Affinity transfer to the archaeal extremophilic Sac7d protein by insertion of a CDR

Sabino Pacheco[1,2,3,4], Ghislaine Béhar[1,2,3], Mike Maillasson[1,2,3,5], Barbara Mouratou[1,2,3], Frédéric Pecorari[1,2,3*]

[1]CRCNA - UMR 892 INSERM, [2]6299 CNRS, [3]University of Nantes, 8 quai Moncousu, BP 70721, 44007 Nantes, Cedex 1, France.

[4]Institut Pasteur, CNRS UMR 3528, Unité de Microbiologie Structurale, 25 rue du Dr. Roux, 72724 Paris Cedex 15, France.

[5]Plate-forme IMPACT biogenouest, 8 quai Moncousu, BP 70721, 44007 Nantes, Cedex 1, France.

[*]Address for correspondence: Université de Nantes, CRCNA - UMR 892 INSERM, 6299 CNRS, 8 quai Moncousu, BP 70721, 44007 Nantes Cedex 1, France. Tel: 33-2 40 41 28 51; E-mail: frederic.pecorari@univ-nantes.fr

Running title: Affinity transfer to the extremophilic Sac7d protein

ABSTRACT

Artificially transforming a scaffold protein into binders often consists of introducing diversity into its natural binding region by directed mutagenesis. We have previously developed the archaeal extremophilic Sac7d protein as a scaffold to derive affinity reagents (Affitins) by randomization of either a flat surface or a flat surface and two short loops with natural lengths. Short loops are believed to contribute to stability of extremophilic proteins, and loop extension has been reported detrimental for the thermal and chemical stabilities of mesophilic proteins. In this work, we wanted to evaluate the possibility of designing target-binding proteins based on Sac7d by using a complementary determining region (CDR). To this aim, we inserted into three different loops a 10 residues CDR from the cAb-Lys3 anti-lysozyme camel antibody. The chimeras obtained were as stable as wild type Sac7d at extreme pH and their structural integrity was supported. Chimeras were thermally stable, but with T_ms from 60.9°C to 66.3°C (*cf.* 91°C for Sac7d) which shows that loop extension is detrimental for thermal stability of Sac7d. The loop 3 enabled anti-lysozyme activity. These results pave the way for the use of CDR(s) from antibodies and/or extended randomized loop(s) to increase the potential of binding of Affitins.

KEYWORDS

Sac7d, affinity transfer, CDR insertion, Affitin, extremophilic

INTRODUCTION

The continuous development of protein based affinity reagents with favorable properties is essential for research, clinical and industrial applications. While antibodies are the most widely used proteins to develop binders, another class of proteins termed "alternative scaffold proteins" has emerged during the last two decades (for reviews see (Binz *et al.*, 2005, Gebauer and Skerra, 2009)). These proteins are usually small (less than 15 kDa), easy to engineer, stable and expressed with high yields in *Escherichia coli*.

The transformation of proteins into binders requires the introduction of diversity into a region that will become the binding site by creating libraries from which variants are isolated via a selection/screening step. This process is well described for various scaffolds such as Affibodies, DARPins, Anticalins, Monobodies and Affitins. The choice of the region in which to introduce diversity in a protein is usually driven by how the protein binds its natural ligand: it can be a surface (Binz *et al.*, 2004, Mouratou *et al.*, 2007, Nord *et al.*, 1997), or loops (Beste *et al.*, 1999, Koide *et al.*, 1998). Thus, scaffold proteins are often obtained by replacement mutagenesis to maintain the natural mode of binding.

However, it is important to know if the natural mode of binding is the most efficient to derive affinity proteins against a wide range of targets. For example, inserting flexible loops, which often guide protein binding events as observed for CDRs of antibodies, onto a scaffold that originally used only a surface for binding can extend its potential to reach buried epitopes in protein grooves or catalytic sites (Correa *et al.*, 2014, Koide *et al.*, 2012, Schilling *et al.*, 2014).

It has been reported that increasing the length of loops in Rop and chymotrypsin inhibitor-2 proteins can result in a decrease in the overall stability of the proteins

(Ladurner and Fersht, 1997, Nagi and Regan, 1997) while for the protein inhibitor of neuronal nitric oxide synthase (PIN-bodies), no significant destabilization could be observed (Bes *et al.*, 2006). For the fibronectin type III domain, also known as the Monobody scaffold, destabilization effects could be observed or not depending on which loop was extended (Batori *et al.*, 2002).

Overall, these studies suggest that the detrimental consequences of extending loops seem at least protein- and loop-dependent, and therefore difficult to predict. This makes even more challenging the design of an unnatural mode of binding, compared to a conservative one, in a protein without altering the favorable properties for which it was chosen as a scaffold.

We have developed artificial affinity proteins (Affitins) derived from extremophilic proteins from the "7kDa DNA-binding" family found in Archaea (*Sulfolobus*, *Acidianus*, and *Metallospharea* genus for examples), such as Sac7d (Behar *et al.*, 2013, Buddelmeijer *et al.*, 2009, Krehenbrink *et al.*, 2008, Mouratou, *et al.*, 2007). This was primarily done by recruiting residues of the β-sheet surface originally involved in the binding of DNA. As an important step in exploring how diversity can be efficiently introduced into this scaffold, we reported that Affitins are able to bind different epitopes of the same target via two different modes of binding: involving a flat surface or involving a flat surface and two short loops with natural lengths (Behar, *et al.*, 2013).

Sac7d is a hyperthermostable protein with a T_m of 90°C (Edmondson and Shriver, 2001, McCrary *et al.*, 1996, Mouratou, *et al.*, 2007) and is chemically resistant from pH = 0 up to pH = 12 (Behar, *et al.*, 2013). Consequently, these unusual stabilities can be expected to be helpful in reshaping loops. However, it is assumed that short loops contribute to the stability of extremophilic proteins (Kumar and Nussinov, 2001,

Thompson and Eisenberg, 1999). Thus, extending loops in Sac7d could be destabilizing.

We reported that the loop 2 of Affitin D1 is involved in the recognition of human IgG (Behar, *et al.*, 2013), and we recently designed a novel generation of binders with enzymatic inhibition properties by artificially extending this loop with four randomized residues. This was achieved while conserving the favorable properties of Affitins (Correa, *et al.*, 2014). In this work we wanted to address three main questions. Can insertion of a CDR into different loops result in a folded Sac7d? Are the chimeric proteins obtained still hyperthermostable? Is the potential of each loop to introduce recognition of a target equivalent? With these aims, we inserted an anti-lysozyme CDR of ten residues from a camel antibody (Desmyter *et al.*, 1996) at three positions where there is a loop in Sac7d and studied the function and stability properties of the resulting chimeric proteins.

MATERIALS & METHODS

Molecular biology

The sequence ^{100}STIYASYYES^{109} was inserted between the residues ^{9}KG^{10}, ^{27}GK^{28} and ^{37}NG^{38} of the Sac7d protein. This sequence corresponds to the CDR3 sequence involved in hen egg white lysozyme (HEWL)-binding of cAb-Lys3 antibody but with a C109S substitution in order to avoid an intermolecular disulfide bond between the thiol groups. Mutants were named mL1, mL2 and mL3, respectively. Mutagenesis was carried out by two-step PCR using the *sac7d* wild-type gene as template (5'-

GGATCCGTGAAAGTGAAATTTAAATATAAAGGCGAAGAAAAAGAAGTGG

ACACCAGCAAGATCAAGAAAGTTTGGCGTGTGGGCAAAATGGTGAGCTTT

ACCTACGACGACAACGGCAAGACCGGCCGTGGCGCCGTGAGCGAGAAAG ATGCCCCGAAAGAGTTATTAGATATGTTAGCGCGTGCGGAACGTGAAAAA AAGCTT-3'). The first PCR was performed with forward primers 5'-GAAAGTTTGGCGTGTGGGC**TCTACCATTTATGCGTCTTATTATGAAAGC** AAAATGGTGAGCTTTACCTACG-3' and 5'-GGTGAGCTTTACCTACGACGACAACT**CTACCATTTATGCGTCTTATTATG** **AAAGC**GGCAAGACCGGCCGTGGCGC-3' to introduce the CDR3 (bold sequence) in mL2 and mL3, respectively. The reverse primer 5'-ATATA<u>AAGCTT</u>TTTCTCGCGTTCCGCAC-3' was common to both mutants and contains the HindIII restriction site (underlined sequence). To obtain the complete gene, a second PCR was performed using the first PCR fragments as reverse "megaprimers" and forward primer 5'-ATATAGGATCCGTGAAAGTGAAATT-3', which contains the BamHI restriction site. To obtain mL1 protein, only one-step PCR was performed using the forward primer 5'-TAT<u>GGATCC</u>GTGAAAGTGAAATTTAAATATAAAT**CTACCATTTATGCGTC** **TTATTATGAAAGC**GGCGAAGAAAAAGAAGTGGAC-3' and the same reverse primer with the HindIII restriction site. PCR products were cloned in pFP1001 vector, which is a modified version from pQE30 that enables the fusion of RGS-6His tag at the N-terminal.

Production and purification of proteins

Proteins were produced in *E. coli* DH5α harboring the plasmid pFP1001 with cloned genes. All cultures were grown to reach a DO_{600} of about 0.6 in 2xYT and the protein expression was induced with 0.5 mM IPTG over night at 30°C. Cells were pelleted by centrifugation at 4,000 g for 15 minutes and resuspended in lysis buffer

(50 mM $NaPO_4$ pH 7.5, 500 mM NaCl, 20 mM imidazole) containing Bugbuster solution (Novagene). Cells were lysed using an EmulsiFlex homogenizer and centrifuged at 18,000 g for 45 min. Supernatant was used for purification by immobilized metal ion affinity chromatography (IMAC). Then, the proteins were injected into a Superdex75 16/60 column (GE Healthcare) equilibrated with 25 mM Tris-HCl pH 8.0, 500 mM NaCl. Monomeric purified proteins were dialyzed overnight at 4°C against PBS buffer pH 7.4 (130 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) and quantified spectrophotometrically at 280 nm using an extinction coefficient of 8250 M^{-1} cm^{-1} for wild type Sac7d and 12090 M^{-1} cm^{-1} for chimeric proteins.

Analysis of secondary structure content of proteins

Far-UV CD spectra were recorded from 190 to 260 nm on a JASCO J-810 instrument (Jasco) using a quartz cell with a path length of 0.2 cm (Hellma). The temperature of the sample was maintained at 20°C with a programmable Peltier single cell holder PTC-423S. Measurements were recorded in step mode (interval, 1 nm; bandwidth, 2 nm; response time, 0.125 s) and averaged over three scans. Proteins were diluted to a final concentration of 0.25 mg/ml in PBS. A baseline correction was made with the buffer. Secondary structure prediction was performed with the Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) using the CDSSTR algorithm (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008).

pH stability of proteins

To study their pH stability, proteins were diluted to 0.33 mg/ml in buffered solution or a strong acid corresponding to each pH unit from 0 to 14. Except for HCl,

the following solutions were adjusted to the required pH with HCl or NaOH: pH 0 = 1 M HCl, pH 1 = 0.1 M HCl, pH 2 = 50 mM NaH_2PO_4, pH 3 = 50 mM NaH_2PO_4, pH 4 = 50 mM acetic acid, pH 5 = 50 mM acetic acid, pH 6 = 50 mM MES, pH 7 = 50 mM phosphoric acid, pH 8 = 50 mM NaH_2PO_4, pH 9 = 50 mM Tris base, pH 10 = 50 mM methylamine, pH 11 = 50 mM methylamine, pH 12 = 50 mM NaH_2PO_4, pH 13 = 0.1 M NaOH, pH 14 = 1 M NaOH. All the buffers contained NaCl 300 mM. Protein samples were incubated overnight in these buffers at room temperature and CD spectra were recorded at 20°C as described above, using a quartz cell with a path length of 0.2 cm (Hellma).

Thermal stability of proteins

Thermally-induced unfolding of proteins was studied by circular dichroism as described by Edmondson and Shriver (Edmondson and Shriver, 2001). Briefly, proteins were dialyzed against a 1 mM potassium acetate buffer (pH 5.5). Proteins were then diluted to 0.02 mg/ml in this buffer. Ellipticity was recorded at 205 nm (bandwidth, 5 nm; response time, 4 s) using a quartz cell with a path length of 1 cm (Hellma) and curves were normalized. The melting temperatures were evaluated from the collected data as midpoints of the thermal transitions using GraphPad Prism software and a two-state transition model assuming that there is a change in heat capacity of the folded and unfolded forms (Greenfield, 2006, Mouratou, *et al.*, 2007).

Detection of DNA binding activity by fluorescence

This assay was performed as previously reported (McAfee *et al.*, 1996) with some modifications. Tryptophan emission fluorescence was recorded with a Tecan Infinite M1000 spectrofluorimeter. Proteins (5µM) were incubated in the absence or presence

of calf thymus DNA (Sigma) at a final concentration of 200 µM in PBS buffer pH 7.4. An extinction coefficient of 6600 M^{-1} cm^{-1} was used to estimate the concentration of calf thymus DNA at 260 nm. The emission spectra were recorded at 25°C from 300 to 400 nm with 2 nm-increments using an excitation wavelength of 285 nm and a flat 96-well microplate (Greiner). PBS buffer spectra were subtracted from data obtained and a normalization of the curves was performed with equation 1:

$$F_i = F - F_{min} / F_{max} - F_{min} \qquad \text{(Eq.1)}$$

where F_{min} and F_{max} are the fluorescence minima and maxima, F is the fluorescence recorded for each data point and F_i the normalized fluorescence intensity.

DNA-binding affinity was determined by incubation of the proteins (5µM) with several concentrations of DNA (10 to 200 µM) and the quenching fluorescence was recorded at 345 nm. Values were normalized with equation 1 and transformed with equation 2:

$$F_Q = 1 - F_i \qquad \text{(Eq.2)}$$

Data were fitted to a nonlinear regression curve with GraphPad Prism software to determine affinity parameters.

Detection of DNA binding activity by electrophoretic mobility shift assay (EMSA)

For this assay a DNA probe of 200 bp was produced by PCR using wild type *sac7d* gene as template. Approximately 200 ng of this probe was incubated with varying quantities of purified wild type Sac7d or chimeric proteins (0, 0.6, 1.2, 2.5, 5, and 10 µM) in PBS buffer pH 7.0 in a total volume of 10 µl at room temperature for 30 min. The samples were resolved by electrophoresis in 6% acrylamide gels with TBE buffer

at constant voltage. After electrophoresis, gels were stained with GelGreen nucleic acid stain (Biotium).

Detection of anti-HEWL activity by ELISA

An ELISA plate was coated with 50 ng/well of HEWL (Sigma) or casein in PBS buffer and blocked with 0.5% BSA. Then, wild type Sac7d or chimeric proteins were added. After incubation, unbound proteins were removed by washing steps. Bound proteins were detected with an anti-RGS-6His antibody conjugated with horseradish peroxidase (Qiagen) diluted 1:5000. Finally, ortho-phenylenediamine (1 mg/ml OPD, 0.05% H_2O_2, 100 mM sodium citrate, pH 5.0) was used as the substrate for detection. Absorbance was recorded at 450 nm with a Tecan Infinite M1000 plate reader. ELISA was performed at 25°C with 1 h of incubation for each step.

Affinity determination of anti-HEWL activity by surface plasmon resonance (SPR)

Affinity binding assays using SPR were performed with a BIAcore 2000 instrument (GE Healthcare). HEWL diluted to 3 µg/ml in a sodium acetate buffer (pH 5.5) was immobilized (300 RU) by amine coupling on a CM5 chip activated previously with N-hydroxysuccinimide and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide. Residual reactive groups on the surface were inactivated with 1 M ethanolamine, pH 8.5. Kinetic measurements for affinity determinations of chimera were performed with size-exclusion purified proteins injected at concentrations ranging from 0.28 to 35.9 µM at a flow rate of 40 µl/min. Association and dissociation times were controlled at 3 and 10 min, respectively. After each binding cycle the chip was regenerated by injection of NaOH 5 mM for 30 s. Data were

evaluated using BIAeval software (BIAcore) with a heterogeneous binding model and a global fitting procedure (Karlsson and Falt, 1997).

RESULTS

To evaluate the possibility of designing target-binding proteins based on Sac7d by using CDR(s), a part of the CDR3 loop (10 residues) from a camel antibody specific for lysozyme was inserted at places where there is a loop in Sac7d, and then the effect of insertions on the chimeric proteins obtained was studied. This approach has previously been deployed for other potential artificial binding proteins, such as neocarzinostatin (NCS), single chain antibody (scFv), green fluorescent protein (GFP) and ubiquitin (Inoue *et al.*, 2011, Kiss *et al.*, 2006, Nicaise *et al.*, 2004, Saerens *et al.*, 2005).

Construction of the Sac7d protein mutants

Previous studies with CDR inserting/grafting have been performed using sequences of CDR3 of different lengths, truncated or not, originating from various camel anti-lysozyme antibodies. The CDR3 of cAb-Lys3 camel antibody specific for HEWL has been reported as ^{99}DSTIYASYYECGHGLSTGGYGYDS^{122} (Desmyter, *et al.*, 1996) with only the N-terminal part of this loop interacting with HEWL (Transue *et al.*, 1998) providing about 70% of contacts. In order to transfer only the region involved in HEWL recognition, the structure of cAb-Lys3 in complex with HEWL (PDB code: 1MEL) was analyzed with the "Protein interfaces, surfaces and assemblies" service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel and Henrick, 2007), which determines interfacing residues in a complex. The peptide ^{100}STIYASYYE^{108}

of the CDR3 region was identified as interacting with HEWL. A visual inspection of the structure of the complex confirmed this analysis, but also that Cys109 is found in the protruding part of this CDR3. To avoid an impaired cysteine, it was replaced with a serine. Thus the peptide ^{100}STIYASYYES^{109} was inserted onto three different loops of Sac7d orientated towards DNA binding site; between the β1β2 (mL1), β3β4 (mL2), and β4β5 (mL3) strands (Figure 1).

The WT Sac7d and chimeras were cloned into the expression plasmid pFP1001 (Mouratou *et al.*, 2012) and produced with His-tags. The proteins accumulated in the cytoplasm of *E. coli* strain DH5α at 30°C after overnight growth and could be purified to homogeneity in two steps by immobilized metal ion affinity chromatography and gel filtration as described previously (Mouratou, *et al.*, 2007). Since Sac7d has a DNA-binding activity, all the proteins were purified with buffers containing 500 mM NaCl in order to prevent co-purification of DNA. The production yields for WT Sac7d, mL2 and mL3 were 10, 15 and 9 mg per liter of shake-flask culture, respectively, while for mL1 the expression yield was drastically affected (0.25 mg/L).

Characterization of proteins

The proteins ran on a 15% acrylamide SDS/PAGE gel at the positions expected for their calculated molecular masses, and their elution volumes from the gel filtration column were consistent with monomeric species (data not shown).

Secondary structure contents of proteins

To evaluate whether anti-HEWL CDR3 insertions affected the structural integrity of Sac7d, far-UV CD spectra were recorded from 190 to 260 nm. All the spectra were

characteristic of mostly beta-sheet proteins (Figure 2A). CD spectra of chimera were all similar to the spectrum of WT Sac7d from this work and from a previous study (Edmondson and Shriver, 2001).

Secondary structure predictions from CD spectra were made using the Dichroweb software (Table 1). All proteins were predicted to contain percentages of β-sheet and α-helix comparable to those deduced from previously resolved crystallographic structures of WT Sac7d (PDB codes: 1AZP, 1CA5, 1WVL, 1WD1).

Overall, the CD analysis suggested that the structures of the three chimeras remained unchanged compared to the structure of WT Sac7d.

Thermal and pH stabilities of chimera compared to WT Sac7d

Temperature-induced unfolding of proteins was carried out by monitoring the ellipticity at 205 nm while gradually increasing the temperature (Figure 2B). Thermal stabilities of mL1, mL2 and mL3 chimeras determined by analysis of transition curves were of 66.3°C, 65.1°C and 60.9°C, respectively (Table 1). A T_m value of 90.4°C for WT Sac7d was previously reported (Edmondson and Shriver, 2001). These results indicate that CDR insertions in loops can be detrimental for Sac7d thermal stability.

Sac7d is an acidophilic (McCrary, *et al.*, 1996) and alkaliphilic protein (Behar, *et al.*, 2013). To investigate whether CDR3 insertion had an effect on Sac7d pH stability, CD spectra were recorded from 200 to 250 nm. It was observed that the CD spectra for all the proteins remained largely unchanged at pH from 0 to at least 12, indicating the presence of stable secondary structures (data not shown). CD spectra recorded at pH 14 were entirely different for all the proteins and were characteristic of disordered peptides. Figure 2C shows the residual unfolded fraction taking into account that the unfolded state is reached at pH 14 for all the proteins. Surprisingly, this study

indicates that, contrary to thermal stability, the pH stability of Sac7d was unaffected by CDR insertion at the three positions tested, further supporting the idea that chimeras were correctly folded.

Interaction of proteins with DNA

Sac7d is a non-specific DNA binder and thus we wanted to check the functional integrity of chimeras by measuring their ability to bind DNA. In addition, as chimeras were shown well structured and stable, this analysis could be indicative of conformations adopted by CDRs that could interfere with DNA binding surface due to steric hindrances.

Detection of DNA-binding activity by fluorescence

DNA-Sac7d interaction occurs by contact between the minor groove of DNA and the β-sheet surface composed of β3β4β5 strands. A hydrophobic motif positioned on this surface, which includes W24, is buried upon DNA-binding and its intrinsic fluorescence is quenched (McAfee*, et al.*, 1996, Su *et al.*, 2000). To determine their DNA-binding ability, all the proteins were incubated in the absence or presence of calf thymus DNA and the emission fluorescence spectra were recorded from 300 to 400 nm (Figure 3A). The tryptophan fluorescence intensity was significantly quenched for Sac7d in presence of DNA and a similar result was observed for mL3, whereas for mL1 and mL2 the quenchings were less important. Moreover, in agreement with structural analysis by CD, all the proteins showed the same maximal emission wavelength, suggesting marginal structural modifications, and concomitantly were slightly blue-shifted due to the burial of tryptophan upon DNA-binding.

To determine the DNA-binding affinity, a constant concentration of proteins was incubated with an increasing concentration of calf thymus DNA and the magnitude of fluorescence quenching was recorded (Figure 3B). A nonlinear regression analysis to fit the data was performed in order to calculate the affinity. A dissociation constant of 4.89 x 10^{-5} M was obtained for Sac7d (Table 1). This value is comparable to the affinity of 5.37 x 10^{-5} M previously reported for Sac7d and calf thymus DNA (McAfee, *et al.*, 1996). The mL3 protein showed a slightly lower dissociation constant of 6.25 x 10^{-5} M. It was not possible to determine the binding parameters for mL1 and mL2 proteins given that saturation was not reached, suggesting low affinities for DNA.

Detection of DNA-binding activity by electrophoretic mobility shift assay (EMSA)

DNA-binding ability was also studied by EMSA (Figure 3C). Increasing quantities of the proteins were incubated with 200 ng of DNA probe followed by electrophoresis. Figure 3C shows that the DNA probe was visibly shifted even for the lowest concentration of WT Sac7d tested (0.6 µM). No significant effect was observed for the same concentration with mL3, but the more the protein concentration increased the more it was retarded on the gel. No significant interaction for mLl and mL2 was observed at the higher concentration assayed.

Interaction of chimeras with HEWL

Finally, the interaction of chimeras with HEWL was investigated by ELISA and SPR. Specific binding was obtained by ELISA between mL3 and HEWL, while no signal was detected with WT Sac7d, mL1 or mL2 (Figure 4A). The affinity of mL3 for HEWL was determined by SPR (Table 1). Successive injections of increasing

concentrations of mL3 were carried out to obtain kinetic parameters. The sensorgram showed a hyperbolic association curve characteristic of an interaction, while the dissociation was fast (Figure 4B). The K_D value obtained was 6.7 μM.

DISCUSSION

In this study, we have demonstrated that the extremophilic protein Sac7d withstands insertion of ten residues at least for three of its short loops. This was achieved by using an element of recognition from another protein, i.e. a CDR3 loop fragment from an anti-HEWL camel antibody. Thereby, affinity transfer was possible for one of the three positions tested.

The secondary structure content and pH stability were similar for chimeras and WT Sac7d. Sac7d and its homologs show remarkable stabilities (temperature, pH, chemicals) which initially guided our choice to use them to derive artificial affinity proteins (Mouratou, *et al.*, 2007, Pecorari and Alzari, 2008). Several factors are known to contribute to the stability of extremophilic proteins, such as hydrophobic interactions, packing efficiency, salt-bridges, hydrogen bonds, reduction in the entropy of folding (Vieille and Gregory Zeikus, 1996) and some of them have been studied for Sac7d (Bedell *et al.*, 2005, Clark *et al.*, 2004, Clark *et al.*, 2007, Kahsai *et al.*, 2005, McCrary *et al.*, 1998). Loop extensions are expected to destabilize proteins because of the entropic cost associated with loop closure (Nagi and Regan, 1997, Scalley-Kim *et al.*, 2003). A comparative analysis of 20 complete genomes reported that thermophilic proteins tend to show deletions in exposed loop regions compared to their mesophilic homologs (Thompson and Eisenberg, 1999). There are no mesophilic homologs available for comparison with Sac7d (Razvi and Scholtz, 2006). In addition, it has been reported for RNase H from *Thermus thermophilus* and

Rubredoxin from *Pyrococcus furiosus*, which are two other well-characterized hyperthermostable proteins, that their short loops do not make an important contribution to stability (Hernandez *et al.*, 2000, Hollien and Marqusee, 1999). Nevertheless, superimposition of seven three-dimensional structures of Sac7d showed that residues in loops have the largest root mean squared deviations (Bosch *et al.*, 2003). Furthermore, an amide hydrogen exchange study reported that the three loop regions of Sac7d where we inserted the anti-HEWL CDR3 are the least stable and most flexible regions (with the end of the C-terminal helix) of the protein (Kahsai, *et al.*, 2005). We found that the midpoint temperature values of thermal unfolding for the chimeras were about 30°C lower than that of WT Sac7d. The drops in thermal stabilities for these mutants with extended loops are greater than for mutants of the hydrophobic core previously reported (Clark, *et al.*, 2004), indicating that short loops contribute to stability. This study demonstrated that CDR3 insertions were significantly destabilizing for the three positions tested. Remarkably, mL3, the least stable chimera (T_m = 60.9 °C) remained 9°C more stable than the mean T_m value determined for a cured dataset of 65 mesostable proteins with known three-dimensional structures (Folch *et al.*, 2010), and can thus still be regarded as a thermostable protein.

Taken together, these results support the idea that the Sac7d scaffold is tolerant to remodeling of its short loops, via introduction of a peptide stretch (here ten amino acids), without considerable structural modifications, but with a destabilizing effect that could be fatal for a protein less stable than Sac7d. As far as we know, this is the first report of CDR3 insertion for an archaeal extremophilic scaffold protein, here Sac7d.

Furthermore, the integrity of the DNA-binding function was demonstrated for mL3 by fluorescence measurements and EMSA, while for the other two chimeras it was drastically reduced. In addition, thermal unfolding of chimeras occurred cooperatively with a sharp transition for all chimeras, which is an indication of native-like structures (Roy and Hecht, 2000). Our results support the idea that alteration of DNA binding for mL1 and mL2 is more probably due to steric hindrances involving the CDR3 loops, which could be oriented toward the DNA-binding surface, than to an overall structural change of Sac7d folding. Supporting this hypothesis, we also observed in experimental three-dimensional structures of proteins with a similar topology to Sac7d that their natural longer loops (~ 6-8 residues) equivalent to mL2 have tendency to fold toward β-sheet 2 (Figure 5).

Another question we wanted to address in this work was: is the potential of each loop to introduce recognition of a target equivalent? The cAb-Lys3 camel antibody binds to HEWL with an affinity of 20 nM (Desmyter, *et al.*, 1996). We inserted the stretch of 10 amino acids identified as interacting with HEWL from the CDR3 of this antibody. According to our results, the three insertion positions tested in Sac7d (^{9}KG^{10}, ^{27}GK^{28} and ^{37}NG^{38}) were not equivalent. Chimera mL3 was able to bind HEWL as indicated by ELISA and SPR experiments with an affinity of 6.7 μM. This result confirms the possibility of transferring CDR3s with specific binding properties to alternative scaffolds such as Sac7d. Affinity values obtained here were similar to those previously recorded (from 0.5 μM to 7.6 μM) for cAb-Lys3 CDR3 grafting or insertion in NCS, scFv and GFP (Kiss, *et al.*, 2006, Nicaise, *et al.*, 2004, Saerens, *et al.*, 2005). Weak affinities were expected considering that other regions of the cAb-Lys3 antibody contributing to HEWL binding were not present in all the chimeras studied. Another explanation could be that the inserted CDR3 region did not have the

possibility of forming a disulfide bridge between the scaffold frameworks and the Cys of the CDR3 sequence (Inoue, *et al.*, 2011, Nicaise, *et al.*, 2004), constraining the CDR3 conformation. We hypothesize that the orientation of CDR3 was determinant to transfer HEWL binding into the mL3 protein. As we assume that loops 1 and 2 were bent causing steric hindrance for DNA binding, they probably did not have the optimal conformation to penetrate the active site of HEWL. The opposite effect was observed for loop 3, which seemed accessible for HEWL binding and did not interfere with DNA binding. Determining the structure of the chimeras might lead to a better understanding of the differences observed.

In summary, we report the successful transfer of binding properties by CDR insertion in an extremophilic scaffold protein. With this approach, together with our recent report (Correa *et al*. 2014), we were able to design binding sites into Sac7d that recognize concave epitopes. Considering the small size of Sac7d (7 kDa), its tolerance to loop remodeling is remarkable. We anticipate that CDR(s) from antibodies in combination with randomized surface(s) and/or loop(s) may be used as potential recognition elements in Sac7d to help selections to converge toward binding of the corresponding epitope on antigen of interest.

ACKNOWLEDGMENTS

We thank Dr. Yves-Henri Sanejouand and Johann Hendrickx (University of Nantes, FRE3478 CNRS) for helpful discussions.

FUNDING INFORMATION

This work was supported by the Pasteur Institute (Paris, France), with a postdoctoral fellowship for SP, and by “La Région des Pays de la Loire”.

CONFLICT-OF-INTEREST DISCLOSURE

F.P. is co- inventor of a patent that has been filed by the Institut Pasteur and CNRS describing generation of binders from Sac7d and its homologs. F.P. is a co-founder of a spin-off company of the Institut Pasteur / CNRS / Université de Nantes that has a license agreement.

REFERENCES

Batori V., Koide A. and Koide S. (2002) *Protein Eng,* **15**, 1015-1020.

Bedell J.L., Edmondson S.P. and Shriver J.W. (2005) *Biochemistry,* **44**, 915-925.

Behar G., Bellinzoni M., Maillasson M., Paillard-Laurance L., Alzari P.M., He X., Mouratou B. and Pecorari F. (2013) *Protein Eng Des Sel,* **26**, 267-275.

Bes C., Troadec S., Chentouf M., Breton H., Dominique Lajoix A., Heitz F., Gross R., Pluckthun A. and Chardes T. (2006) *Biochem Biophys Res Commun,* **343**, 334-344.

Beste G., Schmidt F.S., Stibora T. and Skerra A. (1999) *Proc Natl Acad Sci U S A,* **96**, 1898-1903.

Binz H.K., Amstutz P., Kohl A., Stumpp M.T., Briand C., Forrer P., Grutter M.G. and Pluckthun A. (2004) *Nat Biotechnol,* **22**, 575-582.

Binz H.K., Amstutz P. and Pluckthun A. (2005) *Nat Biotechnol,* **23**, 1257-1268.

Bosch D., Campillo M. and Pardo L. (2003) *Journal of computational chemistry,* **24**, 682-691.

Buddelmeijer N., Krehenbrink M., Pecorari F. and Pugsley A.P. (2009) *J Bacteriol,* **191**, 161-168.

Clark A.T., McCrary B.S., Edmondson S.P. and Shriver J.W. (2004) *Biochemistry,* **43**, 2840-2853.

Clark A.T., Smith K., Muhandiram R., Edmondson S.P. and Shriver J.W. (2007) *J Mol Biol,* **372**, 992-1008.

Correa A., Pacheco S., Mechaly Ariel E., Obal G., Béhar G., Mouratou B., Oppezzo P., Alzari P.M. and Pecorari F. (2014) *PLoS One,* **9**, e97438.

Desmyter A., Transue T.R., Ghahroudi M.A., Thi M.H., Poortmans F., Hamers R., Muyldermans S. and Wyns L. (1996) *Nat Struct Biol,* **3**, 803-811.

Edmondson S.P. and Shriver J.W. (2001) *Methods Enzymol,* **334**, 129-145.

Folch B., Dehouck Y. and Rooman M. (2010) *Biophys J,* **98**, 667-677.

Gebauer M. and Skerra A. (2009) *Curr Opin Chem Biol,* **13**, 245-255.

Greenfield N.J. (2006) *Nature protocols,* **1**, 2527-2535.

Hernandez G., Jenney F.E., Adams M.W. and LeMaster D.M. (2000) *Proc Natl Acad Sci U S A,* **97**, 3166-3170.

Hollien J. and Marqusee S. (1999) *Biochemistry,* **38**, 3831-3836.

Holm L. and Rosenström P. (2010) *Nucleic Acids Res,* **38**, W545-W549.

Inoue H., Ilhara A., Takahashi H., Shimada I., Ishida I. and Maeda Y. (2011) *Protein Sci,* **20**, 1971-1981.

Kahsai M.A., Martin E., Edmondson S.P. and Shriver J.W. (2005) *Biochemistry,* **44**, 13500-13509.

Karlsson R. and Falt A. (1997) *J Immunol Methods,* **200**, 121-133.

Kiss C., Fisher H., Pesavento E., Dai M., Valero R., Ovecka M., Nolan R., Phipps M.L., Velappan N. and Chasteen L. (2006) *Nucleic Acids Res,* **34**, e132-e132.

Koide A., Bailey C.W., Huang X. and Koide S. (1998) *J Mol Biol,* **284**, 1141-1151.

Koide A., Wojcik J., Gilbreth R.N., Hoey R.J. and Koide S. (2012) *J Mol Biol,* **415**, 393-405.

Krehenbrink M., Chami M., Guilvout I., Alzari P.M., Pecorari F. and Pugsley A.P. (2008) *J Mol Biol,* **383**, 1058-1068.

Krissinel E. and Henrick K. (2007) *J Mol Biol,* **372**, 774-797.

Kumar S. and Nussinov R. (2001) *Cellular and Molecular Life Sciences CMLS,* **58**, 1216-1233.

Ladurner A.G. and Fersht A.R. (1997) *J Mol Biol,* **273**, 330-337.

McAfee J.G., Edmondson S.P., Zegar I. and Shriver J.W. (1996) *Biochemistry,* **35**, 4034-4045.

McCrary B.S., Bedell J., Edmondson S.P. and Shriver J.W. (1998) *J Mol Biol,* **276**, 203-224.

McCrary B.S., Edmondson S.P. and Shriver J.W. (1996) *J Mol Biol,* **264**, 784-805.

Mouratou B., Behar G., Paillard-Laurance L., Colinet S. and Pecorari F. (2012) *Methods Mol Biol,* **805**, 315-331.

Mouratou B., Schaeffer F., Guilvout I., Tello-Manigne D., Pugsley A.P., Alzari P.M. and Pecorari F. (2007) *Proc Natl Acad Sci U S A,* **104**, 17983-17988.

Nagi A.D. and Regan L. (1997) *Fold Des,* **2**, 67-75.

Nicaise M., Valerio-Lepiniec M., Minard P. and Desmadril M. (2004) *Protein Sci,* **13**, 1882-1891.

Nord K., Gunneriusson E., Ringdahl J., Stahl S., Uhlen M. and Nygren P.A. (1997) *Nat Biotechnol,* **15**, 772-777.

Pecorari F. and Alzari P.M. (2008) *Patent Publication Nos PCT/IB2007/004388.*

Razvi A. and Scholtz J.M. (2006) *Protein Sci,* **15**, 1569-1578.

Roy S. and Hecht M.H. (2000) *Biochemistry,* **39**, 4603-4607.

Saerens D., Pellis M., Loris R., Pardon E., Dumoulin M., Matagne A., Wyns L., Muyldermans S. and Conrath K. (2005) *J Mol Biol,* **352**, 597-607.

Scalley-Kim M., Minard P. and Baker D. (2003) *Protein Sci,* **12**, 197-206.

Schilling J., Schöppe J. and Plückthun A. (2014) *J Mol Biol,* **426**, 691-721.

Su S., Gao Y.G., Robinson H., Liaw Y.C., Edmondson S.P., Shriver J.W. and Wang A.H. (2000) *J Mol Biol,* **303**, 395-403.

Thompson M.J. and Eisenberg D. (1999) *J Mol Biol,* **290**, 595-604.

Transue T.R., De Genst E., Ghahroudi M.A., Wyns L. and Muyldermans S. (1998) *Proteins,* **32**, 515-522.

Vieille C. and Gregory Zeikus J. (1996) *Trends Biotechnol,* **14**, 183-190.

Whitmore L. and Wallace B. (2004) *Nucleic Acids Res,* **32**, W668-W673.

Whitmore L. and Wallace B.A. (2008) *Biopolymers,* **89**, 392-400.

LEGENDS TO FIGURES

Figure 1. Structure and sequence of Sac7d. The structure of Sac7d, which folds as a SH3-like five-stranded incomplete β-barrel capped by a C-terminal α-helix, is represented in cartoons. Arrows indicate the position where the anti-HEWL CDR3 of the cAb-Lys3 camel antibody was inserted. The β-sheet composed of β3β4β5-strands is the DNA-binding surface.

Figure 2. Analysis of secondary structure and stability of chimeric proteins. Far-UV spectra (A), temperature- (B) and pH-induced unfolding (C) of Sac7d and chimeric proteins analyzed by circular dichroism. Line codes in the graphs correspond to Sac7d (—), mL1 (——), mL2 (···) and mL3 (–·–).

Figure 3. Analysis of the interaction between DNA and chimeric proteins. (A) Intrinsic tryptophan fluorescence in absence (black line) or presence (gray line) of calf thymus DNA. (B) Determination of DNA-binding affinity of Sac7d and chimera. The fluorescence quenching was measured at 345 nm. Quenching values are plotted versus increasing concentration of calf thymus DNA. Line codes in the graphs correspond to Sac7d (—), mL1 (——), mL2 (···) and mL3 (–·–). (C) Electrophoretic mobility shift assays (EMSA). The DNA probe obtained by PCR was incubated with increasing concentrations of proteins and subjected to electrophoresis in polyacrylamide gel.

Figure 4. Analysis of HEWL binding. (A) ELISA binding assay of Sac7d and chimeric proteins (5 μM) against HEWL (black bars) or casein (white bars) immobilized on the plate. The error bars correspond to the interval of confidence of the mean of three measurements. (B) SPR analysis of mL3 binding to HEWL.

Different concentrations of mL3 (0.28, 0.56, 1.13, 2.25, 4.5, 8.97, 17.95, 35.9 μM), followed by washing with buffer flow are shown. The global fit is indicated by dashed lines.

Figure 5. Superimpositions of experimental structures of proteins structurally homologous to Sac7d having a loop equivalent to mL2 with 6-8 residues. (A) Sac7d is colored in green (pdb code 1wd1), homologs in yellow (pdb codes 1dz1, 1e0b, 3kup, and 3q6s), loops of homologs equivalent to mL2 loop in red and DNA bound by WT Sac7d in grey. (B) Zoom-in of (A) to show how are folded loops of homologs. (C) Another orientation of superimpositions, without structures of homologs for clarity. Homologous proteins were identified using Sac7d structure (pdb code 1wd1) and Dali server (Holm and Rosenström, 2010). All structural representations were prepared with PyMOL (www.pymol.org).

TABLES

Table 1. Secondary structure content, parameter of thermal stability and affinity of Sac7d and chimeric proteins.

Protein	α-Helix (%)	β-Sheet (%)	Turns (%)	Unordered (%)	T_m [a] (°C)	DNA binding[a] (x 10^{-5} M)	HEWL binding[b] (x 10^{-6} M)
Sac7d	6	30	29	35	90.4[c]	4.89 (±0.6)	ND
mL1	6	32	32	30	66.3 (±0.4)	ND	ND
mL2	7	35	26	32	65.1 (±0.4)	ND	ND
mL3	10	30	28	32	60.9 (±0.4)	6.25 (±2.4)	6.7

[a]Values calculated from experimental data. The points were fitted to nonlinear regression curves and the quality criteria were r squared >0.98 and small values for the standard deviation indicated in parentheses. Data represent the result of two independent experiments.

[b]K_D value recorded by SPR experiments.

[c] From Edmondson and Shriver (2001)

N.D.: not detectable

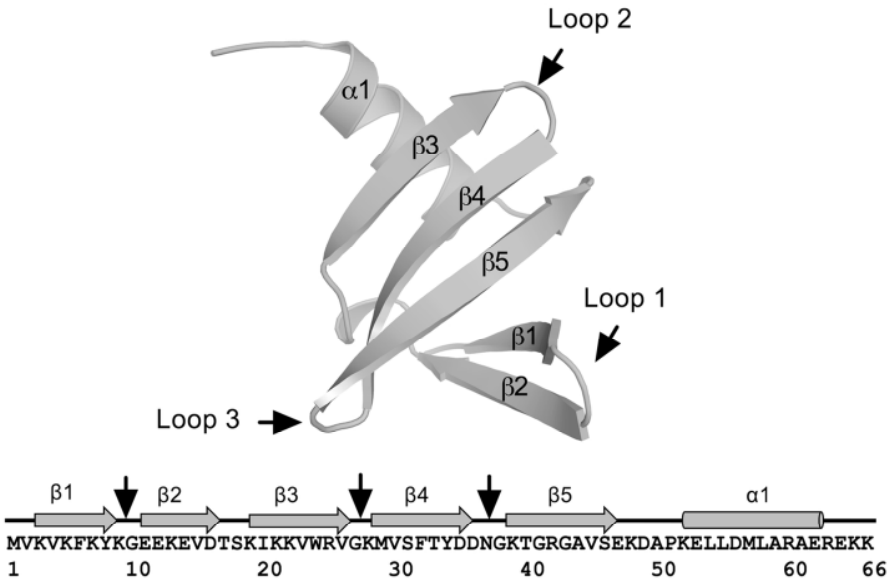

Figure 1. Structure and sequence of Sac7d. The structure of Sac7d, which folds as a SH3-like five-stranded incomplete β-barrel capped by a C-terminal α-helix, is represented in cartoons. Arrows indicate the position where the anti-HEWL CDR3 of the cAb-Lys3 camel antibody was inserted. The β-sheet composed of β3β4β5-strands is the DNA-binding surface.
118x78mm (300 x 300 DPI)

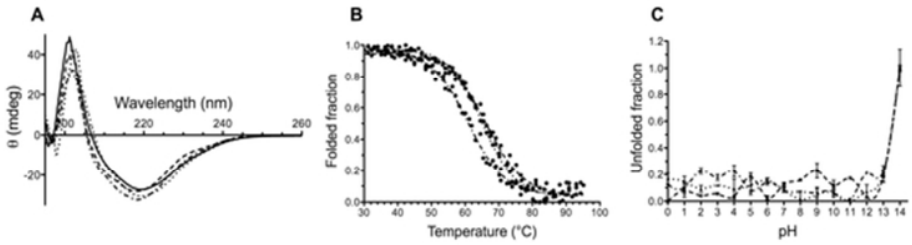

Figure 2. Analysis of secondary structure and stability of chimeric proteins. Far-UV spectra (A), temperature- (B) and pH-induced unfolding (C) of Sac7d and chimeric proteins analyzed by circular dichroism. Line codes in the graphs correspond to Sac7d (—), mL1 (---), mL2 (···) and mL3 (—·—).
50x14mm (300 x 300 DPI)

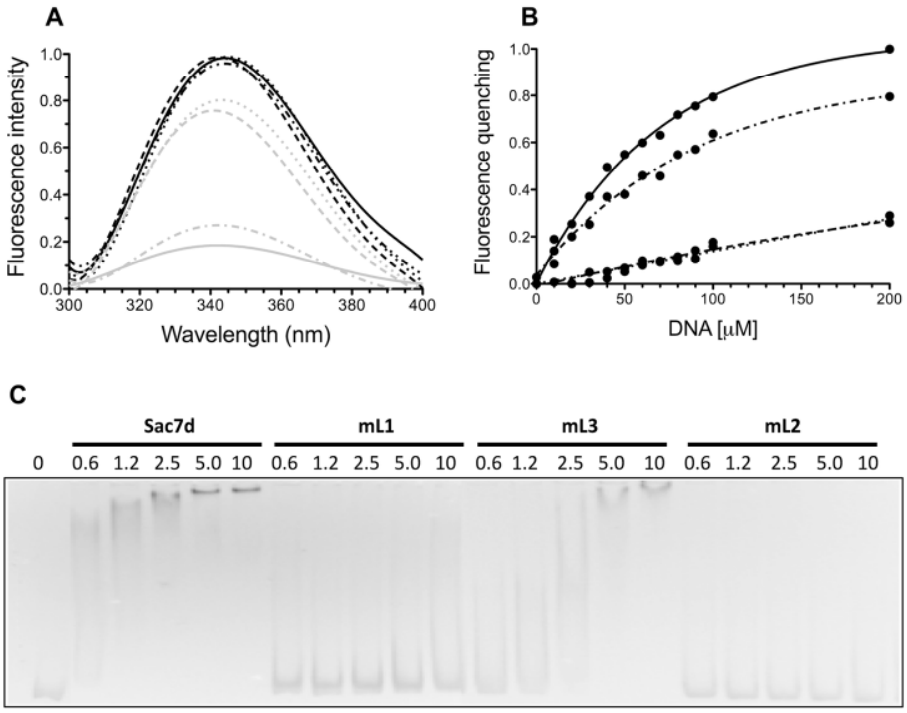

Figure 3. Analysis of the interaction between DNA and chimeric proteins. (A) Intrinsic tryptophan fluorescence in absence (black line) or presence (gray line) of calf thymus DNA. (B) Determination of DNA-binding affinity of Sac7d and chimera. The fluorescence quenching was measured at 345 nm. Quenching values are plotted versus increasing concentration of calf thymus DNA. Line codes in the graphs correspond to Sac7d (—), mL1 (---), mL2 (···) and mL3 (—·—). (C) Electrophoretic mobility shift assays (EMSA). The DNA probe obtained by PCR was incubated with increasing concentrations of proteins and subjected to electrophoresis in polyacrylamide gel.
135x103mm (300 x 300 DPI)

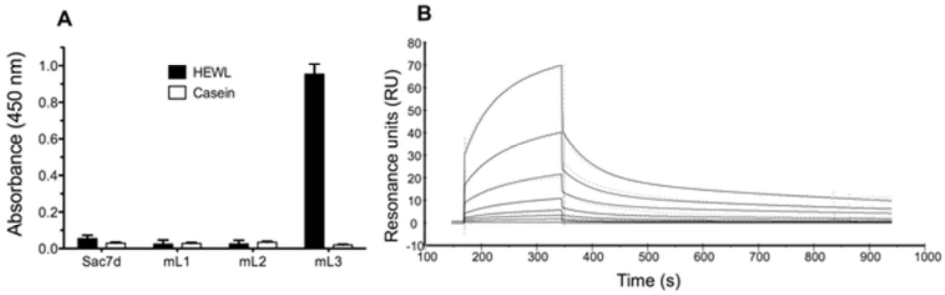

Figure 4. Analysis of HEWL binding. (A) ELISA binding assay of Sac7d and chimeric proteins (5 μM) against HEWL (black bars) or casein (white bars) immobilized on the plate. The error bars correspond to the interval of confidence of the mean of three measurements. (B) SPR analysis of mL3 binding to HEWL. Different concentrations of mL3 (0.28, 0.56, 1.13, 2.25, 4.5, 8.97, 17.95, 35.9 μM), followed by washing with buffer flow are shown. The global fit is indicated by dashed lines.
58x19mm (300 x 300 DPI)

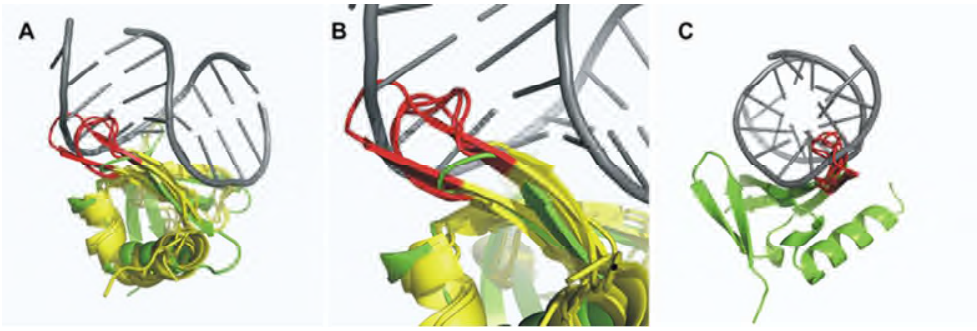

Figure 5. Superimpositions of experimental structures of proteins structurally homologous to Sac7d having a loop equivalent to mL2 with 6-8 residues. (A) Sac7d is colored in green (pdb code 1wd1), homologs in yellow (pdb codes 1dz1, 1e0b, 3kup, and 3q6s), loops of homologs equivalent to mL2 loop in red and DNA bound by WT Sac7d in grey. (B) Zoom-in of (A) to show how are folded loops of homologs. (C) Another orientation of superimpositions, without structures of homologs for clarity. Homologous proteins were identified using Sac7d structure (pdb code 1wd1) and Dali server (Holm and Rosenström, 2010). All structural representations were prepared with PyMOL (www.pymol.org).
60x19mm (300 x 300 DPI)