Pure & Appl. Chem., Vol. 68, No. 11, pp. 2153–2162, 1996.
Printed in Great Britain.

Template assisted protein *de novo* design

G. Tuchscherer and M. Mutter

Institute of Organic Chemistry, University of Lausanne, BCH-Dorigny, CH-1015 Lausanne, Switzerland

Abstract: The ultimate goal in protein de novo design is the construction of artificial proteins exhibiting tailor-made structural and functional properties. To create native-like macromolecules in copying nature's way has proven to be difficult because the mechanism of folding in its complexity has yet to be unravelled. In the present review we describe a conceptually different approach in protein design, the concept of template assembled synthetic proteins (TASP), in order to bypass the folding problem. Topological templates as "built-in" devices for the induction of well-defined folding topologies and progress in synthetic strategies, e.g. chemoselective ligation methods and the use of pseudo-prolines to enhance peptide solvation open the way for the design of more complex, and hence functional TASP molecules.

INTRODUCTION

Constructing novel proteins with tailor made properties similar to natural proteins is the ultimate and one of the most challenging goals in biomimetic chemistry.
In the absence of a detailed knowledge of the folding mechanism of natural proteins, a general design strategy of polypeptide sequences with a high propensity to fold in a predetermined three-dimensional structure appears to be still out of reach. The major obstacle in the construction of artificial proteins rests in the complexity of the folding pathway as well as in the limited diversity of structural motifs in natural proteins, e.g. βαβ folding units or four-helix bundle arrangements. Among the surprisingly small number of recurring secondary structural motifs (α-helix, β-sheet, β-turn) (1-3) the same structural type is adopted by many different sequences, and the formation of small globular units is not confined to a specific amino acid sequence ("degeneracy of the folding code").
There are currently two basic hypotheses on the problem of protein folding mechanisms: One is based on the assumption that folding is initiated by the formation of fluctuating elements of local secondary structure along the unfolded polypeptide chain (4,5); subsequent interactions between these ordered regions and simultaneous rearrangements apparently lead to the formation of a compact globular structure. Alternatively, it is postulated that the unfolded polypeptide chain first collapses into a fairly compact "molten globule" state followed by the formation of specific secondary structures (6). It has been pointed out that these hypotheses are not mutually exclusive and could be reconciled to a unified theory on protein folding (7). It is of utmost interest to derive a general strategy for the "de novo" design of proteins from first principles that exhibit some of the characteristic properties of natural proteins.

THE TEMPLATE CONCEPT

A common feature in all strategies for protein design is the assembly of medium-sized peptide blocks adopting amphiphilic secondary structures in solution to a more complex but well-defined folding topology which in turn is necessary for biological functions. Consequently, a detailed knowledge of secondary structure formation and stability is essential. Numerous studies using homo-oligopeptides (8) or host-guest techniques (9), C^{α}-alkylated residues (10,11) or the side chains of trifunctional

amino acids for metal complexation (12) contributed to our present knowledge of the critical chain length, solvent and sequence dependence of helix and β-sheet formation.
A most attractive approach with regard to their use in protein design is the induction and stabilization of short secondary structure blocks by synthetic devices (templates, Fig. 1).

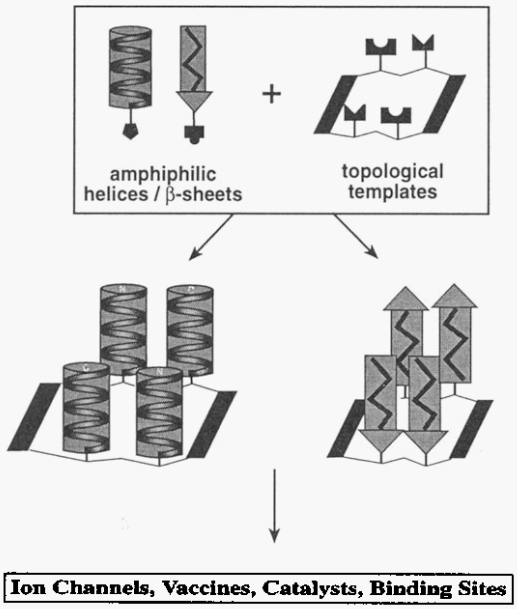
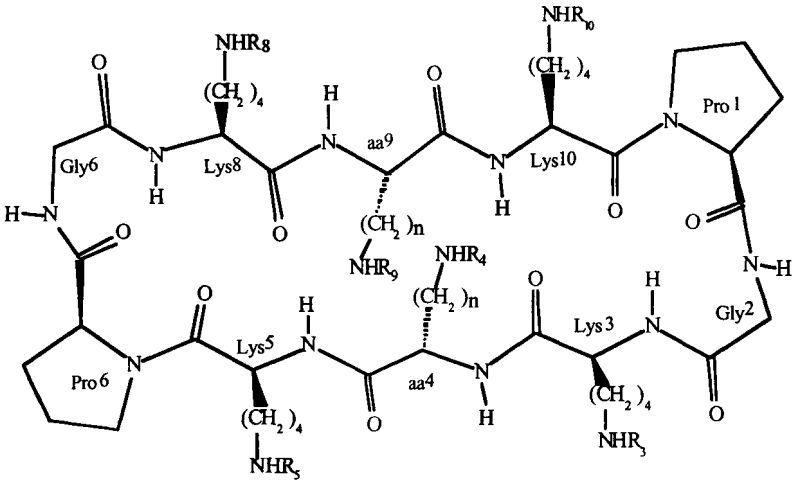


Fig. 1 Topological templates as folding devices have been used for the induction and stabilization of protein folds such as α-helical bundle or β-sheet TASP molecules for mimicking some properties of natural proteins.

The use of conformationally constrained molecules as templates by geometrically fixing the first amino acid in the proper orientation for helix or β-sheet initiation is one way to bypass the entropically unfavourable nucleation step in secondary structure formation. The amphiphilic character of such stabilized helical or β-sheet peptide blocks is the prerequisite for self-association in solution and the major driving force for formation of more complex packing topologies typical for proteins.
Thus, DeGrado designed a membrane channel forming α-helical peptide using only leucine (hydrophobic) and serine (hydrophilic) residues (13); possibly the most consequent application of this "amphiphilic principle" represents the design of polypeptide sequences with potential for 4α-helix bundle formation by using a binary code (14) as a general design strategy.
Since the complex folding mechanism has yet to be unravelled, we proposed a conceptually different approach in protein de novo design to bypass the folding problem: the template assembled synthetic proteins (TASP) concept (15) (Fig. 1). As a key element, a topological template serves as a "built-in" device to induce and reinforce intramolecular interaction of the covalently attached amphipathic peptide blocks thus leading to well-defined packing topologies such as α-helical bundle or β-sheet TASP molecules.
Topological templates may be generally characterized as synthetic devices that orient functional groups or structural units in well-defined spatial arrangements (16). Typically, template molecules represent structural motifs such as constrained peptides, cyclodextrines or polycyclic systems disposing selectively addressable functional groups. As prototype template molecules in the TASP approach cyclic decapeptides derived from the antibiotic gramicidin S containing four lysines as attachment sites were used (Fig. 2). As a second generation of this type of templates, RAFT (regioselectively addressable functionalized templates) molecules exhibit selectively addressable sites due to orthogonal protection techniques or unique chemical reactivity. As depicted in Fig. 2 such a RAFT not only offers the possibility for differentiation of the four lysine side chains but also for binary differentiation on the opposite face of the template (17).
Various examples for template induced tertiary structure formation have been reported in recent literature, e.g. the use of the tetraphenyl porphyrine system as a template for the construction of a

designed hemeprotein (18). Another elegant approach to assemble helical bundles was followed by Ghadiri et al. using transition metals for the complexation of helices via N-terminal ligands. In a further step, a heterodinuclear three helix bundle metalloprotein was synthesized with increased thermodynamic stability (19).
More recently, rigid organic macrocyclic scaffolds with an enforced cavity have been used for the orientation of a 4-helix bundle protein; by varying the length of the spacer between peptide and template the influence of flexibility/rigidity of the scaffold on the stability can be investigated (20). As a common feature, template molecules limit the degrees of freedom of the attached α-helical or β-sheet peptides and thus, promote intramolecular self-association to the desired folding topology.



aa9=aa4=Ala **1** R_8=Dde; R_{10}=Boc; R_3=Aloc; R_5= Fmoc **2** R_8=R_3=Aloc; R_5=R_{10}=Boc
aa9=aa4=Lys **3** R_3=R_5=R_8=R_{10}=Boc; R_4=R_9=Aloc

Fig. 2 Prototype RAFT molecule for the construction of TASP with up to 4 different helices with aa9=aa4=Ala (1) or two- domain TASP molecules with aa9=aa4=Lys (see text).

SYNTHETIC ASPECTS IN PROTEIN DESIGN

Molecular biology has revolutionized the study of protein structure and function. Not only has the microbial production of enzymes and other proteins in useful amounts become routine in recent years but with systematic alteration of protein sequence by site-specific mutagenesis virtually any protein sequence is accessible. The recent development of new methods for ligating peptides (Fig. 3), however, promises to make chemical synthesis of large proteins an attractive alternative to biosynthesis, particularly for the construction of novel molecules containing non-natural amino acids or other structural modifications.
Two strategies have been applied for the synthesis of proteins: stepwise assembly from their constituent amino acid on a solid support (solid phase peptide synthesis, SPPS) (20), or convergent coupling of peptide segments in solution (fragment condensation) (21). Although SPPS has been optimized to the extent that proteins of about 100 amino acids in length can be synthesized, accumulation of side products over the many coupling steps render purification of the target product laborious and time consuming. Convergent strategies have the considerable advantage that synthesis and purification of peptide segments up to 30 residues in length is straightforward but they are limited by the poor solubility of fully protected peptide segments and the tendency of α-carboxy-activated peptides to racemize.

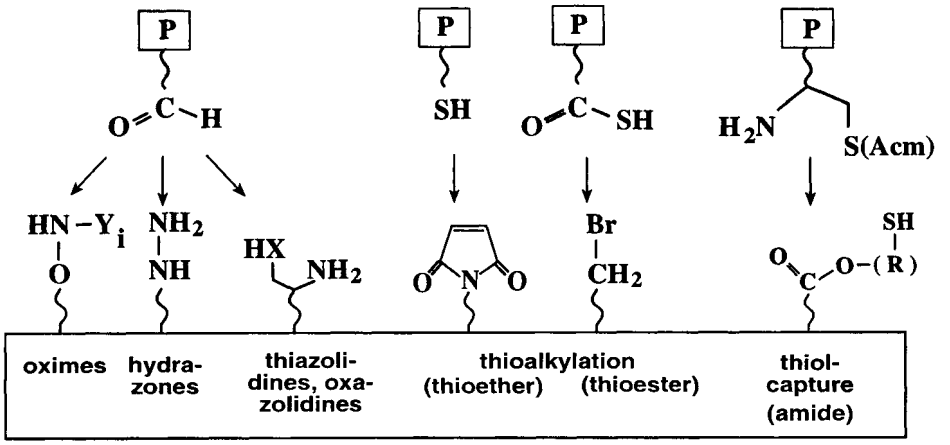


Fig. 3 Chemoselective ligation methods for protein design.

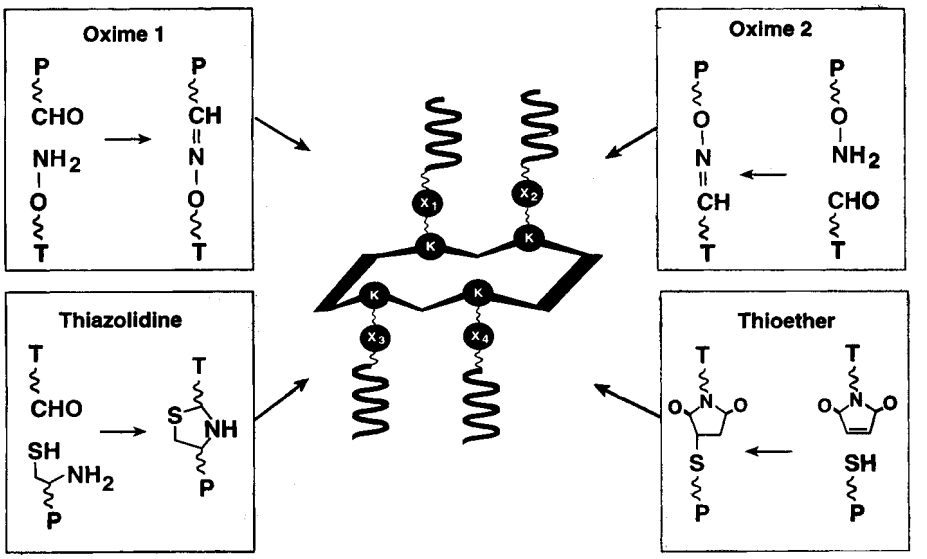


Fig. 4 Regioselectively Addressable Functionalized Template (RAFT) for the construction of TASP molecules by chemoselective ligation procedures (see text).

Many of these difficulties can be circumvented by using the recently introduced chemoselective ligation methods (22-25) which allow for the condensation of completely unprotected peptide fragments in aqueous medium (Fig. 3).

Chemoselective ligation methods appear to be particularly useful in the TASP design. With respect to the elaborated protection chemistry in peptide synthesis, cyclic peptides as topological templates with up to four orthogonal protection groups for the lysine side chains as attachment sites are accessible. Selective cleavage allows for appropriate selective functionalization leading to regioselectively addressable functionalized template (RAFT) (26) molecules as key compounds for the construction of

TASP molecules of higher complexity. As shown in Fig. 4, a prototype TASP with up to four different helices using different chemoselective ligation procedures can be synthesized.
Although the utility of chemical ligation strategies was demonstrated for the synthesis of larger and more complex peptides/proteins (27,28) and it's sole limitation seemed the availability of larger, appropriately modified peptide segments, one fundamental problem has not been taken into account yet. The peptide fragments, designed to be amphiphilic with a high propensity for secondary structure formation have a strong tendency in aqueous solution for self-association (Fig. 5).
Such high molecular weight aggregates are unfavorable in the ligation process which results e.g. in extended reaction times or incomplete reaction. To prevent self-association of unprotected, secondary-structure forming peptide sequences during chain assembly, pseudo-prolines (ψPro) (29) have been introduced as a powerful tool to modify temporarily the intrinsic properties of peptides that are responsible for aggregation and secondary structure formation (Fig. 6). Pseudo-prolines consist of serine- or threonine-derived oxazolidines and cysteine-derived thiazolidines and are obtained by reacting the free amino acid with aldehydes or ketones (Fig. 6a).

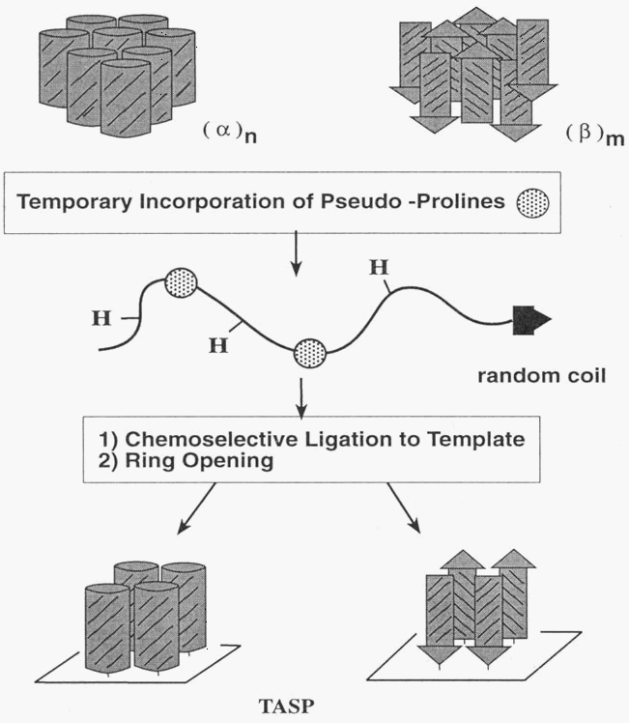


Fig. 5 Prevention of self-association during the ligation process by temporary incorporation of pseudo-prolines (ΨPro) into the peptide fragments .

Due to the presence of a cyclic system (fixed ϕ-angle) in addition to the preference for a cis-amide bond (30) with the preceding residue, the incorporation of a ψPro moiety results in a kink conformation (Fig. 6b), thus preventing peptide aggregation, self-association or β-structure formation.

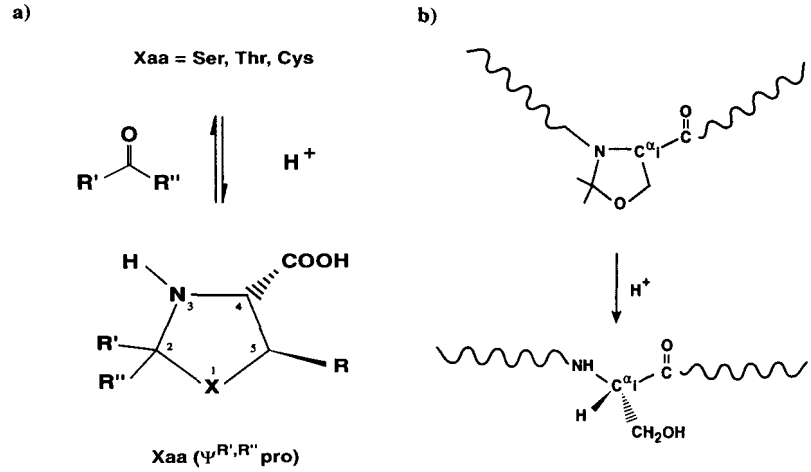


Fig. 6 a) Ser, Thr, Cys derived pseudo-prolines (ΨPro); b) effect of ΨPro upon peptide backbone.

Consequently, pseudo-prolines fulfil two functions simultaneously: they serve (i) as temporary protection for Ser, Thr and Cys and (ii) as solubilizing building blocks to increase solvation and coupling rates during peptide synthesis and in subsequent chain assembly. Finally, ring-opening of the ΨPro by strong acid results in the completely deprotected peptide, restoring the regular Ser, Thr, Cys side chains and the molecule can adopt the designed topology.

TEMPLATE ASSEMBLED SYNTHETIC PROTEINS

Several authors have reported on selective membrane channel forming TASP molecules using topological templates to define and orient membrane spanning helical segments. For example, DeGrado et al. (13) synthesized a four-helix bundle proton channel using a tetraphenylporphyrin system as template; Mutter et al. (31) designed and synthesized a template assembled channel-forming protein derived from the bee venom melittin. As a striking common feature, the membrane channel forming TASP molecules exhibit single channel conductance, ion selectivity and high thermodynamic stability.

Numerous activities in the template based design of functional molecules are being observed in the field of immunology. For example, Tam established the "multiple antigenic peptide" (MAP) (32) approach using branched oligo-lysines as template for the attachment of antigenic peptides. Here, the template acts merely as a support to increase immunogenicity rather than as a structure inducing device. Similarly, Rose used multiple oxime ligations to make a totally synthetic macromolecule of controlled structure and of molecular weight in the protein range (MW ~ 19 000 D) (27). Another chemoselective ligation method was used by Kent for the total synthesis of a 4-helix TASP molecule (24) where the fragments to be joined had complementary reactivity (e.g. a thiol and a bromo acetyl function).

As demonstrated for a number of model peptides the combination of orthogonal protection techniques for resin cleavage and side chain protection with chemoselective ligation reactions represents a versatile tool for constructing TASP molecules of high structural complexity. Furthermore, with the introduction of secondary structure disrupting, solubilizing techniques (25, 29-31), molecules which have been inaccessible so far have become accessible.

We have synthesized a TASP with an antiparallel 4α-helical bundle topology (36) for the evaluation of the overall stability of a 4α-helical bundle in a parallel versus antiparallel arrangement. The modified helical segments derived from hen eggwhite lysozyme have been covalently attached to a selectively addressable template by sequential ligation to the template via oxime bond formation. As followed by HPLC, the oxime bond formation proceeded under mild conditions to completion; interestingly, the ligation of the last two helical blocks did not significantly interfere with the prior

attached two helices. CD spectroscopic investigation revealed high helical content and a significant increase in secondary structure formation due to template-induced long-range interactions of the attached helices. Surprisingly, the parallel arrangement of the helices resulted in higher thermodynamic stability of the TASP compared to the antiparallel arrangement. Amphiphilic β-sheet forming peptides of the type **$(A\text{-}B)_n$; A= Ser,Thr; B= Ala, Leu, Val** and their assembly to a topological template represent an even more challenging example for the solubilizing effect of pseudo-prolines and the versatility of chemoselective ligations (37). As prototypes for this interesting class of peptides we have synthesized a series of peptides with the repetitive sequence **$(Xaa\text{-}Ser)_n$** (Fig. 7). Applying standard protection chemistry, we were not able to achieve quantitative reactions after passing the critical chain length for β-sheet formation (n ~ 3-5). However, insertion of two ΨPro-protected residues resulted in complete disruption of aggregates throughout the synthesis of the peptide which was reflected in high coupling yields.

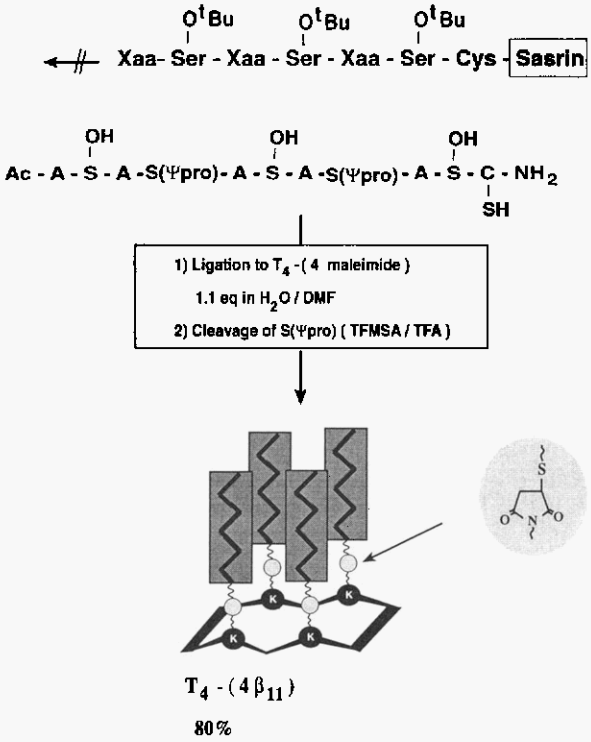


Fig. 7 Synthesis of a β-sheet TASP (see text).

The ΨPro containing peptide was readily soluble in water and various organic solvents and the IR spectrum indicates that it adopts a random coil conformation (amide I band at >1650cm^{-1}). After ring opening the peptide was insoluble in water and most organic solvents and the IR spectrum shows the typical bands of a β-sheet (e.g. amide I band at ~1620 cm^{-1}). The partially deprotected ΨPro-containing peptide was chemoselectively ligated to a maleimide functionalized topological template to give a 4β bundle TASP (Fig. 7). As followed by HPLC, the ligation reaction of the peptide to the template via thioether formation proceeded to completion within less than 4h, indicating the absence of aggregation or β-sheet formation during the ligation process.

Similar effects were observed in the synthesis of a membrane channel forming peptide as depicted in Fig. 8. Here, the ΨPro unit was inserted to induce a reversible kink in the helical peptide (38). Despite the hydrophobic character of the peptide, the presence of the ΨPro resulted in good solvation and the subsequent coupling steps proceeded to completion as followed by HPLC. The peptide was

cleaved from the Rink amide resin under mild acidic conditions, thus preserving the oxazolidine ring structure of the Thr($\Psi^{H,H}$pro) residue. As indicated by CD and ATR-IR studies, the ΨPro indeed distorts the helix to some extent (Fig. 8). The helical transmembrane peptides are subject to chemoselective ligation to topological templates to access membrane active TASP molecules with well-defined three-dimensional structures.

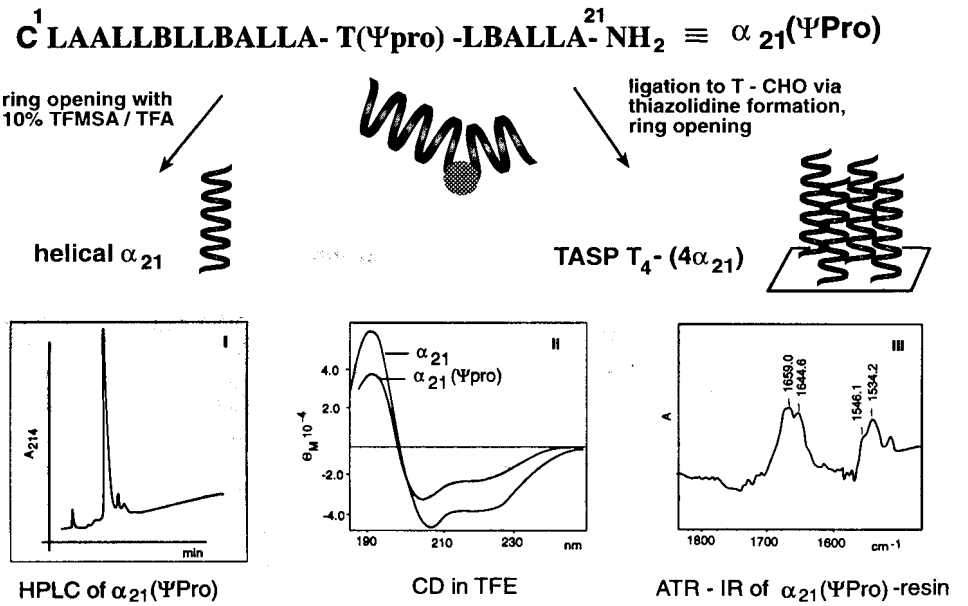


Fig. 8 Synthesis scheme for a membrane spanning hydrophobic peptide using Thr(Ψpro); I: HPLC of the crude, ΨPro containing peptide α_{21}; II: CD of α_{21} and α_{21}(ΨPro) in TFE; III: ATR-IR of the resin bound α_{21}(ΨPro).

SUMMMARY AND PERSPECTIVES

In combining (i) tailor-made orthogonal protection schemes for peptide assembly, (ii) solubilizing techniques to overcome synthetic difficulties such as unfavourable solubility or solvation properties of peptides and (iii) a variety of chemoselective ligation strategies, peptide chemists are no longer bound to copy nature's way of peptide synthesis and protein folding.
The use of unprotected, appropriately functionalized peptide fragments in chemoselective ligation procedures for covalent attachment to regioselectively addressable functionalized templates opens a way to macromolecules with novel chain architectures and functions (Fig. 9). With progress in the synthetic strategy and structural characterization of this new type of macromolecules we are one step closer to more realistically mimic some functional properties of natural proteins (e.g. vaccines, catalysts, ion channels), or even to combine different independent folding topologies which at the same time resemble different functionalities (Fig. 10a). For example, a two-domain TASP exhibiting an ion channel forming 4α-helix bundle and a receptor ligand on the opposite face of the template may be used as a biosensor system; similarly, nonpeptidic molecules (e.g. glycans) can be attached to modulate physicochemical and pharmacokinetical properties of peptide hormones. In a further extension, the self assembly based on noncovalent interactions of individual TASP molecules to supramolecular assemblies (Fig. 10b) can be used in our laboratory as a model system to investigate molecular recognition processes which are essential features in biological functions .

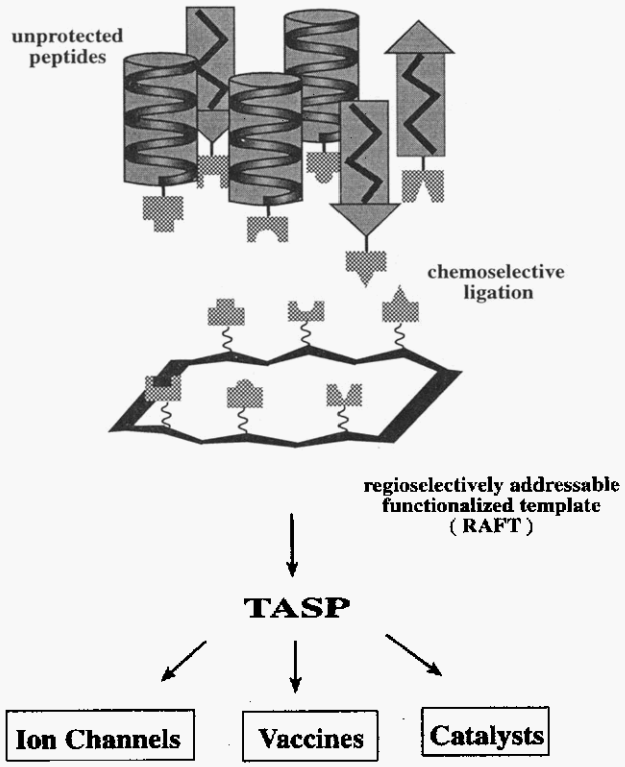


Fig. 9 The use of unprotected peptide segments and selectively addressable templates in chemoselective ligation procedures allows for the construction of more complex TASP molecules resembling ion channels, catalysts or vaccines.

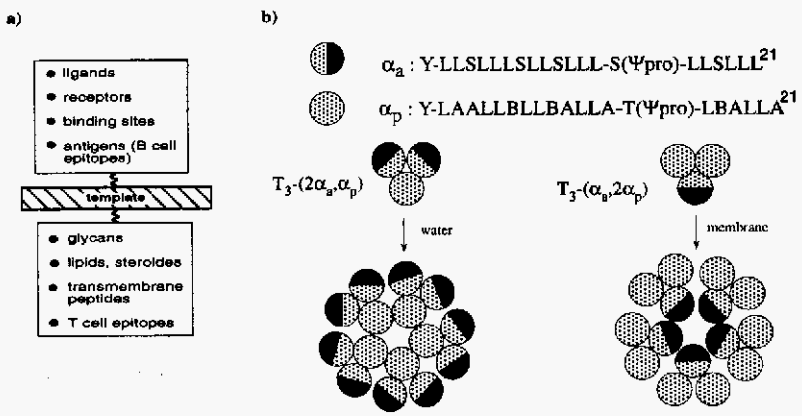


Fig. 10 a) Two-domain TASP molecules; b) Supramolecular assembly of individual TASP molecules.

In conclusion, the combination of existing knowledge and methodology with innovative concepts may eventually turn the ultimate goal of protein de novo design into reality - the construction of new macromolecules with predetermined three-dimensional structure and well-defined functionality.

ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation.

REFERENCES

1. Richardson, J.S., Adv. Prot. Chem. **34**, 167 (1981).
2. Rossmann. M.G. and Argos, P., Ann. Rev. Biochem. **50**, 555 (1981).
3. Presnell, S.R. and Cohen, F.E., Proc. Natl. Acad. Sci. USA **86**, 6592 (1989).
4. Anfinsen, C.B. and Scheraga, H.A., Adv. Prot. Chem. **29**, 205 (1975).
5. Montelione, G.T. and Scheraga, H.A., Acc. Chem. Res. **22**, 70 (1989).
6. Finkelstein, A.V. and Ptytsin, O.B., Prog. Biophys. Mol. Biol. **50**, 171 (1987).
7. Baldwin, R.L., Trends Biochem. Sci. **14**, 291 (1989).
8. Abd El Rahman, S., Anzinger, H. and Mutter, M., Biopolymers **19**, 173 (1980).
9. Toniolo, C, Bonora, G.M., Mutter, M. and Pillai, V.N.R., Makromol. Chem. **182**, 2007 (1981).
10. Toniolo, C., Crisma, M., Formaggio, F. and Caruicchioni, G., Biopolymers **33**, 1061 (1993).
11. Altmann, K.H., Altmann, E. and Mutter, M., Helv. Chim. Acta **75**, 1198 (1992).
12. Ghadiri, M.R. and Choi, C., J. Am. Chem. Soc. **112**, 1630 (1992).
13. Ackerfeldt, K.S., Kim, R.M., Camac, D.,Groves, J.T., Lear, J.D. and DeGrado, W.F., J. Am. Chem. Soc. **114**, 9656 (1992).
14. Hecht, M.H., Protein Gordon Research Conference 1993.
15. Mutter, M. and Tuchscherer, G., Makromol. Chem. Rapid Commun. **9** (6), 437 (1988).
16. Anderson, S., Anderson, H.L. and Sanders, J.K.M., Acc. Chem. Res. **26**, 469 (1993).
17. Dumy, P., Eggleston, I. M., Esposito, G., Nicula, S. and Mutter, M., Biopolymers, 1995, in press.
18. Sasaki, T. and Kaiser, E. T., J. Am. Chem. Soc **111**, 380 (1989).
19. Mezo, A. R. , Gibb, B. C. and Shermann, J. C., Proceed. 13th Am. Pept. Symposium, 1995, in press.
20. Stewart, J.M. and Young, J.D., Solid Phase Peptide Synthesis, Pierce Chemical, 1984.
21. Lloyd-Williams, P., Albericio, F. and Giralt, F., Int. J. Peptide Protein Res. **37**, 58 (1991).
22. Vilaseca, L.A., Rose, K., Werlen, R., Meunir, A., Offord, R.E., Nichols, C.L. and Scott, W.L., Bioconj. Chem. **4**, 515 (1993).
23. Liu, C.-F. and Tam, J.P., J. Am. Chem. Soc. **116**, 4149 (1994).
24. Dawson, P.E. and Kent, S.B.H., J. Am. Chem. Soc. **115**, 7263 (1993).
25. Kemp, D.S. and Carey, R.I., J. Org. Chem. **58**, 2216 (1993).
26. Dumy, P., Eggleston, I.M., Cervigni, S.E., Sila, U., Sun, X. and Mutter, M., Tetrahedron Lett. **36**, 1255 (1995).
27. Rose, K., J. Am. Chem. Soc. **116**, 30 (1994).
28. Dawson, P.E., Muir, T.W., Clark-Lewis, I. and Kent, S.B.H., Science **266**, 776 (1994).
29. Haack, T. and Mutter, M., Tetrahedron Lett. **33**, 1589 (1992).
30. Nefzi, A., Schenk, K. and Mutter, M., Protein and Peptide Lett. **1**, 66 (1994).
31. Pawlak, M., Meseth, U., Dhanapal, B., Mutter, M. and Vogel, H., Protein Sci. **3**, 1788 (1994).
32. Tam, J.P., Proc. Nat. Acad. Sci. USA 85, 5409 (1988).
33. Bayer, E., Henkel, B., Zhang, L., Clausen, N., Goldammer, C. and Panhaus, G., Proceed. 23rd Europ. Pept. Symposuim, p. 201, ESCOM, Leiden (1994).
34. Hyde, H., Johnson T., Owen, D., Quibell, M. and Sheppard, R.C., Int. J. Peptide Protein Res. **43**, 431 (1993).
35. Mutter, M., Oppliger, H. and Zier, A., Makromol. Chem. Rapid Commun. **13**, 151 (1992).
36. Nyanguile, O., Mutter, M. and Tuchscherer, G., Lett. in Peptide Science **1**, 9 (1994).
37. Nefzi, A., Sun, X. and Mutter, M., Tetrahedron Lett. **36**, 229 (1995).
38. Mutter, M., Nefzi, A., Sato, T., Sun, X., Wahl, F. and Wöhr, T., Peptide Research **8** (3), 145 (1995).