Protein deamidation

Noah E. Robinson[†]

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

Communicated by Bruce Merrifield, The Rockefeller University, New York, NY, February 21, 2002 (received for review January 3, 2002)

Protein deamidation

Noah E. Robinson†

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

Communicated by Bruce Merrifield, The Rockefeller University, New York, NY, February 21, 2002 (received for review January 3, 2002)

A completely automatic computerized technique for the quantitative estimation of the deamidation rates of any protein for which the three-dimensional structure is known has been developed. Calculations of the specific deamidation rates of 170,014 asparaginyl residues in 13,335 proteins have been carried out. The calculated values have good quantitative reliability when compared with experimental measurements. These rates demonstrate that deamidation may be a biologically relevant phenomenon in a remarkably large percentage of proteins.

asparaginyl residue deamidation | coefficient of deamidation | deamidation index

Changes in peptide and protein structure through the spontaneous nonenzymatic deamidation of glutaminyl and asparaginyl residues have been observed in many *in vitro* and *in vivo* experiments. Rates of deamidation of individual amide residues depend upon primary sequence, three-dimensional (3D) structure, and solution properties such as pH, temperature, ionic strength, and buffer ions (1–8).

Deamidation at neutral pH introduces a negative charge at the deamidation site and sometimes also leads to β isomerization. These alterations in structure affect the properties of peptides and proteins in chemically and biologically important ways. It has been suggested that *in vivo* deamidation of proteins serves as a molecular timer of biological events and as a mechanism for postsynthetic production of unique proteins of biological significance (2, 4, 6, 7, 9, 10). In the case of *in vivo* protein turnover, the use of deamidation as a molecular timer has been experimentally demonstrated (11–13).

Progress in the understanding of deamidation and its potential biological importance has been impeded by the lack of reliable and useful experimental and theoretical information about the deamidation of most proteins. Experimental studies of the deamidation of individual proteins are laborious and time consuming. Until recently, there were no other means by which to estimate the deamidation rates of specific amides.

The deamidation rates of individual Asn residues in a protein can now be reliably predicted as a result of two recent advances. First, the sequence-controlled Asn deamidation rates of most of the 400 possible near-neighbor combinations in pentapeptide models have been measured (10); the deamidation rates of a representative group of Gln pentapeptides have been determined (N.E.R. & A. B. Robinson, unpublished work); and the relevance of these rate libraries has been established (14). Second, these rates and the 3D structures of proteins with well characterized deamidations have been combined to produce a computation method that correctly predicts the deamidation rates of most Asn residues for which the 3D structure is known (15). This method has been shown to be ≈95% reliable in predicting relative deamidation rates of Asn residues within a single protein, and it is also useful for the prediction of absolute deamidation rates. It has been used to estimate the deamidation rates of 1,371 asparaginyl residues in 126 human proteins (16).

This semiempirical method for estimating deamidation rates (15) depends, however, upon manual observations of a set of 3D characteristics of the protein structure near each amide residue in the protein. These observations are laborious and subject to human error. In using this method, approximately 4 h of work by an experienced investigator with a computer-based 3D-structure viewer is required to determine the deamidation rates of the amides in a typical protein.

The computerized deamidation estimation method reported herein automatically determines essentially the same 3D characteristics as the manual method, with some significant refinements. It is, however, much faster, more reliable, and more convenient. During 4 h of continuous computation with a Pentium IV computer, the deamidation rates for the entire Brookhaven 3D protein database of 13,335 amide-containing protein structures with 170,014 asparaginyl residues have been calculated by this technique. These estimated deamidation rates, which include all amide-containing protein structures in the Brookhaven Protein Data Bank as of April 2001, have been deposited at www.deamidation.org. The computer program is also freely available at this Internet site, so investigators can compute these values for additional proteins.

These calculations provide new insights into the nature of protein deamidation. Moreover, this new method contains subroutines for automatic refinement of the calculation procedure as new experimental deamidation rates become available and also allows automatic calculation of additional estimated protein deamidation rates as new 3D structures are determined.

Materials and Methods

Calculation Method. Deamidation coefficients (C_D) for individual amides and deamidation indexes (I_D) for individual proteins were determined by a fully computerized procedure similar to the manual technique reported (15). Some parts of the calculation method were refined, and all of the adjustable parameters were automatically optimized. The computer program was written with Microsoft C++. The program uses Swiss Protein Data Bank Viewer to detect α-helices and β-sheets. The calculations were made by means of a Pentium IV computer with Microsoft WINDOWS 2000.

The deamidation coefficient, C_D, is defined as $C_D = (0.01)(t_{1/2})[e^{f(Cm,CSn,Sn)}]$, where $t_{1/2}$ is the primary structure half-life (10, 14), C_m is a structure proportionality factor, C_{Sn} is the 3D 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, are those that impede deamidation, including hydrogen bonds, α-helices, β-sheets, and peptide inflexibilities.

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 residue. 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

Abbreviations: 3D, three dimensional; I_D, deamidation index; C_D, deamidation coefficient; D_p, deamidation resolving power.

†To whom reprint requests should be addressed. E-mail: noahr@caltech.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

BIOCHEMISTRY

second residue, and 3 is the third residue. 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 anti-parallel β-sheets:

S_4 = number of residues in the loop. If the flexible region is not bounded on both sides by β-sheets with at least four hydrogen bonds between opposite NH and C=O groups, then this parameter is not used, so $S_4 = 0$.

S_5 = number of hydrogen bonds in the flexible region + 1.

For hydrogen bonds:

S_6 = the number of hydrogen bonds to the Asn sidechain C=O group. Acceptable values are 0, 1, and 2. Hydrogen bonds that are used in S_8 are not counted.

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

S_8 = the number of hydrogen bonds to the backbone N in the peptide bond on the COOH side of Asn. Acceptable values are 0 and 1. This nitrogen is used in the five-membered succinimide ring intermediate.

S_9 = the number of other hydrogen bonds that need to be broken to form the succinimide ring. The number of these bonds is estimated by first determining the angle through which the amide residue bond between the α carbon and the backbone carbonyl carbon would need to be rotated to align the backbone in the optimum configuration for forming the succinimide ring. This angle is then divided by 180° to give a number R between 0 and 1, where 0 is perfectly aligned and 1 is the worst alignment, requiring the greatest change in backbone position.

Residues that require greater chain movement to deamidate usually must break hydrogen bonds farther from the amide residue, so all hydrogen bonds are counted that are less than $(R)(C_R)$ from the C_α of the amide residue to either hydrogen bond atom and that are no more than five residues along the chain from the amide. The value of C_R was optimized at 4.97 Å. Hydrogen bonds that are counted in S_6, S_7, or S_8 are excluded from S_9.

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, and the Asn is 20 residues or less from the chain end:

$S_{10} = 1$ 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 if the Asn is not between a structure and the chain end, then $S_{10} = 0$. This parameter accounts for extra flexibility of the protein chain near the end. It is assumed that, in addition to the factors specifically included in these S_n, residues in the middle of proteins are additionally restrained, and that this part of the restraint is removed for amides near the end of the chain.

If the Asn lies near in sequence 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 between the usual atoms were assumed if the bond length was 4.1 Å or less, and the angle of the hydrogen bond was no more than 180° greater than the optimum nor 55° less than the optimum. This 4.1 Å value was determined by optimization and is greater than normally accepted for a hydrogen bond. The extra length may be caused by hydrogen-bond interactions that occur at different times because of flexibility in the protein. In most cases, the protein backbone must move to some extent to form the succinimide ring. If hydrogen bonds were assigned beyond the space available to accommodate the van der Waals radius of all of the hydrogen atoms, then the surplus bonds were rejected. For these rejections, the longest bonds with the most strained angles were removed.

X-ray crystal structures do not usually give hydrogen atom positions. For consistency, therefore, the hydrogen atom coordinates in NMR and neutron diffraction structures were not used.

The presence of hydrogen atoms was assumed wherever hydrogen atoms are ordinarily found in protein structures. In some cases, particularly –OH and $-NH_2$ groups, the orientation of the hydrogen atoms is uncertain. It was assumed that they would be positioned, where possible, to form the best hydrogen bonds.

The number of hydrogen atoms and donor groups available limits the number of hydrogen bonds. For example, an –OH group cannot have more than one hydrogen bond with atoms that are only hydrogen-bond acceptors. In cases where more than one donor or acceptor was available for forming a particular hydrogen bond, the one with the best characteristics was chosen. This choice involved a compromise between the optimum bond angle and optimum bond length.

All primary structure $t_{1/2}$ values used were those published (10, 14), except for sequences that have not been experimentally measured. Those included Asn with carboxyl-side Pro, Asn, Gln, or residues that are not among the ordinary 20, Asn without a free amide as a result of binding to metals or other moieties, and other unusual sequences. Estimates of $t_{1/2}$ values for these were approximated from those measured (10, 14).

The approximated values were: Asn-Pro, 500; Asn-Hyp, 500; Asn-Asn, 40; Asn-Gln, 60; bound Asn, 500; Asn-Xxx, where Xxx is carboxyl terminal, 1/2 of pentapeptide value; Yyy-Asn-Xxx, where Yyy is not one of the ordinary 20 amino acid residues but is instead Mse, Ini, Llp, Pca, Cgu, Tyq, Tyy, Tpq, Cea, or Cme, mean value for the ordinary amino acids at Yyy (10); Yyy-Asn-Xxx, where Xxx is Tyq, Tyy, or Tpq, same value as for Tyr; Xxx is Cea or Cme, same value as for Cys; where Xxx is Ini, same value as for Arg; Asn as carboxylamide terminal, 40; Asn as carboxyl terminal, 500; Asn as amino terminal, 2/3 pentapeptide values; and Xxx-Asn-Mse, 4/3 Met value.

Structures for 22 proteins from the Brookhaven Protein Data Bank combined with their relative Asn instabilities (15, 16) were used in the optimization of the coefficients C_m and C_{Sn}. These structures included all of the Asn residues in rabbit aldolase chain A, 1ADO; human angiogenin chain A, 1B1I; pig c-amp-dependent protein kinase chain A and chain I, 1CDK; mouse epidermal growth factor, 1EGF; rat fatty acid binding protein, 1LFO; human fibroblast growth factor chain A, 2AFG; *Aspergillus awamorii* glucoamylase, 3GLY; human growth hormone, 1HGU; 1HGU; *Escherichia coli* hpr-phosphocarrier protein, 1HDN; human insulin chain A and chain B, 2HIU; mouse interleukin 1β, 2MIB; chicken lysozyme chain A, 1E8L; bovine ribonuclease A chain A, 1AFK; *Ustilago sphaerogena* ribonuclease U2, 1RTU; bovine seminal ribonuclease chain A, 11BG; human thioltransferase, 1JHB; human triosephosphate isomerase chain A, 1HTI; bovine trypsin, 1MTW; bovine DNase chain A, 2DNJ; and human vascular endothelial growth factor chain A, 2VPF. Horse heart cytochrome *c* (2GIW) was used, except for Lys-Asn 54-Lys, which is reported to be accelerated by a 3D structure change caused by an earlier single deamidation (11, 12). No 3D structure is available for this singly deamidated form. Human hypoxanthine guanine phosphoribosyltransferase chain A (1BZY) was used except for Lys-Asn 106 Asp, which is held by 3D structure in exact position for the succinimide intermediate in the deamidation reaction and is, therefore, accelerated.

The coefficients C_m and C_{Sn} were optimized by means of the deamidation resolving power (D_p) method (15, 17). The optimized values 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%.

The protein deamidation index is defined as $I_D = [\Sigma(C_{Dn})^{-1}]^{-1}$, where C_{Dn} is C_D for the nth Asn residue. Therefore, $(100)(I_D)$ is an estimate of the initial deamidation halftime for the protein with all Asn residues considered.

The coefficients of deamidation, C_{Dn}, and deamidation index, I_D, for each protein were calculated by means of computer programs that are available for use at www.deamidation.org. These can be used for any new protein in the following way.

Primary sequence data for the protein in the format of the "SEQRES" section of the Brookhaven Protein Data Bank and 3D data—including the x, y, and z coordinates for each atom—in the format of the "ATOM" and "CONECT" sections of Brookhaven Protein Data Bank are automatically computer-entered into the program. No manual entry of data are required. Note that if no coordinates are entered for the atoms of a particular residue, it is assumed that this residue is unrestrained and does not participate in a unique 3D structure, which can modulate deamidation. After the data are entered, the computer program calculates the C_{Dn} and I_D values for the protein. The program output includes these values and also the values of S_n for each amide residue, so that the investigator can know the details about the calculation of C_{Dn} for each amide.

This calculation method is conceptually general. The molecular rearrangements necessary for formation of a cyclic deamidation intermediate are usually impeded by protein structure. Those structural impediments have been parameterized. The success of the method does not, however, prove the cyclic imide reaction mechanism, because other mechanisms may be similarly impeded.

Only relative deamidation rates, in the form of reports about which amides have been observed to be deamidated in 22 proteins under a wide variety of conditions, have been used to calibrate the method by optimizing the coefficients, C_m and C_{Sn}. No rate measurements were used for the optimization. Rate of deamidation measurements in 13 proteins at 37°C were used only to test and verify the method (16). Nine of these 13 proteins are in the set of 22 coefficient calibration proteins, and 4 are not.

At no time in its development was this calculation method adapted to any specific aspect of the structures of these 26 individual proteins. The method fails, therefore, for Asn-106 of human hypoxanthine guanine phosphoribosyltransferase chain A, which is a rare example in which deamidation is actually accelerated by 3-D structure. In this case, Asn-106 is held by 3-D structure in exactly the correct position for the deamidation reaction, so the reaction is enhanced. Although a term could be added to the calculation method to allow for this, it would detract from the generality of the method, and Asn-106 is the only case of this type presently known.

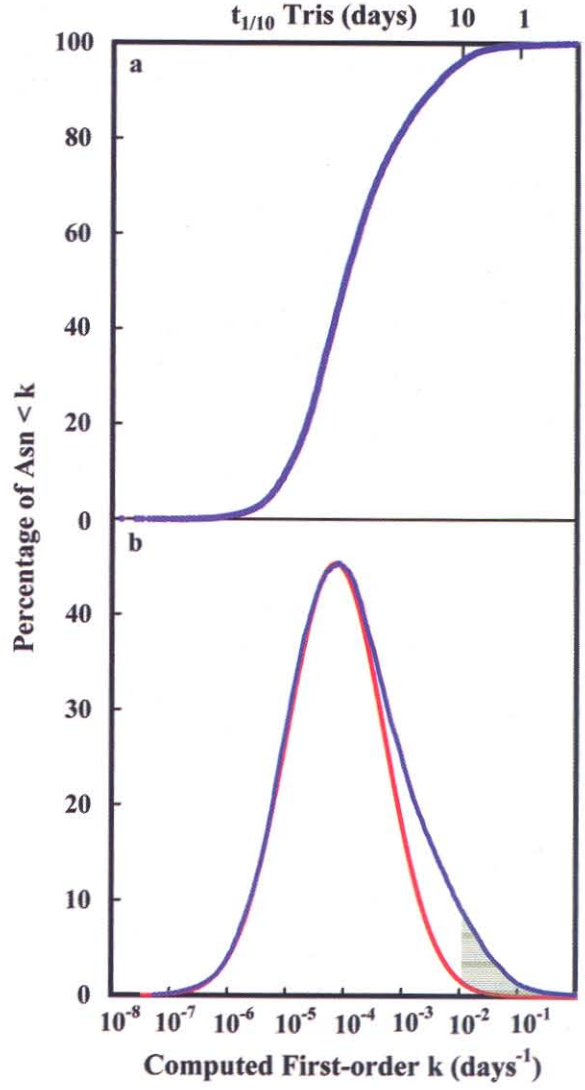

Fig. 1. (*a*) Cumulative distribution function of the calculated first-order rate constants for deamidation of the 131,809 Asn residues in 10,369 proteins used in Table 1. Asn residues involved in the initial deamidation of these proteins are a relatively small part of the complete set. The computed percentage of the Asn residues that are 1/10 deamidated after 10 days in pH 7.4, 37°C, 0.15 M Tris·HCl is 4%, as shown. (*b*) Differentiated values of the distribution function in *a* showing the special class of unstable Asn residues present in these proteins. Also shown with a red line is a Gaussian function that fits the distribution function, except for that part arising from the especially unstable Asn residues. The shaded area contains those Asn residues computed to be 1/10 or more deamidated in 10 days in pH 7.4, 37°C, 0.15 M Tris·HCl.

Results and Discussion

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 Brookhaven protein 3D structure database and are easily performed for any other protein for which a 3D structure becomes available.

In addition, these calculations provide some general insights about deamidation and its prevalence in proteins, as illustrated in Fig. 1, Fig. 2, Fig. 3, Table 1, and Table 2.

These estimated deamidation rates depend upon peptide rates that were experimentally determined in pH 7.4, 37°C, 0.15 M Tris·HCl buffer and upon calculations that estimate the relative contributions of primary and 3D structure. Therefore, the estimated deamidation rates are for pH 7.4, 37°C, 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 three-fold as compared with those in Tris (16). The deamidation halftimes shown in Figs. 1 and 2 are, thus, longer than those expected in living things at 37°C.

Fig. 1*a* shows the cumulative distribution function of the estimated first-order deamidation rate constants in 131,809 asparaginyl residues in 10,369 proteins, whereas Fig. 1*b* shows the derivative of the function in 1*a*, with the unusually unstable amides illustrated. For these figures, 2,966 redundant protein

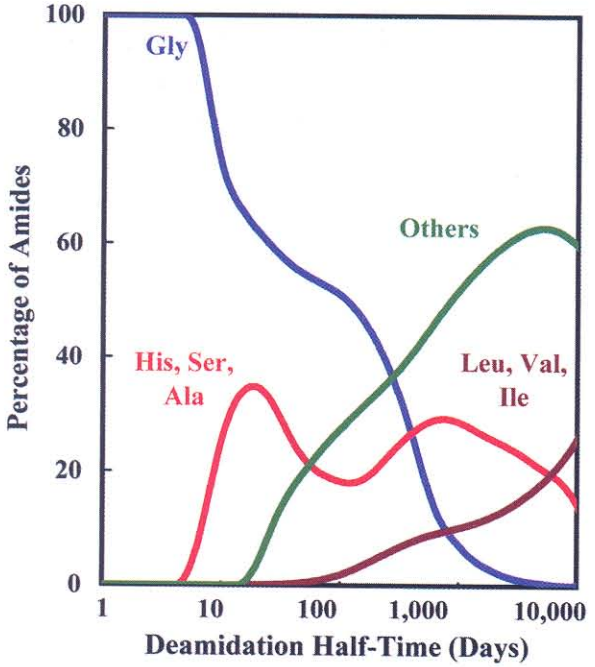

Fig. 2. Percentage of deamidating amides of the 131,809 Asn residues in 10,369 proteins used in Table 1 vs. deamidation half-time for Asn-Gly sequences; Asn-His, Asn-Ser, and Asn-Ala sequences; all sequences other than Asn-Gly, Asn-His, Asn-Ser, Asn-Ala, Asn-Leu, Asn-Val, and Asn-Ile; and Asn-Leu, Asn-Val, and Asn-Ile sequences. As deamidation halftime increases, more sequences contribute to deamidation. Except for a small number of especially sterically unhindered Asn residues, these deamidation halftimes strongly depend upon primary and 3D structures. These values are estimated for pH 7.4, 37°C, 0.15 M Tris·HCl and would be faster *in vivo* at 37°C.

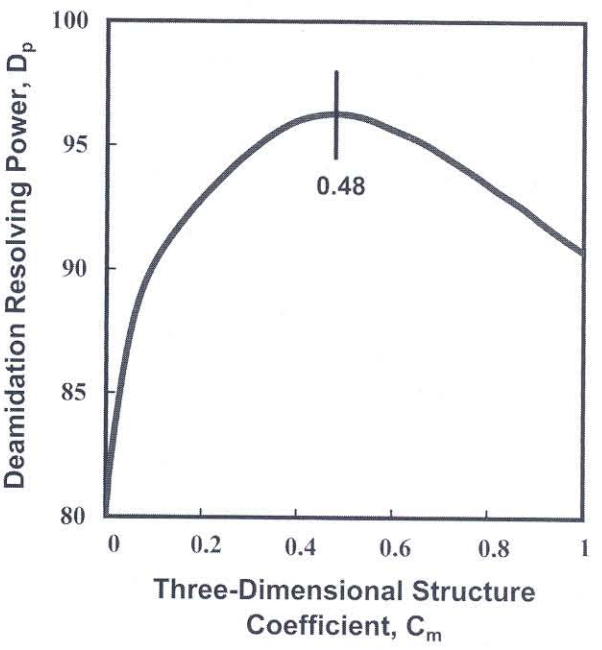

Fig. 3. Deamidation resolving power, D_p vs. 3D structure coefficient C_m for 192 Asn residues in 22 proteins. The optimum value of $C_m = 0.48$ provides the most reliable estimated relative deamidation rates as compared with the experimental values for these proteins.

structures in the Brookhaven Protein Data Base have been excluded. These curves and those for human proteins (16) 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.

As has been reported for human proteins (16), deamidation is not a random consequence of the presence of Asn residues in proteins. The fast deamidations summarized in Fig. 1 result from a set of Asn residues with unusual primary and 3D 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.

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 (10). 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 (18–20). Fig. 2 illustrates the relative importance of the different carboxyl side residues as a function of protein deamidation halftime. Although there are rare reported instances of 3D structures that increase deamidation rates, most rates are determined by primary structure as modulated through slowing by 3D structure. Therefore, only Asn-Gly sequences in locations that are relatively sterically unhindered provide Tris deamidation halftimes of less than about 6 days. Most Asn-Gly sequences have longer halftimes because of 3D effects.

Primary structure deamidation rates in Tris for Asn-His, Asn-Ser, and Asn-Ala peptide sequences range from 6 to 32 days (10, 14), so these sequences and additional, more hindered Asn-Gly sequences dominate this deamidation halftime range and that immediately higher. Within 100-day halftimes, all other sequences except for Asn-Leu, Asn-Val, Asn-Ile, and Asn-Pro significantly contribute. The estimated deamidation halftimes between 100 and 10,000 days are qualitatively and semiquantitatively useful, but these rates lack the direct experimental verification available for the shorter halftimes (16). 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

Table 1. Percentages of residues and proteins in asparagine 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

Table 2. Percentages of carboxyl side residues in asparagine 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			0.10	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	19.0
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	2.80
Ile			0.013	1.3

halftimes are likely to be at least 2- to 3-fold shorter in biological systems at 37°C (16).

Each of the curves in Fig. 2 contains a subgroup of amides that are close to the ends of the proteins. These subgroups arise from Asn sequences with relatively short deamidation halftimes, as a result of their positions near the ends of protein chains, where they are largely free of constraining 3D effects. The Asn residues that have been reported to control the turnovers of rat cytochrome *c* (11, 12) and rabbit muscle aldolase (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 halftimes 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 3D constraints, this type of Asn residue is especially well suited for biological purposes.

Fig. 3 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 (15, 17), all suitable and published experimental relative deamidation observations of proteins are used to optimize each adjustable 3D parameter. The relative proportions of primary and 3D structure effects are determined by optimization of the illustrated parameter C_m as shown in Fig. 3. 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 3D 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 3D structure. When primary structure alone is used, $D_p = 79.8\%$. With 3D structure alone, $D_p = 64.3\%$.

Table 1 shows the percentages of 131,809 asparaginyl residues in 10,369 proteins with 37°C Tris buffer and phosphate buffer deamidation halftimes in various ranges and the percentages of proteins with at least one amide with a deamidation halftime within those ranges. Phosphate is a stronger catalyst of deamidation than is Tris (6, 7, 16, 21). The phosphate deamidation halftimes are estimated for Table 1 as 1/2 the Tris halftimes (16). Under *in vivo* conditions at 37°C, deamidation halftimes would be expected to be, on average, even shorter than those estimated in Table 1 for phosphate (16). 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 (16). 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 2 provides quantitative summaries for 131,809 specific Asn-Xxx sequences. For example, Table 2 shows that 12% of all Asn-His sequences are estimated to be at least 1/10th deamidated after 25 days in 37°C phosphate buffer. In addition to these values, Table 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 Asn-Gly, most deamidation halftimes are substantially determined by a combination of primary and 3D structure.

The calculations reported herein include only Asn residue deamidation. The rates of deamidation of glutaminyl residue, Gln, containing pentapeptides have recently been determined for a set of 60 Gln peptides with a representative range of sequences (N.E.R. & A. B. Robinson, unpublished work). Gln deamidation, like Asn deamidation, depends strongly upon the residue on the carboxyl side of the amide, with Gln-Gly being the fastest to deamidate. Deamidation halftimes for Gln-Gly sequences in pH 7.4, 37°C, 0.15 M Tris·HCl buffer are in the range of 400–600 days (N.E.R. & A. B. Robinson, unpublished work). The values for other sequences are substantially longer. Amino terminal Gln is, however, expected to deamidate more rapidly to the pyrrolidone. Gln residues do not, therefore, significantly affect most of the deamidation index calculations reported herein. As with Asn, Gln deamidation is accelerated by phosphate buffer (5, 6).

Gln deamidations over periods of many years *in vivo* in slow-turnover proteins of long-lived organisms such as human eye lens proteins (22) can be of substantial biological significance. Gln-Gly sequences in sterically unhindered locations of proteins that lack shorter-lived Asn residues also affect the deamidation index. Where buffer ion or other deamidation catalysts are present, Gln deamidation can be more significant. Deamidation is currently being experimentally studied in many

proteins under *in vivo* and *in vitro* circumstances. For examples, see refs. 21–25.

Conclusions

Experimental (10, 14) and semiempirical computational (15, 16) advances have made possible reliable estimates of the deamidation rates of asparaginyl residues in proteins under biologically relevant conditions. These estimates, which require knowledge of the 3D structure of each molecule, are in accord with the known *in vivo* and *in vitro* experimental data and have been previously applied to 126 proteins of human origin (16).

It is reported herein that this method for estimating deamidation rates (15) has been fully computerized, further refined, and completely applied to the entire 170,014 Asn residues of the 13,335 amide-containing protein structures in the Brookhaven Protein Data Bank of 3D structures as of April 2001. These computed deamidation rates are available on the Internet at www.deamidation.org.

These calculation procedures make possible further refinements in the adjustable parameters, as more experimental data becomes available. Fully automatic deamidation estimates for additional proteins also may be made, as their 3D structures are determined. With the computer programs now freely available on the Internet at www.deamidation.org, an investigator can automatically enter the standard sequence and 3D parameters for any protein and calculate the deamidation coefficients and deamidation index for that protein.

The estimated deamidation rates apply directly to pH 7.4, 37°C, 0.15 M Tris·HCl buffer. Deamidation proceeds at a minimal rate under these conditions. There are no reported experimental deamidation rates at 37°C that are not the same as or faster than these computed Tris rates (16). The range of reported experimental values at 37°C is between the Tris rates and about 3-fold faster, depending upon solvent conditions. The average is about 2-fold faster.

The computed rates show that significant amounts of deamidation may be expected to occur in a substantial percentage of proteins under physiological conditions. Deamidation can control *in vivo* protein turnover. This result was found to be the case in cytochrome *c* (11, 12) and in rabbit muscle aldolase (13, 26, 27).

Many other biological processes such as homeostasis, development, and aging require timers, which are, as yet, unknown. Because protein deamidation rates can be genetically programmed throughout the range from a few hours to hundreds of years, and the deamidation rates found in proteins include many that are within the biologically significant range, deamidation is a good candidate for the timing of many biological processes.

Deamidation also provides a means of producing postsynthetic varieties of proteins that are biologically useful and of timing their release into *in vivo* systems.

Moreover, intermolecular effects resulting from *in vivo* protein locations such as in aggregates, membranes, and other organelles can further enhance the versatility exhibited herein in the intramolecular control of deamidation. A possible example of this has been reported for cytochrome *c* (6, 12).

Proteins contain amide residue clocks. These residues are found in almost all proteins, and amide residue clocks are found to be set to timed intervals of biological importance, even though settings to longer times are not only available, but also make up most of the genetically available settings. Deamidation changes protein structures in fundamentally important ways. If deamidation were not of pervasive and positive biological importance, these clocks would have been set to time intervals that are long with respect to the lifetimes of living things. The fact that they are found to be set instead to biologically relevant time intervals strongly supports the original hypothesis (2, 4, 6, 7, 9, 10) that amides play, through deamidation, a special biologically important role.

I thank Drs. H. B. Gray, R. B. Merrifield, and A. B. Robinson for advice and encouragement, the A. Reynolds Morse Foundation for their grant support, and a Caltech Special Institute Fellowship for financial support. Additional information, including computer programs, deamidation coefficients, and deamidation indexes for all proteins in the Brookhaven Protein Data Bank of 3D structures as of April 2001, is available at www.deamidation.org.

1. Flatmark, T. (1964) *Acta Chem. Scand.* **18,** 1656–1666.
2. Robinson, A. B., McKerrow, J. H. & Cary, P. (1970) *Proc. Natl. Acad. Sci. USA* **66,** 753–757.
3. Robinson, A. B., Scotchler, J. W. & McKerrow, J. H. (1973) *J. Am. Chem. Soc.* **95,** 8156–8159.
4. Robinson, A. B. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 885–888.
5. Scotchler, J. W. & Robinson, A. B. (1974) *Anal. Biochem.* **59,** 319–322.
6. Robinson, A. B. & Rudd, C. (1974) *Curr. Top. Cell. Reg.* **8,** 247–295.
7. Robinson, A. B. & Scotchler, J. W. (1974) *Int. J. Pept. Protein Res.* **6,** 279–282.
8. McKerrow, J. H. & Robinson, A. B. (1971) *Anal. Biochem.* **42,** 565–568.
9. Robinson, A. B. & Robinson, L. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8880–8884.
10. Robinson, N. E. & Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 944–949.
11. Flatmark, T. & Sletten, K. (1968) *J. Biol. Chem.* **243,** 1623–1629.
12. Robinson, A. B., McKerrow, J. H. & Legaz, M. (1974) *Int. J. Pept. Protein Res.* **6,** 31–35.
13. McKerrow, J. H. & Robinson, A. B. (1974) *Science* **183,** 85.
14. Robinson, N. E., Robinson, A. B. & Merrifield, R. B. (2001) *J. Peptide Res.* **57,** 483–493.
15. Robinson, N. E. & Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4367–4372.
16. Robinson, N. E. & Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 12409–12413.
17. Robinson, A. B. & Pauling, L. (1974) *Clin. Chem.* **20,** 961–965.
18. Bornstein, P. & Balian, G. (1970) *J. Biol. Chem.* **245,** 4854–4856.
19. Meinwald, Y. C., Stimson, E. R. & Scheraga, H. A. (1986) *Int. J. Pept. Protein Res.* **28,** 79–84.
20. Geiger, T. & Clarke, S. (1987) *J. Biol. Chem.* **262,** 785–794.
21. Yüksel, K. Ü. & Gracy, R. W. (1986) *Arch. Biochem. Biophys.* **248,** 452–459.
22. Takemoto, L. & Boyle, D. (2000) *J. Biol. Chem.* **275,** 26109–26112.
23. Solstad, T. & Flatmark, T. (2000) *Eur. J. Biochem.* **267,** 6302–6310.
24. Sun, A., Yuksel, K. U. & Gracy, R. W. (1995) *Arch. Biochem. Biophys.* **322,** 361–368.
25. Capasso, S. & Salvadori, S. (1999) *J. Pept. Res.* **54,** 377–382.
26. Midelfort, C. F. & Mehler, A. H. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1816–1819.
27. Lai, C. Y., Chen, C. & Horecker, B. L. (1970) *Biochem. Biophys. Res. Commun.* **40,** 461–468.

Deamidation of human proteins

N. E. Robinson*† and A. B. Robinson‡

*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; and ‡Oregon Institute of Science and Medicine, Cave Junction, OR 97523

Deamidation of human proteins

N. E. Robinson*† and A. B. Robinson‡

*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; and †Oregon Institute of Science and Medicine, Cave Junction, OR 97523

Communicated by Frederick Seitz, The Rockefeller University, New York, NY, August 31, 2001 (received for review May 8, 2001)

Deamidation of asparaginyl and glutaminyl residues causes time-dependent changes in charge and conformation of peptides and proteins. Quantitative and experimentally verified predictive calculations of the deamidation rates of 1,371 asparaginyl residues in a representative collection of 126 human proteins have been performed. These rates suggest that deamidation is a biologically relevant phenomenon in a remarkably large percentage of human proteins.

in vivo deamidation | asparaginyl residues

Deamidation of asparaginyl (Asn) and glutaminyl (Gln) residues to produce aspartyl (Asp) and glutamyl (Glu) residues causes structurally and biologically important alterations in peptide and protein structures. At neutral pH, deamidation introduces a negative charge at the reaction site and can also lead to structural isomerization. Early work established that deamidation occurs *in vitro* and *in vivo*, and that the rates of deamidation depend on primary sequence, three-dimensional (3D) structure, pH, temperature, ionic strength, buffer ions, and other solution properties (1–11). It has been hypothesized (3, 5, 7, 12, 13) that Asn and Gln may serve, through deamidation, as molecular clocks which time biological processes such as protein turnover, homeostatic control, and organismic development and aging, as well as mediators of postsynthetic production of new proteins of unique biological value.

Deamidation has been observed and characterized in a wide variety of proteins. It has been shown to regulate some time-dependent biological processes (8, 9) and to correlate with others, such as development and aging. There are many reports of deamidation under physiological conditions in proteins of biological significance. For examples, see refs. 14–18.

Extensive evidence suggests that deamidation of Asn at neutral pH usually proceeds through a cyclic imide reaction mechanism (19–21). Sometimes the Asp produced by deamidation is isomerized to isoAsp. The *in vivo* reversal of this isomerization has been widely reported, but reversal of deamidation itself and of the introduced negative charge has not been observed.

Deamidation rates depend on the amino acid residues near Asn and Gln in the peptide chain with sequence-determined deamidation half-times at neutral pHs and 37°C in the range of 1–500 days for Asn and 100–>5,000 days for Gln (7, 13).

Sequence-determined Asn and Gln deamidation rates are modulated by peptide and protein 3D structures. Deamidation of peptides is observed at both Asn and Gln, largely in accordance with sequence-controlled rates. Deamidation of proteins, which is usually slowed by 3D structure, occurs primarily at Asn except in very long-lived proteins where Gln deamidation is also observed. In a few instances, 3D structure has been reported to increase deamidation rate.

The deamidation rates of individual Asn residues in a protein can be reliably predicted as a result of two recent advances. First, the sequence-controlled Asn deamidation rates of most of the 400 possible near-neighbor combinations in pentapeptide models have been measured (13), and the relevance of this rate library has been established (22). Second, these rates and the 3D structures of proteins with well characterized deamidations have been combined to produce a computation method that correctly predicts the deamidation rates of most Asn residues for which the 3D structure is known (23). This method is more than 95% reliable in predicting relative deamidation rates of Asn residues within a single protein and is also useful for the prediction of absolute deamidation rates.

It is, therefore, now possible to compute the expected deamidation rate of any protein for which the primary and 3D structures are known, except for very long-lived proteins. These proteins require measurement of the 400 Gln pentapeptide rates.

Materials and Methods

Calculation Method. The Brookhaven Protein Data Bank (PDB) was searched to select 126 human proteins of general biochemical interest and of known 3D structure without bias toward any known data about their deamidation, except for 13 proteins (as noted in Table 1) where deamidation has been measured.

The deamidation half-time of each of the 126 proteins was obtained by first computing the deamidation coefficients (C_D) of each Asn and then combining these values into the deamidation index (I_D) by the methods reported (23).

The deamidation coefficient, C_D, is defined as $C_D = (0.01)(t_{1/2})(e^{f(C_m,C_{Sn},S_n)})$, where $t_{1/2}$ is the pentapeptide primary structure half-life (13), C_m is a structure proportionality factor, C_{Sn} is the 3D structure coefficient for the nth structure observation, S_n is that observation, and $f(C_m,C_{Sn},S_n) = C_m[(C_{S_1})(S_1) + (C_{S_2})(S_2) + (C_{S_3})(S_3) - (C_{S_{4,5}})(S_4)/(S_5) + (C_{S_6})(S_6) + (C_{S_7})(S_7) + (C_{S_8})(S_8) + (C_{S_9})(S_9) + (C_{S_{10}})(1 - S_{10}) + (C_{S_{11}})(5 - S_{11}) + (C_{S_{12}})(5 - S_{12})]$. The structure observations, S_n, are those that impede deamidation, including hydrogen bonds, α-helices, β-sheets, and peptide inflexibilities.

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.

Abbreviation: 3D, three-dimensional.

†To whom reprint requests should be addressed. E-mail: noahr@its.caltech.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Computed deamidation half-times for 126 human proteins in pH 7.4, 37°C, 0.15 M Tris·HCl buffer

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 1 (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)	>1000	>1000

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 (23, 26–28) and are designated with an asterisk in the table.

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 N 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 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 if the Asn is 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 are accepted if the bond length is 3.3 Å or less and there is room in the structure to accommodate the van der Waals radius of the hydrogen. All primary structure $t_{1/2}$ values are those published (13), except for Asn with carboxyl-side Pro, Asn, or Gln, and Asn without a free amide as a result of binding to metals or other moieties. Estimated values of $t_{1/2}$ of 500, 40, 60, and 500 days are used for AsnPro, AsnAsn, AsnGln, and bound Asn, respectively.

The coefficients C_m and C_{Sn} were optimized by means of the D_P method (23–25) with an increased set of proteins (26–29). D_P is a measure of the percentage accuracy in classifying the relative deamidation rates of Asn residues in a set of proteins (23). Proteins added to the original set (23) and their Protein Data Bank numbers were bovine DNase I (2DNJ), human hirudin (4HTC), bovine calmodulin (1A29), and human vascular endothelial growth factor (2VPF). Human T cell surface glycoprotein CD4 (1CDJ) was omitted. The Asn 3D environments in all 31 of the calibration proteins were examined and retabulated. These 31 proteins include all of the proteins suitable for this purpose that we have found in the research literature.

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. These values are identical to those found in ref. 23. The deamidation resolving power (D_P) was found to be 95.4%.

The protein deamidation index is defined as $I_D = [\Sigma (C_{Dn})^{-1}]^{-1}$, where C_{Dn} is C_D for the nth Asn residue. Therefore, $(100)(I_D)$ is an estimate of the initial single-residue deamidation half-time for the protein with all Asn residues considered.

Comparison of Calculated Rates with Experimental Rates

The Medline and Citation Index databases were searched for all proteins in which the deamidation rates of identified Asn residues have been reported for 37°C solutions with pH at or near 7.4. Reports were found for a total of 10 individual Asn and 3 combinations of Asn residues in 10 different protein types (17, 18, 23, 29–35). These include 7 proteins that are in the human set of 126 and 3 proteins from other species.

The names, Protein Data Bank numbers, Asn residue position, and computed $(100)(I_D)$ values, respectively, for these proteins are rabbit aldolase: 1ADO, 360, 8.3; human vascular endothelial growth factor: 2VPF, 10, 10; *Escherichia coli* Hpr-phosphocarrier protein: 1HDN, 12–38, 89–22; human fibroblast growth factor: 2AFG, 7, 64; human angiogenin: 1B1I, All, 28; human retinal binding protein: 1BRQ, All, 29; human growth hormone (GH): 1HGU, All, 51; human triose phosphate isomerase: 1HTI, 71, 78; bovine ribonuclease A: 1AFK, 67, 84; and human hemoglobin: 1A3N, with Asn mutants at α50-β80-β82, 47-190-184.

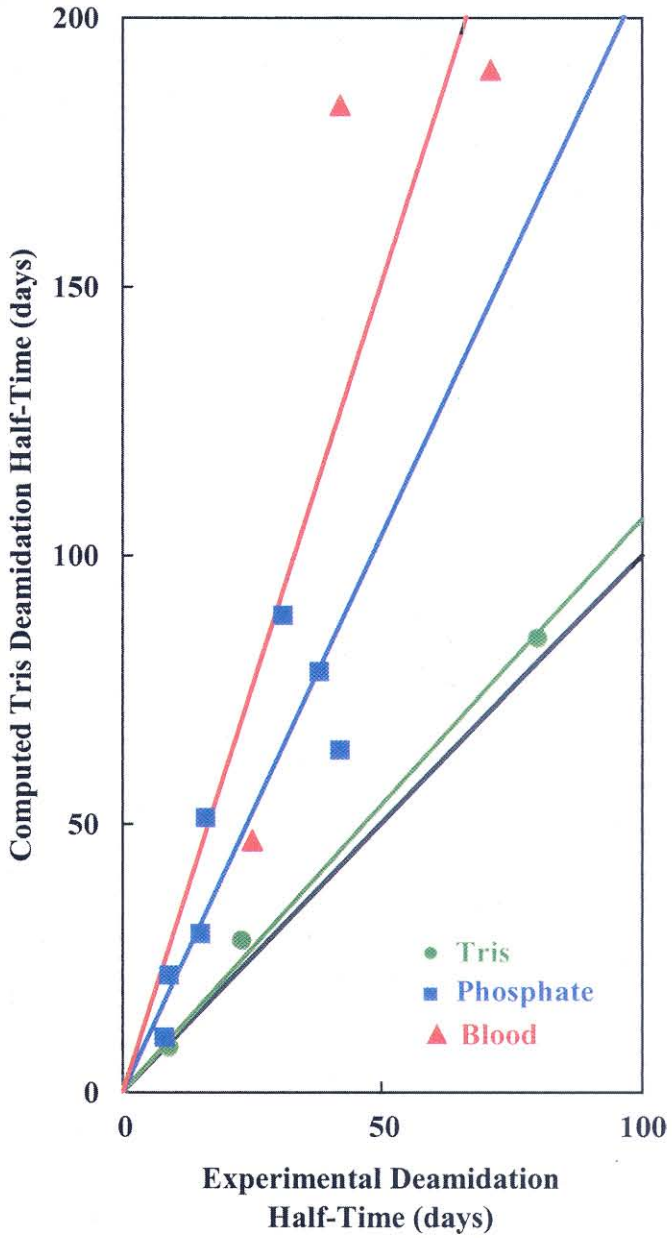

Fig. 1. Calculated single deamidation half-times for 10 individual Asn and 3 combinations of Asn residues in 10 different protein types vs. the corresponding experimentally observed deamidation half-times (17, 18, 23, 29–35). Experiments were *in vitro* in Tris and phosphate buffers and *in vivo* in human blood. Buffer conditions in Tris and phosphate varied among these investigations but were comparable to pH 7.4, 37°C, 0.15 M. Some of the scatter in the figure is probably the result of these variations. If the calculated values and experimental values were identical, the points would lie on the solid black line, as do the values determined in Tris buffers. Catalysis of deamidation is higher by phosphate than by Tris and may be even higher in erythrocytes.

Fig. 1 compares the computed half-times for pH 7.4, 37°C, 0.15 M Tris·HCl buffer with the experimentally observed (17, 18, 29–35) values. The computed values compare favorably with the

experimental values in Tris buffer. In phosphate buffer, the experimental deamidation rates are, on average, 2-fold higher than calculated, and the 3 *in vivo* human blood values average 3-fold higher. This result is entirely as expected because deamidation at neutral pH is subject to catalysis by solution ions. Tris is a very mild catalyst of deamidation. Phosphate is a much stronger catalyst of deamidation in peptides (6, 11) and proteins (17) as compared with Tris. Tissue culture medium contains components even more catalytic of deamidation than phosphate (36). Least-squares lines as shown in Fig. 1 give experimental deamidation rates relative to the computed values in Tris, phosphate, and *in vivo* blood erythrocytes of 1.06, 2.07, and 3.01, respectively.

The agreement between the calculated values and Tris experimental values in Fig. 1 does not arise from computational forcing. The computational method (23) uses experimental sequence-determined pentapeptide deamidation rates in Tris buffer and a parametric 3D structure function with adjustable constants. The optimization method (23–25) for these constants used only the ordered Asn residue instabilities in a wide variety of proteins and buffer types. No experimental absolute deamidation rates were used. The agreement arises because the computational method correctly estimates the relative primary and 3D contributions to the deamidation rate of each Asn, and the primary rates were experimentally determined in Tris.

Results and Discussion

Averaged over all 1,371 Asn, the contributions to the deamidation reaction activation energy from primary and 3D 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 3D structure, which are the same proportions found for a different 24-protein set (23). The cumulative distribution function of the calculated first-order rate constants for deamidation of the 1,371 Asn residues is shown in Fig. 2*a*.

The computed single deamidation half-times in pH 7.4, 37°C, 0.15 M Tris·HCl buffer for the 126 human proteins are shown in Table 1. Table 2 summarizes, on the basis of Table 1, 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 1 and 2. Physiological fluids contain many inorganic, organic, and biochemical substances with deamidase 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 (13, 23) 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 2 result from a set of Asn residues with unusual primary and 3D structures, which comprise about 5% of the total. As illustrated in Fig. 2, 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 Fig. 2 would be expected to be Gaussian. Fig. 2*b* shows the deviation from Gaussian caused by the unusual Asn residues.

Conclusions

The unstable Asn residues that give rise to the deamidation rates shown in Tables 1 and 2 and Figs. 1 and 2 are apparently preferred over the many stable Asn residues that could easily be genetically specified. As shown in Fig. 2, most Asn residues are far more stable. Moreover, even if it were not a result of

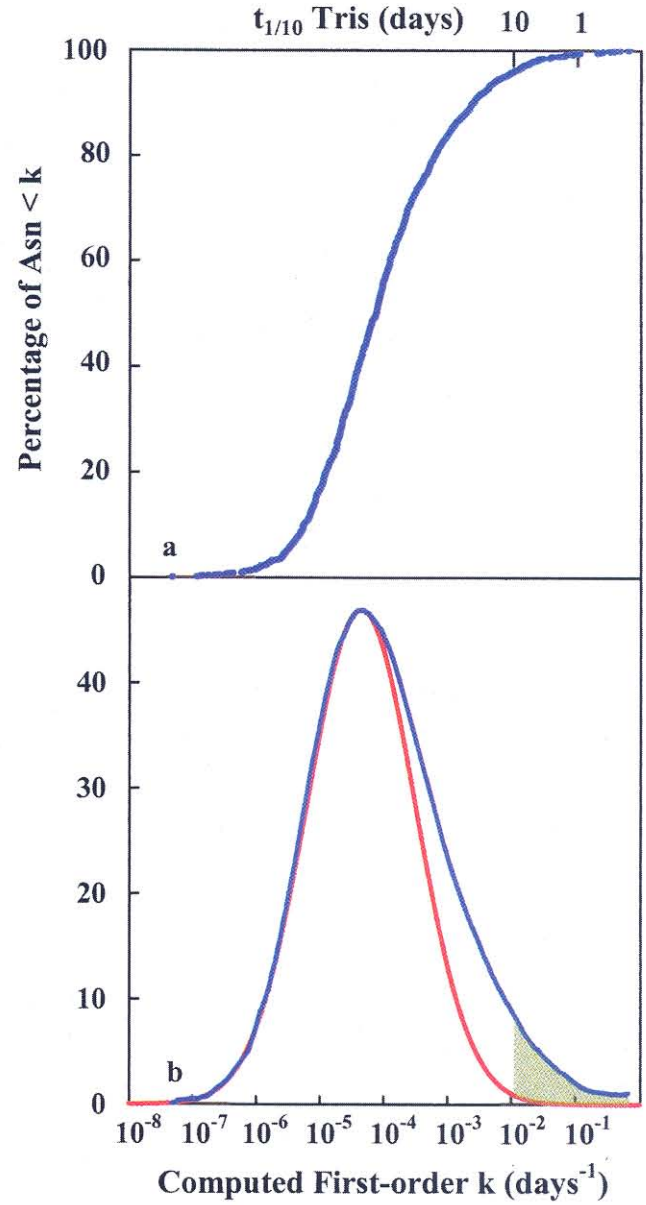

Fig. 2. (*a*) Cumulative distribution function of the calculated first-order rate constants for deamidation of 1,371 Asn residues in 126 human proteins. As indicated, the Asn residues involved in the initial deamidation of these proteins comprise a relatively small part of the complete set. Computed percentages of the Asn residues that are 1/10 deamidated at 1 and 10 days in Tris are 1% and 4%, respectively, as shown. If this deamidation were not of positive biological value, more slowly deamidating sequences and 3D structures could easily have been used. (*b*) Differentiated values of the distribution function in *a* showing the special class of unstable Asn residues present in human proteins. Also shown is a Gaussian function that fits the distribution function except for that part arising from the especially unstable Asn residues. The shaded area contains those Asn residues computed to be one-tenth or more deamidated in 10 days in pH 7.4, 37°C, 0.15 M Tris·HCl. This shaded area for phosphate, physiological fluids, or longer time intervals would be a larger part of the illustrated deviation from Gaussian.

preference, this introduction of negative charges into protein structures would do unacceptable biochemical damage unless it was being used for compensating biological purposes.

Although postsynthetic deamidated proteins are often observed in tissue extracts, their production can be obscured. For example, the *in vivo* steady-state concentrations of the deamidated forms of cytochrome *c* are much lower than expected

Table 2. Percentages of human proteins in Table 1 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	Proteins singly deamidated by >1/10		Proteins singly deamidated by >1/2	
pH 7.4	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)(computed Tris rate) was applied to obtain the phosphate rate of each protein. Percentages produced in physiological solutions may be even higher. Steady-state physiological percentages are lowered by protein turnover.

because they are preferentially degraded (7, 8). Those deamidated forms that are not degraded and, therefore, accumulate in living tissues may have other unique biological purposes. Otherwise, their accumulation would be disadvantageous.

Moreover, the deamidation rates in living tissues are changeable. Through the production of enzymatic deamidases or the control of other physiological parameters that affect the reaction activation energy of deamidation, a living cell could easily increase the overall deamidation rates of its proteins to adapt to changes in physiological circumstances. Decrease of deamidation rates to values below those in Tris summarized in Table 2, however, would be difficult except in specialized structures. Deamidation has been observed in the proteins of many other organisms, too; thus, similar findings may be expected.

In summary, reliable and experimentally verified predictive calculations of the deamidation rates of 1,371 Asn residues in a representative collection of 126 human proteins have been carried out. The results of these calculations show that deamidation of human proteins under physiological conditions is so extensive that it is probably of pervasive and fundamental biological importance. Otherwise, the genetic code would specify stable Asn configurations. Likely uses of deamidation include the timing of biological processes and the postsynthetic production of uniquely useful proteins.

We thank Professor and Mrs. R. B. Merrifield for advice and encouragement, and the Kinsman foundation and other donors to the Oregon Institute of Science and Medicine for financial support. Additional information is available at www.deamidation.org.

1. Flatmark, T. (1964) *Acta Chem. Scand.* **18,** 1656–1666.
2. Flatmark, T. & Sletten, K. (1968) *J. Biol. Chem.* **243,** 1623–1629.
3. Robinson, A. B., McKerrow, J. H. & Cary, P. (1970) *Proc. Natl. Acad. Sci. USA* **66,** 753–757.
4. Robinson, A. B., Scotchler, J. W. & McKerrow, J. H. (1973) *J. Am. Chem. Soc.* **95,** 8156–8159.
5. Robinson, A. B. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 885–888.
6. Scotchler, J. W. & Robinson, A. B. (1974) *Anal. Biochem.* **59,** 319–322.
7. Robinson, A. B. & Rudd, C. (1974) *Curr. Top. Cell. Regul.* **8,** 247–295.
8. Robinson, A. B., McKerrow, J. H. & Legaz, M. (1974) *Int. J. Pept. Protein Res.* **6,** 31–35.
9. McKerrow, J. H. & Robinson, A. B. (1974) *Science* **183,** 85.
10. Robinson, A. B. & Scotchler, J. W. (1974) *Int. J. Pept. Protein Res.* **6,** 279–282.
11. McKerrow, J. H. & Robinson, A. B. (1971) *Anal. Biochem.* **42,** 565–568.
12. Robinson, A. B. & Robinson, L. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8880–8884.
13. Robinson, N. E. & Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 944–949.
14. Solstad, T. & Flatmark, T. (2000) *Eur. J. Biochem.* **267,** 6302–6310.
15. Lindner, H., Sarg, B., Hoertnagl, B. & Helliger, W. (1998) *J. Biol. Chem.* **273,** 13324–13330.
16. Takemoto, L. & Boyle, D. (2000) *J. Biol. Chem.* **275,** 26109–26112.
17. Sun, A., Yuksel, K. U. & Gracy, R. W. (1995) *Arch. Biochem. Biophys.* **322,** 361–368.
18. Capasso, S. & Salvadori, S. (1999) *J. Pept. Res.* **54,** 377–382.
19. Bornstein, P. & Balian, G. (1970) *J. Biol. Chem.* **245,** 4854–4856.
20. Meinwald, Y. C., Stimson, E. R. & Scheraga, H. A. (1986) *Int. J. Pept. Protein Res.* **28,** 79–84.
21. Geiger, T. & Clarke, S. (1987) *J. Biol. Chem.* **262,** 785–794.
22. Robinson, N. E., Robinson, A. B. & Merrifield, R. B. (2001) *J. Pept. Res.* **57,** 483–493.
23. Robinson, N. E. & Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4367–4372.
24. Robinson, A. B. & Westall, F. C. (1974) *J. Orthomolec. Psych.* **3,** 70–79.
25. Robinson, A. B. & Pauling, L. (1974) *Clin. Chem.* **20,** 961–965.
26. Cacia, J., Quan, C. P., Vasser, M., Sliwkowski, M. B. & Frenz, J. (1993) *J. Chromatogr.* **634,** 229–239.
27. Bischoff, R., Lepage, P., Jaquinod, M., Cauet, G., Acker-Klein, M., Clesse, D., Laporte, M., Bayol, A., Dorsselaer, A. V. & Roitsch. C. (1993) *Biochemistry* **32,** 725–734.
28. Potter, S. M., Henzel, W. J. & Aswad, D. W. (1993) *Protein Sci.* **2,** 1648–1663.
29. Goolcharran, C., Cleland, J. L., Keck, R., Jones, A. J. S. & Borchardt, R. T. (2000) *AAPS PharmSci.* **2,** U73–U86.
30. Brennan, T. V., Anderson, J. W., Zongchao, J., Waygood, E. B. & Clarke, S. (1994) *J. Biol. Chem.* **269,** 24586–24595.
31. Volkin, D. B., Verticelli, A. M., Bruner, M. W., Marfia, K. E., Tsai, P. K., Sardana, M. K. & Middaugh, C. R. (1994) *J. Pharm. Sci.* **84,** 7–11.
32. Hallahan, T. W., Shapiro, R., Strydom, D. J. & Vallee, B. L. (1992) *Biochemistry* **31,** 8022–8029.
33. Minic, Z., Hranisavljevic, J. & Vucelic, D. (1997) *Biochem. Mol. Biol. Int.* **41,** 1057–1066.
34. Lewis, U. J., Cheever, E. V. & Hopkins, W. C. (1970) *Biochim. Biophys. Acta* **214,** 498–508.
35. Paleari, R., Paglietti, E., Mosca, A., Mortarino, M., Maccioni, L., Satta, S., Cao, A. & Galanello, R. (1999) *Clin. Chem.* **45,** 21–28.
36. Robinson, A. B. & Scotchler, J. E. (1973) *J. Int. Res. Commun.* **1,** 8.

N.E. Robinson
A.B. Robinson
R.B. Merrifield

Mass spectrometric evaluation of synthetic peptides as primary structure models for peptide and protein deamidation

Authors' affiliations:
N.E. Robinson, Division of Chemistry and Chemical Engineering, California Institute of Technology, USA.

A.B. Robinson, Oregon Institute of Science and Medicine, USA.

R.B. Merrifield, Rockefeller University, USA.

Correspondence to:
Arthur B. Robinson
Oregon Institute of Science and Medicine
2251 Dick George Road
Cave Junction
OR 97523
USA
Tel.: 1-541-592-4142
Fax: 1-541-592-2597
E-mail: art@oism.org

Dates:
Received 5 October 2000
Revised 4 December 2000
Accepted 7 January 2001

To cite this article:
Robinson, N.E., Robinson, A.B. & Merrifield, R.B. Mass spectrometric evaluation of synthetic peptides as primary structure models for peptide and protein deamidation. *J. Peptide Res.*, 2001, **57**, 483–493.

ISSN 1397-002X

Key words: deamidation; mass spectrometry; peptide synthesis; peptides; proteins

Abstract: Solid-phase peptide synthesis and deamidation measurements using a novel mass spectrometric technique were carried out for 94 model asparaginyl peptides from 3 to 13 residues in length. Deamidation rates of these peptides in pH 7.4, 37.0°C, 0.15 M Tris–HCl buffer were measured and evaluated. It was found that they validate the use of pentapeptide models as surrogates for the primary sequence dependence of peptide and protein deamidation rates and the discovery by difference of secondary, tertiary and quaternary structure effects. Deamidation of the pentapeptide models, compared with that of longer peptides of more intricate structure, is discussed, and the application of this technique to deamidation measurement of intact proteins is demonstrated.

Deamidation of amide side-chains of peptides and proteins has been widely observed. This property frequently introduces undesired heterogeneity into peptide and protein preparations. Moreover, it has been hypothesized that glutaminyl and asparaginyl residues in peptides and proteins serve, through spontaneous deamidative transformation into glutamyl and aspartyl residues, as molecular timers of biological events such as protein turnover, development and aging (1–4). The timing of protein turnover by deamidation was first demonstrated by comparison of model peptides with proteins in the cases of cytochrome *c* (5) and aldolase (6). Deamidation of peptide sequences from histones was also reported (7) and discussed in terms of development and aging. Since these initial hypotheses and experiments, much progress has been made. See, for example, recent experiments on deamidation of histone (8),

lens crystallin (9), phenylalanine hydroxylase (10), triosphosphate isomerase (11) and ribonuclease (12).

During the 1970s, the rates of deamidation in 37°C, pH 7.4, ionic strength 0.15–0.2 phosphate buffer of 34 asparaginyl and 30 glutaminyl peptides were measured (3,5–7,13,14). These peptides were primarily of the type GlyXxxAsnYyyGly and GlyXxxGlnYyyGly. Recently, the deamidation rates of an additional 306 asparaginyl pentapeptides of this type were measured (15). It was predicted in 1970 and then demonstrated experimentally (1–5) that deamidation rates depend upon primary structure and upon three-dimensional conformation from secondary, tertiary and quaternary effects. Three-dimensional effects have, to date, largely been inferred by comparison of peptide model deamidation rates with the actual rates of deamidation of other peptides and proteins (see, for example, Ref. 22). There has, however, been no systematic evaluation of the underlying assumption that these pentapeptide models suitably represent most of the primary structure part of the deamidation rates. We have therefore undertaken the synthesis and deamidation measurement of additional model peptides to test and evaluate this assumption.

These 94 peptides were synthesized using solid-phase peptide synthesis (16,17) and the rates of deamidation were measured by a novel method of direct electrospray injection into an ion trap mass spectrometer. The experimental results were consistent with first-order kinetics, so the first-order rate constants and deamidation half-times were calculated assuming first-order dependence upon peptide concentration.

Experimental Procedures

Peptide synthesis

The peptides were synthesized by means of solid-phase peptide synthesis (16,17) in an Advanced Chemtech Model 396 MBS synthesizer with a 96-well reaction block as described previously (15). We used Wang resin substituted with 0.67 meq/g Fmoc-Gly, 0.74 meq/g Fmoc-Ala, or 0.44 meq/g Fmoc-His (Trityl). For peptides blocked on the C-terminal by amidation, 0.4 meq/g Rink Amide MBHA resin was used. For peptides blocked by acetylation of the N-terminal, acetic acid was coupled in the same way as the amino acid derivatives. These N- and C-terminal blocking groups are designated 'Ac' and 'NH_2' in Tables 2 and 8. Derivatives and resin were purchased from Peptides International.

The products were precipitated and washed three times with 15-mL portions of methyl *t*-butyl ether, and vacuum dried. They were then dissolved in purified H_2O, divided into five parts in low-temperature vials, freeze dried, and stored at –80°C. Distilled H_2O was further purified to 18.2 mΩ resistivity in a Labconco 90006-00 ion-exchange purifier before use.

Yields were determined by hydrolysis of portions of the products with 6 N HCl in H_2O at 95°C for 48 h followed by amino acid analysis. This analysis was carried out in a Thermoquest HPLC with a P4000 pump, AS3000 autosampler, and UV6000LP detector combined with a Pickering PCX5200 derivatizer, column, ninhydrin and lithium buffer system. Yields averaged ≈70% with the remaining peptide either not washed from the resin or solubilized during the ether washes. Loss in the ether depended upon the structure of each peptide.

Deamidation of peptides

Peptide deamidation reactions were carried out in pH 7.4, 37.0°C, 0.15 M Tris–HCl buffer. Peptide concentrations were 1.0×10^{-3} M. For the pH dependence experiments with GlySerAsnHisGly, solutions were prepared by adjusting the pH of 0.15 M Tris base with 6 N HCl prior to addition of the peptide. The pH values reported for this experiment are averages of the pH of tubes 1, 9 and 18 of each experiment measured with a calibrated Orion 9803BN micro pH electrode and Orion Model 420A pH meter after all samples had been removed from 37.0°C at the end of the experiment, frozen at –80°C, and then thawed for measurement at 26°C. In the more poorly buffered solutions at pH 3.9, 4.3 and >8, the pH decreased by ≈0.1 units during the deamidation experiments. At the other pH values, the range of pH throughout the experiment was ≈0.02 units. These 26°C pH values were adjusted with Tris buffer tables from 26 to 37°C for all pH values >6, and were used without adjustment for pH values <6.

Deamidation reactions were carried out in sets of 18, 2.0-mL polypropylene centrifuge vials with screw caps and rubber O-rings. Each vial contained 100 µL of solution and was fitted with a hand-made 0.002-in. thick Teflon film liner. This liner was pressed into each tube and a similar liner was placed over the solution and under the vial cap before sealing. The pH experiment was carried out in 1.5-mL polypropylene vials without Teflon liners. Precautions against evaporation were taken by sealing the vials in trays that were hydrated with open vials of water. All 18 samples for all of the experiments were frozen at –80°C until

use and all were placed simultaneously in the 37.0°C reaction chamber. At appropriate timed intervals, sets of these samples were removed from the reaction chamber and frozen at –80°C. All 18 samples from each experiment were analyzed by mass spectrometry on the same occasion and were appropriately alternated to avoid systematic errors from drift during the analysis.

The deamidation reactions took place in an insulated stainless steel incubation chamber controlled at 37.00±0.02°C [15]. Incubation times varied between 4 and 102 days depending upon the individual deamidation rates. These times were adjusted so that each set provided measured points over a range appropriate for the deamidation of the individual peptide.

Titration of GlySerAsnHisGly

A 2×10^{-3} M solution of GlySerAsnHisGly in 0.5 mL H_2O was titrated with 0.1 M NaOH by means of a 250-µL syringe driven by a Harvard Apparatus Model PHD2000 syringe pump. The syringe was connected to a 0.010 in. i.d. Teflon tube, which was withdrawn from the solution during each of the 60 measurements taken between pH 3 and pH 9. The solution was stirred with a Teflon stirring bar while nitrogen flowed into the reaction vial above the solution surface. This nitrogen was introduced through a 0.010 in. i.d. Teflon tube over the stirred solution for 30 min prior to the titration and throughout the titration. Measurements were made with an Orion 9803BN micro pH electrode and Orion Model 420A pH meter. Titration was performed in a 37°C warm room in which the apparatus and solutions were equilibrated prior to the experiment.

Measurement of deamidation by mass spectrometry

Measurements were made in a Thermoquest Model LCQ mass spectrometer fitted with a Thermoquest electrospray source, a Thermoquest AS3500 autosampler and a customized sample delivery system driven by a Harvard Apparatus PHD2000 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. The mass spectrometer was powered by a 240-V system of four Exeltech MX1000 power modules supplied from a battery bank that was charged by two Lorain Flotrol A100F25 rectifiers. The 5-mL syringe contained 1.5% acetic acid in acetone and the 10-mL syringe contained purified H_2O. These were pumped to give a combined flow of 40 µL/min of solution that was 0.5% acetic acid and 33% acetone into the electrospray source. The heated capillary was operated at 180°C. Each run consisted of a 50-µL loop injection, which was monitored continuously for 5–7 min after each injection by a repeating sequence of four high-resolution zoom scans centered on the expected m/z ratio assuming a single charge. One 100–2000 m/z low-resolution scan was also collected for each four zoom scans. Approximately two zoom scans per µL of injected sample were recorded. Seven-minute runs were made only in the case of long or basic peptides that tended to exhibit retarded fractions in the sample delivery system.

Prior to injection, the samples were diluted 500-fold with H_2O by pipetting of 1.5 µL into 749 µL in the injection vials. This permitted direct injection into the electrospray source without fouling of the instrument by nonvolatile constituents. In the case of proteins and larger peptides, H_2O was substituted for acetone in the injection system to avoid precipitation. All solutions throughout these experiments were handled in Teflon or polypropylene containers. With the exception of the pump syringes, no glass containers were used.

Sample injection was into the pumped H_2O stream, which was then mixed with the acetone stream in a U.466 Upchurch PEEK static mixing chamber before entry into the electrospray source. The LCQ was operated in the positive ion mode. Meticulous attention was paid to the materials in contact with the sample, and the hold-up volume and surface area of the sample delivery system was minimized. Teflon and Tefzel were used where available. However, the final delivery capillary was deactivated fused silica, the mixing chamber was PEEK plastic and the sample valve body of the autosampler was also PEEK. Therefore, systematic errors were sometimes introduced by differential adsorption of the reactant and product peptides in the sample delivery system, particularly in the cases of the more basic peptides. Differential adsorption was easily detected because it occasionally introduced slight downward or upward curvatures into the first-order deamidation rate plots, which were accompanied by the expected increases or decreases of total ion current as the deamidation proceeded. Differential ionization in the mass spectrometer would also produce this effect. Most of our first-order rate plots were, however, straight lines, and those that were significantly curved responded to changes in the sample delivery system. In these cases, also, differential composition with time during injection was also observed. Overall mass spectrometer sensitivity drifted somewhat over a period of days, in part because the LCQ mass spectrometer is not fitted with an easily replaced disposable heated capillary.

A typical 4-h mass spectrometry run consisted of 36 samples, with 18 from each of two peptide rate experiments with different masses. These were alternated in order to wash the sample delivery system for each other. Usually, three runs measuring the complete deamidation curves for six peptides were completed each day. We also carried out experiments wherein several peptides of differing mass were measured simultaneously. The throughput of these procedures can be increased 10–20-fold using such combined measurements and even by combined peptide syntheses of appropriately mass selected peptide sets. We carried out such multiple analyses with good results. Precision is, however, compromised as the number of peptides increases. As we wished to keep precision error for these experiments below 1%, we did not combine peptides for the experiments reported here.

Calculation of results

Raw data from all of the sample-containing zoom scans from each LCQ run were summed and recorded in Microsoft EXCEL tables to give results as shown in Fig. 1. The point-by-point values of these curves were then corrected for noise, adjusted for the naturally occurring isotopic ratios, and separated into the contributions from the undeamidated and deamidated peptides. These differ from each other in that deamidation increases the mass by 1 amu. The rates of deamidation were then calculated and plotted as first-order rate curves as shown in Fig. 2. These calculations were performed in MATHCAD 8 PROFESSIONAL with overlays from MACRO EXPRESS 2000 that permit large numbers of calculations to be made sequentially and automatically. Measurement of the 18 samples for the 94 peptides required 1692 loop injections and 200 h of mass spectrometer operation. Total time for analysis was 16 days, although this work could have been completed in 8 days had the mass spectrometer been operated 24 h a day. These measurements have been largely automatic with the autosampler and syringes requiring attention once every 4 h. With the aid of our computer programs, each 4 h of mass spectrometer operation requires ≈ 30 min of manual computer manipulations to calculate the results, largely because our programs are not integrated into the LCQ software.

Deamidation of rabbit muscle aldolase

As a demonstration that this mass spectrometric technique is also suitable for the study of specific deamidations of intact proteins, we carried out an experiment in which the

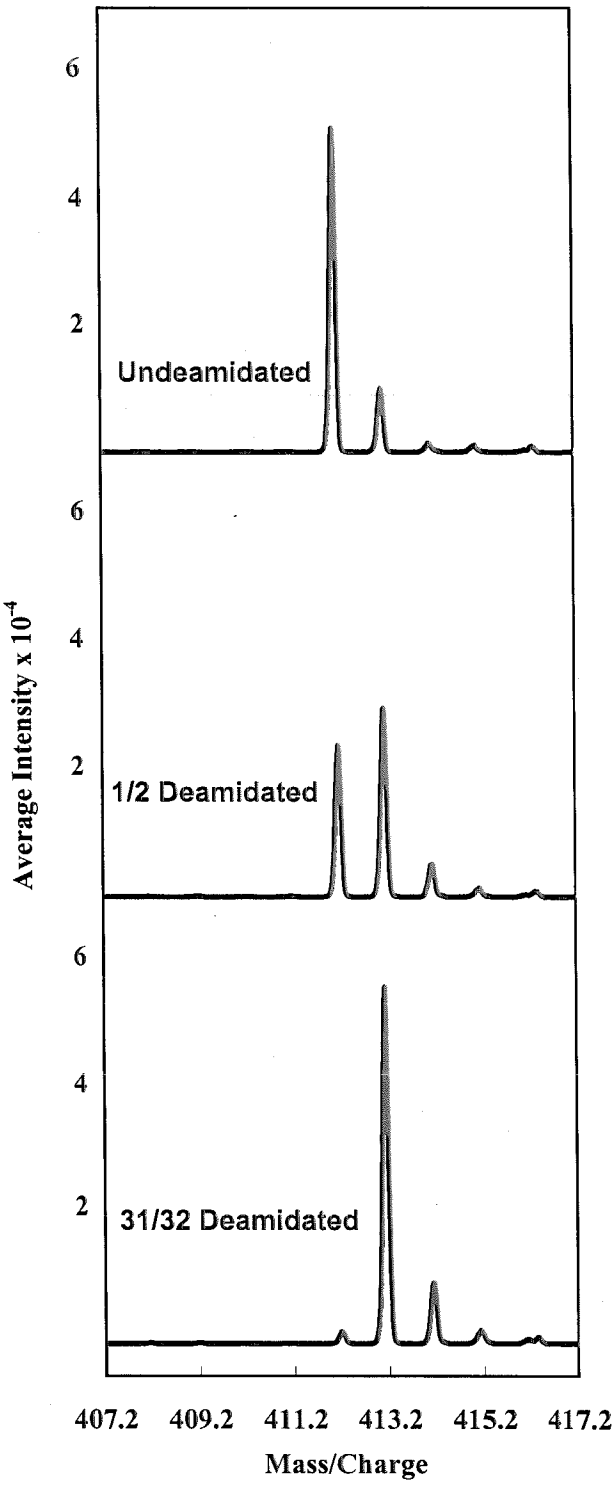

Figure 1. Representative averaged zoom scans from direct 50-μL loop injection of the peptide AsnHisAlaAla into the LCQ mass spectrometer. These are plots of the actual and unsmoothed experimental data as obtained in Excel tables from the LCQ. The graphical base lines have been omitted to show the quality of these results. The individual scans are less smooth, as a result of statistical effects. Scans are from the 1st, 4th and 15th points of the deamidation experiment measured at 0.00, 6.00 and 28.00 days, respectively. Deamidation solutions were 1.0×10^{-3} M peptide, 0.15 M Tris–HCl, pH 7.4, 37.0°C. Injected solution was 2×10^{-6} M peptide and 3×10^{-4} M Tris–HCl buffer.

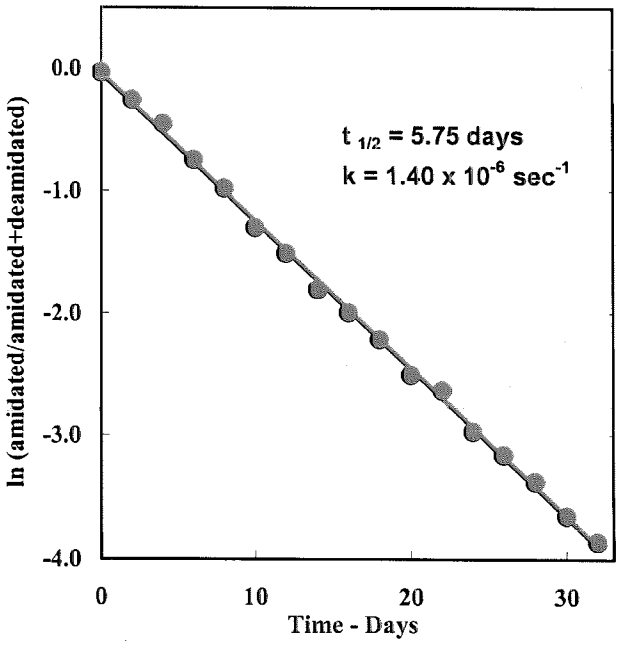

Figure 2. First-order rate plot of the deamidation of the peptide AsnHisAlaAla. Deamidation half-time for the 1.0×10^{-3} M peptide, 0.15 M Tris–HCl, pH 7.4, 37.0°C was calculated to be 5.75 days with a first order rate constant of 1.40×10^{-6} s^{-1}.

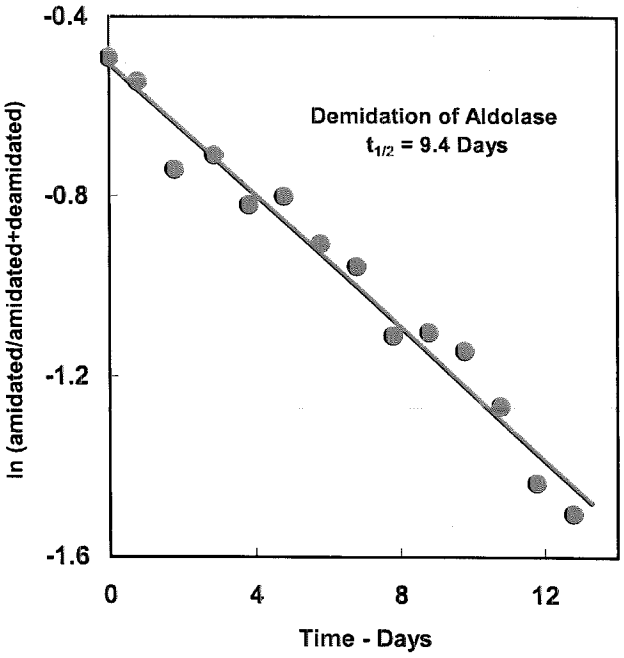

Figure 3. First-order rate plot of the deamidation of rabbit muscle aldolase in the sequence SerAsnHis near the C-terminus of the protein. The deamidation half-time and first-order rate constant for aldolase were found to be 9.4 days and 8.5×10^{-7} s^{-1}, respectively. The solution was 1.0×10^{-3} M in aldolase and 1.0×10^{-3} M in GlySerAsnHisGly and GlyAlaAsnSerGly, 0.15 M Tris–HCl, pH 7.4, 37.0°C. The deamidated solutions were digested with chymotrypsin for 10 min at 25°C immediately before dilution for mass spectrometry.

first two asparaginyl residues near the C-terminus of rabbit muscle aldolase and two model peptides with the same nearest neighbor sequences were exposed to deamidating conditions and measured simultaneously in the same solution [15]. This solution was 1.0×10^{-3} M in peptides and protein, 0.15 M Tris–HCl, pH 7.4, 37.0°C.

Immediately prior to measurement in the mass spectrometer, separate portions of the deamidated solutions were treated with trypsin or chymotrypsin to release the deamidated peptides in aldolase. These peptides and the synthetic peptide models were then scanned simultaneously in the mass spectrometer. The peptides GlySerAsnHisGly and GlyAlaAsnSerGly had deamidation half-times of 8.3 and 11.4 days, respectively. The peptides IleuSerAsnHisAla-Tyr and AlaLeuAlaAsnSerLeuCysGlnGlyLys, which were released by digestion of aldolase after deamidation of the intact protein, had deamidation half-times of 9.4 and >150 days, respectively. The deamidation plot for the fastest deamidation of rabbit muscle aldolase, that of SerAsnHis, as measured in the peptide IleuSerAsnHisAla-Tyr, is shown in Fig. 3.

In accordance with earlier work [6,21], the SerAsnHis sequence deamidates at approximately the same rate in the model peptide and in the protein. Moreover, these rates are in good agreement with the 8 day *in vivo* turnover rate of aldolase. The AlaAsnSer sequence is, however, slowed markedly in aldolase because this sequence occurs in a helical part of the protein [15].

Results

The experimental results are summarized in Fig. 4 and Tables 1–8. We found by titration that the peptide GlySer-AsnHisGly has p*K* values of 3.1, 6.4 and 7.8 for the C-terminal, imidazole and N-terminal groups, respectively. It is to be expected that, for most of the peptides studied herein, except for those with acetyl and/or amide end blocking groups, the C- and N-terminal p*K* values will be in the same general pH range. Therefore, at pH 7.4, the peptides are ionized almost completely at the carboxyl end, but are a mixture of ionized and neutral species at the amino end. Heterogeneity also arises from some of the amino acid residue side-chains.

The precision error of these measurements, as measured by standard deviation of the deamidation half-time, was generally <1% of the reported value. Systematic errors arose, however, in the sample delivery system, which in the case of some peptides discriminately adsorbed different

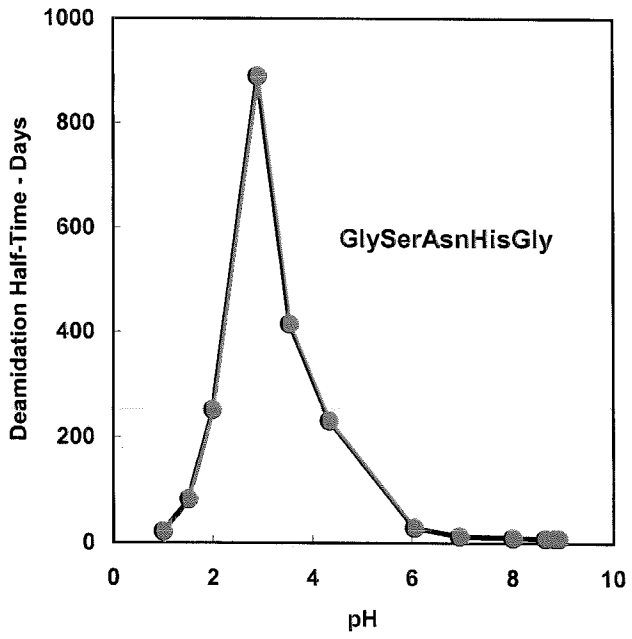

Figure 4. Deamidation half-times of a 1.0×10^{-3} M solution of GlySerAsnHisGly as a function of pH in 37°C, 0.15 M Tris base adjusted to pH with 6 N HCl. The pH values shown are measured averages of the 1st, 9th and 17th points of the deamidation solutions after the deamidation reactions had taken place.

amounts of the amidated and deamidated peptides. This could be seen, in some of the peptides, as differential tailing during the analytical injection accompanied by the expected slight curvature in the first-order rate plots. With 18 high-precision points in each plot, these effects are easily discerned. We therefore estimate that the values reported in Tables 1–8 should be considered reliable to a probable error of ≈ 5%. Many of the values are more accurate than this, and few, if any, are >10% in error. In addition, we are concerned that the very short peptides shown in Table 8 might be especially susceptible to differential ionization, but here we report the values we observed.

The effect upon deamidation rate of pH and of changing degree of protonation is shown in Fig. 4. Deamidation half-time at pH 9, where both the amine and imidazole are neutral, is ≈ 8 days. The half-time increases gradually as protonation of the amine is completed. When, however, the imidazole group is protonated, the half-time increases sharply to a value of ≈ 1000 days. Thereafter, it decreases again to ≈ 12 days at pH 1.0 as the C-terminal is neutralized. It is not known how many different mechanisms of deamidation may be at work throughout this pH range. One mechanism may suffice, or, at low pH another mechanism may predominate. At pH 7.4, however, little general acid or base catalysis by ionized H_2O would be expected. The cyclic imide mechanism (18) or variations of it, as mediated by the peptide structures present, probably predominates at pH 7.4.

Table 1. **_AsnGly_ deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
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

Table 2. **_AsnHis_ deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
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

With these considerations in mind, especially as mixtures of species may cause small variations in the measured rates, consider the deamidation rates of the 94 model peptides summarized in Tables 1–8.

It has been found (15), in deamidation measurements of 306 model peptides of the type GlyXxxAsnYyyGly in pH 7.4, 37.0°C, 0.15 M Tris–HCl buffer, that the median deamidation half-times in days for carboxyl side nearest neighbors (Yyy), in sets comprised of 18 of the ordinary amino acid residues as amino side nearest neighbors (Xxx),

Table 3. **_XxxAsnYyy_ deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
GlyIleAsnIleGly	384	0.0209
GlyAlaAsnIleGly	300	0.0267
GlyAlaAsnGluGly	74.1	0.108
GlyAlaAsnTyrGly	73.9	0.109
GlyTyrAsnTyrGly	70.6	0.114
GlyArgAsnArgGly	67.4	0.119
GlyAlaAsnArgGly	62.4	0.129
GlyGluAsnGluGly	60.3	0.133
GlyAlaAsnLysGly	55.9	0.144
GlyLysAsnLysGly	53.5	0.150
GlyProAsnAlaGly	31.8	0.252
AlaAlaAsnAlaAla	29.9	0.268
AlaAlaSerAsnAlaAla	27.8	0.289
GlyIleAsnAlaGly	25.9	0.310
GlyGluAsnAlaGly	25.8	0.311
GlyHisAsnAlaGly	24.6	0.326
GlyArgAsnAlaGly	24.4	0.329
GlyTyrAsnAlaGly	24.3	0.330
GlySerAsnAlaGly	24.1	0.333
GlyLysAsnAlaGly	23.6	0.340
GlyAlaAsnAlaGly	22.5	0.357
GlySerAsnSerGly	15.1	0.531
GlyAlaAsnSerGly	14.9	0.538
GlyHisAsnHisGly	10.7	0.750
GlyAlaAsnHisGly	9.3	0.863
GlyGlyAsnHisGly	9.2	0.872
GlySerAsnGlyGly	0.96	8.36

Table 4. **AlaXxxAlaAsnAlaYyyAla deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
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

are Gly 1.14, His 10.2, Ser 15.7, Ala 25, Asp 32, Thr 46, Cys 52, Lys 58, Met 60, Arg 62, Glu 64, Phe 70, Tyr 79, Trp 97, Leu 119, Val 243, and Ileu 302. Amino side nearest neighbor effects are (15), in decreasing order of deamidation rate, Gly, Ser, Thr, Cys, Met, Phe, Tyr, Asp, Glu, His, Lys, Arg, Ala, Leu, Val, Ileu, Trp and Pro. The effect upon deamidation of the carboxyl side nearest neighbor is much greater than that of the amino side nearest neighbor. Table 3 includes a few of these pentapeptide rates that are especially relevant to the longer peptide models reported herein.

Several effects are evident in the model peptide deamidation rates.

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 five-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 five-membered imide. As soon as one or more residues intervene, Lys begins to catalyze deamidation, the half-times moving from $\approx$ 50 days, to a range of 14–19 days. Similar effects are seen for Arg and other residues as summarized in Tables 4–7.

2. In the case of Gly as the carboxyl side nearest neighbor, the 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 shown in Table 1. The peptide elongation effect is shown also in Table 2, where the SerAsnHis central region was 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,

Table 5. **AlaXxxAlaAlaAsnAlaAlaYyyAla deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
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 6. **AlaXxxAlaAlaAlaAsnAlaAlaAlaYyyAla deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
AlaSerAlaAlaAlaAsnAlaAlaAlaSerAla	1490	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. **AlaXxxAlaAlaAlaAlaAsnAlaAlaAlaAlaYyyAla deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
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

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 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 2 in which both ends of the peptides have been blocked by Ac and NH_2 or have been elongated by one residue, the Ac and NH_2 deamidation rates are slower than are the single residue additions.

3. 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 4–7 to be: for 7, Arg, His, Tyr, Lys, SerHis, Ser, Ileu, Ala, Glu, Pro; for 9, Lys, SerHis, His, Ser, Ala, Glu, Ileu; for 11, Lys, Glu, His, Ileu, SerHis, Ser; and for 13, Lys, His, Glu, Ileu, Ser and SerHis.

4. There is a sharp reversal of order and slowing of deamidation by one to two orders of magnitude in the

Table 8. **_XxxAsnYyy peptide deamidation rates**

Peptide	$t_{1/2}$ (days)	$k\times10^6$ (sec^{-1})
$AlaAlaAlaAlaNH_2$	n.o.	n.o.
AlaAlaSerAsn	n.o.	n.o.
GlyAsnGly	98	0.082
AlaAsnAla	270	0.030
$AlaAlaSerAsnNH_2$	37.6	0.213
SerAsnHis	39.6	0.203

n.o., none observed.

Ser...Ser, Ser...His and Ileu...Ileu 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 α-helix is also evident in rabbit muscle aldolase. To a lesser degree, 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.

5. The short peptides in Table 8 illustrate the necessity of a carboxyl side nitrogen for peptide deamidation at neutral pH. This is in accord with the proposed imide mechanism (18).

6. It is additionally striking, in Tables 1, 2 and 4–7, 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 rate. Rules arising from these regularities are the subject of a later report. With, however, the results reported herein and in Ref. 15, the known deamidation rates of model peptides have now been increased by 6-fold, so reference to the models themselves will suffice, in most cases, for semi-quantitative predictions and evaluations.

Reaction mechanisms

While statements about reaction mechanism herein are, to some extent, inherently speculative, it would be odd if, with so much new reaction rate data at our disposal, we made no mention of its relevance to the understanding of mechanism.

First, throughout these experiments and in the CPK model building with which we accompanied them, we had most prominently in mind the five-membered cyclic imide mechanism originally proposed to explain the unusually rapid deamidation of peptides with carboxyl side glycyl

residues as nearest neighbors to asparaginyl residues (18, 23). This mechanism suffices to explain most of the carboxyl side nearest neighbor deamidation rates we measured. Sterically facilitated by Gly, catalyzed by His, Ser, Asp, Thr and Cys, and sterically impeded by Trp, Leu, Val and Ileu, this mechanism seems to be robust. All of the 19 other amino acid residues are sterically hindered compared with Gly in the formation of a five-membered imide. The deamidation rates of carboxyl side Ser, Thr and Cys peptides are in the correct order compared with valence bond estimates of their dipole moments, as would be predicted by this mechanism, with the effect of Thr diminished by steric hindrance. Moreover, our pH dependence experiments of GlySerAsnHisGly are in accord with this mechanism, except in the low pH region where general acid catalysis may serve.

We are, however, less satisfied by the five-membered ring explanation as it relates to the longer peptides. While it may be that all of the deamidation rates reported in Tables 1, 2 and 4–7 can be explained by this mechanism, something additional may be at work. For example, we would not expect the deamidation-enhancing effect of Lys to remain so substantial as it moves farther and farther away from the postulated five-membered imide. Certainly, the Lys can easily bend back upon it, but why does this effect not change with distance?

One possibility is that larger rings may be involved in the deamidation of longer peptides. Once an active side-chain is relieved of the steric hindrance of a nearest neighbor position to the amide, it is also free to participate in other ring structures that could enhance deamidation. These possibilities are easily susceptible to experimental tests through specialized model peptides, therefore we plan to measure the deamidation rates of some appropriate peptides with the purpose of distinguishing between them.

Summary Explanations

Deamidations of amide residues clearly serve, as originally proposed primarily on theoretical grounds (1–4), as molecular clocks that time biological processes. A few examples of this have been demonstrated. As yet it is not known how widespread the use of these clocks may be and to what extent they control fundamental biological processes.

The exploration of this effect requires convenient tools for the measurement of deamidation and a substantial number of peptide models to aid in its understanding in the complex environments of naturally occurring peptides, proteins and organelles. We have therefore developed a mass spectrometric method that is suitable for the measurement of deamidation rates in peptides and proteins. In combination with solid-phase peptide synthesis, we have demonstrated this method by measuring the deamidation rates of a series of model peptides.

The pH, ionic strength and temperature chosen for these experiments are similar to those relevant to many biological systems and also to the conditions utilized in the measurement of the asparaginyl peptides reported previously, primarily from our laboratories (3, 15).

We have used Tris–HCl buffer because earlier experiments indicated that this buffer has a lesser effect upon deamidation rates than does phosphate buffer and other constituents frequently used in *in vitro* and *in vivo* studies (3,19,20). It is expected that this will make these deamidation rates more useful as a basis upon which to build a parametric system of deamidation predictors.

We have used the LCQ mass spectrometer in the direct injection mode rather than the usual HPLC-mass spectrometry mode. Problems with buffer ion precipitation in the electrospray source were avoided primarily by sample dilution. While the quantitative dynamic range of this mass spectrometer is relatively narrow, we obtained good results with sample concentrations in the 2×10^{-6} to 1×10^{-7} M range. That this method is convenient and efficient is illustrated by the fact that, after the period of initial development, two scientists (NER and ABR) working for 4 months, without technicians or other support and with additional responsibilities, completed >9000 measurements and calculations determining the rates of deamidation of ≈ 500 peptides.

Direct injection mass spectrometry is of special value in this work for reasons other than efficiency. It is absolutely essential that the proportions of amidated and deamidated peptides should be maintained throughout the experimental measurements. While rate measurements can, in principle, be made by means of either one of the species individually, such measurements are generally much less precise than comparative measurements. However, because the reactant and product peptides have a different charge, they tend to be adsorbed to differing extents by surface interactions. These interactions are a serious problem in the reaction vessels and even in the sample delivery of the peptides to the mass spectrometer during loop injections. If an additional step such as chromatography were utilized, they would be magnified severely. Moreover, without careful work, this sort of systematic error can pass unrecognized. With 18 high-precision points in each experiment, we were able to

guard against this. Fewer points with less precision would render this effect unobservable and the rates of deamidation so measured could be seriously in error.

As a general rule, Teflon surfaces of the smallest possible surface area have proved to be best for this work. The acetone–acetic acid system gives excellent sensitivity. In the case of some of the larger peptides, and with proteins, H_2O must be substituted for acetone to avoid precipitation. This disadvantage is offset, somewhat, because larger peptides tend to produce higher signals in the mass spectrometer.

The LCQ spectrometer has proved excellent for this work with the exception of one serious drawback. The manufacturer changed the heated capillary through which the sample is introduced from an easily replaceable disposable capillary, as originally designed, into an expensive and inconvenient unit. This prevents routine replacement of the capillary on a daily basis. In the experiments reported herein, we made use of a peculiar advantage of Tris–HCl buffer. At the concentrations employed, this buffer is volatile and moves through the electrospray system and into the mass spectrometer as chains of various lengths. These chains are present in large amount and sometimes occur at m/z ratios identical to the peptides of interest. Fortunately, however, they are decomposed, apparently by collisions with helium in the ion trap, before the high-resolution zoom scan, so Tris–HCl buffer ions have essentially no effect on the deamidation measurements. When peptides in nonvolatile buffers or proteins are injected, however, the heated capillary rapidly degrades. Cleaning only partially restores the capillary, and the sensitivity and reliability of the instrument gradually diminishes. For large numbers of precise quantitative experiments with good reliability in a useful variety of solutions, the mass spectrometer should be fitted with a disposable sample introduction system.

Another potentially confounding problem is formation of salts of sodium and potassium during the latter stages of ionization. For this reason, all of our experiments were performed in polypropylene or Teflon containers. Glass was avoided in this study, except for a short length of deactivated fused silica tubing within the electrospray needle. Fused silica was originally used in the sample loop, but results were improved by using a Teflon sample loop, which markedly reduced adsorptive tailing during sample introduction. Also, highly purified H_2O was used throughout. Regardless of these precautions, however, some sodium and potassium is usually present. If significant portions of the peptides are ionized as salts, precision is reduced. We found, however, that operation of the heated capillary at a relatively low temperature of 180°C minimizes the formation of sodium and potassium salts acceptably.

Herein and elsewhere [15], we report the deamidation rates of 400 model peptides of differing sequences and, in some cases, in differing solvent conditions. All of these are asparaginyl peptides and include all possible nearest neighbor configurations with the exceptions of half of the Pro sequences and those containing two or more amides. We synthesized these other peptides and plan to report their deamidation rates in future publications. We also synthesized the complete set of 400 nearest neighbor Gln sequences and have begun deamidation experiments on that set. Because Gln residues tend to have longer deamidation times than Asn residues, many of these peptides need to remain at 37°C for 6–12 months before high precision deamidation rate experiments are completed.

Conclusions

The rates of deamidation of model peptides are subject to exquisite control by primary and three-dimensional structure that has not been heretofore demonstrated. We have demonstrated some aspects of this control in a series of 94 model peptides between 3 and 13 residues in length. This was performed by a combination of automated solid-phase peptide synthesis with a novel method of ion trap mass spectrometry, which is convenient, efficient and versatile.

The measured deamidation rates extend the available framework of understanding of primary structure determined deamidation rates, so that it is more useful in experiments to understand the relative effects of primary, secondary, tertiary and quaternary structure on the deamidation of peptides and proteins.

Also, these quantitative determinations of primary and secondary effects of elongated peptide chains and side-chains several residues removed from the deamidating amide, place in perspective the effects of nearest neighbor interactions in determining deamidation rate. The overall conclusion is that the nearest neighbor interactions play an especially large part in determining the deamidation rate of model peptides, so the relevance of the recently completed [15] rate table for peptides of the type GlyXxxAsnYyyGly is confirmed.

Moreover, because deamidation-mediating effects of moieties farther removed from the nearest neighbors require additional peptide chain flexibility, a limitation that generally retards deamidation in intact proteins, nearest

neighbor interactions are probably of even greater relative importance than is illustrated in these experiments.

Laboratories having need of peptides of any of the 900 amide sequences that have been synthesized to date in the course of this work (Ref. 15 and this study), especially for the purpose of comparative work on protein deamidation, are welcome to contact the corresponding author at the address given on p. 483. All of these synthetic peptides, all 9000 deamidation rate solutions, and all 9000 dilution solutions for mass spectrometry are currently stored at –80°C and are available for further experiments.

Acknowledgements: We thank Professor Brian Chait for his advice about mass spectrometry and the John Kinsman Foundation and other donors to the Oregon Institute of Science and Medicine for financial support.

References

1. Robinson, A.B., McKerrow, J.H. & Cary, P. (1970) Controlled deamidation of peptides and proteins: an experimental hazard and a possible biological timer. *Proc. Natl Acad. Sci. USA* **66**, 753–757.
2. Robinson, A.B. (1974) Evolution and the distribution of glutaminyl and asparaginyl residues in proteins. *Proc. Natl Acad. Sci. USA* **71**, 885–888.
3. Robinson, A.B. & Rudd, C. (1974) Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. *Curr. Topics Cell. Reg.* **8**, 247–295.
4. Robinson, A.B. (1978) Molecular clocks, molecular profiles, and optimum diets: three approaches to the problem of aging. *Mech. Ageing Dev.* **9**, 225–236.
5. Robinson, A.B., McKerrow, J.H. & Legaz, M. (1974) Sequence dependent deamidation rates for model peptides of cytochrome C. *Int. J. Peptide Protein Res.* **6**, 31–35.
6. McKerrow, J.H. & Robinson, A.B. (1974) Primary sequence dependence of the deamidation of rabbit muscle aldolase. *Science* **183**, 85.
7. Robinson, A.B. & Scotchler, J.W. (1974) Sequence dependent deamidation rates for model peptides of histone IV. *Int. J. Peptide Protein Res.* **6**, 279–282.
8. Lindner, H., Sarg, B., Hoertnagl, B. & Helliger, W. (1999) The Microheterogeneity of the mammalian H1(0) histone. Evidence for an age-dependent deamidation. *J. Cancer Res. Clin. Oncol.* **125**, 182–186.
9. Takemoto, L. & Boyle, D. (1998) Deamidation of specific glutamine residues from alpha-A crystallin during aging of the human lens. *Biochemistry* **37**, 13681–13685.
10. Solstad, T. & Flatmark, T. (2000) Microheterogeneity of recombinant human phenylalanine hydroxylase as a result of nonenzymatic deamidations of labile amide containing amino acids: effects on catalytic and stability properties. *Eur. J. Biochem.* **267**, 6302–6310.
11. Sun, A., Yuksel, K.U. & Gracy, R.W. (1995) Terminal marking of triosphosphate isomerase: consequences of deamidation. *Arch. Biochem. Biophysics* **322**, 361–368.
12. Capasso, S. & Salvadori, S. (1999) Effect of the three-dimensional structure on the deamidation reaction of ribonuclease A. *J. Peptide Res.* **54**, 377–382.
13. Robinson, A.B., Scotchler, J.W. & McKerrow, J.H. (1973) Rates of nonenzymatic deamidation of glutaminyl and asparaginyl residues in pentapeptides. *J. Am. Chem. Soc.* **95**, 8156–8159.
14. Robinson, A.B. & Tedro, S. (1973) Sequence dependent deamidation rates for model peptides of hen egg white lysozyme. *Int. J. Peptide Protein Res.* **5**, 275–278.
15. Robinson, N.E. & Robinson, A.B. (2001) Molecular clocks. *Proc. Natl Acad. Sci. USA* **98**, 944–949.
16. Merrifield, R.B. (1963) Solid phase peptide synthesis I. Synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154.
17. Merrifield, R.B. (1995) Solid phase peptide synthesis. In *Peptides: Synthesis, Structures, and Applications* (B. Gutte, ed.). Academic Press, New York, pp. 94–169.
18. Geiger, T. & Clarke, S. (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. *J. Biol. Chem.* **262**, 785.
19. Scotchler, J.W. & Robinson, A.B. (1974) Deamidation of glutaminyl residues: dependence on pH, temperature, and ionic strength. *Anal. Biochem.* **59**, 319–322.
20. Robinson, A.B. & Scotchler, J.W. (1973) Deamidation of glutaminyl residues in tissue culture media. *J. Int. Res. Comm.* **1–8**, 26.
21. Midelfort, C.F. & Mehler, A.H. (1972) Deamidation *in vivo* of an asparagine residue of rabbit muscle aldolase. *Proc. Natl Acad. Sci. USA* **69**, 1816–1819.
22. Robinson, N.E. & Robinson, A.B. (2001) Prediction of protein deamidation rates from primary and three-dimensional structure. *Proc. Natl Acad. Sci. USA*, in press.
23. Meinwald, Y.C., Stimson, E.R. & Scheraga, H.A. (1986) Deamidation of the asparginyl–glycyl sequence. *Int. J. Peptide Protein Res.* **28**, 79–84.

Molecular clocks

Noah E. Robinson* and Arthur B. Robinson†‡

*Department of Chemistry, California Institute of Technology, Pasadena, CA 91125; and †Oregon Institute of Science and Medicine, 2251 Dick George Road, Cave Junction, OR 97523

Communicated by Bruce Merrifield, The Rockefeller University, New York, NY, November 28, 2000 (received for review September 26, 2000)

Molecular clocks

Noah E. Robinson* and Arthur B. Robinson†‡

*Department of Chemistry, California Institute of Technology, Pasadena, CA 91125; and †Oregon Institute of Science and Medicine, 2251 Dick George Road, Cave Junction, OR 97523

Communicated by Bruce Merrifield, The Rockefeller University, New York, NY, November 28, 2000 (received for review September 26, 2000)

A convenient and precise mass spectrometric method for measurement of the deamidation rates of glutaminyl and asparaginyl residues in peptides and proteins has been developed; the rates of deamidation of 306 asparaginyl sequences in model peptides at pH 7.4, 37.0°C, 0.15 M Tris·HCl buffer have been determined; a library of 913 amide-containing peptides for use by other investigators in similar studies has been established; and, by means of simultaneous deamidation rate measurements of rabbit muscle aldolase and appropriate model peptides in the same solutions, the use of this method for quantitative measurement of the relative effects of primary, secondary, tertiary, and quaternary protein structure on deamidation rates has been demonstrated. The measured rates are discussed with respect to the hypothesis that glutaminyl and asparaginyl residues serve, through deamidation, as molecular timers of biological events.

deamidation | biological clocks | peptides | mass spectrometry

The hypothesis that glutaminyl and asparaginyl residues in peptides and proteins serve, through deamidative transformation into glutamyl and aspartyl residues, as molecular timers of biological events such as protein turnover, development, and aging (1–4) was originally based on the suggestion and then experimental proof that the deamidation of cytochrome *c* occurs *in vivo* (5–6) and on reasoning that deamidation is seriously disruptive to biological systems unless it is being used for compensating beneficial biological purposes.

Subsequently, it was shown that the first-order deamidation half times in pH 7.4, 37°C ionic strength 0.15–0.20 phosphate buffer of glutaminyl and asparaginyl residues vary over a range of at least 1 day to 9 years as a function of primary sequence and likely over an even wider range as a function of secondary, tertiary, and quaternary structure (3, 7–11). The dependence of deamidation on pH, temperature, ionic strength, and other solution properties was also demonstrated (3, 12–13). It was additionally shown that the overall *in vivo* compositions and specific sequence distributions of amide residues in peptides and proteins are supportive of the amide molecular clock hypothesis (1–3, 14).

The first two specific amide clocks to be identified were those that control the *in vivo* turnover rates of cytochrome *c* (6, 9) and rabbit muscle aldolase (15–16). Measurements of *in vivo* turnover rates, *in vivo* steady-state concentrations, *in vivo* and *in vitro* deamidation rates of these proteins, and *in vitro* deamidation rates of appropriate model peptides demonstrated both sequence dependence and three-dimensional structure dependence of the deamidation of these two proteins. Sequence-controlled deamidation of the C-terminal sequence . . . Thr-Asn-Glu in cytochrome *c* apparently leads to a changed three-dimensional structure that accelerates a second deamidation of cytochrome *c*. The resulting deamidated forms of cytochrome *c* are rapidly catabolized *in vivo* with rates equivalent to the turnover rate. In the case of rabbit muscle aldolase, sequence-controlled deamidation of the C-terminal sequence . . . Ileu-Ser-Asn-His-Ala-Tyr and *in vivo* turnover proceed at the same rates, the former apparently timing the latter. In both proteins, comparisons to model peptides also showed general suppression of deamidation of other amides in the proteins by means of three-dimensional structure effects. Deamidation of histone peptides was also reported and discussed in terms of development and aging (10). Since this initial work, much progress has been made. See, for example, recent experiments on deamidation of phenylalanine hydroxylase (17), histone (18), crystallin (19), trisphosphate isomerase (20), and ribonuclease (21).

It has been proposed that the deamidation of asparaginyl peptides at neutral pHs proceeds by way of a cyclic five-membered imide formed from the asparaginyl α carbon and side chain and the peptide bond nitrogen of the next residue toward the carboxyl end of the peptide (11, 22). This mechanism explains the production of both aspartyl and isoaspartyl peptides during deamidation, which has been widely observed.

The general approach that was first taken toward understanding the effects of primary, secondary, tertiary, and quaternary peptide and protein structure on deamidation was that of measuring the effects of primary structure in model peptides and then inferring the effects of secondary, tertiary, and quaternary structure by comparison with the deamidation rates of peptides and proteins of interest (2–4, 7–10, 16). Therefore, the deamidation rates of about 60 model peptides and several proteins were determined by means of the labor-intensive techniques available in the 1970s. Although a substantial amount of experimental data, which is beyond the scope of this discussion, has subsequently accumulated concerning deamidation, and new techniques have been used, progress has been impeded by lack of knowledge of the complete library of sequence-controlled deamidation rates and by lack of a fast and accurate means of determining the rates of deamidation of each amide in a protein, with reference to the model peptide rates.

We have, therefore, synthesized, by means of Merrifield solid-phase peptide synthesis (23–24), the entire library of 800 possible nearest-neighbor amide sequences in pentapeptides, which involve the 20 ordinary amino acid residues, and an additional 113 peptides of interest to the effects of residues between 2 and 6 positions distant from the amides in up to 13 residue peptides.

We have devised a means of measuring the rates of deamidation of these peptides by direct loop injections into an ion-trap mass spectrometer. This method permits the determination of an 18-point deamidation curve with an inherent deamidation precision error of less than 1% by means of 18 loop injections requiring about 2 h total mass spectrometer time. Although we have also shown that measuring several peptides of differing molecular weights simultaneously can markedly accelerate these measurements, we prefer to measure the peptides individually except where simultaneous multiple measurements have special value as in work on proteins.

Moreover, we have demonstrated that this method can be used to measure the deamidation rates of specific amides within a protein, while the rates of deamidation of the peptides to be compared with that protein are also simultaneously being measured within the same solution. The model peptides are included

Abbreviations: DMF, dimethylformamide; TFA, trifluoroacetic acid; Fmoc, 9-fluorenylmethoxycarbonyl.

‡To whom reprint requests should be addressed. E-mail: art@oism.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

with the protein during the rate experiment, and the reaction mixture is digested with an appropriate proteolytic enzyme before direct injection into the mass spectrometer. The mass regions of all of the amide-containing peptides are then simultaneously scanned.

We herein report the deamidation rates of 306 asparaginyl pentapeptides. We also report the deamidation rates of the first two asparaginyl residues near the C terminus of rabbit muscle aldolase along with the rates of simultaneously measured analogous model peptides, which allow the determination of the effects of primary and secondary structure on these aldolase deamidations. All of these rates of deamidation were determined at pH 7.4 in 0.15 M Tris·HCl buffer at 37.0°C with peptide and protein concentrations of 1.0×10^{-3} M.

Materials and Methods

Synthesis of Peptides. The peptides were synthesized by means of Merrifield solid-phase peptide synthesis (23, 24) in an Advanced ChemTech Model 396 MBS synthesizer with a 96-well reaction block. Derivatives used 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-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. These were all purchased from Peptides International.

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 h. 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 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, freeze dried, and stored at −80°C.

The acetamidomethyl-blocking groups of cystine were approximately 20% removed in TFA. Deamidation rates of both the blocked and unblocked peptides were measured by mass spectrometry. Deamidation rates for the blocked peptides will be reported elsewhere.

Yields were determined by hydrolysis with 6 N HCl in H_2O at 95°C for 48 h 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. In most cases, impurities in the peptides were so minimal as to be indistinguishable from background in the mass spectrometer.

Deamidation of Peptides. Peptide deamidation reactions were carried out in pH 7.4, 37.0°C, 0.15 M Tris·HCl buffer. Peptide concentrations were 1.0×10^{-3} M. These reactions were carried out in 1.5-ml polypropylene centrifuge vials with screw caps and rubber O-rings. The vials were also sealed in trays in a hydrated environment to prevent evaporation. 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. Each vial contained 100 μl of solution.

The deamidation reactions took place in a specially constructed incubation chamber controlled at 37.00 ± 0.02°C throughout the 3-month period of these experiments. This chamber was continuously monitored by means of three thermisters and one glass thermometer; one of the thermisters and the thermometer were calibrated to ASTM standards.

Incubation times varied between 4 and 102 days. These were adjusted to provide measured points over a range of at least one-fourth and at most six deamidation half times for each individual peptide or protein. At timed intervals, vials were removed from the incubator and frozen at −80°C. All portions for an individual rate were measured in a single set of mass spectrometry runs with the analyses alternated to eliminate drift errors during the course of the mass spectrometry.

In the case of aldolase, ICN rabbit muscle aldolase precipitated from 6 M ammonium sulfate was centrifuged, dissolved in H_2O, and diluted to 1.0×10^{-3} M. The solution also contained the peptides Gly-Ser-Asn-His-Gly and Gly-Ala-Asn-Ser-Gly at 1.0×10^{-3} M and was 0.015 M in pH 7.4 Tris·HCl buffer. This solution was divided into 50-μl portions in polypropylene vials, placed at 37.0°C, removed from the incubator at 14 1-day intervals, and frozen at −80°C. Before mass spectrometric analysis, 3-μl portions of the reaction mixtures were each mixed with 16 μl of a 0.5 mg/ml solution of Sigma T-1426 TPCK treated trypsin or ICN 100478 α-chymotrypsin respectively. The trypsin mixtures were incubated at 37.0°C for 2 hours, and the chymotrypsin mixtures were incubated at 25°C for 10 min before dilution for mass spectrometry. The tryptic and chymotryptic digests were injected separately into the mass spectrometer, so 28 loop injections were performed. Trypsin produced the Ala-Leu-Ala-Asn-Ser-Leu-Cys-Gln-Gly-Lys peptide and chymotrypsin the Ileu-Ser-Asn-His-Ala-Tyr peptide.

Measurement of Deamidation by 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 PHD2000 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. 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).

The 5-ml syringe contained 1.5% acetic acid in acetone, and the 10-ml syringe contained purified H_2O. These were pumped to give a combined flow of 40 μl/minute of solution that was 0.5% acetic acid and 33% acetone into the electrospray source. The heated capillary was at 180°C. The solutions were combined, after sample introduction into the H_2O stream, in a U.466 Upchurch PEEK static mixing T (Upchurch Scientific, Oak Harbor, WA). 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. Meticulous attention was paid to the materials in contact with the sample, and the hold-up volume and surface area of the sample delivery system were minimized. Teflon and Tefzel were used when available.

Before analysis, the samples were diluted 500-fold with H_2O by pipetting of 1.5 μl into 749 μl in the injection vials. In the case of proteins and larger peptides, H_2O was substituted for acetone in the injection system to avoid precipitation. 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.

A typical 4-h mass spectrometry run consisted of 36 samples, with 18 from each of 2 peptide rate experiments. These included peptides of different masses, and the samples were alternated to mutually wash the sample delivery system. 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 by 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 protein experiments.

Calculation of Results. Raw data from all of the sample-containing zoom scans from each LCQ run were summed, averaged, and recorded in Microsoft EXCEL tables to give results as shown in Figs. 1 and 2. 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 as shown in Fig. 3. These calculations were performed in MATHCAD 8 PROFESSIONAL (Math Soft, Cambridge, MA) with overlays from MACRO EXPRESS 2000 (Insight Software Solutions, Bountiful, VT) that permit large numbers of calculations to be made sequentially and automatically. 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 h of mass spectrometer operation. With the aid of our computer programs, each 4 h of mass spectrometer operation requires about 30 min of manual computer manipulations to calculate the results, largely because our programs are not integrated into the LCQ software.

Results and Discussion. The results are summarized in Tables 1 and 2. The first-order deamidation half times for the 306 asparaginyl peptides are between 1 day and 455 days with a distribution function as shown in Fig. 4. Clearly, the side chain of the amino acid residue on the carboxyl side of the asparaginyl residue has a larger effect on the deamidation rate than does the residue on the amino side. The residues on both sides affect the deamidation rate in an ordered way that is explainable from their structures and the proposed cyclic imide mechanism. Table 1 reports deamidation rates for amide-side Pro, but not for carboxyl-side Pro. For all 18 peptides with carboxyl-side Pro, our measurements showed deamidation half times of more than 1,000 days. Because Pro cannot participate in the five-membered imide ring, this result would be expected. The duration of our experiments did not, however, allow the careful and precise measurement of these rates, so the values will be reported later after they have been investigated more carefully.

The precision error of the values reported in Table 1 is generally less than 1%. Systematic errors, however, are also present. In some cases, the amidated and deamidated peptides were differentially retarded and slightly absorbed in the sample delivery system. Usually, the more positively charged peptides were absorbed more strongly, so the undeamidated peptides were absorbed more extensively than were the deamidated products. For this reason, we estimate that the absolute values of these deamidation rates have probable errors of about 5% or less.

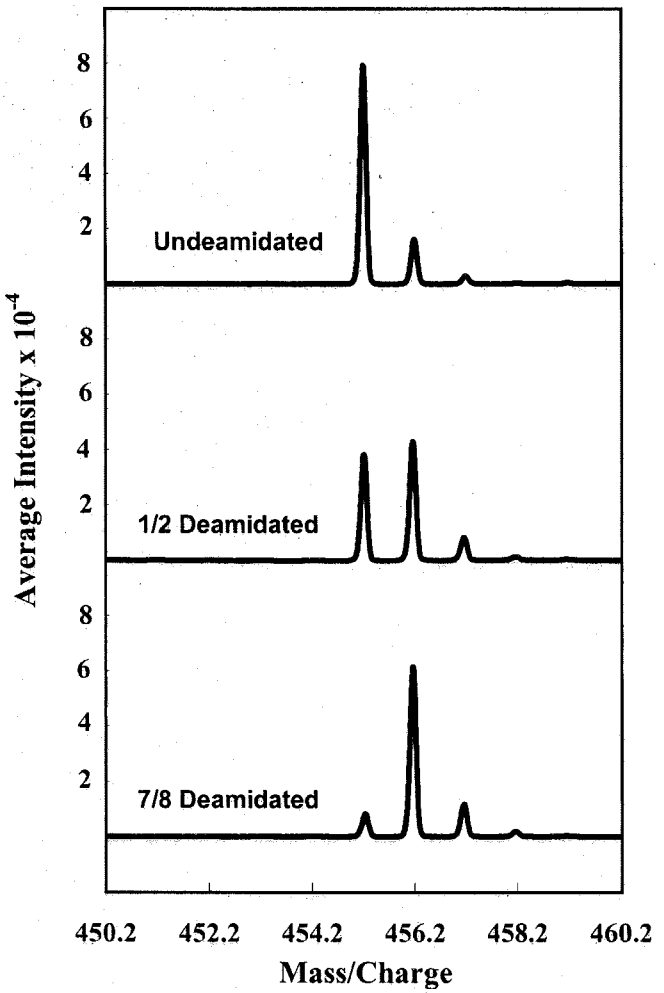

Fig. 1. Representative averaged zoom scans from direct 50-μl loop injection of the peptide Gly-Ala-Asn-His-Gly into the LCQ mass spectrometer. Scans are from the first, sixth, and fifteenth points of the deamidation experiment. Deamidation solutions were 1.0×10^{-3} M peptide, 0.15 M Tris·HCl, pH 7.4, and 37.0°C. Injected solution was 2×10^{-6} M peptide and 3×10^{-4} M Tris·HCl buffer. The graphs shown are actual and typical unsmoothed experimental data. The graphical base lines have been omitted to show the quality of these results.

At pH 7.4, these peptides are not homogeneous molecular species. For example, the pKs of the carboxyl, imidazole, and amino groups in the peptide Gly-Ser-Asn-His-Gly are 3.1, 6.4, and 7.8, respectively (25). Therefore, whereas the carboxyl and imidazole groups are mostly deprotonated at pH 7.4, the amino group, to a significant extent, is an equilibrium mixture of the protonated and deprotonated species, which have different but similar deamidation rates (25). Some of the variability seen in Table 1 is, therefore, likely to be the result of differences in pK of the peptides, but most of the variation arises from peptide sequence.

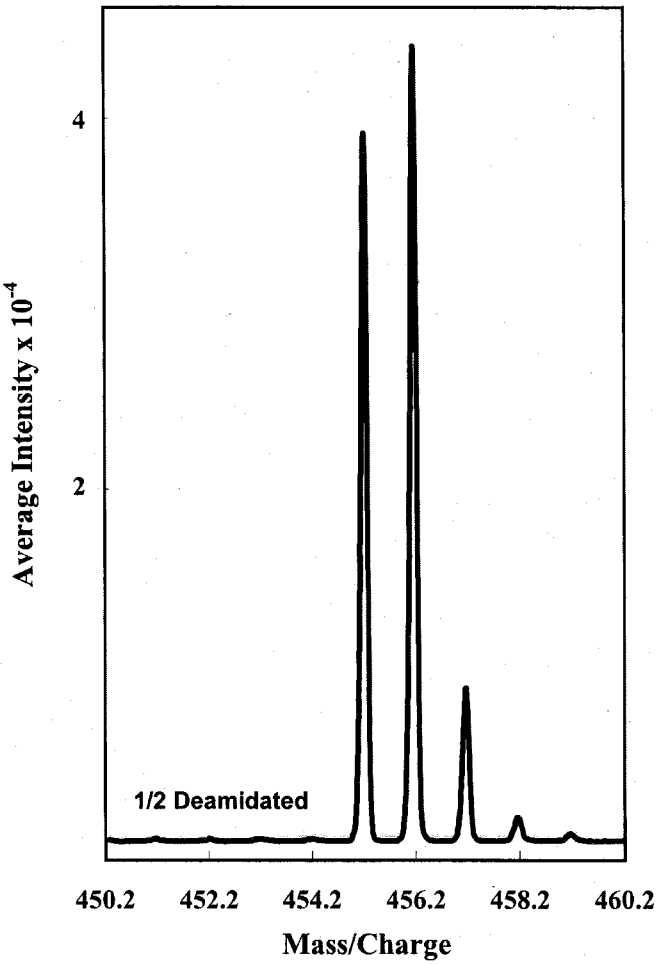

Fig. 2. Unsmoothed and uncorrected experimental data for averaged zoom scans of the sixth deamidation point in the deamidation of Gly-Ala-Asn-His-Gly. This graph is typical of actual experimental data obtained for the 306 peptides in these experiments.

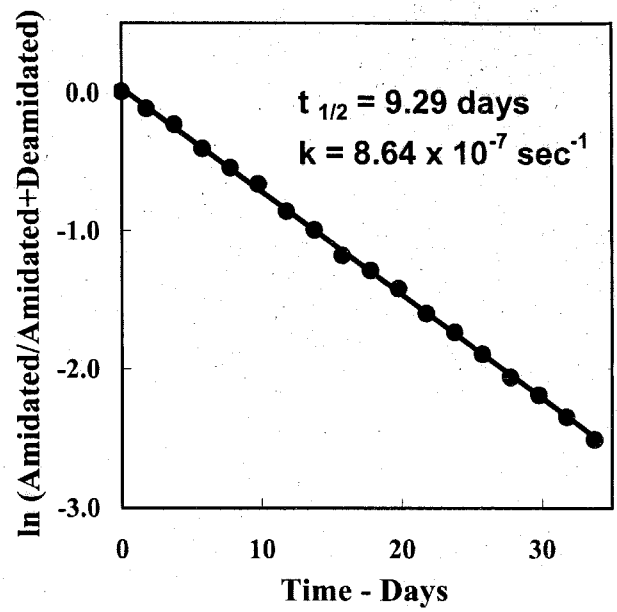

Fig. 3. First-order rate plot of the deamidation of the peptide Gly-Ala-Asn-His-Gly. Deamidation half time for the 1.0×10^{-3} M peptide in 0.15 M Tris·HCl, at pH 7.4, and 37.0°C was calculated to be 9.3 days, and the first-order rate constant $k = 0.86 \times 10^{-6}$·sec^{-1}.

We have chosen Tris·HCl buffer for these experiments, because it is relatively benign in its direct effects on deamidation rate as compared, for example, to phosphate (3, 12). As these studies are expanded, the Tris·HCl rates will serve as a base to which the effects of other solute molecules can be added. Of central importance, however, is the comparison of peptide deamidation rates with those of proteins with corresponding sequences under identical solvent and environmental conditions. For this reason, the results reported in Table 2 are of special importance.

Table 2 shows that, as expected from previous work (15–16), the C-terminal Ser-Asn-His 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. The Tris·HCl value of 8.3 days for the deamidation of Gly-Ser-Asn-His-Gly, as compared with that for aldolase of 9.4 days, is in good agreement with that previously reported of 6.4 days for Gly-Ser-Asn-His-Gly in phosphate buffer (16) and aldolase of 8 days (15). Phosphate buffer is known to accelerate deamidation (12). It is also in good agreement with the reported *in vivo* turnover rate for rabbit muscle aldolase of 8 days (15).

Conversely, however, the second amide sequence from the C-terminal, Ala-Asn-Ser, 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.

We suggest that strategically located asparaginyl residues may serve as useful quantitative detectors of helix formation in model peptide and other systems. This deamidation rate is very sensitive to helix and is easily measurable. Helix dependence of deamidation has also been reported in several model peptides (25).

We are convinced that this method for deamidation rate measurement of particular sequences in proteins, which we have demonstrated here with these two sequences in aldolase, will prove robust. Preliminary experiments that we have performed with several other proteins have shown that the majority of their amide-containing peptides are immediately measurable. Where disulfide bonds are present, it is helpful to include 1 μl of 0.1 M 1,4-dithiothreitol in the enzymatic digestion mixtures. Chromatographic and electrophoretic separations should be avoided entirely because they are time consuming and because they introduce the possibility of large systematic errors from differential losses of the amidated and deamidated peptides during measurement.

Studies of a wide variety of model peptides have shown that pentapeptide models of the sort summarized in Table 1 are good surrogates for longer peptides and proteins of similar sequences, with respect to primary structure effects on deamidation (25). It is evident that the deamidation rates of these asparaginyl pentapeptides at 37.0°C in pH 7.4, 0.15 M Tris·HCl buffer depend primarily on the adjacent carboxyl-side residue, as would be expected from the cyclic imide model.

For carboxyl-side residues without specific functionality, the order of deamidation rates is Gly, Ala, Leu, Val, and Ileu,

Table 1. First-order deamidation half times of pentapeptides GlyXxxAsnYyyGly in days at pH 7.4, 37.0°C, 0.15 M Tris·HCl

	Yyy																		
	Gly	His	Ser	Ala	Asp	Thr	Cys	Lys	Met	Glu	Arg	Phe	Tyr	Trp	Leu	Val	Ile	Mean	Median
Xxx																			
Gly	1.03	9.2	11.8	21.1	28.0	39.8	40.6	48.2	50.4	73.9	57.8	64.0	63.6	77.1	104	224	287	70.6	50.4
Ser	0.96	8.3	15.1	24.1	30.3	45.7	60.2	55.5	54.9	59.7	59.7	52.2	64.7	76.8	110	233	285	72.7	55.5
Thr	1.04	9.6	17.1	24.6	27.9	50.0	55.5	57.6	47.6	60.8	51.2	76.4	80.6	72.5	110	237	279	74.0	55.5
Cys	1.14	10.8	19.0	26.4	30.6	48.7	46.0	46.6	64.5	48.3	83.1	73.9	83.9	111	119	229	304	79.1	48.7
Met	1.04	10.2	15.2	22.1	26.4	43.6	49.6	60.4	56.9	72.4	58.8	61.9	74.0	92.7	113	211	275	73.2	58.8
Phe	1.15	10.2	18.1	24.2	27.4	39.0	46.5	58.2	58.6	62.4	61.2	69.5	75.1	102	118	203	287	74.2	58.6
Tyr	1.49	10.2	11.9	24.3	28.4	38.1	48.6	55.1	64.3	41.0	56.9	58.0	70.6	120	118	241	306	76.1	55.1
Asp	1.53	9.7	17.0	24.0	29.4	52.4	54.1	75.9	57.3	46.8	87.2	70.1	70.4	80.3	111	241	298	78.0	57.3
Glu	1.45	9.0	16.4	25.8	32.0	36.8	44.2	77.8	59.6	60.3	80.9	70.2	94.5	98.4	130	268	279	81.4	60.3
His	1.14	10.7	15.7	24.6	31.2	47.2	43.9	50.2	63.1	69.4	48.9	72.1	82.3	95.4	116	247	327	79.2	50.2
Lys	1.02	10.5	15.6	23.6	34.0	58.1	49.0	53.5	60.9	72.5	57.4	70.1	96.7	98.1	119	246	313	81.1	58.1
Arg	1.00	10.0	14.3	24.4	34.7	50.7	50.5	49.6	74.4	68.3	67.4	68.3	90.0	127	128	247	311	83.4	67.4
Ala	1.05	9.3	14.9	22.5	31.9	43.5	63.7	55.9	59.2	74.1	62.4	65.6	73.9	130	124	254	300	81.5	62.4
Leu	1.08	10.7	16.7	25.1	32.1	46.1	53.5	60.1	62.6	56.7	62.1	72.4	75.7	74.5	155	294	391	87.6	60.1
Val	1.23	10.2	18.2	27.5	33.5	49.9	63.2	63.8	65.7	64.8	67.4	66.6	79.2	88.9	154	291	366	88.9	64.8
Ile	1.26	11.5	14.5	25.9	33.8	46.3	52.7	64.4	58.8	58.6	66.4	61.5	79.3	86.7	154	295	384	87.9	58.8
Trp	1.75	11.3	15.5	30.7	43.6	38.9	83.1	59.4	64.2	75.7	73.9	71.1	92.6	135	133	226	286	84.8	71.1
Pro	1.18	12.8	18.9	31.8	48.6	63.1	60.0	67.8	78.4	92.0	72.9	100	114	122	181	364	455	111	72.9
Mean	1.20	10.2	15.9	25.1	32.4	46.5	53.6	58.9	61.2	64.3	65.3	69.1	81.2	99	128	253	318	81.4	61.2
SD	0.05	0.25	0.50	0.65	1.4	1.7	2.4	2.1	1.8	2.9	2.6	2.4	3.1	5.0	5.0	9.3	11.9	3.1	2.4
%SD	4.5	2.4	3.2	2.6	4.2	3.6	4.5	3.6	2.9	4.6	4.0	3.4	3.8	5.0	3.9	3.7	3.7	3.7	3.7
Median	1.14	10.2	15.7	24.5	31.6	46.2	51.6	57.9	60.3	63.6	62.3	69.8	79.3	97	119	243	302	78.5	60.3

entirely as expected from steric hindrance. Phe, Tyr, and Trp are between Ala and Leu and likely also contribute primarily steric effects. The series Ser, Thr, Cys, and Met follows the order of the dipole moments of the functional groups, which are geometrically available to stabilize imide formation as is the imidazole group of His. The Asp side chain is short enough for it, too, to contribute significantly to deamidation, but the longer Glu, Lys, and Arg side chains are inhibited by their length.

The amino side residues are also listed in Table 1 in order of their deamidation-enhancing effect. Although a smaller effect, this is easily distinguishable. This listing is approximately in the order of number of peptide rates above and below the medians, with allowances made for special structures. The order found is qualitatively similar to that of the carboxyl side.

Some special effects are notable. For example, peptides with paired nearest-neighboring basic and acidic residues, Glu-Asn-Lys, Glu-Asn-Arg, Asp-Asn-Lys, and Asp-Asn-Arg, clearly stand out as having deamidation half times about 50% higher than the similar singular Glu, Asp, Lys, and Arg analogues. In contrast, this same effect can be distinguished for Lys-Asn-Glu, Arg-Asn-Glu, Lys-Asn-Asp, and Arg-Asn-Asp, but it raises deamidation half times in these peptides only by about 10%. Detailed discussion and mechanistic considerations of the deamidation rates reported in Table 1 is beyond the scope of this report and will be discussed elsewhere.

The deamidation rates in Table 1 serve as a basis for the design of model peptides appropriate for the separation of primary, secondary, tertiary, and quaternary structure effects in experiments on proteins. Used in combination with the protein deamidation measurement technique exemplified in Table 2, they are suitable for the evaluation of nearest-neighbor sequence effects. Effects have also been observed from functional groups farther removed from Asn and from chain elongation (25). We are now maintaining a supply of the 800 possible Gly-Xxx-Asn-Yyy-Gly and Gly-Xxx-Gln-Yyy-Gly peptides and 113 other amide-containing peptides in our laboratory and are interested in making them available without cost to investigators wishing to make use of these peptide models in the comparative evaluation of deamidation of specific proteins.

Table 2. Deamidation of rabbit muscle aldolase and model peptides in 1.0×10^{-3} M peptide or protein, 0.15 M Tris·HCl, pH 7.4, and 37.0°C in the same solution

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

Conclusions

The values reported in Table 1 strengthen the hypothesis that deamidations of peptide and protein amides serve as ubiquitous

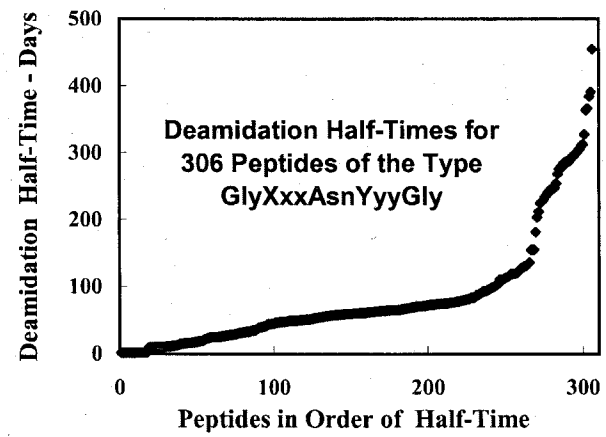

Fig. 4. Distribution function of deamidation half times for the 306 asparaginyl pentapeptides of the type Gly-Xxx-Asn-Yyy-Gly listed in Table 1. Half times are for 1.0×10^{-3} M peptide in 0.15 M Tris·HCl, pH 7.4, and 37.0°C.

and versatile clocks for the timing of biological events. A necessary condition of this hypothesis is that amide clocks be available that are suitable for the timing of events throughout the useful range of biological processes. Although the evidence for this was good (3) on the basis of the approximately 30 asparaginyl peptides and 30 glutaminyl peptides previously synthesized and measured, it is far more robust now that a complete set of nearest-neighbor rates is becoming available. Fig. 4 shows that an abundance of nearest-neighbor asparaginyl sequences are available between deamidation half times of 1 day and 1 year. Glutaminyl sequences extend this range to more than 10 years (3).

Therefore, near-neighbor variations in combination with sequence effects of residues farther away in the peptide chain (25) and three-dimensional effects from secondary, tertiary, and quaternary structure surely provide a versatile set of deamidation rates suitable to the durations of most biological processes.

There is an additional possibility regarding protein alteration and turnover. Amides may serve as molecular indicators that the integrity of each protein molecule has been maintained. Because deamidation rates are suppressed by three-dimensional structures that impede the deamidation reaction, amides contained in these structures may serve as indicators that the structures are still intact. If an individual molecule is altered, deamidation within the altered region may then set off catabolism of that molecule or some other corrective process, or the altered region may deamidate to form a peptide or protein with a new function.

It is likely that the lifetime of a set of protein molecules is determined at the time of synthesis, perhaps through deamidation, as a useful compromise between the time it is expected to remain intact and useful to the organism and the effort that will be required to resynthesize it. For some proteins, of course, the time will be adjusted to remove the protein after it has completed its work because its continued existence would be harmful. Molecular clocks are needed, rather than molecule-by-molecule evaluation of proteins within the living system, because the wide variety of forms arising as the molecules age would be so great as to overwhelm even the versatility of a well-developed immune system. In addition to preprogrammed obsolescence, however, the possibility that amide evaluators in specific regions of each molecule may be keeping watch over the integrity of their regions should be considered.

Moreover, in some cases, deamidated peptides and proteins may have special biological functions wherein timed deamidation controls their delivery to the biological system. In these instances, which could control ordinary homeostatic functions or those related to timed processes such as development or aging, deamidation would serve as a functional molecular clock rather than a terminal marking clock.

Only a few biological events have been definitively shown to be regulated by amide molecular clocks. This is not surprising, because such demonstrations have, so far, required laborious and sophisticated experimentation. The original reasoning from which this hypothesis arose about 30 years ago (1–4) remains intact and is as follows.

The static properties of asparaginyl and glutaminyl residues are not unique and can be easily duplicated by some of the other 18 commonly occurring amino acid residues or others that might be used. The disruptive effect on peptide and protein structure from deamidation reactions, which introduce postsynthetic charged residues, is so great that, were they not especially useful, these amide residues would not be widely present in peptides and proteins. Therefore, it is likely that the instability of asparaginyl and glutaminyl residues is their primary biological function, and that they serve as easily programmable molecular clocks for the timing of biological processes (1–4).

We thank Professor and Mrs. R. B. Merrifield, in whose laboratory and with whose help we synthesized the peptides, and Professor Brian Chait for his advice and counsel concerning the mass spectrometry. We also thank the John Kinsman Foundation and other donors to the Oregon Institute of Science and Medicine for financial support.

1. Robinson, A. B., McKerrow, J. H. & 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. & Rudd, C. (1974) *Curr. Top. Cell. Regul.* **8,** 247–295.
4. Robinson, A. B. (1978) *Mech. Ageing Dev.* **9,** 225–236.
5. Flatmark, T. (1964) *Acta Chem. Scand.* **18,** 1656–1666.
6. Flatmark, T. & Sletten, K. (1968) *J. Biol. Chem.* **243,** 1623–1629.
7. Robinson, A. B., Scotchler, J. W. & McKerrow, J. H. (1973) *J. Am. Chem. Soc.* **95,** 8156–8159.
8. Robinson, A. B. & Tedro, S. (1973) *Int. J. Pept. Protein Res.* **5,** 275–278.
9. Robinson, A. B., McKerrow, J. H. & Legaz, M. (1974) *Int. J. Pept. Protein Res.* **6,** 31–35.
10. Robinson, A. B. & Scotchler, J. W. (1974) *Int. J. Pept. Protein Res.* **6,** 279–282.
11. Geiger, T. & Clarke, S. (1987) *J. Biol. Chem.* **262,** 785–794.
12. Scotchler, J. W. & Robinson, A. B. (1974) *Anal. Biochem.* **59,** 319–322.
13. Robinson, A. B. & Scotchler, J. W. (1973) *J. Int. Res. Commun.* **1–8,** 8.
14. Robinson, A. B. & Robinson, L. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8880–8884.
15. Midelfort, C. F. & Mehler, A. H. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1816–1819.
16. McKerrow, J. H. & Robinson, A. B. (1974) *Science* **183,** 85.
17. Solstad, T. & Flatmark, T. (2000) *Eur. J. Biochem.* **267,** 6302–6310.
18. Lindner, H., Sarg, B., Hoertnagl, B. & Helliger, W. (1998) *J. Biol. Chem.* **273,** 13324–13330.
19. Takemoto, L. & Boyle, D. (2000) *J. Biol. Chem.* **275,** 26109–26112.
20. Sun, A., Yuksel, K. U. & Gracy, R. W. (1995) *Arch. Biochem. Biophys.* **322,** 361–368.
21. Capasso, S. & Salvadori, S. (1999) *J. Peptide Res.* **54,** 377–382.
22. Clarke, S. (1987) *Int. J. Pept. Protein Res.* **30,** 808–821.
23. Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85,** 2149–2154.
24. Merrifield, R. B. (1995) *Peptides: Synthesis, Structures, and Applications*. ed. Gutte, B. (Academic, New York), pp. 94–169.
25. Robinson, N. E., Robinson, A. B. & Merrifield, R. B. (2001) *J. Pept. Protein Res.*, in press.

Prediction of protein deamidation rates from primary and three-dimensional structure

Noah E. Robinson* and Arthur B. Robinson†‡

*Division of Chemistry, California Institute of Technology, Pasadena, CA 91125; and †Oregon Institute of Science and Medicine, 2251 Dick George Road, Cave Junction, OR 97523

Communicated by Bruce Merrifield, The Rockefeller University, New York, NY, February 8, 2001 (received for review January 13, 2001)

Prediction of protein deamidation rates from primary and three-dimensional structure

Noah E. Robinson* and Arthur B. Robinson†‡

*Division of Chemistry, California Institute of Technology, Pasadena, CA 91125; and †Oregon Institute of Science and Medicine, 2251 Dick George Road, Cave Junction, OR 97523

Communicated by Bruce Merrifield, The Rockefeller University, New York, NY, February 8, 2001 (received for review January 13, 2001)

A method for the quantitative estimation of instability with respect to deamidation of the asparaginyl (Asn) residues in proteins is described. The procedure involves the observation of several simple aspects of the three-dimensional environment of each Asn residue in the protein and a calculation that includes these observations, the primary amino acid residue sequence, and the previously reported complete set of sequence-dependent rates of deamidation for Asn pentapeptides. This method is demonstrated and evaluated for 23 proteins in which 31 unstable and 167 stable Asn residues have been reported and for 7 unstable and 63 stable Asn residues that have been reported in 61 human hemoglobin variants. The relative importance of primary structure and three-dimensional structure in Asn deamidation is estimated.

biological clocks | proteins

The spontaneous deamidation of glutaminyl and asparaginyl residues causes experimentally and biologically important changes in peptide and protein structures. In asparaginyl deamidation, the primary reaction products are aspartyl and isoaspartyl. Early work on peptide and protein deamidation (1–10) established that deamidation occurs *in vitro* and *in vivo* and depends on primary sequence, three-dimensional (3D) structure, pH, temperature, ionic strength, buffer ions, and other solution properties.

It was hypothesized (3, 5, 7) and then experimentally demonstrated (2, 8, 9, 11) that deamidation can serve as a biologically relevant molecular clock that regulates the timing of *in vivo* processes. Substantial evidence supports the hypothesis that Asn deamidation at neutral pH proceeds through a cyclic imide reaction mechanism (12–14).

A procedure is needed whereby the stability of individual amides in peptides and proteins can be reliably estimated. Although it was evident to investigators 30 years ago (2–7) that protein deamidation rates depend on primary, secondary, tertiary, and quaternary protein structure, and numerous examples have been found, it was not possible to devise a useful deamidation prediction procedure until a complete library of deamidation rates as a function of primary sequence was available.

A suitable library of sequence-determined Asn rates has now been published (15), and the relevance of this library has been established (16). These rates can now be combined with 3D data to provide a useful deamidation prediction procedure. Each amide residue has an intrinsic sequence-determined deamidation rate, which depends on charge distribution, steric factors, and other aspects of peptide chemistry. This primary rate is modulated by 3D structure, which usually slows the rate. In a few instances, it increases the deamidation rate.

We have devised a simple procedure that is useful for predicting the relative deamidation rates of most protein Asn residues. We have tested this procedure on a complete set of all proteins for which, during a review of the literature, we found experiments specifically identifying one or more labile Asn residues in a protein and also a suitable 3D structure for that same protein. Although our procedure assumes that deamidation proceeds through a cyclic five-membered imide formed by reaction of the Asn amide side chain with the nearest carboxyl-side peptide bond nitrogen, it would likely give good results even if the actual mechanism were different.

When sequence-dependent rates of deamidation first became available (3–10), it was found that most protein deamidation rates were slower than those of corresponding model peptides, except in protein amides located in especially flexible regions such as those that initiate the *in vivo* turnover of cytochrome C (2, 8) and aldolase (9, 11). Deamidation suppression of Asn in α-helices has been demonstrated (15–17), and it is evident that Asn deamidation generally depends on 3D freedom in the peptide chain.

We have limited this Asn deamidation prediction procedure to 3D observations that can easily be made with an ordinary personal-computer-based 3D protein structure viewer and 1–2 hours of work per protein without special computer programs or other aids. Subtle or complicated 3D effects have, therefore, been omitted. Although it is to be expected that sophisticated computerized procedures for this purpose will eventually be devised, there are not yet sufficient experimental data with which to calibrate such procedures.

Materials and Methods

Selection of Proteins. 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 Brookhaven Protein Data Bank (http://www.rcsb.org/pdb) was then searched for a 3D 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 3D structure is included herein. None have been 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.

The selected proteins and their Brookhaven Protein Data Bank identification numbers are: rabbit aldolase 1ADO (11, 18), human angiogenin 1B1I (19, 20), bovine calbindin 4ICB (21, 22), pig cAMP-dependent protein kinase 1CDK (23, 24), horse cytochrome C 2GIW (NMR) (25, 26), mouse epidermal growth factor 1EGF (NMR) (27, 28), rat fatty acid-binding protein 1LFO (29, 30), human fibroblast growth factor 2AFG (31, 32), *Aspergillus awamorii* glucoamylase 3GLY (33, 34), human growth hormone 1HGU (35, 36), human hemoglobin 1A3N (37–44, 45), *Escherichia coli* Hpr-phosphocarrier protein 1HDN (NMR) (46, 47), human hypoxanthine guanine phosphoribosyl

Abbreviation: 3D, three-dimensional.

‡To whom reprint requests should be addressed. E-mail: art@oism.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

transferase 1BZY (48, 49), human insulin 2HIU (NMR) (50, 51), mouse interleukin 1β 2MIB (52, 53), human interleukin 2 3INK (54, 55), chicken lysozyme 1E8L (NMR) (56, 57), bovine ribonuclease A 1AFK (58, 59), *Ustilago sphaerogena* ribonuclease U2 1RTU (60, 61), bovine seminal ribonuclease 11BG (62, 63), human T cell surface protein CD4 1CDJ (64, 65), human thioltransferase 1JHB (NMR) (66, 67), human triosephosphate isomerase 1HTI (68, 69), and bovine trypsin 1MTW (70, 71).

Trypsin is included, but the reported (70) 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 measurements on an undifferentiated solution were reported.

Selection of 3D Parameters. A set of observations of the 3D 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 by one of us (N.E.R.) before any calculations were carried out. The tabulated observations were not changed after calculations began. 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 with MICROSOFT NT 4.0.

The deamidation coefficient, C_D, is defined as $C_D = (0.01)(t_{1/2})(e^{f(C_m, C_{S_n}, S_n)})$, where $t_{1/2}$ is the pentapeptide primary structure half life (15), C_m is a structure proportionality factor, C_{Sn} is the 3D structure coefficient for the nth structure observation, S_n is that observation, and $f(C_m, C_{S_n}, S_n) = C_m[(C_{S_1})(S_1) + (C_{S_2})(S_2) + (C_{S_3})(S_3) - (C_{S_{4,5}})(S_4)/(S_5) + (C_{S_6})(S_6) + (C_{S_7})(S_7) + (C_{S_8})(S_8) + (C_{S_9})(S_9) + (C_{S_{10}})(1 - S_{10}) + (C_{S_{11}})(5 - S_{11}) + (C_{S_{12}})(5 - S_{12})]$. 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 N 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 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. 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 (15), except for Asn with carboxyl-side Pro, Asn, or Gln and N-glycosylated Asn. We used estimated values of $t_{1/2}$ of 500, 40, 60, and 500 days for Asn-Pro, Asn-Asn, Asn-Gln, and N-glycosylated Asn, respectively.

Optimization of the Coefficient of Deamidation. C_D values were optimized (72, 73) by using various values for C_m and C_{S_n} to maximize the value of the deamidation resolving power, D_P. The optimized values were $C_m = 0.48$, $C_{S_1} = 1.0$, $C_{S_2} = 2.5$, $C_{S_3} = 10.0$, $C_{S_{4,5}} = 0.5$, $C_{S_6} = 1.0$, $C_{S_7} = 1.0$, $C_{S_8} = 3.0$, $C_{S_9} = 2.0$, $C_{S_{10}} = 2.0$, $C_{S_{11}} = 0.2$, and $C_{S_{12}} = 0.7$.

For example, the β-Lys-Asn 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

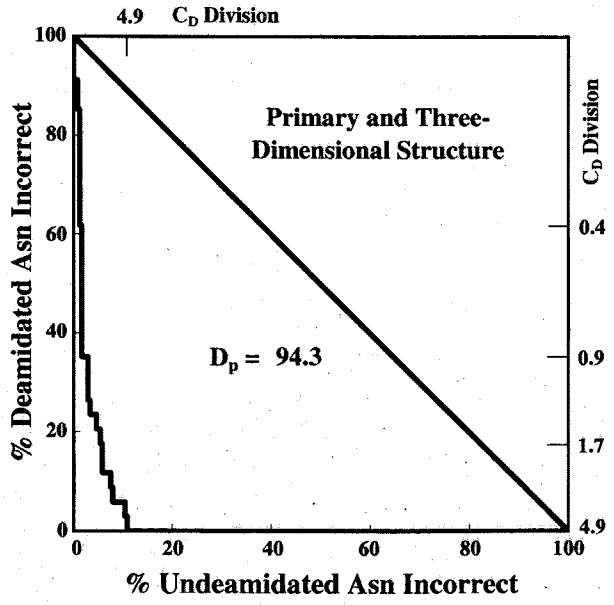

Fig. 1. Classification accuracies of the Asn residues in Tables 1 and 2 with all possible C_D division values used for the classification, excluding four Asn marked ‡ in Table 1 and ‡ and ‡‡ in Table 2, and calculated deamidation resolving power (D_P).

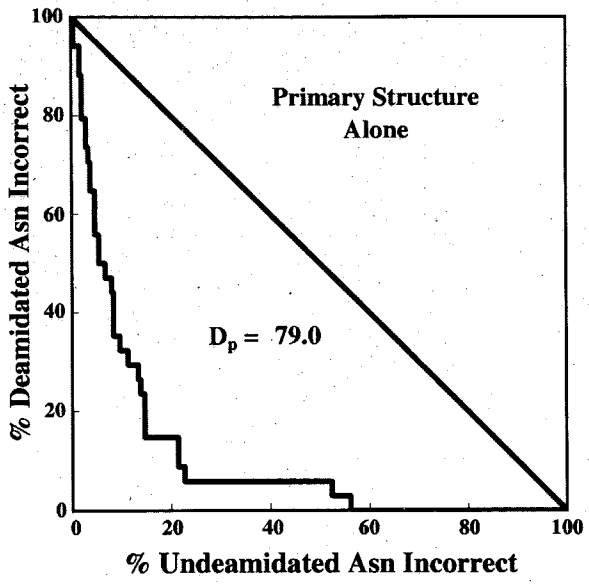

Fig. 2. Tabulation and calculation as in Fig. 1, but by using only the primary structure part of the coefficients C_D. $C_m = 0$.

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 Gly-Lys-Asn-His-Gly half life (15) is 10.5 days. Therefore, $C_D = (0.01)(10.5)e^{(0.48)[(1)(1)+(2)(1)+(2)(1-0)+(0.2)(4)]} = (0.105)e^{(0.48)(5.8)} = (0.105)(16.184) = 1.70$.

The D_P calculation method as developed previously for the evaluation of quantitative procedures in diagnostic medicine (72, 73) was used as illustrated in Figs. 1–3. A total of 264 Asn residues listed in Tables 1 and 2 were arranged in order of calculated C_D values and then divided into all possible two group sets arising from division at all possible C_D values. The errors at these division points for the optimized parameters are graphed

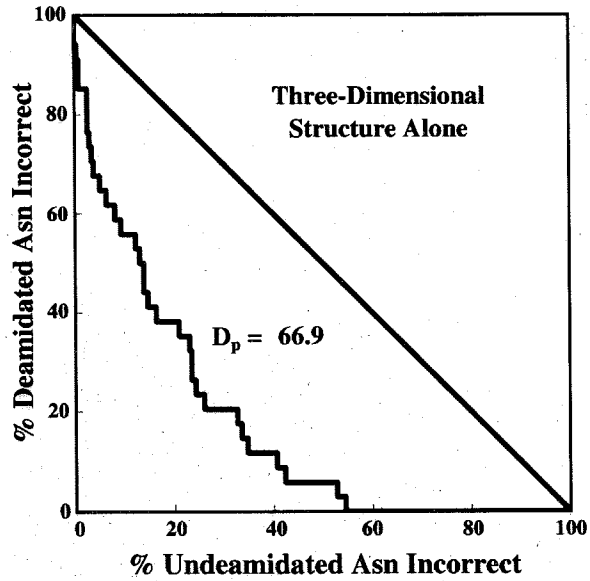

Fig. 3. Tabulation and calculation as in Fig. 1, but by using only the 3D structure part of the coefficients C_D. All $t_{1/2} = 1$.

in Fig. 1. Figs. 2 and 3 show graphs for primary structure and 3D structure alone. If the classification of Asn stabilities were perfect, then the graphs in Figs. 1–3 would be straight lines along the axes, appearing as points in the origin. If there were no correlation between the calculations and the experimental data, the graphs would be along the diagonal lines. D_P is defined as the percentage of the area between the diagonal and the origin that has been successfully removed by the deamidation estimation procedure.

Two of the hemoglobin Asn mutations involve large undetermined structural changes in the protein, one by a frame-shift and the other causing the loss of the heme group, so suitable 3D criteria could not be tabulated. 3D effects apparently markedly accelerate deamidation of Asn 54 in cytochrome C and Asn 88 in interleukin 2. These four Asn were not used in calculating D_P.

Reliability of the Coefficient of Deamidation. In addition to D_P, the Asn ranks within each protein as shown in Table 1 are especially interesting because these ranks avoid the complication that the different proteins were subjected to a wide variety of differing deamidating conditions. All 70 Asn in the hemoglobin set shown in Table 2 were incubated *in vivo* at 37°C for an average of 60 days in human blood.

Although the Asn residues designated as deamidating have been reported from experiments, those designated as undeamidating depend on negative results. In many cases, ammonia evolution or protein separation experiments have shown that additional unstable amides are present in these proteins. This is reflected in the asymmetry seen in Fig. 1, wherein some of the "% deamidated Asn incorrect" at low C_D values are probably correctly assigned but not yet reported. We expect that some of the Asn residues listed in Tables 1 and 2 with low C_D values will eventually be found to significantly deamidate.

The values of C_D depend on 18 x-ray diffraction and 6 NMR structures. Although the deamidation of aldolase Asn-360 is known to be entirely sequence controlled *in vivo* and *in vitro* with no 3D suppression (9, 15, 16), the x-ray crystal structure shows one suppressing hydrogen bond. This aldolase C_D is, therefore, 0.22. It should be 0.08. Solution structures are best used when available.

Multiplication of the coefficient of deamidation (C_D) by 100 provides a semiquantitative prediction of Asn deamidation half times in 37°C, pH 7.4, 0.15 M Tris·HCl buffer, even though C_D does not include all aspects of 3D structure. Table 3 lists those proteins for which experimental deamidation half times at 37°C, pH at or near 7.4, but with a wide range of buffer types and solution properties have been reported (2, 31, 39, 50, 74–78) vs. the corresponding values of $(100)(C_D)$ for those Asn. The overall differences in Table 3 are well within the range expected from variations in buffer type and other solvent conditions.

It is customary to guess which Asn residues may easily deamidate on the basis of primary structure. With the complete rate table (15) and 33 of the deamidating Asn residues in our data set, the sequence assumptions that these types of Asn residues easily deamidate are 49% in error even in the very unstable Asn-Gly sequences, 70% in the Asn-Ser and Asn-His sequences, 83% in Asn-Ala and Asn-Asp, and 91% in Asn-Gln, Asn-Lys, and Asn-Tyr. The converse nondeamidation assumptions are 51%, 30%, 17%, and 9% in error, respectively (see Fig. 4). In comparison, Fig. 1 shows that a division criterion of $C_D \leq 3$ leads to less than 6% error in classification of all easily deamidating and all relatively stable Asn residues, simultaneously. A criterion of $C_D \leq 5$ includes 100% of deamidating Asn residues, except for Asn 54 in cytochrome C and Asn 88 in interleukin 2.

Table 1. Ordered deamidation coefficients and experimentally determined deamidating Asn residues in 23 proteins

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

Squares designate Asn reported as deamidated. ___ designates Deamidation Index, I_D.
‡These two unshaded squares designate unusual protein structures that accelerate deamidation.
‡‡Uses primary $t_{1/2}$ from ref. 8.
‡‡‡Uses primary $t_{1/2}$ from ref. 16.

Deamidation Index. The initial deamidation of a protein at neutral pH causes a unit decrease in charge. We define $I_D = [\Sigma(C_{D_n})^{-1}]^{-1}$, where C_{D_n} is C_D for the nth Asn residue, as the protein "deamidation index." Therefore $(100)(I_D)$ is an estimate of the initial single-residue deamidation half time for the protein with all Asn residues considered, as shown in Tables 1 and 2.

Results and Discussion

This calculation method, based on the sequence-controlled deamidation rates of Asn model peptides and simple aspects of the Asn 3D environment in proteins, permits a useful estimation of the instability with respect to deamidation of Asn in proteins.

Table 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 3D structures are unknown, and C_D derived from wild-type hemoglobin is not applicable. Squares designate Asn reported as deamidated. ___ designates wild-type deamidation index, I_D.

For a diverse group of protein types, this method is at least 94% reliable, as illustrated in Fig. 1. This reliability is underestimated, because the evaluation in Fig. 1 considers all of these protein amides simultaneously even though their deamidations were observed under a wide variety of experimental conditions. Moreover, some experimentally known Asn instabilities in these proteins have not yet been characterized, so the data used in Fig. 1 incorrectly classify some Asn as stable that are actually unstable.

When used to determine the reportedly most unstable Asn residues within a single protein as illustrated in Tables 1 and 2, this method correctly identifies the most unstable Asn residue

Table 3. Deamidation half times in days at 37°C, pH 7.4 vs. estimates by (100)(C_D)

	Experimental*	Calc (100)(C_D)
Hpr–phosphocarrier protein (Asn-38)	10	57
Angiogenin (Asn-61 and Asn-109)	23†	17‡
Hemoglobin (α-Asn-50)	25	18
Growth hormone (Asn-149 and Asn-152)	29†	48
Hpr–phosphocarrier protein (Asn-12)	31	21
Triose phosphate isomerase (Asn-71)	38	78
Hemoglobin (β-Asn-82)	42	19
Fibroblast growth factor (Asn-7)	60	64
Hemoglobin (β-Asn-80)	71	173
Ribonuclease A (Asn-67)	64	85‡
Insulin (B-Asn-3)	135	117

*Buffer conditions vary. pHs at or close to 7.4.
†Reported rate for sum of both Asn residues.
‡Buffer (Tris) identical to that of model peptides used to calculate C_D.

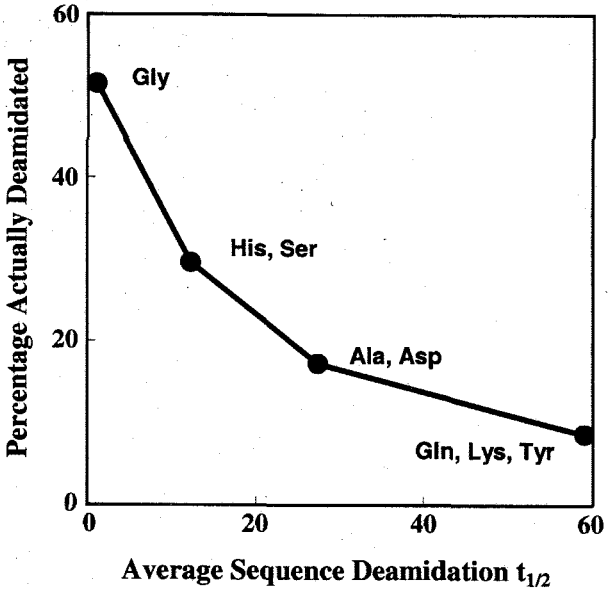

Fig. 4. Percentages of deamidating Asn residues listed in Tables 1 and 2 that would be correctly guessed by simply assuming that Asn residues with COOH-side Gly, His, Ser, Ala, Asp, Gln, Lys, or Tyr deamidate vs. average pentapeptide deamidation half times (15) for those specific Asn sequences.

for 31 of 36 residues in 24 proteins and, in 4 of the remaining 5 cases, is in error by only one residue.

This method does not allow for special 3D structures that change deamidation rates in unusual ways. There are still too few reported instances of these to permit their theoretical estimation. In two Asn sequences encountered here, Lys-Asn 54-Lys in cytochrome C and Ser-Asn 88-Ileu in interleukin 2, the reported experimentally determined protein rates are faster than the sequence determined rates. Also, in two instances, Met-Asn 15-Gly in triosephosphate isomerase (79) and Lys-Asn 54-Lys (7) in cytochrome C, deamidation takes place after a prior deamidation of the protein changes the structure in an accommodating way. Although this calculation method cannot predict these special effects, it aids in their recognition.

Finally, this procedure provides a semiquantitative answer to a previously unanswered question. What are the relative contributions to deamidation rates in proteins from primary structure and 3D structure? Figs. 1–3 serve as a reasonable basis for estimating that Asn deamidation in proteins is, on average, determined approximately 60% by primary structure and 40% by 3D structure. These percentages apply to 3D effects that diminish deamidation rates below those of primary structure alone. In 2 cases out of 36—about 6% of deamidating Asn and 1% of all Asn examined here—3D structure is reported to actually accelerate deamidation.

These calculations demonstrate that most deamidation rates of Asn residues in proteins are approximately equal to the sequence-controlled rates modulated through slowing by 3D structure. The modulated values can be estimated by a remarkably simple calculation. We are now experimentally determining a complete deamidation rate table for Gln residues in pentapeptides, which should allow a similar treatment for Gln residues in proteins. Values of I_D and C_D for many other proteins are available at www.deamidation.org.

We thank Prof. and Mrs. R. B. Merrifield for their advice and encouragement. We also thank the John Kinsman Foundation and other donors to the Oregon Institute of Science and Medicine for financial support.

1. Flatmark, T. (1964) *Acta Chem. Scand.* **18,** 1656–1666.
2. Flatmark, T. & Sletten, K. (1968) *J. Biol. Chem.* **243,** 1623–1629.
3. Robinson, A. B., McKerrow, J. H. & Cary, P. (1970) *Proc. Natl. Acad. Sci. USA* **66,** 753–757.
4. Robinson, A. B., Scotchler, J. W. & McKerrow, J. H. (1973) *J. Am. Chem. Soc.* **95,** 8156–8159.
5. Robinson, A. B. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 885–888.
6. Scotchler, J. W. & Robinson, A. B. (1974) *Anal. Biochem.* **59,** 319–322.
7. Robinson, A. B. & Rudd, C. (1974) *Curr. Top. Cell. Regul.* **8,** 247–295.
8. Robinson, A. B., McKerrow, J. H. & Legaz, M. (1974) *Int. J. Pept. Protein Res.* **6,** 31–35.
9. McKerrow, J. H. & Robinson, A. B. (1974) *Science* **183,** 85.
10. Robinson, A. B. & Scotchler, J. W. (1974) *Int. J. Pept. Protein Res.* **6,** 279–282.
11. Midelfort, C. F. & Mehler, A. H. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1816–1819.
12. Bornstein, P. & Balian, G. (1970) *J. Biol. Chem.* **245,** 4854–4856.
13. Meinwald, Y. C., Stimson, E. R. & Scheraga, H. A. (1986) *J. Pept. Protein Res.* **28,** 79–84.
14. Geiger, T. & Clarke, S. (1987) *J. Biol. Chem.* **262,** 785–794.
15. Robinson, N. E. & Robinson, A. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 944–949.
16. Robinson, N. E., Robinson, A. B. & Merrifield, R. B. (2001) *J. Pept. Protein Res.* **57,** 1–12.
17. Kosky, A. A., Razzaq, U. O., Treuheit, M. J. & Brems, D. N. (1999) *Protein Sci.* **8,** 2519–2523.
18. Blom, N. & Sygusch, J. (1997) *Nat. Struct. Biol.* **4,** 36–39.
19. Hallahan, T. W., Shapiro, R., Strydom, D. J. & Vallee, B. L. (1992) *Biochemistry* **31,** 8022–8029.
20. Leonidas, D. D., Shapiro, R., Allen, S. C., Subbarao, G. V., Veluraja, K. & Acharya, K. R. (1999) *J. Mol. Biol.* **285,** 1209–1233.
21. Chazin, W. J., Kordel, J., Thulin, E., Hofmann, T., Drakenberg, T. & Forsen, S. (1989) *Biochemistry* **28,** 8646–8653.
22. Svensson, L. A., Thulin, E. & Forsen, S. (1992) *J. Mol. Biol.* **223,** 601–606.
23. Jedrzejewski, P. T., Girod, A., Tholey, A., Konig, N., Thullner, S., Kinzel, V. & Bossemeyer, D. (1998) *Protein Sci.* **2,** 457–469.
24. Bossemeyer, D., Engh, R. A., Kinzel, V. Ponstingl, H. & Huber, R. (1993) *EMBO J.* **12,** 849–859.
25. Flatmark, T. (1966) *Acta Chem. Scand.* **20,** 1487–1496.
26. Banci, L., Bertini, I., Huber, J. G., Spyroulias, G. A. & Turano, P. (1999) *J. Biol. Inorg. Chem.* **4,** 21–33.
27. DiAugustine, R. P., Gibson, B. W., Aberth, W., Kelly, M., Ferrua, C. M., Tomooka, Y., Brown, C. F. & Walker, M. (1987) *Anal. Biochem.* **165,** 420–429.
28. Montelione, G. T., Wuthrich, K., Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D. & Scheraga, H. A. (1992) *Biochemistry* **31,** 236–249.
29. Odani, S., Okazaki, Y., Kato, C., Uchiumi, T. & Takahashi, Y. (1993) *Arch. Biochem. Biophys.* **309,** 81–84.
30. Thompson, J., Winter, N., Terwey, D., Bratt, J. & Banaszak, L. (1997) *J. Biol. Chem.* **272,** 7140–7150.
31. Volkin, D. B., Verticelli, A. M., Bruner, M. W., Marfia, K. E., Tsai, P. K., Sardana, M. K. & Middaugh, C. R. (1994) *J. Pharmacol. Sci.* **84,** 7–11.
32. Blaber, M., Disalvo, J. & Thomas, K. A. (1996) *Biochemistry* **35,** 2086–2094.
33. Svensson, B., Larson, K., Svendsen, I. & Boel, E. (1983) *Carlsberg Res. Commun.* **48,** 529–544.
34. Aleshin, A. E., Hoffman, C., Firsov, L. M. & Honzatko, R. B. (1994) *J. Mol. Biol.* **238,** 575–591.
35. Silberring, J., Brostedt, P., Ingvast, A. & Nyberg, F. (1991) *Rapid Commun. Mass Spectrom.* **5,** 579–581.
36. Chantalat, L., Jones, N. D., Korber, F., Navaza, J. & Pavlovsky, A. G. (1995) *Protein Peptide Lett.* **2,** 333.
37. Huisman, T. H. J., Carver, M. F. H. & Efremov, G. D. (1998) *A Syllabus of Human Hemoglobin Variants* (Univ. of Georgia, Augusta, GA), 2nd Ed.
38. Wajcman, H., Kister, J., Vasseur, C., Blouquit, Y., Trastour, J. C., Cottenceau, D. & Galacteros, F. (1992) *Biochim. Biophys. Acta* **1138,** 127–132.
39. Paleari, R., Paglietti, E., Mosca, A., Mortarino, M., Maccioni, L., Satta, S., Cao, A. & Galanello, R. (1999) *Clin. Chem.* **45,** 21–28.
40. Wajcman, H., Vasseur, C., Blouquit, Y., Santo, D. E., Peres, M. J., Martins, M. C., Poyart, C. & Galacteros, F. (1991) *Am. J. Hematol.* **38,** 194–200.
41. 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. & Jones, R. T. (1996) *Am. J. Hematol.* **52,** 305–309.
42. Moo-Penn, W. F., Jue, D. L., Bechtel, K. C., Johnson, M. H. & Schmidt, R. M. (1976) *J. Biol. Chem.* **251,** 7557–7562.
43. Seid-Akhavan, M., Winter, W. P., Abramson, R. K. & Rucknagel, D. L. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 882–886.
44. Blackwell, R. Q., Boon, W. H., Liu, C. S. & Weng M. I. (1972) *Biochim. Biophys. Acta* **278,** 482–490.
45. Tame, J. & Vallone, B. (2000) *Acta Crystallogr. D Biol. Cryst.* **56,** 805–811.
46. Sharma, S., Hammen, P. K., Anderson, J. W., Leung, A., Georges, F., Hengstenberg, W., Klevit, R. E. & Waygood, E. B. (1993) *J. Biol. Chem.* **268,** 17695–17704.
47. Van Nuland, N. A., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J., van Gunsteren, W. F., Scheek, R. M. & Robillard, G. T. (1994) *J. Mol. Biol.* **237,** 544–559.
48. Wilson, J. M., Landa, L. F., Kobayashi, R. & Kelley, W. N. (1982) *J. Biol. Chem.* **257,** 14830–14834.
49. Shi, W., Li, C. M., Tyler, P. C., Furneaux, R. H., Grubmeyer, C., Schramm, V. L. & Almo, S. C. (1999) *Nat. Struct. Biol.* **6,** 588–593.
50. Brange, J., Langkjaer, L., Havelund, S. & Volund, A. (1992) *Pharm. Res.* **9,** 715–726.
51. Hua, Q. X., Gozani, S. N., Chance, R. E., Hoffmann, J. A., Frank, B. H. & Weiss, M. A. (1995) *Nat. Struct. Biol.* **2,** 129–138.
52. Daumy, G. O., Wilder, C. L., Merenda, J. M., McColl, A. S., Geoghegan, K. F. & Otterness, I. G. (1991) *FEBS Lett.* **278,** 98–102.
53. Van Oostrum, J., Priestle, J. P., Grutter, M. G. & Schmitz, A. (1991) *J. Struct. Biol.* **107,** 189–195.
54. Sasaoki, K., Hiroshima, T., Kusumoto, S. & Nishi, K. (1992) *Chem. Pharm. Bull.* **40,** 976–980.
55. Brandhuber, B. J., Boone, T., Kenney, W. C. & McKay, D. B. (1987) *Science* **238,** 1707–1709.
56. Kato, A., Tanimoto, S., Muraki, Y., Kobayashi, K. & Kumagai, I. (1992) *Biosci. Biotechnol. Biochem.* **56,** 1424–1428.
57. Schwalbe, H., Grimshaw, S. B., Spencer, A., Buck, M., Boyd, J., Dobson, C. M., Redfield, C. & Smith, L. J. (2001) *Protein Sci.*, in press.
58. Wearne, S. J. & Creighton, T. E. (1989) *Proteins Struct. Funct. Genet.* **5,** 8–12.
59. Leonidas, D. D., Shapiro, R., Irons, L. I., Russo, N. & Acharya, K. R. (1997) *Biochemistry* **36,** 5578–5588.
60. Kanaya, S. & Uchida, T. (1986) *Biochem. J.* **240,** 163–170.
61. Noguchi, S., Satow, Y., Uchida, T., Sasaki, C. & Matsuzaki, T. (1995) *Biochemistry* **34,** 15583–15591.
62. Di Donato, A., Galletti, P. & D'Alessio, G. (1986) *Biochemistry* **25,** 8361–8368.
63. Vitagliano, L., Adinolfi, S., Sica, F., Merlino, A., Zagari, A. & Mazzarella, L. (1999) *J. Mol. Biol.* **293,** 569–577.
64. Teshima, G., Porter, J., Yim, K., Ling, V. & Guzzetta, A. (1990) *Biochemistry* **30,** 3916–3922.
65. Wu, H., Myszka, D. G., Tendian, S. W., Brouillette, C. G., Sweet, R. W., Chaiken, I. M. & Hendrickson, W. A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 15030–15035.
66. Papov, V. V., Gravina, S. A., Mieyal, J. J. & Biemann, K. (1994) *Protein Sci.* **3,** 428–434.
67. Sun, C., Berardi, M. J. & Bushweller, J. H. (1998) *J. Mol. Biol.* **280,** 687–701.
68. Yuan, P. M., Talent, J. M. & Gracy, R. W. (1981) *Mech. Ageing Dev.* **17,** 151–162.
69. Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A. & Hol, W. G. (1994) *Protein Sci.* **3,** 810–821.
70. Kossiakoff, A. A. (1988) *Science* **240,** 191–194.
71. Stubbs, M. T., Huber, R. & Bode, W. (1995) *FEBS Lett.* **375,** 103–107.
72. Robinson, A. B. & Westall, F. C. (1974) *J. Orth. Psych.* **3,** 70–79.
73. Robinson, A. B. & Pauling, L. (1974) *Clin. Chem.* **20,** 961–965.
74. Brennan, T. V., Anderson, J. W., Zongchao, J., Waygood, E. B. & Clarke, S. (1994) *J. Biol. Chem.* **269,** 24586–24595.
75. Johnson, B. A., Shirokawa, J. M., Hancock, W. S., Spellman, M. W., Basa, L. J. & Aswad, D. W. (1989) *J. Biol. Chem.* **264,** 14262–14271.
76. Yuksel, K. U. & Gracy, R. W. (1986) *Arch. Biochem. Biophys.* **248,** 452–459.
77. Capasso, S. & Salvadori, S. (1999) *J. Peptide Res.* **54,** 377–382.
78. Lewis, U. J., Cheever, E. V. & Hopkins, W. C. (1970) *Biochim. Biophys. Acta* **214,** 498–508.
79. Sun, A., Yuksel, K. U. & Gracy, R. W. (1995) *Arch. Biochem. Biophys.* **322,** 361–368.