Pharmaceutical Research, Vol. 26, No. 11, November 2009 (© 2009)
DOI: 10.1007/s11095-009-9953-8

Research Paper

Multivariate Analysis of the Sequence Dependence of Asparagine Deamidation Rates in Peptides

Andrew A. Kosky,[1,4] Vasumathi Dharmavaram,[2] Gayathri Ratnaswamy,[2] and Mark Cornell Manning[3]

Received January 18, 2009; accepted August 7, 2009; published online September 9, 2009

Purpose. To develop a quantitative scheme to describe and predict asparagine deamidation in polypeptides using chemometric models employing reduced physicochemical property scales of amino acids.
Methods. Deamidation rates for 306 pentapeptides, Gly-(*n−1*)-Asn-(*n+1*)-Gly, with the residues *n−1* and *n+1* varying over the naturally occurring amino acids, were obtained from literature. A multivariate regression technique, called projection to latent structures (PLS), was used to establish mathematical relationships between the physicochemical properties and the deamidation half-lives of the amino acid sequences. Three reduced physicochemical property scales, amide hydrogen exchange rates (to describe the relative acidity of the amide protons) and flexibility parameters for the sequences were evaluated for their predictive capacity.
Results. The most effective descriptors of the deamidation half-lives were reduced-property parameters for amino acids called zz-scores. The PLS models with the reduced property scales, combined with the hydrogen exchange rates and/or flexibility parameters, explained more than 95% of the sequence-dependent variation in the deamidation half-lives. The amide hydrogen exchange rate (i.e., amide proton acidity), hydrophilicity, polarizability, and size of amino acids in position *n+1* were found to be the principal factors governing the rate of deamidation. The effect of amino acids in position *n−1* was found to be negligible.
Conclusions. Chemometric analysis employing reduced physicochemical parameters can provide an accurate prediction of chemical instability in peptides and proteins. The relative importance of these various factors could also be determined.

KEY WORDS: asparagine deamidation rates; chemometrics; multivariate analysis; physicochemical property scales; projection to latent structures.

INTRODUCTION

The assessment of the stability of a new protein drug candidate typically takes place after the protein sequence has been selected, expressed and committed to development. Since the development of stabilizing formulations and manufacturing processes involves expensive and time consuming screening experiments, tools that are predictive of protein stability from primary sequence data would be advantageous to the molecule design and selection process.

The deamidation of asparagine (Asn) residues is a common degradation pathway for peptides and proteins, both *in vitro* and *in vivo* (1–5). The deamidation reaction is spontaneous, can impact the structure and function of the molecule, and occurs at rates that can limit the half-life of the protein. The deamidation reaction is particularly problematic in the development of protein-based pharmaceuticals where degradation during manufacturing and storage reduces the practical shelf-life of a drug product (5–8). Quantitative models of protein deamidation would be of great value in design and development of new protein pharmaceuticals (6) and would increase mechanistic insight into this important degradation process.

In aqueous solution above pH 6, the primary mechanism of deamidation involves the formation of a succinimide intermediate through intramolecular nucleophilic attack of the Asn side-chain carbonyl-carbon by the backbone nitrogen of the next residue (*n+1* position, see Fig. 1) (9,10). In low pH solutions, deamidation via direct hydrolysis of the side-chain amide is the predominant reaction mechanism. The rate of the reaction varies widely with both intrinsic (primary sequence, presence of secondary and tertiary structure) and extrinsic (pH, temperature, buffer composition) factors.

[1] Genentech, Inc., 1 DNA Way, South San Francisco, CA 94010, USA.
[2] Amgen Inc., Thousand Oaks, California 91320, USA.
[3] Legacy BioDesign LLC, Johnstown, Colorado 80534, USA.
[4] To whom correspondence should be addressed. (e-mail: aakosky@gmail.com)

ABBREVIATIONS: 3loc, Flexibility parameter of the *n+1* residue; Asn, Asparagine; HX, Amide hydrogen exchange; ln(HX), Natural logarithm of the *n+1* residue amide proton half-lives; $\ln(t_{1/2})$, Natural logarithm of the deamidation half-lives; PC, Principle component; PCA, Principle component analysis; PLS, Projection to latent structures or partial least squares; RMSEP, Root mean squared error of prediction.

0724-8741/09/1100-2417/0 © 2009 Springer Science + Business Media, LLC

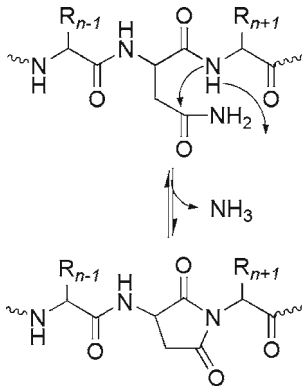

Fig. 1. The rate of base-catalyzed asparagine deamidation is affected by the properties of the neighboring residue side-chains in position 1 ($n-1$) and position 3 ($n+1$).

Among the intrinsic factors, the effects of neighboring amino acids have been widely studied. For instance, the size of the side chain in the *n+1* position has a dramatic effect on reaction rate, presumably by steric hindrance of intramolecular cyclization (2,11–14). For this reason, Asn residues followed by glycine, serine, and alanine have been shown to deamidate much faster than when proximal to bulkier amino acids.

Serine and histidine can accelerate Asn deamidation, apparently by hydrogen bonding to the carbonyl oxygen of the Asn side chain, making it more electrophilic, thereby increasing the rate of intramolecular nucleophilic attack (2). In addition, the acidity of the *n+1* amino acid has been implicated in controlling deamidation rates, as the proton must be removed to render the amide sufficiently nucleophilic (1,15,16). Amide hydrogen exchange and deamidation rates are similarly base-catalyzed, and the relationship between the rates of deamidation of Asn residues and the hydrogen exchange rate of the *n+1* residue has been examined for a small number of sequences (15,16). Those results suggest that the acidity (or its surrogate, amide exchange rate) of the *n+1* residue backbone amide impacts deamidation rate. Recently, Peters and Trout examined the elementary steps in the deamidation reaction of H_2N-Asn-NH-CH_3 using density functional theory (10). Their results indicate that deprotonation of the *n+1* residue lowers the deamidation reaction barrier significantly.

Another chemical property that is likely to affect deamidation rates is the flexibility of the peptide backbone, by permitting or hindering the adoption of the appropriate backbone geometry for intramolecular cyclization (17) (see Fig. 1). The retardation of deamidation rate within regular secondary (18,19) or tertiary (20) structures is a further indication that flexibility plays an important role. Moreover, it has been reported that flexibility affects acidity properties (21), where Gly residues have been found to be more acidic by being able to sample more of conformational space. Thus, both the flexibility and acidity of Gly residues facilitate deamidation.

Therefore, susceptibility to deamidation appears to depend upon a variety of physicochemical properties (amide proton acidity, chain flexibility, steric bulk of the side chains, hydrogen bonding capacity, etc.). Attempts to compose mathematical models focused on any one property at a time are inevitably deficient, because any substitution of one amino acid for another will alter more than one of the physicochemical properties. For example, even the most sophisticated electrostatic models appear to be unable to explain fully the effects of multiple mutations in systems such as the cold shock protein from *Bacillus* species (22–25). Any time multiple properties vary at once, chemometric analysis is warranted. This study seeks to provide a framework for quantitatively determining the relative importance of the physicochemical factors on the deamidation reaction. The PLS models described below provide the first multivariate approach for combining these important factors into a single scheme. This work also reflects the first use of multivariate statistical methods for assessing the stability propensity of peptides and proteins.

A number of quantitative property scales have been developed for amino acids. Based on a wide variety of physicochemical properties, principal component analysis has yielded a small number of 'reduced' properties that are able to reproduce adequately the original properties (26–28). These reduced properties have been used in quantitative structure-activity relationship (QSAR) studies for peptides (26,27). The deamidation of Asn residues makes an excellent subject for such an analysis, because the reaction rates are highly sequence-dependent, and a nearly comprehensive set of intrinsic rate values is available in literature. Robinson *et al*. have published deamidation half-lives for 306 pentapeptide sequence variants, Gly-X-Asn-Y-Gly, where X (*n−1*) and Y (*n+1*) were varied over the naturally-occurring amino acids (29–31).

An important advantage of using reduced properties in PLS models of the deamidation reaction is that multiple response variables can be considered simultaneously. For example, the biological activity of a subset of peptides could be included as a second response variable (32). From such a model, the reduced properties that most favored biological activity and while limiting deamidation could be derived and translated into a protein sequence with optimal activity and stability.

Three different reduced property scales were examined and used to construct projection to latent structures (PLS) models of the deamidation rates. The effects of flexibility and amide exchange rates were also explicitly considered. The PLS models presented here are relevant to the succinimide-mediated (base-catalyzed) route, which is the predominant deamidation reaction mechanism except in highly acidic media. This study represents the first example of using reduced amino acid properties to predict chemical instability in a polypeptide. The use of PLS allows the relative importance of the determinants of the reaction (flexibility, steric bulk, NH acidity, etc.) to be compared quantitatively.

METHODS

Chemometrics. There are many multivariate statistical methods available to identify and quantify correlations in large complex data sets. Projection to latent structures (PLS) is the most widely used chemometric method. PLS allows for the examination of correlations between a data set of properties (in this case, reduced properties describing the

amino acid substitutions) and a quantitative endpoint (e.g., the half-life for deamidation in a pentapeptide).

The descriptive variable matrix (called the X-matrix) was composed of mathematical representations of the peptide sequences with respect to the physicochemical properties of constituent amino acids. The descriptive variable matrix and the deamidation rates reported by Robinson and Robinson were subject to PLS modeling (29–31). The PLS models were analyzed to determine the accuracy, robustness and comprehensiveness of the predictive models. Regression coefficients were extracted from the models to determine which variables had a statistically significant influence on the response variable (e.g. deamidation half-life). Detailed descriptions of the independent steps are provided in the following sections.

Mathematical Representation of Pentapeptide Sequences. In order to perform multivariate analysis on the Asn deamidation-rate data of Robinson and Robinson, descriptor variables for each amino acid in the data set were identified (29–31). Each amino acid (AA) in the sequence was represented by a set of reduced properties. These reduced properties were generated using PCA, considering a wide range of AA properties, such as surface area, mass, volume, molecular weight, pKa(s), pI, aqueous solubility, hydrophobicity, polarizability, etc., thereby condensing a large number of descriptors for each AA to a few principal (reduced) properties (26,27).

Three different sets of physicochemical parameters were considered for their ability to predict Asn deamidation rates. The first set of reduced properties, named z-scores, was reported by Hellberg and coworkers, and included three values, where z1 is related to the hydrophilicity of the amino acid, z2 is related to its steric bulk, and z3 is related to its polarity and electronic properties (27). In subsequent work, a five-parameter set of z-scores was published (26). This extended set of reduced properties, referred to as zz-scores, provided a better fit to the measured properties, but the association of the zz4 and zz5 to a particular chemical or physical property was less clear. The zz4 scale is positively correlated with the heat of formation of the residue and negatively correlated with the electronegativity. The zz5 scale was positively correlated with electrophilicity (energy of the lowest unoccupied molecular orbital) and negatively correlated with the energy of the highest occupied molecular orbital and polarizability of the residue. A third set of parameters was considered, called PP values (28). As with the z-scores, there were three reported values for each amino acid. In general terms, PP1, PP2, and PP3 correspond to polarity, hydrophobicity, and hydrogen bonding capability, respectively. Physicochemical descriptions for each of the reduced property scales are summarized in Table I.

Interaction between each fundamental parameter at each position in the sequence is expected. Thus, all possible cross terms between each reduced property (i.e., z-, zz- or PP-score) within a given residue and between residues were considered explicitly. In the numbering schemes discussed below, the *n−1* residue is labeled position 1, the Asn residue is position 2, and the C-terminal residue is position *n+1*, and is labeled position 3. Since the Asn residue is constant, only values at positions *n−1* and *n+1* were considered in the mathematical models. Quadratic models were also constructed, i.e. squared terms for each reduced property were included.

Table I. Summary of the Reduced Property Scales Examined in the PLS Models (26,28)

Scale	Abbreviation	Physicochemical Properties
z-scores	z1	hydrophilicity
	z2	steric bulk, polarizability
	z3	polarity, electronegativity
zz-scores	zz1	hydrophilicity
	zz2	steric bulk, polarizability
	zz3	polarity, electronegativity
	zz4	heat of formation
	zz5	electrophilicity, hardness
PP scores	PP1	polarity
	PP2	size, hydrophobicity
	PP3	hydrogen bonding capacity (positive for H-bond acceptors, negative for H-bond donors)

The physicochemical properties that correlate positively and negatively with the scales are listed

Calculation of Amide Hydrogen Exchange Rates. One chemical property that is not included in determination of the z-, zz- or PP-scores is the intrinsic amide exchange rate. Since this property is correlated with acidity, the amide exchange rate may be an important predictor of deamidation rates. Therefore, this factor was explicitly considered. Amide hydrogen exchange rate constants of the *n+1* residues were calculated for each sequence according to the empirical equations of Englander and coworkers (33). The rate constants (k) were converted to half-lives in accordance with a first-order reaction at pH 7.5 to match the conditions and scaling of the deamidation data (i.e., $t_{1/2}=0.693/k$).

Flexibility Parameters. Another intrinsic chemical property that was included in some of the models was flexibility of the peptide chain. The proper orientation of certain moieties is essential for succinimide formation, the first step in Asn deamidation (9,17). Flexibility was measured using the location parameters developed by Smith *et al.* (34). They are based on the B-factors from numerous crystal structures of proteins.

PLS Modeling. The deamidation rates, shown in Table II, were obtained from the published values of Robinson *et al.* and corrected for acid-hydrolysis (29–31). All statistical modeling was performed with the Unscrambler® software (CAMO, Corvallis OR). Detailed descriptions of PLS have been published elsewhere (35–37). In this study, PLS was used to quantify the effect of an assortment of descriptive variables (i.e. residue-specific reduced properties, flexibility parameters, and hydrogen exchange rates) of each peptide on the rate of asparagine deamidation. A matrix of the descriptive variables was assembled, where each row of the matrix contained the descriptive variables for one of the sequences based solely on its AA composition (306 rows). This collection of sequence-specific variables formed the X-matrix; the deamidation rate data comprised the dependent variable (Y-variable). Mathematical models can be constructed that explain the greatest amount of variance in the dependent variable(s) of interest (the Y-variable or Y-matrix). The best single description of the relationship

Table II. Half-Lives in Days for Asparagine Deamidation in Gly-Xxx-Asn-Yyy-Gly Pentapeptides Incubated at 37°C (adapted from Robinson and Robinson (4,30)

Yyy	Gly	His	Ser	Ala	Asp	Thr	Cys	Lys	Met	Glu	Arg	Phe	Tyr	Trp	Leu	Val	Ile	Mean	Median
Xxx																			
Gly	1.0	9.2	11.8	21.2	28.1	40.0	40.8	48.5	50.7	74.6	58.2	64.5	64.1	77.8	105.4	230.4	297.6	72.0	54.5
Ser	1.0	8.3	15.1	24.2	30.4	46.0	60.7	55.9	55.3	60.1	60.1	52.5	65.2	77.5	111.5	239.9	295.4	74.1	58.0
Thr	1.0	9.6	17.1	24.7	28.0	50.3	55.9	58.0	47.9	61.3	51.5	77.1	81.4	73.2	111.5	244.2	289.0	75.4	56.9
Cys	1.1	10.8	19.0	26.5	30.7	49.0	46.3	46.9	65.0	48.6	84.0	74.6	84.8	112.6	120.8	235.7	315.9	80.7	57.0
Met	1.0	10.2	15.2	22.2	26.5	43.8	49.9	60.9	57.3	73.1	59.2	62.4	74.7	93.8	114.6	216.7	284.7	74.5	60.0
Phe	1.2	10.2	18.1	24.3	27.5	39.2	46.8	58.6	59.0	62.9	61.9	70.1	75.8	103.3	119.8	208.2	297.6	75.5	60.3
Tyr	1.5	10.2	11.9	24.4	28.5	38.3	48.9	55.5	64.8	41.2	57.3	58.4	71.2	121.8	119.8	248.4	318.1	77.7	56.4
Asp	1.5	9.7	17.0	24.1	29.5	52.7	54.5	76.6	57.7	47.1	88.2	70.9	71.0	81.1	112.6	248.4	309.4	79.5	64.2
Glu	1.5	9.0	16.4	25.9	32.1	37.0	44.4	78.6	60.0	60.8	81.7	70.8	95.6	99.6	132.1	277.2	289.0	83.0	65.8
His	1.1	10.7	15.7	24.7	31.3	47.5	44.1	50.5	63.6	70.0	49.2	72.8	93.1	96.5	117.7	254.8	340.8	80.8	57.1
Lys	1.0	10.5	15.6	23.7	24.1	58.5	49.3	53.9	61.4	73.2	57.8	70.7	97.9	99.3	120.8	253.7	325.6	82.8	59.9
Arg	1.0	10.0	14.3	24.5	34.9	51.0	50.8	49.9	75.1	68.9	68.0	68.9	91.0	129.0	130.1	254.8	323.5	85.0	68.4
Ala	1.1	9.3	14.9	22.6	32.0	43.7	64.2	56.3	59.6	74.8	62.9	66.1	74.6	132.1	125.9	262.3	311.6	83.2	63.5
Leu	1.1	10.7	16.7	25.2	32.2	46.4	53.9	60.6	63.1	57.1	62.6	73.1	76.4	75.2	158.0	305.1	410.9	89.9	61.6
Val	1.2	10.2	18.2	27.6	33.6	50.2	63.7	64.3	66.2	65.3	68.0	67.2	80.0	89.9	157.0	301.9	383.4	91.1	65.8
Ile	1.3	11.5	14.5	26.0	33.9	46.6	53.0	64.9	59.2	59.0	67.0	62.0	80.1	87.6	157.0	306.2	403.2	90.2	60.6
Trp	1.8	11.3	15.5	30.8	43.8	37.1	84.0	59.8	64.7	76.4	74.6	71.7	93.7	137.3	135.2	232.5	296.5	86.4	73.2
Pro	1.2	12.8	18.9	31.9	48.9	63.6	60.5	68.4	79.2	93.1	73.6	101.3	115.6	123.9	185.2	381.2	482.2	114.2	76.4
Mean	1.2	10.2	15.9	25.2	32.6	46.8	54.0	59.3	61.7	64.9	65.9	69.7	82.0	100.6	129.7	261.2	331.9	83.1	63.3
Median	1.1	10.2	15.7	24.7	32.0	46.6	53.0	58.6	61.4	64.9	62.9	70.1	80.1	99.3	120.8	253.7	315.9	82.8	62.1

between the variation in the X-matrix and the endpoint (the Y-matrix) is called the first principal component, PC1. The second most important (in terms of describing the variance in the Y-matrix) component is called the second principal component, PC2, and so on. Each of the PCs contains some contribution from each of the variables in the X-matrix.

The final model is composed of a number of PCs that together provide an adequate description of the Y-matrix. Regression coefficients can be calculated for each variable in the X-matrix. If a variable within the X-matrix contributes heavily to the construction of a given PC, then it is ranked as being significant. In summary, PLS takes information from the X-matrix, calculates the desired number of PCs, and constructs a suitable model. The model that includes all of the samples is termed a calibration model. The overall coefficient of determination (r) indicates the quality of the model.

Due to the diversity of scales, all of the variables were normalized by dividing by the standard deviation using this weighting option in the Unscrambler® software. This gives each X-variable an equal opportunity to influence the model. The jackknife algorithm was used for all PLS models to determine the variables in the X-matrix that were statistically significant with respect to a 95% confidence interval, and Hotelling T2 ellipses were used to demarcate outliers in PLS score plots (38,39).

The models were validated using a full cross validation protocol, which provides a rigorous description of the validity of a PLS model (37,38). Using the full cross validation method, one sample at a time was excluded from the calibration data set, and the model was calibrated using the remaining samples (37,40,41). The value for the excluded sample was then predicted, and the residuals were computed to assess the validity of the model. Regression coefficients are reported for both the original calibration set and for the validation set (generated by full cross validation).

RESULTS

PLS Regression of Condensed Sequence Properties and Deamidation Rates. The initial PLS regression models were constructed using the original deamidation half-lives (Table II) as the endpoint or Y-variable. Approximately 45% of the Y-variance (i.e., the variance in the deamidation half-lives) was explained by a PLS model with the z-scores for the amino acids in positions 1 and 3 (*n−1* and *n+1*) as the X-matrix (Table III). Inclusion of all possible cross terms plus quadratic terms (the squares of the z-scores at each position) produced PLS models that explained 74% of the variance in the deamidation half-lives. Excluding the z-scores for the *n−1* residue slightly increased the amount of the described variance in the deamidation half-lives.

Linearization of Deamidation Rates. The half-lives for asparagine deamidation occur over a wide range, covering more than two orders of magnitude (Table II). Care must be taken when modeling broadly ranged data, because values at either end of the range can dominate curve-fitting. Since PLS employs linear principal components and extreme values in the response variable can have a magnified impact on the quality of the model, the data must be compressed in order to obtain satisfactory mathematical models. For this study, all factors were weighted by the reciprocal of the standard deviation ($1/\sigma$), a type of standardization of variables that is common in PLS modeling (37,40,42). However, the $1/\sigma$ normalization has a limited effect when the range is much wider than the standard deviation or the relationship between the response and predictive variables is non-linear. In order to address this issue, two different approaches were evaluated for linearization of the deamidation half-lives, namely using the square root ($(t_{1/2})^{1/2}$) and the natural logarithm ($\ln(t_{1/2})$) of the half-life (43). Such linearization is often used in

Table III. Results of PLS modeling Using z-Scores to Fit the Asparagine Deamidation Half-Life ($t_{1/2}$) Values

Positions	Interaction Terms	Other factors	r (calibration)	r (validation)	Captured Y-variance	# of PCs	RMSEP (days)
1 & 3	none	none	0.673	0.655	45%	1	63.0
1 & 3	interaction	none	0.808	0.783	65%	2	51.9
1 & 3	square	none	0.735	0.711	54%	2	58.7
1 & 3	interaction & square	none	0.860	0.833	74%	4	46.1
3	interaction & square	none	0.867	0.841	75%	5	45.0
3	interaction & square	ln HX	0.940	0.936	88%	4	29.2
3	interaction & square	flex	0.919	0.909	84%	8	34.7
3	interaction & square	ln HX & flex	0.950	0.947	90%	5	26.6

The correlation coefficients for the calibration and validation models (r) are shown with the percentage of the variance in the deamidation half-life captured by the model, the optimal number of PCs, and the root mean square error of prediction after cross validation (RMSEP).

chemometric studies to allow proper fitting of widely ranging variables and non-linear relationships (42,44–47).

The models that were calculated with $\ln(t_{1/2})$ resulted in significantly less curvature than when calculated with the raw half-life data or the square root scaling, independent of which reduced property scale was used (Fig. 2). Therefore, the remainder of the PLS models presented in this study employ $\ln(t_{1/2})$ as the Y-matrix.

Comparison of Reduced Property Scales. PLS regression results of $\ln(t_{1/2})$ with the three different property scales for the *n+1* amino acid are compared in Fig. 3. The PLS model using the zz-scores captured more of the Y-variance and displayed lowest root mean square error of prediction, while the z-scores were better at predicting the deamidation half-life than the PP-scores (Fig. 3 and Table IV). Note that the zz-score models use five factors per amino acid compared to three for the z- and PP-score reduced properties.

Inclusion of Flexibility and Hydrogen Exchange Parameters. The addition of peptide flexibility parameters and/or the amide proton exchange half-lives (used as the natural logarithm of the exchange rate, denoted as ln(HX)) improves the quality of all of the models (Tables III and IV). PLS models including both ln(HX) and the flexibility parameters capture >94% of the variance in the Y-matrix and markedly reduce the root mean square errors of prediction for each reduced property scale (Table IV). When included along with the z-scores, ln(HX) becomes the most influential variable in the X-matrix (Table V). However, ln(HX) has only a modest effect on the zz-score models, because the five-component zz-scores already capture more than 95% of the Y-variance (the natural logarithm of the deamidation half-lives, $\ln(t_{1/2})$).

Performance of z-Score Models. Regression coefficients for five different PLS models of $\ln(t_{1/2})$ are listed in Table V. Only the parameters that were found to be statistically significant are presented (within the 95% confidence interval from jackknife analysis). In all of the PLS models, the deamidation half-life is primarily governed by the reduced properties in position 3 (*n+1*). The amino acid in position 1 (*n−1*) has virtually no effect.

All three z-scores for the *n+1* residue are statistically significant in models that describe more than 90% of the Y-variance (Models 2–5, Table V). The 3z1 variable, which reflects the hydrophilicity of the *n+1* residue, has the largest regression coefficient and is negatively correlated with deamidation half-life, that is, longer deamidation half-lives arise from decreased hydrophilicity of the *n+1* amino acid. Similarly, 3z3 (describing electronic properties or polarity) is negatively correlated with $\ln(t_{1/2})$. The 3z2 variable is positively correlated with the deamidation half-life, $\ln(t_{1/2})$. In other words, the deamidation reaction was more rapid when the bulk of the *n+1* residue side-chain was small. The squared terms of 3z2 and 3z3 and interaction terms involving 3z2 were significant in every model (Table V).

Response surface plots were created to visualize the interrelationship between the predictive variables (Fig. 4). The contours in these plots represent the response variable, $\ln(t_{1/2})$, which is color-coded according to the horizontal bar at the bottom of the figure. In the top panel, a singular well-defined region of maximal deamidation half-life corresponds to negative values of 3z1 (hydrophilicity) and positive values of 3z2 (size). However, the response of $\ln(t_{1/2})$ to 3z3 (middle and bottom panels, Fig. 4) indicates that the longest deamidation half-lives correspond to values at both the negative and positive extremes of 3z3 (polarity and electro-negativity). The shape of this response surface indicates that there is significant non-linearity in the relationship between ln (t1/2) and 3z3 and reveals how the squared-terms are significant to the models.

The natural logarithm of the amide proton exchange half-lives were significantly positively correlated with the deamidation half-lives. The magnitude of the 3z1 regression coefficient was decreased, but the magnitudes of the $3z1^2$, $3z2^2$ and $3z3^2$ coefficients were increased when ln(HX) was included in the model. In addition, ln(HX)*3z1 and ln(HX)*3z3 are statistically significant. These results suggest that the increased Y-variance explained by the model (97% in Model 2 relative to 81% in Model 1, Table V) is due to amide proton exchange rate for the nitrogen immediately after the Asn residue (i.e., in position 3), a property that is not contained in the z-scores.

Inclusion of the flexibility parameter also influences the magnitudes of the regression coefficients for the z-scores (Table V). Although 3z1, 3z2 and 3z3 were only slightly affected, the flexibility parameter of the third residue, and the interaction terms between the flexibility and z-scores of the third residue (3z1 and 3z3) are statistically significant. Like

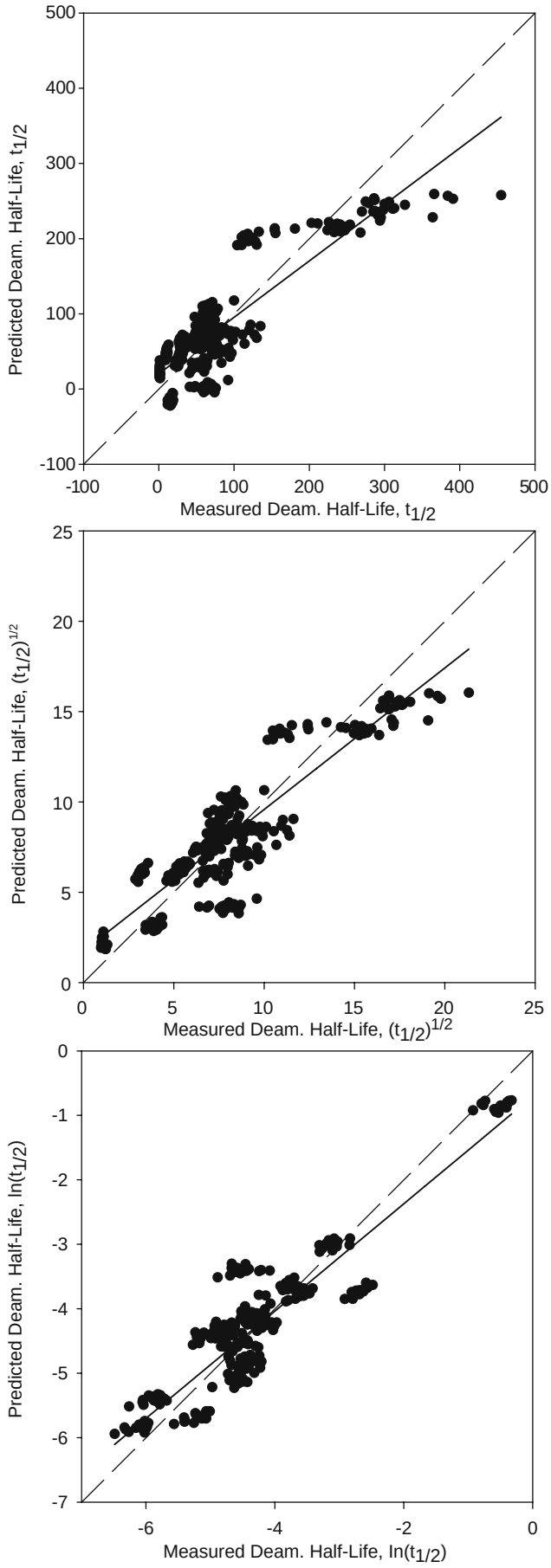

Fig. 2. Effect of different linearization methods for the asparagine deamidation half-life values in quadratic PLS models with z-scores and all interaction terms. The top panel shows the PLS model without linearization of the deamidation half-lives; the PLS models calculated with the square root and natural logarithm linearizations are shown in the middle and bottom panels, respectively.

the amide exchange rate, flexibility of the third residue is another property that is not captured in the z-scores.

When ln(HX) is included along with flexibility parameters and z-scores, the regression coefficient for the flexibility parameter becomes insignificant (Model 4, Table V). The interaction terms of the flexibility parameter with 3z1 and 3z3 are still important to the model. These results suggest that flexibility is less important than the acidity of the succeeding amide proton, although both influence deamidation rates.

Performance of zz-Score Models. Each of the zz-scores contributed statistically significant regression coefficients in the PLS model of $\ln(t_{1/2})$ (Model 5, Table V). The most important parameters in the model were 3zz3 (polarity), 3zz1*3zz2 (product of hydrophilicity and steric bulk), 3zz2*3zz3, 3zz1 and 3zz2*3zz5 (product of steric bulk and electronegativity). Given that z1, z2 and z3 are analogous to zz1, zz2, zz3, the increased Y-variance explained in Model 5 (with respect to Model 1) can be attributed to the contributions of zz4 and zz5. Both 3zz4 (heat of formation) and 3zz5 are negatively correlated with $\ln(t_{1/2})$. As with the z-score models, the properties of the amino acid in position 1 (*n–1*) had little influence on the PLS models, and exclusion of the z-scores for the *n–1* amino acid had minor effects on the amount of Y-variance explained (Tables III & IV).

Performance of PP-Score Models. Regression coefficients for four PLS models of $\ln(t_{1/2})$ are listed in Table VI; only the coefficients that were found to be statistically significant are presented. The PP-scores terms representing the size, hydrophobicity (3PP2) and hydrogen bonding capacity (3PP3) of the *n+1* residue had the most influence on the models. Of the PP-scores, 3PP1 (polarity) had the least influence on the models.

In Model 1 of Table VI, $3PP1^2$ and 3PP3 had the largest regression coefficients, suggesting that the rate of deamidation was positively correlated with the polarity and hydrogen-bond-accepting capability of the *n+1* amino acid. When ln(HX) was included in the calculation (Model 2 Table VI), neither 3PP1 nor 3PP3 alone was statistically significant, but 3PP2 was negatively correlated with $\ln(t_{1/2})$. Each of the PP-scale interaction terms with ln(HX) significantly influenced the model, and 3PP2*ln(HX) had the largest (negative) regression coefficient. The model indicates that deamidation half-lives were shorter when the size/hydrophobicity and amide exchange rate of the *n+1* residue decreased. The size, hydrophobicity and hydrogen bonding capability information provided by 3PP2 and 3PP3 were largely captured by hydrogen exchange half-life term (ln (HX)).

When the flexibility parameters were included in the calculation (Model 3 of Table VI), the Y-variance (in $\ln(t_{1/2})$) explained by the model increased to ~97%. The model only produced three statistically significant X-variable regression

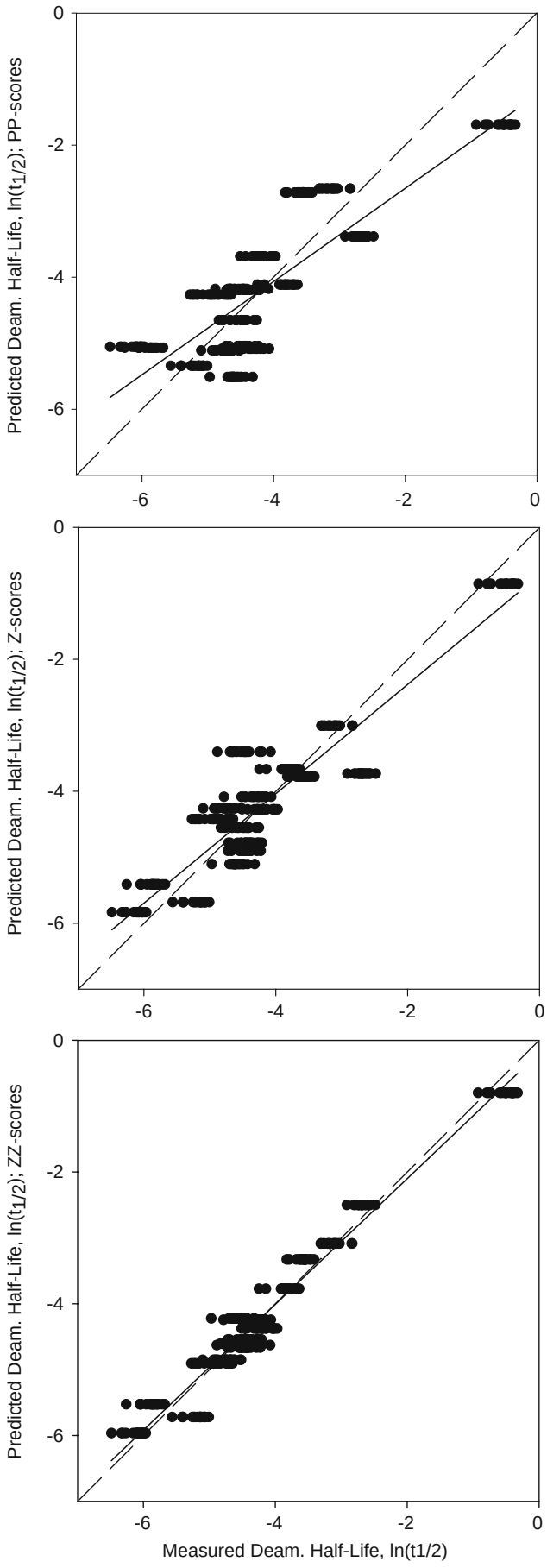

Fig. 3. Comparison of PLS models using different reduced property scales to fit the natural logarithm of the asparagine deamidation half-lives. The top panel is the model produced with the zz-scores; the middle and bottom panels are the z-scores and PP-scores, respectively. Each model was calculated with the reduced properties for the *n*+1 residue only.

coefficients, i.e. 3loc*3PP1, 3PP1*3PP2, and $3PP3^2$. These three variables accounted for approximately 74% of the explained variance in $\ln(t_{1/2})$. The regression coefficient for 3loc*3PP1 was negatively correlated with ln(t1/2) indicating that deamidation half-lives decrease with the flexibility and polarity of the *n+1* residue. The regression coefficient for $3PP3^2$ was large and positive, again indicating that the deamidation half-lives were longer when the *n*+1 residue was a hydrogen bond donor.

When the flexibility and hydrogen exchange parameters were simultaneously included in the calculations, the percent Y-variance explained by the model increased slightly and a larger number of X-variables had significant regression coefficients (Model 4 Table VI). The most important parameters were 3loc*3PP1, $3PP1^2$, ln(HX)*3PP1, and ln(HX)*3loc. As with the z-score models, the deamidation half-lives were positively correlated with decreasing polarity and decreasing flexibility of the *n+1* residue.

DISCUSSION

Comparison of Reduced Property Scales. Three different reduced property scales were considered in this study: z-scores, zz-scores, and PP-scores. Using these reduced properties alone, the PLS models developed here capture as much as 90% of the variance in the deamidation half-life, as described in terms of physicochemical properties of the flanking amino acids (Table III). The captured variance is even greater when one of the linearized forms of the deamidation half-life is used as the Y-variable (see below). The zz-scores, by themselves, can account for 90–97% of the variance in the scaled deamidation half-lives. For the more limited, three-component z- and PP-scores, the z-scores proved to be more effective at predicting chemical reactivity, even though they were never constructed to do so. Together, these results indicate that reduced properties are capable of accurate prediction of peptide chemical reactivity.

Linearization of the Deamidation Half-Life. Since there was clearly some non-linear behavior in the deamidation half-lives, two different schemes (the natural logarithm and the square root) were evaluated for linearizing the deamidation data. This is a common approach in chemometric modeling for data that extends over orders of magnitude (37,42,44–47). The natural logarithm of the deamidation half-lives provided a more effective linearization than the square root of the half-lives (Fig. 3 and compare Tables V and VI). The increased linearity in the natural logarithm normalization can be rationalized by the fact that each variable is then proportional to the corresponding reaction activation energy. Models of deamidation produced by Robinson & Robinson and Capasso (separately) employed logarithmic scales to take

Table IV. Results of PLS Modeling Using z-Scores and All Squared and Interaction Terms to Fit the Natural Logarithm of the Asparagine Deamidation Half-Life ($t_{1/2}$) Values

Scale	Positions	Other factors	r (calibration)	r (validation)	Captured Y-variance	# of PCs	RMSEP (days)
z-scores	1 & 3	None	0.912	0.898	81%	5	1.7
z-scores	3 only	None	0.910	0.905	82%	5	1.7
z-scores	1 & 3	flex	0.966	0.952	91%	7	1.5
z-scores	3 only	flex	0.962	0.959	92%	6	1.4
z-scores	1& 3	ln HX	0.990	0.987	96%	9	1.2
z-scores	3 only	ln HX	0.988	0.987	97%	8	1.2
z-scores	3 only	ln HX & flex	0.988	0.988	98%	7	1.2
PP-scores	3 only	None	0.840	0.829	70%	8	2.0
PP-scores	3 only	ln HX	0.942	0.942	89%	7	1.5
PP-scores	3 only	flex	0.985	0.985	97%	11	1.2
zz-scores	3 only	None	0.977	0.974	96%	5	1.3
zz-scores	1 & 3	None	0.980	0.971	97%	6	1.3
zz-scores	1 & 3	ln HX	0.991	0.987	97%	6	1.2
zz-scores	1 & 3	ln HX & flex	0.990	0.985	97%	6	1.2

The correlation coefficients for the calibration and validation models are presented along with the percent of variance in the deamidation half-life captured by the validated model and the root mean square error of prediction after cross validation (RMSEP).

advantage of the additivity of substituent effects and proportionality with the activation energy (30,48).

Interaction, Squared and Sequence Properties. Inclusion of interaction parameters was essential, especially between the reduced properties describing the residue in position *n+1*. Inclusion of the sequence properties, flexibility and/or amide proton exchange half-lives, improved the models and decreased the error of prediction (RMSEP). Similarly, inclusion of the squared terms for the reduced and sequence properties improved the accuracy of all of the models. This result indicated that the relationship between the predictive and response variables was non-linear even after linearization of the deamidation half-lives (e.g. $(t_{1/2})^{1/2}$ or $\ln(t_{1/2})$).

Effect of the n−1 Residue. The PLS regressions performed here also indicate that none of the reduced properties associated with residue *n−1* are significant predictors of the deamidation half-lives (Table V). Terms associated with the residue at the *n−1* position could be excluded from the calculations without decreasing the accuracy of the mathematical models (Tables III, IV, and V). Robinson and Robinson, the originators of the deamidation data used in this study, note that the *n−1* residue effects are relatively small and subject to

Table V. Regression Coefficients for PLS Models of the Natural Logarithm of the Deamidation Half-Lives with z-Scores or zz-Scores

Model 1		Model 2		Model 3		Model 4		Model 5	
z-scores		z-scores & lnHX		z-scores & flex		z-scores, lnHX & flex		zz-scores	
Y-variance explained: 81%		Y-variance explained: 97%		Y-variance explained: 95%		Y-variance explained: 95%		Y-variance explained: 95%	
R (val.)=*0.898*		R (val.)=*0.987*		R (val.)=*0.986*		R (val.)=*0.986*		R (val.)=*0.974*	
3Z1	−0.642	ln HX	0.588	3Z1	−0.706	ln HX	0.375	3ZZ1	−0.386
3Z3	−0.386	3Z1	−0.439	3Z2	0.102	3Z1	−0.380	3ZZ2	0.200
3Z1*3Z2	0.223	3Z2	0.159	3Z3	−0.416	3Z2	0.177	3ZZ3	−0.497
3Z2*3Z3	0.352	3Z3	−0.310	3loc	−0.229	3Z3	−0.302	3ZZ4	−0.270
$3Z2^2$	−0.206	3Z1*3Z2	−0.433	3Z1*3Z2	0.130	ln HX*3Z2	−0.247	3ZZ5	−0.198
$3Z3^2$	0.249	ln HX*3Z1	0.181	3Z1*3loc	0.431	ln HX*3Z3	−0.075	3ZZ1*3ZZ2	0.442
		ln HX*3Z3	0.446	3Z2*3Z3	0.385	3Z1*3Z3	0.228	3ZZ1*3ZZ3	−0.110
		$(\ln HX)^2$	0.586	3Z3*3loc	0.403	3Z1*3loc	0.295	3ZZ1*3ZZ4	−0.191
		$3Z1^2$	−0.368	$3Z1^2$	−0.402	3Z3*3loc	0.253	3ZZ2*3ZZ3	0.400
		$3Z2^2$	−0.408	$3Z2^2$	−0.173	$(\ln HX)^2$	0.254	3ZZ2*3ZZ5	−0.370
		$3Z3^2$	0.374	$3Z3^2$	0.446	$3Z2^2$	−0.244	3ZZ3*3ZZ4	0.228
				$3loc^2$	0.195	$3Z3^2$	0.248	3ZZ3*3ZZ5	−0.129
								3ZZ4*3ZZ5	−0.111
								$3ZZ2^2$	−0.326
								$3ZZ3^2$	0.339

Only parameters that were found to be statistically significant are shown. The flexibility parameters of the first and third residues are indicated by 1loc and 3loc, respectively.

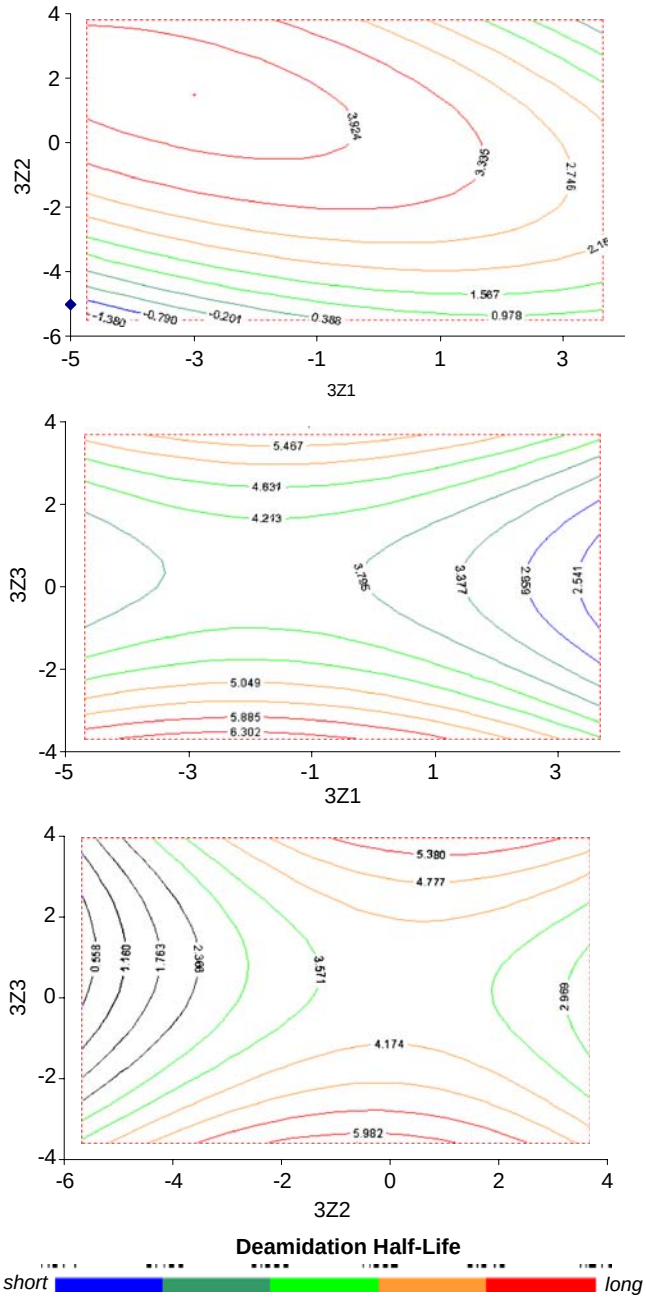

Fig. 4. Response surfaces for PLS model including z-scores, all interaction terms, and ln(HX) as the X-matrix. The response variable was $\ln(t_{1/2})$; values of $\ln(t_{1/2})$ are color-coded according to the horizontal bar in each plot.

uncertainty inherent to the experimental system (including buffer effects) (30).

Physicochemical Determinants of Deamidation. Each of the reduced property scores for the *n+1* residue appears to be a statistically significant predictor of the deamidation half-lives (Table I). When using z-scores, the most important reduced property is the hydrophilicity for the *n+1* residue (3z1), as indicated by the regression coefficients obtained from different models (Table V). The influence of polarity/electronic properties (3z3) and size/bulk (3z2) properties on the deamidation rates was also important, as seen in the squared and interaction terms (Model 2, Table V). An analogous interpretation arises from the models employing the PP-scores, where parameters for polarity (3PP1), hydrogen bonding capacity (3PP3) and size/hydrophobicity (3PP2) scores each contributed significantly to the models.

Regardless of the property scale employed, the physicochemical determinants of deamidation were observed to be highly interrelated and to respond to the deamidation half-lives nonlinearly. In fact, squared terms of the reduced properties were significant in every model. The response surface diagrams in Fig. 4 provide a good visualization of this effect. The area describing the longest deamidation half-lives (shown in red and orange) transverse nearly the entire range of hydrophilicity (3z1) of the *n+1* residue, and for any given value of 3z1, the deamidation half-life response to the polarity of the *n+1* residue appears parabolic.

Robinson and Robinson developed a primary sequence model of asparagine deamidation based on the steric and catalytic effects of the *n+1* residue (30). In their model, the probability that the *n+1* residue prevents the asparagine side chain from adopting reactive conformations (i.e., leading to the formation of the succinimide intermediate) was effectively accounted for by a model that included the steric bulk of the side chain. In the PLS models, 3z2 (the property most closely related to the size of the *n+1* residue side chain) was, as expected, consistently positively correlated with deamidation half-life (Table IV). Thus, the reduced property, z2, provides a meaningful description of the steric effects of *n+1* residue side chains that affect succinimide formation. The squared term of 3z2 was significantly negatively correlated to the natural logarithm of the deamidation half-lives in each of the PLS models, indicating that the deamidation half-life depends on 3z2 in a non-linear fashion.

Inspection of Table II confirms that the mean deamidation half-lives do not increase predictably with increased mass of the *n+1* residue. A number of relatively heavy and/or charged amino acids (e.g. histidine, lysine) have shorter half-lives than some relatively small amino acids (e.g. valine). This may explain the relatively small regression coefficients for the 3z2 variable. The model of Robinson and Robinson accounts for these effects by adjusting for steric interference of the deamidation reaction that is due to side-chain branching (e.g. β-carbon) of the *n+1* residues and catalytic effects specific to particular residues. The z2 variable may not fully account for the effects of side chain branching and is not expected to account for catalytic effects of charge or polar amino acids. Analogously, the interaction terms of z2 with z1 and z3 are more powerful predictors of deamidation half-life than the 3z2 variable. For example, in Model 1 of Table V, 3z2 is insignificant, but the 3z1*3z2 and 3z2*3z3 are significant predictors of the natural logarithm of the deamidation half-lives. A literal interpretation of this result is that the size/bulk of the *n+1* residue side chain is positively correlated with deamidation half-life when the polarity and hydrophilicity are simultaneously taken into account. This result highlights the importance of the interaction terms in generating and interpreting PLS models of chemical reactivity.

The increased hydrophilicity of the *n+1* residue may influence the deamidation half-life by affecting the acidity of the amide proton. In PLS models of the amide proton

Table VI. Regression Coefficients for PLS Models of the Natural Logarithm of the Deamidation Half-Lives With 3PP-Scores

Model 1		Model 2		Model 3		Model 4	
3PP-scores		3PP-scores & lnHX		3PP-scores & flex		3PP-scores, lnHX & flex	
Y-variance explained: 70%		Y-variance explained: 89%		Y-variance explained: 97%		Y-variance explained: 98%	
R (val.)=*0.829*		R (val.)=*0.942*		R (val.)=*0.985*		R (val.)=*0.991*	
3PP2	−0.401	ln HX	0.383	3loc*3PP1	−2.08	ln HX	−0.164
3PP3	−1.748	3PP2	0.180	3PP1*3PP2	−0.459	3loc	−0.633
3PP1*3PP2	0.495	ln HX*3PP1	−0.299	$3PP3^2$	1.839	3PP2	0.494
3PP1*3PP3	1.054	ln HX*3PP2	−0.905			3PP1*3PP2	0.377
3PP2*3PP3	0.755	ln HX*3PP3	0.354			3PP1*3PP3	−0.625
$3PP1^2$	−1.494	3PP1*3PP3	−0.213			3PP2*3PP3	0.308
$3PP2^2$	−0.191	$ln\ HX^2$	0.427			ln HX*3PP1	−1.436
$3PP3^2$	0.181					3loc*3PP1	2.094
						$3loc^2$	−0.160
						$3PP1^2$	−1.495
						$3PP3^2$	−0.572
						3loc*ln HX	0.678
						$Ln\ HX^2$	−0.279

Only parameters that were found to be statistically significant are shown. The flexibility parameters of the first and third residues are indicated by 1loc and 3loc, respectively.

exchange half-lives (not shown), the electronic properties and polarities of the *n+1* residue were found to contribute to the exchange rate as well as the rate of deamidation. Thus, the polarity of the *n+1* residue can affect shortening of the deamidation half-lives by increasing amide exchange rate of that residue. If *n+1* amide deprotonation and formation of the succinimide-intermediate are rate-determining steps in the deamidation reaction, then any property that contributes to the electron-withdrawing nature of the amide nitrogen will facilitate the reaction.

Deprotonation of the backbone amide nitrogen of the *n+1* residue has been hypothesized to be the rate-limiting step and thus a principal factor in governing Asn deamidation rates. Brennan and Clarke showed that unit increases in the logarithm of deamidation rate roughly correlated to unit increases in the logarithm of hydrogen exchange rate for a set of ten dipeptides (15). A similar relationship is seen in the analysis of 306 amino acid sequences used in this study (not shown). The slope of linear relationship between the logarithm of deamidation and hydrogen exchange rates was found to be 0.95 (not shown). The magnitude of observed scatter was largest in peptides with the fastest deamidation rates and with charged residues in the *n+1* position (Gly, Cys, His, Arg, Asp). In *ab initio* studies of the deamidation reaction, Radkiewicz and coworkers found that the deprotonated amide of Asn-Gly sequences was stabilized by the high conformational flexibility of the glycine residue relative to other *n+1* position amino acids (16). This effect would promote the rapid deamidation observed in Asn-Gly sequences, indicating why flexibility and acidity are both influential in controlling Asn deamidation rates.

Taken together, the results presented here with the reports of Robinson *et al.* and Peters *et al.* suggest that the rate-determining steps to the deamidation reaction vary with the identity *n+1* residue (10,30). For some sequences, the deprotonation of the *n+1* residue may reduce the initial energy barrier enough to allow the formation of the tetrahedral intermediate to occur more readily than the loss of ammonia from the tetrahedral intermediate. For other sequences, the amplitude of the activation energy may be set by other properties of the *n+1* residue side chain (e.g. size/bulk), and the activation energy for formation of the tetrahedral intermediate may remain higher than that for the formation of the succinimide regardless of protonation state of *n+1* residue. In either case, the deprotonation of the *n+1* residue amide is presumed to facilitate the reaction by increasing the energy of the reactant sequence and decreasing the relevant transition state energy (10).

Flexibility at residue *n+1* was also found to be important. As mentioned above, there is a steric component to deamidation that proceeds through intramolecular cyclization. Likewise, an optimal set of dihedral angles exists for the peptide backbone to engage in nucleophilic attack of the Asn side chain carbonyl group, as outlined by Clarke (17). The location parameters used here appear to be accurate descriptors of peptide backbone flexibility (34). Even though they were developed based on crystal structures of proteins, they do seem to predict accurately the relative flexibility of the peptide that would allow facile cyclization and subsequent hydrolysis to the deamidated products. Since flexibility of the polypeptide backbone can vary widely in folded proteins, this methodology, although based on flexible peptides, could provide useful predictions for larger globular proteins.

CONCLUSIONS

This study demonstrates that reduced amino acid properties are highly accurate predictors of chemical reactivity in peptides. The z-scores produced the most accurate models of

the deamidation rates, especially if the five-component zz-scores were employed. The most accurate models required linearization of the deamidation half-lives, where the natural logarithm provides the most effective linearization. The inclusion of interaction (i.e., cross) terms and squared terms was also essential for a robust and accurate PLS model.

The reactivity of an asparagine residue is dominated by the properties of the *n+1* residue; the properties of the *n−1* residue were negligible in PLS models of the deamidation reaction rates. The properties that control the rate of deamidation are closely related to the properties that control the rates of amide proton exchange of the *n+1* residue. The extent to which the protonation state of the *n+1* residue governs the rate of deamidation varies with the properties of the *n+1* residue side chain. The reaction rate is also governed by the hydrophilicity, size, polarity, and flexibility of the *n+1* amino acid. While most of these properties have been implicated in controlling deamidation rates, there has been no methodology to quantify or rank order the relative importance of each factor until this study.

The PLS analyses presented here exemplify an important application of multivariate statistical methods. Accurate prediction of chemical reactivity is a valuable asset in developing stable peptides and proteins. This approach requires no *a priori* knowledge of the reaction mechanism, only the primary sequence.

ACKNOWLEDGMENTS

The authors thank Amgen Incorporated for sponsoring this project and Art and Noah Robinson for permission to reproduce their table of deamidation half-times.

REFERENCES

1. Manning MC, Patel K, Borchardt RT. Stability of protein pharmaceuticals. Pharm Res. 1989;6:903–18.
2. Aswad DW. Deamidation and Isoaspartate Formation in Peptides and Protein. In: Aswad D, editor. CRC series in analytical biotechnology. USA: CRC; 1995. p. 1–259.
3. Robinson AB, McKerrow JH, Cary P. Controlled deamidation of peptides and proteins: an experimental hazard and a possible biological timer. Proc Natl Acad Sci U S A. 1970;66:753–7.
4. Robinson NE, Robinson AB. Molecular clocks. Proc Natl Acad Sci U S A. 2001;98:944–9.
5. Wakankar AA, Borchardt RT. Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization. J Pharm Sci. 2006;95:2321–36.
6. Jenkins N. Modifications of therapeutic proteins: challenges and prospects. Cytotechnology. 2007;53(1–3):121–5.
7. Chelius D, Rehder DS, Bondarenko PV. Identification and characterization of deamidation sites in the conserved regions of human immunoglobulin gamma antibodies. Anal Chem. 2005;77:6004–11.
8. Ren D, Ratnaswamy G, Beierle J, Treuheit MJ, Brems DN, Bondarenko PV. Degradation products analysis of an Fc fusion protein using LC/MS methods. Int J Biol Macromol. 2009;44:81–5.
9. Li B, Borchardt RT, Topp EM, Vander Velde D, Schowen RL. Racemization of an asparagine residue during peptide deamidation. J Am Chem Soc. 2003;125:11486–7.
10. Peters B, Trout BL. Asparagine deamidation: pH-dependent mechanism from density functional theory. Biochemistry. 2006;45:5394–2.
11. Wright HT. Sequence and structure determinants of the non-enzymatic deamidation of asparagine and glutamine residues in proteins. Protein Eng. 1991;4:283–94.
12. Stephenson RC, Clarke S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. J Biol Chem. 1989;26:6164–70.
13. Patel K, Borchardt RT. Chemical pathways of peptide degradation. III. Effect of primary sequence on the pathways of deamidation of asparaginyl residues in hexapeptides. Pharm Res. 1990;7:787–93.
14. Tyler-Cross R, Schirch V. Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. J Biol Chem. 1991;266:22549–56.
15. Brennan TV, Clarke S. Effect of adjacent histidine and cysteine residues on the spontaneous degradation of asparaginyl- and aspartyl-containing peptides. Int J Pept Protein Res. 1995;45: 547–53.
16. Radkiewicz JL, Zipse H, Clarke S, Houk KN. Neighboring side chain effects on asparaginyl and aspartyl degradation: an ab initio study of the relationship between peptide conformation and backbone NH acidity. J Am Chem Soc. 2001;123:3499–506.
17. Clarke S. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. Int J Pept Protein Res. 1987;30:808–21.
18. Kosky AA, Razzaq UO, Treuheit MJ, Brems DN. The effects of alpha-helix on the stability of Asn residues: deamidation rates in peptides of varying helicity. Pro Sci. 1999;8:2519–23.
19. Stevenson CL, Friedman AR, Kubik TM, Donlan ME, Borchardt RT. Effect of secondary structure on the rate of several growth hormone releasing factor analogs. Int J Pept Protein Res. 1993;42:497–503.
20. Kossiakoff AA. Tertiary structure is a principal determinant to protein deamidation. Science. 1988;240:191–4.
21. Radkiewicz JL, Zipse H, Clarke S, Houk KN. Neighboring side chain effects on asparagine and aspartyl degradation: an ab initio study on the relationship between peptide conformation and peptide backbone NH acidity. J Am Chem Soc. 2001;123:3499–506.
22. Torrez M, Schultehenrich M, Livesay DR. Conferring thermostability to mesophilic proteins through optimized electrostatic surfaces. Biophys J. 2003;85:2845–53.
23. Dominy BN, Perl D, Schmid FX, Brooks CL. The effects of ionic strength on protein stability: the cold shock protein family. J Mol Biol. 2002;319:541–54.
24. Sanchez-Ruiz JM, Makhatadze GI. To charge or not to charge? Trends Biotechnol. 2001;19:132–5.
25. Wunderlich M, Martin A, Schmid FX. Stabilization of the cold shock proteins CspB from Bacillus subtilis by evolutionary optimization of coulombic interactions. J Mol Biol. 2005;347: 1063–76.
26. Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem. 1998;41:2481–91.
27. Hellberg S, Sjostrom M, Skagerberg B, Wold S. Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem. 1987;30:1126–35.
28. Cruciani G, Baroni M, Carosati E, Clementi M, Valigi R, Clementi S. Peptide studies by means of principle properties of amino acids derived from MIF descriptors. J Chemometrics. 2004;18:146–55.
29. Robinson NE, Robinson AB, Merrifield RB. Mass spectrometric evaluation of synthetic peptides as primary structure models for peptide and protein deamidation. J Pept Res. 2001;57:483–93.
30. Robinson NE, Robinson AB. Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects. J Pept Res. 2004;63:437–48.
31. Robinson NE, Robinson ZW, Robinson BR, Robinson AL, Robinson JA, Robinson ML, *et al.* Structure-dependent non-enzymatic deamidation of glutaminyl and asparaginyl pentapeptides. J Pept Res. 2004;63:426–36.
32. Palermo G, Piraino P, Zucht HD. Performance of PLS regression coefficients in selecting variables for each response of a multi-

variate PLS for omics-type data. Advances and Applications in Bioinformatics and Chemistry. 2009;2009(2):57–70.
33. Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. Proteins. 1993;17:75–86.
34. Smith DK, Radivojac P, Obradovic Z, Dunker AK, Zhu G. Improved amino acid flexibility parameters. Protein Sci. 2003;12:1060–72.
35. Katz MH. Multivariate analysis: a practice guide for clinicians. New York: Cambridge University Press; 1999. p. 158–62.
36. Stahle L, Wold S. Multivariate data analysis and experimental design in biomedical research. Prog Med Chem. 1988;25:291–338.
37. Wold S. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst. 2001;58:109–30.
38. Esbensen K-H. Multivariate data analysis—in practice: an introduction to multivariate data analysis and experimental design. 5th ed. USA: Camo Process; 2001.
39. Martens H, Martens M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Qual Prefer. 2000;11:5–16.
40. Lin TP, Hsu CC. Determination of residual moisture in lyophilized protein pharmaceuticals using a rapid and non-invasive method: near infrared spectroscopy. PDA J Pharm Sci Technol. 2002;56:196–205.
41. Wakeling IN, Morris JJ. A test of significance for partial least squares regression, *J*. Chemometrics. 1993;7:291–304.
42. Elshereef R, Budman H, Moresoli C, Legge RL. Fluorescence spectroscopy as a tool for monitoring solubility and aggregation behavior of β-lactoglobulin after heat treatment. Biotechnol Bioeng. 2006;95:863–74.
43. Verdu-Andres J, Massart DL, Menardo C, Sterna C. Correlation of non-linearities in spectroscopic multivariate calibration by using transformed original variables and PLS regression. Anal Chim Acta. 1997;349:271–82.
44. Berglund A, Wold S. INLR, implicit non-linear latent variable regression. J Chemometrics. 1997;11:141–56.
45. Baeza-Baeza JJ, Ramis-Ramos G. Reduction of the relative standard deviation in the least-squares fitting of limearized equations by using sensitivity weights. Anal Chim Acta. 1995;316:173–84.
46. Spiegelman C, Wikander J, O'Neal P, Coté GL. A simple method for linearizing nonlinear spectra for calibration. Chemom Intell Lab Syst. 2002;60:197–209.
47. Hadjuski L, Geladi P, Hopke P. A comparison of modeling nonlinear systems with artificial neural networks and partial least squares. Chemom Intell Lab Syst. 1999;49:91–103.
48. Capasso S. Estimation of the deamidation rate of asparagine side chains. J Pept Res. 2000;55:224–9.