BIOINFORMATICS ORIGINAL PAPER Vol. 23 no. 21 2007, pages 2816–2822
doi:10.1093/bioinformatics/btm436

Sequence analysis

A novel knowledge-based approach to design inorganic-binding peptides

Ersin Emre Oren[1,2], Candan Tamerler[3], Deniz Sahin[1,3], Marketa Hnilova[1],
Urartu Ozgur Safak Seker[1,3], Mehmet Sarikaya[1,*] and Ram Samudrala[2,*]
[1]Materials Science and Engineering Department, [2]Department of Microbiology, Computational Genomics Group,
University of Washington, Seattle, USA and [3]Molecular Biology and Genetics, Istanbul Technical University,
Istanbul, Turkey

Received on June 6, 2007; revised on August 1, 2007; accepted on August 16, 2007
Advance Access publication September 17, 2007
Associate Editor: Keith Crandall

ABSTRACT

Motivation: The discovery of solid-binding peptide sequences is accelerating along with their practical applications in biotechnology and materials sciences. A better understanding of the relationships between the peptide sequences and their binding affinities or specificities will enable further design of novel peptides with selected properties of interest both in engineering and medicine.

Results: A bioinformatics approach was developed to classify peptides selected by *in vivo* techniques according to their inorganic solid-binding properties. Our approach performs all-against-all comparisons of experimentally selected peptides with short amino acid sequences that were categorized for their binding affinity and scores the alignments using sequence similarity scoring matrices. We generated novel scoring matrices that optimize the similarities within the strong-binding peptide sequences and the differences between the strong- and weak-binding peptide sequences. Using the scoring matrices thus generated, a given peptide is classified based on the sequence similarity to a set of experimentally selected peptides. We demonstrate the new approach by classifying experimentally characterized quartz-binding peptides and computationally designing new sequences with specific affinities. Experimental verifications of binding of these computationally designed peptides confirm our predictions with high accuracy. We further show that our approach is a general one and can be used to design new sequences that bind to a given inorganic solid with predictable and enhanced affinity.

Contact: sarikaya@u.washington.edu or ram@compbio.washington.edu

Supplementary information: Supplementary Material containing, the quartz-binding peptide sequences, additional results and the specific scoring matrices are available at *Bioinformatics* online.

1 INTRODUCTION

The formation and structuring of solid components of biological hard tissues are controlled by proteins leading to their complex and highly functional architectures (Mann, 1988; Sarikaya, 1999; Weiner and Addadi, 1997). These minerals include calcium carbonate polymorphs ($CaCO_3$) in mollusk shells and echinoderm spines and tests, silica-based (SiO_2) skeletal units of single-celled organisms such as radiolarian and spicules of sponges, magnetic (Fe_3O_4) nanoparticles in magnetotactic bacteria and hydroxyapatite nanoparticles in bone and dental tissues in mammalians (Lowenstam, 1981). Proteins are known to control nucleation, growth and structure formation of minerals and provide molecular scaffolds in the formation of hard tissues (Mann, 1988; Paine and Snead, 1997; Sarikaya, 1999; Weiner and Addadi, 1997). There has been a recent surge of research activity in utilizing genetically engineered peptides that could be used for inorganic materials synthesis, assembly and formation under ambient conditions (Ball, 2001; Sarikaya *et al.*, 2003; Seeman and Belcher, 2002).

The peptides used in practical solid materials formation are selected using combinatorial biology techniques based on the developments during last two decades. For example, *in vivo* (e.g. phage Hoess, 2001; Smith, 1985; and cell-surface display, Wittrup, 2001) and *in vitro* (e.g. ribosomal and mRNA display, Amstutz *et al.*, 2001) combinatorial biology protocols originally developed to select peptides with affinity to biological entities such as enzymes, cells, viruses and other proteins. These approaches have recently been adapted by us and others for selecting peptides that specifically bind to desired inorganic material substrates (Brown, 1997; Naik *et al.*, 2002; Sarikaya *et al.*, 2003, 2004; Thai *et al.*, 2004; Whaley *et al.*, 2001). Although the nature of peptide-inorganic interaction is not yet well understood, many short peptide sequences specific to metals (silver, gold, platinum, palladium and titanium), oxides (silica, magnetite and titanium oxide), minerals (calcite, hydroxyapatite and quartz) and semiconductors (cadmium sulfide, zinc sulfide, zinc oxide and cuprous oxide) have been discovered as potential utility for future engineering materials and have been used in the proof-of-principle synthesis, morphogenesis and assembly of inorganics, finding practical applications in a wide ranging and diverse areas of nanotechnology and medicine (Gaskin *et al.*, 2000; Naik *et al.*, 2002; Sano *et al.*, 2005; Sarikaya *et al.*, 2004).

*To whom correspondence should be addressed.

The question of how proteins recognize and bind to minerals and inorganic substrates with specific affinities and specificities has been a long-term issue both for the purpose of understanding hard tissue regeneration (e.g. bone and dental tissues) and also making practical materials using synthetic peptides as nucleators, growth modifiers or as control agents. Whether medical or practical engineering applications, the understanding of the possible mechanism(s) of inorganic formation has to be eventually addressed for rational design and tailoring of these peptides towards specific materials systems (Gaskin *et al.*, 2000; Ratner *et al.*, 1996). Some of these questions, e.g. include what sequence domain or molecular structure act as the catalyzer for inorganic formation (Shimizu *et al.*, 1998), what sequence- or structure-affect-oriented nucleation or control growth are all part of this puzzle. Experimental as well as modeling studies towards this understanding are in their infancy (De Yoreo and Dove, 2004; Mann, 1988; Paine *et al.*, 2001; Weiner and Addadi, 1997), and accurate force field parameters required to model the protein substrate interaction are still under development (Gray, 2004; Oren *et al.*, 2005; Zhou *et al.*, 2003). The diverse range of inorganics that need to be characterized further confounds the solution to this problem.

Our approach in addressing the issue of specific inorganic-binding peptides is based on the observation that, in nature, proteins that perform similar functions usually have similar sequences due to evolutionary, biochemical and biophysical constraints (Attwood, 2000). Protein sequence alignment is a basic tool used by biologists for various analyses, from detecting key functional residues to inferring the evolutionary history of a protein family. Typically, pairs of sequences are aligned using an optimization procedure, such as dynamic programming (DP) (Needleman and Wunsch, 1970; Smith and Waterman, 1981) that finds the best possible relative arrangement of the amino acids maximizing the overall similarity score. Various heuristic methods that aim for speed rather than absolute accuracy have been developed (Altschul *et al.*, 1994, 1997; Lipman and Pearson, 1985; Pearson and Lipman, 1988; Thompson *et al.*, 1994). A scoring matrix is used to obtain the score for aligning two amino acids (match or mismatch) in an alignment of two protein sequences, and the overall score can be considered as a measure of the similarity between sequences. BLOSUM (Henikoff and Henikoff, 1992) and PAM (Dayhoff *et al.*, 1978) are two widely used scoring matrices derived from naturally occurring sequences. These matrices have been extensively evaluated for nucleotide and protein sequence comparisons with the primary goal of inferring homology or evolutionary relationships found in nature.

We hypothesized that a set of peptides generated by directed evolution through *in vivo* selection to recognize a given solid material will have similar sequences, much as evolutionarily related proteins do. Based on this, we defined a metric to assess the global similarities of selected sets of experimentally determined inorganic binders to shed more light into the understanding of the similarities of these peptides. The relationships among the known inorganic binders were then used to bootstrap new scoring matrices (Gonnet *et al.*, 1992; Kann *et al.*, 2000) and to select novel peptides with specific affinities from a pool of randomly generated ones. Using this approach, we are able to identify novel peptides with predictable functionalities including superior- or non-binding to a given inorganic material.

2 MATERIALS AND METHODS

2.1 Computational methods

2.1.1 Data source Peptides binding specific inorganics were produced using phage-display techniques as described in Naik *et al.* (2002), Sarikaya *et al.* (2003), Smith (1985) and Whaley *et al.* (2001). Briefly, a phage library (New England Biolabs Inc., 2006) containing 2.7×10^9 phage clones was exposed to the target substrate and *in vivo* selection or panning was carried out by washing away the unbound phage, and eluting the specifically bound phage. The eluded phage is then amplified and taken through additional binding/amplification cycles to enrich the pool in favor of binding sequences. We used a total of 39 quartz (rhombohedral silica, SiO_2)-binding peptides in the quartz-specific analyses described here (see Supplementary Material). These peptides were further characterized using affinity analysis such as fluorescence microscopy in which the bound phages are visualized and classified into 10 strong, 14 moderate and 15 weak binders. Further details of the generation procedure are given elsewhere (Oren *et al.*, 2007).

2.1.2 Similarity score calculation The Needleman–Wunsch dynamic programming algorithm (Needleman and Wunsch, 1970), which guarantees an optimal scoring alignment with a given scoring matrix, was used for sequence similarity comparisons. Given a scoring matrix, the overall similarity score of a pairwise alignment is defined by the sum of all similarity values of the aligned residue pairs minus a gap penalty for every insertion or deletion introduced into and/or extended in the alignment.

In general, the cost for opening a new gap in a sequence is higher than extending an existing gap, and the accuracy of the alignment heavily depend on the selection of these parameters (Vogt *et al.*, 1995). In the following calculations the affine gap formula, $g(k) = -\mathrm{gop} - (k-1)\mathrm{gep}$ is used to penalize the gaps. Here, k is the gap length ($k = 1, 2, \ldots, n$), gop and gep are the gap opening and gap extension penalties, respectively. In this work, we assumed that $\mathrm{gep} = 0.1\mathrm{gop}$.

Using the Needleman–Wunsch algorithm, the pairwise similarity scores (PSS) of any two peptide sequences with a given scoring matrix and gap opening and extension penalties was first calculated, followed by the total similarity scores (TSS) between two sets of binders, A and B, by using every possible PSS. TSS calculated using the sequences within the same set is referred to as self-TSS otherwise TSS calculated between different sets is referred to as cross-TSS. The TSS are normalized according to following expression:

$$TSS_{A-B}\big([A]_{NA} - [B]_{NB}\big) = \frac{1}{NA \cdot (NB - \delta_{AB})} \sum_{i=1}^{NA} \sum_{j=1}^{NB} PSS_{ij}\big(1 - \delta_{ij}\delta_{AB}\big) \tag{1}$$

where, δ is the usual Kronecker delta function ($\delta_{ij} = 1$ *if* $i = j \wedge \delta_{ij} = 0$ otherwise), NA and NB are the total number of sequences in sets A and B, and PSS_{ij} is the PPS value between the i^{th} sequence of set A and j^{th} sequence of set B.

2.1.3 Scoring matrix generation Scoring matrices such as BLOSUM 62 and PAM 250 are derived from naturally occurring protein sequences and are generally meant to be applied to such sequences. The utility of these scoring matrices may be limited in terms of comparing peptides that bind to inorganic substrates that were selected by directed evolution. A given set of inorganic binders are typically characterized semi-quantitatively through immunofluorescence microscopy into three groups as strong, moderate and weak binders (Sarikaya *et al.*, 2003). Our goal is to use existing scoring matrices as a starting point to derive new ones that capture the relationships within the strong binders while differentiating them from the weak-binding peptides. To accomplish this, we iteratively perturb the matrices using a greedy procedure: after each perturbation we calculated both the self-TSS of the strong binders ($TSS_{S–S}$) and the cross-TSS between the strong and weak binders ($TSS_{S–W}$). The maximization of the difference between $TSS_{S–S}$ and $TSS_{S–W}$ is used as an objective function and any perturbation that increased this difference is accepted and all other perturbations are rejected.

2.1.4 Computational inorganic-binding peptide design Our bioinformatics approach also enables us to generate novel peptide sequences with predictable binding affinities. To accomplish this, we first generated random sequences based on the observed amino acid frequencies in the phage library used for the combinatorial selection (New England BioLabs Inc., 2006). We then calculated the TSS between each of these sequences and the experimentally determined strong binder group. Sequences with the highest and lowest similarity scores were considered to represent the strongest and weakest binders, respectively.

2.2 Experimental methods

Designed peptides were synthesized by and purchased from United Biochemical Research Inc. (Seattle, WA, USA) with a purity >95%. For adsorption characterization of the designed peptides on SiO_x, a Kretschmann configuration surface plasmon resonance (SPR) spectrometer, developed by Radio Engineering Institute Czech Republic, was used. SPR spectroscopy detects the refractive index change at a metal aqueous interface (Jung *et al.*, 1998). Commonly gold (50 nm) and silver is used as sensing layer on SPR chip. We used SiO_x as an additional layer on gold for novel biosensing applications (Szunerits and Boukherroub, 2006). We first coated a gold SPR chip with 4 nm SiO_x using ion-beam sputter coater (Gatan Inc., PA, USA), operated at 6 keV with a 10 mA/cm^2 ion current density and under 6×10^{-5} Torr vacuum. The amount of bound peptide on the surface was determined by the shift in the refractive index dip position. A higher shift reflects high amount of peptide adsorption and a sharp increase reveals a faster binding (Chang *et al.*, 2006; Tamerler *et al.*, 2006).

3 RESULTS AND DISCUSSIONS

3.1 Similarity analysis of experimentally characterized quartz-binding peptides

The experimentally characterized quartz-binding peptides (10 strong, 14 moderate and 15 weak binders; see Methods) were used to develop our bioinformatic approach. Sequence relationships within and between different affinity groups were determined by calculating the TSS for sequences belonging each of the three affinity groups. The TSS were calculated using both BLOSUM 62 (Fig. 1A) and PAM 250 (Fig. 1B) scoring matrices and with various gap opening (1–10) and extension penalties (0.1–1), and the penalties that minimized self- and cross-TSS while remaining positive were chosen for further analysis (see Supplementary Material).

Figure 1A and B shows that the self-TSS of strong-binding quartz sequences (black bars) is higher while the self-TSS of weak quartz binders (light gray bars) is lower compared to the TSS of all quartz binders (white bars with gray base). The similarity scores reveal that a relatively small number of sequences with significant sequence similarity to each other possess strong affinity to quartz. The remaining sequences are either moderate or weak binders and therefore have lower or no significant similarities. This observation provides evidence in favor of our hypothesis that relationships between experimentally discovered sequences can be uncovered using bioinformatics tools developed for naturally occurring protein sequence analysis. Further, it enables one to use the similarity scores to design and characterize new material-specific scoring matrices as well as peptides with specific binding affinities and specificities.

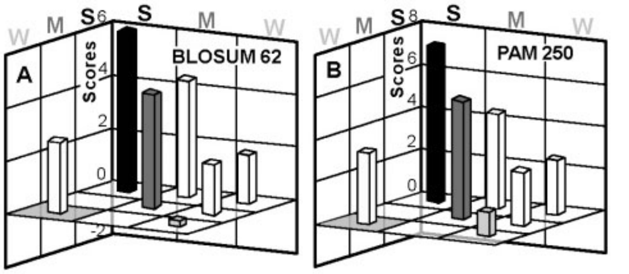

Fig. 1. (**A** and **B**) Total similarity scores (TSS) between the strong (S: black), moderate (M: dark gray) and weak (W: light gray) quartz binders. The TSS were calculated using the (A) BLOSUM 62 and (B) PAM 250 scoring matrices with gap penalties that minimize TSS while ensuring that the sign is positive (see Supplementary Material). The TSS of all the sequences is indicated in white with gray base. Sequences that strongly bind quartz have the highest self-TSS.

3.2 Design of quartz-specific scoring matrices

We hypothesized that the predictive power of approach could be further improved (beyond what is observed in Fig. 1A and B) by developing a new scoring matrix that takes into account the specific sequence patterns responsible for quartz binding. We chose PAM 250 as the seed matrix to optimize a new scoring matrix specific to inorganic quartz-binding peptides (QUARTZ I) (see Methods section). This matrix, along with the BLOSUM 62 and PAM 250 were then used to classify the experimentally characterized peptides.

3.3 Classification of experimentally characterized quartz-binding peptides

To demonstrate the predictive power of our approach to characterize the binding affinities of the quartz-binding peptides, the TSS between each peptide sequence (P) and the strong quartz binder group ($TSS_{P–S}$) were calculated. This was accomplished by removing the peptide being evaluated from the strong quartz binder set, if present, to prevent an artificial inflation of the similarity scores (i.e. leave-one-out cross-validation). The results are illustrated in Figure 2A–C for the three scoring matrices used. In Figure 2, the top bar graphs are the similarity scores between the individual quartz binders and the strong quartz binder set, the middle bar graphs are their corresponding surface coverage values (higher values indicate greater binding affinities), and the bottom bars show the correspondence between the predicted and experimental

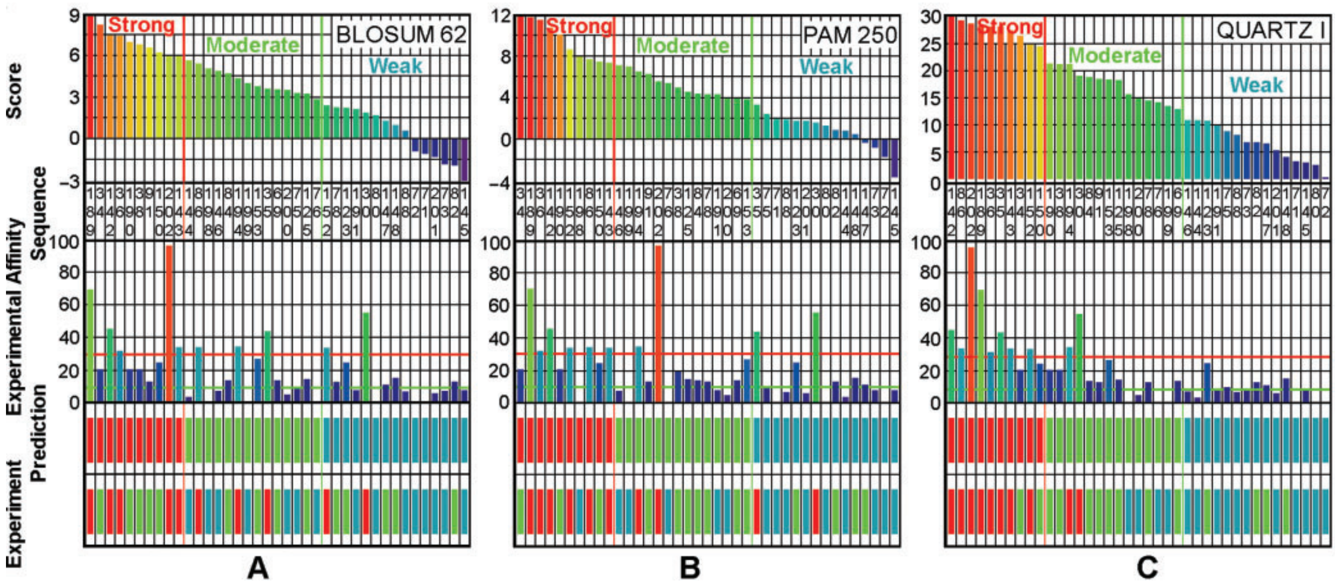

Fig. 2. Classification of experimentally characterized quartz-binding peptides. The top bar graph shows the peptides ordered based on a decreasing TSS_{P-S} score calculated using the (**A**) BLOSUM 62, (**B**) PAM 250 and (**C**) QUARTZ I scoring matrices. The corresponding experimentally determined affinities are shown in the middle bar graphs. The bottom bar graphs show the qualitative (strong (red), moderate (green) and weak (teal)) correspondence between the predicted and experimental affinities. The QUARTZ I matrix is able to best classify the 39 known experimental quartz binders according to their affinities.

affinities. The results indicate that even though the BLOSUM 62 and PAM 250 scoring matrices work well (50 and 60% accuracy, respectively), the QUARTZ I matrix is able to most accurately (80%) classify the known experimental quartz binders according to their affinities.

We also carried out similarity score classification of a given peptide by comparing it with both the strong (TSS_{P-S}) and the weak (TSS_{P-W}) quartz binder sets. The results indicate that comparing the sequences with the strong and weak binders does not improve the prediction accuracy. We expect that this is due to the low similarity score (TSS_{W-W}) of the weak quartz binder group (Fig. 1A–B) (see Supplementary Material, Section 4).

3.4 Computational design of new quartz-binding peptides

To demonstrate the predictive power of our approach, we generated 1 000 000 random sequences (12 000 000 amino acids total) and calculated their TSS to the experimentally known strong quartz-binding sequences.

Figure 3 shows the range of the calculated scores for the sequences and their distributions using the BLOSUM 62, PAM 250 and QUARTZ I scoring matrices. Then, using the cutoff scores between the strong, moderate and weak binders as shown in Figure 2, the percentages of having different affinity groups were calculated. Approximately 6% of the sequences were predicted to be strong binders using the QUARTZ I scoring matrix. Given that our false positive rate in identifying strong quartz binders (Fig. 2C) is 2 out of 10, we would expect that roughly 80% of these sequences are strong quartz binders.

Our analysis illustrates how our matrices can be used to design new peptides possessing specific affinities to quartz. We chose a final set of six strong and four weak predicted quartz binders based on the consensus of the TSS using all three scoring matrices (Fig. 4). We emphasize here that at the time of selection, there was no experimental information available about these 10 peptides, which are distinct and independent from any of the other experimentally characterized peptides.

3.5 Experimental validation of computationally designed quartz-binding peptides

We synthesized the 10 designed peptides and experimentally evaluated their binding characteristics using a surface plasmon resonance spectroscopy assay (see Methods section). We compared the binding affinities of our predicted peptides to the strongest phage display selected peptide previously observed, i.e. DS 202 (RLNPPSQMDPPF). Figure 5 shows that our designed sequences exhibit binding affinities to quartz as predicted. Four of the peptides, especially S1 that has the highest score, are much stronger than the DS202, the strongest quartz binder previously observed (and used as part our bioinformatics approach). The superior binding is achieved due to the accumulation of information from different strong peptides into our scoring matrix and score calculations, and indicates that our approach can be used to obtain second-generation peptides with superior functionality. Further independent experimental studies on these computationally designed peptides, including immunofluorescence analysis and gold quantum-dot immobilization on quartz using biotinylated peptides, extensively verify our design methodology (Oren *et al.*, 2007).

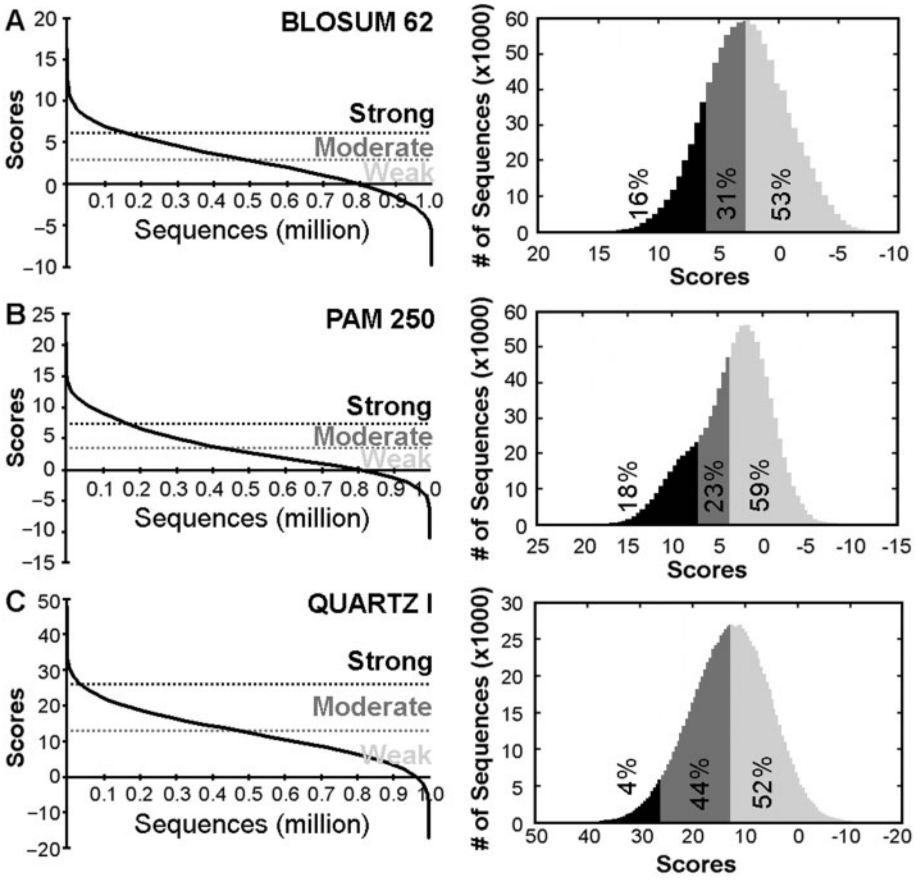

Fig. 3. Computational design of new peptides that bind to quartz. The range of similarity scores (left) and their distributions (right) between the randomly generated sequences and experimentally characterized strong quartz-binding peptides are shown. The peptides are ordered based on a decreasing TSS score using (**A**) BLOSUM 62, (**B**) PAM 250 and (**C**) QUARTZ I scoring matrices. The corresponding sequence distributions with respect to the TSS of the peptides are subdivided using the cutoff scores calculated in Figure 2. The sequences with the highest and lowest TSS are taken to represent the strongest and weakest binders.

3.6 Second generation scoring matrices

The predictive power of scoring matrices is dependent on the quantity and the quality of the initial data from which the scoring matrices are generated. We, therefore, expanded our initial set of 39 quartz binders with the 10 newly designed sequences and used the QUARTZ I matrix to optimize a new scoring matrix as described in the Methods section. As described in Section 3.2, we performed leave-one-out cross-validation to assess the ability of the new QUARTZ II to accurately classify the peptides. Figure 6 shows that the QUARTZ II matrix is slightly better at classifying quartz binders than the QUARTZ I matrix, particularly for the strong binders. The new QUARTZ II matrix can then be used to design more sequences, enabling the accuracy of our approach to be improved in an iterative fashion.

3.7 Analysis of residue preferences of the inorganic-binding peptides

We investigated the amino acid distributions in the strong quartz-binding peptides and compared their relative abundance to the weak-binding peptides (Fig. 7). Since only a small number of sequences are available in a given category, position-independent distributions were compiled and normalized according to the frequencies used to generate them from the phage display technique (New England BioLabs Inc., 2006).

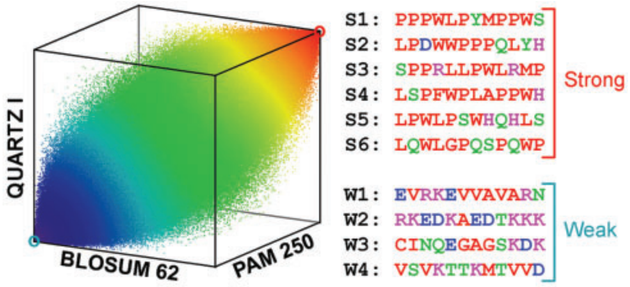

Fig. 4. Correlation between the TSS calculated by different scoring matrices for 1 000 000 peptides. Six strong and four weak predicted binding sequences were chosen based on the agreement of the TSS (indicated by circles). The sequences of designed strong (S) and weak (W) peptides are shown on the right and the amino acids are colored according to their chemical properties (hydrophobic, acidic, basic and polar).

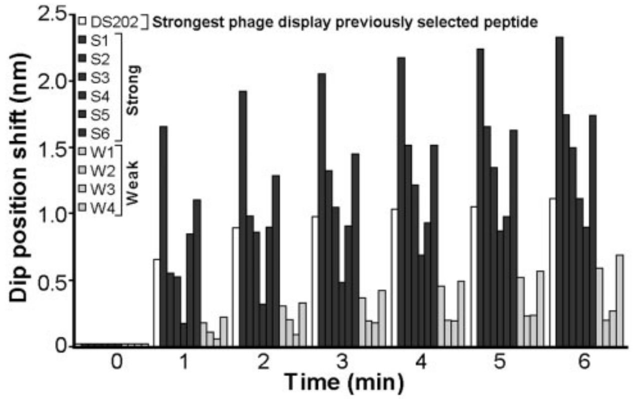

Fig. 5. Experimental validation of computationally designed peptides. Surface plasmon resonance spectral analysis that measures the amount of bound peptide versus time was performed at 4 μM concentrations for six strong (black) and four weak (light gray) designed peptides together with DS202 (white), the strongest phage display selected peptide. The higher the shifts in the dip position at a particular time, the stronger the binding and also the sharper the shift reveals a faster binding. Our designed peptides all exhibit binding affinities as predicted, with all but four of our stronger binders demonstrating equal or higher affinity than the strongest quartz binder previously observed.

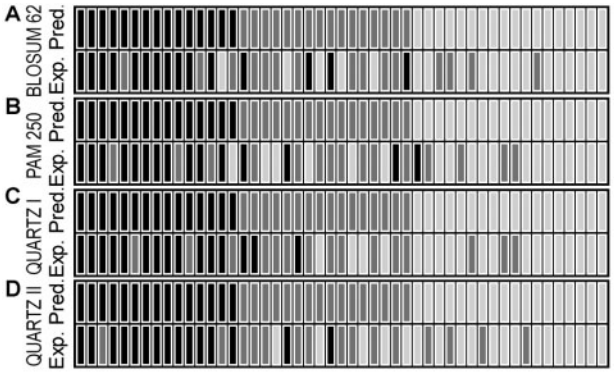

Fig. 6. Performance of the QUARTZ II scoring matrix. The 49 peptides, for which the experimental data are available, were used to design a new matrix and its performance is compared to the previously evaluated matrices. The new matrix has the best performance and can be used to more accurately design quartz binders with more specific affinities.

Based on these preferences (Fig. 7), we speculate that the strong quartz-binding peptides pack against the inorganic surface and reduce exposure to water. The peptides also likely have extended conformations: The bulky hydrophobic side chains of Tryptophan, Phenylalanine and Methionine in a small peptide require adequate spacing, and the Proline residue reduces conformational flexibility. Further, residues that may allow for collapse of the peptides either directly (through the formation of salt bridges between oppositely charged amino acids, disulfide bridges between Cysteines or collapse of the smaller hydrophobes) or indirectly (Glycine, which increases conformational flexibility) are underrepresented. Further investigation using simulation and solid state NMR techniques are being used to investigate the exact conformational nature of these peptides and their binding modes.

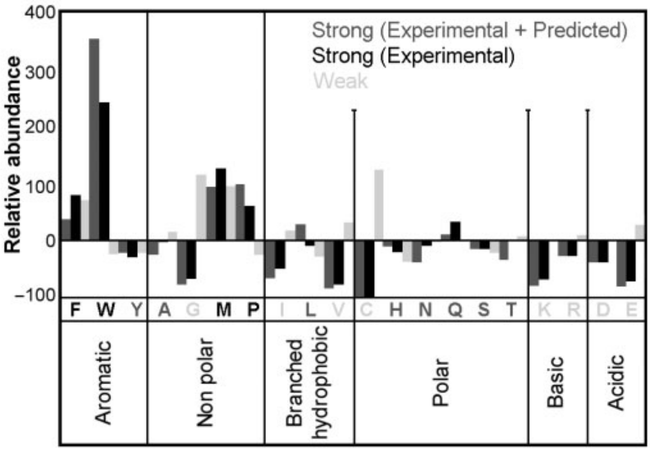

Fig. 7. Amino acid distributions of the quartz binders. Overrepresented amino acids in the strong binders are Tryptophan, Phenylalanine and Methionine, which contain bulky hydrophobic side chains and Proline which also contains a hydrophobic side chain that reduces conformational flexibility. Underrepresented amino acids in the strong binders are the four charged amino acids, the smaller hydrophobes and Glycine which does not have a side chain (thereby increasing main chain flexibility). Unlike the weak binders, strong binders do not contain any Cysteine, which may cause collapse of the binders by forming disulfide bridges. Based on these preferences, we speculate that the strong quartz-binding peptides have extended conformations, pack against the inorganic surface, and reduce exposure to water.

3.8 The importance of the amino acid composition and the sequence order

To demonstrate the relative importance of the simple amino acid composition compared to the sequence, we have also developed a classifier (see Supplementary Material) based on only the amino acid composition in strong and weak binder groups. The comparison of Figure S3 (see Supplementary Material) with the Figure 2C shows that by using simple amino acid relative abundances for classification the accuracy of predicting the strong binders drop from 80 to 50% and also the discrimination shown in Figure 2C between strong and weak binders disappears.

The circular dichroism (CD) spectral analysis of the designed peptides indicates that there are some structural features within the strong-binding peptides: the strong binders were found to adopt a polyproline type II conformation, whereas the weak binders adopt random coil molecular conformation (J.S.Evans, personal communication).

These two sources of evidence indicate that the simple amino acid composition provides some information, but it is not adequate to represent the peptide—inorganic interactions. The crucial information comes from the sequential arrangement of the amino acid residues in a peptide, which also inherently contains the amino acid composition.

4 CONCLUSIONS

Our computational knowledge-based approach provides a general and simple methodology to quantitatively classify and design peptides according to their inorganic-binding properties. We applied our approach by generating new scoring matrices for classification and used it to design new peptides capable of binding specific substrates with predictable affinities. We experimentally characterized the designed sequences and showed excellent correspondence between prediction and experiment. As more experimental data becomes available, we can iteratively improve our approach to generate new scoring matrices and further improve the design of new peptides. Our approach is completely general and may be used to classify and design novel peptides with any arbitrary functional property, such as binding to specific organic substrates (DNA, RNA and other protein), ability to spatially organize quantum dots and to create hybrid molecular constructs, resulting in a wide variety of applications in materials science and biology.

ACKNOWLEDGEMENTS

This work was supported by grants from National Science Foundation (NSF) MRSEC Program through the University of Washington Genetically Engineered Materials Science and Engineering Center (DMR 0520567) and Turkish State Planning Organization (C.T.). This work was also supported in part by NIH grant GM068152, NSF grant DBI-0217241, a NSF CAREER award and a Searle Scholar Award (R.S.).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1994) Issues in searching molecular sequence databases. *Nat. Genet.*, **6**, 119–129.

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Amstutz,P. *et al.* (2001) In vitro display technologies: novel developments and applications. *Curr. Opin. Biotechnol.*, **12**, 400–405.

Attwood,T.K. (2000) The Babel of bioinformatics. *Science*, **27**, 471–473.

Ball,P. (2001) Life's lessons in design. *Nature*, **409**, 413–416.

Brown,S. (1997) Metal recognition by repeating polypeptides. *Nat. Biotechnol.*, **15**, 269–272.

Chang,Y. *et al.* (2006) Highly protein-resistant coatings from well-defined diblock copolymers containing sulfobetaines. *Langmuir*, **17**, 1169–1175.

Dayhoff,M.O. *et al.* (1978) A model of evolutionary change in proteins. In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*. Vol. 5, (Suppl. 3). National Biomedical Research Foundation, Washington DC, pp. 345–352.

De Yoreo,J.J. and Dove,P.M. (2004) Shaping crystals with biomolecules. *Science*, **306**, 1301–1302.

Gaskin,D.J.H. *et al.* (2000) Identification of inorganic crystal-specific sequences using phage display combinatorial library of short peptides: a feasibility study. *Biotechnol. Lett.*, **22**, 1211–1216.

Gonnet,G.H. *et al.* (1992) Exhaustive matching of the entire protein-sequence database. *Science*, **256**, 1443–1445.

Gray,J.J. (2004) The interaction of proteins with solid surfaces. *Curr. Opin. Struct. Biol.*, **14**, 110–115.

Henikoff,S. and Henikoff,J.G. (1992) Amino-acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.

Hoess,R.H. (2001) Protein design and phage display. *Chem. Rev.*, **101**, 3205–3218.

Jung,L.S. *et al.* (1998) Quantitative interpretation of the response of surface plasmon resonance sensors to absorbed films. *Langmuir*, **14**, 5636–5648.

Kann,M. *et al.* (2000) Optimization of a new score function for the detection of remote homology. *Proteins*, **41**, 498–503.

Lipman,D.J. and Pearson,W.R. (1985) Rapid and sensitive protein similarity searches. *Science*, **227**, 1435–1441.

Lowenstam,H.A. (1981) Minerals formed by organisms. *Science*, **211**, 1126–1131.

Mann,S. (1988) Molecular recognition in biomineralization. *Nature*, **332**, 119–124.

Naik,R.R. *et al.* (2002) Silica-precipitating peptides isolated from a combinatorial phage display peptide library. *J. Nanosci. Nanotechnol.*, **2**, 95–100.

Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to search for similarities in amino acid sequence of 2 proteins. *J. Mol. Biol.*, **48**, 443–453.

New England BioLabs Inc. (2006) Ph.D.-12 Phage Display Library Kit Instruction Manual, Version 2.7, p. 21.

Oren,E.E. *et al.* (2005) Metal recognition of septapeptides via polypod molecular architecture. *Nano Lett.*, **5**, 415–419.

Oren,E.E. *et al.* (2007) In silico design of inorganic binding peptides. *Nat. Mater.*, submitted.

Paine,M.L. and Snead,M.L. (1997) Protein interactions during assembly of the enamel organic extracellular matrix. *J. Bone Miner. Res.*, **12**, 221–226.

Paine,M.L. *et al.* (2001) Regulated gene expression dictates enamel structure and tooth function. *Matrix Biol.*, **20**, 273–292.

Pearson,W. and Lipman,D. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

Ratner,B. *et al.* (1996) *Biomaterials Science: Introduction to Materials in Medicine*. Academic Press, San Diego, USA.

Sano,K-I. *et al.* (2005) Specificity and biomineralization activities of Ti-binding peptide-1 (TBP-1). *Langmuir*, **21**, 3090–3095.

Sarikaya,M. (1999) Biomimetics: materials fabrication through biology. *Proc. Natl Acad. Sci. USA*, **96**, 14183–14185.

Sarikaya,M. *et al.* (2003) Molecular biomimetics: nanotechnology through biology. *Nat. Mater.*, **2**, 577–585.

Sarikaya,M. *et al.* (2004) Materials assembly and formation using engineered polypeptides. *Annu. Rev. Mater. Res.*, **34**, 373–408.

Seeman,N.C. and Belcher,A.M. (2002) Emulating biology: building nanostructures from the bottom up. *Proc. Natl Acad. Sci. USA*, **99**, 6451–6455.

Shimizu,K. *et al.* (1998) Silicatein α: cathepsin L-like protein in sponge biosilica. *Proc. Natl Acad. Sci. USA*, **95**, 6234–6238.

Smith,G.P. (1985) Filamentous fusion phage – novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315–1317.

Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.

Szunerits,S. and Boukherroub,R. (2006) Preparation and characterization of thin films of SiOx on gold substrates for surface plasmon resonance studies. *Langmuir*, **22**, 1660–1663.

Tamerler,C. *et al.* (2006) Adsorption kinetics of an engineered gold binding peptide by surface plasmon resonance spectroscopy and a quartz crystal microbalance. *Langmuir*, **22**, 7712–7718.

Thai,C.K. *et al.* (2004) Identification and characterization of Cu2O- and ZnO-binding polypeptides by Escherichia coli cell surface display: toward an understanding of metal oxide binding. *Biotechnol. Bioeng.*, **87**, 129–137.

Thompson,J.D. *et al.* (1994) CLUSTAL-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

Vogt,G. *et al.* (1995) An assessment of amino-acid exchange matrices in aligning protein sequences – the twilight zone revisited. *J. Mol. Biol.*, **249**, 816–831.

Weiner,S. and Addadi,L. (1997) Design strategies in mineralized biological materials. *J. Mater. Chem.*, **7**, 689–702.

Whaley,S.R. *et al.* (2001) Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature*, **405**, 665–668.

Wittrup,K.D. (2001) Protein engineering by cell-surface display. *Curr. Opin. Biotechnol.*, **12**, 395–399.

Zhou,J. *et al.* (2003) Orientation of adsorbed antibodies on charged surfaces by computer simulation based on a united-residue model. *Langmuir*, **19**, 3472–3478.