INVESTIGATIONS OF PEPTIDE AND PROTEIN DEAMIDATION

Thesis by

Noah E. Robinson

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2004

(Defended December 15, 2003)

DEDICATION

I dedicate this thesis to the memory of my mother,

Laurelee R. Robinson.

ACKNOWLEDGEMENTS

Of the many people who have helped in various ways with this work, I wish to especially thank Drs. R. Bruce Merrifield, Harry B. Gray, and Arthur B. Robinson.

Professor Merrifield was very generous in allowing us the use of his peptide synthesis facilities and providing us with essential advice. Both Professor and Mrs. Merrifield have been and continue to be extremely helpful in providing us with encouragement, advice, and assistance.

Professor Gray has been exceedingly kind and supportive in allowing me to pursue this course of research and making possible the essential arrangements required.

Most of the work described here was done in collaboration with my father, Arthur B. Robinson.

Particular thanks are also due my brothers and sisters, Zachary, Arynne, Bethany, Joshua, and Matthew Robinson who participated in some of the laboratory work and provided personal and logistical support which included assisting with the building and equipping of the laboratory. Matthew's dog Rusty's insatiable appetite for sticks, stones, and anything else he can fit in his mouth provided general amusement.

I would also like to thank Professor Brian Chait for his advice and counsel concerning mass spectrometry and Professors Frederick Seitz and R. Bruce Merrifield for communicating four papers to PNAS.

Additionally I would like to thank Professor Martin D. Kamen who was my father's thesis advisor and a family friend for more than 30 years. I will never forget the many enjoyable hours spent with him and listening to him play his viola.

During much of this work the laboratory environment was substantially improved by the wonderful music of George Beverly Shea. I wish to particularly thank him for this and for the ministry which he continues to pursue at the age of 94 with Dr. Billy Graham. I would also like to thank Jeff Cooper whose advice concerning personal awareness was very useful at 2 A.M. one night on the streets of New York City during one of our trips there to make peptides. His further philosophy along these lines should certainly have been followed on that occasion.

I also thank the John Kinsman Foundation, the A. Reynolds Morse Foundation and other donors to the Oregon Institute of Science and Medicine as well as a Caltech Special Institute Fellowship for financial support.

ABSTRACT

This work concerns the peptide and protein deamidation of asparaginyl, Asn, and glutaminyl, Gln, residues to aspartyl and glutamyl residues, respectively. This is accompanied by a change in charge and increase of one unit in mass. Isomerization also often takes place. This reaction–which occurs spontaneously and non-enzymatically in many proteins under physiological conditions–has been hypothesized to be a general molecular clock, which could be used as a timer for protein turnover, development, or aging, and for the controlled conversion of one protein into another.[1–9] Since this hypothesis was first proposed, amide molecular clocks have been demonstrated in several biological systems. These include the timing of protein turnover in rat cytochrome c[10, 11] and aldolase;[12–14] the counting of individual enzyme molecule catalytic cycles in triosephosphate isomerase;[15–17] and the resetable time-dependent monitoring of DNA repair in Bcl-X_L.[18–21]

The objective of this work is to increase the understanding of peptide and protein deamidation, especially the prediction of deamidation rates for specific amides in proteins. This capability has been developed for Asn residues in proteins of known three-dimensional structure, and the sequence information required to do this for Gln residues has also been obtained. This work provides stability information to protein chemists and needed information to those who continue to investigate the molecular clock hypothesis.

The problem was attacked in two phases. The first consisted of experimental work to determine the rate of deamidation as a function of primary sequence near the amide. All 400 possible peptides of the type GlyXxxAsnYyyGly were synthesized and the deamidation rates of 323 of these were measured. An additional 67 asparaginyl containing peptides that had a variety of lengths and forms were also measured in order to assess the efficacy of pentapeptides in this work. Also, a similar set of 400 peptides were synthesized with Gln in place of Asn. The deamidation rates of 52 of the Gln peptides have been measured and, from these rates, the rates of the other Gln peptides have been inferred. With the addition of rates for 34 peptides with blocked Cys, a total of 476 individual rate

experiments on peptides have been carried out in 0.15 M Tris buffer at pH 7.4, 37.00 °C. Additional rate experiments were conducted on solution effects and on properties of the experimental system.

The second phase involved using experimental data from the literature–on relative instabilities of specific amides in proteins under specified solvent conditions and on protein three-dimensional structures–to quantify the effect that secondary, tertiary, and quaternary protein structure has on primary structure deamidation rates.

Factors that determine deamidation rates include structural components as well as electronic effects. It happens that the primary sequence component is so large that it is possible to separate it from the rest of the problem and measure it alone. The secondary, tertiary, and quaternary effects were then considered. It was found that, on average, 60% of the rate is accounted for by primary sequence, while the remaining 40% is ascribable to secondary, tertiary, and quaternary structure in those proteins for which suitable deamidation observations are available.[7] For all proteins, the estimated values are 50% and 50%. For individual amides, these percentages vary over a wide range.

A procedure was developed through which the quantitative deamidation rate of any amide in a protein for which the three-dimensional structure is known can be calculated. These estimated rates are accurate to within about a factor of two or better for most amides with half-times less than 100 days. They can be used to predict the fastest amides in a protein to greater than 95% reliability. This reliability is remarkable, considering the fact that deamidation half-times vary from less than 1 day to over 100 years in 37.00 °C, pH 7.4, 0.15 M Tris buffer. The procedure was calibrated with only relative amide instabilities. It was then evaluated with absolute deamidation rates.

Details of this work are given in the pages which follow. The result was a computerized procedure through which deamidation rates for every Asn of every protein in the PDB Protein Database were calculated and analyzed.

This work provides two discoveries of biological importance.

First, the assumption of genetic determination of deamidation rates, which has until now been dependent upon scattered and largely qualitative observations, has been placed upon a firm quantitative foundation with a thorough understanding of the effects of primary, secondary, and tertiary structure.

Second, this quantitative understanding has permitted reliable estimation of the distribution functions of deamidation rates in thousands of proteins. These rates show that a substantial fraction of proteins are genetically programmed to deamidate in biologically relevant timed intervals even though most of the genetically available structures deamidate more slowly. This quantitative preference for fast deamidation times relative to slow deamidation times is an entirely new discovery. The fact that this property is present within the thousands of proteins of currently known three-dimensional structure markedly strengthens the molecular clock hypothesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Number *Page*

NOMENCLATURE

Asn: Abbreviation for "asparaginyl residue" and indicates a residue in a peptide or protein. Instances of the amino acid asparagine by itself are referred to herein as either free asparagine or free Asn.

Deamidation: The conversion of asparaginyl and glutaminyl residues into aspartyl and glutamyl residues respectively, with the loss of ammonia. This often occurs with isomerization.

Gln: Abbreviation for "glutaminyl residue" and indicates a residue in a peptide or protein. Instances of the amino acid glutamine by itself are referred to herein as either "free glutamine" or "free Gln."

Peptide: A short chain made by linking amino acid residues through peptide bonds. Most of the peptides used in this study were 5 residues, although some longer ones of up to 13 residues were employed.

Protein: A long peptide usually at least 100 residues in length, sometimes also including other chemical groups. A protein typically folds into a special three-dimensional structure by itself or in concert with other macromolecules.

Chapter 1

OVERVIEW OF DEAMIDATION

Introduction

The posttranslational modification of proteins is a widespread phenomenon. Besides enzymatic processes, there are many non-enzymatic processes by which proteins undergo specific changes in charge and conformation after they are physiologically synthesized and folded. Among these changes, the non-enzymatic deamidation of asparaginyl and glutaminyl residues to form negatively charged aspartyl and glutamyl residues is unusual. Each amide residue has a specific rate of deamidation that is genetically determined through the sequence of residues immediately adjacent in the peptide chain and the three-dimensional structure of the protein. The half-times of these reactions under physiological conditions vary from less than one day to more than a century. Some of the other common post-synthetic protein modifications are glycosylation, racemization, oxidation, crosslinking, and chain cleavage.[22–24] Among these reactions, however, deamidation is unique.

Deamidation of Asn or Gln leads directly to Asp, Glu, or their isomers. It does so at a genetically determined rate for each amide that is controlled by the sequence and structure of the protein. Moreover, in peptides this rate varies from less than a day to more than 20,000 days depending on protein sequence. When three-dimensional structure is included, the range of deamidation rates extends from less than 0.5 days to more than a century with almost any rate available depending on the choice of primary, secondary, and tertiary structure. This range includes the *in vivo* lifetimes of a large percentage of proteins.

Deamidation of L-Asn produces L-Asp, L-isoAsp, D-Asp, and D-isoAsp. Gln probably forms similar isomers, so changes in protein structure through isomerization are possible. Always occurring at neutral pH, however, is an increase in negative charge at the

deamidation site. This change in charge can have dramatic consequences for the protein. Not only can this affect its folded structure, but it can also change its binding properties, recognition by enzymes, or almost any other property depending upon where the deamidation occurs and how critical this region is to the property. At the very least, every protein chemist should be aware of possible rapid deamidation sites in the proteins he is studying. A preparation that is undergoing unknown modifications with time can very easily lead to faulty results.

Moreover, *in vivo* experiments without deamidation estimation are especially hazardous. The structural changes that accompany deamidation often lead to substantially increased *in vivo* degradation rates. Therefore, failure to detect deamidated products does not necessarily mean that deamidation is not occurring at a significant rate.

Our primary interest in deamidation, however, relates to its possible use as a biomolecular timer as was initially proposed over 30 years ago.[1–4] The known properties of deamidation make it ideally suited for this purpose.

Why should Asn and Gln be present in proteins? They have no special charges or other structures which make them unusual. It should have been possible to build proteins without them–perhaps with two other amino acids which are more stable in physiological conditions. Consider cysteine. Cysteine is easily oxidized and generally not the most desirable of amino acids, but it has a purpose. In providing the unique function of forming crosslinks, cysteine justifies its existence regardless of its drawbacks. Now consider the amides, especially Asn. What special function does Asn perform to atone for the fact that it is often very unstable and introduces post-synthetic negative charges into proteins? Perhaps the answer is that this instability *is* the function of Asn. It is this reasoning that prompted the initial investigations into the sequence dependence of deamidation.[1–4]

Now that much more is known, these arguments are more compelling. The deamidation rates for Gln peptides turn out to be, on average, about 100 times as long as those of Asn peptides.[3, 25, 26] If an amide side chain is needed, it is available in a more stable form from

Gln. Moreover, the results of the prediction work described here show that most Asn deamidations proceed relatively slowly as well, with fast deamidations limited to about 5% of protein Asn structures. These fast amides occur, however, more frequently in proteins than would be expected by chance[8, 9] even though more stabile structures could easily be genetically specified. It has also been shown that Asn itself appears more frequently in proteins than is expected by chance.[5]

As has been discovered during this work, the deamidation rates of Gln and Asn are mutually complimentary, with the two blending with one another to cover the entire range of biological interest. It may be that this complementarity is, in part, responsible for the unusual property that both of these two amino acid residues are provided in proteins, regardless of the fact that they largely duplicate one another. The differences in deamidation rates of the Asn succinimide mechanism vs. the Gln glutarimide mechanism account for this feature.

An examination of the properties that make deamidation suitable as a timer is illustrative. It was previously known, as a result of early deamidation studies, that a wide range in peptide deamidation rates is available.[1-4, 25, 27, 11, 4, 28] The peptide work reported here shows this range to be even wider and richer than before. What better way to set a timer in a protein than to choose the appropriate amide structure? Amide clocks are simple and elegant and depend to a large extent on the sequence of just two residues.

This work now shows that the deamidation rate can be fine tuned by protein structure to virtually any rate with a half-time greater than a few hours. If amides are timers, how does the organism know when the alarm goes off? The introduction of a negative charge answers this question, with isomerization perhaps providing additional versatility. There is, however, evidence that living systems are better off without the isomerization. Enzymes that reverse Asn isomerization without correcting the change in charge have been shown to be ubiquitous and also necessary for good health.[29, 30] Interestingly, no enzymes have yet been discovered which reverse the charge change.

In summary, almost all proteins contain several amide residue clocks. These clocks are genetically adjustable to any timed interval of more than a few hours, and, as they run, they introduce substantial changes into protein structures. Unless these clocks are useful, they are surely disruptive to biological systems and would not have been used.

There is much interest in the "N-Rule" as a molecular timer of protein turnover through ubiquitination. Deamidation does not necessarily compete with this system. The N-Rule is primarily associated with much shorter time intervals.[31] No amides have yet been found to have half-times shorter than 12 hours under physiological conditions. Also, deamidation can be incorporated into the N-rule pathway. In the N-rule system, Gln and Asn are thought to be involved as tertiary destabilizing residues. They are first converted to Glu and Asp–which could be done non-enzymatically over predetermined time intervals–and are then arginated before being recognized by ubiquitin.[31]

The work that has been done on *in vitro* and *in vivo* deamidation of proteins is likewise interesting. Nonenzymatic deamidation of Asn or Gln *in vivo* or *in vitro* has been reported for more than 200 types of naturally occurring peptides and proteins. Most of these molecules deamidate primarily at asparaginyl residues, including those of special current interest such as ribonuclease[32, 33] and phenylalanine hydroxylase.[34, 35] Deamidation rates of glutaminyl residues are, on average, much slower than those of asparaginyl residues. Consequently, nonenzymatic deamidation of Gln has been observed primarily in very long-lived proteins such as the crystallins of the eye lens.[36–47] We expect that the examples found so far are only a beginning. It is likely that timing of events in biological processes will be found to take place by means of elegant control systems with fundamental timers built into various components of living organisms. The extent to which this will involve amide clocks is as yet unknown.

At this moment amides all over your body are deamidating–at genetically-determined rates. This includes your eye lenses, your enzymes, your structural proteins, your nucleoproteins, and many other molecules. Most of your proteins and many of your peptides are susceptible to this reaction. These molecules are gradually becoming more

negatively charged. This isn't just happening to you, it is happening to every living thing. Could deamidation be timing important processes in your body that are essential to life? Perhaps it is causing you to age. We know that at least some of it is occurring in proteins that will never be replaced. The results to date are both encouraging and alarming. As chemists we have a duty to perform. This deamidation must not go on unchecked. It must either be understood, or it must be stopped!

Previous Studies

In 1806 the first amino acid was isolated from natural sources.[48] This amino acid was asparagine and was named after the asparagus in which it was found[49, 50] in 1826. The first deamidation of asparagine–by hydrolysis–was described[51] in 1832. Optical activity of asparagine was shown[52] in 1851. Glutamine was deduced from the presence of excess ammonia in hydrolosates[53] in 1873, isolated[54] in 1883, and isolated from proteins[55] in 1932. The discovery of the conversion of free glutamine into pyrrolidone-carboxylic acid occurred[56, 57] in 1914. This reaction was found to be catalyzed by phosphate, arsenate, and other basic ions.[58]

The early history of glutamine and asparagine was reviewed[59] in 1961. The structures of asparagine and glutamine are shown as stick figures and space filling CPK models in Figures 1-1 and 1-2.

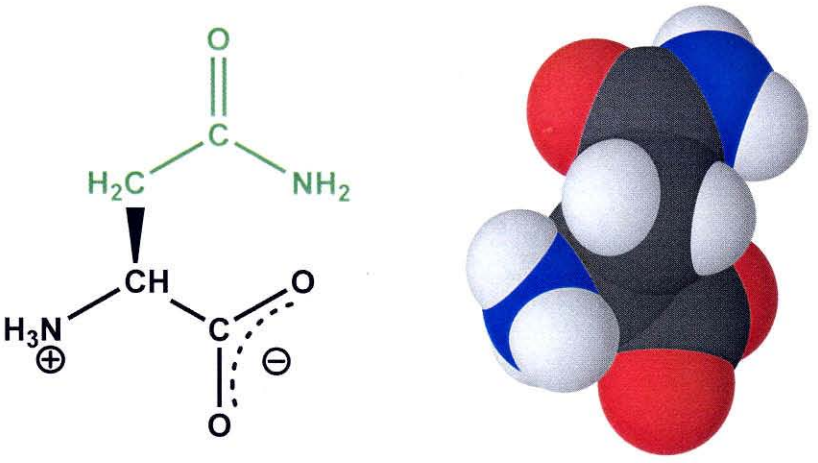


Figure 1-1 – Free Asparagine at Neutral pH

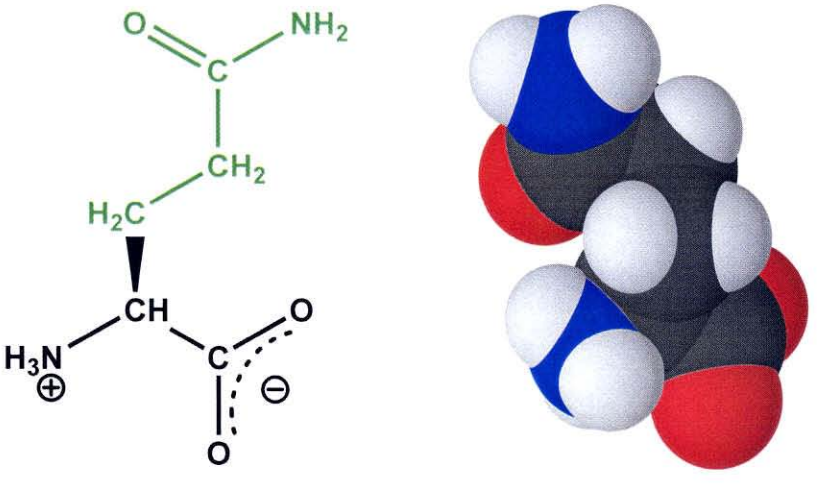
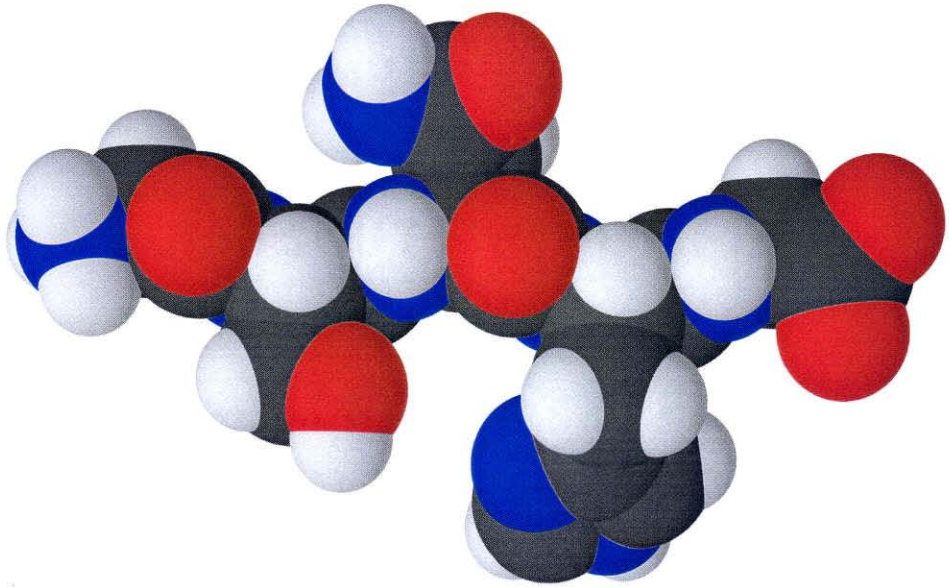


Figure 1-2 – Free Glutamine at Neutral pH

When assembled into peptides, the structures are like the one shown in Figure 1-3 with planar peptide bonds as indicated.[60] This peptide is GlySerAsnHisGly. The conformation of this peptide has been chosen so that the side chains can be clearly seen. It is evident, however, that nearest neighbor side chains can easily interact with Asn. This property is important to the sequence dependence of deamidation.



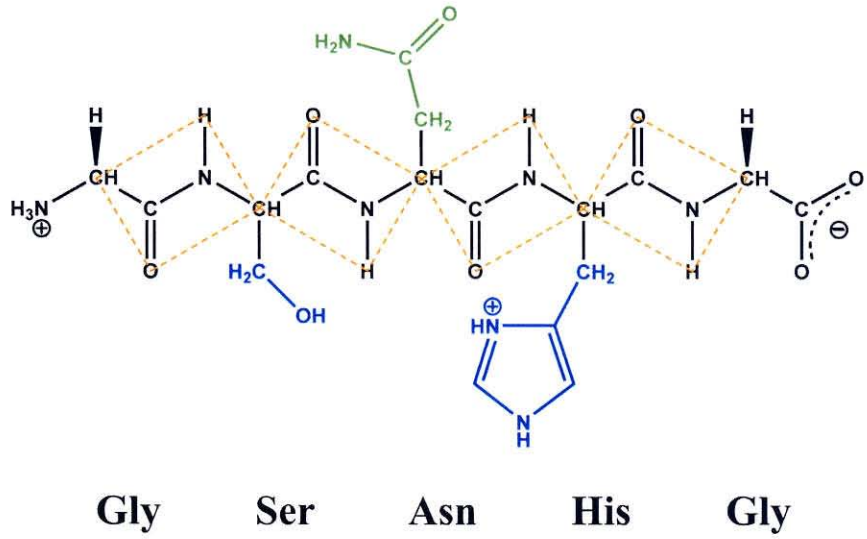


Figure 1-3 – Structure of GlySerAsnHisGly at Neutral pH

Deamidated insulin was the first deamidated polypeptide isolated.[61] This was done by counter current distribution in 1952. For many years after this, the deamidation of proteins was assumed to be an artifact caused by isolation and purification. It was not until the late 1960's that Flatmark[62 - 65, 10] showed that the deamidation of cytochrome c was occurring *in vivo*. Before much peptide work had been done, it was thought that deamidation in proteins was most likely occurring at glutaminyl residues because of the earlier discovery of the rapid deamidation of free glutamine in solution. No such deamidation was observed for free asparagine.

It turns out that free glutamine easily forms a cyclic pyrrolidonecarboxylic acid in solution[66, 67] and also does so when on the N-terminal of a peptide or protein. This can be a problem when working with N-terminal glutamine.[68, 69]

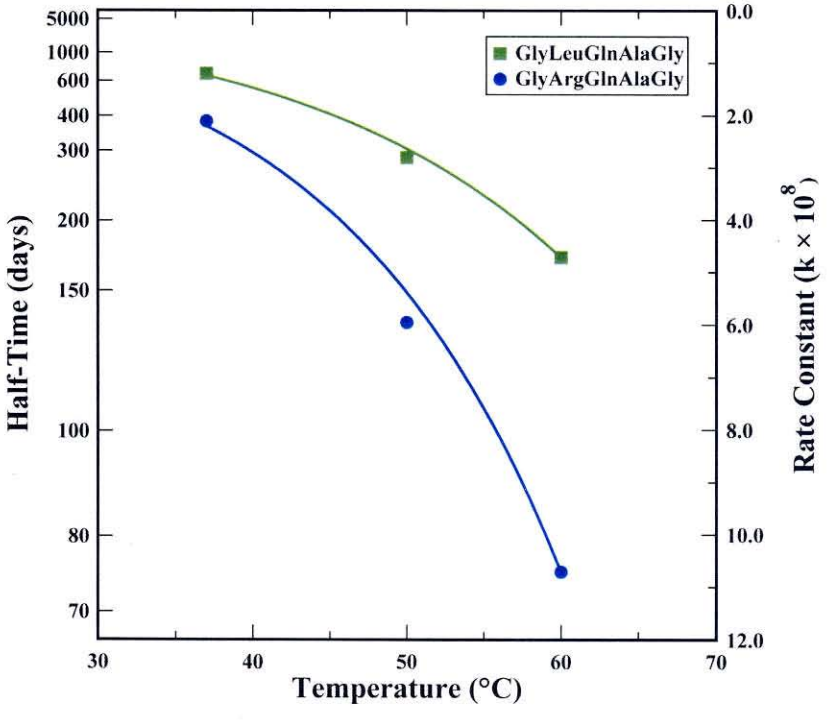


Figure 1-4 – Temperature Dependence of Two Gln Peptides.
Figure adapted from reference 72.

Free asparagine does not form the cyclic acid. In peptides, however, asparaginyl residues usually deamidate much more quickly than do glutamine residues.

The invention of solid-phase peptide synthesis by R. B. Merrifield[70, 71] made possible extensive investigations of sequence dependent deamidation. Almost all peptide work on deamidation has been done on peptides synthesized using solid-phase peptide synthesis. The properties discussed below are all based on this type of peptide work.

Deamidation has been shown to be temperature dependent[72] as illustrated for pH 7.4, KH_2PO_4, Na_2HPO_4, I = 0.2, buffer with 0.001 M peptide, in Figure 1-4. It is also strongly pH dependent[63, 72, 73] as shown in Figure 6-1. Figure 6-4 shows ionic strength dependence.[74]

The buffer dependence of deamidation is best illustrated in Figure 6-3 for GlyLeuGlnAlaGly and GlyArgGlnAlaGly. Three different buffers are shown on this plot and illustrate pH dependence. This buffer dependence is very significant, particularly at physiological pHs. At pH 7.4 the deamidation half-times for these two peptides are about twice as long in Tris buffer than in phosphate buffer. Phosphate buffer catalyzes the deamidation reaction to a greater extent than Tris.

The measurement of fundamental peptide deamidation rates should be done in the most benign buffer available, to avoid confounding catalytic affects of the solution ions. These effects undoubtedly vary with protein structure, so measurement even in a particular physiologic buffer is questionable. Moreover, accurate measurements at pH 7.4 in phosphate buffer could be more difficult. The slope in Figure 6-3 is so steep at this point that minor variations in pH would have a major effect on the rate. In contrast Tris pH dependence is not nearly so pronounced. Therefore, Tris buffer was chosen for the peptide deamidation work reported in this thesis.

The deamidation of AsnGly sequences is known to proceed through a succinimide intermediate.[75–78, 28] This leads, not only to L-Asp, but also to L-isoAsp, D-Asp and D-isoAsp. Figure 1-5 illustrates these possible reaction paths.

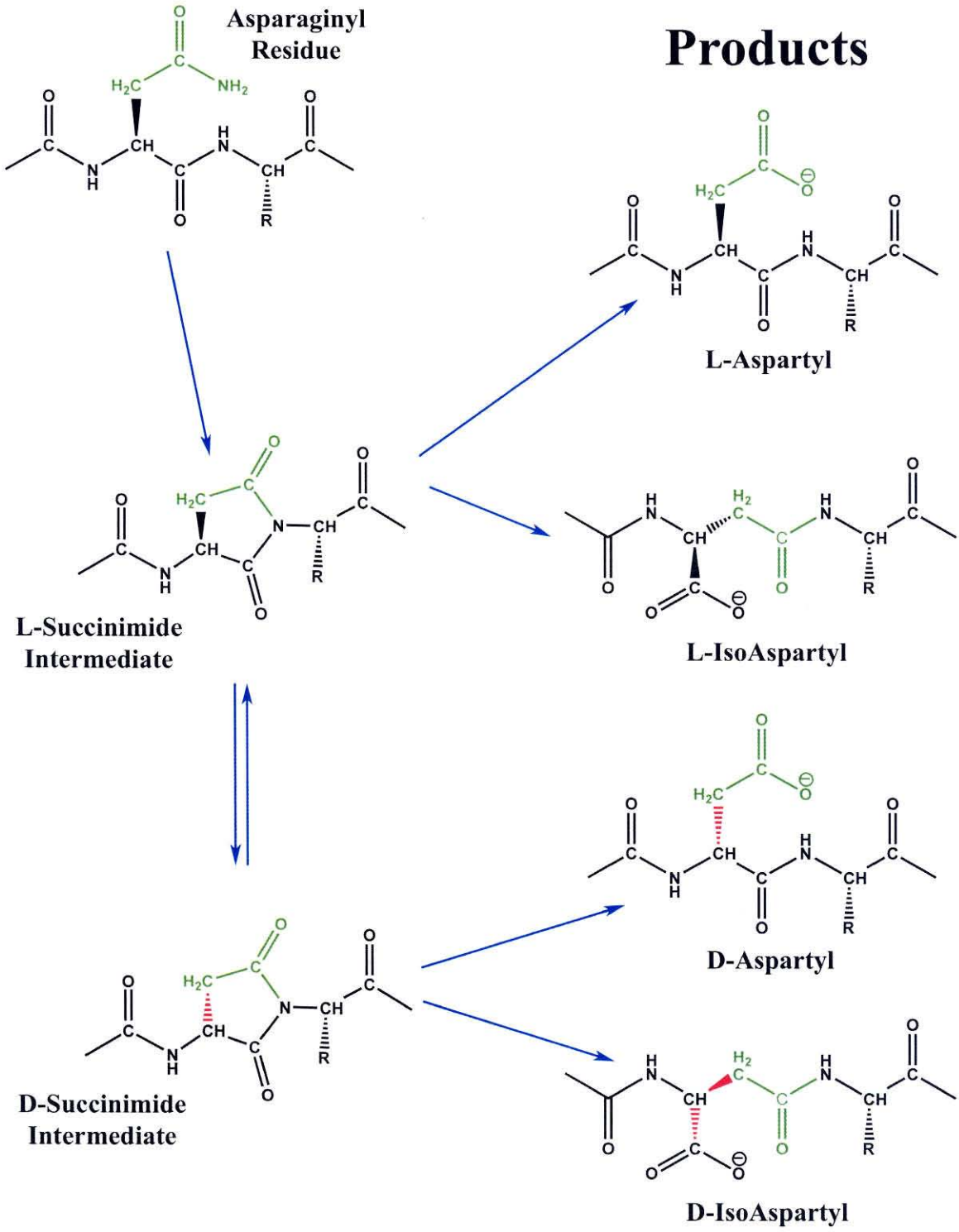


Figure 1-5 – Possible Deamidation Reaction Pathways for Asn Deamidation

An enzyme has been discovered that converts L-isoAsp and D-Asp into L-Asp.[29, 79, 80] This does not correct the most significant change that occurs during deamidation–an increase in charge. It does, however, allow the products of deamidation to be converted back to a single species. This enzyme has been shown to be ubiquitous and important to the health of living things.

The deamidation rate of rat cytochrome c apparently controls its *in vivo* turnover rate.[10, 63, 11] In this case, one Asn deamidates, which changes the structure of the protein and allows another amide to deamidate. The second deamidation is faster than the first. The deamidated products are much more rapidly catabolised then is the undeamidated product.

The turnover rate of rabbit muscle aldolase is identical to the deamidation rate of one of its amides[12, 13]–Asn360. This amide is on the end of the chain and its rate, like that of the first deamidation of cytochrome c, is almost completely sequence controlled.

Triosephosphate isomerase represents another type of *in vivo* deamidation. During the operation of the enzyme, its structure changes briefly and markedly increases its deamidation rate. Thus the number of enzyme cycles determines the overall deamidation rate.[15–17]

Another notable example is the deamidation of Bcl-X_L–a protein involved in monitoring DNA repair. Deamidation of Bcl-X_L is normally suppressed by another protein produced by the DNA. Production of this protein is reduced when the DNA is sufficiently damaged. This allows the deamidation of Bcl-X_L to proceed until the DNA is repaired, at which point the clock is reset. Deamidation of Bcl-X_L triggers cell death. If too much time elapses and enough Bcl-X_L becomes deamidated, the cell is destroyed.[18–21] In this system, an amide clock serves as a monitor of DNA repair. If repair is not sufficiently fast, the cell is eliminated.

The best example of *in vivo* deamidation over very long periods of time is the work that has been done on lens crystallins. Some proteins in the eye lens last for many years.

Deamidation of asparaginyl and glutaminyl residues in human lenses has been shown to occur over several decades. Eye lens crystallin deamidation is being actively studied.[36 - 47]

Deamidation of histone has also been predicted[72] and observed.[81] In addition to these examples, more than 200 species of peptides and proteins have been found deamidated.[82–88] Microheterogeneity in protein preparations is a common occurrence.[3] The negative charge introduced by deamidation facilitates separation by electrophoresis or isoelectric focusing. When more than one amide can deamidate, multiple deamidated different forms of the same protein are possible. This is also true for identical amides in multi-subunit proteins, where one, two, or more deamidated products can be seen as the different subunits deamidate.[89]

Reasons for Current Studies

While about 100 peptide deamidation rates measured under a wide variety of conditions had been reported prior to this work, these rates were not suitable for quantitative predictive work on the deamidation of proteins. The original set of 65 rates[3] was the best in the literature. These rates were measured in 37 °C, pH 7.4, 0.15 M phosphate buffer. Unfortunately, the choice of buffer led to large inconsistencies in these data, which could not be corrected based on the limited data set. These data did, however, provide information on the sequence dependence as well as the half-time range available for peptide deamidation. In some cases, for example, in the deamidation of Aldolase,[12, 13] where there is no secondary or tertiary structure component, these rates were sufficient to explain the protein rates.

Other investigators have measured the deamidation rates of one or several peptides. For review see reference **78**. Most of these were also measured in phosphate buffers and the other conditions varied over a wide range. Apparently as a result of investigator hurry, it has become common practice to use high temperature and/or basic or acidic pH's to accelerate the reaction rates. Attempts are then made to extrapolate these data back to normal conditions. While perhaps of some chemical interest, a deamidation rate derived

from data at extreme pHs or high temperature cannot be said to be representative of physiological conditions.

The result of this was that, prior to our work, the available peptide rates were scattered and poorly understood. AsnGly peptides are known to have unusually fast deamidation rates due to the ease with which they can form the succinimide intermediate, so many investigators have assumed that their proteins are deamidating at AsnGly sequences, an assumption that has now been shown to be frequently in error.[7–9]

A more complete set of rates, measured under identical conditions and with the most benign buffer available, was needed. This work is reported here and in papers which have been published.[6, 73, 26] The results of this work are satisfactory. Not only did these data increase the number of published peptide rates by a factor of more than 5, but the rates themselves are self-consistent and trends in these data are clear. For the first time, it is now possible to know with certainty the primary sequence dependence of deamidation. Moreover, it has been found possible to use these data do develop a method for understanding and predicting primary sequence deamidation rates.[161]

The pentapeptide model is primarily of value as a foundation for understanding the deamidation of proteins. Our next step was, therefore, to use these primary sequence data along with three-dimensional structure to understand and predict the deamidation rates of amides in proteins. Beyond the general observations that flexible regions tend to deamidate faster[90, 91]–as would be expected and had been shown 30 years ago[10–14]–nothing like this had been attempted. The second part of this thesis describes the development of such a method–a procedure that turns out to be remarkably accurate and which depends upon the precise knowledge of the primary sequence dependence of deamidation.

OVERVIEW OF RESULTS

Experimental Results for Primary Structure Rates

The sequence dependence of nonenzymatic pentapeptide deamidation in 37.00 °C, pH 7.4, 0.15 M Tris buffer as determined in this work is summarized in Figure 2-1.

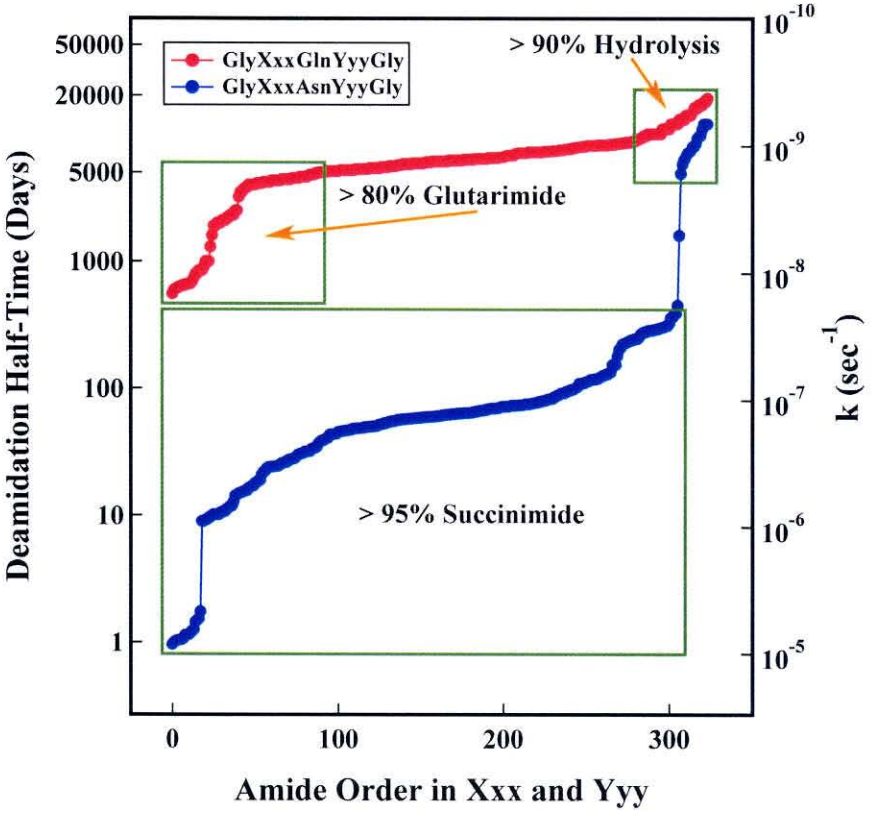


Figure 2-1 – Distribution of Pentapeptide Deamidation Rates for Peptides With Sequences GlyXxxAsnYyyGly and GlyXxxGlnYyyGly

The upper curve is a plot of the ordered deamidation rates for peptides of the type GlyXxx**Gln**YyyGly, while the lower curve is for peptides of the type GlyXxx**Asn**YyyGly. In the case of the Gln peptides, 52 values were actually measured. The remaining rates were determined based on surface fitting to the 52 experimental values. For Asn, 320 sequences were measured and the remaining 4 inferred. Based on the self consistency of the Gln rates and comparison with Asn rates, these interpolations and extrapolations are sufficient.

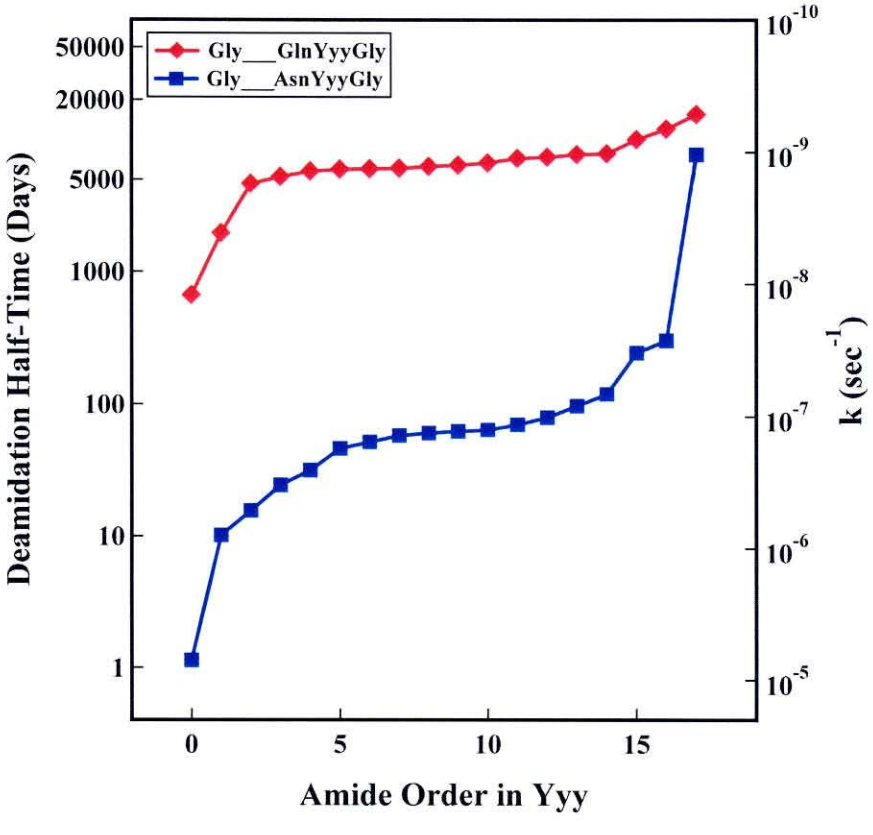


Figure 2-2 – Distribution of Median Pentapeptide Deamidation Rates for Peptides With the Sequences Gly___AsnYyyGly and Gly___GlnYyyGly.

Figure 2-2 is similar to 2-1, except that the left-hand residue effects have been replaced with a median so that the sequences covered are Gly___AsnYyyGly and Gly___GlnYyyGly. The right-hand residue has a much greater effect on the deamidation rate than does the left. Figure 2-3 shows these results on a two-dimensional surface.[163]

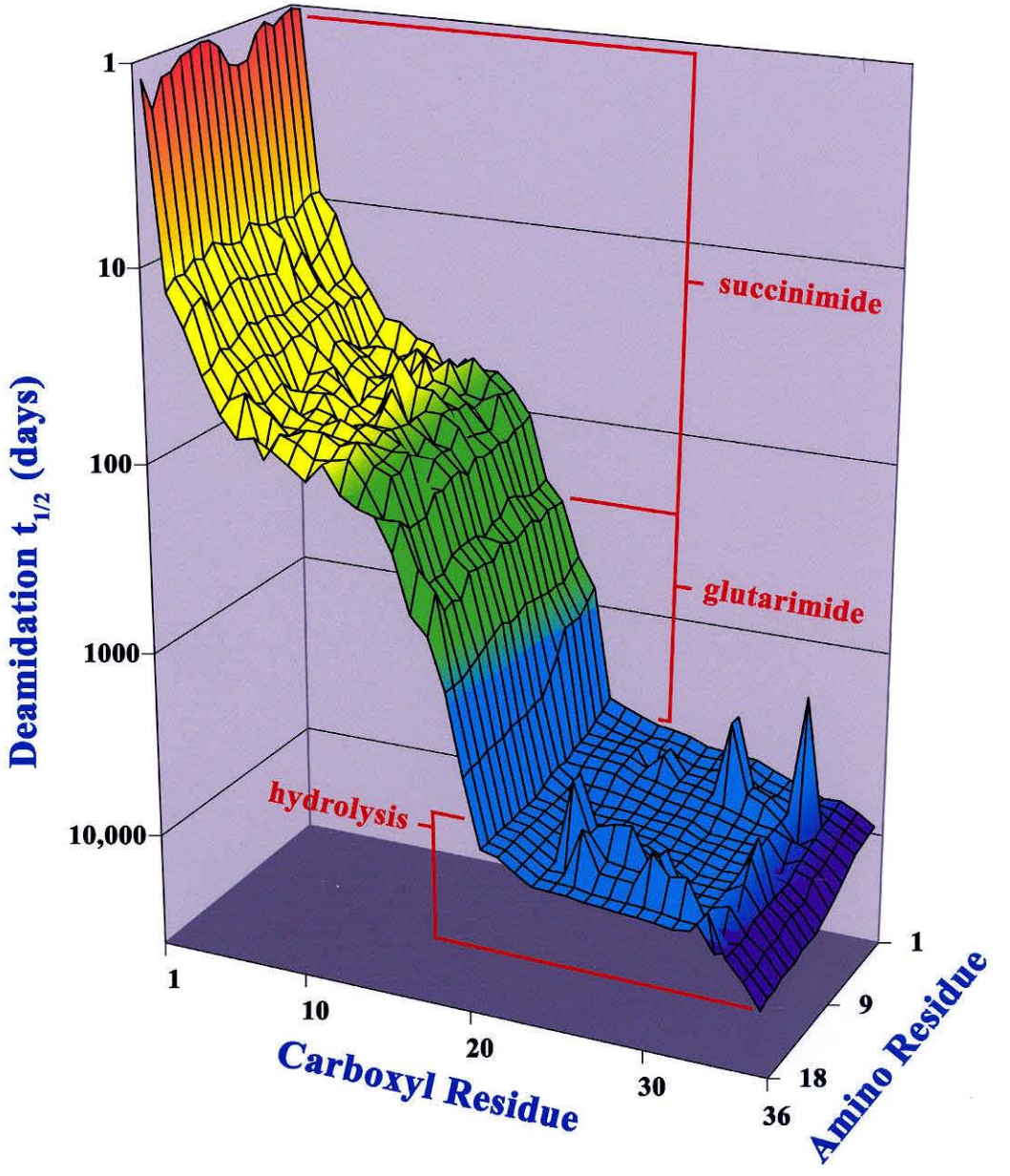


Figure 2-3 – Sequence Dependence of Asn and Gln Pentapeptide Deamidation Rates

Figures 2-1, 2-2, and 2-3 illustrate two important results of this peptide work. First, based only on the sequence of three residues, deamidation half-time can be set over a range of four orders of magnitude, which includes most biologically significant times. Second, this range is thoroughly covered with minor gaps in rate. With modulation by secondary, tertiary, and quaternary structure in proteins, virtually any time period is available for the use of amides as molecular clocks.

The shortest pentapeptide Asn half-time measured was 0.96 days and the longest was 455 days, except for AsnPro rates the shortest of which is 1,600 days. Almost any rate in between can be achieved by correct choice of sequence. Of the 52 Gln half times measured so far, the shortest was 560 days–starting about where Asn leaves off–and the longest was 18,000 days. The highest AsnPro rate is 12,000 days. It is clear from the measured sequence dependence of Asn and Gln deamidation that almost any half-time throughout this range can be selected by correct choice of sequence.

If we assume that a biological event could be triggered by as little as 10% deamidation or as much as 90%, this requires between 1/6 of a half-time and 3 half-times. Thus, with only the sequence dependence of pentapeptides, any interval from 4 hours to 60,000 days (160 years) can be timed–a range of over 5 orders of magnitude. This range is extended and made more versatile by means of secondary, tertiary, and quaternary structure effects.

Prediction of Protein Deamidation Rates from 3-D Structure

The main effect of three-dimensional protein structure is to slow down the primary sequence rates, although there are cases in which the primary rate is accelerated. This fact underlies our development of the 3-D prediction method for Asn rates. The method is basically a way of determining how much the primary rate is suppressed by 3-D structure.

This method has been optimized by comparison with the protein deamidation literature and computerized to allow prediction of deamidation rates in any protein for which the three dimensional structure is known. Figure 2-4 summarizes the results of applying this procedure to the entire Brookhaven PDB protein data bank.[9]

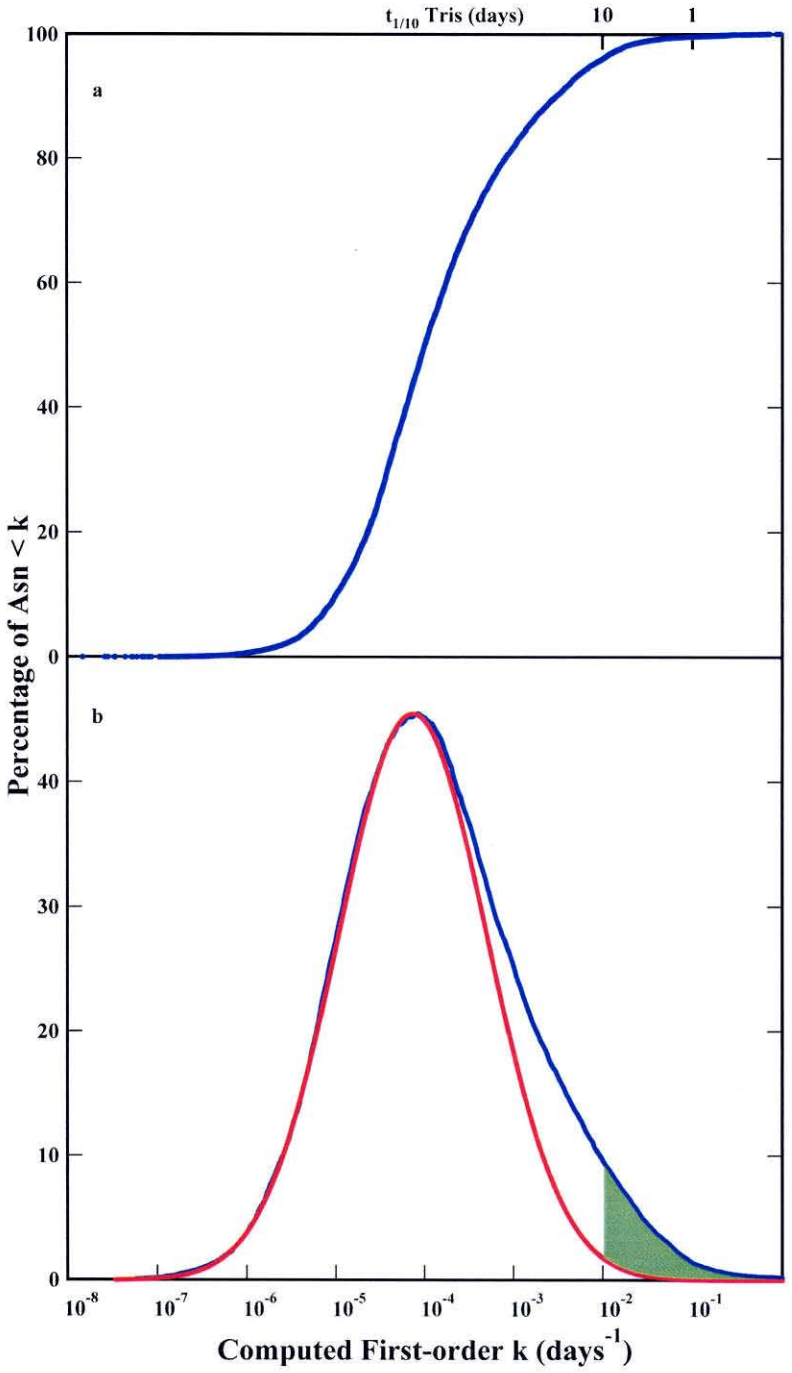


Figure 2-4 – Cumulative Distribution and Derivatized Curve for 170,014 Amides in the Protein PDB Databank.

There are two important features in Figure 2-4 besides the overall distribution function. First, it is possible to slow an amide down to the point were it is essentially no longer deamidating, except, perhaps, by slow hydrolysis. The calculation method is not calibrated in the very long half-time region on the left-hand side of the curve, but it is expected that the qualitative results are about right. This means that the deamidation of an amide can be essentially stopped by three-dimensional structure. Second, a very large number of amides in proteins are not slowed down, but are programmed to deamidate quickly. This may be done for a beneficial purpose, since these deamidations could easily have been suppressed.

The Gaussian fit to the left-hand side of these data follows the calculated values very well until the fast amides are reached. This deviation starts to take place at 1,000 days and becomes very pronounced in the 100 day to 1 day region. This shows that deamidation in this region is not determined by a group of similarly sized independent variables and that, regardless of the fast rates, these amides are genetically preferred. We think that many of these amides were specifically chosen as biological timers in these proteins.

Figure 2-5a shows Asn deamidation half-times for 17,935 peptides and proteins.[163] The half-times are plotted vs. the ordered primary sequence and three-dimensional structure activation energy contributions to I_D in each protein. These genetically determined half-times range between 4 hours and more than 100 years. They have been calculated by means of experimental primary structure deamidation rates[6, 73] and computed secondary, tertiary, and quaternary structure factors derived from known protein amide deamidations.[7-9]

The yellow and orange parts of the surface in Figure 2b include peptides and proteins with single deamidation half-times in 0.15M Tris·HCl, pH 7.4, 37.0 °C of about 10 days or less. The primary sequence rates were determined in Tris-HCl because Tris is a relatively weak catalyst of deamidation. Other buffers such as phosphate are stronger catalysts, and components of physiological fluids also catalyze deamidation. These computations provide, therefore, baseline deamidation rates to which the effects of other substances can be added. *In vivo* rates tend to be about 2 to 3 times faster than Tris rates.[8]

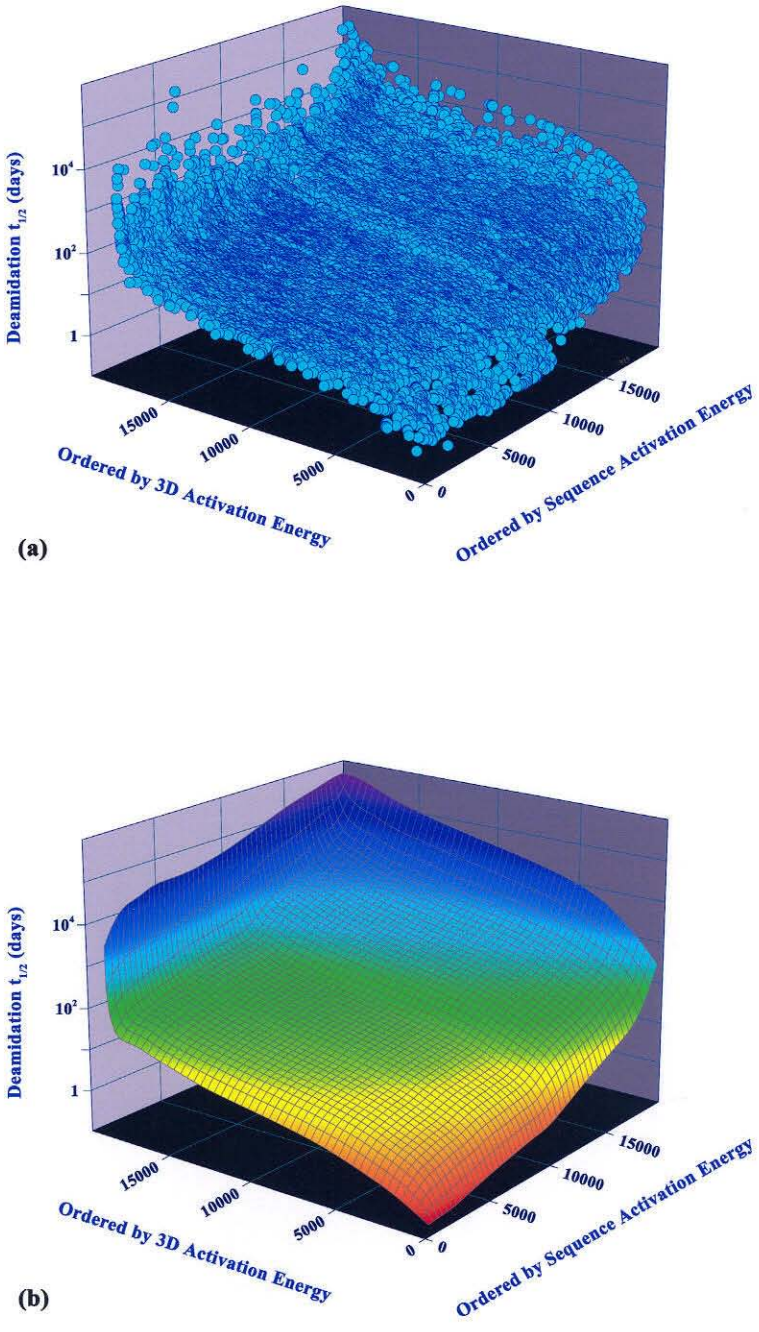


Figure 2-5 – Deamidation Rates for 17,935 Peptides and Proteins Ordered by Primary and Three-Dimensional Effects

Overall Results

With the publication of these data, the peptide work needed in this field is to a large extent complete. A few more actual Gln measurements would improve accuracy of these values and actual rate measurements on the double amides (i.e., GlyXxxAsnAsnGly, GlyAsnGlnYyyGly, etc.) should be carried out. Additionally, measurements of buffer type and concentration dependence, as well as pH and temperature dependence on several representative sets of Asn and Gln peptides would be useful. These are, however, not essential to determining deamidation rates in proteins within the accuracies needed for experimental biological purposes.

Many variables can affect the deamidation rate of an amide in a protein. When a protein is placed in a living a system even more factors enter into its deamidation rate. The precision that would be introduced by more peptide work is probably much less than these factors. When a precise rate for a particular amide in a protein is needed, it should be measured directly. When approximately correct rates are needed, they can now be predicted in the case of Asn with the current methods.

This three-dimensional prediction method permits determination, with greater than 95% reliability, of the relative stabilities toward deamidation of the amides in any protein for which the three-dimensional structure is known. It also permits reliable estimation of absolute deamidation rates. Improvements can be made as more protein data becomes available. Of special interest will be extension to Gln and very long-lived Asn. Further refinement is currently primarily limited by insufficient knowledge of unstable amides in proteins for which crystal structures have been determined.

Chapter 3

MATERIALS

Purified Water

All of the water used in these experiments was first distilled using a Consolidated Model NP-E-1 stainless steel still. It was then purified to 18.2 MΩ using a Labconco Water Pro PS (Catalogue Number 90006-00) with the following cartridges: 1. Carbon Filter, 2. Deionization Cartridge, 3. Deionization Cartridge, 4. Organic Absorption Cartridge, 5. Ultraviolet Light, 6. 0.2 μM Hollow Fiber Filter (Part 90929).

Solid-Phase Synthesis Chemicals

Derivatives used for the Fmoc solid-phase synthesis were 9-fluorenylmethoxycarbonyl (Fmoc)-Ala, Fmoc-Arg(2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl), Fmoc-Asp(O-*t*-butyl), Fmoc-Asn, Fmoc-Cys(acetamidomethyl), Fmoc-Cys(Trityl), Fmoc-Glu(O-*t*-butyl), Fmoc-Gln, Fmoc-Gly, Fmoc-His(Trityl), Fmoc-Ile, Fmoc-Leu, Fmoc-Lys(*t*-butyloxycarbonyl), Fmoc-Met, Fmoc-Phe, Fmoc-Pro, Fmoc-Ser(*t*-butyl), Fmoc-Thr(*t*-butyl), Fmoc-Trp, Fmoc-Tyr(*t*-butyl), Fmoc-Val, and Wang resin substituted with 0.67 meq/g Fmoc-Gly or Fmoc-Ala or Fmoc-His(Trityl) as needed. For c-terminal peptides, Rink Amide MBHA resin was used. These were purchased from Peptides International.

Solvents used for the synthesis included *N*-methylpyrrolidone (NMP), dimethylformamide (DMF), methylene chloride, and methyl-*t*-butyl ether.

General Reagents and Solvents

Standard reagents of the highest purities available were purchased from VWR Scientific.

Chapter 4

SOLID-PHASE SYNTHESIS OF PEPTIDES

Chemistry

The peptides were synthesized by using Merrifield solid-phase peptide synthesis[70, 71] with Fmoc chemistry. They were synthesized in the laboratory of Professor R. B. Merrifield at Rockefeller University, with Professor Merrifield's advice and assistance.

Each well of the synthesizer initially contained 0.1 g of Wang resin. Double couplings for 45 min each with a 3 min 1.5 ml wash of 50:50 *N*-methylpyrrolidone (NMP)/dimethylformamide (DMF) between couplings were used. Reagents for each coupling were 0.5 ml of 0.5 M derivative in NMP; 0.5 ml of coupling reagent that was 0.5 M 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate and 0.5 M 1-hydroxybenzotriazole in DMF; 0.25 ml of neutralizer that was 2 M *N,N*-diisopropylethylamine in NMP; and 0.25 ml of DMF. Double deblocking with 1.5 ml of 20% piperidine in DMF was carried out once for 5 min, followed by a second 15 min deblocking. Resin and side chain protecting groups were removed with 1.5 ml of scavenger-containing trifluoroacetic acid (TFA) solution at room temperature for 2 hours. The TFA solution was TFA/anisole/ethylmethylsulfide/ethanedithiol in the proportions 93:3:3:1, respectively, by volume. The product was filtered and the resin washed once with 1 ml of TFA. Eleven separate syntheses were performed, with 96 peptides synthesized simultaneously in most cases. The acetamidomethyl-blocking groups of cysteine were approximately 20% removed in TFA.

The peptides were precipitated and washed three times with 15 ml of methyl-*t*-butyl ether, vacuum dried, dissolved in 18.2 MΩ distilled and purified H_2O, divided into five parts in low-temperature vials (Nunc Model 375930), freeze dried, and stored at −80 °C.

Procedure and Equipment

The peptides were synthesized in an Advanced ChemTech Model 396 MBS synthesizer with a 96-well reaction block.

This synthesizer operates by means of two robotic arms carrying needles. The needle is inserted into a reagent bottle, a syringe draws the reagent up, and it is then inserted in the appropriate chamber of the reaction block before the contents are forced out of the end of the needle by reversing the syringe. Mixing of the reagents is done by orbital shaking motion of the reaction block. The reaction block is made of Teflon, contains an array of wells 12 × 8, and is covered by a sheet of rubber which is held tightly in place with a stainless steel cover. It is about 8" × 10" in size and holds up to 8 ml per well. Typical loading is 0.1 g of resin per well. Emptying is accomplished by applying nitrogen pressure to the reaction block. This causes the liquid to drain out through a frit at the bottom of each well. A loop in the opening at the bottom of the vessel keeps the liquid from draining out except when pressure is applied.

This synthesizer works well, but it requires careful watching when large numbers of peptides are being synthesized. The reagents are expensive and can be destroyed by a mistake that goes uncorrected. For this reason, the entire synthesis was carefully watched. Each step was monitored. Every time an amino acid derivative was transferred, the sequence was checked to make sure it was the right one. Every time a reagent was delivered it was carefully checked. Each block drainage was also watched to make sure the proper amount of liquid exited the reaction block.

After synthesis, the deblocking and cleavage from the resin was carried out in TFA as described above. Unfortunately the manufacturer of the synthesizer neglected to provide a convenient way of draining the TFA and peptide solution from the reaction block into a centrifuge tube large enough to allow precipitation with ether. This is no problem for a few peptides, but when all 96 wells of the reaction block are being used it is difficult to recover the TFA without mixing the wells.

A rack was needed of the same shape as the reaction vessel block on which the reaction vessel block could be mounted. This rack required holes of the same size as a 15 ml centrifuge tube, but no larger, so that the tubes would be stopped by the threads at the top.

A rack of this type was made from ½" Lucite. It is essentially a square box with 96 round holes in the top–one for each tube. The top is removable, and the height of the bottom of the box is adjustable, so that the length of tube coming out of the top can be adjusted to fit up against the reaction block which mounts on top. A special set of recessed holes at the corners hold the legs of the reaction block firmly in position. The holes in the top were designed for an array of 96 Falcon 352096 polypropylene tubes. The tops of the tubes touch each other when they are inserted. This means that support of the array depends on about a 1/8" strip of plastic between each tube.

Drilling this top had to be done very precisely and without cracking the top. I worked for about half a day in the Rockefeller University machine shop to do this. At that time, Rockefeller University had a beautiful machine shop. There was a pleasant older machinist there who supplied me with the necessary tools and gave helpful advice. It was this machine shop in which parts of Professor Merrifield's first peptide synthesizer were built. This machine shop also built pieces of Stein and Moore's first amino acid analyzer and Lyman Craig's original countercurrent distribution machine. Unfortunately, Rockefeller University recently destroyed the machine shop and disposed of all the machine tools.

The pieces were glued together by using dichloromethane. Four rubber stoppers were cut to size and glued to the bottom as feet. This rack was used for all of the TFA transfers and proved to be invaluable.

Purification Procedures

With so many peptides, purification by the normal methods would be prohibitive. While an automated system could be built with several columns, it would add considerably to the effort required.

In our case, the mass spectrometer is highly selective. Only impurities with the same mass as the peptide can interfere with the analysis. Our reaction conditions are designed to minimize intermolecular effects between peptides. Impurities are substantially lower in concentration than the desired peptide and are unlikely to affect the reaction significantly. For these reasons, we decided not to purify the peptides beyond several ether washes and freeze drying (lyophilization) from water. In most cases, impurities in the peptides were so minimal as to be indistinguishable from background in the mass spectrometer.

After the TFA-containing peptides were transferred to centrifuge tubes, about 15 ml of methyl-*t*-butyl ether were added to each tube and the peptides were allowed to precipitate. The peptides were then centrifuged to the bottom of the tubes and the ether was poured off. This was followed by two more 15 ml ether washes. The peptides were then vacuum dried, dissolved in 18.2 MΩ purified H_2O, divided into five parts in low-temperature vials (Nunc Model 375930), freeze dried, and stored at -80 °C.

Amino Acid Analysis of Peptides

Yields were determined by hydrolysis with 6 N HCl in H_2O at 95 °C for 48 hours followed by amino acid analysis in a ThermoQuest HPLC with a P4000 pump, AS3000 autosampler, and UV6000LP detector combined with a Pickering PCX5200 derivatizer (Pickering Laboratories, Mountain View, CA), column, ninhydrin, and lithium buffer system. Yields averaged about 70% with the remaining peptide either not washed from the resin or solubilized during the ether washes.

These analyses provided the measurements of concentrations and verified each of the peptide compositions.

Problems Encountered during Synthesis

There were essentially no special problems with the synthesis. With Professor Merrifield's advice and assistance, the chemistry worked perfectly. Every peptide we have measured has been present in high yield with good purity.

Summary of Peptide Synthesis

The peptide synthesis was a major part of this research problem. We synthesized approximately 1,000 different individual peptides in Professor Merrifield's laboratories at Rockefeller University. The automation makes this possible, but still we watched about 150,000 liquid transfers to make these peptides. Both Professor Merrifield and Mrs. Merrifield were very generous and helpful during this work.

Chapter 5

DEVELOPMENT OF EXPERIMENTAL APPARATUS AND PROCEDURES

Warm Room and Oven

The temperature chosen for the deamidation reactions was 37 °C. This is the deamidation temperature that has been used historically and was initially chosen as an approximation to physiological conditions.[1, 25, 3]

It was necessary to control the temperature as accurately as possible. To do this, a special warm room was constructed. It consisted of the 10 ft^3 sample oven (Blue M Stabil-Therm) pictured in Figure 5-1 placed in a small heated room about 7' x 9'. The oven was set out from the wall to provide free air circulation around it. The inner glass door provided an easy way to observe the samples and read the internal glass standard thermometer (one of nine temperature sensors used in this system) without disturbing the temperature.

The outside room temperature was kept at 35.5 °C ± 0.5 °C using a Cole-Parmer Model 89000-10 temperature controller connected to a standard 1500 watt water-filled electric floor heater from International Brand. Uniformity in temperature throughout the room was achieved by using a small electric fan aimed at the ceiling (Patton Model U2-1487, 120 VAC, 1.5 Amp set on Low).

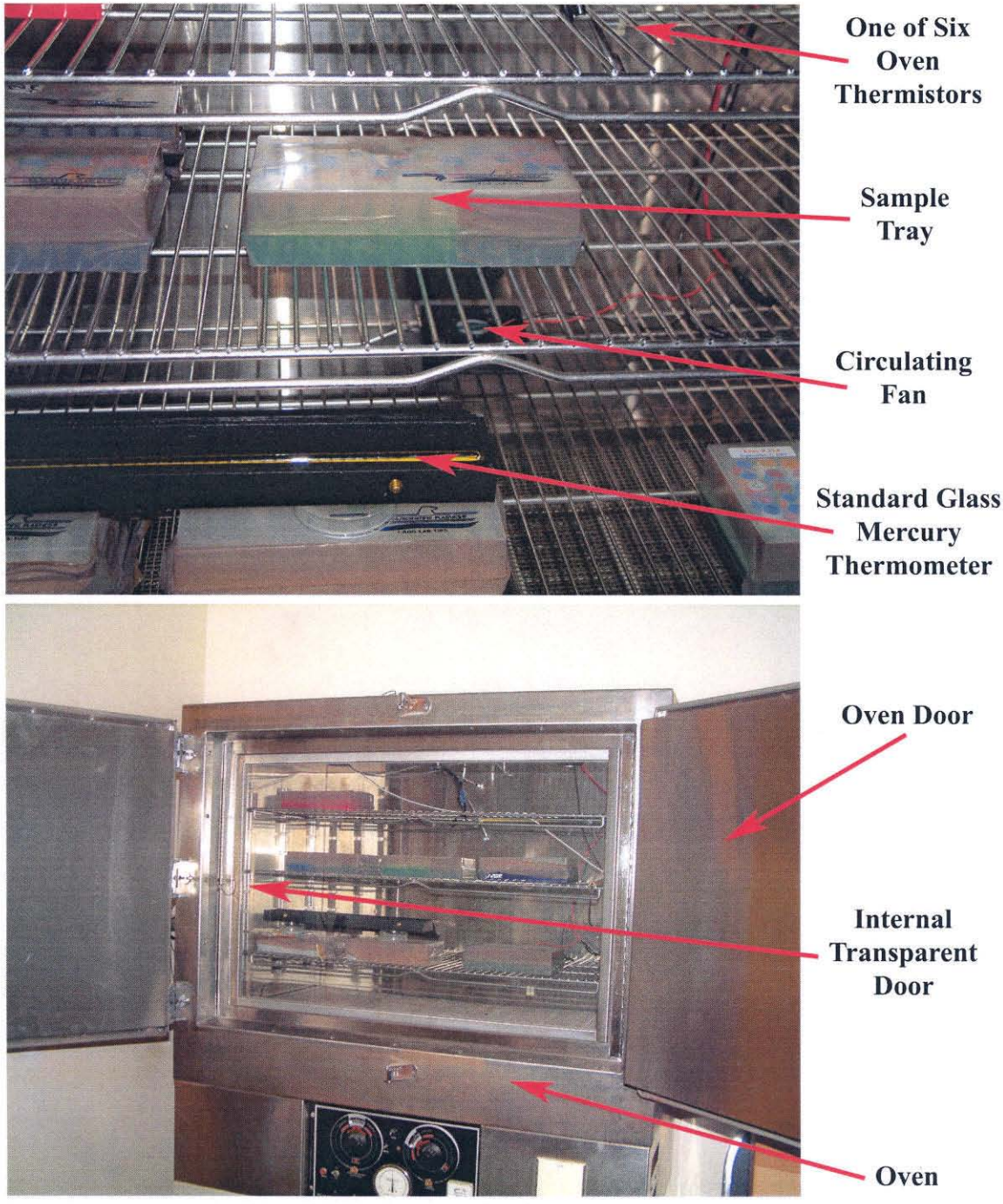


Figure 5-1 – Photographs of Oven. The Room Surrounding this Oven is Kept at 35.5 °C ± 0.5 °C.

The oven was held at exactly 37.00 °C ± 0.02 °C by a separate controller system. This control system presented a special problem because of the high precision required. The only sufficiently accurate commercially available controller we could locate was a YSI Model 72 controller, which was no longer manufactured. The other controllers available were generally only accurate to 0.1 °C and not very stable in this range. YSI built their controller for calibrating thermistors and marketed the controller itself for a few years. We located three units (two from used scientific equipment dealers and one on eBay), had one of these refurbished by the factory, and kept the other two as spares. A YSI model 400 thermistor was used with the controller. Both controllers are pictured in Figure 5-2.

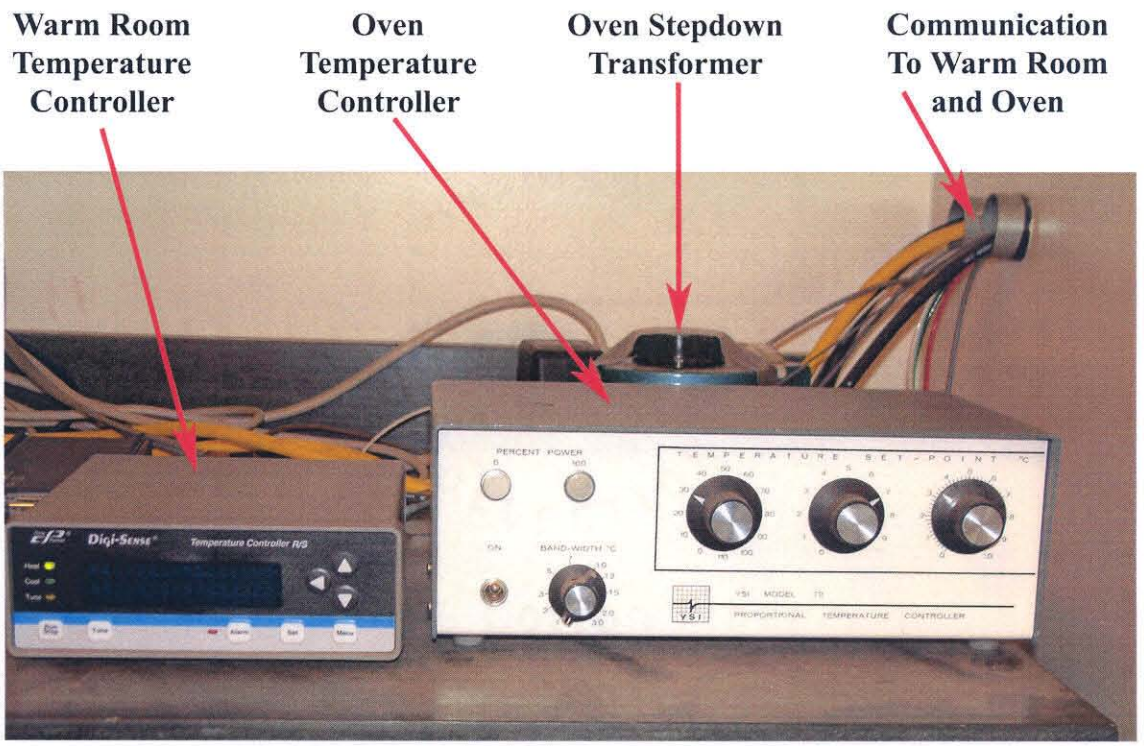


Figure 5-2 – Photograph of Oven Control Systems

The original heating elements from the oven were wired directly into the controller through a step down transformer. The transformer kept the voltage down, assured that the elements were never very hot, and reduced temperature oscillations. In addition, two

ASTM calibrated thermometers were placed in the oven to keep the system calibrated to ASTM standards. One of these was a digital thermometer employing a thermistor (Ertco-Eutechnics Model 4400–Serial Number 302931), and the other one was a glass mercury thermometer (Ertco-Eutechnics ASTM 64C-FC, Total Immersion Glass Mercury Thermometer, Serial Number 5498). The samples were arranged in the oven to allow as much air circulation around them as possible, and a small computer fan kept the air circulating. We used a 5 watt fan. Larger fans add too much energy to the surrounding air and prevent temperature regulation when there is only a 1.5 °C difference between the inside and outside temperatures.

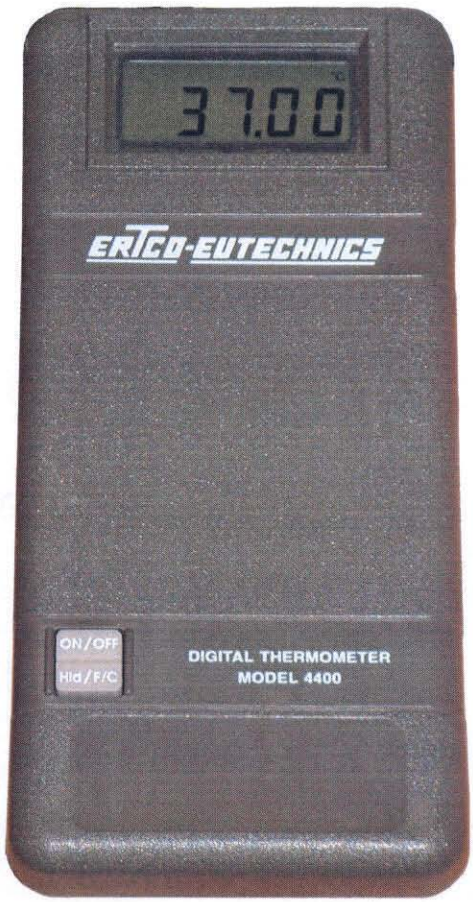


Figure 5-3 – Photograph of Standard Digital Thermometer

This construction had several major advantages over a simple incubator without a heated outer room. First, the temperature in the oven was capable of being controlled much more precisely than would be otherwise possible. This is because the outside air temperature was only 1.5 °C below the inside temperature. Heat transfer was therefore very slow through any part of the oven wall. Small and precisely controlled amounts of energy were added to keep the oven warm. The regular fluctuations in temperature were less than 0.002 °C, leaving only potential systematic error. Secondly, samples could be added and removed from the oven without causing a dramatic drop in the oven temperature. Removing a sample typically causes only about a 0.2 °C drop in the interior air temperature, which is quickly corrected by the controller. Since it takes a while to transfer heat from the air to the samples, which were sealed in special boxes, there was essentially never any substantial change in sample temperature.

In addition, both the room temperature and oven temperature were monitored by three thermistors connected to an IBM computer running Windows 95, through an InstruNet Model 100 analog/digital converter. The computer made a continuous plot of oven and warm room temperatures for monitoring purposes. This system also monitored our mass spectrometer room temperature.

As an additional protection in case of equipment failure, two Goldline Model SP-34 thermostats were connected to the oven. These are shown in Figure 5-4. One of these was set to trigger above 37.5 °C and the other one below 36.5 °C. They were both connected by means of underground wiring to a 120 decibel bull-horn in the roof of a nearby building, which could be easily heard ½ mile away. While these thermostats were occasionally triggered during extensive loading or unloading of the oven, the rest of the equipment was so reliable, that they were never otherwise used. The bull-horn was also connected to a timer that was programmed to sound the horn 5 min prior to the time when sample points were scheduled to be removed, but was usually reset manually before it actually sounded. This was especially useful as a reminder for removing samples from the oven during the middle of the night for short term rate experiments that needed samples every few hours.

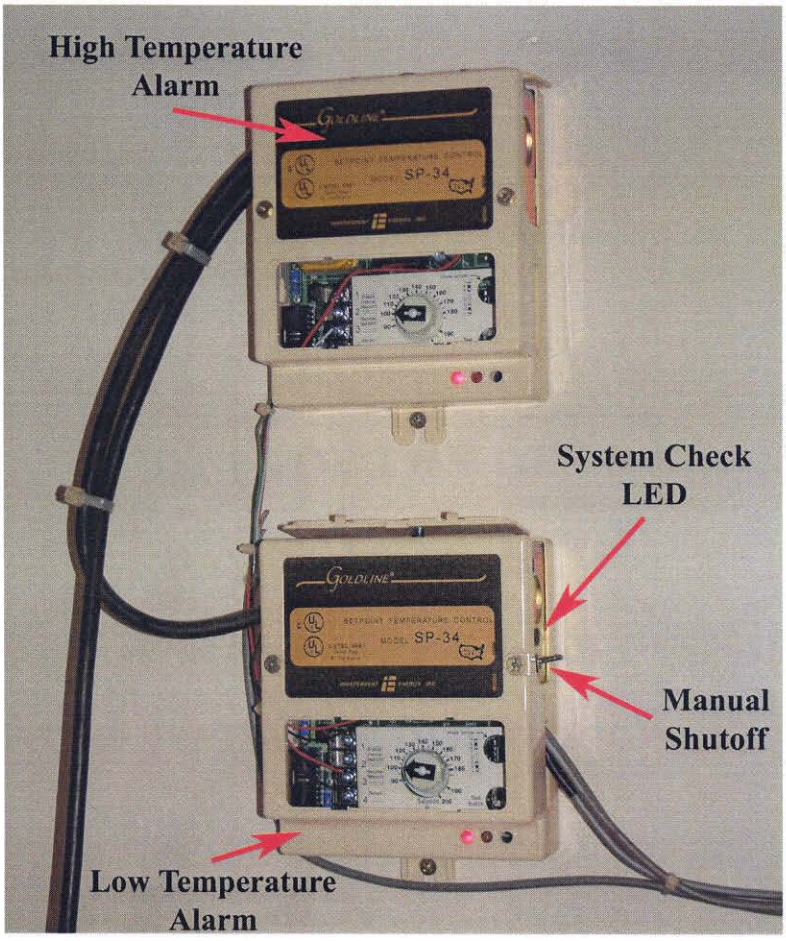


Figure 5-4 – Photograph of Alarm Systems

The main control system was run through an Exeltech XP1100 full sinewave inverter connected to a continuously charged battery bank which improved temperature control and served as a precaution in case of power failure. This is picture in Figure 5-5. The battery bank was capable of running the system for at least 12 hours, without intervention. There were two other large battery banks available and three backup diesel generators capable of supplying power to the system in case of an extended power outage. The main lab generator is shown in Figure 5-6 and one backup battery bank in Figure 5-7.

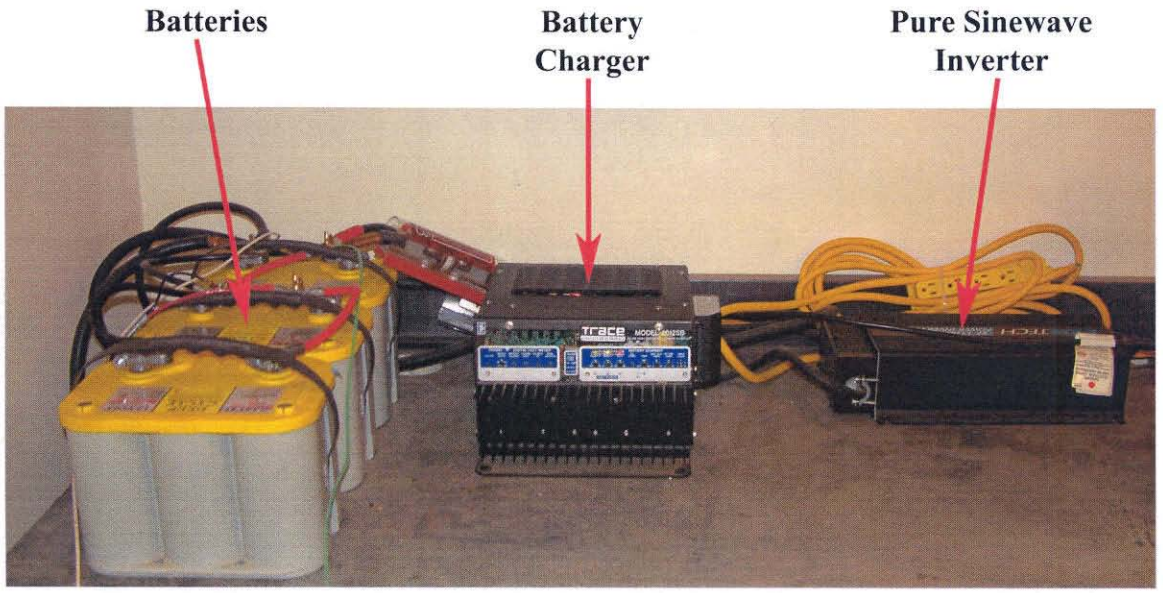


Figure 5-5 – Photograph of Oven Power System

The temperature controllers, computer, and battery bank were placed outside of the warm room and communicated by cables run through the wall. This allowed monitoring and adjusting without disturbing the warm room temperature. The center of the conduit shown for this purpose in Figure 5-2 is filled with insulation, though this is not visible.

Note that this setup requires that the laboratory temperature never exceed 35.5 °C (95.9°F). If this happened, the warm room regulation would shut off since it can only provide heat, and the room temperature would rise. At about 36.5 °C, the energy output from the oven fan would start to exceed the output from the oven control circuit and the oven temperature itself would rise. The outside temperature is, however, only above this point in the middle of the day for a few weeks every year in our location. The laboratory insulation, day to night temperature fluctuations, and laboratory air conditioning prevented this from happening. In a warmer climate, backup air conditioning for the warm room would be advisable and would allow this system to be operated at lower temperatures.

Figure 5-6 – Photograph of Generator

One improvement would be especially valuable for uniformity and stability of the oven. This would be the addition of one more chamber. If the outer room were held at 35.5 °C and another chamber were placed around the current chamber, the space between the inner and outer chambers could be held reliably at 36.9 °C without risk of exceeding 37.0 °C, providing efficient air circulation was used. The innermost chamber could probably then be held within a few thousandths of 37.000 °C with a minor amount of air circulation, if the controller were kept under constant environmental conditions. Our current system varies between 36.99 °C and 37.01 °C with regular oscillations less than 0.002 °C. This system has been running without interruption at 37.00 °C ± 0.02 °C for more than 3 years.

Figure 5-7 – Photograph of Backup Battery Bank

Sample Vials Used in Rate Experiments

Individual points during the rate experiments were taken by filling vials for each point at the beginning and removing and freezing them at -80 °C on set schedules.

The ideal sample vial should be completely gas tight to prevent evaporation. It should be made from a material that does not absorb the peptide. Of particular concern are materials that might absorb the amidated and deamidated peptides to differing extents, thus leading to systematic errors in the analysis. Most glass vials also have the special problem that they introduce sodium and potassium into the solution, which complicates mass spectrometric analyses.

For this work we selected 1.5 ml polypropylene vials with screw caps and rubber O-rings (VWR Scientific Products Cat. No. 201700-213). Half of the vials in each experiment, alternating in reaction times, were fitted with pressed-in 0.002-inch-thick Teflon film liners and tops. The sample sets were arranged with color-coded caps and sample numbers on the cap inserts.

Before use, the Teflon film was washed 3 times with dichloromethane, followed by 3 rinses with acetone and then 3 rinses with concentrated nitric acid. This was followed by 12 rinses with purified H_2O (the same water used in the experiments). The rinses were done in batches with three 1000 ml graduated cylinders filled with rolled Teflon film. The solvents were poured from one cylinder to the next during the procedure and allowed to stand about 5 min for each rinse with dichloromethane, acetone, and nitric acid. The H_2O rinses were only about 2 min each. Afterward the Teflon sheets were hung from the lights in the laboratory, sprayed with more purified H_2O as a final rinse, and left overnight to dry.

The dried Teflon was then cut into 1 inch squares and formed over the bottom of a 1.5 ml ependorf tube that was taped upside down to the table. The bottom of the ependorf tube was carefully polished with a knife to remove all burrs that might puncture the Teflon and then thoroughly washed with purified H_2O before use. These formed Teflon inserts were placed in the tubes and another square of Teflon about ¾ inch on each side was placed on

top before screwing down the cap. The vial inserts were all prepared and the vials labeled before any samples were placed inside, but the caps were not screwed down until after the sample was added. Throughout the procedure the Teflon pieces were kept on large squares of similarly cleaned Teflon to keep them clean, and they were only handled by the edges which do not come in contact with the sample.

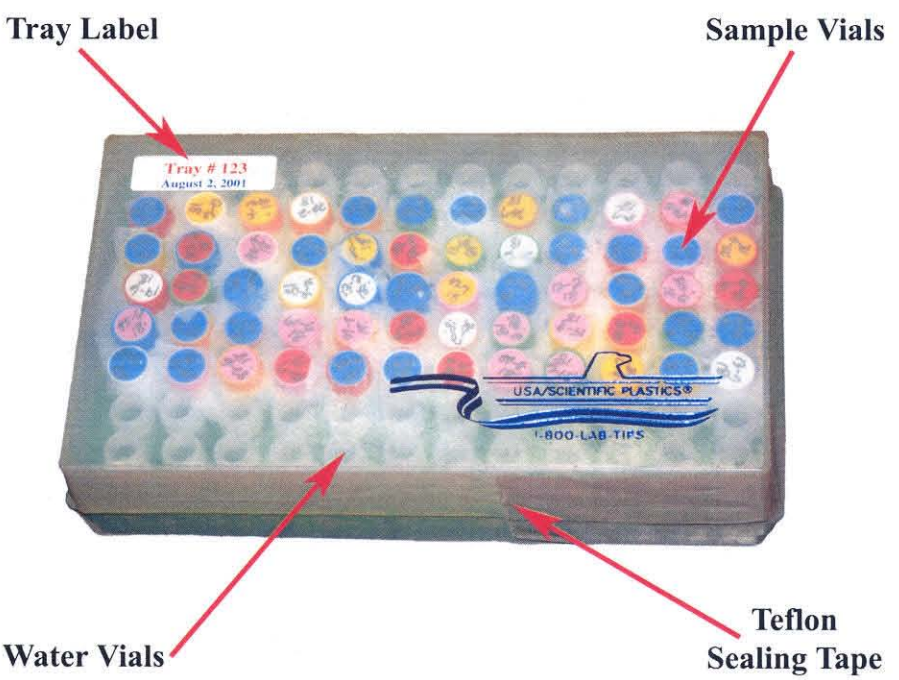


Figure 5-8 – Photograph of Sealed Sample Tray

Each vial contained 100 μl of solution. The vials were placed in sealed 96-well trays along with several open vials of water to keep the atmosphere hydrated and prevent evaporation as shown in Figure 5-8. The tape shown wrapped around the trays is a special Teflon tape and minimizes evaporation. As much as possible, these trays were combined

and labeled in such a way that they could be removed and placed in the freezer without opening them.

Neither the polypropylene nor Teflon appeared to absorb significant amounts of peptide. However, there were sometimes slight differences in deamidation rate between the two. In these cases, there usually were faster rates in the polypropylene vials. This difference was ordinarily less than 1 to 2%, so all 18 points were used in the calculations. The most likely reason for this is a very small amount of differential absorption on the vial walls, but this aspect was not investigated further.

Procedure for Filling Vials

Tris buffer was prepared in large batches for use in the experiments. The pH was adjusted to be exactly 7.40 at 37.00 °C when diluted by addition of the correct amount of peptide solution. Appropriately sized samples of water-dissolved peptide were mixed with this buffer and a correct aliquot of water to produce 2 ml of sample, 0.001 M in peptide and 0.15 M in pH 7.40 Tris buffer. Each preparation was different and was based on the concentration of peptide in the stock solutions as measured by amino acid analysis.

These samples were kept frozen except for the short period of time required to prepare them. The original peptide stock solutions were only unthawed as needed and immediately refrozen. When the solutions were ready, they were unthawed and pipetted into the previously prepared reaction vials. Each vial was then carefully tightened by hand and the sets of vials were then refrozen. Again samples were only thawed out as needed and then immediately refrozen.

Vial tightening was a challenge and physical hazard, which would probably be forbidden by some safety departments. Consistently torquing several thousand vials in a row to the same tightness–particularly the ones with Teflon inserts, the caps of which were very hard to screw down over two layers of Teflon–is very hard on fingers. We filled a total

of about 9,500 of these vials throughout the course of beginning these experiments, and there was not enough time for my hand to recover between sets.

Once the vials were ready, they were unthawed and placed in trays like the one shown earlier in Figure 5-8. Water vials were added to the trays to hydrate the environment and prevent evaporation through any small leaks in the sample vials, and the trays were then sealed as shown with Teflon tape to minimize evaporation of the water. For experiments lasting long periods of time, these trays were opened and the water vials refilled every couple of months. Once the trays were prepared, they were labeled as illustrated in Figure 5-8 with information about the tray and the date and time at which the tray was scheduled to be removed from the oven.

Removal and Storage of Individual Vials

Individual vials were removed from the oven at specified times and frozen at -80 °C. This allowed analysis of all the samples after the rate experiment had been completed, leading to more accurate results.

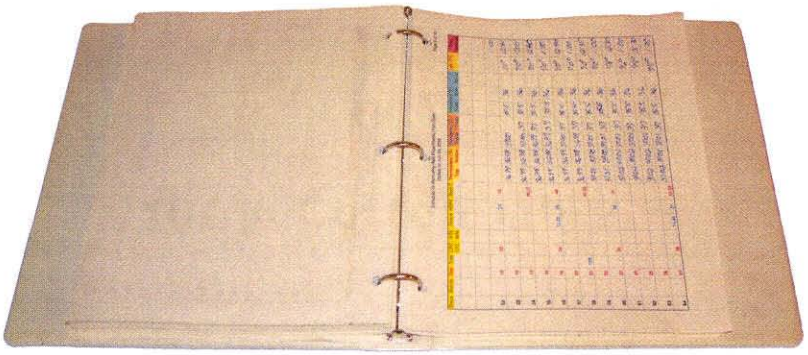


Figure 5-9 – Logbook for Recording Sample Removal Times

A logbook–shown in Figure 5-9–was used to record the actual time at which each sample was removed, the temperature of the ASTM standard digital and glass thermometers, and the computer thermistor readings for the oven and room. The control set point of the oven and room were also recorded along with any adjustments made in the control set point for the oven (removal of trays and corresponding change in air currents required occasional adjustments of about 0.01 °C in the set point). I apologize that this lab book is not bound. Ordinary laboratory recording practices are difficult when taking data on this scale, since everything is organized by computer and most data is taken by computer. We made computer printouts of only essential data and calculations taken during the course of these experiments and still this resulted in over 20,000 pages of laser printouts.

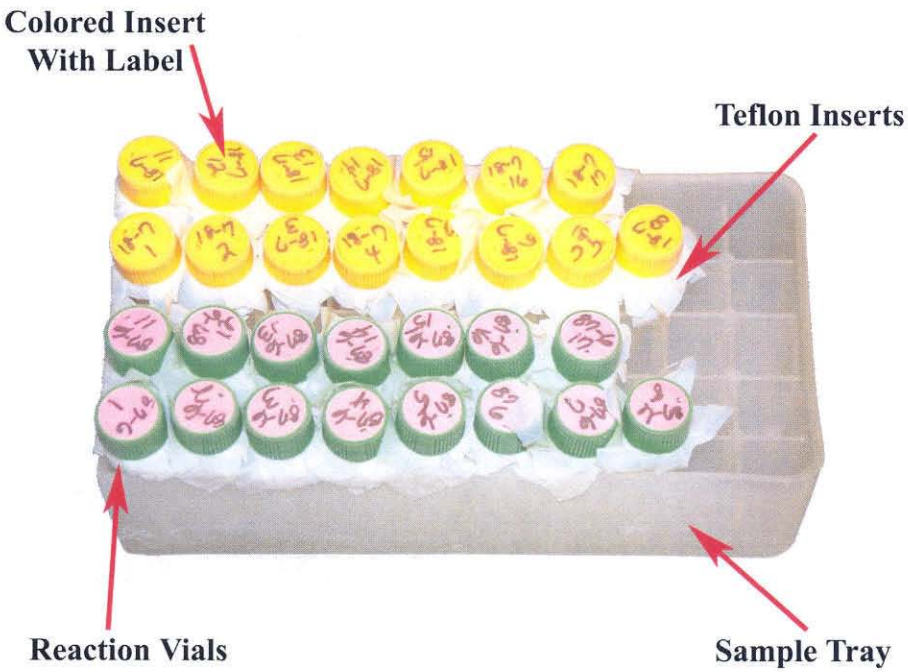


Figure 5-10 – Photograph of Sample Storage Container

Once all the vials in a particular experiment had been removed, they were organized in special containers (Model 235-5000 USA Scientific Plastics) with two rate experiments per container as illustrated in Figure 5-10. Figure 5-10 is a Gln experiment in progress. All of these vials have Teflon inserts, 15 having been removed from the oven. These containers were labeled as shown in the example of Figure 5-11 and stored at -80 °C. They were unthawed briefly to remove a small sample for analysis but have otherwise been kept frozen and are being permanently stored at -80 °C in the boxes shown in Figure 5-11 for future analysis if desired.

Figure 5-11 – Photograph of Sample Storage Box

Preparation of Samples for Analysis

Before analysis, the samples were diluted 500-fold by pipetting 1.5 μl of sample into 749 μl of H_2O in the injection vials. All solutions throughout these experiments were handled in Teflon or polypropylene containers. No glass containers were used except for the syringe pump syringes. The aldolase samples were diluted only 50-fold before analysis; 75-μl injections were used; data were collected for 8 min per sample; and the LCQ automatic gain control was turned off to avoid suppression of detector sensitivity by the large amounts of proteolytic enzyme and aldolase fragments in the sample. The larger samples offset the lesser data collected per peptide, because each of four peptides was being scanned simultaneously.

These diluted samples were paired in boxes like the one shown in Figure 5-12 and are being permanently stored at -80 °C with the original reaction solutions.

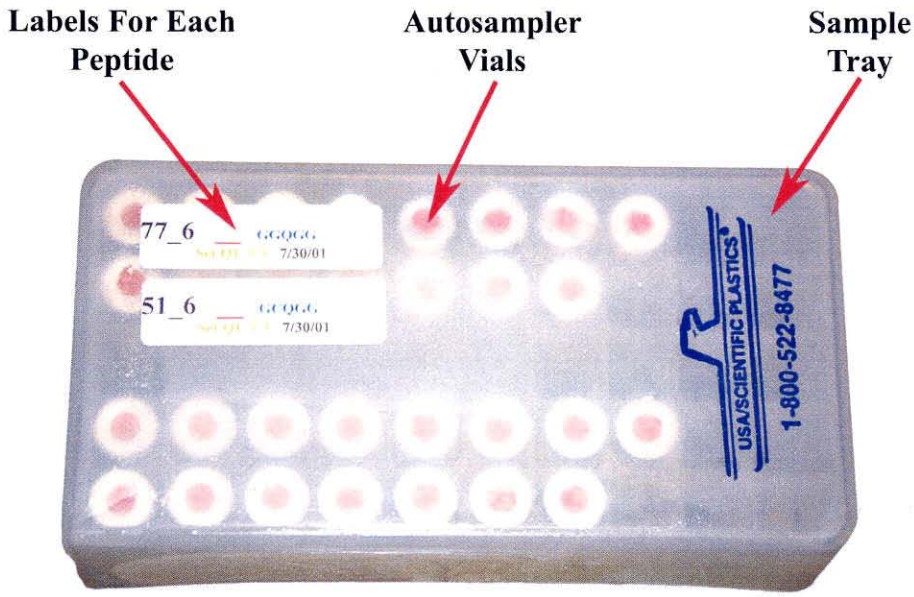


Figure 5-12 – Autosampler Vial Handling and Storage Box

Samples were only thawed out as needed and immediately frozen after preparation. The autosampler was kept at 5 °C over the 4 ½ hour course of the measurements.

For the AsnGly peptides with very short half-times, the samples were kept frozen and only thawed out as needed by the mass spectrometer. This reduced the time any one sample was out during the measurements to about 1 hour at 5 °C. This is probably about the equivalent of 10 min at 37 °C compared to the roughly 0.5–1.5 day half-time and was offset by the time required to come to temperature in the sample oven. No systematic deamidation error was detected in these samples.

The samples were alternated in reaction times over the course of the measurements and were typically done in pairs, with peptides of different masses to wash the system between points. If a single peptide rate were being measured, a set of purified water samples was used in place of another peptide. Thus a typical 18 point rate experiment for peptides A and B with at least a 10 mass unit difference between them would be done in the order: 1A, 1B, 18A, 18B, 2A, 2B, 17A, 17B, 3A, 3B, 16A, 16B, 4A, 4B, 15A, 15B, 5A, 5B, 14A, 14B, 6A, 6B, 13A, 13B, 7A, 7B, 12A, 12B, 8A, 8B, 11A, 11B, 9A, 9B, 10A, and 10B.

Each sample was thoroughly shaken before being placed in the autosampler to guard against separation during freezing. A sharp flick of the wrist drives the sample to the top of the tube and another one drives it down again. Five such cycles were considered sufficient to mix the sample.

A typical 4 ½ hour mass spectrometry run consisted of 36 samples, with 18 from each of 2 peptide rate experiments. We also carried out experiments wherein several peptides of differing mass were simultaneously measured. The throughput of these procedures can be increased by 10- to 20-fold using such combined measurements and/or by combined peptide syntheses of appropriately selected peptide sets. Precision, however, is compromised as the number of peptides rises. As we wished to keep precision error in these experiments below 1%, we did not combine peptides for the experiments reported here, except in the case of the aldolase experiment.

Special Considerations for Mass Spectrometry

Measurements were made in a ThermoQuest LCQ mass spectrometer fitted with a ThermoQuest electrospray source, a ThermoQuest AS3500 autosampler, and a customized sample delivery system driven by a Harvard Apparatus PHD 2000 syringe pump with 5 and 10-ml Hamilton glass syringes. The autosampler had a Teflon-coated needle assembly, a Tefzel valve rotor, and a PEEK valve body. This was done to minimize absorption as much as possible. For the same reason, all sample delivery lines were kept as short as possible and holdup volumes were minimized.

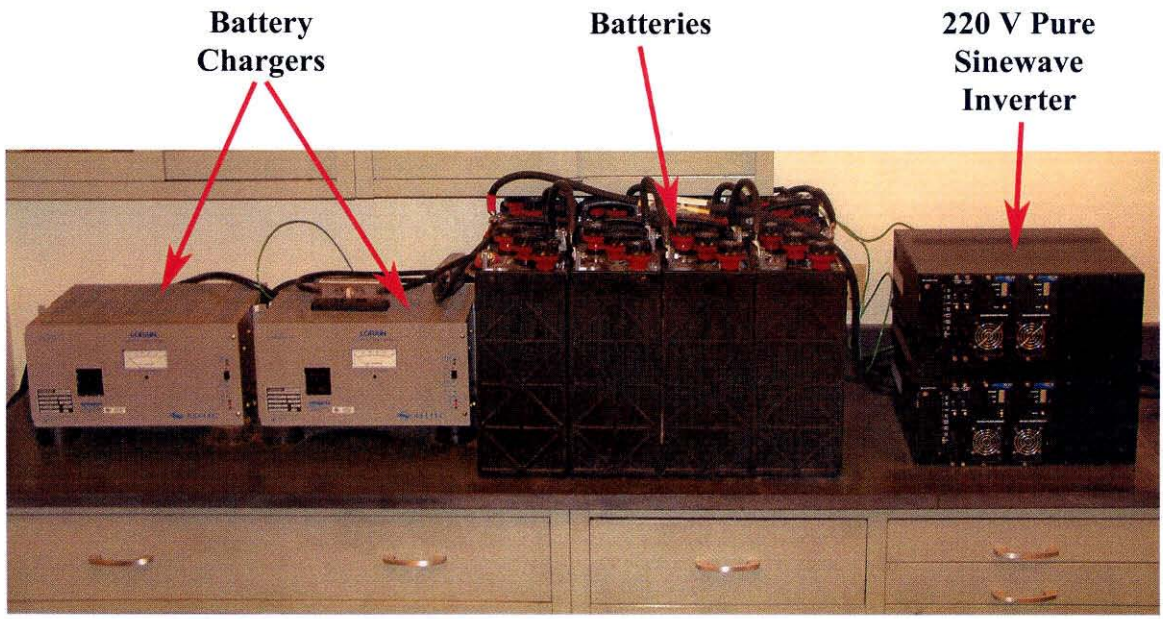


Figure 5-13 – Mass Spectrometer Power System

The LCQ was powered at 240 V by four Exeltech MX1000 power modules connected to a battery bank serviced by two Lorain Flotrol A100F25 rectifiers (Lorain Corporation, Lorain, OH). This system is shown in Figure 5-13. It produced uninterrupted pure sine wave and constant voltage power of much higher quality than that obtainable from the power company.

Design of Mass Spectrometer Sample Introduction Apparatus

The design of the sample handling system was optimized to minimize absorption of the peptide. Differential absorption of deamidated and undeamidated peptide was a special concern. The extra negative charge that is introduced when a peptide deamidates makes differential absorption likely when any absorption is present.

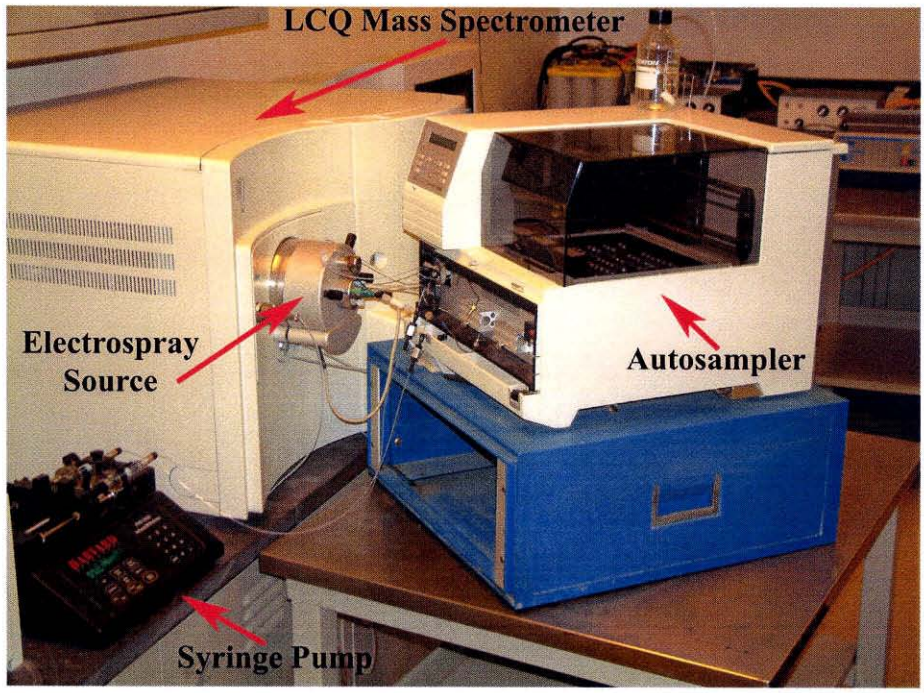
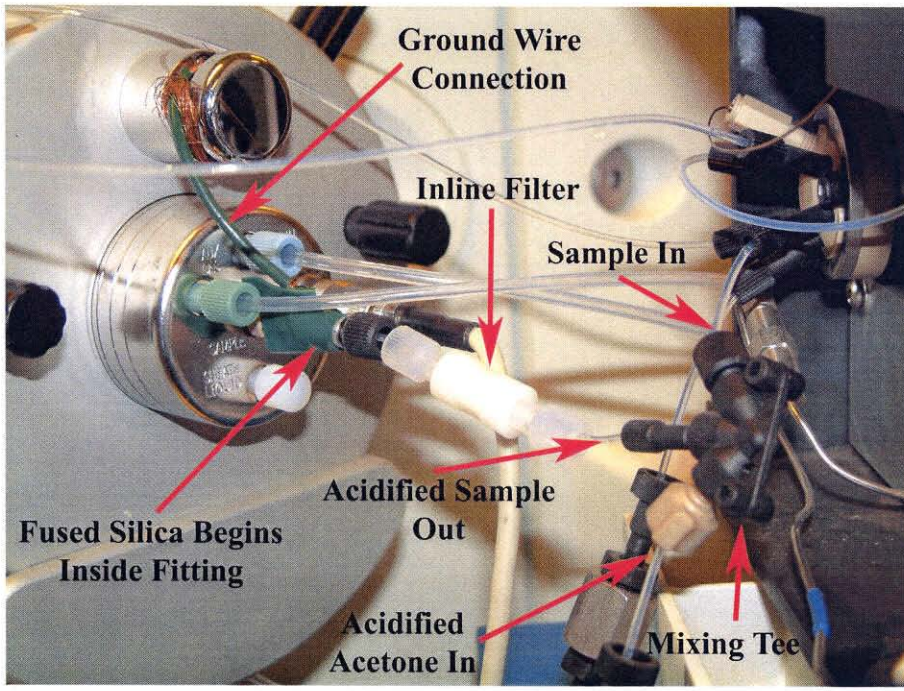


Figure 5-14 – Arrangement of Autosampler with Respect to Mass Spectrometer

Two methods were used to reduce absorption. First, wherever possible, surfaces that contacted the sample were made from materials that minimized absorption. Secondly, the sample lines were kept as short as possible. The autosampler was arranged directly in front of the mass spectrometer source as shown in Figure 5-14. Additionally a special Teflon-coated needle assembly, a Tefzel valve rotor and a PEEK valve body were installed in the autosampler. The line to the mass spectrometer source was made from Tefzel and the deactivated fused silica line that is part of the mass spectrometer source was kept to about 3 inches. The typical length used for a mass spectrometer of this type is about 24 inches.



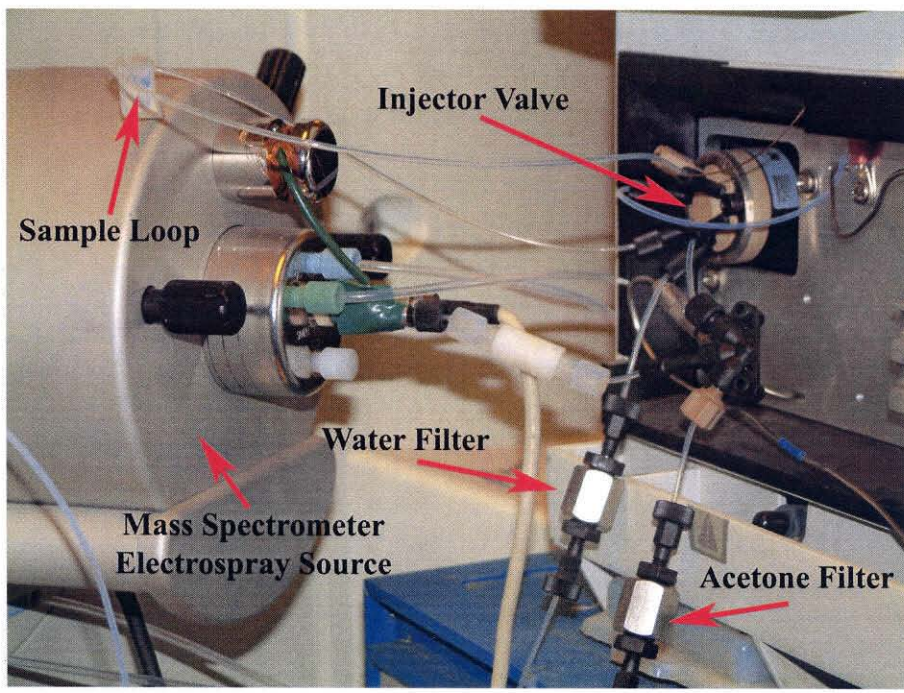


Figure 5-15 – Close-up Photographs of Sample Introduction Apparatus

Photographs of the sample introduction apparatus are shown in Figure 5-15. Only the lines shown exiting the autosampler valve and traveling from the mixing tee to the mass spectrometer were kept short. These are the lines that handle the sample, where the risk of absorption is highest. The Teflon-coated needle assembly and PEEK valve body shown were the best we could arrange given the design of the autosampler. The PEEK valve is better than stainless steel but still not perfect. A small amount of tailing is coming from this valve.

The sample loop shown is a 100 μL loop made from 0.020" ID Teflon. The individual injections were 50 μL. It was critical that the sample travel from the sample vial to the mass spectrometer without tailing and mixing significantly with the liquid column that drives it. This was achieved by drawing the sample into the loop, while leaving a 25 μL air bubble on either side. This air separation remains through the mixing tee and all the way to the mass spectrometer. The auto sampler is not normally meant to work this way, but a manual for the low level firmware settings was obtained from ThermoQuest and the appropriate modifications were made.

The lines coming from the syringe pump are also made from Teflon. The filters–shown in Figure 5-15–are for removing small particulate matter from the solvents. They are Upchurch Scientific model A-431 0.5 micron filters.

The 5 ml syringe contained 1.5% acetic acid in acetone, and the 10 ml syringe contained purified H_2O. The diameters of the two syringes (Hamilton models 81520 (5.0 ml) and 81620 (10.0 ml)) were such that the 10 ml syringe pumped at exactly twice the rate of the 5-ml syringe. The combined flow rate was 40 μL/min In the case of proteins and larger peptides, H_2O was substituted for acetone in the injection system to avoid precipitation.

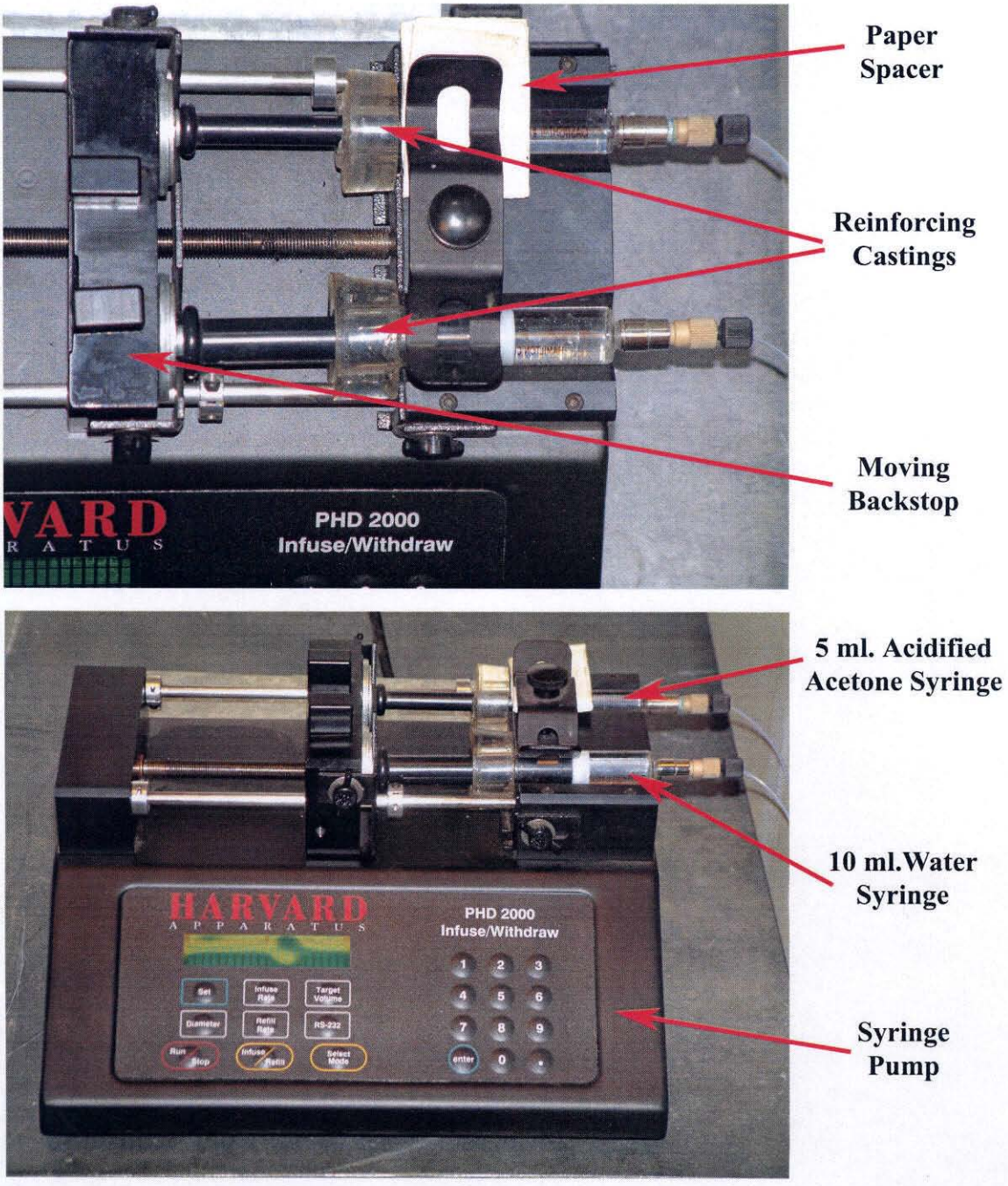


Figure 5-16 – Syringe Pump With Syringes Loaded

The syringe pump is a Harvard Apparatus model PHD 2000. The photograph in Figure 5-16 shows it with the syringes loaded. The smaller diameter of the 5-ml syringe requires the paper spacer shown to apply pressure from the hold down clamp. This is adequate for this system, but if the plastic castings around the syringe are not uniform a lead weight is needed as well. The syringes shown here were also specially modified for use in our system.

It is essential to maintain a steady flow of liquid from the syringes with no bubbles or hesitations from either syringe. Not only would this cause the mass spectrometer sensitivity to oscillate, but a brief cessation of flow from either syringe would cause the other one to backup through the mixing tee, changing the concentrations of everything and generally causing trouble. A steady flow is difficult to maintain reliably at these low flow rates with syringes this large, since the linear velocity of the plunger is only 0.16 mm/minute. Syringes normally used with an LCQ are about 20 times smaller, so this is not a serious problem. We needed, however, to be able to run for at least 4 ½ hours without stopping.

Only glass or stainless steel syringes are suitable. Plastic syringes bend as they are depressed and can lead to oscillations in the flow rate. We selected glass ones as likely to be the most innocuous, despite their tendency to bleed sodium into the system. Air bubbles were carefully removed before loading and the backstop was gently moved up tight against the syringes and then carefully nestled into the driving screw to force a small amount of liquid out of both syringes, while maintaining the pressure. This causes both syringes to start positively at the same time. The tubing is then reconnected and allowed to run for about 10 min to remove any air in the lines before starting the analyses.

The backpressure generated by the flow system is beyond the specifications for these glass syringes. Therefore, the glass ears that hold the syringes in the pump break off. The solution to this problem can be seen in Figure 5-16. The lower portion of the syringe body up to and including the ears is incased in plastic to distribute the force more evenly on the syringe body. This was done with liquid plastic casting resin, obtainable from a hobby store. Care must be taken with pouring this as it could contaminate the mass spectrometer if

it came in contact with the inside of the syringe. The pump drive and syringes in Figure 5-16 are also shown equipped with grippers to allow the syringes to refill automatically by reversing direction. There were several complications with this automatic refilling, however, and it was only of marginal value, so this feature was never implemented.

The line from the 10-ml syringe travels directly into the injector valve. Its purpose is to drive the sample into the mass spectrometer and provide the baseline when no sample is present. The H_2O is never mixed with the sample. After exiting the injector valve, this line continues on to one side of the tee carrying the sample.

The flow from the 5-ml syringe is directly into the other input port of the mixing tee and contains 1.5% acetic acid in acetone. This is to acidify the samples before they enter the mass spectrometer. This procedure allows the sample to be analyzed within 60 seconds of acidification and reduces the possibility of side reactions or accelerated deamidation caused by low pH. The acetone improves the volatility of the solution and therefore increases ionization of the peptide, which improves the analysis.

The mixing tee is an Upchurch model U-466 and is designed for efficient and complete mixing of the two samples.

The final step before introduction into the mass spectrometer is filtration by the Upchurch A-419, 2 micron filter installed in the fitting shown and an additional Upchurch C-409 micro filter installed in the fitting which follows. Without filtering at these points, the surface deactivated-fused silica line which follows would need to be replaced every few sample runs, which would be difficult and time consuming.

The surface deactivated-fused silica line itself is not shown in the figure. The ordinary setup for the LCQ has about 24 inches of this tubing which comes out of the back of the electrospray source with plenty of excess tubing before the Teflon line starts. This allows easy adjustment of the position with respect to the electrospray tip shown in Figure 5-17. To minimize the length of the tubing a different procedure was adopted.

Using a combination of fittings, the Teflon line was mounted directly to the back of the mass spectrometer source as shown in Figure 5-15. The fused silica line runs only through the source itself from the electrospray tip to the fitting in the back of the source. This means that the end of the fused silica is fixed at the back of the electrospray source and therefore that the length of the tubing determines its position with respect to the electrospray tip (ideally a millimeter or so inside the tip).

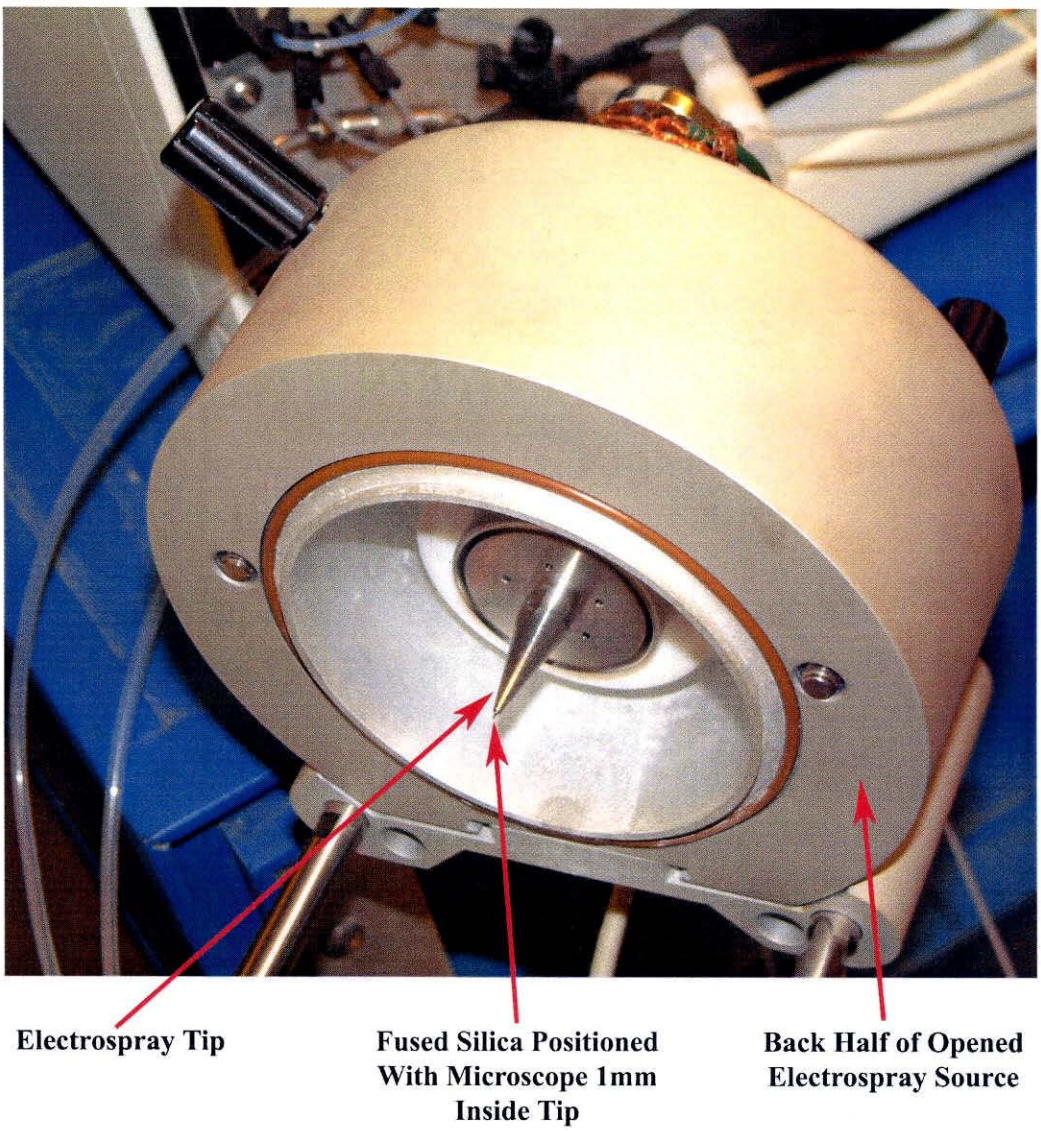


Figure 5-17 – Close-up of Electrospray Source

A microscope was used to make a clean cut on the end of the fused silica tubing and position it precisely with respect to the tip before cutting it to length. With experience, this procedure could be completed in less than an hour. This tubing was replaced about every 50 sample set runs.

With these essential modifications and precautions, the electrospray source was otherwise the standard source shown in Figure 5-17.

Mass Spectrometric Data Collection Procedures

In order to assure that the mass spectrometer sensitivity remained high, a standard sample was run periodically. The standard we used was LeuArgArgAlaAlaLeuGly. Generally this standard was run each morning before starting, though in same cases, especially when we ran overnight without stopping, this pattern was not strictly followed. This type of standard is essential for keeping track of the machine. Various problems can occur to damage the mass spectrometer sensitivity. One particular problem was the heated capillary in the LCQ. The original design called for a disposable and easily replaceable heated capillary tube that could be changed regularly (optimally every day). The manufacturer changed this into a $500 rhodium plated unit with the heater built into it. The result is that the heated capillary gradually degrades with time and is expensive and difficult to change. It is necessary to watch this closely, so that it can be replaced when needed. Substantial loss of signal can cause the mass spectrometer to drift out of the quantitatively linear range of 10^{-6}–10^{-8} M at these flow rates.

After mixing, the solution flowed into the mass spectrometer source at 40 μl/minute. It was 0.5% acetic acid, 33% acetone and 66.5% H_2O. Each run consisted of a 50-μl loop injection of sample, which was monitored continuously for 5–7 min by a sequence of 4 high-resolution zoom scans centered on the expected charge/mass ratios and one 100–2,000 charge/mass low-resolution scan. Approximately two zoom scans per microliter of sample were recorded.

The temperature of the mass spectrometer's heated capillary was held at 180 °C. This essentially eliminated sodium adducts, markedly improving the quantitative analysis. Significant amounts of sodium adduct could preferentially remove one species from the measurement and compromise the rate experiments. There is a side effect of this temperature. Tris buffer (which is still present in small quantities in each sample after dilution) tends to form aggregates which are not broken up in the heated capillary at this low temperature. A capillary temperature of 260 °C is necessary to eliminate the Tris aggregates as shown in Figure 5-18. Multiple Tris aggregates with many different charges can be clearly seen in the total mass spectrum for the run. The time in the LCQ ion trap prior to the zoom scan, however, allows these aggregates to break apart so they do not interfere with the quantitative results.

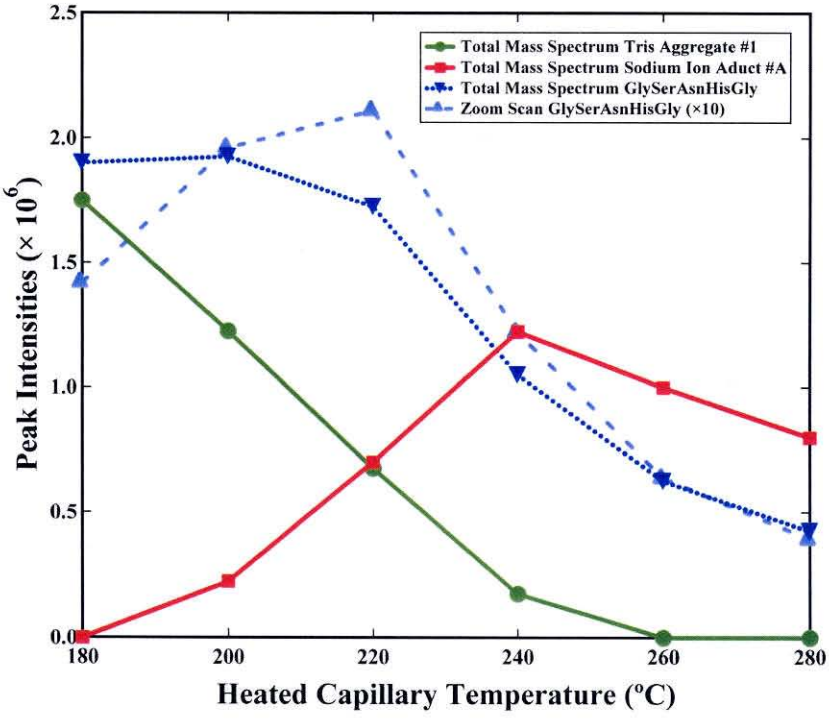


Figure 5-18 – Heated Capillary Temperature vs. Sodium and Tris Adducts (Individual curves are scaled by mass-dependent sensitivities)

A system is necessary to assure that accuracy is maintained. There are multiple steps in this sequence that could cause differential loss of peptide and therefore inaccuracy in the results. Possible places for this to occur are in the reaction vials, on pipette tips, in the measurement vials, in any of the tubing or fittings in the injection system, differential ionization in the electrospray source, or even by differential sensitivity of the mass spectrometer detector. The effect from the detector is probably not too great, since the comparison is being made between two similar substances that are only 1 mass unit different. The precautions described above are meant to minimize these problems. Most mass spectrometers are used only qualitatively or semi-quantitatively. The problem is that deamidation causes a change in charge, so almost any type of absorption or loss is likely to occur differently for the two forms.

An important safety precaution is provided by our large data set. By taking 18 points, we not only assure accuracy, but we can see differential effects. Almost any type of differential effect will cause the first order reaction plot to curve instead of form a straight line. Our plots were perfectly straight lines for peptides with short half-times. A small amount of curvature does start to appear at longer half-times, though not enough to bias our rates significantly.

Each sample was scanned by the mass spectrometer for 5–7 min with an alteration of one low resolution scan from 100–2000 mass units, and four high resolution zoom scans 10 mass units wide and centered on the first isotope peak. The 10 mass unit window, shown in Figure 5-19, includes the first peak and the following 4 isotope peaks to the right of the first peak. It also includes 4 peaks to the left of the main peak, which can be used to estimate and correct for the baseline noise level. Typically about 100 zoom scans are taken over the course of each sample measurement.

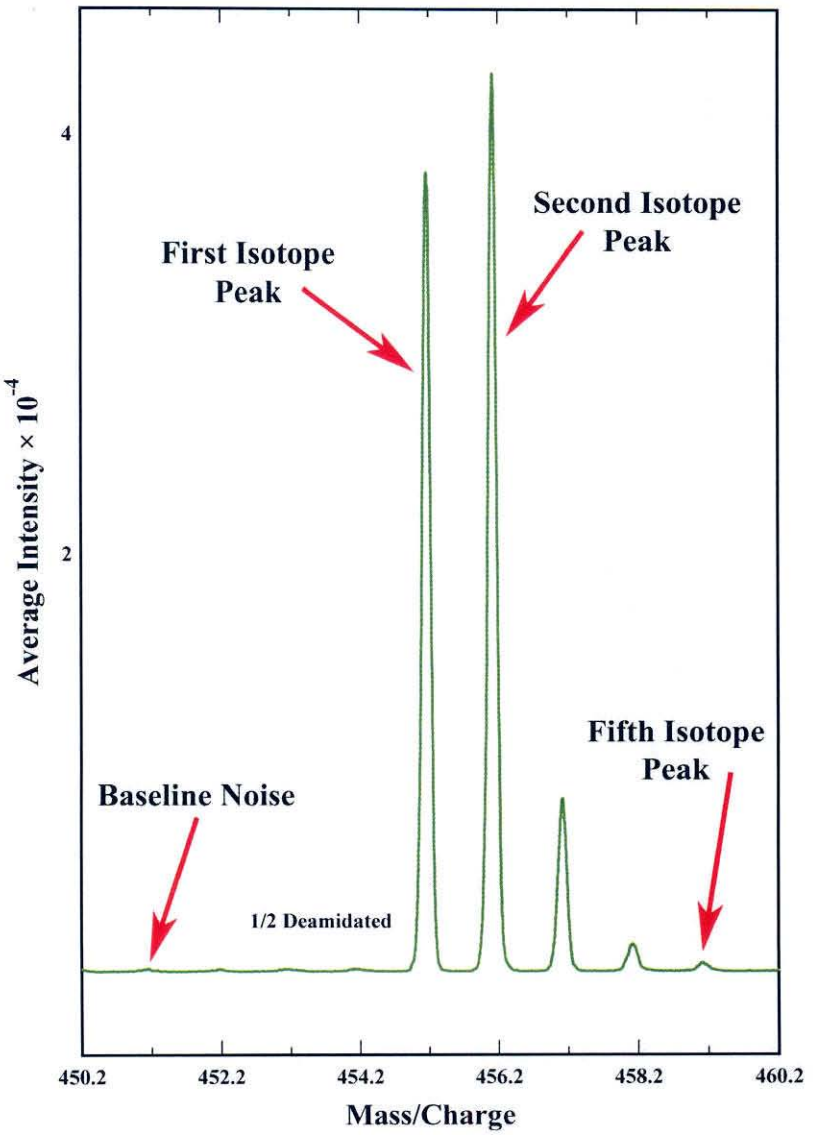


Figure 5-19 – Sample Averaged Zoom Scans from the ½ Deamidated Sample of GlyAlaAsnAlaGly

The ratios of these isotope peaks change over the course of the experiment as peptide deamidates and shifts the mass one unit to the right. This is illustrated in Figure 5-20.

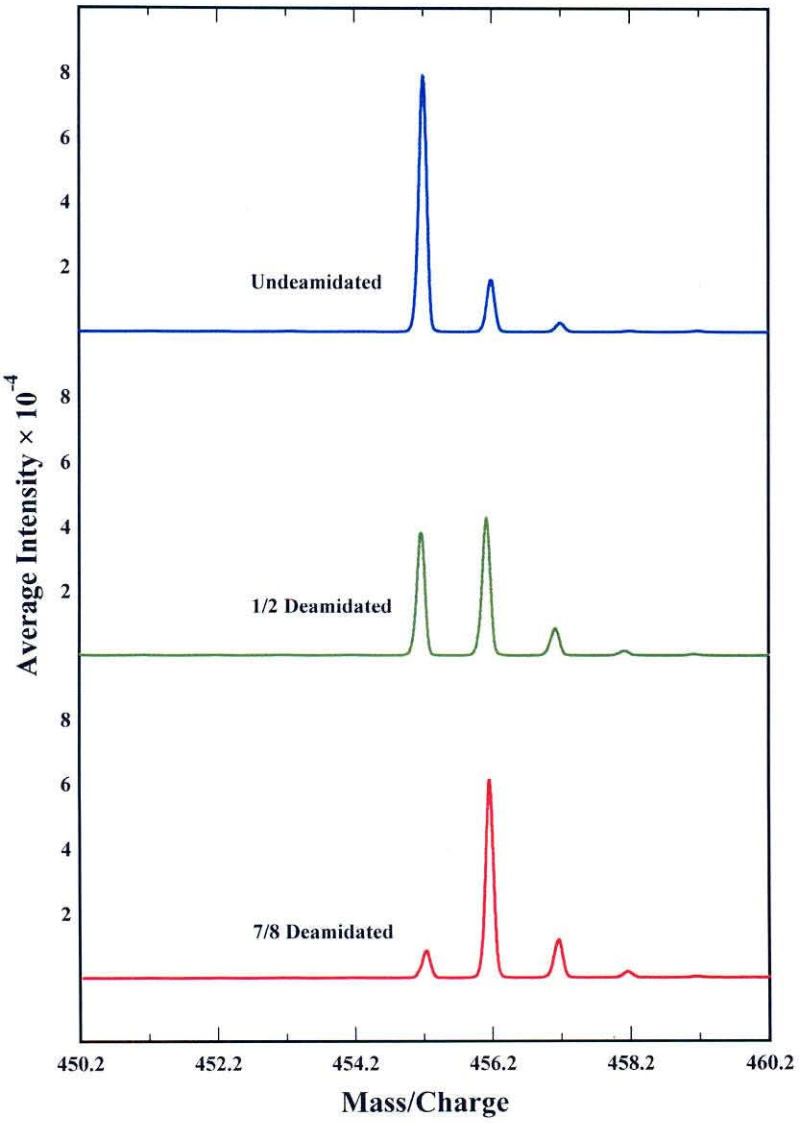


Figure 5-20 – Change in Peak Intensities as Deamidation of GlyAlaAsnAlaGly Progresses

Figure 5-21 shows the total zoom scan ion current over the course of a typical sample run. The roughly square part of the curve is the sample plug traveling through the mass spectrometer.

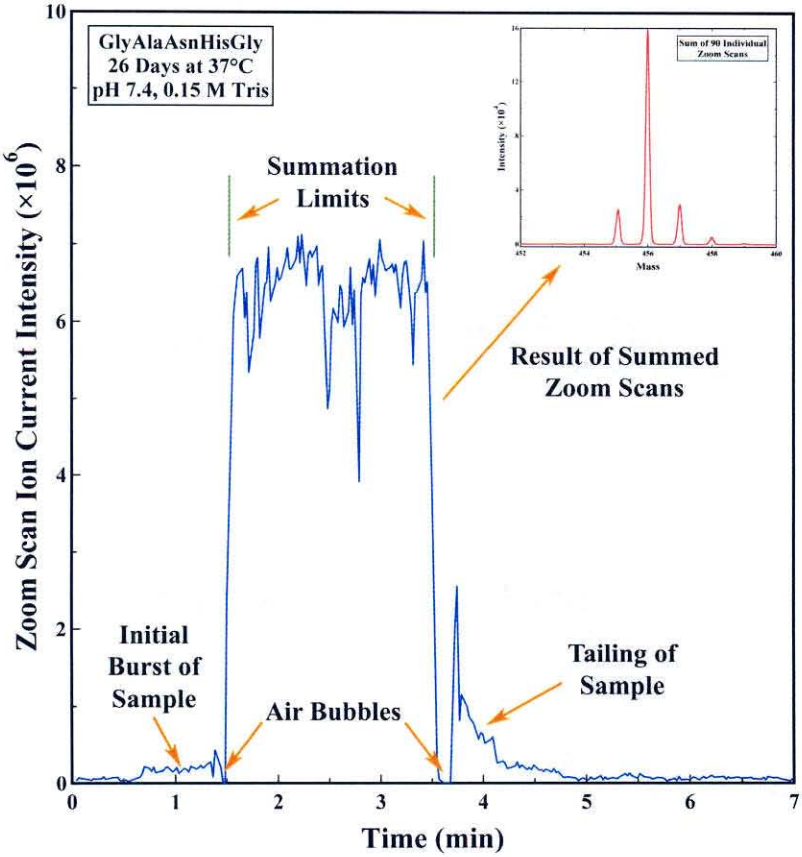


Figure 5-21 – Total Zoom Scan Ion Current as a Function of Time for GlyAlaAsnHisGly after 26 Days of Deamidation

The left part of the curve is taken before the sample reaches the mass spectrometer. At this point we are measuring the signal from the water syringe. When the sample actually reaches the mass spectrometer, there is essentially no water (only a small amount from contamination of the lines) from the syringe in the sample. The left air bubble is a short period with no signal. This occurs because of the separation air bubble inserted by the autosampler. Notice that there is a brief burst of signal just before this air bubble. This comes from a combination of diffusion past the bubble and small amounts of sample caught in the injector valve and placed in front of the air bubble when the valve turns.

To the right of the main signal block, there is a second air bubble. This is the air bubble which comes after the sample plug. In this case there is a much large signal which follows the air plug. Some of this is sample caught in the injector valve, but most of it is due to tailing in the sample lines which gradually tapers off as shown. While this was minimized with careful choice of line materials and construction, the amount of tailing shown is typical of these experiments.

When absorption in sample lines occurs, it nearly always occurs differentially due to the difference in charge between the amidated and deamidated forms of the peptide. In the case of some very long and sticky peptides, this can be seen by looking at different samples taken over the course of the experiment. This is illustrated in Figure 5-22. Chromatography is being preformed on the sample by the small amount of absorbing surfaces still left in our system. This only occurred for a few peptides.

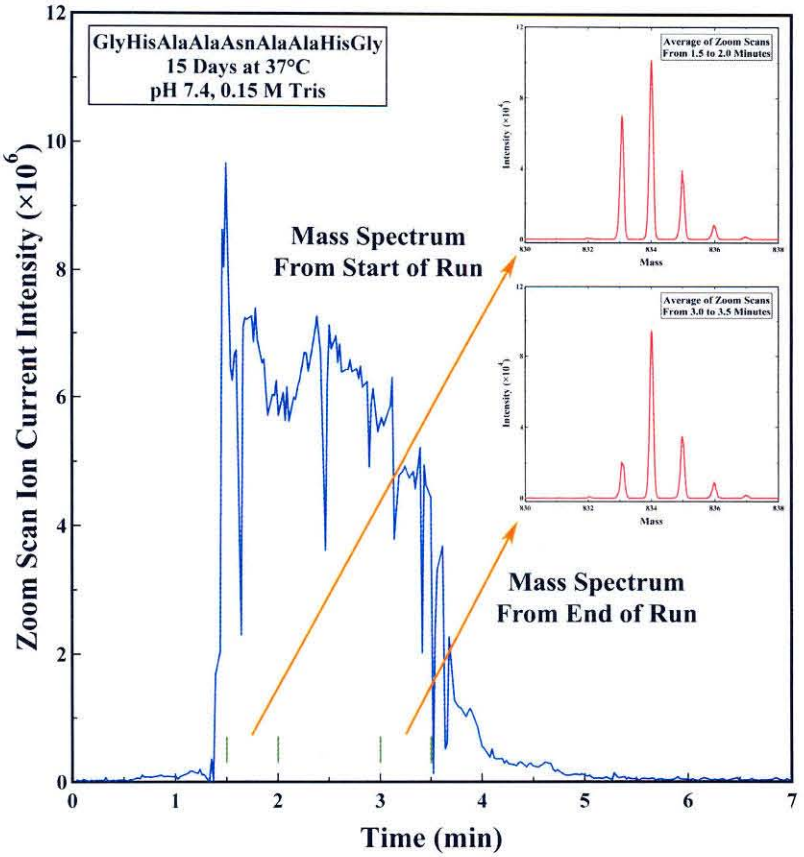


Figure 5-22 – Example of a Badly Tailing Sample. Fifteen Days of Deamidation for GlyHisAlaAlaAsnAlaAlaHisGly

Absorption of peptide is not, however, the only reason for tailing. Any cracks between fittings or other regions where liquid is flowing at a reduced velocity will cause tailing but not chromatography. These regions exchange material with the main liquid column by diffusion. Over the course of the 5-7 minute loop injection they become saturated with sample and take time to be replaced with water. Every piece of tubing in this system was cut carefully square and inserted to the bottom of the fitting to minimize this. Not all holdup cracks were removed, however, and holdup volume in the injector valve as well as small amounts in the other fittings is still present.

Figure 5-21 shows two lines marking the limits over which the sample was summed and the inset shows the resulting mass spectrum. It is important to sum the mass spectra over as wide a range as possible. Inclusion of all data removes artifacts that might arise from separation due to tailing. At the same time, including regions of very low intensity introduces noise into the system and possibly even systematic error, since the linearity of the mass spectrometer response may be compromised at very low concentrations. The limits selected were a compromise between these two factors and were chosen by hand for each sample depending on the shape of the curve and any change in composition with time. Cases like the one shown in Figure 5-22 are rare, so the care taken was primarily only a precaution. A macro was written to bring each curve up on the screen automatically in the ThermoQuest Xcalibur software. The spectra within the chosen limits, with the same number of zoom scans included in each range for consistency, were automatically averaged and saved to an Excel sheet.

Procedures for Handling Data

At the time of peptide synthesis, each peptide was given a number according to the synthesis block in which it was made and the number in the block. Thus the number 26_4 would be the 26th sample in the 4th reaction block. These numbers were carefully maintained throughout the purification procedures and all the way through the mass spectrometer analysis. To provide redundancy, the following convention was used in naming all output from the mass spectrometer. The sequence was substituted for the underscore in 26_4. Thus, since 26_4 was the peptide GlyLeuAsnSerGly, the main directory and calculation sheet names were 4GLNSG26 and the individual sample results were simply given the names GLNSG_1, GLNSG_2 etc.

With the additional check that the masses would very likely be wrong if there were any mix-ups and since the mass was calculated based on the directory name, the possibility of confusing samples was essentially eliminated.

In the entire course of these experiments, with more than 500 rates measured from individual peptides which included programming the synthesizer for more than 1,000 peptides, putting correct resin individually in each reaction well, adding correct reaction reagents to the reagent vessels, placing peptide derivatives in the correct vessels (with no error checking other than watching each transfer), transfer of the peptides from the synthesizer block to labeled precipitation tubes, transfer from precipitation tubes to 5 labeled freezer vials which were freeze dried before use (this expanded the freeze drying and hand labeling of freezer vials to about 5,000 tubes), transfer from one of these vials to a 2 ml tube with the correct buffer and concentration for the rate experiments, transfer from this tube to more than 10,000 individual tubes for the reaction oven, removal of each tube and placing it in a freezer at the correct time, resorting these tubes by hand into individual rate experiments, transfer of a sample from these tubes to a sample tube at the correct concentration for the autosampler, placement of these tubes in the correct positions in the autosampler, programming of the autosampler for the correct samples in experiment pairs, programming of the mass spectrometer for the correct mass from a value calculated using the original Excel sheet telling which sample goes in which well of the synthesizer, output of the mass spectrometer data into individual files (file names were entered separately for each sample), transfer from the original mass spectrometer output to a correctly selected and summed zoom scan file in Excel, and finally selection of the correct Excel sheet for each sample by the Mathcad sheet used in the calculations as well as entry of the correct reaction time for each point, *there was never a single instance of samples being mixed up or a discrepancy in expected mass*. For this reason we are fairly confident of our procedure and certain that no major errors were made. The essential computer output from these experiments is more than 20,000 pages and fills three filing cabinets. Two of these are shown in Figure 5-23.

Figure 5-23 – Filing Cabinets Containing Data

Calculation Programs

As previously mentioned, all data were extracted to Microsoft Excel sheets. The point-by-point values of these curves were then corrected for baseline noise, adjusted for the naturally occurring isotopic ratios, and separated into the contributions from the undeamidated and deamidated peptides, which differ from each other in that deamidation increases the mass by 1 atomic mass unit. The rates of deamidation were then calculated and plotted as first-order rate curves like the one shown in Figure 5-24.

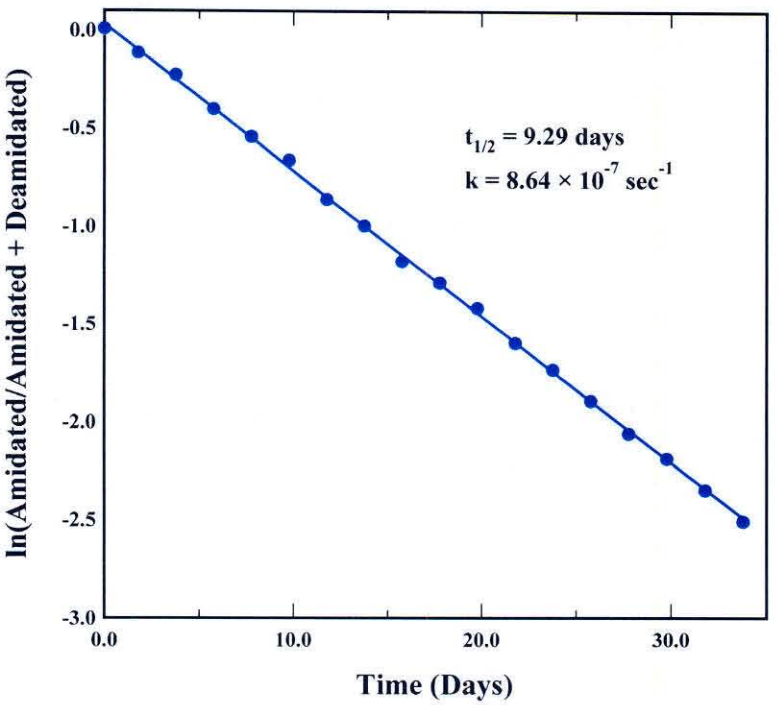


Figure 5-24 – Sample Rate Plot. The Peptide is GlyAlaAsnHisGly.

These calculations were performed in MATHCAD 8 PROFESSIONAL (Math Soft, Cambridge, MA) using an extensive Mathcad sheet written specifically for this purpose. This sheet is 118 pages long and displays graphs of each individual point, a graph of all points overlaid, and the final results with plots. It performs several other functions as well.

First the exact mass is calculated, along with the expected ratios of isotope peaks based on the reported naturally occurring isotope abundances.[92] We were concerned about whether isotope ratios vary from place to place on the earth. If, however, the deamidation rate is based on reduction of the first peak (which is 100% the most abundant isotopes) compared to all of the others, this can only affect the intercept of the curve and not the rate. Most of our plots approximately intercept the zero deamidation

point, so we are confident in the accuracy of the isotope ratios and the lack of significant deamidation during sample preparation.

Data for each point are examined by special routines to insure that correct data are present. The mass is shifted slightly as necessary to make all of the spectra overlay perfectly at the correct mass. The mass calibration on the mass spectrometer was good. A typical shift was about 0.1 mass units. The peak positions are then automatically determined and the peaks are integrated. Expanded plots for each point are made showing exactly were the limits for each peak are positioned. A correction for background noise is made to each peak. This is a small correction and is based on hand selected positions to the left of the first peptide peak shown in Figure 5-19. An additional small correction was made for data beyond the fifth peak shown. Data are then plotted and a straight line fit through the points. A curve is also fitted as a check against second order effects and differential detection of peptide. These data were also fit separately to the points from Teflon inserted vials and polypropylene vials. Calculations were made with a variety of other peak combinations as well, but the standard curve with all points based on reduction of the first peak compared to all the others proved to be the best. It is these values that have been reported.

Entering all of the data into the Mathcad sheet and waiting for the results is time consuming and would need to be repeated as the calculation procedure was improved. A macro was therefore written in MACRO EXPRESS 2000 (Insight Software Solutions, Bountiful, VT) to calculate the entire set automatically. This required more than 50 pages of macro instructions. Mathcad proved to be unstable and crashed occasionally during use with a sheet this extensive. The macro, therefore, not only had to do all the entry, but it contained various checks and safeties. It checked that Mathcad was still running and that the job was completed properly at various stages. If there were any difficulties, it would restart Mathcad and repeat the calculation. If it ran into serious trouble, it stopped the program to prevent havoc in the computer. It repeatedly checked which program was running at each point to do all of this. It was also responsible for entering all of the data

about the peptide from Excel sheets–which included mass spectrometer data, reaction time data, noise correction peak data, sequence data, and any other data needed–as well as the time and date of each calculation along with a unique calculation number and any special comments.

In the course of this work, we have measured and calculated about 500 deamidation rates involving 9,000 loop injections and requiring a total of about 1,000 hours of mass spectrometer operation. With the aid of these computer programs, each 4 hours of mass spectrometer operation requires about 30 min of manual computer manipulations to calculate the results, largely because these programs are not integrated into the LCQ software and rely on the procedure described above.

Chapter 6

CONDITIONS FOR DEAMIDATION

Selection of pH and Temperature

It was essential that the best reaction conditions available be selected before investing the effort required to make the measurements. In this we were assisted by the rate experiments already present in the literature. The choice of temperature was clear. For this we used 37.00 °C. This is a well-established temperature for physiologically interesting experiments.

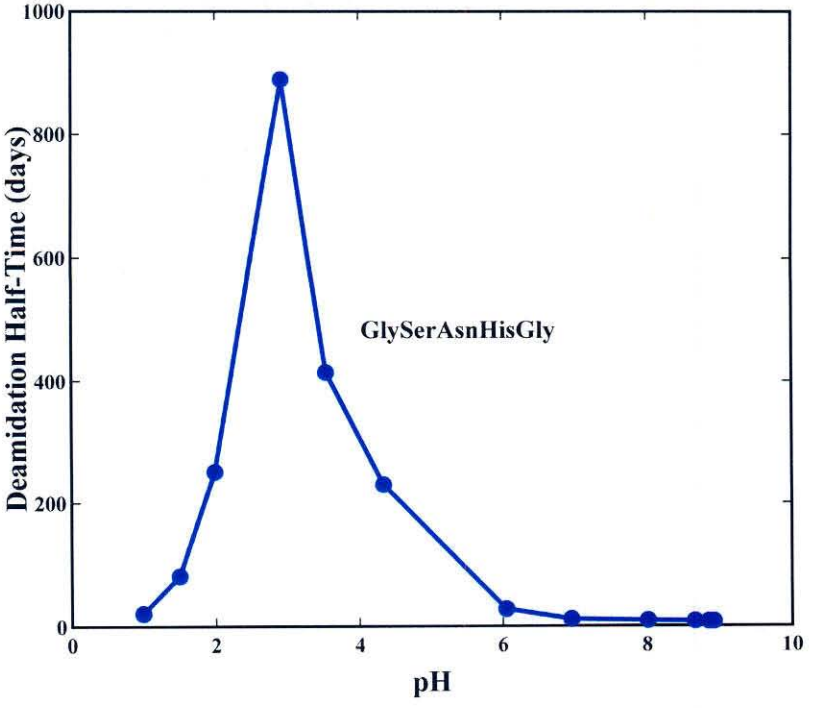


Figure 6-1 – pH Curve for GlySerAsnHisGly in 37.00 °C, 0.15 M Tris Buffer

A pH of 7.40 was chosen. This is approximately the pH of blood. This is a critical parameter, since reaction rate depends so strongly on pH. Figure 6-1 shows a pH curve for the peptide GlySerAsnHisGly. This clearly shows the dependence on pH. The reaction rate actually decreases as the pH is lowered, until direct hydrolysis takes over and the rate is sharply accelerated.

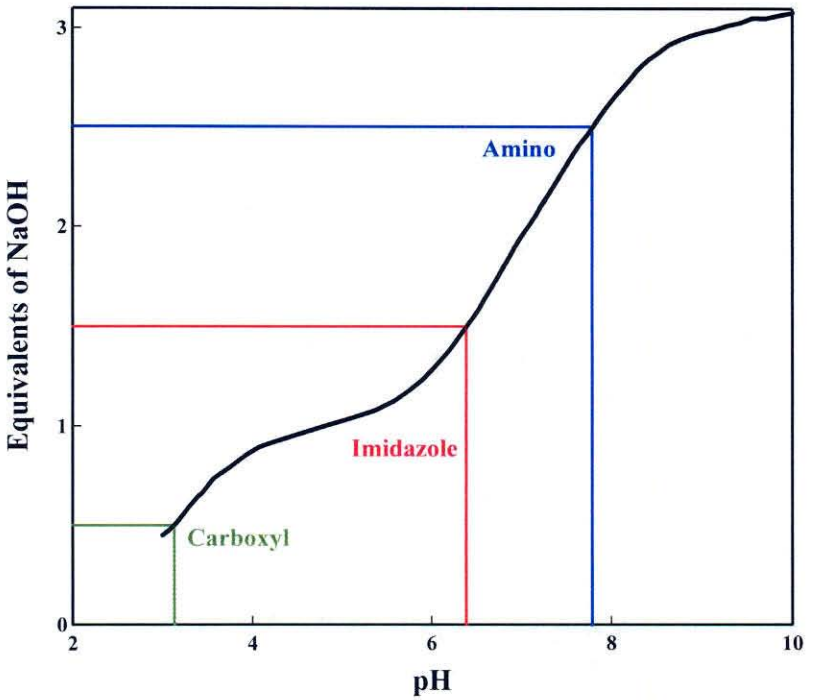


Figure 6-2 – Titration of GlySerAsnHisGly

The titration curve that we measured[73] for this peptide is shown in Figure 6-2. It was found that the pK values were 3.1, 6.4, and 7.8 for the c-terminal, imidazole, and n-terminal groups, respectively. Thus at neutral pH this peptide does not consist of a single species. It is likely that, for AsnHis peptides, one form deamidates at a different rate than the other.

Selection of Buffer Type and Ionic Strength

Selection of buffer type and ionic strength was a more difficult problem. Previous experiments by investigators had been done under a wide variety of conditions, but 0.15 M sodium phosphate and potassium phosphate were the most common.[3] It has been shown however that phosphate markedly accelerates the rate of deamidation. Figure 6-3 shows this for two Gln peptides.[72]

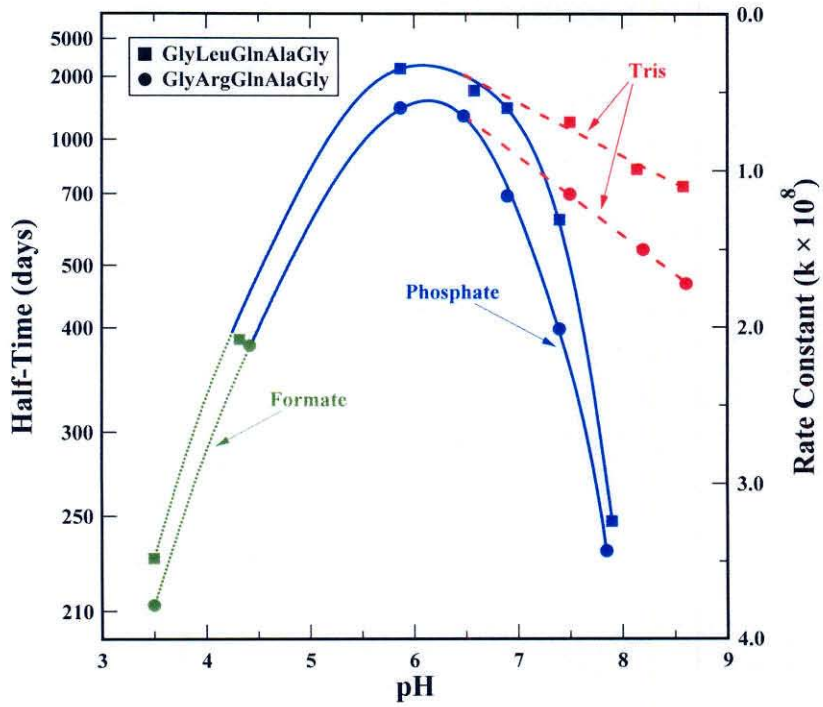


Figure 6-3 – Buffer and pH Dependence of the Peptides GlyLeuGlnAlaGly and GlyArgGlnAlaGly. Reactions were carried out at 37.0 °C, ionic strength 0.2, and were 0.001 M in peptide. *Figure adapted from reference 72.*

Ionic strength can be adjusted by addition of a variety of ions to the solution, but high ionic strength has been shown to accelerate deamidation[3, 74] as is shown in Figure 6-4. It was therefore decided to let the ionic strength be determined by the buffer concentration. Anything that modifies the rates beyond their fundamental values is undesirable in a basic set of rates.

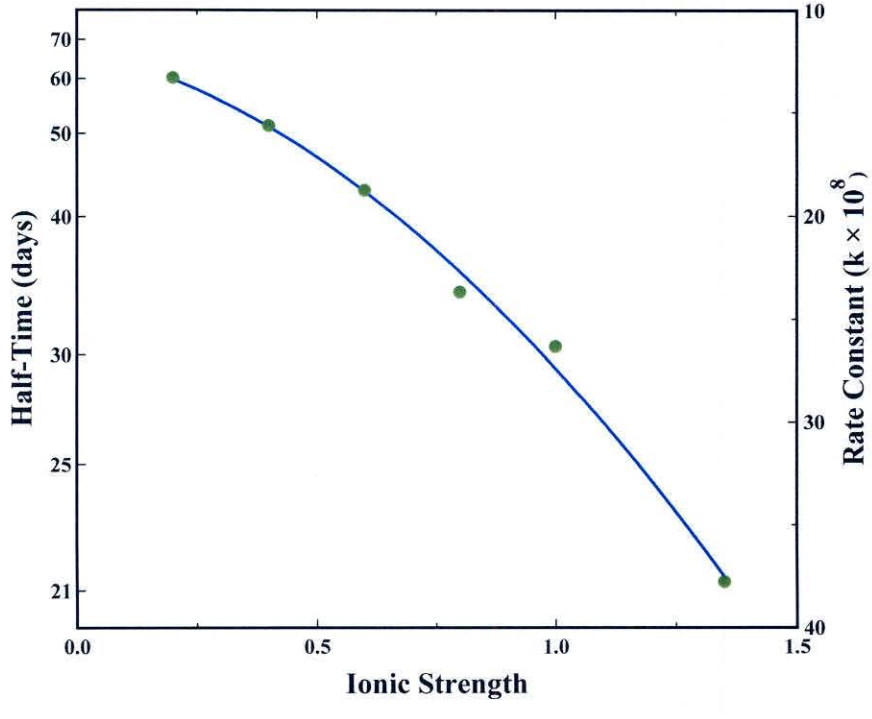


Figure 6-4 – Ionic Strength Dependence on the Deamidation of GlyArgAsnArgGly. The reactions were carried out in pH 10, 37 °C buffer.
Figure adapted from reference 74.

Several buffer types were considered. Tris buffer was selected with a concentration of 0.15 M because it has a minimal catalytic effect on deamidation. Phosphate catalysis has been shown to substantially accelerate deamidation.[3] Tris is still not perfect. Extrapolations to zero buffer concentration have shown that the 60-day half-time of an Asn residue in ribonuclease is accelerated by Tris.[78]

Selection of Peptide Concentration

Peptide concentrations were kept at 0.001 M as was used in the past.[3] This is low enough to minimize intermolecular effects, yet high enough to avoid serious absorption problems. With this peptide concentration, a dilution of a factor of 750 was possible resulting in only 2×10^{-4} M Tris in the solution injected into the mass spectrometer.

Chapter 7

EXPERIMENTAL RESULTS

Deamidation Rates for Asparaginyl Pentapeptides

The experimental data are given in Tables 7-1 through 7-8. Every rate reported in this thesis was carried out individually except for the cysteine and blocked cysteine peptides which were done in the same solution and the simultaneous experiments on rabbit muscle aldolase. All reactions were in 0.15 M, pH 7.4, Tris buffer at 37.00 °C ± 0.02 °C.

Table 7-1 gives the deamidation rates for all asparaginyl containing pentapeptides that we measured. The four values not shown in bold are the only values that were not directly measured. These four were estimated from the deamidation rates of the others. In the case of the singly blocked cysteine peptides, these two values could not be distinguished because they were measured in the same solution.

There are some obvious trends in these rates. It is clear that the right-hand residue is even more important in determining the deamidation rate than was previously known. The columns in the table have been ordered by median half-time of the right-hand residue. The rows are also ordered by the left-hand residue. While it is clear that the left-hand residue has a significant effect on the rate, this effect is much less than the right-hand residue.

Table 7-1 First-Order Deamidation Half-Times of GlyXxxAsnYyyGly in days at pH 7.4, 37.0 °C, 0.15 M Tris HCl

Xxx\Yyy	Gly	His	Ser	Ala	Asp	AmCys	Thr	Cys	Lys	Met	Glu	Arg	Phe	Tyr	Trp	Leu	Val	Ile	Pro	Median†
Gly	**1.03**	**9.2**	**11.8**	**21.1**	**28.0**	**27.6**	**39.8**	**40.6**	**48.2**	**50.4**	**73.9**	**57.8**	**64.0**	**63.6**	**77.1**	**104**	**224**	**287**	**7170**	50.4
Ser	**0.96**	**9.0**	**15.1**	**24.1**	**30.3**	**41.3**	**45.7**	**60.2**	**55.5**	**54.9**	**59.7**	**59.7**	**52.2**	**64.7**	**76.8**	**110**	**233**	**285**	**7060**	55.5
Thr	**1.04**	**9.6**	**17.1**	**24.6**	**27.9**	**34.4**	**50.0**	**55.5**	**57.6**	**47.6**	**60.8**	**51.2**	**76.4**	**80.6**	**72.5**	**110**	**237**	**279**	**6290**	55.5
Cys	**1.14**	**10.8**	**19.0**	**26.4**	**30.6**	38.3	**48.7**	**46.0**	**46.6**	**64.5**	**48.3**	**83.1**	**73.9**	**83.9**	**111**	**119**	**229**	**304**	**1550**	48.7
AmCys	**1.14**	**10.9**	**15.4**	**21.5**	**32.9**	**39.3**	**41.7**	46.6	**48.9**	**56.5**	**45**	**58.8**	**63.3**	**78.8**	**81.3**	**100**	**215**	**250**	**3900**	48.9
Met	**1.04**	**10.2**	**15.2**	**22.1**	**26.4**	**33.8**	**43.6**	**49.6**	**60.4**	**56.9**	**72.4**	**58.8**	**61.9**	**74.0**	**92.7**	**113**	**211**	**275**	**9300**	57.9
Phe	**1.15**	**10.2**	**18.1**	**24.2**	**27.4**	**29.8**	**39.0**	**46.5**	**58.2**	**58.6**	**62.4**	**61.2**	**69.5**	**75.1**	**102**	**118**	**203**	**287**	**7990**	58.6
Tyr	**1.49**	**10.2**	**11.9**	**24.3**	**28.4**	**33.3**	**38.1**	**48.6**	**55.1**	**64.3**	**41.0**	**56.9**	**58.0**	**70.6**	**120**	**118**	**241**	**306**	**9830**	51.8
Asp	**1.53**	**9.7**	**17.0**	**24.0**	**29.4**	**45.8**	**52.4**	**54.1**	**75.9**	**57.3**	**46.8**	**87.2**	**70.1**	**70.4**	**80.3**	**111**	**241**	**298**	**11800**	55.7
Glu	**1.45**	**9.0**	**16.4**	**25.8**	**32.0**	**32.1**	**36.8**	**44.2**	**77.8**	**59.6**	**60.3**	**80.9**	**70.2**	**94.5**	**98.4**	**130**	**268**	**279**	9600	59.9
His	**1.14**	**10.7**	**15.7**	**24.6**	**31.2**	**33.8**	**47.2**	**43.9**	**50.2**	**63.1**	**69.4**	**48.9**	**72.1**	**82.3**	**95.4**	**116**	**247**	**327**	**8440**	50.2
Lys	**1.02**	**10.5**	**15.6**	**23.6**	**34.0**	**36.5**	**58.1**	**49.0**	**53.5**	**60.9**	**72.5**	**57.4**	**70.1**	**96.7**	**98.1**	**119**	**246**	**313**	**4940**	58.1
Arg	**1.00**	**10.0**	**14.3**	**24.4**	**34.7**	**42.3**	**50.7**	**50.5**	**49.6**	**74.4**	**68.3**	**67.4**	**68.3**	**90.0**	**127**	**128**	**247**	**311**	**5790**	67.4
Ala	**1.05**	**9.3**	**14.9**	**22.5**	**31.9**	**40.6**	**43.5**	**63.7**	**55.9**	**59.2**	**74.1**	**62.4**	**65.6**	**73.9**	**130**	**124**	**254**	**300**	**7370**	62.4
Leu	**1.08**	**10.7**	**16.7**	**25.1**	**32.1**	**33.6**	**46.1**	**53.5**	**60.1**	**62.6**	**56.7**	**62.1**	**72.4**	**75.7**	**74.5**	**155**	**294**	**391**	**10500**	60.1
Val	**1.23**	**10.2**	**18.2**	**27.5**	**33.5**	**34.7**	**49.9**	**63.2**	**63.8**	**65.7**	**64.8**	**67.4**	**66.6**	**79.2**	**88.9**	**154**	**291**	**366**	**8030**	64.8
Ile	**1.26**	**11.5**	**14.5**	**25.9**	**33.8**	**33.0**	**46.3**	**52.7**	**64.4**	**58.8**	**58.6**	**66.4**	**61.5**	**79.3**	**86.7**	**154**	**295**	**384**	**11600**	58.8
Trp	**1.75**	**11.3**	**15.5**	**30.7**	**43.6**	**42.9**	**38.9**	**83.1**	**59.4**	**64.2**	**75.7**	**73.9**	**71.1**	**92.6**	**135**	**133**	**226**	**286**	12000	67.6
Pro	**1.18**	**12.8**	**18.9**	**31.8**	**48.6**	**43.7**	**63.1**	**60.0**	**67.8**	**78.4**	**92.0**	**72.9**	**100**	**114**	**122**	**181**	**364**	**455**	**6590**	72.9
Mean	1.19	10.3	15.9	25.0	32.5	36.7	46.3	53.2	58.4	60.9	63.3	65.0	68.8	81.1	98	126	251	315	7000	60.9
St.Dev.	0.05	0.23	0.49	0.67	1.3	1.2	1.7	2.4	2.1	1.8	3.1	2.5	2.3	3.0	4.9	5.1	9.3	12.2	600	2.3
%St.Dev.	4.4	2.2	3.1	2.7	4.1	3.3	3.6	4.5	3.6	2.9	4.8	3.9	3.4	3.7	5.0	4.0	3.7	3.9	8.8	3.7
Median	1.14	10.2	15.6	24.4	31.9	34.7	46.1	50.5	57.6	59.6	62.4	62.1	69.5	79.2	95	119	241	300	7100	59.6

† Median does not include Yyy = AmCys

Bold type values are experimental

The succinimide mechanism of deamidation has been extensively studied in the context of AsnGly deamidation[75–77] and is responsible for the fast deamidation rates shown for AsnGly sequences. Table 7-1 suggests that this mechanism is probably in operation for most Asn sequences–as has been generally supposed. Since AsnPro cannot react through the succinimide mechanism, the deamidation half-times are very long. The rate for AsnPro peptides is probably due to direct acid and base catalysis by hydrolysis of the amide with nearest neighbor residues modifying this rate. Some obvious trends are discussed below, an exact quantitative explanation is described in Chapter 9.

For AsnIle, the rates are still well below the hydrolysis rates shown for AsnPro. At this level, with reference to AsnPro, AsnIle has about a 4% contribution from hydrolysis, assuming that the hydrolysis rate of AsnPro is comparable to AsnIle. The bulky side chain of Ile apparently impedes, but does not prevent, the succinimide mechanism. AsnVal has nearly the same shaped side chain as AsnIle, but with one fewer methyl groups on the end. This produces slightly less impedance and therefore a higher rate of deamidation.

A marked acceleration is evident when the jump to AsnLeu is made. This is easily understood. Ile, Val, and Thr are the only amino acids with methyl groups on the first carbon of their side chain. This methyl group is the closest to the succinimide and the one most likely to interfere in forming the succinimide ring. The hydroxyl of Thr probably has a catalytic effect which overcomes this problem to a certain extent. When this methyl group is removed from Thr to produce Ser the half-time goes up by a factor of three. When the hydroxyl group of Ser is removed to produce Ala, the half-time goes back down by about 65% indicating that the hydroxyl group is indeed catalytic.

The rest of the right-hand residue effects are also easily rationalized. AsnTrp, AsnTyr, and AsnPhe are all bulky and appear just before AsnLeu. AsnHis stands out as the next fastest after AsnGly in spite of its large ring.

The pentapeptide model was designed to minimize secondary structure effects. Some of these are still present and may in fact be generally a part of the sequence dependence of

deamidation. The half-times of AspAsnLys, AspAsnArg, GluAsnLys and GluAsnArg are about 50% longer than the corresponding median values. The reverse sequences have a similar effect, but with only about 10% longer half-times. This is may be caused by a special structure–perhaps only transient–formed between the positively charged side chain (Lys or Arg) and the negatively charged side chain (Asp or Glu).

Deamidation Rates for Other Asparaginyl Containing Peptides

The pentapeptide model GlyXxxAmideYyyGly was initially chosen in the 1970's as a model for primary sequence dependent deamidation.[3] The residues Xxx and Yyy are varied to determine the sequence dependence and the Gly residues on the end move the charges away from the reaction center. For the work described here a similar choice was made. Pentapeptides still appeared to be the logical length. Adding another residue to either end would increase the risk of secondary structures. Also, it did not seem sensible to use anything on the ends with a functional side chain that may introduce secondary structure effects. It was desirable to determine the reliability of this model.

Table 7-2 – _AsnGly_ Deamidation Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
GlyAsnGly	98	0.082
GlyGlyAsnGlyGly	1.083	7.41
GlyGlySerAsnGlyGly	1.003	8.00
GlySerAsnGlyGly	0.957	8.38
GlySerGlyGlyAsnGlyGlyGlyGly	0.935	8.58
GlyGlyGlyAsnGlyGlyGly	0.916	8.76
GlyGlyGlyGlyAsnGlyGlyGlyGly	0.845	9.49
GlySerGlyGlyAsnGlyGlyHisGly	0.828	9.69
GlyGlyGlyGlyAsnGlyGlyHisGly	0.820	9.78
GlySerGlyAsnGlyGlyGly	0.752	10.7
GlySerGlyAsnGlyHisGly	0.648	12.4
GlyGlyGlyAsnGlyHisGly	0.646	12.4

For this reason, 67 model peptides were synthesized early in the experiments. The deamidation rates of these are shown in Tables 7-2 through 7-8. Tables 7-2 and 7-3 show variations around the sequences GlyGlyAsnGlyGly and GlySerAsnHisGly.

Table 7-3 – _AsnHis_ Deamidation Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
SerAsnHis	39.6	0.203
GlyAlaAsnHisGly	9.29	0.86
GlyGlyAsnHisGly	9.18	0.87
AcSerAsnHisNH_2	9.01	0.89
AlaAlaAsnHisAlaAla	8.44	0.95
GlySerAsnHisGly	8.30	0.97
AcAsnHisAlaAla	8.10	0.99
AcAlaAlaSerAsnHisAlaAlaNH_2	8.05	1.00
AlaSerAsnHisAla	7.83	1.02
GlyGlySerAsnHisGlyGly	7.42	1.08
GlyGlyAsnHisGlyGly	7.41	1.08
AcAlaSerAsnHisAlaNH_2	7.35	1.09
AlaAlaSerAsnHisAlaAla	7.27	1.10
AcGlyGlySerAsnHisGlyGlyNH_2	7.05	1.14
AlaAlaAlaSerAsnHisAlaAlaAla	6.78	1.18
AcGlySerAsnHisGlyNH_2	6.63	1.21
GlyGlyGlySerAsnHisGlyGlyGly	6.47	1.24
AsnHisAlaAla	5.75	1.40

It is clear that the residues on either end are essential to this work. Both SerAsnHis and GlyAsnGly with no end protection are substantially slowed down and would not be representative of this sequence. The fastest peptide in Table 7-3 is AsnHisAlaAla. This indicates that the negative charge on the c-terminal side of SerAsnHis and GlyAsnGly is the main reason for their slow rates, since putting a positive charge directly on the n-terminal side accelerates the rate. Qualitatively, the rates are otherwise very similar.

Table 7-4 is a collection of short peptides. The peptide with the slowest rate, $AlaAlaAlaAlaNH_2$, is a c-terminal amide. This amide cannot deamidate through the succinimide, so a slow rate is expected. The same is true of AlaAlaSerAsn. Both peptides probably deamidate through direct hydrolysis.

Table 7-4 – _XxxAsnYyy Peptide Deamidation Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
$AlaAlaAlaAlaNH_2$	6590	0.00116
AlaAlaSerAsn	3440	0.00223
GlyAsnGly	98	0.082
AlaAsnAla	270	0.030
$AlaAlaSerAsnNH_2$	37.6	0.213
SerAsnHis	39.6	0.203
AlaAlaAsnAlaAla	29.9	0.268
AlaAlaSerAsnAlaAla	27.8	0.289

Table 7-5 – AlaXxxAlaAsnAlaYyyAla Deamidation Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
AlaProAlaAsnAlaProAla	195	0.0411
AlaGluAlaAsnAlaGluAla	31.8	0.252
AlaAlaAlaAsnAlaAlaAla	31.2	0.257
AlaIleAlaAsnAlaIleAla	25.9	0.310
AlaSerAlaAsnAlaAlaAla	25.4	0.316
AlaSerAlaAsnAlaSerAla	21.0	0.382
AlaAlaAlaAsnAlaHisAla	16.9	0.475
AlaSerAlaAsnAlaHisAla	16.6	0.483
AlaLysAlaAsnAlaLysAla	14.5	0.553
AlaTyrAlaAsnAlaTyrAla	14.4	0.557
AlaHisAlaAsnAlaHisAla	12.7	0.632
AlaArgAlaAsnAlaArgAla	10.9	0.736

Table 7-6 – AlaXxxAlaAlaAsnAlaAlaYyyAla Deamidation Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
AlaIleAlaAlaAsnAlaAlaIleAla	45.1	0.178
AlaGluAlaAlaAsnAlaAlaGluAla	43.4	0.185
AlaAlaAlaAlaAsnAlaAlaAlaAla	37.0	0.217
AlaSerAlaAlaAsnAlaAlaSerAla	29.8	0.269
AlaSerAlaAlaAsnAlaAlaAlaAla	28.7	0.280
AlaAlaAlaAlaAsnAlaAlaHisAla	25.0	0.321
AlaHisAlaAlaAsnAlaAlaHisAla	23.7	0.339
AlaSerAlaAlaAsnAlaAlaHisAla	21.1	0.380
AlaLysAlaAlaAsnAlaAlaLysAla	14.1	0.569

Table 7-7 – AlaXxxAlaAlaAlaAsnAlaAlaAlaYyyAla Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
AlaSerAlaAlaAlaAsnAlaAlaAlaSerAla	1,490	0.0054
AlaSerAlaAlaAlaAsnAlaAlaAlaHisAla	630	0.0127
AlaIleAlaAlaAlaAsnAlaAlaAlaIleAla	278	0.029
AlaHisAlaAlaAlaAsnAlaAlaAlaHisAla	59.8	0.134
AlaGluAlaAlaAlaAsnAlaAlaAlaGluAla	55.9	0.144
AlaLysAlaAlaAlaAsnAlaAlaAlaLysAla	15.0	0.535

Table 7-8 – AlaXxxAlaAlaAlaAlaAsnAlaAlaAlaAlaYyyAla Rates

Peptide	$t_{1/2}$ *Days*	$k \times 10^6$ *Sec*
AlaSerAlaAlaAlaAlaAsnAlaAlaAlaAlaHisAla	1150	0.0070
AlaSerAlaAlaAlaAlaAsnAlaAlaAlaAlaSerAla	960	0.0084
AlaIleAlaAlaAlaAlaAsnAlaAlaAlaAlaIleAla	150	0.053
AlaGluAlaAlaAlaAlaAsnAlaAlaAlaAlaGluAla	36.5	0.220
AlaHisAlaAlaAlaAlaAsnAlaAlaAlaAlaHisAla	32.6	0.246
AlaLysAlaAlaAlaAlaAsnAlaAlaAlaAlaLysAla	18.8	0.427

Tables 7-5 through 7-8 are variations on the sequence AlaXxx(Ala)$_n$Asn(Ala)$_n$YyyAla, with various residues two or more residues removed from the amide. In general, the half-times are longer as the sequence lengthens, although the initial effect is sometimes acceleration. The longer half-times are likely due to the gradual formation of α-helix or other secondary structures for long peptides.

Figure 7-1 illustrates this for 6 different types of sequences. In all cases except the Lys peptides, there is a marked slowing of the rates as the sequence becomes long.

An interesting case is the peptide AlaProAlaAsnAlaProAla with an unusually long half-time of 195 days. This is probably due to a special structure formed by the bend that Pro introduces into the peptide.

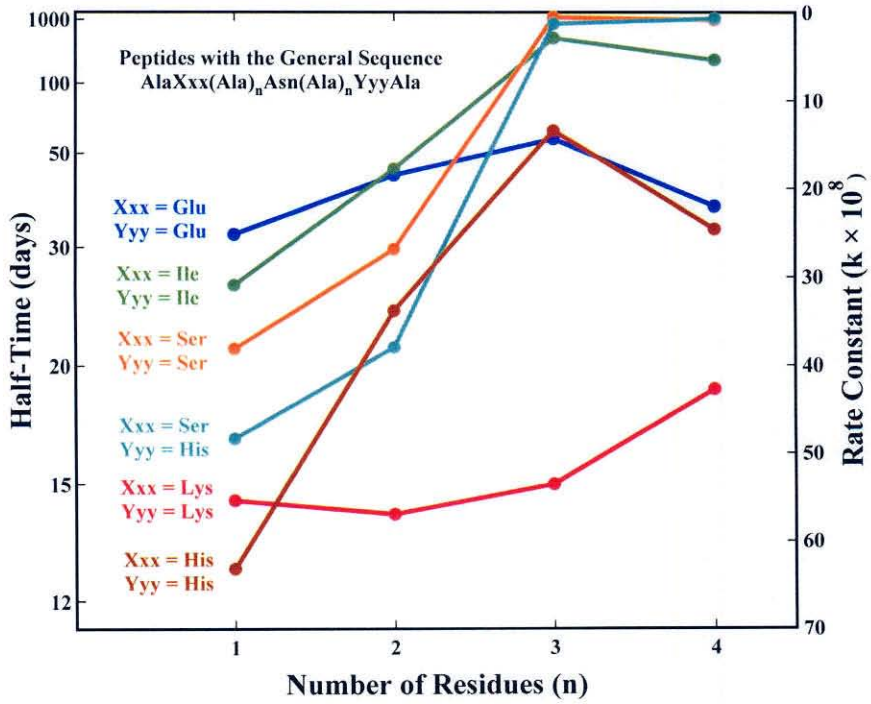


Figure 7-1 – Half-Time as a Function of Sequence Length for Peptides of the Type AlaXxx(Ala)$_n$Asn(Ala)$_n$YyyAla

Several other effects are evident in the model peptide deamidation rates given in Tables 7-2 through 7-8

1. As the carboxyl side nearest neighbors are moved away from the amide by one or more residues, steric factors that inhibited their effects upon deamidation are diminished. As nearest neighbors, Gly is sterically favorable to 5-membered ring imide formation and Ser and His are shaped so that they can catalyze imide-mediated deamidation, probably through dipole effects. Conversely, Lys is too long to bring its side chain to bear on the 5-membered imide. As soon as one or more residues intervene, Lys begins to catalyze deamidation, the half-times moving from about 50 days, to a range of 14 to 19 days. Similar effects are seen for Arg and other residues as summarized in Tables 4 to 7.

2. In the case of Gly as the carboxyl side nearest neighbor, deamidation rate increases as the peptide is elongated by additional Gly residues, and His and Ser also increase this rate, especially carboxyl side His. This effect of His and Ser diminishes with distance from the amide as is shown in Table 7-2. The peptide elongation effect is shown also in Table 7-3, where the SerAsnHis central region has been kept intact, with addition of peptide on both sides beyond this region. Additional Gly residues are more effective than Ala in increasing the deamidation rate, and, in both cases, the addition of the amino acid residues is more effective than simply blocking the ends with acetyl and amide. Removal of the final carboxyl side residue sharply decreases the deamidation rate, as for example in AlaAlaSerAsn. Removal of the final amino side residue accelerates deamidation, as for example in AsnHisAlaAla.

   The percent increases in the first-order rate constants for the additions of Gly pairs to a GlyAsnGly core, Gly pairs to a SerAsnHis core, and Ala pairs to a SerAsnHis core are 18, 9, 13, 15, 8 and 7 respectively, for an average of 12%. Addition of Ala to an AlaAsnAla core decreases the first order rate constant by 10% per Ala pair.

The Ala pair additions, however, produce peptides that are more likely to assume a partially helical structure, which would diminish deamidation.

In all five cases in Table 7-3 where both ends of the peptides have been blocked by Acetyl and Amide or have been elongated by one residue, the Acetyl and Amide lead to slower deamidation rates than do the single residue additions.

3. The large slowing effect apparent in GlyAsnGly and SerAsnHis is most likely due to the nearby carboxyl group. The effect is probably a charge effect, especially since blocking the ends in the case of AcSerAsnHisNH_2 brings the rate back down to normal. Unfortunately peptides were not made for which one side or the other was blocked alone, but the AsnHisAlaAla example indicates that the positive charge on the amino end is actually accelerating the rate rather than decreasing it since AcAsnHisAlaAla has a half-time of 8.10 days and AsnHisAlaAla is only 5.75 days.

4. The orders of increasing deamidation rates in the 7, 9, 11, and 13 residue peptides with the unique side chains moving correspondingly farther away from the amide are shown in Tables 7-5 to 7-8 to be: for 7–Arg, His, Tyr, Lys, SerHis, Ser, Ile, Ala, Glu, Pro; for 9–Lys, SerHis, His, Ser, Ala, Glu, Ile; for 11–Lys, Glu, His, Ile, SerHis, Ser; and for 13–Lys, His, Glu, Ile, Ser, and SerHis.

5. There is a sharp reversal of order and slowing of deamidation by one to two orders of magnitude in the SerSer, SerHis, IleIle peptides in the 11 and 13 residue lengths as shown in Tables 6 and 7. This is likely due to alpha helix or other secondary structure formation, which prevents imide formation. This effect by an alpha helix has been demonstrated in rabbit muscle aldolase.[11–13]

To a lesser extent, the His peptides also show a slowing of deamidation in the 11 and 13 residues lengths that may result from the presence of secondary structure. In marked contrast, this effect is almost entirely absent in the Lys peptides.

6. The short peptides in Table 7-4 illustrate the importance of carboxyl side nitrogen to peptide deamidation at neutral pH. This is in accord with the proposed imide mechanism.

7. It is additionally striking that the effects of chain elongation, inclusion of active side chains on the carboxyl and amino sides, and the introduction of end group blocking agents make well-ordered and quantitative contributions to deamidation rates. With, however, the results reported herein and elsewhere,[6, 26, 73] the known deamidation rates of model peptides have now been increased by more than 5-fold, so reference to the models themselves will suffice, in most cases, for semi-quantitative predictions.

It is clear from this work that the pentapeptide model is a very good choice for studying the primary sequence dependence of deamidation. The rates in the 67 test models were quite similar to the pentapeptide rates except where extraneous factors, not desirable in the sequence set, intervened. The most important effects to keep in mind when comparing peptide rates to protein rates are the special effects on the rate when an amide is very close to the end of a protein and the effect that Arg and Lys can have even when several residues from the amide. It is quite possible, however, that Arg and Lys will not produce as large an effect in most proteins, because of the secondary, tertiary, and quaternary structure constraints present.

Table 7-9 First-Order Deamidation Half-Times of GlyXxxGlnYyyGly in days at pH 7.4, 37.0 °C, 0.15 M Tris HCl

Xxx\Yyy	Gly	Cys	Met	Thr	Ser	Ala	His	Lys	Leu	Ile	Val	Arg	Glu	Asp	Phe	Pro	Tyr	Trp	Median
Cys	**560**	800	3200	3500	3800	4100	4200	4400	4800	4900	5000	5100	5600	6100	6500	7100	7900	9100	4800
Met	600	900	3500	3800	4100	**4400**	4400	4600	5000	5000	5000	5100	5800	6200	6600	7300	8200	9400	5000
Thr	**670**	1000	3700	4000	4200	4300	4500	4800	5200	5300	5100	**5100**	5900	6300	6800	7500	8400	9700	5100
Lys	650	1000	4000	4100	4200	**4300**	**6100**	4000	5300	5400	5700	**2300**	**5400**	5900	7000	7700	8800	**10000**	5300
Arg	**660**	**1000**	4100	4200	4300	4400	4900	4000	5400	5500	5800	2300	5400	5900	7100	8100	9200	11000	4900
Val	640	1300	4200	4300	4400	**4500**	5000	**5200**	5500	**5600**	5900	6100	6500	7000	7200	8500	9700	12000	5500
Pro	630	1600	4500	4600	4600	**4700**	5200	5500	5800	6000	6200	**6400**	6800	7200	7300	8900	10000	13000	5800
Ala	**610**	**1900**	4400	**5100**	5200	**5300**	5500	5700	6100	**6200**	6400	**7200**	7300	**7400**	7500	9300	**10000**	14000	6100
Gly	**650**	1900	**4500**	5200	**5700**	**5900**	5900	6000	6200	6300	**6500**	7200	7300	7600	**7600**	**10000**	12000	**15000**	6200
Leu	**670**	2000	4800	5300	5800	6000	6100	6100	6300	6500	6800	7200	7400	7800	8000	10000	12000	16000	6300
Ile	**620**	2000	**5100**	5300	5800	6200	6100	6100	6300	6500	7100	7200	7700	8100	8100	10000	12000	16000	6300
Phe	660	2000	5100	5300	5900	**6300**	6200	6200	6400	6400	7100	7200	8100	8200	8200	10000	12000	16000	6400
Ser	700	2100	5100	5400	6000	**6400**	**6500**	6300	6100	**5900**	6800	7200	8100	8200	8300	10000	13000	17000	6400
Glu	750	2100	5200	5400	6100	**7100**	2500	4600	4300	**4200**	6400	5200	8200	8300	8400	10000	13000	17000	5400
Asp	800	2100	5200	5400	6200	**7100**	**2500**	**4600**	6200	6400	6600	5200	8200	8400	8500	11000	13000	17000	6200
His	**850**	2200	5200	5500	6300	7200	7200	4000	6600	6700	6800	**4500**	5800	**5600**	8600	11000	14000	18000	6300
Tyr	**850**	2200	5300	5600	6400	7300	7400	7500	7800	7900	8000	8100	8300	8600	8700	11000	14000	**18000**	7800
Trp	850	2300	5300	**5600**	6500	7400	7500	7600	7900	8000	8200	8300	8500	**8800**	**8600**	11000	14000	19000	7900
Mean	690	1700	4600	4900	5300	5700	5400	5400	6000	6000	6400	5900	7000	7300	7700	9400	11200	14300	6000
St.Dev.	22	129	163	169	228	296	352	272	226	233	221	423	273	259	180	329	521	809	246
%St.Dev.	3.2	7.6	3.5	3.4	4.3	5.2	6.5	5.0	3.8	3.9	3.4	7.2	3.9	3.5	2.3	3.5	4.7	5.7	4
Median†	660	1950	4650	5250	5750	5950	6000	6050	6250	6400	6650	7200	7350	7700	7800	10000	12000	15500	6150

† Median without charged residues.

Bold type values are experimental.

Deamidation Rates for Glutaminyl Pentapeptides

Table 7-9 contains the Gln pentapeptide rates, also measured in 0.15 M Tris buffer at pH 7.4, 37.00 °C. The 52 values in bold were actually measured. The rest of the values in the table were interpolated and extrapolated based on the measured rates.[26]

It is clear from these data that GlnGly sequences are especially fast compared to other Gln sequences. This is most likely due to the formation of the glutarimide[26, 93]–a similar ring to the succinimide. The other rates except possibly for Cys are probably determined largely but not entirely by hydrolysis. The rates are comparable to the AsnPro and GlnPro rates, which must be hydrolysis, since the succinimide and glutarimide cannot form.

Deamidation of Aldolase

Table 7-10 and Figure 7-2 show that, as expected from previous work,[12, 13] the c-terminal SerAsnHis sequence in aldolase deamidates at essentially the same rate as does its corresponding model peptide—which was present in the same solution at the same time and was simultaneously measured in the mass spectrometer.

Table 7-10 Deamidation of Rabbit Muscle Aldolase and Model Peptides in 1.0×10^{-3} M Peptide or Protein, 37.00 °C, pH 7.4, 0.15 M Tris HCl in the Same Solution.

Peptide	$t_{1/2}$ Days	k x 10^{-6} Sec
Aldolase - IleuSerAsnHisAlaTyr	9.4	0.85
GlySerAsnHisGly	8.3	0.97
Aldolase - AlaLeuAlaAsnSerLeuCysGlnGlyLys	More than 150 days	
GlyAlaAsnSerGly	11.4	0.70

The Tris value of 8.3 days for the deamidation of GlySerAsnHisGly, as compared with that for aldolase of 9.4 days, is in good agreement with that previously reported of

6.4 days for GlySerAsnHisGly in phosphate buffer[13] and aldolase of 8 days.[12] Phosphate buffer is known to accelerate deamidation.[72] It is also in good agreement with the reported *in vivo* turnover rate for rabbit muscle aldolase of 8 days.[12] Measured with the presence of aldolase, the GlySerAsnHisGly rate in Tris is 9.0 days.

Conversely, however, the second amide sequence from the c-terminal, AlaAsnSer, deamidated with a half-life of 11.4 days in the peptide but was not observed to deamidate at all in the protein during the 14-day duration of the experiment. In the protein, therefore, this sequence has a deamidation rate at least 10- to 20-fold slower than the model. This difference is likely because this amide is located in an α-helix in the protein. The postulated cyclic imide intermediate involved in the deamidation of these asparaginyl peptides requires that the amide nitrogen of the residue toward the carboxyl end of the asparaginyl residue be available for imide formation. This would require disruption of the α-helix in which this nitrogen is participating.

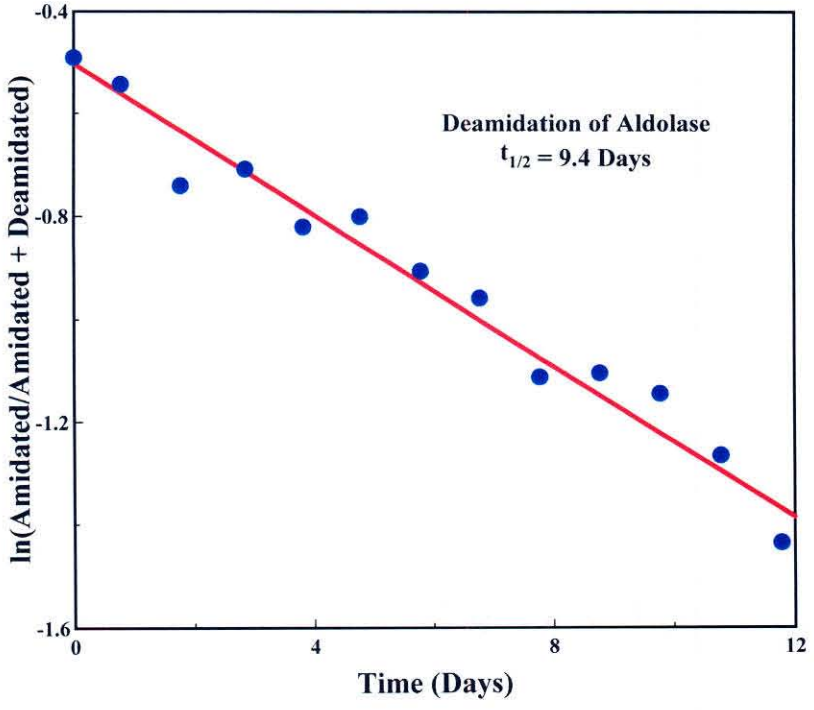


Figure 7-2 – Deamidation of SerAsnHis in Rabbit Muscle Aldolase

Analysis of Data

One interesting feature of the Gln and Asn rate data is that the rates of the shortest Gln peptides are just slightly longer than those of the longest Asn peptides. Thus Gln picks up where Asn leaves off without a gap in rates. Furthermore a wide range of rates are represented by Gln beginning with the GlnGly rates and picking up with other rates that are not very much slower than GlnGly. This is illustrated in Figure 7-3.

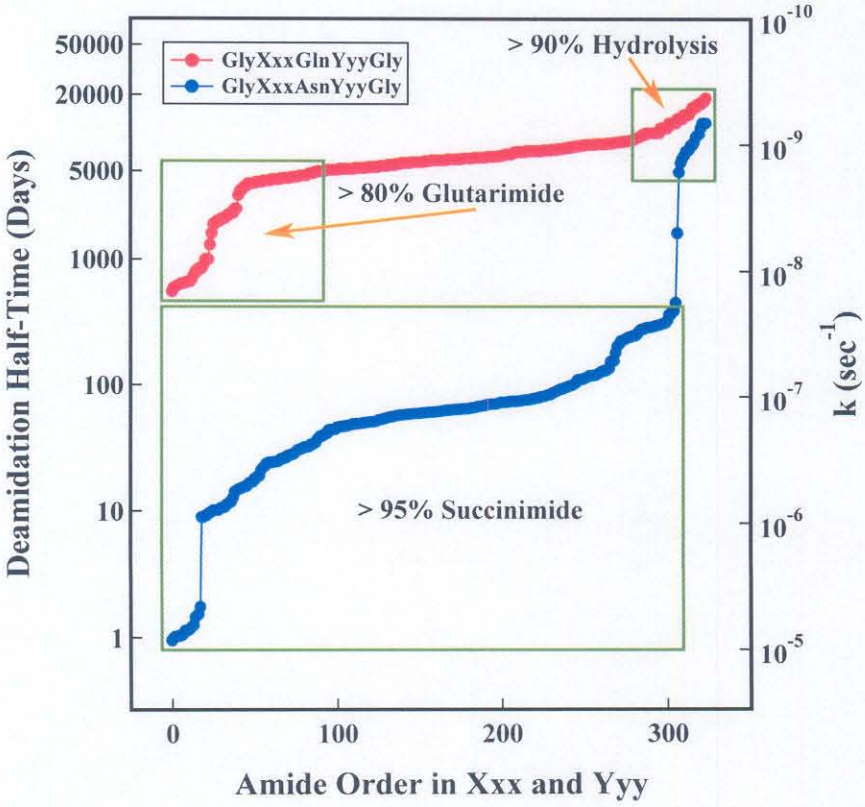


Figure 7-3 – Distribution of Pentapeptide Deamidation Rates for Peptides with the Sequences GlyXxxAsnYyyGly and GlyXxxGlnYyyGly

The sequence dependence of deamidation is illustrated with the three-dimensional plot[163] shown in Figure 7-4. This plot incorporates the sequence dependent deamidation rates for both Asn and Gln with the right and left-hand residue effects. The flattening out of the surface at about 10,000 days is caused by direct hydrolysis.

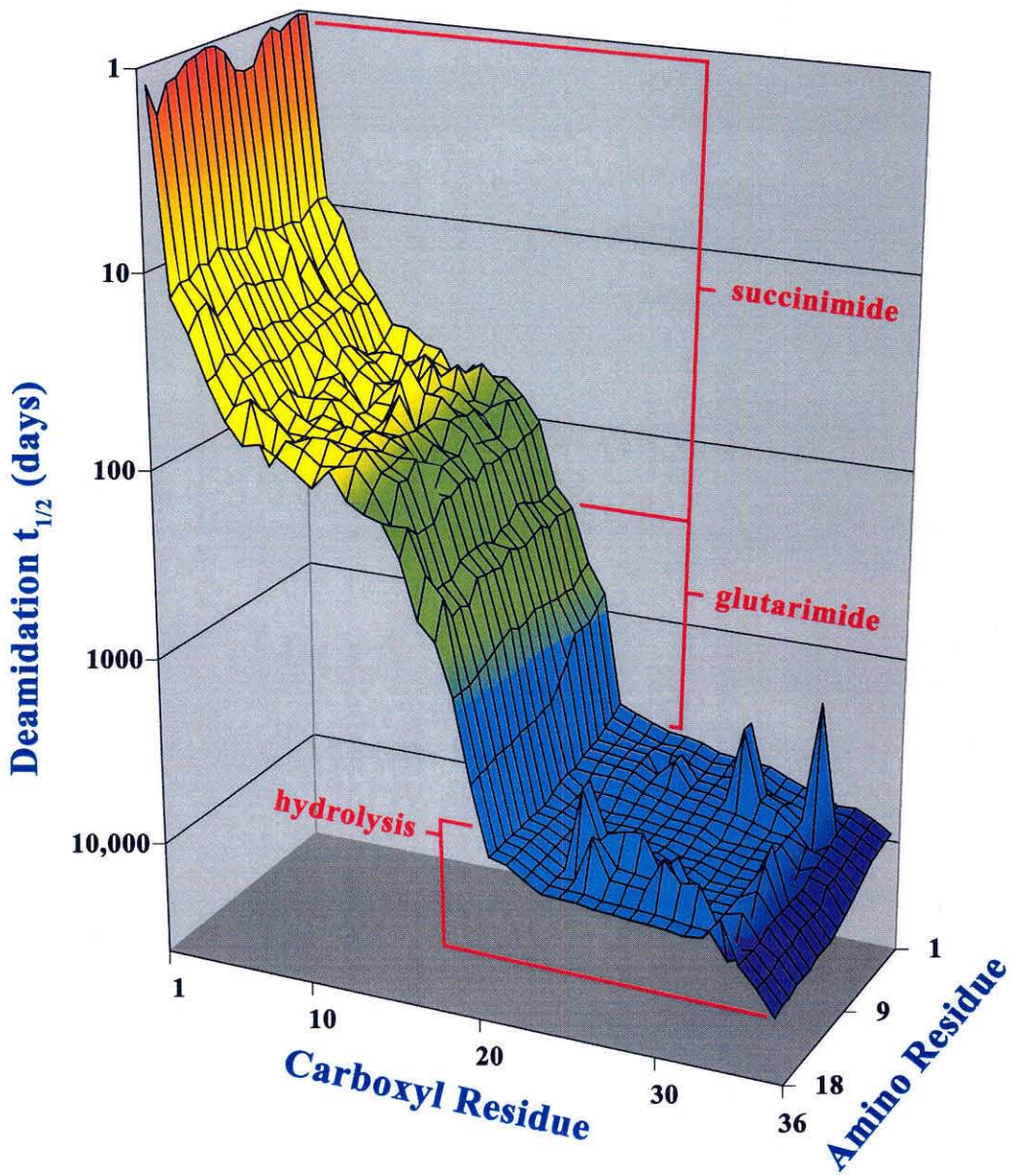


Figure 7-4 – Sequence Dependence of Asn and Gln Pentapeptide Deamidation Rates

Chapter 8

EXPLANATION AND PREDICTION OF PRIMARY SEQUENCE RATES[161]

Introduction to Method

The mechanism of AsnGly deamidation has been extensively studied and appears to deamidate through a succinimide intermediate.[75-77, 28] Very little work on the mechanism of Gln deamidation has been done, but one study on a short blocked peptide with the GlnGly sequence indicates that it deamidates through a similar glutarimide intermediate.[93]

The cyclic intermediate is formed by reaction of the amide in close proximity to the right-hand residue side chain. The absence of any side chain on Gly, reduces steric hindrance and allows the formation of this intermediate with relative ease. The substitution of any other amino acid in this position has been understood to impede the cyclic mechanism, although the extent of this was unknown. The possibility of another mechanism that could operate at relatively short half-times has not been ruled out.

The measurements of the Asn pentapeptides described in Chapter 7 are currently the strongest evidence that the succinimide mechanism is operating over the entire range of Asn sequences in Tris buffer. Without any special analysis, it is clear that rate decreases as the left-hand residue becomes more bulky. Catalytic effects of certain functional groups are also clearly present. The rate of deamidation of AsnPro–which cannot form a succinimide intermediate–is found to be thousands of days, indicating that this fast mechanism is unavailable to AsnPro.

It is also clear that both Asn and Gln also deamidate at a much slower rate through general acid and base catalyzed hydrolysis. Based on the AsnPro and GlnPro data, we have estimated this rate to be *on average* about 8000 days in 37.00 °C, pH 7.4, 0.15 M Tris

buffer. Sequence dependence of hydrolysis is probably also present, but has not yet been measured separately.

For Asn peptides, hydrolysis is a very minor factor. The inherently slower rate of Gln deamidation, however, causes hydrolysis to be a major pathway in Gln deamidation. The limit of about 8000 days is responsible for the gradually leveling off of the Gln data at about this level. Hydrolysis sequence dependence is probably responsible for the few rates longer than 8000 days.

Combining these data into a completely understood and quantitative explanation of primary sequence dependence is a valuable contribution. It is this task which is undertaken here. It is found that a simple semi-empirical steric and catalytic model explains most of these effects quantitatively.

A table of coefficients is produced that allows the prediction of deamidation rates for peptide sequences that have not yet been measured. The method also shows how the measurement of deamidation rates for a selection of non-natural sequences would allow further expansion of this table.

This method permits a better understanding of deamidation, provides a prediction procedure for protein engineering, and facilitates the improved computation of peptide and protein primary, secondary, and tertiary structure deamidation rates.[7-9]

Prelude

The procedure described here is based on the effect of individual side chain groups on the effective activation energy of the reaction. Each atom or group of atoms added to the side chains neighboring to Asn or Gln either increases or decreases the overall activation energy of deamidation.

Most of the groups that increase the activation energy and therefore slow the rate do so through steric hindrance of the reaction. This hindrance is explained very well using the succinimide model for Asn deamidation[75-77, 28] and the glutarimide model for Gln deamidation.[93] In addition, deamidation proceeds through hydrolysis. Hydrolysis is most prevalent at low and high pH, but also proceeds at a reduced rate at neutral pH.[93, 26] We compute herein the sequence dependence of the imide mechanisms. There is not yet sufficient experimental data to calculate sequence dependence of hydrolysis, so we use a median hydrolysis rate.

Some atoms, or groups of atoms, may have both a catalytic effect and a steric effect. For example, an –OH group in the γ position slows deamidation to a small extent, but much less than a hydrogen atom. The steric hindrance of –OH is higher than that of –H, so –OH must also have a catalytic effect.

The deamidation of an individual molecule is a statistical phenomenon. It is not occurring at a particular rate. Each amide can be assigned a probability that it will deamidate during a particular time interval. Deamidation rates result from the combined statistics of an *ensemble* of these molecules. Increasing the activation energy simply lowers the chance that a particular molecule will react. The deamidation half-time of a GlyXxxAsnGlyGly peptide is about 1 day. This means that each individual molecule has a 50% chance of deamidating during 1 day or a 0.00058% chance of deamidating during 1 second. Since the time scale of molecular motion is several orders of magnitude faster than this, it is clear that deamidation of even a fast amide occurs infrequently on the molecular timescale.

It has been hypothesized that the removal of the hydrogen atom from the backbone nitrogen is the first step in the succinimide deamidation mechanism.[162] The low acidity of this proton may explain the low probability of the reaction. The amide could be viewed as waiting for a proton to be removed. Once this occurred, the amide would have a brief period of time during which it could react, providing that it had sufficient energy as governed by the Boltzman distribution to get in position fast enough. If the amide is too far away or collides with a side chain atom, it may lose its opportunity to react. The catalytic effects of certain groups may be caused by transient hydrogen bonds and other structures that cause the amide to spend more of its time in closer proximity to the site at which it must react.

Alternatively, this can be viewed as separate low probabilities, P_1 and P_2 for deprotonation and amide positioning. These probabilities are likely to be largely, but not entirely independent. Reaction occurs with probability $P = (P_1)(P_2)$.

Whether or not the backbone proton must be removed, it is definitely true that the amide needs to move into the correct position if it is to form an imide. Sterically hindering groups can interfere with this in at least two ways. First a group–for example a methyl group–that is fixed in one place can increase the energy required for the amide to move into position, which raises the activation energy. Second, when a group is free to move through rotation around one or more bonds, it will interfere to a significant extent only a certain fraction of the time. Thus, these steric effects depend on size of the group, the variety of conformations available to that group, and the distribution function of occupation of these conformations.

With these points in mind, we now turn to the problem of understanding the carboxyl-side residue sequence dependence of deamidation.

Calculation of Constants

Deamidation half-times for Asn and Gln depend most strongly on the carboxyl-side residue. There is also an amino-side residue effect, but, in 37.00° C, pH 7.4, 0.15 M Tris buffer, this is about a factor of 20 less[26] than that of the carboxyl-side residue effect. The amino-side residue effect is so small that it is partially masked by other experimental variables, such as buffer catalysis. A thorough analysis of the amino-side residue should wait until sufficient deamidation rates have been measured at zero buffer concentrations. Therefore, for analysis of the carboxyl-side effects, we have used medians of 18 amino-side residues. These include all of the ordinary 20 amino acid residues except for Gln and Asn, which have not been measured.

Deamidation at long half-times is also substantially affected by direct hydrolysis of the amides. This occurs in parallel with the imide mechanisms. We use here a median half-time for this hydrolysis of 8000 days, which is the median for GlyXxxAsnProGly peptides (26). This value is utilized to correct the Asn and Gln values used in these calculations. In the case of Asn, this is a small correction of between .01% and 4% of the reaction rate, but for Gln the correction is substantial because hydrolysis rates are comparable to those of the glutarimide mechanism at pH 7.4. There is also an apparent sequence dependence of hydrolysis in the Gln peptides that cannot be completely corrected with the currently available data because it, like the amino-side effect, is partially masked by buffer ions.

Table 8-1 lists the natural log of the median hydrolysis-corrected rate constants times 100 (s^{-1}) for the Asn and Gln pentapeptides as a function of carboxyl-side residue as computed from experimental data.[6, 26, 73] Two sets of values are listed. The first (100)(ln(k)) is directly proportional to the Arrhenius activation energy. The second is normalized to Gly. This removes the effect of everything except the side chain and defines the effect of the Gly hydrogen side chain as zero. Both hydrogen and carbon occur in this β position and have comparable effects.

Table 8-1 (100)ln(k) Deamidation for the Medians of Xxx in GlyXxx(Asn/Gln)YyyGly[#]

Yyy	Asn[§]		Gln[§]	
	Calculated Values	Normalized to Gly	Calculated Values	Normalized to Gly
Gly	-1186.3	0	-1831.1	0
His	-1405.8	219.4	-2181.5	350.4
Ser	-1448.3	262.0	-2165.6	334.4
Ala	-1493.1	306.8	-2178.2	347.1
Asp	-1520.1	333.7	-2393.4	562.3
AmCys	-1528.4	342.1	---	---
Thr	-1556.9	370.6	-2136.5	305.3
Cys	-1566.2	379.8	-1958.8	127.6
Lys	-1579.4	393.1	-2184.9	353.7
Met	-1582.8	396.5	-2104.7	273.5
Glu	-1587.4	401.1	-2313.2	482.0
Arg	-1587.0	400.7	-2290.7	459.5
Phe	-1598.3	411.9	-2433.6	602.5
Tyr	-1611.5	425.1	---	---
Trp	-1630.3	444.0	---	---
Leu	-1652.6	466.3	-2198.9	367.7
Val	-1724.8	538.5	-2230.9	399.7
Ile	-1747.4	561.1	-2210.2	379.0

[§] A correction of 8,000 = half-time for hydrolysis was applied to the experimental values before normalization to Gly. [#] $k = sec^{-1}$.

Effect of Aliphatic Side Chains

First, we consider the effect of hydrocarbon substituents on the Asn activation energy. The α-carbon is the backbone carbon in the peptide chain with successive positions assigned Greek letters as shown in Figure 8-1.

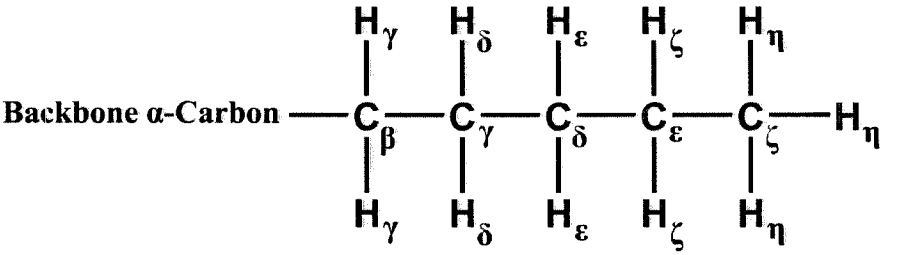


Figure 8-1 – Convention for Designating Positions Along Amino Acid Residue Side Chain. This Convention is Used Regardless of Atom Type and Branching.

The method used is to determine the effects of each atom or group of atoms individually. This is best illustrated by Figure 8-2 which shows a sequence of side chains.

Substitution of a methyl group for hydrogen at the β position occurs in going from Gly to Ala. Thus the effect is computed from Table 8-1 as (Ala – Gly) = 306.8 – 0 = 306.8.

The effect of a methyl at the γ position can be computed in two ways, by moving from Ala to Val or from Ser to Thr. Val contains two γ-CH_3 groups, so we compute it as (Val – Ala)/2 = (538.5-306.8)/2 = 115.9. Computed from Thr – Ser this value is 370.6 – 262.0 = 108.6. The average is 112.3.

For substitution of a δ-CH_3 group, compare Ile and Val, (Ile – Val) = (561.1-538.5) = 22.6. This value can also be obtained from Leu, which has two δ-CH_3 groups. The value of the non-natural group $-CH_2CH_3$ is determined from Ala and the value we determined earlier for a γ-CH_3 group substitution. Thus, for $-CH_2CH_3$, we have (306.8 + 112.3) = 419.1 and (Leu – 419.1)/2 = (466.3 – 419.1)/2 = 23.6. This is almost identical to the previous value of 22.6. The average is 23.1.

Thus, substitution increments for replacement of a hydrogen atom by a methyl group in the β, γ, and δ positions are 306.8, 112.3, and 23.1. This procedure is internally consistent,

with two redundant verifying examples. As would be expected, the values are successively smaller as the methyl groups are placed farther from the reaction site.

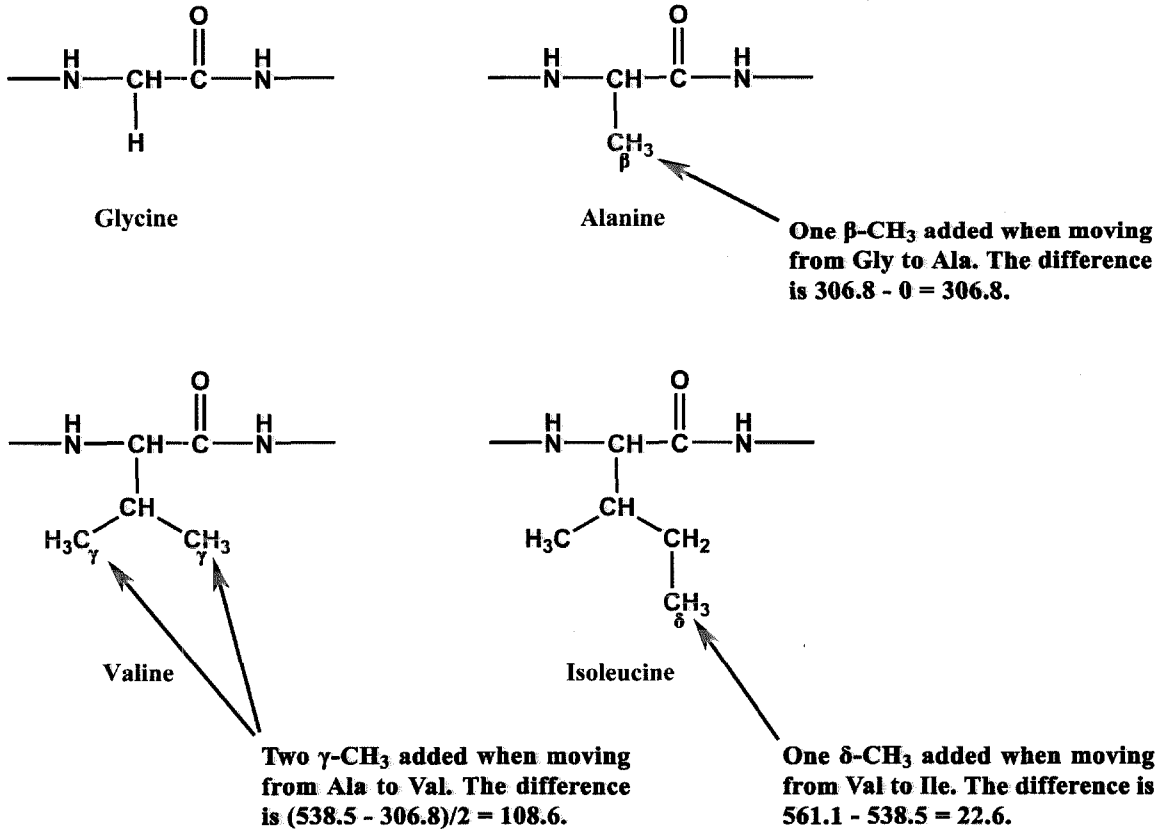


Figure 8-2 – Illustration of Method for Determining Individual Group Effects

These values can also be derived from the statistical probability that the methyl groups will be in a position to interfere with deamidation. Define the effect of the β-CH_3 group as X, and substitute one of its hydrogen atoms with a γ-CH_3 group. Reference to a CPK model shows that the new methyl group will occupy a similarly interfering position for about 120° out of the 360° rotation around the α-β bond. Thus the effect of the second methyl group should be about (1/3)(X). Examination of CPK models with the addition of a δ-CH_3 group

indicates that the interference will be the same as that of the γ-CH_3 except that it occurs for 180° out of the 360° of the α-β bond rotation, but only during 120° of the 360° β-γ rotation. Thus the effect of replacing a hydrogen atom with a δ-CH_3 group is (180/360 – 120/360)(120/360)X = (3/6 – 2/6)(1/3) = (1/18)X.

Since X is 306.8, the effects of the γ and δ methyl groups should be 102.3, and 17.0 respectively. In fact they are 112.3 and 23.1.

It is useful to separate the effect of the hydrogen atoms from the carbon atoms in these calculations. The substitution, for example, of a γ-CH_3 group involves the removal of a γ-H and replacement by a γ-C and 3 δ-H atoms. What is the effect of each of these individual atoms?

Four equations encapsulate this

(Eq. 8-1) $H_\beta = 0$ (By definition from normalizing to Gly)

(Eq. 8-2) $C_\beta + 3\ H_\gamma - H_\beta = 306.8$

(Eq. 8-3) $C_\gamma + 3\ H_\delta - H_\gamma = 112.3$

(Eq. 8-4) $C_\delta + 3\ H_\epsilon - H_\delta = 23.1$

Additionally, the effect of a carbon atom should be approximately a constant times the effect of a hydrogen atom if they are in identical positions. Thus

(Eq. 8-5) $C_\beta = a(H_\beta)$

(Eq. 8-6) $C_\gamma = a(H_\gamma)$

(Eq. 8-7) $C_\delta = a(H_\delta)$

(Eq. 8-8) $C_\epsilon = a(H_\epsilon)$

The value of "a" is determined by plotting the cube root of the carbon atom result vs. the substituent number as shown in Figure 8-3b. The graph becomes perfectly linear when a = 1.038 as shown, which is about (Radius of Carbon/Radius of Hydrogen)$^{1/3}$ = 1.07. Using this value for "a"

(Eq. 8-9) $H_\beta = 0$

(Eq. 8-10) $C_\beta = 0$

(Eq. 8-11) $H_\gamma = (306.8)/3 = 102.3$

(Eq. 8-12) $C_\gamma = a(306.8)/3 = 106.1$

(Eq. 8-13) $H_\delta = (112.3 + (1-a)(306.8)/3)/3 = 36.1$

(Eq. 8-14) $C_\delta = a(112.3 + (1-a)(306.8)/3)/3 = 37.5$

(Eq. 8-15) $H_\epsilon = (23.1 + (1-a)(112.3 + (1-a)(306.8)/3)/3)/3 = 7.2$

(Eq. 8-16) $C_\epsilon = a(23.1 + (1-a)(112.3 + (1-a)(306.8)/3)/3)/3 = 7.5$

This calculation provides values for the ϵ hydrogen and carbon atoms. Further extrapolation of the line from the first three points in Figure 8-3b gives the fourth point at ζ.

The fractional volume occupied by a linear side chain should decrease with the cube of the substituent number. This is the basis for the plot in Figure 8-3b. Also of interest are the unmodified experimental values, which are only known in terms of removal of a hydrogen atom and substitution of a methyl group. This graph should also be linear, and it is as shown in Figure 8-3a. The line in Figure 8-3a has a correlation coefficient of 0.99996. The first three values are experimental, and the last one is computed from the results of our calculation. The position of this fourth value optimizes at the same value of "a" where the line in Figure 8-3b becomes a perfect fit.

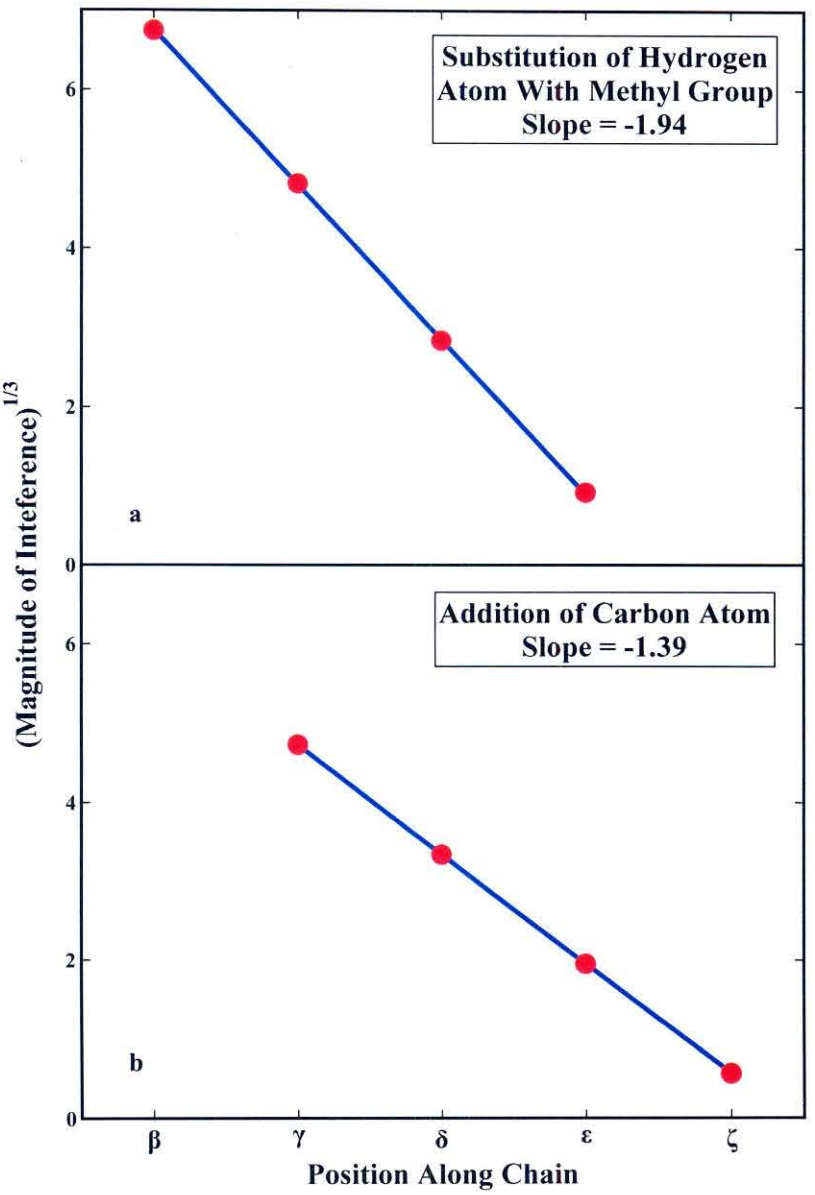


Figure 8-3 – Graph of the cube roots of the methyl substitution for hydrogen coefficients (a) and carbon atom coefficients alone (b) as a function of the position along the carboxyl residue side chain. The β, γ, and δ points for (a) were calculated directly from experimental data. The ϵ point is dependent on the individual atom calculations. The linearity of (b) was used to optimize the individual atom coefficients. As expected in this theoretical treatment, the fractional volume occupied by a linear side chain decreases with the cube root of the substituent number.

Also of interest are the slopes of the lines in Figures 8-3a and 8-3b. Examination of CPK models indicates that each addition of a carbon atom increases the available radius by roughly 0.8 Å, while the effective steric radius of carbon is about 1.1 Å. Thus we would expect the slope in Figure 8-3b to be about $-1.1/0.8 = -1.38$. In fact it is –1.39. Similarly for Figure 8-3a the effective radius of a methyl group is larger, which should give a steeper slope. An estimate of the overall methyl radius is about 1.6 Å which gives a slope of $-1.6/0.8 = -2$. The actual value is –1.94.

These consistencies verify this steric model and also give strong evidence for the succinimide mechanism over the entire range of pentapeptide Asn deamidation sequences.

Aromatic Rings

Three of the 20 amino acids contain aromatic rings–Phe, Tyr, and Trp. Each of these rings is connected in the γ position as illustrated in Figure 8-4.

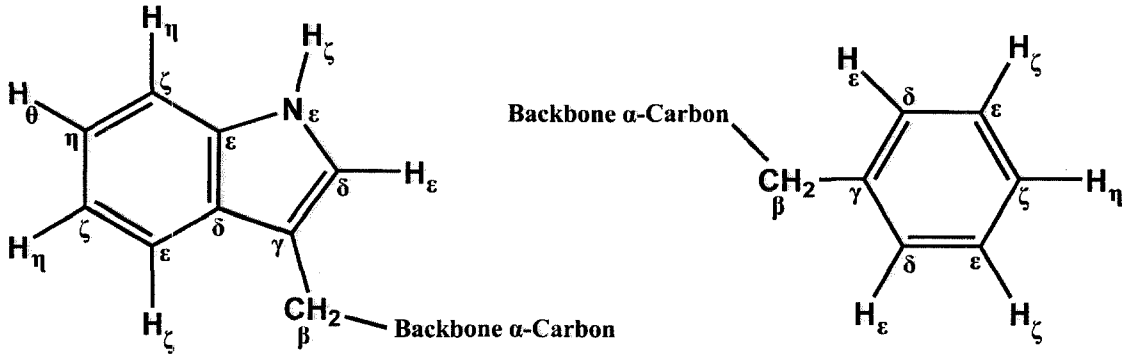


Figure 8-4 – Labeled Trp and Phe Side Chains

Calculation of the effects of these rings in the γ position can be obtained by subtracting the values for one β carbon atom and two γ hydrogen atoms from the appropriate values listed in Table 8-1 as follows

(Eq. 8-17) $\gamma C_5H_5 = \text{Phe} - (C_\beta + 2\ H_\gamma) = 411.9 - (0 + 2(102.3)) = 207.3$

(Eq. 8-18) $\gamma C_5H_5OH = \text{Tyr} - (C_\beta + 2\ H_\gamma) = 425.1 - (0 + 2(102.3)) = 220.5$

(Eq. 8-19) $\gamma C_8H_6N = \text{Trp} - (C_\beta + 2\ H_\gamma) = 444.0 - (0 + 2(102.3)) = 239.4$

These are the experimental results. These values can also be derived from the aliphatic elemental coefficients as follows.

First consider Phe. The γ-C and δ-C are about as free as in a non-aromatic ring. The ϵ-H is not as free to move, but its direction allows it to interfere with deamidation. The remaining atoms are directed away from the reaction site. Therefore

(Eq. 8-20) $\gamma\text{-Phenyl} = C_\gamma + 2\ C_\delta + 2\ H_\epsilon = 106.1 + (2)(37.5) + (2)(7.2) = 195.5$

(Eq. 8-21) If the ϵ carbon atoms are included, γ-Phe = 210.5.

The experimental γ-Phe = 207.3. The result is closer with ϵ-carbon atoms, but remarkably accurate in either case.

For Trp, using the value of 207.3 from Phe, the extra effect of the second ring can be added. Specifically, what is the effect of the ζ-H atom? If averaged over the many conformations available to an aliphatic, the value is about zero, but this atom in Trp is held in one position and directed toward the reaction site. Examination of Figure 8-4 shows that the ζ-H atom occupies a position relative to the reaction site that is almost identical to the nearby δ-C. Using the value for a δ-H and correcting the ϵ-H to a ϵ-C

(Eq. 8-22) $\gamma\text{-Indole} = \gamma\text{-Phenyl} + H_\delta + C_\epsilon - H_\epsilon = 207.3 + 36.1 + 7.5 - 7.2 = 243.7$

The experimental value is 239.4.

If instead of beginning with the experimental value for Phe, addition of the basic atom values is used for all of the atoms, the result is.

(Eq. 8-23) $\gamma\text{-Indole} = C_\gamma + 2\,C_\delta + H_\epsilon + C_\epsilon + H_\delta = 106.1 + (2)(37.5) + 7.2 + 7.5 + 36.1$
$= 231.9$

This is remarkably close to the experimental value and has been determined entirely from individual atom contributions.

Tyr is identical to Phe except for the OH group in the para position on the phenyl ring. Based on steric hindrance, this group should not affect the deamidation rate, so the value should be identical to that for Phe. In fact, it is slightly higher at 220.5 vs. 207.3.

One possible cause of this small difference is hydrogen bonding to other groups, such as the amide side chain. Another possibility is that the hydroxyl changes the electron distribution of the ring.

In summary, the aliphatic side chains can be reduced to individual atomic contributions and summed to predict the deamidation rate for any sp^3 hydrocarbon chain with remarkable accuracy. These atomic coefficients can also be used with good accuracy for aromatic rings. Therefore, the imide deamidation rates for any aliphatic Asn amino-side residue can be accurately computed, and other hydrocarbons can also be estimated if care is taken to allow for the different geometries.

Other Atoms and Charges

Other atoms in the ordinary amino acid residue side chains include oxygen, nitrogen, and sulphur. In addition to steric hindrance, these atoms have additional characteristics such as hydrogen bonding capability and positive and negative charges that can increase or decrease the reaction rate.

With data from the 20 naturally occurring amino acids, a partial understanding of these effects is possible, but these amino acid residues do not provide experimental values for multiple active substituent effects. For example, would the addition of a second hydroxyl group to Ser increase the deamidation rate, which is already catalyzed by the Ser hydroxyl? It is likely that these effects are additive, because deamidation occurs so infrequently that many of the effects are probably transient. A structure that is only formed for a small fraction of the time should be largely independent of other formations that can take place at other times. Experiments to answer this question will be the subject of a later report.

First, let us consider positive charge. Removal of all residues on the amino side of Asn accelerates the deamidation rate,[73] possibly as a result of the nearby positive charge. Also, the imidazole positive charge from carboxyl-side His has a very large catalytic effect. Lys and Arg are the other two positively charged residues. Assuming that the steric hindrance of N is similar to C we calculate

(Eq. 8-24) Catalytic His = His – Phe = 219.5 – 411.9 = **-192.4**

(Eq. 8-25) Catalytic Lys = Lys – (C_β + 2 H_γ + C_γ + 2 H_δ + C_δ + 2 H_ϵ + C_ϵ + N_ζ + All others zero) = 393.1 – [0 + (2)(102.3) + 106.1 + (2)(36.1) + 37.5 + (2)(7.2) + 7.5 + 0.2] = **-49.4**

(Eq. 8-26) Catalytic Arg = Arg – (C_β + 2 H_γ + C_γ + 2 H_δ + N_δ + H_ϵ + C_ϵ + 2 N_ζ + All others zero) = 400.7 – [0 + (2)(102.3) + 106.1 + (2)(36.1) + 37.5 + (7.2) + 7.5 +(2)(0.2)] = **-34.8**

(Eq. 8-27) Catalytic His in 2nd position to the right = [(AAANAHA – AAANAAA) + (ASANAHA – ASANAAA)]/2 = [(269.8 – 331.1) + (268.0 – 310.5)]/2 = **-51.9**

(Eq. 8-28) Catalytic His in 3rd position to the right = [(AAAANAAHA – AAAANAAAA) + (ASAANAAHA – ASAANAAAA)]/2
= [(308.9 – 348.1) + (292.0 – 322.7)]/2 = **-35.0**

Figure 8-5 shows these catalytic effects, without the NH steric effect removed, as a function of the distance from the reaction site. The peptide values for His in the second and third positions are from reference **73**. While the correlation is good, it is likely that other factors are present. In addition to the standard steric effects of the β to ζ atoms, extra steric effects may be important for Lys and Arg because catalysis may involve bringing the side chain closer than its average position.

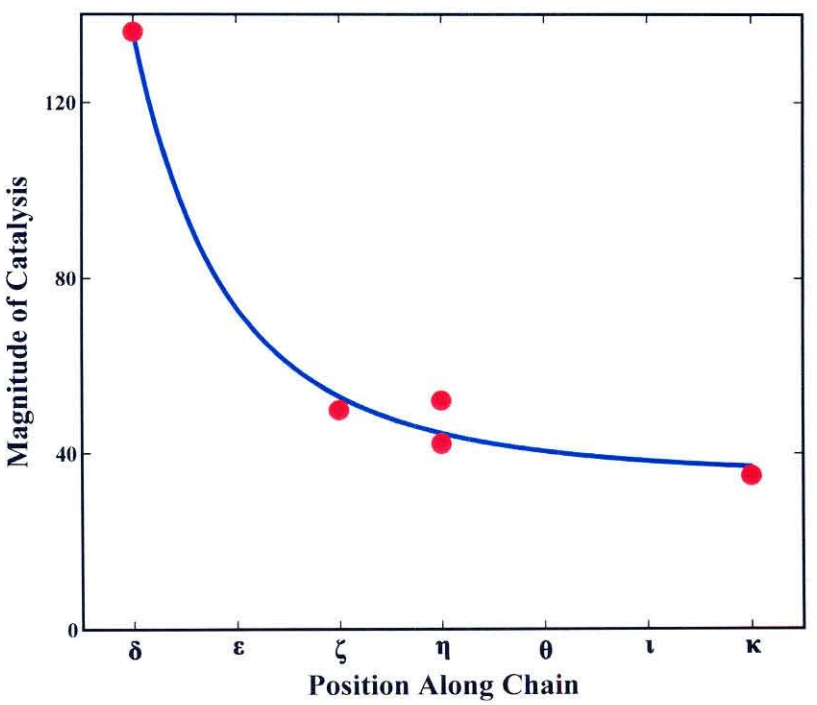


Figure 8-5 – Graph of the NH^+ Catalysis Coefficients for His, Lys, and Arg as a Function of Position Along the Carboxyl-Side Residue Side Chain and Peptide Backbone. Steric Effect of NH is Included.

Ser and Thr contain γ–OH groups. These groups clearly have a catalytic effect in addition to a steric effect, since the net coefficient is 55.6 (Ser is 57.4 and Thr is 53.8, this is the average), while a single hydrogen atom is 102.3 in this position. Assuming a steric effect similar to that of CH, the catalysis is 55.6 – 142.2 = –86.6. How this catalysis occurs is unknown, possibilities include a modification of the backbone nitrogen atom pKa, or transient hydrogen bonds with the deamidating amide, which could keep it in closer proximity to the reaction site. Also, this could be an effect on the structure of water, as could part of the catalytic effects of other residues.

Cys and Met are the only naturally occurring sulfur-containing side chains. Estimation of the sulfur steric effect is complicated by its substantially larger radius than other atoms. We estimate that it is similar to a $–CH_2–$ group. In the case of Cys the net effect of the –SH group is 175.3. Removal of the steric effect gives 175.3 – 178.4 – 36.1 = –39.2 for a catalytic effect. For AmCys (hydrogen atom replaced by acetamidomethyl group) the sulfur atom has a net effect of 77.7 (assuming no special effects from the acetamidomethyl group). Correcting for the steric effect gives 77.7 – 178.4 = –100.7. For Met the net effect for –S– is 5.5. Correcting for its steric effect gives 5.5 – 52.0 = –46.5. In general, there appears to be some catalysis caused by sulfur.

The carboxylic acids, Asp and Glu, both have negative charges as well as hydrogen bonding capabilities. The net effect of the carboxyl is 129.2 for Asp and 18.2 for Glu. Correcting for the steric effect results in a catalytic effect of 129.2 – 106.1 – 2(37.5) = –51.9 for Asp and 18.2 – 37.5 – 2(7.5) = –34.3 for Glu. These catalytic effects are very similar, although a reduction with distance is evident. It is interesting to note that these effects for negative oxygen atoms in the δ and ϵ positions are comparable to those of positive charges in the ϵ and ζ positions, with the same tendency to fall off only very slowly with distance. This is an indication that the catalytic effects may be caused by direct involvement of the side chains, rather than inductive effects.

These increments are listed in Table 8-2. For convenience in computing rates, all values listed include both the steric and catalytic components without separation. Simply subtract

the corresponding hydrocarbon steric effects from the top of the table to isolate catalytic effects. All values listed in bold are computed exactly from experimental rates. The others are computed theoretically. The aromatics are based on the hydrocarbons by the method shown previously to be quite accurate. The –O– values are based on both carboxylic acids and hydroxyl groups which could distort these values, although the progression from 19.5 to 11.5 to –9.6 gives the expected logical progression of –86.6, –26, and –17.1 when steric effects are removed. The –S– value in the γ position is based on AmCys rather than Cys.

Gln Deamidation

This model for Asn succinimide deamidation is applicable to Gln glutarimide deamidation as well. While there may be important differences in the functional group effects for the longer Gln chain, the steric effects are similar as illustrated in Figure 8-6. Figure 8-6a shows the Gln values without hydrolysis correction vs. the Asn values with hydrolysis removed, as listed in Table 8-1. The fundamental difference between succinimide rates vs. glutarimide rates has been removed by subtracting the Gln values from GlyXxxGlnGlyGly and the Asn values from GlyXxxAsnGlyGly.

The difference between GlyXxxGlnGlyGly and GlyXxxAsnGlyGly is 644.8. Much of this difference is probably due to the extra entropy associated with the longer Gln chain. Based on the standard entropy of formation of hydrocarbons (29), the entropy associated with adding an extra -CH_2- is about 40 J/(K mol), which corresponds to about Δ100(ln k) = 480 on our scale. This is 74% of the actual experimental value. Thus the differences between Gln and Asn rates are also based on the probability of the formation of the ring.

When the effect of hydrolysis is removed, the points are scattered around the theoretical 1:1 line in Figure 8-6b. The two-dimensional medians of these values are shown. At the medians of 395 for Asn, the Gln medians are 226 without hydrolysis correction and 361 with hydrolysis correction. Therefore, when amino-side residue effects are removed and carboxyl-side values averaged, the calculated deamidation rates for Asn and Gln peptides differ by only 9%, which is within experimental and computational error.

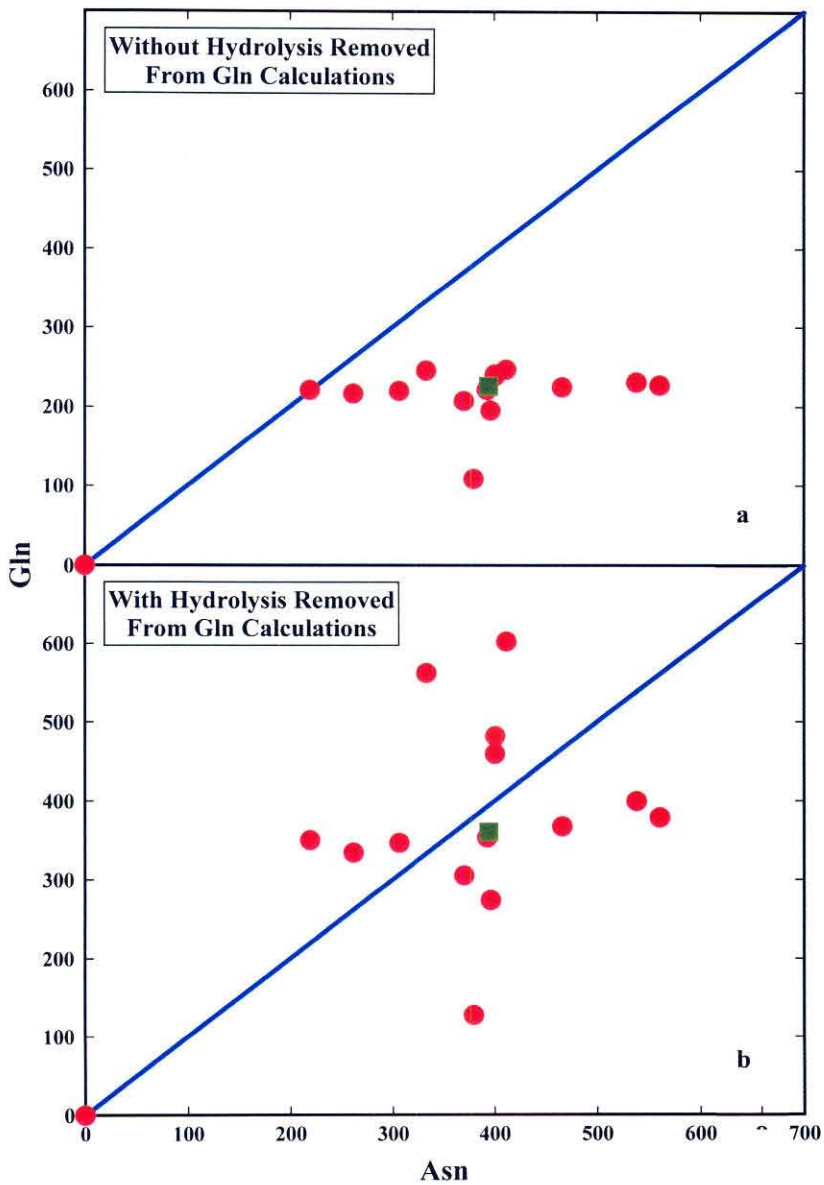


Figure 8-6 – (a) Normalized Gln (100)(ln(k)) without hydrolysis corrections vs. those for Asn (100)(ln(k)) with hydrolysis corrections as listed in Table 8-1. The fundamental difference in succinimide rate vs. glutarimide rate has been removed by normalization, which subtracts the Gln values from GlyXxxGlnGlyGly and the Asn values from GlyXxxAsnGlyGly. (b) As in (a) with both the Asn and Gln values corrected for hydrolysis. The two dimensional medians of the plotted points are indicated by the squares.

The scatter could result from differences in special effects of functional group side chains on the glutarimide vs. the succinimide mechanisms. Substantial scatter is, however, also introduced by the uncorrected and unknown sequence dependence of direct hydrolysis on Gln deamidation. Correction for hydrolysis moves the points from their positions in Figure 8-6a to those in Figure 8-6b, but this single value does not correct for sequence-dependence of hydrolysis.

Prediction of Unknown Deamidation Rates

Using the coefficients derived in the previous sections and applying similar methods to the other amino acids, Table 8-2 was produced. This table can be used for the determination of the primary-sequence determined deamidation rate for *any* sequence for which the carboxyl-side residue–natural or synthetic–can be assembled from substituents found in this table. For convenience, various groups are listed in addition to individual atoms. The coefficients for each of the atoms or groups of atoms in the carboxyl-side residue side chain are summed, using the convention shown in Figure 8-1. Multiply bonded groups should be treated as described for Phe and Trp.

The deamidation half-time is computed as follows

(Eq. 8-29) Asn Peptides–succinimide $t_{1/2} = [(\ln(2))/86400]e^{((\mathbf{Sum/100}) + 11.863)}$

(Eq. 8-30) Gln Peptides–glutarimide $t_{1/2} = [(\ln(2))/86400]e^{((\mathbf{Sum/100}) + 18.311)}$

The computed result for most Asn peptides should be within 10% or less of the experimental value in 37.00 °C, pH 7.4, 0.15 M Tris-HCl buffer. Gln rates are less accurate and are more affected by hydrolysis. For Gln peptides and long-lived Asn peptides, a correction for hydrolysis should be made. Using a hydrolysis value of 8000 days[26]

(Eq. 8-31) $t_{1/2total} = 1/(1/8000 + 1/t_{1/2calculated})$

Table 8-2 – Δ (100)ln(k) Coefficients for Calculating Deamidation Rates*.

	β	γ	δ	ϵ	ζ	η
– H	**0**	**102.3**	**36.1**	**7.2**	0.18	0
– CH_3	**306.8**	**214.5**	**59.2**	**8.1**	0.19	0
– CH_2 –	**204.5**	**178.4**	**52.0**	**7.9**	0.19	0
– CH – –	**102.3**	**142.2**	**44.7**	**7.7**	0.19	0
– – C – –	**0**	**106.1**	**37.5**	**7.5**	0.19	0
– C_5H_5	284.5	**207.4**	52.9	7.9	0	0
– C_5H_5OH	---	**220.6**	---	---	---	---
– C_8H_6N (Indole)	390.6	**239.4**	60.4	8.1	0	0
– $C_3H_3N_2^+$ (Imidazole)	---	**14.9**	---	---	---	---
– S –	---	77.7	**5.5**	---	---	---
– SH	---	**175.3**	---	---	---	---
– O –	---	19.5	11.5	-9.6	---	---
– OH	---	**55.6**	---	---	---	---
– CO_2^-	---	**129.2**	**18.2**	---	---	---
– – NH^+	---	---	-136.0	---	-49.7	-42.1
– $N_3CH_5^+$ (Guanidino)	---	---	---	**-34.2**	---	---
– NH_3^+	---	---	---	---	**-49.7**	---

*Bold-face values based directly on experimental rates and k = sec^{-1}.

Consider for example the sequence GlyXxxAsnThrGly. In order to calculate the deamidation half-time from Table 8-2, it is necessary to consider the Thr side chain. The structure is –$CH(CH_3)OH$. The AsnThr structure is illustrated in Figure 8-7.

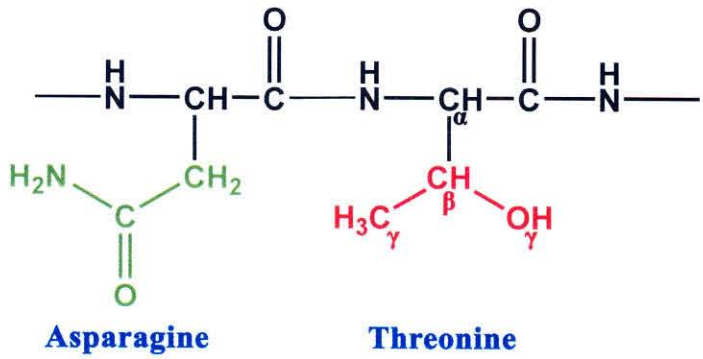


Figure 8-7 – Illustration of AsnThr Sequence

Adding a β-CH, a γ-CH_3, and a γ-OH and computing for Asn in Equation 8-30 and then correcting for hydrolysis in Equation 8-31 yields the result. The experimental value is 46.2 days.[6]

(Eq. 8-32) AsnThr-Suc. $t_{1/2} = [(\ln(2))/86400]e^{((102.3 + 214.5 + 55.6)/100 + 11.863)} =$ **47.2 days**

(Eq. 8-33) AsnThr-Total $t_{1/2} = 1/((1/47.2) + (1/8000)) =$ **46.9 days**

The same procedure can be applied to the non-natural amino acid norleucine in the peptide sequence AsnNle. The norleucine side chain is $-CH_2CH_2CH_2CH_3$ as illustrated in Figure 8-8. Adding the appropriate groups and computing gives

(Eq. 8-34) AsnNle-Suc. $t_{1/2} = [(\ln(2))/86400]e^{((204.5 + 178.4 + 52.0 + 8.1)/100 + 11.863)} =$ **95.6 days**

(Eq. 8-35) AsnNle-Total $t_{1/2} = 1/((1/95.5) + (1/8000)) =$ **94.4 days**

The experimental half-time is unknown, but will be found to be very close to 94.4 days when it is measured.

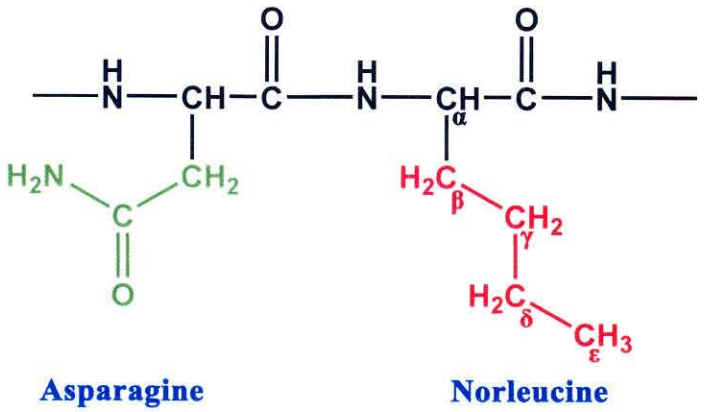


Figure 8-8 – Illustration of AsnNleu Sequence

Verifications

This prediction procedure is based on experimentally determined deamidation half-times. Since some of these calculations are redundant, they provide verification of the method as do additional computations. These verifications include.

1) Almost identical values are obtained for replacement of a γ-H by a γ-CH_3 by comparing Ser to Thr (108.6) and Ala to Val (115.9).

2) Comparison of Ile to Val gives a value of 22.6 for replacing a δ-H by a δ-CH_3. Calculation based on Leu, Ala, and γ-CH_3 results in 23.6.

3) The γ and δ changes in substituting CH_3 for H can be estimated from theory based on CPK modeling. Experimental are 112.3, and 23.1, while theoretical are 102.3, and 17.0.

4) Plot of the cube root of the three experimental values of 306.8, 112.3, 23.1, and of the fourth value of 0.81 obtained from the previous values is a straight line as expected from collision modeling. Therefore, the ratio of volume used to volume available increases with the cube of the chain length. The slopes of these plots

shown in Figure 8-3 closely correspond (–1.38 vs. –1.39 and –2.00 vs. –1.94) to expected values. Figure 8-3b, used to optimize the constant "a" was capable of being straightened to an R^2 of within 10^{-14} of 1.

5) The coefficient value for the phenyl side chain of Phe predicted entirely from its atomic constituents is between 195.5 and 210.5, while the experimental value is 207.3. The hydroxyphenyl of Tyr is similar.

6) The coefficient of the indole side chain of Trp is predicted from its atomic constituents to be 231.9, while prediction based on the Phe experimental rate gives 243.7. The experimental value is 239.4.

7) Median results for Gln and Asn calculations compare favorably.

Buffer and Amino-Side Residue Effects

These calculations were made with deamidation rates recently measured[6, 26, 73] in Tris buffer. Sequence-dependent deamidation rates in the previous scientific literature were measured primarily in phosphate buffer. About 65 of those peptide rates were the rates measured prior to 1974 through which the sequence dependence of deamidation and its approximate range and qualitative determination by steric and catalytic factors was discovered.[25, 3] In the case of Asn, these rates have been reviewed and quantitatively categorized.[158] Unfortunately, however, phosphate strongly catalyses deamidation[3] by general base catalysis of hydrolysis, by the catalysis of the imide mechanism, or by some other as yet undiscovered mechanism. This catalysis introduces a major correction that is, itself, sequence dependent, so phosphate rates are much more difficult to understand with currently available experimental data. The general effects are qualitatively similar to those described herein, but show a large amount of quantitative variation.

Moreover, strong phosphate catalysis also affects deamidation rates in proteins. The quantitative computation of deamidation rates of proteins[7-9] contains about a factor of 2 in uncertainty largely as a result of phosphate-catalyzed measurements. This factor of 2 will be resolved when protein measurements in Tris and other benign buffers or after extrapolation to zero buffer concentrations become more widely available. At present, only one protein, ribonuclease A, has been carefully studied under these conditions.[78] The experimental resolution of this problem will then permit a more accurate calculation of the protein secondary, tertiary and quaternary structure coefficient,[9] C_m.

The measured high-precision deamidation rates for the series GlyXxxAsnGlyGly in Tris were used for this study.[6] These peptides are largely free of both hydrolysis and carboxyl-side effects, and their rates show substantial regularity of Xxx substituent side chain vs. reaction rate. This regularity could, however, reflect inductive effects, direct steric effects, indirect steric effects by thermal energy damping, or indirect effects on the contribution of the charged amino group of the peptide. This charge contribution has been shown in other model peptides.[73] Amino-side effects can also be buffer-ion mediated.

Regardless, therefore, of the regularity of deamidation rates in the series GlyXxxAsnGlyGly, we have chosen not to postulate a specific quantitative amino-side residue explanation at this time. For empirical purposes with peptides, the amino-side effect can be qualitatively estimated directly from the 474 sequence-dependent peptide rates that have been measured[6, 26, 73] in 37.00 °C, pH 7.4, 0.15 M Tris-HCl. In proteins, since the amino-side residue is farther from the reaction center, it is likely to involve, on average, substantial molecular movement to exert its influence and, therefore, to be even more suppressed by secondary, tertiary, and quaternary protein structure than is the carboxyl-side mediated mechanism.

Conclusions

The methods and results described here provide a new quantitative understanding of primary sequence control of deamidation, which has heretofore been understood only qualitatively and empirically. They provide a theoretical foundation that explains the similarities between Asn and Gln deamidation of peptides.

Deamidation can be visualized as a billiard game in which all of the atoms in a particular peptide are in constant motion. The goal is for the amide to reach the reaction site regardless of obstacles that may deflect it from its path. Any other atom in the vicinity of the reaction site can serve as such an obstacle and will be more significant the closer it is tethered to that site. Charged side chains and other functional groups can effect the reaction in special ways–either helping the amide reach its goal or hindering it, depending on the circumstances. They may also affect the acidity of the backbone hydrogen atom or the structure of water. This game continues for each molecule at a very high speed, until the correct configuration is reached. Then the molecule deamidates. Quantitative analysis by means of the coefficients in Table 8-2 provides the following.

1. Improved understanding and simplification of primary sequence control of deamidation.

2. Facilitated protein engineering design of special sequences–both with ordinary and synthetic amino acid residues–that have desired deamidation rates. Rates can be predicted by simply assembling pieces from a table of known constants–in the same way that the NMR structure of a new compound can be estimated based on its constituents.

3. Improved procedures for three-dimensional deamidation rate computations. It aids in estimating deamidation rates for unusual residues for which primary rates have not yet been determined, and it suggests improvements in the three-dimensional computation procedure itself.

Chapter 9

PREDICTION OF PROTEIN DEAMIDATION RATES FROM 3-D STRUCTURE

Introduction to Method

In principle, the deamidation of amides in proteins is completely a three-dimensional phenomenon. Any factor that can influence the three-dimensional structure around an amide can have an effect on the deamidation rate of that amide. These factors include primary sequence as well as secondary, tertiary, and quaternary structure elements. Primary sequence rates are simply one component of a three-dimensional problem.

While the deamidation rate of a peptide is still controlled by the three-dimensional structure of the peptide, the correct choice of sequence simplifies the problem to the point where it can be easily measured and understood and provides values that are generally common to most peptide and protein sequences. The pentapeptide models on which the primary sequence rates were measured were designed with this in mind. They are generally too short to form secondary structures, yet long enough to diminish the effects of the charges on the ends. The choice of a Gly residue on either end reduces the possibility of side chain interaction beyond that of the nearest neighbors, and the choice of Tris buffer keeps intermolecular effects to a minimum. The result is that the pentapeptide primary rates are suitable for protein work.

The real environment to which most proteins are exposed is far more complex. In addition to the three-dimensional structure of the protein itself, there are interactions with other proteins, variations in pH and sometimes temperature, a multitude of solution ions of all sorts which can vary in type and concentration from one location to the next, and possible changes in any of these parameters over time. All of these confounding factors complicate the understanding of *in vivo* as well as *in vitro* protein deamidation, but the

primary structure rate measurements provide a foundation on which to build an understanding of these processes.

As it turns out, the effects of primary sequence and protein secondary, tertiary, and quaternary structure on the deamidation rates of proteins are comparable in size,[7] although in many cases one or the other dominates. Additionally, it turns out that the main effect of protein structure is to slow down the primary sequence deamidation rate. This simplifies the theoretical problem. It is only necessary to determine the nature of and fractional contribution of secondary, tertiary, and quaternary structure to the primary-structure-determined peptide rates. Since we now know from experiment the absolute primary structure rates, a correct evaluation of the amount of slowing from three-dimensional structure gives the absolute protein deamidation rate.

It is important to realize that the calculation procedure we developed (as described below) was *not* calibrated by comparing the results to known deamidation *rates* in proteins. First, the available data would have been unsatisfactory. There are far more reports in the literature of amides which have been qualitatively observed to deamidate than quantitative measurements of actual deamidation rates. Also, the few rates that do exist in the literature are under a wide variety of temperatures, pH's, and buffers, which makes comparison among proteins difficult. Even with only relative qualitative instabilities within a protein as criteria, we were only able to find 31 proteins with both qualitatively observed deamidation of identified Asn and 3-D structures which were suitable for use in our calibration set.

Second, by excluding all measured rates except the primary sequence pentapeptide rates, our absolute predictions–which are very close to known deamidation rates in proteins that have been measured–are more useful. The only quantitative experimental values upon which they depend are the very accurately measured peptide rates. It is satisfying that application of our procedure to several instances of relative deamidation of amides in proteins that we later discovered in the literature produced the correct predictions and that our procedure also gives very good agreement with the 13 known applicable protein absolute deamidation rates, even though these absolute rates were not used to calibrate it.

Our optimization was based on the assumption that, despite the variation in reaction conditions, the amides in proteins that have been reported to deamidate are generally the fastest ones in their respective proteins. The goal was to select parameters and optimize them, so that these amides would be chosen as the fastest ones based solely on rank.

Selection of Literature Rates for Calibration of Procedure

All reports of Asn deamidation in proteins wherein investigators identified the specific deamidating Asn residue were gathered from the Medline and Citation Index databases. The Protein Data Bank (http://www.rcsb.org/pdb) was then searched for a 3-D structure that was identical in protein biological type and primary sequence to each protein in which deamidation had been reported. Every protein for which we found a suitable deamidation report and a corresponding 3-D structure is included herein. None were omitted.

In addition, 44 human hemoglobin mutations that convert another residue into Asn and 16 mutations that change the residue on the carboxyl side of one of the 10 wild-type Asn residues have been reported. This set of 70 Asn residues, of which 7 have been reported to deamidate, is included. Only one hemoglobin structure was used. It was assumed that the hemoglobin structure would not change significantly due to a single mutation, so the modifications were made by substituting the appropriate residue in the hemoglobin structure. This approximation is probably correct for most of these mutations.

The selected proteins and their Protein Data Bank identification numbers are rabbit aldolase[12, 94] 1ADO, human angiogenin[95, 96] 1B1I, bovine calbindin[97, 98] 4ICB, pig cAMP-dependent protein kinase[99, 100] 1CDK, horse cytochrome c[63, 101] 2GIW (NMR), mouse epidermal growth factor[102, 103] 1EGF (NMR), rat fatty acid-binding protein[104, 105] 1LFO, human fibroblast growth factor[106, 107] 2AFG, *Aspergillus awamorii* glucoamylase[108, 109] 3GLY, human growth hormone[110, 111] 1HGU, human hemoglobin[112-120] 1A3N, *Escherichia coli* Hpr-phosphocarrier protein[121-123] 1HDN (NMR), human

hypoxanthine guanine phosphoribosyl transferase[124, 125] 1BZY, human insulin[126, 127] 2HIU (NMR), mouse interleukin 1β[128, 129] 2MIB, human interleukin 2[131, 132] 3INK, chicken lysozyme[133, 134] 1E8L (NMR), bovine ribonuclease A[135, 136] 1AFK, *Ustilago sphaerogena* ribonuclease U2[137, 138] 1RTU, bovine seminal ribonuclease[139,140] 11BG, human T cell surface protein CD4[141, 142] 1CDJ, human thioltransferase[143, 144] 1JHB (NMR), human triosephosphate isomerase[145, 146] 1HTI, and bovine trypsin[147, 148] 1MTW.

Trypsin is included, but the reported[147] relative Asn instabilities are unsuitable. Trypsin was incubated in solution for 1 year while the solution was differentiated through crystal growth into a homogenous fraction that exhibited deamidation at three positions. No deamidation observations of undifferentiated solutions were reported.

The proteins listed above were used for the development of the calculation method.[7] Soon after its development the procedure was applied to 126 human proteins.[8] It was then computerized and applied to the Protein Databank.[9] For this later work the protein calibration set was further refined by removal of Human T cell surface glycoprotein CD4 (because the protein in which deamidation was observed was different from that used for the crystal structure), and the addition of 4 new proteins. These new ones were bovine DNase I[149] 2DNJ, human hirudin[150] 4HTC, bovine calmodulin[151] 1A29, and human vascular endothelial growth factor[152] 2VPF.

Equation for Calculation of C_D Values

Each of the parameters measured during the analysis was assumed to have some effect on the Arrhenius activation energy for the reaction. In most cases this is a positive value. The known mechanism for AsnGly deamidation involves formation of a succinimide ring intermediate (see Figure 1-5). Based on our measurements of primary sequence rates, this mechanism is probably applicable to most Asn deamidation rates. Most of the parameters that we use can be explained in terms of hindering formation of this ring. The success of our method does not, however, confirm this mechanism because the parameterized secondary, tertiary and quaternary structure factors may affect other mechanisms similarly.

The rate constant for primary structure is related to the Arrhenius activation energy. This equation is

(Eq. 9-1) $$k_P = Ae^{-E_P/RT}$$

Substituting an extra term for 3-D activation energy we get the total rate constant (k_T) as shown below

(Eq. 9-2) $$k_T = Ae^{-(E_P + E_{3D})/RT} = A\left(e^{-E_P/RT}\right)\left(e^{-E_{3D}/RT}\right)$$

Substituting the primary rate term produces the following equation

(Eq. 9-3) $$k_T = k_P\left(e^{-E_{3D}/RT}\right)$$

The calculations were done in terms of deamidation half-time. Furthermore we define a new constant, the deamidation coefficient, C_D. By definition C_D is 100 times the predicted half-time. This makes the numbers more reasonable in the region we are studying (the rates in the test set vary from a half-time of a few days to several hundred days). The conversion between rate constant and half-time is

(Eq. 9-4) $$t_{1/2} = ln(2)/k$$

Substituting this into Equation 9-3 for both k_T and k_P and converting to C_D gives

(Eq. 9-5) $$C_D = (0.01)t_{T1/2} = (0.01)t_{P1/2}\left(e^{E_{3D}/RT}\right)$$

This is the equation used to combine the primary half-time with the three-dimensional modifications. For this work, the temperature is assumed to be constant since the primary rates are at 37 °C, and we wish to predict rates at 37 °C. Since R is also a constant, we are not measuring activation energy directly. We simply ignore everything expect the general

functional form. This reduces the equation to the one that is used with our prediction procedures

(Eq. 9-6) $$C_D = (0.01)t_{P1/2}\left(e^{f(C_m, C_{S_n}, S_n}\right)$$

The functional form of $f(C_m, C_{sn}, S_n)$ is described in more detail below. It is simply a weighted sum of the secondary, tertiary, and quaternary effects for that residue.

It is easy to suggest reasons that this equation may not always be applicable. One could argue, for example, that, depending on three-dimensional structure, there could be more than one hump in the reaction energy diagram and that the activation energies cannot therefore be added. Similarly the succinimide may not be the only mechanism for deamidation. We know from our work, however, that this method works in at least 95% of cases. The parameters we are measuring are generally restraining ones and would probably be similar even in the case of a different mechanism.

Development of Parameters for Quantifying Effects

A set of observations of the 3-D environment of each Asn was selected. This set included positions with respect to α-helical or β-sheet regions, hydrogen bonds to the Asn, other hydrogen bonds inhibiting formation of a succinimide intermediate, and relative freedom of the Asn peptide backbone. These observations were made and tabulated before any calculations were carried out. The tabulated observations were not changed after calculations began, although some parameters which were tabulated were found to be redundant and were not ultimately used.

The observations were made with Swiss Protein Data Bank Viewer software, StereoGraphics ENT B and CE-3 viewer hardware (StereoGraphics Corp., San Rafael, CA) and a Pentium III computer running Microsoft NT 4.0.

It is worth noting that, while in a system of this type it is required to make the observations before trying to fit them to the data, the observations are not dependent upon this for this particular system once the general method has been established. It is very difficult to bias most of these parameters. Only the additional hydrogen bonds parameter (S_9 below) is somewhat arbitrarily assigned and even here the value is obvious in most cases. The observations were, however, made before optimization and never changed as a matter of rigor.

The deamidation coefficient, C_D, is defined as $C_D = (0.01)t_{P1/2}\left(e^{f(C_m, C_{S_n}, S_n)}\right)$, where $t_{1/2}$ is the pentapeptide primary structure half life,[6] C_m is a structure proportionality factor, C_{Sn} is the 3-D structure coefficient for the nth structure observation, S_n is that observation, and $f(C_m, C_{Sn}, S_n) = C_m[(C_{S1})(S_1) + (C_{S2})(S_2) + (C_{S3})(S_3) - (C_{S4,5})(S_4)/(S_5) + (C_{S6})(S_6) + (C_{S7})(S_7) + (C_{S8})(S_8) + (C_{S9})(S_9) + (C_{S10})(1 - S_{10}) + (C_{S11})(5 - S_{11}) + (C_{S12})(5 - S_{12})]$. The structure observations, S_n, were selected as those most likely to impede deamidations, including hydrogen bonds, α helices, β sheets, and peptide inflexibilities. The functional form of C_D assumes that each of these structural factors is added to the reaction activation energy. The observed S_n were

For Asn in an α-helical region

S_1 = distance in residues inside the α-helix from the NH_2 end, where $S_1 = 1$ designates the end residue in the helix, 2 is the second residue, and 3 is the third. If the position is 4 or greater, $S_1 = 0$.

S_2 = distance in residues inside the α-helix from the COOH end, where $S_1 = 1$ designates the end residue in the helix, 2 is the second residue, and 3 is the third. If the position is 4 or greater or $S_1 \neq 0$, then $S_2 = 0$.

$S_3 = 1$ if Asn is designated as completely inside the α-helix, because it is 4 or more residues from both ends. If the Asn is completely inside, $S_3 = 1$, $S_1 = 0$, and $S_2 = 0$. If $S_1 \neq 0$ or $S_2 \neq 0$, then $S_3 = 0$.

For flexibility of a loop including Asn between two adjacent antiparallel β sheets

S_4 = number of residues in the loop.

S_5 = number of hydrogen bonds in the loop. $S_5 \geq 1$ by definition.

For hydrogen bonds

S_6 = the number of hydrogen bonds to the Asn side chain C=O group. Acceptable values are 0, 1, and 2.

S_7 = the number of hydrogen bonds to the Asn side chain NH_2 group. Acceptable values are 0, 1, and 2.

S_8 = the number of hydrogen bonds to the backbone nitrogen atom in the peptide bond on the COOH side of Asn. Hydrogen bonds counted in S_6 or S_7 are not included. Acceptable values are 0 and 1. This nitrogen atom is used in the five-membered succinimide ring.

S_9 = additional hydrogen bonds, not included in S_6, S_7, and S_8, that would need to be broken to form the succinimide ring.

For Asn situated so that no α-helix, β-sheet, or disulfide bridge structure is between the Asn and the end of the peptide chain

S_{10} = 1 if the number of residues between the Asn and the nearest such structure is 3 or more. If the number of intervening residues is 2, 1, or 0, or Asn not between structure and chain end, then S_{10} = 0.

If the Asn lies near to any α-helix, β-sheet, or disulfide bridge structures

S_{11} = the number of residues between the Asn and the structure on the NH_2 side, up to a maximum of 5. Values of 0, 1, 2, 3, 4, and 5 are acceptable.

S_{12} = the number of residues between the Asn and the structure on the COOH side, up to a maximum of 5. Values of 0, 1, 2, 3, 4, and 5 are acceptable.

Hydrogen bonds selected by the Swiss Protein Data Bank (PDB) viewer were accepted if the bond length was 3.3 Å or less and there was room in the structure to accommodate the van der Waals radius of the hydrogen. This treatment of H-bonds was improved in the later fully-computerized procedure. The Swiss PDB viewer, according to the customary criteria, selected α helices and β sheets. All primary structure $t_{1/2}$ values were those published,[6] except for Asn with carboxyl-side Pro, Asn, or Gln and N-glycosylated Asn. We used estimated values for $t_{1/2}$ of 500, 40, 60, and 500 days for AsnPro, AsnAsn, AsnGln, and N-glycosylated Asn, respectively. At the time this was done, the AsnPro rates had not yet been determined.

Coefficients Used in Equation

C_D values ("Coefficient of Deamidation") were optimized[153, 154] by using various values for C_m and C_{Sn} to maximize the value of the deamidation resolving power, D_P, as described in the calibration procedure section. The optimized values were $C_m = 0.48$, $C_{S1} = 1.0$, $C_{S2} = 2.5$, $C_{S3} = 10.0$, $C_{S4,5} = 0.5$, $C_{S6} = 1.0$, $C_{S7} = 1.0$, $C_{S8} = 3.0$, $C_{S9} = 2.0$, $C_{S10} = 2.0$, $C_{S11} = 0.2$, and $C_{S12} = 0.7$.

For example, the β-LysAsn(145)His sequence of hemoglobin is not in an α-helix or in a loop between two β sheets, so S_1 through $S_4 = 0$, $S_5 = 1$. There is one hydrogen bond to the amide side chain nitrogen and one other to be broken to form the imide, but there are none to the amide carboxyl or the backbone nitrogen, so $S_6 = 0$, $S_7 = 1$, $S_8 = 0$, and $S_9 = 1$. This Asn is near the carboxyl end of the chain and one residue from an α-helix on the amino side, so $S_{10} = 0$, $S_{11} = 1$, and $S_{12} = 5$. The GlyLysAsnHisGly half life[6] is 10.5 days. Therefore, $C_D = (0.01)(10.5)e^{(0.48)[(1)(1)+(2)(1)+(2)(1-0)+(0.2)(4)]} = (0.105)e^{(0.48)(5.8)} = (0.105)(16.184) = 1.70$.

Figure 9-1 shows a detailed picture of this analysis for Asn(67) in bovine pancreatic ribonuclease A. The experimental deamidation rate of this amide in this protein in 37 °C, pH 7.4, 0.015 M Tris buffer is known to be 64 days[6]. The amide is not in an alpha helix, so S_1 through $S_3 = 0$. The loop between beta sheets is 7 residues long and contains eleven

hydrogen bonds. Thus $S_4 = 7$ and $S_5 = 11$. There is one hydrogen bond in each category so $S_6 = 1$, $S_7 = 1$, $S_8 = 1$, and $S_9 = 1$. The amide is not near the chain end, so $S_{10} = 0$. The nearest fixed structure is three residues away on the left side and two residues away on the right side, so $S_{11} = 2$, and $S_{12} = 3$. The GlyLysAsnGlyGly half life[6] is 1.02 days. Therefore, $C_D = (0.01)(1.02)e^{(0.48)[(-0.5)(11/7)+(1)(1)+(1)(1)+(1)(3)+(1)(2)+(0.2)(5-4)+(0.7)(5-3)]} = (0.0102)e^{(0.48)(7.8143)} = (0.0102)(42.56) = 0.669$.

C_D is multiplied by 100 to give the predicted Tris deamidation half-time in days for the amide. In this case (0.669)(100) = 66.9 or 67 days for the final computerized procedure.[9] The experimentally observed half-time is 64 days. The initial hand-calculation procedure gave 85 days.

While our calculation method does not always provide such close agreement with experiment, it is generally excellent. Note that neither the absolute rate of deamidation of ribonuclease A, nor that of any other protein was used in this calculation or in the determination of its adjustable constants. The actual ribonuclease rate was not considered in any aspect of the calculation development (we actually did not know this rate at the time). Only the identities of the fastest amides in 31 proteins observed separately under varying solvent conditions were utilized in calibration of the constants. Ribonuclease A was one of these 31, but the constants are essentially the same even when it is excluded.

In the case of ribonuclease A, the sequence-determined rate of this amide is approximately 1 day, so the determination of secondary, tertiary, and quaternary structure effects needed to be quite good.

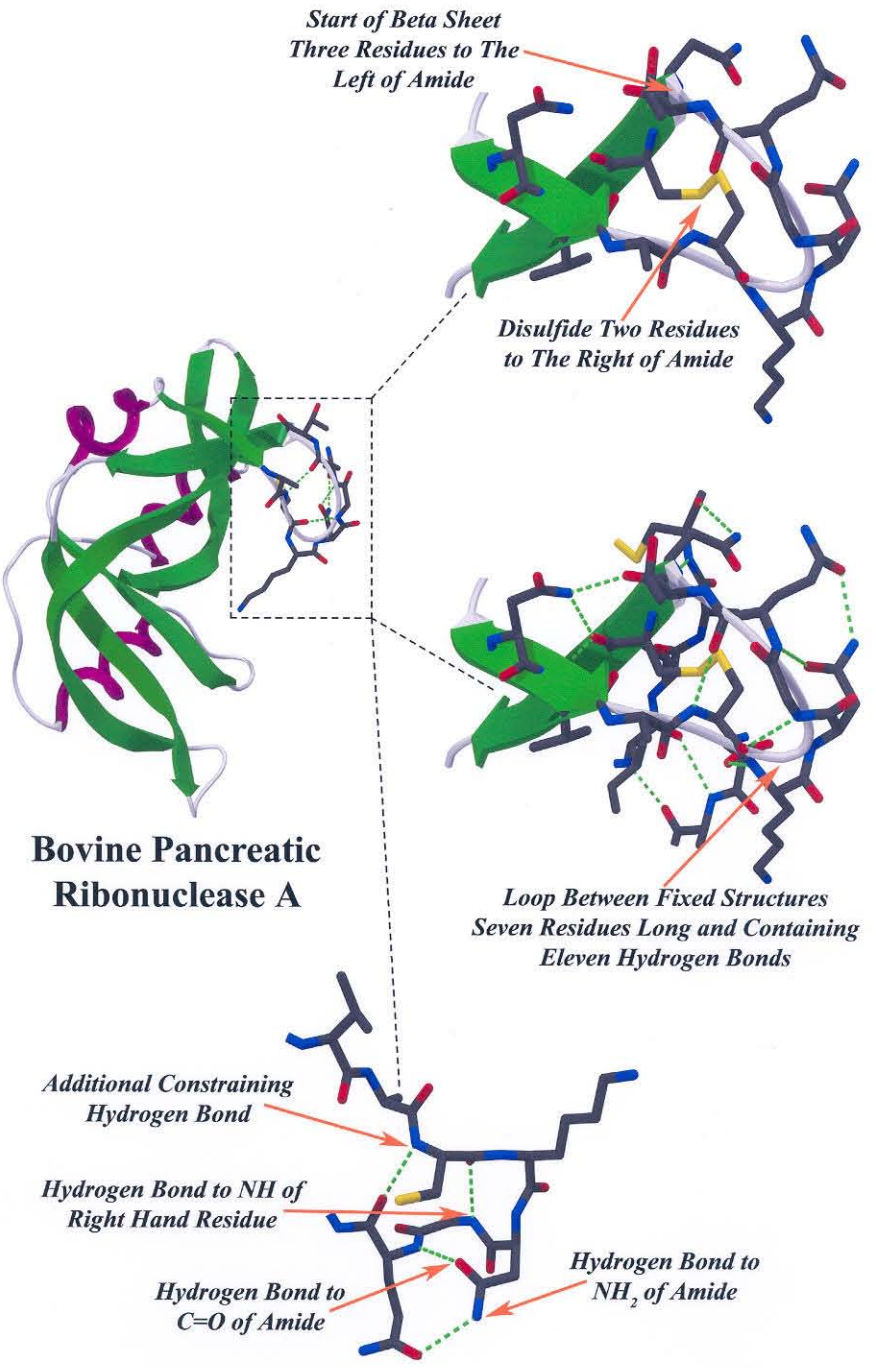


Figure 9-1 – The Deamidation of LysAsn(67)Gly in Ribonuclease-A

In the case of rabbit muscle aldolase, the predicted half-time was 22 days and the actual half-time was 9 days. The deamidation rate of Asn(360) in rabbit muscle aldolase is known to be primary sequence controlled. There is essentially no suppression of the primary sequence rate by three-dimensional structure.

Our calculations agree with this in every way, except for one hydrogen bond that is present in the crystal structure as shown in Figure 9-2. This hydrogen bond results in a predicted rate that is slower then the *in vivo* rate. It is likely that this hydrogen bond is only present in the crystal structure and not in solution. Nevertheless, the predicted rate is close enough to the known experimental rate for most purposes.

Determination of I_D Values

It is often useful to know the initial, cumulative amide deamidation half-time of an entire protein in the case where more than one amide may be deamidating. We define this value as I_D, the "Deamidation Index." Since C_D is inversely proportional to the rate constant, I_D can be calculated as follows

(Eq. 9-7)
$$I_D = \left[\Sigma\left(C_{D_n}{}^{-1}\right)\right]^{-1}$$

Here C_{Dn} is the C_D value for the nth Asn residue in the protein. The deamidation index is a measure of overall protein stability towards deamidation. It can be experimentally measured using electrophoresis or isoelectric focusing of the entire protein without identifying individual amides.

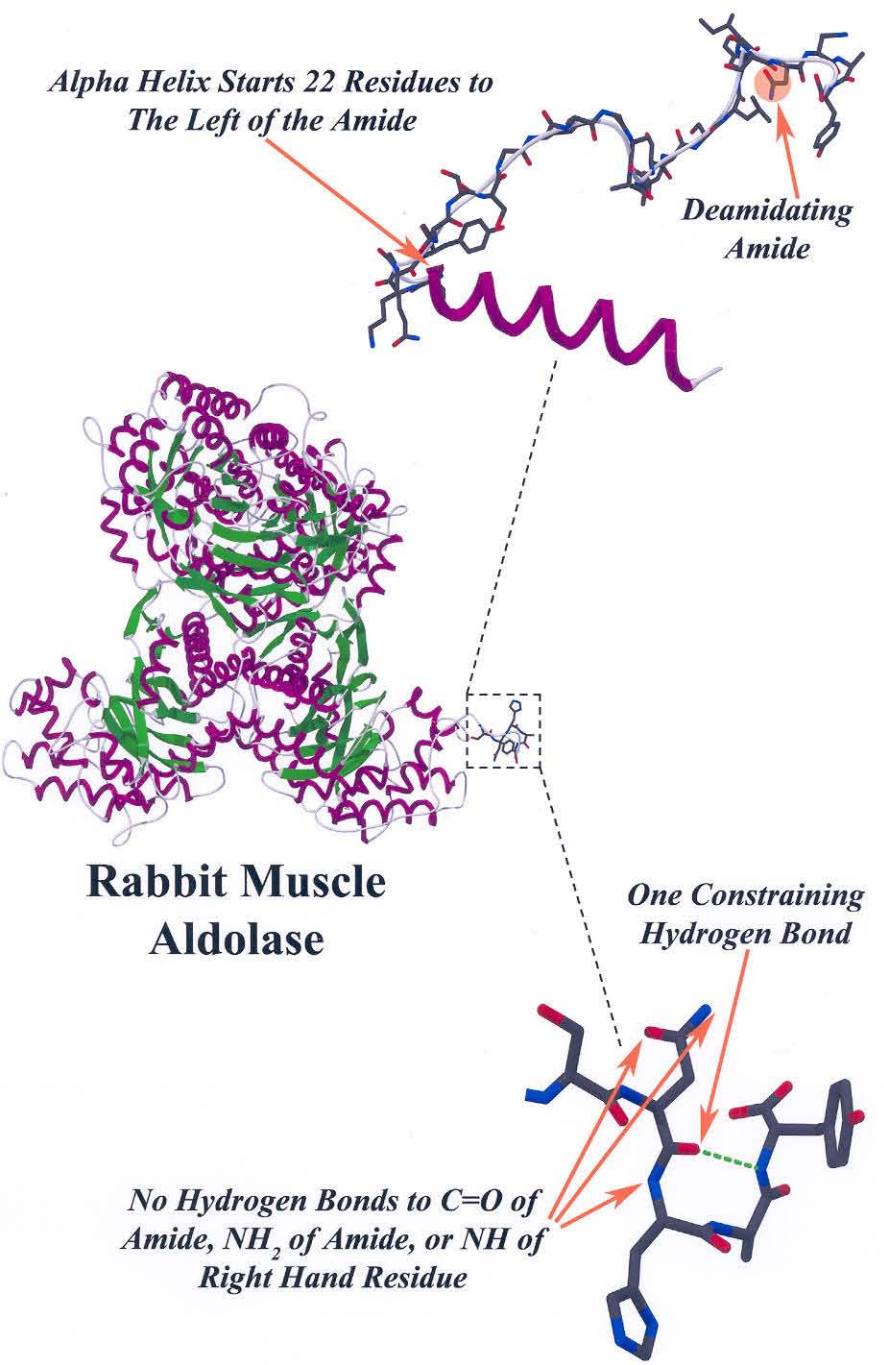


Figure 9-2 – The Deamidation of SerAsn(360)His in Rabbit Muscle Aldolase

Calibration Procedure

Having chosen a set of parameters and measured them in our set of 31 proteins, it was necessary to determine the optimum values for the coefficients C_m and C_{S1}–C_{S12}. The optimized values for these coefficients have already been given in the "Coefficients Used in Equation" section.

Our initial guesses based on intuition were used as a starting point. The final optimized values were usually within a factor of two of these guesses. The next step was to devise a quantitative procedure that would measure the extent of separation between the fast and slow amides. For this we used a D_P calculation method that was initially developed for use in diagnostic medicine.[153, 154]

The predicted C_D values were placed in order from smallest predicted C_D value to largest predicted C_D value. If the prediction method and experimental observations were perfect, all of the amides known to be fast would be at the top of the list, while all of the slow amides would be at the bottom of the list. Experimental variations affect the value of the highest D_P that can be achieved, but the optimization of the prediction coefficients is largely independent. Table 9-1 shows one such list for illustration. A gray box is placed around each amide known to have a fast deamidation rate. Note that this is the optimized list. The initial ordering was not as good as the one shown here and improved with each improvement in the coefficients.

The ordering in Table 9-1 represents a single axis along which the amides have been placed in order of C_D value. Once the order has been established, the numbers are ignored, and we look only at the relative positions along the axis with each amide classified as either fast (the ones in the gray boxes) or slow (all the rest of the amides). Of course there is a wide range in deamidation rate and fast and slow are only qualitative measurements, but we are not using absolute values, so this is the best we can do.

Table 9-1 - Ordered List of Deamidation Coefficients

0.03	1.18	6.11	16.78	42.36	81.53	146.93	355.12	1080.78
0.04	1.28	6.30	18.51	42.93	85.73	158.67	359.92	1215.47
0.06	1.38	6.31	18.66	44.41	89.07	163.60	369.77	1261.63
0.08	1.53	7.31	18.89	45.38	89.07	164.62	395.41	1302.66
0.14	1.57	7.43	19.52	45.64	89.47	169.91	412.55	1309.55
0.18	1.62	8.01	19.54	47.33	92.68	176.70	413.58	1312.63
0.19	1.64	8.19	20.35	47.39	98.18	179.94	429.93	1344.23
0.21	1.69	8.33	20.79	49.03	99.66	189.16	433.03	1394.02
0.21	1.70	8.95	21.10	50.48	99.85	189.46	441.01	1586.97
0.21	1.70	9.01	21.35	51.14	100.00	190.41	464.78	1610.30
0.22	1.73	9.21	21.54	51.46	106.10	191.99	495.93	2105.20
0.26	1.77	10.56	22.01	53.95	107.81	197.07	498.12	2154.93
0.29	1.98	11.09	22.38	55.12	114.83	200.56	536.57	2312.40
0.31	2.10	11.46	22.78	55.12	115.39	202.38	549.84	2360.14
0.35	2.48	11.48	22.80	55.37	115.39	208.03	564.14	2362.13
0.38	2.53	11.53	23.29	58.63	119.39	223.09	581.00	2690.29
0.42	3.06	11.68	23.82	60.09	125.44	239.43	581.83	2830.03
0.42	3.08	11.86	26.65	60.63	126.94	239.61	599.51	
0.57	3.99	12.12	29.76	62.45	126.94	240.60	701.65	
0.58	4.07	12.33	29.78	66.23	128.21	244.68	710.86	
0.64	4.28	12.35	30.22	68.55	128.73	247.20	723.66	
0.67	4.37	12.37	31.47	69.49	130.77	247.20	748.74	
0.68	4.55	12.54	32.71	70.82	130.78	251.35	748.81	
0.78	4.60	12.68	33.06	71.55	130.95	268.74	779.55	
0.81	4.73	12.95	33.11	72.92	132.49	273.69	787.29	
0.85	4.82	13.01	35.04	74.82	133.50	274.38	789.80	
0.93	4.89	13.06	35.70	75.48	135.35	274.60	807.45	
0.93	4.92	13.06	35.87	76.42	141.04	276.96	891.97	
0.96	4.92	14.37	36.82	77.08	141.04	290.79	936.18	
1.14	5.39	14.49	39.25	77.72	141.30	305.00	978.84	
1.17	5.71	15.20	40.71	78.07	144.60	307.72	1076.14	
1.17	6.06	16.09	40.92	80.51	145.04	345.57	1077.46	

Gray boxes indicate amides known to deamidated.

In the ideal circumstance with a perfect calculation method and perfect experimental measurements, the slowest fast amide would be faster than the fastest slow amide–a situation which is close to reality in this circumstance but certainly not perfect. By combining all of the measurements into one list, we are mixing together all of the different proteins. These experimental observations were carried out in many different laboratories under a wide variety of conditions with variations in temperature, pH, buffer type and other

properties that prevent this from being an ideal system. There is also the possibility that, though an observation of a fast amide is not likely to be in error, some of the "slow" amides have simply not yet been discovered to be fast, since "slow" depends upon a negative experiment. All of these uncertainties contribute to noise in the system and make a perfect separation unlikely no matter how good the calculation method.

Now that the values have been ordered for a particular set of coefficients, the column can simply be cut at some point, and we can count the number of fast and slow amides on either side. Calculations of the ordering in Table 9-1 are done to determine the percent errors in classification made by cutting the columns just above each entry and counting the number of correct and incorrect assignments above and below the cut. We then plot these numbers for all possible division points as shown in Figure 9-3.

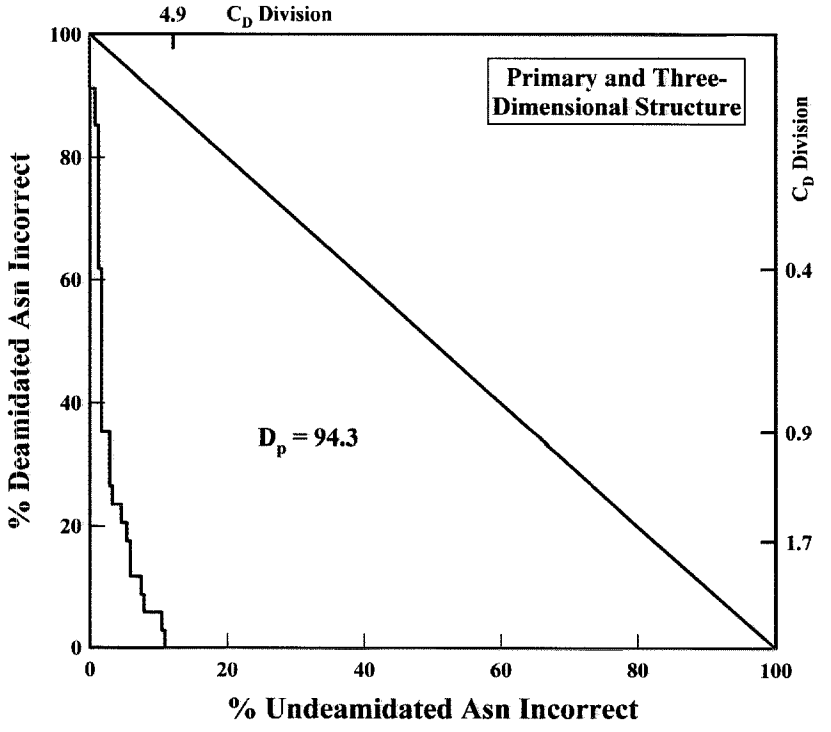


Figure 9-3 – D_P Plot for Optimized 3-D Structure Method.

The first Y value is always 100 and the first X value is always 0 since no fast amides are in the fast section (100% missed) and all the slow ones are in the slow section (0% missed).

If the classification were perfect, the errors in the fast section would diminish as we proceed down the column while the slow section would remain at 0% until we reached the division point between fast and slow amides. At this point the errors in both sections would be 0% since the division would be perfect. Proceeding beyond this point would leave the Y value at 0 and the X value would gradually increase to 100 at the end of the column. The area between the random assignment diagonal shown in the figure and the plot (which would simply follow the X and Y axis) would be 10000, which we divide by 100 to get a D_P of 100%. Conversely, if the calculation were valueless, the plot would follow the diagonal line.

Figure 9-3 for our optimized parameters shows that the division is not perfect. Errors accumulate in the slow section before they have all been removed from the fast section and the value of D_P is only 94.3. For a completely random distribution, D_P would be approximately zero.

This method is illustrated graphically in Figure 9-4. Here the values are ordered along a single axis and the fast and slow amides are given different symbols. Two example points are calculated.

With this method, the optimization procedure was carried out by adjusting the individual parameters and maximizing the value of D_P.

There is the danger in a multivariable system of finding a local maximum in the optimization, while there is a much better maximum that cannot be reached except with a significant change in a particular combination of variables. This was guarded against in two ways. First, the value of 94.3 (later improved to 96.5 in the computerized procedure) is about as high as can be expected given the many uncertainties in the experimental data set, so we know the method cannot be optimized much further. Second, the parameters all

relate to physical data that should intuitively fall within certain limits. We know, for example, that the coefficients of the different types of hydrogen bonds should not be too far different from each other. In fact, the initial guesses of the parameters before optimization produced a surprisingly good result.

This optimization procedure was ideally suited to this system. It allowed us to quantitatively measure the predictive ability of the method, without reference to absolute rates–which were unavailable in sufficient numbers–simply by dividing the experimental data set into two classes, amides reported to be deamidated, and amides not reported to be deamidated. Despite the simplicity of the procedure, it permitted us to optimize our method to a high degree of accuracy. Once optimized, it correctly predicts the amount of three-dimensional suppression of the primary sequence rates. Since the primary sequence rates are based on experiments in 37.00 °C, pH 7.4, 0.15 M Tris buffer, this also produces absolute deamidation rates for amides in proteins under these conditions.

The graph to the right shows the percent errors in classification made by cutting the axis at each possible position.

The normalized area between the diagonal line and the plot (D_p) would be 100 for a perfect division of fast and slow amides and 0 for a completely random distribution.

Two example cuts are shown. The percentages calculated are are based on the number not in the set compared to the total. In the left slice, 12 of the 34 fast amides were missed, giving 35% on the y axis. Only 4 of the 229 slow amides were not put in the slow set, so this is 1.7% on the x axis. This is repeated at all 272 possible positions to produce the plot shown.

The experimental observations were made over a wide variety of pH, temperature and buffer types and probably do not include all deamidating residues. Therefore, a Dp score of 100 may be impossible.

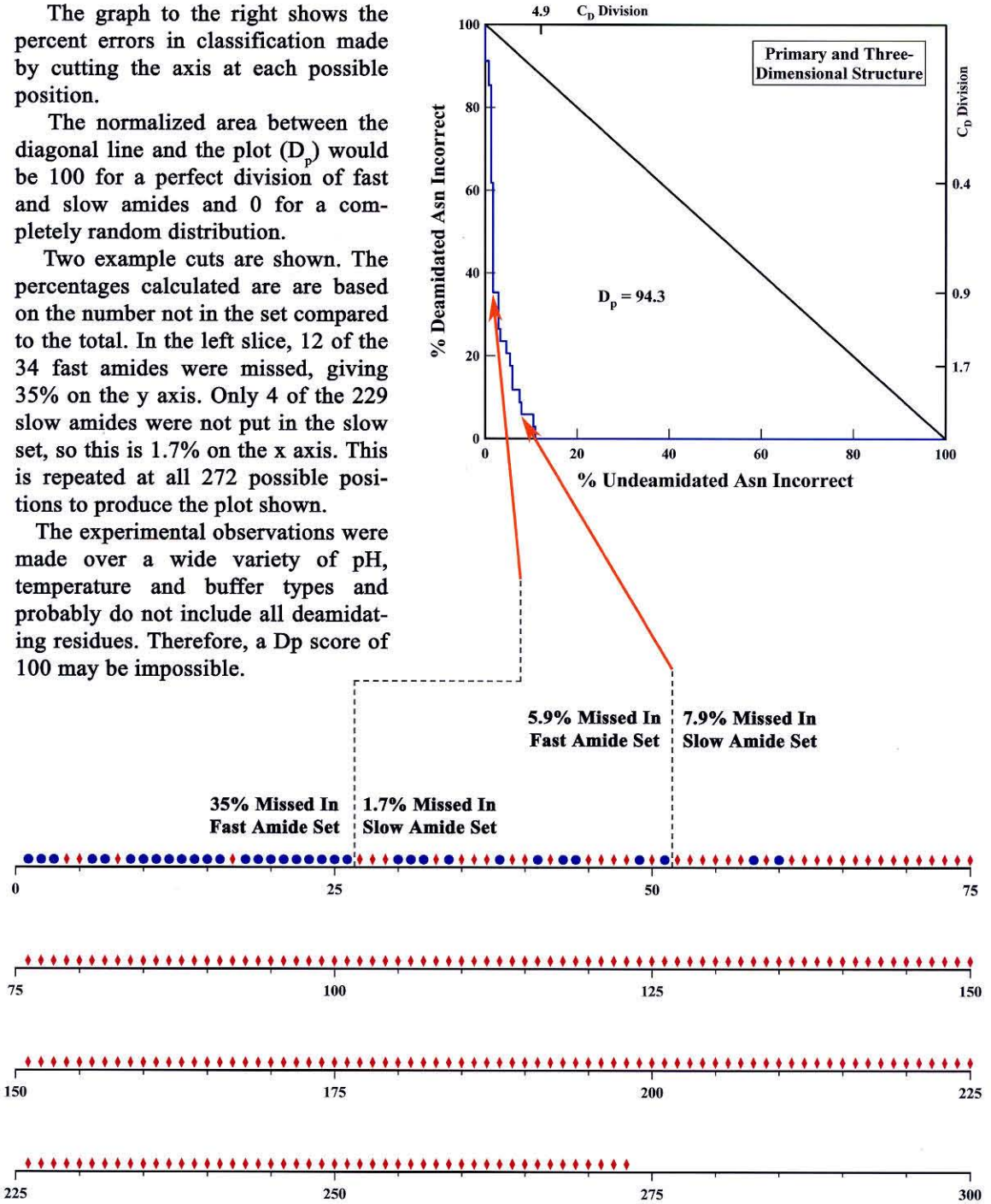


Figure 9-4 – Predicted C_D Values for 273 Amides Ordered By Value – Deamidation Has Been Experimentally Observed Only for Those Amides Indicated by Blue Circles

Results and Analysis of Hand Prediction Procedure

Table 9-2 is a summary of the C_D values for the individual amides in the proteins. Table 9-3 lists the C_D values for the hemoglobin mutants.[7, 130] The gray boxes surround amides that are known to deamidate. As can be seen from these two tables, the accuracy of the procedure is exceptionally high. Not only do the deamidated amides have the lowest C_D values in most cases, but the separation between those and the others is often large. We expect that, in many of the cases where the separation is not large and possibly even in the few cases where another amide was predicted to be faster than the one experimentally determined, further experiments will eventually show that these amides deamidate quickly as well.

There are a few notable exceptions. The worst is SerAsn(88)Ile of Interleukin 2 which was reported in our paper[7] as accelerated by an unknown mechanism, since it was reported to be faster than even its primary sequence rate. This amide was actually excluded from the optimization set, since it was so obviously anomalous. After publication, we found that there is indeed an accelerating effect. We neglected to consider that the deamidation was reported at pH 5,[131] which is far outside the pH range relevant to these calculations.

The other exceptions are the two hemoglobin mutants marked with open squares in Table 9-3. One of these deamidations results from a frame-shift mutation which disrupts the structure. The other comes from a mutation in the pocket holding the heme and results in loss of the heme. Again this must disrupt the structure considerably. Since we do not have crystal structures of these mutants and we can no longer assume that the structure is about the same as in normal hemoglobin, our prediction procedure cannot be applied. The values given are based on a normal hemoglobin structure and are therefore not applicable. These amides were not used in the optimization.

Table 9-2 Ordered Deamidation Coefficients and Experimentally Determined Deamidating Asn Residues in 23 Proteins.

Aldolase	
Ser-Asn360-His	**0.22**
Thr-Asn119-Gly	0.93
Gln-Asn180-Gly	4.89
Pro-Asn231-Met	19.5
Ala-Asn334-Ser	76.4
Leu-Asn284-Ala	129
Glu-Asn166-Ala	346
Glu-Asn50-Thr	724
Ile-Asn282-Leu	790
Glu-Asn319-Leu	979
Ile-Asn287-Lys	1394
Val-Asn70-Pro	1587
Ala-Asn168-Val	2105
Glu-Asn54-Arg	2830
	0.17
Angiogenin	
Lys-Asn61-Gly	**0.29**
Glu-Asn109-Gly	**0.38**
Asp-Asn3-Ser	1.7
Gly-Asn49-Lys	18.5
Glu-Asn59-Lys	20.3
Ile-Asn43-Thr	21.5
Gly-Asn63-Pro	55.1
Glu-Asn68-Leu	71.5
Arg-Asn102-Val	1610
	0.15
Calbindin	
Lys-Asn56-Gly	**0.03**
Pro-Asn21-Gln	8.01
	0.03
cAMP-Dependent Protein Kinase	
Gly-Asn2-Ala	**0.21**
Ile-Asn340-Glu	1.53
Gly-Asn67-His	4.28
Gly-Asn283-Leu	11.5
Val-Asn99-Phe	11.9
Gln-Asn36-Thr	12.7
Tyr-Asn216-Lys	23.3
Ser-Asn326-Phe	39.2
Val-Asn289-Asp	40.7
Lys-Asn293-His	42.4
Asp-Asn113-Ser	53.9
Glu-Asn32-Pro	89.1
Leu-Asn90-Glu	180
Arg-Asn271-Leu	251
Ser-Asn115-Leu	275
Glu-Asn171-Leu	413
	0.16
Cytochrome c	
Thr-Asn103-Glu	**0.68** ‡‡
Pro-Asn31-Leu	12.3
Lys-Asn54-Lys	**51.1** ‡
Ala-Asn52-Lys	749
Glu-Asn70-Pro	1310
	0.64
Epidermal Growth Factor	
Leu-Asn16-Gly	0.08
Asn1-Ser	**0.26** ‡‡‡
Cys-Asn32-Cys	8.19
	0.06
Fatty Acid Binding Protein	
Phe-Asn105-Gly	**0.42**
Asp-Asn89-Lys	1.98
His-Asn61-Glu	4.73
Met-Asn2-Phe	30.2
Thr-Asn111-Thr	159
Glu-Asn14-Phe	202
	0.32
Fibroblast Growth Factor	
Ser-Asn18-Gly	0.21
Gly-Asn7-Tyr	**0.64**
Lys-Asn114-Gly	1.38
Phe-Asn2-Leu	3.08
Glu-Asn92-His	6.3
Pro-Asn80-Glu	9.21
Lys-Asn106-Trp	51.5
Tyr-Asn95-Thr	74.8
	0.13
Glucoamylase	
Val-Asn181-Gly	**0.35**
Arg-Asn69-Gly	0.96
Tyr-Asn313-Gly	1.18
Asp-Asn145-Gly	2.48
Ser-Asn395-Gly	13.1
Trp-Asn171-Gln	13.1
Ala-Asn236-Phe	18.9
Phe-Asn110-Val	22.4
Arg-Asn430-Ser	35.7
Leu-Asn292-Asp	42.9
Gly-Asn315-Pro	55.1
Asp-Asn45-Pro	68.6
Ala-Asn277-His	77.1
Thr-Asn247-Thr	98.2
Leu-Asn20-Asn	99.9
Ala-Asn426-Asn	127
Ser-Asn9-Glu	189
Asn-Asn427-Arg	190
Ser-Asn93-Pro	192
Asn-Asn21-Ile	241
Arg-Asn161-Asp	465
	0.18
Growth Hormone	
His-Asn152-Asp	**0.81**
Thr-Asn149-Ser	**1.17**
Ala-Asn99-Ser	1.64
Ser-Asn63-Arg	4.07
Gln-Asn12-Ala	128
Ser-Asn72-Leu	170
Lys-Asn159-Tyr	496
	0.34
Hypoxanthine Guanine Phosphoribosyl-transferase	
Cys-Asn106-Asp	**3.06**
Leu-Asn202-His	7.31
Pro-Asn25-His	8.33
Lys-Asn128-Val	16.8
Arg-Asn87-Ser	45.4
Tyr-Asn195-Glu	80.5
Tyr-Asn153-Pro	89.1
Leu-Asn85-Arg	1344
	1.45
Insulin	
B-Val-Asn3-Gln	**1.17**
A-Cys-Asn21	6.06 ‡‡‡
A-Glu-Asn18-Tyr	115
	0.97
Interleukin 1β	
Leu-Asn32-Gly	**0.04**
Asn-Asn137-Ser	0.93
Gly-Asn136-Asn	1.69
Ile-Asn37-Gln	13
Ser-Asn53-Asp	22.8
Lys-Asn66-Leu	55.4
Pro-Asn119-Trp	433
Gln-Asn35-Ile	550
Phe-Asn102-Lys	780
	0.04
Interleukin 2	
Asn-Asn30-Tyr	9.01
Lys-Asn77-Phe	45.6
Leu-Asn26-Gly	125
Ile-Asn29-Asn	127
Leu-Asn119-Arg	197
Leu-Asn71-Leu	239
Lys-Asn33-Pro	702
Ile-Asn90-Val	936
Ser-Asn88-Ile	**2362** ‡
	6.21
Lysozyme	
Gly-Asn103-Gly	**0.06**
Met-Asn106-Ala	**0.58**
Ile-Asn59-Ser	4.6
Arg-Asn113-Arg	7.43
Arg-Asn46-Thr	11.5
Asp-Asn19-Tyr	15.2
Arg-Asn74-Leu	22.8
Thr-Asn44-Arg	23.8
Phe-Asn39-Thr	47.4
Cys-Asn65-Asp	60.1
Val-Asn93-Cys	201
Gly-Asn27-Trp	245
Cys-Asn77-Ile	277
Ser-Asn37-Phe	807
	0.05
Ribonuclease-A	
Lys-Asn67-Gly	**0.85**
Ser-Asn24-Tyr	11.5
Val-Asn44-Thr	14.4
Ala-Asn103-Lys	16.1
Thr-Asn71-Cys	19.5
Pro-Asn94-Cys	29.8
Lys-Asn62-Val	70.8
Gly-Asn113-Pro	131
Arg-Asn34-Leu	141
Cys-Asn27-Gln	308
	0.66
Ribonuclease Seminal	
Lys-Asn67-Gly	**0.31**
Gly-Asn17-Ser	2.1
Thr-Asn71-Cys	4.37
Val-Asn44-Thr	60.6
Pro-Asn94-Cys	77.7
Cys-Asn27-Leu	145
Ser-Asn24-Tyr	787
	0.25
Ribonuclease-U2	
Tyr-Asn68-Gly	0.14
Ala-Asn32-Gly	**1.62**
Asp-Asn77-Tyr	12.5
Ser-Asn16-Asp	36.8
Thr-Asn8-Cys	85.7
Thr-Asn91-Thr	131
Gly-Asn12-Val	132
Asp-Asn38-Tyr	145
Ile-Asn20-Thr	147
	0.13
T-Cell Surface Glycoprotein CD4	
Leu-Asn52-Asp	**3.99**
Lys-Asn30-Ser	5.71
Ala-Asn103-Ser	12.3
Lys-Asn137-Ile	21.3
Gly-Asn66-Phe	29.8
Gln-Asn164-Gln	66.2
Ser-Asn32-Gln	69.5
Lys-Asn73-Leu	89.5
Gly-Asn39-Gln	92.7
	1.56
Thioltransferase	
Thr-Asn51-His	**4.92**
Thr-Asn55-Glu	119
Val-Asn7-Cys	1076
	4.70
Triose Phosphate Isomerase	
Thr-Asn71-Gly	**0.78**
Met-Asn15-Gly	**1.77**
Gln-Asn65-Cys	4.55
Ile-Asn245-Ala	31.5
Gly-Asn11-Trp	35.9
Asp-Asn153-Val	165
Leu-Asn29-Ala	208
Ser-Asn195-Val	360
	0.47
Trypsin	
Leu-Asn115-Ser	**1.14**
Tyr-Asn95-Ser	**1.28**
Ile-Asn48-Ser	**4.82**
Leu-Asn34-Ser	6.11
Ser-Asn97-Thr	10.6
Lys-Asn223-Lys	12.1
Ser-Asn245	21.1 ‡‡‡
Ala-Asn25-Thr	32.7
Leu-Asn100-Asn	40.9
Gly-Asn143-Thr	44.4
Gly-Asn79-Glu	50.5
Asp-Asn72-Ile	75.5
Ile-Asn74-Val	99.7
Ser-Asn179-Met	108
Cys-Asn233-Tyr	355
Asn-Asn101-Asp	537
	0.42
Phosphocarrier Protein - Hpr	
Pro-Asn12-Gly	**0.21**
Ser-Asn38-Gly	**0.57**
	0.15

Squares designate Asn reported as deamidated. ___ designates Deamidation Index, I_D.
‡ These two un-shaded squares designate unusual protein structures that accelerate deamidation.
‡‡Uses primary $t_{1/2}$ from Reference 11.
‡‡‡Uses primary $t_{1/2}$ from Reference 73.

Table 9-2 Ordered Deamidation Coefficients for 70 Asn Residues in Wild-Type and Mutant Human Hemoglobins and Experimentally Determined Deamidating Asn Residues.

Hemoglobin - 7.78			
α-Ser-Asn50-Gly	**0.18**	β-Val-Asn61-Ala	**141**
β-Leu-Asn82-Gly	**0.19**	α-Val-Asn11-Ala	**141**
α-Pro-Asn78-Gly	**0.67**	β-Gly-Asn108-Met	**160**
β-Lys-Asn145-His	**1.7**	β-Ala-Asn139-Asp	**164**
β-Asp-Asn80-His	**1.73**	β-Gly-Asn17-Val	**177**
β-Val-Asn19-Gly	**2.53**	β-Ala-Asn139-Thr	**223**
β-Ser-Asn73-Gly	**4.92**	β-Asp-Asn80-Arg	**240**
β-Ala-Asn63-Gly	**5.39**	β-Gly-Asn65-Lys	**247**
α-Val-Asn56-Gly	**6.31**	α-Gly-Asn60-Lys	**247**
α-Pro-Asn78-Ala	**11.1**	α-Asp-Asn7-Thr	**269**
α-Leu-Asn87-Ala	**11.7**	β-Pro-Asn59-Val	**274**
α-Asp-Asn75-Met	**12.4**	β-Leu-Asn92-Cys	**(274)**
α-Phe-Asn47-Leu	**13**	β-Leu-Asn89-Glu	**291**
β-Gly-Asn120-Glu	**14.5**	β-Asp-Asn80-Leu	**305**
β-Val-Asn21-Glu	**18.7**	β-Gly-Asn108-Leu	**330**
α-Val-Asn74-Asp	**20.8**	β-His-Asn117-Phe	**370**
α-Pro-Asn78-Thr	**22**	α-Val-Asn133-Thr	**414**
β-Asp-Asn80-His	**26.7**	α-Leu-Asn126-Lys	**498**
β-Ala-Asn143-Lys	**33.1**	α-Ser-Asn85-Leu	**564**
β-Leu-Asn79-Asn	**33.1**	α-Asp-Asn127-Phe	**581**
β-Cys-Asn94-Lys	**35**	β-Glu-Asn102-Phe	**582**
β-Ser-Asn52-Ala	**47.3**	β-Val-Asn19-Val	**600**
α-His-Asn90-Leu	**49**	β-Gly-Asn108-Val	**711**
β-Gly-Asn47-Leu	**58.6**	α-Ala-Asn6-Lys	**749**
β-Pro-Asn52-Ala	**62.5**	α-Val-Asn94-Pro	**892**
β-Trp-Asn38-Gln	**72.9**	β-Glu-Asn102-Leu	**1077**
α-Thr-Asn68-Ala	**78.1**	β-Val-Asn99-Pro	**1081**
α-Ser-Asn139-Thr	**(81.5)** ‡	α-Thr-Asn9-Val	**1215**
β-His-Asn144-Tyr	**100**	α-Lys-Asn61-Val	**1262**
β-Asp-Asn95-Leu	**106**	β-Ala-Asn139-Val	**1303**
α-Ala-Asn64-Ala	**115**	β-Gly-Asn57-Arg	**1313**
β-Ala-Asn139-Ala	**115**	β-Gln-Asn132-Val	**2155**
α-Val-Asn97-Phe	**131**	β-Glu-Asn102-Ile	**2312**
β-Val-Asn19-Glu	**134**	α-Gly-Asn16-Val	**2360**
β-Val-Asn19-Met	**135**	β-Gly-Asn57-Pro	**2690**

*Frame shift mutation and **heme loss mutation, so three-dimensional structures are unknown and C_D derived from wild-type hemoglobin is not applicable. Squares designate Asn reported as deamidated. ___ designates wildtype deamidation index, I_D.

Note that the observed deamidations in Tables 9-2 and 9-3 were utilized to optimize the coefficients used to calculate the D_P values. The small amount of data available in the literature prohibited us from calibrating and testing with entirely independent sets of proteins. These coefficients, however, depend upon the entire set of proteins and are not much different if each of the individual test proteins is separately excluded. Several proteins were later found that we had overlooked in our initial literature survey. Estimates of the deamidation coefficients of these proteins, with the same adjustable constants determined for this original set, all gave perfect results in prediction of the fastest

deamidating amides. Moreover, the estimates of absolute rates for proteins using this set of constants showed excellent agreement with experiment[8] even though no absolute rates were used in the calibration. In addition, the self-consistency of the results attests to their value.

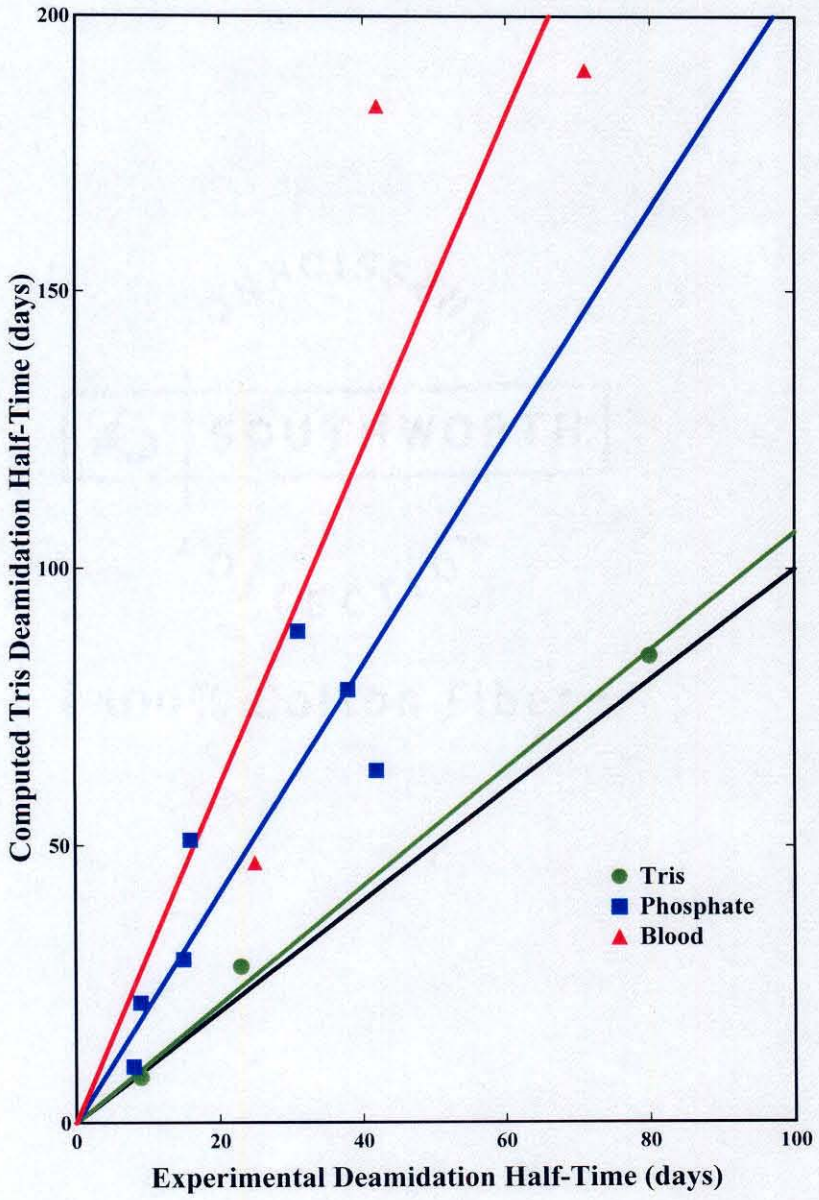


Figure 9-5 – Comparison of Predicted and Experimental Deamidation Half-Times

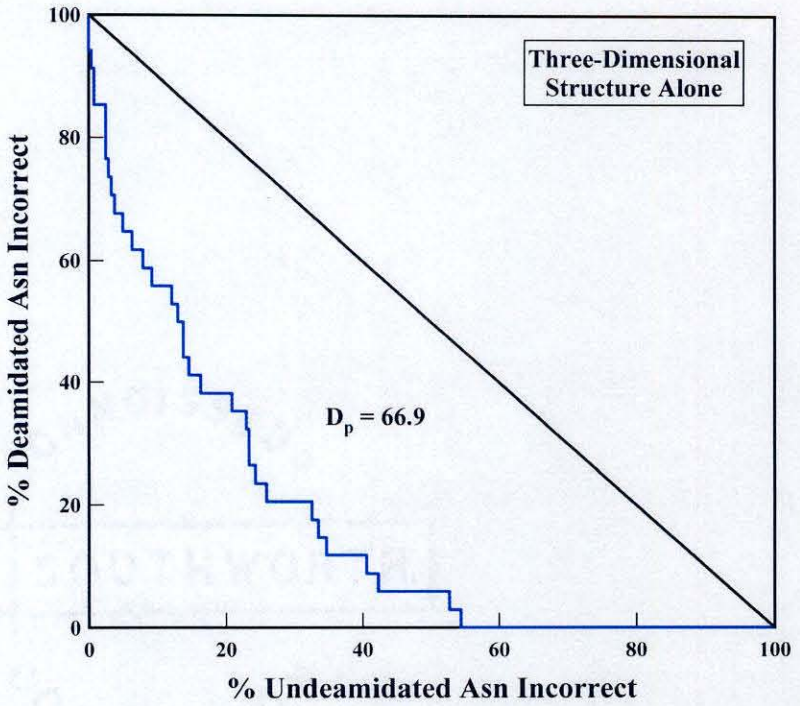
Figure 9-5 shows a comparison of actual deamidation rates to those we have predicted. The calculated rates are applicable to real rates, since they are based on suppression of experimentally determined primary sequence rates by amounts that have been carefully optimized.

Some variation is expected, since the predicted rates depend on the crystal or NMR structures. For example, the exact number of hydrogen bonds could be slightly different in solution, and there is undoubtedly variation in their strength, which has not been considered. A single hydrogen bond difference can change the rate by a factor of two or more. In terms of the overall results on proteins in general, these types of effects should average out.

A plot of predicted half-time versus real half-time for 13 amides in proteins is given in Figure 9-5. The black line is the theoretically perfect line. The predictions would be perfect if these data all lay along this line. The three values that were obtained in Tris buffer lie close to this line. The predicted values are for Tris buffer.

Phosphate is a much stronger catalyst of deamidation in peptides[72, 74, 155] and proteins[16] as compared to Tris. Tissue culture medium contains components even more catalytic of deamidation than phosphate.[156] Least-squares lines as shown in Figure 9-3 give experimental deamidation rates relative to the computed values in Tris, phosphate, and *in vivo* blood erythrocytes of 1, 2, and 3.

It is of interest to consider the fraction of the rate that is determined by primary sequence alone and the fraction determined by three-dimensional structure. The D_P calculation was, therefore, repeated in two different ways. Using the optimized parameters and setting the primary sequence half-time to 1, the separation ability of three-dimensional structure alone can be determined. The D_P calculation can also be done using only the primary sequence rates. These plots are shown in Figure 9-6. See Figure 9-3 for the two elements combined.



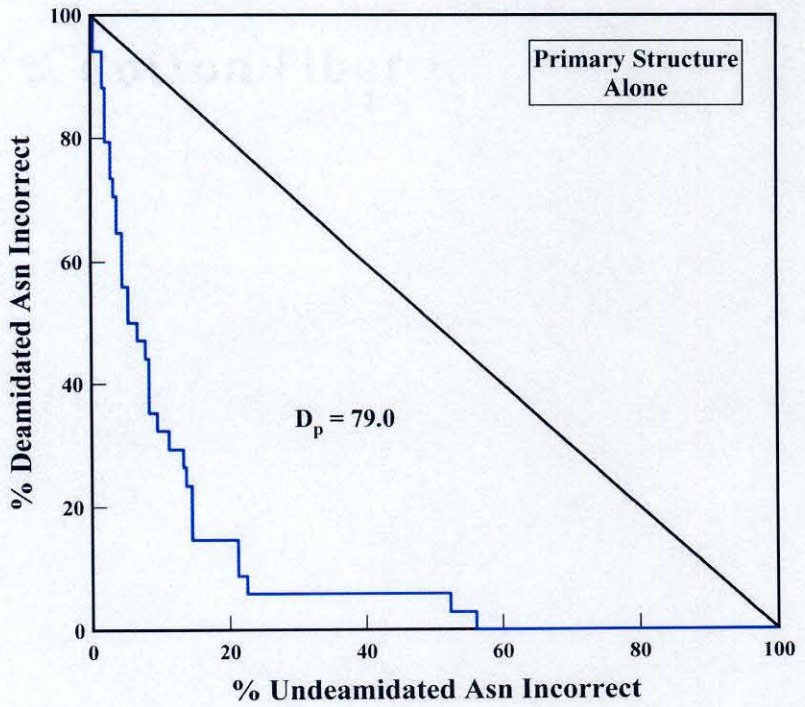


Figure 9-6 – D_P plots for 3-D and Primary Structure Alone

It is clear that both primary and secondary structure are necessary to achieve reliable deamidation rate predictions. One can estimate from these plots that roughly 60% of the deamidation rate is controlled by primary structure, while 40% is controlled by three-dimensional structure in this protein set. Both elements are essential for predicting absolute deamidation rates in proteins.

An illustration of this is also shown in Figure 9-7. Because AsnGly sequences are much faster than other primary sequence deamidation rates, many investigators assume that deamidation occurs at AsnGly sequences. This figure shows the errors of this assumption.

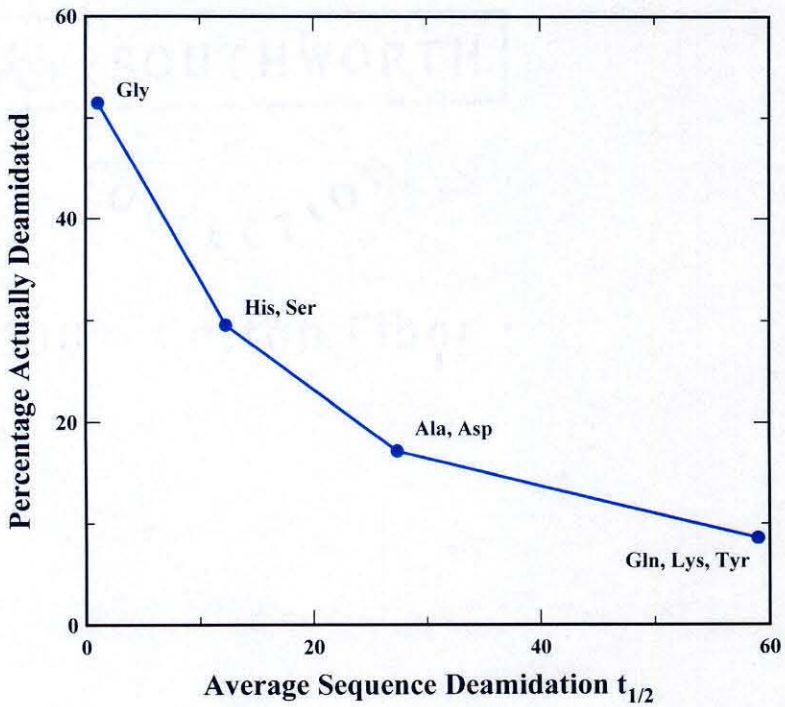


Figure 9-7 – Percentages of deamidating Asn residues listed in Tables 9-2 and 9-3 that would be correctly guessed by simply assuming that Asn residues with COOH-side Gly, His, Ser, Ala, Asp, Gln, Lys or Tyr deamidate versus average pentapeptide deamidation half times[6] for those specific Asn sequences.

In these proteins, only 51% of the AsnGly sequences are observed to deamidate, while 30% of the AsnHis and AsnSer sequences deamidate. For AsnAla and AsnAsp this value is 17%, and only drops to 6% for AsnGln, AsnLys, and AsnTyr.

Both primary and secondary structure are needed to reliably predict deamidation rates in proteins. The procedure described here performs these predictions with remarkable accuracy. This same procedure can be applied to Gln deamidation and probably Asp and Glu isomerization as well as chain cleavage of Asn, Asp, Gln and Glu, once sufficient calibration data are available. The application to Asp isomerization would be especially interesting. Application of a simple constant may be sufficient for many purposes. For example, in the case of chain cleavage of AsnGly, a constant derived from reference **164** could be used.

The next section describes the application of this manual procedure to 126 human proteins. The following chapter is devoted to the automation of this procedure through the use of computer algorithms and extension to the entire PDB protein structure databank.

Application of Hand Prediction Procedure to 126 Human Proteins

Having devised a method for predicting deamidation of any protein for which the three-dimensional structure is known, we next applied this method to a large number of proteins in order to learn how extensive deamidation is in ordinary proteins. The ones we had studied so far were selected because they were known to deamidate. If deamidation is used extensively as a biological timer, a large fraction of proteins should be unstable toward deamidation during their biological lifetimes. Since deamidation accelerates protein turnover, it is not detected to its full extent in protein preparations from *in vivo* sources.

The method was therefore applied to 126 proteins with known three-dimensional structures. The resulting protein I_Ds are tabulated in Table 9-4. The C_D values for each amide in the protein were predicted and combined into the I_D for the protein.

Table 9-4 Computed deamidation half-times for 126 human proteins in 37° C, pH 7.4, 0.15 M Tris•HCl buffer. The proteins were selected without regard to reported deamidation except for 13 proteins that were among those used to develop and test the calculation method and are designated asterisk in the table.

Protein	1/2 Life (Days)	1/10 Life (Days)	Protein	1/2 Life (Days)	1/10 Life (Days)
Uracil-DNA Glycosylase (1LAU)	1.0	0.15	Proinsulin (1EFE)	110	17
Uroporphyrinogen Decarboxylase (1URO)	1.0	0.15	Mitogen-Activated Protein Kinase P38 (1WFC)	110	17
Transaldolase (1F05)	1.4	0.21	Glutathione Reductase (1BWC)	120	18
Urokinase-Type Plasminogen Activator (1LMW)	1.7	0.26	Ribonuclease 4 (1RNF)	130	20
Purine Nucleoside Phosphorylase (1ULA)	1.8	0.27	Aldose Reductase (1EL3)	130	20
Growth Hormone Receptor (1A22)	2.4	0.36	α-Lactalbumin (1B9O)	130	20
Peptidyl-Prolyl Cis-Trans Isomerase (1F8A)	2.4	0.36	Ornithine Transcarbamoylase (1OTH)	130	20
Thymidylate Synthase (1HW3)	2.7	0.41	Malic Enzyme (1EFK)	140	21
Procathepsin B (3PBH)	2.9	0.44	Glucose-6-Phosphate 1-Dehydrogenase (1QKI)	140	21
D-Glyceraldehyde-3-Phosphate Dehydrogenase (3GPD)	4.2	0.64	Procarboxypeptidase A2 (1AYE)	150	23
Karyopherin β2 (1QBK)	5.3	0.81	Apoptosis Regulator Bax (1F16)	170	26
Glutathione S-Transferase (12GS)	5.3	0.81	Ornithine Decarboxylase (1D7K)	170	26
N-Acetylgalactosamine-4-Sulfatase (1FSU)	6.1	0.93	UDP-Galactose 4-Epimerase (1EK6)	180	27
Fructose Bisphosphate Aldolase (4ALD)	7.6	1.2	Stem Cell Factor (1EXZ)*	180	27
Intestinal Fatty Acid Binding Protein (3IFB)	7.6	1.2	Hypoxanthine Guanine Phosphoribosyltransferase (1BZY)*	180	27
Cyclophilin A (1AWQ)	8.7	1.3	Electron Transfer Flavoprotein (1EFV)	190	29
Vascular Endothelial Growth Factor (2VPF)*	10	1.5	Phenylalanine Hydroxylase (1DMW)	220	33
Inositol Monophosphatase (1IMB)	15	2.3	Annexin V (1ANX)	220	33
Pancreatic Inhibitor Variant 3 (1CGI)	16	2.4	Platelet Factor 4 - HPF4 (1RHP)	230	35
D-Glucose 6-Phosphotransferase (1HKC)	16	2.4	Insulin (2HIU)*	260	40
Myeloperoxidase (1MHL)	16	2.4	Prethrombin2 (1HAG)	260	40
α-Chymotrypsinogen (1CGI)	16	2.4	Interleukin-4 (2CYK)	270	41
Lysophospholipase (1LCL)	16	2.4	Interleukin-1β (2I1B)	280	43
Interleukin-16 (1I16)	17	2.6	Neutrophil (Gelatinase) (1DFV)	290	44
C-AMP Dependent Kinase A (1CMK)	19	2.9	O6-Alkylguanine-DNA Alkyltransferase (1EH6)	300	46
Pepsinogen (1HTR)	20	3.0	Glucosamine-6-Phosphate Deaminase Isomerase (1D9T)	320	49
Angiogenin (1A4Y)*	21	3.2	Quinone Reductase Type 2 (1QR2)	330	50
Fibroblast Growth Factor (2AFG)*	21	3.2	Nad(P)H Dehydrogenase (1QBG)	350	53
Gastric Lipase (1HLG)	21	3.2	Serum Albumin (1E7G)	360	55
Calmodulin (1CTR)	21	3.2	Plasminogen Activator Inhibitor-1 (1C5G)	370	56
Bone Morphogenetic Protein 7 (1BMP)	21	3.2	T-Cell Surface Glycoprotein CD4 (1CDJ)*	380	58
Acetylcholinesterase (1F8U)	23	3.5	α-Thrombin (1A3E)	380	58
Retinol Binding Protein (1BRQ)*	24	3.6	Eosinophil Cationic Protein (1QMT)	430	65
Catalase (1QQW)	25	3.8	Ribonuclease Inhibitor (1A4Y)	450	68
Dihydrofolate Reductase (1DRF)	25	3.8	Transforming Growth Factor-β Two (1KLA)	460	70
Interleukin-10 (2ILK)	25	3.8	Thioltransferase (1JHB)*	470	71
Farnesyltransferase (1EZF)	26	4.0	Profilin I (1FIL)	480	73
S-Adenosylhomocysteine Hydrolase (1A7A)	28	4.3	Lithostathine (1LIT)	490	74
Procathepsin K (1BY8)	28	4.3	Phosphatidylethanolamine Binding Protein (1BD9)	680	100
3-Methyladenine DNA Glycosylase (1BNK)	35	5.3	Dihydroorotate Dehydrogenase (1D3G)	720	110
Medium Chain Acyl-Coa Dehydrogenase (1EGE)	36	5.5	Quinone Reductase (1D4A)	750	110
Homeobox Protein PAX-6 (6PAX)	39	5.9	Hemoglobin (1A3N)*	780	120
α1-Antitrypsin (1QLP)	40	6.1	Retinoic Acid Receptor (1BY4)	860	130
Carbonic Anhydrase I (1HCB)	45	6.8	Psoriasin (1PSR)	890	140
GTP-Binding Protein (1DOA)	45	6.8	ADP-Ribosylation Factor 6 (1E0S)	>1000	150
Ferritin (2FHA)	46	7.0	Lectin L-14-II (1HLC)	>1000	180
Procathepsin L (1CS8)	48	7.3	Nucleoside Diphosphate Kinase (1NUE)	>1000	210
Growth Hormone (1HGU)*	51	7.8	L-3-Hydroxyacyl-CoA Dehydrogenase (1F0Y)	>1000	230
Triose Phosphate Isomerase (1HTI)*	52	7.9	Interleukin 2 (3INK)*	>1000	230
Interleukin-6 (1IL6)	56	8.5	Transthyretin (1DVQ)	>1000	290
DNA Polymerase β (1BPX)	58	8.8	Single-Stranded DNA Binding Protein (3ULL)	>1000	290
Glutathione Synthetase (2HGS)	58	8.8	Protein Kinase C Interacting Protein 1 (1KPA)	>1000	300
Fructose-1,6-Bisphosphatase (1FTA)	59	9.0	GTPase Ran (1QBK)	>1000	380
CDK2 Kinase (1BUH)	65	9.9	Annexin III (1AXN)	>1000	430
Ribonuclease A (1AFK)	66	10	Fk506-Binding Protein (1D6O)	>1000	710
Ap Endonuclease (1BIX)	72	11	Interleukin-5 (1HUL)	>1000	760
Carbonic Anhydrase IV (1ZNC)	72	11	Heme Oxygenase (1QQ8)	>1000	780
Branched-Chain α-Keto Acid Dehydrogenase (1DTW)	81	12	Histone H2A.Z (1F66)	>1000	>1000
Argininosuccinate Lyase (1AOS)	83	13	Copper Transport Protein ATOX1 (1FEE)	>1000	>1000
Creatine Kinase (1QK1)	84	13	17 β-Hydroxysteroid Dehydrogenase (1DHT)	>1000	>1000
Carbonic Anhydrase II (1BV3)	90	14	Myoglobin (2MM1)	>1000	>1000
Interleukin-8 (1IL8)	95	14	Ubiquitin (1D3Z)	>1000	>1000
Dihydropteridine Reductase (1HDR)	100	15	Granulocyte Colony-Stimulating Factor (1RHG)		

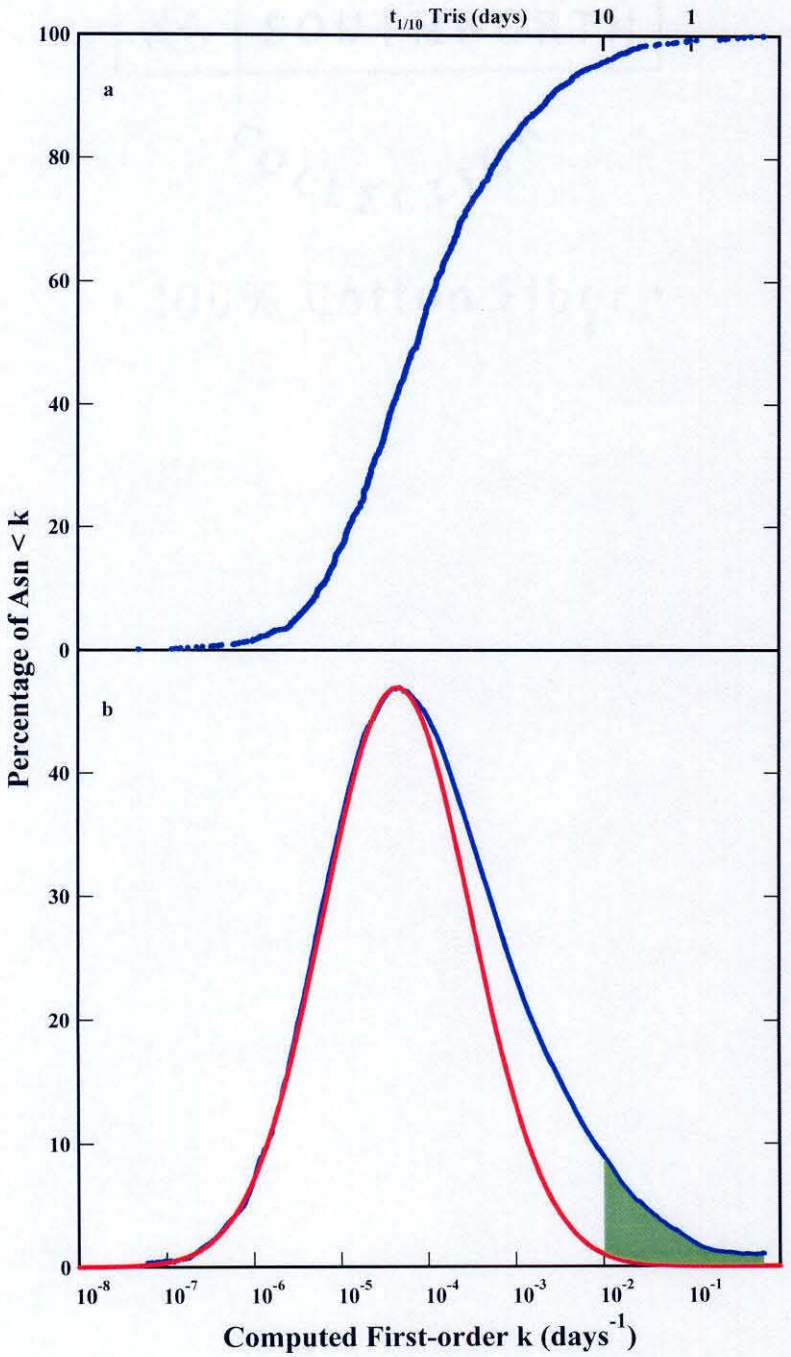


Figure 9-8 – Distribution of C_D Values for 126 Human Proteins

As can be seen from Table 9-4, a large fraction of these proteins have half-times which are certainly fast enough to be biologically significant, while the deamidation in another group of them is suppressed to the point where it is no longer relevant. This distribution is illustrated in Figure 9-8. The upper curve is a cumulative distribution function for the amides tabulated in Table 9-4. The lower curve is the differentiated distribution. It is also clear from this figure that there is no general suppression of fast amides in human proteins.

Table 9-5 – Percentages of Human Proteins in Table 9-4 Computed to be More Than One-Tenth or One-Half Singly Deamidated in Tris or Phosphate Buffer After 1, 5, 10, and 50 Days*.

Days at 37 °C pH 7.4	Proteins Singly Deamidated by >1/10		Proteins Singly Deamidated by >1/2	
	Tris	Phosphate	Tris	Phosphate
1	10%	13%	1.6%	4%
5	31%	43%	8%	13%
10	43%	56%	13%	20%
50	71%	82%	37%	49%

*A phosphate buffer correction of (2.0)(computed Tris rate) was applied to obtain the phosphate rate of each protein. Percentages produced in physiological solutions may be even higher. Steady state in vivo physiological percentages are lowered by protein turnover.

Averaged over all 1,371 Asn, the contributions to the deamidation reaction activation energy from primary and 3-D structures are about equal, although they vary widely for individual Asn residues. The average relative deamidation rates of Asn within single proteins in this 126-protein set are 60% determined by primary and 40% by 3-D structure, which are the same proportions found for a different 24-protein set.[7] The cumulative distribution function of the calculated first-order rate constants for deamidation of the 1,371 Asn residues is shown in Figure 9-8*a*.

The computed single deamidation half-times in 37 °C, pH 7.4, 0.15 M Tris buffer for the 126 human proteins are shown in Table 9-4. Table 9-5 summarizes, on the basis of

Table 9-4, the extent to which deamidation is expected to occur within this set of 126 proteins.

The percentages of deamidation in living tissues are probably higher than shown in Tables 9-4 and 9-5. Physiological fluids contain many inorganic, organic, and biochemical substances with deamidation catalyzing activity. We know of no reported instance, *in vivo* or *in vitro*, of an experimentally measured protein deamidation rate that is slower than its computed Tris rate. All reported rates are the same or faster. There are two instances of individual proteins[6, 7] in which negative results for the detection of deamidation in specific amides indicates that the rates, if measured, might be slower than calculated.

This deamidation is not a random consequence of the presence of Asn residues in proteins. The fast deamidations summarized in Table 9-5 result from a set of Asn residues with unusual primary and 3-D structures, which comprise about 5% of the total. As illustrated in Figure 9-8, most individual Asn deamidation rates are slower. Because a large number of similarly sized partially independent factors determine Asn deamidation rates in proteins, the distribution functions in Figure 9-8 would be expected to be Gaussian. Figure 9-8*b* shows the deviation from Gaussian caused by the unusual Asn residues.

Chapter 10

COMPUTERIZED PREDICTION OF DEAMIDATION RATES

General Design of Program

The calculation procedure was developed entirely by manual observations using a structure viewer program. The accuracy of this procedure is more than 95% and is probably limited more by inaccuracies in the protein calibration set than by deficiencies in the procedure. For this reason, it was not expected that the procedure could be substantially improved by computerizing it. Although some improvements were made, the main focus in the programming was to automate the current procedure rather than attempt to improve upon it.

The first step was to write algorithms that duplicate the work we already were doing manually. In many cases this is straightforward, in others much more difficult. It was necessary to be able to read PDB files and make all calculations that the structure viewer makes for us. All optimization routines (which were done before using sophisticated Microsoft Excel sheets) were built into the program.

Throughout the hand procedure, detection of alpha helices and beta sheets was decided by the Swiss PDB viewer. While not difficult to do, we have so far left this part of the task to the Swiss PDB viewer. This keeps detection of these parameters standardized. A macro is simply used to open every file in the Swiss PDB viewer and individual files are then saved with just the helices, beta sheets, and loop regions. These files are compared by the program with the original file to detect the secondary structures. We plan eventually to remove this dependency from the program by writing our own routine. Other than this, all of the calculations are performed using routines that I wrote specifically for this program. In its current state, the program contains about 70 source files and more than 15,000 lines of C++ programming. It can calculate deamidation rates for the more than 200,000 amides

in the entire approximately 20,000 proteins in the Protein PDB Databank in about 9 hours using a 1.3 GHz Pentium 4 computer running Microsoft Windows 2000.

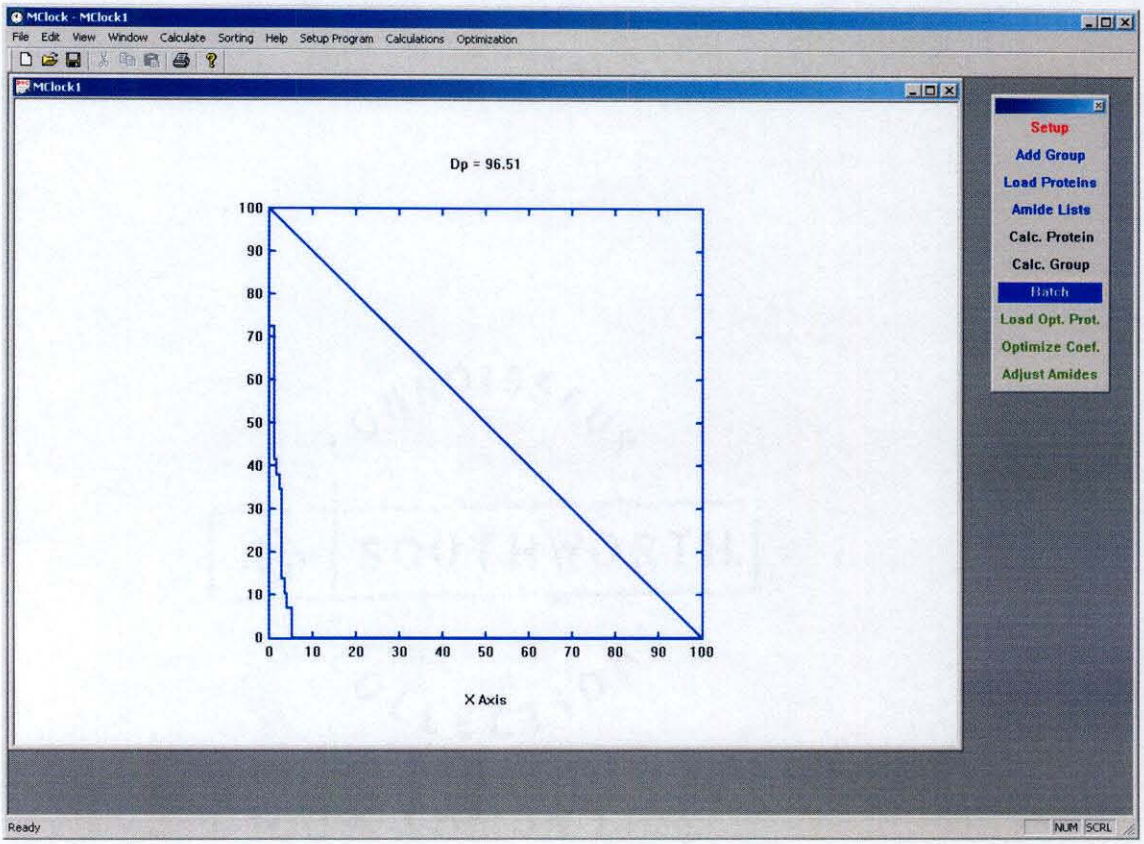


Figure 10-1 – Main Program Interface

Figure 10-1 shows the main program interface. A summary of the design of this program is given in the following pages. It is somewhat more sophisticated than the hand procedure, but not enough to make a great difference in results. The description of the hand procedure from chapter 8 is adequate to understand its function. The main advantage of the computerized procedure is ease in calculation of large numbers of amide rates.

Discussion of Routines and Specialized Problems

The first programming step was to load the PDB files into computer memory and put them into a form which can easily be utilized for calculations. Most of this is routine programming. In this step, all hydrogen bonds are found (later optimization of hydrogen bond length requires recalculation). Alpha-helix and beta-sheets are detected using Swiss PDB output as the detection procedure. Of particular concern was the assignment of primary structure.

Occasionally, when measuring parameters by hand, an amide is found that is near or in a break in an incomplete crystal structure. It is generally best in these cases to assume that the uncrystallized piece is completely free and unrestrained, by assigning no secondary structure or hydrogen bonds. While not always a good approximation, such regions are often difficult to crystallize for the reason that they are free. This is especially true of crystal structures missing short loops. Programming this presents a special problem, particularly since the complete sequence information is contained in a separate section of the PDB file and is poorly correlated to the three-dimensional data section. It was necessary to write an alignment routine in order to make this work. The routine occasionally found differences between the two datasets due to data errors, which had to be reconciled.

Another problem was hydrogen bond detection. The protein must be searched for pairs of atoms that could have hydrogen bonds between them. This must be done without any hydrogen coordinates, since these are not in most structure files. It is then necessary to determine whether the angle is reasonable for a hydrogen bond and also whether or not the maximum number of hydrogen bonds per atom is exceeded. In the manual procedure, it is easy to see when too many hydrogen bonds have been detected (Swiss PDB viewer does not filter for anything except length and angle) and ignore the poorer ones. A reliable routine for doing this proved more difficult, since the number of hydrogen bonds allowed depends on which ones are removed. This can become quite complicated when done according to a strict routine. This problem was, however, solved with a considerable amount of computation per bond.

Once the hydrogen bonds have all been determined, the number of hydrogen bonds in each category is simply counted by the computer. This is easy for everything except the "additional hydrogen bonds parameter" (S_9). The manual procedure for doing this involves looking at the overall structure in the area and making a semi-quantitative guess about which neighboring hydrogen bonds would need to broken in order to form the succinimide. The guesses were based on overall geometry and how far away the reacting groups were from each other. It was easy to imagine the groups coming together and visualize the motion required while taking the nearby structure into account. Sometimes, one part of the chain was assumed to move, sometimes another, depending on that which seemed better. Computers require absolute rules, so a different procedure was substituted.

The only *initial* change in backbone structure required for the formation of the succinimide is rotation around the C-C bond of the amide. Since the rest of the structure is approximately planar, this is one of only two bonds that can normally rotate for each residue. Figure 10-2 illustrates this rotation.

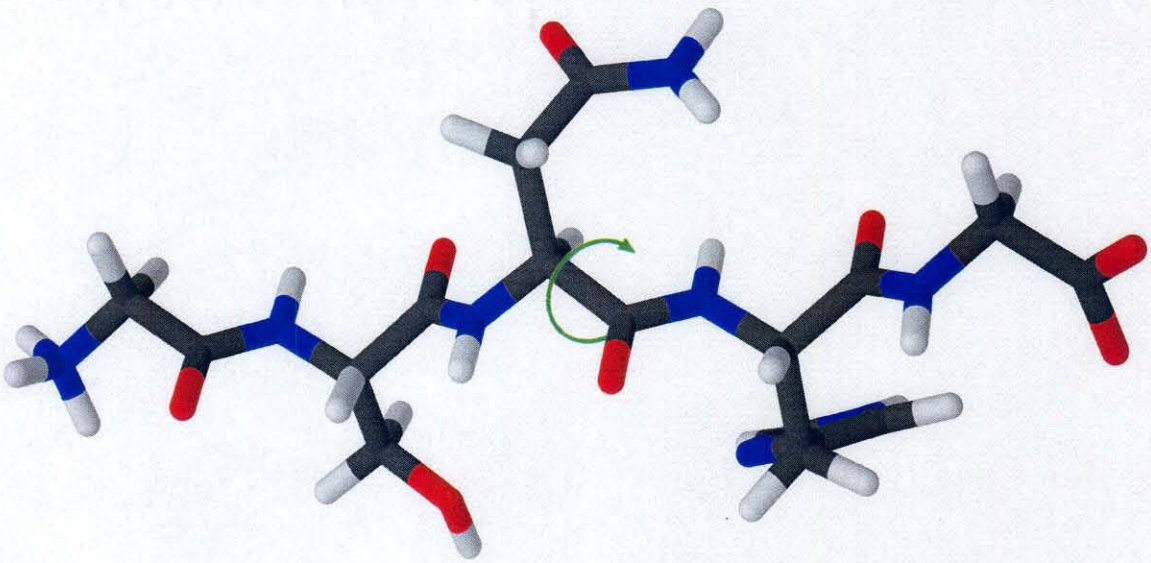


Figure 10-2 – Rotation of Bond Required to Form Succinimide Intermediate.

Once the groups are lined up, the amide side chain has only to swing around into position for the succinimide ring to form. Unless the amide is in a short peptide or a very free section of the protein, the C-C rotation cannot be done without disrupting the neighboring structure. This is the principle used by the computer to determine the number of hydrogen bonds.

It was an easy matter to calculate the angle around the C-C bond, where 0° is perfectly lined up and 180° is completely opposite. The program then divides the angle by an optimized constant and finds all hydrogen bonds within this radius of the amide. Any hydrogen bonds already counted in the other hydrogen bond parameters are ignored. It is also required that the selected hydrogen bonds be connected, at least on one end, to residues that are no more than 5 residues from the amide on either side, counting along the chain.

Improvements Over Hand Procedure

The computerized procedure allows for easier optimization of parameters. It also allows parameters to be optimized that could not be optimized manually. The hydrogen bond minimum and maximum lengths and angles were optimized as well. Attempts were also made to add parameters for direct use of the rotation angle, overall electron density and other things that would be expected to correlate with the deamidation rate. It was found that many of these things did have some correlation, but they were for the most part redundant and did not significantly improve the original accuracy. Ultimately, it was determined that the original set of parameters were the best choice. Overall, there was about a 2% increase in accuracy (measured by the D_P method) - from the original[7] 94.3% to about 95.4% in the updated hand procedure[8] and finally 96.5% with the computerized procedure.[9]

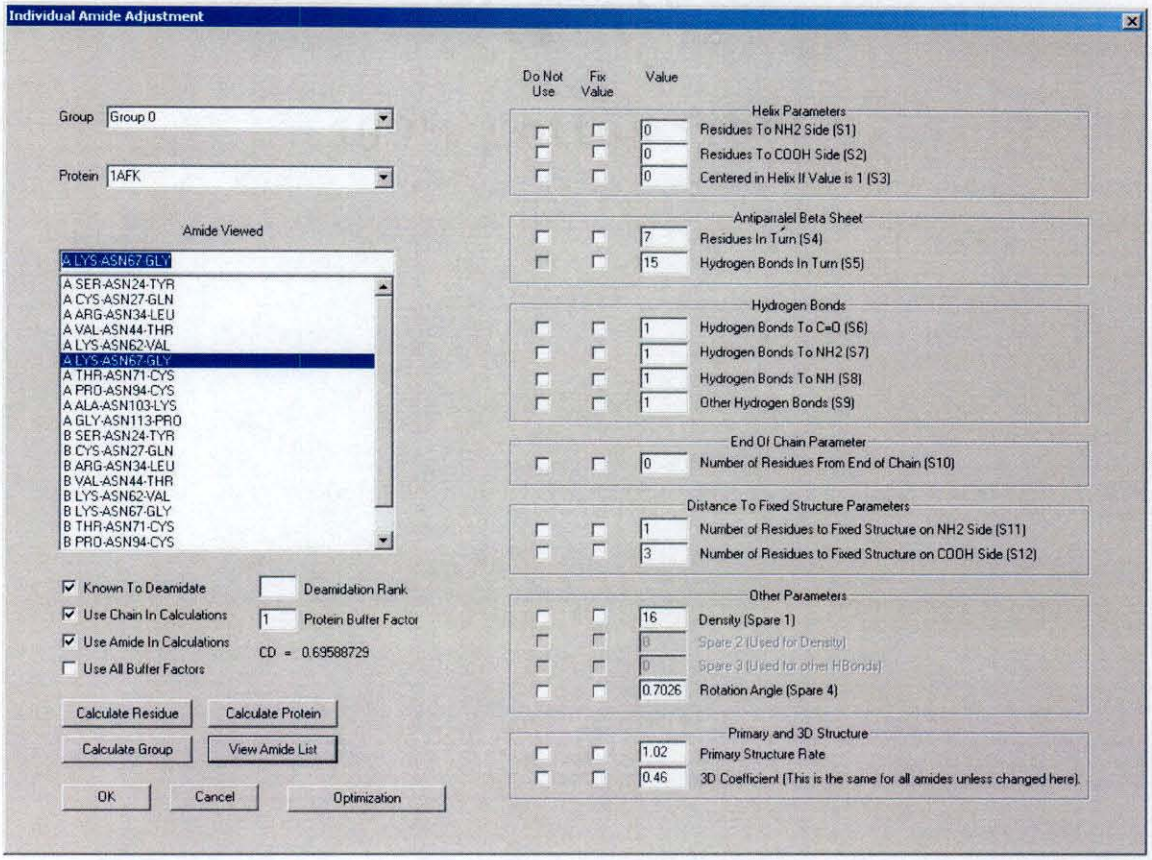


Figure 10-3 – Individual Amide Modification Window

It is of interest to know which parameters are involved in determining the deamidation rate of a particular amide. For convenience in determining the magnitude of these effects, a portion of the program made it easy to change any parameter (for example a hydrogen bond that slows the rate below that expected) in any particular residue and see what effect this has on the C_D value. Figure 10-3 shows this dialog. None of our published optimization or reported C_D values were ever changed in this way, even in cases were there is clearly an extra hydrogen bond that may not be present in solution.

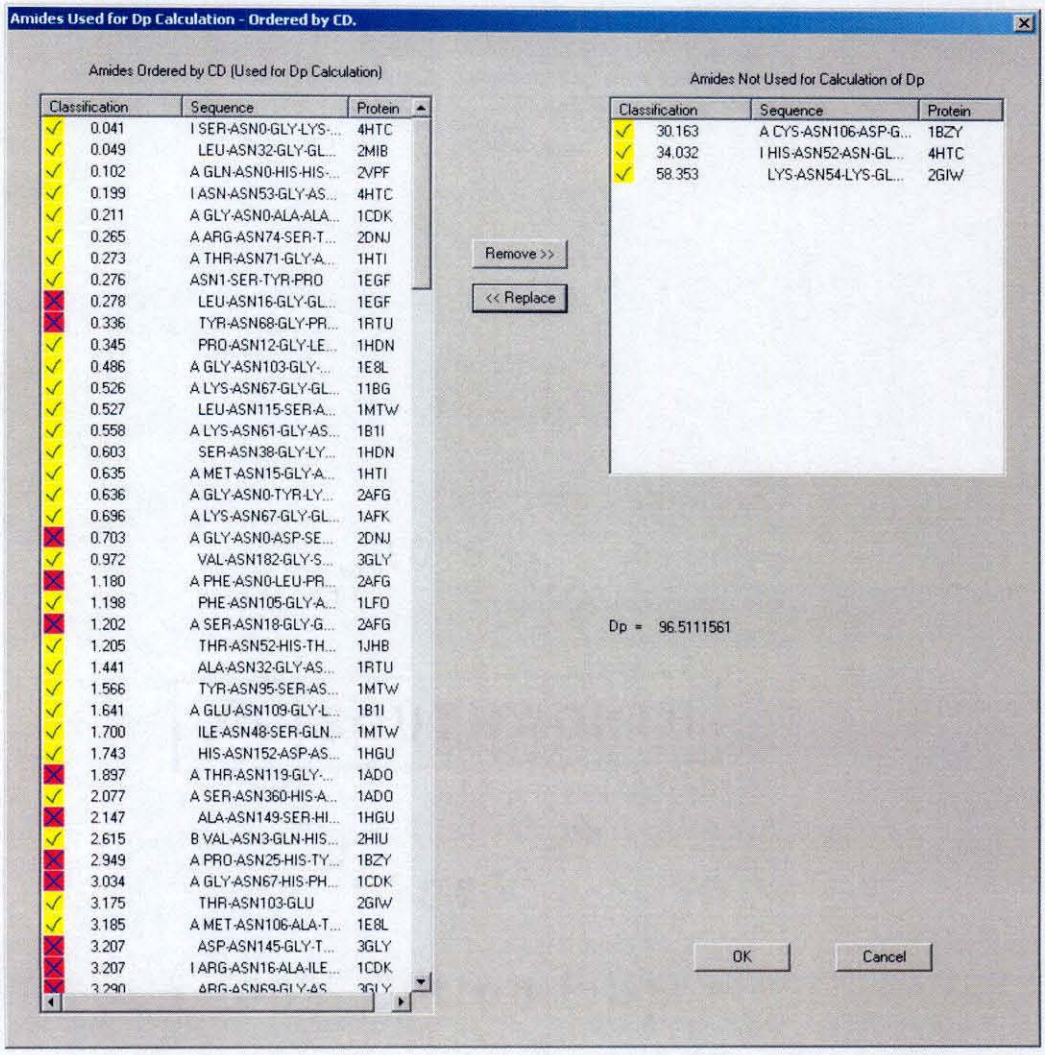


Figure 10-4 – Individual Amide Viewing and Sorting Window

Viewing the C_D rankings and removing an amide from the optimizations is made possible by the dialog shown in Figure 10-4.

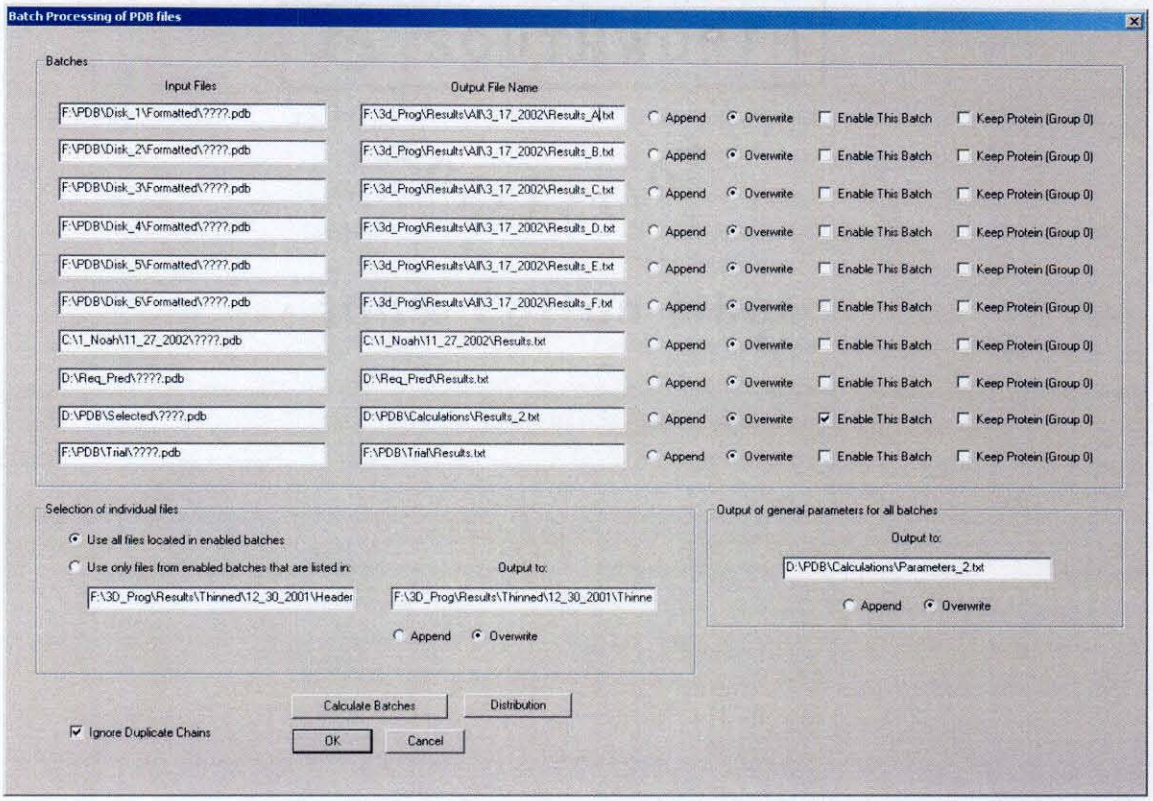


Figure 10-5 – Batch Calculations Window

Proteins can be individually loaded into the program for calculation of their C_D and I_D values. For large numbers of proteins–particularly when calculating the entire the protein databank–the batch window shown in Figure 10-5 is used.

Optimization of Parameters

Optimization of parameters was done using the parameters optimization window shown in Figure 10-6. This made it possible to recalculate D_P for a range of variations in one parameter at a time. The data are plotted and the optimum point automatically selected. Sometimes a value slightly different from the actual maximum is manually selected

because of noise in the data. In Figure 10-6 the optimum point shown is 0.46. From the trend of the graph it is clear that 0.48 is closer to the center, so this value was input manually. The smoothed data are shown in Figure 10-7.

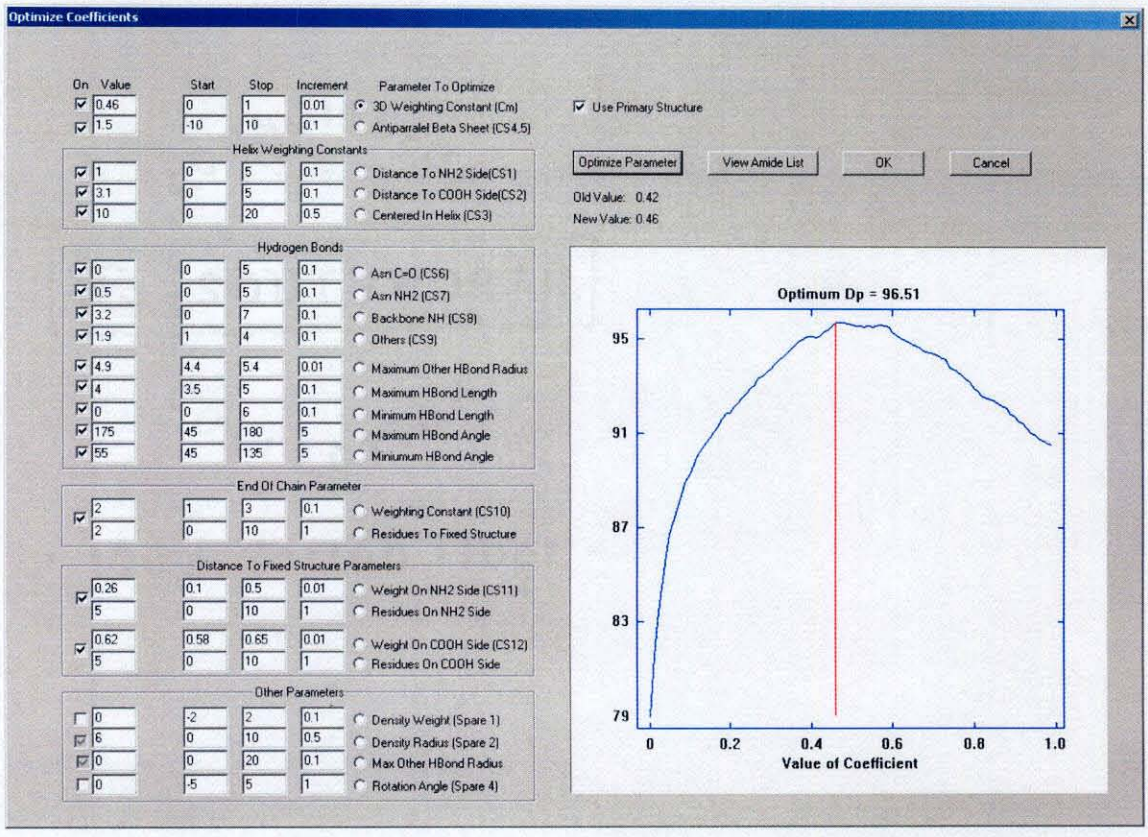


Figure 10-6 – Parameter Optimization Window

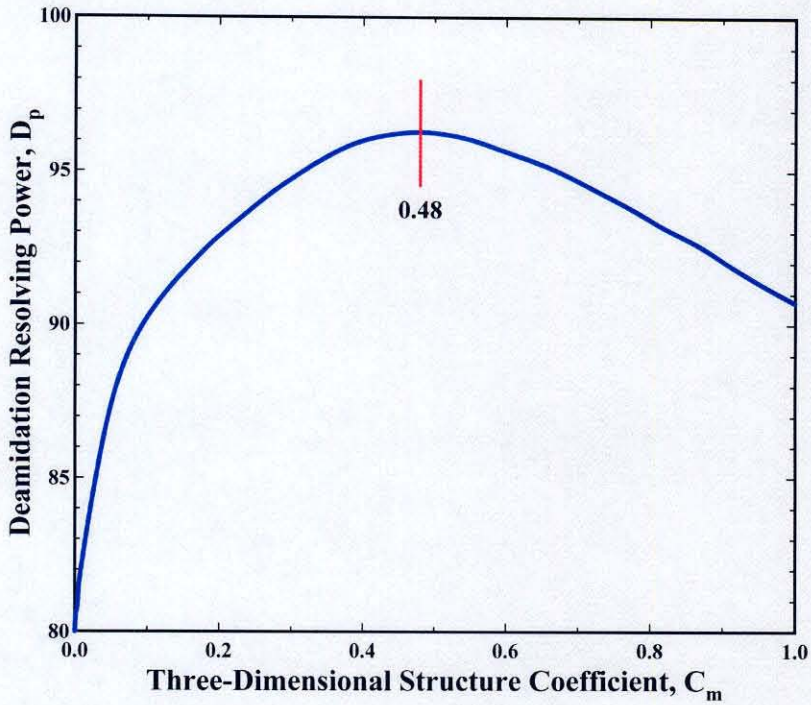


Figure 10-7 – Deamidation Resolving Power as a Function of C_m

After careful optimization of all parameters, the values chosen for the coefficients were $C_m = 0.48$, $C_{S1} = 1.0$, $C_{S2} = 3.1$, $C_{S3} = 10$, $C_{S4,5} = 1.5$, $C_{S6} = 0$, $C_{S7} = 0.5$, $C_{S8} = 3.2$, $C_{S9} = 1.9$, $C_{S10} = 2.0$, $C_{S11} = 0.26$, and $C_{S12} = 0.62$. The D_P was found to be 96.5%.

Application of Program to Protein Data Bank

Using this procedure, the deamidation rates were predicted for every amide in the protein databank.[9] At the time of publication, we calculated values for 131,809 asparagine residues in 10,369 proteins. The data analysis in the next section is based on this data set. The size of the protein databank has increased to over 20,000 proteins since that time, and a new set has been calculated. The most recent version is available at http:\\www.deamidation.org.

Results and Analysis of Computerized Prediction Procedure

The most important results of this work are in the estimated deamidation rates of individual proteins. These calculations provide reliable estimates of instability with respect to deamidation for the Asn residues in all of the amide-containing proteins in the protein 3-D structure database and are easily performed for any other protein for which a 3-D structure becomes available.

These estimated deamidation rates depend upon peptide rates that were experimentally determined in 37.00 °C, pH 7.4, 0.15 M Tris·HCl buffer and upon calculations that estimate the relative contributions of primary and 3-D structure. Therefore, the estimated deamidation rates are for 37.00 °C, pH 7.4, 0.15 M Tris·HCl buffer. These baseline solution conditions were chosen because Tris catalyzes deamidation to a much lesser extent than most other buffers. Phosphate buffers and the solute mixtures typically found in living things usually increase deamidation rates at least two or threefold as compared with those in Tris.[8] The deamidation half-times shown in Figures 10-8 and 10-9 are, thus, longer than those expected in living things at 37 °C.

In addition, these calculations provide some general insights about deamidation and its prevalence in proteins, as illustrated in Figures 10-8 and 10-9, and Tables 10-1, and 10-2.

Figure 10-8a shows the cumulative distribution function of the estimated first-order deamidation rate constants in 170,014 asparaginyl residues in 17,935 proteins, whereas Fig. 10-8b shows the derivative of the function in 10-8a, with the unusually unstable amides illustrated. For these figures, 2,966 redundant protein structures in the Protein Data Base have been excluded. These curves and those for human proteins[8] (Figure 9-8) are qualitatively and quantitatively similar. In the course of this work, these functions also have been calculated for mouse, rat, chicken, *Bacillus subtilis*, and *Escherichia coli* proteins. Although not identical, these distributions are all quite similar, with mouse, *B. subtilis*, and *E. coli* having somewhat higher numbers of unstable amides in their proteins as compared with human, rat, and chicken.

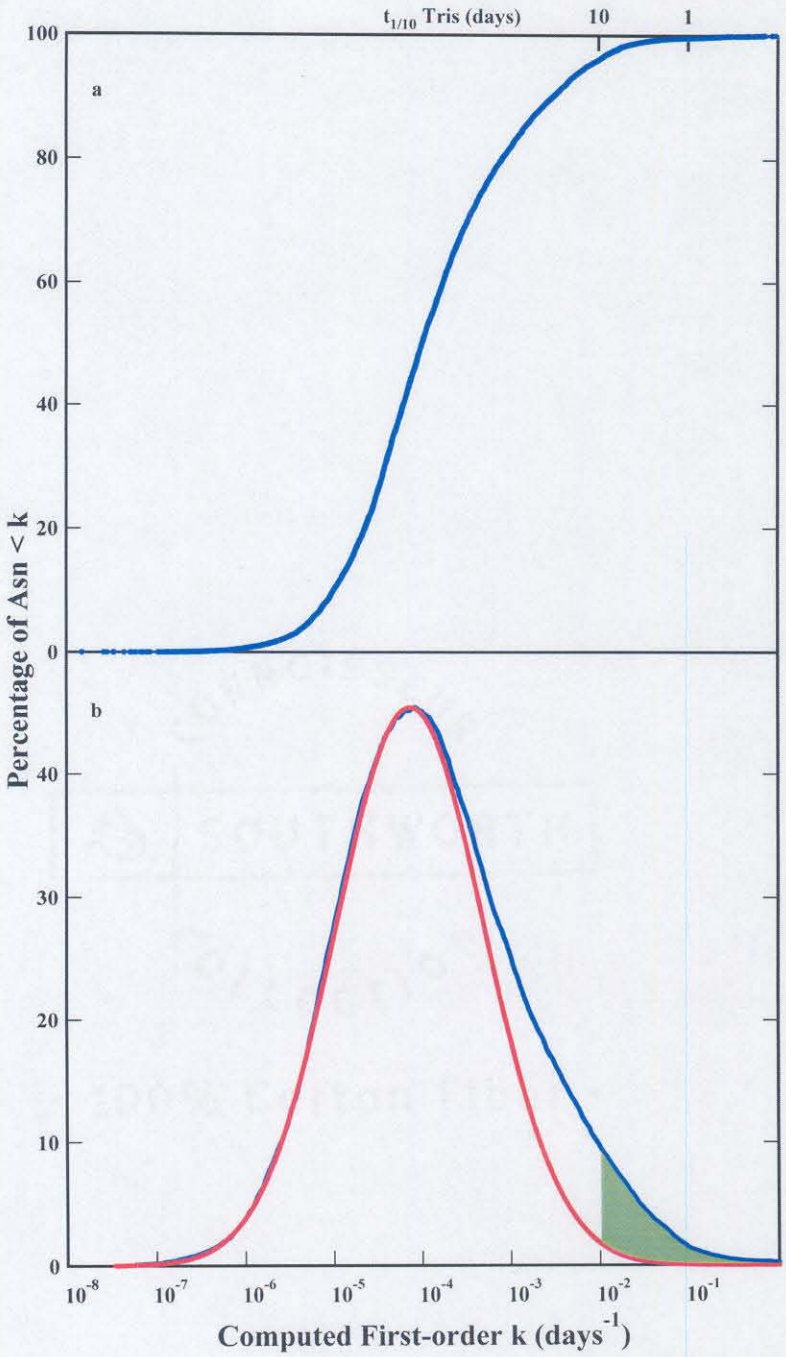


Figure 10-8 – Cumulative Distribution and Derivatized Curve for 170,014 Amides in the PDB Databank.

As has been reported for human proteins,[8] deamidation is not a random consequence of the presence of Asn residues in proteins. The fast deamidations summarized in Figure 10-9 result from a set of Asn residues with unusual primary and 3-D structures, which comprise about 5% of the total. These need not have been incorporated in protein structures, because most individual Asn deamidation rates are slower.

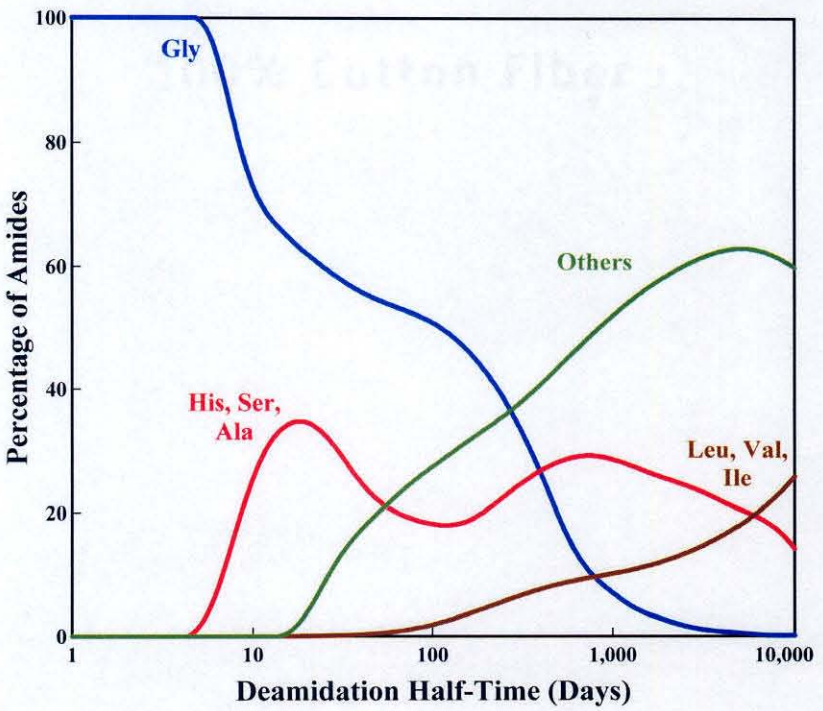


Figure 10-9 – Cumulative Distribution Plots for 131,809 Amides in PDB Databank for Predicted Half-Times in 37.00 °C, pH 7.4, 0.15 M Tris With Various Right-Hand Residues

Table 10-1 – Percentages of Residues and Proteins in Asn Deamidation Ranges

Timer	Amides, cumulative no.	Amides, %	Proteins, %
Tris 1/2 time, days			
<1	20	0.02	0.19
<2.5	162	0.12	1.5
<5	473	0.36	3.7
<10	695	0.53	5.6
<25	1659	1.3	13
<50	3591	2.7	25
<100	7080	5.4	41
Phos. 1/10 time, days			
<1	538	0.41	4.3
<2.5	1101	0.84	8.8
<5	2208	1.7	17
<10	5029	3.8	32
<25	10798	8.2	52
<50	16340	12	64
<100	23652	18	75

Whereas both the amino-side and carboxyl-side residues immediately adjacent to the amide residues affect deamidation rates, the carboxyl side residue is more important.[6] Asn deamidation at neutral pH has been reported to proceed primarily by means of a succinimide mechanism, which involves an intermediate ring structure on the carboxyl side of the deamidating residue.[75–78, 28] Figure 10-9 illustrates the relative importance of the different carboxyl side residues as a function of protein deamidation half-time. Although there are rare reported instances of 3-D structures that increase deamidation rates, most rates are determined by primary structure as modulated through slowing by 3-D structure. Therefore, only AsnGly sequences in locations that are relatively sterically unhindered provide Tris deamidation half-times of less than about 6 days. Most AsnGly sequences have longer half-times because of 3-D effects.

Table 10-2 – Percentages of Carboxyl Side Residues in Asn Deamidation Ranges

Residue	%, < 5 days	%, < 10 days	%, < 25 days	%, < 100 days
Tris 1/2 time				
Gly	4.2	6.0	11	38
His		0.46	2.4	7.7
Ser		0.19	2.3	9.8
Ala			1.3	5.6
Cys			0.23	1.1
Thr			0.1	2.0
Asp			0.060	4.8
Lys			0.014	1.8
Glu				3.0
Gln				1.7
Arg				1.6
Asn				1.5
Phe				1.1
Met				0.93
Tyr				0.89
Trp				0.37
Leu				0.33
Phosphate 1/10 time				
Gly	14	27	53	87.3
His	3.6	5.9	12	33.2
Ser	2.9	8.1	14	34.6
Ala	1.9	4.2	7.2	17.9
Cys	0.23	0.84	2.6	7.3
Thr	0.14	1.5	4.2	12.3
Asp	0.93	1.9	6.8	19
Lys	0.071	1.2	4.4	10.4
Glu	0.057	1.6	5.3	13.1
Gln	0.041	1.4	3.9	10.1
Arg	0.035	1.0	3.4	9.4
Asn	0.15	1.4	5.0	13.5
Phe		0.46	2.3	7.3
Met	0.036	0.75	2.4	5.4
Tyr	0.038	0.36	1.2	5.4
Trp		0.16	0.74	5.0
Leu		0.26	1.3	5.4
Val			0.01	2.8
Ile			0.013	1.3

Primary structure deamidation rates in Tris for AsnHis, AsnSer, and AsnAla peptide sequences range from 6 to 32 days,[6, 73] so these sequences and additional, more hindered AsnGly sequences dominate this deamidation half-time range and that immediately higher. Within 100-day half-times, all other sequences except for AsnLeu, AsnVal, AsnIle, and AsnPro significantly contribute. The estimated deamidation half-times between 100 and 10,000 days are qualitatively and semi-quantitatively useful, but these rates lack the direct experimental verification available for the shorter half-times.[8] Other processes such as additional deamidation mechanisms, side chain oxidation, peptide bond cleavage, and racemization also become more important at the longer time intervals. These deamidation half-times are likely to be at least 2- to 3-fold shorter in biological systems at 37 °C.[8]

Each of the curves in Figure 10-9 contains a subgroup of amides that are close to the ends of the proteins. These subgroups arise from Asn sequences with relatively short deamidation half-times, as a result of their positions near the ends of protein chains, where they are largely free of constraining 3-D effects. The Asn residues that have been reported to control the turnovers of rat cytochrome *c*[10, 11] and rabbit muscle aldolase[12, 13] are of this type. In both of these cases, there is marked preferential *in vivo* degradation of the deamidated protein molecules. Because the estimated range of Asn deamidation half-times available in these subgroups is between 0.2 days and more than 200 days under most *in vivo* conditions at 37 °C, and they are free of 3-D constraints, this type of Asn residue is especially well suited for biological purposes.

Figure 10-7 illustrates, by means of the parameter C_m, the optimization routines that were used for all of the parameters in these calculations. By using the deamidation resolving power technique,[7, 8] all suitable and published experimental relative deamidation observations of proteins are used to optimize each adjustable 3-D parameter. The relative proportions of primary and 3-D structure effects are determined by optimization of the illustrated parameter C_m as shown in Figure 10-7. At the optimum value of $C_m = 0.48$, the experimental deamidation rates are 96.5% correctly ordered by

the calculations. The remaining 4% of disorder includes amides wherein the inhibitory effects of 3-D structure are imperfectly estimated by this calculation method and those where the experiments themselves are in error. Moreover, most of the deamidation rates of the imperfectly ordered 4% of Asn residues are estimated approximately correctly.

Optimization of the deamidation resolving power at 96.5% requires both primary and 3-D structure. When primary structure alone is used, D_P = 79.8%. With 3-D structure alone, D_P = 64.3%.

Table 10-1 shows the percentages of 131,809 asparaginyl residues in 10,369 proteins with 37 °C Tris buffer and phosphate buffer deamidation half-times in various ranges and the percentages of proteins with at least one amide with a deamidation half-time within those ranges. Phosphate is a stronger catalyst of deamidation than is Tris[3, 4, 8, 157]. The phosphate deamidation half-times are estimated for Table 10-1 as 1/2 the Tris half-times.[8] Under *in vivo* conditions at 37 °C, deamidation half-times would be expected to be, on average, even shorter than those estimated in Table 10-1 for phosphate.[8] As shown, about 17% of all proteins are estimated to have at least one amide that is 1/10 or more deamidated after 5 days in phosphate at 37 °C. Approximately 4.3% are 1/10 or more deamidated after 1 day. At 37 °C *in vivo*, these percentages are expected to be higher.[8] Accelerated protein turnover of deamidated forms and other factors can obscure the observation of these deamidations in ordinary experiments on protein preparations extracted from living things.

Table 10-2 provides quantitative summaries for 131,809 specific AsnXxx sequences. For example, Table 10-2 shows that 12% of all AsnHis sequences are estimated to be at least 1/10th deamidated after 25 days in 37 °C phosphate buffer. In addition to these values, Table 10-2 also illustrates that it is not possible to usefully determine the relative deamidation rates of specific amides within a protein on the basis of primary sequence alone. Whereas the effects of primary sequence are evident, even in the simplest case of AsnGly, most deamidation half-times are substantially determined by a combination of primary and 3-D structure.

Figure 10-10 illustrates the percentage of primary structure vs. three-dimensional structure that make up deamidation rates in 131,809 amides from the Protein Databank. The rise on the left-hand side is due to amides that have no three-dimensional structure constraints. These amides deamidate at essentially sequence controlled rates. Most amides, however, lie toward center of the graph with a median of 53.2% secondary, tertiary, and quaternary contributions.

The estimate based on amides that have been found to deamidate in proteins gives 40% secondary, tertiary, and quaternary contributions,[7] as mentioned earlier. The higher value for amides taken from the Protein Databank is understandable, since these amides have not been selected as especially fast ones.

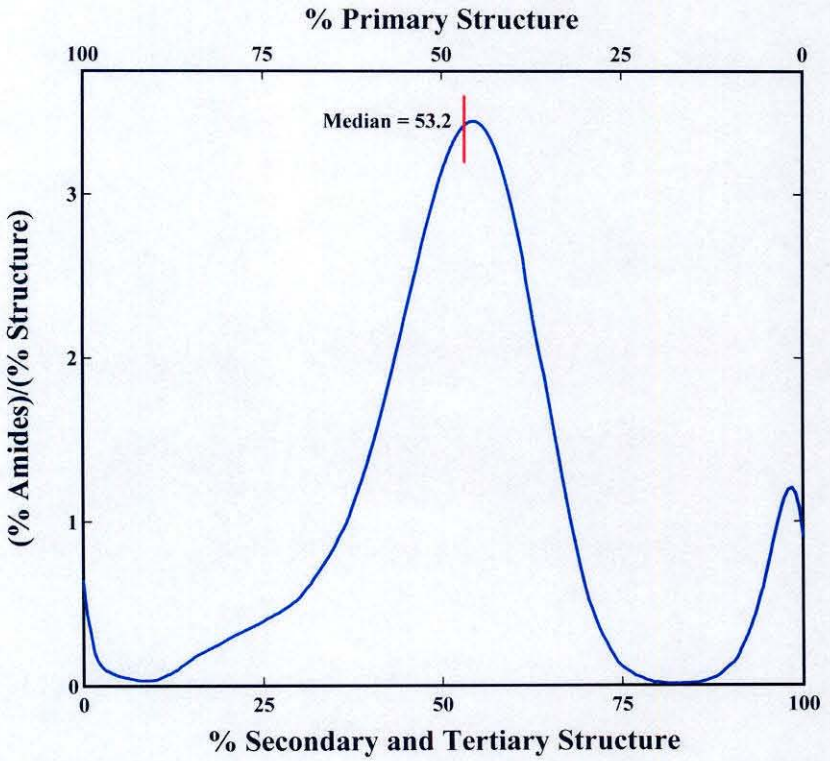


Figure 10-10 – Plot of Contributions of Primary vs. Secondary, Tertiary, and Quaternary Structure to Overall Deamidation C_D

Gln Deamidation Predictions

A similar treatment for Gln deamidation in proteins has not yet been done. The first impediment was in determining the Gln sequence dependence, which was not completed until recently.[26] Deamidation rates for 30 Gln peptides had been reported,[3, 25] but these measurements were carried out in phosphate buffer, which substantially complicates their use. It has been found that the reported deamidation rates for Asn peptides in phosphate can be self-consistently described,[158] although they would not have been suitable for accurate predictive work, but phosphate effects on deamidation rates of Gln peptides are apparently much larger than for Asn.[72]

The Gln pentapeptide rates reported here are consistent with the *in vivo* rates found for long-lived proteins.[36–39] It is still necessary to develop a predictive method, although some preliminary steps have been taken and are very promising. A direct application of the current method–which uses constants derived based on short-lived Asn in proteins–to both long-lived Asn and Gln produced rates in a few crystallin proteins that are roughly 1/3 the correct *in vivo* values. Until enough protein data exist, it will be difficult to determine correctly the necessary Gln coefficients.

An interesting example of the comparison between Gln and Asn deamidation is the deamidation of the protein Interleukin-1β. The mouse protein contains an amide with the sequence LeuAsn(32)Gly and deamidates at this position[128] with a half-time of 1.5 days in 37 °C, pH 8.5, 0.1 M Tris. There is one hydrogen bond as illustrated in Figure 10-11. The computerized procedure predicts a C_D value of 0.0388 and therefore a half-time of 3.9 days (the original hand procedure gave a predicted half-time of 3.8 days). This is excellent agreement, especially considering the pH difference.

In human Interleukin-1β this Asn residue has been replaced by a Gln and the half-life has been measured[159] as 240 days at 30 °C, pH 7.0, 0.01M Tris. A prediction procedure for Gln has not yet been established, but Figure 10-11 shows that this amide is essentially free and unconstrained in the loop and has no hydrogen bonds, so an approximate

sequence controlled rate would be expected. In fact the GlyLeuGlnGlyGly rate has been measured as 670 days.

The comparison between 240 days and 670 days is good, especially since there are additional things to consider. First, the conditions are not quite the same, but a lower temperature for the protein should provide a longer half-time, and peptide work indicates that this pH change would as well.[72, 73] A more important effect is shown in Figure 10-10. Although somewhat difficult to visualize without a three-dimensional view, the backbone conformation is ideal for forming a glutarimide. It is only necessary for the side chain–which is completely free–to swing into the correct position in order to react. This may very well accelerate the protein rate beyond that of the peptide.

It is worth noting that the backbone conformations of the human and mouse protein are essentially the same. While this conformation is not unusually unfavorable for forming a succinimide in the mouse protein and may cause some acceleration, the difference in length of the Gln chain probably allows the glutarimide to form with a greater relative acceleration.

This is one example of Gln deamidation and is easily understood based on the current prediction model and the peptide deamidation rates. A prediction procedure can be developed for Gln as soon as enough of these examples are available.

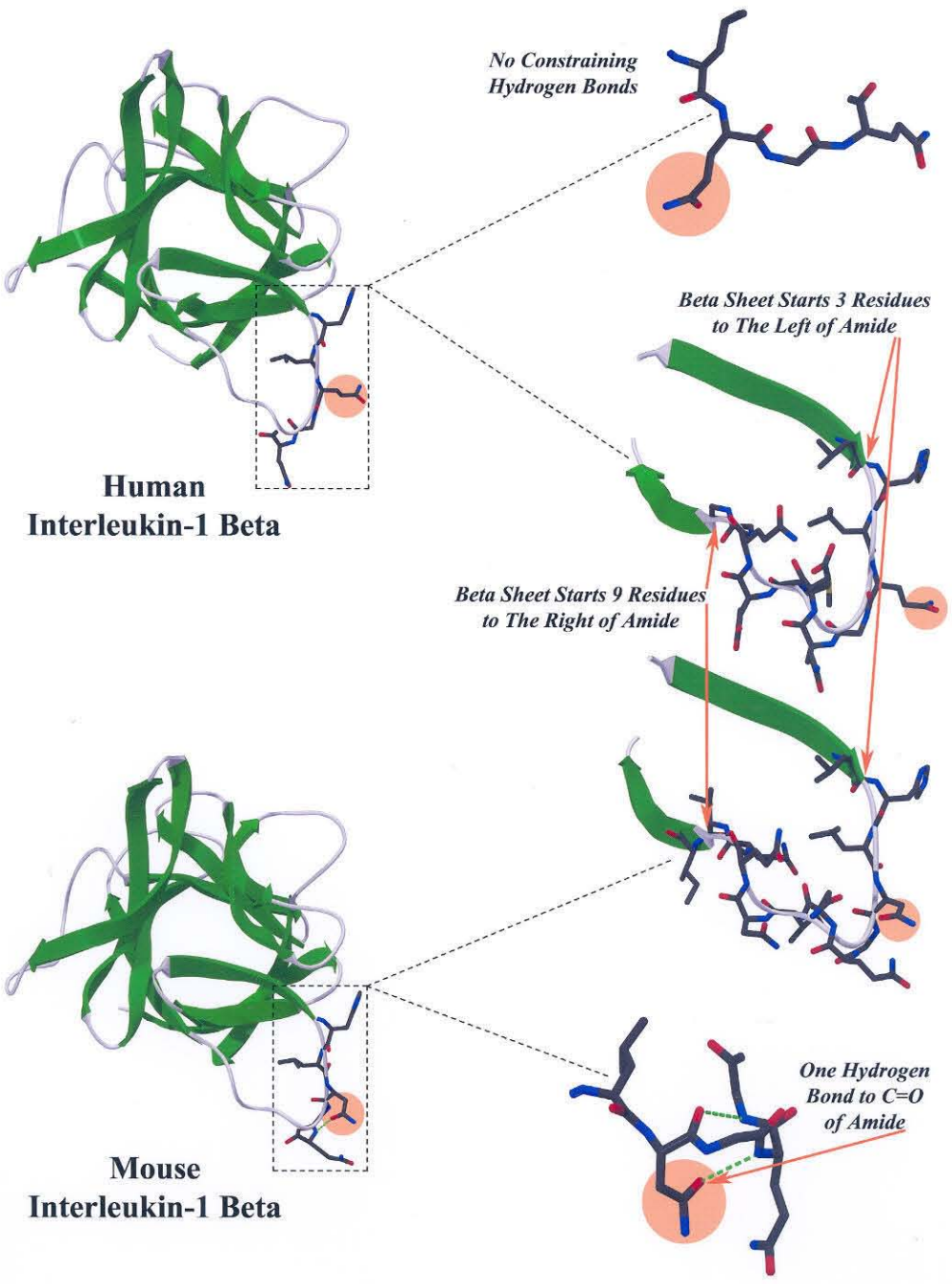


Figure 10-11 – Comparison Between Human (5I1B) and Mouse (8I1B) Interleukin-1β

of less than about one hundred days, where the procedure has been tested. Preliminary work with Gln and long-lived Asn residues in proteins indicates that this procedure is also accurate for long Asn rates and extendable to Gln once the calibration data are available.

Applications to Protein Chemistry

Deamidation is a chemical reaction of significance to the reactions that occur in living things. A general but inadequate knowledge of sequence controlled deamidation was present before this work. The peptide work described here substantially expanded–by a factor of more than 5–the available rates and utilized reaction conditions that allowed extension to protein deamidation. Sequence controlled deamidation is, as a result of the work reported herein, fairly well understood and is shown to cover a wide half-time range of more than 4 orders of magnitude.

The prediction procedure developed for Asn and expandable to Gln shows how this sequence dependence on deamidation can be combined with protein structure to determine protein deamidation rates. Not only does three-dimensional structure modify the sequence controlled rates, but this work shows that it is also possible essentially to completely turn off or turn on a deamidation site by means of three-dimensional structure. While this was generally evident 30-years ago[3] and statements to this effect with reference to specific cases have occasionally appeared in the literature since then, no quantitative treatment has previously been attempted.

At the very least, it is essential that protein chemists be aware of the extent to which deamidation is occurring in the proteins they are studying. This work now makes it possible to predict deamidation rates in any protein for which the three-dimensional structure is known.

The possibility that glutaminyl and asparaginyl residues may serve as ubiquitous molecular timers of biological processes increases the relevance of this work. This work demonstrates more than ever before the fine control and wide range of time intervals which

are available. A choice of two amino acids in a row, first an amide–either glutamine or asparagine–followed by another residue can set this clock to almost any value within a range of four orders of magnitude. The addition of a third residue to the left of the amide provides finer control over the exact rate. Three-dimensional structure provides additional control over the rate and can even turn the clock completely off. This feature allows protein deamidation rates to respond to changes in three-dimensional structure initiated by other factors. This last adjustment is a function of protein structure and is open to modification at any point during the life of the protein. In effect, amide sentinels can be imbedded throughout the protein, ready to send their signals of deamidation should a change of structure occur.

We believe, based on the many features of this reaction, that peptide and protein deamidation will eventually be found to be a ubiquitous part of the general timing and regulation mechanisms in biological systems. Very little is currently known about these systems.

Applications to Biomolecular Clock Hypothesis

We now state this hypothesis is as follows

Two of the twenty ordinary amino acid residues in peptides and proteins, Asn and Gln, especially Asn, are uniquely unstable under physiological conditions. The non-enzymatic deamidation half-times of these residues are genetically programmable throughout the range from less than a day to more than a century with many within the biological lifetimes of the proteins, organelles, and organisms of which they are a part.

Deamidation of Asn or Gln introduces negative charges into proteins and, to a lesser but significant extent, isomerization as well.

Why do peptides and proteins include residues that are unstable and cause fundamental changes in the structures of peptides and proteins during the biological lifetimes of those molecules?

These timed changes of structure must be of substantial value to living things. Otherwise, they would be unnecessarily disruptive to the integrity of the molecules essential to life, and would not be permitted.

We suggest that this value resides in the ability of amide residues to serve as molecular clocks for the timing of biological processes.

Originally, this hypothesis presupposed that there were not other undiscovered properties of Asn and Gln of such great biological importance that their presence in proteins is necessary regardless of instability. It is now known that a large percentage of proteins deamidate during their in vivo lifetimes, but that this deamidation occurs in a small percentage of amide structures. If Asn and Gln were needed for other purposes, many stable configurations are available.

Amide residues serve as biological molecular clocks. A few examples of these clocks are now known. We suggest that amide clocks will be found to be ubiquitous in living things.

While this general hypothesis, without the included specifics provided by the work described herein, predates this thesis project,[1-3, 5, 160] the experimental results herein add two very important discoveries.

First:

The assumption of genetic determination of deamidation rates, which has until now been dependent upon scattered and largely qualitative observations, has been placed upon a firm quantitative foundation with a thorough understanding of the effects of primary, secondary, tertiary, and quaternary structure.

Second:

This quantitative understanding has permitted reliable estimation of the distribution functions of deamidation rates in thousands of proteins. These rates show that a substantial fraction of proteins are genetically programmed to deamidate in biologically relevant timed intervals even though most of the genetically available structures deamidate more slowly. This quantitative preference for fast deamidation times relative to slow deamidation times is an entirely new discovery. The fact that this property is present within the thousands of proteins of currently known three-dimensional structure markedly strengthens the molecular clock hypothesis.

Publications

1. Robinson, N. E. and Robinson, A. B. (2001) Molecular Clocks. *Proc. Natl. Acad. Sci. USA* **98**, 944-949.

2. Robinson, N. E., Robinson, A. B., and Merrifield, R. B. (2001) Mass Spectrometric Evaluation of Synthetic Peptides as Primary Structure Models for Peptide and Protein Deamidation. *J. Pept. Res.* **57**, 483-493.

3. Robinson, N. E. and Robinson, A. B. (2001) Prediction of Protein Deamidation Rates from Primary and Three-Dimensional Structure. *Proc. Natl. Acad. Sci. USA* **98**, 4367-4372.

4. Robinson, N. E. and Robinson, A. B. (2001) Deamidation of Human Proteins. *Proc. Natl. Acad. Sci. USA* **98**, 12409-12413.

5. Robinson, N. E. (2002) Protein Deamidation. *Proc. Natl. Acad. Sci. USA* **99**, 5283-5288.

6. Carvalho, R. M. N., Solstad, T., Robinson, N. E., Robinson, A. B., and Flatmark, T. (2002) Possible Contributions of Labile Asparagine Residues to Difference in Regulatory Properties of Human and Rat Phenylalanine Hydroxylase. *Proceedings of the 12th International Symposium on the Chemistry and biology of Pteridines and Folates.*

7. Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, J. A., Robinson, M. L., and Robinson, A. B. (2004) Structure-Dependent Nonenzymatic Deamidation of Glutaminyl and Asparaginyl Pentapeptides. *J. Pept. Res.* **63**, 426-436.

8. Robinson, N. E. and Robinson A. B. (2004) Prediction of Primary Structure Deamidation Rates of Asparaginyl and Glutaminyl Peptides Through Steric and Catalytic Effects. *J. Pept. Res.* **63**, 437-448.

9. Robinson, N. E. and Robinson A. B. (2004) Amide Molecular Clocks In Drosophila Proteins: Potential Regulators of Aging and Other Processes. *Mechanisms of Ageing and Development.* **125**, 259-267.

10. Robinson, N. E. and Robinson A. B. (2003) Amide Molecular Clocks are Ubiquitous in Proteins. *Prepared.*

11. Robinson, N. E. (2003) Non-Enzymatic Deamidation Mediated Protein Turnover As a Possible Element of the Ubiquitin Pathway. *Submitted to Journal of Theoretical Biology.*

References

1. Robinson, A. B., McKerrow, J. H., and Cary, P. (1970) *Proc. Natl. Acad. Sci. USA* **66**, 753-757.
2. Robinson, A. B. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 885-888.
3. Robinson, A. B. and Rudd, C. J. (1974) *Curr. Top. Cell. Regul.* **8**, 247-295.
4. Robinson, A. B. and Scotchler, J. W. (1974) *Int. J. Pept. Protein Res.* **6**, 279-282.
5. Robinson, A. B. and Robinson, L. R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8880-8884.
6. Robinson, N. E. and Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 944-949.
7. Robinson, N. E. and Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 4367-4372.
8. Robinson, N. E. and Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 12409-12413.
9. Robinson, N. E. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5283-5288.
10. Flatmark, T. and Sletten, K. (1968) *J. Biol. Chem.* **243**, 1623-1629.
11. Robinson, A. B., McKerrow, J. H., and Legaz, M. (1974) *Int. J. Pept. Protein Res.* **6**, 31-35.
12. Midelfort, C. F. and Mehler, A. H. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1816-1819.
13. McKerrow, J. H. and Robinson, A. B. (1974) *Science* **183**, 85.
14. Lai, C. Y., Chen, C., and Horecker, B. L. (1970) *Biochem. Biophys. Res. Commun.* **40**, 461-468.
15. Gracy, R. W. and Yuan, P. M. (1980) *Federation Proceedings* **39**, 1690.
16. Sun, A., Yuksel, K. U., and Gracy, R. W. (1995) *Arch. Biochem. Biophys.* **322**, 361-368.
17. Gracy, R. W., Talent, J. M., and Zvaigzne (1998) *J. Exp. Zoology* **282**, 18-27.
18. Deverman, B. E., Cook, B. L., Manson, S. R., Niederhoff, R. A., Langer, E. M., Rosová, I., Kulans, L. A., Fu, X., Weinberg, J. S., Heinecke, J. W., Roth, K. A., and Weintraub, S. J. (2002) *Cell* **111**, 51-62.
19. Takehara, T. and Takahashi, H. (2002) *Cancer Research* **63**, 3054-3057.

20. Takehara, T. and Takahashi H. (2000) *Gastroenterology* **118**, 2433.

21. Yarnell A. (2003) *Chemical and Engineering News* **81**, 47-49.

22. Helfman, P. M. and Bada, J. L. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2891-2894.

23. Stadtman, E. R. and Levine, R. L. (2000) *Ann. NY Acad. Sci.* **899**, 191-208.

24. Harding, J. J. (1985) *Advances in Protein Chemistry* **37**, 247.

25. Robinson, A. B., Scotchler, J. W., and McKerrow, J. H. (1973) *J. Am. Chem. Soc.* **95**, 8156-8159.

26. Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, J. A., Robinson, M. L., and Robinson, A. B. (2004) *J. Pept. Res.* **63**, 426-436.

27. Robinson, A. B. and Tedro, S. (1973) *Int. J. Pept. Protein Res.* **5**, 275-278.

28. Geiger, T. and Clarke, S. (1987) *J. Biol. Chem.* **262**, 785-794.

29. Johnson, B. A., Murray, E. D. Jr., Clarke, S., Glass, D. B., and Aswad, D. W. (1987) *J. Biol. Chem.* **262**, 5622-5629.

30. Johnson, B. A., Shirokawa, J. M., Hancock, W. S., Spellman, M. W., Basa, L. J., and Aswad, D. W. (1989) *J. Biol. Chem.* **264**, 14262-14271.

31. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234**, 179-185.

32. Manjula, B. N., Acharya, A. S., and Vithayathil, P. J. (1976) *Int. J. Pept. Protein Res.* **8**, 275-282.

33. Capasso, S. and Salvadori, S. (1999) *J. Peptide Res.* **54**, 377-382.

34. Solstad, T. and Flatmark, T. (2000) *Eur. J. Biochem.* **267**, 6302-6310.

35. Solstad, T., Carvalho, R. N., Andersen, O. A., Waidelich, D., and Flatmark, T. (2003) *Eur. J. Biochem.* **270**, 929-938.

36. Palmer, W. G. and Papconstantinou, J. (1969) *Proc. Natl. Acad. Sci. USA* **64**, 404-410.

37. Takemoto, L. and Boyle, D. (1998) *Biochemistry* **37**, 13681-13685.

38. Takemoto, L. (1998) *Current Eye Res.* **17**, 247-250.

39. Lampi, K. J., Oxford, J. T., Bachinger, H. P., Shearer, T. R., and David, L. L. (2001) *Exp. Eye Res.* **72**, 279-288.

40. Lapko, V. N., Purkiss, A. G., Smith, D. L., and Smith, J. B. (2002) *Biochemistry* **41**, 8638-8648.

41. De Jong, W. W., Mulders, J. W. M., Voorter, C. E. M., Berbers, G. A. M., Hoekman, and Bloemendal, W. A. (1988) *Adv. Exp. Med. Biol.* **231**, 95-108.

42. Groenen, P. J. T. A., van Dongen, M. J. P., Voorter, C. E. M., Bloemendal, H., and de Jong, W. W. (1993) *Fed. European Biochem. Soc.* **322**, 69.

43. Krichevskaia, A. A., Lucash, A. I., Pushkina, N. V., Shepotinovskaia, I. V., and Sherstnev, K. B (1984) *Nauchnye Doki Vyss Shkoly Biol Nauki*, 23-28.

44. Ohguro, H., Konno, S., Konari, K., Kitamura, K., Sohma, H., Nakagawa, T., and Akino, T. (1996) *Sapporo Medical Journal* **65**, 349.

45. Takemoto, L. and Emmons, T. (1991) *Curr. Eye Res.* **10**, 865-869.

46. van Kleen, F. S. M., de Jong, W. W., and Henders, H. J. (1975) *Nature* **258**, 264-266.

47. Takemoto, L. and Boyle, D. (2000) *J. Biol. Chem.* **275**, 26109-26112.

48. Vauquelin, L. N. and Robiquet, P. J. (1806) *Ann. Chim. (Paris)* **57**, 88.

49. Dulong A. (1826) *J. Pharm.* **12**, 278.

50. Plisson, A. (1827) *J. Pharm.* **13**, 477.

51. Liebig, J. and Wohler (1832) *J. L. Ann. Chem.* **3**, 268.

52. Pasteur, L. (1851) *Ann.* **80**, 146.

53. Hlasiwetz, H. and Habermann, J. (1873) *Ann.* **169**, 150.

54. Schulze, E. and Bosshard, E. (1883) *Ber.* **16**, 312.

55. Damodarm, M., Jaaback, G., and Chibnall, A. C. (1932) *Biochem. J.* **26**, 1704.

56. Foreman, F. W. (1914) *Biochem. J.* **8**, 463.

57. Melville, J. (1935) **29**, 179.

58. Gilbert, J. B., Price, V. E., and Greenstein, J. P. (1949) *J. Biol. Chem.* **180**, 209.

59. Greenstein, J. P. and Winitz, M. Chemistry of the Amino Acids 3, Wiley, New York, 1856 and 1929 (1961).

60. Pauling, L., The Nature of the Chemical Bond, 3rd Edition, Cornell, Ithaca, 281 (1962)

61. Harfenist, E. J., and Craig, L. C. (1952) *J. Am. Chem. Soc.* **74**, 3083.

62. Flatmark, T. (1964) *Acta Chem. Scand.* **18**, 1656-1666.

63. Flatmark, T. (1966) *Acta Chem. Scand.* **20**, 1487-1496.

64. Flatmark, T. and Sletten, K. (1967), 413-421.

65. Flatmark, T. and Vesterberg, O. (1966) *Acta Chemica Scandinavica* **20**, 1497-1503.

66. Forman, F. W., (1914) *Biochem. J.* **8**, 463.

67. Vickery, H. B., Pucher, G. W., Clark, H. E., Chibnall, A. C., and Westall, R. G. (1935) *Biochem. J.* **29**, 2710.

68. Lee, J. C., Gray, H. B., and Winkler, J. R. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 7760-7764.

69. Lee, J. C., Engman, K. C., Tezcan, F. A., Gray, H. B. and Winkler, J. R. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 14778-14782.

70. Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85**, 2149-2154.

71. Merrifield, R. B. (1995) Solid-phase peptide synthesis. In Peptides: Synthesis, Structures, and Applications (B. Gutte, ed.). Academic Press, New York, pp. 94-169.

72. Scotchler, J. W. and Robinson, A. B. (1974) *Anal. Biochem.* **59**, 319-322.

73. Robinson, N. E., Robinson, A. B., and Merrifield, R. B. (2001) *J. Pept. Res.* **57**, 483-493.

74. McKerrow, J. H. and Robinson, A. B. (1971) *Anal. Biochem.* **42**, 565-568.

75. Bornstein, P. and Balian, G. (1970) *J. Biol. Chem.* **245**, 4854-4856.

76. Meinwald, Y. C., Stimson, E. R., and Scheraga, H. A. (1986) *Int. J. Pept. Protein Res.* **28**, 79-84.

77. Tam, J. P., Rieman, M. W., and Merrifield, R. B. (1988) *Peptide Research* **1**, 6-18.

78. Capasso, S., Mazzarella, L., and Zagari, A. (1991) *Peptide Res.* **4**, 234-238.

79. Lowenson, J. D. and Clarke, S. (1991) *Gerontology* **37**, 128-151.

80. McFadden, P. N. and Clarke, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2460-2464.

81. Lindner, H., Sarg, B., Hoertnagl, B., and Helliger, W. (1998) *J. Biol. Chem.* **273**, 13324-13330.

82. Johnson, B. A., Shriokawa, J. M., and Aswad, D. W. (1989) *Archives of Biochemistry and Biophysics* **268**, 276-286.

83. Lewis, U. J., Cheever, E. V., and Hopkins, W. C. (1970) *Biochim. Biophys. Acta* **214**, 498-508.

84. Lorentz, K (1979) *Clinica Chimica Acta* **93**, 161-162.

85. Lukash, A. I., Shepelev, A. P., Pushkina, N. V., Tsybul'skii, I. E., and Al'perovich, D. V. (1985) *Vopr Med Khim* **31**, 104-108.

86. Minic, Z., Hranisavljevic, J., and Vucelic, D. (1997) *Biochem. Mol. Biol. Int.* **41**, 1057-1066.

87. Raba, R. E. and Aaviksaar, A. A. (1982) *Bioorganicheskaya Khimiya* **8**, 707-709.

88. Zaman, G., Saphier, P. W., Loveridge, N., Kiumura, T., Sakakibara, S., Bernier, S. M., and Hen, G. N. (1991) *Endocrinology* **128**, 2583-2590.

89. Perkins, M., Theiler, R., Lunte, S., and Jeschke, M. (2000) **17**, 1110-1117.

90. Stevensen, C. L., Manning, M. C., and Borchardt, R. T. (1989) *Biophysical Journal* **55**, 357A.

91. Kosky, A. A., Razzaq, U. O., Treuheit, M. J., and Brems, D. N. (1999) *Protein Sci.* **8**, 2519-2523.

92. Lide, D. R. Editor (1995) *Handbook of Chemistry and Physics 75th Edition.*

93. Capasso, S., Mazzarella, L, Sica, F., and Zagari, A. (1991) *Journal of the Chemical Society Chemical Communications* **23**, 1667-1668.

94. Blom, N. and Sygusch, J. (1997) *Nat. Struct. Biol.* **4**, 36-39.

95. Hallahan, T. W., Shapiro, R., Strydom, D. J., and Vallee, B. L. (1992) *Biochemistry* **31**, 8022-8029.

96. Leonidas, D. D., Shapiro, R., Allen, S. C., Subbarao, G. V., Veluraja, K., and Acharya, K. R. (1999) *J. Mol. Biol.* **285**, 1209-1233.

97. Chazin, W. J., Kordel, J., Thulin, E., Hofmann, T., Drakenberg, T., and Forsen, S. (1989) *Biochemistry* **28**, 8646-8653.

98. Svensson, L. A., Thulin, E., and Forsen, S. (1992) *J. Mol. Biol.* **223**, 601-606.

99. Jedrzejewski, P. T., Girod, A., Tholey, A., Konig, N., Thullner, S., Kinzel, V., and Bossemeyer, D. (1998) *Protein Sci.* **2**, 457-469.

100. Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., and Huber, R. (1993) *EMBO J.* **12**, 849-859.

101. Banci, L., Bertini, I., Huber, J. G., Spyroulias, G. A., and Turano, P. (1999) *J. Biol. Inorg. Chem.* **4**, 21-33.

102. DiAugustine, R. P., Gibson, B. W., Aberth, W., Kelly, M., Ferrua, C. M., Tomooka, Y., Brown, C. F., and Walker, M. (1987) *Anal. Biochem.* **165**, 420-429.

103. Montelione, G. T., Wuthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D., and Scheraga, H. A. (1992) *Biochemistry* **31**, 236-249.

104. Odani, S., Okazaki, Y., Kato, C., Uchiumi, T., and Takahashi, Y. (1993) *Arch. Biochem. Biophys.* **309**, 81-84.

105. Thompson, J., Winter, N., Terwey, D., Bratt, J., and Banaszak, L. (1997) *J. Biol. Chem.* **272**, 7140-7150.

106. Volkin, D. B., Verticelli, A. M., Bruner, M. W., Marfia, K. E., Tsai, P. K., Sardana, M. K., and Middaugh, C. R. (1994) *J. Pharmacol. Sci.* **84**, 7-11.

107. Blaber, M., Disalvo, J., and Thomas, K. A. (1996) *Biochemistry* **35**, 2086-2094.

108. Svensson, B., Larson, K., Svendsen, I., and Boel, E. (1983) *Carlsberg Res. Commun.* **48**, 529-544.

109. Aleshin, A. E., Hoffman, C., Firsov, L. M., and Honzatko, R. B. (1994) *J. Mol. Biol.* **238**, 575-591.

110. Silberring, J., Brostedt, P., Ingvast, A., and Nyberg, F. (1991*) Rapid Commun. Mass Spectrom.* **5**, 579-581.

111. Chantalat, L., Jones, N. D., Korber, F., Navaza, J., and Pavlovsky, A. G. (1995) *Protein Peptide Lett.* **2**, 333.

112. Huisman, T. H. J., Carver, M. F. H., and Efremov, G. D. (1998) in A Syllabus of Human Hemoglobin Variants 2nd Ed. (Univ. of Georgia, Augusta, GA)

113. Wajcman, H., Kister, J., Vasseur, C., Blouquit, Y., Trastour, J. C., Cottenceau, D., and Galacteros, F. (1992) *Biochim. Biophys. Acta* **1138**, 127-132.

114. Paleari, R., Paglietti, E., Mosca, A., Mortarino, M., Maccioni, L., Satta, S., Cao, A., and Galanello, R. (1999) *Clin. Chem.* **45**, 21-28.

115. Wajcman, H., Vasseur, C., Blouquit, Y., Santo, D. E., Peres, M. J., Martins, M. C., Poyart, C., and Galacteros, F. (1991) *Am. J. Hematol.* **38**, 194-200.

116. Hutt, P. J., Donaldson, M. H., Khatri, J., Fairbanks, V. F., Hoyer, J. D., Thibodeau, S. N., Moxness, M. S., McMorrow, L. E., Green, M. M., and Jones, R. T. (1996) *Am. J. Hematol.* **52**, 305-309.

117. Moo-Penn, W. F., Jue, D. L., Bechtel, K. C., Johnson, M. H., and Schmidt, R. M. (1976) *J. Biol. Chem.* **251**, 7557-7562.

118. Seid-Akhavan, M., Winter, W. P., Abramson, R. K., and Rucknagel, D. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 882-886.

119. Blackwell, R. Q., Boon, W. H., Liu, C. S., and Weng, M. I. (1972) *Biochim. Biophys. Acta* **278**, 482-490.

120. Tame, J. and Vallone, B. (2000) *Acta Crystallogr. D Biol. Cryst.* **56**, 805-811.

121. Sharma, S., Hammen, P. K., Anderson, J. W., Leung, A., Georges, F., Hengstenberg, W., Klevit, R. E., and Waygood, E. B. (1993) *J. Biol. Chem.* **268**, 17695-17704.

122. Brennan, T. V., Anderson, J. W., Jia, Z., Waygood, E. B., and Clarke, S. (1994) *J. Biol. Chem.* **269**, 24586-24595.

123. Van Nuland, N. A., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J., van Gunsteren, W. F., Scheek, R. M., and Robillard, G. T. (1994) *J. Mol. Biol.* **237**, 544-559.

124. Wilson, J. M., Landa, L. E., Kobayashi, R., and Kelley, W. N. (1982) *J. Biol. Chem.* **257**, 14830-14834.

125. Shi, W., Li, C. M., Tyler, P. C., Furneaux, R. H., Grubmeyer, C., Schramm, V. L., and Almo, S. C. (1999) *Nat. Struct. Biol.* **6**, 588-593.

126. Brange, J., Langkjaer, L., Havelund, S., and Volund, A. (1992) *Pharm. Res.* **9**, 715-726.

127. Hua, Q. X., Gozani, S. N., Chance, R. E., Hoffmann, J. A., Frank, B. H., and Weiss, M. A. (1995) *Nat. Struct. Biol.* **2**, 129-138.

128. Daumy, G. O., Wilder, C. L., Merenda, J. M., McColl, A. S., Geoghegan, K. F., and Otterness, I. G. (1991) *FEBS Lett.* **278**, 98-102.

129. Van Oostrum, J., Priestle, J. P., Grutter, M. G., and Schmitz, A. (1991) *J. Struct. Biol.* **107**, 189-195.

130. Robinson, N. E. (2000) *Unpublished*

131. Sasaoki, K., Hiroshima, T., Kusumoto, S., and Nishi, K. (1992) *Chem. Pharm. Bull.* **40**, 976-980.

132. Brandhuber, B. J., Boone, T., Kenney, W. C., and McKay, D. B. (1987) *Science* **238**, 1707-1709.

133. Kato, A., Tanimoto, S., Muraki, Y., Kobayashi, K., and Kumagai, I. (1992) *Biosci. Biotechnol. Biochem.* **56**, 1424-1428.

134. Schwalbe, H., Grimshaw, S. B., Spencer, A., Buck, M., Boyd, J., Dobson, C. M., Redfield, C. and Smith, L. J. (2001) *Protein Sci., in press.*

135. Wearne, S. J. and Creighton, T. E. (1989) *Proteins Struct. Funct. Genet.* **5**, 8-12.

136. Leonidas, D. D., Shapiro, R., Irons, L. I., Russo, N., and Acharya, K. R. (1997) *Biochemistry* **36**, 5578-5588.

137. Kanaya, S. and Uchida, T. (1986) *Biochem. J.* **240**, 163-170.

138. Noguchi, S., Satow, Y., Uchida, T., Sasaki, C., and Matsuzaki, T. (1995) *Biochemistry* **34**, 15583-15591.

139. Di Donato, A., Galletti, P., and D'Alessio, G. (1986) *Biochemistry* **25**, 8361-8368.

140. Vitagliano, L., Adinolfi, S., Sica, F., Merlino, A., Zagari, A., and Mazzarella, L. (1999) *J. Mol. Biol.* **293**, 569-577.

141. Teshima, G., Porter, J., Yim, K., Ling, V., and Guzzetta, A. (1990) *Biochemistry* **30**, 3916-3922.

142. Wu, H., Myszka, D. G., Tendian, S. W., Brouillette, C. G., Sweet, R. W., Chaiken, I. M., and Hendrickson, W. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 15030-15035.

143. Papov, V. V., Gravina, S. A., Mieyal, J. J., and Biemann, K. (1994) *Protein Sci.* **3**, 428-434.

144. Sun, C., Berardi, M. J., and Bushweller, J. H. (1998) *J. Mol. Biol.* **280**, 687-701.

145. Yuan, P. M., Talent, J. M., and Gracy, R. W. (1981) *Mech. Ageing Dev.* **17**, 151-162.

146. Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A., and Hol, W. G. (1994) *Protein Sci.* **3**, 810-821.

147. Kossiakoff, A. A. (1988). *Science* **240**, 191-194.

148. Stubbs, M. T., Huber, R., and Bode, W. (1995) *FEBS Lett.* **375**, 103-107.

149. Cacia, J., Quan, C. P., Vasser, M., Sliwkowski, M. B., and Frenz, J. (1993) *J. Chromatogr.* **634**, 229-239.

150. Bischoff, R., Lepage, P., Jaquinod, M., Cauet, G., Acker-Klein, M., Clesse, D., Laporte, M., Bayol, A., Dorsselaer, A. V., and Roitsch, C. (1993) *Biochemistry* **32**, 725-734.

151. Potter, S. M., Henzel, W. J., and Aswad, D. W. (1993) *Protein Sci.* **2**, 1648-1663.

152. Goolcharran, C., Cleland, J. L., Keck, R., Jones, A. J. S., and Borchardt, R. T. (2000) *AAPS PharmSci.* **2**, U73-U86.

153. Robinson, A. B. and Westall, F. C. (1974) *J. Orthomolec. Psych.* **3**, 70-79.

154. Robinson, A. B. and Pauling, L. (1974) *Clin. Chem.* **20**, 961-965.

155. Hamilton, P. B. (1945) *J. Biol. Chem.* **158**, 375.

156. Robinson, A. B. and Scotchler, J. W. (1973) *J. Int. Res. Commun.* **8**, 1-8.

157. Yuksel, K. U. and Gracy, R. W. (1986) *Arch. Biochem. Biophys.* **248**, 452-459.

158. Capasso, S. (2000) *J. Peptide Res.* **55**, 224-229.

159. Gu, L. C., Erdös, E. A., Chiang, H., Calderwood, T., Tsai, K., Visor, G. C., Duffy, J., Hsu, W. C., and Foster, L. C. (1991) *Pharm. Res.* **8**, 485-490.

160. Robinson, A. B. (1979) *Mech. Ageing Dev.* **9**, 225-236.

161. Robinson, N. E. and Robinson A. B. (2004) *J. Pept. Res.* **63**, 437-448.

162. Capasso, S., Mazzarella, L., Sic, F., Zagari, A., and Salvadori, S. (1993) *J. Chem. Soc. Perkin. Trans. 2*, 679-682.

163. Robinson, N. E. and Robinson A. B. (2004) *Mechanisms of Ageing and Development.* **125**, 259-267.

164. Capasso, S., Mazzarella, L., Sorrentino, G., Balboni, G., and Kirby, A. (1996) *Peptides.* **17**, 1075-1077.