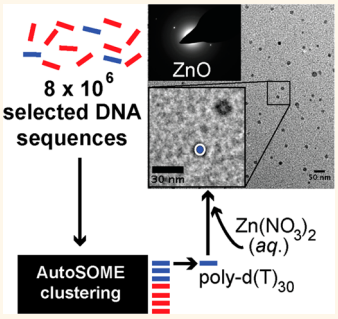
# Efficient Selection of Biomineralizing DNA Aptamers Using Deep Sequencing and Population Clustering

Lukmaan A. Bawazer,[†,§,⊥] Aaron M. Newman,[†,§,⊥] Qian Gu,[†] Abdullah Ibish,[†,⊥] Mary Arcila,[†] James B. Cooper,[†] Fiona C. Meldrum,[‡] and Daniel E. Morse[†,*]

[†]Department of Molecular, Cellular and Developmental Biology, Institute for Collaborative Biotechnologies, and Biomolecular Science and Engineering Program, University of California, Santa Barbara, California 93106, United States, and [‡]School of Chemistry, University of Leeds, Leeds, United Kingdom LS2 9JT. [§]These authors contributed equally to this work. [⊥]Present address: L.A.B, University of Leeds, Leeds, UK, LS1 9JT; A.M.N, Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA, 94305; A.I., Georgetown University School of Medicine, Washington, DC 20057.

**ABSTRACT** DNA-based information systems drive the combinatorial optimization processes of natural evolution, including the evolution of biominerals. Advances in high-throughput DNA sequencing expand the power of DNA as a potential information platform for combinatorial engineering, but many applications remain to be developed due in part to the challenge of handling large amounts of sequence data. Here we employ high-throughput sequencing and a recently developed clustering method (AutoSOME) to identify single-stranded DNA sequence families that bind specifically to ZnO semiconductor mineral surfaces. These sequences were enriched from a diverse DNA library after a single round of screening, whereas previous screening approaches typically require 5−15 rounds of enrichment for effective sequence identification. The consensus sequence of the largest cluster was poly d(T)$_{30}$. This consensus sequence exhibited clear aptamer behavior and was shown to promote the synthesis of crystalline ZnO from aqueous solution at near-neutral pH. This activity is significant, as the crystalline form of this wide-bandgap semiconductor is not typically amenable to solution synthesis in this pH range. High-resolution TEM revealed that this DNA synthesis route yields ZnO nanoparticles with an amorphous−crystalline core−shell structure, suggesting that the mechanism of mineralization involves nanoscale coacervation around the DNA template. We thus demonstrate that our new method, termed Single round Enrichment of Ligands by deep Sequencing (SEL-Seq), can facilitate biomimetic synthesis of technological nanomaterials by accelerating combinatorial selection of biomolecular−mineral interactions. Moreover, by enabling direct characterization of sequence family demographics, we anticipate that SEL-Seq will enhance aptamer discovery in applications employing additional rounds of screening.




**KEYWORDS:** SELEX · massively parallel sequencing · next-generation sequencing · DNA aptamers · zinc oxide · biomimetic mineralization · AutoSOME · population clustering

DNA-based high-throughput screening and *in vitro* evolution encompass an array of biotechnological strategies that mimic molecular aspects of natural evolution. Recently demonstrated as powerful new tools for materials engineering,[1,2] these strategies generate individual DNA sequences,[3] RNA sequences,[4] short polypeptides,[5,6] or full proteins (such as enzymes)[7] with novel activities. An example of such activity is aptamer behavior, in which a short biopolymer (DNA, RNA, or polypeptide) exhibits tight binding affinity to a molecular or mineral target. The method for producing nucleic acid aptamers is generally known as SELEX, for SElection of Ligands by EXponential enrichment.[3,4] SELEX is typically utilized to screen single-stranded (ss) DNA libraries for individual ssDNA aptamer sequences. Toward establishing efficient screening protocols, ssDNA aptamer discovery *via* SELEX is experimentally advantageous relative to other approaches such as phage-display or cell surface-display polypeptide screening or RNA library screening.[2,3] This is because ssDNA aptamers simultaneously exhibit both target binding and information encoding functionality, allowing both genotype and phenotype to be expressed from a single molecule. As a result, bound ssDNA aptamer sequences can be directly enzymatically amplified (after elution from the target), obviating experimental steps needed to link genotype and phenotype that are required in other screening approaches.[2] All *in vitro* evolution strategies,

* Address correspondence to d_morse@lifesci.ucsb.edu.

Received for review August 26, 2013 and accepted December 16, 2013.

**Published online December 16, 2013**
10.1021/nn404448s

including SELEX, rely on the handling of large, diverse libraries of DNA molecules. The molecular population demographics of these libraries change over successive rounds of screening or selection, as families of high activity-encoding DNA sequences become increasingly enriched and prevalent in the library through each successive round of selection. Obtaining sequence information from sampled DNA subpopulations of the screened library is critical for identifying specific subsets of active biomolecules. Until a few years ago, the bottleneck of traditional DNA sequencing methods limited such sampling analyses to a small number of DNA molecules (often <0.01% of the total population[8]), which, in conventional SELEX, would necessitate 5−15 rounds of screening to progressively enrich the library with a sufficiently high proportion of active biomolecules before these active variants could be identified.

Here we employ next-generation DNA sequencing technologies[9−12] and AutoSOME cluster analysis[13] to broadly survey molecular population demographics of a single-stranded (ss) DNA library that has been screened for ZnO-binding activity (Figure 1). This approach, termed Single round Enrichment of Ligands by deep Sequencing (SEL-Seq), allowed us to successfully identify ZnO-binding aptamers after a single round of screening. The largest clustered sequence family we thus identified is rich in thymine content and its consensus sequence, poly d(thymidine), behaves as an aptamer and also facilitates ZnO mineralization from a hydrated $Zn(NO_3)_2$ precursor in a near-neutral aqueous environment. While similar next-generation DNA sequencing approaches have recently been demonstrated to improve SELEX efficiency for isolating aptamers against protein targets,[14,15] the present report is the first to describe the use of high-throughput sequencing to accelerate selection of biomolecular-mineral interactions; it also represents the first time a population clustering algorithm has been used to discover new aptamers based on consensus sequences of clustered oligonucleotides from a SELEX-enriched synthetic gene pool.

To reveal population diversity in next-generation sequencing data, a practical yet accurate unsupervised computational strategy is needed. One method, commonly employed in comparative genomics studies, involves computing all possible pairwise alignments within the sequence population.[16] Although sensitive and specific, such an approach is intractable with very large data sets.[16] More sophisticated strategies have employed *seed* sequences to efficiently identify clusters (*e.g.*, refs 17 and 18), or probabilistic/statistical techniques to detect over-represented sequence patterns (*e.g.*, ref 19). Unfortunately, seed-based methods achieve improvements in running time at the expense of classification accuracy,[16] while motif discovery methods typically impose run-time constraints on motif size and/or number,[19] limiting their utility for broad demographic analyses. In this study, we used AutoSOME,[13] a novel strategy for revealing natural cluster structure in large data sets. As was recently demonstrated,[20] AutoSOME can identify representative sequence populations without the need for *seed* sequences, and without prior knowledge of motif length or cluster number. Thus, the approach described here should both accelerate combinatorial biomolecular discovery for materials engineering applications, and be generally applicable to multiple variations of *in vitro* evolution.

## RESULTS AND DISCUSSION

Each molecule in the starting ssDNA library population (commercially synthesized by Integrated DNA Technologies) consists of 30 random nucleotides flanked by universal priming sequences; these regions permit sequence amplification by PCR (*e.g.*, Figure 1b). As a pilot study, we used polycrystalline nanoparticles of ZnO (Sigma), a technologically important semiconductor,[21] as the SELEX target. In a typical round of SELEX screening, a diverse library of DNA sequences is incubated with the supplied target (here, ZnO nanoparticles) and nonbinding sequences are washed away. We incubated 11.4 nmol of the ssDNA library with 200 ng/mL ZnO nanoparticles (35 nm diameter average particle size) in 1 mL of binding buffer (300 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5), and washed the nanoparticles by centrifugation and resuspension eight times with water and binding buffer. After each wash, nanoparticles were recovered by pelleting with centrifugation and aspiration to remove the supernatant. After particles were resuspended for the subsequent wash, they were transferred to a new test tube before centrifugation to ensure that any plastic-binding aptamers were removed at each wash cycle. The DNA sequences that remained associated with the target ZnO were eluted by heating the nanoparticles at 95 °C in water, centrifuging, and aspirating the supernatant for use as a heterogeneous template in the polymerase chain reaction (PCR) to amplify the originally bound DNA sequences (see Supporting Information for further method details). Because extraneous library sequence diversity is removed at the washing step, this amplified DNA population is enriched with ZnO-binding aptamers relative to the starting library. Typically, this enriched population would be incubated again with the target in one or more iterations of screening to further enhance the ratio of high-affinity aptamers to nonbinding or weakly binding sequences within the library (Figure 1). However, in the present study, this first-generation selected library was subjected to massively parallel sequencing with an Applied Biosystems SOLiD (AB SOLiD) system. Such next generation sequencing strategies have already proved valuable for

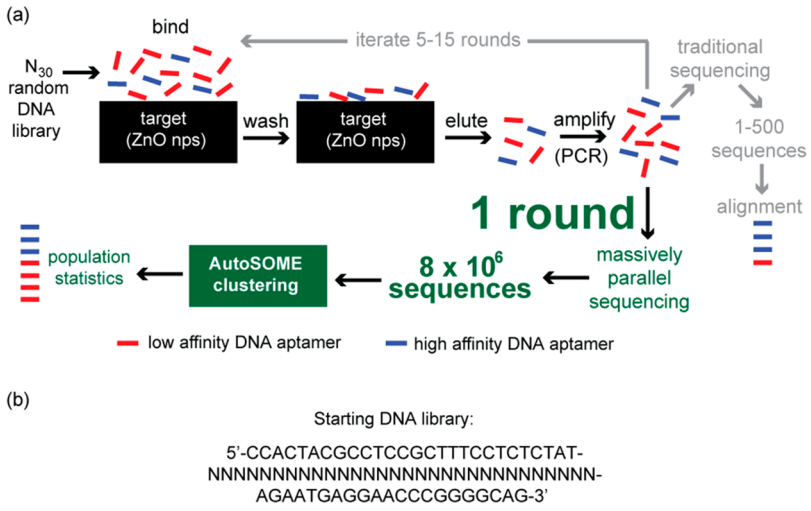



**Figure 1. (a) Schematic comparing the traditional SELEX screening approach using 5−15 rounds of screening (gray) with Single round Enrichment of Ligands by deep Sequencing (SEL-Seq) (green). ZnO nanoparticles (nps) were used as the SELEX target in this study (see text for details). (b) Starting library used in this study: single-stranded DNA with random 30-nucleotide stretches are flanked with universal priming sequences.**

numerous applications, including genome sequencing for biomedical research.[9,10,22,23]

AB SOLiD sequencing yielded an information pool of 8 million individual sequences that describe the screened library. From the 8 million total sequences, a random sample of 300 000 unique DNA molecules was selected for cluster analysis. This smaller sample population reduced the computational load during analysis while ensuring that sequence duplicates generated during PCR amplification were removed. Within this randomly chosen subpopulation, families of related sequences were subsequently identified using AutoSOME (see Methods). A summary of results from AutoSOME cluster analysis is presented in Figure 2a, which shows both the most highly populated as well as less populated sequence families from the screened library. Those less-populated sequence families were randomly selected for display from a large number of statistically rare sequence families. All identified families, together with AutoSOME sequence clustering software and documentation, are available online at http://jimcooperlab.mcdb.ucsb.edu/autosomeseq.

Of the sequence clusters identified, the largest cluster (family 1, Figure 2a) represents ∼2% of the analyzed population and is characterized by the consensus sequence poly $dT_{30}$. Due to the filter applied prior to cluster analysis, which excluded duplicate sequences, this consensus sequence represents an averaged biochemical composition from a family of ∼6,000 distinct individual sequences (Figure 2a). The three other largest clusters are also characterized by consensus sequences that include mononucleotide repeats, namely poly d(C), poly d(G), and poly d(A), respectively (Figure 2a). To determine whether consensus sequences from these large clusters exhibit aptamer behavior for ZnO, we assessed their mineral binding performance. We tested sequences derived from the two most prevalent families (1 and 2, respectively, of Figure 2a), two consensus sequences from smaller families (3 and 4 of Figure 2a, respectively), and the starting random library population (L) as a control.

Each of the five sequences selected for further study was commercially synthesized (Integrated DNA Technologies) to include a 5′-biotin and a 3′-Alexa-488 fluorophore. Each sequence was exposed to a ZnO surface, and nonbinders were removed by washing (see Methods). Bound sequences were captured through their biotin handle with an excess of streptavidin-coated polystyrene beads, which were then analyzed by flow cytometry (see Methods). The results, presented in Figure 2b, show that the most populated sequences in the screened library (sequences 1 and 2) bind with approximately 10-fold higher affinity relative to the starting library (L). Sequences 1 and 2 showed an even greater (∼100-fold) binding difference relative to sequence 4, the least populated sequence family studied, which bound more weakly than the starting library. These results demonstrate that SEL-Seq efficiently identifies aptamer activity after a single round of SELEX screening.

The differential binding activity observed among the studied ssDNA sequences indicates that the sequence composition is important for mineral binding. No stem-loop secondary structures, which usually characterize ssDNA aptamers, were present in the strongest binders studied. The high mononucleotide repeat content in these aptamers suggests that their binding mechanism may involve chelation of mineral metal centers by regularly spaced hydrogen-bonding moieties (*i.e.*, hydroxyl or carbonyl oxygens and amine nitrogens) presented by the DNA molecule, rather than a purely electrostatic interaction which could be achieved by

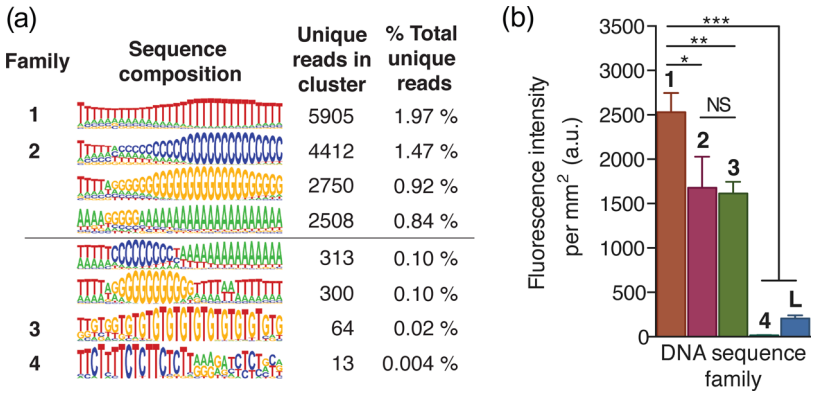



**Figure 2. (a) Results from AutoSOME clustering on 300 000 unique sequence reads isolated after one round of SELEX screening against ZnO nanoparticles. Shown are 8 clustered sequence families. The top four clusters shown are the four largest clusters, while the bottom four were randomly selected for display and from a pool of less populated cluster families. All clusters may be viewed online at http://jimcooperlab.mcdb.ucsb.edu/autosomeseq. Consensus sequences from families 1, 2, 3, and 4 as well as the starting library (L) were selected for further study in mineral binding (b) or mineralization assays (Figures 3 and S1). (b) Results of mineral binding assay with sequences indicated in panel (a) on ZnO surfaces show that highly represented sequence families in the screened library exhibit mineral-binding activity. Values are presented as means ± SEM, and statistical significance was assessed by a two-sided unequal variance *t*-test (*$P = 0.06$; **$P = 0.0026$; ***$P < 1.3 \times 10^{-7}$; NS = not significant).**

many sequences. However, while secondary structures do not characterize the consensus sequences, they may characterize individual sequences within the clustered families. This is suggested in the composition of family 1, in which deoxyadenosine is the second most represented nucleotide (after deoxy-thymine) in that cluster, and would likely permit intramolecular A:T base pairing to occur in some of the sequences. Measuring the binding activity of a consensus sequence relative to its constituents is thus an important subject for future research.

Family 3, while exhibiting an order of magnitude larger representation than the lowest-binding sequence cluster studied (family 4), is still much less frequently represented than families 1 and 2. However, family 3 binds ZnO with strength similar to family 2, and with two-thirds the affinity of family 1. This result might be explained by the particular alternating dG/dT biochemistry that is prevalent in this sequence family, which could yield non-Watson−Crick G:T base pairing. Such 'wobble' base pairing mismatches have previously been shown to occur intramolecularly in DNA, analogously to the G:U pairing that is common in RNA.[24] If regular spacing of metal chelating groups is indeed the mechanism for the aptamer-surface binding we selected, then folded structures dependent on wobble base pairing in this family may have led to removal of a large fraction of this sequence type during the washing rounds, accounting for this cluster's lower prevalence relative to families 1 and 2. Regardless, we note that the largest family, Family 1, bound significantly better than Family 3 ($P = 0.0026$; two-sided unequal variance *t*-test; Figure 2b), consistent with a general pattern toward enhanced ZnO binding activity in larger sequence populations.

Sequences 1 and 4 were further investigated for ZnO-synthesizing activity. The sequences were reacted with 10 mM $Zn(NO_3)_2$ overnight (17 h) in binding buffer and the reaction solutions were vacuum filtered through 10,000 kDa molecular weight cutoff filters and washed with 1 mL $H_2O$ through the same filters. Washed products were resuspended in 70% ethanol, dried on grids by evaporation in air, and analyzed by transmission electron microscopy (TEM) with electron dispersive spectroscopy (EDS) and selected area electron diffraction (SAED). The reaction with sequence 4 showed no Zn signal by EDS and no crystalline materials by SAED in the recovered products (data not shown). In contrast, products from the reaction with sequence 1 showed inorganic−DNA composite materials (Figure 3a) that at high magnification show lattice fringes (Figure 3b) indicating the presence of crystalline mineral. EDS revealed both Zn and P signals, indicative of a Zn−DNA composite (Figure 3c). The selected area diffraction pattern matches closely to crystalline wulfingite $Zn(OH)_2$ (Figure 3d). In a second mineralization assay, DNA sequences were bound to polystyrene bead surfaces prior to mineralization and subsequently analyzed by flow cytometry (Figure S1). Fluorescence measurements assayed for fluorophore-labeled DNA associated (*via* binding or incorporation) with semiconductor material synthesized on bead surfaces. These data confirm that sequence 1 facilitates synthesis of zinc mineralization whereas sequence 4 and the starting library L do not.

We conducted a third set of zinc-based mineralization assays with poly $dT_{30}$ to explore the effect of DNA concentration on mineral formation. In these experiments, 20 $\mu$L reactions were prepared in 1.5 mL Eppendorf tubes, including 10 mM $Zn(NO_3)_2$, binding buffer, and either 1 $\mu$M or 100 $\mu$M of poly $dT_{30}$ DNA. Copper-Formvar TEM grids were laid flat on the surface of each reaction solution (held in place by surface

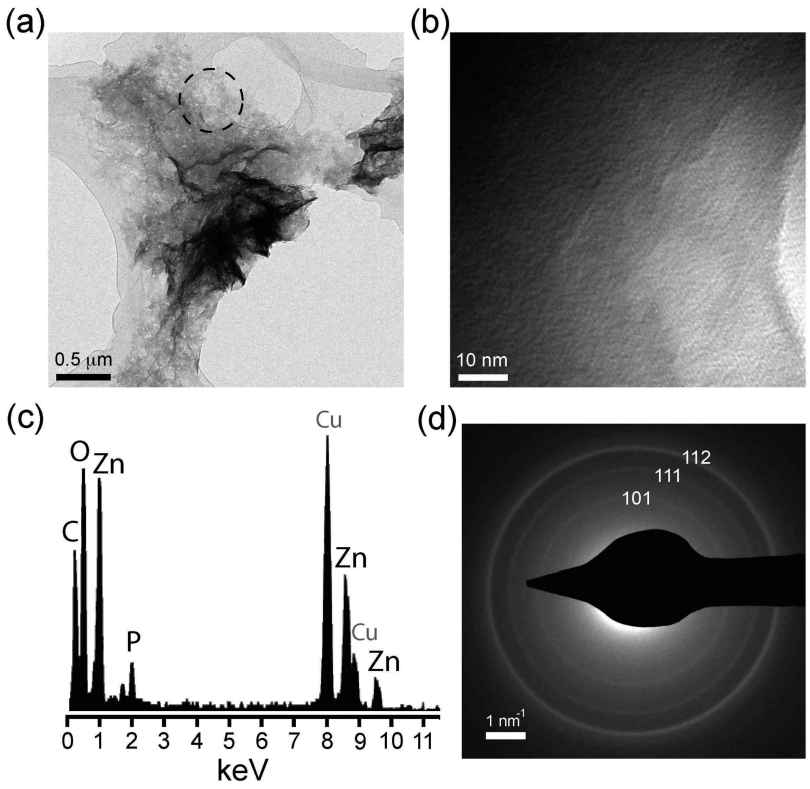



**Figure 3. Products from reaction of consensus sequence 1 with $Zn(NO_3)_2$ for 17 h at near-neutral pH, recovered and washed by filtration and analyzed by TEM with EDS and SAED. (a) TEM image of products. The selected area for electron diffraction shown in (d) is indicated. (b) TEM image of products at higher magnification, showing lattice fringes suggestive of a crystalline product. (c) EDS spectrum of materials shown in (a) and (b). (d) SAED pattern of material shown in (a) and (b), indexing to crystalline $Zn(OH)_2$.**

tension) and remained there during reaction incubation. After 24 h, the grids were collected and immediately rinsed with 1 mL $H_2O$, dried, and then imaged by TEM. The results show that crystalline ZnO nanoparticles are formed on the TEM grids at low (1 $\mu$M) DNA concentration (Figure 4a), while at higher concentration only amorphous agglomerates are formed (Figure 4b). Interestingly, at high magnification the ZnO nanoparticles are seen to comprise an amorphous−crystalline core−shell structure (Figure 4a, middle). These findings suggest that ZnO crystallization is initiated *via* interactions of ions with the surfaces of individual DNA molecules (Figure 4c), as the core regions of the nanoparticles are in the approximate size range of the average hydrodynamic diameter, ~7 nm, that has been previously determined for poly $d(T)_{30}$.[25] The formation of nanoscale metal−DNA coacervates during the mineralization process, as suggested by the data in Figure 4, would allow thymine bases to spatially template $Zn^{2+}$ ions (Figure 4c, left) and may further permit structural ordering of water molecules around each mineral nucleus. In contrast, coacervate formation, which is sensitive to the concentration of constituent species,[26] is not apparent in the mineralization products on the TEM grid from reaction with the higher DNA concentration (Figure 4b).

To determine the specific hydrodynamic diameter of poly $d(T)_{30}$ in this system and further investigate the possibility of a nanocoacervate-driven mineralization mechanism, we conducted dynamic light scattering (DLS) analysis on poly $d(T)_{30}$. At 100 $\mu$M, poly $d(T)_{30}$ is found by DLS to exhibit a bimodal size distribution, with average hydrodynamic diameters of 4.78 and 269 nm, respectively (Figure S2a). The smaller average diameter (4.78 nm) likely represents the size of individual poly $d(T)_{30}$ molecules, which comprise 32.6% of the overall molecular population, while molecular agglomerates (average diameter 269 nm) comprise 67.4% of the population. When $Zn(NO_3)_2$ is added to the sample (to 10 mM final concentration), the bimodal population persists and the average diameters decrease to 4.55 and 210 nm, representing 22.7% and 77.3% of the population, respectively (Figure S2a). No change was thereafter observed in this distribution over 90 min of analysis (with DLS readings taken every 15 min). Thus, three main changes occur initially through the addition of zinc: (1) zinc induces a conformational contraction of individual poly $d(T)_{30}$ molecules, (2) a fraction of individual poly $d(T)_{30}$ molecules are shifted toward agglomerated structures, and (3) the agglomerated structures also contract in size (Figure S2a). This is consistent with the presence of zinc-induced nanocoacervates in the reaction solution, because if zinc contributed only to cross-linking (*versus* nanocoacervation), an increase in agglomerate size

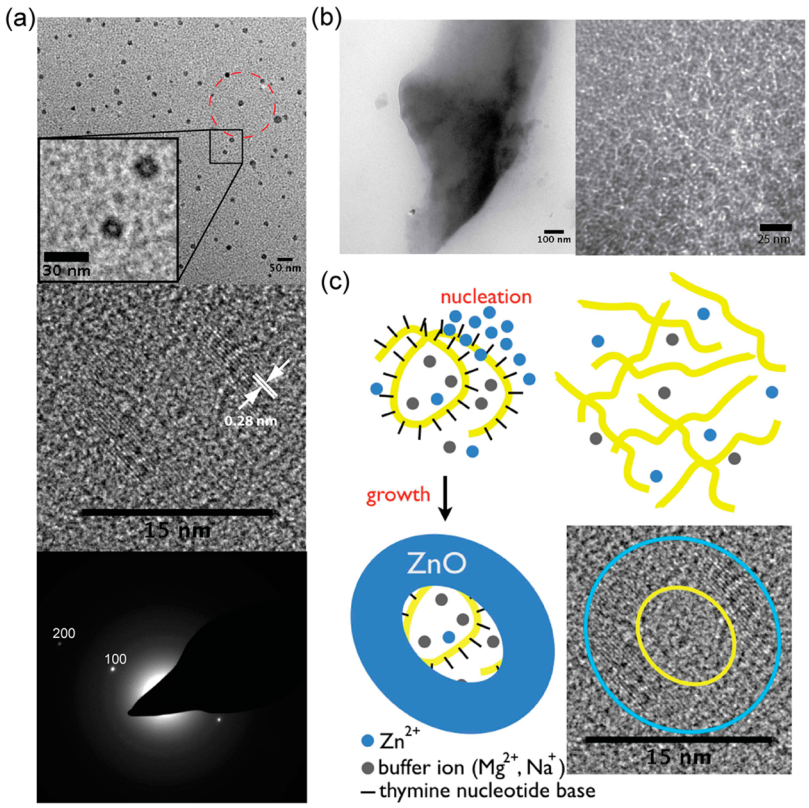



**Figure 4. Products from reaction of consensus sequence 1 with $Zn(NO_3)_2$ for 24 h at near-neutral pH, with products collected directly on the TEM grid during mineralization (see text). (a) Reaction with 1 $\mu$M DNA yields nanoparticles (top) composed of an amorphous core with a crystalline ZnO shell, as indicated by relatively electron-transparent nanoparticle cores (top, inset), lattice fringes at the nanoparticle shells (middle), and selected area electron diffraction (bottom), confirming the presence of ZnO. The area selected for electron diffraction is indicated in red (top). (b) Reaction with 100 $\mu$M DNA yields amorphous sheets composed of DNA-mineral agglomerates. (c) Schematics of mineralization models in which nanoscale coacervation around individual DNA molecules (upper left) templates ZnO nuclei *via* regular spacing of thymine bases, yielding crystalline ZnO nanoshells around an amorphous core (bottom). A higher concentration of DNA would produce an interpenetrating network of DNA molecules (upper right; bases, while not shown, would bridge DNA molecules *via* interactions with divalent ions), disrupting sites for ordered mineral templating.**

would be expected, in contrast to the observed average size decrease for this population; and if zinc had no effect, no size change would be expected. The lower poly $d(T)_{30}$ reaction concentration of 1 $\mu$M, which yields ZnO nanoparticles (Figure 4a), was found to be too dilute to generate a detectable scattering signal, precluding DLS analysis of the DNA at this concentration. As an alternate approach to examine DNA structure at 1 $\mu$M poly $d(T)_{30}$, we sought to image $Zn^{2+}$-stained DNA by conducting TEM analysis on early stage (45 min) reaction products from 1 $\mu$M poly $d(T)_{30}$ reacted with 10 mM $Zn(NO_3)_2$. This revealed the presence of amorphous nanopoarticles in the range of 7−12 nm (Figure S2b), which we interpret as poly $d(T)_{30}$-$Zn^{2+}$ nanocoacervates at early stages of mineralization. This particle size agrees with the previous study that reported the size of poly $d(T)_{30}$ to be ~7 nm, a size which was determined under dilute ssDNA conditions (1 nM).[25] Further, as shown in Figure 4c and highlighted further in Figure S3, in the core−shell structure formed after a 24 h reaction the ZnO lattice fringes are separated by a ~7 nm amorphous core. In some cases, core−shell amorphous-hexaganol-ZnO nanoparticles are also observed in the early stage (45 min) reaction products (Figure S2c), further supporting the conclusion that the stained amorphous nanoparticles represent a precursor structure to ZnO formation. It is possible that the amorphous regions of these core−shell structures include both the templating DNA as well as an amorphous or partially ordered zinc-mineral seed layer that contributes to nanoshell ZnO crystallization. However, even in this case that early stage mineralization leads to a diameter size increase of individual poly $d(T)_{30}$ molecules at 1 $\mu$M concentration (to ~7 nm, *vs* ~4.5 nm at 100 $\mu$M), we note again that no increase in diameter was observed by DLS for poly $d(T)_{30}$ at 100 $\mu$M for the same early stage reaction period in the presence of zinc (Figure S2a). Thus, overall, these data agree with a scenario in which zinc- poly $d(T)_{30}$ nanocoacervates formed at 1 $\mu$M DNA are uniquely suited to template wurtzite ZnO.

Overall, these data collectively suggest that zinc and poly $d(T)_{30}$ complex with one another to form diverse yet structurally defined precursor architectures that

drive mineralization, and that hexagonal ZnO is formed only when mineralization occurs at structurally independent (*vs* agglomerated) poly $d(T)_{30}$ molecules that achieve a specific hydrodynamic size. Since wulfingite $Zn(OH)_2$ (and not ZnO) was observed from the bulk solution form a reaction between 1 $\mu$M poly $d(T)_{30}$ and $Zn(NO_3)_2$ (Figure 3), the Formvar film of the TEM grid may contribute to the formation of the ZnO nanoshells (Figure 4). However, no nanoshells were found on the grid at 100 $\mu$M, confirming the importance of DNA concentration to the crystallization process. Also, if ZnO nanoshells were also formed in bulk solution at 1 $\mu$M poly $d(T)_{30}$, these could have been potentially lost by filtration during preparation of the materials shown in Figure 3. While the possible role of the Formvar surface requires further study, we confirmed that buffer conditions are critical to ZnO formation, as a reaction conducted in HEPES buffer produced only amorphous products (Figure S4) under conditions that otherwise yielded ZnO nanoshells when binding buffer (Tris-based) is used (Figure 4). Additional controls using either 1 $\mu$M of DNA family 4 (consensus sequence, Figure 2) or thymine monomer (30 $\mu$M) produced no observable reaction products of any kind on the TEM grids, confirming the importance of both sequence and DNA structure to the mineralization process. Further, on the basis that wulfingite was observed only from the reaction with 1 $\mu$M (and not 100 $\mu$M) poly $d(T)_{30}$, we infer that agglomerate structures formed at 1 $\mu$M poly $d(T)_{30}$ are structurally distinct from those formed at higher concentration such that they are able to promote zinc hydroxide crystallization. Finally, as an initial test to explore directed tuning of poly $d(T)_{30}$ conformation, we found that larger zinc-based structures of $\sim$100 $\mu$m, as observed by SEM (Figure S5) could be mineralized when poly $d(T)_{30}$ is chemically conjugated with cholesterol to promote DNA self-assembly (100 $\mu$M of DNA was used and no crystalline products were observed by TEM). Clearly, the poly $d(T)_{30}$–$Zn^{2+}$ system represents a rich mineralization reaction space that is sensitive to changes in buffer conditions, DNA concentration, and DNA conformation.

Among this complex reaction space, the ZnO nanoshells represent the most interesting observed product based on the technological value of this material and the difficulty of forming it under near-neutral aqueous conditions, as has been achieved here. In addition to the above TEM and DLS data, poly deoxythymidine-based DNA templating of ZnO crystallization is also supported by the closely matched spacing that occurs between adjacent DNA bases in a given oligomer (0.34 nm)[25] and the *a* lattice constant (corresponding to the Zn–Zn spacing) in the basal plane of wurzite ZnO (0.325 nm).[27] Additionally, the positioning of nucleobase ligands has been shown important toward achieving ordered DNA–metal complexes.[28] Thus, because thymine presents the most symmetric configuration of potential Zn-bonding sites (carbonyl, amine, carbonyl) of the four DNA bases, it could preferentially order $Zn^{2+}$ ions into a spatial configuration relevant to crystal nucleation. Further studies are required to gain a better understanding of both ZnO nanoshell crystallization and the extensive phase diagram of this mineralization system over a variety of reactant and buffer conditions. It is clear, however, that zinc–[poly d(T)] mineralization is a sequence-dependent process that was quickly discovered in this study through the new biotechnologies of high-throughput sequence and bioinformatic (AutoSOME) cluster analysis.

Collectively, the above results confirm that mineral-binding aptamers, rapidly identified by SEL-Seq, can serve as useful and flexible tools for mineral synthesis. While the employed starting library was not sequenced in this initial study, the largest sequence cluster, Family 1, exhibited the most significant binding of all families tested (Figure 2b), and the mineral synthesis (Figure S1) comparisons between the AutoSOME-identified sequences *versus* the starting library demonstrate that the identified aptamers are enriched relative to the starting population, and exhibit superior performance. We estimate that $\sim 5 \times 10^6$ DNA molecules were contained in the eluate after screening (*i.e.*, prior to PCR amplification; see Supporting Information), which is only $\sim 10^{-9}$ of the initial DNA pool utilized for screening (11.4 nmol or $\sim 10^{15}$ DNA molecules). This suggests a low amount of background binding was represented in the eluate. Further, of the 300 000 unique eluted sequences analyzed, the fact that the largest cluster represented only 2% of the number of molecules in the entire eluate suggests that the enriched library was still highly diverse. Thus, the fact that the identified aptamers exhibit significant mineral binding and synthesis activity demonstrates that our approach efficiently accelerates SELEX screening.

While beyond the scope of the current study, SEL-Seq could be applied to gain additional information about the SELEX screening process. Such applications might include analyses of synthesized libraries, such as investigations of how sequence populations compositionally and architecturally evolve in iterative rounds of SELEX, and novel implementations of multistage screening strategies to quickly identify even tighter-binding mineral aptamers. As a possible example of the latter approach, with current DNA synthesis capabilities (*e.g.*, ref 29) it would be possible to synthesize the entire set of individual sequences in a cluster (*e.g.*, family 1 in Figure 2a) for a final screening round that is likely to yield (given the performance of the consensus sequences shown here) very highly performing aptamer sequences. Such analyses could ultimately yield new insights into the molecular mechanisms by which exquisitely complex biological systems, such as those that direct biomineralization, have been developed through combinatorial selection in nature.

## CONCLUSION

Inspired by the diverse biominerals found in nature, combinatorial biomolecular screening approaches are being increasingly used to identify laboratory-generated biomolecules with mineralization activity.[2,30,31] Polypeptide libraries,[6,30] RNA libraries,[32] and, most recently, mutant enzyme libraries[33] have been used in this context, but the present study is the first in which DNA libraries have been screened for defined DNA−mineral interactions. DNA is a ubiquitous engineering substrate in bionanotechnology due to its relative stability, ease of manipulation with recombinant enzymes, and amenability to chemical synthesis. By demonstrating that a DNA aptamer identified with SEL-Seq catalyzes synthesis of its target mineral from solution, the present study introduces SELEX-generated DNA aptamers, high-throughput DNA sequencing, and sequence population clustering as new tools for use in combinatorially engineering biomimetic nanomaterials.

## METHODS

The starting ssDNA library, prepared *via* hand-mixing synthesis randomization, was confirmed by the manufacturer (Integrated DNA Technologies) to be free of nucleotide bias (25 ± 2% representation of each nucleotide). Prior to the mineral-binding SELEX screening step, the library was heated at 95 °C for 5 min and slowly cooled to room temperature to allow secondary structures to form. 11.4 nmol of the ssDNA library was then incubated with 200 ng/mL ZnO nanoparticles (Sigma; 35 nm diameter average particle size) in 1 mL of binding buffer (300 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.6) using low-binding Axygen Maxy-Clear 1.5 mL test tubes. The mixture was incubated with gentle mixing in a tube turner for 1 h at 21 °C. To remove nonbinding sequences the nanoparticles were washed four times with 1 mL binding buffer and four times with 1 mL water by centrifugation and removal of the supernatant by aspiration. After the final wash, nanoparticles were resuspended in 100 $\mu$L buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). To elute nanoparticle-associated ssDNA sequences, this mixture was heated at 95 °C for 5 min, centrifuged at 20 000*g* for several seconds, and the supernatant was then transferred to a new test tube. The entire eluted library was used as a template for gene amplification by PCR, using conditions described in the Supporting Information. AB SOLiD sequencing was conducted according to the manufacturer's instructions, and details of the surface-binding assay are described in the Supporting Information of this paper.

For AutoSOME cluster analysis, DNA sequences were randomly chosen without replacement from the pool of 8 million sequences until 300 000 unique sequences were obtained (451 016 total sequences were needed). To capture putative context dependencies between adjacent nucleotides, each DNA sequence was converted into a dinucleotide compositional vector (*i.e.*, a vector of length 16 to accommodate all possible dinucleotides) prior to clustering. All 300 000 compositional vectors were then clustered using the AutoSOME algorithm, and the resulting clusters were saved to memory (AutoSOME parameters: unit variance normalization, $P \leq 0.01$, 20 ensemble iterations, otherwise default parameters as described in ref 13). Next, to align possible ZnO binding motifs located in different DNA sequence regions, sequences within each compositional cluster were exhaustively compared to all cyclical permutations of every other sequence in the same cluster, and the highest quality alignment was identified. To represent the alignment, a master consensus sequence was generated (most frequent nucleotide in each column), and a consensus error *E* was computed (number of mismatches to the consensus sequence divided by total number of aligned nucleotides). If *E* was ≤0.4, the cluster was saved to disk. Otherwise, AutoSOME was rerun on the compositional cluster to identify subclusters. This second stage was recursively repeated until each cluster either satisfied a more stringent consensus error threshold ($E \leq 0.3$) or the size of the cluster fell below a minimum size threshold (5 sequences). Finally, population statistics were tabulated for each sequence cluster relative to the 300 000 randomly sampled unique sequences. Conditions for mineralization assays directly from solution are given in the main text (above). Corroborating mineralization assays conducted with DNA displayed from polystyrene microbeads and analyzed by FACS are described in the Supporting Information. Dynamic light scattering data was conducted using a Malvern Zetasizer Nano instrument.

To statistically assess the differences in binding affinity between families in Figure 2b, we first performed outlier analysis using the ROUT method,[34] which identified and removed two measurements with a 3 $mm^2$ target mineral surface area. Samples were subsequently determined to have a normal distribution *via* the D'Agostino and Pearson omnibus normality test, which justified the application of a two-sided *t*-test with unequal variance to evaluate the statistical significance between fluorescence intensity measurements.

*Conflict of Interest:* The authors declare no competing financial interest.

*Acknowledgment.* All aspects of this work were supported by a grant from the U.S. Dept. of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering under Award #DE-FG03-02ER46006 to D.E.M. In addition, L.A.B. was partially supported by University of California Systemwide Biotechnology Research & Education Program GREAT Training Grant 2007-19. We thank K. Kosik (UC-Santa Barbara) for access to the AB-SOLID instrument, L. Fielding and S. Armes (University of Sheffield, U.K.) for assistance with DLS measurements, and D. Kolodin, J. León, and G. Mantalas for their helpful experimental assistance.

*Supporting Information Available:* This includes five supporting figures and supporting experimental methods. This material is available free of charge *via* the Internet at http://pubs.acs.org.

## REFERENCES AND NOTES

1. Lee, Y. J.; Yi, H.; Kim, W. J.; Kang, K.; Yun, D. S.; Strano, M. S.; Ceder, G.; Belcher, A. M. Fabricating Genetically Engineered High-Power Lithium-Ion Batteries Using Multiple Virus Genes. *Science* **2009**, *324*, 1051–1055.
2. Bawazer, L. A. From DNA to Genetically Evolved Technology. *MRS Bull.* **2013**, *38*, 509–518.
3. Ellington, A. D.; Szostakz, J. W. Selection of *in Vitro* Single-Stranded DNA Molecules that Fold into Specific Ligand-Binding Structures. *Nature* **1992**, *355*, 850–852.
4. Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment—RNA Ligands to Bacteriophage-T4 DNA Polymerase. *Science* **1990**, *249*, 505–510.
5. Pasqualini, R.; Ruoslahti, E. Organ Targeting *in Vivo* Using Phage Display Peptide Libraries. *Nature* **1996**, *380*, 364–366.
6. Whaley, S. R.; English, D. S.; Hu, E. L.; Barbara, P. F.; Belcher, A. M. Selection of Peptides with Semiconductor Binding Specificity for Directed Nanocrystal Assembly. *Nature* **2000**, *405*, 665–668.
7. Shao, Z. X.; Arnold, F. H. Engineering New Functions and Altering Existing Functions. *Curr. Opin. Struct. Biol.* **1996**, *6*, 513–518.

8. Tok, J. B. H.; Fischer, N. O. Single Microbead SELEX for Efficient ssDNA Aptamer Generation against Botulinum Neurotoxin. *Chem. Commun.* **2008**, *16*, 1883–1885.
9. Shendure, J.; Ji, H. L. Next-Generation DNA Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.
10. Kircher, M.; Kelso, J. High-Throughput DNA Sequencing—Concepts and Limitations. *BioEssays* **2010**, *32*, 524–536.
11. Glenn, T. C. Field Guide to Next-Generation DNA Sequencers. *Mol. Ecol. Resour.* **2011**, *11*, 759–769.
12. Rothberg, J. M.; Hinz, W.; Rearick, T. M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J. H.; Johnson, K.; Milgrew, M. J.; Edwards, M.; *et al.* An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing. *Nature* **2011**, *475*, 348–352.
13. Newman, A. M.; Cooper, J. B. AutoSOME: A Clustering Method for Identifying Gene Expression Modules without Prior Knowledge of Cluster Number. *BMC Bioinf.* **2010**, *11*, 117.
14. Cho, M.; Xiao, Y.; Nie, J.; Stewart, R.; Csordas, A. T.; Oh, S. S.; Thomson, J. A.; Soh, H. T. Quantitative Selection of DNA Aptamers through Microfluidic Selection and High-Throughput Sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15373–15378.
15. Zimmermann, B.; Gesell, T.; Chen, D.; Lorenz, C.; Schroeder, R. Monitoring Genomic Sequences during SELEX Using High-Throughput Sequencing: Neutral SELEX. *PLoS One* **2010**, *5*, e9169.
16. Sun, Y.; Cai, Y.; Huse, S. M.; Knight, R.; Farmerie, W. G.; Wang, X.; Mai, V. A Large-Scale Benchmark Study of Existing Algorithms for Taxonomy-Independent Microbial Community Analysis. *Briefings Bioinf.* **2012**, *13*, 107–121.
17. Li, W.; Godzik, A. Cd-hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22*, 1658–1659.
18. Edgar, R. C. Search and Clustering Orders of Magnitude Faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461.
19. Hu, J.; Li, B.; Kihara, D. Limitations and Potentials of Current Motif Discovery Algorithms. *Nucleic Acids Res.* **2005**, *33*, 4899–4913.
20. Newman, A. M.; Cooper, J. B. Global Analysis of Proline-Rich Tandem Repeat Proteins Reveals Broad Phylogenetic Diversity in Plant Secretomes. *PLoS One* **2011**, *6*, e23167.
21. Klingshirn, C. ZnO: Material, Physics and Applications. *ChemPhysChem* **2007**, *8*, 782–803.
22. Kahvejian, A.; Quackenbush, J.; Thompson, J. F. What Would You Do if You Could Sequence Everything? *Nat. Biotechnol.* **2008**, *26*, 1125–1133.
23. Ng, S. B.; Turner, E. H.; Robertson, P. D.; Flygare, S. D.; Bigham, A. W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E. E.; *et al.* Targeted Capture and Massively Parallel Sequencing of 12 Human Exomes. *Nature* **2009**, *461*, 272.
24. Kennard, O. Structural Studies of DNA Fragments − The G. T Wobble Base Pair in A-DNA, B-DNA and Z-DNA; The G.A Base Pair in B-DNA. *J. Biomol. Struct. Dyn.* **1985**, *3*, 205–226.
25. Doose, S.; Barsch, H.; Sauer, M. Polymer Properties of Polythymine as Revealed by Translational Diffusion. *Biophys. J.* **2007**, *93*, 1224–1234.
26. de Kruif, C. G.; Weinbreck, F.; de Vries, R. Complex Coacervation of Proteins and Anionic Polysaccharides. *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 340–349.
27. Morko, H.; Ozgir, O. *Zinc Oxide: Fundamentals, Materials and Device Technology*; Wiley-VCH: Weinheim, Germany, 2009.
28. Tanaka, K.; Clever, G. H.; Takezawa, Y.; Yamada, Y.; Kaul, C.; Shionoya, M.; Carell, T. Programmable Self-Assembly of Metal Ions Inside Artificial DNA Duplexes. *Nat. Nanotechnol.* **2006**, *1*, 190–194.
29. Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R.-Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010**, *2*, 52.
30. Brown, S. Engineered Iron Oxide-Adhesion Mutants of the Escherichia-coli Phage-Lamdba Receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8651–8655.
31. Feldheim, D. L.; Eaton, B. E. Selection of Biomolecules Capable of Mediating the Formation of Nanocrystals. *ACS Nano* **2007**, *1*, 154–159.
32. Gugliotti, L. A.; Feldheim, D. L.; Eaton, B. E. RNA-Mediated Metal-Metal Bond Formation in the Synthesis of Hexagonal Palladium Nanoparticles. *Science* **2004**, *304*, 850–852.
33. Bawazer, L. A.; Izumi, M.; Kolodin, D.; Neilson, J. R.; Schwenzer, B.; Morse, D. E. Evolutionary Selection of Enzymatically Synthesized Semiconductors from Biomimetic Mineralization Vesicles. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1705–1714.
34. Motulsky, H. J.; Brown, R. E. Detecting Outliers When Fitting Data with Nonlinear Regression—A New Method Based on Robust Nonlinear Regression and the False Discovery Rate. *BMC Bioinf.* **2006**, *7*, 123.