

Delineation of structural domains and identification of functionally important residues in DNA repair enzyme exonuclease VII

Katarzyna Poleszak¹, Katarzyna H. Kaminska¹, Stanislaw Dunin-Horkawicz^{1,2},
Andrei Lupas², Krzysztof J. Skowronek¹ and Janusz M. Bujnicki^{1,3,*}

¹Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ks. Trojdena 4, PL-02-109 Warsaw, Poland, ²Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany and ³Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, PL-61-614 Poznan, Poland

Received February 25, 2012; Revised May 12, 2012; Accepted May 14, 2012

ABSTRACT

Exonuclease VII (ExoVII) is a bacterial nuclease involved in DNA repair and recombination that hydrolyses single-stranded DNA. ExoVII is composed of two subunits: large XseA and small XseB. Thus far, little was known about the molecular structure of ExoVII, the interactions between XseA and XseB, the architecture of the nuclease active site or its mechanism of action. We used bioinformatics methods to predict the structure of XseA, which revealed four domains: an N-terminal OB-fold domain, a middle putatively catalytic domain, a coiled-coil domain and a short C-terminal segment. By series of deletion and site-directed mutagenesis experiments on XseA from *Escherichia coli*, we determined that the OB-fold domain is responsible for DNA binding, the coiled-coil domain is involved in binding multiple copies of the XseB subunit and residues D155, R205, H238 and D241 of the middle domain are important for the catalytic activity but not for DNA binding. Altogether, we propose a model of sequence–structure–function relationships in ExoVII.

INTRODUCTION

Environmental agents and endogenous metabolic processes involving DNA constantly challenge the chemical structure and stability of the genome. Lesions that are not repaired and errors that are not corrected may lead to mutations, disease and cell death, however

they are also the main source of genetic variability and therefore a driving force for evolution. To maintain genome integrity and control genetic variability, living organisms have evolved various biochemical systems for DNA repair (1–4). The key players in these systems are enzymes that catalyze reactions leading from damaged DNA to a repaired molecule. The knowledge of DNA repair enzymes is critical to our understanding of how cells control the integrity of their genomes.

One of the primary pathways of DNA repair is mismatch repair (MMR), a post-replicative process that removes errors introduced during replication (mismatched nucleotides, small loops, insertions and deletions) (5). In *Escherichia coli*, the first step of MMR is the recognition of the error in the newly replicated DNA by the MutS protein, followed by binding of a ‘molecular matchmaker’ MutL, which then recruits the MutH endonuclease to the complex. MutH introduces a nick in the unmethylated daughter strand at the nearest hemimethylated GATC site, which can be located up to 1000 bp from the mismatch. The nicked DNA is then unwound by helicases and excised by single-strand (ssDNA)-specific exonucleases (6,7). The final step is the re-synthesis of DNA by polymerase III and ligation of the nick by DNA ligase.

MMR in *E. coli* engages four ssDNA-specific exonucleases: ExoI and ExoX degrade DNA from the 3′ end (8,9), RecJ from the 5′ end (9) and finally, exonuclease VII (ExoVII) from both the 5′ and 3′ ends (10). These enzymes remove the DNA strand that bears the mismatch and depending on the localization of the nick in the DNA, exonucleases with different polarities are required (11). The repair system is impaired when all four exonucleases are inactive, and consequently, the bacteria are unable to correct errors resulting from

*To whom correspondence should be addressed. Tel: +48 22 597 0750; Fax: +48 22 597 0715; Email: iamb@genesilico.pl

mismatches as well as frameshifts (11–13). *E. coli* strains without two (ExoI and ExoVII) or three (ExoI, ExoVII and RecJ) active exonucleases show an increased rate of frameshift mutations while the number of mismatches does not change. There are two alternative hypotheses that try to explain this observation. The first one suggests that in the absence of exonucleases specific for the 3' end, the frameshift intermediates are not degraded, resulting in an increased frequency of frameshift mutations (13). The second one considers that the frameshift mutator phenotype could be due to the excess of ssDNA in the cell, which causes the induction of the SOS system (14).

ExoVII also plays a role in homologous recombination and depending on the genetic background it can decrease or increase the frequency of recombination. In *recD recJ* double mutants, ExoVII activity is crucial for recombination, suggesting that it can substitute for RecJ. In mutants lacking 3' → 5' exonucleases (ExoI and ExoX), ExoVII decreases the rate of recombination, which supports the observation that the 5' → 3' exonuclease activity of ExoVII is more efficient than its 3' → 5' activity (15).

Despite the importance of ExoVII, its structure and mechanism of action have remained largely unknown since its discovery in 1974 by Chase and Richardson (16,17). ExoVII from *E. coli* is a protein complex comprising two subunits: a large subunit XseA (51.8 kDa) and a small subunit XseB (10.5 kDa), encoded by the *xseA* and *xseB* genes, respectively (10,18). It has been estimated that the complex is composed of one XseA subunit and four XseB subunits from densitometric analysis of protein bands in Coomassie-stained polyacrylamide gels (10). ExoVII from *E. coli* catalyses the degradation of ssDNA in the absence of metal ions and is active in the presence of EDTA (10). ExoVII from *Thermotoga maritima*, however, which is composed of subunits homologous to those in *E. coli*, has been found to require Mg²⁺ for activity (19). Structural information would help characterize and understand the ExoVII mechanism of action in related enzymes from this family.

MATERIALS AND METHODS

Bioinformatic analyses

Searches for XseA homologs in the non-redundant (nr) sequence database at NCBI were carried out using PSI-BLAST (20) with the e-value threshold of 1e⁻³⁰. In order to verify the reliability of the selected cut-off value and to visualize sequence similarities, we chose the clustering tool cluster analysis of sequences (CLANS), which uses the *P*-values of high scoring segment pairs (HSPs) obtained from an all-against-all BLAST search to compute attractive and repulsive forces between each sequence pair and to move the sequences according to the force vectors resulting from all pairwise interactions (21). The cluster of closest homologs of *E. coli* XseA was extracted. The multiple sequence alignment of the XseA family was calculated using PROMALS (22) with default parameters and refined by hand to ensure that no unwarranted gaps had been introduced within α -helices and

β -strands. Based on the alignment, the phylogenetic tree was calculated using MEGA 4.0 (23), employing the Minimum Evolution method with the JTT model of substitutions. The stability of individual nodes was calculated using the interior branch test (1000 replicates) and confirmed by the bootstrap test (data not shown).

Protein structure prediction (including identification of domains, prediction of coiled coils, disorder, secondary structure and fold-recognition (FR), i.e. alignment with proteins of known structures) was carried out via the GeneSilico metaserver gateway (24). Initial predictions were done for the whole XseA protein, and subsequently for the individual domains. Modeling of the XseA structure was carried out with the homology modeling approach using Modeller 9v7 (25), followed by the optimization with REFINER (26). For the region with no template available (residues 260–396), a provisional *de novo* model was built depicting its secondary structure, albeit without tertiary interactions. Model quality was assessed by MetaMQAP (27) and ProQ (28). Mapping of sequence conservation onto the XseA model and XseB crystal structure was done using the corresponding XseA and XseB multiple sequence alignments with the ConSurf server (29). The multiple sequence alignment and the model were also used to plan site-directed mutagenesis experiments.

XseA–XseB interactions

We assigned the hydrophobic core of XseB based on the investigation of the available structure from *Bordetella pertussis* (PDB code 1VP7, doi:10.2210/pdb1vp7/pdb). Using the buildali.pl script (with -cov 90 option) from the HHsearch 1.5.0 package (30), we obtained homologs and multiple sequence alignments for the XseB family (query sequence gi:67462835, residues 11–69). The same procedure was used for the XseA coiled-coil domain (gi:16130434, residues 266–394). Based on these alignments, we calculated the average hydrophobicity score using the scale of Kyte and Doolittle (31). In order to assess the amount of variability in the number of coiled-coil helices in the coiled-coil regions of XseA proteins, we measured the lengths of sequences localized between the putative catalytic domain and the conserved C-terminal domain.

Cloning and mutagenesis of *xseA* and *xseB* genes

The *E. coli* *xseA* and *xseB* genes have been cloned previously with non-cleavable N-terminal His6 tags into the recombinant expression plasmid pCA24N by Saka *et al.* (32), resulting in the pXseANHis and pXseBNHis plasmids which were obtained from the ASKA re-cloned library [NBRP (NIG, Japan): *E. coli*]. Both *xseA* and *xseB* have also been inactivated in the *E. coli* K-12 strain BW25113 by Baba *et al.* (33), and the knock-out strains (Δ *xseA* and Δ *xseB*) were obtained from the Keio library. The *xseA* gene was amplified in a PCR reaction and cloned into pXseBNHis as a HindIII–SalI fragment, resulting in a construct (pXseABNHis), expressing full-length XseB with an N-terminal His6 tag and XseA without a tag. Site-directed mutagenesis of the *xseA* gene in pXseABNHis construct was performed by a PCR-based

technique. The mutants were sequenced and found to contain only the desired mutations. The pXseACHisB plasmid, expressing full-length XseA with a C terminal His6 tag and XseB without a tag, was constructed by modifying pXseABNHis in a PCR reaction. Deletion mutagenesis of *xseA* gene in pXseABNHis and pXseACHisB was also carried out by PCR. Plasmids pXseANHis Δ 104-C-term and pXseANHis Δ 1–103 were constructed by amplifying *xseA* fragments corresponding to amino acid residues 1–103 and 104–456 of XseA and recloning these as a HindIII–SalI inserts into the pCA24N vector, resulting in constructs expressing selected domains of XseA C-terminally fused to His6 tag. All constructs are described in Table 1.

Protein expression and purification

Proteins were expressed from plasmids carrying *xseA*, *xseB* and both genes together, in *E. coli* strains Δ *xseA* or Δ *xseB* after 1 mM IPTG induction at 25°C overnight. Cells were harvested by centrifugation (4000 g for 15 min, 4°C) and pelleted. The cell pellet was first washed with STE buffer (10 mM Tris–HCl, pH 8.0, 150 mM NaCl and 1 mM EDTA), re-suspended and lysed by sonication in binding buffer [BB; 50 mM sodium phosphate, pH 8.0, 0.3 M NaCl, 10 mM imidazole, pH 8.0, 10% (v/v) glycerol, 10 mM BME and 1 mM PMSF]. Proteins were purified using Ni-NTA

Agarose beads (Sigma-Aldrich). The purification was carried out at 4°C. Proteins from the clarified lysates were bound to the Ni-NTA resin, the beads were washed with BB, wash buffer 1 (BB supplemented with 2 M NaCl) and wash buffer 2 (BB with 20 mM imidazole). The protein was eluted with elution buffer (BB with 250 mM imidazole, pH 8.0). To confirm the presence and identity of proteins in the eluted fractions, samples of eluates were resolved in 15% SDS–PAGE and screened in western blots with anti-His6 antibody–horseradish peroxidase conjugate (Sigma-Aldrich) followed by chemiluminescent detection. Protein concentration was determined based on densitometry of protein bands in Coomassie Brilliant Blue–stained SDS–PAGE gels.

In vitro cleavage assay

In vitro activity assays were carried out in a 10 μ l reaction volume in cleavage buffer (70 mM Tris–HCl, pH 8.0, 8 mM EDTA, 10 mM BME and 50 μ g/ml BSA) with 1 μ M of 70 nt oligonucleotide with random sequence (70N) as a substrate and 1 μ M of purified protein variant. Digestion was performed for 30 min at 37°C.

Electrophoretic mobility shift assay

70 N oligonucleotide was end-labeled with [γ -³²P]ATP (PerkinElmer, Life Science) using T4 polynucleotide

Table 1. Constructs and their applications

Protein name	Plasmid	Expressed proteins	Expressed in strain	Variant	Experiment
XseA	pXseANHis	His6–XseA	Δ <i>xseB</i>	–	<i>In vitro</i> activity assay Filter binding
XseB	pXseBNHis	His6–XseB	Δ <i>xseA</i>	–	<i>In vitro</i> activity assay Filter binding
ExoVII wt	pXseABNHis	XseA His6–XseB	Δ <i>xseA</i>	–	<i>In vitro</i> activity assay
ExoVII variants	pXseABNHis	XseA	Δ <i>xseA</i>	D155A R205A H238A D241A Q177A D246A D250A T255A Δ 397–456	<i>In vitro</i> activity assay Filter binding
ExoVII wt	pXseACHisB	His6–XseB XseA–His6 XseB	Δ <i>xseA</i>	–	<i>In vitro</i> activity assay XseB binding Gel filtration
ExoVII deletion variants	pXseACHisB	XseA–His6	Δ <i>xseA</i>	Δ 267–301 Δ 307–349 Δ 353–393 Δ 267–301, 307–349 Δ 267–301, 353–393 Δ 307–349, 353–393 Δ 267–393	<i>In vitro</i> activity assay XseB binding
ExoVII lacking the OB-fold domain (Δ 1–103)	pXseACHisB Δ 1–103	XseB Δ 1–103 XseA	Δ <i>xseA</i>	Δ 1–103	<i>In vitro</i> activity assay Filter binding
XseA variant Δ 104-C-term	pXseANHis Δ 104-C-term	His6– Δ 104-C-term XseA	Δ <i>xseA</i>	F63A R64E, R68E, R69E Q96A	<i>In vitro</i> activity assay EMSA Filter binding
XseA variant Δ 1–103	pXseANHis Δ 1–103	His6– Δ 1–103 XseA	Δ <i>xseB</i>	–	<i>In vitro</i> activity assay Filter binding

kinase. Following incubation for 1 h at 37°C, oligonucleotide was purified by ethanol precipitation. Binding reactions contained 200 nM labeled oligonucleotide and a series of concentrations (5–200 nM) of XseA variant Δ 104-C-term, encompassing only the OB-fold domain, in the cleavage buffer. The reaction mixtures were incubated for 30 min on ice and subjected to electrophoresis (60 V, 4 h, 8°C) on 8% polyacrylamide gels with 10% glycerol and Tris–borate–EDTA buffer.

Nitrocellulose filter binding assay

70 N oligonucleotide was end-labeled with [γ -³³P]ATP (PerkinElmer, Life Science) using the same procedure as described for the preparation of 70 N for the electrophoretic mobility shift assay (EMSA) experiment. Binding reactions were carried out in 50 μ l cleavage buffer without BSA, they contained 0.1 nM labeled oligonucleotide and 20 nM of one of the following proteins: wild type (wt) XseA, XseA variants: Δ 104-C-term, Δ 1–103, wt XseB, ExoVII (XseA–XseB complex) variants with one of the following substitutions in XseA: D155A, R205A, H238A, D241A or with the deletion of the N-terminal domain Δ 1–103. The binding reactions were incubated for 30 min on ice and filtered through 0.22 μ m nitrocellulose filters (Whatmann) in a Dot-Blot apparatus (Bio-Rad) (34). Each well was washed three times with 200 μ l of cleavage buffer without BSA. Dried filters were exposed to a phosphorimager screen overnight. Images were scanned on a Storm PhosphorImager and the retained radioactivity was quantified using ImageQuant software (Amersham). The measurements were performed three times for each protein.

Gel filtration

Gel filtration was performed for ExoVII protein expressed from a plasmid carrying genes encoding XseA–His6 and XseB (Table 1) in a Δ *xseA* strain. The protein was purified by Ni-NTA Agarose column chromatography. Next, it was loaded on a Superose 6L 3.2 PC column (GE Healthcare). The analysis was carried out in gel filtration buffer (GF; 50 mM sodium phosphate, pH 8.0, 150 mM NaCl and 10% glycerol) using an ÄKTA™ Purifier chromatography system (GE Healthcare). The protein was also analyzed on a Superose 6L 3.2 PC column using a GF buffer with 2 M urea. For this analysis, the protein was prepared by incubation with 2 M urea for 30 min at 4°C. Samples of eluates were assayed for activity. The Superose 6L 3.2 PC column was calibrated using Sigma-Aldrich protein standards (albumin 66 kDa, alcohol dehydrogenase 150 kDa, amylase 200 kDa, apoferritin 443 kDa and thyroglobulin 669 kDa).

RESULTS

Sequence analysis of the XseA family

Searches of the nr sequence database with PSI-BLAST (as of September 2011) revealed 2298 homologs of XseA that exhibited conservation over the essentially entire protein length, with the lowest similarity in the C-terminal helical region (see Figure 1 for a multiple sequence

alignment of representative sequences and Supplementary File 1 for a complete alignment). In these searches, the N-terminus of XseA showed remote sequence similarity to other proteins containing OB-fold domains, including members of the RecJ nuclease family. This was expected, as the XseA sequence entry in GenBank is already annotated with the N-terminal OB-fold domain (ExoVII_LU_OBF). However, the remaining parts of the protein sequence exhibited no significant similarity to any known protein domains; hence we attempted to predict their structure using the FR approach.

The GeneSilico protein fold-recognition metasever (24) confirmed the assignment of the OB-fold to the N-terminal domain (residues 1–119); according to the consensus predictor PCONS, all top predictions had scores 0.9973–0.3147 and reported the structure of human replication protein A (RPA; PDB code: 1quq) as the best template for modeling.

For the region following the OB-fold domain (residues 120–260), none of the individual methods queried via the GeneSilico metasever returned highly scored predictions. However, PCONS indicated with high confidence (all top 10 predictions, with scores from 1.57 to 1.10) that this region is structurally similar and potentially homologous to the catalytic domain of oxidoreductases from the dehydroquinase synthase-like superfamily (SCOP code: e.22.1.2). Such scores indicate that the structural fold has been correctly identified with \gg 95% confidence and that the software identified no serious alternative among known 3-D folds. Based on the consensus prediction, the structure of iron-containing alcohol dehydrogenase (TM0920) from *Thermotoga maritima* (PDB code: 1o2d) was selected as the best template for modeling the XseA middle domain. The core of this domain is formed by a parallel 4-stranded β -sheet, with a relative strand order 2134, flanked by four α -helices that connect the β -strands with each other. The architecture of this domain resembles the Rossmann fold commonly found in dinucleotide binding enzymes; however, its architecture is distinct due to the different topology of connections between secondary structure elements. The analysis of the sequence conservation in the predicted oxidoreductase-like domain revealed the presence of an universally conserved glycine-rich motif (residues 206–208 in *E. coli*), which forms a loop between strand β 3 and helix α 3 of the oxidoreductase-like domain. The corresponding motif present in iron-containing alcohol dehydrogenase from *T. maritima* is formed by residues at positions 94–96 and occurs in the loop connecting strand β 4 and helix α 6. Interestingly, a similar glycine-rich turn and a following α -helix constitute a characteristic nucleotide recognition locus in dehydrogenases (35). The glycine-rich motif can be further extended to include surrounding residues conserved in both template and target sequences fitting the consensus motif DxxxVGxGGGSxxD (DVLIVGRGGGSLED: residues 199–212 in XseA from *E. coli*, DLVMIVRGGGSKED: residues 193–206 in XseA from *T. maritima* and DFVVLGGGSPMD: residues 88–100 in the iron-containing alcohol dehydrogenase). Sequence comparison with *T. maritima* iron-containing alcohol dehydrogenase revealed the

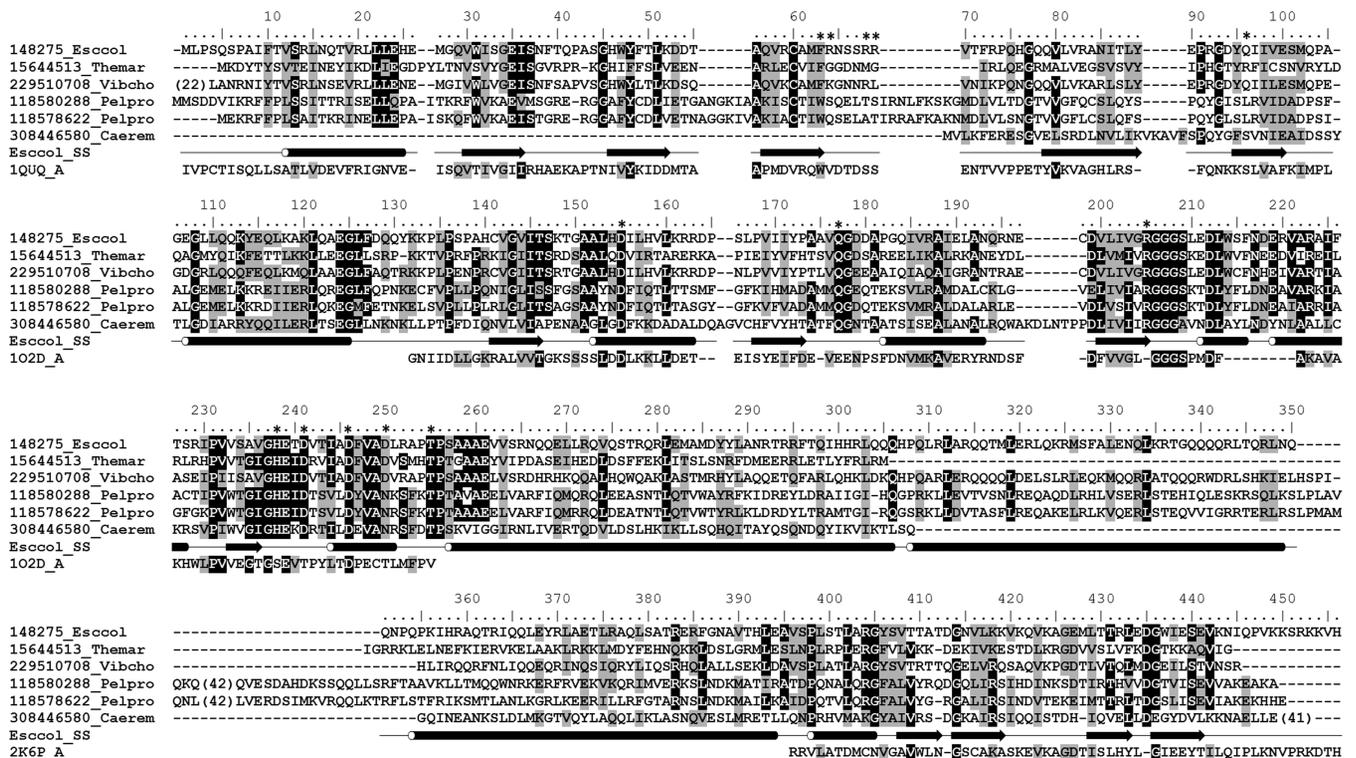


Figure 1. Multiple sequence alignment of five representative members of the XseA family. Proteins are indicated by the NCBI GI number and the abbreviated genus and species name. Conserved residues or residues whose physicochemical character is conserved in >51% of sequences in the XseA family (the complete alignment of 132 representatives is available as a Supplementary File 1) are highlighted in black and gray, respectively. Consensus secondary structure (SS) calculated by the GeneSilico metaserver is indicated below the alignment as tubes (helices) and arrows (strands) for the *E. coli* XseA protein. Residues of *E. coli* XseA selected as targets for alanine substitutions are marked by asterisks (*). Below the alignment of XseA representatives, template structures used for the modeling of the three domains are indicated (1QUQ:4-107, 1O2D: 33-178 and 2K6P: 15-75). For the clarity of the presentation, template residues without counterparts in XseA sequences are not shown.

presence of two other highly conserved motifs: VxTxK (residues 144–148 and 33–37 in XseA and iron-containing alcohol dehydrogenase, respectively) and xPVV (residues 230–234 in XseA and 129–132 in the template structure).

The region C-terminal to the oxidoreductase-like domain (residues 261–389) is predicted to form three long α -helices that have a propensity to form coiled-coils, according to PCOILS (36). FR methods failed to assign any reliable template for homology modeling. Therefore, for this part of the protein, the provisional model has been built based solely on the secondary structure prediction.

For the C-terminal region (residues 390–456), several FR servers reported, albeit with low scores, matches to the alpha-L domain, in particular to the functionally uncharacterized protein HP1423 from *Helicobacter pylori*—PDB code: 2k6p. The canonical structure of this domain is composed of two β -hairpins and two α -helices that form an L-shaped meander together with the loop between β 2 and β 3 strands. According to the metaserver results, the alpha-L domain and the C-terminal domain of XseA share similar secondary structure composition and length (~60 aa), and harbor a conserved GD box motif placed in the loop joining two-hairpins (37).

In order to provide a structural framework for further analyses of XseA proteins, we built a molecular model comprising the OB-fold domain, the oxidoreductase

domain, and the alpha-L domain generated by homology (template-based) modeling. According to MetaMQAP, the predicted global root mean square deviations (RMSD) of the modeled domains with respect to the (unknown) native counterparts are: 3.8 (OB-fold domain), 2.6 (oxidoreductase domain) and 4.5 (alpha-L domain). According to ProQ, predicted LGscore of the three domains are: 2.4, 6.1 and 1.8, respectively (LGscore > 1.5 > 3.0 > 5 means that a model is likely to be correct, good or very good, respectively). Such scores indicate that the global fold of the two first domains and the mutual position of most residues are likely to be accurate, but the atomic details (e.g. conformations of side-chains) should be taken with a grain of salt. The remaining part of the molecule has been generated according to the predicted secondary structure (see ‘Materials and Methods’ section for details). The relative position of domains is currently unknown and must be considered purely arbitrary. Figure 2 shows the resulting model, and illustrates the distribution of sequence conservation.

Analysis of the model in the light of features such as fold and sequence conservation, suggested that the N-terminal domain is likely to be involved in DNA binding, while the middle domain (residues 120–260) may be the catalytic core of ExoVII. These two functional predictions were subsequently tested by a series of

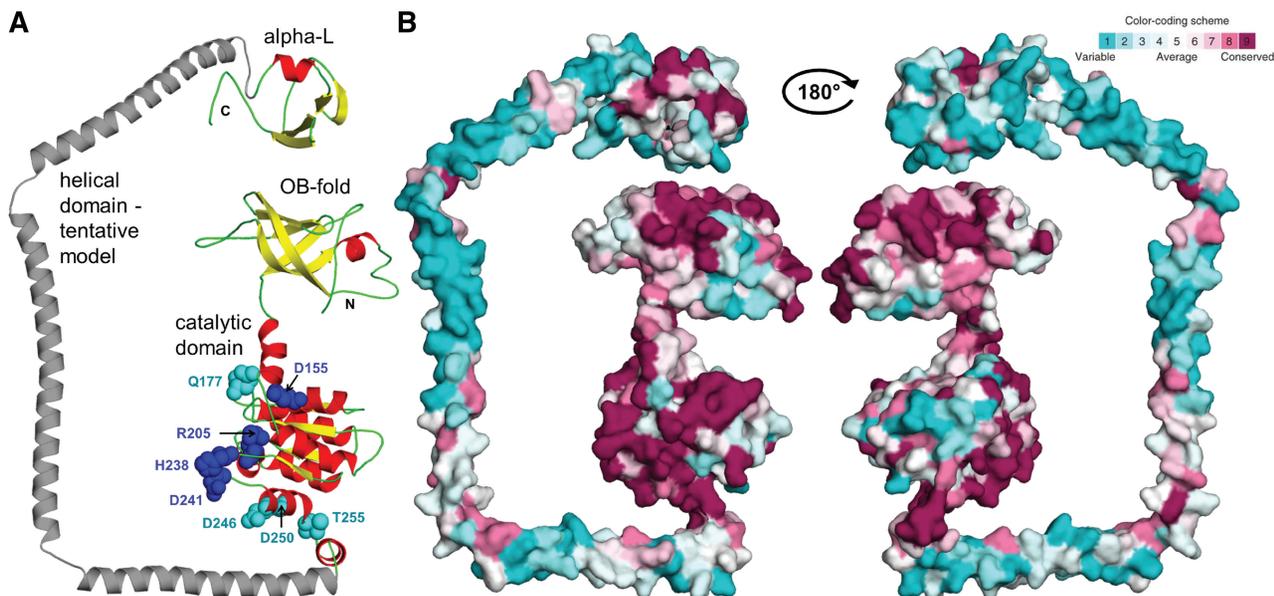


Figure 2. Structural model of XseA. (A) The model in the 'cartoon' representation, with helices, strands and loops colored red, yellow and green, respectively. The tentatively modeled helical region (residues 261–389) is shown in gray. Residues of XseA substituted to alanine are shown as spheres. Potentially catalytic residues are colored in blue. Substitution of residues, colored in cyan, to alanine reduced the activity to the level of about 60–80% of the wt enzyme (see 'Discussion' section). The relative position of domains is currently unknown and must be considered purely arbitrary. Coordinates are available for download from the FTP server <ftp://genesilico.pl/iamb/models/XseA/>. (B) The model in two orientations, shown in the surface representation, illustrating conservation of surface-exposed residues (with the color-coding scheme presented in the upper-right corner of the figure).

experiments (see below). The function of the C-terminal domain could not be unambiguously assigned, as the homologs reported with similar scores are involved in a variety of molecular functions.

Phylogenetic analyses of XseA proteins

In order to elucidate phylogenetic relationships in the XseA family, we calculated the Minimum Evolution phylogenetic tree based on an alignment of 132 representative sequences (see Supplementary File 2). The tree shown in Figure 3 reveals that the XseA family can be divided into two unequal subfamilies. The major subfamily comprises members from many bacterial taxa (such as Proteobacteria, Firmicutes, Deferribacteres, Euryarcheota, Tenericutes and others), and it includes the two experimentally characterized members of the XseA family, from *E. coli* and *T. maritima*. Therefore, position of sequences in the tree roughly agrees with the taxonomy of the host organisms (e.g. with *Thermotoga* XseA located on a relatively deep branch, remote from all other members, including *E. coli* XseA and most of Proteobacteria separated from Firmicutes). However, numerous horizontal transfer events are evident from well-supported branches that group together sequences from distant taxa; in particular, Deltaproteobacteria appear promiscuously, suggesting multiple independent transfers. An alternative hypothesis for the observed distribution of species in the major branch of the XseA tree is an extremely uneven rate of divergence, leading to separate long-branch attraction events.

The minor branch comprises only a few, experimentally uncharacterized sequences from Proteobacteria and,

interestingly, one eukaryotic member of the XseA family, a protein from the nematode *Caenorhabditis remanei*. Horizontal gene transfer from bacteria to nematode genomes and its functional relevance has been documented (38). However, we cannot exclude the possibility that the respective XseA-like gene has been derived from a bacterial DNA impurity during the *C. remanei* sequencing project. The nematode XseA homolog lacks the N-terminal OB-fold domain, and its genome encodes no detectable XseB homolog. Thus, it is unclear if this protein is enzymatically active, although it possesses all conserved residues characteristic of the XseA family.

Most organisms encode only one member of the XseA family (either from the major or from the minor branch), among the 132 organisms analyzed here only *Pelobacter propionicus* DSM 2379 encodes three XseA homologs (two from the minor branch, one from the major one), and there are five cases with two members, one each in the major and minor branches: *Polaromonas naphthalenivorans* CJ2, *Saccharophagus degradans* 2–40, *Sorangium cellulosum* 'So ce 56', *Burkholderia vietnamiensis* G4 and *Geobacter lovleyi* SZ. In all fully sequenced genomes we analyzed, the number of XseB homologs is equal to the number of XseA homologs (data not shown), e.g. *P. propionicus* DSM 2379 encodes three XseB-like proteins, suggesting that the evolution of XseA and XseB is strongly correlated.

Prediction of XseA–XseB interactions

To gain more information about the structural arrangement of the ExoVII complex, we attempted to predict how its subunits might interact with each other. The crystal

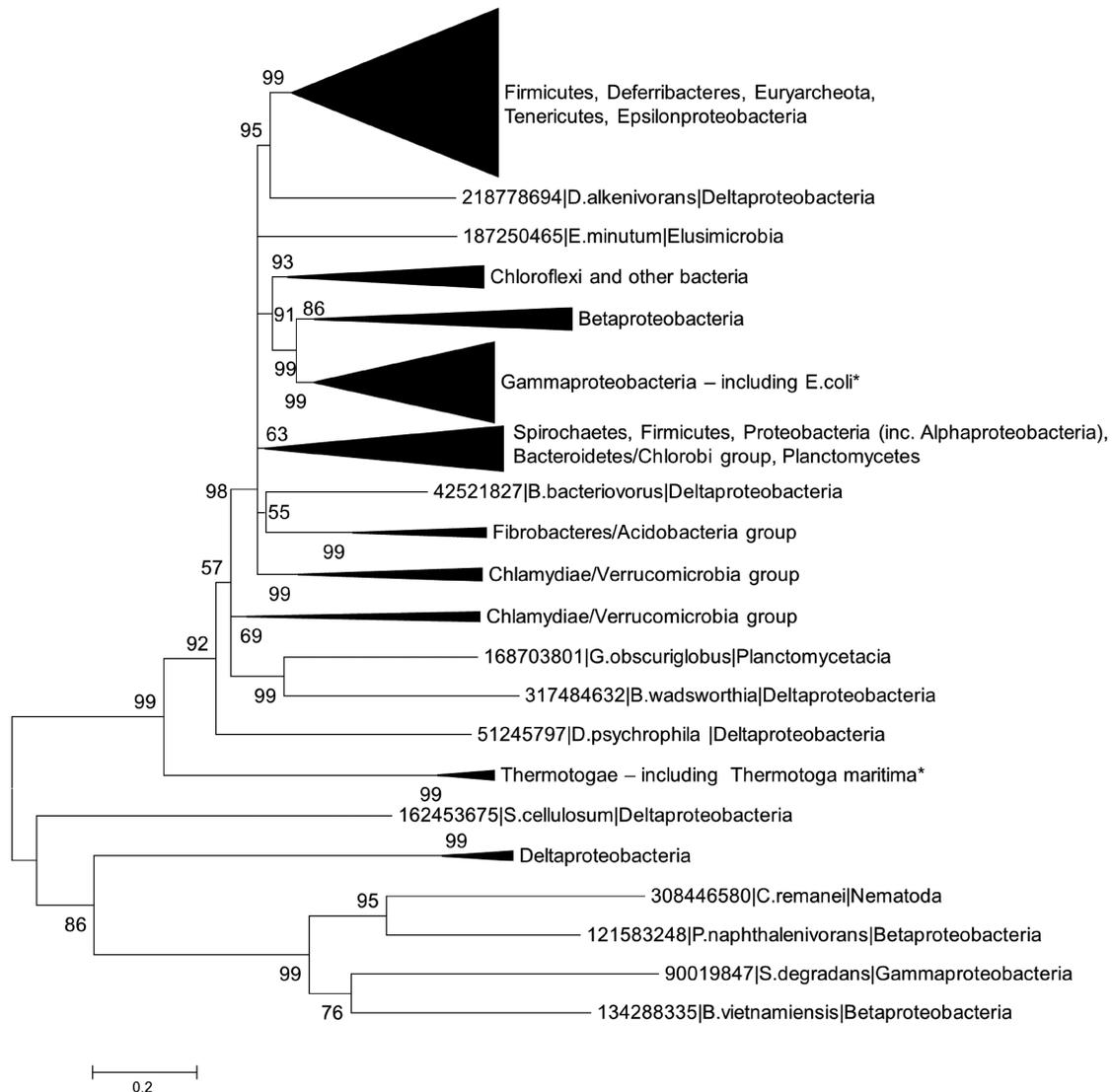


Figure 3. A minimum evolution phylogenetic tree of the XseA family. The branches of the tree are indicated by their representatives. Values at the nodes indicate the percent value of the statistical support. Branches dominated by particular phyla have been collapsed and are shown as triangles. Two experimentally characterized members of the family (as of February 2012) are shown by asterisks (*). A full (non-collapsed) version of the tree is available in the MEGA format as a Supplementary File 2.

structure of XseB from *B. pertussis* has been determined (PDB ID 1vp7, doi:10.2210/pdb1vp7/pdb), revealing a dimer with four helices arranged in an unusual three-helical bundle, in which two shorter helices are stacked on one side (Figure 4B). The hydrophobic core of this bundle adopts a knobs-to-knobs packing that is characteristic for a subclass of coiled-coil structures (39). In these coiled coils, the hydrophobic core is formed by three positions (canonical coiled coils use two), which assume two distinct geometries: an *x* layer, where the side-chains point towards the core of the bundle, and a *da* layer, where the two side-chains point side-ways, enclosing a central cavity. The results of sequence analysis indicate that this packing mode is a conserved feature of the XseB family. Further investigation of the XseB structure revealed that some of the hydrophobic residues are exposed to the solvent and form a conserved

hydrophobic surface (Figure 4A). Due to its solvent-exposed nature, this area is the most probable site of interaction with XseA.

E. coli XseA contains three consecutive helices with elevated coiled-coil forming propensity. The number of these helices varies among members of the XseA family and they do not exhibit any correlated sequence substitutions (data not shown), which suggests that they have not evolved to interact with each other, but may interact with other coiled-coil domains, e.g. XseB. Sequence analysis revealed that these helices in XseA show a hydrophobicity pattern that strongly resembles the one observed in XseB (Figure 4A). In addition, we found that each helix in XseA contains a centrally located single-residue insertion. In coiled coils such insertions, called skip residues, can be accommodated without disruption of the helices by delocalization over three heptads to produce two

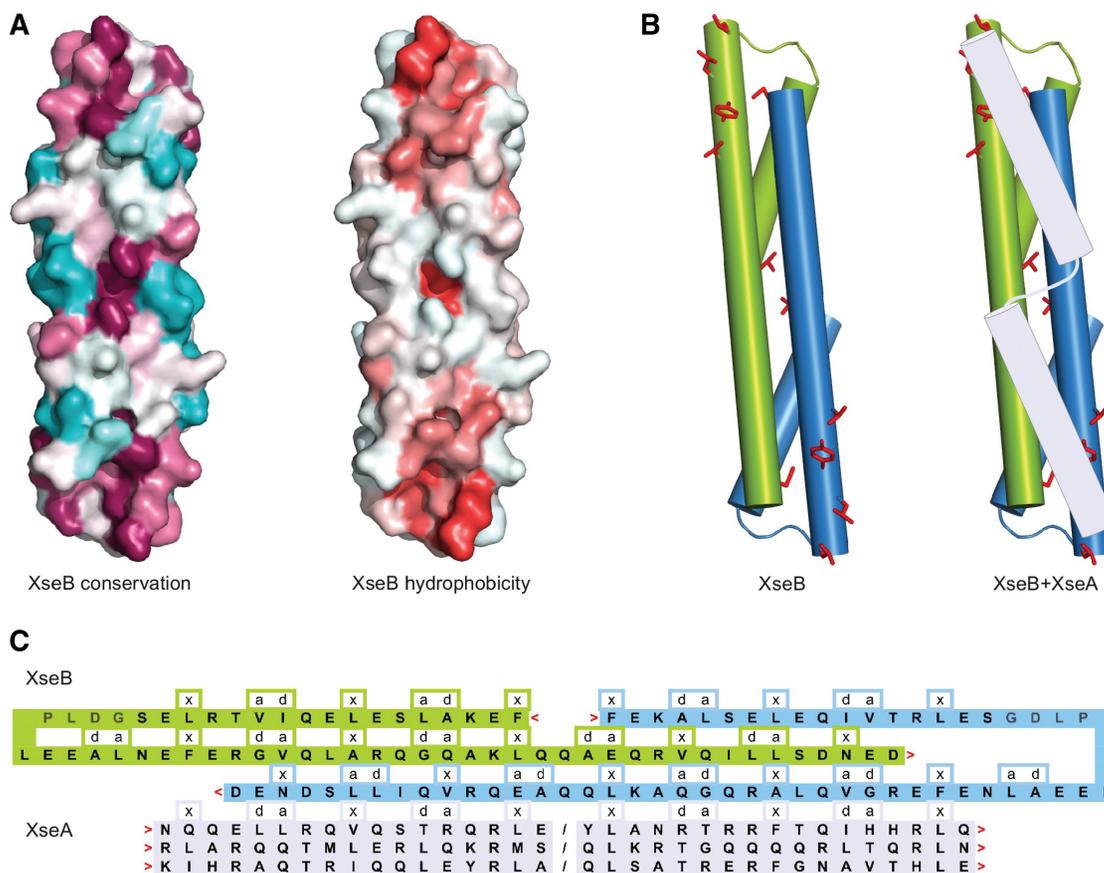


Figure 4. Properties of the coiled-coil regions of *E. coli* XseA and XseB subunits. **(A)** Surface representation of the XseB dimer structure colored according to the sequence conservation and the average hydrophobicity. **(B)** Schematic representation of the topology of the XseB dimer (left) and a putative model of XseA–XseB complex (right). Conserved hydrophobic residues of XseB predicted to be involved in XseA binding are shown in red. **(C)** Sequences of the XseB dimer and the three coiled-coil forming XseA helices. Letters indicate positions that participate in the formation of the hydrophobic core. Relative positioning of the sequences reflects their structural arrangement.

hendecads [$3 \times (7/2) + 1 = 22/6 = 2 \times (11/3)$] or by delocalization over two heptads to produce one pentadecad [$2 \times (7/2) + 1 = 15/4$], both of which have characteristic patterns of hydrophobic residues and cause a local shift in the handedness of the supercoil (40). Skip residues can however also be accommodated by a disruption in the conformation of the helix backbone and, since the observed pattern of hydrophobicity of XseA does not support a local 11/3 or 15/4 periodicity, we conclude that the skip residue disrupts the helical structure. We thus propose a model for the XseA–XseB interaction, in which the helices of XseA do not form a stable arrangement by themselves; rather, each of them individually interacts with an XseB dimer and forms a four-helical bundle (Figure 4). This model explains the presence of the skip residues in the XseA helices: the break caused by them allows XseA helices to adopt a structure that complements the two short helices of XseB.

Structure–function relationships in *E. coli* XseA

XseA, XseB and the ExoVII complex were expressed from plasmids coding for either XseA–His6 and XseB or for XseA and His6–XseB (Table 1), and were purified to ~80% homogeneity. The activity of these proteins was tested in the *in vitro* cleavage assay, which showed that

the subunits separately are inactive, but when co-expressed and co-purified they do form an active complex (Supplementary Figure S1).

Based on the analysis of residues that are conserved in the XseA family and are on the same face of the model of the purported oxidoreductase-like domain, we selected candidates that were likely to be important for DNA cleavage by that domain (due to their involvement in DNA binding and/or catalysis). To test the model-based predictions, we carried out a biochemical characterization of ExoVII variants obtained by site-directed mutagenesis (Table 1). The *in vitro* activity assay revealed that ExoVII variants: D155A, R205A, H238A, D241A lost nucleolytic activity (Figure 5).

In order to gain more information about the function of the conserved C-terminal region of XseA, we constructed a deletion variant $\Delta 397$ –456 (Table 1). Results of the *in vitro* activity assay showed that this deletion inactivates the ExoVII enzyme (Figure 5).

Our bioinformatics analyses supported the prediction of an N-terminal OB-fold domain in XseA, presented earlier by Larrea *et al.* (19). In order to validate the functional prediction that this domain is responsible for DNA binding, XseA variants comprising the isolated OB-fold domain ($\Delta 104$ -C-term) or XseA without the OB-fold

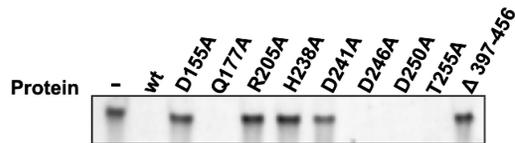


Figure 5. Activity of wt and variants of ExoVII (variants of XseA in complex with wt XseB). Gels showing DNA degradation by ExoVII wt and variants in an *in vitro* activity assay. 70 N DNA oligonucleotide was digested with purified enzymes, resolved in 8% polyacrylamide gels with 8 M urea and stained with ethidium bromide. Amino acids substitutions and regions deleted (Δ) in XseA are shown above each line, '-' refers to the untreated substrate.

domain ($\Delta 1-103$) were assayed for their ability to bind ssDNA. The OB-fold domain alone (XseA variant $\Delta 104$ -C-term) is able to bind ssDNA, according to both the EMSA and the filter-binding analysis (Figure 6). On the other hand, XseA with the OB-fold domain removed ($\Delta 1-103$) is unable to bind DNA either alone or in complex with XseB (Figure 6). In the *in vitro* activity assay, the ExoVII variant lacking the OB-fold domain ($\Delta 1-103$) showed a complete loss of the nucleolytic activity. Additionally, the filter-binding assay revealed that the full-length XseA alone or XseB alone are also unable to bind DNA, while the ExoVII variants with substitutions in the putative catalytic domain of XseA (and nucleolytically inactive) do retain the DNA-binding ability (Figure 6B). To test the functional relevance of residues predicted to take part in DNA binding according to the model, we measured the DNA binding capacity of the OB-fold ($\Delta 104$ -C-term) variants: F63A, Q96A (single substitutions) and a triple substitution variant (R64E/R68E/R69E). The OB-fold variant with triple substitutions and F63A variant showed a decreased DNA-binding activity (Figure 6B) whereas the Q96A substitution had a mild effect.

To validate the prediction that the helical segment of XseA (region 267–393) is involved in XseB binding and to determine which helices are involved in the interaction, we constructed a series of ExoVII deletion variants that lack one of the helices: $\Delta 1$ ($\Delta 267-301$ aa), $\Delta 2$ ($\Delta 307-349$ aa), $\Delta 3$ ($\Delta 353-393$ aa), variants with two helices removed: $\Delta 12$ ($\Delta 267-301$, $307-349$ aa), $\Delta 13$ ($\Delta 267-301$, $353-393$ aa), $\Delta 23$ ($\Delta 307-349$, $353-393$ aa) and a variant without all three helices: $\Delta 123$ ($\Delta 267-393$ aa; Figure 7A). Since the C-terminally His6-tagged XseA co-purifies with XseB, we assayed the XseA deletion variants for this ability. XseB does not co-purify with XseA without the helical domain ($\Delta 123$). Variants with a single helix deleted ($\Delta 1$, $\Delta 2$ and $\Delta 3$) bound about 70% and variants with two helices deleted ($\Delta 12$, $\Delta 13$ and $\Delta 23$) about 50% of the XseB levels compared to the full-length XseA (Figure 7B). All ExoVII variants mentioned above exhibited a complete loss of exonucleolytic activity *in vitro*. These results are in line with the proposed model and suggest that each helix of XseA binds one XseB dimer (possibly with some co-operativity). The fact that homologs of XseA frequently contain different numbers of helices in the corresponding region further supports the notion of their modular function.

It has been concluded from densitometric analysis of protein bands in Coomassie-stained polyacrylamide gels

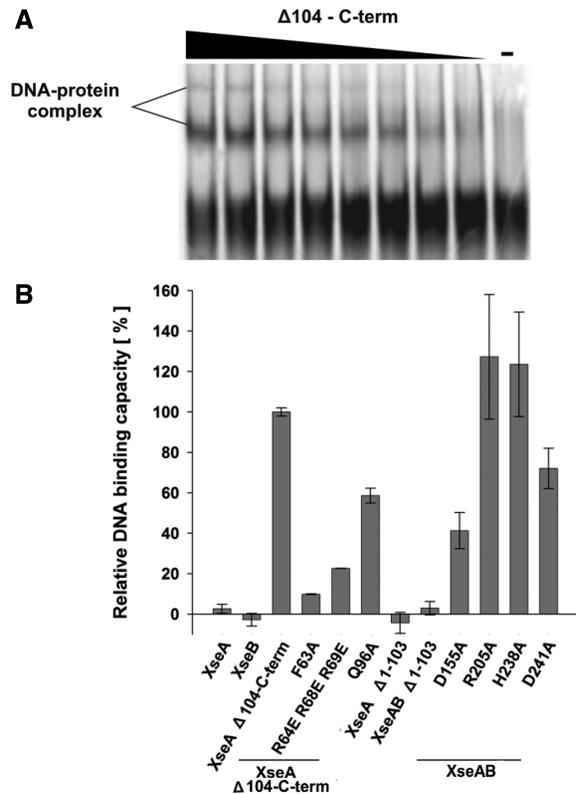


Figure 6. DNA-binding activity of wt and variants of ExoVII (XseA and XseB complex). (A) DNA binding of XseA variant $\Delta 104$ -C-term. EMSA experiments were performed with 70 N oligonucleotide end-labeled with $[\gamma\text{-}^{33}\text{P}]$ ATP and a series of XseA variant $\Delta 104$ -C-term concentrations (5–200 nM), '-' refers to untreated substrate. (B) Relative DNA binding of XseA variants and ExoVII variants. The DNA-binding capacity was measured in a filter binding assay. Binding activities are expressed in relation to that of XseA variant $\Delta 104$ -C-term (which was set to 100%) and standard deviations (error bars) were calculated from three independent measurements. Background binding level was measured in a reaction without protein and this was subtracted from our results.

that ExoVII from both *E. coli* and *T. maritima* consists of one large subunit and four small subunits (10,19). In contrast, our size exclusion experiments performed on Superose 6 L 3.2 PC columns yielded large oligomers for ExoVII, with an estimated molecular weight about 660 kDa. We were able to disrupt these by incubation with 2 M urea (the presence of urea at this concentration did not affect the enzymatic activity of ExoVII, data not shown). Subsequent chromatography in a buffer containing 2 M urea yielded an additional peak with a molecular weight corresponding to ~ 109 kDa. This may correspond to the complex of a single XseA subunit and six XseB subunits (calculated molecular weight 115 kDa; Figure 8). All fractions containing ExoVII collected from size exclusion chromatography exhibited the nuclease activity.

DISCUSSION

ExoVII was discovered 35 years ago, but its structure and molecular mechanism have remained substantially unexplored (16). ExoVII comprises two subunits, XseA

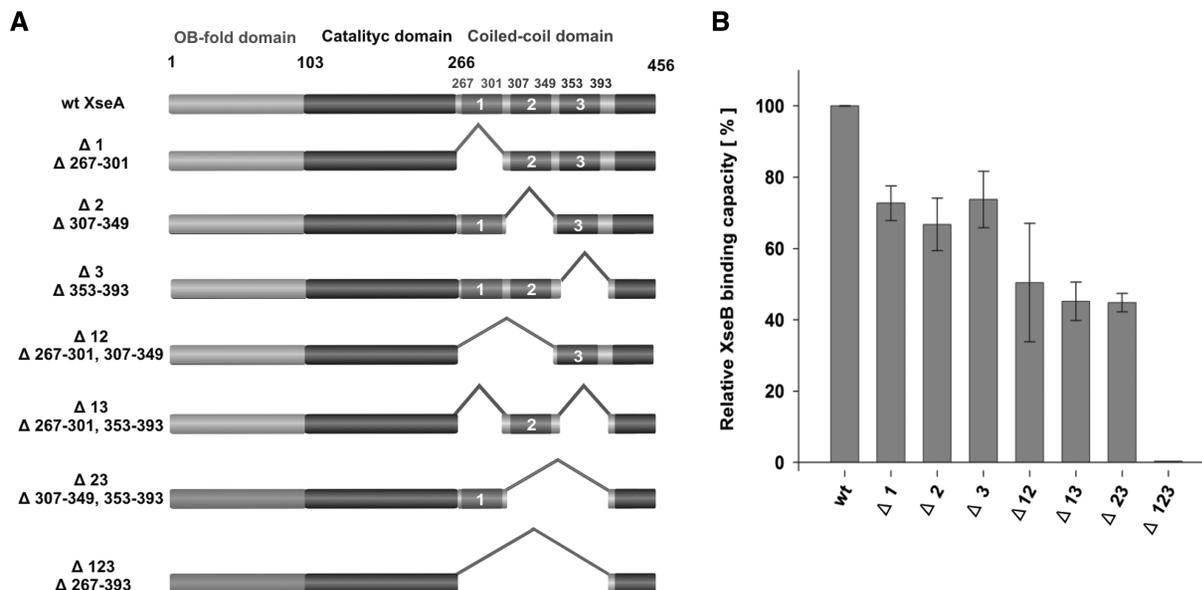


Figure 7. XseB-binding capacity of the coiled-coil domain of XseA. (A) Schematic representation of XseA deletion variants used for XseB interaction mapping. (B) Relative XseB-binding capacity of XseA wt and deletion variants. XseB binding was determined by the extent of XseB co-purification with His6 tagged XseA variants, and measured by the densitometry analysis of purified proteins in Coomassie Brilliant Blue-stained SDS-PAGE gels. The relative binding capacity was calculated in relation to that of wt XseA (which was regarded as 100%).

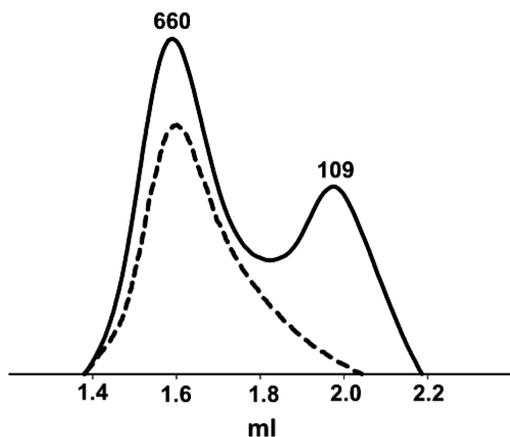


Figure 8. Analytical gel filtration of ExoVII. Elution profiles of ExoVII on Superose 6L 3.2 PC column in GF buffer (dashed line) and GF buffer with the addition of 2M urea (continuous line).

and XseB. The crystal structure of XseB from *B. pertussis* has been determined (doi:10.2210/pdb1vp7/pdb). Recently, Larrea *et al.* (19) characterized experimentally the XseA/B homologs from *Thermatoga maritima*, TM1768 and TM1769. They predicted that the large subunit of ExoVII is composed of two domains: the N-terminal OB-fold domain and a C-terminal domain termed 'ExoVII_Large'. Here, we present a structural model for XseA, as well as for the XseA–XseB interactions that lead to formation of the ExoVII complex. Our results show that XseA consists of four domains: the OB-fold domain, the catalytic domain, a helical domain and a C-terminal extension.

We demonstrated that residues D155, R205, H238 and D241 are essential for the nuclease activity of *E. coli*

ExoVII. The substitution of these residues inactivated the enzyme, but they did not hinder DNA binding. In the structural model of the putative catalytic domain, they are located on the same face of the protein; they do not form a very tight cluster, which can be attributed to the relative precision of the model. Thus, we propose that all or some of these residues belong to the active site of XseA. Recently, it has been shown that D235 and D240 residues are essential for the nuclease activity of *T. maritima* ExoVII (19). Based on these findings and sequence analyses of XseA family, the authors showed that these residues are conserved in *E. coli* (D241, D246) but also in other XseA homologs, and that the conserved region has a motif RGGG(x)nGHxxDxxxxD. The motif identified by Larrea *et al.* (19) includes residues R205, H238 and D241 of *E. coli* XseA. In this study, we have shown that D155, which is located outside of the conserved motif, is also important for the activity of the *E. coli* ExoVII enzyme. On the other hand, we demonstrated that a substitution of residue D246 in *E. coli* XseA did not inactivate the enzyme. This residue corresponds to D240 in *T. maritima*, which is essential for the activity of that enzyme. Larrea *et al.* substituted both *T. maritima* XseA residues (D234 and D240) simultaneously; therefore, it is difficult to ascertain the role of individual residues. It is currently unknown if residues in *T. maritima* that correspond to *E. coli* D155, R205, H238 are essential for its nuclease activity.

The role of Q177, D246, D250 and T255 in catalysis remains unclear. Sequence analyses revealed that these residues are highly conserved in the XseA family and substitution of these residues to alanine reduced the activity to the level of about 60–80% of the wt enzyme (data not shown). In the structural model of the XseA catalytic domain, these residues are located around the putative

active site; therefore, their substitution could destabilize the active site or the interactions with the DNA, without direct interference with catalysis.

Despite the fact that magnesium ions are crucial for the activity of ExoVII from *T. maritima*, the activity of ExoVII from *E. coli* does not depend on the presence of metal ions. Larrea *et al.* postulated that ExoVII family can be divided into two groups: *E. coli*-like (resistant to EDTA) and *T. maritima*-like (sensitive to EDTA) (19). Our phylogenetic analysis indicates that *T. maritima*-like proteins are outliers of the family and suggests that the majority of members are *E. coli*-like. The distribution of magnesium dependence in the XseA family remains to be analyzed experimentally on members of the most divergent branches, whose selection can be aided by our phylogenetic tree.

The OB-fold domain alone (XseA variant Δ 104-C-term) is capable of DNA binding, while the ExoVII variant without the OB-fold domain in XseA (Δ 1–103) lost this capability. The OB-fold variants: F63A and R64E/R68E/R69E showed almost complete loss of DNA binding. This result supports the model-based prediction that F63 is indeed important for DNA binding. The predicted role of R64, R68 and/or R69 is also supported, although at this stage, the role of individual Arg residues remains unknown. Substitution Q96A decreased the binding of DNA to about 50% in comparison to that of XseA variant Δ 104-C-term (which was used as a reference). This residue probably also takes part in binding DNA, as predicted with the help of the model, but clearly it is not essential for this process. Surprisingly, the full-length XseA alone (without XseB) does not form a complex with DNA, while inactive ExoVII variants (with substitutions in the XseA catalytic domain, and in the presence of XseB) are able to bind DNA. We speculate that the XseA protein without XseB does not fold properly and therefore is unable to bind DNA. We were able to isolate and purify XseA and XseB separately, but when the two subunits were mixed together, denatured and refolded, they failed to form a catalytically active complex (data not shown), suggesting that the complex formation between the subunits may begin already at the stage of the protein synthesis.

We attempted to identify the interaction site(s) between XseA and XseB and gain information about the structural arrangement of the complex. The bioinformatic analyses showed that XseA contains a region consisting of three α -helices predicted to be involved in coiled-coil-like interactions. This has led us to a hypothesis that these helices may be involved in interactions with XseB, which has been confirmed experimentally. We did not observe XseB binding to the XseA variant that had all three helices deleted (Δ 123), and XseB binding was decreased for other XseA variants that had one or two helices deleted. While deletions of individual helices in XseA only decreased the XseB binding, they all abolished the nuclease activity of ExoVII. This result is surprising, given the observation that XseA from *T. maritima* is active, yet natively contains only two coiled-coil helices and hence corresponds structurally to a variant of *E. coli* XseA with one helix deleted. It may indicate that XseA

forms a functionally active complex only when it interacts with a precise number of XseB subunits, a possibly different number in the case of ExoVII enzymes in different species. Our attempts to express and purify isolated helices were not successful; therefore, we could not examine how many XseB subunits are bound by one helix.

It has been suggested, based on the results of size exclusion chromatography, sedimentation in sucrose gradient (16), and native gel electrophoresis, that ExoVII from *E. coli* and *T. maritima* are pentamers, composed of one XseA subunit and four XseB subunits (10,19). The results from our size exclusion experiments suggested that *E. coli* ExoVII is actually a heptamer, which consists of one XseA subunit and six XseB subunits. This result agrees with the bioinformatic-based prediction that a single helical segment of XseA can bind one XseB dimer. Consequently, we predict that *T. maritima* XseA, which has only two coiled-coil units, should bind only four XseB subunits.

We found that the full-length XseA, which possesses the OB-fold domain responsible for DNA binding, is not able to form a complex with DNA, unless it is also complexed with XseB. Earlier, it has been shown that ExoVII activity was reduced when the XseB protein was overexpressed (10). It was demonstrated that the transcription of *xseB* gene was induced upon interaction of *Neisseria meningitidis* with host cells (41). In this case, the up-regulation of XseB resulted in an induction of a DNA repair system and an increase of frequency of phase variation. That XseB expression is regulated, which in turn may influence the ExoVII activity, is suggested by the finding that a transcription factor SlyA that contributes to the virulence of *Salmonella typhimurium*, binds upstream of the *xseB* gene (42). Thus, we hypothesize that the binding of XseB to XseA is the key element that regulates the activity of ExoVII. The exact nature of XseA–XseB interactions and the structure of the ExoVII complex remain to be elucidated.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure 1 and Supplementary Files 1–2.

ACKNOWLEDGEMENTS

We would like to thank Michal Boniecki for help with REFINER. We are also grateful to Agata Kamaszewska and Elzbieta Purta for useful comments and suggestions regarding the manuscript.

FUNDING

The 6th and 7th Framework Programmes of the EU (initially ‘DNA ENZYMES’ [MRTN-CT-2005-019566] and subsequently ‘HEALTH-PROT’ [229676]); The Max Planck Society (institutional funds to S.D.-H. and A.L.); The Polish Ministry of Science [MNiSW, N N401 585738] and the START fellowship from the Foundation for Polish Science (to K.H.K.); The European Research Council [ERC, StG RNA+P = 123D to K.S. and J.M.B.]; ‘Ideas for Poland’ fellowship from the FNP

(to J.M.B.). Funding for open access charge: EC FP7 contract number 229676 (HEALTHPROT) and ERC (RNA+P = 123D).

Conflict of interest statement. None declared.

REFERENCES

1. Brissett, N.C. and Doherty, A.J. (2009) Repairing DNA double-strand breaks by the prokaryotic non-homologous end-joining pathway. *Biochem. Soc. Trans.*, **37**, 539–545.
2. Vaisman, A., Lehmann, A.R. and Woodgate, R. (2004) DNA polymerases eta and iota. *Adv. Protein Chem.*, **69**, 205–228.
3. Robertson, A.B., Klungland, A., Rognes, T. and Leiros, I. (2009) DNA repair in mammalian cells: base excision repair: the long and short of it. *Cell Mol. Life Sci.*, **66**, 981–993.
4. Milanowska, K., Krwawicz, J., Papaj, G., Kosinski, J., Poleszak, K., Lesiak, J., Osinska, E., Rother, K. and Bujnicki, J.M. (2011) REPAIRtoire: a database of DNA repair pathways. *Nucleic Acids Res.*, **39**, D788–D792.
5. Joseph, N., Duppatla, V. and Rao, D.N. (2006) Prokaryotic DNA mismatch repair. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 1–49.
6. Schofield, M.J. and Hsieh, P. (2003) DNA mismatch repair: molecular mechanisms and biological function. *Annu. Rev. Microbiol.*, **57**, 579–608.
7. Kunkel, T.A. and Erie, D.A. (2005) DNA mismatch repair. *Annu. Rev. Biochem.*, **74**, 681–710.
8. Lehman, I.R. and Nussbaum, A.L. (1964) The deoxyribonucleases of *Escherichia coli*. V. On the specificity of exonuclease I (phosphodiesterase). *J. Biol. Chem.*, **239**, 2628–2636.
9. Lovett, S.T. and Kolodner, R.D. (1989) Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **86**, 2627–2631.
10. Vales, L.D., Rabin, B.A. and Chase, J.W. (1982) Subunit structure of *Escherichia coli* exonuclease VII. *J. Biol. Chem.*, **257**, 8799–8805.
11. Harris, R.S., Ross, K.J., Lombardo, M.J. and Rosenberg, S.M. (1998) Mismatch repair in *Escherichia coli* cells lacking single-strand exonucleases ExoI, ExoVII, and RecJ. *J. Bacteriol.*, **180**, 989–993.
12. Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P. and Lovett, S.T. (2001) Redundant exonuclease involvement in *Escherichia coli* methyl-directed mismatch repair. *J. Biol. Chem.*, **276**, 31053–31058.
13. Viswanathan, M. and Lovett, S.T. (1998) Single-strand DNA-specific exonucleases in *Escherichia coli*. Roles in repair and mutation avoidance. *Genetics*, **149**, 7–16.
14. Hersh, M.N., Morales, L.D., Ross, K.J. and Rosenberg, S.M. (2006) Single-strand-specific exonucleases prevent frameshift mutagenesis by suppressing SOS induction and the action of DinB/DNA polymerase IV in growing cells. *J. Bacteriol.*, **188**, 2336–2342.
15. Dermic, D. (2006) Functions of multiple exonucleases are essential for cell viability, DNA repair and homologous recombination in recD mutants of *Escherichia coli*. *Genetics*, **172**, 2057–2069.
16. Chase, J.W. and Richardson, C.C. (1974) Exonuclease VII of *Escherichia coli*. Purification and properties. *J. Biol. Chem.*, **249**, 4545–4552.
17. Chase, J.W. and Richardson, C.C. (1974) Exonuclease VII of *Escherichia coli*. Mechanism of action. *J. Biol. Chem.*, **249**, 4553–4561.
18. Chase, J.W., Rabin, B.A., Murphy, J.B., Stone, K.L. and Williams, K.R. (1986) *Escherichia coli* exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (xseA). *J. Biol. Chem.*, **261**, 14929–14935.
19. Larrea, A.A., Pedroso, I.M., Malhotra, A. and Myers, R.S. (2008) Identification of two conserved aspartic acid residues required for DNA digestion by a novel thermophilic Exonuclease VII in *Thermotoga maritima*. *Nucleic Acids Res.*, **36**, 5992–6003.
20. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
21. Frickey, T. and Lupas, A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, **20**, 3702–3704.
22. Pei, J. and Grishin, N.V. (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, **23**, 802–808.
23. Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.*, **24**, 1596–1599.
24. Kurowski, M.A. and Bujnicki, J.M. (2003) GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.*, **31**, 3305–3307.
25. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
26. Boniecki, M., Rotkiewicz, P., Skolnick, J. and Kolinski, A. (2003) Protein fragment reconstruction using various modeling techniques. *J. Comput. Aided Mol. Des.*, **17**, 725–738.
27. Pawlowski, M., Gajda, M.J., Matlak, R. and Bujnicki, J.M. (2008) MetaMQAP: a meta-server for the quality assessment of protein models. *BMC Bioinformatics*, **9**, 403.
28. Wallner, B. and Elofsson, A. (2003) Can correct protein models be identified? *Protein Sci.*, **12**, 1073–1086.
29. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.*, **33**, W299–W302.
30. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
31. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
32. Saka, K., Tadenuma, M., Nakade, S., Tanaka, N., Sugawara, H., Nishikawa, K., Ichiyoshi, N., Kitagawa, M., Mori, H., Ogasawara, N. et al. (2005) A complete set of *Escherichia coli* open reading frames in mobile plasmids facilitating genetic studies. *DNA Res.*, **12**, 63–68.
33. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
34. Conlan, L.H. and Dupureur, C.M. (2002) Multiple metal ions drive DNA association by PvuII endonuclease. *Biochemistry*, **41**, 14848–14855.
35. Wierenga, R.K., De Mayer, M.C.H. and Hol, W.G.J. (1985) Interactions of pyrophosphate moieties with alpha helices in dinucleotide binding proteins. *Biochemistry*, **24**, 1346–1357.
36. Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
37. Alva, V., Dunin-Horkawicz, S., Habeck, M., Coles, M. and Lupas, A.N. (2009) The GD box: a widespread noncontiguous supersecondary structural element. *Protein Sci.*, **18**, 1961–1966.
38. Mayer, W.E., Schuster, L.N., Bartelmes, G., Dieterich, C. and Sommer, R.J. (2011) Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC Evol. Biol.*, **11**, 13.
39. Dunin-Horkawicz, S. and Lupas, A.N. (2010) Measuring the conformational space of square four-helical bundles with the program samCC. *J. Struct. Biol.*, **170**, 226–235.
40. Lupas, A.N. and Gruber, M. (2005) The structure of alpha-helical coiled coils. *Adv. Protein Chem.*, **70**, 37–78.
41. Morelle, S., Carbonnelle, E., Matic, I. and Nassif, X. (2005) Contact with host cells induces a DNA repair system in pathogenic *Neisseria*. *Mol. Microbiol.*, **55**, 853–861.
42. Stapleton, M.R., Norte, V.A., Read, R.C. and Green, J. (2002) Interaction of the *Salmonella typhimurium* transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. *J. Biol. Chem.*, **277**, 17630–17637.