Identification and Characterization of Cu$_2$O- and ZnO-Binding Polypeptides by Escherichia coli Cell Surface Display: Toward an Understanding of Metal Oxide Binding

Corrine K. Thai, Haixia Dai, M. S. R. Sastry, Mehmet Sarikaya, Daniel T. Schwartz, François Baneyx

1Department of Chemical Engineering, University of Washington, Seattle, Washington 98195; telephone: 206-685-3451; fax: 206-685-3451; e-mail: baneyx@u.washington.edu
2Materials Sciences and Engineering, University of Washington, Seattle, Washington 98195

Received 24 November 2003; accepted 23 March 2004
Published online 16 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/bit.20149

Abstract: We have used the FliTrx cell surface display system to identify disulfide-constrained dodecapeptides binding to the semiconducting metal oxides Cu$_2$O and ZnO. Sequence analysis of the inserts revealed that the two populations exhibit similar, yet subtly different patterns of amino acid usage. Both sets of binders were enriched in arginine, tryptophan, and glycine with a higher degree of positional preference in the case of Cu$_2$O binders. Tyrosine, proline, and serine were underrepresented in both populations. Peptides binding electrodeposited Cu$_2$O or ZnO with high avidity could be subdivided into two classes based on pI and hydrophilicity. In the hydrophilic and positively charged Class I binders, the Arg–X–X–Arg tetrapeptide appears to be implicated in metal oxide binding, whereas Arg–Arg and Arg–Lys pairs allow for discrimination between Cu$_2$O and ZnO. Molecular dynamics simulations of the disulfide-constrained peptides suggest that the aforementioned motifs are important to properly orient two basic residues that are likely to contact the metal oxides. The implications of our results in understanding the rules governing the interaction between peptides and inorganic compounds and in their use for the design of hybrid nanoarchitectures are discussed. © 2004 Wiley Periodicals, Inc.

Keywords: display; FliTrx; inorganic binding; molecular biomimetics

INTRODUCTION

The observation that mineralizing organisms make use of macromolecules such as proteins and lipids to control the nucleation, assembly, and architecture of inorganic phases has opened the door to a new paradigm in nanomaterials design, which has been called molecular biomimetics or biomimetic nanotechnology (Sarikaya, 1999; Sarikaya et al., 2003; Seeman and Belcher, 2002). The central premise of this emerging field is that inorganic-binding peptides, either in isolation or when inserted within the structural framework of proteins displaying useful characteristics (e.g., self-assembly or ability to bind other macromolecules such as DNA), can be used as molecular erector sets to direct the assembly of hybrid materials with control of composition and topology (Sarikaya et al., 2003). Because naturally occurring biomineralization proteins may only be useful to regenerate the inorganic material that they are associated with, most efforts have been directed at identifying small polypeptides that bind with high affinity to inorganic materials of engineering interest using cell surface or phage display technologies. In typical experiments, a library of random amino acid sequences, usually varying in size between 7 and 15 residues, and either unconstrained or constrained by the formation of a disulfide bond between flanking and invariant cysteine residues, is displayed within a cell surface protein of E. coli or a bacteriophage coat protein (Azzazy and Highsmith, 2002; Dani, 2001; Samuelson et al., 2002). Cells or viruses are contacted with the inorganic material and binding sequences of interest are enriched by multiple rounds of biopanning. To date, cell surface display has been used to identify peptides that recognize iron oxide (Brown, 1992), gold (Brown, 1997), and zinc oxide (Kjærgaard et al., 2000), whereas phage display has been used to isolate peptides binding to gallium arsenide (Whaley et al., 2000), silica (Naik et al., 2002a), silver (Naik et al., 2002b), zinc sulfide (Lee et al., 2002), calcite (Li et al., 2002), and cadmium sulfide (Mao et al., 2003). Although growing, the available “toolbox” of inorganic binding sequences remains small and has been
obtained using a limited number of display technologies (chiefly the M13 Ph.D. phage display system commercialized by New England Biolabs, Beverly, MA). Furthermore, little is known about the “rules” governing the interaction between short polypeptides and inorganic compounds.

The FliTrx cell surface display library (Lu et al., 1995) positions random sequences of 12 amino acids as disulfide-constrained loops within Thioredoxin 1 (Trx), which is itself inserted into FlIC, the major E. coli flagellar protein. The resulting fusion proteins are exported to the cell surface where they assemble into partially functional flagella. The FliTrx library has an estimated diversity of $1.77 \times 10^6$, which is higher than that of previously used cell surface display systems (Brown, 1997; Kjærgaard et al., 2000), contains no predefined structural motif, and has not been exploited for the selection of inorganic-binding sequences. Here, we report on the use of this library for the identification of Cu$_2$O- and ZnO-binding peptides, and employ a combination of statistical analysis, microfabrication, fluorescence microscopy, and molecular dynamics simulations to gain insight into the nature of the rules governing the interaction between the selected peptides and metal oxides.

**MATERIALS AND METHODS**

**Media**

M9 salts (10×; pH 7.4) were prepared by dissolving 60 g of Na$_2$HPO$_4$, 30 g of KH$_2$PO$_4$, 5 g of NaCl, and 10 g of NH$_4$Cl in 1 L of deionized water. IMC medium is M9 supplemented with 0.2% casamino acids, 0.5% glucose, and 1 mM MgCl$_2$. RM medium is M9 supplemented with 2% casamino acids, 1% glycerol, and 1 mM MgCl$_2$.

**Materials**

A potentiosstat/galvanostat (Model 263, Princeton Applied Research, Princeton, NJ) was used in all electrochemical experiments. Copper oxide substrates were synthesized electrochemically as previously described (Dai et al., 2004). Briefly, a copper foil working electrode (10 × 10 × 0.127 mm; 99.99% purity; Alfa Aesar) was immersed in 1 M KOH. Cu$_2$O and Cu(OH)$_2$ films were anodically grown on the foil by control of the sweep potential range during cyclic voltammetry. Cu$_2$O films were formed by sweeping the potential from −1000 to −450 mV, whereas Cu(OH)$_2$ films were formed in the −400- to −125-mV potential range (against SCE). For ZnO substrates, a gold-coated glass working electrode was immersed in 0.1 M Zn(NO$_3$)$_2$ with air bubbling for 2 h and a piece of Zn sheet was used as the counter-electrode. ZnO was cathodically deposited for 30 min on the gold surface at −800 mV and 60°C. All deposited films were thoroughly rinsed with deionized water and air dried. Composition was verified by Raman spectroscopy and/or monochromatic XPS. Surfaces of indicated dimensions were manually cut from the slides.

Five ZnO single crystals (10 × 10 × 0.5 mm) were purchased from Marketech International (Port Townsend, WA) and used as provided. Patterned ZnO/Cu$_2$O/Au substrates were made by modification of previously described protocols (Wang et al., 2001). A first tape mask was laid to protect stripes of Au and a ZnO thin film was deposited through the mask openings. A second tape mask was laid to protect ZnO stripes and the substrate was dipped into 1 M H$_2$SO$_4$ to dissolve the accessible ZnO and reveal the underlying Au. The surface was cleaned with deionized water and placed into a Cu$_2$O plating bath (0.4 M CuSO$_4$, 3 M lactic acid, 5 M NaOH [pH 9.0]). The counter-electrode was a piece of Cu sheet. Cu$_2$O was cathodically deposited onto the Au through the mask openings for 1 h at −350 mV and 60°C. After electrodeposition, the masks were peeled off and the patterned metal surface was rinsed with acetone, methanol, and deionized water, and then air dried. The compositional purity of patterned substrates was verified by Raman spectroscopy.

**Selection of Inorganic-Binding Peptides**

One vial of the FliTrx library (Invitrogen, Carlsbad, CA) was used to inoculate 50 mL of IMC supplemented with 50 µg/mL carbenicillin in a sterile 250-mL shake flask. Cells were grown for 15 to 18 h at 25°C with shaking (250 rpm) to saturation ($A_{600} ≈ 3$). Aliquots corresponding to 10$^{10}$ cells (assuming 10$^9$ cells per milliliter per absorbance unit at 600 nm) were transferred to 50 mL of IMC medium supplemented with carbenicillin and 100 µg/mL L-tryptophan to induce expression of the modified flagellar protein. Cells were grown for 5 or 6 h as just described. Cu$_2$O and ZnO surfaces (10 × 10 mm) were positioned in the center of 4-cm-diameter polystyrene culture dishes (Corning), washed with 10 mL of sterile deionized water for 2 to 3 min, and incubated with gentle shaking with 10 mL of freshly prepared blocking solution (100 mg dry milk, 300 µL 5 M NaCl, 500 µL 20% methyl α-D-mannopyranoside, 9.2 mL IMC, and 10 µL of 50 mg/mL carbenicillin) for 5 min at 25°C. After decanting the liquid, 10 mL of induced cultures supplemented with 100 mg dry milk, 300 µL of 5 M sterile NaCl, and 500 µL of sterile 20% methyl α-D-mannopyranoside were added to the dish. Cells were allowed to contact the inorganic surface at 25°C for 1 min with agitation at 50 rpm and without agitation for 45 min. At the end of the incubation period, the surface was transferred to a new culture dish containing 10 mL of freshly prepared wash solution (2.5 mL 20% methyl α-D-mannopyranoside, 47.5 mL IMC, 50 µL of 50 mg/mL carbenicillin) using sterile tweezers. The dish was agitated at room temperature and 50 rpm for 5 min. The surface was transferred to a fresh dish as described earlier and the wash operation was repeated a total of five times. To recover binders, the surface was moved to a sterile culture dish containing 10 mL of IMC supplemented with carbenicillin and the plate was vortexed for 30 s to shear off the flagella. The solution was recovered by pipetting,
transferred to a sterile 125-mL shake flask, and cells were allowed to grow for 15 to 18 h at 25°C. This completed a round of panning. The entire panning process was repeated five times to enrich for tight binders. Dilutions from the fifth panning round were plated onto RMG plates (M9 medium, 2% casamino acids, 0.5% glucose, 1 mM MgCl₂, 1.5% agar) containing carbenicillin and incubated overnight at 30°C.

**Sequencing**

Randomly selected colonies were used to inoculate 5 mL of RM medium supplemented with carbenicillin and cultured for 16 to 24 h at 30°C. Plasmid DNA was recovered using the QIAprep Spin Miniprep Kit (Qiagen) and the nucleotide sequences of the inserts were determined using the Perkin-Elmer Big Dye PCR sequencing kit and either 5’-ATTGCACCTGACTGACGAC-3’ as a forward primer or 5’-CCCTGATATTCGTCAGCG-3’ as a reverse primer.

**Statistical Analysis**

Statistical significance testing was used to compare the observed number of occurrences of individual amino acids and/or motifs in the sequenced clones (denoted \( o \)) to their expected number of occurrences in the naive FliTrx population (denoted \( e \)), for a given number of independent trials (denoted \( N \)). Because the presence or absence of a particular residue or motif is dichotomous, mutually exclusive, and independent, the probability of obtaining a particular outcome is governed by the binomial distribution. Here, we are mainly concerned with the cumulative binomial probability function:

\[
P(o, N, p) = \sum_{i=o}^{N} \frac{N!}{i!(N-i)!} p^i (1-p)^{N-i}
\]

(1)

where \( p \) is the probability of a particular amino acid or motif occurring in any one trial (Box et al., 1978). The number of independent trials is dictated by the number of random amino acids in the displayed polypeptide (denoted \( \ell \)), the number of amino acids in the motif of interest (denoted \( m \)), and the number of sequenced clones (denoted \( c \)) according to the relationship:

\[
N = (\ell - (m - 1)) \times c
\]

(2)

For all the experiments described here, \( \ell = 12 \), but the number of sequenced clones differs depending on the inorganic substrate, with \( c = 31 \) for Cu₂O and \( c = 33 \) for ZnO binders. The motif length is \( m = 1 \) for single amino acids, \( m = 2 \) for paired amino acids (we evaluated the RR, KK, RK, and KR pairs), and \( m = 4 \) for the motifs RXRX, RXRR, where X denotes any amino acid. Of course, the constraint \( m \leq \ell \) must hold.

Eqs. (1) and (2) can be used to set 95% confidence limits for the observed number of occurrences of a particular residue or motif (\( o \)). The procedure involves choosing the residue or motif of interest to determine \( m \), as well as the specific inorganic substrate used (so \( c \) is known), thereby allowing \( N \) to be determined. The lower 95% confidence limit is found by setting \( p = 0.025 \) and solving for \( p \) in Eq. (1). This value of \( p \) is used to calculate the lower limit (denoted \( o_\text{L} \)) via the equation: \( o_\text{L} = p \cdot N \). The upper 95% confidence limit is found by setting \( p = 0.975 \) and solving Eq. (1) to find a new value of \( p \). This new \( p \) is used to calculate the upper limit (denoted \( o_\text{U} \)) via \( o_\text{U} = p \cdot N \).

Error bars shown in figures for the number of occurrences of a particular amino acid or motif were calculated in this manner.

The expected distribution of amino acids and polypeptide motifs in naive FliTrx dodecapeptides is straightforward to calculate assuming that the library is random (Invitrogen). If we reject any clones where translation is interrupted by a stop codon, the 61 remaining codons specify all 20 amino acids, with some amino acids encoded by a single codon (M, W), some by two codons (F, Y, H, Q, E, D, C, N, K), and others by three (I), four (V, P, T, A, G), and six codons (L, S, R). We label the codon degeneracy of a particular amino acid \( d_j \), where \( j \) can denote any of the 20 amino acids. Thus, for amino acid \( j \), the expected probability for finding \( j \) in the FliTrx library (\( p_{e,j} \)) is given by:

\[
p_{e,j} = \frac{d_j}{61}
\]

(3)

with the expected number of occurrences being \( e_j = p_{e,j} \cdot N \). For motifs with multiple amino acids, calculation of the 95% confidence limits is identical, but the expected probability for finding the motif (\( p_M \)) is now given by:

\[
p_M = \prod_{j=1}^{m} p_{e,j}
\]

(4)

where \( \Pi \) is the product operator and any motif with unspecified spacer amino acids has \( p_{e,j} = 1 \) for those locations. For example, the expected probability for the motif RXRX in the naive FliTrx library is \( p_M = (6/61) \cdot (1) \cdot (1) \cdot (6/61) \). The expected number of occurrences for this motif in the library is \( e_M = p_M \cdot N \).

The computation of 95% confidence intervals for the positional analysis of Figure 2 is identical to the analysis of single amino acid confidence intervals, but in this case the number of independent trials is \( N = c \) (i.e., \( N = 31 \) for Cu₂O and \( N = 33 \) for ZnO), because each clone has only one opportunity to display a specific amino acid at a specific location. As a result of the small number of independent trials, the confidence intervals are quite large.

**Fluorescence Microscopy**

Single colonies of GIB26 cells (Invitrogen; F⁻ lacE⁻ ampC⁻::P<sub>raf</sub>L ΔfliC ΔmotB eda::Tn10) freshly transformed with plasmids encoding Cu₂O- or ZnO-binding sequences were grown for 15 to 18 h in 5 mL of IMC supplemented
with carbenicillin at 25°C with shaking to $A_{600} \approx 3$. Cells (10$^{10}$) were transferred to 50 mL of IMC supplemented with carbenicillin and tryptophan and cultures were grown as described earlier for 5 to 6 h. Aliquots corresponding to 10$^6$ cells were combined at room temperature with 250 µL of a 6 µM stock solution of the nucleic acid-specific fluorescent dye SYTO9 (Molecular Probes, Eugene, OR) in an Eppendorf tube. Samples were mixed by inversion and held in the dark for 15 to 25 min. A surface of pure Cu$_2$O, Cu(OH)$_2$, or ZnO or a patterned substrate was placed in the tube, which was held in the dark for 15 min. Cells were removed by pipetting and the surface was gently washed three times with 1 mL of IMC. Surfaces were blotted with laboratory tissues to remove moisture and transferred to the stage of a Nikon E800 upright microscope fitted with a rhodamine cube. Fluorescent cells were visualized following illumination with a mercury lamp and images were captured using METAMORPH software (Universal Imaging). Binding affinities were estimated by counting the number of fluorescent cells in at least three randomly selected fields. All experiments were carried out at least twice.

**Simulations**

The dodecapeptide sequences of Table I were modified by the addition of a cysteine–glycine–proline (CGP) tripeptide at their N-terminus and a cysteine–proline–glycine (CPG) tripeptide at their C-terminus to reproduce the loop displayed by the FliTrx system. The flanking cysteine residues were manually constrained with a disulfide bridge and all atom structures were built using the CHARMM27b4 package at pH 7.0. Structures were optimized by the steepest descent method for 3000 cycles and by the conjugate gradient method for another 3000 cycles. The peptides were solvated in a water box of 50 Å × 25 Å × 35 Å with 1462 water molecules described by the TIP3P model. Water molecules with an oxygen atom closer than 2.8 Å to the peptides were removed. To keep the system neutral, some chlorine and sodium ions modeled with Beglov and Roux potentials were added to the simulation box. The energy of the solvated systems was minimized by the steepest descent method for 500 cycles and by the conjugate gradient method for 500 cycles using CHARMM27b4. Each peptide was heated to 300K and equilibrated for 10,000 steps. The Brendsen method was used to keep to the temperature constant. Systems were coupled with a heat bath using a coupling time of 0.1 ps. Bonds containing hydrogen were held rigid using the SHAKE method with geometric tolerance of 10$^{-6}$. Short-range, nonbonded interactions were calculated by a switched potential with a switching function starting at 10 Å and reaching zero at a distance of 12 Å. Electrostatic interactions were calculated by the shifted potential with a cutoff distance of 12 Å. Three-dimensional periodic boundary conditions were used in the simulations. Molecular surfaces were calculated and rendered using Swiss-PdbViewer.

**Table I.** Characteristics of selected Cu$_2$O- and ZnO-binding sequences.

<table>
<thead>
<tr>
<th>Binder</th>
<th>Sequence$^a$</th>
<th>Adhering cells$^b$</th>
<th>Charge$^c$</th>
<th>pI$^d$</th>
<th>Hydrophilicity$^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu$_2$O Class I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN225</td>
<td>RHTDGLRRRAAR</td>
<td>100 ± 29</td>
<td>+3</td>
<td>12.0</td>
<td>+0.79</td>
</tr>
<tr>
<td>CN85</td>
<td>RARRQGVDVSRD</td>
<td>62 ± 23</td>
<td>+2</td>
<td>11.5</td>
<td>+1.38</td>
</tr>
<tr>
<td>CN86</td>
<td>KPRRSSAAARGSEG</td>
<td>53 ± 8</td>
<td>+3</td>
<td>12.0</td>
<td>+1.22</td>
</tr>
<tr>
<td>CN46</td>
<td>ADRTRGRGNC</td>
<td>46 ± 10</td>
<td>+3</td>
<td>11.5</td>
<td>+0.96</td>
</tr>
<tr>
<td>Cu$_2$O Class II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN44</td>
<td>NTVWRLNSSCGM</td>
<td>84 ± 31</td>
<td>+1</td>
<td>8.2</td>
<td>−0.45</td>
</tr>
<tr>
<td>CN88</td>
<td>EKWGHQECYRH</td>
<td>47 ± 18</td>
<td>0</td>
<td>7.0</td>
<td>0.27</td>
</tr>
<tr>
<td>CN93</td>
<td>TMEPRWWCNPIN</td>
<td>34 ± 8</td>
<td>0</td>
<td>5.7</td>
<td>−0.41</td>
</tr>
<tr>
<td>ZnO Class I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN179</td>
<td>RIGHGRLQRXPL</td>
<td>100 ± 17</td>
<td>+4</td>
<td>12.3</td>
<td>+0.53</td>
</tr>
<tr>
<td>CN155</td>
<td>VRTRDDARKH</td>
<td>69 ± 5</td>
<td>+3</td>
<td>11.6</td>
<td>+1.48</td>
</tr>
<tr>
<td>CN120</td>
<td>PASRVEKNGVAR</td>
<td>13 ± 6</td>
<td>+3</td>
<td>11.7</td>
<td>+1.00</td>
</tr>
<tr>
<td>ZnO Class II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN146</td>
<td>MRHSSSEPRL</td>
<td>84 ± 12</td>
<td>+1</td>
<td>9.4</td>
<td>+0.38</td>
</tr>
<tr>
<td>CN111</td>
<td>PAGLOQGVFAVEV</td>
<td>78 ± 13</td>
<td>−1</td>
<td>4.0</td>
<td>−0.55</td>
</tr>
<tr>
<td>CN185</td>
<td>RTDDGVAGRTWL</td>
<td>61 ± 7</td>
<td>0</td>
<td>6.0</td>
<td>+0.33</td>
</tr>
</tbody>
</table>

$^a$Paired basic residues are in **bold italics**. The RXRR and RXXRXR motifs are underlined.

$^b$Number of adhering cells obtained as described in Materials and Methods were normalized to the highest binder tested (CN225 for Cu$_2$O-specific clones and CN179 for ZNO-specific clones).

$^c$Calculated by subtracting the number of basic residues (R and K) from the number of acidic residues (D and E).

$^d$Obtained using the pI/MW calculator at Expsys (http://us.expasy.org/).

$^e$Calculated using the hydrophilicity index of Hopp and Woods where a score of +1.5 corresponds to a very hydrophilic protein and a score of −1.5 to a very hydrophobic one (Hopp and Woods, 1981).
RESULTS

Identification and Compositional Characterization of Cu$_2$O- and ZnO-Binding Peptides

Two metal oxides of engineering interest, the excitonic semiconductor, Cu$_2$O, and the wide-band-gap semiconductor, ZnO, were selected for this study. Cu$_2$O substrates were prepared by anodic electrodeposition on copper foil. The identity of the deposited film was routinely verified by Raman spectroscopy and the purity was $\approx 99\%$, as judged by XPS analysis. Polished ZnO single crystals were used as supplied by the manufacturer. The FliTrx library, which expresses random dodecapeptides flanked by cysteine–glycine–proline/glycine–proline–cysteine sequences (CGP/GPC) on the cell surface as FliC–Trx fusions (Lu et al., 1995) was panned against $\approx 1\text{-cm}^2$ inorganic surfaces using the protocol outlined previously. After five successive rounds of selection conducted to enrich for tight binders, 31 (Cu$_2$O) or 33 (ZnO) randomly selected clones were sequenced to determine the amino acid composition of the inserts.

Because neither Cu$_2$O- nor ZnO-binding sequences converged toward a consensus, we first compared the observed number of occurrences of each amino acid (Fig. 1; bars) to their theoretical chance of occurrence (horizontal bars), assuming that the FliTrx library is random. To assess the

Figure 1. Analysis of Cu$_2$O- and ZnO-binding sequences. Bars show the number of occurrences of each of the 20 amino acids in 31 (Cu$_2$O binders, light gray bars) or 33 (ZnO binders, dark gray bars) sequenced clones. Error bars represent the calculated 95% confidence interval in these values and are asymmetric due to the use of a binomial distribution. Horizontal bars correspond to the theoretical number of occurrences of each of the 20 amino acids based on the assumption that the FliTrx library is fully random. Statistically significant positional enrichment occurs when error bars do not intersect the horizontal bars; this is indicated by up arrows on top of the panels. Dark arrows indicate that both Cu$_2$O and ZnO binders are enriched for a particular residue. Gray arrows indicate that either Cu$_2$O or ZnO binders are enriched or depleted for a particular residue.

As a further test of the enrichment of the selected clones, the positional dependency of arginine (A), tryptophan (B), and glycine (C) was assessed. Error bars represent the calculated 95% confidence interval in these values and are asymmetric due to the use of a binomial distribution. Horizontal bars correspond to the theoretical number of occurrences of each of the three amino acids based on the assumption that the FliTrx library is fully random. Statistically significant positional enrichment occurs when error bars do not intersect the horizontal bars; this is indicated by up arrows on top of the panels. Dark arrows indicate that both Cu$_2$O and ZnO binders are enriched for a particular residue. Gray arrows indicate that either Cu$_2$O or ZnO binders are enriched or depleted for a particular residue.

Figure 2. Positional dependency of arginine (A), tryptophan (B), and glycine (C) occurrences along dodecapeptides selected for binding to Cu$_2$O (light gray bars) or ZnO (dark gray bars). Error bars represent the calculated 95% confidence interval in these values and are asymmetric due to the use of a binomial distribution. Horizontal bars correspond to the theoretical number of occurrences of each of the three amino acids based on the assumption that the FliTrx library is fully random. Statistically significant positional enrichment occurs when error bars do not intersect the horizontal bars; this is indicated by up arrows on top of the panels. Dark arrows indicate that both Cu$_2$O and ZnO binders are enriched for a particular residue. Gray arrows indicate that either Cu$_2$O or ZnO binders are enriched or depleted for a particular residue.

Because neither Cu$_2$O- nor ZnO-binding sequences converged toward a consensus, we first compared the observed number of occurrences of each amino acid (Fig. 1; bars) to their theoretical chance of occurrence (horizontal bars), assuming that the FliTrx library is random. To assess the
serine (S), and threonine (T). In the case of ZnO binders, tryptophan (W), glycine (G), and arginine (R) were overrepresented, whereas tyrosine (Y), proline (P), and serine (S) were underrepresented (Fig. 1). We conclude that the two metal oxide-binding populations exhibit similar compositional biases, but there are subtle differences between Cu2O and ZnO binders.

Because both populations exhibited an overrepresentation of arginine, tryptophan, and glycine residues, we next compared the positional occurrence of these amino acids along the dodecapeptides in the two clonal populations (Fig. 2). In the case of Cu2O binders (Fig. 2A, light gray bars), arginines (R) were enriched along most of the sequence with the highest rate of incidence at positions 3, 4, 5, and 11. Overrepresentation at these locations, as well as a more modest enrichment at positions 7 and 12, was statistically significant. For ZnO binders (Fig. 2A, dark gray bars) there was a lower positional bias, and the greatest chance of encountering an arginine was at locations 7, 11, and 12. However, only overrepresentation at position 7 was statistically significant. Similarly, tryptophan (W) exhibited little positional bias in ZnO binders, although it was specifically excluded at location 5 (Fig. 2B, dark gray bars). By contrast, the same residue was enriched in the peptide core (positions 6 and 7), and to a lesser extent at positions 3 and 4 in the case of Cu2O binders (light gray bars). Glycine residues (G) occurred most frequently in the central region of both sets of dodecapeptides, with a preference for positions 4, 6, 8, and 10 in Cu2O binders and for positions 3, 5, 6, 7, and 8 in ZnO binders, all of which were statistically significant (Fig. 2C).

High-Avidity Cu2O and ZnO Binders

To gain insight into the relationship between amino acid composition and strength of binding, GI826 transformants were labeled with SYTO9, a fluorescent dye that binds nucleic acids in metabolically active E. coli. Culture aliquots were contacted with slides of electrodeposited Cu2O or ZnO and the materials were examined by fluorescence microscopy after three cycles of washing to remove weakly bound cells. Whereas plasmid-free control cells did not adhere to the substrate (Fig. 3;GI826), those expressing Cu2O (Fig. 3) or ZnO (data not shown) binding sequences on their surface could be visualized as uniformly distributed bright rods on a dark background. The avidity of various clones for the metal oxides was estimated by counting the number of adhesive cells in three randomly chosen fields in at least duplicate experiments. These results as well as the amino acid sequences of a subset of strong binders and their physicochemical characteristics are reported in Table I.

Although CLUSTALW (Higgins et al., 1994) and MULTALIN (Corpet, 1988) analysis failed to reveal any significant homology between dodecapeptides when more than two sequences were aligned, two classes of binders could be distinguished in the case of Cu2O. Class I sequences are hydrophilic, highly basic, carry a net charge of at least +2, and often contain twin arginines, whereas Class II sequences contain tryptophan, tend to be more hydrophobic, and carry a 0 or +1 net charge. ZnO binders could also be divided into two subclasses based on pI and net charge. Class I ZnO sequences were reminiscent of Class I Cu2O sequences, except that arginine–lysine (RK) pairs were more common than arginine–arginine (RR; Table I). Class II ZnO binders had a lower incidence of tryptophan (W), but were enriched in the hydrophobic residues leucine (L) and valine (V) relative to Cu2O Class II members.

To probe the significance of paired basic residues identity, we reexamined the two populations for the presence of RR and RK motifs and calculated 95% confidence intervals for their likelihood of occurrence. Figure 4 shows that, although the overrepresentation of RR pairs in Cu2O binders was significant, that of RK pairs was not. The converse was true in the case of ZnO binders. Interestingly, the RXRR sequence (where X represents any amino acid) was encountered twice in high-avidity Class I Cu2O binders, whereas the RXXKR motif was found twice in Class I ZnO binders (Table I). Statistical analysis of the sequenced clones revealed that enrichment for the RXRR motif was highly significant in both metal oxide-binding populations (Fig. 4). These results suggest that RXRR sequences may be important for Cu2O and ZnO binding by disulfide-constrained dodecapeptides and that RR pairs may confer specificity for Cu2O, whereas RK pairs may be important for ZnO binding in Class I sequences.

Binding Specificity

Because of the high degree of similarity between the Cu2O- and ZnO-binding sequences (Table I), we evaluated the specificity of ZnO binding using a single gold substrate
electrochemically patterned with stripes of Cu$_2$O and ZnO. Four of the six clones tested recognized ZnO more efficiently than Cu$_2$O (Fig. 5). However, the ZnO to Cu$_2$O selectivity (defined as the ratio of the number of cells adhering to ZnO to those bound to Cu$_2$O) was variable with values of 3.1 for CN179, 2.2 for CN146, 1.5 for CN155, and 1.3 for CN185. More surprisingly, CN120 and CN111 bound Cu$_2$O better than ZnO, leading to selectivities of <1.0 (0.3 and 0.7, respectively). Three of the six clones examined were also capable of adhering to gold with CN146 being approximately as efficient at either gold or ZnO recognition, and CN155 and CN179 exhibiting ZnO to gold selectivities of 6.8 and 2.1, respectively. On the other hand, none of the ZnO or Cu$_2$O binders listed in Table 1 conferred cells the ability to bind to Cu(OH)$_2$ (data not shown).

Molecular Dynamics Simulations

To gain further insights into the nature of inorganic binding discrimination, we used molecular dynamics to model the structures of the CN179, CN155, and CN120 sequences flanked by CGP at the N-terminus, GPC at the C-terminus, and constrained by a disulfide bridge between cysteine residues. The computed molecular surfaces of these three Class I ZnO binders is shown in Figure 6. All peptides exhibited corrugation on the “top” face that should contact the inorganic. CN179 and CN120 had the most regular projecting features with an approximate periodicity of 5 Å, which may be important for matching the material crystal lattice. The location of basic residues in the RXXRK motif was mapped onto the structure of CN179 and CN155. In both cases, the first arginine of the motif (R6 or R8) lined the waist of the top face. Lysines K10 and K12 projected outward at 45° (CN179) or 135° (CN155) angles from this feature. The second arginine of the motif (R9 or R11) projected toward the “bottom” of both structures, which should preclude it from interacting with inorganic compounds. Rather, this arrangement may be optimal for the proper orientation of the (presumably inorganic-contacting) lysine residues. In the case of CN120, R11 and R12 were fully solvent exposed and at an ≈90° angle from each electrochemically patterned with stripes of Cu$_2$O and ZnO. Four of the six clones tested recognized ZnO more efficiently than Cu$_2$O (Fig. 5). However, the ZnO to Cu$_2$O selectivity (defined as the ratio of the number of cells adhering to ZnO to those bound to Cu$_2$O) was variable with values of 3.1 for CN179, 2.2 for CN146, 1.5 for CN155, and 1.3 for CN185. More surprisingly, CN120 and CN111 bound Cu$_2$O better than ZnO, leading to selectivities of <1.0 (0.3 and 0.7, respectively). Three of the six clones examined were also capable of adhering to gold with CN146 being approximately as efficient at either gold or ZnO recognition, and CN155 and CN179 exhibiting ZnO to gold selectivities of 6.8 and 2.1, respectively. On the other hand, none of the ZnO or Cu$_2$O binders listed in Table 1 conferred cells the ability to bind to Cu(OH)$_2$ (data not shown).

Molecular Dynamics Simulations

To gain further insights into the nature of inorganic binding discrimination, we used molecular dynamics to model the structures of the CN179, CN155, and CN120 sequences flanked by CGP at the N-terminus, GPC at the C-terminus, and constrained by a disulfide bridge between cysteine residues. The computed molecular surfaces of these three Class I ZnO binders is shown in Figure 6. All peptides exhibited corrugation on the “top” face that should contact the inorganic. CN179 and CN120 had the most regular projecting features with an approximate periodicity of 5 Å, which may be important for matching the material crystal lattice. The location of basic residues in the RXXRK motif was mapped onto the structure of CN179 and CN155. In both cases, the first arginine of the motif (R6 or R8) lined the waist of the top face. Lysines K10 and K12 projected outward at 45° (CN179) or 135° (CN155) angles from this feature. The second arginine of the motif (R9 or R11) projected toward the “bottom” of both structures, which should preclude it from interacting with inorganic compounds. Rather, this arrangement may be optimal for the proper orientation of the (presumably inorganic-contacting) lysine residues. In the case of CN120, R11 and R12 were fully solvent exposed and at an ≈90° angle from each other.

Figure 4. Significance of motifs involving basic residues. Bars show the number of occurrences of twin arginines (RR), arginine–lysine (RK) pairs, and of the RXXR motif (where X represents any amino acid) in 31 (Cu$_2$O binders, light gray bars) or 33 (ZnO binders, dark gray bars) sequenced clones. Error bars represent the calculated 95% confidence interval in these values (see Materials and Methods) and are asymmetric due to the use of a binomial distribution. Horizontal bars correspond to the theoretical number of occurrences of these motifs based on the assumption that the FliTrx is fully random. Statistically significant enrichment for a particular motif occurs when error bars do not intersect the horizontal bars.

Figure 5. Cross-specificity of ZnO-binding sequences. GI826 transformants harboring plasmids specifying the indicated ZnO recognition motifs were contacted with a gold surface patterned with ZnO and Cu$_2$O and the number of adhering cells was counted in at least three fields. Values shown are normalized to the number of adhering cells for GI826/CN179 that was arbitrarily set at 100.

Figure 6. Simulated molecular surfaces of selected metal oxide binders. Structures on the left correspond to a “front” view of the disulfide-constrained peptides, whereas structures on the right correspond to a “top” view rotated by 45°. In the “front” view, disulfide bridges are at the “bottom” of the structure. Basic residues are dark-shaded. The numbering scheme assigns position 1 to the first residue of the random dodecapeptide rather than to the first cysteine of the modeled 18-mer.
DISCUSSION

In this study, we have used the FliTrx cell surface display system to isolate disulfide-constrained dodecapeptides that confer E. coli the ability to bind Cu$_2$O or ZnO surfaces when presented on the flagellar protein FliC. Because selected sequences do not converge toward a consensus when phage or cell surface display libraries are panned against inorganic compounds (this work; see also Brown, 1997; Brown et al., 2000; Kjærgaard et al., 2000; Lee et al., 2002; Li et al., 2002; Naik et al., 2002a, 2002b; Whaley et al., 2000), we used a combination of compositional analysis, cross-spectrivity analysis, and simulations to gain insight into the nature of the interactions that govern the binding of polypeptides to metal oxides.

Cu$_2$O- and ZnO-binding sequences were found to exhibit similar, although subtly different patterns of amino acid utilization (Fig. 1). The two populations were enriched in the basic amino acid arginine and the aromatic residue tryptophan with a more pronounced positional bias in the case of Cu$_2$O binders (Fig. 2). Glycine was overrepresented in the central region of both sets of dodecapeptides, presumably because this small amino acid allows for flexible orientation of neighboring residues. We also observed a bias against tyrosine, proline, and serine residues in both populations, as well as an enrichment for methionine and a depletion for leucine and threonine residues that was specific to Cu$_2$O binders. Although the reasons for the latter compositional differences remain obscure, leucine may play a role in oxide discrimination, because it was underrepresented in Cu$_2$O binders but overrepresented in ZnO binders (Fig. 1).

Clustering of high-avidity binders revealed the existence of two distinct classes of metal oxide binders (Table I). In the positively charged and hydrophilic Class I sequences, twin arginines (RR) were often encountered in Cu$_2$O binders, whereas arginine-lysine (RK) pairs were more common in ZnO binders. Furthermore, two Class I Cu$_2$O binders contained an RXRR sequence (where X represents any amino acid), the counterpart of which appears to be RXXRK in Class I ZnO binders. The significance of these patterns was confirmed by statistical analysis of the clonal populations (Fig. 4), suggesting a role for the RXRR motif in metal oxide binding and for paired basic residues in discrimination between Cu$_2$O and ZnO. In support of the former proposal, the RXXRS motif was previously identified in an independent selection for ZnO-binding peptides (Kjærgaard et al. 2000). On the other hand, the existence of more neutral and hydrophobic Class II binders that do not necessarily contain basic amino acids (e.g., CN111) highlights the fact that there may be quite different solutions to the problem of metal oxide binding.

In an attempt to correlate spatial arrangement of basic residues with binding specificity, we used molecular dynamics to predict the surface of three Class I ZnO binders that had been tested for their ability to bind Cu$_2$O, ZnO, and Au (Figs. 5 and 6). Comparisons of simulation results and avidity data suggest that the most efficient ZnO interactions occur when the first arginine and the lysine residue of the RXXXRK motif are positioned at an $\approx 135^\circ$ or $45^\circ$ angle and that efficient Cu$_2$O binding takes place when paired arginines are at an $\approx 90^\circ$ angle. Although these predictions should be viewed with care because other side chains, surface features, and the carrier protein framework almost certainly modulate peptide–inorganic interactions, it is of interest to note that the arrangement of basic residues in CN179 and CN155 matches the ZnO crystal structure (Wurzite structure with lattice parameters $a_o = 3.2495 \text{ Å}$ and $c_o = 5.2069 \text{ Å}$ at 300K) and that of twin arginines in CN120 and CN86 matches the structure of Cu$_2$O (cubic with unit cell edge $a = 4.2696 \text{ Å}$ at 299K). Finally, it is worthwhile to point out that, whereas interactions between basic residues and the inorganic appear to be important for ZnO and Cu$_2$O binding, the hydration state of the surface also plays a role, because none of the binders tested adhered to the chemically related compound Cu(OH)$_2$.

Why ZnO binders exhibit highly variable affinities for gold (Fig. 5) remains unclear. It has been observed that the hydroxyl side chain amino acids, threonine and serine, are enriched in gold-binding peptides (Brown, 1997). Indeed, CN146, a Class II ZnO binder that contains a string of three successive serines, bound ZnO and Au with comparable efficiency (Fig. 5). However, among the ZnO binders modeled, the most efficient Au binder (CN179) does not contain either of these residues, whereas both CN155 (a weak Au binder) and CN120 (a non-Au binder) do.

Direct testing of the role of basic residues in Cu$_2$O/ZnO binding was not possible with synthetic peptides owing to their surface activity that resulted in the dissolution of oxide films (unpublished data). To further our understanding of peptide-oxide binding, we have engineered the CN225 sequence within the framework of a DNA-binding protein and demonstrated that it endows it with the ability to bind Cu$_2$O with nanomolar affinity. A combination of site-directed mutagenesis and molecular dynamics simulation is being used to pinpoint the contribution of basic amino acids and overall structure to metal oxide binding. From a practical standpoint, the availability of such bifunctional proteins paves the way to explore the concept that genetically engineered proteins for functional inorganics (GEPI) can be used to control the composition and architecture of inorganic materials at the nanoscale level (Sarikaya, 1999; Sarikaya et al., 2003).
CONCLUSIONS

This study demonstrates the suitability of the FliTrx cell surface display system for the isolation of inorganic binding polypeptides and how a combination of statistical analysis, microfabrication, fluorescence microscopy, and simulations can provide insights on the nature of the rules governing peptide–inorganic interactions. From a practical standpoint, our observation that sequences selected for ZnO binding exhibit moderate to high affinity for Cu₂O suggests that complete discrimination between two related compounds may be difficult to achieve in a single panning experiment. These results also underscore the fact that care will be required if using published sequence information to assemble hybrid nanomaterials containing more than one inorganic component.

The authors are grateful to Jian Zhou and Shaoyi Jiang for help with molecular dynamics simulations. We also thank one of the reviewers of this manuscript for very helpful suggestions on statistical analysis of the data.

References
