De Novo Tubular Nanostructure Design Based on Self-Assembly of β-Helical Protein Motifs

Nurit Haspel,[1] David Zanuy,[2] Carlos Alemán,[2]
Haim Wolfson,[1] and Ruth Nussinov[3,4,*]
[1] School of Computer Science
Faculty of Exact Sciences
Tel Aviv University
Tel Aviv 69978
Israel
[2] Department of Chemical Engineering
ETSEIB-UPC
Av. Diagonal 647
08028 Barcelona
Spain
[3] Basic Research Program
SAIC-Frederick, Inc.
Center for Cancer Research Nanobiology Program
National Cancer Institute
NCI-Frederick
Frederick, Maryland 21702
[4] Sackler Institute of Molecular Medicine
Department of Human Genetics
Sackler Faculty of Medicine
Tel Aviv University
Tel Aviv 69978
Israel

Summary

We present an approach for designing self-assembled nanostructures from naturally occurring building block segments obtained from native protein structures. We focus on structural motifs from left-handed β-helical proteins. We selected 17 motifs. Copies of each of the motifs are stacked one atop the other. The obtained structures were simulated for long periods by using Molecular Dynamics to test their ability to retain their organization over time. We observed that a structural model based on the self-assembly of a motif from *E. coli* galactoside acetyltransferase produced a very stable tube. We studied the interactions that help maintain the conformational stability of the systems, focusing on the role of specific amino acids at specific positions. Analysis of these systems and a mutational study of selected candidates revealed that the presence of proline and glycine residues in the loops of β-helical structures greatly enhances the structural stability of the systems.

Introduction

Nanotechnology aims to design novel materials and molecular devices, often via self-assembly. Among the applications one can count targeted drug delivery systems, computational devices, and scaffolding tissues (Ferrari, 2005a, 2005b). In nature, protein domains often self-assemble, spontaneously organizing in stable structures through noncovalent interactions. These may create large complexes of well-defined structure and function. Exploiting the natural ability of protein molecules to self-assemble can be a very useful approach in the design and construction of novel molecular structures (Ai et al., 2003; Bae et al., 2005; Haino et al., 2005; Pargaonkar et al., 2005; Porod et al., 2004). Recently, there have been numerous reports describing the use of natural and artificial peptides, DNA segments, and RNA segments in nanodesign (Badjic et al., 2004; Chworos et al., 2004; Claussen et al., 2003; Liao and Seeman, 2004; Percec et al., 2004; Rajagopal and Schneider, 2004; Rathore and Sogah, 2001; Valery C. et al., 2003; Yan et al., 2003; Zhang S., 2003). Due to advances in peptide synthesis and molecular engineering techniques (Bang and Kent, 2005; Kent, 2004; Koide et al., 2000, 2005; Li et al., 2005; Miller and Raines, 2005; Smith et al., 2005), self-assembly of peptide segments can become a favorable route by which to obtain nanostructures, particularly those consisting of single or associated tubes, fibers, and vesicles.

Computations are increasingly becoming a major tool in nanobiology and nanostructure design. The use of advanced simulation methods and efficient modeling algorithms, in addition to the abundance of data that can be found in protein databases, can considerably accelerate the design process via fast probing of many models. The aim is for experimental testing to be performed only on models that are a priori computationally predicted to be likely candidates.

Some naturally occurring proteins contain a tubular or fibrillar motif in their folds. A good example of proteins whose fold is naturally tubular is the β helix protein fold. 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 especially useful: the tubular structure is regular and symmetrical and is often stabilized by a network of interactions between similar residues in consecutive coils (see Figure 1A) (Jenkins et al., 1998).

The common types of interactions in β helices include:

1. Asparagine (or glutamine) ladders that stabilize the helical structure through hydrogen bonds between residues in consecutive rungs (see Figure 1B).
2. Stacking of aromatic (Phe, Tyr, His) and aliphatic (Pro) rings (see Figure 1C).
3. Hydrophobic interactions (especially Val, Ile, Leu) (see Figure 1D).

The tubular nature of left-handed β-helical 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. In addition, their helical and symmetric structure makes them good candidates to be excised and tested as modules.

Here, we present a general approach to the design of nanostructures based on the potential assembly property of protein segments, in which the segments are taken from naturally occurring proteins and have

*Correspondence: ruthn@ncifcrf.gov
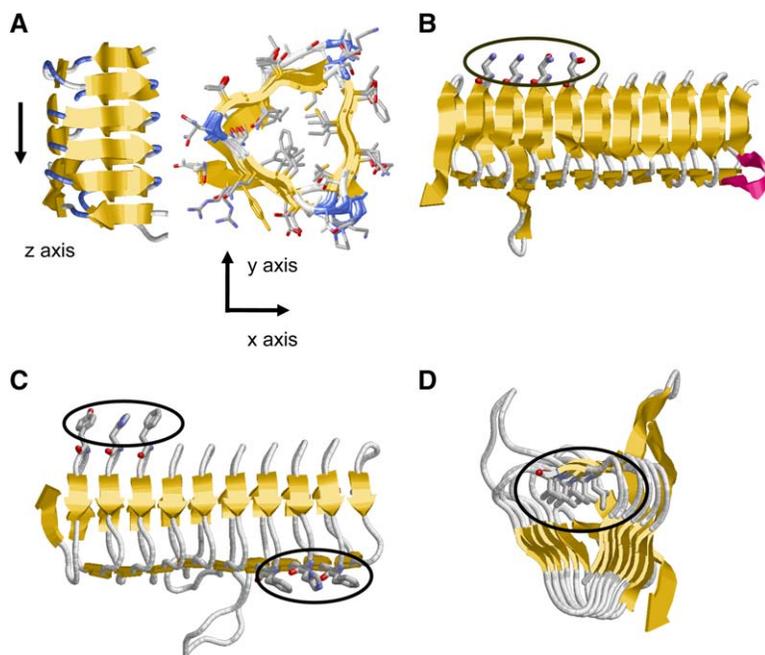
Figure 1. An Example of β-Helical Proteins and Their Typical Interactions

(A) An example of a left-handed β-helical protein. Left: front view; right: top view. The z axis is in the direction of the fiber growth. The protein is thermal hysteresis from Spruce budworm (*Choristoneura fumiferana*), shown in five-turn isoforms (PDB code: 1l0s). The atoms are represented as solid ribbons. In the front view, the individual atoms are omitted for clarity.

(B–D) Examples of typical left-handed β helix interactions. In each case, the interacting residues are emphasized in wire-frame and are circled in black. (B) An example of an asparagine ladder. The protein is N-acetyl-glucosamine 1-phosphate uridyltransferase GlmU, specifically the C-terminal domain from *Streptococcus pneumoniae* (PDB code: 1hm0). (C) An example of aromatic ring stacking. The protein is UDP N-acetylglucosamine acyltransferase from *E. coli*, gene *lpxA* (PDB code: 1lxa). (D) An example of hydrophobic residue interactions. The protein is tetrahydrodipicolinate-N-succinyltransferase, THDP-succinyltransferase, DapD from *Mycobacterium bovis* (PDB code: 1kgq).

preferred conformational tendencies. In this work, we design nanoconstructs based on left-handed β-helical proteins. To construct our nanosystems, we select short (two turns), repetitive motifs and extract the corresponding coordinates from the PDB (Bernstein et al., 1977). We assemble copies of the motifs on top of one another (see Figure 2 for an example of a system) and simulate them for long periods of time (≥ 20 ns) at a temperature of 300ºK by using NAMD (Kale et al., 1999), a high-performance molecular dynamics program, to test their structural stability over time (see Experimental Procedures for details). Of the 17 systems we tested, the construct based on the assembly of copies of residues 131–165 of galactoside acetyltransferase from *E. coli* (PDB code: 1krr, chain A) showed a remarkable structural stability over a long period of simulation time (20–40 ns) under all of the tested temperature and ionic strength conditions. The main criterion we used to assess the structural stability of our tested models is the retention of the bulk organization over time, and we largely focused on the organization of the loop regions since these are typically the least stable structural components.

We also studied the effect of specific amino acids and chemical interactions on the conformational stability of the overall structure, again, especially in loop regions. Through a mutational study, we found that apart from the characteristic interstrand interactions of β-helical proteins, the presence of proline residues around the loop areas greatly contributes to the retention of the loop structure and hence to the stability of the overall conformation. In addition, in many cases we found a relatively large number of glycines in loop areas. These glycines are involved in hydrogen bonds with the side chains of other residues in their vicinity and hence contribute to maintaining the conformation of the loops. In the future, we aim to further enhance the stability of the system by inserting specific point mutations of natural and synthetic amino acids whose structures

are available. The choice of the synthetic residues and the positions of insertion are guided by our mutational observations on the stabilizing effects of naturally occurring residues on the entire system.

Results and Discussion

System Description
The systems we constructed and simulated consisted of four repeats of a basic two-turn unit, containing altogether eight strand-loop motifs stacked on one another. Each basic repetitive unit was taken from the PDB file

```
PITIGNNVWIGSHVVINP
GVTIGDNSVIGAGSIVT-
```


Basic repetitive unit
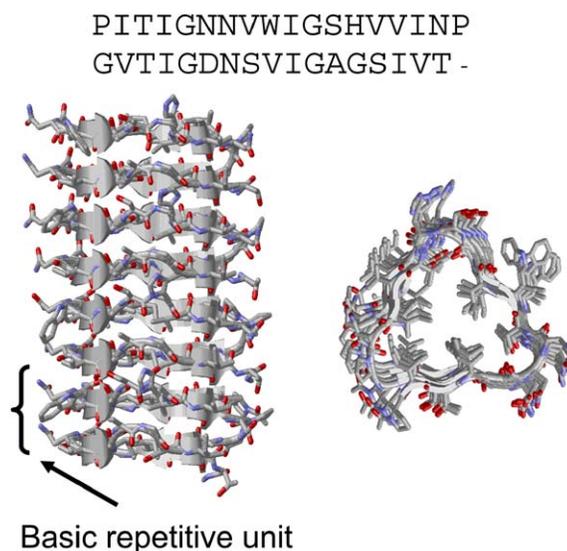
Figure 2. The krr1 System: Residues 131–165 of Galactoside Acetyltransferase

Left: front view; right: top view. The repetitive unit is marked in the front view of the system. The sequence alignment of the rungs of the repeat is shown at the top.

Table 1. The Sequences Used to Create the Simulated Nanotubes

Name	PDB	Protein Name	Residues	Sequence
hv91	1hv9	N-acetylglucosamine 1-phosphate uridyltransferase GlmU, C-terminal domain from *E. coli*	259–294	**RFDLRGTLTHGRD**VEIDT NVIIEGNVTLGHRVKIGT
hv92			296–329	CVIKNSVIGDDCEISPY TVVEDANLAAACTIGPF
hv92_mut1				CVIKNSNIGDDCEISPY TVVEDANLAAACTIGPF
hv92_mut2				CV**G**KNSVIGDDCEISPY TVVEDANLAA**P**CTIGPF
hv92_mut3				CVIKNSNIGDDCEISPY TVVEDANLAA**K**CTIGPF
hv92_mut4				CVIKNSVIG**P**DCEISPY TVVEDANLA**P**ACTIGPF
kap1[a]	1kap	metalloprotease from *Pseudomonas aeruginosa*	338–374	LLIGNDVANVLKGG**A**GND ILYGGLG**A**DQLWGG**A**GAD
kgq1	1kgq	tetrahydrodipicolinate-N-succinlytransferase, THDP-succinlytransferase, DapD from *Mycobacterium bovis*	104–138	RVVPP**A**TVRQG**A**FI**A**RNT VLMPSYVNIG**A**YVDEGT**T**
kgq2	1kgq		129–164	NIG**A**YVDEGT**M**VDTWATV GSCA**Q**IGKNVHLSGGVGI
krr1	1krr	galactoside acetyltransferase from *E. coli*	131–165	PIT**I**GNNVWIGSHVVINP GVT**I**GDNSVIGAGSIVT-
krr1_mut1				PIT**I**GNNVWIGSHVVINP GVT**I**GDN**V**VIGAGSIVT-
krr1_mut2				PIT**I**GNNVWIGSHVVINP GVT**I**GDNSVIGAG**V**IVT-
krr1_mut3				PIT**I**GNNVWIGSHVVIN**A** GVT**I**GDNSVIGAGSIVT-
l0s1	1l0s	thermal hysteresis protein from spruce budworm (*Choristoneura fumiferana*), five-turn isoforms	27–57	VDKSEVFGTT**C**TGSRFDGVT**I**TTSTSTGSRI
lxa1	1lxa	UDP N-acetylglucosamine acyltransferase from *E. coli*, gene *lpxA*	33–69	GPHV**E**IGEGTVLKSHVVV-NGHTK**I**GRDNE**I**YQFAS**I**G
lxa2	1lxa		109–144	TK**V**GSDNLLMINAHI**A**HD CTVGNRC**I**LANNATLAGH
lxa3	1lxa		142–177	AGHV**S**VDDFA**I**IGGMTAV HQFC**I**IGAHV**M**VGGCSGV
qre1	1qre	γ-carbonic anhydrase from the Archaeon *Methanosarcina thermophila*	119–153	SQVHGPA**A**VGDDTFIGMQ AFVFKSKVGNNCVLEPRR

The loop and turns regions are set in bold. Loops and turns were determined by using the DSSP program (Kabsch and Sander, 1983). In the mutated systems, the mutated residues are set in bold. For secondary structure composition, see the wild-type sequence.
[a] This protein is a right-handed β helix.

of a left-handed β-helical protein and was represented at the atomic level. The units were selected by their structural and sequential regularity: the loops in each turn should be regular and should contain no protruding parts, and the sequence should be as regular as possible, containing as many typical β-helical interactions (as described in the Introduction) as possible.

Since a one-turn unit may not be stable enough and may tend to break up around the edges, we used a repetitive unit that consisted of two complete helical turns for all of our systems. In addition, in a one-stranded repetitive system, all of the interstrand interactions would be between identical amino acids, since it means stacking copies of the same sequence one atop of another. This type of organization is not found in naturally occurring peptides in which the amino acids that interact between two consecutive turns of the helix are different.

Larger repetitive units (with 3–4 complete turns) may result in a system that is too large for long simulations, since we wish to examine systems containing at least four repetitive units in order to study the self-elongation properties of our systems. A system consisting of four repeats of a two-turn basic unit appears to be a good compromise between these two conflicting demands, and it results in a reasonably sized system (∼17,000 atoms including the solvent). Furthermore, in a four-unit system, the two middle units are completely surrounded by adjacent units, allowing us to study the interstrand interactions, and the two other units are exposed to the solvent on one side, allowing both the solute-solute and solute-solvent interactions to be studied.

We tested 17 different systems taken from 7 different β-helical proteins. Because we had a hard time finding structurally regular right-handed systems, six of the proteins were left-handed β-helical proteins and only one protein was a right-handed β-helical protein. The only family of right-handed β helices we encountered that suited our purposes was the β-roll family. All of the β roll proteins belong to one family and therefore are very similar structurally and sequentially. We simulated three variants of metalloprotease from *Pseudomonas aeruginosa* (PDB code: 1kap), composed of residues 338–374 ± 1 residue. The three systems fell apart very quickly, and thus the simulation was stopped. Table 1 lists the tested systems and their sequences. All of the systems were simulated in a water box for a period of 20 ns or longer, at 300°K. The simulations were performed by using the NAMD program (see Experimental Procedures for a detailed description of the simulation conditions). We chose to simulate our systems for a period of at least 20 ns, because this period provides a compromise between the need for a long enough simulation time that ensures the equilibration of the system and the accumulation of enough data, but does not resort to excessive computational efforts needed to carry out longer-term simulations. Further, since our goal here is to validate the stability of some of our systems and to select the best candidates for experimental tests, we need to differentiate between the best models and other models by means of their ability to maintain their structure under different conditions. Thus, we performed additional simulations at the temperature of 360°K. If our models collapse when the temperature is increased by 60°, their structural organization is too sensitive to temperature changes, since it is easily lost under thermal stress. In addition, in order to validate our results in an environment that is more reminiscent of physiological conditions, we simulated some of our models by adding
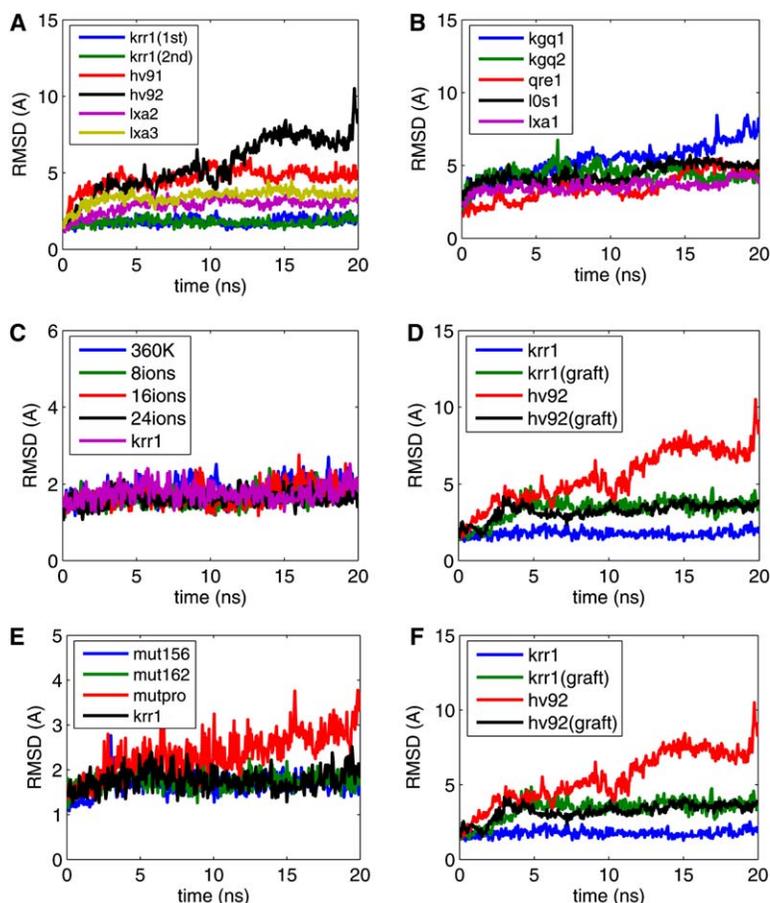
Figure 3. Structural Analysis of the Simulated Systems

(A and B) The evolution of the rmsd of the C-α atoms of the simulated models compared to the minimized structures for all of the left-handed wild-type sequences. krr1 was simulated twice. The plot was divided into two subplots for clarity. See Table 1 for the corresponding sequences.

(C) The evolution of the rmsd of the C-α atoms of the simulated models compared to the minimized structures for all of the krr1 systems that were simulated under different temperatures and ionic strength conditions. The different simulation conditions are specified in Experimental Procedures.

(D) The evolution of the rmsd of the C-α atoms of the simulated models compared to the minimized structures for all of the hv92 mutants. See Table 1 for the sequences of the wild-type and mutated sequences.

(E) The evolution of the rmsd of the C-α atoms of the simulated models compared to the minimized structures for all of the krr1 mutants. See Table 1 for the sequences of the wild-type and mutated sequences.

(F) The evolution of the rmsd of the C-α atoms of the wild-type versus grafted krr1 and hv92. The krr1 (graft) system contains the backbone of krr1 and the sequence of hv92, while hv92 (graft) contains the backbone of hv92 and the sequence of krr1. For clarity, the scales are different in each section.

ions to the solution and by using EWALD (Darden et al., 1993) particle mesh summation to evaluate the effect of ionic strength on the structural organization of our models. We ran three different simulations with a different number of ions, keeping the overall charge of the system neutral to enable the EWALD summation (details of the different ionic concentrations are provided in Experimental Procedures).

Figures 3A and 3B show the evolution of the rmsd with respect to the initial (minimized) structure of all of the wild-type, left-handed simulated systems over time. We considered only the C-α atoms in the calculation. As can be seen, most of our models changed their structure during the heating and equilibration stages, so that at the beginning of the simulation the rmsd between their minimized structure and the heated and equilibrated structure is ~1 Å. Except for this initial change, some of our simulated models changed only slightly during the simulation, and the rmsd of their C-α atoms hardly exceeded 2 Å. However, other systems lost their initial organization during the simulation and became amorphous. In this paper, we focus on the system that, as Figure 3 shows, was able to retain its structure better than all of the other systems. It tested to be very well organized structurally over time in the sense that it retained its bulk structure and characteristic interactions throughout the simulation. This system was built of residues 131–165 of E. coli galactoside acetyltransferase (PDB code: 1krr). It was denoted krr1 and appeared to be a good candidate for nanofiber construction. Our

results showed that krr1 remained very well organized under different temperature and ionic strength conditions. We simulated krr1 twice for 20 ns each run to verify the results. Both simulations are shown in Figure 3A. As seen, the two simulations are consistent with one another. Figure 2 shows the initial structure of the system and its sequence.

Different Temperature and Ionic Strength Simulation
In order to test the ability or krr1 to remain structurally organized under a condition change, we simulated it under the following conditions:

1. 360°K.
2. Addition of sodium and chloride ions as specified in Experimental Procedures (0.23%, 0.5%, and 0.8% w/w, which constituted 8, 16, and 24 ions, respectively).

Figure 3C shows the rmsd with respect to the initial structure of all of the above-mentioned simulations of krr1 (including the 300°K simulation). As can be seen, the tested changes of simulation conditions did not have a significant effect on the structural organization of krr1, which remained very stable regardless of the changed conditions.

Structural Analysis
As seen from the data described above, all of our models contained assembled copies of structurally regular

motifs taken from β-helical proteins, and they all had very similar initial structures. Yet, some models maintained their structures during the simulations, while others lost them and became disorganized. We assume therefore, that the key to understanding what makes certain self-assembling systems more stable than others with similar structures lies in their sequences, especially in the sequential motifs located in key structural positions.

When analyzing the β-helical structures, one of our assumptions was that the loop regions, being more flexible than β strands, contribute significantly to the retention of the structural organization of the entire system. This assumption is based on the observation that loop areas tend to be less conserved structurally and lose their organization faster. Therefore, we assume that stable loops are essential for overall stable structures. We also assume that specific residues and interactions at the loop areas have a special contribution to the overall stability of the β-helical structure of our systems. In order to quantitatively characterize the structural retention of the loop areas, we measured the evolution of the structure of the loops in the simulated models throughout the simulation. The main criteria that characterize the retention of the loop structure are the following:

1. The distribution of the ϕ and ψ dihedral angles of the loop residues throughout the simulation.
2. The rmsd of the main chain atoms of loop residues in a given structure with respect to the initial structure.

Figure 4 shows the distribution of the main chain dihedral angles of all of the loop residues of krr1 throughout the simulation. The loop was defined as the two-turn residues plus another residue in each direction (see Table 1 for the secondary structure assignment). As can be seen, the main chains of these residues remained in a very narrow range of ϕ and ψ angles, showing that the loop region was able to maintain its initial backbone conformation throughout the simulation. The rmsds of the loop residues of krr1 (see the Supplemental Data available with this article online) show a similar tendency. Figure 5A provides the structures of the krr1 system at the beginning (left) and the end (right-hand side) of the simulation, respectively. It can be seen that the overall structure of the krr1 loops changed only slightly during the simulation.

Mutant Study and the Effect of Specific Residues on Structural Organization

To elucidate the effects of specific residues on the structural retention of the loop areas in general, and to determine which residues contribute the most to the stability of the system, we performed sequence and structure analysis of krr1. The sequence alignment of the two turns in krr1, displayed in Figure 2, shows that the loop areas are rich in proline and glycine residues. This can be explained by the following observations: Proline is a rigid amino acid whose main chain conformational space is very limited; therefore, it restricts the loop main chain movements and contributes to its stability. Glycine tends to reside in loop areas, especially in the vicinity of β strands, breaking β structures and absorbing many of the loop movements due to its flexibility. In addition, it restricts the loop movements and contributes to the overall structural stability due to its involvement in stabilizing interactions with the side chains of other residues in its vicinity.

In order to assess the effect of proline and glycine residues on the retention of the overall structures, we simulated mutant derivatives of the most stable model, krr1, and of the least stable one, hv92 (see Figure 3A). We simulated three mutants of krr1 and four mutants of hv92. The mutants are shown in Table 1. In the hv92 mutants, we aimed at forming typical β-helical interactions such as asparagine ladders and salt bridges by inserting prolines or involving glycines in hydrogen bonds. In the krr1 mutants, we either replaced Pro148 with alanine or substituted Ser156 or Ser162 with valine, preventing the glycines in the loop of the mutated serine from forming hydrogen bonds. Figure 3D plots the rmsd of the hv92 derivatives with respect to their initial structures during the simulations (including the wild-type), and Figure 3E shows the same for the derivatives of krr1. As seen in Figure 3D, nearly all of the mutations inserted in hv92 caused the system to be more organized structurally. In addition to the expected increase of stability caused by the insertion of typical β-helical interactions, our mutant simulations showed that a dramatic increase in the structural organization was attained in the mutants in which prolines were inserted into the loops. Neither of the hv92 mutants was as structurally stable as krr1, but their structural organization was much better preserved than the organization of the wild-type hv92. In addition, as seen in Figure 3E with respect to the krr mutants, the mutant with the lowest structural organization (which has the largest rmsd with respect to the initial structure) was the mutant in which Pro148 was replaced by alanine. The other mutants, in which one serine residue was replaced by valine, exhibited a much smaller effect on the rmsd. This led us to conclude that, in addition to the typical β-helical interactions, proline bears a great importance in helping loops retain their conformations.

Below, we provide a detailed analysis of the hv92 mutant in which prolines were inserted, denoted hv92_mut4. In this mutant system, Asp305 and Ala322, both located on consecutive turns in the same loop, were mutated to proline. Figures 6A and 6B show the distribution of the main chain conformations of the loop residues for one of the loops in hv92. As seen, the dihedral angles distribute very widely, which shows that, in this system, the loop structure tends to melt during the simulation, and explore a wider range of conformations for a given residue. Figures 6C and 6D provide the distribution of the main chain dihedral angles of one of the loop residues of hv92_mut4 throughout the simulation. The loop shown here is equivalent to the loop shown in Figures 6A and 6B for the wild-type, and it is the loop we mutated. Comparing Figures 6A and 6B with Figures 6C and 6D, respectively, we see that the distribution of the φ and ψ angles is much narrower in the mutant system. Figure 6 also shows the rmsd of the initial structure, specifically of the loop residues of hv92 and hv92_mut4, over time. As seen in Figures 6A and 6B, the residues in this case fluctuate considerably, and they can move up to ~15 Å from the original position, which shows that this particular loop turned into an amorphous aggregate. Figures 6C and 6D show the evolution of the rmsd of the loop residues
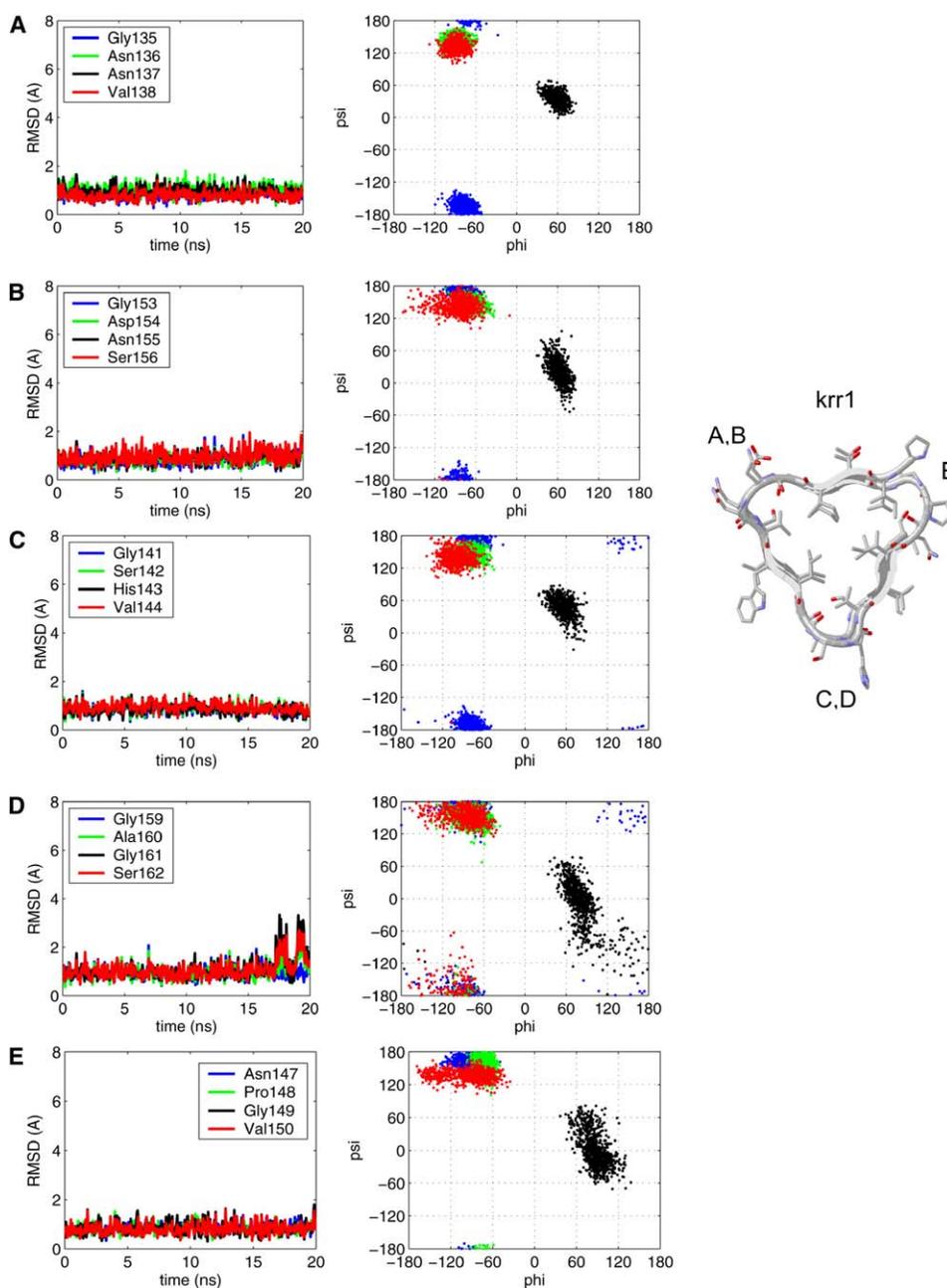
Figure 4. The Distribution of the Main Chain Dihedral Angles, φ and ψ, for the Loop Residues of krr1 throughout the Simulation

(A–E) The loops are designated (A)–(E) according to their location on the structure (see the right side for the designation of each loop). Subfigures (A)–(E) show the distribution of main chain angles (right) for loops (A)–(E), respectively. Each residue is displayed in a different color (see inset legends).

with respect to the initial structure for hv92_mut4. As can be clearly seen, the rmsd of the loop residues in the mutant is much smaller than it is in the wild-type. The main chain dihedral angle distribution and evolution of the rmsd of all of the loops of hv92 and hv92_mut4 can be found in the Supplemental Data. The final structure of hv92_mut4, shown in Figure 5C on the right, albeit less organized than the initial structure, was still able to retain its initial conformation in a much better way than in the wild-type structure, shown in Figure 5B. We can conclude that the presence of proline in loops has a significant contribution to the ability of the system to

maintain its bulk structure due to its backbone being conformationally restricted. However, the models we simulated that contained proline, such as the wild-type hv92, were not always structurally organized enough to maintain their bulk structures through 20 ns of simulation; thus, other conditions should also be fulfilled in order to ensure the structural stability of a system. Our studies point to several additional factors that contribute to the structural organization. These include:

1. Regularity of the backbone conformation: as discussed below, one of the hv92 loops is one amino
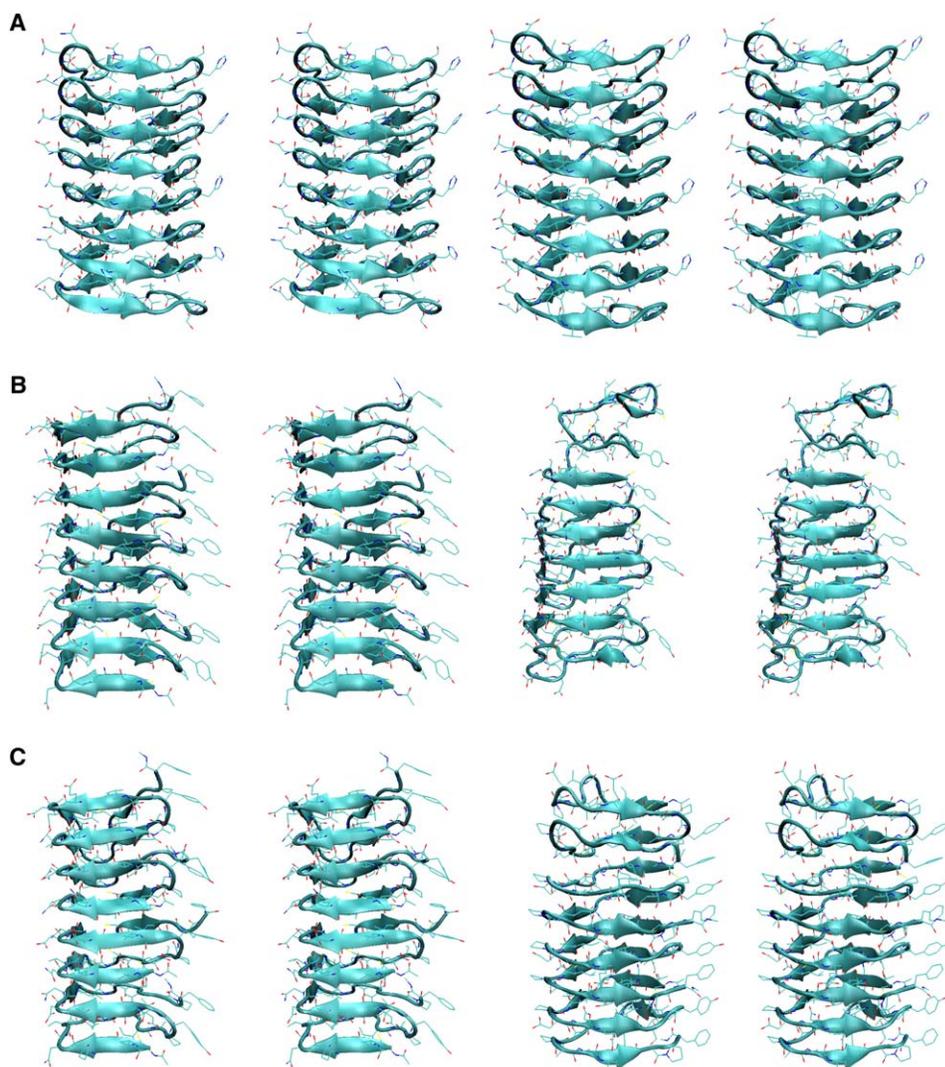
Figure 5. The Evolution of the Structure of Three Simulated Systems

(A) Left: the simulated system krr1 at the beginning of the simulation. Right: the simulated system krr1 after 20 ns of the simulation.

(B) Left: the simulated system hv92 at the beginning of the simulation. Right: the simulated system hv92 after 20 ns of the simulation.

(C) Left: the simulated system hv92_mut4 at the beginning of the simulation. Right: the simulated system hv92_mut4 after 20 ns of the simulation.

The figures are in stereovision.

acid shorter, making the backbone not completely regular as compared to krr1. Other systems, such as kgq1 and qre1, were also less regular in their backbone conformation, especially in their secondary structure assignment (see Table 1), and they were indeed less structurally stable than krr1 by the end of their simulations.

2. Charge repulsion: as seen in Table 1, hv92, kgq1, and qre1 contain proline at the loops, but all three contain similarly charged residues placed near one another somewhere along the sequence. For example, the hv92 system contains two adjacent aspartic acid residues in one of the loops. One of the additional prolines in hv92_mut4 was inserted in place of one of these aspartates, thus reducing the charge repulsion in addition to the structural stabilizing effect of the proline.

3. The presence or lack of characteristic β-helical interactions such as asparagine ladders and

stacking. Virtually all of our systems contain hydrophobic residue stacking in the interior of the structure; however, hv92 contains no asparagine ladder. krr1 contains an asparagine ladder and a threonine ladder.

Main Chain-Side Chain Hydrogen Bonds

The krr1 system is characterized by two networks of hydrogen bonds around two of the three loop areas not containing proline, and these networks apparently contribute to the stabilization of the system. In one loop, the hydrogen bonds are between the main chain oxygens of Gly135 and Gly153, which reside on consecutive coils, and the OH group of Ser156, residing at the other edge of the same loop. The other network of hydrogen bonds is between the main chain oxygens of Gly141 and Gly159, which also reside on consecutive coils, and the OH group of Ser162. There is a serine residue only every other coil; thus, altogether there are four
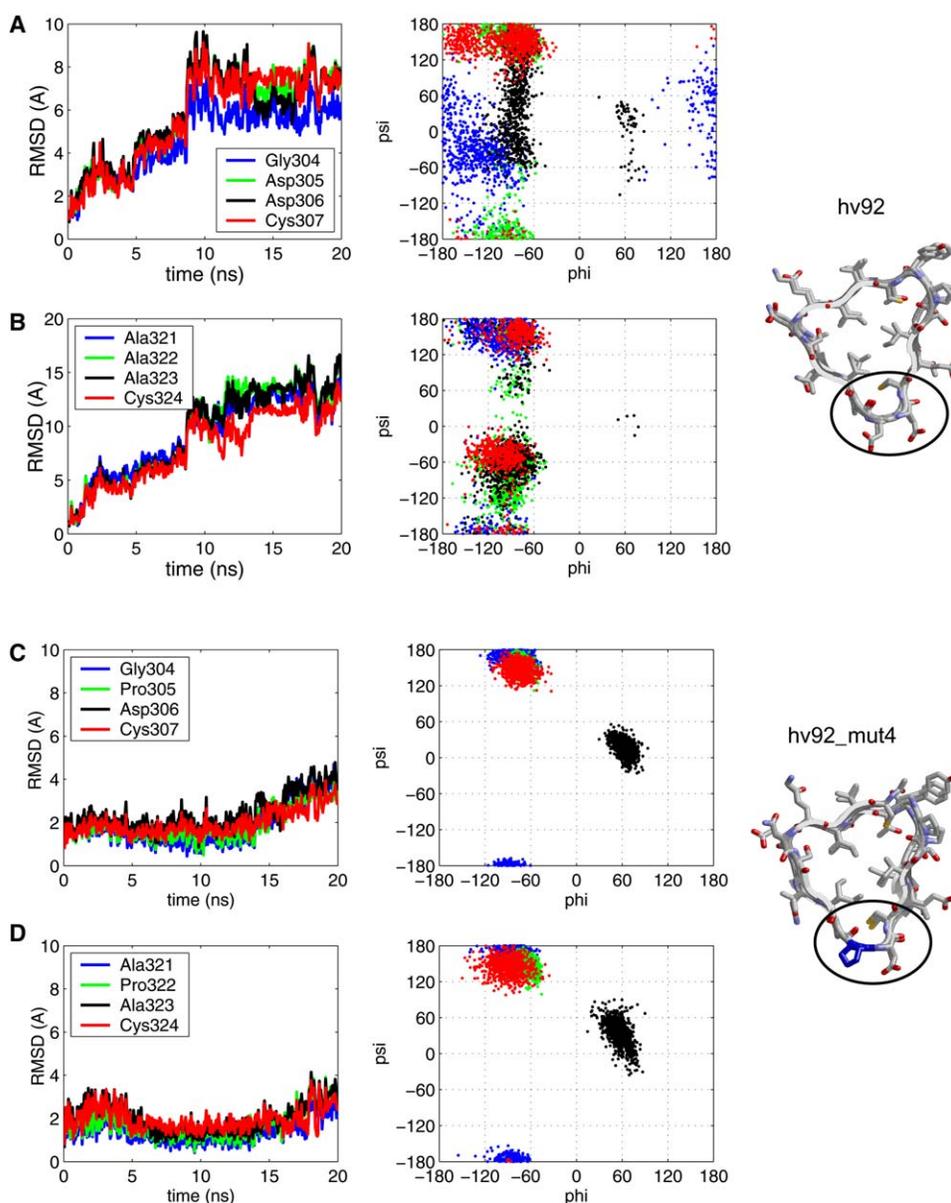
Figure 6. The Distribution of the Main Chain Dihedral Angles, φ and ψ, and the Evolution of the Rmsd for Part of the Loop Residues of hv92 and hv92_mut4 throughout the Simulation

(A–D) The systems are displayed on the right side, and the displayed loop is circled in black. In hv92_mut4, the mutated residues are emphasized in blue. Subfigures (A) and (B) show the evolution of the rmsd (left) and the distribution of main chain angles (right), respectively, for hv92. Subfigures (C) and (D) show the evolution of the rmsd (left) and the distribution of main chain angles (right), respectively, for hv92_mut4. Each residue is displayed in a different color (see inset legends).

serines and eight glycines in each loop over the four repeats. Figure 7 shows the evolution of the two networks of hydrogen bonds along the structure. As can be seen, hydrogen bonds are forming and breaking throughout the simulation, but most of the time at least four hydrogen bonds are present, which implies that all of the serines may be involved in hydrogen bonds most of the time. Other systems we tested, which were less structurally stable, also contained glycine residues around the loops, but in lesser numbers, and the glycine residues were not always positioned in places that enabled them to hydrogen bond with the side chains of other residues. We can conclude that the hydrogen bonds formed

by glycine and serine contribute to the structural organization of a β-helical system. As seen in Figure 3E, when Ser156 was replaced by valine, the rmsd of the system with respect to the initial structure increased compared to the wild-type. However, no significant change in the rmsd was observed when Ser162 was replaced by valine. Judging by the results, we can conclude that the mere presence of glycine residues is not always sufficient to ensure the stability of a loop region of β-helical-based structures. It may assist in stabilizing structures, but only in addition to other interactions. This observation should be taken into account when designing mutants or ab initio structures based on the β-helical fold.
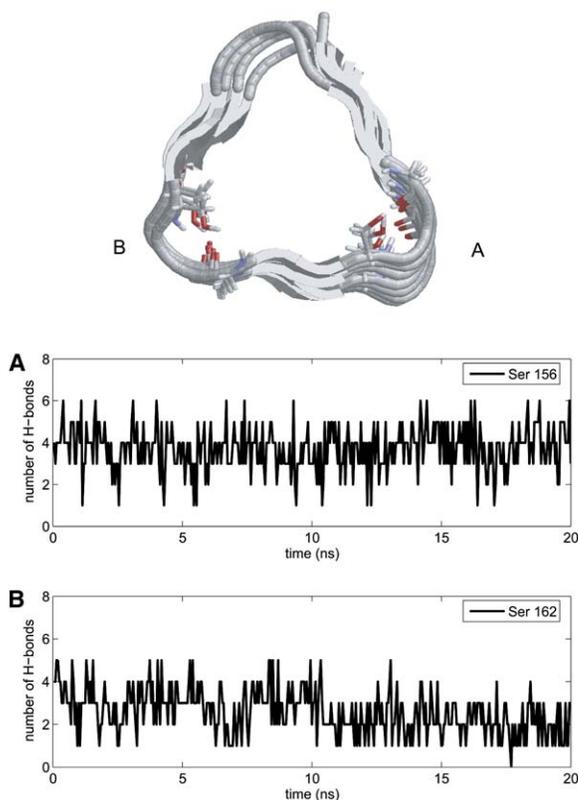
Figure 7. Hydrogen Bond Interactions in the krr1 System

Top: an illustration of the two chains of hydrogen bonds around the loop areas of the krr1 system. The two chains are emphasized in wire-frame. Left: Gly159 and Gly141 hydrogen bonding with the HG of Ser162 is shown. Right: Gly135 and Gly153 hydrogen bonding with the HG of Ser156 is shown.

Bottom: (A) The number of hydrogen bonds formed by Gly135 and Gly153 with Ser156 over time. (B) The number of hydrogen bonds formed by Gly141 and Gly159 with Ser162 over time.

Sequential versus Structural Effects: Grafting Simulations

As discussed above, despite the fact that all of the systems had very similar initial structures, some systems exhibit higher structural stability than others. In order to further try to assess the dependence of the structural maintenance on the sequence versus the initial backbone structure, we grafted the sequence of hv92 onto the backbone conformation of krr1 and repeated the simulation. We also did the opposite grafting, placing the sequence of krr1 onto the backbone of hv92. If indeed only the sequence is responsible for the structural organization of our systems, we should expect the system with the krr1 backbone and the hv92 sequence to be less stable than the original krr1, while the system with the hv92 backbone and the krr1 sequence should be more stable than the original hv92. We tried to place each residue of one system in its respective position on the other, paying special attention to retaining the secondary structure assignment—that is, placing loop residues in loop regions and strand residues in strand regions. However, unfortunately, the grafting was not full, since by examining the backbone structure of krr1 versus hv92, one can see that one hv92 loop is shorter by one amino acid than the other loops, making the tri-

angular structure not equilateral. Consequently, the basic repetitive unit of hv92 contains 34 residues, while krr1 contains 35. As a result, when grafting the krr1 sequence onto the hv92, 1 residue per turn was omitted (a Gly). This happened in two turns. The converse also holds: when grafting the hv92 onto the krr1 backbone, 1 residue in two turns was missing. Consequently, we left the krr1 Gly in its original position in each of those two turns. In addition, the C-terminal loop of hv92 is 1 residue longer. Consequently, when grafting the krr1, the original hv92-terminal glycine and proline were retained. And, again, the converse also holds. Figure 8 shows the sequence and structure of the two grafted systems versus the wild-type. Despite these shortcomings, interestingly, as shown in Figure 3F, both grafted structures were much more structurally organized than hv92 and less organized than krr1; over time, their structural stability is more or less the same. Thus, this result suggests that both backbone and side chain effects contribute to the structural stability of the β-helical systems.

Conclusions

β-helical proteins are promising candidates for nanostructure design for several reasons: they are repetitive, tubular, and symmetrical and thus suitable for designing new nanomaterials of a repetitive nature such as nanofibers and tubes. Their structure is available in protein structure databases, a fact that facilitates the computational design process; they crystallize easily and are readily available from naturally occurring proteins, which makes them easy to produce since we can harvest them from existing molecules in simple methods. Alternatively, they can be synthesized, so their sequences can be better manipulated (Bang and Kent, 2005; Nilsson et al., 2005). Their tube-like structure makes them of potential use for targeted small-molecule delivery, fiber construction, and electron transfer (S. Kent, personal communication). In this work, we constructed nanofibers based on motifs taken from left-handed β-helical proteins and simulated them to test their stability. Our goal was to design nanosystems that show structural stability and can be used as templates for experimental construction of novel systems. We found that a system constructed of four replicas of residues 131–165 of galactoside acetyltransferase exhibited remarkable stability under the simulated conditions, including temperature increase and the addition of ions. We characterized the specific residues and interactions that keep the system structurally stable and found that, apart from the known stacking interactions that characterize left-handed β helices, the presence of proline and glycine residues in the loop regions contributes to the stability of the systems we constructed. The proline residues are very rigid and make the loop region more stable, while the glycine residues are flexible and thus sustain most of the loop movements and help stabilize it. In addition, if positioned in the "right" place, as in the case of krr1, they further contribute by being involved in hydrogen bonds with nearby residues.

In the future, we plan to enhance the stability of the structures by inserting nonnative amino acids. In addition, we plan to exploit our advanced computational abilities to further facilitate the structural design. We are currently developing efficient computational algorithms

A Native hv92 sequence

CVIKNSVIGDDCEISPY
TVVEDANLAAACTIGPF

B Native krr sequence

PITIGNNVWIGSHVVINP
GVTIGDNSVIGAGSIVT-

C Grafted hv92 sequence

PCVIGKNSVIGDDCDISP
YTVVGEDANLGAACTIG-

D Grafted krr sequence

ITINNVWIGSHVVINPG
VTIDNSVIGAGSIVGPP
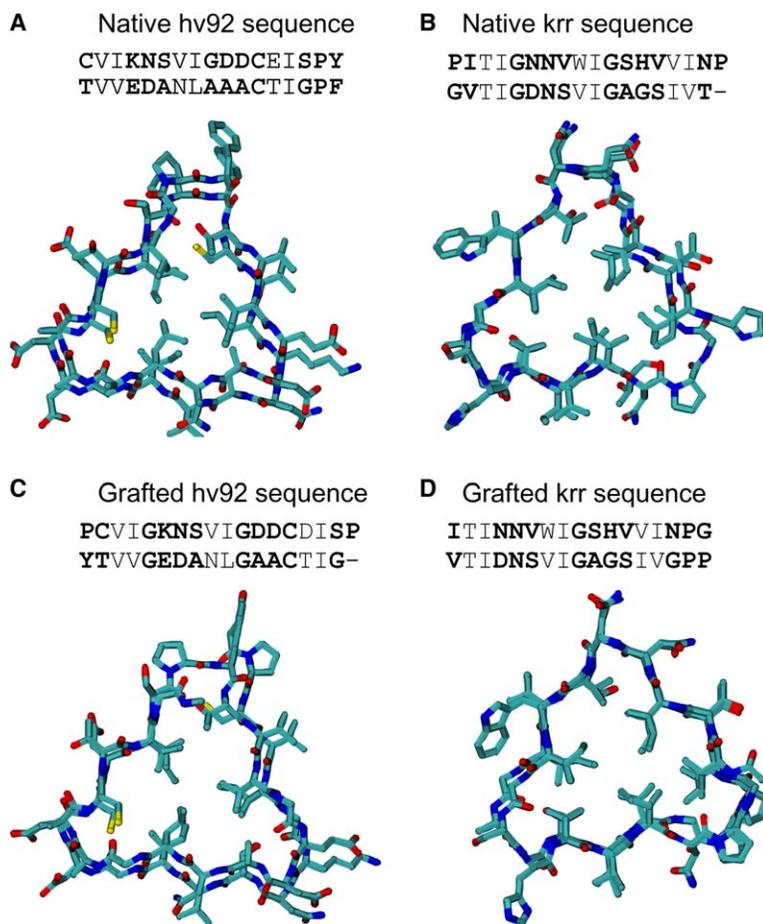
Figure 8. The Sequences and Structures of the Grafted and Wild-Type krr1 and hv92

(A) The sequence and structure of the repetitive unit of the original hv92.
(B) The sequence and structure of the repetitive unit of the original krr1.
(C) The sequence and structure of the repetitive unit of the grafted hv92 (with the krr1 backbone).
(D) The sequence and structure of the repetitive unit of the grafted krr1 (with the hv92 backbone).

for fast prediction of the structure and stability of β sheet-rich systems. The algorithms already give encouraging results in predicting the structure of β-amyloids. We plan to adapt them toward predicting the potential stability of β-helical systems. This would further speed the design, since potentially unstable systems may not even be simulated. Overall, this work emphasizes the advantage of interdisciplinary research that combines computational methods with experimental ones. Computational methods can accelerate the nanodesign process, since they allow many potential molecular assemblies to be probed in a relatively short period of time, thus allowing unstable candidates to be screened quickly.

Experimental Procedures

The calculations were performed by using the NAMD package (Kale et al., 1999). All of the atoms of the system were considered explicitly, and the energy was calculated by using the CHARMM22 force field (MacKerell et al., 1998). The water molecules were represented explicitly, by using the TIP3 model (Jorgensen et al., 1982). The simulations were performed by using the NVT ensemble in an orthorhombic simulation box. We chose constant volume simulations because all of the trajectories were obtained at high temperature. By these means, we could assure that a proper density distribution would not be lost due to thermal effects. Periodic boundary conditions were applied by using the nearest image convention. The box size was adjusted to fit the complex size, so that infinite dilution conditions would be maintained. The box dimensions were 50 × 50 × 70Å to ensure infinite dilution. Each system contained ~15,000–20,000 atoms, including the solvent. The starting molecu-

lar structures were built by using the INSIGHTII molecular package (2000, Accelrys, San Diego, CA). For any given arrangement, we fixed the interturn distance of adjacent repetitive units to match the interstrand distance within each unit, which was ~4.5 Å. The charge of all potential titratable groups was fixed to those values corresponding to neutral pH, such that all aspartic acid side chains were represented in their anionic form and all lysine side chains in their acidic positively charged form. Both peptide edges were capped to avoid interactions between adjacent termini.

The Simulation Conditions
We performed the simulations under the following conditions:

1. No ions in the solution, 300ºK.
2. No ions in the solution, 360ºK.
3. Ionic strength of 0.23% w/w, 300ºK (~8 ions).
4. Ionic strength of 0.5% w/w, 300ºK (~16 ions).
5. Ionic strength of 0.8% w/w, 300ºK (~24 ions).

In the case of ionized solution, we kept the overall charge of the system neutral for the use of EWALD (Darden et al., 1993) particle mesh summation to calculate the electrostatic charges. The ions were chloride and sodium.

Before running each molecular dynamics simulation, the potential energy of each system was minimized by using 5000 conjugate gradient steps. The heating protocol included 15 ps of increasing the temperature of the system from 0ºK to the final temperature of 300ºK (or 18 ps of increasing the temperature from 0ºK to 360ºK) plus 100 ps of an equilibration period. We perform the simulations at 300ºK and 360ºK in order to enhance the stability differences between the models by means of thermal stress. Furthermore, using high temperature allowed us to infer some kinetic tendencies. Residue-based cutoff was applied at 14 Å, i.e., if any two molecules have any atoms within 14 Å, the interaction between them is evaluated. A numerical integration time step of 1 fs was used for all of the

simulations. The nonbonded pair list was updated every 20 steps, and the trajectories were saved every 1000 steps (1 ps) for subsequent analysis. Each simulation was run for a period of 20 ns. Potentially stable systems were run for an additional 20 ns.

This protocol has been used in our research group for a few years (Haspel et al., 2005; Ma and Nussinov, 2002; Zanuy et al., 2003, 2004; Zanuy and Nussinov, 2003) and has been proven to correspond to the experimental results (Luehrs et al., 2005; Petkova et al., 2002; Reches et al., 2002; Zanuy et al., 2004). In order to further validate the results presented in this paper, we also simulated an entire β-helical protein in its native condition (as mentioned in the PDB file. The protein is illustrated in Figure 1A.). The details of the simulations can be found in the Supplemental Data.

Structural Analysis
We calculated the structural conservation in the following ways:

1. Conservation of the size of the structure with respect to the minimized structure: the trajectories were aligned with the initial structure, and the rmsd was calculated with respect to C-α atoms.
2. Conservation of the loops was defined as the rmsd of the C-α of each residue of the loop with respect to the initial minimized structure. In addition, the distribution of the ϕ and ψ dihedral angles was plotted.
3. Sequence alignment and analysis were performed with the CLUSTALX software (Thompson et al., 1997).

Supplemental Data
Supplemental Data include simulations conducted to verify the protocol used by the authors. They also include analyses that are less important for the understanding of the work done in the paper. They are available at http://www.structure.org/cgi/content/full/14/7/1137/DC1/.

References

Ai, H., Jones, S.A., de Villiers, M.M., and Lvov, Y.M. (2003). Nano-encapsulation of furosemide microcrystals for controlled drug release. J. Control. Release *86*, 59–68.

Badjic, J.D., Balzani, V., Credi, A., Silvi, S., and Stoddart, J.F. (2004). A molecular elevator. Science *303*, 1845–1849.

Bae, J., Choi, J.H., Yoo, Y.S., Oh, N.K., Kim, B.S., and Lee, M. (2005). Helical nanofibers from aqueous self-assembly of an oligo(p-phenylene)-based molecular dumbbell. J. Am. Chem. Soc. *127*, 9668–9669.

Bang, D., and Kent, S.B. (2005). His6 tag-assisted chemical protein synthesis. Proc. Natl. Acad. Sci. USA *102*, 5014–5019.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M.

(1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. *112*, 535–542.

Chworos, A., Severcan, I., Koyfman, A.Y., Weinkam, P., Oroudjev, E., Hansma, H.G., and Jaeger, L. (2004). Building programmable jigsaw puzzles with RNA. Science *306*, 2068–2072.

Claussen, R.C., Rabatic, B.M., and Stupp, S.I. (2003). Aqueous self-assembly of unsymmetric peptide bolaamphiphiles into nanofibers with hydrophilic cores and surfaces. J. Am. Chem. Soc. *125*, 12680–12681.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N*log(N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092.

Ferrari, M. (2005a). Cancer nanotechnology: opportunities and challenges. Nat. Rev. Cancer *5*, 161–171.

Ferrari, M. (2005b). Nanovector therapeutics. Curr. Opin. Chem. Biol. *9*, 343–346.

Haino, T., Matsumoto, Y., and Fukazawa, Y. (2005). Supramolecular nano networks formed by molecular-recognition-directed self-assembly of ditopic calix[5]arene and dumbbell [60]fullerene. J. Am. Chem. Soc. *127*, 8936–8937.

Haspel, N., Zanuy, D., Ma, B., Wolfson, H., and Nussinov, R. (2005). A comparative study of amyloid fibril formation by residues 15–19 of the human calcitonin hormone: a single β-sheet model with a small hydrohpobic core. J. Mol. Biol. *345*, 1213–1227.

Jenkins, J., Mayans, O., and Pickersgill, R. (1998). Structure and evolution of β-helical proteins. J. Struct. Biol. *122*, 236–246.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1982). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999). NAMD2: greater scalability for parallel molecular dynamics. J. Comput. Phys. *151*, 283–312.

Kent, S.B. (2004). Novel forms of chemical protein diversity–in nature and in the laboratory. Curr. Opin. Struct. Biol. *15*, 607–614.

Koide, S., Huang, X., Link, K., Koide, A., Bu, Z., and Engelman, D.M. (2000). Design of single-layer sheets without a hydrophobic core. Nature *403*, 456–460.

Koide, S., Yang, X., Huang, X., Dunn, J.J., and Luft, B.J. (2005). Structure-based design of a second-generation Lyme disease vaccine based on a C-terminal fragment of *Borrelia burgdorferi* OspA. J. Mol. Biol. *350*, 290–299.

Li, W., Metcalf, D.G., Gorelik, R., Li, R., Mitra, N., Nanda, V., Law, P.B., Lear, J.D., DeGrado, W.F., and Bennett, J.S. (2005). A push-pull mechanism for regulating integrin function. Proc. Natl. Acad. Sci. USA *102*, 1424–1429.

Liao, S., and Seeman, N.C. (2004). Translation of DNA signals into polymer assembly instructions. Science *306*, 2072–2074.

Luehrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Doebeli, H., Schubert, D., and Riek, R. (2005). 3D structure of Alzheimer's amyloid β-(1–42) fibrils. Proc. Natl. Acad. Sci. USA *102*, 17342–17347.

Ma, B., and Nussinov, R. (2002). Stabilities and conformations of Alzheimer's β-amyloid peptide oligomers (A-β16–22, A-β16–35, A-β10–35): sequence effects. Proc. Natl. Acad. Sci. USA *99*, 14126–14131.

MacKerell, J.A.D., Bashford, D., Bellott, M., Dunbrack, R.L.J., Evanseck, J., and Field, M.J. (1998). All-hydrogen empirical potential for molecular modeling and dynamic studies of proteins using the CHARMM22 force field. J. Phys. Chem. Ser. B *102*, 3586–3616.

Miller, B.G., and Raines, R.T. (2005). Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases. Biochemistry *44*, 10776–10783.

Nilsson, B.L., Soellner, M.B., and Raines, R.T. (2005). Chemical synthesis of proteins. Annu. Rev. Biophys. Biomol. Struct. *34*, 91–118.

Pargaonkar, N., Lvov, Y.M., Li, N., Steenekamp, J.H., and de Villiers, M.M. (2005). Controlled release of dexamethasone from

microcapsules produced by polyelectrolyte layer-by-layer nanoassembly. Pharm. Res. *22*, 826–835.

Percec, V., Dulcey, A.E., Balagurusamy, V.S., Miura, Y., Smidrkal, J., Peterca, M., Nummelin, S., Edlund, U., Hudson, S.D., Heiney, P.A., et al. (2004). Self-assembly of amphiphilic dendritic dipeptides into helical pores. Nature *430*, 764–768.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. (2002). A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. Proc. Natl. Acad. Sci. USA *99*, 16742–16747.

Porod, W., Werblin, F., Chua, L.O., Roska, T., Rodriguez-Vazquez, A., Roska, B., Fay, P., Bernstein, G.H., Huang, Y.F., and Csurgay, A.I. (2004). Bio-inspired nano-sensor-enhanced CNN visual computer. Ann. N Y Acad. Sci. *1013*, 92–109.

Rajagopal, K., and Schneider, J.P. (2004). Self-assembling peptides and proteins for nanotechnological applications (review). Curr. Opin. Struct. Biol. *14*, 480–486.

Rathore, O., and Sogah, D.Y. (2001). Self-assembly of β-sheets into nanostructures by poly(alanine) segments incorporated in multiblock copolymers inspired by spider silk. J. Am. Chem. Soc. *123*, 5231–5239.

Reches, M., Porat, Y., and Gazit, E. (2002). Amyloid fibril formation by pentapeptide and tetrapeptide segmentd of human calcitonin. J. Biol. Chem. *277*, 35475–35480.

Smith, B.D., Soellner, M.B., and Raines, R.T. (2005). Synthetic surfaces for ribonuclease adsorption. Langmuir *21*, 187–190.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *24*, 4876–4882.

Valery, C., Paternostre, M., Robert, B., Gulik-Krzywicki, T., Narayanan, T., Dedieu, J.C., Keller, G., Torre Percec, V., Dulcey, A.E., Balagurusamy, V.S., et al. (2003). Biomimetic organization: octapeptide self-assembly into nanotubes of viral capsid-like dimension. Proc. Natl. Acad. Sci. USA *100*, 10258–10262.

Yan, H., Park, S.H., Finkelstein, G., Reif, J.H., and LaBean, T.H. (2003). DNA-templated self-assembly of protein arrays and highly conductive nanowires. Science *301*, 1882–1884.

Zanuy, D., and Nussinov, R. (2003). The sequence dependence of fiber organization: a comparative molecular dynamics study of the Islet amyloid polypeptide segments 22–27 and 22–29. J. Mol. Biol. *329*, 565–584.

Zanuy, D., Ma, B., and Nussinov, R. (2003). Short peptide amyloid organization: stabilities and conformations of the Islet amyloid peptide NFGAIL. Biophys. J. *84*, 1–11.

Zanuy, D., Porat, Y., Gazit, E., and Nussinov, R. (2004). Peptide sequence and amyloid formation: molecular simulations and experimental study of a human Islet amyloid polypeptide fragment and its analogs. Structure *12*, 439–455.

Zhang, S. (2003). Fabrication of novel biomaterials through molecular self-assembly (review). Nat. Biotechnol. *21*, 1171–1178.