Chimia 54 (2000) 552–557

ISSN 0009–4293

Evolution *versus* Design: Template-Directed Self-Assembly of Peptides to Artificial Proteins (TASP)

Manfred Mutter and Gabriele Tuchscherer*

Abstract: Protein design and mimicry combines elements of synthetic organic chemistry with structural and functional aspects of biological relevance in a unique way. Due to progress in this interdisciplinary research field, access to molecules featuring some essential properties of native proteins appears to be within reach, enabling the complex mechanisms in molecular recognition processes to be deciphered. Since its introduction by the authors, the template approach in protein *de novo* design (Template Assembled Synthetic Proteins, TASP) has experienced a broad conceptual diversification. Starting from today's state-of-the-art in protein design, we present here some ongoing work in the Lausanne laboratories focusing on the use of regioselectively addressable templates and TASP scaffolds for addressing fundamental questions in peptide assembly, protein folding and mimicry. It is shown that the developed concepts can ideally reconcile evolutionary and rational design principles for creating molecules of biological and therapeutic interest.

Keywords: Peptide assembly · Protein design · Protein mimicry · Supramolecular chemistry · TASP · Topological templates

Introduction

The formation of polypeptides that fold into unique three-dimensional entities of immense structural and functional complexity represents one of the most fascinating events in the evolutionary process. While random arrays of amino acid residues linked by amide bonds would adopt multiple conformations in solution, Nature selected some well-defined polypeptide sequences exhibiting the potential for folding into molecular-scale machines of diverse, highly sophisticated functionality.

Actually, there is some controversial discussion on whether or not the long-standing protein-folding problem has been resolved to some extent [1][2], culminating in the question: Is it possible to predict the native structure of a protein starting uniquely from a given primary sequence? In view of the enormous body of sequence information from gene analysis, the answer to this question will be of utmost importance for fulfilling the high ambitions and expectations in today's genome and proteome research activities. A tedious, but reliable way for unraveling a protein's complexity represents the construction of models of well-defined structural parameters. Here, the study of peptides with potential for adopting secondary structure and their subsequent assembly into folding units ('supersecondary structures') mimicking some essential properties of native proteins, *e.g.* globular, monomeric state, cooperative unfolding, buried hydrogen bonds *etc.* proved to be a powerful approach. If general rules for the hierarchical onset of secondary and tertiary structure formation could be delineated, the construction of novel proteins with predetermined structural properties would be within reach; even more appealing, the complex interplay between structure and function could become a target of utmost theoretical and practical relevance. For some unknown reasons, Nature has developed the ribosomal machinery for producing exclusively linear polypeptide chains and during the time scale of evolution, learned how to fold them into unique three-dimensional macromolecules. In the absence of this 'know-how', *de novo* design aims to define its own rules for generating protein-like constructs of tailored properties which resulted in some conceptually different approaches (Fig. 1). What is the minimal chain length of a polypeptide for folding into a thermodynamically stable unit? First results on supersecondary structure motifs (Fig. 1a, 1b) [3][4] indicate a critical size of about 30 amino acid residues, which is in harmony with strategies aiming at 'downsizing' native pro-

*Correspondence: PD. Dr. G. Tuchscherer
University of Lausanne
Institute of Organic Chemistry; BCH
CH-1015 Lausanne
Tel.: +41 21 692 40 63
Fax: +41 21 692 39 55
E-Mail: Gabriele.Tuchscherer@ico.unil.ch

teins and still retaining their major functions (Fig. 1c, 1h) [5][6]. To avoid the intrinsic protein-folding problem, we have proposed the use of topological templates as a built-in device for directing peptide blocks to a well-defined structural entity (Fig. 1d) [7][8]. Meanwhile, this template-assembled synthetic proteins (TASP) concept has found broad acceptance for mimicking some complex properties of native proteins such as metal or substrate binding, electron transfer (Fig. 1e) [9], catalysis or membrane channel formation (Fig. 1f) [10]. As an extension of the template approach, secondary structure elements have been assembled to multi-branched constructs of reduced conformational space ('locked-in folds' such as the zinc finger motif, (Fig. 1g) [11]. Finally, the use of stable scaffolds ('minibodies', Fig. 1h) [6] or regioselectively addressable templates for the assembly of binding loops (Fig. 1i) [12] opened new perspectives in protein mimicry. In the present article, we focus on some recent work in template-directed protein design ongoing in the Lausanne laboratories.

The Template Concept: TASP for the Study of Supramolecular Assembly

Topological templates can be regarded as synthetic chaperonins, directing covalently attached amphipathic peptide blocks to well-defined folding units such as 4α-helical bundles [13]. Despite the variety of templates proposed so far (*e.g.* porphyrins, cyclodextrins, calixarenes, steroids), cyclic peptides have some distinct advantages due to their convenient synthetic accessibility and their broad structural and functional variability. Typically, cyclic decapeptides consisting of two antiparallel β-strands connected by two β-turns serve as regioselectively accessible functional templates (RAFT) [14], featuring up to four orthogonal protected attachment sites in well-defined spatial orientations (Fig. 2).

Recently, the solution and solid state conformation of such a prototype template has been resolved by NMR and X-ray [15][16] spectroscopy, confirming the structural parameters as postulated by molecular modeling simulation (Fig. 2). Up to now, this type of template has been used for the construction of a number of mainly 4α-helical bundle TASP molecules that have been designed with the aim to mimic some of the essential structural and functional properties of native proteins [17].

Meanwhile the major principles governing secondary structure formation are well understood, but the interplay of energetic factors such as the hydrophobic effect, van der Waals packing or Coulomb forces in the assembly process of the constituting structural elements into entities of higher complexity is still the central question in today's research on protein folding and design [18]. Here, the template concept offers a versatile tool for the experimental evaluation of these energetic parameters. For example, competitive trapping of pre-assembled helical peptides on reactive templates allows for the investigation of the role of the helix macrodipole, internal packing or Coulomb interactions in four-helix bundle formation, providing valuable insight in protein folding processes, self-assembly phenomena and supramolecular structure formation (Fig. 3).

In our ongoing studies in collaboration with Jane Richardson from Duke University on this subject we start from a four-helical bundle TASP derived from ROP (Repressor of Primer) protein, featuring optimal internal packing parameters. In a typical experiment, two competing helices of differential packing capacity are assembled on a scaffold consisting of a two-helical TASP [19]. First experiments point to a pronounced effect of the packing parameters upon helix-orientation and thermodynamic stability of the

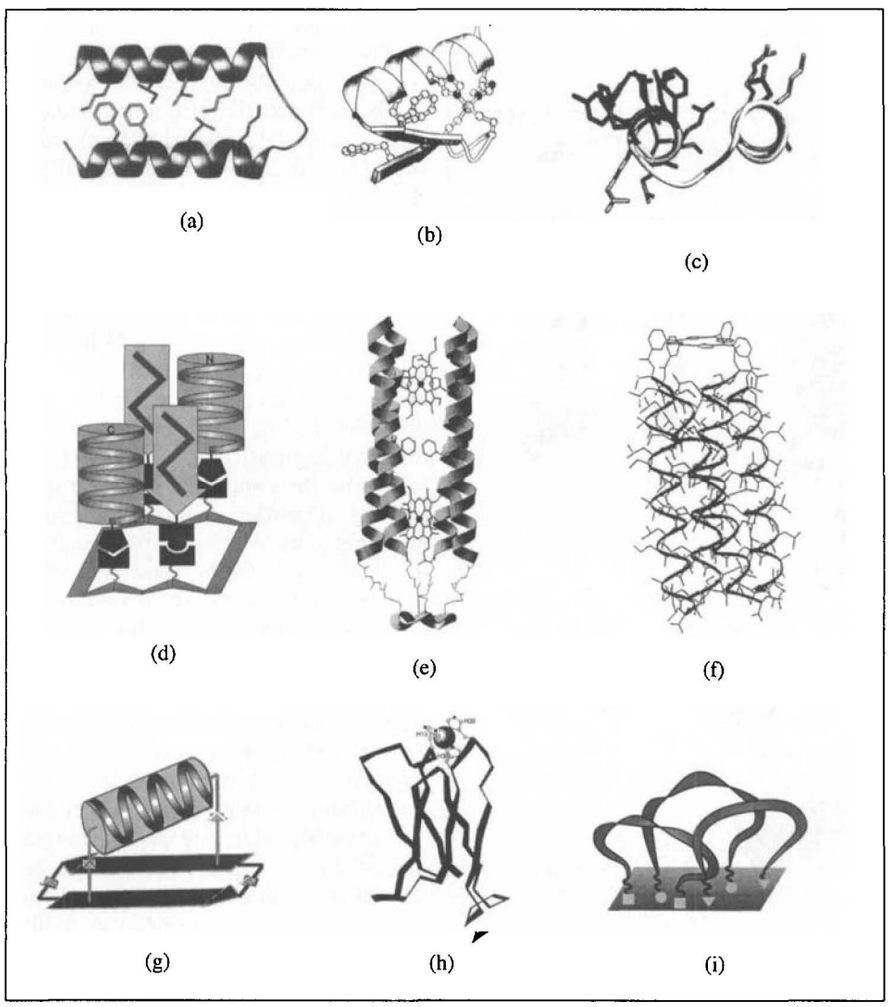

Fig. 1. Some representative examples of concepts for the *de novo* design of folding motifs exhibiting tailored functional properties. a) design of helical bundles from scratch, *e.g.* helix-loop-helix motifs [3]; b) redesign of zinc finger folding motifs [4]; c) downsizing of protein folds to the minimal functional domain on the example of protein A [5]; d) *de novo* design of 4α-helical bundles of non-native chain connectivities, termed template assembled synthetic proteins (TASP) for bypassing the protein folding problem [7] [8]; e) helical bundles as redox proteins [9]; f) membrane ion channel forming 4α-helical bundle TASP [10]; g) zinc finger fold as an oligocyclic locked-in tertiary fold [11]; h) protein subdomains as scaffolds for inserting ligand binding loop [6]; i) receptor mimetics using regioselectively addressable functional templates (RAFT) for the assembly of binding loops [12]

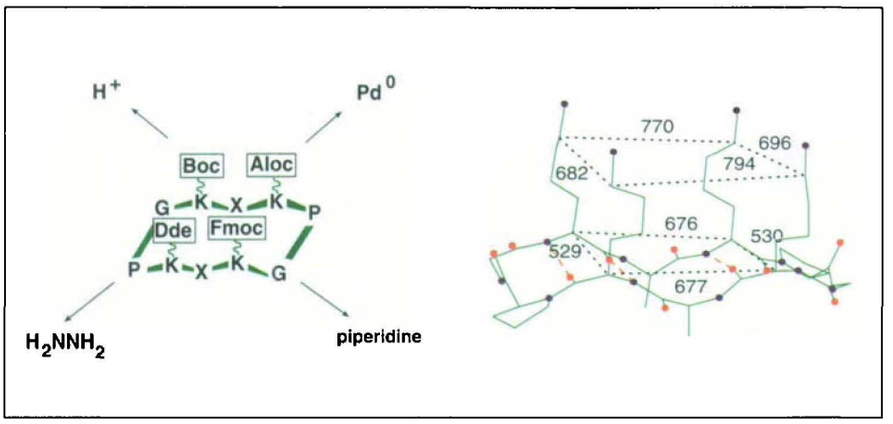

Fig. 2. Left: Schematic representation of a prototype decameric template molecule featuring four orthogonal protecting groups selectively removable with acid (Boc), piperidine (Fmoc), hydrazine (Dde), and Pd catalysis (Aloc). Right: Side-view of a molecular modeling picture of the template in a low energy conformation with the four Lys side chains as attachment sites oriented above the plane of the template in the appropriate geometry for ligation of four peptide fragments. Approximate distances of the C^{α} and C^{ε}-atoms of the lysine residues are given in nanometers. This hypothetical structure has recently been confirmed by X-ray and NMR [16].

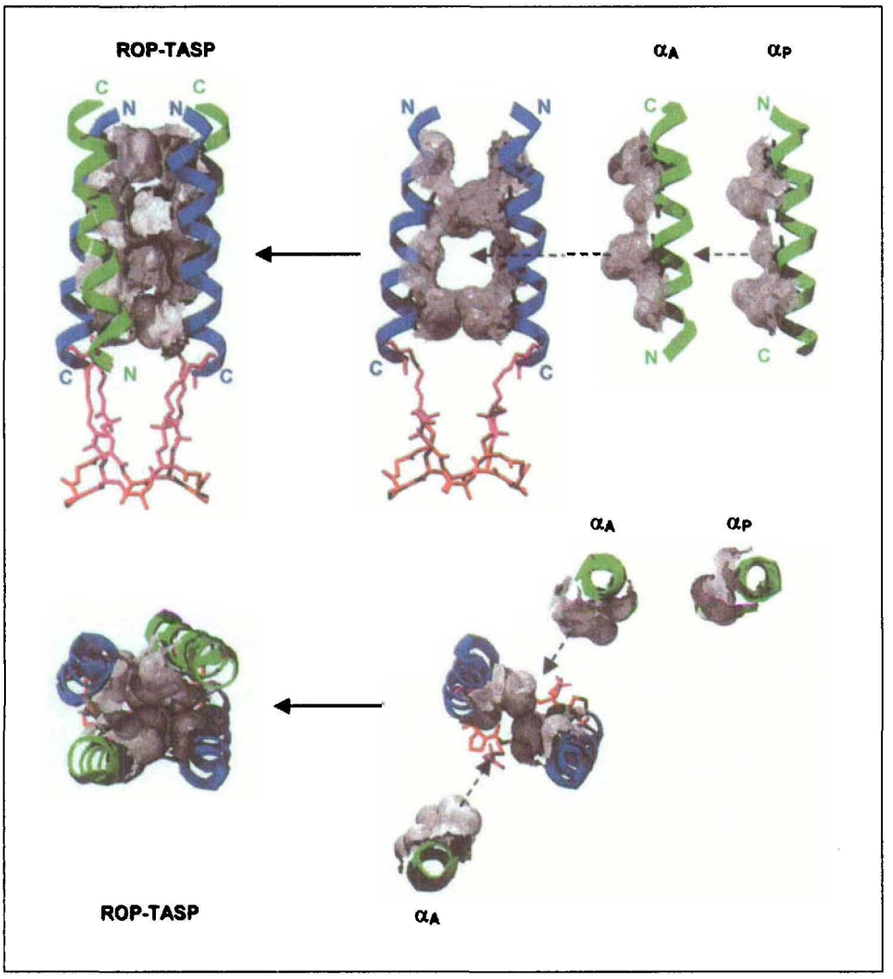

Fig. 3. Template trapping [19]: Ribbon representation of the competitive association process of helices to a two-helical TASP scaffold for evaluating the importance of complementary packing in peptide/protein assembly. Helices (α_A) designed to optimally complement the 2-helix TASP scaffold are strongly favored in the assembly process over helices (α_P) with suboptimal packing parameters (top: side view; bottom: top view). The pre-assembled bundle arrangements are stabilized by covalent fixation ('trapping') of the helices allowing for detailed structural analysis. Side chain surfaces of the core residues are shown to highlight the complementarity of packing.

resulting 4-α bundle TASP, giving important clues for future protein design. In combination with combinatorial methods, the present methodology of template trapping contributes significantly to our understanding of supramolecular assembly in molecular recognition processes.

Locked-in Folds as Functional Protein Mimetics

Due to recent progress in the methodology of peptide synthesis, protection group chemistry and chemoselective ligation techniques, the assembly of peptides to macromolecules of well-defined architecture and tailored functions becomes increasingly appealing. In completely bypassing Nature's complex way of protein folding, the chemoselective ligation of peptide fragments to locked-in folds (LIF) according to a molecular kit system allows the construction of molecular scaffolds exhibiting increased thermodynamic stability compared to their native counterparts (Fig. 4). As a prototype of this new generation of TASP molecules, the DNA-binding $\beta_2\alpha$-folding motif of a zinc finger protein (Zif) was represented as a locked-in fold featuring a cyclic decapeptide as scaffold for the covalent attachment at both chain ends of the functional helix applying oxime and thioether ligation procedures [11]. The results on the conformational and metal binding properties of this locked-in Zif encouraged us to design a trimeric Zif of minimal size with the final goal of accessing synthetic Zif molecules of tailored DNA-binding properties (Fig. 5) [20].

To this end, the 26-mer helical block of the native Zif was reduced to a 13-mer helix in keeping the DNA-binding residues as well as the two histidines for zinc complexation. Simultaneously, the β-strand-turn-β-strand motif was reduced to a standard cyclic decapeptide template comprising two zinc binding Cys residues that are directed towards the helical block. Again, chemoselective ligation procedures allow for efficient convergent syntheses, *i.e.* solid phase synthesis of linear segments, assembly of the unprotected fragments to monomeric Zif modules and subsequent trimerization *via* oxime bond formation. Preliminary experiments on the DNA-binding capacity suggest that synthetic Zif mimetics may become versatile targets in DNA recognition studies of tremendous therapeutic potential.

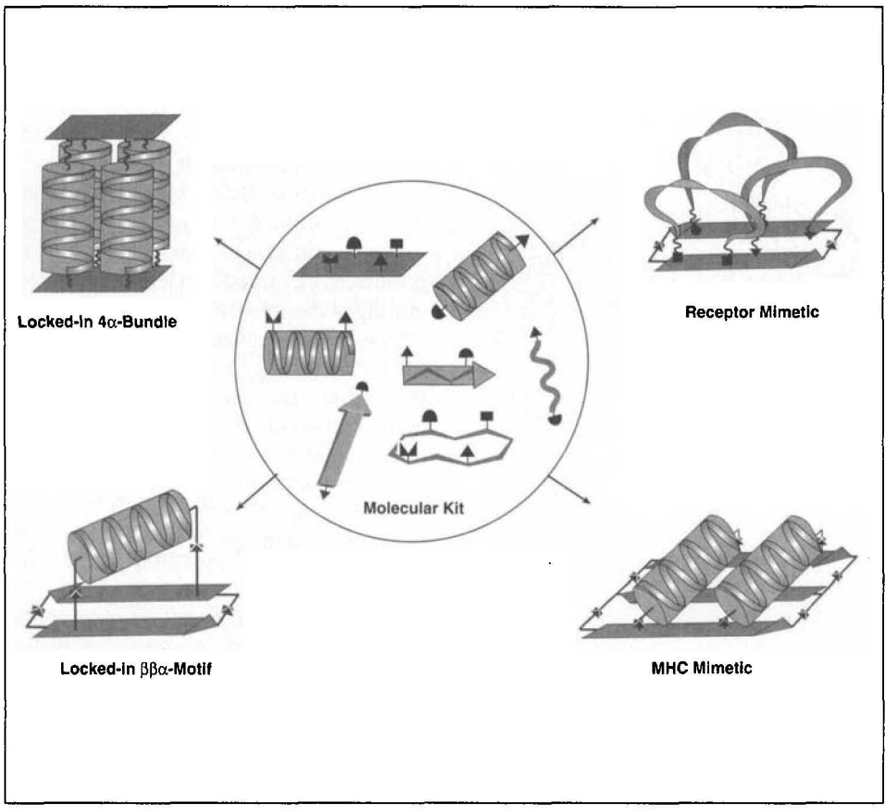

Fig. 4. Locked-in tertiary folds as an extension of the TASP concept are molecules with a built-in pathway for folding [8]. Based on the principles of a molecular kit, helices, β-sheets, turns and loops as molecular bricks are assembled covalently *via* both chain ends to appropriately functionalized templates resulting in multibridged molecules, *e.g.* locked-in four helix bundles, template-assembled recognition loops (receptor mimetics), zinc finger or MHC structural motifs.

Grafting Recognition Sites to TASP Scaffolds

The construction of protein-like folding motifs as structurally stable scaffolds for the introduction of 'function' represents one of the major objectives in protein design. This concept is based on the presentation of random sequences in a geometrically restricted manner, *e.g.* through grafting of randomized positions onto conformationally constrained scaffold-proteins [21]. For engineering novel binding sites into proteins, foreign loop sequences [22–25] or metal binding sites [26] have been transferred to non-related proteins but still maintaining their native functional properties. However, only proteins with sufficiently large regions at the molecular surface that are highly tolerant to substitutions or insertions without losing the overall three-dimensional structure are suitable scaffolds. But rather than recruiting naturally existing proteins or domains for further engineering, the design of such presentation scaffolds *de novo* has become an attractive target [3–5][8–11][17].

In conceptually separating structure from function, we have, in close collaboration with the Division of Hematology at the University Hospital of Lausanne, designed and synthesized a chimeric 4-helix bundle TASP derived from ROP protein and the cell adhesion glycoprotein E-selectin (Fig. 6). This mimetic is aimed at inhibiting an early stage in cell adhesion processes, in particular leukocyte adhesion to endothelial cells. For this purpose, the core region of the antiparallel homodimeric ROP protein was redesigned to ensure formation of a stable 4-helix bundle structure independently of the structurally non-related binding surface of E-selectin to be grafted [18]. The helices as fundamental building blocks have been designed on the basis of general features for helix stabilization, *e.g.* incorporation of helix-stabilizing residues, side chain-side chain electrostatic interactions, C- and N-caps and support of the helix macrodipole by adding positively or negatively charged residues near the C- or N-termini. These idealized α helices are arranged in an antiparallel fashion with the core residues of ROP (Leu, Ala), to ensure favorable hydrophobic interactions in the interior of the four-helix bundle TASP as scaffold for the transfer of the molecular surface onto the surface of such a bundle structure (Fig. 6). Crucial residues of E-selectin for binding to its counter-receptor PSGL-1 (P-selectin glycoprotein ligand-1) as determined by X-ray and mutagenesis studies have been matched onto the surface formed by two neighboring helices in the TASP molecule in the appropriate geometry based on the coordination of the Cα atoms, including the Ca^{2+} complexing site. As a special feature, the resulting TASP contains two identical surfaces each resembling the binding site as found in native E-selectin comprising 13 residues that are involved either in ligand or calcium binding.

The above example demonstrates the potential of using TASP molecules as three-dimensional scaffolds for grafting receptor or acceptor sites of complex native proteins. Further refinements of the system are in progress for accessing native-like protein models of biomedical interest.

Template-assembled Binding Loops as Receptor Mimetics

The described hierarchical approach in reducing complex native proteins to their minimal structural elements and maintaining the desired functional properties (see also examples in Fig. 1) is of pivotal importance for understanding biological recognition processes on a molecular level. Based on the principles of a Lego system with *e.g.* helices, templates and loops as molecular bricks, the template approach can be applied for downsizing functional proteins to active site mimetics. In separating the structural and functional part of a protein as recently proposed, the binding loops of a receptor or antibody molecule are grafted onto a regioselectively addressable template (RAFT) molecule mimicking the structure supporting framework of the native molecule [12]. Topological templates featuring up to three sequentially different binding loops on tailored RAFT molecules seem to be within reach applying state-of-the-art peptide chemistry [27]. This opens a new route for the construc-

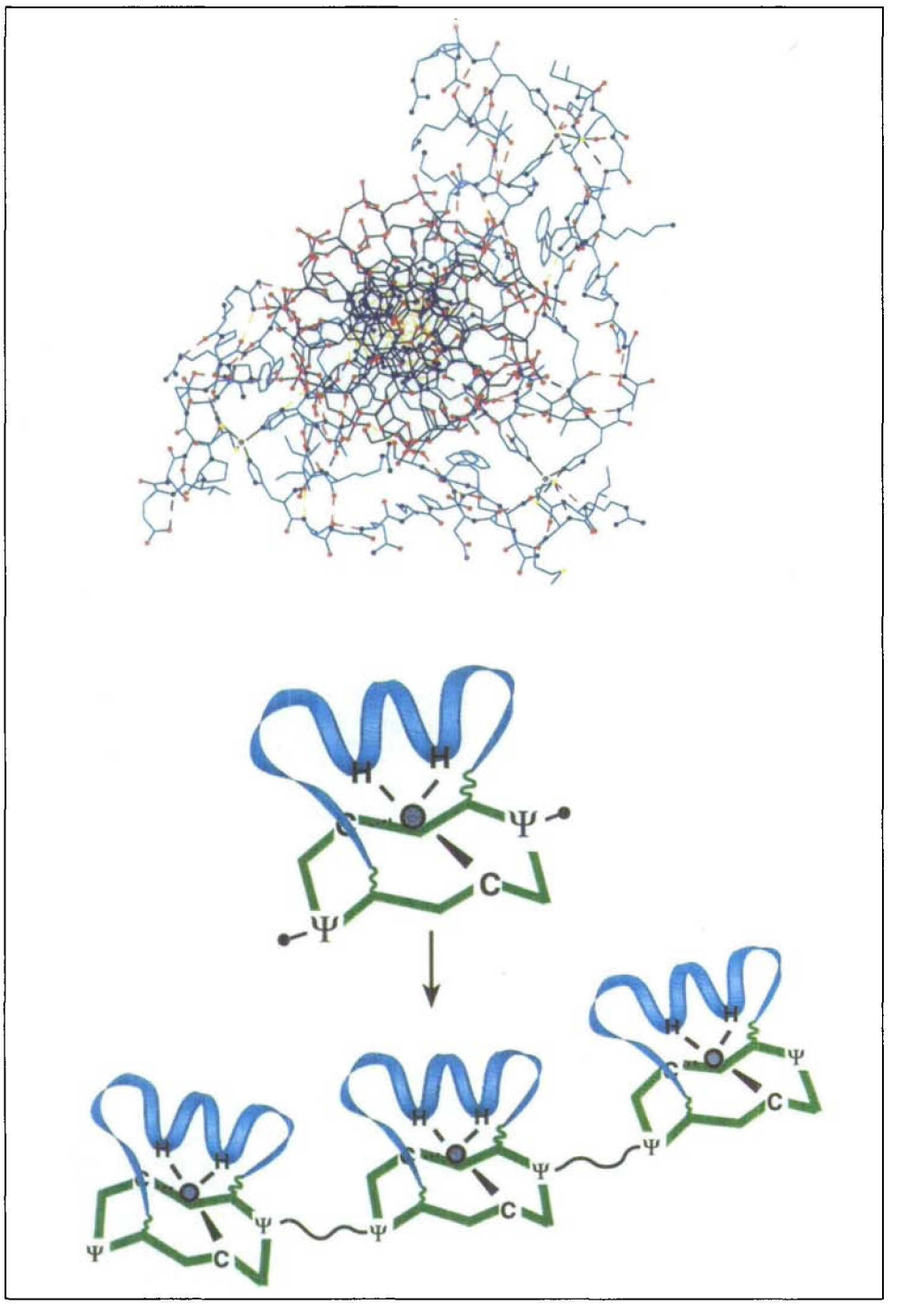

Fig. 5. Top: X-ray structure of the trimeric zinc finger (Zif) of the transcription factor IIIa in the DNA bound state (view along the DNA axis). Bottom: Scheme for the chemical synthesis of a trimeric locked-in Zif; two pseudo-prolines featuring functionalities for chemoselective ligation replace the proline residues in the turns of the template molecule; subsequently, individual Zif modules are covalently ligated *via* the pseudo-proline building blocks to form the trimer.

tion of TASP molecules mimicking just some of the essential features of antibodies, receptors or even enzymes.

Reconciling Evolution and Design

Rational design and combinatorial chemistry still appear as two conceptually diverging rather than converging approaches in the search for therapeutically relevant molecules. Only recently, ideas evolve to reconcile both strategies in creating 'intelligent libraries'. In combining rational recognition elements such as helices, binding loops, RAFT molecules with combinatorial principles (*e.g.* ligand-directed self-assembly to constructs of higher complexity) in applying affinity selection as functional screening, we presently evaluate these innovative concepts (Fig 7) [28].

Where are the limits of mimicking native proteins? More explicitly: Did Nature economize on the way to create proteins of the well-known complexity? It seems to be realistic to assume that the price for simplification of structure is loss of function such as specificity, selectivity or activity. Consequently, protein mimicry and design should at present be preferably regarded as a learning process aiming at deciphering Nature's complexity in molecular recognition processes. Still, in targeting specific properties such as binding, inhibition or catalysis, synthetic constructs as described above may become more than creative visions in diagnostics and drug development. Furthermore, apart from representing an enormous intellectual challenge of considerable practical potential, the evolutionary principles supplementing the state-of-the-art chemistry in biomimetic chemistry reconciles in a perfect way Nature's rules and chemist's tools [29].

Acknowledgements

The Swiss National Science Foundation and the Canton Vaud support our work in this field.

Received: August 16, 2000

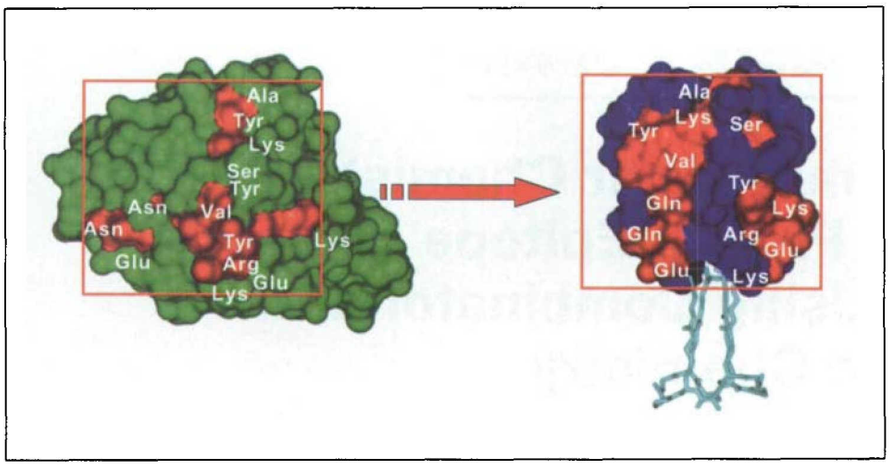

Fig. 6. Chimeric TASP [18]: Grafting of a molecular surface onto a four-helix bundle TASP as scaffold. Connolly surface of E-selectin (green) is depicted with important residues for binding to PSGL-1 and for Ca^{2+} complexation in red. Two neighboring helices (blue) in the TASP molecule provide a sufficiently large area to accommodate the same residues (red) in the appropriate geometry of the native molecule.

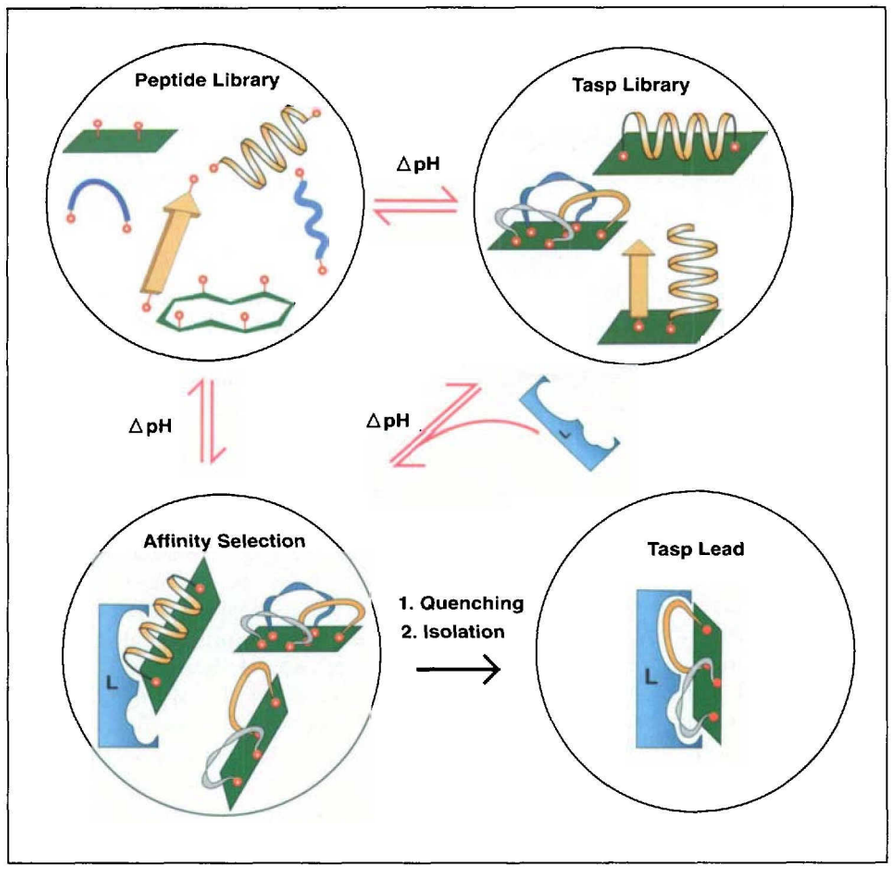

Fig. 7. Evolutionary principles and design [28]: In general, reversible self-association of a pool of peptide fragments (peptide library) leads to TASP assemblies of all possible combinations (TASP library). In the presence of a ligand (L), the statistical distribution of the members of the TASP library is driven towards selection of the highest affinity-binding motif. The dynamic process of random assembly, disassembly and re-assembly is no longer arbitrary but governed by the recognition step and results ultimately, in the affinity-directed enrichment of the most potent candidate. Subsequent quenching and isolation allows for the biostructural investigation of such lead compounds.

[1] M.J. Sipple, *Structure* **1999**, *7*, R81.
[2] K.W. Plaxco, D.S. Riddle, V. Gantcharova, D. Bake, *Curr. Opin. Struct. Biol.* **1998**, *8*, 80.
[3] S. Olofsson, G. Johansson, L. Baltzer, *J. Chem. Soc., Perkin. Trans.* **1995**, 2, 2047.
[4] M.D. Struthers, R.P. Cheng, B. Imperiali, *Science* **1996**, *271*, 342.
[5] M.A. Starovasnik, A.C. Braisted, J.A. Wells, *Proc. Natl. Acad. Sci.* **1997**, *94*, 10080.
[6] A. Pessi, *Nature* **1993**, *362*, 367.
[7] M. Mutter, G. Tuchscherer, *Makromol. Chem. Rapid Comm.* **1988**, *9*, 437.
[8] M. Mutter, G. Tuchscherer, *Cell. Mol. Life Sci.* **1997**, *53*, 851.
[9] H.K. Rau, N. DeJonge, W. Haehnel, *Proc. Natl. Acad. Sci.*, USA **1998**, *95*, 11526.
[10] K.S. Akerfeldt, J.D. Lear, Z.R. Wasserman, L.A. Chung, W.F. DeGrado, *Acc. Chem. Res.* **1993**, *26*, 191.
[11] M. Mathieu, C. Lehmann, G. Tuchscherer, *Angew. Chem. Int. Ed.* **1998**, *37*, 2990.
[12] M. Mutter, P. Dumy, P. Garrouste, C. Lehmann, M. Mathieu, C. Peggion, S. Peluso, A. Razaname, G. Tuchscherer, *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1482.
[13] M. Mutter, G. Tuchscherer, C. Miller, K.H. Altmann, R. Carey, D. Wyss, A. M. Labhardt, J. Rivier, *J. Am. Chem. Soc.* **1992**, *114*, 1436.
[14] P. Dumy, I.M. Eggleston, S. Cervigni, U. Sila, X. Sun, M. Mutter, *Tetrahedron Lett.* **1995**, *36*, 1255.
[15] P. Dumy, I.M. Eggleston, S. Nicula, M. Mutter, *Biopolymers* **1996**, *39*, 297.
[16] S. Peluso, T. Rückle, C. Lehmann, M. Mutter, C. Peggion, M. Crisma, *Nature Struct. Biol.*, submitted.
[17] G. Tuchscherer, L. Scheibler, P. Dumy, M. Mutter, *Biopolymers (Pept. Sci.)* **1998**, *47*, 63.
[18] D. Grell, J. Fernandez, S. Giraud, O. Spertini, G. Tuchscherer, in preparation.
[19] D. Grell, J. Richardson, M. Mutter, *Nature Struct. Biol.*, submitted.
[20] B. Rohwedder, R. Pillai, M. Mutter, G. Tuchscherer, *Angew. Chem.*, submitted.
[21] P.-A. Nygren, M. Uhlén, *Curr. Opin. Struct. Biol.* **1997**, *7*, 463.
[22] C. Vita, C. Roumestand, F. Toma, A. Menez, *Proc. Natl. Acad. Sci., USA*, **1995**, *92*, 6404.
[23] G. Lee, W. Chan, M.R. Hurle, R.L. DesJarlais, F. Watson, G.M. Sathe, R. Wetzel, *Protein Engineering*, **1993**, *6*, 745.
[24] E. Kellenberger, G. Mer, C. Kellenberger, G. Marguerie, J.F. Lefevre, *Europ. J. Biochem.* **1999**, *260*, 810.
[25] C.C. Arico-Muendel, A. Patera, T.C. Pochapsky, M. Kuti, A.J. Wolfson, *Protein Engin.* **1999**, *12*, 189.
[26] A. Tramontano, E. Bianchi, S. Venturini, F. Martin, A. Pessi, *J. Mol. Recognition* **1994**, *7*, 9.
[27] S. Peluso, P. Dumy, C. Nkubana, Y. Yokokawa, M. Mutter, *J. Org. Chem.* **1999**, *64*, 7114.
[28] L. Patiny, D. Banfi, M. Mutter, in preparation.
[29] M. Mutter, *Trends Biochem. Sci.* **1988**, *13*, 261.