Journal of Chromatography A, 1217 (2010) 191–198

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

# Investigation of protein binding affinity in multimodal chromatographic systems using a homologous protein library

Wai Keen Chung[a,c], Ying Hou[a,c], Melissa Holstein[a,c], Alexander Freed[a,c], George I. Makhatadze[b,c], Steven M. Cramer[a,c,*]

[a] Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[b] Department of Biology, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[c] Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, USA

ARTICLE INFO

*Article history:*
Available online 7 August 2009

*Keywords:*
Multimodal chromatography
Mixed mode chromatography
Homologous protein library
Structure–property relationships
Protein–ligand interactions
Coarse-grained simulations

ABSTRACT

A library of cold shock protein B mutant variants was employed to examine differences in protein binding behavior in ion exchange and multimodal chromatography. Single site mutations introduced at charged amino acids on the protein surface resulted in a homologous protein set with varying charge density and distribution. The retention times of the mutants varied significantly during linear gradient chromatography in both systems. The majority of the proteins were more strongly retained on the multimodal cation exchange resin as compared to the traditional cation exchanger. Further, the elution order of the mutants on the multimodal resin was different from that obtained with the ion exchanger. Quantitative structure–property relationship models generated using a support vector regression technique were shown to provide good predictions for the retention times of protein mutants on the multimodal resin. A coarse-grained ligand docking package was employed to examine the various interactions between the proteins and ligands in free solution. The multimodal ligand was shown to utilize multiple interaction types to achieve stronger retention on the protein surface. The use of this protein library in concert with the qualitative and quantitative analyses presented in this paper provides an improved understanding of protein behavior in multimodal chromatographic systems.



## 1. Introduction

The development of efficient bioseparation processes for the production of high-purity biopharmaceuticals is one of the most pressing challenges facing the pharmaceutical and biotechnology industries today. In addition, high-resolution separations for complex bioanalytical applications are becoming increasingly important. While it is generally accepted that non-specific interactions can often complicate single mode chromatographic separations (e.g. ion exchange, reversed phase), these additional interactions can also result in unexpected selectivities [1,2]. Recent advances in the design of multimodal chromatographic systems have produced new classes of chromatographic materials which can provide alternative and improved selectivities as compared to traditional single mode chromatographic materials [3–9]. Johansson et al. have developed a library of multimodal ligands that can be employed for the capture of charged proteins under conditions of high salts [4,5]. Liu and Pohl have developed a silica-based multimodal resin capable of weak anion exchange and reversed phase interactions for the simultaneous separation of acidic, basic and neutral pharmaceutical compounds [9]. Small ligand and pseudo-affinity chromatographic materials such as those used for hydrophobic charge induction chromatography have resulted in new classes of multimodal ligands which offer selectivities due more to multimodal interactions than to specific binding to certain classes of proteins. In addition, several libraries of multimodal ligands have been recently developed and employed on chromatographic resins for screening with biological mixtures. Applications of this technology range from preparative protein purifications [10,11] to front end separations for mass-spectrometry analysis [12,13]. While these new materials offer potential for bioseparations, there is a lack of fundamental understanding of the nature of binding of these ligands to protein surfaces.

The proteins used in the current study were cold shock protein B (CspB) variants. CspB is a small, monomeric globular protein that is found in bacteria. As it does not form multimers in solution, it reduces the complexity when attempting to study protein behavior in a range of systems. The small size of the protein (67 amino acid residues) also makes it easier to carry out simulations and modeling studies. A thermostable variant (CspB-TB) has also been developed

* Corresponding author at: Chemical Engineering, 3211 CBIS Building, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590, United States. Tel.: +1 518 276 6198; fax: +1 518 276 4030.
*E-mail address:* crames@rpi.edu (S.M. Cramer).

0021-9673/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2009.08.005

**Table 1**
Classes of CspB-TB residue substitutions; retention time of variants on SP Sepharose FF; amino acid sequence of CspB-TB templates.

| Native | | Positive → negative | | Positive → neutral | | Negative → neutral | | Negative → positive | |
|---|---|---|---|---|---|---|---|---|---|
| Variant | Retention time (mins) | Variant | Retention time (mins) | Variant | Retention time (mins) | Variant | Retention time (mins) | Variant | Retention time (mins) |
| Native | 11.10 | R3E[a] | 10.13 | R3Q[a] | 9.93 | D24N | 11.26 | D10K | 14.30 |
| | | K12E | 1.31 | K5Q | 9.28 | E43Q | 10.23 | E21K | 9.67 |
| | | K13E | 1.47 | K7Q | 9.40 | E48Q[a] | 10.71 | D24K | 10.36 |
| | | K39E | 8.26 | K12Q | 9.00 | E50Q[a] | 9.82 | E43K | 12.20 |
| | | K42E | 8.74 | K13Q | 8.22 | E53Q | 10.32 | E50K | 10.28 |
| | | K55E | 3.81 | K20Q[a] | 9.41 | | | E53K | 12.30 |
| | | | | K39Q | 9.33 | | | | |
| | | | | K42Q | 2.45 | | | | |
| | | | | K55Q | 8.78 | | | | |

Native CspB-TB sequence: MLRGKVKWFD-SKKGFGFITK-EGQDDVFVHF-SAIQEMGFKT-LKEGQAVEFE-IVEGKRGPQA-AHVKVVE. M36G Sequence: LLRGKVKWFD-SKKGFGFITK-EGQDDVFVHF-SAIQEGGFKT-LKEGQAVEFE-IVEGKRGPQA-AHVKVVE.

[a] Mutants created using the original CspB-TB as a template. All other mutants were created using M36G as the template.

by Gribenko and Makhatadze [14]. It was found to be structurally similar to the cold shock protein in B. subtilis (CspB-Bs). Modifications to the charged regions on the surface of CspB and CspB-TB result in a library of mutants that have similar structure but varying charge density and distribution.

In this study, this homologous library of CspB-TB mutants is employed to investigate the effects of charge modification on protein retention in multimodal and traditional cation exchange systems and a detailed comparison is carried out. Quantitative structure–property relationship (QSPR) models are generated using a support vector regression technique to enable prediction of the binding affinity trends and coarse-grained simulations are used to examine the synergistic interactions brought about by the different chemical species on the multimodal ligand. The use of this protein mutant library in concert with the qualitative and quantitative analyses presented in the paper provides an improved understanding of protein binding affinity in multimodal chromatography as well as an elucidation of possible binding regions on the protein surface.

## 2. Materials and methods

### 2.1. Materials

CspB-TB mutant variants were generated, overexpressed in *E. coli* and purified as described elsewhere [14]. Mutations were carried out by altering specific charged amino acids on the surface of the protein. Table 1 lists the mutations performed. As can be seen in the table, there were four subclasses of mutations carried out: positive to negative, positive to neutral, negative to positive and negative to neutral. As described elsewhere, this library was generated from two initial forms of Csp-B-TB, mutant M36G which has a leucine at position 1 and a glycine at position 36 and the original Csp-B-TB construct which had methionine residues at the N-terminus and position 36 (sequence information for both forms are shown at the bottom of Table 1). Mutants R3E, R3Q, K20Q, E48Q and E50Q were created using the original CspB-TB construct while mutants K12E, K13E, K39E, K42E, K55E, K5Q, K7Q, K12Q, K13Q, K39Q, K42Q, K55Q, D10K, E21K, D24K, E43K, E50K, E53K, E21Q, D24N, E43Q, and E53Q were constructed using mutant M36G. The difference between mutants expressed from the original CspB-TB construct and M36G is that all mutants created from the CspB-TB plasmid will have a methionine residue at position 1 and 36 while mutants created from the M36G plasmid will possess a glycine residue at these positions. A sulfopropyl Sepharose Fast Flow (SP Sepharose FF) column (1 ml) and a salt-tolerant multimodal cation exchanger (Capto MMC) column (1 ml) were used to study the retention behavior of the homologous mutant library on the two resin surfaces.

### 2.2. Equipment

Analytical linear gradient experiments were performed using a Waters HPLC system comprising of a 600 multisolvent delivery system, a 712 WISP autoinjector, and a 996 photodiode array detector controlled by a Millennium chromatography software manager.

### 2.3. Linear gradient experiments

The mutant variants were analyzed on a GE Healthcare Capto MMC column at pH 6.25. Linear gradient elution runs from 100% buffer A (20 mM sodium phosphate, pH 6.25) to 100% buffer B (20 mM sodium phosphate containing 1.5 M of sodium chloride, pH 6.25) in 60 column volumes at a flow rate of 1 ml/min were carried out to obtain retention time data on each mutant variant. Injection volumes ranging from 25 to 50 μl of sample were used depending on the initial concentration of each mutant variant. Before being injected onto the column, all protein samples were filtered using a 0.2 μm pore diameter syringe filter and diluted to a concentration of 0.5 mg/ml using buffer A. The column effluent was monitored at UV wavelength 280 nm and fluorescence (Ex: 300 nm, Em: 349 nm). The variants were also analyzed on the SP Sepharose column under the same experimental conditions, with the exception of buffer B (20 mM sodium phosphate with 0.5 M sodium chloride, pH 6.25).

### 2.4. Quantitative structure–property relationship (QSPR) modeling and SVM regression model

In order to construct a predictive QSPR model for the retention times of the homologous protein library, the protein crystal structure of each mutant was modified using the homology modeling program in the Molecular Operating Environment (MOE) software from CCG (Chemical Computing Group, Inc.) [15]. Ten intermediate homology structure models for each mutated protein were then built using the homology modeling program in MOE based on the corresponding FASTA sequences. These different intermediate homology models were the result of a permutational selection of different loop candidates and side chain rotamers. The intermediate homology model which had the highest score according to MOE's packing evaluation function was chosen as the homology model for further optimization. Finally, a sequence of energy minimizations was performed to obtain a final optimal homology model for each mutant in the library. Since these mutants are based on the stable small protein CspB, it is likely that these homology models resulted in accurate representation of the mutant 3-D structures. The detailed procedure for the homology modeling process has been described previously [16]. A variety of descriptors were calculated based on the three-dimensional protein structures, including traditional two-dimensional and three-dimensional descriptors

computed using the MOE software package [15]. In addition, pH-dependent electrostatic potential (EP) descriptors and molecular lipophilicity potential (MLP) descriptors were calculated as described elsewhere [17]. Invariant descriptors were then removed by using a partial least squares for PC (PLS-PC) software package [18]. Subsequently, a sparse linear $l_1$-norm support vector machine (SVM) regression algorithm [19] was applied in the feature selection of the hybrid set of molecular descriptors to identify a subset of descriptors relevant for each response. Finally, independent non-linear predictive retention SVM-QSPR models were generated using the descriptors from the feature selection step and were examined for quality and consistency (given by the cross-validated $r^2$ and RMSE for the training set) and predictive ability (for an external test set of molecules not used in the feature selection or generation of the model). In this work, the original training set of proteins is randomly subdivided into a validation set, with the remaining proteins used as a training subset. Models are then created for different sets of training proteins and are used to make predictions on the validation set of proteins left out of the training set. This procedure is repeated approximately 200 times and the cumulative results are used to construct 12 distinct but similar models. This is termed "bootstrapping" in quantitative structure–property relationship modeling literature and is known to provide better model generalization. A more detailed background on QSPR theory and regression procedures are described elsewhere [20–23].

### 2.5. Coarse-grained ligand–protein docking simulations in free solution

Coarse-grained docking simulations of the protein with representative ion exchange and multimodal chromatographic ligands in free solution were preformed using the Autodock package developed by Morris and co-workers [24,25]. To find a suitable binding conformation a genetic search algorithm is applied to reduce the free energy of a randomized ligand starting position. At each step of the search algorithm, an empirical function, based on the weighted summation of different energy functions, is used to predict the free energy of an adopted protein–ligand confirmation. The energy functions used include van der Waals, hydrogen bonding, electrostatics, desolvation, and torsional free energies. Each of these properties is given empirical weightings in the total energy calculation determined from a fitting of known ligand–protein interactions. While this and other docking applications are typically used for high energy binding at a known binding site, the versatility of the genetic search algorithm makes it ideal for use in blind-docking simulations where there may be several low energy binding sites at unknown locations on the protein's surface.

For the current study the protein was protonated according to the experimental pH and then energy minimized using AMBER 94 force field parameters [26]. The ligand was built as a truncated form of the resin ligand moiety and is shown in Fig. 1a and b for the MM and ion exchange ligand, respectively. Docking simulations were carried out with one protein and one ligand type at a time for both ion exchange and MM chromatography. Each docking simulation was performed with a large number of energy calculations (approximately 27,000 generations in the genetic algorithm) in order to assure sites of minimum energy. Further, each of these energy calculations were carried out multiple times at random ligand starting positions (100 for cation exchange and 200 for multimodal ligands). The results of these simulations were then analyzed for visualization, energetics, and cluster analysis using the AutodockTools package of the Python Molecular Viewer program [27,28].

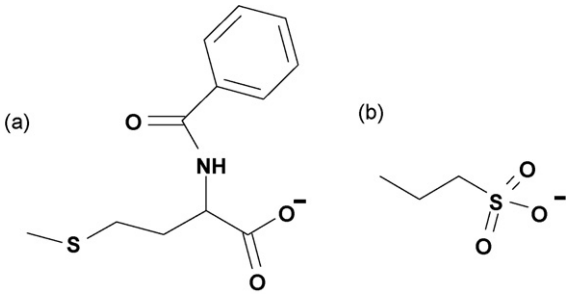


**Fig. 1.** Diagrams of (a) representative multimodal chromatographic ligand, N-Benzoyl-Methionine; (b) representative cation exchange chromatographic ligand, 1-propanesulfonic acid.

## 3. Results and discussion

### 3.1. Retention behavior of mutants under linear gradient conditions on SP Sepharose FF

A library of cold shock protein B mutant variants was employed to examine differences in protein binding behavior in ion exchange and multimodal chromatography. All mutants were created through single site mutations targeted at specific charged sites on the surface of the native protein as indicated in Table 1. In order to compare the results it is first important to examine the behavior in traditional cation exchange chromatography which was carried out at the same pH conditions (pH 6.25) as the multimodal experiments. Table 1 presents the retention times under linear gradient cation exchange conditions. In order to examine the elution order of the various mutants relative to the native protein, the data is also presented in Fig. 2.

Despite the use of a shallow salt gradient (60CV), many variants eluted with a similar retention to the native CspB at this pH. As expected, the reversal or neutralization of a positively charged residue in general resulted in weaker retention as compared to neutralization or reversal of a negatively charged residue. Reversal of the positively charged residues at positions 12, 13 and 55 caused more significant reductions in retention time as compared to other mutants in this subclass. This indicates that these three lysines may play an important role in determining protein retention on the cation exchanger and could possibly be part of a preferred binding region on the protein surface. At position 42, an unusual phenomenon was observed whereby neutralization of the positively charged residue at this position caused more significant reduction in retention time as compared to a reversal in the charge. This is in

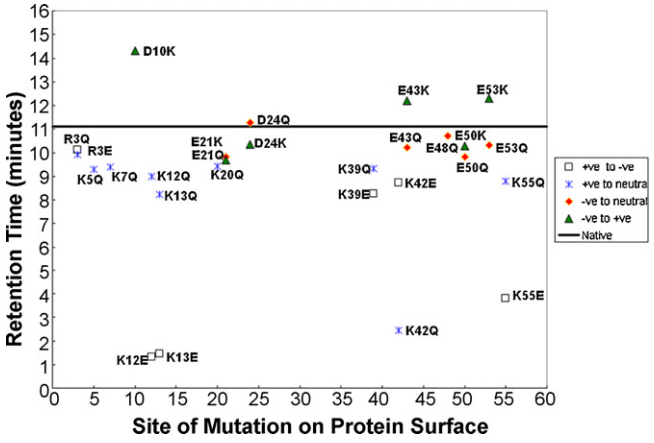


**Fig. 2.** Retention time profile of mutant variants on a SP Sepharose FF cation exchange column [experimental conditions: pH 6.25, 60CV linear salt gradient (Buffer B: Buffer A +0.5 M NaCl)].

**Table 2**
Classes of CspB-TB residue substitutions; retention time of variants on Capto MMC.

| Native | | Negative → positive | | Negative → positive | | Negative → positive | | Negative → positive | |
|---|---|---|---|---|---|---|---|---|---|
| Variant | Retention time (mins) | Variant | Retention time (mins) | Variant | Retention time (mins) | Variant | Retention time (mins) | Variant | Retention time (mins) |
| Native | 38.24 | R3E[a] | 1.33 | R3Q[a] | 1.58 | D24N | 43.83 | D10K | 33.30 |
| | | K12E | 1.5 | K5Q | 36.68 | E43Q | 48.94 | E21K | 42.19 |
| | | K13E | 19.96 | K7Q | 31.95 | E48Q[a] | 43.20 | D24K | 46.68 |
| | | K39E | 24.5 | K12Q | 24.72 | E50Q[a] | 42.1 | E43K | (Not eluted) |
| | | K42E | 28.48 | K13Q | 13.62 | E53Q | 45.25 | E50K | 49.08 |
| | | K55E | 19.82 | K20Q[a] | 36.7 | | | E53K | (Not eluted) |
| | | | | K39Q | 29.75 | | | | |
| | | | | K42Q | 1.622 | | | | |
| | | | | K55Q | 23.8 | | | | |

[a] Mutants created using the original CspB-TB as a template. All other mutants were created using M36G as the template.

contrast to previous results from our laboratory with these same mutants which were carried out at pH 5[16].

While most mutations where a negative charge was either neutralized or replaced with a positive charge had minimal effect on the elution time, the mutant D10K exhibited a strong increase in retention as compared to the wild type.

### 3.2. Retention behavior of mutants under linear gradient conditions on Capto MMC

In contrast to the weak retention of the protein library on the cation exchange column, the proteins exhibited strong retention on the Capto MMC column at the same pH conditions. In fact, the retention was so strong that a higher salt concentration (1.5 M NaCl) was required in the B buffer in order to elute a majority of the variants from the multimodal column (note: the same number of column volumes was employed in both gradients). Table 2 presents the retention times of the variants under the linear gradient Capto MMC conditions. In addition, the elution order of the CspB mutant library relative to the native protein is presented in Fig. 3.

As can be seen in Fig. 3, significant variation in the elution times was observed for the mutants on the Capto MMC surface. In general, variants with more positive surface charge were more strongly retained and the effects of charge reversal on protein retention were more pronounced than the effects of charge neutralization. However, there were subtle differences in the retention behavior that could be observed with the multimodal resin system. Although mutants where a positive charge was reversed exhibited weaker retention as compared to native CspB, the elution order of this subclass of mutants was significantly different from that observed on the cation exchange resin. For example, mutant R3E which was moderately retained on the ion exchange resin was found to elute in the flow through in the multimodal system. On the other hand, mutant K13E which eluted in the flow through on the ion exchanger was retained on the multimodal surface, although to a lesser extent than native CspB.

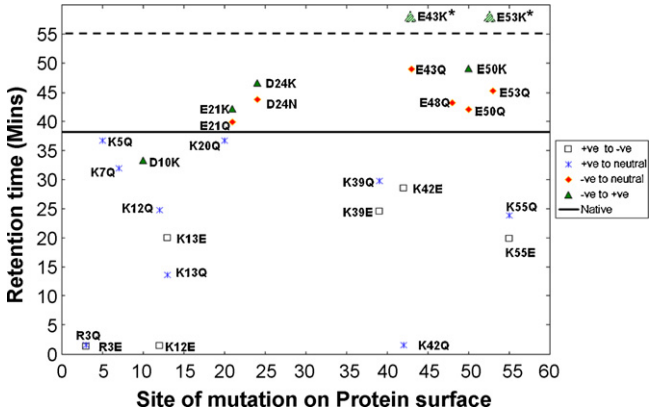


**Fig. 3.** Retention time profile of mutant variants on a multimodal weak cation exchange Capto MMC column [experimental conditions: pH 6.25, 60CV linear salt gradient (Buffer B: Buffer A +1.5 M NaCl)] {*Note*: the dotted line and the green dashed triangles are used to emphasize mutants E43K and E53K not eluting off the column under the experimental conditions} (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

While weaker retention was observed for the neutralization of positive charge as compared to reversal of charge at position 42 for both resin systems, the multimodal system also exhibited this behavior for mutations at position 13. In fact, K42Q had such a significant drop in retention on the MMC column that it eluted in the flow through.

In contrast to other variants where a negative charge was reversed, mutant D10K exhibited weaker retention on the Capto MMC surface as compared to the native protein. This is in stark contrast to the retention behavior observed for this variant on the cation exchanger. Further, while D10K was the most strongly bound variant on the cation exchanger, variants E43K and E53K exhibited the strongest retention on the MMC stationary phase. In fact, these variants did not elute off the MMC column under the given experimental conditions.

Clearly, these results with both the cation and the multimodal cation exchange resins indicate that both affinity and selectivity can be significantly different in these two systems. The fact that the CspB library was significantly more strongly bound in the MMC column indicates that electrostatics is likely not the sole mode of interaction between the protein surface and the MMC resin. While the effects of altering surface charge on protein retention in ion exchange can be understood by considering electrostatic potential maps [16], in multimodal systems these changes can also alter hydrophobic and hydrogen bonding contributions to binding from that region of the protein. In order to examine these multimodal effects in more detail, two theoretical approaches were employed, namely, QSPR and coarse-grained docking calculations.

### 3.3. Structure–property modeling of mutant library on Capto MMC

In order to enable the *a priori* prediction of protein retention times in the homologous protein library, a QSPR model was generated for two "unknown" mutants (E50K and K55E), which were not included in the model generation. In the training of the QSPR model, twenty-one of the twenty-seven proteins in the homologous library were used since six of the proteins were either eluted in the flow through or were retained too strongly on the column to be eluted during the gradient. Fig. 4a shows the results of the QSPR model. The open circles represent the training set molecules and the dark squares represent the predictions made using test set molecules. The error bars represent the standard deviation of all the bootstrapping values of the predictions. The cross-validated $r^2$

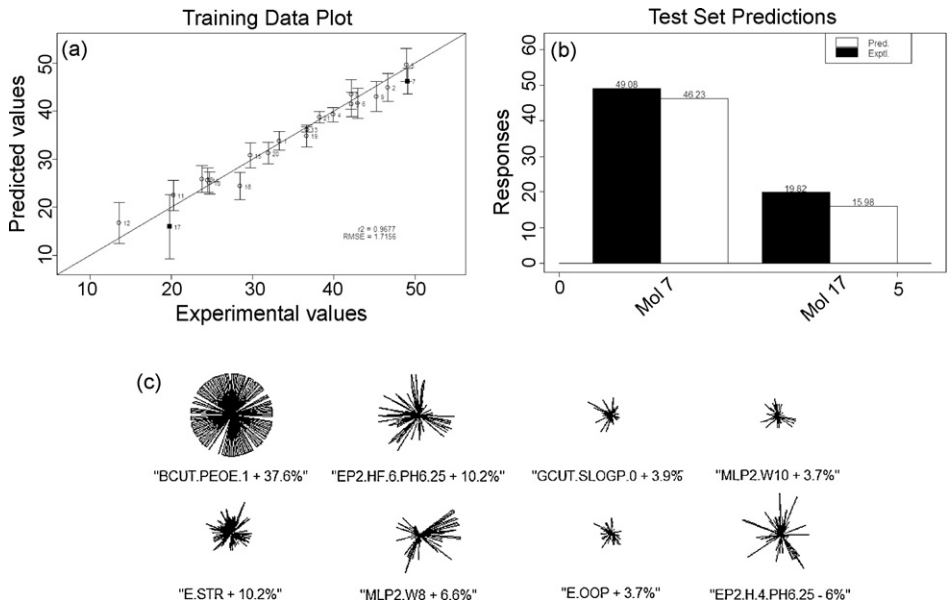


**Fig. 4.** (a) SVM QSPR model for prediction of the entire data set, where the open squares indicate the results for the training data set, cross-validated $r^2 = 0.9677$, the solid squares are the results for the test data. (b) Comparison of predicted and experimental retention times for the test set. (c) Star plots for the eight descriptors selected by the feature selection algorithm.

value was 0.9677, indicating good agreement between experimental retention times and the predicted results. In addition, the QSPR model also provided reasonably good predictions for test mutants E50K (molecule 7) and K55E (molecule 17) as can be seen from Fig. 4b.

### 3.4. Model interpretation

Understanding the underlying relationship between protein retention in multimodal chromatography and the physicochemical chemical properties of proteins is critical for the development of multimodal chromatographic processes. In addition to using QSPR models for the *a priori* prediction of protein chromatographic behavior, the descriptors selected in the model can also be employed to provide insight into factors that affect the retention time of proteins in the library. The descriptors resulting from the feature selection step were employed in a "Star plot" analysis to facilitate interpretation. Fig. 4c shows the star plots for the most significant descriptors with a threshold of >3% in the model. In this analysis, each star corresponds to a single selected molecular descriptor, where the line in each star represents one bootstrap and the radius of each line indicates the weight of that descriptor in the relevant sparse SVM model (a more detailed explanation of bootstrapping and star plots can be found in the references [20–23]. As seen in the figure, eight descriptors were obtained in the 3 general categories of charge/electrostatic potential, molecular lipophilicity/hydrophobicity and bond interactions. The definitions of those descriptors are listed in Table 3.

#### 3.4.1. Charge and electrostatic potential (EP) related descriptors

Descriptor "BCUT.PEOE.1" is an adjacency and distance matrix descriptor calculated by the value of PEOE partial charges from the MOE software. It is a charge related descriptor and exhibited the highest weight in the model (37.6%), which indicates the importance of the protein partial charge in binding in this multimodal chromatographic system.

The EP2.H and EP2.HF bin type descriptors refer to discrete ranges of EP values that are calculated on the protein Van der Waals surface at pH 6.25. Both types of EP descriptor values are sorted into 10 bins, and each bin contains the sum of the number of positions in a particular EP range, where bin 0 is the range of lowest values and bin 9 is the range of largest values in the protein library. As seen in Fig. 4c, the "EP2.HF.6.PH6.25" bin descriptor had a weight of 10.29%, which is a major positive contributor, whereas the "EP2.H.4.PH6.25" bin descriptor had a negative weight of −6%. This is due to the fact that bin 6 contains more positive EP values resulting in attraction in the cationic MM system and bin 4 contains more negative EP values resulting in repulsion (note: EP descriptors based on higher bins (>6) or lower bins (<4) represented relatively small surface areas on the proteins which is probably the reason that they were not selected in the QSRR model generation procedure). The relatively high weights of these EP descriptors suggest that the electrostatic interaction between the proteins and the MM resin play an important role in this MMC system.

#### 3.4.2. Molecular lipophilicity/hydrophobicity related descriptors

Since hydrophobic interactions can also play an important role in MMC chromatography, it is not surprising that descriptors from the molecular lipophilicity/hydrophobicity category were selected during the QSPR feature selection process. MLP2.W8 and MLP2.W10 are molecular lipophilicity potential descriptors which had weights of 6.6% and 3.7%, respectively in the QSPR model. In addition, GCUT.SLOGP.0 is a molecular hydrophobicity descriptor calculated from the MOE software using atomic contributions to log *P* (Wildman and Crippen SlogP method) which had a weight of 3.9%. Although these descriptors indicate that hydrophobic interactions played a role in protein retention in the MMC system, the effect was not as pronounced as electrostatic interactions.

#### 3.4.3. Bond interactions

"E.STR" represents the bond stretch potential energy and "E.OOP" represents the out-of-plane potential energy, with the weight of 10.2% and 3.7%, respectively. E.STR is reflective of the type of residue that is being mutated. For example, arginine and lysine have relatively higher values whereas asparagine and glutamine have lower values. E.OOP represents the conformational degree of freedom of the aromatic rings on the protein. Although no mutations were made that include the aromatic rings, other mutations can also effect the local protein conformation. Further, as will be shown below, aromatic rings on the protein surface play

**Table 3**
Definitions of selected descriptors.

| Categories | Descriptors | Definition of descriptors | Weight |
|---|---|---|---|
| Partial charge and electrical potential (EP) | BCUT.PEOE.l | PEOE partial charge descriptor (MOE) | +37.6% |
| | EP2.HF.6.PH6.25 | Sum of atoms with EP within bin 6 at pH 6.25 divided by the number of aid points on surface | +10.29% |
| | EP2.H.4.PH6.2? | Sum of atoms with EP within bin 4 at pH 6.25 | −6.0% |
| Molecular lipopliilicity/hydropliobicity | MLP2W8 | Molecular lipophilicity potential (MLP). Relates to lipophilicity hydrophobicity properties on a molecular surface | +6.6% |
| | GCUT.SLOGP.O | SlogP descriptor, log of octanol water partition coefficient (using the Wildman and Crippen SlogP method) (MOE) | +3.9% |
| | MLP2W10 | Molecular lipophilicity potential (MLP) descriptor [l] | +3.7% |
| Bond | E.STR | Bond stretch potential energy (MOE) | +10.2% |
| | E.OOP | Out-of-plane potential energy (MOE) | +3.7% |

an important role in high affinity interaction sites of the MM ligand.

### 3.5. Coarse-grained ligand docking analysis of CspB

Coarse-grained docking simulations were carried out as described in the theoretical section. These simulations produced a set of local minima binding confirmations between the ligands and protein which were then grouped into local binding sites, each of which had a range of possible configurations. The cluster analysis results can be seen in Fig. 5a–c. In Fig. 5a and b the colored points (blue = Capto MMC, yellow = SP) represent the center of mass of the ligand for the resulting configuration. Since the traditional cation exchange ligand has only one major interaction site it showed a much tighter distribution, only localizing at a few highly charged regions on the protein surface. In contrast, the multimodal ligand had a larger distribution around localized binding sites. This is probably due to the many favorable confirmations that can be obtained when there are multiple modes of interaction at play.

Comparing the local distribution of the binding sites to the experimental results shows a good correlation between the binding sites and residues shown to have large contributions to binding. For example in Fig. 5b the SP ligand localized near residues 5, 12, 13, 55, and 56, all of which showed a loss of retention in the cation exchanger upon neutralization or reversal. In Fig. 5a the MMC ligand is shown to adsorb favorably near residues 3, 12, 13, 39, 42, 55, and 56; residues that showed significant loss of retention in the MMC system. By studying the local distribution of binding sites

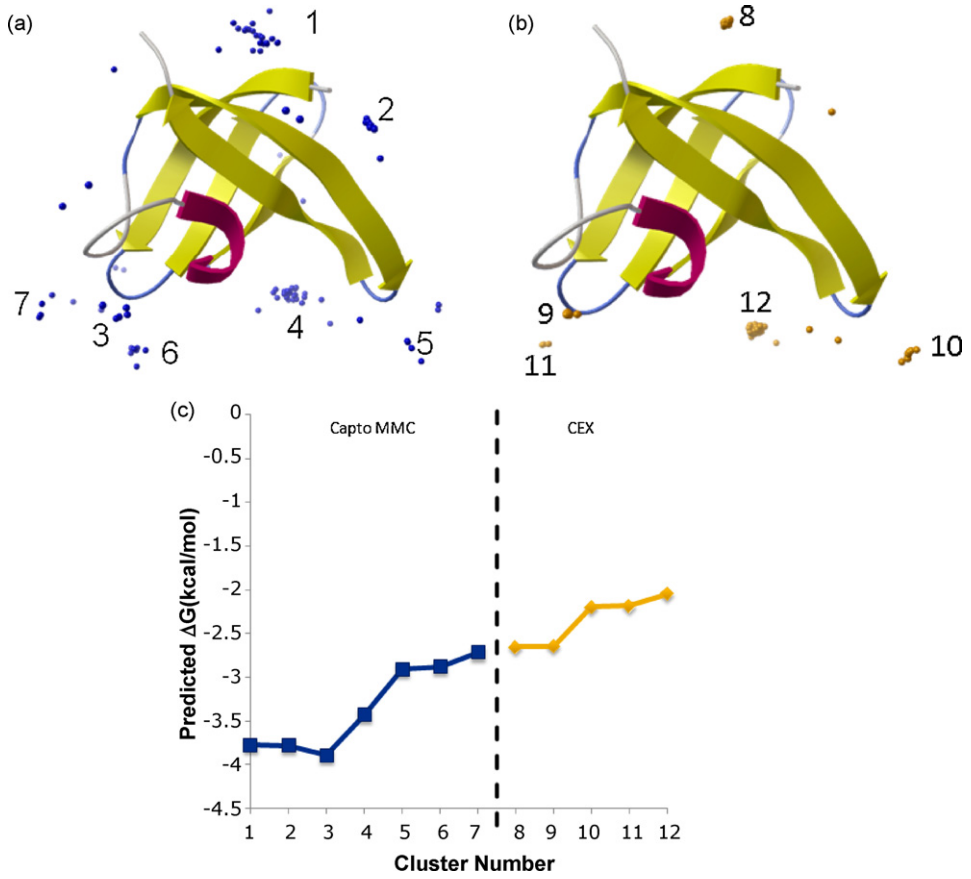


**Fig. 5.** Autodock cluster analysis for representative (a) Capto MMC Ligands (blue spheres). (b) CEX Ligands (yellow spheres) binding to the surface of Native CspB. Spheres represent the center of mass of the ligand for the resulting configurations. (c) Energy values for the associated clusters (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

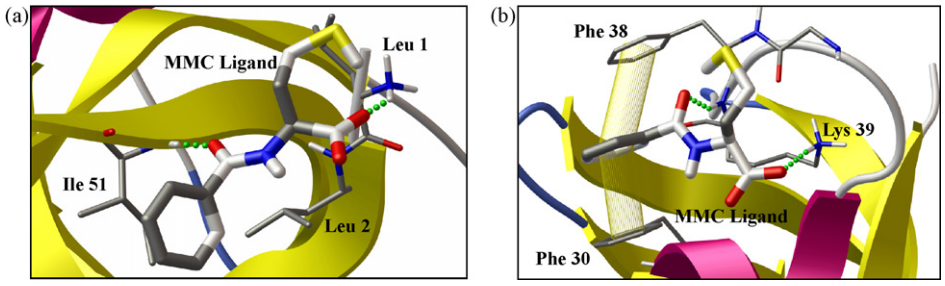


**Fig. 6.** Lowest energy MMC ligand adsorption confirmation at (a) Binding Site 2. The ligand has a charge reinforced H-bond with the N-terminus, an H-bond with the backbone of Ile 51, and is hydrophobically associated with the sidechains of Leu 2 and Ile 51. (b) Binding Site 3. The ligand has a charge reinforced H-bond with the sidechain of Lys 39, an H-bond with the backbone of Lys 39, and forms pi–pi stacking between Phe 36 and 38. Hydrogen bonds are highlighted in green dots, pi–pi stacking effects are shown in wire cylinders (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

it can also be seen why mutations on certain residues resulted in significant changes in retention while others did not. For example, residue 3 showed a dramatic change in retention on the MMC resin upon neutralization, while the neighboring residues 5 and 7 did not. Upon examination of the local binding distribution around residues 3, 5, and 7 the ligand clearly favored interactions with residue 3 over the other two. As a result, the loss of the residues 5 and 7 does not affect the binding of the protein while the removal of residue 3 would cause an important binding site to disappear.

It is also important to look at how the energetics of interactions plays a role in the binding of the ligands to the protein. Fig. 5c shows the average binding free energy for the different clusters on the protein's surface. As can be seen in the figure, the multimodal ligand showed stronger binding relative to the traditional cation-exchange ligand, which agrees with the stronger retention of the protein observed on the multimodal resin. This is probably due to the combined energies from multiple interaction types. The traditional ion-exchanger showed the expected interaction energy resulting from a single charge reinforced hydrogen bond (approximately 2 kcal/mol). The multimodal ligand displayed a wider range of interaction potentials indicate of multiple possible modes of interaction. While the higher energy site clusters (1–4) showed the effects of additional hydrophobic interactions and hydrogen bonding sites, the lower energy site clusters (5–7) were very similar to the traditional ion exchanger since they were mostly charger–charge interactions with minor additional van der Waals contacts.

A better understanding of the way multiple interactions can play a role in the increased binding energies can be seen by examining the ligand confirmations within a binding site on the protein surface. Fig. 6a and b shows representative views of the ligand in the high energy binding sites 2 and 3 defined in Fig. 5a (note: in this figure only the portion of the protein surface that directly interacts with the ligand is shown). The ligand in Fig. 6a is electrostatically interacting with the N-terminus, has an additional hydrogen bond with the backbone of isoleucine 51, and is hydrophobically associated with leucine 2 and isoleucine 51. Fig. 6b shows the ligand in its next lowest energy binding site. Here the ligand has a charge–charge bond with lysine 39, additional hydrogen bond with the backbone of residue 39, and is able to form a pi–pi stacking between phenylalanines 36 and 38. By forming strong interactions with several residues, the ligand is able to attach with more than double the strength of a traditional cation exchange ligand. Further, the addition of interaction types beyond simple electrostatics alters how the protein-resin system reacts to the addition of salts or other mobile phase modifiers. The importance of electrostatics and hydrophobic interactions shown in these ligand docking calculations was also indicated in the molecular descriptors selected in the QSPR analysis. However, it should be noted that these coarse-grained docking simulations were performed using representative chromatographic ligands in free solution. Ligand immobilization onto the resin surface will likely pose steric limitations and these will be studied in the future using a variety of experimental (EPR and NMR) and simulation (molecular dynamics) techniques.

## 4. Conclusions

A homologous library of mutant variants was used to examine the effects of charge modification on protein retention on a multimodal cation exchange surface. The protein library was observed to bind more strongly on the multimodal cation exchange resin as compared to the traditional cation exchanger due to the synergistic effects arising from the multiple modes of interaction between the ligand and the protein surface. Significant differences in elution order of the variants were also seen on the two chromatographic surfaces indicating that the proteins may have different preferred binding regions on the two resin surfaces. On the cation exchanger, charge modification altered protein retention to different extents, depending on the location and microenvironment of the mutated amino acid. However, on the multimodal resin, the effects of charge modifications were more complex due to the various protein–ligand interactions that can take place within a single binding region on the protein surface. QSPR was shown to be able to provide good predictions of protein retention time in the MM system and molecular descriptors selected during the model generation indicated that electrostatic and hydrophobic interactions were the primary modes of interaction. The use of Autodock provided insight into various protein–ligand interactions that can occur within a single binding site resulting in stronger binding energetics for a multimodal ligand as compared to the cation exchange ligand. In addition, the nature of the multimodal ligand enables it to interact with a larger proportion of the protein surface unlike the cation exchange ligand which is restricted to concentrated charged regions of the protein surface. Although ligand–protein docking programs like Autodock can provide insight into multimodal ligand–protein interactions, future work in our lab on using NMR with labeled proteins and full MD simulations will provide further insight in the nature of selectivity in MM chromatographic systems. Further, future work will employ a variety of experimental (EPR and NMR) and simulation (molecular dynamics) techniques to study protein interactions in solid phase MM systems.

## Acknowledgements

This work was supported by National Science Foundation grants CBET 0418413 and MCB 0416746. The authors would like to acknowledge Prof. Curtis Breneman from Rensselaer for his advice and the use of the MOE program software.

## References


[1] W.R. Melander, Z. El Rassi, C. Horvath, J. Chromatogr. 469 (1989) 3.
[2] W.S. Hancock, J.T. Sparrow, J. Chromatogr. 206 (1981) 71.
[3] L.W. Mclaughlin, Chem. Rev. 89 (1989) 309.
[4] B.L. Johansson, M. Belew, S. Eriksson, G. Glad, O. Lind, J.L. Maloisel, N. Norrman, J. Chromatogr. A 1016 (2003) 35.
[5] B.L. Johansson, M. Belew, S. Eriksson, G. Glad, O. Lind, J.L. Maloisel, N. Norrman, J. Chromatogr. A 1016 (2003) 21.
[6] D. Gao, D.Q. Lin, S.J. Yao, Chem. Eng. Sci. 61 (2006) 7260.
[7] S.C. Burton, N.W. Haggarty, D.R.K. Harding, Biotech. Bioeng. 56 (1997) 45.
[8] S.C. Burton, D.R.K. Harding, J. Chromatogr. A 775 (1997) 39.
[9] X.D. Liu, C. Pohl, Lc Gc Europe (2007) 33.
[10] X.D. Liu, C. Pohl, Am. Lab. 39 (2007) 22.
[11] D. Gao, S.J. Yao, D.Q. Lin, J. App. Polym. Sci. 107 (2008) 674.
[12] W. Bicker, M. Lammerhofer, W. Lindner, Anal. Bioanal. Chem. 390 (2008) 263.
[13] P.J. Boersema, N. Divecha, A.J.R. Heck, S. Mohammed, J. Proteome Res. 6 (2007) 937.
[14] A.V. Gribenko, G.I. Makhatadze, J. Mol. Biol. 366 (2007) 842.
[15] MOE, Chemical Computing Group, Montreal, Canada, 2006.
[16] W.K. Chung, Y. Hou, A. Freed, M. Holstein, G.I. Makhatadze, S.M. Cramer, Biotech. Bioeng. 102 (2009) 869.
[17] T. Yang, M.C. Sundling, A.S. Freed, C.M. Breneman, S.M. Cramer, Anal. Chem. 79 (2007) 8927.
[18] C. Heckler, PLS-PC: PLS Toolkit; V4.4, Eastman Kodak, 2000 (PLS).
[19] M. Song, C.M. Breneman, J. Bi, N. Sukumar, K.P. Bennett, S. Cramer, N. Tugcu, J. Chem. Inform. Comput. Sci. 42 (2002) 1347.
[20] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1998.
[21] A. Ladiwala, F. Xia, Q.O. Luo, C.M. Breneman, S.M. Cramer, Biotech. Bioeng. 93 (2006) 836.
[22] K. Rege, A. Ladiwala, N. Tugcu, C.M. Breneman, S.M. Cramer, J. Chromatogr. A 1033 (2004) 19.
[23] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, Langmuir 19 (2003) 8443.
[24] G.M. Morris, D.S. Goodsell, W.E. Hart, R.K. Belew, A.J. Olson, Comp. Chem. 19 (1998) 1639.
[25] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, J. Comp. Chem. 28 (2007) 1145.
[26] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, J. Am. Chem. Soc. 117 (1995) 5179.
[27] M.F. Sanner, J. Mol. Graph. Model 17 (1999) 57.
[28] M.F. Sanner, A.J. Olson, J.C. Spehner, Biopolymers 38 (1996) 305.