PROTEINS: Structure, Function, and Bioinformatics 68:1–12 (2007)

# Review Article

# Principles of Nanostructure Design With Protein Building Blocks

**Chung-Jung Tsai,[1] Jie Zheng,[1] David Zanuy,[2] Nurit Haspel,[3] Haim Wolfson,[3] Carlos Alemán,[2] and Ruth Nussinov[1,4]***

[1]*Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, Maryland 21702*
[2]*Departament d'Enginyeria Química, E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politécnica de Catalunya, Diagonal 647, Barcelona E-08028, Spain*
[3]*School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel*
[4]*Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel*

***ABSTRACT*** **Currently there is increasing interest in nanostructures and their design. Nanostructure design involves the ability to predictably manipulate the properties of the self-assembly of autonomous units. Autonomous units have preferred conformational states. The units can be synthetic material science-based or derived from functional biological macromolecules. Autonomous biological building blocks with available structures provide an extremely rich and useful resource for design. For proteins, the structural databases contain large libraries of protein molecules and their building blocks with a range of shapes, surfaces, and chemical properties. The introduction of engineered synthetic residues or short peptides into these can expand the available chemical space and enhance the desired properties. Here we focus on the principles of nanostructure design with protein building blocks. Proteins 2007;68:1–12.** © **2007 Wiley-Liss, Inc.**

**Key words: nanostructure design; protein building blocks; self-assembly; protein folding; repeat proteins**

## INTRODUCTION

Nanostructure design offers promise. The ability to shape matter on the molecular scale is expected to lead to significant advances in diagnostics, imaging, and drug delivery.[1–8] To date, much of the nanotechnological design has been based on synthetic organic chemistry combined with polymer science. Complex macromolecules may have distinctive properties and functions,[5–9] which are potentially useful on various surfaces, in thin films, encapsulation, and drug delivery. Designs also include nanoscale sensors, which exploit the surface area of carbon nanotubes and semiconductors. Yet, the scale of biological systems is also inherently nano. Biology has evolved intricate pathways, elegant structures, and well-orchestrated mechanisms through which processes are regulated, and molecules associate and dissociate, all at the nanoscale level. Nanoscale machines in living cells span molecular motors, protein synthesis, chromosomal replication and transcription, and regulated membrane channels; they span the machinery of the cellular organization, transport and migration, and the fine control of the dynamics of these processes. Precise nanoscale associations are observed in a broad range of assemblies, including the complex ribosomal machine, the viral capsid, and the complicated system allowing the T4 virus to inject its genetic material into the cell. Despite the immense documentation of nano associations and their accurate performance in the living cell, much is still not understood, including the principles of the specific interactions and the precise quantitative regulation of the orchestrated processes and of the individual machines. Our lack of knowledge encompasses the characteristics of the cellular components, their quantities, interactions, and affinities; they encompass the inter-relationship between their variability, relative performance and precision; and they relate to flexibility and internal motion. Yet, in parallel to filling these essential gaps in our knowledge, we may start to use parts of the biological repertoire in design, with the challenging goal of predictable qualitative

Grant sponsor: National Institute of Health; Grant number: NO1-CO-12400; Grant sponsors: National Cancer Institute, Center for Cancer Research.

*Correspondence to: Ruth Nussinov, Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: ruthn@ncifcrf.gov

Received 5 September 2006; Revised 5 December 2006; Accepted 24 December 2006

Published online 3 April 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21413

manipulation of their combinations and expansion of their chemical space and repertoire toward specific functional roles. Eventually, the overall aim is to precisely quantitatively control the supramolecular assemblies.

Hence, biology offers a broad range of functional systems and proteins which already form nanoscaffolds and nanomachines. They perform with high precision and few mistakes, suggesting that they may serve as a natural material for design. Protein and nucleic acid molecules and their parts may be assembled in a bottom-up approach into novel architectures, which may differ from those found in nature.[10,11] To be able to control and manipulate these assemblies into stable functional devices the key is to follow the engineering principles observed in the living cell.[6,12–18] Successful nanostructure design with proteins follows the physical principles of folding, stability, and protein–protein interactions.[19–21] In hierarchical design strategies, the assembled protein monomers need to be in their preferred conformational states.[22] This is challenging, as we need to consider not only the assembly of the monomers into intermediate structures; in addition, some insight into the mechanisms through which the monomers assemble correctly is essential.

## PROTEIN FOLDING AND PROTEIN BUILDING BLOCKS

Inspection of the protein structural database (PDB)[23] illustrates a broad range of molecular shapes that can be used to create diverse scaffolds. The variability will increase if we take not only entire proteins but also their conformationally independent building block parts. For a successful design, the conformation of the building block when it is embedded in the nanostructure should reflect its natural tendency. A library of such building blocks would be useful for nanostructure design.

A few years ago, we have developed the building block folding model and a scheme to cut protein structures into their building block components.[24] The building block folding model is a "practical" model for protein folding. The model states that protein folding is a hierarchical process.[25] The basic unit from which a fold is constructed is the outcome of a combinatorial assembly of a set of building blocks. The building block is a highly populated fragment in a given protein structure. If we were to cut the building block from the protein chain, the most highly populated conformation of the resulting peptide in solution would likely be similar to that of the building block when in the native protein. However, while the conformations of most building blocks are preserved in the native structure, the mutually stabilizing association between the building blocks may nevertheless result in alternate conformations. In such cases, the conformations of the building blocks that we observe in the native protein structure differ from their original stand-alone conformations. A major challenge in nanodesign based on building block selection and association is to ensure that within the designed structure, the building block is in its preferred conformational state.

## STRUCTURAL DESIGN WITH PROTEIN BUILDING BLOCKS

The dissection of all the proteins in the structural data bank yields a library of protein fragments, ranging from complete tertiary folds to short building block parts of the chain. Each building block fragment has an associated favorable conformation. Clearly, proteins in the same family yield very similar building blocks. However, according to the building block folding model, since a building block fragment is a conformationally independent entity, building blocks from different protein families can also share similar building block structures.[26] This suggests that building blocks may be useful in designing proteins. Previously, we have tested this idea in a design of single chain proteins.[22] Our goal was to devise an algorithm to engineer proteins with naturally occurring folds and low sequence homology (Fig. 1). The sequence identity was kept as low as possible to avoid a homology bias. The algorithm implemented the strategy of designing a protein using relatively stable fragments, with a high population time. The fragments were selected by searching for local minima along the polypeptide chain. The two engineered proteins were constructed using protein building blocks sharing $\sim$20% and $\sim$25% amino acid sequence identities with their native counterparts, respectively. The stabilities of the engineered proteins have been tested by explicit water molecular dynamics simulations.

In protein design, we select a native protein scaffold, and substitute native building blocks by alternative ones, which may be taken from globally dissimilar protein structures. Hence, the local conformation of the substituted building block is similar between the target and the engineered protein; however, the global protein structures may differ (Fig. 1). In the nano design application, the shape may or may not occur naturally. The goal is to be able to devise a useful shape, and construct it by putting together building block parts. These building blocks will be extracted from the rich library of parts. Since the parts are conformationally independent units, they may be used in various constructs, and they are likely to retain their native conformations. In particular, this will be the case if they have a strong compact hydrophobic core. Their sizes may vary from $\sim$20 to more than 200 residues. Most frequently the building block size is around 40–60 residues. However, interestingly, even some short 8-mer sequence fragments were shown to have preferred conformations,[27] suggesting that these can be used to fill in cases with incomplete structural coverage in the PDB.

What are the advantages in using conformationally independent naturally-occurring units in design? From the technical standpoint, they should be simpler to synthesize and handle; smaller sizes afford a greater number of combinations, shapes, and tuning; faster synthesis is advantageous, since one can introduce more easily mutations with variants of native residues, or insert nonnative residues. Using libraries of building blocks with a range of shapes, surfaces, and chemical properties is attractive. Hence, a major criterion in nanodesign is a stable build-

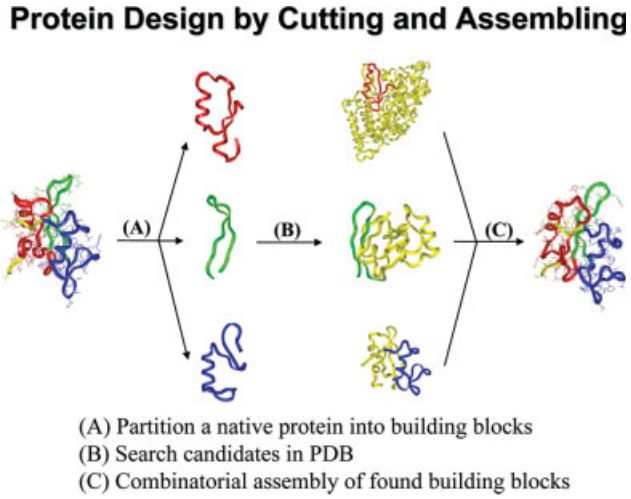


Figure 1.

Fig. 1. An illustration of single protein design as instructed by the building block folding model. Starting with a native protein scaffold, we substitute native building blocks by alternative ones which may be taken from globally dissimilar protein structures. It clearly shows in the figure that while the local conformation of the substituted building block is similar between the target and the engineered protein, the global structures of the proteins from which the building blocks are picked can be different. At the end, the overall sequence identity between the engineered protein and the native protein is less than 25%.

Fig. 2. (**a**) Two different molecular arrangements that allow the formation of Asn ladders: *On the left*, equatorial and axial projection of the chains arrangements in the protofibril of DFNKF aggregates. On the right, equatorial and axial projection of a left-handed β-helix (Chain A of *E. Coli* Galactoside acetyltransferase); (**b**) Schematic representation of the scheme followed to select β-helix based building blocks for designing nanostructures: highly symmetrical segments are taken in their native conformation, replicated, and used as modular templates to probe *in silico* the stabilities their potential assemblies.

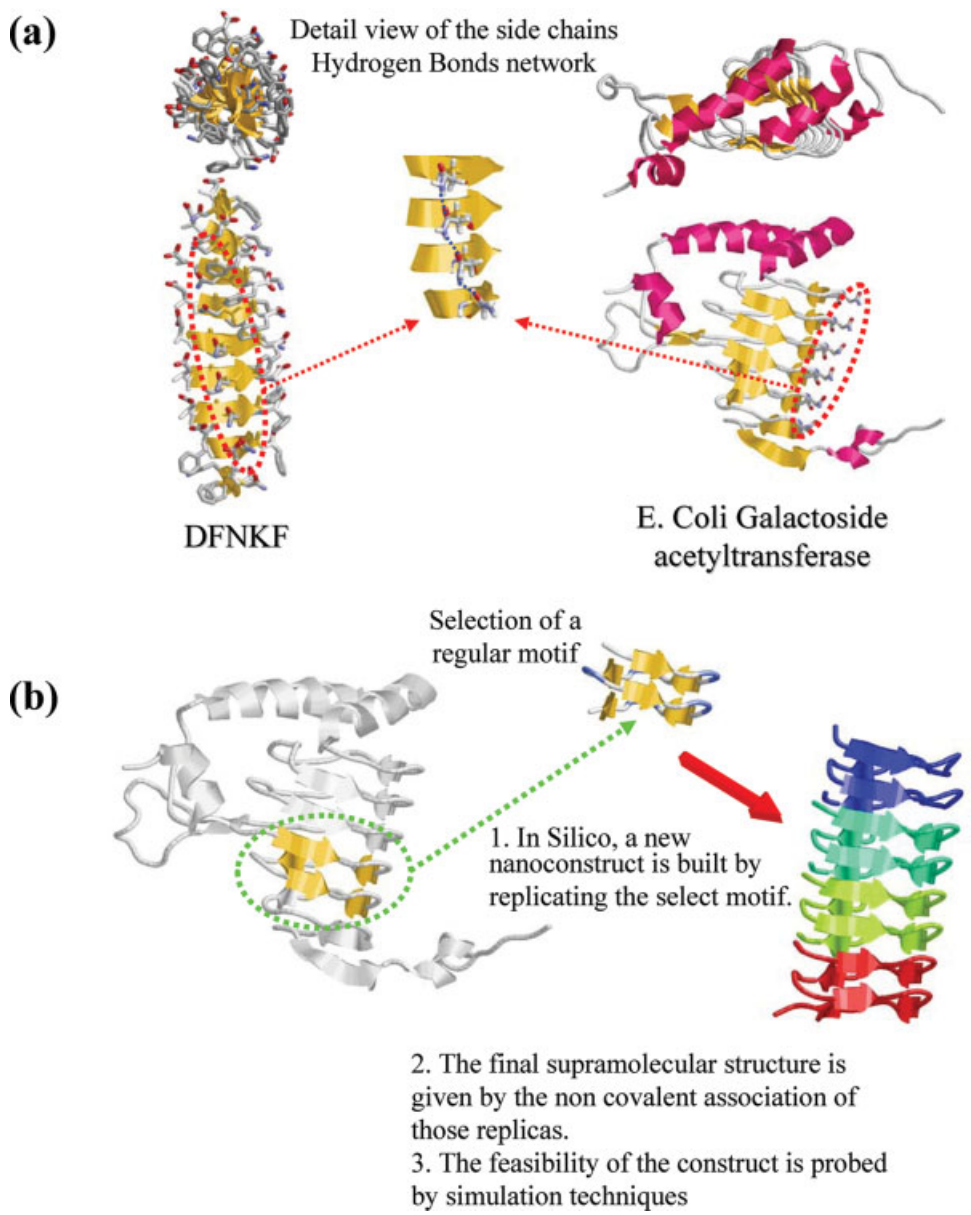


Figure 2.

ing block: the higher the stability of the building block, the higher the chances of success.

Nevertheless, in nanostructure engineering the requirement that building blocks be stable is insufficient for a successful design. The association between the building blocks should also be favorable. Building blocks mutually stabilize each other. Unlike the engineering of protein chains, in self assembly of stand-alone parts, there is no chain linkage. In single chains, while the association has to be stable to begin with, the covalent linkage further stabilizes the interaction between the building blocks and drives the protein to a particular conformation.

Experiment and computation have already shown that building blocks can be modulated, leading to new properties, and to creation of new binding surfaces for different engineered functions. In a remarkable experiment, Kohl et al.[28] have succeeded in creating artificial proteins from an ankyrin repeat, which were sufficiently stable to allow their crystallization and structural determination. The two consensus-designed ankyrin repeats with 88% sequence identity have different stabilities, surface charge distributions, and extents of hydrophobic residues exposure.[29] Simulations of these artificial proteins have shown that the difference in stabilities largely reflects the difference in the optimization of the electrostatic interactions.[30] Thus, a combination of experiment and simulations provides insight into the principles of building block design. Moreover, using the cloning protocol of Shiba et al.[31] these can be straightforwardly polymerized to further enhance the stability. Our on-going simulations are in agreement with this experimental observation for different β-helical repeats.

## WHAT ARE THE ELEMENTS THAT ARE IMPORTANT IN STABILIZING THE INTERACTION BETWEEN THE BUILDING BLOCKS?

The key elements in stabilizing the building blocks' associations are universal. As in protein–protein interactions, they involve hydrophobic patches, tight packing in some of the interface regions, good geometric complementarity, H-bonds, electrostatic interactions, and absence of unsatisfied charges. Since, however, proteins are flexible molecules, with frequent hinges between domains or building blocks, the interactions between the building blocks are usually not too tight. Evolution has optimized the proteins for function. In contrast, in nanodesign the inter-building block interactions should lend additional stability, since the function of these designs frequently necessitates rigidity. Hence, for successful nanodesign there are two criteria: First, the building block has to have a high population time. Second, the association needs to fit tightly geometrically and chemically. To further enhance the stability, substitutions by other natural residues or by conformationally constrained, nonnatural residues may be introduced in a chemical-biology strategy. As illustrated by the amyloid self assembly, strong favorable intermolecular "zipper" association is a key to rigidity, resistance, and durability under harsh conditions such as high temperature and nonphysiological pH. Designed nanostructure implies a regular ordered self-assembly as compared to an amorphous aggregate, implying a preferred mode of interaction. In the ordered amyloid assembly, the preferred inter-strand interactions within sheets derive from the backbone H-bonds and side-chain interactions. If the sheets consist of strands oriented in a parallel fashion, identical residues stack on top of each other, forming ladders. In the simulations, we have observed the Asn ladders to be the bottle-neck in the formation of the DFNKF amyloid aggregates.[32] A detailed picture of the supramolecular arrangement that makes possible the formation of such interaction network is shown in Figure 2(a). The amphiphilic nature of the peptide side chains promotes a rapid aggregation.[33] The presence of Asn residues conditions the strands orientation: When two strands associate in parallel, the Asn side chains are perfectly aligned and form hydrogen bonds. The extra stability induced by the formation of such interactions determines the final geometry of the fibril. Another example is the polyglutamine repeats in the huntingtin protein[34] and in the GNNQQNY peptide derived from the sup35.[35] In the parallel-stranded β-helices where the geometrical reference is defined by the strand direction and the main chain amide groups that form hydrogen bonds, consecutive rungs similarly display a tendency to have residues with similar chemistry stack on top of each other, in addition to the inter-strand hydrogen bonds, as can be observed in Figure 2(a), where the Asn residues help the formation of β-strands.

## WHICH BUILDING BLOCKS TO SELECT FOR THE NANODESIGN?

Here, we focus on three classes of design: In *Classes I* and *II* the building blocks are derived from repeat proteins. In *Class I* the design uses duplicates of a repeat, and assembles them following the native geometry to create a stack. No new interface is created. Here we have an essentially one-dimensional tube [Fig. 3(a)]. In *Class II* the assembly uses both the original native interface and creates a second interface [Fig. 3(b)]. Building blocks from repeat proteins appear a good choice, as they may be expected to *a priori* be inherently stable. Nevertheless, for this type, there is still the question of how favorable are their interactions.

An example of the *Class I* is the left-handed β-helix repeat. β-helices, provide an example of naturally occurring proteins, which contain a tubular or fibrillar motif in their folds. The fold of β-helical proteins contains a repetitive helical strand-loop motif, where each repeat contributes a strand to one or more parallel β sheet(s). The left handed β-helical fold is particularly useful: the tubular structure is regular and symmetric and is often stabilized by a network of interactions between similar residues in consecutive coils. The tubular nature of left handed β-hel-

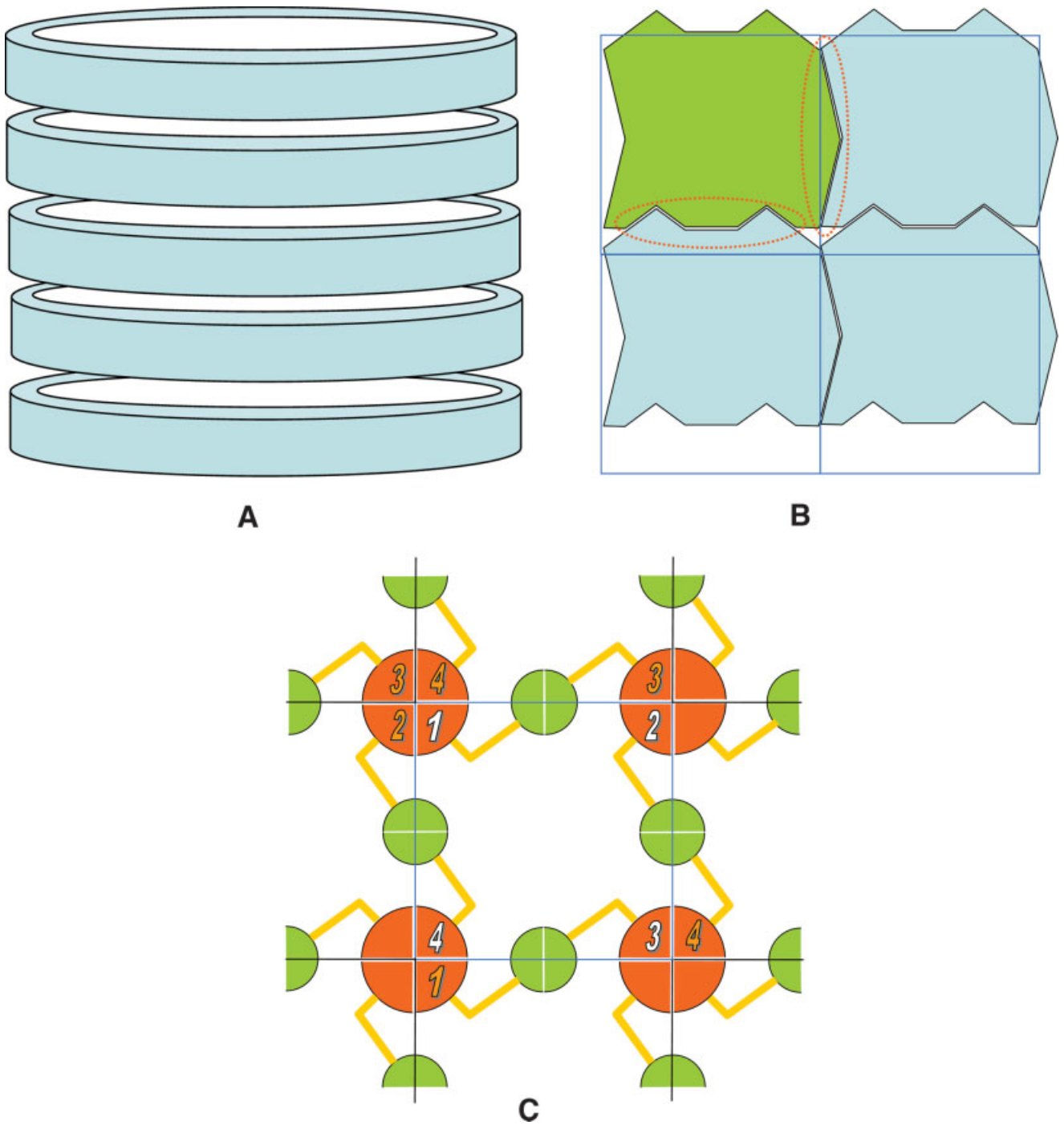


Fig. 3. A simplified diagram to illustrate the general scheme of arrangements in protein nanotubes. In all of the three Classes, only one unique entity was involved in the tube construction. Figure 3(**A**) depicts a tube with circular stacking between repeated units (*Class I*). The tube is simply built by an expansion in one-dimension with a repeating association interface along the tube axis. The (-helix nanotube falls in this category. In Figures 3(**B**) (*Class II*) and 3(**C**) (*Class III*), a tube is constructed by wrapping a planar sheet shaped by a 2-D lattice (highlighted in color blue). Here, there is only one entity in the lattice and without any symmetry. As illustrated in Figure 3(**B**), the tube requires two distinct interfaces (indicated by the two dotted orange ovals) to complete the 2-D expansion of a planar sheet. A hydrophobic sheet appears suitable to satisfy such a strict requirement due to the nonspecific characteristics of hydrophobic interactions. Both the FF nanotube and the surfactant-like nanotube belong to this category. Some of the repeat proteins could fall in this category with one native interface preserved and the other created by engineering. The third category utilizes oligomers being optimized by nature to craft the framework of a tube. In Figure 3(**C**), each chain contains two domains (represented in color orange and green) fused by a linker (yellow). There are four chains in a unit cell with *P4* symmetry and the asymmetric unit contains only one chain. The associations within the entire tube are the repeated tetramers and dimers. The tetramer is made out of four domains numbered from 1 to 4. Those domains which are inside the specified 2-D lattice are highlighted in white color. The dimers are seen between units (1,3) and (2,4). The diagram shown here is just a construct of many other topologies. The HIV-1 capsid CA protein nanotube is a representative member of this category. It gives a similar construct with a hexameric ring formed by the N-terminal domain of the CA protein instead of a tetramer as illustrated in Figure 3(C).

ical proteins makes them excellent candidates to be used as building blocks to construct fibrillar or tubular nanostructures without the need to perform many structural manipulations, as can be seen in Figure 2. In addition, their helical and symmetric structure makes them good candidates to be excised and tested as modules. Thus, nanotube design is guided by a systematic search for building blocks, selected based on their fold symmetry [Fig. 2(b)]. Selected segments are extracted and a tubular nanoconstruct is generated *in silico* by assembling several copies of the selected module. Using molecular dynamics simulations, we have constructed and tested 17 β-helix-based systems. Systems that are stable in the simulations still require experimental observation; however, systems that are unstable in the simulations are unlikely to observe experimentally. Simulations provide a first screen, limiting the range of constructs to be tested experimentally. Among the tested systems,[36] the construct based on the assembly of repeats of residues 131–165 of Galactoside acetyltransferase from *E. coli* (PDB 1krr, chain A) showed a remarkable ability to retain the assembled organization over a long period of simulation time (20–40 ns) under all the tested temperature and ionic strength conditions. Residues with similar chemical properties are stacked on top of each other, creating favorable interactions. In addition, backbone hydrogen bonds further lend stability to the compact repeats, each of which possesses a strong hydrophobic core. Thus, to ensure the association of such building blocks we first need to increase the population of the desired conformation of the monomer. Figure 4(a) provides the behavior of a single krr1 repeat, which shows a relative stability in terms of retaining its regular organization. In an attempt to further stabilize the repeat, we introduce a synthetic residue, 1-Amino-2,2-diphenylcyclopropanecarboxylic acid ($c_3$Dip), a cyclopropane analog of phenylalanine bearing two geminal phenyl substituents. The highly strained cyclopropane ring of this nonproteogenic amino acid[37,38] and the interactions between the rigidly positioned aromatic side chains and the peptide backbone induce strong stereochemical constraints.[38–43] This constrained amino acid is able to induce a double γ-turn (an incipient $2.2_7$-helix) in a linear dipeptide in the crystalline state.[43] Previously, we had parameterized it and fully characterized the conformational space of the trimer Ac-Gly-(*S*)c3Dip-Gly-NHMe (Zanuy et al., unpublished). The insertion of the Gly-(*S*)c3Dip-Gly segment in the β-helix repeat loop, instead of the residues 159–161, proved to provide an extra degree of conformational stability to the turn [Fig. 4(b)]. On the other hand, not all assemblies consisting of stacked β-helix repeats are stable. For instance when Gly-(*S*)c3Dip-Gly mutant repeats are assembled, the nanotube construct disintegrates, apparently due to steric hindrance of the consecutive repeats. Even repeats consisting of wild type sequences often lack the ability to assemble: Our simulations of four stacked repeats from another left-handed β-helix, *N*-acetylglucosamine 1-phosphate uridyltransferase GlmU, C-terminal domain from *E. coli*, PDB 1hv9, residues 296–329, indicate that this construct is unstable. 1hv9 single repeat is not a good building block by itself: as can be seen in Figure 4(c) after 4 ns the single repeat unfolds. Different single point mutations with nonproteogenic amino acids were tested to improve both the population of the proper fold and the assembly rate: Insertion of cyclopropane (1-amino-2-cyclopronanecarboxylic acid) at position 26 of the repeat, leads to the stabilization of the repeat [Fig. 4(d)]. Figure 4(e) illustrates that even though the conformation of the loop is restricted without steric hindrance, the association is still unstable; however, positioning the cyclopropane substituent at position 27 leads to a stable construct. This illustrates that for the association to be stable, both the building block should have a high population time, and the assembly should be favorable.

For *Class II* we provide two examples, the tetracopeptide repeats (TPR)[44,45] and the Toll-like receptor 3 ligand binding domain.[46] The potential of the repeat proteins in the design of mechanotransduction and nanodevices has been illustrated very elegantly. Applying atomic force microscopy on Ankyrin repeats, Marszalek and coworkers[47] have shown that tandem ankyrin repeats exhibit tertiary-structure-based elasticity and behave as a linear and fully reversible spring in single-molecule measurements. Further, they measured the refolding force of the unfolded domains. The TPR repeat[44,45] consists of two α-helices. The protein is composed of several covalently-linked repeats, forming a superhelix. As in the β-helix, sequentially adjacent repeats interact with each other. There are no interactions between sequentially-distant units. Of particular interest, while each repeat is degenerate, with no position invariance, a consistent pattern of key residues that are essential for the stability has been observed. These consensus residues have been used to design repeat proteins of high stability, with the stability increasing with the number of repeats.[44] Nevertheless, while each repeat appears stable on its own, unlike the β-helices which have a continuous hydrophobic core, the interactions between successive repeats are limited. The interfaces between the helices in the stacked repeats are limited in size, with discrete hydrophobic patches, and they are not as compact as in the β-helix. What holds these repeats together in addition to favorable inter-repeat interactions is the chain linkage, which stabilizes the association and drives it to a particular conformation. Cutting the chain linkages between the repeats of TPR, and stacking isolated repeats is likely to lead to unstable associations. However, it should be possible to enhance the interactions through interface design [Fig. 3(b)]. For the toll-like receptor 3 ligand binding protein the situation is different. Rather than forming a super-helical structure, the 23 repeating LRR units form a flat platform.[46] This arrangement raises the possibility of engineering horizontally-associated motifs into flat discs, which might be stacked on top of each other.

In *Class III* the building blocks are units (subunits or domains) taken from oligomeric proteins and covalently fused by a linker. In this case, the nanodesign consists of a designed repeat made of (at least) three units, the N-ter-

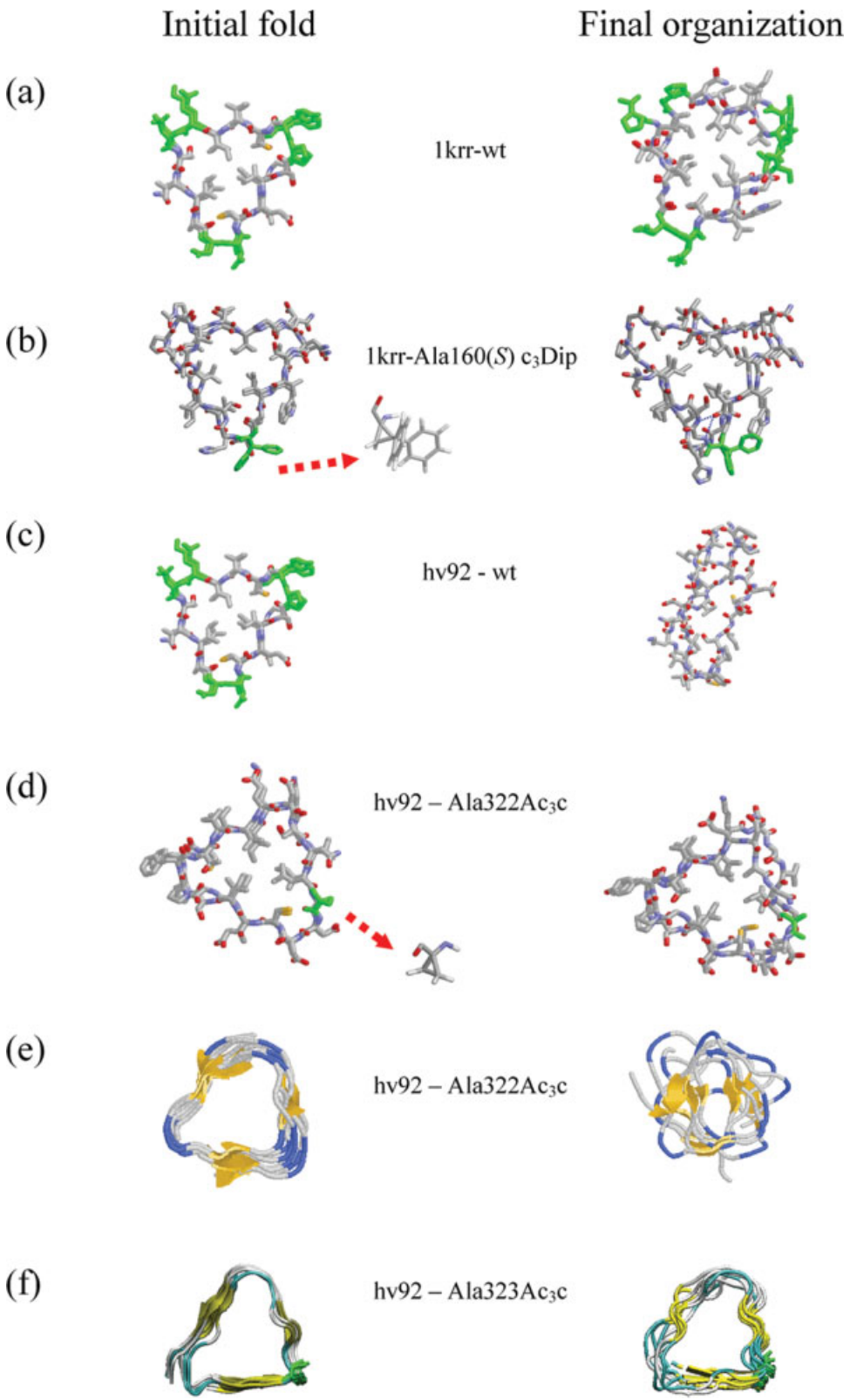


Fig. 4. Nanotubes created from β-helix repeats. In all cases, the image prior to the simulations is at the left and following the simulations on the right. (**a**) A single repetitive unit of wild type $1KRR_{131\text{-}165}$. (**b**) A single repetitive unit of $1KRR_{131\text{-}165}$ mutant A160(S)-$c_3$dip. See the extra stabilization introduced by the insertion of (S) c3Dip residue in 1krr segment: the restrained residue favors the formation of intraturn $C_{10}$ and $C_{13}$ hydrogen bonds formed between the C=O of Ile158 and the N-H of both Gly161 and Ser162. Hydrogen bonds are depicted using a blue dashed line. (**c**) A single repetitive unit of wild type $1HV92_{296\text{-}329}$. (**d**) A single repetitive unit of $1HV92_{296\text{-}329}$ mutant A322$Ac_3$c. (**e**) Four repetitive units of $1HV92_{296\text{-}329}$ mutant A322$Ac_3$c. Each unit has a single point $Ac_3$c mutation at Ala322. (**f**) Four repetitive units of $1HV92_{296\text{-}329}$ mutant A323$Ac_3$c. Each unit has a single point $Ac_3$c mutation at Ala323.

minal part of the repeat, the C-terminal part, and the linker. The two terminal parts will retain their native oligomeric interfaces in the design [Fig. 3(c)]. For cases like the one shown in Figure 3(c) (*Class III*), similar to the HIV-1 capsid (CA) protein[48] (see figure legend) two native interfaces are used. In this case, each building block is a domain of the protein, and the two domains are connected by a linker, also belonging to the CA protein. The N-terminal domain forms a hexamer association and the C-terminal domain a dimeric association. In the more general case, the building blocks are monomers taken from different oligomeric proteins. The design connects the two building blocks by a linker also selected from the PDB. Each of the monomers retains its native interactions with its sister oligomers, making it very stable. The remarkable work of Padilla et al.[49] provides experimental corroboration to the principles of nanostructure design with protein building blocks using such a strategy. It outlines and validates a scheme of fusing via a rigid linker, a subunit A which *in vivo* forms an oligomer $A_n$ to a subunit B which *in vivo* functions when in an oligomeric form, $B_m$. The fusion fixes the spatial orientation of A and B with respect to each other. Each subunit associates with its native partners, preserving the native interfaces. The authors illustrate the design of two fusion proteins leading to a cage and to a filament. This design is attractive since each of the units from which the repeat is constructed is in its conformationally preferred state and the interfaces are all native. Each unit would assemble to create its respective *in vivo* oligomeric state leading to the desired energy gap between this assembly and all alternate associations. This strategy resembles design with DNA. Each partner binds its native associate, whether specified by H-bonds (DNA) or by a combination of molecular interactions forces.

## PROTEIN FOLDING, REPEAT PROTEINS, AND PROTEIN BUILDING BLOCKS

The principles of folding and design are similar: as in folding, protein design relates to preferred conformational states and their interactions. In two interesting articles, Mello and Barrick[50] and Kajander et al.[51] have made the proposition that the folding of Ankyrin and TPRs (and probably all) repeat proteins can be understood in terms of the Ising model. This model describes systems that are composed of repeated units where each unit has only two possible states, either an up or a down state. Originally, the Ising model was used to describe systems of electrons, where each electron can have either a spin up or a spin down. For repeat proteins, in terms of the Ising model, each repeat unit is in either in a folded or an unfolded state and extensive contacts are only within a unit and between sequentially adjacent units. The Ising model describes the system by its density of states: thus, for example, since the model describes each of the units as either folded or unfolded, for a protein with four repeats, one repeat can be unfolded and three folded.

While the Ising model does not consider the folding/unfolding pathway, the building block folding model[24] describes the folding of the building blocks and the pathway of their association. Thus, in the building block folding model, we may have one building block which is folded and the other three which are also folded; however, the interactions between the first and its spatially adjacent ones might be nonnative. Such a situation is not treated by the Ising model: if all are folded, the possibility of nonnative interactions between them is not considered. Since the probability of such nonnative interactions between sequentially connected repeat units is low, the Ising model which essentially says that if the units are correctly folded it follows that the interactions between them are native-like, is highly successful for repeat proteins.[50,51]

In our case, for large nanostructure designs which need to be rigid, the mutual stabilization is crucial. This stabilization would be achieved via the self-assembly, where stabilization increases with the number of units in the assembly.

## SELF-ASSEMBLY: WORKING EXAMPLES

Above, we have used repeat proteins as working examples in a design with protein building blocks scenario. In the first β-helix repeat example (*Class I*), the designs are essentially one-dimensional. The β-helical shape and the native inter-repeat interface are retained, and no new interfaces are created [Fig. 3(a)]. Our goal is to improve existing interfaces to rigidify the association. For the other repeat proteins in nature,[31] each repeat also has only two neighbors, up and down (or, horizontally left and right, in the flat association of the LRR repeats[46]), similarly making it one-dimensional. However, as we show in the nanotube example of Figure 5, when a tube is constructed with an engineered shape, additional neighbors are needed for the construction making it essentially two-dimensional. Hence, if we are to use repeat proteins as units in *de novo* design, we have to introduce mutations to obtain additional favorable interactions [*Class II,* Fig. 3(b)]. The number of preferred associations should be limited, with an energy gap between the desired association and alternates, driving the self assembly toward the engineered nanotube shape. This is the native situation for *Class III* [Fig. 3(c)].

The tube construction scheme resembles that of carbon nanotubes.[48] There, tubes made of carbon atoms are linked together in rings of six to form hexagons. The flat sheets of hexagons are joined and rolled up to make a hollow tube (Fig. 5). There is an inherent geometrical distortion when wrapping a planar 2-D lattice onto the surface of a cylinder, an inevitable outcome of the geometrical transformation. After wrapping, the planar 2-D lattice is no longer planar; rather, it is a twisted 2-D lattice. The magnitude of the "distortion at the tube surface" is proportional to the curvature of the constructed tube. While we may introduce molecular distortion correction to create an atomic model closer to that on an actual realistic

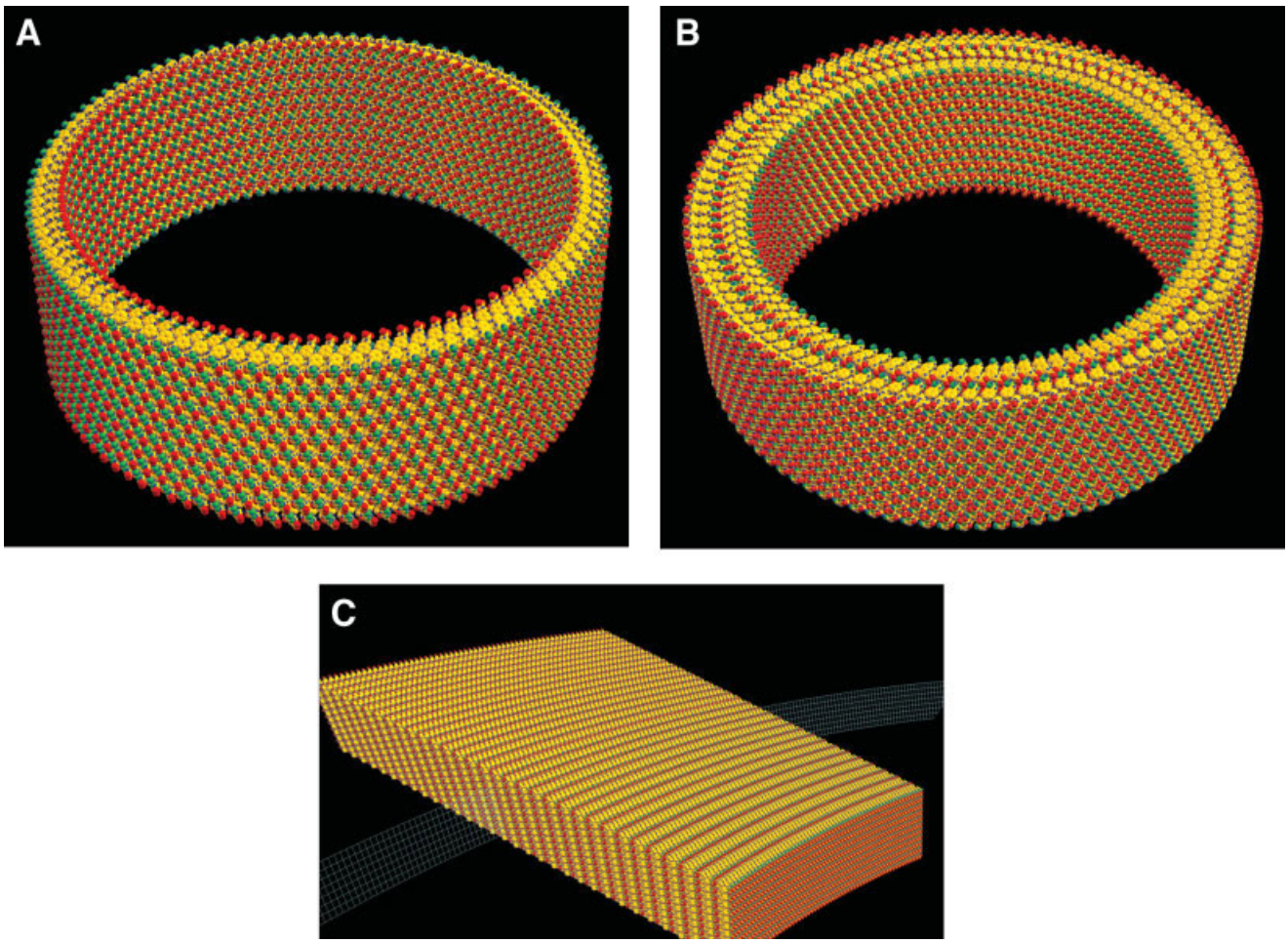


Fig. 5. Three nanotubes constructed with diphenylalanine peptides. (**a**) The nanotube shown here has a total of 2,000 diphenylalanine molecules with 10 circular stacking of 200 dipeptides in one complete round. The wall thickness of tube (about 16 Å) corresponds to the length of two dipeptides with four hydrophobic aromatic rings buried inside the tube. (**b**) With a total of 4,000 molecules shown here, the nanotube has a wall thickness corresponding to the length of four dipeptides. This tube was constructed by a procedure of tube lamination, in which the two diphenylalanines were duplicated along the axis perpendicular to the tube axis to double the wall thickness. (**c**) A slice of tube shown here with a total of (64*61*11) molecules is very close to the smallest observed diphenylalanine nanotube with an outer diameter of 4,000 Å, an inner diameter of 3,000 Å, and a wall thickness of 500 Å. The tube was created by repeating the thickness doubling lamination procedure 5 times with an enlarged tube radius.

nanotube, the tube will always display a higher density toward the tube center and a lower density away from the tube center, explaining why the diameter of a protein nanotube always has a lower and upper limit. If the diameter of the tube is too short and the wall thickness of the tube is too wide, the portion inside the tube will be too crowded. On the other hand, the packing of the outside part of the tube will be too loose to hold the structure. This leads to the tube-construction scheme: If the wrapping system can be used to describe the structure of the repeat protein, additional interactions may be introduced. However, these interactions may not be favorable otherwise a nanostructure might have formed which would create a problem for the organism. A wrapping system that preserves the native repeat unit association and at the same time allows mutations to enhance favorable interactions into a self-assembled nanotube may lead to favorable design.[48,52] On its own, using the wrapping system, with a single interface a repeat such as the TPR's would create a superhelix. To fill the tube up, a second interface is needed (see Fig. 6), as in the case of the lipid nanotubes[53–55] A possible second interface [Fig. 3(b)] to select is a crystal packing interface.

## PRINCIPLES OF NANOSTRUCTURE DESIGN WITH PROTEIN BUILDING BLOCKS: CONCLUSIONS

In the β-helix based design, the shape is given: it is that of the β-helix. The design attempts to modulate the β-helix properties, depending on the functional target. In contrast, for the shape-specified nanotube construction scheme we develop a new, geometry-guided construction strategy. In both cases, the nanodesign uses knowledge obtained from nature. For the nanotube design the success rate of the design should be higher if we select building blocks retaining their native associations. For the *Class II* repeat proteins, we retain one native interface.

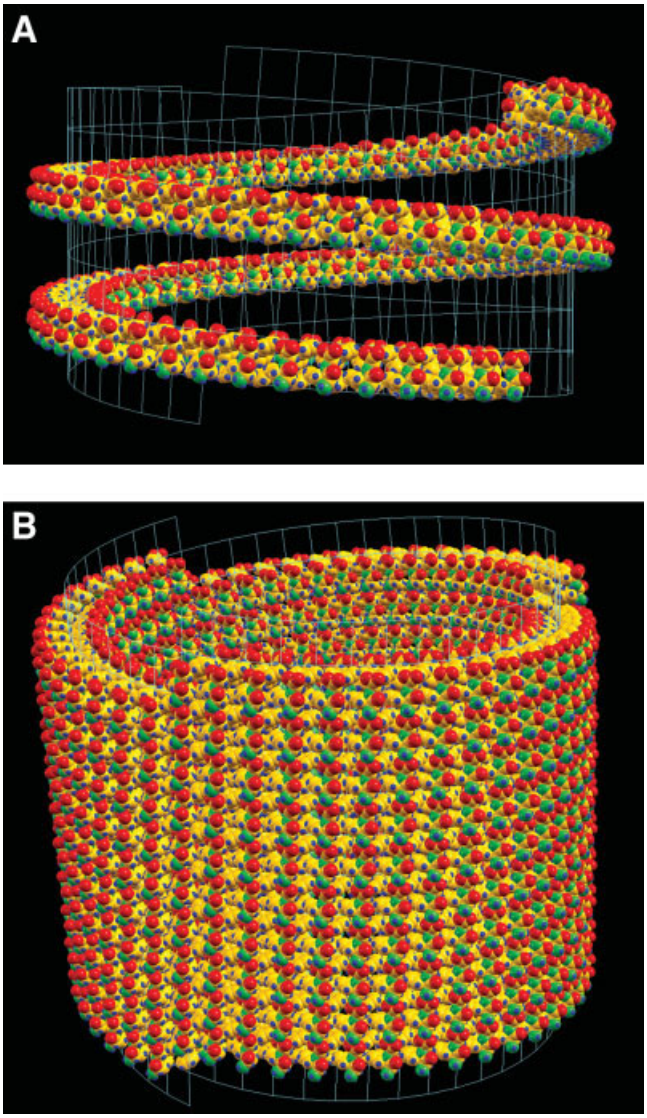


Fig. 6. An illustration to emphasize that two repeating interfaces are needed to construct a nanotube under the 2-D lattice wrapping scheme for *Classes II* and *III*. In Figure 6(**A**), a superhelical structure [embedded in a nanotube as shown in Figure 6(**B**)] is clearly seen with the structure constructed by using only one repeating interface as indicated in the simplified drawing of Figure 3(B). This is similar to the one-dimensional association in repeat proteins which also exhibit a superhelical structure. The nanotube shown in Figure 6(B) is made out of three such superhelical structures as shown in Figure 6(A). The association between the three superhelical structures is exactly a second repeating interface as implied by the wrapping 2-D lattices (line drawn in cyan). To take advantage of repeat proteins in protein nanotube design, one has to focus on creating such a second interface while preserving the native interface that provides the superhelical structure.

The challenge is to create new interactions to form the tube while preserving the repeat interactions in the design.

For a successful building block-based nanodesign two criteria should be satisfied: first, the conformation of the building block in the constructed scaffold should be similar to its preferred conformation when isolated in solution; that is, the building block should be in its native conformational state and have a high population time. Even if unstable, association will enhance its stability increasing with the number of repeats.[44] An *a priori* "good" building block could be further enhanced in a chemical biology strategy. By selectively substituting natural residues for synthetic ones whose preferred conformations fit the natural tendencies of the target polypeptide backbone it is possible to constrain backbone flexibility, rigidifying, and stabilizing the building blocks. At the same time the choice of the synthetic residue should not destabilize the association. Second, nanotubes essentially involve two dimensional lattices, resembling a flat sheet rolled into a tube shape. Thus, when using building blocks in the creation of the sheet new interfaces need to be engineered. The designed new interface between the building blocks should be highly favored, with few preferred alternates, leading to an energy gap driving the self-assembly into the desired state. In the DNA origami,[16] there is a single highly likely association specified by the base pairing. That, however, is not the case for proteins.[56,57]

Using protein building blocks in nanodesign presents a number of advantages. Proteins and their building blocks provide a vast repertoire of shapes and chemical properties. Their self-assembly may create potential scaffolds for numerous purposes. The structures of an increasing number can be found in the structural databases. The building blocks we observe in natural proteins and their interfaces have been optimized by evolution through billions of years. Nevertheless, their optimization has been for particular functions under given sets of living (or, physiological) conditions. For designed targeted function, while a good starting point, it may not suffice. Among the potential building blocks, repeat units taken from repeat proteins appear a good choice, as they already fulfill one requirement, favorable interaction in one dimension, although this interaction may not be sufficiently stable when the backbone linkage is cut. Selected building blocks with high population times enhanced by synthetic residues, particularly when the native interfaces are retained, may provide a reasonable start. With these conditions, a set of candidate designs may be created and tested by simulations, limiting the otherwise large number of potential constructs and accelerating the nanodesign.

Hence, protein-based nanotube design may involve (i) circular stacking preserving the native interfaces (the β-helix serves as an example; [*Class I,* Fig. 3(a)]; (ii) a repeat picked from a repeat protein with two interfaces, one of which is native [*Class II,* Fig. 3(b)]; (iii) creating a repeat by fusing building blocks [*Class III,* Fig. 3(c)]. Here, the terminal building blocks are picked from oligomeric proteins, retaining their native oligomeric interfaces. We note that case (ii) may allow tube lamination, as in the diphenylalanine peptide [Fig. 5(c)].

## ACKNOWLEDGMENTS

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400. The content of this publication does not necessar-

ily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This Research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

## REFERENCES

1. Ferrari M. Cancer nanotechnology: opportunities and challenges. Nature Rev Cancer 2005;5:161–171.
2. Ellis-Behnke RG, Liang YX, You SW, Tay DK, Zhang SG, So KF, Schneider GE. Nano neuro knitting: peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. Proc Natl Acad Sci USA 2006;103:5054–5059.
3. Lomander A, Hwang WM, Zhang SG. Hierarchical self-assembly of a coiled-coil peptide into fractal structure. Nano Lett 2005;5: 1255–1260.
4. Yokoi H, Kinoshita T, Zhang SG. Dynamic reassembly of peptide RADA16 nanofiber scaffold. Proc Natl Acad Sci USA 2005;102: 8414–8419.
5. Zhang SG. Emerging biological materials through molecular self-assembly. Biotechnol Adv 2002;20:321–339.
6. Zhang SG. Fabrication of novel biomaterials through molecular self-assembly. Nat Biotechnol 2003;21:1171–1178.
7. Reches M, Gazit E Designed aromatic homo-dipeptides: formation of ordered nanostructures and potential nanotechnological applications. Phys Biol 2006;3:S10–S19.
8. Yemini M, Reches M, Gazit E, Rishpon J. Peptide nanotube-modified electrodes for enzyme-biosensor applications. Anal Chem 2005;77:5155–5159.
9. Hawker CJ, Wooley KL. The convergence of synthetic organic and polymer chemistries. Science 2005;309:1200–1205.
10. Nam KT, Kim DW, Yoo PJ, Chiang CY, Meethong N, Hammond PT, Chiang YM, Belcher AM. Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. Science 2006;312:885–888.
11. Huang Y, Chiang CY, Lee SK, Gao Y, Hu EL, De Yoreo J, Belcher AM. Programmable assembly of nanoarchitectures using genetically engineered viruses. Nano Lett 2005;5:1429–1434.
12. Nanda V, Rosenblatt MM, Osyczka A, Kono H, Getahun Z, Dutton PL, Saven JG, DeGrado WF. De novo design of a redox-active minimal rubredoxin mimic. J Am Chem Soc 2005;127: 5804–5805.
13. Dwyer MA, Hellinga HW. Periplasmic binding proteins: a versatile superfamily for protein engineering. Curr Opin Struct Biol 2004;14:495–504.
14. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. Computational redesign of protein–protein interaction specificity. Nat Struct Mol Biol 2004;11:371–379.
15. Lassila JK, Keeffe JR, Oelschlaeger P, Mayo SL. Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. Protein Eng Des Sel 2005;18: 161–163.
16. Rothemund PW. Folding DNA to create nanoscale shapes and patterns. Nature 2006;440:297–302.
17. Seeman NC. Structural DNA nanotechnology: an overview. Methods Mol Biol 2005;303:143–166.
18. Seeman NC. DNA enables nanoscale control of the structure of matter. Q Rev Biophys 2006;6:1–9.
19. Rajagopal K, Ozbas B, Pochan DJ, Schneider JP. Probing the importance of lateral hydrophobic association in self-assembling peptide hydrogelators. Eur Biophys J 2006;35:162–169.
20. Valery C, Artzner F, Robert B, Gulick T, Keller G, Grabielle-Madelmont C, Torres ML, Cherif-Cheik R, Paternostre M. Self-association process of a peptide in solution: from β-sheet filaments to large embedded nanotubes. Biophys J 2004;86:2484–2501.
21. Tang M, Waring AJ, Hong M. Intermolecular packing and alignment in an ordered β-hairpin antimicrobial peptide aggregate from 2D solid-state NMR. J Am Chem Soc 2005;127:13919–13927.
22. Tsai HH, Tsai CJ, Ma BY, Nussinov R. In silico protein design by combinatorial assembly of protein building blocks. Protein Sci 2004;13:2753–2765.
23. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58:899–907.
24. Tsai CJ, Maizel JV, Nussinov R. Anatomy of protein structures: visualizing how a one-dimensional protein chain folds into a three-dimensional shape. Proc Natl Acad Sci USA 2000;97: 12038–12043.
25. Baldwin RL, Rose GD. Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci 1999;24:26–33.
26. Haspel N, Tsai CJ, Wolfson H, Nussinov R. Reducing the computational complexity of protein folding via fragment folding and assembly. Protein Sci 2003;12:1177–1187.
27. Ho BK, Dill KA. Folding very short peptides using molecular dynamics. PLo S Comp Biol 2006;2:e27.
28. Kohl A, Binz HK, Forrer P, Stumpp MT, Pluckthun A, Grutter MG. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci USA 2003;100:1700–1705.
29. Binz HK, Kohl A, Pluckthun A, Grutter MG. Crystal structure of a consensus-designed ankyrin repeat protein: implications for stability. Proteins Struct Funct Genet 2006;65:280–284.
30. Yu H, Kohl A, Binz HK, Pluckthun A, Grutter MG, van Gunsteren WF. Molecular dynamics study of the stabilities of consensus designed ankyrin repeat proteins. Proteins Struct Funct Genet 2006;65:285–295.
31. Shiba K, Takahashi Y, Noda T. Creation of libraries with long ORFs by polymerization of a microgene. Proc Natl Acad Sci USA 1997;94:3805–3810.
32. Tsai HH, Reches M, Tsai CJ, Gunasekaran K, Gazit E, Nussinov R. Energy landscape of amyloidogenic peptide oligomerization by parallel-tempering molecular dynamics simulation: significant role of Asn ladder. Proc Natl Acad Sci USA 2005;102:8174–8179.
33. Zanuy D, Haspel N, Tsai HH, Ma BY, Gunasekaran K, Wolfson HJ, Nussinov R. Side chain interactions determine the amyloid organization: a single layer beta-sheet molecular structure of the calcitonin peptide segment 15–19. Phys Biol 2004;1:89–99.
34. Zanuy D, Gunasekaran K, Lesk AM, Nussinov R. Computational study of the fibril organization of polyglutamine repeats reveals a common motif identified in β-helices. J Mol Biol 2006;358:330–345.
35. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D. Structure of the cross-β spine of amyloid-like fibrils. Nature 2005;435:773–778.
36. Haspel N, Zanuy D, Aleman C, Wolfson H, Nussinov R. De novo tubular nanostructure design based on self-assembly of β-helical protein motifs. Structure 2006;14:1137–1148.
37. Toniolo C, Crisma M, Formaggio F, Peggion C. Control of peptide conformation by the Thorpe-Ingold effect (C-α-tetrasubstitution). Biopolymers 2001;60:396–419.
38. Aleman C, Jimenez AI, Cativiela C, Perez JJ, Casanovas J. Influence of the phenyl side chain on the conformation of cyclopropane analogues of phenylalanine. J Phys Chem B 2002;106: 11849–11858.
39. Jimenez AI, Cativiela C, Aubry A, Marraud M. β-turn preferences induced by 2,3-methanophenylalanine chirality. J Am Chem Soc 1998;120:9452–9459.
40. Jimenez AI, Vanderesse R, Marraud M, Aubry A, Cativiela C. Folding types of dipeptides containing the diastereoisomeric cyclopropanic analogues of phenylalanine. Tetrahedron Lett 1997;38: 7559–7562.
41. Jimenez AI, Cativiela C, Marraud M. A γ-turn induced by a highly constrained cyclopropane analogue of phenylalanine (c(3)diPhe) in the solid state. Tetrahedron Lett 2000;41:5353–5356.
42. Casanovas J, Jimenez AI, Cativiela C, Perez JJ, Aleman C. *N*-acetyl-*N*′-methylamide derivative of (2S,3S)-1-amino-2,3-diphenylcyclopropane-carboxylic acid: Theoretical analysis of the conformational impact produced by the incorporation of the second phenyl group to the cyclopropane analogue of phenylalanine. J Org Chem 2003;68:7088–7091.

43. Jimenez AI, Ballano G, Cativiela C. First observation of two consecutive γ turns in a crystalline linear dipeptide. Angew Chem Int Ed 2005;44:396–399.
44. Main ER, Lowe AR, Mochrie SG, Jackson SE, Regan L. A recurring theme in protein engineering: the design, stability and folding of repeat proteins. Curr Opin Struct Biol 2005;15:464–471.
45. Main ER, Stott K, Jackson SE, Regan L. Local and long-range stability in tandemly arrayed tetratricopeptide repeats. Proc Natl Acad Sci USA 2005;102:5721–5726.
46. Bell JK, Botos I, Hall PR, Askins J, Shiloach J, Segal DM, Davies DR. The molecular structure of the Toll-like receptor 3 ligand-binding domain. Proc Natl Acad Sci USA 2005;102: 10976–10980.
47. Lee G, Abdi K, Jiang Y, Michaely P, Bennett V, Marszalek PE. Nanospring behaviour of ankyrin repeats. Nature 2006;440:246–249.
48. Tsai CJ, Zheng J, Nussinov R Designing a nanotube using naturally occurring protein building blocks. PLo S Comp Biol 2006;2:e42.
49. Padilla JE, Colovos C, Yeates TO. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc Natl Acad Sci USA 2001;98:2217–2221.
50. Mello CC, Barrick D. An experimentally determined protein folding energy landscape. Proc Natl Acad Sci USA 2004;101: 14102–14107.
51. Kajander T, Cortajarena AL, Main ERG, Mochrie SGJ, Regan L. A new folding paradigm for repeat proteins. J Am Chem Soc 2005;127:10188–10190.
52. Tsai CJ, Zheng J, Alemán C, Nussinov R. Structure by design: from single proteins and their building blocks to nanostructures. Trends Biotechnol 2006;24:449–454
53. Selinger JV, Spector MS, Schnur JM. Theory of self-assembled tubules and helical ribbons. J Phys Chem B 2001;105:7157–7169.
54. Thomas BN, Lindemann CM, Clark NA. Left- and right-handed helical tubule intermediates from a pure chiral phospholipid. Phys Rev E 1999;59:3040–3047.
55. Schnur JM. Lipid tubules—a paradigm for molecularly engineered structures. Science 1993;262:1669–1676.
56. Song G, Lazar GA, Kortemme T, Shimaoka M, Desjarlais JR, Baker D, Springer TA. Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion. J Biol Chem 2006;281:5042–5049.
57. Huang PS, Love JJ, Mayo SL. Adaptation of a fast Fourier transform-based docking algorithm for protein design. J Comput Chem 2005;26:1222–1232.