

Copyright
by
Erin Elaine Gooch
2004

The Dissertation Committee for Erin Elaine Gooch Certifies that this is the approved version of the following dissertation:

Biomolecule Interactions on Calcium Carbonate and Stoichiometrically Similar Biomedical, Optical and Electronic Materials

Committee:

Angela M. Belcher, Supervisor

Karen Browning

David Hoffman

Brian Korgel

John McDevitt

**Biomolecule Interactions on Calcium Carbonate and
Stoichiometrically Similar Biomedical, Optical and Electronic
Materials**

by

Erin Elaine Gooch, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2004

Dedication

“If I have seen further, it by standing on the shoulders of giants.”

- Sir Isaac Newton

To: Mom, Dad and Seth,
the giants upon whose shoulders I have had the privilege of standing.

Acknowledgements

Thank you to all who have supported me through my tenure as a graduate student: Dr. Angela Belcher, who gave me the opportunity to work in this exciting field and the freedom to explore the many avenues of the science; Harshal Gupta, my diligent assistant, whose help with experiments is immensely appreciated; Dr. Glenn Martyna, whose peptide modeling program gave me the answers to the questions *I* was seeking and an introduction to the exciting field of computational biochemistry; Dr. Andrew Hayhurst, Roz Sweeney, and Ioana Pavel, who helped me learn the many details of protein engineering; Dr. Klaus Linse and Dr. Esther Ryan, who helped me with my many mass spectrometry experiments; Dr. David Hoffman, who agreed to run my peptide through his NMR even though he knew the results wouldn't tell me much and who spent much time helping me decipher my data; Nina Antikainen and Lindsay Pell Burns, my wonderful friends, who helped me so much with not only science but also life! A special thank you to my families, the Gooches, Carruthers, and Nicols, whose love and support means the world to me!

**Biomolecule Interactions on Calcium Carbonate and
Stoichiometrically Similar Biomedical, Optical and Electronic
Materials**

Publication No. _____

Erin Elaine Gooch, Ph.D.

The University of Texas at Austin, 2004

Supervisor: Angela M. Belcher

A combinatorial approach has been successfully used to characterize peptides that bind to four different surfaces of CaCO_3 and six different oxide substrates that are chemically and stoichiometrically similar to CaCO_3 . A standard screening and a single substrate screening method were employed. Both methods used a bacteriophage combinatorial library with a complexity of 10^9 different random sequences. While geological (104) calcite screenings do not seem to show strong consensus, peptides screened against geological aragonite exhibit extraordinary consensus. Overall, peptides screened against aragonite were highly basic. Peptides screened against (110) geological aragonite show [LAIVG]P[WF][RKH] and triple [RKH] patterns. The most significant binding peptide, A21 (LPPWKHKTSOVA) was found in 4 separate screenings. The (110)

geological aragonite sequences are highly enriched in prolines. Coincidentally, peptides screened against the optical material LiNbO_3 +Z show patterns similar to those found in the aragonite screenings. Other substrates that did not exhibit strong consensus include powdered LiNbO_3 , powdered BaTiO_3 , powdered PbTiO_3 , single crystal LiNbO_3 -Z, and hydroxyapatite. A database and analysis program was written to catalog and evaluate the statistical significance of the many sequences. One peptide that was determined to be extremely significant in binding to aragonite, A20 (LPKWQERQMLSA), was modeled using umbrella sampling molecular dynamics techniques as well as analyzed by NMR. It was determined that there is a good probability that the peptide's conformation is helical, but it is able to interconvert between α -helical, 3-10 helical and extended conformations fast enough that a stable secondary structure is not detectable by NMR. Engineered phage were constructed to display these significant peptides on the pVIII surface, increasing the expression level and the long range order in the hopes of building an aragonite nucleation surface. Hybrid organic-inorganic materials were grown on the phage resulting in a mixture of the three forms of CaCO_3 : calcite, vaterite, and aragonite. In addition, hybrid materials were grown on the optical waveguide material LiNbO_3 . Doubly engineered phage were shown to bind preferentially to the LiNbO_3 +Z surface through the interaction of the A1 displayed peptide as well as nucleate semiconducting ZnS particles via the A7 peptide selected previously against ZnS.

Table of Contents

List of Tables.....	xii
List of Figures	xxvi
Chapter 1: Introduction	1
1.1: Biomineralization.....	1
1.1.1: Biomineralization to Biomimetic Crystal Growth.....	2
1.1.2: Natural Nacre-Associated Proteins	6
1.2: Scope of Project	8
1.2.1: Combinatorial Approach.....	9
1.2.2: Characterization of Peptides	10
1.2.3: Peptide Structure Determination	10
1.2.4: Engineering of Peptides	11
1.2.5: Hybrid Organic-Inorganic Materials.....	11
1.3: Characterization of Materials.....	11
1.3.1: Geological Calcite.....	13
1.3.2: Geological Aragonite	15
1.3.3: Nacreous Aragonite.....	18
1.3.4: Hydroxyapatite.....	19
1.3.5: Barium Titanate.....	21
1.3.6: Lead Titanate.....	22
1.3.7: Lithium Niobate	24
References:.....	26
Chapter 2: Peptide Selection Using Phage Display	35
2.1: Phage Display Theory	35
2.1.1: pIII Structure and Function	36
2.1.2: Phage Assembly in Bacteria.....	40
2.1.3: Commercial Phage Display System – New England Biolabs..	42

2.2: Phage Display Experimental	44
2.2.1: Standard Elution Method	45
2.2.2: Single Substrate Screening Method	47
2.2.3: Determining Random Phage Peptide Sequences	48
2.3: Sequence Analysis	48
2.3.1: Amino Acid Percents by Position	49
2.3.2: Pattern Discovery Algorithm – Teiresias by IBM	49
2.3.3: Statistical Significance of Discovered Patterns.....	51
2.3.4: Helical wheel.....	52
2.3.5: Homology to Biomineral Associated Proteins.....	53
References:.....	54
Chapter 3: Phage Display Results: Calcium Carbonate	58
3.1: Geological Aragonite	61
3.1.1: Overall Geological Aragonite Patterns from Teiresias	61
3.1.2: Geological Aragonite Analysis by Library	64
3.1.3: Effect of Screening pH for Geological Aragonite Sequences..	69
3.1.4: Face-Specific Elution for Geological Aragonite.....	74
3.1.5: Effect of Elution Method for Geological Aragonite Sequences.....	80
3.1.6: Helical Wheel for Statistically Significant Aragonite Peptides	86
3.2: Nacreous Aragonite.....	90
3.2.1: Nacreous Aragonite Analysis by Library.....	91
3.2.2: Similarities to Geological Aragonite 001 Sequences.....	95
3.3: Geological Calcite	96
3.3.1: Effect of Screening pH on Calcite Sequences.....	97
3.3.2: Effect of Elution Method on Calcite Sequences	99
3.4: Homology to Natural Biomineralization Proteins.....	103
3.5: Conclusions	106

References:.....	107
Chapter 4: Phage Display Results: Similar Materials	110
4.1 Hydroxyapatite	111
4.2: Barium Titanate, Lead Titanate and Lithium Niobate Powders	114
4.3: Lithium Niobate Single Crystal	120
4.4: Conclusions.....	126
Chapter 5: Peptide Structure Determination	128
5.1: Choice of Model Peptide and Surface.....	128
5.2: Umbrella Sampling MD Method	128
5.2.1: Theory	128
5.2.2: Coordinate System and Peptide Set Up	129
5.2.3: Molecular Dynamics Parameters	131
5.2.4: Modeling Results	132
5.3: Peptide NMR.....	134
5.4: Conclusions	137
References:.....	138
Chapter 6: Protein Engineering and Materials Growth.....	140
6.1: Display of Peptide on M13 Phage Outer Coat Protein	140
6.2: Phage-Mediated Calcium Carbonate Crystal Growth.....	146
6.2.1: pVIII Mediated Crystal Growth.....	147
6.3: Semiconducting Film Growth on Optical Waveguide Material	153
6.3.1: Double Display Phage System	154
6.3.2: Phage Film Growth	155
6.4: Conclusions	158
References:.....	160
Chapter 7: Conclusions	162
7.1: Peptide Screening, Pattern Analysis and Homology.....	162
7.2: Structure Determination.....	163

7.3: Protein Engineering.....	165
7.4: Peptide Mediated Crystal Growth.....	165
Appendix A: Sequences	167
Appendix B: Database and Analysis Program Source Codes	197
B.1: Query.pl.....	197
B.2: ForwardSequences.pl.....	211
B.3: CalcProbability.pl.....	213
B.4: CountEqPatterns.pl.....	223
B.5: ParseNacreProteins.pl.....	226
Appendix C: NMR Spectra of the A20 peptide	229
Bibliography.....	233
Vita	247

List of Tables

Table 1.1: Nacre-associated proteins that have been sequenced and their functions or homologies.	7
Table 1.2: Unit cell parameters for calcite (PDF#86-0174).	13
Table 1.3: Unit cell parameters of aragonite [55-57].	16
Table 1.4: Unit cell parameters for hydroxyapatite (PDF#09-0432).	19
Table 1.5: Unit cell parameters for BaTiO ₃ (PDF#34-0129).	22
Table 1.6: Unit cell parameters of lead titanate (PDF#48-0105).	23
Table 1.7: Lithium niobate unit cell parameters (PDF#20-0631).	24
Table 3.1: Summary of the screening conditions for the calcium carbonate materials. The dots indicate selection conditions under which these materials were screened. Geological aragonite was screened as a whole and with separate faces (001 and 110).	60
Table 3.2: Overall short aragonite patterns found by IBM's Teiresias pattern recognition program. The significance is calculated from the observed frequencies of amino acids in the NEB libraries and the number (#) of times the pattern was found.	62
Table 3.3: Long patterns for overall aragonite found by IBM's Teiresias pattern recognition program. The significance is calculated from the observed frequencies of amino acids in the NEB libraries and the number (#) of times the pattern was found.	63

Table 3.4: Number of multiple sequences pulled out of each separate screening where A20 and A21 were found. The percentage is the percentage of total PhD12 sequences screened from geological aragonite.....	63
Table 3.5: Summary of geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the <i>screenings</i> are tabulated next to the overall <i>library</i> average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.....	65
Table 3.6: Highest percentages for each position in each of the three libraries screened against geological aragonite. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, pralines are italicized, aromatic groups are underlined.....	66
Table 3.7: High significance patterns for each library for geological aragonite found by IBM's Teiresias pattern recognition program. The significance is calculated from the observed frequencies of amino acids in the NEB libraries and the number (#) of times the pattern was found.....	68

Table 3.8: Summary of PhD12 geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the screenings are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted. 71

Table 3.9: Highest percentages for each PhD12 position for the sequences screened against geological aragonite at different pHs. The average of all of the geological aragonite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined. 72

Table 3.10: Top patterns for sequences from the PhD12 library screening of geological aragonite at pH 7.5 and pH 8.5. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found. . 74

Table 3.11: Top ten patterns for sequences from the geological aragonite (110) screenings. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found. 75

Table 3.12: Sequences found multiple times in the same geological aragonite (110) screening and the number of times found. Std is the standard acid elution screening and SSS is the single substrate screening. The percentage is calculated using the total number of sequences selected against (110) geological aragonite. 76

Table 3.13: Summary of the (110) and (001) geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted. 77

Table 3.14: Top ten patterns for sequences from the geological aragonite 001 screenings. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias. 78

Table 3.15: Sequences found multiple times in the same geological aragonite 001 screening, the number of times found, and the percentage of the total sequences selected against (001) geological aragonite. 79

Table 3.16: Summary of the different elution methods for the PhD12 geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the <i>screenings</i> are tabulated next to the overall <i>library</i> average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.....	82
Table 3.17: Highest percentages for each PhD12 position for the sequences screened against geological aragonite using the three different elution methods. The average of all of the geological aragonite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.....	84
Table 3.18: Patterns from the PhD12 geological aragonite sequences eluted with acid, base and bacteria, respectively. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found. .	86
Table 3.19: All patterns from the nacreous aragonite sequences. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.	90

Table 3.20: Summary of nacreous aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted. 92

Table 3.21: Highest percentages for each position for the sequences screened against nacreous aragonite for each library. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined. 93

Table 3.22: Patterns from the nacreous aragonite sequences. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found. . 94

Table 3.23: Patterns from the combined geological aragonite (001) surface and nacreous aragonite sequences. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found..... 95

Table 3.24: Summary of geological calcite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted. 98

Table 3.25: Highest percentages for each PhD12 position for the sequences screened against geological calcite at pH 7.5 and pH 8.5. The average of all of the geological calcite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined..... 99

Table 3.26: Summary of the different elution methods of the PhD12 geological calcite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted..... 101

Table 3.27: Highest percentages for each PhD12 position for the sequences screened against geological calcite and eluted with acid or base. The average of all of the geological calcite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined..... 102

Table 3.28: Patterns from the geological calcite sequences. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.....	103
Table 3.29: Protein fragments that have been shown to bind to calcite and hinder mineral overgrowth. Charged residues are in bold, hydroxyl residues are in gray italics and amides are underlined....	104
Table 3.30: The number of patterns common to the aragonite screening results found for each known nacre protein sequence. The highlighted proteins have numerous patterns in common with the screening population.	105
Table 4.1: Summary of the screening conditions for the biomedical (hydroxyapatite – HA), electronic (PbTiO ₃ and BaTiO ₃) and optical (LiNbO ₃) materials. The dots indicate selection conditions under which these materials were screened. LiNbO ₃ single crystal is poled in the Z direction, giving two different screening surfaces.	110
Table 4.2: Highest percentages for each position in each library for the sequences screened against hydroxyapatite. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.	112

Table 4.3: Summary of hydroxyapatite screenings broken down to the amino acid level. The percentage of amino acids in the <i>screenings</i> are tabulated next to the overall <i>library</i> average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.....	113
Table 4.4: Patterns from the hydroxyapatite sequences. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.....	114
Table 4.5: Summary of the PhD7C sequences for the powdered BaTiO ₃ , PbTiO ₃ and LiNbO ₃ screenings broken down to the amino acid level. The percentage of amino acids in the <i>screenings</i> are tabulated next to the overall <i>library</i> average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.....	117
Table 4.6: Highest percentages for each PhD7C position for the sequences screened against powdered BaTiO ₃ , PbTiO ₃ and LiNbO ₃ . Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.....	118

Table 4.7: Patterns from the BaTiO₃, PbTiO₃ and LiNbO₃ powder sequences.
The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias. 120

Table 4.8: Summary of the different elution methods for the PhD12 single crystal LiNbO₃ -Z and +Z screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences (> ±3.0%) from the expected library percentages are highlighted. 122

Table 4.9: Highest percentages for each PhD12 position for the sequences screened against single crystal LiNbO₃ A) -Z surface and B) +Z surface. The averages of all PhD12 single crystal LiNbO₃ -Z and +Z surface sequences are included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined. 124

Table 4.10: Patterns from the -Z and +Z LiNbO₃ single crystal sequences.
The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias. 125

Table 4.11: LiNbO ₃ sequences show similarity to geological aragonite sequence motifs. Charged residues are bold, prolines are italicized, Aromatic residues are underlined.....	126
Table 5.1: Proton NMR chemical shifts for the A20 synthetic peptide residues.....	136
Table 6.1: A1 sequence compared to similar high significance LiNbO ₃ patterns. A7 sequence screened against and shown to capture ZnS [7].....	155
Table A.1: PhD12 sequences from geological aragonite screened at pH 7.5 and eluted with acid.....	167
Table A.2: PhD12 sequences from geological aragonite screened at pH 7.5 and eluted with acid; screening 2.....	168
Table A.3: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with acid.....	169
Table A.3 Continued: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with acid.....	170
Table A.4: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with base.....	170
Table A.5: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with bacteria.....	171
Table A.6: PhD7 sequences from geological aragonite screened at pH 8.5 and eluted with acid.....	172

Table A.7: PhD7C sequences from geological aragonite screened at pH 8.5 and eluted with acid.....	173
Table A.8: PhD12 sequences from the 110 face of geological aragonite screened at pH 8.5 and eluted with bacteria.....	174
Table A.9: PhD12 sequences from the 110 face of geological aragonite screened at pH 8.5 and eluted with acid.....	175
Table A.10: PhD12 sequences from the 001 face of geological aragonite screened at pH 8.5 and eluted with bacteria.....	176
Table A.12: PhD12 sequences from the 001 face of geological aragonite screened at pH 8.5 and eluted with acid.....	177
Table A.13: PhD12 sequences from geological calcite screened at pH 7.5 and eluted with acid.	178
Table A.14: PhD12 sequences from geological calcite screened at pH 8.5 and eluted with acid.	179
Table A.15: PhD12 sequences from geological calcite screened at pH 8.5 and eluted with base.....	180
Table A.16: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 1.	180
Table A.17: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria.; screening 2	181
Table A.18: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 3.	181

Table A.19: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 4.	182
Table A.20: PhD7 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria.	183
Table A.21: PhD7C sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 1.	183
Table A.22: PhD7C sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 2.	184
Table A.23: PhD12 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria; screening 1.	185
Table A.24: PhD12 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria; screening 2.	186
Table A.25: PhD7 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria.	187
Table A.25 Continued: PhD7 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria.	188
Table A.26: PhD7 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria.	189
Table A.27: PhD7C sequences from barium titanate screened at pH 7.5 and eluted with acid.	190
Table A.28: PhD7C sequences from lead titanate screened at pH 7.5 and eluted with acid.	191

Table A.29: PhD7C sequences from lithium niobate screened at pH 7.5 and eluted with acid.	192
Table A.30: PhD12 sequences from single crystal lithium niobate -Z screened at pH 7.5 and eluted with acid.	193
Table A.31: PhD12 sequences from single crystal lithium niobate -Z screened at pH 7.5 and eluted with bacteria; screening 1.	193
Table A.32: PhD12 sequences from single crystal lithium niobate -Z screened at pH 7.5 and eluted with bacteria; screening 2.	194
Table A.33: PhD12 sequences from single crystal lithium niobate +Z screened at pH 7.5 and eluted with acid.	195
Table A.34: PhD12 sequences from single crystal lithium niobate +Z screened at pH 7.5 and eluted with bacteria.	196

List of Figures

Figure 1.1: SEM images of aragonitic nacre tablets. A) View from the top B) View from the side C) View of organic material between tablets.	5
Figure 1.2: A) Geological calcite rhombohedron [50]. B) Calcite (104) surface; calcium is blue, carbon is grey, oxygen is red (Molecular Simulations, Inc.).	14
Figure 1.3: XRD of calcite showing orientation along the (104) face.	15
Figure 1.4: XRD of the non-twinned face of the geological aragonite crystal, showing 002 and 004 parallel planes.	17
Figure 1.5: XRD of the twinned face of the aragonite crystal, showing 110, 220, and 330 planes.	17
Figure 1.6: A.) Illustration of the aragonite (001) surface, calcium is blue, carbon is grey, oxygen is red (Molecular Simulations, Inc.). B.) SEM image of nacreous aragonite tablets; {001} surface. C.) Illustration of the aragonite (001) surface (Molecular Simulations, Inc.). D.) Image of geological aragonite [58]; there are 6 pseudo-hexagonally-twinned {110} faces parallel to a single {001} surface.	18
Figure 1.7: XRD of 001 oriented aragonite from black lipped oyster nacre.	19
Figure 1.8: XRD of sintered polycrystalline hydroxyapatite pellet.	21
Figure 1.9: XRD of single crystal LiNbO_3	25
Figure 1.10: XRD of powdered polycrystalline LiNbO_3	25

Figure 2.1: Diagram of the M13 filamentous phage coat proteins and single-stranded circular DNA, with gene numbers marked [2].	35
Figure 2.2: A.) Diagram of the phage pIII infecting its host <i>E. coli</i> . D2 binds the <i>E. coli</i> f-pilus, freeing D1, which interacts with the TolA-DIII B.) Ribbon representation [11] of the pIII D1-D2 “horseshoe,” N-terminal of D1 is green; C-terminal of D2 is red; f-pilus binding residues on D2 are dark blue; TolA binding residues on D1 are orange; (PBD IDs: 1g3p, 2g3p) [8, 9, 12, 13].	37
Figure 2.3: Diagram of the pIII showing the domains and functions of each. Note that the D2 interacts with the TolA binding site on D1.....	42
Figure 2.4: Representations of the three types of New England Biolabs fused peptide library configurations and lengths.	43
Figure 2.5: Evolutionary biopanning protocol steps [32].	45
Figure 2.6: Titer plates showing phage plaques from phage still bound to A) calcite and B) aragonite after acid elution.....	47
Figure 3.1: Helical wheel representations of three significant and interesting peptides selected against geological aragonite. The sequences are presented as α -helices and 3-10 helices. Residue key: bold = basic, underlined = acidic.....	89
Figure 5.1: The A20 peptide in a helical conformation. r is the average distance from the i to $i + 4$ residue.....	131

Figure 5.2: A) 2-D plot of the mean-standard deviation probability distribution function $P(r,\sigma;N,\beta)$ of the A20 peptide *in vacuo*. Maximum $\approx r_{bar} = 4.2$, $\sigma = 1.6$. B): Sampling of different A20 peptide conformers at r and σ values near the highest probability peak. 133

Figure 5.3: A) 2-D plot of the mean-standard deviation probability distribution function $P(r,\sigma;N,P_{ext},\beta)$ of the A20 peptide in computer water. B) Sampling of different A20 peptide conformers at r and σ values near the three highest probability peaks in solution. 134

Figure 6.1: Phage diagram with the molecular structure [4] of the phage pVIII (PDB ID: 2CPB) shown with the positively charged C-terminal (lysine residues in red) helix facing the negatively charged phage DNA. Engineering of the pVIII is accomplished by fusing a peptide onto the N-terminus. 141

Figure 6.2: pMoPac33 vector (~6760 bp) used for cloning hybrid peptide fusions on the N-terminus of the pVIII. The insert region is shown in black and the gene VIII is striped. 143

Figure 6.3: DNA sequences inserted into the pMopac33 vector for the two aragonite peptides A20 and A1. Both sequences contain the leading Sfi I site and pesidase cleavage site before the peptide codons. 144

Figure 6.4: A) High resolution TEM image of crystals grown with the A20-pVIII phage. B) FFT image of selected area in A, showing the various planes of vaterite.....	148
Figure 6.5: A) High resolution TEM image of crystals grown with the A20-pVIII phage. B) FFT image of selected area in A, showing the various planes of aragonite.....	149
Figure 6.6: A) High resolution TEM image of crystals grown with the A1-pVIII phage. B) FFT image of selected area in A, showing the various planes of aragonite.....	151
Figure 6.7: A) High resolution TEM image of crystals grown without phage. B) FFT image of selected area in A, showing the (006) plane of calcite.	152
Figure 6.8: Diagram of a LiNbO ₃ optical waveguide with semiconductor nanocrystal emitters linked by bifunctional peptides, which bind and nucleate LiNbO ₃ and ZnS, respectively.	154
Figure 6.9: SEM image of hybrid materials growth using doubly displayed phage. Control phage (far left) and experiment phage (center and far right) were exposed to both LiNbO ₃ +Z (row 1) and LiNbO ₃ -Z (row 2). ZnS (white crystals) were grown using the A7 peptide to mediate growth. All scale bars 1 μm.....	157
Figure C.1: 2-dimensional Nuclear Overhauser Effect spectrum (NOESY) of the A20 peptide.	229
Figure C.2: 2-dimensional Correlation Spectrum (COSY) of the A20 peptide.	230

Figure C.3: 2-dimensional Total Correlated Spectrum (TOCSY) of the A20 peptide.....	231
Figure C.4: Rotating-frame Overhauser Enhancement Spectrum (ROESY) of the A20 peptide.....	232

Chapter 1: Introduction

1.1: BIOMINERALIZATION

Biom mineralization is the ability of many organisms to build inorganic structural elements into or around their bodies. The organisms that display this ability range from organisms as simple as single celled coccolithophorids, enclosed in their intricate calcite cages, to animals and human beings, supported by strong bones and nourished by food ground down by sharp teeth, both made up of biom mineralized hydroxyapatite. Although nature has perfected its mechanism for growing biologically induced minerals, this molecular mechanism remains elusive to the scientists who have been studying these processes for years.

The study of biom mineralization is important for several reasons. On the most basic level, we will gain a better understanding of these remarkable organisms and how they function in their respective environments. In addition, these materials offer several desirable properties, such as fracture toughness, self-correction, uniform single crystal synthesis, as well as size, shape and polymorph control at the nanometer “building block” level all the way up to the macroscopic “heterostructure” level. Furthermore, these materials are synthesized at ambient pressure and temperature, a feat not yet matched by typical materials syntheses. Harnessing and manipulating the specific recognition of these proteins can lead to self-assembling tethering of molecules or particles to inorganic substrates or growth of specific polymorphs, orientations or shapes of inorganic crystals.

This dissertation will review the properties of these biominerals as well as the organic biomolecules that have been shown to be essential in the formation of these biominerals. In addition, a novel method of determining and characterizing the interactions between these biomolecules and biominerals will be presented as well as methods used to engineer biomolecules to mimic natural biomineralization processes.

1.1.1: Biomineralization to Biomimetic Crystal Growth

Many biomineralizing systems have been studied in order to determine the exact mechanisms of crystal polymorph nucleation, size and shape control, and regulation of crystal orientation. In many of these systems, biomolecules or mineralization vesicles have been discovered to be critical to inorganic crystal growth. In these studies, it has been shown that biomineralization may indeed be a complicated growth process with several different molecules contributing different functions in a highly orchestrated mineralization scheme. Further studies have even revealed that synthetic molecules can be designed to perform some of these features of natural biomineralizing systems. This is known as biomimetic crystal growth or crystal binding and is an important goal in inorganic science as nature's ability to control many features of crystal growth at the nanometer all the way up to the centimeter scale is extraordinary. This work presented in this dissertation also attempts to discover biomolecules that have potential as biomimetic mineral nucleators, inhibitors, and stabilizers.

One example of natural biomineralization is the ability of organisms ranging from bacteria to humans to make magnetic minerals, usually magnetite,

Fe_3O_4 or greigite, Fe_3S_4 , by either biologically induced precipitation or by organically controlled biomineralization [1]. Although there is an extremely diverse population of magnetic material-producing organisms, the most thoroughly studied are the magnetotactic bacteria species. These bacteria can produce 35-120 nm long, single-domain permanent magnetic particles, allowing them to migrate specifically along magnetic field lines [2].

Although the organic phase responsible for magnetite formation in magnetotactic bacteria remains vague, crystal growth is thought to be directed by specific proteins in the phospholipid membrane of a specialized mineralization vesicle, the magnetosome. It is known that these bacteria produce low molecular weight chelating molecules, siderophores, which bind and solubilize the typically insoluble Fe(III) [2]. Furthermore, these siderophores are transported between the bacterial membranes by way of siderophore binding proteins [3]. This exemplifies nature's use of different specialized proteins to recruit precursors, move them across membranes, and create a specialized environment for optimal crystal growth.

There are several examples of biomimetic iron oxide crystal growth. Nanocrystalline iron oxide has been successfully templated within the hollow channel of the lumazine synthase enzyme [4]. In addition, a biomimetic matrix of polyvinyl alcohol has been used as a substrate to grow iron oxide crystals of varying morphology [5].

Another example of highly-investigated, biologically-controlled inorganic materials growth is marine silica (SiO_2) formation. Natural silica deposition in

diatoms, specifically the *Cylindrotheca fusiformis*, has been linked to several proteins produced by the organism. While there are several families of proteins found in diatom cell walls, it is the silaffins [6, 7], low molecular weight, hydrofluoric acid-extractable proteins that appear to affect silica precipitation. Silaffins are denoted by several hydroxyl amino acids (serines and tyrosines) as well as N-methylated lysines, lending fodder to the theory that the peptide matches the chemistry of the inorganic material. In fact, the silaffins precipitate silica from silicic acid much faster and abundantly than in the absence of silaffins [8].

As an example of biomimetic silica precipitation, a 19-amino acid synthetic peptide has been made using one of the silaffin repeat segments as its base [8]. It was found that this peptide, even without the lysine modifications found in natural silaffins, could indeed precipitate as much silica as the silaffin-1A.

Another silica-associated protein found in sponge spicules, silicatein, has been found to hydrolyze the precursor tetraethoxysilane and condense silica at ambient temperatures and pressures [9]. This protein was sequenced and analyzed for possible reactive functionalities [10]. Subsequently, a specially designed block copolymer was synthesized and shown to have the ability to function in the hydrolysis of the silica precursor and condensation of silica [11].

The most relevant biomineralization example for this research is the extensive study of mollusk shell architecture. Mollusks, which build their shells out of calcium carbonate, have the remarkable ability to deposit two distinct

polymorphs of calcium carbonate, calcite and aragonite, in separate layers. The minerals in each of these layers are deposited in a specific orientation, size and shape [12].

In the well-studied red abalone (*Haliotis rufescens*), the calcite mineral layer is prismatic in morphology and is oriented in the 001 direction [13]. The organic molecules of the aragonite layer consist of, most probably, silk fibronin-like molecules, β -chitin and soluble acidic glycoproteins [14], although the majority of the components have yet to be precisely identified. Subsequently, the brick and mortar structure of the aragonite layer is built (Figure 1.1). Between each aragonite tablet as well as occluded within the crystals, there are several acidic glycoproteins (Figure 1.1 C). The aragonite tablets are also oriented in the 001 direction (along the c-axis). This brick and mortar nacreous aragonite is a microlaminate material that imparts the material a fracture resistance 3000 times greater than geological aragonite [15].

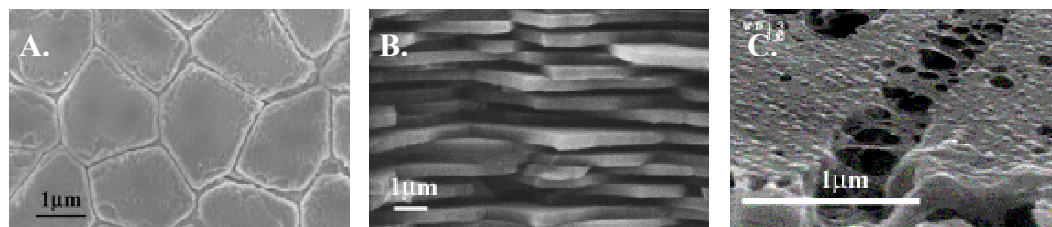


Figure 1.1: SEM images of aragonitic nacre tablets. A) View from the top B) View from the side C) View of organic material between tablets.

In contrast, minerals formed by solution precipitation are predominantly the thermodynamically stable phase: rhombohedral calcite crystals with 104 orientation. Aragonite is a meta-stable phase of calcium carbonate, usually only

formed at high temperatures and high pressures and is typically elongated in the c-axis direction. In mollusks, aragonite is formed at ambient temperatures and pressures. In addition, the aragonite tablet growth is elongated in the ab-plane and arrested in the c-direction [16], producing tablets about 10-15 μm in diameter and 500 nm thick.

In both the calcite and aragonite layers, there are proteins associated with the mineral phase. These proteins, once extracted, have been shown to control the phase and morphology of the mineral precipitated [17, 18]. This ability to preferentially precipitate a specific phase of inorganic material without high pressures, temperatures, or other additives, such as magnesium, is a materials synthesis feat. The total weight percent of these mineralizing organic material is only ~2-5% [14]. In the aragonitic layers, the sheets between the mineral tablets is about 50 nm thick and contains a core of insoluble material, made of chitin and insoluble protein [19]. Adsorbed to this core organic material are presumably acidic glycoproteins responsible for the growth of the aragonite [14].

1.1.2: Natural Nacre-Associated Proteins

Several nacre-associated proteins have been extracted from aragonitic mollusk shell layers and sequenced using cDNA technology [20-26] or N-terminal sequencing [27]. From these studies, nacre-associated proteins have been found to have a wide range of functions. Some studies indicate that these proteins most likely work in concert with each other in order to crystallize the meta-stable aragonite mineral [17, 18, 26]. These proteins and functions are listed in Table 1.1. In addition, it is feasible that aragonite nucleation and growth

proteins may also have aragonite recognition/adsorption domains, matrix recognition/adsorption domains as well as aragonite nucleating domains.

GI	Name	Organism	Functions
12655873	AP24 precursor [28]	<i>H. rufescens</i>	Frustrates calcite growth
12655875	AP7 precursor [28]	<i>H. rufescens</i>	Frustrates calcite growth
18202416	Perlustrin [27, 29]	<i>H. laevigata</i>	Insulin-like growth factor binding activity
25090912	Perlucin [27, 30]	<i>H. laevigata</i>	C-type lectin domain
13309811	Mucoperlin [25]	<i>P. nobilis</i>	Mucin-like protein, inhibits CaCO ₃
2204081	MSI60 [23]	<i>P. fucata</i>	Insoluble nacre sheet protein
7511706	Lustrin A [22]	<i>H. rufescens</i>	Frustrates calcite growth, elastomeric protein
1480031	Nacrein [24]	<i>P. fucata</i>	Carbonic anhydrase domain (HCO ₃ ⁻ production) & induces CaCO ₃ crystallization
20372975	Nacrein [31]	<i>T. marmoratus</i>	Homologous to Nacrein
11362266	N66 [26]	<i>P. maxima</i>	Homologous to Nacrein
5821239	N16#1 [20]	<i>P. fucata</i>	Induces CaCO ₃ crystallization
5821243	N16#2 [32]	<i>P. fucata</i>	Homologous to N16
5821245	N16#3 [33]	<i>P. fucata</i>	Homologous to N16
5821247	N16#4 [34]	<i>P. fucata</i>	Homologous to N16
5821249	N16#5 [35]	<i>P. fucata</i>	Homologous to N16
5821253	N16#7 [36]	<i>P. fucata</i>	Homologous to N16
11362265	N14 [26]	<i>P. maxima</i>	Homologous to N16
4519446	Pearlin [20, 21]	<i>P. fucata</i>	Homologous to N14, N16

Table 1.1: Nacre-associated proteins that have been sequenced and their functions or homologies.

Biomimetic calcium carbonate growth has been accomplished in a few different ways. Various orientations of calcite have been growth on self-assembled monolayers, merely by changing the functional group [37-39]. Patterned single crystal calcite has been grown using these specialized self-assembled monolayers [37]. Amorphous calcium carbonate is nucleated using a

mixture of functional groups, and single crystal calcite is nucleated at a single point, crystallizing the amorphous calcium carbonate into calcite as the crystallization front advances [40]. Polymer matrices have been developed that can chemically induce oriented (001) calcite rather than the thermodynamically induced (104) calcite [41]. In addition, it has been shown that calcite growth can be modified using specially designed chemical [42] and peptide [43] additives.

Biomimetic aragonite growth, a much more difficult task, has been accomplished using a biomimetic polypeptide that mimics the acidic biomolecules found in natural aragonite growth [44]. It has been shown that the secondary structure, β -sheet rather than α -helix, as well as specific amino acids, aspartic acid rather than glutamic acid, are important in the preferential growth of aragonite over calcite.

1.2: SCOPE OF PROJECT

There are five facets to this project. The first aspect is the screening of the inorganic materials, using a combinatorial method, to find peptides that bind to each of the material faces. Secondly, a database was built by the author to store and analyze the discovered sequences for statistically significant patterns. Thirdly, a structural analysis of a very statistically significant peptide was undertaken to determine the possibility of any peptide secondary structure that may contribute to binding. Fourthly, using statistically significant peptide sequences, the phage coat proteins were engineered to express the sequence of interest 5-20 times more than the original combinatorial library phage. Finally,

these engineered phage were used to grow hybrid organic-inorganic materials with a range of functions.

1.2.1: Combinatorial Approach

Using a well-characterized combinatorial phage display system, this research set out to pioneer a new application of this technique to screen for biological interactions with inorganic materials. Phage display is a thoroughly tested method of screening for potential biological interactions, such as molecule-receptor, protein-inhibitor, and other peptide-molecule or peptide-peptide interactions. Most of these studies are done with the intention to find molecules, proteins, or peptides that may be able to act as drugs to inhibit destructive biological processes or enhance healing processes. The success of these phage display screenings has lead our group to postulate that they might also be successful in screening against non-biological substrates.

Phage display libraries are designed to harness the efficiency of massive parallel experimentation. Phage, also known as bacteriophage, are viruses that infect bacteria. Specifically, the M13 filamentous phage is about a micron long, 6 nm wide and infects *E. coli* bacteria by injecting its DNA into the host, causing the *E. coli* to non-lytically replicate, package, and release thousands more copies of the M13 phage. The phage itself is made up of circular, single stranded DNA, surrounded by several different coat proteins. Phage display libraries have been constructed by engineering the gene that codes for one of these minor coat proteins, the pIII. The library contains 10^9 different phage, each with a different, random peptide displayed on the N-terminus of the pIII [45]. These random

segments have been shown to be easily accessible to substrate molecules and the phage are easily replicated by infecting *E. coli*. Therefore, for each round of screening, the substrate-binding phage can be isolated by washing off non-binding phage and amplified to create a population of phage with a higher affinity for the specific substrate. With several rounds of screening, this becomes an evolutionary selection process to find the phage that bind very strongly with the substrate. Chapter two gives a more detailed description of phage display and the screening methods employed in these experiments.

1.2.2: Characterization of Peptides

The screening of these materials has resulted in over 700 peptide sequences for 10 different inorganic substrates. These sequences have also been screened for patterns within the subgroups determined by the screening parameters. Patterns and trends have been found for each material group and each method of screening employed. These patterns have been further analyzed to determine the statistical significance of the discovered patterns. Chapters three and four describe the patterns, trends and statistics of the peptide screenings for the calcium carbonate and stoichiometrically similar materials, respectively.

1.2.3: Peptide Structure Determination

Based on these discovered patterns, further characterization of these peptides has been done on a select few, very statistically significant, peptides. The peptides were synthesized using standard peptide synthesis methods. Further characterization included peptide NMR studies to determine secondary and

tertiary structure, which was then compared to computationally-determined molecular dynamics model. These results are discussed in chapter five.

1.2.4: Engineering of Peptides

The progression of this project has ultimately led to the genetic engineering of proteins to express the aragonite binding peptides. Using the M13 phage as scaffold, the aragonite binding peptides have been fused onto the major coat protein, increasing their expression level and allowing for studies on the effects of long-range order on crystal growth.

Further genetic engineering has also been done to synthesize doubly displayed peptides on the phage surface. These doubly engineered phage express peptides that recognize both lithium niobate (LiNbO_3) and zinc sulfide (ZnS) semiconductor material. Both of these protein engineering projects are described in chapter six.

1.2.5: Hybrid Organic-Inorganic Materials

Chapter six will also show that these genetically engineered phage can nucleate or capture nanoparticles of CaCO_3 . Furthermore, the doubly-displayed phage have been shown to recognize and bind to LiNbO_3 as well as serve as a support for ZnS nanoparticle growth, proving that protein-coupled hybrid materials is a viable application. Chapter six details the results of these materials growth experiments using the engineered phage

1.3: CHARACTERIZATION OF MATERIALS

The choice of substrates for this project began as an attempt to understand the specific interactions of biomolecules and the minerals they so elegantly

control in nature. Despite several decades of research and vast resources devoted to this problem, the specific sequence, structure or function of these biomolecules has remained essentially elusive, with a few notable exceptions. As mentioned previously, it has been shown that specific molecules are responsible for size, shape and polymorph control of the inorganic phase. It has also been shown that self-assembled monolayers of regularly repeating chemical groups can also control the size, shape and polymorph of certain biominerals [37, 39, 46]. However, the basic unit of interaction of a protein or peptide with its corresponding biomineral remains to be determined. If this basic unit of recognition can be determined, it may lead to the enhancement of theories of biomineralization mechanisms, which may subsequently lead to the bioengineering of controlled mineralization or growth of both biological and non-biological materials.

The project began with two of the most common biominerals found in nature, calcite and aragonite. These are the two minerals that make up mollusk shells, with the shell biomolecules controlling the deposition of regular, uniform crystals of both the thermodynamically stable phase, calcite and the meta-stable phase, aragonite.

Hydroxyapatite (HA) is another common biomineral, found in its carbonated form in both bones and teeth. It is structurally similar to aragonite, with carbonates being replaced by hydroxyl and phosphate groups, in the ratio of $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$. Determining the basic unit of interaction of HA, will open the door for a wide range of applications, such as composite inorganic-organic,

biocompatible materials growth, protein-coupled drug delivery systems, and protein recruitment of bone growth factors.

Finally, this project has been extended to include electronic and optical oxide materials that are stoichiometrically similar to calcium carbonate. Barium titanate (BaTiO_3) and lead titanate (PbTiO_3) are perovskite materials used in ferroelectric and piezoelectric applications as well as in optical materials. Lithium niobate (LiNbO_3) is an optical material used in laser waveguides. It also has acoustic, piezoelectric and pyroelectric properties.

Each material was characterized by powder X-Ray Diffraction (XRD) to verify the crystal structure and/or orientation using a Bruker-Nonius D8 Advance Theta-2Theta Powder Diffractometer (TMI Core Facilities, University of Texas at Austin).

1.3.1: Geological Calcite

The crystal structure of calcite, as determined by W. L. Bragg [47], can be described as a rhombohedral crystal, in both shape and axial system, but it is most often classified today as a hexagonal system. The carbonate, CO_3 , groups are arranged as planar, near-equilateral triangles oriented perpendicular to the c-axis. The O-C-O angles deviate little from the expected 120° and the C-O bond length has a mean value of 1.28 \AA [48]. Table 1.2 gives the unit cell parameters for calcite.

a	b	c	α	β	γ
4.988	4.988	17.068	90°	90°	120°

Table 1.2: Unit cell parameters for calcite (PDF#86-0174).

Calcite has a cleavage plane along the (104) face [49], giving rise to rhombohedral crystals with all six faces having identical (104) surfaces. Calcite has R-3c symmetry. Each calcium is coordinated by six oxygen molecules, belonging to six different CO₃ groups; three from the layer below it and three from the layer above. Each consecutive layer of carbonates is rotated 180° from the layer before it. Figure 1.2 shows both the calcite crystal and its molecular structure. The (104) surface is relatively flat, giving it a fairly even charge distribution.

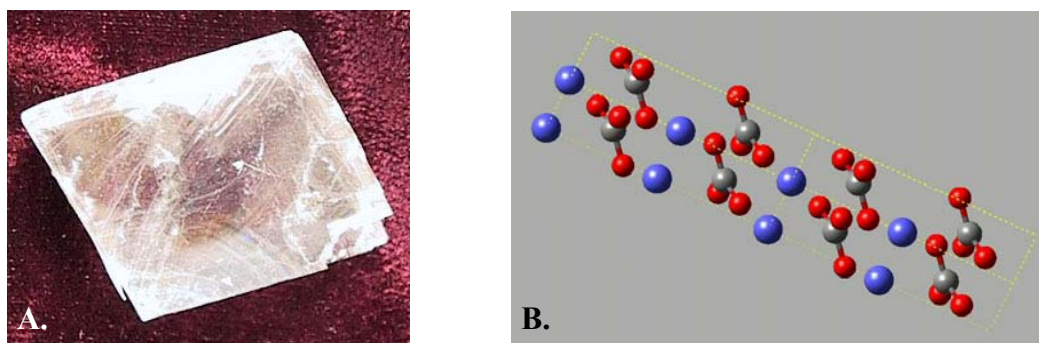
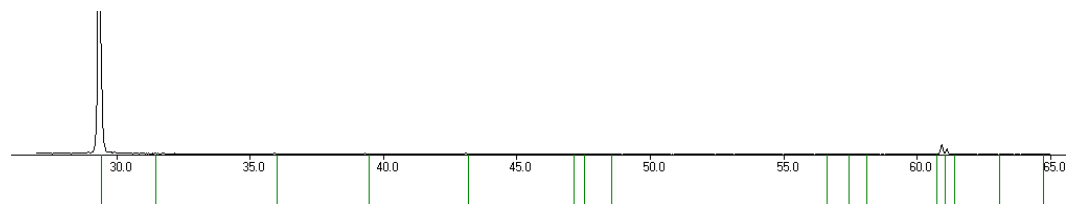


Figure 1.2: A) Geological calcite rhombohedron [50]. B) Calcite (104) surface; calcium is blue, carbon is grey, oxygen is red (Molecular Simulations, Inc.).

Calcite is the thermodynamically stable form of calcium carbonate, and calcite rhombohedrons can easily be precipitated out of a supersaturated solution of calcium carbonate at room temperature and at ambient pressure. However, this precipitation is often preceded by the formation of the meta-stable vaterite polymorph of calcium carbonate, which converts relatively quickly to calcite.

Vaterite is the most soluble form of calcium carbonate, and according to the Ostwald step rule, it should precipitate first.

The calcite crystals used in our experiments were geological samples, cleaved to achieve a rhombohedron with as few step edges as possible and with dimensions of 0.5 – 1.0 cm on each side. The crystals were washed in sterilized 20 mM NH_4HCO_3 buffer to remove any surface contaminants. The NH_4^+ escapes as gaseous ammonia, leaving a basic solution ($\sim\text{pH } 9.0$) that will not etch the pH vulnerable calcite surface. Calcium carbonate is susceptible to etching at pHs below about 8.0. Figure 1.3, below, is the X-ray diffraction of a typical calcite crystal used in these experiments.



Peak	h k l	2θ
1	1 0 4	29.398
2	2 0 8	60.992

Figure 1.3: XRD of calcite showing orientation along the (104) face.

1.3.2: Geological Aragonite

Aragonite is a meta-stable form of calcium carbonate, usually formed at high temperatures and/or high pressures. The crystal structure of aragonite, also determined by W. L. Bragg [51], is orthorhombic or pseudo-hexagonal. Here, the nearly-planar carbonate ions also lie perpendicular to the c-axis, but in a

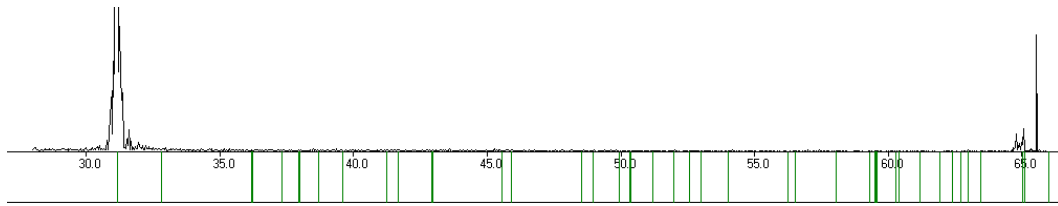
corrugated fashion (Figure 1.6 A). In aragonite, the carbonates are more closely packed than in calcite. Therefore, aragonite is a denser material. In addition, each calcium atom is coordinated to nine oxygen atoms, instead of six [52].

The most commonly ascribed space group for aragonite is the Pmcn, where $a < c < b$; c is aligned along the pseudo-hexagonal axis. However, the original space group as determined by Wyckoff [53] was Pnam, where $c < a < b$. Two faces of aragonite were screened in this project due to their mutual and natural occurrence in many geological samples (Figure 1.6 D). The samples used were mined from Spain and grew as pseudo-hexagonal off-growths emerging from a bulk crystal. These pseudo-hexagonal crystals are formed naturally by twinning through the existence of a glide plane [54], resulting in a crystal with six (110) faces. These crystals are elongated in the c -direction, terminating in a (001) surface (Figure 1.6 D) perpendicular to the (110) faces.

a	b	c	α	β	γ
4.9598	7.9641	5.7379	90°	90°	90°

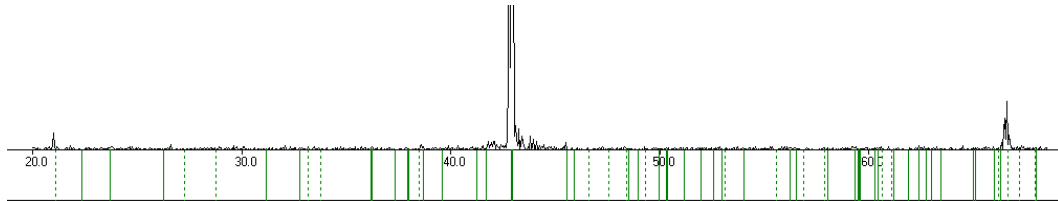
Table 1.3: Unit cell parameters of aragonite [55-57].

The aragonite samples were cleaved from the bulk crystal with preference given to crystals with the fewest step edges on both (110) and (001) surfaces. They too were washed in 20 mM NH_4HCO_3 as they are also susceptible to etching. The X-ray diffraction patterns below confirm the two different major faces of the geological aragonite samples.



Peak	h k l	2θ
1	0 0 2	31.149
2	0 0 4	64.956

Figure 1.4: XRD of the non-twinned face of the geological aragonite crystal, showing 002 and 004 parallel planes.



Peak	h k l	2θ
1	1 1 0	21.084
2	2 2 0	42.965
3	3 3 0	66.580

Figure 1.5: XRD of the twinned face of the aragonite crystal, showing 110, 220, and 330 planes.

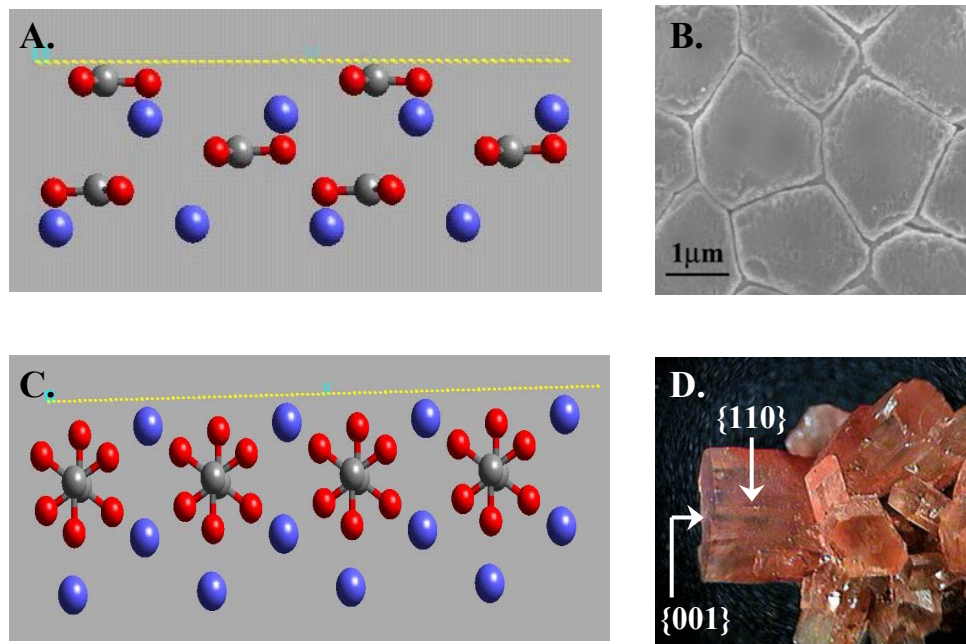
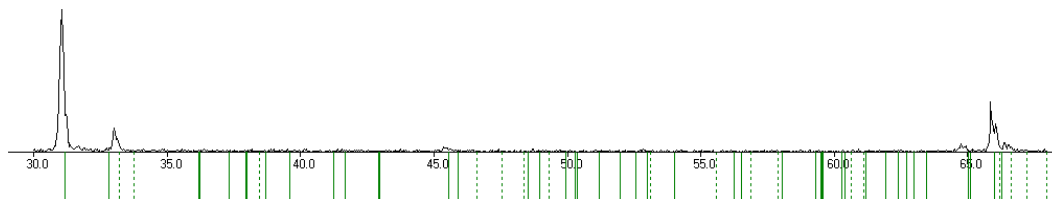


Figure 1.6: A.) Illustration of the aragonite (001) surface, calcium is blue, carbon is grey, oxygen is red (Molecular Simulations, Inc.). B.) SEM image of nacreous aragonite tablets; {001} surface. C.) Illustration of the aragonite (001) surface (Molecular Simulations, Inc.). D.) Image of geological aragonite [58]; there are 6 pseudo-hexagonally-twinned {110} faces parallel to a single {001} surface.

1.3.3: Nacreous Aragonite

Biological nacre aragonite was taken from the mother-of-pearl layer of the shell of a black-lipped oyster (*Pinctada margaritifera*) harvested from a pearl farm in the Cooke Islands. The nacre was bleached several times in strong bleach (10-12%) to remove any surface organic material. The aragonite tablets in this layer are arranged in the typical “brick and mortar” arrangement and are about 10 μm across and 500 nm thick. The tablets are elongated in the ab direction, while the c-direction (001) is perpendicular to the surface of the shell. Each tablet layer is

separated from the next by a thin organic layer. However, since the bleaching was done without sonication, the removal of the organic layer only applies to the top few layers, allowing the nacre sample to remain a solid $\sim 1 \text{ cm}^2$ piece.



Peak	h k l	2θ
1	0 0 2	31.149
2	1 2 1	32.767
3	0 0 4	64.956
4	2 2 3	65.932

Figure 1.7: XRD of 001 oriented aragonite from black lipped oyster nacre.

1.3.4: Hydroxyapatite

Hydroxyapatite (HA), $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$, is a commonly studied bone implant material due to the fact that it is biocompatible. Furthermore, it has been found that HA actually stimulates bone growth and does not elicit a negative immune response [59, 60]. HA is even more interesting when its crystal structure is taken into consideration. HA is a hexagonal crystal with $\text{P6}_3/\text{m}$ symmetry and lattice parameters quite similar to aragonite (Table 1.4) [61], and has even been hydrothermally converted from biological and geological aragonite [13].

a	b	c	α	β	γ
9.418	9.418	6.884	90°	90°	120°

Table 1.4: Unit cell parameters for hydroxyapatite (PDF#09-0432).

Actual bone is 60-70% dry weight [62] carbonate apatite mineral, also called dahllite and 30–40% organic material. The percentage of carbonate ranges from about 4-6% [63, 64]. Dahllite has hexagonal symmetry [63] and grows in thin plates with the c-axis in the plane of the plates [65]. The plate faces themselves are (100) surfaces. The dimensions of the plates are roughly 500 x 250 x 25 Å, although the length and width vary widely (100 – 1000 Å).

Like shells, bone has a biomolecule framework that is intricately associated with the inorganic crystals. In bone, the organic scaffold is made up of type I collagen. The mineral plates are arranged within the fibrils and can be aligned over neighboring fibrils up to distances of about 1 μm [66].

Biocompatible HA ceramics are typically made by high temperature syntheses, resulting in high crystallinity, but relatively poor fracture toughness as compared to natural bone [67]. This can be overcome by introducing additives that mechanically reinforce the structure, but high temperature sintering is still required [68-70]. In addition, the nacre-HA conversion resulted in a material that was not as hard as native nacre, although the coral-HA conversion did result in material that was harder than native coral [71]. Since natural bone is a composite organic-inorganic material made by protein-directed growth at biological temperatures and pressures, it follows that studying the basic interaction of proteins with natural or synthetic bone material may someday lead to a biomimetic approach to bone growth.

The polycrystalline, sintered HA blocks (Berkeley Advanced Biomaterials, Inc.) were studied under the same experimental conditions at CaCO_3 .

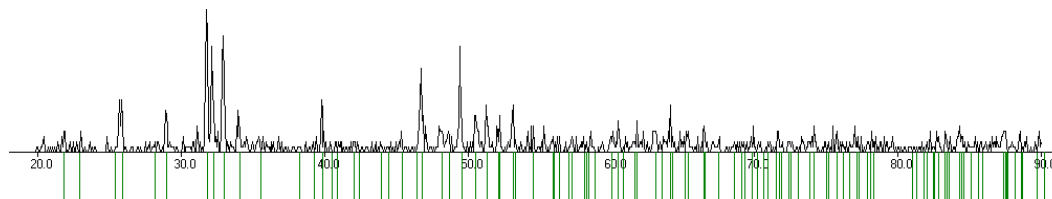


Figure 1.8: XRD of sintered polycrystalline hydroxyapatite pellet.

1.3.5: Barium Titanate

Barium Titanate, BaTiO_3 , is a high dielectric constant, perovskite material that is also ferroelectric and has therefore been used in capacitors as well as piezoelectric transducers [72, 73]. In addition, there has also been research in the areas of chemical sensors and thermistors [74, 75]. Often, the application requires thin films, but typical methods of growing these films necessitate high temperatures and sometimes high pressures. This can lead to interdiffusion and reactions at the film interface. Furthermore, the thermal stress of heating and cooling can cause cracking and affect the long-term stability of the device [76, 77]. These thin films of barium titanate are typically synthesized using metal-organic deposition laser ablation, sputtering, or sol-gel processing [78-81]. Currently, hydrothermal syntheses are being investigated as a low temperature route to thin film growth [82]. Still, there remains difficulty in attaining the requisite uniform thickness, orientation and lack of surface roughness. Film synthesis of materials similar to calcium carbonate (EuCoCO_3) at ambient

pressure and temperature has been attained by using proteins extracted from mollusk shells [83]. Further studies of protein mediated crystal growth at ambient temperatures and pressures could lead to better surface morphology and crystal orientation, provided that the film still retains its ferroelectric properties.

Further applications are composite thin films used in optical devices such as waveguides and optical fibers for optical switching or amplification of optical signals [84]. Self-assembled nanoparticle-embedded thin films have been made using dielectric films, including BaTiO₃ and metal nanoparticles [85-89]. Thus, harnessing the self-assembling nature of proteins to enhance this process would be a great achievement indeed. The barium titanate crystals used in these screenings were in polycrystalline powder form (Aldrich Chemical Co., 99.995%)

a	b	c	α	β	γ
5.7248	5.7248	13.967	90°	90°	120°

Table 1.5: Unit cell parameters for BaTiO₃ (PDF#34-0129).

1.3.6: Lead Titanate

Lead titanate, PbTiO₃, also known as macedonite, is another perovskite material with piezoelectric and ferroelectric properties. Thin film PbTiO₃ has been used in piezoelectric devices, infrared sensors, optical devices and dynamic random access memories (DRAMs) [90]. Like its perovskite relative, BaTiO₃, thin films are made by sputtering, sol-gel, molecular chemical vapor deposition (MOVCD), spray pyrolysis, electrostatic spray deposition (ESD), as well as hydrothermal methods [90] Similarly, most of these methods necessitate high

temperatures at both the deposition step and the annealing step. Thus, thin film applications of lead titanate must also deal with the issue of layer interdiffusion, uniform thickness, and surface roughness.

Today, these ceramic oxide devices are predominantly made with zirconium-doped lead titanate, or PZT ($\text{PbZr}_x\text{Ti}_{1-x}\text{O}_3$). The zirconium atoms replace the titanium atoms that occupy the central tetragonal position. This material has the same perovskite structure, but has advantages in its high electromechanical coupling factor, excellent frequency-temperature characteristics and can be made with good quality [91].

Making PbTiO_s thin films by protein directed growth has the same advantages as making BaTiO_3 in this way. While studying PbTiO_s may seem less practical than PZT, it is most likely a better starting place given its simpler composition. Substituting Zr may be just a matter of changing the concentration of the starting materials. In addition, PbTiO_s has the potential in nonlinear optical applications just like nanoparticle- BaTiO_3 composite films [92]. The PbTiO_s screened in this procedure was polycrystalline powder (Alfa Aesar, 99.9%).

a	b	c	α	β	γ
12.358	12.358	14.541	90°	90°	90°

Table 1.6: Unit cell parameters of lead titanate (PDF#48-0105).

1.3.7: Lithium Niobate

Lithium Niobate, LiNbO_3 is also a ferroelectric, piezoelectric and pyroelectric material, although it does not have a true perovskite structure. Instead, it is a rhombohedral crystal with hexagonal lattice coordinates (Table 1.7) and can be polarized due to the existence of a distorted octahedrons, with alternating Li, Nb, and vacant centers, as viewed in the Z-direction. LiNbO_3 is often used as a nonlinear optical material for optical parametric oscillators and frequency doublers. It is also a photoreactive material that can be used to record holograms when coupled with a laser [93]. Although, it is typically doped with iron to increase the photoreactive response [94]. Furthermore, LiNbO_3 is also used in surface acoustic wave devices and waveguides due to its large acousto-optic and electro-optic coefficients [chen, xia].

a	b	c	α	β	γ
5.1494	5.1494	13.862	90°	90°	120°

Table 1.7: Lithium niobate unit cell parameters (PDF#20-0631).

Waveguides are fabricated by creating refractive index barriers. This can be done by etching the waveguide material or by ion diffusion, usually with titanium. Often these waveguides are doped with erbium (Er) ions that amplify the signal over long distances. However, if quantum dots are coupled to the evanescent field of the waveguide, meaning that they are located very close to the optical mode of the waveguide, these quantum dots could act as amplifiers [95].

Therein lies the potential for self-assembled protein coupled quantum dot amplifiers.

Both polycrystalline powder LiNbO_3 (Alfa Aesar, Puratronic, 99.9995%) and single crystal Z-poled LiNbO_3 wafer pieces (Bell Labs) were used in these screenings. The Z-direction was noted for each screening as +Z and -Z, corresponding to the direction of polarization. The single crystal samples were broken into pieces of about 1-2 cm^2 and both powder and wafer pieces were washed with ultrapure water.

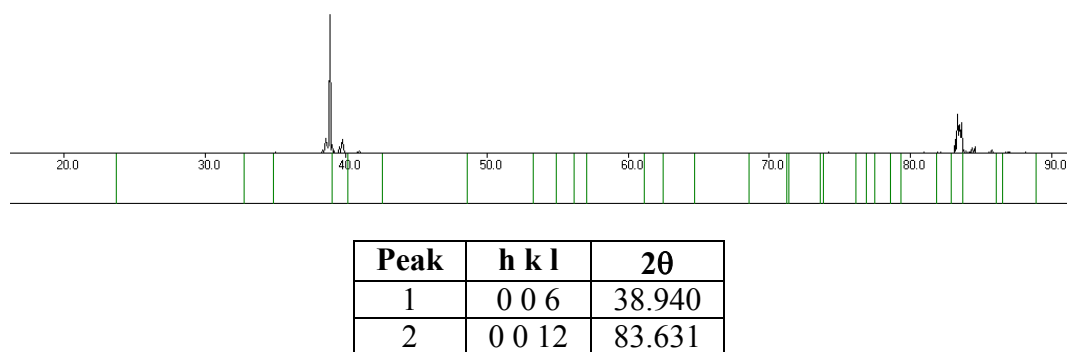


Figure 1.9: XRD of single crystal LiNbO_3 .

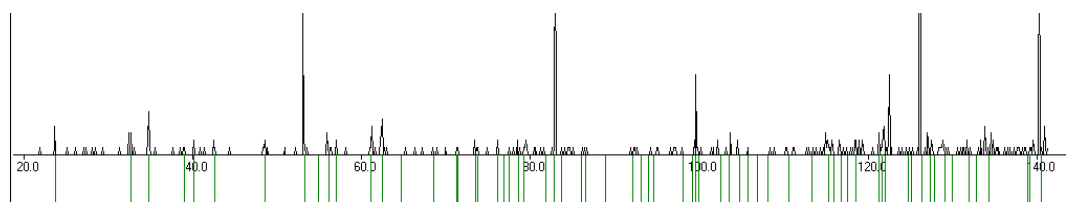


Figure 1.10: XRD of powdered polycrystalline LiNbO_3 .

REFERENCES:

1. *On Biomineralization*, ed. H.A. Lowenstam and S. Weiner. 1989, Oxford: Oxford University Press.
2. Bazylinski, D.A. and R.B. Frankel, *Magnetic Iron Oxide and Iron Sulfide Minerals within Microorganisms*, in *Biomineralization: From Biology to Biotechnology and Medical Application*, E. Baeuerlein, Editor. 2000, Wiley-VCH: Weinheim. p. 25-46.
3. Ferguson, A.D., et al., *Siderophore-Mediated Iron Transport: Crystal Structure of FhuA with Bound Lipopolysaccharide*. *Science*, 1998. **282**(5397): p. 2215-2220.
4. Shenton, W., et al., *Synthesis of nanophase iron oxide in lumazine synthase capsids*. *Angewandte Chemie-International Edition*, 2001. **40**(2): p. 442-445.
5. Chakraborty, J., et al., *Behaviour of iron ion in the morphosynthesis of magnetite particles*. *Materials Transactions*, 2003. **44**(6): p. 1124-1127.
6. Kroger, N. and M. Sumper, *Diatom cell wall proteins and the cell biology of silica biomineralization*. *Protist*, 1998. **149**(3): p. 213-219.
7. Kroger, N., R. Deutzmann, and M. Sumper, *Polycationic peptides from diatom biosilica that direct silica nanosphere formation*. *Science*, 1999. **286**(5442): p. 1129-1132.
8. Kroger, N. and M. Sumper, *The Biochemistry of Silica Formation in Diatoms*, in *Biomineralization*, E. Baeuerlein, Editor. 2000, Wiley-VCH: Weinheim. p. 151-170.
9. Shimizu, K., et al., *Silicatein alpha: Cathepsin L-like protein in sponge biosilica*. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**(11): p. 6234-6238.
10. Cha, J.N., et al., *Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones in vitro*. *Proceedings of the National Academy of Sciences of the United States of America*, 1999. **96**(2): p. 361-365.

11. Cha, J.N., et al., *Biomimetic synthesis of ordered silica structures mediated by block copolypeptides*. *Nature*, 2000. **403**(6767): p. 289-292.
12. Fritz, M., et al., *Flat pearls from biofabrication of organized composites on inorganic substrates*. *Nature (London, United Kingdom)*, 1994. **371**(6492): p. 49-51.
13. Zaremba, C.M., et al., *Critical Transitions in the Biofabrication of Abalone Shells and Flat Pearls*. *Chemistry of Materials*, 1996. **8**(3): p. 679-90.
14. Weiner, S. and W. Traub, *Macromolecules in Mollusk Shells and Their Functions in Biomineralization*. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 1984. **304**(1121): p. 425-&.
15. Jackson, A.P., J.F.V. Vincent, and R.M. Turner, *Comparison of nacre with other ceramic composites*. *Journal of Materials Science*, 1990. **25**(7): p. 3173-8.
16. Manne, S., et al., *Atomic Force Microscopy of the Nacreous Layer in Mollusc Shells*. *Proceedings: Biological Sciences*, 1994. **256**(1345): p. 17-23.
17. Belcher, A.M., et al., *Control of crystal phase switching and orientation by soluble mollusk-shell proteins*. *Nature (London)*, 1996. **381**(6577): p. 56-58.
18. Falini, G., et al., *Control of aragonite or calcite polymorphism by mollusk shell macromolecules*. *Science*, 1996. **271**(5245): p. 67-69.
19. Schaeffer, T.E., et al., *Does Abalone Nacre Form by Heteroepitaxial Nucleation or by Growth through Mineral Bridges?* *Chemistry of Materials*, 1997. **9**(8): p. 1731-1740.
20. Samata, T., et al., *A new matrix protein family related to the nacreous layer formation of *Pinctada fucata**. *Febs Letters*, 1999. **462**(1-2): p. 225-229.
21. Miyashita, T., et al., *Complementary DNA cloning and characterization of pearlins, a new class of matrix protein in the nacreous layer of oyster pearls*. *Marine Biotechnology*, 2000. **2**(5): p. 409-418.

22. Shen, X.Y., et al., *Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of Haliotis rufescens*. Journal of Biological Chemistry, 1997. **272**(51): p. 32472-32481.
23. Sudo, S., et al., *Structures of mollusc shell framework proteins*. Nature, 1997. **387**(6633): p. 563-564.
24. Miyamoto, H., et al., *A carbonic anhydrase from the nacreous layer in oyster pearls*. Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(18): p. 9657-9660.
25. Marin, F., et al., *Mucins and molluscan calcification - Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel Pinna nobilis (Bivalvia, Pteriomorpha)*. Journal of Biological Chemistry, 2000. **275**(27): p. 20667-20675.
26. Kono, M., N. Hayashi, and T. Samata, *Molecular mechanism of the nacreous layer formation in Pinctada maxima*. Biochemical and Biophysical Research Communications, 2000. **269**(1): p. 213-218.
27. Weiss, I.M., et al., *Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc Haliotis laevigata*. Biochemical and Biophysical Research Communications, 2000. **267**(1): p. 17-21.
28. Michenfelder, M., et al., *Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains*. Biopolymers, 2003. **70**(4): p. 522-533.
29. Weiss, I.M., et al., *Perlustrin, a Haliotis laevigata (abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N-terminal module of vertebrates*. Biochemical and Biophysical Research Communications, 2001. **285**(2): p. 244-249.
30. Blank, S., et al., *The nacre protein perlucin nucleates growth of calcium carbonate crystals*. Journal of Microscopy-Oxford, 2003. **212**: p. 280-291.
31. Yano, M., Miyashita, T. and Miyamoto, H., *Shell matrix Nacrein, a novel family of carbonic anhydrase, is conserved in bivalve and gastropod*. 2001, Entrez.

32. Hayashi, N.a.S., T., *N14#2 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
33. Hayashi, N.a.S., T., *N14#3 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
34. Hayashi, N.a.S., T., *N14#4 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
35. Hayashi, N.a.S., T., *N14#5 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
36. Hayashi, N.a.S., T., *N14#7 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
37. Aizenberg, J., A.J. Black, and G.M. Whitesides, *Control of crystal nucleation by patterned self-assembled monolayers*. *Nature*, 1999. **398**(6727): p. 495-498.
38. Aizenberg, J., *Patterned crystallization of calcite in vivo and in vitro*. *Journal of Crystal Growth*, 2000. **211**(1-4): p. 143-148.
39. Aizenberg, J., A.J. Black, and G.H. Whitesides, *Oriented growth of calcite controlled by self-assembled monolayers of functionalized alkanethiols supported on gold and silver*. *Journal of the American Chemical Society*, 1999. **121**(18): p. 4500-4509.
40. Aizenberg, J., et al., *Direct fabrication of large micropatterned single crystals*. *Science*, 2003. **299**(5610): p. 1205-1208.
41. Addadi, L., et al., *A chemical model for the cooperation of sulfates and carboxylates in calcite crystal nucleation: relevance to biomineralization*. *Proc. Natl. Acad. Sci. U. S. A.*, 1987. **84**(9): p. 2732-6.
42. Didymus, J.M., et al., *Influence of low-molecular-weight and macromolecular organic additives on the morphology of calcium carbonate*. *Journal of the Chemical Society*, 1993. **Faraday Transactions** **89**(15): p. 2891-900.
43. DeOliveira, D.B. and R.A. Laursen, *Control of Calcite Crystal Morphology by a Peptide Designed To Bind to a Specific Surface*. *Journal of the American Chemical Society*, 1997. **119**(44): p. 10627-10631.

44. Levi, Y., et al., *Control over aragonite crystal nucleation and growth: an in vitro study of biomineralization*. Chemistry--A European Journal, 1998. **4**(3): p. 389-396.
45. *Ph.D.-12™, Ph.D.-7™, Ph.D.-C7C™ Phage Display Peptide Library Kit Instruction Manuals*: New England Biolabs.
46. Aizenberg, J., *Patterned crystallisation on self-assembled monolayers with integrated regions of disorder*. Journal of the Chemical Society-Dalton Transactions, 2000(21): p. 3963-3968.
47. Bragg, W.L., *The analysis of crystals by the x-ray spectrometer*. Proceedings of the Royal Society of London, 1914(A89): p. 468-489.
48. Zemann, J., *Zur Stereochemie der Karbonate*. Fortschr. Mineral, 1981. **59**: p. 95-116.
49. *Carbonates: Mineralogy and Chemistry*. Reviews in Mineralogy, ed. R.J. Reeder. Vol. 11. 1983.
50. Ojard, D., http://www.d.umn.edu/geology/pagnucco/Carbonates/smallpics/Calcite-2881_small.jpg.
51. Bragg, W.L., *The Structure of Aragonite*. Proceedings of the Royal Society of London, 1924. **105**: p. 16-39.
52. Lippmann, F., *Sedimentary Carbonate Minerals*. 1973, New York: Springer-Verlag.
53. Wyckoff, R.W.G., *Orthorhombic space group criteria and their application to aragonite*. American Journal of Science, 1925. **5th Series**(1X): p. 145-175.
54. Bragg, W.L., *Atomic Structure of Minerals*. 1937, Oxford, England: Oxford University Press.
55. Dal Negro, A. and L. Ungaretti, *Refinement of the crystal structure of aragonite*. American Mineralogist, 1971. **56**: p. 768-772.
56. DeVilliers, J.P.R., *Crystal structures of aragonite, strontianite and witherite*. American Mineralogist, 1971. **56**: p. 758-772.

57. Dickens, B. and J.S. Bowen, *Refinement of the crystal structure of the aragonite phase of CaCO₃*. Journal of Research National Standards A. Phys. Chem., 1971. **75A**: p. 27-32.
58. <http://www.theimage.com/mineral/aronite/aronite.jpg>.
59. White, E. and E.C. Shors, *Biomaterial Aspects of Interpore-200 Porous Hydroxyapatite*. Dental Clinics of North America, 1986. **30**(1): p. 49-67.
60. Ohgushi, H., et al., *Bone-Formation Process in Porous Calcium-Carbonate and Hydroxyapatite*. Journal of Biomedical Materials Research, 1992. **26**(7): p. 885-895.
61. Eysel, W. and D.M. Roy, *Topotactic Reaction of Aragonite to Hydroxyapatite*. Zeitschrift Fur Kristallographie, 1975. **141**(1-2): p. 11-24.
62. Posner, A.S., N.C. Blumenthal, and F. Betts, in *Phosphate Minerals*, L.O. Nriagu and P.B. Moore, Editors. 1984, Springer-Verlag: Berlin; New York.
63. Weiner, S. and W. Traub, *Bone-Structure - from Angstroms to Microns*. Faseb Journal, 1992. **6**(3): p. 879-885.
64. McConnell, D., Journal of Dental Research, 1952. **31**: p. 53.
65. Selvig, K.A., *Periodic lattice images of hydroxyapatite crystals in human bone and dental hard tissues*. Calcified Tissue Research, 1970. **6**: p. 227-238.
66. Traub, W., T. Arad, and S. Weiner, *3-Dimensional Ordered Distribution of Crystals in Turkey Tendon Collagen-Fibers*. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(24): p. 9822-9826.
67. Suchanek, W. and M. Yoshimura, *Processing and properties of hydroxyapatite-based biomaterials for use as hard tissue replacement implants*. Journal of Materials Research, 1998. **13**(1): p. 94-117.
68. Ahn, E.S., et al., *Nanostructure processing of hydroxyapatite-based bioceramics*. Nano Letters, 2001. **1**(3): p. 149-153.
69. Li, J., B. Fartash, and L. Hermansson, *Hydroxyapatite Alumina Composites and Bone-Bonding*. Biomaterials, 1995. **16**(5): p. 417-422.

70. Ioku, K., M. Yoshimura, and S. Somiya, *Microstructure and Mechanical-Properties of Hydroxyapatite Ceramics with Zirconia Dispersion Prepared by Post-Sintering*. Biomaterials, 1990. **11**(1): p. 57-61.
71. Roy, D.M. and S.K. Linnehan, *Hydroxyapatite formed from coral skeletal carbonate by hydrothermal exchange*. Nature (London, 1974. **United Kingdom**) **247**(5438): p. 220-2.
72. Newnham, R.E., *Electroceramics*. Reports on Progress in Physics, 1989. **52**(2): p. 123-156.
73. Cross, L.E., *Dielectric, Piezoelectric, and Ferroelectric Components*. American Ceramic Society Bulletin, 1984. **63**(4): p. 586-590.
74. Yeh, Y.C. and T.Y. Tseng, *Humidity-Sensitive Electrical-Properties of Ba_{0.5}Sr_{0.5}TiO₃ Porous Ceramics*. Journal of Materials Science Letters, 1988. **7**(7): p. 766-768.
75. Ishihara, T., et al., *Mixed-Oxide Capacitor of CuO-BaTiO₃ as a New Type CO₂ Gas Sensor*. Journal of the American Ceramic Society, 1992. **75**(3): p. 613-618.
76. Kwok, C.K. and S.B. Desu, *Low-Temperature Perovskite Formation of Lead Zirconate Titanate Thin-Films by a Seeding Process*. Journal of Materials Research, 1993. **8**(2): p. 339-344.
77. Chen, S.Y. and I.W. Chen, *Cracking During Pyrolysis of Oxide Thin-Films - Phenomenology, Mechanisms, and Mechanics*. Journal of the American Ceramic Society, 1995. **78**(11): p. 2929-2939.
78. Kotecki, D.E., et al., *(Ba,Sr)TiO₃ dielectrics for future stacked-capacitor DRAM*. Ibm Journal of Research and Development, 1999. **43**(3): p. 367-382.
79. Yoshimura, T., N. Fujimura, and T. Ito, *The initial stage of BaTiO₃ epitaxial films on etched and annealed SrTiO₃ substrates*. Journal of Crystal Growth, 1997. **174**(1-4): p. 790-795.
80. Sharma, H.B. and A. Mansingh, *Phase transition in sol-gel-derived barium titanate thin films*. Journal of Physics D-Applied Physics, 1998. **31**(13): p. 1527-1533.

81. Xu, J.J., A.S. Shaikh, and R.W. Vest, *High K BaTiO₃ Films from Metalloorganic Precursors*. Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control, 1989. **36**(3): p. 307-312.
82. Chien, A.T., et al., *Low-Temperature Low-Pressure Hydrothermal Synthesis of Barium-Titanate - Powder and Heteroepitaxial Thin-Films*. Journal of Materials Research, 1995. **10**(7): p. 1784-1789.
83. Belcher, A.M., *Personal communication: Inorganic film growth using Abalone nacre proteins*. 1997: Santa Barbara, CA.
84. Liu, Y., et al., *Third-order nonlinear optical response of Au-core CdS-shell composite nanoparticles embedded in BaTiO₃ thin films*. Applied Physics B-Lasers and Optics, 2003. **76**(4): p. 435-439.
85. Ricard, D., P. Roussignol, and C. Flytzanis, *Surface-Mediated Enhancement of Optical-Phase Conjugation in Metal Colloids*. Optics Letters, 1985. **10**(10): p. 511-513.
86. Flytzanis, C., et al., *Nonlinear Optics in Composite-Materials*. Journal of the Optical Society of America B-Optical Physics, 1986. **3**(8): p. P93-&.
87. Hache, F., et al., *The Optical Kerr Effect in Small Metal Particles and Metal Colloids - the Case of Gold*. Applied Physics a-Materials Science & Processing, 1988. **47**(4): p. 347-357.
88. Tokizaki, T., et al., *Subpicosecond Time Response of 3rd-Order Optical Nonlinearity of Small Copper Particles in Glass*. Applied Physics Letters, 1994. **65**(8): p. 941-943.
89. Liao, H.B., et al., *Large third-order optical nonlinearity in Au:SiO₂ composite films near the percolation threshold*. Applied Physics Letters, 1997. **70**(1): p. 1-3.
90. Huang, H., et al., *Ferroelectric PbTiO₃ thin films prepared by electrostatic spray deposition (ESD)*. Microelectronic Engineering, 2003. **66**(1-4): p. 688-694.
91. Ren, T.L., et al., *Piezoelectric and ferroelectric films for microelectronic applications*. Materials Science and Engineering B-Solid State Materials for Advanced Technology, 2003. **99**(1-3): p. 159-163.

92. Tang, L., et al., *Preparation of silver dispersed PbTiO₃ film by sol-gel method*. Materials Science and Engineering B, 2003. **99**(1-3): p. 370-373.
93. Gunter, P. and J.P. Huignard, *Photorefractive Materials and Their Applications .I. Introduction*. Topics in Applied Physics, 1988. **61**: p. 1-5.
94. Hesselink, L., et al., *Photorefractive materials for nonvolatile volume holographic data storage*. Science, 1998. **282**(5391): p. 1089-1094.
95. Hu, E.L., *Personal communication: Quantum dot amplifiers for lithium niobate waveguides*. 2002: Santa Barbara, Ca.

Chapter 2: Peptide Selection Using Phage Display

2.1: PHAGE DISPLAY THEORY

A combinatorial approach has been employed to elucidate the interactions between biopolymers and the inorganic materials they so elegantly nucleate and shape. We have used a fused-peptide library displayed on the coat of M13 bacteriophage, also known as phage. M13 phage are filamentous phage belonging to the same family as fd phage, f1 phage and Ike phage [1]. A phage is a virus (900nm x 6nm) [2] that infects *E. coli* bacteria. It is made up of a single-stranded, circular piece of DNA, which is protected by a layer of coat proteins. The proteinaceous phage coat (Figure 2.1) is made up of primarily pVIII. However, the ends of the phage are capped with other proteins, pVII and pIX at one end and pVI and pIII at the infectious end [2].

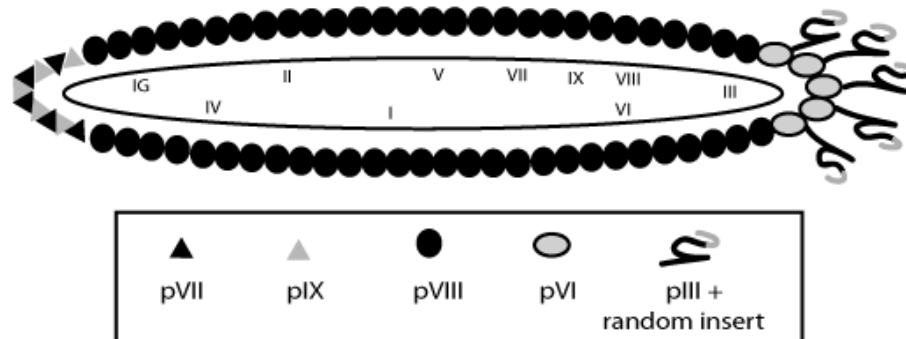


Figure 2.1: Diagram of the M13 filamentous phage coat proteins and single-stranded circular DNA, with gene numbers marked [2].

It is the pIII that is responsible for binding to the f-pilus of the *E. coli* bacteria and initiating infection. Peptides fused to this pIII protein are displayed on the surface of the virion particle and have been shown to be accessible to antibodies in solution [3]. Using combinatorial genetics, extremely large libraries of random peptides can be fused to the N-terminus of the phage pIII. Subsequently, these libraries can be screened simultaneously against a solid phase target, typically tethered antibodies or cellular receptors. Much of this work is done in the vein of “drug discovery,” and represents a powerful technique due to the physical link between phage phenotype and genotype. This work and work in this lab [4-7] represents some of the first applications of this system to screen against solid inorganic surfaces.

2.1.1: pIII Structure and Function

The mature pIII protein consists of 406 residues making up three domains, separated by glycine-rich repeats [8]. There is a C-terminal transmembrane region followed by D3, D2, and D1 at the N-terminus (Figure 2.2 and 2.3). The structure of the N-terminal domains D1 and D2 have been solved by X-ray crystallography for both M13 and fd filamentous phage [8, 9], although the C-terminal region and D3 domain have not. The N-terminal D1 segment has also been studied by NMR [10]. The M13 phage and fd phage D1-D2 fragments of the pIII differ by only two amino acids.

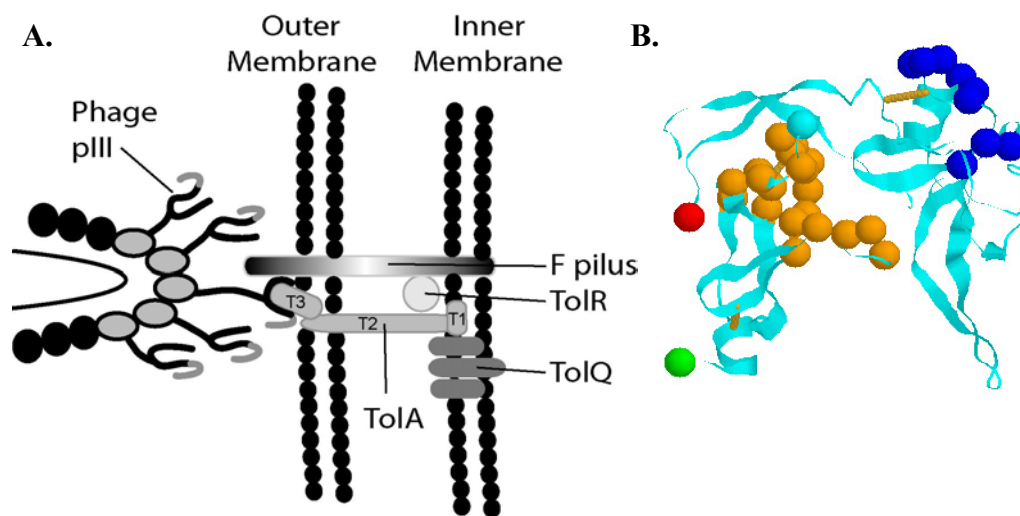


Figure 2.2: A.) Diagram of the phage pIII infecting its host *E. coli*. D2 binds the *E. coli* f-pilus, freeing D1, which interacts with the TolA-DIII B.) Ribbon representation [11] of the pIII D1-D2 “horseshoe,” N-terminal of D1 is green; C-terminal of D2 is red; f-pilus binding residues on D2 are dark blue; TolA binding residues on D1 are orange; (PBD IDs: 1g3p, 2g3p) [8, 9, 12, 13].

The D1-D2 structure resembles a horseshoe shape, with a hydrophobic core running between the two domains (Figure 2.2) [8, 9]. There is evidence in the two crystal structures that there is a hinge segment that allows D1 and D2 to separate, as if widening the horseshoe. Spatially, the N-terminus and the C-terminus of the D1D2 fragment are very close, with both ending up on the outside tip of the D1 domain, away from the hydrophobic inner core. In the peptide display libraries used in these experiments, the displayed peptide would be fused directly to the N-terminal alpha helix (Figure 2.2 B), separated only by the flexible SSSG linker.

Phage infection begins with the pIII binding to the bacterial f-pilus, causing the f-pilus to retract into the *E. coli* [14]. Then, the pIII also binds to the

bacterial TolA protein [15, 16], which is anchored in the cytoplasmic membrane, but is thought to span the periplasm so that it can interact with the phage pIII as it penetrates the outer membrane [17]. Subsequent steps in the infection process are yet undetermined, but necessarily involve phage DNA being imported into the cytoplasm and dissolution of the phage coat [18].

The phage pIII D2 is the domain that is responsible for bacterial f-pilus recognition, with binding being irreversible and triggering a retraction of the f-pilus towards the bacterium [14]. The bacterial f-pilus is made up of the protein pilin, which consists of 70 residues and is predominantly hydrophobic [19]. It is also suggested that the bacterial f-pilus may have pentameric symmetry [20]. This pentameric symmetry is also known to exist in the pVIII phage coat protein packing [21], and is likely to be the symmetry of the pIII arrangements, given that the pIII is present in approximately five copies per phage particle [22, 23].

Prior to infection, NMR and crystallography studies have shown that the phage pIII D2 interacts with the TolA binding site of D1, thereby making it inaccessible and the phage relatively non-infectious (Figure 2.3) [9, 15]. In the presence of the f-pilus, the TolA binding site on D1 becomes accessible. Therefore, it is thought that the binding of the f-pilus to D2 initiates a conformational change that allows D1 to be accessible to TolA. The dynamics of this conformational change are not known, however, thermodynamic studies have shown that at higher temperatures ($>45^{\circ}\text{C}$), phage are more infectious to *E. coli* that do not bear f-pili (f-) and bind TolA better in ELISA assays [8]. CD melting curves confirm conformational changes at temperatures greater than 45°C , and

give evidence that the thermal unfolding of the D1D2 fragment mimics the f-pilus binding conformational change necessary for TolA binding and infection. In addition, mutagenesis experiments have determined which amino acids on D2 are implicated in f-pili binding [12] (Figure 2.2). The residues that contribute the most to f-pili binding are located in a continuous segment (residues 178 – 199), on what is presumed to be the solvent accessible side of the D1-D2 complex. In other words, it is on the outer portion of the horseshoe. Since D1 and D2 interact with each other on the inside of the horseshoe, it seems even more likely that the D2-f-pilus interaction frees the inner portion of the horseshoe so that the second step of the infection process, TolA binding, can occur.

After f-pilus binding and f-pilus retraction, the now-accessible D1 domain of the pIII is brought into close proximity to the TolAQR proteins, which are all necessary for successful phage infection [17]. These three proteins form a complex in the cytoplasmic membrane of *E. coli*. However, it is the TolA protein that has been shown to bind to the phage pIII during infection [15, 16]. The TolA protein has three domains (Figure 2.2), with the N-terminal T1 being the anchor domain in the cytoplasmic membrane [24]. T2 is an extended helical domain that most likely spans the periplasmic space, allowing T3 to be positioned in or near the outer bacterial membrane. It is this C-terminal domain, T3, that is necessary for phage infection [15]. The pIII-D1 binds both TolA in the presence of the f-pilus and pIII-D2 in its absence. The pIII-D1 binding domains of TolA and pIII-D2 have no structural similarity, yet it has been determined that both TolA and pIII-D2 bind to nearly the same pIII-D1 residues [13].

Subsequent steps of infection are not well defined, although it has been shown that phage coat proteins can be found in the bacterial membrane after dissolution of the phage [17], and may in fact be reused in the progeny phage. After dissolution of the phage coat, the DNA is somehow imported into the bacteria, where the bacterial proteins replicate the DNA and phage proteins.

2.1.2: Phage Assembly in Bacteria

After phage infection and replication in the host bacterium, the newly formed DNA and phage proteins are packaged and extruded through the bacterial membrane. This occurs non-lytically, enabling the host to be very efficient amplifiers of the phage particles. There are 11 proteins coded by the phage genome [2]. As stated previously, there are four minor coat proteins (pIII, pVI, pVII, pIX), present in 3-5 copies per phage and one major coat protein (pVIII), present in about 2700 copies per phage. The remaining proteins are responsible for DNA replication (pII, pX), ssDNA binding (pV), and phage assembly (pI, pIV, pXI).

Phage assembly begins with the pV molecules binding to the single stranded DNA [25] to form the filamentous shape. Packaging then begins with the pVII and pIX coat proteins with assistance from the pI, an assembly protein [26]. Subsequently, the positively-charged pVIII proteins are complexed with the phage DNA, while the pV DNA-sequestering proteins are removed. The phage length is proportional to length of the DNA due to the charge coupling of the negatively charged DNA and the positively charged pVIII. The elongating phage

are then thought to be extruded through a channel formed by the phage pIV [27, 28]. Phage replication is completed by capping with the pIII and pVI [29].

Release of phage from the membrane is thought to entail a two step process involving the pIII [30]. The pIII and pVI must initially be well anchored in the membrane and able to interact with the pVIII coat proteins. The second step is a conformational change in the pIII that disrupts the pIII and pVI membrane interactions, allowing the phage particle to be released. It seems as though the pIII “flips” its orientation during the capping process, allowing for the protection of the pVI, which is known to be susceptible to degradation in the absence of the pIII [31].

It has been discovered that the C-terminal region (D3) of the pIII is responsible for the membrane incorporation, stabilization of phage, and release from membranes [30]. This region has been divided into four significant regions. First, there is a glycine-rich spacer that separates D2 from D3. Following this, there is a region called C1 that has been shown to be necessary in making the phage particles stable. Next to this region is the C2 portion that is essential for phage release from the bacterial membranes. Phage particles that contain the pIII C2 region but not the C1 region will release phage, but they are not stable in detergents, whereas wild type phage are quite detergent resistant. Finally, there is a hydrophobic membrane-anchoring region at the very end of the C-terminus (Figure 2.3).

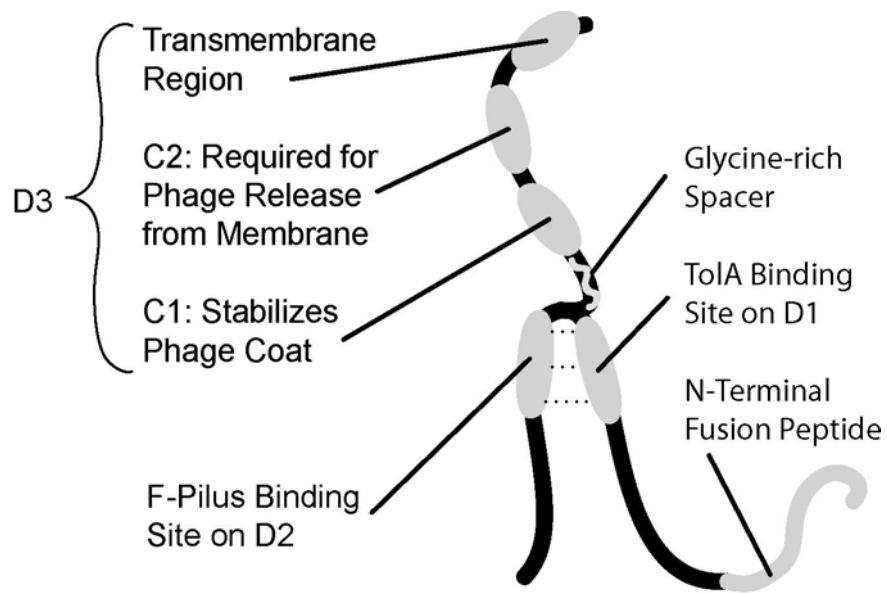


Figure 2.3: Diagram of the pIII showing the domains and functions of each. Note that the D2 interacts with the Tola binding site on D1.

Since the engineered segment of the pIII lies at the N-terminus, it is unlikely that an engineered pIII would affect the ability of the phage to replicate or to be released from the bacterial membrane. However, since one cannot test all possible (10^9) random peptide fusions for their ability to incorporate into the membrane and be released from the bacterium, we cannot presume that all engineered phage will replicate with the same efficiency.

2.1.3: Commercial Phage Display System – New England Biolabs

The New England Biolabs Phage Display Libraries, PhD12, PhD7 and PhD7C have been used as our combinatorial peptide libraries. Each library displays random peptides on the N-terminus of the pIII coat protein after a GGGS spacer designed to give the displayed peptide some flexibility. The distinguishing

factor between the libraries is the length and shape of the peptide displayed. The “12” designates 12 inserted amino acids, while the “7” designates 7 inserted amino acids. The “C” refers to the constrained or loop configuration of the genetically engineered insert (Figure 2.4), which is accomplished using a disulfide bridge between two cysteine residues which flank the random 7-mer insert. There are approximately five copies of the pIII protein on a single phage. The genetically engineered segment of the pIII protein is the only difference between the different phage in the library. The libraries contain approximately 10^9 different inserts [32].

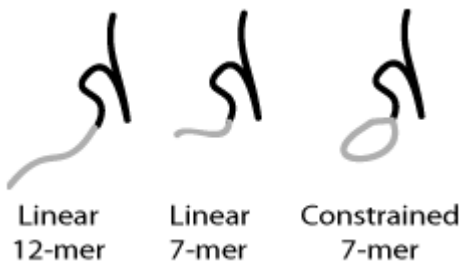


Figure 2.4: Representations of the three types of New England Biolabs fused peptide library configurations and lengths.

The phage genome has also been engineered to include a lac Z gene insert. This gene encodes for the β -galactosidase enzyme that is induced by the molecule Isopropyl- β -D-thiogalactopyranoside (IPTG) to hydrolyze the molecule 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal), an analogue of lactose. The product of this reaction is blue, allowing for quick identification and quantification of phage plaques grown on plates that contain both molecules. This quantification process, called titering, is accomplished by plating several dilutions of our phage solutions [32].

2.2: PHAGE DISPLAY EXPERIMENTAL

Although the selection process has several variables, the basic process is accomplished by exposing 10 μL of our library to the substrate and allowing it to interact for one hour, while rocking at room temperature. At the end of the exposure period, the non-binding phage are washed from the surface using a detergent solution, 0.1% TBST (Tris-buffered saline with 0.1% tween-20). In later rounds of selection, the detergent solution is increased to a concentration of 0.5% TBST to increase the selective pressure.

Two methods of amplification of binding phage have been investigated. The first is the standard acid/base elution method, followed by amplification in *E. coli* bacteria. The second is the single substrate screening method, which works under the assumption that the best binding phage will not be eluted from the surface even when exposed to a denaturing agent such as an acid or a base.

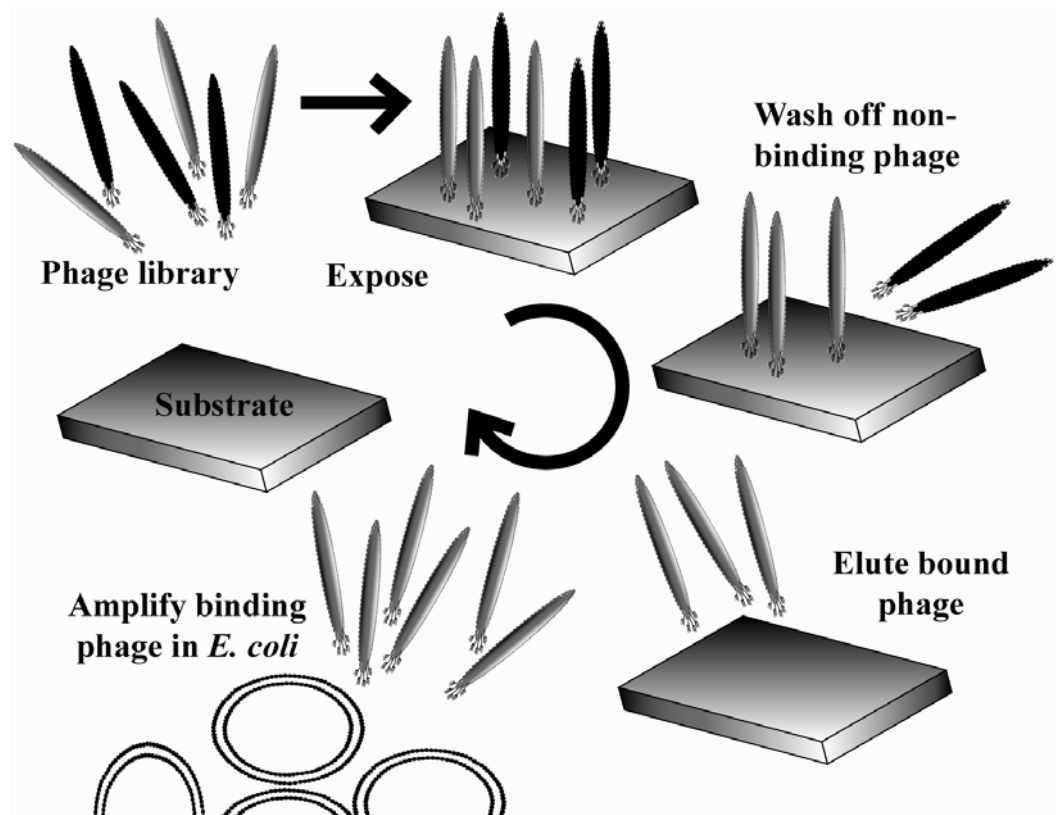


Figure 2.5: Evolutionary biopanning protocol steps [32].

2.2.1: Standard Elution Method

The standard acid elution method involves exposing the substrate with the bound phage to an acidic buffer (glycine-HCl pH 2.2) for 10 minutes, followed by removal and neutralization (Tris-HCl pH 9.1) of the phage and buffer. The eluted phage are then amplified for 4 hours in an early log ($A \approx .010$) *E. coli* culture and then purified by centrifugation to remove the bacteria. Each subsequent round of selection involves exposing a fresh substrate to the same amount of phage ($\sim 10^9$) and repeating the process with the amplified, higher surface-affinity phage. This

selection process is carried out 3-8 times in order to find convergence in the sequences found to bind to the substrate.

This method has been shown to give very good consensus, as will be presented later. This is probably due to the fact that acidic conditions will etch the surface of most of the materials screened here, effectively pulling off the bound phage. However, it has been hypothesized that tight binding phage may actually stabilize the surface when exposed to an acidic environment, preventing etching and thus removal of the phage. This theory was tested with both aragonite and calcite by adding the crystals from which phage had been eluted to an *E. coli* solution. These mixtures were then plated on IPTG/X-gal plates (Figure 2.6). The resulting bacterial lawns showed a significant number of plaques, many uncountable and overlapping near the substrate itself (approximately $10^2 - 10^4$ p.f.u.). Therefore, another screening method was designed to attempt to preserve these most-tightly bound phage.

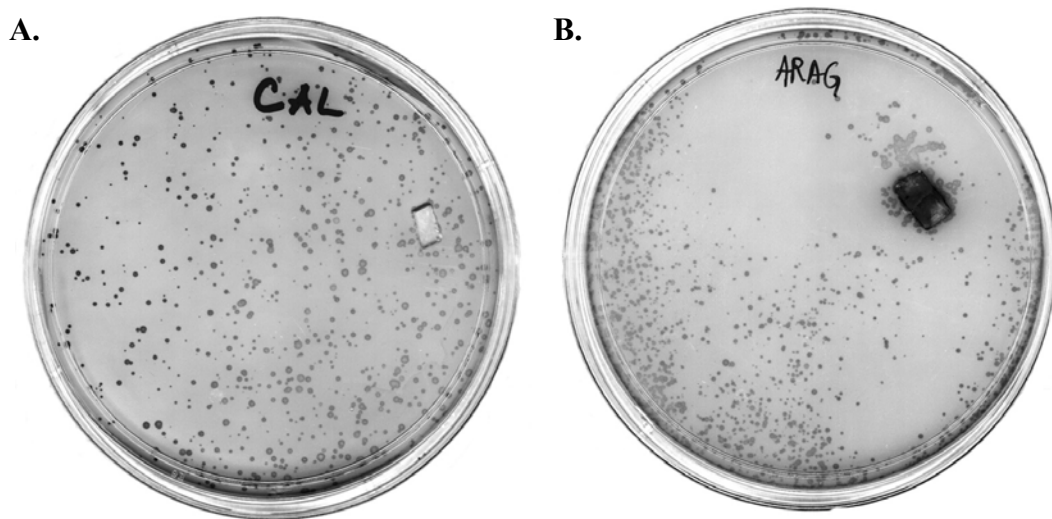


Figure 2.6: Titer plates showing phage plaques from phage still bound to A) calcite and B) aragonite after acid elution.

2.2.2: Single Substrate Screening Method

The single substrate panning and elution method is based on the fact that it has been shown that there is a large number of tightly bound phage that are discarded during each round of panning. In this method, the substrate is exposed to the same amount of library phage (10 μ L) and rocked for one hour. The non-binding phage are washed off the surface with 0.1% or 0.5% TBST, depending on the selection pressure desired, just as before. However, the bound phage are not eluted off the surface with acid, but are allowed to remain bound to the substrate. Then, the entire substrate is incubated with an early-log *E. coli* bacteria culture for 4 hours. During the amplification, theoretically, the phage should be easily eluted from the surface due to the shaking of the incubator, the elevated temperature (37°C), long period of time, large volume for dissociation (20 mL), and high

affinity of the phage for the *E.coli* f-pilus. After amplification, the substrate is then washed again with TBST to remove the bacteria and non-binding phage. Since the incubation with the bacteria both amplified the phage and exposed our substrate to the amplified phage population, the amplification/exposure step is then repeated 3-5 times. The substrate is washed with TBST between each amplification. In this method, the phage were eluted with bacteria. One microliter of mid-log (A= 0.500) bacteria was placed on the surface, immediately removed and diluted in more bacteria. Several dilutions were plated in order to get plates with well-separated plaques.

2.2.3: Determining Random Phage Peptide Sequences

Both methods result in phage with a higher affinity for the substrate used. These phage were then sequenced from the plates obtained after the elution step. Well-separated blue phage plaques were picked from the elution titer plates using a sterile toothpick. The toothpick was then placed into 1 mL of early log *E. coli* bacteria and allowed to amplify for 4 hours. The phage were then separated from the bacteria using centrifugation and purified. The phage were then lysed open with iodide buffer and the DNA was selectively precipitated in the presence of ethanol. The DNA was then sequenced using the -96 GIII sequencing primer (NEB). The sequences were then translated and analyzed for binding patterns.

2.3: SEQUENCE ANALYSIS

A secure, web-based database was developed by the author to store, edit, sort, and analyze the thousands of sequences found by our lab. The sequences can

be found in Appendix A and the source codes can be found in Appendix B. The sequences were analyzed in several complimentary methods presented here.

2.3.1: Amino Acid Percents by Position

First, the sequences were analyzed based on the position of each amino acid. Homologous amino acids were grouped by chemical similarities, and the residue or homology group that dominated each amino acid position was determined. The amino acid percents were also compared to the library percents to determine if the binding sequences were significant or if they were merely a reflection of the library.

2.3.2: Pattern Discovery Algorithm – Teiresias by IBM

The sequences were also analyzed using the IBM pattern recognition program, Teiresias. The Teiresias program finds the maximum number of patterns within a designated window of amino acids, by use of a two-phase algorithm [33]. The first phase, known as the scanning phase, creates a compilation of elementary patterns. The second, convolution, phase then recursively combines the original patterns into longer patterns, by overlapping the elementary patterns within the restrictive limits of the pattern window, adjacent sequences, equivalent amino acid definitions, and “don’t care” (symbolized by a period) amino acid positions. The final patterns are maximal in length and composition. Equivalent amino acid groups can be defined by the user and can reflect the user’s desire to find patterns based on the chemical similarities of different amino acids, structural similarities, or even overlapping substitution matrices, such as BLOSSUM 50, etc.

In these analyses, Teiresias was run using amino acid equivalencies. The window (W) was set to be 12 amino acids long, so as to encompass all of the library lengths. The number of literals to match (L) was set to a minimum of 3, so that even short patterns could be discovered. The number of sequences containing the pattern (K) was variable, since only the most often occurring patterns are desired, but longer significant patterns may only appear twice in a population. All other parameters were left as the default values, since we are looking to find only patterns and not families of homologous proteins.

The input sequences for Teiresias were filtered to negate the effects of preferred amplification of certain selected sequences. It is very often the case that during these screenings, one or two clone sequences will dominate later rounds of selection, with instances where one clone will be pulled out 10-20 times. It is possible that this is evidence of supreme binding and thus consensus, but it is also known that certain sequences can amplify better than others. For example, wild-type phage that occur due to excision of the DNA insert can easily take over a screening population due to its intrinsic ability to amplify better than phage with engineered pIIIs. Other contaminants have proven to amplify extremely well, costing us much time and energy decontaminating equipment and materials. Theoretically, it is not likely that the pIII insert affects *E. coli* infection due to the fact that the insert is at the N-terminus, which lies on the opposite face of the binding domains (f-pilus and TolA) necessary for infection [15]. However, it has been shown that diminished pIII production can lead to non-infectious phage particles that do not release from the bacterial membrane [30]. It is not likely that

this has a profound effect, since the phage particle termination and membrane release domains of the pIII are on the C-terminus [30]. However, the effects of pIII engineering on phage amplification have not been studied in depth and cannot be completely disregarded, especially in light of the many screenings that have resulted in dominate sequences.

A more convincing way to determine the binding motif is to look at the statistics of the selected population without including all of the duplicate sequences, unless the duplicate sequences were selected from completely separate screenings. Therefore, the analysis program filters out all duplicate sequences from each screening, allowing Teiresias to find the best patterns. By utilizing “discovery using equivalences,” Teiresias will find the both patterns using single amino acids and chemically or structurally similar amino acids. Therefore, Teiresias will allow us to define amino acids that are literally conserved in a pattern as well as amino acids that are conserved only in their general chemical or structural nature. Of course, it should be noted that the results of this analysis are bolstered by several screenings of the same substrate and by insuring that contamination is not a factor.

2.3.3: Statistical Significance of Discovered Patterns

The patterns found by Teiresias were statistically analyzed using the amino acid frequencies of each of the three New England BioLabs libraries to determine the probability of finding these patterns randomly in the library’s population. The patterns were then ranked in order of the highest significance, which corresponds to the lowest probability. First, the frequencies of each of the

user-defined equivalence sets were calculated by summing the constituent amino acid frequencies. This was done for each library, as the amino acid frequencies are slightly different.

$$\text{Frequency}_{\text{eq set}} = \sum(\text{Frequency}_{\text{AAs}}) \quad (2.1)$$

Then, the probability, ρ , of finding the specific pattern in each PhDX library was calculated. This was done by multiplying the frequencies of each literal in the pattern and then multiplying by the number of ways the specific pattern could be found in a sequence the length of the library insert.

$$\rho_{(\text{PhDX})} = \prod(\text{Frequency}_{\text{Literal}}) \times \# \text{ of patterns/sequence} \quad (2.2)$$

The overall probability, P , for finding the specific pattern at exactly the number of times it was found in each library, is calculated by the following equation, where a , b , and c are the number of times the pattern was found in each library, PhD12, PhD7, PhD7C, respectively:

$$P = \rho_{(\text{PhD12})}^a \times \rho_{(\text{PhD7})}^b \times \rho_{(\text{PhD7C})}^c \quad (2.3)$$

Finally, the significance of the patterns is easily ranked by converting the probability into an integer.

$$\text{Significance} = -\log(P) \quad (2.4)$$

2.3.4: Helical wheel

Some of the sequences deemed significant in the pattern discovery process were also analyzed for helical conformation probability. In the absence of time consuming and computer intensive molecular modeling for all of the significant peptides, a helical wheel evaluation can give a first hand approximation of the

likelihood the conserved amino acids are spatially oriented on the same side of the peptide.

The analysis is based on the fact that α -helical protein segments have 3.6 residues per turn. Therefore, when viewed along the helix axis, each residue's α -carbon will be rotated 100° from its neighbor's α -carbon. For 3-10, or extended helices, the rotation is about 120° . This type of analysis gives a quick identification of probable hydrophobic, hydrophilic, or target binding regions.

2.3.5: Homology to Biomineral Associated Proteins

The selected sequences for aragonite have also been compared to natural proteins found in aragonitic layers of mollusk organisms. Several proteins have been successfully extracted, although only a fraction of these have been characterized by sequencing. In phage display screenings, homology is difficult to define due to the fact that any library only represents a very small fraction ($\leq 0.0000015\%$) of all of the possible combinations of amino acids. Therefore, a program has been written by the author to compare the patterns found in the aragonite screening population to the sequences of known aragonite-associated proteins. The aragonite-associated proteins were scanned for the seven most significant patterns found for geological aragonite. These patterns were tallied and similarities between the screenings and the known proteins were noted.

REFERENCES:

1. NCBI, The Universal Virus Database of the International Committee on Taxonomy of Viruses: Index of Viruses from http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_inovi.htm. 2002.
2. Phage Display of Peptides and Proteins: A Laboratory Manual, ed. B.K. Kay, J. Winter, and J. McCafferty. 1996, San Diego: Academic Press, Inc.
3. Smith, G.P., Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science*, 1985. 228(4705): p. 1315-1317.
4. Whaley, S.R., et al., Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature (London)*, 2000. 405(6787): p. 665-668.
5. Lee, S.-W., et al., Ordering of quantum dots using genetically engineered viruses. *Science (Washington, DC, United States)*, 2002. 296(5569): p. 892-895.
6. Mao, C.B., et al., Viral assembly of oriented quantum dot nanowires. *Proceedings of the National Academy of Sciences of the United States of America*, 2003. 100(12): p. 6946-6951.
7. Belcher, A.M. and E.E. Gooch, Protein components and inorganic structure in shell nacre. *Biomaterialization*, 2000: p. 221-249.
8. Holliger, P., L. Riechmann, and R.L. Williams, Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 angstrom: Evidence for conformational lability. *Journal of Molecular Biology*, 1999. 288(4): p. 649-657.
9. Lubkowsky, J., et al., The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nature Structural Biology*, 1998. 5(2): p. 140-147.
10. Holliger, P. and L. Riechmann, A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure*, 1997. 5(2): p. 265-275.

11. Martz, E., Protein Explorer: Easy Yet Powerful Macromolecular Visualization. Trends in Biochemical Sciences, 2002. 27(February): p. 107-109 <http://proteinexplorer.org>.
12. Deng, L.W. and R.N. Perham, Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli. Journal of Molecular Biology, 2002. 319(3): p. 603-614.
13. Lubkowski, J., et al., Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure with Folding & Design, 1999. 7(6): p. 711-722.
14. Jacobson, A., Role of F pili in the penetration of bacteriophage f1. Journal of Virology, 1972. 10: p. 835-843.
15. Riechmann, L. and P. Holliger, The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E-coli. Cell, 1997. 90(2): p. 351-360.
16. Click, E.M. and R.E. Webster, Filamentous phage infection: required interactions with the TolA protein. Journal of Bacteriology, 1997. 179(20): p. 6464-6471.
17. Click, E.M. and R.E. Webster, The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. Journal of Bacteriology, 1998. 180(7): p. 1723-1728.
18. Trenkner, E., F. Bonhoeffer, and A. Gierer, The fate of the protein component of bacteriophage fd during infection. Biochemical and Biophysical Research Communications, 1967. 28: p. 932-939.
19. Frost, L.S., K. Ippenihler, and R.A. Skurray, Analysis of the Sequence and Gene-Products of the Transfer Region of the F-Sex Factor. Microbiological Reviews, 1994. 58(2): p. 162-210.
20. Marvin, D.A. and W. Folkhard, Structure of F-Pili - Reassessment of the Symmetry. Journal of Molecular Biology, 1986. 191(2): p. 299-300.
21. Glucksman, M.J., S. Bhattacharjee, and L. Makowski, 3-Dimensional Structure of a Cloning Vector - X-Ray-Diffraction Studies of Filamentous Bacteriophage-M13 at 7-Angstrom Resolution. Journal of Molecular Biology, 1992. 226(2): p. 455-470.

22. Grant, R.A., et al., Structure of the Filamentous Bacteriophage-F1 - Location of the Minor Coat Protein-a, Protein-C and Protein-D. *Journal of Biological Chemistry*, 1981. 256(1): p. 539-546.
23. Lin, T.C., R.E. Webster, and W. Konigsberg, Isolation and Characterization of the C-Protein and D-Protein Coded by Gene-Ix and Gene-Vi in the Filamentous Bacteriophage-F1 and Bacteriophage-Fd. *Journal of Biological Chemistry*, 1980. 255(21): p. 331-337.
24. Levensgood, S.K., W.F. Beyer, Jr., and R.E. Webster, TolA: a membrane protein involved in colicin uptake contains an extended helical region. *Proceedings of the National Academy of Sciences of the United States of America*, 1991. 88(14): p. 5939-43.
25. Gray, C.W., 3-Dimensional Structure of Complexes of Single-Stranded DNA-Binding Proteins with DNA - Ike and Fd Gene-5 Proteins Form Left-Handed Helices with Single-Stranded-DNA. *Journal of Molecular Biology*, 1989. 208(1): p. 57-64.
26. Russel, M. and P. Model, Genetic-Analysis of the Filamentous Bacteriophage Packaging Signal and of the Proteins That Interact with It. *Journal of Virology*, 1989. 63(8): p. 3284-3295.
27. Kazmierczak, B.I., et al., Piv, a Filamentous Phage Protein That Mediates Phage Export across the Bacterial-Cell Envelope, Forms a Multimer. *Journal of Molecular Biology*, 1994. 238(2): p. 187-198.
28. Linderoth, N.A., M.N. Simon, and M. Russel, The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science*, 1997. 278(5343): p. 1635-1638.
29. Lopez, J. and R.E. Webster, Morphogenesis of Filamentous Bacteriophage-F1 - Orientation of Extrusion and Production of Polyphage. *Virology*, 1983. 127(1): p. 177-193.
30. Rakonjac, J., J.-n. Feng, and P. Model, Filamentous Phage are Released from the Bacterial Membrane by a Two-step Mechanism Involving a Short C-terminal Fragment of pIII,. *Journal of Molecular Biology*, 1999. 289(5): p. 1253-1265.
31. Rakonjac, J. and P. Model, Roles of pIII in Filamentous Phage Assembly,. *Journal of Molecular Biology*, 1998. 282(1): p. 25-41.

32. Ph.D.-12™, Ph.D.-7™, Ph.D.-C7C™ Phage Display Peptide Library Kit Instruction Manuals: New England Biolabs.
33. Rigoutsos, I. and A. Floratos, Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm (vol 14, pg 55, 1998). Bioinformatics, 1998. 14(2): p. 229-229.

Chapter 3: Phage Display Results: Calcium Carbonate

Presented here are the results of the phage display sequence analyses for calcium carbonate. The patterns discovered for each grouping are arbitrarily limited due to the fact that there is an almost indefinite number of patterns that can be generated. Therefore, each group was analyzed individually for the most significant patterns in a reasonable sampling size of patterns. The probability is calculated using the frequency of the specific New England Biolabs (NEB) libraries used as described in chapter two. This calculated probability is the probability of finding a certain pattern X number of times, given the complexity of the library. The significance, described in chapter two as well, is directly derived from the probability and allows an easy ranking of the best patterns

Patterns will also be described by the use of some specific terms. Literals are the amino acids that are unchanging in a pattern, while wild cards, typically denoted by periods, are positions in the pattern that could be filled by any amino acid. The window of a pattern is the length of the pattern, including all literals and wild card positions. Equivalency brackets are groups of similar amino acids defined by the user.

Since we are looking for a binding motif for the unit cells of these inorganic materials, we must consider two possible classes of results; short, high-frequency patterns and longer, lower-frequency patterns. The random peptide fused to the pIII is considered to be flexible and accessible to the solvent and potential binding surfaces [1], so it is likely that the peptides do not have any

secondary or tertiary structure. However, it can be hypothesized that it may be plausible for these peptides to adopt at least a semi-stable structure either prior to or upon binding to the surface. If one considers this possibility, it would follow that these patterns would probably be longer in length, with relatively few wildcard positions and conserved over a smaller number of sequences. In contrast, peptides that were free to bend and twist, would give general patterns at high frequencies, presumably with several wild-card positions. The probability and significance method allows us to find both of these classes of sequences.

The equivalence sets used in these screenings are somewhat arbitrary and therefore, may require some explanation. The hydrophobic group contains all amino acids with hydrocarbon chain functional groups, alanine, isoleucine, leucine, valine and also glycine. This group typically constitutes 22-28% of all amino acids in the NEB phage display libraries. The basic group contains all amino acids with a positively charged amino functional group. These include arginine, lysine and histidine and are responsible for 10-15% of the amino acids in the libraries. The acidic residues are defined by their negatively charged carboxylic acid functional groups. These are aspartic acid and glutamic acid and make up 5-7% of the library residues. The amides include asparagine and glutamine and constitute 10-14% of the libraries. The hydroxyl group makes up 24-27% of the library residues and includes the alcohols serine, threonine and tyrosine. Methionine (~3%), proline (11-13%), and cysteine (0-0.5%) are considered unique. Phenylalanine and tryptophan are considered the aromatics, although there are members of other equivalence groups that are also aromatic.

Therefore, these groupings are more of a convenient way to describe the patterns than they are a stringent classification of functional groups.

Several screenings of each material were undertaken. Table 3.1 summarizes the selection conditions for the screenings carried out for calcium carbonate. Certain screening environments have been shown to make a difference in the resulting peptide sequences. These will be summarized in each of the following sections.

Screening pH	Elution	Library	Geo.	Aragonite			Calcite
				001	110	Nacre	104
7.5	acid	PhD12 PhD7 PhD7C	•				•
8.5	acid	PhD12 PhD7 PhD7C	• • •	•	•		•
8.5	base	PhD12 PhD7 PhD7C	•				•
8.5	bacterial	PhD12 PhD7 PhD7C	•	•	•	• • •	

Table 3.1: Summary of the screening conditions for the calcium carbonate materials. The dots indicate selection conditions under which these materials were screened. Geological aragonite was screened as a whole and with separate faces (001 and 110).

3.1: GEOLOGICAL ARAGONITE

Geological aragonite has given the best consensus results of any of the substrates screened, and has therefore been studied in great detail and under several screening conditions. First, the overall patterns from all of the screenings will be presented, as they tell an exceptionally compelling story. Then, the different screening parameters will be broken down and analyzed. These parameters include the three different phage display libraries, the pH of the screening, the specific aragonite face of binding, and the elution method.

3.1.1: Overall Geological Aragonite Patterns from Teiresias

As stated previously, geological aragonite gives the best patterns and statistics of any of the other substrates studied here. In fact, the patterns for aragonite are 46 orders of magnitude better than any other substrate screened here. The best patterns for aragonite can indeed be categorized into the two classes mentioned above: the high-frequency, short patterns and the lower-frequency, long patterns.

The short patterns dominate the list of aragonite patterns, which is not surprising, considering that there are 272 total sequences eluted from the geological aragonite screenings. It is overwhelming evident that the most significant patterns for geological aragonite are combinations of the [RKH] triplet. Another significant short pattern cluster is the [LAIVG], P, [WF], [RKH] pattern, or various combinations thereof. The importance of these patterns is reiterated in high significance, long patterns presented later. Table 3.2 shows the highest significance short patterns. These short patterns were grouped by

similarity, with all combinations of the ordered pattern within the window size calculated together (i.e. [RKH][RKH][RKH] or [RKH].[RKH][RKH] or [RKH]..[RKH][RKH], etc.) These significance values are amazingly high and represent a general combined motif that is apparently very effective in binding aragonite: [LAIVG]P[WF][RKH][RKH][RKH].

Significance	#	Short Patterns	Window
237	274	[RKH][RKH][RKH]	6
154	98	PW[RKH]	6
122	149	P[RKH][RKH]	6
113	97	[WF][RKH][RKH]	6
70	145	[LAIVG]P[RKH]	6
68	77	[LAIVG]P[WF]	6

Table 3.2: Overall short aragonite patterns found by IBM's Teiresias pattern recognition program. The significance is calculated from the observed frequencies of amino acids in the NEB libraries and the number (#) of times the pattern was found.

Also noteworthy are the sets of extended patterns (Table 3.3). As can be seen in the table, these patterns are found only a few times (≤ 5) in the general geological aragonite sequence population. However, these patterns have two important statistical properties. They are long patterns (≥ 7 literals), and they have relatively few equivalency brackets. These patterns emphasize the importance of the previous group of short patterns, as most of these include some combination of [LAIVG], P, W, as well as multiple [RKH] literals.

Significance	#	Long Patterns
60	4	LPPWKHKTSOVA
39	5	P.WKH...G[LAIVG][LAIVG]
38	6	LP.W..[RKH]..[LAIVG].A
37	5	PWKH[RKH].S..[LAIVG]
36	6	P.WKH.[STY]..[LAIVG]
36	5	LP.WKH.[STY]..[LAIVG]
31	2	LPKWQERQMLSA

Table 3.3: Long patterns for overall aragonite found by IBM's Teiresias pattern recognition program. The significance is calculated from the observed frequencies of amino acids in the NEB libraries and the number (#) of times the pattern was found.

By far the most interesting patterns within this group of geological aragonite patterns are the two full 12-mer sequences pulled from multiple screenings, A20: LPKWQERQMLSA (x2) and A21: LPPWKHKTSOVA (x4). These sequences were pulled out of separate screenings with multiple occurrences in each of these discrete screenings (Table 3.4). Since only counting multiple sequences once per separate screening negates the possible effect of preferential amplification, the significance of these sequences may be even higher than is apparent here. As seen in Table 3.4, A21 makes up 21.6% of all PhD12 sequences found for geological aragonite, while A20 makes up 2.3%.

Name	Sequence	#₁	#₂	#₃	#₄	%
A20	LPKWQERQMLSA	4	1			2.3
A21	LPPWKHKTSOVA	2	10	22	14	21.6

Table 3.4: Number of multiple sequences pulled out of each separate screening where A20 and A21 were found. The percentage is the percentage of total PhD12 sequences screened from geological aragonite.

Not only are these the only two sequences pulled out of multiple, separate screenings, but they are also very similar to each other. The both start with a hydrophobic leucine followed by proline(s). Then, there is a charged/polar region in the middle of the sequence. Roughly at the boundary between these first two distinct sections typically lies an aromatic tryptophan. Finally, these sequences are terminated in a relatively hydrophobic region. The major difference between the two sequences is the charged/polar region in the center. In A20, there are two amide groups, two positively charged residues and one negatively charged amino acid. In A21, this region has three positively charged groups followed by two hydroxyl groups. A21, while definitively being the most significant long pattern, also encompasses two of the most significant short patterns, the [RKH] triplet and the L, P, W, [RKH] pattern. Thus, this is the most significant sequence encountered for geological aragonite.

3.1.2: Geological Aragonite Analysis by Library

The sequences screened against geological aragonite are quite basic overall, but the percent analysis (Table 3.5) shows some interesting trends. The most dramatic differences from the library averages are the increase in basic amino acids, and the concurrent decrease in hydroxyl amino acids. This effect increases as the degrees of freedom decrease, from PhD12 to PhD7 to PhD7C.

AAs	PhD12	Lib. Avg.	Δ	PhD7	Lib. Avg.	Δ	PhD7C	Lib. Avg.	Δ
A	5.8	6.0	-0.2	6.9	6.7	0.2	0.6	6.5	-5.9
C	0.3	0.5	-0.2	1.1	0.6	0.5	0.0	0.0	0.0
D	2.4	2.8	-0.4	2.9	3.1	-0.2	1.1	4.1	-3.0
E	1.8	3.1	-1.3	2.3	2.2	0.1	0.6	3.1	-2.5
F	2.5	3.3	-0.8	2.3	2.2	0.1	0.6	2.1	-1.5
G	3.2	2.6	0.6	4.0	4.9	-0.9	2.3	2.2	0.1
H	8.1	6.3	1.8	11.4	3.7	7.7	22.3	6.9	15.4
I	2.9	3.4	-0.5	2.9	2.9	0.0	1.1	2.1	-1.0
K	7.6	2.8	4.8	6.3	2.4	3.9	25.1	3.8	21.3
L	8.0	9.3	-1.3	8.6	9.4	-0.8	2.3	9.6	-7.3
M	2.3	2.6	-0.3	0.0	2.7	-2.7	1.7	3.3	-1.6
N	3.3	4.6	-1.3	4.0	4.7	-0.7	9.7	6.4	3.3
P	14.3	12.2	2.1	17.7	13.1	4.6	9.1	10.7	-1.6
Q	4.7	5.1	-0.4	4.0	5.1	-1.1	4.6	7.1	-2.5
R	4.5	4.7	-0.2	8.0	3.7	4.3	4.0	4.3	-0.3
S	7.9	10.0	-2.1	6.3	11.8	-5.5	4.6	8.6	-4.0
T	8.3	11.1	-2.8	5.7	10.8	-5.1	3.4	13.1	-9.7
V	3.9	3.9	0.0	2.3	4.2	-1.9	1.7	1.9	-0.2
W	5.7	2.2	3.5	2.3	1.8	0.5	3.4	1.9	1.5
Y	2.2	3.6	-1.4	1.1	3.9	-2.8	1.1	2.4	-1.3
X	0.4	0.0	0.4	0.0	0.0	0.0	0.6	0.0	0.6
LAIVG	23.8	25.2	-1.4	24.6	28.1	-3.5	8.0	22.3	-14.3
RKH	20.2	13.8	6.4	25.7	9.8	15.9	51.4	15.0	36.4
WF	8.2	5.5	2.7	4.6	4.0	0.6	4.0	4.0	0.0
STY	18.4	24.7	-6.3	13.1	26.5	-13.4	9.1	24.1	-15.0
DE	4.1	5.9	-1.8	5.1	5.3	-0.2	1.7	7.2	-5.5
NQ	8.0	9.7	-1.7	8.0	9.8	-1.8	14.3	13.5	0.8

Table 3.5: Summary of geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

The consensus aspect of this analysis is important to note. Table 3.6 shows the percent position summary for all three libraries. Without making any

concessions for starting positions, there is a strong pattern that emerges naturally for the PhD12 library: [LAIVG]P[RKH]W[RKH][RKH] with a primarily hydroxyl and hydrophobic C-terminus. The percentages for each position are high (most between 30% and 45%), considering that there are a large number of sequences in this population. As expected from the overall percentages (Table 3.5, above) found in the geological aragonite screenings, the basic nature of the percent position analysis increases in the PhD7 and PhD7C screenings, with the PhD7C sequences being highly basic (28-64%) in each position.

Position	PhD12	%	PhD7	%	PhD7C	%
1	LAIVG	45.3	LAIVG	24.0	C	100.0
2	<i>P</i>	35.2	RKH/P	24.0	RKH	64.0
3	RKH	35.8	RKH	36.0	RKH	48.0
4	<u>WF</u>	22.6	LAIVG	48.0	RKH	56.0
5	RKH	34.6	RKH	32.0	RKH	52.0
6	RKH	30.2	RKH	36.0	RKH	28.0
7	STY	25.2	RKH	28.0	RKH	56.0
8	STY	25.2			RKH	56.0
9	LAIVG	28.3			C	100.0
10	LAIVG	28.9				
11	LAIVG	32.7				
12	LAIVG	34.6				

Table 3.6: Highest percentages for each position in each of the three libraries screened against geological aragonite. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, pralines are italicized, aromatic groups are underlined.

The sequences from the geological aragonite PhD7 screening tend to be similar to the overall patterns for geological aragonite. The percent position summary (Table 3.6) shows the general trends for this screening. The basic

groups (+16%) and prolines (+4.6%) are over-represented compared to the library averages. Conversely, there is a lack of hydrophobic, amides, and, most dramatically hydroxyl groups (-13%). Individually, all of the basic amino acids, are found at higher rates. However, it is the histidines (+7.7%) that are mostly responsible for this trend as they are found at roughly twice the rate of arginine (+3.9%) and lysine (+4.3%) and three times the library average. A typical PhD7 geological aragonite sequence based on the percent position summary (Table 3.6), would look like: [LAIVG]P[RKH][LAIVG][RKH][RKH][RKH]. Not surprisingly, this is quite similar to the overall trends for geological aragonite and the PhD12 library. The most noticeable difference is the lack of tryptophan.

The sequences found in the PhD7C screening are even more dominated by basic amino acids. In fact, the percent position analysis (Table 3.5) shows that basic amino acids constitute over 50% of all of the amino acids found in these sequences, as compared to the typical 15% found in the library. According to the percent position summary (Table 3.6), there is a 48-64% chance of finding a basic group in six of the seven positions in the sequence. The exception to this is the 5th position, which is typically a basic group (28%), a proline (24%) or an amide (20%). The increase in basic groups is mainly due to the dramatic increase in histidine (+15%) and lysine (+21%), as the percentage of arginine residues is similar to the library average (-0.3%). This increase in basic groups occurs at the expense of hydroxyl, hydrophobic and acidic residues, in that order.

Significance	#	PhD12 Patterns
60	4	LPPWKHKTSOVA
43	18	W[RKH][RKH]
42	22	P..[RKH].....[LAIVG]
41	16	P.W.....[LAIVG]
41	18	W[RKH]...[STY]

Significance	#	PhD7 Patterns
13	2	H.K[LAIVG][RKH]H[LAIVG]
12	4	K.[RKH][RKH]
12	4	[RKH][RKH]..H
10	3	K.H[RKH]
10	2	APK..[LAIVG]P

Significance	#	PhD7C Patterns
28	16	[RKH][RKH][RKH]
27	13	H[RKH][RKH]
22	12	[RKH].[RKH][RKH]
22	11	[RKH][RKH]..[RKH]
22	10	[RKH][RKH]...[RKH]

Table 3.7: High significance patterns for each library for geological aragonite found by IBM's Teiresias pattern recognition program. The significance is calculated from the observed frequencies of amino acids in the NEB libraries and the number (#) of times the pattern was found.

The percent position analysis is an excellent preview of the Teiresias patterns. In fact, most of the top aragonite patterns (Table 3.2 and 3.3) fit into the representative sequence based on percent position. Table 3.7 gives the highest-ranking patterns for each library. These patterns do converge with the overall geological aragonite patterns, both extended and short. Therefore, we might conclude that while a long pattern may have significant binding potential, shorter,

higher frequency patterns may be able to do the job just as well. Furthermore, these binding motifs may be flexible, explaining why there are so many high significance short patterns.

The patterns found by Teiresisas for the PhD7 geological aragonite sequences (Table 3.7) are also quite similar to some of the shorter patterns found for the overall geological aragonite population. They are dominated by basic groups, with some having the familiarly positioned P, W or [LAIVG].

The patterns found for the PhD7C library screened against geological aragonite are, not surprisingly, a direct reflection of the basic group dominance. In fact, almost all of the top patterns are some combination of [RKH] residues. Table 3.7 shows the top five patterns, but this trend continues far into the list of patterns. In the top 50 patterns, only 3 have non-[RKH] residues, with those residues being the familiar W and P and also N.

3.1.3: Effect of Screening pH for Geological Aragonite Sequences

The pH of the screenings does appear to have a slight influence on the type of sequences found. Calcium carbonate is sensitive to etching at relatively neutral pHs. In order to ensure that the surface was not etched away under the binding sequences, most screenings were done at pH 8.5. This implies that the surface is most likely carbonate rich, since any surface calcium atoms are susceptible to dissociation in the basic aqueous environment. Therefore, it is not surprising that the binding sequences at pH 8.5 are enriched in basic groups, mostly lysine, and slightly deficient in acidic groups. However, this trend also

applies to the pH 7.5 binding amino acids as well, but the difference from the general library averages is slightly lower.

It is also interesting to note the concurrent decrease in hydroxyl amino acids, especially serine. As the pH increases, the number of basic groups increases, while the hydroxyl residues decrease at nearly the same rate. It can almost be hypothesized that these hydroxyl amino acids may act as binding residues to the calcium atoms, through the electronegative hydroxyl oxygen. While it is only a hypothesis, it may help explain central to c-terminal hydroxyls found to bind to aragonite.

AAs	Lib. Avg.	All PhD12	Δ	pH 7.5	Δ	pH 8.5	Δ
A	6.0	5.8	-0.2	6.7	0.7	5.5	-0.5
C	0.5	0.3	-0.2	0.4	-0.1	0.2	-0.3
D	2.8	2.4	-0.4	2.1	-0.7	2.5	-0.3
E	3.1	1.8	-1.3	2.3	-0.8	1.6	-1.5
F	3.3	2.5	-0.8	1.9	-1.4	2.7	-0.6
G	2.6	3.2	0.6	2.5	-0.1	3.4	0.8
H	6.3	8.1	1.8	8.1	1.8	8.1	1.8
I	3.4	2.9	-0.5	2.3	-1.1	3.2	-0.2
K	2.8	7.6	4.8	7.1	4.3	7.8	5.0
L	9.3	8.0	-1.3	7.3	-2.0	8.2	-1.1
M	2.6	2.3	-0.3	2.9	0.3	2.1	-0.5
N	4.6	3.3	-1.3	3.3	-1.3	3.3	-1.3
P	12.2	14.3	2.1	13.8	1.6	14.5	2.3
Q	5.1	4.7	-0.4	4.4	-0.7	4.8	-0.3
R	4.7	4.5	-0.2	3.3	-1.4	4.9	0.2
S	10.0	7.9	-2.1	9.8	-0.2	7.2	-2.8
T	11.1	8.3	-2.8	7.9	-3.2	8.5	-2.6
V	3.9	3.9	0.0	4.0	0.1	3.9	0.0
W	2.2	5.7	3.5	7.1	4.9	5.3	3.1
Y	3.6	2.2	-1.4	2.5	-1.1	2.1	-1.5
X	0.0	0.4	0.4	0.4	0.4	0.4	0.4
LAIVG	25.2	23.8	-1.4	22.7	-2.5	24.2	-1.0
RKH	13.8	20.2	6.4	18.5	4.7	20.7	6.9
WF	5.5	8.2	2.7	9.0	3.5	7.9	2.4
STY	24.7	18.4	-6.3	20.2	-4.5	17.8	-6.9
DE	5.9	4.1	-1.8	4.4	-1.5	4.1	-1.8
NQ	9.7	8.0	-1.7	7.7	-2.0	8.1	-1.6

Table 3.8: Summary of PhD12 geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

Although there are slight differences between the amino acid percentages in the different pH screenings, the patterns of the two types of screenings are not

significantly different. Table 3.9 shows that the order of highest percentage amino acids is not significantly different, nor are the percentages at which these amino acids occur in these positions. Although it is not evident in this summary chart, the full percent position charts show two notable aspects. First, the second most common amino acid group in position 4 in the pH 8.5 screenings is [WF] at 20.2%. This gives more credence to the LP[RKH]W[RKH][RKH] motif. Secondly, the [STY] equivalency set is the first or second most commonly encountered amino acid in 10 of the 12 positions at pH 7.5, while the same is true for pH 8.5 at only 5 of the 12 positions.

	All PhD12	%	pH 7.5	%	pH 8.5	%
1	LAIVG	45.3	LAIVG	45.0	LAIVG	45.4
2	<i>P</i>	35.2	<i>P</i>	45.0	<i>P</i>	31.9
3	RKH	35.8	RKH	30.0	RKH	37.8
4	<u>WF</u>	22.6	<u>WF</u>	30.0	RKH	21.8
5	RKH	34.6	RKH	40.0	RKH	32.8
6	RKH	30.2	RKH	25.0	RKH	31.9
7	STY	25.2	STY	30.0	STY	23.5
8	STY	25.2	LAIVG	22.5	STY	26.9
9	LAIVG	28.3	LAIVG	22.5	LAIVG	30.3
10	LAIVG	28.9	LAIVG	30.0	LAIVG	28.6
11	LAIVG	32.7	LAIVG	40.0	LAIVG	30.3
12	LAIVG	34.6	LAIVG	25.0	LAIVG	37.8

Table 3.9: Highest percentages for each PhD12 position for the sequences screened against geological aragonite at different pHs. The average of all of the geological aragonite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

The patterns for the different pH screenings are also similar (Table 3.10). The major distinction between the two types of screenings is the magnitude of the significance. However, this is a direct reflection of the number of sequences examined for each of the different pH screenings. There are 44 different sequences sampled from the pH 7.5 screening, while there are 119 sampled from the pH 8.5 screenings. Therefore, the highest significance patterns for pH 7.5 are long, many-literal patterns. Not surprisingly, most of these fit within the best-percent position template. It is also interesting to note that while they were only counted once in the pH 7.5 screening, A20 was found four times and A21 was found twice. The importance of the A20 pattern is reflected in the highest-ranking pattern for pH 7.5, since it includes one other sequence that has almost 60% sequence identity with A20. The pH 8.5 sequence patterns from the PhD12 library are quite similar to the overall PhD12 patterns, due to the fact that they make up the majority of the total geological aragonite sequences. Like the overall patterns, they too reiterate the importance of the A21 sequence, the P, W, [RKH] motif and the [RKH] triplet motif.

Significance	#	pH 7.5 Patterns
17	2	LPKW....MLS
17	2	KP..[RKH].YS.[LAIVG]LS
16	6	W[RKH]....P
16	7	[WF][RKH]....P
16	7	P.....[STY].[LAIVG][LAIVG]
Significance	#	pH 8.5 Patterns
45	3	LPPWKHKTSOVA
36	15	W[RKH][RKH]
34	9	WK[RKH]..[STY]
34	17	[WF][RKH][RKH]
33	13	P.W.[RKH]

Table 3.10: Top patterns for sequences from the PhD12 library screening of geological aragonite at pH 7.5 and pH 8.5. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.

3.1.4: Face-Specific Elution for Geological Aragonite

Aragonite binding was further investigated using face-specific elution methods to explore the differences in binding to the two major faces of geological aragonite, (110) and (001). The face-specific screenings were done twice, once using the single substrate screening method and once using the standard acid elution method. The differences in patterns from each of the elution methods are significant and are discussed in the next section. The (110) face is the twinned face of the geological aragonite crystal, therefore there are six such faces in the geological pseudohexagonal crystal. The (001) face is the terminating face perpendicular to the (110) face and is also the same face that constitutes the nacreous surface of shell aragonite. Each of the geological aragonite crystal sides

is roughly the same surface area. So, depending on how it is cut from the bulk crystal the ratio of (110) to (001) surface area is approximately 6:1.

Significance	#	110 Surface Patterns
30	2	LPPWKHKTSQVA
22	3	LP.WKH.[STY].[LAIVG]
17	3	P.WK.....V[LAIVG]
17	4	P.WK.....[LAIVG]
17	5	P..[RKH][RKH].[STY].[LAIVG]
17	3	LP.WKH
16	4	P..[RKH][RKH].[STY].[LAIVG][LAIVG]
15	3	P.WK.....V
15	3	PW[RKH][RKH]...[LAIVG][LAIVG]
15	3	P..[RKH][RKH][STY]S[LAIVG].[LAIVG]

Table 3.11: Top ten patterns for sequences from the geological aragonite (110) screenings. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.

Table 3.11 shows the top 10 patterns for (110) geological aragonite. There are several long patterns, and most patterns include the A21 sequence. The percent position analysis (Table 3.13) shows that tryptophan and lysine appear in this population at a higher frequency than in the library. The general (110) pattern includes residues: L, P, W, [RKH]₂, [STY]₁₋₂, and [LAIVG]₁₋₂, roughly in that order. Not coincidentally, this general pattern is quite similar to the general pattern found for the overall geological aragonite sequences. This is likely due to the high (110):(001) surface area ratio. A21 is even more significant in these screenings in that it was found on the (110) surface in both the acid elution screening and the single substrate screening. In fact, it was found several times in each screening. A21 and another sequence found twice in the acid elution screening are listed in Table 3.12 below. It seems extremely significant that both

of these sequences have the P.WK....V[IA] motif and account for over 60% of all sequences found for 110.

Name	Sequence	Std	SSS	%
A21	LPPWKHKTSOVA	14	16	57.7
110-A-2-7	HPTWKTPIQNVI	2	0	3.8

Table 3.12: Sequences found multiple times in the same geological aragonite (110) screening and the number of times found. Std is the standard acid elution screening and SSS is the single substrate screening. The percentage is calculated using the total number of sequences selected against (110) geological aragonite.

AA	Lib. Avg.	110 Total	Δ	001 Total	Δ
A	6.0	6.9	0.9	4.4	-1.6
C	0.5	0.0	-0.5	0.6	0.1
D	2.8	2.9	0.1	1.7	-1.1
E	3.1	2.5	-0.6	0.6	-2.5
F	3.3	1.1	-2.2	3.3	0.0
G	2.6	2.9	0.3	5.6	3.0
H	6.3	5.8	-0.5	7.2	0.9
I	3.4	4.3	0.9	4.4	1.0
K	2.8	7.6	4.8	5.0	2.2
L	9.3	9.1	-0.2	8.9	-0.4
M	2.6	3.6	1.0	2.8	0.2
N	4.6	3.6	-1.0	3.9	-0.7
P	12.2	11.6	-0.6	13.3	1.1
Q	5.1	3.6	-1.5	7.8	2.7
R	4.7	2.9	-1.8	2.8	-1.9
S	10.0	11.2	1.2	7.8	-2.2
T	11.1	9.4	-1.7	10.6	-0.5
V	3.9	2.9	-1.0	3.9	0.0
W	2.2	5.8	3.6	2.2	0.0
Y	3.6	1.8	-1.8	3.3	-0.3
X	0.0	0.4	0.4	0.0	0.0
LAIVG	25.2	26.1	0.9	27.2	2.0
RKH	13.8	16.3	2.5	15.0	1.2
WF	5.5	6.9	1.4	5.6	0.1
STY	24.7	22.5	-2.2	21.7	-3.0
DE	5.9	5.4	-0.5	2.2	-3.7
NQ	9.7	7.2	-2.5	11.7	2.0

Table 3.13: Summary of the (110) and (001) geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

The patterns found for the (001) surface of geological aragonite are listed in Table 3.14. The patterns discovered for the (001) aragonite sequences are distinctly different than the (110) sequences, despite the fact that the A21

sequence was also found in this screening. It was found on the (001) surface only in the single substrate screening where both the (110) and (001) surfaces were screened simultaneously. A21's presence on the (001) surface may be an artifact of the strong binding to the (110) surface, which increased its amplification. However, A21 does contain a motif seen in several other (001) sequences: the multiple proline motif.

Significance	#	001 Surface Patterns
12	2	WLQPP...[RKH]
10	2	WLQPP
9	2	[DE]..QP.LT
9	2	L...K..T.[LAIVG][LAIVG]
8	3	[DE]..Q..L
8	3	[LAIVG]PP[WF]
8	2	[DE]Q.Q..L
8	3	N.[STY][LAIVG][NQ]
8	2	P..[RKH][RKH].[STY].[LAIVG][LAIVG]
8	2	D.[RKH]Q..L

Table 3.14: Top ten patterns for sequences from the geological aragonite 001 screenings. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.

Multiple prolines account for 75% of all of the (001) sequences (Table 3.15) when multiple occurrences of sequences are taken into account. More specifically, multiple prolines coupled with W, K, and Q make up the majority of this subset. Not only is the frequency of proline high (17-45%) for each of these sequences in Table 3.15, but the frequency of proline in the entire (001) population is double the expected frequency from this library.

Name	Sequence	#	%
001-A-2-1	IPPWLQPPLPQK	13	36.1
A21	LPPWKHKTSOVA	6	16.7
001-A-3-1	GKWKLPPFPMTH	5	13.9
Atop-5	VSSLPRATHPPC	1	2.8
Atop-10	FIPGAPTLVQTG	1	2.8
Atop-11	DLRQPALTLHPG	1	2.8
Atop-14	QWLQPPSQGRSY	1	2.8
Total multiple proline sequences		28	77.9
Total 001 proline residues		106	24.5
Library proline frequency			12.2

Table 3.15: Sequences found multiple times in the same geological aragonite 001 screening, the number of times found, and the percentage of the total sequences selected against (001) geological aragonite.

Polyproline peptides are well-known biological recognition motifs [2] and even form a special secondary structure known as the left-handed polyproline II helix (PPII). There is a right-handed helix, but it is only found in apolar solvents and not yet in any known natural proteins [3]. The PPII structure is a flexible helix due to the fact that there is no backbone hydrogen in proline available for intra- or inter-molecular hydrogen bonding. The rise per residue of these helices is greater than that of alpha helices. This causes the peptide to be a long, stretched-out helix, with the backbone of the peptide itself more exposed to the solvent or binding surface than most secondary structures [3]. In fact, the backbone carbonyls extend from the helix at regular intervals [2] and are known to hydrogen bond with water molecules in a regular pattern [3]. In addition, these

polyproline recognition sequences are known to be able to bind to relatively flat binding surfaces due to its unique structure. Furthermore, these PPII structures have a similar topography $N \rightarrow C$ and $C \rightarrow N$, since the helix has a two-fold pseudosymmetry about an axis perpendicular to the helical axis [2], allowing binding to be less energetically hindered. As recognition sequences, these PPII helices are frequently found on the surface of proteins. Unlike other secondary structures, PPII helices do not often span more than a few residues (usually 4 - 6), since they are not stabilized by hydrogen bonding.

Other proteins that exhibit PPII helices and are of interest to this research include collagen as well as at least one antifreeze glycoproteins [3]. Collagen is a major component of mineralized bone. Bone mineral grows in parallel (100)-oriented plate-like crystals within the collagen fibril [4], similar to the structure of aragonite plates in nacre. Furthermore, aragonite has similar lattice parameters, ionic positions and ionic contents [5] as hydroxylapatite, the carbonated form of which is the major inorganic component of human teeth and bone. Additionally, nacreous aragonite has been converted into hydroxylapatite using hydrothermal methods [5]. It has also been found that protein degradation of filaments found in the interlamellar sheet of abalone has shown that these proteins primarily contain glycine and proline, which are the major components of collagen [6].

3.1.5: Effect of Elution Method for Geological Aragonite Sequences

The elution method used in these screenings is an important factor in achieving consensus. While there are many methods to remove proteins from the surfaces to which they bind, only three were used in these screenings (Table

3.16). Bacteria elution refers to the single substrate screening method. Since acid elution was the most often employed method to elute the binding proteins, the total percentages of amino acids of this subset is almost identical to the overall percentages. There is the familiar over-representation of [RKH] and [WF].

Elutions using a base were not significantly different from those using an acid. The frequencies of lysines, histidines, prolines and tryptophans were higher than the library, while the frequency of threonine and serine were lower. The trends are the same as the acid-eluted sequences, but the magnitude of the differences is approximately double that of the acid-elutions.

The bacteria-eluted sequences deviate quite a bit from the acid and base elutions. The frequencies of all of the amino acids are quite similar to the average frequencies found for the PhD12 library in general. In fact, no residue or equivalence group deviates more than $\pm 3\%$ from the library values. This suggests that the single substrate screening/bacterial elution method is not as effective as changing the pH to disrupt the best-binding sequences or the selection method itself may not provide enough selection pressure to achieve consensus.

AAs	Lib. Avg.	All PhD12	Δ	Acid	Δ	Base	Δ	Bacterial	Δ
A	6.0	5.8	-0.2	6.2	0.2	4.3	-1.7	5.8	-0.2
C	0.5	0.3	-0.2	0.3	-0.2	0.0	-0.5	0.2	-0.3
D	2.8	2.4	-0.4	2.3	-0.5	2.2	-0.6	2.7	-0.1
E	3.1	1.8	-1.3	2.0	-1.1	1.8	-1.3	1.3	-1.9
F	3.3	2.5	-0.8	2.4	-0.9	2.2	-1.1	2.7	-0.6
G	2.6	3.2	0.6	2.9	0.3	2.9	0.3	4.2	1.6
H	6.3	8.1	1.8	8.4	2.1	12.3	6.0	4.8	-1.5
I	3.4	2.9	-0.5	3.0	-0.4	1.4	-2.0	3.5	0.1
K	2.8	7.6	4.8	8.4	5.6	11.2	8.4	3.5	0.7
L	9.3	8.0	-1.3	7.6	-1.7	6.5	-2.8	9.8	0.5
M	2.6	2.3	-0.3	2.7	0.1	1.4	-1.2	1.9	-0.7
N	4.6	3.3	-1.3	3.2	-1.4	2.5	-2.1	4.0	-0.6
P	12.2	14.3	2.1	13.7	1.5	18.5	6.3	13.3	1.1
Q	5.1	4.7	-0.4	4.3	-0.8	1.8	-3.3	7.1	2.0
R	4.7	4.5	-0.2	4.5	-0.2	3.6	-1.1	5.0	0.3
S	10.0	7.9	-2.1	7.7	-2.3	6.9	-3.1	8.8	-1.3
T	11.1	8.3	-2.8	7.8	-3.3	5.1	-6.0	11.5	0.4
V	3.9	3.9	0.0	3.6	-0.3	5.4	1.5	4.0	0.1
W	2.2	5.7	3.5	6.5	4.3	7.6	5.4	2.7	0.5
Y	3.6	2.2	-1.4	1.8	-1.8	1.8	-1.8	3.3	-0.3
X	0.0	0.4	0.4	0.6	0.6	0.4	0.4	0.0	0.0
LAIVG	25.2	23.8	-1.4	23.2	-2.0	20.7	-4.5	27.3	2.1
RKH	13.8	20.2	6.4	21.4	7.6	27.2	13.4	13.3	-0.5
WF	5.5	8.2	2.7	8.9	3.4	9.8	4.3	5.4	-0.1
STY	24.7	18.4	-6.3	17.4	-7.3	13.8	-10.9	23.5	-1.2
DE	5.9	4.1	-1.8	4.3	-1.6	4.0	-1.9	4.0	-1.9
NQ	9.7	8.0	-1.7	7.6	-2.1	4.3	-5.4	11.0	1.3

Table 3.16: Summary of the different elution methods for the PhD12 geological aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

Table 3.17 shows the highest percentage representative sequences for each of the elution methods and the overall PhD12 sequences for geological aragonite.

As previously stated, the acid-eluted sequences are almost identical to the overall PhD12 sequences eluted from geological aragonite due to the fact that they make up the majority of the overall sequences. The base-eluted sequences are also very similar. The trend is the same as the acid-eluted sequences, with minor exceptions, in which the second most frequent residue in these cases matches the acid eluted residues. The bacteria-eluted sequences show quite a different trend, with the exception of the hydroxyl and hydrophobic C-terminus. There are still some charged residues in the center of the sequence. However, the basic groups occur at a slightly lower frequency than in the acid and base elutions. In addition, the amide content of the bacterial elution sequences is somewhat greater than the acid or base elutions. This trend can be seen in the percent summary (Table 3.16) and in the patterns generated from the bacterial elution sequences (Table 3.17).

	All PhD12	%	Acid Elution	%	Base Elution	%	Bacterial Elution	%
1	LAIVG	45.3	LAIVG	49.0	LAIVG	43.5	LAIVG	37.5
2	<i>P</i>	35.2	<i>P</i>	44.8	<i>P</i>	34.8	LAIVG/STY	25.0
3	RKH	35.8	RKH	34.4	K	56.5	RKH	27.5
4	<u>WF</u>	22.6	<u>WF</u>	29.2	RKH	30.4	NQ	25.0
5	RKH	34.6	RKH	39.6	RKH	47.8	LAIVG	30.0
6	RKH	30.2	RKH	33.3	H/P	26.1	RKH	25.0
7	STY	25.2	STY	26.0	RKH	21.7	STY	27.5
8	STY	25.2	LAIVG	22.9	STY	26.1	STY	40.0
9	LAIVG	28.3	LAIVG	22.9	LAIVG	34.8	LAIVG	37.5
10	LAIVG	28.9	LAIVG	31.3	<i>P</i>	39.1	LAIVG	25.0
11	LAIVG	32.7	LAIVG	35.4	RKH	26.1	LAIVG	35.0
12	LAIVG	34.6	LAIVG	34.4	LAIVG	30.4	LAIVG	37.5

Table 3.17: Highest percentages for each PhD12 position for the sequences screened against geological aragonite using the three different elution methods. The average of all of the geological aragonite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

The patterns from the different elution screenings are given in Table 3.18. The acid elution sequences show the familiar L, P, W, K motif appearing, at least in part, in all of the most significant patterns. As a result of the high number of acid-eluted sequences for aragonite and the exceptional consensus, the significance values are extremely high.

The base-eluted sequence patterns are not that different than the acid-eluted sequence patterns. The significance values are not as high, but the W, [RKH] motif is definitely evident as well as the multiple [RKH] motif.

The bacteria-eluted sequence patterns are quite different. There is no evidence of the familiar W, [RKH]; L, P, W, K; or multiple [RKH] motifs. Instead it appears that the most significant motifs include [NQ], [STY], [LAIVG]. It should also be noted that the significance values are also low for the bacteria eluted sequence patterns, despite the fact that there were almost double the number of sequences eluted using bacteria compared to those eluted using base. Thus, having a larger population does not necessarily guarantee appreciably high significance figures.

Despite, the theoretical benefits of single substrate screening, in practice it is difficult to achieve consensus. Most likely, this method does not present a strong enough selection pressure to achieve evolutionary conditions. With that said, however, it should be noted that even the bacterial elution/single substrate screening method was successful in finding the A21 sequence on the surface of geological aragonite. It does not appear in the patterns, due to the fact that the significance calculation does not take into account the multiple occurrences. However, it was found 16 times on the (110) surface and 6 times on the (001) surface, when the surfaces were screened together. Therefore, even under low selection pressure, A21 is found to be extremely significant in binding to aragonite.

Significance	#	Acid Elution Patterns
45	3	LPPWKHKTSGVA
36	19	P..[RKH]....[LAIVG]
36	14	P.W....[LAIVG]
33	13	P.W.[RKH]
33	6	P.WKH....[LAIVG]

Significance	#	Base Elution Patterns
17	4	WKH..[STY]
15	5	[RKH]WK
15	4	[RKH]W..[RKH]....P
15	5	HK..H
15	2	S[STY]HW[RKH].[LAIVG]PR

Significance	#	Bacterial Elution Patterns
12	2	[LAIVG].TP[WF]P[LAIVG]....[LAIVG]
11	2	[NQ]..I..G..R[STY]
11	9	[NQ].[STY][LAIVG]
11	2	VQ[NQ]T[LAIVG]....[STY]
11	2	D.H....QP..[STY]

Table 3.18: Patterns from the PhD12 geological aragonite sequences eluted with acid, base and bacteria, respectively. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.

3.1.6: Helical Wheel for Statistically Significant Aragonite Peptides

The helical wheel analysis is a simple method of analyzing potential peptide structural motifs. Often it is the case that natural helical peptides will have residues with differing binding characteristics on opposite sides of the helix. Therefore, this quick analysis of statistically significant peptides can help us determine if this is the case, and if so, which type of helical structure is possible upon binding to the substrate. The peptides are viewed along the helical axis, with the α -helix having residues spaced 100° apart and the 3-10 helix having

residues spaced 120° apart. Here, we look at three interesting peptides A20, A21, and A1, all found to bind to geological aragonite. In addition, these peptides all have a proline near the N-terminus, a location favored in α -helices. The N-cap + 1 position of an α -helix favors proline residues $\sim 2.6:1$ [7]. In both A20 and A21, there is a proline in the N + 1 position, while in A1, the proline lies at the N + 3 position.

The A20 and A21 peptide sequences have been shown above to be extremely statistically significant in binding to aragonite. The A1 peptide sequence is a highly charged sequence found to bind to aragonite. Although it was not found to be statistically significant, its alternating charge sequence is interesting theoretically as such sequences would supposedly bind very well to an alternating ionic surface. These three peptides sequences, when forced in to these helical constructs, all show very different characteristics (Figure 3.1).

A20 exhibits the expected dual domain distribution of residues. The charged residues (KER) are clustered together within 100 - 120° for each helical model, respectively. Opposite these residues are the amide groups, which may hydrogen bond with water. In the α -helix model, between these two areas are the well-conserved proline and tryptophan.

The A21 sequence, when forced into a helical configuration, exhibits much more spacing between the charged residues. The basic residues (KHK) are nearly evenly spaced around the helix in both models. Also, the two proline residues are found between these three basic amino acids, giving more credence to the theory that prolines may bind to calcium through their backbone carbonyl

oxygens. It is interesting to note that in comparison with the A20 helical wheel, there are several conserved amino acids and equivalent residues: LP.W..[RK]..[LG].A (50% homology).

The A1 sequence is highly charged with all of these charged residues being fairly evenly distributed around the helix. The two glutamic acid residues are found nearly opposite each other with two closely clustered groups of basic amino acids (HKR and RK) between the two acidic groups. A sequence such as this may not preferentially bind with specific amino acids, but may be more flexible to bind at nearly any rotation of the helical axis parallel to the ionic mineral surface.

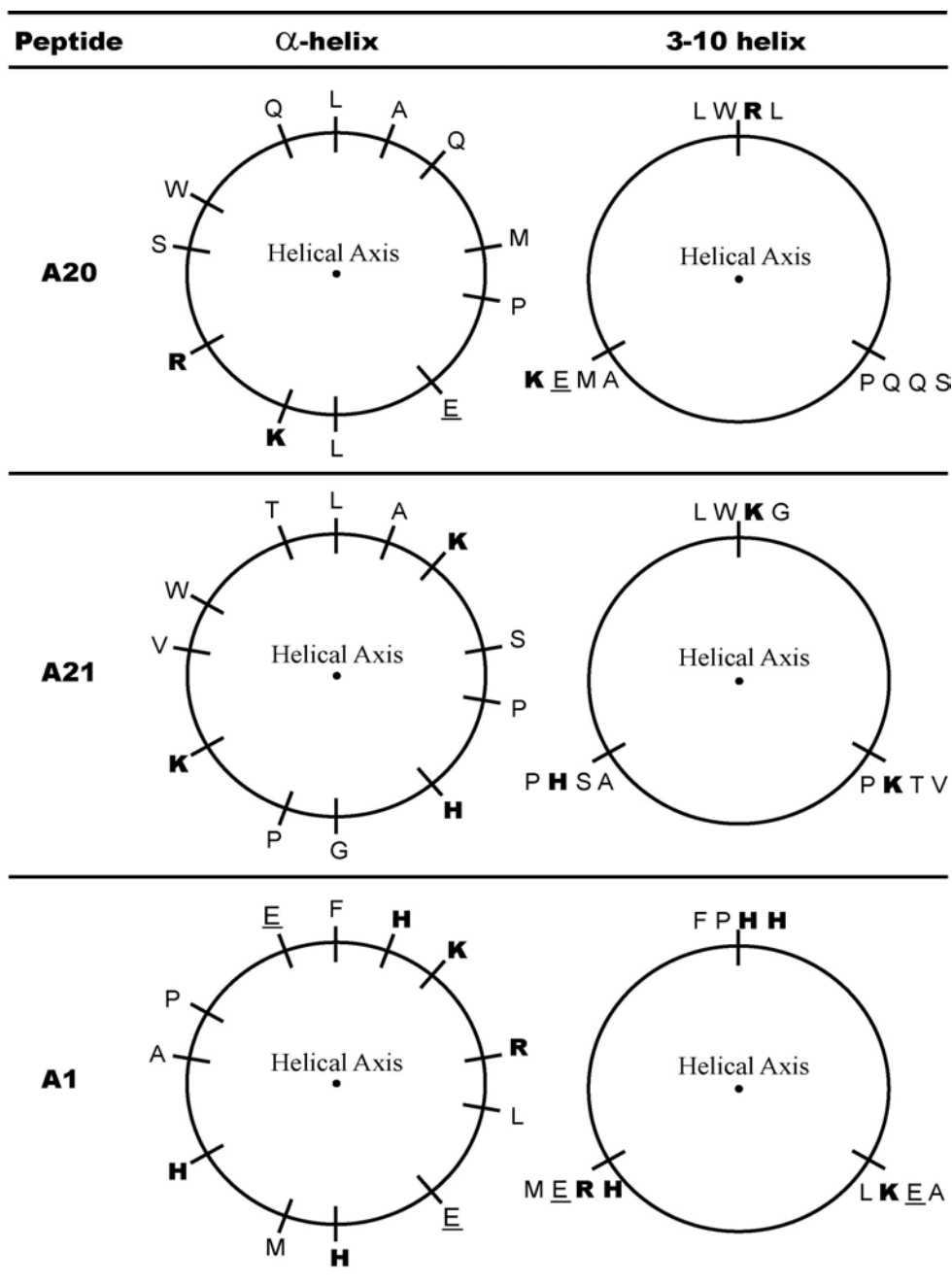


Figure 3.1: Helical wheel representations of three significant and interesting peptides selected against geological aragonite. The sequences are presented as α -helices and 3-10 helices. Residue key: bold = basic, underlined = acidic.

3.2: NACREOUS ARAGONITE

Nacreous aragonite was screened solely using the single substrate screening method with all three phage libraries. The phage were eluted from the bleached (001) surface of black-lipped oyster nacre. As can be seen from Table 3.19, the significance values are on the same order of magnitude as the (001) sequences from geological aragonite. However, they don't appear to show such interesting patterns. While there are several sequences with multiple prolines, there are only 2 (~6%) with adjacent proline residues. Most other multiple-proline sequences were spaced apart and did not lead to patterns with more than one proline. The general motif for nacreous aragonite is [STY][LAIVG] combined with P or [RKH]. Since either combination makes up 57-68% of all amino acids in the respective libraries, these patterns may not be very useful in describing binding to the 001 surface of nacre.

Significance	#	Pattern
13	2	[LAIVG]QLS[RKH][STY]...[LAIVG]L
13	11	[STY][LAIVG]P
12	10	[STY][LAIVG][RKH]
12	9	[STY][RKH]....[LAIVG]
12	9	[STY].[STY].[RKH]
12	13	[STY][LAIVG].[LAIVG]
12	8	S[LAIVG]..[LAIVG]
12	2	[STY][RKH]...V.[LAIVG].[DE].R
12	2	[STY][RKH][NQ]..[LAIVG].L.D[LAIVG]
11	2	S.....P.[STY]S[LAIVG]F

Table 3.19: All patterns from the nacreous aragonite sequences. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.

3.2.1: Nacreous Aragonite Analysis by Library

As is typical for bacteria-eluted sequences, the percentages for each of the amino acids and equivalence groups are very similar to the library averages. It is also worth noting that there are no consistent trends in over- or under-represented amino acids. The only major diversion from the respective library averages is the higher frequency of basic groups and serines at the expense of acidic groups and threonines within the PhD7C screening.

AA s	PhD12	Lib. Avg.	Δ	PhD7	Lib. Avg.	Δ	PhD7C	Lib. Avg.	Δ
A	7.8	6.0	1.8	5.7	6.7	-1.0	4.0	6.5	-2.5
C	0.2	0.5	-0.3	0.0	0.6	-0.6	0.0	0.0	0.0
D	2.7	2.8	-0.1	4.8	3.1	1.7	0.8	4.1	-3.3
E	2.0	3.1	-1.1	2.9	2.2	0.7	1.6	3.1	-1.5
F	3.9	3.3	0.6	3.8	2.2	1.6	1.6	2.1	-0.5
G	2.7	2.6	0.1	5.7	4.9	0.8	0.8	2.2	-1.4
H	7.6	6.3	1.3	4.8	3.7	1.1	7.9	6.9	1.0
I	4.2	3.4	0.8	2.9	2.9	0.0	4.0	2.1	1.9
K	2.9	2.8	0.1	1.9	2.4	-0.5	4.8	3.8	1.0
L	9.1	9.3	-0.2	10.5	9.4	1.1	7.9	9.6	-1.7
M	4.4	2.6	1.8	1.9	2.7	-0.8	1.6	3.3	-1.7
N	4.2	4.6	-0.4	3.8	4.7	-0.9	6.3	6.4	-0.1
P	12.0	12.2	-0.2	10.5	13.1	-2.6	11.1	10.7	0.4
Q	4.2	5.1	-0.9	3.8	5.1	-1.3	8.7	7.1	1.6
R	4.7	4.7	0.0	2.9	3.7	-0.8	5.6	4.3	1.3
S	10.3	10.0	0.3	13.3	11.8	1.5	15.9	8.6	7.3
T	9.1	11.1	-2.0	9.5	10.8	-1.3	9.5	13.1	-3.6
V	3.2	3.9	-0.7	5.7	4.2	1.5	3.2	1.9	1.3
W	2.0	2.2	-0.2	1.9	1.8	0.1	2.4	1.9	0.5
Y	2.9	3.6	-0.7	3.8	3.9	-0.1	2.4	2.4	0.0
X	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LAIVG	27.0	25.2	1.8	30.5	28.1	2.4	19.8	22.3	-2.5
RKH	15.2	13.8	1.4	9.5	9.8	-0.3	18.3	15.0	3.3
WF	5.9	5.5	0.4	5.7	4.0	1.7	4.0	4.0	0.0
STY	22.3	24.7	-2.4	26.7	26.5	0.2	27.8	24.1	3.7
DE	4.7	5.9	-1.2	7.6	5.3	2.3	2.4	7.2	-4.8
NQ	8.3	9.7	-1.4	7.6	9.8	-2.2	15.1	13.5	1.6

Table 3.20: Summary of nacreous aragonite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

The representative sequences for each of the libraries are listed in Table 3.21, along with the percentages of each equivalence group found at each

position. The percentages are not very descriptive as the most common amino acids in the libraries are found in the screenings in each position. However, in each library, the equivalence group with the highest percent position is [STY]. Therefore, it may be an important binding residue. In addition, the PhD12 library exhibits central [RKH] residues, a motif common to the geological aragonite screenings.

	PhD12	%	PhD7	%	PhD7C	%
1	STY	47.1	STY	33.3	C	100.0
2	LAIVG	26.5	LAIVG	40.0	STY	44.4
3	LAIVG	26.5	LAIVG/STY	26.7	LAIVG	33.3
4	LAIVG	23.5	LAIVG	40.0	STY	33.3
5	STY	35.3	LAIVG	40.0	STY	27.8
6	RKH	23.5	LAIVG	26.7	LAIVG	27.8
7	LAIVG	32.4	STY	46.7	RKH	33.3
8	RKH	32.4			STY	27.8
9	LAIVG	26.5			C	100.0
10	LAIVG	35.3				
11	LAIVG	44.1				
12	LAIVG	35.3				

Table 3.21: Highest percentages for each position for the sequences screened against nacreous aragonite for each library. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

Table 3.22 gives the best patterns for each library screened against nacreous aragonite. Most are long patterns that only occur twice. However the two more frequent patterns do not resemble the highly significant patterns found for geological aragonite, although they do contain the seemingly important [RKH]

and P residues. The groups that appear most often in the PhD12 library patterns are [STY], [LAIVG] with [RKH], P and occasionally [NQ]. The patterns for the PhD7 library screened against nacreous aragonite are also mainly made up of patterns that only occur twice. The main equivalence groups that appear most often in these patterns are [LAIVG] and [STY]. The patterns for the PhD7C sequences of nacreous aragonite are again infrequent, but short patterns. The main equivalence groups that show up in these patterns are [STY] and [NQ].

Significance	#	PhD12 Nacre Patterns
13	2	[LAIVG]QLS[RKH][STY]...[LAIVG]L
12	9	[STY][RKH]....[LAIVG]
12	2	[STY][RKH]...V.[LAIVG].[DE].R
11	2	S.....P.[STY]S[LAIVG]F
11	10	[STY][LAIVG]P
Significance	#	PhD7 Nacre Patterns
8	2	N...LF
6	2	[STY][LAIVG].[LAIVG][STY]S
6	2	I[LAIVG][LAIVG].S
6	2	A.V[RKH]
5	4	[LAIVG].[LAIVG].[STY]
Significance	#	PhD7C Nacre Patterns
8	3	SSQ
7	2	[STY].T.H.[NQ]
7	2	I..RT
7	2	[LAIVG][LAIVG]Q.R
6	4	[STY][LAIVG].[NQ]

Table 3.22: Patterns from the nacreous aragonite sequences. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.

3.2.2: Similarities to Geological Aragonite 001 Sequences

Table 3.23 lists the patterns generated using the combined geological aragonite (001) face sequences and the nacreous aragonite (001) sequences. These patterns are the highest significance patterns that are unique to this combined population or that are found in both constituent populations. The contribution of the nacre sequences is greater than the contribution of the geological sequences in these patterns. The main motif of these patterns is [STY] and [LAIVG] occasionally combined with [RKH] or P.

Significance	#	Pattern	Nacre 001	
15	17	[STY][STY][LAIVG]	12	5
15	8	[RKH]...[LAIVG].L	6	2
14	11	[STY]S[LAIVG]	8	3
14	15	[STY][LAIVG].[LAIVG]	13	2
14	11	S[STY][LAIVG]	8	3
14	2	YN.[STY][NQ]...R	1	1
14	13	[STY][LAIVG]..[LAIVG]	10	3
14	10	[STY][RKH]....[LAIVG]	9	1
14	12	[STY][LAIVG]P	11	1
14	10	[RKH]...[LAIVG].[LAIVG]	7	3

Table 3.23: Patterns from the combined geological aragonite (001) surface and nacreous aragonite sequences. The significance is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found.

The patterns generated for the (001) surface of the two different aragonite samples (geological and nacreous) are quite different when taken separately. The geological aragonite samples showed a tendency toward multiple proline sequences, with over 70% of all samples showing a PP motif. The patterns for

nacreous aragonite do not show this multiple proline proliferation in the selected sequences. The patterns for nacreous aragonite are on approximately the same order as the (001) geological aragonite patterns. However, the significance of these nacreous patterns improved with the addition of the (001) sequences.

There are several reasons these results may have presented themselves. First, nacre was only screened using bacterial elution. It has been shown that bacterial elution usually gives quite different sequences than standard acid or base elutions for these substrates. Therefore, it would be interesting to see what kinds of sequences would be eluted from nacreous aragonite using acid. Second, the nacreous aragonite has proteins occluded within it [8]. These proteins were removed using strong bleach, but there is always the possibility that some of the proteins remained, giving a very different (organic) binding surface.

3.3: GEOLOGICAL CALCITE

Geological calcite was screened several times at two different pHs and eluted using both acid and base. All surfaces of geological calcite are 104 faces, and therefore, face specific elution was not necessary. The calcite screenings yielded sequences that mainly reflect the diversity of the library, as the percentages of each amino acid fairly closely mirrors that of the library (Table 3.24). In the overall calcite sequence population, serine is the only amino acid found at a significantly higher frequency than the library. With the lack of both basic and acidic groups in the overall calcite population, it appears that binding to calcite is not as electrostatic as binding to aragonite. The increased serine content mirrors a similar selection by Li et al. [9]. Therefore, peptide binding to calcite

may be either hydrogen bonding in nature, or the tightest binding sequences, composition unknown, may be lost due to the inability of acid to remove all binders (Figure 2.5 A).

3.3.1: Effect of Screening pH on Calcite Sequences

There are some slight differences in the percentages of amino acids found in the different pH screenings of geological calcite. Table 3.24 shows the average amino acid percents for all calcite sequences together as well as the constituent groups of calcite sequences screened at the two different pHs, 7.5 and 8.5. The pH 7.5 sequences are low in leucine as well as acidic amino acids. In contrast, the pH 8.5 sequences have an average frequency of leucines and acidic residues. Although the pH 7.5 sequences are deficient in basic groups, the pH 8.5 sequences are even more deficient. The pH 8.5 sequences also have fewer glutamines than the pH 7.5 sequences, which have slightly more than the library average. Both pH groups contribute to the overall increase in serine and proline in the overall calcite sequences, although the serine increase in the pH 8.5 population (+4.7%) is twice that of the pH 7.5 population (+2.3%). Interestingly these trends are exactly the opposite of the trends for aragonite, which showed more [RKH] and fewer [STY] at pH 8.5.

AA	Lib. Avg.	All Calcite	Δ	pH 7.5	Δ	pH 8.5	Δ
A	6.0	7.5	1.5	8.3	2.3	7.1	1.1
C	0.5	0.2	-0.3	0.0	-0.5	0.3	-0.2
D	2.8	2.9	0.1	2.0	-0.8	3.2	0.4
E	3.1	1.8	-1.3	0.4	-2.7	2.3	-0.8
F	3.3	2.9	-0.4	3.2	-0.1	2.7	-0.6
G	2.6	3.4	0.8	2.8	0.2	3.6	1.0
H	6.3	5.0	-1.3	6.0	-0.3	4.7	-1.6
I	3.4	2.4	-1.0	1.6	-1.8	2.7	-0.7
K	2.8	2.4	-0.4	1.2	-1.6	2.9	0.1
L	9.3	7.7	-1.6	5.2	-4.1	8.6	-0.7
M	2.6	4.1	1.5	5.2	2.6	3.6	1.0
N	4.6	4.7	0.1	4.0	-0.6	5.0	0.4
P	12.2	14.1	1.9	13.9	1.7	14.2	2.0
Q	5.1	4.7	-0.4	7.5	2.4	3.6	-1.5
R	4.7	4.2	-0.5	5.2	0.5	3.8	-0.9
S	10.0	14.0	4.0	12.3	2.3	14.7	4.7
T	11.1	8.6	-2.5	9.9	-1.2	8.0	-3.1
V	3.9	4.6	0.7	6.7	2.8	3.8	-0.1
W	2.2	1.5	-0.7	1.6	-0.6	1.5	-0.7
Y	3.6	3.2	-0.4	2.8	-0.8	3.3	-0.3
X	0.0	0.2	0.2	0.4	0.4	0.2	0.2
LAIVG	25.2	25.5	0.3	24.6	-0.6	25.9	0.7
RKH	13.8	11.6	-2.2	12.3	-1.5	11.4	-2.4
WF	5.5	4.4	-1.1	4.8	-0.7	4.2	-1.3
STY	24.7	25.8	1.1	25.0	0.3	26.1	1.4
DE	5.9	4.6	-1.3	2.4	-3.5	5.5	-0.4
NQ	9.7	9.4	-0.3	11.5	1.8	8.6	-1.1

Table 3.24: Summary of geological calcite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

The representative sequences for the calcite sequences and the constituent pH 7.5 and pH 8.5 populations are shown in table 3.25. Only the pH 7.5

sequences have positions where the highest percentage group is not [STY] or [LAIVG]. There are two proline positions and one LAIVG/STY/NQ. It is difficult to describe binding on calcite, except for the fact that electrostatic binding does not seem to be a dominating factor at either pH.

	All Calcite	%	pH 7.5	%	pH 8.5	%
1	STY	32.9	LAIVG	42.9	STY	36.4
2	STY/ <i>P</i>	23.7	<i>P</i>	28.6	STY	25.5
3	STY	26.3	STY	33.3	LAIVG	27.3
4	STY	25.0	LAIVG/STY	28.6	STY	23.6
5	STY	26.3	LAIVG	33.3	STY	30.9
6	LAIVG	27.6	<i>P</i>	23.8	LAIVG	30.9
7	STY	26.3	STY	38.1	LAIVG	25.5
8	LAIVG	31.6	LAIVG	33.3	LAIVG/STY	30.9
9	LAIVG	30.3	LAIVG/RKH/NQ	19.0	LAIVG	34.5
10	LAIVG	32.9	STY	38.1	LAIVG	36.4
11	LAIVG	38.2	LAIVG	33.3	LAIVG	40.0
12	STY	34.2	STY	33.3	STY	34.5

Table 3.25: Highest percentages for each PhD12 position for the sequences screened against geological calcite at pH 7.5 and pH 8.5. The average of all of the geological calcite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

3.3.2: Effect of Elution Method on Calcite Sequences

In the aragonite screenings, the sequences eluted with base followed the same trend as the acid-eluted sequences; the trends were just twice as intense. Here, the only consistent trend is the fact that serines are eluted more frequently in both elution methods for the calcite screenings (Table 3.26). However, they are significantly more frequent in the base elutions. Other trends include amino acids

that are encountered at roughly the rate of the library in one elution method and slightly higher or lower in the other elution method. With the exception of serine, it is difficult to determine binding trends based on elution method.

AA	Lib. Avg.	All Calcite	Δ	Acid	Δ	Base	Δ
A	6.0	7.5	1.5	7.2	1.2	8.0	2.0
C	0.5	0.2	-0.3	0.3	-0.2	0.0	-0.5
D	2.8	2.9	0.1	2.5	-0.3	3.6	0.8
E	3.1	1.8	-1.3	1.6	-1.5	2.2	-0.9
F	3.3	2.9	-0.4	3.3	0.0	1.8	-1.5
G	2.6	3.4	0.8	3.9	1.3	2.2	-0.4
H	6.3	5.0	-1.3	5.0	-1.3	5.1	-1.2
I	3.4	2.4	-1.0	2.2	-1.2	2.9	-0.5
K	2.8	2.4	-0.4	2.2	-0.6	2.9	0.1
L	9.3	7.7	-1.6	7.7	-1.6	7.6	-1.7
M	2.6	4.1	1.5	4.9	2.3	2.2	-0.4
N	4.6	4.7	0.1	4.1	-0.5	6.2	1.6
P	12.2	14.1	1.9	13.8	1.6	14.9	2.7
Q	5.1	4.7	-0.4	5.5	0.4	2.9	-2.2
R	4.7	4.2	-0.5	4.4	-0.3	3.6	-1.1
S	10.0	14.0	4.0	13.4	3.4	15.6	5.6
T	11.1	8.6	-2.5	8.5	-2.6	8.7	-2.4
V	3.9	4.6	0.7	5.2	1.3	3.3	-0.6
W	2.2	1.5	-0.7	1.6	-0.6	1.4	-0.8
Y	3.6	3.2	-0.4	2.4	-1.2	5.1	1.5
X	0.0	0.2	0.2	0.3	0.3	0.0	0.0
LAIVG	25.2	25.5	0.3	26.3	1.1	23.9	-1.3
RKH	13.8	11.6	-2.2	11.6	-2.2	11.6	-2.2
WF	5.5	4.4	-1.1	4.9	-0.6	3.3	-2.2
STY	24.7	25.8	1.1	24.2	-0.5	29.3	4.6
DE	5.9	4.6	-1.3	4.1	-1.8	5.8	-0.1
NQ	9.7	9.4	-0.3	9.6	-0.1	9.1	-0.6

Table 3.26: Summary of the different elution methods of the PhD12 geological calcite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

Again, the percent position analysis summary (Table 3.27) reveals that the most often encountered amino acids are [STY] and [LAIVG] for most positions.

However, some of the [STY] positions are relatively well conserved in the base elution population (STY in position 1 and S in position 6). In addition, the base-eluted sequences have two positions in which prolines are the most common residue.

	All Calcite	%	Acid Elution	%	Base Elution	%
1	STY	32.9	LAIVG	32.1	STY	56.5
2	STY/P	23.7	STY/ LAIVG	18.9	STY	34.8
3	STY	26.3	LAIVG	28.3	STY	39.1
4	STY	25.0	STY	28.3	<i>P</i>	26.1
5	STY	26.3	STY	26.4	LAIVG/STY	26.1
6	LAIVG	27.6	LAIVG	28.3	S	30.4
7	STY	26.3	STY	30.2	LAIVG	26.1
8	LAIVG	31.6	LAIVG	32.1	LAIVG/STY	30.4
9	LAIVG	30.3	LAIVG	24.5	LAIVG	43.5
10	LAIVG	32.9	LAIVG	34.0	LAIVG/P	30.4
11	LAIVG	38.2	LAIVG	41.5	LAIVG	30.4
12	STY	34.2	STY	32.1	STY	39.1

Table 3.27: Highest percentages for each PhD12 position for the sequences screened against geological calcite and eluted with acid or base. The average of all of the geological calcite PhD12 sequences is included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

The patterns for the calcite sequences are dominated by [STY], [LAIVG], and sometimes proline. Even the two long patterns are highly populated with hydroxyl groups. Overall, these are relatively high significance values, although, they are not nearly as good as aragonite values. It is interesting that different polymorphs of calcium carbonate, when screened in the same manner, yield patterns and significance values of very different orders of magnitude.

Significance	#	Pattern
18	2	F.MQ..[STY]T.SQP
17	2	F.MQ...T.SQP
15	12	[STY][STY]...P
15	18	[LAIVG][STY].[LAIVG]
15	2	V..T.A.S.STS
14	13	[LAIVG][LAIVG]P
14	11	[STY].S.[STY]
14	16	[STY]..[LAIVG][LAIVG]
14	16	[STY].[LAIVG].[LAIVG]
14	11	S..[LAIVG][LAIVG]

Table 3.28: Patterns from the geological calcite sequences. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.

3.4: HOMOLOGY TO NATURAL BIOMINERALIZATION PROTEINS

Other protein binding experiments with calcite have shown that, in nature, proteins that bind to calcite and hinder crystal growth often have a flexible, open conformation. In contrast to the selected sequences presented here, these calcite-binding sequences typically are quite electrostatic in nature. In simulation studies, it has been shown that as the protein approaches the calcite surface, it unfolds, forms hydrogen bonds with the surface followed by electrostatic interactions [10]. In addition, it has been shown that proteins that bind to calcite and inhibit mineral growth bind so strongly that they cannot be easily eluted with high or low ionic strength solutions [11]. Therefore, it is quite possible that the best binding calcite sequences are those that are not eluted with high or low pH solutions. Then, the best binders may remain bound to the substrate, while the

medium strength binders are amplified and sequenced. Therefore, it is likely that a new elution method may need to be developed for calcite screenings.

Name and origin	Sequence
Lithostathine protein fragment from human pancreas inhibits calcite growth [10]	E E A Q T E L P Q A R
D4 domain of Lustrin A from <i>H. rufescens</i> frustrates calcite overgrowth and cannot be easily displaced from surface [11]	G K G A S <i>Y</i> <i>D</i> <i>T</i> <i>D</i> <i>A</i> <i>D</i> <i>S</i> <i>G</i> <i>S</i> <i>D</i> <i>N</i> <i>R</i> <i>S</i> <i>P</i> <i>G</i> <i>Y</i> <i>L</i> <i>P</i> <i>Q</i>
AP24-1 from <i>H. rufescens</i> frustrates calcite overgrowth and has random coil configuration [12]	A D D E D A S S G L C <i>N</i> <i>Q</i> <i>Y</i> <i>N</i> <i>Q</i> <i>N</i> <i>V</i> <i>T</i> <i>T</i> <i>R</i> <i>P</i> <i>N</i> <i>N</i> <i>K</i> <i>P</i> <i>K</i> <i>M</i> <i>F</i>
AP7-1 from <i>H. rufescens</i> frustrates calcite overgrowth and is unfolded [12]	D <i>D</i> <i>N</i> <i>G</i> <i>N</i> <i>Y</i> <i>G</i> <i>N</i> <i>G</i> <i>M</i> <i>A</i> <i>S</i> <i>V</i> <i>R</i> <i>T</i> <i>Q</i> <i>G</i> <i>N</i> <i>T</i> <i>Y</i> <i>D</i> <i>D</i> <i>L</i> <i>A</i> <i>S</i> <i>L</i> <i>I</i> <i>S</i> <i>Y</i> <i>L</i>

Table 3.29: Protein fragments that have been shown to bind to calcite and hinder mineral overgrowth. Charged residues are in bold, hydroxyl residues are in gray italics and amides are underlined.

In contrast to calcite, the sequences selected against aragonite provide much more information and tentative homology to known aragonite-associated proteins. The only exceptions to this statement are the sequence patterns found for nacreous aragonite. This is probably due to the use of the less efficient single substrate screening and the possibility that surface proteins were not completely removed during the bleaching treatment. Known nacre proteins were analyzed for

the same patterns found to be highly significant in the aragonite screening process. Again, the patterns were calculated using all ordered combinations of the pattern over a six amino acid window. The results are presented in Table 3.30.

GI	Name	Organism	[RKH][RKH][RKH]	[WF][RKH][RKH]	P[RKH][RKH]	[LAIVG]P[RKH]	[LAIVG]P[WF]	P[WF][RKH]	PP
12655873	AP24 precursor [12]	<i>H. rufescens</i>	5	2	2	4	0	1	3
12655875	AP7 precursor [12]	<i>H. rufescens</i>	5	3	1	0	0	0	0
18202416	Perlustrin [13, 14]	<i>H. laevigata</i>	4	4	0	4	1	0	2
25090912	Perlucin [13, 15]	<i>H. laevigata</i>	1	4	0	0	1	1	1
13309811	Mucoperlin [16]	<i>P. nobilis</i>	1	0	0	26	1	0	28
2204081	MSI60 nacre [17]	<i>P. fucata</i>	21	0	1	2	3	0	0
7511706	Lustrin A [18]	<i>H. rufescens</i>	6	2	7	28	8	3	188
20372975	Nacrein [19]	<i>T. marmoratus</i>	13	12	6	15	12	1	7
1480031	Nacrein [20]	<i>P. fucata</i>	43	9	11	12	3	1	3
11362266	N66 [21]	<i>P. maxima</i>	60	9	13	15	3	1	5
4519446	Pearlin [22, 23]	<i>P. fucata</i>	2	2	1	1	3	3	0
11362265	N14 [21]	<i>P. maxima</i>	0	1	0	0	1	1	0
5821239	N16#1 [23]	<i>P. fucata</i>	6	0	3	1	1	0	0
5821243	N16#2 [24]	<i>P. fucata</i>	7	1	3	1	1	1	0
5821245	N16#3 [25]	<i>P. fucata</i>	5	0	3	1	1	0	0
5821247	N16#4 [26]	<i>P. fucata</i>	5	0	3	1	1	0	0
5821249	N16#5 [27]	<i>P. fucata</i>	2	2	1	1	3	3	0
5821253	N16#7 [28]	<i>P. fucata</i>	2	2	1	1	3	3	0

Table 3.30: The number of patterns common to the aragonite screening results found for each known nacre protein sequence. The highlighted proteins have numerous patterns in common with the screening population.

Mucoperlin is found in the nacre of the bivalve mollusc, *Pinna nobilis* [16] and exhibits two patterns found in the aragonite screening population:

[LAIVG]P[RKH] and PP. Thirteen tandem repeats (LSSGMPLSKPMPDSTQVVEIPIQSASLLQPK) and a 222 residue C-terminal domain make up this potentially highly glycosylated protein. Mucoperlin is thought to be located in the nacre matrix, surrounding the aragonite tablets, and, in combination with other acetic acid-soluble nacre extracts, has been shown to inhibit calcium carbonate growth *in vitro*.

Nacrein [19, 20] and the homologous N66 protein [21] both share several of the patterns found in the aragonite screening population. These proteins, in combination with other nacreous protein extracts, can indeed induce platy aragonite crystals from an aragonitic crystallization solution, which includes magnesium ions [21].

Lustrin A, from *Haliotis rufescens* presents a very interesting comparison due to the fact that it is a nacreous protein that contains eight proline rich domains [18]. The proline rich domains in Lustrin A (PP, PXP, PXXP) have been postulated to be spacers for the cysteine rich domains that separate them. In addition to having many polyproline patterns, Lustrin A also contains several [LAIVG]P[RKH] patterns, which are found quite often in the aragonite binding peptide population.

3.5: CONCLUSIONS

Phage display screening for biological and geological inorganic materials is an effective method for discovering peptides that bind to these types of materials. However, a new method developed from the two screening methods tested here would be beneficial for two main reasons: 1) the best binders are

likely still bound to the surface even after acid or base elution and 2) single substrate screening compensates for this fact, yet it has not proven to be efficient in discovering binding motif consensus.

Calcite sequences produce short patterns that are quite common to any peptide population. Aragonite sequences show extraordinary consensus overall as well as in the face specific subgroup populations. The (110) aragonite surface consensus motif is [LAIVG]PW[RKH][RKH][RKH][STY][LAIVG]. The (001) aragonite surface motifs consist of various polyproline sequences. Both of these motifs show putative homology to known nacre proteins, especially Lustrin A, Mucoperlin, and Nacrein. Nacre screenings did not generate any consensus motifs, most likely due to the less effective method of screening used.

REFERENCES:

1. Smith, G.P., *Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface*. Science, 1985. **228**(4705): p. 1315-1317.
2. Zarrinpar, A., R. Bhattacharyya, and W. Lim, *The Structure and Function of Proline Recognition Domains*. Science's STKE, 2003. **re8**.
3. Szabo, Z., *Polyproline helices*. 1997.
4. Weiner, S. and W. Traub, *Bone-Structure - from Angstroms to Microns*. FASEB Journal, 1992. **6**(3): p. 879-885.
5. Zaremba, C.M., et al., *Critical Transitions in the Biofabrication of Abalone Shells and Flat Pearls*. Chemistry of Materials, 1996. **8**(3): p. 679-90.
6. Schaeffer, T.E., et al., *Does Abalone Nacre Form by Heteroepitaxial Nucleation or by Growth through Mineral Bridges?* Chemistry of Materials, 1997. **9**(8): p. 1731-1740.

7. Richardson, J.S. and D.C. Richardson, *Amino-Acid Preferences for Specific Locations at the Ends of Alpha-Helices*. Science, 1988. **240**(4859): p. 1648-1652.
8. Belcher, A.M., et al., *Control of crystal phase switching and orientation by soluble mollusk-shell proteins*. Nature (London), 1996. **381**(6577): p. 56-58.
9. Li, C.M., G.D. Botsaris, and D.L. Kaplan, *Selective in vitro effect of peptides on calcium carbonate crystallization*. Crystal Growth & Design, 2002. **2**(5): p. 387-393.
10. Gerbaud, V., et al., *Mechanism of calcite crystal growth inhibition by the N-terminal undecapeptide of lithostathine*. Journal of Biological Chemistry, 2000. **275**(2): p. 1057-1064.
11. Wustman, B.A., et al., *Characterization of a Ca(II)-, mineral-interactive polyelectrolyte sequence from the adhesive elastomeric biomineralization protein lustrin A*. Langmuir, 2003. **19**(22): p. 9373-9381.
12. Michenfelder, M., et al., *Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains*. Biopolymers, 2003. **70**(4): p. 522-533.
13. Weiss, I.M., et al., *Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc Haliotis laevigata*. Biochemical and Biophysical Research Communications, 2000. **267**(1): p. 17-21.
14. Weiss, I.M., et al., *Perlustrin, a Haliotis laevigata (abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N-terminal module of vertebrates*. Biochemical and Biophysical Research Communications, 2001. **285**(2): p. 244-249.
15. Blank, S., et al., *The nacre protein perlucin nucleates growth of calcium carbonate crystals*. Journal of Microscopy-Oxford, 2003. **212**: p. 280-291.
16. Marin, F., et al., *Mucins and molluscan calcification - Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel Pinna nobilis (Bivalvia, Pteriomorpha)*. Journal of Biological Chemistry, 2000. **275**(27): p. 20667-20675.
17. Sudo, S., et al., *Structures of mollusc shell framework proteins*. Nature, 1997. **387**(6633): p. 563-564.

18. Shen, X.Y., et al., *Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of Haliotis rufescens*. Journal of Biological Chemistry, 1997. **272**(51): p. 32472-32481.
19. Yano, M., Miyashita, T. and Miyamoto, H., *Shell matrix Nacrein, a novel family of carbonic anhydrase, is conserved in bivalve and gastropod*. 2001, Entrez.
20. Miyamoto, H., et al., *A carbonic anhydrase from the nacreous layer in oyster pearls*. Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(18): p. 9657-9660.
21. Kono, M., N. Hayashi, and T. Samata, *Molecular mechanism of the nacreous layer formation in Pinctada maxima*. Biochemical and Biophysical Research Communications, 2000. **269**(1): p. 213-218.
22. Miyashita, T., et al., *Complementary DNA cloning and characterization of pearlilin, a new class of matrix protein in the nacreous layer of oyster pearls*. Marine Biotechnology, 2000. **2**(5): p. 409-418.
23. Samata, T., et al., *A new matrix protein family related to the nacreous layer formation of Pinctada fucata*. Febs Letters, 1999. **462**(1-2): p. 225-229.
24. Hayashi, N.a.S., T., *N14#2 - 14 kD matrix protein family in nacreous layer of Japanese pearl oyster, Pinctada fucata*. 1999, Entrez.
25. Hayashi, N.a.S., T., *N14#3 - 14 kD matrix protein family in nacreous layer of Japanese pearl oyster, Pinctada fucata*. 1999, Entrez.
26. Hayashi, N.a.S., T., *N14#4 - 14 kD matrix protein family in nacreous layer of Japanese pearl oyster, Pinctada fucata*. 1999, Entrez.
27. Hayashi, N.a.S., T., *N14#5 - 14 kD matrix protein family in nacreous layer of Japanese pearl oyster, Pinctada fucata*. 1999, Entrez.
28. Hayashi, N.a.S., T., *N14#7 - 14 kD matrix protein family in nacreous layer of Japanese pearl oyster, Pinctada fucata*. 1999, Entrez.

Chapter 4: Phage Display Results: Similar Materials

Presented here are the results of the screenings for oxide materials that are chemically or stoichiometrically similar to calcium carbonate. Table 4.1 shows the different screening parameters for these materials. Hydroxyapatite, $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$, is a material of biomedical interest as a bone replacement material. The hydroxyapatite samples used were sintered polycrystalline pellets, so there is no preferential crystal face. The two perovskite electronic materials, also screened in polycrystalline powder form, were barium titanate (BaTiO_3) and lead titanate (PbTiO_3). Finally, the optical material lithium niobate (LiNbO_3) was screened here in both polycrystalline powder and single crystal Z-poled form.

Screening			HA	PbTiO ₃	BaTiO ₃		LiNbO ₃	
pH	Elution	Library	Powder	Powder	Powder	Powder	+Z	-Z
7.5	acid	PhD12					•	•
		PhD7						
		PhD7C		•	•	•		
7.5	bacterial	PhD12					•	•
		PhD7						
		PhD7C						
8.5	bacterial	PhD12	•					
		PhD7	•					
		PhD7C	•					

Table 4.1: Summary of the screening conditions for the biomedical (hydroxyapatite – HA), electronic (PbTiO_3 and BaTiO_3) and optical (LiNbO_3) materials. The dots indicate selection conditions under which these materials were screened. LiNbO_3 single crystal is poled in the Z direction, giving two different screening surfaces.

4.1 HYDROXYAPATITE

All hydroxylapatite (HA) screenings were done using the single substrate screening/bacterial elution method and all three libraries. The percent position analysis summary (Table 4.2) shows that [STY] and [LAIVG] are the most commonly encountered equivalence groups. This is probably not a very descriptive account of HA binding, since with the exception of the over-representation of [LAIVG] residues in the Phd7 population, [STY] and [LAIVG] are not found at high rates in these screenings (Table 4.3). Therefore, these percent position trends are likely a reflection of the library percents rather than evidence of binding residues. There are two factors here that are likely to prevent pattern elucidation. First, the single substrate screening method tends to select sequences that are more like the library than acid-eluted sequences. Second, the presence of multiple crystal faces in the hydroxylapatite pellet make consensus difficult, due to the multiple mechanisms of binding possible.

	PhD12	%	PhD7	%	PhD7C	%
1	STY	34.0	LAIVG	39.5	C	100.0
2	STY	30.2	STY	31.6	LAIVG/STY	18.8
3	STY	32.1	LAIVG	39.5	STY	31.3
4	LAIVG	22.6	LAIVG	26.3	LAIVG	25.0
5	LAIVG	34.0	LAIVG	28.9	STY	34.4
6	LAIVG	22.6	LAIVG	42.1	STY	31.3
7	LAIVG	22.6	LAIVG	39.5	LAIVG	28.1
8	LAIVG	22.6			LAIVG	37.5
9	LAIVG	28.3			C	100.0
10	STY	30.2				
11	LAIVG	28.3				
12	LAIVG	34.0				

Table 4.2: Highest percentages for each position in each library for the sequences screened against hydroxyapatite. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

Since HA, like calcium carbonate, is an ionic material, one would expect binding to be electrostatic in nature. However, only the PhD7C library shows any evidence of a higher frequency of charged residues in the selected sequences (Table 4.3). Yet, acidic or basic residues do not seem to be conserved enough to appear in the percent position analysis or highly significant patterns (Table 4.4) for any library screened against HA.

AAs	PhD12	Lib. Avg.	Δ	PhD7	Lib. Avg.	Δ	PhD7C	Lib. Avg.	Δ
A	5.8	6.0	-0.2	9.8	6.7	3.1	5.8	6.5	-0.7
C	0.6	0.5	0.1	0.0	0.6	-0.6	0.0	0.0	0.0
D	2.7	2.8	-0.1	1.9	3.1	-1.2	3.6	4.1	-0.5
E	1.3	3.1	-1.8	2.6	2.2	0.4	0.9	3.1	-2.2
F	3.1	3.3	-0.2	3.0	2.2	0.8	3.1	2.1	1.0
G	3.8	2.6	1.2	2.6	4.9	-2.3	4.5	2.2	2.3
H	4.4	6.3	-1.9	5.3	3.7	1.6	7.1	6.9	0.2
I	2.7	3.4	-0.7	3.0	2.9	0.1	2.7	2.1	0.6
K	3.6	2.8	0.8	2.3	2.4	-0.1	2.2	3.8	-1.6
L	8.5	9.3	-0.8	12.4	9.4	3.0	5.4	9.6	-4.2
M	3.9	2.6	1.3	2.6	2.7	-0.1	2.7	3.3	-0.6
N	5.0	4.6	0.4	4.1	4.7	-0.6	7.6	6.4	1.2
P	13.5	12.2	1.3	13.2	13.1	0.1	8.9	10.7	-1.8
Q	6.9	5.1	1.8	5.6	5.1	0.5	6.7	7.1	-0.4
R	5.2	4.7	0.5	1.9	3.7	-1.8	7.1	4.3	2.8
S	12.4	10.0	2.4	6.8	11.8	-5.0	11.2	8.6	2.6
T	7.2	11.1	-3.9	11.7	10.8	0.9	10.7	13.1	-2.4
V	3.3	3.9	-0.6	6.4	4.2	2.2	4.0	1.9	2.1
W	2.8	2.2	0.6	1.9	1.8	0.1	1.3	1.9	-0.6
Y	2.0	3.6	-1.6	2.3	3.9	-1.6	3.1	2.4	0.7
X	1.1	0.0	1.1	0.8	0.0	0.8	1.3	0.0	1.3
LAIVG	24.1	25.2	-1.1	34.2	28.1	6.1	22.3	22.3	0.0
RKH	13.2	13.8	-0.6	9.4	9.8	-0.4	16.5	15.0	1.5
WF	6.0	5.5	0.5	4.9	4.0	0.9	4.5	4.0	0.5
STY	21.7	24.7	-3.0	20.7	26.5	-5.8	25.0	24.1	0.9
DE	3.9	5.9	-2.0	4.5	5.3	-0.8	4.5	7.2	-2.7
NQ	11.9	9.7	2.2	9.8	9.8	0.0	14.3	13.5	0.8

Table 4.3: Summary of hydroxyapatite screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences ($> \pm 3.0\%$) from the expected library percentages are highlighted.

The pattern significance values are among the best for a “non-aragonite” population (Table 4.4). The patterns differ from the percent position analysis only

with the presence of several patterns that include amide groups, giving significant patterns with some combination of [STY], [LAIVG] and sometimes [NQ].

Significance	#	Pattern
19	17	[STY].[STY].[LAIVG]
19	18	[STY][LAIVG].[LAIVG]
18	18	[LAIVG].[LAIVG][LAIVG]
16	12	[NQ][LAIVG][LAIVG]
16	11	[STY].[NQ][LAIVG]
16	11	T.[STY].[LAIVG]
16	18	[LAIVG][LAIVG][LAIVG]
16	12	[NQ][STY][LAIVG]
16	15	[STY].[LAIVG][LAIVG]
16	11	[NQ][STY].[LAIVG]

Table 4.4: Patterns from the hydroxyapatite sequences. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.

The screenings of sintered HA would likely have improved with the use of acid elution, especially since acid would etch HA, expectantly removing the bound phage. In addition, the selection pressure of the single substrate method is even less effective with polycrystalline substrates.

4.2: BARIUM TITANATE, LEAD TITANATE AND LITHIUM NIOBATE POWDERS

Table 4.5 shows the percent summary for each of the amino acids and equivalence groups for the three powder screenings. All of the powder screenings were done with the PhD7C library and all were eluted using the standard acid elution method. The constrained library has fewer degrees of freedom and

therefore, helps to limit the number of variables during the screening process since the powdered substrates have many different surfaces.

Barium titanate (BaTiO_3) shows trends similar to the aragonite sequences. The basic groups are high, specifically lysine and arginine (+3.7% together), as are the proline residues (+4.4%). Overall, the hydrophobic amino acids are low in frequency as are the amide groups. Even the percent position summary (Table) looks similar to the N-terminus of the PhD12 aragonite sequences. There is proline near the N-terminus and a polar/charged region in the center. It is interesting to note the very high frequency of [STY] in the third position. This is the highest percentage of any position in any substrate screened. However, it is unlikely that the two populations are truly similar since the barium titanate peptides are constrained in a disulfide loop that would likely prevent these amino acids from interacting with the surface in the same manner that a non-constrained, PhD12 aragonite peptide would.

The patterns for barium titanate have quite low significance (Table 4.7). In fact, these are the lowest significance values found for any substrate screened. Most patterns are only found twice and do not seem to be very important. However, there does seem to be a slight trend toward patterns with P, [STY] and [RKH].

These trends are underscored by the fact that there were two sequences pulled out multiple times. This duplicate data was not included in the percent position calculation, due to the desire to negate amplification biasing. However, the two sequences, **TPLHLQM**, pulled out twice, and **PTSAHPS**, pulled out

three times, do indeed reflect the proline, hydroxyl group pairing as well as the central a hydrophobic residue and a charged group, histidine, in both of these cases. Therefore, this may be an important trend despite the low significance values.

	Lib. Avg.	BaTiO ₃	Δ	PbTiO ₃	Δ	LiNbO ₃	Δ
A	6.5	8.4	1.9	8.8	2.3	6.0	-0.5
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	4.1	4.2	0.1	1.4	-2.7	1.8	-2.3
E	3.1	1.7	-1.4	2.7	-0.4	4.2	1.1
F	2.1	1.7	-0.4	0.0	-2.1	1.8	-0.3
G	2.2	1.7	-0.5	2.0	-0.2	3.0	0.8
H	6.9	6.7	-0.2	8.2	1.3	12.5	5.6
I	2.1	0.0	-2.1	0.7	-1.4	0.6	-1.5
K	3.8	5.9	2.1	4.1	0.3	15.5	11.7
L	9.6	7.6	-2.0	11.6	2.0	6.5	-3.1
M	3.3	3.4	0.1	1.4	-1.9	3.6	0.3
N	6.4	4.2	-2.2	8.2	1.8	2.4	-4.0
P	10.7	15.1	4.4	9.5	-1.2	10.1	-0.6
Q	7.1	5.0	-2.1	5.4	-1.7	3.0	-4.1
R	4.3	5.9	1.6	4.8	0.5	2.4	-1.9
S	8.6	10.9	2.3	10.9	2.3	10.7	2.1
T	13.1	10.9	-2.2	15.0	1.9	7.1	-6.0
V	1.9	0.8	-1.1	2.7	0.8	4.8	2.9
W	1.9	0.8	-1.1	0.7	-1.2	1.2	-0.7
Y	2.4	4.2	1.8	2.0	-0.4	3.0	0.6
X	0.0	0.8	0.8	0.0	0.0	0.0	0.0
LAIVG	22.3	18.5	-3.8	25.9	3.6	20.8	-1.5
RKH	15.0	18.5	3.5	17.0	2.0	30.4	15.4
WF	4.0	2.5	-1.5	0.7	-3.3	3.0	-1.0
STY	24.1	26.1	2.0	27.9	3.8	20.8	-3.3
DE	7.2	5.9	-1.3	4.1	-3.1	6.0	-1.2
NQ	13.5	9.2	-4.3	13.6	0.1	5.4	-8.1

Table 4.5: Summary of the PhD7C sequences for the powdered BaTiO₃, PbTiO₃ and LiNbO₃ screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences (> ±3.0%) from the expected library percentages are highlighted.

Lead titanate, PbTiO₃, generates slightly different amino acid and equivalence group percentages (Table 4.5). The hydrophobic groups are higher here than any of the powdered substrates. This is mainly due to the increase in

leucines and alanines. The polar hydroxyl groups, especially serines and threonines, as well as the basic groups may be responsible for binding, since they both are present in high percentages. The percent position summary (Table 4.6) shows that there are several hydroxyl or basic positions. It also shows a predominately amide position at position two. The C-terminal residues of the lead titanate sequences are mainly hydrophobic. There are noticeably fewer acidic groups as well as aromatic tryptophans and phenylalanines.

Lead titanate patterns (Table 4.7) have the best significance of any of the powders, although they too are lower than calcite, hydroxyapatite and, of course, aragonite. The top patterns for lead titanate are all derivatives of the best significance pattern. Therefore, the consensus order seems to be [STY][NQ][STY][LAIVG] with some variable spacing.

	BaTiO ₃	%	PbTiO ₃	%	LiNbO ₃	%
1	C	100.0	C	100.0	C	100.0
2	<i>P</i>	35.3	STY	38.1	STY	33.3
3	LAIVG	29.4	NQ	28.6	RKH/LAIVG	29.2
4	STY	70.6	STY	38.1	RKH	33.3
5	<u>RKH/DE</u>	23.5	STY	33.3	RKH	25.0
6	<u>STY/L/RKH</u>	17.6	LAIVG/RKH	33.3	RKH	37.5
7	RKH	35.3	LAIVG	33.3	RKH	41.7
8	STY	29.4	LAIVG	42.9	STY	25.0
9	C	100.0	C	100.0	C	100.0

Table 4.6: Highest percentages for each PhD7C position for the sequences screened against powdered BaTiO₃, PbTiO₃ and LiNbO₃. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

The lithium niobate, LiNbO_3 , powder sequences give some very interesting total amino acid and equivalence group percentages (Table 4.5). These sequences are extremely high in basic lysines and histidines and deficient in almost all other amino acids. The most deficient amino acids are the amides, threonine and leucine. The percent position summary (Table 4.6) demonstrates that the representative sequence for the lithium niobate substrate is almost completely basic. The only non-basic positions are near the constrained cysteines. This is extremely reminiscent of the PhD7C sequences for aragonite. More similarities to aragonite will be explored in the single crystal lithium niobate screenings.

The best patterns for lithium niobate powder sequences are remarkably similar to the patterns for the same library screened against geological aragonite. In fact, the highest significance pattern for both populations is exactly the same: the [RKH] triplet. In general, the lithium niobate powder patterns are highly basic and occasionally include hydrophobic groups and/or hydroxyl groups.

Significance #	BaTiO3 Patterns
7	2 PTS...[STY]
7	2 [RKH]T[DE].P
6	3 [STY]P.H
6	2 [LAIVG]MY
6	4 [STY].[STY][RKH]
Significance #	PbTiO3 Patterns
13	2 T[NQ][STY]SA[STY]L
10	4 T[NQ][STY].[LAIVG]
9	5 [STY]..S[LAIVG]
9	4 T[NQ].S
8	6 [STY]..[STY][LAIVG]
Significance #	LiNbO3 Patterns
11	6 [RKH][RKH][RKH]
10	6 [LAIVG].[RKH][RKH]
10	2 THEP.[RKH]
10	3 S[LAIVG].[RKH]H
10	4 [LAIVG]..[RKH]K

Table 4.7: Patterns from the BaTiO₃, PbTiO₃ and LiNbO₃ powder sequences. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.

4.3: LITHIUM NIOBATE SINGLE CRYSTAL

In addition to the powdered lithium niobate substrate, lithium niobate was also screened in single crystal form. +Z and -Z refer to the direction of polarization within the crystal. These two faces were screened using both acid elution and single substrate screening. Like most other substrates screened using the single substrate screening method, the sequences found using this method are much more similar to the library than the acid-eluted sequences. These sequences

often include several serines. The acid-eluted sequences seem to give more distinct patterns.

The trends in amino acid distributions are quite different for the two different elution methods, as seen in the aragonite screenings as well. Table 4.8 shows the amino acid percentages for all of the lithium niobate screenings. It is quite evident that the bacterial elution method mirrors the library percentages with the exception of serines, which are found more frequently in all bacterial screenings, without regards to the substrate. The acid eluted sequences show a much more diverse distribution of amino acids in the selected peptide populations.

The $-Z$ lithium niobate sequences show a trend of higher than average, P, [RKH], and [LAIVG] at the expense of [STY] residues. The $+Z$ lithium niobate sequences exhibit the same trend with a more dramatic divergence from the library averages. LiNbO_3 $+Z$ residues are highly basic, almost double the library frequency. Like the $-Z$ sequences, there is also a higher than average amount of [LAIVG] residues and a lower than average amount of [STY].

AA	Lib.	LiNbO ₃ -Z						LiNbO ₃ +Z					
		Avg.	All -Z	Δ	Acid	Δ	Bact.	Δ	All +Z	Δ	Acid	Δ	Bact.
A	6	7.4	1.4	8.9	2.9	6.9	0.9	7.5	1.5	7.1	1.1	7.7	1.7
C	0.5	0.5	0	0	-0.5	0.7	0.2	0.2	-0.3	0	-0.5	0.3	-0.2
D	2.8	1.6	-1.2	1.7	-1.1	1.4	-1.4	3.3	0.5	3	0.2	3.5	0.7
E	3.1	1.6	-1.5	2.8	-0.3	0.7	-2.4	1.5	-1.6	1.8	-1.3	1.3	-1.8
F	3.3	3.2	-0.1	2.8	-0.5	3.3	0	3.1	-0.2	2.4	-0.9	3.5	0.2
G	2.6	3.4	0.8	2.8	0.2	4	1.4	3.3	0.7	3.6	1	3.2	0.6
H	6.3	7.2	0.9	8.3	2	6.2	-0.1	8.3	2	11.9	5.6	6.4	0.1
I	3.4	3.6	0.2	3.9	0.5	3.3	-0.1	3.5	0.1	4.8	1.4	2.9	-0.5
K	2.8	4.3	1.5	6.1	3.3	2.9	0.1	4.6	1.8	8.3	5.5	2.6	-0.2
L	9.3	8.8	-0.5	9.4	0.1	8.3	-1	10.2	0.9	10.7	1.4	9.9	0.6
M	2.6	2.9	0.3	1.1	-1.5	4.3	1.7	1.9	-0.7	1.2	-1.4	2.2	-0.4
N	4.6	2.5	-2.1	2.8	-1.8	2.2	-2.4	3.8	-0.9	1.2	-3.4	5.1	0.5
P	12.2	13.3	1.1	16.1	3.9	11.6	-0.6	12.1	-0.1	13.7	1.5	11.2	-1
Q	5.1	7	1.9	6.7	1.6	7.2	2.1	4.4	-0.7	3	-2.1	5.1	0
R	4.7	4.1	-0.6	2.8	-1.9	4.7	0	4.2	-0.5	4.2	-0.5	4.2	-0.5
S	10	13.1	3.1	11.1	1.1	14.5	4.5	10.8	0.8	6.5	-3.5	13.1	3.1
T	11.1	9	-2.1	6.7	-4.4	10.5	-0.6	7.9	-3.2	7.7	-3.4	8	-3.1
V	3.9	2.7	-1.2	3.3	-0.6	2.2	-1.7	3.8	-0.2	4.8	0.9	3.2	-0.7
W	2.2	1.8	-0.4	0.6	-1.6	2.9	0.7	2.1	-0.1	2.4	0.2	1.9	-0.3
Y	3.6	2	-1.6	2.2	-1.4	1.8	-1.8	3.3	-0.3	1.8	-1.8	4.2	0.6
X	0	0.2	0.2	0	0	0.4	0.4	0.2	0.2	0	0	0.3	0.3
LAIVG	25.2	25.9	0.7	28.3	3.1	24.6	-0.6	28.3	3.1	31	5.8	26.9	1.7
RKH	13.8	15.5	1.7	17.2	3.4	13.8	0	17.1	3.3	24.4	10.6	13.1	-0.7
WF	5.5	5	-0.5	3.3	-2.2	6.2	0.7	5.2	-0.3	4.8	-0.7	5.4	-0.1
STY	24.7	24.1	-0.6	20	-4.7	26.8	2.1	22.1	-2.6	16.1	-8.6	25.3	0.6
DE	5.9	3.2	-2.7	4.4	-1.5	2.2	-3.7	4.8	-1.1	4.8	-1.1	4.8	-1.1
NQ	9.7	9.5	-0.2	9.4	-0.3	9.4	-0.3	8.1	-1.6	4.2	-5.5	10.3	0.6

Table 4.8: Summary of the different elution methods for the PhD12 single crystal LiNbO₃ -Z and +Z screenings broken down to the amino acid level. The percentage of amino acids in the *screenings* are tabulated next to the overall *library* average amino acid percents. Large differences (> ±3.0%) from the expected library percentages are highlighted.

The representative sequences for the $-Z$ (Table 4.9 A) and $+Z$ (Table 4.9 B) surface of lithium niobate are shown below. In agreement with the overall bacterial elution trends, the $-Z$ and $+Z$ LiNbO_3 percent position analyses show that most positions are designated by a [LAIVG] or [STY] residue, although there are a couple positions where [RKH] is indicated. There is no consensus between the two elution groups on the position of these amino acids. When analyzing the acid elution trends, some residues with more binding potential emerge. Both LiNbO_3 $-Z$ and $+Z$ show several positions where [RKH] is the most abundant group. In addition, the $-Z$ sequences indicate that [NQ] may be important for binding, while the $+Z$ sequences show 2 proline positions that may be significant.

The most interesting aspect of the percent position summary for the lithium niobate $+Z$ sequences (Table 4.9 B) is the resemblance of the acid-eluted sequences to the geological aragonite sequences. There are four basic positions and two proline positions, flanked by hydrophobic residues. The bacterial elution sequences hint at this similarity, but the most often encountered amino acids, [STY] and [LAIVG], dominate all but one position. This similarity will be further explored in the pattern analysis for these sequences.

A.	All -Z	%	-Z/ Acid	%	-Z/Bacterial	%
1	LAIVG	29.7	RKH/LAIVG	40	STY	30.4
2	LAIVG	32.4	<u>NQ</u>	26.7	LAIVG	39.1
3	LAIVG	29.7	LAIVG	33.3	STY	30.4
4	LAIVG	24.3	LAIVG	26.7	RKH	30.4
5	LAIVG	29.7	LAIVG	33.3	STY	34.8
6	STY/LAIVG	24.3	RKH/LAIVG	26.7	STY	26.1
7	STY	24.3	RKH/STY/LAIVG	20	STY	26.1
8	LAIVG	32.4	LAIVG	40	LAIVG	26.1
9	LAIVG	35.1	STY/LAIVG	33.3	LAIVG	34.8
10	STY	29.7	STY	26.7	RKH/STY	30.4
11	STY	32.4	STY	40	STY/LAIVG	26.1
12	STY	27	LAIVG	26.7	STY	39.1

B.	All +Z	%	+Z/ Acid	%	+Z/Bacterial	%
1	RKH	30	RKH	28.6	RKH/STY	30.8
2	LAIVG	27.5	RKH	42.9	STY/LAIVG	26.9
3	STY/LAIVG	27.5	STY	35.7	LAIVG	30.8
4	LAIVG	25	RKH/STY	28.6	LAIVG	23.1
5	LAIVG	25	RKH	35.7	STY/LAIVG	26.9
6	RKH/LAIVG	27.5	LAIVG	42.9	STY	23.1
7	STY/LAIVG	27.5	<i>P</i> /LAIVG	28.6	STY	34.6
8	LAIVG	30	LAIVG	28.6	LAIVG	30.8
9	LAIVG	35	LAIVG	50	LAIVG	26.9
10	LAIVG	35	LAIVG	57.1	STY	30.8
11	STY	25	<i>P</i>	21.4	STY	30.8
12	LAIVG	40	STY/LAIVG	28.6	LAIVG	46.2

Table 4.9: Highest percentages for each PhD12 position for the sequences screened against single crystal LiNbO₃ A) -Z surface and B) +Z surface. The averages of all PhD12 single crystal LiNbO₃ -Z and +Z surface sequences are included for comparison. Hydrophobic residues are in bold, basic residues are shaded gray, hydroxyl groups are bold and underlined, prolines are italicized, aromatic groups are underlined.

The Teiresias patterns for $-Z$ and $+Z$ lithium niobate are listed in Table 4.10. All of the top patterns for $\text{LiNbO}_3 -Z$ are long patterns only encountered twice, and are therefore, probably not highly significant. The $\text{LiNbO}_3 +Z$ patterns are somewhat more significant and seem to indicate that pairs or alternating basic and acidic amino acids are important for binding. In these patterns, the importance of the [RKH] residues can be seen in almost all of the top patterns.

Significance	#	$\text{LiNbO}_3 -Z$ Patterns
12	2	[LAIVG]PLI..A..H
11	2	L.H[STY]...[LAIVG].MT
11	2	A.AQ[STY].TP
11	2	QL.[RKH]..S.[LAIVG]L
11	2	GQ.[RKH].H..T
Significance	#	$\text{LiNbO}_3 +Z$ Patterns
14	2	FKHQ..[LAIVG][RKH][LAIVG]
13	7	[RKH].[DE]...[LAIVG]
12	4	[RKH].D.[RKH].[LAIVG]
12	3	[RKH]SD.[RKH].[LAIVG]
11	11	[STY].[STY]...[LAIVG]

Table 4.10: Patterns from the $-Z$ and $+Z$ LiNbO_3 single crystal sequences. The Probability is calculated using the amino acid frequencies of the specific libraries. The # is the number of instances the pattern was found, and the Significance is calculated by Teiresias.

It is also interesting to note that the highest-ranking pattern has a very similar motif to the two most often encountered sequences for aragonite, A20 and A21. The order is not conserved in any of these patterns, however, the motif is [WF], [RKH]₂, Q. It is not clear why exactly these $+Z$ sequences would be similar to aragonite sequences, but there do exist similarities on several levels.

Table 4.11 gives the aragonite sequences and several examples of lithium niobate +Z sequences with multiple charged residues, prolines, and aromatic groups. It should be noted that the LPZR5-26 sequence, KHMHWHPALNT, was found nine times in these screenings and exhibits quite a noticeable resemblance to A20 and A21.

A1	A20	A21
<u>F</u> KMPKE <u>H</u> ER <u>H</u> A <u>H</u>	LPKWQERQMLSA	LPPW <u>K</u> H <u>K</u> TSGVA
LiNbO ₃ +Z Sequences		
<u>W</u> HSD <u>K</u> HPLATRV	<u>K</u> H <u>M</u> H <u>W</u> HPPALNT	<u>H</u> LKAPSAPVKVK
G <u>F</u> <u>K</u> HQELVKGQE	KPTLY <u>A</u> PH <u>H</u> QPW	<u>H</u> HSPRHPPPLPR
<u>Y</u> PN <u>F</u> <u>K</u> HQSKL <u>H</u> L		

Table 4.11: LiNbO₃ sequences show similarity to geological aragonite sequence motifs. Charged residues are bold, prolines are italicized, Aromatic residues are underlined.

4.4: CONCLUSIONS

It was found that phage display screening is effective to various degrees in finding sequence patterns of amino acids that bind to materials that are chemically or stoichiometrically similar to calcium carbonate. It was found that polycrystalline powders or sintered substrates were difficult to achieve consensus with the presumed multiplicity of binding sites present. The powdered lithium niobate and single crystal lithium niobate did exhibit quite a few similarities to the patterns seen for geological aragonite, and showed potential for this similarity to be exploited in other experiments. In addition, it seems that the polarization of

the lithium niobate crystal does not strongly influence the types of charged residues found for the respective screenings. Both the LiNbO₃ -Z and +Z sequences were enriched in basic groups.

When broken down into the two different elution methods, there are some noteworthy trends. The acid elution method gives more dramatic divergences from the library frequencies than the bacterial elution method. This is true for almost all substrates screened here. For some reason, the single substrate screening/bacterial elution method does not elicit enough selection pressure to be effective in finding binding consensus.

Chapter 5: Peptide Structure Determination

In order to better understand possible mechanisms of peptide binding to calcium carbonate, it was necessary to describe the structure of the peptide. This was accomplished using a peptide, A20, found in multiple screenings against aragonite. Umbrella sampling molecular dynamics methods were used to probe the probability of the peptide occupying various conformations.

5.1: CHOICE OF MODEL PEPTIDE AND SURFACE

The A20 clone sequence (LPKWQERQMLSA) was selected for atomistic computer simulation study because A20 was the most commonly observed peptide in the early stages of screening for selective binding to the (110) surface of aragonite. It was later determined that the A21 phage clone (LPPWKHKTSVGA) has higher affinity. However, the A20 peptide, the only clone other than A21 that eluted multiple times, has high sequence similarity with A21 and is representative of peptides which bind well to aragonite. In the atomistic modeling studies, the A20 peptide was capped with an acetyl group (CH_3CO) on the N-terminus and a methylamine group (NHCH_3) on the C-terminus.

5.2: UMBRELLA SAMPLING MD METHOD

5.2.1: Theory

The techniques employed in modeling this A20 peptide are multiple time step, umbrella sampling methods developed for helical or semi-helical peptides of

intermediate length (5-20 residues) [1]. This method has been shown to be effective at simulating experimental peptide results both *in vacuo* and in water solution. This is an extended system multiple time step molecular dynamics scheme in conjunction with two dimensional umbrella sampling techniques.

Peptides studied experimentally [2-4] are best described as residing in a distribution of conformational states. Furthermore, these states are constantly interconverting. Experimental and theoretical studies of these conformers are of great importance to understanding how structure flexibility contributes to protein function as well as protein-related diseases, such as Alzheimer's disease. In the studies done here, it will be important to gaining insight on how structure flexibility contributes to or hinders face-specific inorganic crystal recognition by biopolymers.

5.2.2: Coordinate System and Peptide Set Up

The umbrella sampling molecular dynamics calculations of the A20 peptide in the gas phase and in water solution were performed using the techniques described by Samuelson and Martyna [1]. An all-atom CHARM22 force field [5] was employed to treat the interactions between the atoms in the system while the PINY_MD software [6] was used to perform the calculations. In umbrella sampling molecular dynamics, a model reaction coordinate is introduced to enhance barrier-crossing events between the many thermally accessible conformations available to the peptide. This is accomplished by performing a series of simulations that evolve the system along the reaction coordinate through the use of a biasing potential that encourages the system to sample a particular

region of the reaction coordinate. The effect of the biasing potential can be removed, exactly, after all the simulations are completed, by a self-consistent procedure and unbiased/correct results generated.

In this study, the “mean helical bond distance mean-standard deviation” two dimensional reaction coordinate was employed. The mean alpha helical hydrogen bond distance (oxygen of “i” residue to the hydrogen of “i + 4” residue) is denoted \bar{r} and the standard deviation of this distance is denoted, σ (Figure 5.1). If the mean of the hydrogen bond distances of the peptide is small ($\sim 2\text{\AA}$) and the standard deviation is small, the peptide is a perfect α -helix. If, in contrast, the mean is large, ($\sim 8\text{\AA}$) and the standard deviation is small the peptide is a perfect beta-strand. Large standard deviations indicate that the helical bonds are disordered, some long and some short.

In both gas phase and solution phase, the peptide was initially placed in a perfect α -helical conformation with the biasing potential set to encourage sampling of α -helices (small \bar{r} and σ). The peptide was then allowed to equilibrate under the dynamics described below and then calculations were extended to successively larger values of \bar{r} and σ using the equilibrated conformation of a previous run (with smaller \bar{r} and/or σ) to initiate the next calculation. In this way, the peptide is encouraged to sample a variety of conformational states ranging from helices to strands.

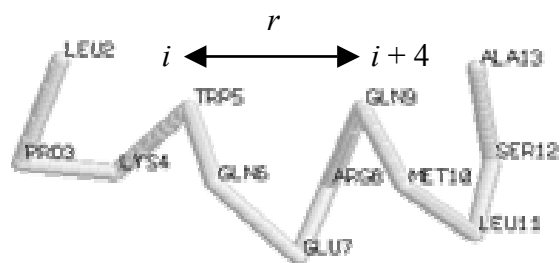


Figure 5.1: The A20 peptide in a helical conformation. \bar{r} is the average distance from the i to $i + 4$ residue.

5.2.3: Molecular Dynamics Parameters

Gas phase calculations were performed in the canonical ensemble, NVT ($T = 300$ K), using the Nose-Hoover chain extended system canonical dynamics method for a total simulation time of 60 ns spanning seventeen rbar parameters, $1.5 < \text{rbar} < 9.5 \text{ \AA}$, and seven σ parameters, $0.0 < \sigma < 3.0 \text{ \AA}$. Therefore, 119 simulations were run to sample the peptide phase space. In order to ensure convergence, the calculation was repeated three times, using the equilibrated conformers from the previous run at the same (rbar, σ) as the initial conformer in a new run.

In the solution phase calculations, the A20 peptide and one counter ion (Cl^-) was placed in the center of a computer water box, consisting of 2872 water molecules that did not overlap with the peptide. The water molecules were treated as rigid bodies and the water box was $\sim 46 \text{ \AA}$ on edge. The system was first equilibrated for 30 ps under NVT conditions ($T = 300$ K) with the peptide, taken from the gas phase calculation at appropriate value of the biasing parameter, kept frozen. Next, the full system was equilibrated for 60 ps under NPT conditions

($T=300$ K, $P_{\text{ext}}=1$ atm) using the extended system isothermal-isobaric dynamics scheme implemented in the PINY_MD simulation package [6]. Again, three runs at each r_{bar} and σ parameter value were performed for eighteen r_{bar} parameters, $1.5 < r_{\text{bar}} < 10.0$ Å, and seven sigma parameters, $0.0 < \sigma < 3.0$ Å, resulting in 126 simulations per run. The total simulation time per run was 30 ns or 240 ps of simulation time at each biasing parameter value.

5.2.4: Modeling Results

The A20 peptide was modeled *in vacuo* and in water using the 2-D umbrella sampling method. The unbiased probability plot of the A20 peptide *in vacuo*, shown in Figure 5.2, reveals a single peak centered around $r_{\text{bar}} = 4.2$ Å, $\sigma = 1.6$ Å. A perfect α -helix would have r_{bar} values between 1 and 3 Å and σ values near 0 Å. Therefore, at these r_{bar} and σ values, the helix is most likely a 3-10 helical chain with possible unraveling at either end. However, both the N-cap + 1 proline and the C-terminal alanine are consistent with helix stability [7, 8]. The apparent tightness of this helix is most likely due to the fact that there are no water molecules or ions to disrupt the hydrogen bonding or the possible interaction of the lysine and glutamic acid. These two residues would be very close in space due to the looser 3-10 helical structure.

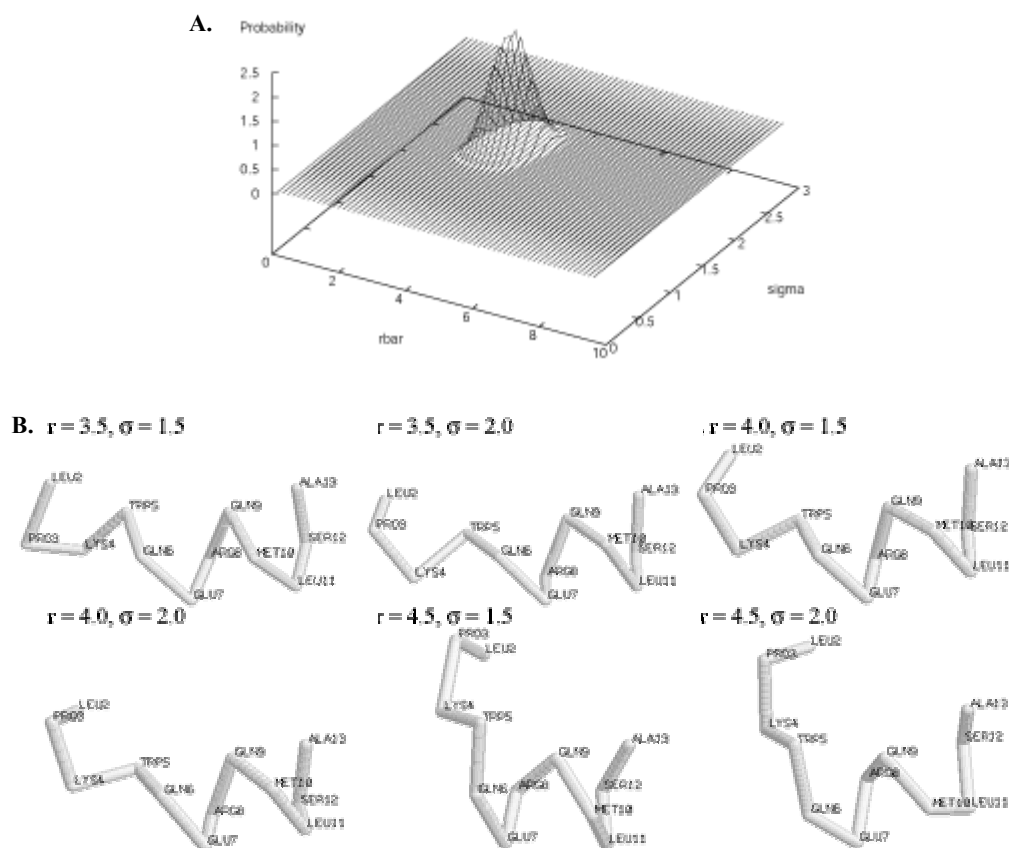


Figure 5.2: A) 2-D plot of the mean-standard deviation probability distribution function $P(r, \sigma; N, \beta)$ of the A20 peptide *in vacuo*. Maximum \approx $r_{\text{bar}} = 4.2$, $\sigma = 1.6$. B): Sampling of different A20 peptide conformers at r and σ values near the highest probability peak.

The A20 peptide exhibits solvent effects upon immersion in water. A20 unfolds slightly as it interacts with water molecules, giving rise to three main populations, which appear as peaks in the probability distribution function (Figure 5.3 A). There is an α -helical region with $r_{\text{bar}} = 3.0$ and $\sigma = 1.5$, a 3-10 helical

region, $r_{\text{bar}} = 5.25$ with high $\sigma = 2.3$, and an extended chain region, $r_{\text{bar}} = 8.3$, with lower sigma, $\sigma = 1.3$ (Figure 5.3 B). In addition, there is population in the probability distribution between the three main peaks, indicating that the peptide likely moves relatively easily between conformational states.

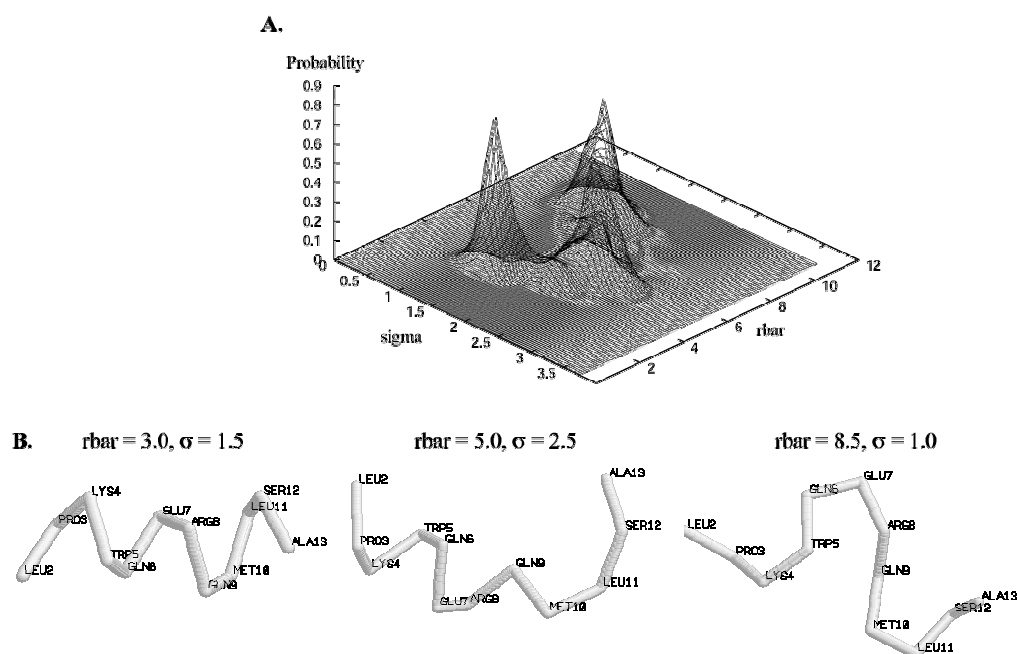


Figure 5.3: A) 2-D plot of the mean-standard deviation probability distribution function $P(r, \sigma; N, P_{\text{ext}}, \beta)$ of the A20 peptide in computer water. B) Sampling of different A20 peptide conformers at r and σ values near the three highest probability peaks in solution.

5.3: PEPTIDE NMR

The structure of the A20 peptide was also probed experimentally by use of peptide proton nuclear magnetic resonance (NMR). The synthetic N-terminally acetylated A20 peptide was synthesized by ICMB Core Facilities, Austin, Texas,

HPLC-purified, and lyophilized. The NMR sample was prepared in a 50 mM NaCl buffer at a peptide concentration of 10 mg/mL and analyzed with a 500 MHz Varian Inova spectrometer at 20°C and 30°C [Dr. David Hoffman, UT]. Spectra, which only show evidence of interaction only between the expected neighboring protons, are included in Appendix C.

This flexible description of the peptide in solution is confirmed by NMR data, which gives no evidence of stable secondary structure. The chemical shifts of the protons of the A20 peptide are listed in Table 5.1.

Residue	¹ H	σ _{1H}	σ _{1H-2}	Residue	¹ H	σ _{1H}	σ _{1H-2}	Residue	¹ H	σ _{1H}	σ _{1H-2}	
Ace	CH ₃	1.80										
1 Leu	HN	8.12		5 Gln	HN	7.69		9 Met	HN	8.20		
	Hα	4.36			Hα	3.96			Hα	4.30		
	Hβ	1.33	1.84		Hβ	1.79			Hβ	1.86	1.95	
	Hχ	1.53			Hχ	1.96			Hχ	2.37	2.46	
	Hδ	0.75			Hδ	6.95	7.38		CH ₃	2.56		
2 Pro	Hα	4.21		6 Glu	HN	7.91		10 Leu	HN	8.12		
	Hβ	1.33	1.99		Hα	4.02			Hα	4.22		
	Hχ	1.74	1.82		Hβ	1.86	1.94		Hβ	1.55		
	Hδ	3.38	3.68		Hχ	2.28			Hχ	1.51		
								Hδ	0.73	0.78		
3 Lys	HN	8.30		7 Arg	HN	8.10		11 Ser	HN	8.11		
	Hα	3.99			Hα	4.09			Hα	4.27		
	Hβ	1.61			Hβ	1.64			Hβ	3.70	3.75	
	Hχ	1.25			Hχ	1.70			OH	5.65		
	Hδ	1.51			Hδ	1.48						
	Hε	2.81			Hε	3.03						
	NH ₃	7.40			NH	7.04						
4 Trp	HN	7.81		8 Gln	HN	8.18		12 Ala	HN	8.13		
	Hα	4.44			Hα	4.12			Hα	4.16		
	Hβ	3.13	3.21		Hβ	1.85	1.94		Hβ	1.26		
	Hχ	7.16			Hχ	2.21						
	HN	10.11			Hδ	6.73	7.36					
	Hr ₁	7.35							NH ₂	HN	6.66	
	Hr ₂	7.10								HN	7.17	
	Hr ₃	7.01										
	Hr ₄	7.42										

Table 5.1: Proton NMR chemical shifts for the A20 synthetic peptide residues.

5.4: CONCLUSIONS

The A20 peptide is extremely significant as a peptide that binds to geological aragonite. Helical wheel analyses (Chapter 3.1.6) estimate that the three charged residues of A20 are positioned in space in a roughly triangular arrangement, with the K, E, and R residues on the same side of the helix. Umbrella sampling molecular dynamics simulations force the peptide into various helical conformations in order to determine the probability of the peptide occupying such conformations.

In vacuo, the A20 peptide adopts a helix with 3-10 helical properties, verifying the high probability of the three charged residues occupying spatial positions in a generally triangular fashion as predicted. In computer water, the A20 peptide unravels due to solvent effects, which allow the peptide backbone and side chains to interact with water, making the bonds to itself less effective. However, the conformations the A20 peptide in water can exhibit in this model are quite extensive, ranging from an α -helix to a 3-10 helix to an extended chain conformation. Furthermore, there is evidence in this model that the peptide easily introconverts between all of these states. NMR experiments confirm only the fact that there is no dominant conformation or ensemble of conformations that can be detected.

Peptides with these characteristics that are known to bind to inorganic materials in nature are typically mineral growth inhibitor proteins or peptides. The flexibility in the backbone structure allows the peptide to remain close to the surface even if one or more bonds with the surface are disrupted [9, 10].

However, when fused onto the phage proteins, there is a greater chance that the helical nature of the peptide is conserved, since both the N-termini of the pIII and the pVIII are α -helical in nature. Furthermore, the N-cap + 1 proline and alanine are known helix-stabilizing residues. Peptides known to be mineral growth enhancing are much more rigid in their structure [11]. Therefore, in order to increase the nucleation properties of these selected peptides, a more rigid peptide structure may be necessary to develop. Conversely, in order to increase the inhibiting properties, it may be necessary to develop a longer peptide with more binding residues, as peptides known to inhibit calcite in nature have 2-3 times the number of charged residues as A20 [9, 10, 12].

REFERENCES:

1. Samuelson, S.O. and G.J. Martyna, *Two dimensional umbrella sampling techniques for the computer simulation study of helical peptides at thermal equilibrium: The 3K(I) peptide in vacuo and solution*. Journal of Chemical Physics, 1998. **109**(24): p. 11061-11073.
2. Martinez, G. and G. Millhauser, *Ftir Spectroscopy of Alanine-Based Peptides - Assignment of the Amide I' Modes for Random Coil and Helix*. Journal of Structural Biology, 1995. **114**(1): p. 23-27.
3. Rohl, C.A. and R.L. Baldwin, *Exchange Kinetics of Individual Amide Protons in N-15-Labeled Helical Peptides Measured by Isotope-Edited Nmr*. Biochemistry, 1994. **33**(25): p. 7760-7767.
4. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually Stable Helix Formation in Short Alanine-Based Peptides*. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(14): p. 5286-5290.
5. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.

6. Tuckerman, M.E., et al., *Exploiting multiple levels of parallelism in Molecular Dynamics based calculations via modern techniques and software paradigms on distributed memory computers*. Computer Physics Communications, 2000. **128**(1-2): p. 333-376.
7. Richardson, J.S. and D.C. Richardson, *Amino-Acid Preferences for Specific Locations at the Ends of Alpha-Helices*. Science, 1988. **240**(4859): p. 1648-1652.
8. Rohl, C.A., W. Fiori, and R.L. Baldwin, *Alanine is helix-stabilizing in both template-nucleated and standard peptide helices*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(7): p. 3682-3687.
9. Gerbaud, V., et al., *Mechanism of calcite crystal growth inhibition by the N-terminal undecapeptide of lithostathine*. Journal of Biological Chemistry, 2000. **275**(2): p. 1057-1064.
10. Wustman, B.A., et al., *Characterization of a Ca(II)-, mineral-interactive polyelectrolyte sequence from the adhesive elastomeric biomineralization protein lustrin A*. Langmuir, 2003. **19**(22): p. 9373-9381.
11. Cerini, C., et al., *Nucleation of calcium oxalate crystals by albumin: Involvement in the prevention of stone formation*. Kidney International, 1999. **55**(5): p. 1776-1786.
12. Michenfelder, M., et al., *Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains*. Biopolymers, 2003. **70**(4): p. 522-533.

Chapter 6: Protein Engineering and Materials Growth

In order to investigate the crystal nucleation potential of the peptides found to be significant in crystal binding, we have used the phage itself as a scaffold for protein-mediated crystal growth. By engineering the phage pVIII to display the selected peptides and increasing the expression level, we can achieve longer-range order than peptide display on the phage pIII. The engineered phage were then used to grow hybrid organic-inorganic materials.

6.1: DISPLAY OF PEPTIDE ON M13 PHAGE OUTER COAT PROTEIN

The pVIII is the phage major coat protein, which is present in approximately 2700 copies [1]. Overlapping pVIII encase the phage DNA, allowing a somewhat flexible coat. X-ray studies have shown that the pVIII encasement is continuously alpha-helical [2] with the positively charged C-terminal region interacting with the phosphate backbone of the DNA (Figure 6.1). Mutation studies have shown that the charge ratio appears to dictate the number of pVIII encasing the DNA [3], generating longer phage particles using engineered pVIII with fewer positive lysine residues.

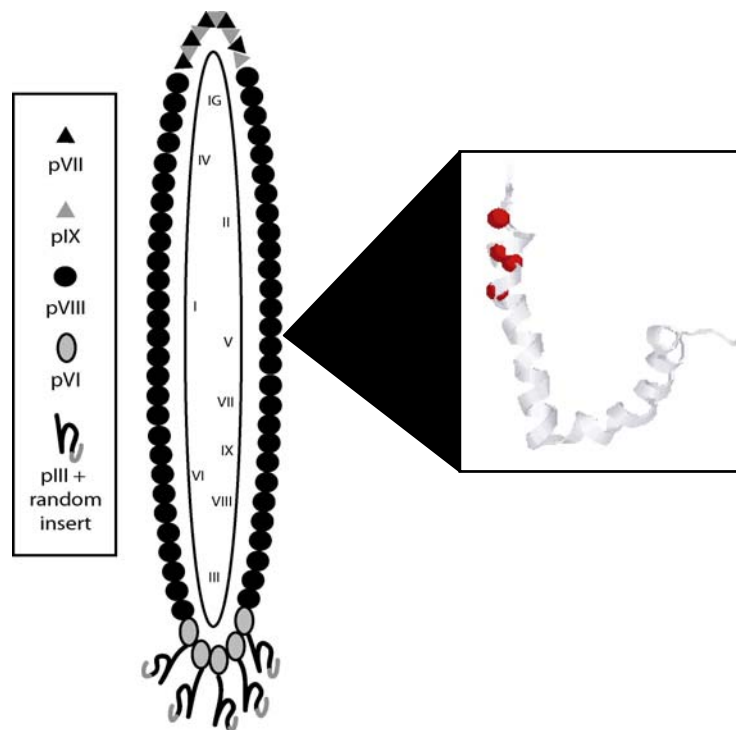


Figure 6.1: Phage diagram with the molecular structure [4] of the phage pVIII (PDB ID: 2CPB) shown with the positively charged C-terminal (lysine residues in red) helix facing the negatively charged phage DNA. Engineering of the pVIII is accomplished by fusing a peptide onto the N-terminus.

The advantage of peptide display on the pVIII is that a much larger number of displayed peptides can be attained [5] compared to the pIII, of which there are approximately 5 copies. There are two methods of pVIII display: recombinant phage and hybrid phage. Recombinant phage have all ~2700 copies of the pVIII engineered, whereas hybrid phage have both the wild type and engineered pVIII displayed. Recombinant phage systems are limited in their ability to display large peptide. Typically the displayed peptide is limited to about 6-8 amino acids in length [6]. Depending on the sequence of the displayed

peptide, there can also be a loss of stability of the phage coat, leading to phage particles that easily fall apart [unpublished data – Hayhurst, Flynn]. Hybrid phage particles can tolerate larger peptides, but can only express the engineered pVIII at about 1-10% compared to the wild type pVIII.

In order to increase the valency of the selected peptides on the phage particle, we have employed the hybrid 8 + 8 cloning system. In this system, the pVIII of the major coat are produced by both the bacterial plasmid containing the engineered gene VIII and the phage wild type gene VIII. The pMoPac33 vector [unpublished - Hayhurst, Georgiou, Iverson] has been employed to express our selected peptides (Figure 6.2). The vector contains the M13 gene VIII (g8p) with SfiI sites inserted between the N-terminal signal sequence and the N-terminal mature pVIII sequence. The vector also contains an ampicillin resistance gene (amp) as well as the lacZ , which allows the expression of the engineered protein to be controlled by induction with IPTG. The DNA that codes for the inserted peptide sequence is inserted at the SfiI site, so that after cleavage of the signal sequence, the inserted sequence is positioned at the N-terminus of the mature major coat protein.

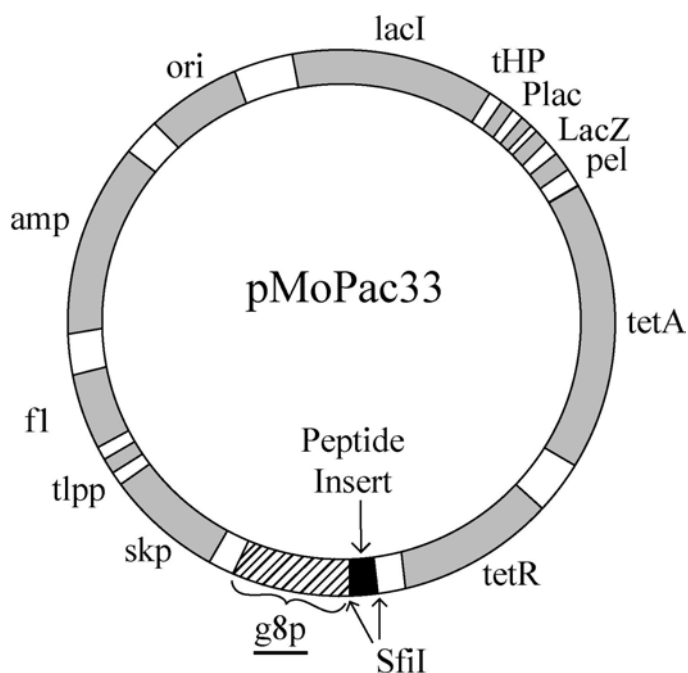


Figure 6.2: pMoPac33 vector (~6760 bp) used for cloning hybrid peptide fusions on the N-terminus of the pVIII. The insert region is shown in black and the gene VIII is striped.

The two sequences that were cloned into the pVIII phage coat were the A1 and the A20 sequences, both selected against aragonite. The insert DNA was ordered (Integrated DNA Technologies) with 5' phosphorylated sticky ends designed to match the SfiI sites in the pMopac33 (Figure 6.3). The plasmid was digested with SfiI and the ~5kb piece was extracted from a 1% agarose gel and then ligated with the insert DNA (1:10 mole ratio) using T4 DNA ligase. The DNA was then washed and desalted before electroporation into the Tg1 *E. coli* bacterial strain. The undigested pMopac33 plasmid was electroporated as a positive control, and the digested plasmid ligated with no insert was

electroporated as a negative control. All were grown on LB-ampicillin plates (100ug/mL ampicillin) in order to ensure the selection of the pMopac33 plasmid.

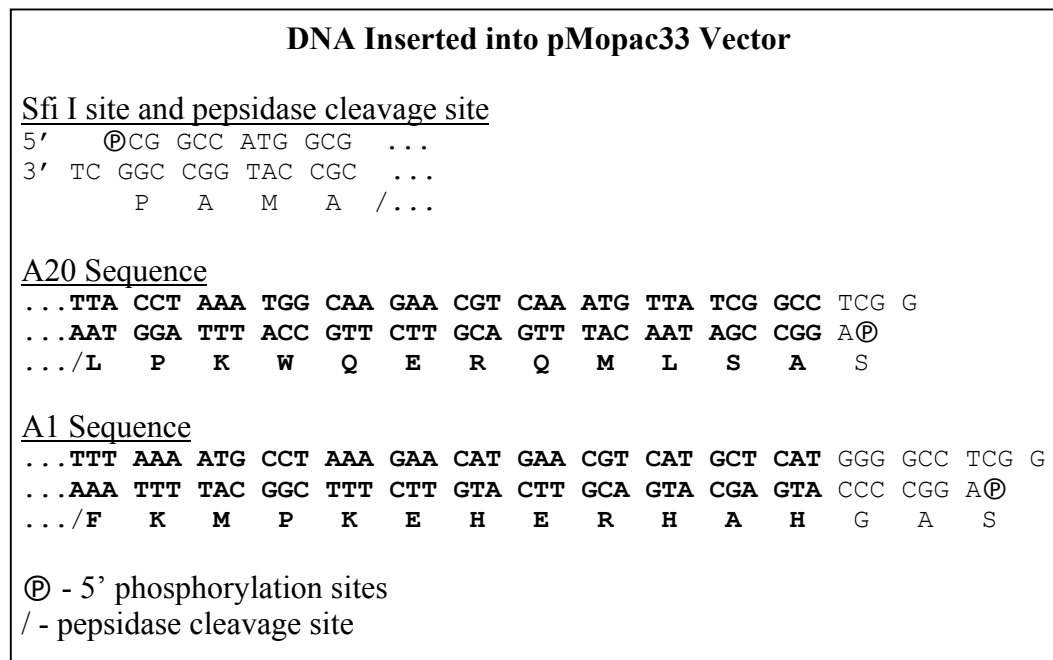


Figure 6.3: DNA sequences inserted into the pMopac33 vector for the two aragonite peptides A20 and A1. Both sequences contain the leading Sfi I site and pepsidase cleavage site before the peptide codons.

The transformed plasmid was purified and sequenced (Institute for Cellular and Molecular Biology, UT) to confirm the proper ligation of the insert into the plasmid. The transformed cells were grown in Terrific Broth, 2% glucose with 100 µg/mL ampicillin to an OD₆₀₀ of 0.500. The cells were then infected with helper phage at a multiplicity of infection of 10:1. Some preparations used an anti-streptavidin pIII NEB phage (obtained from Seung-wuk Lee), while other preparations used M13K07 helper phage (kanamycin resistant M13 strain obtained from NEB) and 35 µg/mL kanamycin to ensure superselection. The

culture was allowed to incubate at 37°C without shaking for one hour. Then, the bacteria were centrifuged and resuspended in Terrific Broth (no glucose), 100 µg/mL ampicillin, 1mM IPTG and 35 µg/mL kanamycin, if needed. The hybrid phage were grown overnight with shaking. The bacteria were then centrifuged out and the phage were pelleted with PEG-NaCl and resuspended in TBS pH 7.5.

The phage were then tested for expression of the engineered pVIII by SDS protein gel electrophoresis and by Matrix Assisted Laser Desorption Ionization (MALDI) Mass Spectroscopy. The two gels used were NuPage 12% bis-tris gels with SDS-MES running buffer were run at 200 V, 100 mA and Novex 16% tricine gel with SDS running buffer 125 V, 80 mA. These were all run with the Invitrogen Mark12 standard and stained with either Coomassie Blue or silver stain (Invitrogen). MALDI samples were denatured using heat and trifluoroacetic acid (TFA), mixed with a synapinic acid matrix solution, and analyzed in the Perseptive Biosystems Voyager MALDI-Time of Flight Mass Spectrometer (Institute for Cellular and Molecular Biology, UT).

The protein gels were determined to be unsatisfactory for determining engineered pVIII expression. Even with the more sensitive silver staining, the larger protein band was either too faint to detect or was smeared by the wild type band when higher concentrations were used. When the same samples were analyzed by MALDI-MS, the little to no engineered protein was detected, indicating that the expression level is very low, probably less than 1%. The engineered A7-pVIII provided in the pAKdisplay.A7 vector (gift of Andrew Hayhurst, UT), was used as a control and expressed at a detectable level.

6.2: PHAGE-MEDIATED CALCIUM CARBONATE CRYSTAL GROWTH

Calcium carbonate nanoparticles were grown from solution on the engineered phage coat proteins. Crystals were grown by solution diffusion across a membrane (regenerated cellulose, MWCO 5 kDa) in a DispoEquilibrium Dialyzer (Harvard/AmiKa). A 1 mM calcium carbonate (CaCl_2) solution (75 μL) was placed in one chamber of the DispoEquilibrium Dialyzer along with $\sim 5 \times 10^8$ p.f.u. of phage and 75 μL of a 1 mM ammonium bicarbonate (NH_4HCO_3) solution was placed in the other chamber. The ions were allowed to diffuse across the membrane for 12-24 hours and the reaction was quenched by twice removing the solution from the non-phage chamber and adding 75 μL of 1 mM NH_4HCO_3 to ensure that the final solution pH was above 8.0. Two different engineered phage were used as scaffolds for crystal growth: A1-pVIII and A20-pVIII.

All samples (1-10 μm of nucleation solution) were placed on carbon-coated copper TEM grids (Ted Pella) after the removal of the formvar backing. Samples were vacuum desiccated overnight and imaged in the JEOL 2010F Transmission Electron Microscope (Texas Materials Institute). The images were analyzed using fast Fourier transformations (FFT) in order to determine the d-spacings and therefore crystal structure. Since the d-spacings of the three forms of calcium carbonate are very similar, the angles between the planes were calculated using the crystal unit cell parameters and the Miller indices (hkl values) (Equations 6.1 and 6.2).

$$\text{Orthorhombic: } \cos \theta = \frac{\left(\frac{h_1 h_2}{a^2} + \frac{k_1 k_2}{b^2} + \frac{l_1 l_2}{c^2} \right)}{\sqrt{\left(\frac{h_1^2}{a^2} + \frac{k_1^2}{b^2} + \frac{l_1^2}{c^2} \right) \left(\frac{h_2^2}{a^2} + \frac{k_2^2}{b^2} + \frac{l_2^2}{c^2} \right)}} \quad (6.1)$$

$$\text{Hexagonal: } \cos \theta = \frac{\left[h_1 h_2 + k_1 k_2 + \frac{1}{2} (h_1 k_2 + h_2 k_1) + \frac{3a^2}{4c^2} l_1 l_2 \right]}{\sqrt{\left(h_1^2 + k_1^2 + h_1 k_1 + \frac{3a^2}{4c^2} l_1^2 \right) \left(h_2^2 + k_2^2 + h_2 k_2 + \frac{3a^2}{4c^2} l_2^2 \right)}} \quad (6.2)$$

6.2.1: pVIII Mediated Crystal Growth

The crystals grown using the A20-pVIII phage construct show many nanoparticles under high-resolution microscopy. These crystals are well crystallized and range in size from 4-10 nm. While there are many crystals with only one d-spacing, and thus impossible to identify, it was found that there was a mixture of all three phases present. Figure 6.4 shows the crystals and FFT indexing for a vaterite pattern. Figure 6.5 shows the crystals and FFT indexing of an aragonite pattern.

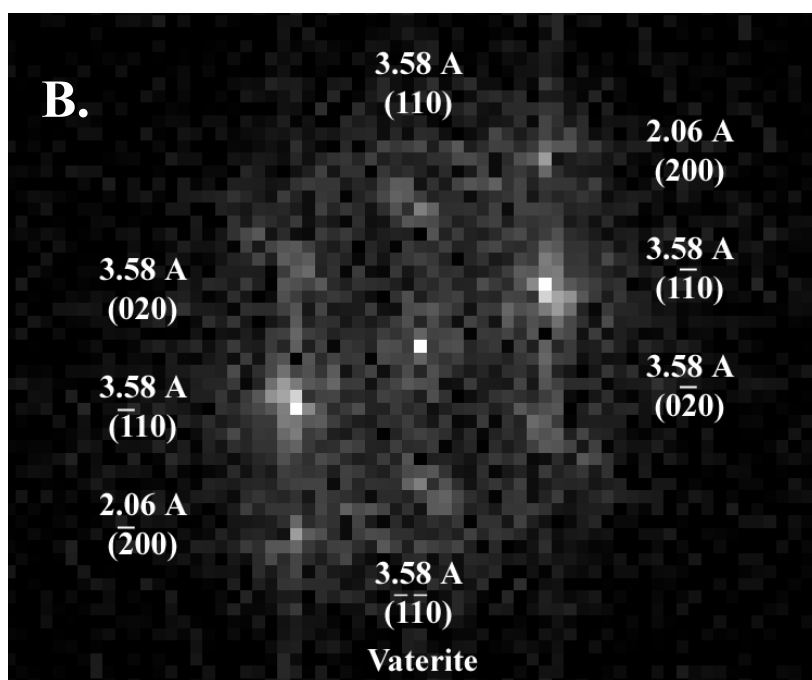
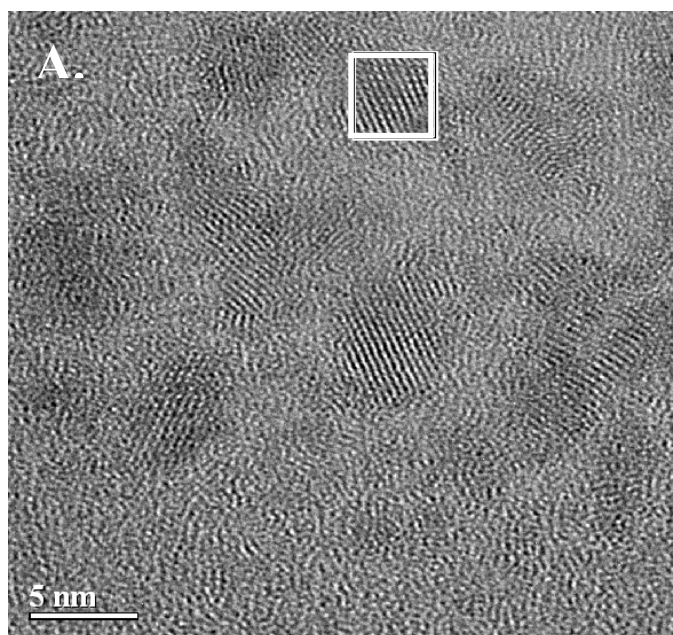


Figure 6.4: A) High resolution TEM image of crystals grown with the A20-pVIII phage. B) FFT image of selected area in A, showing the various planes of vaterite.

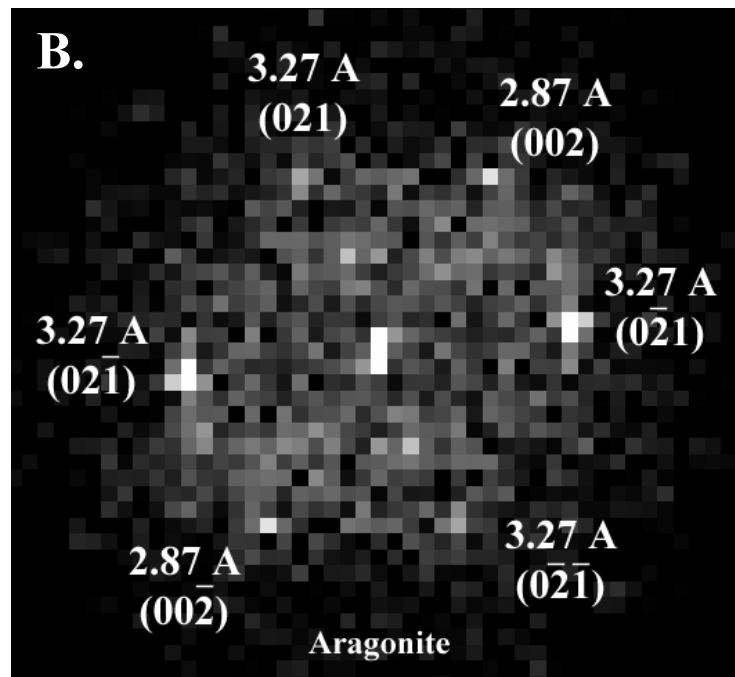
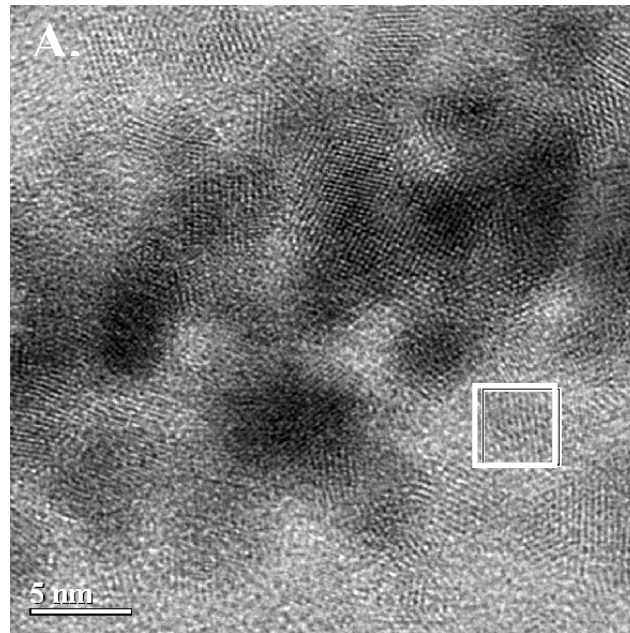


Figure 6.5: A) High resolution TEM image of crystals grown with the A20-pVIII phage. B) FFT image of selected area in A, showing the various planes of aragonite.

Crystals grown using the A1-pVIII phage as a support, also exhibited good nanocrystal growth. These crystals, about the same magnitude as the A20-pVIII crystals, also showed a mixture of the three phases. Figure 6.6 shows one example of an aragonite crystal and its indexed FFT pattern.

Control nucleation experiments, $\text{CaCl}_2 + \text{NH}_4\text{HCO}_3$ with no phage, did not have an abundance of crystals. Most were single crystal, single d-spacing crystals, most likely calcite (Figure 6.7). There was no evidence of aragonite in these samples.

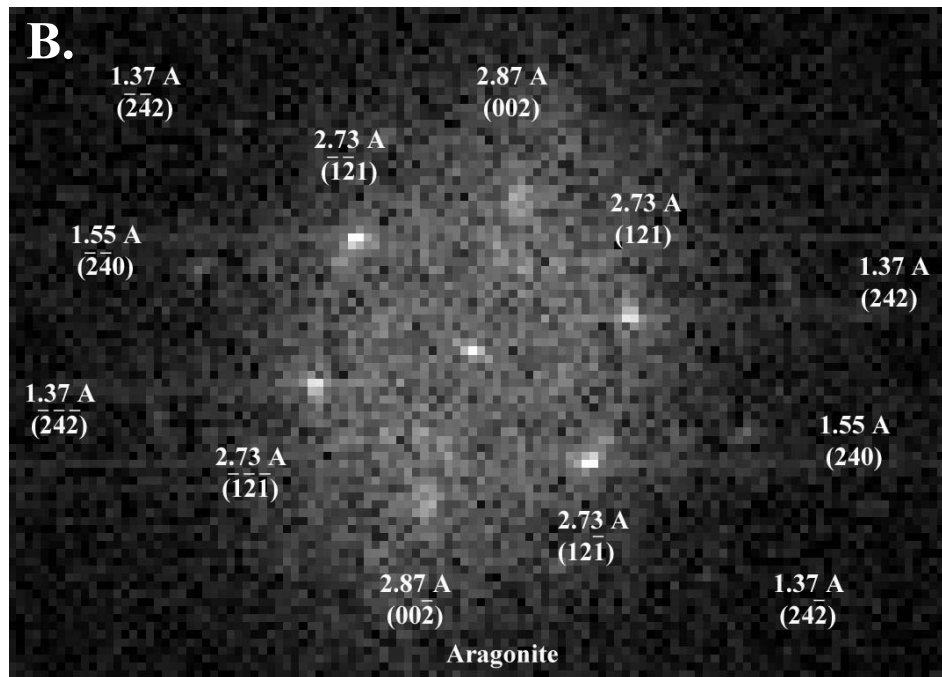
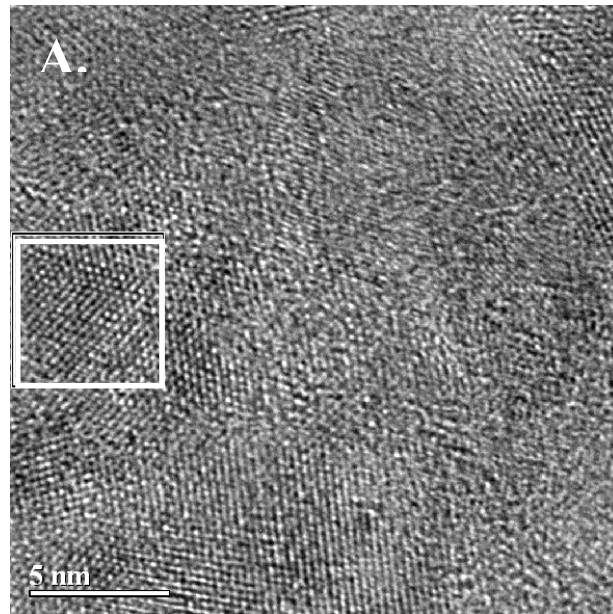


Figure 6.6: A) High resolution TEM image of crystals grown with the A1-pVIII phage. B) FFT image of selected area in A, showing the various planes of aragonite.

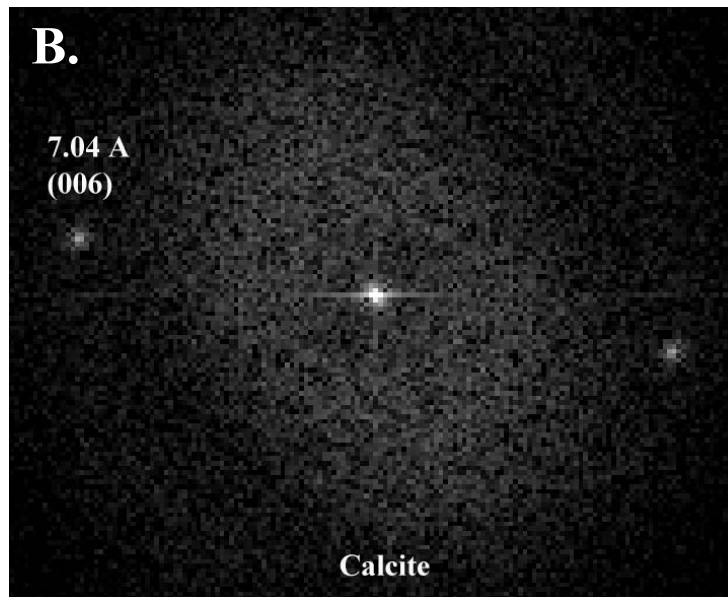
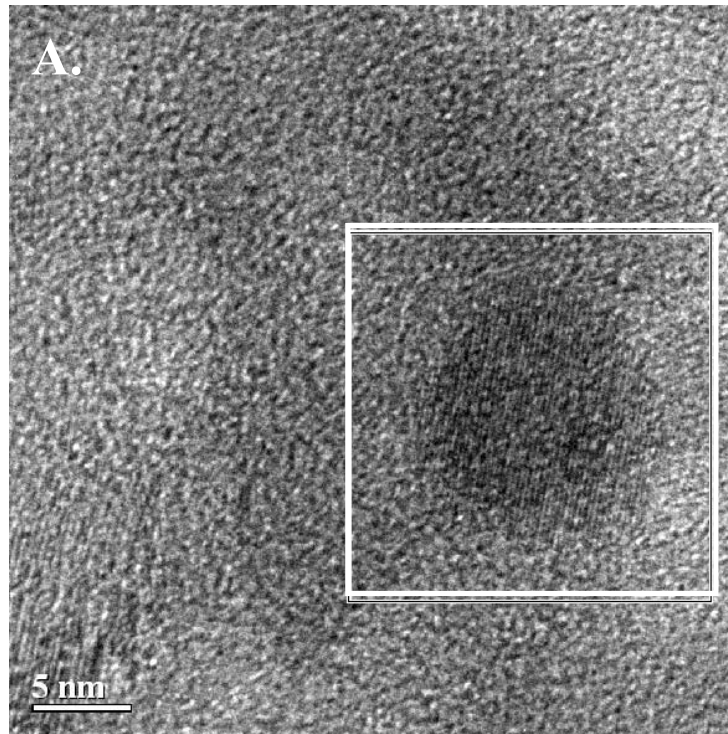


Figure 6.7: A) High resolution TEM image of crystals grown without phage. B) FFT image of selected area in A, showing the (006) plane of calcite.

6.3: SEMICONDUCTING FILM GROWTH ON OPTICAL WAVEGUIDE MATERIAL

Lithium niobate is an optical material used in laser waveguides. The light is contained within the optical mode of the waveguide due to the difference in refractive indices between the waveguide and the surrounding material or air (Figure 6.8). The waveguide can be fabricated by ion bombardment of the surrounding material to change the refractive index or by etching away the surrounding material. Over long distances, the light intensity can diminish, thus necessitating an optical amplifier to absorb and re-emit the light at a higher intensity. Waveguide amplifiers are made with semi-conducting quantum dots embedded in the waveguide itself. A potential future use of this hybrid inorganic-organic materials growth is a self-assembling, optical waveguide amplifier, using the specific recognition of the selected peptides to bind to the waveguide and also grow semi-conducting nanoparticles that couple to the optical mode.

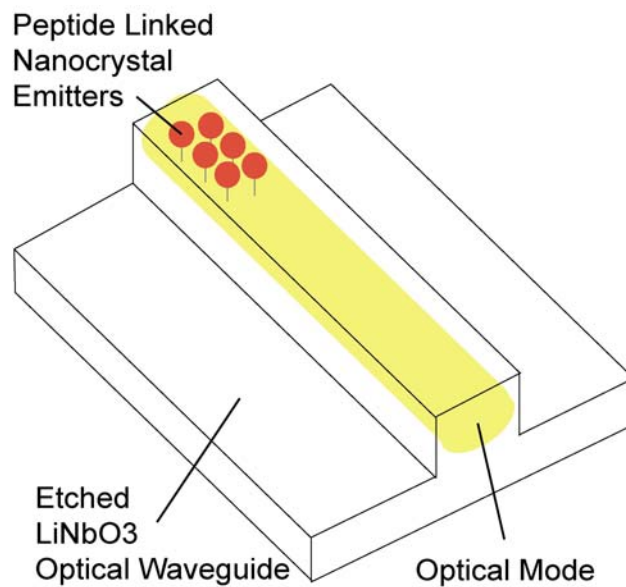


Figure 6.8: Diagram of a LiNbO_3 optical waveguide with semiconductor nanocrystal emitters linked by bifunctional peptides, which bind and nucleate LiNbO_3 and ZnS , respectively.

6.3.1: Double Display Phage System

The concept used in these experiments is an engineered phage that displays both peptides that recognize lithium niobate and peptides that capture ZnS nanoparticles grown in solution. The lithium niobate recognition peptide was chosen to be the A1 peptide selected from the aragonite screenings. This peptide also has a pattern similar to those found to be significant in the lithium niobate powder and single crystal +Z screenings. Overall the pattern is basic, but contains some alternating acidic residues (Table 6.1).

A1 Sequence	LiNbO ₃ Patterns	A7 ZnS Sequence
FKMPKEHERHAH	[RKH].[DE]...[LAIVG] [RKH][RKH][RKH]	CNPMHQNC

Table 6.1: A1 sequence compared to similar high significance LiNbO₃ patterns. A7 sequence screened against and shown to capture ZnS [7].

Tg1 *E. coli* cells transformed with the A1-pVIII sequence in the pMopac33 vector as described above were used as the host cells for the second transformation. The A7-pVIII sequence, specific for ZnS [8], was provided in the pAKdisplay.A7 vector (gift of Andrew Hayhurst, UT), which contains the lac operon and a chloramphenicol resistance gene. These cells were amplified, the plasmid vector was purified, and electroporated into the Tg1-pMopac33 *E. coli* cells.

The doubly transformed cells were grown in Terrific Broth 2% glucose with chloramphenicol and ampicillin selection to an OD₆₀₀ of 0.500. The bacteria were then infected with M13K07 helper phage at a multiplicity of infection of 10:1. The culture was allowed to incubate at 37 °C without shaking for one hour and then the bacteria were centrifuged and resuspended in Terrific Broth (no glucose) with chloramphenicol, ampicillin and kanamycin selection. The hybrid phage were grown overnight and the bacteria were then centrifuged out. The phage were pelleted with PEG-NaCl and resuspended in TBS pH 7.5.

6.3.2: Phage Film Growth

Phage films were grown on both sides of a polarized single crystal LiNbO₃ using the doubly displayed A1-A7 phage. Control films were also grown using

singly displayed A7 phage. The films were grown using downward meniscus force due to evaporation until dry (16 days) starting with 0.1 mg/mL phage in TBS pH 7.5 solutions [9]. Each of the four surfaces, A1-A7 on LiNbO₃ +Z, A1-A7 on LiNbO₃ -Z, A7 on LiNbO₃ +Z, A7 on LiNbO₃ -Z, was then incubated with 50 μL 1mM ZnCl₂ for 12 hours. Then, 50 μL 1mM NaS was added to each surface, and the surfaces were incubated on ice (0–2°C) for 12 hours. The surfaces were removed and rinsed briefly with 100 μL DDH₂O. Each surface was mounted with carbon tape on a SEM stub, coated with chromium and imaged using a Hitachi S-4500 field emission SEM (Texas Materials Institute, UT). The resulting ZnS nanoparticle coated LiNbO₃ substrate images are shown in Figure 6.9.

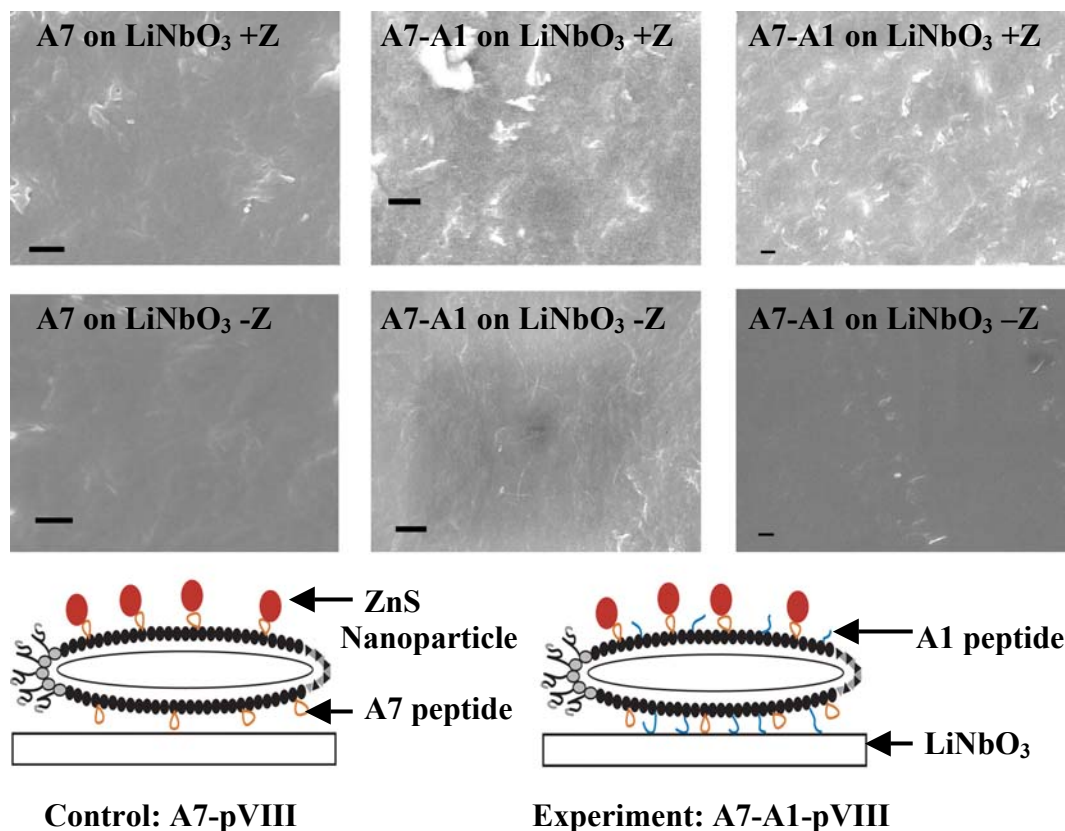


Figure 6.9: SEM image of hybrid materials growth using doubly displayed phage. Control phage (far left) and experiment phage (center and far right) were exposed to both LiNbO₃ +Z (row 1) and LiNbO₃ -Z (row 2). ZnS (white crystals) were grown using the A7 peptide to mediate growth. All scale bars 1 μm.

The phage films do show some selectivity for the lithium niobate +Z surface, although there appears to be some non-specific binding of the A7-pVIII phage on the same surface. Phage binding to the lithium niobate -Z surface is minimal for both the A7-pVIII and A7-A1-pVIII phage. The A7-A1-pVIII phage binding to the lithium niobate +Z surface is quite extensive.

6.4: CONCLUSIONS

It has been shown that it is indeed possible to grow specific phases of calcium carbonate using genetically engineered phage that express the peptides selected against the specific phase. Both the A20-pVIII construct and the A1-pVIII construct appear to have the ability to grow aragonite under ambient pressures and temperatures and without additives such as magnesium to affect the solvation of the calcium and carbonate ions.

It is unclear whether the aragonite growth is due to the long-range order of the peptide expressed on the major phage coat or if a small number of peptides could work as effectively. Since the expression level of the engineered pVIII particles was very low (~1%), long-range ordering of these peptides may not be necessary for aragonite growth. This has been shown to be true for other phage-mediated semiconductor particle growth [10]. However, other synthetic polypeptide routes to aragonite growth have indicated that longer-range order is important, as acidic β -sheets are better able to grow aragonite than acidic α -helices of the same overall amino acid composition [11]. The flexibility of the peptide fusions may be reduced when expressed at the end of the stable pVIII α -helix, and this may indeed facilitate ion complexation at the right distances for aragonite growth. It should also be noted that the peptides used in these studies were not highly acidic as are most proteins or peptides associated with aragonite growth. In fact, these peptides are more basic in nature, indicating that carbonate templating may be a possible mechanism for aragonite growth.

The use of the doubly hybrid phage constructs exemplifies the enormous potential of this system to bind one surface and simultaneously influence the growth of a separate crystal structure. While it is unlikely that A1 binding to $\text{LiNbO}_3 + \text{Z}$ is extremely specific due to the fact that it was selected for aragonite, it does show the possibility of deriving such a system using these techniques. This example demonstrates not only a proof of concept, but also a highly useful potential device.

This system also has many potential uses outside of optical and electronic hybrid materials growth. Other experiments were attempted using aragonite phage to grow hydroxylapatite from solution, since the chemical composition and structure of the two materials are so similar. Both engineered pVIII phage constructs as well as phosphorylated coat protein phage construct were used as materials growth templates. The samples did not produce enough material to be detected using X-ray diffraction. However, this represents a promising route for advances in synthetic bone material production.

Biological aragonite has already proven to be an effective bone replacement material. Powdered nacre of the giant oyster, *Pinctada maxima*, has actually been used as an implant in animal and human bones. It was shown to biodissolve non-toxically and even stimulate new bone growth [12]. Nacre has three extremely important advantages to conventional bone replacement materials: 1) it can convert to natural bone, 2) the organic component gives nacre a fracture resistance much better than conventional hydroxylapatite and 3) soluble

proteins from nacreous aragonite stimulate osteoblast differentiation [13-15]. These attributes are also attainable using our synthetic phage system.

REFERENCES:

1. *Phage Display of Peptides and Proteins: A Laboratory Manual*, ed. B.K. Kay, J. Winter, and J. McCafferty. 1996, San Diego: Academic Press, Inc.
2. Marvin, D.A., et al., *Molecular-Models and Structural Comparisons of Native and Mutant Class-I Filamentous Bacteriophages Ff (Fd, F1, M13), If1 and Ike*. Journal of Molecular Biology, 1994. **235**(1): p. 260-286.
3. Greenwood, J., G.J. Hunter, and R.N. Perham, *Regulation of Filamentous Bacteriophage Length by Modification of Electrostatic Interactions between Coat Protein and DNA*. Journal of Molecular Biology, 1991. **217**(2): p. 223-227.
4. Papavoine, C.H.M., et al., *Solution structure of the M13 major coat protein in detergent micelles: A basis for a model of phage assembly involving specific residues*. Journal of Molecular Biology, 1998. **282**(2): p. 401-419.
5. Perham, R.N., et al., *Engineering a Peptide Epitope Display System on Filamentous Bacteriophage*. Fems Microbiology Reviews, 1995. **17**(1-2): p. 25-31.
6. Greenwood, J., A.E. Willis, and R.N. Perham, *Multiple Display of Foreign Peptides on a Filamentous Bacteriophage - Peptides from Plasmodium-Falciparum Circumsporozoite Protein as Antigens*. Journal of Molecular Biology, 1991. **220**(4): p. 821-827.
7. Mao, C.B., et al., *Viral assembly of oriented quantum dot nanowires*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(12): p. 6946-6951.
8. Lee, S.-W., et al., *Ordering of quantum dots using genetically engineered viruses*. Science (Washington, DC, United States), 2002. **296**(5569): p. 892-895.

9. Lee, S.-W., B.M. Wood, and A.M. Belcher, *Chiral Smectic C Structures of Virus-Based Films*. Langmuir, 2003. **19**(5): p. 1592-1598.
10. Flynn, C.E., et al., *Synthesis and organization of nanoscale II-VI semiconductor materials using evolved peptide specificity and viral capsid assembly*. Journal of Materials Chemistry, 2003. **13**(10): p. 2414-2421.
11. Levi, Y., et al., *Control over aragonite crystal nucleation and growth: an in vitro study of biomineralization*. Chemistry--A European Journal, 1998. **4**(3): p. 389-396.
12. Lopez, E., et al., *Demonstration of the Capacity of Nacre to Induce Bone-Formation by Human Osteoblasts Maintained Invitro*. Tissue & Cell, 1992. **24**(5): p. 667-679.
13. Rousseau, M., et al., *The water-soluble matrix fraction from the nacre of Pinctada maxima produces earlier mineralization of MC3T3-E1 mouse pre-osteoblasts*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 2003. **135**(1): p. 1-7.
14. Mouries, L.P., et al., *Bioactivity of nacre water-soluble organic matrix from the bivalve mollusk Pinctada maxima in three mammalian cell types: fibroblasts, bone marrow stromal cells and osteoblasts*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 2002. **132**(1): p. 217-229.
15. Almeida, M.J., et al., *Comparative effects of nacre water-soluble matrix and dexamethasone on the alkaline phosphatase activity of MRC-5 fibroblasts*. Journal of Biomedical Materials Research, 2001. **57**(2): p. 306-312.

Chapter 7: Conclusions

7.1: PEPTIDE SCREENING, PATTERN ANALYSIS AND HOMOLGY

The peptide screening and statistical significance method of pattern analysis has proven to be an extremely useful tool in determining peptide binding motifs for inorganic substrates. This method negates the effects of preferential amplification of the phage and makes pattern ranking straightforward and objective.

However, the screening methods employed in this dissertation leave much room for further studies. The standard pH elution method leaves many phage bound to the surface in the elution step. It is unclear if these phage are so tightly bound that they prevent the crystal from etching or if the phage are just re-binding to the newly etched surface. It would be interesting to design new screening protocols to determine this. The single substrate screening method may hold potential, but the selection pressure does not appear to be high enough to attain consensus.

As for the substrates themselves, only sequences selected against geological aragonite gave patterns of high significance. These aragonite sequences were coincidentally very similar to many lithium niobate +Z sequences, despite the fact that the significance values were not extremely high. Substrates screened using only the single substrate screening/bacterial elution method (nacre and hydroxyapatite) did not exhibit high significance patterns due to the screening method, leaving the question of potential consensus still unanswered.

Polycrystalline materials, such as barium titanate, lead titanate, powdered lithium niobate and hydroxyapatite provided a plethora of binding surfaces and crevices, thus limiting the statistical significance of the patterns found. Calcite patterns were high in hydroxyl group amino acids. However, the patterns found for calcite were very short and generic in nature. It is not clear why calcite patterns were not easily elucidated, although it may be due to the fact that the best binders remained bound to the substrate even after elution or that the calcite (104) surface does not exhibit features conducive to binding, such as corrugation or charge separation of the ions.

The screenings of geological aragonite did show good relative homology to natural nacre-associated proteins. Since these libraries only represent a very small fraction ($\leq 0.0000015\%$) of all of the possible combinations of amino acids, exact homology is extremely difficult to achieve. Yet, when the known biomineral-associated proteins were screened for patterns found in the geological aragonite sequence population, many similarities were found. The main motifs of interest are the L, P, W, [RKH]; the [RKH]₃; and the multiple proline patterns. The first two are found mainly in the (110) aragonite population, while the multiple proline sequences are found in the (001) aragonite population. Nevertheless, these three motifs show good putative homology to the known nacre proteins Lustrin A, Mucoperlin, and Nacrein.

7.2: STRUCTURE DETERMINATION

The statistically significant A20 peptide was found to be a flexible peptide by both molecular dynamics modeling and by experimental NMR. Molecular

dynamics simulations show that on very small time scales, there are three high probability conformer states in solution. However, there is also probability between these three states, which would enable the peptide to interconvert easily between the conformations. Experimentally, though, these states are not populated at long enough time scales for detection using NMR. Therefore, the peptide is assumed to be flexible in nature as a synthetic 12-mer peptide. Nevertheless, when the peptide is fused to the phage coat proteins (pIII and pVIII), more stability may be bestowed upon the structure, as more interactions with the anchoring α -helical coat protein are possible.

The charged residues of the A20 peptide are most likely responsible for binding and in most of the molecular dynamics models are located on the same side of the molecule. Even in the more extended conformations, the K, E, and R residues form a triangular arrangement, which may indeed match up to the surface ion arrangement. Further modeling experiments may resolve this remaining question.

The apparent flexibility of these peptides leads one to assume that these peptides may function best as mineral growth inhibitor peptides. Nucleating proteins tend to be more rigid. By fusing these peptides to the coat proteins, there may be a higher degree of stability, however, this cannot be easily proven. With the apparent success of the aragonite nucleation, a higher degree of stability seems likely. Future experiments for this project could include the rational design of more stable peptides using the data discovered in these screenings.

7.3: PROTEIN ENGINEERING

Protein engineering of the phage major coat protein, the pVIII, is advantageous for two main reasons. First, the peptide of interest can be expressed at 5-10 times the level it is when fused to the minor coat protein, pIII, increasing the potential of crystal growth and the probability of finding such peptide-associated crystals. Second, the long-range order of the peptide expression along the phage coat may be necessary for peptide-mediated crystal growth.

Here, the peptide fusions were successfully engineered into the pMoPac33 vector. However, the peptide expression was not detectable using mass spectrometry or protein gel electrophoresis. Therefore, the expression level is assumed to be very low (~1%). Control experiments on a shorter peptide pVIII fusion were detectable and did express correctly when grown simultaneously under the exact same conditions. The success of the pVIII construct crystal growth experiments gives further evidence of effective pVIII fusion expression.

7.4: PEPTIDE MEDIATED CRYSTAL GROWTH

Using the pVIII engineered phage constructs, it was indeed possible to grow all three phases of calcium carbonate. Calcite is formed thermodynamically and is usually present in nucleation experiments. Vaterite, a meta-stable form of calcium carbonate, is also present. Vaterite is typically formed in nucleation experiments as a precursor to the more stable calcite according to the Ostwald step rule, which states that the most soluble phase will form first in a supersaturated solution. In these experiments, it may be possible that the vaterite crystals are stabilized by the phage proteins since these crystals were aged in a

vacuum desiccator overnight. It is clear that the aragonite crystals found in the TEM samples are most likely due to stabilization by phage proteins, since aragonite does not nucleate under ambient conditions without the addition of magnesium, which was absent in these experiments. Both the A20-pVIII and the A1-pVIII samples showed evidence of aragonite crystal growth, while the control samples gave evidence of calcite and possibly vaterite, but no aragonite. Furthermore, all other peptide-mediated crystal growth experiments have utilized acidic polypeptides. Here, we have shown that predominately basic peptides can also serve as a template for aragonite growth.

Peptide mediated crystal growth is not limited to calcium carbonate. Hybrid inorganic-organic-inorganic materials were grown in a sandwich fashion in order to make a prototype self-assembled waveguide amplifier. Doubly engineered phage were constructed to express both the A1 (aragonite, also similar to lithium niobate) peptide and the A7 (ZnS semiconductor) peptide as pVIII fusions. The A1 peptide was shown to preferentially bind to the lithium niobate +Z surface, with minimal non-specific binding. Furthermore, SEM images did in fact show phage-shaped ZnS crystal growth on the lithium niobate surface. This growth system has many potential uses in hybrid electronic, optical and biomedical materials development.

Appendix A: Sequences

Clone Name	Round	Sequence	Index	Total
EEG-A1	2	FKMPKEHERHAH	1	1
EEG-A2	2	TDTPWSAWVTYV	1	1
EEG-A3	2	QNEEWTPLDHAL	1	1
EEG-A4	2	TDTGSSTPSRES	1	1
EEG-A5	2	RISMTTTTKPSM	1	1
EEG-A6	3	VVSPVWRAPASP	1	1
EEG-A7	3	AWKHAHPPPTTS	1	1
EEG-A8	3	TARTSHAHWTSK	1	1
EEG-A9	3	SETMIFSAGKGA	1	1
EEG-A10	3	QHNGHLLTPMEQ	1	1
EEG-A11	3	SVTFNKAQQQGF	1	1
EEG-A12	3	GIKMWVPPRPY	1	1
EEG-A13	3	YYPQWSDASWIL	1	1
EEG-A15	3	KSNWHQIESPYL	1	1
EEG-A16	5	LPWQKTPRLHHD	1	1
EEG-A17	5	QAYWKTAPIQWH	1	1
EEG-A18	5	SWKWAXXKPMPV	1	1
EEG-A19	5	HDNSTLSPRSGP	1	1
EEG-A20	5	LPKWQERQMLSA	1	4
EEG-A21	5	LPPWKHKTSGVA	1	2
EEG-A22	5	YPNWKHSTHLLM	1	1
EEG-A23	5	LPKWSSAWPTPP	1	1
EEG-A24	5	LPKWQERQMLSA	2	4
EEG-A25	5	IPHAKFKALPQP	1	1
EEG-A26	5	LPKWQERQMLSA	3	4
EEG-A27	5	LPPWKHKTSGVA	2	2
EEG-A28	5	LPKWTPSSMLSD	1	1
EEG-A29	5	LPKWQERQMLSA	4	4
EEG-A30	5	LPTKYWKLQSTH	1	1

Table A.1: PhD12 sequences from geological aragonite screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-A31	2	APWNTNQTHLLM	1	1
EEG-A32	2	LPPWSKGNVAQP	1	1
EEG-A33	2	APWWASPIASGP	1	1
EEG-A34	2	IPRHHHTVLSLS	1	1
EEG-A35	2	HAIKWHTVTYPS	1	1
EEG-A36	2	AFSKHPKNSFVT	1	1
EEG-A37	2	KPHVKPYSWGLS	1	1
EEG-A38	2	KPWMHFYSDLLS	1	1
EEG-A39	2	VPRAHCTTKWVC	1	1
EEG-A40	2	LPRHWLPNNPTH	1	1
EEG-A41	2	WTPKHATRLQNL	1	1
EEG-A42	2	APSKHADRHQIH	1	1
EEG-A43	2	MPWHWKQMPSP	1	1
EEG-A44	2	QNEVPRVENGR	1	1
EEG-A45	2	HYDFHLPPSLNP	1	1

Table A.2: PhD12 sequences from geological aragonite screened at pH 7.5 and eluted with acid; screening 2.

Clone Name	Round	Sequence	Index	Total
EEG-A46	2	AREHVQLHISYA	1	1
EEG-A48	2	TYHYEDTGITPT	1	1
EEG-A49	2	STRDLTMPLVLL	1	1
EEG-A50	2	IVPAIPNRVNWL	1	1
EEG-A51	2	FEKREPPXMGMA	1	1
EEG-A52	2	STKIEQTIQHDV	1	1
EEG-A53	2	AQQSSKTTTPLT	1	1
EEG-A54	2	HMCATVCPLKYH	1	1
EEG-A56	3	LPHNKFPFSPNR	1	1
EEG-A57	3	NHNQHQQFKIHI	1	1
EEG-A58	3	GPWKKDPTWQKI	1	1
EEG-A59	3	GPFWRWVDRGHM	1	1
EEG-A60	3	APKWQKTGVPLP	1	1
EEG-A61	3	SHKSAHHPRFLQ	1	1
EEG-A62	3	SPPHWKRPPSHW	1	1
EEG-A63	3	TPPWAKASRWTI	1	1
EEG-A64	3	TPNWKHWRLGLG	1	1
EEG-A65	3	IVDHKHPFKSRK	1	1
EEG-A66	5	LPPWKHKTSOVA	1	10
EEG-A67	5	LPKWQERQMLSA	1	1
EEG-A68	5	APKRKLPVTATL	1	1
EEG-A69	5	LPPWKHKTSOVA	2	10
EEG-A70	5	GEKAKHHPFLXX	1	1
EEG-A71	5	LPRWANTSVLLR	1	1
EEG-A72	5	HHKLAHNTPWRF	1	1
EEG-A74	5	TPKWDRPLRVQL	1	1
EEG-A75	5	AKFDHKPHPRFN	1	1
EEG-A78	5	HMPHNHPKKSQW	1	1
EEG-A80	5	FPTWKASHPRQS	1	1
EEG-A106	7	LPPWKHKTSOVA	3	10
EEG-A107	7	LPPWKHKTSOVA	4	10
EEG-A108	7	VPITKPRHTTIV	1	1
EEG-A109	7	TLRRHKTRHWTP	1	1
EEG-A110	7	QPWHKPPRVQAL	1	1
EEG-A111	7	LPKHWQFHHRGP	1	1
EEG-A112	7	LPPWKHKTSOVA	5	10
EEG-A113	7	GWKXKHTPPPWP	1	1

Table A.3: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-A114	7	FHKPNHHRPAFA	1	1
EEG-A115	7	LPPWKHKTSOVA	6	10
EEG-A116	7	APGHMKSPRWSG	1	1
EEG-A117	7	LPPWKHKTSOVA	7	10
EEG-A118	7	LPPWKHKTSOVA	8	10
EEG-A119	7	LPPWKHKTSOVA	9	10
EEG-A120	7	LPPWKHKTSOVA	10	10

Table A.3 Continued: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-A81	3	KWNPKPPAWPHL	1	1
EEG-A82	3	MHKSDHASKHQH	1	1
EEG-A83	3	KPWKHHMSETLV	1	1
EEG-A85	3	LPSKKPLRVFPF	1	1
EEG-A87	3	APSTHWKMVPRS	1	1
EEG-A88	3	VVGPIPQTLFPF	1	1
EEG-A89	3	GWKHQLTTSPVH	1	1
EEG-A90	3	GWKHGPTNPKHA	1	1
EEG-A91	5	SHKIHWWQLESL	1	1
EEG-A92	5	HWKHTASPKVPF	1	1
EEG-A93	5	HWKDRSIFAPVS	1	1
EEG-A94	5	HVKWSAPNINLR	1	1
EEG-A95	5	HWKTKLPWPPEV	1	1
EEG-A96	5	LPPWWPHSKERP	1	1
EEG-A97	5	VMAKPWHKHPQV	1	1
EEG-A98	5	GHKLRHFPVLHN	1	1
EEG-A99	5	APKWKPDAPPRL	1	1
EEG-A100	5	LPKWYSHPVDTF	1	1
EEG-A101	5	DHKSSHWRYLPR	1	1
EEG-A102	5	FVYAPGPPPLNP	1	1
EEG-A103	5	TPWHKSKPHLPP	1	1
EEG-A104	5	HHKPEHGTRGTY	1	1
EEG-A105	5	WPNWKYSATTDX	1	1

Table A.4: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with base.

Clone Name	Round	Sequence	Index	Total
EEG-ABE-1	2	SNPPTPPRLPPP	1	1
EEG-ABE-2	2	DYHYAPPQPQRY	1	1
EEG-ABE-3	2	VQNTVKHYRLTP	1	1
EEG-ABE-4	2	IPTPFPLVTHQL	1	1
EEG-ABE-6	2	SHHRPNSWSLLT	1	1
EEG-ABE-7	2	DWNPWWRSGRSL	1	1
EEG-ABE-8	2	KVDVGRFKQPYA	1	1
EEG-ABE-9	2	QFKDSQPGDPYR	1	1
EEG-ABE-10	2	ATTPWPAPPPVG	1	1
EEG-ABE-12	3	YTKDTSLARKGT	1	1
EEG-ABE-13	3	FVARTANLLTWD	1	1
EEG-ABE-14	3	AIQLPGRIPMQE	1	1
EEG-ABE-15	3	INTQKLAVRNIW	1	1
EEG-ABE-16	3	LQTQITTGAPRS	1	1
EEG-ABE-17	3	RPSLPIYPHQLS	1	1
EEG-ABE-18	3	TTRQTLPDFAWW	1	1
EEG-ABE-19	3	NVQQTISFPYPS	1	1
EEG-ABE-20	3	LELRTTQTDAGA	1	1
EEG-ABE-21	3	GLVTPKPPLSPR	1	1
EEG-ABE-22	3	ALESLQSTYLTH	1	1
EEG-ABE-23	3	KAQLRHPTFNLL	1	1
EEG-ABE-24	3	QPMHNDPLASLN	1	1
EEG-ABE-25	3	TMSFLQPSATLP	1	1

Table A.5: PhD12 sequences from geological aragonite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-A7-1	3	AQTAPPF	1	1
EEG-A7-2	3	SNSARTP	1	1
EEG-A7-3	3	HSGPYHR	1	1
EEG-A7-4	3	SPACIRC	1	1
EEG-A7-5	3	APKAPVP	1	1
EEG-A7-6	3	ALPLTLH	1	1
EEG-A7-7	3	TPPLPPV	1	1
EEG-A7-8	3	WDGLPGN	1	1
EEG-A7-9	3	NLPGHTQ	1	1
EEG-A7-10	3	KLHPHPT	1	1
EEG-A7-11	5	PNLAVQH	1	1
EEG-A7-12	5	FPKWHRW	1	1
EEG-A7-13	5	KEEKNHK	1	1
EEG-A7-14	5	PRQGSRP	1	1
EEG-A7-15	5	HHKLHHL	1	1
EEG-A7-16	5	QTLNDPP	1	1
EEG-A7-17	5	AQTEIYT	1	1
EEG-A7-18	5	FNLPSPP	1	1
EEG-A7-19	5	APKWRLP	1	1
EEG-A7-20	5	FHKDSHR	1	1
EEG-A7-21	5	GHLARIR	1	1
EEG-A7-22	5	PRRSGHQ	1	1
EEG-A7-23	5	TPVPLSE	1	1
EEG-A7-24	5	HSKIKHA	1	1
EEG-A7-25	5	DHRDISR	1	1

Table A.6: PhD7 sequences from geological aragonite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-A7C-1	3	CKEQHPRKC	1	1
EEG-A7C-2	3	CSHHHHKRC	1	1
EEG-A7C-3	3	CPNHHKWKC	1	1
EEG-A7C-4	3	CAKHPWKSC	1	1
EEG-A7C-5	3	CKSKMQPHC	1	1
EEG-A7C-6	3	CNNHKHKMC	1	1
EEG-A7C-7	3	CKHTV NKSC	1	1
EEG-A7C-8	3	CGPQKPQYC	1	1
EEG-A7C-9	3	CRLTNGTQC	1	1
EEG-A7C-10	3	CKSPNHRHC	1	1
EEG-A7C-11	5	CRHTHHKHC	1	1
EEG-A7C-12	5	CKIKYHKDC	1	1
EEG-A7C-13	5	CHKNNIKHC	1	1
EEG-A7C-14	5	CRPNKLSNC	1	1
EEG-A7C-15	5	CGKHFXHKC	1	1
EEG-A7C-16	5	CHPHKNMKC	1	1
EEG-A7C-17	5	CHHHKPHKC	1	1
EEG-A7C-18	5	CKWHSPNKC	1	1
EEG-A7C-19	5	CGKHTWQRC	1	1
EEG-A7C-20	5	CKHKHPQNC	1	1
EEG-A7C-21	5	CKLPVKKPC	1	1
EEG-A7C-22	5	CHKKDPKVC	1	1
EEG-A7C-23	5	CKLNHQSKC	1	1
EEG-A7C-24	5	CNKHHTWKC	1	1
EEG-A7C-25	5	CPHWKNHNC	1	1

Table A.7: PhD7C sequences from geological aragonite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-Aside-1	3	LPPWKHKTSGVA	1	16
EEG-Aside-2	3	LPPWKHKTSGVA	2	16
EEG-Aside-3	3	LPPWKHKTSGVA	3	16
EEG-Aside-4	3	LPPWKHKTSGVA	4	16
EEG-Aside-5	3	LPPWKHKTSGVA	5	16
EEG-Aside-6	3	LPPWKHKTSGVA	6	16
EEG-Aside-7	3	LPPWKHKTSGVA	7	16
EEG-Aside-8	3	LPPWKHKTSGVA	8	16
EEG-Aside-9	3	LPPWKHKTSGVA	9	16
EEG-Aside-10	3	LPPWKHKTSGVA	10	16
EEG-Aside-11	3	LPPWKHKTSGVA	11	16
EEG-Aside-12	3	LPPWKHKTSGVA	12	16
EEG-Aside-13	3	LPPWKHKTSGVA	13	16
EEG-Aside-14	3	MSSTSMPTINRA	1	1
EEG-Aside-15	3	SISALPTSSSTT	1	1
EEG-Aside-16	3	LPPWKHKTSGVA	14	16
EEG-Aside-17	3	LPPWKHKTSGVA	15	16
EEG-Aside-18	3	LPPWKHKTSGVA	16	16
EEG-Aside-20	3	GEHQVPETWLPL	1	1

Table A.8: PhD12 sequences from the 110 face of geological aragonite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-110-A-2-1	2	MHKPSHDRPLIR	1	1
EEG-110-A-2-2	2	LNHKTASAQTTE	1	1
EEG-110-A-2-3	2	GQYWLLGSSIAP	1	1
EEG-110-A-2-4	2	TASWKSTVNKWN	1	1
EEG-110-A-2-6	2	LPQWKHSSLTLH	1	1
EEG-110-A-2-7	2	HPTWKTPIQNVI	1	2
EEG-110-A-2-8	2	TDTFIPLIAPTI	1	1
EEG-110-A-2-9	2	ESMYKPTNTSLQ	1	1
EEG-110-A-2-10	2	TIAASKHSRMLS	1	1
EEG-110-A-2-11	2	AHQPWVLRSDPA	1	1
EEG-110-A-2-12	2	LPWHRQPLFGLT	1	1
EEG-110-A-2-13	2	NPNAPSMATLFS	1	1
EEG-110-A-2-14	2	HPTWKTPIQNVI	2	2
EEG-110-A-2-16	2	DAMPYIPQRPEY	1	1
EEG-110-A-2-17	2	LPPWKHKHTSGVA	1	14
EEG-110-A-2-18	2	REEDHLHKSDL	1	1
EEG-110-A-2-19	2	KWVDMKMWGSSP	1	1
EEG-110-A-2-20	2	AAPLATNVSPKV	1	1
EEG-110-A-3-1	3	LPKHSWIKYGWH	1	1
EEG-110-A-3-10	3	LPPWKHKHTSGVA	2	14
EEG-110-A-3-2	3	LPPWKHKHTSGVA	3	14
EEG-110-A-3-3	3	LPPWKHKHTSGVA	4	14
EEG-110-A-3-4	3	LPKTDPKVKNQM	1	1
EEG-110-A-3-5	3	LPPWKHKHTSGVA	5	14
EEG-110-A-3-6	3	LPPWKHKHTSGVA	6	14
EEG-110-A-3-7	3	LPPWKHKHTSGVA	7	14
EEG-110-A-3-8	3	LPPWKHKHTSGVA	8	14
EEG-110-A-3-9	3	LPPWKHKHTSGVA	9	14
EEG-110-A-5-1	5	LPPWKHKHTSGVA	10	14
EEG-110-A-5-2	5	LPPWKHKHTSGVA	11	14
EEG-110-A-5-3	5	LPPWKHKHTSGVA	12	14
EEG-110-A-5-4	5	LPPWKHKHTSGVA	13	14
EEG-110-A-5-5	5	LPPWKHKHTSGVA	14	14

Table A.9: PhD12 sequences from the 110 face of geological aragonite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-Atop-1	3	LSGNKNSTIALP	1	1
EEG-Atop-2	3	LPPWKHKTSOVA	1	6
EEG-Atop-3	3	YNDTINFGHIRT	1	1
EEG-Atop-4	3	TSQIPHVAGFYT	1	1
EEG-Atop-5	3	VSSLPRATHPPC	1	1
EEG-Atop-6	3	LPPWKHKTSOVA	2	6
EEG-Atop-8	3	LPPWKHKTSOVA	3	6
EEG-Atop-9	3	DQHQEQLQPSLT	1	1
EEG-Atop-10	3	FIPGAPTLVQTG	1	1
EEG-Atop-11	3	DLRQPALTLHPG	1	1
EEG-Atop-13	3	YTKASNFTVQPM	1	1
EEG-Atop-14	3	QWLQPPSQGRSY	1	1
EEG-Atop-15	3	LPPWKHKTSOVA	4	6
EEG-Atop-16	3	LPPWKHKTSOVA	5	6
EEG-Atop-17	3	INSSLNTHLHAV	1	1
EEG-Atop-18	3	LPPWKHKTSOVA	6	6
EEG-Atop-19	3	TTHYHKSPRMHT	1	1
EEG-Atop-20	3	MFHQVKMSYTIG	1	1

Table A.10: PhD12 sequences from the 001 face of geological aragonite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-001-A-2-1	2	IPPWLQPPLPQK	1	13
EEG-001-A-2-2	2	IPPWLQPPLPQK	2	13
EEG-001-A-2-3	2	IPPWLQPPLPQK	3	13
EEG-001-A-2-4	2	IPPWLQPPLPQK	4	13
EEG-001-A-3-1	3	GKWKLPPFPMTH	2	5
EEG-001-A-3-2	3	IPPWLQPPLPQK	5	13
EEG-001-A-3-3	3	IPPWLQPPLPQK	6	13
EEG-001-A-3-4	3	GKWKLPPFPMTH	3	5
EEG-001-A-3-5	3	IPPWLQPPLPQK	7	13
EEG-001-A-3-6	3	IPPWLQPPLPQK	8	13
EEG-001-A-3-7	3	IPPWLQPPLPQK	9	13
EEG-001-A-3-8	3	IPPWLQPPLPQK	10	13
EEG-001-A-3-9	3	IPPWLQPPLPQK	11	13
EEG-001-A-3-10	3	GKWKLPPFPMTH	1	5
EEG-001-A-5-1	5	IPPWLQPPLPQK	12	13
EEG-001-A-5-2	5	IPPWLQPPLPQK	13	13
EEG-001-A-5-4	5	GKWKLPPFPMTH	4	5
EEG-001-A-5-5	5	GKWKLPPFPMTH	5	5

Table A.12: PhD12 sequences from the 001 face of geological aragonite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-C1	2	APLTSVMIQTAT	1	1
EEG-C2	2	MPAVMSSAQVPR	1	1
EEG-C3	2	TTMLAPPPRSND	1	1
EEG-C4	2	VPHMSMTALSVS	1	1
EEG-C5	2	TFFGXAASRPQV	1	1
EEG-C6	3	IPSRPPSMQMRQ	1	1
EEG-C7	3	GHLHQTPPLWPR	1	1
EEG-C8	3	FNMQGLTTASQP	1	1
EEG-C10	3	VGPTLAHSMSTS	1	1
EEG-C14	3	VQNRQNIVHATW	1	1
EEG-C17	5	SQNHQHFKLLP	1	1
EEG-C18	5	HSTPQPTSFPDS	1	1
EEG-C19	5	VHTTAPTAPTAM	1	1
EEG-C21	5	SVSVGMKPSRP	1	1
EEG-C23	5	APNTYHIRPGLG	1	1
EEG-C25	5	TNPVAWSVSPYR	1	1
EEG-C26	5	KPYVHQYSQPAL	1	1
EEG-C27	5	HSAQQWDRASMS	1	1
EEG-C28	5	RVTNAFDPYPET	1	1
EEG-C29	5	AYHSQTHLFVPH	1	1
EEG-C30	5	QNTSRPDYFSL	1	1

Table A.13: PhD12 sequences from geological calcite screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-C36	2	TSAHNPFWSPTT	1	1
EEG-C37	2	EQLPSLMRKVAP	1	1
EEG-C38	2	SFPEPLSPLVRP	1	1
EEG-C40	2	SNKSDNLIPDAS	1	1
EEG-C42	3	MEPESRPALNGD	1	1
EEG-C44	3	NTPNITNARLLM	1	1
EEG-C45	3	QLATSDMKHFLP	1	1
EEG-C46	3	LEAGTPVYSAPV	1	1
EEG-C48	3	HGAQHQPSFLSP	1	1
EEG-C49	3	TPQISGSSAFKP	1	1
EEG-C50	3	VESTQAMSNSTS	1	1
EEG-C51	5	MNWNWMESGVVA	1	1
EEG-C52	5	HPISHQLQAPVR	1	1
EEG-C55	5	ISSMPSKGSCHAR	1	1
EEG-C56	5	TGGSQAPAPTLP	1	1
EEG-C62	5	AGPHYLYVGPGL	1	1
EEG-C64	5	SHFSPVLLSSRS	1	1
EEG-C65	5	LRVPHGMQQSMR	1	1
EEG-C91	7	DVPPSIAPLLRS	1	1
EEG-C92	7	CPMLPWGCLPLP	1	1
EEG-C94	7	QTMSPITPRLSW	1	1
EEG-C95	7	IRHMSPRTGAPY	1	1
EEG-C96	7	HHGHGSAXYFVV	1	1
EEG-C97	7	DYLNYPTRPAS	1	1
EEG-C98	7	TFHQKPESEATF	1	1
EEG-C99	7	LINNTDTVSLTT	1	1
EEG-C100	7	DSPFMLLHHKVK	1	1
EEG-C101	7	FPMQNDSTSSQP	1	1
EEG-C102	7	MNSGLSALDGKT	1	1
EEG-C103	7	NLQTLKYPSPPT	1	1
EEG-C104	7	FPILSDMPATRK	1	1
EEG-C105	7	LPSSMPTFTLAP	1	1

Table A.14: PhD12 sequences from geological calcite screened at pH 8.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-C66	3	SQTSYSDSPPNY	1	1
EEG-C67	3	NPTPPNPDAVRR	1	1
EEG-C68	3	ATSRWSQEMLPN	1	1
EEG-C69	3	FPKTTNYYVNPR	1	1
EEG-C70	3	SSSPHSQLPPPP	1	1
EEG-C71	3	EIVSAVALDMAR	1	1
EEG-C74	3	YLLDFSPALPNM	1	1
EEG-C75	3	YPYEPANNRAPT	1	1
EEG-C76	5	TTPNLNMHSITT	1	1
EEG-C77	5	ISKDSQETWSTP	1	1
EEG-C78	5	RLPYDDAASWYT	1	1
EEG-C79	5	LHKPQGGLAPAS	1	1
EEG-C80	5	RFVDFEPSSTTS	1	1
EEG-C81	5	YSNDHSAWDPLT	1	1
EEG-C82	5	SMPHLSK GALIS	1	1
EEG-C83	5	SPSLSRPHVHST	1	1
EEG-C84	5	YTSPTPPNRHIG	1	1
EEG-C85	5	STSNQAHPVPLK	1	1
EEG-C86	5	HYHPFKGSALVP	1	1
EEG-C87	5	LPQHIMSSSAKP	1	1
EEG-C88	5	SAAIAITYTPHN	1	1
EEG-C89	5	TPQPLSLLSEA	1	1
EEG-C90	5	SVSNHAGKLT DN	1	1

Table A.15: PhD12 sequences from geological calcite screened at pH 8.5 and eluted with base.

Clone Name	Round	Sequence	Index	Total
EEG-12N-1	3	SPTNFSPFSSIF	1	1
EEG-12N-2	3	SMFESHARLIGP	1	1
EEG-12N-3	3	DIQLSKTARMAL	1	1
EEG-12N-4	3	IADTRNNPTLPM	1	1

Table A.16: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 1.

Clone Name	Round	Sequence	Index	Total
EEG-12N2-1	3	APRDNFHTVNLF	1	4
EEG-12N2-2	3	APRDNFHTVNLF	2	4
EEG-12N2-3	3	APRDNFHTVNLF	3	4
EEG-12N2-4	3	APRDNFHTVNLF	4	4

Table A.17: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria.; screening 2

Clone Name	Round	Sequence	Index	Total
EEG-Nacre-1	3	TLQLSHSSTALL	1	1
EEG-Nacre-3	3	SRNWDVSLPDVR	1	1
EEG-Nacre-4	3	SGHAYNFPWYSA	1	2
EEG-Nacre-6	3	SPSGPLVEPIIP	1	1
EEG-Nacre-9	3	WPLQKMATTHFM	1	1
EEG-Nacre-10	3	TKTMHVPAFEPR	1	1
EEG-Nacre-11	3	SQPAGNPRYSLF	1	1
EEG-Nacre-12	3	SGHAYNFPWYSA	2	2
EEG-Nacre-13	3	SFASTTCHPSGR	1	1
EEG-Nacre-14	3	YPTHGTGFGDQP	1	1
EEG-Nacre-15	3	MYPHIPTVRLFT	1	1
EEG-Nacre-16	3	LLADTTHHRPWT	1	1
EEG-Nacre-19	3	HTHQTMLKLM DG	1	1
EEG-Nacre-20	3	HFTKYAPKDASG	1	1

Table A.18: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 3.

Clone Name	Round	Sequence	Index	Total
EEG-N12-4-1	4	HRPMTAPSPLSN	1	1
EEG-N12-4-3	4	HSLSTKLHTQTI	1	1
EEG-N12-4-5	4	YNSSIQSTLPRR	1	1
EEG-N12-5-1	5	AVAYETVAHTAP	1	1
EEG-N12-5-2	5	LQQNISLPQAPI	1	1
EEG-N12-5-3	5	EPPVNWNPWLWT	1	1
EEG-N12-5-4	5	QNLDQASSPPNI	1	1
EEG-N12-5-8	5	YSVPYKMOVITPI	1	1
EEG-N12-5-9	5	SVIVPQALEPHR	1	1
EEG-N12-5-10	5	WHPLTHARMAAL	1	1
EEG-N12-5-12	5	STTENPAKRANM	1	1
EEG-N12-5-14	5	HLPKPQMHRTA	1	1
EEG-N12-5-16	5	SLKMPHWPHELLP	1	1
EEG-N12-5-17	5	SMYTFSSAFELA	1	1
EEG-N12-5-18	5	QQMHLMSYAPGP	1	1
EEG-N12-5-20	5	KSLSRHDHIIHHH	1	1

Table A.19: PhD12 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 4.

Clone Name	Round	Sequence	Index	Total
EEG-N7-4-1	4	SPYVPGT	1	1
EEG-N7-4-2	4	VTNTAIY	1	1
EEG-N7-4-3	4	VDPLHKS	1	1
EEG-N7-4-4	4	NQDVPLF	1	1
EEG-N7-4-5	4	TLSYATS	1	1
EEG-N7-5-1	5	SIILGSS	1	1
EEG-N7-5-2	5	HPERATL	1	1
EEG-N7-5-3	5	STFTHPR	1	1
EEG-N7-5-4	5	QGSLSNL	1	1
EEG-N7-5-6	5	NLQEFLF	1	4
EEG-N7-5-8	5	NLQEFLF	2	4
EEG-N7-5-10	5	DLMSYMG	1	1
EEG-N7-5-11	5	SATVHHP	1	1
EEG-N7-5-12	5	NLQEFLF	3	4
EEG-N7-5-14	5	GPGQAPT	1	1
EEG-N7-5-15	5	WPADVRS	1	1
EEG-N7-5-17	5	NLQEFLF	4	4
EEG-N7-5-19	5	KSLDWEP	1	1

Table A.20: PhD7 sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-7CN-1	3	CSAKYVKNC	1	1
EEG-7CN-2	3	CHTPWSRIC	1	1

Table A.21: PhD7C sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 1.

Clone Name	Round	Sequence	Index	Total
EEG-N7C-4-1	4	CSHLHSPLC	1	1
EEG-N7C-4-3	4	CPKSSQSFC	1	1
EEG-N7C-5-1	5	CMRNDVLQC	1	1
EEG-N7C-5-2	5	CSSQHFPQC	1	1
EEG-N7C-5-3	5	CISNHISSC	1	1
EEG-N7C-5-4	5	CNLSPKPQC	1	2
EEG-N7C-5-5	5	CQSSQSPRC	1	1
EEG-N7C-5-7	5	CYTPHWNC	1	1
EEG-N7C-5-9	5	CTLAQWRAC	1	1
EEG-N7C-5-10	5	CTPSSLRHC	1	1
EEG-N7C-5-11	5	CNLSPKPQC	2	2
EEG-N7C-5-13	5	CAGIQPRTC	1	1
EEG-N7C-5-14	5	CTLTLHNQC	1	1
EEG-N7C-5-16	5	CNPIMQRTC	1	1
EEG-N7C-5-17	5	CKEVHTAC	1	1
EEG-N7C-5-19	5	CHKPTSLTC	1	1
EEG-N7C-5-20	5	CSLENVPYC	1	1

Table A.22: PhD7C sequences from the 001 face of nacreous aragonite screened at pH 8.5 and eluted with bacteria; screening 2.

Clone Name	Round	Sequence	Index	Total
EEG-HA-1	3	ERYDNVWQLPLG	1	1
EEG-HA-2	3	SWFPLMPVSTAL	1	1
EEG-HA-3	3	SSLNCAQNCMDR	1	1
EEG-HA-4	3	VSNPQLLMSSAP	1	1
EEG-HA-5	3	GKPQASTQPIRS	1	1
EEG-HA-6	3	ANTFKAKSPSHP	1	1
EEG-HA-7	3	QITTSPMRYPYI	1	1
EEG-HA-8	3	NSWMTPTPHSDR	1	1
EEG-HA-9	3	WSMSHTWTSSLM	1	1
EEG-HA-10	3	QPPDFPPDSPLL	1	1
EEG-HA-11	3	NMAHVNILQPPE	1	1
EEG-HA-12	3	QVHDRFSAVYSH	1	1
EEG-HA-13	3	SSPLPQMRSQPV	1	1
EEG-HA-14	3	FSDGQMHRISRW	1	1
EEG-HA-15	3	QVLGKHPKEIFM	1	1
EEG-HA-16	3	TQSFGCRACHSP	1	1
EEG-HA-17	3	HVKVQMPTFPSL	1	1
EEG-HA-18	3	DSKPMQNPIQSW	1	1
EEG-HA-19	3	AQMNLMLGNLT	1	1

Table A.23: PhD12 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria; screening 1.

Clone Name	Round	Sequence	Index	Total
EEG-H12-4-1	4	HWSSSFQTGWQN	1	1
EEG-H12-4-2	4	MAEDRPTGPSWA	1	1
EEG-H12-4-4	4	NTVAKMLQHSLP	1	1
EEG-H12-4-5	4	IQPHNSKRAFEL	1	1
EEG-H12-4-6	4	XTPPNPSLPATS	1	1
EEG-H12-4-7	4	TLSDVNVTRQAR	1	1
EEG-H12-4-8	4	SPLHLLRPGLQT	1	1
EEG-H12-4-9	4	DSQRGPFMVHRL	1	1
EEG-H12-4-10	4	QMLHTKPRMPNF	1	1
EEG-H12-4-12	4	TPSAAPNNWRYL	1	1
EEG-H12-4-13	4	ISSTGLSLHGGK	1	1
EEG-H12-4-14	4	SHYSAFPLPSPH	1	1
EEG-H12-4-15	4	TTKQTLVALTRV	1	1
EEG-H12-4-17	4	TWAGTTVTRWNL	1	1
EEG-H12-4-18	4	SIPRMASNFNQL	1	1
EEG-H12-4-19	4	SNSLHKAPQPTP	1	1
EEG-H12-4-20	4	TATLLQRMQAFS	1	1
EEG-H12-5-2	5	SLTPLTLKPTMS	1	1
EEG-H12-5-3	5	IPSSHALGSKPI	1	1
EEG-H12-5-4	5	HHIRPSQTNQPP	1	1
EEG-H12-5-5	5	WSQTLTSDGLW	1	1
EEG-H12-5-6	5	NWPVSERYPPPR	1	1
EEG-H12-5-7	5	QSEPQQKPGSLL	1	1
EEG-H12-5-8	5	QSTSPNIQASRT	1	1
EEG-H12-5-9	5	SLPQGSWSWQKR	1	1
EEG-H12-5-10	5	TWTPAMAPSAGP	1	1
EEG-H12-5-11	5	VKYQPPNHLPPR	1	1
EEG-H12-5-12	5	NSYDFQKPRTPG	1	1
EEG-H12-5-13	5	ADNAAAIGPNRY	1	1
EEG-H12-5-14	5	XXXXXXFQSLGY	1	1
EEG-H12-5-17	5	FPGYVRHISLD	1	1
EEG-H12-5-18	5	SEHKPPTPPFPA	1	1
EEG-H12-5-19	5	HPASKPDDARIS	1	1
EEG-H12-5-20	5	SLPNGNKPKLYV	1	1

Table A.24: PhD12 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria; screening 2.

Clone Name	Round	Sequence	Index	Total
EEG-H7-4-1	4	GAPPXAH	1	1
EEG-H7-4-2	4	YSPPSLE	1	1
EEG-H7-4-3	4	QTALQIV	1	1
EEG-H7-4-4	4	SFLTDVA	1	1
EEG-H7-4-5	4	VEIEALL	1	1
EEG-H7-4-6	4	TEPLPSG	1	1
EEG-H7-4-7	4	KLSIEHP	1	1
EEG-H7-4-8	4	HSLPILW	1	1
EEG-H7-4-9	4	WSRPAST	1	1
EEG-H7-4-10	4	GMANTSV	1	1
EEG-H7-4-12	4	VRHNAV	1	1
EEG-H7-4-13	4	AYPTQAA	1	1
EEG-H7-4-14	4	QVLPAQQ	1	1
EEG-H7-4-15	4	STMAFFW	1	1
EEG-H7-4-16	4	LHTPPSP	1	1
EEG-H7-4-17	4	TTVNMLP	1	1
EEG-H7-4-18	4	ALGTPQN	1	1
EEG-H7-4-19	4	QTYTVMT	1	1
EEG-H7-4-20	4	FTNTFAK	1	1
EEG-H7-5-1	5	TPSLPVP	1	1
EEG-H7-5-2	5	GLALXPP	1	1
EEG-H7-5-3	5	LTSLAPP	1	1
EEG-H7-5-4	5	KDTWMDT	1	1
EEG-H7-5-5	5	HPHTLQK	1	1
EEG-H7-5-6	5	AAAPWIR	1	1
EEG-H7-5-7	5	LPTSYQL	1	1
EEG-H7-5-8	5	FGSSLTE	1	1
EEG-H7-5-9	5	ATFHLVL	1	1
EEG-H7-5-10	5	TVIQNRL	1	1
EEG-H7-5-11	5	QTLPNTV	1	1

Table A.25: PhD7 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-H7-5-12	5	VPPYFNS	1	1
EEG-H7-5-13	5	NPTPTQA	1	1
EEG-H7-5-14	5	TQTLLVL	1	1
EEG-H7-5-16	5	HQLVLLG	1	1
EEG-H7-5-17	5	AIKAPLH	1	1
EEG-H7-5-18	5	IHPHYMN	1	1
EEG-H7-5-19	5	NKVEPDA	1	1
EEG-H7-5-20	5	SHHRMP	1	1

Table A.25 Continued: PhD7 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-H7C-4-1	4	CSWHSQNMCM	1	1
EEG-H7C-4-2	4	CMPSQRNVC	1	1
EEG-H7C-4-3	4	CEQLGQRAC	1	1
EEG-H7C-4-4	4	CTHRTPGFC	1	1
EEG-H7C-4-6	4	CSQLRTGFC	1	1
EEG-H7C-4-7	4	CDRTNYAAC	1	1
EEG-H7C-4-8	4	CSTVSPGSC	1	1
EEG-H7C-4-9	4	CHTQAFRIC	1	1
EEG-H7C-4-10	4	CFVMSSIPC	1	1
EEG-H7C-4-11	4	CHAVQAAYC	1	1
EEG-H7C-4-12	4	CGLSDSWRC	1	1
EEG-H7C-4-13	4	CNHTPGHIC	1	1
EEG-H7C-4-14	4	CHLRSGNSC	1	1
EEG-H7C-4-15	4	CHARDDYLC	1	1
EEG-H7C-4-16	4	CDSPRRFTC	1	1
EEG-H7C-4-17	4	CIDQHRVPC	1	1
EEG-H7C-4-18	4	CPGPSTNAC	1	1
EEG-H7C-4-19	4	CTTHTHPPC	1	1
EEG-H7C-4-20	4	CLHKPPMAC	1	1
EEG-H7C-5-1	5	CNHVDTHQC	1	1
EEG-H7C-5-2	5	CLQNHQTRC	1	1
EEG-H7C-5-3	5	CPTRSSYIC	1	1
EEG-H7C-5-4	5	CNFGPWQVC	1	1
EEG-H7C-5-6	5	CSTSHPRQC	1	1
EEG-H7C-5-7	5	CLFSKYNSC	1	1
EEG-H7C-5-9	5	CPYNQVLMC	1	1
EEG-H7C-5-11	5	CERLSGKNC	1	1
EEG-H7C-5-12	5	XXXXSNYAC	1	1
EEG-H7C-5-13	5	CPNPKQTTC	1	1
EEG-H7C-5-15	5	CLTVATITC	1	1
EEG-H7C-5-16	5	CNTDTAPMC	1	1
EEG-H7C-5-19	5	CNTSNSKLC	1	1

Table A.26: PhD7 sequences from hydroxylapatite screened at pH 8.5 and eluted with bacteria.

Clone Name	Round	Sequence	Index	Total
EEG-B7C-1	2	CPNLPQPKC	1	1
EEG-B7C-2	2	CMASDPRQC	1	1
EEG-B7C-3	2	CPTSAHPSC	3	3
EEG-B7C-4	2	CPAHPFNSC	1	1
EEG-B7C-5	2	CTPLHLQMC	2	2
EEG-B7C-6	2	CPRTDLPRC	1	1
EEG-B7C-7	2	CGMYQPYTC	1	1
EEG-B7C-8	2	CYASLMYKC	1	1
EEG-B7C-9	2	CGASSQKAC	1	1
EEG-B7C-10	2	CWKTETPLC	1	1
EEG-B7C-11	3	CTPLHLQMC	1	2
EEG-B7C-12	3	CHDTKTRAC	1	1
EEG-B7C-13	3	CPTSPDHTC	1	1
EEG-B7C-15	3	CTPSHRAQC	1	1
EEG-B7C-16	3	CPTSAHPSC	1	3
EEG-B7C-18	3	CXHTLSKNC	1	1
EEG-B7C-19	3	CPTSAHPSC	2	3
EEG-B7C-20	3	CPSVYRDAC	1	1
EEG-B7C-23	3	CNASHFPRC	1	1
EEG-B7C-25	3	CTNLELKSC	1	1

Table A.27: PhD7C sequences from barium titanate screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-P7C-1	2	CTPANHTRC	1	1
EEG-P7C-2	2	CENLNAKTC	1	1
EEG-P7C-3	2	CLKTDHNNC	1	1
EEG-P7C-4	2	CSNPTHHSC	1	1
EEG-P7C-6	2	CTQTSAYLC	1	1
EEG-P7C-7	2	CPHLLHTTC	1	1
EEG-P7C-8	2	CPGQQTGQC	1	1
EEG-P7C-10	2	CGAKPLMLC	1	1
EEG-P7C-11	3	CNEWHPRTC	1	1
EEG-P7C-12	3	CVPTQSRIC	1	1
EEG-P7C-14	3	CHHSSPAHC	1	1
EEG-P7C-15	3	CTNTSLAYC	1	1
EEG-P7C-16	3	CPLVTSRVC	1	1
EEG-P7C-17	3	CLEPAHKAC	1	1
EEG-P7C-18	3	CQLERVLSC	1	1
EEG-P7C-19	3	CPPRNLPAC	1	1
EEG-P7C-20	3	CKSAQSLYC	1	1
EEG-P7C-21	3	CTNPSTNLC	1	1
EEG-P7C-22	3	CTNSSASLC	1	1
EEG-P7C-24	3	CSRTDHALC	1	1
EEG-P7C-25	3	CTTTMKAQC	1	1

Table A.28: PhD7C sequences from lead titanate screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-L7C-1	2	CVAPHKTHC	1	1
EEG-L7C-2	2	CAGKSANMC	1	1
EEG-L7C-3	2	CSKPSHDRC	1	1
EEG-L7C-4	2	CKQDHHRNC	1	1
EEG-L7C-5	2	CKPLTRTQC	1	1
EEG-L7C-6	2	CSQGTTVSC	1	1
EEG-L7C-7	2	CTGYLGHYC	1	1
EEG-L7C-8	2	CKKPPHAVC	1	1
EEG-L7C-9	2	CWVDQYMMC	1	1
EEG-L7C-10	2	CLETPSPYC	1	1
EEG-L7C-11	3	CSLKHHKSC	1	1
EEG-L7C-12	3	CKTHEPHKC	1	1
EEG-L7C-13	3	CVSLKKHLC	1	1
EEG-L7C-14	3	CAGYLPVIC	1	1
EEG-L7C-15	3	CTHEPMRNC	1	1
EEG-L7C-16	3	CTSAFKHKC	1	1
EEG-L7C-18	3	CSLAMPSEC	1	1
EEG-L7C-19	3	CKHMSLKSC	1	1
EEG-L7C-20	3	CEKKPHSPC	1	1
EEG-L7C-21	3	CLHKHPTPC	1	1
EEG-L7C-22	3	CVSANL AFC	1	1
EEG-L7C-23	3	CTFKAQHKC	1	1
EEG-L7C-24	3	CPHKWSKVC	1	1
EEG-L7C-25	3	CKEKHEPSC	1	1

Table A.29: PhD7C sequences from lithium niobate screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-LWMZ12-16	3	GHKVQINLIPPP	1	1
EEG-LWMZ12-17	3	GVHVRNHSPAPL	1	1
EEG-LWMZ12-18	3	FHKPSKYSPFSP	1	1
EEG-LWMZ12-19	3	VKPLHAALPTEP	1	1
EEG-LWMZ12-20	3	SQPDASSAATSP	1	1
EEG-LWMZ12-21	3	SMASAQSQTPSL	1	1
EEG-LWMZ12-22	3	KPANYDHWPSA	1	1
EEG-LWMZ12-23	3	HVQPQTMIALST	1	1
EEG-LWMZ12-24	3	LELPLSQEVLLR	1	1
EEG-LWMZ12-25	3	GQIKPHRLTYPP	1	1
EEG-LWMZ12-26	3	KPTQNFPLLERH	1	1
EEG-LWMZ12-27	3	IPLIAKAQTHTP	1	1
EEG-LWMZ12-28	3	KQTPLGAYFKAH	1	1
EEG-LWMZ12-29	3	HISSPHFTSSQD	1	1
EEG-LWMZ12-30	3	HNETRLPHSPAG	1	1

Table A.30: PhD12 sequences from single crystal lithium niobate –Z screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-12LM-4	3	GSTQAWMSPPLA	1	1

Table A.31: PhD12 sequences from single crystal lithium niobate –Z screened at pH 7.5 and eluted with bacteria; screening 1.

Clone Name	Round	Sequence	Index	Total
EEG-LMZR3-6	3	SLPHSIPILRMT	1	1
EEG-LMZR3-7	3	QGSLMPSIGSIQ	1	1
EEG-LMZR3-8	3	FSTDATNHSTHS	1	1
EEG-LMZR3-10	3	SWQHHTKLQPTR	1	1
EEG-LMZR3-11	3	TMGFTAPRFPHY	1	1
EEG-LMZR3-12	3	QLIKPNSCLLGC	1	1
EEG-LMZR3-15	3	AHXFSEPSARSS	1	1
EEG-LMZR3-18	3	QLTHTWSWALMT	1	1
EEG-LMZR5-21	5	ELFKTPADKRFF	1	3
EEG-LMZR5-22	5	ARAPLIPFASQH	1	1
EEG-LMZR5-23	5	TLGTMAMFPHTS	1	1
EEG-LMZR5-25	5	YLPYSASPPRL	1	2
EEG-LMZR5-26	5	WSPGQQRLHNST	1	1
EEG-LMZR5-27	5	NWNAGKGTMSPP	1	1
EEG-LMZR5-28	5	YLPYSASPPRL	2	2
EEG-LMZR5-29	5	SDTISRLHVSMT	1	2
EEG-LMZR5-30	5	ITPPGMWSQSQR	1	1
EEG-LMZR5-31	5	ELFKTPADKRFF	2	3
EEG-LMZR5-32	5	VLPPKPMRQPVA	1	1
EEG-LMZR5-33	5	WAHPLSSLRPAP	1	1
EEG-LMZR5-34	5	SDTISRLHVSMT	2	2
EEG-LMZR5-36	5	QDVHLTQQSRYT	1	1
EEG-LMZR5-37	5	ATIGPTTQMHPQ	1	1
EEG-LMZR5-38	5	ELFKTPADKRFF	3	3
EEG-LMZR5-39	5	YSTSQKAVSSKQ	1	1
EEG-LMZR5-40	5	HSSHTSNTLPSV	1	1

Table A.32: PhD12 sequences from single crystal lithium niobate –Z screened at pH 7.5 and eluted with bacteria; screening 2.

Clone Name	Round	Sequence	Index	Total
EEG-LWPZ12-1	3	HHSPRHPPPLPR	1	1
EEG-LWPZ12-2	3	WALTPPTRPLLT	1	1
EEG-LWPZ12-3	3	MKTHAASAIGSY	1	1
EEG-LWPZ12-4	3	GFKHQELVKGQF	1	1
EEG-LWPZ12-5	3	NIDIGIPFHLPR	1	1
EEG-LWPZ12-6	3	WHSDKHPLATRV	1	1
EEG-LWPZ12-7	3	EIHRMKQELLD	1	1
EEG-LWPZ12-8	3	TKAPHVHLLSPT	1	1
EEG-LWPZ12-9	3	YPNFKHQSKLHL	1	1
EEG-LWPZ12-10	3	SHVKPDWHIVLP	1	1
EEG-LWPZ12-11	3	HLKAPSAPVKVK	1	1
EEG-LWPZ12-12	3	KSDLHLAPRTPI	1	1
EEG-LWPZ12-13	3	GHTTGIATVSTS	1	1
EEG-LWPZ12-14	3	KPTLYAPHHQPW	1	1

Table A.33: PhD12 sequences from single crystal lithium niobate +Z screened at pH 7.5 and eluted with acid.

Clone Name	Round	Sequence	Index	Total
EEG-LPZR3-1	3	HSPSRIWALPSI	1	1
EEG-LPZR3-2	3	HSNEPNRLPTNL	1	1
EEG-LPZR3-3	3	LGNQSPQVEDTT	1	1
EEG-LPZR3-4	3	HLQDLHRPSRFY	1	1
EEG-LPZR3-5	3	QSSHDHSDNRQL	1	1
EEG-LPZR3-6	3	SQAVARSVNQDM	1	1
EEG-LPZR3-7	3	SATXTPAYLGFL	1	1
EEG-LPZR3-8	3	LSPQAYAELKLL	1	1
EEG-LPZR3-9	3	NMLGYRTSQSGN	1	1
EEG-LPZR3-10	3	SHALPLTWSTAA	1	1
EEG-LPZR4-1	4	KMHWHPPALNT	1	9
EEG-LPZR4-2	4	STPWFSYFRTL	1	2
EEG-LPZR4-3	4	NIAVGPNLRVSL	1	1
EEG-LPZR4-4	4	YLPYSASPPRL	1	1
EEG-LPZR4-5	4	KMHWHPPALNT	2	9
EEG-LPZR4-8	4	HSSHTSNTLPSV	1	2
EEG-LPZR4-9	4	KMHWHPPALNT	3	9
EEG-LPZR5-21	5	YRTDPLLYKTSV	1	1
EEG-LPZR5-22	5	TLGTMAMFPHTS	1	1
EEG-LPZR5-23	5	ADIMSQTTNGPH	1	1
EEG-LPZR5-24	5	LALQPNGASCGQ	1	1
EEG-LPZR5-25	5	KMHWHPPALNT	4	9
EEG-LPZR5-26	5	KMHWHPPALNT	5	9
EEG-LPZR5-27	5	APFIHSFAPYVP	1	1
EEG-LPZR5-28	5	KMHWHPPALNT	6	9
EEG-LPZR5-29	5	HWDPFSL SAYFP	1	1
EEG-LPZR5-30	5	STPWFSYFRTL	2	2
EEG-LPZR5-31	5	KMHWHPPALNT	7	9
EEG-LPZR5-32	5	HSFTPDLPPKV	1	1
EEG-LPZR5-33	5	IPSQITSSPKYN	1	1
EEG-LPZR5-34	5	KMHWHPPALNT	8	9
EEG-LPZR5-35	5	HEKHADYWTSQR	1	1
EEG-LPZR5-36	5	KMHWHPPALNT	9	9
EEG-LPZR5-37	5	SDTISRLHVSMT	1	1
EEG-LPZR5-38	5	HSSHTSNTLPSV	2	2
EEG-LPZR5-39	5	PQAPAKFQGAP	1	1

Table A.34: PhD12 sequences from single crystal lithium niobate +Z screened at pH 7.5 and eluted with bacteria.

Appendix B: Database and Analysis Program Source Codes

B.1: QUERY.PL

```
#!/usr/bin/perl -w

#set equivalence set to use for percent position &
statistics

my $eqlist = "LAIVG\nRKH\nWF\nSTY\nDE\nNQ";

#set search parameters (1 of 3)

    my $searchstring1 = "aragonite";

# my $query1 = "user";
# my $query1 = "substrate";
    my $query1 = "description";
# my $query1 = "round";
# my $query1 = "library";
# my $query1 = "elution";
# my $query1 = "seqname";
# my $query1 = "sequence";
# my $query1 = "comment";

    my $search2 = "and";
# my $search2 = "or";

#set search parameters (2 of 3)

    my $searchstring2 = "any";

# my $query2 = "user";
# my $query2 = "substrate";
    my $query2 = "description";
# my $query2 = "round";
# my $query2 = "library";
# my $query2 = "elution";
# my $query2 = "seqname";
```

```

# my $query2 = "sequence";
# my $query2 = "comment";

# my $search3 = "and";
  my $search3 = "or";

#set search parameters (3 of 3)

  my $searchstring3 = "any";

# my $query3 = "user";
# my $query3 = "substrate";
  my $query3 = "description";
# my $query3 = "round";
# my $query3 = "library";
# my $query3 = "elution";
# my $query3 = "seqname";
# my $query3 = "sequence";
# my $query3 = "comment";

##### END OF USER INPUT #####
#open temp files for downloading
my $tempdir =
"$searchstring1"."_"."$searchstring2"."_"."$searchstring3";
mkdir ("/home/Erin Gooch/database/$tempdir", 0777) ||
die "Can't make temp directory: $!";
open (ERROR, ">/home/Erin
Gooch/database/$tempdir/errors.txt") || die "Can't
open errors: $!";
my $tempfile1 = "fasta_with_doubles.txt";
my $tempfile4 = "fasta_no_doubles.txt";
my $tempfile2 = "excel.xls";
open (TEMP1, ">/home/Erin
Gooch/database/$tempdir/$tempfile1") || print ERROR
"Can't open $tempfile1: $!";
open (TEMP2, ">/home/Erin
Gooch/database/$tempdir/$tempfile2") || print ERROR
"Can't open $tempfile2: $!";

```

```

open (TEMP4, ">/home/Erin
Gooch/database/$tempdir/$tempfile4") || print ERROR
"Can't open $tempfile4: $!";
open (INPUT, ">/home/Erin
Gooch/database/$tempdir/inputfile.txt") || print ERROR
"Can't open inputfile: $!";

@inputeq = split("\n", $eqlist);
print INPUT "eqlist\t@inputeq\n";
print ERROR "$tempfile1\n";
print ERROR "$tempfile2\n";
print ERROR "$search2, $search3\n";

#set up search classification
my @class;
if ($searchstring1 eq ""){$searchstring1 = "any";}
#set 1st class bit
    if ($searchstring1 =~ /^any/) {$class[0] = 0;}
    else {$class[0] = 1;}

if ($search2 eq "or") {$class[1] = 0;}
#set 2nd class bit
    elsif ($search2 eq "and") {$class[1] = 1;}

if ($searchstring2 eq ""){$searchstring2 = "any";}
#set 3rd class bit
    if ($searchstring2 =~ /^any/) {$class[2] = 0;}
    else {$class[2] = 1;}

if ($search3 eq "or") {$class[3] = 0;}
#set 4th class bit
    elsif ($search3 eq "and") {$class[3] = 1;}

if ($searchstring3 eq ""){$searchstring3 = "any";}
#set 5th class bit
    if ($searchstring3 =~ /^any/) {$class[4] = 0;}
    else {$class[4] = 1;}

my $class = join ("", @class);
print ERROR "class is $class\n";

```

```

#organize query parameters into arrays
my @strings;
$strings[0] = $searchstring1;
$strings[1] = $searchstring2;
$strings[2] = $searchstring3;

my @fields;
$fields[0] = $query1;
$fields[1] = $query2;
$fields[2] = $query3;

#set match for classes based on logic = (A and/or B)
and/or C
my @boolean = "";
#class 1: match all
if ($class eq "01000"){@boolean = ("001", "010",
"100", "110", "101", "011", "111", "000");}
elsif ($class eq "00010"){@boolean = ("001", "010",
"100", "110", "101", "011", "111", "000");}
elsif ($class eq "01010"){@boolean = ("001", "010",
"100", "110", "101", "011", "111", "000");}
elsif ($class eq "00000"){@boolean = ("001", "010",
"100", "110", "101", "011", "111", "000");}
#class 2: match any A
elsif ($class eq "11000"){@boolean = ("100", "110",
"101", "111");}
elsif ($class eq "10010"){@boolean = ("100", "110",
"101", "111");}
elsif ($class eq "11010"){@boolean = ("100", "110",
"101", "111");}
elsif ($class eq "10000"){@boolean = ("100", "110",
"101", "111");}
#class 3: match any B
elsif ($class eq "01100"){@boolean = ("010", "110",
"011", "111");}
elsif ($class eq "00110"){@boolean = ("010", "110",
"011", "111");}
elsif ($class eq "01110"){@boolean = ("010", "110",
"011", "111");}
elsif ($class eq "00100"){@boolean = ("010", "110",
"011", "111");}

```

```

#class 4: match any C
elsif ($class eq "01001"){@boolean = ("001", "101",
"011", "111");}
elsif ($class eq "00011"){@boolean = ("001", "101",
"011", "111");}
elsif ($class eq "01011"){@boolean = ("001", "101",
"011", "111");}
elsif ($class eq "00001"){@boolean = ("001", "101",
"011", "111");}
#class 5: match A or B
elsif ($class eq "10110"){@boolean = ("010", "100",
"110", "101", "011", "111");}
elsif ($class eq "10100"){@boolean = ("010", "100",
"110", "101", "011", "111");}
#class 6: match A and B
elsif ($class eq "11100"){@boolean = ("110", "111");}
elsif ($class eq "11110"){@boolean = ("110", "111");}
#class 7: match A or C
elsif ($class eq "10011"){@boolean = ("001", "100",
"110", "101", "011", "111");}
elsif ($class eq "10001"){@boolean = ("001", "100",
"110", "101", "011", "111");}
#class 8: match A and C
elsif ($class eq "11001"){@boolean = ("101", "111");}
elsif ($class eq "11011"){@boolean = ("101", "111");}
#class 9: match B or C
elsif ($class eq "01101"){@boolean = ("001", "010",
"110", "101", "011", "111");}
elsif ($class eq "00101"){@boolean = ("001", "010",
"110", "101", "011", "111");}
#class 10: match B and C
elsif ($class eq "00111"){@boolean = ("011", "111");}
elsif ($class eq "01111"){@boolean = ("011", "111");}
#class 11: match A and B or C
elsif ($class eq "11101"){@boolean = ("001", "110",
"101", "011", "111");}
#class 12: match A or B and C
elsif ($class eq "10111"){@boolean = ("101", "011",
"111");}
#class 13: match A or B or C

```

```

elseif ($class eq "10101"){@boolean = ("001", "010",
"100", "110", "101", "011", "111");}
#class 14: match A and B and C
elseif ($class eq "11111"){@boolean = ("111");}
print ERROR "boolean = @boolean\n";

#start search
my $index;
my $x;
my %matchlist;
my $grab;
my $boo;

open (DATABASE, "/home/Erin Gooch/database/NEWDB.txt")
|| die "Can't open database: $!";
$/ = "\r\n";
while (my $line = <DATABASE>){
    chomp $line;
    my @line = split (/:/, $line);
    my @check = (0,0,0);
    for ($x = 0; $x <= $#strings; $x++){
        if ($fields[$x] =~ /seqname/){$index = 0;}
        elsif ($fields[$x] =~ /user/){$index = 1;}
        elsif ($fields[$x] =~ /substrate/){$index =
2;}
        elsif ($fields[$x] =~ /description/){$index =
3;}
        elsif ($fields[$x] =~ /round/){$index = 4;}
        elsif ($fields[$x] =~ /library/){$index = 5;}
        if ($strings[$x] =~ /PhDC7/i){$strings[$x] =
"PhD7C";}}
        elsif ($fields[$x] =~ /elution/){$index = 6;}
        elsif ($fields[$x] =~ /sequence/){$index = 7;}
        $strings[$x] =~ tr/Xx/./;
        elsif ($fields[$x] =~ /comment/){$index = 8;}
        if ($index == 5){
            if ($line[$index] =~ /$strings[$x]\b/i)
            {$check[$x] = 1;}
            else {$check[$x] = 0;}
        }
        else {

```



```

        if ($line[$index] =~ /$strings[$x]/i)
    {$check[$x] = 1;}
#     elsif ($strings[$x] eq "") {$check[$x] = 1;}
#     elsif ($strings[$x] =~ /^any/i) {$check[$x] = 1;}
        else {$check[$x] = 0;}
    }
#         print ERROR "match field: $field[$x]
index: $index  match: $line[$index] with:
$strings[$x]\n";
}#end of match line loop, grab if match is made

    my $binary = join ("", @check);
#     print ERROR "match $binary\n";
    foreach $boo (@boolean) {
        if ($boo =~ /$binary/) {$grab = 1; last;}
        else { $grab = 0;}
    }
    if ($grab == 1) {$matchlist{$line} = 1;}
}#end of read database loop
close DATABASE;

#make result list
my @matches = keys(%matchlist);
@matches = sort (@matches);

#find duplicate sequences
my %doubles;
my %names;
foreach my $m (@matches){
    my @entry = split(/:/, $m);
    print ERROR "comment = $entry[8]!\n";
    my $no_amp =
"$entry[1]".$entry[2]".$entry[3]".$entry[5]".$entr
y[6]".$entry[7]";
    $doubles{$no_amp} += 1;
    $names{$no_amp} = "$m";
}#end of find only duplicates instructions
my @doubles = keys(%doubles);
my @doublelist = values(%names);

my @total;

```

```

my @index;
my %index;
for (my $p = 0; $p <= $#matches; $p++){
    my @entry = split(/:/, $matches[$p]);
    my $no_amp2 =
"$entry[1]"."$entry[2]"."$entry[3]"."$entry[5]"."$entry[6]"."$entry[7]";
    foreach my $d (@doubles){
        if ($no_amp2 =~ /$d/){
            $index{$d} +=1;
            $index[$p] = $index{$d};
            $total[$p] = $doubles{$d};
        }
    }
}

#write to fasta, xls files: with doubles
print TEMP2 "With duplicate sequences\n";
print TEMP2 "Clone Name\tUser
Name\tSubstrate\tDescription\tRound\tLibrary\tElution
Method\tSequence\tComment\tIndex\tTotal\n";
#my $count = 0;
my @seqlist;
my @liblist;
for (my $p = 0; $p <= $#matches; $p++){
    my $list = $matches[$p];
    # print ERROR "match = $list:end\n";
    my @entry = split(/:/, $list);
    $seqlist[$p] = $entry[7];
    $liblist[$p] = $entry[5];
    print TEMP1 ">$entry[0]
$index[$p]/$total[$p]\n$entry[7]\n";
    foreach my $entry (@entry){
        print TEMP2 "$entry\t";
    }
    print TEMP2 "$index[$p]\t$total[$p]\n";
#    $count++;
}

#write to files: without doubles
$countout = 2*($#doublelist+1);

```

```

print INPUT "count\t$countout\n";
my @seqlist2;
my @liblist2;
my $short;
my $short2;
my @reverse;
print TEMP2 "\n\n\nWithout duplicate sequences\n";
print TEMP2 "Clone Name\tUser
Name\tSubstrate\tDescription\tRound\tLibrary\tElution
Method\tSequence\tComment\n";
for (my $p = 0; $p <= $#doublelist; $p++){
    my $list2 = $doublelist[$p];
    print ERROR "doubles = $list2:end\n";
    my @entry2 = split (/:/, $list2);
    $seqlist2[$p] = $entry2[7];
    $liblist2[$p] = $entry2[5];
    if ($entry2[5] =~ /PhD12/) {$send = 12;}
    elsif ($entry2[5] =~ /PhD7C\b/) {$send = 9;}
    elsif ($entry2[5] =~ /PhD7\b/) {$send = 7;}
print ERROR "end = $send\n";
$sendm1 = $send - 1;
if ($#reverse > $sendm1){
    until ($#reverse = $sendm1) {
        $trash = pop(@reverse);
    }
    for (my $raa = 0; $raa < $send; $raa++){
        my $new = $send - 1 - $raa;
#print ERROR "end = $send, raa = $raa, new = $new\n";
        $reverse[$new] = substr($entry2[7], $raa, 1);
    }
    my $reverse = join ("", @reverse);
    if ($entry2[5] =~ /PhD7C\b/) {
        $short = substr($entry2[7], 1, 7);
        $short2 = substr($reverse, 1, 7);
        print ERROR "short seq = $short\nentry2[7] =
$entry2[7]\n";
        print TEMP4 ">$entry2[0] 1\n$short\n";
        print INPUT
"$entry2[0]\t1\t$short\t$entry2[5]\n";
        print TEMP4 ">$entry2[0] 2\n$short2\n";

```

```

        print INPUT
"$entry2[0]\t2\t$short2\t$entry2[5]\n";
    }
    else {print TEMP4 ">$entry2[0] 1\n$entry2[7]\n";
        print INPUT
"$entry2[0]\t1\t$entry2[7]\t$entry2[5]\n";
        print ERROR "regular seq = $entry2[7]\n";
        print TEMP4 ">$entry2[0] 2\n$reverse\n";
        print INPUT
"$entry2[0]\t2\t$reverse\t$entry2[5]\n";
    }

        foreach my $entry2 (@entry2){
            print TEMP2 "$entry2";
            print TEMP2 "\t";
        }
    print TEMP2 "\n";
}

close TEMP1;
close TEMP2;

##### START OF PERCENT POSITION CODE
#####

#my $seqlist = join(":", @seqlist);
#my $liblist = join(":", @liblist);
my $seqlist2 = join(":", @seqlist2);
my $liblist2 = join(":", @liblist2);
#my $liblist = "PhD12:PhD12:PhD12:PhD12";
#my $seqlist =
"LKIPRDEWANRL:EWQKINYPSLRH:WMIRTQALDCPF:KHYRGTPVSWLA";

#check library list
print ERROR "liblist is $liblist2\n";
my @libs = split(":", $liblist2);
my $libflag = 0;
my $end = $#libs - 1;
my $lib;
for (my $l = 0; $l <= $end; $l++){
    my $n = $l + 1;

```

```

    if ($libs[$1] =~ /$libs[$n]/){
        $libflag = 0;
        $lib = $libs[$n]}
    else {$libflag = 1; last;}
}
#print ERROR ("library = $libs[$n]");

#open temp files for downloading
my $tempfile3 = "percentposition.xls";
open (TEMP3, ">/home/Erin
Gooch/database/$tempdir/$tempfile3") || print ERROR
"Can't open $tempfile2: $!";

#warning for non-matching libraries
print ERROR "libflag is $libflag\n";
if ($libflag == 1){
print ERROR ("Your search generated matches of
different libraries. Percent position is only
meaningful when you compare sequences from the same
library. Go back.");
}

#instructions for matching libraries
elsif ($libflag == 0){

#set sequence length
print ERROR "lib = $lib\n";
if ($lib =~ /PhD12/){$end = 12;}
#elsif($lib =~ /PhD7C/){$end = 9;}
elsif($lib =~ /PhD7C\b/){$end = 7;}
elsif($lib =~ /PhD7\b/){$end = 7;}
else {print ERROR ("Error in determining sequence
length");}

#add table row labels
my $endline = $end+3;
my @AAlist = qw(A C D E F G H I K L M N P Q R S T V W
Y X);
my @table1;
for (my $g = 0; $g <=#AAlist; $g++){
    my $f = $g*$endline;

```

```

    $table1[$f] = $AAlist[$g];
    for (my $e = 1; $e<$endline; $e++){
        my $e2 = $f+$e;
        $table1[$e2] = 0;
    }
}
my @eqlist = split ("\n", $eqlist);
my @table2;
for (my $g = 0; $g <=$#eqlist; $g++){
    my $f = $g*$endline;
    $table2[$f] = $eqlist[$g];
    for (my $e = 1; $e<$endline; $e++){
        my $e2 = $f+$e;
        $table2[$e2] = 0;
    }
}

#tally each sequence
my @seqs = split (":", $seqlist2);
my $total = $#seqs+1;
foreach my $seq (@seqs){
    my @AAs;
    if ($lib =~ /PhD7C\b/){
        for (my $a = 0; $a < 7; $a++){
            my $b = $a+1;
            $AAs[$a] = substr($seq, $b, 1);
            for (my $x = 0; $x <= $#AAlist; $x++){
                if ($AAs[$a] =~ /$AAlist[$x]/){
                    my $p = $x*($endline)+$a+1;
                    $table1[$p] += 1*(100/$total);
                    my $q = $x*($endline)+($endline-2);
                    $table1[$q] += 1*(100/($total*$end));
                    if ($AAs[$a] =~ /C/) {
                        print ERROR "a = $a\np = $p\nq =
$q\ntable[p] = $table1[$p]\ntable[q] =
$table1[$q]\n";}
                }
            }#end of for each AA in AAlist
        } #end of for each position
    }#end of if PhD7C
else {

```

```

for (my $a = 0; $a < $end; $a++){
    $AAs[$a] = substr($seq, $a, 1);
    for (my $x = 0; $x <= $#AAList; $x++){
        if ($AAs[$a] =~ /$AAList[$x]/){
            my $p = $x*($endline)+$a+1;
            $table1[$p] += 1*(100/$total);
            my $q = $x*($endline)+($endline-2);
            $table1[$q] += 1*(100/($total*$end));
        }
    }
    }#end of for each AA in AAList
} #end of for each position
}#end of else

}#end of for each sequence

#set library percentages
my @libavg;
if ($lib =~ /PhD12/){
    @libavg = qw(6.0 0.5 2.8 3.1 3.3 2.6 6.3 3.4 2.8 9.3
2.6 4.6 12.2 5.1 4.7 10.0 11.1 3.9 2.2 3.6 0.0);
}#end of if phd12 lib average loop
elsif ($lib =~ /PhD7C\b/){
    @libavg = qw(6.5 0.0 4.1 3.1 2.1 2.2 6.9 2.1 3.8 9.6
3.3 6.4 10.7 7.1 4.3 8.6 13.1 1.9 1.9 2.4 0.0);
} #end of phd7c library average loop
elsif ($lib =~ /PhD7\b/){
    @libavg = qw(6.7 0.6 3.1 2.2 2.2 4.9 3.7 2.9 2.4 9.4
2.7 4.7 13.1 5.1 3.7 11.8 10.8 4.2 1.8 3.9 0.0);
} #end of phd7 library average loop
for (my $t = 0; $t<=#libavg; $t++){
    my $p = $t*($endline)+($endline-1);
    $table1[$p] = $libavg[$t];
#    $table1x[$p] = $libavg[$t];
}

#make excel array/convert tally to percentage
#print TEMP "$total\tSequence total\n";
#print TEMP "$end\tlibrary residues\n\n";
print TEMP3 "Percent Position\n\n";

#set up html and excel tables

```

```

if ($end == 12){
    print TEMP3
    "AA\t1\t2\t3\t4\t5\t6\t7\t8\t9\t10\t11\t12\tTotal\tLib
    . Avg.";
}
elseif ($end == 7){
    print TEMP3 "AA\t1\t2\t3\t4\t5\t6\t7\tTotal\tLib.
    Avg.";
}
elseif ($end ==9){
    print TEMP3
    "AA\t1\t2\t3\t4\t5\t6\t7\t8\t9\tTotal\tLib. Avg.";
}
for (my $r = 0; $r <= $#table1; $r++){
    if ($table1[$r] =~ /[A-Z]/){
        print TEMP3 "\n$table1[$r]\t";
    }
    else {
        print TEMP3 "=$table1[$r]\t";
    }
}

#add equivalent AA percentages
for (my $e = 0; $e <= $#eqlist; $e++){
    for (my $d = 0; $d <= $#table1; $d++){
        if ($table1[$d] =~ /[$eqlist[$e]]/){
            for (my $r = 1; $r < $endline; $r++){
                my $f = $e*$endline + $r;
                my $g = $d + $r;
                $table2[$f] += $table1[$g];
            }
        }
    }
}

print TEMP3 "\n\nEquivalent Sets";
if ($end == 12){
    print TEMP3
    "\n\nAAs\t1\t2\t3\t4\t5\t6\t7\t8\t9\t10\t11\t12\tTotal
    \tLib. Avg.";
}

```



```

elsif ($end == 7){
    print TEMP3
    "\n\nAAs\t1\t2\t3\t4\t5\t6\t7\tTotal\tLib. Avg.";
}
elsif ($end ==9){
    print TEMP3
    "\n\nAAs\t1\t2\t3\t4\t5\t6\t7\t8\t9\tTotal\tLib.
Avg.";
}

for (my $r = 0; $r <= $#table2; $r++){
    if ($table2[$r] =~ /[A-Z]/){
        print TEMP3 "\n$table2[$r]\t";
    }
    else {
        print TEMP3 "$table2[$r]\t";
    }
}
}#end of if library matches instructions
close TEMP1;
close TEMP2;
close TEMP3;
close TEMP4;
close INPUT;
close ERROR;
close DATABASE;
die "End of query sequences\n";

```

B.2: FORWARDSEQUENCES.PL

```

#!/usr/bin/perl -w

opendir (OLD, "old3") || die "can't open old
directory:$!";
while ($name = readdir(OLD)){
    unless ($name =~ /\./){
        open (IN1, "</home/Erin
Gooch/database/old3/$name/fasta_no_doubles.txt") ||
die "can't open $name/fasta_no_doubles.txt: $!";

```

```

        open (IN2, "</home/Erin
Gooch/database/old3/$name/inputfile.txt") || die
"can't open $name/input.txt: $!";
        open (FASTA, ">/home/Erin
Gooch/database/old3/$name/fasta2.txt") || die "can't
open $name/fasta2.txt: $!";
        open (OUT, ">/home/Erin
Gooch/database/old3/$name/inputfile2.txt") || die
"can't open $name/inputfile2.txt: $!";

print "$name\n";
while ($line1 = <IN1>){
    @line = split (/\\s+/, $line1);
    if ($line[1] == 1){
        $line2 = <IN1>;
        print FASTA "$line1$line2";
    }
}
close IN1;
close FASTA;

$line3 = <IN2>;
print OUT "$line3";
$line3 = <IN2>;
@line = split (/\\s+/, $line3);
$a = $line[1]/2;
print OUT "count\\t$a\\n";
while ($line3 = <IN2>){
    @line = split (/\\s+/, $line3);
    if ($line[1]== 1){
        print OUT "$line3";
    }
}

close IN2;
close OUT;

}}
die "End of run";

```

B.3: CALCPROBABILITY.PL

```
#!/usr/bin/perl -w
$/ = "\n";

open (OUT, ">/home/Erin
Gooch/database/old3/allprobs.xls") || die "can't open
allprobs.xls: $!";
open (OUT2, ">/home/Erin
Gooch/database/old3/alllits.xls") || die "can't open
alllits.xls: $!";

#AA frequencies for each library
@freq12 = qw(6.0 0.5 2.8 3.1 3.3 2.6 6.3 3.4 2.8 9.3
2.6 4.6 12.2 5.1 4.7 10.0 11.1 3.9 2.2 3.6 0.0);
@freq7C = qw(6.5 0.0 4.1 3.1 2.1 2.2 6.9 2.1 3.8 9.6
3.3 6.4 10.7 7.1 4.3 8.6 13.1 1.9 1.9 2.4 0.0);
@freq7 = qw(6.7 0.6 3.1 2.2 2.2 4.9 3.7 2.9 2.4 9.4
2.7 4.7 13.1 5.1 3.7 11.8 10.8 4.2 1.8 3.9 0.0);
print OUT2 "frequencies PhD12\t";
foreach $amino (@freq12){
    print OUT2 "$amino\t";
}
print OUT2 "\nfrequencies PhD7C\t";
foreach $amino (@freq7C){
    print OUT2 "$amino\t";
}
print OUT2 "\nfrequencies PhD7\t";
foreach $amino (@freq7){
    print OUT2 "$amino\t";
}
print OUT2
"\n\nLiterals\tPattern\tWindow\tCount12\tCount7C\tCoun
t7\tFolder Name\n";

opendir (OLD, "old3") || die "can't open old
directory:$!";
while ($name = readdir(OLD)){
    unless ($name =~ /\./){
        print "$name\n";
    }
}
```

```

    open (ERROR, ">/home/Erin
Gooch/database/old3/$name/errorsprobF.txt") || die
"Can't open errors: $!";
    open (ERROR2, ">/home/Erin
Gooch/database/old3/$name/errorsprobF.xls") || die
"Can't open errors: $!";

    open (OUTFILE, ">/home/Erin
Gooch/database/old3/$name/ProbabilityF.xls") || die
"Can't open probability file: $!";
    print OUTFILE
"Probability\tNum\tSignificance\tPattern\tL\t" ||
print "can't print to out file\n";
    open (IN, "</home/Erin
Gooch/database/old3/$name/inputfile2.txt") || die
"Can't open inputfile: $!";
    open (OLDPATT, "</home/Erin
Gooch/database/old3/$name/TeirResults2.txt") || die
"Can't open pattern2 file: $!";
    open (PATTERNS, ">/home/Erin
Gooch/database/old3/$name/TeirResults3.txt") || die
"Can't write pattern3 file: $!";
print ERROR "directory: $name\n";

#remove duplicate patterns
@line = <OLDPATT>;
for ($g = 0; $g <= $#line; $g++){
    $print = 1;
    $sub = $line[$g];
    while ($sub =~ /\|/){$sub =~ s/\|/1/; $sub =~
s/\|/1/;}
    while ($sub =~ /\../){$sub =~ s /\../0/;}
#    print ERROR "*****\n
current sequence is $sub There are $#line elements in
array\n";
    for ($p = 0; $p < $g; $p++) {
        $sub2 = $line[$p];
        while ($sub2 =~ /\|/){$sub2 =~ s/\|/1/; $sub2 =~
s/\|/1/;}
        while ($sub2 =~ /\../){$sub2 =~ s /\../0/;}
#    print ERROR "matching with $sub2";

```

```

        if ($sub =~ /$sub2/ ) {
            $print = 0;
#            print ERROR "found duplicate pattern:
$line[$g] and $line[$p], $g =\ $p\n";
        }
    }
    if ($print == 1) {print PATTERNS "$line[$g]";}
}
close OLDPATT;
close PATTERNS;
if ($#line > 0) {
    until ($#line < 1){
        $trash = pop (@line);
    }
}
open (PATTERNS, "</home/Erin
Gooch/database/old3/$name/TeirResults3.txt") || die
"Can't read pattern3 file: $!";

#read in sequences from input file
@in2 = <IN>;
@AAlist = qw(A C D E F G H I K L M N P Q R S T V W Y
X);
print "Read input file\n";

#determine eq list
@eqlist = split (/\\s+/, $in2[0]);
$trash = shift (@eqlist);
@eqlist2 = @eqlist;
my $chk = 0;
foreach $aa (@AAlist){
    $chk = 0;
    foreach $eq (@eqlist2){
        if ($aa =~ /[$eq]/){$chk = 1; last;}
    }
}
if ($chk == 0) {push @eqlist2, $aa;}
}
for ($eq2 = 0; $eq2 <= $#eqlist2; $eq2++){
    $eqarray[$eq2] = 0;
    $eqtotal[$eq2] = 0;
    print OUTFILE "$eqlist2[$eq2]\t";
}

```

```

}print OUTFILE "\n";
print ERROR "eq list is @eqlist2\n";

#determine number of sequences
@countlist = split (/s+/, $in2[1]);
$count2 = $countlist[1];
#print "number of input sequences is $count2\n";

#initialize tables
for ($q = 0; $q < $count2; $q++){
    $namef[$q] = "";
    $indexf[$q] = 0;
    $seqf[$q] = "";
    $libf[$q] = "";
}#end of initialize arrays
if ($#namef >= $count2){
print ERROR "$#namef is larger than $count2, removing
extra array elements\n";
    until ($#namef < $count2) {
        $trash = pop (@namef);
        $trash = pop (@indexf);
        $trash = pop (@seqf);
        $trash = pop (@libf);
    }
}#end of get rid of extra array elements

#sort sequence data into tables
$count = 0;
@sq = @eqtotal;
for ($w = 2; $w <= $#in2; $w++) {
    @input = split (/s+/, $in2[$w]);
    $namef[$count] = $input[0];
    $indexf[$count] = $input[1];
    $seqf[$count] = $input[2];
    $libf[$count] = $input[3];

    #tally amino acids for query
    $leng =length($seqf[$count]) - 1;
    if ($#sq > $leng){

```

```

        until ($#sq = $leng) {
            $trash = pop (@sq);
        }
    }#end of get rid of extra array elements

    for ($ab = 0; $ab <= $leng; $ab++){
        $sq[$ab] = substr($seqf[$count], $ab, 1);
        for ($ac = 0; $ac <= $#eqlist2; $ac++){
            if ($sq[$ab] =~ /[$eqlist2[$ac]]/) {
                $eqtotal[$ac]++;
            }
        }
    }#match AAs to eqlist
    $count++;
}#end of sort sequences from input file
print ERROR "@eqlist2\n@eqtotal\n";
print ERROR "number of seqs is $count2 and seq list
($#namef = $#libf) is @namef\n";

#read in patterns
$count = 0;
while (defined($in = <PATTERNS>)){ $in3[$count] = $in;
$count++;}
print "Read patterns file\n";
print ERROR "Number of patterns is $count - 1\n";

if ($#in3 >= $count){
    print ERROR "$#in3 is larger than $count, removing
extra array elements\n";
    until ($#in3 < $count) {
        $trash = pop (@in3);
        $trash = pop (@l);
        $trash = pop (@w);
        $trash = pop (@k);
        $trash = pop (@number);
        $trash = pop (@signif);
        $trash = pop (@pattern);
        $trash = pop (@sequences);
        $trash = pop (@positions);
        $trash = pop (@warn);
    }
}

```

```

    }
}#end of get rid of extra array elements

$count = 0;
for ($eachline = 0; $eachline <=#in3; $eachline++) {
#print ERROR "*****\n";
    $warn[$eachline] = "";
    if ($in3[$eachline] =~ /LWK/i) {
        $inpt = substr($in3[$eachline], 4, 99);
        @inpt = split (/,/, $inpt);
        $l = $inpt[0]; print ERROR "l = $l, ";
        $w = $inpt[1]; print ERROR "w = $w, ";
        $k = $inpt[2]; print ERROR "k = $k\n";
    }
    else {
        $l[$count] = $l;
        $w[$count] = $w;
        $k[$count] = $k;
        @line = split (/s+/, $in3[$eachline]);
        $number[$count] = $line[0];
        $signif[$count] = $line[2];
        $pattern[$count] = $line[3];
        $h = 4; $seq = ""; $position = "";
        while ($h<=#line) {
            $seq .= "$line[$h],";
            $h++;
            $position .= "$line[$h],";
            $h++;
        }#end of while reading pattern sequence and
position indices
        $sequences[$count] = $seq;        chop
$sequences[$count];
        $positions[$count] = $position;  chop
$positions[$count];
print ERROR "seq list is $sequences[$count]\n";

#Calculate number of literals in pattern
    $length = length($pattern[$count]);
    for ($a = 0; $a < $length; $a++){
        $lit[$a] = substr($pattern[$count], $a, 1);

```



```

}#end of initialize literals array
$l2 = $length - 1;
if ($#lit > $l2){
    until ($#lit = $l2) {
        $trash = pop(@lit);
    } }#delete extra array elements for short seqs
$join = 0;
$lpatt = 0;
$splice = "";
    for ($a = 0; $a < $length; $a++){
        if ($lit[$a] =~ /\[/) {
            $t2 = $a;
            $temp = "";
            $join = 1;
        }#end of if [
        elsif ($lit[$a] =~ /\]/) {
            $lit[$t2] = $temp;
            $splice .= "$a,";
            $join = 0;
            $lpatt++;
        }#end of elsif ]
        elsif ($lit[$a] =~ /\./) {
            $lpatt++;
            $splice .= "$a,";
        }#end of elsif .
        elsif ($lit[$a] =~ /[A-Z]/){
            if ($join == 1){
                $temp .= "$lit[$a]";
                $splice .= "$a,";
            } #end of if join yes
            elsif ($join == 0) {
                $lpatt++;
            }#end of if join no
        }#end of if literal is an AA
    }#end of for each pattern element
chop $splice;
@splice = split(/,/ , $splice);
for ($b = $#splice; $b >= 0; $b--){
#    $sp = 0;
    $a = $splice[$b];
    splice (@lit, $a, 1);

```

```

    }
    print ERROR "pattern is @lit\n";
    @lit2 = sort(@lit);

#AA frequencies for each library
@freq12 = qw(6.0 0.5 2.8 3.1 3.3 2.6 6.3 3.4 2.8 9.3
2.6 4.6 12.2 5.1 4.7 10.0 11.1 3.9 2.2 3.6 0.0);
@freq7C = qw(6.5 0.0 4.1 3.1 2.1 2.2 6.9 2.1 3.8 9.6
3.3 6.4 10.7 7.1 4.3 8.6 13.1 1.9 1.9 2.4 0.0);
@freq7 = qw(6.7 0.6 3.1 2.2 2.2 4.9 3.7 2.9 2.4 9.4
2.7 4.7 13.1 5.1 3.7 11.8 10.8 4.2 1.8 3.9 0.0);

#check libraries
$count12 = 0;
$count7C = 0;
$count7 = 0;
@seqlist = split (/,/, $sequences[$count]);
for ($g = 0; $g <=#seqlist; $g++){
    $h = $seqlist[$g];
    if ($libf[$h] =~ /PhD12/){
        $count12++;
        $warn[$count] .= "PhD12,";
    }
    elsif ($libf[$h] =~ /PhD7C/){
        $count7C++;
        $warn[$count] .= "PhD7C,";
    }
    elsif ($libf[$h] =~ /PhD7/){
        $count7++;
        $warn[$count] .= "PhD7,";
    }
}
chop $warn[$count];
print ERROR "Counts: $count12, $count7C,
$count7\n";
print ERROR2
"$pattern[$count]\t$count12\t$count7C\t$count7\t";
print OUT2
"@lit2\t$pattern[$count]\t$lpatt\t$count12\t$count7C\t
$count7\t$name\n";

```

```

#calc frequencies for Eq's
for ($x = 0; $x<=#eqlist; $x++){
    $frereq12[$x] = 0;
    $frereq7C[$x] = 0;
    $frereq7[$x] = 0;
}

for ($x = 0; $x <= $#eqlist; $x++){
    for ($y = 0; $y <= $#AAlist; $y++){
        if ($AAlist[$y] =~ /[$eqlist[$x]]/){
            $frereq12[$x] += $freq12[$y];
            $frereq7C[$x] += $freq7C[$y];
            $frereq7[$x] += $freq7[$y];
        } } }

@freq12 = (@freq12, @frereq12);
@freq7C = (@freq7C, @frereq7C);
@freq7 = (@freq7, @frereq7);
@freqAA = (@AAlist, @eqlist);
for ($rr = 0; $rr <=#freqAA; $rr++){
    $freq12[$rr] /= 100;
    $freq7C[$rr] /= 100;
    $freq7[$rr] /= 100;
}
print ERROR "AA-eq list is @freqAA\n";
print ERROR "freqs are @freq12\n";

#calculate probability - send to array:
$prob[$count]
$prob12 = 1;
$prob7C = 1;
$prob7 = 1;
for ($lt = 0; $lt <= $#lit; $lt++) {
    for ($ad = 0; $ad <= $#eqlist2; $ad++){
        if ($lit[$lt] =~ /[$eqlist2[$ad]]/){
            $eqarray[$ad]++;
        } }
    for ($aa3 = 0; $aa3 <= $#freqAA; $aa3++) {
        if ($freqAA[$aa3] =~ /\b$lit[$lt]\b/){
            $prob12 *= $freq12[$aa3];
            $prob7C *= $freq7C[$aa3];
        }
    }
}

```

```

        $prob7 *= $freq7[$aa3];
    } } }
print ERROR2 "$prob12\t$prob7\t$prob7C\t";
$prob12 *= (12 - $lpatt + 1);
$prob7C *= (7 - $lpatt + 1);
$prob7 *= (7 - $lpatt + 1);
print ERROR2 "$prob12\t$prob7\t$prob7C\t";

$prob12 = $prob12**$count12;
$prob7C = $prob7C**$count7C;
$prob7 = $prob7**$count7;
print ERROR2 "$prob12\t$prob7\t$prob7C\t";

$prob[$count] = $prob12 * $prob7C * $prob7;
print ERROR2 "$prob[$count]\n";

#print results to outfile
print OUT
"$prob[$count]\t$number[$count]\t$signif[$count]\t$pat
tern[$count]\t$l[$count]\t$name\n";
print OUTFILE
"$prob[$count]\t$number[$count]\t$signif[$count]\t$pat
tern[$count]\t$l[$count]";
    foreach $eq (@eqarray) {
        print OUTFILE "\t$eq";
    }
print OUTFILE "\n";

    foreach $eq (@eqarray) {
        $eq = 0;
    }

    $count++;
}#end of else line is pattern
}#end of read in patterns
print ERROR "pattern list ($#in3) is @l\n";

print OUTFILE "\n\n\t\t\t\t\t";
foreach $eq3 (@eqlist2){

```

```

    print OUTFILE "$eq3\t";
}
print OUTFILE "\n\t\t\t\t\t\t";
foreach $eq4 (@eqtotal){
    print OUTFILE "$eq4\t";
}

#reset array element
for ($r = 0; $r<=$count; $r++){
    $print[$r] = 2;
    $l[$r] = 0;
    $w[$r] = 0;
    $k[$r] = 0;
    $number[$r] = 0;
    $signif[$r] = 0;
    $pattern[$r] = "";
    $sequences[$r] = "";
    $positions[$r] = "";
    $warn[$r] = "";
}
print ERROR "cleared pattern arrays: @l\n";

close ERROR;
close OUTFILE;
close IN;
close PATTERNS;
}}
close OUT;
die "End of program";

```

B.4: COUNTSEQPATTERNS.PL

```

#!/usr/bin/perl -w

#@array1 = qw([RKH][RKH][RKH] [RKH][RKH].[RKH]
[RKH].[RKH][RKH] [RKH][RKH]..[RKH] [RKH].[RKH].[RKH]
[RKH]..[RKH][RKH] [RKH][RKH]...[RKH]
[RKH]...[RKH][RKH] [RKH]..[RKH].[RKH]
[RKH].[RKH]..[RKH]);
#@array2 = qw(3 4 4 5 5 5 6 6 6 6);

```

```

#@array1 = qw(PW[RKH] P.W[RKH] PW.[RKH] P..W[RKH]
PW..[RKH] P.W.[RKH] P...W[RKH] PW...[RKH] P..W.[RKH]
P.W..[RKH]);
#@array2 = qw(3 4 4 5 5 5 6 6 6 6);

#@array1 = qw([WF][RKH][RKH] [WF].[RKH][RKH]
[WF][RKH].[RKH] [WF]..[RKH][RKH] [WF][RKH]..[RKH]
[WF].[RKH].[RKH] [WF]..[RKH].[RKH] [WF].[RKH]..[RKH]
[WF]...[RKH][RKH] [WF][RKH]...[RKH]);
#@array2 = qw(3 4 4 5 5 5 6 6 6 6);

#@array1 = qw(P[RKH][RKH] P.[RKH][RKH] P[RKH].[RKH]
P..[RKH][RKH] P[RKH]..[RKH] P.[RKH].[RKH]
P..[RKH].[RKH] P.[RKH]..[RKH] P...[RKH][RKH]
P[RKH]...[RKH]);
#@array2 = qw(3 4 4 5 5 5 6 6 6 6);

#@array1 = qw([LAIVG]P[RKH] [LAIVG].P[RKH]
[LAIVG]P.[RKH] [LAIVG]..P[RKH] [LAIVG]P..[RKH]
[LAIVG].P.[RKH] [LAIVG]...P[RKH] [LAIVG]P...[RKH]
[LAIVG]..P.[RKH] [LAIVG].P..[RKH]);
#@array2 = qw(3 4 4 5 5 5 6 6 6 6);

#@array1 = qw([LAIVG]P[WF] [LAIVG].P[WF] [LAIVG]P.[WF]
[LAIVG]..P[WF] [LAIVG]P..[WF] [LAIVG].P.[WF]
[LAIVG]...P[WF] [LAIVG]P...[WF] [LAIVG]..P.[WF]
[LAIVG].P..[WF]);
#@array2 = qw(3 4 4 5 5 5 6 6 6 6);

@array1 = qw(PP P.P P..P P...P P....P P.....P P.....P
P.....P P.....P P.....P P.....P P.....P);
@array2 = qw(2 3 4 5 6 7 8 9 10 11 12);

open (OUT, ">/home/Erin
Gooch/database/P_P_eq_patts.xls") || die "Can't open
out file: $!";
#open (OUT2, ">/home/Erin
Gooch/database/fasta_PWBsc.txt") || die "Can't open
out file: $!";
open (DATABASE, "/home/Erin Gooch/database/NEWDB.txt")
|| die "Can't open database: $!";

```

```

$/ = "\r\n";
while ($line = <DATABASE>){
    chomp $line;
    @line = split (/:/, $line);
    $name = $line[0];
    $des = $line[3];
    $sequence = $line[7];
    $library = $line[5];
    $no_amp =
"$line[1]".$line[2]".$line[3]".$line[5]".$line[6]"
".$line[7]";
    $doubles{$no_amp} += 1;

    if ($doubles{$no_amp} <= 1) {
#        print OUT
"$name\t$no_amp\t$doubles{$no_amp}\n";
        #loop around find seq

        if ($des =~ /001|nacritic/ ) {

#print OUT2 ">$name\n$sequence\n";
            for ($x = 0; $x <= $#array1; $x++) {
#                print OUT "$array1[$x]\n";
                $length = $array2[$x];
                if ($library =~ /PhD12/) {$a = 0; $b =
13 - $length;}
                elsif ($library =~ /PhD7\b/) {$a = 0;
$b = 8 - $length;}
                elsif ($library =~ /PhD7C\b/) {$a = 1;
$b = 9 - $length;}
                for ($c = $a; $c < $b; $c++) {
                    $shortseq = substr($sequence, $c,
$length);
                    if ($shortseq =~ /$array1[$x]/) {
                        print OUT
"$name\t$des\t$sequence\t$library\t$array1[$x]\n";
#                            print OUT "$shortseq\n";
                    }
                }
            }
        }
    }
}

```

```

    }
}
close DATABASE;
close OUT;
#close OUT2;

```

B.5: PARSENAcrePROTEINS.PL

```

#!/usr/bin/perl -w

@array2 = qw(PP);

@array3 = qw([RKH][RKH][RKH] PW[RKH] [WF][RKH][RKH]
P[RKH][RKH] [LAIVG]P[RKH] [LAIVG]P[WF] P.P);

@array4 = qw([RKH][RKH].[RKH] [RKH].[RKH][RKH]
P.[WF][RKH] P[WF].[RKH] [WF].[RKH][RKH]
[WF][RKH].[RKH] P.[RKH][RKH] P[RKH].[RKH]
[LAIVG].P[RKH] [LAIVG]P.[RKH] [LAIVG].P[WF]
[LAIVG]P.[WF] P..P);

@array5 = qw([RKH][RKH]..[RKH] [RKH].[RKH].[RKH]
[RKH]..[RKH][RKH] P..[WF][RKH] P[WF]..[RKH]
P.[WF].[RKH] [WF]..[RKH][RKH] [WF][RKH]..[RKH]
[WF].[RKH].[RKH] P..[RKH][RKH] P[RKH]..[RKH]
P.[RKH].[RKH] [LAIVG]..P[RKH] [LAIVG]P..[RKH]
[LAIVG].P.[RKH] [LAIVG]..P[WF] [LAIVG]P..[WF]
[LAIVG].P.[WF] P...P);

@array6 = qw([RKH][RKH]...[RKH] [RKH]...[RKH][RKH]
[RKH]..[RKH].[RKH] [RKH].[RKH]..[RKH] P...[WF][RKH]
P[WF]...[RKH] P..[WF].[RKH] P.[WF]..[RKH]
[WF]..[RKH].[RKH] [WF].[RKH]..[RKH] [WF]...[RKH][RKH]
[WF][RKH]...[RKH] P..[RKH].[RKH] P.[RKH]..[RKH]
P...[RKH][RKH] P[RKH]...[RKH] [LAIVG]...P[RKH]
[LAIVG]P...[RKH] [LAIVG]..P.[RKH] [LAIVG].P..[RKH]
[LAIVG]...P[WF] [LAIVG]P...[WF] [LAIVG]..P.[WF]
[LAIVG].P..[WF] P....P);

```



```

open (OUT, ">/home/Erin
Gooch/database/Nacre_Protein_Patterns2.xls") || die
"Can't open out file: $!";
open (FASTA, "/home/Erin
Gooch/database/Nacre_Proteins.txt") || die "Can't open
protein file: $!";
$/ = "\r\n";

```

```

$index = 0;
while ($line = <FASTA>){
  chomp $line;
  if ($line =~ /^>/) {
    $name[$index] = $line;
    $seq = <FASTA>;
    chomp $seq;
    $sequence[$index] = $seq;

    for ($x = 0; $x <= $#array6; $x++) {
      $b = length($sequence[$index])-5;
      for ($c = 0; $c <= $b; $c++) {
        $short6 = substr($sequence[$index], $c,
6);
        if ($short6 =~ /$array6[$x]/) {print OUT
"$name[$index]\t$c\t$short6\t$array6[$x]\n";}
      }
    }
    for ($x = 0; $x <= $#array5; $x++) {
      $b = length($sequence[$index])-4;
      for ($c = 0; $c <= $b; $c++) {
        $short5 = substr($sequence[$index], $c,
5);
        if ($short5 =~ /$array5[$x]/) {print OUT
"$name[$index]\t$c\t$short5\t$array5[$x]\n";}
      }
    }
    for ($x = 0; $x <= $#array4; $x++) {
      $b = length($sequence[$index])-3;
      for ($c = 0; $c <= $b; $c++) {
        $short4 = substr($sequence[$index], $c,
4);

```

```

        if ($short4 =~ /$array4[$x]/) {print OUT
"$name[$index]\t$c\t$short4\t$array4[$x]\n";}
    }
}
for ($x = 0; $x <= $#array3; $x++) {
    $b = length($sequence[$index])-2;
    for ($c = 0; $c <= $b; $c++) {
        $short3 = substr($sequence[$index], $c,
3);
        if ($short3 =~ /$array3[$x]/) {print OUT
"$name[$index]\t$c\t$short3\t$array3[$x]\n";}
    }
}
for ($x = 0; $x <= $#array2; $x++) {
    $b = length($sequence[$index])-1;
    for ($c = 0; $c <= $b; $c++) {
        $short2 = substr($sequence[$index], $c,
2);
        if ($short2 =~ /$array2[$x]/) {print OUT
"$name[$index]\t$c\t$short2\t$array2[$x]\n";}
    }
}
    $index++;
}
close FASTA;
close OUT;
die "End of program\n";

```

Appendix C: NMR Spectra of the A20 peptide

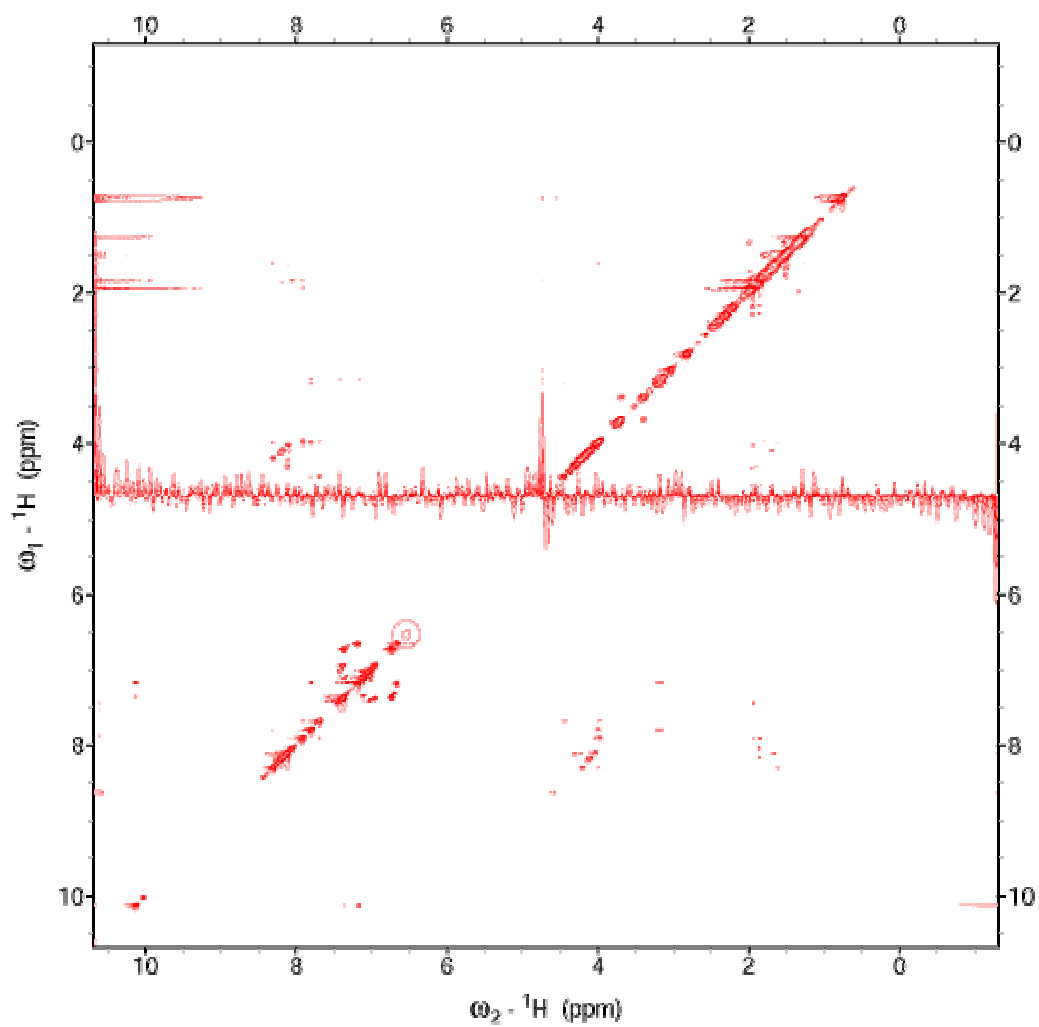


Figure C.1: 2-dimensional Nuclear Overhauser Effect spectrum (NOESY) of the A20 peptide.

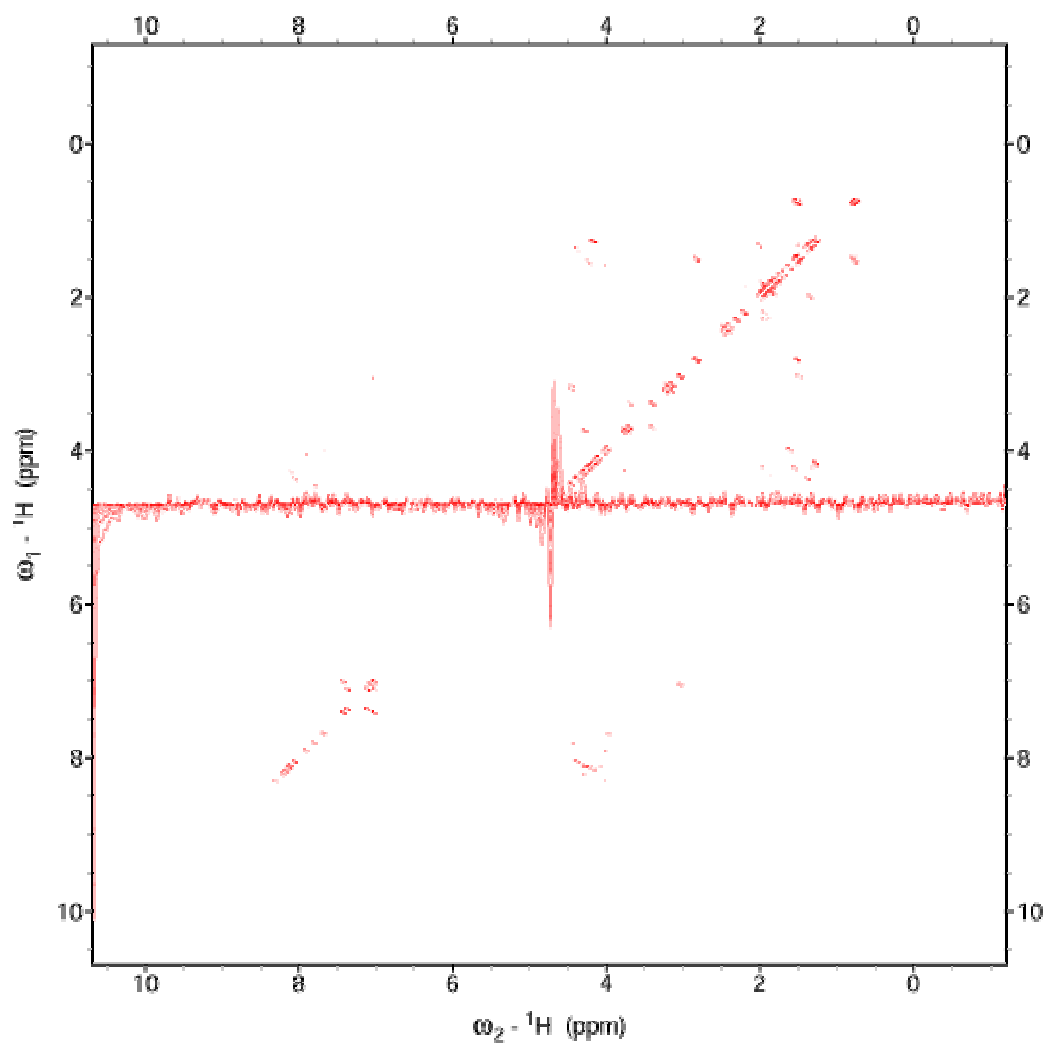


Figure C.2: 2-dimensional Correlation Spectrum (COSY) of the A20 peptide.

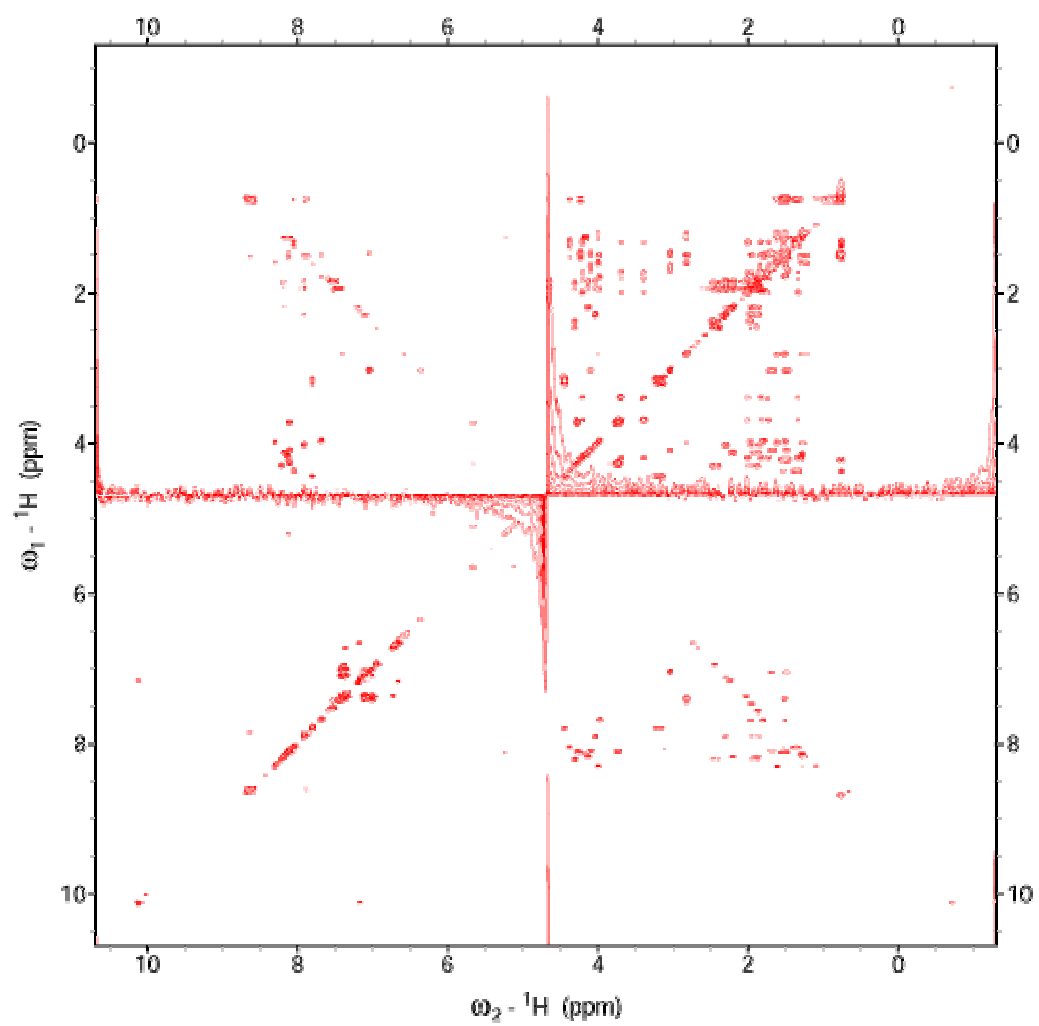


Figure C.3: 2-dimensional Total Correlated Spectrum (TOCSY) of the A20 peptide.

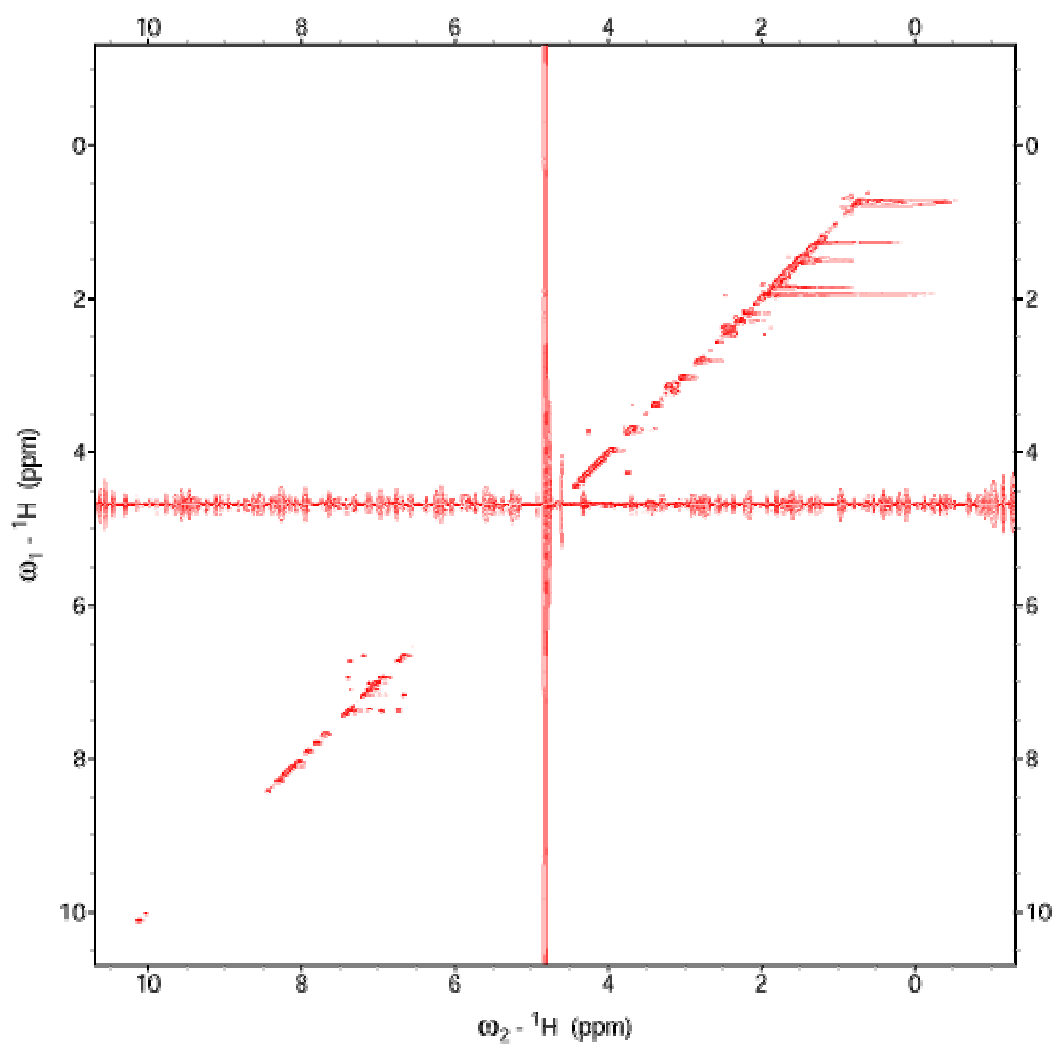


Figure C.4: Rotating-frame Overhauser Enhancement Spectrum (ROESY) of the A20 peptide.

Bibliography

- Addadi, L., et al., *A chemical model for the cooperation of sulfates and carboxylates in calcite crystal nucleation: relevance to biomineralization*. Proc. Natl. Acad. Sci. U. S. A., 1987. **84**(9): p. 2732-6.
- Ahn, E.S., et al., *Nanostructure processing of hydroxyapatite-based bioceramics*. Nano Letters, 2001. **1**(3): p. 149-153.
- Aizenberg, J., A.J. Black, and G.M. Whitesides, *Control of crystal nucleation by patterned self-assembled monolayers*. Nature, 1999. **398**(6727): p. 495-498.
- Aizenberg, J., *Patterned crystallization of calcite in vivo and in vitro*. Journal of Crystal Growth, 2000. **211**(1-4): p. 143-148.
- Aizenberg, J., A.J. Black, and G.H. Whitesides, *Oriented growth of calcite controlled by self-assembled monolayers of functionalized alkanethiols supported on gold and silver*. Journal of the American Chemical Society, 1999. **121**(18): p. 4500-4509.
- Aizenberg, J., et al., *Direct fabrication of large micropatterned single crystals*. Science, 2003. **299**(5610): p. 1205-1208.
- Aizenberg, J., *Patterned crystallisation on self-assembled monolayers with integrated regions of disorder*. Journal of the Chemical Society-Dalton Transactions, 2000(21): p. 3963-3968.
- Almeida, M.J., et al., *Comparative effects of nacre water-soluble matrix and dexamethasone on the alkaline phosphatase activity of MRC-5 fibroblasts*. Journal of Biomedical Materials Research, 2001. **57**(2): p. 306-312.
- Bazylinski, D.A. and R.B. Frankel, *Magnetic Iron Oxide and Iron Sulfide Minerals within Microorganisms*, in *Biomineralization: From Biology to Biotechnology and Medical Application*, E. Baeuerlein, Editor. 2000, Wiley-VCH: Weinheim. p. 25-46.

- Belcher, A.M., et al., *Control of crystal phase switching and orientation by soluble mollusk-shell proteins*. Nature (London), 1996. **381**(6577): p. 56-58.
- Belcher, A.M. and E.E. Gooch, Protein components and inorganic structure in shell nacre. *Biomaterialization*, 2000: p. 221-249.
- Belcher, A.M., *Personal communication: Inorganic film growth using Abalone nacre proteins*. 1997: Santa Barbara, CA.
- Blank, S., et al., *The nacre protein perlucin nucleates growth of calcium carbonate crystals*. Journal of Microscopy-Oxford, 2003. **212**: p. 280-291.
- Bragg, W.L., *The analysis of crystals by the x-ray spectrometer*. Proceedings of the Royal Society of London, 1914(A89): p. 468-489.
- Bragg, W.L., *The Structure of Aragonite*. Proceedings of the Royal Society of London, 1924. **105**: p. 16-39.
- Bragg, W.L., *Atomic Structure of Minerals*. 1937, Oxford, England: Oxford University Press.
- Cha, J.N., et al., *Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones in vitro*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(2): p. 361-365.
- Cha, J.N., et al., *Biomimetic synthesis of ordered silica structures mediated by block copolypeptides*. Nature, 2000. **403**(6767): p. 289-292.
- Chakraborty, J., et al., *Behaviour of iron ion in the morphosynthesis of magnetite particles*. Materials Transactions, 2003. **44**(6): p. 1124-1127.
- Chen, S.Y. and I.W. Chen, *Cracking During Pyrolysis of Oxide Thin-Films - Phenomenology, Mechanisms, and Mechanics*. Journal of the American Ceramic Society, 1995. **78**(11): p. 2929-2939.
- Chien, A.T., et al., *Low-Temperature Low-Pressure Hydrothermal Synthesis of Barium-Titanate - Powder and Heteroepitaxial Thin-Films*. Journal of Materials Research, 1995. **10**(7): p. 1784-1789.

- Cerini, C., et al., *Nucleation of calcium oxalate crystals by albumin: Involvement in the prevention of stone formation*. *Kidney International*, 1999. **55**(5): p. 1776-1786.
- Click, E.M. and R.E. Webster, Filamentous phage infection: required interactions with the TolA protein. *Journal of Bacteriology*, 1997. **179**(20): p. 6464-6471.
- Click, E.M. and R.E. Webster, The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage ϕ 1 during infection. *Journal of Bacteriology*, 1998. **180**(7): p. 1723-1728.
- Cross, L.E., *Dielectric, Piezoelectric, and Ferroelectric Components*. *American Ceramic Society Bulletin*, 1984. **63**(4): p. 586-590.
- Dal Negro, A. and L. Ungaretti, *Refinement of the crystal structure of aragonite*. *American Mineralogist*, 1971. **56**: p. 768-772.
- Deng, L.W. and R.N. Perham, Delineating the site of interaction on the pIII protein of filamentous bacteriophage ϕ d with the F-pilus of *Escherichia coli*. *Journal of Molecular Biology*, 2002. **319**(3): p. 603-614.
- DeVilliers, J.P.R., *Crystal structures of aragonite, strontianite and witherite*. *American Mineralogist*, 1971. **56**: p. 758-772.
- DeOliveira, D.B. and R.A. Laursen, *Control of Calcite Crystal Morphology by a Peptide Designed To Bind to a Specific Surface*. *Journal of the American Chemical Society*, 1997. **119**(44): p. 10627-10631.
- Dickens, B. and J.S. Bowen, *Refinement of the crystal structure of the aragonite phase of CaCO₃*. *Journal of Research National Standards A. Phys. Chem.*, 1971. **75A**: p. 27-32.
- Didymus, J.M., et al., *Influence of low-molecular-weight and macromolecular organic additives on the morphology of calcium carbonate*. *Journal of the Chemical Society*, 1993. **Faraday Transactions** **89**(15): p. 2891-900.
- Eysel, W. and D.M. Roy, *Topotactic Reaction of Aragonite to Hydroxyapatite*. *Zeitschrift Fur Kristallographie*, 1975. **141**(1-2): p. 11-24.
- Falini, G., et al., *Control of aragonite or calcite polymorphism by mollusk shell macromolecules*. *Science*, 1996. **271**(5245): p. 67-69.

- Ferguson, A.D., et al., *Siderophore-Mediated Iron Transport: Crystal Structure of FhuA with Bound Lipopolysaccharide*. *Science*, 1998. **282**(5397): p. 2215-2220.
- Flynn, C.E., et al., *Synthesis and organization of nanoscale II-VI semiconductor materials using evolved peptide specificity and viral capsid assembly*. *Journal of Materials Chemistry*, 2003. **13**(10): p. 2414-2421.
- Flytzanis, C., et al., *Nonlinear Optics in Composite-Materials*. *Journal of the Optical Society of America B-Optical Physics*, 1986. **3**(8): p. P93-&.
- Fritz, M., et al., *Flat pearls from biofabrication of organized composites on inorganic substrates*. *Nature (London, United Kingdom)*, 1994. **371**(6492): p. 49-51.
- Frost, L.S., K. Ippenihler, and R.A. Skurray, *Analysis of the Sequence and Gene-Products of the Transfer Region of the F-Sex Factor*. *Microbiological Reviews*, 1994. **58**(2): p. 162-210.
- Gerbaud, V., et al., *Mechanism of calcite crystal growth inhibition by the N-terminal undecapeptide of lithostathine*. *Journal of Biological Chemistry*, 2000. **275**(2): p. 1057-1064.
- Glucksman, M.J., S. Bhattacharjee, and L. Makowski, *3-Dimensional Structure of a Cloning Vector - X-Ray-Diffraction Studies of Filamentous Bacteriophage-M13 at 7-Angstrom Resolution*. *Journal of Molecular Biology*, 1992. **226**(2): p. 455-470.
- Grant, R.A., et al., *Structure of the Filamentous Bacteriophage-F1 - Location of the Minor Coat Protein-a, Protein-C and Protein-D*. *Journal of Biological Chemistry*, 1981. **256**(1): p. 539-546.
- Gray, C.W., *3-Dimensional Structure of Complexes of Single-Stranded DNA-Binding Proteins with DNA - Ike and Fd Gene-5 Proteins Form Left-Handed Helices with Single-Stranded-DNA*. *Journal of Molecular Biology*, 1989. **208**(1): p. 57-64.
- Greenwood, J., A.E. Willis, and R.N. Perham, *Multiple Display of Foreign Peptides on a Filamentous Bacteriophage - Peptides from Plasmodium-Falciparum Circumsporozoite Protein as Antigens*. *Journal of Molecular Biology*, 1991. **220**(4): p. 821-827.

- Greenwood, J., G.J. Hunter, and R.N. Perham, *Regulation of Filamentous Bacteriophage Length by Modification of Electrostatic Interactions between Coat Protein and DNA*. Journal of Molecular Biology, 1991. **217**(2): p. 223-227.
- Gunter, P. and J.P. Huignard, *Photorefractive Materials and Their Applications .1. Introduction*. Topics in Applied Physics, 1988. **61**: p. 1-5.
- Hache, F., et al., *The Optical Kerr Effect in Small Metal Particles and Metal Colloids - the Case of Gold*. Applied Physics a-Materials Science & Processing, 1988. **47**(4): p. 347-357.
- Hayashi, N.a.S., T., *N14#2 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
- Hayashi, N.a.S., T., *N14#3 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
- Hayashi, N.a.S., T., *N14#4 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
- Hayashi, N.a.S., T., *N14#5 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
- Hayashi, N.a.S., T., *N14#7 - 14 kD matrix protein family in nacreous layer of Japanese peal oyster, Pinctada fucata*. 1999, Entrez.
- Hesselink, L., et al., *Photorefractive materials for nonvolatile volume holographic data storage*. Science, 1998. **282**(5391): p. 1089-1094.
- Holliger, P. and L. Riechmann, A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure, 1997. **5**(2): p. 265-275.
- Holliger, P., L. Riechmann, and R.L. Williams, Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 angstrom: Evidence for conformational lability. Journal of Molecular Biology, 1999. **288**(4): p. 649-657.

<http://www.theimage.com/mineral/aragonite/aragonite.jpg>.

- Hu, E.L., *Personal communication: Quantum dot amplifiers for lithium niobate waveguides*. 2002: Santa Barbara, Ca.
- Huang, H., et al., *Ferroelectric PbTiO₃ thin films prepared by electrostatic spray deposition (ESD)*. *Microelectronic Engineering*, 2003. **66**(1-4): p. 688-694.
- Ioku, K., M. Yoshimura, and S. Somiya, *Microstructure and Mechanical-Properties of Hydroxyapatite Ceramics with Zirconia Dispersion Prepared by Post-Sintering*. *Biomaterials*, 1990. **11**(1): p. 57-61.
- Ishihara, T., et al., *Mixed-Oxide Capacitor of Cuo-Batio₃ as a New Type Co₂ Gas Sensor*. *Journal of the American Ceramic Society*, 1992. **75**(3): p. 613-618.
- Jackson, A.P., J.F.V. Vincent, and R.M. Turner, *Comparison of nacre with other ceramic composites*. *Journal of Materials Science*, 1990. **25**(7): p. 3173-8.
- Jacobson, A., *Role of F pili in the penetration of bacteriophage fl*. *Journal of Virology*, 1972. 10: p. 835-843.
- Kay, B.K, J. Winter, and J. McCafferty, 1996. *Phage Display of Peptides and Proteins: A Laboratory Manual*, ed., San Diego: Academic Press, Inc.
- Kazmierczak, B.I., et al., *Piv, a Filamentous Phage Protein That Mediates Phage Export across the Bacterial-Cell Envelope, Forms a Multimer*. *Journal of Molecular Biology*, 1994. 238(2): p. 187-198.
- Kono, M., N. Hayashi, and T. Samata, *Molecular mechanism of the nacreous layer formation in Pinctada maxima*. *Biochemical and Biophysical Research Communications*, 2000. **269**(1): p. 213-218.
- Kotecki, D.E., et al., *(Ba,Sr)TiO₃ dielectrics for future stacked-capacitor DRAM*. *Ibm Journal of Research and Development*, 1999. **43**(3): p. 367-382.....
- Kroger, N. and M. Sumper, *Diatom cell wall proteins and the cell biology of silica biomineralization*. *Protist*, 1998. **149**(3): p. 213-219.
- Kroger, N., R. Deutzmann, and M. Sumper, *Polycationic peptides from diatom biosilica that direct silica nanosphere formation*. *Science*, 1999. **286**(5442): p. 1129-1132.

- Kroger, N. and M. Sumper, *The Biochemistry of Silica Formation in Diatoms, in Biomineralization*, E. Baeuerlein, Editor. 2000, Wiley-VCH: Weinheim. p. 151-170.
- Kwok, C.K. and S.B. Desu, *Low-Temperature Perovskite Formation of Lead Zirconate Titanate Thin-Films by a Seeding Process*. Journal of Materials Research, 1993. **8**(2): p. 339-344.
- Lee, S.-W., et al., *Ordering of quantum dots using genetically engineered viruses*. Science (Washington, DC, United States), 2002. **296**(5569): p. 892-895.
- Lee, S.-W., B.M. Wood, and A.M. Belcher, *Chiral Smectic C Structures of Virus-Based Films*. Langmuir, 2003. **19**(5): p. 1592-1598.
- Levengood, S.K., W.F. Beyer, Jr., and R.E. Webster, TolA: a membrane protein involved in colicin uptake contains an extended helical region. Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(14): p. 5939-43.
- Levi, Y., et al., *Control over aragonite crystal nucleation and growth: an in vitro study of biomineralization*. Chemistry--A European Journal, 1998. **4**(3): p. 389-396.
- Li, C.M., G.D. Botsaris, and D.L. Kaplan, *Selective in vitro effect of peptides on calcium carbonate crystallization*. Crystal Growth & Design, 2002. **2**(5): p. 387-393.
- Li, J., B. Fartash, and L. Hermansson, *Hydroxyapatite Alumina Composites and Bone-Bonding*. Biomaterials, 1995. **16**(5): p. 417-422.
- Liao, H.B., et al., *Large third-order optical nonlinearity in Au:SiO₂ composite films near the percolation threshold*. Applied Physics Letters, 1997. **70**(1): p. 1-3.
- Lin, T.C., R.E. Webster, and W. Konigsberg, Isolation and Characterization of the C-Protein and D-Protein Coded by Gene-Ix and Gene-Vi in the Filamentous Bacteriophage-FI and Bacteriophage-Fd. Journal of Biological Chemistry, 1980. **255**(21): p. 331-337.
- Linderoth, N.A., M.N. Simon, and M. Russel, The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. Science, 1997. **278**(5343): p. 1635-1638.

- Lippmann, F., *Sedimentary Carbonate Minerals*. 1973, New York: Springer-Verlag.
- Liu, Y., et al., *Third-order nonlinear optical response of Au-core CdS-shell composite nanoparticles embedded in BaTiO₃ thin films*. Applied Physics B-Lasers and Optics, 2003. **76**(4): p. 435-439.
- Lopez, E., et al., *Demonstration of the Capacity of Nacre to Induce Bone-Formation by Human Osteoblasts Maintained Invitro*. Tissue & Cell, 1992. **24**(5): p. 667-679.
- Lopez, J. and R.E. Webster, Morphogenesis of Filamentous Bacteriophage-F1 - Orientation of Extrusion and Production of Polyphage. Virology, 1983. **127**(1): p. 177-193.
- Lowenstam, H.A. and S. Weiner. 1989, ed., *On Biomineralization*, Oxford: Oxford University Press.
- Lubkowski, J., et al., The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nature Structural Biology, 1998. **5**(2): p. 140-147.
- Lubkowski, J., et al., Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure with Folding & Design, 1999. **7**(6): p. 711-722.
- MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
- Manne, S., et al., *Atomic Force Microscopy of the Nacreous Layer in Mollusc Shells*. Proceedings: Biological Sciences, 1994. **256**(1345): p. 17-23.
- Mao, C.B., et al., *Viral assembly of oriented quantum dot nanowires*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(12): p. 6946-6951.
- Marin, F., et al., *Mucins and molluscan calcification - Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel Pinna nobilis (Bivalvia, Pteriomorpha)*. Journal of Biological Chemistry, 2000. **275**(27): p. 20667-20675.

- Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually Stable Helix Formation in Short Alanine-Based Peptides*. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(14): p. 5286-5290.
- Martinez, G. and G. Millhauser, *Ftir Spectroscopy of Alanine-Based Peptides - Assignment of the Amide I' Modes for Random Coil and Helix*. Journal of Structural Biology, 1995. **114**(1): p. 23-27.
- Martz, E., Protein Explorer: Easy Yet Powerful Macromolecular Visualization. Trends in Biochemical Sciences, 2002. 27(February): p. 107-109 <http://proteinexplorer.org>.
- Marvin, D.A., et al., *Molecular-Models and Structural Comparisons of Native and Mutant Class-I Filamentous Bacteriophages Ff (Fd, F1, M13), If1 and Ike*. Journal of Molecular Biology, 1994. **235**(1): p. 260-286.
- Marvin, D.A. and W. Folkhard, Structure of F-Pili - Reassessment of the Symmetry. Journal of Molecular Biology, 1986. 191(2): p. 299-300.
- McConnell, D., Journal of Dental Research, 1952. **31**: p. 53.
- Michenfelder, M., et al., *Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains*. Biopolymers, 2003. **70**(4): p. 522-533.
- Miyashita, T., et al., *Complementary DNA cloning and characterization of pearlín, a new class of matrix protein in the nacreous layer of oyster pearls*. Marine Biotechnology, 2000. **2**(5): p. 409-418.
- Miyamoto, H., et al., *A carbonic anhydrase from the nacreous layer in oyster pearls*. Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(18): p. 9657-9660.
- Mouries, L.P., et al., *Bioactivity of nacre water-soluble organic matrix from the bivalve mollusk Pinctada maxima in three mammalian cell types: fibroblasts, bone marrow stromal cells and osteoblasts*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 2002. **132**(1): p. 217-229.
- NCBI, The Universal Virus Database of the International Committee on Taxonomy of Viruses: Index of Viruses from http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fs_inovi.htm. 2002.

- Newnham, R.E., *Electroceramics*. Reports on Progress in Physics, 1989. **52**(2): p. 123-156.
- Ohgushi, H., et al., *Bone-Formation Process in Porous Calcium-Carbonate and Hydroxyapatite*. Journal of Biomedical Materials Research, 1992. **26**(7): p. 885-895.
- Ojard, D.,
http://www.d.umn.edu/geology/pagnucco/Carbonates/smallpics/Calcite-2881_small.jpg.
- Papavoine, C.H.M., et al., *Solution structure of the M13 major coat protein in detergent micelles: A basis for a model of phage assembly involving specific residues*. Journal of Molecular Biology, 1998. **282**(2): p. 401-419.
- Perham, R.N., et al., *Engineering a Peptide Epitope Display System on Filamentous Bacteriophage*. Fems Microbiology Reviews, 1995. **17**(1-2): p. 25-31.
- Ph.D.-12™, Ph.D.-7™, Ph.D.-C7C™ Phage Display Peptide Library Kit Instruction Manuals*: New England Biolabs.
- Posner, A.S., N.C. Blumenthal, and F. Betts, in *Phosphate Minerals*, L.O. Nriagu and P.B. Moore, Editors. 1984, Springer-Verlag: Berlin; New York.
- Rakonjac, J., J.-n. Feng, and P. Model, Filamentous Phage are Released from the Bacterial Membrane by a Two-step Mechanism Involving a Short C-terminal Fragment of pIII,. Journal of Molecular Biology, 1999. **289**(5): p. 1253-1265.
- Rakonjac, J. and P. Model, Roles of pIII in Filamentous Phage Assembly,. Journal of Molecular Biology, 1998. **282**(1): p. 25-41.
- Reeder, R.J, 1983, *Carbonates: Mineralogy and Chemistry*. Reviews in Mineralogy, ed... Vol. 11.
- Ren, T.L., et al., *Piezoelectric and ferroelectric films for microelectronic applications*. Materials Science and Engineering B-Solid State Materials for Advanced Technology, 2003. **99**(1-3): p. 159-163.

- Ricard, D., P. Roussignol, and C. Flytzanis, *Surface-Mediated Enhancement of Optical-Phase Conjugation in Metal Colloids*. Optics Letters, 1985. **10**(10): p. 511-513.
- Richardson, J.S. and D.C. Richardson, *Amino-Acid Preferences for Specific Locations at the Ends of Alpha-Helices*. Science, 1988. **240**(4859): p. 1648-1652.
- Riechmann, L. and P. Holliger, The C-terminal domain of ToIA is the coreceptor for filamentous phage infection of E-coli. Cell, 1997. **90**(2): p. 351-360.
- Rigoutsos, I. and A. Floratos, Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm (vol 14, pg 55, 1998). Bioinformatics, 1998. **14**(2): p. 229-229.
- Rohl, C.A. and R.L. Baldwin, *Exchange Kinetics of Individual Amide Protons in N-15-Labeled Helical Peptides Measured by Isotope-Edited Nmr*. Biochemistry, 1994. **33**(25): p. 7760-7767.
- Rohl, C.A., W. Fiori, and R.L. Baldwin, *Alanine is helix-stabilizing in both template-nucleated and standard peptide helices*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(7): p. 3682-3687.
- Rousseau, M., et al., *The water-soluble matrix fraction from the nacre of Pinctada maxima produces earlier mineralization of MC3T3-E1 mouse pre-osteoblasts*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology, 2003. **135**(1): p. 1-7.
- Roy, D.M. and S.K. Linnehan, *Hydroxyapatite formed from coral skeletal carbonate by hydrothermal exchange*. Nature (London, 1974. **United Kingdom**) **247**(5438): p. 220-2.
- Russel, M. and P. Model, Genetic-Analysis of the Filamentous Bacteriophage Packaging Signal and of the Proteins That Interact with It. Journal of Virology, 1989. **63**(8): p. 3284-3295.
- Samata, T., et al., *A new matrix protein family related to the nacreous layer formation of Pinctada fucata*. Febs Letters, 1999. **462**(1-2): p. 225-229.

- Samuelson, S.O. and G.J. Martyna, *Two dimensional umbrella sampling techniques for the computer simulation study of helical peptides at thermal equilibrium: The 3K(I) peptide in vacuo and solution*. Journal of Chemical Physics, 1998. **109**(24): p. 11061-11073.
- Schaeffer, T.E., et al., *Does Abalone Nacre Form by Heteroepitaxial Nucleation or by Growth through Mineral Bridges?* Chemistry of Materials, 1997. **9**(8): p. 1731-1740.
- Selvig, K.A., *Periodic lattice images of hydroxyapatite crystals in human bone and dental hard tissues*. Calcified Tissue Research, 1970. **6**: p. 227-238.
- Sharma, H.B. and A. Mansingh, *Phase transition in sol-gel-derived barium titanate thin films*. Journal of Physics D-Applied Physics, 1998. **31**(13): p. 1527-1533.
- Shen, X.Y., et al., *Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of Haliotis rufescens*. Journal of Biological Chemistry, 1997. **272**(51): p. 32472-32481.
- Shenton, W., et al., *Synthesis of nanophase iron oxide in lumazine synthase capsids*. Angewandte Chemie-International Edition, 2001. **40**(2): p. 442-445.
- Shimizu, K., et al., *Silicatein alpha: Cathepsin L-like protein in sponge biosilica*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(11): p. 6234-6238.
- Smith, G.P., *Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface*. Science, 1985. **228**(4705): p. 1315-1317.
- Suchanek, W. and M. Yoshimura, *Processing and properties of hydroxyapatite-based biomaterials for use as hard tissue replacement implants*. Journal of Materials Research, 1998. **13**(1): p. 94-117.
- Sudo, S., et al., *Structures of mollusc shell framework proteins*. Nature, 1997. **387**(6633): p. 563-564.
- Szabo, Z., *Polyproline helices*. 1997.

- Tang, L., et al., *Preparation of silver dispersed PbTiO₃ film by sol-gel method*. Materials Science and Engineering B, 2003. **99**(1-3): p. 370-373.
- Tokizaki, T., et al., *Subpicosecond Time Response of 3rd-Order Optical Nonlinearity of Small Copper Particles in Glass*. Applied Physics Letters, 1994. **65**(8): p. 941-943.
- Traub, W., T. Arad, and S. Weiner, *3-Dimensional Ordered Distribution of Crystals in Turkey Tendon Collagen-Fibers*. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(24): p. 9822-9826.
- Trenkner, E., F. Bonhoeffer, and A. Gierer, The fate of the protein component of bacteriophage fd during infection. Biochemical and Biophysical Research Communications, 1967. 28: p. 932-939.
- Tuckerman, M.E., et al., *Exploiting multiple levels of parallelism in Molecular Dynamics based calculations via modern techniques and software paradigms on distributed memory computers*. Computer Physics Communications, 2000. **128**(1-2): p. 333-376.
- Weiner, S. and W. Traub, *Macromolecules in Mollusk Shells and Their Functions in Biomineralization*. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 1984. **304**(1121): p. 425.
- Weiner, S. and W. Traub, *Bone-Structure - from Angstroms to Microns*. Faseb Journal, 1992. **6**(3): p. 879-885.
- Weiss, I.M., et al., *Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc Haliotis laevis*. Biochemical and Biophysical Research Communications, 2000. **267**(1): p. 17-21.
- Weiss, I.M., et al., *Perlustrin, a Haliotis laevis (abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N-terminal module of vertebrates*. Biochemical and Biophysical Research Communications, 2001. **285**(2): p. 244-249.
- Whaley, S.R., et al., Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. Nature (London), 2000. **405**(6787): p. 665-668.

- White, E. and E.C. Shors, *Biomaterial Aspects of Interpore-200 Porous Hydroxyapatite*. Dental Clinics of North America, 1986. **30**(1): p. 49-67.
- Wustman, B.A., et al., *Characterization of a Ca(II)-, mineral-interactive polyelectrolyte sequence from the adhesive elastomeric biomineralization protein lustrin A*. Langmuir, 2003. **19**(22): p. 9373-9381.
- Wyckoff, R.W.G., *Orthorhombic space group criteria and their application to aragonite*. American Journal of Science, 1925. **5th Series**(1X): p. 145-175.
- Xu, J.J., A.S. Shaikh, and R.W. Vest, *High K Batio3 Films from Metalloorganic Precursors*. Ieee Transactions on Ultrasonics Ferroelectrics and Frequency Control, 1989. **36**(3): p. 307-312.
- Yano, M., Miyashita, T. and Miyamoto, H., *Shell matrix Nacrein, a novel family of carbonic anhydrase, is conserved in bivalve and gastropod*. 2001, Entrez.
- Yeh, Y.C. and T.Y. Tseng, *Humidity-Sensitive Electrical-Properties of Ba_{0.5}sr_{0.5}tio₃ Porous Ceramics*. Journal of Materials Science Letters, 1988. **7**(7): p. 766-768.
- Yoshimura, T., N. Fujimura, and T. Ito, *The initial stage of BaTiO₃ epitaxial films on etched and annealed SrTiO₃ substrates*. Journal of Crystal Growth, 1997. **174**(1-4): p. 790-795.
- Zaremba, C.M., et al., *Critical Transitions in the Biofabrication of Abalone Shells and Flat Pearls*. Chemistry of Materials, 1996. **8**(3): p. 679-90.
- Zarrinpar, A., R. Bhattacharyya, and W. Lim, *The Structure and Function of Proline Recognition Domains*. Science's STKE, 2003.
- Zemann, J., *Zur Stereochemie der Karbonate*. Fortschr. Mineral, 1981. **59**: p. 95-116

Vita

Erin Elaine Gooch was born in Austin, Texas on April 2, 1976 to Larry Wayne Gooch and Cynthia Nicol Gooch. She grew up in both Austin and Darnestown, Maryland. Erin attended The College of William and Mary in Williamsburg, Virginia and graduated with high honors in chemistry in 1998. She then went to pursue her graduate studies in chemistry at the University of Texas at Austin in the Biomolecular Materials Group, headed by Professor Angela Belcher, now at the Massachusetts Institute of Technology. Upon completion of the dissertation, she has accepted a post-doctoral position in the Pathology Department at the Duke University Medical Center with Professor Daniel Kenan.

Permanent address: 12532 Triple Crown Road, Darnestown, Maryland 20878

This dissertation was typed by the author.