Chimia 54 (2000) 558–563

ISSN 0009–4293

The Supramolecular Chemistry of Proteins – Protein Epitope Mimetics Prepared Using Combinatorial Biomimetic Chemistry

Luyong Jiang, Michel Favre, Kerstin Moehle, and John A. Robinson*

Abstract: Some new approaches are described for the synthesis of protein epitope mimetics, in particular, mimics of β-hairpin structures found in antibodies, growth factors, and other proteins involved in macromolecular recognition in Nature. The mimetics are derived by transplanting the hairpin loop from the protein of interest to an organic template that functions to constrain the loop into a hairpin conformation. Furthermore, the methods are amenable to parallel synthesis methods, which allows the production of combinatorial libraries of β-hairpin mimetics. Such libraries are amenable to high-throughput purification, analysis and screening for biological activity, and so may prove to be a valuable source of ligands for chemical biology, and drug and vaccine research.

Keywords: Conformation · β-Hairpin · Peptides · Peptidomimetics · Secondary structure · Supramolecular chemistry

Introduction

The three-dimensional structures of molecular machines such as the proteasome [1][2], the molecular chaperone GroEL [3], the photosynthetic reaction centre [4], F1 ATPase [5], the cytochrome bc1 complex [6], or RNA polymerase-II [7], testify to the enormous scale and complexity of the 'supramolecular chemistry' achieved by Nature using the polypeptide scaffold. The modular construction of these molecular machines is based on the natural tendency of peptide chains to adopt units of regular secondary structure (the α-helix and β-sheet), which may be combined in countless different ways, some of which fold into stable supramolecular assemblies. The current explosive growth in protein sequence data from genome sequencing projects, and the increasing number of protein 3D structures, fuel interest in exploring new modular approaches to small molecule synthetic protein mimics. For example, technologies that allow surface epitopes to be mimicked in small synthetic molecules, perhaps made by combinatorial and parallel chemistry approaches, might provide access to interesting new classes of biologically active molecules, and be a source of ligands for biotechnology, and drug and vaccine discovery [8].

One approach to protein mimetics is to construct linear polypeptide chains, incorporating residues deemed important for biological activity. However, the conformation and dynamics of folded proteins are often intimately linked with biological activity, and these properties can be difficult to mimic adequately in linear flexible polypeptide chains in aqueous solution. In principle, this folding problem can be circumvented by developing new ways to constrain peptide chains into stable, well-defined conformations, such as α-helices and β-hairpins. β-Hairpins are frequently involved in protein-protein recognition, and in this context are particularly interesting targets for mimetic design, since two anti-parallel β-strands as well as a β- (or related) turn may be incorporated within one and the same molecule. Moreover, if the hairpin conformation is stable in aqueous solution, the relative positions of backbone and side-chain groups may be accurately mapped in three-dimensional space. This information can be decisive in understanding and developing structure-activity relationships. A further significant step can be envisaged, however, in which combinatorial libraries of hairpin mimetics are generated by parallel synthesis and evaluated for biological activity. A similar principle is used, of course, *in vivo* during antigen-driven selection, amplification and maturation of antibodies in the immune system.

Antibodies are well known for their ability to recognise and bind virtually any organic molecule, ion, solid surface or polymer. The immune system exploits the immunoglobulin fold for the generation of a large combinatorial library of proteins, each having a binding site com-

*Correspondence: Prof. J.A. Robinson
Institute of Organic Chemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Tel.: +41 1 635 42 42
Fax: +41 1 635 68 33
E-Mail: robinson@oci.unizh.ch
http://www.unizh.ch/oci/

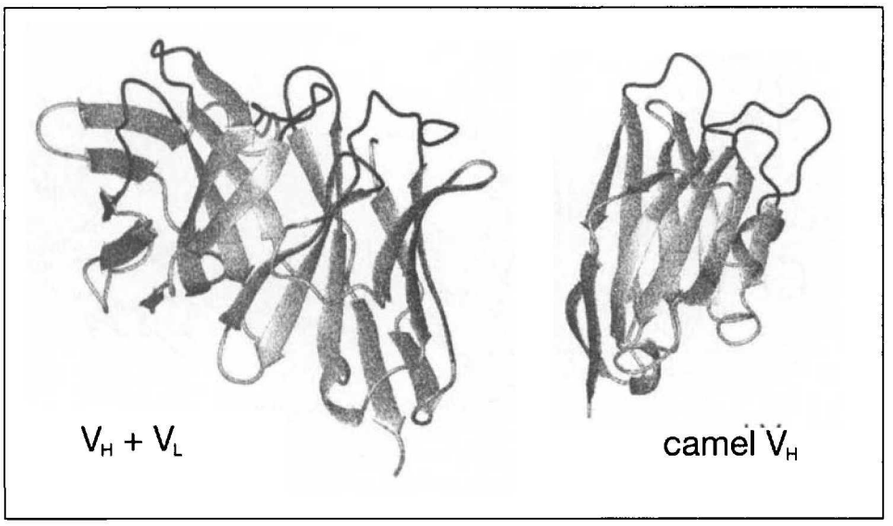


Fig. 1. Ribbon diagrams of the variable regions of the heavy and light chains of an IgG antibody (left), as well as the V-region of a camel heavy chain antibody (right).

posed of variable surface loops (Fig. 1), whose chemical and physical properties together confer a unique ability to bind antigens. Human and mouse antibodies have antigen binding sites constructed from six so-called hypervariable loops, or complementarity determining regions (CDRs), three in the heavy (H) and three in the light (L) chain. Despite the combinatorial diversity that can be incorporated in these loops, all CDRs except H3 appear to adopt only a limited number of allowed or canonical conformations [9]. Moreover, the CDRs-2 and -3 of the H- and L-chains adopt well-defined β-hairpin structures. The antibodies from camelids, on the other hand, contain only a H-chain (Fig. 1), and their antigen-binding sites are composed of only three CDR loops [10]. Yet these camelid antibodies are also able to bind a diverse array of antigens [11]. Of course, the considerable simplification in structure, compared to IgG antibodies, make camelid antibodies interesting targets for biomimetic and supramolecular chemistry.

We describe below some results of our recent studies of β-hairpin mimetics, which include loop structures derived from antibody CDRs [12], recognition loops in cytokine receptors [13] and growth factors such as platelet-derived growth factor [14], as well as epitopes found on proteins from the malaria parasite *Plasmodium falciparum* [15–17] and the human immunodeficiency virus (HIV). Mimetics of epitopes on viral and parasite proteins are of great interest in the design of novel synthetic vaccine candidates.

Nomenclature of β-Hairpins

A systematic classification of β-hairpins found in a collection of high-resolution protein crystal structures has been introduced [18][19], that takes into account the length of the loop region connecting the two antiparallel β-strands and the inferred hydrogen-bonding pattern between peptide amide groups flanking the turn (Fig. 2). For short loops, a restricted number of backbone conformations are observed at the tip of the hairpin, which include two-residue type-I, -I', and -II' β-turns [20]. The class 2:2 β-hairpin is found most frequently in protein crystal structures, and these hairpins most often contain type-I' and -II' β-turns. As the tip of the loop becomes longer (includes more residues), so the number of conformations found for each in the data-set increases. This may be accompanied by increased conformational flexibility of the loop in solution. Also, it is interesting to note that β-hairpins in globular proteins are usually not flat [21]. When viewed in the direction of their strands they are seen to twist in a right-handed direction. Furthermore, the twist of the type-I' and -II' turns is compatible with the twist observed in β-hairpins, which explains why the type-I' and -II' turns are favoured over type-I and -II turns (which have a mirror-image backbone conformation) at the tip of the loop [19].

Template-Stabilised β-Hairpin Mimetics

One approach to the synthesis of mimetics of β-hairpin loops in folded proteins is to transplant the hairpin from the protein onto an organic template that stabilises the hairpin secondary structure. The primary function of the template is to fix the bond vectors at each end of the hairpin in the correct relative positions.

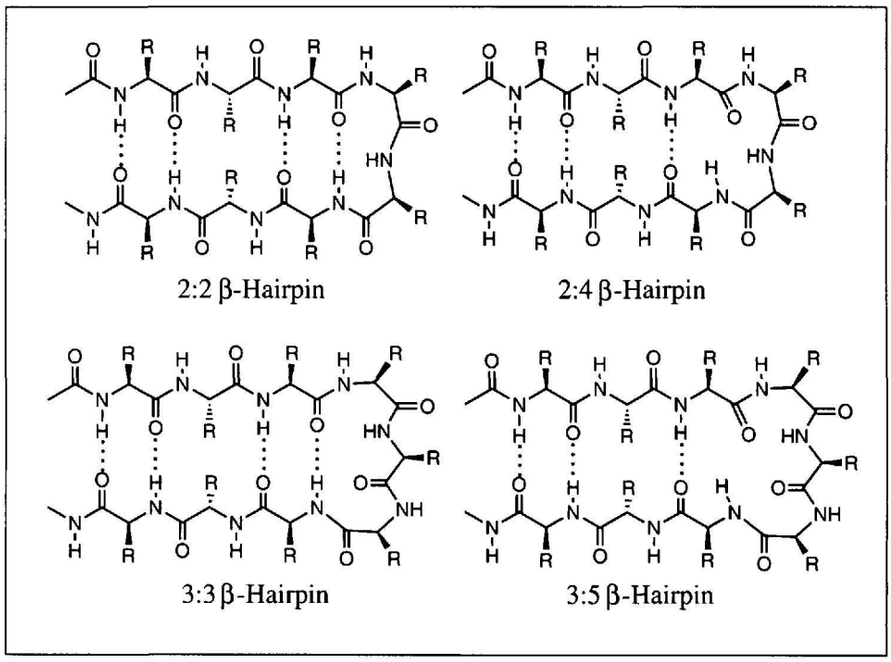


Fig. 2. The classification of β-hairpins depends on whether the end residues of the β-strands are linked by a single (2:4, 3:5, *etc.* or X:Y loops where Y=X+2) or double H-bonds (2:2, 3:3, *etc.* or X:Y loops where X=Y), see [18][19] for more details.

The rest of the loop may then adopt in the mimetic the same hairpin conformation seen in the intact protein. The approach is illustrated for an antibody CDR-mimetic in Fig. 3, where the dipeptide unit D-Pro-L-Pro is used as a template. It has been known for some time [22][23], that the D-Pro-L-Pro dipeptide adopts a quite rigid type-II' β-turn, which makes it ideal as a β-hairpin mimetic template [24].

Other templates we have developed to stabilise hairpin loop conformations are shown in Fig. 4. For example, the bicyclic diketopiperazine **1** has the required geometry and functional groups to fix the ends of a hairpin loop in the correct relative positions. Closely related is the diketopiperazine **2**, which also contains an additional carboxylic acid group for coupling the mimetic to other molecules [16][17]. Recently we also described the use of a D-Pro-L-Apro dipeptide template (**8**), where Apro is (2*S*,4*S*)-4-aminoproline, for the synthesis of a CDR mimetic [25]. The use of L-Apro also opens the possibility to link this mimetic to other molecules through the new amino group in the template.

The hairpin mimetics can be synthesised using solid-phase Fmoc peptide chemistry, by assembling first a linear peptide precursor, and subsequent macrocyclisation either on the solid-phase or in solution. A procedure used recently [25] for the synthesis of a CDR mimetic (**9**), with the D-Pro-L-Apro template, is shown in the Scheme. The hairpin-constraining template is incorporated near the middle of the peptide chain, so as to enforce a backbone conformation in the linear precursor that is favourable for cyclization. Typically, the macrocyclisation proceeds in close to quantitative yield.

The solution conformation of the hairpin mimetics can be elucidated by NMR and dynamic simulated annealing. For example, with the CDR mimetic **9**, the ^{1}H NMR spectrum recorded at 600 MHz in d_6-DMSO reveals a large spectral dispersion of the H-C(α) protons and also the NH resonances, indicative of a defined conformation. The ^3J (α,HN) values for residues in the β-strands are > 8.5 Hz, consistent with the presence of a β-structure. The three backbone NH groups of Leu^1, Tyr^3 and Val^8 have relatively slow H/D exchange rates, consistent with their involvement in H-bonding between amide groups across the hairpin, as predicted by the structure calculations and MD simulations.

The most striking evidence for a stable β-hairpin conformation, however, comes from NOESY spectra. These reveal a network of long-range NOE connectivities between residues far apart in the sequence, but spatially close together on opposite sides of the hairpin. This includes, for example, a strong NOE between the H-C(α) protons in Trp^2 and Trp^7 (distance restraint 2.4 Å). Using NOE-derived distance restraints, a family of low-energy structures can be calculated by simulated annealing. These structures (Fig. 5) fulfil almost all the distance restraints (see below), and possess a well-defined β-hairpin backbone conformation. An interesting feature of these structures is the close stacking of the indole groups of Trp^2 and Trp^7 on one side of the β-hairpin, as seen earlier in a related mimetic [12]. The interaction of these indole rings is also supported by CD spectra recorded in MeOH and H_2O:MeOH 9:1, which include a characteristic strong exciton couplet at 225 nm.

The mimetic, however, is not rigid. Rather the NMR structures represent an average about which the molecule fluctuates. In particular, not all the observed NOEs involving the Trp^2 and Trp^7 side chains can be satisfied by one average NMR structure. The Trp^7 HD1 proton, for example, shows NOEs to both Val^8 HN and to Trp^2 HA, which can best be

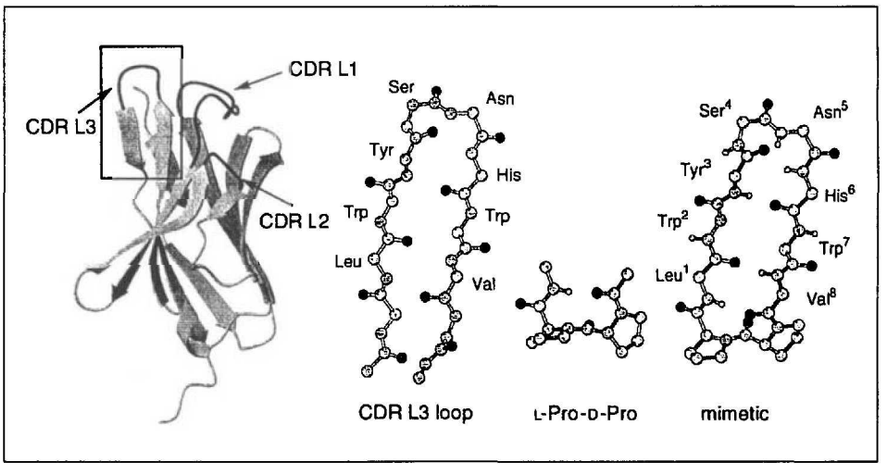


Fig. 3. The variable region of the light chain of the antibody is shown as a ribbon diagram (PDB structure file 1GIG). The L3 loop to be mimicked is shown in the centre. The residues Leu-Trp-Tyr-Ser-Asn-His-Trp-Val are transplanted from the protein to the template to give the mimetic [12].

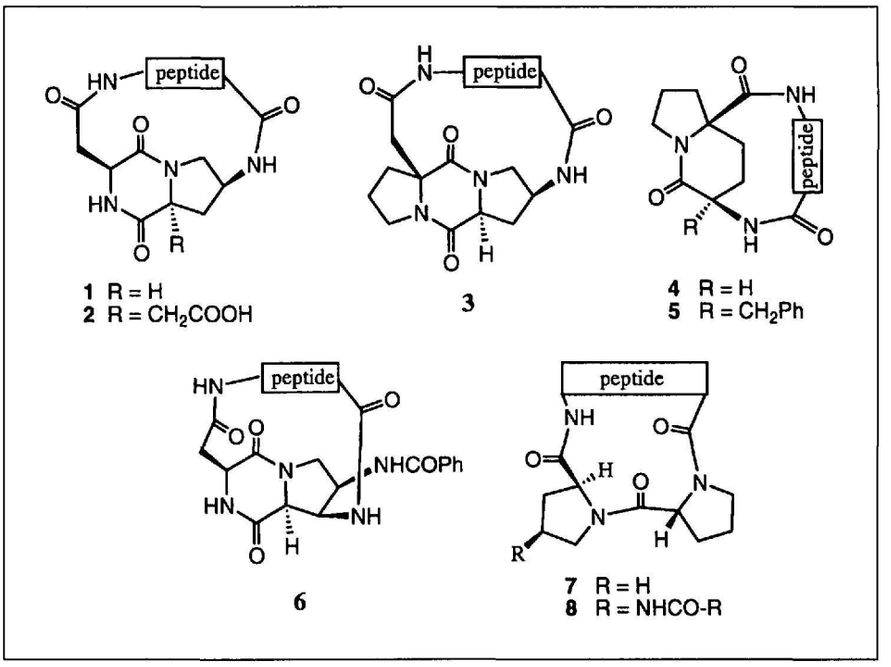
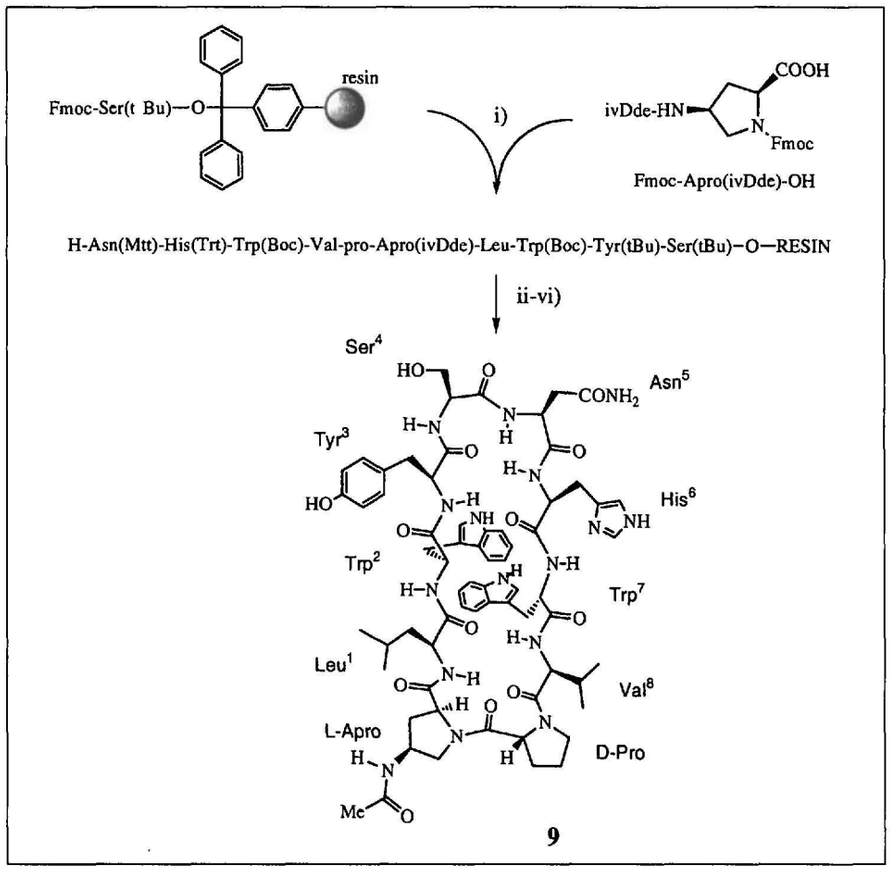


Fig. 4. Some templates used for the synthesis of loop mimetics (reviewed in [24]).



Scheme. Synthesis of **9**, a mimetic of the L3 CDR loop in antibody HC19 (see Fig. 3). i) Solid-phase peptide synthesis using amino acid and HBTU/HOBt (4 equiv.), iPr_2NEt, in DMF; and 20% piperidine in DMF for Fmoc removal; ii) AcOH:CH_2Cl_2:MeOH 5:4:1; iii) HATU, HOAt, iPr_2NEt, DMF; iv) NH_2NH_2 in DMF; v) Ac_2O, DMF; vi) TFA:iPr_3SiH:H_2O 95:2.5:2.5.

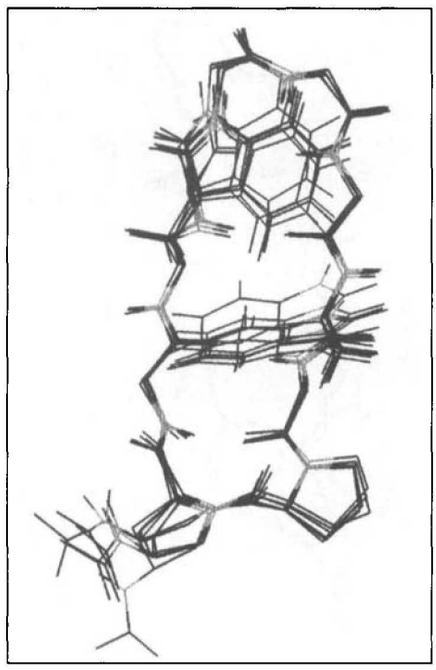


Fig. 5. Superimposition of a group of low-energy NMR structures of mimetic **9** determined by simulated annealing. N atoms in grey. The D-Pro-L-Apro template is at the bottom.

explained if the Trp^7 indole is undergoing a 180° flipping motion about the C(β)–(Cγ) bond which is fast on the NMR time-scale.

In order to explore how the structural properties inferred from NOE data might be influenced by motional averaging, molecular dynamics calculations were performed, both with and without distance restraints. These simulations show, that at least over several nanoseconds, the backbone of the mimetic remains in a well-defined β-hairpin conformation, with only small fluctuations of backbone ϕ and ψ angles, as illustrated in Fig. 6. Interestingly, the flipping of the Trp^7 indole role inferred from NOE data (*vide supra*) was observed in the MD simulations.

The goal of this work was to synthesise an accurate conformational mimetic of an antibody CDR loop [25]. A superimposition of a typical NMR solution structure of the mimetic, and the loop taken from the crystal structure of the antibody Fab fragment [26][27] is shown in Fig. 7. This demonstrates that the mimetic adopts essentially the same hairpin loop conformation as seen in the intact antibody, and augers well for the design of other biologically interesting protein epitope mimetics. For example, using similar methods we have shown recently [14] that accurate hairpin loop mimetics can also be made of a protruding loop in the human platelet-derived growth factor B (PDGF-B) (see below and Fig. 8). Currently, efforts are underway to determine whether these mimetics are useful in the discovery of novel PDGF receptor antagonists.

Combinatorial Biomimetic Chemistry

With methods in hand to synthesise conformationally well-defined template-bound β-hairpin mimetics, it becomes interesting to prepare combinatorial libraries by parallel synthesis. From a certain perspective, this would bring us one step closer to mimicking an aspect of antibody production in the humoral immune system, which also involves the production of a library of hairpin loop structures, but presented on an immunoglobulin scaffold (Fig. 1).

Using the methods outlined above, a technology has now been established for the parallel solid-phase synthesis of β-hairpin mimetic libraries [14]. Since the emphasis is on mimicry of naturally occurring structures, the designation 'combinatorial biomimetic chemistry' seems appropriate.

For example, a mimetic was first produced of a hairpin loop in human PDGF-B. PDGF belongs to the cystine-knot family of growth factors [28], which includes also transforming growth factor-β (TGF-β), nerve growth factor (NGF) and vascular endothelial growth factor (VEGF). The extended and protruding loop-III in PDGF was transplanted to a D-Pro-L-Pro template, either as an 8-residue loop, or as a 12-residue loop, to afford the mimetics **10** and **11**, which were synthesised by methods described above. The structures of both mimetics in aqueous solution were shown by NMR and MD to be very similar to that seen in the crystal structure of PDGF-B (for **11** see Fig. 8).

Using the hairpin structure of **10** as a scaffold, a small 24-member library was then designed in which the four residues at the tip of the loop Val-Arg-Lys-Lys (VRKK) were held constant, and the residues at positions 1, 2, 7 and 8 were varied (Fig. 9). Of course, the diversity that can potentially be incorporated into even a

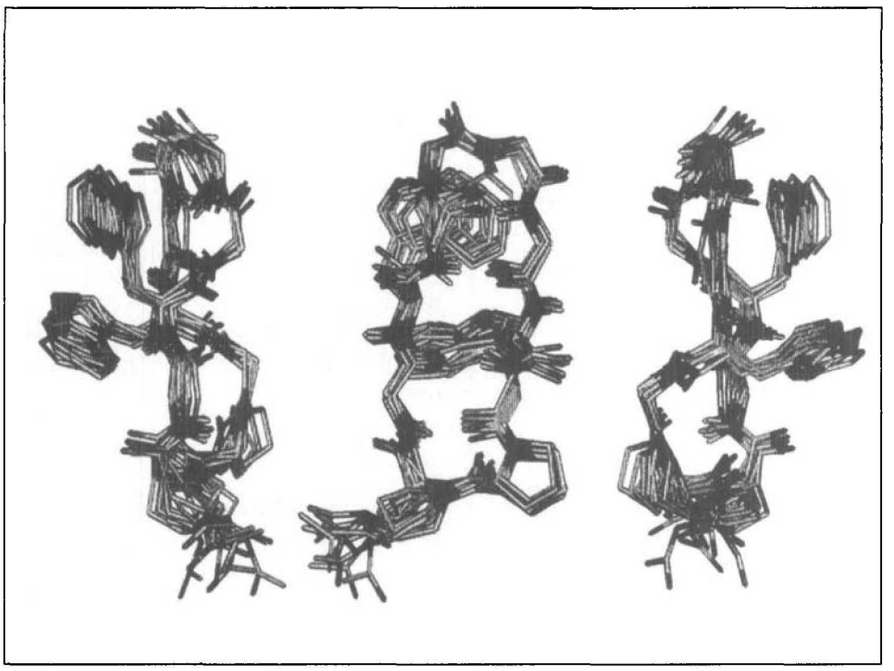


Fig. 6. *Centre*, superimposition of ten frames, one taken every 600 ps, from an MD simulation (in explicit DMSO solvent) of the mimetic **9**; *left*, rotated 90° to the left; *right*, rotated 90° to the right. Only the side-chains of Trp^2 and Trp^7 are shown. The D-Pro-L-Apro template is at the bottom.

relatively small 8-membered hairpin loop mimetic is enormous. Even using only the 20 proteinogenic amino acids, over 8^{20} different hairpin loops are possible. This number can be rapidly expanded by including non-proteinogenic amino acids, and there is also scope to introduce new functional groups into the template (*vide supra*). An important question, however, is whether the well-defined hairpin structure seen in the starting mimetic **10** is maintained when the sequence is varied. The answer to this question seems to be yes, with some qualifications.

The parallel synthesis of the 24 hairpin loop mimetics was performed using a manual 24-reactor workstation, essentially using methods described above. The purity of the library was assayed by reverse-phase HPLC, which showed for > 95% of the members a single major product of *ca.* 60–90% purity. These major products were purified and characterised by MS, CD and NMR spectroscopy. The conformation of the mimetics in

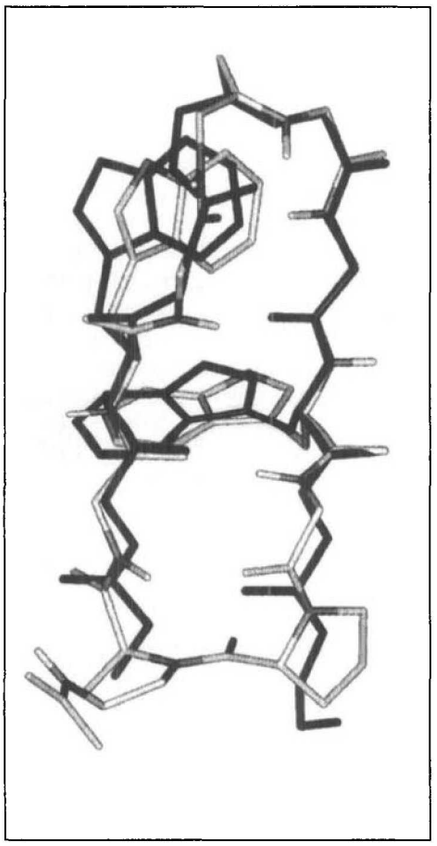


Fig. 7. Superimposition of the solution structure of the mimetic **9** (grey) determined by NMR, and the crystal structure (PDB file 1GIG) of the antibody CDR L3 loop (black) (Fig. 3).

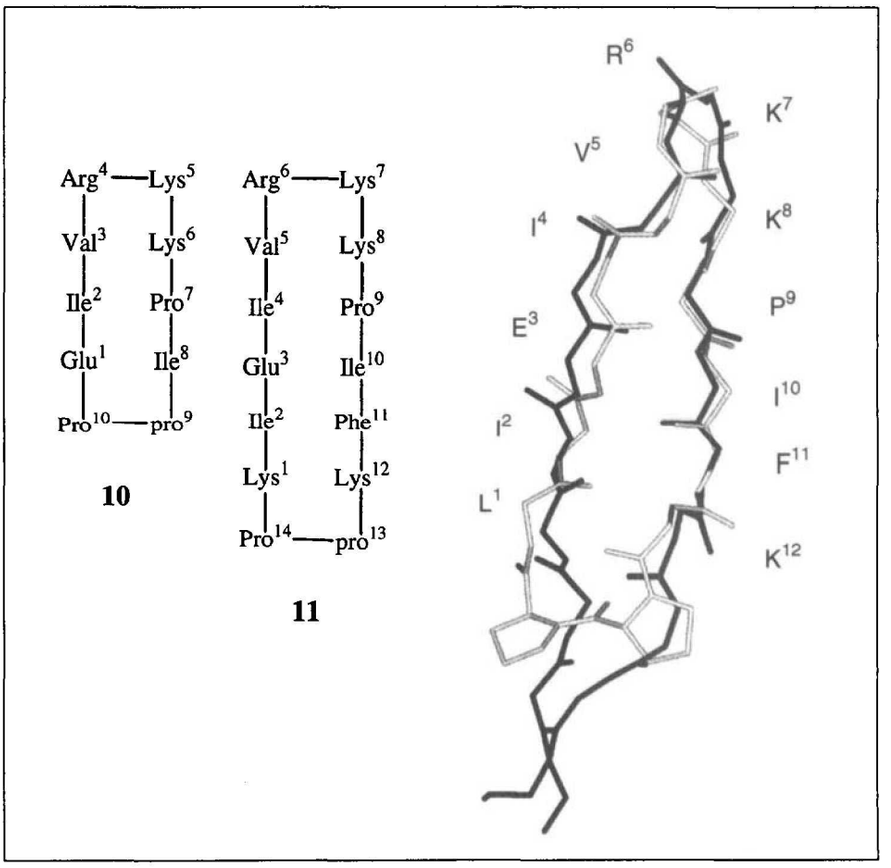


Fig. 8. *Right*, superimposition of the solution structure of the mimetic **11** (grey) determined by NMR, and the crystal structure of the protruding loop-III from the crystal structure of PDGF-BB [14].

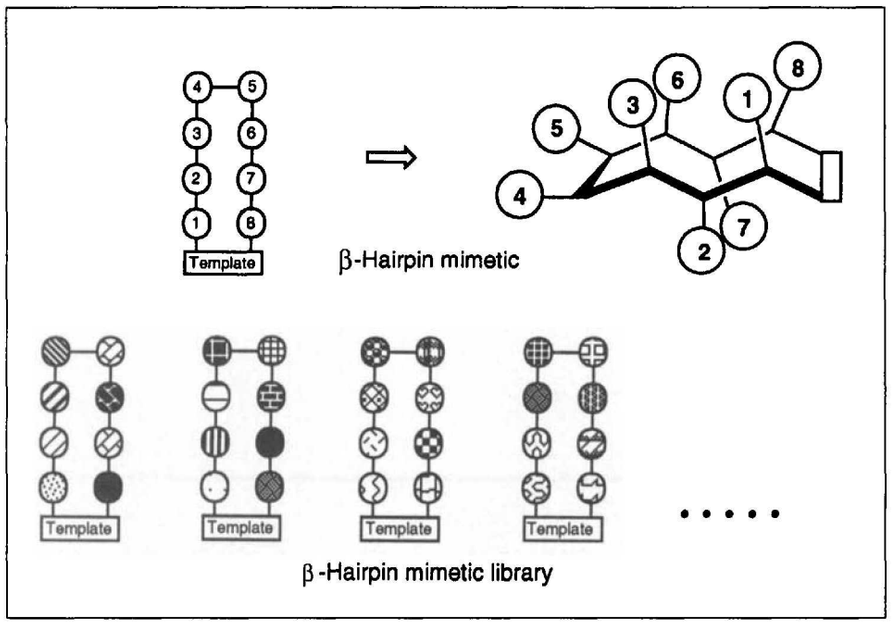


Fig. 9. Starting with an 8-residue β-hairpin mimetic (**10**), a combinatorial library can be assembled (bottom). The side chains of residues-1, -3, -6 and -8 cluster on one side of the hairpin and those in residues -2 and -7 on the other side. The β-hairpin conformation should be maintained by the template.

aqueous solution was assessed by analysing 2D TOCSY, DQ-COSY and NOESY spectra. The NOESY spectra, in particular, showed for most of the library, long-range NOEs characteristic of the expected hairpin conformation [14].

The results of this study are encouraging from the perspective of preparing much larger libraries of protein epitope mimetics, for drug and vaccine research. With use of robotic workstations, high-throughput HPLC-MS systems for purification and analysis, and high-throughput bioassays, the quality of the libraries should be high enough to allow rapid and reliable determination of structure-activity relationships in biologically active mimetics. As a first step in evaluating this strategy, new libraries of PDGF mimetics are currently being prepared, and will be evaluated as potential PDGF receptor antagonists. Finally, the perspective also exists of using the hairpin mimetics as building blocks for the assembly of more complex supramolecular assemblies. And this will be a topic of future work.

Acknowledgements

The production of β-hairpin mimetic libraries by combinatorial biomimetic chemistry is being developed in collaboration with POLYPHOR AG. The authors thank Dr. D. Obrecht for insightful discussions.

The authors thank the Swiss National Science Foundation, and the Commission for Technology and Innovation for supporting this work.

Received: August 16, 2000

[1] M. Groll, L.J. Ditzel, D. Stock, M. Bochtler, H.D. Bartunik, R. Huber, *Nature* **1997**, *386*, 463–471.
[2] J. Löwe, D. Stock, B. Jap, P. Zwickl, W. Baumeister, R. Huber, *Science* **1995**, *268*, 533–539.
[3] K. Braig, Z. Otwinowski, R. Hegde, D.C. Boisvert, A. Joachimiak, A.L. Horwich, P.B. Sigler, *Nature* **1994**, *371*, 578–586.
[4] J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, *Nature* **1985**, *318*, 618–624.
[5] J.P. Abrahams, A.G.W. Leslie, R. Lutter, J.E. Walker, *Nature* **1994**, *370*, 621–628.
[6] S. Iwata, J.W. Lee, K. Okada, J.K. Lee, M. Iwata, B. Rasmussen, T.A. Link, S. Ramaswamy, B.K. Jap, *Science* **1998**, *281*, 64–71.
[7] P. Cramer, D.A. Bushell, J. Fu, A.L. Gnatt, B. Maier-Davis, N.E. Thompson, R.R. Burgess, A.M. Edwards, P.R. David, R.D. Kornberg, *Science* **2000**, *288*, 640–649.
[8] D. Obrecht, M. Altorfer, J.A. Robinson, *Adv. Med. Chem.* **1999**, *4*, 1–68.
[9] B. AlLazikani, A.M. Lesk, C. Chothia, *J. Mol. Biol.* **1997**, *273*, 927–948.
[10] A. Desmyter, T.R. Transue, M.A. Ghahroudi, M.H.D. Thi, F. Poortmans, R. Hamers, S. Muyldermans, L. Wyns, *Nat. Struct. Biol.* **1996**, *3*, 803–811.
[11] V.K. Nguyen, R. Hamers, L. Wyns, S. Muyldermans, *EMBO J.* **2000**, *19*, 921–930.
[12] M. Favre, K. Moehle, L. Jiang, B. Pfeiffer, J.A. Robinson, *J. Am. Chem. Soc.* **1999**, *121*, 2679–2685.
[13] J. Späth, F. Stuart, L. Jiang, J.A. Robinson, *Helv. Chim. Acta.* **1998**, *81*, 1726–1738.
[14] L. Jiang, K. Moehle, B. Dhanapal, D. Obrecht, J.A. Robinson, *Helv. Chim. Acta.* **2000**, *83*, in press.
[15] M.E. Pfeifer, J.A. Robinson, *Chem. Commun.* **1998**, 1977–1978.
[16] C. Bisang, L. Jiang, E. Freund, F. Emery, C. Bauch, H. Matile, G. Pluschke, J.A. Robinson, *J. Am. Chem. Soc.* **1998**, *120*, 7439–7449.
[17] C. Bisang, C. Weber, J.A. Robinson, *Helv. Chim. Acta.* **1996**, *79*, 1825–1842.
[18] B.L. Sibanda, J.L. Thornton, *Methods Enzymol.* **1991**, *202*, 59–82.
[19] B.L. Sibanda, T.L. Blundell, J.M. Thornton, *J. Mol. Biol.* **1989**, *206*, 759–777.
[20] E.G. Hutchinson, J.M. Thornton, *Protein Sci.* **1994**, *3*, 2207–2216.
[21] C. Chothia, *J. Mol. Biol.* **1983**, *163*, 107–117.
[22] C.M. Nair, M. Vijayan, Y.V. Venkatachalapathi, P. Balaram, *J. Chem. Soc., Chem. Comm.* **1979**, 1183–1184.
[23] C.M.K. Nair, M. Vijayan, *J. Chem. Soc., Perkin II* **1980**, 1800–1804.
[24] J.A. Robinson, *Synlett.* **2000**, 429–441.
[25] M. Favre, K. Möhle, J.A. Robinson, in '4th International Electronic Conference on Synthetic Organic Chemistry', **2000**, *http://www.unibas.ch/mdpi/ecsoc-4.htm*
[26] T. Bizebard, R. Daniels, R. Kahn, B. Golinelli-Pimpaneau, J.J. Skehel, M. Knossow, *Acta. Cryst.* **1994**, *D50*, 768–777.
[27] T. Bizebard, B. Gigant, P. Rigolet, B. Rasmussen, O. Diat, P. Bösecke, S.A. Wharton, J.J. Skehel, M. Knossow, *Nature* **1995**, *376*, 92–94.
[28] P.D. Sun, D.R. Davies, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 269–291.