

KNOWLEDGE-BASED DESIGN OF INORGANIC-BINDING PEPTIDES

Ram Samudrala,¹ Ersin Emre Oren,^{1,2} Gong Cheng,¹ Jeremy Horst,^{1,3} Brady Bernard,¹ Mustafa Gungormus,² Marketa Hnilova,² Hanson Fong,² Candan Tamerler,^{2,4} and Mehmet Sarikaya^{2,4}

¹*Computational Biology Group, Department of Microbiology, School of Medicine*

²*Materials Science & Engineering,*

³*Department of Oral Biology, School of Dentistry,*

University of Washington, Seattle, Washington 98195, USA

⁴*Molecular Biology and Genetics, Istanbul Technical University, Maslak, Istanbul, 80626, Turkey*

How proteins/peptides bind to and recognize the inorganic materials with different affinities and specificities has primary importance in biomineralization, hard tissue regeneration (e.g., bone and dental tissues) and making practical materials using synthetic peptides as nucleators, growth modifiers or as control agents [1-4]. The understanding of the possible mechanism(s) of inorganic formation eventually has to be addressed for rational design and tailoring of these peptides towards specific materials systems for medical and/or practical engineering applications. Here we introduce a method that combines experimental knowledge with computational biology tools and enables design of new peptides efficiently with superior binding affinities and multiple material specificities as a potential tool for molecular engineering in materials and medicine.

In nature, proteins that perform functions that are similar to each other usually have similar sequences and structures due to biochemical, biophysical and evolutionary constraints [5]. Founded on this observation, we hypothesized that inorganic binding peptides, generated by directed evolution through in vivo selection (e.g. phage and cell surface display techniques), recognizing the same inorganic material will have alike sequences, much as evolutionarily related proteins do. Based on this hypothesis, we defined a metric that measures the sequence similarities between the experimentally selected sets of inorganic binding peptides. We derived 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 this method. The scoring matrices thus generated were used to computationally design novel peptides with enhanced binding affinities to a given inorganic material using the experimental knowledge on the existing inorganic binding peptides [6].

Experimental verifications of binding of these computationally designed peptides confirm our predictions with nearly 100% accuracy (Fig. 1). We then developed our approach and designed peptides having multiple functionalities capable of binding to quartz, hydroxyapatite, both or neither inorganic materials with proper experimental verification [6].

This procedure can be generalized to any number of inorganics as long as there is sufficient data to "train" to design a substrate specific scoring matrix. The multifunctional peptides have utility in developing surface engineering of inorganics including metals and oxides, combining several nanomaterials, nanoparticles, nanowires, or molecular erectors when conjugated to other proteins, enzymes or DNA.

We further expanded our approach to find the functional domains of natural proteins involved in hard tissue formation (e.g. magnetite-binding proteins from magnetotactic bacteria, hydroxyapatite binding proteins in mammalian bone and dental tissues, silica binding proteins from sponge spicules and skeletons of radiolarian, and, calcium carbonate-binding proteins from mollusk shells and echinoderm skeletal units). We chose amelogenin, which is the most abundant protein involved in all stages of enamel mineral formation in the tooth, as the test case as described below:

Similarity analysis: Based on the data obtained from two independent phage display selection (using 12 aa and 7 aa libraries) on hydroxyapatite, we derived two different sequence scoring matrices and used them to compare amelogenin sequence with the experimentally derived strong hydroxyapatite binding peptides. This comparison provides us two regions having high sequence similarity to the experimental strong hydroxyapatite binding peptides.

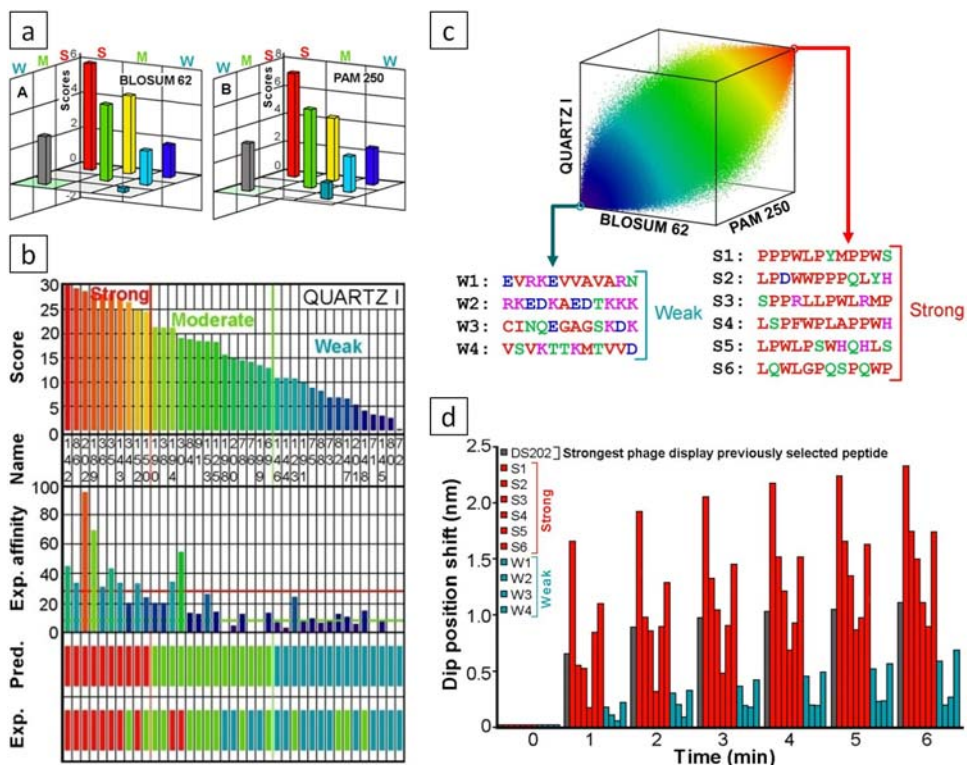


Fig. 1: Design of inorganic binding peptides. **(a)** Total similarity scores between the strong (S: red), moderate (M: green) and weak (W: teal) quartz binders calculated using the BLOSUM 62 (left) and PAM 250 (right) scoring matrices. **(b)** Classification of experimentally characterized quartz binding peptides using QUARTZ I scoring matrix. The top bar graph shows peptides ordered based on a decreasing similarity score. The corresponding experimentally determined affinities are shown in the middle bar graphs. The bottom graphs show qualitative (strong, moderate, and weak) correspondence between the predicted and experimental affinities. **(c)** Correlation between the similarity scores of random peptide sequences to strong quartz binders calculated by different scoring matrices. The sequences of designed strong (S) and weak (W) peptides are shown and the amino acids are colored according to their chemical properties (hydrophobic, acidic, basic and polar). **(d)** Experimental validation of computationally designed peptides using SPR spectroscopy analysis that measures the amount of bound peptide versus time (performed at 4 mM) for six strong (red) and four weak (teal) peptides along with DS202 (black), the strongest phage display selected peptide.

Geometrical knowledge based scoring function: To predict the naturally occurring protein/ion binding activity, we developed a knowledge based scoring function from geometrical parameters observed for inorganic single ions coordinating small molecule organics found in many diffraction structures publicly available through the protein data

bank [7-9]. We first verified this scoring function by accurately ranking the binding affinities between proteins and ions, predicting the naturally occurring ion binding sites of proteins and recovering the native conformation of protein-substrate binding (<http://protinfo.compbio.washington.edu/soak>; Fig. 2a). We then performed an exhaustive grid-based search placing ions into all solvent exposed sites in the predicted amelogenin structure. The sites are scored with a grid based exhaustive sampling algorithm, clustered using a variant of an unsupervised hierarchy clustering algorithm, and refined with Monte Carlo minimization [10].

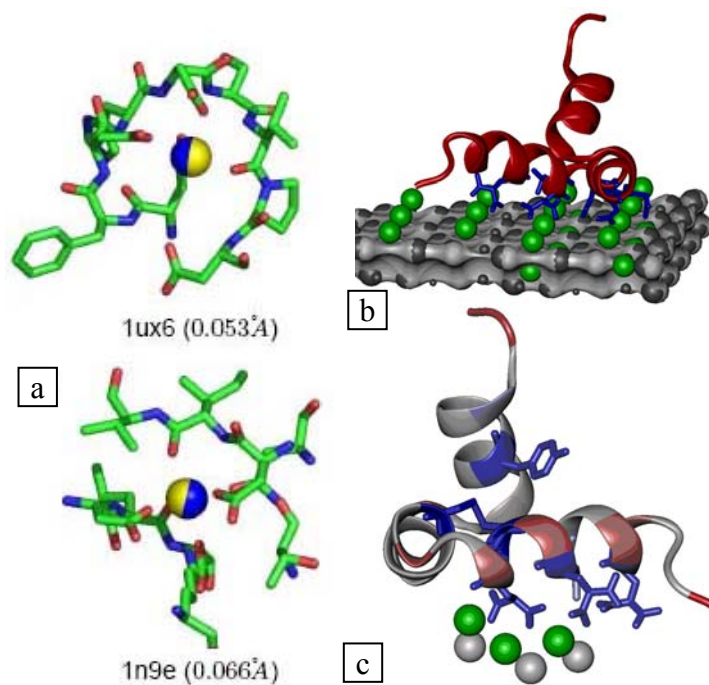


Fig. 2: Structure based predictions of protein to inorganic interactions. **(a)** The geometrical knowledge based scoring function produces high accuracy predictions (yellow spheres) of known metal ion binding sites (blue spheres) within protein structures (interfacing side chains shown with carbons: green, oxygens: red, nitrogens: blue). **(b)** The conformation sampling method is used to predict the interface of Osteocalcin (protein backbone shown in red cartoon, γ -carboxylated glutamic acids known to interface with hydroxyapatite shown in blue) to a hydroxyapatite surface (calciums: green, phosphates: dark grey, hydroxides: light grey). **(c)** Comparing calcium ions from the docked model (green spheres) to those from the known crystal structure (white spheres) indicates a match to the putative hydroxyapatite binding mechanism of Osteocalcin [11].

Conformation sampling: Hydroxyapatite is composed of phosphate and calcium ions. If we assume that the partial covalent effect dominant protein/ion interactions is similar to the effect that governing the protein/hydroxyapatite interface, the method of modeling the protein/ion interaction can be transferred to predict the protein/hydroxyapatite interactions. The orientation between the protein and the hydroxyapatite are sampled using a Metropolis-Hasting algorithm [12] with iterative simulated annealing. The conformation of the protein is treated as rigid body and perturbed in all degrees of rotational and translational freedom. The converged conformation minimas are collected and clustered (Fig. 2b and 2c).

The comparison of amelogenin with the experimentally selected hydroxyapatite binding peptides, together with the amelogenin structure predictions [13-15] and Ca ion binding regions provide us two putative functional domains that may primarily be involved in enamel biomineralization. We then isolated these domains, detected key secondary structural signatures via circular dichroism spectral analysis and finally used them in hydroxyapatite binding and biomineralization experiments. The understanding of amelogenin function in enamel formation is crucial for regenerative medicine.

ACKNOWLEDGEMENTS

This work was primarily supported by the NSF-MRSEC program at the University of Washington and also by US ARO-DURINT Program (EEO, CT, MS, RS). We also acknowledge support from NIH GM068152, NSF DBI-0217241, a NSF CAREER award, and a Searle Scholar Award (RS); and the Turkish State Planning Organization (CT).

REFERENCES

- [1] S. Brown, M. Sarikaya, E. Johnson, *J. Mol. Biol.*, **299**, 725 (2000).
- [2] S.R. Whaley, D.S. English, E.L. Hu, P.F. Barbara, A.M. Belcher, Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly *Nature*, **405**, 665 (2000).
- [3] R.R. Naik, S.J. Stringer, G. Agarwal, S.E. Jones, M.O. Stone, Silica-precipitating peptides isolated from a combinatorial phage display peptide library. *Nat. Mater.*, **1**, 169 (2002).
- [4] M. Sarikaya, C. Tamerler, A.K.Y. Jen, K. Schulten, F. Baneyx, Molecular biomimetics: nanotechnology through Biology *Nat. Mater.*, **2**, 577 (2003).
- [5] T.K. Attwood, *Science* **27**, 471 (2000).
- [6] E.E. Oren, C. Tamerler, D. Sahin, M. Hnilova, U.O.S. Seker, M. Sarikaya, R. Samudrala, A novel knowledge-based approach to design inorganic-binding peptides. *Bioinformatics*, **23**, 2816 (2007).

- [7] G. Cheng, R. Samudrala, An all-atom geometrical knowledge-based scoring function to predict protein metal ion binding sites, affinities and specificities. *To be submitted*, (2008).
- [8] B. Bernard, G. Cheng, R. Samudrala, A generalized knowledge-based scoring function for biomolecular interactions. *To be submitted*, (2008).
- [9] R. Samudrala, J. Moult, An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J. Mol. Biol.*, **275**, 893 (1998).
- [10] N. Metropolis and S. Ulam, The Monte Carlo Method, *Journal of the American Statistical Association*, **44**, 335 (1949).
- [11] Q.Q. Hoang, F. Sicheri, A.J. Howard, D.S. Yang, Bone recognition mechanism of porcine osteocalcin from crystal structure. *Nature*, **425**, 977 (2003).
- [12] W.K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", *Biometrika*, **57**(1), 97 (1970).
- [13] L.-H. Hung, S.-C. Ngan, T. Liu, R. Samudrala, PROTINFO: New algorithms for enhanced protein structure prediction. *Nucleic Acids Research* **33**, W77 (2005).
- [14] L.-H. Hung, R. Samudrala, PROTINFO: Secondary and tertiary protein structure prediction. *Nucleic Acids Research*, **31**, 3296 (2003).
- [15] Y. Zhang, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40 (2008).