CHROM. 19 254

REVERSED-PHASE LIQUID CHROMATOGRAPHY OF SITE-SPECIFIC MUTAGENIZED STAPHYLOCOCCAL NUCLEASE

GRADIENT RETENTION AND CORRELATIONS WITH AMINO ACID-BASED PREDICTIVE SCALES

JOHN C. FORD

Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114 (U.S.A.)

and

JOHN A. SMITH*

*Department of Pathology, Harvard Medical School and *Departments of Molecular Biology and Pathology, Massachusetts General Hospital, Boston, MA 02114 (U.S.A.)*

(Received November 12th, 1986)

SUMMARY

   Comparison of the gradient reversed-phase chromatographic retentions of twelve Staphylococcal nuclease mutants and the naturally occurring protein showed that the chain location and the chemical nature of the substituted amino acid(s) were equally significant in determining the retention. Correlations between the retention times of these nuclease mutants and of previously published data for interleukin 2 mutants and insulin variants with nineteen amino acid-based predictive scales revealed retention time to be significantly correlated to several scales. The use of mutagenized proteins allowed a more sensitive analysis of the individual amino acid contributions to retention than can be achieved by utilizing a more diverse set of proteins.

INTRODUCTION

   Reversed-phase liquid chromatography (RPLC) is an important technique for the isolation and characterization of proteins. While the general principles of RPLC retention are known[1], the details of the mechanism are unclear, particularly for proteins and other macromolecules. Although surface tension and other solvophobic effects are known to control retention[2,3], recent work has shown the importance of conformational changes in determining retention and peak shape[4-6].
   Typically, protein RPLC mechanistic studies are performed using a set of commercially available proteins chosen to have selected properties (*e.g.*, molecular weights, isoelectric points, etc.) varying over a broad range[7,8]. The difficulty with this strategy is that it is unlikely that any two members differ in only one respect.

Therefore, assignment of the origins of an observed behavior to a particular property of the proteins is problematic.

To circumvent such experimental problems, Lau *et al.*[9] used chemically defined, synthetic peptides in their study of the effects of typical RPLC conditions on secondary and quaternary structure. While their approach allows any desired sequence to be studied, the maximum practical size for synthetic peptides is currently limited to approximately 50 residues, due to chemical limitations of the synthetic technique and to purification problems encountered[10]. Additional problems include verification of the purity of these synthetic products and determination of their physical properties.

Alternatively, species variant proteins, having a single difference (or a limited number of differences) in the primary structure, have been used, such as the study by McLeod and Wood[11], using species variant and chemically-modified insulins. Fausnaugh and Regnier[12] have also taken this approach, using lysozymes from a variety of avian species to study the mechanism of hydrophobic interaction chromatography. Although many such variant proteins are available naturally (*e.g.*, the various lysozyme or hemoglobin variants), the use of such proteins requires their isolation from diverse species, many of which may not be readily available, as well as verification of protein sequence differences.

With the advent of genetic engineering techniques, specifically oligodeoxynucleotide-directed site-specific mutagenesis[13], it has become possible to create specific, single-site mutagenized proteins. Thus, the ability to design and generate selected sets of mutants, having a range of properties, yet retaining the same molecular weight, is readily attainable. Kunitani *et al.*[14] have reported the use of RPLC in the purification and characterization of genetically engineered mutants of human recombinant interleukin 2 (IL-2). Their study was an examination of the effects of altering disulfide bonds on the structure and retention of various IL-2 mutants. Such sets of proteins have great potential for investigations of the mechanisms of protein RPLC.

The information gained from such an investigation is maximized when the protein chosen is simple and well-characterized. Thus, we used Staphylococcal nuclease (Nase) [ribonucleate(deoxyribonucleate)-3'-nucleotidohydrolase, EC 3.1.4.7] for an investigation of amino acid retention contributions and of structural effects in RPLC. Naturally occurring, wild-type (WT) Nase is a low-molecular-weight, nonspecific exonuclease. It consists of a single chain of 149 residues without disulfide bonds. A 1.5 Å X-ray diffraction pattern is also available [15]. Numerous mechanistic, immunological, and structural studies have been performed on WT Nase, making it one of the most thoroughly characterized globular proteins[16-19]. Recently, the gene for WT Nase was isolated from *Staphylococcus aureus*[20], and a protocol for the isolation of WT Nase from an *Escherichia coli* expression system was developed[21]. The availability of the Nase gene, site-specific mutagenesis techniques, and facile expression and isolation procedures make Nase an excellent model protein for many physicochemical studies.

We report here an investigation of the gradient RPLC retention behavior of a set of twelve Nase mutants —this set of proteins consists of ten single-site mutations, two multiple-site mutations— and WT Nase. While not an exhaustive array of permutations, this set has several interesting features: three members are tyrosine-to-phenylalanine mutations (residues 85, 113, 115), and two sites (residues 43 and

85) have three different modifications each. Thus, the influence of the chain location of a mutation, as well as the effect(s) of the chemical nature of the mutation may be examined. Additionally, the physicochemical similarity among the set of mutagenized proteins allows the determination of retention correlations which might otherwise be masked by the use of unrelated proteins.

EXPERIMENTAL

The mutagenized Nases were prepared by the two-primer oligodeoxynucleotide-directed site-specific mutagenesis procedure of Zoller and Smith[13]. We shall use the term "mutant" in the following discussion to refer to a mutagenized enzyme. The expression plasmid (pFOG405) containing the WT Nase gene was a kind gift of Dr. David Shortle (Johns Hopkins University). The bacterial expression and isolation of the WT and mutant Nases were as described by Serpersu et al.[21], using E. coli strain SE6004, which was also a gift from Dr. Shortle. The presence of a single band in sodium dodecylsulfate polyacrylamide gel electrophoresis confirmed the purity of the WT Nase and the mutants.

The chromatography was performed using an HP1090 LC system (Hewlett-Packard, Waldbronn, F.R.G.) with simultaneous detection at 280 and 220 nm. The column used throughout was a Vydac TP C_4 (5 μm particles, 150 × 4.6 mm I.D.) purchased from The Separations Group (Hesperia, CA, U.S.A.). The acetonitrile was HPLC grade from J. T. Baker (Phillipsburg, NJ, U.S.A.), the trifluoroacetic acid (TFA), HPLC grade from Pierce (Rockford, IL, U.S.A.), and the water used was purified by passage through an Ultrapure cartridge (No. 90343, Fisher Scientific, Pittsburg, PA, U.S.A.), followed by distillation in an all-glass still. The mobile phase was: A, 0.1% TFA in water; and B, pure acetonitrile. Gradients were run from 0 to 60% B at a flow-rate of 1 ml/min, with gradient times of 10, 20 and 60 min being used.

All samples were dissolved in 0.1% TFA at approximately 1 mg/ml; the injection volume was 10 μl. Variation of the injected mass by a factor of two in either direction resulted in no measurable change in retention time, indicating that the chromatography was being performed in a linear (or nearly linear) portion of the adsorption isotherm. All Nase retention times are the average of three measurements and are presented as the average value ± the standard deviation. The gradient retention time was taken as the time of the peak maximum, as indicated by a Model 3390 A recording integrator (Hewlett-Packard, Avondale, PA, U.S.A.).

The values of the amino acid coefficients for the various predictive scales are presented in the Appendix. Five values are estimated, because the corresponding values were not given in the cited references. The proline and arginine values in the vapor pressure scale were arbitrarily assigned to be −1.61 and −19.92, respectively. Since proline is involved in only two of the substitutions examined here, murine insulin I and turkey insulin, and arginine is involved only in the Glu43Arg* Nase

---

* The naming convention is as follows: the first three letters are the three letter code for the amino acid in the WT Nase; the number is the residue number; and the final three letters are the three letter code for the amino acid substituted. Thus, Glu43Arg specifies the mutant in which the wild-type glutamate residue 43 is replaced by an arginine (i.e., Glu⁴³ → Arg⁴³).

mutant, the impact of these estimates is minimal. Arginine and lysine in the fraction 100% buried scale and arginine in the fraction 95% buried scale were assigned the value(s) corresponding to maximum exposure. These assignments affect the correlations of the Glu43Arg and Glu43Lys Nase mutants, as well as the murine I insulin (affected by the lysine value only).

Correlation coefficients were calculated using standard techniques[22–24]. For each calculation, the retention times of the mutant (or variant) groups were correlated to the numbers derived by subtracting the value(s) for the original amino acid(s) from the value(s) of the substituted one(s) for each member of the group. Obviously, the specific values for the amino acids were dependent on the predictive scale used. No attempt was made to calculate correlations across groups of proteins.

RESULTS AND DISCUSSION

Table I lists the gradient retention time of each Nase mutant for the gradient times used in this study. All of the proteins eluted at similar times, consistent with the single- and multiple-site mutants being structurally and chemically similar to one another. All of the tyrosine-to-phenylalanine mutations, the Met26Leu mutant, and the multiple-site mutations eluted later then WT in every gradient. The mutations at residue 43, which involve replacing an uncharged (*i.e.*, at the operating pH of *ca.* 2) glutamic acid by positively charged amino acids, eluted earlier than WT, as did the Ala60Gly mutant. Replacement of the tyrosine at position 85 by either serine or glutamic acid likewise resulted in earlier elution. Fig. 1 is a plot of the relative retention (defined as relative to the retention of the WT Nase) *versus* the gradient time and shows that no selectivity changes occurred over the range of gradient times used.

TABLE I

GRADIENT RETENTION TIMES OF MUTAGENIZED AND NATURALLY OCCURRING STAPHYLOCOCCAL NUCLEASE

| *Mutant*	*Gradient retention time (min)*		
	Gradient time		
	10 min	*20 min*	*60 min*
Tyr85Ser	7.61 ± 0.02	12.70 ± 0.01	32.98 ± 0.03
Ala60Gly	7.63 ± 0.00	12.71 ± 0.01	32.99 ± 0.04
Tyr85Glu	7.63 ± 0.00	12.71 ± 0.03	33.02 ± 0.01
Glu43Lys	7.65 ± 0.01	12.77 ± 0.02	33.17 ± 0.03
Glu43Arg	7.66 ± 0.00	12.80 ± 0.01	33.21 ± 0.02
Glu43His	7.66 ± 0.01	12.80 ± 0.02	33.28 ± 0.03
Wild Type	7.67 ± 0.01	12.87 ± 0.01	33.44 ± 0.00
Tyr113Phe	7.70 ± 0.01	12.88 ± 0.01	33.59 ± 0.02
Met26Leu	7.73 ± 0.00	12.96 ± 0.01	33.68 ± 0.00
Tyr85Ser + Tyr113Phe + Tyr115Phe	7.76 ± 0.01	13.01 ± 0.02	33.92 ± 0.04
Tyr85Phe	7.77 ± 0.01	13.01 ± 0.02	33.92 ± 0.01
Tyr115Phe	7.81 ± 0.01	13.10 ± 0.01	34.24 ± 0.02
Tyr113Phe + Tyr115Phe	7.84 ± 0.01	13.17 ± 0.01	34.39 ± 0.05
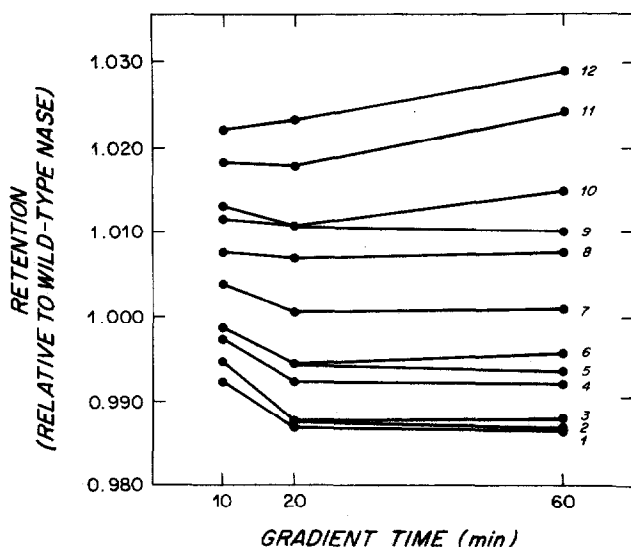
Fig. 1. The retention of mutagenized Staphylococcal nucleases (Nase) relative to the retention of the wild-type (*i.e.*, naturally occurring) protein *versus* gradient time. The chromatographic system was: A = 0.1% TFA in water, B = acetonitrile; 0–60% B at 1 ml/min; Vydac C4 TP column (5 μm particle diameter, 150 × 4.6 mm I.D.). All gradients run at room temperature. Identification: 1 = Tyr85Ser (a Ser is substituted for the naturally occurring Tyr at residue 85); 2 = Ala60Gly; 3 = Tyr85Glu; 4 = Glu43Lys; 5 = Glu43Arg; 6 = Glu43His; 7 = Tyr113Phe; 8 = Met26Leu; 9 = Tyr85Ser + Tyr113Phe + Tyr115Phe; 10 = Tyr85Phe; 11 = Tyr115Phe; 12 = Tyr113Phe + Tyr115Phe.

Before discussing these results, it is appropriate to describe briefly the known structure and solution behavior of WT Nase. WT Nase is hydrophilic at pH 7, containing 28 positively charged and 21 negatively charged residues, and is flexible, being without disulfide bonds[15,20]. It is also known to be reversibly denaturable by temperature, by extremes of pH, and by chaotropic agents (*e.g.*, urea or guanidine hydrochloride[25])[16–19]. The crystalline form shows about 24% α-helix and 14% β-sheet[15]. Further, there are no long stretches of strongly hydrophobic residues.

Because WT Nase is denatured at pH < 4[26], as well as by organic additives (and possibly by adsorption to hydrocarbonaceous surfaces[4–6]), the Nase proteins are expected to be denatured during RPLC in the present study. In conditions known to promote true random coil behavior (*i.e.*, in 6 *M* guanidine hydrochloride), the intrinsic viscosity of WT Nase is 17.0 ml g^{-1} (ref. 27). [The intrinsic viscosity of native WT Nase is 1.9 ml g^{-1} (ref. 27)]. However, Epstein *et al.* [26], in their study of the refolding kinetics, reported the intrinsic viscosity of acid-denatured Nase to be 8.0 ml g^{-1} in 0.1 *M* sodium chloride and found the measured value was sensitive to the ionic strength. These data indicate that pH-denatured Nase is not completely random coil. Additionally, spectroscopic studies suggested that other, non-Nase polypeptides undergoing RPLC are not generally random coil but rather contain secondary structure[9,28].

The differences among these possible structures is shown in Fig. 2a–c. Fig. 2a represents the peptide chain of the native, folded form of WT Nase. The organization of the secondary and tertiary structures is characteristic of the native state. Fig. 2b
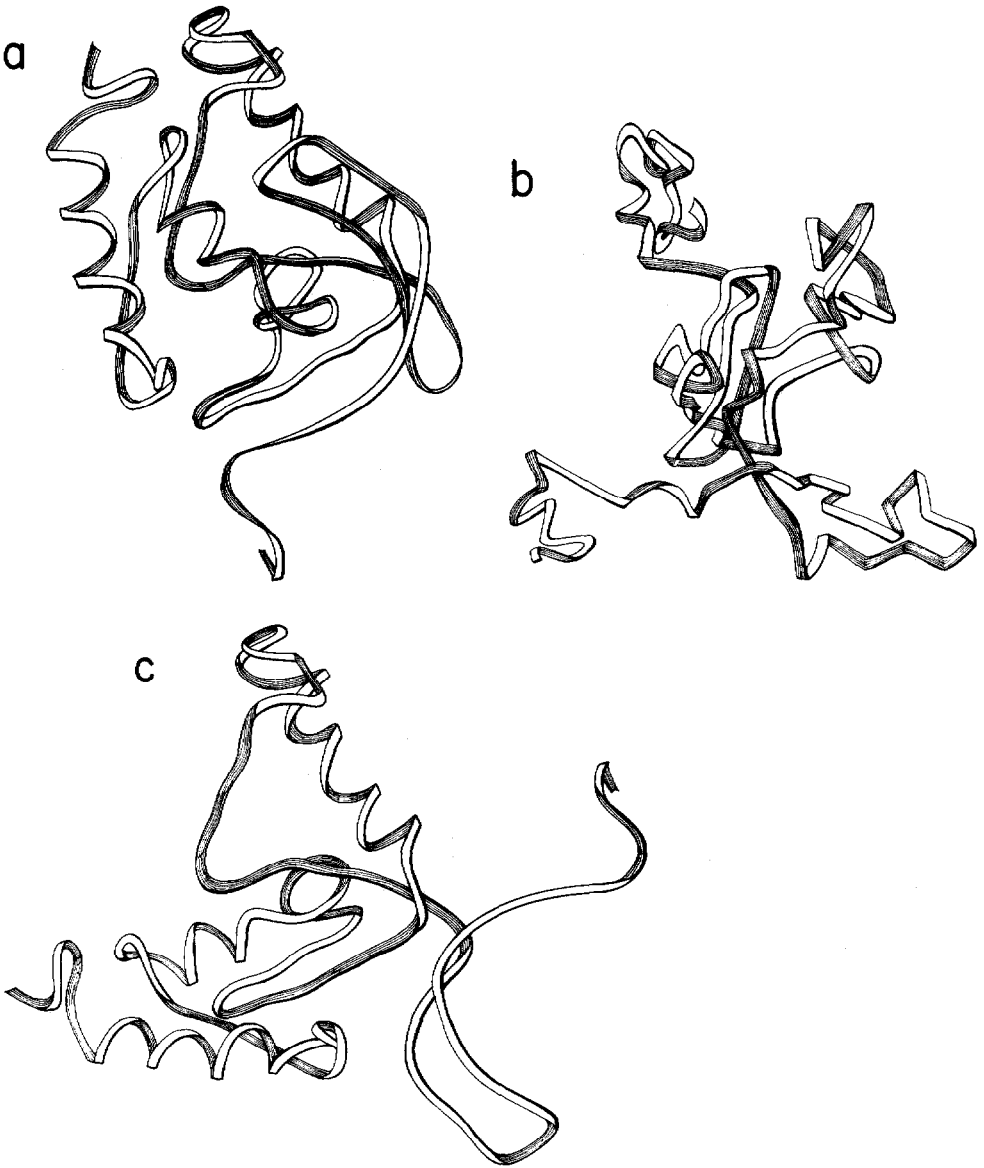
Fig. 2. Representations of possible protein structural states relevant to reversed-phase chromatography: (a) native, folded conformation; (b) random coil conformation; (c) expanded, ordered conformation. (Drawings are not to scale. Intrinsic viscosity data indicate the relative volumes, a:b:c, to be approximately $1:8:4$[26,27].

represents a random coil. The conformational space available to a residue in a random coil is limited by the energetics of rotation about bonds and the physical exclusion from areas occupied by other segments of the chain. Thus, Fig. 2b represents only one of many possibilities, and the solution behavior of a random coil would be

expected to show little or no positional or structural sensitivity to amino acid com-
position. (The flexibility/rigidity is affected by differing conformational freedom
available to different amino acids[29].) Fig. 2c represents a possible denatured form of
the protein containing secondary structure, which may or may not be identical to
that found in the native protein. This model is consistent with the existence of higher
order structure in the acid-denatured protein as demonstrated by intrinsic viscosity
measurements[26].

McLeod and Wood[11] explained certain details of the retention behavior of
insulins derived from different species by comparison to the crystal structure of por-
cine insulin. They justified this reliance on the X-ray derived structure by citing the
circular dichroism measurements of Ettinger and Timasheff[30], who earlier demon-
strated only minor conformational changes in native insulin in the presence of organic
solvent (albeit acidified methanol and chloroethanol). By contrast, intrinsic viscosity
measurements of acid-denatured WT Nase indicate significant conformational
changes[26]. Thus, interpretation of our retention data in terms of the Nase crystal
structure is not possible. Furthermore, other proteins have been demonstrated to be
denatured by RPLC under these or similar conditions[6], although the exact nature of
these denatured states was not characterized.

Retention behavior of singly substituted mutants

If the contributions of the individual residues to the overall retention were
simply additive, then the tyrosine-to-phenylalanine mutants (residues 85, 113 and
115) should exhibit identical retention. However, the retention variations observed
indicate the presence of residual (or induced) ordered structure in these proteins
under RPLC conditions, because the residue location of the substitution in the chain
is significant in determining retention behavior (Table I). This is inconsistent with a
random coil conformation, because positional effects would be expected to be small
or non-existent due to the conformational flexibility inherent in such a random coil
state (Fig. 2b). In agreement with the idea of residual or induced structure in poly-
peptides, other workers have observed that retention prediction methods generally
do not apply accurately to peptides of twenty or more residues and have attributed
this to unspecified structural effects[31–33]. The model of a denatured protein, contain-
ing structured regions (Fig. 2c), explains the observed retention variations of the set
of tyrosine-to-phenylalanine mutants and also accommodates the strong denaturing
effects of RPLC.

The three mutants at residue 43 (substitution of Lys, Arg or His for Glu) are
chemically similar, while those at position 85 (Phe, Ser, or Glu substituted for Tyr)
are more dissimilar. The variation in retention time for the latter set was statistically
indistinguishable from the variation for the entire set of mutant proteins, while that
for the mutants at residue 43 was distinguishable at the 90% confidence level. Thus,
chemically dissimiliar substitutions at residue 85 are sufficient to cause retention vari-
ation as great as that seen for the entire set of twelve mutants, while chemically
similar substitutions at residue 43 are not sufficient.

Retention behavior of multiply substituted mutants

Retention prediction methods have been developed which allow reasonably
accurate calculation of the retention time of a peptide, based on a knowledge of the

TABLE II

INFLUENCE OF MULTIPLE SUBSTITUTIONS ON THE CHROMATOGRAPHIC RETENTION OF MUTAGENIZED STAPHYLOCOCCAL NUCLEASE

Abbreviations: Exp. = experimentally observed retention shift, Calc. = calculated retention shift; Diff. = difference; NA = not applicable.

Mutant	Gradient retention time shift* (min)								
	10 min gradient			20 min gradient			60 min gradient		
	Exp.	Calc.	Diff.	Exp.	Calc.	Diff.	Exp.	Calc.	Diff.
Tyr85Ser	−0.08 ± 0.03	NA	NA	−0.17 ± 0.02	NA	NA	−0.46 ± 0.03	NA	NA
Tyr113Phe	0.03 ± 0.02	NA	NA	0.01 ± 0.02	NA	NA	0.15 ± 0.02	NA	NA
Tyr115Phe	0.14 ± 0.02	NA	NA	0.23 ± 0.02	NA	NA	0.80 ± 0.02	NA	NA
Tyr113Phe + Tyr115Phe	0.17 ± 0.02	0.17 ± 0.04	0.00	0.30 ± 0.02	0.24 ± 0.04	0.06	0.95 ± 0.05	0.95 ± 0.04	0.00
Tyr85Ser + Tyr113Phe + Tyr115Phe	0.09 ± 0.02	0.09 ± 0.05	0.00	0.14 ± 0.03	0.13 ± 0.04	0.01	0.48 ± 0.04	0.49 ± 0.08	0.01

* The gradient retention time shift is defined as the retention time of the mutagenized protein minus the retention time of the naturally occurring protein.

amino acid composition[31-33]. In comparing the experimental retention time to the calculated value, Guo *et al.*[34] found a maximum error of 4.6 min for a set of 58 peptides having 2–16 residues. Beyond 20 residues, the decreasing accuracy of the predictions was attributed to the influence of molecular weight. Our retention data for the tyrosine-to-phenylalanine mutations show that structural effects must be considered in attempting to apply predictive systems of this type to larger polypeptides. All of the tyrosine-to-phenylalanine mutants have identical amino acid compositions and identical molecular weights, yet the retention variation is as great as for the entire set of thirteen proteins.

The fundamental idea of such retention prediction methods is that the retention of a peptide can be considered to be the sum of individual, independent contributions from each residue. Extending this idea and taking into account the positional effects (presumably due to secondary structure), the shift in retention relative to the retention of the WT Nase for a multiply substituted Nase mutant should be equal to the sum of the retention shifts of each of the corresponding singly substituted mutants (*e.g.*, the retention shift of the Tyr113Phe + Tyr115Phe mutant would be equal to the sum of the retention shifts of the Tyr113Phe and the Tyr115Phe mutants).

The excellent agreement between the calculated and experimental retention shifts shown in Table II suggests that the solution structure of the multiply substituted mutants is similar to that of the singly substituted protein in the region of each mutation (*i.e.*, each mutation probably represents only a local structural perturbation). The logical implication is that polypeptide retention prediction methods in which individual contributions are summed without considering polypeptide chain location will fail for sufficiently large polypeptides. Further, any peptide retention prediction methods which account for structural effects in a polypeptide may fail when the structure of the polypeptide varies significantly from the model polypeptide.

Correlation with amino acid-based predictive methods

A number of methods are used for characterizing the behavior of amino acid residues within peptides and proteins, and it seemed reasonable that the RPLC retention might correlate to one or more of these. In addition to the retention data for the Nase mutants, we also calculated correlation coefficients* for the RPLC retention of the species variants of insulin reported by McLeod and Wood[11] and the human recombinant IL-2 mutants reported by Kunitani *et al.*[14]. The latter included data for both oxidized and reduced proteins. In each case, the naturally occurring form (or for insulin, the bovine form) is used as the reference compound, and only the altered residues are used to form the correlation. These proteins are listed in Table III (insulins) and IV (IL-2s).

Before presenting the results of correlation analysis, it is appropriate to describe the differences in these proteins. As mentioned above, Nase is flexible, without disulfide bridges, and hydrophilic, with charged residues distributed throughout the

---

* The correlation coefficient, r, does not directly indicate the degree of correlation; it is necessary to compare the value to that derived from a completely uncorrelated population[22,24]. This interpretation yields the probability of the existence of a linear correlation. As such, values of r much smaller than 1.00 can indicate significant correlation, depending on the number of data.

TABLE III

AMINO ACID SUBSTITUTIONS IN SPECIES VARIANT INSULINS

Data from McLeod and Wood[11]. Bovine insulin is reference protein.

Name	Residue* number	Residue in reference protein	Residue in species variant
Human insulin	A8	Ala	Thr
	A10	Val	Thr
	B30	Val	Ile
Porcine insulin	A8	Ala	Thr
	B30	Val	Ile
Murine insulin I	A4	Glu	Asp
	A8	Ala	Thr
	A10	Val	Ile
	B3	Asn	Lys
	B9	Ser	Pro
	B30	Ala	Ser
Murine insulin II	A4	Glu	Asp
	A8	Ala	Thr
	A10	Val	Ile
	B3	Asn	Lys
	B29	Lys	Met
	B30	Ala	Ser
Turkey insulin	A8	Ala	His
	A9	Ser	Asn
	A10	Val	Pro
	B1	Phe	Ala
	B2	Val	Ala
	B27	Thr	Ser

* A and B refer to the insulin A and B chains respectively.

molecule. In contrast, insulin consists of two peptide chains (A and B chains) joined by interchain disulfide bonds and has a relatively hydrophobic core. Moreover, in describing circular dichroism changes observed for insulin in the presence of organic solvents, McLeod and Wood[11] say "these have been attributed to changes in the environment of tyrosine residues on dissociation of the dimer by the solvent, some minor conformational alterations and an increase in the α-helical content." The primary structure of IL-2 is presented by Kunitani et al.[14]. Examination of the protein sequence shows a relatively large number of charged residues and that a 54 residue sequence (from residues 80 to 133) contains eleven proline, isoleucine and valine residues, which have restricted conformational space available[29]. By contrast, WT Nase has eighteen Pro, Ile and Val residues distributed within the entire 149-residue protein. The presence of the Pro, Ile and Val residues should result in a relatively rigid structure for reduced (i.e., lacking the disulfide bridge) IL-2, although presumably not so rigid as the intramolecularly crosslinked insulin and oxidized IL-2. Additionally, the set of oxidized IL-2 mutants is likely to contain major structural perturbations, since misalignment of the disulfide bridge would be expected to distort

TABLE IV

AMINO ACID SUBSTITUTIONS FOUND IN MUTAGENIZED INTERLEUKIN 2 (IL-2)

Data from Kunitani et al.[14]. Human IL-2 is reference protein.

Name	Residue number	Residue in reference IL-2	Residue in mutant IL-2
[Ser^{125}]IL-2	125	Cys	Ser
[desAla^1,Cys^{125}]IL-2	1	Ala	Deleted
[desAla^1,Ser^{125}]IL-2	1	Ala	Deleted
	125	Cys	Ser
[desAla^1,Ala^{125}]IL-2	1	Ala	Deleted
	125	Cys	Ala
[desAla^1,Ala^{104},Ser^{125}]IL-2	1	Ala	Deleted
	104	Met	Ala
	125	Cys	Ser
[desAla^1,Ser^{58}]IL-2	1	Ala	Deleted
	58	Cys	Ser
[desAla^1,Ser^{105}]IL-2	1	Ala	Deleted
	105	Cys	Ser

significantly the protein's structure. (In the native IL-2, the disulfide bond exists between Cys^{58} and Cys^{105}, and there is a free sulfhydryl at Cys^{125}.) In this sense, the oxidized IL-2 mutants are more dissimilar to each other than the reduced IL-2 mutants, the Nase mutants or the insulin variants. Thus, the flexibility of the proteins might be expected to decrease in the order Nase > reduced IL-2 > oxidized IL-2 > insulin.

Table V is a listing of the correlation coefficients for the retention times of each of the protein groups correlated to predictive scales developed by Meek[31], Meek and Rossetti[32] and Guo et al.[33]. As mentioned previously, the value of the correlation coefficient does not indicate the significance of the correlation. Thus, the per cent probability of these coefficients being derived from a totally uncorrelated distribution is also shown in Table V. In the following discussion, "significantly correlated" is defined as having at least a 90% confidence for the value of the correlation coefficient (i.e., at most 10% probability of a random correlation). While correlation does not imply causality, the examination of correlations for sets of such closely related proteins reveals the nature of individual residue contributions to retention by minimizing the diverse structural effects which would be encountered by studying sets of less similar proteins.

The scales presented in Tables V are based on analyses of the RPLC retention of peptides. All three of the scales by Meek and co-workers[31,32] are derived from data collected using low pH mobile phases and octadecyl-modified RPLC columns. The first two scales are based on a perchlorate–perchloric acid buffer and differ by the number of peptides used to determine the individual amino acid contributions. Generally, the correlation was better with the third scale, which used data from a phosphate–phosphoric acid buffer. The oxidized forms of IL-2 showed the poorest correlation and were not significant at the 90% confidence level. The Nase data was significantly (>99.9%) correlated, and the reduced IL-2 was correlated at >90%

TABLE V

CORRELATION OF REVERSED-PHASE RETENTION OF PROTEIN GROUPS TO CHROMATOGRAPHICALLY DERIVED AMINO ACID-BASED PREDICTIVE SCALES

Abbreviations: r is the correlation coefficient; P_u is the probability of the value of the coefficient being obtained from uncorrelated data.

Scale	Protein group (number of variants)							
	Nuclease*(13)		Interleukin 2 (8)				Insulin (6)	
			Reduced form		Oxidized form			
	r	P_u	r	P_u	r	P_u	r	P_u
ClO₄								
25 Peptide Database**	0.636	< 2%	0.808	< 2%	0.267	>10%	0.728	>10%
100 Peptide Database***	0.821	<0.1%	0.638	<10%	0.561	>10%	0.709	>10%
PO₄ 100 Peptide Database***	0.901	<0.1%	0.657	<10%	0.590	>10%	0.730	<10%
ClO₄, pH 7§	0.835	<0.1%	0.652	<10%	0.507	>10%	0.796	<10%
0.1% TFA, pH 2§	0.938	<0.1%	0.639	<10%	0.493	>10%	0.806	<10%

    * Staphylococcal nuclease (Nase).
   ** Data from Meek[31].
 *** Data from Meek[31] and Meek and Rossetti[32].
    § Data from Guo et al.[33].

significance to each scale. The insulin retention correlated significantly only to the phosphate-based scale. The scales of Guo et al.[33] are derived from data acquired using wide-pore, octyl-modified RPLC columns, with either a perchlorate buffer, pH 7 or a 0.1% TFA buffer, pH 2. Both systems showed significant correlation to all the protein groups except that of the oxidized IL-2 mutants; the Nase retentions are the most significantly correlated (>99.9%). These data suggest that protein chain flexibility enhances correlations to scales of this type. However, since the scales are derived from chromatographic data, the correlations suggest only that the increased flexibility allows the individual residues to interact with the stationary phase in a fashion more similar to small peptides.

Table VI lists the corresponding correlation values for two scales used to profile protein hydrophilicity and the four scales from which they were derived. First presented is the solvent parameter scale of Levitt[35], based on the octanol–water partitioning data of Nozaki and Tanford[36]. Although O'Hare and Nice[37] found that their peptide retention data did not generally correlate well with solvent parameter data, we observe a significant correlation between retention and the solvent parameter scale for all of the protein groups except for the Nase mutants. Negative correlation coefficients are found for all of the proteins using this scale, indicating that the more hydrophilic substitutions result in shorter retentions.

The hydrophilicity scale of Hopp and Woods[38] (Table VI) was proposed to predict hydrophilic regions in a protein, which would presumably correlate with sur-

face-exposed, antigenic regions. Their scale is a modification of the solvent parameter scale, with increased hydrophilicity values for proline, aspartic acid and glutamic acid. Although these modifications improved the ability of the scale to predict antigenic determinants[38], they decrease somewhat the correlation to RPLC retention.

Again, the retention of the highly flexible Nase group is not significantly correlated to this scale, while the retention of the presumably flexible reduced IL-2 mutants is correlated. The significance of the correlation for the more inflexible proteins (i.e., the oxidized IL-2 mutants and the insulin variants) was between that for the reduced IL-2 and Nase mutants. Thus, flexibility is not obviously related to the degree of correlation for either of these scales.

Kyte and Doolittle[39] also proposed a hydrophilicity scale and introduced the term "hydropathy" to indicate the hydrophobicity/hydrophilicity of the peptide chain. Their scale was based on data from Wolfenden et al.[40] and from Chothia[41]. The former data were vapor pressure measurements of side-chain analogues to amino acids, and the latter were measurements of the likelihood of side-chain burial in model globular proteins, based on X-ray data. Kyte and Doolittle's hydropathy scale[39] correlated significantly only to the retention times of the Nase mutants with the positive sign of the correlation coefficient reflecting the nature of the scale's definition (i.e., increasing hydropathy corresponds to decreasing hydrophilicity) (Table VI). In contrast, the vapor pressure scale[40] correlated significantly to the retention data of the Nase mutants, the oxidized IL-2 mutants and the insulins. However, the sign of the correlation coefficient for the insulin retention times was positive; that for the other solute sets was negative. Thus, in the case of insulin, a substitution which

TABLE VI

CORRELATION OF REVERSED-PHASE RETENTION OF PROTEIN GROUPS TO AMINO ACID-BASED, HYDROPHILICITY PREDICTIVE SCALES

Scale	Protein group (number of variants)							
	Nuclease* (13)		Interleukin 2 (8)				Insulin (6)	
			Reduced form		Oxidized form			
	r	P_u	r	P_u	r	P_u	r	P_u
Solvent Parameter**	−0.359	>10%	−0.750	< 5%	−0.652	<10%	−0.760	<10%
Hydrophilicity***	−0.309	> 10%	−0.750	< 5%	−0.652	<10%	−0.716	>10%
Hydropathy§	0.841	<0.1%	0.513	>10%	0.198	>10%	0.712	>10%
Vapor Pressure§§	−0.728	< 1%	−0.458	>10%	−0.679	<10%	0.880	< 5%
Fraction Buried§§§								
100% Buried	0.746	< 1%	0.262	>10%	0.744	< 5%	0.216	>10%
95% Buried	0.766	< 1%	0.530	>10%	0.676	<10%	0.688	>10%

 * Staphylococcal nuclease (Nase).
 ** Data from Levitt[35].
 *** Data from Hopp and Woods[38].
 § Data from Kyte and Doolittle[39].
 §§ Data from Wolfenden et al.[40].
 §§§ Data from Chothia[41], as given in Kyte and Doolittle[39].

would correspond to a decrease in vapor pressure (*i.e.*, greater attraction to an aqueous environment) coincided with greater retention, which is opposite to the expected behavior. The scale of Chothia[41] correlated to the retention times of the Nase mutants and the oxidized IL-2 mutants, for both the 100% and the 95% buried scales. If this probability of a residue being buried is inversely related to that residue's hydrophilicity, then substituting an inherently hydrophobic residue (*i.e.*, a residue which is likely to be buried) into either the flexible Nase structure or the grossly perturbed oxidized IL-2 structure should result in an increased retention. However, since the naturally occurring insulin species variants are required to be biologically active, they might be expected to be folded to minimize the exposure of inherently hydrophobic groups (and thereby provide the greatest conformational stabilization). The lack of correlation of the insulin variant retention to either of these last two scales is possibly due to the subtle variations in the folding of these variants.

Table VII gives the correlation coefficients and their significance for several diverse scales. Grantham[42] proposed that the evolutionary distance between proteins could be measured by quantitating the differences in their sequences. This quantitation involves three separate contributions: polarity, composition and molecular side-chain volume. The polarity scale was significantly and inversely correlated to the retentions of the Nase mutants and the insulin variants. This inverse correlation is understandable, since more polar substitutions are expected to be more hydrophilic. The lack of significance of such a correlation in the case of the IL-2 mutants, particularly for the reduced forms, is not readily explicable. (In general, the correlations

TABLE VII

CORRELATION OF REVERSED-PHASE RETENTION OF PROTEIN GROUPS TO ADDITIONAL AMINO ACID-BASED PREDICTIVE SCALES

Scale	*Protein group (number of variants)*							
	Nuclease[*] *(13)*		*Interleukin 2 (8)*				*Insulin (6)*	
			Reduced form		*Oxidized form*		*r*	*P_u*
	r	*P_u*	*r*	*P_u*	*r*	*P_u*		
Polarity[**]	−0.510	<10%	−0.468	>10%	0.052	>10%	−0.846	< 5%
Composition[**]	−0.440	>10%	−0.042	>10%	0.098	>10%	−0.661	>10%
Molecular side-chain volume[**]	0.110	>10%	0.508	>10%	0.250	>10%	0.860	< 5%
Glycine difference[**]	0.373	>10%	0.428	>10%	0.442	>10%	0.803	<10%
Standard state[***]	−0.061	>10%	0.207	>10%	0.262	>10%	0.714	>10%
Mean exposed surface area[***]	−0.666	< 2%	−0.529	>10%	−0.235	>10%	−0.524	>10%
Mean area buried[***]	0.413	>10%	0.360	>10%	0.345	>10%	0.839	< 5%
Mean fractional area change[***]	0.788	< 1%	0.130	>10%	0.432	>10%	0.801	<10%

    [*] Staphylococcal nuclease (Nase).
  [**] Data from Grantham (1974)[42].
[***] Data from Rose *et al.* (1985)[43].

of the reduced and oxidized IL-2 mutants are the least satisfactorily explainable of these protein groups. The possible origins of this phenomenon will be discussed later.)

The composition scale (which is defined as the atomic weight ratio of non-carbon elements in "end groups or rings" to carbons in the amino acid side-chain[42]) did not significantly correlate to any of the retention data (Table VII). This may reflect the arbitrary nature of the scale's definition (*i.e.*, some heteroatoms will be charged, as for Lys, and others neutral, as for Asn, yet the scale scores only by the relative number of heteroatoms). Similarly, the molecular side-chain volume scale ignores the charge differences among the side chains, which may explain why the retention times of the Nase mutants is not significantly correlated to the side-chain volume. This lack of correlation was surprising, since the exposed area is expected to increase with increasing volume of the substituted residue, and therefore the retention is also expected to increase. The presence of a significant correlation for the insulin variants is presumably due to the generally conservative nature of the changes involved. The glycine difference scale[42], a composite of the three previously mentioned scales, correlated significantly only to the insulin variants, although the correlation of these variants to the glycine difference scale was less significant than the correlation to either the polarity or the molecular side-chain volume scales.

Also presented in Table VII are the correlations to four scales described by Rose *et al.*[43]. These are also derived from X-ray data, using a set of 23 globular proteins (WT Nase is included in the set). Their first scale is termed the "standard state area" and is the weighted average accessible area of the residue Z in the tripeptide Gly-Z-Gly. These standard state areas do not correlate significantly to the retention of any of the protein groups.

Their second scale is the mean solvent-accessible surface area (*i.e.*, the exposed area of the residue in the protein). This correlated inversely to the retention times of all protein sets except the Nase mutants. The inverse correlation is intuitively satisfying, as solvophobic theory predicts that a larger exposed area will result in greater retention[1].

Their third scale is the mean area buried upon folding, which is given by the standard state area minus the mean exposed surface area. Rose *et al.*[43] state "The area a residue buries upon folding is proportional to its hydrophobic contribution to the conformational free energy, ΔG_{conf}". In contrast to the exposed surface area, the area buried correlated significantly only to the retention times of the insulin variants. This may be due to the generally poor correlation of the retention data and the standard state areas, since the mean area buried is defined in terms of the standard state area. Alternatively, this scale, which is defined in terms of hydrophobically stabilized folding, may be inappropriate for the flexible Nase mutants and the IL-2 mutants.

Their fourth scale is the mean fractional area change upon folding and is defined as the area buried upon folding divided by the standard state area. This is characterized by Rose *et al.*[43] as "the fractional accessibility is an intrinsic measure of hydrophobicity". Therefore, this scale should be applicable to either folded or unfolded proteins. As expected, the scale was positively and significantly correlated to the retention times of the Nase mutants and the insulin variants. The lack of correlation to the oxidized IL-2 mutants may be due to the probable structural variations among these proteins, while the lack of correlation to the reduced IL-2 mutants

has no apparent explanation. The significant correlation for both the denatured Nase mutants and the native-like insulin variants may reflect the similarity of the retentive forces in RPLC to those forces causing the formation of higher-order structure in proteins. In other words, substitution of residues which produce the greatest reduction in exposed area upon folding also produced the greatest increase in retention.

Several times we have pointed out the difficulty in interpreting the correlation data for the IL-2 mutants, due to the limited chemical variation in the set and the likelihood of gross structural variations among the oxidized IL-2 mutants. Therefore, the oxidized IL-2 mutants do not represent a set of proteins with limited and well-defined differences. We have included these proteins because they represent the only other set of mutagenized proteins for which RPLC retention data is reported.

Comparing the quality of the correlations across the various amino acid-based predictive scales, it is immediately apparent that the RPLC-based scales[31-33] (Table V) are overall the most significantly correlated. Only the oxidized IL-2 mutants are correlated at less than 90% significance. Although in general protein retention does not correlate well to such scales[31-33], the Nase mutants and the reduced IL-2 mutants show significant correlation, demonstrating the close physicochemical similarity among the members of each group of proteins and consequently confirming the validity of using mutagenized proteins in RPLC mechanistic studies.

The use of multiple sets of mutants (and/or species variants) does allow an assessment of the influence of flexibility on the correlation of retention to these predictive scales. The RPLC scales, for example, correlate in approximately the same order as the flexibility: Nase mutants > reduced IL-2 mutants > insulin variants. This suggests that, with the appropriate mutagenized proteins, precise studies of conformational effects in RPLC are possible. Previously, such studies have been performed using well-characterized proteins and have contributed greatly to our understanding of protein RPLC[3-8], but as the mechanistic details are worked out, it will become increasingly necessary to use more subtly varied proteins such as the Nase mutants described in this paper. Mutagenesis techniques can provide such protein models.

Parker et al.[44] have recently described the use of amino acid retention coefficients in the prediction of antigenic determinants, using a modified Hopp and Woods[45] algorithm. This method located both known antigenic determinants and surface-accessible sites (as determined by X-ray analysis) in a number of globular proteins. Our results suggest that their method will be most successful with those proteins which unfold significantly under the conditions of RPLC, as we observed significantly better correlation for the Nase mutants than for the less flexible insulin variants. It also suggests that RPLC may be useful in estimating the structural integrity of proteins, by measuring the deviation of the retention from the value calculated using the amino acid composition and the RPLC retention coefficients.

As stated above, the correlations between the X-ray derived scales of Rose et al.[43] suggest a relationship between the forces causing folding of proteins and those forces causing RPLC retention. These correlation data provide the most direct demonstration of the relationship between solvophobic retention[1] and hydrophobic forces generating higher order structure[46] currently available. Katzenstein et al.[47] have reported a direct relationship between RPLC retention and protein conformation. They have observed spectroscopic changes in alkyl-modified silica surface-

bound proteins in the presence of mixed 1-propanol–aqueous solutions and have demonstrated a correlation between the RPLC elution composition and the midpoint of the spectrally observed structural transitions. They state[47] "A direct role for 1-propanol induced conformational changes in the HPLC elution process... is suggested by the linear relationship between the amount of 1-propanol necessary to elute the protein and the amount needed to produce one-half of the total spectral change." Our data suggest that RPLC elution is due to forces which necessarily perturb higher order structure. Therefore, although the elution compositions and conformational transitions correlated[47], the correlation is due to both phenomena originating from the weakening of the hydrophobic forces caused by the organic modifier. A detailed spectral analysis of the Nase mutants (or another set of mutagenized proteins) in the adsorbed and solution states would be useful in establishing whether or not the

TABLE AI

CHROMATOGRAPHICALLY DERIVED AMINO ACID CHARACTERIZATION COEFFICIENTS

n = The number of peptides in the retention database used to calculate the retention coefficients.

Amino acid	ClO_4[*] ($n = 25$)	ClO_4[**] ($n = 100$)	PO_4[**] ($n = 100$)	ClO_4, pH 7[***] ($n = 20$)	0.1% TFA, pH 2[***] ($n = 20$)
Ala	−0.1	1.1	1.0	2.2	2.0
Arg	−4.5	−0.4	−2.0	0.9	−0.6
Asn	−1.6	−4.2	−3.0	−0.8	−0.6
Asp	−2.8	−1.6	−0.5	−2.6	0.2
Cys	−2.2	7.1	4.6	2.6	2.6
Gln	−2.5	−2.9	−2.0	−0.2	0.0
Glu	−7.5	0.7	1.1	−1.3	1.1
Gly	−0.5	−0.2	0.2	0.0	−0.2
His	0.8	−0.7	−2.2	2.2	−2.1
Ile	11.8	8.5	7.0	8.3	7.4
Leu	10.0	11.0	9.6	9.0	8.1
Lys	−3.2	−1.9	−3.0	0.0	−2.1
Met	7.1	5.4	4.0	6.0	5.5
Phc	13.9	13.4	12.6	9.0	8.1
Pro	8.0	4.4	3.1	2.2	2.0
Ser	−3.7	−3.2	−2.9	−0.5	−0.2
Thr	1.5	−1.7	−0.6	0.3	0.6
Trp	18.1	17.1	15.1	9.5	8.8
Tyr	8.2	7.4	6.7	4.6	4.5
Val	3.3	5.9	4.6	5.7	5.0
Tyr-SO$_4$	6.5	2.4	3.7		
Amino–	−0.4	4.6	0.9	−2.4	−6.9
–Amide	5.0	4.4	4.9		
Pyroglutamyl–	−2.8	2.8	2.9		
Carboxyl–	6.9	2.2	1.6	−5.2	−0.8
N-Acetyl–	3.9	6.6	3.8		

* Data from J. L. Meek[31].
** Data from J. L. Meek and Z. L. Rossetti[32].
*** Data from D. Guo et al.[33].

relationship between the elution composition and the structural transition is causal or coincidental.

Site-specific mutagenesis has been shown to be a useful tool for the study of protein RPLC. We have demonstrated the utility of using sets of similar proteins to probe protein adsorption as well as factors influencing retention. We have shown that, for proteins with precisely the same amino acid composition, the retention depends on primary and higher order structure, and that a variety of mutations at a single site can have as great an effect on the chromatographic retention as substitutions at different positions throughout the polypeptide chain. Our results suggest that

TABLE AII

NON-CHROMATOGRAPHIC AMINO ACID CHARACTERIZATION COEFFICIENTS

Residue (number in set)	Solvent parameter[a]	Hydrophilicity[b]	Hydropathy[c]	Vapor pressure[d]	Fraction buried[e]	
					100% Buried	95% Buried
Ala (397)	−0.5	−0.5	1.8	1.94	5.3	1.6
Arg (137)	3.0	3.0	−4.5	(−19.92)[f]	(−4.5)[f]	(−4.7)[f]
Asn (221)	0.2	0.2	−3.5	−9.68	−3.1	−2.7
Asp (239)	2.5	3.0	−3.5	−10.92	−2.5	−2.3
Cys (98)	−1.0	−1.0	2.5	−1.24	6.0	3.2
Gln (164)	0.2	0.2	−3.5	−9.38	−4.0	−3.6
Glu (217)	2.5	3.0	−3.5	−10.19	−2.8	−1.7
Gly (435)	0.0	0.0	−0.4	2.39	4.2	1.3
His (99)	−0.5	−0.5	−3.2	−10.23	−3.6	−1.9
Ile (255)	−1.8	−1.8	4.5	2.15	4.5	5.2
Leu (297)	−1.8	−1.8	3.8	2.28	3.2	2.8
Lys (288)	3.0	3.0	−3.9	−9.52	(−4.5)[f]	−4.2
Met (66)	−1.3	−1.3	1.9	−1.48	1.0	1.9
Phe (135)	−2.5	−2.5	2.8	−0.76	2.5	3.5
Pro (152)	−1.4	0.0	−1.6	(−1.61)[f]	−2.4	−1.8
Ser (341)	0.3	0.3	−0.8	−5.06	−0.7	−1.0
Thr (265)	−0.4	−0.4	−0.7	−4.88	−0.5	−1.0
Trp (75)	−3.4	−3.4	−0.9	−5.89	−2.4	−0.3
Tyr (181)	−2.3	−2.3	−1.3	−6.11	−3.3	−2.2
Val (348)	−1.5	−1.5	4.2	1.99	4.5	5.2

[a]   Data from M. Levitt[35].
[b]   Data from T. P. Hopp and K. R. Woods[38].
[c]   Data from J. Kyte and R. F. Doolittle[39].
[d]   Data from C. Chothia[41] as given in Kyte and Doolittle[39].
[e]   Data from R. V. Wolfenden et al.[40].
[f]   Data in parentheses are estimated.
[g]   Data from R. Grantham[42].
[h]   Polarity of the free amino acid.
[i]   Defined as the atomic weight ratio of noncarbon atoms in "end groups" to carbon atoms in the side chain.
[j]   Data from G. D. Rose et al.[43].
[k]   The standard state area is the average exposed area when glycines are on either side of the residue.
[l]   The mean exposed surface area of the residue in the reference set of 23 gobular proteins.
[m]   The mean area buried upon folding: the standard state area minus the mean exposed surface area.
[n]   The mean fractional area change upon folding: the mean area buried upon folding / the standard state area.

even with the harsh solvent conditions (low pH, organic solvents, etc.) used in this study, the presence of some higher order structure in the denatured protein contributes appreciably to the retention mechanism.

Finally, we have demonstrated that the empirical amino acid predictive scales of Wolfenden et al.[40], of Meek[31], of Meek and Rossetti[32], of Kyte and Doolittle[39], of Guo et al.[33] and of Rose et al.[43] are significantly correlated to the gradient RPLC retention times. In particular, this last-named correlation suggests a similarity between the forces responsible for maintenance of the folded structure of globular proteins and those responsible for RPLC retention. The ability to determine such

Polarity[g,h]	Composition[g,i]	Molecular side-chain volume[g]	Glycine difference[g]	Standard state area[j,k]	Mean exposed surface area[j,l]	Mean area buried[j,m]	Mean fractional area change[j,n]
8.1	0.00	31	60	118.1	31.5	86.6	0.74
10.5	0.65	124	125	256.0	93.8	162.2	0.64
11.6	1.33	56	80	165.5	62.2	103.3	0.63
13.0	1.38	54	94	158.7	60.9	97.8	0.62
5.5	2.75	55	159	146.1	13.9	132.3	0.91
10.5	0.89	85	87	193.2	74.0	119.2	0.62
12.3	0.92	83	98	186.2	72.3	113.9	0.62
9.0	0.74	3	0	88.1	25.2	62.9	0.72
10.4	0.58	96	98	202.5	46.7	155.8	0.78
5.2	0.00	111	135	181.0	23.0	158.0	0.88
4.9	0.00	111	138	193.1	29.0	164.1	0.85
11.3	0.33	119	127	225.8	110.3	115.5	0.52
5.7	0.00	105	127	203.4	30.5	172.9	0.85
5.2	0.00	132	153	222.8	28.7	194.1	0.88
8.0	0.39	32.5	42	146.8	53.7	92.9	0.64
9.2	1.42	32	56	129.8	44.2	85.6	0.66
8.6	0.71	61	59	152.5	46.0	106.5	0.70
5.4	0.13	170	184	266.3	41.7	224.6	0.85
6.2	0.20	136	147	236.8	59.1	177.7	0.76
5.9	0.00	84	109	164.5	23.5	141.0	0.86

correlations is a consequence of the limited changes effected by the single-site mutations and demonstrates the utility of using site-directed mutagenesis for unravelling the intricacies of protein-RPLC interactions.

APPENDIX

Amino acid characterization coefficients

The data presented in Tables AI and AII are the amino acid coefficients for the predictive scales used in this work. These data are collected from the original publications, with the exception of the 95% and 100% fraction buried scales of Chothia[41], which are as given by Kyte and Doolittle[39].

Only five data are estimates, as discussed in the Experimental. The value for arginine and proline in the vapor pressure scale are estimated to be -19.92 and -1.61, respectively. The values for lysine and arginine in the 95% fraction buried scale and the value for arginine in the 100% fraction buried scale are estimated to be the maximally exposed values of -4.5, -4.5 and -4.7, respectively.

REFERENCES

1 C. Horváth, W. Melander and I. Molnár, *J. Chromatogr.*, 125 (1976) 129.
2 W. Melander and C. Horváth, *Arch. Biochem. Biophys.*, 183 (1977) 200.
3 R. A. Barford, B. J. Sliwinski, A. C. Breyer and H. L. Rothbart, *J. Chromatogr.*, 235 (1982) 281.
4 S. A. Cohen, K. Benedek, Y. Tapuhi, J. C. Ford and B. L. Karger, *Anal. Biochem.*, 144 (1985) 275.
5 X. M. Lu, K. Benedek and B. L. Karger, *J. Chromatogr.*, 359 (1986) 19.
6 K. Benedek, S. Dong and B. L. Karger, *J. Chromatogr.*, 317 (1984) 227.
7 X. Geng and F. E. Regnier, *J. Chromatogr.*, 296 (1984) 15.
8 K. A. Cohen, *Ph. D. Dissertation*, Northeastern University, Boston, MA, 1982.
9 S. Y. M. Lau, A. K. Taneja and R. S. Hodges, *J. Chromatogr.*, 317 (1984) 129.
10 B. Merrifield, *Science (Washington, D.C.)*, 232 (1986) 341.
11 A. McLeod and S. P. Wood, *J. Chromatogr.*, 285 (1984) 319.
12 J. L. Fausnaugh and F. E. Regnier, *J. Chromatogr.*, 359 (1986) 131.
13 M. J. Zoller and M. Smith, *DNA*, 3 (1984) 479.
14 M. Kunitani, P. Hirtzer, D. Johnson, R. Halenbeck, A. Boosman and K. Koths, *J. Chromatogr.*, 359 (1986) 391.
15 F. A. Cotton and E. E. Hazen, Jr., in P. D. Boyer (Editor), *The Enzymes*, Vol. 4, Academic Press, New York, 3rd. ed., 1971, p. 153.
16 P. W. Tucker, E. E. Hazen, Jr. and F. A. Cotton, *Mol. Cell. Biochem.*, 22 (1978) 67.
17 P. W. Tucker, E. E. Hazen, Jr. and F. A. Cotton, *Mol. Cell. Biochem.*, 23 (1979) 3.
18 P. W. Tucker, E. E. Hazen, Jr. and F. A. Cotton, *Mol. Cell. Biochem.*, 23 (1979) 67.
19 P. W. Tucker, E. E. Hazen, Jr. and F. A. Cotton, *Mol. Cell. Biochem.*, 23 (1979) 131.
20 D. Shortle, *Gene*, 22 (1983) 181.

21 E. H. Serpersu, D. Shortle and A. S. Mildvan, *Biochemistry*, 25 (1986) 68.
22 P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw Hill, New York, 1969.
23 J. D. Hinchen, *Practical Statistics for Chemical Research*, Methuen, London, 1969.
24 H. D. Brunk, *An Introduction to Mathematical Statistics*, Xerox College Publishing, Lexington, MA, 2nd ed., 1965.
25 D. Shortle, *J. Cellular Biochem.*, 30 (1968) 281.
26 H. F. Epstein, A. N. Schechter, R. F. Chen and C. B. Anfinsen, *J. Mol. Biol.*, 60 (1971) 499.
27 H. Taniuchi and C. B. Anfinsen, *J. Biol. Chem.*, 244 (1969) 3864.
28 A. J. Sadler, R. Micanovic, G. E. Katzenstein, R. V. Lewis and C. R. Middaugh, *J. Chromatogr.*, 317 (1984) 93.
29 G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.*, 23 (1968) 283.
30 M. J. Ettinger and S. N. Timasheff, *Biochemistry*, 10 (1971) 831.
31 J. L. Meek, *Proc. Natl. Acad. Sci., U.S.A.*,77 (1980) 1632.
32 J. L. Meek and Z. L. Rosetti, *J. Chromatogr.*, 211 (1981) 15.
33 D. Guo, C. T. Mant, A. K. Taneja, J. M. R. Parker and R. S. Hodges, *J. Chromatogr.*, 359 (1986) 499.
34 D. Guo, C. T. Mant, A. K. Taneja and R. S. Hodges, *J. Chromatogr.*, 359 (1986) 519.
35 M. Levitt, *J. Mol. Biol.*, 104 (1976) 59.
36 Y. Nozaki and C. Tanford, *J. Biol. Chem.*, 246 (1971) 2211.
37 M. J. O'Hare and E. C. Nice, *J. Chromatogr.*, 171 (1979) 209.
38 T. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci., U.S.A.*, 78 (1981) 3824.
39 J. Kyte and R.F. Doolittle, *J. Mol. Biol.*, 157 (1982) 105.
40 R. V. Wolfenden, P.M. Cullis and C. C. F. Southgate, *Science (Washington, D.C.)*, 206 (1979) 575.
41 C. Chothia, *J. Mol. Biol.*, 105 (1976) 1.
42 R. Grantham, *Science (Washington, D.C.)*, 185 (1974) 862.
43 G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee and M. H. Zehfus, *Science (Washington, D.C.)*, 229 (1985) 834.
44 J. M. R. Parker, D. Guo and R. S. Hodges, *Biochemistry*, 25 (1986) 5425.
45 T. P. Hopp and K. R. Woods, *Molec. Immunol.*, 20 (1983) 483.
46 W. Kauzmann, *Adv. Protein Chem.*, 16 (1959) 1.
47 G. E. Katzenstein, S. A. Vrona, R. J. Wechsler, B. L. Steadman, R. V. Lewis and C. R. Middaugh, *Proc. Natl. Acad. Sci., U.S.A.*, 83 (1986) 4268.