Defining the property space for chromatographic ligands from a homologous series of mixed-mode ligands

James A. Woo^a, Hong Chen^b, Mark A. Snyder^b, Yiming Chai^a, Russell G. Frost^b and Steven M. Cramer^{a†}

^aDepartment of Chemical and Biological Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute

^bLife Science Group, Bio-Rad Laboratories

^cProcess Chromatography Division, Bio-Rad Laboratories

† corresponding author:

Steven M. Cramer

E-mail: crames@rpi.edu

Phone: (518) 276-6198

Fax: (518) 276-4030

Abstract

A homologous ligand library based on the commercially-available Nuvia cPrime ligand was generated to systematically explore various features of a multimodal cation-exchange ligand and to identify structural variants that had significantly altered chromatographic selectivity. Substitution of the polar amide bond with more hydrophobic chemistries was found to enhance retention while remaining hydrophobically-selective for aromatic residues. In contrast, increasing the solvent exposure of the aromatic ring was observed to strengthen the ligand affinity for both types of hydrophobic residues. An optimal linker length between the charged and hydrophobic moieties was also observed to enhance retention, balancing the steric accessibility of the hydrophobic moiety with its ability to interact independently of the charged group. The weak pKa of the carboxylate charge group was found to have a notable impact on protein retention on Nuvia cPrime at lower pH, increasing hydrophobic interactions with the protein. Substituting the charged group with a sulfonic acid allowed this strong MM ligand to retain its electrostatic-dominant character in this lower pH range. pH gradient experiments were also carried out to further elucidate this pH dependent behavior. A single QSAR model was generated using this accumulated experimental data to predict protein retention across a range of multimodal and ion exchange systems. This model could correctly predict the retention of

proteins on resins that were not included in the original model and could prove quite powerful as an *in silico* approach towards designing more effective and differentiated multimodal ligands.

Keywords: Multimodal chromatography; Hydrophobic interaction; pH gradients; Protein surface properties; Quantitative Structure-Activity Relationship

1. Introduction

Multimodal chromatographic systems have developed in a variety of forms including mixed-mode, hydrophobic charge induction, mixed ligands and mixed bed chromatographic systems, with many permutations of ligand structures within each category [1-6]. The modes of interaction in these systems are typically either a combination of electrostatic and hydrophobic interactions or a mixture of positive and negative charges which can present unique advantages in selectivity over traditional single mode chromatographic separations [2-4,7]. Mixed-mode chromatography and hydrophobic charge induction chromatography are the predominant methods utilized in preparative scale multimodal separations, largely due to their superior resolution of impurities or the ability to capture proteins directly from high ionic strength cell culture fluid [8-13]. In these forms of multimodal chromatography, the orthogonal modes of interaction are combined into a single molecular entity, improving the homogenous distribution of both interaction moieties across the surface of the chromatographic support.

There is a growing set of publications in the literature that investigate the chemical and structural diversity of multimodal ligands and have begun to identify structural characteristics that lead to significant functional diversity [14]. In the work of Johansson et al., a comprehensive set of mixed-mode and mixed-ligand media was synthesized to create cation-exchange and anion-exchange libraries and the results indicated that ligands containing aromatic moieties demonstrated increased salt-tolerant adsorption as compared to ligands with aliphatic chain groups [4,15,16]. Mountford et al. created a series of heterocyclic aromatic rings systems with a variety of substituents and geometric arrangements and observed that the more polar ligands tended to be more selective when capturing a target antibody and resolving it from cell culture fluid contaminants [17]. Molecular dynamics simulations with MEP HyperCel, a pyridine-based ligand, showed that this ring forms both hydrophobic and hydrogen bonding interactions that help it form a tight interaction with the target hydrophobic pocket on the Fc-region of an

antibody [18]. This ligand also incorporated a thioether group and was developed as part of a class of thiophilic ligands that utilize hydrophobic π -donor/acceptor interactions to form strong interactions with aromatic groups and were observed to specifically adsorb immunoglobulins from a background of host cell impurities [1,19-21]. In the first paper in this series [22], it was observed that spatial organization of hydrophobic and charged moieties on two multimodal cation-exchange ligands (Capto MMC and Nuvia cPrime) proved to have a substantial effect on the retention behavior of certain proteins with clusters of surface-exposed aliphatic residues while having similar affinities to charged and aromatic moieties.

However, many more variables in multimodal ligand design have yet to be characterized, three of which are addressed in the current study. These variables include the role of geometric constraints (the distance between two functional groups and the relative steric accessibility of these functional groups), the effect of charge density and ligand pKa, and the presence of a polar substituent near the hydrophobic moiety. In the current work, these variables are characterized using a homologous series of nine prototype ligands that are based on a commercial multimodal resin template (Nuvia cPrime) so that alternate sources of variation (base matrix chemistry, immobilization chemistry and ligand density) are greatly reduced and any differences can be associated with changes in the chemical and structural properties of these ligands. In addition, these ligands are screened across a diverse set of protein chemistries and structures which can then be used to identify class-specific differences in protein adsorption that are related to a particular change in ligand chemistry. Finally, a single QSAR model is generated using this accumulated experimental data to predict protein retention across a range of multimodal and ion exchange systems.

2. Materials and Methods

2.1 Materials

Glacial acetic acid and guanidine hydrochloride were purchased from Thermo Fisher Scientific (Pittsburgh, PA). Sodium chloride, sodium acetate, sodium phosphate monobasic, sodium phosphate dibasic, sodium hydroxide, hydrochloric acid, L-arginine HCl, urea, ovalbumin (chicken egg white albumin), α -lactalbumin (bovine), albumin (bovine, human),

conalbumin (chicken egg white), β -lactoglobulin A (bovine milk), β -lactoglobulin B (bovine milk), trypsin (bovine and porcine), α -chymotrypsin (bovine pancreas), α -chymotrypsinogen A (bovine pancreas), ribonuclease A (bovine pancreas), ribonuclease B (bovine pancreas), cytochrome C (horse heart), aprotinin (bovine lung), lysozyme (chicken egg white), papain (papaya latex), and avidin (egg white) were purchased from Sigma-Aldrich (St. Louis, MO). Recombinant human ubiquitin was purchased from Boston Biochem, Inc. (Cambridge, MA). Capto MMC, CM Sepharose Fast Flow and SP Sepharose Fast Flow chromatography media were purchased from GE Healthcare (Uppsala, Sweden). MX-Trp-650M chromatographic media was a gift from Tosoh Biosciences LLC (King of Prussia, PA). Nuvia cPrime and the various prototype chromatography media were provided by our collaborator, Bio-Rad Laboratories (Hercules, CA). 5mm x 50mm glass columns and adapters were purchased from Pharmacia Biotech (Uppsala, Sweden).

2.2 Column Packing Procedure

Chromatographic resin was first equilibrated in deionized water and then was resuspended in a 50% (v/v) slurry in deionized water. 2.2 mL of slurry was poured into a 5 mm (ID) x 50 mm column and flow-packed in deionized water at 0.5 mL/min for 30 minutes. The flow adapter was adjusted onto the surface of the resin bed and flow was adjusted to 1mL/min and packed for another 30 minutes. The adapter was again adjusted onto the bed surface at the final bed volume of ~1 mL.

2.3 Resin Titration Experiments

Chromatographic resin was first equilibrated in deionized water and then rinsed with an equal volume of 0.1 M HCl. The resin was then resuspended in an equal volume of 0.1 M HCl and equilibrated for 2 hours with mild agitation to maintain the suspension of resin particles. Afterwards, the solution was allowed to settle and the supernatant was removed. An equal volume of 0.1 M HCl was added to the settled resin and the solution was resuspended. This solution was then titrated with 0.1 M NaOH. The solution was thoroughly mixed after each addition of base and the solution pH was recorded after a delay of 5 minutes.

2.4 Protein Library Screening Experiments

Linear gradient experiments were performed on an Äkta Explorer 100 (Amersham Biosciences, Uppsala, Sweden). Running buffers for all experiments were prepared from a 25x concentrate (500 mM acetate, pH 5 or 500 mM phosphate, pH 6) and diluted to the desired concentration without further pH adjustment. Buffers containing co-solutes (urea, guanidine-HCl and L-arginine HCl) were pH adjusted as needed using 2 M NaOH or 2 M HCl stock solutions.

1mL columns were equilibrated at 1 column volume (CV)/min with 5 CV of 1% Buffer B (Buffer A + 1.5 M NaCl) in Buffer A prior to the start of each experiment. Proteins were dissolved in the equilibration buffer (1% Buffer B) to 3 mg/mL and pipetted into 96-well UV transparent well plates. 50 µL of protein solution was aspirated by the A-905 autosampler (Amersham Biosciences, Uppsala, Sweden) and injected into the column. A linear salt gradient from 1 to 100% Buffer B was generated over 45 CV and held at 100% B for 8 mL (to account for the dead volume of the chromatography system). The column was then re-equilibrated with 7 CV prior to the next injection. Absorbance at the column effluent was measured at 280 nm and 215 nm using a 10 mm UV flow cell. Retention times were determined by calculating the center-of-mass for each peak. The conductivity in mS/cm was determined for that retention time and the conductivity was used to determine the elution salt concentration value.

2.5 pH Gradient Experiments

Linear gradient experiments were performed on an Äkta Explorer 100 (Amersham Biosciences, Uppsala, Sweden). Running buffers for all experiments were prepared from a 20x concentrate (400 mM each of citrate, phosphate, tris base and glycine, titrated to either pH 4.0 or pH 11.0) and diluted to the desired concentration without further pH adjustment.

1 mL columns were equilibrated at 1 column volume (CV)/min with 5 CV of Buffer A (pH 4.0 buffer) prior to the start of each experiment. Proteins were dissolved in the equilibration buffer to 3 mg/mL and deposited into 96-well UV transparent well plates. 50 µL of protein solution was aspirated by the A-905 autosampler (Amersham Biosciences, Uppsala, Sweden) and injected into the column. A linear pH gradient from 0 to 100% Buffer B (pH 11.0 buffer)

was generated over 45 CV and held at 100% B for 8 mL (to account for the dead volume of the chromatography system). The column was then re-equilibrated with 7 CV of Buffer A prior to the next injection. Absorbance at the column effluent was measured at 280 nm and 215 nm using a 10 mm UV flow cell. Retention times were determined by calculating the center-of-mass for each peak. The conductivity was adjusted in these experiments by adding equal amounts of NaCl (50, 100, 150, 250, 500, 1000, 1500 mM) to both Buffer A and Buffer B in order to generate a range of ionic strengths. For each experiment, both the pH and conductivity were recorded at the maximum of the protein elution peak.

2.6 Preparation of Protein 3D structures

All protein structures were obtained from the RCSB Protein Data Bank; the corresponding PDB codes can be found in Table 1. Water molecules and co-solutes were removed from the structure file and homology modeling (Molecular Operating Environment (MOE), Montreal, Québec, Canada) was performed to replace any segments of the polypeptide missing from the structural file. Structures were protonated at pH 7 using the Protonate3D function in MOE and subjected to three rounds of tethered energy minimization using the Amber99 forcefield.

2.7 Calculation of Residue Cluster Descriptors

Using the prepared protein structure file, the solvent-accessible surface area (ASA) of each atom was calculated using MOE. The surface area of all side chain atoms corresponding to a particular residue were summed together and the % exposure of this residue was calculated as the ratio of ASA_{residue} to ASA_X for a Gly-X-Gly tripeptide. pKa values for all titratable groups were calculated using PROPKA 2.0 [23]. The location of each residue in the protein structure was recorded as the center-of-mass for the residue's side chain. Residues with a % exposure < 0.15 were considered to be buried and excluded from the descriptor calculations. Uncharged residues, which were defined as basic residues where the $pK_a < pH$ and acidic residues where the $pK_a > pH$ were also excluded.

Residue clusters were calculated using the same methodology employed by Hou et al. [7]. After compiling a list of selected residues of a particular property (e.g. charged acidic

residues or exposed aliphatic residues), distances were computed between each pair of residues and any pairs falling within the 2-10 Å range were recorded. For clusters of two properties (e.g. acidic-aromatic clusters), distances were calculated only between residues of different properties. From these pairs, any that shared a common residue were considered linked and then grouped into a single cluster. Finally, two descriptor values were calculated from each list of clusters; the number of clusters, and the largest cluster size (which is equal to the number of residues in the largest cluster).

2.8 Calculation of Individual Property Map and Overlapping Clusters Descriptors

Using the prepared protein structure files, electrostatic potential maps were generated using the Adaptive Poisson-Boltzmann Solver (APBS) at the desired pH [24]. Protonation states for these electrostatic potential maps were determined by PROPKA [23]. Hydrophobic potential maps of the protein were also generated based on the spatial-aggregation propensity (SAP) algorithm as first published by Chennamsetty et al. [25] and using the Black and Mould hydrophathy index [26] to assign hydrophobic potentials to each surface-exposed protein atom. A uniform grid of points was placed at 1 Å distances throughout the volume of the PDB file. At each grid point, a hydrophobic potential was assigned based on the SAP potential of all nearby atoms weighted by a decay rate of $1/r$.

Two other hydrophobic potential maps were based on using the accessible surface area of atoms within the sidechains of aromatic residues (Phe, Tyr, Trp and His), or those within aliphatic residues (Ala, Leu, Ile, Val and Met) to define atomic potential values. Local maxima were identified based on a contour analysis of the potential map and all adjacent grid points to that maxima were assigned as a cluster. For each cluster, a center of mass (COM) and a potential-weighted average radius for the cluster were determined from the potentials of the grid points within that cluster as follows:

The sum of potential values of grid points that are assigned to a cluster and within a distance of r_{avg} of the COM was defined as the total strength of the cluster. Strong clusters were defined as the top 50% of clusters from a potential map of a protein. Four descriptors were generated for each potential map: the number of clusters, and the total strength, normalized strength and mean strength of the top 25% of clusters. The normalized strength was defined as the total strength divided by the volume of those clusters. The mean strength was defined as the total strength divided by the number of these clusters. Descriptors were also generated for the number of pairs of overlapping strong-strong, strong-weak and weak-weak clusters, defined as those clusters overlapping within r_{avg} from their COMs.

2.9 Calculation of Ligand Descriptors and the Preparation of QSAR Models

Ligand structures were assembled in MOE and underwent energy minimization using the Amber99 forcefield. All available 2D, i3D and x3D descriptors available in the MOE software package were then calculated for each ligand. These descriptors include measures of molecular shape, hydrophilic/hydrophobic volumes/surface areas/moments, and partial charges. Next, a training set was assembled by compiling the various protein and ligand descriptors, the solution pH, and the response value (elution salt concentration) into a comma delimited file (.csv). This training set was loaded into the Yet Another Modeling Software (YAMS) hosted by the Rensselaer Exploratory Center for Cheminformatics Research (RECCR) [27]. Within this program, recursive feature elimination was used to select descriptors for the final model over 12 iterations of selection using intermediate SVM models where the lowest weighted 20% of the descriptors were eliminated after each iteration. The best of four final models (MLR, PLS, SVM and Random Forest) generated by the YAMS software was selected for each dataset. The fitness of the final model was evaluated by the model R^2 (as determined by 1-PRESS/SSR), y-scrambling, the R^2 of the cross-validated model and then finally by measuring the R^2 from the predicted values of an external dataset that was not used to train the model. Acceptable performance in the y-scrambling test was defined as a maximum r^2 of 0.45 for the 20 scrambled-response models as compared to the final model performance constraint where r^2 was required to be greater than 0.9. This ensured that the final model was three standard deviations outside the variation of the scrambled-response models, which would mean that there was a <0.1% chance

that a random training set based on the descriptors selection could have the same performance as the true training set. Acceptable model performance was also defined by achieving an R^2 of 0.85 after averaging the results of 10 rounds of 10-fold crossvalidation, which evaluated the dependence of the model on single data points in the training set.

3. Results and Discussion

3.1 Assembling the Homologous Ligand Library

A library of 9 multimodal cation-exchange prototype resins was assembled to identify chemical moieties and structural motifs with orthogonal modes of selectivity relative to the original Nuvia cPrime ligand. As can be seen in Table 2, these were sorted into groups of ligands examining linker length, linker chemistry, charged group chemistry and solvent exposure of the phenyl ring. While these prototypes varied in ligand density from 60-120 $\mu\text{mol/mL}_{\text{resin}}$, previous work with Nuvia cPrime has shown that both protein selectivity and retention were relatively invariant with ligand density ($R^2 = 0.93$) over the range from 76-126 $\mu\text{mol/mL}_{\text{resin}}$ [22]. Thus, the variation in ligand density between the different prototypes was expected to have minimal effect on the retention behavior in these systems, an observation which has also been noted in several ion-exchange resin systems [28,29]. All of these ligands were immobilized on the same acrylamido gel matrix, which ensured an additional degree of comparability between the various resin materials.

3.2 Separation Distance between Charged and Hydrophobic Moieties

In order to create these charged and hydrophobic multimodal ligands, a linker group is necessary to connect but separate these two chemical groups. The length of this linker will determine the extent to which each moiety can interact independently. A very short separation ($<5 \text{ \AA}$) would force the hydrophobic ring to interact in a region closer to a protein charge. The local environment around this protein charge would likely be preferentially hydrated, which may weaken hydrophobic interactions with the adjacent hydrophobic moiety. For longer linkers, the two moieties could interact more independently as the ligand would have more degrees of conformational freedom. Of the ligand variants in this library, three of them (Prototypes 3, 8 and 9) were designed to examine the effect that the length of the hydrophilic linker arm exerts on protein selectivity.

The chromatographic retention data for the commercial protein library presented in the methods section is given in Figure 1 for both pH 5 and pH 6. As seen in the Figure, the 6 Å linker arm on the original Nuvia cPrime ligand appeared to be the optimal length for enhancing the salt-tolerant retention of proteins on these multimodal surfaces. While this optimal length was observed at both pHs, the differences in retention were more pronounced at pH 5. For acidic proteins ($pI < 6$ in Table 1), the difference between the shorter and longer linker lengths were minimal, with a sharp maximum in protein retention at a linker length of 6 Å (Nuvia cPrime). While longer linker length may enable the hydrophobic moiety to interact more independently of the charged group (minimizing charge repulsion effects on this interaction), this hydrophilic linker could also create a steric barrier to hydrophobic interactions because the hydrophobic moiety was immobilized to the resin surface. Thus, an optimal length would maximize the independence of each moiety while minimizing the steric influence of the linker.

For those basic proteins with minimal hydrophobicity (horse/bovine cytochrome C and ribonuclease A/B), increasing the length of the linker resulted in small increases in the retention of the protein relative to the shortest length, however, a sharp optimum was still evident at 6 Å. As the linker became longer, additional ligands could potentially interact with the adsorbed protein and increase its footprint on the resin surface relative to the short linker resin, thus enhancing the electrostatically-dominant retention of these proteins. Interestingly, for the other basic proteins where hydrophobic interactions were more important, a smaller reduction in protein retention was observed for the short linker length (Prototype 3) resin as compared to Nuvia cPrime. For these proteins, positive charge would be more prevalent on the protein surface, so the ligand would have more freedom to interact with a region that is also adjacent to hydrophobic residues. In addition, a shorter linker would have reduced the steric barrier to hydrophobic associations of the aromatic ring, thus enhancing retention relative to the longest linkers.

3.3 Polar vs. Non-polar Substituents near the Hydrophobic Moiety

The amide bond in the Nuvia cPrime linker group was relatively hydrophilic and could also have affected the electronic properties of the adjacent aromatic ring by extending the delocalized π -bond system towards more electronegative atoms. Both of these effects may have

increased the solubility of the aromatic moiety and thus reduced the potential for hydrophobic associations. This hydrophilic linker group was modified in Prototypes 2 and 5 to increase its hydrophobic potential by substituting the amide bond with a thioether or aliphatic linker, respectively. Since sulfur has a similar electronegativity to carbon, the thioether bond (Prototype 2) is non-polar and would be expected to reduce the solvation of the linker group. Because it lacked any π -bonds, it would also have had no effect on the electronic properties of the adjacent aromatic ring. However, since sulfur has electron lone pairs, the hydrophobic thioether group should also be a compatible π -donor for aromatic interactions, which was proposed as the mechanism of selectivity behind thiophilic chromatographic ligands that included this thioether group. The aliphatic linker (Prototype 5) was non-polar and should not have influenced the electronic properties of the adjacent aromatic ring; however, this carbon-based linker lacked lone pairs of electrons that could have enhanced hydrophobic interactions with the adjacent phenyl ring. A sulfone variant (Prototype 6) was also included in this set to change the chemistry of the linker group while remaining a hydrophilic linker. The sulfone linker is an oxidized thioether bond which makes the group more hydrophilic as well as a π -acceptor in aromatic interactions.

The chromatographic retention data for this ligand set is given in Figure 2. Both of the hydrophobic linker variants (Prototypes 2 and 5) selectively increased the retention of certain hydrophobic proteins of the library, particularly at pH 5. Previous work has indicated that the Nuvia cPrime ligand had a propensity for interacting with proteins displaying exposed aromatic residues [22]. Notably, retention on the thioether linker variant (Prototype 2) was enhanced for those proteins with clusters of exposed aromatic residues and comparable to the performance of the Nuvia cPrime ligand for those proteins without significant aromaticity (lactoglobulins, ribonuclease, cytochrome C, ovalbumin, avidin). While this increased affinity for aromatic residues could be attributed to both the increased hydrophobicity of the linker and the addition of a π -donor group, the former was more likely as the retention of these proteins was also enhanced for the aliphatic linker (Prototype 5) which does not have a π -donor group. For the trypsins, chymotrypsins and α -lactalbumin, retention was significantly higher on the aliphatic linker variant (Prototype 5) as compared to the thioether variant (Prototype 2).

Interestingly, the difference between these two ligands was greatly reduced at pH 6, suggesting that there was an increase in hydrophobicity that was more prominent at low pH. To investigate this further, pH titration curves were calculated for Nuvia cPrime and Prototypes 2 and 5.

As seen in Figure 3, the inflection point in the titration curve is higher for Prototype 5, indicating that the pKa of this ligand is ~6.2, 0.7 pH units above the Nuvia cPrime and Prototype 2 ligands. At pH 6, the effect of this pKa shift was minimal as both the Nuvia cPrime and Prototype 5 resins were significantly charged (76% and 39% respectively). However at pH 5, Prototype 5 was significantly less charged (6%) than the Nuvia cPrime or Prototype 2 ligands (24%), which would result in a more hydrophobic resin surface that could interact with regions of the protein that would have repelled the charged ligand. In contrast, proteins with low hydrophobicity (cytochrome C, ribonuclease, and ovalbumin) were observed to be more weakly retained on Prototype 5, while retention on Nuvia cPrime and Prototype 2 was comparable. The significantly lowered charge density on the Prototype 5 resin at pH 5 would have impacted the retention of these proteins as they are thought to adsorb primarily via electrostatic interactions. Interestingly, the sulfone variant (oxidized version of thioether linker) behaved nearly identical to the Nuvia cPrime ligand with a high degree of correlation in retention behavior for all of the proteins at both pH conditions ($R^2 = 0.94-98$). However, slight increases in retention were observed for lysozyme in Prototype 6 which may suggest that an additional mechanism of interaction contributed to the adsorption of this protein (e.g. aromatic associations). These results suggest that the hydrophobicity of the linker was more influential than any additional interactions afforded by the different atom types or geometries.

3.4 Organization of Substituents around the Aromatic Ring

In addition to the hydrophilicity of the Nuvia cPrime linker group, hydrophobic associations with the aromatic ring could be sterically hindered since the para- position on the ring is used as the resin attachment point. By moving this attachment point closer to the linker group (i.e. ortho-position), more of the aromatic ring surface would be exposed and thus hydrophobic associations potentially enhanced. Since aromatic interactions often involve ring-face conformations [30], the increased exposure of the ring's edge could also increase the

propensity for aromatic interactions to form. To test this hypothesis, two ligand variants (Prototypes 4 and 7) were created where the resin was attached to the ortho-position on the aromatic ring, thus increasing solvent exposure and potentially hydrophobic interactions.

As seen in Figure 4, this geometric re-arrangement of the ligand was observed to enhance the retention of nearly all proteins at pH 5 while being more selective at pH 6. As compared to the thioether variant (Prototype 2), which also increased retention through hydrophobic associations, retention on the ortho-conformation of Nuvia cPrime (Prototype 4) was generally stronger but less selective for proteins with exposed aromatic residues. For those proteins with enhanced retention on both resins, retention was similar which may suggest that both of these modifications are suitable routes to improve ligand affinity for aromatic residues.

An ortho-conformation of Prototype 2 (Prototype 7) was also synthesized and studied to determine whether this geometric rearrangement could be combined with chemical modifications to the linker group to further increase the hydrophobic character of the ligand. As seen in the figure, all of the hydrophobic proteins were increasingly enhanced on this resin, validating the earlier observation that this ortho-conformation of the ligand increases its propensity to interact with all types of hydrophobicity on the protein surface. This ligand appeared to better differentiate those proteins with minimal hydrophobicity (ovalbumin, cytochrome C and ribonuclease) as compared to the ortho-conformation of Nuvia cPrime (Prototype 4), as no difference in retention was observed between the para-conformation ligand (Prototype 2) and ortho-conformation ligand (Prototype 7) for these proteins. This suggested that hydrogen bonding may have contributed to the interaction energy in the case of Prototype 4.

3.5 Strong and Weak Charged Groups in Multimodal Ligands

Figure 5 presents protein retention data for Nuvia cPrime and Prototype 1 that replaces the weak carboxylic acid moiety with a strong sulfonic acid group. In addition, results are presented with weak (CM) and strong (SP) cation exchangers as a control. Strong and weak ion-exchangers are defined not by the strength of interaction, but the pH range over which they are charged. As a result of the carboxylate groups becoming protonated at low pH, the charge density on a CM surface would be lower and thus reduce the potential energy of electrostatic

interactions (both attraction and repulsion) as compared to a strong cation exchanger. At pH 5, the strong SP resin was observed to have similar or enhanced retention of proteins in the library as compared to the weak CM resin. In contrast to IEX, the strong MM resin (Prototype 1) was observed to have generally weaker retention of most protein species at pH 5. At pH 6, both the strong IEX and MM resins demonstrated higher retention of select proteins in the library, although the overall effect was much less pronounced. It is important to note that the linker for the strong MM ligand (Prototype 1) was slightly longer at 7Å, which may also have contributed to the reduction in protein retention in view of the results presented above in Figure 1 for Prototype 8. While the sulfonic acid has a pKa ~2.3 and remains fully charged at both pH 5 and 6, the pKa of the weak carboxylic acid is ~5.5 which means that only 25% of the ligands are charged at pH 5. This leaves many uncharged, hydrophobic ligands that are available to interact with regions of the protein surface where electrostatic repulsion (regions of negative EP) would have reduced the potential of the charged ligand to interact. These uncharged ligands should have little affinity for basic proteins with minimal hydrophobicity (e.g. horse/bovine cytochrome C) and indeed there is no difference in adsorption between the weak and strong MM ligands for these proteins at pH 5. In addition, it would be expected that the charged ligand should have a greater affinity than the uncharged ligand for protein surfaces where positive charges are adjacent to hydrophobic regions (ubiquitin [31], avidin, aprotinin and lysozyme). As a result, no difference should be observed between the strong and weak MM ligands for these proteins at pH 5, which was confirmed by the experimental data. For many of the other proteins, hydrophobic regions are adjacent to negative charges (i.e. α -chymotrypsinogen A [32]) or the net charge is still negative (i.e. ovalbumin), which would repel the charged ligand and prevent it from forming hydrophobic interactions with these surfaces on the protein. Since these surfaces can be accessed by the uncharged ligand, the retention of these proteins should be increased on the weak MM resin relative to the strong Prototype 1 resin.

To further investigate the contribution of uncharged MM ligands, pH gradient studies were performed with a representative protein from each category (α -chymotrypsinogen A, horse cytochrome C and lysozyme) over a range of ionic strength to determine the relationship between pH and salt concentration on the elution behavior for these resins. During these pH experiments, the protein surface potential becomes progressively more negative with increasing

pH which induces electrostatic repulsion with the negative surface of the resin and facilitates elution. An increase in the ionic strength of the solution would be expected to lower the elution pH by reducing the strength of electrostatic attraction between the protein and the surface. Conversely, increased ionic strength would strengthen hydrophobic interactions and could raise the elution pH at high salt concentrations. The results of this screen for Nuvia cPrime, Prototype 1, CM and SP Sepharose Fast Flow can be found in Figure 6, where dashed lines indicate the limits of the pH gradient from pH 4 to pH 11.

As expected, on the ion-exchange resins since no hydrophobic interactions can occur with these ligands, elution pH was quickly lowered by increasing the ionic strength until the proteins were no longer retained on the column. On the Nuvia cPrime resin, this relationship was observed at low ionic strength where the protein eluted at high pH. Beyond a critical pH (~4.3), the elution pH of α -chymotrypsinogen A and horse cytochrome C became insensitive to further increases in the ionic strength of the solution. For lysozyme, the elution pH began to increase at high ionic strength, which indicated that hydrophobic interactions were now dominant and increasing electrostatic repulsion was needed to facilitate elution of the protein. As expected, the strong multimodal ligand (Prototype 1) exhibited hybrid behavior between the ion-exchange ligands and Nuvia cPrime. For α -chymotrypsinogen A, hydrophobic regions were separated from positive charges and thus the protein was no longer retained once the electrostatic interactions were mitigated by a high solution ionic strength. For horse cytochrome C, hydrophobic regions were expected to be insignificant and thus the protein was also unretained at high ionic strength. This occurred at a higher ionic strength than with α -chymotrypsinogen A as the charge density on the surface of horse cytochrome C was much higher. Since lysozyme was thought to have hydrophobic regions with adjacent positive charges, the strong multimodal ligand could interact hydrophobically with the protein while remaining charged. Since strong hydrophobic interactions could occur while the ligand was still charged, the elution pH never reached that critical pH where the ligand becomes fully uncharged. In addition, the elution conductivities of the salt-based linear gradient separations were also plotted at pH 5 and pH 6 (hollow points on Figure 6). Interestingly, they appear to be quantitatively comparable with the data obtained from the pH gradient experiments, notably predicting the failure to elute lysozyme in any salt concentration at pH 5. This indicates that the data obtained using either pH or salt-based linear

gradients could be considered interchangeable and could be used in tandem to efficiently scan a given design space.

3.6 Development of a Unified QSAR Model for the Prediction of Multimodal Resin Library

Using the protein surface descriptors and the ligand molecular descriptors described in the experimental section, a Quantitative Structure-Activity Relationship (QSAR) model was generated that encompassed full chemical diversity of the prototype resins in this homologous ligand library. In addition, datasets from several commercially-available cation-exchange resins (SP and CM Sepharose Fast Flow) and other multimodal cation-exchange resins (Capto MMC and Toyopearl MX-Trp 650M) were added to increase the diversity of the ligand dataset. Of the original dataset of 19 proteins, 14 resins and 2 pH conditions, 4 resins were selected at random (Nuvia cPrime, Prototypes 4 and 8, and SP Sepharose Fast Flow) and reserved as an external test set for the model.

Using recursive feature elimination based on SVM regression, the initial set of 172 protein descriptors and 8 ligand descriptors was reduced to a concise set of 8 protein descriptors and 3 ligand descriptors (Figure 7,) which was found to be the optimal set that maximized model accuracy while minimizing the potential overfitting of the model parameters to the training set as was confirmed by the internal model validation methods. From this concise descriptor set, an SVM training model was generated ($R^2 = 0.90$) that sufficiently predicted the data within the external test set ($R^2 = 0.91$) (Figure 8a and 8b). Internal validation methods (10-fold cross-validation: $R^2 = 0.82$, and y-scrambling: $R^2_{\text{max}} = 0.32$) also confirmed the accuracy of the SVM training model. As can be seen in Figure 8, the model was in general well suited for predicting the data within the external test set. While the model accuracy was quite good for most of the resins, the predictive ability was weaker for ligands at the extreme ends of this ligand property space (Capto MMC, SP Sepharose Fast Flow and Prototype 7). This could potentially be attributed to the relative abundance of cases where proteins were unable to bind (e.g. SP resin) or were not recovered in the gradient (e.g. Capto MMC and Prototype 7). Assigning the maximum or minimum concentration of the linear gradient to these proteins may not be a close enough approximation to represent the true strength of protein interactions with the resin surface. For

example, the pH gradient experiments for lysozyme (which was fully retained on most resins at pH 5) showed that the protein would never desorb from the Nuvia cPrime and Prototype 1 resins at pH 5, therefore it is not surprising that the QSAR model was unable to accurately predict an elution concentration for this protein on many of these multimodal resins. Another possible explanation for the lower predictive performance of the model for these three resins is that these ligands were further away in property space (Figure 7) than the ligands employed in the training set. This is a common phenomenon in machine-learning models, where interpolation is generally more accurate than extrapolation at generating correct predictions of the experimental phenomena. Excluding these fully retained proteins (which constitute 7% of the training set and 2% of the test set) for the aforementioned reasons, 95% of both the training set and test set predictions fell within ± 200 mM NaCl of the actual data values, while the equivalent 95% confidence interval for replicates in the experimental data was ± 100 mM NaCl.

The protein descriptors selected by the current model (Figure 7) were very similar in character to those selected for the Capto MMC and Nuvia cPrime models reported in the first paper in this series [22]. This suggests that all of these multimodal cation-exchange ligands recognize similar protein surface features (albeit to varying degrees). The current model includes descriptors for both aliphatic and aromatic clusters on protein surfaces, which were previously shown to be effective in classifying differences in protein retention behavior on the Capto MMC and Nuvia cPrime systems. The two selected descriptors that measure aromatic clusters quantified clusters either in proximity to basic residues (which can form highly favorable interactions with charged ligands) or next to acidic residues (which could form stronger interactions with uncharged ligands). The distinction between hydrophobic regions based on the adjacent electrostatic potential was also thought to be important in explaining protein selectivity trends for Prototypes 1 and 5 which had noticeably different proportions of charged ligands at pH 5 as compared to the original Nuvia cPrime ligand. Descriptors for both negative and positive EP, basic residue clusters and solution pH were also included in the model to account for the attractive and repulsive forces generated between the protein and the resin surface at a given pH condition.

The most important ligand descriptor (Figure 7) identified during feature selection was the capacity factor at -0.5 kcal/mol as defined by Cruciani et al. [33]. This descriptor measures the ratio of hydrophilic surface area to total surface area and was assigned a negative weight in the model. This descriptor is inversely proportional to the hydrophobic surface area of the ligand, indicating that an increase in surface area or exposure of the aromatic ring increases protein retention in these MM systems. This corroborated the data obtained with the ortho-conformation ligands (Prototypes 4 and 7), where an increase in the hydrophobic surface area of the ligand indeed increased protein retention relative to their corresponding para-conformation ligands (Nuvia cPrime and Prototype 2 respectively). The critical packing parameter (the ratio of hydrophobic to hydrophilic volume) was also selected by the model (Figure 7) and was able to recognize the increased hydrophobicity introduced by changes in the linker groups present in Prototypes 2, 3 and 5. These changes reduced the hydrophilic volume of the ligand by substituting/removing polar atoms found in the amide bond. The final and least important ligand descriptor selected in the model was the fraction of the accessible surface area consisting of negatively charged atoms. As can be seen in Figure 7, this descriptor indicated that Prototypes 3, 5 and 8 were less negatively charged than Nuvia cPrime. This could explain why positively charged, but hydrophilic proteins (e.g. ribonuclease, cytochrome C) were more weakly retained on these resins as the electrostatic attraction may have been weaker at these lower negative charge densities.

4. Conclusions

From the current study, one could speculate on some potential guidelines for the design of future multimodal ligands. It appeared that optimizing the charged properties of the ligand will have minimal effect because the main driver for enhanced protein selectivity in this ligand library came from thoughtful augmentation of hydrophobic properties to the protein-ligand association. Choosing modalities with more defined interaction states (aromatics, π -donor/acceptors) allowed ligand geometry to play a larger role in defining the selective behavior of this complex ligand for proteins with similar modalities (i.e. exposed aromatic residues). In contrast, simply increasing the solvent exposure of the aromatic ring was observed to strengthen the ligand affinity for both types of hydrophobic residues. Further enhancement of hydrophobic properties (e.g. fused ring systems or an increased number of interaction modalities) should be

viewed with caution as these modifications will likely enhance affinity but may also increase its promiscuity for different protein targets, reduce the recovery of adsorbed species, or risk the hydrophobic collapse of immobilized ligands onto the matrix support. Studies using pH gradients at various ionic strengths showed that the elution of most proteins became increasingly insensitive to ionic strength at low pH on the weak MM-CEX Nuvia cPrime resin, while the strong MM-CEX Prototype 1 resin and both strong and weak ion-exchange resins remained sensitive to ionic strength. These findings demonstrate that these weak MM ligands can be used in a hydrophobic charge induction chromatography mode, creating new avenues for generating selectivity between proteins.

Finally, a QSAR model was trained on this experimental data, which identified numerical descriptors that quantified critical ligand properties for multimodal cation-exchange resins in addition to important protein property descriptors and could correctly predict the retention of proteins on multimodal and ion-exchange resins that were not included in the original model. The development of QSAR models for the prediction of protein retention behavior in a range of multimodal and ion exchange systems could be extremely useful for facilitating methods development for the purification of protein biologics. In addition, this QSAR model could be used to screen a wider array of novel multimodal ligands and the most promising candidates with superior resolution could then be synthesized to experimentally confirm the predicted performance. The resultant data could be fed back into the training set of the model, thus expanding the ligand property space and improving the prediction of future ligand performance. This ligand property map could also be used to select a concise set of ligands that effectively cover the useful property space without dramatically expanding experimental design space for developing a new multimodal separation process.

5. Acknowledgments

This work was supported by NSF Grant CBET 1150039 and Bio-Rad Laboratories Inc. The authors thank Prof. Curt Breneman and Dr. Michael Krein for their assistance in using the YAMS software and validating the final models.

572 6. References

- 573 [1] E. Boschetti, Antibody separation by hydrophobic charge induction chromatography,
574 Trends in Biotechnology 20 (2002) 333-337.
- 575 [2] Z. el Rassi, C. Horvath, Tandem columns and mixed-bed columns in high-performance
576 liquid chromatography of proteins, J Chromatogr 359 (1986) 255-264.
- 577 [3] B. Buszewski, R.M. Gadzala-Kopiuch, M. Jaroniec, Chromatographic Properties of
578 Mixed Chemically Bonded Phases with Alkylamide and Aminopropyl Ligands, Journal
579 of Liquid Chromatography & Related Technologies 20 (1997) 2313-2325.
- 580 [4] B.L. Johansson, M. Belew, S. Eriksson, G. Glad, O. Lind, J.L. Maloisel, N. Norrman,
581 Preparation and characterization of prototypes for multi-modal separation aimed for
582 capture of positively charged biomolecules at high-salt conditions, J Chromatogr A 1016
583 (2003) 35-49.
- 584 [5] D. Gao, D.-Q. Lin, S.-J. Yao, Protein adsorption kinetics of mixed-mode adsorbent with
585 benzylamine as functional ligand, Chemical Engineering Science 61 (2006) 7260-7268.
- 586 [6] K. Zhao, L. Yang, X. Wang, Q. Bai, F. Yang, F. Wang, Preparation of a novel dual-
587 function strong cation exchange/hydrophobic interaction chromatography stationary
588 phase for protein separation, Talanta 98 (2012) 86-94.
- 589 [7] Y. Hou, C.J. Morrison, S.M. Cramer, Classification of protein binding in hydroxyapatite
590 chromatography: synergistic interactions on the molecular scale, Anal Chem 83 (2011)
591 3709-3716.
- 592 [8] J. Chen, J. Tetrault, Y. Zhang, A. Wasserman, G. Conley, M. DiLeo, E. Haimen, A.E.
593 Nixon, A. Ley, The distinctive separation attributes of mixed-mode resins and their
594 application in monoclonal antibody downstream purification process, Journal of
595 Chromatography A 1217 (2010) 216-224.
- 596 [9] K.A. Kaleas, C.H. Schmelzer, S.A. Pizarro, Industrial case study: Evaluation of a mixed-
597 mode resin for selective capture of a human growth factor recombinantly expressed in E.
598 coli, Journal of Chromatography A 1217 (2010) 235-242.
- 599 [10] T. Arakawa, Y. Kita, H. Sato, D. Ejima, MEP chromatography of antibody and Fc-fusion
600 protein using aqueous arginine solution, Protein Expression and Purification 63 (2009)
601 158-163.
- 602 [11] J. Paul, S. Jensen, A. Dukart, G. Cornelissen, Optimization of a preparative multimodal
603 ion exchange step for purification of a potential malaria vaccine, Journal of
604 Chromatography A 1366 (2014) 38-44.
- 605 [12] K.A. Kaleas, M. Tripodi, S. Revelli, V. Sharma, S.A. Pizarro, Evaluation of a multimodal
606 resin for selective capture of CHO-derived monoclonal antibodies directly from harvested
607 cell culture fluid, Journal of Chromatography B 969 (2014) 256-263.
- 608 [13] L.S. Wolfe, C.P. Barringer, S.S. Mostafa, A.A. Shukla, Multimodal chromatography:
609 Characterization of protein binding and selectivity enhancement through mobile phase
610 modulators, Journal of Chromatography A 1340 (2014) 151-156.
- 611 [14] G. Zhao, X.-Y. Dong, Y. Sun, Ligands for mixed-mode protein chromatography:
612 Principles, characteristics and design, Journal of Biotechnology 144 (2009) 3-11.
- 613 [15] B.L. Johansson, M. Belew, S. Eriksson, G. Glad, O. Lind, J.L. Maloisel, N. Norrman,
614 Preparation and characterization of prototypes for multi-modal separation media aimed
615 for capture of negatively charged biomolecules at high salt conditions, J Chromatogr A
616 1016 (2003) 21-33.

- [16] T. Yang, C.M. Breneman, S.M. Cramer, Investigation of multi-modal high-salt binding ion-exchange chromatography using quantitative structure–property relationship modeling, *Journal of Chromatography A* 1175 (2007) 96-105.
- [17] S.J. Mountford, R. Daly, E.M. Campi, A. Bhagwat, A.J. Robinson, M.T.W. Hearn, N-Heterocyclic-based adsorbents for antibody purification—effect of ligand structure, *Journal of Molecular Recognition* 27 (2014) 285-295.
- [18] D.-Q. Lin, H.-F. Tong, H.-Y. Wang, S.-J. Yao, Molecular insight into the ligand–IgG interactions for 4-mercaptoethyl-pyridine based hydrophobic charge-induction chromatography, *The Journal of Physical Chemistry B* 116 (2012) 1393-1400.
- [19] K.L. Knudsen, M.B. Hansen, L.R. Henriksen, B.K. Andersen, A. Lihme, Sulfone-aromatic ligands for thiophilic adsorption chromatography: Purification of human and mouse immunoglobulins, *Analytical biochemistry* 201 (1992) 170-177.
- [20] J. Porath, F. Maisano, M. Belew, Thiophilic adsorption - a new method for protein fractionation, *FEBS Letters* 185 (1985) 306-310.
- [21] A. Schwarz, F. Kohen, M. Wilchek, Novel sulfone-based thiophilic ligands for the high-performance liquid chromatographic purification of antibodies, *Reactive Polymers* 22 (1994) 259-266.
- [22] J.A. Woo, S. Parimal, M.R. Brown, R. Heden, S.M. Cramer, The effect of geometrical presentation of multimodal cation-exchange ligands on selective recognition of hydrophobic regions on protein surfaces, *Journal of Chromatography A* (2015).
- [23] H. Li, A.D. Robertson, J.H. Jensen, Very fast empirical prediction and rationalization of protein pKa values, *Proteins: Structure, Function, and Bioinformatics* 61 (2005) 704-721.
- [24] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences* 98 (2001) 10037-10041.
- [25] N. Chennamsetty, V. Voynov, V. Kayser, B. Helk, B.L. Trout, Prediction of Aggregation Prone Regions of Therapeutic Proteins, *The Journal of Physical Chemistry B* 114 (2010) 6614-6624.
- [26] S.D. Black, D.R. Mould, Amino acid scale: Hydrophobicity of physiological L-alpha amino acids, *Analytical biochemistry* 193 (1991) 72-82.
- [27] M. Krein, T.-W. Huang, L. Morkowchuk, D.K. Agrafiotis, C.M. Breneman, Developing Best Practices for Descriptor-Based Property Prediction: Appropriate Matching of Datasets, Descriptors, Methods, and Expectations, *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, Wiley-VCH Verlag GmbH & Co. KGaA, 2012, p. 33-64.
- [28] D. Wu, R.R. Walters, Effects of stationary phase ligand density on high-performance ion-exchange chromatography of proteins, *Journal of Chromatography A* 598 (1992) 7-13.
- [29] H. Aono, I. Iliescu, D. Cecchini, S. Wood, J.T. McCue, Mitigation of chromatography adsorbent lot performance variability through control of buffer solution design space, *Journal of Chromatography A* 1318 (2013) 198-206.
- [30] S. Burley, G. Petsko, Aromatic-aromatic interaction: a mechanism of protein structure stabilization, *Science* 229 (1985) 23-28.
- [31] M.A. Holstein, W.K. Chung, S. Parimal, A.S. Freed, B. Barquera, S.A. McCallum, S.M. Cramer, Probing multimodal ligand binding regions on ubiquitin using nuclear magnetic resonance, chromatography, and molecular dynamics simulations, *Journal of Chromatography A* 1229 (2012) 113-120.

[32] S. Parimal, Chemical and Biological Engineering, Rensselaer Polytechnic Institute, 2014, p. 147.

[33] G. Cruciani, P. Crivori, P.A. Carrupt, B. Testa, Molecular fields in quantitative structure–permeation relationships: the VolSurf approach, Journal of Molecular Structure: THEOCHEM 503 (2000) 17-30.

Figure 1: Chromatographic retention data of the protein library on Nuvia cPrime and linker arm prototypes (P3, P8 and P9) under linear salt gradient conditions (1.5 M sodium chloride over 45 column volumes). Non-eluting proteins are indicated by an asterisk. (a) Retention at pH 5. (b) Retention at pH 6.

Figure 2: Chromatographic retention data of the protein library on Nuvia cPrime (low) and linker group prototypes (P2, P5 and P6) under linear salt gradient conditions (1.5 M sodium chloride over 45 column volumes). Non-eluting proteins are indicated by an asterisk. (a) Retention at pH 5. (b) Retention at pH 6.

Figure 3: Titration of Nuvia cPrime resin and Prototypes 2, 5 and 6

Figure 4: Chromatographic retention data of the protein library on Nuvia cPrime (low) and hydrophobic group prototypes (P2, P4 and P7) under linear salt gradient conditions (1.5 M sodium chloride over 45 column volumes). Non-eluting proteins are indicated by an asterisk. (a) Retention at pH 5. (b) Retention at pH 6.

Figure 5: Chromatographic retention data of the protein library on Nuvia cPrime, Prototype 1, CM and SP Sepharose Fast Flow under linear salt gradient conditions (1.5 M sodium chloride over 45 column volumes). Non-eluting proteins are indicated by an asterisk. (a) Retention at pH 5. (b) Retention at pH 6.

Figure 6: pH gradient retention data for Nuvia cPrime, Prototype 1, CM and SP Sepharose Fast Flow (pH 4.0 to pH 11.0 over 45 column volumes) at various NaCl concentrations (0, 100, 150, 250, 500, 1000 and 1500 mM). Non-eluting proteins are indicated by an asterisk. Hollow points indicate retention at pH 5 or pH 6 in a 0-1.5 M NaCl linear gradient. (a) α -chymotrypsinogen A, (b) horse cytochrome C, (c) α -chymotrypsin and (d) lysozyme.

695
696 Figure 7: (a) and (b) Diagram of multimodal ligands in property space defined by selected ligand
697 descriptors. Ligands selected for test set predictions are circled in black. (c) Protein surface
698 property descriptors selected for the QSAR model.
699
700 Figure 8: Predictive QSAR model for the multimodal and ion-exchange resin library.
701 (a) Training model. (b) External test set. Dashed lines indicate error bars of $\pm 200\text{mM NaCl}$.

Figure 1

Figure 2

Figure 3

Figure 3: Titration of Nuvia cPrime resin and Prototypes 2, 5 and 6

Figure 4

Figure 5

Figure 6

Figure 7

C. Protein Surface Property Descriptors	
Aliphatic Clusters (Mean Strength)	Negative EP Clusters (Total Strength of Top 25%)
Aromatic/Acidic Residues (Largest Cluster)	Positive EP Patches (Separated by 5 - 7 Å)
Aromatic/Basic Residues (Largest Cluster)	Solution pH
Basic Residues (Number of Clusters)	

Figure 8

Table 1

Table 1: Summary of proteins used in the linear gradient retention studies.

Protein	PDB Code	pI	Size (kDa)	Mean Potential of Aromatic Clusters	Mean Potential of Aliphatic Clusters	Mean Potential of Hydropathy Clusters
Aprotinin	1PIT	10.5	6.5	1.88	3.87	1.23
Avidin	1VYO	9.69	28.7	2.15	3.96	1.79
Bovine Serum Albumin	3V03	5.82	66.3	4.67	3.99	2.22
Conalbumin	1AIV	6.69	75.8	2.73	4.06	1.17
Horse Cytochrome C	1HRC	10.25	11.7	0.83	1.71	0.51
Human Serum Albumin	1AO6	5.67	66.4	2.28	3.24	1.18
Lysozyme	1AKI	11.35	14.3	2.05	2.60	0.94
Ovalbumin	1OVA	4.9	42.7	1.54	3.38	0.54
Papain	9PAP	8.88	23.4	5.02	1.31	2.21
Ribonuclease A	1RBX	9.45	13.7	1.24	1.67	0.68
Ribonuclease B	1RBB	8.9	13.7	1.24	1.67	0.68
Bovine Trypsin	1S0Q	10.3	23.3	1.93	2.18	1.17
Porcine Trypsin	1S81	10.5	23.5	1.72	2.85	0.93
Ubiquitin	1UBQ	6.79	8.6	1.70	5.40	2.23
α -Chymotrypsin	5CHA	9.17	25.2	2.44	5.01	1.46
α -chymotrypsinogen A	2CGA	8.52	25.7	2.24	3.40	1.07
α -lactalbumin	1F6S	5	14.1	1.38	3.28	0.94
β -lactoglobulin A	1B0O	5.1	18.2	2.27	5.63	1.88
β -lactoglobulin B	1BSQ	5.1	18.3	0.49	3.29	0.88

Table 2

Table 2: Summary of ligand structures and characteristics for mixed-mode resin prototypes.

Length-Based Derivatives			
Prototype 3	Nuvia cPrime	Prototype 8	Prototype 9
4-aminophenyl acetic acid	4-aminohippuric acid	2-(2-(4-aminophenyl)acetamido)acetic acid	2-(2-(4-aminobenzamido)acetamido)acetic acid
92 µmol/mL	126 µmol/mL	92 µmol/mL	80 µmol/mL
Average Distance between Charged and Hydrophobic Moieties			
4 Å	6 Å	7 Å	8.5 Å
Linker Group Variants			
Nuvia cPrime	Prototype 5	Prototype 2	Prototype 6
4-aminohippuric acid	4-(4-aminophenyl)butyric acid	2-((4-aminophenyl)thio)acetic acid	2-((4-aminophenyl)sulfonyl)acetic acid
126 µmol/mL	114 µmol/mL	86 µmol/mL	60 µmol/mL
Charged Group Derivatives		Solvent-Exposure Derivatives	
Nuvia cPrime	Prototype 1	Prototype 4	Prototype 7
4-aminohippuric acid	2-(4-aminobenzamido)ethanesulfonate	2-aminohippuric acid	2-((2-aminophenyl)thio)acetic acid
126 µmol/mL	120 µmol/mL	88 µmol/mL	70 µmol/mL