INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY

Phys. Biol. **3** (2006) S54–S62 doi:10.1088/1478-3975/3/1/S06

Concepts and schemes for the re-engineering of physical protein modules: generating nanodevices via targeted replacements with constrained amino acids

Carlos Alemán[1], David Zanuy[1], Ana I Jiménez[2], Carlos Cativiela[2], Nurit Haspel[3], Jie Zheng[4], Jordi Casanovas[5], Haim Wolfson[3] and Ruth Nussinov[4,6]

[1] Departament d'Enginyeria Química, ETS d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, Diagonal 647, Barcelona E-08028, Spain
[2] Departamento de Química Orgánica, Instituto de Ciencia de Materiales de Aragón, Universidad de Zaragoza—CSIC, Zaragoza E-50009, Spain
[3] School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
[4] SAIC-Frederick, Center for Cancer Research Nanobiology Program, NCI-Frederick, MD 21702, USA
[5] Departament de Química, Escola Universitària Politècnica, Universitat de Lleida, C/Jaume II No 69, Lleida E-25001, Spain
[6] Department of Human Genetics, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

E-mail: carlos.aleman@upc.edu and ruthn@ncifcrf.gov

Received 14 October 2005
Accepted for publication 2 February 2006
Published 22 February 2006
Online at stacks.iop.org/PhysBio/3/S54

Abstract

Physically building complex multi-molecular structures from naturally occurring biological macromolecules has aroused a great deal of interest. Here we focus on nanostructures composed of re-engineered, natural 'foldamer' building blocks. Our aim is to provide some of the underlying concepts and schemes for crafting structures utilizing such conformationally relatively stable molecular components. We describe how, via chemical biology strategies, it is further possible to chemically manipulate the foldamer building blocks toward specific shape-driven structures, which in turn could be used toward potential-designed functions. We outline the criteria in choosing candidate foldamers from the vast biological repertoire, and how to enhance their stability through selected targeted replacements by non-proteinogenic conformationally constrained amino acids. These approaches combine bioinformatics, high performance computations and mathematics with synthetic organic chemistry. The resulting artificially engineered self-organizing molecular scale structures take advantage of nature's nanobiology toolkit and at the same time improve on it, since their new targeted function differs from that optimized by evolution. The major challenge facing nanobiology is to be able to exercise fine control over the performance of these target-specific molecular machines.

1. Introduction

Modern chemistry and biology face a major challenge: obtaining fully functional molecular devices for biomedical purposes [1]. The complex nature of a broad variety of diseases, such as fibril-related pathologies, AIDS and cancer, has shown the scientific community the necessity of designing and producing molecular machines able to perform targeted

tasks in multi-cellular organisms. Nanotechnology aims to target the cellular and even the molecular level, rather than the organ or tissue. For this goal to materialize into designed molecular shapes implies a multidisciplinary effort encompassing the physical, chemical and biological sciences over the next decade: it first requires control over the formation of nanodevices and complete characterization of their physical properties; next, it is essential to test the nanostructures *in vitro* under physiological conditions and, finally, *in vivo* in complex multi-cellular organisms.

Success in building such nanodevices requires judicious selection of the appropriate chemical compounds. Yet, the huge number of recently reported nanomaterials makes the task of choosing candidate chemical blocks that can function as molecular scaffolds an overwhelming challenge [2]. Assuming that biological macromolecules are compatible with multi-cellular organisms, their modification appears to be a logical path. However, to simultaneously control the formation of complex molecular structures and their required chemical functions may be a utopian exercise, unless some initial selective steps are taken: first to choose between the two aspects of nano-building, either search for specific functions or focus on nano-organization; and second to generate databases on which correlations between complex organization, stability of the molecular devices and specific functions would be performed. This information should yield the minimal chemical requisites needed to obtain simple building blocks and/or functionality (in terms of chemical activity). Information extracted from known biological structures should indicate where to look for biological applications of the designed devices. Finally, whether we start from one specific molecular structure or from a generic (bio)chemical function, newly desired features can be added by a rational redesign of known simple features which have been characterized previously.

In this paper, we introduce some basic concepts crucial for engineering protein-based molecular devices. We describe some previous achievements combined with new ideas that we believe will impact new molecular designs.

2. From nature to nanotechnology

2.1. Controlling nanostructure organization: from self-assembly to biological functional blocks

Building nanodevices implies the ability to control the organization of complex structures at the molecular level, joining selected building blocks into new coherent assemblies with known shapes. Thus, nano-organization transforms into a more extensive concept: it is not simply the geometrical ways of assembling molecular modules but understanding the global shapes that different molecular motifs can present after their association. Such modular assembly leads to organizations that are beyond the molecular level. Stable associations of molecules at size levels between nanometer and micrometer are known as supramolecular complexes [3] and can be found in nature in numerous types of shapes ranging from ribosomes to enzyme complexes [4, 5].

How can we exercise control over the assembly process? That is, how can we guide a minimal block into the 'right' supramolecular organization? One potentially simple way to obtain new materials is through a self-assembly process [2, 6]. Thus, formation of new compounds does not derive from complex chemical reactions but through the spontaneous association of small parts via non-covalent interactions. The resulting compounds have similar physical features to those obtained by chemical synthesis, and the driving forces that bring the parts together are simple physical interactions such as hydrogen bonds, electrostatic attraction and the hydrophobic effect. Actually, nature presents many examples of such self-assembled complexes, with the cell membrane organization [7] and the DNA double-helix [8] among the most well-known cases. Self-assembly of short peptides was even proposed to have a role in the very early stages of life [9]. Still, within the context of protein nanotechnology, self-assembled structures whose constituents are short peptides form amyloid-like fibrils [10–13] and nanotube assemblies [14, 15]. Remarkably, peptides as short as dipeptides have been shown to form nanotubes [15, 16]. Surfactant-like peptides have also been shown to form nanotube and nanovesicles [17].

Finally, it remains to select structural motifs for a particular complex task in the biological medium. Again, the potential number of choices could be close to infinite if the search starts from scratch [18]. Nonetheless, in the biological context we may not need to search among the many potential building blocks to achieve specific functional architectures: nature has already done this work and natural macromolecules have been selected over eras of life to perform very accurate and specific intra- and extra-cellular functions [19]. From this 'natural library', we can pick those specific pieces that potentially suit our structural requirements, for instance sequences that have a very strong tendency to form a given regular conformation in water. Under such circumstances, to innovate molecular constructs, we may start from architectures of known shapes that could constitute basic motifs. By combining the natural blocks, the self-assembly and altering some chemical features of the target molecules, it should be possible to design and build nanostructures with controlled physical properties. The last step would involve implementation, adaptation and optimization of the chemical function in the new molecular architectures, with the ultimate challenge of making them molecular machines.

2.2. The choice between structures versus molecular shapes

Naturally occurring biological macromolecules provide a rich repertoire of molecular segments. However, the question arises as to how to initiate the selection process: should we start by looking for secondary structure motifs [20], or focus only on protein interfaces, the actual surfaces of the molecular assembly? We believe that to make nanotechnology a successful endeavor, the key is a shape-oriented search: we need to look for geometrical features that would fulfill the structural requirements. A good example can be found in the viral capsid organizations, which are based on the assembly of regular geometrical blocks [21]. We may again take advantage

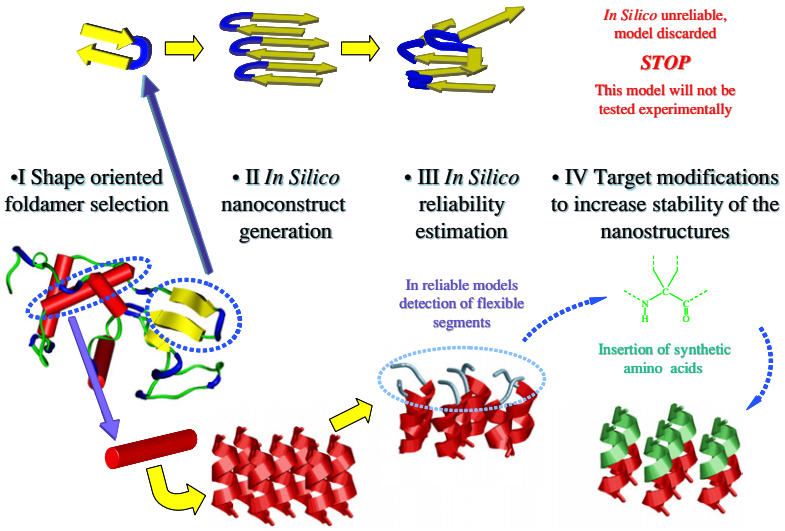

Figure 1. Pictorial flowchart of the newly presented strategy for designing biocompatible stable nanostructures: starting from a natural protein fold, conformational motifs are selected on the basis of shape-oriented mathematical descriptors. The reliability of the segments as potential foldamers is tested *in silico* by energy-based molecular simulations. Those nanoconstructs that are not reliable *in silico* (top) are discarded. Those nanostrustrures that do not disintegrate are subjected to structural and physical characterizations (bottom): the segments of the nanoconstruct that might present excessive conformational flexibility can further be targets for the incorporation of conformationally restricted amino acids.

of the clues provided by nature to build new devices that would combine our needs and the natural adaptation of the selected molecular blocks.

The guidelines for the shape selection have not yet been established (see section 4) and mathematical formulation is crucial. Once a systematic set of mathematical descriptors of shapes are assembled, the selection should be based on three basic criteria (see figure 1): (i) which geometrical shape can perform the desired chemical activities; (ii) which molecules with this precise shape can be found in nature and finally, the most critical, (iii) what are the possible geometrical combinations allowed for these motifs so that they can form the self-assembled organizations *in silico* and *in vitro*. Therefore, the shape-oriented search should include consideration of the potential association of molecular motifs through the investigation of the sequence–structure relationships on the interacting surfaces of their complexes.

The choice of shapes, instead of explicit structural motifs, presents clear advantages: it is possible to expand the building blocks search to include a broad spectrum of macromolecules that might significantly differ in terms of sequence, structure and function but may nevertheless share molecular shape features. This strategy may allow a combinatorial association of natural blocks for the design of new potential molecular associations. Searching for proper shapes fulfills two prerequisites: it is a function-oriented search and it may detect potential assembly sites, since to adjust natural building blocks to perform new functions we would probably need to assemble them in a new way. Obviously, selecting complementary shapes will improve the efficiency of the design. Complementarity implies going beyond geometrical characteristics, extending to chemical and physical features.

Shape-oriented search implies that we do not need to be concerned with achieving correct folding since we have already started from stable conformational blocks. We would only need *to reduce the conformational freedom of the most flexible residues*. However, targeting specific points of a stable shape may allow us to introduce a low degree of interference in the overall organization of the motif and in the inter-motif association.

In summary, the novelty of this approach is based on modification of naturally selected functions of certain molecular motifs by using tools offered by physicochemistry, organic synthesis and molecular biology.

3. Selection of molecular blocks

3.1. Looking for the proper foldamers

If we explore the literature, we may find different definitions of the *foldamer* concept. We believe that the one given by Hill *et al* [18] addresses perfectly what should drive the search for the nanodevice scaffolds (we quote): ‘. . . foldamer is any oligomer that folds into a conformationally ordered state in solution, the structures of which are stabilized by a collection of non-covalent interactions between nonadjacent monomer units . . . .’. Thus, the development of nanomaterials can be described as the design (or redesign) of foldamers that would contain the necessary structural motifs to allow complex chemical functions. Hence, if naturally occurring proteins are to be used as the material source, the segments selected based on shape must also be foldamers. These peptide foldamer modules fold into conformationally ordered states in solution and are further stabilized by non-covalent interactions. Foldamers may be likened to domains, subdomains, foldons [18] or building blocks of the protein structure. They have at least some hydrophobic core which increases the population time of the native conformation. As we have recently

shown via molecular simulations, these scaffolds are relatively stable on their own in water. Their hydrophobicity limits the exploration of local minima beyond the compact native conformation, ensuring the lowest amount of contacts between the polar solvent and the hydrophobic core [22].

Finally, if those segments do not favor the formation of coherent organizations, systematic chemical modifications must be introduced in order to transform those conformationally 'unreliable' segments into suitable foldamers. A plausible way to carry out such a modification is outlined below. Proteins are not the unique source of natural foldamers. Other research groups use similar principles, searching other biomacromolecules: for instance, Jaeger and his colleagues [23] have undertaken a conceptually related approach using RNA.

3.2. Improving nature's work

Nature has selected protein conformations to perform specific tasks. Flexibility, dynamics, stiffness and molecular shape, among others, are features that can be associated with natural building blocks and that have been chosen by evolution. When choosing one specific structural shape to fabricate nanoconstructs, we should remember that in nature it was selected to perform functions that may be unrelated with what we intend to accomplish, and that the environment may be different. That is, the molecular piece we want to transform needs to have a conformational behavior that suits our goals: in order to grow a nanodevice we need to select a 'good' foldamer, able to drive the assembly of these blocks. However, the crystallographic data that would guide the search could lead toward blocks that cannot exist on their own when isolated from the whole protein, or whose natural conformational properties are not optimal for the self-assembly processes. Nature may select sequences in which the native conformation may have a low population time, and it is stabilized through interactions with its binding partners [24].

Even though these conformational features are well tailored for the native protein function, they may be an impediment in molecular engineering. Here, we are often interested in small segments extracted from large proteins, which would probably reduce the conformational stability of the chosen segments. In consequence, it is necessary to adopt some systematic strategy to select those minimal blocks that present foldamer features and that are suitable for the self-assembly requirements. Equally important are the design of methodologies to enhance the stability of those foldamer conformations.

Since the early steps in protein engineering [25], in order to improve the thermodynamic stability of a given protein conformation the first choice has been the targeted mutagenesis. Key residues in the most mobile segments have been selectively mutated by similar, naturally occurring amino acids with slightly different conformational properties. For instance, if the presence of an α-helix was required, amino acids which are less frequent in helical conformations would be substituted by others with similar chemical features but with higher propensity to form helices. This strategy is not, however, devoid of complications. First, when introducing other proteinogenic amino acids, except for Gly (which lacking a side chain, is very flexible), the degrees of conformational freedom are not changed drastically [26]. That is, the intrinsic preference of a given amino acid to adopt particular conformations is unchanged when inserting it into the new sequence. Under these conditions, the interactions between the side chains of the rest of the protein and the new amino acid could dramatically disturb the final organization [27]. Thus, using other proteinogenic amino acids may imply changing more degrees of freedom than experimentalists can control. Moreover, these strategies are always functional-group oriented, which imposes structural manipulation below the nanometer level. This is where *chemical biology* can be enormously advantageous, allowing us to vastly increase and, in principle, manipulate at will the required conformational properties at targeted points.

3.3. Conformationally restricted amino acids

Substitution of natural amino acids in protein sequences implies a generic choice from a prescribed set of other 19. However, we may further modify and tailor the conformational preferences of specific sequences/segments: recognizing the technical limitations, we can reduce the degrees of freedom by using blocks that present much lower conformational flexibility, i.e. non-proteinogenic amino acids with restricted conformational properties. Our goal would be to engineer these to have their restricted conformational preference matching the target site in the fold and at the same time, considering its sequence environment. Although this area has been the center of an intense investigation in the field of bioorganic chemistry [28], it is less known in the biological disciplines. For this reason, in this section we briefly summarize the essential characteristics of the most representative examples of constrained α-amino acids.

Stereochemical constraints can easily be achieved by introducing modifications in natural proteinogenic amino acids. Such modifications may involve changes in the amino acid skeleton or alteration of the peptide bond. The most commonly modified peptide bond is the retroamide. The conformational impact of retromodification on different amino acids has been reported earlier [29, 30] (figure 2). In spite of the fact that other alterations of the peptide bond have received less attention than retromodification, the conformational impact produced by the replacement of the amide link by the *N*-amino amide [31], *N*-hydroxamide [32], thioamide [33] and sulfonamido [34] groups in peptides has also been investigated (figure 2). The overall results revealed that amino acids with modified amide groups differ from the standard amino acids not only in the conformational preferences but also in the hydrogen bonding capabilities [35]. These results appear very useful toward the design of new building blocks with potential applications in nanobiology.

Among the modified α-amino acids, α,α-dialkylated analogs are probably the most extensively studied. This is because the substitution of the α-carbon hydrogen by an alkyl group induces drastic changes in the conformational

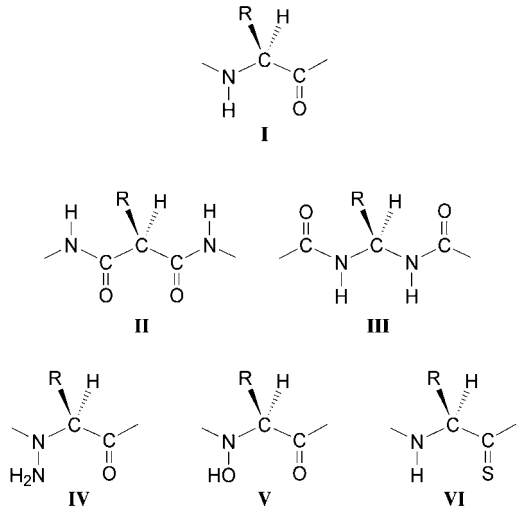

Figure 2. Schematic representation of different modifications of the peptide bond: (**I**) general structure of a proteinogenic amino acid; (**II**, **III**) retroamide; (**IV**) *N*-amino amide; (**V**) *N*-hydroxamide; (**VI**) thioamide.

properties. Indeed, such a modification is frequently associated with the incorporation of strong stereochemical constraints. The simplest α,α-dialkylated amino acid is the α-aminoisobutyric acid (Aib), also denoted as α-methylalanine (figure 3), in which a methyl group replaces the α-carbon hydrogen of Ala. The strong propensity of Aib-based peptides to adopt helical conformations (of either the 3_{10}- or the α-type) was demonstrated in a number of early theoretical and experimental studies [36–40]. Comparison between the potential energy surfaces of Aib and Ala indicates that the incorporation of the second methyl group severely reduces the flexibility of the backbone, with the conformational freedom of the former amino acid being much more restricted than that of the latter [40]. α-methyl derivatives of other proteinogenic amino acids (denoted in general as (αMe)Xaa, figure 3) have been the subject of several studies [36, 37], showing a high preference for helical conformations, as the prototype Aib.

The cyclic homologue of Aib, 1-aminocyclopropanecarboxylic acid (Ac_3c, figure 3), exhibits strong stereochemical constraints induced by the highly strained cyclopropane ring. The strong influence of the bond between the two β-carbon atoms on the conformational freedom of the backbone can be understood by comparing the conformational preferences of Aib and Ac_3c [41]. As for Aib, folded conformations are preferred over extended dispositions, but the particular electronic properties of the three-membered ring lead to a preference for the *bridge* ($\psi = 0^\circ$) instead of the helical region of the conformational map [36, 37, 41]. Higher homologues of Ac_3c, 1-aminocycloalkanecarboxylic acids (Ac_nc with $n = 4, 5, 6$) have been shown to parallel the behavior of Aib [36, 37].

In addition to the stereochemical constraints associated with α,α-dialkylation, substituted derivatives of Ac_3c (c_3Xaa, figure 3) allow control of side chain orientation. The incorporation of the α- and β-carbons in a cyclic structure prohibits rotation about the C^α–C^β bond and, as a consequence, the side chain adopts a well-defined orientation, which is fixed by the stereochemistry of the cyclopropane ring. The selectively oriented side chains may play a critical role in directing the backbone folding, as demonstrated for the cyclopropane analogs of phenylalanine [41–45]. The most remarkable example found to date is a cyclopropane amino acid able to induce a double γ-turn (an incipient 2.2_7-helix) in a linear dipeptide in the crystalline state [45]. Such a structural motif was predicted but never characterized before in linear peptides based exclusively on proteinogenic amino acids. The ability of the side chain to modulate the backbone conformation has also been evidenced for the cyclohexane analogs of phenylalanine [46].

α,β-dehydroamino acids (ΔXaa, figure 3) are another interesting family of constrained amino acids. The double bond linking the α- and β-carbons affects the electronic and stereochemical properties of these compounds and is responsible for their conformational behavior [47–52]. The smallest member of the family, α,β-dehydroalanine (ΔAla, figure 3), accommodates the fully extended conformation, giving rise to planar structures [49, 50]. In contrast,

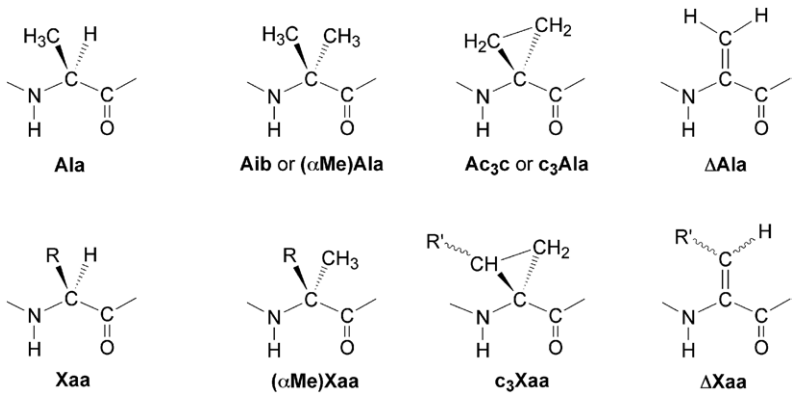

Figure 3. Restriction of the conformational flexibility of a proteinogenic amino acid (Xaa) by α-methylation [(αMe)Xaa], cyclopropanation (c_3Xaa) or α, β-dehydrogenation (ΔXaa). The prototype of each family, i.e. the corresponding alanine analog, is represented.

the other extensively studied unsaturated amino acid, α, β-dehydrophenylalanine, has been shown to induce folded helical conformations, since the fully extended arrangement leads to a steric clash between the rigidly oriented side chain and the peptide backbone [51, 52].

Finally, an extra stabilization of a given conformation can be achieved by introducing simultaneously two or more modifications in a natural proteinogenic amino acid. For instance, a 3_{10}-helix conformation was designed using a sequence that combines α,α-dialkylated amino acids and retroamide bonds [53]. This was performed taking advantage of the intrinsic helix forming tendency of Aib and the hydrogen bonding network imposed by retromodification, which is compatible with a 3_{10}-helix but not with an α-helix.

Thus, we may expect that in the near future chemical compounds that have for decades relegated to organic chemistry curiosities, would become basic structural motifs that would allow nanoengineers the control over the conformational preferences of natural protein blocks. In order to efficiently fabricate nanodevices, we need to control the conformational behavior of the macromolecules that would form the molecular assemblies. This control becomes unreliable when dealing with natural proteins since nature needed hundreds thousands years to develop the proper molecular machinery. However, we could bypass this limiting step by rationally inserting such synthetic residues in the natural sequences and shift the conformational equilibrium toward a particular conformation.

4. Bioinformatics: emerging tools in nanobiology

4.1. Catching shapes in the PDB

To design new shapes from protein building blocks, we need to be able to *a priori* explore whether the candidate block shapes match each other and can be put together. Since the number of building blocks can be very large, matching their shapes, in principle in all rotations and all translations, is an extremely time-consuming chore. To perform this task, we need to consider two ingredients: molecular representation and geometrical matching. The third ingredient is the chemical matching which we address in detail below. Depending on the resolution we wish to attain, there are three possible ways in which we can describe (and match) shapes from the Protein Data Bank (PDB). The first, highest resolution is an all-atom surface description, using the coordinates of points deposited on the protein surface. The second is the atomic coordinates of the backbone of the protein. Finally, the coarsest approach uses the folds in their secondary structure vector representation. This latter description allows matching protein folds. Here, we focus on how one can efficiently match proteins using the first, detailed protein surface description. However, we note that a similar approach can be applied to the coarser representations. Since coarser representations have fewer shape descriptors, their matching will be considerably faster.

For the first, highest resolution, the protein surface is usually represented by its geometric features. Connolly laid the foundation for the detailed protein surface description. The Connolly surface consists of the part of the van der Waals surface of the atoms that is accessible to the probe sphere (contact surface) connected by a network of convex, concave and saddle shape surfaces that smoothes over the crevices and pits between the atoms. On the basis of the Connolly analysis [54, 55], the surface may be described by sparse critical points [56] defined as the projection of the gravity center of a Connolly face. To align the shape of the surfaces of two molecules in a complementary manner, we need to compute a rigid transformation that superimposes the surfaces without allowing one molecule to deeply penetrate, or overlap the other. To quickly obtain hypotheses for such transformations, it suffices to align a triplet of ordered non-collinear points (congruent triangles) from both molecules. The points are those describing the molecular surfaces. These are computed to accurately represent the maxima (*holes*) and minima (*knobs*) of the shape function [57]. Sparseness is important, since the complexity of the matching algorithm depends on the number of points. A surface normal, associated with each point, is also computed. These points are dubbed *critical points*. The strategy then reduces to matching only pairs of critical points with the additional geometric information of their surface normals. In order to compute a candidate rigid transformation, we need to detect a pair of critical points in both molecules that share the same internal distance, and, if superimposed, have opposing surface normals. This reduces the number of potential-matched configurations, and concomitantly reduces the run-time complexity of the program.

The PatchDock algorithm [58] provides an example of how this description can be used to efficiently match the protein surfaces. PatchDock exploits the fact that in order to compute a rigid motion between a pair of molecules, it is enough to detect local complementarity between the shapes. That is, we can focus directly on aligning complementary features, without the need to perform an exhaustive search in the six-dimensional (6D) space of rotations and translations. In practice, the surfaces of the molecules are first divided into patches according to rough surface shape (concave, convex or flat), using the sparse-point molecular shape representation, and a surface segmentation algorithm. Complementary patches are identified and superimposed by shape-matching techniques.

Databases of protein shapes have already been created [59]. However, for the re-engineering of the protein modules we need an efficient clustering of the building blocks using the various types of their shape features. This will enhance the combinatorial matching toward newly designed shapes.

4.2. Application of molecular simulation techniques to predict the structure of self-assembled systems

Obtaining nanostructures based uniquely on a random combination of complementary foldamer shapes may be inefficient. In this context, any methodology that would allow the rational evaluation of relative and/or absolute stabilities of any new potential nanoconstructs would become a main tool to enhance the efficiency of the designing and manufacturing process.

Molecular simulation techniques, i.e. essentially molecular dynamics and Monte Carlo, have turned out to be the emerging methodological approaches that fulfill the demand for rational molecular engineering [60]. The combination of the exponential increase in the available computational power and the development of fast and relatively simple methods for computing the cohesion energy of molecular systems puts these computational strategies in the front line of the nanotechnological challenge. If the computational simulations are accurate and their reliability and limitations are under control, the development of molecular machines with specific medical and technological applications can be consistently accelerated by rapid testing of new ideas prior to experimental assays. Currently, it is becoming possible to systematically estimate *in silico* the structural cohesion stability, the structural features and the weak points of the engineered shape by means of energy-based methodologies [61].

This would then be the missing link of our working flowchart: the computational exploration of the potential assemblies of the blocks under conditions that mimic natural environments (see figure 1). Thus, those assemblies consisting of complementary building blocks and whose molecular shapes fit a particular function would be tested under the main environmental conditions: thermal stress, solvent influence (disassembly due to solubilization), ionic strength and hydrophobic effects. Instead of massively testing those molecular entelechies through tedious, slow and expensive experimental assays, we could accurately select potential candidates with higher chances of success using molecular simulation methods. This strategy does not ensure success, although its reliability was recently demonstrated when studying different types of molecular complexes: the stabilities of the assemblies of both proteinogenic and polypeptidic aggregates have already been successfully probed and understood by means of molecular simulations [62, 63].

However, such tests should not be the only aspects explored *in silico*. Detailed atomistic simulations are powerful techniques, since they allow mapping and identifying molecular parts that could be re-engineered using chemical tools which account for conformational flexibility [64]. Some molecular segments might have been selected by nature to be very flexible; however, in the construct, they may need to be rigidified to avoid amorphous aggregations. The simulations may point to the residues that should be mutated with the conformationally constrained amino acids as discussed above.

In principle, such a strategy can go one step further. These techniques can be used not only to indicate positions to be substituted but also to simulate the effect of the substitutions on the conformational stability and assembly features. With a unique set of simulations, we can estimate the relative stabilities of engineered complex series, their conformational features selectively, and target those points that may enhance the nanoconstruct cohesion. Therefore, the design can be accelerated by reducing the time spent on the trial and error process when modifying natural molecules. A combination of theory and experimental assays should lead to higher chances of success in nanotechnological projects.

5. Summary

Devising molecular machines is increasingly becoming a key challenge in modern science. Here, we provide an overview of concepts and schemes for the re-engineering of nanodevices based on molecular structures. We outline a new approach: starting from *pre-existing* molecular shapes, i.e. from parts of natural biomacromolecules whose structures are available, we computationally select shapes that can potentially be manipulated to allow spontaneous formation of supramolecular assemblies. Supramolecular arrangements meeting the stipulated prerequisites would be subject to an *a posteriori* alteration through the introduction of chemical derivatives, toward the desired molecular functionality. In order to obtain stable building-block foldamers suitable for nanomolecular architecture, conformationally restricted amino acids are suggested as a useful tool. These residues increase the spontaneous coherent organization of the building blocks. Such a strategy integrates chemical biology which allows fine targeted manipulations of biological macromolecules with polymer science, biology, physics and mathematics. Further, combining mathematical descriptors for molecular shapes with massive use of molecular simulations would speed up the selection process. The energy-based molecular simulations can point molecular positions where synthetic amino acids with particular conformational preferences should be inserted into the natural sequences. Finally, only those nanostructures shown to favor coherent multimolecular organization *in silico* would be tested *in vitro*, saving human and material resources that random combinatorial search would imply and accelerating nanodesign.

Acknowledgments

The authors are indebted to CESCA (Centre de Supercomputació de Catalunya) and CEPBA (Centre Europeu de Paral.lelisme de Barcelona) for computational facilities. We gratefully acknowledge the support by the Center of Excellence, funded by the Israel Academy of Sciences. 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 NO1-CO-12400. 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 US Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

[1] Balzani V, Credi A and Venturi M 2003 *Molecular Devices and Machines—A Journey into the Nano World* (Weinheim: Wiley)

[2] Forster S and Plantenberg T 2002 From self-organizing polymers to nanohybrid and biomaterials *Angew. Chem. Int. Ed.* **41** 688–714

[3] Lawrence D S, Jiang T and Levett M 1995 Self-assembling supramolecular complexes *Chem. Rev.* **95** 2229–60

[4] Fromont-Racine M, Senger B, Saveanu C and Fasiolo F 2003 Ribosome assembly in eukaryotes *Gene* **313** 17–42

[5] Bamford D H, Gilbert R J C, Grimes J M and Stuart D I 2001 Macromolecular assemblies: greater than their parts *Curr. Opin. Struct. Biol.* **11** 107–13

[6] Sastry M, Rao M and Ganesh K N 2002 Electrostatic assembly of nanoparticles and biomacromolecules *Acc. Chem. Res.* **35** 847–55

[7] Anderson P W and Stein D L 1987 *Self-organizing Systems* ed I Yates and F Eugene (New York: Plenum)

[8] Seeman N C and Belcher A M 2002 Emulating biology building nanostructures from the bottom up *Proc. Natl Acad. Sci. USA* **99** 6451–5

[9] Panchenko A R, Luthey-Schulten Z, Cole R and Wolynes P G 1997 The foldon universe: a survey of structural similarity and self-recognition of independently folding units *J. Mol. Biol.* **272** 95–105

[10] Reches M, Porat Y and Gazit E 2002 Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin *J. Biol. Chem.* **277** 35475–80

[11] Naito A, Kamihira M, Inoue R and Saito H 2004 Structural diversity of amyloid fibril formed in human calcitonin as revealed by site-directed 13C solid-state NMR spectroscopy *Magn. Reson. Chem.* **42** 247–57

[12] Maji S K, Drew M G and Banerjee A 2001 First crystallographic signature of amyloid-like fibril forming beta-sheet assemblage from a tripeptide with non-coded amino acids *Chem. Commun. (Camb.)* **7** 1946–7

[13] Nelson R, Sawaya M R, Balbirnie M, Madsen A O, Riekel C, Grothe R and Eisenberg D 2005 Structure of the cross-beta spine of amyloid-like fibrils *Nature* **435** 773–8

[14] Lu K, Jacob J, Thiyagarajan P, Conticello V P and Lynn D G 2003 Exploiting amyloid fibril lamination for nanotube self-assembly *J. Am. Chem. Soc.* **125** 6391–3

[15] Song Y, Challa S R, Medforth C J, Qiu Y, Watt R K, Pena D, Miller J E, van Swol F and Shelnutt J A 2004 Synthesis of peptide-nanotube platinum-nanoparticle composites *Chem. Commun. (Camb.)* **9** 1044–5

[16] Reches M and Gazit E 2003 Casting metal nanowires within discrete self-assembled peptide nanotubes *Science* **300** 625–7

[17] Vauthey S, Santoso S, Gong H, Watson N and Zhang S 2002 Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles *Proc. Natl Acad. Sci. USA* **99** 5355–60

[18] Hill D J, Mio M J, Prince R B, Hughes T S and Moore J S 2001 A field guide to foldamers *Chem. Rev.* **101** 3893–4011

[19] Petsko G A and Ringe D 2003 *Protein Structure and Function* (London: New Science)

[20] Zhao X and Zhang S 2004 Fabrication of molecular materials using peptide construction motifs *Trends Biotechnol.* **22** 470–6

[21] Morgan G J 2003 Historical review: viruses, crystals and geodesic domes *Trends Biochem. Sci.* **28** 86–90

[22] Tsai H H, Tsai C J, Ma B and Nussinov R 2004 *In silico* protein design by combinatorial assembly of protein building blocks *Protein Sci.* **13** 2753–65

[23] 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** 2068–72

[24] Gunasekaran K, Tsai C J, Kumar S, Zanuy D, Ma B and Nussinov R 2003 Extended disordered proteins: targeting function, with less scaffold *TiBS* **28** 81–5

[25] Drexler K E 1981 Molecular engineering: an approach to the development of general capabilities for molecular manipulation *Proc. Natl Acad. Sci. USA* **78** 5275–8

[26] Kent S 2004 Novel forms of chemical protein diversity—in nature and in the laboratory *Curr. Opin. Biotechnol.* **15** 607–14

[27] Dobson C M, Sali A and Karplus M 1998 Protein folding: a perspective from theory and experiment *Angew. Chem. Int. Ed.* **37** 868–93

[28] Venkatraman J, Shankaramma S C and Balaram P 2001 Design of folded peptides *Chem. Rev.* **101** 3131–52

[29] Fletcher M D and Campbell M M 1998 Partially modified retro-inverso peptides: development, synthesis, and conformational behavior *Chem. Rev.* **98** 763–96

[30] Alemán C 2001 A computational study of partially modified retro-inverso valine dipeptides: effect of the side chain on the conformational preferences of malonyl and gem-diamino residues *J. Chem. Phys.* B **105** 860–6

[31] Alemán C 2002 N-amination of peptides: a theoretical study *J. Phys. Chem.* A **106** 1441–9

[32] Devin C, Bernad N, Cristau M, Artis-Noel A M, Heitz A, Fehrentz J A and Martinez J 1999 Synthesis and biological evaluation of c-terminal hydroxamide analogues of bombesin *J. Pept. Sci.* **5** 176–84

[33] Artis D R and Lipton A 1998 Conformations of thioamide-containing dipeptides: a computational study *J. Am. Chem. Soc.* **120** 12200–06

[34] Baldauf C, Gunther R and Hofmann H J 2004 Conformational properties of sulfonamido peptides *J. Mol. Struct. (Theochem.)* **675** 19–28

[35] Alemán C 2001 On the ability of modified peptide links to form hydrogen bonds *J. Phys. Chem.* A **27** 6717–23

[36] Benedetti E 1996 X-ray crystallography of peptides: the contribution of the italian laboratories *Biopolymers (Pept. Sci.)* **40** 3–44

[37] Toniolo C, Crisma M, Formaggio F and Peggion C 2001 Control of peptide conformation by the Thorpe–Ingold effect (Cα-tetrasubstitution) *Biopolymers (Pept. Sci.)* **60** 396–419

[38] Alemán C, Roca R, Luque F J and Orozco M 1997 Helical preferences of alanine, glycine, and aminoisobutyric homopeptides *Proteins* **28** 83–93

[39] Karle I L and Balaram P 1990 Structural characteristics of helical peptide molecules containing Aib residues *Biochemistry* **29** 6747–56

[40] Toniolo C, Crisma M, Formaggio F, Peggion C, Broxterman Q B and Kaptein B 2004 Molecular spacers for physicochemical investigations based on novel helical and extended peptide structures *Biopolymers (Pept. Sci.)* **76** 162–76

[41] Alemán C, Jiménez A I, Cativiela C, Pérez J J and Casanovas J 2002 Influence of the phenyl side chain on the conformation of cyclopropane analogues of phenylalanine *J. Phys. Chem.* B **106** 11849–58

[42] Jiménez A I, Cativiela C, Aubry A and Marraud M 1998 β-turn preferences induced by 2,3-methanophenylalanine chirality *J. Am. Chem. Soc.* **120** 9452–9

[43] Moye-Sherman D, Jin S, Ham I, Lim D, Scholtz J M and Burgess K 1998 Conformational preferences of RNase A C-peptide derivatives containing a highly constrained analogue of phenylalanine *J. Am. Chem. Soc.* **120** 9435–43

[44] Royo S, De Borggraeve W M, Peggion C, Formaggio F, Crisma M, Jiménez A I, Cativiela C and Toniolo C 2005 Turn and helical peptide handedness governed exclusively by side-chain chiral centers *J. Am. Chem. Soc.* **127** 2036–7

[45] Jiménez A I, Ballano G and Cativiela C 2005 First observation of two consecutive γ turns in a crystalline linear dipeptide *Angew. Chem. Int. Ed.* **44** 396–9

[46] Jiménez A I, Cativiela C, Gómez-Catalán J, Pérez J J, Aubry A, Paris M and Marraud M 2000 Influence of side chain restriction and NH$\cdots\pi$ interaction on the β-turn

folding modes of dipeptides incorporating phenylalanine cyclohexane derivatives *J. Am. Chem. Soc.* **122** 5811–21
[47] Alemán C and Casanovas J 1995 Molecular conformational-analysis of dehydroalanine analogs *Biopolymers* **36** 71–82
[48] Thormann M and Hofmann H-J 1998 Conformational properties of peptides containing dehydroamino acids *J. Mol. Struct. (Theochem.)* **431** 79–96
[49] Crisma M, Formaggio F, Toniolo C, Yoshikawa T and Wakamiya T 1999 Flat peptides *J. Am. Chem. Soc.* **121** 3272–8
[50] Palmer D E *et al* 1992 Effects of dehydroalanine on peptide conformations *J. Am. Chem. Soc.* **114** 5634–42
[51] Jain R and Chauhan V S 1996 Conformational characteristics of peptides containing α,β-dehydroamino acids residues *Biopolymers (Pept. Sci.)* **40** 105–19
[52] Mathur P, Ramakumar S and Chauhan V S 2004 Peptide design using α,β-dehydro amino acids: from β-turns to helical hairpins *Biopolymers (Pept. Sci.)* **76** 150–61
[53] Alemán C 1997 Chain conformation in polyretropeptides: III. Design of a 3_{10}-helix using α,α-dialkylated amino acids and retropeptide bonds *Proteins* **29** 575–82
[54] Connolly M L 1983 Solvent-accessible surfaces of proteins and nucleic acids *Science* **221** 709–13
[55] Connolly M 1983 Analytical molecular surface calculation *J. Appl. Cryst.* **16** 548–58
[56] Lin S L and Nussinov R 1996 Molecular recognition via face center representation of a molecular surface *J. Mol. Graph.* **14** 78–90
[57] Norel R, Petrey D, Wolfson H and Nussinov R 1999 Examination of shape complementarity in docking of unbound proteins *Proteins* **35** 403–19
[58] Schneidman-Duhovny D, Inbar Y, Nussinov R and Wolfson H J 2005 Geometry-based flexible and symmetric protein docking *Proteins* **60** 224–31
[59] Binkowski T A, Naghibzadeh S and Liang J 2003 Computed atlas of surface topography of proteins *Nucleic Acids Res.* **31** 3352–5
[60] Landman U 2005 Materials by numbers: computations as tools of discovery *Proc. Natl Acad. Sci. USA* **102** 6671–8
[61] Jorgensen W L and Tirado-Rives J 2005 Potential energy functions for atomic-level simulations of water and organic and biomolecular systems *Proc. Natl Acad. Sci. USA* **102** 6665–70
[62] Zanuy D, Porat Y, Gazit E and Nussinov R 2004 Peptide sequence, molecular assembly and amyloid formation: molecular simulations and experimental study of a human islet amyloid polypeptide fragment and its analogues *Structure* **12** 439–55
[63] Zanuy D, Casanovas J and Alemán C 2004 Thermal stability of the secondary structure of poly(α,l-glutamate) in self-assembled complexes as studied by molecular dynamics in chloroform solution *J. Am. Chem. Soc.* **126** 704–5
[64] Paschek D, Gnanakaran S and Garcia A E 2005 Chemical theory and computation special feature: simulations of the pressure and temperature unfolding of an α-helical peptide *Proc. Natl Acad. Sci. USA* **102** 6765–70