BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

*Biochemistry* **2010,** *49,* 1549–1555 1549
DOI: 10.1021/bi9021044

# Identification and Sequence Composition Characterization of Chondroitin Sulfate-Binding Peptides through Peptide Array Screening†

Karen Chao Butterfield,‡ Michael Caplan,‡ and Alyssa Panitch*,§

‡*Harrington Department of Bioengineering, Arizona State University, Tempe, Arizona 85287-9709, and* §*Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47907-2032*

*Received December 8, 2009; Revised Manuscript Received January 22, 2010*

ABSTRACT: Chondroitin sulfate (CS) is an important glycosaminoglycan that has been implicated in several disease processes, such as cancer and spinal cord injury. However, few studies have characterized CS-binding protein and peptide sequences for diagnostic and therapeutic use. In this study, peptide array screening, affinity capillary electrophoresis, and statistical analysis were used to both identify and characterize C6S-binding peptides for sequence composition. The compositional characterization results showed that Phe, Arg, and Tyr all had a significantly high rate of occurrence in the "high binding" affinity peptides, while tryptophan and lysine were significantly underrepresented in this population. Peptides modified with alanine point mutations for Phe, Arg, and Tyr all had lower C6S-binding affinities than the original peptides, demonstrating that these amino acids are all important for C6S binding. Several peptides were designed that substituted Arg for Lys and Phe or Tyr for Trp to create peptides with higher binding affinity. The peptides with the Arg substitution all had improved binding affinities while the Phe/Tyr substitution decreased C6S-binding affinity. Further analysis showed that the increased occurrence of Phe and Tyr in the "high affinity" peptides was dependent upon their positions both within the peptide sequence and in relation to other critical amino acids. Finally, a motif (ABBAA) was suggested for C6S-binding peptides where A represents any aromatic amino acid and B any basic amino acid. The results demonstrate that the methodology developed in this study for sequence composition analysis is an effective technique for the characterization of the interaction between peptides and CS.

Chondroitin sulfate (CS) is a member of the glycosaminoglycan (GAG) family. GAGs are highly charged linear polysaccharides that are generally found within the extracellular matrix (ECM) or on the cell surface. GAGs play an important role in several physiological processes from cancer progression (*1*, *2*) to spinal cord injury (*3*−*5*) and act mainly through their ability to specifically bind and modulate proteins to influence cell behavior (*6*, *7*). Studies have also shown that growth factor activity can be regulated by CS binding through sequestration to prevent degradation and through localization of growth factor activity to the cell surface (*8*). CS also interacts with various ECM molecules, such as collagen and elastin, to influence ECM organization which in turn affect cell behavior (*9*−*11*).

Even though CS is the most prevalent GAG, few studies have focused on both identifying and characterizing the peptide sequence of CS-binding proteins. Identification of CS-binding proteins and peptide sequences has been difficult due to the heterogeneous population of polysaccharide chains with variable chain length, sulfation patterns, and conformations found throughout the body. The diverse sulfation patterns of CS have been postulated to function as molecular recognition motifs for growth factors, chemokines, and other proteins (*7*). Several studies focused on the sulfation pattern by identifying specific GAG structures that bind to specific proteins (*12*, *13*). Other studies have employed immunological methods (*14*−*16*) or phage display (*17*−*20*) to produce antibodies or antibody fragments that bind specifically to different GAGs. However, these studies lack sufficient information for sequence analysis to determine which amino acids are important for CS binding. The identification of amino acids critical for CS binding can lead to a better understanding of the key chemical interactions involved in CS-binding proteins. Furthermore, the results from the sequence composition analysis can be used to design improved CS-binding peptides for diagnostic and therapeutic use.

Identification and sequence composition characterization of CS-binding peptides were accomplished through peptide array screening. This technique has emerged as a fast high-throughput method for identifying proteins and peptides that bind to a target molecule (*21*−*23*). Even though phage display and other directed evolution technologies sample from a larger starting library ($\sim 10^7$), peptide array has the advantage of fewer rounds of selection and does not require cloning for sequence identification. In this study, chondroitin 6-sulfate (C6S) was screened on a peptide array to identify CS-binding peptides. Several peptides were selected from the peptide array and characterized through affinity capillary electrophoresis (ACE). Statistical analysis was performed for sequence composition analysis to identify several amino acids that were significantly overrepresented in the "high binding" affinity peptides. Alanine point mutations in the CS-binding peptides confirmed that the amino acids identified in the compositional analysis contributed to CS-binding affinity. This information was utilized to design peptides with increased

†This research was funded in part by the Flinn Foundation, American Heart Association Pacific Mountain Affiliate, Arizona State University Graduate College, and the National Science Foundation CAREER: CBET 0651643.

*Corresponding author: e-mail, apanitch@purdue.edu; telephone, 765-496-1313; fax, 765-496-1459.


Published on Web 01/22/2010

Table 1: Characteristics of C6S-Binding Peptides

| peptide name | sequence | charge[a] | fluorescence values[b] | | dissociation constant ($K_d$)[c] | |
|---|---|---|---|---|---|---|
| | | | initial binding | after wash (0.4 M NaCl) | C6S[d] ($\mu$M) | C4S[d] ($\mu$M) |
| C6S-1 | EKRIWFPYRRF | +3 | 39207 | 33383 | 1.14 ± 0.13 | 16.29 ± 1.16 |
| C6S-2 | YKTNFRRYYRF | +4 | 38996 | 39918 | 2.31 ± 0.10 | 10.41 ± 0.78 |
| C6S-3 | VLIRHFRKRYY | +4 | 42472 | 32883 | 13.39 ± 1.14 | 58.01 ± 2.62 |
| C6S-4 | RRFMQYSARRF | +4 | 34478 | 25644 | 12.83 ± 0.54 | 88.68 ± 2.74 |
| C6S-5 | YIKRKHRYFQR | +5 | 22931 | 18576 | 12.21 ± 0.38 | 44.22 ± 2.69 |
| C6S-6 | RRMQKTRLQHR | +5 | 23782 | 18827 | 55.18 ± 3.80 | 71.02 ± 2.13 |
| C6S-7 | KNYKFRFGTRV | +4 | 26767 | 16204 | 63.68 ± 2.39 | 142.9 ± 8.32 |

[a]Calculated by subtracting the number of basic residues from the number of acidic residues. [b]Unitless relative fluorescent intensities. [c]$K_d$ reported as mean ± SE. [d]All peptides had statistically significant dissociation constants between C6S and C4S. Statistical significance was determined by nonoverlapping 95% CI for the two dissociation constants.

CS-binding affinity. This study has shown that sequence composition analysis is a powerful tool for elucidating the interaction between CS and peptides and can be applied toward the design of improved CS-binding peptides for diagnostic and therapeutic use.

## MATERIALS AND METHODS

*Peptide Array Screening for C6S-Binding Peptides*. Chondroitin 6-sulfate (MW∼20000 g/mol; Sigma-Aldrich, St. Louis, MO) from shark cartilage was modified with a glycosylamine reaction and conjugated to Alexa dye 647 (Invitrogen, Eugene, OR) before shipping to LC Sciences (Houston, TX) for peptide array screening. C6S (∼10 mg) was suspended in a saturated ammonium bicarbonate solution and shaken at room temperature for 4−5 days. The solution was then lyophilized several times to remove excess ammonium bicarbonate. The amine-modified C6S was fluorescently labeled through conjugation to Alexa Fluor 647 carboxylic acid succinimidyl ester following the Invitrogen protocol (A32757). Briefly, the Alexa dye was resuspended in 10 $\mu$L of $H_2O$, added to a solution of amine-modified C6S (400 $\mu$g) in 1 M sodium bicarbonate buffer (100 $\mu$L), and incubated in the dark at room temperature. After 1 h, 5 $\mu$L of 1.5 M hydroxylamine (pH 5.5) was added and incubated for 15 min. The reaction mixture was diluted in 500 $\mu$L of 1 × PBS, and excess dye was removed through spin filtration (Microcon MWCO 10000; Millipore). The Alexa dye 647 conjugated C6S was stored in 1 × PBS at −20 °C until shipment.

Peptide array synthesis and screening were performed at LC Sciences. The peptide array consisted of 1831 unique 11-mer peptide sequences with the probability of each amino acid occurring at each position within the peptide sequence determined by an initial amino acid distribution. To increase the probability of identifying CS-binding peptides of high binding affinity, the initial amino acid distribution was derived from compositional analysis of known GAG-binding peptide sequences (*24*, *25*). Previous studies have shown that basic residues are important for heparin binding so the presence of Arg and Lys residues were highest in the amino acid distribution. Tyr is also an enriched amino acid found in heparin-binding protein sites but has a smaller presence than the basic residues leading to a relatively lower probability in the initial distribution (*26*). Acidic residues (Glu, Asp) and small nonpolar residues (Ala, Gly) were assumed to play minor roles in CS binding so the probability for their occurrence in the amino acid distribution was very low. Finally, cysteines were not included in the distribution to inhibit disulfide binding within the linear peptide sequence. Several quality control probes were included on the chip to ensure the fidelity of peptide synthesis.

For peptide array screening, the chip was initially placed in blocking buffer (1 × PBS, 1% BSA, 0.05% Tween-20, pH 7.4) overnight at 4 °C to reduce nonspecific binding. After blocking, binding buffer (1 × PBS, 1% BSA, pH 7.4) containing 10 nM Alexa Fluor 647 conjugated C6S was circulated over the chip at 25 °C for 1 h. The chip was washed with 1 × PBS buffer and scanned for fluorescence intensity after the initial binding. To reduce nonspecific binding due mainly to electrostatic interactions, the chip was washed with 0.4 M NaCl in 1 × PBS. After the initial binding and salt wash, the chip was scanned for fluorescence, and the detected fluorescence value was correlated to a peptide sequence; all prewash and postwash fluorescent intensity data are provided in Supporting Information.

*Peptide Synthesis and Purification*. Several peptide sequences (Table 1) that were chosen from the peptide array of varying fluorescence values were synthesized and characterized for binding affinity through affinity capillary electrophoresis (ACE). Peptides were synthesized through standard Fmoc-solid-phase chemistry on a Symphony peptide synthesizer (Protein Technologies, Inc., Tucson, AZ) at 200 $\mu$M with rink amide resin. Fmoc-protected amino acids were activated by HBTU and added to the growing peptides in 5 M excess. After peptide synthesis, the peptides were cleaved from the resin using a cleave cocktail containing 92.5% trifluoroacetic acid (TFA), 2.5% water, 2.5% triisopropylsilane, and 2.5% ethanedithiol. Peptides were precipitated in 10× excess ice-cold diethyl ether and centrifuged at 5000 rpm for 30 min. The ether was decanted off, and the peptide was resolubilized in water before lyophilization.

Peptide purification was performed using reverse-phase chromatography with an AKTA Explorer FPLC (Amersham Biosciences, Piscataway, NJ) equipped with a C18 column (Grace Vydac, Hesperia, CA; 22 mm internal diameter, 250 mm length, 10−15 $\mu$m particle size). After the column was equilibrated with 5 column volumes of water containing 0.1% TFA, the peptide was loaded onto the column and subjected to an increasing linear gradient from 0% to 60% acetonitrile containing 0.1% TFA over 12 column volumes. The collected peptides were lyophilized, and the mass of each peptide was confirmed with MALDI-TOF mass spectrometry on a Voyager-DE STR spectrometer (Applied Biosystems, Foster City, CA) using an α-cyano-4-hydroxycinnamic acid matrix (Sigma-Aldrich).

*Affinity Capillary Electrophoresis*. Affinity capillary electrophoresis (ACE) was performed to characterize the binding affinity of CS-binding peptides by determining the dissociation

constant between the peptides and C6S and chondroitin 4-sulfate (C4S, MW~2000 g/mol; Wako Chemicals, Richmond, VA). ACE was performed on a Beckman PACE M/DQ (Fullerton, CA) following a protocol described by McKeon and Holland (*27*). The capillary consisted of a 75 $\mu$m uncoated fused-silica capillary of total length 40.2 cm that was prerinsed with 0.1 M NaOH and $H_2O$ before each separation. The peptide (500 $\mu$M) with a neutral marker (0.01% DMF) was injected in the reverse direction at 1 psi for 5 s and monitored through UV detection at 200 nm. Separations were carried out in the reverse direction at 8 kV at 20 °C in 50 mM sodium phosphate buffer (pH 7.5) with varying concentration of CS. Since the separations were in the reverse direction, the length to the detector was 10.2 cm. For each peptide, several concentrations of CS were tested in triplicate, and the peak migration time was determined for the neutral marker and peptide. The change in mobility ($\delta_{\text{mobility}}$) due to CS binding was determined by eq 1:

$$\delta_{\text{mobility}} = \frac{L_w L_t}{V}\left(\left(\frac{1}{t_{cs}} - \frac{1}{t_{eofcs}}\right) - \left(\frac{1}{t_{free}} - \frac{1}{t_{eof}}\right)\right) \quad (1)$$

where $L_w$ is the length to the detector, $L_t$ is the total length, $V$ is the separation voltage, $t_{cs}$ is the migration time of the peptide in the presence of CS, $t_{free}$ is the migration time of the peptide in the absence of CS, $t_{eofcs}$ is the migration time of the neutral marker in the presence of CS, and $t_{eof}$ is the migration time of the neutral marker in the absence of CS. The dissociation constant ($K_d$) was obtained by fitting the data to the one-site binding model using GraphPad Prism v5.0 (GraphPad Software, San Diego, CA).

*Statistical Analysis*. Spearman's rank correlation coefficient was used to determine the statistical significance of correlation between the fluorescence values and the $K_d$ values between CS and the selected CS-binding peptides as seen in Table 1. The calculation was performed by MATLAB (R2007a, Natick, MA), and the test statistic ($r_s$) was calculated by eq 2:

$$r_s = 1 - \frac{\sum_{i=1}^{n}[R(X_i) - R(Y_i)]^2}{n(n^2-1)} \quad (2)$$

where $n$ is the number of peptides tested, $R(X_i)$ is the rank of peptide due to the fluorescence value, and $R(Y_i)$ is the rank of peptide due to the dissociation constant of the interaction between the peptide and CS. The correlation was deemed statistically significant at $p < 0.05$.

Compositional analysis of the CS-binding peptides was performed following a method described in Thai et al. (*28*). When sorted by fluorescence values after the 2 M NaCl wash, the peptide sequences found in the top 2% were designated as "high affinity" C6S-binding peptides. A particular amino acid or motif was statistically significant when the occurrence in the "high affinity" group of peptides was outside the 95% confidence interval (CI) of the cumulative binomial distribution ($P$) which was derived from the expected number of occurrences due to the initial amino acid distribution. The number of occurrences ($x$) that correspond to the observed number of occurrences ($x_o$) and upper ($x_{upper}$) and lower ($x_{lower}$) limits for the 95% CI were calculated in Excel (Microsoft 2007, Seattle, WA) using the cumulative distribution function shown in eq 3:

$$P(x, N, p) = \sum_{i=0}^{x} \frac{N!}{i!(N-i)!} p^i (1-p)^{N-i} \quad (3)$$

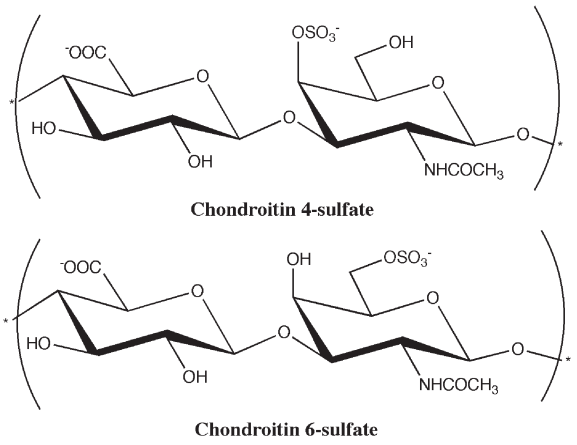


FIGURE 1: Repeating dissacharide units of C4S and C6S.

where $x$ is the number of occurrences of a particular amino acid or motif, $N$ is the number of independent trials, and $p$ is the probability of a particular amino acid ($p_a$) or motif ($p_m$) occurring in any one trial. The probability for the occurrence of the motif ($p_m$) was calculated using eq 4:

$$p_m = \prod_{j=1}^{m} p_a \quad (4)$$

where $m$ is the number of amino acids in the motif of interest and $p_a$ is the probability of the amino acid in the initial distribution. With a 95% CI ($\alpha = 0.05$), the upper and lower critical number of occurrences for the amino acid or motif interest was determined by solving for $x_{lower}$ and $x_{upper}$ when $P_{lower} = 0.025$ and $P_{upper} = 0.0975$. The observed number of occurrences for a particular amino acid or motif ($x_o$) is statistically significant if $P(x_o)$ is less than or equal to $P_{lower}$ and greater than or equal to $P_{upper}$.

The statistical significance of the occurrence of each individual amino acid and several motifs in the top 2% C6S-binding peptides was determined. The testing for statistical significance of one amino acid or motif is considered one hypothesis. Since several hypotheses were tested for statistical significance on a single data set, the probability of data falsely appearing significant is very high. Therefore, a Bonferroni correction was employed to produce a stricter threshold for statistical significance where the $p$-value was divided by the number of hypotheses to discourage the appearance of false positives ($P_{lower} = 0.000862$, $P_{upper} = 0.999138$).

## RESULTS

*Correlation of Fluorescence Values to CS-Binding Affinity*. Fluorescently labeled C6S was screened on the peptide array consisting of 1831 unique 11-mers. The array was washed with buffer that contained 0.4 M NaCl, and the fluorescence of each peptide spot was detected after each wash. Several peptides were chosen from the peptide array and characterized for binding affinity to C6S and C4S (Figure 1) through ACE.

Table 1 shows the peptide sequences, charge, fluorescence values after the initial binding and 0.4 M NaCl wash, and dissociation constants ($K_d$) for the interaction between the peptide and C6S and C4S. Spearman's rank correlation coefficient was used to determine the significance of the relationship between the fluorescence values to the charge and $K_d$ of the peptides to C6S and C4S. For a perfect relationship, $r_s$ would be equal to 1 whereas if $r_s = 0$, there would be no significant

relationship between the two trends. There was no statistically significant ($p > 0.05$) correlation between the fluorescence values and the overall charge content of the CS-binding peptides. However, the charge range was limited (+3 to +5), making it difficult to conclude that charge was not a significant factor affecting C6S-binding affinity.

The correlation of the initial binding fluorescence values and $K_d$ to C6S was not statistically significant ($p > 0.05$), whereas the relationship between the fluorescence values after the 0.4 M NaCl wash and dissociation constants to C6S was statistically significant with a test statistic $r_s = -0.762$ ($p < 0.05$). In the initial binding assay, the CS−peptide interaction may have had a significant contribution from nonspecific electrostatic interactions which led to fluorescence values that did not significantly correlate to C6S-binding affinity. After the salt wash, the nonspecific interactions that occurred in the initial binding were reduced, yielding a fluorescence intensity profile that corresponded with C6S-binding affinity. As a result, peptides with high fluorescence values after the 0.4 M NaCl wash correlate to peptides with higher binding affinities to C6S.

All peptides had statistically significant ($\alpha = 0.05$) higher binding affinities to C6S than to C4S. The correlation between the trends for fluorescence values after the salt wash and $K_d$ between the peptides and C4S was not statistically significant ($p > 0.05$). These results show that the fluorescence values after the NaCl wash corresponded significantly to C6S-binding affinity and not to C4S-binding affinity. This suggests that the technique was selective for peptides that bind to C6S rather than to the closely related C4S.

*Compositional Characterization of C6S-Binding Peptide Sequences*. The Spearman's rank correlation analysis demonstrated that peptides with high fluorescence values after the 0.4 M NaCl wash have high C6S-binding affinity. Therefore, peptides found in the top 2% in fluorescence values after the salt washes were chosen as a representative sample of "high affinity" peptides and compared to the entire peptide population for compositional analysis to determine which amino acids are important for C6S binding. Figure 2 shows the observed number of individual amino acids in the "high affinity" peptides compared to the 95% CI derived from the initial amino acid distribution. If the observed number of amino acids (open circles) falls outside the 95% CI (half-filled boxes), then the occurrence of that amino acid is deemed statistically significant. The compositional analysis showed that the "high affinity" peptides had a significant increase in occurrence of Arg (R), Phe (F), and Tyr (Y) while Lys (K) and Trp (W) significantly decreased in occurrence.

*Point Mutations*. The increased occurrence of certain amino acids in the "high affinity" binding peptides shows that these amino acids may be important for C6S binding. To determine if these amino acids (F, R, and Y) are important for C6S binding, alanine point mutations were made on C6S-2, a "high affinity" peptide. Alanine was subsequently substituted for R, F, and Y throughout the peptide sequence for C6S-2 (YKTNFRRYYRF) as seen in Table 2. The C6S-2 variants were synthesized, purified, and characterized for binding affinity through ACE. The $K_d$ of the C6S-2 alanine variants were significantly ($\alpha = 0.05$) higher than the original sequence (C6S-2). The F and Y substitution decreased the binding affinity moderately, whereas the R substitution drastically decreased the binding affinity of the C6S-binding peptide. Therefore, the alanine point mutation results show that F, R, and Y are all important for C6S binding with R possibly having a greater contribution to C6S-binding affinity.

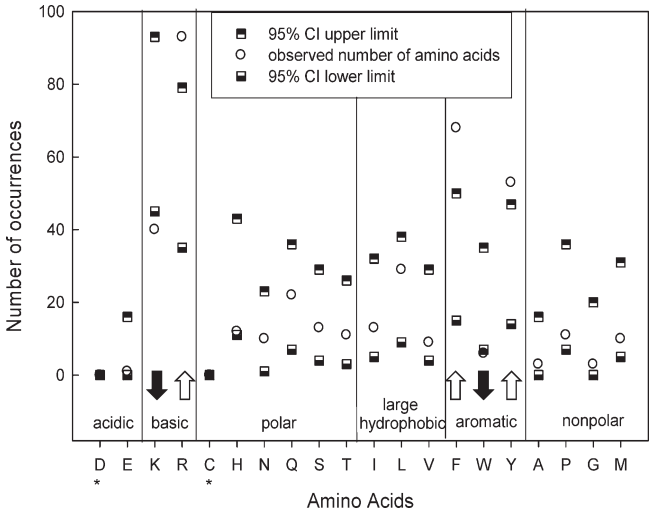


FIGURE 2: Compositional analysis of C6S-binding peptide sequences. The occurrence of each individual amino acid in the "high affinity" peptide sequences was compared to the occurrence of each amino acid in the initial amino acid distribution. The open circles represent the observed number of amino acids in the top 2% C6S-binding peptides. The 95% CI is represented by the half-filled boxes: upper limit (upper filled) and lower limit (lower filled). The arrows represent amino acids that have statistically significant occurrences (white arrows, increase; dark arrows, decrease). *, $p_a = 0$ for C and D in initial amino acid distribution.

Table 2: C6S-Binding Affinity of C6S-2 Alanine Variants

| peptide | sequence | C6S $K_d$ ($\mu$M) |
|---|---|---|
| C6S-2 | YKTNFRRYYRF | $2.31 \pm 0.10$ |
| C6S-2-AF | YKTNARRYYRA | $96.23 \pm 3.62$ |
| CS6-2-AR | YKTNFAAYYAF | $707.5 \pm 116.1$ |
| C6S-2-AY | AKTNFRRAARF | $96.72 \pm 3.44$ |

Table 3: C6S-Binding Affinity of Trp (W → F or Y) or Lys (K → R) Point Mutations

| peptide | sequence | C6S $K_d$ ($\mu$M) |
|---|---|---|
| C6S-1 | EKRIWFPYRRF | $1.14 \pm 0.13$ |
| C6S-1-WF | EKRIFFPYRRF | $14.08 \pm 1.14$ |
| C6S-1-WY | EKRIYFPYRRF | $15.82 \pm 1.94$ |
| C6S-1-KR | ERRIWFPYRRF | $0.45 \pm 0.02$ |
| C6S-1-7 | KNYKFRFGTRV | $63.68 \pm 2.39$ |
| C6S-7-KR | RNYRFRFGTRV | $36.34 \pm 1.91$ |

The compositional analysis and alanine point mutation results have shown that F, R, and Y are all important amino acids for C6S binding. With this knowledge, the C6S-binding peptide sequences can be manipulated to create peptides with higher binding affinity to C6S. Several peptide array variants were synthesized and characterized for binding affinity to C6S through ACE. Table 3 shows the original peptide sequences, the change in sequence (W → F or Y, K → R), and the corresponding $K_d$ for each peptide to C6S. The substitution of R for K improved the C6S-binding affinity for both peptides tested (CS6-1 and CS6-7). However, the change of W to F or Y (CS6-1) did not improve C6S binding and actually decreased binding affinity. It is possible that W is an important amino acid within the CS6-1 peptide sequence and the removal caused a decrease in C6S-binding affinity. On the other hand, the ability of F and Y to increase C6S-binding affinity may be influenced by other factors such as

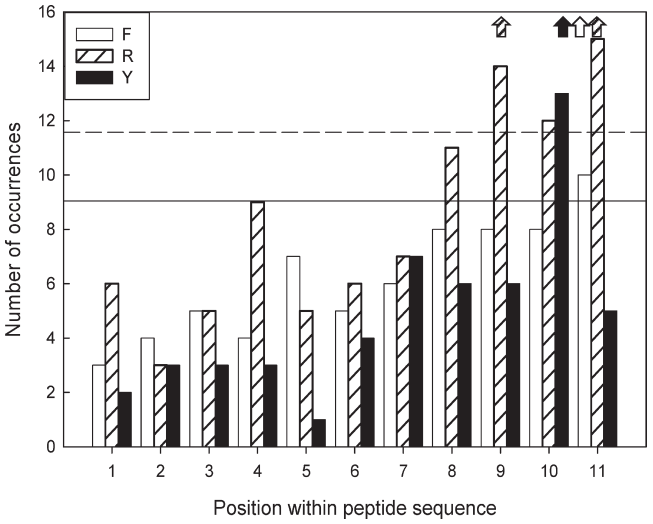


FIGURE 3: Positional analysis of increased occurrence of F, R, and Y in C6S-binding peptides. The white (F), striped (R), and black (Y) bars represent the number of occurrences of each amino acid in the "high affinity" C6S-binding peptides at each position within the peptide sequence. The solid (F, Y) and dashed (R) horizontal lines represent the upper limits for the 95% CI. The arrows (white, F; shaded, R; black, Y) indicate positions where the number of occurrences for each amino acid in the "high affinity" C6S-binding peptides was statistically significant where the occurrence for each amino acid was greater than the 95% CI (horizontal line) for each position.

positional dependency of the incorporated amino acid both within the peptide sequence and in relation to neighboring amino acids. From a purely compositional perspective, the incorporation of R for K into the peptide sequence can increase C6S-binding affinity, whereas the substitution of F and Y for W does not increase C6S-binding affinity in the CS6-1 peptide sequence.

*Position Dependency Analysis*. The previous results have shown that while R, F, and Y all show an increased occurrence in C6S-binding peptides, further increasing their presence can either increase (R) or decrease (F and Y) C6S-binding affinity. To elucidate this phenomenon, further analysis was pursued to determine if the increased occurrence of these amino acids was position dependent both within the peptide sequence and in relation to neighboring amino acids. Figure 3 shows the positional analysis of the increased occurrence of each amino acid (F, R, and Y) in the top 2% of C6S-binding peptides compared to the initial amino acid distribution for each position within the peptide sequence. The increased occurrence of F and Y was statistically significant at positions 11 and 10, respectively, while R was statistically significant at positions 9 and 11. These results demonstrate that the increased occurrence of Y and F was relatively specific for position while the increased occurrence of R was not as specific. Furthermore, the increased occurrence of F, R, and Y in the top 2% of C6S-binding peptides is clustered together, demonstrating that the effects of the individual amino acids might have a cooperative effect in increasing C6S-binding affinity.

*Motif Analysis*. The clustering of amino acids that contribute to C6S binding in the positional analysis was evidence for the presence of a recurring motif within the top 2% of C6S-binding peptides. Previous studies that have identified GAG-binding motifs mainly focus on the position of the basic amino acids in relation to other amino acids. In this study, the motif analysis (Table 4) evaluated the distribution of basic (B), aromatic (A), and nonbasic (X) amino acids in 5-mers found in the top 2%. Three motifs with 1B (dark gray) had a statistically significant high occurrence in the "high affinity" peptide population. Since the B was not position specific within these motifs (BAAAA, AABAA, AAAAB), these motifs lost their relevance as a specific C6S-binding motif. In the other overrepresented motif (ABBAA), the position of the BB was evaluated in relation to both aromatic and nonbasic amino acids. The majority of the motifs with the BB did not have a significant presence in the top 2%, and two motifs (light gray) actually had significantly lower occurrences in the "high affinity" peptides, suggesting that these motifs (XBXBB, XBBBX) may be unfavorable for C6S-binding affinity. These results indicate that the motif ABBAA may be an important C6S-binding motif; however, validation of the importance of the identified motif requires a significantly larger peptide population than sampled in the present study.

Table 4: Motif Analysis[a]

| 1B | A: aromatic | X: non-basic |
|---|---|---|
| | BAAAA | BXXXX |
| | ABAAA | XBXXX |
| | AABAA | XXBXX |
| | AAABA | XXXBX |
| | AAAAB | XXXXB |
| **2B** | | |
| *B + B* | BAAAB | BXXXB |
| | ABAAB | XBXXB |
| | AABAB | XXBXB |
| | ABABA | XBXBX |
| | BAABA | BXXBX |
| | BABAA | BXBXX |
| *BB* | *ABBAA* | XBBXX |
| | AABBA | XXBBX |
| | AAABB | XXXBB |
| | BBAAA | BBXXX |

| 3B | A: aromatic | X:non-basic |
|---|---|---|
| *B + B + B* | BABAB | BXBXB |
| *BB + B* | BBAAB | BBXXB |
| | ABBAB | XBBXB |
| | BAABB | BXXBB |
| | BBABA | BBXBX |
| | ABABB | XBXBB |
| | BABBA | BXBBX |
| *BBB* | BBBAA | BBBXX |
| | ABBBA | XBBBX |
| | AABBB | XXBBB |
| **4B** | | |
| *BB + BB* | BBABB | BBXBB |
| *BBB + B* | BABBB | BXBBB |
| | BBBAB | BBBXB |
| *BBBB* | BBBBA | BBBBX |
| | ABBBB | XBBBB |

[a](B) basic, (A) aromatic, and (X) nonbasic amino acids in 5-mer motifs were evaluated for statistical significance ($\alpha = 0.05$). Shaded boxes represent motifs with a significantly high (dark gray) or low (light gray) occurrence in the "high affinity" binding population when compared to the initial amino acid distribution.

## DISCUSSION

In this study, novel C6S-binding peptides were identified through peptide array screening and characterized for binding affinity through ACE. Compositional analysis was performed to determine which amino acids are important for high binding affinity to C6S. For the analysis, it was necessary to first identify a population of peptides designated as "high binding" affinity to C6S. The correlation analysis showed that the fluorescence values after the 0.4 M NaCl wash from the peptide array screening corresponded to the $K_d$ values for the interaction of the peptides to C6S. Therefore, peptide spots with high fluorescence values corresponded to peptides with high C6S-binding affinity. The top 2% of the array peptides when sorted according to fluorescence intensity values after 0.4 M NaCl were designated as "high affinity." This sample size (37 peptide sequences) was large enough for statistical analysis yet small enough so that the fluorescence values were considerably different than the fluorescence values of the entire peptide array (see Supporting Information).

The peptide array consisted of 1831 peptides which represents a small fraction of possible peptide sequences for 11-mers ($20^{11}$). To increase the probability of identifying CS-binding peptides with high binding affinity, the peptide library initial amino acid

distribution was biased toward amino acids commonly found in GAG-binding peptides/proteins. C6S-binding peptides of varying affinity ($1.14 \pm 0.13$ $\mu$M to $63.68 \pm 2.39$ $\mu$M) were identified. The "high binding" affinity peptides ($K_d$ 1−10 $\mu$M) have comparable binding affinity to designed heparin-binding peptides (3.85−6 $\mu$M) (*25*). However, the peptide sequences identified from heparin-binding proteins, such as antithrombin III, have much higher binding affinity (10−100 nM) (*6*). The goal of this study was to take the first step in identifying CS-binding peptides. Further analysis beyond compositional analysis combined with more comprehensive screening techniques such as phage display may yield CS-binding peptide sequences with higher binding affinity.

Compositional analysis was performed by comparing the amino acid distribution of the "high affinity" peptide sequences to the initial amino acid distribution. Results showed that certain amino acids had a statistically significant increase in occurrence (F, R, Y) while other amino acids had a significant decrease in occurrence (K, W). Alanine point mutations were made within the peptide C6S-2 to assess the contribution of F, R, and Y to C6S-binding affinity. The C6S-2 alanine variants all had higher $K_d$ values to C6S than the original C6S-2 peptide sequence. Thus, the amino acids F, R, and Y are all important for the interaction of the C6S-2 peptide to C6S with R possibly have a greater contribution to high C6S-binding affinity.

To date, this study is the first to identify the amino acids that contribute to the binding interaction of proteins/peptides to C6S. Several studies have shown that basic amino acids are important for GAG-binding proteins (*25*, *29*). However, few studies have focused on the importance of aromatic amino acids for GAG-binding proteins (*30*). Bae et al. have postulated that Y may play an important role in the heparin−antithrombin III interaction (*26*). Since F is similar in structure to Y, it is possible that the increased occurrence of F is due to the high binding affinity of C6S to the benzyl side chain. On the other hand, the indole group of W has a bulkier structure with two rings which can lead to steric hindrance and lower C6S-binding affinity.

The majority of GAG−protein structure−function studies have focused on heparin which is a highly sulfated polysaccharide (*8*, *31*−*33*). Since CS has fewer sulfation groups, which results in unsulfated regions along the chain, it is possible that the aromatic side chains on the peptide interact through hydrophobic forces to the unsulfated stretches of CS. Therefore, hydrophobic forces may be more significant for C6S binding than for heparin binding.

It is important to consider what type of interactions contributes to GAG−protein interactions when characterizing protein binding affinity. Affinity liquid chromatography is the most commonly used technique for studying GAG-binding affinity. In this assay, the GAG is immobilized on a solid matrix, and the protein is applied to the column. The column is subjected to an increasing salt concentration which competes with the electrostatic interactions to elute the protein from the GAG column. This approach fails to take into account the hydrophobic and hydrogen-bonding contributions to the GAG−protein interaction. Preliminary assays with C6S-modified beads showed that the CS-binding peptides could not be removed from the column with a high salt concentration up to 2 M NaCl (data not shown). Since the results from this study have shown that hydrophobic forces contribute to C6S-binding affinity, future studies must use a more holistic approach, such as gold nanoparticles or affinity capillary electrophoresis, which takes into account all physical interactions to determine CS-binding affinity (*34*).

The compositional analysis results were utilized to design peptides with higher binding affinity to C6S. The conversion of K to R in the peptides C6S-1 and C6S-7 improved C6S-binding affinity. Previous studies have shown that heparin-binding peptide sequences that contain R have higher heparin-binding affinity than sequences with K (*35*). The greater affinity of R for GAGs may be due to the formation of more hydrogen bonds with the guanidino group of R than the ammonium cation of K with the GAG sulfate groups. Also, the guanidino cation may form an inherently stronger electrostatic interaction with the sulfate anion (*35*). For the basic amino acids, CS has similar binding characteristics to heparin with stronger affinity to R than K.

The substitution of F or Y for W in the C6S-1 peptide did not improve C6S-binding affinity. Further analysis showed that the contribution of the amino acids F and Y was fairly position specific for high binding affinity to C6S and is only significant when found adjacent to other critical amino acids (F, R, and Y). Since this study only investigates the absolute position of certain amino acids (F, R, and Y), further studies must be performed in order to understand the contribution of amino acids relative to one another in any position within the peptide sequence. It is also possible that W is an important amino acid for C6S binding in the C6S-1 peptide; therefore, the substitution of any amino acid for W at this position would lead to lower C6S-binding affinity. This study is one of only a few to have characterized the contributions of aromatic amino acids to GAG binding. Additional studies must be performed to understand the role of specific aromatic amino acids in C6S binding.

The motif analysis further confirmed the importance of aromatic amino acids for high binding affinity to C6S. The motif ABBAA, where A represents any aromatic amino acid, had a statistically significant increase in occurrence when compared to the motif XBBXX, where X represents any nonbasic amino acid. The position of BB in relation to the aromatic amino acids was also significant, suggesting an important role for the ABBAA motif. Since there are few known C6S-binding proteins, this motif may be utilized in the identification of novel C6S-binding proteins. Shen et al. have recently identified the first CSPG receptor involved in neural regeneration following spinal cord injury (*36*). It is possible that, with more sequence information derived from native CS-binding proteins, peptide array screening may be used to identify other CSPG receptors.

This study has identified C6S-binding peptides and characterized the C6S−peptide interaction through peptide array screening, affinity capillary electrophoresis, and statistical analysis. The compositional analysis showed that certain aromatic and basic amino acids are favored for high binding affinity to C6S which led to the design of peptides with higher C6S-binding affinity. The characterization of the mutated peptides showed that compositional modifications for increased binding affinity can be accomplished more generally with basic amino acids but may be limited with aromatic amino acids. Finally, a potential C6S-binding motif (ABBAA) was identified. The understanding of which amino acids and motifs play an important role in C6S binding will lead to a better understanding of key chemical interactions involved in C6S binding and the development of improved C6S-binding peptides for diagnostic and therapeutic applications.

## SUPPORTING INFORMATION AVAILABLE

Peptide sequences and their fluorescence intensities both after the initial screening and after the 0.4 M NaCl wash. This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES


1. Timar, J., Ladanyi, A., Lapis, K., and Moczar, M. (1992) Differential expression of proteoglycans on the surface of human melanoma cells characterized by altered experimental metastatic potential. *Am. J. Pathol. 141*, 467–474.
2. Nikitovic, D., Assouti, M., Sifaki, M., Katonis, P., Krasagakis, K., Karamanos, N. K., and Tzanakakis, G. N. (2008) Chondroitin sulfate and heparan sulfate-containing proteoglycans are both partners and targets of basic fibroblast growth factor-mediated proliferation in human metastatic melanoma cell lines. *Int. J. Biochem. Cell Biol. 40*, 72–83.
3. Properzi, F., Carulli, D., Asher, R. A., Muir, E., Camargo, L. M., van Kuppevelt, T. H., ten Dam, G. B., Furukawa, Y., Mikami, T., Sugahara, K., Toida, T., Geller, H. M., and Fawcett, J. W. (2005) Chondroitin 6-sulphate synthesis is up-regulated in injured CNS, induced by injury-related cytokines and enhanced in axon-growth inhibitory glia. *Eur. J. Neurosci. 21*, 378–390.
4. Gilbert, R. J., McKeon, R. J., Darr, A., Calabro, A., Hascall, V. C., and Bellamkonda, R. V. (2005) CS-4,6 is differentially upregulated in glial scar and is a potent inhibitor of neurite extension. *Mol. Cell. Neurosci. 29*, 545–558.
5. Bradbury, E. J., Moon, L. D., Popat, R. J., King, V. R., Bennett, G. S., Patel, P. N., Fawcett, J. W., and McMahon, S. B. (2002) Chondroitinase ABC promotes functional recovery after spinal cord injury. *Nature 416*, 636–640.
6. Capila, I., and Linhardt, R. J. (2002) Heparin-protein interactions. *Angew. Chem., Int. Ed. Engl. 41*, 391–412.
7. Gama, C. I., Tully, S. E., Sotogaku, N., Clark, P. M., Rawat, M., Vaidehi, N., Goddard, W. A., III, Nishi, A., and Hsieh-Wilson, L. C. (2006) Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat. Chem. Biol. 2*, 467–473.
8. Raman, R., Sasisekharan, V., and Sasisekharan, R. (2005) Structural insights into biological roles of protein-glycosaminoglycan interactions. *Chem. Biol. 12*, 267–277.
9. Bingley, J. A., Hayward, I. P., Campbell, G. R., and Campbell, J. H. (2001) Relationship of glycosaminoglycan and matrix changes to vascular smooth muscle cell phenotype modulation in rabbit arteries after acute injury. *J. Vasc. Surg. 33*, 155–164.
10. Theocharis, A. D., Theocharis, D. A., De Luca, G., Hjerpe, A., and Karamanos, N. K. (2002) Compositional and structural alterations of chondroitin and dermatan sulfates during the progression of atherosclerosis and aneurysmal dilatation of the human abdominal aorta. *Biochimie 84*, 667–674.
11. Pins, G. D., Christiansen, D. L., Patel, R., and Silver, F. H. (1997) Self-assembly of collagen fibers. Influence of fibrillar alignment and decorin on mechanical properties. *Biophys. J. 73*, 2164–2172.
12. Keiser, N., Venkataraman, G., Shriver, Z., and Sasisekharan, R. (2001) Direct isolation and sequencing of specific protein-binding glycosaminoglycans. *Nat. Med. 7*, 123–128.
13. Tully, S. E., Rawat, M., and Hsieh-Wilson, L. C. (2006) Discovery of a TNF-alpha antagonist using chondroitin sulfate microarrays. *J. Am. Chem. Soc. 128*, 7740–7741.
14. Sorrell, J. M., Carrino, D. A., Baber, M. A., Asselineau, D., and Caplan, A. I. (1999) A monoclonal antibody which recognizes a glycosaminoglycan epitope in both dermatan sulfate and chondroitin sulfate proteoglycans of human skin. *Histochem. J. 31*, 549–558.
15. Resende, M., Nielsen, M. A., Dahlback, M., Ditlev, S. B., Andersen, P., Sander, A. F., Ndam, N. T., Theander, T. G., and Salanti, A. (2008) Identification of glycosaminoglycan binding regions in the *Plasmodium falciparum* encoded placental sequestration ligand, VAR2CSA. *Malar. J. 7*, 104.
16. Pothacharoen, P., Kalayanamitra, K., Deepa, S. S., Fukui, S., Hattori, T., Fukushima, N., Hardingham, T., Kongtawelert, P., and Sugahara, K. (2007) Two related but distinct chondroitin sulfate mimetope octasaccharide sequences recognized by monoclonal antibody WF6. *J. Biol. Chem. 282*, 35232–35246.
17. Purushothaman, A., Fukuda, J., Mizumoto, S., ten Dam, G. B., van Kuppevelt, T. H., Kitagawa, H., Mikami, T., and Sugahara, K. (2007) Functions of chondroitin sulfate/dermatan sulfate chains in brain development. Critical roles of E and iE disaccharide units recognized by a single chain antibody GD3G7. *J. Biol. Chem. 282*, 19442–19452.
18. ten Dam, G. B., van de Westerlo, E. M., Purushothaman, A., Stan, R. V., Bulten, J., Sweep, F. C., Massuger, L. F., Sugahara, K., and van Kuppevelt, T. H. (2007) Antibody GD3G7 selected against embryonic glycosaminoglycans defines chondroitin sulfate-E domains highly up-regulated in ovarian cancer and involved in vascular endothelial growth factor binding. *Am. J. Pathol. 171*, 1324–1333.
19. Smetsers, T. F., van de Westerlo, E. M., ten Dam, G. B., Clarijs, R., Versteeg, E. M., van Geloof, W. L., Veerkamp, J. H., van Muijen, G. N., and van Kuppevelt, T. H. (2003) Localization and characterization of melanoma-associated glycosaminoglycans: differential expression of chondroitin and heparan sulfate epitopes in melanoma. *Cancer Res. 63*, 2965–2970.
20. Smetsers, T. F., van de Westerlo, E. M., ten Dam, G. B., Overes, I. M., Schalkwijk, J., van Muijen, G. N., and van Kuppevelt, T. H. (2004) Human single-chain antibodies reactive with native chondroitin sulfate detect chondroitin sulfate alterations in melanoma and psoriasis. *J. Invest. Dermatol. 122*, 707–716.
21. Gao, X., Pellois, J. P., Na, Y., Kim, Y., Gulari, E., and Zhou, X. (2004) High density peptide microarrays. In situ synthesis and applications. *Mol. Divers. 8*, 177–187.
22. Frank, R. (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications. *J. Immunol. Methods 267*, 13–26.
23. Kramer, A., and Schneider-Mergener, J. (1998) Synthesis and screening of peptide libraries on continuous cellulose membrane supports. *Methods Mol. Biol. 87*, 25–39.
24. Margalit, H., Fischer, N., and Bensasson, S. A. (1993) Comparative analysis of structurally defined heparin-binding sequences reveals a distinct spatial distribution of basic residues. *J. Biol. Chem. 268*, 19228–19231.
25. Fromm, J. R., Hileman, R. E., Caldwell, E. E., Weiler, J. M., and Linhardt, R. J. (1997) Pattern and spacing of basic amino acids in heparin binding sites. *Arch. Biochem. Biophys. 343*, 92–100.
26. Bae, J., Desai, U. R., Pervin, A., Caldwell, E. E., Weiler, J. M., and Linhardt, R. J. (1994) Interaction of heparin with synthetic antithrombin III peptide analogues. *Biochem. J. 301* (Part 1), 121–129.
27. McKeon, J., and Holland, L. A. (2004) Determination of dissociation constants for a heparin-binding domain of amyloid precursor protein and heparins or heparan sulfate by affinity capillary electrophoresis. *Electrophoresis 25*, 1243–1248.
28. Thai, C. K., Dai, H., Sastry, M. S., Sarikaya, M., Schwartz, D. T., and Baneyx, F. (2004) Identification and characterization of Cu(2)O- and ZnO-binding polypeptides by *Escherichia coli* cell surface display: toward an understanding of metal oxide binding. *Biotechnol. Bioeng. 87*, 129–137.
29. Cardin, A. D., and Weintraub, H. J. (1989) Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis 9*, 21–32.
30. Caldwell, E. E., Nadkarni, V. D., Fromm, J. R., Linhardt, R. J., and Weiler, J. M. (1996) Importance of specific amino acids in protein binding sites for heparin and heparan sulfate. *Int. J. Biochem. Cell Biol. 28*, 203–216.
31. Tyler-Cross, R., Sobel, M., McAdory, L. E., and Harris, R. B. (1996) Structure-function relations of antithrombin III-heparin interactions as assessed by biophysical and biological assays and molecular modeling of peptide-pentasaccharide-docked complexes. *Arch. Biochem. Biophys. 334*, 206–213.
32. Sasisekharan, R., Raman, R., and Prabhakar, V. (2006) Glycomics approach to structure-function relationships of glycosaminoglycans. *Annu. Rev. Biomed. Eng. 8*, 181–231.
33. Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998) Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *BioEssays 20*, 156–167.
34. Jeong, K. J., Butterfield, K., and Panitch, A. (2008) A novel assay to probe heparin-peptide interactions using pentapeptide-stabilized gold nanoparticles. *Langmuir 24*, 8794–8800.
35. Fromm, J. R., Hileman, R. E., Caldwell, E. E. O., Weiler, J. M., and Linhardt, R. J. (1995) Differences in the interaction of heparin with arginine and lysine and the importance of these basic-amino-acids in the binding of heparin to acidic fibroblast growth factor. *Arch. Biochem. Biophys. 323*, 279–287.
36. Shen, Y., Tenney, A. P., Busch, S. A., Horn, K. P., Cuascut, F. X., Liu, K., He, Z., Silver, J., and Flanagan, J. G. (2009) PTPsigma is a receptor for chondroitin sulfate proteoglycan, an inhibitor of neural regeneration. *Science 326*, 592–596.