# Combinatorial Approaches to Constructing Protein Nanostructures

Ed Frank*, Yao Lin** and Lee Makowski**

*Mathematics and Computer Science Division, efrank@northwestern.edu
**Biosciences Division, lin@anl.gov, lmakowski@anl.gov
Argonne National Laboratory
Argonne, Illinois 60439

## ABSTRACT

A major bottleneck in the use of proteins for the construction of complex nanostructures is the design of protein interfaces that will bind to one another with high-affinity and high specificity. The interaction of subunits with one another may be mediated by non-covalent, high-affinity protein-protein contacts, as is the case in many naturally occurring multi-subunit complexes. However, there are currently no well developed strategies for the design of complementary binding surfaces such as those required for the engineering of multi-subunit nanostructures. In the absence of a design paradigm, a promising alternative is the selection of tight binders from large, combinatorial libraries. The design specifications for these libraries include stability and physical robustness of the scaffold molecule; diversity of the displayed peptides; and ease of synthesis/expression.

***Keywords***: phage display, nanostructures, proteins, hybrid structures

## 1 INTRODUCTION

The properties of inorganic nanoparticles have been extensively studied and exhibit properties that are not observed in bulk samples of the compounds from which they are made. In complex nanostructures, adjacent nanoparticles may exhibit proximity effects that are expected to create emergent phenomena unexpected on the basis of individual particle behavior. For instance, a small number of gold nanoparticles of progressively decreasing size may act as a 'nanolens' to focus surface plasmons into very significant field gradients when exposed to light [1]. Experimental exploration of multiple particle effects and their utilization in nanomaterials and devices will require precision synthesis of large numbers of complex nanostructures, virtually identical to one another. However, the synthesis of complex nanostructures composed of multiple particles represents a significant challenge.

Biomolecular systems constantly carry out construction of complex, nanoscale assemblies. Proteins constitute a critical fraction of virtually all of these assemblies. Monodisperse populations of proteins can be produced at very large scales. Their structural and mechanical diversity and extraordinary recognition properties – providing for binding to both organic and inorganic materials - make them strong candidates for use as subunits in the construction of complex nanostructures. The potential of proteins as scaffolds for construction of hybrid nanostructures has been clear for several years [2,3]. Practical implementation of this approach has been limited to repeating structures such as linear arrays of proteins or two-dimensional lattices [4] or the use of biomolecules as spacers in the formation of nanostructured materials [5]. These structures utilize one or two inter-particle interactions over and over again. As such, they fall short of a demonstration of the synthesis of a complex nanostructure that incorporates multiple nanoparticles into a precise, non-periodic arrangement. To accomplish that requires the use of multiple, distinct inter-particle interactions. The design, discovery or synthesis of the interacting pairs of surfaces required for these constructs represents a major barrier for formation of complex nanostructures. Perhaps the most promising approach to overcoming this barrier is the use of combinatorial libraries for selection of surfaces that will bind to target molecules.

Combinatorial libraries of peptides or proteins, either extending from the C- or N-terminus of a protein, or incorporated within loops of a protein, have proven highly effective for identifying reagents with affinity to specific target molecules. If the target molecules are chosen to be structural or functional components of a nanostructure, the approach may provide an easy path to identification of interacting surfaces. This capability greatly enhances the potential of using proteins as scaffolds by providing an almost limitless expansion of the targets to which proteins can be engineered to bind.

A naturally occurring protein has been designed through evolution to support the survival of a particular organism and it is unlikely that any protein, in unmodified form, will represent the ideal design for use in a nanostructure. Consequently, to make nanostructures, we must either resort to *de novo* design or protein modification. Modification of an existing protein to satisfy the engineering specifications of a nanostructure will require the attachment of functional groups to the protein *and* the formation of interaction surfaces that can attach with high affinity and specificity to other structural elements of the structure. Since attachment of a functional group (such as

an inorganic nanoparticle) may involve the same kind of conceptual problem as formation of an interaction surface, we will focus only on the later here.

Effective methods for design of protein surfaces that will interact with one another have not yet been developed. The alternate approach considered here is to construct a combinatorial library of 'surfaces' [6] and from that library select those that interact most strongly with a target molecule. Functionally, the library is a population of proteins each exhibiting one of a large number of possible amino acid sequences that constitute a terminal peptide, a loop (or several loops), or perhaps the surface of an $\alpha$-helix. The power of phage display is that each sequence in the library is physically linked to the DNA that encodes it. When a particular member of the library is affinity selected it can be amplified in bulk because of that linkage.

Affinity selection is carried out to identify a surface that is an integral part of one protein and will interact strongly and specifically with a selected surface of a target molecule. The target could be another protein, the surface of an inorganic nanoparticle or some other element to be incorporated into the nanostructure. Phage display technology has identified peptides or proteins with high affinity to many organic [7] and inorganic targets [8]. Here, we describe experiments with the goal of applying combinatorial methods for the construction of complex nanostructures in which protein subunits act as a scaffold for the precise placement of inorganic nanoparticles relative to one another.

In these experiments we utilize a single protein, the fibronectin FN3 domain, as a structural subunit. The FN3 domain was chosen because it is small (less than 100 amino acids), its three-dimensional structure is well known and it appears to be highly amenable to re-engineering in a way that does not disrupt the overall structure of the protein [9]. It has three flexible loops at each end. Two have been replaced with a large diversity of peptide sequences resulting in the construction of large combinatorial libraries [10] and the other four appear to be amenable to similar, if somewhat more limited, modification.

.

## 2 METHODS

In order to construct pairs of binding surfaces that will enable variants of FN3 to bind to one another, we modified wild-type (WT) FN3 in two ways; construction of random libraries of peptides on loops BC and FG (adjacent to one another at one end of the protein); and construction of individual amino acid substitutions along the 'waist' of the molecule. The randomization of two adjacent loops produced a very large variety of potential binding surfaces. The single amino acid substitution produced a small target on the surface of FN3 that was distinct from WT. The purpose of randomization was to create a library so large that by random chance some members that bind tightly to the region of FN3 exhibiting the single amino acid substitution could be found. Selection from the library using phage display methods would then make possible the identification of loop sequences that exhibited high affinity binding to the mutation site.

Once these modifications were characterized, it would then be possible to construct hybrid FN3 particles carrying two modifications - (i) a particular pair of loop peptides specific for binding to a single site mutation and (ii) a single site mutation. Here we will designating by the letter 'A' a library member specific for binding to a specific mutation, 'a'. Using this notation, a hybrid FN3 exhibiting mutation 'b' and peptide loops 'A' might be designated as [A-b]. [A-b] will exhibit high affinity to [B-c] (to form the dimer, [A-b]:[B-c], through the interaction of peptides 'B' with mutation 'b') and [C-a] (to form the dimer [C-a]:[A-b], through the interaction of peptides 'A' with mutation 'a'). A variant exhibiting both 'A' and 'a', that is [A-a], will self-assemble into a homopolymer, …:[A-a]:[A-a]:[A-a]:… and would quite possibly be very difficult to express in large quantities since the formation of this polymer *in vivo* could disrupt the host cell carrying out the protein expression. However, two constructs such as [A-b] and [B-a] can be expressed and purified separately, without forming a polymeric structure. When mixed together, they should polymerize into …:[A-b]:[B-a]:[A-b]:[B-a]:… This polymer would be of indeterminate length and, as such, does not represent an advance over many existing methods for the formation of one- and two-dimensional arrays of proteins. The power of the approach becomes evident when used in combination with a directed assembly process.

Subunits constructed in this way can be utilized in a surface-directed assembly process that mimics solid-state polymer synthesis and provides complete control over the order in which subunits are added. For instance, assume that a subunit, [D-b], can be immobilized on a surface in such a way that the binding potential of site 'b' is maintained. It is then possible to incubate this surface with a solution containing [B-a], forming a set of immobilized dimers, |[D-b]:[B-a]. After washing the surface of excess [B-a], the surface can then be incubated with [A-b], forming a set of immobilized trimers, |[D-b]:[B-a]:[A-b]. This process can be continued for any number of cycles, limited only by the error rate of the process and the stability of the resulting constructs. In the most general case, each subunit, [X-y], could have a different geometry, and carry a different functional nanoparticle. Even with only two or three binding pairs that do not cross react, the number of possible constructs is huge. Formation of branching structures requires at least one subunit to have a third binding surface.

A surface-directed assembly process of this kind would result in the massively parallel synthesis of a very large number (~$10^{10}$ - $10^{11}$ per sq. cm.) of virtually identical nanostructures. The first step in implementation of this process is to construct the binding pairs.

Phage displayed libraries of FN3 loops BC and FG were constructed following the work of Koide [9] and Kay [10]. Single site mutations in WT FN3 were constructed using standard methods. Libraries of FN3 were screened for members that bound specifically to mutants of FN3 by first pre-screening against WT FN3 to remove members that bound to FN3 and then using standard phage display methods for selection of library members that bound to the mutants. Mutants of FN3 were immobilized for screening either by direct binding to a plastic plate or by *in vivo* biotinylation followed by immobilization on a streptavidin-coated plate. Selected library members were sequenced and the sequences of the peptide inserts determined in order to aid in the identification of consensus sequences. Selected library members were then characterized for affinity to both mutant and WT FN3 as well as control molecules that included streptavidin and bovine serum albumin (BSA).

## 3 RESULTS

FN3 domains expressed using a vector that results in the *in vivo* biotinylation of the protein [11] were immobilized on streptavidin coated plates and streptavidin coated magnetic beads. Members of the FN3 library that bind to WT FN3 were removed from the library by attaching biotinylated WT FN3 to magnetic beads and incubating them long enough to remove all library members that bind wild type from the library. The fractionation was repeated and the depleted library was titered after each depletion to insure that all members with affinity to wild type were removed.

Four mutants of FN3 were constructed (E9V; R6Q; R33Q and E47V). These amino acid substitutions are in the 'waist' of FN3, remote from the loops at either end of the molecule. Of these, E9V displayed instability, running as a diffuse band at an elevated molecular weight in SDS-PAGE. This suggested that the mutant was not folding properly, and it was dropped from further analysis.

Two of the remaining clones (R6Q and E47V) were chosen as targets for screening of the FN3 library. WT FN3 was used as a control target for these experiments. These constructs were bound to a streptavidin coated plate through biotin attached to the mutant *in vivo*. Members of the library that bind the mutant FN3 were selected using standard phage-display techniques.

In order to evaluate the selected members of the library, eight clones were chosen at random and their DNA sequenced. The sequences of the random inserts into loop BC and loop FG were compared for library members selected against the two mutants and the wild type. These sequences all exhibited a high abundance of basic residues suggesting that binding specificity was low. In order to test this hypothesis, ELISAs were carried out for selected clones to determine the relative binding affinity of the selected library members for the different target proteins. Results suggested both low affinity and low specificity. In some cases clones did not exhibit substantial selectivity for target over BSA. Additional controls indicated that in most cases the selected library member bound to SA tighter than to target protein, suggesting that most library members had bound to the SA coated plate through direct interaction with the SA rather than interaction with FN3. Additional binding experiments indicated that the binding of library members to target proteins were weak and called into question the integrity of the target proteins.

Analysis of target proteins on native gels indicated that in the presence of a biotinylation site, the FN3 domains were forming oligomers. Oligomerization might well block the site of mutation, making it impossible to select library members for affinity to that site. Consequently, the experiments were repeated with a vector that did not cause the biotinylation of FN3. In this case, mutant FN3 domains were immobilized directly onto plastic, and the selection experiments repeated. Analysis of selected library members again suggested low affinity and low specificity.

From these results, we conclude that the FN3 mutants used as targets in these experiments are not sufficiently distinct in structure from WT to offer an adequate target for library selection. Recall that the library from which members were being selected was depleted of members with high affinity for WT. The results of our experiments suggest that the depleted library contains few if any members with high affinity and specificity for the mutations constructed from WT. Consequently, we chose to re-design the mutants and repeat the selection experiments rather than move forward with the assembly experiments on the basis of poorly binding clones.

## 4 DISCUSSION

The long term goal of this work is the synthesis of large populations of nanostructures that are identical, functional, and exhibit novel properties – complex, multi-subunit nanostructures that incorporate multiple inorganic nanoparticles. To achieve this a synthetic strategy that maintains complete control over the incorporation of each subunit must be implemented. A surface-directed strategy such as that first suggest by Eddie Goldberg [12] provides a platform for incorporation of unique subunits at each assembly cycle with the need to utilize a minimum number of interacting pairs. The use of combinatorial techniques for the selection of these interacting pairs of protein surfaces appears to be a promising approach to the

identification of surfaces that will bind with high affinity and specificity to a target molecule.

The stability of interfaces constructed using combinatorial techniques needs to be investigated. It is quite possible that the non-covalent interactions formed will not be stable over the anticipated functional lifetime of the resultant nanostructure. If this is the case, the possibility of stabilization using chemical cross-linking will need to be explored. As long as the interfaces are stable over the period of time required to assemble the nanostructure, chemical cross linking may make it possible to extend their shelf life far beyond what would be possible with only non-covalent interactions.

In spite of the power of combinatorial techniques, their utility has certain limitations. We have shown here that in some cases it may not be possible to select from a combinatorial library a binding surface that can distinguish between two target molecules differing by a single amino acid. The mutations used in this study all involve substitutions of surface amino acids that are part of β-strands. These are the most rigid portions of FN3. Except in the case of mutation E9V, which led to de-stabilization of the domain, it is likely that all of the mutants constructed led to only local structural alterations. Extensive study of protein-protein interactions suggests that a more successful strategy would be to engineer mutations into flexible regions of the target proteins. This would result in a potentially larger structural impact of the mutation and provide additional flexibility for the formation of protein-protein interactions.

## REFERENCES


[1] K. Li, ,M. I. Stockman, and D.J. Bergman (2003) Self-Similar Chain of Metal Nanospheres as an Efficient Nanolens. Phys. Rev. Lett. 91, 227402-1-4.

[2] N.C. Seeman, and A. M. Belcher (2002) Emulating biology: Building nanostructures from the bottom up. Proc. Natl. Acad. Sci. 99,6451-6455.

[3] M. Sarikaya, C. Tamerler, A.K.-Y. Jen, K. Schulten and F. Baneyx (2003) Molecular biomimetics: nanotechnology through biology. Nature Materials 2, 577-585.

[4] S. Howorka (2007) Creating regular arrays of nanoparticles with self-assembling protein building blocks. J. Materials Chem. 17: 2049-2053.

[5] Y. Huang,, C.-Y. Chiang, S.K. Lee, Y. Gao,E. L. Hu, J. De Yoreo,. and A. M. Belcher (2005) Programmable Assembly of Nanoarchitectures Using Genetically Engineered Viruses Nanoletters 5, 1429-1434.

[6] V.A. Petrenko, G.P. Smith, X.Gong and T. Quinn (1996) A library of organic landscapes on filamentous phage. Protein Eng. 9, 797-801.

[7] D.J. Rodi, and L. Makowski (1999) Phage display technology: Finding a needle in a vast molecular haystack. *Current Opinion Biotech.* **10**, 87-93.

[8] S.R. Whaley, D.S. English, E.L. Hu, P.F. Barbara, and A.M. Belcher (2000) Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. Nature. 405:665-8.

[9] A. Koide, C.W. Bailey, X. Huang, and S. Koide (1998) The fibronectin type III domain as a scaffold for novel binding proteins. J. Mol. Biol. 284, 1141–1151.

[10] E Karatan, M. Merguerian, Z. Han, M. D. Scholle, S. Koide, and B. K. Kay (2004) Molecular Recognition Properties of FN3 Monobodies that Bind the Src SH3 Domain. Chemistry & Biology, 11, 835–844.

[11] M. D. Scholle, F. R. Collart and B. K. Kay (2004) *In vivo* biotinylated proteins as targets for phage-display selection experiments. Protein Expression and Purification 37, 243–252

[12] E. B. Goldberg (1999) Materials for the production of nanometer structures and se thereof. U.S. Patent 5,864,013.