1. DEAMIDATION IN PEPTIDES AND PROTEINS

Asparagine (Asn) and glutamine (Gln), two of the 20 amino acid residues that ordinarily occur in proteins, are inherently unstable under physiological solvent conditions. Asn and Gln were shown to be normal constituents of proteins [1, 2] and first studied with respect to deamidation [3]. The chemistry of the free amino acids was thoroughly understood by 1961 [4] and, by 1974 a partial understanding of peptide deamidation had been gained.

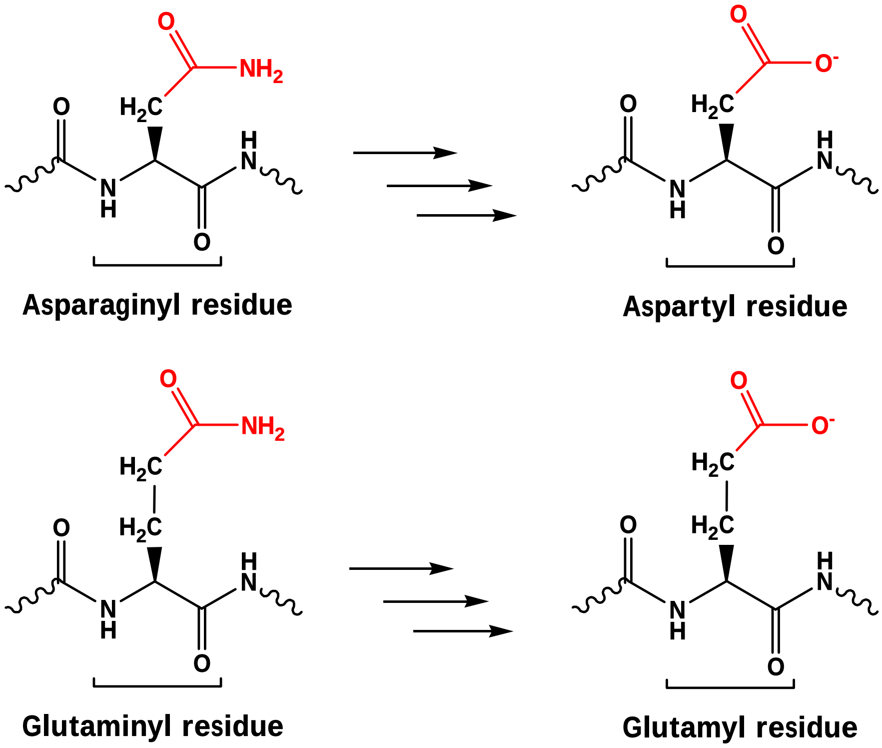


Figure 1.1. Deamidation of asparaginyl and glutaminyl residues.

Gln and Asn spontaneously and nonenzymatically deamidate into glutamyl, (Glu) and aspartyl (Asp) residues (Figure 1.1), with half-times that vary from a few hours to more than 100 years at 37 °C. Gln deamidation is usually substantially slower than Asn deamidation. In addition to peptide and protein structure, non-enzymatic deamidation rates depend upon pH, temperature, ionic strength, and buffer type [5, 6]. Enzymatic deamidation also occurs in living things.

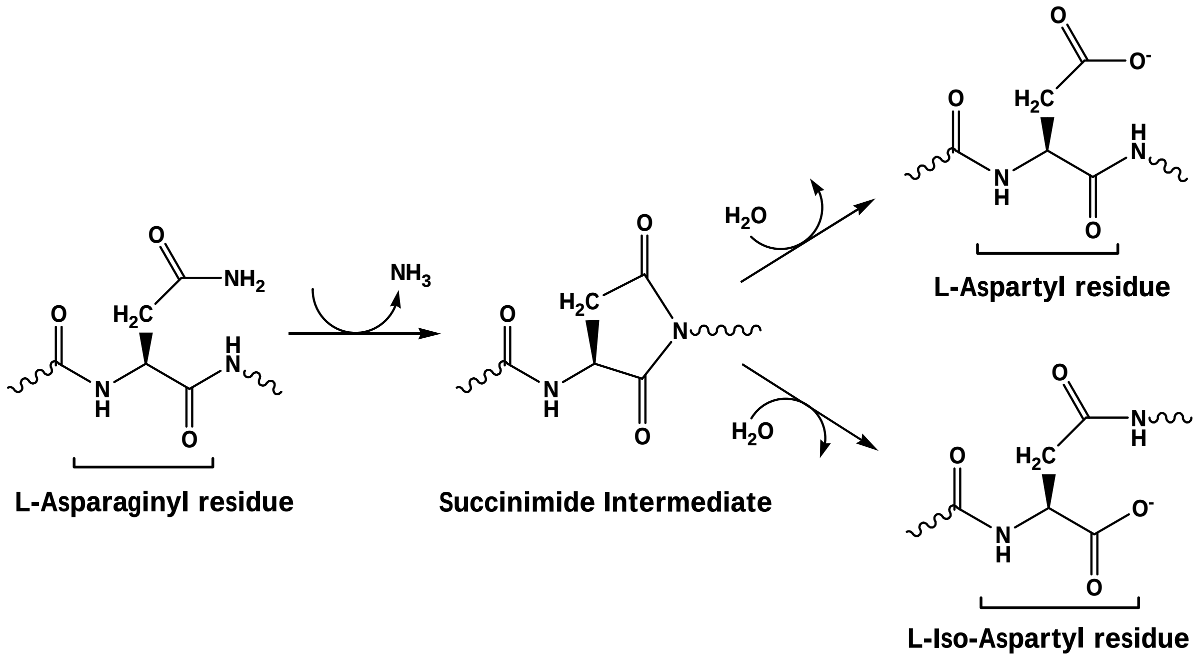


Figure 1.2. Deamidation through succinimide intermediate.

Although deamidation probably occurs through several reaction mechanisms, there is substantial evidence that deamidation of relatively unrestrained Asn residues proceeds through a succinimide intermediate near neutral pH (Figure 1.2), leading to significant amounts of iso-aspartyl residues [7-9].

Three-dimensional structure near the amide is important in determining the deamidation rate; restrictive structures like α-helices substantially slow down the rate while more flexible structures allow faster rates. It was found that most protein Asn deamidation rates are determined by primary structure and slowed down by three-dimensional structure. There are rare examples of acceleration by protein structure.

The deamidation of Asn and Gln in peptides and proteins is of significant biological interest, because it can often produce substantial structural changes. At neutral pH, a negative charge is added to the molecule, and, in some cases, the resulting Asp is isomerized.

It has been hypothesized that deamidation serves as a *molecular clock* for the timing of biological processes [10, 11]. The timed processes of protein turnover, development, and aging have been suggested as possible roles for deamidation. The biological turnover rates of rat cytochrome c [6, 10] and rabbit muscle aldolase [12], have been shown to be controlled by deamidation. Increased amounts of deamidated proteins have been found in some aged and diseased tissues, such as human eye lens cataracts [13] and Alzheimer's plaques [14].

About 1,700 research papers on various aspects of the deamidation of peptides and proteins have been published since the biological importance of deamidation was first emphasized [6, 11]. This literature is, however, mostly fragmented into special studies of individual peptides and proteins in a wide variety of conditions. Most of this work has been hindered by the fact that experimental studies of protein deamidation with available techniques are laborious, time-consuming and lack a means of reliably estimating the instability of a particular peptide or protein with respect to non-enzymatic deamidation of the amide residues.

Since the original suggestion in 1970 [11] that deamidation plays a positive biological role, especially as a molecular timer, some evidence has accumulated to support this hypothesis. It was found that deamidation rates could be varied over a wide range by changing primary sequence [15]. The distribution functions of naturally occurring sequences around amide residues were found to be non-random as were the amide compositions of proteins [11, 16]. Specific roles for some deamidations were found [17] and, in two cases, it was shown that deamidation regulates the rate of protein turnover [6, 10, 12]. Studies of the occurrence of deamidation in a wide variety of proteins have been reported [13, 18, 19]. The question remains, however, as to whether or not deamidation is an interesting property of proteins that occasionally has biological usefulness, or is it of widespread biological importance.

The estimated deamidation rates of proteins in the Brookhaven Protein Data Bank show that deamidation may be expected to occur in a substantial percentage of proteins under physiological conditions and within biologically interesting time intervals. These estimates agree well with the actual protein deamidation experiments that have been reported. Moreover, since instances are known wherein deamidation increases protein susceptibility to biological degradation, even more deamidation may be occurring than is ordinarily seen in biological preparations.

Most Asn residues in proteins do not have fast deamidation rates. The amides that are most interesting biologically consist of only a few percent of the entire set. Therefore, biologically stable amides could easily be genetically provided. Unstable amides need not be present in proteins unless their instability is biologically functional.

The change in charge that accompanies deamidation has a substantial effect on protein structure, as does the isomerization that sometimes occurs. There are many reports that these changes markedly affect protein function or stability or both. Amide residues genetically programmed for the proteins of biological systems would, therefore, be expected to be stable with respect to deamidation unless the shorter-lived amides have positive biological uses.

To summarize, one can recall the conclusions from a recent work of Robinson: "*Proteins contain amide residue clocks. These residues are found in almost all proteins and amide residue clocks are found to be set to timed intervals of biological importance, even though settings to longer times are not only available, but also make up most of the genetically available settings. Deamidation changes protein structures in fundamentally important ways. If deamidation were not of pervasive and positive biological importance, these clocks would be set to time intervals that are long with respect to the lifetimes of living things The fact that they are found to be set instead to biologically relevant time intervals strongly supports the original hypothesis that amides play, through deamidation, a special biologically important role*".

Deamidation is obviously being used for some widespread and fundamental biological purpose. If this were not so, it would be genetically suppressed since it is otherwise very disruptive to protein structures and would not be tolerated in living systems.

2. OBJECTIVE AND SCOPE

The previous chapter was meant to provide a brief overview on deamidation in peptides and proteins and emphasize its biological importance. As mentioned earlier, mechanistic aspects of deamidation have been previously investigated both experimentally and theoretically, but many aspects have not been elucidated yet. The present study aims to get a deeper insight on the mechanisms leading to deamidation of asparaginyl residues using computational techniques to investigate the feasibility of alternative reaction mechanisms. Some of these mechanisms have been proposed in the literature. Others are put forward here for the first time. The following chapter presents the basic principles of the theoretical approaches employed throughout the course of this study; pertinent methodological details as well as practical aspects of the computational methods utilized are presented.

This study aims to tackle four different topics (Chapters 4 – 7) related to deamidation of Asn in peptides and proteins. These chapters were intended to be full articles, hence they were written in a standalone, independent format, where each chapter includes a detailed introduction and methodology section. The first topic of interest (Chapter 4) is the effect of solvent molecules on deamidation mechanism, where the difference in mechanistic details and energetics of water-catalyzed deamidation is investigated. The second topic (Chapter 5) is based on a comparative study of alternative routes for deamidation, more specifically a comparative analysis of direct hydrolysis of Asn versus the cyclic imide-mediated route. A possible side reaction –non-enzymatic peptide backbone cleavage at Asn and Asp residues– is studied and the relative feasibility of peptide fragmentation near these residues is discussed (Chapter 6). The influence of protein primary structure on deamidation rates is investigated (Chapter 7), where the effect of the Asn carboxyl-side residue's identity on deamidation rates is explored. Important conclusions drawn from each chapter are summarized (Chapter 8) and specific suggestions for future work are discussed (Chapter 9).

3. THEORETICAL BACKGROUND

3.1. Molecular Quantum Mechanics

All electronic structure methods seek an approximate solution to the Schrödinger equation. For small systems consisting of a few particles, a very accurate solution for the Schrödinger equation can be found but, unfortunately, there is no simple way of obtaining exact solutions for many electron atoms and molecules. Therefore some assumptions have to be made.

The molecular orbital (MO) programs build a set of MO's to be occupied by the electrons assigned to the molecule. The MO calculation then simply involves finding the combination of the atomic orbitals that have the proper symmetries and that give the lowest electronic energy. This is known as the linear combination of atomic orbitals (LCAO) [20].

Many simple MO methods are based on one-electron treatment, in that the electron is considered not to interact with others in the molecule. The Hückel and Extended Hückel theory is based on this approach.

The SCF method takes electron-electron terms into account by considering the interaction between an electron in a given orbital and the mean field of the other electrons in the molecule. It involves an iterative process in which the orbitals are improved from cycle to cycle until the electronic energy reaches a constant minimum value and the orbitals no longer change. This situation is described as self-consistent.

At the SCF level the electron-electron interaction is actually overestimated. The theory does not allow the electrons to avoid each other but assumes that their instantaneous positions are independent of one another. However, the error is reasonably consistent, so that its effects can be made to cancel by the use of proper comparison. The SCF method is also known as the Hartree-Fock or single determinant theory.

Both ab initio and semi-empirical calculations treat the linear combination of orbitals by iterative computations, which establish a self-consistent electric field (SCF) and minimize the energy of the system. In ab initio calculations, electron-electron repulsion is specifically taken into account, while the semi-empirical method appears as an intermediate approach.

3.2. Semi-empirical Methods

Semi-empirical methods are characterized by their use of parameters derived from experimental data in order to simplify the approximation to the Schrödinger equation. As such, they are relatively inexpensive and can be practically applied to very, very large molecules.

The semi-empirical quantum-mechanical methods developed by Dewar and coworkers [21-24] have been successful at reproducing molecular energies, replicating molecular structures and interpreting chemical reactions [25, 26]. To overcome some of the computational difficulties, approximations are made in which several of the integrals involving core orbitals are replaced by parameters. The number of two-electron integrals calculated is reduced, by simply ignoring them or calculating them in an approximate fashion. Three levels of approximation have been defined by Pople and Beveridge, in which certain two-electron integrals are neglected [27].

The first is known as the complete neglect of differential overlap (CNDO) [28]. It assumes the atomic orbitals to be spherically symmetrical when evaluating electron repulsion integrals. The directionality of p-orbitals was included only via the one-electron resonance integrals, the size of which depend on the orientations and distances of the orbitals and on a constant assigned to each type of bond.

The second, known as intermediate neglect of differential overlap (INDO), contains all terms that CNDO contains and includes all one-center two-electron integrals. The third is known as neglect of diatomic differential overlap (NDDO) in which all two-electron two-center integrals involving charge clouds arising from pairs of orbitals on an atom were retained.

In 1975, Dewar and coworkers published the MINDO/3 method, which is a modified version of the INDO method. MINDO/3 uses a set of parameters in approximation. These parameters, along with the constants used to evaluate the resonance integrals, allow the results to be fitted as closely as possible to experimental data.

The first practical NDDO method was introduced by Dewar and Thiel in 1977 [29] called modified neglect of diatomic overlap (MNDO), the model was parameterized on experimental molecular geometries, heats of formation, dipole moments and ionization potentials. The orbital exponents and the core integral were again treated as empirical parameters to be determined in the fitting procedure.

The inability of MNDO has lead to a reexamination of the model, leading to the Austin Model 1 (AM1) [24]. In this model a term was added to MNDO to correct for the excessive repulsions at van der Waals distances. For this purpose, each atom was assigned a number of spherical gaussians, which were intended to mimic long-range correlation effects.

The third parameterization of MNDO is the Parametric Method Number 3 (PM3), AMl being the second. In PM3, the parameters were optimized using a large set of reference molecular data. This allowed 12 elements to be optimized simultaneously [30]. The PM3 method has been used in this study for the semi-empirical level calculations.

Semi-empirical and ab initio methods differ in the trade-off made between computational cost and accuracy of the result. Semi-empirical calculations are relatively inexpensive and provide reasonable qualitative descriptions of molecular systems and fairly accurate quantitative prediction of energies and structures for systems where good parameter sets exist. Semi-empirical methods may only be used for systems where parameters have been developed for all of their component atoms. In addition to this, semi-empirical models have a number of well-known limitations. Types of problems on which they do not perform well include hydrogen bonding, transition structures, and molecules containing atoms for which they are poorly parameterized.

3.3. Ab Initio Methods

The main difficulty in solving the Schrödinger equation is the presence of the electron-electron interactions terms. It is very difficult to find analytical solutions to the Schrödinger equation that has this as part of its potential energy term, but computational techniques are available that give very detailed and reliable numerical solutions for the wavefunctions and energies [31, 32].

The time-independent Schrödinger equation is:

$$\hat{H}\psi = E\psi \tag{3.1}$$

where E is the electronic energy, ψ is the wave function describing the system and $\hat{H}$ is the Hamiltonian operator.

Nuclei are much heavier than electrons. Hence the electrons move much faster than the nuclei, and to a good approximation, one can regard the nuclei as fixed while the electrons carry out their motion. The Born-Oppenheimer approximation neglects the kinetic energy of nuclei and the Hamiltonian operator in atomic units is:

$$\hat{H} = \sum_{i=1}^{N}(-\frac{1}{2}\nabla_i^2) - \sum_{i=1}^{N}\frac{1}{r_i} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{\alpha=1}^{N}\sum_{\beta>\alpha}^{N}\frac{1}{r_{\alpha\beta}} \tag{3.2}$$

where the first term is the kinetic energy operator for electrons, the second term is the operator for the attraction between electrons and nuclei, the third term is the interelectronic repulsion operator and the final term is the internuclear repulsion operator for an N electron system.

Then approximating ψ as an antisymmetrized product of n orthonormal orbitals $\Psi_i(\vec{x})$ with a Slater determinant, $\vec{x}$ denoting both spin and spatial coordinates of an electron, the electronic energy (called Hartree-Fock (HF) energy at this level of approximation) may be written as:

$$E_{HF} = \langle \psi_{HF} | \hat{H} | \psi_{HF} \rangle = 2\sum_{i=1}^{n} H_i + \sum_{i,j=1}^{n} (2J_{ij} - K_{ij}) \tag{3.3}$$

$$H_i = \int \Psi_i^*(\vec{x}) \left[-\frac{1}{2}\nabla^2 + v(\vec{x}) \right] \Psi_i(\vec{x}) d\vec{x} \tag{3.4}$$

where $v(\vec{x})$ is the external potential due to the nuclei or other sources, J_{ij} is the Coulomb integral and K_{ij} is the exchange integral:

$$J_{ij} = \iint \Psi_i(\vec{x}_1)\Psi_i^*(\vec{x}_1)\frac{1}{r_{12}}\Psi_j^*(\vec{x}_2)\Psi_j(\vec{x}_2) d\vec{x}_1 d\vec{x}_2 \tag{3.5}$$

$$K_{ij} = \iint \Psi_i^*(\vec{x}_1)\Psi_j(\vec{x}_1)\frac{1}{r_{12}}\Psi_i(\vec{x}_2)\Psi_j^*(\vec{x}_2) d\vec{x}_1 d\vec{x}_2 \tag{3.6}$$

The average value of the energy, including the normality conditions, is given by equation 3.7. Then, using the variation principle, one finds the wave function that minimizes the energy.

$$\langle E \rangle = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \tag{3.7}$$

Hartree-Fock theory is very useful for providing initial, first-level predictions for many systems. It is also reasonably good at computing the structures and vibrational frequencies of stable molecules and some transition states. As such, it is a good base-level theory. However, the Hartree-Fock method does not take into account the correlation energy resulting from the instantaneous interactions of electrons. Its neglect of electron correlation makes it unsuitable for the accurate modeling of the energetics of reactions and bond dissociation.

Since electrons are correlated with each other the instantaneous electron correlation should be included into the wavefunction. Methods that include correlation energy such as configuration interaction methods (CI), coupled cluster (CC) and perturbation theory are called post-Hartree-Fock methods [33]. There are other correlated methods [34], such as multiconfiguration SCF (MCSCF) and multi-reference configuration interaction (MRCI).

The Møller-Plesset perturbation theory (MP) treats the effect of electronic interactions as a perturbation to a system consisting of non-interacting electrons [27]. The first order perturbation introduces the interaction between the electrons in the ground state and is equal to the Hartree-Fock theory. The second order perturbation (MP2) takes into account the interaction of the doubly excited configurations with the ground state configuration. The third order perturbation (MP3) adds the contribution of doubly excited configurations interacting with each other. The fourth order perturbation (MP4) brings in interactions involving single, double and quadruple excitations. Although the MP4 method is very accurate it is very expensive computationally. Post-SCF methods are not suitable for this project, since there are many structures to optimize; high computational expenses demand an alternative method.

Fortunately, the density functional theory (DFT) provides a cheaper method to treat correlation energy.

3.4. Density Functional Theory

The DFT method computes electron correlation via general functionals of the electron density. DFT functionals partition the electronic energy into several components which are computed separately: the kinetic energy, the electron-nuclear interaction, the Coulomb repulsion, and an exchange-correlation term accounting for the remainder of the electron-electron interaction (which is itself divided into separate exchange and correlation components in most actual DFT formulations.)

The density functional theory is based on the Kohn-Hohenberg theorems proposed in 1964 [34-35]. The first theorem states that the electron density $\rho(r)$ determines the external potential $v(r)$, i. e. the potential due to the nuclei. The second theorem introduces the variational principle. Hence, the electron density can be computed variationally and the position of nuclei, energy, wave function and other related parameters can be calculated.

The electron density is defined as:

$$\rho(x) = N\int \ldots \int \left|\Psi(x_1, x_2, \ldots x_n)\right|^2 dx_1 dx_2 \ldots dx_n \tag{3.8}$$

where x represents both spin and spatial coordinates of electrons.

The electronic energy can be expressed as a functional of the electron density:

$$E[\rho] = \int v(r)\rho(r)dr + T[\rho] + V_{ee}[\rho] \tag{3.9}$$

where $T[\rho]$ is the kinetic energy of the interacting electrons and $V_{ee}[\rho]$ is the interelectronic interaction energy. The electronic energy may be rewritten as:

$$E[\rho] = \int v(r)\rho(r)dr + T_s[\rho] + J[\rho] + E_{xc}[\rho] \tag{3.10}$$

with $J[\rho]$ being the coulomb energy, $T_s[\rho]$ being the kinetic energy of the non-interacting electrons and $E_{xc}[\rho]$ being the exchange-correlation energy functional. The exchange-correlation functional is expressed as the sum of an exchange functional $E_x[\rho]$ and a correlation functional $E_c[\rho]$, although it contains also a kinetic energy term arising from the kinetic energy difference between the interacting and non-interacting electron systems. Kinetic Energy term, being the measure of the freedom, and exchange-correlation energy, describing the change of opposite spin electrons (defining extra freedom to an electron), are the favorable energy contributions. The Coulomb energy term describes the unfavorable electron-electron repulsion energy and therefore disfavors the total electronic energy.

In Kohn-Sham density functional theory, a reference system of independent non-interacting electrons in a common, one-body potential V_{KS} yielding the same density as the real fully-interacting system is considered. More specifically, a set of independent reference orbitals ψ_i satisfying the following independent particle Schrödinger equation are imagined:

$$\left[-\frac{1}{2}\nabla^2 + V_{KS}\right]\psi_i = \varepsilon_i\psi_i \tag{3.11}$$

with the one-body potential V_{KS} defined as:

$$V_{KS} = v(r) + \frac{\partial J[\rho]}{\partial\rho(r)} + \frac{\partial E_{xc}[\rho]}{\partial\rho(r)} \tag{3.12}$$

$$V_{KS} = v(r) + \frac{\rho(r')}{\left|r-r'\right|}dr' + v_{xc}(r) \tag{3.13}$$

where $v_{xc}(r)$ is the exchange-correlation potential. The independent orbitals ψ_i are known as Kohn-Sham orbitals and give the exact density by:

$$\rho(r) = \sum_i^N \left|\psi_i\right|^2 \tag{3.14}$$

if the exact form of the exchange-correlation functional is known. However, the exact form of this functional is not known and approximate forms are developed starting with the local density approximation (LDA). This approximation gives the energy of a uniform electron gas, i. e. a large number of electrons uniformly spread out in a cube accompanied with a uniform distribution of the positive charge to make system neutral. The energy expression :

$$E[\rho] = T_s[\rho] + \int \rho(r)v(r)dr + J[\rho] + E_{xc}[\rho] + E_b \tag{3.15}$$

where E_b is the electrostatic energy of the positive background. Since the positive charge density is the negative of the electron density due to uniform distribution of particles, the energy expression is reduced to:

$$E[\rho] = T_s[\rho] + E_{xc}[\rho] \tag{3.16}$$

$$E[\rho] = T_s[\rho] + E_x[\rho] + E_c[\rho] \tag{3.17}$$

The kinetic energy functional can be written as:

$$T_s[\rho] = C_F \int \rho(r)^{5/3} dr \tag{3.18}$$

where C_F is a constant equal to 2.8712. The exchange functional is given by:

$$E_x[\rho] = -C_x \int \rho(r)^{4/3} dr \tag{3.19}$$

with C_x being a constant equal to 0.7386. The correlation energy, $E_c[\rho]$, for a homogeneous electron gas comes from the parametrization of the results of a set of quantum Monte Carlo calculations.

The LDA method underestimates the exchange energy by about 10 per cent and does not have the correct asymptotic behavior. The exact asymptotic behavior of the exchange energy density of any finite many-electron system is given by:

$$\lim_{x\to\infty} U_x^{\sigma} = -\frac{1}{r} \tag{3.20}$$

U_x^{σ} being related to $E_x[\rho]$ by:

$$E_x[\rho] = \frac{1}{2}\sum_{\sigma}\int \rho_{\sigma} U_x^{\sigma}\, dr \tag{3.21}$$

A gradient-corrected functional is proposed by Becke:

$$E_x = E_x^{LDA} - \beta \sum_{\sigma}\int \rho_{\sigma}^{4/3} \frac{x_{\sigma}^2}{1 + 6\beta x_{\sigma} \sinh^{-1} x_{\sigma}}\, dr \tag{3.22}$$

where σ denotes the electron spin, $x_{\sigma} = \frac{|\nabla \rho_{\sigma}|}{\rho_{\sigma}^{4/3}}$ and β is an empirical constant (β=0.0042). This functional is known as Becke88 (B88) functional [36].

The adiabatic connection formula connects the non-interacting Kohn-Sham reference system (λ=0) to the fully-interacting real system (λ=1) and is given by:

$$E_{xc} = \int_0^1 U_{xc}^{\lambda} d\lambda \tag{3.23}$$

where λ is interelectronic coupling-strength parameter and U_{xc}^{λ} is the potential energy of exchange-correlation at intermediate coupling strength. The adiabatic connection formula can be approximated by:

$$E_{xc} = \frac{1}{2} E_x^{exact} + \frac{1}{2} U_{xc}^{LDA} \tag{3.24}$$

since $U_{xc}^{0} = E_x^{exact}$, the exact exchange energy of the Slater determinant of the Kohn-Sham orbitals, and $U_{xc}^{1} = U_{xc}^{LDA}$ [37].

The closed shell Lee-Yang-Parr (LYP) correlation functional [38] is given by:

$$E_c = -a \int \frac{1}{1 + d\rho^{-1/3}} \left\{ \rho + b\rho^{-2/3} \left[C_F \rho^{5/3} - 2t_w + (\frac{1}{9} t_w + \frac{1}{18} \nabla^2 \rho) \right] e^{-c\rho^{-1/3}} \right\} dr \tag{3.25}$$

where

$$t_w = \frac{1}{8} \frac{\left|\nabla \rho(r)\right|^2}{\rho(r)} - \frac{1}{8} \nabla^2 \rho \tag{3.26}$$

and *a*=0.04918, *b*=0.132, *c*=0.2533 and *d*=0.349.

The mixing of LDA, B88, E_x^{exact} and the gradient-corrected correlation functionals to give the hybrid functionals [39] involves three parameters:

$$E_{xc} = E_{xc}^{LDA} + a_0\left(E_x^{exact} - E_x^{LDA}\right) + a_x \Delta E_x^{B88} + a_c \Delta E_c^{non-local} \tag{3.27}$$

where ΔE_x^{B88} is the Becke's gradient correction to the exchange functional. In the B3LYP functional, which was utilized in this study, the gradient-correction ($\Delta E_c^{non-local}$) to the correlation functional is included in LYP. However, LYP contains also a local correlation term which must be subtracted to yield the correction term only:

$$\Delta E_c^{non-local} = E_c^{LYP} - E_c^{VWN} \tag{3.28}$$

where E_c^{VWN} is the Vosko-Wilk-Nusair correlation functional, a parametrized form of the LDA correlation energy based on Monte Carlo calculations. The empirical coefficients are a_0=0.20, a_x=0.72 and a_c=0.81 [40].

3.5. Basis Sets

A basis set is the mathematical description of the orbitals within a system, (which in turn combine to approximate the total electronic wavefunction) used to perform the theoretical calculation. Larger basis sets more accurately approximate the orbitals by imposing fewer restrictions on the locations of the electrons in space.

In 1951, Roothaan proposed the Hartree-Fock orbitals as linear combinations of a complete set of known functions, called basis functions. There are two types of set of basis functions for atomic Hartree-Fock calculations, Slater-type functions and Gaussian-type functions.

In the simplest Hartree-Fock model, the number of basis functions on each atom is as small as possible that is only large enough to accommodate all the electrons and still maintain spherical symmetry. As a consequence, the molecular orbitals have only limited flexibility. If larger basis sets are used, the number of adjustable coefficients in the variational procedure increases, and an improved description of the molecular orbitals is obtained. Very large basis sets will result in nearly complete flexibility. The limit of such an approach termed the Hartree-Fock limit represents the best that can be done with a single electron configuration.

There are numerous different Gaussian basis sets with which SCF calculations can be carried out. The most widely used are those developed by Pople and co-workers. The simplest and lowest basis set is called STO-3G. This means that the Slater-type orbitals are represented by three gaussian functions. It is a minimal basis set which means that it has only as many orbitals as are necessary to accommodate the electrons of the neutral atom.

The next level of basis sets developed by Pople is referred to as the split-valence basis sets. The problem of any minimal basis set is its inability to expand or contract its orbitals to fit the molecular environment. One solution to the problem is to use split valence or double zeta basis sets in which the basis are split into two parts, an inner, compact orbital and an outer, more diffuse one. Thus the size of the atomic orbital that contributes to the molecular orbital can be varied within the limits set by the inner and outer basis functions. Split-valence basis set splits only the valence orbitals in this way, whereas double zeta basis also have split core orbitals.

For greater flexibility the split-valence basis set can be augmented with polarization functions. In polarization basis sets, which are the next level of improvement in basis set, d orbitals are added to all heavy atoms is designated with a * or (d). Polarization can be added to hydrogen atoms as well, this would be done by **.

Diffuse functions are large-size versions of s- and p-type functions (as opposed to the standard valence-size functions). They allow orbitals to occupy a larger region of space. Basis sets with diffuse functions are important for systems where electrons are relatively far from the nucleus: molecules with lone pairs, anions and other systems with significant negative charge, systems in their excited states, and so on. The 6-31+G (d) basis set is the 6-31G (d) basis set with diffuse functions added to heavy atoms. The double plus version, 6-31++G (d), adds diffuse functions to the hydrogen atoms as well. The 6-31+G** split valence basis set has been used for gas phase calculations in this study, with the addition of polarization and diffused functions on heavy atoms as well as polarization functions on hydrogens.

3.6. Continuum Solvation Models

In continuum solvation models [41, 42], the solvent is represented as a uniform polarizable medium characterized by its static dielectric constant ε. In basic continuum solvation models, the solute is described at a homogenous quantum mechanical (QM) level and the solute-solvent interactions are limited to those of electrostatic terms.

The total solvation free energy may be written as

$$\Delta G_{solvation} = \Delta G_{cavity} + \Delta G_{dispersion} + \Delta G_{electrostatic} \tag{3.29}$$

In this representation, ΔG_{cavity} is the energetic cost of creating a cavity in the medium producing a destabilization effect. Dispersion interactions between solvent and solute add stabilization to solvation free energy term expressed as $\Delta G_{dispersion}$. The latter electrostatic term, $\Delta G_{electrostatic}$, has a stabilization effect and furthermore it appears to be responsible for the main structural changes of the solute.

The solute charge distribution within the cavity induces a polarization of the surrounding medium, which in turn induces an electric field within the cavity called the *reaction field*. This field then interacts with solute charges, providing additional stabilization. The effect of the reaction field may be modeled by an appropriately distributed set of induced polarization charges on the surface S of the dielectric. The charge density on the surface of the cavity, $\sigma(\vec{r}_S)$, is given by the standard electrostatics in terms of the dielectric constant, ε, and the electric field perpendicular to the surface, $\vec{F}$, generated by the charge distribution within the cavity

$$4\pi\varepsilon\sigma(\vec{r}_S) = (\varepsilon - 1)F(\vec{r}_S) \tag{3.30}$$

Once $\sigma(\vec{r}_S)$ is determined, the associated potential is added as an extra term to the Hamiltonian operator

$$H = H_0 + V_\sigma \tag{3.31}$$

$$V_\sigma(\vec{r}) = \int \frac{\sigma(\vec{r}_S)}{|\vec{r} - \vec{r}_S|} d\vec{r}_S \tag{3.32}$$

The potential from the surface charge V_σ is given by the molecular charge distribution but also enters the Hamiltonian and thus influences the molecular wave function, the procedure is therefore iterative.

In the Polarized Continuum solvation Models (PCM), the solute is embedded in a cavity defined by a set of spheres centered on atoms (sometimes only on heavy atoms), having radii defined by the van der Waals radius of the atoms multiplied by a predefined factor (usually 1.2). The cavity surface is then subdivided into small domains (called tesserae), where the polarization charges are placed. Among the many solutions to the electrostatic problem is the Integrated Equation Formalism (IEF) originally formulated by Cancés and Menucci [43-45]. The IEF-PCM method is a recent development in the polarized continuum models and has been utilized for solvent calculations in this study. It is based on the use of operators largely exploited in the theory of integral equations. The concept of cavity and of its tessellation is conserved. The IEF formalism is in fact able to treat a larger class of electrostatic problems.

Apart from the apparent surface charge (ASC) methods, other solutions for calculating the electrostatic solute-solvent interactions in continuum models have been proposed such as, multipole expansion (MPE) methods, generalized Born approximation (GBA), image charge (IMC) methods, finite element methods (FEM) and finite difference methods (FDM).

3.7. Molecular Dynamics

Molecular dynamics (MD) simulations is one of the principal methods in the theoretical investigation of biological molecules that provides information on the time dependent behavior of molecular systems. MD simulations are routinely used to examine the structure, dynamics and thermodynamics of biological molecules and their complexes.

Molecular dynamics is a *deterministic* method; the state of the system at any future time can be predicted from its current state. Successive configurations of the system are generated by integrating Newton's law of motion. The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time.

At the heart of molecular mechanics lies the force field which describes the potential energy surface of the system. The force field is composed of various contributions like bonded or valence terms (bond stretching, angle bending and torsion angle) and non bonded terms (van der Waals and Coulomb forces) all of which contain empirical parameters fitted to results of experimental studies or high level calculations.

In the light of these contributions the potential energy V(R) is defined as:

$$V(\mathbf{R}) = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{non-bonded}} \tag{3.39}$$

The specific contributions to the potential energy can be described as follows:

$$E_{\text{bond}} = \sum_{\text{bonds}} \frac{k_b}{2}(l - l_0)^2 \tag{3.40}$$

where k_b is specific force constant, l is bond length and l_0 is equilibrium bond distance

$$E_{\text{angle}} = \sum_{\text{angles}} \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{3.41}$$

where k_θ is specific force constant, θ is bond angle and θ_0 is equilibrium bond angle

$$E_{\text{torsion}} = \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) \tag{3.42}$$

where V_n is the amplitude, n is the number of minima on the potential energy surface, ω is the torsion angle and γ is the phase factor

$$E_{\text{non-bonded}} = \underbrace{\sum_i \sum_{j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)}_{\text{van der Waals}} + \underbrace{\sum_i \sum_{j>i} \frac{q_i q_j}{\varepsilon r_{ij}}}_{\text{Coulomb}} \tag{3.43}$$

where van der Waals interaction between two atoms i and j separated by distance r_{ij} is described by Lennard Jones potential with parameters A_{ij} and B_{ij}, and Coulomb potential is described by electrostatic interaction between a pair of atoms i and j using q_i and q_j as charges on atom and Є as the dielectric constant of medium.

The state of any classical system can be completely described by means of specifying the positions and momenta of the all particles:

$$\begin{aligned} \mathbf{q} &= (x_1, y_1, z_1, x_2, y_2, z_2, \ldots) \\ \mathbf{p} &= (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, p_{y,2}, p_{z,2}, \ldots) \end{aligned} \tag{3.44}$$

Since a phase point is defined by the positions and momenta of all particles, it determines the location of the next phase point in the absence of outside forces acting upon the system. Therefore, the relationship between two positions in any time interval is given:

$$q(t_2) = q(t_1) + \int_{t_1}^{t_2} \frac{p(t)}{m} dt \tag{3.45}$$

$$v = \frac{p}{m} \tag{3.46}$$

Similarly, using Newton's Second Law of Motion, the relationship between any two momentum vectors is:

$$p(t_2) = p(t_1) + m \int_{t_1}^{t_2} a(t)dt \tag{3.47}$$

$$a = \frac{F}{m} \tag{3.48}$$

MD simulations generate information (atomic positions and velocities) at the microscopic level. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics.

In statistical mechanics, time averages are defined as ensemble averages. The average value of any property during time evolution is:

$$\langle A \rangle = \frac{1}{M} \sum_{i}^{M} A(t_i) \tag{3.49}$$

where M is the number of times the property is sampled.

If the system is allowed to evolve in time indefinitely, it will eventually pass through all possible states and the above equation becomes [34, 46]:

$$\langle A \rangle = \lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(\tau)d\tau \tag{3.50}$$

assuming Ergodic hypothesis to be valid and independent of choice of t_0.

4. EFFECT OF SOLVENT MOLECULES ON MECHANISM OF ASPARAGINE DEAMIDATION

In this chapter, the effect of solvent molecules on the deamidation mechanism will be investigated. A cyclic imide-mediated route has been previously suggested, as the pathway for deamidation of Asn residues in relatively unrestrained peptides; previous computational studies on Asn deamidation have focused on modeling the aforementioned route without solvent assistance. The following study shows the favorable contribution of explicit solvent molecules on the imide-mediated mechanism of deamidation. In addition, a novel route which involves tautomerization of the Asn side chain amide into an amidic acid has been suggested as an alternative for the formation of the cyclic imide.

Deamidation of proteins occurs spontaneously under physiological conditions. Asparaginyl (Asn) residues may deamidate into aspartyl (Asp) residues, causing a change in both the charge and the conformation of peptides. It has been previously proposed by Capasso *et al.* that deamidation of relatively unrestrained Asn residues proceeds through a succinimide intermediate. This mechanism has been modeled by Konuklar *et al.* and the rate determining step for the deamidation process in neutral media has been shown to be the cyclization step leading to the succinimide intermediate. In the present study, possible *water-assisted* mechanisms, for both concerted and stepwise succinimide formation, were computationally explored using the B3LYP method with 6-31+G** basis set. Single point solvent calculations were carried out in water, by means of integral equation formalism-polarizable continuum model (IEF–PCM) at the B3LYP/6-31++G** level of theory. A novel route leading to the succinimide intermediate via tautomerization of the Asn side chain amide functionality has been proposed. The energetics of these pathways has been subject to a comparative study in order to identify the most probable mechanism for the deamidation of peptides in solution.

4.1. Introduction

The deamidation of proteins may occur under physiological conditions and is known to limit the lifetime of proteins. Deamidation is the conversion of the amide group on the side chain of an amino acid residue to a carboxylate or carboxylic acid depending on the pH of the medium. Asparagine (Asn) and glutamine (Gln), two of the 20 amino acid residues that ordinarily occur in proteins, are uniquely unstable under physiological conditions; they spontaneously and non-enzymatically deamidate into glutamyl (Glu) and aspartyl (Asp) residues [15] (Figure 4.1). Gln deamidation is usually substantially slower than Asn deamidation.

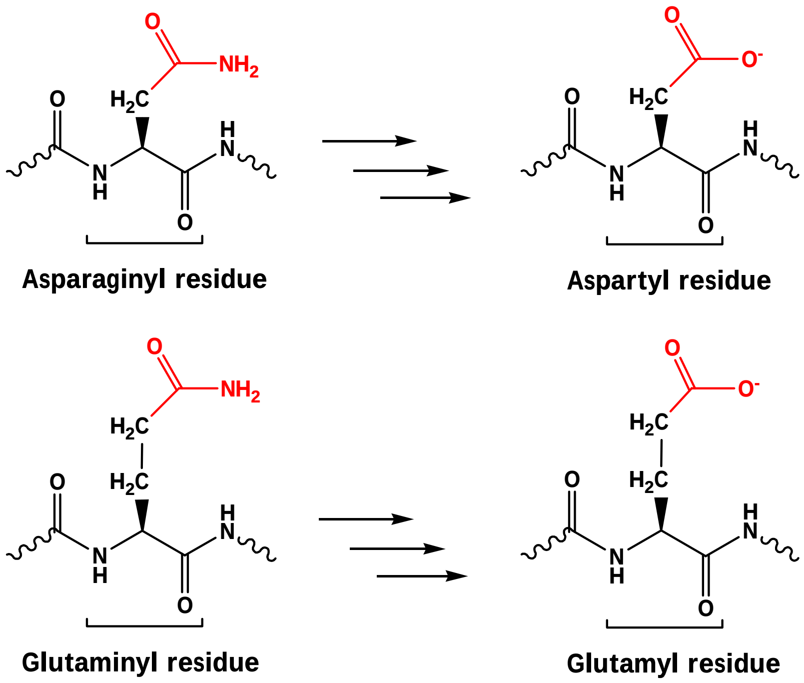


Figure 4.1. Deamidation of asparaginyl and glutaminyl residues.

The deamidation of Asn and Gln in peptides and proteins is of significant biological interest. Over 1,700 research papers on various aspects of the deamidation of peptides and proteins have been published [47] since the biological importance of deamidation was first emphasized. Deamidation of asparaginyl and glutaminyl residues causes time-dependent changes in charge and conformation of peptides and proteins [48, 49]. Deamidation rates of 1371 asparaginyl residues in a representative collection of 126 human proteins have been calculated [50]. Deamidation half-times for these proteins were shown to range from 1 – 1000 days. These rates have suggested that deamidation is a biologically relevant phenomenon in a remarkably large percentage of human proteins.

The timed processes of protein turnover, development, and aging have been suggested as possible roles for deamidation [51]. It has been hypothesized by Robinson *et al*. that deamidation serves as a molecular clock for the timing of biological processes [52]. The fact that static properties of asparaginyl and glutaminyl residues are not unique and can be easily duplicated by some of the other 18 commonly occurring amino acid residues suggestively indicates the essence of their disruptive effect on peptide and protein structure by deamidation reactions. Furthermore, it has been proposed by Robinson *et al*. [52] that the instability of asparaginyl and glutaminyl residues is their primary biological function, and that they serve as easily programmable molecular clocks.

Capasso *et al*. have proposed that the deamidation of relatively unrestrained Asn residues proceeds through a succinimide intermediate [53, 54, 19] (Figure 4.2). Acylation of the amino group on the neighboring ($n+1$) carboxyl side residue by the β-carbonyl (side chain carbonyl) group of the L-Asn residue produces a five membered cyclic imide, namely the succinimide intermediate. The first step in the mechanism (Figure 4.2) includes the cyclization and consecutive loss of an NH_3. The succinimide intermediate then hydrolyzes at either one of the two carbonyls. Experimental findings indicate that the hydrolysis reaction gives L-Asp and L-β-Asp (L-iso-Asp) in a 3:1 ratio [55].

The deamidation of Asn residues via succinimide intermediates has been previously investigated theoretically [56] using the Density Functional Theory (B3LYP/6-31G*). The cyclization, deamination (loss of NH_3) and hydrolysis reactions that lead to the deamidation of a model peptide have been studied and the succinimide intermediate suggested by Capasso *et al.* [53] has been confirmed. The formation of the succinimde intermediate has been proposed to be a multi-step process [56], in which the initial cyclization step forms a tetrahedral intermediate (Figure 4.3). This is followed by the deamination step to produce the succinimide ring. Finally, hydrolysis of the succinimide ring leads to L-Asp or L-iso-Asp residues [57, 58].The rate determining step in neutral media has been proposed to be the formation of the tetrahedral intermediate [56].

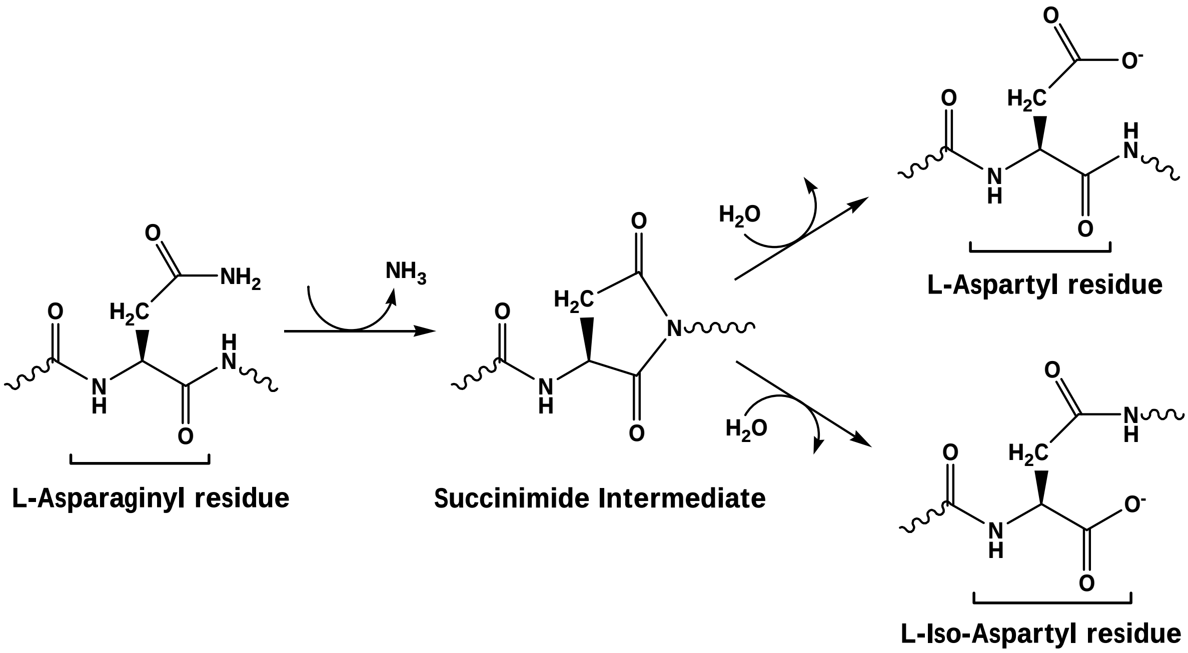


Figure 4.2. Deamidation mechanism.

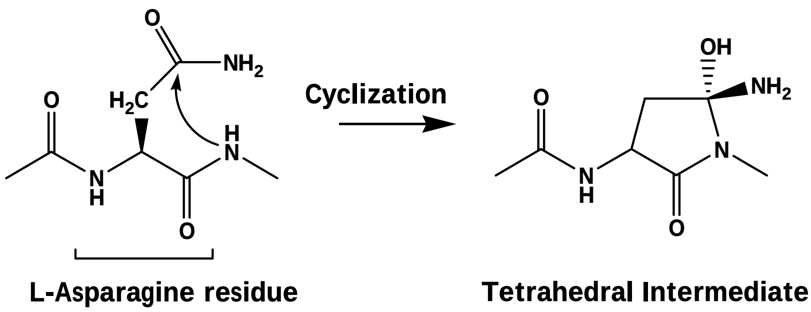


Figure 4.3. Rate determining step for deamidation in neutral media.

An experimental study has been performed by Robinson *et al.* on the complete 800-pentapeptide set of all possible combinations of the sequences Gly-Xxx-Asn-Yyy-Gly and Gly-Xxx-Gln-Yyy-Gly, where Xxx and Yyy are any of the 20 normally occurring amino acid residues [59]. This study has indicated that deamidation rate is controlled primarily by the carboxyl side residue (Yyy), also commonly referred to as the *n+1* residue, with smaller effects from the amino side residue (Xxx). This is consistent with the succinimide reaction mechanism that was originally proposed [53] to explain the deamidation rates of Asn-Gly sequences and iso-Asp formation.

In the present study, alternative pathways, namely *water-assisted* deamidation mechanisms have been investigated. The effect of explicit H_2O molecules on the mechanism and energetics of deamidation has been explored using computational techniques. Both concerted and stepwise reaction mechanisms leading to the succinimide intermediate have been taken into account. As an alternative mechanism the tautomerization of the Asn side chain amide functionality has been explored. Activation barriers for these pathways have been used for comparative purposes and to identify the most probable mechanism for deamidation of peptides in solution.

4.2. Computational Methodology

Preliminary analysis of the potential energy surfaces (PES) for all proposed mechanisms were carried out at a semi-empirical level (PM3) [24]. Further geometry optimizations were performed using the density functional theory (DFT) [34] at the B3LYP/6-31+G** [37, 38, 40] level of theory. Geometries of stationary points were optimized without any constraints. All stationary points have been characterized by a frequency analysis from which zero-point energy and thermal corrections have also been attained using the ideal gas approximation and standard procedures. Local minima and first order saddle points were identified by the number of imaginary vibrational frequencies. The intrinsic reaction coordinate (IRC) approach [60, 61], followed by full geometry optimization, has been used to determine the species reached by each transition structure. Note that the resulting energy minima do not necessarily correspond to the lowest conformation of the system (particularly for solute-water complexes), though differences are not expected to be large. Free energies of activation ($\Delta G^{\ddagger}$) are calculated as the difference of free energies between transition states and reactive conformers reached by IRC calculations. All energy values for gas phase optimized structures listed throughout the discussion include thermal free energy corrections at 298 K and 1 atm.

The effect of a polar environment on the reaction path has been taken into account by use of the self-consistent reaction field (SCRF) theory. Single-point energies in water ($\varepsilon = 78.5$) utilizing the integral equation formalism-polarizable continuum (IEF-PCM) model [41, 43-45] at the B3LYP/6-31++G** level were calculated on gas phase B3LYP/6-31+G** optimized structures. Bondi radii [62] scaled by a factor of 1.2 have been used for solvent calculations. All solvent energies reported include thermal corrections to free energies, obtained from gas phase optimizations and non-electrostatic corrections.

All gas phase optimizations and single point solvent calculations have been carried out using the Gaussian 03 program package [63]. Reaction mechanisms shown in figures throughout the text contain gas phase optimized geometries (B3LYP/6-31+G**) of reactive conformers, transition states and products, respectively. All distances shown in the figures are in Angstroms (Å).

4.3. Results and Discussion

In the first part of this study, the succinimide formation mechanism proposed by Capasso *et al*. and previously modeled by Konuklar *et al*. has been computationally revisited with a slightly larger model compound (Figure 4.4), in order to more efficiently mimic Asn and its neighboring residues. More specifically, instead of mimicking the backbone NH with an NH_2 group as in the previous model, an acetyl ($CH_3C{=}O$) group has been added to the Asn backbone NH, to help prevent unrealistic intramolecular H-bonding that can be formed by a less restricted NH_2. The basis set has been improved by the addition of diffuse functions on heavy atoms and polarization functions on hydrogen atoms, the method was preserved. The concerted mechanism will serve as a benchmark for comparison with energetics of the newly proposed mechanisms in this study.

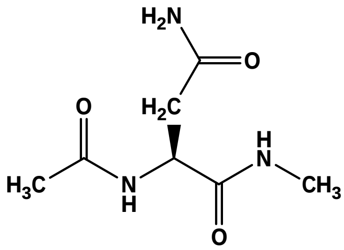


Figure 4.4. Model peptide with L-asparaginyl residue.

The previously suggested concerted mechanism has been employed on the new model compound using B3LYP/6-31+G** (Figure 4.5). This is a four-centered asynchronous concerted mechanism, where the hydrogen on the backbone NH, which belongs to the *n+1* residue, is transferred to the Asn side chain carbonyl group early in the reaction as seen in the transition state. Later in the concerted step, the backbone nitrogen attacks the carbonyl carbon and the five-membered tetrahedral intermediate forms. The free energy of activation ($\Delta G^{\ddagger}$) for this concerted reaction is quite high, 49.7 kcal/mol in gas phase, for a reaction that spontaneously and non-enzymatically occurs under physiological conditions. The same concerted reaction was shown to have an activation barrier of 50.3 kcal/mol in gas phase (47.5 kcal/mol in solution) in the previous study where both a smaller model compound and basis set were used.

Hence, there is reason to believe that this reaction may occur through different reaction mechanisms with lower barriers, assisted by solvent molecules that are in the vicinity of the asparaginyl residue. In the next part of this study, three different reaction mechanisms for the cyclization leading to the tetrahedral intermediate are proposed and computationally explored, all incorporating the effect of explicit H_2O molecules on the reaction mechanism. In the last part, the deamination mechanism (loss of NH_3) and the effect of solvent molecules on the reaction barrier and energetics are discussed.

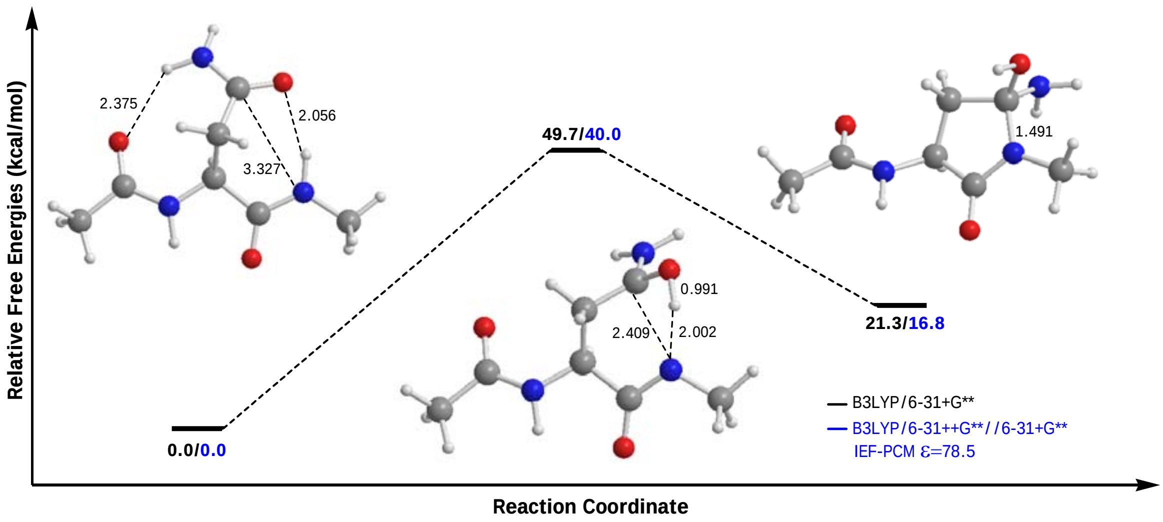


Figure 4.5. Potential free energy profile for the concerted cyclization reaction.

4.3.1. Cyclization

4.3.1.1. Concerted - *Water Assisted* – Cyclization. The first water assisted cyclization mechanism studied is the concerted cyclization step assisted by a single water molecule. The water molecule helps the transfer of the hydrogen from the backbone NH of the *n+1* residue, onto the Asn side chain carbonyl oxygen (Figure 4.6).

In the optimized geometry for the reactive conformer, the water molecule is shown to form two H-bonds with the model peptide, one with the carbonyl of the Asn side chain, and another with the backbone NH of the neighboring carboxyl side residue. In the transition state structure of this concerted reaction, the hydrogen transfer occurs early in the transition state and ring closure via attack of backbone *n+1* N to carbonyl C of Asn side chain is observed later in the concerted step.

The energetics of this reaction shows that the concerted reaction can go through a lower barrier with a water assisted pathway. The water molecule reduces the H---O distance for the proton transfer, but has little effect on the C---N distance, thus a slightly reduced but still very high barrier. The free energy of activation for the water assisted concerted cyclization step is 4.2 kcal/mol (2.9 kcal/mol in solution) lower than the concerted reaction (Figure 4.5).

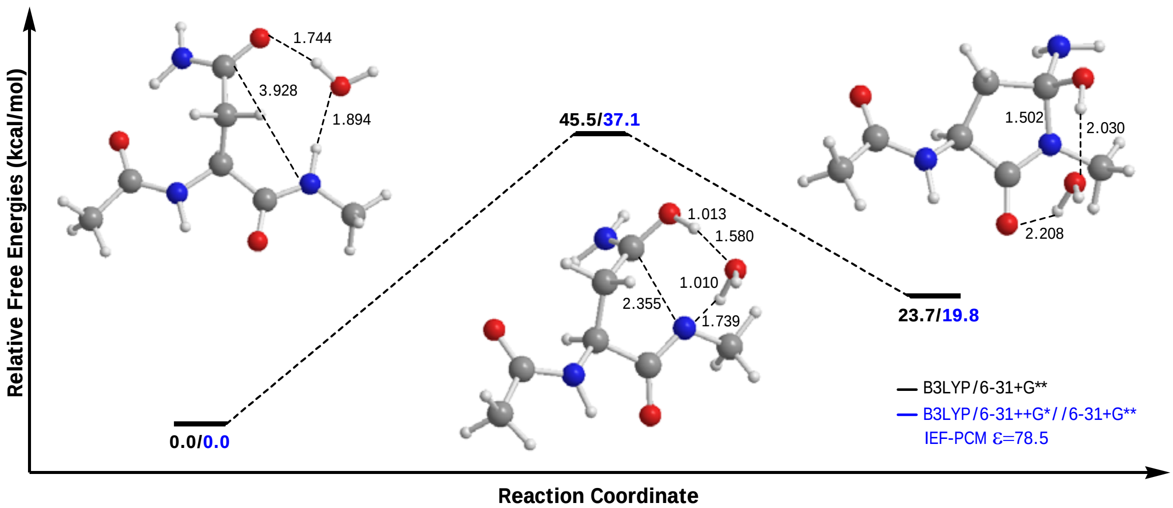


Figure 4.6. Potential free energy profile for the water assisted concerted cyclization reaction via 1 H_2O.

Another water assisted concerted cyclization mechanism -assisted by two H_2O molecules- has been modeled (Figure 4.8). In this mechanism, two H_2O molecules are actively involved in the reaction, while a third H_2O is in a passive mode. The proton is again transferred from the backbone NH of the carboxyl side residue onto the carbonyl of the Asn side chain, however this time the proton transfer takes place over two water molecules instead of one. The third water molecule is hydrogen bonded with the Asn backbone carbonyl and the Asn side chain NH_2, though this water molecule does not actively participate in the reaction, it serves as a stabilizing agent by providing hydrogen bonds. The free energy of activation for this mechanism is approximately 8 kcal/mol lower than the one water assisted concerted cyclization (Figure 4.6) and 12.1 kcal/mol lower than the originally proposed concerted cyclization (Figure 4.5). The presence of two water molecules significantly reduces the C---N distance, thus a considerable decrease in the activation barrier is observed.

4.3.1.2. Stepwise - *Water Assisted* – Cyclization. The second pathway explored was the water assisted stepwise cyclization, which consists of two consecutive steps. The first of these steps is the simultaneous deprotonation of the backbone *n+1* NH and protonation of the Asn side chain carbonyl oxygen with the assistance of water molecules. Proton transfer takes place through the water molecules via the hydrogen-bond network. The second step is the ring closure, which involves the negatively charged backbone nitrogen's attack onto the positively charged side chain carbonyl carbon, hence forming the tetrahedral intermediate (Figure 4.7).

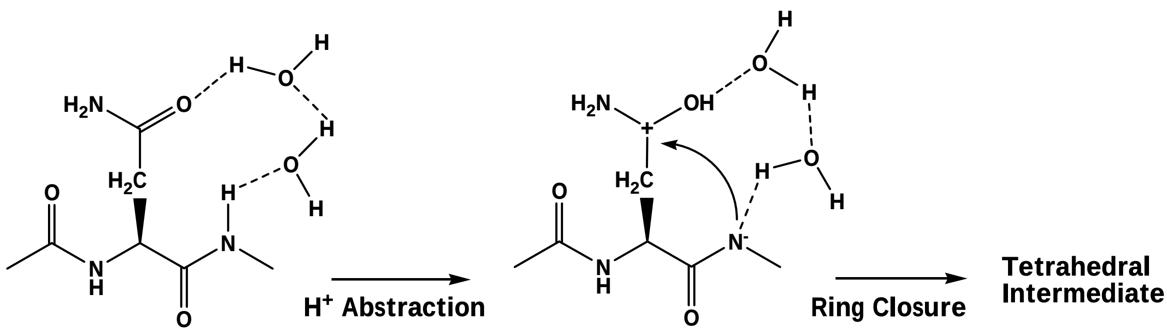


Figure 4.7. Mechanism for stepwise water assisted cyclization.

B3LYP/6-31+G** optimized geometries and free energies for both steps in the water assisted stepwise cyclization have been modeled (Figure 4.9). The first barrier shows the proton transfer over two water molecules (deprotonation/protonation). The third water molecule is not actively participating in the reaction, however it serves as a stabilizing agent, similar to the case in Figure 4.8. A zwitterionic compound is formed at the end of the first step and is stabilized by H-bonds with water molecules. The free energy of activation for this step is 26.8 kcal/mol in gas phase. The effect of a polar environment somewhat stabilizes the transition state and free energy of activation in solvent is lower, 23 kcal/mol, for this step.

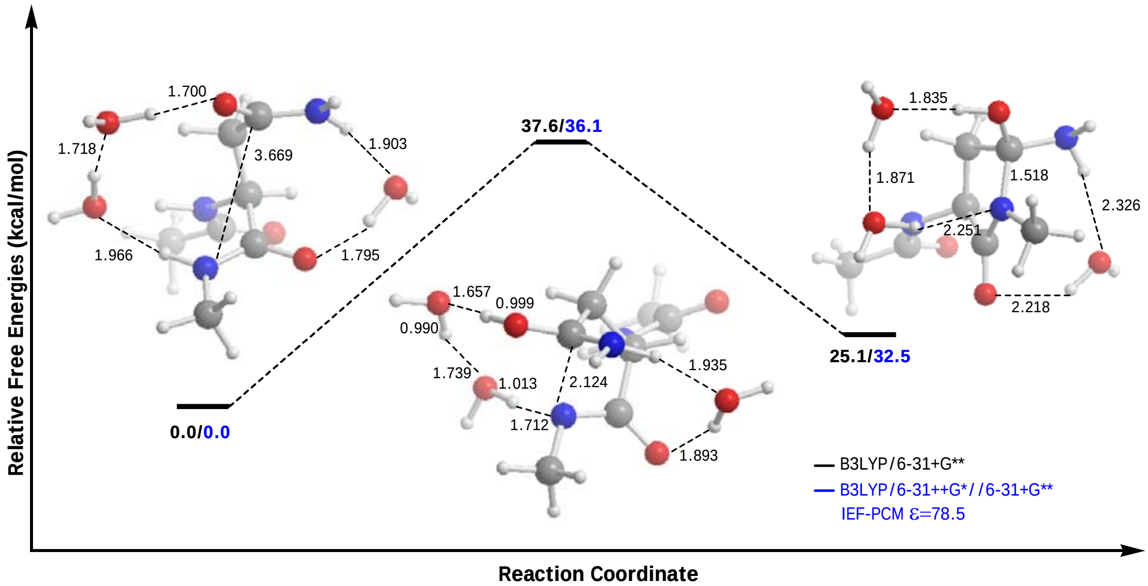


Figure 4.8. Potential free energy profile for the water assisted concerted cyclization reaction via 2 H_2O's.

It should be noted that the attempt to locate the transition state and the product for the deprotonation/protonation step via *one* peripheral water molecule failed. This is possibly due to a very flat potential energy surface between the transition state and the zwitterion formed. A similar case can be seen in the first step of Figure 4.9, where the transition state and its product have very similar geometries and the potential energy surface is quite flat. This gives us reason to believe that the stepwise cyclization mechanism is not likely to proceed towards the next step but fall back to the reactant. Nevertheless, assuming the zwitterion could be stable enough to complete the ring closure, we have investigated the next step. The formation of the tetrahedral intermediate could not be found starting from the zwitterion formed at the end of the first step. However, by a simple change in the backbone, the zwitterion can be transformed into a reactive conformer. This conformational change is a rotation around the Asn C_α – CO bond to form the reactant of the second step shown in Figure 4.9.

The backbone rotation mentioned has not been modeled in this study and it has been assumed that it will not affect the overall barrier of the water assisted stepwise cyclization appreciably, since the two intermediates are very close in energy, the latter being a little more stable. Second step in the water assisted cyclization is the ring closure. This step is expected to have a low barrier, since a negatively charged nitrogen atom is attacking a positive carbon, hence the barrier for ring closure alone is 8 kcal/mol in gas phase (12.3 kcal/mol in solution). There is an effective H-bond network between the peptide and the three peripheral water molecules throughout the reaction, with H-bond distances ranging from 1.420 – 1.965 Å.

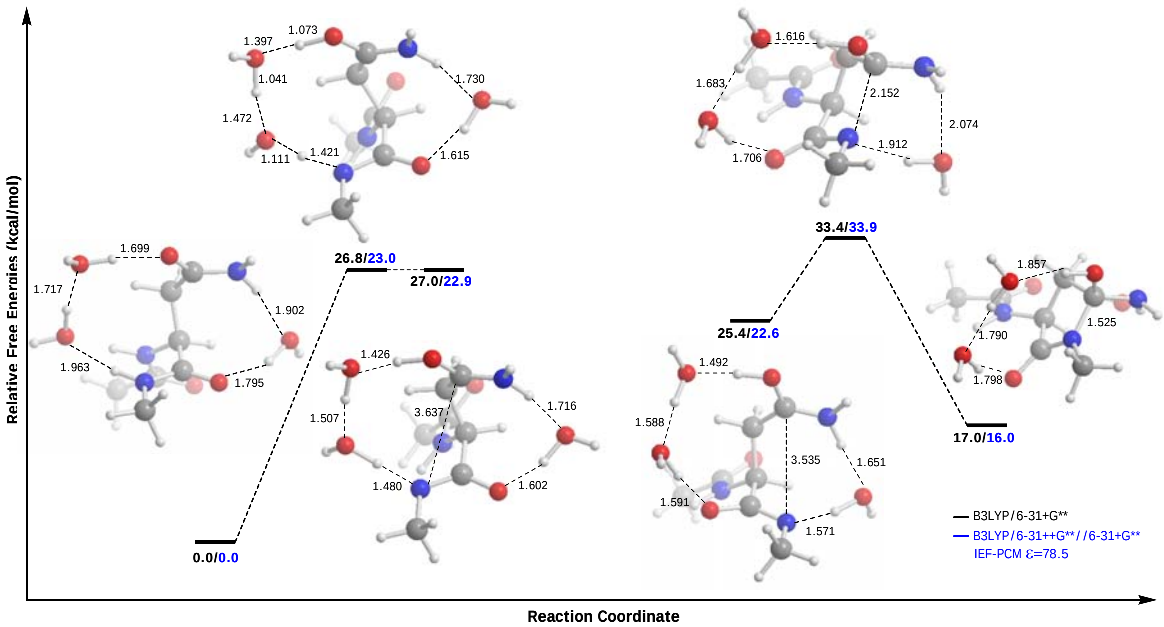


Figure 4.9. Potential free energy profile for the water assisted stepwise cyclization reaction.

The overall free energy of activation for the water assisted stepwise cyclization is shown to be 33.4 kcal/mol in gas phase (33.9 kcal/mol in solution). This energy barrier is more than 15 kcal/mol lower than the free energy of activation for the gas phase concerted reaction previously reported. It is also lower in energy than the concerted water assisted cyclization mechanisms previously discussed in this text.

4.3.1.3. Asparagine Side Chain Tautomerization. The asparagine side chain, like any other amide functionality, can tautomerize into an amidic acid tautomer, with the transfer of a proton from the side chain NH_2 group to the side chain carbonyl oxygen. The tautomerization reaction is depicted together with the ring closure that leads to the tetrahedral intermediate (Figure 4.10). This is an alternative pathway leading to the tetrahedral intermediate, which can then be converted to the succinimide intermediate through expulsion of NH_3.

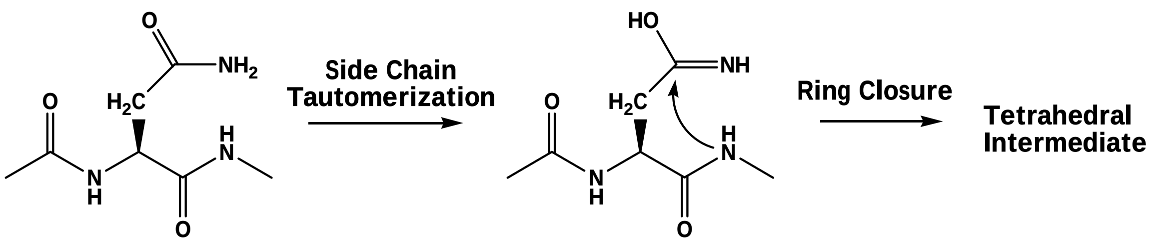


Figure 4.10. Asparagine side chain tautomerization followed by cyclization to tetrahedral intermediate.

Side chain tautomerization may occur with or without the assistance of water molecules. Asn side chain tautomerization with no help from surrounding water molecules has been modeled (Figure 4.11). This is a four-centered concerted reaction with an expected high $\Delta G^{\ddagger}$, which was calculated to be approximately 45 kcal/mol in both gas phase and solution. The tautomerization of the amide functional group to the amidic acid tautomer is not likely to proceed through a concerted four-centered reaction but rather through a water-assisted concerted one, where the proton transfer occurs via peripheral water molecules in the vicinity of the amide functionality.

Asn side chain tautomerization through one peripheral water molecule has been explored (Figure 4.12). The free energy of activation for this reaction is significantly lower than the tautomerization reaction which lacked the assistance of a water molecule (Figure 4.11). The water assisted tautomerization barrier is less than half the barrier for the four-centered reaction. The proton transfer from the NH_2 group to the carbonyl group goes through a water molecule, which is initially H-bonded to both functional groups (Figure 4.12).

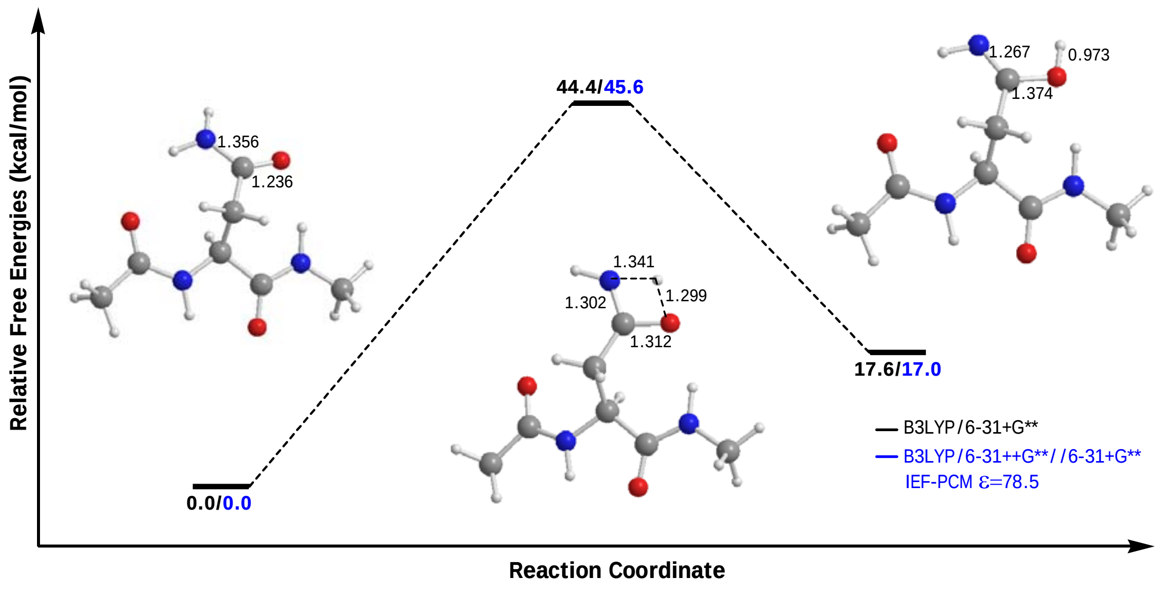


Figure 4.11. Potential free energy profile for the asparagine side chain tautomerization without explicit H_2O molecule.

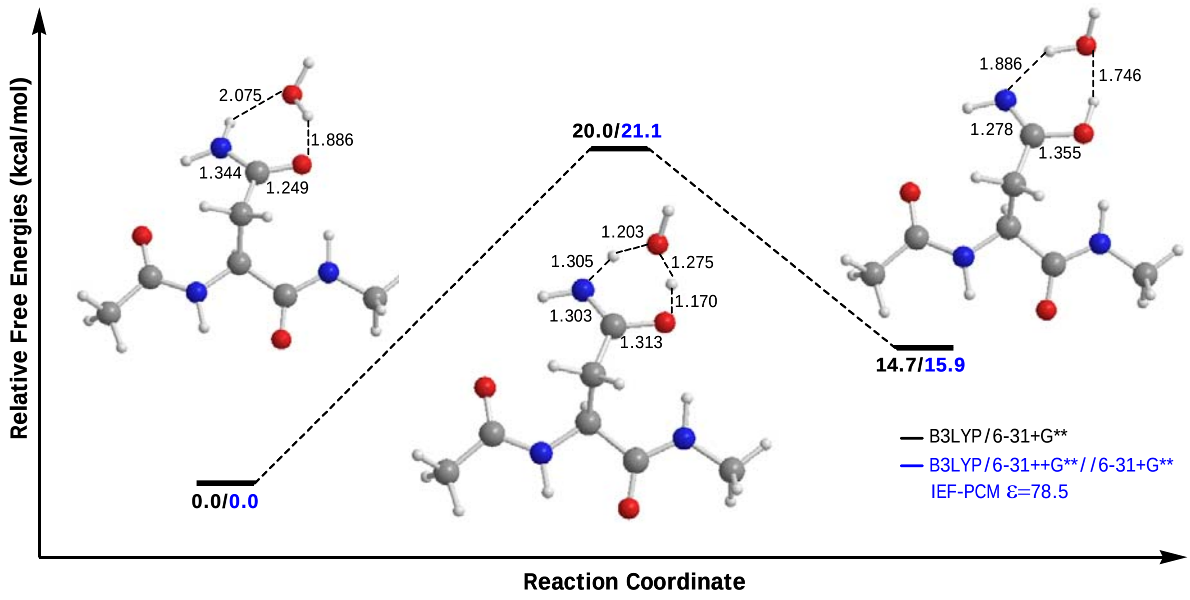


Figure 4.12. Potential free energy profile for the asparagine side chain tautomerization through 1 peripheral H_2O.

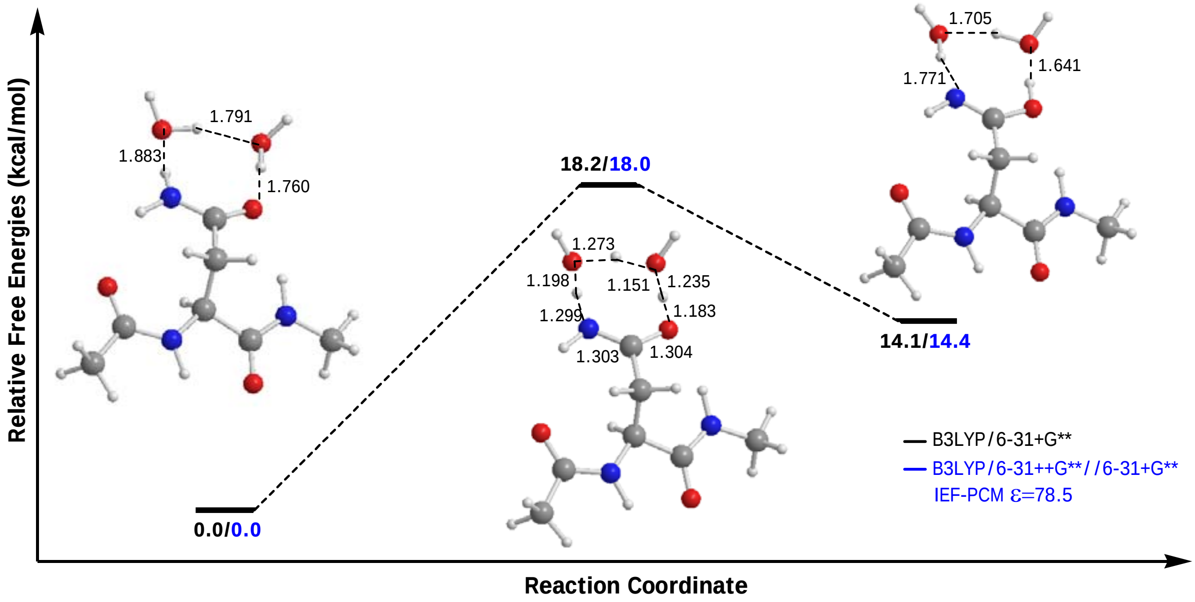


Figure 4.13. Potential free energy profile for the asparagine side chain tautomerization through 2 peripheral H_2O.

Asn side chain tautomerization through two peripheral water molecules (Figure 4.13) was shown to have an activation barrier slightly lower than the one water case (Figure 4.12). $\Delta G^{\ddagger}$ values for both the one water assisted and two water assisted tautomerizations are comparable and their difference with the barrier of Figure 4.11 indicates the necessity of modeling these reactions with explicit water molecules to better mimic the reaction in solution.

Amide tautomerization has been subject to some recently published theoretical studies on smaller amides [64, 65]. Energetics for the water assisted tautomerization reaction for Asn calculated in this study are in good agreement with these studies, which showed formamide tautomerization to be approximately 20 kcal/mol with water assistance.

<u>4.3.1.4. Ring Closure of Amidic Acid Tautomer.</u> The amidic acid tautomer formed through tautomerization of the Asn side chain may undergo ring closure to yield the tetrahedral

intermediate (Figure 4.10). The energy required for a concerted ring closure starting from the amidic acid tautomer is approximately 27 kcal/mol in gas phase, but the effect of a polar environment significantly reduces this barrier to 17.5 kcal/mol in solution (Figure 4.14)._A variant of the concerted ring closure mechanism with two peripheral water molecules was also modeled (Figure 4.15). These solvent molecules are not actively involved in the reaction, however they lower the barrier by stabilizing the transition state through a hydrogen bond network.

Ring closure of the amidic acid tautomer can also occur through a stepwise mechanism with active water assistance (Figure 4.16). In the first step of the stepwise ring closure, the backbone NH proton is transferred to the Asn side chain through a water molecule, forming a zwitterionic intermediate, which then undergoes ring closure to form the tetrahedral intermediate. The overall free energy of activation for the stepwise ring closure of the amidic acid tautomer is 4.4 kcal/mol lower than the concerted ring closure with no water assistance (Figure 4.14). The favorable interactions in Figure 4.15 and the stepwise mechanism in Figure 4.16, which prove to be helpful in reducing the activation barrier, have been combined to model the stepwise ring closure of the amidic acid tautomer with three peripheral water molecules (Figure 4.17), where only one of the water molecules is actively involved in the reaction mechanism. The three water-assisted ring closure mechanisms (Figures 4.15, 4.16, and 4.17) have comparable barriers.

It is imperative to indicate that the activation barriers associated with the ring closure of the amidic acid tautomer (Figures 4.14, 4.15, 4.16, and 4.17) cannot be directly compared with barriers of concerted (Figures 4.5, 4.6 and 4.8) or stepwise (Figure 4.9) cyclizations. It should be noted that the amidic acid tautomer must first be formed before it can undergo a ring closure to give the tetrahedral intermediate. The mechanism and energetics for the formation of the amidic acid tautomers were previously shown in Figures 4.11, 4.12, and 4.13. Hence, the reactants in Figures 4.14, 4.15, 4.16, and 4.17 are higher in energy than the reactants in Figures 4.5, 4.6, 4.8 and 4.9.

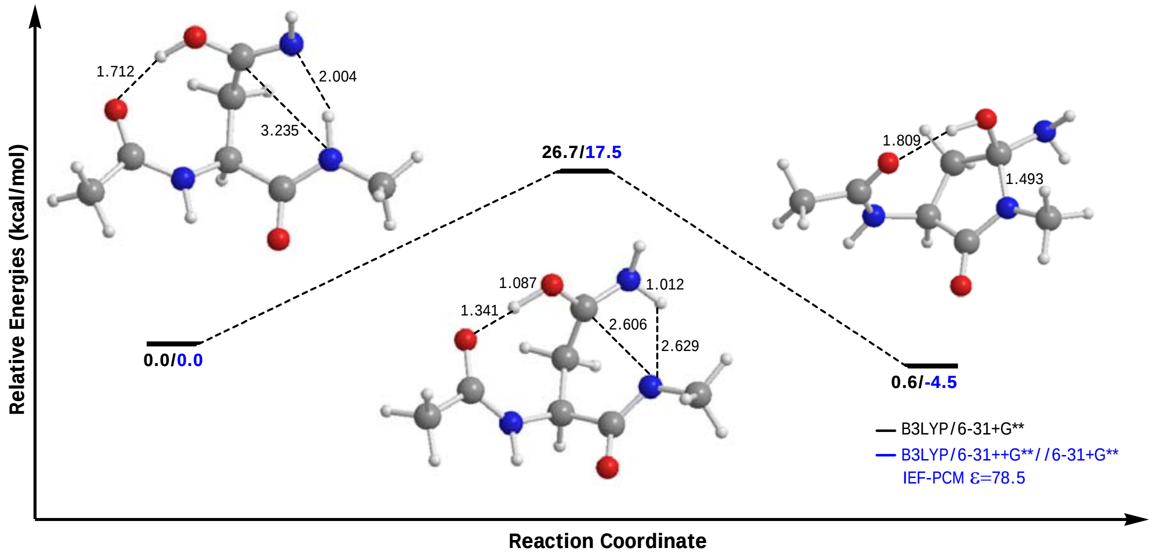


Figure 4.14. Potential free energy profile for the concerted ring closure in amidic acid tautomer.

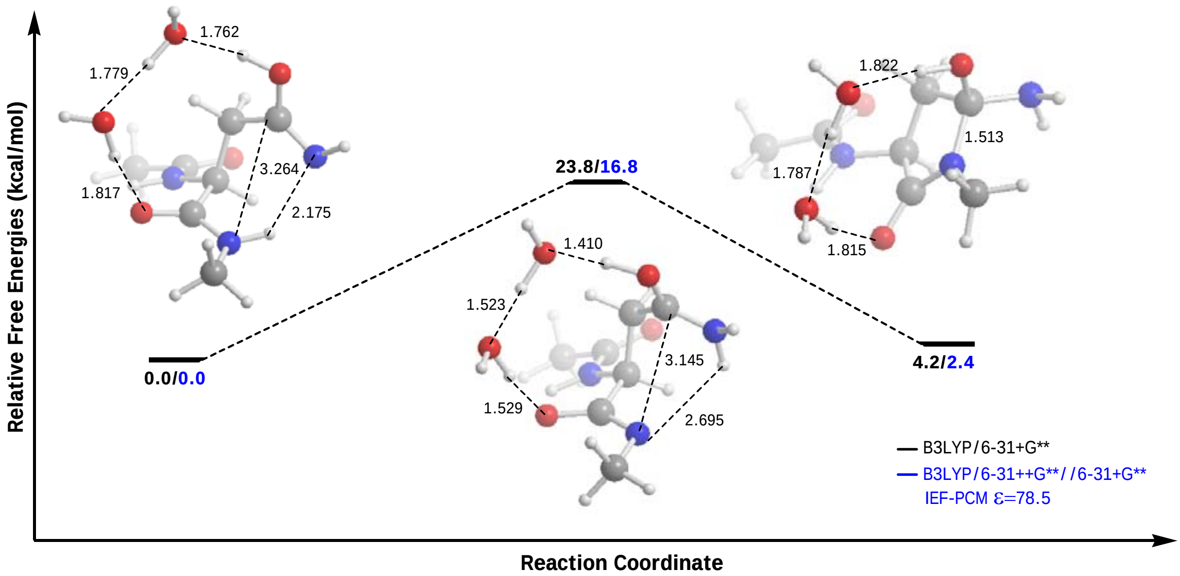


Figure 4.15. Potential free energy profile for the concerted ring closure in amidic acid tautomer with 2 H_2O passive assistance.

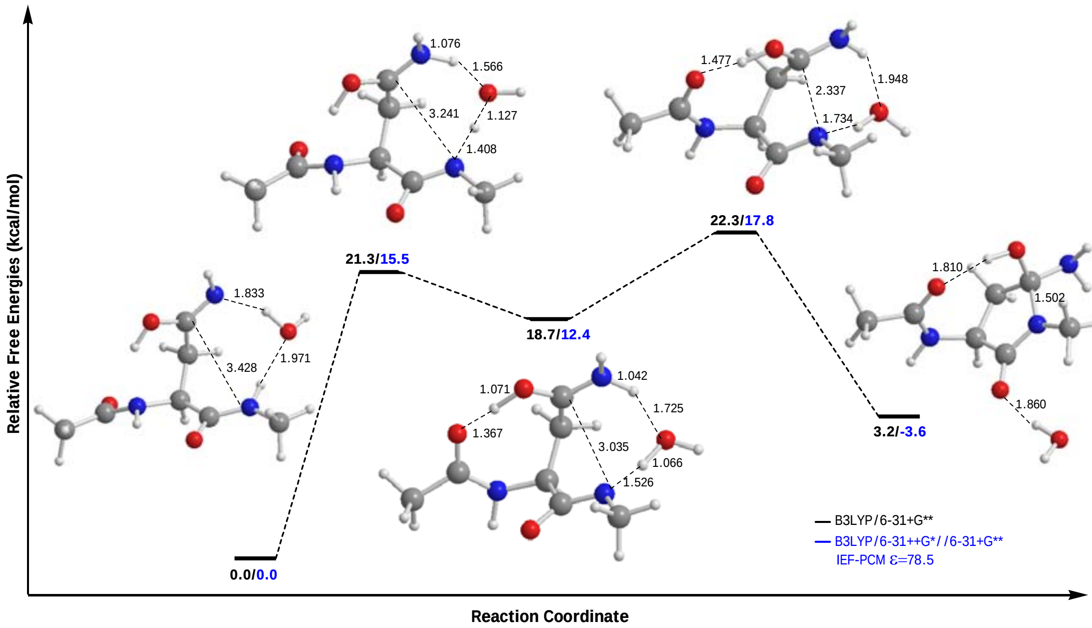


Figure 4.16. Potential free energy profile for the stepwise ring closure in amidic acid tautomer with active assistance of 1 H_2O molecule.

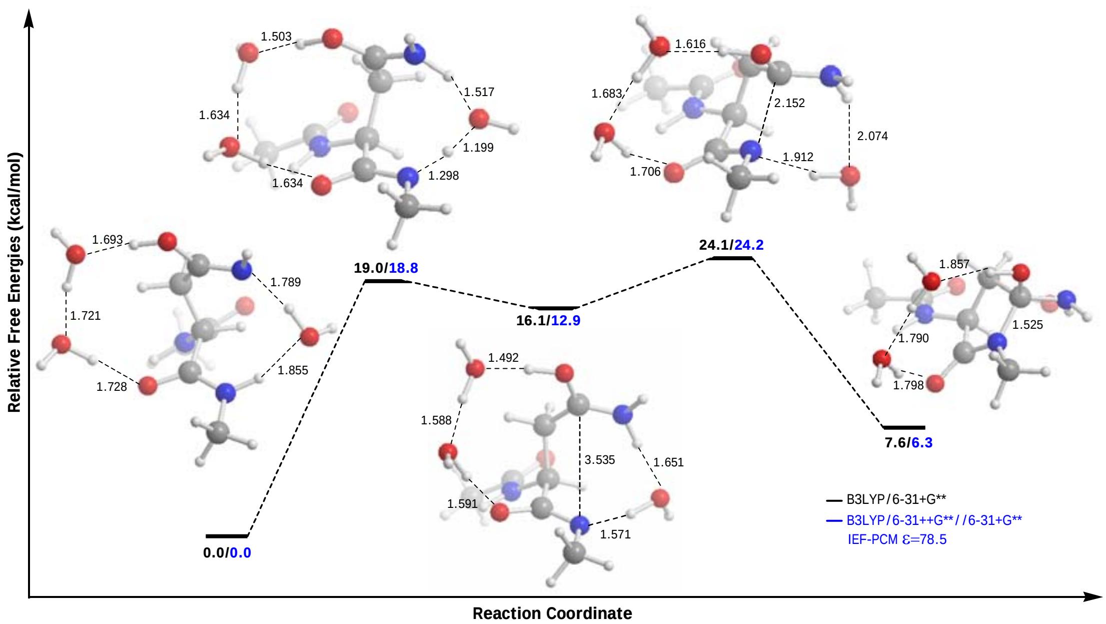


Figure 4.17. Potential free energy profile for the stepwise ring closure in amidic acid tautomer with active assistance of 2 H_2O molecule.

The concerted (Figure 4.8) and stepwise (Figure 4.9) water assisted cyclization and the stepwise water assisted ring closure of the amidic acid tautomer (Figure 4.17) all have three peripheral H_2O molecules, i.e. identical molecularity, enabling a legitimate comparison of the free energy of activation for these mechanisms. The reactants in Figures 4.8 and 4.9 have similar geometries and are energetically almost identical. For an appropriate comparison of the energetics for these three mechanisms, the overall barrier for the mechanism in Figure 4.17 was re-evaluated, taking the reactants of Figures 4.8 and 4.9 as the starting point. The overall barrier for the cyclization through the tautomerization route (Figure 4.17) starting from the reactants of Figures 4.8 and 4.9 was computed to be 33.4 kcal/mol. In light of this result, it is important to point out that the second step in both the stepwise water assisted cyclization (Figure 4.9) and the stepwise water assisted ring closure of the amidic acid tautomer (Figure 4.17) are identical. Whether the reactant undergoes a deprotonation/protonation followed by a backbone rotation (Figure 4.9) or a side chain tautomerization followed by a deprotonation/protonation step (Figure 4.17), it ends up at the same zwitterionic intermediate, which explains the identical barriers of activation for these two mechanisms. Although, it should be noted that the stepwise water assisted cyclization mechanism (Figure 4.9) is probably less likely to occur, since the first step of this mechanism is highly reversible. The concerted water assisted mechanism (Figure 4.8) has a slightly higher barrier among the three mechanisms compared. As a result, the tautomerization route could be suggested to be the most probable mechanism for the formation of the tetrahedral intermediate. However, since the difference in activation energies are small, the most prominent outcome of these results is that all water assisted mechanisms are more than 10 kcal/mol lower in energy than the previously proposed waterless concerted cyclization mechanism.

4.3.2. Deamination

The deamination step is the loss of an NH_3 molecule from the tetrahedral intermediate to give the succinimide intermediate that Capasso *et al*. have previously proposed (Figure 4.18). The NH_2 group in the tetrahedral intermediate was originally part of the amide functional group in the Asn side chain. The carbon atom bearing the –OH group is the carbonyl carbon of the Asn side chain, which has been protonated throughout the course of the reaction. A proton transfer from the –OH to the NH_2 causes consecutive departure of the amine (NH_3). In addition, as the proton is abstracted a second C=O forms on the ring, expanding the surface for the delocalization of electrons through resonance. In this way, a much more stable intermediate known as the succinimide is formed. The proton transfer may take place with or without the assistance of surrounding solvent molecules. However, proton transfer through water molecules is expected to have a substantially lower barrier of activation, as it has been shown throughout this study.

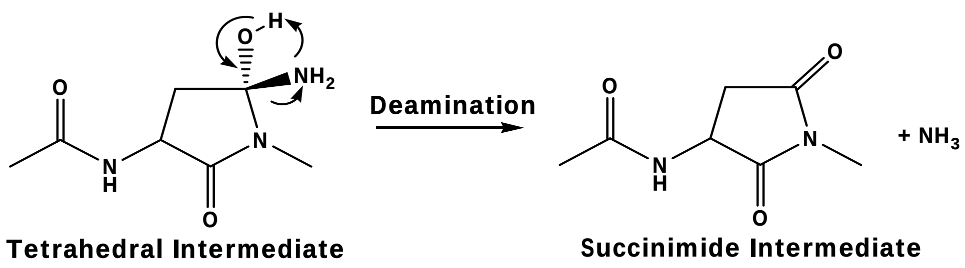


Figure 4.18. Deamination (loss of nh_3) in the tetrahedral intermediate to yield the succinimide intermediate.

In this part of the study, the deamination of the tetrahedral intermediate with no H_2O assistance, as previously proposed, has been re-evaluated with the new model compound (Figure 4.19) and a higher basis set, similar to the case for the cyclization reaction. This is a concerted four-centered step with a barrier of 31.5 kcal/mol in gas phase. If deamination is considered to go through a four-centered mechanism, comparison of energetics between cyclization and deamination suggests that deamination may as well be the rate determining step in neutral media rather than the cyclization step.

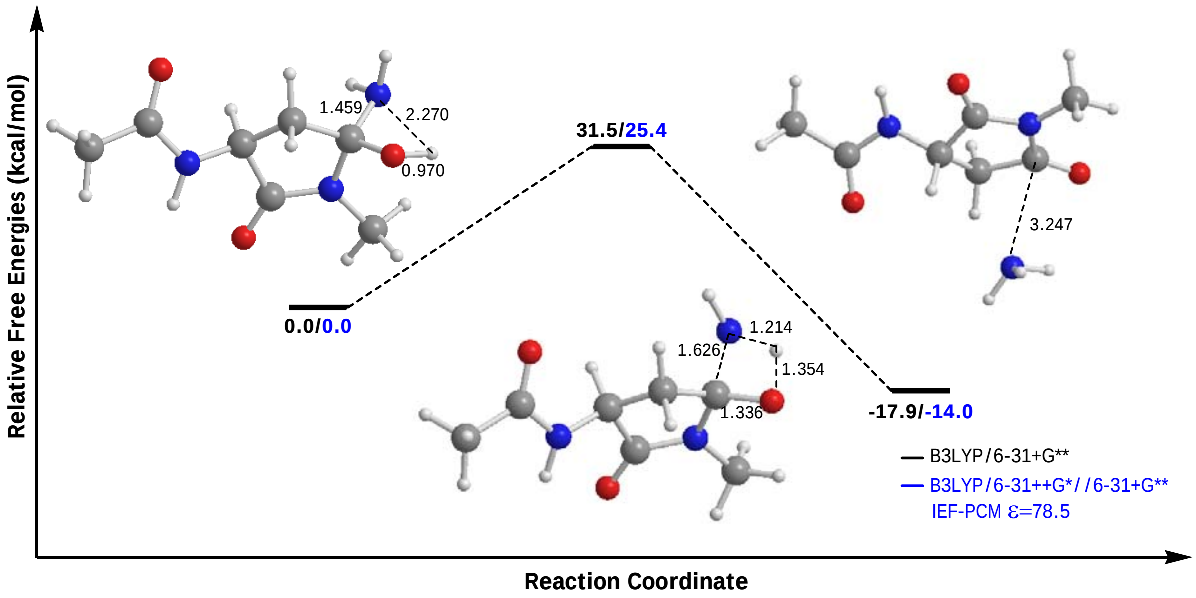


Figure 4.19. Potential free energy profile for the deamination without explicit H_2O molecules.

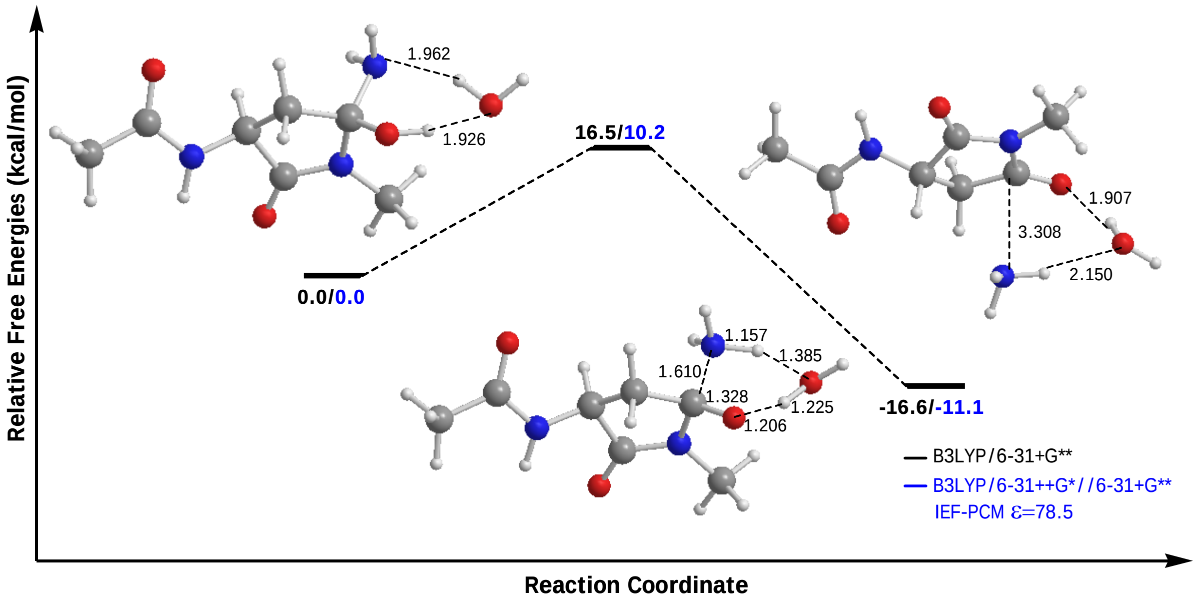


Figure 4.20. Potential free energy profile for the deamination with 1 H_2O molecule.

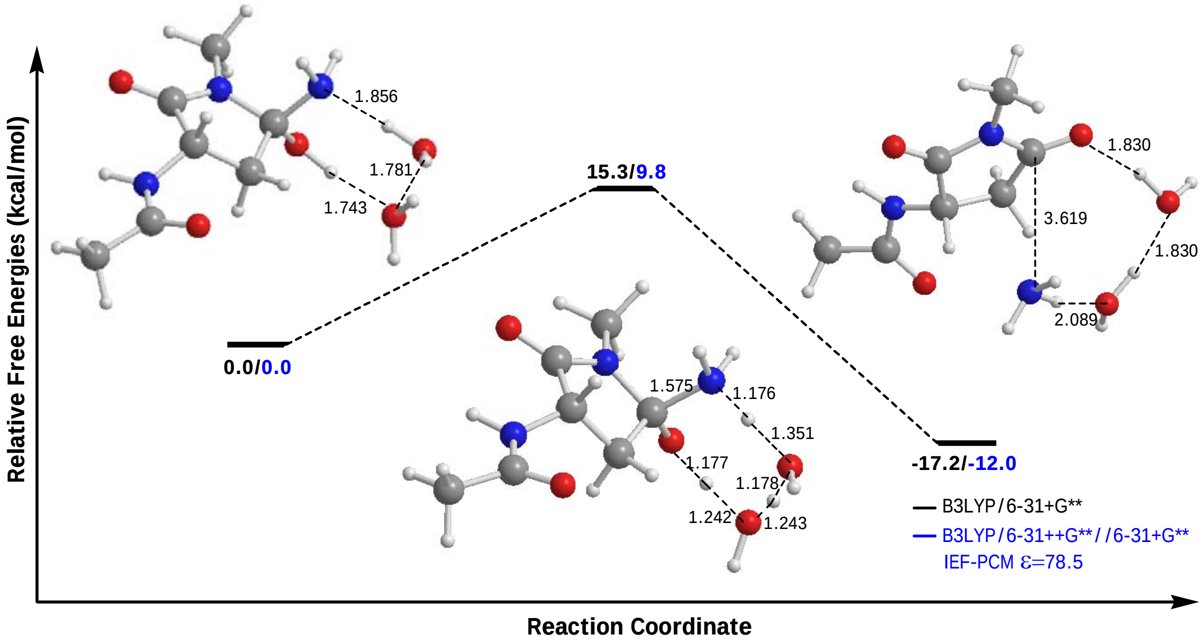


Figure 4.21. Potential free energy profile for the deamination with 2 H_2O molecules.

However, water assisted deamination mechanisms prove otherwise. The proton transfer goes through a H_2O molecule which is H-bonded to both the $–NH_2$ and the –OH involved in the reaction (Figure 4.20). The $\Delta G^{\ddagger}$ for the one H_2O assisted deamination in gas phase is half that for the 4-centered mechanism (Figure 4.19). Furthermore, the activation barrier in solution for the one H_2O assisted deamination process is as low as 9.1 kcal/mol. The deamination process through two H_2O molecules, in which the proton transfer takes place via two solvent molecules has also been explored (Figure 4.21). The energetics for both H_2O assisted deamination mechanisms are comparable. The exothermicity of the deamination reaction step is also noteworthy, indicating the stabilization in the ring.

Once the succinimide ring forms the reversal to the Asn residue is considered not feasible. NH_3 attack on the succinimide ring is not foreseen, due to the negligibly small amount of NH_3 produced during the reaction. At this point, hydrolysis is inevitable and either an Asp or an iso-Asp residue will form.

4.4. Conclusion

The aim of this study was to reinvestigate the energetics of the deamidation mechanism in peptides with the effect of solvent molecules, suggesting different pathways involving the assistance of explicit H_2O molecules.

Three different mechanisms were suggested for the cyclization step of the deamidation reaction, which was previously proposed to be the rate determining step. All water assisted cyclization reactions investigated in this study were shown to have a significantly lower barrier than the previously proposed concerted waterless mechanism. Three water assisted mechanisms with identical molecularity (Figures 4.8. 4.9 and 4.17) were shown to have overall barriers in the range of 34 – 37 kcal/mol approximately 15 kcal/mol lower in energy than the concerted waterless mechanism previously proposed. The most probable mechanism for the formation of the tetrahedral intermediate is proposed to be the tautomerization route, nevertheless since the barrier differences between these three mechanisms is small, the other two should also be considered as competitive reaction pathways.

This study has established the effect of water molecules on the deamination step of the deamidation reaction, verifying that the cyclization step, with a substantially higher barrier for activation, is the rate determining step. It is also noteworthy that the involvement of water molecules in the deamination step has lowered the barrier to half.

This study has shown that water molecules in the vicinity of asparaginyl residues serve as a catalyst in deamidation reactions. Deamidation in proteins may as well be enhanced by other residues, which are capable of accepting or donating protons. However, experimental results have shown that the three-dimensional structure of the protein accelerates deamidation in only 6% of the cases [50, 66]. Therefore, one may conclude that deamidation in proteins or enzymes will be more probable for those potential deamidating sites that exhibit the largest accessibility by solvent molecules.

This investigation also suggests that a quantitative description of the process may require carrying out a detailed statistical treatment of the solvent effect. This will be done in future studies using molecular dynamics simulations and combined QM/MM potentials. Further investigations will also include examination of 1) solvent assisted mechanisms in enzymes having sites with different deamidation rates and 2) the effect of the identity of the *n+1* residue on deamidation, which is known to have different half-times for different amino acids.

5. DIRECT HYDROLYSIS VERSUS SUCCINIMIDE-MEDIATED DEAMIDATION MECHANISMS

The previous chapter has introduced important insight on the catalytic effect of solvent molecules on the mechanism as well as energetics of the steps leading up to the succinimide intermediate. Previous computational studies on succinimide formation and succinimide hydrolysis have suggested that the rate determining step for the overall deamidation process is the cyclization step leading to the tetrahedral intermediate. However, since these studies did not take into account the effect of water molecules and activation barriers were calculated for the waterless reaction, it is imperative to re-analyze these barriers using explicit solvent molecules and to check whether water-assistance modifies the rate determining step for deamidation process. The following study, therefore, takes into account all steps of deamidation including succinimide hydrolysis. This chapter also includes the analysis of another potential mechanism for deamidation, which could compete with the succinimide-mediated pathway. Thus the direct hydrolysis of the Asn side chain amide has been investigated, in order to allow a comparative analysis between both deamidation mechanisms. The following article has recently been submitted for publication in *Chemistry – A European Journal*.

This study provides an overview of plausible mechanisms leading to the deamidation of asparagine residues and comparatively discusses relative energetics and feasibilities of these pathways. Mechanisms leading to deamidation were computationally investigated using density functional theory (B3LYP/6-31+G**). The catalytic effect of solvent molecules on barrier heights was explicitly analyzed. Mechanism and energetics of direct hydrolysis of asparagine (Asn) to aspartic acid (Asp) were compared with several succinimide-mediated mechanisms. Direct hydrolysis was indeed shown to be competitive with the succinimide-mediated deamidation routes in the absence of acid or base catalysis. Based on comparison of energetics data, the tautomerization route is the most plausible mechanism for the formation of the succinimide intermediate. The rate determining step for the formation of the succinimide intermediate is the cyclization step, regardless of mechanism. However, the hydrolysis of the succinimide intermediate is shown to have a higher barrier than succinimide formation and therefore, is suggested to be the actual rate determining step for the complete deamidation process.

5. 1. Introduction

Asparagine (Asn) and glutamine (Gln) residues are known to undergo spontaneous non-enzymatic deamidation to form aspartic acid (Asp) and glutamic acid (Glu) residues under physiological conditions [7-9, 15]. The conversion of the neutral amide side chain to the negatively charged carboxylate causes time-dependent changes in conformation and limits the lifetime of peptides and proteins [48, 49, 51]. Deamidation half-times in human proteins were shown to occur over a wide range of biologically relevant time intervals [50], and have been associated with the timed process of protein turnover, development and ageing [51]. Robinson has proposed the *molecular clock hypothesis*, which suggests that deamidation is a biological molecular timing mechanism that could be set to any desired time interval by genetic control of the primary, secondary, and tertiary structure surrounding the amide [50]. Recent experiments [19, 54, 55, 59, 66-68] and computations [56-58, 69, 70] have been in accord with this hypothesis and provided compelling evidence of its significance.

The deamidation reaction mechanism was initially believed to be an acid or base catalyzed direct hydrolysis, with a minimum reaction rate near neutral pH. Deamidation products for L-Asn and L-Gln were expected to be L-Asp and L-Glu with little racemization to D-Asp and D-Glu at basic pH. However, the pH minimum for deamidation was actually observed to be around 5 for both peptides and proteins and an L-iso-Asp product was observed in addition to the L-Asp [55]. The succinimide-mediated deamidation mechanism (Figure 5.1) is suggested to be responsible for shifting the minimum to pH 5 and for the variety of reaction products observed.

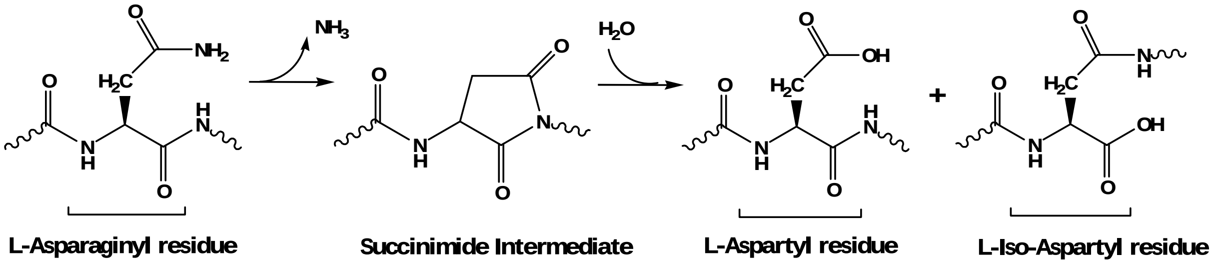


Figure 5.1. Succinimide-mediated deamidation of Asn residues.

Capasso et al. have proposed that deamidation of relatively unrestrained Asn residues goes through a succinimide intermediate (Figure 5.1) [19, 54]. The cyclic imide then hydrolyzes at either one of the two carbonyls to give Asp and iso-Asp. The ratio of L-Asp to L-iso-Asp was experimentally found to be 3:1 [55].

Experimental evidence indicates that, at low pH, hydrolysis of the side chain amide functionality occurs with ease and imide intermediates are not observed, contrary to neutral and basic pH conditions where, succinimide derivatives have almost always been identified. As pH decreases below 5, direct hydrolysis via acid catalysis takes place at an increasing rate. The fact that direct hydrolysis is most prevalent is indicated by the marked drop of the iso-Asp:Asp ratio [8]. Ordinary base catalysis also occurs at high pH, but the rapidity of the imide mechanism at high pH usually obscures this.

However, since reaction rates are often related to transient chemical species that are difficult to observe and are subject to influences of solvent and other factors, there is often uncertainty concerning a proposed mechanism, as is the case with deamidation. Nevertheless, the succinimide-mediated mechanism is supported by many experimental observations.

Previous computational studies on deamidation of Asn include modeling the formation of the succinimide as a two-step process; cyclization followed by deamination [58]. Subsequent computations on this mechanism established the fact that water molecules in the vicinity of Asn residues catalyze deamidation [71]. Radkiewicz et al. have computationally explored the racemization of Asp and Asn via succinimide intermediates [69] and have studied the effect of neighboring side-chains on backbone NH acidity [70].

Experimental findings have shown that the rate of deamidation of Asn residues is primarily controlled by the carboxyl side residue (*n* +1) with smaller effects from the amino side residue (*n* -1) [59]; this is also consistent with the succinimide reaction mechanism. However, the relationship between the size and/or charge of the *n* +1 residue and rate of deamidation is not clear. In peptides with substantial freedom of movement, sequence dependence has also been detected for other residues further along the peptide chain in both directions [59]. The effects of more distant residues are probably largely suppressed in proteins.

In light of the succinimide-mediated mechanism, deamidation reaction rates of ordinary Asn residues in physiological solvent conditions are largely affected by several factors. Firstly by the intrinsic acidity of the *n* +1 backbone nitrogen, which depends on inductive and electrostatic effects, and, therefore, upon peptide sequence. Second, by the amount of available conformational space and steric hindrance in the vicinity of the Asn residue, which may enhance or inhibit the formation of the cyclic intermediate. This is especially important in proteins where rearrangements in the protein three-dimensional structure may be necessary to allow proper alignment of the side chain and/or the backbone for ring closure. Finally, by the availability of water molecules or a proton donor, which is crucial for the decomposition of the cyclic tetrahedral intermediate that may otherwise revert to the open form. While protein structure usually inhibits deamidation, there are many instances in which protein structure near the amide allows deamidation to occur at its primary sequence controlled rate [59]. There are also relatively rare instances in which protein structure actually increases the deamidation rate [66].

This study aims to get a deeper insight on plausible mechanisms leading to deamidation of Asn residues and comparatively discuss relative energetics and feasibilities of these pathways at physiological conditions. For this purpose, several succinimide-mediated deamidation paths together with a direct hydrolysis mechanism will be considered. Some of these mechanisms are described here for the first time. Catalysis of these reactions by different number of explicit solvent molecules will be investigated.

5. 2. Computational Methodology

All gas-phase geometry optimizations were performed using the density functional theory (DFT) [34] at the B3LYP/6-31+G** level [37-40]. Diffuse and polarization functions are included on heavy atoms, since utilization of diffuse functions is especially necessary in the optimization of anionic systems; polarization functions were also added on hydrogen atoms in order to account for the presence of hydrogen-bonds. Geometries of stationary points were optimized without any constraints. All stationary points were characterized by a frequency analysis from which zero-point energy and thermal corrections were attained using the ideal gas approximation and standard procedures. Local minima and first order saddle points were identified by the number of imaginary vibrational frequencies. The intrinsic reaction coordinate (IRC) approach [60, 61], followed by full geometry optimization, was used to determine the species connected by each transition structure. Energy values for gas phase optimizations listed throughout the discussion include thermal free energy corrections at 298 K and 1 atm.

The energetics of pathways leading to deamidation investigated herein will be comparatively discussed with Asn deamidation mechanisms previously studied by the authors [71]. All structures whose energy values are reproduced from previous deamidation studies are labeled with an asterisk (*). The same model peptide, level of theory and basis set are used for a legitimate comparison with aforementioned studies. The catalytic effect of water molecules on deamidation was previously established and has been taken into account in this study.

For clarity, comparative discussion of energetics between different mechanisms is always made among species with identical molecularity, i.e. for species that are associated with the same n in the initial peptide model – $(H_20)_n$ complex. This choice minimizes the errors in the estimation of entropic contributions [72]. Relative free energies of activation ($\Delta G^{\ddagger}$) are calculated as the difference of free energies between transition states and the initial reactant. Following previous studies on amide hydrolysis [72] in water-assisted mechanisms, the initial reactant is taken as the solute-water complex with the relevant number of water molecules. Nomenclature of each transition state corresponds to the name indicated in the relevant figures and the explicit number of water molecules used in that particular step.

All calculations were carried out using the Gaussian 03 program package [63]. Structures shown throughout the text are gas-phase optimized geometries (B3LYP/6-31+G**). Distances and free energies listed in the discussion are in angstroms (Å) and kcal/mol, respectively.

5. 3. Results and Discussion

In the first part of this study, we describe different mechanisms for deamidation. The main characteristics of the molecular structures will be commented. The direct hydrolysis of the Asn side chain amide was modeled, to set a benchmark for comparison with mechanisms involving a succinimide intermediate, which are thoroughly discussed thereafter. Finally, the energetics and feasibilities of different mechanisms leading to deamidation are discussed.

Different number of explicit water molecules was used (0, 1 or 2) to analyze the effect of solvent on reaction mechanism and energetics. The initial structure for each system is the model compound or model compound-water complex (Figure 5.2). The following abbreviations are used throughout the discussion: asparagine (***asn***); succinimide intermediate (***suc***); tetrahedral intermediate (***tet***); amidic acid tautomer (***taut***); gemdiol (***gem***); aspartic acid (***asp***).

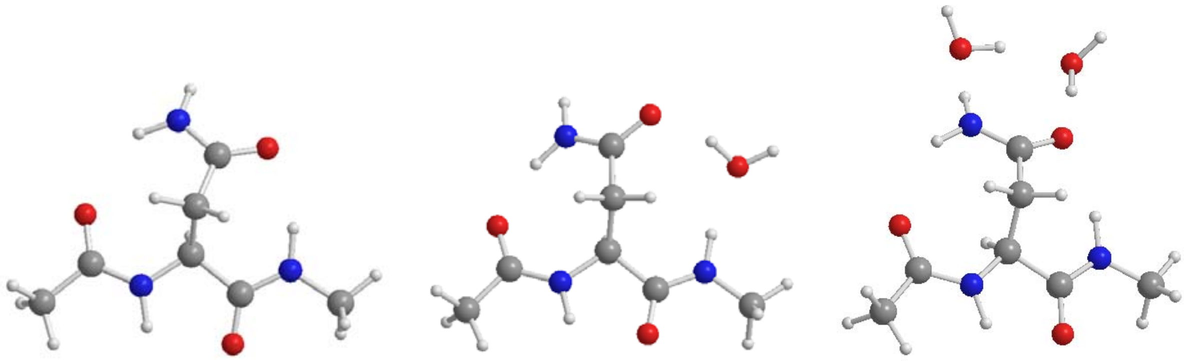


Figure 5.2. Optimized structures model peptides with Asn residue.

5.3.1. Direct Hydrolysis of Asparagine to Aspartate (*asn* → *asp*)

Amide hydrolysis has been extensively studied by computational methods, formamide hydrolysis in particular [72]. The hydrolysis of the Asn side chain to the carboxylic acid via a concerted mechanism was modeled with one or two water molecules. In the two-water case, the reaction involves a proton relay mechanism where solvent molecules serve as a conduit; amide hydrolysis involves the expulsion of an NH_3 group and deamidation takes place in a single step (Figure 5.3).

Optimized structures are shown in Figure 5.4. Given that the water molecule is a reactant, **TS1-1H$_2$O** is in fact an "unassisted" case, while **TS1-2H$_2$O** shows one-water assistance. Transition state structures show that the proton transfer from the water molecule to the leaving group (-NH_2) is underway in both cases. The C-N amide distances are rather short (approximately 1.5 Å), indicating reactant-like transition states.

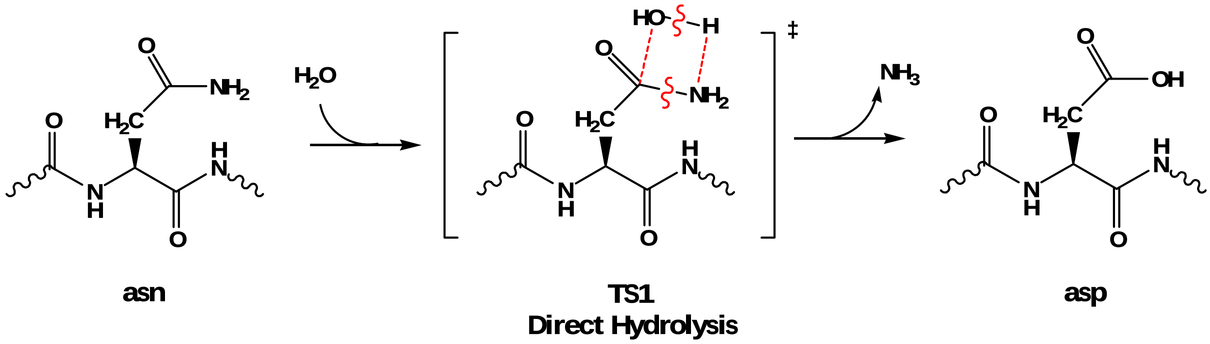


Figure 5.3 Direct hydrolysis of Asn to Asp.

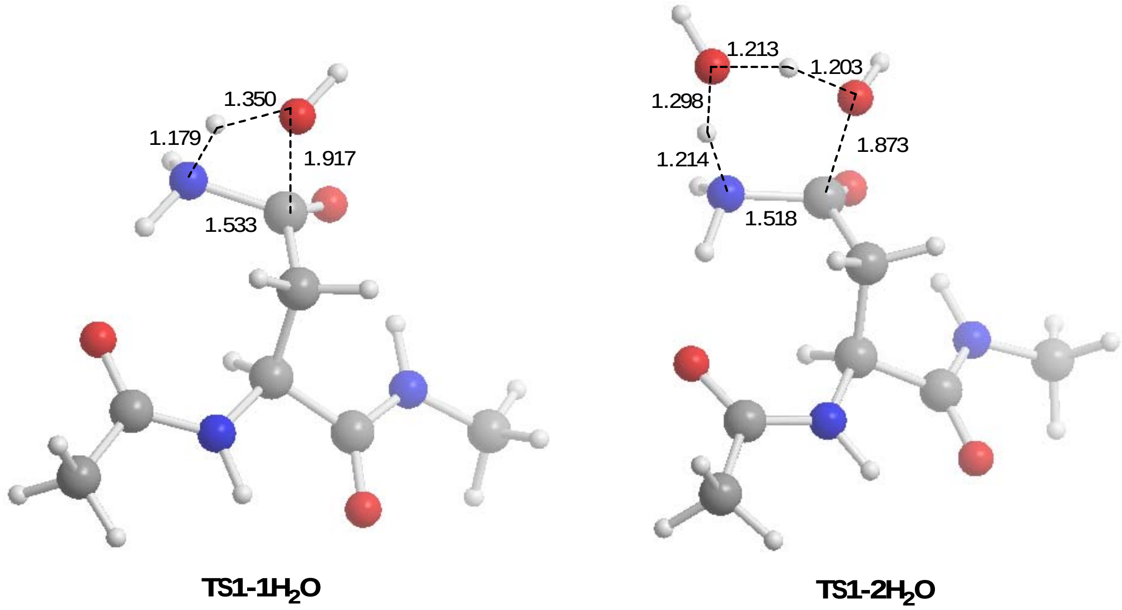


Figure 5.4. Optimized geometries and free energies of activation for the transition state of direct hydrolysis of Asn to Asp (***asn*** → ***asp***) with one (**TS1-1H$_2$O)** and two (**TS1-2H$_2$O)** water molecules, respectively.

5.3.2. Succinimide-Mediated Deamidation of Asn to Asp

Succinimide-mediated deamidation consists of two main steps; the formation of the succinimide intermediate followed by the hydrolysis of the imide. The formation of the succinimide intermediate may be achieved via several mechanisms involving a different number of steps that are discussed below. Some of these mechanisms have already been described in the literature and will be only outlined in this section. The detailed structures corresponding to these mechanisms can be found in the original references. The energies will be included in the next section (marked by an asterisk).

5.3.2.1. Succinimide Formation - *Single Step (asn → suc).* The first mechanism is the concerted ring closure accompanied by deamination, in which the n + 1 backbone NH transfers its H to the Asn side chain NH_2. Ejection of an ammonia molecule and formation of the cyclic imide occur in a single step (Figure 5.5).

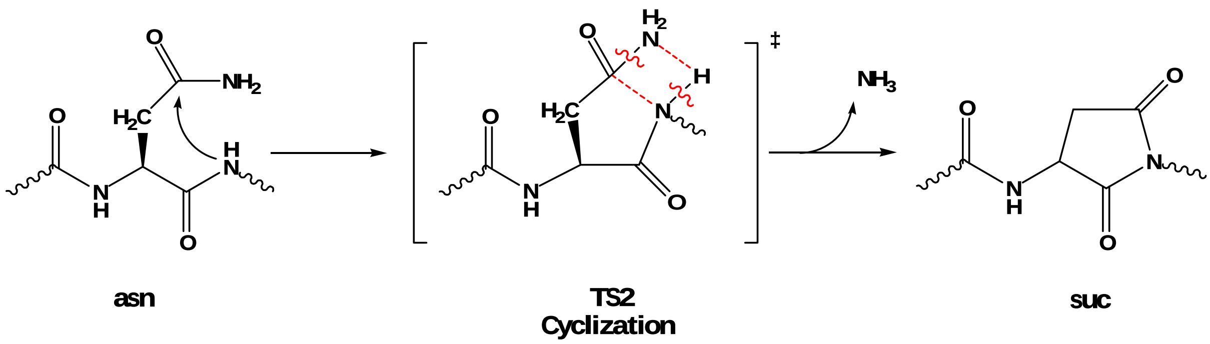


Figure 5.5. Succinimide formation – single-step mechanism (***asn*** → ***suc***).

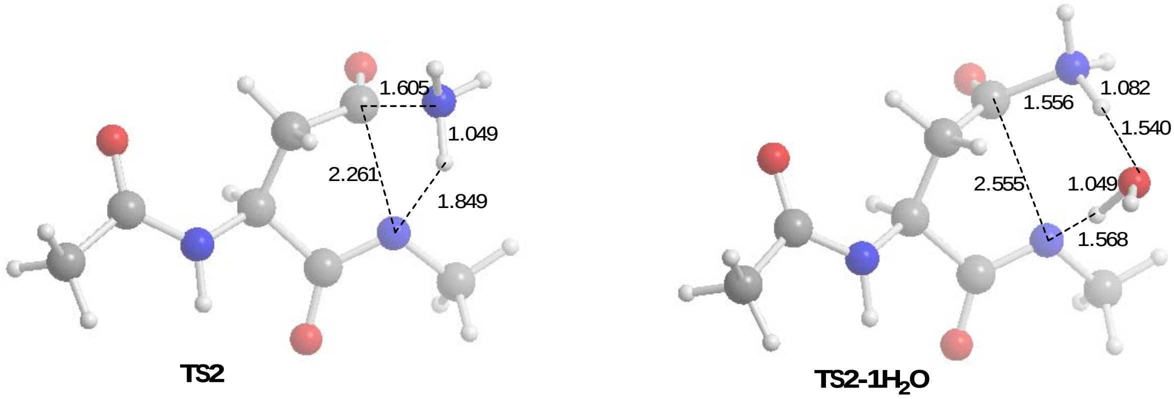


Figure 5.6. Optimized geometries for the transition state of succinimide formation through a single step (***asn*** **→** ***suc***), waterless (**TS2)** and one-water (**TS2-1H$_2$O**) assisted mechanisms, respectively.

The main difference between the water assisted (**TS2-1H$_2$O**) and non-assisted (**TS2**) cases (Figure 5.6) is the extent of ring closure, the C-N distance is considerably shorter in **TS2** (2.261 Å as opposed to 2.555 Å in **TS2-1H$_2$O**). In addition, the leaving group is further detached from the side-chain carbonyl carbon (1.605 Å as opposed to 1.556 Å in **TS2-1H$_2$O**).

5.3.2.2. Succinimide Formation - *Two Step (asn → tet → suc).* This mechanism has been previously described [26]. The succinimide intermediate can form via a two-step mechanism, in which a tetrahedral intermediate is involved (Figure 5.7). A proton is transferred from the *n* +1 backbone NH to the Asn side chain carbonyl oxygen, instead of the side chain NH_2 as in the previous mechanism. In the second step (deamination) the tetrahedral intermediate transforms into a succinimide.

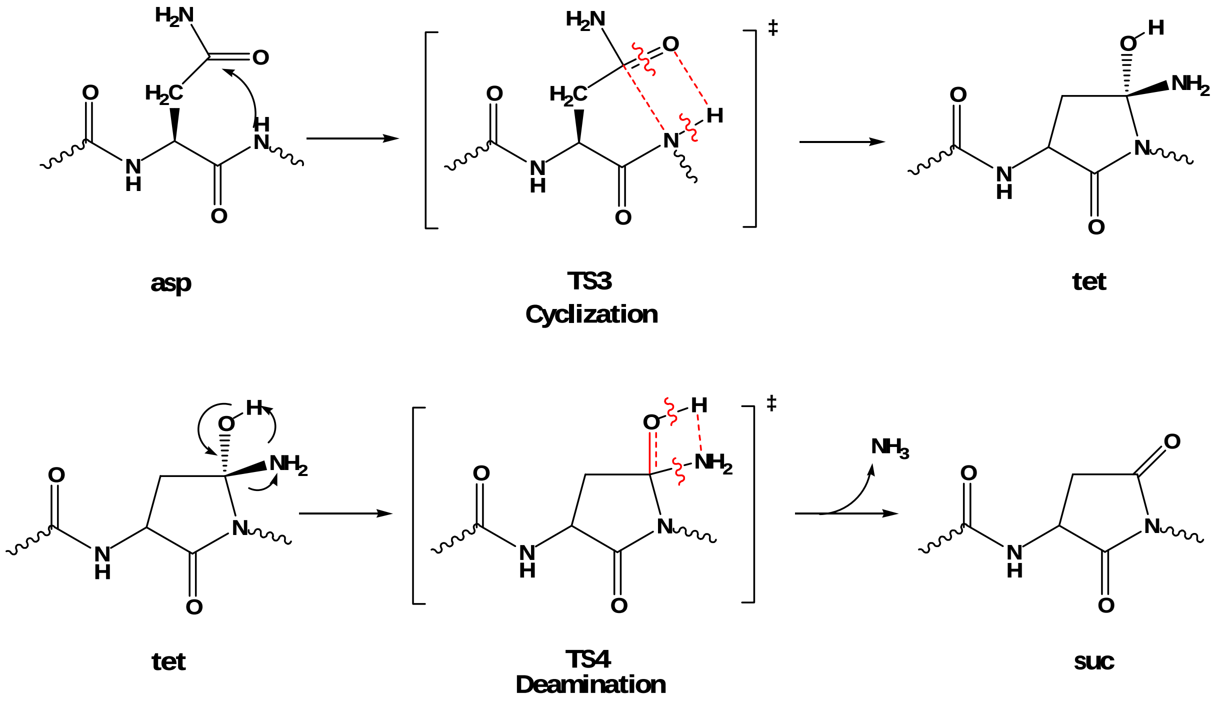


Figure 5.7. Succinimide formation – two-step mechanism (***asn*** → ***tet*** → ***suc***).

<u>5.3.2.3. Succinimide Formation - *Three Step (asn → taut → tet → suc).*</u> This mechanism has also been previously described [71]. The succinimide intermediate may also be formed via a tautomerization route; this mechanism requires three steps and goes through two intermediates. The first step is the tautomerization (***asn*** → ***taut***) of the Asn side chain amide into an amidic acid tautomer (Figure 5.8). The tautomer then undergoes ring closure (***taut*** → ***tet***) to give the tetrahedral intermediate. Finally, the cyclic tetrahedral intermediate deaminates to form the succinimide intermediate (***tet*** → ***suc***) through a process identical to the one discussed in the previous section (Figure 5.7).

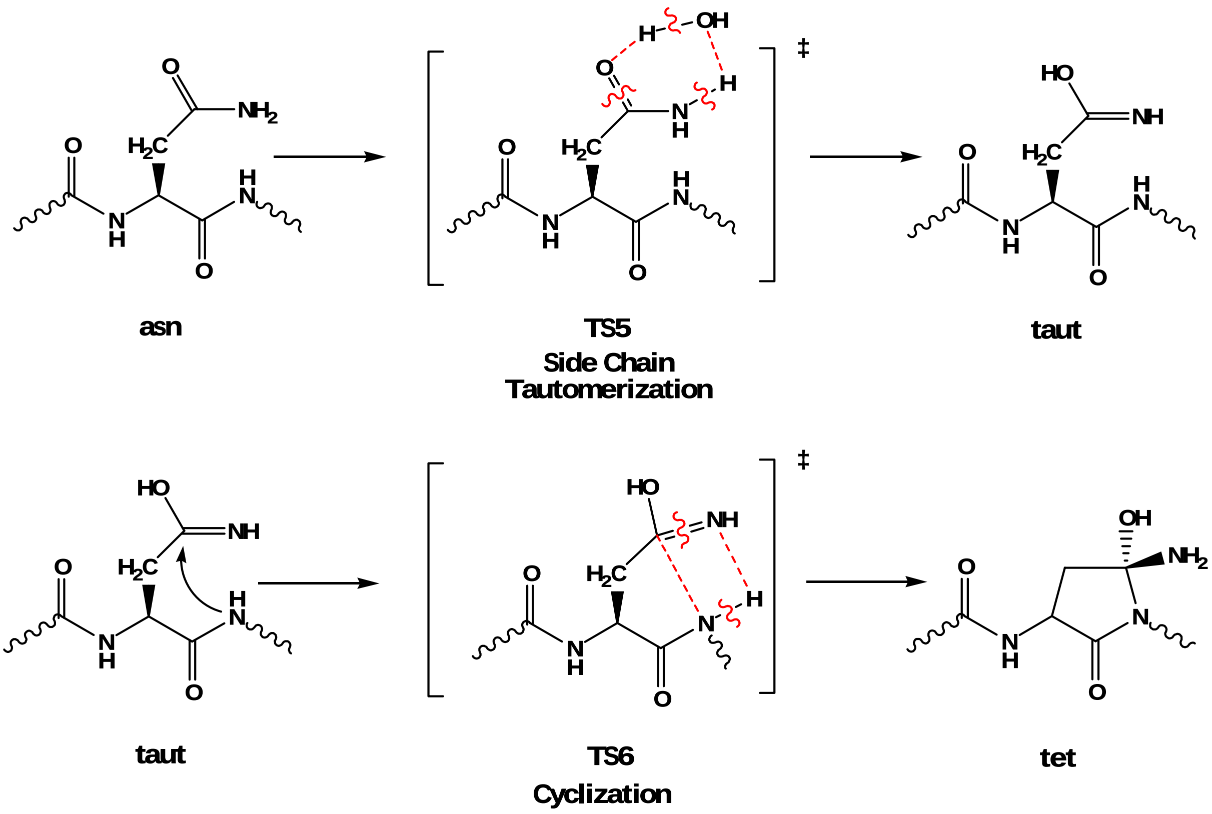


Figure 5.8. Succinimide formation – three-step mechanism (***asn*** → ***taut*** → ***tet*** → ***suc***).

5.3.2.4. Hydrolysis of the Succinimide Intermediate (*suc* → *asp*). Deamidation is completed when the succinimide intermediate undergoes hydrolysis and an Asp residue forms. It is noteworthy to indicate that the hydrolysis may take place at either one of the carbonyl groups on the succinimide. However products will be different, Asp and iso-Asp may both form; when iso-Asp forms, the peptide backbone is altered (Figure 5.1). An extra atom on the peptide chain will be an unstabilizing factor in a protein three-dimensional structure, as it might disrupt intermolecular interactions within the chain. The formation of iso-Asp has been previously investigated computationally [57] and will not be discussed in this study, since the aim is to find the most plausible pathway for the complete deamidation of Asn into Asp.

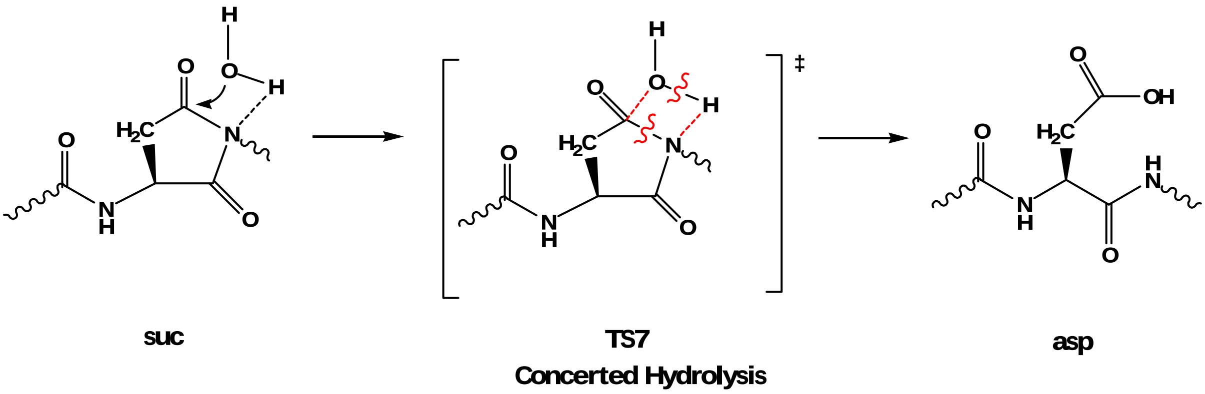


Figure 5.9. Concerted hydrolysis of the succinimide intermediate (***suc*** → ***asp***).

Hydrolysis of the succinimide intermediate can take place via a concerted ring opening (Figure 5.9) or through a gemdiol intermediate (Figure 5.10); both will give the same product. Previous studies on amide hydrolysis have shown that a stepwise mechanism going through a gemdiol intermediate has a considerably lower barrier than a concerted reaction. These studies have also revealed that water molecules catalyze hydrolysis of an amide, as mentioned earlier.

In the concerted amide hydrolysis reaction, a water molecule attacks the ring carbonyl, the N-C bond breaks, as a proton is transferred from the water molecule to the ring NH. As a result, a carboxylic acid forms (Figure 5.9). However, in the gemdiol-mediated stepwise mechanism (Figure 5.10), the initial step is the addition of a water molecule to the ring carbonyl, forming a gemdiol intermediate, which consequently undergoes ring opening in the second step to reveal the same product. In this way, the transformation of the Asn side chain amide into a carboxylate group is complete.

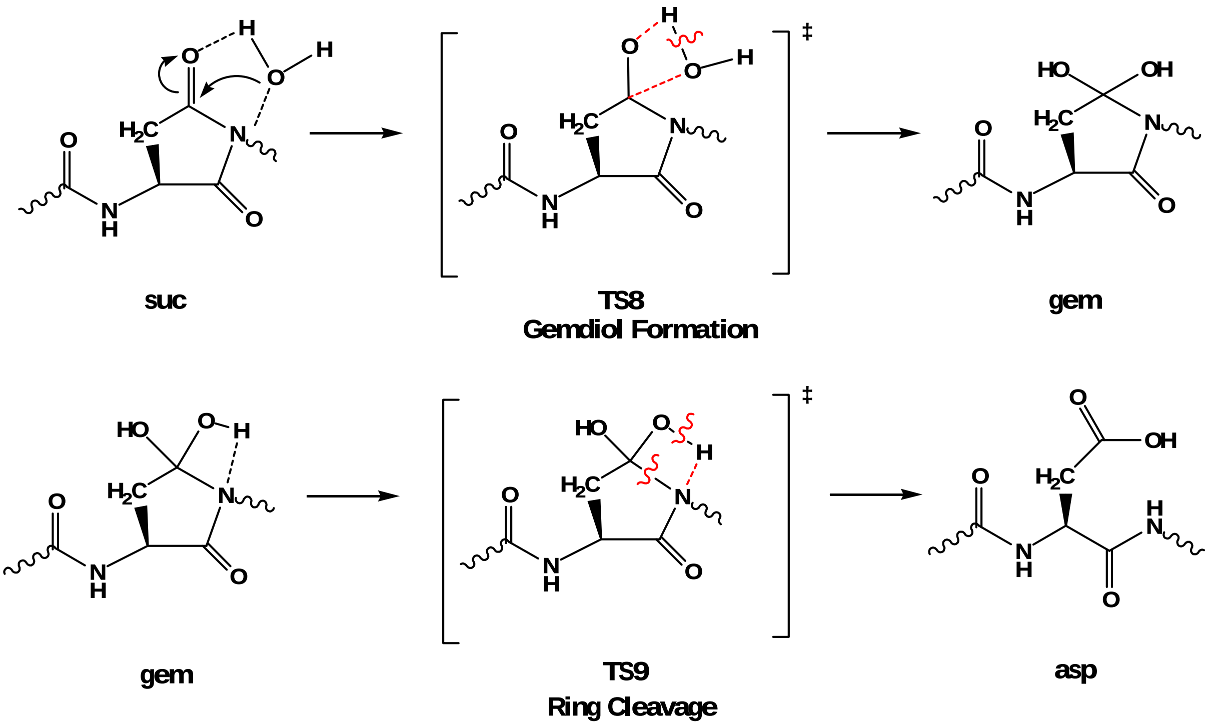


Figure 5.10. Stepwise hydrolysis of the succinimide intermediate (***suc*** → ***gem*** → ***asp***).

Optimized geometries for the transition structures of concerted amide hydrolysis are depicted in Figure 5.11. Given that this is a *hydrolysis* reaction, transition state **TS7-1H$_2$O** corresponds to an unassisted process and **TS7-2H$_2$O** corresponds to a one-water *assisted* concerted hydrolysis reaction. The two transition state structures differ in several aspects. The unassisted hydrolysis of the imide (**TS7-1H$_2$O**) is a four-centered concerted -yet asynchronous- transition state, where the proton transfer from the water molecule to the ring nitrogen has already occurred (N-H distance 1.126 Å), the lengthening in the C-N distance is substantial (1.684 Å), however the attacking –OH is still rather far (C-O distance 1.954 Å). The one-water assisted mechanism (**TS7-2H$_2$O**), however, shows different geometrical features; in the six-centered transition state, the proton transfer is still not complete, the nucleophilic –OH group is quite close (C-O distance 1.670 Å) to the carbonyl under attack, but the ring opening is still premature (C-N distance 1.693 Å).

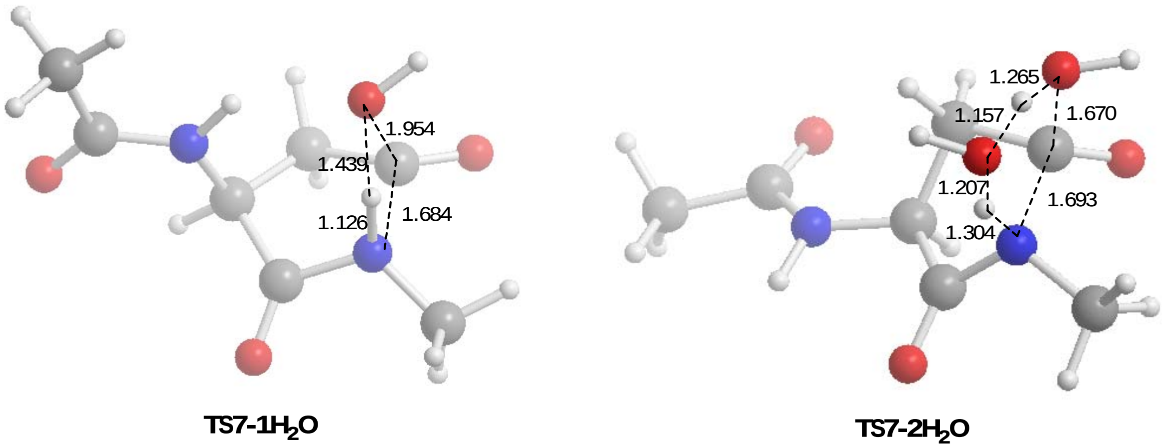


Figure 5.11. Optimized geometries for the transition state of concerted hydrolysis of the succinimide intermediate into Asp (***suc* → *asp***), "unassisted" (**TS7-1H$_2$O**) and one-water assisted (**TS7-2H$_2$O**) mechanisms, respectively.

Stepwise succinimide hydrolysis was modeled in the same manner, with one and two water molecules (Figure 5.12). In the unassisted stepwise mechanism, a water molecule is added to the ring in step 1 (**TS8-1H$_2$O**), the second step **(TS9)** is the ring-opening. In the two-water case, one of the solvent molecules is added to the ring in step 1 (**TS8-2H$_2$O**), while the second solvent molecule assists this process. The following ring opening step (**TS9-1H$_2$O**) is also assisted by a water molecule. The two water case (**TS8-2H$_2$O** and **TS9-1H$_2$O**) is undoubtedly a better model to study the gemdiol mechanism, due to the water-assistance, which is absent in the former case. This will be more apparent in the next section, when energetics of these reactions are comparatively discussed. The main difference in the geometries of the gemdiol formation transition states (**TS8-1H$_2$O** and **TS8-2H$_2$O**), is the extent of proton transfer. In the water-assisted case, proton transfer from the attacking water molecule to the carbonyl oxygen is almost complete, contrary to the unassisted case. However, the carbonyl C-O distance is longer in **TS8-1H$_2$O** (1.313 Å as opposed to 1.293 Å in **TS8-2H$_2$O**). A strong H-bond network can be seen in the water-assisted case. In the second step, water-assistance (**TS9-1H$_2$O**) has enhanced the extent of proton transfer compared to the unassisted case (**TS9**). Ring opening is slightly more achieved in **TS9-1H$_2$O** (2.238 Å as opposed to 2.220 Å in **TS9**).

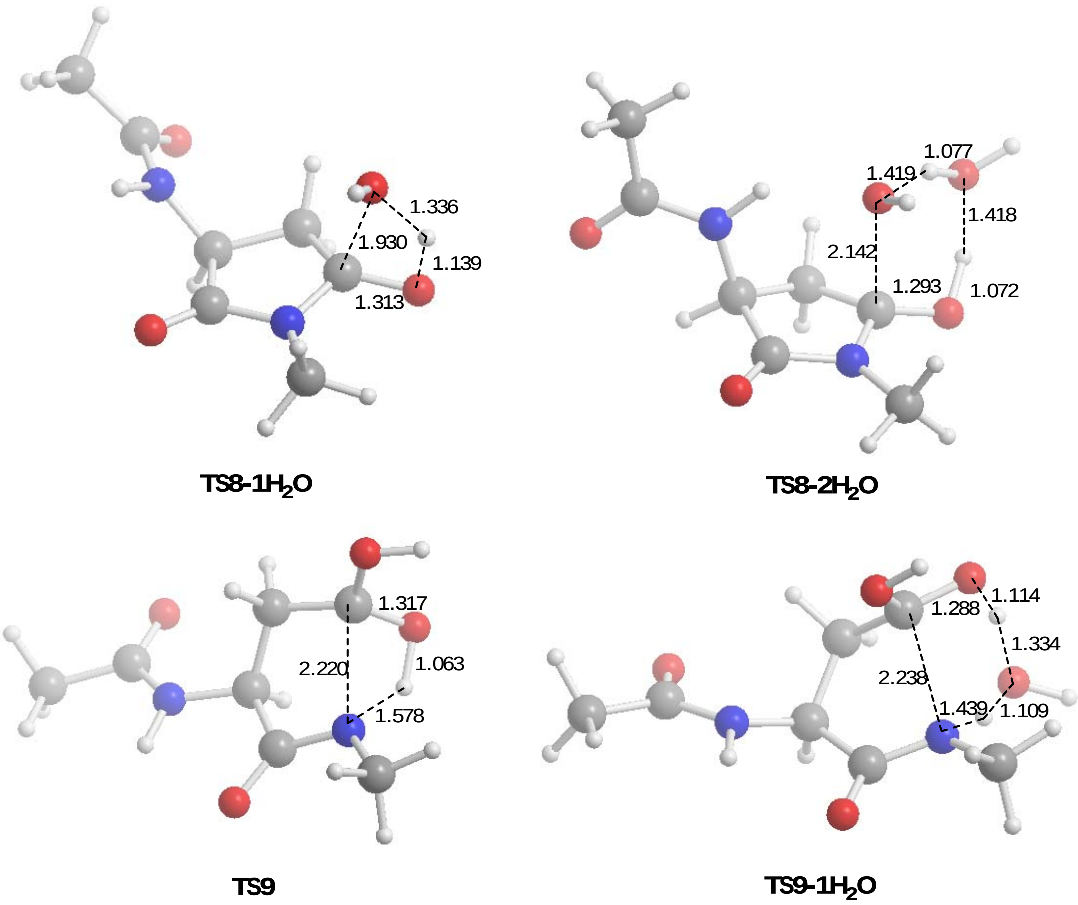


Figure 5.12. Optimized geometries for the transition state of gemdiol-mediated hydrolysis of the succinimide intermediate into Asp (***suc*** **→** ***gem*** **→** ***asp***), “unassisted” (**TS8-1H$_2$O** and **TS9**) and one-water assisted (**TS8-2H$_2$O** and **TS9-1H$_2$O**) mechanisms, respectively.

5.3.3. Direct Hydrolysis Versus Succinimide-Mediated Deamidation: Comparison Of Energetics And Mechanisms

In this part, imide-mediated routes will be energetically compared amongst each other as well as against the direct hydrolysis pathway. Relative barriers and feasibilities will be discussed. Competing mechanisms with identical molecularity (same number of atoms, depending on the number of water molecules in the initial complex, 0, 1 or 2) are grouped and presented in Figures 5.13 (no water), 5.14 (one-water) and 5.15 (two-water) along with energetics. All structures whose energy values are reproduced from previous deamidation studies are labeled with an asterisk (*).

Although previously explored by Konuklar et al. [56, 58], the energetics for 1) the formation of the succinimide intermediate and 2) its decomposition into aspartic acid by means of water hydrolysis has not been analyzed with the same number of water molecules and within the same energetic scale. These reactions have been previously explored separately and barrier heights have been calculated with respect to the initial structure of each individual reaction. Therefore, activation energies are not relative to one another, but to the reactant of each step. In this study, we have evaluated all steps of the deamidation process with respect to a single reference point, the model peptide (Figure 5.2). Activation barriers calculated in earlier studies for waterless, one- and two-water cases were re-calculated with respect to this reference point for a legitimate comparison; slight changes in free energies of activation have been observed for some steps, but differences are within the range of 2 kcal/mol. The free energy of a single ammonia molecule was added to each component in the succinimide hydrolysis mechanism for scaling purposes.

The waterless mechanism (Figure 5.13) shows all three possibilities for succinimide formation, ***asn* → *suc***, ***asn* → *tet* → *suc*** and ***asn* → *taut* → *tet* → *suc*** (Figures 4.4, 5.7 and 5.8, respectively). The concerted mechanism (***asn* → *suc***) has the highest barrier (58.7 kcal/mol). The tautomerization route (and ***asn* → *taut* → *tet* → *suc***) is the most plausible pathway for succinimide formation in the absence of solvent assistance. The deamination (***tet* → *suc***) step is rate-determining (50.4 kcal/mol) in both stepwise mechanisms. Please note that the waterless mechanism depicted in Figure 6, does not include the hydrolysis of the imide (***suc***) into an Asp, since this requires a water molecule. Waterless mechanisms have been modeled as a benchmark, in order to see the effect of water catalysis on each step but the fact that a succinimide intermediate can form even in the absence of water molecules is remarkable.

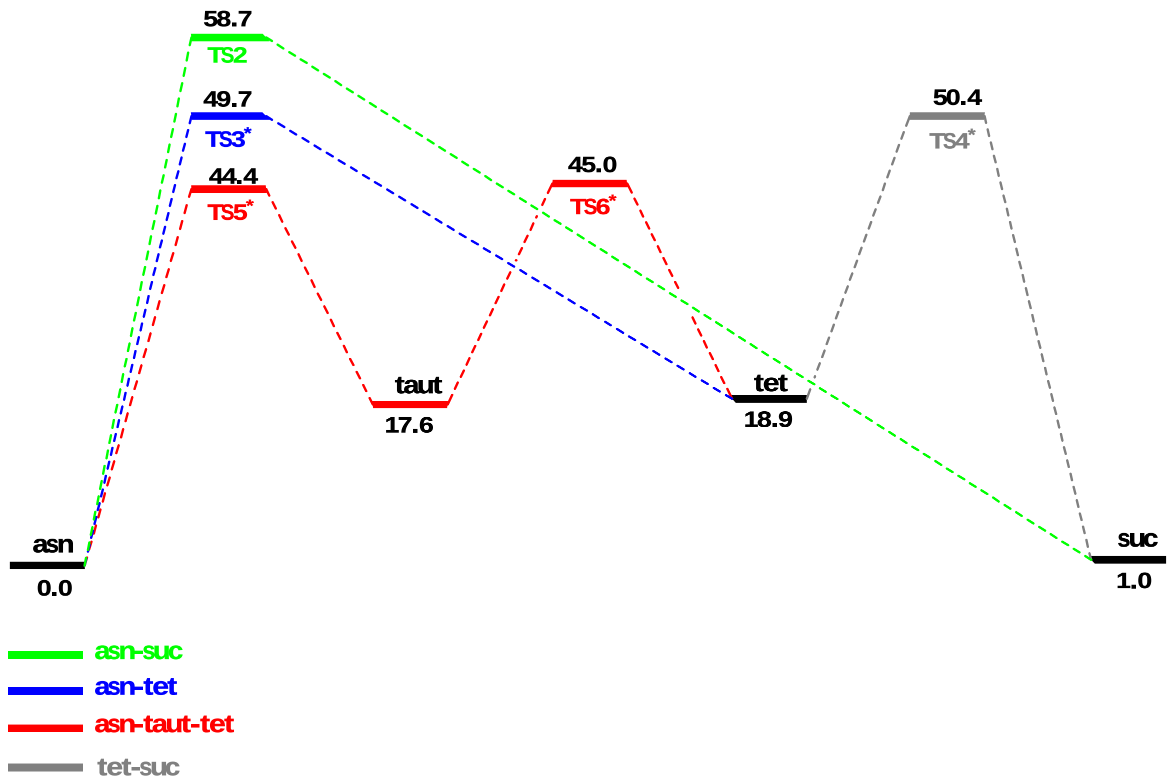


Figure 5.13. Reaction coordinate for deamidation -no water.

Deamidation mechanisms involving one water molecule are depicted in Figure 5.14. Three possibilities for succinimide formation are shown; once again the concerted mechanism (***asn*** → ***suc***) has the highest barrier (57.6 kcal/mol) and the tautomerization route (***asn*** → ***taut*** → ***tet*** → ***suc***) is the most plausible pathway for succinimide formation with one water assistance. The deamination (***tet*** → ***suc***) step is no longer rate-determining (37.5 kcal/mol) for succinimide formation in the stepwise pathways (***asn*** → ***tet*** → ***suc*** and ***asn*** → ***taut*** → ***tet*** → ***suc***); the cyclization step is the bottleneck for the formation of the imide.

However, the hydrolysis of the succinimide also shows relatively high barriers (Figure 5.14). The concerted hydrolysis (***suc*** → ***asp***) was expected to have a higher barrier (58.2 kcal/mol) than the gemdiol-mediated stepwise route (***suc*** → ***gem*** → ***asp***); the barrier difference between the two mechanisms is approximately 5 kcal/mol. The ring-opening step (***gem*** → ***asp***) is rate-determining for the stepwise pathway (53.9 kcal/mol as opposed to 48.5 kcal/mol in ***suc*** → ***gem***).

When imide-mediated deamidation is considered, the most plausible pathways for succinimide formation and succinimide hydrolysis are the tautomerization and gemdiol mechanisms, respectively. The rate-determining step for the complete imide-mediated deamidation process seems to be the ring opening step (***gem*** → ***asp***) in succinimide hydrolysis with a barrier much higher (53.9 kcal/mol) than the cyclization step. However, it should be noted that the succinimide hydrolysis barriers involve mechanisms with only one water molecule, i.e. the hydrolysis step depicted in Figure 5.14 is unassisted, unlike the succinimide formation steps. The water molecule is a reactant in hydrolysis, as mentioned earlier (Figure 5.11 and 5.12); therefore the one-water case does not constitute for a fair comparison of energetics between the formation and hydrolysis of the succinimide.

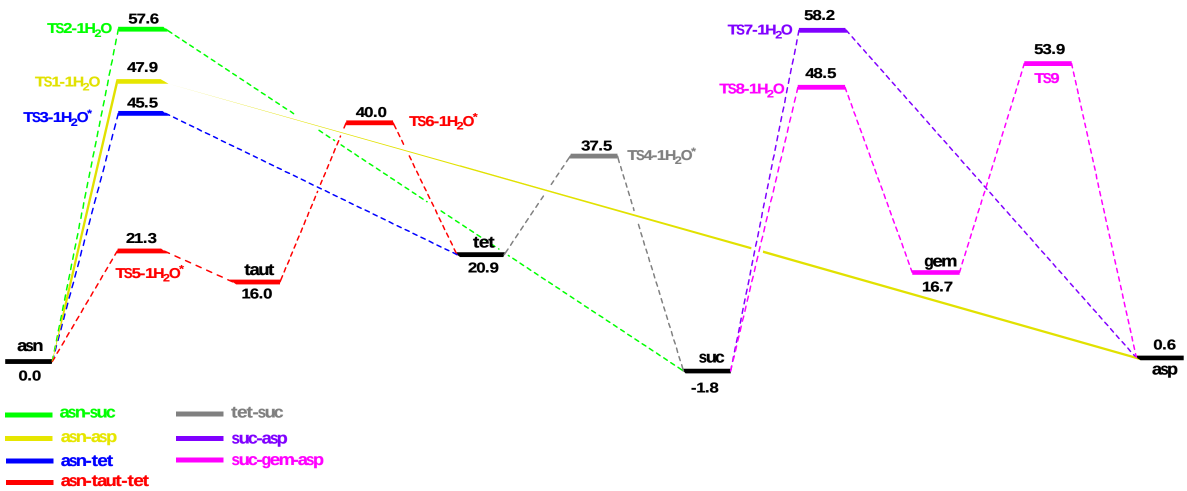


Figure 5.14. Reaction coordinate for deamidation -one water.

The direct hydrolysis mechanism (***asn* → *asp***) with one-water molecule is relatively slower (47.9 kcal/mol) than the succinimide formation steps, but like the imide hydrolysis steps, direct hydrolysis with one-water is unassisted, and therefore a legitimate comparison cannot be made.

In the two-water case (Figure 5.15), all steps of the imide-mediated deamidation reaction and direct hydrolysis are water-assisted; activation barriers of all steps are lower than their uncatalyzed counterparts (Figure 5.13 and 5.14). The concerted ***asn* → *suc*** mechanism for the formation of the succinimide intermediate was not further investigated with two explicit water molecules since the barrier height is considerably higher than the stepwise routes (***asn* → *tet* → *suc*** and ***asn* → *taut* → *tet* → *suc***) as shown in Figures 5.13 and 5.14. The tautomerization route ***(asn* → *taut* → *tet* → *suc***) is the most plausible pathway for succinimide formation with two water assistance. The cyclization step is rate-determining for succinimide formation in both stepwise pathways (***asn* → *tet* → *suc*** and ***asn* → *taut* → *tet* → *suc***).

The activation barriers for the stepwise hydrolysis of the succinimide with two-water molecules have much lower barriers than the previous one-water -unassisted- case (Figure 5.14). Again the concerted hydrolysis (***suc* → *asp***) has a higher barrier (58.4 kcal/mol) than the gemdiol-mediated stepwise route (***suc* → *gem* → *asp***); the barrier difference between the two mechanisms is approximately 10 kcal/mol. The ring-opening step (***gem* → *asp***) is rate-determining for the stepwise pathway (46.3 kcal/mol as opposed to 36.9 kcal/mol in ***suc* → *gem***).

In imide-mediated deamidation, the most plausible pathways for succinimide formation and succinimide hydrolysis are the tautomerization (***asn* → *taut* → *tet* → *suc***) and gemdiol (***suc* → *gem* → *asp***) mechanisms, respectively. The rate-determining step for the complete imide-mediated deamidation process seems to be the ring-opening step (***gem* → *asp***) in succinimide hydrolysis with a barrier (46.3 kcal/mol as opposed to 39.7 kcal/mol) almost 6 kcal/mol higher than the cyclization step (***taut* → *tet***) in succinimide formation.

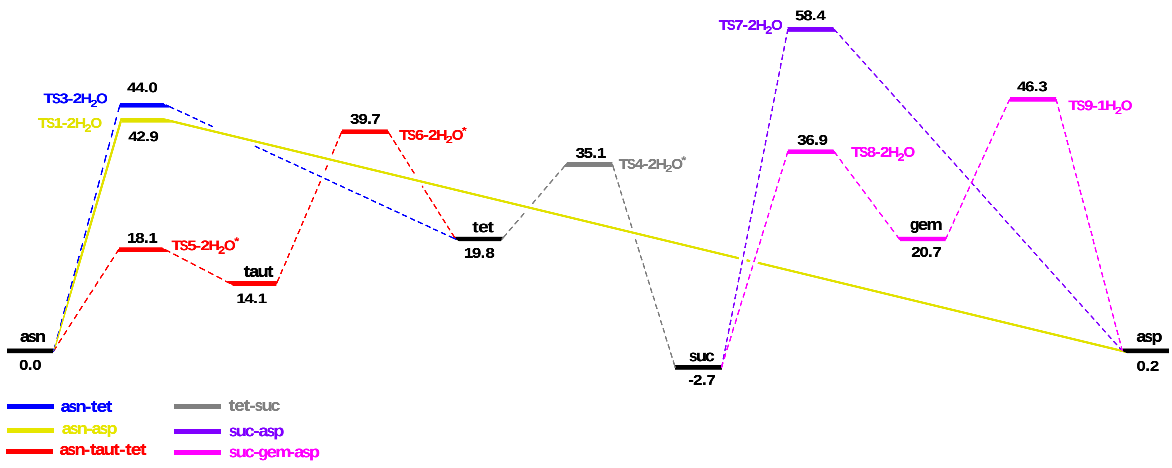


Figure-5.15. Reaction coordinate for deamidation -two water.

For all mechanisms (waterless and water-assisted) the tautomerization route (***asn* → *taut* → *tet* → *suc***) has the lowest barrier for the formation of the succinimide intermediate. The cyclization step is the rate-determining step for succinimide formation in all water assisted mechanisms. The deamination step is faster than cyclization, except for the waterless deamination case (Figure 5.13).

Succinimide hydrolysis barriers are higher than cyclization barriers, indicating that the bottleneck of the deamidation process is in fact the hydrolysis step. This may in fact explain the observation and isolation of succinimide intermediates during deamidation experiments. Previous computational studies on succinimide formation and succinimide hydrolysis have indicated that the rate determining step for the overall deamidation process is the cyclization step leading to the tetrahedral intermediate, however, these studies did not take into account the effect of water molecules and activation barriers were calculated for the waterless reactions. Nevertheless, the identity of neighboring residues or backbone orientation may change the position of the bottleneck in the imide-mediated deamidation process.

In addition, direct hydrolysis seems to be a competitive reaction to the imide-mediated route even in the absence of acid or base catalysis. The activation barrier for direct hydrolysis (***asn* → *asp***) with two-water molecules (42.9 kcal/mol) is slightly lower than the rate-determining step (***gem* → *asp***) in the imide mechanism (46.3 kcal/mol). Although, it is well-known that direct hydrolysis is the dominant mechanism for deamidation of Asn under acidic conditions, it has been suggested that at neutral pH the succinimide mechanism is in effect. The fact that these barriers are only 3 kcal/mol apart suggests that they are rather competitive and neither mechanism can be ruled out for the deamidation of Asn. Availability of solvent molecules, position of water bridges, intramolecular interactions within the peptide or protein, as well as spatial distribution of neighboring residues and three-dimensional aspects, such as hindrance by secondary structure may effect which mechanism will be at play. The absence of iso-Asp as a product of deamidation in some cases [8] may be a result of secondary structure inhibiting cyclization; in that case direct hydrolysis is destined to be the major pathway for deamidation.

5.4. Conclusion

Main conclusions that can be drawn from this study are as follows: 1) water assistance increases the rate of deamidation, a fact already established 2) the tautomerization route has the lowest barrier for the formation of the succinimide intermediate regardless of the number of water molecules that assist the reaction, including the waterless mechanism; 3) cyclization is the rate-determining step for succinimide formation in all water assisted mechanisms; 4) hydrolysis of the succinimide intermediate is likely to go through a stepwise mechanism, where a gemdiol intermediate is formed; 5) more importantly, succinimide hydrolysis barriers are higher than those for succinimide formation.

These conclusions contrast in part with previous calculations that had shown that cyclization is the rate determining step for the formation of the succinimide intermediate. Our study suggests that when the entire deamidation process is considered, the hydrolysis step is the actual rate determining step. As shown in Figure 5.15, the stepwise hydrolysis barrier is the rate determining step for the overall water-assisted deamidation process, which is likely to proceed through the tautomerization route. This also explains the isolation of succinimide intermediates during deamidation reactions. The bottleneck of this process is therefore, proposed to be the hydrolysis of the succinimide intermediate.

Another important finding is the relative ease of direct hydrolysis. In all mechanisms involving the use of explicit solvent molecules (Figures 5.14 and 5.15) direct hydrolysis seems to be a competitive reaction to the imide-mediated route even in the absence of acid or base catalysis.

6. NON-ENZYMATIC PEPTIDE BOND CLEAVAGE AT ASPARAGINE AND ASPARTIC ACID

The previous chapter showed that, as long as water-assistance is possible, the cyclization step is not likely to be the rate-determining step for the overall deamidation process. The bottleneck of the imide-mediated route was instead suggested to be the hydrolysis of the succinimide intermediate; this also explains the isolation and observation of succinimide derivatives in some deamidation reactions. In addition, the direct hydrolysis mechanism was shown to be as feasible as the imide-mediated mechanism, indicating why iso-aspartate is not always observed as a result of deamidation under physiological conditions. However, deamidation is not the only fate of asparaginyl residues; many deamidation experiments have yielded fragmentation products as well. This chapter deals with the cleavage of the peptide backbone near Asn and Asp. Peptide fragmentation has been experimentally observed near these residues and the aim of this study is to propose a plausible pathway for backbone cleavage. The details of the mechanism and the energetics will also provide insight on the feasibility of cleavage near Asn and Asp. It is in fact important to predict the likelihood of deamidation preceding backbone cleavage at Asn residues. The following article has recently been submitted for publication in the *Journal of Organic Chemistry*.

Non-enzymatic peptide bond cleavage at asparagine (Asn) and glutamine (Gln) residues has been observed during peptide deamidation experiments; cleavage has also been reported at aspartic acid (Asp) and glutamic acid (Glu) residues. Although, peptide backbone cleavage at Asn is known to be slower than deamidation, fragmentation products are often observed during peptide deamidation experiments. In this study, mechanisms leading to the cleavage of the carboxyl-side peptide bond of Asn and Asp residues were investigated using computational methods (B3LYP/6-31+G**). Single point solvent calculations at the B3LYP/6-31++G** level, were carried out in water, utilizing the integral equation formalism-polarizable continuum (IEF-PCM) model. Mechanism and energetics of peptide fragmentation at Asn were comparatively analyzed with previous calculations on deamidation of Asn. When deamidation proceeds through direct hydrolysis of the Asn side chain or through cyclic imide formation –via a tautomerization route– it exhibits lower activation barriers than peptide bond cleavage at Asn. The fundamental distinction between the mechanisms leading to deamidation –via a succinimide– and backbone cleavage was found to be the difference in nucleophilic entities involved in the cyclization process (backbone versus side chain amide nitrogen). If deamidation is prevented by protein three-dimensional structure, cleavage may become a competing pathway.

Fragmentation of the peptide backbone at Asp was also computationally studied, in order to understand the likelihood of Asn deamidation preceding backbone cleavage. The activation barrier for backbone cleavage at Asp residues is much lower (approximately 15 kcal/mol) than that for Asn. This suggests that peptide bond cleavage at Asn residues is more likely to take place after it has deamidated into Asp.

6. 1. Introduction

Asparagine (Asn) and glutamine (Gln), two of the 20 most common natural amino acids, are known to be uniquely unstable under physiological conditions; they spontaneously and non-enzymatically deamidate into aspartic acid (Asp) and glutamic acid (Glu), respectively [7, 10, 11]. However, deamidation is not the only possible fate for these two amino acids; non-enzymatic peptide bond cleavage at the carboxyl-side of Asn and Gln residues has also been experimentally observed [9]. The multitude of products observed is most likely due to the tendency of Asn and Gln to form rings.

Deamidation of Asn at acidic pH is known to occur exclusively through a direct hydrolysis mechanism, where the neutral side chain amide is transformed into a carboxylic acid, forming an Asp residue [8]. However, Asn deamidation at neutral pH has been suggested to take place via a cyclic imide, further experimental [19, 54, 55, 59, 67, 68] and computational [56-58, 69, 70, 73] studies have supported this idea. In the succinimide-mediated deamidation mechanism, a nucleophilic attack of the carboxyl-side backbone NH to the Asn side chain carbonyl occurs, forming a cyclic tetrahedral intermediate (Figure

6.1). The cyclization step is then followed by deamination, where an ammonia molecule is ejected, to form an even more stable intermediate, the succinimide. Subsequent hydrolysis of the amide bonds on the succinimide ring forms the Asp and iso-Asp products.

The rate determining step of this mechanism has been suggested to be the initial cyclization step [54, 56] where the backbone amide acts as a nucleophile. However, a recent theoretical study on the deamidation mechanism of a small model peptide has shown that the barrier for the cyclization step is much lower with water-assistance and succinimide hydrolysis is the actual rate determining step of the overall deamidation reaction [71], noting that results may vary for different n +1 residues. On the other hand, it was also shown that for the same system deamidation through direct hydrolysis of the Asn side chain at neutral pH is at least as feasible as deamidation through a cyclic imide.

Capasso et al. have studied peptide bond cleavage at Asn [74, 75]. At elevated temperatures, only backbone cleavage without deamidation has been observed with Asn-Pro sequences, where succinimide formation cannot occur [9, 75, 76] while Asn-Ile gave both deamidation and peptide fragmentation products. Cleavage products are formed in most peptide deamidation experiments, but the reaction is usually much slower than deamidation. Asn peptide deamidation half-times range from about 1 to 400 days, Asn cleavage rates range from about 200 to more than 10,000 days [68]. Cleavage of Asn-Pro is the fastest sequence and, since its backbone nitrogen lacks a proton, Asn-Pro deamidates by slow hydrolysis; hence cleavage is the principal degradative pathway for Asn-Pro sequences [9, 75].

Peptide bond cleavage at Asp and Glu residues has also been reported in peptides and proteins [77, 78] and the occurrence rate is higher than that for Asn and Gln. The side-chain carboxylic acid group is suggested to play a key role in the cleavage process; this is referred to as the “aspartic acid side chain effect” [79]. Since Asn deamidation occurs more readily than backbone cleavage, it is important to investigate the probability of deamidation preceding cleavage, in which case, mechanism and energetics of Asp cleavage becomes essential.

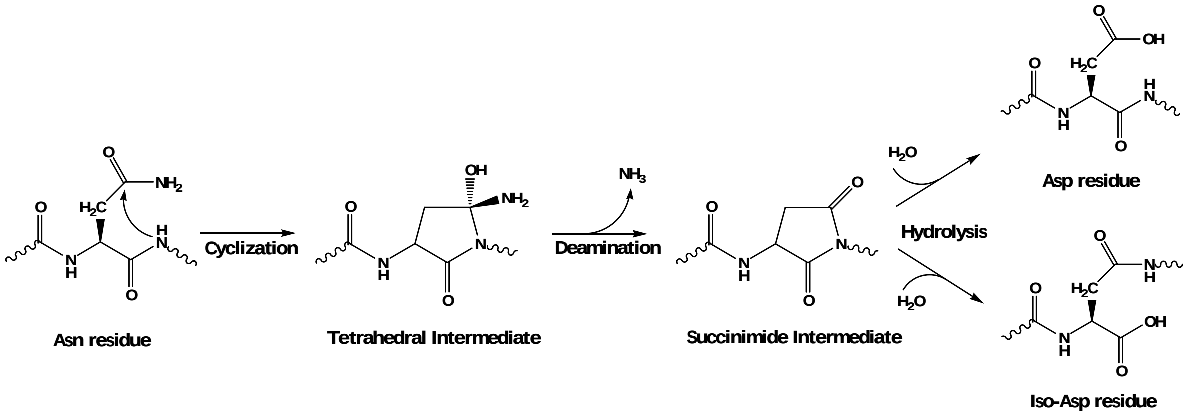
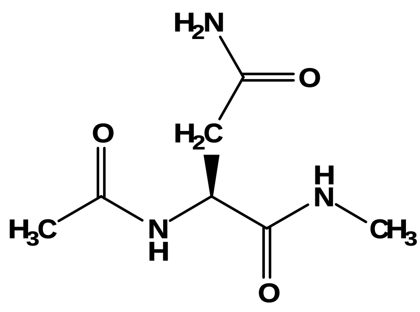


Figure 6.1. Succinimide-mediated deamidation pathway.

The aim of this study is to computationally explore the mechanistic differences between deamidation and peptide bond cleavage at Asn and investigate backbone cleavage at Asp. In order to account for the rate difference between cleavage and deamidation, energetics of peptide fragmentation will be compared with previous calculations [71] on deamidation of Asn. Moreover, mechanistic and energetic information on Asp backbone cleavage will enable a reasonable prediction for the likelihood of Asn deamidation preceding cleavage. Instead of Asn itself undergoing cleavage, it may convert to Asp via deamidation and cleavage may take place henceforth. Knowledge on the mechanism and energetics of Asn cleavage may also be useful in cases where deamidation is hindered by protein tertiary structure, in which case cleavage may become a competing reaction.

6.2. Computational Methodology

As mentioned earlier, Asn peptide bond cleavage mechanism studied herein will be compared with the Asn deamidation mechanism previously studied by the authors [71] and therefore the same methodology was employed in this study. Moreover, the same model peptide (Figure 6.2) was used for a fair analysis. All structures and energy values reproduced from previous deamidation studies are indicated by an asterisk (*). Previous computational studies have shown that water molecules have a catalytic effect on deamidation, whereby they serve as efficient proton conduits. The backbone cleavage mechanism was investigated in light of this information.



Model Compound

Figure 6.2. Model peptide with Asn residue.

Full geometry optimizations were performed in gas-phase –without any constraints– using the density functional theory (DFT) [34] at the B3LYP/6-31+G** level [36-40]. The use of this basis set and method in similar peptide systems is well established. Stationary points were characterized by a frequency analysis. Zero-point energy and thermal corrections were attained using the ideal gas approximation and standard procedures. Local minima and first-order saddle points were identified by the number of imaginary vibrational frequencies. The species reached by each transition structure was determined by intrinsic reaction coordinate (IRC) calculations [60, 61]. Relative free energies of activation ($\Delta G^{\ddagger}$) are calculated as the difference of free energies between transition states and reactants (reactant-water complex where applicable) of each step. Energy values for gas-phase optimizations listed throughout the discussion include thermal free energy corrections at 298 K and 1 atm.

The self-consistent reaction field (SCRF) theory, utilizing the integral equation formalism-polarizable continuum (IEF-PCM) model [41-45] in H_2O (ε = 78.0) at the B3LYP/6-31++G** level was used to account for the effect of a polar environment. Bondi radii [62] scaled by a factor of 1.2 were used for all solvent calculations. Single point energies for solvent calculations include nonelectrostatic and thermal free energy corrections obtained from gas-phase optimizations.

All calculations were carried out using the Gaussian 03 program package [63]. All distances and free energies listed in the discussion are in angstroms (Å) and kcal/mol, respectively. Single point solvent energies are given in parenthesis.

6.3. Results and Discussion

The first part of the discussion elaborates on the mechanism of Asn peptide bond cleavage and comparatively analyzes mechanism and energetics of peptide fragmentation with deamidation. Alternative fates for intermediates formed during backbone cleavage are discussed in detail. In the second part, backbone cleavage at Asp is investigated and predictions are made based on energetic comparisons.

6.3.1. Peptide Fragmentation at Asn Residues

6.3.1.1. Cyclization. Cleavage of the peptide bond by attack of the Asn side chain NH_2 is assumed to proceed stepwise, i.e. through preliminary attack to Asn backbone carbonyl followed by bond cleavage. Therefore, in the first step, a cyclic intermediate forms (Figure 6.3). Although the side chain NH_2 group is not likely to be as nucleophilic as the backbone NH, it is considerably more mobile and has a better chance of assuming a reactive conformation. The product (Intermediate I) of this ring closure is a cyclic tetrahedral intermediate, much like the one that forms in the succinimide-mediated deamidation pathway (Figure 6.1).

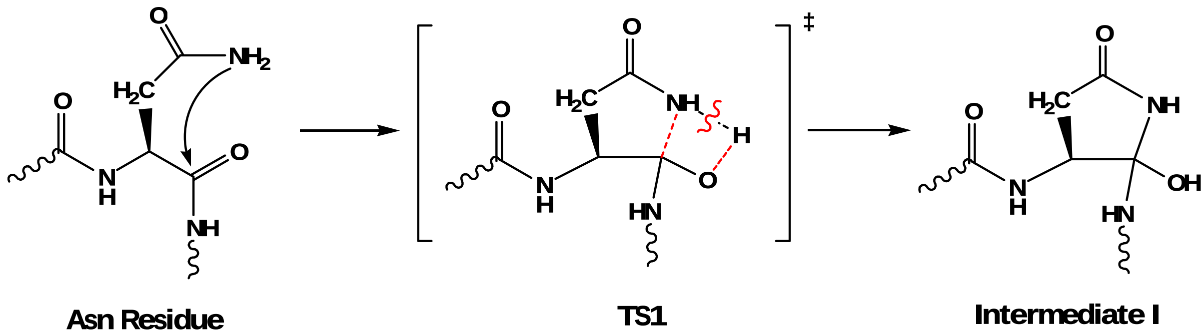


Figure 6.3. Side-chain NH_2 attack on Asn backbone carbonyl to form a cyclic tetrahedral intermediate.

Both, non-assisted and water-assisted mechanisms were modeled for the concerted ring closure via NH_2 attack (Figure 6.4). The non-assisted attack (**TS1**) involves the early transfer of a proton from the nucleophile (side chain NH_2) to the Asn backbone carbonyl's oxygen, as indicated by the rather short O-H distance (1.003 Å) and the elongation in the backbone carbonyl (1.340 Å) compared to the ground state C=O distance of 1.230 Å. Meanwhile, ring formation is underway (N-C distance 2.238 Å).

The water assisted mechanism (**TS1-H₂O**), where the proton transfer from the side chain to the backbone goes through a solvent molecule, shows the H^+ is transferred to the carbonyl group (O-H distance 1.015 Å) while the ring is yet to close (N-C distance 2.254 Å). The main difference between the concerted waterless (**TS1**) and water-assisted (**TS1-H_2O**) transition state is the distance of the breaking N-H bond (1.798 Å versus 1.689 Å, respectively), which is probably due to the larger flexibility of the latter system.

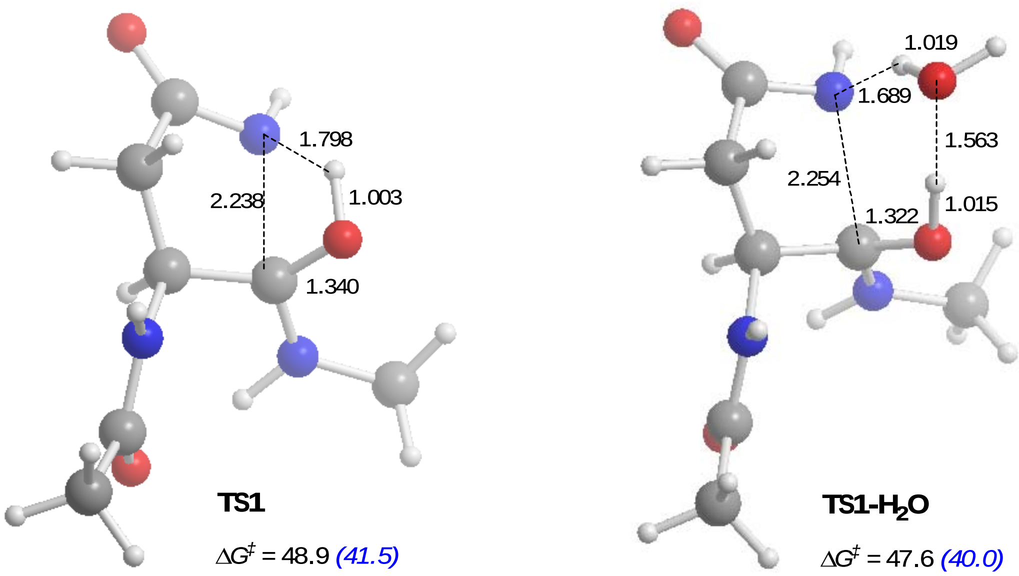


Figure 6.4. Optimized geometries and free energies of activation for the transition state of side chain NH_2 attack on backbone Asn carbonyl, concerted non-assisted (**TS1**) and concerted water-assisted (**TS1-H$_2$O**) mechanisms, respectively.

Previous studies on the deamidation mechanism have revealed the catalytic effect of water molecules. In this particular reaction step, the assistance of the water molecule has a favorable effect, however it is not substantial, as may be seen in the barrier heights reported (Figure 1). Single point solvent energies in water indicate the stabilizing effect of a polar environment. Both transition states have a somewhat zwitterionic structure, and therefore benefit from electronic interaction with the dielectric continuum. This is indicated by approximately 8 kcal/mol decrease in barrier height for both ring closure reactions involving **TS1** and **TS1-H$_2$O** (Figure 6.4).

Although, the attacking species and the products are different, the free energy of activation for the concerted non-assisted ring closure in deamidation and cleavage pathways is rather close, with less than 1 kcal/mol free energy difference (Figure 6.5).

Solvent calculations have a similar effect on both mechanisms. The tetrahedral intermediate, formed via the deamidation path, is more stable (by approximately 3 kcal/mol) than the one (Intermediate I) formed during cleavage.

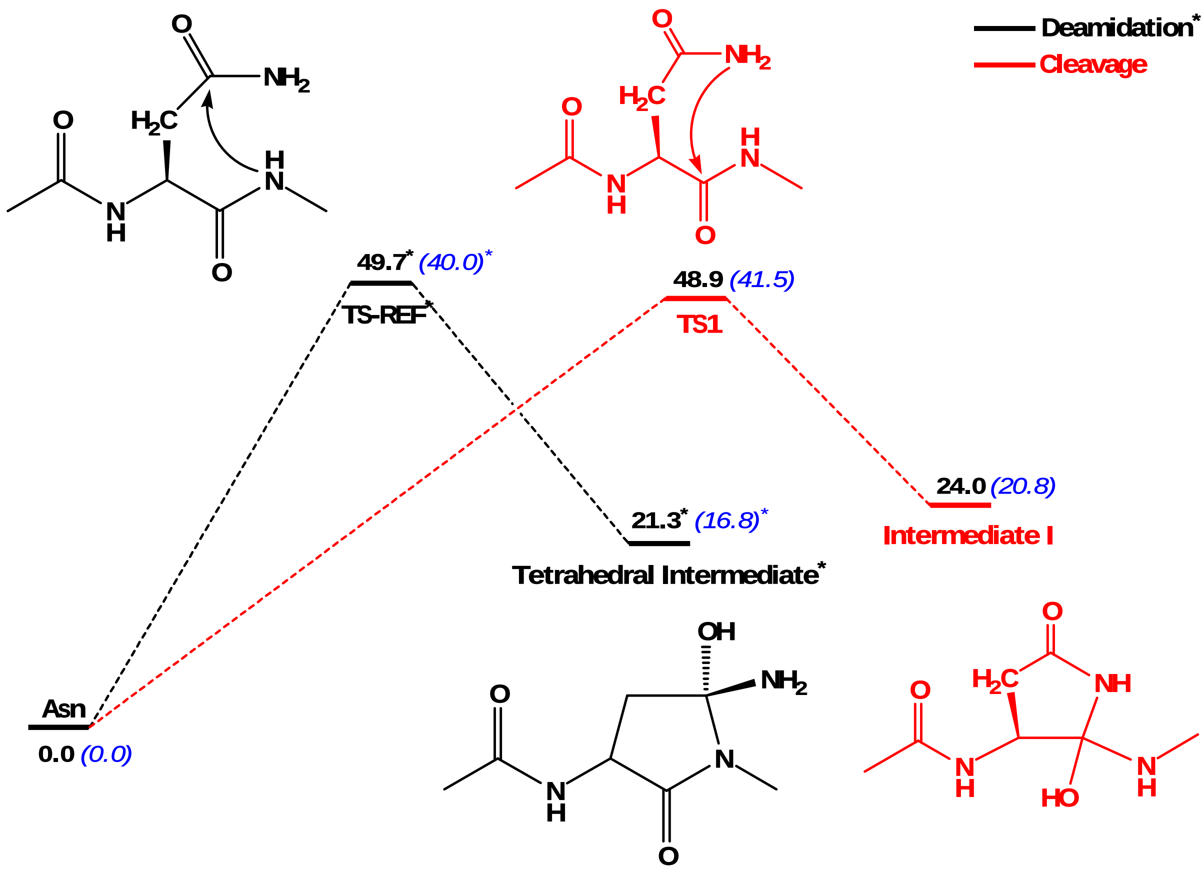


Figure 6.5. Relative free energies for the concerted non-assisted cyclization step in deamidation and backbone cleavage of Asn.

The energetics of the water-assisted mechanism for the ring closure steps of cleavage and deamidation are depicted in Figure 6.6. Water assistance has a more favorable effect on the deamidation cyclization step, lowering the barrier by almost 5 kcal/mol, whereas it only causes a decrease of 1.3 kcal/mol for the cyclization step in the cleavage mechanism. Nevertheless, activation energies for cyclization in both pathways are still comparable.

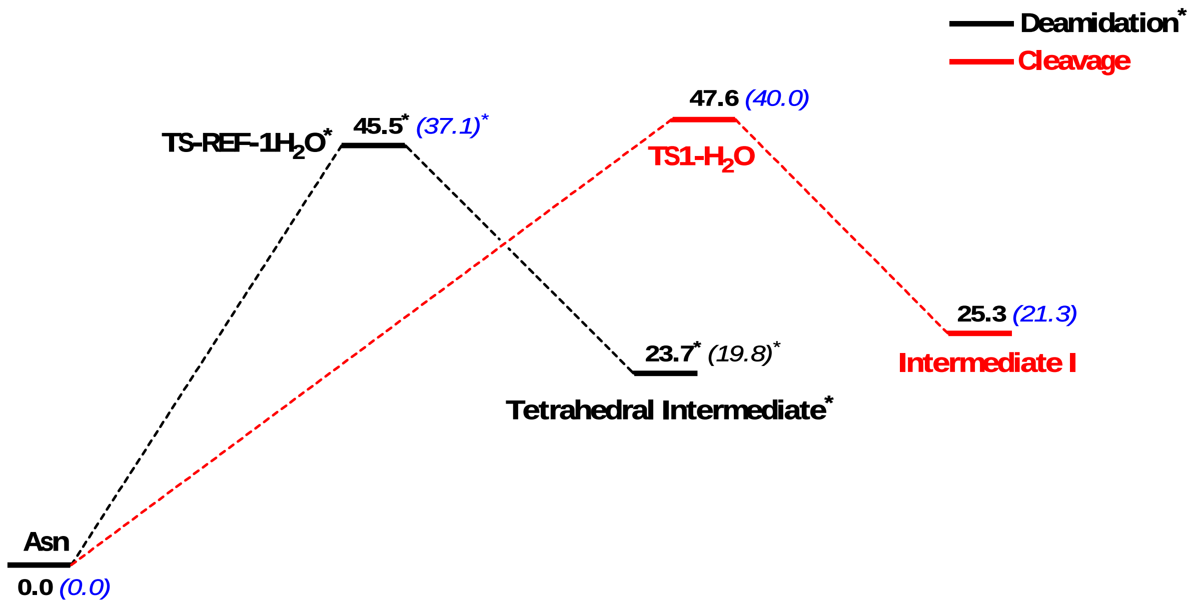


Figure 6.6. Relative free energies for the concerted water-assisted cyclization in deamidation and backbone cleavage of Asn.

<u>6.3.1.2. Fates of Intermediate I.</u> While the deamidation pathway proceeds with a water assisted deamination (expulsion of NH_3), Intermediate I does not have a free NH_2 group for this to occur. However, it may undergo a water-assisted cleavage of the peptide backbone and form fragmentation products. Besides the ring amide may be hydrolyzed to give a non-cyclic intermediate. Both reactions are studied below.

Water molecules assist the fragmentation of the peptide bond on the carboxyl-side (Figure 6.7) by allowing a proton relay from the –OH on the ring to the carboxyl-side backbone NH group. In this way, the alcohol functionality will be converted into a carbonyl group enhancing conjugation in the cyclic structure, while the backbone of the peptide breaks. This will lead to two fragments, one of which bears a free NH_2 group and the other will be carrying a succinimide ring at its tail end.

Transition state **TS2-H_2O** shows a slight lengthening in the peptide backbone; the C–N bond is in the range of 1.550 Å (Figure 6.8), quite close to the ground state bond distance of 1.470 Å, suggesting an early (reactant-like) transition state. The proton transfer is incomplete, as opposed to the case in **TS1** and **TS1-H_2O** (Figure 6.4). The activation barrier is significantly low (approximately 30 kcal/mol lower) compared to the initial cyclization step (Figures 6.5 and 6.6).

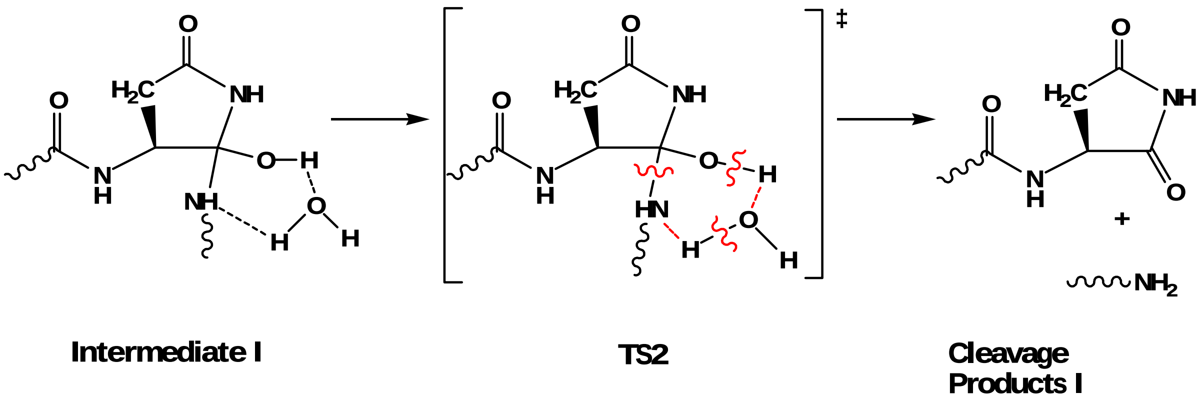


Figure 6.7. Peptide backbone cleavage at Intermediate I.

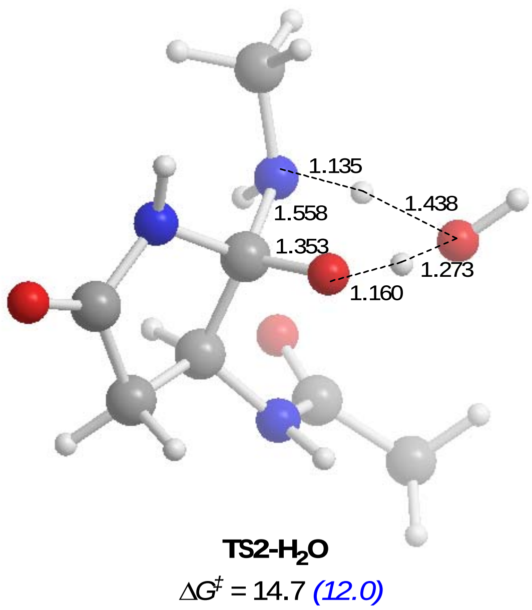


Figure 6.8. Optimized geometry and free energy of activation for the transition state of cleavage at Intermediate I, water-assisted (**TS2-H_2O**).

The transition state geometries for the backbone cleavage step reveal some prerequisites for peptide fragmentation (Figure 6.8). Considering backbone rigidity in a peptide or protein, the proton transfer from the ring –OH to the backbone NH must take place via a solvent molecule. In addition, proper alignment of these two entities is crucial; the –OH should not be involved in any other intermolecular interactions, the NH group should be in the proper conformation to accept the proton and above all, a solvent molecule must form a water bridge between these two entities in order to facilitate proton transfer. These requirements are not too challenging for a peptide chain with considerable flexibility and access to solvent molecules. However, inside the three-dimensional structure of a protein the possibility of proper alignment and water bridge availability may be limited.

Another possible fate for Intermediate I is the hydrolysis of the ring amide by nearby water molecules. Hydrolysis can take place via a concerted ring opening (Figure 6.9) or through a gemdiol intermediate (Figure 6.10). Both will give the same product, in which the Asn side chain has been transformed into an Asp.

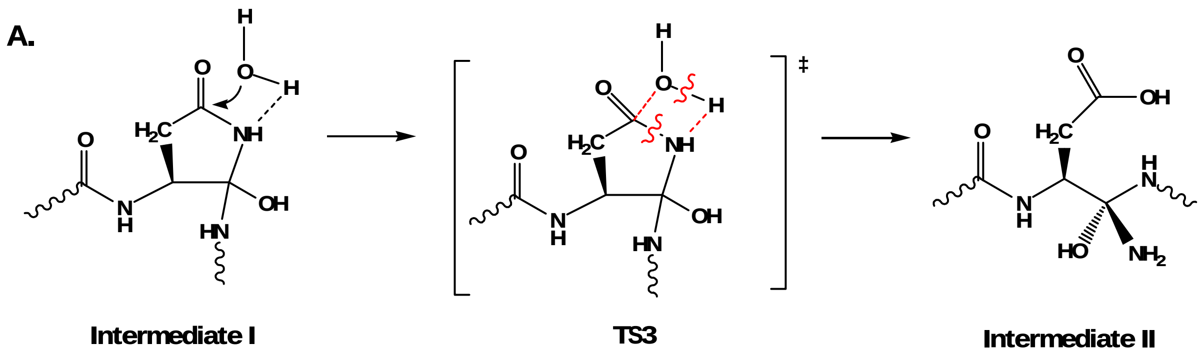


Figure 6.9. Concerted ring amide hydrolysis at Intermediate I.

In the concerted amide hydrolysis reaction, a water molecule attacks the ring carbonyl, the N-C bond breaks, as a proton is transferred from the water molecule to the ring NH. As a result, a carboxylic acid and an NH_2 group emerge (Figure 6.9). In the gemdiol-mediated stepwise mechanism (Figure 6.10), the initial step is the addition of a water molecule to the ring carbonyl, forming a gemdiol intermediate, which consequently undergoes ring opening to reveal the same product (Intermediate II).

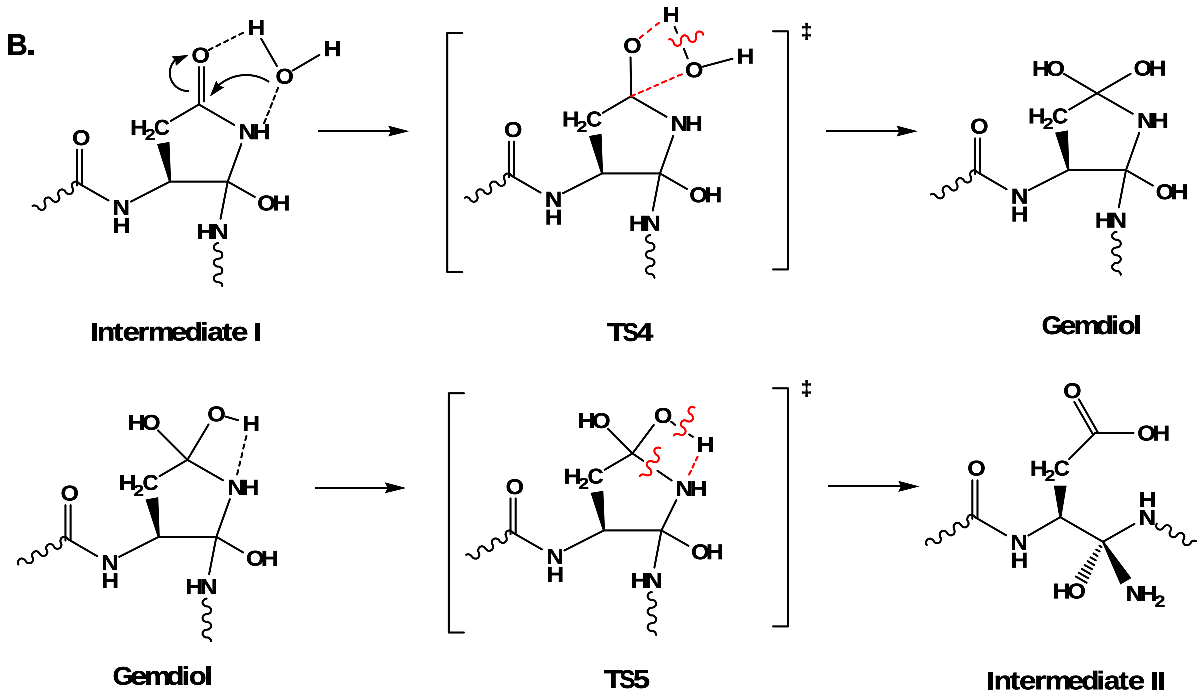


Figure 6.10. Stepwise ring amide hydrolysis at Intermediate I via a gemdiol intermediate.

The water-assisted concerted hydrolysis of the ring amide requires two water molecules (**TS3-H$_2$O**) assisted by one H_2O only, since one of the water molecules acts as a reactant. As shown in Figure 6.11, a proton-relay occurs, from the water attacking the amide carbonyl to the amide nitrogen, through a water bridge. At the transition state, the proton transfer is incomplete; however the C-N bond distance (1.578 Å) is substantially lengthened compared to 1.380 Å in the ground state. In addition, the nucleophilic –OH group is quite close to the carbonyl carbon (C-O distance 1.814 Å) that it is about to attack.

The first step of the water-assisted gemdiol-mediated mechanism (**TS4-H$_2$O**) also involves two water molecules. In this step, the addition of a water molecule to the amide carbonyl occurs. The transition state structure (**TS4-H$_2$O**) shows the proton has been transferred from the water molecule to the amide carbonyl oxygen (O-H distance 1.018 Å) through a water-bridge, with a rather strong hydrogen-bond network (1.530 Å and 1.581 Å). Incidentally, the C=O double bond is partially broken as indicated by the C-O bond distance of 1.316 Å (1.230 Å in the ground state). The nucleophilic attack distance is 2.104 Å. The cleavage of the ring bearing the gemdiol is depicted in **TS5-H$_2$O** (Figure 6.12). In this step, a proton is transferred from the gemdiol to the adjacent nitrogen via a water molecule. **TS5-H_2O** may be described as an early transition state (reactant-like), since the proton transfer is not complete and the cleavage of the ring is not quite advanced, as indicated by the relatively short C-N bond (1.562 Å compared to the ground state C-N distance of 1.466 Å).

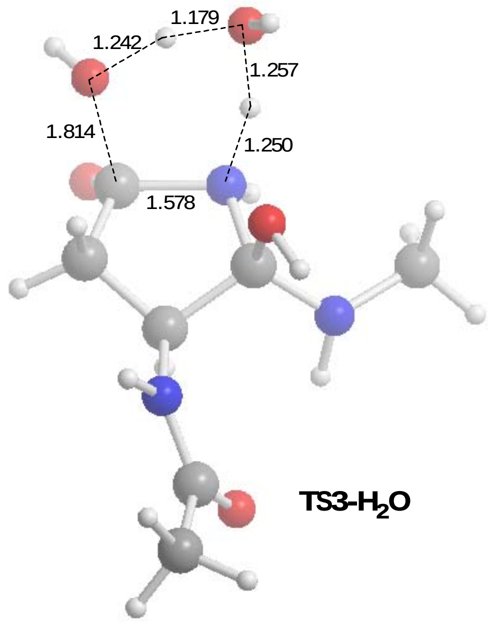


Figure 6.11. Optimized geometry for the transition state of concerted ring amide hydrolysis, one water-assisted.

Activation energies for the stepwise amide hydrolysis (Figure 6.13) shows that both steps are energetically rather close and the rate-determining step of the two-step process is the ring opening step. Previous studies on amide hydrolysis have suggested that a stepwise mechanism going through a gemdiol intermediate has a considerably lower barrier than a concerted reaction [72]. Energetics of the water-assisted amide hydrolysis modeled herein is in accord with this expectation; the stepwise (gemdiol-mediated) route is more favorable than the concerted pathway by almost 10 kcal/mol.

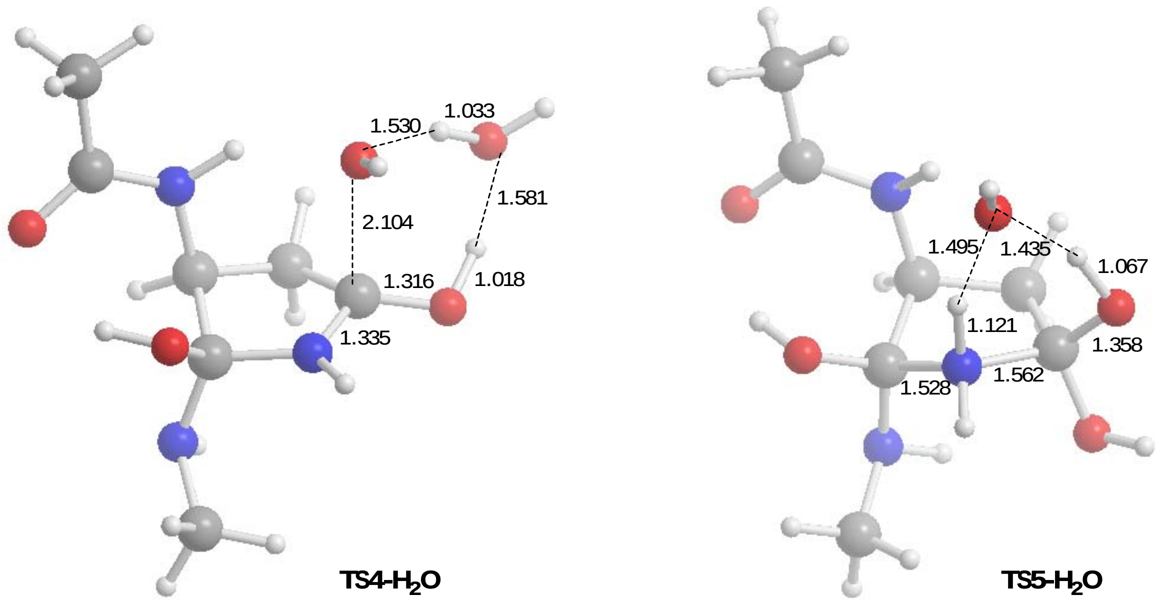


Figure 6.12. Optimized geometries for the transition states of gemdiol-mediated stepwise ring amide hydrolysis, one water-assisted.

The energetic comparison of the possible fates of Intermediate I reveals a barrier difference of approximately 25 kcal/mol between carboxyl-side peptide bond cleavage (Figure 6.8, $\Delta G^{\ddagger}$=14.7 kcal/mol) and ring amide hydrolysis (Figure 6.13, $\Delta G^{\ddagger}$=38.2 kcal/mol). It is apparent that fragmentation is much more favorable than amide hydrolysis at the Intermediate I stage. However, backbone cleavage may be prohibited inside a protein, where the conformational space around the Asn residue is severely restricted by spatial distribution of neighboring residues and/or secondary structure, such as intramolecular H-bonds that form β-sheets or α-helices. In that case, at Intermediate I, ring amide hydrolysis may take place to form Intermediate II which in a subsequent step could undergo deamidation to reveal Asp and/or cleavage at the carboxyl-side peptide bond. This is discussed in the next section.

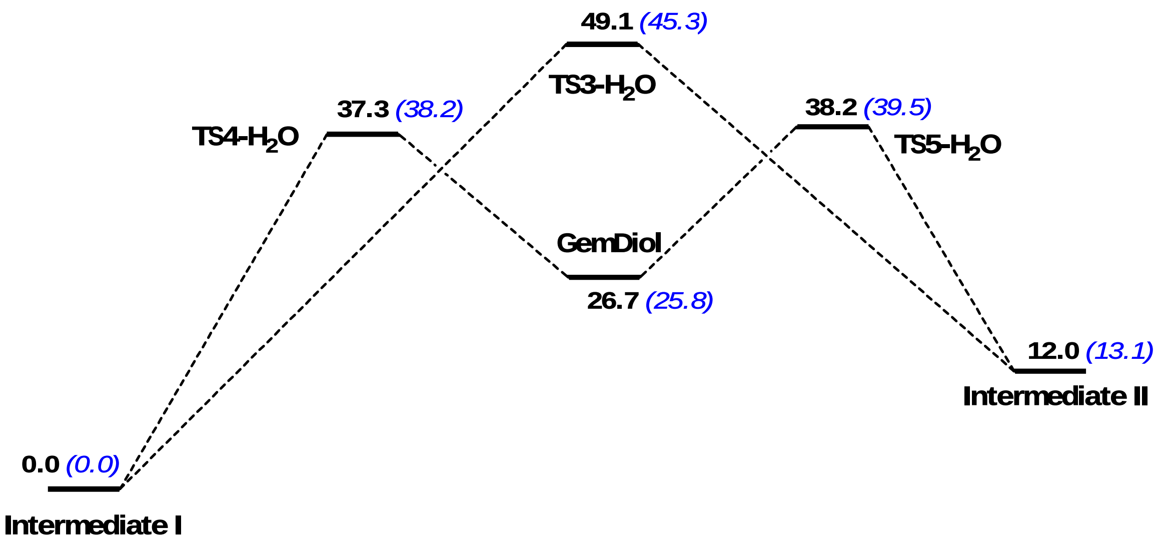
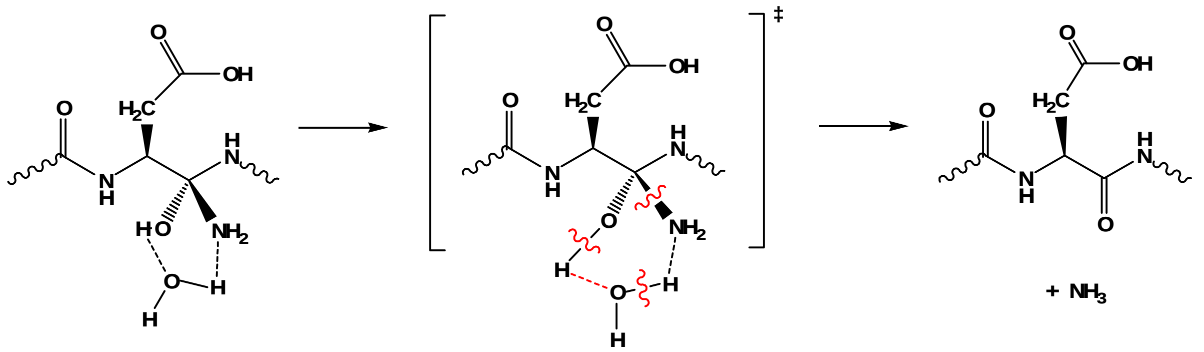


Figure 6.13. Relative free energies for concerted and gemdiol-mediated ring amide hydrolysis at Intermediate I, water-assisted.

<u>6.3.1.3. Fates of Intermediate II.</u> Intermediate II may go through a water-assisted deamination step (Figure 6.14) that will reform the backbone peptide bond and an Asp residue is formed. This route may be suggested as an alternative pathway to the traditional deamidation mechanism (Figure 6.1).

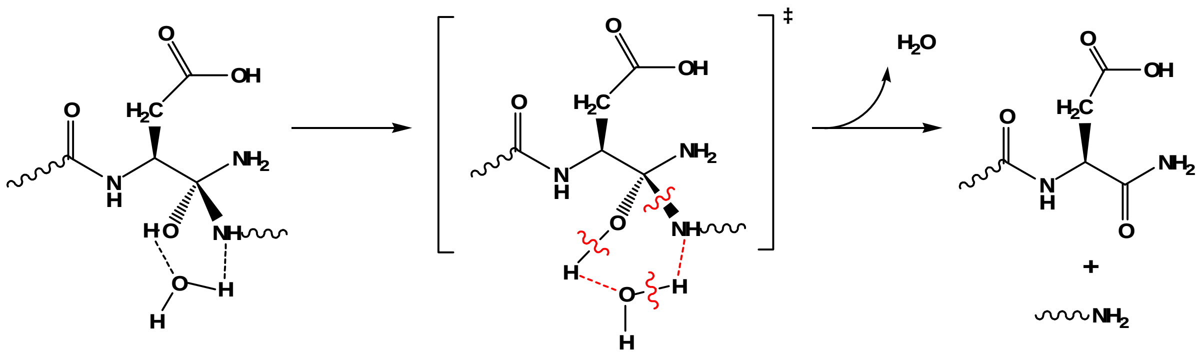
Intermediate II may as well undergo peptide backbone fragmentation to reveal cleavage products, given that proper orientation of the backbone is feasible for this reaction (Figure 6.15). Cleavage products are similar to those at the Intermediate I stage (Figure 6.7), however, in this case there are no rings at any of the tail ends. The availability of a water bridge between these two entities is essential and depends on the position and proper alignment of the backbone NH group.

Relative free energies for deamination and cleavage at the Intermediate II stage are shown in Figure 6.16, together with optimized structures of transition states for these two reactions. The deamination transition state (**TS6-H_2O**) shows that the ammonia is almost expelled (C-N bond distance 1.695 Å). In the cleavage case (**TS7-H_2O**), however, the backbone lengthening has just started (C-N bond distance 1.583 Å). The cleavage route is energetically favorable by almost 6 kcal/mol.



Intermediate II TS6 Aspartic Acid Residue

Figure 6.14. Deamination (NH_3 expulsion) at Intermediate II.



Intermediate II TS7 Cleavage Products II

Figure 6.15. Backbone cleavage at Intermediate II.

<u>6.3.1.4. Deamidation versus Peptide Backbone Cleavage at Asn .</u> The overall energetics of peptide bond cleavage at Asn is depicted in Figure 6.17. The initial cyclization step, with an activation barrier of 47.6 kcal/mol, is incidentally the rate-determining step for peptide bond cleavage at Asn.

Deamidation may proceed through different pathways, as suggested by the large range of experimental reaction rates even for the same primary sequence of amino acids. In general, it has been assumed that the first step in deamidation consists of a concerted cyclization.

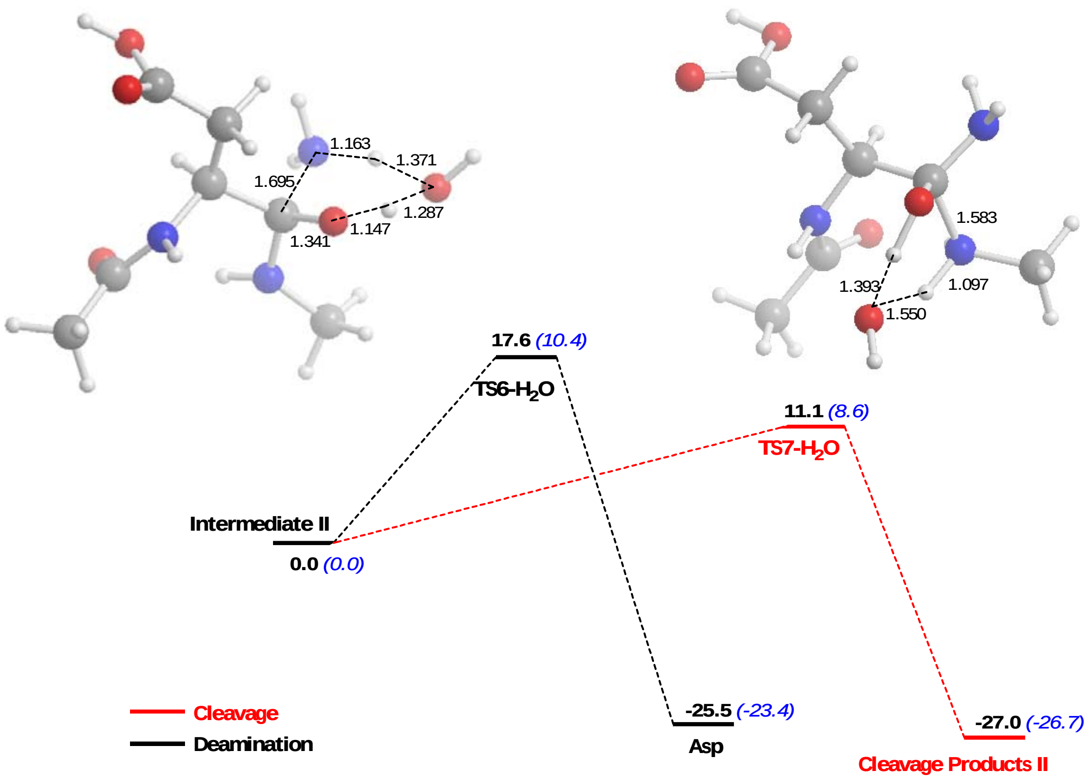


Figure 6.16. Relative free energies and optimized structures for transition states of the possible fates of Intermediate II, one water-assisted.

We have shown that the corresponding activation barrier is comparable to that found for the equivalent concerted cyclization step in backbone cleavage (either in non-assisted or water-assisted mechanisms, see Figures 6.5 and 6.6 respectively). Thus, if deamidation involves concerted cyclization as the first reaction step, deamidation and backbone cleavage would display comparable reaction rates, in contrast with known experimental facts. Actually, we have reported previously that deamidation may proceed through alternative pathways, specifically direct hydrolysis of the Asn side chain or cyclic imide formation via a tautomerization route. Both mechanisms involve lower activation barriers than concerted cyclization provided water assistance is accessible. Comparison between theoretical and experimental data for deamidation versus backbone cleavage suggests therefore that deamidation proceeds through one of these water-assisted processes.

Note that in case backbone cleavage at Intermediate I is inhibited, the hydrolysis of the ring amide may eventually occur giving rise to an Asp residue (Figure 6.9), i. e. the major product of the deamidation reaction. However, since Intermediate I only has one carbonyl functionality (unlike the succinimide ring), its hydrolysis cannot lead to the other deamidation product, the iso-Asp. Direct hydrolysis of the Asn side chain will also result in Asp formation, exclusively. Indeed, there are many instances where only Asp is formed as a result of deamidation.

6.3.2. Peptide Fragmentation at Asp Residues

Asp can also undergo peptide chain cleavage through formation of an anhydride intermediate and by means of a mechanism similar to peptide bond cleavage at Asn. The Asp side chain carboxylic acid group is a better candidate for nucleophilic attack than the Asn side chain NH_2, since the Asp side chain is expected to be present as carboxylate at physiological pH. However, the effective pK_a of the acid may depend substantially on the molecular environment. In the present work, the reaction has been modeled with the carboxylic acid form of Asp rather than the carboxylate anion.

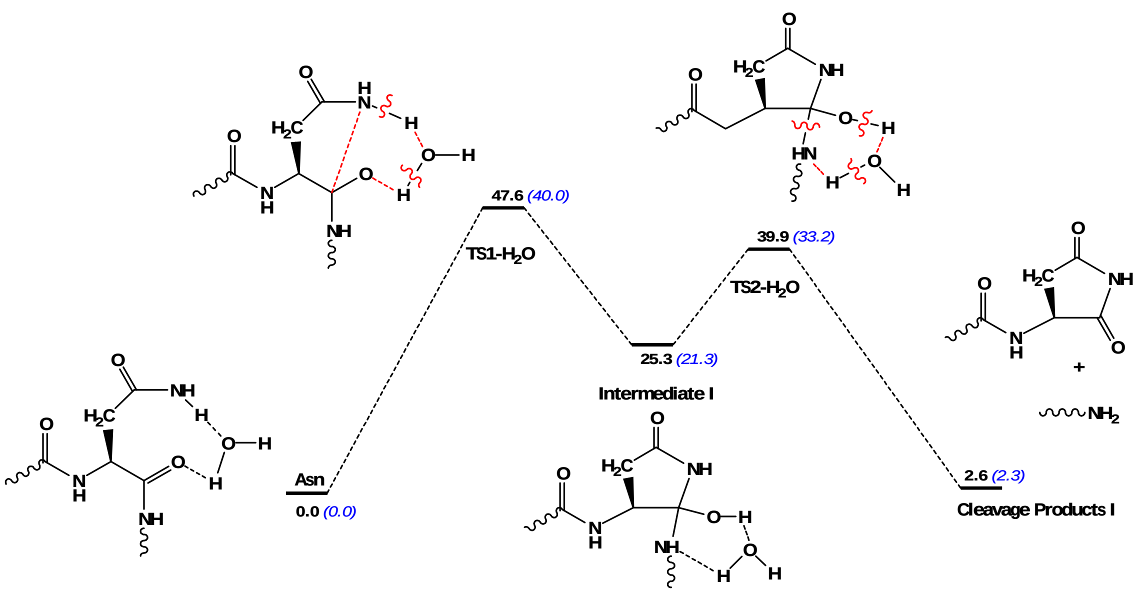


Figure 6.17. Relative free energies for peptide bond cleavage at Asn, water-assisted.

The Asp side chain attacks the backbone carbonyl and a cyclic tetrahedral intermediate (**Asp-Intermediate I**) forms in the initial step (Figure 6.18), much like the case in Asn (Figure 6.3). The ring closure step is followed by the actual cleavage step (Figure 6.18). Cleavage products are similar to those for Asn, except an anhydride ring forms at one of the peptide tail ends, instead of a succinimide ring (Figure 6.7).

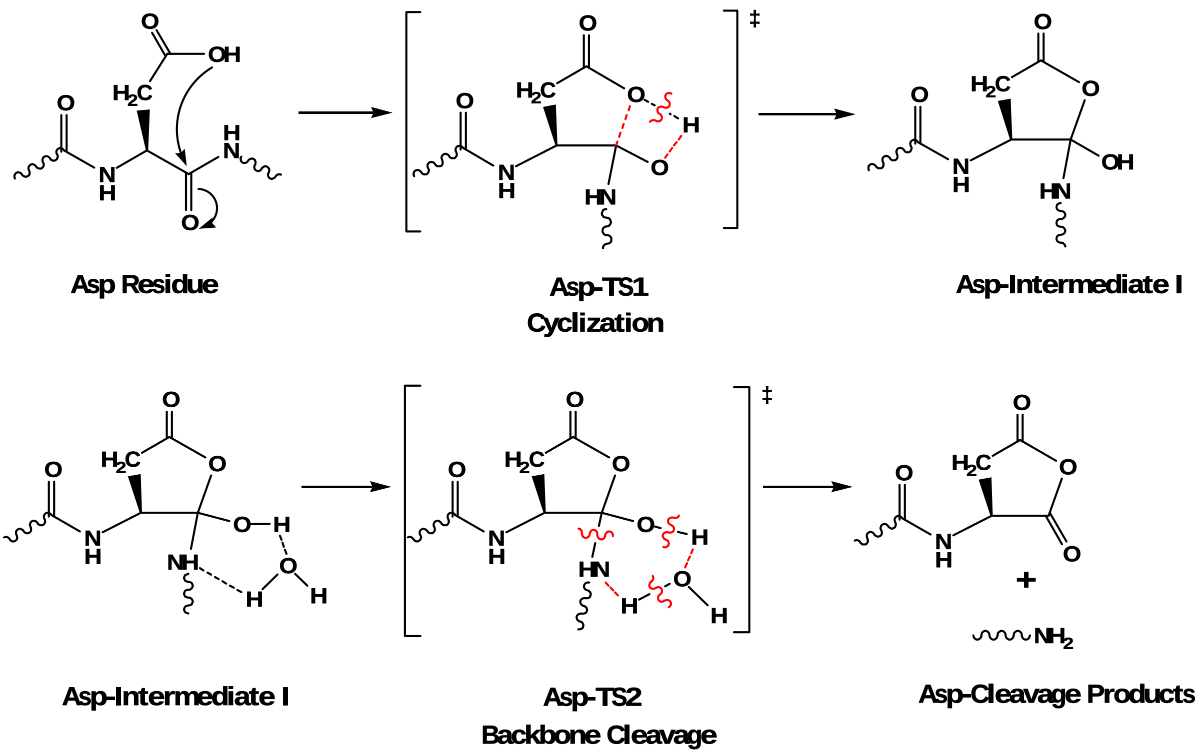


Figure 6.18. Peptide bond cleavage mechanism for Asp residues.

There are other similarities between Asn and Asp reactions. The transition state structure for the non-assisted concerted cyclization mechanism (**Asp-TS1**, Figure 6.19) involves an asynchronous proton transfer accompanied by ring closure. In the water-assisted transition structure (**Asp-TS1-H_2O**, Figure 6.19) proton transfer via the water molecule takes place prior to actual ring closure. Finally, water assistance has slightly decreased the energetics of this transition state.

Cyclization to **Asp-Intermediate I** is followed by backbone cleavage (Figure 6.18) and again there are similarities with the Asn case. The activation barrier for peptide fragmentation step is quite low (Figure 6.20). Hydrolysis of **Asp-Intermediate I** will not be discussed, since the outcome of such a reaction will be the starting structure itself.

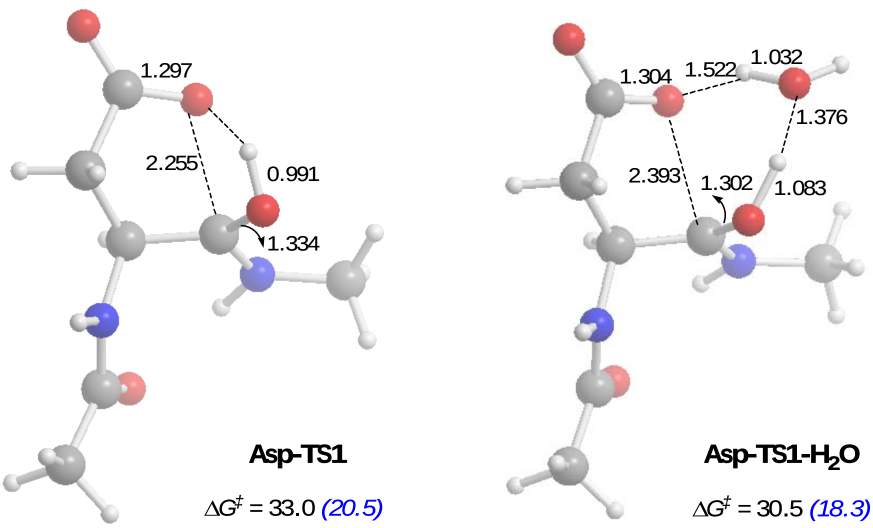


Figure 6.19. Optimized geometries and free energies of activation for the transition state of side chain –OH attack on backbone Asp carbonyl, concerted waterless (**Asp-TS1**) and concerted one water-assisted (**Asp-TS1-H_2O**) mechanisms, respectively.

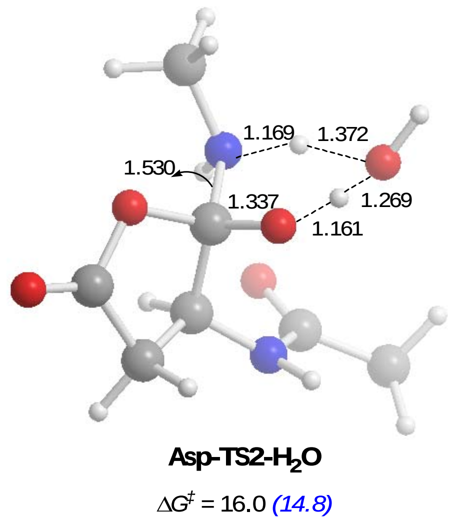


Figure 6.20. Optimized geometry and free energy of activation for transition state of backbone cleavage at Asp-Intermediate I, concerted one water-assisted (**Asp-TS2-H_2O**) mechanisms.

The overall energetics of peptide bond cleavage at Asp is depicted in Figure 6.21. Although the activation barrier for the initial cyclization step is higher (30.5 kcal/mol) than the following cleavage step (16 kcal/mol), the latter is rate-determining in contrast with the Asn case.

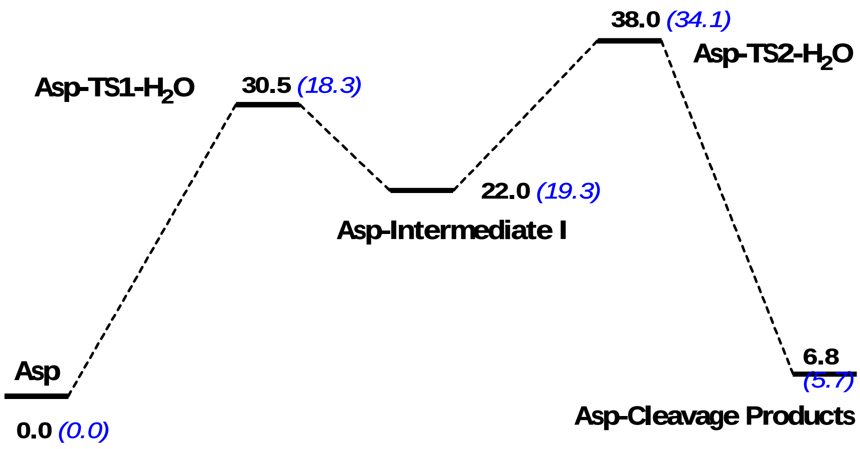


Figure 6.21. Relative free energies for peptide bond cleavage at Asp, water-assisted mechanism.

Analysis of the energetics of backbone fragmentation at Asp residues might shed light on whether fragmentation is more likely to occur subsequent to deamidation of Asn into Asp, rather than cleavage at Asn itself. It is important to emphasize the difference in activation barriers (taken roughly as the relative free energy of the highest TS) of Asn (47.6 kcal/mol, Figure 6.17) and Asp (38.0 kcal/mol, Figure 6.21) cleavage. Peptide bond cleavage is energetically much more favorable for Asp. From the energetic data acquired in this study and previous computational studies done on deamidation of Asn, it is therefore reasonable to suggest that in relatively flexible peptides, backbone cleavage at Asn residues may be preceded by deamidation into an Asp.

6.4. Conclusion

In this study, mechanisms leading to non-enzymatic peptide bond cleavage at Asn and Asp have been investigated using computational methods. Mechanism and energetics of peptide fragmentation at Asn has been comparatively analyzed with previous calculations on deamidation of Asn. The cyclization step was shown to be the rate determining step for backbone cleavage at Asn (Figure 6.17).

Although concerted cyclization for deamidation and cleavage have comparable activation barriers, previous computational studies have shown that deamidation does not necessarily involve such a reaction step. Direct hydrolysis of the Asn side chain and cyclic imide formation via a tautomerization route have lower activation barriers. These processes are water assisted and therefore require the presence of water molecules. Backbone cleavage is unlikely to be competitive with deamidation in peptides with access to solvent molecules.

An important conclusion of this study is the energetics of backbone cleavage at Asp residues. The peptide fragmentation barrier is indeed much lower (approximately 10 kcal/mol) for Asp than that for Asn. We therefore, suggest that cleavage at Asn residues takes place after an Asn residue has deamidated into an Asp. Since the cleavage products differ for Asn (Figure 6.7) and Asp (Figure 6.18), experimental verification of this proposal appears to be quite feasible.

Cases where deamidation is prevented by protein three-dimensional structure, due to hindrance of backbone rotation caused by sterics, secondary structure and/or spatial distribution of neighboring residues, backbone cleavage of the Asn residue may ensue as a competitive reaction.

7. PRIMARY SEQUENCE DEPENDANCE OF ASPARAGINE DEAMIDATION RATES

Previous chapters mainly focused on the mechanistic details, energetics and feasibility of the formerly suggested succinimide-mediated deamidation route, which involves the Asn *n* +1 residue. As a result the most plausible pathways for deamidation have been outlined. A recent experimental study on pentapeptides has shown that Asn deamidation rates are directly related to primary structure (peptide sequence) near Asn, with a more prominent effect from the carboxyl-side (*n* +1) residue. This is in accord with the succinimide mechanism; however, the effect of the identity of the *n* +1 residue is unclear. The objective of the following study is to explore the correlation between experimental deamidation rates and primary structure of peptides and proteins, in order to identify the factor(s) causing this dependence. Quantum-mechanical calculations were performed in order to compare activation barriers for systems with different *n* +1 residues. On the other hand, molecular dynamic simulations were used to better understand the differences in the dynamic structure. In addition to the previously suggested succinimide-mediated deamidation pathway that was used as a benchmark during this study, an alternative route for deamidation –involving the Asn amino-side– has been proposed here for the first time. This paper is still a work in progress and the final form of the paper is intended for publication in *Biochemistry*.

Deamidation rates in peptides and proteins are known to be influenced by primary structure, i.e. the peptide sequence near the asparagine (Asn) residue. In a current study, Robinson et al. showed that Asn deamidation half-times for pentapeptides are primarily affected by the carboxyl-side (*n* +1) residue, with less influence from the amino (*n* -1) side. This is consistent with the previously suggested succinimide-mediated mechanism, which involves the attack of the carboxyl-side backbone NH to the Asn side-chain carbonyl to form a cyclic intermediate. However, the effect of the identity of the *n* +1 residue is unclear. This study uses computational techniques in an attempt to identify the dominant factor(s) causing the variation in deamidation rates of pentapeptides with different primary sequence. The succinimide-mediated mechanism previously studied with a smaller model peptide (ACE-Asn-NME) was modeled (B3LYP/6-31+G**) using a larger system (ACE-Asn-Yyy-NME) and different n +1 (Gly, Ala and Val) residues. The succinimide-mediated reaction barriers for all three systems were energetically compared, in an effort to see whether activation barriers are similar for Asn with different carboxyl-side residues. In a separate attempt to understand the dynamic differences between these systems, molecular dynamic (MD) simulations were performed on three pentapeptides, Gly-Gly-Asn-Yyy-Gly, –where Yyy is Gly, Ala and Val– with significantly different experimental deamidation half-times. MD simulations were analyzed in terms of solvent accessibility, backbone structure, intramolecular hydrogen-bonds, water bridges and per cent occurrence of potentially reactive conformations. In addition to the previously suggested succinimide-mediated deamidation pathway that was used as a benchmark during this work, an alternative route for deamidation –involving the Asn amino-side– has been proposed here for the first time.

7.1. Introduction

Non-enzymatic asparagine (Asn) deamidation occurs spontaneously under physiological conditions [7-9, 15] and is known to limit the lifetime of peptides and proteins [48, 49]. The conversion of Asn to aspartate (Asp) has been suggested to go through a stepwise pathway involving a cyclic imide [19, 54](Figure 7.1). Detection and isolation of succinimide derivatives as well as isomerization of Asp to iso-aspartate (iso-Asp) during some peptide deamidation experiments has supported this idea [19, 54, 55, 59, 67, 68].

The succinimide-mediated Asn deamidation mechanism has been subject to several computational studies and computational [56-58, 69-71, 73]. A recent computational study has shown that the rate-determining step of this mechanism is the hydrolysis of the cyclic intermediate; this explains the relatively long lifetime of the succinimide intermediate and its detection/isolation during peptide deamidation experiments. However, there are many cases where an intermediate was not observed and deamidation did not lead to iso-Asp formation; this has been explained through the presence of another mechanism, *direct hydrolysis*, which was shown to be operating at ease under physiological conditions.

Direct hydrolysis is the conversion of the Asn side chain amide into a carboxylate through a non-cyclic concerted step. It is well-known that the hydrolysis of amides is catalyzed in the presence of acid or base [72]; however, it was shown that direct hydrolysis is at least as feasible as the succinimide-mediated route even under physiological conditions.

All Asn residues in peptides and proteins are inherently unstable and are prone to deamidation. The conversion of the neutral amide side chain (Asn) into a negatively charged carboxylate (Asp) causes substantial conformational changes in peptides and proteins. There are many reports that these changes markedly affect protein function or stability or both [68]. Robinson et al. have studied the primary sequence dependence of deamidation rates on pentapeptides [59]. They found that deamidation rates could be varied over a wide range by changing primary sequence. The timed processes of protein turnover, development, and aging have been suggested as possible roles for deamidation [51]. It has been hypothesized that deamidation serves as a *molecular clock* for the timing of biological processes [52].

Robinson et al. have measured the rates of deamidation of 64 peptides between 3 to 13 residues in length [68]. They showed that a pentapeptide model, where Asn is the central amino acid, is adequate to analyze the primary sequence dependence of Asn deamidation. The model Gly-Xxx-Asn-Yyy-Gly was designed so that the charges on the peptide ends are moved away from the amide by the Gly residues, without making the peptide so long that it might adopt special three-dimensional configurations.

The complete 800-pentapeptide set of all possible combinations of the sequences Gly-Xxx-Asn-Yyy-Gly and Gly-Xxx-Gln-Yyy-Gly, where Xxx and Yyy are any of the 20 naturally occurring amino acid residues were synthesized [68]. Deamidation rate was shown to be controlled primarily by the carboxyl side residue (Yyy) with smaller effects from the amino side residue (Xxx) (Figure 7.2). This is consistent with the succinimide reaction mechanism that was originally proposed to explain the unusually rapid deamidation rates of Asn-Gly sequences and iso-Asp formation [19, 54].

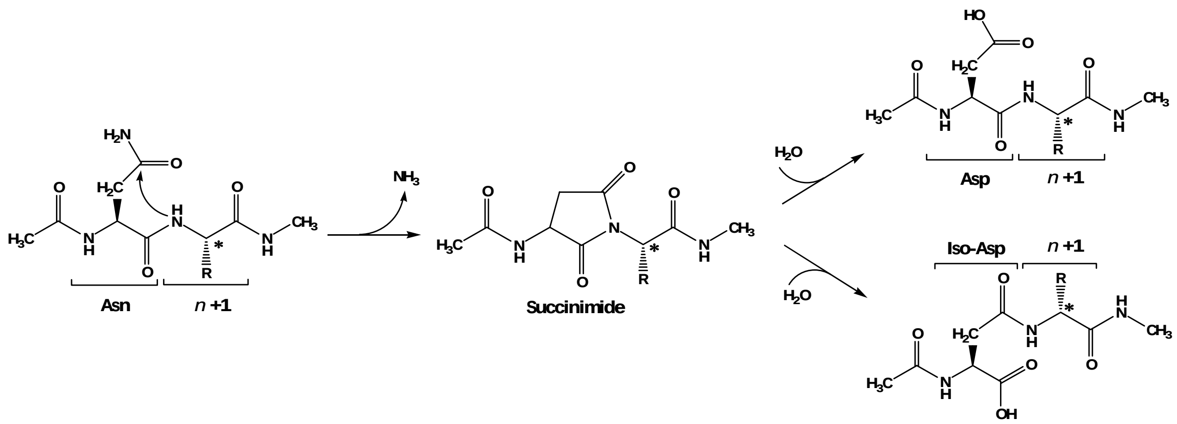


Figure 7.1. Deamidation *via* a succinimide intermediate.

Individual deamidation rates of these peptides revealed some remarkable characteristics (Table 1). Glycine (Gly) –with no interfering side chain– exhibited an unusually short half-time of 1 day; this was associated with the ease of succinimide ring formation. The additional methyl group in alanine (Ala) increased the half-time to 25 days; larger aliphatic groups increased the half-time even more, with the most extreme example being isoleucine (Ile), which has a mean half-time of about 320 days. Increase in steric hindrance was suggested as the reason for the difference in half-times.

However, in light of previous computational studies, which have proved the catalytic effect of water molecules, the size of the hydrophobic side chain could also effect deamidation rates by inhibiting solvent accessibility; this would impair deamidation whether it proceeds through the cyclic intermediate or through direct hydrolysis.

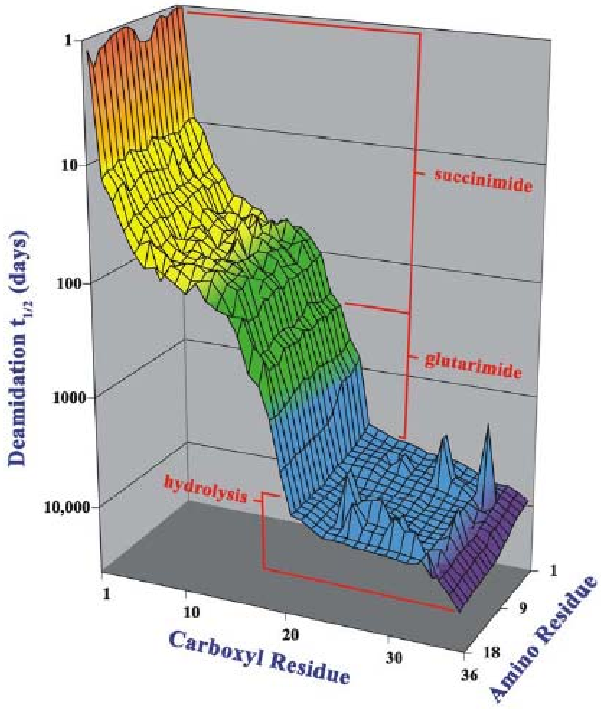


Figure 7.2. Sequence dependence of Asn pentapeptide half-times.

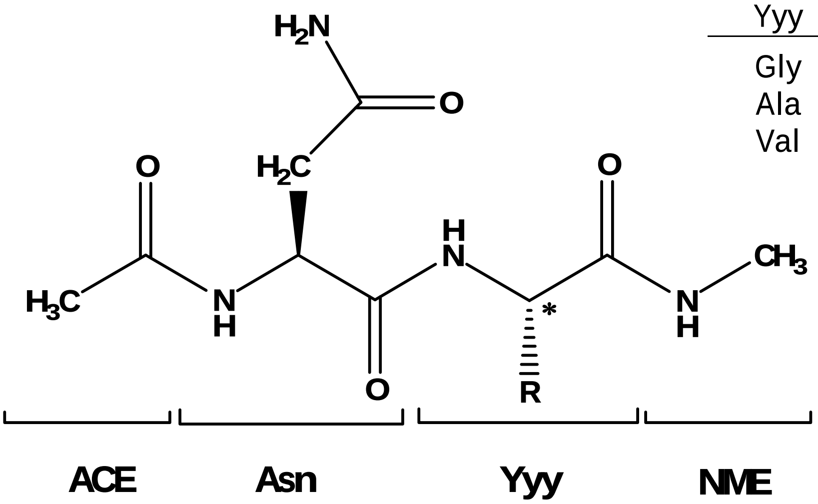
Significant deamidation of Asn-Pro (proline) sequences were not observed even after 1000 days. Since the backbone nitrogen is covalently bound to the side chain, Asn-Pro is not expected to deamidate through the succinimide mechanism. On the other hand, polar and charged side chains were shown to cause an increase in the deamidation rate.

Histidine (His), which is expected to have similar steric hindrance to that of phenylalanine (Phe), deamidates 7 times faster. Similarly threonine (Thr), which is similar to valine (Val) in terms of shape and size, has a deamidation rate that is 5 times faster, possibly due to the presence of the OH group, which facilitates H-bonding with nearby solvent molecules, rather than repel them.

Table 7.1. Mean deamidation half-times for Asn-Yyy sequences in pentapeptides.

Carboxyl-Side Residue	[a]Half-Time (Days)
Gly	1.2
His	10
Ser	16
Ala	25
Asp	32
Thr	47
Cys	54
Lys	59
Met	61
Glu	64
Arg	65
Phe	69
Tyr	81
Trp	99
Leu	130
Val	250
Ile	320
Pro	> 1000

The fact that the identity of the *n* +1 residue has a major effect on deamidation half-times of pentapeptides –while the *n* -1 residue's effect is comparably negligible– is anticipated due to the succinimide-mediated mechanism, where the initial cyclization step involves the attack of the carboxyl-side NH. The size and charge of the *n* +1 side chain is expected to affect the mode of attack, hence enhancing or inhibiting deamidation. However, there is no direct correlation between size and charge of the carboxyl side residue and deamidation half-times listed for pentapeptides (Figure 2). Phenylalanine, tryptophan (Trp) and tyrosine (Tyr) –three amino acids with aromatic side chains– have much lower deamidation half-times than Val, which bears a small isopropyl group. It is important to understand the factors that cause the differences in deamidation rates, keeping in mind that there may be several parameters operating simultaneously.



Yyy	R
Gly	H
Ala	CH_3
Val	$CH(CH_3)_2$

Figure 7.3. Model peptide with Asn-Yyy residue.

The objective of this study is to get a deeper insight on primary sequence dependence of pentapeptide deamidation rates, using computational methods. The succinimide-mediated mechanism will be modeled with different n +1 residues and energetically compared, in an effort to see whether activation barriers are similar for different carboxyl-side residues. This requires quantum-mechanical calculations on a model compound which encompasses the carboxyl side residue. For this purpose, three end-blocked dipeptides, ACE-Asn-Yyy-NME (Figure 7.3) –where Yyy is Gly, Ala and Val– were chosen and modeled with regard to the succinimide-mediated mechanism, in light of the knowledge gained from previous work.

Pentapeptide deamidation half-times (Table 1) are quite different for Gly, Ala and Val (1, 25 and 250 days, respectively); this is expected to be reflected in deamidation energetics, in case the difference is caused by variance in activation barriers. However, there may be an alternative explanation for the large difference in rates of these three pentapeptides. We have shown before that water molecules play a crucial role in deamidation mechanisms by assisting proton transfer and hydrolysis steps. Accordingly, the increase in deamidation half-times for the three pentapeptides might be correlated to the increase in hydrophobicity of the three side chains (Gly, Ala and Val, respectively), since water access to the backbone NH may be increasingly hindered as the side chain gets larger. To explore this hypothesis, molecular dynamic (MD) simulations were performed on these pentapeptides, Gly-Gly-Asn-Yyy-Gly –where Yyy is Gly, Ala and Val– in order to investigate the difference in solvent distribution. Molecular dynamic simulations will also enable a comparison with respect to the per cent occurrence of the reactive conformers observed in quantum-mechanical calculations, also known as near attack conformers (NACs). For this purpose, MD simulations will be studied in detail, with respect to accessibility of solvent molecules, backbone structure, intramolecular H-bonds and water bridges.

Apart from the results of the pentapeptide experiment, there are several experimental studies where the identity of the *n* -1 residue is influential on deamidation rate. This may be due to a different mechanism that is in effect, which involves the amino-side of the Asn residue. Exploring the presence and/or feasibility of such a mechanism through quantum-mechanical calculations is an additional aim in this study.

7.2. Computational Methodology

There are two main theoretical approaches in this study; energetic comparison among three different dipeptides (Figure 7.3) will be facilitated by quantum-mechanical (QM) calculations utilizing the Density Functional Theory (DFT); the alternative route involving the Asn amino-side will also be explored with this methodology. In addition, molecular dynamic (MD) simulations will be used in order to investigate the time-dependent behavior of three pentapeptides (Gly-Gly-Asn-Yyy-Gly, where Yyy is Gly, Ala and Val) and their dynamic differences. Methodological details and practical aspects of the two methods are described below in detail.

7.2.1. QM Calculations

The same level of theory and basis set are used with aforementioned computational studies on deamidation. Full geometry optimizations were performed in gas-phase – without any constraints– using the density functional theory (DFT) [34] at the B3LYP/6-31+G** level [36-40]. The use of this basis set and method in similar peptide systems is well established. Stationary points were characterized by a frequency analysis. Zero-point energy and thermal corrections were attained using the ideal gas approximation and standard procedures. Local minima and first-order saddle points were identified by the number of imaginary vibrational frequencies. The species reached by each transition structure was determined by intrinsic reaction coordinate (IRC) calculations [60, 61]. Free energies of activation ($\Delta G^{\ddagger}$) are calculated as the difference of free energies between transition states and reactants (reactant-water complex) of each step. Energy values for gas-phase optimizations listed throughout the discussion include thermal free energy corrections at 298 K and 1 atm. All calculations have been carried out using the Gaussian 03 program package [63]. All distances and free energies listed in the discussion are in angstroms (Å) and kcal/mol, respectively.

7.2.2. MD Simulations

Molecular dynamic simulations have been performed using the AMBER 9.0 program package. All simulations were performed in water, using the TIP3P water model. Calculations were employed at 300 K in the isobaric-isothermal (NPT) ensemble, using a single pentapeptide molecule in a cubic box with periodic boundary conditions. Each system contained approximately 12 600 atoms and 4180 water molecules. The step size was chosen to be 0.001 ps. Nonbonded interactions were cut off at 9.0 Å for the direct sum and Particle Mesh Ewald was employed to account for long-range electrostatic interactions. For each pentapeptide, a 5 ns simulation was performed and coordinates were saved every 0.2 ps for analysis.

7.3. Results and Discussion

The first part of this study is based on quantum-mechanical calculations on the succinimide-mediated pathway (Figure 7.1) using a new and larger model compound (Figure 7.3) that will mimic the Asn residue together with its carboxyl-side residue. In addition, calculations on the deprotonation of the backbone NH for different *n* +1 residues will be discussed. The second part of the discussion consists of analysis of molecular dynamic simulations performed on three pentapeptides, with significantly different deamidation half-times, in order to get a dynamic –rather than a static– perspective. Finally, an alternative deamidation mechanism, involving the Asn amino-side, will be discussed.

7.3.1. Energetics for Different *N* +1 Residues

Primary sequence dependence of Asn deamidation was explored through the succinimide-mediated mechanism, which was modeled with different *n* +1 residues (Figure 7.3) in order to see the differences in activation barriers. For this purpose, three end-blocked dipeptides, ACE-Asn-Yyy-NME –where Yyy is Gly, Ala and Val– were used and modeled in light of the knowledge gained from previous computational studies on deamidation.

Succinimide formation was previously shown to proceed through several mechanisms; the tautomerization route (Figure 7.4) was suggested to be the most plausible pathway; this route consists of three steps, the first being the tautomerization of the Asn side-chain amide. Asn side-chain tautomerization is not expected to be noticeably effected by the *n* +1 side chain, at least in the case of peptides with considerable flexibility and large solvent accessibility. One-water assisted tautomerization revealed identical barriers (19.8 kcal/mol) for all Asn-Yyy dipeptides studied, as anticipated.

However, if the hydrophobicity of the carboxyl-side R group defers water access to the Asn side chain, this may effect the ease of tautomerization, hence deamidation half-times. Therefore, inside the three-dimensional structure of a protein, a large aliphatic group nearby the Asn residue may significantly decrease deamidation rates.

The cyclization step leading up to the tetrahedral intermediate was previously proposed to be the rate-determining step for the formation of the succinimide intermediate (Figure 7.4). This step is followed by water-assisted deamination to reveal the succinimide. In order to see the effect of the *n* +1 side chain on reaction barriers, the concerted cyclization step was modeled with no water, one-water and two-water assistance (Table 2).

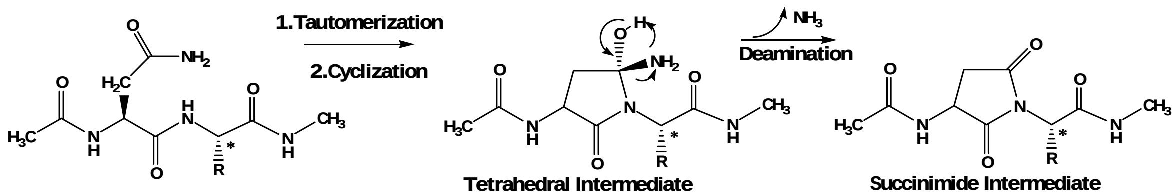


Figure 7.4. Tautomerization route for succinimide formation.

Water-assistance has lowered the activation barriers for all Asn-Yyy sequences, the catalytic effect of solvent molecules on deamidation rates has already been previously suggested. The effect of the *n* +1 side chain is apparent in the cyclization barriers listed (Table 2); as the size of the Yyy side chain increases cyclization becomes more difficult, possibly due to steric hindrance. The experimental deamidation half-times listed for pentapeptides with Yyy as Gly, Ala and Val correspond to barrier differences of 1.9 and 3.3 kcal/mol for Gly/Ala and Gly/Val, respectively (calculated using the relationship $\delta\Delta G = -RT\ln(t_2/t_1)$ where t is half-time, T=310K, and we assume the same number of assisting water molecules for all pentapeptide reactions). Calculated barrier differences –although consistent in terms of trend– do not reproduce the same results as experiment, this suggests that the rate difference does not solely depend on the difference in activation barriers for the three Asn-Yyy sequences under study. Nonetheless, molecular dynamic simulations will provide further information on solvent accessibility and per cent occurrence of near attack conformations (NAC).

Table 7.2. Activation barriers for cyclization in Asn-Yyy sequences.

Yyy	Deamidation [a]Half-times	[b]$\Delta G^{\ddagger}$		
		no H_2O	1 H_2O	2 H_2O
Gly	1	44.1	40.0	37.1
Ala	25	45.6	42.4	38.0
Val	250	48.0	47.2	40.3

A current computational study has suggested that the rate-determining step for the overall deamidation process is in fact the succinimide hydrolysis step (Figure 7.1) and not the cyclization step leading to the formation of this intermediate (Figure 7.4.). However, the aforementioned study was performed on a smaller model system and it has been suggested that the identity of the *n* +1 residue might reverse the order of the rate-determining step, considering that the differences were not very significant. On the other hand, the effect of the carboxyl-side R group is not expected to increase the barriers for hydrolysis of the ring carbonyl and we have not carried out computations for such a reaction step. Thus, if hydrolysis is the rate-determining step in deamidation of pentapeptides, the role of the *n* +1 residue should probably be related to the change in water access to the reaction site, i.e. to the hydrophobicity of the side chain near the Asn residue.

7.3.1.1. Deprotonation of the *n* +1 residue for different Asn-Yyy sequences. The ease of deprotonation of the *n* +1 backbone NH is essential for deamidation studies. It could be considered as an important measure of susceptibility to cyclization; once the deprotonated structure is formed the cyclization step will be much easier. A previous computational study has shown that the barriers for cyclization are as low as 10 kcal/mol for the anionic nitrogen case.

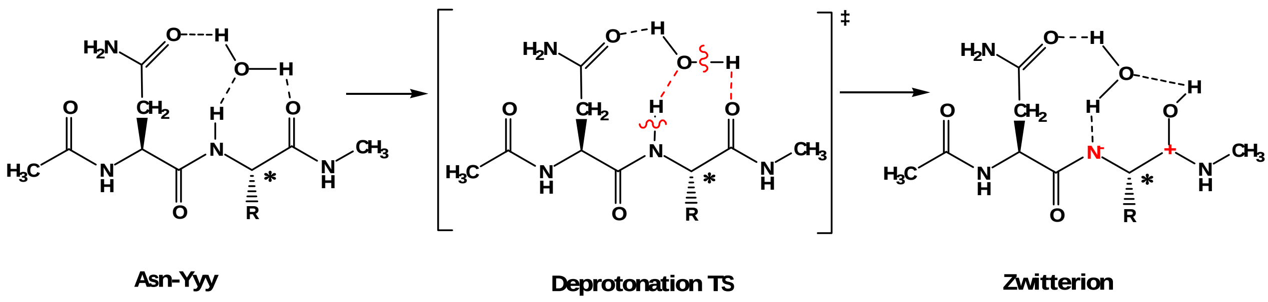


Figure 7.5. Deprotonation of the *n* +1 backbone NH.

The deprotonation reaction (Figure 7.5) involves a proton transfer –through a water molecule relay– from the NH to the C=O group in the *n* +1 residue. The product is a zwitterion, where the backbone NH has been deprotonated, making it more nucleophilic and in this way, favoring the cyclization step. Activation barriers for the deprotonation of the backbone NH of ACE-Asn-Yyy-NME (where Yyy is Gly, Ala and Val, respectively) were calculated to be 31.2, 32.6 and 33.0 kcal/mol, respectively. This is consistent with the experimental deamidation trend for the corresponding pentapeptides.

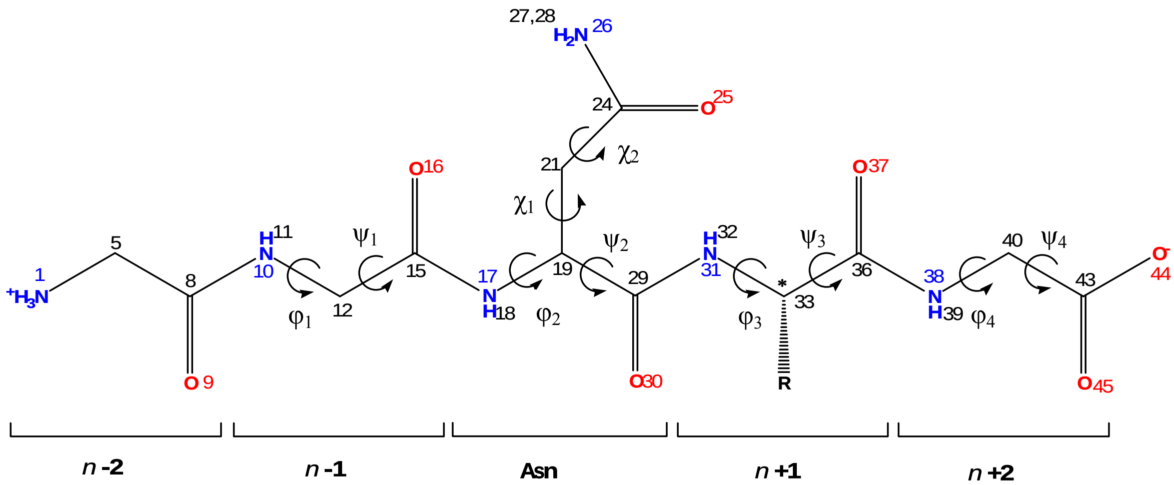


Figure 7.6. Pentapeptide Gly-Gly-Asn-Yyy-Gly.

Quantum-mechanical calculations discussed in this section give an idea of how the increase in size of the hydrophobic side chain on the carboxyl-side of Asn will change activation barriers and the difference in free energy of deprotonation for the *n* +1 backbone NH, which are closely related to the deamidation process. However, there are dynamic aspects in reactions that cannot be accounted for with this methodology. For this purpose, MD simulations have been performed in order to better understand the differences in time-dependent behavior of these systems in solution.

7.3.2. Molecular Dynamic Simulations of Pentapeptides

In this section, results of molecular dynamic (MD) simulations performed on three pentapeptides (Gly-Gly-Asn-Yyy-Gly, where Yyy is Gly, Ala and Val) will be discussed. MD simulations will be analyzed in terms of solvent accessibility, intramolecular water bridges, backbone structure, and per cent occurrence of near attack conformations (NAC).

7.3.2.1. Intramolecular Hydrogen-bonds. Information on intramolecular hydrogen-bonds between proton-acceptor and proton-donor sites is of substantial value; these interactions will be an indicator for the occurrence of reactive conformers (near-attack-conformers) throughout the course of the simulation. Prominent H-bonding interactions have been listed in Table 3; occurrence rates below 1% were discarded. Distances for H-bonds between proton-acceptor/proton-donor were taken as < 2.3 Å. Atomic numbering (Figure 7.6) is identical for all three pentapeptides regardless of the identity of the *n* +1 residue.

Intermolecular H-bonds are diverse for the three pentapeptides studied (Table 3). These peptides are quite small and rather flexible, therefore binding secondary structure is not expected, however, there seems to be a reasonable amount of long-lasting prominent intermolecular interactions, especially for the Ala and Val cases, as seen from the interaction between H^{39}-O^{16}. Long-lasting intramolecular H-bonds between sites on the peptide backbone suggest that a level of secondary structure exists. High per cent occurrence of H-bonds for a particular atom suggests that it is not readily available to interact with other groups, as is the case for H^{32}. Nonexistence of H-bonds also provides valuable information; the absence of a H-bond between O^{25}-H^{32} is noteworthy. This set of data does not provide a meaningful explanation for the deamidation rate difference observed in the pentapeptide experiment, however, it is important to emphasize that H-bonding pattern is rather different for the three pentapeptides that only differ in terms of an R group on the *n* +1 residue.

Table 7.3. Per cent occurrence of intramolecular H-bonds among proton-accepting and donating sites.

H-bonds	% occurrence		
	Gly	Ala	Val
H^{11}-O^{16}	3.8	6.4	15.8
H^{18}-O^{9}	5.0	5.8	3.6
H^{28}-O^{37}	2.5	-	-
H^{32}-0^{16}	2.3	1.8	3.0
H^{32}-O^{37}	12.9	3.9	1.6
H^{32}-O^{9}	16.5	11.9	14.6
H^{39}-O^{16}	5.4	34.2	29.5
H^{39}-O^{9}	2.5	7.7	15.5

7.3.2.2. Intramolecular Water-bridges. Previous computational studies on deamidation have revealed the importance of intramolecular water-bridges, which enable transfer of protons and catalyze steps in the deamidation process. In light of this knowledge, MD simulations have been analyzed and long-lasting, prominent water-bridges among potential proton-accepting and donating sites were listed for all three pentapeptides. Per cent occurrences of water-bridges were evaluated for the entire simulation (Table 4). Distances for water bridges between proton-acceptor/proton-donor were taken as < 2.3 Å.

The significant per cent occurrence of water-bridges between O^{25}-H^{27} indicates that favorable structures for water-assisted side-chain tautomerization are frequently encountered for all three pentapeptides. The nonexistence of certain water-bridges throughout the simulation is also very relevant. The absence of a water bridge between O^{25}-H^{32} indicates that the reactive conformer (near-attack-conformers) for one of the prominent water-assisted cyclization mechanisms is not readily encountered.

Table 4.4. Per cent occurrence of water-bridges among proton-accepting and donating sites.

Water-bridges	% occurrence		
	Gly	Ala	Val
H^{11}-O^{16}	3.9	4.3	8
H^{11}-O^{40}	4.2	1.4	1.5
H^{18}-O^{37}	5.0	1.1	-
H^{18}-O^{9}	3.7	1.9	2.2
H^{27}-O^{25}	13.5	16.8	12.8
H^{27}-O^{30}	4.8	5.0	3.4
H^{28}-O^{16}	6.9	4.9	3.4
H^{28}-O^{30}	3.5	2.3	1.6
H^{28}-O^{37}	6.3	-	-
H^{32}-O^{37}	6.7	2.0	2.0
H^{32}-O^{9}	8.1	7.5	1.7
H^{39}-O^{16}	1.5	3.3	1.5
H^{39}-O^{9}	3.7	1.5	-

7.3.2.3. Backbone Structure (Near-Attack-Conformation). Pentapeptide MD simulations were also analyzed in terms of critical distances and dihedrals. The aim was to pinpoint differences in the evolution of three-dimensional structure for the three pentapeptides under study. Atoms with potential reactivity were particularly important; in light of the deamidation mechanisms previously studied, distances between nucleophilic atoms and electrophilic centers were analyzed. In addition, backbone dihedrals were studied, in order to deduce the three-dimensional structure of the peptide backbones as well as the relationship between Asn side-chain amide and the peptide backbone. Graphs shown in Figures Figure 7.8-7.19 are for Gly, Ala and Val, respectively. Distances are in Å, dihedrals are in degrees.

In terms of the cyclization step in the succinimide-mediated deamidation reaction the time-dependent behavior of the nucleophilic and electrophilic sites are of great importance. The evolution of the distances between C^{24}-N^{31} an O^{25}-H^{32} during the course of the simulation is given in Figure 7.8 and 7.9. An average distance of 4 Å is observed for C^{24}-N^{31}; in terms of the nucleophilic attack leading to ring closure, there is no apparent effect of the difference in *n* +1 residues. In addition the O^{25}-H^{32} distances are rather large for proton transfer to occur.

Peptide backbone dihedrals are important in terms of understanding the orientation of reactive groups with respect to one another. For this purpose, critical dihedrals have been plotted for the three pentatpeptides (Figure 7.10-7.19).

The most important observation in terms of the peptide backbone is the orientation of N^{31} that can be seen in the evolution of the dihedral ψ_2 as well as the orientation of the Asn side-chain which is indicated by dihedral χ_1 (Figure 7.18). A newman projection (Figure 7.7) better demonstrates the orientation of N^{31} with respect to the Asn side-chain. The progress of dihedrals ψ_2 and χ_1 –throughout the MD simulation– indicate that N^{31} and C^{24} are never on the same side of the peptide backbone, except for a short instance for Gly (see χ_1 in Figure 7.18). This is true for all three pentapeptides and it is important to emphasize that if cyclization is to occur a major rotation must take place either on the peptide backbone (ψ_2) or on the Asn side chain (χ_1).

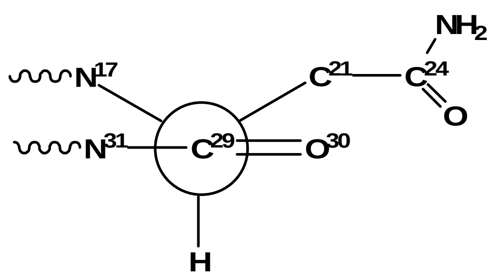


Figure 7.7. Newman projection for C^{29}-C^{19}

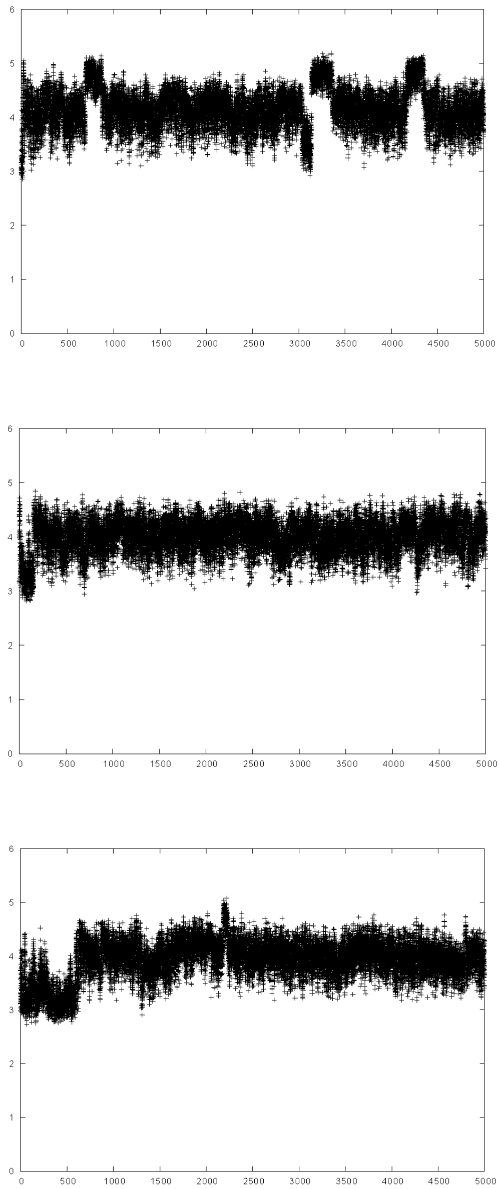


Figure 7.8 Critical distances $\mathbf{C^{24}}$-$\mathbf{N^{31}}$.

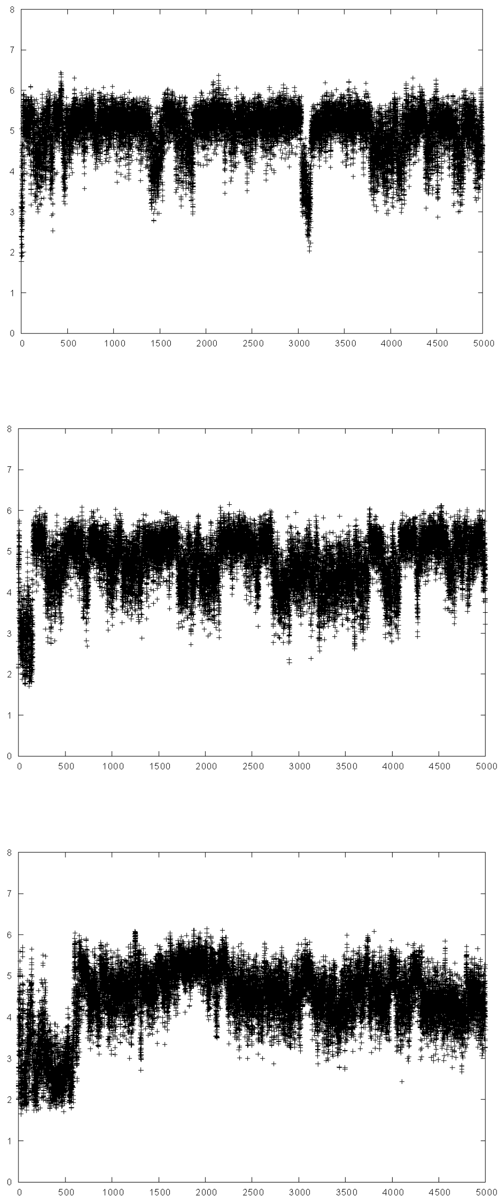


Figure 7.9. Critical distances **O^{25}-H^{32}**.

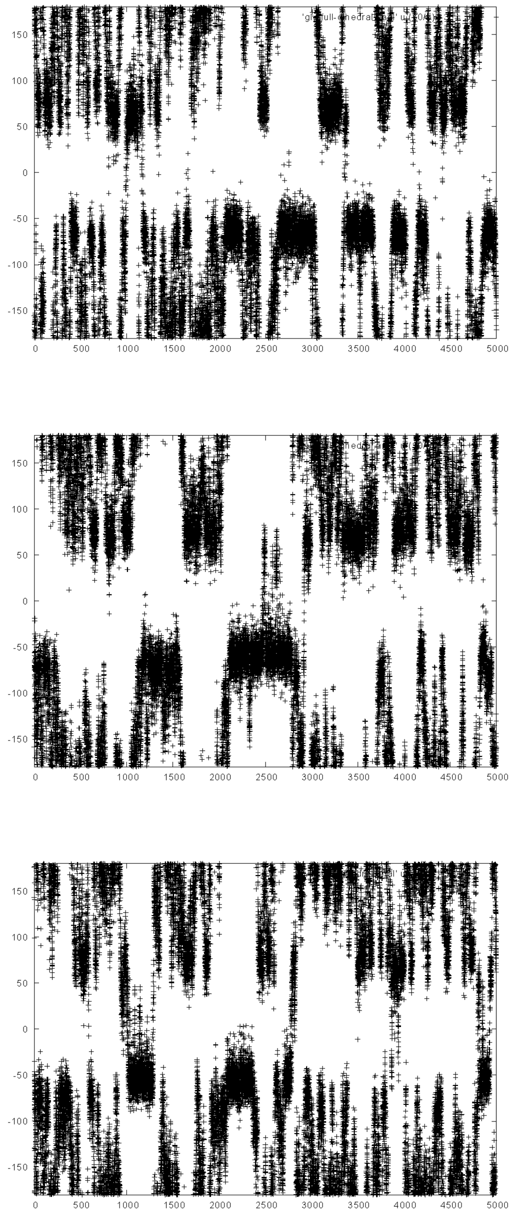


Figure 7.10. Critical dihedrals φ_1.

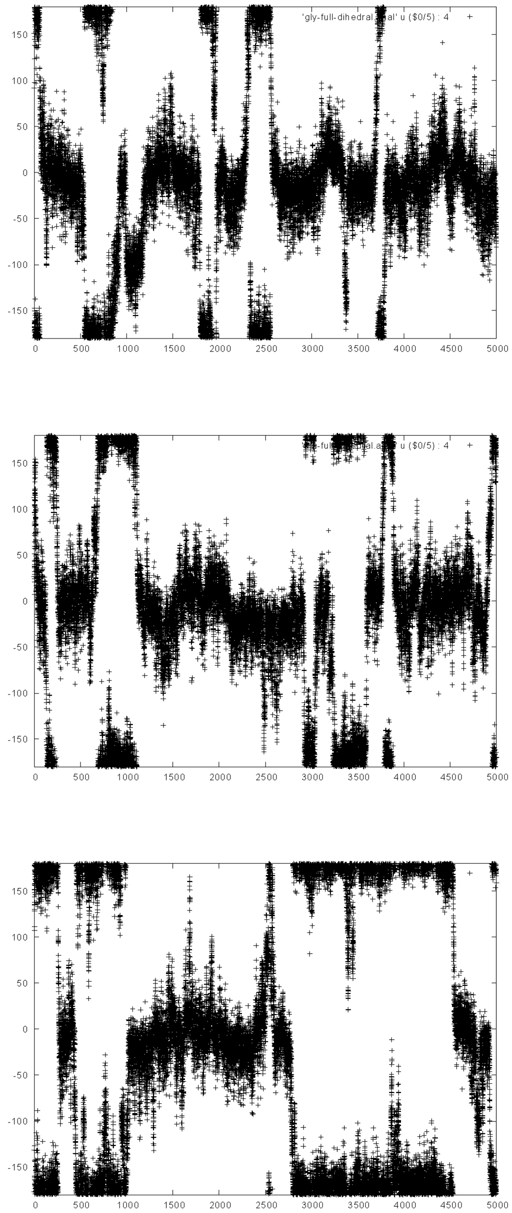
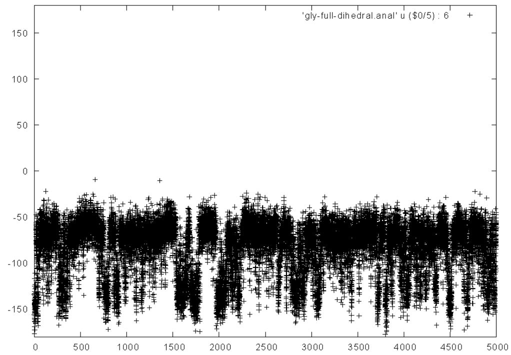
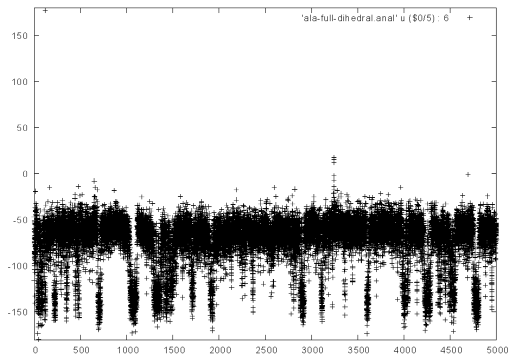


Figure 7.11. Critical dihedrals $\boldsymbol{\psi_1}$.





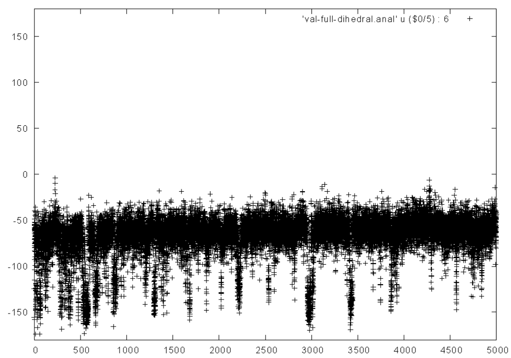


Figure 7.12. Critical dihedrals φ_2.

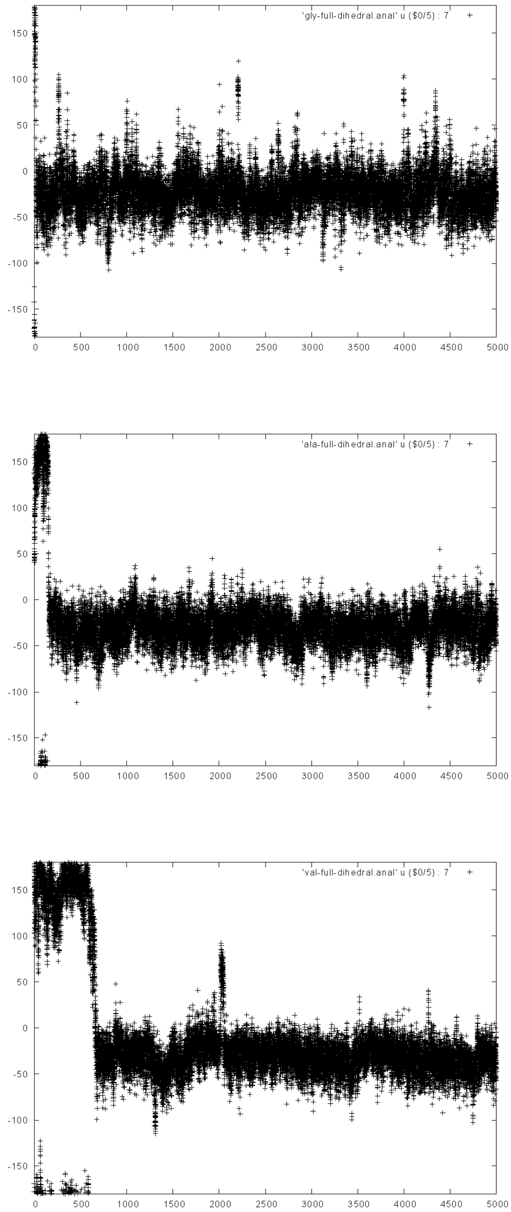


Figure 7.13. Critical dihedrals $\boldsymbol{\psi_2}$.

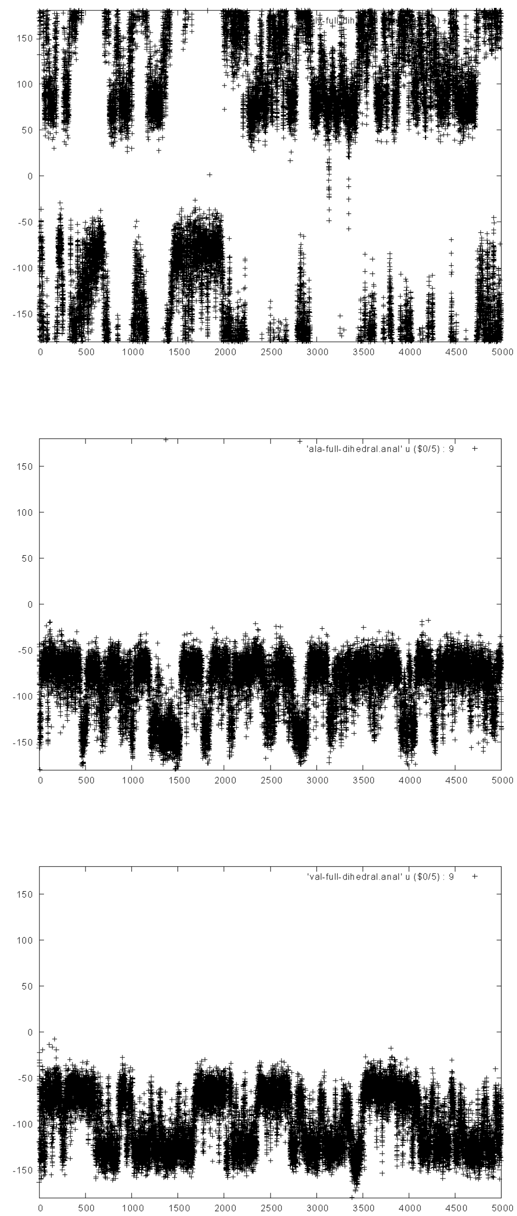


Figure 7.14. Critical dihedrals φ_3.

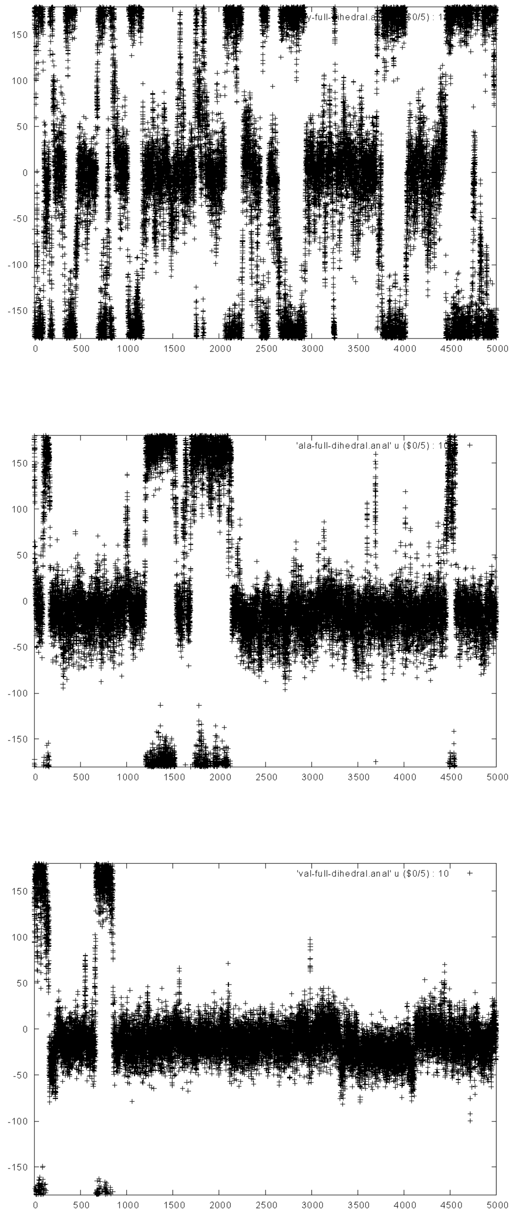


Figure 7.15. Critical dihedrals ψ_3.

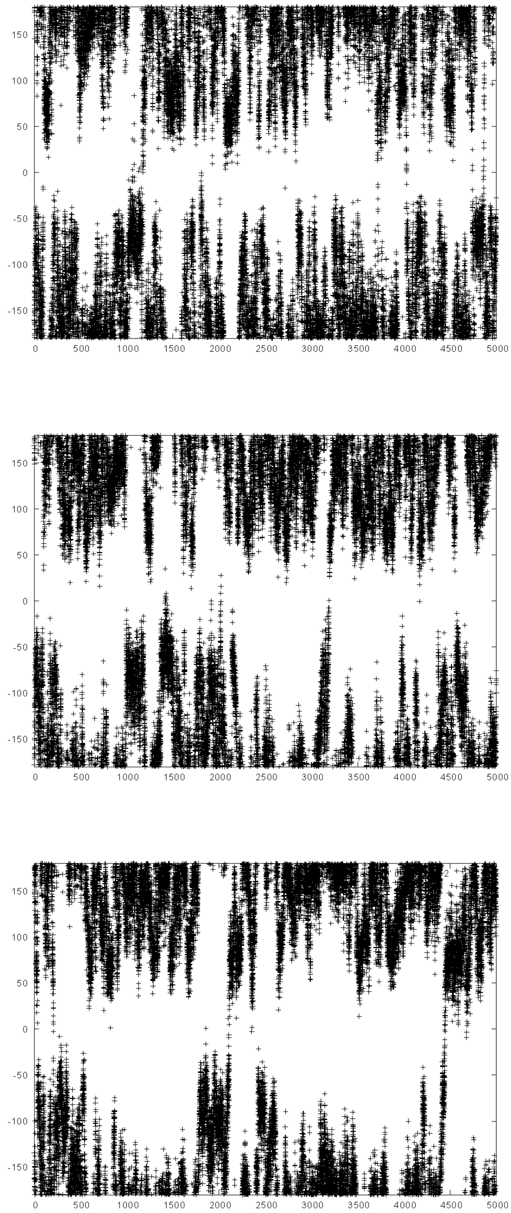


Figure 7.16. Critical dihedrals $\boldsymbol{\varphi_4}$.

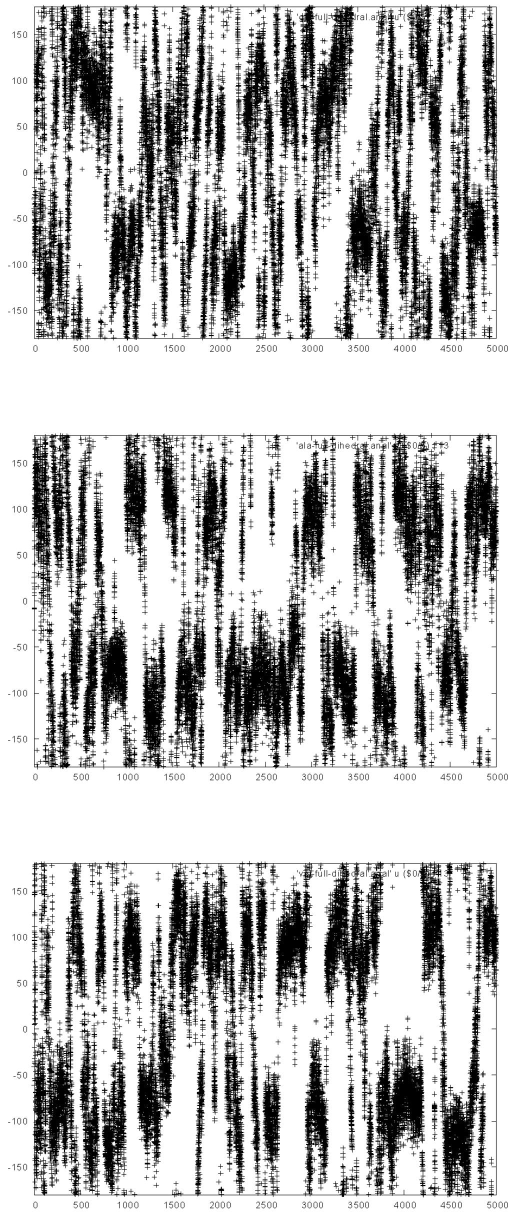


Figure 7.17. Critical dihedrals $\boldsymbol{\psi_4}$.

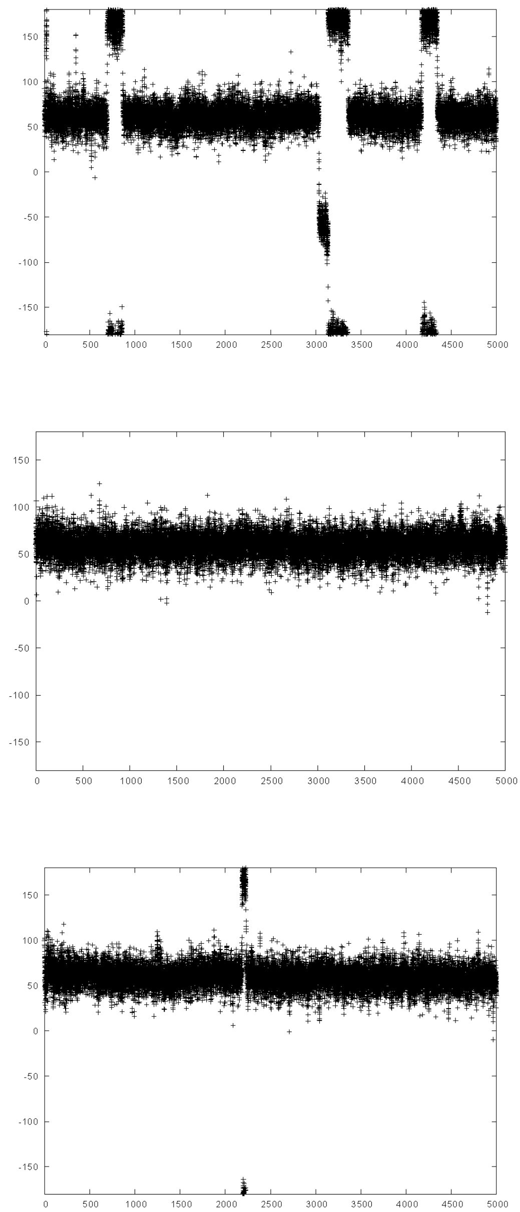


Figure 7.18. Critical dihedrals χ_1.

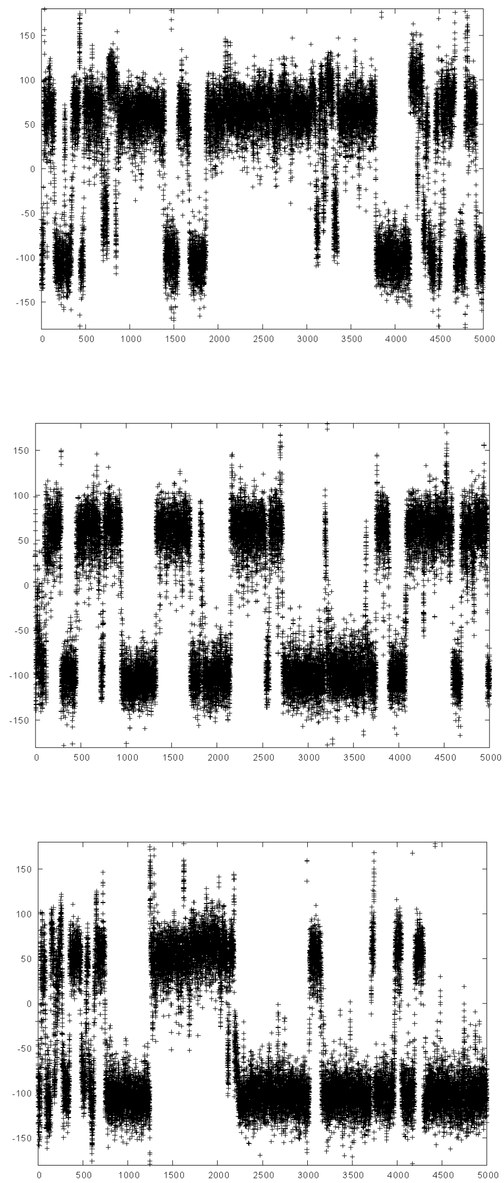


Figure 7.19. Critical dihedrals χ_2.

7.3.3. Alternative Deamidation Pathway - Deamidation via the Amino-side

As mentioned previously, contrary to the results of the pentapeptide experiment, there are several experimental studies where the identity of the *n* -1 residue is influential on deamidation rate; this may be due to a different mechanism that is in effect. This part of the discussion will be based on exploring an alternative deamidation pathway, which involves the amino-side of the Asn residue. The feasibility of this mechanism will be compared, in terms of energetics, with the succinimide-mediated mechanism previously modeled. Previous computational studies on the succinimide mechanism were performed on a smaller model compound (ACE-Asn-NME). For comparative purposes, the alternative mechanism investigated herein was modeled using the same model.

As depicted in Figure 7.20, an alternative pathway, which goes through a cyclic intermediate instead of a succinimide, was modeled. This mechanism involves the attack of the *n* -1 residue's backbone carbonyl to the Asn side chain carbonyl forming an intermediate with a six-membered ring (Int I). Consequently, deamination and hydrolysis of this intermediate leads to the formation of the aspartyl residue (Figure 7.20).

In the first step of this mechanism, the backbone carbonyl of the *n* -1 residue acts as a nucleophile and attacks the carbonyl carbon of the amide on the Asn side chain. Meanwhile, a water molecule assists the transfer of a proton from the NH adjacent to the aforementioned backbone carbonyl to the oxygen of the Asn side chain carbonyl. Proton transfer takes place prior to ring closure, indicating an asynchronous transition state (Figure 7.21); the Asn backbone NH is essentially deprotonated in the transition state, while the side chain carbonyl carbon bears a proton (0.994 Å). In fact, proton transfer enhances the nucleophilic attack by two-fold. First by forming a resonance structure where the backbone carbonyl carries a more negative charge and therefore is a much better nucleophile and secondly, by forming a more positive center on the side chain carbonyl carbon, further enhancing the nucleophilic attack. Meanwhile, ring closure is shown to be underway, as indicated by the O—C distance of 1.951 Å. As a result of this step the cyclic intermediate (Int I) forms. The barrier for ring closure is rather high (41.3 kcal/mol), but is comparable to succinimide formation barriers previously reported.

In the second step, Int I undergoes deamination –with water-assistance– to lose an ammonia molecule and form a ring (Int II) that is more stable due to conjugation. The water molecule helps proton transfer from the ring –OH to the –NH_2 group. The transition state shows that proton transfer is incomplete and ejection of the NH_3 is yet to occur, as seen in the C-N distance, which is still quite short (1.580 Å) The deamination step is straightforward and enables better conjugation throughout the ring atoms (Figure 7.22); hence, relative energies reveal that Int II is approximately 10 kcal/mol more stable than Int I. The activation barrier for the deamination step of Int I (17 kcal/mol) is basically identical to the corresponding step in the succinimide-mediated deamination.

The last step of this mechanism is the hydrolysis of Int II to give a peptide with an Asp residue (Figure 7.23). This particular hydrolysis step requires more than one water molecule. In fact, the proton relay mechanism takes place through three water molecules. A solvent molecule attacks the ring carbonyl carbon (O—C distance 2.575 Å); the O-(C=O) bond on the ring lengthens (1.545 Å); meanwhile, through a proton relay, solvent molecules donate a H^+ to the backbone N (N-H distance 1.032 Å) which had originally lost a proton in the first step (Figure 7.21). In this way, the ring opens and the peptide backbone is restored. The side chain now bears a carboxylic acid group instead of the amide group, deamidation is complete. The energetics of the overall process is comparable with the succinimide mechanism previously studied.

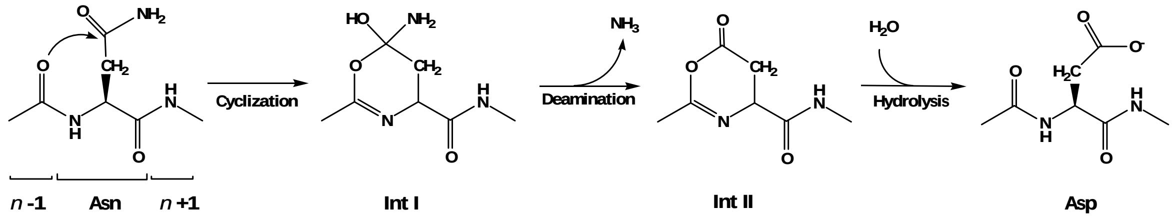


Figure 7.20. Alternative pathway for deamidation of Asn.

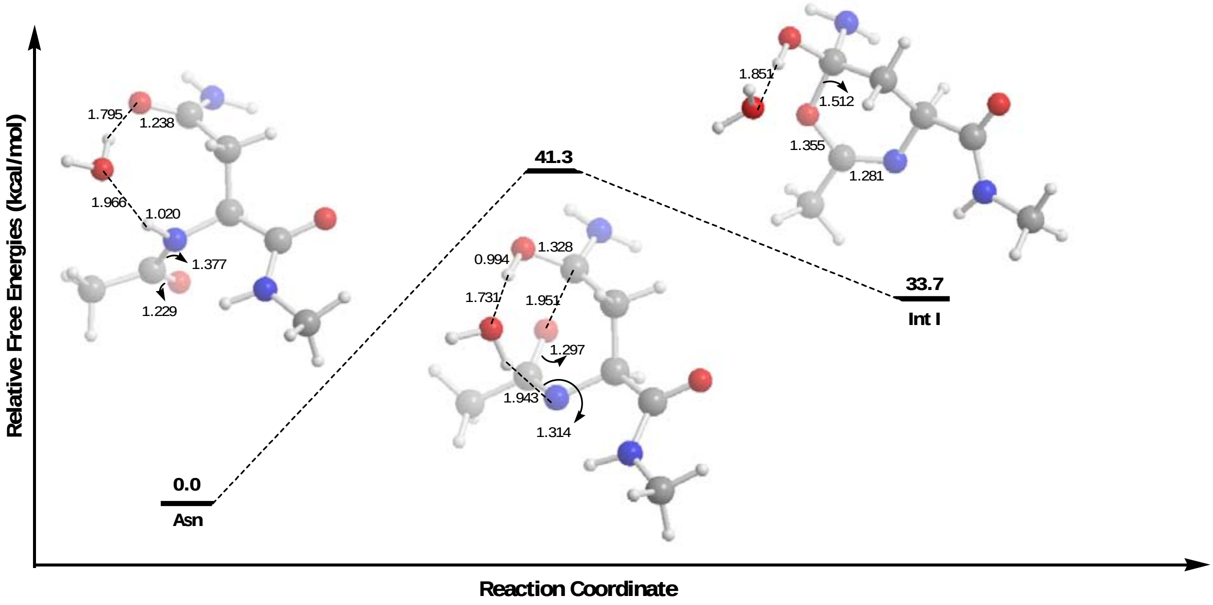


Figure 7.21. Relative free energies (kcal/mol) and optimized geometries for the cyclization step of alternative route, reactant, TS and product, respectively.

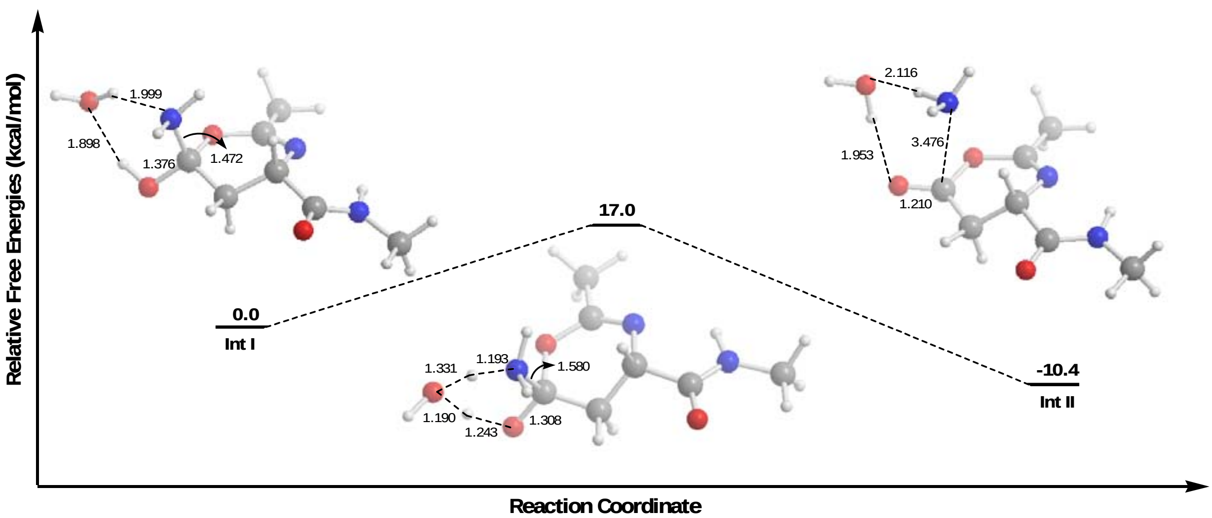


Figure 7.22. Relative free energies (kcal/mol) and optimized geometries for the deamination step of alternative route, reactant, TS and product, respectively.

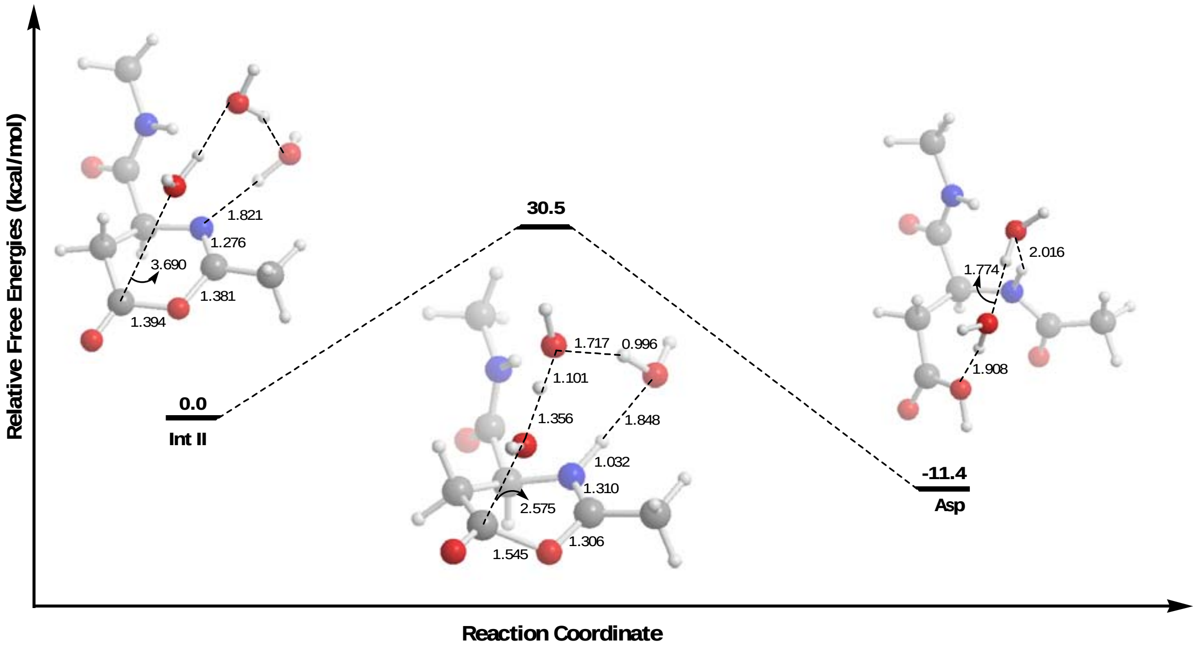


Figure 7.23 Relative free energies (kcal/mol) and optimized geometries for the hydrolysis step of alternative route, reactant, TS and product, respectively.

7.4. Conclusion

The aim of this study was to better understand the factor(s) causing primary sequence dependence in Asn deamidation, in particular, the reasons behind the significantly different Asn deamidation rates in pentapeptides with different *n* +1 residues. Calculated activation barriers for the cyclization step leading to the succinimide formation, which was previously suggested to be the rate-determining step, was consistent with the trend dictated by experimental half-times.

In addition, deprotonation of the *n* +1 NH reproduced similar results, demonstrating that different R groups on the side-chain of the *n* +1 residues influence the ease of deprotonation, indicating that deamidation may be effected as well. On the other hand, MD simulations of pentapeptides with the sequence Gly-Gly-Asn-Yyy-Gly, where Yyy is Gly, Ala and Val, revealed that near-attack-conformations, which are predicted from QM calculations, were not encountered during the simulation. In addition, some of the critical water-bridges and intramolecular H-bonds were non-existent during the simulations due to the absence of the correct orientation of reactive groups. This suggests that either a different mechanism for deamidation may be in effect or major backbone rotation is necessary to achieve the reactive conformation.

8. GENERAL CONCLUSION

The first part of this study has shown that deamidation is catalyzed by water molecules and that several different mechanisms exist for the formation of the succinimide intermediate. Three different mechanisms were suggested for the cyclization step of the deamidation reaction, which was previously proposed to be the rate determining step. All water assisted cyclization reactions investigated were shown to have overall barriers in the range of 34 – 37 kcal/mol approximately 15 kcal/mol lower in energy than the concerted waterless mechanism previously proposed. The most probable mechanism for the formation of the tetrahedral intermediate is proposed to be the tautomerization route. The effect of water molecules on the deamination step of the deamidation reaction has also been established, verifying that the cyclization step, with a substantially higher barrier for activation, is the rate determining step for the succinimide formation. It is also noteworthy that the involvement of water molecules in the deamination step has lowered the barrier to half. It can be suggested that water molecules in the vicinity of asparaginyl residues serve as a catalyst in deamidation reactions. Therefore, one may conclude that deamidation in proteins or enzymes will be more probable for those potential deamidating sites that exhibit the largest accessibility by solvent molecules.

Main conclusions drawn from the second study are: 1) water assistance increases the rate of deamidation, by catalyzing both the formation and hydrolysis of the succinimide intermediate; 2) the tautomerization route has the lowest barrier for the formation of the succinimide intermediate regardless of the number of water molecules that assist the reaction, including the waterless mechanism; 3) hydrolysis of the succinimide intermediate is much more likely to go through a stepwise mechanism, where a gemdiol intermediate is formed. More importantly, the hydrolysis barriers are higher than those for succinimide formation. Previous calculations had shown that the cyclization step is the rate determining step for the formation of the succinimide intermediate, however, when the entire deamidation process is considered, the hydrolysis step is the actual rate determining step. The stepwise hydrolysis barrier is the rate determining step for the overall deamidation process, which is likely to proceed through the tautomerization route. This also explains the isolation of succinimide intermediates during deamidation reactions. The bottleneck of this process is therefore, proposed to be the hydrolysis of the succinimide intermediate. Another rather important finding was the competitiveness between direct hydrolysis and the imide-mediated route at neutral pH. It was shown that direct hydrolysis is as plausible as imide-mediated deamidation of Asn even in the absence of acid or base catalysis.

In the third study, mechanisms leading to non-enzymatic peptide bond cleavage at Asn and Asp have been investigated and it has been suggested that backbone cleavage is unlikely to be competitive with deamidation in peptides with access to solvent molecules. However, the fact that peptide cleavage products at Asn residues are quite often encountered in experimental studies was explained through the ease of backbone cleavage near Asp. Peptide fragmentation barriers are much lower (approximately 10 kcal/mol) for Asp than those for Asn, it is suggested that cleavage at Asn residues may take place after an Asn residue has deamidated into an Asp. In other words, Asn cleavage may be a result of a deamidation reaction followed by a consequent cleavage of the Asp. This proposal may be subject to verification by experiments, since the cleavage products differ for Asn and Asp.

The last study aimed to better understand the reasons behind the significantly different Asn deamidation rates in pentapeptides with different n +1 residues were explored. Calculated activation barriers for the cyclization step leading to the succinimide formation, which was previously suggested to be the rate-determining step, was consistent with the trend dictated by experimental half-times. In addition, deprotonation of the n +1 NH reproduced similar results, demonstrating that different R groups on the side-chain of the n +1 residues influence the ease of deprotonation.

On the other hand, MD simulations of pentapeptides with the sequence Gly-Gly-Asn-Yyy-Gly, where Yyy is Gly, Ala and Val, revealed that near-attack-conformations, which are predicted from QM calculations, were not encountered during the simulation. In addition, some of the critical water-bridges and intramolecular H-bonds were non-existent during the simulations due to the absence of the correct orientation of reactive groups. This suggests that either a different mechanism for deamidation may be in effect or major backbone rotation is necessary to achieve the reactive conformation.

9. FUTURE WORK

As a result of the studies mentioned herein, detailed mechanistic insights of several deamidation routes for Asn have been acquired. In light of the work done on Asn, one of the next goals is to investigate the mechanistic aspects of Gln deamidation, in order to rationalize the slower deamidation rate of Gln. This will require a quantum mechanical approach analyzing differences in reaction barriers as well as a molecular dynamic simulation which will enable a comparison of the dynamic differences between Asn and Gln. In this way, pinpointing the factor(s) that cause Gln deamidation to be substantially slower than Asn deamidation may be possible.

The next challenge in the pursuit of understanding deamidation will be studying a protein which is known to deamidate, preferentially at two or more sites with the same primary structure but different deamidation rates due to secondary structure. This will enable a better understanding of the effect of secondary structure on deamidation. A good candidate for this study is the protein Bcl-x_L. Bcl-xL is an antiapoptotic member of the Bcl-2 family, which inhibits apoptosis initiated by various cellular stresses, and has a pivotal role in the survival of tumor cells. This protein is currently a major topic of interest. It is known to deamidate at two of its Asn residues exclusively. Modeling the deamidation reaction in this system will require the use of molecular dynamics simulations and combined QM/MM potentials.

REFERENCES

1. Damodarm, M., "The isolation of glutamine from an enzymic digest of gliadin", *Biochem. J.*, Vol. 26, pp. 235-247, 1932.

2. Damodarm, M., Jaaback, G. and A. C. Chibnall, "The isolation of glutamine from an enzymic digest of gliadin", *Biochem. J.*, Vol. 26, pp. 1704-1713, 1932.

3. Chibnall, A. C. and R. G. Westall, "Determination of Glutamine in the Presence of Asparagine", *Biochem. J.*, Vol. 26, pp. 122-132, 1932.

4. Greenstein, J. P. and M. Winitz, *Chemistry of the Amino Acids*, Wiley, NY, 1961.

5. Melville, J., "Sequence and structure determinants of the nonenzymatic deamidation of asparagine and glutamine residues in proteins", *Biochem. J.*, Vol. 29, pp. 179-186, 1935.

6. Flatmark, T. and K. Sletten, "Deamidations in Recombinant Human Phenylalanine Hydroxylase", *J. Biol. Chem.*, Vol. 243, pp. 1623-1629, 1968.

7. Bornstein, P. and G. Balian, "The Specific Nonenzymatic Cleavage of Bovine Ribonuclease with Hydroxylamine", *J. Biol. Chem*., Vol. 245, pp. 4854-4856, 1970.

8. Meinwald, Y. C., Stimson, E. R. and H. A. Scheraga, "Deamidation of the asparaginyl-glycyl sequence", *Int. J. Peptide Protein Res*., Vol. 28, pp. 79-84, 1986.

9. Geiger, T. and S. Clarke, "Deamidation, isomerization at asparaginyl and aspartyl residues in peptides", *J. Biol. Chem.*, Vol. 262, pp. 785-794, 1987.

10. Robinson, A. B., "Evolution and the Distribution of Glutaminyl and Asparaginyl Residues in Proteins", *Proc. Natl. Acad. Sci. USA*, Vol. 71, pp. 885-888, 1974.

11. Robinson, A. B., McKerrow, J. H. and P. Cary, "Controlled Deamidation of Peptides and Proteins: An Experimental Hazard and a Possible Biological Timer", *Proc. Natl. Acad. Sci. USA*, Vol. 66, pp. 753-757, 1970.

12. McKerrow, J. H. and A. B. Robinson, "Primary Sequence Dependence of the Deamidation of Rabbit Muscle Aldolase", *Science*, Vol. 183, pp. 85, 1974.

13. Takamoto, L. and D. Boyle, "The X-ray Crystal Structure of Human γS-crystallin C-terminal Domain", *J. Biol. Chem.*, Vol. 275, pp. 26109-26112, 2001.

14. Watanabe, A., Takio, K. and Y. Ihara, "Deamidation and Isoaspartate Formation in Smeared Tau in Paired Helical Filaments", *J. Biol. Chem.*, Vol. 274, pp. 7368-7378, 1999.

15. Robinson, A. B., Scotchler, J. W. and J. H. McKerrow, "Rates of nonenzymic deamidation of glutaminyl and asparaginyl residues in pentapeptides", *J. Am. Chem. Soc.*, Vol. 95, pp. 8156-8159, 1973.

16. Robinson, A. B. and L. R. Robinson, "Distribution of Glutamine and Asparagine Residues and Their Near Neighbors in Peptides and Proteins", *Proc. Natl. Acad. Sci. USA*, Vol. 88, pp. 8880-8884, 1991.

17. Yüksel, K. U. and R. W. Gracy, "Prediction of protein deamidation rates from primary and three-dimensional structure", *Arch. Biochem. Biophys.*, Vol. 248, pp. 452-459, 1986.

18. Solstad, T. and Flatmark, T., "Microheterogeneity of recombinant human phenylalanine hydroxylase as a result of nonenzymatic deamidations of labile amide containing amino acids", *Eur. J. Biochem.*, Vol. 267, pp. 6302-6310, 2000.

19. Capasso, S. and S. Salvadori, “Mechanism for Intein C-terminal Cleavage: A Proposal from Quantum Mechanical Calculations”, *J. Peptide Res.*, Vol. 54, pp. 377-382, 1999.

20. Clark, T., *A Handbook of Computational Chemistry*, John Willey&Sons, Canada, 1985.

21. Bingham, R.C., Dewar, M. J. S. and D. H. Lo., “An Improved Version of the MINDO Semiempirical SCF-MO Method”, *J. Am. Chem. Soc.*, Vol. 97, pp. 1285-1293, 1975.

22. Dewar, M. J. S. and W. Thiel, “The MNDO Method. Approximations and Parameters”, *J. Am. Chem. Soc.*, Vol. 99, pp. 4899-4907, 1977.

23. Dewar, M. J. S., Zoebisch, E. G. Healy, E. F. and J. J. P. Stewart, “AM1: A New General Purpose Quantum Mechanical Molecular Model”, *J. Am. Chem. Soc.*, Vol. 107, pp. 3902-3909, 1985.

24. Stewart, J. J. P., “Optimization of parameters for semi-empirical methods I-Method.”, *J. Comp. Chem.*, Vol. 10, pp. 209-220, 1989.

25. Pilar, F. L., *Elementary Quantum Chemistry,* Mc-Graw-Hil1, New York, 1990.

26. Hirst, D. M. A., *Computational Approach to Chemistry,* Blackwell Scientific Publications, Oxford, 1990.

27. Pople, J. A. and R. Seeger, “Electron Density in Møller-Plesset Theory”, J. Chem. Phys., 62, 4566, 1975.

28. Pople, J. A. and D. L. Beveridge, *Approximate Molecular Orbital Theory,* Mc-Graw-Hill, New York, 1970.

29. Stewart, J. J. P., "Optimization of Parameters for Semiempirical Methods Applications", *J. Comp. Chem.*, Vol. 10, pp. 221-264, 1990.

30. Levine, I. N., *Quantum Chemistry,* Prentice-Hall, New Jersey, 1991.

31. Hehre, W. J., L. Radom, P. R. Schleyer and J. A. Pople, *Ab Initio Molecular Orbital Theory,* John Willey and Sons, New York, 1986.

32. Lowe, J. P., *Quantum Chemistry,* 2nd Ed., Academic Press, 1993.

33. Wang, L. C. and R. J. Boyd, "The effect of electron correlation on the electron density distributions of molecules: Comparison of perturbation and configuration interaction methods", *J. Chem. Phys.*, Vol. 90, pp. 1083, 1983.

34. Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons Ltd., USA, 2004.

35. Handy, N. C., "Density Functional Theory", in: B. O. Roos (ed.), *Lecture Notes in Quantum Chemistry*, Vol. 2, pp. 91-123, Springer-Verlag, Berlin, 1994.

36. Becke, A. D., "Density-Functional Exchange Energy Approximation with Correct Asymptotic Behavior", *Phys. Rev. A*, Vol. 38, pp. 3098-3103, 1988.

37. Becke, A. D., "A New Mixing of Hartree-Fock and Local Density Functional Theories", *J. Chem. Phys.*, Vol. 38, pp. 1372-1377, 1993.

38. Lee, C., Yang, W. and R. G. Parr, "Development of the Colle-Salvetti conelation energy formula into a functional of the electron density", *Phys. Rev. B*, Vol. 37, pp. 785-789, 1988.

39. Becke, A. D., "Density Functional Thermochemistry. III. The Role of Exact Exchange", *J. Chem. Phys.*, Vol. 98, pp. 5648-5652, 1993.

40. Madura, J. and W. L. Jorgensen, "Ab initio and Monte Carlo calculations for a nucleophilic addition reactionin the gas phase and in aqueous solution", *J. Am. Chem. Soc.*, Vol. 108, pp. 2517-2527, 1986.

41. Tomasi, J., Mennucci, B. and E.J. Cancès, "The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM ab initio level ", *Mol. Struct. (THEOCHEM)*, Vol. 464, pp. 211-226, 1999.

42. Tomasi, J., Benedetta, M. and R. Cammi, "Quantum Mechanical Continuum Solvation Models", *Chem. Rev.*, Vol. 105, pp. 2999-3093, 2005.

43. Cancès, E. T., Mennucci, B. and J. Tomasi, "Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications", *J. Chem. Phys.*, Vol. 107, pp. 3032-3041, 1997.

44. Mennucci, B., Cancès, E. and J. Tomasi, "evaluation of solvent effects in isotropic and anisotropic dielectrics andin ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications", *J. Phys. Chem. B*, Vol. 101, pp. 10506-10517, 1997.

45. Mennucci, B. and J. Tomasi, "Continuum Solvation Models: A New Approach To The Problem Of Solute's Charge Distribution And Cavity Boundaries", *J. Chem. Phys.*, Vol. 106, pp. 5151-5158, 1997.

46. Leach, A. R., Molecular Modelling Principles and Applications, Prentice Hall, England, 2001.

47. Robinson, N. E. and A. B. Robinson, "Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects", *J. Peptide Res.*, Vol. 63, pp. 437-448, 2004.

48. Kim, E., Lowenson, J. D., MacLaren, D. C., Clarke, S. and S. G. Young, "Deficiency of a protein-repair enzyme results in the accumulation of altered proteins, retardation of growth, and fatal seizures in mice", *Proc. Natl. Acad. Sci. USA,* Vol. 94, pp. 6132-6137, 1997.

49. Solstad, T., Carvalho, R. N., Anderson, O. A., Waidelich, D. and T. Flatmark, *Eur. J. Biochem.*, Vol. 270, pp. 929-938, 2003.

50. Robinson, N. E. and A. B. Robinson, "Prediction of protein deamidation rates from primary and three-dimensional structure", *Proc. Natl. Acad. Sci. USA*, Vol. 98, pp. 4367-4372, 2001.

51. Weintraub, S. J. and S. R. Manson, "Asparagine deamidation: A regulatory hourglass", *Mech. Age. Dev.*, Vol. 125, pp. 255-257, 2004.

52. Robinson, N. E. and A. B. Robinson, "Deamidation of human proteins", *Proc. Natl. Acad. Sci. USA*, Vol. 98, pp. 12409-12413, 2001.

53. Capasso, S., Mazzarella, L., Sica, F., Zagari, A. and S. Salvadori, "Spontaneous cyclization of the aspartic acid side chain to the succinimide derivative", *J. Chem. Soc. Chem. Comm.*, pp. 919-921, 1992.

54. Capasso, S., Mazzarella, L., Sica, F., Zagari, A. and S. Salvadori, "Effect of the three - dimensional structure on the deamidation reaction of ribonuclease A", *J. Chem. Soc. Perkin Trans. 2,* pp. 679-682, 1993.

55. Capasso, S., Mazzarella, L., Sica, F. and A. Zagon, "Effect of the three - dimensional structure on the deamidation reaction of ribonuclease A", *J. Peptide Res.*, Vol. 2, pp. 195-197, 1989.

56. Konuklar, F. A., Aviyente, V., Sen, T. Z. and I. Bahar, "Modeling the Deamidation Reaction of Asparagine Residues via Succinimide Intermediates", *J. Mol. Model.,* Vol. 7, pp. 147-160, 2001.

57. Konuklar, F.A., Aviyente, V. and M. F. Ruiz-Lopez, “A Theoretical Study on Modeling the Neutral and Alkaline Hydrolysis of Isoaspartate”, *J. Phys. Chem. A*., Vol. 106, pp. 11205-11214, 2002.

58. Konuklar, F. A. and V. Aviyente, “Theoretical Analysis of Hydrolysis Reaction in Succinimide: Formation of Aspartate”, *Org .Biomol.Chem*., Vol. 1, pp. 2290-2297, 2003.

59. Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, M. R., Robinson, N. E. and A. B. Robinson, “Use of Merrifield solid phase peptide synthesis in investigations of biological deamidation of peptides and proteins”, *J. Peptide Res.*, Vol. 63, pp. 437-448, 2004.

60. Gonzalez, C. and H. B. Schlegel, “An improved algorithm for reaction path following”, *J. Chem. Phys.*, Vol. 90, pp. 2154-2161, 1989.

61. Gonzalez, C. and H. B. Schlegel, “An improved algorithm for reaction path following”, *J. Phys. Chem.*, Vol. 94, pp. 5523-5527, 1990.

62. Bondi, A., “van der Waals Volumes and Radii Bondi”, *J. Phys. Chem.*, Vol. 68, pp. 441-451, 1964.

63. Gaussian 03, Revision B.05, Frisch, M. J., Gaussian, Inc., Wallingford CT, 2004.

64. Hazra, M. K. and T. Chakraborty, “Formamide Tautomerization: Catalytic Role of Formic Acid”, *J. Phys. Chem. A*., Vol. 109, pp. 7621-7625, 2005.

65. Liang, W., Li, H., Hu, X. and S. Han, “Proton Transfer of Formamide + nH_2O (n = 0-3): Protective and Assistant Effect of the Water Molecule”, *J. Phys. Chem. A*., Vol. 108, pp. 10219-10224, 2004.

66. Robinson, N. E. and A. B. Robinson, “Molecular Clocks”, *Proc. Natl. Acad. Sci. USA*, Vol. 98, pp. 944-949, 2001.

67. Tam, J. P., Riemen, M. W. and R. B. Merrifield, “Accelerated chemical synthesis of peptides and small proteins”, *Peptide Res.*, Vol. 1, pp. 6-18, 1988.

68. Robinson, N. E., Robinson, A. B., Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins, Althouse Press, Cave Junction, OR, 2004.

69. Radkiewicz, J. L., Zipse, H., Clarke, S. and K. N. Houk, “Accelerated racemization of aspartic acid and asparagine residues via succinimide intermediates: An ab initio theoretical exploration of mechanism”, *J. Am. Chem. Soc.*, Vol. 118, pp. 9148-9155, 1996.

70. Radkiewicz, J. L., Zipse, H., Clarke, S. and K. N. Houk, “Neighboring Side Chain Effects on Asparaginyl and Aspartyl Degradation: An Ab Initio Study of the Relationship Between Peptide Conformation and Backbone NH Acidity”, *J. Am. Chem. Soc.*, Vol. 123, pp. 3499-3506, 2001.

71. Catak, S., Monard, G., Aviyente, V. and M. F. Ruiz-López, “Reaction Mechanism of Deamidation of Asparaginyl Residues in Peptides: Effect of Solvent Molecules”, *J. Phys. Chem. A*, Vol. 110, pp. 8354-8365, 2006.

72. Gorb, L., Asensio, A., Tunon, I., M. F. Ruiz-Lopez, “The mechanism of formamide hydrolysis in water from ab initio calculations and simulations”, *Chem. Eur. J.*, Vol. 11, pp. 6743-6753, 2005.

73. Peters, B. and B. L. Trout, “Asparagine Deamidation: pH-Dependent Mechanism from Density Functional Theory”, *Biochemistry*, Vol. 45, pp. 5384-5392, 2006.

74. Capasso, S., Mazzarella, L., Sorrentino, G., Balboni, G. and A. J. Kirby, “Kinetics and Mechanism of the Cleavage of the Peptide Bond Next to Asparagine”, *Peptides*, Vol. 17, pp. 1075-1077, 1996.

75. Landon, M., *Methods in Enzymology*, Vol. 47, pp. 145-147, 1977.

76. Tarelli, E. and P. H. Corran, “Ammonia cleaves polypeptides at asparagine proline. bonds”, *J. Peptide Res.*, Vol. 62, pp. 245-251, 2003.

77. Joshi, A. B., Rus, E. and L. E. Kirsch, “The degradation pathways of glucagon in acidic solutions”, *Int. J. Pharm.*, Vol. 203, pp. 115–125, 2000.

78. Joshi, A. B. and L. E. Kirsch, “The relative rates of deamidation at asparagine and glutamine”, *Int. J. Pharm.*, Vol. 273, pp. 213–219, 2004.