Calculations of Reaction Mechanisms and Entropic Effects in Enzyme Catalysis

MASOUD KAZEMI
Abstract

Ground state destabilization is a hypothesis to explain enzyme catalysis. The most popular interpretation of it is the entropic effect, which states that enzymes accelerate biochemical reactions by bringing the reactants to a favorable position and orientation and the entropy cost of this is compensated by enthalpy of binding. Once the enzyme-substrate complex is formed, the reaction could proceed with negligible entropy cost.

Deamination of cytidine catalyzed by *E.coli* cytidine deaminase appears to agree with this hypothesis. In this reaction, the chemical transformation occurs with a negligible entropy cost and the initial binding occurs with a large entropy penalty that is comparable to the entropic cost of the uncatalyzed reaction. Our calculations revealed that this reaction occurs with different mechanisms in the cytidine deaminase and water. The uncatalyzed reaction involves a concerted mechanism and the entropy cost of this reaction appears to be dominated by the reacting fragments and first solvation shell.

The catalyzed reaction occurs via a stepwise mechanism in which a hydroxide ion acts as the nucleophile. In the active site, the entropy cost of hydroxide ion formation is eliminated due to pre-organization of the active site. Hence, the entropic effect in this reaction is due to a pre-organized active site rather than ground state destabilization.

In the second part of this thesis, we investigated peptide bond formation and peptidyl-tRNA hydrolysis at the peptidyl transferase center of the ribosome. Peptidyl-tRNA hydrolysis occurs by nucleophilic attack of a water molecule on the ester carbon of peptidyl-tRNA. Our calculations showed that this reaction proceeds via a base catalyzed mechanism where the A76 O2' is the general base and activates the nucleophilic water.

Peptide bond formation occurs by nucleophilic attack of the α-amino group of aminoacyl-tRNA on the ester carbon of peptidyl-tRNA. For this reaction we investigated two mechanisms: i) the previously proposed proton shuttle mechanism which involves a zwitterionic tetrahedral intermediate, and ii) a general base mechanism that proceeds via a negatively charged tetrahedral intermediate. Although both mechanisms resulted in reasonable activation energies, only the proton shuttle mechanism found to be consistent with the pH dependence of peptide bond formation.

Keywords: Enzyme catalysis, Entropy, Cytidine deamination, Ribosome, Peptidyl-tRNA hydrolysis, Peptide bond formation, Empirical valence bond method, Density functional theory

Masoud Kazemi, Department of Cell and Molecular Biology, Computational Biology and Bioinformatics, Box 596, Uppsala University, SE-751 24 Uppsala, Sweden.

© Masoud Kazemi 2017

ISSN 1651-6214
urn:nbn:se:uu:diva-316497 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-316497)
Dedicated to Mehdi, Mahrokh, and Mahsa
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>Enzyme-substrate complex</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>Turnover number</td>
</tr>
<tr>
<td>$\Delta G$</td>
<td>Gibbs free energy difference</td>
</tr>
<tr>
<td>RRHO</td>
<td>Rigid rotor harmonic oscillator</td>
</tr>
<tr>
<td>EVB</td>
<td>Empirical valence bond</td>
</tr>
<tr>
<td>CDA</td>
<td>Cytidine deaminase</td>
</tr>
<tr>
<td>DFT</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>TI</td>
<td>Tetrahedral intermediate</td>
</tr>
<tr>
<td>PTC</td>
<td>Peptidyl transferase center</td>
</tr>
<tr>
<td>KIE</td>
<td>Kinetic isotope effect</td>
</tr>
<tr>
<td>KSIE</td>
<td>Kinetic solvent isotope effect</td>
</tr>
<tr>
<td>ZPE</td>
<td>Zero point energy</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamic</td>
</tr>
<tr>
<td>FEP</td>
<td>Free energy perturbation</td>
</tr>
</tbody>
</table>
Introduction

Life from a chemical point of view

Biochemical reactions govern every aspect of living organisms (Figure 1). The metabolic pathways meet the energy demands of various processes in cells and provide the building blocks that constitute the cellular structures. The flow of genetic information is also, in essence, a series of biochemical reactions in which genetic information is either transferred to the next generation—replication—or used to maintain the normal operations of organisms by transcription and translation\textsuperscript{1,3}

![Figure 1. Chemical reactions in different compartments of the cell.](image-url)

The cells’ interaction with their environment is another example where an external stimulus triggers cascades of biochemical reactions and creates a
response inside the cell. Many of these responses are either biochemical reactions or are dependent on these processes. For the living organisms to function properly, these reactions must occur rapidly. However, under physiological conditions many molecules are too stable to participate in chemical reactions sufficiently fast\textsuperscript{1-3}.

To remedy this, nearly all chemical reactions in the living organisms are catalyzed by enzymes. While most enzymes are proteins, catalytic RNAs (ribozymes) also play a crucial role in the processes that are connected with the flow of genetic information. Regardless of their nature, all enzymes contain an active site pocket where the chemical reaction occurs (Figure 2). The active site is made of chemical groups that interact with the substrate via weak interactions during the initial binding process. The specificity of an enzyme for a particular substrate is dictated by these interactions. That is, the correct substrate is able to maximize these interactions which result in effective binding. This is important for catalysis under biological conditions, since enzymes need to recognize and bind the correct substrate among large numbers of molecules in the cell. In the active site, catalytic groups participate directly in the chemical reaction. For instance, in the case of serine proteases, a serine residue acts as a nucleophile and forms a covalent bond to the substrate\textsuperscript{2-4}.

![Figure 2. Cytidine deaminase in complex with a substrate analog. The active site cavity is highlighted.](image)

The role of enzymes is not limited to the catalysis of biochemical reactions; they also provide high level control mechanisms to maintain cell homeostasis. This is achieved by regulation of enzyme concentration which in turn controls the chemical reactions. In addition to this, the activity of many enzymes is also modulated by chemical compounds that bind to them. Inhibitors are compounds that are similar to normal substrates but cannot undergo chemical reactions. These molecules bind to the active site and block the enzyme activity. Another mode of control is provided by allosteric compounds that regulate the enzyme activity by inducing conformational change.
These molecules bind to the allosteric sites that are normally different from the active site. Moreover, many enzymes are sequestered in specific locations inside the cell. In this way, the chemical reactions that are catalyzed by these enzymes occur only in the designated locations. For example, the respiratory enzymes, which are involved in electron transfer reactions, are mainly localized in the mitochondria to protect cells from free radical species that are produced as by-products.2-4

Enzyme action

Kinetic characterization

To understand enzyme catalysis, we require a way of quantifying enzyme activity. Since enzymes catalyze chemical reactions, kinetics provides the most natural description. An enzyme catalyzed reaction—in its simplest form—can be described as a two-step process.

\[
\begin{align*}
E + S & \rightleftharpoons_{k_i}^{k_{-i}} ES \\
& \rightarrow_{k_2} E + P
\end{align*}
\]

An enzyme-substrate (ES) complex forms in the first step, and, in the second step, the ES complex breaks down to product (P) and free enzyme (E). The kinetics of this reaction is then expressed in a system of coupled differential equations as

\[
\begin{align*}
\frac{d[S]}{dt} &= -k_4[S][E] + k_1[ES] \\
\frac{d[E]}{dt} &= -k_4[S][E] + (k_{-1} + k_2)[ES] \\
\frac{d[ES]}{dt} &= k_1[S][E] - (k_{-1} + k_2)[ES] \\
\frac{d[P]}{dt} &= k_2[ES]
\end{align*}
\]

where square brackets represent the concentration of species involved in the reaction, and \(k_i\) is the rate constant. For many practical cases, the ES formation is a fast reversible step compared with the chemical transformation \((k_1 \sim k_{-1} \gg k_2)\). Using this postulate and taking into account that—under physiological conditions—the enzyme concentration is lower than the substrate, the numerical solutions of differential equations 1 to 4 show three distinct regimes for this reaction (Figure 3).
Figure 3. Numerical simulation of a two-step enzymatic reaction in which the substrate binding is a fast reversible process and the chemical transformation is a slow irreversible step.

During the initial phase, substrate binding to the free enzymes increases ES concentration rapidly (Figure 3). Subsequently, ES formation reaches equilibrium. Hence, the ES concentration remains constant (steady state) while the concentration of substrate and product change linearly. As substrate depletes, the system can no longer maintain the steady state and ES concentration decays. The steady state quasi-equilibrium approximation, results in the simplification of differential equations 1 to 4 and the product formation rate—turnover rate ($V$)—can now be expressed in the closed form as

$$V = \frac{d[P]}{dt} = k_2[ES] = \frac{k_2[E_0][S]}{k_{-1} + k_2 \frac{k_1}{[E_0]}} + [S]$$  \hspace{1cm} (5)$$

where $[E_0]$ is the total amount of enzyme ($[E_0] = [E] + [ES]$). The term $k_2[E_0]$ is the maximum velocity ($V_{max}$) and $(k_{-1} + k_2)/k_1$ corresponds to the Michaelis constant ($K_M$). Using these abbreviations we can rewrite equation 5 as

$$V = \frac{V_{max} [S]}{K_M + [S]}$$  \hspace{1cm} (6)$$

This rate equation, which is named after Leonor Michaelis and Maud Menten as the Michaelis-Menten equation, expresses the reaction rate as a function of the substrate concentration. The constants $V_{max}$ and $K_M$ are obtained by measuring the enzymatic reaction rate at different concentrations of substrate, and fitting the measured rates to the equation 6. The parameter $K_M$ is an equilibrium constant that describes the quasi-equilibrium between the reactants (E+S) and enzyme-substrate complex (ES). In the case of our sim-
ple reaction, if the chemical transformation rate is much slower than the rate of substrate dissociation \( k_2 \ll k_{-1} \), \( K_M \) would reduce to the dissociation constant \( (K_d) \). However, for more complex reactions with multiple intermediate states, this interpretation is not applicable as other rate constants may contribute to \( K_M \).

The maximum reaction rate \( (V_{\text{max}}) \) is achieved when all enzyme molecules are in complex with the substrate \([S] \gg [E]_0\). In this condition, for a given enzyme concentration, the reaction rate is then restricted by the slowest reaction step \( (k_2 \text{ in our reaction}) \) which is conventionally called turnover number \( (k_{\text{cat}} \text{ in s}^{-1}) \). Under physiological conditions, however, the substrate concentration is normally lower than \( K_M \). This, in turn, implies that the product formation rate depends also on the substrate concentration as

\[
\frac{d[P]}{dt} = \frac{k_{\text{cat}}}{K_M} [S][E] \tag{7}
\]

The second order rate constant \( k_{\text{cat}}/K_M \text{ (units } M^{-1}s^{-1}) \) is the catalytic efficiency of enzymes and is relevant to enzyme activity under physiological conditions\(^2\text{-}^4\).

**Thermodynamic characterization**

The experimentally measured kinetic parameters—\( K_M \) and \( k_{\text{cat}} \)—can be expressed in terms of thermodynamic potentials, namely the Gibbs free energy. For the enzyme-substrate dissociation, the change in Gibbs free energy is given by

\[
\Delta G = \Delta G^\circ + RT\ln\left(\frac{[S][E][ES]}{C^0/C^0/C^0}\right) \tag{8}
\]

where \( \Delta G^\circ \) is the standard Gibbs free energy change, and \( C^0 \) is the standard state concentration. The change in Gibbs free energy is zero at equilibrium, and if we assume that the equilibrium is achieved rapidly, we can express \( K_M \) in terms of \( \Delta G^\circ \) as

\[
K_M = e^{-\frac{\Delta G^\circ}{RT}} \tag{9}
\]

where \( R \) is the gas constant, and \( T \) is the absolute temperature. The \( k_{\text{cat}} \) is also related to the change of Gibbs free energy based on the transition state theory \( \text{(TST)} \text{\(^6\text{-}^7\)} \) by

\[
k_{\text{cat}} = \kappa \frac{k_BT}{h} e^{-\frac{-\Delta G^\dagger}{RT}} \tag{10}
\]

where \( \kappa \leq 1 \) is the transmission coefficient, \( k_B \text{ is the Boltzmann constant, } h \) is the Plank constant, and \( \Delta G^\dagger \) is the activation free energy. The TST as-
sumes quasi-equilibrium between the transition and reactant states. The exponent on the right-hand side of equation 10 is the equilibrium constant corresponding to this assumption and the transmission coefficient takes into account the probability that a reactive trajectory in the transition state (TS) returns to the reactant state. In biochemical reactions, the kinetic energy of the system is often well below the activation barrier which implies that the transmission coefficient can be taken to be unity. Hence, the TST rate is the upper limit for the absolute rate. Using these relations, it is then possible to analyze chemical reactions in terms of Gibbs free energy change (Figure 4).

![Figure 4. Free energy profile of a hypothetical catalyzed and uncatalyzed reaction.](image)

The first step in an enzyme catalyzed reaction is the substrate binding to the enzyme active site. The free energy change of this process ($\Delta G_{bind}$) is approximately related to the inverse of $K_M$ according to equation 9. The $\Delta G_{bind}$ of enzymes is normally negative. This implies that enzymes can bind effectively to the correct substrate. After the ES complex formation, the system must climb the activation barrier ($\Delta G_{cat}^{\ddagger}$) to undergo the chemical transformation. This free energy cost is related to the $k_{cat}$ according to equation 10. The catalytic efficiency of enzymes ($k_{cat}/K_M$) is the overall free energy penalty that the system must pay to reach to the TS from the reactant state ($\Delta G^{\ddagger} = \Delta G_{bind}^{\ddagger} + G_{cat}^{\ddagger}$). The equilibrium between the reactant and product concentrations is described in terms of the reaction free energy ($\Delta G_r$) using equation 9. Since enzymes have no effect on the final equilibrium, the $\Delta G_r$ of the enzyme catalyzed reaction and the corresponding solution reaction are identical (Figure 4).

To be able to quantify enzyme catalysis, we need to compare the catalyzed reaction to a reference reaction. The reference reaction is identical to the enzymatic reaction, except that it occurs in water (Figure 4).
reference reaction, we further introduce the concept of reaction cage which is important for multi-component reactions. The reaction cage is the state in which the reactants are in proximity of each other. This state corresponds to bringing the reactants from 1M standard state concentration to an effective \(~55M\) concentration (here we assumed that the molar volume of the reacting functional groups of the reactants to be the same as that of water). In the case of bimolecular reactions, the free energy change of reaching this state \((\Delta G_{\text{cage}})\) can be calculated by fixing one of the reactants and confining the other molecule from the initial volume \((v_1)\) to the cage volume \((v_2)\) around our fixed molecule, \(\Delta G_{\text{cage}} = -RT\ln(v_2/v_1)\). If the reacting groups have molar volumes similar to that of water, the \(\Delta G_{\text{cage}}\) is approximately \(2.4\) kcal mol\(^{-1}\). The cage activation free energy \((\Delta G_{\text{cage}}^\ddagger)\) is the free energy cost of reaching the TS from the state which the reactants are already in the reaction cage (Figure 4). The overall activation free energy \((\Delta G^\ddagger)\) of the uncatalyzed reaction, hence, is the sum of the concentration dependent contribution \((\Delta G_{\text{cage}})\) and the cage activation free energy \((\Delta G_{\text{cage}}^\ddagger)\). This treatment of the reference reaction provides a way to compare different steps in the enzymatic reaction to the corresponding ones in a reference reaction\(^8\).

**Enzyme catalysis**

In the previous sections, we introduced the tools for quantifying enzyme catalysis. The catalytic efficiency \((k_{\text{cat}}/K_M)\) of enzymes\(^9\) on average are \(10^7\) M\(^{-1}\) s\(^{-1}\), while the rate constant of the corresponding uncatalyzed reactions\(^9\) are in the range \(10^{-9}-10^{-16}\) s\(^{-1}\). How do enzymes achieve such rate accelerations?

**Chemical effect**

Enzymes increase the rate of biochemical reactions partly by altering the reaction mechanisms. The amino acids that form the active site provide functional groups that may not be available to the uncatalyzed reaction. This in turn can promote mechanisms that are energetically more favorable\(^4\). Serine protease enzymes are an example of such a process. In these enzymes, peptide bond cleavage first occurs by nucleophilic attack of an activated serine residue, whereas, in the uncatalyzed reaction, a water molecule acts as the nucleophile. The serine residue is activated by donating one proton to a nearby histidine residue that is stabilized electrostatically by a negatively charged aspartic acid. In this case, the histidine residue acts as a general base. Many enzymes, in addition to the naturally occurring amino acids, also employ a variety of organic and inorganic compounds (cofactors) to increase their support for different chemical reactions. For instance, metal proteases
use a zinc cation to activate a nucleophilic water molecule during peptide hydrolysis reaction. The ability of enzymes in providing more efficient catalytic groups is the chemical effect in enzyme catalysis\textsuperscript{2,4}.

**Active site effect**

Enzyme catalysis is not limited only to clever chemistry and many enzymatic reactions occur more rapidly than the corresponding uncatalyzed ones despite an identical mechanism. For instance, the triple mutant of bacterial serine protease, where the catalytic residues are mutated to alanine, retains a residual catalysis and is able to accelerate peptide bond hydrolysis $10^3$ times faster than the uncatalyzed reaction\textsuperscript{10}. The pioneering hypothesis in explaining these effects is Pauling’s TS stabilization theory\textsuperscript{11} which asserts that enzymes are evolved to bind tightly to the TS rather than the reactant or product. The high affinity of the enzyme active site for the TS then results in lowering the activation energy. Since TSs are short lived species, they cannot be observed directly. However, the inhibition of enzymes by TS analogues, that are designed to resemble the TS, provide a proof of this concept\textsuperscript{12}.

Although TS stabilization provide a simple framework for describing enzyme catalysis, the exact mechanisms by which enzymes achieve such stabilization is not addressed.

*Figure 5. A pre-organized active site stabilizes the TS effectively without the reorganization energy penalty.*

Early computer simulations revealed that the electrostatic interactions are the most important factor in enzyme catalysis\textsuperscript{13-15}. This implies that the enzymes active site is better in solvating the TS than the bulk water. Although solution phase provides more interaction partners compared with the enzymes active site, the waters in the first solvation shell are not able to optimally
align themselves relative to the TS\textsuperscript{15}. That is, the orientation of waters in the first solvation shell is restricted due to interactions with the bulk solvent (Figure 5). In contrast, the enzyme active site is evolved to effectively stabilize the TS and the pre-organized structure\textsuperscript{16} of the active site provides optimal electrostatic interactions for this purpose. In addition, the enzyme active site minimizes the reorganization energy cost of the environment. During a chemical reaction the charge distribution of the reacting fragments change as they move along the reaction coordinate. In the case of uncatalyzed reactions, this means that the water molecules have to orient themselves differently in the reactant state and TS. This reorganization is kinetically unfavorable since the existing interactions have to be broken prior to formation of the new ones. In the enzymes active site, the reorganization energy cost is paid during folding process; hence, the chemical reaction can proceed without this energy cost. The preorganization concept, proposed initially by Warshel\textsuperscript{16}, offers a general explanation for the active site effect on enzyme catalysis and reveals how enzymes are able to stabilize the TS.

Alternative views
So far, the picture that has been depicted of enzyme catalysis is the author’s view on the matter. Nonetheless, many alternative suggestions have been put forward to explain the effect of an enzyme active site on their catalysis. Ground state destabilization is an alternative hypothesis that asserts the catalytic effect of enzymes is achieved by destabilization of the reactant-enzyme complex\textsuperscript{17}. It is hypothesized that this destabilization can occur in a number of ways, such as desolvation in which the reactant is placed in an electrostatically unfavorable environment or conformational strain which suggests that the enzyme active site distorts the equilibrium geometry of the substrate. Although ground state destabilization may affect \( k_{cat} \), the enzyme efficiency \( (k_{cat}/K_M) \) would not be affected by such processes (Figure 4). \( k_{cat}/K_M \) is the only kinetic parameter that is relevant to enzymes that operate under physiological conditions, hence, it is reasonable to assume that enzymes are evolved to optimize this parameter. Therefore the destabilization hypothesis provides no evolutionary advantage for enzyme catalysis. The entropy effect\textsuperscript{17}, which suggests enzymes operate by reducing the entropy penalty of chemical reactions, is also one of the popular hypotheses. This hypothesis will be analyzed more extensively in the next chapter.

Alternatively, quantum tunneling\textsuperscript{18} and enzyme dynamics\textsuperscript{19} were invoked to explain enzyme catalysis. Although, the quantum tunneling effect is present in a number of enzymes, for instance hydrogen transfer reactions, the magnitude of this effect is likely to be similar in the reference and enzyme catalyzed reactions\textsuperscript{20}. Therefore this effect cannot explain enzyme catalysis in general. The enzyme dynamics does not appear to play a major role in catalysis either. That is, according to this hypothesis, certain enzyme mo-
tions are coupled to the reaction coordinate and cannot be treated by TST which assumes that these motions are at equilibrium. This suggestion then implies that $\kappa$ in the TST should not be unity, since $\kappa$ encapsulates any deviations from equilibrium statistics and quantum effects. However, multiple computational studies showed that the values of $\kappa$ are similar in both the solution and enzyme catalyzed reactions\textsuperscript{21-23}. 
Entropy and enzyme catalysis

One of the early attempts to quantitatively explain the origin of enzyme catalysis is the comparison between intra- and intermolecular reactions\textsuperscript{17,24}. Intramolecular reactions often occur faster than their intermolecular counterparts. For instance, formation of succinic anhydride from succinate half esters is $\sim 10^5 \text{ M} (~7 \text{ kcal mol}^{-1})$ more favorable than the corresponding intermolecular reaction of acetate with acetate esters\textsuperscript{24,25} in a 1M standard state.

The higher rate of the intramolecular reaction was attributed to the entropic advantage\textsuperscript{17}. That is, the reactants in the intramolecular reaction are already in a favorable position; therefore, there is no entropy penalty due to the loss of the translational degrees of freedom. Imposing restraints on the reactants’ rotational degrees of freedom was shown to further increase the reaction rate by a factor of $\sim 10^3 \text{ M} (~4 \text{ kcal mol}^{-1})$.

Hence, it was suggested that eliminating the translational and rotational entropy penalties could yield up to $10^8 \text{ M} (~11 \text{ kcal mol}^{-1})$ rate acceleration\textsuperscript{17}. Based on these experiments, Jencks proposed\textsuperscript{17} that one of the ways by which enzymes enhance the reaction rate is by bringing the reacting fragments into a favorable position and orientation. The entropy penalty of this positioning and alignment is hypothesized to be compensated by a favorable enthalpy of binding. Once the Michaelis complex (ES) is formed, the reaction can then proceed with almost negligible entropy cost (Figure 6).
Figure 6. Schematic representation of an idealized enzymatic reaction based on the entropy effect.

Therefore, for a given reaction (Figure 6), the enthalpy of activation should be similar in the catalyzed and uncatalyzed reactions (enthalpy contribution to $k_{\text{non}}$ and $k_{\text{cat}}$) and the activation entropy of the uncatalyzed reaction should be similar to the entropy penalty of the initial binding process in the catalyzed reaction (entropy contribution to $k_{\text{non}}$ and $K_M$).

A critical assumption in this hypothesis is that the observed rate acceleration in the intramolecular reaction is attributed solely to the entropic effects. This assumption, however, is disputed and computational studies suggested that the higher rate of the intramolecular reactions is related to the activation enthalpy rather than entropy\textsuperscript{26}. Hence, intramolecular reactions are not necessarily a reasonable model for enzyme catalysis. In fact, the temperature dependence of a series of catalyzed and uncatalyzed reactions shows deviations from Jencks predictions\textsuperscript{27,28}. For the tested reactions, the rate enhancement was consistently found to be achieved by reducing activation enthalpy. In contrast, the activation entropy—entropy contribution to $k_{\text{cat}}$—shows an irregular behavior and is raised by enzymes for some reactions and lowered in other cases\textsuperscript{28}.

This evidence suggests that Jencks entropy argument does not provide a general explanation for the effect of entropy in enzyme catalysis. Nevertheless, the role of entropy cannot be ignored, since the experimental observations suggest that the entropy contribution to the activation free energy is different in the catalyzed and uncatalyzed reactions.

Calculating entropy

To study the effect of entropy in enzyme catalysis and interpret the experimental results, we should be able to calculate the entropy change of the whole system, \textit{i.e.}, the reacting fragments and the environment—solvent or
enzyme. One way to approach this is by using approximate analytical functions for the partition functions of the reactant and TS which can be related to the change in entropy by

\[ \Delta S_{i\rightarrow f} = k_B \ln \frac{Q_f}{Q_i} + k_B T \frac{\partial}{\partial T} \left( \ln \frac{Q_f}{Q_i} \right) \]  

(11)

where \( Q_i \) and \( Q_f \) are the partition functions of initial and final states respectively. The partition function for a given state can be obtained by the rigid rotor harmonic oscillator (RRHO) approximation as is customary in quantum chemistry. In this treatment, the translation, rotation and vibrations degrees of freedom are assumed to be uncoupled which implies that the total partition function for a given state can be calculated as

\[ Q = Q_t Q_r Q_v Q_e \]  

(12)

where subscripts \( t, r, v, \) and \( e \) indicate, respectively, translation, rotation, vibration and electronic contributions to the partition function. Although the RRHO approximation can be used for obtaining thermodynamic parameters for systems with a dominant minimum in the potential energy, for instance gas phase reactions, extending this method to biologically relevant systems is not trivial. The potential energy for these systems is a complex hyper-surface with multiple minimua that may contribute to a given state (Figure 7).

\[ Figure 7. \text{In complex systems, multiple minima may contribute to the partition function of a given state.} \]

Therefore, to obtain an accurate estimate of entropy, the contribution of all thermally accessible minima to the partition function must be taken into account. However, characterizing all minima is a challenging task due to the high dimensionality of the underlying potential energy surface.

The empirical valence bond (EVB) method is an approach that offers extensive sampling of the potential energy surface. The reason for this is that the potential energy function of EVB is computationally inexpensive. In the EVB framework, each point along the reaction coordinate is a linear combination of diabatic bases that are described by analytical force fields (see
Methods summary). This offers a computationally cheap potential function that could be used for extensive sampling of the chemical reactions. Based on this method, a potential of mean force is constructed ($\Delta G(\chi)$) as a function of the reaction coordinate. From $\Delta G(\chi)$ the activation and reaction free energy can readily be calculated. The entropy contribution to the activation free energy, in analogy to the experimental approach, can be obtained from the temperature dependency of the activation free energy. That is, by repeating the simulations at different temperatures and plotting the $\Delta G^\ddagger / T$ versus $1/T$, one obtains the Arrhenius plot from which the activation parameters—enthalpy and entropy—can be extracted by linear fitting to the equation

$$\frac{\Delta G^\ddagger}{T} = \frac{\Delta H^\ddagger}{T} - \Delta S^\ddagger$$

(13)

Here it is assumed that the change in the heat capacity ($\Delta C_p$) is constant over the temperature interval. Therefore, the temperature dependency of enthalpy is canceled and the observed temperature dependency of the activation free energy can be attributed to only the entropy.

Deamination of cytidine, a prime example

Despite the evidence against Jencks entropy hypothesis, this argument is still invoked to rationalize enzyme catalysis. This is partly due to the fact that some chemical reactions appear to behave exactly according to Jencks entropy argument.

![Figure 8. Hydrolytic deamination of cytidine to uridine.](image)

Deamination of cytidine is a prime example of such reactions. In this reaction, hydrolytic deamination of cytidine converts this nucleotide to uridine (Figure 8). Deamination of cytidine$^{29}$ can occur spontaneously ($k_{non} = 2.7 \times 10^{-10} \text{ s}^{-1}$) which results in DNA instability$^{30}$. *E. coli* cytidine deaminase$^{31}$ (CDA) is one of the enzymes that also catalyzes this reaction ($k_{cat}/K_M = 2.8 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$). In general, cytidine deaminases are of clinical interest due to their ability in degrading cytidine-based anti-tumor drugs. CDA also provides a unique opportunity to test the role of entropy in enzyme catalysis since the thermodynamic parameters of this reaction have been experimentally determined$^{30}$ for both catalyzed and uncatalyzed reactions (Figure 9).
The thermodynamic parameters of cytidine deamination in solution (blue) and in E. coli cytidine deaminase (red) at 25 °C.

For the catalyzed reaction, the entropy contribution to $k_{cat}$ is negligible ($T\Delta S^\ddagger = 0.9$ kcal mol$^{-1}$) while the entropy penalty for binding is $T\Delta S^0 = -7.6$ kcal mol$^{-1}$. This entropy penalty matches the entropy contribution to the activation free energy of the uncatalyzed reaction ($T\Delta S^\ddagger = -8.3$ kcal mol$^{-1}$). This is exactly the behavior one would expect if Jencks entropy hypothesis were in effect (Figure 6). To understand the origin of this entropy change, we studied the deamination of cytidine in solution (paper I) and CDA (paper II) using density functional theory (DFT) and EVB methods.

Spontaneous deamination of cytidine

Deamination of cytidine in solution occurs by formation of a tetrahedral intermediate (TI) at neutral pH (Figure 10). In this reaction, formation of the tetrahedral intermediate (TI) is the rate limiting step which could either occur via a concerted or a stepwise mechanism.

In the concerted mechanism, nucleophile attack of a water molecule on C4 of cytidine occurs simultaneously with proton transfer to N3. This proton
transfer could be facilitated by one or two waters resulting in a six- or an eight-membered ring TS respectively. In the stepwise mechanism, on the other hand, the attacking water molecule initially protonates N3 of cytidine and, subsequently, the TI is formed by nucleophilic attack of the resulting hydroxide ion on C4. Each of these TSs was studied by DFT method and our calculations showed that the concerted eight-membered and the stepwise mechanisms result in a similar barrier of approximately 30 kcal mol⁻¹.

For each of these mechanisms, we then calculated the thermodynamic activation parameters—entropy and enthalpy—by constructing the Arrhenius plot as described above. The EVB Hamiltonian, for all mechanisms, was parameterized to the experimental activation free energy (Δ‡ GLES = 30.4 kcal mol⁻¹). Remarkably, each mechanism resulted in a different enthalpy-entropy partitioning. In these calculations, the concerted mechanism with the eight-membered ring TS yielded the activation entropy that was consistent with the experimental value. Hence, we concluded that the spontaneous deamination of cytidine occurs via this mechanism. Parameterization of EVB to the DFT calculations resulted in the same conclusion.

It is interesting to explain the origin of enthalpy-entropy partitioning in our calculations, since this provides an intuitive way of understanding the entropic effects. The entropy is related to the curvature of the potential energy functions describing the reactant and TSs in each mechanism. This could be illustrated by a simple harmonic oscillator (SHO), for which the partition function can be evaluated analytically. For a SHO with a potential function of the form \( V(x) = E_0 + k(x - x_0)^2 \), the classical configurational partition function is given by

\[
Q = \mathcal{N} \left( \frac{2\pi \beta k}{\mathcal{N}} \right)^{1/2} e^{-\beta E_0}
\]

where \( \mathcal{N} \) is a constant to make the partition function unitless, \( \beta = 1/k_B T \). If we assume, for a given reaction, that the initial and final states are described by two SHOs, the change in entropy can then be calculated as

\[
\Delta S^\ddagger = k_B \ln \left( \frac{k_i}{k_f} \right)^{1/2}
\]

where the \( k_i \) and \( k_f \) are the curvatures of the potential functions describing the initial and final states respectively. Hence, if \( (k_i/k_f) > 1 \), the chemical transformation has a favorable entropy contribution, and if \( (k_i/k_f) < 1 \), the entropy contribution would be unfavorable. Although the SHO is a crude simplification of the real systems which often deviate from harmonicity, this model provides a simple framework to rationalize the results of our entropy calculations.

According to our calculations, the initial step in the stepwise reaction—proton transfer from water to cytidine—is associated with large unfavorable
entropy. In this reaction, the system moves from a weakly interacting neutral state to a charged state with stronger electrostatic interaction. Therefore, this reaction can be thought of as a transformation in which $k_i$ is less than $k_f$. The entropy penalty for this reaction is dominated by the solvent since they interact more strongly with the localized charges in the product state than reactant state. The same logic can be applied to the second step of this reaction, the nucleophilic attack hydroxide ion on the protonated cytidine. The entropy contribution to this step is favorable because the charges of the reacting fragments delocalize as the reaction proceeds toward TI. Therefore, for this step, $k_i$ is greater than $k_f$. The entropy effect of this step is also dominated by the solvent rather than reacting fragments. The six- and eight-membered ring concerted mechanisms show a different entropic effect compared to that of the stepwise mechanism. In these mechanisms, the entropy contributions are unfavorable and are dominated by the reacting fragments as is evident from the correlation between the entropy penalty and the number of water molecules involved in the TS. In these cases, the solvent entropies are probably not as dominant since the charge distribution of the reactant and TS are more similar.

Overall, these results suggest that Jencks entropy argument is a special case of a bigger picture. For reactions that are not associated with appreciable charge redistribution the entropy change, as predicted by Jencks entropy argument, is dominated by the reacting fragments. The solvent contribution to the entropy change, however, becomes a dominating factor for the reactions that involve charge separation.

Cytidine deamination by cytidine deaminase

*E. coli* cytidine deaminase (EC 3.5.4.5) converts both cytidine and 2'-deoxycytidine to uridine and 2'-deoxouridine respectively. The zinc cation coordinates with three residues of the enzyme (Cys129, Cys132 and His102) and a water molecule in the fourth position (Figure 11). In the active site, Glu104 interacts with both the zinc bound water and substrate. Mutating this residue severely reduces the enzyme activity which indicates it is involved in chemical steps. Although this reaction—similar to the spontaneous deamination—occurs via a TI, its mechanism is different. In the CDA catalyzed reaction, TI formation occurs by nucleophilic attack of a hydroxide ion on C4 of the protonated cytidine (Figure 12, TS2). The hydroxide ion is produced by proton transfer from the catalytic water to Glu104 and stabilized by electrostatic interaction with the zinc cation.
Figure 11. The active site of *E.coli* cytidine deaminase (PDB 1ALN) with substrate analogue (orange). The active site position in the dimer is shown in inset.

Subsequently, the TI undergoes another proton transfer, which is mediated by Glu104 (Figure 12, TS3 and TS4). This produces a zwitterionic intermediate that collapses to uridine and ammonia (Figure 12, TS5). In this mechanism, TS2 and TS5 result in similar activation barriers of 15.3 and 14.7 kcal mol$^{-1}$. Hence, according to our calculations, deamination of cytidine by CDA occurs with a different mechanism compared to that of the solution reaction.

Figure 12. Schematic representation of cytidine deamination catalyzed by *E.coli* cytidine deaminase. The figure is adapted from reference 38.

In the next step, we proceeded to calculate the entropy contribution to the activation free energy following the same strategy as outlined above. The entropy contribution to the TS of TI formation (TS2) was slightly unfavorable. This effect is due to entropic cost of the bimolecular reaction and higher affinity of the active site for TI. In fact, all the intermediates with the tetrahedral geometry showed unfavorable entropies. Such observations are consistent with the high binding affinity of the transition-state analog for the active site. Collapse of the zwitterionic intermediate, on the other hand, has
slightly favorable entropy as one expects from a unimolecular decomposition reaction. Overall, these observations suggested that the entropy contribution to the catalyzed reaction is dominated by the configurational entropy of the reacting fragment\(^{38}\).

The picture that emerges from these results, at least in the case of CDA, does not support Jenck entropy argument for enzyme catalysis. Since the spontaneous and CDA catalyzed reactions occur with different mechanisms, a direct comparison of these reactions is misleading. In solution, the concerted mechanism is favorable because the solvent entropy penalty of the stepwise pathway is prohibitive. In the enzyme, on the other hand, the reaction proceeds via the stepwise mechanism. This mechanism is promoted by the enzyme active site, since the entropy penalty of solvent reorganization is eliminated. In the active site, the key residues are pre-organized in a favorable orientation to stabilize the TS; therefore, the stepwise reaction can proceed without reorganization entropy cost. This conclusion supports Warshel’s preorganized active site view\(^{16}\).
Ribosome, a catalytic RNA

Ribosomes are ribonucleoprotein complexes that are involved in protein synthesis. The prokaryotic ribosomes consist of two subunits, 50S and 30S (Figure 13).

![Figure 13. The structure of the prokaryotic 70S ribosome\(^{39}\) (PDB code 4V5J) in complex with RF2 in the A-site.]

These two subunits bind together and form the 70S complex that sandwiches the messenger RNA molecule\(^{40}\). The 70S complex is composed of three tRNA binding sites (Figure 13). The A-site is the position where the new amino acyl-tRNA enters the complex while the P-site binds to the peptidyl-tRNA, and the E-site is the exit position. During the elongation phase of protein synthesis, the sequence of the nascent peptide chain is determined by Watson-Crick interactions between mRNA and A-site tRNA in the codon.
recognition site (Figure 13). Once the correct tRNA is recognized, the nascent peptide chain in the P-site forms a peptide bond with aminoacyl-tRNA in the A-site and the 70S complex slides to the next codon on the mRNA. This pushes the P-site tRNA to the E-site and peptidyl-tRNA from A- to P-site. Hence, at the end of each cycle, the A-site is empty to accept a new aminoacyl-tRNA. Termination of protein synthesis is marked by stop codons in the mRNA sequence. These codons are recognized by the release factors (RFs) that bind to the A-site and trigger the hydrolysis of the ester bond between the nascent peptide and the P-site tRNA. Both peptide bond formation and peptidyl-tRNA hydrolysis occur at the peptidyl transferase center (PTC) of the ribosome. The PTC is located in the 50S subunit and is entirely made of ribonucleotides (Figure 14).

**Figure 14.** The peptidyl transferase center of the 50S subunit during peptide bond formation (A) and peptidyl-tRNA hydrolysis (B). The conserved nucleotides are colored in yellow.

The nucleotides C2063, A2451, U2503, and U2585 compose the first solvation shell of the PTC. During the elongation phase, the acceptor arms of the A- and P-site tRNAs are extended into the PTC. This positions the reacting fragments—peptidyl-tRNA and aminoacyl-tRNA—in close proximity. During the termination phase, on the other hand, the A-site is occupied by the RFs that insert a universally conserved methylated Gln residue in the PTC. This Gln residue triggers peptidyl-tRNA hydrolysis resulting in the release of the peptide chain from the P-site tRNA (Figure 14).

The reactions catalyzed by the PTC lie at the heart of biology. However, despite the wealth of experimental results, the mechanisms of these reactions are still controversial. This encouraged us to investigate the reaction mechanism of peptide hydrolysis (paper III) and peptide bond formation (paper IV).
Peptidyl-tRNA hydrolysis

During peptidyl-tRNA hydrolysis, the nucleophilic attack of a water molecule on the ester carbonyl of peptidyl-tRNA results in cleavage of the ester bond between the P-site A76-O3' and nascent peptide chain (Figure 15). This reaction is triggered by the conserved GGQ sequence of RFs inserted in the PTC. The conserved Glu240 residue is methylated at the N5 position. This residue accelerates the hydrolysis reaction by positioning the nucleophile and by stabilizing the developing negative charge on the carbonyl oxygen of the substrate41-45.

For this reaction, the experimentally measured activation free energy46-48 is $\Delta G^\ddagger = 18.0$ kcal mol$^{-1}$ and the entropy contribution to the activation barrier is negligible46 ($T\Delta S^\ddagger = -1.6$ kcal mol$^{-1}$ at 25 °C). The hydrolysis reaction shows a linear pH dependence which indicates that an ionizable group with $pK_a > 9$ is implicated in the rate limiting step46. Based on this observation, it was hypothesized that a hydroxide ion is the nucleophile rather than water46,47. Furthermore, kinetic solvent isotope effect (KSIE) and proton inventory studies of the hydrolysis reaction are consistent with a mechanism in which only one proton is transferred in the rate limiting step46.

Prior to our study, several mechanisms were proposed for this reaction49,50. However, the suggested mechanisms appear to be inconsistent with the experimental observations46-48. That is, the calculated activation energies were too high, and the proposed mechanisms involve a concerted proton transfer between the reacting fragments which is clearly not supported by the pH-rate profile of the hydrolysis reaction46-48.

To investigate the underlying reason for the pH dependence of the hydrolysis reaction, we constructed a cluster model of the PTC which was used to study various mechanisms using the DFT method. Our model included the residues in the first solvation shell of the PTC. Although these residues do not participate directly in the chemical steps, mutagenesis experiments indicated that they are important for the hydrolysis reaction51. MD simulations41,45 also revealed that these residues participate in establishing the hydrogen bond network in the PTC. Using this model, we initially tested a mechanism in which a hydroxide ion, instead of water, acts as the nucleo-
phile. This stationary state, however, was not stable and the P-site A76 2’-hydroxyl donated a proton to the hydroxide ion forming a water molecule. Since, the $pK_a$ of the 2’-hydroxyl ($pK_a = 13.7$)\(^{52}\) is lower than that of water ($pK_a = 15.7$), the P-site A76 2’-hydroxyl seems to be a better candidate for deprotonation (Figure 16, R’). Here, we assumed that the PTC environment does not perturb the $pK_a$ of 2’-hydroxyl. This approximation appears to be reasonable as is evident from the small $pK_a$ shift of the nucleophilic amine in the peptide transfer reaction\(^{53}\). Using this approximation for the $pK_a$ of the 2’-hydroxyl, the energetic cost of the initial deprotonation was calculated to be about $\sim 8$ kcal mol\(^{-1}\) at pH 7.5.

![Figure 16](image)

*Figure 16. Schematic representation of the general base mechanism for the peptidyl-tRNA hydrolysis.*

Once the A76 2’-hydroxyl is deprotonated, the nucleophilic attack by the catalytic water to the substrate’s carbonyl carbon results in the formation of a negatively charged tetrahedral intermediate (Figure 16, IM). In this step, concurrently with the nucleophilic attack, the water molecule donates one proton to the A76 O2’. According to our DFT calculations, this step is rate limiting with an activation energy of $\Delta E^\ddagger = 15.8$ kcal mol\(^{-1}\). Correcting for the entropy, based on the experimental value, the activation free energy ($\Delta G^\ddagger = 17.4$ kcal mol\(^{-1}\)) is also in good agreement with experiments\(^{46-48}\). The agreement between the calculated KSIE (3.4) and the experimental measurement further supports this mechanism.

Hence, peptidyl-tRNA hydrolysis catalyzed by the PTC appears to occur via a base catalyzed mechanism and the deprotonation of A76 O2’ can explain the origin of its pH dependence\(^{54}\). In this mechanism, the deprotonated A76 O2’ acts as the general base by activating the nucleophilic water. In fact, this is consistent with the experiments in which the A76 O2’ is mutated. In these experiments, the activation free energy increased by approximately 6 kcal mol\(^{-1}\) which could partly be attributed to a change of the general base from A76 O2’ to a water molecule\(^{55}\).
Peptide bond formation

In the PTC, peptide bond formation occurs by the nucleophilic attack of the \( \alpha \)-amino group of aminoacyl-tRNA on the ester carbon of peptidyl-tRNA (Figure 17).

![Figure 17. Peptide bond formation in the PTC where the \( \alpha \)-amino group of aminoacyl-tRNA acts as the nucleophile.](image)

The activation free energy \( (\Delta C^\ddagger = 16.5 \text{ kcal mol}^{-1}) \) for this reaction, similar to peptidyl-tRNA hydrolysis, is dominated by the enthalpic term \( (\Delta H^\ddagger = 17.2 \text{ kcal mol}^{-1}) \) and the entropy contribution to the reaction barrier is negligible \( (\Delta S^\ddagger = 0.7 \text{ kcal mol}^{-1}) \)\(^{56,57}\). The reaction rate of peptide bond formation is only dependent on the ionization state of the nucleophilic amine indicating that, in contrast to peptidyl-tRNA hydrolysis, this reaction does not involve base catalysis\(^{53,58}\). For this reaction, the magnitude of the KSIE suggests a concerted proton transfer with three protons in flight\(^{46}\). Furthermore, heavy atom KIE experiments of peptide bond formation on the 50S subunit with a substrate analog indicate an early TS in which the nucleophilic attack of the \( \alpha \)-amino group on the ester carbon was suggested to be the rate limiting step\(^{59,60}\). Peptide bond formation on the 50S subunit, however, is much slower than the 70S ribosome reaction. Hence, it is not clear whether these results are applicable to peptide bond formation catalyzed by the 70S ribosome.

Two main mechanisms are proposed for peptide bond formation (Figure 18). One possibility for this reaction is the proton shuttle mechanism\(^{61-65}\) which involves formation of a zwitterionic tetrahedral intermediate (T\(^+\)). It was suggested that the T\(^+\) formation is fast and the rate limiting step corresponds to the collapse of T\(^+\). High resolution crystal structures\(^{66}\) and MD simulations\(^{62,63}\) further indicated that, during the T\(^+\) decomposition, the P-site A76 O2' and a water molecule mediate proton transfer from the \( \alpha \)-amino group to the P-site A76 O3', via an eight-membered ring TS. This mechanism is consistent with the KSIE and pH dependence of peptide bond formation\(^{67}\). The proton shuttle mechanism, however, does not appear to be consistent with the heavy atom KIE results\(^{59,60}\).

Alternatively, it was proposed that peptide bond formation proceeds by forming a negatively charged tetrahedral intermediate (T\(^-\))\(^{60}\). The heavy atom
KIE results suggested the rate limiting step corresponds to the T\(^-\) formation, hence, predicting an early TS. For peptide bond formation, the normal \(^{15}\text{N} \ KIE^{59,60}\) and near-zero Brønsted coefficient\(^59\) implies that, at the rate limiting step, one proton should be transferred simultaneously from the \(\alpha\)-amino group to an unknown base. Invoking such a base catalyzed mechanism, however, is not consistent with the pH-rate profile of peptide bond formation.

Figure 18. Schematic representation of the main mechanisms proposed for peptide bond formation.

One way to explain these results is the possibility of two competing mechanisms. Such a hypothesis seems reasonable, considering that peptide bond formation and peptidyl-tRNA hydrolysis are similar reactions. That is, they involve ester bond cleavage that occurs with different nucleophiles. Hence, it is not clear why these reactions should occur with different mechanisms as is evident from experimental results\(^46\). Here, similar to our previous study on peptide hydrolysis, we constructed a large cluster model of the PTC and tested various mechanisms using the DFT method.

We initially re-examined the previously proposed proton shuttle mechanism. In agreement with previous studies\(^62-65\), our calculations also suggested that the T\(^-\) formation is fast and the rate limiting step corresponds to the T\(^-\) breakdown. In our calculations, the T\(^-\) decomposition via the eight-membered ring TS overestimated the activation energy of peptide bond formation by \(\sim 8 \text{ kcal mol}^{-1}\). We further tested a concerted proton transfer via a ten-membered ring TS. The activation energy of this TS was 22.0 kcal mol\(^{-1}\) which is still higher than the experimental value. This is partly due to the fact that we only consider the ZPE contributions in our calculations. The
estimated activation energy is expected to improve if the thermally accessible vibrational degrees of freedom are also included. However, we are less certain about the magnitude of such a correction due to the limitations of RRHO approximations used in our method. We also investigated this TS by calculating the magnitude of the KSIE which was found to be in agreement with the experiment. Overall, both the KSIE and calculated activation energy of this mechanism are in semi-quantitative agreement with the corresponding experimental values indicating that the proton shuttle is a viable mechanism for peptide bond formation.

To investigate the possibility of a general base catalysis for peptide bond formation we removed one proton from the P-site A76 O2′ forming a negatively charged system. Such a strategy is useful, since no assumption is made regarding the identity of the proton acceptor. For this mechanism, the activation energies of the T′ formation and decomposition found to be 16.4 kcal mol⁻¹ and 19.2 kcal mol⁻¹ respectively. The rate limiting step cannot be identified confidently, since the energy difference between the calculated activation energies is not large enough and alternative conformations of the nascent peptide could potentially affect the estimated energy of the later TS. Nevertheless, the estimated activation energies, regardless of the rate limiting step, found to be in good agreement with experiments. This is not surprising as such a mechanism is operational during peptide hydrolysis. Interestingly, our calculations revealed that the T′ formation does not result in a normal ^15N KIE as was suggested based on the heavy atom KIE results^60. That is, such an early TS would result in a normal ^15N KIE only if the nucleophilic attack of the α-amino group occurs concurrently with the proton transfer from this group to either the bulk solvent or a base. However, we found no evidence of such a concurrent proton transfer even when A76 2′-hydroxyl is deprotonated. This is reasonable considering the higher pK_a of the α-amino group relative to the A76 2′-hydroxyl. The general base mechanism is clearly not consistent with the pH-rate profile of peptide bond formation and our calculation further predicted that the magnitude of the KSIE for this mechanism is much smaller than the experimental value^46.

Hence, according to our calculations, the proton shuttle and the general base mechanisms are both kinetically relevant to peptide bond formation. However, it is not clear why only the proton shuttle mechanism is operational in this reaction. In the general base mechanism, one proton should be transferred from the α-amino group to either the bulk solvent or a base. On the other hand, the proton shuttle mechanism occurs, if the water molecules that participate in proton transfer are in favorable positions. Hence, one explanation for this behavior is that the H-bond network of the PTC could potentially modulate the mechanism by which the peptide bond formation occurs. In fact, the early MD simulations highlighted the importance of a stable H-bond network for the PTC activity^62,63. Therefore, the presence of aminoacyl-tRNA in the A-site of the 70S ribosome could stabilize the H-bond
network that promotes the proton shuttle mechanism and inhibits general base catalysis, *e.g.*, by blocking proton transfer to the bulk. Such a hypothesis has direct implications for the interpretation of the experimental results since different experimental conditions could potentially change the mechanism of peptide bond formation.
Methods summary

Empirical valence bond method

The empirical valence bond method (EVB) employs a similar framework to valence bond theory with some simplifications. That is, each point along a reaction coordinate is described by a linear combination of valence structures\textsuperscript{8,68-70}. In EVB, the valence structures are defined by standard force fields of the form

\[
U(r) = \sum_{t} D_{t}^{l} \left(1 - e^{-a_{t}(b_{-} - b_{0})^{2}} \right) + \sum_{t} \frac{1}{2} k_{b}^{l} (\theta - \theta_{0}^{l})^{2} \\
+ \sum_{t} k_{\phi}^{l} \left[1 + \cos \left( n_{t} \phi_{t}^{l} - \delta_{t}^{l} \right) \right] + \sum_{t} \frac{1}{2} k_{\zeta}^{l} (\zeta - \zeta_{0}^{l})^{2} \\
+ \sum_{i<j} \frac{1}{4\pi\varepsilon_{0}} \frac{q_{i} q_{j}}{r_{ij}} + \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{11}
\]

where the terms on the r.h.s. describe bonds, angles, torsions, improper, electrostatic and van der Waals interactions respectively. For instance, a chemical reaction—in its simplest form—can be described in terms of two valence structures corresponding to the reactant and product states. In this way, for each point along the reaction coordinate, one obtains two potential energies from each valence structure, \( U^{R} \) and \( U^{P} \) (Figure 19). These are the diabatic potential energy surfaces whose intersection defines the TS. In EVB, the reaction coordinate is characterized by the difference between diabatic energies \( \Delta \varepsilon = U^{R} - U^{P} \). The adiabatic ground state energy surface is then obtained by mixing the diabatic energies according to the eigenvalue equation

\[
HC = CE \tag{12}
\]

where \( E \) is the diagonal eigenvalue matrix, the columns in matrix \( C \) correspond to the eigenvectors that define the weight of each diabatic state for a given eigenvalue, and \( H \) is the EVB Hamiltonian matrix. In the \( H \) matrix, the diagonal elements \( (H_{ii}) \) are the diabatic potential energies and the off-diagonal elements \( (H_{ij}) \) are the coupling between diabatic states. The lowest
eigenvalue in matrix $E$ corresponds to the ground state and the rest are the excited states (Figure 19)

![Figure 19](image)

*Figure 19.* The diabatic potential energy profile of valence structures (thin lines) and the adiabatic potential energy profile of the ground state (bold line) and excited state (dashed line) along an arbitrary reaction coordinate.

In EVB, the off-diagonal terms of matrix $H$ are defined empirically and diagonal elements are also corrected by a parameter ($\alpha$), which corresponds to the difference in the heat of formation between valence structures, based on experimental results or high level calculations. This parameterization is necessary since the classical potential functions describing the system do not take into account electronic effects.

Having defined the ground state potential energy surface, it is then desirable to obtain the free energy profile along the reaction coordinate. One way to construct this free energy profile is from the ensemble average, which can be obtained from molecular dynamics (MD) simulations, for each point along the reaction coordinate ($\chi = \Delta \varepsilon$) according to

$$
\Delta G(\chi_n) = -\beta^{-1} \ln \left[ \frac{\int \delta(\chi_n - \chi_n) e^{-\beta U^{VB}} d\Gamma}{\int e^{-\beta U^{VB}} d\Gamma} \right]
$$

(13)

where $\Delta G(\chi_n)$ is the free energy cost of being at the reaction coordinate $\chi_n$, and $\chi_n$ is the $n$th energy interval along the reaction coordinate, and $d\Gamma$ is the volume element in configurational space.

This direct approach, however, suffers from convergence problems since the high energy regions of energy surface will not be sampled. To circumvent this, in the spirit of free energy perturbation (FEP), a mapping potential is introduced to drive the system across the potential energy surface in discreet sampling windows (Figure 20). The mapping potential is given by

$$
U_{i}^{m} = (1 - \lambda_{i})U^{R} + \lambda_{i}U^{P}
$$

(14)
where the mapping parameter $\lambda$ varies in the $[0,1]$ interval and $U_i^m$ is the $i$th sampling potential. In this way, the system is restrained to the relevant energy neighborhood that is dictated by the mapping potential ($U_i^m$).

**Figure 20.** The free energy profile along the reaction coordinate is obtained by sampling on the mapping potential ($U^m$).

In this formulation, the free energy cost of being in the reaction coordinate $\chi_n$ can now be expressed as

$$
\Delta G(\chi_n) = \Delta G_{\lambda_1 \rightarrow \lambda_i} + \Delta G_{U_i^m \rightarrow U_i^B}
$$  \hspace{1cm} (15)

where the first term on the r.h.s. is the free energy cost of reaching the $i$th sampling window on the mapping potential and the second term is the free energy cost of switching from the mapping potential to the EVB ground state potential energy. The $\Delta G_{\lambda_1 \rightarrow \lambda_i}$ can be expressed as the ratio of partial partition functions along the discretized sampling path\textsuperscript{68,69} by

$$
\Delta G_{\lambda_1 \rightarrow \lambda_i} = -\beta^{-1} \sum_{j=1}^{i-1} \ln \left[ \frac{\int e^{-\beta U_{j+1}^m} d\Gamma}{\int e^{-\beta U_j^m} d\Gamma} \right]
$$

$$
= -\beta^{-1} \sum_{j=1}^{i-1} \ln \left[ \frac{\int e^{-\beta U_{j+1}^m} d\Gamma}{\int e^{-\beta(U_j^m-U_{j+1}^m)} e^{-\beta U_{j+1}^m} d\Gamma} \right]
$$

$$
= +\beta^{-1} \sum_{j=1}^{i-1} \ln (e^{-\beta(U_j^m-U_{j+1}^m)})_{j+1}
$$

$$
= \Delta G^{\text{FEP}} (\lambda_i)
$$  \hspace{1cm} (16)
The same strategy can also be used to express the second term on the r.h.s. of equation 15. That is, the EVB ground state energy could be considered as a perturbation to the mapping potential. Hence, the free energy of the EVB ground state energy is obtained by sampling on the mapping potential and is given by

\[
\Delta G^m_{i} \rightarrow v^{VB}_{i}(\chi_n) = -\beta^{-1} \ln \left[ \frac{\int \delta(\chi - \chi_n) e^{-\beta u^{VB}_{i}} \, d\Gamma}{\int e^{-\beta u^{m}_{i}} \, d\Gamma} \right]
\]

\[
= -\beta^{-1} \ln \left[ \frac{\int \delta(\chi - \chi_n) e^{-\beta (u^{VB}_{i}-u^{m}_{i})} e^{-\beta u^{m}_{i}} \, d\Gamma}{\int e^{-\beta u^{m}_{i}} \, d\Gamma} \right]
\]

\[
= -\beta^{-1} \ln \left( e^{-\beta \left( u^{VB}_{i} - u^{m}_{i} \right) \chi_n} \right)
\]

(17)

By substituting equations 16 and 17 into equation 15, we obtain the final expression for EVB free energy as

\[
\Delta G(\chi_n) = \sum_{\{i \mid \chi(i) \in \chi_n\}} p_i \left[ \Delta G^{PEP}(\lambda_i) - \beta^{-1} \ln \left( e^{-\beta \left( u^{VB}_{i} - u^{m}_{i} \right) \chi_n} \right) \right]
\]

(18)

where the sum runs over all the sampling windows that contribute to \( \chi_n \), and \( p_i \) is the normalized statistical weight of the \( i \)th sampling window. For each sampling window, this statistical weight is proportional to the number of contributions that is made to a given value of reaction coordinate.

EVB offers an inexpensive potential energy surface that could be calibrated to reproduce a known free energy profile and sampled efficiently. To be able to capture the catalytic effect of an enzyme active site in a chemical reaction, the EVB Hamiltonian is initially calibrated to reproduce the free energy profile of the reference reaction (Figure 4) based on the experimental values or high level calculations. Following this, the reaction is simulated in the enzyme active site using the same parameters that is obtained from the reference reaction simulation. In this approach, we assume that the catalytic effect of the enzyme is due to the interactions of the reacting fragments with the active site and that these interactions can be described adequately by a classical force field potential. Such an assumption is valid as long as these interactions involve no major electronic effects.
DFT cluster model

The main idea in this approach is to represent the enzyme active site by a small model based on reliable crystallographic structures\textsuperscript{71-73}. Therefore, the key residues can be included in the model to capture the main catalytic effects. Some atoms in the model system are then constrained to their crystallographic positions to mimic the effect of the protein backbone. In this way we make sure that the integrity of the active site is preserved and that we avoid artificial conformational changes. The electrostatic effect of the protein environment is taken into account by adding a continuum solvent with an appropriate dielectric constant. Once the cluster model is constructed, it is then possible to test different mechanistic hypotheses systematically. In our work, we employed the DFT method for studying chemical reactions since it provides a computationally manageable method for describing electronic structure.

The central object in DFT is the electron density, $\rho(\mathbf{r})$. This is so since the first theorem of Hohenberg and Kohn\textsuperscript{74} guarantees that $\rho(\mathbf{r})$ uniquely defines the Hamiltonian of a system and that the ground state energy of a many-particle system is a functional of the electron density, $E_0[\rho_0]$. Furthermore, the second Hohenberg-Kohn theorem proves that the electron density is amenable to the variational principle. That is, any trial density will result in an energy that is higher than the ground state energy, $E_0 < E[\tilde{\rho}]$ where $\tilde{\rho} \neq \rho_0$. Therefore, the electron density contains all information that is required to describe a many-particle system and the ground state energy can be obtained by minimizing the energy functional with respect to the electron density.

One of the difficulties in using the electron density for describing a molecular system is the calculation of electron kinetic energies. The Kohn-Sham formulation\textsuperscript{75} of DFT provides a practical method for tackling this problem. Here, it is assumed that there exists a hypothetical system in which electrons do not interact and this hypothetical system has the same electron density as the real one. In this way, in analogy to Hartree-Fock theory, the main contribution to the kinetic energy can be calculated explicitly from the single determinant wavefunction. Therefore, the energy functional in atomic units is given by\textsuperscript{76,77}

$$E[\rho(\mathbf{r})] = \sum_{i}^{N} \left( \langle \chi_i | - \frac{1}{2} \nabla^2_{\mathbf{r}_i} | \chi_i \rangle - \langle \chi_i | \sum_{k}^{\text{nuclei}} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} | \chi_i \rangle \right) \\
+ \sum_{i}^{N} \left( \langle \chi_i | \frac{1}{2} \int \frac{\rho(\mathbf{r})}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' | \chi_i \rangle + E_{xc}[\rho(\mathbf{r})] \right)$$ (19)
where the terms on the r.h.s. are, respectively, the kinetic energy of single non-interacting electrons, the nuclear-electron interaction, the classical electronic interactions, and the exchange-correlation energies. The last term takes into account the corrections to the kinetic energy due to electron interactions, the non-classical contributions to the coulomb interactions and self-interaction. In equation 19, the electron density is described in terms of orbitals and is given by

\[ \rho(r) = \sum_{i=1}^{N} |\chi_i|^2 \]  

(20)

where \( N \) is the number of electrons, and \( \chi_i \) represent the \( i \)th spin orbital. Minimizing the energy functional (equation 19) with respect to the orbitals—under the constraint of orthonormality—results in the Kohn-Sham (KS) equations

\[
-\frac{1}{2} \nabla^2 - \sum_{k} \frac{Z_k}{|r_i - r_k|} + \int \frac{\rho(r')}{|r_i - r'|} d\mathbf{r}' + \frac{\delta E_{xc}}{\delta \rho} \chi_i = \epsilon_i \chi_i
\]

(21)

where \( \delta E_{xc}/\delta \rho \) is the exchange-correlation potential and \( \epsilon_i \) is the KS orbital energy. The KS equations are a set of coupled self-consistent equations since the operator \( h_i^{KS} \) depends on the total density. The orbitals that satisfy these equations yield the exact ground state electron density of the real system. If the exchange-correlation functional was known, the ground state energy could then be obtained by plugging in these orbitals into equation 19. In practice, the spatial part of spin orbitals are expanded in a set of \( m \) basis functions as

\[
\chi_i(r) = \sum_{r=1}^{m} c_{ri} \phi_r(r)
\]

(22)

This expansion converts the problem of finding the orbitals to the optimization of the basis functions coefficients \( (c_{ri}) \). Hence, for a closed shell system where the orbitals are doubly occupied, the system of KS equations (equation 21) can now be written in the matrix form as an eigenvalue problem

\[
h^{KS} C = S C \epsilon
\]

(23)

where \( C_{m \times n} \) is the coefficient matrix of \( n \) orbitals expanded in \( m \) basis functions, \( S \) is the overlap matrix with elements \( S_{rs} = \int \phi^*_r(r) \phi_s(r) d\mathbf{r} \), \( \epsilon \) is diagonal matrix of KS orbital energies, and \( h^{KS} \) is the KS matrix

\[
h^{KS} = h + J + v_{xc}
\]

(24)
here $h$ is the one-electron integral matrix for kinetic energy and the electron-nuclear attraction which are given by

$$h_{rs} = -\frac{1}{2} \int \phi_i^*(r) \nabla^2 \phi_s(r) \, dr - \int \phi_i^*(r) \left[ \sum_{k}^{\text{nuclei}} \frac{Z_k}{|r - r_k|} \right] \phi_s(r) \, dr$$ \hspace{1cm} (25)

and $J$ is the Coulomb matrix with matrix elements

$$J_{rs} = \sum_{t,u=1}^{m} D_{tu} \int \frac{\phi_i^*(r_1) \phi_s(r_1) \phi_i^*(r_2) \phi_u(r_2)}{|r_1 - r_2|} \, dr_1 \, dr_2$$ \hspace{1cm} (26)

where $D_{tu}$ are the density matrix elements that are defined by coefficient matrix, $D_{tu} = 2 \sum_{i=1}^{N/2} c_{ti} c_{ui}^*$. The last term in equation 24 ($v_{xc} = \delta E_{xc}/\delta \rho$) is the matrix of exchange-correlation potential with matrix elements

$$v_{xc,rs} = \int \phi_i^*(r) v_{xc}(r) \phi_s(r) \, dr$$ \hspace{1cm} (27)

This equation is solved numerically and $v_{xc}(r)$ is evaluated at the electron density that is calculated from

$$\rho(r) = \sum_{r,s=1}^{m} D_{rs} \phi_i^*(r) \phi_s(r)$$ \hspace{1cm} (28)

Since equation 23 is dependent on the density matrix through matrices $J$ and $v_{xc}$, this eigenvalue equation should be solved self-consistently. The resulting optimized density can, in principle, be used to calculate the exact ground state energy according to

$$E = Tr(Dh) + \frac{1}{2} Tr(DJ) + E_{xc}$$ \hspace{1cm} (29)

For complex molecular systems, however, the explicit form of the exchange-correlation functional is not known which implies that these equations should be solved approximately. There are many flavors of the exchange-correlation functional depending on the level of approximations used in deriving them. In our work we employed the B3LYP and M06-2X exchange-correlation functionals. The B3LYP is one of the widely used hybrid functionals. The basic idea, here, is that part of the exchange energy can be computed exactly from the hypothetical non-interacting reference system. Hence, the B3LYP functional is defined as

$$E_{xc}^{B3LYP} = a E^{HF} + (1 - a) E^{LDA} + b E^{BB} + (1 - c) E^{LDA} + c E^{LYP}$$ \hspace{1cm} (29)
where the exchange energy is obtained by mixing the exact Hartree-Fock exchange ($E_{x}^{HF}$), local density approximation (LDA) exchange that is derived from the uniform electron gas ($E_{x}^{LDA}$), and the Becke’s generalized-gradient approximation (GGA) exchange which is dependent on the gradient of density in addition to the local density ($E_{x}^{B88}$). The correlation functional in B3LYP is defined as a combination of the uniform electron gas correlation ($E_{c}^{LDA}$), and the LYP functional derived from the correlation energy of the helium atom. Here, the empirical parameters $a$, $b$, and $c$ are optimized to reproduce some experimental data. In the same spirit, the M06-2X functional is a combination of LDA, PBE, VS98 and B97 functionals with more than 40 empirical parameters that are optimized to reproduce a set of experimental data\textsuperscript{80}. 


I den andra delen av denna avhandling undersökte vi bildandet av peptidbindningar och hydrolysen av peptidyl-tRNA i ribosomen. Ribosomen är ribonukleoproteinkomplex som sköter proteinsyntesen. Den prokaryota ribosomen består av två subenheter, 50S och 30S. Dessa två subenheter binds samman och bildar ett 70S-komplex som i sin tur binder en mRNA-molekyl.
70S-komplexet består av tre tRNA-bindningsställen. A-sätet är den position där det nya aminoacyl-tRNA kommer in i komplexet medan P-sätet binder till peptidyl-tRNA och E-sätet är det sista läget för tRNA innan det lämnar ribosomen. Under eloneringsfasen av proteinsyntesen bestäms proteinkedjens sekwensen hos den framväxande peptidkedjan genom Watson-Crick-interaktioner mellan mRNA-kodonet och A-sätets tRNA-antikodon. När rätt tRNA känns igen, bildar den begynnande peptidkedjan i P-sätet en peptidbindung med aminoacyl-tRNA i A-sätet och 70S-komplexet flyttar till nästa kodon på mRNA. Terminering av proteinsyntes kännetecknas av stoppkodon i mRNA-sekwensen. Dessa kodon igenkänns av termineringsfaktorer (RFs) som binder till A-sätet och triggar hydrolysvande av esterbindningen mellan peptidkedjan och P-sätets tRNA.

Bildandet av peptidbindung och hydrolysvande av peptidyl-tRNA sker vid peptidyltransferascentret (PTC) på ribosomen. PTC är lokaliserat på 50S-subenheten och består helt av ribonukleotider. Under eloneringsfasen riktar acceptor-armarna hos A- och P-sätets tRNA in mot PTC. Detta placering av peptidyl-tRNA och aminoacyl-tRNA i närheten av varandra. Under termingeringsfasen är A-sätet dock ockuperat av RFs som infogar en universellt konservaderad Gln aminosyra i PTC. Denna aminosyra aktiverar hydrolys av peptidyl-tRNA, vilket resulterar i frisättning av peptidkedjan från tRNA i P-sätet. Vi undersökte reaktionsmekanismen av peptidhydrolysvande i artikel III och bildandet av peptidbindningar i artikel IV.

Våra beräkningar visade att peptidyl-tRNA-hydrolysvande katalyserad av PTC verkar ske genom en substratassisterad general-bas-mekanisem och att deprotonering av substrat kan förklara uppkomsten av dess pH-beroende. I denna mekanisem fungerar A76 O2' hos substratet som bas och aktiverar en nukleoofil vattenmolekyl. Det aktiverade vattnet attackerar substratets karbonylkol, vilket resulterar i klyvning av esterbindningen mellan peptidkedjan och P-sätets tRNA. Bildning av peptidbindung sker genom en nukleoofil attack av α-aminogruppen hos aminoacyl-tRNA på esterkolet hos peptidyl-tRNA. För denna reaktion undersökte vi två mekanismer, nämligen den tidigare föreslagna ”proton shuttle”-mekanismen som innebär bildandet av en zwitterjonisk tetraedrisk intermediär, samt general-bas-mekanismen som sker genom att bilda en negativt laddad tetraedrisk intermediär. Även om båda mekanismerna resulterade i rimliga aktiveringsenerger är endast proton shuttle-mekanismen konsekvent med pH-beroendet av peptidbindnings bildning. Det är oklart hur generell-bas-mekanismen hämmas i ribosomen, men denna fråga kan vara i fokus i framtida forskningsprojekt.
I thank my supervisor Johan Åqvist for giving me the chance to do what I like the most and the freedom to find my own way. Over the years I came to appreciate his Samurai style of supervision, and even more so now that I am about to continue this journey on my own. I am deeply grateful to Fahmi Himo for hours of scientific discussions and friendship. I cannot thank Terese Bergfors enough for showing me the art of writing and stimulating discussions. I thank David van der Spoel for interesting discussions and answering my impetuous questions. I thank Lynn Kamerlin for her help and support. This thesis would not have been possible if it were not for the support of friends and colleagues in the ICM department. I refrain from naming them since it requires pages and takes hours, but I acknowledge their friendship and help.
References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1482

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-316497