

A new division of bacterial UvrA homologues

MARTA MARSZAŁKOWSKA¹, MAGDALENA BIL², ŁUKASZ KREFT¹, MARCIN OLSZEWSKI^{1*}

¹Faculty of Chemistry, Gdańsk University of Technology, Gdańsk, Poland

²Department of Biology, Katholieke Universiteit Leuven, Leuven, Belgium

* Corresponding author: molszewski@pg.gda.pl

Abstract

The UvrA protein is a DNA-binding and damage-recognition enzyme which participates in the prokaryotic type nucleotide excision repair (NER) pathway. It has recently been noted that some bacterial genomes comprise additional *uvrA* genes which encode five distinct types of UvrA homologue. We investigated the sequences of over 2400 bacterial genomes and found 130 examples of bacteria containing *uvrA2* genes. The sequence analyses conducted on these UvrA homologues revealed that the previously established division of UvrA proteins might be based on some incorrect assumptions. In this paper, we present the reasons for our creation of a new division of UvrA homologues and a description of the four UvrA classes we have created.

Introduction

The nucleotide excision repair (NER) pathway operating throughout the prokaryotic genome is a versatile DNA repair system that can remove a wide variety of lesions which often differ in their chemical structure. Damage recognition is performed by the UvrA protein, which belongs to the ABC ATPase superfamily. Bacterial UvrA is a dimeric protein, unique among DNA repair enzymes. This feature enables UvrA to detect various DNA lesions by using an indirect readout mechanism (Jaciuk et al. 2012). Together with other components of NER, namely UvrB, UvrC, and UvrD, the UvrA protein is essential to the survival of almost every living bacterium (Truglio et al. 2006).

It has recently been discovered that chromosomes in some prokaryotic organisms, such as *Xanthomonas axonopodis* (Shen et al., 2007), *Pseudomonas putida* (Tark et al., 2008) and *Deinococcus radiodurans* (Tanaka et al., 2005), contain additional *uvrA* genes coding UvrA homologues. The analysis of some UvrA homologues has revealed that these proteins can be divided into five different classes. Class I proteins are full-length UvrA proteins which are present in most bacterial species, including *E. coli*. Class II and III proteins contain large deletions, each of a different region. Class IV proteins are duplications of the full-length UvrA protein and Class V proteins are duplications with a deletion in the C-termi-

nal half (Goosen & Moolenaar 2008). It is notable that the authors of this last publication suggest that, due to different deletions, some classes of UvrA proteins may have lost their ability to participate in NER.

Methods

The amino acid (aa) sequences of the UvrA proteins were analyzed using standard protein-protein BLAST and RPS-BLAST. Multiple sequence alignments were created using the MAFFT program and the results were analyzed and edited using the GeneDoc editor program (copyright by Karl Nicholas).

Results and discussion

We searched for additional *uvrA* genes in an analysis of over 2400 bacterial genomes and found 130 examples of bacteria containing *uvrA2* genes. As the next step, we decided to verify whether sequences of these genes contain mutations or deletions in significant domains, and might thus be assigned to some of the previously described classes of UvrA homologues. We have found that some of the rules underlying the first division of UvrA homologues might be incorrect, as a third zinc finger motif and UvrB-binding domain have been discovered (Pakotiprapha et al., 2009) and the proteins assigned to classes III and V do not contain certain deletions. Thus,

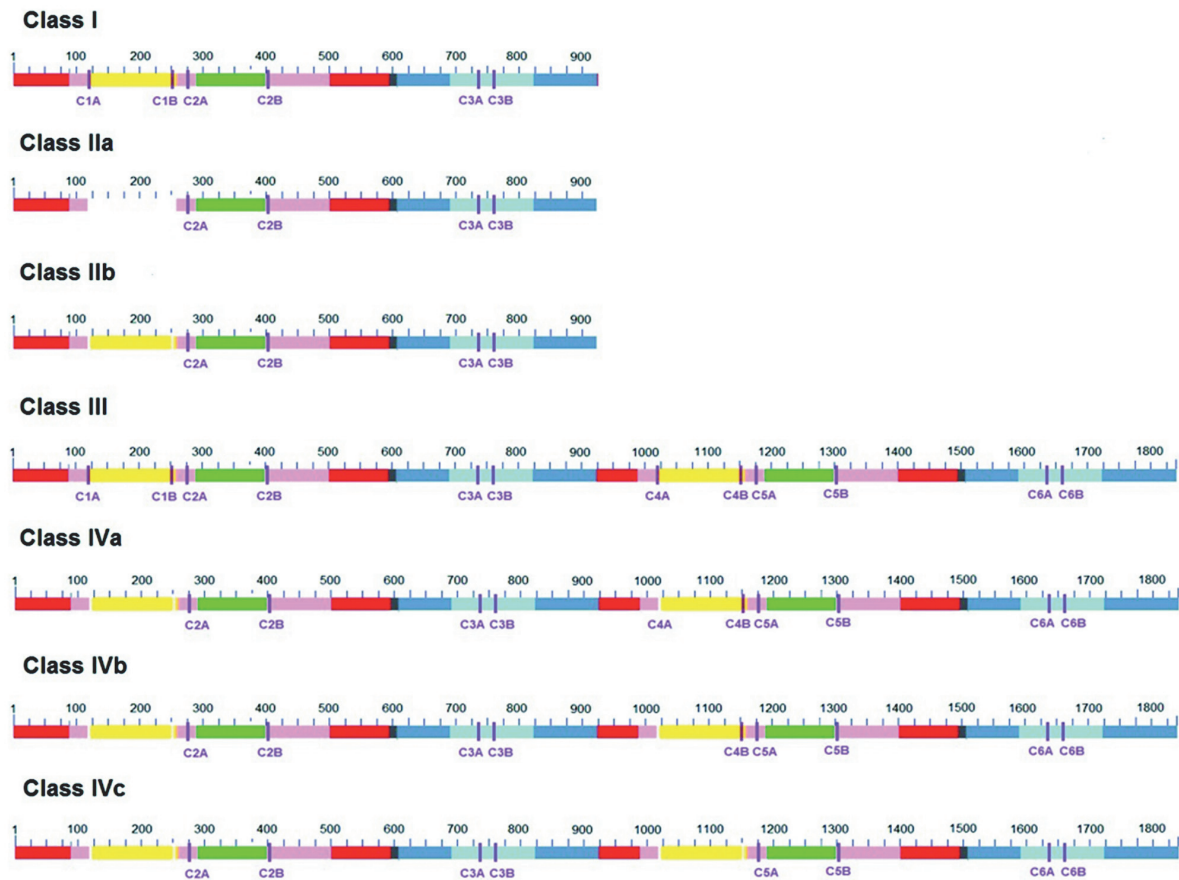


Fig. 1. Sequences are colored according to the domain: ATP-binding domain I, III, red; signature motif I, III, pink; ATP binding domain II, IV, blue; signature motif II, IV, cyan; UvrB-binding domain I, II, yellow; insertion sequence I, II, green; cysteine clusters (C1A – C6B), violet; lack of cysteine clusters or UvrB-binding domain indicates deletions

we decided to create a new division founded on both our *uvrA* sequence base and current knowledge regarding UvrA proteins.

The new division consists of four main classes of UvrA proteins and five subclasses (Fig. 1).

- Class I (900-1050 aa): the basic UvrA protein type, present in most bacteria; sequences of this type of UvrA homologue contain two ABC-ATPase sites, a UvrB-binding domain and three zinc finger motifs.
- Class IIa (740-880 aa): this class of UvrA homologue contains the deletion of the UvrB-binding domain and the first zinc finger motif; in other words, there are no 120-140 amino acids residues; proteins ascribed to this class are present in different species of Actinobacteria, Firmicutes, Bacteroidetes and Proteobacteria.
- Class IIb (800-900 aa): this type of UvrA sequence contains the deletion of the first zinc finger motif; all other domains, including the UvrB-binding domain, are present; Class IIb proteins can be found in all species of Chlorobi and in some species of Firmicutes and Proteobacteria.
- Class III (1750-2012 aa): this type of UvrA protein comprises class I homologues which have been duplicated to form a protein of over 200 kDa; all ATPase sites are present, as are three zinc finger motifs and two UvrB-binding domains; these types of proteins are present in Proteobacteria, particularly in many species of Betaproteobacteria and Planctomycetes.
- Class IVa (1800-1950 aa): similar to class III, sequences of this type of UvrA protein have been duplicated, but they lack the first zinc finger motif; these homologues can be found in several species of Betaproteobacteria and Chlamydiae.
- Class IVb (1900-2000 aa): these UvrA homologues differ from classes III and IVa in that they contain deletions of the first zinc finger motif and the first cysteine

cluster of the fourth zinc finger motif; they are present in many species of Betaproteobacteria.

- Class IVc (1780-1900 aa): like all the other subtypes of class IV UvrA proteins, the sequence of class IVc has been duplicated, but is characterized by the lack of the first and fourth zinc finger motifs; these homologues can be found in several species of Verrucomicrobia and Proteobacteria; in Chlamydiae they are considered to be the only type of UvrA homologue present in the cell.

To date, three different UvrA protein crystal structures have been identified. Two are class I proteins, *Bacillus stearothermophilus* (Pakotiprapha et al. 2009) and *Thermotoga maritima* (Jaciuk et al., 2012), and one is a class IIa homologue, *Deinococcus radiodurans* (Timmins et al., 2009). The crystal structures of classes IIb, III and IVa-c have not yet been obtained.

It is notable that only the function of class I proteins has been thoroughly examined and fully proven (Goosen and Moolenaar, 2008). The exact role of class II-IV UvrA homologues in bacteria cells remains unclear. However, analyses conducted on the sequences of two proteins responsible for antibiotic resistance, DrrC from *Streptomyces peucetius* (Lomovskaya et al., 1996) and SnrO from *Streptomyces nogalater* (Ylihonko et al., 1996), have revealed that they belong to the group of class IIa UvrA proteins. This example demonstrates that the function of additional *uvrA* genes might not be connected with DNA repair.

Conclusions

The new division of UvrA proteins, created on the basis of our *uvrA* sequence base, consists of four main UvrA

classes and five subclasses. The homologues subscribed to classes I, IIa and IIb are characterized by a length of between 750-1050 aa and a widespread occurrence among different species of bacteria. However, their sequences differ in terms of the presence of certain domains, such as the UvrB-binding domain and the first zinc finger motif. In contrast, the UvrA homologues of classes III and IVa-c, the sequences of which have been duplicated to form a protein containing between 1750 and 2000 aa, are only present in some species of Proteobacteria, Chlamydiae and Verrucomicrobia. Apart from class III, they are characterized by deletions in the cysteine clusters of the first and fourth zinc finger motifs.

Our analyses demonstrate that bacteria cells contain a great variety of UvrA homologues for which the structures and functions are still unknown.

References

- Goosen N., Moolenaar G.F. (2008) DNA repair 7: 353-379.
- Jaciuk M., Nowak E., Skowronek K., Tańska A., Nowotny M. (2012) Nat. Struct. Mol. Biol. 18: 191-197.
- Pakotiprapha D., Inuzuka Y., Bowman B., Moolenaar G., Goosen N., Jeruzalmi D., Verdine G. (2008) Mol. Cell 29: 122-133.
- Shen C.H., Chiang Y.C., Hsu C.H., Yang M. (2007) Mol. Genet. Genomics 277: 149-160.
- Tanaka M., Narumi I., Funayama T., Kikuchi M., Watanabe H., Matsunaga T., Nikaido O., Yamamoto K. (2005) J. Bacteriol. 187: 3693-3697.
- Tark M., Tover A., Koorits L., Tegova R., Kivisa M. (2008) DNA repair 7: 20-30.
- Timmins J., Gordon E., Caria S., Leonard G., Acajjaoui S., Kuo M., Monchois V., McSweeney S. (2009) Cell 17: 547-558.
- Truglio J.J., Croteau L., Van Houten B., Kisker C. (2006) Chem. Rev. 106: 233-252.