1$), and a coordination number between the atoms of the hydrogen molecule.

The free energy calculations have been done with metadynamics within its extended Lagrangian formalism [39]. The width of the Gaussian functions has been chosen for 0.1 in both dimensions, while the height of the hills was 0.5 kcal/mol. The minimum free energy path on the free energy surface was localised by applying the zero temperature free energy string method. The images of the optimized pathway have been used to construct a reaction path-CV which has been used as reaction coordinate ($\lambda$) [13]. The chosen CVs have been tested through committor analysis. The ensemble of shooting points has been sampled through two dimensional umbrella potential. The reaction coordinate was used to define the stable reactant and product regions, and to analyse the committor trajectories. Configurations from the reactant and product states and from the saddle region of the free energy surface were aligned with an algorithm based on their minimal RMSD (Eq. 1.17) and visualized for each system (*vide infra*).

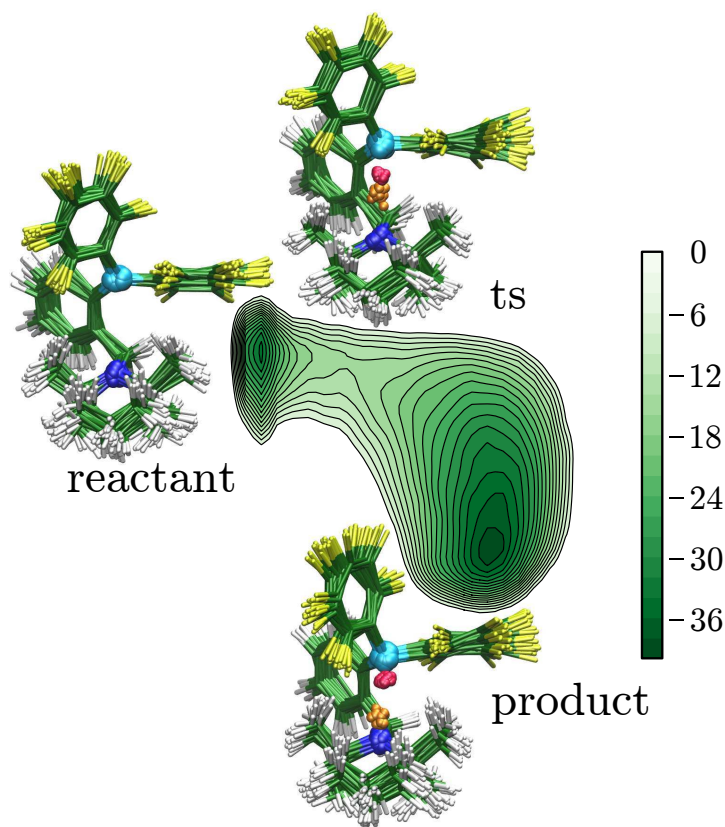### 2.4.3 Nitrogen/boron "molecular tweezer" by Repo and Rieger et al. [122]



**Figure 2.17:** The free energy surface of the "molecular tweezer" system with respect to $CV_1$ (horizontal axis) and $CV_2$ (vertical axis) and an ensemble of geometries for reactant, transition and product states. Colour of the atoms: N-blue, B-cyan, $H_B$-gold, $H_N$-red, C-green, F-yellow, $H_{FLP}$-white. Energy contours are in kcal/mol. $H_2$ molecules in the reactant state are not shown.

The free energy surface of the studied molecular tweezer system is clearly characterized by two well defined minima and the connecting transition state region. Apart from the displacements of the reacting four atoms, the ensembles of configurations in the reactant, transition and product states show only small variations, which can be associated with equilibrium vibrations and internal rotation of methyl groups. This indicates that the contributions of internal rotors to the free energy along the reaction path cancel out to a large extent.

The position of the hydrogen atoms in the TS is surprisingly localised. Note, that on Fig. 2.17 the atoms of molecular hydrogen are omitted from the reactant state

ensemble for sake of simplicity. Comparison of the TS and product state ensembles clearly shows that the bonding to boron atom is almost complete at the TS, while the proton binds to the nitrogen in concert with the cleavage of the H–H bond after the TS.
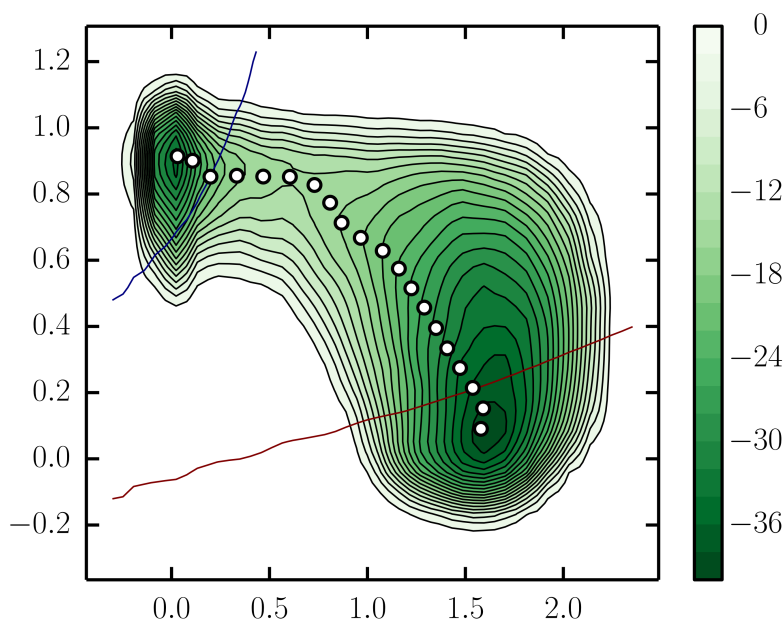


**Figure 2.18:** The minimum free energy pathway and the definition of stable reaction and product states. Energy contours are in kcal/mol.

The shape and the quite uniform slope of the minimum energy pathway indicates that a single CV is able to describe the reaction sufficiently. However, very fine details, such as the pronounced curvature at the TS, would not be possible to capture with a simple linear combination of our present CVs.
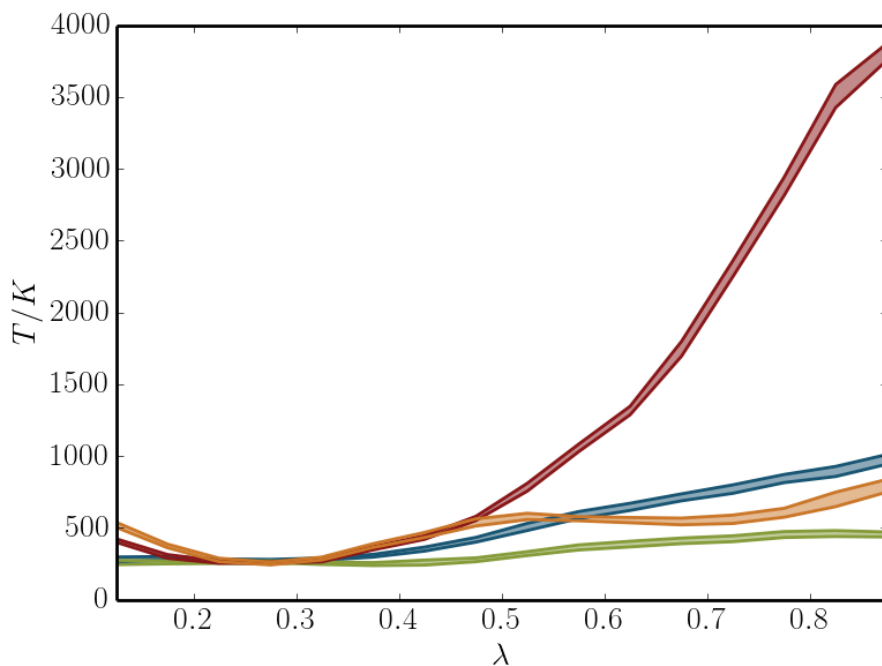
**Figure 2.19:** Temperature of B (blue), N (green) and activated hydrogen atoms $H_B$ (orange), $H_N$ (red). The shaded area corresponds to the 90% confidence interval calculated from the standard error of the mean. The temperature of other atoms remained constant at the 300 K set for the TS ensemble.

The instantaneous temperature of each atom has been calculated and averaged along the reaction path from the committor simulations. Since these trajectories are NVE trajectories, we can deduce the temperature distribution of a hypothetical reactant ensemble where all configurations have a temperature distribution sufficient to reach the TS. Fig. 2.19 indicates that the necessary thermal energy for the reaction is dominantly stored in the molecular hydrogen. Interestingly, the two atoms of the molecule contribute differently. Even larger differences can be observed at approaching the product state. The H-B bond formation is significantly more advanced then that of H-N at the TS. Therefore a large part of the H-N bond energy is released after the TS, which contributes to the significantly higher temperature of the H atom on the nitrogen.

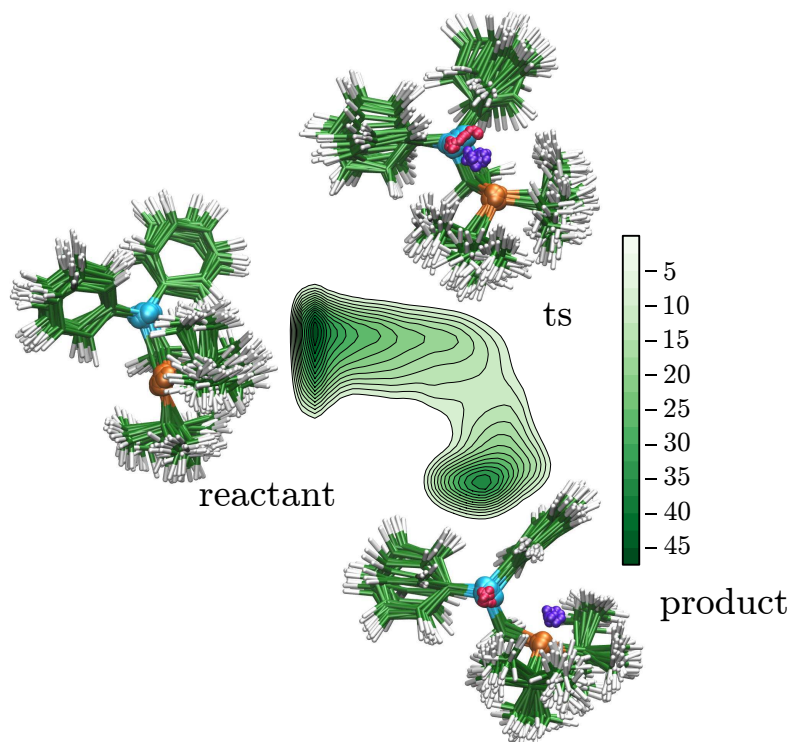### 2.4.4 The "preorganized" FLP of Slootweg, Lammertsma et al. [139]



**Figure 2.20:** The free energy surface of the "preorganized" FLP system with respect to $CV_1$ (horizontal axis) and $CV_2$ (vertical axis) and an ensemble of geometries for reactant, transition and product states. Colour of the atoms: P-gold, B-cyan, $H_B$-purple, $H_N$-red, C-green, $H_{FLP}$-white. Energy contours are in kcal/mol. The $H_2$ molecules in the reactant state are not shown.

This preorganized system is one of the rare FLP systems that activate hydrogen molecule in spite of the absence of fluorinated groups or other strongly electron withdrawing groups.

An interesting feature of the reaction path (Fig. 2.20) is the remarkable variation of the mobility of the phenyl rings. At the reactant state both rings fluctuate around the equilibrium arrangement. In contrast, at the TS the phenyl rings rotate around the B-C $\sigma$ bonds. At the product state, one of the phenyl rings can still rotate, while the other is restrained by the $^t$Bu-groups. The high mobility at the TS is counter-intuitive, as we expect a more pyramidal B arrangement by the sp$^2$ $\rightarrow$sp$^3$ hybridization change upon $H_2$ coordination. Such pyramidalization is expected to increase the steric congestion around the boron atom. However after filling the empty

orbital of the boron atom, no delocalization can occur between the phenyl groups
and the boron atom. This effectively reduces the torsion potential around the B-C $\sigma$
bonds. Although the HBCPH core is quite rigid, the orientation of the $^t$Bu groups
provides room for the rotation of one of the phenyl groups in the product state.

Another interesting feature of the $H_2$ activation of this FLP is that the orientation
of the simultaneously developing B–H and P–H bonds deviate from the usual planar
alignment. This may be attributed to the steric repulsion between the ligands of the
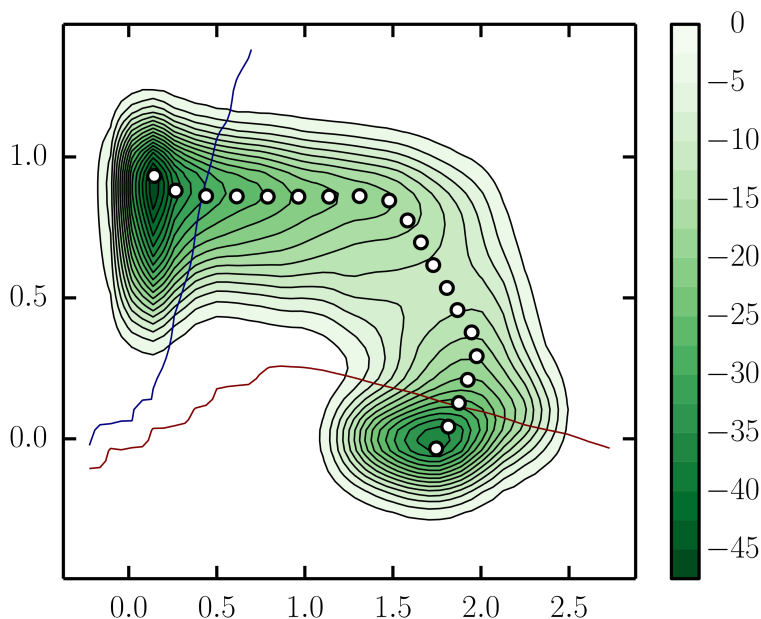Lewis centers.



**Figure 2.21:** The minimum free energy pathway and the definition of reaction and product
states. Energy contours are in kcal/mol.

The minimum free energy surface of the system is also interesting (See Fig. 2.21.
It features three almost perfectly straight segments. The first includes the approach
of the hydrogen molecule to the reactive center. This part is responsible for the major
part of the free energy barrier. The second segment shows a synchronous heterolytic
hydrogen cleavage and the attachment of the hydride and the proton to the boron
and the phosphorous, respectively. During the third segment the final hidride–proton
distance is reached. It is important to note that at variance with the previous system
the iso-lines defining the reactant and the product state are close to be orthogonal to
each other. This means that no linear combination of the two coordinates would be
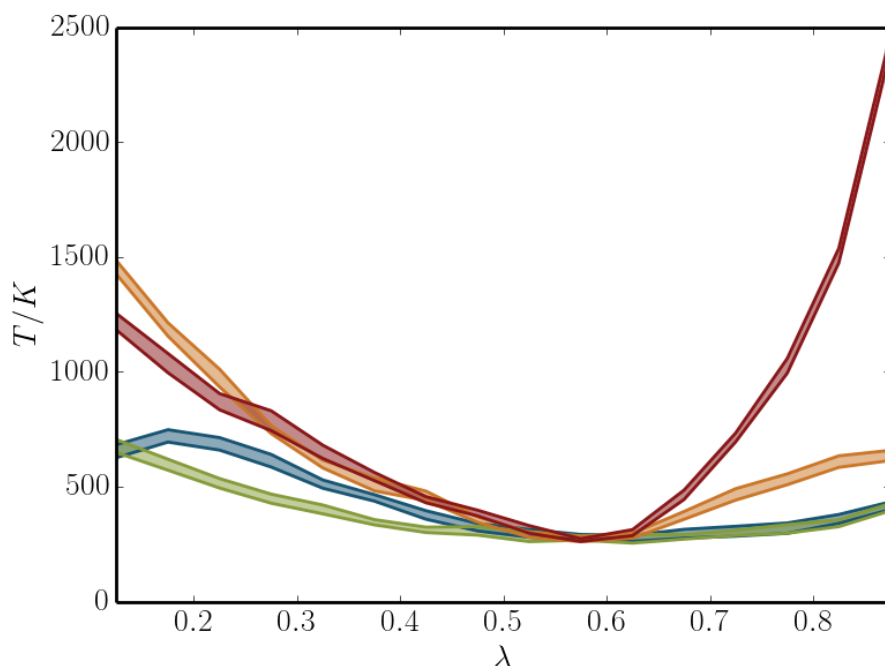sufficient for the description of the reactive event.

**Figure 2.22:** Temperature of B (blue), P(green) and activated hydrogen atoms H$_B$ (orange), H$_P$ (red). The shaded area corresponds to the 90% confidence interval calculated from the standard error of the mean. The temperature of other atoms remained constant (300 K).

The instantaneous atomic temperature shows a dramatically different picture from what we have seen for the molecular tweezer (See Fig. 2.22). The energy is more uniformly distributed among the atoms involved in the reaction. In the reaction side not only the hydrogen atoms have high temperature, but also the boron and phosphorus atoms have comparable temperature. This can be understood by assuming that the twist around the CH$_2$–P and CH$_2$–B bond requires excess velocity on the boron and phosphorous atoms. This torsional movement is crucial to form the strained transitional state structure with a five-member ring. In the product side the excess heat from the P–H bond formation again contributes to the temperature of the proton, while the B–H bond formation releases much smaller amount of heat.

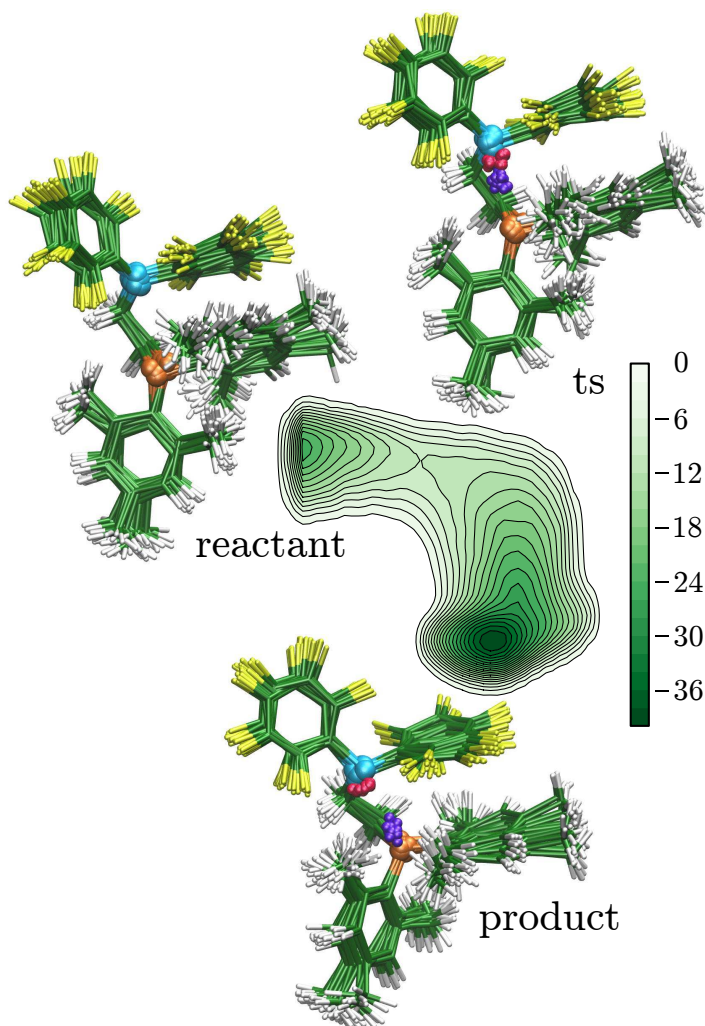## 2.4.5 Ethylene-linked phosphine–borane system by Erkel at al. [138]



**Figure 2.23:** The free energy surfaceof the ethylene-linked FLP system with respect to the $CV_1$ (horizontal axis) and $CV_2$ (vertical axis) and an ensemble of geometries for reactant, transition and product states. Colour of the atoms: Colour of the atoms: P-gold, B-cyan, $H_B$-purple, $H_N$-red, C-green, F-yellow, $H_{FLP}$-white. Energy contours are in kcal/mol. The $H_2$ molecules in the reactant state are not shown.

The free energy surface (Fig. 2.23) indicates that before reaching the TS, only minor H–H bond dissociation occurs. The ensembles of the FLP systems are fairly uniform, they show only oscillations around a well-defined structure in the reactant,

transition and product states. Again hydride coordination to the boron atom precedes the proton attachment to the phosphorous. The product zwitter-ionic species is stabilized by the Coulombic interaction between the FLP counterparts.
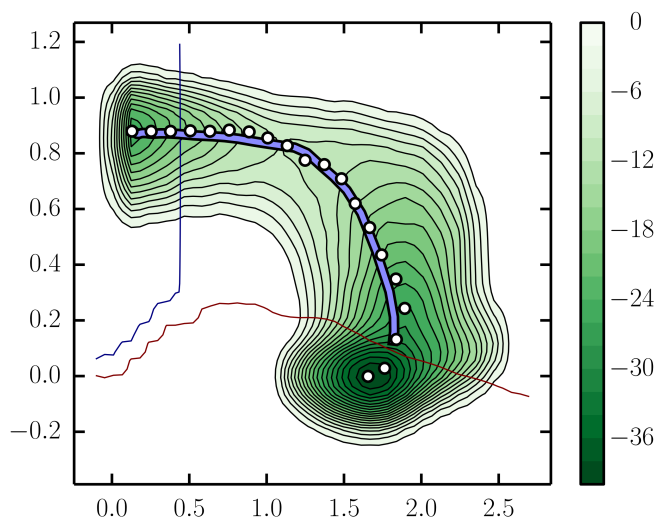


**Figure 2.24:** The minimum free energy pathway (white circles) and the projection of the intrinsic reaction coordinate (purple line). The definition of reaction and product states are indicated with thin blue and red lines, respectively. Energy contours are in kcal/mol.

Comparison of the minimum free energy path and the 0 K IRC curve provides additional insight into the reaction. To this end we have projected the IRC configurations into our two dimensional CV space. Fig. 2.24 shows that the two curves nicely coincide. This finding is important, as it suggests that the zero temperature description is qualitatively correct in terms of the important coordinates.

Although the minimum energy path is in good agreement with the minimum free energy path in these two CVs, this does not necessarily mean that structural parameters calculated along the IRC path are representative for the reaction path ensemble. We have therefore calculated the distribution of some structural parameters along the commitment trajectories and plotted their distribution with respect to the reaction coordinate (See Figs. 2.25–2.30).

Fig 2.25 presents the distribution of $BH_B$ distance along the reaction coordinate. The distribution is quite narrow and it is in excellent agreement with the IRC values. We can also see that the evolution of the B–H distance supports the early bond formation picture.

The $H_P P$ distance distribution on Fig. 2.26 is wider than the distribution of $BH_B$ distance. Still, the IRC gives satisfactory description for the progress of this variable

along the reaction coordinate. The asyncronous B–H vs. P–H bond formation can be also understood by comparing the slopes of the curves in Figs. 2.25 and 2.26.
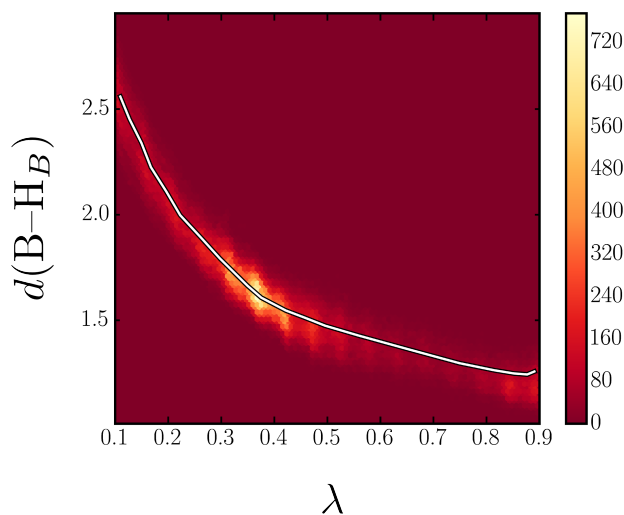


**Figure 2.25:** Distribution of the B–H$_B$ distance with respect to $\lambda$. The IRC is designated with white line.
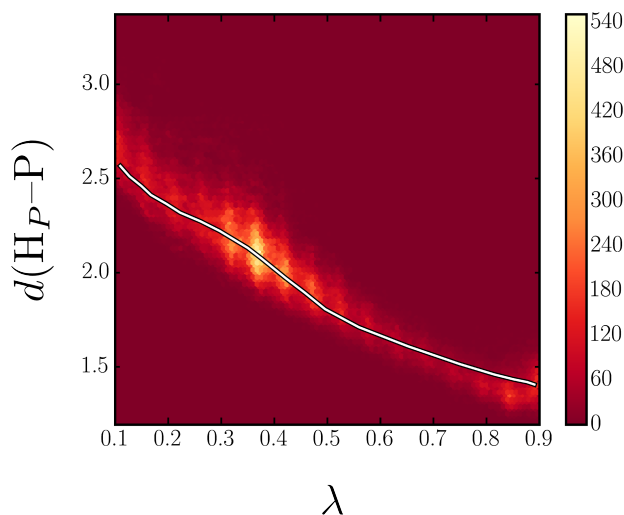


**Figure 2.26:** Distribution of the H$_P$–P distance with respect to $\lambda$. The IRC is designated with white line.

Fig. 2.25 presents the distribution of $BH_B$ distance along the reaction coordinate. The distribution is quite narrow and it is in excellent agreement with the IRC values. We can also see that the evolution of the B–H distance supports the early bond formation picture.

The $H_P P$ distance distribution on Fig. 2.26 is wider than the distribution of $BH_B$ distance. Still, the IRC gives satisfactory description for the progress of this variable along the reaction coordinate. The asyncronous B–H vs. P–H bond formation can be also understood by comparing the slopes of the curves in Figs. 2.25 and 2.26.
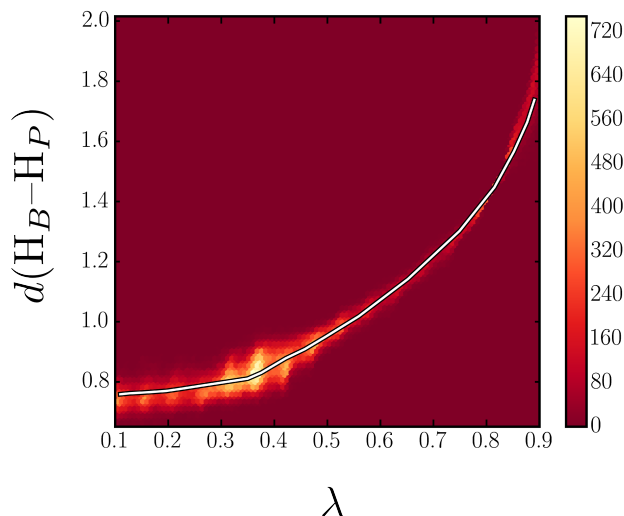


**Figure 2.27:** Distribution of the $H_B H_P$ distance with respect to the reaction coordinate. The IRC is designated with white line.

As the $H_B H_P$ distance is in direct connection with $CV_2$, its good correlation with the IRC curve is not surprising. The distribution of the reactive trajectory snapshots is also presented on Fig. 2.27 in order to demonstrate the power of IRC curve in the qualitative description. The distribution is very narrow along the reaction coordinate, especially in the product side of the reaction.

Another set of relevant coordinates for the description of the FLP systems is the two angles $BH_B H_P$ (Fig. 2.28) and $H_B H_P P$ (Fig. 2.29). The angles are often used to interpret molecular orbital overlaps along the reaction path. Although the IRC gives again a reasonable guess for these parameters, the distributions are more spread than those of the distances. Similarly, the torsional angle defined by the atoms $BH_B H_P P$ proved to be quite undetermined, and they scatter quite considerably along the reaction coordinate (Fig. 2.30). The IRC provides only strongly qualitative picture for its evolution.

The static calculations also suggest that formation of the $H_P P$ and $BH_B$ bonds are correlated. In order to describe this correlation we have plotted the distribution

of the $H_P$P distance with respect to the $BH_B$ distance. The IRC curve has been projected again onto these two coordinates. The distribution shows that the correlation is significant and follows the picture provided by the IRC. We also see, that this correlation of the two distances is stronger at the product side.



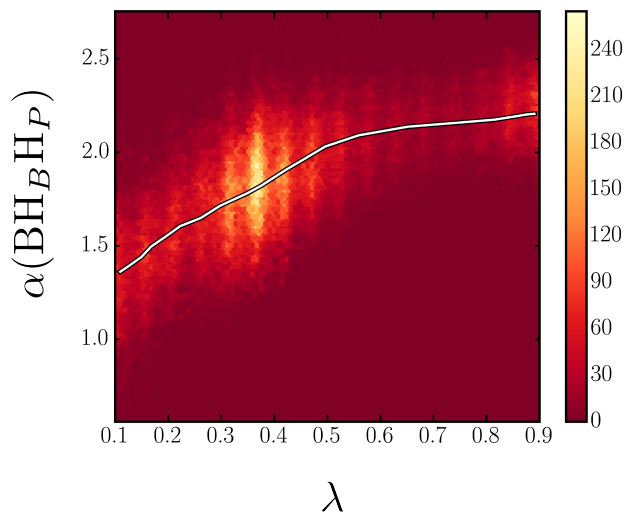**Figure 2.28:** Distribution of the $BH_BH_P$ angle with respect to the reaction coordinate. The IRC is designated with white line.

**Figure 2.29:** Distribution of the $H_B H_P P$ angle with respect to the reaction coordinate. The IRC is designated with white line.
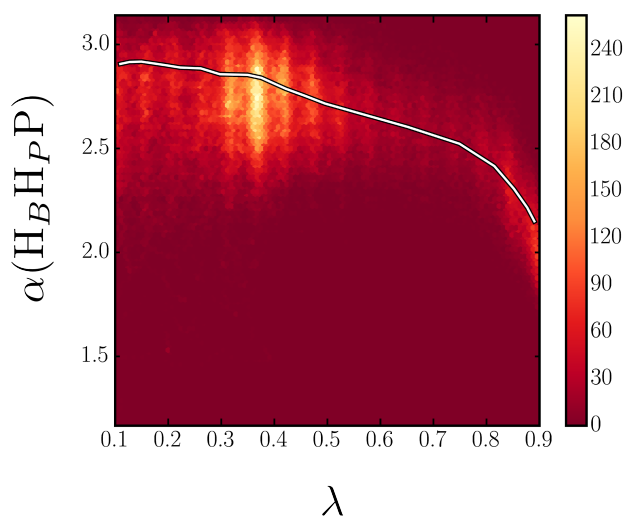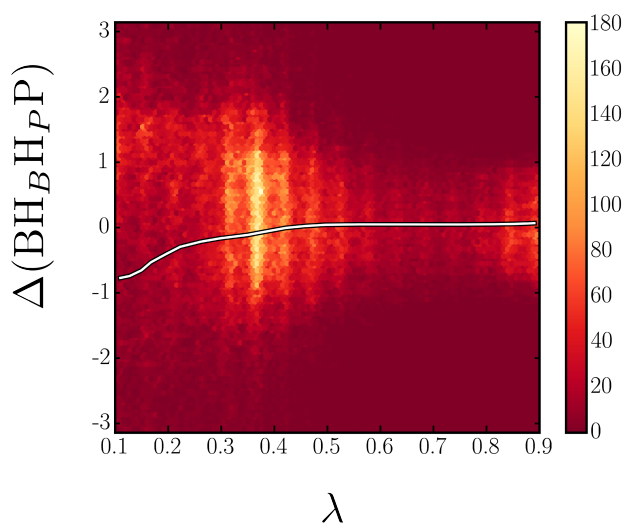


**Figure 2.30:** Distribution of the $BH_B H_P P$ dihedral with respect to the reaction coordinate. The IRC is designated with white line.
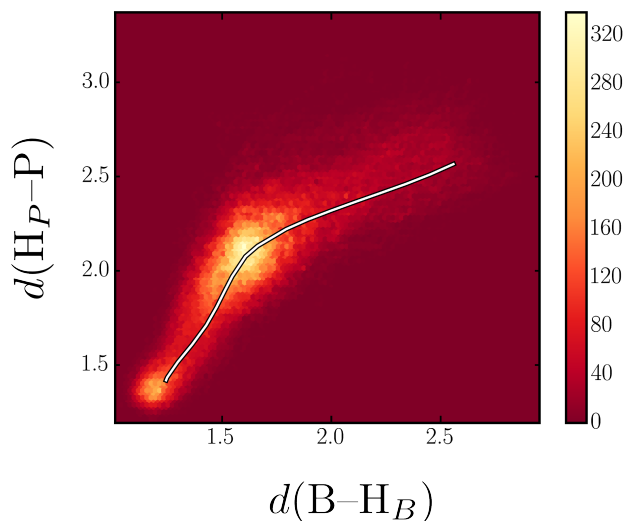
**Figure 2.31:** Distribution of the $H_P P$ distance with respect to the $BH_B$ distance. The IRC is designated with white line.

Analysis of the temperature curves (Fig. 2.32) yields similar conclusions to those obtained for the previous systems. Again we see that the largest amount of heat is localized on the P–H unit. This feature can be observed for each system and points to an important reactivity issue regarding subsequent reactions. An extremely heated proton on the Lewis-base phosphorous or nitrogen atom is likely to participate in subsequent non-equilibrium reactions such as degradation of the complex or fast deprotonation of the system. However we note that these issues require simulations with explicit solvent, because heat dissipation might be significantly effected by the solvent cage.
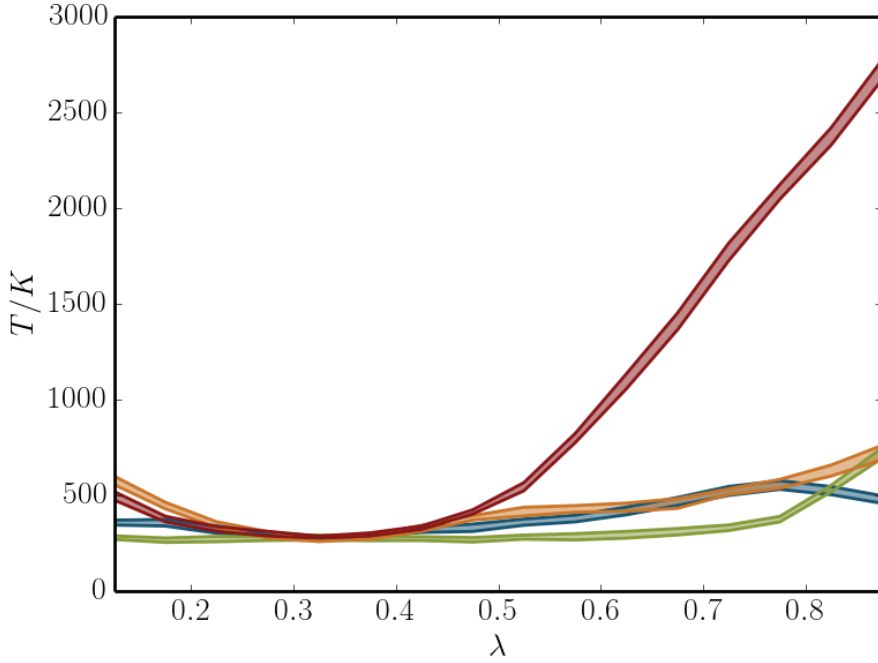
**Figure 2.32:** Temperature of B (blue), P (green) and the activated hydrogen atoms $H_B$ (orange), $H_P$ (red). The shaded area corresponds to the 90% confidence interval calculated by the standard error of the mean.

In the reactant side the kinetic energy of the hydrogen atoms is significantly large for all three systems. Detailed insight can be obtained by decomposing the velocities of the H atoms into vibrational, rotational and translational components. The center of mass for the two atoms defined as:

$$\boldsymbol{r}_{com} = \frac{m_H \boldsymbol{r}_1 + m_H \boldsymbol{r}_2}{2m_H} \tag{2.2}$$

$$= \frac{\boldsymbol{r}_1 + \boldsymbol{r}_2}{2} \tag{2.3}$$

The center of mass velocity can be obtained by simple derivation:

$$\boldsymbol{v}_{com} = \frac{\boldsymbol{v}_1 + \boldsymbol{v}_2}{2}. \tag{2.4}$$

The translational kinetic energy corresponding to the center of mass motion is provided by the following expression:

$$E_{kin}^{trans} = 2m_H \frac{v_{com}^2}{2}. \tag{2.5}$$

The instantaneous temperature can be calculated the translation as:

$$T_{trans} = \frac{2E_{kin}^{trans}}{3k_B} \tag{2.6}$$

The velocity of the atoms in the center of mass coordinate system is defined as:

$$\boldsymbol{v}_i^{com} = \boldsymbol{v}_i - \boldsymbol{v}_{com} \tag{2.7}$$

This velocity can be projected into the axis defined by the two hydrogen atoms, in order to obtain the velocity for vibrations of the hydrogen molecule:

$$\boldsymbol{v}_{vibr} = (\boldsymbol{v}_i^{com} \cdot \boldsymbol{e}) \cdot \boldsymbol{e}, \tag{2.8}$$

where $\boldsymbol{e}$ is unit vector of the H–H bond. The corresponding kinetic energy contribution is:

$$E_{kin}^{vibr} = 2m_H \frac{v_{vibr}^2}{2} \tag{2.9}$$

The instantaneous vibrational temperature for this single degree of freedom is:

$$T_{vibr} = \frac{2E_{kin}^{vibr}}{k_B} \tag{2.10}$$

The contribution of the two rotational degrees of freedom to the kinetic energy can be calculated from the energy conservation:

$$E_{kin}^{rot} = E_{kin}^{tot} - E_{kin}^{trans} - E_{kin}^{vibr}, \tag{2.11}$$

where $K_{tot}$ is the total kinetic energy of the hydrogen atoms. The rotational temperature is:

$$T_{rot} = \frac{2E_{kin}^{rot}}{2k_B} \tag{2.12}$$

The kinetic energy distribution between the different degrees of freedom of the $H_2$ molecules reaching the TS region shows an interesting picture (Fig. 2.33). The initial translational kinetic energy of the hydrogen molecules is much larger than in the transition state. In contrast, the kinetic energy of the vibrational and rotational degrees of freedom vary negligibly. This implies that the major part of the free energy necessary to bring the system to activated state concentrated in the translational degree of $H_2$.

The instantaneous temperature distribution shows the same picture (Fig 2.34). Indeed the temperature of the translation becomes approximately three times higher than the temperatures of the other degrees of freedom.

In summary we have explored the mechanism of the hydrogen activation of three FLP systems in finite temperature picture. The free energy exploration has supported the single-step mechanistic picture from previous static calculations. Statistical analysis of the structural parameters of the committor calculations has shown that the IRC configurations are relevant in the finite temperature as well. The velocity evolution of
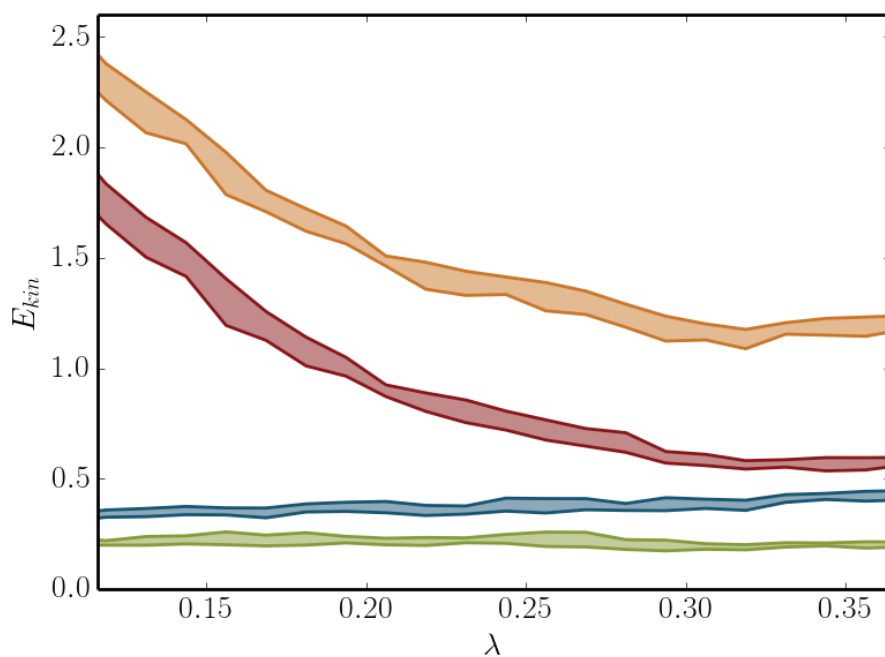
**Figure 2.33:** Kinetic energy distribution between the different degrees of freedom of the $H_2$ molecule with respect to the reaction coordinate: total (orange), translational (red), vibrational (green) and rotational (blue). The shaded area corresponds to the 90% confidence interval calculated from the standard error of the mean. All data sets are depicted in $10^{-20}$ J unit.

the committor trajectories provided additional informations about the reactions. The product side heat increase of the proton was significant in all systems. In addition the reaction side decomposition of the kinetic energy revealed the importance of the translational kinetic energy in hydrogen activation.

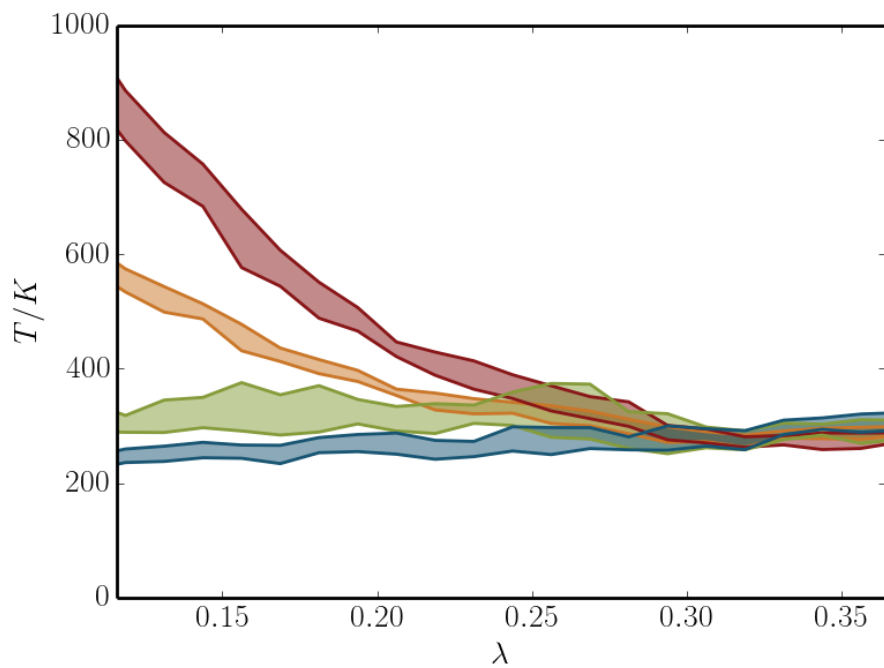**Figure 2.34:** Temperature distribution between different degrees of freedom of the $H_2$ molecule along the reaction coordinate: total (orange), translational (red), vibrational (green) and rotational (blue). The shaded area corresponds to the 90% confidence interval calculated from the standard error of the mean.

# Chapter 3

# Methodological development

This chapter presents the methodological developments we have done recently. First a rate theory [5] is presented, then the development of a new reaction coordinate optimizing method is discussed. [6]

# 3.1 Divided Saddle Theory – new method for calculating rate constants [5]

## 3.1.1 Introduction

As it was pointed out in Section 1.6.2, most methods for rate constant calculation are based on the assumption that the dividing surface (defined by the half-half splitting probability) is known, a priori. Methods such as the Bennett-Chandler [68,69], the EPF [71], or the expression from Hummer [73] use trajectory calculations that have been initiated from the dividing surface. The corresponding initial structures can be harvested from constrained molecular dynamics simulations if the reaction coordinate and its value corresponding for the dividing surface are known. However searching for the right reaction coordinate might include committor calculations as explained in section 1.4.3. Even if the proper reaction coordinate is known searching for the dividing surface itself implies expensive committor calculations along the reaction coordinate. Even if both the reaction coordinate and the dividing surface are known, generating equilibrium ensemble of the dividing surface is a challenging computational task. As it was already mentioned for transition path sampling simulations, transition pathways between reactant and product states might be divided by large barriers. In such cases equilibration with an additional constrain bounds the system into a particular reaction channel. Additionally, constrained simulation in CV space is not always available in molecular dynamics codes.

One way to overcome these problems is to generate geometries by umbrella sampling instead of constrains especially by the Hamiltonian exchange or parallel tempering variants. If umbrella potential is applied the geometries can be sampled by choosing the configurations in the vicinity of the dividing surface. Besides the arbitrariness of considering a configuration close or far from the surface, this approach is rather inefficient. It includes only a small portion of the generated configurations into the sample. Naturally the efficiency of the configuration generation can be enhanced by applying larger force constant. However, it also requires smaller time-step.

Beside the technical difficulties, there is an inconsistency behind the physical picture of the methods that are based on the dividing surface ensemble. This aspect can be enlightened by the example of an *out-of-equilibrium* reaction. Forward rate constant is calculated in BC-like methods from the assumption of equilibrium between reactant state and the dividing surface. The backward rate constant is calculated from the equilibrium assumption between the product state and the dividing surface. For an out-of-equilibrium reaction, the assumed equilibrium relations are against the zeroth law of thermodynamics, which defines equilibrium as a reflexive and transitive

relation. Namely the $RS$ should be in equilibrium with $PS$ if $RS$ is in equilibrium with $ds$ and $PS$ is in equilibrium with $ds$.

The difficulties and inconsistency of the BC-like methods and the computational difficulties of other methods (e.g. TPS) encouraged us to develop a new theory for rare events that can facilitate the development of new methods for rate constant calculations.

### 3.1.2   The theory

Our theory is based on the assumption, that the reaction coordinate $\lambda$ is known. We define the reactant and product states as adjoint ranges of the reaction coordinate. The two regions meet at the dividing surface. We assume that the system is in equilibrium *within* the reactant state. If the backward rate constant is also desired, the same assumption for the product state is also necessary. An equilibrium between the reactant and product states is not required.

We have started the derivation of the rate constant similarly to the Bennett-Chandler [68, 69] procedure by obtaining the microscopic rate constant from a sufficiently long equilibrium simulation. For such a long trajectory the macroscopic rate constant (inverse of mean life time ($\tau$)) can be calculated in a straightforward approach:

$$k = 1/\tau = \frac{N^t}{t^t_{RS}} = \frac{1}{\langle t_{RS} \rangle}, \tag{3.1}$$

where $t^t_{RS}$ is the total time spent by the system in the reactant state ($RS$) and $N^t$ is the total number of forward reactions. In order to define reactions we have introduced the definition of stable reactant and stable product regions similarly to EPF [71] and the method of Hummer [73]. We have divided the trajectory into segments at its crossing points with the borderlines of the stable reactant and product states (see Fig. 3.1). We consider a trajectory segment forward reactive if it leaves the stable reactant region and enters to the stable product region without revisiting the stable reactant region. Time reversion converts any forward reactive event to a backward reactive event and vice versa. Eq. 3.1 is a proper definition for the rate constant. However, it is only efficient for fast reactions. $\tau$ is frequently many orders of magnitude larger than the typical time scale of simulations. Therefore calculation of Eq. 3.1 with proper statistics is not generally possible.

In order to derive a computationally feasible method for rate constant calculations we have factorized expression 3.1. To this end we have introduced an auxiliary region, the *saddle domain* ($SD$) (See Fig. 3.1). The saddle domain extends from the dividing surface into the reactant state to arbitrary length. The same region can be defined for the backward region likewise. The factorization is performed as:
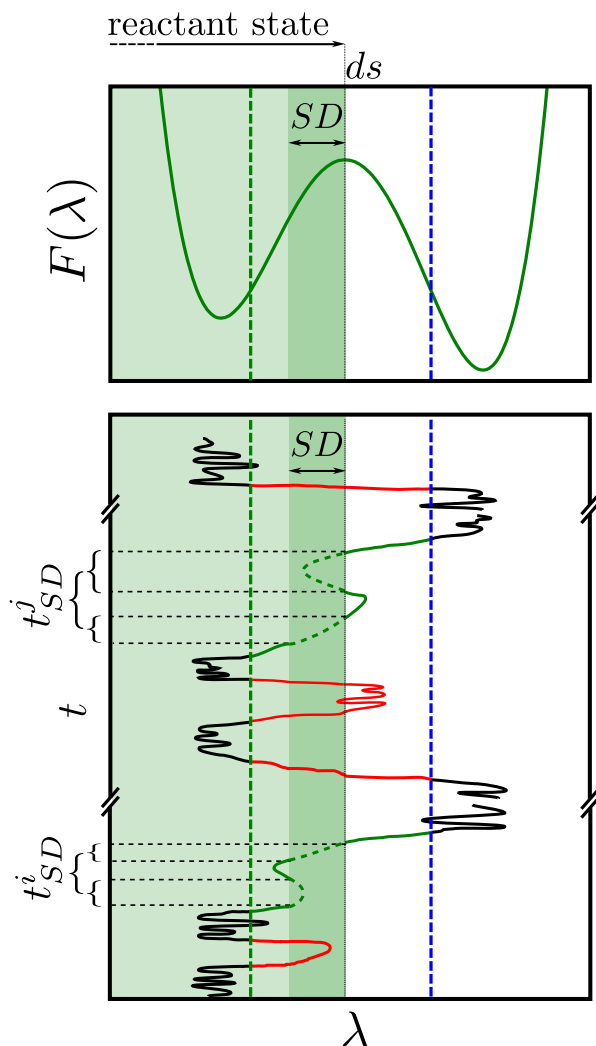
**Figure 3.1:** Top: definition of reactant state (*RS*) and *Saddle Domain SD* on a typical free energy profile. Both *RS* and *SD* extends from the dividing surface *ds*. The stable reactant and product regions are defined by green and blue dashed lines. Bottom: typical time evolution of the reaction coordinate for a two state system. Forward reactive trajectory segments are green, while non reactive trajectory segments are indicated with red lines. Note, that some of the red segments can be reactive in the backward direction. $t_{SD}^i$ and $t_{SD}^j$ denotes the total time spent in the *SD* by the *i*-th and *j*-th trajectory segments respectively. These visits are indicated by dashed green lines.

$$k \quad = \quad \frac{N^t}{t_{RS}^t} \tag{3.2}$$

$$k_{DST} \quad = \quad \frac{N^t}{t_{SD}^t} \cdot \frac{t_{SD}^t}{t_{RS}^t} \tag{3.3}$$

$$= \quad k_{SD} \cdot \alpha_{RS}^{SD}. \tag{3.4}$$

In Eq. 3.4 we have introduced $k_{SD}$, the rate constant from the $SD$ which is defined analogously to the forward rate constant in Eq. 3.2:

$$k_{SD} = \frac{N^t}{t_{SD}^t} \tag{3.5}$$

The remaining function is the conversion factor, which often appears in chemistry:

$$\alpha_{RS}^{SD} \quad = \quad \frac{t_{SD}^t}{t_{RS}^t} \tag{3.6}$$

$$= \quad \frac{P_{SD}}{P_{RS}} \tag{3.7}$$

$$= \quad \frac{c_{SD}}{c_{RS}} \tag{3.8}$$

So far the equilibrium assumption has not been employed. The equations above remain unchanged even in the case of steady state kinetics of a catalytic cycle. If equilibrium within the reactant state is assumed, $\alpha_{RS}^{SD}$ can be calculated by applying the definition of the free energy surface:

$$\alpha_{RS}^{SD} = \frac{\int\limits_{\lambda \in SD} e^{-\beta F(\lambda)} d\lambda}{\int\limits_{\lambda \in RS} e^{-\beta F(\lambda)} d\lambda} \tag{3.9}$$

### 3.1.3  The method

The remaining task is to find an efficient algorithm to evaluate Eq. 3.5. To this end we have expressed it as an ensemble average on the trajectory segments.

$$k_{SD} \quad = \quad \frac{N^t}{t_{SD}^t} \tag{3.10}$$

$$= \quad \frac{\sum N^i}{\sum t_{SD}^i} \tag{3.11}$$

$$= \quad \frac{\langle N \rangle_{segments}}{\langle t_{SD} \rangle_{segments}} \tag{3.12}$$

The reactivity of the $i^{\text{th}}$ segment is denoted with $N^i$, which is unity for reactive segments, and zero for non-reactive segments. The time spent by a particular segment

in $SD$ is denoted with $t_{SD}^i$. The trajectory segments can be generated by sampling configurations in $SD$ and initiating trajectories with random velocities, integrating the equations of motion for each random initial velocity forward and backward in time until they reach one of the stable states. Note, that trajectories with longer residence time in $SD$ are more likely to be sampled. Therefore this bias needs to be compensated. This can be analytically accomplished by reweighting each segment with the factor of $1/t_{SD}$:

$$k_{SD} \quad = \quad \frac{\langle N \rangle_{segments}}{\langle t_{SD} \rangle_{segments}} \tag{3.13}$$

$$= \quad \frac{\langle N/t_{SD} \rangle_{NVT}}{\langle t_{SD}/t_{SD} \rangle_{NVT}} \tag{3.14}$$

$$= \quad \left\langle \frac{N}{t_{SD}} \right\rangle_{NVT}. \tag{3.15}$$

Naturally, if the sampling is generated using other biases, additional reweighting is necessary for calculating the expressions correctly. The final expression for $k_{DST}$ can be obtained by including the definitions in Eqs. 3.15, and 3.9 into Eq. 3.4:

$$k_{DST} = \left\langle \frac{N}{t_{SD}} \right\rangle_{NVT} \frac{\int\limits_{\lambda \in SD} e^{-\beta F(\lambda)} d\lambda}{\int\limits_{\lambda \in RS} e^{-\beta F(\lambda)} d\lambda}. \tag{3.16}$$

### 3.1.4 Comparison with other methods

The clear advantage of our method is that it is not necessary to know the dividing surface a priori. One can sample the geometries around the approximate position of the dividing surface with any biased sampling method. One of the most natural choices is to apply umbrella sampling. The committor or the average reactivity along the reaction coordinate can be calculated from the short dynamical trajectories. In this way the dividing surface defined by the isocommittor or the maximal reactivity can be localised.

All trajectories can be utilized in the calculation of the rate constant through Eq. 3.16. Once the position of $ds$ is determined the rate constant can be calculated by a simple postprocessing of the trajectories. At this point the width of the $SD$ is arbitrary. We emphasize that although the width of $SD$ might affect the efficiency, its variation does not change the calculated rate constant. According to our experience the $SD$ can be chosen to cover a $\lambda$-range within 1-2 $kT$ from the $ds$.

It is interesting to examine our theory and the method in the limits of the $SD$ width. First let us extend it to infinity. $\alpha_{RS}^{SD}$ becomes unity, and $k_{SD}$ will be equal to $k$. In this case our method will recover the brute force calculation of the rate constant. However the efficiency still depends on the applied bias potential for sampling. If the bias potential is constant, the efficiency will be the same as performing undisturbed

long simulations and counting the rare events. With a moderate strength umbrella potential the efficiency can be enhanced.

In the zero $SD$-width limit our rate expression transforms:

$$k_{DST} = \left\langle \frac{N}{t_{SD}} \right\rangle_{NVT} \cdot d\lambda \cdot \frac{e^{-\beta F(\lambda = ds)}}{\int\limits_{\lambda \in RS} e^{-\beta F(\lambda)} d\lambda} \qquad (3.17)$$

Since the total time spent in the $SD$ can be expressed as the sum of the individual visits of $SD$ by a given segment, our expression now recovers the expression of Hummer [73]:

$$k_{DST} = \left\langle \frac{N}{t_{SD}} \right\rangle_{NVT} \cdot d\lambda \cdot \frac{e^{-\beta F(\lambda = ds)}}{\int\limits_{\lambda \in RS} e^{-\beta F(\lambda)} d\lambda} \qquad (3.18)$$

$$= \left\langle \frac{N}{\sum_i t_i} \right\rangle_{NVT} \cdot d\lambda \cdot \frac{e^{-\beta F(\lambda = ds)}}{\int\limits_{\lambda \in RS} e^{-\beta F(\lambda)} d\lambda} \qquad (3.19)$$

$$= \left\langle \frac{N}{\sum_i |v_i|^{-1}} \right\rangle_{NVT,ds} f_{eq}(\lambda_{ds}) \qquad (3.20)$$

The last line is the same expression as in Eq. 1.105. One can recognize the unnormalized transmission coefficient from the equations above:

$$R_{DST} = \left\langle \frac{N}{t_{SD}} \right\rangle_{NVT} \cdot d\lambda, \qquad (3.21)$$

which can be compared with the corresponding equation of EPF in Eq.1.104. By taking into account the non-vanishing terms only, the EPF expression can be rewritten as:

$$R_{EPF} = \left\langle \frac{\chi^{EPF}}{t_{SD}^f} \right\rangle_{NVT} \cdot d\lambda \qquad (3.22)$$

Here $t_{SD}^f$ denotes the infinitesimal time period spent by a reactive trajectory with velocity $d\lambda/t_{SD}^f$ during the *first* crossing of the $SD$ with $d\lambda$ width.

The two expressions are very similar and indeed equivalent in infinite sampling. Note, that DST accumulates the true dynamical information from a reactive segment irrespectively which configuration of the reactive trajectory segment has been sampled. In contrast EPF provides the same information from a specific reactive segment, if it has been sampled repetitively in all $ds$ crossing points. EPF gives non-zero contribution only at the first crossing point of a given trajectory, where the contribution is $1/t_{SD}^f$. In contrast DST gives $1/t_{SD}$ contribution with $t_{SD}/t_{SD}^f$ higher probability for a given trajectory. This means that DST is more efficient in terms of the number of simulations (for example if the trajectories are already available from a committor analysis). However, if one intends to calculate only rate

constant without calculating the committor, EPF is more efficient. The efficiency is obtained by the termination of the trajectories that attempt a recrossing during backward integration. It is worth emphasizing, that in this comparison we neglected the simulations and computational cost necessary to find the $ds$, which is itself in the order of a DST calculation. In terms of overall computational time DST is at least as effective as EPF. In the most favourable case DST is twice as efficient as EPF.

### 3.1.5 Numerical tests

In order to assess the performance of our method we have performed long unbiased simulations and DST calculations for two systems and then compared the calculated rate constants. All calculations have been performed by employing the CP2K simulation package [136]. The first system was the equilibrium of the alanine-dipeptide between the C5 and the C7 equatorial conformations. The PES was described by the Amber forcefield [97], and all NVT simulations have been carried out with Nose-Hoover global thermostat at 300 K and with 0.3 fs time-steps.

This transformation is known to be sufficiently described by the two torsional angles $\Phi$ and $\Psi$. The free energy with respect to these two CVs was explored by multiple-walker metadynamics simulation. A linear combination of these two CVs was chosen as an approximate reaction coordinate ($\lambda$):

$$\lambda = a \cdot \Phi + b \cdot \Phi + c \tag{3.23}$$

where a, b and c were chosen to yield a vector (see Fig.3.2 ) that connects the minima of the reactant and product and to provide continuous transition from 0 to 1 from reactant to product ($a = 0.30$, $b = -0.36$, $c = 1.79$). $\lambda$ is simply the projection of the $(\Phi, \Psi)$ points to this vector (See Fig. 3.2). As a reference we have performed a 500 ns unbiased simulation. The free energy along the reaction coordinate has been calculated through Boltzmann inversion.

In order to identify the true dividing surface, DST calculation has been carried out. 1167 configurations have been sampled from the region $\lambda \in [0.4, 0.67]$, which encloses the top of the barrier on the estimated free energy surface. The committor along $\lambda$ has been calculated from the DST calulations, and the dividing surface was assigned to the $\lambda = 0.57$ value. The definitions of the reactant side saddle domain and product side saddle domain are: $ds_R = \lambda \in [0.4, 0.57]$ and $ds_P = \lambda \in [0.57, 0.67]$. With the definition of $ds$ and $SD$-s in hand, we have calculated the rate constant from both the unbiased simulations and from DST trajectories. Moreover, as the free energy barrier heights are often used for TST expression (Eq. 1.114) in the literature (See eg. [140–146]), we calculated the rate constant in this manner as well. The results of the different approaches have been summarized in Table 3.1.

The rate constants calculated in the framework of DST are in excellent agreement with the unbiased simulation results. Moreover, the equilibrium constants from the reference and DST are also in perfect agreement. However it must be noted, that the
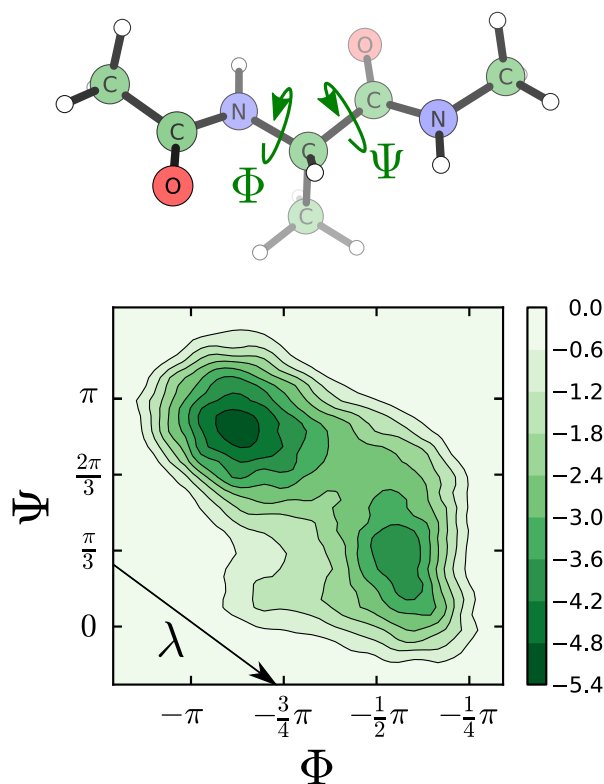
**Figure 3.2:** Top: Defintion of the collective variables ($\Phi$, $\Psi$) describing the C5 $\rightleftharpoons$ C7$_{\text{eq}}$ transformation of alanine dipeptide. Bottom: Definition of the reaction coordinate $\lambda$ and the free energy surface. Contours are depicted in kcal/mol units.

standard error of the mean is consistently larger for our technique than for the direct calculations. At variance, applying the TST expression in Eq. 1.114 for free energy barrier heights resulted in discouraging results. The rate constants are 5-7 times larger than the reference values, and their ratio also show the limitations of this approach.

In order to provide quantitative assessment of the efficiency of DST, we compare it numerically with the very efficient EPF method, employing the alanine-dipeptide test system. We have performed a new set of simulations for DST and a separate set for EPF with NVT ensemble of shooting points at $SD$ and $ds$ (determined from our previous calculation) respectively. Fig. 3.3 shows, that DST provides a slightly faster convergence in terms of necessary configurations. Within the performed number of computations EPF's convergence is even problematic. However EPF uses a magnitude smaller computational effort as it follows only a small portion of trajectories until they reach one of the stable states. Nevertheless we should bear in mind that the proper estimation of the overall efficiency of the rate constant calculations

**Table 3.1:** Comparison of rate constants and equilibrium constants from unbiased simulations, and DST calculations. Rate constants are in $10^{11}$ s$^{-1}$ units. Standard error is given in parenthesis.

| Alanine-dipeptide | direct | present method | TST |
|---|---|---|---|
| $k_1$ | 0.266 (0.011) | 0.257 (0.014) | 1.474 (0.063) |
| $k_{-1}$ | 1.629 (0.033) | 1.564 (0.068) | 11.866 (0.464) |
| $K$ | 0.163 (0.006) | 0.164 (0.011) | 0.124 (0.007) |
| Barbaralene | direct | present method | TST |
| $k$ | 19.26 (0.51) | 18.63 (1.65) | 49.63 (6.25) |

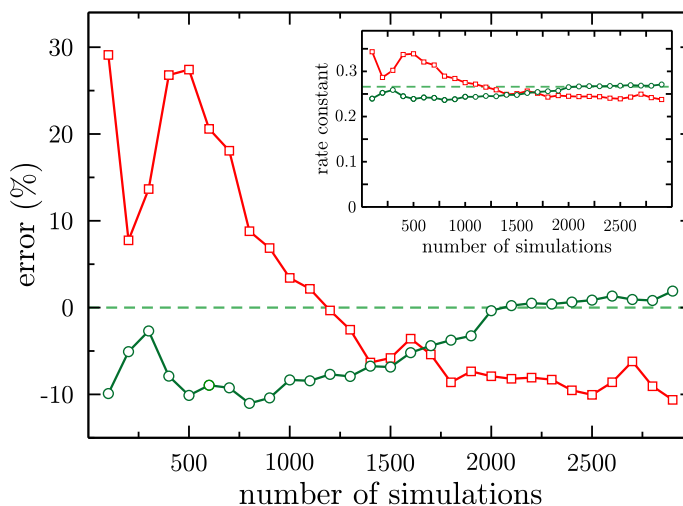of different methods should include the cost of the identification of the dividing surface.



**Figure 3.3:** The convergence of the calculated rate constants for the C5 → C7eq transformation in terms of relative error from best estimate from direct MD. (dashed line). Inset: estimated rate constants (in $10^{11}$ 1/s units) as a function of the number of the generated trajectory segments. Values for DST and EPF are depicted in green and red respectively.

The convergence of our method is surprisingly fast, as approximately 2000 simulations and 1 ns of total simulation time were necessary for the converged results. This amount of computation is feasible with the present-day simulation techniques, especially bearing in mind that the trajectory calculations can be trivially parallelized. These results encouraged us to perform similar calculations with a much more CPU demanding forcefield.

Our next example is the Cope rearrangement of barbaralene at 1000 K. The high temperature was necessary to obtain sufficient statistics with direct calculation. The BLYP functional have been chosen for the description of the PES. The solution for the electronic structure problem has been performed with a double $\zeta$ basis sets with additional set of polarization functions. The Goedecker-Teter-Hutter [137] pseudopotentials have been applied and plane-wave basis set an with energy cutoff of 300 Ry has been employed to expand the electron densities. The periodic cubic simulation cell parameter was 20.8 bohr. The interaction between periodic images has been decoupled with the Martyna-Tuckermann method [147]. Since the reaction is degenerate, the choice of proper reaction coordinate did not require preliminary free energy exploration. We have chosen an anti-symmetric reaction coordinate:

$$\lambda = (d_1 - d_2), \tag{3.24}$$

where $d_1$ and $d_2$ denote the forming and breaking C–C bonds lengths (see Fig. 3.4). The $SD$ region has been defined as $\lambda \in [-0.20 \text{ bohr}, 0.00 \text{ bohr}]$. 200 configurations have been collected from $SD$. The short trajectory simulations took 16.3 ps and their result compares well with the values of the unbiased MD with 1 ns long simulation time. The results of the simulations are presented ion Table 3.1. The rate constant calculated in the DST framework and that by the unbiased simulations compare well. The difference between the two approaches is approximately 3.3%. Similarly to the case of the alanine dipeptide, the application of the TST formula for the barrier heights resulted in an overestimated rate constant.

In summary we have developed a theory and and method for the calculation of rate constants from molecular dynamics simulations. The method employs a free energy profile along an appropriate reaction coordinate and short committor calculations. The method has been validated by numerical simulations. We have analitically showed how our theory relates to Bennett-Chandler type methods.

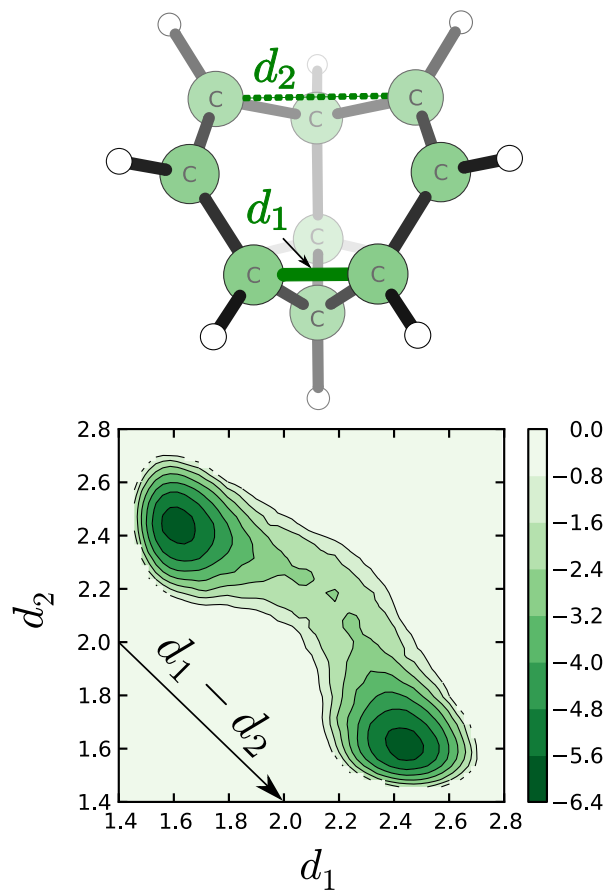**Figure 3.4:** Top: Definition of the distances ($d_1$ and $d_2$) for barbaralene. Bottom free energy surface corresponding to $d_1$ and $d_2$ and the definition of reaction coordinate for the Cope rearrangement of barbaralene. Units are kcal/mol and bohr.

# 3.2 Committor map collective variable [6]

## 3.2.1 Introduction

We have emphasized in Section 1.4.3, finding appropriate CVs and representing the reaction coordinate is a complex computational challenge. We have already pointed out, the description of a reaction can be validated through committor calculations. This is the state-of-the-art approach to decide whether our set of coordinates is suitable for the description. Often the number of important coordinates is too large to explore the free energy surface in their space. Therefore, in practice dimensional reduction techniques such as path CV, property maps or linear combinations of CVs are applied. Even if the right set of CVs are in use, dimension reduction techniques might lead to insufficient description of the reactive event. For example, it is possible that the committor distribution at the dividing surface is not singly peaked at the $\phi = 0.5$ value. If a committor calculation provides unsatisfactory results, there is no general procedure to utilise the enormous amount of information to find a better reaction coordinate.

In this project we aimed to develop a methodology to find a suitable reaction coordinate as a function of CVs. Naturally, if the set of coordinates used for the description is incomplete, this approach will not be able to find the proper coordinate. Therefore we assume that a set of CVs is available which describes properly the rare event in question. Our simple approach follows the work of Ma and Dinner [49]. The attractive features from their work are the followings: i) the committor is mapped onto the space of the CVs on the basis of previous committor calculations. ii) As the same quantity (committor) is represented that is used to validate the dividing surface, the quality of this coordinate only depend on the set of CVs that are in use for the mathematical construction. Since a committor calculation is a simulation technique with the full configuration as an input, the committor can also be represented with respect to any set of CVs. iii) If the initial configurations are stored, the committor can be mapped onto a new set of CVs as well by simple post processing. The approach of Ma and Dinner also suffers from serious drawbacks. Firstly they apply neural networks for the representation of the committor function. This approach requires strong parametrization and careful testing. In their approach they have used 2000-2700 configurations for the description of the alanine dipeptide system and the method required precise committor value for each selected configuration. This has been achieved by the calculation of 100-400 committor trajectories per configurations. Moreover, the initial configurations were generated from expensive transition path sampling algorithm. The authors also emphasized the importance of uniform sampling density for the initial configurations. Their complicated, transition path sampling based algorithm only meets this requirement along the committor. The uniform sampling in orthogonal degrees of freedom is not guaranteed by their approach. Another disadvantage of the method is that it employs a complicated functional form for the representation.

They applied a multilevel neural network, which employs the Fermi-like (Eq. 1.10) function at each level. Due to the complex functional form they could not apply simple umbrella sampling with respect to this coordinate. They had to apply instead multidimensional umbrella sampling simulations with respect to the original CVs.

### 3.2.2  Description of the method

In order to keep the advantages of Ma and Dinner's approach but to overcome the difficulties of their method we applied a different flexible representation for the committor. We chose the property-map (Eq. 1.13) function for this purpose. The property-map provides numerous advantageous features for this application. First, it is an interpolation formula. The advantages of an interplation formula can be demonstrated by the following set-up. Let us suppose that 0 committor values are assigned for configurations at the stable reactant state and 1 committor values are assigned for the molecules at the stable product state. With such a set-up every molecule that is different from both reactant and product states according to the distance metric will be assigned with a value between zero and one. If a molecule is characterised by the same similarity (a distance value according to the metric) for the stable reactant state and for the stable product state, it will be assigned with the value of 0.5. This aspect of the function is very close to the chemical intuition, which attempts to assign characteristics for new molecules on the basis of similarity or dissimilarity to known substances.

Another interesting property of the functional form is that it performs local averaging and smoothing. The function gives similar results irrespective of averaged committor values or the individual zero and one committor values are employed as represented properties. If the same number of committor analyses have been performed for each configuration, the individual and averaged committors provide the same function. The function performs averaging, with weights that depend on the vicinity of the points. This aspect of the property map turns it a sort of smoothing method. Therefore it is not important to feed the function with converged committor values. This facilitates a better exploration of the configurational space, since only limited number of simulations is necessary for an initial configuration.

Previously we have defined the representation of the reaction coordinate as the property-map representation of a dictionary of committor results and the corresponding vectors of the CV space. In order to design an efficient method we need to define an algorithm for the generation of the starting configurations for committor calculations. We intend to keep this part of the algorithm flexible to meet the various needs of molecular simulations. We assume that the property map is initialized by zero and one values from the reactant and product stable states with the application of distance metric based on CVs. At this point the CV space is mapped with a simple interpolation. Alternatively the function initialization might be performed with additional committor values from the vicinity of the free energy saddle. This feature is especially useful if the

committor analysis of a preliminary reaction coordinate has not provided satisfactory results.

After the initialization of the coordinate any continuous biasing technique can be used for the generation of new configurations. Any static or adaptive biasing potential or force method is suitable for this purpose that applies potentials or forces instead of constrains. Afterwards a histogram is calculated along the trial coordinate ($f(\lambda)$). By assigning $1/f(\lambda)$ sampling probability for each saved configuration, the equal sampling with respect to $\lambda$ can be ensured. However, this sampling cannot balance the uneven sampling from the initialization from the previous iterations. The sampling of unexplored regions of the CV space would be advantageous. It can be accomplished by the calculation of the number of neighbours from the dictionary of the previous committor calculations and CV results for each possible configuration. The number of the neighbours are calculated as follows:

$$N_n(\boldsymbol{q'}) = \sum_i \exp\left(-\gamma D_i(\boldsymbol{q}_i, \boldsymbol{q'})\right) \tag{3.25}$$

Note that this quantity is simple function of the distance function of the property map ($Z(\boldsymbol{q'})$ in Eq. 1.14):

$$N_n(\boldsymbol{q'}) = \exp\left(-\gamma Z(\boldsymbol{q'})\right) \tag{3.26}$$

Our final equation for the weights of the configurations to be sampled is:

$$w(\boldsymbol{q'}) = \frac{1}{N_n(\boldsymbol{q'}) \cdot f(\lambda(\boldsymbol{q'}))}. \tag{3.27}$$

Each iteration of our method consists of the foregoing generation of the new configurations, random sampling according to the $w(\boldsymbol{q'})$ weights and committor calculations from each configurations. Finally the dictionary of the committor results and the corresponding initial CV values are updated. Accordingly, the definition of the approximate reaction coordinate is also updated.

### 3.2.3 Numerical tests

In order to assess the applicability of our methodology, we have performed numerical test simulations for the $C5 \rightleftharpoons C7_{eq}$ conformational equilibrium of alanine dipeptide. For this work we have applied the Gromacs 4.6.5 [148–152] simulation package in combination with Plumed 2.0. [10] The simulations have used the Amber forcefield [97] with 0.3 fs time step. The simulation temperature of 300 K was provided by the CSVR thermostat [22] with a 100 fs time constant.

As a first step of the validation of our method we have to calculate a reference mapping of the committor for the $\Psi, \Phi$ plane corresponding to the $C5 \rightleftharpoons C7_{eq}$. To this end we have performed metadynamics simulation with the two CVs. The two valleys have been filled up with 20000 hills with dimensions of 0.1·0.1 rad$^2$ and height of 0.05 kJ/mol. The surface has been further smoothed with additional 20000 hills with 0.05·0.05 $rad^2$ dimensions and 0.025 kJ/mol height. The biasing potential hills have
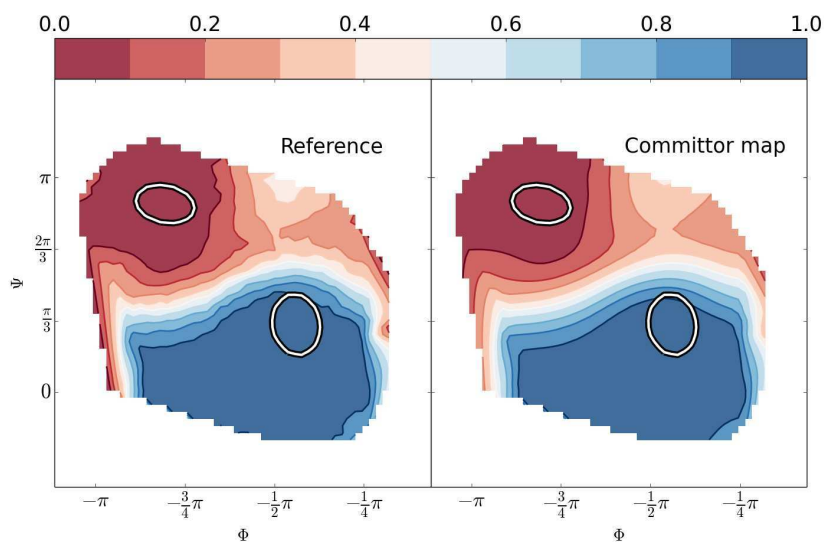
**Figure 3.5:** Comparison of the cubic spline represented reference data (left) and the property map representation (right) for the same data set

been deposited at every 500 steps. Afterwards the system was equilibrated without further bias deposition for an additional 12 ns. The CV range has been divided into a 50 x 50 grid, and each grid with higher than 20 geometries has been randomly sampled. In total 1090 grids with 20 geometries in each has been sampled. From each geometry 40 committor calculations have been initiated with random velocities. The results of each grid have been averaged. The corresponding mapping from the over 800000 committor calculations has been plotted through two dimensional cubic spline interpolation and with the suggested property map representation (applying $\gamma = 20$) on Fig. 3.5.

The comparison of the two representations shows promising results. The committor map (i.e. the property map representation of comittor) provides excellent agreement with the two dimensional spline representation. The commitor map is able to describe each important feature of the committor function. The mapping of commitor onto these two CVs show an interesting feature. Namely that the committor range between 0.2 and 0.3 is mapped into two disjunct regions of the CV space. This feature is due to the high potential energy region at $(-0.5\pi,\pi)$. It is remarkable that the committor map also describes this feature of the committor function. It is important to note, that generation of the coordinates with multiple dimension free energy exploration, and such a large number of committor calculations are very is computationally demanding for realistic systems. Therefore we studied the applicability of our algorithm with two different biasing techniques for configuration generation.

First we have applied replica exchange umbrealla sampling simulations for the

configuration generation. The umbrella sampling simulations were performed for 15 ps, at 300 K, with 0.3 fs simulation timestep. The umbrella potentials were applied at the following committor map values: 0.2, 0.35, 0.5, 0.65, 0.8. The force constants for the simulations have been set as 350 kJ/mol, while $\gamma$ was kept at the value of 20. Random exchange between the neighbouring simulations were applied at every 2000 MD step. In each iteration 100 configurations were chosen, and from each configuration 5 committor calculations were performed. Fig. 3.6 presents the results of this simulation. The important qualitative features of the reference committor function appear after 5 iterations. However the interesting feature of the 0.2-0.3 committor region appears only at the $20^{\text{th}}$ iteration.

Then we have tested our algorithm in combination with multiple worker metadynamics simulation. Similarly to the replica exchange umbrella sampling simulations, the geometries were generated by 5 simulations with an overall 75 ps simulation time. The metadynamics employed the committor map as a CV and used hill potentials with 0.1 width and 0.05 kJ/mol height. The hills were deposited at every 100 MD steps.

The results of this simulation is depicted in Fig. 3.7. Interestingly the results converge significantly faster with this geometry generation technique. All important qualitative features of the committor function appear already in the $5^{\text{th}}$ iteration. Both the reactant and the product regions and the transition state region is sufficiently described. Moreover, the interesting disjunct nature of the 0.2-0.3 committor range is also captured by this method. In the following 25 iterations the function remains considerably stable with small oscillations.

In summary we have introduced a new dimensionality reduction approach by representing the committor of the system by property map. In order to achieve an efficient mapping, we have developed an algorithm for the initial configuration generation and iterative reaction coordinate optimization. The algorithm can be combined with different free energy methods. As a test of the approach two free energy methods have been tested in combination with our algorithm. Our method provided proper mapping of the committor with both Hamiltonian replica exchange and metadynamics, with a faster convergence for the latter.
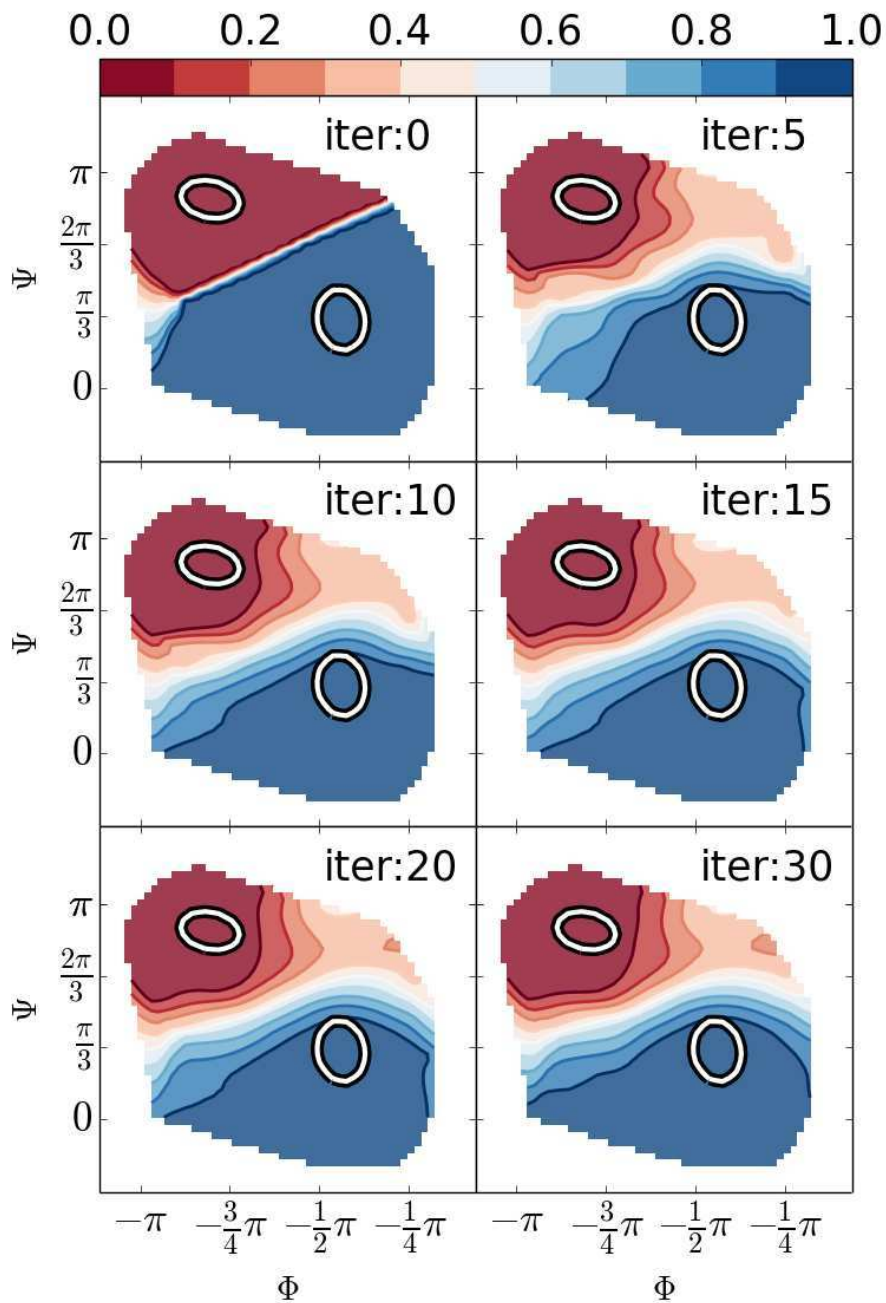
**Figure 3.6:** Convergence of the committor map CV with configuration generation through Hamiltonian-replica exchange molecular dynamics simulations. Angles are in radian.
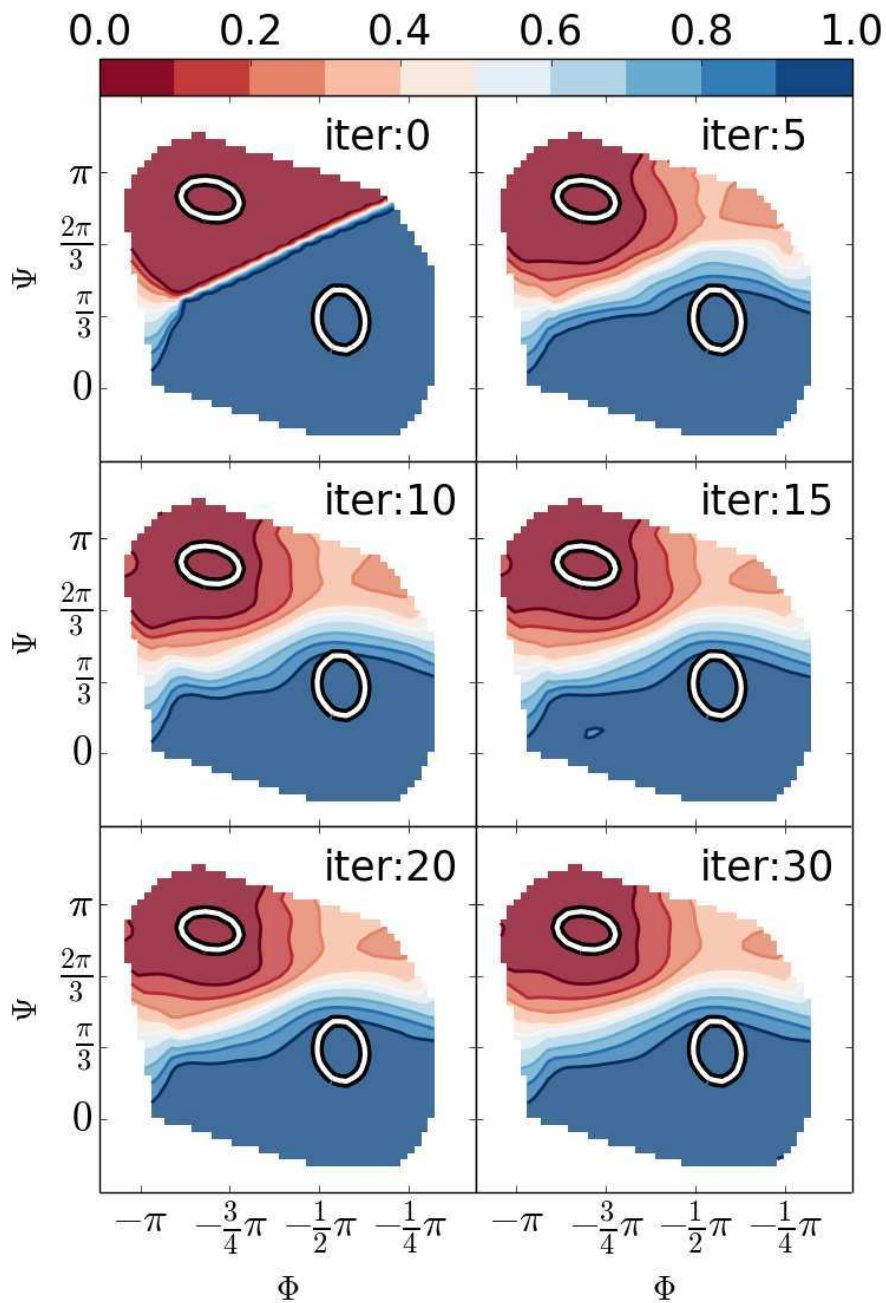
**Figure 3.7:** Convergence of the committor map CV with configuration generation through multiple-worker metadynamics simulations. Angles are in radian.

# Summary

The present thesis addresses reaction mechanisms by theoretical methods. The thesis can be divided into two thematically distinct directions: applications and method development. Along the first direction the thesis summarizes our mechanistic studies on four reaction families which are very important and promising from synthetic point of view. In these studies we have employed density functional methods which are essential to describe the bond breaking and bond formation processes. Regarding the second directions we have developed a theory for rate constants and a new method to efficiently calculate them. In general, rate constant calculations require the knowledge of the reaction coordinate. We also developed a protocol to optimize reaction coordinate employing dynamical information.

Our mechanistic studies provided a deeper understanding of palladium catalyzed C–H activation reaction and revealed the role of an interesting bimetallic catalytic species in the subsequent C–C coupling. The reaction path identified by our calculations is the rate-determining C–H activation step followed by a C–C coupling via a bimetallic complex and a Pd(III)-Pd(I) reductive elimination.

Silver catalysis is an emerging field in synthetic heterocyclic chemistry. Our theoretical exploration of a recently developed silver-mediated oxidative C-H/C-H functionalization could help to elucidate the mechanism of an efficient synthetic route toward substituted furans. The calculations have pointed out the dual role of silver in the mechanism: it is a reactant to initiate the reaction in a radical route and a catalyst in the later stages of the path to drive the reaction to the furan-ring formation. Some of our findings have also been verified experimentally.

The predictive power of our computation has been nicely confirmed in a joint experimental-theoretical study of direct trifluoro-ethylation of indoles. The calculations have identified not only the the reaction channel leading to C3 trifluoro-ethylation of indoles, but also explained the side reactions observed in experiment and the important role of the bases. On the basis of the results we have shown that the efficiency of a given substrate-base combination can be assessed by comparing the activation barriers of the competing N- and C-alkylations.

In the study of $H_2$ activation by Frustrated Lewis Pairs we compared two mechanistic pictures: a static and a dynamic view of the cooperative hydrogen splitting by FLPs. Combination of enhanced sampling methods with molecular dynamics has

been proved to be successful and yielded new mechanistic insights. On one hand, we could confirm the cooperative, concerted heterolytic $H_2$ picture emerged from earlier, static calculations. However, new, remarkable aspects of the reaction paths have been identified from the reactive trajectories. In particular, the temperature distributions at the reactant and product states disclosed new features of the reactions. At the reactant state the temperature distribution pointed out the role of $H_2$ translation energy in the activation. The thermally highly excited P–H unit at the product state is a strong confirmation of the asynchronity in the activation and may have crucial role in subsequent reactions.

We have developed a new and efficient method to calculate rate constants. The method uses information from the free energy profile along a suitable reaction coordinate and dynamical information from short unbiased trajectories. The main idea of the underlying theory is to introduce the concept of Saddle Domain and a factorization of the rate constant expression accordingly. We have shown that the method has good convergence properties and provides accurate rate constant values.

We have also developed a new approach to obtain an efficient CV where we represent the committor of the system by property map. The method is based on an efficient initial geometry generation algorithm and an iterative optimization. Combination of the method with enhanced sampling techniques has been shown to provide very promising results. A natural continuation of this direction is the application of the this new reaction coordinate in a rate constant calculation employing our new method for realistic systems.

# References

## R.1  Papers Forming the Basis of the Dissertation

[1] F. Szabó, J. Daru, D. Simkó, T. Zs. Nagy, A. Stirling, Z. Novák, "Mild Palladium-Catalyzed Oxidative Direct ortho-C–H Acylation of Anilides under Aqueous Conditions", *Adv. Synth. Catal.* **2013**, *355*, 685.

[2] J. Daru, Zs. Benda, Á. Póti, Z. Novák, A. Stirling, "Mechanistic Study of Silver-Mediated Furan Formation by Oxidative Coupling", *Chem. Eur. J.* **2014**, *20*, 15395.

[3] G. L. Tolnai, A. Székely, Z. Makó, T. Gáti, J. Daru, T. Bihari, A. Stirling, Z. Novák, "Efficient Direct 2,2,2-Trifluoroethylation of Indoles via C-H Functionalization", Chem. Commun. **2015** DOI: 10.1039/c5cc00519a

[4] J. Daru, I. Bakó, A. Stirling, I. Pápai, "AIMD modelling of frustrated Lewis pairs", manuscript in preparation

[5] J. Daru, A. Stirling, "Divided Saddle Theory: A New Idea for Rate Constant Calculation", *J. Chem. Theory Comput.* **2014**, *10*, 1121.

[6] J. Daru, G. Tóth, "Committor map collective variable", manuscript in preparation

## R.2  References

[7] D. Moroni, *Efficient sampling of rare event pathways: from simple models to nucleation.* Ph.D. Thesis, University of Amsterdam, (2005).

[8] P. G. Bolhuis and C. Dellago, in Rev. Comp. Chem., edited by K. B. Lipkowitz (Wiley, New York, 2010), Vol. 27, pp. 111–202

[9] M. Ceriotti, G. Tribello, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13023.

[10] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comp. Phys/ Commun.* **2014**, *185*, 604.

[11] M. Bonomi, D. Branduardi, F. Gervasio, and M. Parrinello, *J. Am. Chem. Soc.* **2008**, *130*, 13938.

[12] J. M. Park, A. Laio, M. Iannuzzi, M. Parrinello, *J. Am. Chem. Soc.* **2006**, *128*, 11318.

[13] D. Branduardi, F. L. Gervasio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 054103.

[14] V. Spiwok, B. Králová, *J. Chem. Phys.* **2011**, *135*, 224504.

[15] P. Atkins, J. De Paula, *Physical Chemistry*, 9th edition, Oxford University Press, Oxford, (2010)

[16] M. Tuckerman, *Statistical Mechanics and Molecular Simulations*, Oxford University Press, Oxford, (2008)

[17] H. F. Trotter, *Proc. Natl. Acad. Sci. USA* **1959**, *10*, 545.

[18] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *J. Chem. Phys.* **1982**, *76*, 637.

[19] S. Nosé, *J. Chem. Phys.* **1984**, *81*, 511.

[20] W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695.

[21] G. J. Martyna, M. L. Klein, M. Tuckerman, *J. Chem. Phys.* **1992**, *97*, 2635.

[22] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.

[23] D. Reith, M. Pütz, F. Müller-Plathe, *J. Comput. Chem.* **2003**, *24*, 1624.

[24] M. Rosenblatt, *Ann. Math. Statist.* **1956**, *27*, 832.

[25] E. Parzen, *Ann. Math. Statist.* **1962**, *33*, 1065.

[26] C. Chipot, A. Pohorille, *Free energy calculations–Theory and Applications in Chemistry and Biology.* Springer series in chemical physics 86, Springer Berlin Heidelberg New York, (2007)

[27] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187.

[28] A. M. Ferrenberg, R. H. Swendsen, *Phys. Rev. Lett.* **1989**, *63*, 1195.

[29] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *J. Comput. Chem.* **1992**, *13*, 1011.

[30] M. R. Shirts, J. D. Chodera, *J. Phys. Chem.* **2008**, *129*, 124105.

[31] T. Lee, B. K. Radak, A. Pabis, D. M. York, *J. Chem. Theory Comput.* **2013**, *9*, 153.

[32] T. Lee, B. K. Radak, M. Huang, K. Wong, D. M. York, *J. Chem. Theory Comput.* **2014**, *10*, 24.

[33] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **2000**, *329*, 261.

[34] W. Schommers, *Phys. Rev. A* **1983**, *28*, 3599.

[35] P. Maragakis, A. van der Vaart, M. Karplus, *J. Phys. Chem. B* **2009**, *113*, 4664.

[36] T. Huber, A. E. Torda, W. F. Gunsteren, *J. Comput. Aided Mol. Des.* **1994**, *8*, 695.

[37] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562.

[38] A. Laio, F. L. Gervasio, *Rep. Prog. Phys.* **2008**, *71*, 126601.

[39] M. Iannuzzi, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **2003**, *90*, 238302.

[40] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.

[41] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 30.

[42] M. Mezei, *Mol. Sim.* **1993**, *10*, 225.

[43] J. G. Kirkwood, *J. Chem. Phys.* **1934**, *2*, 767.

[44] E. A. Carter, G. Ciccotti, J. T. Hynes, R. Kapral, *Chem. Phys. Lett.* **1989**, *156*, 472.

[45] E. Darve, A. Pohorille, *J. Chem. Phys.* **2001**, *115*, 9169.

[46] W. Li, A. Ma, *Mol. Sim.* **2014**, *40*, 784.

[47] W. E, W. Ren, E. Vanden-Eijnden, *J. Phys. Chem. B* **2005**, *109*, 6688.

[48] D. Branduardi, J. D. Faraldo-Gómez, *J. Chem. Theory Comput.* **2013**, *9*, 4140.

[49] A. Ma, A. R. Dinner, *J. Phys. Chem. B* **B**, *2005*, 109.6769

[50] *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6732.

[51] B. Peters, B. L. Trout, *J. Chem. Phys.* **2006**, *125*, 054108.

[52] B. Peters, *J. Chem. Phys.* **2007**, *127*, 034109.

[53] C. Dellago, P. G. Bolhuis, F. S. Csajka, D. Chandler, *J. Chem. Phys.* **1998**, *108*, 1964.

[54] T. J. H. Vlugt, B. Smit, *Phys. Chem. Comm.* **2001**, *2*, 1.

[55] W. Lechner, J. Rogal, J. Juraszek, B. Ensing, P. G. Bolhuis, *J. Chem. Phys.* **2010**, *133*, 174110.

[56] D. A. McQuarrie, J. D. Simon, *Physical chemistry: a molecular approach*, Vol. 1. Sausalito, CA, University science books (1997).

[57] W. E, W. Ren, E. Vanden-Eijnden, *Phys. Rev. B* **2002**, *66*, 052301.

[58] C. G. Broyden, *J. Inst. Math. Appl.* **1970**, *6*, 76.

[59] R. Fletcher, *Comput. J.* **1970**, *13*, 317.

[60] D. Goldfarb, *Math. Comput.* **1970**, *24*, 23.

[61] D. F. Shanno, *Math. Comput.* **1970**, *24*, 647.

[62] B. Murtagh, R. W. H. Sargent, *Comput. J.* **1972**, *13*, 185.

[63] Ödön Farkas, H. B. Schlegel, *Phys. Chem. Chem. Phys.* **2002**, *4*, 11.

[64] P. Pulay, *cpl* **1980**, *73*, 393.

[65] P. Császár, P. Pulay, *J. Mol. Struct.* **1984**, *114*, 31.

[66] X. Li, M. Frisch, *J. Chem. Theory Comput.* **2005**, *2*, 835.

[67] K. N. Kudin, G. E. Scuseria, E. Cancés, *J. Phys. Chem.* **2002**, *116*, 8255.

[68] C. H. Bennett, in *Algorithms for Chemical Computations*; R. Christofferson, Ed.; ACS Symposium Series No. 46 ACS, Washington, D.C., (1977)

[69] D. Chandler, *J. Chem. Phys.* **1978**, *68*, 2959.

[70] J. P. Bergsma, J. R. Reimers, K. R. Wilson, J. T. Hynes, *J. Chem. Phys.* **1986**, *85*, 5625.

[71] T. S. van Erp, D. Moroni, and P. G. Bolhuis, *J. Chem. Phys.* **2003**, *118*, 7762.

[72] J. B. Anderson, *Adv. Chem. Phys.* **1995**, *91*, 381.

[73] G. Hummer, *J. Chem. Phys.* **2004**, *120*, 516.

[74] M. J. Ruiz-Montero, D. Frenkel and J. J. Brey, *Mol. Phys.* **1997**, *90*, 925.

[75] M. S. Green, *J. Chem. Phys.* **1954**, *22*, 398.

[76] R. Kubo, *J. Phys. Soc. Jpn.* **1957**, *12*, 570.

[77] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, New York, (1987)

[78] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, 864.

[79] W. Kohn, L.J. Sham, *Phys. Rev. A* **1965**, *140*, 1133.

[80] A. D. Becke, *J. Chem. Phys.* **2014**, *140*, 18A301.

[81] A. D. Becke, M. R. Roussel *Phys. Rev. A* **1989**, *39*, 3761.

[82] P. Mori-Sánchez, A. J. Cohen, W. Yang, *J. Chem. Phys.* **2006**, *125*, 201102.

[83] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372.

[84] I. C. Gerber and J. G. Ángyán, *Chem. Phys. Lett.* **2005**, *415*, 100.

[85] Y. Zhao, B. J. Lynch, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 4786.

[86] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[87] H. Casimir, D. Polder, *Phys. Rev.* **1948**, *73*, 360.

[88] B. M. Axilrod, E. Teller, *J. Chem. Phys.* **1943**, *11*, 299.

[89] J. Muto, *Proc. Phys. Math. Soc. Jpn.* **1943**, *17*, 629.

[90] A. Becke, *Phys. Rev. A* **1988**, *38*, 3098.

[91] C. Lee, W. Yang, R. Parr, *Phys. Rev. B* **1988**, *37*, 785.

[92] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.

[93] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.

[94] Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2009**, *5*, 324.

[95] J. Chaia, M. Head-Gordon *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

[96] J. Chaia, M. Head-Gordon, *J. Chem. Phys.* **2008**, *128*, 084106.

[97] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr, D. M Ferguson, D. C. Spellmeyer, T Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.

[98] W. Y. Chan, Z. Zhou, W. Yu, *Adv. Synth. Catal.* **2011**, *353*, 2999.

[99] C. Li, L. Wang, P. Li, W. Zhou, *Chem. Eur. J.* **2011**, *17*, 10208.

[100] Y. Wu, B. Li, F. Mao, X. Li, F. Y. Kwong, *Org. Lett.* **2011**, *13*, 3258.

[101] X. Zhao, C. S. Yeung, V. M. Dong, *J. Am. Chem. Soc.* **2010**, *132*, 5837.

[102] D. C. Powers, M. A. L. Geibel, J. E. M. N. Klein, T. Ritter, *J. Am. Chem. Soc.* **2009**, *131*, 17050.

[103] D. C. Powers, T. Ritter, *Nat. Chem.* **2009**, *1*, 302.

[104] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.

[105] W. R. Wadt, P. J. Hay, *J. Chem. Phys.* **1985**, *82*, 284.

[106] A. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6379.

[107] D. L. Davies, S. M. A. Donald, S. A. Macgregor, *J. Am. Chem. Soc.* **2005**, *127*, 13754.

[108] M. Lafrance, C. N. Rowley, T. K. Woo, K. Fagnou, *J. Am. Chem. Soc.* **2006**, *128*, 8754.

[109] S. I. Gorelsky, D. Lapointe, K. Fagnou, *J. Am. Chem. Soc.* **2008**, *130*, 10848.

[110] A. R. Katritzky, *Comprehensive Heterocyclic Chemistry III*, 1st ed., Elsevier: Amsterdam; New York (2008)

[111] A. F. Pozharskii, A. R. Katritzky, A. T. Soldatenkov,*Heterocycles in Life and Society: An Introduction to Heterocyclic Chemistry, Biochemistry, And Applications*, 2nd ed., Wiley, Chichester, West Sussex (2011)

[112] C. He, S. Guo, J. Ke, J. Hao, H. Xu, H. Chen, A. Lei, *J. Am. Chem. Soc.* **2012**, *134*, 5766.

[113] J. Ke, C. He, H. Liu, M. Lia, A. Lei, *Chem. Commun.* **2013**, *49*, 7549.

[114] M. Gao, C. He, H. Chen, R. Bai, B. Cheng, A. Lei, *Angew. Chem., Int. Ed.* **2013**, *52*, 6958.

[115] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.

[116] L. E. Roy, P. J. Hay, R. L. Martin, *J. Chem. Theory Comput.* **2008**, *4*, 1029.

[117] J. D. Cox, D. D. Wagman, V.A. Medvedev, CODATA Key Values for Thermodynamics, Hemisphere Publishing Corp., New York, **1984**, 1

[118] R. L. Hinman, E. B. Whipple, *J. Am. Chem. Soc.* **1962**, *84*, 2534.

[119] N. Otero, M. J. G. Moa, M. Mandado, R. A. Mosquera, *Chem. Phys. Lett.* **2006**, *428*, 249.

[120] P. A. Chase, G. C. Welch, T. Jurca, D. W. Stephan, *Angew. Chem., Int. Ed.* **2007**, *46*, 8050. Erratum: *Angew. Chem., Int. Ed.* **2007**, *46*, 9136.

[121] P. Spies, S. Schwendemann, S. Lange, G. Kehr, R. Fröhlich, G. Erker, *Angew. Chem., Int. Ed.* **2008**, *47*, 7543.

[122] V. Sumerin, F. Schulz, M. Atsumi, C. Wang, M. Nieger, M. Leskelä, T. Repo, P. Pyykkö, B. Rieger, *J. Am. Chem. Soc.* **2008**, *130*, 14117.

[123] H. Wang, R. Fröhlich, G. Kehr, G. Erker, *Chem. Commun.* **2008**, *45*, 5966.

[124] S. J. Geier, P. A. Chase D. W. Stephan, *Chem. Commun.* **2010**, *46*, 4884.

[125] J. Spielmann, F. Buch, S. Harder, *Angew. Chem., Int. Ed.* **2008**, *47*, 9434.

[126] K. Chernichenko,, Á. Madarász, I. Pápai, M. Nieger, M.a Leskelä, T. Repo, *Nat. Chem.* **2013**, *5*, 718.

[127] M. Lindqvist, N. Sarnela, V. Sumerin, K. Chernichenko, M. Leskelä, T. Repo, *Dalton Trans.* **2012**, *41*, 4310.

[128] T. A. Rokob, I. Pápai, *Top. Curr. Chem.* **2013**, *332*, 157.

[129] A. Hamza, A. Stirling, T. A. Rokob, I. Papai, *Int. J. Quantum Chem.* **209**, *109*, 2416.

[130] A. T. Rokob, I. Bakó, A. Stirling, A. Hamza, I. Pápai, *J. Am. Chem. Soc.* **2013**, *135*, 4425.

[131] M. Pu, T. Privalov, *Inorg. Chem.* **2014**, *53*, 4598.

[132] M. Pu, T. Privalov, *Int. J. Quantum Chem.* **2014**, *114*, 289.

[133] M. Pu, T. Privalov, *J. Chem. Phys.* **2013**, *138*, 154305.

[134] M. Pu, T. Privalov, *ChemPhysChem.* **2014**, *15*, 2936.

[135] M. Pu, T. Privalov, *ChemPhysChem.* **2014**, *15*, 3714.

[136] CP2K version 2.2.372; CP2K Developers Group, 2010. CP2K is freely available from http://www.cp2k.org/.

[137] S. Goedecker, M. Teter, J. Hutter, *Phys. Rev. B* **1996**, *54*, 1703.

[138] P. Spies, G. Erker, G. Kehr, K. Bergander, R. Fröhlich, S. Grimme, D. W. Stephan, *Chem. Commun.* **2007**, *47*, 5072.

[139] F. Bertini, V. Lyaskovskyy, B. J. J. Timmer, F. J. J. Kanter, M. Lutz, A. W. Ehlers, J. C. Slootweg, K. Lammertsma, *J. Am. Chem. Soc.* **2012**, *134*, 201.

[140] M. Boero, T. Ikeda, E. Ito, K. Terakura, *J. Am. Chem. Soc.* **2006**, *128*, 16798.

[141] M. De Vivo, M. Dal Peraro, M. L. Klein, *J. Am. Chem. Soc.* **2008**, *130*, 10955.

[142] E. Schreiner, N. N. Nair, D. Marx, *J. Am. Chem. Soc.* **2009**, *131*, 13668.

[143] A. Comas-Vives, A. Stirling, G. Ujaque, A. Lledós, *Chem. Eur. J.* **2010**, *16*, 8738.

[144] G. Santarossa, A. Vargas, M. Iannuzzi, A. Baiker, *Phys. Rev. B* **2010**, *81*, 174205.

[145] G. Kovács, A. Stirling, A. Lledós, G. Ujaque, *Chem. Eur. J.* **2012**, *18*, 5612.

[146] P. Dopieralski, J. Ribas-Arino, P. Anjukandi, M. Krupicka, J. Kiss, D. Marx, *Nat. Chem.* **2013**, *5*, 685.

[147] G. J. Martyna, M. E. Tuckerman, *J. Chem. Phys.* **1999**, *110*, 2810.

[148] H. Bekker, H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, H. Keegstra, B. Reitsma, M. K. R. Renardus, *Gromacs: A parallel computer for molecular dynamics simulations.* In Physics Computing 92 de Groot, R. A., Nadrchal, J., eds. World Scientific, Singapore, (1993)

[149] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comp. Phys. Comm.* **1995**, *91*, 43.

[150] E. Lindahl, B. Hess, D. van der Spoel, *J. Mol. Mod.* **2001**, *7*, 306.

[151] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comp. Chem.* **2005**, *26*, 1701.

[152] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435.

a doktori értekezés nyilvánosságra hozatalához

**I. A doktori értekezés adatai**
A szerző neve: **Daru János**
MTMT-azonosító: **10032746**
A doktori értekezés címe és alcíme:
**Theoretical investigations of reaction mechanisms – from methodological development to applications**
DOI-azonosító[39]: **10.15476/ELTE.2015.029**
A doktori iskola neve: **ELTE Kémia Doktori Iskola**
A doktori iskolán belüli doktori program neve:
**Elméleti és fizikai kémia, anyagszerkezetkutatás**
A témavezetők neve és tudományos fokozata:
**Stirling András**, az MTA doktora
**Tóth Gergely**, PhD

**II. Nyilatkozatok**
A doktori értekezés szerzőjeként[40]
a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom a Természettudományi Kar Tudományszervezési és Egyetemközi Kapcsolatok Osztályának ügyintézőjét, Bíro Évát, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.
b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[41]
c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (dátum)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;[42]
d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követőn egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.[43]

**2.** A doktori értekezés szerzőjeként kijelentem, hogy a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;

b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.

**3.** A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: 2014. február 11.

......................................................
a doktori értekezés szerzőjének aláírása

[38] Beiktatta az Egyetemi Doktori Szabályzat módosításáról szóló CXXXIX/2014. (VI. 30.) Szen. sz. határozat. Hatályos: 2014. VII.1. napjától.

[39] A kari hivatal ügyintézője tölti ki.

[40] A megfelelő szöveg aláhúzandó.

[41] A doktori értekezés benyújtásával egyidejűleg be kell adni a tudományági doktori tanácshoz a szabadalmi, illetőleg oltalmi bejelentést tanúsító okiratot és a nyilvánosságra hozatal elhalasztása iránti kérelmet.

[42] A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a minősített adatra vonatkozó közokiratot.

[43] A doktori értekezés benyújtásával egyidejűleg be kell nyújtani a mű kiadásáról szóló kiadói szerződést.