the FEBS Journal

Grafting of functional motifs onto protein scaffolds identified by PDB screening – an efficient route to design optimizable protein binders

Rym Tlatli, Hervé Nozach, Guillaume Collet, Fabrice Beau, Laura Vera, Enrico Stura, Vincent Dive and Philippe Cuniasse

Service d'Ingénierie Moléculaire des Protéines, Institut de Biologie et Technologies de Saclay (IBITEC-S), Commissariat à l'Energie Atomique, Gif-sur-Yvette, France

Keywords
functional motif; grafting; MMP inhibitor; protein binder design; protein–protein interactions

Correspondence
P. Cuniasse, Service d'Ingénierie Moléculaire des Protéines, Institut de Biologie et Technologies de Saclay (IBITEC-S), Commissariat à l'Energie Atomique, F-91191 Gif-sur-Yvette, France
Fax: +33 1 69 08 90 71
Tel: +33 1 69 08 56 35
E-mail: philippe.cuniasse@cea.fr

(Received 21 August 2012, revised 19 October 2012, accepted 30 October 2012)

doi:10.1111/febs.12056

Artificial miniproteins that are able to target catalytic sites of matrix metalloproteinases (MMPs) were designed using a functional motif-grafting approach. The motif corresponded to the four N-terminal residues of TIMP-2, a broad-spectrum protein inhibitor of MMPs. Scaffolds that are able to reproduce the functional topology of this motif were obtained by exhaustive screening of the Protein Data Bank (PDB) using STAMPS software (search for three-dimensional atom motifs in protein structures). Ten artificial protein binders were produced. The designed proteins bind catalytic sites of MMPs with affinities ranging from 450 nM to 450 μM prior to optimization. The crystal structure of one artificial binder in complex with the catalytic domain of MMP-12 showed that the inter-molecular interactions established by the functional motif in the artificial binder corresponded to those found in the MMP-14–TIMP-2 complex, albeit with some differences in geometry. Molecular dynamics simulations of the ten binders in complex with MMP-14 suggested that these scaffolds may allow partial reproduction of native inter-molecular interactions, but differences in geometry and stability may contribute to the lower affinity of the artificial protein binders compared to the natural protein binder. Nevertheless, these results show that the *in silico* design method used provides sets of protein binders that target a specific binding site with a good rate of success. This approach may constitute the first step of an efficient hybrid computational/experimental approach to protein binder design.

Introduction

In recent years, considerable efforts have been made to design alternatives to antibodies for development of therapeutic or diagnostic compounds. Approaches based on production and screening of large libraries involving randomization of restricted regions in moderate size protein scaffolds with a stable core were proposed. Impressive successes were obtained by these approaches [1], and some of these proteins are currently being investigated for therapeutic applications [2], demonstrating the potential of moderate-size proteins in this context. However, these approaches do not offer precise control of the binding location of the protein binder at the surface of the target or its binding orientation. This may be an important drawback in the context of drug development, where specific interactions are mandatory.

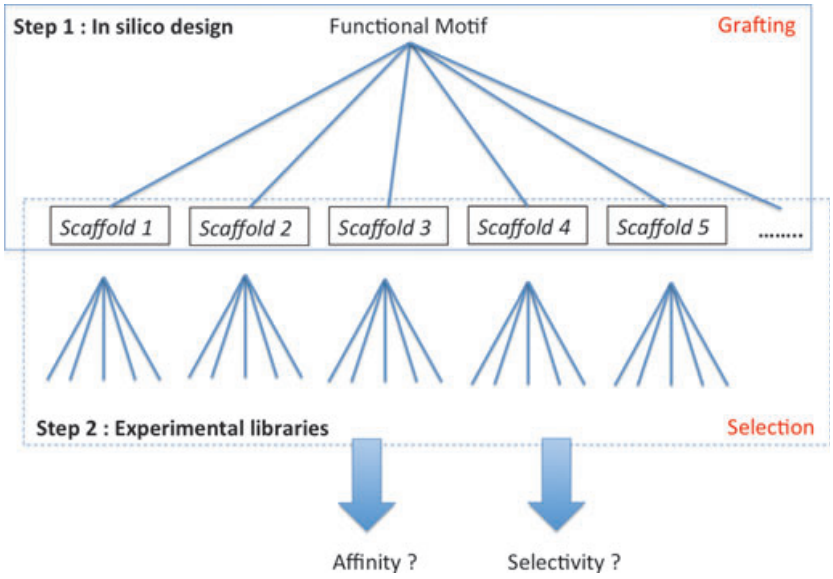
Abbreviations
MMP, matrix metalloproteinase; rmsd, root mean square deviation; TIMP, tissue inhibitor of metalloproteinases.

In silico functional motif grafting methods have been proposed as an alternative to generic scaffolds for structure-based design of artificial protein binders. These approaches are based on the observation that a limited set of residues, referred to as hot spots, located at the surface of a protein binder, contribute predominantly to the binding free energy [3]. Artificial protein binders may be obtained by grafting such hot spots on protein scaffolds that are able to reproduce the functional topology of these residues. In contrast to library-based approaches, grafting methods allow the design of protein binders that interact at a specific location because they are based on reproduction of precise interactions. In early studies, the grafted motifs were restricted to residues belonging to a single structural element that were reproduced on a different protein [4–9]. Although important successes were obtained using this principle, its potential is clearly limited by the availability of scaffolds. In addition, functional epitopes are often discontinuous, and the residues that constitute the motif may be located in separated structural elements. To overcome these limitations, the backbone grafting approach was recently proposed [10,11]. In this method, linear epitopes are transplanted onto a protein scaffold by replacing portions of native backbone by the linear segments to be grafted. This rather aggressive method may result in an important proportion of non-viable proteins. In addition, due to limited scaffold availability, this approach will also be restricted to simultaneous transfer of a very small number of linear epitopes.

A secondary structure-independent approach for side chain grafting may offer an interesting alternative to the above methods [12,13]. It is a generalization of secondary structure-driven side-chain grafting, based on the observation that proteins that possess radically different folds may bind structurally similar protein interfaces [3,14,15]. Furthermore, it was shown that structurally dissimilar proteins that interact at one binding site used the same contact residues at the surface of the target [3]. These observations led to the design of artificial protein binders by grafting a functional motif onto a scaffold possessing a 3D structure unrelated to that of the natural protein binder but capable of reproducing the functional topology of the motif. In this case, scaffold identification is achieved by detecting a set of residues in a protein structure that adopt a topology similar to that adopted by the selected motif in the natural protein binder, but not necessarily in the same structural context. This may be achieved if only Ca and Cb coordinates of the residues are taken into account for the scaffold search. This principle also renders the screening independent of the type of amino acids that are natively present on the scaffolds examined. To fully exploit the potential of the method, systematic examination of the complete Protein Data Bank (PDB; more than 75 000 protein structures in April 2012) rather than a limited subset of pre-selected structures is required. For this purpose, software capable of analysing protein structures at the atomic scale and of detecting sets of Ca, Cb atoms in a defined topology were developed [16,17]. It should be noted that this search principle, implemented in STAMPS software (search for three-dimensional atom motifs in protein structures) [16], allows identification of scaffolds for continuous and discontinuous epitopes. We previously used this approach and STAMPS to design three miniproteins that are capable of binding the Kv1.2 channel with affinity between 500 nM and 1.6 μM [13], without further optimization. The grafting of a discontinuous epitope of erythropoietin (EPO) at the surface of the rat pleckstrin homology domain of phospholipase Cδ1 led to an artificial erythropoietin receptor (EPOr) binder with binding affinity of 24 nM after optimization [18]. However, applications of this approach are still scarce, and the available data do not yet permit evaluation of its robustness.

The secondary structure-independent *in silico* design approach may constitute the first step of a general hybrid computational/experimental strategy (Scheme 1), in which initial protein binders showing measurable albeit weak affinities for the target are obtained (step 1). In a second step, beyond the scope of the present paper, focused libraries of these protein binders may be designed to select miniproteins with the desired affinity/selectivity (step 2). The coupling of *in silico* approaches with powerful experimental methods for testing and optimizing the initial protein binder candidates has already resulted in important successes in the field of protein binder design [10,19]. However, these studies also suggested that the limiting step of this hybrid approach remains discovery of the initial binders. The global success rate of the design approach may be significantly improved if the *in silico* design step were able to produce a large set of initial protein binders rather than a single one.

The goal of the present paper is to evaluate the capacity of the secondary structure-independent grafting approach to respond to the challenge of designing a set of initial protein binders that target a specific binding site (step 1). As a test case, we applied it to the design of miniproteins targeting the catalytic site of matrix metalloproteinases (MMPs). MMPs are involved in many pathological conditions in humans, such as cancer, arthritis and cardiovascular diseases [20,21]. This family contains more than 20 members characterized by high structural similarity of their



Scheme 1. General strategy for protein binder design.

catalytic domains. Artificial miniprotein binders may offer an interesting alternative to synthetic inhibitors [22–26] for obtaining compounds with high affinity/selectivity. Catalytic domains of MMPs are characterized by a high degree of similarity of their sequence and 3D structures [27,28]. Their action is regulated by TIMPs (tissue inhibitors of metalloproteinases), endogenous potent inhibitors proteins with sub-nanomolar affinities for MMPs [29]. Mutagenesis [30–33] and structural studies [34–37] revealed the particular role of the first four residues of TIMPs that interact in the S1–S3′ region of the catalytic domain of MMPs. The structural studies also confirmed interaction of the α-amino and carbonyl groups of residue in position 1 of TIMPs with the catalytic Zn of MMPs, as previously evidenced via chemical modification of the N-terminal residue in TIMP-2 [38] and TIMP-4 [39]. Thus, these four residues were used as the motif for grafting onto miniprotein scaffolds to design miniprotein MMP inhibitors. The scaffolds were identified by systematic screening of the PDB using STAMPS [16]. Ten protein binder candidates were produced and evaluated as MMP inhibitor. We also sought to evaluate whether these scaffolds permitted reproduction of the interactions established by the natural protein binder with the targets.

Results

Motif to be grafted

The atom coordinates of residues C1, S2, C3 and S4 of TIMP-2 selected as the functional motif for grafting were taken from the 3D structure of the catalytic domain of MMP-14 in interaction with TIMP-2 (PDB code 1BQQ) [35]. In this structure, residues 2–4 exhibit a series of hydrogen bonds between main-chain donor and acceptor groups and residues of the catalytic domain of the enzyme (Fig. 1). In the present work, this structure was considered as the prototype of MMP–TIMP interaction.

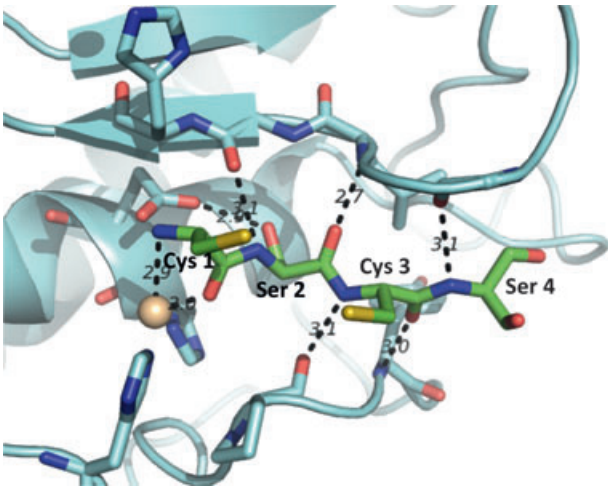


Fig. 1. Inter-molecular interactions between the four N-terminal residues of TIMP-2 (shown as sticks) and residues of the catalytic domain of MMP-14 (also shown as sticks) (PDB code 1BQQ). The catalytic zinc atom of the enzyme is shown in orange (sphere representation). Residues are coloured by atom type (N, blue; O, red; S, yellow; C, green for TIMP-2 residues; cyan for the catalytic domain of MMP-14).

Scaffold search with STAMPS

Scaffold searches were performed using the developmental version of STAMPS version 3.1.1. Several constraints were imposed during these searches. The residue equivalent to C1 in the scaffold must be located in the N-terminal region of the scaffolds to be able to reproduce a similar interaction as observed with the catalytic Zn atom of MMPs in TIMP complexes. Taking into account the limited size of the motif (four residues) and its widespread structure, the number of solutions may be very large. This was confirmed by preliminary STAMPS searches performed without restriction on the type of the equivalent residues (data not shown). To limit the number of solutions, two searches were performed using restraints on the native residues present in the scaffolds. In search 1 (S_Cys1), the native residue in the position equivalent to C1 in the scaffold must be a cysteine, the other - residues (2–4) being one of the 20 amino acids. In search 2 (S_Cys3), we restricted the selected scaffolds to those naturally possessing a cysteine in the position equivalent to C3 of the motif, with residues 1, 2 and 4 being one of the 20 amino acids. Other parameters were included to limit the number of scaffolds identified by STAMPS. We searched for the equivalent motif in the first five residues of the proteins, with the possibility of deleting one residue in the N-terminal position of the native sequence of the scaffold. The size of the scaffolds was limited to a range of 15–60 residues, and the maximum root mean square deviation (rmsd) of the Ca–Cb motifs between the topology observed in the complex 1BQQ and that of the selected scaffold was set to 1.25 Å. Searches S_Cys1 and S_Cys3 resulted in 21 and 198 solutions, respectively. As STAMPS exhaustively examines all PDB files, the solutions identified include some redundancies (Table S1). Figure 2A shows that the sizes of the scaffolds identified are distributed between 15 and 60 residues. The topological similarity between the searched motif and those identified in these scaffolds is high, as shown by the rmsd distribution for the 219 solutions (Fig. 2B). A similar conclusion may be drawn from the distribution of the mean Ca–Cb angle after superimposition on the TIMP-2 motif (Fig. 2C). Furthermore, the distribution of steric factors for the 219 solutions in the binding mode required to reproduce the position of the functional motif in TIMP-2 relative to the cat-

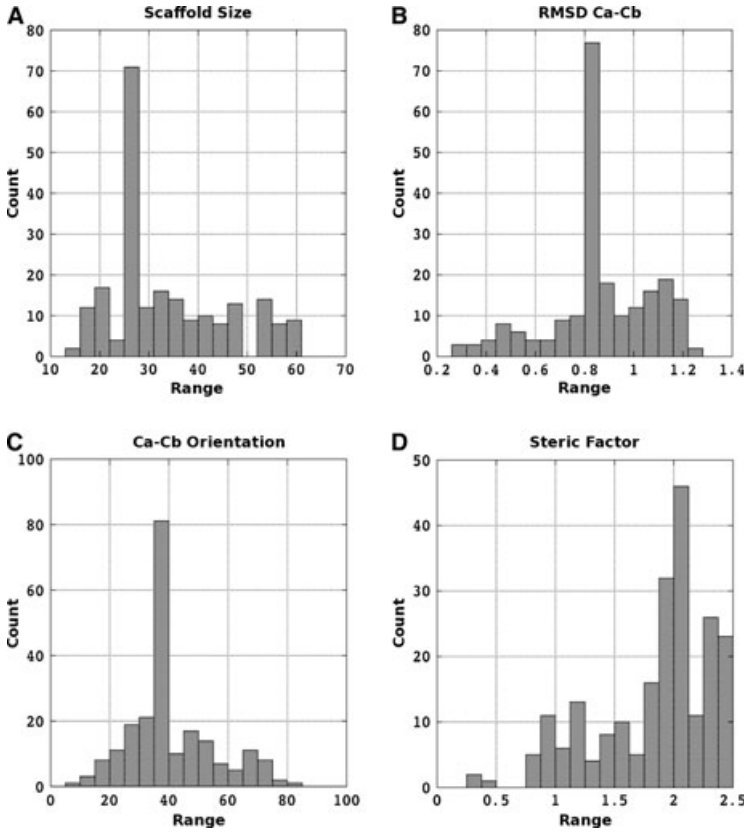


Fig. 2. Global analysis of the scaffolds identified by STAMPS. (A) Distribution of the protein scaffold sizes. (B) rmsd of the positions of the Ca,Cb atoms after superimposition on the corresponding atoms in the functional motif. (C) Geometric deviation measured on the basis of the mean value of the Ca–Cb angles between the functional motif and the corresponding vectors in the considered scaffold. (D) Geometric deviation measured on the basis of the inter-molecular steric factor.

alytic domain of MMP-14 shows satisfactory steric compatibility with the target (Fig. 2D). Visual examination of the 219 complexes identified by STAMPS was performed to check for possible minor steric hindrance not detected by the crude approach implemented in STAMPS and to eliminate scaffolds in which the topologically equivalent motif was located in an unstructured region. We finally selected a set of ten scaffolds that were further investigated (Table 1). Their 1D structures and those of the mutated forms are shown in Fig. 3. The 3D structures of the ten selected scaffolds are shown in Fig. 4. This illustrates well the diversity of the secondary/tertiary structures of these scaffolds and the structural similarity of the functional motifs in the various contexts.

Electrostatic and shape similarity between the protein binder candidates and the natural binder TIMP-2

We compared the shapes of the ten protein binder candidates with that of the natural inhibitor TIMP-2 in the region involved in interaction with the MMP using PIPSA software [40]. Shape similarity is quantified by calculation of the Hodgkin shape similarity index (SSI_H), which varies from 0 in the absence of shape similarity to 1 when the shapes are identical. Figure 5A shows the values obtained for the ten protein binder candidates. The shape similarity in the binding region between the natural ligand TIMP-2 and the ten scaffolds selected is satisfactory, as SSI_H_part ranges from 0.56 to 0.75.

Electrostatic potential plays an important role in binding affinity. This long-range interaction is determined by the full charge distribution of the protein scaffold. As no restriction on the electrostatic potential was applied during the scaffold search, it is important to check for electrostatic similarity between the selected scaffolds and the natural binder. We calculated the electrostatic potential for the ten scaffolds using the Poisson–Boltzmann equation in APBS software [41]. The structures used were those derived from STAMPS that included mutations of the motif residues when necessary (see below). The resulting electrostatic potential grids were compared with that calculated for the natural inhibitor TIMP-2 in a region restricted to the binding region with the target (see Experimental procedures). The results indicate that most scaffolds exhibit a similar electrostatic potential in their binding region, with the exception of 1EDP_A and 2F3C_I_m, which showed uncorrelated potentials compared to the natural binder (Fig. 5B). However, such a situation also occurs in some cases in natural binders (P. Cuniasse, unpublished data). As none of the ten scaffolds

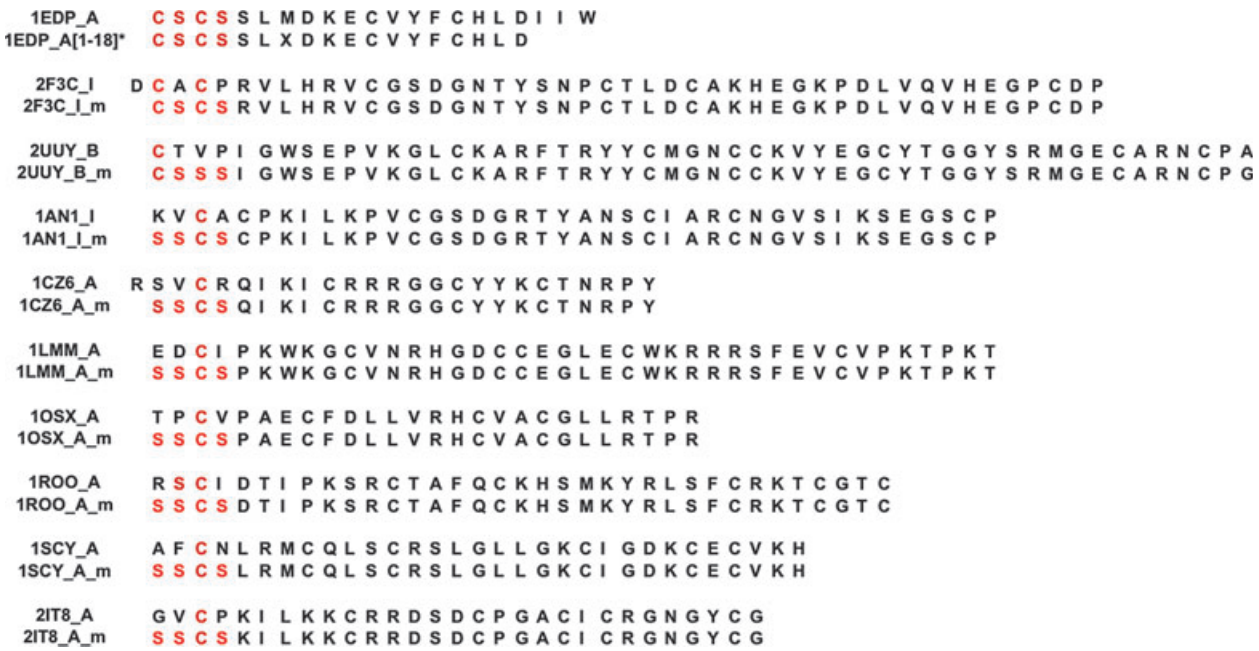


Fig. 3. Amino acid sequences of the ten protein binders selected for molecular modelling and experimental evaluation. The residues of the functional motif are shown in red. For each scaffold, the first line indicates the amino acid sequence found in the PDB file and the second line indicates the sequence of the protein produced. X, norleucine.

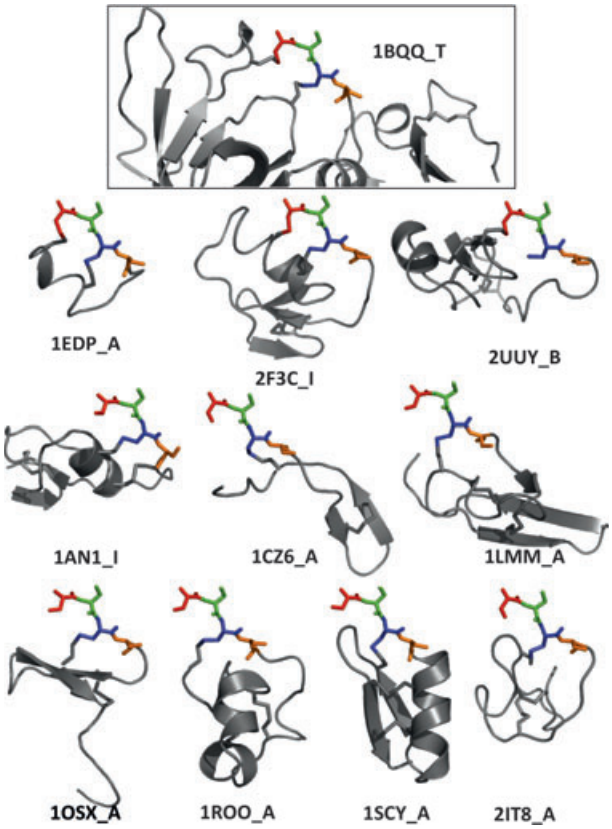


Fig. 4. 3D structures of the scaffolds identified by STAMPS via screening of the PDB and selected for further evaluation. The inset at the top of the figure shows the structure of TIMP-2 in its complex with MMP-14 (PDB code 1BQQ). The structures are shown in cartoon representation in grey. Native disulfides and the residues constituting the grafted motif are indicated by stick representation. Hydrogen atoms are omitted.

displayed anti-correlated electrostatic potentials in the binding region with respect to TIMP-2, they were all retained for further evaluation.

Molecular dynamics refinement of the selected candidates

The ten initial complexes identified by STAMPS were refined by energy minimization and restrained molecular dynamics *in vacuo*. The resulting refined structures were analysed, and a series of parameters characterizing protein–protein interactions were determined. For each complex, positional restraints were applied to all non-hydrogen atoms of the catalytic domain of MMP-14 to preserve its 3D structure and to the four residues of the protein binder designs to maintain their position with respect to the catalytic domain of MMP-14. A second type of restraint was applied to the scaffolds during refinement to preserve their overall 3D structure while allowing global motion of the scaffold with respect to the target. This allowed optimization of surface complementarity of the two partners. These restraints correspond to a set of distances joining the Ca of each residue to the Ca atoms of all other residues in the scaffold. This distance matrix was automatically generated at the start of the refinement protocol, and the reference distances were taken from the coordinates of the X-ray/NMR structure of the scaffold obtained from the PDB (see Experimental procedures). The initial step of the refinement protocol consisted of an energy minimization to correct for possible unfavourable local steric or electrostatic interactions in the starting complex. This step is followed by 15 cycles of molecular dynamics (25 000 steps each). At the end of each cycle, the structure was subjected to energy minimization under positional restraints. The resulting structure was saved for further analysis. Parameters characterizing these protein complexes were determined (Table 2). The total variation in accessible surface area upon binding of the complex (ΔASA_bind) ranges from 1086 to 2182 $Å^2$ for 1OSX_A_m and 1AN1_I_m, respectively. With the

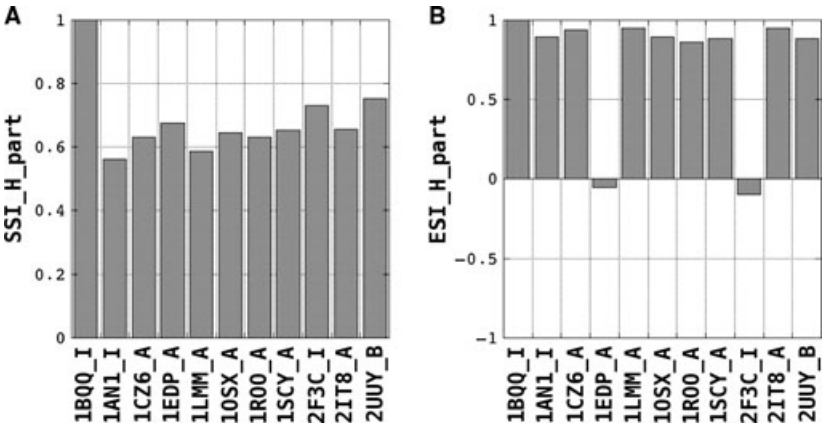


Fig. 5. Shape (A) and electrostatic (B) Hodgkin similarity indices calculated between the reference ligand TIMP-2 and the ten selected protein binder candidates in the region involved in interaction with the target MMP-14. SSI_H_part, Hodgkin shape similarity index; ESI_H_part, Hodgkin electrostatic similarity index.

Table 1. Scaffolds identified by STAMPS after exhaustive screening of the PDB, and selected for further *in silico* and experimental evaluation.

Search	Name	Protein size	Motif size	rmsd[a]	Orientation	Steric (inter/intra)	Residue 1		Residue 2		Residue 3		Residue 4	
S_Cys1	1EDP_A	17	4	0.61	21.39	2.12/0.11	Cys	1	Ser	2	CYS	3	SER	4
	2F3C_I_m	46	4	0.48	16.76	1.18/0.30	Cys	6	Ala	7	CYS	8	PRO	9
	2UUY_B_m	52	4	0.36	9.71	1.05/0.25	Cys	24	Thr	25	VAL	26	PRO	27
S_Cys3	1AN1_I_m	40	4	1.05	47.38	2.27/0.28	Lys	2	Val	3	CYS	4	ALA	5
	1CZ6_A_m	25	4	0.54	17.80	1.76/0.17	Ser	2	Val	3	CYS	4	ARG	5
	1EDP_A_m	17	4	0.61	21.39	2.12/0.11	Cys	1	Ser	2	CYS	3	SER	4
	1LMM_A_m	40	4	1.13	53.52	1.97/0.24	Glu	1	Asp	2	CYS	3	ILE	4
	1OSX_A_m	26	4	1.13	51.76	1.21/0.23	Thr	17	Pro	18	CYS	19	VAL	20
	1ROO_A_m	35	4	1.21	51.94	0.80/0.13	Arg	1	Ser	2	CYS	3	ILE	4
	1SCY_A_m	31	4	1.14	48.18	0.49/0.16	Ala	1	Phe	2	CYS	3	ASN	4
	2F3C_I_m	46	4	0.48	16.76	1.18/0.30	Cys	6	Ala	7	CYS	8	PRO	9
	2IT8_A_m	29	4	0.71	16.40	1.17/0.30	Gly	1	Val	2	CYS	3	PRO	4

[a] Calculated based on the position of Ca,Cb atoms.

exception of scaffold 1OSX_A_m, the values observed for these scaffolds are in the range reported for natural complexes (1600 ± 400 Å^2) [42]. A similar conclusion may be drawn from the variation in accessible surface area of the protein scaffolds upon binding (ΔASA_interf_lig), which ranges from 592 Å^2 for 1OSX_A_m to 1073 Å^2 for 1AN1_I_m. The number of residues of the protein binder located at the interface with the target varies from 11 to 22. This variability is also illustrated in Fig. 6, in which the binding surfaces of the protein target in the modelled complexes are shown in orange. In addition, the protein binder candidates interact with slightly different regions on the surface of the target. The ten protein binder candidates presented a predominantly polar interface, with the non-polar fraction varying between 8% and 30% (Table 2). The gap volume indices characterizing surface complementarity between the protein binders and the target varied between 1.61 and 2.15 Å, which is in the range observed for natural protein–protein complexes [43]. We also analysed the inter-molecular hydrogen bonds established between the two partners. This number ranges from 5 in the case of 1EDP_A to 13 in the case of 1LMM_A_m (Table 2). When the number of hydrogen bonds were normalized to the interface area, the resulting hydrogen bond surface densities varied between 4.0 and 8.2×10^{-3} Å^{-2}, which is also in the range observed for natural hetero-dimeric protein–protein complexes [43,44]. In addition, salt bridges and cation–π interactions were observed in the models of the MMP-14/designs complexes (Table 2). The above observations show that the selected scaffolds are able to establish a set of inter-molecular interactions, some of which involve residues of the grafted motif and others corresponding to residues located at the interface with the target. In addition, this modelling study showed that the parameters characterizing the ten artificial interfaces are within the ranges observed for natural protein–protein complexes.

Production and evaluation of the ten protein binder candidates

The above results prompted us to produce the ten protein binder candidates in order to evaluate their ability to bind a series of MMPs (MMP-2, MMP-8, MMP-9, MMP-12, MMP-13 and MMP-14). Some of the designed artificial protein binders were produced by solid-phase peptide chemical synthesis, and others were produced by recombinant expression in *Escherichia coli*. The mutations depended on the native composition of each particular scaffold (Fig. 3). In the

Table 2. Characterization of refined MMP-14–protein binder candidate complexes. ΔASA_bind, total variation of accessible surface area upon binding; ΔASA_interf_lig, Protein–binder interface area; Nb_Res_Inter_lig, number of residue on the protein-binder side of the interface;%Res_NP_lig, proportion of apolar residues on the protein-binder side of the interface;% Surf_NP_lig, proportion of atoms belonging to apolar groups on the protein binder side of the interface; Gap_Vol_In_ligand, gap volume index; Nb_interprot_HB, number of inter-molecular hydrogen bonds; Nb_Salt_Bridges, number of inter-molecular salt bridges; Nb_CP, number of inter-molecular cation–π interactions; Interprot_HB_SD, inter-molecular hydrogen bond surface density (number of hydrogen bonds/total buried accessible surface area upon binding).

Scaffold	1BQQ_T	1EDP_A	2F3C_I_m	2UUY_B_m	1AN1_I_m	1CZ6_A_m	1LMM_A_m	1OSX_A_m	1ROO_A_m	1SCY_A_m	2IT8_A_m
ΔASA_bind (Å^2)	2966	1230	1514	2029	2182	1561	1821	1086	1204	1445	1370
ΔASA_interf_lig (Å^2)	1552	664	826	1043	1073	891	917	592	672	771	786
Nb_Res_Inter_lig	36	12	15	22	20	14	19	11	12	15	13
%Surf_NP_lig	42	22	8	30	27	16	21	17	15	15	18
Gap_Vol_In_ligand (Å)	1.92	1.72	1.84	1.87	2.15	1.62	1.61	1.84	1.67	1.91	1.70
Nb_interprot_HB	17	5	13	12	13	12	15	6	6	11	9
Nb_Salt_Bridges	0	0	0	1	1	1	1	0	1	2	0
Nb_CP	2	0	0	1	1	1	0	2	1	0	1
Interprot_HB_SD (10^{-3} Å^{-2})	5.7	4.0	6.6	5.9	5.9	7.7	8.2	5.5	4.9	7.3	6.5

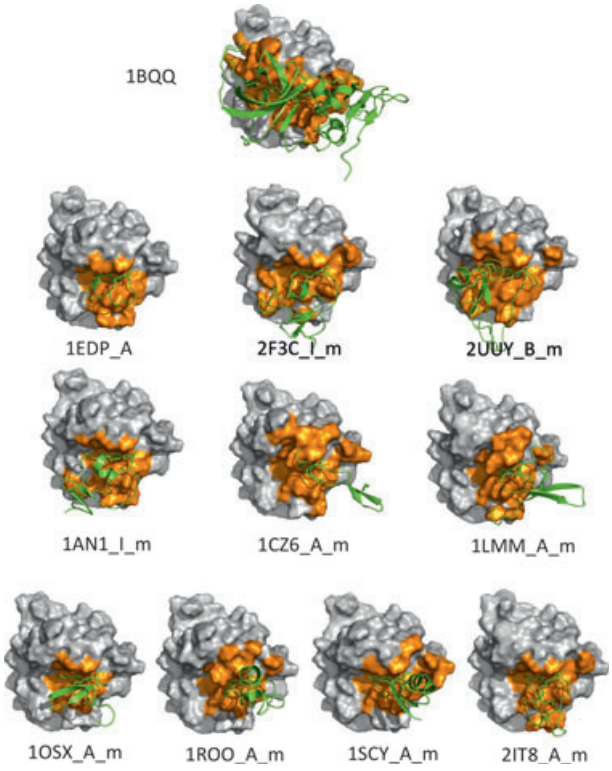


Fig. 6. Graphical representation of the complex between the catalytic domain of MMP-14 and the natural inhibitor TIMP-2 or the designed protein binders. The catalytic domain of MMP-14 is shown in surface represented in grey. Protein ligands are shown in cartoon representation in green. The MMP-14 interface (residues at a distance < 3.5 Å) is coloured in orange.

case of endothelin-1 (1EDP_A), no mutation was necessary as the four native residues identified by STAMPS in this scaffold had the CSCS sequence. In the case of scaffold 2F3C_I_m, the residue in position 1 of the native scaffold was deleted. This left the native cysteine in position 2 of the scaffold as the N-terminal residue in 2F3C_I_m. The residues in positions 3 and 5 of the native sequence (corresponding to positions 2 and 4 of the motif) were mutated to serine. Similarly to 2F3C_I_m, residue R1 present in the native structure of 1CZ6_A was deleted, making native residue C4 residue C3 of the motif. In the other scaffolds, cysteines present in positions 1 and 3 in the native structure were preserved. When the residue in position 1 or 3 in the native scaffold was not a cysteine, it was replaced by a serine, a residue that is sterically similar to cysteine. This strategy avoids unwanted effects of the presence of a free cysteine during the protein production step. In all cases, residues corresponding to positions 2 and 4 of the motif were replaced by serine.

Table 3. Binding affinity for MMPs (K_i values in μM; mean ± standard error) of the ten artificial protein binders. ND, not determined.

Protein binders	Scaffold name	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1EDP_A	Endothelin-1	7 ± 1	2 ± 0.2	7 ± 0.9	1 ± 0.1	0.45 ± 0.03	ND
2F3C_I_m	Thrombin inhibitor infestin	32 ± 3	6 ± 1	36 ± 5	31 ± 2	73 ± 7	100 ± 15
2F3C_I[a]	Thrombin inhibitor infestin	56 ± 6	56 ± 19	51 ± 5	77 ± 7	143 ± 17	217 ± 39
2UUY_B_m	Tick-derived tryptase inhibitor	69 ± 8	24 ± 3	39 ± 6	160 ± 12	158 ± 20	ND
2UUY_B[a]	Tick-derived tryptase inhibitor	10 ± 1	4 ± 0.2	20 ± 2	8 ± 0.9	9 ± 0.7	129 ± 19
1AN1_I_m	Leech-derived tryptase inhibitor	22 ± 2	5 ± 0.7	30 ± 4	136 ± 10	38 ± 7	132 ± 11
1CZ6_A_m	Androctonin	92 ± 17	40 ± 6	45 ± 6	>500	183 ± 18	213 ± 15
1LMM_A_m	Psalmotoxin-1	18 ± 1	11 ± 3	32 ± 3	34 ± 1	68 ± 5	24 ± 3
1OSX_A_m	Extracellular domain of BLyS R3	97 ± 13	54 ± 9	58 ± 10	197 ± 12	206 ± 29	>500
1ROO_A_m	SHK toxin	182 ± 21	54 ± 8	102 ± 14	457 ± 35	381 ± 44	>500
1SCY_A_m	Scyllatoxin	150 ± 15	75 ± 5	130 ± 13	160 ± 47	440 ± 46	ND
2IT8_A_m	Trypsin inhibitor 2	173 ± 15	70 ± 8	157 ± 20	>500	>500	>500

[a] Native N-terminal sequence.

Table 4. Binding affinity for MMPs (K_i values in μM; mean ± standard error) of the analogues of 1EDP_A and 1AN1_I. ND, not determined; hF, homo-phenyl alanine; FF, a pseudo residue with a methyl-biphenyl moiety.

Protein binder	Scaffold name/mutation	MMP-2	MMP-8	MMP-9	MMP-12	MMP-13	MMP-14
1EDP_A	Endothelin-1	7 ± 1	2.0 ± 0.2	7.0 ± 0.9	1.0 ± 0.1	0.45 ± 0.03	ND
1EDP_[1-18][a]		20 ± 3	4.0 ± 0.3	8.0 ± 1.0	6.0 ± 0.3	4.0 ± 0.3	ND
1EDP_[1-18]S2hF[a]	S2hF	0.6 ± 0.4	0.7 ± 0.6	0.4 ± 0.3	0.2 ± 0.2	0.3 ± 0.2	100 ± 12
1EDP_[1-18]S2FF[a]	S2FF	1.0 ± 0.4	1.0 ± 0.5	0.7 ± 0.3	0.6 ± 0.4	0.4 ± 0.3	ND
1AN1_I	Leech-derived tryptase inhibitor	22 ± 2	5 ± 1	30 ± 4	136 ± 10	38 ± 7	132 ± 11
1AN1_I_m1	(K2S, V3I, A5S, L10Y)	26 ± 4	16 ± 2	24 ± 3	85 ± 6	64 ± 7	51 ± 6
1AN1_I_m2	(K2S, V3I, A5S, A22M)	19 ± 2	18 ± 2	33 ± 5	122 ± 100	169 ± 43	70 ± 3
1AN1_I_m3	(K2S, V3I, A5S, K8W)	17 ± 1	7 ± 0.2	31 ± 4	85 ± 3	42 ± 4	93 ± 6
1AN1_I_m4	(K2S, V3I, A5S, K8W, L10Y, A22M)	16 ± 2	3.0 ± 0.3	20 ± 3	71 ± 4	23 ± 2	99 ± 8
1AN1_I_m5	(K2S, V3M, A5S, K8W, L10Y, A22M)	13 ± 1	2.0 ± 0.1	20 ± 3	36 ± 4	11 ± 1	59 ± 2

[a] Residue M7 was replaced by a norleucine.

As the functional motif is taken from the natural broad-spectrum inhibitor of MMPs, TIMP-2, the protein binders are expected to interact with all MMPs. The binding of the ten protein binder candidates to a series of six MMPs was thus evaluated by determining their ability to inhibit the cleavage of a synthetic fluorogenic substrate (see Experimental procedures). The corresponding inhibition constants (K_i) are reported in Table 3. In the case of MMP-2, MMP-8 and MMP-9, all the designed binders were able to inhibit the catalytic activity of these metalloproteinases, with K_i values ranging from 2 to 182 μM. In the case of MMP-12 and MMP-13, 1EDP_A, 2F3C_I_m, 2UUY_B_m, 1AN1_I_m and 1LMM_A_m displayed K_i values ranging from 450 nM to 158 μM. 1CZ6_A_m showed a K_i value of 183 μM for MMP-13, and no inhibition of catalytic activity was observed for MMP-12 at 500 μM. 1ROO_A_m displayed weak inhibition of MMP-12 and MMP-13 (K_i > 300 μM), and 2IT8_A_m displayed no inhibition at 500 μM for these two enzymes. Only 2F3C_I_m, 1AN1_I_m, 1CZ6_A_m and 1LMM_A_m were able to inhibit MMP-14, with K_i values ranging from 24 to 213 μM. In the case of 2F3C_I and 2UUY_B, we also evaluated binding of the native scaffolds (Table 3). 2F3C_I_m displayed binding affinities that were slightly improved by factors of 2–10 for the MMPs tested, compared to the native sequence. In contrast, 2UUY_B_m showed lower binding affinity for these MMPs compared to 2UUY_B, by factors ranging from 5 to 17. In this study, the interactions between the grafted motif and the target involved main-chain chemical functions. It is therefore expected that native scaffolds reproduce these interactions. However, the contrasting results obtained with 2UUY_B_m and 2F3C_I_m as compared to their native analogues show that the consequences of mutating the native residues of the scaffold to graft the functional motif may depend on the particular context of each scaffold.

Analogues of some protein binders

1EDP_A and 1AN1_I_m were selected for optimization. We produced four analogues of 1EDP_A. Exami-

Table 5. Inter-atomic distances (Å) between the CSCS motif in the artificial binder 1EDP_A and the catalytic domain of MMP-12 deduced from the X-ray structure of this complex. The distances (Å) observed for the CSCS motif in the X-ray structure of the natural binder TIMP-2 in complex with MMP-14 (PDB code 1BQQ) are also shown.

	Residue name	Residue number	Atom		Residue name	Residue number	Atom	1BQQ	1EDP_A
Binder	Cys	1	N	MMP	Zn2	289	ZN	2.9	2.0
Binder	Cys	1	O	MMP	Zn2	289	ZN	2.0	2.1
Binder	Cys	1	N	MMP	Glu	240/127	OE1	2.7	3.7
Binder	Ser	2	N	MMP	Ala	200/87	O	3.1	3.5
Binder	Ser	2	O	MMP	Ala	200/87	N	3.8	3.7
Binder	Ser	2	OG	MMP	Glu	240/127	OE1	2.9	4.5
Binder	Ser	2	O	MMP	Leu	199/86	N	2.7	2.7
Binder	Cys	3	N	MMP	Pro	259/146	O	3.1	3.7
Binder	Cys	3	O	MMP	Tyr	261/148	N	3.0	3.6
Binder	Ser	4	N	MMP	Gly	197/84	O	3.1	3.0

nation of the model of this compound in interaction with the catalytic domain of MMP-14 suggested that the last three residues may be removed. We thus synthesized 1EDP_[1-18]* and two analogues in which residue S2 was replaced by non-natural amino acids possessing bulky aromatic side chains. This strategy is based on the specific structure of the S1′ sub-site of most MMPs, which is able to accommodate long non-natural side chains. This has influenced the development of synthetic inhibitors of MMPs and generally led to improved binding affinity compared to natural side chains [23]. The 1EDP analogues were produced by solid-phase peptide synthesis. Table 4 shows their binding affinity for six MMPs. As expected, the presence of homo-phenyl alanine in position 2 led to improved binding affinities by factors ranging from 5 for MMP-8 to 32 for MMP-2 compared to a serine in this position (1EDP_[1-18]*). When the side chain of S2 was replaced by a methyl-biphenyl moiety (compound 1EDP_[1-18]S2FF), we also observed improved affinities for the targets by factors up to 20 compared to 1EDP_[1-18]*. These observations are consistent with a binding mode whereby the residue in position 2 in the 1EDP_A analogue occupies the S1′ sub-site of the MMPs tested. We observed binding constants of up to 290 nM for MMP-12 (compound 1EDP_[1-18] S2hF*, Table 4). We also produced a series of analogues of 1AN1_I_m in which the residues in positions 2, 3, 5, 8, 10 and 22 (Table 4) were substituted. These mutations were suggested by the FOLDX software [45] based on a modelled structure of 1AN1_I_m in interaction with MMP-8. This complex was selected as 1AN1_I_m displayed highest affinity for the catalytic domain of MMP-8. However, these substitutions only led to modest improvements in affinity for some MMPs by factors up to $\sim$ 4, and the best binder displayed a K_i value of 2 μM for MMP-8 (Table 4).

Structural characterization of 1EDP_A in complex with MMPs

The above results strongly suggested that the designed binders interact with the MMPs in the expected binding mode. In order to characterize the binding mode of these artificial binders and determine to what extent they are able to reproduce the interactions established by the natural binder TIMP-2 with MMPs (PDB code 1BQQ), crystallization assays were performed. For this study, we selected 1EDP_A. Soaking experiments performed with 1EDP_A on MMP-12 crystals caused lattice disorder, and data were collected for the one crystal that was still viable. Despite the difficulties with the experiment, the surviving crystal diffracted to a resolution of 2.8 Å, and the data were used to calculate an electron density map in which it is possible to discern unambiguously a peptide fragment interacting in the S1–S3′ region of the enzyme. The electron density is strong for residues 1–4 of 1EDP_A, so the position of the atoms and characterization of their interaction with the catalytic domain of the enzyme may be considered reliable. The electron density corresponding to regions of the scaffold not directly involved in the interaction with MMP-12 is less well defined. The reasons for the weaker electron density may relate to the short length of the soak, partial cleavage of the scaffold and conformational variability of the scaffold within the crystal lattice. Conformational variability is also observable in the various structures of endothelin-1 deposited in the PDB, which display important conformational variability from each other. The sequence that matches the electron density for the four-residue stretch corresponds well to that for 1EDP_A. The presence of stronger electron density for the sulfur atoms that correspond to the cysteines in positions 1 and 3 of 1EDP_A is also reassuring. Residues C1 to S4 adopt a conformation within the

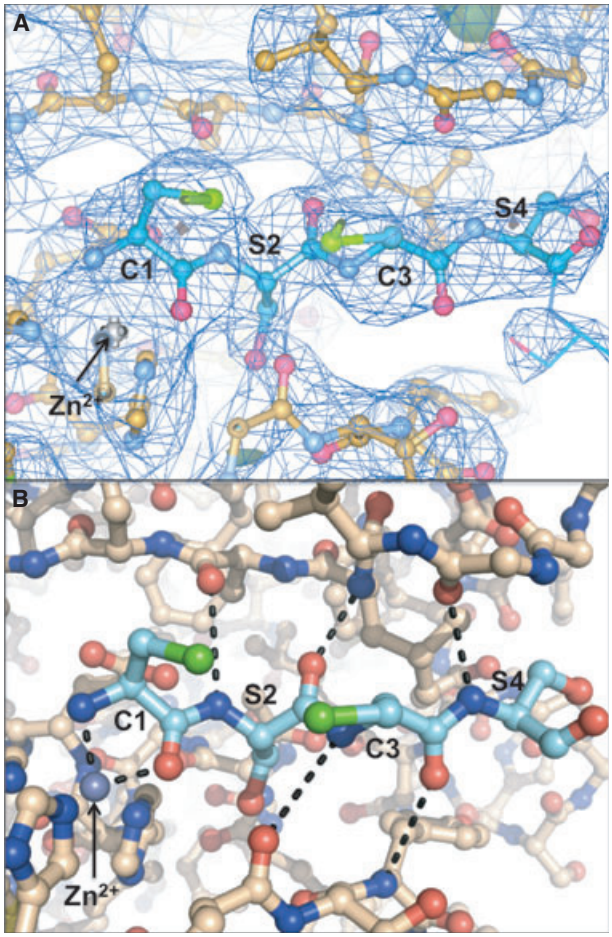


Fig. 7. X-ray structure of 1EDP_A in complex with MMP-12. The structures are shown in ball and stick representation and coloured by element type (N in dark blue, O in red and S in green). The carbon atoms are coloured in orange for the catalytic domain of MMP-12 and in blue for 1EDP_A. (A) The electron density is shown in mesh representation coloured in blue at 0.66 σ. (B) Only residues 1–4 of 1EDP_A are shown. The electron density is omitted and the inter-molecular interactions between residues 1–4 of 1EDP_A and the catalytic domain of MMP-12 are indicated by dashed lines.

active site of MMP-12 that is similar to that observed for the equivalent residues from the natural binder, TIMP-2, in complex with MMP-14 (Fig. 7). The N-terminal nitrogen and carbonyl oxygen of residue C1 interact with the catalytic Zn^{2+} atom, and the corresponding distances are close to that observed in the TIMP-2–MMP-14 complex (1BQQ, Table 4). Several distances between hydrogen bond donors and acceptors belonging to residues 1–4 of 1EDP_A and the catalytic domain of MMP-12 are consistent with the presence of inter-molecular hydrogen bonds (Table 5). Comparison of these distances also showed that some differences exist between the artificial binder 1EDP_A and the natural one. Four distances between the two partners involving hydrogen bond donors or acceptors were longer by 0.6–1.6 Å in the MMP-12–1EDP_A complex compared to the MMP-14–TIMP-2 complex (1BQQ). The distance separating the hydroxyl oxygen of S2 in 1EDP_A and the carboxylate oxygen of residue E240 in the enzyme showe an increase of 1.6 Å. This is due to the different rotamers adopted by residue S2 in the two structures. Finally, four distances in the MMP-12–1EDP_A complex were similar within 0.4 Å to those observed in the MMP-14–TIMP-2 complex.

Determination of the stability of the inter-molecular interactions in the artificial binders by molecular dynamic simulation

To further evaluate the possible differences in geometry and dynamic stability of the interactions consequent to grafting of the functional motif on the various scaffolds, we performed molecular dynamics simulations of the ten binders in complex with MMP-14 in water solution. For each complex, a 1 ns molecular dynamics trajectory was calculated using the complex identified by STAMPS as the starting structure, and further refined by energy minimization and molecular dynamics under positional restraints (see Experimental procedures). Twelve mean distances characterizing interaction of the selected motif with the target were calculated from these simulations (Table S2). We observed excellent agreement between the mean distances determined in simulation of the natural complex TIMP-2–MMP-14 and the corresponding distances observed in the X-ray structure of this complex (1BQQ). Binders 1CZ6_A_m, 1LMM_A_m, 1OSX_A_m, 2F3C_I_m and 2IT8_A_m reproduce these interatomic distances well. According to the simulations, other binders reproduce only partially the distances observed in the natural complex. 1EDP_A displayed nine distances that were similar within 0.5 Å to those found in the natural complex (Table S2). We also noted a good agreement between the distances determined in this simulation and that obtained from the X-ray structure of 1EDP_A in complex with MMP-12. In the case of 1SCY_A_m and 2UUY_B_m, eight distances are well reproduced. Designs 1ROO_A_m and 1AN1_I_m showed seven and six distances, respectively, that were similar within 0.5 Å to the corresponding ones in the TIMP-2–MMP-14 complex. We further examined, along the molecular dynamics trajectories of the ten complexes, the stability of seven interactions that we wished to reproduce. Two of them involve the N-terminal residue of the binder interacting with the catalytic Zn^{2+} ion. Five hydrogen bonds

Table 6. Occupancies of a set of inter-molecular interactions determined in the molecular dynamics simulation of the complexes between the catalytic domain of MMP-14 and the natural ligand TIMP-2 (1BQQ) or the ten artificial protein binders. The atoms involved in these interactions are indicated at the top of each column for the binder and the target MMP-14. Occupancies below 0.8 are shown in bold.

Protein binder atom Protein target atom	N_1 ZN_1	O_1 ZN_1	N_2 ALA_200_O	O_2 ALA_200_N	3_N PRO_259_O	3_O TYR_261_N	4_N GLY_197_O
1BQQ_T	1.00	1.00	**0.62**	1.00	0.99	1.00	1.00
1AN1_I_m	**0.00**	1.00	**0.02**	0.98	0.97	**0.62**	**0.00**
1CZ6_A_m	1.00	1.00	**0.79**	1.00	0.98	0.97	**0.00**
1EDP_A	1.00	1.00	0.96	1.00	**0.72**	1.00	1.00
1LMM_A_m	1.00	1.00	0.96	0.99	0.99	0.96	0.89
1OSX_A_m	1.00	1.00	**0.57**	1.00	0.99	**0.66**	1.00
1ROO_A_m	**0.00**	1.00	**0.00**	**0.67**	0.95	1.00	0.96
1SCY_A_m	**0.00**	1.00	**0.02**	0.91	0.99	1.00	**0.08**
2F3C_I_m	1.00	1.00	0.81	1.00	0.95	**0.27**	0.92
2IT8_A_m	1.00	1.00	**0.35**	**0.57**	0.94	0.99	0.99
2UUY_B_m	1.00	1.00	**0.52**	1.00	**0.64**	**0.47**	1.00

involving donors and acceptors of main-chain residues of the motif were also analysed (Table 6). rmsds for the positions of the protein atoms and the interface area between the partners indicated that equilibrium was reached in all cases in < 500 ps. We thus determined the occupancies (fraction of the time where the interaction is observed on the total time analyzed) of the selected inter-molecular interactions in the last 500 ps of the trajectories. To this end, we used a maximum distance between the acceptor atom and the hydrogen atom bound to the donor of 2.8 Å, and a maximum donor/hydrogen acceptor angle of 120°. In the case of the natural complex TIMP-2–MMP-14, six interactions were stable in the time frame analysed (occupancy of $\sim$ 1). One hydrogen bond between the amide nitrogen of residue S2 and the carbonyl oxygen of residue A200 is formed only about 60% of the time. When considering the stabilities of these interactions in the ten protein binders, we found only three cases in which the interaction between the nitrogen of the residue in position 1 and the Zn^{2+}+ ion was not observed in the time frame (1AN1_I, 1ROO_A and 1SCY_A). In these structures, the hydrogen bond between the amide of the residue in position 2 of the binders and the carbonyl oxygen of residue Ala200 of the enzyme was also absent. Examination of these trajectories indicated rotation of the ψ angle of the residue in position 1 and the ϕ angle of the residue in position 2 of the three binders compared to the starting structure, and that the nitrogen of residue 2 establishes alternative intra-molecular hydrogen bonds (data not shown). It should be noted that these three scaffolds do not possess a cysteine involved in a disulfide bond in position 1 that would have precluded adoption of this conformation. However, this movement is not only due to the presence of a serine in position 1, as this conformation was not observed in scaffolds 1CZ6_A_m, 1LMM_A_m, 1OSX_A_m and 2IT8_A_m. Overall, Table 6 shows that, according to the simulations, some protein binders reproduce only partially the set of hydrogen bonds observed in the natural complex. In addition, several scaffolds compatible with reproduction of the hydrogen bonds observed in the natural complex displayed differences in occupancy compared to this complex. These differences may play a role in the lower affinity observed for the designed protein binders compared to the natural one, but other factors, specific for each scaffold, also contribute to this observation.

Discussion

Recent studies have shown that high-affinity protein binders may be obtained using *in silico* design approaches based on grafting of a functional motif onto protein scaffolds [10,18,19]. These studies clearly demonstrated the role of computational design in defining the framework within which mutagenesis or molecular evolution will enable achievement of the desired affinity [46]. However, these authors point out that, despite the power of the method, directed evolution/selection is unlikely to be successful unless the starting structure already possesses, even at a very low level, the activity that one wishes to select. These studies also suggested that computational design remains the limiting step of these approaches. Obtaining initial protein binders is clearly not an easy task. For instance, in a study reporting the design of protein binders targeting the b12 antibody, only one of 62 proteins designed was shown to possess measurable affinity, albeit at a low level [10]. This low success rate may be due, at least in part, to the rather aggressive

backbone grafting approach used in that study that resulted in a large number of deletions and mutations in the protein scaffolds. This may also explain why only 25 of the 62 protein designed could be expressed in soluble form and purified. In a study describing the design of a protein binder targeting the conserved stem region of the influenza hemagglutinin, the authors grafted hot spots derived from computational docking onto protein scaffolds identified using a protocol that has similarity with the secondary structure-independent approach used in the present paper [19]. Seventy-three of the 88 proteins designed were evaluated by yeast display. However, only two of these proteins displayed measurable but weak affinity for the target. This shows that computational protein binder design is not a solved problem. We still lack a protein design method allowing a high success rate.

Here, we used an *in silico* secondary structure-independent approach based on grafting of a functional motif onto miniprotein scaffolds to design binders targeting the catalytic sites of MMPs. This approach has already shown its ability to create binders that bind in the micromolar to nanomolar range prior to optimization [13,18]. The scaffold search was performed by systematic screening of the PDB using STAMPS software [16]. The search principle implemented in STAMPS allows identification of both continuous and discontinuous functional motifs. The grafted motif corresponded to the continuous N-terminal tetrapeptide of TIMP-2, a natural broad-spectrum protein inhibitor of MMPs. The scaffold search relied on identification of a set of side chains with a topology similar to that of the functional motif. This principle allows us to take advantage of the existing conformational state of the scaffold. It also avoids aggressive protein structure manipulation, with grafting of the functional site being achieved via a limited number of mutations at the surface of the selected scaffold. Here we sought to evaluate the yield of the method in terms of generating a set of active protein binders. To this end, we produced ten of the protein binder candidates obtained by the *in silico* approach. We showed that the ten proteins were able to target most of the MMP catalytic domains tested with affinities of up to 450 nM prior to optimization. This observation is consistent with the fact that the interactions established by the selected motif are expected to be similar with all MMPs because their catalytic domains possess a high structural similarity in the targeted region. Furthermore, none of the additional criteria used for scaffold selection (structural complementarity with the target and electrostatic similarity with the natural broad-spectrum protein inhibitor TIMP-2) permit discrimination between different MMP catalytic domains.

The ten selected scaffolds possess various 3D structures. Endothelin-1 was identified as a suitable scaffold. This protein naturally contains the CSCS N-terminal motif. A previous study showed that a truncated analogue of the related miniprotein sarafotoxin 6b possessing a similar fold to that of endothelin-1 and a CSCS N-terminal motif was capable of inhibiting MMP-1, MMP-2 and MMP-9 with affinities between 1 and 20 μM [47]. In this work, the authors identified sarafotoxin 6b as a suitable scaffold based on its N-terminal 1D sequence homology with TIMPs. It should be noted that we also identified sarafotoxin 6b as a suitable scaffold (pdb1srbA_00, Table S2), but it was not selected for experimental evaluation due to its fold similarity with endothelin-1. In the present work, nine other scaffolds whose 3D structures were unrelated to that of with endothelin-1 were identified by systematic screening of the PDB and selected for experimental evaluation. Scaffold 2F3C_I possess cysteine residues in positions 1 and 3. It displayed a fold similarity with endothelin-1 in the region of the two disulfides. The corresponding protein displayed binding affinities for the MMPs tested ranging from 6 to 100 μM. One scaffold naturally possessing a cysteine residue in the N-terminal position (2UUY_B) gave a protein binder with affinities between 4 and 129 μM for the MMPs tested. All other scaffolds possessed a cysteine residue in position 3 and interact with most MMPs tested, albeit with moderate affinities. Overall, as compared to previously reported studies, the yield of the approach may be considered as good as the ten proteins designed displayed measurable binding affinities for the targets. In addition, all these binders were produced by solid-phase peptide synthesis or recombinant techniques.

In the present work, we also wished to evaluate to which extent the selected scaffold permitted reproduction of the interactions established by the functional motif in the natural complex. This question has been addressed in some studies via structural characterization of the designed proteins in complex with their protein target in the case of secondary structure-driven motif grafting [48] and backbone grafting [10]. These studies revealed a high degree of structural mimicry for compounds displaying high affinity and specificity similar to that of the natural binder. In the present study, we showed high structural mimicry of the functional motif in the context of 1EDP_A, but also found some differences in the precise geometry of the interactions. We could not solve the 3D structure of other binders in interaction with one of the MMPs. Molecular dynamics simulations of the ten designed binders in interaction with the catalytic domain of MMP-14

suggested that some scaffolds allow reproduction of the interactions observed in the natural ligand. However, depending on the scaffolds, some differences in the dynamics of these interactions were observed. In this regard, the conformational control due to residue C1 may be important in the present case. These observations suggest that, in addition to topological considerations guiding scaffold selection, it may be important to take their dynamics in the bound state into account. This will enable selection of scaffolds offering the best conformational stability of the grafted residues.

The present results also illustrate one limit of grafting-based design methods. Any mutation performed to graft the motif may alter the conformational ensemble of the protein scaffold in the free state. This may prevent reproduction of the inter-molecular interactions with the target. This could also lead to an important free energy penalty for returning to the correct conformation at the bound state. In addition, increasing the conformational ensemble of the protein binder in the free state may lead to an increased unfavourable entropy variation upon binding. These phenomena may contribute to the moderate affinity obtained for most of the initial protein binders designed by grafting methods. However, several remarks may be made about our method. First, our approach entails a limited number of mutations (maximum four in this case) to graft the functional motif on the protein scaffold. The present approach is thus far less aggressive for protein structures than backbone grafting approaches, thereby increasing the possibility that the designed proteins actually bind the target. Second, the presence of disulfides in some scaffolds may contribute to preserving both the structure and the native conformational ensemble after mutation. Third, our method was designed to produce a series of protein binder candidates rather than a single one, thus increasing the possibility of obtaining at least one binder with the desired property. This is the reason why we are continuing to improve the STAMPS software to take advantage of the complete structural diversity of the PDB. Further efforts to correctly sample the conformational ensemble accessible to the molecule in the free state and the bound state by simulation approach are also required. This may permit the introduction of new scaffold selection criteria.

One advantage of the grafting approach is that it permits control of the location where the binder interacts at the target surface. Furthermore, knowledge of the binding mode of the initial moderate-affinity binder may be used to guide its optimization. This permitted production of an analogue of endothelin-1 in which the residue at position 2 was replaced by a bulky non-natural amino acid. This led to an increased binding affinity compared to the parent analogue for all MMPs tested (up to 290 nM for MMP-12). This is consistent with previous reports showing that bulky hydrophobic non-natural side chains interacting at the S1′ sub-site may lead to improved affinity of pseudo-peptide inhibitors. However, optimization of initial binders by single mutations often gives disappointing results, as observed in the case of scaffold 1AN1_I. Alternatively, coupling of efficient computational design methods with rapid experimental techniques for testing and evolving the proteins designed *in silico* could become a standard to develop artificial protein binders. The present work permitted the design of ten artificial protein binders targeting the catalytic domain of six MMPs, with affinities between 450 nM and 450 μM. Molecular models of these ten artificial protein binders in interaction with MMP-14 indicated that they possess interface areas in the range 700-1000 $Å^2$. These initial protein binder candidates may be used to create libraries focused on specific positions at these interfaces. This would permit selection of variants with improved affinity. Furthermore, differential selection may also be used to identify variants of these artificial protein binders displaying specificity for any particular MMP. Phage display and selection based on the scaffold of the broad-spectrum natural MMP inhibitor TIMP-2 [49] permitted us to obtain variants of this natural protein binder selective for MMP-1. The ten artificial miniprotein binders designed in the present study offer new starting points to develop selective artificial miniprotein inhibitors for various members of the matrixin family.

Experimental procedures

Scaffold search in the PDB

The scaffold search was performed using developmental version 3.1.1 of the STAMPS software (search for three-dimensional atom motifs in protein structures) [16]. This software permits identification of proteins possessing a set of residues in a topology similar to that of the functional motif used as input. STAMPS is able to screen all PDB files (> 75 000 protein structures, April 2012) very efficiently. Some flexibility was introduced via the STAMPS parameters used for the searches. The maximum rmsd for the position of the Ca,Cb atoms of a selected scaffold compared to the corresponding atoms in the selected motif was 1.25 Å (MAXRMSD). The difference between one of the Ca,Cb distances in the selected motif and the

corresponding distance in the motif examined must not exceed 1.85 Å (MAXDDIST). In addition, the maximum deviation between one of the Ca–Cb vectors in the selected motif and the corresponding vector in the examined scaffold must be lower than 130° (MAXANGLE), and the mean deviation for all Ca–Cb angles was limited to 90° (DELTANGLE). The size of the protein scaffolds was limited to a minimum of 15 residues (MINSIZE) and a maximum of 60 residues (MAXSIZE). The intermolecular steric hindrance criterion (STERICERR) allowing evaluation of steric compatibility of the scaffold with the target MMP-14 was set to 2.5. The parameter permitting evaluation of the intra-molecular steric hindrance after automated grafting of the motif residues on the scaffold (STERICINTRA) was set to 1.0. The 3D structures of scaffolds satisfying the above criteria in complex with the target MMP-14 were saved by STAMPS in PDB format. The search was performed using the full PDB.

Molecular modelling refinement of the selected scaffolds

The structures of the MMP-14–scaffold complexes identified by STAMPS were subjected to relaxation by molecular dynamics and energy minimization under positional restraints using CHARMM software version 35b1 [50] and CHARMM force field version 27 [51]. For each complex, after reconstruction of the hydrogen atoms absent in the original PDB structures, the initial structure was minimized under restraints. Standard harmonic restraints were applied to the Cα atoms of the target MMP-14 with a force constant of 50 kcal·mol^{-1}·A^{-2} except for residues of the target located at a distance less than 7 Å from the protein scaffold, for which the harmonic constant was set to 0.5 kcal·mol^{-1}·A^{-2}. The secondary structure of the protein scaffold was preserved using all intra-molecular Ca,Ca distance restraints, with the harmonic constant set to 50 kcal·mol^{-1}·A^{-2}. These distance restraints were defined via the Nuclear Overhauser Effect (NOE) facility of CHARMM. This allowed preservation of the 3D structure of the scaffold without preventing global movement of the scaffold with respect to the target. A set of positional restraints was also applied to the side chain of the grafted functional motif to preserve its position relative to the target. In all cases, we observed that the initial minimization allowed correction of some local steric hindrance between the two partners that resulted from the crude steric evaluation in STAMPS. After 300 initial energy minimization steps, 375 000 molecular dynamics steps (duration 0.0005 ps) at 300 K were performed. These calculations were performed *in vacuo*. After each 25 000 steps, the structure of the complex was quenched using a 3000 steps energy minimization, and the minimized structure was saved for further analysis.

Electrostatic analysis of the protein binder candidates

The electrostatic potentials of the various protein binder-target complexes were calculated using the Poisson–Boltzmann model with APBS software version 1.3 [41]. The coordinates of the structures used in these calculations are those of the X-ray structures after reconstruction of hydrogen atoms and, if necessary, of the missing coordinates of non-hydrogen atoms. Residues were considered to be in a protonation state similar to that of free amino acids without a pK shift at pH 7. After superimposing the structures of the common protein–target complex in each set, the resulting coordinates of the protein binder-target complexes were saved. Then the electrostatic potentials of the protein binder complexes were calculated in the absence of the target. The charge and radius used were those of the PARSE parameter set [52]. The grid dimensions and their spacing were defined in order to encompass all the scaffolds studied. This allowed further comparison of the electrostatic potential using PIPSA software [40]. The electrostatic potential was calculated using the focusing method. The non-linear version of the Poisson–Boltzmann equation was used with an ionic strength of 150 mM, and the radius of these ions was set to 2.0 Å. The dimensions of a coarse grid in the three directions were set to 97 points, with spacing of 2.0 Å. The solvent radius was set to 1.4 Å. The solute and solvent dielectric constants were set to 1 and 80, respectively. During the first calculation, we used the multiple Debye–Hückel boundary conditions. During the focusing step, the grid dimensions were set to 97 points in each direction, the spacing was set to 0.8 Å, and the boundary condition was obtained from the coarse grid calculation. The electrostatic potentials calculated were written in UHBD format [53] to allow further analysis using PIPSA software. The electrostatic potential was compared using version 2 of the PIPSA software. This program calculates the Hodgkin similarity indices to compare the molecular potential M of two proteins a and b [54,55]. The molecular potentials are calculated at point (i, j, k) on a three-dimensional grid superimposed on the molecule analysed, and the Hodgkin similarity index ESI_H$_{a,b}$, is defined as:

$$\mathrm{ESI\_H}_{a,b} = \frac{2(M_a, M_b)}{(M_a^2 + M_b^2)}$$

In the case of comparison of the electrostatic potentials Φ_a and Φ_b:

$$(M_a, M_b) = \sum_{i,j,k} \Phi_a(i,j,k)\Phi_b(i,j,k)$$

Therefore, the value of ESI_H varies from −1, when the electrostatic potentials are anti-correlated, to 1 when they are identical, with a value of 0 corresponding to uncorrelated potentials.

For greater physical relevance, the points (i, j, k) are selected in a region called 'skin' around each protein. In the present case, the skin is defined as a layer located between a distance of 3 Å (probe radius) and a distance of 7 Å from the van der Waals surface. This corresponded to a layer of thickness 4 Å. The Hodgkin similarity indices for comparison of the electrostatic potential of proteins a and b are calculated using the grid points within the intersection of their skins. The ESI_H calculation is performed either on the complete molecular skins (ESI_H_all) or is restricted to the binding site region (ESI_H_part). This region is defined as the intersection of the skin with a sphere of radius 10 Å centred on the coordinates of the N atom of residue C1 in TIMP-2 (structure 1BQQ).

The Hodgkin similarity index may be used to compare the shape of two proteins a and b. In this case, the shape similarity index $SSI\_H_{a,b}$, is defined as:

$$SSI\_H_{a,b} = \frac{2(n_{inter})}{(n_a + n_b)}$$

where n_a and n_b are the number of grid points in the skin of proteins a and b, respectively, and n_{inter} is the number of grid points in the intersection of skins of proteins a and b. Therefore, the value of SSI_H varies from 0 when the shapes of the proteins are dissimilar, to 1 when they are identical. This shape similarity index may be calculated in the complete intersection of proteins a and b (SSI_H_all) or may be restricted to the binding region. This region was defined as the intersection of the skins in the binding region (SSI_H_part). This intersection was defined as a sphere of 10 Å centred on the coordinates of the N atom of residue C1 in TIMP-2 (structure 1BQQ).

DNA constructs and protein purification

Synthetic genes optimized for recombinant expression of miniproteins in *E. coli* were obtained from Geneart AG (Regensburg, Germany). These genes contain the sequence of a tobacco etch virus (TEV) protease cleavage site (ENLYFQ), followed by sequence corresponding to miniproteins containing the functional motif, with Gateway recombination sites at each extremity of the gene. These synthetic constructs were cloned by Gateway™ LR (Life Technologies, Grand Island, NY, USA) cloning technology using pETG-82A (EMBL, Heidelberg, Germany) as the destination vector. The resulting expression plasmids encode the fusion protein of interest. Variants of miniproteins were obtained by performing site-directed mutagenesis using the PCR technique. Nucleotide sequences of all expression plasmids were verified by sequencing (Eurofins MWG, Ebersberg, Germany).

For production of recombinant proteins, 1 L of LB medium containing 1 mL ampicillin (100 mg·mL^{-1}) was inoculated with 10 mL of an overnight culture of the BL21 Star™ (DE3), (Life technologies) *Escherichia coli* strain transformed with expression plasmids, and incubated at 37 °C. When the attenuance at 600 nm was 0.8, the culture was incubated at 20 °C for 30 min, and expression of recombinant proteins was induced by adding 1 mL of 1 M isopropyl thio-β-D-galactoside, and incubation was continued overnight at 20 °C. Cells were then harvested by centrifugation (4500 ***g***, 30 min, 4 °C), and the pellet was resuspended in 50 mL lysis buffer (100 mM Tris/HCl, pH 8, 150 mM NaCl, 5% glycerol); cells were then disrupted using a cell disrupter (Constant System Ltd, Daventry, UK). The lysate was cleared by centrifugation (39 000 ***g***, 30 min, 4 °C) and the supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare, Freiburg, Germany). The His6-tagged fusion proteins were then eluted using a linear gradient of buffer B (100 mM Tris/HCl, pH 8, 150 mM NaCl, 5% glycerol, 500 mM imidazole) in buffer A (100 mM Tris/HCl, pH 8, 150 mM NaCl, 5% glycerol, 40 mM imidazole) (0–100% B over 30 min at a flow rate of 2 mL·min^{-1}). The fractions containing the His6-tagged fusion protein were pooled and dialysed for 3 h against buffer C (50 mM Tris/HCl, pH 8) using a Spectra/Por® (Spectrum Europe BV, Breda, the Netherlands) dialysis membrane (molecular weight cut-off 3500). The protein of interest was then cleaved using 10% w/w TEV protease overnight at 4 °C. Then the protein of interest was purified by reversed-phase HPLC on a C4 column (Vydac 214TP1010 Grace Discovery Science, Columbia, MD, USA; particle diameter 10 μm; pore size 300 Å; column width and length 10 × 250 mm) using a non-linear gradient of solvent B [90% acetonitrile, 10% H_2O, 0.09% trifluoroacetic acid (TFA)] in solvent A (95% H_2O, 5% acetonitrile 0.1% TFA) at a flow rate of 4 mL·min^{-1}.

Solid-phase peptide synthesis

Proteins 1EDP_A, 2UUY_B, 2F3C_I, 2IT8_A and mutants of 1EDP_A were obtained by solid-phase synthesis on an automated peptide synthesizer (model 433A; Applied Biosystems, Life Technologies, Carlsbad, CA, USA) using the fluorenylmethyloxycarbonyl/tert-butyl ether (Fmoc/tBu) strategy and a modified program. As a standard procedure, 0.1 mmol of Fmoc–C-terminal rink amide methylbenzhydrylamine (MBHA) resin (loading rate 0.66 mmol·g^{-1}), a tenfold excess of each Fmoc-amino acid (Novabiochem, MERCK BIOSCIENCES AG, Läufelfingen, Switzerland), DCCI (dicyclohexylcarbodiimide)/HOBt (1-hydroxybenzotriazole) activation and HBTU (O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate)/NMP (N-Methyl-2-pyrrolidone) coupling conditions were used. After drying under vacuum, the miniprotein was liberated from the resin, and protecting groups were removed by acidic cleavage with 10 mL TFA/triisopropylsilane (TIS)/H_2O (95/2.5/2.5 v/v/v) for 2 h at room temperature. The resin was then filtered, and the free miniprotein was precipitated in 100 mL ice-cold methyl *tert*-butyl ether (TBME). After centrifugation (3000 ***g***, 30 min,

4 °C), the pellet was dissolved in 30 mL 10% acetic acid and freeze-dried.

Purification by reversed-phase HPLC

The crude extracts were reduced in denaturing buffer (0.5 M Tris, pH 8, 6 M guanidine) with 10 mM dithiothreitol for 90 min at 37 °C, and the solution was then acidified to pH 2 using TFA/H_2O (50/50 v/v) and subjected to reversed-phase HPLC purification on a Discovery® BIO Wide Pore C18 semi-preparative column (Supelco, Sigma-Aldrich, St Louis, MO, USA; particle diameter 5 μm; pore size 300 Å; column width 10 mm and length 250 mm) using non-linear gradients of solvent B (90% acetonitrile, 10% H_2O, 0.09% TFA) in solvent A (90% H_2O, 10% acetonitrile, 0.1% TFA) at a flow rate of 5 mL·min^{-1}. Detection was performed at 214 nm using a Merck (Whitehouse Station, NJ, USA) L4000 UV detector equipped with a 5 mm cell.

Disulfide pairing

Disulfide pairing was performed in denaturing buffer containing 5 mM reduced and 0.5 mM oxidized glutathione and the protein at a final concentration of 1 mg·mL^{-1}, at 4 °C for 24–48 h. Monitoring of disulfide pairing was performed by following the disappearance of the fully reduced protein by analytical reversed-phase HPLC. The solution was purified as described above.

Characterization of proteins

After HPLC purification, chemically and biologically produced proteins were lyophilized and then solubilized in the appropriate buffer for further studies. Purity of proteins was assessed by analytical HPLC coupled to a high-resolution ion trap mass spectrometer (Esquire HCT; Bruker Daltonics, Billerica, MA, USA). The correct mass after disulfide bond formation was confirmed by MALDI-TOF/TOF 4800 analysis (Applied Biosystems), and correct disulfide pairing was checked by binding assays using natural targets of miniproteins, circular dichroism or crystallography. Concentrations of miniproteins were determined using a JLC-500/V AminoTac amino acid analyser (JEOL, Tokyo, Japan) and absorbance measurement at 280 nm on a Beckmann (Fullerton, CA, USA) spectrometer using a molar extinction coefficient calculated on the basis of the known amino acid content. Disulfide pairing was tested indirectly by checking the native function of the scaffold when possible and when the grafting did not entail mutation of residues involved in the native function. Scaffold 2F3C_I is a trypsin inhibitor. We observed that 2F3C_I_m inhibits trypsin with a K_i value of 0.5 ± 0.1 nM, which is close to the value reported in the literature (2 nM) [56]. Similarly, scaffold 1AN1_I is a trypsin inhibitor. We found that 1AN1_I_m was able to inhibit trypsin with a K_i value of 1.7 ± 0.2 nM, consistent with the value reported in the literature (1.8 nM [57]). The K_i value determined for inhibition of trypsin activity by 2UUY_B_m was 3.0 ± 0.2 nM, again consistent with the value reported in the literature (< 10 nM [58]). Finally, the inhibition constant for 2IT8_A_m was 73.0 ± 0.2 nM, a factor that is ~20 times higher than the value reported in the literature [59]. However, in this case, three residues involved in the native function of the scaffold were mutated for grafting the functional motif. Nevertheless, the high inhibition constant for trypsin suggests that the mutated protein is correctly folded.

MMP inhibition assays

MMP inhibition by the miniproteins was tested by competition experiments using 1.8 μM of the fluorogenic substrate McaMat (Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2, Calcobiochem, Merck) and human MMPs in a nanomolar range concentration (from R & D Systems, Minneapolis, MN, USA, or produced in the laboratory as described previously [60]), in 50 mM Tris/HCl buffer pH 6.8, 10 mM $CaCl_2$. The substrate and enzyme concentrations for the experiments were chosen so as the substrate utilization remains well below 10% of the total amount and to observe the initial rates. For each inhibitor, the percentage of inhibition was determined in duplicate experiments. Four inhibitor concentrations were chosen to observe a 20–80% range of inhibition. After 45 min incubation under shaking at 25 °C, inhibition assays were performed by recording the fluorescence increase induced by cleavage of the fluorogenic substrate, using 96-well non-binding surface plates (Corning-Costar, Lowell, MA, USA). Fluorescence signals were monitored using a Fluoroscan Ascent photon counter spectrophotometer (Thermolab) equipped with a temperature control device and a plate shaker. K_i values were determined as described previously [61].

Crystallization

The protein solution for crystallization consisted of 643 μM of the catalytic domain of the F67D mutant of human MMP-12 (residues 106–263) with 1 mM acetohydroxamic acid to prevent self-degradation of the proteinase during crystallization in 0.1 M glycine pH 8.5 and 27% polyethylene glycol 10 000. Crystals of MMP-12 were grown in the presence of acetohydroxamic acid by sitting-drop vapor diffusion experiments at 20 °C in a cooled incubator. The crystals were harvested and transferred to a cryo-solution (90 mM Tris/HCl pH 8, 27% polyethylene glycol 8000, 15% monomethyl ether/polyethylene glycol 550, 10% glycerol) for 5 min to remove acetohydroxamic acid, then transferred to the same cryo-solution with 1 mM 1EDP_A for 5 min before the crystals were picked up with a loop and plunged into liquid nitrogen.

Structure determination and refinement

The data for crystals were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beam line ID14-2 at 100 K from a single crystal. Most MMP-12 crystals soaked with 1EDP_A were badly disordered, and only one crystal gave usable data to only 2.8 Å resolution. Similar crystals that had followed the same soaking procedure but with small nanomolar inhibitors diffracted to 1.6–2.3 Å. Data reduction was performed using MOSFLM [62]. The crystals belong to tetragonal space group $P2_12_12$, with cell parameters $a = 69.1$ Å, $b = 63.2$ Å, $c = 37.0$ Å, with one molecule in the asymmetric unit. The structure was solved by rigid-body refinement using REFMAC [63], starting with PBD entry 3LIR [64]. The CCP4 suite of programs [65] was used for manipulation of structure factors and coordinates. Final fitting and stereochemical analysis of the refined model were performed using COOT [66]. The figures were created using PyMOL (Schrödinger Inc., New York, NY, USA).

Molecular dynamics simulation in water solution

The calculations were carried using CHARMM version 35b1 [50] and version 27 of the CHARMM parameter set for proteins [51]. For each artificial binder, the initial MMP-14 complex was obtained after relaxation by energy minimization and molecular dynamics under positional restraints *in vacuo* of the complexes identified by STAMPS (see above). This structure was immersed in a cubic box of pre-equilibrated TIP3P water model. After removal of the water molecules overlapping the protein complex, a short minimization was performed to optimize the water–protein interface. During this step, the coordinates of the proteins were fixed. This was followed by a second energy minimization under positional restraints applied to the protein atoms. Then, molecular dynamics simulation was performed. The system was heated from 0 to 298 K in 100 ps in steps of 50 K. During this equilibration step, harmonic restraints (5 kcal·mol^{-1}·Å^{-2}) were applied to the positions of zinc, calcium and non-hydrogen backbone atoms. After equilibration, all positional restraints were removed. Molecular dynamics simulation was performed using the leapfrog Verlet algorithm at constant pressure (1 atm) and temperature (298 K). Electrostatic interactions were calculated using the particle mesh Ewald summation method. The time step was set to 1 fs, and the coordinates were saved every 0.1 ps.

Acknowledgements

Gael Debret is gratefully acknowledged for her work in development of the STAMPS software.

References

1 Gebauer M & Skerra A (2009) Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Cheml Biol* **13**, 245–255.
2 Beck A, Wurch T, Bailly C & Corvaia N (2010) Strategies and challenges for the next generation of therapeutic antibodies. *Nat Rev Immunol* **10**, 345–352.
3 DeLano WL, Ultsch MH, de Vos AM & Wells JA (2000) Convergent solutions to binding at a protein–protein interface. *Science* **287**, 1279–1283.
4 Vita C, Drakopoulou E, Vizzavona J, Rochette S, Martin L, Menez A, Roumestand C, Yang YS, Ylisastigui L, Benjouad A *et al.* (1999) Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc Natl Acad Sci USA* **96**, 13091–13096.
5 Domingues H, Cregut D, Sebald W, Oschkinat H & Serrano L (1999) Rational design of a GCN4-derived mimetic of interleukin-4. *Nat Struct Biol* **6**, 652–656.
6 Martin L, Stricher F, Misse D, Sironi F, Pugniere M, Barthe P, Prado-Gotor R, Freulon I, Magne X, Roumestand C *et al.* (2003) Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat Biotechnol* **21**, 71–76.
7 Sia SK & Kim PS (2003) Protein grafting of an HIV-1-inhibiting epitope. *Proc Natl Acad Sci USA* **100**, 9756–9761.
8 Cobos ES, Pisabarro MT, Vega MC, Lacroix E, Serrano L, Ruiz-Sanz J & Martinez JC (2004) A miniprotein scaffold used to assemble the polyproline II binding epitope recognized by SH3 domains. *J Mol Biol* **342**, 355–365.
9 Li C, Liu M, Monbo J, Zou GZ, Li CQ, Yuan WR, Zella D, Lu WY & Lu WY (2008) Turning a scorpion toxin into an antitumor miniprotein. *J Am Chem Soc* **130**, 13546–13548.
10 Azoitei ML, Correia BE, Ban YEA, Carrico C, Kalyuzhniy O, Chen L, Schroeter A, Huang PS, McLellan JS, Kwong PD *et al.* (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science* **334**, 373–376.
11 Azoitei ML, Ban YEA, Julien JP, Bryson S, Schroeter A, Kalyuzhniy O, Porter JR, Adachi Y, Baker D, Pai EF *et al.* (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J Mol Biol* **415**, 175–192.
12 Liang SD, Li WZ, Xiao L, Wang JS & Lai LH (2000) Grafting of protein–protein interaction epitope. *J Biomol Struct Dynamics* **17**, 821–828.
13 Magis C, Gasparini D, Lecoq A, Le Du MH, Stura E, Charbonnier JB, Mourier G, Boulain JC, Pardo L, Caruana A *et al.* (2006) Structure-based secondary structure-independent approach to design

protein ligands: application to the design of Kv1.2 potassium channel blockers. *J Am Chem Soc* **128**, 16190–16205.
14 Keskin O & Nussinov R (2005) Favorable scaffolds: proteins with different sequence, structure and function may associate in similar ways. *Protein Eng Des Sel* **18**, 11–24.
15 Keskin O & Nussinov R (2007) Similar binding sites and different partners: implications to shared proteins in cellular pathways. *Structure* **15**, 341–354.
16 Debret G, Martel A & Cuniasse P (2009) Rasmot-3d Pro: a 3D motif search webserver. *Nucleic Acids Res* **37**, W459–W464.
17 Zhang CS & Lai LH (2012) AutoMatch: target-binding protein design and enzyme design by automatic pinpointing potential active sites in available protein scaffolds. *Proteins* **80**, 1078–1094.
18 Liu S, Liu S, Zhu X, Liang H, Cao A, Chang Z & Lai L (2007) Nonnatural protein–protein interaction-pair design by key residues grafting. *Proc Natl Acad Sci USA* **104**, 5330–5335.
19 Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA & Baker D (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821.
20 McCawley LJ & Matrisian LM (2001) Matrix metalloproteinases: they're not just for matrix anymore! *Curr Opin Cell Biol* **13**, 534–540.
21 Rydlova M, Holubec L, Ludvikova M, Kalfert D, Franekova J, Povysil C & Ludvikova M (2008) Biological activity and clinical implications of the matrix metalloproteinases. *Anticancer Res* **28**, 1389–1397.
22 Whittaker M, Floyd CD, Brown P & Gearing AJH (1999) Design and therapeutic application of matrix metalloproteinase inhibitors. *Chem Rev* **99**, 2735–2776.
23 Cuniasse P, Devel L, Makaritis A, Beau F, Georgiadis D, Matziari A, Yiotakis A & Dive V (2005) Future challenges facing the development of specific active-site-directed synthetic inhibitors of MMPs. *Biochimie* **87**, 393–402.
24 Fisher JF & Mobashery S (2006) Recent advances in MMP inhibitor design. *Cancer Metast Rev* **25**, 115–136.
25 Tu GG, Xu WF, Huang HM & Li SH (2008) Progress in the development of matrix metalloproteinase inhibitors. *Curr Med Chem* **15**, 1388–1395.
26 Yiotakis A & Dive V (2008) Synthetic active site-directed inhibitors of metzincins: achievement and perspectives. *Mol Aspects Med* **29**, 329–338.
27 Bode W (2003) Structural basis of matrix metalloproteinase function. *Biochem Soc Symp* **70**, 1–14.
28 Tallant C, Marrero A & Gomis-Ruth FX (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim Biophys Acta* **1803**, 20–28.
29 Brew K, Dinakarpandian D & Nagase H (2000) Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochim Biophys Acta* **1477**, 267–283.
30 Butler GS, Hutton M, Wattam BA, Williamson RA, Knauper V, Willenbrock F & Murphy G (1999) The specificity of TIMP-2 for matrix metalloproteinases can be modified by single amino acid mutations. *J Biol Chem* **274**, 20391–20396.
31 Meng Q, Malinovskii V, Huang W, Hu YJ, Chung L, Nagase H, Bode W, Maskos K & Brew K (1999) Residue 2 of TIMP-1 is a major determinant of affinity and specificity for matrix metalloproteinases but effects of substitutions do not correlate with those of the corresponding P1' residue of substrate. *J Biol Chem* **274**, 10184–10189.
32 Wei S, Chen Y, Chung L, Nagase H & Brew K (2003) Protein engineering of the tissue inhibitor of metalloproteinase 1 (TIMP-1) inhibitory domain – in search of selective matrix metalloproteinase inhibitors. *J Biol Chem* **278**, 9831–9834.
33 Hamze AB, Wei S, Bahudhanapati H, Kota S, Acharya KR & Brew K (2007) Constraining specificity in the N-domain of tissue inhibitor of metalloproteinases-1; gelatinase-selective inhibitors. *Protein Sci* **16**, 1905–1913.
34 GomisRuth FX, Maskos K, Betz M, Bergner A, Huber R, Suzuki K, Yoshida N, Nagase H, Brew K, Bourenkov GP *et al.* (1997) Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **389**, 77–81.
35 Fernandez-Catalan C, Bode W, Huber R, Turk D, Calvete JJ, Lichte A, Tschesche H & Maskos K (1998) Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J* **17**, 5238–5248.
36 Morgunova E, Tuuttila A, Bergmann U & Tryggvason K (2002) Structural insight into the complex formation of latent matrix metalloproteinase 2 with tissue inhibitor of metalloproteinase 2. *Proc Natl Acad Sci USA* **99**, 7414–7419.
37 Maskos K, Lang R, Tschesche H & Bode W (2007) Flexibility and variability of TIMP binding: X-ray structure of the complex between collagenase-3/MMP-13 and TIMP-2. *J Mol Biol* **366**, 1222–1231.
38 Higashi S & Miyazaki K (1999) Reactive site-modified tissue inhibitor of metalloproteinases-2 inhibits the cell-mediated activation of progelatinase A. *J Biol Chem* **274**, 10497–10504.
39 Troeberg L, Tanaka M, Wait R, Shi YE, Brew K & Nagase H (2002) *E. coli* expression of TIMP-4 and comparative kinetic studies with TIMP-1 and TIMP-2: insights into the interactions of TIMPs and matrix

metalloproteinase 2 (gelatinase A) . *Biochemistry* **41**, 15025–15035.
40 Blomberg N, Gabdoulline RR, Nilges M & Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins* **37**, 379–387.
41 Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**, 10037–10041.
42 Lo Conte L, Chothia C & Janin J (1999) The atomic structure of protein–protein recognition sites. *J Mol Biol* **285**, 2177–2198.
43 Jones S & Thornton JM (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci USA* **93**, 13–20.
44 Xu D, Tsai CJ & Nussinov R (1997) Hydrogen bonds and salt bridges across protein–protein interfaces. *Protein Eng* **10**, 999–1012.
45 Guerois R, Nielsen JE & Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369–387.
46 Peisajovich SG & Tawfik DS (2007) Protein engineers turned evolutionists. *Nat Methods* **4**, 991–994.
47 Lauer-Fields JL, Cudic M, Wei S, Mari F, Fields GB & Brew K (2007) Engineered sarafotoxins as tissue inhibitor of metalloproteinases-like matrix metalloproteinase inhibitors. *J Biol Chem* **282**, 26948–26955.
48 Huang CC, Stricher F, Martin L, Decker JM, Majeed S, Barthe P, Hendrickson WA, Robinson J, Roumestand C, Sodroski J *et al.* (2005) Scorpion-toxin mimics of CD4 in complex with human immunodeficiency virus gp120: crystal structures, molecular mimicry, and neutralization breadth. *Structure* **13**, 755–768.
49 Bahudhanapati H, Zhang YN, Sidhu SS & Brew K (2011) Phage display of tissue inhibitor of metalloproteinases-2 (TIMP-2): identification of selective inhibitors of collagenase-1 (metalloproteinase 1 (MMP-1)). *J Biol Chem* **286**, 31761–31770.
50 Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S & Karplus M (1983) CHARMM – a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* **4**, 187–217.
51 MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586–3616.
52 Sitkoff D, Sharp KA & Honig B (1994) Accurate calculation of hydration free-energies using macroscopic solvent models. *J Phys Chem* **98**, 1978–1988.
53 Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR *et al.* (1995) Electrostatics and diffusion of molecules in solution – simulations with the University of Houston Brownian dynamics program. *Comput Phys Commun* **91**, 57–95.
54 Wade RC, Gabdoulline RR & De Rienzo F (2001) Protein interaction property similarity analysis. *Int J Quantum Chem* **83**, 122–127.
55 Hodgkin EE & Richards WG (1987) Molecular similarity based on electrostatic potential and electric field. *Int J Quantum Chem* **32**, 105–110.
56 Campos ITN, Amino R, Sampaio CAM, Auerswald EA, Friedrich T, Lemaire HG, Schenkman S & Tanaka AS (2002) Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem Mol* **32**, 991–997.
57 Stubbs MT, Morenweiser R, Sturzebecher J, Bauer M, Bode W, Huber R, Piechottka GP, Matschiner G, Sommerhoff CP, Fritz H *et al.* (1997) The three-dimensional structure of recombinant leech-derived tryptase inhibitor in complex with trypsin – implications for the structure of human mast cell tryptase and its inhibition. *J Biol Chem* **272**, 19931–19937.
58 Valdez-Cruz NA, Conde R, Zamudio F & Possani LD (2011) Anticoagulants from scorpion venoms. In Toxins and Homeostasis: From Bench to Bedside (Marklan FS, ed.), pp. 255–266, Springer, NY, USA.
59 Sommerhoff CP, Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Diederichsen U & Kolmar H (2010) Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase β. *J Mol Biol* **395**, 167–175.
60 Devel L, Rogakos V, David A, Makaritis A, Beau F, Cuniasse P, Yiotakis A & Dive V (2006) Development of selective inhibitors and substrate of matrix metalloproteinase-12. *J Biol Chem* **281**, 11152–11160.
61 Horovitz A & Levitzki A (1987) An accurate method for determination of receptor ligand and enzyme-inhibitor dissociation constants from displacement curves. *Proc Natl Acad Sci USA* **84**, 6654–6658.
62 Leslie AGW (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D* **62**, 48–57.
63 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240–255.
64 Devel L, Garcia S, Czarny B, Beau F, Lajeunesse E, Vera L, Georgiadis D, Stura E & Dive V (2010) Insights from selective non-phosphinic inhibitors of MMP-12 tailored to fit with an S-1' loop canonical conformation. *J Biol Chem* **285**, 35900–35909.
65 Bailey S (1994) The Ccp4 suite – programs for protein crystallography. *Acta Crystallogr D* **50**, 760–763.

66 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* **66**, 486–501.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Table S1. STAMPS output for the two searches of the PDB.

Table S2. Mean inter-atom distances determined in molecular dynamics simulations of the complexes between the protein binder candidates and the catalytic domain of MMP-14.