Gene 418 (2008) 41–48

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Is crab duplex-specific nuclease a member of the *Serratia* family of non-specific nucleases?

Veronika E. Anisimova [a,b,1], Alex S. Shcheglov [a,*,1], Ekaterina A. Bogdanova [a], Denis V. Rebrikov [b,2], Alexey N. Nekrasov [a], Ekaterina V. Barsova [a], Dmitry A. Shagin [a,b], Sergey A. Lukyanov [a]

[a] *Shemiakin and Ovchinnikov Institute of Bioorganic Chemistry RAS, Miklukho-Maklaya 16/10, 117871 Moscow, Russia*
[b] *Evrogen JSC, Miklukho-Maklaya 16/10, 117871 Moscow, Russia*

ARTICLE INFO

Article history:
Received 3 January 2008
Received in revised form 3 March 2008
Accepted 6 April 2008
Available online 16 April 2008

Received by S.M. Mirkin

Keywords:
Kamchatka crab
Arthropod
Sugar non-specific nucleases
Mutagenesis
Phylogenetic analysis

ABSTRACT

Kamchatka crab duplex-specific nuclease (Par_DSN) has been classified as a member of the family of DNA/RNA non-specific beta-beta-alpha metal finger (bba-Me-finger) nucleases, the archetype of which is the nuclease from *Serratia marcescens*. Although the enzyme under investigation seems to belong to the family of *S. marcescens* nucleases, Par_DSN exhibits a marked preference for double-stranded DNA as a substrate and this property is unusual for other members of this family. We have searched other Arthropod species and identified a number of novel Par_DSN homologs. A phylogenetic analysis demonstrates that the Par_DSN-like enzymes constitute a separate branch in the evolutionary tree of bba-Me-finger nucleases. Combining sequence analysis and site-directed mutagenesis, we found that Par_DSN and its homologs possess the nuclease domain that is slightly longer than that of classic *Serratia* relatives. The active site composition of Par_DSN is similar but not identical to that of classic *Serratia* nucleases. Based on these findings, we proposed a new classification of Par_DSN-like nucleases.

1. Introduction

Superfamily of beta-beta-alpha metal finger (bba-Me-finger) nucleases combines structurally different groups including DNase colicins, H–N–H- and His-Cys box-endonucleases, DNA/RNA non-specific nucleases, structure-specific phage T4 endonuclease VII, etc. (Kuhlmann et al., 1999). These enzymes are different in overall fold and share no sequence similarity, but contain the so-called bba-Me-finger, a stretch of 22 amino acids composed of two beta-strands and an alpha-helix. In addition all these enzymes contain divalent cation that is located in the centre of bba-Me-finger in position conservative for all enzymes belonging to this group (Kuhlmann et al., 1999).

DNA/RNA non-specific bba-Me-finger nucleases, the archetype of which is the nuclease from *Serratia marcescens* (*Serratia* family of nucleases, SFN), typically hydrolyze both double-stranded (ds) and single-stranded (ss) DNA, as well as RNA at equal or similar rates (Rangarajan and Shankar, 2001). Several nucleases belonging to this family preferentially hydrolyze DNA, for example, *Syncephalastrum racemosum* nuclease (Ho and Liao, 1999). SFN members share a common structure, specifically, a DNA/RNA non-specific endonuclease (NUC) domain containing a conserved Asp-Arg-Gly-His (DRGH) motif with an active site histidine (Friedhoff et al., 1994, 1996a,b; Meiss et al., 2000).

Similar mechanisms of action and active site structures have been postulated for all SFN members (Rangarajan and Shankar, 2001). However, detailed analyses of structure–function correlations have been performed on a restricted number of SFN members, including nucleases from *S. marcescens* (SmNase), *Anabaena* sp. (Anasp_NucA), and bovine endonuclease G (endoG) (Friedhoff et al., 1994; Miller et al., 1994; Miller and Krause, 1996; Lunin et al., 1997; Meiss et al., 2000; Schafer et al., 2004). Inclusion of other nucleases are in the SFN family is based on sequence similarity with SmNase.

Among putative SFN members are group of Arthropod nucleases including recently described duplex-specific nuclease (Par_DSN) from Kamchatka crab hepatopancreas (Shagin et al., 2002) and CuquEndo from southern house mosquito (Calvo and Ribeiro, 2006). These proteins comprise NUC domain with motive characteristic for SFN members, however, they have substrate specificity unusual for DNA/RNA non-specific nucleases, i.e. they display a strong preference for double-stranded (ds) DNA as compared with single-stranded (ss) DNA and RNA.

Abbreviations: Par_DSN, duplex-specific nuclease from Kamchatka crab *Paralithodes camtschaticus*; bba-Me-finger, beta-beta-alpha metal finger; SFN, *Serratia* family of nucleases; ds, double stranded; ss, single stranded; SmNase, nuclease from *Serratia marcescens*; NUC domain, DNA/RNA non-specific endonuclease domain; Anasp_NucA, nuclease from *Anabaena* sp.; IU, information units; DS_NUCs, duplex-specific nucleases; IDIC-sites, sites of increased degree of information coordination between residues; RACE, rapid amplification cDNA ends.

* Corresponding author. Tel./fax: +7 495 330 7056.
E-mail address: jukart@mail.ru (A.S. Shcheglov).

[1] These authors contributed equally to the work.

[2] Present address: DNA-technology JSC, Kashirskoe highway 24, building 2, office 222, 115478, Moscow, Russia.

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.04.005

Public resources indicate that at least two more Arthropod nucleases demonstrate a marked preference for ds DNA, specifically, mitochondrial nuclease from *D. melanogaster* embryos (Harosh et al., 1992) and commercially available nuclease from arctic shrimp *Pandalus borealis* (USB, http://www.usbweb.com/category.asp?%20cat=mbe&id=78314). Despite the lack of sequence data for *Drosophila* and arctic shrimp nucleases, similarity of their enzymatic properties allow to suppose that both these nucleases are homologous to Par_DSN.

Unusual substrate specificity of the nucleases makes it useful in development of advanced technologies in many areas of biotechnology and biomedicine. In particular, Par_DSN found use in modern technologies of cDNA normalization (Zhulidov et al., 2004, 2005), full-length subtractive hybridization (Peng et al., 2008), single nucleotide polymorphism detection (Shagin et al., 2002; Al'tshuler et al., 2005), and quantitative determination of telomeric overhang (Zhao et al., 2008). Studies of the mechanism of action of these nucleases are important for refining and improving their specific cleavage properties.

In this paper, we analyzed the relationships of Par_DSN and its Arthropod homologs with other members of *Serratia* family of nucleases and analyzed the structure of Par_DSN active site. We proposed a new classification of Par_DSN-like nucleases and shown that they represent an interest example of divergent evolution of enzymatic function.

2. Materials and methods

2.1. Par_DSN activity analysis

DNAse activity was measured using the modified Kunitz assay (Kunitz, 1950; Liao et al., 1974) in 1×reaction buffer (7 mM $MgCl_2$, 50 mM Tris–HCl, pH 8.0) containing 400 μM calf thymus DNA. Substrate preference was examined as described earlier (Shagin et al., 2002).

2.2. Cloning of Par_DSN homologs

Total RNA was isolated from gammarus, glass shrimp and mangrove fiddler crab, according to the protocol of Matz (Matz et al., 2003). First-strand cDNA synthesis and RACE procedures with degenerative primers were performed using the SMART RACE cDNA Amplification Kit (Clontech), according to manufacturer's instructions. Primers were generated on the basis of multiple sequence alignments of Par_DSN, kuruma shrimp nuclease and SFN members from *S. marcescens* (1583130), *Glossina morsitans* (AAF82097), *Homo sapiens* (XP_002889), *S. racemosum* (P81204), *C. echinulata* (P81203), and *D. melanogaster* (AAF49206). For cloning of the gammarus nuclease, the AFN2 primer (5′-cctcagtggca(g/a)gcttt(c/t)aac-3′) was used for PCR with the Step-Out system. In total, 34 cycles of amplification were performed under the following conditions: 95 °C for 10 s, 62 °C for 10 s, and 72 °C for 2 min. For isolation of glass shrimp and mangrove fiddler crab putative nucleases, the AFN2 primer was used for initial PCR that included 32 cycles with the following program: 95 °C for 10 s, 60 °C for 10 s, and 72 °C for 2 min, and the GNW1 primer (5′-caggcctttaataatggtaattgg-3′) was employed for the nested PCR involving 25 cycles (95 °C for 10 s, 64 °C for 10 s, and 72 °C for 2 min). The cDNAs containing the full coding sequences of nucleases were isolated with the aid of a SMART RACE cDNA Amplification Kit (Clontech) and gene-specific primers for fiddler crab nuclease (5′-ggattgccattaatgtcgtc-3′; 5′-ccactgtacacccgaaggtc-3′; 5′-aaccaaggctcgccaagtcc-3′), gammarus nuclease (5′-caatggtccgaattctgttctc-3′; 5′-gtgactacgcgcagagtggc-3′), and glass shrimp nuclease (5′-ccagcactccccaacctcc-3′; 5′-gtcaggtcagtgccgtgggc-3′).

2.3. Sequence analysis

Sequences were analyzed with BLAST software (http://www.ncbi.nlm.nih.gov/BLAST). Domain organization was verified using the web-based SMART tool (http://smart.embl-heidelberg.de) and SignalP 3.0 software (http://www.cbs.dtu.dk/services/SignalP/). The putative N- and/or C-terminal terminal ends were searched in the genomic regions using FGENESH software (http://sun1.softberry.com). Multiple sequence alignments were prepared with the aid of MAFFT software (http://timpani.genome.ad.jp/~mafft/server/), with following optimization by visual inspection. Phylogenetic analysis was performed using TREE-PUZZLE, version 5.0 (Strimmer and von Haeseler, 1997). Maximum likelihood (ML) trees were inferred using quartet puzzling with 10,000 puzzling steps. We employed the discrete G-distribution model (with eight categories) for site heterogeneity (Yang, 1996), and the Whelan and Goldman model of substitution (Whelan and Goldman, 2001). The neighbor-joining tree was applied to estimate the parameters.

The Par_DSN information structure was analyzed as described in (Nekrasov, 2004). Briefly, the Par_DSN amino acid sequence was presented as "information units" (IU) composition, whereby IU is a short (up to 6 residues) overlapping fragments. A profile of information content was generated for the Par_DSN sequence on the basis of previously obtained statistic parameters (Nekrasov, 2002, 2004). Sites of increased degree of information coordination between residues (IDIC-sites) were detected by decomposing the information content profile into Gaussian curves, and locating maxima of the decomposition (Gaussian) functions with the greatest values for the protein sequence in question. IDIC-sites of different lengths were identified by varying the half-width of the Gaussian curves. Blocks of IDIC-sites were used for plotting a IDIC-diagram where centers of the bordering IDIC-sites are joined. Thus, hierarchical elements (IDIC-branches) that form IDIC-trees and IDIC-associations were selected.

2.4. Site-directed mutagenesis

Site-directed mutagenesis was performed using a pET plasmid containing the Par_DSN coding sequence (pET 22b-DSN) by PCR with the T7 promoter or T7 terminator primers (Novagen) and specific primers comprising nucleotide substitutions (Supplemental Table 1). Fragments were purified by agarose gel electrophoresis, and extracted using the QIAquick gel extraction kit (Qiagen). The fragments were annealed and elongated using *Taq* polymerase under the following conditions: 5 cycles of 94 °C for 20 s, 58 °C for 2 min, and 72 °C for 1 min. Full-length coding sequences were amplified using the T7 promoter and terminator primers, and cloned into the NdeI and HindIII restriction sites of pET 22 b+. The cloned sequences were verified by sequencing. Than Par_DSN mutants were expressed and purified as described earlier (Anisimova et al., 2006).

2.5. Cloning of truncated Par_DSN variants

To clone the truncated Par_DSN variants, coding sequences were amplified from pET 22b-DSN plasmid by PCR with the following gene-specific primers: P1 (5′-acgtcggatcctgggttcttcagcgtctccatg-3′) and P2 (5′-acgtcaagcttcccgtgagccaggtcgtcc-3′) for the Par_DSN-t1 variant; P3 (5′-acgtcggatccttatgaactgatcagcgtgtgttt-3′) and the T7 terminator for the Par_DSN-t2 variant; P4 (5′-acgtcggatcctacaaccttgcgcaccgagca-3′) and the T7 terminator for the Par_DSN-t3 variant; and P5 (5′-acgtcggatcctaaccacgtgaaggagacc-3′) and T7 terminator for the Par_DSN-t4 variant. Fragments were purified and cloned into pET 22 b+, as described above. The absence of additional mutations was confirmed by sequencing. Than truncated Par_DSN variants were expressed and purified as described earlier (Anisimova et al., 2006).

3. Results

3.1. Collection and structure of Arthropod nucleases

Determination of the structural features responsible for the unusual properties of some Arthropod nucleases requires accurate

Table 1
Sequence collection of non-specific nucleases

	Protein name	Protein ID	Species	Method of finding
Crustacea				
1	Kamchatka crab duplex-specific nuclease (Par_DSN)	AAN86143	*Paralithodes camtschaticus*	HC
2	Gammarus putative nuclease (Gam_NUC)	DQ862539	*Gammarus* sp.	HC
3	Kuruma shrimp nuclease (Pen_NUC)	CAB55635	*Penaeus japonicus*	BLAST search
4	Glass shrimp putative nuclease (Pal_NUC)	DQ862538	*Palaemonidae* sp.	HC
5	Fiddler crab putative nuclease (Uca_NUC)	DQ862540	*Uca crassipes*	HC
6	Green crab putative nuclease (Car_NUC)	DN738756; DV944764	*Carcinus maenas*	BLAST search
7	Lobster putative nuclease (Hom_NUC)	CN853059	*Homarus americanus*	BLAST search
Insect				
8	ENSANGP00000017934 (An_17934)	XP_320813	*Anopheles gambiae str. PEST*	BLAST search
9	ENSANGP00000010690 (ANO_10690)	XP_318056	*A. gambiae*	BLAST search
10	ENSANGP00000028890 (An_28890)	XP_560379	*A. gambiae*	BLAST search
11	ENSANGP00000019783 (An_19783)	XP_308089	*A. gambiae*	BLAST search
12	ENSANGP00000019760 (An_19760)	XP_308088	*A. gambiae*	BLAST search
13	ENSANGP00000002180 (An_02180)	XP_307627	*A. gambiae*	BLAST search
14	ENSANGP00000018575 (An_18575)	XP_317654	*A. gambiae*	BLAST search
15	CG8862 (Dr_CG8862)	NP_610737	*D. melanogaster*	BLAST search
16	CG14120 (Dr_CG14120)	NP_648610	*D. melanogaster*	BLAST search
17	CG6839 (Dr_CG6839)	NP_649076	*D. melanogaster*	BLAST search
18	CG3819 (Dr_CG3819)	NP_649078	*D. melanogaster*	BLAST search
19	CG14118 (Dr_CG14118)	NP_648612	*D. melanogaster*	BLAST search
20	CG9989 (Dr_CG9989)	AAF56806	*D. melanogaster*	BLAST search
21	CG33346 (Dr_CG33346)	AAS65221	*D. melanogaster*	BLAST search
22	IP10440 (Dr_IP10440)	AAY55811	*D. melanogaster*	BLAST search
23	LP23408 (Dr_LP23408)	AAS77426	*D. melanogaster*	BLAST search
24	GA19896 (Dr_GA19896)	EAL30413	*D. pseudoobscura*	BLAST search
25	GA12772 (Dr_GA12772)	EAL31004	*D. pseudoobscura*	BLAST search
26	GA17708 (Dr_GA17708)	EAL30415	*D. pseudoobscura*	BLAST search
27	GA12733 (Dr_GA12733)	EAL27007	*D. pseudoobscura*	BLAST search
28	GA11906 (Dr_GA11906)	EAL25371	*D. pseudoobscura*	BLAST search
29	Tsal1 (Gl_Tsal1)	AAF82097	*Glossina morsitans*	BLAST search
30	Tsal2 (Gl_Tsal2)	AAF82098	*G. morsitans*	BLAST search
31	CuquEndo (Culex NUC)	AAR18449	*Culex pipiens quinquefasciatus*	BLAST search
32	43.7 kDa salivary protein (Lut_NUC))	AAS16916	*Lutzomyia longipalpis*	BLAST search
Other				
33	Anasp_NUCA	P38446	*Anabaena variabilis*	BLAST search
34	SmNase	P13717	*Serratia marcescens*	BLAST search
Other				
35	mitochondrial nuclease CaO19.8582	XP_716865	*Candida albicans SC5314*	BLAST search
36	Major mitochondrial nuclease Nuc1 (Sac_NUC)	NP_012327	*Saccharomyces cerevisiae*	BLAST search
37	Sr-nuclease (Sr_NUC)	P81204	*Syncephalastrum racemosum*	BLAST search
38	Major nuclease C1A isoform precursor (Cun_NUC)	AAC78769	*Cunninghamella echinulata*	BLAST search
39	Endonuclease G-like 1, mitochondrial precursor (Hum_EndoGL)	NP_005098	*Homo sapiens*	BLAST search
40	Ced-3 protease suppressor protein 6 (Caen_cps-6)	AAL08030	*Caenorhabditis elegans*	BLAST search

HC — homolog cloning.

comparative analysis with SFN. The accuracy of multiple alignments depends on numbers of homologs available for analysis (Katoh et al., 2005). Consequently, we initially focused on identification of collection of Arthropod non-specific nucleases, particularly, close Par_DSN homologs. To date, only the kuruma shrimp nuclease sequence having 64% amino acid identity with Par_DSN has been reported (Wang et al., 2000).

Using degenerative primers corresponding to the most conserved amino acid sequences in the DNA/RNA non-specific nuclease family and Step-out RACE technology (Matz et al., 1999), we isolated the complete coding sequences of novel nucleases from three Crustacean species, including gammarus (*Gammarus* sp.), glass shrimp (*Palaemonidae* sp.), and fiddler crab (*Uca crassipes*). Additionally, partial nucleotide coding sequences of putative non-specific nucleases from green crab (*Carcinus maenas*) and lobster (*Homarus americanus*) were identified from BLAST searches of the nr database (Table 1). All predicted amino acid sequences of the novel nucleases displayed 50–65% identity with Par_DSN, except gammarus nuclease (Gam_NUC) with 32% identity.

Analysis using SMART and SignalP 3.0 softwares indicates that all Crustacean nucleases, including novel and recently described (Wang et al., 2000; Shagin et al., 2002), contain a signal peptide, that is probably important for nuclease secretion and domains homologous to NUC of SmNase. No domains other than NUC were identified in the mature proteins. In all cases, the NUC domain (determined with SMART software) was connected to the signal peptide by a 120–180 aa linker with unknown function.

To perform accurate alignments of Crustacean and other SFN nucleases, we expanded the group for comparison with the sequences of putative insect non-specific nucleases identified with a BLAST homologue search (Table 1). In the cases where the amino acid sequence looked incomplete, an additional search of the putative N- and/or C-terminal ends was performed in the respective genomic regions using FGENESH software (Salamov and Solovyev, 2000).

Before alignment, the protein structures of insect nucleases were analyzed. Signal peptides were predicted with high probability for majority of insect nucleases. Again no functional domains other than NUC were identified in all mature proteins.

Multiple sequence alignments were generated using MAFFT software (Katoh et al., 2005) with subsequent optimization by visual inspection (Supplemental Fig. 1). Conserved regions of the NUC domains from the alignment data were used for maximum likelihood (ML) tree construction with TREE-PUZZLE software. The NUC domains examined were divided into two main subgroups, the first containing SFN members and two insect enzymes, and the second (that we called duplex-specific nucleases, DS_NUCs) consisting of other insect and Crustacean nucleases. The second group was subdivided into several

subclades whereby Crustacean nucleases (except Gam_NUC) are grouped together (Fig. 1).

On the basis of multiple alignments, we found that the NUC domain of DS_NUCs is longer by at least 50 residues than the NUC domain of SFN predicted with SMART software and previous alignments (Rangarajan and Shankar, 2001). The sequence homology of DS_NUCs was detected far from the putative N-terminal end of the NUC domain that was definite based on SmNase sequence (Fig. 2). For instance, several conserved Cys residues typical of DS_NUCs were detected in this region. These data indicate that these nucleases either contain an additional domain of unknown function or an N-termimal extension of the NUC domain. The main difference between the NUC

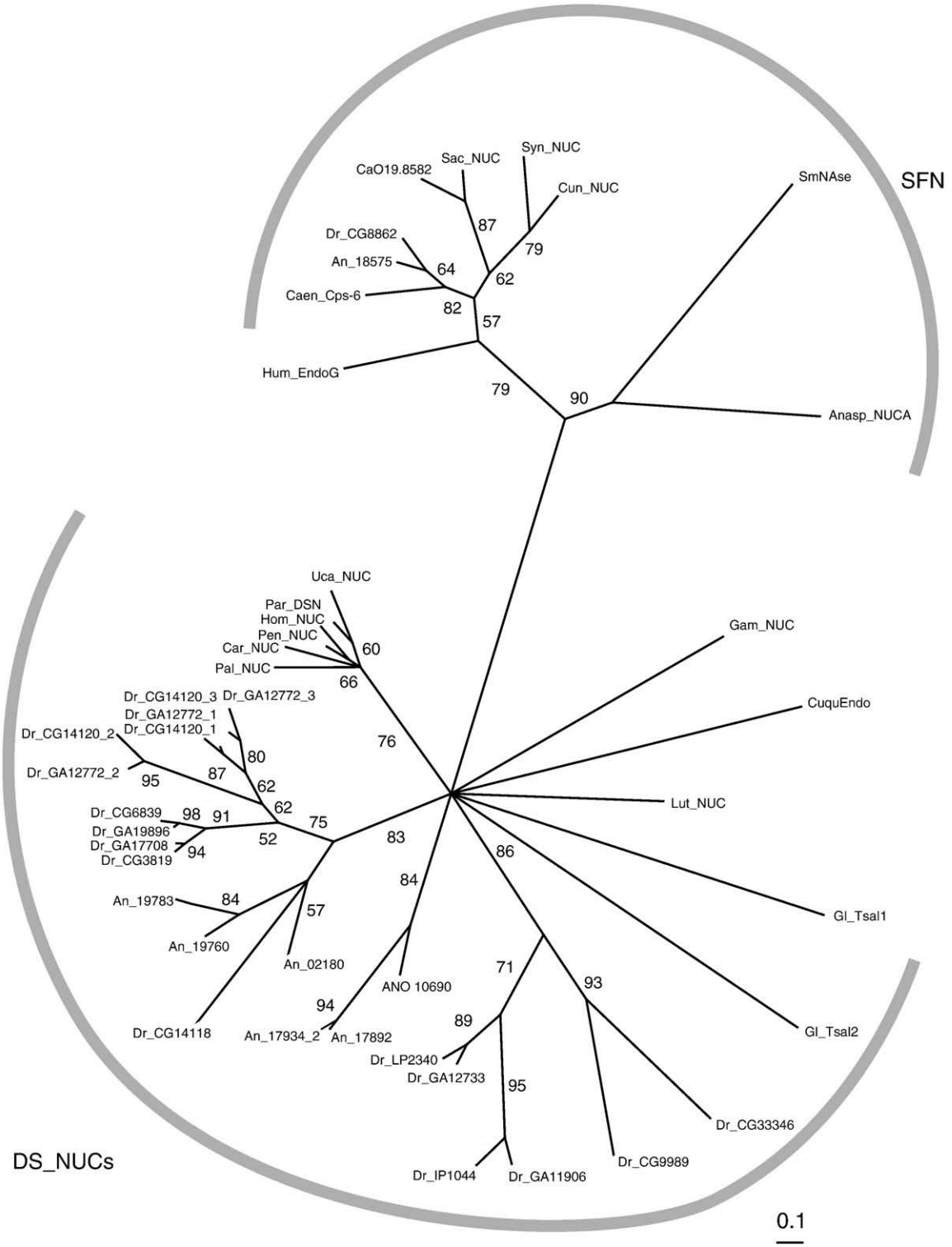

Fig. 1. Unrooted maximum likelihood (ML) tree constructed with TREE-PUZZLE software, based on multiple alignments of mature non-specific nucleases, as shown in Supplemental Fig. 1. The ML tree was inferred using quartet puzzling with 10,000 puzzling steps. The numbers at the nodes indicate support values for each branch.

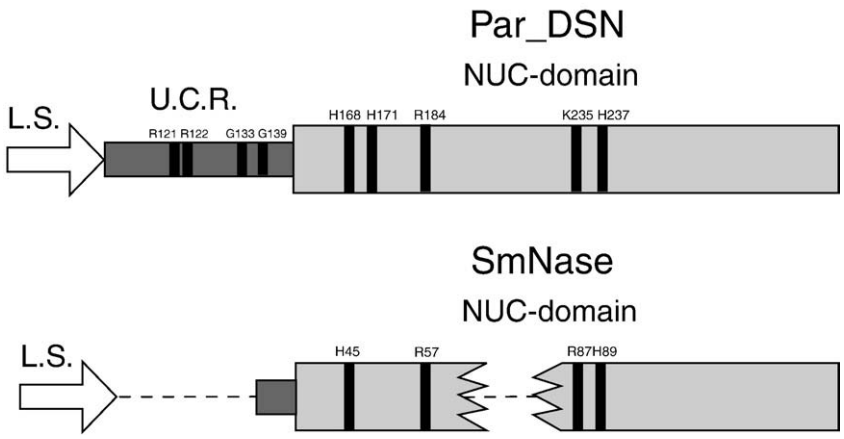

Fig. 2. Sketchy alignment and domain structure of Par_DSN and SmNase. Par_DSN residues, changed by mutagenesis, marked as bold lines. Corresponding residues at SmNase also are bold. L.S. — leader signal, U.C.R. — unknown region, conserved within DS_NUCs.

domains of DS_NUCs and typical SFN members involves an insertion of 16 (or more) residues immediately upstream of the RGH motif that is characteristic of both subgroups (Fig. 2).

We also analyzed amino acid residues of Par_DSN corresponding to those identified in the active sites of SFN nucleases. The existence of a common catalytic site for DNA and RNA substrates has been suggested for SFN members, including His89, Arg57, Arg87, Asn119, Glu127, Trp123, and Arg131 amino acid residues in SmNase (His124, Arg93, Arg122, Asn155, Glu163, Trp159, and Arg167 in Anasp_NucA, respectively) (Friedhoff et al., 1996a,b; Meiss et al., 2000). His89 has been postulated as a general base in catalytic activity, while Asn119 as a primary ligand for metal binding (Miller et al., 1999; Meiss et al., 2000). In addition, Miller and colleagues showed that Asp86 (Asp 121 in Anasp_NucA) interacts indirectly with the metal binding site via hydrogen bonds with the amide nitrogens of Asn119 and Gln120 (Miller et al., 1999). The main difference in the action of the nucleases analyzed has been found in transition state stabilization. For instance, Arg57 has been supposed to be required for neutralization of the extra negative charge of the transition state in SmNase (Friedhoff et al., 1999), while non-conserved Asp95 plays this role in Anasp_NucA (Meiss et al., 2000). Moreover, In the case of endoG Cys115 located in the position corresponding to Arg57 seems to be tot well suited for transition state stabilization (Schafer et al., 2004).

The majority of DS_NUCs examined contain conserved Glu, Trp, Asn, His and Arg residues corresponding to Glu127, Trp123, Asn119, His89 and Arg57 of SmNase. Only Tsal proteins from the salivary glands of *G. morsitans* (Haddow et al., 2002) and second NUC domains of CG14120-PA and GA12772-PA proteins contain one or more substitutions at these positions. Similarly, Arg residues corresponding to Arg167 of Anasp_NucA are conserved throughout the DS_NUCs, except of CG9989, CuquEndo and Tsal2 proteins. The residue corresponding to Anasp_NucA Arg122 has not been identified in Tsal2 proteins, and is substituted with Lys in four of the six Crustacean nucleases. Asp86 of SmNase is not conserved in the DS_NUCs, and is replaced with Ala in all Crustacean nucleases.

3.2. Structure of Par_DSN NUC domain

To establish the region encompassing the complete functional catalytic domain of Par_DSN, we analyzed the protein information structure, as described by Nekrasov (2002, 2004). This analysis allows the detection of the spatial structural units of the protein in which residues relate more to each other than in a casual amino acid sequence (so-called IDIC-associations). IDIC-associations represent entire structural elements that form functional domains (Nekrasov, 2004), and are divided by regions with decreased degree of information. Such regions represent linkers between spatial structural elements of the protein, and some may be used as sites for excision of functional domains.

Analysis of the Par_DSN information structure (Fig. 3) revealed the presence of six IDIC-associations, the first corresponding to the signal peptide region. The second association is located between residues 35 and 114, the third between 114 and 238, the fourth between 238 and 260, the fifth between 265 and 325, and the sixth commencing from residue 336 (here and below numeration of Par_DSN residues is given for unprocessed protein). Position 114 was identified as the putative start of a catalytic domain.

A number of Par_DSN variants with truncated N-termini were prepared for exact identification of the nuclease domain. The first (Par_DSN-t1) included residues from positions 192 to 379 of the Par_DSN precursor (GenBank ID AAN86143) linked with a bacterial signal peptide and 6×His tag. This sequence corresponds to the NUC domain identified with SMART software. We purified and renatured this protein, however unfortunately this protein was insoluble and inactive.

The second variant, Par_DSN-t2, comprising residues from positions 149 to 406, corresponded to the NUC domain identified by multiple sequence alignment. After renaturation and purification, Par_DSN-t2 was obtained as a soluble protein lacking DNase activity. In contrast to the full-length protein (either natural or recombinant), it was sensitive to proteinase K and was completely cleaved by this protease. The Par_DSN-t3 variant contained residues 162–406. On this sequence we deleted unpaired Cys residues. The Par_DSN-t3 was soluble, was sensitive to proteinase K and was lacking DNAse activity. Finally, a Par_DSN-t4 variant including residues 114–406 (predicted from Par_DSN structure information) was purified as an active soluble enzyme. However the mutant protein displayed dramatically different properties from the wild-type Par_DSN. For example, Par_DSN-t4 has considerably lower specific activity (6 Kunitz-units vs 6000 Kunitz-units for intact Par_DSN), temperature optimum (37 °C vs 60 °C for intact Par_DSN) and temperature stability. Activity of Par_DSN-t4 completely disappeared after incubation at 70 °C for 20 min, while full-length Par_DSN retained about 25% activity under these conditions. Moreover, the truncated enzyme was degraded by proteinase K. Finally, the ds DNA:ss DNA preference ratio for Par_DSN-t4 was equal to 10, whereas the corresponding ratio for intact Par_DSN was equal to 1000.

3.3. Site-directed mutagenesis data

In addition to truncated variants, several point mutants of Par_DSN were obtained by site-directed mutagenesis. All mutant

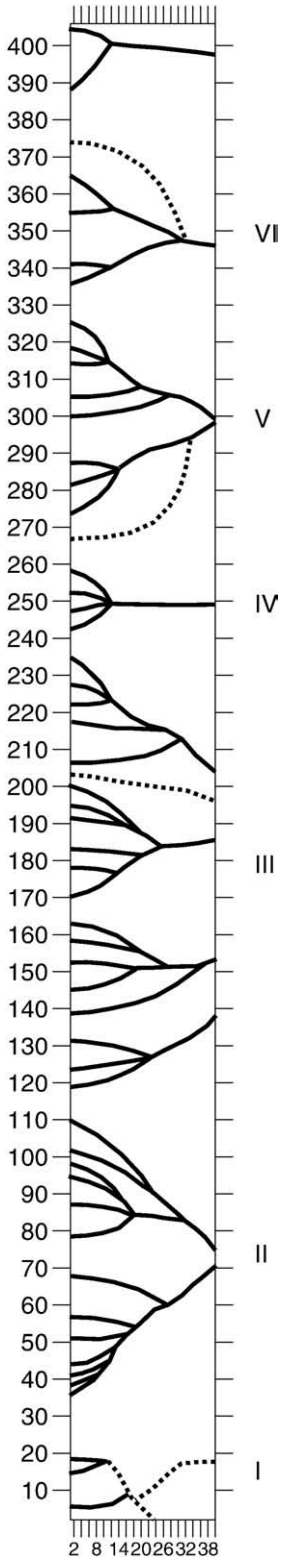

Fig. 3. The information structure (IDIC-diagram) of Par_DSN. The axes represent the serial numbers of amino acid residues in the protein sequence and half-width values of decomposition function.

proteins were expressed in *E. coli*, renatured from inclusion bodies, and purified to homogeneity. The enzymatic activity of mutant proteins was measured using the modified Kunitz assay with calf thymus DNA (Table 2). The enzymatic activity was determined with both ds and ss DNA and in the last case we used synthetic oligonucleotides. We additionally examined the stability of mutant proteins to proteinase K. Stability to proteolysis reflects proper formation of disulfide bonds.

As expected, substitution of His237 (corresponding to His89 of SmNase) to Ala resulted in almost complete loss of enzymatic activity, suggesting a crucial role of this residue in catalysis. It is worthwhile to mention that this mutation does not affect stability of enzyme to proteolysis by proteinase K.

We also performed mutagenesis of Lys235 corresponding to Arg87 of SmNase that is presumably involved in the substrate binding. The Lys235Ala substitution led to a sharp decrease of nuclease activity (Table 2). Analogous effects were observed for the SmNase Arg87Ala substitution (Friedhoff et al., 1996a). In contrast, the Lys235Arg mutant displayed comparable catalytic activity to wild-type Par_DSN. These findings suggest similar roles of Lys235 in Par_DSN and Arg87 in SmNase. Moreover, we concluded that Arg-Lys substitution is not involved in the determination of Par_DSN substrate specificity.

According to our multiple alignment analysis, Arg184 of Par_DSN corresponds to SmNase Arg57. The Arg57Lys SmNase mutant retained only 6% of activity, and the Arg57Ala mutant displayed less than 1% activity (Friedhoff et al., 1996a). In contrast, the Arg184Lys mutant of Par_DSN had similar enzymatic properties as the wild-type protein, and activity of the Arg184Ala mutant was only 3 times less than the corresponding activity of the wild-type protein. Thus, unlike Arg57 of SmNase, Arg184 of Par_DSN appears unimportant for catalysis. All Arg184 and Lys235 mutants were stable to proteinase K. Next, we focused on the His residues in the region corresponding to the environment of His45 of SmNase that is conserved in typical SFN members. A conserved His residue is present in DS_NUCs (His168 in Par_DSN). Another His residue (His171 in Par_DSN) is characteristic of Crustacean nucleases only. Mutagenesis of His168 and His171 in Par_DSN resulted in inactive mutants with loss of stability to proteinase K. These data indicate that these residues may be important for correct protein folding.

Similarly, Gly139Val and Gly133Val mutants were inactive and were rapidly degraded by proteinase K. According to our alignments, Gly 139 of Par_DSN is conserved within all Arthropod nucleases examined, except Tsal1.

Arg121Ala and Arg122Ala mutants displayed comparable activity and stability to proteinase K as the wild-type Par_DSN. However, a double mutant (Arg121Ala and Arg122Ala) was inactive, but stable to proteinase K. These Arg residues are located upstream of the NUC domain, as determined by multiple alignments, but the majority of DS_NUCs contain Arg or Lys at one or at both corresponding positions.

Table 2
Par_DSN mutagenesis and mutant activity studies

Position	Corresponding position in SmNase	Amino acid exchange	Activity against ds DNA, Kunitz unit/mg.	Yield after proteinase K treatment (%)
Wild type	–	–	6070	20%
His 237	His 89	Ala	n.d.	19%
Lys 235	Arg 87	Ala	110	21%
		Arg	5980	20%
Arg184	Arg 57	Lys	6100	18%
		Ala	1810	22%
Arg122		Lys	6050	20%
		Ala	6030	
Arg121		Lys	5990	19%
		Ala	5960	
Arg 121, 122		Ala	n.d.	4%
His168	His45	Ala	n.d.	0
His 171		Ala	n.d.	0
Gly 133		Val	n.d.	0
Gly 139		Val	n.d.	0

4. Discussion

A number of Arthropod nucleases exhibit divalent cation-dependent DNAse activity with a marked preference for ds DNA. Par_DSN requires at least 10 base pairs of perfect duplex (with no mismatches) for effective cleavage of DNA, and possesses only slight activity with ss DNA and RNA as substrates (Shagin et al., 2002; Anisimova et al., 2006). Activity against ss DNA can be detected only when high concentrations of both Par_DSN enzyme and substrate are used (Zhao et al., 2008); and is changed proportionally to the activity against ds DNA under various conditions including ionic strength, pH, addition of synergetic agents (Ca^{2+}-ions), detergents, chaotropic agents, and polyamines (data not shown).

Like Par_DSN, commercially available nuclease from arctic shrimp *P. borealis*, mitochondrial nuclease from *D. melanogaster* embryos (Harosh et al., 1992), and recombinant nuclease CuquEndo from southern house mosquito (Calvo and Ribeiro, 2006) have high specificity to ds DNA as a substrate.

Among these nucleases, Par_DSN and CuquEndo contain a NUC domain that shares characteristic structural features of the corresponding domain of SFN. Although, CuquEndo shows less similarity to Par_DSN (25% identity and 40% similarity at the amino acid level) than SmNase (26% identity and 42% similarity), both are combined into a separate clade from typical SFN by means of phylogenetic analyses. Analysis of other Arthropod nucleases revealed that majority of these enzymes belongs to the same clade as Par_DSN, whereas the others are the members of SFN family (e.g. CG8862 and ENSANGP00000018575). Therefore we may postulate divergent evolution of DS_NUCs and SFN nucleases that support their different biological functions and substrate specificities.

Detailed analysis of Arthropod nucleases belonging to the DS_NUC group reveals that these enzymes have larger NUC domain than the enzymes belonging to the SFN family. In good agreement with this conclusion the site-directed mutagenesis revealed several residues outside the SFN NUC domain sequence essential for Par_DSN activity (Arg121 and Arg122). In addition, this conclusion agrees well with the results of analysis of Par_DSN information structure and is supported by the fact that all attempts to express the Par_DSN NUC domain predicted using either SMART software or multiple sequence alignments with SFN members resulted in obtaining very unstable proteins lacking nuclease activity. For instance the shortest mutant Par_DSN-t1 variant was completely insoluble and inactive. Longer variants (Par_DSN-t2 and Par_DSN-t3) were solubilized during refolding procedure, but both were inactive and were easily cleaved by proteinase K. Only Par_DSN-t4 mutant containing complete sequence of catalytic domain predicted by the analysis of Par_DSN information structure was correctly folded and possessed nuclease activity. Since according to the data of size-exclusion chromatography DSN is presented in the form of monomer, we can exclude suggestion that truncated mutants are lacking enzymatic activity due their inability to form any kinds of oligomers.

Multiple sequence alignments of DS_NUCs and SFN members resulted in the identification of several essential differences within NUC domains. However, the NUC domains of the most DS_NUCs contain the largest part of characteristic residues conserved in SFN NUC domains (a few exceptions are Tsal proteins and second NUC domains of CG14120-PA and GA12772-PA). Mutational analysis of Par_DSN residues at positions 235 and 237 corresponding to Arg87 and His89 of SmNase confirms their crucial involvement in formation of active site.

It is important to underline that in addition to conservative residues described for SFN NUC, the corresponding domain of Par_DSN contains additional residues important for enzymatic activity. For instance, the mutant 121Ala–122Ala mutant is completely inactive whereas the single mutants Arg121Ala and Arg122Ala mutants possess high hydrolytic activity comparable with the wild-type protein. This finding indicates that at least one of these Arg residues is somehow involved either in formation or regulation of catalytic activity of Par_DSN active site. It is worthwhile to mention that the majority of Arthropoda nucleases contain one or both Arg (or Lys) residues at the corresponding positions of their structure.

Many properties of isolated NUC domain of Par_DSN were significantly different from those of intact full-size enzyme. For example, Par_DSN-t4 mutant had significantly lower specific activity, lower stability and specificity to ds DNA than the full-size enzyme. The rate of hydrolysis of ds DNA by Par_DSN-t4 was 10 times higher than that of ss DNA, while for intact Par_DSN, the ds DNA:ss DNA ratio was approximately 1000. Therefore we may suggest that the region containing residues 28–113 is important for selective ds DNA binding. BLAST searches revealed no homologous sequences to this region apart from Arthropod nucleases. We suggest that this region is linked to the non-specific NUC domain in Arthropod ancestors, and is currently involved in the modulation of Par_DSN substrate specificity. Future X-ray experiments will probably provide detailed information on the interaction between this region and nuclease domain of Par_DSN.

In addition, residues that play a subsidiary role in catalysis are different in Par_DSN and SmNase. For instance, Asp86 of SmNase, that seems to be indirectly involved in the interaction with the cation-binding site (Miller et al., 1999), is conserved in SFN. However, the majority of Arthropod nucleases contain another residue at the corresponding position, for example Ala in all Crustacea proteins. According to multiple sequence alignments, Arthropod nucleases contain an insertion of several residues near the RGH motif. It is possible that some residues within this insert somehow indirectly participate in the interaction with the cation-binding site. It was also found that Par_DSN Arg184 corresponding to the SmNase Arg57 is not involved in stabilization of the transition state.

Thus, Par_DSN and its homologs from Arthropod have high specificity towards ds DNA, contain nuclease domains that are longer and divergent from the corresponding domain of SFN. In addition the active site of Par_DSN and its homologs is similar, but not identical to the active site of the NUC domain of non-specific *Serratia*-like nucleases. Moreover, the nuclease domains of Arthropod enzymes are linked to peptide consisting of about 90 aa that seems to be important for modulation of substrate specificity and enzyme stability. Therefore we suppose that Arthropod nucleases should be classified as a novel subfamily of bba-Me-finger nucleases — duplex-specific nucleases.

Acknowledgements

We are grateful to Prof. N.B. Gusev for comments on this manuscript. This work was supported by the Evrogen JSC (Russia) and Presidium RAS (Grant No. 34-405/06-60).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2008.04.005.

References

Al'tshuler, I.M., Zhulidov, P.A., Bogdanova, E.A., Mudrik, N.N., Shagin, D.A., 2005. Application of the duplex-specific nuclease preference method to the analysis of point mutations in human genes. Bioorg. Khim. 31, 627–636.

Anisimova, V.E., Rebrikov, D.V., Zhulidov, P.A., Staroverov, D.B., Lukyanov, S.A., Shcheglov, A.S., 2006. Renaturation, activation, and practical use of recombinant duplex-specific nuclease from Kamchatka crab. Biochemistry (Mosc.) 71, 513–519.

Calvo, E., Ribeiro, J.M., 2006. A novel secreted endonuclease from *Culex quinquefasciatus* salivary glands. J. Exp. Biol. 209, 2651–2659.

Friedhoff, P., et al., 1994. A procedure for renaturation and purification of the extracellular *Serratia marcescens* nuclease from genetically engineered *Escherichia coli*. Protein Expr. Purif. 5, 37–43.

Friedhoff, P., Kolmes, B., Gimadutdinow, O., Wende, W., Krause, K.L., Pingoud, A., 1996a. Analysis of the mechanism of the *Serratia* nuclease using site-directed mutagenesis. Nucleic Acids Res. 24, 2632–2639.

Friedhoff, P., et al., 1996b. Kinetic analysis of the cleavage of natural and synthetic substrates by the *Serratia* nuclease. Eur. J. Biochem. 241, 572–580.

Friedhoff, P., Franke, I., Krause, K.L., Pingoud, A., 1999. Cleavage experiments with deoxythymidine 3′,5′-bis-(*p*-nitrophenyl phosphate) suggest that the homing endonuclease I-PpoI follows the same mechanism of phosphodiester bond hydrolysis as the non-specific *Serratia* nuclease. FEBS Lett. 443, 209–214.

Haddow, J.D., Poulis, B., Haines, L.R., Gooding, R.H., Aksoy, S., Pearson, T.W., 2002. Identification of major soluble salivary gland proteins in teneral *Glossina morsitans morsitans*. Insect Biochem. Mol. Biol. 32, 1045–1053.

Harosh, I., Mezzina, M., Harris, P.V., Boyd, J.B., 1992. Purification and characterization of a mitochondrial endonuclease from *Drosophila melanogaster* embryos. Eur. J. Biochem. 210, 455–460.

Ho, H.C., Liao, T.H., 1999. Protein structure and gene cloning of *Syncephalastrum racemosum* nuclease. Biochem. J. 339 (Pt 2), 261–267.

Katoh, K., Kuma, K., Miyata, T., Toh, H., 2005. Improvement in the accuracy of multiple sequence alignment program MAFFT. Genome Inform. 16, 22–33.

Kuhlmann, U.C., Moore, G.R., James, R., Kleanthous, C., Hemmings, A.M., 1999. Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? FEBS Lett. 463, 1–2.

Kunitz, M., 1950. Crystalline desoxyribonuclease; digestion of thymus nucleic acid; the kinetics of the reaction. J. Gen. Physiol. 33, 363–377.

Liao, T.F., Pattison, M.L., Chen, C.L., 1974. Prolongation of pseudopregnancy of rats by gonadotropins implanted in the ovarian bursae. Endocrinology 95, 1234–1239.

Lunin, V.Y., et al., 1997. Three-dimensional structure of *Serratia marcescens* nuclease at 1.7 A resolution and mechanism of its action. FEBS Lett. 412, 217–222.

Matz, M., et al., 1999. Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res. 27, 1558–1560.

Matz, M.V., Alieva, N.O., Chenchik, A., Lukyanov, S., 2003. Amplification of cDNA ends using PCR suppression effect and step-out PCR. Methods Mol. Biol. 221, 41–49.

Meiss, G., Gimadutdinow, O., Haberland, B., Pingoud, A., 2000. Mechanism of DNA cleavage by the DNA/RNA-non-specific *Anabaena* sp. PCC 7120 endonuclease NucA and its inhibition by NuiA. J. Mol. Biol. 297, 521–534.

Miller, M.D., Krause, K.L., 1996. Identification of the *Serratia* endonuclease dimer: structural basis and implications for catalysis. Protein Sci. 5, 24–33.

Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J., Krause, K.L., 1994. 2.1 A structure of *Serratia* endonuclease suggests a mechanism for binding to double-stranded DNA. Nat. Struct. Biol. 1, 461–468.

Miller, M.D., Cai, J., Krause, K.L., 1999. The active site of *Serratia* endonuclease contains a conserved magnesium-water cluster. J. Mol. Biol. 288, 975–987.

Nekrasov, A.N., 2002. Entropy of protein sequences: an integral approach. J. Biomol. Struct. Dyn. 20, 87–92.

Nekrasov, A.N., 2004. Analysis of the information structure of protein sequences: a new method for analyzing the domain organization of proteins. J. Biomol. Struct. Dyn. 21, 615–624.

Peng, R.H., et al., 2008. Kamchatka crab duplex-specific nuclease-mediated transcriptome subtraction method for identifying long cDNAs of differentially expressed genes. Anal. Biochem. 372, 148–155.

Rangarajan, E.S., Shankar, V., 2001. Sugar non-specific endonucleases. FEMS Microbiol. Rev. 25, 583–613.

Salamov, A.A., Solovyev, V.V., 2000. Ab initio gene finding in *Drosophila* genomic DNA. Genome Res. 10, 516–522.

Schafer, P., et al., 2004. Structural and functional characterization of mitochondrial EndoG, a sugar non-specific nuclease which plays an important role during apoptosis. J. Mol. Biol. 338, 217–228.

Shagin, D.A., et al., 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res. 12, 1935–1942.

Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U. S. A. 94, 6815–6819.

Wang, W.Y., Liaw, S.H., Liao, T.H., 2000. Cloning and characterization of a novel nuclease from shrimp hepatopancreas, and prediction of its active site. Biochem. J. 346 (Pt 3), 799–804.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699.

Yang, 1996. The among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–372.

Zhao, Y., Hoshiyama, H., Shay, J.W., Wright, W.E., 2008. Quantitative telomeric overhang determination using a double-strand specific nuclease. Nucleic Acids Res. 36, e14.

Zhulidov, P.A., et al., 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. 32, e37.

Zhulidov, P.A., et al., 2005. A method for the preparation of normalized cDNA libraries enriched with full-length sequences. Bioorg. Khim. 31, 186–194.