Weitzman, P. D. J., & Kinghorn, H. A. (1978) *FEBS Lett. 88*, 255–258.
Wolodko, W. T., Brownie, E. R., & Bridger, W. A. (1980) *J. Bacteriol. 143*, 231–237.
Wolodko, W. T., O'Connor, M. D., & Bridger, W. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2140–2144.
Wolodko, W. T., Brownie, E. R., O'Connor, M. D., & Bridger, W. A. (1983) *J. Biol. Chem. 258*, 14116–14119.

# New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites†

J. M. R. Parker,‡ D. Guo, and R. S. Hodges*

*Medical Research Council Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7*

*Received September 19, 1985; Revised Manuscript Received April 25, 1986*

ABSTRACT: A new set of hydrophilicity high-performance liquid chromatography (HPLC) parameters is presented. These parameters were derived from the retention times of 20 model synthetic peptides, Ac-Gly-X-X-$(Leu)_3$-$(Lys)_2$-amide, where X was substituted with the 20 amino acids found in proteins. Since hydrophilicity parameters have been used extensively in algorithms to predict which amino acid residues are antigenic, we have compared the profiles generated by our new set of hydrophilic HPLC parameters on the same scale as nine other sets of parameters. Generally, it was found that the HPLC parameters obtained in this study correlated best with antigenicity. In addition, it was shown that a combination of the three best parameters for predicting antigenicity further improved the predictions. These predicted surface sites or, in other words, the hydrophilic, accessible, or mobile regions were then correlated to the known antigenic sites from immunological studies and accessible sites determined by X-ray crystallographic data for several proteins.

It is now thought that the entire surface of a protein can be considered antigenic when peptide fragments of the protein surface are used as immunogens or a different species is used as the immunizing host (Green et al., 1982; Benjamin et al., 1984). Antigenic sites are defined as those residues of a native protein that are bound by antibodies raised to a native protein, native protein fragments, or synthetic peptides. By definition, since antigenic sites are those recognized by antibodies, it is most likely that these sites are accessible or on the surface of a protein, and these regions are probably more mobile than interior regions. Since these sites are on the surface, they are probably hydrophilic. Indeed, algorithms for hydrophilicity and accessibility have been used to predict antigenicity.

We have experimentally determined a new hydrophilicity scale derived from the contribution in high-performance liquid chromatography (HPLC)[1] of each amino acid side chain to the retention time of model synthetic peptides, Ac-Gly-X-X-$(Leu)_3$-$(Lys)_2$-amide, where X was substituted by the 20 amino acids found in proteins. This new set of hydrophilicity parameters was used in a modified Hopp and Woods (1978) algorithm to predict which areas of a protein are on the surface. It was found that these predicted surface sites correlate well with the known antigenic sites for myoglobin, lysozyme, cytochrome *c*, and influenza hemagglutinin. An excellent correlation was also shown for accessible sites determined by X-ray data for myoglobin, lysozyme, cytochrome *c*, bovine trypsin, rat mast cell protease, and *Streptomyces griseus* trypsin.

## Experimental Procedures

*Materials.* Unless otherwise stated, chemicals and solvents were reagent grade. Diisopropylethylamine (DIEA), dichloromethane ($CH_2Cl_2$), and trifluoroacetic acid (TFA) were redistilled prior to use. Picric acid was dissolved in $CH_2Cl_2$ and dried over magnesium sulfate. Acetonitrile (HPLC grade) was obtained from Fisher Scientific, Fairlawn, NJ. Double-distilled water was purified by passage through a Milli-Q water purification system (Millipore Corp., Bedford, MA). Poly(styrene-*co*-divinylbenzene) benzydrylamine hydrochloride resin (0.75 mmol of $NH_2$/g of resin) and poly(styrene-*co*-divinylbenzene) chloromethyl resin (~1.0 mmol of Cl/g of resin) were purchased from Beckman Instruments, Inc., Palo Alto, CA, and Pierce Chemical Co., Rockford, IL, respectively. *tert*-Butyloxycarbonyl (Boc) amino acids were purchased from Vega Biochemicals (Tucson, AZ), Bachem Fine Chemicals, Inc. (Torrance, CA), Beckman Instruments, Inc. (Palo Alto, CA), and the Protein Research Foundation (Japan).

*Apparatus.* Peptide synthesis was carried out on a Beckman peptide synthesizer, Model 990. The HPLC instrumentation was composed of a Spectra-Physics SP8700 solvent delivery system and SP8750 organizer module, combined with a

† This work was supported by the Medical Research Council of Canada and equipment grants from the Alberta Heritage Foundation for Medical Research.
* Correspondence should be addressed to this author.
‡ Recipient of an Alberta Heritage Foundation for Medical Research Postdoctoral Fellowship.

[1] Abbreviations: HPLC, high-performance liquid chromatography; RPC, reversed-phase high-performance liquid chromatography; DIEA, diisopropylethylamine; TFA, trifluoroacetic acid; BOC, *tert*-butyloxycarbonyl.

0006-2960/86/0425-5425$01.50/0 

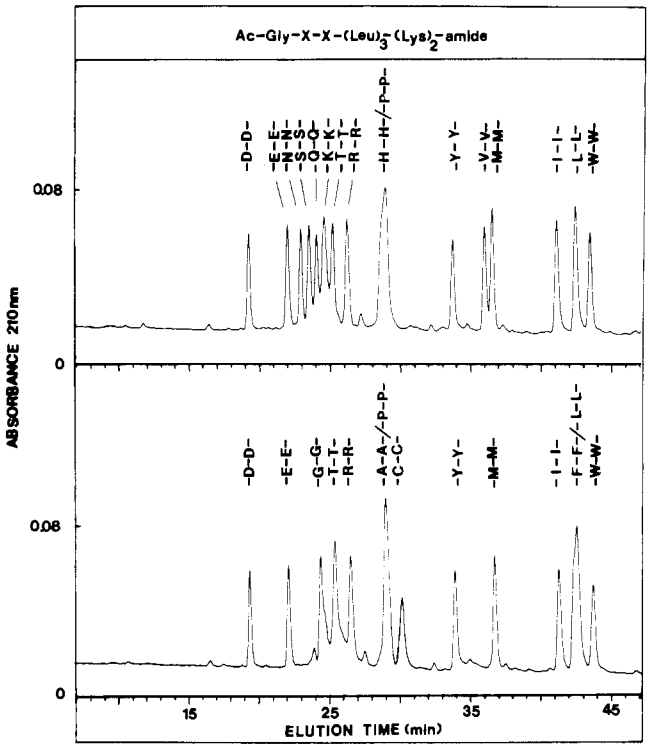


FIGURE 1: Representative RPC profile at pH 7.0 of synthetic peptides with the sequence $N^{\alpha}$-acetyl-Gly-X-X-$(Leu)_3$-$(Lys)_2$-amide, where X is substituted by the 20 amino acids found in proteins. Elution profiles were obtained with a SynChropak RP-P C-18 column (250 × 4.1 mm i.d.) using a linear AB gradient of 1.67% solvent B/min and a flow rate of 1 mL/min. Solvent A was aqueous 10 mM $(NH_4)_2HPO_4$/0.1 M $NaClO_4$ buffer, and solvent B was 0.1 M $NaClO_4$ in aqueous acetonitrile ($H_2O/CH_3CN$, 40:60).

Hewlett-Packard HP 1040A detection system, HP3390A integrator, HP85 computer, HP9121 disk drive, and HP7470A plotter. Samples were injected into a 500-μL injection loop (Model 7125, Rheodyne Inc., Cotati, CA).

*Columns.* Peptide retention times were determined on a reversed-phase SynChropak RP-18 (C-18) column (250 × 4.1 mm i.d., 6.5-μm particle size, 300-Å pore size, carbon loading ~7.5%).

*Peptide Synthesis.* The peptide analogues were synthesized by using the general procedure for solid-phase synthesis described by Merrifield (Erickson & Merrifield, 1976) with modifications described previously (Parker & Hodges, 1985).

*Peptide Purification.* The crude peptides were purified on a SynChropak RP-P C18 reversed-phase column (250 × 4.1 mm i.d.). Solvent A was 0.1% TFA/$H_2O$, and solvent B was 0.05% TFA/acetonitrile. Linear gradients varied from 0.5% to 1% solvent B/min with a flow rate of 1 mL/min. Peptide analogues containing Asn, Asp, Gln, Glu, Arg, Lys, and His residues were checked for purity and correct net charge by high-voltage paper electrophoresis at pH 6.5. Aliquots (10 μL) of stock solutions of the individual purified peptides were hydrolyzed in 100–200 μL of 6 M hydrochloric acid at 110 °C for 24 h in evacuated sealed tubes. The hydrolysates were subsequently analyzed on a Durrum 500 amino acid analyzer to confirm peptide composition. Methanesulfonic acid [containing 0.2% 3-(2-aminoethyl)indole (4 M)] was used as a hydrolyzing agent for the Trp-containing peptide (Simpson et al., 1976). The hydrolysate was partially neutralized with an equal volume of 3.5 M NaOH prior to analysis.

## Results and Discussion

The HPLC hydrophilicity scale presented in this paper provides the first experimentally derived set of parameters

Table I: HPLC Retention Coefficients Derived from Retention Times of Model Peptides[a]

| amino acid | retention coefficient (min) | amino acid | retention coefficient (min) |
|---|---|---|---|
| Trp | +9.5 | Arg | +0.9 |
| Phe | +9.0 | Thr | +0.3 |
| Leu | +9.0 | Lys | +0.0 |
| Ile | +8.3 | Gly | +0.0 |
| Met | +6.0 | Gln | −0.2 |
| Val | +5.7 | Ser | −0.5 |
| Tyr | +4.6 | Asn | −0.8 |
| Cys | +2.6 | Glu | −1.3 |
| Ala | +2.2 | Asp | −2.6 |
| Pro | +2.2 | α-amino | −2.4, 0[b] |
| His | +2.2 | α-COOH | −5.2 |

[a] These are the retention coefficients determined by Guo et al. (1986). [b] The charged α-amino group on an N-terminal Arg residue had a smaller effect than it did on an N-terminal residue with an uncharged side chain.

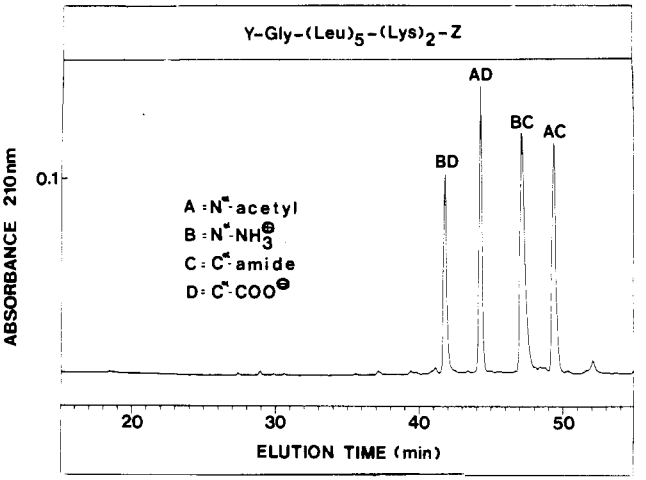


FIGURE 2: Representative RPC profile at pH 7.0 of four synthetic peptides with the sequence Y-Gly-$(Leu)_5$-$(Lys)_2$-Z where Y = $N^{\alpha}$-acetyl or α-amino and Z = $C^{\alpha}$-amide or carboxyl. The chromatographic conditions are described in Figure 1.

using synthetic peptides. It was assumed that the retention time of a peptide was directly related to the summed hydrophobicity and hydrophilicity of the amino acid residues in that peptide.

*Retention Coefficients.* When a model peptide mixture was chromatographed on the C-18 column at pH 7.0, the addition of sodium perchlorate ($NaClO_4$) to the gradient buffers [buffer A = aqueous 10 mM $(NH_4)_2HPO_4$/0.1 M $NaClO_4$; buffer B = 0.1 M $NaClO_4$ in 40% $H_2O$ and 60% acetonitrile] was essential to provide the excellent resolution shown in Figure 1. Prior to use, the pH 7.0 gradient buffer A and the aqueous component of buffer B were passed through a preparative Synchropak C-18 column to remove impurities associated with the buffer salts. In the absence of perchlorate, the peptides exhibited high retention times and peak broadening. Reversed-phase silica gel columns may contain surface silanols which act as weak acids and are ionized above pH 3.5–4. At pH 7.0, the $NaClO_4$ may be suppressing ionic interactions between the peptides and column support through ion-pair formation (Meek, 1980).

Retention coefficients for N- and C-terminal groups (Table I) were determined from the separation of a mixture of model peptides with the sequence Y-Gly-$(Leu)_5$-$(Lys)_2$-Z, where Y = $N^{\alpha}$-acetyl (A) or α-amino (B) and Z = $C^{\alpha}$-amide (C) or α-carboxyl (D). A chromatogram of a mixture of the four peptides is shown in Figure 2. The effect of changing the

acetyl moiety at the N-terminus to a free $\alpha$-amino group is shown by comparing peptides AD and BD or AC and BC (Figure 2). It is very common for the $pK_a$ of a peptide $\alpha$-amino group to be near 7, and this partial deprotonation would explain the smaller effect at pH 7.0 compared to that observed by Guo et al. (1986) at a lower pH. In contrast, the effect of changing the C-terminal amide to an $\alpha$-carboxyl group shows a large decrease in retention time (compare AC and AD or BC and BD; Figure 2). The C-terminal $\alpha$-carboxyl group would be competely ionized ($COO^-$) under these conditions (pH 7.0) and would be expected to have a large effect on the hydrophilicity of the peptide. Thus, at pH 7.0, the $\alpha$-carboxyl group is fully ionized, and at lower pH the $\alpha$-amino group is fully protonated. These charged end groups, in contrast to blocked and uncharged end groups, can drastically affect the retention time of a peptide.

Of the three organic solvents most commonly used in RPC, the order of effectiveness in eluting peptides has been shown to be 2-propanol > acetonitrile > methanol (Mahoney & Hermodsen, 1980; Wilson et al., 1981; Hermodsen & Mahoney, 1983). However, the much superior resolution and selectivity obtained with acetonitrile compared to the alcohols confirms its value as the best organic solvent for most practical purposes (Guo et al., 1986). The usefulness of the alcohols is generally limited to special cases where, with very hydrophobic peptides, a more nonpolar solvent (2-propanol) or, with very hydrophilic peptides, a more polar solvent (methanol) may be advantageous.

To show that organic modifiers have little effect on the retention coefficients previously run in acetonitrile (Guo et al., 1986) we carried out the separation of a peptide mixture in 2-propanol, acetonitrile, and methanol (Figure 3). The most important conclusion drawn from Figure 3 is that the relative elution order of the peptides using all three solvents is almost identical. This was shown to be true even for different *n*-alkyl supports. The RPC profile using a C-18 *n*-alkyl column (see Figure 1) was almost identical with the profile using a C-8 *n*-alkyl support (middle panel, Figure 3). The discrepancies observed for a few residues were attributed to variations in column selectivity due to batch variation and column aging. Because of the superior resolution and selectivity with acetonitrile as the eluting organic solvent, we have used the hydrophilicity parameters derived from acetonitrile to predict antigenic sites.

The retention coefficients (in minutes) were determined from the retention times on reversed-phase HPLC (see Figure 1). The retention time of each peptide was assigned by recording the retention time of the pure peptide and peptide mixtures and spiking these mixtures with a calculated amount of a known peptide. The retention coefficients for each amino acid residue (X) in the eight-residue sequence Ac-Gly-X-X-$(Leu)_3$-$(Lys)_2$-amide were determined by subtracting the retention time of the core peptide Ac-Gly-$(Leu)_3$-$(Lys)_2$-amide from the retention time of the eight-residue sequence and dividing the result by two. Two approaches were used to obtain a value for the core peptide. In the first approach, the retention coefficients for Gly, Leu, and Lys were determined by using simultaneous equations of retention data for three of the eight-residue peptide sequences where -X-X- was Gly, Leu, or Lys. In the second approach, the observed retention time of the synthetic core peptide Ac-Gly-$(Leu)_3$-$(Lys)_2$-amide was used. The results of these two approaches were identical. All parameters are calculated for $N^{\alpha}$-acetylated and C-terminal amide peptides. These retention coefficients are shown in Table I.

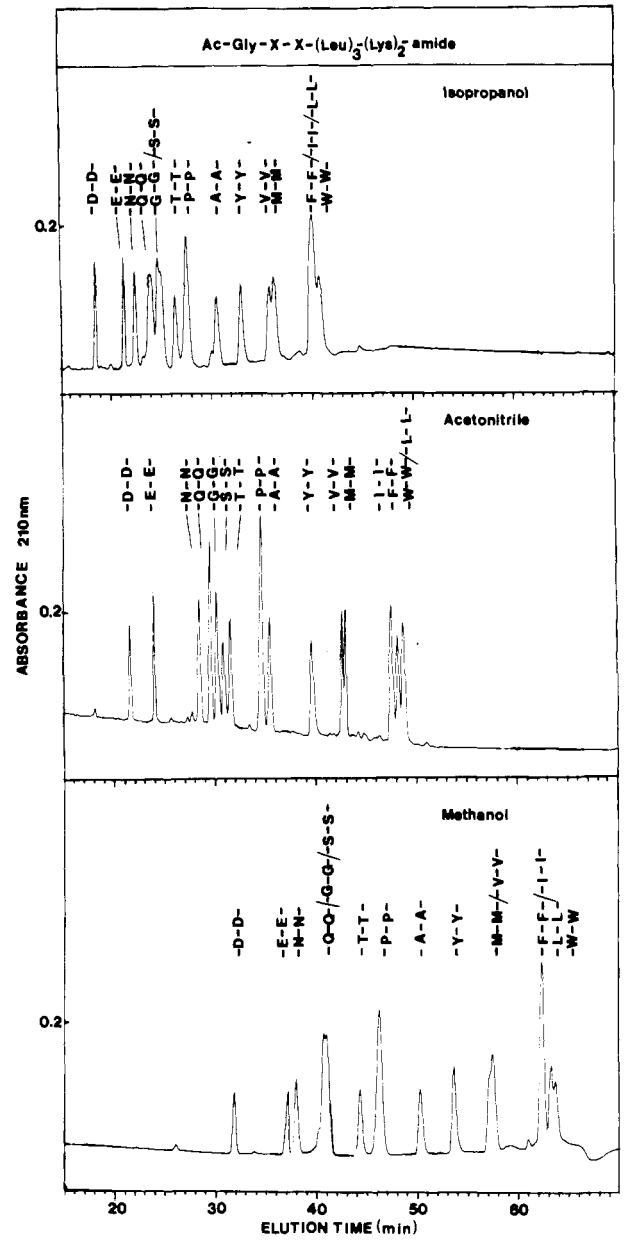


FIGURE 3: Representative RPC profile at pH 7.0 of 16 synthetic peptides using the chromatographic conditions described in Figure 1 with 3 different eluting solvents on a SynChropak RP-P C-8 column (250 × 4.1 mm i.d.). This column is different than the column used to give the profile in Figure 1. The top panel shows the profile when solvent B contained 2-propanol; the middle panel, acetonitrile; and the bottom panel, methanol.

*Hydrophilicity Scale.* The peptide retention times (from Figure 1) provided a set of HPLC retention coefficients (Table I) for all 20 amino acid residues found in proteins. The HPLC hydrophilicity scale shown in Table II assigns the amino acid residue with the maximum retention coefficient from Table I a hydrophobic value of −10 and the amino acid residue showing the minimum retention coefficient a hydrophilic value of +10. The remaining amino acid residues were scaled proportionally. In all previously reported hydrophilicity scales, no account is made for the contribution of the N-terminal free amino group or C-terminal carboxyl group. Using the data in Table I, we have determined these values to be 9.7 and 14.3, respectively. We have scaled all other parameters used in this paper in a similar manner so that they can be compared directly (Tables II and III).

*Correlation to Other Parameters.* A comparison of these new HPLC parameters with nine other parameters is shown

Table II: HPLC Parameters and Parameters Derived from X-ray Crystallographic Data Used To Predict Hydrophilicity and Antigenicity

| | parameters[a] | | | | | |
|---|---|---|---|---|---|---|
| | | | | *B* values[f] | | |
| amino acid | HPLC[b] | accessibility[c,d] | bulk[e] hydrophobic character | set 1 | set 2 | set 3 |
| Trp | −10.0 | +3.2 | −2.6 | −10.0 | −5.1 | −10.0 |
| Phe | −9.2 | +0.5 | −2.9 | −9.6 | −5.6 | −1.2 |
| Leu | −9.2 | −0.3 | −6.6 | −6.5 | −1.3 | −0.7 |
| Ile | −8.0 | −3.4 | −9.8 | −3.7 | −7.3 | −4.6 |
| Met | −4.2 | +1.9 | −4.6 | −8.2 | −10.0 | −9.9 |
| Val | −3.7 | −2.5 | −10.0 | −5.3 | −4.2 | −1.3 |
| Tyr | −1.9 | +8.0 | −0.6 | −7.0 | −4.0 | −7.3 |
| Cys | +1.4 | −10.0 | −5.5 | −7.1 | −8.6 | −0.4 |
| Ala | +2.1 | +2.7 | +1.3 | −0.5 | −2.6 | −3.0 |
| Pro | +2.1 | +7.5 | +7.9 | +0.6 | +9.6 | +0.1 |
| His | +2.1 | +6.7 | +4.6 | −5.3 | −2.1 | −2.8 |
| Arg | +4.2 | +9.8 | +6.4 | −0.7 | +4.6 | −2.3 |
| Thr | +5.2 | +7.1 | +6.5 | +2.1 | +6.6 | +0.3 |
| Lys | +5.7 | +10.0 | +7.9 | +3.8 | +9.4 | +10.0 |
| Gly | +5.7 | +2.3 | +3.5 | +7.8 | +5.8 | −0.5 |
| Gln | +6.0 | +8.9 | +6.2 | +9.6 | +4.6 | −3.5 |
| Ser | +6.5 | +6.7 | +8.4 | +10.0 | +6.4 | −0.5 |
| Asn | +7.0 | +8.4 | +7.6 | +5.7 | +2.7 | +0.0 |
| Glu | +7.8 | +8.9 | +5.7 | +3.8 | +5.3 | +0.7 |
| Asp | +10.0 | +8.4 | +10.0 | −1.1 | +10.0 | +0.1 |
| α-amino | +9.7 | | | | | |
| α-COOH | +14.3 | | | | | |

[a] These parameters are derived from peptide and protein data whereas those in Table III are derived from free amino acids and their derivatives. [b] This paper and Guo et al. (1986). [c] Janin (1979). [d] These parameters are very similar to those of Chothia (1976). [e] Manavalan & Ponnuswamy (1978). [f] Karplus & Schulz (1985).

Table III: Partition and Free Energy Transfer Parameters of Amino Acids Used To Predict Regions of Hydrophilicity and Antigenicity in Proteins

| | parameters[a] | | | | | |
|---|---|---|---|---|---|---|
| amino acid residues | hydrophobicity[b] | global[c] | hydrophobicity[d] | π[e] | hydropathy[f] | hydration[g] potential |
| Trp | −10.0 | −10.0 | −6.6 | −10.0 | +2.0 | +2.4 |
| Phe | −7.2 | −5.3 | −9.0 | −7.2 | −6.2 | −5.2 |
| Leu | −5.0 | −3.7 | −10.0 | −6.6 | −8.4 | −9.8 |
| Ile | −5.0 | −3.9 | −8.5 | −7.2 | −10.0 | −9.6 |
| Met | −3.4 | −2.2 | −2.4 | −3.7 | −4.2 | −4.2 |
| Val | −4.1 | −3.1 | −3.1 | −3.7 | −9.3 | −9.4 |
| Tyr | −6.6 | −5.0 | −8.3 | −2.1 | +2.9 | +2.7 |
| Cys | −2.5 | −1.7 | +5.3 | −2.2 | −5.5 | −4.5 |
| Ala | −0.9 | −0.1 | +7.2 | +1.9 | −4.0 | −9.3 |
| Pro | +0.6 | +1.1 | +1.3 | −0.6 | +3.5 | +1.0[h] |
| His | −0.9 | −0.2 | +7.9 | +3.0 | +7.1 | +8.9 |
| Arg | +10.0 | +10.0 | +7.9 | +10.0 | +10.0 | +10.0[h] |
| Thr | −0.6 | −0.1 | +4.8 | +2.2 | +1.5 | +0.9 |
| Lys | +10.0 | +9.2 | +6.1 | +9.8 | +8.6 | +7.9 |
| Gly | +0.6 | +1.1 | +8.8 | +3.8 | −0.9 | −10.0 |
| Gln | +1.25 | +1.6 | +10.0 | +5.1 | +7.8 | +7.7 |
| Ser | +1.5 | +2.1 | +5.8 | +4.0 | +1.8 | +1.2 |
| Asn | +1.25 | +1.6 | +9.4 | +7.5 | +7.8 | +8.1 |
| Glu | +10.0 | +9.2 | +6.5 | +7.7 | +7.8 | +8.9 |
| Asp | +10.0 | +8.7 | +7.2 | +8.5 | +7.8 | +10.0 |

[a] These parameters are derived experimentally from free amino acids or their derivatives, whereas those in Table II are derived from peptide and protein data. [b] Hopp & Woods (1981). [c] Fraga (1982). [d] Bull & Breese (1974). [e] Fauchere & Pliska (1983). [f] Kyte & Doolittle (1982). [g] Wolfenden et al. (1979). [h] Arbitrarily assigned.

in Tables II and III. They were calculated to have the following correlations with the HPLC parameters: accessibility, 0.58; bulk hydrophobic character, 0.84; Hopp and Woods hydrophobicity, 0.82; global, 0.82; Bull and Breese hydrophobicity, 0.89; Faushere and Pliska π parameter, 0.93; Kyte and Doolittle hydropathy, 0.72; and hydration potential, 0.59. Although the correlation with the π parameter was the highest, large discrepancies were observed for the hydrophilic residues arginine and lysine and the nonpolar residues proline, glycine, and cysteine.

*Surface Profile.* The surface profile for a protein was determined by summing the parameters for each residue of a seven-residue segment and assigning this sum to the fourth residue. This procedure was repeated by shifting the segment by one residue from the N- to the C-terminus. A plot of these values against the residue number using either HPLC, accessibility, bulk hydrophobic character, hydrophobicity, global, π, or hydropathy parameters provided a surface profile. The only exception to this procedure was that for *B* values. The algorithm using *B* values was described in the paper by Karplus and Schultz (1985). To objectively interpret all of these profiles, we have used the following arbitrary of rules: (1) The average surface hydrophilicity is defined as the mean of the profile values for a protein using a particular parameter set. (2) Any residues with a profile value greater than 25% above the average surface hydrophilicity value were defined as

surface sites. The 25% value was calculated as 25% of the difference between the maximum value in the plot and the average surface hydrophilicity value. (3) In determining the profile of a protein with a free N-terminal amino group or a C-terminal carboxyl group, the experimental values of 9.7 (amino group) and 14.3 (carboxyl group) are added directly to the value for the N- or C-terminal amino acid residue. The coefficients listed in Table I were calculated for the $N^{\alpha}$-acetyl and $C^{\alpha}$-amide groups. For all parameters other than the HPLC parameters determined in this study, no account is made for free N- or C-terminal residues.

Surface profiles of four proteins, myoglobin, lysozyme, cytochrome *c*, and influenza hemagglutinin, were calculated by using all the parameters listed in Tables II and III. Examples of these profiles using the HPLC, Janin accessibility, and Karplus and Schulz flexibility parameters are shown in Figure 4A–C for lysozyme. The general patterns of these three profiles and those of all parameters studied were similar. However, the intensity of all the maximum profile values varied with every parameter set. To interpret and correlate these profiles to antigenicity, residues 25%, 50%, and 75% above the average surface hydrophilicity (see Experimental Procedures) were examined.

Antigenic sites were defined as those sites on a protein that were recognized by antibodies raised to that protein or peptide fragments. Antigenic sites for myoglobin, lysozyme, cytochrome *c* (Atassi, 1984; Benjamin et al., 1984), and influenza (Wiley et al., 1981; Green et al., 1982; Muller et al., 1982) have been reported. Generally, for the four proteins investigated, the HPLC parameters predicted more of these known antigenic sites than the other sets of parameters listed in Tables II and III. It was apparent that no single-parameter set was able to predict all antigenic sites. This is illustrated in the plots shown for lysozyme (see Figure 4A–C). In Figure 4B, a strong antigenic site including residues 15–22 is predicted by the Janin parameters. A very weak antigenic site was predicted for the same region by using the HPLC parameters (Figure 4A) while the Karplus and Schulz parameters (Figure 4C) did not show a value above the 25% cutoff. In contrast, both the HPLC hydrophilicity parameters and those of Karplus and Schulz predicted strong antigenic sites for residues 98–102 and 115–120 whereas very weak antigenic sites were predicted in the same regions using the Janin parameters.

*Composite Surface Profile.* A combination of various parameters was investigated to improve the prediction of antigenic sites. Three parameters were chosen that individually provided the highest scores at predicting antigenic sites. These were the HPLC hydrophilicity (this study), Janin accessibility, and Karplus and Schultz flexibility parameters. Three surface profile plots using the HPLC (this paper), accessibility (Janin), and *B*-value parameters (Karplus and Schulz) were used to generate a composite surface profile. The surface sites (defined in rule 2 under Surface Profile) for each profile (hatched areas in Figure 4A–C) were scaled from 0 to 100 where the maximum surface site value in each plot was set equal to 100 and the 25% surface site value was equal to 0. Each scaled plot was then superimposed on the other, and the maximum value of each residue was used to give the composite profile value. This program is available on diskettes for the IBM-PC or APPLE-MacIntosh computer from the Alberta Peptide Institute (Biochemistry Department, University of Alberta). An example of such a composite plot is shown in Figure 5 for lysozyme. Composite surface profiles were then plotted for myoglobin, lysozyme, cytochrome *c* (horse and albacore sequences), influenza HA1, bovine trypsin, rat mast cell protease,

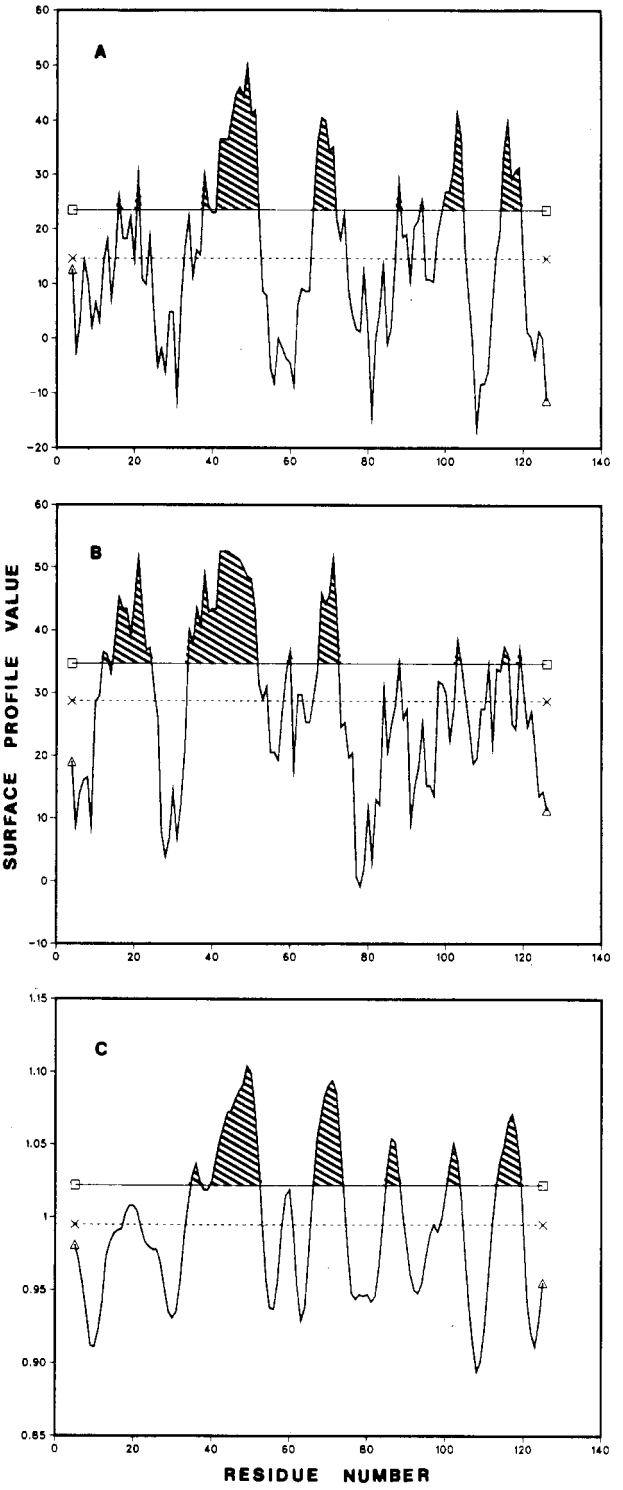


FIGURE 4: Representative surface profiles for lysozyme. (A) Surface profile calculated with the HPLC parameters described in this paper (Table II). (B) Surface profile calculated with the accessibility parameters (Table II). (C) Surface profile calculated with the *B*-value parameters (Table II). The dotted line represents the mean protein hydrophilicity value, and the solid line represents the 25% cutoff value described under Experimental Procedures. The hatched areas defined those regions of the primary sequence used to obtain the composite surface profile plot (see Experimental Procedures).

and *Streptomyces griseus* trypsin. These predicted surface sites are presented in Tables IV–X. Inspection of these tables shows that all of the antigenic sites for myoglobin and cytochrome *c* and influenza are predicted as surface sites. Only the N- and C-terminal antigenic sites were not predicted for lysozyme. The first X-ray diffraction study of a lysozyme–antibody complex (Amit et al., 1985) has confirmed several

Table IV: Myoglobin

| predicted surface residues | exptl antigenic sites determined by immunological studies[a] | surface-accessible residues determined by X-ray data[b] |
|---|---|---|
| 5–6 | 4 | 3–4, 8–9 |
| 18–27 | 15–22 | 15–16, 18–20 |
| 34–45 | 34 | 34–35, 37–38, 41–42 |
| 47–63, 65 | 56–62 | 44–45, 47–54, 56–57, 62–63 |
| 77–85, 87–90 | 72–88 | 77–79, 83–84, 87–88 |
| 93–98 | 94–99 | 95–96 |
| 118–131 | 113–119 | 116–118, 120–122, 125–126 |
| 140, 142–150 | 145–151 | 144–149, 147–150, 152–153 |

[a] Atassi (1984); Benjamin et al. (1984). [b] Brookhaven Data Bank, 168 (deoxy sperm whale myoglobin).

Table V: Lysozyme

| predicted surface residues | exptl antigenic sites determined by immunological studies[a] | surface-accessible residues determined by X-ray data[b] |
|---|---|---|
| | 1, 5, 7 | 1–2 |
| 12–13, 15–24 | 13–14, 19–24 | 13–14, 18–22 |
| 34–52 | 33, 34, 45, 48 | 33–34, 43–48 |
| 60 | 62 | 61–62 |
| 66–73 | 64–80 | 67–68 |
| 85–88 | 87–89 | 85–87 |
| 94, 100–104 | 93–97, 102, 103 | 101, 103 |
| 113–119 | 113, 114, 116, 117–121 | 112–114, 116–119 |
| | 125 | 125–126, 128–129 |

[a] Atassi (1984); Benjamin et al. (1984); Amit et al. (1985). [b] Brookhaven Data Bank, 253 (RS16 hen egg-white lysozyme).

of the predicted surface sites for lysozyme. Other than the known antigenic sites, only one extra surface site (119–131) was predicted for myoglobin. Two surface sites (15–30 and 68–78) other than the known antigenic sites were predicted for cytochrome *c*. Four additional surface sites (153–159, 168–175), 227–264, and 282–299) were predicted for influenza. It is interesting to note that these sites are in areas where several mutations have been found in the wild-type influenza (Wiley et al., 1981).

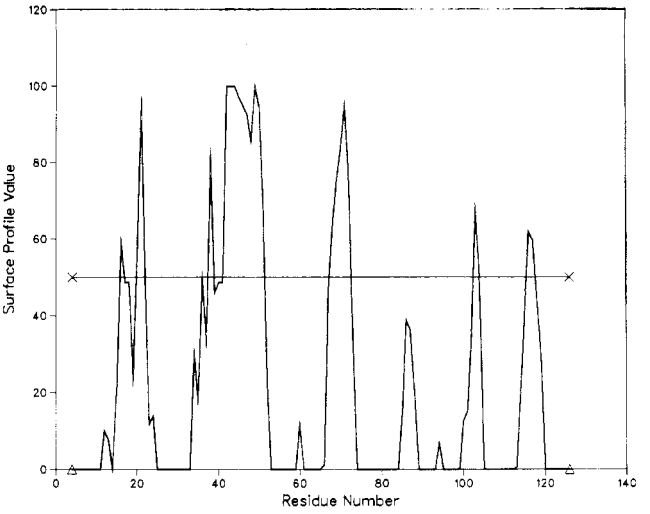


FIGURE 5: Predicted composite surface profile for lysozyme obtained by combining the hatched areas in Figure 4A–C as described under Experimental Procedures. The solid line represents a 50% cutoff value. Residues with composite profile values greater than this value would be used to select potential synthetic peptides capable of producing antipeptide antibodies which would recognize linear antigenic determinants on a protein.

*Surface-Accessible Residues from X-ray Data.* These predicted surface sites were also found to correlate well with surface-accessible residues determined by X-ray crystallography. All surface-accessible residues were calculated by using the method of Lee and Richards (1971) using a probe radius of 1.4 Å. Two or more consecutive residues with accessible areas greater than 50 Å$^2$ each were defined as being on the protein surface. Coordinates for the following proteins were from the Brookhaven Protein Data Bank (April, 1985): myoylobin (268), lysozyme (253), cytochrome *c* albacore (149), bovine trypsin (329), rat mast cell protease (GSPRAT, 298), *Streptomyces griseus* trypsin (SGT, personnel communication). The accessible residues for these proteins are shown in Tables IV–VI and VIII–X. It should be noted that the predicted sites for myoglobin and cytochrome *c* which did not correlate with immunological data were shown to be on the

Table VI: Cytochrome *c* (from Horse and Albacore)

| predicted surface residues | | exptl antigenic sites determined by immunological studies,[a] | surface-accessible residues determined by |
|---|---|---|---|
| albacore | horse[c] | horse | X-ray data,[b] albacore |
| 4–8 | 4–7 | 1–4 | 2–5, 7–8 |
| 10, 15–16, 19–28 | 15–16, 19–30 | | 11–13, 21–23, 25–28 |
| 38–55 | 38–44, 50–56 | 44, 46–50 | 44–45, 53–55 |
| 63–64 | 61–64 | 60–62 | 60–62 |
| 68–78 | 68–73, 76–78 | | 72–73, 76–77 |
| 86–93 | 86–92 | 89–92, 96 | 86–88 |
| 100 | 100–101 | 103, 104 | 99–100 |

[a] Atassi (1984); Benjamin et al. (1984). [b] Brookhaven Data Bank, 149 (oxidized albacore cytochrome *c*). [c] Primary sequence from Dickerson et al. (1971).

Table VII: Influenza HA1

| predicted surface residues | exptl antigenic sites determined by immunological studies[a] | predicted surface residues | exptl antigenic sites determined by immunological studies[a] |
|---|---|---|---|
| 4–10, 21–24, 29–30, 32–35, 38 | 1–36 | 185–192 | 187–196 |
| 44–52 | 52 | 200, 202, 206–212, 218–220 | 205–220 |
| 74–75, 80, 82–83, 91–93, 98, 101–104, 106–107 | 76–111 | 227–228, 239–240, 247–249, 256–265, 267–268 | |
| 115–117 | 105–140 | 271–274, 278 | 277 |
| 131–145 | 140–146 | 282–296 | |
| 153, 156–160 | | 304, 309–313, 323–325 | 306–329 |
| 168–175 | | | |

[a] Wiley et al. (1981); Green et al. (1982); Muller et al. (1982).

Table VIII: Bovine Trypsin

| predicted surface residues | surface-accessible residues determined by X-ray data[a] | predicted surface residues | surface-accessible residues determined by X-ray data[a] |
|---|---|---|---|
| 6–9, 11–14, 16–17, 20 | 5–6, 21–22 | 122–133 | 125–129 |
| | | 145–152, 154–157 | 145–146, 149–150, 153–155 |
| 42, 44 | 42–45 | | |
| 50–52, 54, 59–62 | 56–61 | 167–180 | 165–167 |
| | | 185–186 | |
| 75–82 | 74–80 | 197–205 | 200–202 |
| 92, 94–96 | 91–93, 95–99 | 220 | 217–218, 221–223 |
| 109–113 | 107–109 | | |

[a] Brookhaven Data Bank, 329 (DIP-inhibited trypsin).

Table IX: Rat Mast Cell Group Specific Protease

| predicted surface residues | surface-accessible residues determined by X-ray data[a] | predicted surface residues | surface-accessible residues determined by X-ray data[a] |
|---|---|---|---|
| 6, 9–16 | 5–6, 8–11 | 149–155 | 150–152 |
| 22–25 | 23–24, 26–28 | 158–163, 169–174 | 155–159, 163–164, 171–172, 174–175 |
| 45–50 | 49–50 | | |
| 59–73 | 61–63, 73–75 | 180–185 | |
| 76–77, 80–86 | 79–80, 82–85 | 191 | 189–190 |
| 95–100 | 96–98, 100–101 | 199–205 | 200–202 |
| | 106–107 | 212 | |
| 111–115 | 113–116, 118–119 | | 220–221, 223–224 |
| 128–142 | 133–137 | | |

[a] Brookhaven Data Bank, 298 (rat mast cell proteinase II).

Table X: *Streptomyces grieseus* Trypsin

| predicted surface residues | surface-accessible residues determined by X-ray data[a] | predicted surface residues | surface-accessible residues determined by X-ray data[a] |
|---|---|---|---|
| 5–9 | 5–6, 8–9 | 139–141, 145 | 142–143, 147–148 |
| 22 | 19–20 | 151, 155 | |
| 40–52 | 46–47 | 159–175 | 159–162 |
| 59, 61, 66–69 | 60–61, 63–64 | 177–184 | 178–179, 181–185 |
| 74–81 | 74–77 | 201–202 | 199–200 |
| 92–96 | 89–91, 93–96 | 205, 209 | |
| 103–109 | 101–104 | 220 | 221–222 |
| 119–129 | 120–121, 123–124, 126–127 | | |

[a] Personal communication with Dr. R. Read, Department of Biochemistry, University of Alberta.

protein surface by X-ray crystallographic data.

In conclusion, we have presented a new set of hydrophilicity parameters derived from HPLC of synthetic peptides. We have also scaled nine other parameters so that a direct comparison to these new HPLC parameters can be made. Since no single parameter was able to predict all antigenic sites, we have combined the HPLC, accessibility, and flexibility parameters to produce a composite surface profile. This composite profile was shown to correlate well with the known antigenic and X-ray-determined accessible sites for several proteins. Previously, hydrophilicity algorithms had been used extensively to select potential synthetic peptides from a protein sequence capable of producing antipeptide antibodies which recognize linear antigenic determinants on the protein. Antibodies specific for synthetic peptides can be prepared by coupling this peptide to a protein carrier (Parker & Hodges, 1985) and immunizing animal hosts with these peptide conjugates. We suggest that sites with surface profiles greater than a value of 50% (see Figure 5) be chosen first, followed by those sites with successively lower cutoff values.

## ACKNOWLEDGMENTS

We thank M. Fujinaga and J. Moult from this department for programs to calculate chain flexibility and surface residues from X-ray data. We also thank R. Reed for providing his refinement for the structure of SGT.

**Registry No.** L-W, 73-22-3; L-F, 63-91-2; L-L, 61-90-5; L-I, 73-32-5; L-M, 63-68-3; L-V, 72-18-4; L-Y, 60-18-4; L-C, 52-90-4; L-A, 56-41-7; L-P, 147-85-3; L-H, 71-00-1; L-R, 74-79-3; L-T, 72-19-5; L-K, 56-87-1; G, 56-40-6; L-Q, 56-85-9; L-S, 56-45-1; L-D, 56-84-8; L-E, 56-86-0; L-Asn, 70-47-3; AC, 102623-33-6; AD, 103437-43-0; BC, 103437-44-1; BD, 103437-42-9; Ac-Gly-D-D-$(Leu)_3$-$(Lys)_2$-amide, 102623-43-8; Ac-Gly-E-E-$(Leu)_3$-$(Lys)_2$-amide, 102623-41-6; Ac-Gly-N-N-$(Leu)_3$-$(Lys)_2$-amide, 102623-48-3; Ac-Gly-S-S-$(Leu)_3$-$(Lys)_2$-amide, 102623-45-0; Ac-Gly-Q-Q-$(Leu)_3$-$(Lys)_2$-amide, 102623-44-9; Ac-Gly-K-K-$(Leu)_3$-$(Lys)_2$-amide, 102623-50-7; Ac-Gly-T-T-$(Leu)_3$-$(Lys)_2$-amide, 102623-42-7; Ac-Gly-R-R-$(Leu)_3$-$(Lys)_2$-amide, 102623-47-2; Ac-Gly-H-H-$(Leu)_3$-$(Lys)_2$-amide, 102623-49-4; Ac-Gly-P-P-$(Leu)_3$-$(Lys)_2$-amide, 102623-39-2; Ac-Gly-Y-Y-$(Leu)_3$-$(Lys)_2$-amide, 102623-37-0; Ac-Gly-V-V-$(Leu)_3$-$(Lys)_2$-amide, 102623-36-9; Ac-Gly-M-M-$(Leu)_3$-$(Lys)_2$-amide, 102623-35-8; Ac-Gly-I-I-$(Leu)_3$-$(Lys)_2$-amide, 102623-34-7; Ac-Gly-W-W-$(Leu)_3$-$(Lys)_2$-amide, 102644-25-7; Ac-Gly-G-G-$(Leu)_3$-$(Lys)_2$-amide, 102623-46-1; Ac-Gly-A-A-$(Leu)_3$-$(Lys)_2$-amide, 102623-40-5; Ac-Gly-C-C-$(Leu)_3$-$(Lys)_2$-amide, 102623-38-1; Ac-Gly-F-F-$(Leu)_3$-$(Lys)_2$-amide, 102623-32-5; cytochrome *c*, 9007-43-6; lysozyme, 9001-63-2; trypsin, 9002-07-7; protease, 9001-92-7.

## REFERENCES

Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., & Pojak, R. J. (1985) *Nature (London) 313*, 156–158.

Atassi, M. Z. (1984) *Eur. J. Biochem. 145*, 1–20.

Benjamin, D. C., Berzofsky, J. A., East, I. J., Gurd, F. R. N., Hannum, C., Leach, S. J., Margoliasch, E., Michael, J. G., Miller, A., Prager, E. M., Reichlin, M., Sercarz, E. E., Smith-Gill, S. J., Todd, P. E., & Wilson, A. C. (1984) *Annu. Rev. Biochem. 2*, 67–101.

Bull, H. B., & Bresse, K. (1974) *Arch. Biochem. Biophys. 161*, 665–670.

Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., & Margoliash, E. (1971) *J. Biol. Chem. 246*, 1511–1535.

Erickson, B. W., & Merrifield, R. B. (1976) *Proteins (3rd Ed.) 2*, 255–257.

Fauchere, J. L., & Pliska, V. (1983) *Eur. J. Med. Chem. 18*, 369–375.

Fraga, S. (1982) *Can. J. Chem. 60*, 2606–2610.

Green, N., Alexander, H., Olson, A., Alexander, S., Schinnick, T. M., Sutcliffe, J. G., & Lerner, R. A. (1982) *Cell (Cambridge, Mass.) 28*, 477–487.

Guo, D., Mant, C. T., Taneja, A. K., Parker, J. M. R., & Hodges, R. S. (1986) *J. Chromatogr. 359*, 499–517.

Hermodsen, M., & Mahoney, W. C. (1983) *Methods Enzymol. 91*, 352.

Hopp, T. P., & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824–3828.

Janin, J. (1979) *Nature (London) 277*, 491–492.

Karplus, P. A., & Schulz, G. E. (1985) *Naturwissenschaften 72*, S.212.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199.

Manavalan, P., & Ponnuswamy, P. K. (1978) *Nature (London) 275*, 673–674.

Meek, J. L. (1983) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1632.

Muller, G. M., Shapira, M., & Arnon, R. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 569–573.
Parker, J. M. R., & Hodges, R. S. (1985) *J. Protein Chem. 3*, 465–478.
Simpson, R. J., Neuberger, M. R., & Ler, T. Y. (1976) *J. Biol. Chem. 251*, 1936.
Wiley, D. C., Wilson, I. A., & Skehel, J. J. (1981) *Nature (London) 289*, 373–378.
Wilson, K. J., Honegger, A., Stotzel, R. P., & Hughes, G. J. (1981) *Biochem. J. 199*, 31–41.
Wolfenden, R. V., Cullis, P. M., & Southgate, C. C. F. (1979) *Science (Washington, D.C.) 206*, 575–577.

# Why Does Ribonuclease Irreversibly Inactivate at High Temperatures?†

Stephen E. Zale and Alexander M. Klibanov*

*Department of Applied Biological Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

*Received February 21, 1986; Revised Manuscript Received May 16, 1986*

ABSTRACT: The mechanism of irreversible thermoinactivation of bovine pancreatic ribonuclease A in the pH range relevant to enzymatic catalysis has been elucidated. At 90 °C and pH 4, the enzyme inactivation is caused by hydrolysis of peptide bonds at aspartic acid residues (the main process) and deamidation of asparagine and/or glutamine residues. At 90 °C and neutral pH (pH 6 and 8), the enzyme inactivation is caused by a combination of disulfide interchange (the main process), β-elimination of cystine residues, and deamidation of asparagine and/or glutamine residues. These four processes appear to demarcate the upper limit of thermostability of enzymes.

Enzymes (proteins) are exposed to high temperatures in a variety of real-life situations including thermophilic microorganisms living in hot springs and hydrothermal vents, industrial biochemical reactors, food preparation, heat sterilization of fermentation media, etc. Therefore, elucidation of molecular events taking place in proteins upon heating is of fundamental significance for both pure and applied biochemistry. Knowledge of why and how proteins lose their biological activity is crucial for understanding thermophilic behavior (Zuber, 1976; Friedman, 1978) and also should demarcate the upper limit of thermostability of proteins and hence of life (Brock, 1985). Continuing controversy over whether bacteria can live at 250 °C (Baross & Deming, 1983, 1984; Walsby, 1983; Trent et al., 1984; White, 1984) further underscores the timeliness and importance of such research. In addition, a mechanistic understanding of enzyme stability and inactivation is required for the emerging area of stabilization of enzymes by protein engineering (Ulmer, 1983; Perry & Wetzel, 1984; Estell et al., 1985).

Upon heating in aqueous solutions, the following processes occur in enzymes (Klibanov, 1983): First, the enzyme molecule partially unfolds due to heat-induced disruption of the delicate balance of various noncovalent interactions (Schulz & Schirmer, 1979; Creighton, 1983) that maintain the native conformation at ambient temperature. This process, which results in enzyme inactivation, is completely *reversible*: the enzymatic activity is fully regained if the enzyme solution is promptly cooled down. However, upon prolonged heating, a gradually decreasing fraction of the enzymatic activity returns after cooling, indicating that another, *irreversible* process also takes place. Thus, the overall phenomenon can be depicted by the classical scheme of Lumry and Eyring (1954):

$$N \rightleftharpoons U \rightarrow I \qquad (1)$$

where N is the native, catalytically active enzyme, U is the reversibly unfolded, catalytically inactive enzyme, and I is the irreversibly inactivated enzyme.

A great deal is known about the first step in eq 1 (Kauzmann, 1959; Tanford, 1968; Lapanje, 1978; Privalov, 1979; Jaenicke, 1981; Pfeil, 1981) which is often called thermal denaturation or melting of proteins. The origin and mechanisms of this process are conceptually straightforward and well understood. On the other hand, the nature of the second, irreversible step in eq 1 has remained somewhat mysterious for decades. This process does not render itself to well-defined thermodynamic studies and therefore is usually viewed as simply a nuisance by enzymologists and protein chemists, and yet, this process is the cause of enzyme deterioration in commercial biocatalytic reactors and hence plagues further progress of enzyme technology (Klibanov, 1983). Irreversible thermoinactivation of enzymes has been ascribed to a variety of diverse mechanisms (Baldwin, 1975; Whitaker, 1972; Tombs, 1985), of which none is generally applicable and many have not withstood experimental tests (Klibanov, 1983).

Recently, we have succeeded in elucidating the mechanisms of irreversible thermal inactivation of hen egg white lysozyme (Ahern & Klibanov, 1985). The processes leading to thermoinactivation were found to be deamidation of Asn/Gln residues, hydrolysis of peptide bonds at Asp residues, destruction of disulfide bonds, and formation of incorrectly folded and kinetically trapped structures; their relative contributions depended on the pH. To test the generality of the uncovered mechanisms, in the present work we have investigated irreversible thermoinactivation of another, unrelated enzyme, bovine pancreatic ribonuclease A. Ribonuclease was chosen as a model because it is a relatively small, monomeric enzyme, containing no non-protein components, whose structure is well established (Richards & Wyckoff, 1971; Blackburn & Moore, 1982) and whose reversible thermal denaturation (Hermans & Scheraga, 1961; Brandts, 1965) and conformational dy-

† This research was supported by Grant PCM-8316020 from the National Science Foundation.

0006-2960/86/0425-5432$01.50/0 