Published in final edited form as:

Protein Pept Lett. 2011 April ; 18(4): 362–372. doi:10.2174/092986611794653996.

Symmetry-Based Self-Assembled Nanotubes Constructed Using Native Protein Structures: The Key Role of Flexible Linkers

Idit Buch¹, Chung-Jung Tsai², Haim J. Wolfson³, Ruth Nussinov^{1,2,*}

¹Department of Human Genetics, Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

²SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI – Frederick, Bldg 469, Frederick, MD 21702, USA

³School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

Abstract

We construct nanotubes using native protein structures and their native associations from structural databases. The construction is based on a shape-guided symmetric self-assembly concept. Our strategy involves fusing judiciously-selected oligomerization domains via peptide linkers. Linkers are inherently flexible, hence their choice is critical: they should position the domains in three-dimensional space in the desired orientation while retaining their own natural conformational tendencies; however, at the same time, retain the construct stability. Here we outline a design scheme which accounts for linker flexibility considerations, and present two examples. The first is HIV-1 capsid protein, which *in vitro* self-assembles into nanotubes and conical capsids, and its linker exists as a short flexible loop. The second involves novel nanotubes construction based on antimicrobial homodimer Magainin 2, employing linkers of distinct lengths and flexibility levels. Our strategy utilizes the abundance of unique shapes and sizes of proteins and their building blocks which can assemble into a vast number of combinations, and consequently, nanotubes of distinct morphologies and diameters. Computational design and assessment methodologies can help reduce the number of candidates for experimental validation. This is an invited paper for a special issue on protein dynamics, here focusing on flexibility in nanotube design based on protein building blocks.

Keywords

Building blocks; molecular dynamics simulations; nanotube design; self-assembly; oligomerization domain; symmetry

1. INTRODUCTION

Bio-nanotechnology aims to design novel biomaterials and molecular devices, often via self-assembly. In nature, proteins often self-assemble, spontaneously organizing in stable

* Address correspondence to this author at the SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, Bldg 469, Rm 151, NCI-Frederick, Frederick, Maryland 21702, USA; Tel: 301-846-5579; Fax: 301-846-5598; ruthnu@helix.nih.gov.

structures through non-covalent interactions. These create large complexes with well-defined structures that have specific functions [1–4]. Exploiting the ability of proteins to self-assemble in a specific manner can be useful for construction of novel structures [5–8]. Recently, there have been numerous reports describing the use of natural and artificial peptides in nanodesign [9–12]. Due to advances in peptide synthesis and molecular engineering techniques [13,14], peptide self-assembly can become a favorable route to obtain single or associated tubes, fibers, micelles and vesicles. Simulation methods and efficient modeling algorithms coupled with protein databases can accelerate the design via fast probing of many models, limiting experimentations to those with higher probability of success. While the number of methods increases, few general strategies address structural architectures and symmetries.

Padilla *et al.* [15] combined symmetric protein components for engineering self-assembling nanostructures with the assumption that natural proteins form from preferred self association of domains and that dimers, trimers, and tetramers are common. Hence, following certain geometric rules, two such domains can be covalently connected through a peptide linker into a single larger monomer. Such a fused monomer is therefore made of three segments, where those at both ends have a strong tendency to associate with identical copies of themselves into a symmetric nanostructure. In this manner, a vast range of symmetric nanostructures can be designed, such as cages, shells, unbounded layers, and filaments. The final 3D geometry of the molecular architecture depends on the oligomeric order of the original domains and on the preferred orientation of a helical linker. Raman *et al.* [16] substantiated this strategy, but introduced a new approach for fusing homo-oligomers: two monomers can be fused if both have cysteines adjacent to their termini. The termini are linked covalently via two glycines, and oxidative conditions lead to formation of a disulphide bond between the monomers. This bond rigidifies a desired orientation between the original monomers. Usui *et al.* incorporated cysteines into the binding elements by site-directed mutagenesis [17].

Tsai *et al.* presented an *in silico* strategy for constructing a protein nanotube [18]. The strategy involves optimal mapping of candidate protein building blocks (BBs) into a nanotube shape, leading to a complete atomic model of the protein nanotube. The procedure involves wrapping a planar sheet onto a tube surface. If the arrangement of the building blocks has a 2D repeating pattern, then the planar sheet can be shaped by the repeating 2D lattice. The lattice can be described by the three lattice constants $\{a, b, \gamma\}$. Two additional integers $\{n_1, n_2\}$ define the wrapping of the planar sheet onto the tube surface, where n_1 defines the number of cells needed to wrap a full round of the tube, and n_2 defines the number of cells shifted after one complete round. This computational strategy was based on known experimental observed data [19–21].

Since self-assembled nanostructures are generally huge systems, earlier assessment methods that used MD simulations were limited by running time, and the protein region of interest [22,23]. Recently, we devised an all-atom molecular dynamics (MD) simulation method for validating protein nanotubes [24]. The method utilizes the observation from nature that a protomeric unit-cell self-assembles according to a predetermined symmetry. The mathematical transformations of the unit-cell to its identical neighbors are calculated in the preprocessing stage. The full atomic-coordinates of the unit-cell only are preserved

throughout the simulation, and the cell non-bonded interactions with its neighboring images are constantly considered. This method allows accurate simulation of protein nanotubes in reasonable times.

Here we use a new protocol for designing protein nanotubes from peptide building blocks based on experimental [15–17] and computational [18,24] works. Our method is based on two assumptions: (1) that a self-assembled peptide nanostructure can be virtually disassembled into identical asymmetric units that interact with each other through symmetry operations [25,26]; and (2) that *most* self-assembled lamellar sheets tend to wrap into cylindrical or fibrillar structures since a closed geometrical shape is energetically more favorable than its open planar state [1,27,28]. The revolving of a lamellar sheet into a tube, fibril or vesicle is dependent on the monomer concentration. The higher the concentration, the more globular is the geometry [29,30]. Since at low concentration a lamellar sheet would form a tube or a fibril, we focus on these geometries. Fig. (1) outlines the construction protocol. We demonstrate the strong relationship between our generalized protocol and the construction procedure of the human immunodeficiency virus type 1 (HIV-1) capsid protein [21], and construct novel protein nanotubes, based on the antimicrobial protein Magainin 2 [31]. Whereas in the HIV-1 capsid protein the linker is a short six amino-acid flexible loop, in our designed nanotube constructs we model different linkers of distinct lengths, flexibilities and dynamics, aiming to optimize geometrical and thermodynamic stability.

2. METHODS

Our design and construction protocol has several stages. First, we define a set of symmetries of a 2D lattice cell. This cell will eventually wrap around a cylinder surface. For each lattice symmetry, we select homo-oligomers from the PDB [32] and a specific linker to fuse the two oligomerization domains (ODs). In the last setup stage, we construct a unit-cell from the fused ODs and optimize the cell structure before wrapping it with identical entities into a tube. Finally, using CHARMM [33] we assess the stability of the model [24]. All minimizations, simulations and analysis use CHA-RMM22 force-field.

2.1. Defining Main Lattice Symmetries

Fig. (2) presents symmetric schemes of a unit-cell made of fused building blocks. In these typical schemes, the homo-oligomers that build a lattice cell include dimers, trimers, tetramers and hexamers. Table 1 summarizes the properties and constraints of each scheme.

2.2. Selecting Oligomers and Linkers

The chain size of the selected oligomers is relatively small (here < 40 amino acids) due to the nanotube assessment times required by MD simulations. The linker that binds two ODs is short (here < 15 amino acids). In the setup phase the domains are placed on a plane according to the chosen symmetric scheme (Fig. (2)) with a minimal Euclidean distance between two adjacent termini, and no steric clashes. To model a flexible linker, we used ModLoop [34], which models a short peptide sequence between two fixed termini, or Modeller [35], such that the linker retains its predetermined conformation. We repeated this binding process for the construction of the whole unit-cell, and then minimized using

“Steepest Decent” followed by “ABNR” (Adopted Basis Newton-Raphson). These were run by CHARMM [33].

2.3. Determining the Cell Parameters

The lattice constants $\{a, b, \gamma\}$ affect the non-bonded interactions of the unit-cell with its identical neighboring images; setting too high values would cause weak non-bonded interactions, whereas setting lower values would cause geometric clashes or overlaps. We placed the unit-cell construct on the x-y plane so that its principal geometric axis coincides with the X-axis and the next largest coincides with the Y-axis. We then set $a = X_{\max} - X_{\min}$, $b = Y_{\max} - Y_{\min}$ and $c = Z_{\max} - Z_{\min}$. The constant γ is defined by the chosen lattice symmetry; $\gamma = 90^\circ$ for C2 or C4 symmetry (Fig. (2a and b)), and $\gamma = 120^\circ$ for P6 symmetry (Fig. (2c and d)). We solvated the constructed unit-cell in bulk water with dimensions corresponding to $\{a, b\}$ while adding 9Å to both sides of c and counter-ions to a neutral pH. Then, after a few hundreds of local minimization steps, we ran constant pressure simulations with standard periodic boundary conditions for 0.5ns using NAMD [36]. We used the Nosé-Hoover Langevin piston method for pressure control [37], and the Langevin temperature control at 300K. This allowed a fine approximation of the volume of the unit-cell, from which we drew the refined lattice constants $\{a, b\}$. The result was used as an input for the next stage.

To wrap the unit-cell with its identical neighboring images into a cylinder, two additional wrapping parameters n_1 and n_2 had to be determined [18]. Based on the obtained lattice cell parameters $\{a, b, \gamma\}$, we ran local minimizations using cylindrical periodic boundary conditions with ascending n_1 values. For each n_1 we calculated the potential energy of the unit-cell. Fig. (3) depicts the potential energy values taken for different values of n_1 in two examples of observed nanotubes; the HIV-1 CA protein, and the surfactant-like peptide AcV6D [19]. As n_1 increases the total energy decreases exponentially, indicating that a planar arrangement of cells is more favorable than cylindrical. However, our local minimization does not consider the non-bonded interactions between the two marginal cells that account for the closure of the cylinder. Instead, the cylindrical periodic boundary conditions only consider the non-bonded interactions of our unit-cell with its closest neighboring cells. Therefore, we picked the minimal value of n_1 for which the energy nearly reaches a plateau, as seen in both examples of Fig. (3). Once n_1 was obtained, we used the above method to find n_2 . Since n_2 defines the helicity of the tube, its value is limited by n_1 ; a higher value suggests a more fibril-like structure rather than a tube. Finally, we obtained all five parameters that define the transformation of the unit-cell to its neighboring images. These parameters also define the outer radius of the tube, according to the following equation [18]:

$$r_{(\text{tube})} = \frac{\left[(n_1 a + n_2 b \cos \gamma)^2 + (n_2 b \sin \gamma)^2 \right]^{\frac{1}{2}}}{2\pi} \quad (1)$$

2.4. All-Atom MD Simulations

After obtaining the three lattice constants and the two wrapping parameters, we used our computational validation protocol [24] to assess the construct. This protocol uses explicit solvent together with cylindrical periodic boundary conditions to simulate a solvated unit-cell into a tube. Whilst the full atomic coordinates of the solvated unit-cell are preserved, the coordinates of the identical neighboring cells are calculated based on preprocessed transformation matrices. This enables measuring the non-bonded interactions of the unit-cell along the simulation. We carry out all-atom MD simulations into a nanotube, using the Image Facility of CHARMM [33].

We ran local minimization steps as described above to adjust the solvated cell into a wedge of a cylinder. Langevin Dynamics was used to restrain the water molecules at the exterior boundaries from escaping to the outside and inside vacuum with a low friction coefficient of 1.0 at 300°K, applying image centering. SHAKE was added to restrain the carbon-hydrogen bonds, with a 1fs timestep. The Verlet leap-frog integrator was used, and an isotropic periodic sum method was employed for calculating the long-range interactions. The non-bonded lists and the image lists were updated as necessary. These were calculated using the input transformation matrices. The minimum analysis period was set to 20ns and was varied according to the system size; small systems were run for a longer period of time. The CHARMM utility [33] was also used for trajectory analyzes. The system images were viewed by VMD [38].

3. RESULTS

3.1. P6 Plane Group Symmetry of Dimers and Hexamers

In vitro, the HIV-1 CA protein was observed to self-assemble into conical shapes as well as single-walled cylinders [39]. Recently, a 9Å resolution atomic model of the full-length wild-type monomer was reconstructed from cryoEM (cryo-electron microscopy) images [21]. Each monomer has two domains connected covalently by a short linker. The N-terminal domain (NTD) is the largest domain with seven α -helices, the linker is a loop of six amino acids and the C-terminal domain (CTD) contains four α -helices. The monomers are arranged in an approximate P6 lattice symmetry [40], such that the NTDs self-associate to form a hexameric ring, whereas the CTDs self-associate to form a dimer. The dimerization of two CTDs connects each ring to its six neighboring hexameric rings. Thus, this natural example fully complies with the scheme illustrated in Fig. (2d).

The actual lattice constants of the HIV-1 CA protein hexamer were determined by Ganser-Pornillos and coworkers [21] to be $\{a=92.7\text{\AA}, b=92.7\text{\AA}, \gamma=120^\circ\}$. Experimental observations also revealed a wide range of nanotube radii (20nm-25nm). Altogether, using the tube radius equation [18], we set the wrapping parameters to $\{n_f=14, n_z=0\}$, which corresponds to an exterior radius of ~20.6nm. The construction of the HIV-1 CA protein nanotube was discussed in [24]. Fig. (4) presents the construction of the unit-cell according to our protocol and approximately a fourth round of the resulting nanotube.

3.2. C4 Plane Group Symmetry of Anti-Parallel Dimers

We constructed a single-walled protein nanotube with a C4 symmetry lattice cell, comprised of dimers only. Fig. (2a) illustrates the corresponding 2D scheme. The building block is cyclic, and is made of four antimicrobial Magainin 2 dimers (PDB code 1dum) [41]. Each monomer is an α -helix, and all helices are anti-parallel. Each monomer has 23 amino acids. The structure is shown in Fig. (5). In the Magainin 2 dimeric crystal structure, the monomers bind to each other through hydrophobic residues, whereas the charged and polar residues point toward the exterior environment (Fig. (5a)). Most of the charged residues are positive. Therefore, a C4 symmetry construct of four monomers ensures high level of electrostatic repulsion in the central cavity of the cell (Fig. (5b)). Such repulsion may aid in the geometrical stabilization of the cell (and hence nanotube) over time. Three different types of Magainin 2 based systems were constructed. These types differ by the linker secondary structure. In type 1 we modeled a poly-glycine loop to reflect a very flexible linker. Type 2 linker is comprised of poly-prolines and type 3 is comprised of a short helix from the PDB - both types rigidify the system. Fig. (5c) presents the three types of linkers.

3.2.1. Glycine Loops—Based on the Euclidean distance between the termini of two adjacent monomers, we constructed a cyclic peptide of four monomers connected by four poly-glycine linkers. Five residues were the minimum required to close the gaps. After solvation, ionization and minimization, we applied Langevin Dynamics keeping the pressure fixed for 0.5ns. The lattice constants obtained were $\{41\text{\AA}, 41\text{\AA}, 90^\circ\}$, and the wrapping parameters - $\{16,0\}$. Unfortunately the subsequent simulation stopped after 12ns due to steric hindrance within the cyclic unit-cell. Repeating the procedure with very flexible poly-glycine linkers of 12 residues obtained the lattice constants $\{42\text{\AA}, 42\text{\AA}, 90^\circ\}$ and wrapping parameters $\{16,0\}$. However, the simulation halted after 5ns, again due to steric hindrance. At this stage we decided to introduce a linker of higher rigidity, a poly-proline loop, to allow a relaxed conformation of the cyclic peptide, and at the same time, constrain the motion within the unit-cell.

3.2.2. Proline Loops—A left-handed poly-proline II helix (PPII) is formed when sequential residues adopt backbone dihedral angles of roughly ($\phi=-75^\circ$, $\psi=150^\circ$) and have *trans* isomers of their peptide bonds [42,43]. The PPII helix contains exactly three residues per turn, and the rise per turn is approximately 3.1 $\text{\AA}$. After measuring the Euclidean distance between two adjacent monomer termini, we placed four poly-proline linkers of five residues each to fill the gaps between the termini and create a cyclic peptide.

Following the procedure as above, lattice constants of $\{38.2\text{\AA}, 38.2\text{\AA}, 90^\circ\}$ were obtained. Using local minimization with cylindrical boundary conditions, we obtained the wrapping parameters $\{16,0\}$. Finally, Langevin Dynamics was run for 30ns. Fig. (6a) presents snapshots along the simulation. As can be seen during the simulation, adjacent or opposite fragments of the cyclic peptide tend to create intra-molecular interactions. Thus the initial 2D structure of the unit-cell which possesses a central pore is distorted. Nevertheless, the inter-molecular interactions with the neighboring images prevail, and therefore, the nanotube preserves its physicochemical stability. Approximately half round of the resulting nanotube is seen in Fig. (6b). The coordinates were produced by the Image Facility of CHARMM

using the lattice constants and wrapping parameters of this specific tube. Fig. (6c) highlights the unit-cell position of the charged residues with regard to its neighboring cells. Almost all of these hydrophilic side-chains still point towards the central cavity of the cell, thus maintaining the hydrophobic interactions between the cells. To further rigidify the initial structure of the unit-cell and preserve the central cavity of the cyclic peptide, we introduced helical linkers [15].

3.2.3. Helical Linkers—Insertion of short linker helices to create a cyclic peptide is challenging since the covalent links at both ends are constrained by the limited range of backbone dihedral angles of α -helix. We cut a short α -helix of 13 residues from the structure of Follistatin (PDB code: 2p6a:c). The helix was chosen at random, and its length was determined by the Euclidean distance between the adjacent monomers in the unit-cell. We used Modeller [35] to bind each helix on both ends, while keeping the Magainin monomers still and allowing only the linkers to move. According to Table 1 the cyclic peptide illustrated in Fig. (2a) is organized in anti-parallel fashion. As a result, the covalent linkage of four helices to create a cyclic peptide produces an asymmetric square (Fig. (7)).

We followed our protocol by obtaining the lattice constants as $\{43.55\text{\AA}, 43.55\text{\AA}, 90^\circ\}$, followed by the wrapping parameters $\{16,0\}$. Finally, we applied Langevin Dynamics using cylindrical boundary conditions for 30ns. Fig. (7) presents the simulation snapshots every 10ns. The asymmetric 3D structure of the cyclic peptide maintains its structure along the simulation. As a result, the central pore does not collapse or gets distorted. Moreover, the fused monomers, which are initially classified as helices, also maintain their initial conformation. However, as in the poly-proline case, there is still a slight tendency along the simulation to create intra-molecular interactions while preserving the inter-molecular interactions with the neighboring images.

We compared the root mean-square deviations (RMSD) of poly-proline linkers and helical linkers systems. (Fig. (8a)) depicts the C RMSD of the unit-cells along the simulation. Since the reference structure is the minimized system, the high RMSD value at time 0 reflects a structural change the unit-cell has gone through during the first few picoseconds of equilibration. Neglecting the first few nanoseconds, it is clear that the helical linkers system preserves its original structure and orientation the most. This is also evident from the non-bonded energy of the unit-cell with its neighboring cells (Fig. (8b)). Here, the poly-proline linkers system fluctuates more than the helical linkers system and therefore exhibits a lower physicochemical stability over time.

4. DISCUSSION

We devised a shape-guided computational strategy to construct protein nanotubes without limitations on their size or diameter. The strategy is based on creating monomers by fusing two identical or distinct domains from native homo-oligomers and retaining their native association. The resulting fused monomers therefore possess two domains which self-assemble with identical copies of themselves. Via self-assembly a vast range of nanostructures can be obtained. This mechanism can be tailored to create nanotubes with variable shapes, sizes and diameters for various biological and non-biological applications.

Our strategy utilizes the natural repertoire of peptides and protein building blocks whose structures are available. This repertoire is extensive, and the combinations it can produce are vast. The challenge is to correctly engineer these in predefined geometries and chemical bonding while retaining their preferred associations. The innovative scheme of Padilla *et al.* [15] coupled with the powerful strategy of Tsai *et al.* [18] and our computational protocol [24] suggest a useful strategy to explore.

In this work we examined the already observed HIV-1 CA protein which self-assembles into conical shapes and single-walled nanotubes *in vitro*. Noteworthy, other retroviruses capsid proteins have also been observed to form nanotubes, *in vitro*. These include the Rous Sarcoma Virus (RSV) [44], the Murine Leukemia Virus (MLV) [45] and the Equine Infectious Anemia Virus (EIAV) [46]. Together with HIV-1 CA protein, these proteins share a common structure of a monomer comprised of a large NTD and small CTD which are covalently linked by a short flexible loop. Moreover, all these monomers were observed to self-assemble in the same manner, where the NTDs form a homo-hexamer and the CTDs form a homo-dimer. Thus, we hypothesize that this family of retroviruses can be attributed to our 2D assembly scheme, presented in Fig. (2d). To assemble them correctly we need the 3D structures of both interfaces: hexamer and dimer. Currently, both structures exist for HIV-1 CA protein only. Once both interfaces of the other retroviruses are available, their nanotubes could similarly be constructed as described here.

We addressed the challenge of *ab initio* construction of novel monomers which self-assemble into novel nanostructures focusing on the choice of the linker between the two ODs. The linker choice is critical, since it needs to fixate the fused domains in their relative positions. Therefore, as an example of *ab initio* construction, we chose the Magainin 2 homodimer [31] and followed a C4 symmetry scheme of fusing a cyclic peptide, while introducing three different types of linkers, one at a time. These types differ in their rigidity, and thus, in their ability to maintain their 3D structure throughout the simulation. Even though our simulations were successful with either poly-proline linkers or helical linkers, the helical linker maintained the initial 3D structure of the system best. Structural and thermodynamic studies validate this conclusion.

Nonetheless, we observed that in the natural case of the HIV-1 CA protein nanotube the linker is a short random loop. Looking at the HIV-1 CA nanotube Fig. (4) it is clear that the self-assembly of a single monomer creates a very compact nanostructure, with no cavities through which a small molecule could penetrate. Thus, the high flexibility of the linker has almost no dynamical effect on the overall conformation of the self-assembled nanotube. On the other hand, the Magainin 2 based cyclic peptide has a central pore which may allow such penetration into and out of the tube. To maintain the initial geometry of the pore during dynamics and to prevent a collapse, it is essential to rigidify the linkers which are the weak fragments of the cyclic peptide. Thus, it appears that when the 3D protomeric structure in the unit-cell is compact, a flexible dynamic linker may also be suitable, whilst a porous unit-cell requires a more rigid linker.

During the construction process, we faced the challenge of determining the optimal radius of the self-assembled tube. We realized that as the radius increments, the total potential energy

of the system decreases exponentially. However, this behavior is deceiving since the cylindrical periodic boundary conditions method does not allow considering the non-bonded interaction energy between the two marginal cells. Hence, to overcome this problem, we chose radii values for which the energy nearly reached a plateau. This concept was reinforced by the analysis of the radii of the HIV-1 CA protein nanotube as well as the surfactant-like peptide nanotube AcV₆D which were both observed experimentally.

Here we constructed a fused protein monomer from Magainin 2 homo-dimer, and simulated its self-assembly into a single-walled nanotube. In principle, if there are identical protomeric monomers and the concentration of the solute is high, structural propagation can be indefinite. The surfactant-like peptide AcV₆D self assembles into a tube [19]; another example of tube formation is the microtubulin [47]. There, the interdigitation of the microtubulin building blocks prevents closure into a vesicle. A similar situation exists in other proteins, like the bacterial flagellar [48]. However, the non-functional diphenylalanine (FF) was observed in the test tube to form both a tube and a vesicle [49]. Thus, some monomers favor an association into either tubes or vesicles, whereas others can form both. For the latter case, it can be expected that the preferred organization is concentration-dependent. In the case of FF, experiments revealed that at high peptide concentration, a tube was formed; at low concentration, a vesicle. Moreover, a tube has a one-dimensional curvature, whereas a vesicle has two. At a low concentration, the limited extent of non-polar buried surface area in a short tube drives the system to closure since a vesicle has a larger buried surface area. On the other hand, when the concentration is high, there is no such chemical pressure on the system. A long tube buries extensive non-polar surface. Thus, when the concentration is high, there is a high chance of vesicles fusing together to form a long tube, or an elongated vesicle, as shown by coarse-grained Monte Carlo simulations [50]. Therefore, based on our simulations that exhibited geometrical and thermodynamic stability, we hypothesize that at a high enough concentration, our fused cyclic peptides can form long nanotubes.

Our *ab initio* construction was applied to systems with approximate local C₄ lattice symmetry, but it is not limited to this symmetry and can be extended to assess other types of symmetrical systems where all the fused monomers are identical, and where these are assumed to assemble into cylinders. A future development of generic simulation protocols for handling other geometric structures might further facilitate the assessment of a broader range of novel protomeric structures which are assembled from such fused homo-oligomers.

Constructing novel protein nanostructures with desired properties is challenging. There is a great necessity for experimental and computational strategies toward this aim. A computational strategy is a valuable starting point, since it saves enormous time and resources by ruling out unlikely solutions. However, experimental validation is the next required step for any computational solution.

ACKNOWLEDGEMENTS

We thank Dr. Bernard R Brooks for fruitful discussions and support. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily 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. This study utilized the high-performance computational capabilities of the LoBoS supercomputers as well as the Biowulf PC/Linux cluster at the National Institutes of Health, Bethesda, MD (<http://biowulf.nih.gov>).

LIST OF ABBREVIATIONS

MD	Molecular dynamics simulations
HIV-1	Human immunodeficiency virus type 1
OD	Oligomerization domain

REFERENCES

- [1]. Elgersma RC; Meijneke T; Posthuma G; Rijkers DT; Liskamp RM Self-assembly of amylin(20–29) amide-bond derivatives into helical ribbons and peptide nanotubes rather than fibrils. *Chemistry*, 2006, 12, 3714–3725. [PubMed: 16528792]
- [2]. Yeates TO; Tsai Y; Tanaka S; Sawaya MR; Kerfeld CA Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells. *Biochem. Soc. Trans*, 2007, 35, 508–511. [PubMed: 17511640]
- [3]. Zheng J; Jang H; Ma B; Nussinov R Annular structures as intermediates in fibril formation of Alzheimer A beta(17–42). *J. Phys. Chem. B*, 2008, 112, 6856–6865. [PubMed: 18457440]
- [4]. Zhao Z; Singh R; Barman A; Johnson NF; Prabhakar R Modeling the Self-Assembly Dynamics of Macromolecular Protein Aggregates Underlying Neurodegenerative Disorders. *J. Comput. Theor. Nanosci*, 2009, 6, 1338–1351.
- [5]. Haspel N; Zanuy D; Aleman C; Wolfson H; Nussinov R De novo tubular nanostructure design based on self-assembly of beta-helical protein motifs. *Structure*, 2006, 14, 1137–1148. [PubMed: 16843895]
- [6]. Ruiz R; Kang H; Detcheverry FA; Dobisz E; Kercher DS; Albrecht TR; de Pablo JJ; Nealey PF Density multiplication and improved lithography by directed block copolymer assembly. *Science*, 2008, 321, 936–939. [PubMed: 18703735]
- [7]. Gazit E Molecular self-assembly: bioactive nanostructures branch out. *Nat. Nanotechnol*, 2008, 3, 8–9. [PubMed: 18654439]
- [8]. Haspel N; Zanuy D; Zheng J; Aleman C; Wolfson H; Nussinov R Changing the charge distribution of beta-helicalbased nanostructures can provide the conditions for charge transfer. *Biophys. J*, 2007, 93, 245–253. [PubMed: 17416628]
- [9]. Ballano G; Zanuy D; Jimenez AI; Cativiela C; Nussinov R; Aleman C Structural analysis of a beta-helical protein motif stabilized by targeted replacements with conformationally constrained amino acids. *J. Phys. Chem. B*, 2008, 112, 13101–13115. [PubMed: 18811190]
- [10]. Cui H; Muraoka T; Cheetham AG; Stupp SI Self-assembly of giant peptide nanobelts. *Nano Lett*, 2009, 9, 945–951. [PubMed: 19193022]
- [11]. Lim YB; Lee E; Lee M Cell-penetrating-peptide-coated nanoribbons for intracellular nanocarriers. *Angew Chem. Int. Ed. Engl*, 2007, 46, 3475–3478. [PubMed: 17385811]
- [12]. Garcia-Fandino R; Granja JR; D'Abramo M; Orozco M Theoretical characterization of the dynamical behavior and transport properties of alpha,gamma-peptide nanotubes in solution. *J. Am. Chem. Soc*, 2009, 131, 15678–15686. [PubMed: 19860480]
- [13]. Amblard M; Fehrentz JA; Martinez J; Subra G Methods and protocols of modern solid phase Peptide synthesis. *Mol. Biotechnol*, 2006, 33, 239–254. [PubMed: 16946453]
- [14]. Kent SB Total chemical synthesis of proteins. *Chem. Soc. Rev*, 2009, 38, 338–351. [PubMed: 19169452]
- [15]. Padilla JE; Colovos C; Yeates TO Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U S A*, 2001, 98, 2217–2221. [PubMed: 11226219]

- [16]. Raman S; Machaidze G; Lustig A; Aebi U; Burkhard P Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. *Nanomedicine*, 2006, 2, 95–102. [PubMed: 17292121]
- [17]. Usui K; Maki T; Ito F; Suenaga A; Kidoaki S; Itoh M; Taiji M; Matsuda T; Hayashizaki Y; Suzuki H Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci*, 2009, 18(5), 960–9. Erratum in: *Protein Sci.*, 2009, 18(7), 1571. [PubMed: 19384998]
- [18]. Tsai CJ; Zheng J; Nussinov R Designing a nanotube using naturally occurring protein building blocks. *PLoS Comput. Biol*, 2006, 2, e42. [PubMed: 16683021]
- [19]. Vauthey S; Santoso S; Gong H; Watson N; Zhang S Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. *Proc. Natl. Acad. Sci. USA*, 2002, 99, 5355–5360. [PubMed: 11929973]
- [20]. Reches M; Gazit E Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 2003, 300, 625–627. [PubMed: 12714741]
- [21]. Ganser-Pornillos BK; Cheng A; Yeager M Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell*, 2007, 131, 70–79. [PubMed: 17923088]
- [22]. Aksimentiev A; Brunner R; Cohen J; Comer J; Cruz-Chu E; Hardy D; Rajan A; Shih A; Sigalov G; Yin Y; Schulten K Computer modeling in biotechnology: a partner in development. *Methods Mol. Biol*, 2008, 474, 181–234. [PubMed: 19031067]
- [23]. Klein ML; Shinoda W Large-scale molecular dynamics simulations of self-assembling systems. *Science*, 2008, 321, 798–800. [PubMed: 18687954]
- [24]. Buch I; Brooks BR; Wolfson HJ; Nussinov R Computational Validation of Protein Nanotubes. *Nano Lett*, 2009, 9, 1096–1102. [PubMed: 19199488]
- [25]. Van Workum K; Douglas JF Symmetry, equivalence, and molecular self-assembly. *Phys. Rev. E. Stat. Nonlin Soft Matter Phys*, 2006, 73, 031502. [PubMed: 16605527]
- [26]. Andre I; Bradley P; Wang C; Baker D Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. USA*, 2007, 104, 17656–17661. [PubMed: 17978193]
- [27]. Carlson JC; Jena SS; Flenniken M; Chou TF; Siegel RA; Wagner CR Chemically controlled self-assembly of protein nanorings. *J. Am. Chem. Soc*, 2006, 128, 7630–7638. [PubMed: 16756320]
- [28]. Hess H; Clemmens J; Brunner C; Doot R; Luna S; Ernst KH; Vogel V Molecular self-assembly of “nanowires” and “nanospools” using active transport. *Nano Lett*, 2005, 5, 629–633. [PubMed: 15826099]
- [29]. Gorbitz CH Microporous organic materials from hydrophobic dipeptides. *Chemistry*, 2007, 13, 1022–1031. [PubMed: 17200919]
- [30]. Cardone G; Purdy JG; Cheng N; Craven RC; Steven AC Visualization of a missing link in retrovirus capsid assembly. *Nature*, 2009, 457, 694–698. [PubMed: 19194444]
- [31]. Wakamatsu K; Takeda A; Tachi T; Matsuzaki K Dimer structure of magainin 2 bound to phospholipid vesicles. *Biopolymers*, 2002, 64, 314–327. [PubMed: 12124849]
- [32]. Berman HM; Westbrook J; Feng Z; Gilliland G; Bhat TN; Weissig H; Shindyalov IN; Bourne PE The Protein Data Bank. *Nucleic Acids Res*, 2000, 28, 235–242. [PubMed: 10592235]
- [33]. Brooks BR; Brooks CL 3rd; Mackerell AD Jr.; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S; Caflisch A; Caves L; Cui Q; Dinner AR; Feig M; Fischer S; Gao J; Hodoscek M; Im W; Kuczera K; Lazaridis T; Ma J; Ovchinnikov V; Paci E; Pastor RW; Post CB; Pu JZ; Schaefer M; Tidor B; Venable RM; Woodcock HL; Wu X; Yang W; York DM; Karplus M CHARMM: the biomolecular simulation program. *J. Comput. Chem*, 2009, 30, 1545–1614. [PubMed: 19444816]
- [34]. Fiser A; Sali A ModLoop: automated modeling of loops in protein structures. *Bioinformatics*, 2003, 19, 2500–2501. [PubMed: 14668246]
- [35]. Eswar N; Webb B; Marti-Renom MA; Madhusudhan MS; Eramian D; Shen MY; Pieper U; Sali A Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci*, 2007, Chapter 2, Unit 2.9.
- [36]. Phillips JC; Braun R; Wang W; Gumbart J; Tajkhorshid E; Villa E; Chipot C; Skeel RD; Kale L; Schulten K Scalable molecular dynamics with NAMD. *J. Comput. Chem*, 2005, 26, 1781–1802. [PubMed: 16222654]

- [37]. Feller SE; Zhang Y; Pastor RW; Brooks BR Molecular Dynamics Simulation at Constant Pressure: The Langevin Piston Method. *J. Chem. Phys.*, 1995, 103, 4613.
- [38]. Humphrey W; Dalke A; Schulten K VMD: visual molecular dynamics. *J. Mol. Graph.*, 1996, 14, 33–38, 27–38. [PubMed: 8744570]
- [39]. Ganser-Pornillos BK; Yeager M; Sundquist WI The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.*, 2008, 18, 203–217. [PubMed: 18406133]
- [40]. Li S; Hill CP; Sundquist WI; Finch JT Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature*, 2000, 407, 409–413. [PubMed: 11014200]
- [41]. Hara T; Kodama H; Kondo M; Wakamatsu K; Takeda A; Tachi T; Matsuzaki K Effects of peptide dimerization on pore formation: Antiparallel disulfide-dimerized magainin 2 analogue. *Biopolymers*, 2001, 58, 437–446. [PubMed: 11180056]
- [42]. Adzhubei AA; Sternberg MJ Left-handed polyproline II helices commonly occur in globular proteins. *J. Mol. Biol.*, 1993, 229, 472–493. [PubMed: 8429558]
- [43]. Cubellis MV; Caillez F; Blundell TL; Lovell SC Properties of polyproline II, a secondary structure element implicated in protein-protein interactions. *Proteins*, 2005, 58, 880–892. [PubMed: 15657931]
- [44]. Campos-Olivas R; Newman JL; Summers MF Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *J. Mol. Biol.*, 2000, 296, 633–649. [PubMed: 10669613]
- [45]. Yeager M; Wilson-Kubalek EM; Weiner SG; Brown PO; Rein A Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms. *Proc. Natl. Acad. Sci. USA*, 1998, 95, 7299–7304. [PubMed: 9636143]
- [46]. Jin Z; Jin L; Peterson DL; Lawson CL Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. *J. Mol. Biol.*, 1999, 286, 83–93. [PubMed: 9931251]
- [47]. Gardner MK; Hunt AJ; Goodson HV; Odde DJ Microtubule assembly dynamics: new insights at the nanoscale. *Curr. Opin. Cell Biol.*, 2008, 20, 64–70. [PubMed: 18243676]
- [48]. Minamino T; Imada K; Namba K Molecular motors of the bacterial flagella. *Curr. Opin. Struct. Biol.*, 2008, 18, 693–701. [PubMed: 18848888]
- [49]. Yan X; Cui Y; He Q; Wang K; Li J; Mu W; Wang B; Ou-Yang ZC Reversible transitions between peptide nanotubes and vesicle-like structures including theoretical modeling studies. *Chemistry*, 2008, 14, 5974–5980. [PubMed: 18478616]
- [50]. Ji S; Ding J Spontaneous formation of vesicles from mixed amphiphiles with dispersed molecular weight: Monte Carlo simulation. *Langmuir*, 2006, 22, 553–559. [PubMed: 16401102]

Figure 1.

An illustration of the procedure of fusing protein domains into a monomer that self-assembles into symmetric nanostructures. In this example, a single monomer taken from a dimeric protein (top left, in light gray) is covalently linked to a single monomer taken from a trimeric protein (top right, in dark gray) via a short linker (black). The fused monomer is now comprised of N-terminal and C-terminal domains. The monomer self-assembles into a unit-cell of P6 lattice symmetry, which finally self-assembles with identical cells to create a nanotube.

Figure 2.

Typical plane group symmetries of a 2D unit-cell constructed from the fusion of two ODs. A light gray circle represents a dimer; a dark gray circle represents a trimer, tetramer or hexamer; a thick connector line represents a linker between two monomers. Light-gray numbers indicate the monomers that construct the cell, and dark-gray numbers show the monomers of the neighboring cells. **(a)** C4 symmetry built of dimers. **(b)** C4 symmetry built of tetramers and dimers. **(c)** P6 symmetry made of trimers and dimers. **(d)** P6 symmetry made of hexamers and dimers.

Figure 3.

Obtaining the tube wrapping parameter n_1 using two examples of experimentally observed nanotubes; the HIV-1 CA protein nanotubes, and the surfactant-like AcV₆D nanotube. The potential energy of the unit-cell was measured for different values of the wrapping parameter n_1 which controls the tube curvature. The parameter n_2 was set to zero. In the HIV-1 CA protein case, the actual range of tube radii was obtained from the reconstruction of cryoEM images [21]. In the AcV₆D case, the range of tube radii was obtained from quick-freeze / deep-etch TEM image [19].

Figure 4.

The construction of the HIV-1 CA Protein single-walled nanotube, according to the design scheme presented in figure (2d). (a) The monomer (PDB code:3dik) is comprised of a large N-terminal domain (NTD, in red) and a smaller C-terminal domain (CTD, in green) which are connected covalently by a short linker (blue) of six amino acids. (b) The unit-cell consists of six monomers. The NTDs (red) are on the tube outer surface, and the CTDs (green) protrude toward the tube interior. The lattice constants $\{a, b, \gamma\}$ are depicted. (c) A perspective view of three rows of the CA protein tube as seen from the tube exterior. The

NTDs are in red whereas the CTDs are in green. The wrapping parameters are $\{n_1=14, n_2=0\}$ and the lattice constants are $\{a=92.7\text{\AA}, b=92.7\text{\AA}, \gamma=120^\circ\}$. The outer tube radius is therefore $206\text{\AA}$. A unit-cell is highlighted in a dashed line. **(d)** Approximately a fourth round of the tube, as seen from the tube axis perspective.

Figure 5.

The construction of a fused cyclic peptide from the antimicrobial homodimer Magainin 2 (PDB code: 1dum) according to the scheme shown in figure (2a). The cyclic peptide has C4 plane group symmetry. **(a)** The homo-dimer crystal structure. Each monomer is in different color (red and black). The side-chains are colored by their polarity; basic – blue, acidic – red, polar – green, hydrophobic – light gray. The inter-monomer interactions are hydrophobic, whereas the charged and polar side-chains point to the surface. **(b)** The four monomers are on a planar surface such that their center of mass lies on the origin. The black monomers are to be covalently linked through identical short linkers. The charged side-chains point towards the central cavity of the cell. **(c)** Three types of linkers were used to join adjacent monomers: poly-glycine (left), poly-proline (center) and α -helix (right). These linkers represent different levels of backbone flexibility. The backbone atoms are in green ribbons and the residual atoms are in blue.

Figure 6.

A cyclic peptide made of monomers from the antimicrobial Magainin 2 (PDB code: 1dum) joined by identical poly-proline linkers and simulated for 30ns. (a) The results of MD simulation into a nanotube. At time=0ns the unit-cell appears solvated in water (blue beads) with neutralizing counter-ions (yellow beads). The four monomers of the cyclic peptide are in red ribbon, whereas the poly-proline linkers are in gray. The left perspective is as viewed from the tube exterior, whereas the right perspective exhibits the unit-cell with the four-fold axis vertical. The central pore in the unit-cell changes its 3D shape along the first 20ns of

simulation and preserves it later on. **(b)** Approximately half of the nanotube after 30ns of simulation. Schematic illustrations exhibit the corresponding perspectives. Each cell is in different color. **(c)** A close-up on the charged residues of the unit-cell (in blue ribbons) at 30ns, viewing from the tube exterior perspective. The positively charged residues are in blue sticks, negatively charged in red.

Figure 7.

A cyclic peptide comprised of monomers from the antimicrobial Magainin 2 (PDB code: 1dum) joined by helical linkers. The four identical linkers (in gray) are short -helices of 13 amino acids extracted from the structure of Follistatin (PDB code: 2p6a:c). The unit-cell appears in two perpendicular perspectives; as seen from the tube exterior (left) and when the four-fold axis is vertical (right). At 0ns, the unit-cell is locally minimized, and presented in bulk water (blue beads) with neutralizing ions (yellow beads). The cell shape is asymmetric due to the anti-parallel constraint induced by the design of a cyclic peptide (Table 1).

Figure 8. Structural and energetic analyzes of the poly-proline linkers system and the helical linkers system along 30ns of simulation. (a) The C RMSD calculated with regard to the systems structures after minimization. (b) The non-bonded energy of the unit-cell with its neighboring images.

Table 1.

Properties of the 2D Unit-Cells Shown in Fig. (2)

Sub-Figure	Symmetry of Cell	Oligomers in Cell	Parallelism	Building Blocks
A	C4	4 dimers	anti-parallel	1 ^a
B	C4	4 tetramers 4 dimers	identical	4
C	P6	2 trimers 5 dimers	The trimers are parallel	6
D	P6	4 hexamers 5 dimers	identical ^b	6

^aThe building block is cyclic

^bThe parallelism constraint is imposed by the 1–4, 2–5, and 3–6 hexamer links