Pergamon

Tetrahedron 56 (2000) 9451–9460

TETRAHEDRON

Engineering Novel Bioactive Mini-Proteins on Natural Scaffolds

Loïc Martin,[a] Philipe Barthe,[b] Olivier Combes,[a] Christian Roumestand[b] and Claudio Vita[a,*]

[a]*CEA, Département d'Ingénierie et d'Etudes des Protéines, C.E. Saclay, 91190 Gif-sur-Yvette Cedex, France*
[b]*Centre de Biologie Structurale, CNRS UMR 9955-INSERM U414 Faculté de Pharmacie, 34060 Montpellier, France*

Received 10 March 2000; accepted 18 August 2000

Abstract—Novel bioactive mini-proteins have been engineered by a rational approach, implying the structural and functional reproduction of protein binding sites on stable natural protein structures, functioning as scaffolds. Such molecules, possessing a well-defined three-dimensional structure and a specific biological activity, represent new tools in biology, biotechnology, medical sciences, and also precious intermediates useful in drug design, facilitating the conversion of a protein functional epitope into a classical pharmaceutical.

Introduction

Very often nature uses common structural motifs and protein domains in different biological contexts, to express a large variety of specific functions. The immunoglobulin fold, for example, is one of the most widespread fold found in nature, since it is present in antibodies and in many enzymes and receptors.[1] It can support a large sequence variety, such that two components of the immunoglobulin superfamily can have as low as 10% sequence homology. This finding suggests that new sequences can be incorporated into this structural motif and new binding specificity artificially engineered. As a matter of fact, the immunoglobulin fold has been one of the first motif to be equipped with new antigen specificity by exchanging the sequence of its complementarity-determining region (CDR) loops with that from a different monoclonal antibody.[2] In more recent applications, new antibody specificity has been engineered by displaying the Fab portion at the surface of bacteriophages, followed by random mutagenesis of its CDR loop sequences and functional selection (biopanning) against new antigens.[3]

In the last decade or so, similar approaches, leading to incorporation of new binding functions into predetermined structural scaffolds, have been also applied to several small-size common and stable structural motifs, like the α/β scorpion toxin fold, protease inhibitors, the module of leucine zipper, EGF and zinc finger, the B and Z domain of protein A, knottins and tendamistat.[4,5] De novo designed templates, representing simplified versions of natural structures with natural (Minibody,[6] helical coiled-coil[7]) or artificial (TASP[8]) connectivities, have also been equipped with new binding specificity. In all these applications, new functions have been introduced either by the transfer of functional protein epitopes to specific and structurally compatible regions of the host structures,[9] or by randomization of particular scaffold regions, displayed at the surface of phages, followed by a functional selection against a selected protein target.[4] In a third approach, structural information was linked to a combinatorial approach to reduce a binding domain to a minimum size and to optimize structural stability and function, producing novel bioactive minimized protein folds.[10] The concept of using stable structural motifs as scaffolds in order to reproduce protein functional epitopes or stabilize bioactive conformations is nowadays recognized as one of the most successful approaches in protein engineering.[5,9] Scaffolds of limited size (with fewer than 70 amino acids) are particularly useful, since the newly engineered mini-proteins may be obtained by chemical synthesis, thus allowing any desired chemical modification. Thus, non-natural amino acids, fluorescent and affinity tags, NMR or X-ray crystallography labels or other chemical moieties can be incorporated within specific positions not disturbing function. These possibilities allow transformation of the engineered mini-proteins into new tools that can be extremely useful in biology, biotechnology and medical sciences. In addition, novel bio-active mini-proteins, representing minimized versions of much larger and complex natural proteins, may be useful not only in fundamental science, but particularly in drug discovery. Hence, structural and functional information on the mini-protein functional epitopes can be easily obtained by NMR spectroscopy: this information can lead to the definition of the structure of pharmacophoric groups to be included in peptidomimetics or small-molecular-weight organic compounds that may represent useful candidates as drugs.

Keywords: engineering mini-protein; natural scaffolds; protein structure; peptidomimetic; drug design.
* Corresponding author. Fax: +33-1-69-089137;
e-mail: claudio.vita@cea.fr

0040–4020/00/$ - see front matter
PII: S0040-4020(00)00829-2

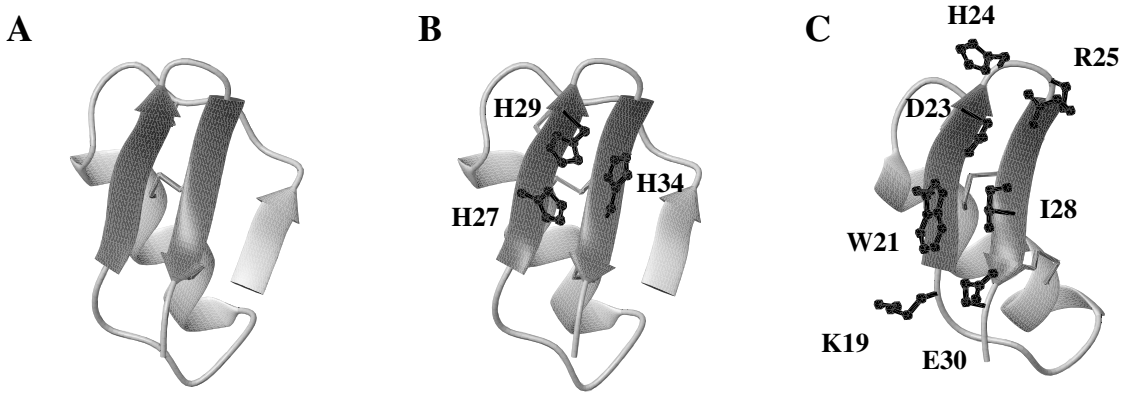

Figure 1. MOLMOL[39] structural model of scorpion toxin scaffold charybdotoxin (A, PDB code, 2crd), the engineered mini-protein binding metals (B) and the curaremimetic mini-protein (C, PDB code, 1cmr). Active side chains are labeled and represented in black sticks and balls.

Multidisulfide containing natural mini-proteins are particularly interesting as structural scaffolds, since these covalent cross-links represent the major determinants of structural stability, allowing substantial sequence mutations of many solvent exposed side chains, without perturbing folding efficiency. Within disulfide stabilized natural scaffolds, the natural structural motif found in scorpion toxins is one of the most attractive for protein engineering.[5,11] This motif contains two canonical secondary structures, an antiparallel β-sheet and a short α-helix, which are joined by three disulfide bridges in the interior of the structure (Fig. 1) that remarkably stabilize such structure to extremely denaturant conditions like high guanidine hydrochloride concentrations or boiling water.[11,12] Furthermore, in small-size multidisulfide scaffolds, particular spacing between cysteine residues have been indicated to play a role in facilitating formation of specific native disulfides, avoiding non-native ones:[13] this may explain the persistent folding efficiency of some engineered constructions based on this fold.[5] All known scorpion toxins contain the α/β structural motif, irrespective of their size, amino acid sequence and function.[14] Interestingly, the same structural motif is also present in insect defensins,[15] plant γ-thionins[16] and in a sweet-tasting mini-protein (brazzein).[17] A comparison between all known toxin sequences reveals that the six cysteines involved in the three disulfide bridges are the only residues strictly conserved.[14] Thus, this simple, compact, and well organized structural motif seems to have naturally evolved to include high sequence permissiveness (as shown by its compatibility with hundreds of different sequences) and functional versatility (as shown by its compatibility with different ion channel blockage activity in scorpion toxins, antimicrobial activity in defensins and interaction with receptors responsible for sweet tasting in brazzein).

The high structural stability, sequence permissiveness and functional versatility of the scorpion toxin fold suggested that this motif could be utilized as template to incorporate new sites in its β-sheet or α-helix regions and therefore engineer novel functions.[11] The potential of the scorpion scaffold was first tested by engineering a metal binding site on the structure of charybdotoxin, a 37-amino acid toxin isolated from the venom of the Israeli *Leiurus Quinquestriatus* scorpion.[11] The essential features of the Zn^{2+} binding site of human carbonic anhydrase B, three histidine residues on two antiparallel β-strands, were incorporated in the β-sheet of charybdotoxin (Fig. 1). Other substitutions in the vicinity of the putative active site were suggested after modeling studies: nine substitutions in all were introduced into the native 37-residue charybdotoxin sequence.[11] The new mini-protein was obtained by automatic solid phase synthesis and the three disulfide bridges formed efficiently, as in the case of charybdotoxin.[18] The newly engineered mini-protein was shown to bind metal ions, Cu^{2+}, Zn^{2+}, Cd^{2+}, Ni^{2+}, Mn^{2+}, with binding constants in the range 10^{-8}–10^{-5} M, in an order reflecting that observed in carbonic anhydrase B.[11] Structure analysis by CD,[18] ^{1}H NMR[11] and EPR (unpublished results) revealed that, in spite of the nine substitutions introduced, the chimeric mini-protein was correctly folded, presented the expected α/β motif of charybdotoxin and metals were bound by the three imidazole groups introduced. This work clearly demonstrated that the scorpion toxin structural motif could function as a template for mini-protein engineering and, following the same strategy of transfer of active sites to specific structural regions of this scaffold, new functions were then engineered.

A *curaremimetic protein*, designed to bind the nicotinic acetylcholine receptor, was obtained by incorporating, within this α/β motif, functionally important residues of the neurotoxin toxin α, that, as curare, binds to this receptor and blocks the opening of the ion channel, provoking flaccid paralysis.[12] By systematic mutagenesis analysis, the active site of toxin α was mainly identified in the concave face of the central loop of its β-sheet.[19] On the basis of a good superposition of the neurotoxin β-hairpin 25–36 to the charybdotoxin β-hairpin 25–36 (with a calculated 1.1 Å RMS deviation for the backbone atoms), eight side chains located on the solvent exposed face of the β-hairpin of charybdotoxin were mutated into the side chains found to be critical for the binding of toxin α to the acetylcholine receptor. Furthermore, the 6 N-terminal residues of charybdotoxin were deleted in order to allow greater solvent accessibility to the side chains of the β-sheet. The designed 31-residue chimeric sequence was synthesized by solid-phase synthesis and folded efficiently and correctly, as evidenced by HPLC and CD analysis.[12] The mini-protein, in competition ELISA, was shown to bind the nicotinic acetylcholine receptor with an IC_{50} of 5×10^{-5} M, which is 10^5 fold higher than that of the extremely potent neurotoxin

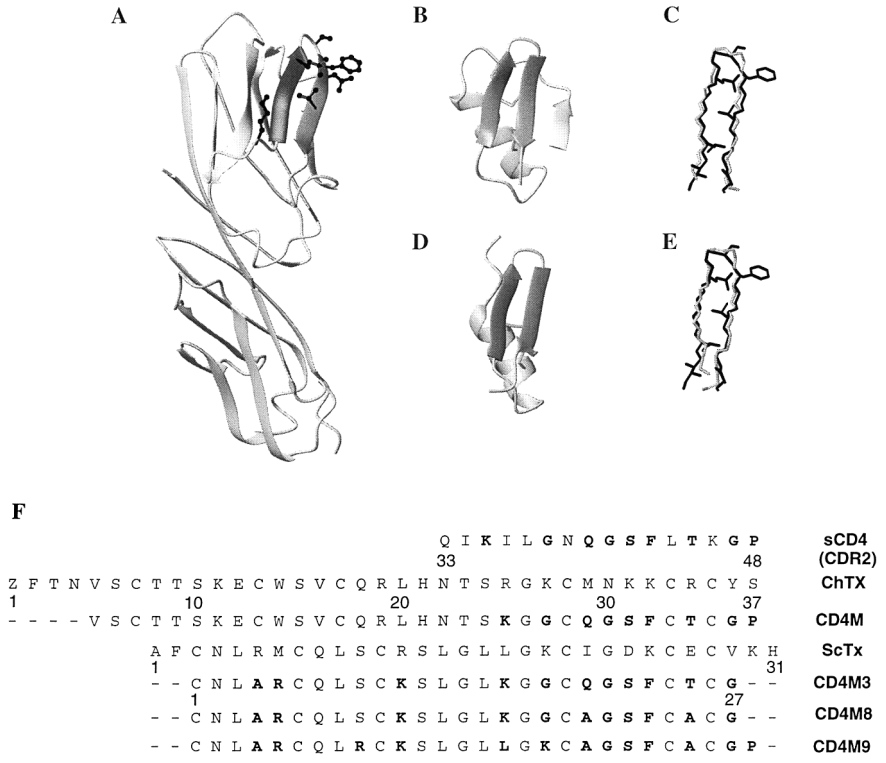

Figure 2. MOLMOL[39] structural model of D1-D2 domains of human CD4 (A, PDB code, 1cdh), compared to charybdotoxin (B) scyllatoxin (D, PDB code, 1scy) scaffold; structurally similar β-hairpins are shaded in darker gray and the side chains of CD4 CDR2-like loop, involved in gp120 binding, are in sticks and balls. The CD4 (36–48) β-hairpin (in black sticks) is superimposed on the charybdotoxin 25–37 (C) and scyllatoxin 18–30 (E) backbone atoms (in gray). F, Sequence (in one-letter code, Z is pyroglutamic acid) of CDR2-like loop of human CD4, charybdotoxin scaffold (ChTx), mini-protein CD4M, scyllatoxin scaffold (ScTx), mini-protein CD4M3, CD4M8 and CD4M9. Functional residues are in bold.

toxin α.[12,19] Structure determination of the *curaremimetic mini-protein* by ^{1}H NMR[20] revealed the presence of an α/β motif (Fig. 1) typical of the toxin scaffold, a good overall resemblance of the transferred site of the chimera with the original one in the parent neurotoxin. However, some divergences at the most solvent exposed (and more mobile) side chains were also observed. The structural resemblance was confirmed by the fact that antibodies, elicited in rabbits against the chimera, were able to recognize the parent neurotoxin and prevented its binding to the acetylcholine receptor.[12] To explain the relatively low affinity shown by the chimera for the acetylcholine receptor, we made theoretical calculations on the expected contribution to the receptor binding energy of the neurotoxin site that was transferred, as compared to neurotoxin total binding energy. This analysis indicated a good agreement between the estimated contribution and the binding energy of the chimera, determined from ELISA.[12] Thus, this explanation suggested that the structural graft was correctly achieved and the transferred site represented only a portion of the functional epitope. Probably, further improvement in receptor affinity could be obtained by submitting the mini-protein to combinatorial approach and functional selection, to rescue additional binding energy lost in the size reduction of the original binding site.

In this report we describe recent works on the engineering of a mini-protein, inhibitor of human immunodeficiency virus type-1 (HIV-1) infection, which has been designed to reproduce the essential features of the CD4 surface interacting with the viral envelope glycoprotein gp120. This example also illustrates how the binding affinity of the first engineered mini-protein was effectively improved on the basis of structure–function relationships, emphasizing the usefulness of using a stable and sequence permissive scaffold.

Results

The interaction of the gp120 envelope glycoprotein of HIV-1 with the CD4 protein represents the initial step of virus entry into target cells.[21] This interaction triggers a conformational change in the envelope glycoprotein gp120 that favors anchoring of the virus to the chemokine receptors, recently discovered co-receptors essential for entry; this in turn leads to exposure of the fusogenic domain of the envelope glycoprotein gp41, fusion of viral and cell membranes and finally cell infection.[22] The recently reported crystallographic structure of gp120, in complex with CD4 and the Fab portion of a neutralizing monoclonal antibody,[23] has demonstrated that a large surface (742 Å^2) of the domain D1 of CD4 binds to a large (800 Å^2) depression on gp120. The CDR2-like loop of CD4 is central in this interaction: CD4 Phe-43 side chain plugs the entrance of a deep cavity in gp120 and CD4 Arg59 side-chain is implicated in a double H-bond with Asp-368 side-chain of gp120. This structure justifies the critical functional role played by most of the residues of the CDR2-like loop, suggested by previous mutagenesis experiments.[24] We envisioned that a mini-protein reproducing the structural features of the CDR2-like loop could function as an

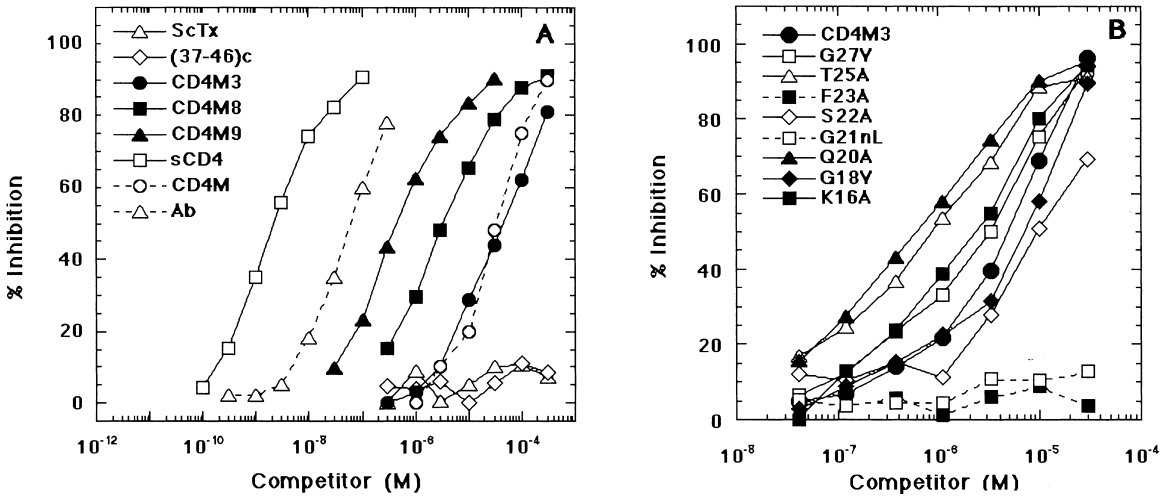

Figure 3. (A) ELISA of inhibition of gp120 binding to coated CD4 by recombinant CD4 (sCD4), mini-protein CD4M, CD4M3, CD4M8, CD4M9 and control scyllatoxin (ScTx), cyclic CD4(37–46) peptide. Inhibition of gp120 binding to coated CD4 by purified anti-CD4M antibodies (Ab, dashed line) is also shown. (B) ELISA of inhibition of gp120 binding to coated CD4 by CD4M3 mutants; residue mutated are indicated by standard one letter code (nL is norleucine).

inhibitor of CD4-gp120 interaction and, consequently, of virus attachment to cells and infection.

According to the strategy used in previous examples, we looked for a mini-protein scaffold that could function as template for the structural reproduction of the CD4 site: we found that the scorpion toxin charybdotoxin contained a solvent exposed β-hairpin (sequence 25–37) that could be superposed on the CDR2-like region (sequence 36–48) of human CD4, with an RMS deviation for backbone atoms of 1.3 Å only (Fig. 2). In order to allow access of the β-hairpin of the scaffold to gp120 binding, the N-terminus was also shortened by four residues. The resulting 33-residue CD4 mimetic, denominated CD4M, containing nine CD4 residues in structurally equivalent regions of the scaffold, was synthesized by automated solid-phase synthesis and folded correctly, as evidenced by CD spectroscopy.[25] The capability of this mini-protein to bind gp120 and to inhibit CD4-gp120 interaction was evaluated by competition ELISA. In this system, CD4M was able to specifically inhibit the binding of soluble recombinant gp120 to coated recombinant CD4, with 3×10^{-5} M IC_{50} (Fig. 3A). In addition, this mini-protein was also injected in rabbits, and the elicited antibodies were purified by affinity-chromatography. These antibody sub-population recognized human CD4 effectively (not shown) and inhibited gp120 binding to CD4 (Fig. 3A). This demonstrates that the designed mini-protein, despite its low affinity for gp120, reproduced the target CD4 site in a way good enough to elicit antibodies specific for the CD4 CDR2-like loop, thus preventing gp120 interaction with CD4.

In order to produce a better CD4 mimic, we then adopted a different α/β framework, the scorpion scyllatoxin (31 residues only). This mini-protein appeared superior as structural scaffold, as it contained a β-hairpin, sequence 18–30, which superimposed its backbone atoms to the CDR2-like loop of CD4, sequence 36–48, with an RMS deviation of only 1.1 Å; what's more, it presented a shorter loop connecting the helix to the first β-strand and no N-terminal β-strand (Fig. 2), presumingly allowing a sterically more favorable gp120 binding. Thus, a new chimeric mini-protein was designed which preserved the structurally important Cys residues of the scyllatoxin scaffold, but included the solvent exposed CDR2 side chains, Gly38, Gln40, Gly41, Ser42, Phe43, Thr45 and Gly47 into structurally equivalent regions of scyllatoxin (Fig. 2). Furthermore, in order to increase CD4 structural mimicry, we also included an Arg7 and a Lys18, topologically equivalent to the functional Arg59 and Lys35 of CD4. To abolish the toxic activity of the toxin, two more residues were mutated, Arg6 to Ala, Arg13 to Lys, and two residues at the N- and C-terminus were deleted. The final 27-amino acid sequence contained 9 residues topologically positioned as in human CD4 and maintained only 16 of the 31 amino acids of the scyllatoxin sequence. The new CD4 mimic, named CD4M3,[26] was chemically synthesized, folded efficiently and presented a CD spectrum similar to that of scyllatoxin, in spite of many mutations in the native sequence. In competitive ELISA, the mini-protein was able to specifically bind gp120 with 3.0×10^{-5} M IC_{50}, which is identical to that shown by CD4M and four order of magnitude higher than that shown by sCD4 (Fig. 3A, Table 1).

The biological performance of this mini-protein was then effectively improved on the basis of a 'rational' structure–function relationship approach. In the first place, its

Table 1. gp120 Binding activity of engineered mini-proteins and recombinant sCD4

Competitor[a]	IC_{50}[b] (M)
sCD4	$1.4\pm0.1\times10^{-9}$
CD4M	$3.0\pm0.5\times10^{-5}$
CD4M3	$3.0\pm0.5\times10^{-5}$
CD4M8 (CD4M3(Q20A, T25A))	$2.3\pm0.2\times10^{-6}$
CD4M14 (CD4M8(K16L, G18K))	$1.5\pm0.5\times10^{-6}$
CD4M16 (CD4M8(K16L, G18K, P28))	$1.0\pm0.6\times10^{-6}$
CD4M17 (CD4M8(S9R))	$9.0\pm0.5\times10^{-7}$
CD4M9 (CD4M8(S9R, K16L, G18K,P28))	$4.0\pm0.1\times10^{-7}$

[a] Mini-protein mutants are indicated by the residue mutation.
[b] Concentration (±SD) of competitor required for 50% inhibition of sCD4 binding to fixed gp120, in competitive ELISA.

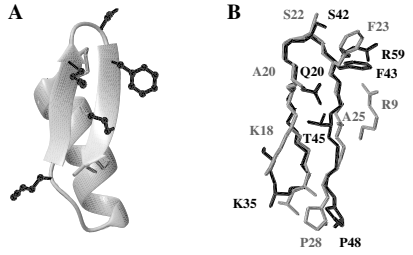

Figure 4. (A) Average NMR structure of mini-protein CD4M3, showing the functional side chains as sticks and balls. (B) Superposition of the CD4 35–48 β-hairpin (black sticks) and modeled 15–28 β-hairpin (gray sticks) of CD4M9.

three-dimensional structure was determined by ^{1}H NMR spectroscopy:[26] this analysis showed that the mini-protein contained a very well-defined structure (only 0.2 Å RMS deviation between the backbones of the 20 final structures) and an α/β fold characteristic of the scorpion scaffold, with a helix in the region 2–13 and an antiparallel β-sheet in the region 16–26 (Fig. 4A). Most importantly, this analysis revealed that the putative binding site, transferred from CD4, was very well defined and that the backbone atoms of the sequence 17–26 of CD4M3 could be superposed on the corresponding atoms of the sequence 37–46 of native CD4 with a RMS deviation of 0.61 Å only.[26] Furthermore, the side chains of Gln20, Ser22, Phe23 and Thr25 had an orientation very similar to that of the corresponding side chains in CD4. In particular, the Phe23 side chain, which was very well defined because of many long range contacts observed in CD4M3, stack out into the solvent in a conformation which is unusual for a hydrophobic moiety, but was reminiscent of that of Phe43 of CD4 (Figs. 2A and 4A). This moiety, in the crystal structure of the CD4-gp120 complex, is seen to plug the entrance of the 'Phe43 cavity' of gp120.[23] Lys16, Arg5 side chains and C-terminal Gly27, however, diverged from the structure of CD4 Lys35, Arg59 and β-strand 43–47, respectively. Secondly, a functional analysis was performed on the mini-protein, by substituting each putative active side chain by an alanine residue and the three glycine residues by a valine (or norleucine). The effect of these substitutions on gp120 binding (Fig. 3B) clearly indicated that each transferred residue played a different functional role and pointed to a Phe residue present at the tip of the β-hairpin, as a 'hot spot' of the chimera active surface, in agreement with the literature data on mutagenesis of recombinant hCD4.[24] Interestingly, this analysis revealed that two substitutions, Gln20Ala and Thr25Ala, increased its apparent gp120 binding affinity by 5-fold (Fig. 3B): a similar increase in binding affinity was also observed in sCD4.[27] Thus, the performed structural and functional analysis, pointing to a remarkable structural and functional resemblance between the CD4 and the mini-protein binding sites, suggested also some changes that could increase its structural and functional mimicry with the CD4 site. The two mutations enhancing affinity, Gln20Ala and Thr25Ala, were incorporated in a double mutant, denominated CD4M8: the mini-protein affinity increased by about 10-fold to 2.3×10^{-6} M IC_{50} (Table 1, Fig. 3A). Additional mutations, suggested by the structural analysis, were then modeled on the CD4M3 mini-protein structure, and the molecule then synthesized. Thus, Lys16 was moved to position 18, to better mimic the side chain of CD4 Lys35: this mutant (CD4M14) showed increased

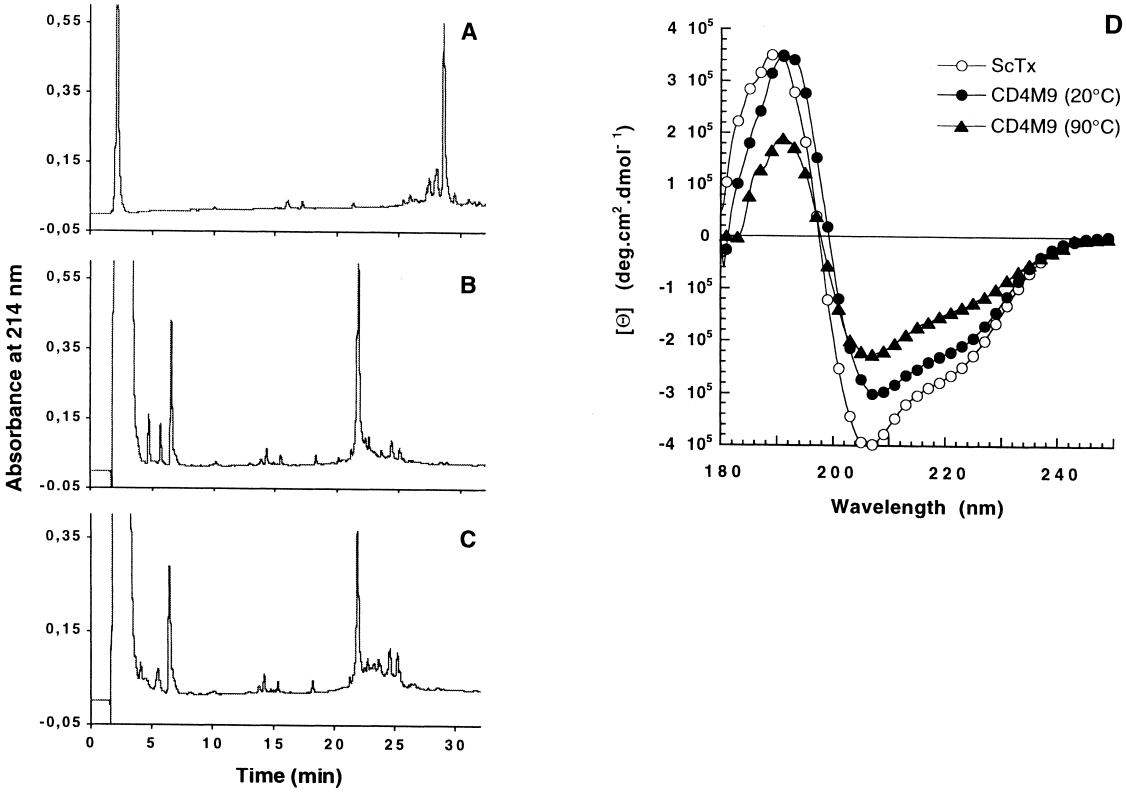

Figure 5. Reverse-phase HPLC of crude (A) and refolded CD4M9 mini-protein, in benign buffer (B) and in the presence of 6.0 M guanidine hydrochloride (C). Far UV circular dichroism spectra (D) of scorpion scyllatoxin (ScTx) and engineered CD4M9, recorded at 20°C and 90°C.

binding affinity (Table 1). In order to reduce the flexibility of the mini-protein C-terminus, as evidenced by the structural analysis, a Pro residue was added in position 28 (mutant CD4M16): this sequence extension also produced some increase in affinity (Table 1). Given the important role of CD4 Arg59 in gp120 binding,[23,24] we placed a guanidinium group in the more favorable position 9 of the mini-protein structure: this mutant (CD4M17) also showed an increase in binding affinity (Table 1). Then, modeling studies indicated that the charybdotoxin scaffold contained a Cys7 residue, with its α-amino group in a more favorable position to present a guanidinium group mimicking the Agr59 side chain. Therefore, we designed, and then produced, a CD4M mutant, deleted of two N-terminal residues and containing either a guanidine-acetyl or guanidine-propionyl at the N-terminus. Unfortunately, the presence of a positively charged guanidinium group at that position seemed to destabilize the scaffold significantly, as shown by CD spectroscopy (not shown), and the binding affinity of these mini-proteins were 100-fold down as compared to previous scyllatoxin mini-proteins. At the end, we modeled a new mini-protein, based on the CD4M3 structure and denominated CD4M9, including all the mutations described to enhance affinity, Ser9Arg, Lys16Leu, Gly18Lys, Gln20Ala, Thr25Ala, Pro28, (Fig. 4B). The new 28-residue peptide was obtained by solid-phase peptide synthesis in good yields and, in the presence of redox buffer (see Methods), folded efficiently as shown by HPLC analysis (Fig. 5B). Furthermore, folding efficiency was insensitive to the presence of 4.0 M guanidine-hydrochloride and only slightly perturbed by 6 M concentration of this potent protein denaturant (Fig. 5C), emphasizing the high folding efficiency and conformational stability of the engineered mini-protein. The purified mini-protein presented the expected molecular mass, and exhibited a far-UV CD spectrum (Fig. 5D), very similar to that presented by native scyllatoxin, in agreement with its preserved folding efficiency observed in HPLC. The conformational stability of the CD4M9 mini-protein was also remarkable, as shown by its CD spectrum obtained at 90°C (Fig. 5D), emphasizing the great sequence permissiveness of the scaffold chosen, even after multiple substitutions. When tested in ELISA, the CD4M9 mini-protein exhibited a 4.0×10^{-7} M IC_{50}: this represents a remarkable 100-fold increase in apparent gp120 binding affinity, as compared to the first CD4M3 mini-protein.

Both CD4M3 and the improved CD4M8 and CD4M9 mini-proteins were also examined for their ability to prevent infection of HeLa cells,[26] stably expressing CD4, and the CCR5, CXCR4 co-receptors of HIV-1. The mini-proteins were not toxic for cells as verified by trypan blue exclusion.[26] Laboratory-adapted virus strains, HIV-1_{LAI} and HIV-1_{BaL}, using CXCR4 and CCR5, respectively, as co-receptors for entry, were used. Independently of co-receptor usage, the CD4M3 mini-protein inhibited infection with 1.0×10^{-5} M IC_{50}, complete inhibition was obtained at 1.0×10^{-4} M (J.C. Gluckman, personal communication). Furthermore, in agreement with the ELISA, the CD4M8 and CD4M9 mutants exhibited higher antiviral activity. In particular, the optimized CD4M9 mini-protein fully inhibited HIV-1_{LAI} infection at 1.0×10^{-6} M. Protection from infection was effective only when the mini-proteins were added before but not after the virus, suggesting that they prevented virus binding to CD4, the primary receptor of virus entry. Finally, the optimized CD4M9 mini-protein had a large spectrum antiviral activity, inasmuch as it protected also peripheral blood lymphocytes (PBL) from infection by the laboratory-adapted strains HIV-1_{LAI} and HIV-1_{BaL}, as well as two primary syncitium-inducing HIV-1 isolates (J. C. Gluckman, personal communication).

Discussion

This work illustrates a rational approach for reproducing the core of a protein–protein interacting surface in a mini-protein system. In this approach, the transfer of a protein binding site to a small protein scaffold is followed by an optimization process, implying: (i) three-dimensional structure determination of the first designed mini-protein, (ii) incorporation of mutations aiming to increase the structural mimicry of mini-protein binding surface with that of the target protein, and (iii) functional characterization of mutants in a recursive process. Utilization of structural and functional data resulted in a remarkable 100-fold increase of binding affinity of the first engineered mini-protein. This approach can be probably applied to other protein–protein interaction systems, thus may be of wide use, since is based on the similarity between the structural motif of a protein binding domain and that of a small size natural scaffold. The wealth of small-size structures present in protein structure data bases and the fact that common structural motifs are used by nature to express all needed proteins and functions, strongly suggests that a protein binding epitope may find its structural equivalent in a small-size natural structure: this may then function as scaffold and artificially reproduce the large protein binding epitope in a much simpler system. The remarkable structural stability of the chosen scaffold, allowing introduction of multiple mutations not perturbing folding efficiency and structural integrity, represents a further advantage; thus, mutations could be reliably modeled, new mutants structurally analyzed, then produced and tested in a short time. In addition, in order to introduce useful flexibility in particular structural regions of the scaffold, disulfide bonds can be also replaced by non-covalent interactions (as suggested by previous work on charybdotoxin scaffold[13c]), this may result in structural adjustments that might be required to achieve optimal binding. Furthermore, we expect that further structural mimicry and functional improvement can be once more obtained on the basis of structural data derived from the determination of the structure of the mini-protein–gp120 complex. Given the difficulty in obtaining crystals of such complex or determining its structure by NMR, starting from the known crystallographic structure of CD4 bound to gp120, we recently modeled and energy minimized a putative miniprotein–gp120 complex: the analysis of such model provided effective useful information that allowed further structural mimicry and remarkable improvements of binding affinity (to be published).

Our results demonstrate that, by reproducing the core of the gp120-binding surface of CD4 in a mini-protein system, we could engineer an inhibitor of CD4-gp120 interaction and HIV-1 attachment and entry into cells. The engineered

optimized mini-CD4 differs from previously reported peptide constructs based on CDR loops of CD4[28] by its stable and well-defined CD4-like structure, which fully explains its gp120 binding affinity and antiviral activity. Linear peptides corresponding to the helical region of the viral envelope glycoprotein gp41, such as DP107, DP-178 and more recent D-amino acid containing versions,[29] have been also proposed and shown to inhibit virus–cell membrane fusion and cell infection, presumably by interfering with the gp41 coil-coiled structure formation. Even if the multi-disulfide mini-proteins here described may present useful pharmacological properties, such as low antigenicity and resistance to proteolytic degradation, the peptide nature and low antiviral potency may limit their application in therapy. However, because of their small size, these optimized mini-proteins may be easily manipulated chemically and equipped with useful chemical probes (fluorescent labels, photo-activable groups, affinity tags, isotopically enriched amino acids,…) without disturbing function, thus, they may represent useful and unique tools in the study of the complex process of HIV-1 entry into target cells. Furthermore, incorporation of appropriate radioactive probes on optimized mini-CD4 may also result in new tracers, useful for in vivo imaging in diagnostic applications.

Short protein fragments constrained into cyclic peptides,[30] small size linear or cyclic peptides derived from library selection[31] have been shown to bind to large proteins and receptors, as antagonists and even agonists, confirming that small molecules may act as surrogates of large natural proteins. Likewise, in the case of growth hormone–receptor interaction, in spite of the presence of a large protein–protein contact area, which defines the structural epitope of interaction, few critical side chains have been shown to dominate the energy of binding: these side chains form the core of the functional epitope, also called the 'hot spot' of the protein–protein interaction surface.[32] However, the definition of the 'hot spot' of a protein–protein interface is not always obvious and unanimous, depending on the experimental methods used in functional analysis. This uncertainty emphasizes the usefulness to link the structure-based approach based on a structural scaffold with a systematic mutational analysis, which may optimize the inserted site, once incorporated into the new scaffold. The protein binding domain of protein A and ANP hormone have been minimized to about half their original size with retention of activity, by combining a structure-based approach and phage display methods.[33] Obviously, the combinatorial approach can be also applied to the stable mini-protein scaffold that we used, and critical positions of its active site randomized to produce libraries that may be screened for mutants binding gp120 with highest affinity. The small size, stability and folding efficiency of this scaffold, may allow synthesis of chemical libraries, with incorporation of non-natural amino acids, thus significantly extending the molecular diversity of the libraries.

Finally, the ability to produce a mini-protein presenting a specific active site, within a well-defined conformation constrained in an appropriate small size scaffold, represents a substantial step forward towards the design of low-molecular-weight non-peptide pharmaceuticals.[10,34] The derivation of non-proteinaceous small molecules (preferably <1000 Da) from large natural proteins or protein domains is a difficult task in structure-based drug design, implying recursive cycles of modeling, synthesis and testing. Despite the fact that, presently, the advent of combinatorial chemistry coupled to high-throughput screening has shifted the attention of drug companies away from structure-based structure design, the engineering of active sites on small and constrained scaffolds may represent a 'rational' alternative, applicable to the discovery of leads that can be converted in a final useful drug. One of the most difficult steps in lead discovery, after identification of a suitable protein target, is the identification of the chemical groups forming the pharmacophore and their spatial relationships. Engineered mini-CD4 may represent an interesting intermediate facilitating the design of new antiviral compounds. An interesting peptidomimetic reproducing the β-turn of the CDR2-like region of CD4 has been previously described.[35] However, testing this peptidomimetic in different laboratories has revealed that antiviral potency was lower and that it did not act according to a competitive inhibition mechanism.[36] Detailed structural and functional analysis of the novel mini-protein, mimicking a portion of the CD4 CDR2-like region which is larger than the β-turn tetrapeptide mimic, may provide new unprecedented information on the pharmacophoric groups and their mutual spatial positioning, to be included in a novel potent and specific peptidomimetics or organic compounds, active as inhibitor of HIV-1 entry into target cells.

Conclusion

This work demonstrates that the transfer of functional sites to stable small presentation scaffolds, followed by an optimization process based on structure–function analysis, represents the means to reproduce the structural and functional features of the 'hot spot' of protein–protein interaction surfaces. Such stable and functional mini-proteins constitute a significant minimization step of a large and intricate protein–protein interface and, because of their well-defined three-dimensional structure, their specific biological activity, the ease with which they can be chemically manipulated, may represent new exquisite tools in biology, biotechnology, and new generation tracers for diagnostic applications. Finally, they also represent structural intermediates useful to determine the bio-active conformation of pharmacophoric groups and to identify frameworks for peptidomimetic design, or directly lead to new small organic structures representing novel drug candidates. Thus, novel bio-active mini-proteins reproducing protein–protein interfaces may become useful components of the armamentarium used in medicinal chemistry, accelerating the development of new drugs.

Experimental

Peptide synthesis

Mini-proteins and peptides were synthesized on an Applied Biosystems 433A Peptide Synthesizer, by the solid phase method, utilizing fluorenylmethyloxycarbonyl

(Fmoc)-protected amino acids, polyethylene glycol-polystyrene resin, equipped with a peptide amide linker (PAL-PEG-PS resin, Perseptive Biosystems) and 2-(1-H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) coupling (FastMoc protocol on 0.1 mmol scale; with amino acid derivatives suggested by the Synthesizer manufacturer). Ala and Val mutants were simultaneously synthesized on an Advanced Chemtech 357 Multi synthesizer, by using the same chemical approach. Guanidine-acetyl and guanidine-propionyl moieties were incorporated to the N-terminus of peptide-resin, by manual coupling its hydrochloride salts as 1-hydroxybenzotriazole (HOBt) ester (10-fold molar excess) for 45 min in dimethylformamide. After peptide deprotection,[37] disulfide bonds were formed directly on the crude products dissolved at 0.1 mg/ml in 50 mM phosphate buffer, 0.1 M NaCl, pH 7.8, in the presence of 5 mM/0.5 mM reduced/oxidized glutathione. Synthetic peptides were purified by reverse-phase HPLC, on a preparative Vydac C18 (2.5×25 cm) column, by using a 90 min 0–50% linear gradient of acetonitrile/water, containing 0.1% trifluoroacetic acid, at 10 ml/min. Peptide purity was assessed by analytical HPLC, and identity was verified by amino acid analysis and electrospray mass spectrometry.

Circular dichroism

Circular dichroism spectra were recorded on a Jobin Yvon CD6 dichrograph, driven by an IBM-PC operating with a CD Max data acquisition and manipulation program. Spectra were recorded in a 0.1 cm pathlength quartz cell, with a protein concentration of $1.5–2.10^{-5}$ M in 5 mM phosphate buffer, pH 7.5, by accumulating 4 scans with an integration time of 0.5 s every 0.2 nm.

Structure calculations

Structures were determined from 2D ^{1}H NMR experiments (NOESY, TOCSY and DQF-COSY) recorded at 600 MHz (Bruker AMX-600) on a 4 mM CD4M3 mini-protein sample, pH 3.5, either in 95% H_2O/5% D_2O or 100% D_2O solutions, at 25°C. Standard methods were used for recording and analyzing spectra, deriving distance and angle restraints.[38] Three-dimensional structures were generated from 285 NOEs and 46 angular restraints using the standard force-field parameters of X-PLOR 3.1. From forty structures generated, twenty had any violation of NOE or angular restraints exceeding 0.2 Å and 5°, respectively, and were thus selected for analysis. Global RMS deviation as low as 0.2 Å were calculated for all the backbone atoms (0.8 Å for all the heavy atoms), and all structures had good energy (E_{total}=−105.5 kcal, E_{imp}=3.16 kcal, E_{bond}=3.523 kcal, E_{VdW}=−28.256 kcal, E_{NOE}=9.331 kcal, E_{angle}=9.035 kcal and E_{elect}=−103.1 kcal, for the best structure).

Structure analysis and modeling

All structures were displayed, analyzed, and compared on a Silicon Graphics 4D/25 station using the Sybyl package (Tripos Associates, Inc.). CD4M9 model was obtained by Sybyl software, by using the coordinates of the CD4M3 average NMR structure, introducing the Ser9Arg, Lys16Leu, Gly18Lys, Gln20Ala, Thr25Ala and Pro28 mutations, and finally energy minimizing the resulted model by Sybyl force field (without electrostatics). Molecular graphics were produced using Molmol.[39]

Rabbit immunization

Antibodies directed against the CD4M mini-protein were obtained by immunizing, subcutaneously and at multiple sites, two rabbits (blanc du Bouscat, Wiss) with 220 μg of mini-protein in complete Freund's adjuvants (CFA). Rabbits were re-immunized three times in incomplete Freund's adjuvants (IFA) with the same quantity of peptide at 21 days interval. Animals were bled 15 days after each boosting and their sera were tested for specific antibody production by enzyme-linked immunosorbent assay (ELISA). Elicited antibodies were titrated against either CD4M and parent CD4, coated on microtiter plates. Titers were 10^5 and 10^4 for CD4M and sCD4, respectively. Pooled anti-CD4M sera were purified by affinity chromatography through a CD4-Sepharose 4B column (3.5×1 cm), prepared by coupling 3.0 mg of the hybrid HSA-CD4 protein (a genetic construction of human serum albumin fused to D1-D2 domain of CD4) to 0.8 g of CNBr-activated Sepharose 4B (Pharmacia).

Competitive binding assays

Inhibition of gp120-CD4 interaction was measured in a competitive enzyme-linked immunosorbent assay (ELISA) as described.[25] Briefly, 250 ng/well of sCD4 (D1 to D4 domains; gift of R. W. Sweet, SmithKline–Beecham Pharmaceuticals, King of Prussia, PA) were coated overnight at 4°C in 96-well plates (Maxisorb, Nunc, Denmark); 80 ng/well of $rgp120_{LAI}$ (Intracel, Issaquah, WA) were then added, followed by addition of different concentrations of soluble competitors (CD4, mini-proteins and anti-CD4M3 antibodies), anti-gp120 NEA mAb (NEN-DuPont, Wilmington, DE), a goat anti-mouse peroxidase-conjugated antibody (Jackson ImmunoResearch, West Grove, PA) and the 3,3′,5,5′-tetramethylbenzidine substrate (Sigma, St Louis, MO) for revelation. Inhibition of binding was calculated from the 450 nm OD using the formula: % inhibition= $100\times(OD_{gp120}-OD_{gp120+comp})/OD_{gp120}$. Results are means of triplicate experiments.

Acknowledgements

We thank the French 'Agence Nationale de Recherches sur le SIDA (ANRS)' and 'Ensemble Contre le SIDA-SIDACTION' for financial support.

References

1. (a) Bork, P.; Holm, L.; Sander, C. *J. Mol. Biol.* **1994**, *242*, 309–320. (b) Gerstein, M.; Levitt, M. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11911–11916.
2. Jones, P. T.; Dear, P. H.; Foote, J.; Neuberg, M. S.; Winter, G. *Nature* **1986**, *321*, 522–525.
3. For recent review, see: (a) Hudson, P. J. *Curr. Opin. Biotechnol.* **1998**, *9*, 395–402. (b) Dall'Acqua, W.; Carter, P. *Curr. Opin.*

Struct. Biol. **1998**, *8*, 443–450. (c) Rader, C.; Barbas, III C. F. *Curr. Opin. Biotechnol.* **1997**, *8*, 503–508.
4. Nygren, P. A.; Uhlén, M. *Curr. Opin. Struct. Biol.* **1997**, *7*, 463–469.
5. Vita, C.; Vizzavona, J.; Drakopoulou, E.; Zinn-Justin, S.; Gilquin, B.; Ménez, A. *Biopolymers (Peptide Science)* **1998**, *47*, 93–100.
6. Pessi, A.; Bianchi, E.; Crameri, A.; Venturini, S.; Tramontano, A.; Solazzo, M. *Nature* **1993**, *362*, 367–369.
7. (a) Houston, Jr. M. E.; Wallace, A.; Bianchi, E.; Pessi, A.; Hodges, R. S. *J. Mol. Biol.* **1996**, *262*, 270–282. (b) Myszka, D. F.; Chaiken, I. M. *Biochemistry* **1994**, *33*, 2363–2372. (c) Adamson, J. G.; Zhou, N. E.; Hodges, R. S. *Curr. Opin. Biotechnol.* **1993**, *4*, 428–437.
8. (a) Mutter, M.; Vuilleumier, S. *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 535–554. (b) Tuchscherer, G.; Dömer, B.; Sila, U.; Kamber, B.; Mutter, M. *Tetrahedron* **1993**, *49*, 3559–3575. (c) Tuchscherer, G.; Grell, D.; Mathieu, M.; Mutter, M. *J. Peptide Res.* **1998**, *54*, 185–194. (d) Tuchscherer, G.; Scheibler, L.; Dumy, P.; Mutter, M. *Biopolymers* **1998**, *47*, 63–73.
9. Vita, C. *Curr. Opin. Biotechnol.* **1997**, *8*, 429–434.
10. Cunningham, B. C.; Wells, J. A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 457–462.
11. Vita, C.; Roumestand, C.; Toma, F.; Ménez, A. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6404–6408.
12. Drakopoulou, E.; Zinn-Justin, S.; Guenneugues, M.; Gilquin, B.; Ménez, A.; Vita, C. *J. Biol. Chem.* **1996**, *271*, 11979–11987.
13. (a) Zhang, R.; Snyder, G. H. *J. Biol. Chem.* **1989**, *264*, 18472–18479. (b) Zhang, R.; Snyder, G. H. *Biochemistry* **1991**, *30*, 11343–11348. (c) Drakopoulou, E.; Vizzavona, J.; Neyton, J.; Aniort, V.; Bouet, F.; Virelizier, H.; Ménez, A.; Vita, C. *Biochemistry* **1997**, *37*, 1292–1301.
14. (a) Bontems, F.; Roumestand, C.; Gilquin, B.; Ménez, A.; Toma, F. *Science* **1991**, *254*, 1521–1523. (b) Ménez, A.; Bontems, F.; Roumestand, C.; Gilquin, B.; Toma, F. *Proc. R. Soc. Edinburgh* **1992**, *99B*, 83–103. (c) Darbon, H.; Blanc, E.; Sabatier, J. M. *Perspectives Drug Discovery Design* **1999**, *15*, 41–60.
15. Bonmatin, J. M.; Bonnat, J. L.; Gallet, X.; Vovelle, F.; Ptak, M.; Reichhart, J. M.; Hoffmann, J. A.; Keppi, E.; Legrain, M.; Achstetter, T. *J. Bio. NMR* **1992**, *2*, 235–256.
16. Bruix, M.; Jiménez, M. A.; Santoro, J.; Gonzàlez, C.; Collila, F. J.; Méndez, E.; Rico, M. *Biochemistry* **1993**, *32*, 715–724.
17. Caldwell, J. E.; Abildgaad, F.; Dzakula, Z.; Ming, D.; Hellekant, G.; Markley, J. L. *Nature Struct. Biol.* **1998**, *5*, 427–431.
18. Pierret, B.; Virelizier, H.; Vita, C. *Int. J. Peptide Protein Res.* **1995**, *46*, 471–479.
19. Trémeau, O.; Lemaire, C.; Drevet, P.; Pinkasfeld, P.; Ducancel, F.; Boulain, J. C.; Ménez, A. *J. Biol. Chem.* **1995**, *270*, 9362–9369.
20. Zinn-Justin, S.; Guenneugues, M.; Drakopoulou, E.; Gilquin, B.; Vita, C.; Ménez, A. *Biochemistry* **1996**, *35*, 8535–8543.
21. (a) Dalgleish, A. G.; Beverley, A. C.; Clapham, P. R.; Crawford, D. H.; Greaves, M. F.; Weiss, R. A. *Nature* **1984**, *312*, 763–767. (b) Klatzmann, D.; Champagne, E.; Chamaret, S.; Gruest, J.; Guetard, D.; Hercend, T.; Gluckman, J. C.; Montagnier, L. *Nature* **1984**, *312*, 767–768.
22. For recent review, see: (a) Berger, E. A. *AIDS* **1997**, *11*, S3–S16. (b) Littman, D. R. *Cell* **1998**, *93*, 677–680. (c) Chan D. C.; Kim, P. S. *Cell* **1998**, *93*, 681–684. (d) Choe, H.; Martin, K. A.; Farzan, M.; Sodroski, J.; Gerard, N. P.; Gerard, C. *Seminar Immunol.* **1998**, *10*, 249–257. (e) Cairns, J. S.; D'Souza M. P. *Nature Med.* **1998**, *4*, 563–568.
23. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 648–659.
24. For review, see: (a) Sweet, R. W.; Truneh, A.; Hendrickson, W. A. *Curr. Opin. Biotechnol.* **1991**, *2*, 622–633. (b) Ryu, S. E.; Trueh, A.; Sweet, R. W.; Hendrickson, W. A. *Structure* **1994**, *2*, 59–74.
25. Drakopoulou, E.; Vizzavona, J.; Vita, C. *Lett. Pep. Sci.* **1998**, *5*, 241–245.
26. Vita, C.; Drakopoulou, E.; Vizzavona, J.; Rochette, S.; Martin, L.; Ménez, A.; Roumestand, C.; Yang, Y. S.; Ylistagui, L.; Benjouad, A.; Gluckman, J. C. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13091–13096.
27. (a) Arthos, J.; Deen, K. C.; Chaikin, M. A.; Fornwald, J. A.; Sathe, G.; Sattentau, Q. J.; Clapham, P. R.; Weiss, R. A.; McDougal, J. S.; Pietropaolo, C.; Axel, R.; Truneh, A.; Maddon, P. J.; Sweet, R. W. *Cell* **1989**, *57*, 469–481. (b) Ashkenazi, A.; Presta, L. G.; Marsters, S. A.; Camerato, T. R.; Rosenthal, K. A.; Fendley, B. M.; Capon, D. J. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7150–7154. (c) Moebius, U.; Clayton, L. K.; Abraham, S.; Harrison, S. C.; Reinherz, E. L. *J. Exp. Med.* **1992**, *176*, 507–517.
28. (a) Lifson, J. D.; Hwang, K. M.; Nara, P. L.; Fraser, B.; Padgett, M.; Dunlop, N. M.; Eiden, L. E. *Science* **1988**, *241*, 712–716. (b).Zhang, X.; Gaubin, M.; Briant, L.; Srikantan, V.; Ramachandran, M.; Saragoni, U.; Weiner, D.; Devaux, C.; Autiero, M.; Platier-Tonneau, D.; Greene, M. I. *Nat. Biotechnol.* **1997**, *15*, 150–154.
29. (a) Wild, C.; Oas, T.; McDanal, C.; Bolognesi, D.; Matthews, T. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10537–10541. (b) Wild, C. T.; Shugars, D. C.; Greenwell, T. K.; McDanal, C.; Bolognesi, D.; Matthews, T. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9770–9774.
30. (a) Wei, L.; Gibbs, B. F.; Saragovi, H. U. *J. Biol. Chem.* **1995**, *270*, 6564–6569. (b) Monfardini, C.; Kieber-Emmons, T.; Voet, D.; Godillot, A. P.; Weiner, D. B.; Williams, W. V. *J. Biol. Chem.* **1996**, *271*, 2966–2971. (c) Zhang, X.; Piatier-Tonneau, D.; Auffray, C.; Muraly, R.; Mahapatra, A.; Zhang, F. Q.; Maier, C. C.; Saragovi, H. U.; Greene, M. *Nature Biotechnol.* **1996**, *14*, 472–475.
31. (a) O'Neil, K. T.; Hoess, R. H.; Jackson, S. A.; Ramachandran, N. S.; Mousa, S. A.; DeGrado, W. *Proteins Struct. Funct. Genet.* **1992**, *14*, 509–515. (b) Wrighton, N. C.; Farrell, F. X.; Chang, R.; Kashyap, A. K.; Barbone, F. P.; Mulcahy, L. S.; Johnson, D. L.; Barret, R. W.; Jolliffe, L. K.; Dower, W. J. *Science* **1996**, *273*, 458–464. (c) Livnah, O.; Stura, E. A.; Johnson, D. L.; Middleton, S. A.; Mulcahy, L. S.; Wrighton, N. C.; Dower, W. J.; Jolliffe, L. K.; Wilson, I. A. *Science* **1996**, *273*, 464–471. (d) Hirschmann, R.; Yao, W.; Cascieri, M. A.; Strader, C. D.; Maeschler, L.; Cichy-Knight, M. A.; Hynes, Jr. J.; van Rijn, R. D.; Sprengler, P. A.; Smith, III A. B. *J. Med. Chem.* **1996**, *39*, 2441–2448. (e) McBride, J. D.; Freeman, N.; Domingo, G. J.; Leatherbarrow, R. J. *J. Mol. Biol.* **1996**, *259*, 819–827.
32. Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383–386.
33. (a) Li, B.; Tom, J. Y. K.; Oare, D.; Yen, R.; Fairbrother, W. J.; Wells, J. A.; Cunningham, B. C. *Science* **1995**, *270*, 1657–1660. (b) Braisted, A.; Wells, J. A. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5688–5692.
34. Solazzo, M.; Bianchi, E.; Felici, F.; Cortese, R.; Pessi, A. *Combinatorial libraries. Synthesis, screening and application potential*; Cortese, R. Ed.; Walter de Gruyter: New York, 1996, pp 127–143.
35. Chen, S.; Chrusciel, R. A.; Nakanishi, H.; Raktabutr, A.; Johnson, M. E.; Sato, A.; Weiner, D.; Hoxie, J.; Saragovi, H. U.; Greene, M. I.; Khan, M. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5872–5876.

36. Moore, J. P.; Sweet, R. W. *Perspectives Drug Discovery Design* **1993**, *1*, 235–250.
37. King, D. S.; Fields, C. G.; Fields, G. B. *Int. J. Peptide Protein Res.* **1990**, *36*, 255–265.
38. Wütrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
39. Koradi, R.; Billeter, M.; Wütrich, K. *J. Mol. Graphics* **1996**, *14*, 51–55.