Effects of Protein Microstructure on the Retention Time of T4 Lysozyme Variants in Cation Exchange Chromatography

ARTICLE *in* ANALYTICAL CHEMISTRY · FEBRUARY 2009

Impact Factor: 5.64 · DOI: 10.1021/ac8019812 · Source: PubMed

3 AUTHORS, INCLUDING:

Chenming Zhang
Virginia Polytechnic Institute and State Uni…

59 PUBLICATIONS **1,048** CITATIONS

SEE PROFILE

Article

Effects of Protein Microstructure on the Retention Time of T4 Lysozyme Variants in Cation Exchange Chromatography

Lakshmi P. Pathange, David R. Bevan, and Chenming Zhang

Anal. Chem., **2009**, 81 (2), 649-655• DOI: 10.1021/ac8019812 • Publication Date (Web): 10 December 2008

Anal. Chem. **2009,** *81,* 649–655

Effects of Protein Microstructure on the Retention Time of T4 Lysozyme Variants in Cation Exchange Chromatography

Lakshmi P. Pathange,[†,‡] David R. Bevan,[§] and Chenming Zhang*[,†]

Department of Biological Systems Engineering and Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

The structure of a protein directly affects its function. Therefore, characterization of recombinant protein structures is important but is a challenging task. One of the important forces that play a major role in maintaining both structural and functional properties of proteins is electrostatic interactions among different amino acid residues. In this article, cation exchange chromatography was used to study how the microstructure of some charged amino acid residues may affect a protein's retention. Two sets of T4 lysozyme variants were generated. The first set included seven variants that varied in their charge distribution. These variants were obtained by replacing a charged amino acid residue at different sites on lysozyme. The second set included ten variants that varied in both net charge and charge distribution, and these variants were obtained by replacing charged and neutral amino acid residues at different sites on the protein. The microstructure was quantified by calculating the relative hydrophilicity around the replacing amino acid residue. The retention times of all variants were compared with the retention time of the respective control variant. Among the first set of variants, there was a direct correlation ($R^2 = 0.93$) between the relative hydrophilicity of the replaced amino acid and the protein's retention time, except for two variants (K83H and K124H) whose replacing amino acid residue was involved in intramolecular interactions. For the second set, there was a direct correlation ($R^2 = 0.97$) between the change in net charge (−2 to +2 units) and the retention times. However, the retention times of two variants (R76D and R76S) did not follow the correlation. We hypothesize that the structure around the replacing charged group is responsible for the deviated protein retention pattern.

Characterization of protein structure is one of the most important issues in the development of biotherapeuticals.[1,2] The rapid growth of the biopharmaceutical industry has resulted in a greater demand to develop/evolve new and reliable techniques to characterize recombinant therapeutic proteins. For example, because of the lack of proper characterization techniques and rising health care costs in the U.S., there is an intense debate over the need to use "biogenerics" or "off-patent biotech products".

Unlike small compounds, biopharmaceuticals are an ensemble of large, flexible, and complex molecules. As a result, the process of producing biopharmaceuticals is a complex process.[3,4] The inherent challenges associated with biopharmaceutical production are (1) the type of the extraneous expression host, which expresses proteins that are a heterogeneous mixture, consisting of variants of the natural protein,[5] (2) the type of manufacturing process, which is very complex compared to those for conventional pharmaceutical products, involving specialized techniques such as fermentation, extraction, and purification,[6,7] and (3) the need to conduct expensive clinical trials to determine the efficacy and toxicity of the biopharmaceuticals.[1] For example, the biopharmaceutical IFN-β_{1a}, used for treating multiple sclerosis, when produced at two different manufacturing sites (with the same process conditions), elicited different immunogenic response in humans during clinical trials.[8] Evidently, such results have reinforced the argument that "process is the product"[6] in the biopharmaceutical industry. Consequently, characterizing these biopharmaceuticals is extremely important.

Therefore, there is a urgent need to augment the current array of analytical techniques by either developing new techniques or extending the currently available techniques such that they can characterize subtle structural changes taken place in proteins.[4,5,7] Currently, several analytical techniques (e.g., ion exchange and gel chromatography, mass spectrometry, gel electrophoresis) are used for protein characterization based on its physicochemical parameters (e.g., size, charge, molecular weight).[7] Previously we were able to quantify subtle structural variations in a model protein, T4 lysozyme, and its variant using immobilized metal affinity chromatography (IMAC).[9,10] For IMAC studies, two categories of structural variants were generated via site-directed

* To whom correspondence should be addressed. E-mail: cmzhang@vt.edu. Phone: (540) 231-7601. Fax: (540) 231-3199.

† Department of Biological Systems Engineering.

‡ Current address: Bayer Healthcare, 800 Dwight Way, Berkeley, California 94710.

§ Department of Biochemistry.

(1) Koths, K. *Curr. Opin. Biotechnol.* **1995**, *6*, 681–687.
(2) Kleinberg, M.; Mosdell, K. W. *Am. J. Health-Syst. Pharm.* **2004**, *61*, 695–708.
(3) Schellekens, H. *Trends Biotechnol.* **2004**, *22*, 406–410.
(4) Schellekens, H. *Nephrol., Dial., Transplant.* **2005**, *20*, 31–36.
(5) Chicz, R. M.; Regnier, F. E. *Anal. Chem.* **1989**, *61*, 2059–2066.
(6) Herrera, S. *Nat. Biotechnol.* **2004**, *22*, 1343–1346.
(7) Chirino, A. J.; Mire-Sluis, A. *Nat. Biotechnol.* **2004**, *22*, 1383–1391.
(8) Schellekens, H.; Ryff, J. C. *Trends Pharmacol. Sci.* **2002**, *23*, 119–121.
(9) Pathange, L. P.; Bevan, D. R.; Larson, T. J.; Zhang, C. M. *Anal. Chem.* **2006**, *78*, 4443–4449.

10.1021/ac8019812 CCC: $40.75
Published on Web 12/10/2008

protein mutagenesis based on two parameters: (1) variation in relative surface area of a surface amino acid residue (variants with the same pI)[9] and (2) variation in the microenvironment of different surface accessible amino acid residues (change in net charge resulting in different pI).[10] The experimental results demonstrated a direct correlation between the subtle changes in protein structure (the two structural parameters) and its binding capability in IMAC.[9,10]

The overall goal of this article is to elucidate structure−function relationships (a protein's ability to interact with oppositely charged groups) using ion exchange chromatography (IEC) by correlating the same aforementioned two structural parameters and the resultant functional behavior (binding) in an IEC column. Understanding this correlation is important because the mechanism of protein binding in IEC is different from that in IMAC. The general trend in IEC is that the elution is dependent on the net charge of a protein (or its pI). However, it has been demonstrated that net charge is not the only criterion for protein elution in IEC, as indicated by the stoichiometric displacement model (SDM).[11,12] Though the whole protein surface modulates the protein binding process, the specific binding interactions are dominated by certain contact regions.[13] Thus, all charged amino acid residues do not contribute equally to the protein binding process in IEC.[11] The main goal of our experiments was to analyze subtle variation in protein charge including its position on the protein's surface by IEC. Such studies not only will reveal the ability of IEC in capturing subtle changes in protein structure but also may provide insight into the protein binding mechanism in IEC.

IEC is a widely used protein purification technique at both preparative and analytical scale. In fact, it is considered as the work horse of the biopharmaceutical industry.[14] In IEC, the partition of a solute onto a chromatographic column primarily depends on the physicochemical characteristics of both the solute as well as the stationary phase in the column. The main components that control the retention behavior of a protein are[15] (1) protein structure including charge density, (2) type of stationary phase, (3) pH and type of buffer, and (4) elution conditions (e.g., gradient or stepwise increase of salt concentration in the elution buffer). On the other hand, mechanism wise, a protein's retention behavior is determined by its thermodynamic adsorption/desorption equilibrium for a given stationary phase and mass transport within a column.[16,17] The thermodynamic equilibrium, directly correlated with the protein retention time, is primarily influenced by the protein structure, particularly its electrostatic state in IEC. Generally, thermodynamic equilibrium processes are characterized by change in the Gibbs free energy (ΔG), which is composed of two factors, (a) change in enthalpy (bond energy) (ΔH) and (b) change in entropy (water structure, $T\Delta S$). Therefore the presence of an attractive or a repulsive force (that contributes to ΔH) due to a given protein (net protein charge) in a given type of (anion or cation[18]) IEC column affects the ΔG (retention times). The term representing ΔS (desolvation) is generally not a dominating factor in IEC, unlike in the other chromatographic techniques such as reverse phase. Thus, when the structure of a protein is altered, then there should be an associated variation in its IEC behavior.

The other aspect that contributes to the binding mechanism is mass transport. Though this aspect does not directly contribute to the retention times, a brief insight would elucidate some subtle aspects that facilitate or control the precise thermodynamic equilibrium. The adsorption process involves three major steps, they are (1) mass transfer from mobile to stationary phase, (film resistance) controlled by convection (flow rate), (2) pore or surface diffusion controlled by concentration gradient of the protein, and (3) reversible binding (thermodynamic equilibrium) controlled by the nature of the protein structure.[17,19] Generally, the rate limiting step is believed to be the second step, which is of the order of 10^{-7} cm^2/s for proteins.[16,17]

According to Johnson et al.,[18] the diffusion of a protein in charged gels is dependent on the following equation:

$$D_e = \Phi D_b$$

where D_e is the effective diffusivity in stationary phase, Φ is the partition coefficient, and D_b is the diffusivity in the membrane gels. In charged gels, effective diffusivity is mainly due to the partitioning effect, which is primarily based on the electrostatic properties of both the stationary media and the protein molecule[17,20] for a particular mobile phase. Therefore, electrostatics determines Φ, which in turn controls the driving force (protein concentration gradient) for diffusion. It is important to recognize that electrostatics is dependent not only on the charges of both the protein and the stationary media but also on their spatial arrangements. The effects of these spatial arrangements are referred as steric affects. In some cases when proteins differ in their molecular weight, size, or shape, the steric effects play a significant role because they affect the physical accessibility between proteins and the stationary media. In other cases, such as this report, where a single protein and its variants are considered, steric effects might affect electrostatics but do not affect the physical accessibility factor because all the variants have almost identical molecular weight, size, and shape. Thus, in our experiments, since the stationary phase is kept the same and protein variants are employed, Φ is primarily dependent on electrostatics of the protein variants. Thus, whenever there is a subtle variation (in charge or in microenvironment) in the protein structure, there ought to be a discernible difference in the binding behavior of protein in an IEC column.

In this report, protein structure, that is, its electrostatic state and the microenvironment around a charged residue, is thoroughly investigated by generating a series of point mutations on a model protein, T4 lysozyme, via site-directed mutagenesis. We

(10) Pathange, L. P.; Bevan, D. R.; Zhang, C. M. *Anal. Chem.* **2008**, *80*, 1628–1640.
(11) Kopaciewicz, W.; Rounds, M. A.; Fausnaugh, J.; Regnier, F. E. *J. Chromatogr.* **1983**, *266*, 3–21.
(12) Zhang, C. M.; Glatz, C. E. *J. Chromatogr. A* **2005**, *1069*, 113–118.
(13) Xu, W. S.; Zhou, H.; Regnier, F. E. *Anal. Chem.* **2003**, *75*, 1931–1940.
(14) Liapis, A. I.; Grimes, B. A. *J. Sep. Sci.* **2005**, *28*, 1909–1926.
(15) Yao, Y.; Lenhoff, A. M. *Anal. Chem.* **2004**, *76*, 6743–6752.
(16) Wheelwright, S. M. *Protein Purification: Desigh and Scale up of downstream processing*; Wiley-Interscience: New York, 1991.
(17) Carta, G.; Ubiera, A. R.; Pabst, T. M. *Chem. Eng. Technol.* **2005**, *28*, 1252–1264.
(18) Johnson, E. M.; Berk, D. A.; Jain, R. K.; Deen, W. M. *Biophys. J.* **1995**, *68*, 1561–1568.
(19) Ladisch, M. R. *Bioseparation Engineering: Principles, Practice and Economics*; Wiley-Interscience: New York, 2001.
(20) Hirota, N.; Kumaki, Y.; Narita, T.; Gong, J. P.; Osada, Y. *J. Phys. Chem. B* **2000**, *104*, 9898–9903.

Table 1. Relative Surface Accessibility (rSA)[a] of Both Original and Replaced Amino Acid, Net Microenvironmental Charge around the Surface Accessible Amino Acid with 15 Å, Relative Hydrophilicity Values (rH_y)[b] and the Retention Times (RT, including for C-WT) of All Seven Variants

lysozyme variants	rSA (%) (original amino acid)	rSA (%) (histidine residue)	net micro-environmental charge	rH_y	rRT (%)
C-WT	N/A	N/A	N/A	N/A	0
K19H	62	50	+1	0.54	5.0
K43H	26	15	+2	0.27	11.0
R80H	62	74	+1	0.74	2.1
K83H	60	62	+5	0.62	9.8
K124H	48	47	+3	0.52	0.2
K135H	69	63	+5	0.64	0.7
K147H	49	47	+2	0.51	5.0

[a] The percentage of surface accessibility of the residue (side chain) in the folded protein divided by the SA of the residue (side chain) in the unfolded state, of lysozyme variants calculated by using a probe with a radius of 1.93 Å. [b] Microenvironments with rHy values (1) >0.25 are fairly hydrophilic, (2) < 0.25 are hydrophobic, and (3) < 0.0 are extremely hydrophobic.

attempt to establish a correlation between the protein structure and its respective retention time in an IEC column and use the binding abnormality to show that IEC may be used to detect subtle structural changes on proteins.

MATERIALS AND METHODS

In Silico Methods. The model protein chosen for our experiments is a cysteine-free T4 lysozyme (C-WT).[9] Two categories of mutants representing two structural parameters were generated by homology modeling using the SYBYL 7.1 software.[9] In category I, seven sites were chosen for site-directed mutagenesis based on their relative surface accessibility (rSA) (Table 1). All the seven original amino acid residues were replaced by histidine residue, which varies in the location on the protein surface.[9] The control variant for this category is C-WT. The microenvironment around a surface accessible amino acid residue was characterized by its net charge (Table 1), which was calculated by simple sum of charged residues within a 15 Å distance around the amino acid residue (Table 1). In category II, two point mutants with histidine at sites 135 and 80 (from category I) were chosen as two control variants pWT-135 and pWT-80, respectively. These control variants have similar histidine surface accessibility with high binding strength on IMAC-Cu^{2+} column.[9] Five sites were chosen on these two control variants, three in pWT-135 (subset-1) and two in pWt-80 (subset-2).[10] In total, 10 variants (for Category II) were generated, which varied in net charge and location (Table 2).

Relative Hydrophilicity (rH_y). The relative hydrophobicity/hydrophilicity value (rH_y), which characterizes the microstructure around all the individual residues or sites, was calculated using a modification of a published method.[21] The total hydrophilicity (tH_y) is first calculated by equation 1a, wherein sH_y is the sum of all the hydrophobic/hydrophilic constants of the moieties around these sites (4.25 Å) and is calculated manually from the X-ray crystal structure of cysteine free T4 lysozyme (1L63).[21] The values of rH_y representing the hydrophilicity nature of the microstructures here are obtained from the equation 1b.[21]

Table 2. Change in Net Charge and Relative Retention Times (rRT) for All Ten Variants Including the Two Control Variants, pWT-135 and pWT-80

protein variant	change in net protein charge	relative retention time (rRT) (%)
Control Variant (His-135)	0	0
D127S	+1	14.3
D127K	+2	35.1
S136D	−1	−21.9
S136K	+1	6.8
S136M	−0.15	5.5
R137S	−1	−19.2
R137D	−2	−28.5
Control Variant (His-80)	0	0
R76S	−1	−3.8
R76D	−2	−6.9
E108K	+2	27.3

Table 3. Relative Surface Accessibility[a] and Relative Hydrophilicity Values (rH_y)[b] of All the Five Sites Chosen for Single Point Mutations

sites	relative surface accessibility	relative hydrophilicity (rH_y)
Arg-76	0.60	0.84
Glu-108	0.42	0.55
Asp-127	1.00	1.02
Ser-136[c]	0.10	0.70
Arg-137	0.63	0.78

[a] The relative surface area of each amino acid side was directly obtained from the SYBYL program. [b] Microstructure with rHy values (1) >0.25 are fairly hydrophilic, (2) <0.25 are hydrophobic, and (3) <0.0 are extremely hydrophobic. [c] The aHy value of tyrosine was used to calculate the rHy of Ser-136.

$$tH_y = (1 - SA) \cdot sH_y + (SA) \cdot aH_y \quad (1a)$$

$$rH_y = \frac{tH_y}{aH_y} \quad (1b)$$

where, SA is surface accessibility, and aH_y is the absolute hydrophilicity in water obtained from a method by Mehler et al.[22] The rSA and rH_y values for categories I and II are reported in Table 1 and Table 3, respectively. Strongly hydrophilic microstructures are characterized by higher rH_y values.

Experimental Details. All variants were overexpressed in *Escherichia coli* and purified by a published method using cation exchange chromatography.[9,10] From earlier studies,[9,10] based on circular diochroism (CD) data and enzymatic assays, it was demonstrated that all the 17 lysozyme variants had intact three-dimensional (3D) structures. The nucleotide and amino acid sequences of all the variants were also verified.

Determination of rRT in IEC. A strong cation exchange (100 × 4.6 mm) HPLC column (Analytical Sales and Services Inc., Pompton Plains, NJ) having a polyamide coating with sulfonicpropyl functional groups, with particle size of 6 μm, was used to obtain the retention time of all variants under identical experimental

(21) Mehler, E. L.; Guarnieri, F. *Biophys. J.* **1999**, *77*, 3–22.

(22) Mehler, E. L. *J. Phys. Chem.* **1996**, *100*, 16006–16018.

conditions. The sample concentration used was 0.15 mg/mL. After the sample (45 μL) was loaded on to the column, the equilibrating buffer (20 mM sodium phosphate at pH 7.0) was run for four column volumes (CV). Then, a two step linear gradient was employed using an elution buffer (1 M NaCl, 20 mM sodium phosphate at pH 7.0). In the first step 10% of the elution buffer was reached in three CV. In the second step, wherein all 17 variants and the control variants (C-WT, pWt-135 and pWT-80) were eluted, 45% of the elution buffer was reached in 16 CV. The flow rate of the mobile phase was 1 mL/min. The retention time (RT) values (elution peak maxima), an average of two measurements, were determined at 280 nm. The rRT (relative retention time) was determined by the following equation:

$$\mathrm{rRT} = \frac{\mathrm{RT_C} - \mathrm{RT_V}}{\mathrm{RT_C}} \times 100 \qquad (2)$$

where RT_V stands for the retention time of variants, and RT_C stands for retention time of the control variants. For category I variants, RT_C is the retention time of C-WT, and for category II it is the retention time of pWT-135 for subset-1 and pWT-80 for subset-2.

RESULTS

Effect of Protein Structure on Its Behavior in IEC. The major components of the protein structure that affect its binding behavior in an IEC column are (1) overall net charge on the protein, (2) the type of individual surface amino acids, (3) the microenvironment, and (4) the microstructure around the surface amino acids (including its own rSA) that are involved in the binding process. In this report, all these four components were theoretically quantified by (1) change in net charge, (2) positively charged, negatively charged, or neutral amino acids, (3) net charge in the microenvironment, and (4) rH_y, respectively. Experimentally, all these four components were investigated by replacing selected original amino acid residues with different amino acid substitutions such as histidine, lysine, and serine. The category I variants varied specifically in factor 3 and 4, wherein they varied in their molecular microenvironments (same net charge or same pI). The category II variants varied in all four structural components. The effect of these structural components on the binding behavior of the protein was quantified in terms of relative retention time (rRT). The rRT values for categories I and II are reported in Table 1 and Table 3, respectively.

Category I. In cation exchange chromatography, the elution time of the proteins is directly proportional to the total net positive charge. All the seven variants eluted earlier than C-WT indicating that the lowered pI (or net positive charge), because of the removal of +1 charge amino acid residue (lysine or arginine) and replacement with +0.5 charge (histidine at pH 7), had an expected impact on the rRT values. However, although the apparent change of the net charge is the same for all variants, the impact was not uniform. The retention times for the variants ranged from 28.44 to 31.90 min, while the C-WT retention time was 31.97 min (Figure 1). The change in rRT values ranges from 0.2% to 11% from the C-WT rRT value (Table 1). This non-uniform retention displayed by category I variants can be attributed to the difference in the microstructure of the original amino acids because the replaced

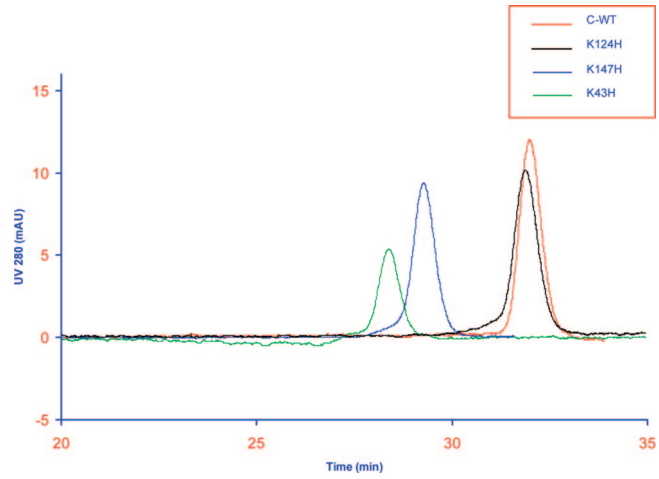

Figure 1. Chromatograms of category I variants, including the wild type lysozyme (C-WT) and three variants K12H, K147H, and K43H.

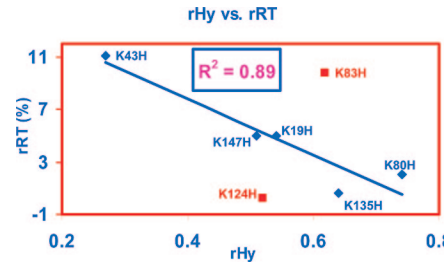

Figure 2. Correlation between the hydrophilic microenvironment (rH_y) and the retention time of all seven variants belonging to category-I.

amino acid is same in all the cases. Thus, rH_y and the net charge in the microenvironment are the two parameters that differ for all seven sites in the seven variants (Table 1). A correlation between the characterized microstructure (rH_y) and its effects on the protein behavior (rRT) is shown in Figure 2. A direct correlation (R^2 = 0.93) between the rH_y and the rRT was obtained when the two outlier variants K124H and K83H were excluded. These results illustrate that all these sites excluding the outliers seem to be a part of either a contact region[23] or an individual interacting site. There was no observable trend between the net charge in the microenvironment and the binding behavior for all seven variants (Table 1). Thus, despite having the same net charge or pI, the variants elute at different times during cation exchange chromatography according to the histidine's microstructure (hydrophilicity).

Excluded Variants K124H and K83H. These two variants deviate significantly from the above-mentioned correlation (Figure 2). For both proteins, the histidine microstructure did not seem to play an important role in their chromatographic behavior. This indicates that (1) these sites (124 and 83) are not situated in one of the contact regions or (2) there are some other factors that need to be considered such as the location of sites (steric effects) and its involvement in intramolecular interactions. In the case of K124H, the original lysine residue (at 124) in the C-WT variant was involved in a salt-bridge (Figure 3), and therefore its contribution to the binding may be negligible. In addition the orientation of Lys124 is projected into the protein interior. On the

(23) Regnier, F. E. *Science* **1987**, *238*, 319–323.

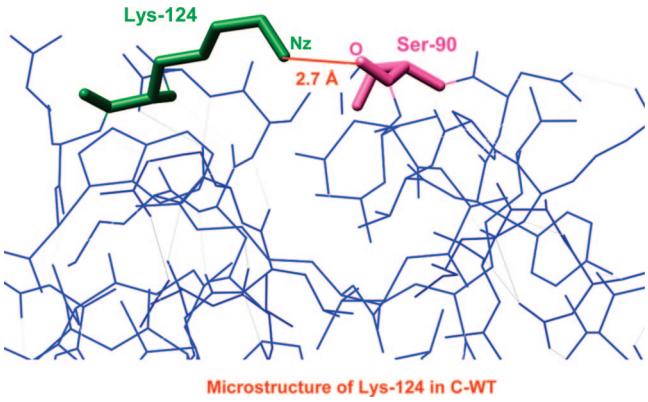

Figure 3. Microstructure of the variant C-WT, wherein the side chain of Lys-124 has involved in a salt bridge with the Ser-90 (oxygen) and is oriented inward into the protein. Thus, there is a high probability that this Site (Lys-124) is a non-interacting site in an IEC column. This figure was generated using UCSF Chimera.

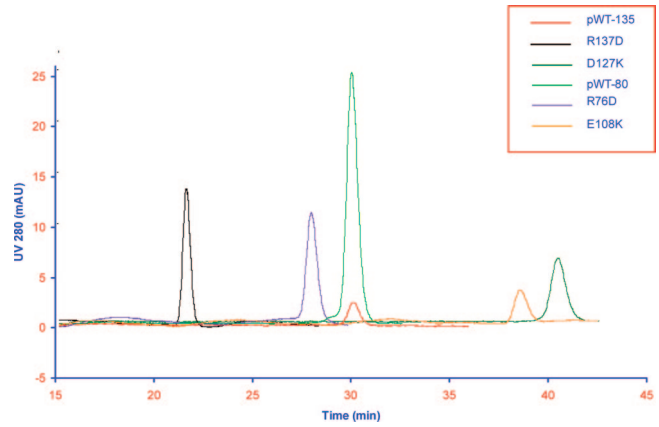

Figure 4. Chromatograms of category II variants including control variants pWT-135 and pWT-80 and four variants R137D, D127K, R76D, and E108K.

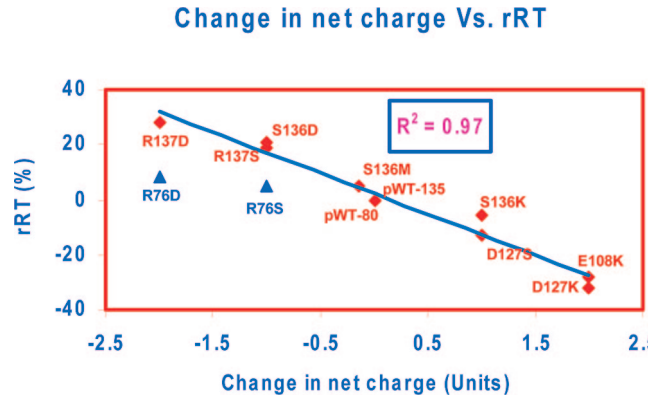

Figure 5. Correlation between the change in net charge and the relative retention time of variants in category-II, excluding variants R76D and R76S.

other hand, by assuming that the orientation of histidine in the K124H variant is similar to that of lysine in the C-WT variant, we can predict that the histidine's contribution to protein binding in K124H might also be lower than expected. Thus, the Lys124 and His124 sites are non-interacting and less-interacting sites, respectively.

In the case of K83H, the Lys83 site in C-WT might be an interacting site because of (1) its high rSA including its favorable orientation, to binding, pointing outward away from the protein's interior and (2) its non-involvement in any intramolecular interactions. However, its replacement amino acid, histidine, is involved in a hydrogen bond,[9] and therefore its contribution to the binding may be lower (less-interactive site) than that of other variants such as K135H or R80H. Thus, the presence of an intramolecular bond, in addition to rH_y, affects the binding behavior of the lysozyme variants in IEC.

Category II. As stated earlier, the elution of proteins in cation exchange chromatography is dependent on their pI. Proteins with higher pI elute later than proteins with lower pI. All variants eluted faithfully according to their pI (net charge) when compared to the pWT (Table 2). These results are similar to those reported in literature.[5] The retention times for the variants ranged from 21.56 to 40.75 min, while the pWT-135 retention time was 30.17 min (Figure 4). The change in rRT values ranges from −28.5% to 35.1% from the pWT (control variant) rRT value (Table 2). Assuming negligible microstructure (rH_y) effect, the variation in rRT values is approximately 15% per unit change in net charge, but from Table 2, it is evident that the effect of unit net charge is not uniform. When the two variants R76D and R76S were excluded, there is a strong correlation ($R^2 = 0.97$) between change in net charge and rRT values (Figure 5), and the strong correlation demonstrates the importance of overall net charge on the proteins. Though the variant S136K is included in this correlation, its rRT value is conspicuously very low when compared to another ± 1 change in the net charge variant (D127S or R137S). This noticeable difference can be attributed to the replaced lysine residue's low rH_y value (0.08), indicating a very hydrophobic microenvironment. The main reason for such a low rH_y value in case of lysine is due to (1) the low surface accessibility of its side chain,[10] which is pointed into the protein interior, and (2) its microstructure, which lowers its pK_a value; thus, its contribution to total net charge is minimal.

Excluded Variants R76S and R76D. The variants at site 76 (R76S and R76D) are similar to the variants at site 137 (R137S and R137D). These two sites can be characterized as having similar microstructural environment because of the following factors: (1) equivalent rSA values, (2) involvement of their arginine residue in a salt bridge with Asp72 (Arg137 is engaged in a salt bridge with Glu22),[10] (3) no significant difference between the microenvironmental net charge value of −1 for site 137 and +1 for site 76 (when compared with the category I values in Table 1), and (4) equivalent rH_y values (Table 2). As expected, because of the replacement of the positively charged arginine at these two sites by serine and aspartic acid, the rRT values are negative. But the rRT values of variants at site 76 are inordinately lower than the values at site 137 (Table 2). This variation can be explained by the formation of intramolecular bonds.

In case of variant R76D, Asp76 is located between His80 and Asp72. Because of the presence of Asp72 and its repulsive force, Asp76 likely forms a salt bridge with His80.[10] Consequently, the pK_a of aspartate is raised, and the total net negative charge on the protein is reduced. Thus, the individual interacting site 76 is now rendered as a less-interacting site. Similarly in case of R76S, as demonstrated by molecular modeling (Figure 6), there likely is a weak hydrogen bond between Ser76 and Asp72

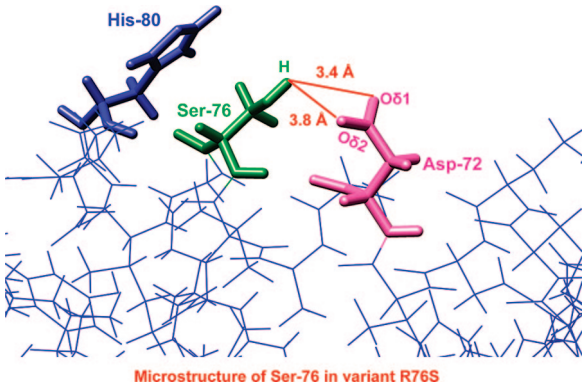

Figure 6. Microstructure of the variant R76S, wherein the side chain of Ser-76 is involved in a (weak) hydrogen bond with the side chain (Oxygen atoms) of Asp-72. This figure was generated using UCSF Chimera.

rendering Asp72 as a less interacting site. Therefore, because of the compromised net negative charge because of the presence of intramolecular interactions, the variants at site 76 eluted much later than the variants at 137.

DISCUSSION

Although it is conceivable that the whole protein surface plays a major role in the chromatographic behavior of a protein, Regnier et al. have demonstrated that there are certain regions or patches on the protein surface known as "contact regions" that are predominantly involved in the chromatographic binding process.[23,24] They have also indicated that the regions that are not the contact regions seem to contribute very little to the binding process. Yao et al.[15] were able to use a continuous electrostatic modeling program to elucidate the presence of contact regions and their possible effects on the protein behavior in IEC. By measuring the interacting free energy (IFE) values of a few regions, they were able to confirm that the front face of cytochrome *c*, which has abundant positive charges, was the contact region. They also demonstrated a direct correlation between the local surface potentials of a contact region and the retention times in IEC. Therefore it can be inferred that in IEC the presence of anisotropic charge on the protein surface results in contact regions that have high charge densities and thus can contribute to higher enthalpic energies. Therefore, any variation in the contact region, such as altering the charge density, affects its thermodynamic favorability in an IEC column.

Another factor, in addition to charge density, which can affect the contact region, is the shape of the protein molecule which induces steric effects during the binding process.[5,23] These effects can be explained by the heterogeneous physical accessibility, that is, the steric effects, of all sites on the protein surface to the chromatographic binding (media) sites. The exact nature of this steric effect is still not fully understood because of the poor knowledge of structure−function relationships in protein molecules. The steric effects are experienced because of the inherent heterogeneous nature of the protein molecules, which is also partially responsible to the anisotropic charge distributions on the protein surface. Thus, it is hard to differentiate the effects of charge density and steric contributions to the protein's chromatographic behavior.

Here, we have used the total charge within 15 Å of the surface amino acid as a measure to understand the effect of microenvironment and its location on the 3D structure of lysozyme. Our results indicate that the microenvironment effect on the binding process in cation exchange chromatography do not follow any particular trend (Table 1). A closer look at these contact regions, which generally consists of one or more charged amino acid residue, might help in elucidating the relationship between steric and charge effects. We have employed rH_y as a parameter to account for the microstructure around the to-be-studied amino acids. For cation exchange chromatography, the contact region mainly consists of charged residues (e.g., lysine and aspartate). These residues that take part in the binding process can be referred to as the interacting residues or sites. It is unclear what factors determine whether a site is interacting, less-interacting, or non-interacting. However, from our results we list three important parameters that might play a major role in determining whether a site is interacting or non-interacting. They are (1) its charge, (2) its side chain surface accessibility, mobility and orientation, and (3) its pK_a, which is influenced by (a) its microstructure (rH_y) and (b) its involvement in intramolecular interactions. Here we have demonstrated that all these three factors determine the binding process in cation exchange chromatography. From the results obtained from category II variants, we have demonstrated that there is a general trend that is illustrated by a direct correlation between net charge and the protein retention times (Figure 5). The surface, mobility, and orientation play an important role in determining the retention times of variant K124H. Meanwhile, the influence of microstructure and intermolecular bonds has been illustrated by variants S136K, K83H, R76S, and R76D.

On the other hand, all the above three parameters are dependent on the physical location of a particular site (steric effect) on the 3D structure of a protein. To understand the importance of the location, the IEC binding mechanism needs to be revisited. In IEC, which is a dynamic process because of the forced convective flow (mostly outside the pores), attaining a macroscopic equilibrium (in the whole column) is not possible.[25] However, as discussed in the introductory section, the electrostatic partitioning takes place in the diffusion controlled step inside the pores, and therefore at the molecular level there might be a possibility of a microscopic equilibrium.[25] Since diffusion is inherently a random process and because of the molecular internal energy, there is a strong possibility that all surface amino acids interact equally probably with the chromatographic media.[24] In addition, because of the heterogeneous stationary phase ligand distribution and an inherently heterogeneous protein (including anisotropic charge) structure, there is a random interaction between the stationary phase and the protein.

When the binding mechanism in cation exchange chromatography is considered in terms of individual interacting sites, their position on the 3D structure of the protein seem to play a minor role, and the major role is played by the charges on the protein

(24) Fausnaughpollitt, J.; Thevenon, G.; Janis, L.; Regnier, F. E. *J. Chromatogr.* **1988**, *443*, 221–228.

(25) Li, X. P.; Hupp, A. M.; McGuffin, V. L. The Thermodynamic and Kinetic Basis of Liquid Chromatography. In *Advances in Chromatography*; Grushka, E., Grinberg, N., Eds.; CRC Press: Boca Raton, FL, 2007; Vol. 45, pp 1−88.

surface and their (anisotropic) location on the protein molecule (electrostatics). Other factors, whose contribution is hard to characterize, also affect chromatographic behavior of a protein in IEC. Some of these factors are (1) hydration effects (entropy), (2) hydrophobic forces that are a part of the microenvironment around the surface amino acids, and (3) intermolecular forces such as van der Waals forces.[26] Another aspect that is not well understood is the change in protein configuration when it is bound to the stationary matrix.

CONCLUSIONS

Two categories of protein variants were generated using site-directed protein mutagenesis. Category I variants primarily varied only in their microstructure. The protein microstructure was theoretically characterized in terms of relative hydrophilicity. Category II variants primarily varied in their net charge. All 17 variants, which consist of the two categories, had different rRT values indicating that the underlying thermodynamic process also varies mainly because of variation in the protein structure.

For category I variants, there is a direct correlation between the protein microstructure and the rRT values in IEC. There is a direct correlation ($R^2 = 0.93$) between the microstructure of the charged group and the protein's retention time, except for two variants (K83H and K124H) whose histidine residue was involved in intramolecular bonds. These two excluded variants have a noninteracting (K124H) and a less-interacting site (K83H). Hence, the variation in rRT values can be attributed to the characterized protein microstructure.

For category II variants, there is a direct correlation ($R^2 = 0.97$) between the change in net charge (−2 to +2 units) and the retention times, except for two variants (R76D and R76S) whose retention times do not follow the correlation. However, there is no clear correlation between the difference in microstructure (rH_y) and the rRT values, because the major overriding or dominating parameter in this category is the change in the net charge. On the other hand, because of the compromised net negative charge, which in turn is due to the presence of intramolecular interactions, the variants at site 76 eluted much later than the other similar variants (e.g., site 137). Consequently, this site rendered as a less-interacting site. These excluded variants at site 76 provide insight into the relationship between protein structure and its ability to interact with charged ligands. Evidently, the presence of intramolecular bonds and the unfavorable orientation of the amino acid side chain, in addition to rH_y, affect the chromatographic binding behavior of the lysozyme variants. From an analytical point of view, IEC may be used as a technique to discover subtle structural variations (charge or microstructure) including intramolecular bonds in proteins by measuring variation in the rRT values.

In this paper we have demonstrated that when all the chromatographic factors are kept constant, the behavior of the protein variants in IEC can be predicted based on their net charge, charge density, and microstructure except for the variants with the specialized microstructures (e.g., presence of intramolecular bonds or steric hindrances). The correlation graphs including the outliers indicate that protein structure plays an important role in protein function. Our previous two papers[9,10] and the results of this paper indicate that it is possible to develop an analytical model or a matrix, which comprises orthogonal and complementary analytical methods, to characterize protein structures based on their protein function. Such an inverse characterization of protein structure based on their function will be a valuable guide for rational protein design or reverse protein engineering.

ACKNOWLEDGMENT

The work is partly funded by Grant J-706 from the Jefferess Memorial Trust.

Received for review September 18, 2008. Accepted November 11, 2008.

AC8019812

(26) Yao, Y.; Lenhoff, A. M. *Anal. Chem.* **2005**, *77*, 2157–2165.