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.

BIOCHEMISTRY

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

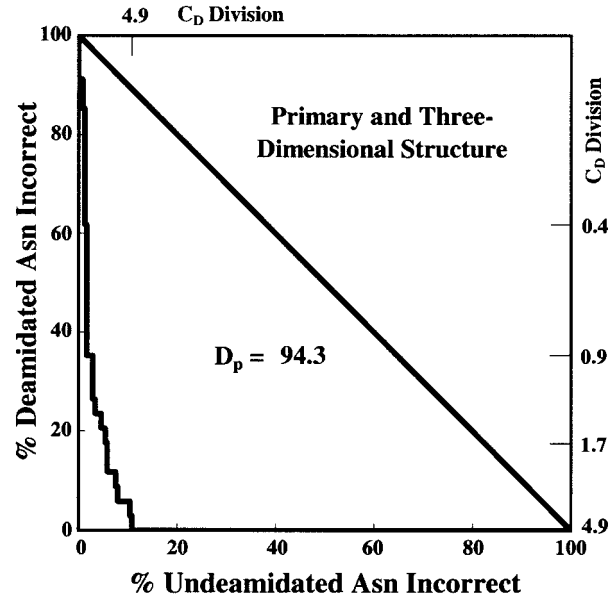

Fig. 1. Classification accuracies of the Asn residues in Tables 1 and 2 with all possible C_D division values used for the classification, excluding four Asn marked ‡ in Table 1 and ‡ and ‡‡ in Table 2, and calculated deamidation resolving power (D_P).

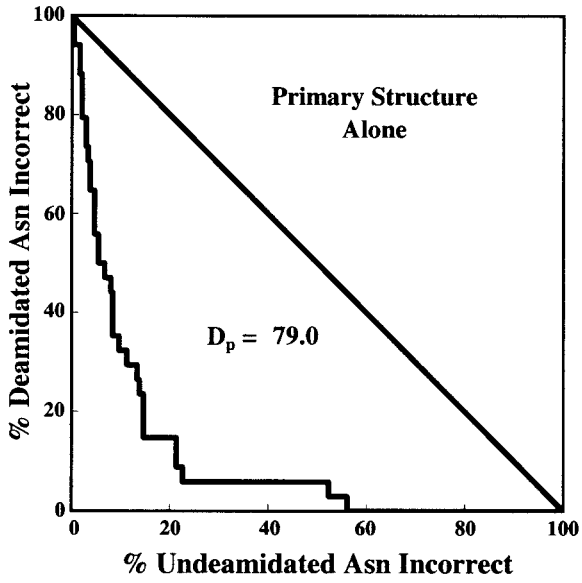

Fig. 2. Tabulation and calculation as in Fig. 1, but by using only the primary structure part of the coefficients C_D. $C_m = 0$.

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 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.

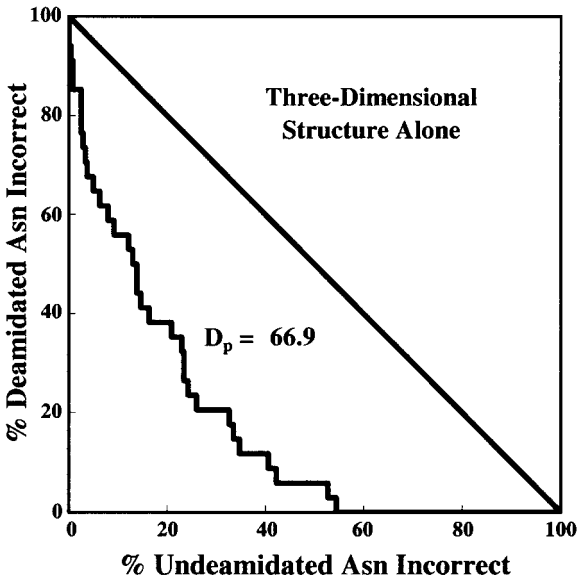

Fig. 3. Tabulation and calculation as in Fig. 1, but by using only the 3D structure part of the coefficients C_D. All $t_{1/2} = 1$.

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.

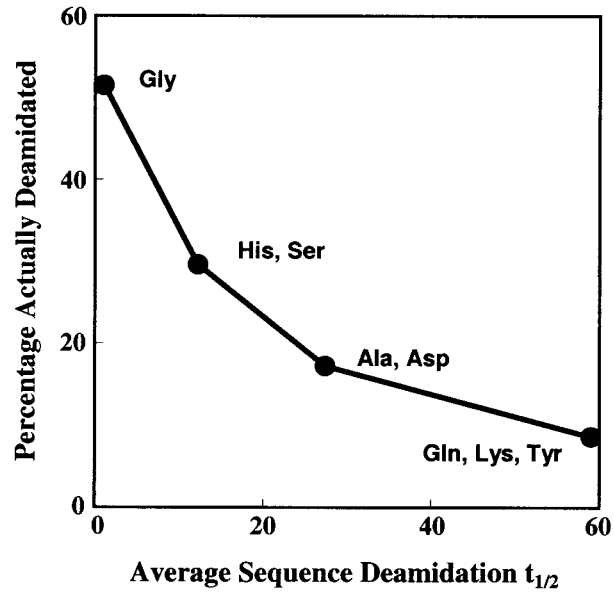

Fig. 4. Percentages of deamidating Asn residues listed in Tables 1 and 2 that would be correctly guessed by simply assuming that Asn residues with COOH-side Gly, His, Ser, Ala, Asp, Gln, Lys, or Tyr deamidate vs. average pentapeptide deamidation half times (15) for those specific Asn sequences.

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.

BIOCHEMISTRY

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. Wajeman, 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. E., 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.