Journal of Chromatography, 296 (1984) 15–30
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 379

RETENTION MODEL FOR PROTEINS IN REVERSED-PHASE LIQUID CHROMATOGRAPHY

XINDU GENG* and FRED E. REGNIER*
Department of Biochemistry, Purdue University, West Lafayette, IN 47907 (U.S.A.)

SUMMARY

This paper presents a retention model for proteins on an reversed-phase chromatography support in which retention is a function of the number (Z) of solvent molecules required to displace the solute from the surface. An equation is derived that relates the capacity factor of a protein to the displacing agent concentration and the stoichiometry of solvent–solute displacement. Experimental tests of the model indicate that each protein has a unique Z value and that Z is directly proportional to the molecular weight of a series of proteins when 60% formic acid is used as the mobile phase additive. This relationship is attributed to a direct relationship between Z and the contact surface area between polypeptide solutes and the support. Desorption curves for proteins also become more convex with increasingly molecular weight, as predicted by the retention model. In the solvent series of methanol, ethanol, propanol, the Z number decreases from the C_1 to C_3 alcohol. The Z number for any particular solvent is also related to other mobile phase additives, such as acids, and the concentration of additives.

INTRODUCTION

Although reversed-phase chromatography (RCP) is widely used in the preparation of polypeptides, there are still a number of questions regarding the retention mechanism. The best and most rigorous description of RPC retention is the thermodynamic model of Horváth *et al.*[1], which is based on solvophobic theory[2]. Unfortunately, a number of the physicochemical constants used in this treatment are either not available or only available for low-molecular-weight, non-polar solutes. A simpler model for RPC would be useful.

An additional question is how the mobile phase interacts with the surface of the RPC packing. For example, Riedo *et al.*[3] has suggested that the alkyl-bonded phase of RPC packings might act as a liquid crystal with a specific molar ratio of solvent to alkyl residues. It has also been suggested[4,5] with multiple component mobile phases that the alkyl silane stationary phase imbibes the more hydrophobic com-

* Visiting Professor from the Department of Chemistry, The Northwestern University, Xian, China.

0021-9673/84/$03.00

ponents to form a conventional stationary phase of the type encountered in liquid–liquid partition chromatography. Another theory[6] regarding the retention process holds that the alkyl silane bonded phase may be treated as if it were an equivalent mechanically held liquid phase. This approach[6] leads to "reasonable predictions of retention as a function of solute structure and mobile phase composition".

The question of how macromolecular solutes interact with the support surface is a second major question in RPC. It would be expected in the case of proteins that the geometry and distribution of hydrophobic groups on the surface of both the support and protein would influence adsorption; especially since the protein has three-dimensional structure. The large number of hydrophobic ligands exposed at a proteins surface make it highly probable that adsorption will occur through multiple amino acid residues. Since molecular dimensions of many proteins are substantially larger than those of the alkyl silane bonded phase, it is probable that only a part of a protein contacts and interacts with the stationary phase. Retention would be determined by the amino acids of the protein in this contact area. If only a portion of the protein surface is determining retention, it is questionable whether general properties of proteins such as their contribution to surface tension in the mobile phase[7] will broadly correlate with RPC retention.

The concept that proteins interact with a surface at multiple amino acid residues was first proposed by Boardman and Partridge[8] from studies of cytochrome *c* adsorption isotherms. Cytochrome *c* was estimated to interact with a substituted polystyrene cation exchange support through 6 amino acid residues. It has recently been shown that this concept is general in ion-exchange chromatography; multiple ions are required to displace most proteins from ion exchange columns[9,10]. The number of ions required was found to be unique to the protein being chromatographed and had no relationship to the total number of charges in the protein. For example, at its isoelectric point (pH 5) β-lactoglobulin behaved as if it had a negative charge of 4. The interpretation of these findings was that only areas of high charge density on the surface of a protein are involved in the electrostatic binding process responsible for retention in ion exchange chromatography. Three-dimensional structure and the arrangement of ligands in space were both implicated in the retention process.

Jennissen[11,12] has also demonstrated the involvement of multiple amino acids in the adsorption of proteins on phenyl-Sepharose columns by examining retention characteristics as a function of phenyl ligand density. Unfortunately this type of study has not been carried out on RPC columns because of the technical difficulties of varying ligand density without exposing surface silanols. The fact that hydrophobic ligand density is even greater in RPC than that of phenyl-Sepharose would indicate that multiple site association is certainly possible in RPC. However, it has not currently been established that polymers are adsorbed on RPC supports through multiple monomers.

One of the most unique characteristics of proteins adsorbed on a RPC support is the fact that changing the concentration of the organic displacing agent by a few percent can result in immediate elution[7,13]. As a result, plots of the capacity factor (k') *versus* organic displacing agent concentration (D_0) are very concave. It appears that this property of proteins is related in some gross way to their size. For example, a change of a few percent in mobile phase composition can move a 500 kilodalton (kD) protein from infinite retention to non-retention. In contrast, a decapeptide

might require a 10% change in D_0 to effect the same degree of change in retention. This unusual behavior has not been reported for small molecules (< 1 kD).

Snyder[14], Soczewinsky and Golkiewicz[15], and Slaats *et al.*[16] consider the mechanism of normal-phase chromatography to be a competition between solvent and solute for adsorption sites on the adsorbent surface, *i.e.*, adsorption of solute from the mobile phase is accompanied by displacement of solvent from the adsorbent. The concept that solvent is displaced upon solute binding has recently been proposed for RPC[17]. It remains to be determined that such stoichiometry actually exists in chromatographic processes and whether it applies to the adsorption of large molecules in RPC.

This paper examines the possibility that adsorption of a protein onto the surface of an RPC packing is accompanied by the direct displacement of a stoichiometric amount (Z) of solvent. An equation is derived that relates displacing agent stoichiometry (Z) and displacing agent concentration (D_0) to retention, based on known phenomena concerning the adsorption of both organic solvents and alkyl ligands to proteins. The paper concludes with an examination of the influence of the chemical nature of displacing agents, ion-pairing agents and temperature on solvent stoichiometry.

THEORY

It has been suggested[18] that the three-dimensional structure of a protein can be a major determinant of retention in surface-mediated separations. As a consequence, this theoretical treatment of retention in RPC will be prefaced by a discussion of protein structure as it might relate to chromatography. RPC is a surface mediated separation process in which retention is controlled by the interaction between solutes and an alkyl silane derivatized surface. Since polypeptides are often large molecules with three dimensional structure and an asymmetric distribution of amino acid residues within this structure, it may be concluded that (1) multiple hydrophobic areas or sites may occur both within proteins and at their surface, (2) only those residues that have access to the external surface may contribute to the retention process, (3) steric limitations prevent all of the groups at the surface of a protein from interacting with an RPC support simultaneously, (4) various groups and even hydrophobic areas at the surface of protein may not associate with the support with the same affinity, and (5) alterations of the three dimensional structure of a protein could change retention by modifying surface hydrophobicity.

It is generally accepted that the mobile phases used in RPC alter the three-dimensional structure of most proteins during the chromatographic process. Furthermore, during the course of gradient elution it is possible that additional alterations of solute structure may occur as the solute and support are solvated. These possibilities are treated in the development of a retention model by the following assumptions: (1) only the structural form(s) of the protein existing when it begins to be eluted from the column contribute to chromatographic retention, (2) no further structural alteration of the protein occur during elution, (3) as a protein is being eluted from an RPC column it will be associated with an average of n alkyl ligands, (4) n will be proportional to ligand density, (5) n will be proportional to the hydrophobic contact area between the solute and RPC column, (6) each ligand (L_0) on the

support surface will be solvated with an average of r solvent molecules at some solvent concentration (D_0), (7) solvation of proteins may occur while they are adsorbed on the support surface, (8) all hydrophobic residues of a protein molecule in solution taken together are solvated by a total of m solvent molecules, (9) solvation of adsorbed proteins at residues other than those in the contact area will influence retention only when protein structure and n are altered, (10) displacement of protein from an RPC support requires a stoichometric amount (Z) of solvent, and (11) protein displacement is accompanied by solvation of the contact areas on both the RPC support and protein.

Retention and elution of a protein from a RPC column will be treated as a series of equilibria between the three components of the system; (1) an alkyl silane bonded phase (L_0) on the support surface, (2) a protein (P_0) free in solution without adsorbed solvent and (3) free organic solvent (D_0). Association of these three components is represented in the equilibria outlined below.

Affinity of the desorbing agent for the hydrophobic surface of RPC supports is responsible for the formation of the ligand–solvent complex L_d.

$$L_0 + rD_0 \rightleftharpoons L_d \qquad \text{(A)}$$

The possibility that an average alkyl silane may associate with multiple (r) solvent molecules is accommodated in the equilibrium. In the case of water–methanol mobile phases, $r = 1$ with an octyl silane bonded phase over a broad range of methanol concentration[19]. With acetonitrile the number is larger.

Association of the organic solvent (D_0) with free protein (P_0) in solution forms the protein solvent complex P_m as shown in equilibrium B.

$$P_0 + mD_0 \rightleftharpoons P_m \qquad \text{(B)}$$

The number of solvent molecules adsorbed by a protein in the acidic medium used to elute RPC columns is represented by m. It should be noted that under these conditions it is anticipated that most proteins have undergone alterations of their native three-dimensional structure. A more extensive discussion of the adsorption of organic solvents at hydrophobic sites in a protein may be found in ref. 20.

A third binary complex, P_a, is formed when the protein associates with the RPC support, as shown in equilibrium C.

$$P_0 + nL_0 \rightleftharpoons P_a \qquad \text{(C)}$$

This equilibrium accommodates the possibility that a protein is adsorbed on a RPC support at multiple alkyl residues. The number of alkyl ligands associated with a molecule of protein is designated by n. As noted above in the assumptions, changes in the three-dimensional structure of the protein during the course of elution may change n.

The concept has been advanced above that the contact area between a protein and a RPC support is a fraction of the total surface area of the protein and that solvation of a protein may occur while it is adsorbed. Solvation of protein on the surface of a RPC support may be envisioned as a two-step process, the first step being solvation of hydrophobic residues not involved in adsorption and the second

being solvation of amino acids and alkyl silanes that control concomitantly adsorption and chromatographic retention. This requires that the affinity of protein for alkyl silane be greater than for solvent and is the basis for the assumption that solvation of the protein outside of the contact area plays no role in retention unless it changes protein structure. Desorption of a protein from an RPC support may be envisioned as the result of the solvation of two surfaces in the contact area; that of the support and that of the protein. The number of solvent molecules (q) that are adsorbed in the contact area on the protein when it is desorbed is a fraction (f) of the total number (m) of solvent molecules that associate with the surface of a protein in solution, *i.e.*, $q = fm$.

Solvation of the protein while it is adsorbed on the RPC support is represented by the microequilibrium D.

$$P_a + (m - q)\, D_0 \rightleftharpoons P_b \qquad \text{(D)}$$

The number of molecules of solvent required to solvate an adsorbed protein will be equal to the number (m) that solvate the molecule in solution minus the number (q) that are required to solvate the contact area.

Solvation of an RPC support surface loaded with protein will be similar to solvation of adsorbed proteins. Affinity of alkyl silane ligands for proteins will be greater than for organic solvents and result in a two step solvation process. In the first step, alkyl ligands not in the contact area will be saturated with solvent without causing desorption. The second step will displace the solute from the RPC support with concomitant solvation of alkyl groups in the contact area of the support surface. Since solvation of groups in the contact area will be identical to solvation of the support in the absence of protein (equilibrium A), there is no need to treat solvation of alkyl groups in the contact area with a separate microequilibrium.

When the concentration of organic solvent reaches some critical level, solvated protein (P_b) will be desorbed from the n surface ligands of the RPC support by some stoichometric amount (Z) of solvent, as represented in equilibrium E.

$$P_b + (nr + q)\, D_0 \rightleftharpoons P_m + nL_d \qquad \text{(E)}$$

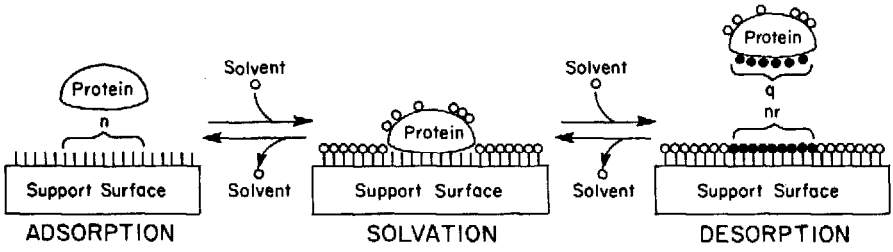

Fig. 1. A schematic diagram of the equilibria involved in gradient elution of a protein from a RPC support. This illustration shows the process to be composed of three parts: adsorption, solvation and desorption. The "bristles" on the support surface represent alkyl silane ligands, while the small circles on the surface of the protein represent adsorbed solvent molecules. Unshaded circles designate solvent molecules that make no contribution to retention when they are adsorbed on either the support or protein. In contrast, shaded circles designate solvent molecules that desorb proteins by being adsorbed in the contact area between the RPC support and protein. The symbols q, n and r are as designated in the text. The diagram shows adsorbed solvent to be in the form of a monolayer for simplicity. In real systems, solvent could be present in multiple layers.

A schematic diagram of the equilibria involved in the adsorption-desorption process can be seen in Fig. 1.

Equilibria A–E may be represented by eqns. 1–5 below.

$$K_1 = [L_d]/([L_0]\,[D_0]^r) \quad (1)$$
$$K_2 = [P_m]/([P_0]\,[D_0]^m) \quad (2)$$
$$K_3 = [P_a]/([P_0]\,[L_0]^n) \quad (3)$$
$$K_4 = [P_b]/[P_a]\,[D_0]^{m-q} \quad (4)$$
$$K_5 = [P_m]\,[L_d]^n/[P_b]\,[D_0]^{(nr+q)} \quad (5)$$

Combining eqns. 1–4 produces the expression

$$K_6 = [P_b]\,[D_0]^{(nr+q)}/[P_m]\,[L_d]^n \quad (6)$$

where K_6 is composed of the cluster of constants, $K_2K_3K_4/K_1^n$. The fact that the adsorption–desorption process is a cycle accounts for the fact that eqn. 6 is the inverse of eqn. 5. The equations indicate that displacement requires a quantity $(nr + q)$ of solvent to effect desorption that is released on readsorption. If

$$Z = nr + q \quad (7)$$

then eqn. 6 may be reduced to

$$K_6 = [P_b]\,[D_0]^Z/[P_m]\,[L_d]^n \quad (8)$$

It has been assumed in eqns. 1–4 that the structure of the protein does not change during the course of solvation. In those cases where it does, it is probable that the contact surface area will change with concomitant changes in n and q; equilibria C and D will require alterations to accomodate these changes and change the derivation of eqn. 6. It is also possible that only a portion of the solvent in the contact surface areas of both the protein and support will be desorbed on solute adsorption. However, this does not alter the fact that there will still be a quantity $(nr + q)$ of solvent molecules required to displace a protein from an RPC support. The concept that a stoichiometric quantity (Z) of solvent is displaced is still valid.

Although eqn. 8 describes the equilibrium between the solvent, a protein and an RPC column, there is no direct relationship between the concentration of these components and chromatographic retention. Retention is usually described by the capacity factor (k'). Capacity factor and solute distribution coefficient (K_d) are related by the equation

$$k' = K_d\varphi \quad (9)$$

where φ is the volume ratio of stationary to mobile phase. Since K_d is the ratio of solute concentration between the phases, it will be seen that

$$K_d = [P_b]/[P_m] \quad (10)$$

Combining eqns. 8–10 produces the expression

$$k' = ([L_d]^n\varphi K_6)/[D_0]^Z \quad (11)$$

The constants K_6 and φ may be combined to form a new constant (Q) and eqn. 11 may be rewritten as

$$k' = Q[L_d]^n/[D_0]^Z \tag{12}$$

This expression relates k' to the stoichiometry of both solvent displacement (Z) and number (n) of alkyl residues associated with a protein.

One of the unique characteristics of RPC of macromolecules is that the range $\Delta[D_0]$ of solvent concentration between which a protein first begins to migrate $[D_1]$ in a column and the concentration $[D_h]$ at which it is non-retained is small.

$$\Delta[D_h] = [D_h] - [D_1] \tag{13}$$

This fact justifies the assumption that (1) protein structure does not change and (2) the number of alkyl ligands (n) is a constant over the concentration range $\Delta[D_0]$. In addition, it is known for methanol and acetonitrile that L_d is a constant over a wide range of solvent concentration. The assumption will be made that this is also true for ethanol and propanol, which allows L_d to be treated at a constant in RPC of proteins. These assumptions allow eqn. 12 to be reduced to

$$k' = I/[D_0]^Z \tag{14}$$

where $I = Q/[L_d]^n$. The Z term is proportional to the amount of solvent required for displacement.

Experimental verification of this model and determination of Z values for proteins is facilitated by using the logarithmic form of equation 11,

$$\log k' = Z \log 1/[D_0] + \log I \tag{15}$$

where Z is the slope of a plot of the logarithm of $\log k'$ *versus* the logarithm of $1/[D_0]$.

There can be little doubt that retention of proteins at surfaces is far more complex than the simple model outlined above would indicate. For example, it is probable in the adsorption of complex macromolecules with a large number of hydrophobic regions that multiple molecular orientations can occur on the support surface. During the course of gradient elution those orientations that are retained with lower affinity would be desorbed and the molecule readsorbed in an orientation of higher binding affinity. It is envisioned that this process would continue until a smaller number of highest affinity orientations remain. There is also the matter of adsorption kinetics. Adsorption of complex macromolecules at multiple sites is probably not an instaneous process. Some degree of molecular "searching" must occur that is time dependent. When these phenomena are considered together, it is apparent that the Z and $\log I$ terms of eqn. 15 represent average values of the very complex phenomena occuring at the support surface. It is the average behavior of molecules that controls chromatographic retention.

EXPERIMENTAL

Equipment

Separations were achieved with a gradient pumping system fitted with a 50 × 4.6 mm octylsilane bonded phase column and a variable wavelength absorbance

detector. Components of the chromatographic system were two Beckman-Altex (Berkeley, CA, U.S.A.) Model 110 A pumps, an Altex solvent programmer, an Altex solvent mixer, a Rheodyne (Berkeley, CA, U.S.A.) Model 7120 injection valve fitted with a 100-μl loop, a SynChropak RP-P column (SynChrom, Linden, IN, U.S.A.), a Varian (Palo Alto, CA, U.S.A.) UV-50 Detector and a Fischer (Austin, TX, U.S.A.) Recordall 5000 recorder. The detector was operated at 278 nm with the 2-propanol–formic acid–water mobile phase and at 254 nm with other mobile phases. Column temperature was controlled within 0.5°C with a Precision Scientific Lo-Temp 154 water bath (Chicago, IL, U.S.A.).

Reagents

HPLC grade isopropanol (IPA) and methanol were obtained from Mallinckrodt (Paris, KY, U.S.A.). Ethanol was supplied by U.S. Industrial Chemicals (NY, U.S.A.). Formic acid (HF_0) was obtained from EM Industries (Gibbstown, NJ, U.S.A.).

All proteins used in this study were obtained from Sigma (St. Louis, MO, U.S.A.). The proteins used and their biological origin are as follows; ribonuclease A (bovine pancreas, type 1-A), insulin (bovine pancreas), albumin (bovine serum), cytochrome *c* (horse heart, type III), carbonic anhydrase (bovine erythrocyte), ovalbumin (chicken egg), lysozyme (chicken egg white), β-lactoglobulin (bovine) and glucagon (bovine).

Methods

Organic solvent concentrations intermediate between those of the stock solutions A and B were obtained with the solvent programmer. The column was equilibrated with 40 ml of mobile phase after each change in mobile phase concentration. During solute retention measurements the column was operated at a flow-rate of 1.0 ml/min. Recorder speeds of 1.0 and 5.0 in./min were used for measuring solute retention volume and column dead volume respectively.

Column dead volume (R_0) for each solute ($k' = 0$) was determined from plots of retention time measurements made at a series of solvent concentration between pure A and B. Minimum solute retention time was taken as the case where $k' = 0$. Capacity factor under any other set of experimental conditions was calculated from the equation

$$k' = (R_v - R_0)/R_0$$

where R_v is the retention volume under experimental conditions. All k' values used in this work were greater than 1.0 to eliminate the inherent inaccuracy of determining small values of k'. Z values were computed on the basis of solvent molarity. Regression analyses were carried out by computer.

RESULTS AND DISCUSSION

Testing the model

Retention data for 7 proteins, plotted according to eqn. 15, are presented in Fig. 2. These polypeptides were chosen because they cover a 10-fold range in molec-

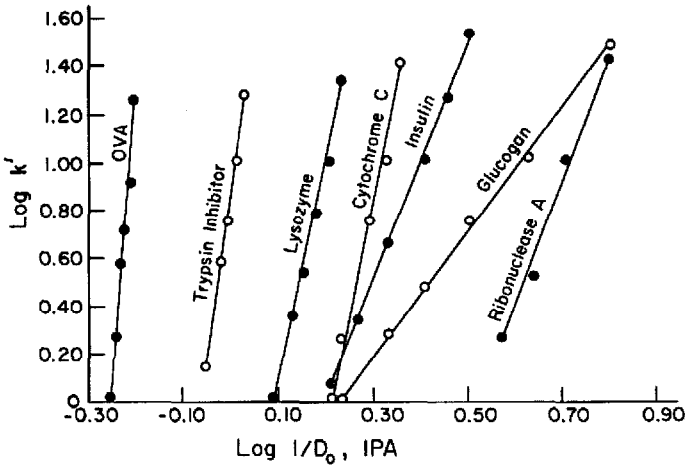

Fig. 2. Retention characteristics of a series of proteins as a function of organic solvent concentration. Fifty percent formic acid was used in all cases with 2-propanol (IPA) as the displacing agent and a 50 × 4.1 mm RP-P (C_8) column operated at 1 ml/min. The concentration of (D_0) was in terms of molarity.

ular weight and include the retention extremes encountered in RPC. A linear relationship was observed for all proteins, as predicted by the model developed above. The amount of solvent (Z) displaced upon solute adsorption, intercept (log I), correlation coefficient (C) and standard deviation of the estimate (S) for these proteins are shown in Table I. Correlation coefficients exceeded 0.99 in all cases. Slopes (Z) of all the curves were distinctly different, except for the insulin–ribonuclease A and cytochrome c–lysozyme pairs. Although the slopes of these pairs are very similar, the intercepts (log I) were quite different. Other cases will be shown in which the interrecepts of

TABLE I

RETENTION PROPERTIES OF SELECTED PROTEINS IN RPC

The displacing agent was 2-propanol in 50% formic acid. Z is the number of solvent molecules required to displace a protein from a RPC support. I is a constant related to the affinity of a protein for a RPC support. C is the correlation coefficient. S is the standard deviation of the estimate. D_1 is the solvent concentration (2-propanol) at which the protein begins to elute. D_h is the solvent concentration (2-propanol) at which a protein is no longer retained. $\Delta D_0 = D_h - D_1$. Column: 50 × 4.1 mm RP-P (C_8); 300 Å pore diameter; 6.7 μm. Mobile phase: solvent A, HF_0 (88%)–water, 50:50; solvent B, HF_0 (88%)–2-propanol–water, 50:20:30. Flow-rate: 1.0 ml/min. Temperature: 35 ± 0.5°C.

Protein	*Molecular weight*	*Z*	*Log I*	*C*	*S*	D_1 *(M)*	D_h *(M)*	$\Delta[D_0]$ *(M)*	$\frac{[D_h]}{[D_1]}$
Glucagon	3335	2.59	−0.565	0.999	0.025	0.156	0.580	0.424	3.72
Insulin	6000	4.79	−0.910	0.999	0.017	0.312	0.624	0.312	2.00
Cytochrome *c*	12,200	8.54	−1.73	0.999	0.026	0.429	0.624	0.195	1.46
Ribonuclease A	13,700	5.18	−2.72	0.995	0.013	0.156	0.390	0.234	2.50
Lysozyme	14,000	9.15	−0.820	0.995	0.052	0.585	0.8190	0.234	1.40
Trypsin inhibitor	35,000	13.8	0.890	0.999	0.028	0.897	1.131	0.234	1.26
Ovalbumin	44,000	23.8	6.08	0.995	0.052	1.600	1.794	0.194	1.12

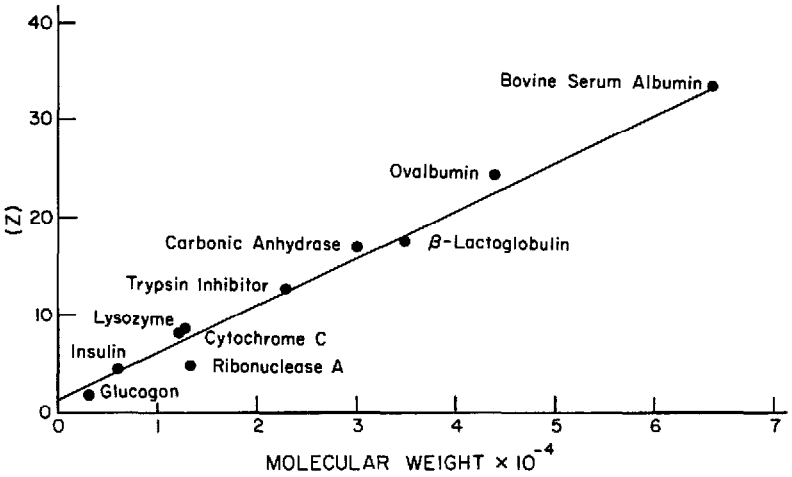

Fig. 3. The relationship between Z number and molecular weight for a series of proteins. The mobile phase contained 60% formic acid with isopropanol as the displacing agent. Experimental conditions were the same as in Fig. 2 except that the formic acid concentration in the mobile phase was increased to 60%.

two retention curves are similar but their Z values are sufficiently different to cause resolution of the proteins. Also presented in Table I is the range of desorbing agent concentration $\Delta[D_0]$ used in determining Z values for a particular protein. The term $[D_1]$ represents the lowest concentration and $[D_h]$ the highest concentration of desorbing agent required to determine a Z value. It is seen that there is an inverse relationship between the ratio $[D_h]/[D_1]$ and Z values. This means that, as the Z number of a substance increases, the range of solvent in which it will be eluted from a column becomes smaller.

For proteins of roughly the same relative percentage of hydrophobic residues it is expected that there would be a proportional relationship between Z number and molecular weight when the molecules are completely denatured. Fig. 3 shows that this is generally true, except for ribonuclease. The correlation coefficient and standard error of the estimate are 0.9968 and 0.850, respectively (except nuclease A). It should be noted that 60% formic acid was used to induce denaturation. The deviation of ribonuclease is probably due to the exceptional resistance of its three-dimensional structure to denaturing conditions. It is probable that a number of species would be found that deviate from linearity if a much larger population of proteins were tested. Such deviations could be due to either an unusually stable tertiary structure or the presence of either a very high or low percentage of hydrophobic amino acids.

Factors influencing Z and I

Solvent strength. It is known from numerous sources[21–23] that the solvent strength of alcohols increases in the order methanol < ethanol < 2-propanol. An examination of the influence of these three alcohols on Z and log I of selected proteins is seen in Table II. Both Z and I decrease with increasing solvent strength of the displacing agents. Apparently, Z decreases as the molecular volume of the alcohols becomes larger. It may be rationalized that steric effects allow fewer molecules to

TABLE II

THE INFLUENCE OF ORGANIC SOLVENTS ON Z AND log I

Column: the same as in Table I. Mobile phase: solvent A, 2-propanol–water (10:90) + 0.1% TFA, solvent B, 2-propanol–water (50:50) + 0.1% TFA; solvent A, ethanol–water (20:80) + 0.1% TFA, solvent B, ethanol–water (80:20) + 0.1% TFA; solvent A, methanol water (20:80) + 0.1% TFA, solvent B, methanol–water (80:20) + 0.1% TFA. Injected volume: 10.0 μl of a 1.0 mg/ml solution of each protein except serum albumin (bovine) which was at 3 mg/ml. Flow-rate: 1.0 ml/min. Temperature: 25 ± 0.5°C.

Protein	*Methanol*		*Ethanol*		*2-Propanol*	
	Z	*log I*	*Z*	*log I*	*Z*	*log I*
Insulin	24.4	27.9	18.5	14.4	16.6	8.13
Cytochrome *c*	51.2	59.8	43.2	35.4	32.1	17.7
Lysozyme	47.0	56.3	37.4	31.6	34.9	19.6
Serum albumin (bovine)	208	258.1	125	110	96.5	57.3

accumulate at the surface of a protein in solution, and a smaller number are displaced upon adsorption of the protein on a RPC column.

Pairing agents. Acids have been widely used in the separation of polypeptides on silica-based RPC supports for two reasons: (1) they minimize the interaction of cationic residues with the support by suppressing the ionization of underivatized surface silanols and (2) they serve as ion-pairing agents that impart new properties to polypeptides. For example, phosphoric acid has been reported[24,25] to form ion pairs with cationic groups in polypeptides that make the molecule more soluble. Acids of this type are hydrophilic pairing agents. In contrast, trifluoroacetic acid (TFA) is said to be a hydrophobic pairing agent because its ion pairs are reported[26–28]

TABLE III

THE INFLUENCE OF ACID ON Z

Column: same as in Table I. Mobile phase: (1) Phosphoric system: solvent A, 2-propanol–0.05 M KH_2PO_4 (pH 2.50) (10:90); solvent B, 2-propanol–0.5 M KH_2PO_4 (pH 2.50) (50:50). (2) Trifluoroacetic acid system: solvent A, 2-propanol–water (10:90) + 0.1% TFA; solvent B, 2-propanol–water (50:50) + 0.1 TFA. (3) Formic acid system: solvent A, HF_0 (88%)–water (60:40); solvent B, HF_0 (88%)–2-propanol–water. Injected volume: the same as in Table II. Flow-rate: 1.0 ml/min. Temperature: 35 ± 0.5°C.

Protein	*Phosphoric (0.05 M)*			*Trifluoroacetic acid (0.1%)*			*Formic acid (60%)*		
	Z	*log I*	*C*	*Z*	*log I*	*C*	*Z*	*log I*	*C*
Insulin	15.0	6.52	0.990	16.6	8.13	0.999	4.26	−0.832	0.999
Cytochrome *c*	30.5	15.2	0.992	32.1	17.7	0.996	8.12	−1.94	0.994
Lysozyme	–	–	–	34.9	19.6	0.990	8.32	−0.614	0.994
Carbonic anhydrase	45.6	28.9	0.990	36.3	23.3	0.994	17.1	2.22	0.997
β-Lactoglobulin	49.5	30.7	0.998	–	–	–	17.5	2.30	0.997
Serum albumin (bovine)	117	63.8	0.999	96.5	57.3	0.991	33.1	−0.747	0.990

to be more strongly retained on RPC columns. Although formic acid has been used very successfully in peptide separations[29,30], it has not been studied sufficiently as a polypeptide pairing agent for it to be categorized.

Table III shows the influence of phosphoric, trifluoroacetic and formic acid on both Z and log I values for a series of proteins when 2-propanol was used as the displacing agent. The results presented in the table do not agree with the currently held concepts of the function of these two agents. On the basis of the reported properties of the two pairing agents, it would be expected that Z values obtained with trifluoroacetic acid would be larger than those obtained with phosphoric acid. Contrary to what is expected, the Z values with TFA are equal to or lower than those of phosphoric acid. Carbonic anhydrase provides the greatest discrepancy.

Although it is used at much higher concentration than phosphoric and trifluoroacetic acid, the behavior of formic acid is even more anomalous. Sixty percent formic acid is reported[30] to produce much greater disruption of the three-dimensional structure of proteins than the other two acids. This would cause a much larger number of hydrophobic acids to come in contact with alkyl residues on the support and should result in the highest Z values of the three acids. Table III shows that the Z values for formic acid are 2 to 4 times lower than with the other acids.

The influence of acid concentration on both Z and log I is shown in Table IV. Z values vary inversely with acid concentration when 2-propanol is used as the displacing agent. That Z values are so much lower with formic acid and continue to decrease even when formic acid concentration is raised from 40 to 60% could possibly be due to two reasons: (1) pairing continues to increase even at high acid concentration or (2) formic acid at sufficiently high concentration is adsorbed on the surface

TABLE IV

THE EFFECT OF ACID CONCENTRATION ON Z

2-Propanol was used as the organic solvent in all cases. Column: the same as in Table I. Mobile phase: solvent A, HF_0 (88%)–water (50:50); solvent B, HF_0 (88%)–2-propanol–water (50:20:30). Injected volume: the same as in Table II. Flow-rate: 1.0 ml/min. Temperature: 35 ± 0.5°C.

Protein	*Formic acid (%)**						*Trifluoroacetic acid (%)**			
	40		*50*		*60*		*0.1*		*0.3*	
	Z	*log I*	*Z*	*log I*	*Z*	*log I*	*Z*	*log I*	*Z*	*log I*
Ribonuclease A	6.38	−2.26	5.18	−2.72	4.37	−2.1				
Insulin	5.96	−0.690	4.79	−0.910	4.26	−0.832	16.6	8.13	16.6	8.51
Cytochrome *c*	11.8	−1.40	8.54	−1.73	8.12	−1.94	32.1	17.7	3.14	18.3
Lysozyme	13.0	−0.098	9.15	−0.82	8.32	−0.614	34.9	19.6	26.5	15.8
Serum albumin (bovine)					33.1	−0.747	96.5	57.3	69.7	43.0
Carbonic anhydrase					17.1	2.22	36.3	23.3	8.13	6.01
β-Lactoglobulin	20.1	3.93			17.5	2.30				
Trypsin inhibitor	19.2	2.46	13.8	0.890	12.6	0.0926				
Glucagon	3.46	–	2.59	−0.565	1.84	0.480				
Ovalbumin		0.0550			24.5	5.90				

* Percent acid was on a v/v basis.

of the RPC support and begins to act as a displacing agent in concert with 2-propanol. If the atypical behavior of formic acid is due to a pairing effect, it must be assumed that formic acid is pairing with the solute by some non-ionic mechanism, such as hydrogen bonding at peptide bonds. This assumption is based on the fact that ion pairing with cationic species in the polypeptide would have been complete at a few percent formic acid. This would result in a substantial solubilizing effect on polypeptides and decrease the number of solvent molecules (Z) required for desorption.

If formic acid is acting as a solvent, it is expected that it would have a sparing effect on the amount of 2-propanol required for polypeptide desorption and elution. Unfortunately the experimental data fit both possibilities for the role of formic acid in RPC of polypeptides. Elucidation of the actual mechanism requires further study.

Even small changes in the concentration of TFA caused a large change in the Z values of some proteins (Table IV). Increasing the concentration of TFA from 0.1 to 0.3% caused a major reduction of Z values with bovine serum albumin (BSA) and carbonic anhydrase when 2-propanol was used as the displacing agent. In contrast, the Z values of insulin and cytochrome c were almost unchanged. These effects might be explained in any of three ways: (1) TFA continues to form more ion pairs as its concentration is increased, (2) at higher concentrations of TFA the acid is acting as a displacing agent and competing with TFA–protein ion pairs for alkyl residues on the support or (3) the hydrophobic portion of TFA begins to pair with hydrophobic regions of the protein and acts as a hydrophilic pairing agent at higher concentration. The first of these explanations is eliminated by the fact that cytochrome c is richer in basic amino acids than the other proteins but experienced a very small change in Z value. The second explanation is eliminated by the fact that TFA had a different effect on different proteins. If TFA were functioning as a displacing agent, it would have had the same relative effect on all proteins. The fact that the effect is differential suggests that some unique property of the proteins themselves is involved. Since both BSA and carbonic anhydrase are known[31,32] to be relatively hydrophobic molecules and there is a greater decrease in their Z values, the third explanation would seem to be the best.

If TFA actually has such amphiphilic properties, the anomaly of the similarity of Z values between phosphoric and trifluoroacetic acid is easily explained. At the same time that TFA is acting as a hydrophobic pairing agent with cationic species in polypeptides, it is serving as a hydrophilic pairing agent with hydrophobic residues in the protein.

Temperature. It is to be expected that the relationship between chromatographic retention and temperature will be more complex for proteins than for small molecules. Changing column temperature would alter the equilibria involved in the retention process by changing either the forces of interaction between molecular species or the conformation of the protein. Changing the force of interaction would be manifested in changes in I values alone. In contrast, alteration of the protein structure could cause both Z and I to vary.

The influence of a 50-degree temperature change on isocratic retention was found to be so large for most proteins that it was difficult to investigate changes in total free energy. Incremental effects of temperature on Z, log I and total free energy were obtained by calculating the retention of a given protein at a particular temper-

TABLE V

THE INFLUENCE OF TEMPERATURE ON Z AND I

Formic acid (50%) and 2-propanol were used as the mobile phase. Column: the same as in Table I. Mobile phase: solvent A, HF_0 (88%)–water (50:50); solvent B, HF_0 (88%)–2-propanol–water (50:20:30). Flow-rate: 1.0 ml/min. Injected volume: the same as in Table II.

Protein	*Temperature (°C)*											
	0		*10*		*20*		*30*		*40*		*50*	
	Z	*log I*	*Z*	*log I*	*Z*	*log I*	*Z*	*log I*	*Z*	*log I*	*Z*	*log I*
Ribonuclease A	6.62	1.41	6.60	1.83	6.29	2.18	6.40	2.32	4.28	2.25	2.68	1.76
Insulin	6.00	0.662	6.02	0.790	5.52	0.781	5.14	0.814	4.41	0.801	4.12	1.15
Cytochrome *c*	10.6	0.433	10.3	0.785	9.43	1.15	8.80	1.56	7.50	1.90	5.70	2.02
Lysozyme	11.4	1.18	10.7	0.728	10.6	0.220	10.2	0.334	9.05	0.935	7.82	1.55

ature from eqn. 15. These effects are shown in Table V. As expected, there is not a linear relationship between ln k' and $1/T$, but there are some regular changes. The total free energy of the four proteins investigated decreased in inverse proportion to temperature, whereas the slope of the curve increased in proportion to temperature. In addition, total free energy changed in proportion to molecular weight. Plots of Z values from Table V *versus* $1/T$ (figures not shown) are convex and of increasing slope with the elevation of temperature. This is attributed to accelerated molecular motion in the displacing agent (2-propanol) at elevated temperatures and a general decrease in the force of attraction between all of the components of the system. Of course, the effect would be expected to increase with molecular size. Plots of the log I values from Table V *versus* $1/T$ (figures not shown) are more complex. Since log I is composed of a cluster of constants, many of which are subject to alteration by temperature, it is not possible to attribute the changes in log I to any one effect. Log I was inversely related to temperature in all cases except ribonuclease A in the temperature range 40–50°C.

CONCLUSIONS

A stoichiometric displacement model has been developed in which chromatographic retention of proteins in RPC is a function of the displacement of a stoichiometric amount (Z) of solvent from the solute and support surface. It is concluded that (1) retention in an RPC column is described by the simple equation $k' = I/[D]^Z$, (2) that the stoichiometric displacement model is consistent with the experimental findings presented in this paper, (3) that Z is unique for each protein and (4) that Z is proportional to the molecular weight of the proteins with the same ratio of hydrophobic to hydrophilic amino acids.

Although the experimental findings are consistent with the stoichiometric retention model, it will need to be corroborated by several other methods.

GLOSSARY OF TERMS

C = Correlation coefficient.
D_h = Displacing agent concentration at which k' approaches zero.
D_1 = Displacing agent concentration at which solute begins to migrate in columns.
D_0 = Displacing agent concentration (moles/l) in the mobile phase.
f = q/m.
I = $Q/[L_d]^n$ is a constant that is unique for each substance.
k' = $K_d\varphi$ = Capacity factor.
K_d = The distribution coefficient for a substance.
L_0 = Ligand density (moles/m^2) on a support surface.
L_d = Density (moles/m^2) of ligand–solvent complex on the support surface.
m = Number of solvent molecules adsorbed by a protein in the mobile phase.
P_0 = Concentration (moles/l) of non-solvated protein in the mobile phase.
P_m = Concentration (moles/l) of solvated protein in the mobile phase.
P_a = Concentration (moles/m^2) of non-solvated protein adsorbed on the support surface.
P_b = Concentration (moles/m^2) of solvated protein adsorbed on the support surface.
q = The number of solvent molecules adsorbed at the contact surface area of the protein.
Q = A constant = $I\,[L_d]^n$.
r = Number of solvent molecules associated with an alkyl silane residue.
R_0 = Column dead volume.
R_v = Solute retention volume.
S = Standard deviation.
T = Temperature (°K).
Z = Number of solvent molecules displaced when a protein adsorbs.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance of Randy Drager, Ed Pfannkoch, Bill Kopaciewicz and Jodi Fausnaugh in the development of the model presented here. This research was funded by NIH Grant No. GM 25431. This is Journal Paper No. 9826 from the Purdue University Agricultural Experiment Station.

REFERENCES

1 Cs. Horváth, W. Melander and I. Molnár, *J. Chromatogr.*, 125 (1976) 129.
2 O. Sinanoğlu, in B. Pullman (Editor), *Molecular Associations in Biology*, Academic Press, New York, 1968, pp. 427–445.
3 F. Riedo, M. Czencz, O. Liardon and E. S. Kováts, *Helv. Chim. Acta*, 61 (1978) 1912.
4 J. J. Kirkland, *J. Chromatogr. Sci.*, 9 (1971) 206.
5 J. H. Knox and A. Pryde, *J. Chromatogr.*, 112 (1975) 171.
6 L. R. Snyder and J. J. Kirkland, *Introduction to Modern Liquid Chromatography*, Wiley-Interscience, New York, 2nd ed., 1979.
7 R. A. Barford, B. J. Sliwinski, A. C. Breyer and H. L. Rothbart, *J. Chromatogr.*, 235 (1982) 281.
8 N. K. Boardman and S. M. Partridge, *Biochem. J.*, 59 (1955) 543.

9 W. Kopaciewicz, M. A. Rounds, J. Fausnaugh and F. E. Regnier, *J. Chromatogr.*, 266 (1983) 3.
10 M. A. Rounds and F. E. Regnier, *J. Chromatogr.*, 283 (1984) 37.
11 H. P. Jennissen, *Biochemistry*, 15 (1976) 5683.
12 H. P. Jennissen, *J. Chromatogr.*, 159 (1978) 71.
13 M. T. W. Hearn and B. Grego, *J. Chromatogr.*, 255 (1983) 125.
14 L. R. Snyder, *Principles of Adsorption Chromatography*, Marcel Dekker, New York, 1968.
15 E. Soczewinski and W. Golkiewicz, *Chromatographia*, 6 (1973) 269.
16 E. H. Slaats, J. C. Kraak, W. J. T. Brugman and H. Poppe, *J. Chromatogr.*, 149 (1978) 255.
17 X. Geng and F. E. Regnier, *Proc. Am. Pept. Assoc.*, (1984) in press.
18 F. E. Regnier, *Science*, 222 (1983) 445.
19 R. M. McCormick and B. L. Karger, *Anal. Chem.*, 52 (1980) 2249.
20 C. Tanford, *The Hydrophobic Effect*, Wiley-Interscience, New York, 2nd ed., 1980.
21 A. F. M. Barton, *Chem. Rev.*, 75 (1975) 731.
22 B. L. Karger, L. R. Snyder and C. Eon, *J. Chromatogr.*, 125 (1979) 71.
23 L. R. Snyder, J. W. Dolan and J. R. Gant, *J. Chromatogr.*, 165 (1979) 3.
24 W. S. Hancock, C. A. Bishop, L. J. Meyer and D. R. K. Harding and M. T. W. Hearn, *J. Chromatogr.*, 161 (1978) 291.
25 M. T. W. Hearn, W. S. Hancock, J. G. R. Hurrell, R. J. Fleming and B. Kemp, *J. Liq. Chromatogr.*, 2 (1979) 919.
26 W. C. Mahoney and M. A. Hermodson, *J. Biol. Chem.*, 255 (1980) 11199.
27 H. P. J. Bennett, A. M. Hudson, C. McMartin and G. E. Purdon, *Biochem. J.*, 168 (1977) 9.
28 C. E. Dunlap III, S. Gentleman and L. I. Lowney, *J. Chromatogr.*, 160 (1978) 191.
29 G. E. Gerber, R. J. Anderegg, W. C. Herlihy, C. P. Gray, K. Biemann and H. G. Khorana, *Proc. Nat. Acad. Sci. U.S.A.*, 76 (1979) 227.
30 J. Heukeshoven and R. Dernick, *J. Chromatogr.*, 252 (1982) 241.
31 Dayhoff, *Atlas of Protein Sequence Structure*, 5 (1978) 306.
32 R. J. Tanis, R. E. Ferrell and R. E. Tashian, *Biochim. Biophys. Acta*, 371 (1974) 534.