Dottorato in "Produzioni Animali, Biotecnologie Veterinarie, Qualità e Sicurezza degli Alimenti" **XXI° ciclo, Università degli Studi di Parma**

and

Doctorat en "Bioinformatique, Biologie Structurale et Génomique"
Ecole Doctorale des Sciences de la Vie et de la Santé,
Université de la Méditerranée - Aix-Marseille II

Co-tutoring PhD Thesis

"Biochemical, biophysical and structural characterization of *Lactococcus lactis* bacteriophage proteins involved in Homologous Recombination and Abortive Infection mechanisms"

Candidate: Scaltriti Erika   Doctorate Coordinator: Prof Primo Mariani

Supervisor: Prof. Stefano Grolli   Co-supervisor: Prof. Mariella Tegoni

Examinators:
Prof. Christian Cambillau
Prof. Sylvain Moineau
Prof. Roberto Ramoni
Prof. Douwe Van Sinderen

Index

Charter 2: Biochemical, biophysical and structural characterisation of Lactococcus lactis bacteriophage proteins involved in Homologous Recombination and Abortive Infection Mechanisms

Abbreviations

Abi: Abortive Infection

AFM: Atomic Force Microscopy

BIR: Break-Induced Replication

dsDNA: double strand DNA

DSB: Double-Strand Break

DSBR: Double-Strand Break Repair

EM: Electron Microscopy

EMSA: Electrophoretic Mobility Shift Assay

HPLC: High Pressure Liquid Chromatography

HR: Homologous Recombination

LAB: Lactic Acid Bacteria

MALS/UV/RI: Multiangle Laser Light Scattering/ UV Absorbance /Refractive Index

OB fold: Oligonucleotide/Oligosaccaride Binding fold

ORF: Open Reading Frame

RDR: Recombination Dependent Replication

RMP: Replication/Recombination Mediators Proteins

RPA: Replication Protein A

RT: Reverse Transcriptase

SDSA: Synthesis-Dependent Strand Annealing

SEC: Size Exclusion Chromatography

SPR: Surface Plasmon Resonance

SSA: Single Strand Annealing

SSAP: Single Strand Annealing Protein

SSB protein: Single Strand DNA Binding Protein

ssDNA: single strand DNA

Summary

Virulent phages of the 936 and P335 group are predominant in infections to *Lactococcus lactis*, a Gram-positive bacterium widely used in the dairy industry. Among the mechanisms adopted by the bacterium against virus infection, abortive infection mechanisms (Abi) were found, but little is known about their molecular way of action. Under the selective pressure of Abi systems, homologous recombination with phage-related sequences present in the host chromosome (Bouchard et al 2000 and 2002, Moineau et al 1995) produces phage mutants which are found in unsuccessful fermentations that lead to huge economical looses. In particular for the AbiK system, Bouchard *et al* (Bouchard et al 2004) have identified four non similar proteins involved in the sensibility to AbiK named Sak. Among these proteins, Ploquin et al (Ploquin et al 2008) demonstrated that ORF252 of phage ul36 (species P335), renamed Sak, is a DNA Single-Strand Annealing Protein (SSAP), homologue of the eukaryotic protein RAD52. Sak3 from phage p2 (species 936) is reported to be the reference for the third group of identified *sak* genes that do not show significant sequence similatity to SSAPs. However, Sak proteins were suspected to be involved in homologous recombination due to their genetic localization in "recombination cassette" with other genes of proteins clearly involved in these processes (Bouchard and Moineau, 2004).The first part of our studies present the biochemical and structural characterization of two Sak proteins: the Sak of phage UL36 and the Sak3 of phage p2. In particular, multi-technique approaches were performed to investigate the oligomerization state and to visualise ring structures (Cryo Electron Microscopy), while functional tests were used to elucidate the SSAP family membership. DNA binding properties were also tested by gel mobility shift assays (EMSA), atomic force microscopy (AFM) and surface plasmon resonance (SPR), and the relationship with their partners in homologous recombination was determined. In the second part of our studies the phage p2 single-strand binding protein (SSB), a Sak3 partner in homologous recombination/replication, was studied. Expression level during phage infection, binding properties to ssDNA, involvement in homologous recombination processes and tridimensional structure were elucidated. Our studies might complete and help to understand some aspects of the Sak function leading to the definition of a linkage between DNA processing events and the antiviral system AbiK, and might elucidate the implication of correlated recombinant proteins like SSBs in these processes.

Resumé

Les bacteriophages virulents du groupe 936 et P335 sont prédominants dans les infections au *Lactococcus lactis*, une bactérie Gram-positive largement utilise dans la société laitière. Parmi les mécanismes que la bactérie utilise contre l'infection du virus, nous retrouvons le mécanisme abortif antivirale (Abi), mais très peu est connu sur son action à niveau moléculaire. Sous la pression sélective des systèmes Abi, la recombinaison homologue avec des séquences que nous retrouvons dans le chromosome (Bouchard et al 2000 et 2002, Moineau et al 1995) de la bactérie produit des phages qui sont mutés et qui sont retrouvés dans les fermentation qui ont échoué. Pour le système AbiK, Bouchard *et al* (Bouchard et al 2004) ont identifié quatre protéines appelées Sak, qui n'ont pas de similitude mais qui sont impliqués dans la sensibilité vers AbiK. Parmi ces protéines, Ploquin et al (Ploquin et al 2008) ont démontré que le ORF252 du phage ul36 (espèce P335), appelé Sak, est une "DNA Single Strand Annealing Protein (SSAP)", homologue de la protéine eukaryote RAD52. Sak3 du phage p2 (espèce 936) est la protéine référence du troisième groupe des gènes *sak* qui ne montrent aucune similarité de séquence avec SSAPs. Cependant, les protéines Sak sont sûrement impliquées dans la recombinaison homologue en raison de leur localisation dans une "recombination cassette" avec les autres gènes des protéines clairement impliquées dans ces mécanismes (Bouchard and Moineau, 2004). La première partie de notre étude a été focalisée sur la caractérisation biochimique et structurale de deux Sak protéines : Sak du phage UL36 et Sak3 du phage p2. Nous avons étudié leur state d'oligomerisation au moyen d'une approche multi-technique, nous avons visualisé les structures des anneaux (Cryo Microscopie Electronique) et utilisé des tests fonctionnels pour élucider leur appartenance à la famille de SSAP. La capacité de lier l'ADN a été prouvée et étudiée par des Gels retard (« Electrophoretic Mobility Shift Assays, EMSA »), par Microscopie à Force Atomique (AFM), et Resonance Plasmonique de Surface (SPR); la relation avec d'autres partenaires dans la recombinaison homologue a aussi été testée par SPR. Dans la deuxième partie de notre étude, nous avons caractérisé la "Single-Strand Binding protein (SSB)" du phage p2, partenaire de Sak3 dans la réplication-recombinaison homologue. Nous avons étudié le niveau d'expression de SSB dans l'infection du phage, sa capacité à lier l'ADN simple brin, son implication dans le mécanisme de recombinaison homologue et nous avons résolu sa structure tridimensionnelle. Nos résultats sont des éléments qui vont aider à la compréhension de quelques fonctions inconnues des protéines Sak, en particulier le lien entre les événements du processing de l'ADN et le système AbiK, et l'implication de protéines telles que les SSB dans ces mécanismes.

Riassunto

I batteriofagi virulenti del gruppo 936 e P335 sono predominanti nelle infezioni verso *Lactococcus lactis*, un batterio Gram-positivo molto diffuso nell'industria casearia. Tra i meccanismi adottati dal batterio contro l'infezione virale, ritroviamo i meccanismi abortivi antivirali (Abi), di cui poco si conosce riguardo alla loro modalità d'azione. Sotto la pressione selettiva dei sistemi Abi, la ricombinazione omologa con sequenze presenti nel cromosoma batterico (Bouchard et al 2000 et 2002, Moineau et al 1995) produce fagi mutati che sono stati ritrovati in fermentazioni ad esito negativo e perciò recanti enormi perdite economiche. Riguardo il sistema AbiK, Bouchard *et al* (Bouchard et al 2004) hanno identificato quattro proteine denominate Sak, a bassa similarità di sequenza, ma tutte implicate nella sensibilità verso AbiK. Tra queste proteine, Ploquin et al (Ploquin et al 2008) hanno dimostrato che la ORF252 del fago ul36 (specie P335), denominata Sak, é una "DNA Single Strand Annealing Protein (SSAP)", omologa della proteina eucariotica RAD52. Sak3 del fago p2 (specie 936) é invece la proteina di riferimento del terzo gruppo di geni *sak* che non mostrano alcuna similarità di sequenza con le SSAPs. Indipendentemente da questo, le proteine Sak sono sicuramente coinvolte nel meccanismo di ricombinazione omologa poiché i loro geni sono localizzati in « recombination cassette » insieme a geni di altre proteine chiaramente implicate in questi meccanismi (Bouchard and Moineau, 2004).

La prima parte dei nostri studi é stata focalizzata sulla caratterizzazione biochimica e strutturale delle due proteine Sak : Sak del fago UL36 e Sak3 del fago p2. Di queste ultime abbiamo studiato lo stato di oligomerizzazione attraverso un approccio multi-tecnica, abbiamo visualizzato le tipiche strutture ad anello (Cryo Microscopie Electronique) ed allestito test funzionali per chiarire la loro appartenenza alla famiglia delle SSAP. La loro capacità di legare il DNA é stata inoltre testata e studiata attraverso « Electrophoretic Mobility Shift Assays, EMSA », Microscopia a Forza Atomica (AFM) e « Surface Plasmon Resonance, SPR » ; la relazione con altri partners nel processo di ricombinazione omologa é stata inoltre testata tramite SPR. Nella seconda parte del nostro studio, abbiamo caratterizzato la "Single-Strand Binding protein (SSB)" del fago p2, partner di Sak3 nella ricombinazione omologa. A questo proposito sono stati studiati il livello di espressione di SSB durante l'infezione fagica, la sua capacità di legare il DNA a singolo filamento, le sue implicazioni nel meccanismo di ricombinazione omologa ed abbiamo determinato la sua struttura tridimensionale. I nostri risultati costituiscono gli elementi che permetteranno la comprensione di alcune funzioni sconosciute delle proteine Sak, in particolare del legame tra gli avvenimenti del processamento del DNA ed il sistema AbiK, dell'implicazione di altre proteine correlate come le SSBs in questi meccanismi.

Preface:

The subject of my thesis is part of a collaboration program between the AFMB-CNRS Laboratory (Christian Cambillau's group) and the University of Laval (Sylvain Moineau's group) directed to the study of *Lactococcus lactis* phage proteins. In particular the abilities in microbiology and genetics of our colleagues of the Laval University and those of biochemistry, molecular and structural biology of our group in AFMB are the basis for a detailed characterization of both the structure and function of these proteins and of their involvement in phage infection mechanisms and in general in phage-host relationship.

In my work of thesis, I have studied proteins involved both in phage replication/recombination processes and in the Antiviral Abortive Infection (Abi) mechanisms adopted by *Lactococcus lactis* to prevent phage attack. In the bibliography that follows this preface, I will try to give an overview, as complete as possibile, of these two important themes. This introduction is supposed to present the three papers reported in Chapter 2 resulting from my thesis work.

Chapter 1: Introduction

1.1 Lactococcal phages: state of the art

1.1.1 <u>Inside the word of bacteriophages: Lactococcal phages</u>

1.1.1.1 Bacteriophages

Bacteriophages were firstly described in 1915 and 1917, respectively, by Frederick William Twort and Félix Hubert d'Hérelle that separately observed the lysis of bacteria on agar and in broth cultures. Subsequently, the International Committee for Taxonomy of Viruses (ICTV) classified phages into six morphological group based on gross morphology and nucleic acid corresponding to Bradley's basic types (Bradley, 1967). At present the ICTV classified bacteriophages in one order, 13 families and 30 genera (Tab. 1). Bacteriophages virions can be tailed, polyhedral, filamentous or pleomorphic (Fig. 2a). By the year 2006 more than 5500 phages from over 150 eubacterial and archaeal genera had been examined by electron microscopy for morphological description. Most of these examined phages were tailed (96%) (Ackermann 2007), constituting an absolute majority of "organisms" on our home planet in sheer numbers ($>10^{30}$ tailed phages in the biosphere) (Fig. 1).

Earth with 10^{30} beetles

Earth with 10^{30} phages
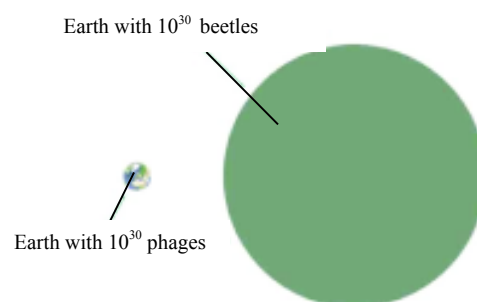
Fig.1. "The Creator Must have an inordinate fondness for beetles" (J.B.S. Haldane). Tailed phages are vastly more abundant than we had imagined. The figure illustrates the consequences if 10^{30} phages were to be transmogrified into Haldane's beetles (Brussow and Hendrix, 2002).

Tailed phage are classified into the order *Caudovirales* and are divided into three large, phylogenetically related, families: *Myoviridae* (contractile tail, 25%), *Siphoviridae* (long noncontractile tail, 61%) and *Podoviridae* (short non contractile tail, 14%). Usually, in virology families are defined by the nature of nucleic acid and particle morphology and the classificatio is mainly based on tail morphology. Tailed phages virions consist of a head with cubic symmetry (prevalently icosahedral) containing the nucleic acid, and a tail of variable lengths (10 to 800 nm) ended by structures , such as base plates, spikes or fibers (Fig. 2b), involved in the interaction with the host. The DNA is a single linear, double-stranded filament and has a size variable between 17 and over 700 kb (Fig. 2a). Replicating DNA usually forms large branched intermediates or "concatamers", that are subsequently cut to a size that fits the preformed capsids (Ackermann HW, 2003). For practical reasons, tailed phages were subdivided according to a set of properties, some of which may be absent in a given member (Van Regenmortel M H V,1990).

Classification and basic properties of bacteriophages [a]

Shape	Nucleic acid	Order and families	Genera	Examples	Members	Characteristics
Tailed	DNA, ds, L	*Caudovirales*	15		4950	
		Myoviridae	6	T4	1243	Tail contractile
		Siphoviridae	6	λ	3011	Tail long, noncontractile
		Podoviridae	3	T7	696	Tail short
Polyhedral	DNA, ss, C	*Microviridae*	4	φX174	40	
	ds, C, T	*Corticoviridae*	1	PM2	3?	Complex capsid, lipids
	ds, L	*Tectiviridae*	1	PRD1	18	Internal lipoprotein vesicle
	RNA, ss, L	*Leviviridae*	2	MS2	39	
	ds, L, S	*Cystoviridae*	1	φ6	1	Envelope, lipids
Filamentous	DNA, ss, C	*Inoviridae*	2	fd	57	Filaments or rods
	ds, L	*Lipothrixviridae*	1	TTV1	6?	Envelope, lipids
	ds, L	*Rudiviridae*	1	SIRV1	2	Resembles TMV
Pleomorphic	DNA, ds, C, T	*Plasmaviridae*	1	L2	6	Envelope, lipids, no capsid
	ds, C, T	*Fuselloviridae*	1	SSV1	8?	Spindle-shaped, no capsid

[a]Modified from reference [4]. With permission of John Libbey-Eurotext. Phage numbers are from reference [3]. C, circular; L, linear; S, segmented; T, superhelical; 1, single-stranded; 2, double-stranded.

Tab.1 Classification and basic properties of bacteriophages (Ackermann H W, 2003)
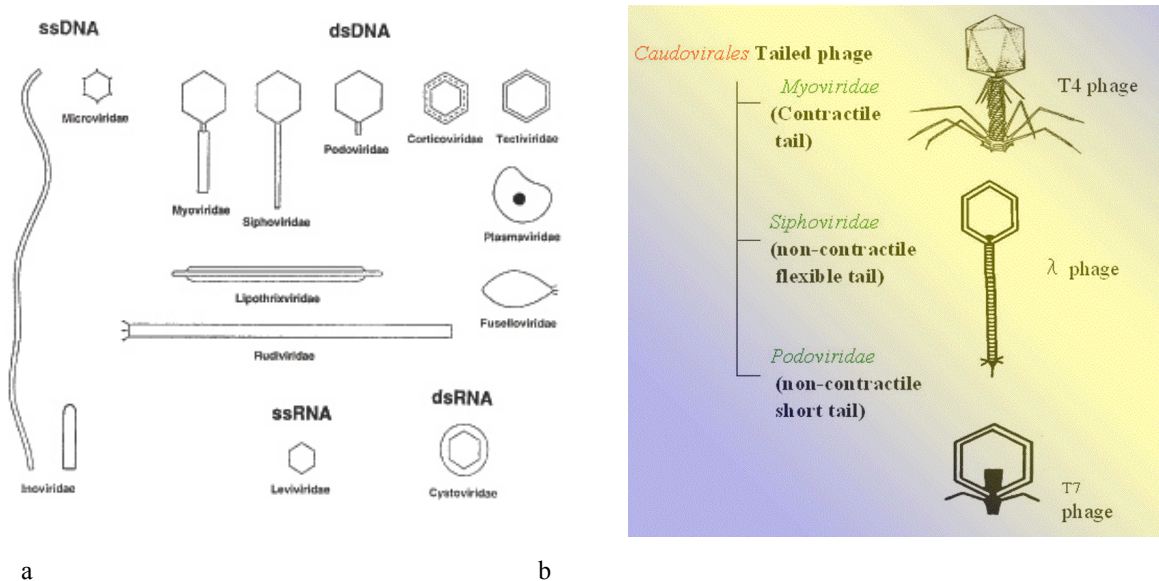
Fig. 2. Schematic representation of the major phage groups (a) (Ackermann HW, 2003). The order Caudovirales and the three related families (b)

(From www.protein.osakau.ac.jp/rcsfp/supracryst/vsrf/virus.bio.html).

1.1.1.2   Phage life cycle

Phage life cycle can be divided into several steps that are quite different in the case of litic or lysogenic phages (Fig.3). Lytic or virulent phages are those which can only multiply in bacteria and kill the cell, by lysis, at the end of the life cycle. Lysogenic or temperate phages are those that can either multiply via the lytic cycle or enter and remain in a quiescent state in the bacterial cell. In this quiescent state most of the phage genes are not transcribed and the phage genome remains in a repressed state called prophage. In this state, the phage DNA can persist as a free plasmid or can be integrated into the host chromosome and replicated along with the host DNA thus passing to the daughter cells. The cell harboring a prophage (lysogen) are not adversely affected by the presence of their DNA and the lysogenic state may persist even indefinitely.

For example the phage lambda, whose DNA is integrated into its host genome by recombination at specific sites (Weisberg and Landy, 1983), enters the lytic or lysogenic cycle depending on the relative levels of two antagonist phage regulatory proteins, Cro and cI. The Cro protein turns off the synthesis of the cI repressor and

thus prevents the establishment of lysogeny. Environmental conditions that favor the production of Cro will lead to the lytic cycle, while those that favor the production of the cI repressor will favor lysogeny (Ackermann and DuBow, 1987).

Independently on the phage type and on the cycle, (lytic or lysogenic), as shown in Fig. 3, at the beginning of the infection, the phage absorbs to the bacterial cell surface using tail fibers or analogous structure for those lacking tail fibers. These bind to specific receptors which are responsible of the host specificity, that is driven by molecular characteristics of the same tail fibers. The nature of the bacterial receptors, that usually play other functional roles, varies for different bacteria and includes proteins on the outer surface of the bacterium, lipopolysaccharides (LPS), pili, and lipoproteins.
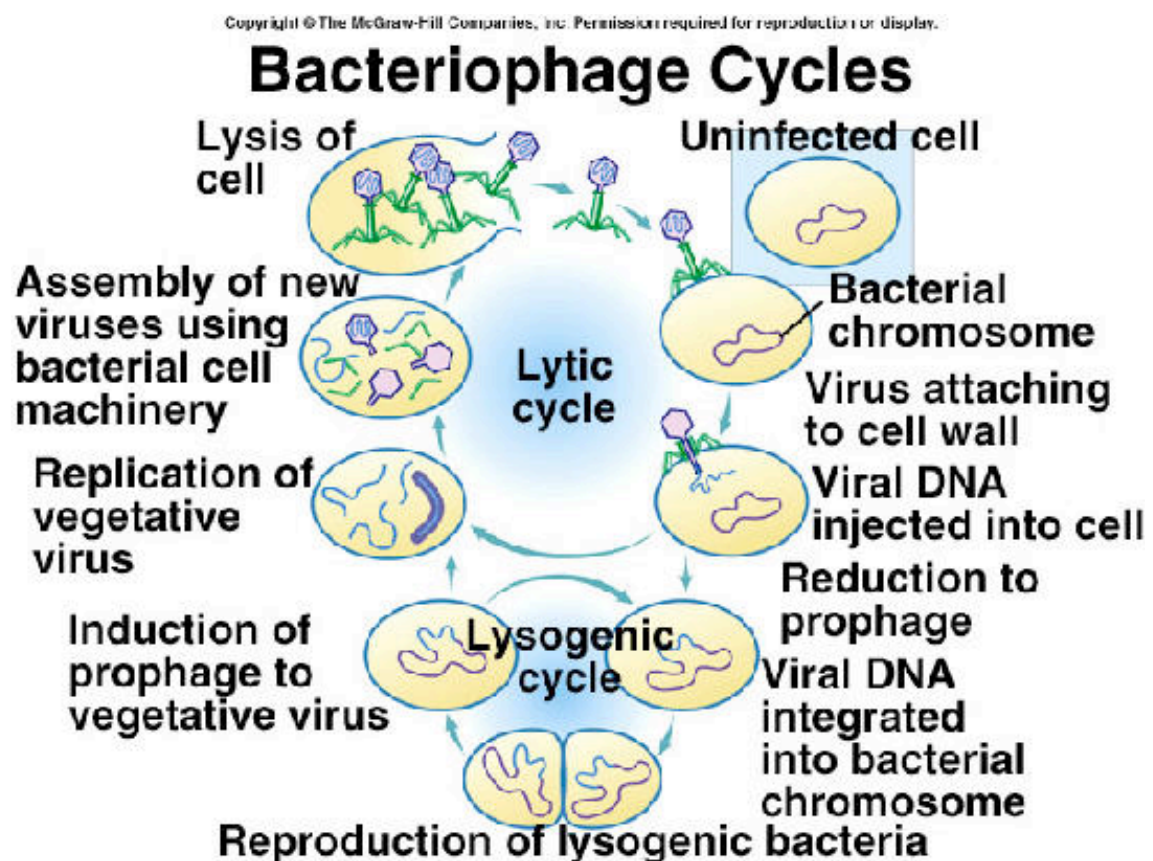

Fig. 3. Phage life cycle (The Mc Graw-Hill Companies, Inc).

The binding of the phage to the bacterium, via the tail fibers alone, is usually weak and is reversible. The irreversible binding of phage to a bacterium, thereby, is mediated by the interaction of one or more components of the base plate. This interaction results in the contraction of the sheath (for those phages which have a sheath) and the hollow tail fiber is pushed through the bacterial envelope. Phages lacking contractile sheaths use other mechanisms, like enzymatic digestion, to penetrate the bacterial envelope. After the perforation of the bacterial envelope, the nucleic acid from the head passes through the hollow tail and enters the bacterial cell, starting an eclipse phase. During this phase, no infectious phage particles can be found either inside or outside the bacterial cell. At this level, phages can adopt litic or lysogenic behaviour. In the case of litic phages, the transcription of phage early genes takes over the host biosynthetic machinery and phage early genes, coding for proteins involved in phage DNA synthesis and in shutting off of host DNA, RNA and protein biosynthesis, are transcribed. Subsequently late m-RNAs are transcribed and late proteins, coding for structural elements, are traduced. This transcriptional phase is followed by an intracellular accumulation phase in which the nucleic acid and structural proteins are accumulated within the bacterial cell and then infectious phage particles are assembled. After a DNA maturation and packaging phase, the accumulation of intracellular phages leads to the release of phages as a consequence of bacterial cell lysis (Lansing M. et al 1993).

1.1.1.3   Phage Genomic organization

Bacteriophages infecting Gram-positive and Gram-negative bacteria share a similar genetic organization; even sequence similarities can be found to some degree (Desiere et al 1999, Hendrix et al 1999). Genes for related functions are clustered together and in general the minimal genome of tailed phages is divided into several regions that encode proteins for DNA packaging, head, tail and tail fibers, DNA replication, transcription regulation, lysogenic conversion and lysis (Fig.4). These multigenic elements (modules) are considered as functional units and can be squeezed into a 20 kb DNA genome as demonstrated by c2 Siphovirus and φ29 Podovirus. As the genome size increases, the virion morphology gets more

complicated, and the phage interferes more with cellular activities. However we are still far from understanding these changements in details (Brussow and Hendrix, 2002). The modular genome organization is at the basis of phage evolution; where each member of a phage family is the product of interchangeable genetic elements (Brussow and Desiere, 2001).

Comparative phage genomics can retrace part of the evolutionary history of the modules encoding phage-specific functions that can be the topological basis of genome organization. With this approach, two new genus (Sfi21-like and Sfi11-like) represented by two unrelated configurations characteristic of *cos-* or *pac*-site phages belonging to *Siphoviridae* family, have been proposed (Lucchini et al 1999). Also the gene maps from λ-, ψ-M2, L5-, Sfi21-, Sfi11-, f-C31, sk1-, and TM4-like phages showed common features of their structural genes defining a λ supergroup within *Siphoviridae*. Tailed phages are the result of both vertical and horizontal evolution and in particular a hierarchy of relatedness within the λ supergroup suggested elements of vertical evolution in the capsid module of *Siphoviridae* (Brussow and Desiere 2001). On the basis of studies of comparative genomics, Proux et al propose to base phage taxonomy on a single structural gene module (head or tail genes), but the peculiar modular nature of phage evolution creates ambiguities in the definition of phage taxa by comparative genomics (Proux et al 2002).
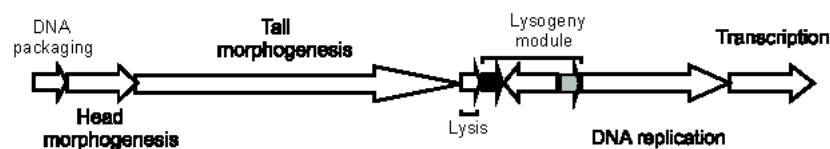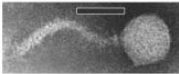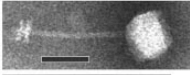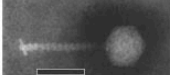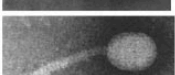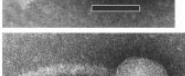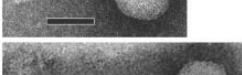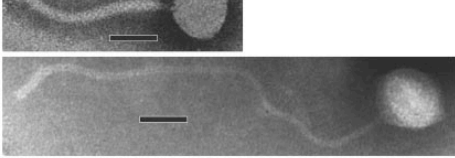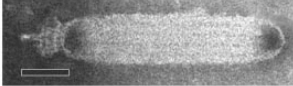
Fig. 4. General genome organization of r1t-, Sfi11-, Sfi21-like phages (modified from Desiere et al 2002).

1.1.1.4   *Lactococcal phages*

Lactic acid bacteria (LAB) belong to the low G+C content (55% or less) class of Gram positive bacteria, and together with *Clostridium, Bacillus, Listeria and Staphylococcus* constitute a diverse group of micro-organisms known for their economic importance in fermentation and processing of milk, meat, alcoholic

beverages and vegetables (Carr et al 2002). In particular some species of lactobacilli, lactococci and streptococci are widely used as starters in dairy fermentation processes. Among commercially used LAB one of the most important is *Lactococcus lactis*. Thi*s* is a gram-positive bacterium, usually isolated from plant material but also present in milk and dairy products probably by the way of cow food (Teuber M 1995). They are non-motile, coccus-shaped, homofermentative bacteria that grow between 10°C and 40°C but not at 45°C; they grow in the presence of 4% NaCl and produce L (+)-lactic acid from lactose (Carr et al, 2002). Strains of *Lactococcus lactis* are used in dairy industry to acidify milk in the preparations of fermented products, such as cheese, buttermilk and sour cream. The use of various *L.lactis* strains in the same starter culture is essential for controlling virulent phages that are responsible of most milk fermentation collapses (Moineau and Levesque, 2005). Lactococcal phages are ubiquitous in the dairy environment, as they are found in raw milk and survive pasteurization (Madera et al, 2004). As bacteriophage infection leads to the lysis of the starter cells and thereby to the blockage of the transformation of lactose into lactic acid, the consequences are a fermentation delay, an alteration of the product quality and, in severe cases, the complete loss of the product (Brussow, 2001). Although phage biodiversity increases rapidly, many different lactococcal phages have been isolated and are classified into 12 species (Jarvis et al 1991), all belonging the *Caudovirales* order (Ackermann, 1999). The same investigation have shown that industrial phage ecology is dominated by 3 phage species (Fig.5). Most of these, representing about half of all isolates, were virulent, small-sized isometric-headed *Siphoviridae* of the 936-species. A quarter of the isolates were classified into the c2-species of virulent lactococcal phages, *Siphoviridae* with prolate heads. A further quarter of the isolates were represented by small isometric phages of the P335 species. These are heterogeneous and comprise virulent and temperate phages and phages with two DNA packaging mechanisms (*co*hesive end *s*equence, *cos*-site and *pac*kaging sequence, *pac*-site, phages) (Brussow, 2001). Common features among these three phage species include a double-stranded DNA, a long noncontractile tail, and their affiliation to the *Siphoviridae* family.

Family	Phage	Capsid diameter (nm)	Tail width (nm)	Tail length (nm)	Electron micrograph[a]
Species					
Siphoviridae					
936	bIL170	50	11	126	
P335	ul36	49	7	104	
1358	1358	45	10	93	
c2	c2	54 X 41	10	95	
Q54	Q54	56 X 43	11	109	
P087	P087	59	14	163	
949	949	70	12	490	
1706	1706	58	11	276	
Podoviridae					
P034	P369	57 x 40	5	19	
KSY1	KSY1	223 X 45	6	32	

[a] Bars, 50 nm.

Fig.5. Morphological diversity of bacteriophages infecting *Lactococcus lactis* (Deveau et al, 2006).

The majority of lactococcal phages are lytic, in particular phages belonging to 936- and c2-like species, whereas the P335 species includes both lytic and temperate phages (Djordjevic and Klaenhammer, 1997; Labrie and Moineau, 2000).

To date, more than 20 complete lactococcal phage genome sequences, belonging to the three abundant lactococcal phage species and to other rare species like Q54 (Fortier at al, 2006), KSY1 (Chopin et al, 2007) and 1706 (Garneau et al, 2008) are available in public databases (Fig.6).

According to the genetic organization of the conserved head cluster recently proposed as a new basis for phage classification (see section 1.1.1.3, Brussow and Desiere 2001, Proux et al 2002), the sequenced lactococcal phages could be divided into five genera: c2-, sk1-, Sfi11-, Sfi21- and r1t-like phages (Tab 2) (Proux et al 2002).

Bacterial host	Strain	Phage	Origin	Genome size (bp)	Old classification	New classification	Lytic or lysogenic	Reference
L. casei	ATCC 393	A2	Cheese	43,411	Unclassified	Sfi21 group	Lysogenic	This work
L. lactis subsp. cremoris	UC509	Tuc2009	Cheese	38,347	P335-type	Sfi11 group	Lysogenic	Accession no. AF109874[a]
	H2L	BK5-T	Cheese	40,003	BK5-T	Sfi21 group	Lysogenic	20
	901-1	TP901-1	Cheese	37,667	P335 type	Sfi11 group	Lysogenic	9
	R1	r1t	Cheese	33,350	P335 type	r1t group	Lysogenic	53
	IL1403	bIL67	Cheese	22,195	c2 type	c2 group	Lytic	49a
L. lactis subsp. lactis	MG 1363	c2	Cheese	22,163	c2 type	c2 group	Lytic	37
L. lactis subsp. cremoris	HID113	sk1	Cheese	28,451	936 type	sk1 group	Lytic	15a
L. lactis subsp. lactis	IL1403	bIL170	Cheese	31,754	936 type	sk1 group	Lytic	17
	IL1403	bIL309	Cheese	36,949	P335 type	Sfi21 group	Lysogenic	16
	IL1403	bIL285	Cheese	35,538	P335 type	Sfi21 group	Lysogenic	16
	IL1403	bIL286	Cheese	41,834	P335 type	Sfi21 group	Lysogenic	16
S. aureus	ATCC 49775	PVL	PVL-producing *S. aureus*	41,401	Unclassified	Sfi21 group	Lysogenic	32
S. thermophilus	B106	7201	Yogurt	35,466	Unclassified	Sfi21 group	Lytic	35
	SFi19	Sfi19	Yogurt	37,370	Sfi21 group	Sfi21 group	Lytic	41
	SFi21	Sfi21	Yogurt	40,739	Sfi21 group	Sfi21 group	Lysogenic	41
L. innocua	CLIP11262	Prophage 5		36,779	Unclassified	Sfi21 group	Lysogenic	29

[a] GenBank accession no.

Tab.2. Characterization of the bacteriophages discussed in Proux et al. In the circumscribed zone lactococcal phages (Proux et al 2002).

The phages sk1 and bIL170 of the 936 species share >80% nucleotide (nt) sequence identity, but differ by point mutations and numerous small insertions/delections. In particular phage sk1 has a 28.4 kb long DNA genome with cohesive ends, divided into two segment: a 16 kb one encoding the structural gene cluster followed by the lysis module. The early genes are encoded on opposite strand segments covering a major part of the right half of the genome. This head-to-head constellation of late and early genes is not found in any other dairy phage. Despite some protein similarities with phage sk1, phage c2 shows a distinct genome organization (Fig.6) and a shorter genome size (22 kb). The origin of replication, located 7 kb downstream of the left *cos*-site, divides the phage c2 genome in leftward-transcribed early genes covering DNA replication and recombination functions and a larger rightward-transcribed late gene cluster covering the lysis module and the structural genes. Phage r1t of the P335 species shows a 33.3 kb genome with cohesive ends. The structural genes were

localized next to the *cos*-site, followed by the lysis genes, the lysogeny and DNA replication modules. The lysogeny genes are the only genes located on the opposite strand. Two temperate *pac*-site phages from the P335 species, that have been completely sequenced (Tuc2009 and TP901-1), show similar structural gene clusters, but different from any other lactococcal phage, including *cos*-site phage from the P335-species (Brussow, 2001). This concise overview on lactococcal phage genome organization gives us an idea of the diversity and flexibility of lactococcal phage genomes and of the correlated variability in essential processes of phage life such as transcription, replication and DNA packaging.
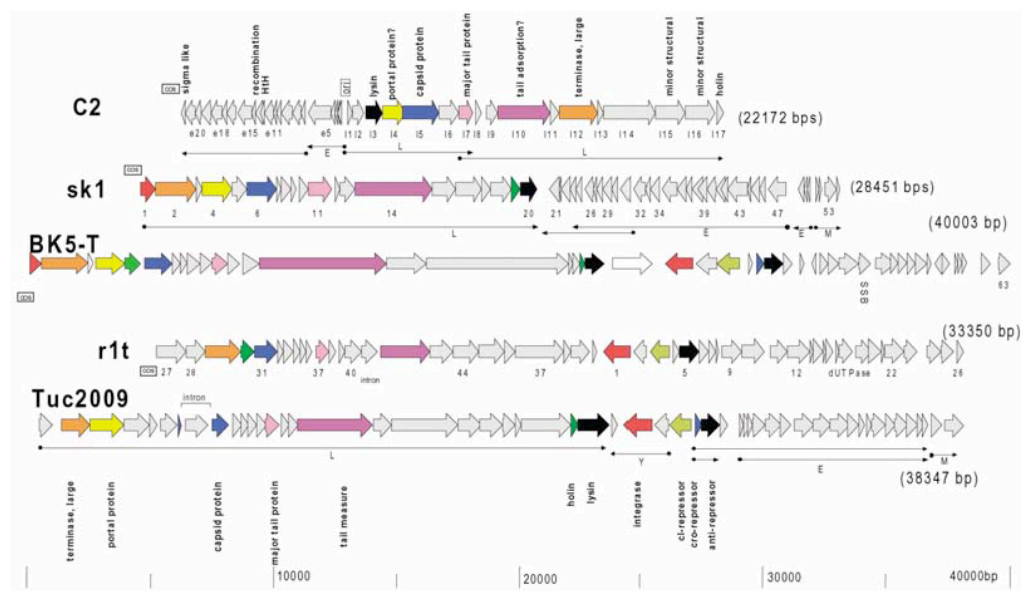
Fig. 6. Alignment of the genetic maps of four distinct L.lactis phage species represented by phages c2, sk1, BK5-T and r1t. Transcription maps are indicated with arrows. Corresponding genes are marked with the same colour (Brussow 2001, Desiere et al 2002).

In LAB phages, the genomic complexity previously described is complicated by horizontal and lateral gene transfer, a phenomenon that is particularly evident for the replication genes of lactococcal phages (Brussow and Desiere, 2001). Gene transfer occurs, not only between bacteriophages belonging to the same DNA homology group or interbreeding phage populations (Horizontal Gene Transfer) (Casjens et al, 1992), but also between the phage genomes and cryptic plasmids or host

chromosomes (Lateral Gene Transfer). New genomic combinations can be created by, at least, four different mechanisms: homologous, microhomologous, site-specific and illegitimate recombinations (Casjens et al, 1992). Homologous recombination, which is also used by temperate phage to alternate lytic growth and prophage status, is the most frequent mode of DNA exchange (Campbell, 1994). In particular, within the lactococcal phage P335 species, large DNA fragments can be exchanged, through homologous recombination, with (cryptic) prophages present within the host chromosome (Bouchard and Moineau, 2000, Durmaz and Klaenhammer, 2000). The acquisition of a methylase gene from plasmids has also been reported for P335 phages (Hill et al, 1991). A similar genetic exchange has also been in CRISPR sequences (Clustered regularly interspaced short palindromic repeats) which are DNA modules characterizing the genomes of most Bacteria and Archea, included lactic acid bacteria (Horvath et al 2008). In these CRISPR sequences, the removal or addition of particular spacers, derived from phage genomic sequences, modifie the phage-resistance phenotype of the bacterial cell and, in association with *cas* genes, provide resistance against phages. This resistance specificity is determined by spacer-phage sequence similarity (Barrangou et al 2007).

1.1.2 *Antiviral systems*

Bacteria have developed many different types of mechanisms against bacteriophages. These, in fact, are ubiquitous and frequently outnumber bacteria in the environment (Chopin et al, 2005). Due to the dynamic phage pressure encountered within mesophilic dairy fermentation, lactococci are an example of bacteria that have developed a variety of strategies to interfere with different stages of phage infection cycle, including prevention of phage absorption, inhibition of DNA injection, phage DNA replication and transcription, restriction of incoming phage DNA, interference with synthesis of phage proteins, phage particle assembly and bacterial lysis. Some *L.lactis* strains were found to naturally possess plasmids coding for anti-phage systems. More than 40 different lactococcal anti-phage plasmids have been reported in the literature (Daly et al 1996). The antiviral mechanisms, based on their mode of action, are divided into four groups: adsorption inhibition, blocking DNA injection,

restriction/modification (R/M) of phage DNA and abortive infection (Abi) (Allison and Klaenhammer 1998, Josephsen and Neve 1998). In particular the factors that prevent phage adsorption and penetration, that act by modifying cell-associated carbohydrate moieties and receptor proteins, are encoded by plasmids as in the case of phage c2 (Garvey et al 1996). Restriction/modification systems, on the other hand, utilize restriction endonuclease that digest specific recognition site in phage DNA (Allison and Klaenhammer 1998).

Abortive infection (Abi), a phage resistance mechanism that is activated after a normal start of infection (e.i. phage adsorbs and inject its DNA into the host cell), determines the interruption of phage development. In particular it leads to the release of few or no progeny particles and to the death of the infected cell. These mechanisms, that have been found in many bacterial species including *Escherichia coli* (Snyder 1995), *Lactococcus lactis* (Forde and Fitzgerald 1999), *Bacillus subtilis* (Hemphili et al 1975) *Streptococcus pyrogenes* (Behnke and Malke, 1978) and *Vibrio cholerae* (Chowdhury et al 1989), indicate that they are most probably widespread in bacteria. Most of the known Abis are plasmid encoded. In particular, in dairy bacteria, plasmids carry most of the genes required for optimal growth in milk and often can be readily transferred by conjugation between strains (Mckay LL. 1983). For this reason, most plasmid-carried *Abi* genes were identified in plasmid transfer experiments. Most of the known Abis have been found in *L.lactis*; the reason is the serious economic consequences of phage attack in dairy industry, where *Lactococcus* is largely used. Thereby, many efforts have been devoted to the study of lactococcal phages and related defence mechanisms, leading to the knowledge of 23 Abi systems that have been isolated, characterized and designed as AbiA through AbiZ (Chopin et al 2005). In Tab. 3 are listed all Abi systems found in *L. lactis* except for AbiZ, which speeds the lysis clock to cause premature lysis of phage-infected *L. lactis* (Durmaz and Klaenhammer 2007) and AbiV that prevents cleavage of the replicated phage DNA of 936-like phages (Haaber et al 2008).

Main characteristics of lactococcal Abis.

Name	Source	Sequence accession number	Groups of sensitive phages[a]	Effects on phage cycle
AbiA	pTR2030	U17233	936, c2, P335	No DNA replication of P335 phages, interference with a phage recombinase
	pCI829	AAA25159		Unknown
AbiB	?	M77708	936	Decay of phage transcripts
	pHP003	AF247159		Unknown
AbiC	pTN20	M95956	936, P335	Normal DNA replication of 936 phages
AbiD	pBF61	U10992	936, c2	Unknown
AbiD1	pIL105	L35176	936, c2	Induction by a phage protein, interference with a phage RuvC-like endonuclease
AbiE	pNP40	U36837	936	Normal DNA replication of 936 phages
AbiF	pNP40	U36837	936, c2	Delayed DNA replication of 936 phages
AbiG	pCI750	U60336	936, (c2), P335	Normal DNA replication of 936 and c2 phages
AbiH	chromosome ?	X97651	936, (c2)	Unknown
AbiI	pND852	U38973	936, (c2)	Unknown
AbiJ	pND859	U41294	(936)	Unknown
AbiK	pSRQ800	U35629	936, (c2), P335	No DNA replication of P335 phages, interference with a phage Erf or Rad52-like recombinase. Normal DNA replication of 936 phages.
AbiL	pND861	U94520	936, (c2)	Unknown
AbiN	prophage	Y11901	936, c2	Unknown
AbiO	pPF144	I61427	936, c2	Unknown
AbiP	pIL2614	U90222	936	Early arrest of DNA replication, absence of early transcripts switch-off
AbiQ	pSRQ900	AF001314	936, c2	Normal DNA replication of 936 and c2 phages
AbiR	pKR223	AF216814	c2	Lowered DNA replication of c2 phages
AbiT	pED1	AF483000	936, P335	Lowered DNA replication of 936 and P335 phages
AbiU	pND001	AF188839	(936), c2, (P335)	Delayed transcription of 936 and c2 phages

[a] Groups in parentheses are relatively less sensitive.

Tab.3. Main characteristics of lactococcal Abis. (Chopin et al 2005).

The Abi phenotype is most frequently conferred by a single gene, but the involvement of two genes has been proposed in a few cases (i.e. AbiE, G, L and T). Very high sequence homology is rarely observed between the genes of lactococcal Abis: for example, AbiD, AbiD1 and AbiF share 28 to 46% identity. Similarly AbiA and AbiK share 23% identity and AbiC and AbiP share 22% identity. The identity between different Abis corresponds to similar domains essential to the Abi activity. AbiA and AbiK share a leucine-rich repeat and a reverse transcriptase motif (Dinsmore et al 1998, Fortier et al 2005), while a transmembrane motif was found both in AbiC , AbiP and AbiT (Domingues et al 2004, Bouchard JD et al 2002,). In AbiP is also present an OB-fold motif, typical of single stranded RNA or DNA binding proteins (Domingues et al 2004). Some Abi systems, like AbiN and AbiP, have also been shown to be toxic for *L. lactis*, when cloned in a high copy number plasmid, under the control of a foreign promoter (Prévots et al 1998, Prévots and Ritzenthaler 1998). Another line of evidence suggests that all lactococcal Abi products might be toxic to the cell. Interestingly, the *abi* genes may be the result of an horizontal gene transfer from another species (O'Connor et al 1996). Moreover, Abi proteins strongly differ in their codon preference, which results in the use of rare

tRNAs species, despite the same amino acids composition of different lactococcal proteins. The use of these rare tRNA would limit Abi's synthesis making consistent the hypothesis of their toxicity (Chopin et al 2005).

A given Abi can be active against phages from one, two or three of the representative three groups shown in Tab.3. As DNA and protein homologies are extremely limited between phages of different groups, it has been proposed that the phage element involved in phage/Abi coupling has to be found among these conserved phage DNA sequences or proteins. This is the case of AbiD1, but recent findings on AbiK suggest that Abis could also interfere with non-homologous proteins having similar functions (see below) (Chopin et al 2005). More information has been gained from the studies on phages resistant to Abis. P335 phages have been shown to acquire resistance to Abi by either exchange of a DNA segment with a resident prophage or point mutations (Bouchard et al 2002, Bouchard and Moineau 2000 and 2004, Moineau et al 1994). By contrast, virulent phage of the 936 group, which has no homologous prophage on the host chromosome, evolved resistance to Abi by point mutations (Forde and Fitzgerald 1999, Dinsmore and Klaenhammer 1997, Bidnenko et al 1995, Bidnenko et al 2002) or, alternatively by recombination during co-infections (Domingues 2004). Despite several bacterial resistance strategies have been developed (Tab. 4) against phage infections, actual efforts have been developed to minimize the appearance of phage mutants by using DNA recombinant strategies based on phage genome as a potential source of novel resistance traits (Chopin et al 2005).

Strategies	Authors	Year	Patent numbers
L. lactis			
PER (phage origin of replication)	Hill & Klaenhammer	1996	EP0474464, US5538864
Phage defense rotation strategy	Klaenhammer et al.	1997	AU8356291, AU664537, CA2050533, EP0474463, US5593885
Phage-triggered cell suicide	Klaenhammer et al.	1998	US5792625
S. thermophilus			
Pediococcus Lac+	Broadbent et al.	1997	US5677166
Phage DNA fragment	Mollet et al.	1998	EP0748871, JP9000274, US5766904
L. lactis R/M system	Moineau et al.	1998	AU4497296, CA2208107, EP0805861, US5824523, WO9621017

Tab. 4. List of patented phage resistance strategies (Chopin et al 2005).

1.1.3 *Lactococcal genes involved in sensibility to AbiK and Sak proteins*

1.1.3.1 The Abortive Infection System AbiK of Lactococccus lactis.

In previous studies, Edmond et al showed that in *Lactococcus lactis* the plasmid pSRQ800 coded for the phage abortive infection system AbiK and conferred strong resistance against small isometric phages of the 936 and P335 species (Edmond et al 1997). A database query, based on amino acid composition, suggested that AbiK and AbiA might be in the same protein family. Moreover in AbiK$^+$-infected cells, as in AbiA$^+$-infected cells, replication of phage ul36 (P335 species) DNA and production of its major capsid protein were not observed (Edmond et al 1997).

Distinct observations on the lytic cycle of phages p2 and P008 (936 species) and of phage P335 (P335 species), resulting from the presence of AbiK, were also reported by Boucher and co-workers (Boucher et al 2000). This study showed that, unlike phage ul36, replication of phage p2 DNA was observed in the AbiK$^+$, but only immature forms (concatemeric and circular DNA) of phage p2 DNA were found, indicating that the presence of AbiK prevented phage DNA maturation (Boucher et al 2000). The antiphage activity was demonstrated to be dependent on the level of expression of a single *abiK* gene and consequently on the intracellular concentration of its protein product. At present it is not known wether the *abiK* gene is espressed constitutively or it must be activated by a phage component (Fortier et al 2005).

The AbiK protein has a predicted size of 599 amino acids with an estimated molecular weight of 71.4 KDa, a pI of 7.98 (Edmond et al 1997) and a reverse transcriptase (RT) motif in the N terminal half of the deduced sequence (Fortier et al 2005). By site directed mutagenesis, some key residues of the RT motif were modified thus causing the loss of the resistance phenotype (Fortier et al 2005). These results and the low sequence similarity with other members of the RT family, show that RT motif is essential for phage resistance and AbiK. On the basis of these considerations, a mechanism of action for AbiK was proposed: through its RT activity, AbiK might retro transcribe a cDNA strand from a currently unknown template RNA (viral mRNAs), that can be subsequently degraded by a cellular

RNase and an exonuclease, with the result of preventing translation and production of phage proteins (Fortier et al 2005).

1.1.3.2 Isolation and characterization of AbiK insensitive phage mutants

Recent studies have shown a remarkable genome plasticity of lytic lactococcal phages that can produce mutants resistant to one or two Abi systems; this is the case of both AbiK and AbiT resistance in phage ul36 (Labrie and Moineau 2007). Usually the selective pressure of Abi Systems generates two general classes of phage mutants. The class I mutants have been obtained only with P335-like phages as a result of homologous recombination with phage-related sequences present in the host chromosome and have been isolated during the study of AbiA, AbiC, AbiK and AbiT (Bouchard et al 2002, Bouchard and Moineau 2000, Durmaz and Klaenhammer 1997, Dinsmore and Klaenhammer 2000, Moineau et al 1994). The class II mutants belong to all three groups of phages and result from point mutations (Bouchard and Moineau 2004). To date, two AbiK-insensitive class I phage mutants (ul36.1 and ul36.2 of the species P335) have been characterized, and one of them is characterized by a reduced burst size and a new origin of replication (Bouchard and Moineau 2000). Ten AbiK-insensitive class II mutants (five derivates of phage ul36, three of phage p2 (species 936) and one of phage φ31 (species P335) were also isolated and subjected to microbiological characterization and genomic analysis (Bouchard and Moineau 2004). For class II phage mutants, AbiK- and AbiK+ *L.lactis* stains were infected at the same time, with wild-type and mutants of phage UL36 to determine the efficiency with which they form a center of infection (ECOI). These results showed that, as expected for wild type phage ul36, AbiK reduces drastically the percentage of cells releasing at least one infectious particle. In the case of class II mutants almost all infected AbiK+ cells release phages showing that mutant phages are not sensible to AbiK mechanism (Edmond et al 1997, Bouchard et al 2000). Similar results were obtained with phage p2, a small-isometric-headed phage of the 936-like group (Boucher et al, 2000).

1.1.3.3 Identification of sak in phage ul36

Phage ul36 was isolated in cheese whey from a failed fermentation during a study on phage ecology in Canada (Moineau et al 1992). This lactococcal bacteriophage (species P335) has a small isometric head of 55 nm in diameter and a long noncontractile tail of 150 nm; it is considered as a reference lytic phage of the P335 species on phage-host interactions, phage ecology and phage detection assays (Boucher and Moineau, 2001). The complete genomic sequence of the *Lactococcus lactis* virulent phage ul36 shows a division of genes into four functional distinct units: lysogeny, early, late and lysis. (Fig. 7).
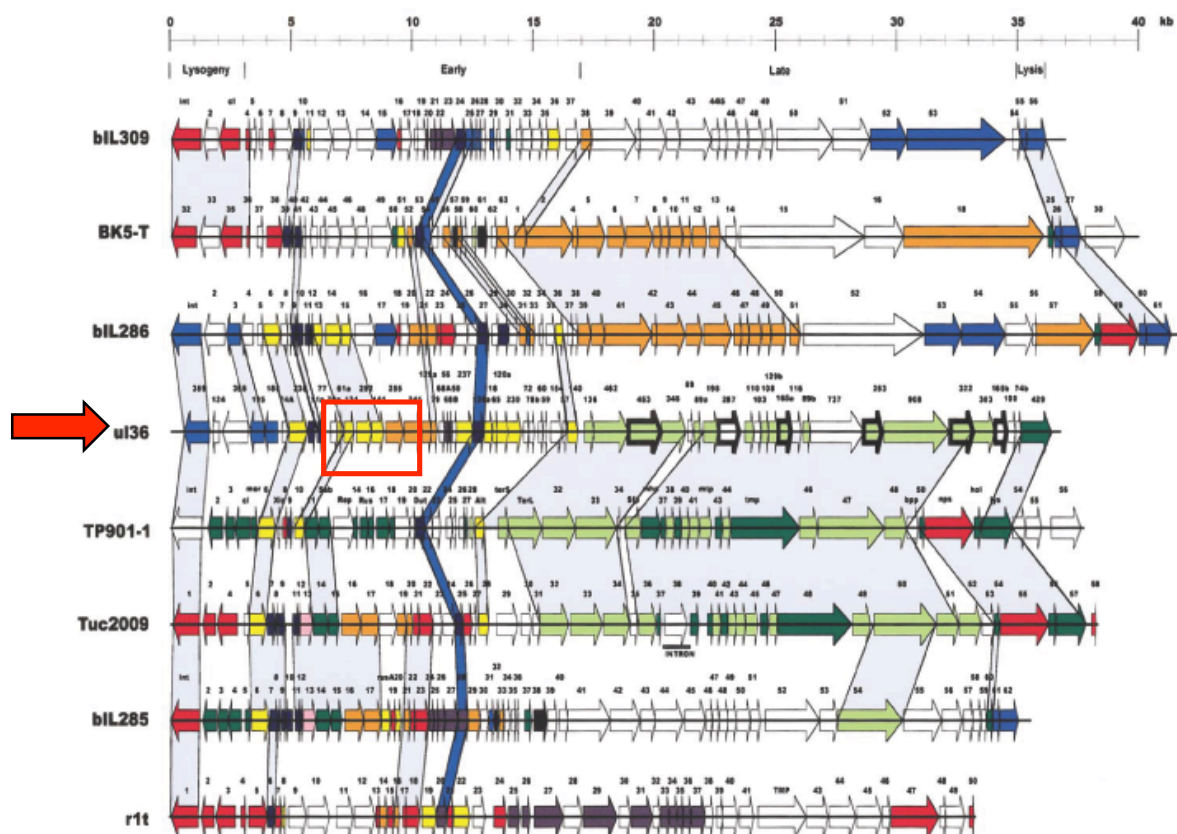
Fig. 7. Alignement of the genetic maps of all completely sequenced lactococcal P335-like and BK5T bacteriophages. The deduced proteins sharing more than 60% aa identity are represented using the same color and linked using gray shading when possible. ORFs with unique sequences are displayed in white. The dUTPase gene, present in all genomes, is linked by blue shading. The set of three genes (*orf*131, *orf*252 and *ssb*) exchanged by homologous recombination between phage ul36 and its host (Bouchard and Moineau, 2000) is highlight in red square. (Labrie and Moineau 2002).

Although lysogens of ul36 have not been found, a lysogeny module comprising six ORFs ascribed to integrase (orf359), repressor (orf188) and Cro repressor (orf74A) is present at one side of the genome. Transcribed genes, coding for morphogenesis and DNA packaging proteins (structural proteins and proteins involved in host recognition), followed by a lysis module (putative holin and lysine) are located at the other genome extremity. Early transcribed genes associated to DNA recombination are found in the middle of UL36 genome. Among these, putative functions were assigned only to five ORF: single-stranded binding protein (*orf*141), replisome organizer (*orf*255), RusA-like resolvase (*orf*129) and activator of the late trascription (*orf*140) (Labrie and Moineau, 2002). In this early-transcribed DNA regulation region, two DNA exchanges, involving phage ul36 and its *L.lactis* host by homologous recombination have been described (Bouchard and Moineau 2000). The two mutants generated by these recombination events have been labelled as class I mutants (section 1.1.3.1) and their genome analysis showed the replacement of a set of three genes (*orf*131, *orf*252 and *ssb*) (Fig.7 and Fig 8) that conferred insensibility to AbiK (Bouchard and Moineau 2000).
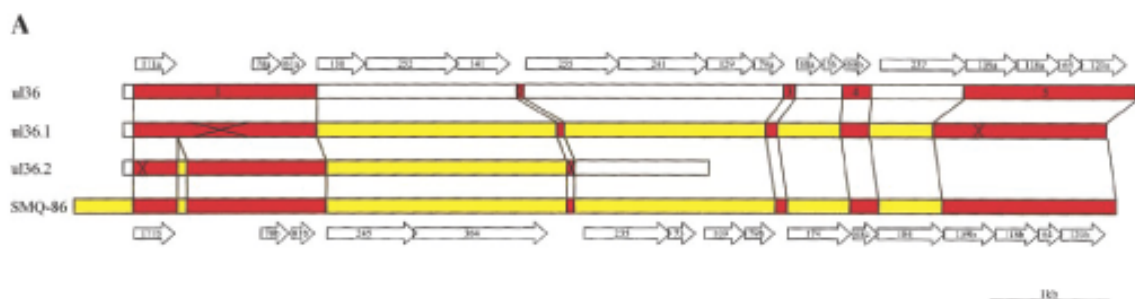
Fig.8. DNA homology between ul36, ul36.1, ul36.2 and the chromosome of *L.lactis* SMQ-86. In white, DNA of ul36; in yellow, chromosomal DNA; in red, regions of homology between ul36 and the chromosome. X, Location of the junction points are determined by the last divergent nucleotide between the phage mutant and ul36 or the chromosome (Bouchard and Moineau, 2000).

This same genomic region was sequenced in five class II ul36 AbiK-insensitives mutants. Sequence data showed that each mutant had one point mutation in *orf*252, renamed *sak* (mnemonic for sensibility to AbiK) (Bouchard and Moineau, 2004). Three point mutations causing amino acid substitutions in the deduced protein were

found in *orf*252 nucleotide sequence (V20I, A41V, A92S) (Table 5). Ploquin et al (Ploquin et al 2008) demonstrated that Sak protein of phage ul36 is a DNA Single-Strand Annealing Protein (SSAPs), as previously proposed by Iyer et al (Iyer et al 2002). Sak is also a functional homolog of the eukaryotic RAD52, a protein with sequence similarities involved in homologous recombination activities. In particular the first 209 aa of human RAD52 protein, that promote single-strand annealing *in vitro* (Singleton et al 2002), correspond to the first 166 aa in Sak ul36 as shown in Fig.9. Moreover, the point mutations, that in Sak confer sensibility to AbiK, are essentially located in nonconserved residues of the RAD52 DNA binding and ring formation domain that corresponds to Sak N terminal region (first 100 aa).
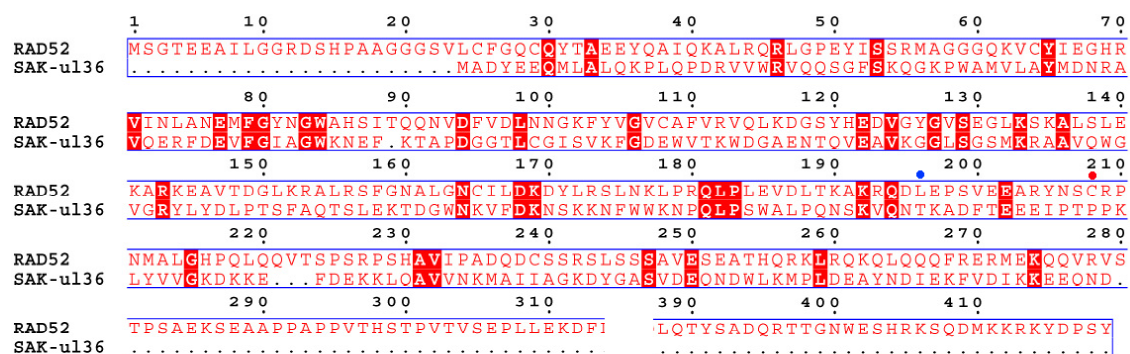
```
              1        10        20        30        40        50        60        70
RAD52    MSGTEEAILGGRDSHPAAGGGSVLCFGQCQYTAEEYQAIQKALRQRLGPEYISSRMAGGGQKVCYIEGHR
SAK-ul36 .......................MADYEEQMLALQKPLQPDRVVWRVQQSGFSKQGKPWAMVLAYMDNRA
                   80        90       100       110       120       130       140
RAD52    VINLANEMFGYNGWAHSITQQNVDFVDLNNGKFYVGVCAFVRVQLKDGSYHEDVGYGVSEGLKSKALSLE
SAK-ul36 VQERFDEVFGIAGWKNEF.KTAPDGGTLCGISVKFGDEWVTKWDGAENTQVEAVKGGLSGSMKRAAVQWG
                  150       160       170       180       190       200       210
RAD52    KARKEAVTDGLKRALRSFGNALGNCILDKDYLRSLNKLPRQLPLEVDLTKAKRQDLEPSVEEARYNSCRP
SAK-ul36 VGRYLYDLPTSFAQTSLEKTDGWNKVFDKNSKKNFWWKNPQLPSWALPQNSKVQNTKADFTEEEIPTPPK
                  220       230       240       250       260       270       280
RAD52    NMALGHPQLQQVTSPSRPSHAVIPADQDCSSRSLSSSAVESEATHQRKLRQKQLQQQFRERMEKQQVRVS
SAK-ul36 LYVVGKDKKE...FDEKKLQAVVNKMAIIAGKDYGASVDEQNDWLKMPLDEAYNDIEKFVDIKKEEQND.
                  290       300       310                 390       400       410
RAD52    TPSAEKSEAAPPAPPVTHSTPVTVSEPLLEKDFI  LQTYSADQRTTGNWESHRKSQDMKKRKYDPSY
SAK-ul36 .................................  ..................................
```

**Fig. 9**. Alignment of Sak ul36 with RAD52 (ClustalW, Higgings et al 1994).

Two other homologues of the N terminal region of Sak were also found in *Salmonella enterica serovar Typhimurium* and *Escherichia coli* phagic proteins. In particular the latest 56 amino acids are similar to the C terminal portion of the SSAP of Salmonella phage P22 named Erf (Tab. 5.) (Bouchard and Moineau, 2004).

A possible mechanism, proposed by Fortier et al, could explain Sak involvement in sensibility to AbiK. In infected Abi+ cells, the single-stranded cDNA released after degradation of the complementary template RNA could be annealed by Sak, which protects it from degradation by cellular exonucleases. Phages harbouring a mutation in *sak* gene would have an altered activity, resulting in failure of AbiK system (Fortier et al 2005).

| Gene	Phage (species)	Size (aa)	Mutation(s)	Homologues [phage/host (protein)]	% Similarity (aa)	GI number[a]
sak	ul36 (P335)	252	V20I	bIL286/*L. lactis* (Orf14)	99 (251/252)	13095757
			A41V	ST64T/*S. enterica* serovar Typhimurium (Orf235)	73 (111/152)	24371550
			A92S	1639/*E. coli* (L12)	75 (86/116)	17977996
				RM378/*Rhodothermus marinus* (Gp18)	54 (85/156)	30044008
sak2	P335 (P335)	202	M94I	H-19B/*E. coli* (Erf)	68 (102/150)	4153829
				HK022/*E. coli* (Erf)	68 (102/150)	9634148
				P22/*S. enterica* serovar Typhimurium (Erf)	63 (108/170)	9635505
				HK97/*E. coli* (Gp40)	67 (101/150)	9634188
				bIL67/*L. lactis* (Erf)	54 (90/165)	522281
				c2/*L. lactis* (Erf)	59 (83/139)	9628668
				MM1/*Streptococcus pneumoniae* (Erf)	49 (59/119)	15088753
sak3	p2 (936)	207	V16A	sk1/*L. lactis* (Orf35)	98 (201/205)	9629689
			M90I	bIL66MI/*L. lactis* (E12)	97 (201/207)	29893202
			S144Y	bIL170/*L. lactis* (E12)	95 (195/205)	9630613
				Tuc2009/*L. lactis* (Orf14)	75 (153/202)	13487813
				bIL285/*L. lactis* (Orf14)	75 (153/202)	13095694
				TP901-1/*L. lactis* (Orf11)	75 (153/202)	13786542
sak4	φ31 (P335)	245	A71S	TPW22/*L. lactis* (Orf11)	99 (239/240)	6465915
				2389/*Listeria monocytogenes* (Gp47)	63 (148/234)	17488552
				O1205/*Streptococcus thermophilus* (Orf9)	61 (148/241)	23455857
				Sfi21/*Streptococcus thermophilus* (Orf233)	60 (146/241)	9632970
				A2/*Lactobacillus casei* (Orf31)	58 (141/242)	22296553
				DT1/*Streptococcus thermophilus* (Orf32)	60 (146/241)	9632448
				Sfi19/*Streptococcus thermophilus* (Orf233)	60 (146/241)	9634922
				Sfi11/*Streptococcus thermophilus* (Gp233)	60 (145/241)	9634992
				315.2/*Streptococcus pyogenes* (0962)	60 (141/231)	21910498
				BK5-T/*L. lactis* (Orf234)	58 (140/240)	14251170
				phiadh/*Lactobacillus gasseri* (Gp14)	53 (123/228)	9633018
				Lp2/*Lactobacillus plantarum* (Gp17)	45 (97/213)	28379004
				phigle/*Lactobacillus plantarum* (He1)	44 (97/217)	23455783

[a] GI, GenInfo Identifier.

Tab. 5 Characteristics of the four phage genes involved in sensibility to AbiK (Bouchard, 2004).
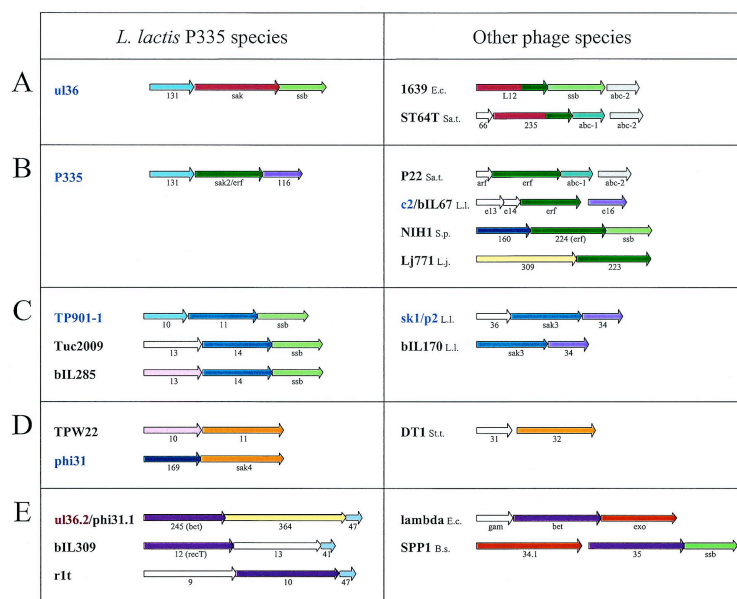
Fig. 10. Putative modules of recombination according to bioinformatics analyses, genome organization and classification of SSAPs into three superfamilies. (A) ORFs encoding homologues of Sak (RAD52 superfamily); (B) homologous of Sak2 or Erf; (C) homologous of Sak3; (D) homologues of Sak4; (E) members of the RecT/Beta superfamily. Neighboring genes represented by arrows of the same color have amino acid sequence similarity, except for ORFs represented in white, which have no significant similatity with other ORFs in figure. Phages sensitive to AbiK are in blue, and those that are insensitive to AbiK are in red. The sensitivity to AbiK was not determined for phage names typed in black. The hosts of the bacteriophages are *E.coli* (E.c.), *S. enterica Typhimurium* (Sa. T.), *L.lactis* (L.l), *Streptococcus pyogenes* (S.p.), *Lactobacillus johnsonii* (L.j), *Streptocccus thermophilus* (St.t.) and *Bacillus subtilis* (B.s.) (Bouchard and Moineau, 2004).

30

1.1.3.4 Identification of other sak genes in lactoccoccal phages

At least four non-similar proteins are involved in sensibility to AbiK (Table 5). Sak2 of phage P335 does not share significant amino acid sequence similarity with that fromul36, but is similar to P22 Erf SSAP and to a protein of c2-like phages, which are also sensible to AbiK (Fig 10 and Table 5) (Bouchard and Moineau, 2004). Other "*sak*-like" genes are not homologous to s*ak* or *sak2* genes, but possess similar *sak* flaking genes. These flaking genes were used to search for other potential *sak*-like genes located within the same genomic region. With this strategy, some genes, named *orf*11 in phage TP901-1 and *orf*35 in the virulent lactococcal phage sk1 and p2 of species 936 were found at the usual location of *sak*$_s$. Fig. 11 shows that phage p2 genome that was not only subjected to mutations in the gene *sak*, but also in the gene *sav* (*orf26*) which is involved in sensibility to AbiV (Haaber et al 2009).
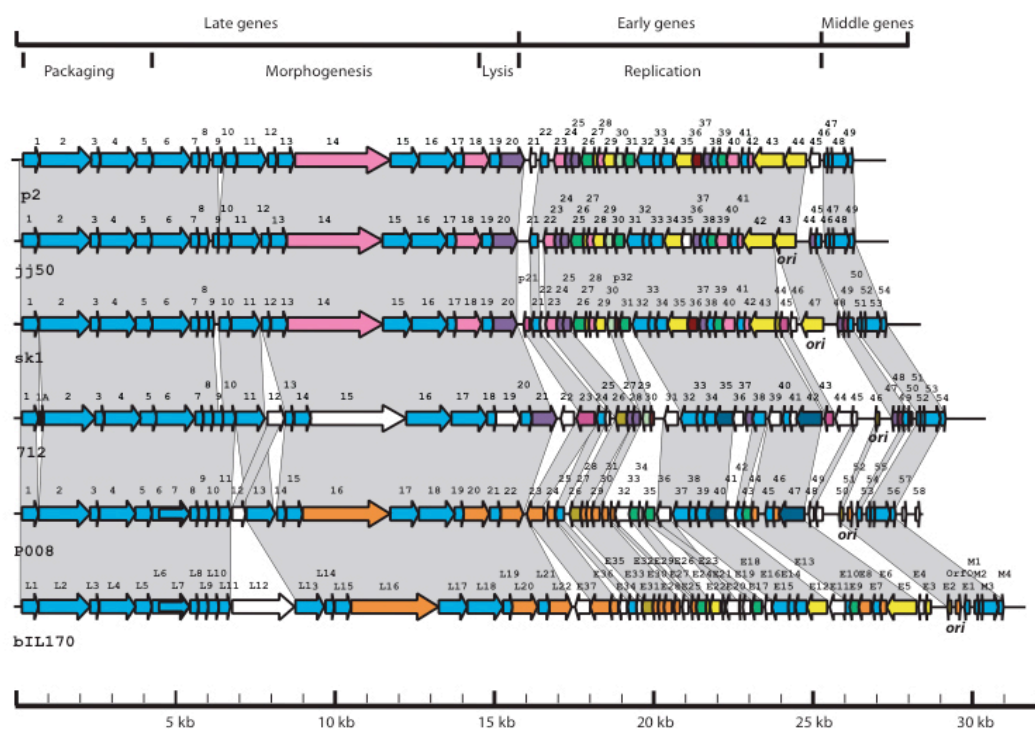
Fig. 11. Schematic representation of the genomic organization of the 936-type phages jj50, sk1, 712, P008 and bIL170. Genomic regions connected by a grey block are indicative of those possessing at least 60% identity at the amino acid level. ORFs of the same colour represent those with at least 80% identity at the amino acid level (modified from Mahony et al 2006)

By sequencing *orf*35 genomic regions of three AbiK-insensitive phage p2 derivates and comparing them to that of wild type phage p2, each of the AbiK-insensitive mutants shows one distinct point mutation in the *sak3* gene. In particular, *sak3* involvement in the sensibility to the phage defense mechanism has been confirmed by the plasmidic overexpression of wild type and mutated Sak3 located downstream to a strong constitutive lactococcal promoter. Finally, by using a similar strategy, *sak4* of phage φ31 was also identified (Fig. 10 and Tab. 5) (Bouchard and Moineau, 2004).

Microbiological studies on wild type and mutated alleles of Sak proteins showed clearly that mutated Sak, Sak2 and Sak3 proteins allow the phage, on one side to circumvent AbiK and on the other to mantein its recombinant activities (Bouchard and Moineau, 2004).

Assuming that *sak* genes encode proteins with similar functions or biochemical activities, the flanking genes of unknown functions are likely to play roles in homologous recombination or DNA repair (Poteete et al 1984 and 1993).

1.2 Homologous Recombination: proteins involved and models

1.2.1 Biochemical mechanisms

Homologous recombination (HR) is an important mechanism for the repair of both prokaryotic and eukaryotic damaged DNA, that prevents the demise of damaged replication forks, and several other aspects of DNA integrity. HR involves the pairing and exchange of genetic material between two homologous chromosomes or DNA molecules. Homologous recombination reactions promote repair of DNA ends formed by double-strand breaks (DSBs) and by replication fork collapse. Recombinational repair also allows cells to replicate past DNA lesions that block the progress of DNA polymerase. In phage T4, recombination is critical for initiating replication. In eukaryotes, homologous recombination is critical for preservation of replication forks, for telomere maintenance and for the accurate reductional segregation of chromosomes during meiosis (San Filippo et al 2008). In prokaryotes, recombination allows horizontal transfer of alleles between bacteria and phages (Gasior et al 2001).

DSBs can be repaired by several homologous recombination HR-mediated pathways, including double-strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA), which is represented in Fig.12.

In general biochemical mechanisms for homologous recombination on DSBs is similar in reference organisms such as *E.coli*, Bacteriophage T4, *S.cerevisiae*, and can be separated into four principal steps: initiation, homologous pairing and DNA strand exchange, DNA heteroduplex extension (branch migration) and resolution in the case of DSBR.
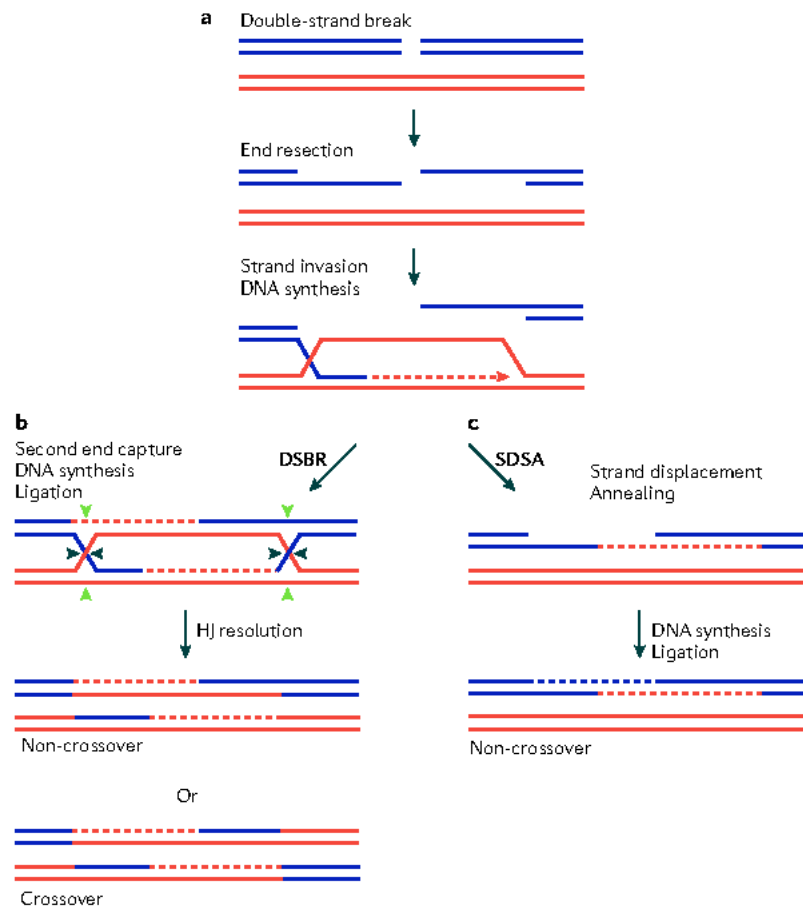
Fig. 12. Schematic representation of the model showing repair of DSBs by DSBR and SDSA (Sung and Klein 2006).

In both pathways, repair is initiated by resection of a DSB, generated by a programmed *in vivo* process or by environmental circumstances, and processed by a helicase, a nuclease, or both, to yield a free 3'-ssDNA overhang(s). The tails represent the substrate for nucleoprotein filament formation with a recombinase (RecA-like protein). Strand invasion by these 3'ss overhangs into a homologous sequence results in formation of the D-loop. The invading strand primes DNA by using the displaced strand as a template (Fig .12.a). After strand invasion and synthesis in DSBR, the second DSB end can be captured to form an intermediate with two Holliday junctions (Hjs). After gap-repair DNA synthesis and ligation, the structure is processed by a resolvase at the Hjs in a non-crossover (black arrow heads

at both Hjs) or crossover mode (green arrow heads at one HJ and black arrow heads at the other HJ) (Fig.12.b). Alternatively the reaction can proceed to SDSA by strand displacement, annealing of the extended single-strand end to the ssDNA on the other break end, followed by gap-filling DNA synthesis and ligation. The repair product from SDSA is always non-crossover (Fig.12.c) (Sung and Klein, 2006). If a DSB occurs between repeated sequences, it can also be repaired by the HR process of single-strand annealing (SSA), where the DSB ends are processed to form ssDNA tails that can anneal each other. SSA does not require the full repertoire of HR genes, as the repair process does not require strand invasion (Syminton 2002, Paques et al 1999).

At last, when a DSB has only one end, as might occur at a collapsed replication fork in which one branch of the fork has been lost, the single end can participate in a HR reaction that is known as break-induced replication (BIR). BIR can result in a non-reciprocal traslocation which joins part of one chromosome to a different chromosome and provides a mean for the elongation of shortened telomers (Syminton 2002, Paques et al 1999, McEachern et al 2006).

In Tab 6 are listed general functions and characteristics of protein involved in genetic recombination. Despite the fact that this table is not recent, it gives us a general idea of common functions and similar proteins in reference organisms, even if much more has to be discovered about functional domains and HR accessory proteins.

| GENERAL FUNCTION	PROTEIN	BIOCHEMICAL FUNCTIONS
E. coli		
Initiating protein(s)	RecBCD	ATP-dependent dsDNA and ssDNA exonuclease, ATP-stimulated ssDNA endonuclease, DNA helicase, recombination hotspot χ-recognition
DNA strand exchange	RecA	DNA-dependent ATPase, DNA- and ATP-dependent coprotease, DNA renaturation, DNA strand exchange
ssDNA-binding protein	SSB	ssDNA binding, stimulates DNA strand exchange
Accessory protein(s)	RecF	ssDNA, dsDNA binding, weak ATPase, interacts with RecR protein
	RecO	ssDNA, dsDNA binding; interacts with RecR; RecOR prevents end-dependent disassembly of RecA filaments; interacts with SSB
	RecR	Interacts with RecF; RecFR complex attenuates RecA filament extension into dsDNA regions
Branch migration	RecG	DNA helicase, branch migration of Holliday junctions
	RuvA	Binds to Holliday-, cruciform-, and four-way junctions, interacts with RuvB protein
	RuvB	DNA helicase, branch migration of Holliday junctions, interacts with RuvA protein
Holliday Junction cleavage	RuvC	Binds to four-way junctions, cleaves Holliday junctions
Other proteins	DNA gyrase	Type II topoisomerase
	DNA topoisomerase I	Type I topoisomerase, ω-protein
	DNA ligase	DNA ligase
	DNA polymerase I	DNA polymerase, 5'→3' exonuclease, 3'→5' exonuclease
Bacteriophage T4		
Initiating protein(s)	gp46	Interacts with gp47; endo- and exonuclease
	gp47	Interacts with gp46; endo- and exonuclease; stimulates gp46 action
	gp41	DNA-dependent NTPase, ssDNA binding, ATP- or GTP-dependent DNA helicase
	gp59	ssDNA binding, stimulates ATPase and helicase activities of gp41, interacts with gp32 and gp41
DNA strand exchange	UvsX	DNA-dependent ATPase, DNA renaturation, DNA strand exchange
ssDNA-binding protein	gp32	ssDNA binding, stimulates DNA strand exchange, interacts with UvsY and UvsX
Accessory protein(s)	UvsY	Stimulates DNA strand exchange, interacts with UvsX and gp32
Branch migration	Dda	DNA helicase, stimulates branch migration by UvsX protein
	UvsW	DNA helicase, branch migration of Holliday junctions, functional analogue of RecG
	gp41	DNA-dependent NTPase, ssDNA binding, ATP- or GTP-dependent DNA helicase
	gp59	ssDNA binding, stimulates ATPase and helicase activities of gp41, interacts with gp32 and gp41
Holliday Junction cleavage	gp49	Binds to and cleaves Y-junctions and Holliday junctions
S. cerevisiae		
Initiating protein(s)	Mre11	Forms complex with Rad50 and Xrs2 which is possibly responsible for resection of double-strand DNA breaks; with Rad50, ssDNA endo- and 3' to 5' dsDNA exonuclease
	Rad50	Forms complex with Mre11 and Xrs2 which is possibly responsible for resection of double-strand DNA breaks, ATP-dependent binding to dsDNA, contains ATP-binding Motif
	Xrs2	Forms complex with Mre11 and Rad50 which is possibly responsible for resection of double-strand DNA breaks
	Spo11	Binds DNA, likely catalytic subunit responsible for double-strand break formation
DNA strand exchange	Rad51	DNA-dependent ATPase, DNA strand exchange, interacts with Rad52, Rad54, and Rad55 proteins
ssDNA-binding protein	RPA	ssDNA binding, stimulates DNA strand exchange, interacts with Rad52 protein
Accessory protein(s)	Rad52	Stimulates DNA strand exchange, interacts with Rad51 and RPA proteins
	Rad54	Contains both ATP-binding, DNA helicase motifs, interacts with Rad51, hydrolyzes ATP, stimulates DNA strand exchange
	Rad55	Stimulates DNA strand exchange, contains Walker ATP-binding motif, interacts with Rad51 protein, forms stable heterodimer with Rad57 protein; shows homology to Rad51
	Rad57	Stimulates DNA strand exchange, contains Walker ATP-binding motif, forms stable heterodimer with Rad55 protein; shows homology to Rad51
Other proteins	Rad59	Shows homology to Rad52 protein, function is unknown

Tab.6. Functions of Proteins involved in Genetic Recombination (Bianco et al 1998).

1.2.2 *Proteins involved in Homologous Recombination Mechanism*

1.2.2.1 *RecA-like Recombinases*

Clark and Margulies first isolated mutants of *Escherichia coli* recA gene on the basis of the their inability to give recombination and of their sensitivity to DNA-damaging agents (Clark and Margulies, 1965). After this finding, the biochemical functions of RecA have been studied intensely in many laboratories over a period of two decades and we now know that RecA protein, which is a DNA-dependent ATPase and an ATP dependent DNA binding protein, is essential to the primary pathway of homologous genetic recombination, and is the first-known member of a list of proteins sharing the capacity to catalyse DNA strand exchange reactions.

The RecA family includes the UvsX protein of bacteriophage T4, the archeal RadA protein, and the eukaryotic Dmc1 and Rad51 proteins, which are both involved in eukaryotic homologous recombination (Bishop, 1994). Although considered structural and functional homologs, RecA, Rad51 and all the other mentioned proteins exhibit relevant structural and functional differences. Only the ATP binding core of these RecA-like proteins has been conserved (Fig 13). This same core has also been found as a common core fold (Story et al, 1992) within DNA helicases, F1-ATPase, ATP-binding cassette (ABC) membrane transporters and the DNA repair enzyme Rad50 (Ye et al 2004). The bacterial RecA protein presents a carboxyl-terminal domain, not found in other RecA-like recombinases, that is comparable to Rad51 N-terminal domain. Both these domains bind DNA and form lobes within the helical RecA filament. Despite this only example of convergent evolution, it has been reported that this C terminal domain of RecA binds double stranded DNA, while N terminal domain of hRad51 binds both ssDNA and dsDNA (Yu et al 2001).
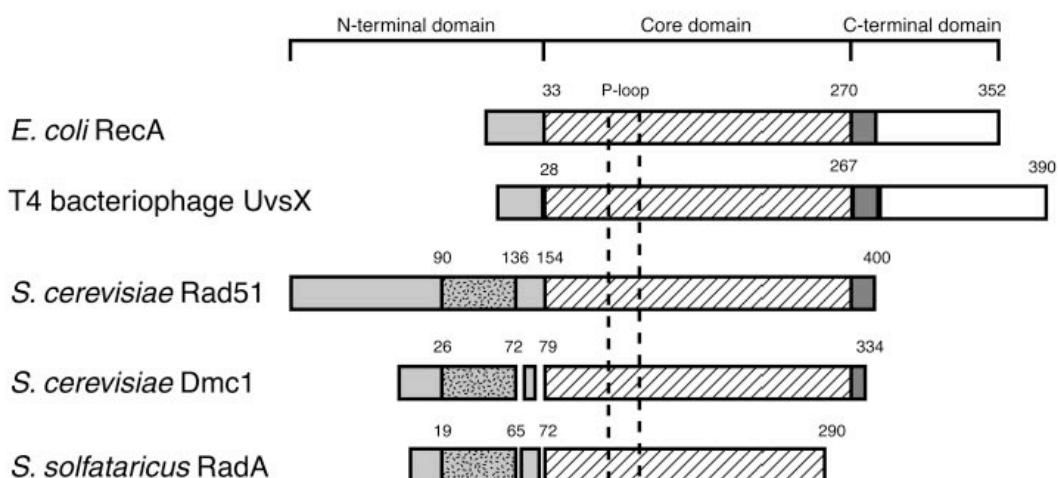
Fig. 13. Comparison of structural domains of RecA-like proteins. Core domain conservation is depicted with diagonal lines. N terminal domain conservation between Rad51, Dmc1 and RadA is shown as spotted region. Regions with no sequence homology include the light gray-shared region of the N-terminal domains and all regions C-terminal to the core domain (Lusetti et al 2002).

RecA-ATP first polymerizes on ssDNA to form a right-handed helical filament with one protomer for three nucleotides of DNA and about six protomers for helical turn. EM studies of RecA-DNA filament complexes reveal, in general, two conformational states (Egelman and Stasiak, 1993): a compressed conformation (inactive state), with an average helical pitch of about 73 Å in absence of DNA or with ADP, and an extended configuration, with an average pitch of about 95 Å (VanLoock et al 2003a). The T4 UvsX, the ScRad51 and the hRad51 protein induce in DNA the same unusual conformation induced by the RecA protein (Fig 14). In the case of hRad51, the unique difference is that the extended filaments, visible with RecA bound to either ssDNA and dsDNA, have been observed only on dsDNA, while those on ssDNA were relatively compressed (Benson et al 1994). The X-ray crystal structure of RecA (Story et al,1992), determined in the absence of ATP and DNA, revealed a helical 6_1-symmetric polymer with a pitch of 83 Å (Fig.15).
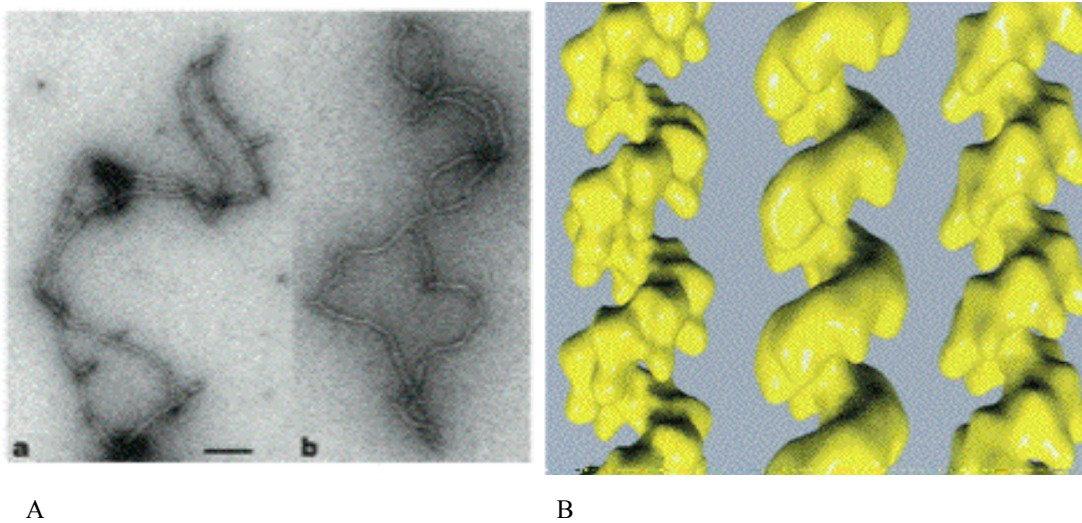
Fig. 14. A: Electron micrograph hRad51 after incubation with ssDNA in the presence of ADP AIF$^-_4$ (a) and after incubation with ssDNA and ATP-γ-S (b). B: From the left to the right: Surfaces of reconstructions of filaments formed by Rad51, ScRad51 and *E.coli* RecA on DNA (Yu et al 2001).
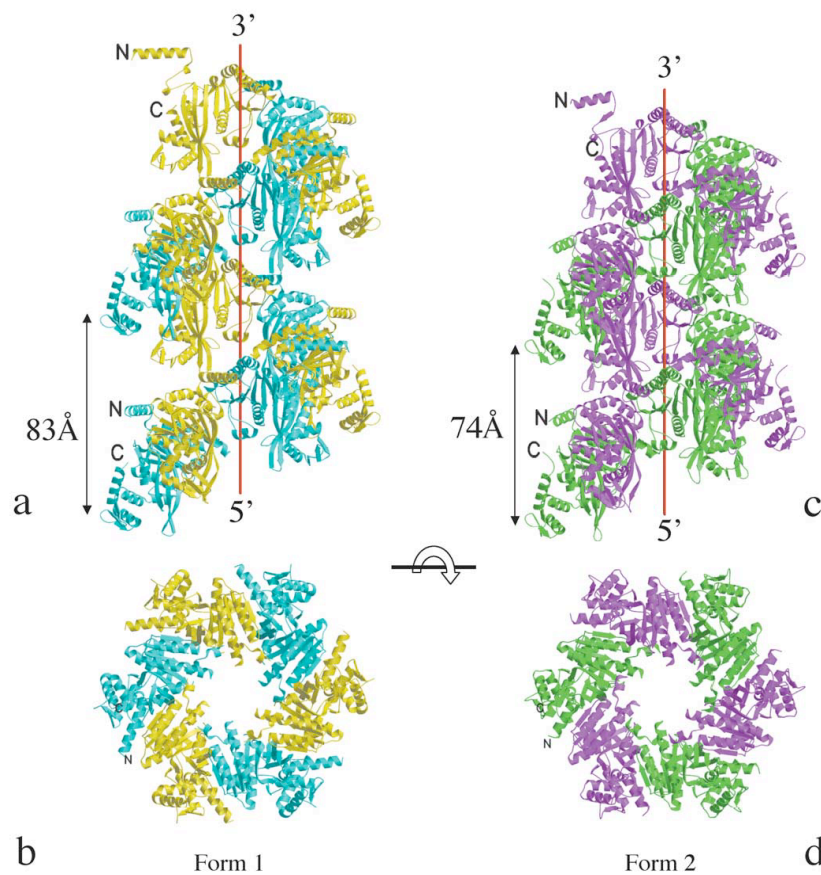
Fig. 15. Comparison of compressed and extended RecA filaments. **a**, The form 1 crystal filament (PDB code 2REB), with a helical pitch of 82.7 Å. **b**, View from the top of a, looking down the 6_1 screw axis. **c**, The form 2 crystal filament with a pitch of 73.5 Å. **d**, View of the top of c. (Xing and Bell, 2004).

The RecA monomer (Mr 38 kDa; 352 amino acids) folds into three domains. The N terminal domain (residues 6-33) forms an α-helix and a β-strand that pack against a neighboring subunit in the filament. The core domain, residues 34-269, folds into a central β-sheet and six α-helices that form the ATP-binding site, which faces the interior of the filament. Two disordered loops of the core L1 (residues 157-164) and L2 (residues 195-209), also point to the central axis of the filament and constitute the likely site for binding ssDNA (Malkov and Camerini-Otero, 1995). The C terminal domain (residues 270-328), which folds into a three stranded β-sheet and three α-helices, is located on the outer surface of the filament and has a site for dsDNA binding (Aihara et al 1997). Residues 329-352 form a disordered C-terminal tail with several negatively charged residues, whose flexible position is influenced by the crystal packaging interactions. Thus, we know relevant structural details on the compressed, inactive state of the RecA filament, but no high resolution structure of the extended, active state observed by EM in complex with ATP and DNA. More recently, Chen et al (Chen et al 2008) were able to construct *E. coli* RecA-DNA complexes that represent finite segments of the filament. Using this approach they could determine a 2.8 Å crystal structure of the active presynaptic RecA-ssDNA filament complex, a 3.15 Å structure of postsinaptic filament containing the new heteroduplex DNA lacking the displaced strand, and a 4.3 Å structure of the inactive filament. Within their work, they show that ssDNA and ATP bind to RecA-RecA interfaces cooperatively, explaining the ATP dependence of DNA binding. The ATP γ-phosphate is placed across the RecA-RecA interface by two lysine residues that also stimulate ATP hydrolysis, providing a mechanism for DNA release (Chen et al 2008).

Here I summarize some structural features of RecA, as a reference recombinase. RecA is quite similar to Rad51 as it has an hexameric shape and forms similar nucleoprotein filaments (Yu et al 2001), but shows relevant differences from others recombinase like RadA and Dmc1. In particular, RadA can bind DNA in the absence of cofactor as an octameric ring and in the presence of ATP as a helical filament (Yang S et al 2001, McIlwraith et al 2001). On the other hand, Dmc1 forms a nucleoprotein complex composed of stacked octameric rings on DNA, and no helical nucleoprotein filament has been detected (Masson J Y et al 1999, Passy S I et

al 1999). As described in the last sub-chapter, some differences are also found in the pattern of interactions with other proteins involved in HR.

1.2.2.2 *Proteins involved in binding with RecA-like proteins*

In general, recombinases alone are able to promote strand exchange to form hybrid DNA *in vitro,* without the help of additional proteins, but accessory factors can stimulate strand exchange. These factors can be divided into two classes: those that act before homology search, by promoting assembly of recombinase filaments (assembly factors), and those that act during homology search and strand exchange. Among the assembly factors we can distinguish ssDNA-binding protein (SSB) and assembly "mediators" (Gasior et al 2001).

1.2.2.2.1 SSB as an Assembly Factor

Single strand binding proteins (SSB) coat and protect single strand DNA (ssDNA) intermediates and facilitate pairing of homologous strands by eliminating DNA secondary structures. For this reason they are involved in multiple pathways of DNA metabolism, including replication, recombination and repair. SSBs are abundant in cells and have been fully characterized in reference organisms as T4 gp32, *E.coli* SSB and eukaryotic Replicative Protein A (RPA). SSBs bind ssDNA preferentially, in unspecific sequences and in a cooperative manner, forming filaments that can readily saturate long stretches of ssDNA (Lohman and Ferrari, 1994). Most SSBs exist as multimeric proteins and, in spite of their different shape and oligomerization, have a common mechanism as they all interact with many similar additional proteins; this interaction both regulates and is regulated by an interaction with ssDNA (Wold et al 1997). Two significant examples are SSBs from *E.coli* and human RPA (eukaryotic SSB) are described below.

The ssDNA-binding protein from *E.coli* is an homotetramer (Fig. 16 a) which can bind long ssDNA in two major binding modes, referred to as $(SSB)_{35}$ and $(SSB)_{65}$, where the subscripts reflect the average number of ssDNA nucleotides occluded by the SSB tetramer. In the $(SSB)_{35}$ mode, an average of two subunits of the stable

tetramer interact with ssDNA, whereas in the (SSB)$_{65}$ mode, the ssDNA wraps around the tetramer interacting with all the four subunits (Lohman T M and Overman L.B.1985). In addition to the interactions with ssDNA, a second essential aspect of SSB function is the interaction with proteins involved in DNA methabolism such as exonuclease I or RecO. An important consensus that has emerged is that SSB uses its C terminal domain to interact with these partners. In contrast with the well-ordered core domain, the C terminal domain is disordered even when bound to DNA (Savvides et al, 2004).

The human replicative protein A (RPA), like other eukaryotic SSBs, is a heterotrimer with three subunits of 70, 32 and 14 KDa. A comparison of free and bound forms of RPA revealed that ssDNA binding is associated with major reorientation between the structural modules-OB-folds, including the DNA binding domain. Two OB folds, stabilized by the presence of DNA in a tandem orientation, adopt multiple conformations in its absence. Within the OB folds, extended loops implicated in DNA binding significantly changed conformation in the absence of DNA. The analysis of intermolecular contacts suggested that other RPA molecules and/or other proteins could compete with DNA for the same binding site. By this competition mechanism, protein-protein interactions can regulate and/or be regulated by DNA binding (Bochkareva et al 2001).

A   B
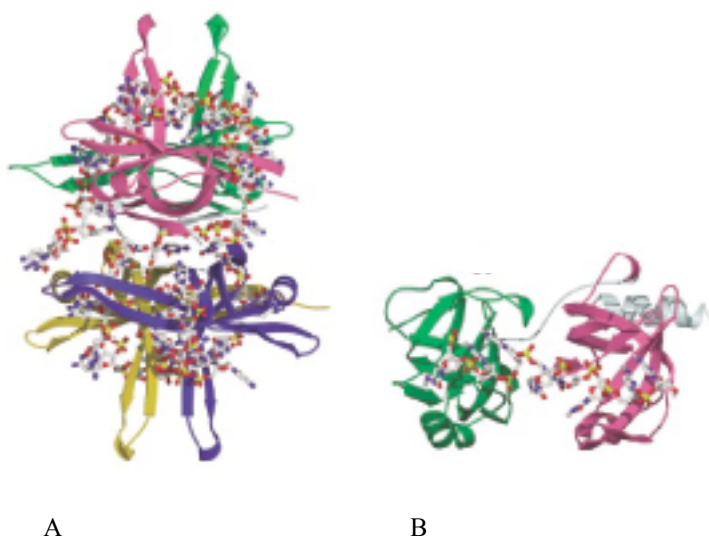
Fig. 16. Structures of OB-fold/ nucleic acid complexes. Ecoli SSB (A) and hRPA (B) (Theobald et al 2003).

For their capacity to interact with both DNA and proteins, SSBs play also an essential role in strand exchange reactions of homologous recombination. It has been showed that the recombinase filaments, formed as a result of SSB-recombinase interactions, are more efficient in strand exchange reactions than filaments formed by recombinase alone because of the persisting secondary structure in ssDNA. Since SSB can outcompete recombinases as a consequence of their higher affinity and faster binding kinetics, the stimulation of strand exchange reaction in presence of SSB can be explained as follow: upon addition of a saturating amount of SSB before the addition of recombinase, the reaction is inhibited, inversely if SSB is added after the recombinase, the latter initiates filaments on ssDNA and then extends sites of secondary structure with the help of SSB. The initial binding of RecA on SSB-coated ssDNA is mediated by "mediator " proteins that cause a local remodelling of the SSB filament (Gasior et al 2001). The process is schematised in Fig. 16.
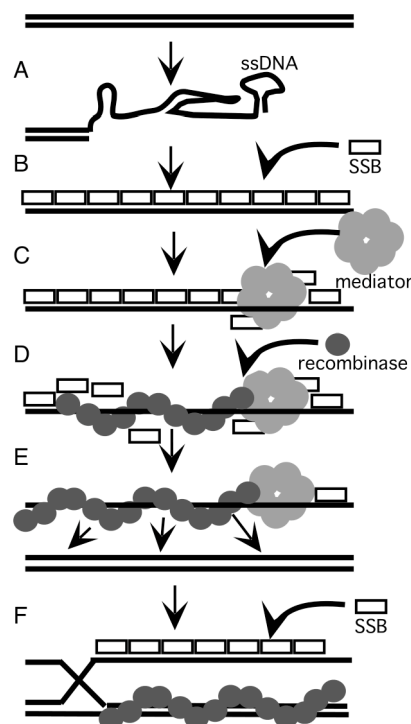
Fig. 17. Generic model for assembly of recombinase on ssb-coated ssDNA. (A) Tract of ssDNA form because of resection at DSB sites or stalling of polymerase. (B) SSB assembles into oligomeric filamento on tracts of ssDNA,. (C) Mediator protein binds to ss-coated DNA. (D) Recombinase initiates filament formation at sites of mediator-SSB-ssDNA (E) The elongated recombinase filament searches for homologous sequence (F) Strand exchange occurs. (Gasior et al 2001).

A more accurate description of SSBs structural and functional features is presented in the introduction of the article "*Crystal Structure and Characterization of a New Type of Single Stranded DNA Binding Protein from the Lactococcal phage p2*" (Section 2.2.1).

1.2.2.2.2 Recombination mediator Proteins

As already mentioned, recombination/replication mediators proteins (RMPs) appear to play an important role in overcoming the inhibitory effect of SSB on recombinase assembly. Studies on RMP activities in bacteriophage T4, *S.cerevisiae* and *E.coli* recombination systems indicate a common biological function for these mediator proteins: a key and common property of UvsY, Rad52 and RecO/R is their ability to promote efficient formation of recombinase oligomers on ssDNA that are coated with SSB. The bacteriophage T4 Uvs Y was the first protein to be shown to have mediator activity *in vitro*. In addition to overcoming inhibitory effects of gp32 on UvsX, UvsY can stabilize UvsX filaments when formed (Harris and Griffith 1989, Kodadek T et al 1989) (see section "Homologous Recombination in Bacteriophages").

E.coli appears to have two distinct mechanisms for promoting the assembly of RecA. RecO and RecR act together as mediator to overcome SSB-dependent inhibition of RecA-mediated strand exchange, while RecF is capable of limiting the assembly of RecA to regions of ssDNA *in vitro*. The second RecA assembly mechanism involves the multifunctional recombination protein RecBCD, which has both exonuclease and helicase activity. In particular, RecB subunit of the RecBCD harbors both helicase and endonuclease activities and the RecD subunit also possesses a helicase activity, while RecC recognizes a specific sequence in DNA called χ and has a scaffolding function in protein complex assembly. In general RecBCD promotes assembly of RecA at DNA ends, whereas RecFOR controls assembly of RecA at single-strand gaps, but this functional specialization is not strict (Gasior et al 2001).

Both the *S.cerevisiae* Rad52 and heterodimeric Rad55/Rad57 have been shown to have mediator function *in vitro* (Breuit et al 2001) and overcome inhibition of

Rad51-mediated strand exchange by RPA (Mcllwraith et al 2000). Although Rad52 stimulates assembly of Rad51, other observations indicate that it has additional activities that can promote recombination. Rad52 has been shown to promote annealing of complementary ssDNA (Mortensen et al 1996, Shinohara et al 1998) and to introduce positive supercoils into double-stranded DNA by a second DNA binding site (Kagawa et al, 2008).

Human Rad52 has been visualized by electron microscopy and a low-resolution three-dimensional structure has been determinated showing a ring-shaped heptameric structure, but initial efforts to crystallize the full-length protein were unsuccessful (Singleton et al, 2002). Only the N terminal truncated protein was crystallized and shows a structure in undecameric rings. The monomer subunit has two domains as shown in Fig. 18. The N terminal domain forms the core of the structure and has a mixed a/b-structure, whereas the smaller C terminal domain II consists of a flexible extended linker that ends in an helix. Domain II extends from the surface to interact with domain I in an adjacent subunit. The subunits are assembled as an undecamer in an 11-fold symmetric ring with a large channel running through the centre. The N terminal domain catalyses the homologous pairing; it has the ability to self associate and bind ssDNA, which probably lies in an exposed groove on the surface of protein ring. The interaction domain of Rad51 is thought to be located in the C terminal region of Rad52 (Singleton et al 2002).
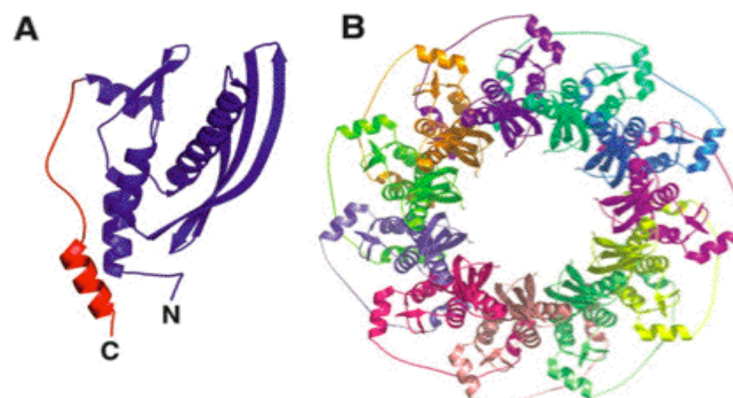
Fig.18 Structure of Rad52 N terminal domain (1-209). A: monomer and B: undecamer (Singleton et al 2002).

Rad52 has been classified as the archetype of a new class of DNA single-strand annealing proteins (SSAPs) such as RecT, RedB, ERF, that have been shown to form similar helical quaternary superstructures (Iyer et al 2002).

A more accurate description of Rad52 structural and functional features is present in the introduction of the article *"Sak protein from Lactococcal phage ul36: structure function study of protein and domains."* and *"Lactococcal Phage p2 ORF35-Sak3 is Involved In a Novel Type of Homologous Recombination"* (Section 2.1.1.1 and 2.1.1.2).

1.2.2.2.3 Regulation of HR by DNA helicases

An important function in HR is also played by DNA helicase. Helicase attaches to ssDNA at the lagging strand of a DNA replication fork and traslocates directionally through double-stranded nucleic acid (dsNA) substrates to catalyse the separation of the complementary NA strands. Furthermore, in conjunction with other components of the HR macromolecular machines, the helicases also catalyze the transfer of single–stranded nucleic acid (ssNA) products of the dsNA opening reaction to other proteins or to different complementary ssNA partners, or release the ssNA products directly in solution. These operations are driven by the consumption of energy derived from the hydrolysis of nucleotide triphosphate (ATPase activity) (Von Hippel and Delagoutte, 2001). In Tab.7, *E.coli* helicases characteristics are reported as an example of their involvement in various biological processes such as DNA replication, recombination, repair and transcription.

Helicase	Function	Specificity	Directionality	Association State	"Coupling" Factors[b]
DnaB	Replication	DNA-DNA	$5' \rightarrow 3'$	Hexamer (6 Identical subunits)[a]	DNA polymerase III ssDNA binding protein (primase)
RNA polymerase (core)	Transcription (elongation)	DNA-DNA	$5' \rightarrow 3'$	Tetramer (4 different subunits)	Next required NTP (see Section 5)
Rho	Transcription (termination)	RNA-DNA	$5' \rightarrow 3'$	Hexamer (6 identical subunits)[a]	(NusG)
RecBCD	Recombination	DNA-DNA	$3' \rightarrow 5'$	Trimer (3 different subunits)	(RecA) (ssDNA binding protein)
UvrD	DNA Repair	DNA-DNA DNA-RNA	$3' \rightarrow 5'$	Not fully defined	DNA polymerase I ssDNA binding protein

[a] Subunits identical in sequence, though not necessarily in conformation.
[b] Here we include both factors that alter the kinetics of the helicase reaction by direct "trapping" mechanisms and (in parentheses) factors that facilitate helicase function (e.g., loading) in other ways (see text).

Tab.7. *E.coli* Helicases involved in various biological processes (Von Hippel and Delagoutte 2001)

Bacteriophages T4 gene 4 proteins provide both the helicase and primase functions for T7 DNA replication and both proteins have been shown to form stable hexamers (Fig. 19) (Egelman et al 1995). Hexameric helicases such as bacteriophage T7 gp4 (Egelman et al 1995), SV40 large T antigene (San Martin et al 1997) and DnaB (Yang et al 2002) contain the conserved RecA nucleotide-binding core and are ubiquitous in DNA recombination, replication and transcription (Stasiak et al 2000). Normally the six subunits of hexameric helicases are arranged in a cyclic symmetry around a central channel and show a highly modular architecture. In various organisms, helicases require "loaders", hexameric proteins that deliver the helicase to its site of action on the DNA template. For example in *E.coli* the "loader" DnaC position the replicative helicase DnaB in preparation for subsequent polymerase assembly, while gp59 loads gp41 helicase in bacteriophage T4. Normally, loaders require specific contacts with both ssDNA and the hexameric helicase and share the same helicase shape for promoting the loading (San Martin et al 1998).
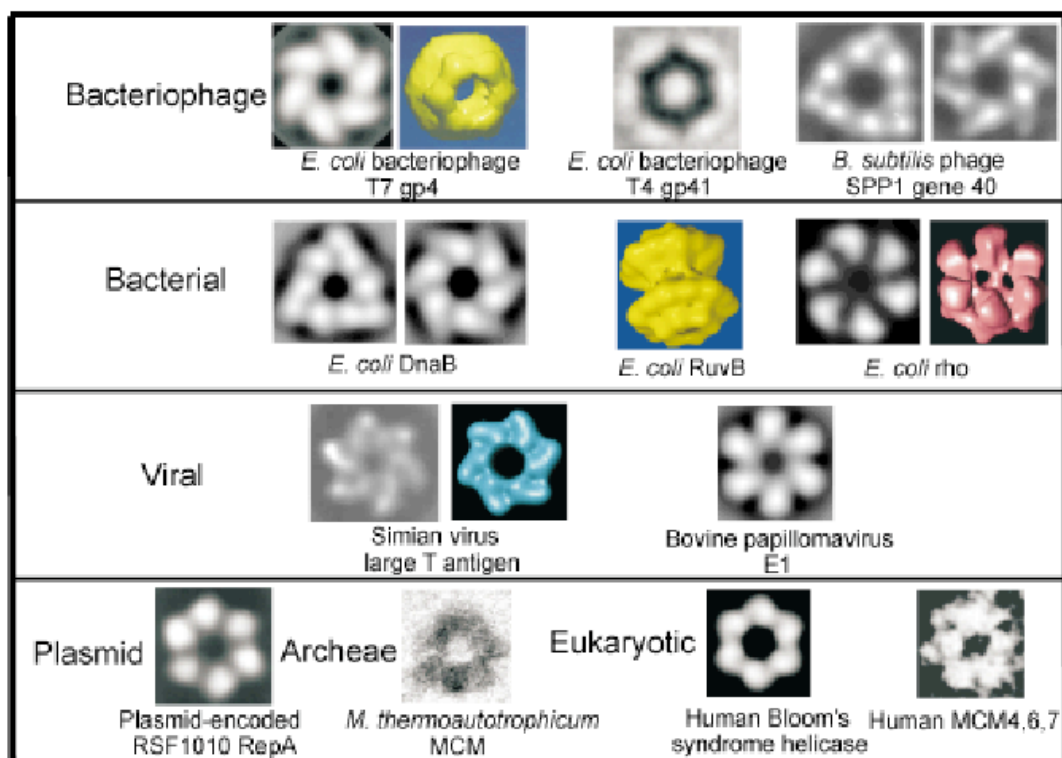

Fig. 19. Electron microscopy images of hexameric helicases (Pate and Picha, 2000).

Interestingly, helicase activity can be associated to topoisomerase activity as in the case of reverse gyrase. This enzyme induces positive supercoiling and comprises a topoisomerase IA fused to a helicase-like domain (Perugino et al 2009). This helicase domain contains motifs I, Ia, II, III, V, VI and Q found in helicases of the superfemilies 1 and 2. Motifs I and II (so called Walker motifs A and B) are mainly devoted to ATP binding and hydrolysis, whereas motifs V and VI in different enzyme can either be involved in nucleic acid binding, NTP hydrolysis or both (Tuteja et al 2004).

Although HR is a major DNA-repair apparatus and helps to prevent replication fork demise, inappropriate or untimely HR events can have deleterious consequences.
It is not unreasonable to suggest that a DNA helicase can both promote homologous recombination and prevent it from occurring, depending on the context and the substrate. Supporting this dual role is the finding that RecQ helicase of *E.coli* helps to initiate homologous recombination intermediates and also disrupts recombinant intermediates (Harmon and Kowalczykowski, 1998). Recombination substrates, in the form of broken DNA molecule formed during replication, can be repaired by re-establishing of the replication fork or by rearrangement of the molecule. The RecQ helicase might act as an antirecombinase to prevent the formation of the rearranged molecule. Genetic analyses in *S.cerevisiae* have also implicated the Srs2 and Sgs1 helicases in the prevention of undesirable HR events (Klein 2000). Srs2 protein has a 3'$\rightarrow$5' DNA-helicase activity and is capable to disrupt the Rad51 presynaptic filament (Fig. 20). Biochemical studies have shown that the Srs2 helicase, at the expense of ATP hydrolysis, dismantles the Rad51 presynaptic filament in the 3'$\rightarrow$5' direction. The single-stranded DNA that becomes available as a result of Rad51 eviction is immediately occupied by the RPA to prevent the re-loading of Rad51 (Fig. 20a). In Fig. 20b, the two Holliday junction structures, that arise from the dissolution of an HR intermediate, are pushed inward by a DNA helicase to form a hemicatenane, which is then dissolved by a topoisomerase to give non-crossover recombinants (Sung and Klein, 2006).
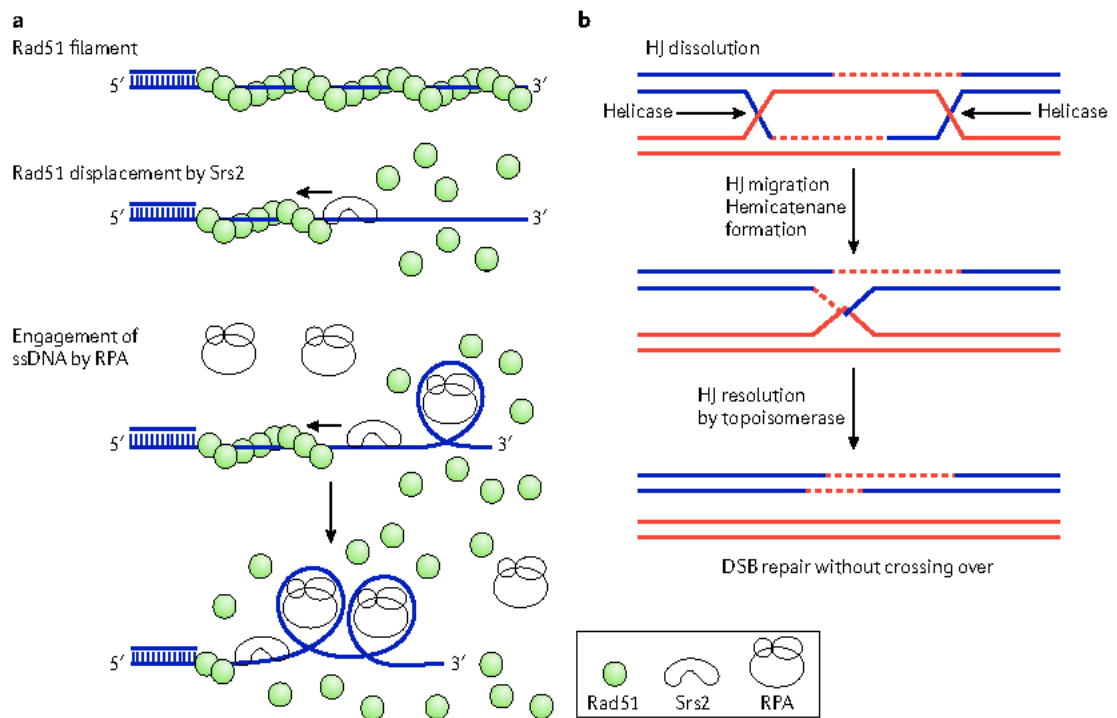
Fig. 20. Regulation of HR by Srs2 helicases. A: Disruption of the Rad51 presynaptic filament. B: Dissolution of an HR intermediate that harbours two Holliday junctions (Sung and Klein, 2006).

1.2.3 Homologous Recombination in Bacteriophages

1.2.3.1 *Generality and systems-model*

In contrast to the enzymatically complex and cumbersome mechanisms of double-strand end repair in the *E.coli* chromosome, some bacteriophages connect homologous double-strand ends by a simple and effective trick. However, while saving on enzymes, phages are wasteful with their DNA. This strategy would be unacceptable for *E.coli* chromosome but is quite affordable for phage genomes, which are only 1 to 2 % of the length of the *E.coli* genome and, by the end of the infection, are present in cells in multiple copies. In particular it was shown that Double Strand Repair (DSR) repair, DNA replication and recombination are tightly linked throughout much of the phage life cycle and also that the major mode of phage DNA replication depends on recombination proteins and can be stimulated by DSBs. In this manner phages need only a little set of ubiquitous recombinant proteins and in some cases they can dig in the host repertoire of recombination proteins in order to increase their possibilities.

Here, I present three well-studied examples of repair/replication/recombination in T4, SPP1 and λ bacteriophage taken as reference organism for Recombination-Dependent-Replication (RDR), and Theta-sigma type DNA Replication and Recombination associated with SSA repair, respectively. In Tab. 8 some information about their genomes and DNA replication strategies are given.

Phage	Incoming Virion DNA	Replication Strategy	End State of Replicated DNA
λ	linear with 5' 12-base complementary ends (cohesive ends)	closed circle switching to rolling circle	linear concatamer
P22	linear with 104% terminal redundancy	recombination with extension via direct repeats	linear concatamer
ø29	linear with covalently attached gp3 at 5' ends	gp3-primed extension, strand displacement	unit length with gp3 covalently attached at 5' ends
T3 (T7)	linear with 230 (160) base pair direct repeats	recombination with extension via direct repeats	linear concatamer
T4	linear with 102% terminal redundancy	invasive strand initiated via terminal redundancy	branched concatamer
SPP1	linear with 104% terminal redundancy	unknown	linear concatamer

Tab. 8. Viral DNA and DNA replication strategies of various dsDNA phages (Chapter 6, The Bacteriophages).

1.2.3.1.1   T4 Bacteriophage

The relationship between Rad51, Rad52 and RPA in eukaryotes and between RecA, RecOR complex and SSB in *E.coli* is reminiscent of that between UvsX, UvsY and gp32 proteins in T4 phage. Bacteriophage T4 provides an excellent model system for biochemical and genetic studies of Recombination-Dependent Replication (RDR). In particular, two possible mechanisms are known: a semiconservative replication mechanism that is the major pathway of T4 RDR (Mosig et al, 1983) and "bubble-migration synthesis" mechanism that is one of the models for DBS repair (Formosa and Alberts, 1986), but the results argue that RDR and DSB repair are simply two different ways of viewing the same reaction: the assembly of semiconservative replication forks at D-loops created from double-strand ends (George et al, 2001). Here I report a general model of RDR, only based on "bubble-migration" mechanism for simplicity.

After infecting *E.coli* cell, T4 first replicates its genome *via* an origin-dependent replication initiation pathway in which branched recombination intermediates generated by the phage homologous recombination machinery are captured and converted into semiconservative DNA replication forks. T4 RDR requires all the major phage-encoded DNA replication and recombination enzymes including: gp43 (DNA polymerase), gp45 (sliding clamp), gp44/62 (clamp loader), gp61 (primase), gp41 (DNA helicase), gp59 (helicase loader; replication mediator protein or RMP), UvsX (general recombinase), UvsY (recombination mediator protein), and gp 46/47 (recombination exonuclease). Figure 21A shows the enzymatic steps in the T4 RDR in vitro reaction and Figure 21B focuses on UvsY-mediated assembly.

A   **B**
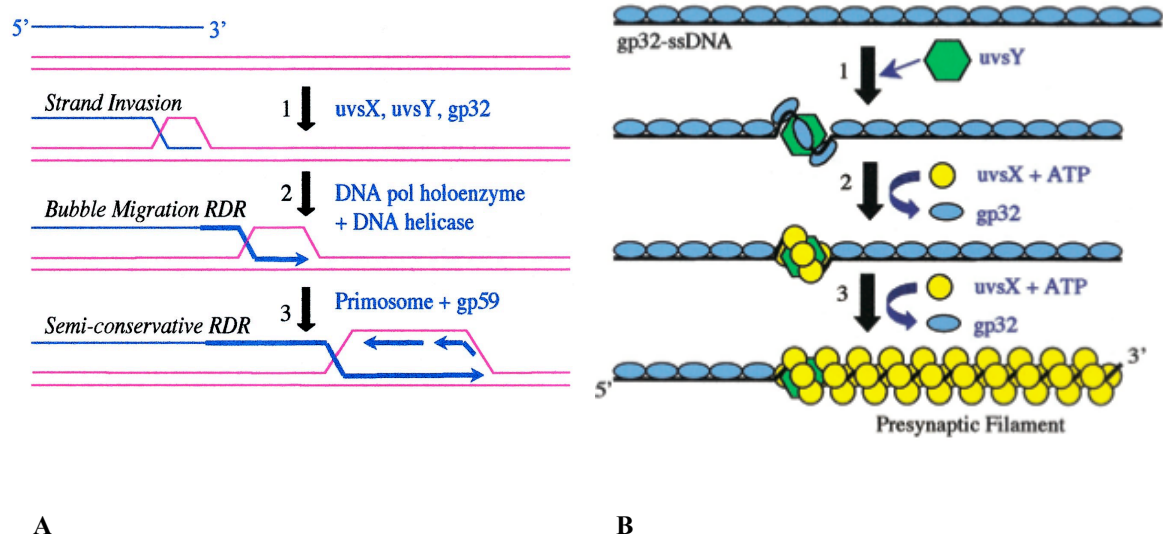
Fig.21. A: T4 *in vitro* general system for RDR. B: Biochemical model for UvsY-mediated assembly of the T4 presynaptic filament. Hexameric UvsY protein binds to gp32-ssDNA complex and destabilizes gp32-ssDNA interactions (Step1). UvsX recombinase is recruited to the UvsY-gp32-ssDNA intermediate and locally displaces gp32 to nucleate a filament (Step2). UvsX-ssDNA filament assembly propagates in the 5'→ 3' direction while displacing gp32 (Step3). (Bleuit et al 2001).

The priming event that initiates the leading strand synthesis component of RDR is catalysed by the T4 RecA homolog, UvsX protein. UvsX forms a presynaptic filament on a ssDNA "primer", which usually consists of a 3' ssDNA tail formed during origin-dependent replication of the phage linear duplex. Under physiological conditions, this reaction also requires the T4 UvsY and gp32 proteins. The T4 recombination mediator protein, UvsY, is necessary for the proper assembly of the T4 presynaptic filament, leading to the recombination-primed initiation of leading strand DNA synthesis and is necessary to stabilize UvsX recombinase, which cooperatively binds to ssDNA. As UvsX, UvsY protein appears to exist as a hexamer in solution. UvsY binds tightly but non-cooperatively to ssDNA and has weaker affinity for double-stranded DNA (dsDNA). It exhibits specific protein-protein interactions with other T4 recombination proteins including UvsX, gp32, and gp46/47. UvsY stimulates the DNA strand exchange activity of UvsX protein and is essential for UvsX activity at elevated concentrations of salt and/or gp32, conditions that approximate the physiological situation encountered by the T4 recombination system *in vivo* (Yassa et al 1997, Harris and Griffith, 1989).

The D-loop structure resulting from strand exchange contains a primed template capable of initiating DNA synthesis by the T4 DNA polymerase holoenzyme.

Addition of holoenzyme plus either of the two T4 DNA helicases (gp41/gp59) results in extensive leading strand DNA synthesis, initially by a conservative, "bubble migration" mechanism (Fig. 22A and B) (Bleuit et al, 2001).
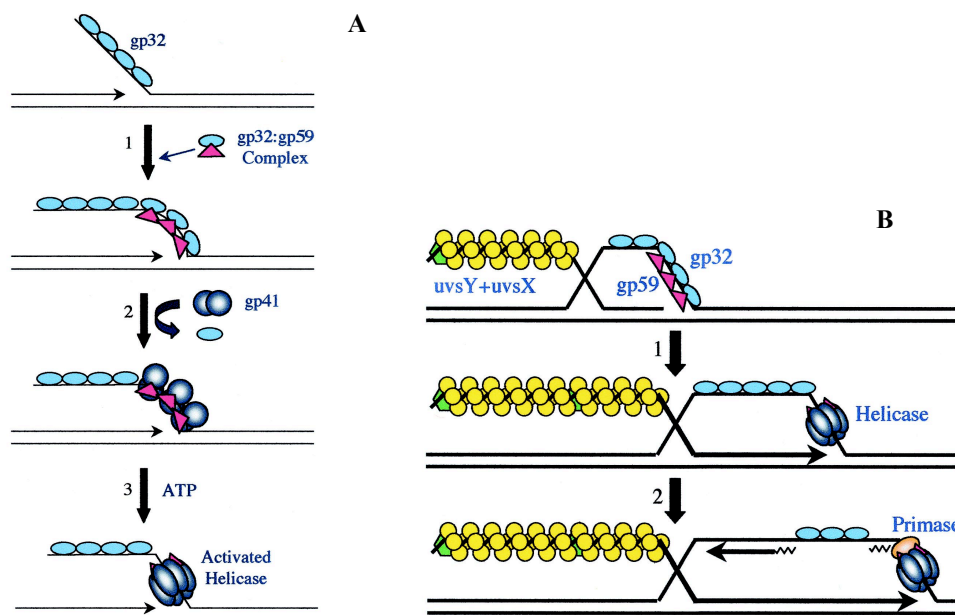

Fig. 22. A: Biological model for gp59-mediated helicase assembly at T4 replication fork, where cluster of gp32-gp59 complexes is incorporated at growing end of gp32 lagging-strand complex (Step 1), gp59 recruits dimers of gp41 helicase to the cluster and promotes displacement of gp32 (Step 2), gp59 stimulates ATP binding by gp41, triggering ring-hexamer formation by the helicase and the gp41-gp59 complex traslocates with the replication fork, ready to recruit primase. B: Enzyme partitioning model for strand-specific priming of Okazaki fragments during T4 recombination-dependent replication.

1.2.3.1.2   *Bacillus subtilis* SPP1 Bacteriophage

Bacillus subtilis bacteriophage SPP1 encapsidates its dsDNA into an empty procapsid by a processive headful packaging mechanism, using a linear head-to-tail concatemer as a substrate (Chai et al 1995). Previously, it has been shown that G38P, G39P and G40P are the only SPP1-encoded functions necessary and sufficient to drive theta replication from the cis-acting *ori*L region in an otherwise non-replicative element in *B.subtilis* cells (Missich et al 1997), but accumulation of SPP1 replication intermediates has not been observed by electron microscopy examination of SPP1 infected cells (Ganesan et al 1976). The generation of concatemeric SPP1 DNA is at least dependent on phage-encoded (G38P, G39P, G40P, G34.1P, G35P and

G36P) and host-encoded (DNA PolIII, ,DnaG and DNA topoisomerases) replication proteins, but is independent of host-encoded components of the primosome (e.g. DnaB, DnaD and DnaI) and recombination proteins (RecA, AddAB and RecF) (Burger et al 1978, Weise et al 1994).

Subsequently, the initiation of theta-type replication at SPP1 *ori*L has been defined. *In vitro* studies reveal that the helicase loader, GP39P, specifically interacts with the replisome organizer, G38P, and with GP40P. GP40P-ATP is a hexameric DNA helicase assembled to G38P-bound *ori*L and activated by G39P (Ayora et al 1999). GP35P and G34.1P share an overall identity of 40 and 18% with *E.coli* recombination proteins RecT and RecE, where RecT is a Single Stand Annealing protein and RecE is an ATP-independent 5'→ 3' exonuclease. Genetic evidence suggests that G35P is essential for phage DNA replication. Purified G35P binds single-strand DNA (ssDNA) and double strand (dsDNA) and specifically interacts with DNA helicase G40P and SSB protein G36P. G35P promotes joint molecule formation between a circular ssDNA and a homologous linear dsDNA with an ssDNA tail. Electron microscopic analysis shows that G35P forms a multimeric ring structure in ssDNA tails of dsDNA molecules and left-handed filaments on ssDNA. G35P promotes strand annealing at the AT-rich region of SPP1 *ori*L on a supercoiled template. These results altogether are consistent with the hypothesis that GP35P might be involved in Recombination-dependent-replication and might direct the assembly of the hexameric replicative helicase G40P at a D-loop structure by a new primosome assembly mechanism that does not require the primosome assembly proteins of *B.subtilis* as shown in Figure 23 (Ayora et al 2002).
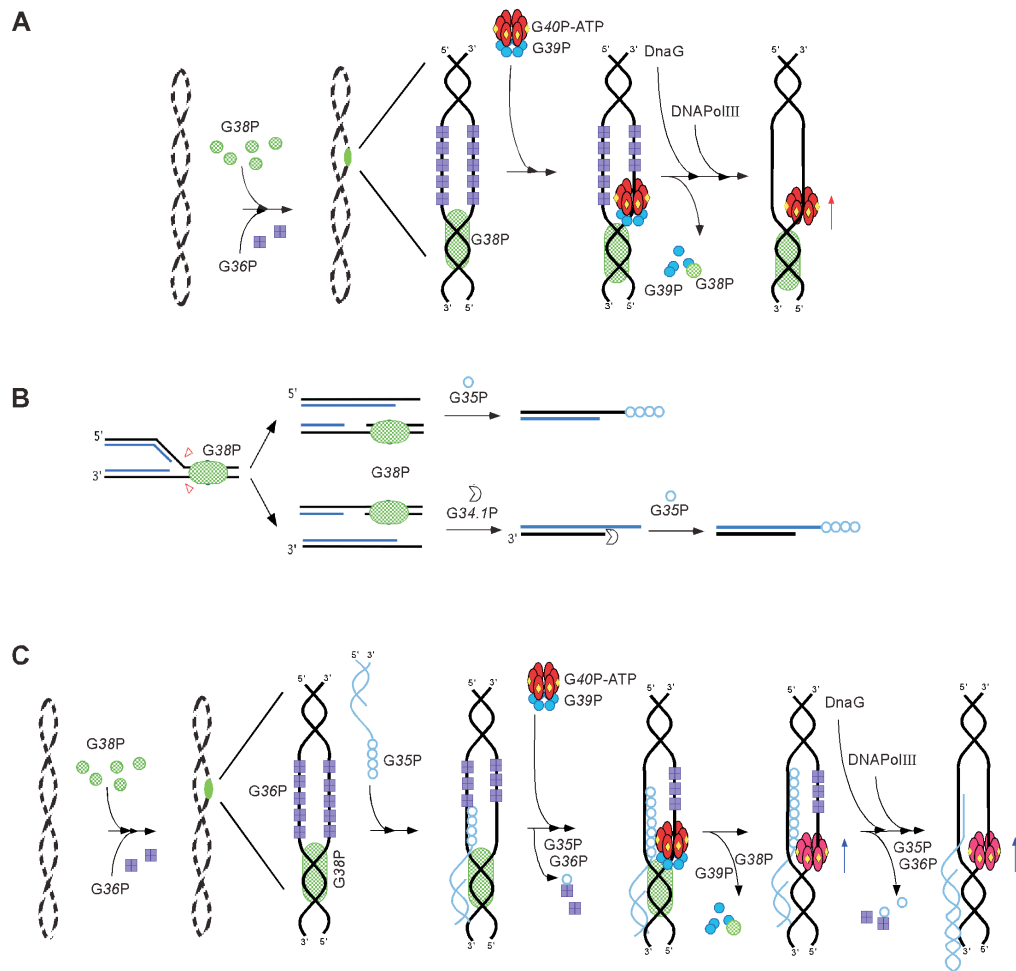
Fig. 23 Model for SPP1 initiation of DNA replication. A, Model for SPP1 initiation of theta type DNA replication. B, Roadblock as a model for the shift from theta to sigma replication. C, Model for SPP1 initiation of sigma type DNA replication (Ayora et al, 2002).

1.2.3.1.3   λ Bacteriophage

Recombination associated with Single-strand annealing (SSA) repair is the mechanism of choice for concatemeric phage and plasmids genomes. Both phage λ and the lamboid Rac prophage encode two recombinational repair enzymes; the repair reactions promoted by these purified enzymes are of the SSA type.

Soon after the discovery of the *recA* gene in *E.coli*, it was found that λ recombination is not affected by *recA* mutations. The λ genes *redα* and *redβ* were found to be required for this RecA-independent λ recombination; the corresponding proteins are an exonuclease and an ssDNA-annealing protein. If unchecked, the host RecBCD nuclease degrades the linear concatemeric products of λ rolling-circle DNA replication, but the Gam protein binds to RecBCD and inhibits all the known

activities of this enzyme. Gam analogs are produced by many bacteriophages of *E.coli* replicating their genomes as linear DNA. λ exonuclease, the gp redα, degrades the 5'-ending strand of linear duplex DNA, generating long 3' overhangs and also slowly degrades short ssDNA. During isolation from infected cells, half of the λ exonuclease activity is purified in a complex with another phage protein, called β. Beta protein, the gp redβ, is a 30-kDa ssDNA-binding protein which promotes the annealing of complementary DNAstrands. If the annealing reaction runs into a duplex region, Beta catalyzes strand displacement, also known as branch migration or strand exchange. The strand exchange reaction reveals that Beta promotes annealing with 5' and 3' polarity relative to the ssDNA to which it is bound. Beta does not bind duplex product of the annealing reaction. Another possible function of Beta is protection of single-stranded overhangs, generated by λ Exo, from the host ssDNA-specific exonucleases. This idea is substantiated by the in vitro finding that 3'ends stimulate Beta binding to ssDNA and double strand ends with 3' overhangs are protected by Beta from nuclease degradation four times better than are double-strand ends with 5' overhangs (Kuzminov et al, 1999). β protein exists in three structural states. In the absence of DNA, β protein forms inactive rings with about 12 subunits. The active form of the β protein in presence of oligonucleotides or single-stranded DNA is a ring, composed of about 15-18 subunits. The double-stranded products of the annealing reaction catalyzed by the ring are bound by β protein in a left-handed helical structure, which protects the products from nucleolytic degradation. These observations suggest structural homology for a family of proteins named SSAP (single strand annealing proteins), including the phage P22 Erf, the bacterial RecT and the eukaryotic Rad52 (Passy et al 1999).

1.2.3.2 Studies on lactococcal phages

Studies concerning Replication/Recombination, and proteins involved in these processes, to date have been not extensively developed in lactococcal phages.
In temperate lactococcal bacteriophage TP901-1, a single strand DNA binding protein homologue (ORF12) and the putative replication protein (ORF13) have been

identified in the early transcribed region of the phage genome. The putative origin of replication has been identified as a series of repeats within *orf13* and has been shown to confer to TP901-1 a resistance phenotype when present in trans. In particular, ORF13 could be the replication initiator protein, which would initiate assembly of the replication complex in phage TP901-1. This initiation could occur through the binding to specific repeats as previously shown in other phage systems, such as the *E.coli* λ phage (Kuzminov et al, 1999). With regard to TP901-1 phage, relevant is the identification of a *sak* gene (*orf11*), homologous of Sak3 (*orf35* of phage p2) (see section 1.1.3.4 and Fig.10). Considering its position in this replication cassette, *orf11* might encode for a Single Strand Annealing Protein (SSAP) or in general for another protein involved in recombinational mechanisms (Bouchard and Moineau, 2004). In a similar manner a replication cassette containing open reading frames, whose deduced proteins exhibited similarities to proteins known to be involved in DNA replication and modification, was found in temperate lactococcal phage Tuc2009 where an SSB protein (Orf15) a replisome organizer protein (Orf16), a methylase (Orf18) and a topoisomerase I (Orf14) were identified (McGrath et al 1999).

Recombination cassettes, such as the one previously described, were also found in *L.lactis* recombinant phages as a results of recombinational events between the phage and *L.lactis* harboring plasmids encoding Abi systems, as shown in phage ϕ31 (Durmaz and Klaenhammer, 2000) and in phage ul36 (see section 1.1.3.3). In phage ϕ31, this region shows homology with the DNA of various lactococcal temperate phages and the proteins encoded show homology with factors that may contribute to efficient recombination processes as the λ recombination protein BET and *E.coli* Hollyday junction resolvase Rus (Durmaz and Klaenhammer, 2000).

In the literature, to date, characterizations of lactococcal phage proteins involved in recombination mechanisms have not yet been realized.

Szczepanska and collegues characterized a SSB protein encoded by the *L.lactis* phage bIL67 *orf14* gene, thus defining a novel cluster of phage SSBs proteins. The purified SSB of bIL67 binds aspecifically to ssDNA via a putative OB-fold, whose structure has been predicted by three-dimensional modelling (Szczepanska et al, 2007).

A DNA Single-Strand Annealing Protein of lactococcal phage ul36 has been also characterized by Ploquin et al (Ploquin et al 2008): ORF252, renamed Sak (mnemonic for sensibility to AbiK) (see section 1.1.3.3) is a functional homolog of the eukaryotic RAD52, a protein involved in homologous recombination activities, with which Sak has a significant sequence homology. A more accurate description of Sak structural and functional features is present in the introduction of the article *"Sak protein from Lactococcal phage ul36: structure function study of protein and domains."* (Section 2.1.1.1).

1.3   *Aim of this study*

The project of my thesis work is focused on the characterisation of SAKs, the proteic products of specific lactococcal phage ORFs, that, when mutated, would confer insensibility to the Abortive Infection mechanism AbiK (Bouchard and Moineau, 2004). As already reported in Section 1.1.3.3, on the basis of their primary sequence, a SSAP function has been suggested for some of these ORFs, but, to date, very little is known for most of them. Thus the aim of my study has been to characterize some of them, in terms of structure and function relationship.

My first target has been the Sak protein of the *L.lactis* phage UL36. When I started my work, the same Sak was under investigation by our collaborators at the University of Laval (Masson's group), whose purpose was the identification of its the physiological function. To this aim, their starting point was the homologue protein Rad52 (Ploquin et al, 2008). On my side, and in complement to their functional studies, my work was oriented to the structural characterization of the protein by X-Ray crystallography techniques. Due to the poor diffraction of the Sak crystals obtained, we re-oriented the structural characterization on the single domains, that on the basis of the homology with Rad52 are supposed to be responsible of the functions of Sak (Article *"Sak protein from Lactococcal phage ul36: structure function study of protein and domains."* -Section 2.2.3). The structural investigation was carried out, in parallel, on wild type Sak and on a mutant (A92S) insensitive to the Antiviral Abortive mechanism AbiK (Section 2.2.3.2).

Afterwards I approached the characterization of Sak3 from phage p2, that is reported to be the reference for the third group of identified *sak* genes *(*Article *"Lactococcal Phage p2 ORF35-Sak3 is Involved In a Novel Type of Homologous Recombination"-* Section 2.2.2). Among the four proteins involved in sensitivity to AbiK (Tab. 10, Section 1.1.3.4), only Sak from UL36 and Sak2 from P335 show a significant sequence homology to single-strand annealing proteins (Bouchard and Moineau, 2004). Thereby I was interested in Sak3, as a case of unsuspected Sak protein, probably having the same functions of SSAPs.

The characterization of possible partners of Sak3 came later, through the study of the proteins encoded in the same "recombination cassette", like *orf35* (*sak3*).   In

particular I decided to clone and characterize *orf34* of phage p2, a sequence located downstream *sak3* that is supposed to be a single-stranded DNA binding protein (SSB) (Bouchard and Moineau, 2004) (Article *"Crystal Structure and Characterization of a New Type of Single Stranded DNA Binding Protein from the Lactococcal phage p2"* (Section 2.2.1). Due to the difficulties encountered during the experiments on Sak proteins (both Sak from phage UL36 and p2) , articles on Sak reported in Section 2.1.1.1 and 2.1.2.1 are in an early state of preparation, while the article on SSB protein has been already submitted to Molecular Microbiology Journal (Section 2.2.1).

The results on the two Sak proteins analyzed, on the ORF34$_{p2}$ and on their interactions have allowed elucidating some aspects of the lactococcal phage recombination/replication mechanism, and of the involvement of these specific proteins.

Chapter 2: Biochemical, biophysical and structural characterisation of Lactococcus lactis bacteriophage proteins involved in Homologous Recombination and Abortive Infection Mechanisms

2.1 SAK proteins

2.1.1 <u>Sak of L.lactis bacteriophage UL36</u>

The first object of my thesis was the comparison between Sak wild type and A92S mutant. It was previously shown that a mutation of Alanine to Serine 92 in lactococcal phage protein Sak makes phage ul36 insensitive to AbiK. The experimental work was realized in parallel with Masson's group and both proteins were subjected to a biochemical (stability in solution, DNA binding) and structural (degree of oligomerization) characterization utilizing large variety of different techniques (HPLC-MALS/UV/RI spectroscopy, Blue Native PAGE gels and Surface Plasmon Resonance). Our experimental results show a substantial identity between Sak wild type and A92S (Section 2.1.1.2) and were similar to these obtained in parallel by our colleagues of the Laval University (personal communication). In order to obtain good crystals for the determination of the 3D structure, we performed, on both Sak forms, several cristallogenesis screenings that unfortunately did not produce crystals with dimensions suitable for X-Ray analysis. For these reason, and taking into account that Rad52 structure was obtained from crystals of a truncated form, corresponding to an N terminal domain, I discomposed the protein into the N-terminal (both Sak wild type and A92S mutant - residues from 1 to 172) and the C terminal domains (residues from 173 to 252).

The domains were subjected to crystallization tests and and crystals with dimensions suitable for X-Ray analysis were achieved. Unfortunately the crystals, that diffracted only at 20 Å resolution, did not allow a good interpretation of the electron density truncated mutants (1-163, 1-181, 1-184, 1-192). Crystals of these mutants were

analized, but today no 3D structure was yet determined. Even though the 3D characterization was not succesfull a comparison between Sak full length and its domains was carried out in solution with the aim of comparing functional and structural aspects of the Sak protein (Article *"Sak protein from Lactococcal phage ul36: structure function study of protein and domains."* -Section 2.1.1.1).

2.1.1.1 Article "**Sak protein from Lactococcal phage ul36: structure function study of protein and domains.**"

Erika Scaltriti[1,2], Sylvain Moineau[3,4,5], Hélène Launay [7], Jean-Yves Masson[7], Claudio Rivetti[6] , Roberto Ramoni[2], Valérie Campanacci[1], Mariella Tegoni[1*] and Christian Cambillau[1*] .

[1]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, case 932, 13288 Marseille cedex 09, France;

[2]Istituto di Biochimica Veterinaria, Facoltà di Medicina Veterinaria, Università di Parma, Via del Taglio 8, 43100 Parma, Italy.

[3]Groupe de Recherche en Écologie Buccale (GREB), Faculté de Médecine Dentaire;

[4]Félix d'Hérelle Reference Center for Bacterial Viruses;

[5]Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Université Laval, Québec City, Québec, Canada, G1K 7P4.

[6]Istituto di Scienze Biochimiche, Università di Parma, 43100, Parma, Italy.

[7]Genome Stability Laboratory Laval University Cancer Research Center, Hôtel-Dieu de Québec 9 McMahon, Québec City, Québec, Canada, G1R2J6.

*Correspondence to Christian Cambillau (cambillau@afmb.univ-mrs.fr)
Tel. +34 491 82 55 90, Fax. +34 491 26 67 20

Abstract

Virulent phages are responsible for most fermentation failures due to infections to *Lactococcus lactis* in the dairy industry. In several lactococcal phages, putative Single-strand Annealing Proteins (SSAPs) were recently found, among which a Sak protein was identified in the virulent *L.lactis* phage ul36 [1]. Recently it was demonstrated that Sak is a functional homolog of RAD52, a SSAP involved in homologous recombination [2]. A comparison between Sak full length and its N- and C-terminal domains was carried out in solution with in mind to elucidate their functional and structural characteristics. We performed Blue Native PAGE, HPLC-SEC and AFM analysis to investigate their oligomerization state, while EMSA and SPR experiments were carried out to compare their capacity to bind DNA. Our results show that the N-terminal domain is essential and sufficient for oligomerization and binding to DNA, while the C-terminal domain does not bind DNA nor oligomerize. DNA binding, single or double-strand, occurs in a negatively charged crevice running externally to the ring, and similar to that of RAD52. Functional activities as annealing and stimulation of the RecA strand exchange indicate that only the N-terminal domain is capable of single strand annealing, while both domains do not stimulate RecA strand exchange reaction. We propose therefore that Sak N-terminus (residues 1-172) is involved both in DNA binding and function, *i.e.* annealing and stimulation of the RecA strand exchange.

Introduction

Lactococcus lactis, a lactic acid bacterium added to milk to control fermentation of dairy products, is susceptible to infection by virulent bacteriophages which are principally members of three species named 936, c2 and P335 that belong to *Siphoviridae* family [3]. Among defence barriers that bacteria act against virus infection, the Abi systems represent important mechanisms that include all cell defences that act after DNA ejection into the cytoplasm and lead to cell death, but little is known about the molecular mechanism used by proteins encoded from these systems to block infection. In particular AbiK system is encoded by the lactococcal plasmid pSRQ800 [4] and AbiK protein has a reverse transcriptase (RT) motif which can be involved in antiphage activity by synthesizing cDNAs strand from phage mRNAs templates, so blocking their translation, DNA replication and late gene expression [5]. Through the selective pressure of Abi systems, phage mutants, resulting of homologous recombination with phage-related sequences present in the host chromosome [6; 7], are found in unsuccessful fermentations. Regions replaced by homologous recombination from host genome are sequenced showing that each class II mutant has one point mutation in particular ORFs renamed Sak (sensitive to AbiK). Bouchard et al [1] have identified four non similar proteins involved in the sensibility to AbiK .

Sak protein of our interest has been identified as ORF252 (Sak) in phage ul36 (P335) [1] and has been recently studied by Ploquin et al [2]. Interestingly, Sak possesses the conserved binding domain of the eukaryotic protein RAD52, a single strand annealing protein (SSAP) [8] involved in DNA double-strand breaks repair mediated by homologous recombination. As Sak, many SSAPs are found in phages: RecT in *E.coli* prophage, Red beta in *E.coli* phage lambda and ERF in Salmonella phage P22. All these SSAPs are primarily of bacteriophage origin and have been acquired by numerous phylogenetically distant cellular genomes. Phage SSAPs are involved in genome circularization, concatamers DNA repair and DNA replication [9]. These functions have been also hypothesize for Sak of the phage ul36 in a first characterization in which is demonstrate that Sak is an homologous of human

RAD52 [2]. Sak in fact binds single-stranded DNA (ssDNA) preferentially over (dsDNA) promoting the renaturation of long complementary ssDNAs [2] and binds the recombinase RecA stimulating strand exchange reactions. From a functional point of view, Sak electron-microscopic reconstruction reveals an undecameric subunit ring, similar to the crystal structure of the N-terminal fragment of human RAD52 [10; 11] that has been extensively studied in all structural and functional aspects.

RAD52 has a dual role in recombination: first it stimulates RAD51-mediated strand invasion through direct interaction with RAD51 recombinase [12] and the ssDNA binding protein replication protein A (RPA); secondly it promotes single strand annealing (SSA) of complementary ssDNA independently of RAD51 [13; 14]. Both human and yeast RAD52 proteins bind to DNA and form a unique ring-like structure by itself and on DNAs [13; 15]. Initial efforts to crystallize the full length human RAD52 were unsuccessful, but a low resolution three dimensional structure has been determined showing that RAD52 forms heptametrical rings [16]. On the contrary, N terminal truncated protein (RAD52$_{1-212}$) has been crystallized and appears as an undecameric rings with a large channel running through the centre [10; 11]. The overall structure of the undecameric ring consists of a "stem" and a "domed cap" with a flat top: the stem region shares a β−β−β−α fold, present in a large number of RNA and DNA binding proteins, while the domed cap is constituted by amino acids residues flaking both ends (N and C terminal parts). [11]. In the RAD52 monomer, two helices, α1 and α2, in the N-terminal part make hydrophobic contacts with the top portion of the β−β−β−α fold in the stem region. On the other hand, the C terminal part (residues 157-205), that consists of a flexible extended linker that ends in an helix, stretches onto the N terminal domain of the neighbouring monomer. From a functional point of view, the N terminal domain has the ability to self associate and bind ssDNA which probably lies in an exposed groove on the surface of protein ring, whereas Rad51 interaction domain is thought to be located in the C terminal region of RAD52 [10]. Through this region, RAD52 interacts also with replication protein A (RPA), a eukaryotic single stranded DNA binding protein (SSB), which is involved in homologous recombination by destabilizing DNA secondary structures. Two binding sites for RAD52 have been identified on RPA [17], while a region on RAD52,

which includes residues 218-303, binds RPA70 as well as RPA32 subunits. RPA binding to RAD52 inhibits the higher-order self-association of RAD52 rings [17]. In human RAD52 there are in fact two experimentally separable self association domains: one in the N terminus (residues 1-192) responsible for assembly of individual subunits into rings, and one in the C terminus (residues 218-418) responsible for higher order oligomerization of rings [18].

On the basis of DNA binding and self-associating domain proposed for RAD52, we performed a series of analyses comparing Sak full-length with N terminal domain mutants and a mutant (residues 173-252) that spans the C terminal portion of Sak including the proposed RPA and RAD51-binding domain in RAD52.

Blue Native PAGE gels, HPLC-SEC analysis and Atomic Force Microscopy (AFM) were performed to investigate oligomerization state, while Electrophoretic Mobility Shift Assay (EMSA) and Surface Plasmon Resonance (SPR) experiments to compare the capacity to bind DNA. Functional activities as annealing and stimulation of the strand exchange were also tested and structural considerations were proposed on the basis of 3D modelling.

MATERIALS AND METHODS

Cloning, espression and purification

Cloning of full length phage ul36 ORF252 (Sak) in pET16b were previously described by Ploquin *et al.* [2]. The supposed ORF252 N-terminal domain (residues 1-172) and mutants 162,180,183,191 (respectively residues 1-162, 1-180, 1-183, 1-191) were recloned introducing a stop codon on pET16b-Sak sequence using QuickChange® Site-Directed Mutagenesis Kit (Stratagene). The C terminal region of phage ul36 Sak (residues 173-252) were amplified by PCR from pET16b-Sak using specific Gateway ^{TM} primers containing attB sequences at both ends, a ribosome binding site, and a N terminal His$_{10}$ tag coding sequence and then cloned by recombination in a Gateway ^{TM} pDEST17 vector (Invitrogen).

Protein expression experiments were carried out with Rosetta (DE3)pLyS (Novagen) and T7Rosetta (DE3) strains. After an overnight induction with 0.5 mM isopropyl 1-

thio-beta–D-galactopyranoside at 25 or 17°C depending on the target, cells were harvested by centrifugation for 10 min at 4000 X g. Bacterial pellets were resuspended in 40 μl/liter of culture of lysis buffer (Tris 50 mM, NaCl 300 mM, imidazole 10 mM, pH 8.0) supplemented with 0.25 mg/ml lysozyme, 1 μg/ml DNAse, 20 mM MgS0$_4$, and antiproteases (complete EDTA-free antiproteases, Roche) and frozen at –80°C. After thawing and sonication, lysates were cleared by a 30-min centrifugation at 12000 x g. Overexpressed proteins were purified on a Pharmacia Akta FPLC by nickel affinity chromatography (His –Trap 5 ml column, GE Healthcare) using a step gradient of imidazole followed by a preparative Sephacryl s500 HR26/60 gel filtration in 10 mM Bicine, 500 mM NaCl, pH 8.5. Only the C terminal region of phage ul36 Sak after nickel affinity chromatography was purified using a preparative Sephacryl s100 HR26/60 gel filtration in 10 mM Tris, 300 mM NaCl, pH 8.0. Purified proteins were concentrated using Amicon Ultra-15 ml (Millipore) and characterized by SDS-PAGE, matrix assisted laser desorption ionization time-of-flight mass spectrometry (Bruker Autoflex), trypsin peptide mass fingerprint, far –UV circular dichroism (Jasco J-810) and dynamic light scattering (Zetasizer Nano-S, Malvern). *Lactococcus lactis* recombinase RecA and E. coli RecA were purified as described previously [2; 19].

Blue Native PAGE

Blue native PAGE was performed using a nativePAGE[TM] Novex ® 3-12% BisTris gel (Invitrogen). Gels were run following the supplier's instructions loading 5 μg of each protein. Native gels were stained with Coomassie ® R-250. The NativeMark[TM] Unstained Protein Standard (Invitrogen) was used to estimate the molecular weight of proteins after native electrophoresis by plotting R_f (retardation factor) values versus $\log_{10}$ molecular weight. Protein migration was performed in the presence and absence of ssDNA 40 mer at saturating concentration.

SEC with On-line Multiangle Laser Light Scattering, Absorbance and Refractive Index (MALS/UV/RI) Detectors

SEC was carried out on an Alliance 2695 HPLC system (Waters) using a Superose 6 column (GE, Healthcare) eluted with 10 mM Bicine and 500 mM NaCl at pH 8.5 at a flow of 0.5 ml/min. Detection was performed using a triple angle light scattering detector (Mini-DAWN ™ TREOS, Wyatt Technology), a quasi-elastic light scattering instrument (Dynapro ™, Wyatt Technology), and a differential refractometer (Optilab® rEX, Wyatt Technology). Molecular weight and hydrodynamic radius determination was performed by ASTRA V software (Wyatt Technology) using a *dn/dc* value of 0.185 ml/g. Proteins were loaded at final concentration of 2 mg/ml.

Atomic Force Microscopy

Sak proteins were dialysed in buffer Hepes 20 mM pH 7.5, NaCl 250 mM and diluted to a concentration of 0.6 µM in a 10x solution. Glutaraldehyde was added, as a crosslinking agent at a final concentration of 0.11% [20]. The crosslinking reaction was incubated for 3 minutes at 4°C to permit the maintenance of the protein oligomers in the deposition step. The reaction is terminated by the addition of a 1/10 of the reaction volume of 1 M Tris-Hcl, pH 8.0. In the deposition step, the 10x concentred reaction was diluted in a deposition buffer containing Hepes 4 mM pH 7.5, NaCl 10 mM, MgCl2 4 mM. A 20 µl droplet of the binding reaction was deposited onto freshly cleaved ruby mica (Mica New York, NY) for 40 seconds. The mica disk was rinsed with water milliQ and dried with weak flux of nitrogen.

AFM images were obtained in air with a Nanoscope III microscope (Digital Instruments Inc., Santa Barbara, CA) operating in tapping mode. All operation were done at room temperature. Commercial diving board silicon cantilevers (Nanosensor or Olympus) were used. The microscope was equipped with a type E scanner (12 mm x 12 mm). Images (512 x 512 pixels) were collected with a scan size of 2 or 4 µm at a scan rate varying between two and four scan lines per second. Measurements of the oligomer dimensions were done using the Nanoscope III microscope Software. Data were plotted using Sigma Plot (Systat Software, Inc., San Jose, CA). The Nanoscope III images were imported into the Gwyddion software (AFM free data analysis

program. <http://gwyddion.net>) to obtain the average height and width, and the volumes measurements.

Polyacrylamide Electrophoretic Mobility Shift Assays (EMSA)

The complexes between Sak and DNA 40mer (ssDNA and dsDNA) were prepared by adding at a constant DNA concentration (10 µM) increasing amount of Sak to a final ratio of (0.4-1)-(8-1) (monomer:monomer) in 5 mM Hepes buffer, 100 mM NaCl, pH 8.3 containing 30% glycerol. In order to confirm the interaction, the complexes were also prepared at constant Sak concentration and increasing amount of DNA, at the same final ratio as above.

After a 15 minutes incubation at room temperature, samples were loaded on a 5% acrylamide gel (50 mM Hepes pH 8.3, 5% glycerol) and let run for 90 minutes at 50 V and 4°C, in 5 mM Hepes, 100 mM NaCl, pH 8.3 and 5% glycerol, as running buffer [21]. dsDNA alone and complexes Sak-dsDNA were visualized by staining the gel in the running buffer containing 0.5 mg/ml ethidium bromide for 10 minutes, whereas staining with 5000X SYBR Green II (Molecular Probes) in TBE (89 mM Tris Base, 89 mM boric acid, 1 mM EDTA pH 8.0) for 30 minutes was used to visualize ssDNA (Fig. 1. Supp.).

Binding reactions (10 µl) were performed in buffer (25 mM MOPS pH 7.0, 60 mM KCl, Tween 0.2%, 2 mM DTT, 5 Mg(CH$_3$COO)$_2$) containing 100 nM DNA. After 5 min at 37°C, the indicated amount of Sak full length or Sak mutants was added (2 µl) to the binding buffer and the incubation was continued for a further 10 minutes. Protein-DNA complexes were fixed by the addition of 0.2% glutaraldehyde followed by 15 min incubation at 37°C and then were analyzed on a 4.2% PAGE using Tris-Glycine buffer (50 mM Tris-Cl pH 8.8, 50 mM glycine). The gels were dried on DE81 filter paper followed by autoradiography. DNA substrates were prepared by annealing a 32P-labeled oligonucleotide (100 nucleotides in length) with appropriate complementary sequences. The 100-mer dsDNA, were purified by 10% PAGE. The sequence of the 100-mer was 5'-GGGCGAATTGGGCCCGACGTCGCATGCTCCT CTAGACTCGAGGAATTCGGTACCCCGGGTTCGAAATCGATAAGCTTACAG TCTCCATTTAAAGGACAAG-3'.

Surface Plasmon Resonance

Surface plasmon resonance experiments were performed on a Biacore 1000 (Biacore Inc. Piscataway, NJ) at 25°C. The chip Streptavidin (SA) was first washed with 10 mM NaOH, 1 M NaCl (3 x 100 µl at 40 µl/min) to eliminate streptavidin loosely bound. Biotinylated-DNA 40mer (ssDNA and dsDNA; 45 µg/mL) in 10 mM Bicine buffer, 150 mM NaCl, pH 8.0, 0.005% (vol/vol) P20 were fixed as ligand. Wild type and mutants Sak at concentration varying from 0.78 to 50 nM in 10 mM Bicine, 500 mM NaCl, pH 8.5, 0.005% (vol/vol) P20 were used as analyte (80 µl at 10 ml/min). High salt concentration was found to be necessary for Sak stability and helped decreasing aspecific interaction. Regeneration was achieved by injection of 4 M MgCl$_2$ (5 µl at 40 µl/min). The RU signal at different concentration of protein was corrected for the buffer contribution in the same flow-cell and for the aspecific interaction on the reference flow-cell. The K$_{DISS}$ values were estimated using 1:1 Langmuir Model (BIAevaluation Software).

In order to check interactions between Sak domains, Sak N terminal (1-172) was immobilized on a chip CM5 by amine-coupling and the C terminal domain (10 µM and 100 µM in 10 mM Tris buffer, 300 mM NaCl, pH 8.0) was used as analyte. No variation of the signal was detected upon the injection of the C-terminal domain.

The same approach and conditions were used to test the interactions with *Lactococcus lactis* recombinase RecA, a protein is involved in homologous recombination processes: firstly RecA was used as analyte on fixed Sak N terminal (1-172) and then RecA was fixed on a CM5 chip (deltaRU=727), using N and C terminal domains as analytes (10 µM and 100 µM in 10 mM Tris buffer, 300 mM NaCl, pH 8.0).

Single strand annealing and strand exchange reaction

Single strand annealing reactions (10 µl) contained denatured 5'-end labeled pPB4.3 NdeI-HindIII 400 bp (450 nM) with Sak in binding buffer (20 mM Hepes pH 7.5, 6.5 mM Mg(CH3COO)2, 2 mM ATP, 5 mM DTT, 100 mM NaCl and 100 µg/ml BSA). Incubation was performed at 20°C for 5 minutes. The reaction products were deproteinized by addition of one-tenth volume of stop buffer (10% SDS and 10 mg/ml proteinase K) followed by a 15 minutes incubation at 20°C. Labeled DNA

products were analyzed by electrophoresis through a 4% TBE1X/PAGE gel run at 175 V for 2.5 hours, dried onto DE81 filter paper and visualized by autoradiography. Reactions of stimulation of the strand exchange (10 µl) contained purified single-stranded pPB4.3 DNA (15 µM) with the indicated concentrations of *E. coli* RecA and Sak domains in TD buffer (25 mM Tris-Acetate pH 7.5, 8 mM MgCl2, 1 mM DTT, 1 mM ATP, 20 mM creatine phosphate, 5 U/ml phosphocreatine kinase). Sak domains were added first to the reaction. After 5 min at 37°C, 32P-end labeled pPB4.3 DNA (400bp fragment, 1.38 µM) was added and incubation was continued for 90 min. Reaction products were deproteinized by addition of one-fifth volume of stop buffer (0.1 M Tris-HCl, pH 7.5, 0.1 M MgCl2, 3% SDS, 5 µg/ml ethidium bromide and 10 mg/ml proteinase K) followed by 45 min incubation at 37°C. Labeled DNA products were analyzed by electrophoresis through 0.8% TAE agarose gels containing 1 µg/ml ethidium bromide, run at 4.3 V/cm, dried onto DE81 filter paper and visualized by autoradiography.

Crystallization and 3D modelling.

Crystallogenesis studies of SAK – Phage ul36 Sak full length and N terminal domain (1-172) were concentrated to 25 and 12 mg/ml, respectively and subjected to crystallization screening with a Cartesian nanodrop-dispensing robot [22]. Sak 1-172 crystals were obtained at 20°C by mixing 300 nl protein in 10 mM Bicine, pH 8.5, 500 mM NaCl with 100 nl 0.1 M Na2HP04, 0.2 M NaCl, pH 4.5 – 7.8% PEG 3000. ul36 Sak crystals were fished using a nylon loop and frozen in a synthetic solution of crystallization medium containing 18% PEG 4000 as cryoprotectant.

Homology modelization of Sak N-terminal domain. We generated a 3D model of Sak ul36 N-terminal domain 1-172, starting from the crystal structure of RAD52, Sak *vs* RAD52 alignment, and Sak secondary structure prediction. We used the program Turbo-Frodo [23] to mutate on display each residue and choose the most adequate side-chain conformation. The model was subjected to energy minimization with Refmac5 [24].

RESULTS AND DISCUSSION

Production of Sak proteins

We cloned and expressed Sak full length as well as a series of mutants of different length starting from N-terminus. These mutants were designed based on the secondary structure prediction and using the sequence alignment with RAD52 (Fig. 1A). All these mutants, except for 1-162, were successfully produced and were soluble up to several mgs/ml. Sak 1-252 was expressed as a soluble protein and could be concentrated up to 30 mg/ml. It was susceptible to protease cleavage and time-dependent degradation was detected by SDS PAGE. N-terminal sequencing showed that protease cleavage was at the C-terminus, generating two N-terminal fragments of 20 and 10 kDa, respectively. In other proteins bearing similar function or shape to Sak, such as rabies virus nucleoprotein (RVN), the carboxy-terminal part, which is necessary for protein-protein interaction with the viral phosphoprotein, is often cleaved off in vivo [25]. Full length protein, subjected to crystallogenesis assays, did not yield diffraction-quality crystals as in the case of human RAD52 [10], probably due to the co-existence of a mixture of different sizes of molecules or oligomers, as already observed by Ploquin et al. [2]. For this reason, and on the basis of secondary structure prediction and sequence alignment with RAD52 , we cloned the N- and C-terminal domains separately (Fig. 1). Initially we cloned the N-terminal mutant 1-172 inserting a stop codon in Sak 1-252 sequence at position 173 and we also generated the complementary C-terminal domain (173-252). The N-terminal domain 1-172 was purified soluble up to several mgs/ml, as the full length protein. High salt concentration (500 mM) were required, however, to keep this domain stable and devoid of time-dependent cleavage degradation. As observed in the case of the N-terminal domain of RAD52 [10; 11], the Sak N-terminal domain (1-172) crystallized readily (Fig.8, see below). However, these crystals diffracted poorly as observed also with recombinant RVN crystals [26]. In order to try to improve the crystals quality, we cloned a series of mutants of different size (1-162, 1-180, 1-183, 1-191). Unfortunately, no significant improvement was obtained, and all mutants showed physical and chemical characteristics similar to Sak 1-172. Only the mutant 1-162 was insoluble, probably because an important region, helix α5 (Fig 1), was cut

off as reported for RAD52 insoluble mutants lacking L10 and α5 [11]. Considering these observations, we kept Sak 1-172 as the reference N-terminal domain for the rest of the study.

The Sak C-terminal domain mutant (173-252) was expressed and purified, and stored as a soluble protein up to 10 mg/ml. Circular dichroism studies showed that this domain is mainly formed by a-helices, as proposed by secondary structure prediction, and is capable of partial refolding (T_m=45°C).

Sak N-terminal domain homology structure

No suitable crystals could be obtained from full-length protein, but the N-terminal domain 1-172 led to hexagonal crystals that diffracted up to 20 Å at best, and could not be improved. Sak mutant 1-180 and 1-183 crystals were obtained and showed similar diffraction. Since crystallization proved unsuccessful, we generated a 3D model of Sak ul36 N-terminal domain 1-172, starting from the crystal structure of RAD52, Sak *vs* RAD52 alignment, and Sak secondary structure prediction. Despite the low sequence identity (17% for the N-termini), modelling was straightforward, because only one residue insertion exists between the two sequences. The generated model is in agreement with secondary structure predictions and does not display clashes. Interestingly, the strong negative crevice observed in RAD52 is also present in Sak ul36, and is the obvious candidate for DNA harbouring (Fig. 2). The centre of the ring in Sak ul36 model exhibits, however a stronger negative surface compared to RAD52, despite the fact that both proteins display close pI (8.2 and 8.9, respectively).

Characterization of the oligomerization state

We performed blue native gradient gel (polyacrylamide 3-12%) analysis of Sak full-length and mutant 1-172 in the presence and absence of saturating amount of ssDNA. Both proteins exhibit several bands at high molecular weight, whereas they appear as a unique band at the monomer molecular weight in SDS gels (Fig 3). This result indicates a high oligomerization state in solution for the full length Sak [2] but also for the N-terminal domain. The presence of several bands indicates that Sak full length and N-terminus 1-172 form different ring populations. Moreover the presence of

saturating amounts of ssDNA makes it possible to the N-terminus 1-172 to enter in the gel, probably by diminishing its surface positives charges (Fig 3. lane 4-5) . This Sak behaviour suggests strongly that ssDNA could bind along the surface of the ring as suggested for RAD52 by Singleton et al [10] and by our homology model (see previous chapter).

Full length Sak and its N terminal domain (1-172) were also subjected to weight and size analysis using the MALS/UV/RI spectroscopy, which confirmed oligomerization state in solution. Sak 1-252 has a theoretical mass of 30275 Da and a MS measured mass of 31723 Da. MALS/UV/RI analysis revealed the presence of two major population of rings of 426 and 761 kDa, respectively. These MW account for ~13 and ~24 monomers for each species, respectively. Mutant 1-172, has a theoretical mass of 19558 Da and a MS measured mass of 21869 Da. It also exhibits a similar behaviour as the wt Sak, forming two major population of ring respectively of 280 and 560 kDa, accounting for ~13 and ~26 monomers, respectively (Tab.1 and Fig.4). Both Sak 1-172 and Sak 1-252 population of oligomers are formed of *n* and *2n* monomers, where *n* corresponds approximately to 13 units. Indeed, these numbers are only estimates, since we could not find a chromatographic column that could separate the two ring populations. The ring of full length Sak was estimated to contain 11 subunits, as shown by electron microscopy [2]. The presence of two major population of rings under the conditions tested that may correspond to single ring (*n*) and stacked ring populations (*2n*) as already reported in previous studies [2].

To characterize further the particle dimensions of the full length Sak and the Sak N-terminal domain, we have employed Atomic Force Microscopy (AFM) imaging. The deposition of either full length Sak or Sak N-terminal domain onto freshly cleaved mica caused disintegration of the oligomers (data not shown). Disintegration was completely prevented when the oligomers were crosslinked with 0.11% glutaraldehyde [20].

As shown in Figure 5A and B both proteins appears as a mixed population of globular species with different size. Analysis of the particles height (Fig. 5C a and b) performed on 40 oligomers for each protein (Sak 1-252 or Sak 1-172) shows the presence of two populations of particles with an average height of 2.9 ± 0.5 nm and

6.5 ± 0.9 nm for Sak 1-252 and of 0.8 ± 0.1 nm and 1.3 ± 0.1 nm for Sak 1-172 (Tab.1). In addition to height measurements we have also performed an analysis of the particle width (Fig. 5C c and d) showing that Sak-252 and Sak 1-172 have similar width (30.0 ± 4.3 nm and 26.1 ± 6.0 nm respectively). For a more accurate and complete analysis of the particle dimensions, we performed an analysis of the volume and the results are represented in Tab.1. Also in this case the particle volume segregates into two distinct populations with a volume of 2420 nm^3 and 6520 nm^3 for Sak 1-252 and of 810 nm^3 and 2770 nm^3 for Sak 1-172 respectively. These results support our working hypothesis that under this experimental conditions there exist two particle populations, one corresponding to a single oligomeric ring of about 13 subunits, and the other corresponding to two stacked oligomeric rings comprising a total of about 26 subunits.

Gathering all results on oligomeric state characterization (MALS/UV/RI spectroscopy and AFM), we established that in these experimental conditions a 1st population is formed by single ring, while a second one contains stacked rings. Ring populations with more than two stacked rings were also found, but they are very minor in our samples. This characteristic of forming rings of different size has been already observed for RAD52 that forms larger particles (80 nm of diameter), half-spheres (50 nm of diameter) and numerous 10 nm rings [18].

These results indicate clearly that full length protein and N-terminal domain 1-172 show similar oligomerization states. The C-terminal domain itself does not play any role in oligomerization. This domain (173-252) is produced as a monomeric protein with a small Rh value of 2.5 nm and does not self-associate in oligomeric forms. Furthermore, AFM studies on the width of oligomers and SEC-MALS Rh values of the N-terminal domain are smaller than in the full length protein. Considering a similar degree of oligomerization (Tab.1), this suggests that the C-terminal domain occupies a position at the periphery of the ring, accounting for the larger radius of the wt Sak. This behaviour of Sak is in contrast with that of RAD52: its full length has been shown by electron microscopy to be an heptamer [16], while X-ray studies identified an undecamer for its N-terminal domain [10; 11].

Interaction studies with DNA or between proteins or domains.

Polyacrylamide Electrophoretic Mobility Shift Assays (EMSA)

Gel shift assays were performed to analyze binding of Sak full length (1-252) and Sak-N-terminal (1-172) to dsDNA and ssDNA. In our experiments the amount of residual un-bound DNA decreased upon increasing Sak concentration, suggesting that Sak and DNA form more stable complex when the concentration of Sak increases. It has previously been shown by Ploquin et al (Ploquin et al 2008) that, in agreement with the behaviour of RAD52, Sak full length binds ssDNA with better affinity than dsDNA. Our results confirm a better affinity for ssDNA and strongly suggest an implication of Sak in strand annealing reactions [15]. Despite the fact that Sak 1-172 do not enter the gel as Sak 1-252, it is noteworthy that Sak N-terminal binds DNA similarly to the full-length protein (Fig. 6A-B) whereas no DNA binding was observed with the C terminal domain (173-252)(Fig. 6A-B). We propose that the DNA binding domain is located at the N-terminal portion of Sak, probably in the positively charged cleft of Sak model (Fig. 2) and the absence of the C-terminal domain has no effect on DNA binding. Experiments carried out with a constant Sak concentration and increasing DNA show a similar pattern (data not shown).

Surface Plasmon Resonance

SPR results show that Sak binds ssDNA and dsDNA with nanomolar affinity (Tab. 2). In particular Kd value of Sak full length (1-252) and Sak N-terminal for dsDNA are comparable, 2.2 and 2.7 nM, respectively. The affinity of both Sak wt and N-terminal domain for ssDNA is larger: slightly for Sak wt (1.8 nM), and 3-4 times better for the N-terminal domain (0.8 nM). The values measured for k_{on} and k_{off} are in the fast range, $k_{on} \approx 4 \pm 0.78 \times 10^5$ $M^{-1}s^{-1}$ and $k_{off} \approx 5 \pm 1.5 \times 10^{-4}$ s^{-1}, and are comparable in all cases excepted for the complex Sak 1-172/ dsDNA, for which, in spite of the very close value of Kd, the k_{off} was 2 order of magnitude slower than that with Sak full-length ($k_{off} = 5 \times 10^{-6}$ s^{-1}) (Fig 7). The very slow k_{off} is at the origin of higher error in the estimation of the constant. No binding with ssDNA nor dsDNA has been detected for Sak C-terminal domain.

Our interpretation of these results is that the N-terminal domain is essential and sufficient for binding DNA, as RAD52 N-terminal domain (Singleton et al 2002). However, the DNA/Sak interaction seems more stable in the presence of the C-terminal domain. The C terminal domain may stabilize dsDNA binding and may be involved in preferentially binding of ssDNA over dsDNA in a marginal way. Sak mutants (1-180, 1-183 and 1-191) DNA binding also give us evidences in that sense (data not shown).

In order to check interactions between Sak domains, some interaction experiments were performed by immobilizing Sak N-terminal (1-172) and passing over the C-terminal domain. No signal variation was detected, indicating that, when separated, both domains have no affinity. The interaction with *Lactococcus lactis* RecA was also tested both using RecA as analyte on fixed Sak N terminal (1-172) and using N and C terminal domains as analytes on fixed RecA. Also in this case no variation of the signal was detected upon the injection of analytes showing that none interaction of the single domains with RecA was present (data not shown).

Functional reactions of Single strand annealing and stimulation of the RecA Strand Exchange

In order to elucidate what region of Sak is delegate to annealing function and stimulation of RecA Strand Exchange, we carried out some reactions of single strand annealing and of stimulation of the RecA Strand Exchange. From these studies, we deduced that Sak N terminal domain is capable of single strand annealing, while C terminal domain do not lead to annealing of complementary oligonucleotides (Fig.8) showing that the annealing function is carry out by the N terminal portion of Sak. Moreover Sak N- or C-terminal domains do not stimulate RecA strand exchange reaction (data not shown) showing that the full-length protein is required. This result is in line with the lack of interaction between Sak domains and RecA in SPR experiments.

CONCLUSIONS

Our studies lead to a better characterization of Sak of the virulent bacteriophage ul36. We have shown that the region delegate to DNA binding is located in the N terminal first 172 aminoacids of the protein. Homology modelling of Sak using RAD52 structure indicates that Sak possess a negatively charged crevice along the external face of the undecamering ring, similar to that of RAD52. This portion of the Sak protein is also sufficient for oligomerization, while the C-terminal domain may protrude radially from the structure. Moreover the N-terminal domain is sufficient for annealing of complementary oligonucleotides, but it needs the C -terminal domain of the protein for RecA strand-exchange stimulation to occur. As previously shown, the full length protein is required for RecA-mediated homologous recombination and binding to RecA itself, unlike RAD52 that presents its interaction site in the C terminal domain [10].

Figure Legends

Fig 1. A: Secondary structure of Sak predicted by the PSIPRED protein structure prediction server [27]. Stop position of N terminal mutant 1-172 is indicated with the red bar. B: Primary sequence Alignment of Sak and Rad52 [28; 29].

Fig 2. Homology model of Sak ul36 (right) compared to the original crystal structure of RAD52 (left). The red arrows point to the rims of the positively charged crevice, the putative DNA binding site.

Fig 3. Blue Native PAGE gel of Sak 1-252 and Sak 1-172 in the absence (lane 2 and 4 respectively) and in the presence of saturating amount of ssDNA (lane 3 and 5 respectively). In lane 1: Protein Molecular Weight Standards.

Fig 4. HPLC-SEC profiles analysed using the MALS/UV/RI spectroscopy of Sak 1-252 (A), Sak 1-172 (B). Arrows refer to the two ring's population described in the text.

Fig 5. AFM images of the two population of rings of the Sak full length and the N terminal (1-172) crosslinked with glutaraldehyde (A and B respectively). The colour code corresponds to a height range of 2 nm from dark to clear. Image size: 1 µm. In C, The distribution of ring's heights and width of Sak 1-252 (a and c, respectively) and Sak 1-172 (b and d, respectively). Arrows refer to the two ring's populations schematized at the top of the panel.

Fig 6. Electrophoretic Mobility Shift Assays of Sak proteins with ssDNA (A) and dsDNA (B). Both gels (A and B) were run with Sak 1-252 (lane 2 to 5), Sak N terminal (1-172) (lane 6 to 9) and C terminal (lane 10 to 13) domains. ssDNA and dsDNA controls are in lane 1 of gel A and B respectively.

Fig 7 Kinetic analysis of interaction of DNA with Sak proteins. Representative sensograms showing the binding of ssDNA (left) and dsDNA (right) respectively to Sak 1-252 (A and B) and Sak 1-172 (C and D). The resonance signal is reported on the Y axis, the time on the X axis.

Fig. 8. Stimulation of the RecA Single-strand annealing by Sak N terminal (1-172) (lane 9 to 13) and C terminal (lane 2 to 6) domains. In lane 7: Sak 1-252 control.

Fig. 1 Supp. Electrophoretic Mobility Shift Assays of Sak 1-252 (left) and N terminal domain (1-172) (right) in complex with ssDNA (A) and dsDNA (B). Arrows refer to the protein-DNA complexes (top) and free DNA (bottom).

Tables:

Table 1. Oligomerization studies on Sak full length (1-252) and Sak N terminal (1-172): SEC-MALS/UV/RI Spectroscopy and AFM analysis.

Protein	Population	MALS/UV/RI spectroscopy			AFM	
		Refractometer mass (kDa)	Rh (nm)	Hypothetical Subunits	Height (nm)	Volume (nm3)
SAK 1-252	1° population (n)	426	8.7	13	2.9 ± 0.5	2420 ± 700
	2°population (2n)	761	9.6	24	6.5 ± 0.9	6520 ± 1300
SAK 1-172 N Domain	1° population (n)	280	6.5	13	0.9 ± 0.1	810 ± 100
	2° population (2n)	560	8	26	1.3 ± 0.1	2770 ± 580

Table 2. Surface Plasmon Resonance evaluation of the binding of Sak full length and mutants (N and C terminal domains) with ssDNA and dsDNA. ND=not detected.

Ligand	Analyte	Kd (nM)
ssDNA	Sak full length	1.85±0.77
	Sak N term	0.82±0.27
	Sak3 C term	ND
dsDNA	Sak full length	2.21±0.96
	Sak N term	2.72±1.18
	Sak3 C term	ND

REFERENCES

1.   Bouchard, J. D. & Moineau, S. (2004). Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J Bacteriol* **186**, 3649-52.

2.   Ploquin, M., Bransi, A., Paquet, E. R., Stasiak, A. Z., Stasiak, A., Yu, X., Cieslinska, A. M., Egelman, E. H., Moineau, S. & Masson, J. Y. (2008). Functional and structural basis for a bacteriophage homolog of human RAD52. *Curr Biol* **18**, 1142-6.

3.   Moineau, S. (1999). Applications of phage resistance in lactic acid bacteria. *Antonie Van Leeuwenhoek* **76**, 377-82.

4.   Boucher, I., Emond, E., Parrot, M. & Moineau, S. (2001). DNA sequence analysis of three Lactococcus lactis plasmids encoding phage resistance mechanisms. *J Dairy Sci* **84**, 1610-20.

5.   Fortier, L. C., Bouchard, J. D. & Moineau, S. (2005). Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. *J Bacteriol* **187**, 3721-30.

6.   Bouchard, J. D. & Moineau, S. (2000). Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. *Virology* **270**, 65-75.

7.   Bouchard, J. D., Dion, E., Bissonnette, F. & Moineau, S. (2002). Characterization of the two-component abortive phage infection mechanism AbiT from Lactococcus lactis. *J Bacteriol* **184**, 6325-32.

8.   Iyer, L. M., Koonin, E. V. & Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**, 8.

9.   Ayora, S., Missich, R., Mesa, P., Lurz, R., Yang, S., Egelman, E. H. & Alonso, J. C. (2002). Homologous-pairing activity of the Bacillus subtilis bacteriophage SPP1 replication protein G35P. *J Biol Chem* **277**, 35969-79.

10.   Singleton, M. R., Wentzell, L. M., Liu, Y., West, S. C. & Wigley, D. B. (2002). Structure of the single-strand annealing domain of human RAD52 protein. *Proc Natl Acad Sci U S A* **99**, 13492-7.

11.   Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T. & Yokoyama, S. (2002). Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol Cell* **10**, 359-71.

12.   Shen, Z., Cloud, K. G., Chen, D. J. & Park, M. S. (1996). Specific interactions between the human RAD51 and RAD52 proteins. *J Biol Chem* **271**, 148-52.

13.   Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S. & Ogawa, T. (1998). Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* **3**, 145-56.

14.   Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci U S A* **93**, 10729-34.

15.   Van Dyck, E., Hajibagheri, N. M., Stasiak, A. & West, S. C. (1998). Visualisation of human rad52 protein and its complexes with hRad51 and DNA. *J Mol Biol* **284**, 1027-38.

16.   Stasiak, A. Z., Larquet, E., Stasiak, A., Muller, S., Engel, A., Van Dyck, E., West, S. C. & Egelman, E. H. (2000). The human Rad52 protein exists as a heptameric ring. *Curr Biol* **10**, 337-40.

17.   Jackson, D., Dhar, K., Wahl, J. K., Wold, M. S. & Borgstahl, G. E. (2002). Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA. *J Mol Biol* **321**, 133-48.

18.   Ranatunga, W., Jackson, D., Lloyd, J. A., Forget, A. L., Knight, K. L. & Borgstahl, G. E. (2001). Human RAD52 exhibits two modes of self-association. *J Biol Chem* **276**, 15876-80.

19.   Eggleston, A. K., Mitchell, A. H. & West, S. C. (1997). In vitro reconstitution of the late steps of genetic recombination in E. coli. *Cell* **89**, 607-17.

20.   Dodson, M. E., H. (1991). Electron Microscopy of Protein-DNA complexes. *Methods Enzymol.* **208**, 168-196.

21.   Ausubel, L. J., Chodos, A., Bekarian, N., Abbas, A. K. & Walker, L. S. (2002). Functional tolerance is maintained despite proliferation of CD4 T cells after encounter with tissue-derived antigen. *Dev Immunol* **9**, 173-6.

22.   Sulzenbacher, G., Gruez, A., Roig-Zamboni, V., Spinelli, S., Valencia, C., Pagot, F., Vincentelli, R., Bignon, C., Salomoni, A., Grisel, S., Maurin, D., Huyghe, C., Johansson, K., Grassick, A., Roussel, A., Bourne, Y., Perrier, S., Miallau, L., Cantau, P., Blanc, E.,

Genevois, M., Grossi, A., Zenatti, A., Campanacci, V. & Cambillau, C. (2002). A medium-throughput crystallization approach. *Acta Crystallogr D Biol Crystallogr* **58**, 2109-15.

23. Roussel, A. & Cambillau, C. (1991). *The TURBO-FRODO graphics package.*, 81, *Silicon graphics Geometry Partners Directory*, Mountain View, USA.

24. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55.

25. Schoehn, G., Iseni, F., Mavrakis, M., Blondel, D. & Ruigrok, R. W. (2001). Structure of recombinant rabies virus nucleoprotein-RNA complex and identification of the phosphoprotein binding site. *J Virol* **75**, 490-8.

26. Albertini, A. A., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2007). Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. *J Struct Biol* **158**, 129-33.

27. McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-5.

28. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, 3320-3.

29. Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-90.
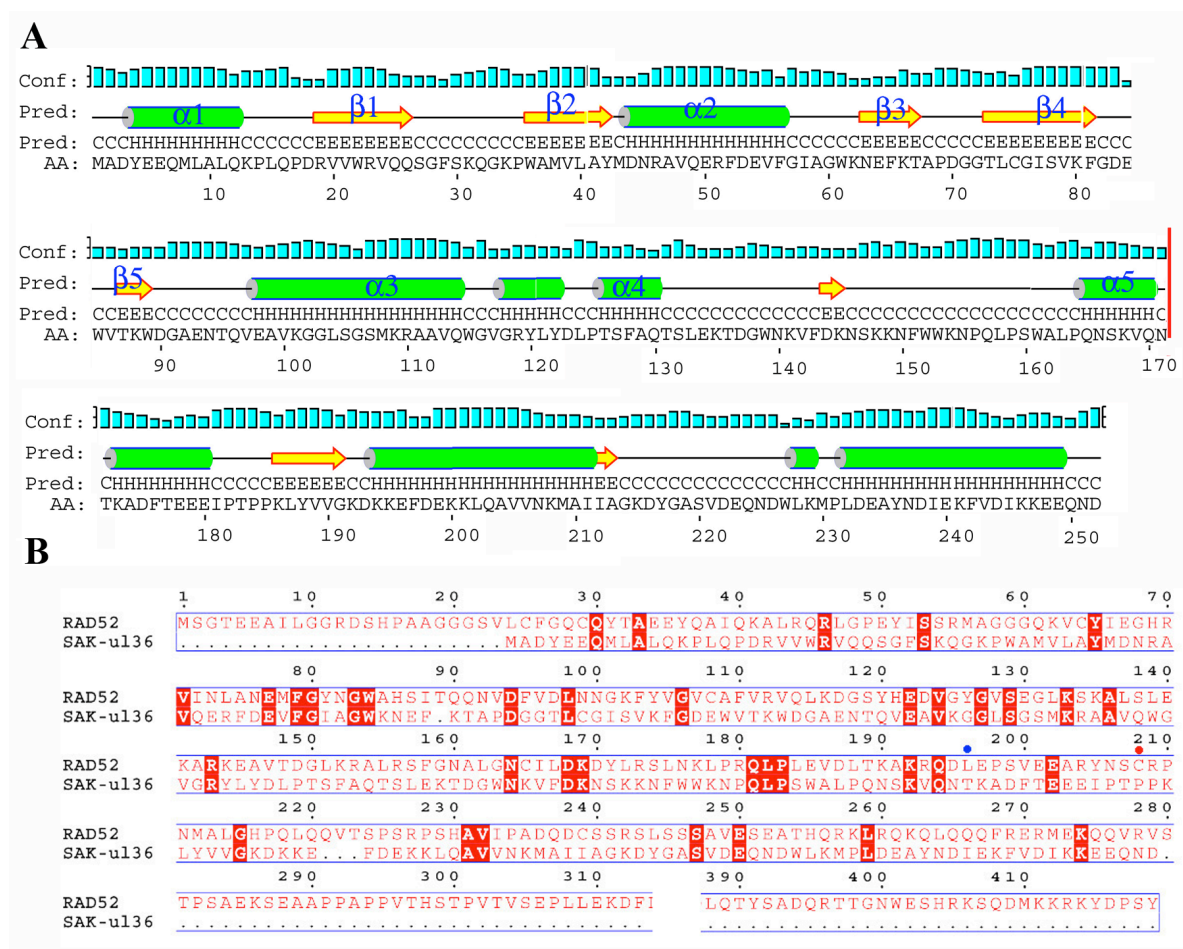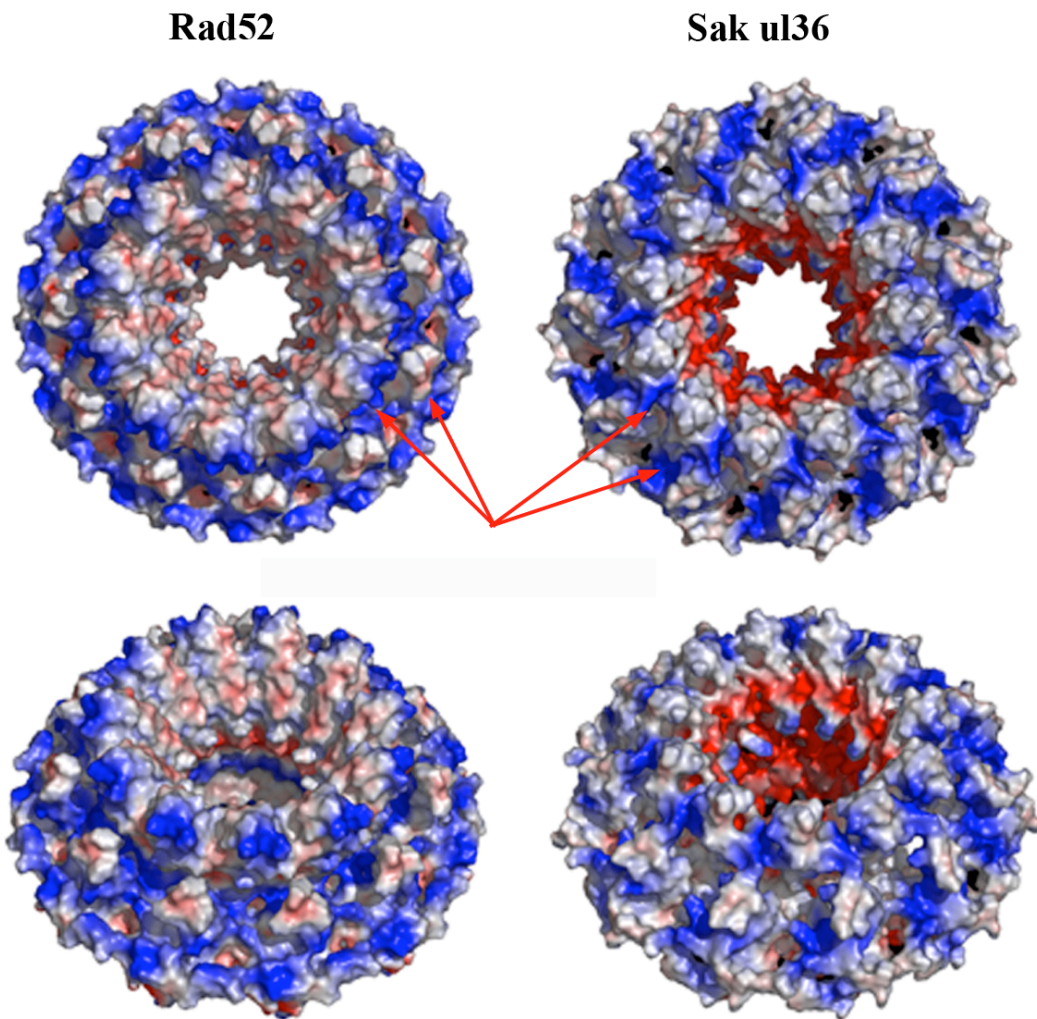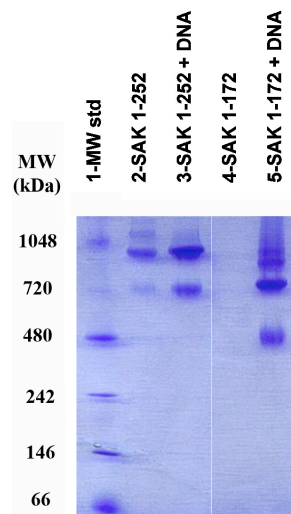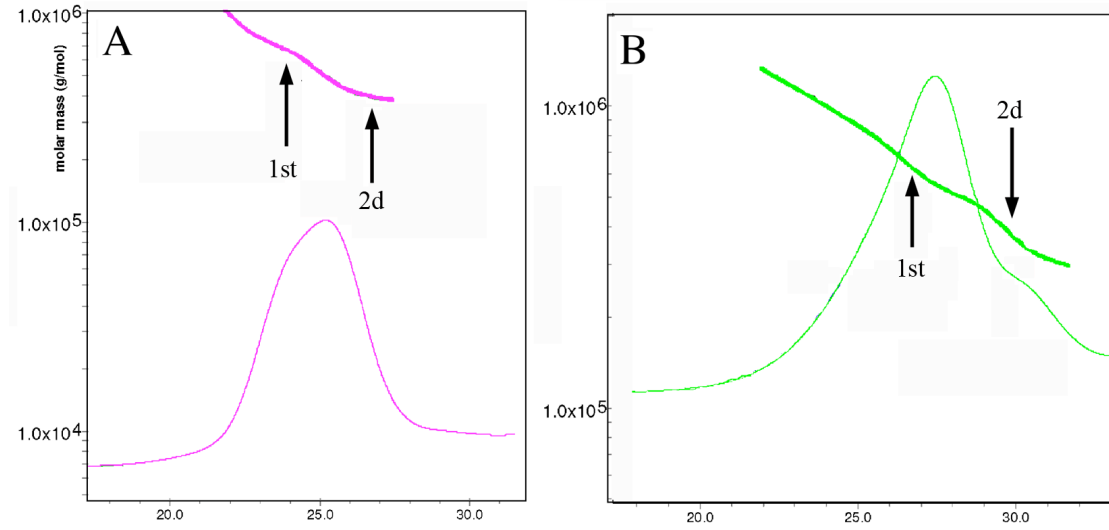
Figures

Figure 1
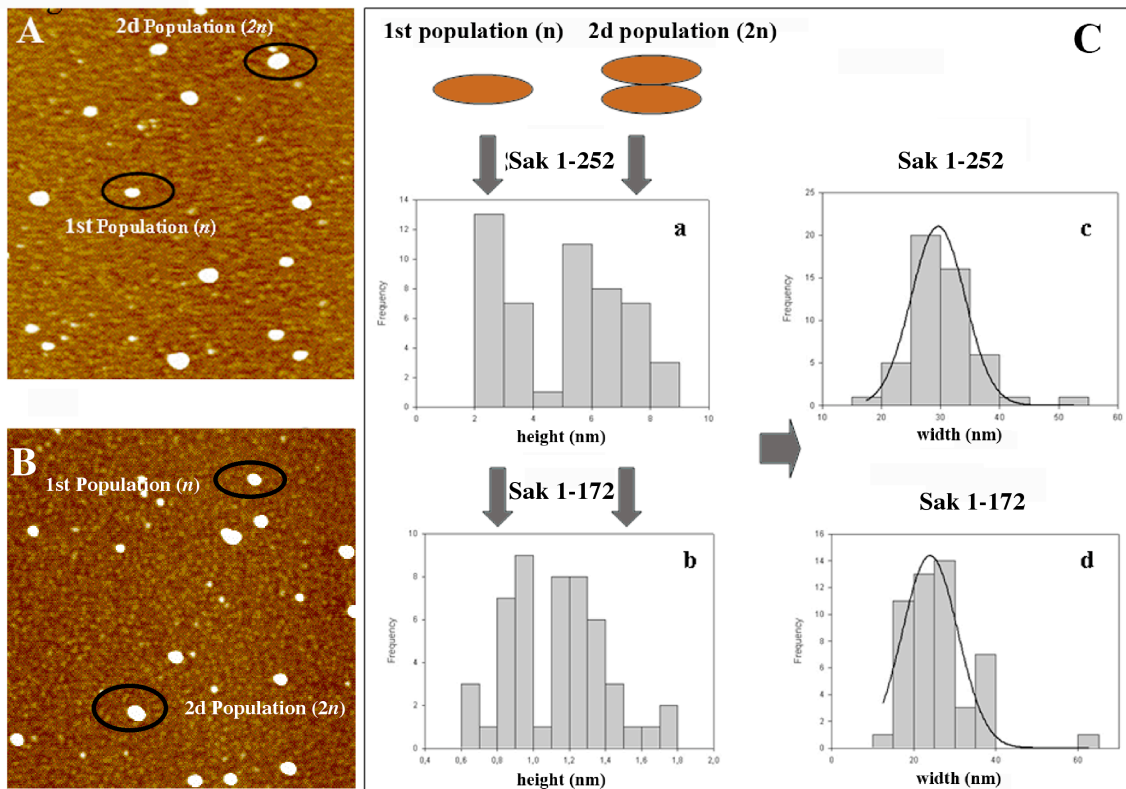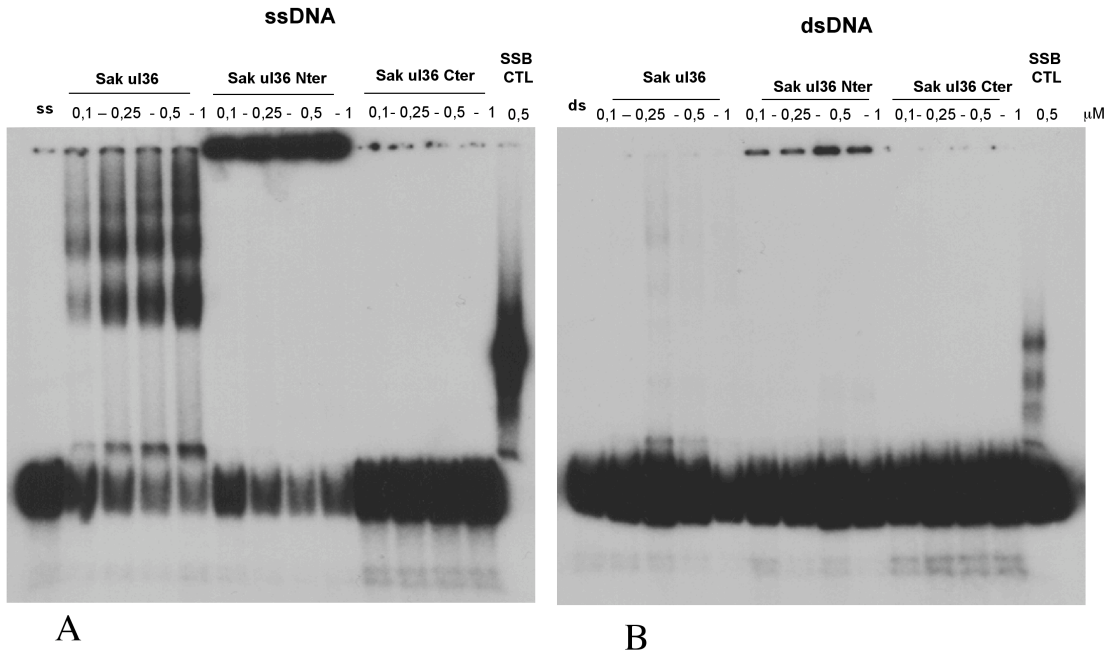
Figure 2

Rad52   **Sak ul36**

Figure 3

Figure 4
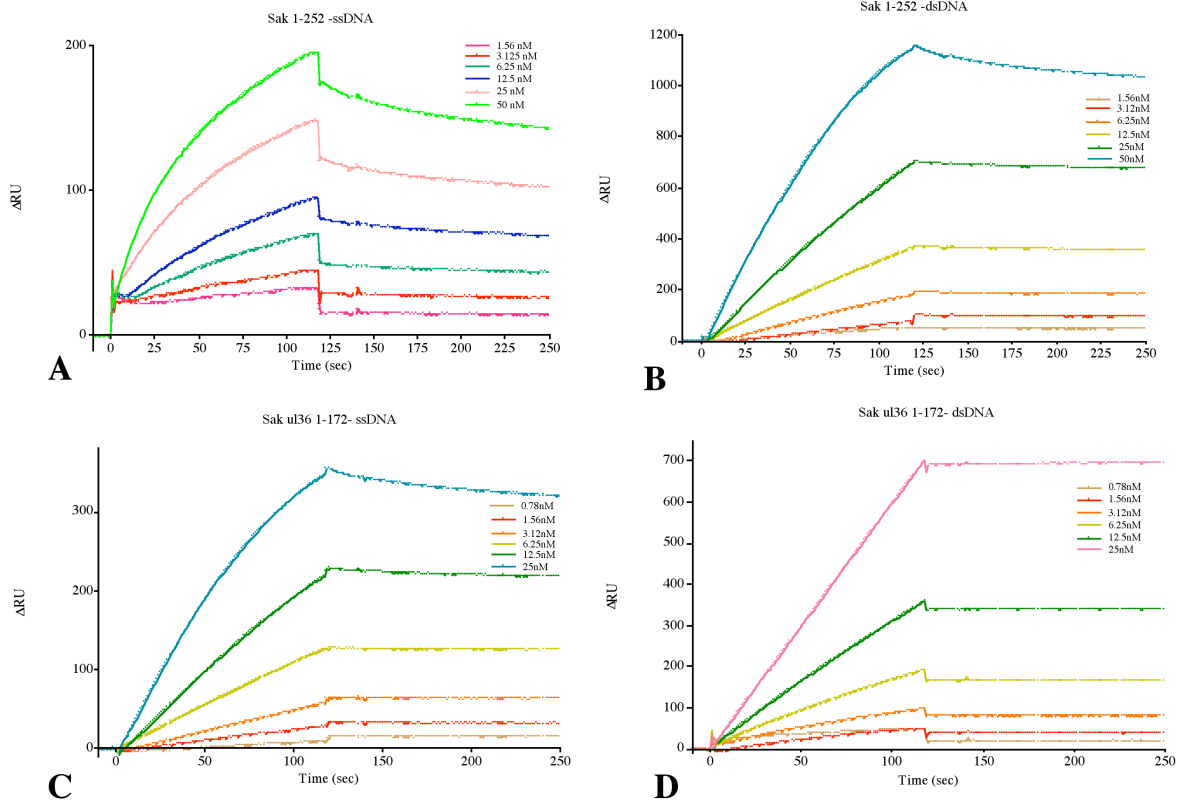

Figure 5

Figure 6
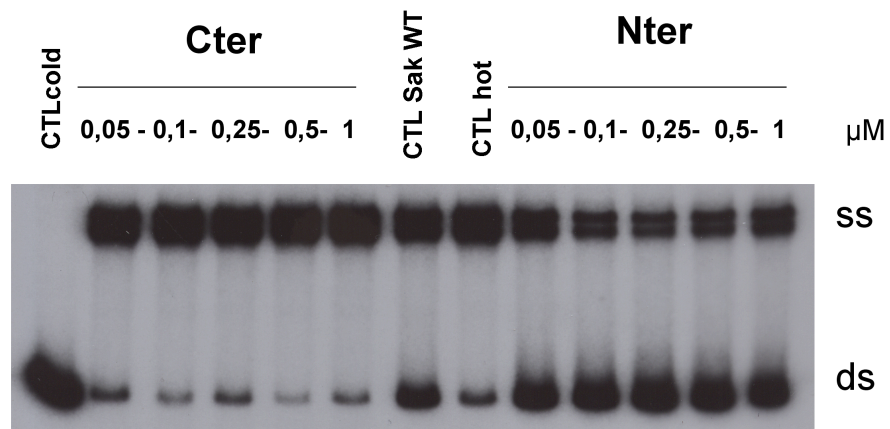

A

B
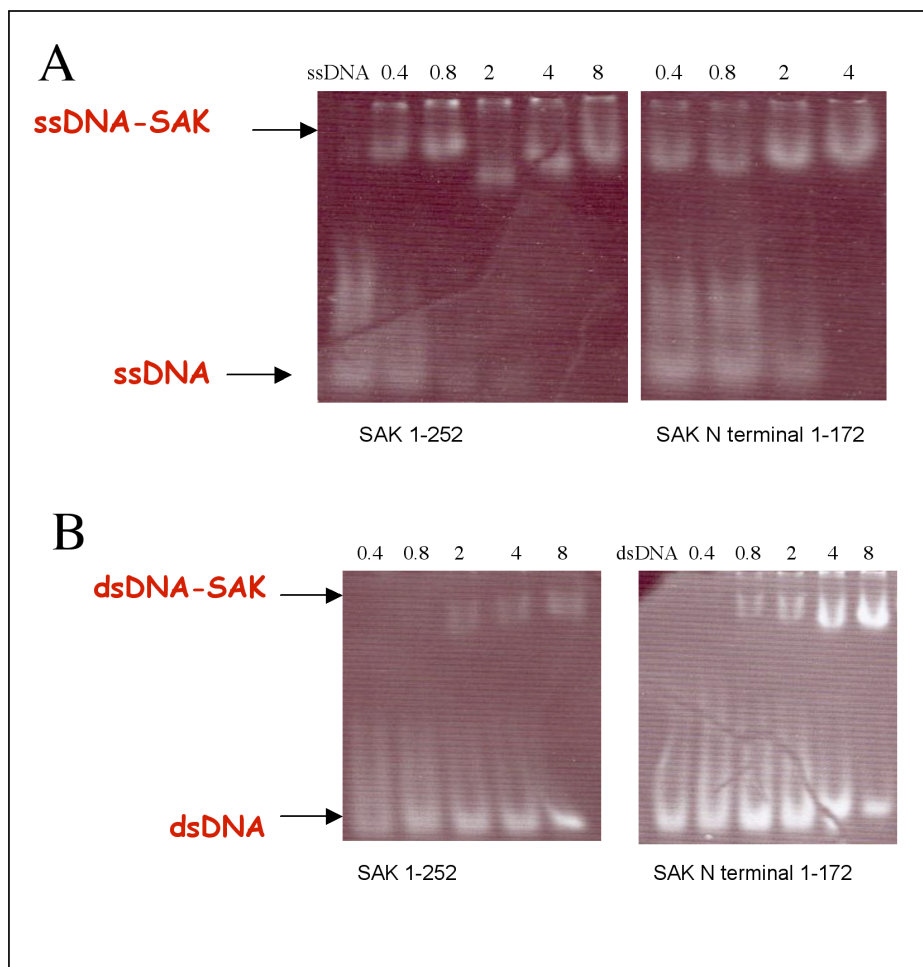
**Figure 7**

A

B

C

D

Figure 8

Supplementary Figure 1

2.1.1.2 Comparison between Sak wt and the mutant A92S

Sak wt and A92S mutant are both expressed in large amount in heterologous bacterial systems (25 mg/L of cellular *E.coli* culture) and are purified, with a similar procedure by an affinity and gel filtration chromatography. For this latest step, a Sephacryl s500 HR26/60 column was used; a 10 mM Bicine, 500 mM NaCl, pH 8.5 resulted ideal for protein stability. During purification, Sak wt and A92S mutant showed a similar behaviour in term of elution profiles (data not shown). HPLC-MALS/UV/RI studies, using a Superose 6 column in the same gel filtration buffer, showed that both Sak wt and A92S mutant are present in similar oligomeric populations and present comparable Rh values (Fig. 24A). This similar behaviour was confirmed by Blue Native PAGE gels that display the same pattern of bands for both Saks (Fig.24B).
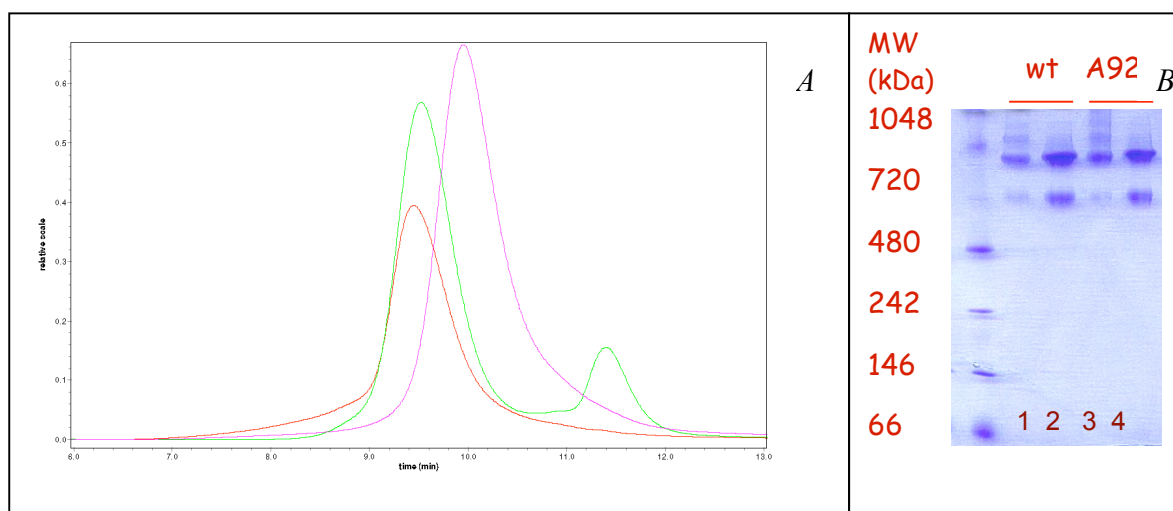

Fig. 24. Characterization of the oligomeric state. A) HPLC-MALS/UV/RI profiles of Sak wt (red line), A92S mutant (green line) and Sak N terminal (rose line). B) Blue Native PAGE gel. Lane 1:Sak wt, Lane 2: Sak wt-ssDNA at saturation, Lane 3: Sak A92S, Lane 4: Sak A92S-ssDNA at saturation.

Both Sak wt and A92S were also characterized for their capacity to bind DNA. Using Surface Plasmon Resonance (Biacore Technology), it turned out that both proteins bind ssDNA and dsDNA with nanomolar affinity (Tab.9). No significant difference in ssDNA and dsDNA dissociation constant could be detected using this Surface Plasmon Resonace approach.

LYGAND	ANALYTE	Kd (M)
ssDNA	Sak wt	$1.46 \pm 0.58 \ 10^{-9}$
	Sak A92S	$2.24 \pm 0.96 \ 10^{-9}$
dsDNA	Sak wt	$1.62 \pm 0.90 \ 10^{-9}$
	Sak A92S	$2.81 \pm 1.02 \ 10^{-9}$

Tab. 9 Binding of Sak wt and A92S mutant to ssDNA and dsDNA.

Polyacrylamide Electrophoretic Mobility Shift Assays confirmed that both Saks bind ss and dsDNA, but it was clear that they bind ssDNA preferentially over dsDNA (Fig. 25). As shown in Fig. 25 for Sak wt, the amount of unbound dsDNA is higher than unbound ssDNA.
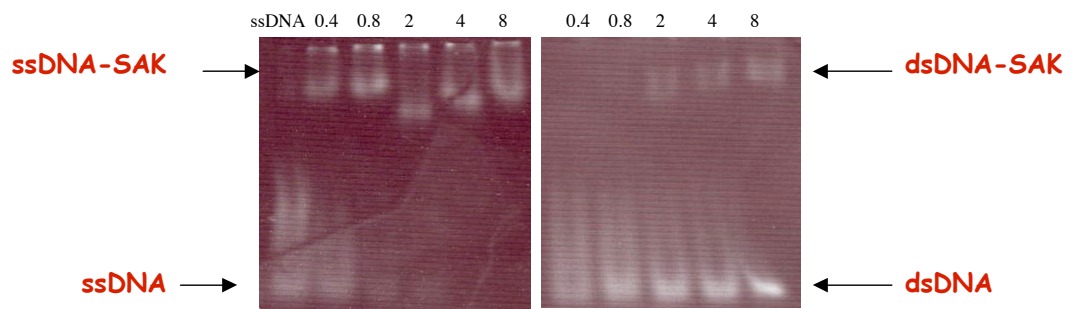
Fig. 25. Polyacrylamide Electrophoretic Mobility Shift Assays of Sak wt-ssDNA (left) and Sak wt-dsDNA (right).

As already reported, our results matched those obtained by our colleagues of the Laval University. In addition they found that both Sak and the A92S mutant promote the renaturation of long complementary single-strand DNA in a similar manner. Electron microscopy revealed also that both Sak proteins form rings that assemble into stacks on ssDNA (personal comunication).

On the basis of these experiments, we can assert that Sak wt and the A92S mutant seem to posses similar features, although the A92S mutation is shown to affect the phage lytic cycle (Bouchard and Moineau, 2004).

2.1.2 *Sak of L.lactis bacteriophage p2*

In order to describe the different types of Sak protein and understand their linkage with AbiK mechanism, we focused our studies on Sak3, a case of unknown Sak protein, which probably has similar SSAP functions. This protein is encoded by *orf35* of phage p2 genome, near the *ssb* (*orf34*), like Orf11 of phage TP901-1 that is located in proximity of an SSB protein (*orf12*) and a replicative protein (*orf13*) (Ostergaard et al 2000).

The position of sak3 gene in this "recombination cassette" suggests a Sak3 involvement in phage replication/recombination mechanism.

In the article that follows, we present a biochemical and structural characterization of Sak3 protein and mutants corresponding to N and C terminal domains. We have performed CryoEM studies to visualise ring structures, and functional tests, as annealing of complementary strands and stimulation of strand exchange of RecA, to elucidate the SSAP family membership. DNA binding properties were also tested by gel mobility shift assays (EMSA), atomic force microscopy (AFM) and surface plasmon resonance (SPR), and the relationship with their partners in homologous recombination was determined.

2.1.2.1 Article "Lactococcal Phage p2 ORF35-Sak3 is Involved In a Novel Type of Homologous Recombination"

Erika Scaltriti[1,2], Patrick Bron[8] , Sylvain Moineau[3,4,5], Hélène Launay [7], Jean-Yves Masson[7], Claudio Rivetti[6], Stefano Grolli[2], Valérie Campanacci[1], Mariella Tegoni[1]*, and Christian Cambillau[1]*.

[1]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, case 932, 13288 Marseille cedex 09, France;

[2]Istituto di Biochimica Veterinaria, Facoltà di Medicina Veterinaria, Università di Parma, Via del Taglio 8, 43100 Parma, Italy.

[3]Groupe de Recherche en Écologie Buccale (GREB), Faculté de Médecine Dentaire;

[4]Félix d'Hérelle Reference Center for Bacterial Viruses;

[5]Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Université Laval, Québec City, Québec, Canada, G1K 7P4.

[6]Istituto di Scienze Biochimiche, Università di Parma, 43100, Parma, Italy.

[7]Genome Stability Laboratory Laval University Cancer Research Center, Hôtel-Dieu de Québec 9 McMahon, Québec City, Québec, Canada, G1R2J6.

[8] Centre de Biochimie Structurale, Département "Structure, Cancer et Virulence", INSERM U554 / CNRS UMR5048, 29 rue de Navacelles, 34090 Montpellier, France

*Correspondence to Christian Cambillau (cambillau@afmb.univ-mrs.fr)
Tel. +34 491 82 55 90, Fax. +34 491 26 67 20

Abstract

Virulent phages of the 936 group, including phage p2, are predominant in infections to *Lactococcus lactis*, a Gram-positive bacterium widely used by the dairy industry. The study of the biology of lactococcal phages gains special significance in relation to the negative impact of phage infection on milk fermentation and the resulting economical looses. In this study, the product of the *orf35* gene from lactococcus phage p2 was characterized. ORF35 is also named Sak3 for its sensibility to the Antiviral Abortive mechanism AbiK[1]. It was considered to be involved in homologous recombination due to its localization upstream of a gene coding for a single-strand binding protein [2] and to belong to Sak proteins among which Single-strand Annealing Proteins (SSAPs) are found. Electron microscopy showed that $ORF35_{p2}$ has a ring shape typical of a superfamily of proteins such as the conserved RecA nucleotide-binding core that are ubiquitous in DNA recombination, replication and transcription. Gel-shift assays and surface plasmon resonance demonstrated that $ORF35_{p2}$ interacts preferentially with ssDNA with nanomolar affinity. Moreover atomic force microscopy showed that it binds DNA sticky over blunt ends. Despite the fact that these features are typical of a SSAP, our functional assays demonstrated that $ORF35_{p2}$ is unable to anneal complementary strands and inhibits instead of stimulate *E. coli* RecA-mediated homologous recombination. Although our results confirm that $ORF35_{p2}$ is involved in homologous recombination, the absence of typical SSAP functions suggests that Sak3 has novel functions probably associated with helicase/topoisomerase activities involved also in AbiK mechanism.

INTRODUCTION

Lactococcus Lactis, a lactic acid bacterium added to milk to control fermentation of dairy products, is susceptible to infection by virulent bacteriophages which are principally members of three groups named 936, c2 and P335 [3]. Among the mechanisms adopted by the bacterium against virus infection, there are abortive infection mechanisms (Abi) that include all cell defences that act after DNA ejection into the cytoplasm and lead to cell death [4], but unfortunately little is known about the molecular mechanism used by proteins encoded from these systems to block infection. In particular AbiK system is encoded by the lactococcal plasmid pSRQ800 [5] and AbiK protein has a reverse transcriptase (RT) motif which can be involved in antiphage activity by synthesizing cDNAs strand from phage mRNAs templates, so blocking their translation, DNA replication and late gene expression [6].

Through the selective pressure of Abi systems, phage mutants, resulting of homologous recombination with phage-related sequences present in the host chromosome [7; 8; 9], are found in unsuccessful fermentations. Bouchard *et al* [1] have identified four non similar proteins involved in the sensibility to AbiK . Among these proteins, Ploquin et al [10] demonstrated that ORF252 of phage ul36 (P335 group) renamed *Sak* is a DNA Single-Strand Annealing Protein (SSAPs) as previously proposed by Iyer et al [11] and that is a functional homolog of the eukaryotic protein RAD52, a protein involved in homologous recombination activities.

ORF35 of the *Lactococcus Lactis* phage p2, is one of the proteins involved in the sensibility to AbiK and was named Sak3 [1]. The primary sequence of Sak3 does not show any significant similarity to SSAPs as Rad52 or Sak, but a similar function was proposed [1].

The SSAPs are involved in DNA double-strand breaks repair and replication mediated by homologous recombination and function in a RecA-dependent or RecA-independent DNA recombination pathways. These proteins are primarily of bacteriophage origin and have been acquired by numerous phylogenetically distant cellular genomes [11]. In particular phage SSAPs are involved in genome circularization, concatemers DNA repair and DNA replication [12]. Genetic evidence suggests that the SPP1-encoded gene 35 product (G35P) is essential for phage DNA

replication. Purified G35P binds single-strand DNA (ssDNA) and double strand (dsDNA) and promotes joint molecule formation between a circular ssDNA and a homologous linear dsDNA with an ssDNA tail. Electron microscopic analysis shows that G35P forms a multimeric ring structure [12]. The structure of beta protein from bacteriophage lambda has also been studied by electron microscopy showing that the protein alone forms rings of about 12 subunits [13]. Beta protein promotes a single-strand annealing reaction that is central to Red-mediated recombination at double-strand DNA breaks and chromosomal ends and binds to ssDNA of minimal length of about 36 nucleotides with a site size of approximately one monomer per five nucleotides [14]. ssDNA binding is stimulated by 3' ends, while double-strand ends with 3' or 5' overhangs are protected by Beta proteins from nuclease degradation [15]. Although a large number of SSAPs are already identified, one of the well characterized SSAPs in terms of DNA binding, protein interactions and activity remains the eukaryotic Rad52. Rad52 has a dual role in recombination: first it stimulates RAD51-mediated strand invasion through direct interaction with RAD51 recombinase [16] and the ssDNA binding replication protein A (RPA); secondly it promotes single strand annealing (SSA) of complementary ssDNA independently of RAD51[17] [18]. Rad52 is an oligomeric protein in which the N and C terminal domains of each monomer can be clearly distinguished: the N terminal domain has the ability to self associate and to bind ssDNA which probably lies in an exposed groove on the surface of protein ring, whereas Rad51 interaction domain is thought to be located in the C terminal region of Rad52 [19]. The sequence independent interactions between Rad52 and DNA occur by single-strand and tailed duplex DNA molecules via precise interactions with the terminal base, in particular with 5' or 3' terminus of the ssDNA [20].

From a structural point of view, electron microscopic (EM) and atomic force microscopic (AFM) studies show that oligomerization and ring structures, formed by a superfamily of proteins containing the conserved RecA nucleotide-binding core, are ubiquitous in DNA recombination, replication and transcription [21]. These include the octameric Dmc1 [13], the δ^I subunits of the *E.coli* clamp loader complex [22] and hexameric helicases such as bacteriophage T7 gp4 [23], SV40 large T antigene [24],

DnaB [25] and the heptameric full length human Rad52 [21]. In this latter case, it should be noticed that the N-terminal DNA binding domain exhibits an 11 or 12 fold symmetry [21], and has been crystallized as an undecamer [19] [26]. These conserved structural features are unexpected in proteins so far from the functional points of view and with poor or none sequence homology [21].

The Dmc1 protein is a meiotic homolog of RecA, where octameric ring formation does not require ATP, Mg^{2+} or the presence of DNA [13; 27], while other members of this ring superfamily are involved in different recombination activities such as single strand annealing or separation of duplex oligonucleotides. In particular human Rad52 has an heptameric ring structure with a central channel and a ring diameter of about 40 Å and 130 Å respectively [21; 28].

Also the six subunits of hexameric helicases are arranged in a cyclic symmetry around a central channel and show a highly modular architecture: DnaB was reconstructed in both 6-fold and 3-fold symmetric states in the absence or in presence of nucleotide cofactors showing that N-terminal domain, which is important for hexamerization and is involved in trimer of dimers formation, exists in two alternate conformations [25]; Electron microscopic studies on gp4 helicase-primase of bacteriophage T7 also have shown that the protein forms hexameric rings with both C6 and C3 symmetry [29] suggesting that a great structural and mechanistic diversity may exist among the hexameric helicases.

Despite the low similarities to SSAPs as Rad52 or Sak, we hypothesize that Sak3 has SSAPs features and is involved in homologous recombination as other ORFs near *orf35* in phage p2 like *orf34* that encodes for a Single Strand Binding Protein recently characterized[2]. With this in mind, we performed a biochemical and structural characterization of Sak3 and mutants corresponding to N or C terminal domains. We have performed CryoEM studies to visualise ring structures and functional tests as annealing of complementary strands and stimulation of strand exchange of RecA, that are essential SSAPs features. Sak3-DNA binding properties were also tested by Electrophoretic Gel Mobility Shift Assays (EMSA), Atomic Force Microscopy (AFM), and Surface Plasmon Resonance (SPR). This latest technique also gave us some information about the interactions of Sak3 and its mutants with their partners in homologous recombination.

MATERIALS AND METHODS

Cloning, expression and purification

Phage p2 ORF35 (Sak3) gene and its C-terminal (residues 160-207) coding region were amplified by PCR from phage p2 genome using specific Gateway [TM] primers containing attB sequences at both ends and cloned by recombination in the Gateway [TM] pETG-20A vector (Dr Arie Geerlof, EMBL Hamburg). The final constructs encoded thioredoxin (Trx) fusion proteins containing a N-terminal hexahistidine tag followed by a tobacco etch virus (TEV) protease recognition site.

C-terminal deletion mutants of Sak3 (residues 1-159 and 1-176) were obtained by introducing a stop codon on pETG-20A-Sak3 using the QuickChange® Site-Directed Mutagenesis Kit (Stratagene). All constructs were checked by sequencing (GATC Biotech, France). Protein production was carried out in Rosetta (DE3) pLysS (Novagen) and T7 Express Iq pLysS (New England Biolabs) strains. After an overnight induction with 0.5 mM isopropyl 1-thio- beta – D-galactopyranoside at 17°C, cells were harvested by centrifugation for 10 min at 4000 x g. Bacterial pellets were resuspended in 40 ml/liter of culture of lysis buffer (Tris 50 mM, NaCl 300 mM, imidazole 10 mM, pH 8.0) supplemented with 0.25 mg/ml lysozyme, 1 µg/ml DNAse, 20 mM $MgSO_4$, and antiproteases (complete EDTA-free antiproteases, Roche) and frozen at –80°C. After thawing and sonication, lysates were cleared by a 30-min centrifugation at 12,000 x g. Over-expressed proteins were first purified on a Pharmacia Akta FPLC by nickel affinity chromatography (His –Trap 5 ml column, GE Healthcare) using a step gradient of imidazole (between 25 and 250 mM) followed by N-terminal thioredoxine cleavage using recombinant Tev protease. A second Ni affinity column was performed to eliminate Tev protease and thioredoxine followed by a preparative Sephacryl S500 HR26/60 gel filtration in 10mM Bicine, 500 mM NaCl, pH 8.5. Purified proteins were concentrated using Amicon Ultra-15 ml (Millipore) and characterized by SDS-PAGE, matrix assisted laser desorption ionization time-of-flight mass spectrometry (Bruker Autoflex), and trypsin peptide mass fingerprint.

Phage p2 Single Strand Binding Protein (SSB) was produced as previously described [2], while *Lactococcus Lactis* recombinase RecA and E. coli RecA were purified as described previously [10; 30].

Blue Native PAGE

Blue native PAGE was performed using a nativePAGE[TM] Novex ® 3-12% BisTris gel (Invitrogen) and following the supplier's instructions loading 5 μg of each protein. Native gels were stained with Coomassie ® R-250. The NativeMark[TM] Unstained Protein Standard (Invitrogen) was used to estimate the molecular weight of proteins after native electrophoresis by plotting R_f (retardation factor) values versus $\log_{10}$ molecular weight. Sak3 migration was performed with and without ssDNA 40 mer at saturating concentration.

SEC with On-line Multiangle Laser Light Scattering, Absorbance and Refractive Index (MALS/UV/RI) Detectors

SEC was carried out on an Alliance 2695 HPLC system (Waters) using a Superose 6 column (GE, Healthcare) eluted with 10mM Bicine and 500 mM NaCl at pH 8.5 at a flow of 0.5ml/min. Detection was performed using a triple angle light scattering detector (Mini-DAWN [TM] TREOS, Wyatt Technology), a quasi-elastic light scattering instrument (Dynapro [TM], Wyatt Technology), and a differential refractometer (Optilab® rEX, Wyatt Technology). Molecular weight and hydrodynamic radius determination was performed by ASTRA V software (Wyatt Technology) using a *dn/dc* value of 0.185 ml/g. Proteins were loaded at final concentration of 2 mg/ml.

Structural studies

Crystallogenesis studies of SAK – Phage p2 Sak3 protein was concentrated to 9.7 mg/ml and subjected to crystallization screening with a Cartesian nanodrop-dispensing robot [31]. Sak3 crystals were obtained at 20°C by mixing 3 μl protein in 10 mM Bicine, pH 8.5, 500 mM NaCl with 2 μl 66 mM Potassium Acetate, 33 mM HEPES, 66 mM Cacodylic Acid Sodium salt, pH 7.5 30% MME PEG 2000 obtaining a final concentration of 6 mg/ml. Crystal appeared after 3 days.

Electron microscopy

Samples for EM study were prepared from crystals as previously reported [32]. Fifteen crystals were harvested from the crystallization drop, washed with 33mM HEPES, 66mM Potassium Acetate, 66mM Sodium Cacodylate pH 7.5 buffer, then resuspended in 6 ml of the same buffer and crushed.

The sample was diluted to a final concentration of 0.05 mg/ml. Three microliters of protein complexes suspension were deposited on glow-discharged carbon-coated cupper grid and leaved 2 min in contact with carbon film. The excess of solution was blotted and 4 µl of 1% uranyl acetate were applied on grid for 1 min. The grids were then dried and keep in a dessicator cabinet until observation.

Electron micrographs were recorded under low-dose conditions with a FEI Tecnai Sphera LaB6 200 kV microscope. Images were collected using a 2K X 2K slow-scan CCD at x 50,000 magnification with a defocus range of 0.4 to 1.0 µm.

The 2D image processing was performed using the IMAGIC V software [33]. Single molecule images were extracted semi-automatically from images using Boxer [34]. The phase-contrast-transfer function was corrected by phase-flipping. Images were auto-centered using CENTER-IMAGE program using the average of raw images as reference. Some preferential views were selected by visual inspection and chose as references using the MRA (multi reference alignment) program included in IMAGIC V software. Images were then grouped into classes and averaged using the MSA (multi-statistical alignment) procedure. The best class averages were selected visually and used as new references for new alignment cycles. The class averages were rotationally averaged using IMPOSE-SYMMETRY included in IMAGIC.

Gel Retardation Assays

Binding reactions (10 µl) were performed in buffer (25 mM MOPS pH 7.0, 60 mM NaCl, Tween 1%, 2 mM DTT) containing 100 nM DNA. After 5 min at 37°C, the indicated amount of Sak3 or Sak3 mutants (Fig. 4) was added (2 µl) to the binding buffer and the incubation was continued for a further 10 minutes. Protein-DNA complexes were fixed by the addition of 0.2% glutaraldehyde followed by 15 min incubation at 37°C and then were analyzed on a 4.2% PAGE using Tris-Glycine

buffer (50 mM Tris-Cl pH 8.8, 50 mM glycine). The gels were dried on DE81 filter paper followed by autoradiography. DNA substrates were prepared by annealing a ^{32}P-labeled oligonucleotide (100 nucleotides in length) with appropriate complementary sequences. The 100-mer dsDNA, were purified by 10% PAGE. The sequence of the 100-mer was 5'-GGGCGAATTGGGCCCGACGTCGCATGCTCCT CTAGACTCGAGGAATTCGGTACCCCGGGTTCGAAATCGATAAGCTTACAG TCTCCATTTAAAGGACAAG-3'.

Surface plasmon resonance (SPR)

SPR experiments were performed using a Biacore 1000 instrument (Biacore Inc. Piscataway, NJ) at 25°C. The streptavidine (SA) chip was first washed with 10 mM NaOH 1M NaCl (3 x100 ul at 40 µl/min) to eliminate streptavidine loosely bound. Biotinylated-DNA 40mer (ssDNA and dsDNA; 45 µg/mL) in 10 mM Bicine buffer, 150 mM NaCl, pH 8.0, 0.005% (vol/vol) P20 were fixed as ligand (deltaRU=530 and 542, respectively). Wild type and mutants Sak3 at concentration varying from 1.5 to 312 nM in 10 mM Bicine, 500 mM NaCl, pH 8.5, 0.005% (vol/vol) P20 were used as analyte (80 ul at 10 µl/min). High salt concentration was found to help in decreasing the aspecific interaction between Sak3 and streptavidine covered dextran. Regeneration was achieved by injection of 4 M $MgCl_2$ (5 µl at 40 µl/min). The sensorgrams at different concentration of protein were corrected for the buffer contribution in the same flow-cell and for the aspecific interaction on the reference flow-cell. The K_D values were estimated using 1:1 Langmuir Model (BIAevaluation Software). In some cases, where the very slow k_{off} made it difficult to obtain a good global fitting, a local fitting or a simulation of the curves were used.

In order to check the network of Sak3-protein interactions we also immobilized on a CM5 chip possible partners of Sak3 in homologous recombination like phage p2 SSB and *Lactococcus lactis* RecA. In order to test Sak3 and SSB interaction, SSB was immobilized as ligand (deltaRU=515). Native Sak3 and mutants of different size were passed as analyte (from 0.15 to 10 µM, 80 ul at 10 µl/min) using 10 mM Tris, 150 mM NaCl pH 8.0 as running buffer. Regeneration was achieved by washing the flow cell with 4 M $MgCl_2$ (5 µl at 40 µl/min). SSB interaction with Sak3-ssDNA

complexes (1 Sak oligomer: 1 ssDNA molecule) was also carried out (0.1 µM to 2 µM, 80 ul at 10 µl/min) under the same buffer conditions as above.

The same approach and conditions were used to test the interactions with *Lactococcus lactis* recombinase RecA fixed on a CM5 chip (delta RU=727). The sensorgrams at different concentration of analyte were corrected for the buffer contribution in the same flow-cell and for the aspecific interaction on the reference flow-cell. The K_D values were estimated using Steady State Affinity model (BIAevaluation Software). In these experiments, in order to check for possible sterical hindrance due to covalent coupling and masking protein interaction, we have tested each protein alternatively as ligand and analyte.

Single strand annealing reactions

Sak3 was incubated with denatured 5'-end ^{32}P labeled 400 bp fragment (pPB4.3 digested with NdeI-HindIII, 450 nM) in binding buffer (20 mM Hepes pH 7.5, 6.5 mM Mg(CH$_3$COO)$_2$, 2 mM ATP, 5 mM DTT, 100 mM NaCl and 100 µg/ml BSA).Incubation was performed at 20°C for 5 minutes. The reaction products were deproteinized by addition of one-tenth volume of stop buffer (10% SDS and 10 mg/ml proteinase K) followed by a 15 minutes incubation at 20°C. Labeled DNA products were analyzed by electrophoresis through a 4% TBE1X/PAGE gel run at 175 V for 2.5 hours, dried onto DE81 filter paper and visualized by autoradiography.

Strand exchange

Reactions (10 µl) contained purified single-stranded pPB4.3 DNA (15 µM) with the indicated concentrations of E. coli RecA and Sak3 in TD buffer (25 mM Tris-Acetate pH 7.5, 8 mM MgCl$_2$, 1 mM DTT, 1 mM ATP, 20 mM creatine phosphate, 5 U/ml phosphocreatine kinase). Sak3 P2 was added first to the reaction. After 5 min at 37°C, ^{32}P-end labeled pPB4.3 DNA (400bp fragment, 1.38 µM) was added and incubation was continued for 90 min. Reaction products were deproteinized by addition of one-fifth volume of stop buffer (0.1 M Tris-HCl, pH 7.5, 0.1 M MgCl$_2$, 3% SDS, 5 µg/ml ethidium bromide and 10 mg/ml proteinase K) followed by 45 min incubation at 37°C. Labeled DNA products were analyzed by electrophoresis

through 0.8% TAE agarose gels containing 1 μg/ml ethidium bromide, run at 4.3 V/cm, dried onto DE81 filter paper and visualized by autoradiography.

Atomic Force Microscopy

Single-stranded (ss) circular M13 mp18 ssDNA was purchased from Sigma Aldrich (Saint Louis, Missouri USA). The 349 base pairs (bp) long dsDNA with 5'-phosphate blunt ends was obtained by restriction digesting of pNEB plasmid with PvuII endonuclease (New England Biolabs). The 641 bp long dsDNA with 5'-OH blunt ends was obtained by PCR using pDE13 plasmid as template, unphosphorylated primers and Deep-Vent DNA polymerase (New England Biolabs). The 656 bp long dsDNA with a single stranded gap of twenty T near the middle of the fragment was produced as described in ref. [35]. The 1004 bp long dsDNA fragment with one nucleotide (nt) overhang at the 3' was obtained by PCR using pNEB193 plasmid as a template and Taq DNA Polymerase (Fermentas). The 835 bp long dsDNA fragment with four nucleotide overhang at the 3' was obtained by restriction digesting of plasmid DNA with Sac I endonuclease (New England Biolabs). All DNA substrates were purified on a 1% TAE agarose gel. DNA was recovered from the excised band with PCR Clean-up Nucleospin Extract II (Macherey-Nagel).

Sak3-DNA binding reactions were assembled in 20 mM Hepes pH 7.5, 250 mM NaCl with a final Sak3-monomer:DNA molar ratio of 16:1. The reactions were incubated at 25°C for 15 minutes, after which glutaraldehyde, to a final concentration of 0.11%, was added [36]. Crosslinking was carried out at 4°C with an average incubation time of 3 minutes to prevent DNA-DNA aggregation. The crosslinking reaction was terminated by addition of 1/10 of the reaction volume of 1 M Tris-HCl pH 8.0. For AFM imaging, samples were diluted ten times in deposition buffer (4 mM Hepes pH 7.5, 50 μM spermidine for sample containing M13 ssDNA or 4 mM Hepes pH 7.5, 10 mM NaCl, 2 mM $MgCl_2$ for samples containing dsDNA) and a 20 μl drop was deposited onto freshly cleaved ruby mica (Mica New York, NY) for 40 seconds. The mica disk was immediately rinsed with water milliQ and dried with weak flux of nitrogen. AFM images were collected in air with a Nanoscope III microscope (Digital Instruments Inc., Santa Barbara, CA) operating in tapping

mode. All operations were done at room temperature. Commercial diving board silicon cantilevers (MikroMasch) were used. The microscope was equipped with a type E scanner (12 mm x 12 mm) and images (512 x 512 pixels) were collected with a scan size of 2 or 4 µm at a scan rate varying between two and four scan lines per second. Scoring of Sak3 oligomers bound to DNA and measurements of the oligomer dimensions were done manually using the Nanoscope III Software. In order to compensate for the tip broadening effect, the width of the oligomers was measured at half of their height. Data were plotted using Sigma Plot (Systat Software, Inc., San Jose, CA).

RESULTS

ORF35 production, biochemical and structural characterization of the oligomeric state

Phage p2 ORF35 (Sak3) was produced as a soluble protein in fusion with thioredoxine. After Tev protease cleavage and purification, Sak3 could be concentrated up to 10 mg/ml. HPLC-SEC with MALS/UV/RI (Wyatt technology) and Blue Native PAGE results showed that Sak3 exhibits an oligomeric structure comparable to that of RAD52 [19] and phage ul36 Sak [10]. In particular, in Blue Native PAGE (Fig.1A) Sak3 migrates at very high molecular weight giving a confirmation of oligomeric Sak3 structure. Moreover the presence of saturating amounts of ssDNA (Fig.1A lane 3) stabilizes different populations that might correspond to ring's population as shown for RAD52 by EM [37] and for Sak proteins [10]. HPLC-SEC experiments confirm that Sak3 is an oligomeric protein as Rad52 (Fig.1B), despite a low similarity to Rad52 (21% of primary sequence homology) (Fig.2A) [38; 39]. MALS/UV/RI analysis shows the presence of a single population of protein of 353 kDaltons with an Hydrodynamic Radius (Rh) of 7.7 nm. As Sak3 has a theoretical mass of 23308 Da, we suggest that each ring is formed of about 15 subunits.

Oligomeric structure determined by electron microscopy.

ORF35 native protein. ORF35 has been submitted to extensive crystallization assays. Although crystals were obtained readily (Fig 3a), we could never obtain crystals

diffracting beyond 20 Å resolution, despite tremendous optimization efforts. Preliminary screening with electron microscopy showed, however, that the crystal content was leading to more homogeneous material than the material issued directly from purification.

Crystals (Figure 3A) were collected from crystallization drops and crushed into buffer. Sak3 were negatively stained using 1% uranyl acetate and imaged by transmission electron microscopy as shown in Figure 3B. Rings are observed in images. Preliminary screening with electron microscopy showed, however, that the crystal content was leading to more homogeneous material than the material issued directly from purification. Although not well contrasted there are easily visible and indicated by an arrow in Figure 3B. To go further in the understanding of Sak3 organization, image processing was carried out. From 6 images recorded with a slow-scan CCD at 50000x magnification, 500 particles were extracted semi-automatically with a pixel size corresponding to 2.1 Å. Images were aligned initially on representative views, grouped into homogeneous classes which were averaged, named "class averages". Class averages showing good signal to noise ratio and particles having homogeneous shapes were used as new references for a new image alignment cycle. This procedure was iteratively applied. The first row in Figure 3C presents main class averages after 5 iterative alignment cycles of images. Particles have a "donut"-like shape with approximately 150 Å outer diameter and 60 Å inner diameter respectively. The donuts display protruding domains with about 3.4 Å long. The best symmetry related to class averages corresponds to C3 and C6 symmetries. The unsymmetrized views of class averages and their corresponding C3,C6 symmetrized views are shown in Figure 3C. Apart from the first class average which presents some differences between the C3 and C6 views, all C3 symmetrized views display a 6 fold organization which is highly consistent with their related C6 symmetry view and their initial view. This strongly suggests that Sak3 adopts a 6 fold symmetry in which the ring is consist of six domains while protruding domains are located at their interfaces.

ORF35 N- and C-terminal domains. In order to investigate the minimal secondary structure elements necessary for the oligomerization and the functional activity, we

have produced truncated mutants of Sak3. On the basis of secondary structure prediction, we have cloned N and C terminal domains separately generating two N terminal mutants (1-159 and 1-176) and one C terminal mutant (159-207). All these mutants were successfully produced and were soluble up to several mg/ml, which makes them good candidates to perform biophysical and functional studies. The characterization of the oligomeric state by Blue Native PAGE and HPLC-SEC shows that the two N terminal mutants (1-159 and 1-176) have a similar behaviour in forming oligomeric structures, while C terminal domain (159-207) is present in a monomeric state (data not shown).

DNA Binding studies

Gel retardation Assays - It was previously shown by Ploquin and colleagues [10] that phage ul36 Sak binds ssDNA better than dsDNA like RAD52. Despite the lack of primary sequence homology with phage ul36 Sak and SSAPs in general, the DNA binding behaviour of Sak3 is similar. In gel retardation assays we have seen that the amount of ssDNA or dsDNA bound to Sak3 increase upon increasing the concentration of Sak protein. Complex formation with ssDNA (Fig 4A) takes place at lower molar ratio than dsDNA (Fig 4B).

The Sak3 of lactococcal phage p2 bound preferentially to ssDNA over dsDNA (Fig 4A and B). Experiments performed by increasing the DNA molar ratio and keeping the Sak concentration constant also confirm that Sak3 form a stable complex with ssDNA and dsDNA (data not shown). Similar experiments performed with Sak3 mutants show that the N terminal domain binds DNA similarly to Sak3, but at higher concentration than the full-length protein. The C terminal domain does not bind DNA, demonstrating that it is not essential for binding function (Fig 4A and B).

SPR experiments - In order to confirm the qualitative results of gel shift assays and to calculate dissociation constant for the protein-DNA binding, we have performed Surface Plasmon Resonance experiments. We have coated ssDNA and dsDNA on a Streptavidine chip and then we performed a kinetic analysis. Our results show that Sak3 binds ssDNA and dsDNA with equal affinity. Excepted for the C-terminal domain (160-207), which does not show any binding to ssDNA and dsDNA, all other

mutants tested (N-term 1-159, N-term 1-176) bind ssDNA and dsDNA with the same pattern as the full-length protein. In all cases, the association rate constant k_{on} is 4-8 10^4 $M^{-1}s^{-1}$. The dissociation rate constant is very slow and very difficult to estimate with precision (Fig.5). Our efforts to improve the determination by experimentally recording the dissociation for hours, and trying local fitting or simulation, were also unsuccessful. Based on the saturation behaviour of the binding kinetic at 150 nM of Sak3, we can roughly estimate the $K_{diss} \leq 15$ nM. An average value of $k_{on} = 5$ 10^4 $M^{-1}s^{-1}$ gives an estimation of k_{off} at 7.5 $10^{-4}s^{-1}$.

SPR experiments with Sak3 mutants confirm that the N terminal mutants have the same behaviour as Sak3, while the C terminal domain does not bind DNA as previously observed in gel shifts.

AFM imaging of DNA-Sak3 complexes

To further characterize the interaction of Sak3 oligomers with single and double stranded DNA, we have employed atomic force microscopy imaging. From a preliminary inspection of Sak3 oligomers deposited onto freshly cleaved mica and imaged in air with the AFM, we found that adsorption onto the negatively charged mica caused disaggregation of the oligomers (data not shown). Disaggregation was completely prevented when the oligomers were crosslinked with 0.11% glutaraldehyde [36]. Under these conditions Sak3 oligomers were imaged as uniform particles with dimensions compatible with their molecular weight. For this reason, Sak3-DNA complexes used for AFM imaging were crosslinked with glutaraldehyde before deposition.

Figure 6A shows that in the presence of M13 ssDNA most of the Sak3 particles are bound to the nucleic acid confirming the gel shift assays results. Due to the folded structure of the M13 ssDNA it is not possible to determine the mode of protein-DNA of interaction of these complexes but it appears clear that most of the oligomers are seen bound to ssDNA. Because it has been hypothesized that Sak3 is involved in double strand break repair mechanisms [1] we have performed experiments with different DNA substrates that mimic double strand breaks. In particular we have analyzed the Sak3 binding behaviours to linear dsDNA fragments having different termini or harbouring a 20 nt single stranded gap. The data show that Sak3 has little

or no affinity for blunt-ended dsDNA fragments either in their unphosphorylated or phosphorylated form (Figure 6B,C). Conversely, dsDNA fragments with termini having one of four nt overhang show a preferential binding of Sak3 to the DNA ends (Figure 6D,E). In the case of DNA harbouring a 20 nt single-stranded gap and bearing blunt ends, it was observed that Sak3 binds preferentially the single-stranded gap over dsDNA or blunt ends (Figure 6F).

A quantification of these observations is shown in figure 6G, which reports the percentage of end-bound and unbound Sak3 oligomers. For each DNA template more than 1000 Sak3 oligomers were scored. In the case of blunt-ended DNA, only 8 or 15% (unphosphorylated or phosphorylated 5' ends respectively) of the total Sak3 oligomers were end-bound, whereas, in the case of single-stranded overhangs, 57 or 63% (one or four nt overhangs respectively) of the total Sak3 oligomers were end-bound. In addition, 31% of the Sak3 oligomers bind a single-stranded gap comprised between two long stretches of dsDNA. This reduced binding may be due to the reduced accessibility of Sak3 oligomers versus ssDNA of a gap compared to ssDNA of the DNA termini.

Functional activity of ORF35-Sak3

Annealing and strand exchange - With in view to determine that Sak3 is a SSAP, $ORF35_{p2}$ was tested for its capacity to anneal complementary oligonucleotides (data not shown) and to stimulate the strand exchange activity of RecA (Fig.7) as shown for other SSAPs like RAD52 and Sak UL36 [10; 17; 18].

Surprisingly, Sak3 do not anneal complementary oligonucleotides and inhibits instead of stimulate the strand exchange activity of RecA (Fig.7). Stimulation of the strand exchange reaction mediated by RecA was also performed with Sak3 mutants, but no significant differences where found showing that probably both N and C terminal domain are necessary for RecA inhibition (data not shown).

ORF35 interaction studies with SSB and RecA - Despite the lacking of typical SSAP functions, we were interested in understanding the involvement of Sak3 in homologous recombination activities by analysing its interaction with other recombination proteins by SPR. The interaction between Sak and SSB was studied

by fixing SSB protein on a chip CM5 and passing as analyte increasing concentrations of Sak3 and of ssDNA-Sak3 complex. The k_{on} and the k_{off} are particularly fast indicating the formation of transitory complexes. The K_{diss} values were calculated on the equilibrium levels at different concentration of analyte.

Sak3 and SSB protein interact with a micromolar affinity (Table 1 and Fig.8A), in line with results obtained for other interaction with SSBs [40]. The Sak3-ssDNA pre-formed complex does not bind to SSB protein covalently coupled. The formation of the complex with ssDNA prevents the interaction of both the component of the complex with SSB, probably because of steric hindrance. *L.lactis* RecA was also fixed as ligand on a CM5 chip and the interaction with Sak3 and its mutants was analysed. RecA interacts with Sak3; the affinity is 5-6 times higher than that of SSB (Table 1 and Fig.8B).

We have also tested the interactions of Sak3 mutants with SSB and RecA. We could not evidence any interaction of the N terminal portion with SSB and RecA (Tab.1); on the contrary the interaction with the C terminal region is detectable although non calculable because of the very low signal. Indeed the low molecular weight of C terminal mutant (5.6 kDa) may possibly account for this result. Nevertheless, our results clearly indicate that the N-terminus is not involved in the interaction with homologous recombination partners, and that the C-terminal is probably the interaction site.

DISCUSSION

In this study, we performed a biochemical and structural characterization of ORF35, renamed Sak3 for its sensibility to the Antiviral Abortive mechanism AbiK [1]. Sak3 was considered to belong to the Sak proteins family, among which Single-strand Annealing Proteins (SSAPs) are found, and to be involved in homologous recombination due to its localization upstream of a gene coding for a single-strand binding protein [2].

As starting hypothesis, we investigated the possibility that Sak3 is a SSAP as Rad52 by studying its oligomerization state and structural features, its DNA binding and its

functional activities as annealing of complementary strands and stimulation of strand exchange of RecA. Our preliminary results on the oligomerization state were performed using HPLC-SEC with MALS/UV/RI (Wyatt technology) and Blue Native PAGE gels and showed that Sak has an oligomeric structure. Subsequently these results were confirmed by the CryoEM structure obtained from Sak3 crystals. In particular CryoEM shows that ORF35 shows a ring structure, typical of the superfamily containing the conserved RecA nucleotide-binding core, diffused in DNA recombination, replication and transcription proteins [21].The 6-fold symmetry of ORF35 rings remembers the ring structure of hexameric helicases such as bacteriophage T7 gp4 [23], SV40 large T antigene [24] and DnaB [25].

DNA binding properties were also tested by gel mobility shift assays (EMSA), atomic force microscopy (AFM) and surface plasmon resonance (SPR). EMSA results showed that Sak3 binds DNA in an unspecific sequence manner with a preferential binding to ssDNA over dsDNA. On the other hand, in SPR experiments Sak3 binds ssDNA and dsDNA with equal affinity. Using both technique, the C-terminal domain (160-207) does not bind DNA, while N terminal mutants (N-term 1-159, N-term 1-176) bind ssDNA and dsDNA with the same pattern as the full length protein, showing that the DNA binding domain is located in the first 159 residues of the protein as in the case of Rad52 [19]. The analysis of Sak3-DNA complexes by atomic force microscopy also confirmed the results previously obtained. The use of different DNA templates with sticky or blunt ends, that mimic double strand breaks, showed that Sak3 binds preferentially the single-stranded gap over dsDNA or blunt ends. In a similar manner Rad52 binds single-strand and tailed duplex DNA molecules via precise interactions with the terminal base, in particular with 5' or 3' terminus of the ssDNA [20]. Beta proteins of phage lambda also have a reduced filament formation on blunt-ended dsDNA [13]. Further SSAPs, most of protein essential in recombination/replication processes like helicases [23], topoisomerases [41] and recombinases [42] have a similar behaviour in binding preferentially ssDNA.

We suspected that Sak3 might be a SSAP, but, surprisingly, Sak3 did not anneal complementary oligonucleotides and inhibited instead of stimulate the strand exchange activity of RecA. This result indicates that ORF35 exhibits a clear anti-recombinase activity already detected in Srs2 and Sgs1 helicases of *S.cerevisiae* to

prevent undesirable homologous recombination events [43]. Recent studies have found that both helicases and topoisomerases are involved in antirecombinase activity and dissolution of the homologous recombination intermediates [44].

After the discovery that Sak3 does not show SSAP features and in order to determine if it is involved in homologous recombination, the relationship of Sak3 with putative homologous recombination partners was studied using SPR. A micromolar affinity between Sak3 and SSB protein was found as in other interactions between SSBs and partners [40]. Our results indicate that Sak-binding site on SSB may be located close to the DNA-binding domain, excluding DNA-SSB interaction as in RAD52 [45]. The N-terminal domain is not involved in the interaction with homologous recombination partners, while the C-terminal domain is probably the interaction site. It is therefore likely that SSB recrutes Sak3, as shown for different prokaryotic SSBs and eukaryotic RPAs [18], to carry out some homologous recombination function, the nature of which is still unknown. These unknown functions involve the interaction of Sak3 with RecA as shown in SPR experiments and in inhibition of the RecA strand exchange reaction.

A more accurate investigation among the members of DUF1071 superfamily, to whom Sak3 belongs, showed that some lactococcal phage *orf$_s$* were located in a similar position to *sak3* in phage p2 genome. In particular *orf11* of phage TP901-1 and *orf14* of both bIL285 and Tuc2009 phages are located in early transcribed region of the genome near genes encoding for replication proteins [1]. *orf11* is located upstream a single-strand DNA binding protein analogue (*orf12*) and the putative replication initiation protein (*orf13*) [46]. In a similar manner a replication cassette containing open reading frames, whose deduced proteins exhibited similarities to proteins known to be involved in DNA replication and modification, was found in temperate lactococcal phage Tuc2009; in particular an SSB protein (*orf15*) a replisome organizer protein (*orf16*), a methylase (*orf18*) and a topoisomerase I (*orf14*) genes were identified [47]. The deduced protein product of *orf14* shows a low degree of similarity to topoisomerase I from two different Mycoplasma species and also shows significant similarity to *orf35* of phage p2 (*sak3*), *orf14* of phage bIL170 and *orf11* of phage TP901-1 (Fig 2B) [47]. These similarities strongly suggest that

Sak3 might have an additional function that could explain why Sak3 cannot be considered a SSAP.

Such additional functions could be associated to proteins that belong a topoisomerase I activity. Among these proteins an interested role is played by the reverse gyrase that comprises a type IA topoisomerase fused to a helicase-like domain. In particular Sak3 showed some features that can be attributed both to a topoisomerase I and to a helicase. A topoisomerse I activity might be retrieved in the higher affinity for ssDNA than for dsDNA and in the initial binding to short ssDNA regions for the initiation of the activity [48]. On the other hand, an helicase activity was found in the EM ring structure and in the capacity to inhibit RecA strand exchange , typical of human Srs2 or yeast Sgs1 helicases that block illegitimate recombination [44]. Reverse gyrase interacts also with SSB like Sak3 and might be involved in the cell response to DNA damage. These features might enhance the idea that Sak3 has a reverse gyrase similar behaviour. This supposition is enhanced by its primary sequence analysis that show some helicase/topoisomerase motifs as the conserved glutamine located upstream of Walker A motif I (<u>GKT/S</u>), the unusual Walker B motifs II (<u>DD</u>GK) (Fig.2B, red asterisk) and the presence of putative active site tyrosine in the C terminal domain [49]. A leucine zipper motif usually present in many gene regulatory proteins was also found between residues 151 and 172 (Fig.2B, red cassette). It is likely that evolution led to the final production of a multifunctional chimeric protein that is involved in phage DNA maturation. The involvement in DNA maturation was shown thanks to the AbiK+ mutant in which the phage DNA maturation was blocked by the presence of AbiK and was restored with mutations on the Sak3 gene [1; 7]. However, questions remain open concerning the exact function and role of Sak3 in phage replication/recombination process and in related resistance mechanisms.

REFERENCES

1. Bouchard, J. D. & Moineau, S. (2004). Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J Bacteriol* **186**, 3649-52.
2. Scaltriti, E., Tegoni, M, Rivetti, C, Launay, H, Masson, J-Y, Tremblay, D, Magadan, A H, Moineau, S, Ramoni, R, Lichière, J, Campanacci, V, Cambillau, C and Ortiz-Lombardía, M. (2009). Crystal Structure and Characterization of a New Type of Single Stranded DNA Binding Protein from the Lactococcal phage p2. *Molecular Microbiology* **Submitted June 12, 2009**.
3. Moineau, S. (1999). Applications of phage resistance in lactic acid bacteria. *Antonie Van Leeuwenhoek* **76**, 377-82.
4. Coffey, A. & Ross, R. P. (2002). Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. *Antonie Van Leeuwenhoek* **82**, 303-21.
5. Emond, E., Holler, B. J., Boucher, I., Vandenbergh, P. A., Vedamuthu, E. R., Kondo, J. K. & Moineau, S. (1997). Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from Lactococcus lactis. *Appl Environ Microbiol* **63**, 1274-83.
6. Fortier, L. C., Bouchard, J. D. & Moineau, S. (2005). Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. *J Bacteriol* **187**, 3721-30.
7. Bouchard, J. D. & Moineau, S. (2000). Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. *Virology* **270**, 65-75.
8. Bouchard, J. D., Dion, E., Bissonnette, F. & Moineau, S. (2002). Characterization of the two-component abortive phage infection mechanism AbiT from Lactococcus lactis. *J Bacteriol* **184**, 6325-32.
9. Moineau, S., Walker, S. A., Vedamuthu, E. R. & Vandenbergh, P. A. (1995). Cloning and sequencing of LlaDCHI [corrected] restriction/modification genes from Lactococcus lactis and relatedness of this system to the Streptococcus pneumoniae DpnII system. *Appl Environ Microbiol* **61**, 2193-202.
10. Ploquin, M., Bransi, A., Paquet, E. R., Stasiak, A. Z., Stasiak, A., Yu, X., Cieslinska, A. M., Egelman, E. H., Moineau, S. & Masson, J. Y. (2008). Functional and structural basis for a bacteriophage homolog of human RAD52. *Curr Biol* **18**, 1142-6.
11. Iyer, L. M., Koonin, E. V. & Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**, 8.
12. Ayora, S., Missich, R., Mesa, P., Lurz, R., Yang, S., Egelman, E. H. & Alonso, J. C. (2002). Homologous-pairing activity of the Bacillus subtilis bacteriophage SPP1 replication protein G35P. *J Biol Chem* **277**, 35969-79.
13. Passy, S. I., Yu, X., Li, Z., Radding, C. M., Masson, J. Y., West, S. C. & Egelman, E. H. (1999). Human Dmc1 protein binds DNA as an octameric ring. *Proc Natl Acad Sci U S A* **96**, 10684-8.
14. Mythili, E., Kumar, K. A. & Muniyappa, K. (1996). Characterization of the DNA-binding domain of beta protein, a component of phage lambda red-pathway, by UV catalyzed cross-linking. *Gene* **182**, 81-7.

15. Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**, 751-813, table of contents.

16. Shen, Z., Cloud, K. G., Chen, D. J. & Park, M. S. (1996). Specific interactions between the human RAD51 and RAD52 proteins. *J Biol Chem* **271**, 148-52.

17. Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci U S A* **93**, 10729-34.

18. Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S. & Ogawa, T. (1998). Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* **3**, 145-56.

19. Singleton, M. R., Wentzell, L. M., Liu, Y., West, S. C. & Wigley, D. B. (2002). Structure of the single-strand annealing domain of human RAD52 protein. *Proc Natl Acad Sci U S A* **99**, 13492-7.

20. Parsons, C. A., Baumann, P., Van Dyck, E. & West, S. C. (2000). Precise binding of single-stranded DNA termini by human RAD52 protein. *EMBO J* **19**, 4175-81.

21. Stasiak, A. Z., Larquet, E., Stasiak, A., Muller, S., Engel, A., Van Dyck, E., West, S. C. & Egelman, E. H. (2000). The human Rad52 protein exists as a heptameric ring. *Curr Biol* **10**, 337-40.

22. Guenther, B., Onrust, R., Sali, A., O'Donnell, M. & Kuriyan, J. (1997). Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III. *Cell* **91**, 335-45.

23. Egelman, E. H., Yu, X., Wild, R., Hingorani, M. M. & Patel, S. S. (1995). Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proc Natl Acad Sci U S A* **92**, 3869-73.

24. San Martin, M. C., Gruss, C. & Carazo, J. M. (1997). Six molecules of SV40 large T antigen assemble in a propeller-shaped particle around a channel. *J Mol Biol* **268**, 15-20.

25. Yang, S., Yu, X., VanLoock, M. S., Jezewska, M. J., Bujalowski, W. & Egelman, E. H. (2002). Flexibility of the rings: structural asymmetry in the DnaB hexameric helicase. *J Mol Biol* **321**, 839-49.

26. Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T. & Yokoyama, S. (2002). Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol Cell* **10**, 359-71.

27. Chang, Y. C., Lo, Y. H., Lee, M. H., Leng, C. H., Hu, S. M., Chang, C. S. & Wang, T. F. (2005). Molecular visualization of the yeast Dmc1 protein ring and Dmc1-ssDNA nucleoprotein complex. *Biochemistry* **44**, 6052-8.

28. D'Souza, J., Dharmadhikari, J.A., Dharmadhikari, A.K., Navadgi, V., Mathur, D. & Rao, B. (2006). Human Rad52 binding renders ssDNA unfolded: image and contour length analyses by atomic force microscopy. *Current Sciance* **91**, 12-25.

29. VanLoock, M. S., Chen, Y. J., Yu, X., Patel, S. S. & Egelman, E. H. (2001). The primase active site is on the outside of the hexameric bacteriophage T7 gene 4 helicase-primase ring. *J Mol Biol* **311**, 951-6.

30. Eggleston, A. K., Mitchell, A. H. & West, S. C. (1997). In vitro reconstitution of the late steps of genetic recombination in E. coli. *Cell* **89**, 607-17.

31. Sulzenbacher, G., Gruez, A., Roig-Zamboni, V., Spinelli, S., Valencia, C., Pagot, F., Vincentelli, R., Bignon, C., Salomoni, A., Grisel, S., Maurin, D., Huyghe, C., Johansson, K., Grassick, A., Roussel, A., Bourne, Y., Perrier, S., Miallau, L., Cantau, P., Blanc, E., Genevois, M., Grossi, A., Zenatti, A., Campanacci, V. & Cambillau, C. (2002). A medium-throughput crystallization approach. *Acta Crystallogr D Biol Crystallogr* **58**, 2109-15.

32. Albertini, A. A., Clapier, C. R., Wernimont, A. K., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. (2007). Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. *J Struct Biol* **158**, 129-33.

33. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J Struct Biol* **116**, 17-24.

34. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82-97.

35. Rivetti, C., Walker, C. & Bustamante, C. (1998). Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J Mol Biol* **280**, 41-59.

36. Dodson, M. E., H. (1991). Electron Microscopy of Protein-DNA complexes. *Methods Enzymol.* **208**, 168-196.

37. Ranatunga, W., Jackson, D., Lloyd, J. A., Forget, A. L., Knight, K. L. & Borgstahl, G. E. (2001). Human RAD52 exhibits two modes of self-association. *J Biol Chem* **276**, 15876-80.

38. Soding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951-60.

39. Soding, J., Biegert, A. & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-8.

40. Reddy, M. S., Guhan, N. & Muniyappa, K. (2001). Characterization of single-stranded DNA-binding proteins from Mycobacteria. The carboxyl-terminal of domain of SSB is essential for stable association with its cognate RecA protein. *J Biol Chem* **276**, 45959-68.

41. Champoux, J. J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70**, 369-413.

42. Lusetti, S. L. & Cox, M. M. (2002). The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* **71**, 71-100.

43. Klein, H. L. (2000). A radical solution to death. *Nat Genet* **25**, 132-4.

44. Sung, P. & Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* **7**, 739-50.

45. Jackson, D., Dhar, K., Wahl, J. K., Wold, M. S. & Borgstahl, G. E. (2002). Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA. *J Mol Biol* **321**, 133-48.

46. Ostergaard, S., Brondsted, L. & Vogensen, F. K. (2001). Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. *Appl Environ Microbiol* **67**, 774-81.

47. McGrath, S., Seegers, J. F., Fitzgerald, G. F. & van Sinderen, D. (1999). Molecular characterization of a phage-encoded resistance system in Lactococcus lactis. *Appl Environ Microbiol* **65**, 1891-9.

48. Perugino, G., Valenti, A., D'Amaro, A., Rossi, M. & Ciaramella, M. (2009). Reverse gyrase and genome stability in hyperthermophilic organisms. *Biochem Soc Trans* **37**, 69-73.

49. Bouthier de la Tour, C., Amrani, L., Cossard, R., Neuman, K. C., Serre, M. C. & Duguet, M. (2008). Mutational analysis of the helicase-like domain of Thermotoga maritima reverse gyrase. *J Biol Chem* **283**, 27395-402.

50. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402.

51. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics* **20**, 426-7.

Figure Legends

Fig .1. Characterization of ORF35-Sak3 wt oligomeric state. A) Blue Native PAGE gel. Lane 1:MW Standard, Lane 2 Sak3 and Lane 3 Sak3-ssDNA complex. B) HPLC-SEC profiles of Sak3.

Fig. 2. Sequence alignments of Sak3. A: Graphical Hitlist of Sak3 alignment with HHpred (Homology detection & structure prediction by HMM-HMM comparison http://toolkit.tuebingen.mpg.de/hhpred [38; 39]. B. Multiple Alignment of Sak3 with some members of the DUF1071 group performed by PSI-PRED [50] and displayed using Jalview [51].

Fig. 3.. Negatively stained Sak3 observed by TEM. A) Crystal of Sak3 which were collected and crushed, which served for observation of Sak3 by TEM after negative stain. Scale bar 100 μm. B)Image of negatively stained Sak3. The arrows show some ring structures present in the image. Scale bar, 30 nm. C)Class averages obtained after images alignment cycles and their related C3 and C6 symmetrized views. Each image corresponds to 315 Å X 315 Å

Fig. 4. Polyacrylamide Electrophoretic Mobility Shift Assays of Sak3 and mutants and ssDNA (A) or dsDNA (B).

Fig. 5 Surface Plasmon Resonance. Sensorgram of Sak3-ssDNA complex. Reponses Units (RU) are represented on Y axis, while Time (s) is represented on X axis.

Fig. 6. AFM images of Sak3-DNA complexes. A) Image of Sak3 bound to M13 ssDNA (inset: protein free M13 ssDNA). Notice the paucity of unbound oligomers over the mica surface. Sak3 oligomers in the presence of DNA with phosphorylated (B) and unphosphorylated (C) blunt ends. D) Sak3 oligomers bound to DNA fragments with 1 nt overgang at the 3' termini. E) Sak3 oligomers bound to DNA fragments with 4 nt overgang at the 3' termini. F) Sak3 oligomers bound to a 20 nt single stranded gap located in the middle of the fragment. G) Bar plot of the fraction of bound (blue) and unbound (magenta) Sak3 oligomers to different DNA templates. All images represent a scan size of 1 mm.

Fig. 7. Inhibition of RecA strand exchange by Sak3. Gel on the left side shows a titration of RecA (0.75-6 μM), while gel on the right side shows reaction with constant RecA.(2μM, lane 2) and increasing concentration of Sak3 p2 (0.1-2 μM)

Tables:

Tab.1 CM5 chip results of Sak3 interaction. ND, determination not possibile. W, low interaction.

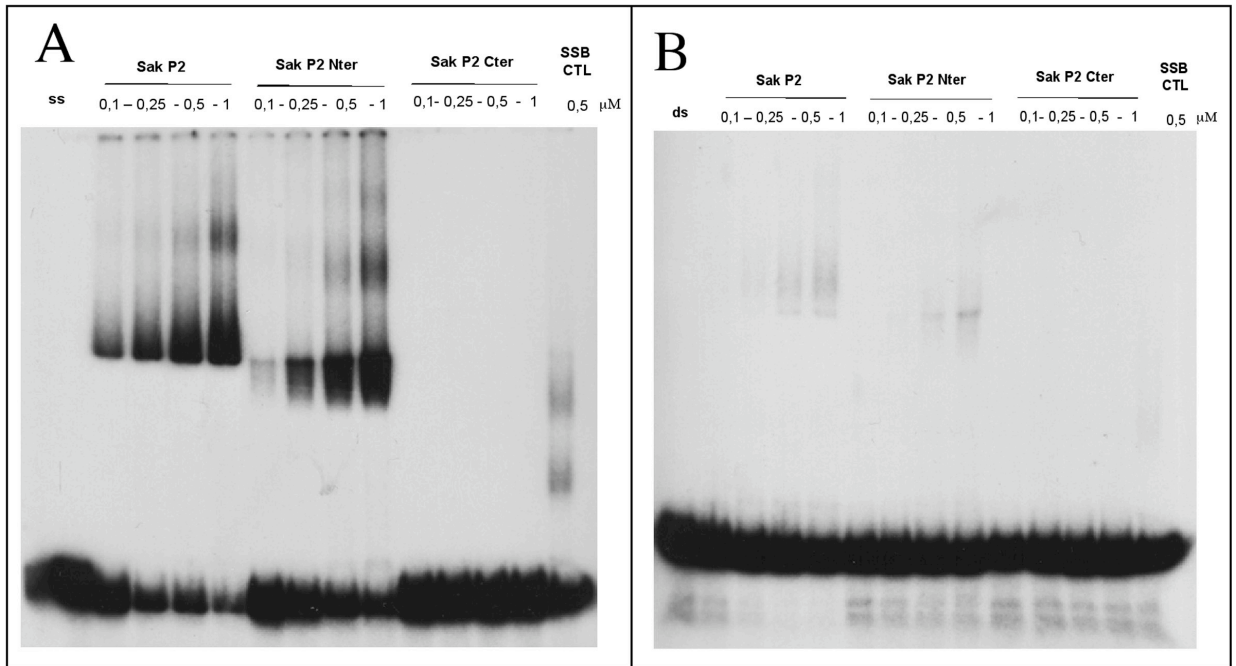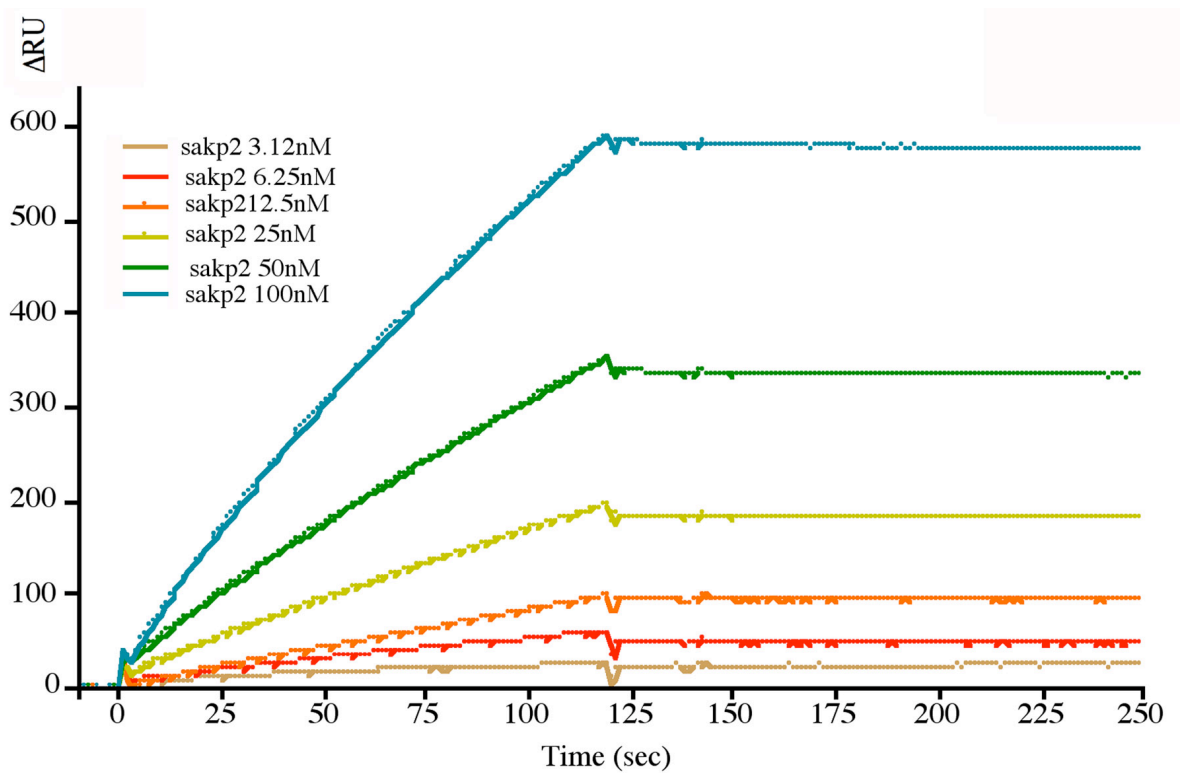| Ligand	Analyte	Kd (μM)
p2 SSB	Sak3	1.7±0.18
	Sak3 N-terminus (1-159)	ND
	Sak3 N-terminus (1-176)	ND
	Sak3 C-terminus (160-207)	W
	Sak3-ssDNA	W
RecA	Sak3	0.12±0.02
	Sak3 N-terminus (1-159)	ND
	Sak3 N-terminus (1-176)	ND
	Sak3 C-terminus (160-207)	W

Figures:

Figure 1
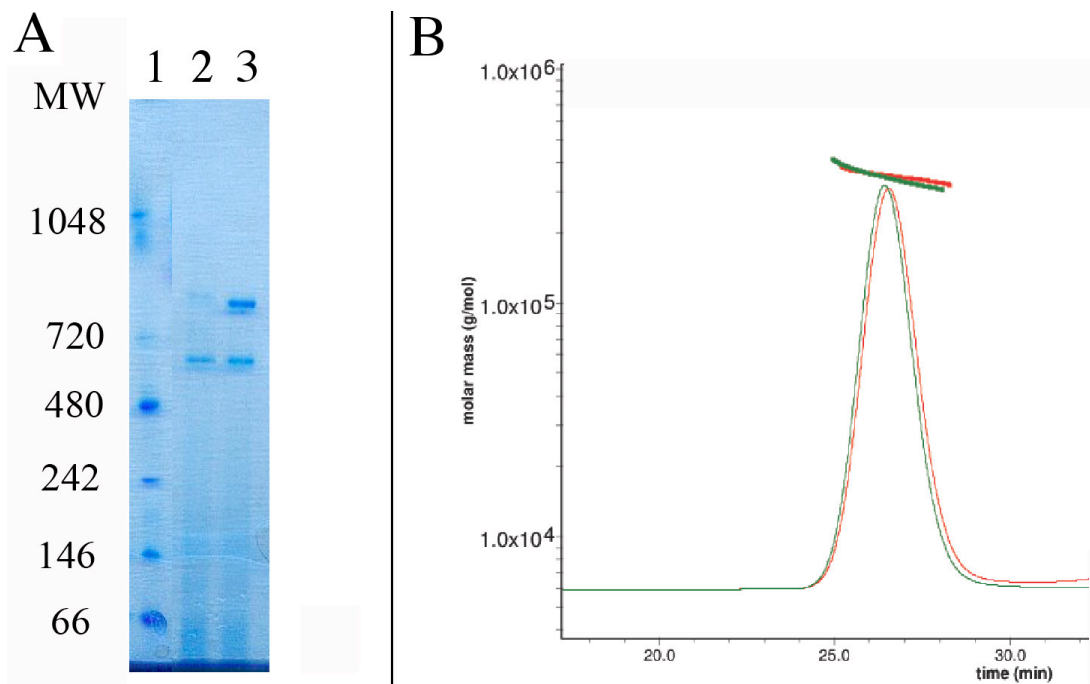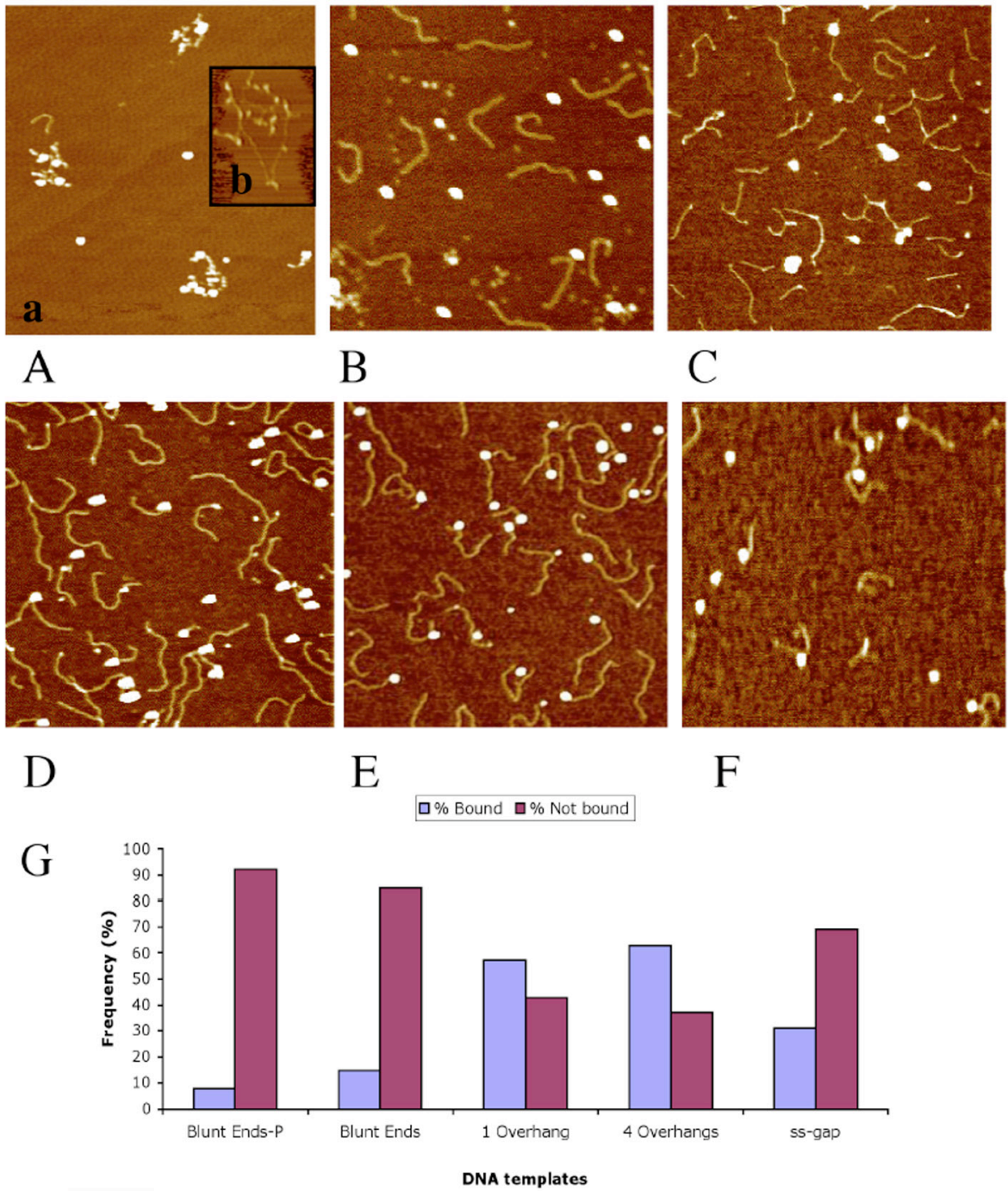
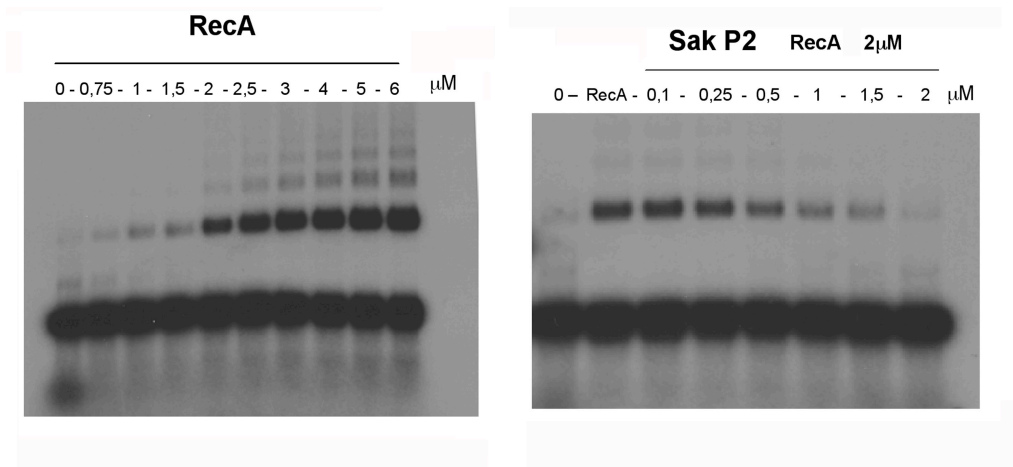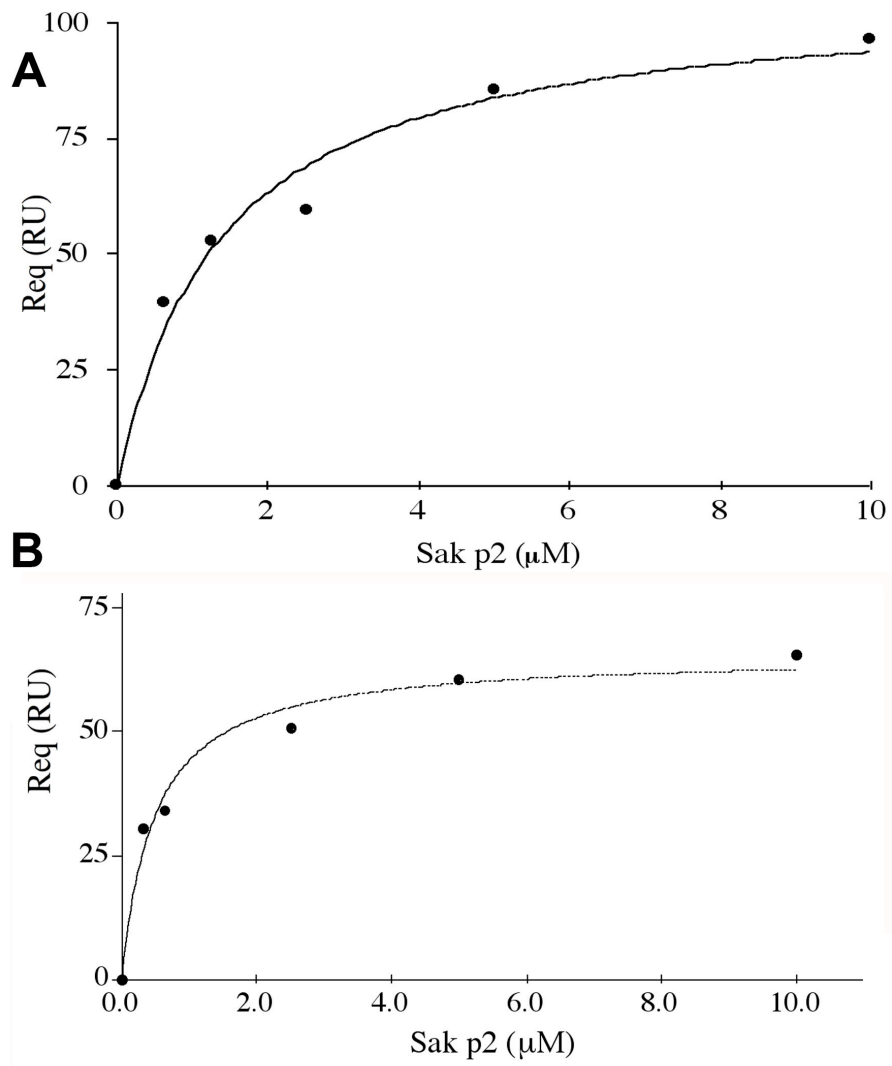**Figure 2**
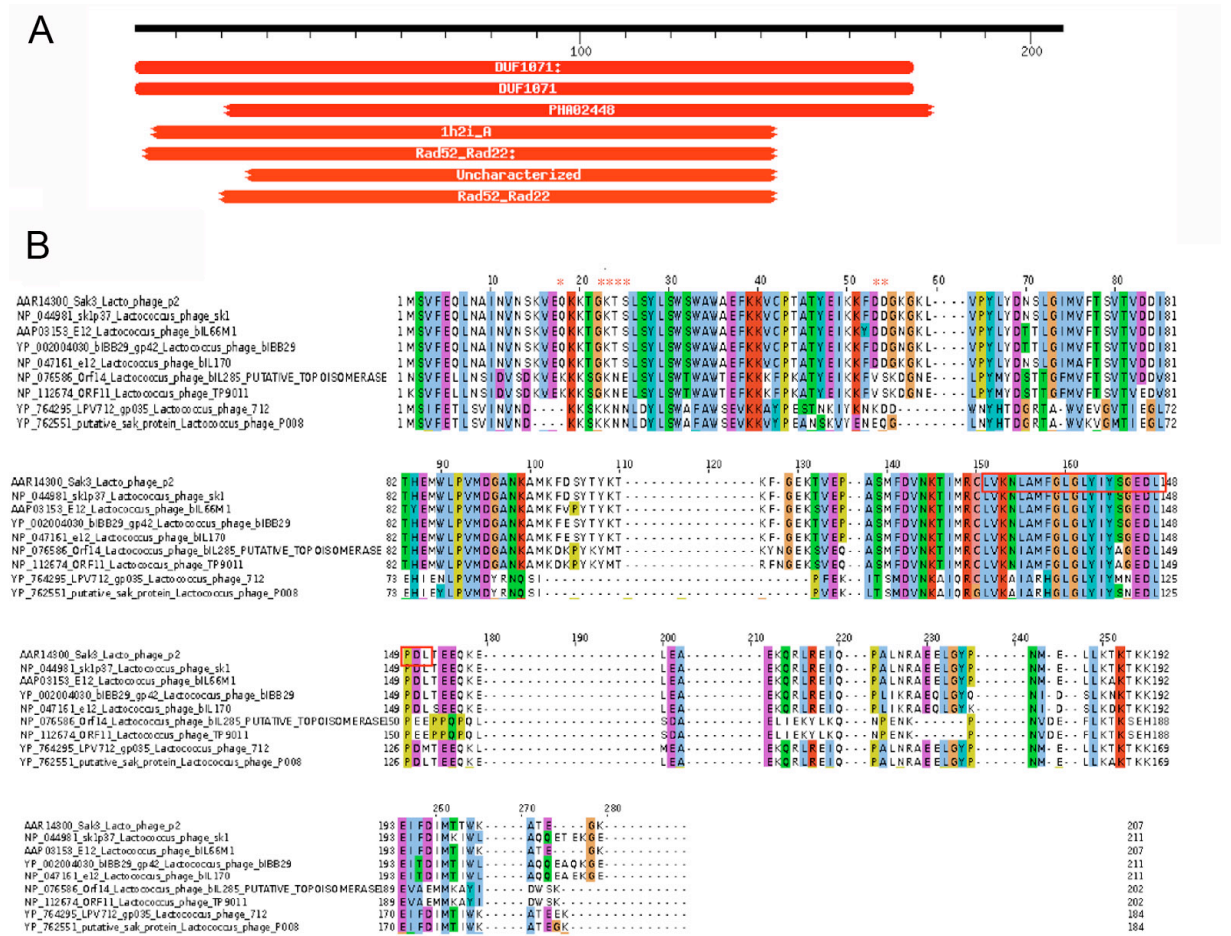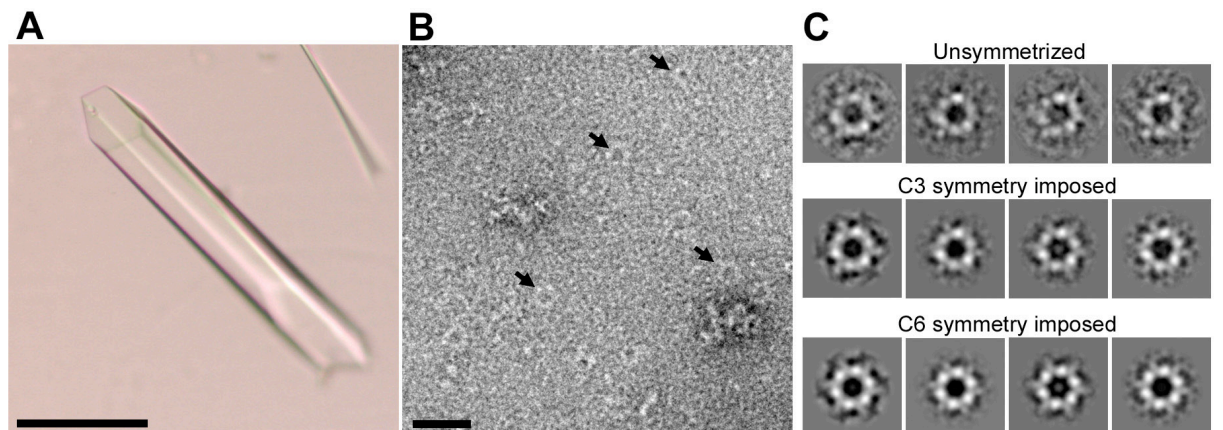
Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8

2.1.3 Conclusions on Sak proteins

From the analysis of Sak wild type and A92S of phage UL36 (P335 group) performed by our colleagues of the University of Laval, it turns out that the Sak mutant binds RNA more efficiently than the wild-type protein. The latter, on the other hand, binds with higher affinity DNA than RNA when compared to the A92S mutant. Sak wild type shows also a higher efficiency into the assembling of RNA/DNA hybrids than Sak mutant. As Sak wild type, Rad52 can bind RNA, but shows higher affinity for DNA. In addition, Rad52 can also form complexes with RNA/DNA hybrids, but the affinity is lower than in the case of Sak wild type (J.Y. Masson, personal communication). The enhanced RNA binding capacity of the Sak A92S mutant and the equivalent binding affinity of Sak wild type and mutant with respect to ssDNA and dsDNA , led our colleagues to speculate that the binding of RNA by Sak A92S mutant could interfere with the AbiK antiviral activity and provide insensitivity to this bacterial mechanism of defense (J.Y. Masson, personal communication). This behavior may be related to the experimental evidence showing that the replication of phage ul36 genome was not detected in the presence of AbiK (Boucher et al, 2000). Nevertheless, the relationship between Sak insensitivity and AbiK's mechanism is still not clear and, to date, it is difficult to formulate some hypothesis on this particular point. Additional information on AbiK mechanism might arise from studies on Sak3 protein from phage p2 (936 group), that does not show significant similarities to Sak and/or Rad52. In fact, Sak3 has a primary sequence related to phage putative topoisomerase and has some structural features typical of helicases and of other hexameric proteins involved in replication, recombination and maturation of DNA. Sak3 is also involved in sensitivity to AbiK *via* a diverse mechanism than Sak (Boucher et al, 2000). In particular, in Abi+ cells infected by phage p2, the processing of the phage DNA toward its mature form is severely reduced, despite the presence of DNA concatemers found at later stages of infection and a few head-like structures observed around a phage DNA molecules (Boucher et al, 2000). The evidence that concatenated DNA of 936 phages is not efficiently processed into mature genome suggests that AbiK halted the phage DNA

maturation and packaging, possibly as the result of defective concatemeric DNA intermediates that are not suitable substrates for the synthesis of mature phage DNA (Boucher et al, 2000). In bacteriophage lambda, the catalytic terminase subunits gpA is responsible for the maturation of the genome before it has been packed and translocated into the capsid and therefore exhibits the following catalytic activities: site specific endonuclease, strand separation (helicase) capacity that is required to excise individual genomes from the concatemer (DNA maturation activity), DNA traslocase (packaging) and ATPase activity (Ortega et al, 2007). Despite the ATPase function, also the bacteriophage ϕ29 DNA packaging ATPase gene product 16 (gp16) has an hexameric ring structure similar to Sak3 and has been demonstrated to bind a prohead RNA (pRNA) that is an essential packaging motor component (Koti et al, 2008). Similarly, our Sak3 might also have an ATPase activity related to DNA maturation and packaging.

Another interesting point is that the positive supercoils introduced by Rad52 into dsDNA during replication/recombination processes can be removed by a topoisomerase I, an enzyme which can relax supercoils in DNA and that, in terms of primary sequence, can be assimilated to Sak3 (Kagawa et al 2008). It is possible that specific evolutionary pressures led to the selection of a multifunctional chimeric protein involved in one or both of these functions that, at a certain extent, could be biologically related. In particular, as the large size of the *abiK* gene supports the possibility of a multifunctional protein with multiple enzymatic activities (Gething, 1997), Sak and Sak3 can both have a dissimilar, but related function that can explain the different consequences on phage life cycle perturbation in the presence of Abi+.

2.2 SSB protein of L.lactis bacteriophage p2

As previously mentioned, the partial characterisation of a SSB protein, encoded by the *L.lactis* phage bIL67 *orf14,* has allowed to define a novel cluster of phage SSB proteins. Nevertheless, the 3D structure and the involvement of these proteins in homologous recombination mechanism have not yet been determined to date (Szczepanska et al, 2007).

In this article we characterize the product of the *orf34* from *L.lactis* phage p2, which was considered as a candidate of a single stranded DNA binding protein (SSB) by its localization, downstream of a gene coding for a putative single-strand annealing protein (Sak3) (Bouchard and Moineau, 2004). We determined ORF34p2 expression level during phage infection, and we have characterized its binding properties to ssDNA (Gel-shift assays, surface plasmon resonance and atomic force microscopy), its involvement in homologous recombination processes (strand exchange stimulation by RecA). We also determined the crystal structure of ORF34p2 and compared it to other SSBs OB folds. This multi-technique approach provides a clear evidence that ORF34p2 is a *bona fide* SSB with specific structural and biochemical properties that distinguish it from the others SSBs

2.2.1 *Article* "Crystal Structure and Characterization of a New Type of Single Stranded DNA Binding Protein from the Lactococcal phage p2"

Erika Scaltriti[1,2], Mariella Tegoni[1], Claudio Rivetti[3], Hélène Launay [4] , Jean-Yves Masson[4], Denise Tremblay[5,6,7], Alfonso H. Magadan[4,5,6,7], Sylvain Moineau[5,6,7] , Roberto Ramoni[2], Julie Lichière[1], Valérie Campanacci[1], Christian Cambillau[1]* and Miguel Ortiz-Lombardía[1]*

[1]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, case 932, 13288 Marseille cedex 09, France;

[2]Istituto di Biochimica Veterinaria, Facoltà di Medicina Veterinaria, Università di Parma, Via del Taglio 8, 43100 Parma, Italy.

[3]Istituto di Scienze Biochimiche, Università di Parma, 43100, Parma, Italy.

[4]Genome Stability Laboratory Laval University Cancer Research Center, Hôtel-Dieu de Québec, 9 McMahon, Québec City, Québec, Canada, G1R 2J6.

[5]Groupe de Recherche en Écologie Buccale (GREB), Faculté de Médecine Dentaire;

[6]Félix d'Hérelle Reference Center for Bacterial Viruses;

[7]Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Université Laval, Québec City, Québec, Canada, G1K 7P4.

*Correspondence to:

Christian Cambillau (christian@afmb.univ-mrs.fr)

Tel. +34 491 82 55 90, Fax. +34 491 26 67 20

or

Miguel Ortiz-Lombardía (miguel@afmb.univ-mrs.fr)

Tel. +34 491 82 55 93, Fax. +34 491 26 67 20

Running title: Structure-function of lactococcus phage p2 ORF34

127

Abstract

Lactococcus lactis, a Gram-positive bacterium widely used by the dairy industry, is subject to infection by a diverse population of virulent phages, among which those of the 936 group, including phage p2, are predominant. With the negative impact of phage infection on milk fermentation and the resulting economical burden, the study of the biology of lactococcal phages gains special significance. We decided to characterize the product of the orf34 gene from lactococcus phage p2, which was considered as a candidate single stranded ssDNA binding protein (SSB) due to its localization downstream of a gene coding for a putative single-strand annealing protein. Two-dimensional gel electrophoresis showed that ORF34$_{p2}$ is expressed in large amounts during the early phases of phage infection, suggesting a major role in this process. Gel-shift assays, surface plasmon resonance and atomic force microscopy demonstrated that ORF34$_{p2}$ interacts with ssDNA with nanomolar affinity. We also determined the crystal structure of ORF34$_{p2}$ and showed that it bears a variation of the typical OB-fold of SSBs. Finally, we found that ORF34$_{p2}$ is able to stimulate E. coli RecA-mediated homologous recombination. Altogether, our results confirm ORF34$_{p2}$ as an SSB protein with specific structural and biochemical properties that distinguish it from the host SSB.

Introduction

Single-stranded DNA binding (SSB) proteins are present in all organisms and play an essential role in their growth and metabolic activities. They are involved in multiple pathways of DNA metabolism including replication, recombination and repair in Bacteria, Archaea, Eukarya and viruses (Lohman & Ferrari, 1994; Wold, 1997). In particular, they coat and protect single strand DNA (ssDNA) intermediates and facilitate pairing of homologous strands by preempting DNA secondary structures. SSB proteins can bind ssDNA in a nonspecific and cooperative manner and form filaments that can saturate long stretches of ssDNA (Lohman & Ferrari, 1994). Besides, SSBs can interact with a large number of heterologous proteins and these interactions both regulate and can be regulated by the interaction with ssDNA (Wold, 1997). Under particular conditions, SSBs can stimulate the assembly of recombinases on ssDNA. Moreover, the filaments formed from SSB-recombinase interactions are in general more efficient in strand-exchange reactions than filaments formed by recombinase alone (Sung, 1997). Indeed, the presence of *Eco*SSB permits the formation of a continuous filament of RecA protein by destabilizing the secondary structure of ssDNA (Meyer & Laine, 1990). In *S. cerevisiae*, the SSB protein RPA (Replication Protein A) stimulates the formation of nucleoprotein filaments including the recombinase Rad51 and, in combination with Rad52, stimulates Rad51-mediated DNA strand exchange (Shinohara et al., 1998; Sung, 1997).

From a structural viewpoint SSBs exist as monomeric or multimeric proteins and share a structural domain named OB-fold (oligonucleotide/oligosaccharide binding-fold) involved in nucleic acid recognition. The OB-fold domains are variable in length and functions, mainly because of differences in the length of the loops forming the nucleic acid interaction surfaces. EcoSSB is the reference SSB protein within the OB-fold structural family (Theobald et al., 2003). The structure of *Eco*SSB consists of one α-helix and two three-stranded antiparallel β-sheets that form a semi-open barrel from where two long loops protrude, one of them involved in ssDNA interaction (Murzin, 1993). The crystal structures of *Eco*SSB with (Raghunathan et al., 2000) and without ssDNA (Matsumoto et al., 2000; Raghunathan et al., 1997) confirmed that this protein forms homotetramers. Each

monomer of the *Eco*SSB homotetramer consists of an N-terminal DNA-binding, the OB-fold domain, followed by a C-terminal region, including a glycine-rich loop and an amphipathic tail. The C-terminal region is proposed to be involved in the interaction with enzymes of the DNA metabolism such as polymerases and recombinases (Genschel et al., 2000; He et al., 2003; Jackson et al., 2002; Reddy et al., 2001; Sun & Shamoo, 2003).

In bacteriophages, SSBs proteins play critical roles in DNA replication and recombination. Both mechanisms are often closely coupled throughout much of the phage lytic cycle during recombination-dependent replication processes (Bleuit et al., 2001). In some coliphages, like T4 and T7, phage-encoded SSBs are essential for DNA replication and cannot be replaced by their host counterparts with which they share no homology (Kim & Richardson, 1993; Miller et al., 2003). On the contrary, coliphage P1 (Hay & Cohen, 1983) and Bacillus phage SPP1 (Ayora et al., 1998) encode proteins highly homologous to the bacterial-host SSB. For these reasons, a few phage SSBs have been identified by their sequence similarity to bacterial SSBs whereas others were recognized on the basis of their activity. The crystal structures of phage SSBs, such as the monomeric T4 gp32 (Shamoo et al., 1995), the dimeric T7 gp2.5 (Hollis et al., 2001) and GN5 from *E. coli* filamentous phage φ1 (Skinner et al., 1994), confirmed their OB-fold structural and functional motif (Murzin, 1993).

Virulent phages of Lactococcus lactis have a negative effect on milk fermentations, which explains the increasing interest on the biology of these phages, including their structure (Ricagno et al., 2006; Sciara et al., 2008; Spinelli et al., 2006), phage DNA replication and homologous recombination processes, as well as on their interaction with their hosts (Bouchard & Moineau, 2004; Moineau, 1999; Ploquin et al., 2008). However, there are no structures of lactococcal SSBs available yet.

In this work, we present the crystal structure of the ORF34$_{p2}$ protein, encoded by the virulent *L. lactis* phage p2 (936 group). The *orf34* gene is located downstream a putative single strand annealing protein named Sak3 (Bouchard & Moineau, 2004). ORF34$_{p2}$ probably belongs to a set of genes involved in phage-promoted DNA recombination, as postulated for the SSB protein of the lactococcal phage bIL67 ORF14$_{bIL67}$ (Szczepanska et al., 2007). Indeed, ORF34$_{p2}$ bears 33% sequence identity with ORF14$_{bIL67}$ (Fig. 1). Of note, virulent phages bIL67 and p2 belong to

genetically-distinct groups of lactococcal phages, namely c2 and 936, respectively (Deveau et al., 2006). Despite the lack of homology with SSB proteins present in databases, ORF34$_{p2}$ shares some characteristics generally observed in this protein family, notably a C-terminal region rich in hydrophilic non-charged amino acids and proline residues, the five-residues consensus sequence DLFGG, as well as the conserved three terminal residues LPF. The presence of these features, which constitute a typical SSB fingerprint (He et al., 2003), prompted us to explore this hypothesis by the study of the biochemical and structural properties of ORF34$_{p2}$.

Results and Discussion

ORF34$_{p2}$ is expressed early and abundantly during phage p2 infection.

To investigate the expression course of ORF34$_{p2}$ during infection, we infected the host strain *L. lactis subsp. cremoris MG1363* with phage p2 and took culture samples at 5 min intervals during 20 minutes. This is the minimum time reported before new p2 virions are released (Haaber et al., 2008). Intracellular proteins were extracted and run on 2D poly-acrylamide gels. Proteomic profiles were compared between infected and non-infected cells in a time course manner (Fig. 2). Based on the theoretical molecular weight (13 kDa) and pI (4.92) of ORF34$_{p2}$, a spot corresponding to an early expressed phage protein with apparent mass 15.4 kDa and pI 5.1 (Fig. 2) was picked out of the gels and analyzed by LC/MS/MS. The mass spectrometry matched ORF34$_{p2}$ protein with 97% of coverage and more than 95% of confidence. The analysis of the time course experiment revealed that ORF34$_{p2}$ was detected as early as 5 minutes post-infection and that its concentration increased significantly over time (Fig. 2). In fact, ORF34$_{p2}$ was one of the most expressed phage p2 proteins (data not shown). Early and abundant expression are the landmarks of proteins playing an important role in the phage infection process. It should be noted that SSBs in general are essential proteins, which cannot be deleted or replaced.

Recombinant ORF34$_{p2}$ forms dimers and tetramers in solution.

An N-terminally His-tagged ORF34$_{p2}$ protein was expressed in *E. coli* and purified as a soluble protein that could be concentrated to 10 mg/ml. Recombinant ORF34$_{p2}$ is prone to time-dependent proteolysis, since SDS-PAGE gels show signs of cleavage after keeping the protein for 5 days at 4 °C (data not shown). SEC-MALS/UV/RI (Wyatt Technology) experiments suggest that ORF34$_{p2}$ is a dimer, like T7 gp 2.5 (Hollis et al., 2001) or phage Pf3 SSB (Folmer et al., 1994). However, size-exclusion chromatography showed that the protein would also form tetramers, like EcoSSB (Raghunathan et al., 2000; Raghunathan et al., 1997) and the phage bIL67 SSB protein ORF14$_{bIL67}$ (Szczepanska et al., 2007). The (SeMet) M2B-ORF34$_{p2}$ mutant prepared for structural studies (see below and in Experimental procedures) showed characteristics similar to the native protein, but with a faster time-dependent degradation.

ORF34$_{p2}$ binds ssDNA in vitro.

Electrophoretic Mobility Shift Assays (EMSA) were used to investigate the ssDNA- and dsDNA-binding properties of ORF34$_{p2}$ (Fig. 3). Quantification of the DNA-protein complexes at 1 μM ORF34$_{p2}$ concentration indicated that 40 % of the input ssDNA was stably bound compared to 1 % for dsDNA (Fig. 3a, lanes 5 and 11). Thus, consistently with its presumed SSB function, ORF34$_{p2}$ nonspecifically binds ssDNA and does not bind dsDNA. This property is necessary to coating and protecting ssDNA intermediates during DNA replication, repair and recombination (Gasior et al., 2001).

ORF34$_{p2}$ binding to ssDNA was also monitored by EMSA using circular M13 ssDNA. As shown in Fig. 3b, titration of M13 ssDNA with increasing amounts of protein resulted in a band shift that reaches a plateau at a 400 ORF34$_{p2}$:M13 ssDNA molar excess (Fig. 3b, lane 6). Given the length of the M13 ssDNA (7249-nt) and assuming a tetrameric organization of ORF34$_{p2}$, this result implies a site size of ~72 nucleotides, similar to the 65 nucleotides wrapped around the EcoSSB homotetramer in one of its binding modes (Lohman & Overman, 1985).

To confirm the qualitative results of EMSA and to measure a dissociation constant for the ORF34$_{p2}$-ssDNA complexes, we performed Surface Plasmon Resonance

(SPR) experiments. This method has already been used to evaluate SSB-ssDNA interactions in real time (Wang et al., 2000). We coated a streptavidin (SA) sensor chip with ssDNA and recombinant ORF34$_{p2}$ was injected over the chip surface at increasing concentrations, between 23 and 375 nM. The mean dissociation constant (K_D) value, based on three independent experiments, was $2.5 \pm 0.6 \ 10^{-8}$ M with an association ($k_{on,}$) and a dissociation (k_{off}) rates of $1 \pm 0.2 \ *10^5$ M^{-1}s^{-1} and $2.5 \pm 0.1*10^{-3}$ s^{-1}, respectively. This result confirms that ORF34$_{p2}$ has a high affinity for ssDNA, albeit lower that other SSB proteins like EcoSSB ($K_D = 2.5 \ 10^{-9}$ M with faster k_{on} and k_{off}, $1.4 \ 10^5$ M^{-1}s^{-1} and $3.4 \ 10^{-4}$ s^{-1} respectively) and Mycobacterium tuberculosis SSB ($K_D = 3.6 \ 10^{-9}$ M with a faster k_{on} and k_{off}, $6.3 \ *10^4$ M^{-1}s^{-1} and $2.3*10^{-4}$ s^{-1} respectively) (Reddy et al., 2001).

The crystal structure of ORF34$_{p2}$ shows overall similarity with bacterial SSBs along with significant differences.

We determined the crystal structure of ORF34$_{p2}$ to gain further insight into its possible role as an SSB protein. If that hypothesis could be confirmed, the structure of ORF34$_{p2}$ would have the bonus of giving the first information at atomic detail of a lactococcal phage SSB protein, a family that shares no significant homology with any SSBs of known structure.

The recombinant (native) ORF34$_{p2}$ protein was crystallized with space group symmetry P2$_1$2$_1$2$_1$ (Table 1). Since low sequence identity with possible models precluded the possibility of solving the structure of the protein by the molecular replacement method and due to the absence of methionine residues in its sequence, we mutated leucines at positions 48 and 54 into methionines thus generating the L48M-L54M double mutant M2B-ORF34$_{p2}$. Seleno-methionyl (SeMet) M2B-ORF34$_{p2}$ crystallizes under conditions and with different space group symmetry (P6$_3$22, Table 1) from the native protein. The structure of (SeMet) M2B-ORF34$_{p2}$ was solved by the single-wavelength anomalous dispersion (SAD) method. The structure of native ORF34$_{p2}$ was subsequently solved by molecular replacement using the structure of (SeMet) M2B-ORF34$_{p2}$ as search model. There is one ORF34$_{p2}$ molecule per asymmetric unit in the (SeMet) M2B-ORF34$_{p2}$ hexagonal crystals and

four in the orthorhombic crystals of native $ORF34_{p2}$. Thus, we had at our disposal five independent monomers of $ORF34_{p2}$ for structural analysis.

As expected for an SSB protein, the N-terminal DNA-binding domain of $ORF34_{p2}$ has the overall structure of an oligonucleotide/oligosaccharide (OB) fold (Figure 4a). Therefore, the nomenclature established for EcoSSB by Raghunathan and co-workers (Raghunathan et al., 1997) will be used here for consistency. Two secondary-structure elements in the structure of $ORF34_{p2}$ differ significantly from other structures of SSB proteins. Namely, the 1' strand is unusually long and, remarkably, the capping α-helix is absent, a unique feature among SSB proteins. Instead, strands 3 and 4 are connected by a eight residues-long loop (P50-D57).

It is also noteworthy that the orientations of the 1', 2 and 3 strands deviate from their orientation in EcoSSB and other SSB proteins. In $ORF34_{p2}$ the 1'-2 β-hairpin is farther and the L_{23} loop closer to the 45_1-45_2 β-hairpin of the same monomer, respectively (Fig. 4a). As a result, the structure of the $ORF34_{p2}$ monomer is more compact than that of the monomer of *Eco*SSB. Furthermore, these three structural motifs are also those exhibiting a higher variability within the five independent monomers presented in this study. Globally, the five independent $ORF34_{p2}$ monomers can be superposed with a main-chain root mean square deviation (r.m.s.d.) of 0.8 Å for the alpha carbons of the 88 residues aligned. In comparison, superposition of the four independent monomers of the *Eco*SSB structure (PDB:1QVC) onto monomer A from then native $ORF34_{p2}$ structure, gives a main-chain r.m.s.d. of 3.9 Å for the alpha carbons of the 74 residues aligned. In both cases a 2.0 Å lower limit for structural similarity was enforced. If the most flexible motifs are excluded and the rigid parts of the five $ORF34_{p2}$ monomers are aligned separately, the main-chain alpha carbon r.m.s.d. falls to 0.55 Å for 81 residues aligned, as determined by the RAPIDO server (Mosca & Schneider, 2008). When the same approach is used with the *Eco*SSB monomers, their r.m.s.d. with respect to $ORF34_{p2}$ falls to 1.3 Å for 51 residues aligned.

Well-defined electronic density is lacking for the C-terminal region beyond residues 88-93 in all subunits of both the (SeMet) M2B-$ORF34_{p2}$ and the native $ORF34_{p2}$ structures. This result is in line with previous data on *Eco*SSB (Savvides et al., 2004), *Mycobacterium tuberculosis* SSB (Saikrishnan et al., 2003) and *Thermotoga*

*maritim*a SSB (DiDonato et al., 2006). Savvides and co-workers have proposed that the C-terminus of *Eco*SSB may protrude away from the DNA-binding domain as a disordered/unstructured region (Savvides et al., 2004). Indeed, the low sequence complexity of the C-terminal domain of SSB proteins, heavily populated by proline, glycine and glutamine residues had hinted at the possibility of the intrinsic disorder of this region (Sancar et al., 1981). Electron density was also too weak to support a model in parts of the L_{23} loop for one of the monomers in the native structure (residues K36-G37, monomer D) and for the (SeMet) M2B-ORF34$_{p2}$ monomer (residues K36-N39). Similarly, residues S70-E72 (monomer D, native structure) and A69-V74 (SeMet) M2B-ORF34$_{p2}$ monomer) belonging to the 45_1-45_2 β-hairpin could not be modeled.

Structure of the ORF34$_{p2}$ dimer.

SSB proteins can be found in a variety of oligomeric forms: monomeric (e.g. bacteriophage T4 gp32 (Shamoo et al., 1995), dimeric (e.g. T7 gp2.5 (Hollis et al., 2001), heterotrimeric (e.g. human RPA (Iftode et al., 1999) and homotetrameric, as in bacterial SSB proteins such as *Eco*SSB. ORF34$_{p2}$ protein forms dimers bearing the same structural disposition in both its native and the (SeMet) M2B-ORF34$_{p2}$ crystal forms. The asymmetric unit of the native ORF34$_{p2}$ crystal includes two dimers while in the (SeMet) M2B-ORF34$_{p2}$ crystal one dimer is formed by the interaction between two symmetry-related monomers (Fig. 4b). The fact that these crystalline forms belong to two different crystallographic space groups strongly supports the notion that the observed dimer has a physiological entity.

In the ORF34$_{p2}$ dimer, the extended, antiparallel β-sheet formed by the 1, 4 and 5 strands from each monomer observed in *Eco*SSB is also present and contributes eight hydrogen bonds to the dimer interface. However, beyond this common feature, the differences observed between the ORF34$_{p2}$ and the *Eco*SSB monomers translate into different arrangements of their respective dimers. The most salient difference between the ORF34$_{p2}$ dimer and the dimers of bacterial SSB proteins is the close interaction between the L_{23} loop of each monomer and the 45_1-45_2 β-hairpin of the other monomer (Fig. 4b). This unique interaction implies that the central prominence that in other SSB proteins is made by the 45_1-45_2 β-hairpins opens up, creating a

continuous groove along the dimer longest dimension (Fig. 4b and 4d). This shallow groove communicates the two transversal channels where ssDNA binds in *Eco*SSB (Raghunathan et al., 2000). Although the interface between L_{23} and the 45_1-45_2 β-hairpin is mainly of hydrophobic nature, this interaction contributes, on average, eight further hydrogen bonds to the dimer, including possible ionic interactions between residue pairs E33-R64 and E35-K68. The actual bonding scheme varies in each case reflecting the flexibility of the structural motifs involved in the interaction, and consistently with the fact that these regions were not completely modeled in some of the monomers (see above). This variability is also observed when we compare the interface surfaces of the dimers. In the three independent dimers found (two in the native structure and one in that of (SeMet) M2B-ORF34$_{p2}$) each monomer buries a surface ranging from 945 Å^2 in the (SeMet) M2B-ORF34$_{p2}$ structure to 1085 Å^2 in one of the dimers in the native structure, as reported by the PISA server at the EBI (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (Krissinel & Henrick, 2004).

Architecture of the ORF34$_{p2}$ tetramer.

The ambiguity observed in solution with respect to the oligomeric state of ORF34$_{p2}$ is echoed by its crystals forms. Whereas the native structure unequivocally displays a head-to-head tetrameric layout, the (SeMet) M2B-ORF34$_{p2}$ structure clearly appears as a dimer.

Similarly to other SSB proteins, the ORF34$_{p2}$ tetramer exhibits pseudo-D$_2$ symmetry (Fig. 4c). However, the principal axes of the two dimers form an angle of ~ 85°, substantially bigger than the angles found in *Eco*SSB (~ 70°) or *Mtb*SSB (~ 35°) tetramers. This almost orthogonal disposition of the dimers allows them to pack more closely than in other homotetrameric SSB proteins, making of the ORF34$_{p2}$ tetramer the most compact homotetrameric SSB structure described until now.

The ORF34$_{p2}$ tetramer is held together mainly by hydrophobic interactions ensuing from the further 3425 Å^2 buried upon its formation. The tetramerization interface, which involves strands 1, 4 and 5 of each monomer, includes only six hydrogen bonds, mostly involving residues at both ends of the well-defined structure, namely T1, T6, Q8, E84 and T89. Contrary to *Eco*SSB and *Hmt*SSB, there are no ionic

interactions to strengthen the tetramer. Therefore, we conclude that under the conditions tested, ORF34$_{p2}$ can form strong dimers and weak tetramers.

Furthermore, a higher-order organization is suggested by the native structure of ORF34$_{p2}$, where a crystallographic interaction is found between monomers A and C of adjacent tetramers. Each monomer buries $\sim$ 980 Å^2 in this interface, which includes more than 10 hydrogen bonds and a possible salt bridge between E57 in monomer A and R15 in monomer C. This inter-tetrameric interface occludes the putative ssDNA binding groove of monomer C but leaves monomer A free to bind to ssDNA. If, as known for *Eco*SSB (Lohman & Overman, 1985), ORF34$_{p2}$ supports different ssDNA binding modes, this tetramer-tetramer interaction could be the basis for one of those putative modes. However, this hypothesis must be tempered by the fact that the (SeMet) M2B-ORF34$_{p2}$ structure does not show this type of interaction.

The structure of ORF34$_{p2}$ suggests putative ssDNA binding sites.

OB-fold proteins tend to use a common ligand-binding interface centered on β-strands 2 and 3, with contributions from the L$_{12}$, L$_{34}$ and L$_{45}$ loops (Murzin, 1993). In the case of EcoSSB, this interface is further extended so that the ssDNA binds along the groove defined in each monomer by the 1'-2 and the 45$_1$-45$_2$ β-hairpins, with further interactions involving the L$_{23}$ loop and the β-helix. The particular architecture of the ORF34$_{p2}$ tetramer makes that the clefts expected to direct ssDNA binding form an almost continuous trench that has a positive electrostatic potential (Fig. 4d). Lining this trench, several positively charged and aromatic residues are interspaced that might play a role in ssDNA binding: K12, R15, K21, K24, K34, K36, R64, K68, K85 and Y44, F48, F52, Y73, F80 and F87, respectively. Residues marked in bold-italics are conserved, or replaced by an amino-acid of the same type, among the homologues of ORF34$_{p2}$ in lactococcal phages (Fig. 1) and thus constitute potential candidates to intervene in ssDNA binding. With a complementary approach, site-directed mutagenesis and subsequent electrophoretic mobility-shift assays performed by Szczepanska and co-workers on Orf14$_{bIL67}$ have shown that residues V17, K65 and Y70, especially the latter two, are involved in ssDNA binding (Szczepanska et al., 2007). Of these residues, only V17 is conserved in ORF34$_{p2}$. However, the other two residues would indeed localize in the ssDNA binding groove. Further

experiments will be required to determine the actual residues involved in direct interaction with ssDNA.

Whereas the crystal structure of ORF34p2 clearly corresponds to that of an SSB protein, its unique features strongly support the notion that lactococcal SSB proteins constitute a new family within the SSB protein superfamily, as it has been proposed elsewhere (Szczepanska et al., 2007).

Imaging of SSB-ssDNA complexes by atomic force microscopy

Atomic force microscopy has been previously exploited to confirm the capacity of SSB proteins to bind ssDNA, e.g. for T7 gp2.5 (He et al., 2003) and for *Eco*SSB (Hamon et al., 2007). Compared to dsDNA, ssDNA is considerably more difficult to image by AFM because it is thinner and forms secondary structures which fold the molecules into complex shapes, or even into blobs (Fig. 5a). To curtail this inconvenience we substituted the trivalent cation spermidine for magnesium in the deposition buffer (Hamon et al., 2007). Spermidine was also efficient in spreading the SSB-ssDNA complexes on the surface at all the protein concentrations used in the experiments presented herein.

We titrated the circular coliphage M13 genomic ssDNA with $ORF34_{p2}$ (Fig. 5). Under our experimental conditions and in the absence of $ORF34_{p2}$, M13 ssDNA displayed several regions with secondary structure organization, spaced by extended stretches (Fig. 5b). We did not observe fully extended M13 ssDNA molecules.

Incubation of M13 ssDNA with increasing concentrations of $ORF34_{p2}$ (Fig. 5c-f) resulted in the formation of nucleoprotein complexes that reached saturation at a $ORF34_{p2}$: M13 ssDNA molar ratio R of 400:1, in full agreement with the EMSA data obtained under very similar experimental conditions (Fig. 3b). In AFM experiments, saturation was established by the appearance of many unbound $ORF34_{P2}$ particles spread over the mica surface and by the observation of fully coated M13 molecules (Fig. 5f). Under such conditions, the $ORF34_{p2}$-ssDNA complexes formed packed "beads-on-a-string" structures that were more extended than the M13 ssDNA alone. However, from inspection of the images, it was clear that the overall contour length of the $ORF34_{p2}$-ssDNA nucleoprotein complexes was significantly less than what it would be expected for the fully-extended ssDNA (Fig. 5e,f). This result strongly

indicated that the DNA binding mode of ORF34$_{p2}$ involves wrapping of the ssDNA around the protein, as anticipated for an SSB protein. The analysis of 30 M13 ssDNA molecules in the presence of saturating amounts of ORF34$_{p2}$ (Fig. 5g) showed that, on the average, there were about 63 ORF34$_{p2}$ particles, possibly tetramers, bound to a single M13 ssDNA molecule. Given the size of M13 ssDNA (7249-nt), these data suggest that binding of an ORF34$_{p2}$ particle to ssDNA involves ~115 nucleotides, either wrapped around the protein or constituting the DNA linker between adjacent particles. Unfortunately, the compact structure of the complexes and the broadening effect of the tip did not permit direct visualization of the DNA linker. On a similar study on *Eco*SSB (Griffith & Shores, 1985), stretches of about 30-nt protein-free DNA were found.

In order to better quantify the length of ssDNA occupied by a single ORF34$_{p2}$ particle and to determine the degree of DNA compaction consequent to the binding, we performed experiments with a linear ssDNA fragment (1204-nt) obtained by asymmetric PCR. This procedure results in a mixture of single-stranded and double-stranded DNA fragments that can easily be discerned from AFM images because dsDNA is thicker and more extended than ssDNA (data not shown). Figure 5h shows a set of images obtained with this DNA substrate under saturating concentrations of ORF34$_{p2}$ (R=400:1). On the one hand, these images show that ORF34$_{p2}$ does not bind dsDNA, confirming the EMSA experiments. On the other hand, ssDNA fragments were completely covered by ORF34$_{p2}$ and formed more extended conformations than those observed with M13 ssDNA. In this case, the limited size of the ssDNA fragment allowed a more accurate determination of the nucleoprotein filament contour length and a more precise evaluation of the number of ORF34$_{p2}$ particles bound to a single ssDNA fragment (Fig. 5i). In particular, in a sample of 170 ORF34$_{p2}$-ssDNA complexes the average number of SSB tetramers bound to the 1204-nt long ssDNA was 9 ± 0.3 (Fig. 5k) , which implies an average SSB nucleo-particle spacing of 134 ± 25 nucleotides. Comparison with site sizes for *Eco*SSB (up to 65 nucleotides/homotetramer (Lohman & Overman, 1985)) and our own results from EMSA (72 nucleotides/homotetramer) this result might indicate that the observed ORF34$_{p2}$ nucleo-particles could correspond to ssDNA bound to ORF34$_{p2}$ octamers.

Contour length measurements of these 170 ORF34$_{p2}$-ssDNA complexes (Fig. 5i) resulted in an average contour length value of 164 ± 4.6 nm. Assuming that the nucleotide step in ssDNA is about 0.6 nm (Murphy et al., 2004), the expected contour length of the 1204-nt long ssDNA fragment would be 722 nm. Therefore, under in vitro conditions of protein saturation, binding of ORF34$_{p2}$ results in a ~77% compaction of the ssDNA.

ORF34$_{p2}$ stimulates RecA-mediated homologous recombination in an in vitro strand exchange reaction

An important feature of SSB proteins is that they stimulate homologous recombination reactions. Thus, *Eco*SSB is required to remove secondary structures from ssDNA thereby allowing RecA to efficiently form presynaptic complexes (Kowalczykowski et al., 1987). Similarly, *S. cerevisiae* RPA (the eukaryotic SSB homologue) does not stimulate Rad51 ATPase activity on ssDNA devoid of secondary structure such as poly(dT), while it can stimulate Rad51 by ~1.4 fold on phiX174 ssDNA (Sugiyama et al., 1998). We therefore addressed the question of whether ORF34$_{p2}$ could stimulate RecA. First, a titration with RecA was performed (Fig. 6a) on a series of strand-exchange reactions. We then kept a limiting concentration of RecA (1 μM) while adding increasing concentrations of ORF34$_{p2}$. At 1 μM RecA, only 5% of displaced ssDNA were produced. Addition of 3 μM ORF34$_{p2}$ resulted in a 3-fold increase in activity compared to RecA alone (Fig. 6b). Therefore, we conclude that ORF34$_{p2}$, like other SSB proteins, can stimulate homologous recombination mediated by the RecA recombinase. However, we have not determined if this stimulation requires direct interaction of ORF34$_{p2}$ with RecA, as reported for RPA in *Saccharomyces cerevisiae* (Sung, 1997) and SSBs from mycobacteria (Reddy et al., 2001).

Conclusions

Our studies lead to the characterization of ORF34$_{p2}$ as a bona fide SSB protein. We have shown that ORF34$_{p2}$ is an early and abundantly expressed protein during phage p2 infection that can bind ssDNA with high affinity, but not dsDNA. AFM experiments have shown that upon binding, ssDNA wraps around ORF34$_{p2}$, as

expected for an SSB protein. The structure of ORF34$_{p2}$, which we have determined in two different space groups, shows that this protein forms dimers and tetramers structurally similar to those found in bacterial SSB proteins. Furthermore, ORF34$_{p2}$ can also stimulate RecA-mediated homologous recombination. Finally, these observations are consistent with the genomic context of the *orf34$_{p2}$* gene, downstream a putative single strand annealing protein named Sak3 (Bouchard & Moineau, 2004).

The unique features that we have found in the structure of ORF34$_{p2}$ strongly back up the idea that lactococcal SSB proteins constitute a new family within the SSB protein superfamily (Szczepanska et al., 2007). In this respect, the structure that we present could serve as a basis to understand the selective advantage for lactococcal phage of expressing their own SSB proteins.

Further functional studies on the proteins coded in the same "recombination cassette" than ORF34$_{p2}$, as well as on their interactions will lead to a better understanding of the implication of lactococcal phage-coded SSB proteins in DNA recombination processes associated with host-phage genetic exchanges and DNA replication.

Experimental Procedures

Bacterial strains and phage. Lactococcus lactis subsp. cremoris MG1363 (Gasson, 1983) was grown at 30°C in M17 supplemented with 0.5% glucose (GM17). Propagation of phage p2 (Moineau et al., 1995) and determination of phage titers were performed as described previously (Emond et al., 1997). In phage assays, GM17 was supplemented with 10 mM calcium chloride (GM17-Ca). Phage DNA was isolated as reported elsewhere (Deveau et al., 2006).

Cloning, expression and purification of SSB. Phage p2 orf34 gene was amplified by PCR from phage p2 genomic DNA using specific Gateway™ primers containing attB sequences at both ends, a ribosome binding site, and an N-terminal His_6-tag coding sequence, and was cloned by recombination in a Gateway™ pETG-20A vector (Invitrogen). Due to the lack of methionines in the SSB protein sequence and in order to introduce selenium for diffraction studies, SSB was recloned introducing two leucine-to-methionine substitutions: L48M, L54M (M2B-ORF34$_{p2}$) using the QuickChange® Site-Directed Mutagenesis Kit (Stratagene).

Protein expression experiments were carried out with the Rosetta (DE3)pLyS (Novagen) strain. After an overnight induction with 1 mM isopropyl 1-thio-beta-D-galactopyranoside at 25 °C, cells were harvested by centrifugation for 10 min at 4,000 x g. Mutant M2B was expressed in minimal conditions in presence of seleno-methionine.

Bacterial pellets were resuspended in 40 ml/liter of culture of lysis buffer (Tris 50 mM, NaCl 300 mM, imidazole 10 mM, pH 8.0) supplemented with 0.25 mg/ml lysozyme, 1 µg/ml DNAse, 20 mM $MgS0_4$, and antiproteases (complete EDTA-free antiproteases, Roche) and frozen at −80°C. After thawing and sonication, lysates were cleared by a 30-min centrifugation at 12,000 x g. Over-expressed proteins were purified on a Pharmacia Äkta FPLC by nickel-affinity chromatography (His-Trap 5 ml column, GE Healthcare) using a step gradient of imidazole (at 25 and 250 mM). The N-terminal His_6-tag was removed by digestion with TEV protease followed by Ni-NTA affinity chromatography, to exclude both the tag peptide and the protease. A

final purification step was conducted on a preparative size-exclusion column (Superdex200 HR26/60) equilibrated in 10 mM Tris, 300 mM NaCl pH 8.0. In the case of the selenomethionyl (SeMet) M2B-ORF34$_{p2}$, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was added to the gel filtration buffer to stabilize the redox state of the protein. Purified proteins were concentrated using Amicon Ultra-15 ml filtration systems (Millipore) and characterized by SDS-PAGE and trypsin peptide mass fingerprint by matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Bruker Autoflex). *E. coli* RecA was purified as described previously (Eggleston et al., 1997), while *E. coli* SSB was purchased from Promega.

Size-exclusion chromatography and light scattering. SEC-MALS/UV/RI experiments were carried out on an Alliance 2695 HPLC system (Waters) using a silica-based KW402.5 column (Shodex) equilibrated in Tris 10 mM, 300 mM NaCl, pH 8.0 at a flow of 0.35 ml/min. The protein was injected at a concentration of 3 mg/ml. Detection was achieved by using a triple angle light scattering detector (Mini-DAWN^{TM} TREOS, Wyatt Technology), a quasi-elastic light scattering instrument (Dynapro^{TM}, Wyatt Technology), and a differential refractometer (Optilab$^{®}$ rEX, Wyatt Technology). Molecular weight and hydrodynamic radius determination was performed by ASTRA V software (Wyatt Technology) using a dn/dc value of 0.185 ml/g.

Phage infection experiments. *L. lactis subsp. cremoris MG1363* was grown in GM17-Ca until the optical density at 600 nm reached 0.2, then the virulent phage p2 was added at a multiplicity of infection of 5. Samples were taken at 5 min intervals and flash-frozen (–80 °C). Cell pellets were resuspended in 10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 0.3 % SDS, 60 mM DTT, protease inhibitors (Roche Diagnostics) and glass beads (106 μm, Sigma-Aldrich). The mixture was vortexed with a Mini-Beadbeater-8 cell (BioSpec Products) five times for 1 min each. Between treatments, cell suspensions were chilled on ice for 1 min. The lysed culture was centrifuged and the cytoplasmic extracts were dosed by RC DC protein assay (Bio-Rad). Then, 350 μg of total protein were precipitated by the methanol/chloroform/water method of

Wessel and Flügge (Wessel & Flügge, 1984). Pellets were dried and reconstituted overnight with 450 µl of rehydratation buffer (7 M urea, 2 M thiourea, 50 mM DTT, 4% CHAPS and traces of bromophenol blue).

Two-dimensional gel electrophoresis (2-DE). 2-DE was performed using the Immobiline/polyacrylamide system. Isoelectric focusing (IEF) was performed with 24 cm IPG strips pH 4-7 (GE Healthcare,). The reconstituted protein sample (see above) was applied on IPG DryStrips (IPGphor II, Isoelectric Focusing Unit, GE Healthcare) using the in-gel sample rehydration technique according to the manufacturer's instructions. After rehydration at 20 ºC for 12 h, IEF (50 µA/strip) was conducted at 500 V for 1 h, 1000 V for 1 h, at gradient steps from 1000 V to 8000 V for 3h and then at 8000 V for 6 h 30 min, to reach a total of 60–70 kVh. The temperature was maintained at 20 ºC. After IEF, each strip was equilibrated for 15 min in 10 ml of equilibration buffer 1 (7 M urea, 1 % DTT, 36 % glycerol, 2 % SDS, 50 mM Tris-Cl, pH 8.8) and then in 10 ml of equilibration buffer 2 (7 M urea, 2.5% iodoacetamide, 36% glycerol, 2% SDS, 50 mM Tris-Cl, pH 8.8) for 15 min. For the second dimension, vertical gradient 10-20 % SDS-PAGE was run for about 4 h (17 W/gel; Ettan DALTsix, GE Healthcare). The gels were stained using Colloidal CBB (0.1% CBB G-250, 20% methanol, 10% ammonium sulfate, 2% ortho-phosphoric acid). Scanning was carried out with an ImageScanner (GE Healthcare) and image analysis was performed using ImageMaster 2D Elite software. Spots were cut out of the gel, digested with trypsin, and identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) at the Centre Protéomique de l'Est du Québec (Quebec city, Quebec, Canada).

Gel retardation assays. Electrophoretic mobility shift assay experiments with 100-nt ssDNA or 100 bp dsDNA fragments were performed as follows: in a volume of 10 µl binding buffer (25 mM MOPS pH 7.0, 60 mM NaCl, 1% Tween, 2 mM DTT) ^{32}P-labeled DNA was added to a final concentration of 100 nM and the solution was incubated for 5 minutes at 37°C. Either ORF34$_{p2}$ or E. coli SSB were added to final concentrations of 100, 250, 500 and 1000 nM and the reactions were incubated for other 10 minutes at 37 °C. Complexes were fixed by the addition of 0.2 %

glutaraldehyde followed by a 15 minutes incubation at 37 °C. The reactions were run onto a 4.2% non-denaturing poly-acrylamide gel on 1x Tris-Glycine buffer (50 mM Tris-Cl, 50 mM glycine, pH 8.8). Gels were dried on DE81 filter paper followed by autoradiography. The dsDNA substrate was prepared by annealing two complementary oligonucleotides 100-nt long and by purifying the annealed product on a 10% non-denaturing poly-acrylamide gel. The sequence of the 100-mer was 5'-GGGCGAATTGGGCCCGACGTCGCATGCTCCTCTAGACTCGAGGAATTCGG TACCCCGGGTTCGAAATCGATAAGCTTACAGTCTCCATTTAAAGGACAAG -3'.

Electrophoretic mobility shift assay experiments with circular M13 ssDNA (7249-nt long) (Sigma Aldrich) were performed under the same conditions used for AFM experiments. Namely, ORF34$_{p2}$-ssDNA complexes were assembled using 100 ng of M13 ssDNA (0.5 nM final concentration) and increasing amounts of protein (15, 30, 50, 100, 200, 400 nM) in 20 mM Tris-HCl pH 7.5, 20 mM NaCl, 50 μM Spermidine. The reactions were incubated for 10 min at 37 °C and loaded in a 1% agarose gel after addition of a gel loading buffer containing glycerol. Electrophoresis was run at 4 °C in TBE buffer at 3 V/cm. After a four hours run, the gel was stained with ethidium bromide (0.5 μg/ml) for 60 minutes.

Surface plasmon resonance (SPR). Experiments were performed using a Biacore 1000 instrument (Biacore Inc.) at 25 °C. The chip Streptavidine (SA) was first washed with 10 mM NaOH, 1M NaCl (three times 100 μl at 40 μl/min) to eliminate loosely bound streptavidine. Biotinylated-DNA 40-mer (ssDNA; 45 μg/mL) in 10 mM Tris buffer, 150 mM NaCl, pH 8.0, 0.005% (vol/vol) P20 were fixed as ligand at 550 RU (Resonance Units). ORF34$_{p2}$ at concentration from 23 to 375 nM in 10 mM Tris, 300 mM NaCl, pH 8.0, 0.005% (vol/vol) P20, was used as analyte (80 μl at 10 μl/min). Regeneration was achieved by injection of 4 M MgCl$_2$ (5 μl at 40 μl/min). The RU signal at different analyte concentrations was corrected for the buffer contribution in the same flow-cell and for aspecific interactions on the reference flow-cell. The K$_D$ values were estimated assuming a 1:1 Langmuir model (BIAevaluation Software).

Crystallization, data collection, structure determination and refinement. The chromatographic peaks of native ORF34$_{p2}$ and (SeMet) M2B-ORF34$_{p2}$ matching the theoretical molecular weight of a tetramer were concentrated in each case to 7 mg/ml and subjected to crystallization screening with a Cartesian nanodrop-dispensing robot (Sulzenbacher et al., 2002). Native ORF34$_{p2}$ and (SeMet) M2B-ORF34$_{p2}$ crystals were obtained in 0.1 M Sodium Cacodylate, pH 6.2, 45 % MME PEG 2000 and 0.1 M Hepes, pH 7.2, 60% MME PEG 550, respectively. In both cases crystals appeared after 3 days and were fished using a nylon loop and, since the crystallization condition was found to be cryoprotectant, directly flash-cooled in liquid nitrogen. Native diffraction data were collected to 2.6 Å resolution at the ESRF (Grenoble, France) on beamline ID23-2. Data were integrated and reduced using MOSFLM and then scaled with the program SCALA from CCP4 suite (Collaborative-Computational-Project-Number-4, 1994).

Diffraction data from a (SeMet) M2B-ORF34$_{p2}$ crystal were collected to 2.1 Å resolution at Soleil (Saint-Aubin, Essonne, France) on beamline Proxima 1. These data were integrated using XDS (Kabsch, 1993) and scaled and reduced with the program XSCALE (Kabsch, 1993). The program SHELXD (Uson & Sheldrick, 1999) was used to locate the Se atoms and SHELXE (Sheldrick, 2002) to produce the initial phases. The first model of the (SeMet) M2B-ORF34$_{p2}$ protein was built using the program ARP-wARP (Cohen et al., 2004). This model was improved by successive cycles of manual rebuilding with COOT (Emsley & Cowtan, 2004) and refinement with REFMAC5 (Murshudov et al., 1997).

The native ORF34$_{p2}$ structure was solved by molecular remplacement with MOLREP (Vagin & Teplyakov, 1997) using the best defined parts of the (SeMet) M2B-ORF34$_{p2}$ structure as the search model. The native model was completed using RESOLVE (Terwilliger, 2002), manually rebuilt using COOT and refined with REFMAC5. TLS (Traslation/Libration/Screw) segments (Winn et al., 2001), defined with the help of the TLS Motion Determination server (http://skuld.bmsc.washington.edu/~tlsmd/), were used during refinement. In the native structure, non-crystallographic symmetry (NCS) restraints were applied to all protein backbone at the initial stages of refinement. These restraints were loosened in further refinement cycles and the most flexible loops excluding from them.

Analysis of the stereochemical quality of the native model was performed using Molprobity (Lovell et al., 2003). Data collection and refinement statistics are summarized in Table 1. All structural figures were prepared with CHIMERA (Pettersen et al., 2004). Atomic coordinates and experimental structure factors of native ORF34$_{p2}$ and (SeMet) M2B-ORF34$_{p2}$ have been deposited within the PDB and are accessible under the codes 2WKC and 2WKD, respectively.

Atomic force microscopy (AFM). ORF34p2-M13 ssDNA complexes were prepared by mixing M13 ssDNA at a final concentration of 0.5 nM in buffer containing 20 mM Tris pH 8.0 and 20 mM NaCl with increasing concentrations of ORF34$_{p2}$ protein, from 30 nM up to 200 nM. Reactions were incubated at 25 °C for 5 minutes and then spermidine (Fluka) was added to a final concentration of 50 μM. Spermidine was used as deposition agent instead of magnesium because it allows a better spreading of the nucleoprotein filaments onto mica (Hamon et al., 2007). A 5 μl droplet of the reaction was deposited onto freshly cleaved ruby mica for 1 minute, followed by rinsing with water milliQ (Millipore) and drying with a weak flux of nitrogen.

Linear ssDNA fragment (1204-nt long) were obtained by asymmetric PCR using pNEB193 plasmid as a template and an unequal primer concentration. The PCR mixture contained: 20 ng dsDNA template, 25 pmol of pNEB-For2 primer, 0.25 pmol pNEB-Rev primer, 50 μM of dNTPs (Fermentas), 1.5 mM MgCl$_2$, 10 x PCR buffer and 2.5 U Taq DNA Polymerase (Fermentas) was thermally cycled for 35 times as follows: 95°C for 60 s, 58°C for 30 s, 72°C for 2 min. The DNA product was purified using PCR Clean-up Nucleospin Extract II (Macherey-Nagel).

AFM images were obtained in air with a Nanoscope III microscope (Digital Instruments Inc.) operating in tapping mode. All operations were done at room temperature. Commercial diving board silicon cantilevers (μmasch) were used. The microscope was equipped with a type E scanner (12 mm x 12 mm). Images (512 x 512 pixels) were collected with a scan size of 2 or 4 μm at a scan rate varying between two and four scan lines per second.

The AFM images were analyzed using software written in Matlab (MathWorks Inc.). Contour length measurements of the nucleoprotein filaments were performed as described elsewhere (Rivetti & Codeluppi, 2001).

Counting of ORF34$_{p2}$ particles coating ssDNA was done manually. Only molecules that were completely visible in the image and did not have an ambiguous shape were considered. Data were plotted using Sigma Plot (Systat Software, Inc.).

Strand exchange. Reactions (10 µl) contained purified circular single-stranded pPB4.3 DNA (15 µM) with the indicated concentrations of E. coli RecA and ORF34$_{p2}$ in TD buffer (25 mM Tris-Acetate pH 7.5, 8 mM MgCl$_2$, 1 mM DTT, 1 mM ATP, 20 mM creatine phosphate, 5 U/ml phosphocreatine kinase). ORF34$_{p2}$ was added first to the reaction. After 5 min at 37 °C, ^{32}P-end labeled pPB4.3 DNA (400 bp fragment, 1.38 µM) was added and incubation was continued for 90 min. Reaction products were deproteinized by addition of one-fifth volume of stop buffer (0.1 M Tris-HCl, pH 7.5, 0.1 M MgCl$_2$, 3% SDS, 5 µg/ml ethidium bromide and 10 mg/ml proteinase K) followed by 45 min incubation at 37 °C. Labeled DNA products were analyzed by electrophoresis through 0.8% TAE agarose gels containing 1 µg/ml ethidium bromide, run at 4.3 V/cm, dried onto DE81 filter paper and visualized by autoradiography.

References

Ayora, S., Langer, U., and Alonso, J. C. (1998) Bacillus subtilis DnaG primase stabilises the bacteriophage SPP1 G40P helicase-ssDNA complex. FEBS Lett, 439: 59-62.

Bleuit, J. S., Xu, H., Ma, Y., Wang, T., Liu, J., and Morrical, S. W. (2001) Mediator proteins orchestrate enzyme-ssDNA assembly during T4 recombination-dependent DNA replication and repair Proc Natl Acad Sci U S A, 98: 8298-305.

Bouchard, J. D. and Moineau, S. (2004) Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins J Bacteriol, 186: 3649-52.

Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., et al. (2004) Towards complete validated models in the next generation of ARP/wARP Acta Crystallogr D Biol Crystallogr, 60: 2222-9.

Collaborative Computational Project Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography Acta Cryst D, 50: 760-3.

Deveau, H., Labrie, S. J., Chopin, M., and Moineau, S. (2006) Biodiversity and classification of lactococcal phages Appl Environ Microbiol, 72: 4338-46.

DiDonato, M., Krishna, S. S., Schwarzenbacher, R., McMullan, D., Jaroszewski, L., Miller, M. D., et al. (2006) Crystal structure of a single-stranded DNA-binding protein (TM0604) from Thermotoga maritima at 2.60 A resolution. Proteins, 63: 256–260.

Eggleston, A. K., Mitchell, A. H., and West, S. C. (1997) In vitro reconstitution of the late steps of genetic recombination in E. coli Cell, 89: 607-17.

Emond, E., Holler, B. J., Boucher, I., Vandenbergh, P. A., Vedamuthu, E. R., Kondo, J. K., and Moineau, S. (1997) Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from Lactococcus lactis Appl Environ Microbiol, 63: 1274-83.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr, 60: 2126–2132.

Folmer, R. H., Folkers, P. J., Kaan, A., Jonker, A. J., Aelen, J. M., Konings, R. N., and Hilbers, C. W. (1994) Secondary structure of the single-stranded DNA binding protein encoded by filamentous phage Pf3 as determined by NMR Eur J Biochem, 224: 663-76.

Gasior, S. L., Olivares, H., Ear, U., Hari, D. M., Weichselbaum, R., and Bishop, D. K. (2001) Assembly of RecA-like recombinases: distinct roles for mediator proteins in mitosis and meiosis Proc Natl Acad Sci U S A, 98: 8411-8.

Gasson, M. J. (1983) Plasmid complements of Streptococcus lactis NCDO 712 and other lactic streptococci after protoplast-induced curing J Bacteriol, 154: 1-9.

Genschel, J., Curth, U., and Urbanke, C. (2000) Interaction of E. coli single-stranded DNA binding protein (SSB) with exonuclease I. The carboxy-terminus of SSB is the recognition site for the nuclease Biol Chem, 381: 183-92.

Griffith, J. and Shores, C. G. (1985) RecA protein rapidly crystallizes in the presence of spermidine: a valuable step in its purification and physical characterization Biochemistry, 24: 158-62.

Haaber, J., Moineau, S., Fortier, L., and Hammer, K. (2008) AbiV, a novel antiphage abortive infection mechanism on the chromosome of Lactococcus lactis subsp. cremoris MG1363 Appl Environ Microbiol, 74: 6528-37.

Hamon, L., Pastre, D., Dupaigne, P., Le Breton, C., Le Cam, E., and Pietrement, O. (2007) High-resolution AFM imaging of single-stranded DNA-binding (SSB) protein--DNA complexes. Nucleic Acids Res, 35: e58.

Hay, N. and Cohen, G. (1983) Requirement of E. coli DNA synthesis functions for the lytic replication of bacteriophage P1 Virology, 131: 193-206.

He, Z., Rezende, L. F., Willcox, S., Griffith, J. D., and Richardson, C. C. (2003) The carboxyl-terminal domain of bacteriophage T7 single-stranded DNA-binding protein modulates DNA binding and interaction with T7 DNA polymerase J Biol Chem, 278: 29538-45.

Hollis, T., Stattel, J. M., Walther, D. S., Richardson, C. C., and Ellenberger, T. (2001) Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7 Proc Natl Acad Sci U S A, 98: 9557-62.

Iftode, C., Daniely, Y., and Borowiec, J. A. (1999) Replication protein A (RPA): the eukaryotic SSB Crit Rev Biochem Mol Biol, 34: 141-80.

Jackson, D., Dhar, K., Wahl, J. K., Wold, M. S., and Borgstahl, G. E. O. (2002) Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA J Mol Biol, 321: 133-48.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants Journal of Applied Crystallography, 26: 795–800.

Kim, Y. T. and Richardson, C. C. (1993) Bacteriophage T7 gene 2.5 protein: an essential protein for DNA replication Proc Natl Acad Sci U S A, 90: 10173-7.

Kowalczykowski, S. C., Clow, J., and Krupp, R. A. (1987) Properties of the duplex DNA-dependent ATPase activity of Escherichia coli RecA protein and its role in branch migration Proc Natl Acad Sci U S A, 84: 3127-31.

Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr, 60: 2256–2268.

Lohman, T. M. and Ferrari, M. E. (1994) Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities Annu Rev Biochem, 63: 527-70.

Lohman, T. M. and Overman, L. B. (1985) Two binding modes in Escherichia coli single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration J Biol Chem, 260: 3594-603.

Lovell, S. C., Davis, I. W., Arendall, 3rd, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., et al. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation Proteins, 50: 437-50.

Matsumoto, T., Morimoto, Y., Shibata, N., Kinebuchi, T., Shimamoto, N., Tsukihara, T., and Yasuoka, N. (2000) Roles of functional loops and the C-terminal segment of a single-stranded DNA binding protein elucidated by X-Ray structure analysis. J Biochem, 127: 329–335.

Meyer, R. R. and Laine, P. S. (1990) The single-stranded DNA-binding protein of Escherichia coli Microbiol Rev, 54: 342-80.

Miller, E. S., Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Durkin, A. S., Ciecko, A., et al. (2003) Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage J Bacteriol, 185: 5220-33.

Moineau, S. (1999) Applications of phage resistance in lactic acid bacteria Antonie Van Leeuwenhoek, 76: 377-82.

Moineau, S., Walker, S. A., Vedamuthu, E. R., and Vandenbergh, P. A. (1995) Cloning and sequencing of LlaDCHI [corrected] restriction/modification genes from Lactococcus lactis and relatedness of this system to the Streptococcus pneumoniae DpnII system Appl Environ Microbiol, 61: 2193-202.

Mosca, R. and Schneider, T. R. (2008) RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes Nucleic Acids Res, 36: W42-6.

Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy Biophys J, 86: 2530-7.

150

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr, 53: 240–255.

Murzin, A. G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J, 12: 861–867.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem, 25: 1605–1612.

Ploquin, M., Bransi, A., Paquet, E. R., Stasiak, A. Z., Stasiak, A., Yu, X., Cieslinska, A. M., et al. (2008) Functional and structural basis for a bacteriophage homolog of human RAD52 Curr Biol, 18: 1142-6.

Raghunathan, S., Kozlov, A. G., Lohman, T. M., and Waksman, G. (2000) Structure of the DNA binding domain of E. coli SSB bound to ssDNA. Nat Struct Biol, 7: 648–652.

Raghunathan, S., Ricard, C. S., Lohman, T. M., and Waksman, G. (1997) Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-A resolution. Proc Natl Acad Sci U S A, 94: 6652–6657.

Reddy, M. S., Guhan, N., and Muniyappa, K. (2001) Characterization of single-stranded DNA-binding proteins from Mycobacteria. The carboxyl-terminal of domain of SSB is essential for stable association with its cognate RecA protein J Biol Chem, 276: 45959-68.

Ricagno, S., Campanacci, V., Blangy, S., Spinelli, S., Tremblay, D., Moineau, S., et al. (2006) Crystal structure of the receptor-binding protein head domain from Lactococcus lactis phage bIL170 J Virol, 80: 9331-5.

Rivetti, C. and Codeluppi, S. (2001) Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica Ultramicroscopy, 87: 55-66.

Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U., and Vijayan, M. (2003) Structure of Mycobacterium tuberculosis single-stranded DNA-binding protein. Variability in quaternary structure and its implications. J Mol Biol, 331: 385–393.

Sancar, A., Williams, K. R., Chase, J. W., and Rupp, W. D. (1981) Sequences of the ssb gene and protein Proc Natl Acad Sci U S A, 78: 4274-8.

Savvides, S. N., Raghunathan, S., Futterer, K., Kozlov, A. G., Lohman, T. M., and Waksman, G. (2004) The C-terminal domain of full-length E. coli SSB is disordered even when bound to DNA. Protein Sci, 13: 1942–1947.

Sciara, G., Blangy, S., Siponen, M., Mc Grath, S., van Sinderen, D., Tegoni, M., et al. (2008) A topological model of the baseplate of lactococcal phage Tuc2009. J Biol Chem, 283: 2716–2723.

Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg, W. H., and Steitz, T. A. (1995) Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. Nature, 376: 362–366.

Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE Z. Kristallogr., 217: 644-650.

Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S., and Ogawa, T. (1998) Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing Genes Cells, 3: 145-56.

Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., et al. (1994) Structure of the gene V protein of bacteriophage f1 determined by multiwavelength x-ray diffraction on the selenomethionyl protein Proc Natl Acad Sci U S A, 91: 2071-5.

Spinelli, S., Desmyter, A., Verrips, C. T., de Haard, H. J. W., Moineau, S., and Cambillau, C. (2006) Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses Nat Struct Mol Biol, 13: 85-9.

Sugiyama, T., New, J. H., and Kowalczykowski, S. C. (1998) DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. Proc Natl Acad Sci U S A, 95: 6049–6054.

Sulzenbacher, G., Gruez, A., Roig-Zamboni, V., Spinelli, S., Valencia, C., Pagot, F., et al. (2002) A medium-throughput crystallization approach Acta Crystallogr D Biol Crystallogr, 58: 2109-15.

Sun, S. and Shamoo, Y. (2003) Biochemical characterization of interactions between DNA polymerase and single-stranded DNA-binding protein in bacteriophage RB69 J Biol Chem, 278: 3876-81.

Sung, P. (1997) Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase J Biol Chem, 272: 28194-7.

Szczepanska, A. K., Bidnenko, E., Plochocka, D., McGovern, S., Ehrlich, S. D., Bardowski, J., et al. (2007) A distinct single-stranded DNA-binding protein encoded by the Lactococcus lactis bacteriophage bIL67. Virology, 363: 104–112.

Terwilliger, T. C. (2002) Automated structure solution, density modification and model building Acta Crystallogr D Biol Crystallogr, 58: 1937-40.

Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) Nucleic acid recognition by OB-fold proteins. Annu Rev Biophys Biomol Struct, 32: 115–133.

Uson, I. and Sheldrick, G. M. (1999) Advances in direct methods for protein crystallography. Curr Opin Struct Biol, 9: 643–648.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement Journal of Applied Crystallography, 30: 1022–1025.

Wang, M., Mahrenholz, A., and Lee, S. H. (2000) RPA stabilizes the XPA-damaged DNA complex through protein-protein interaction Biochemistry, 39: 6433-9.

Wessel, D. and Flügge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids Anal Biochem, 138: 141-3.

Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr, 57: 122–133.

Wold, M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem, 66: 61–92.

Tables

Table 1. Data collection and refinement statistics

Data collection	native ORF34$_{p2}$	(SeMet) M2B-ORF34$_{p2}$
Space group	P2$_1$2$_1$2$_1$	P6$_3$22
Unit cell (Å,°)	a = 48.28, b = 71.71, c = 109.73 α = β = γ = 90	a = 75.74, b = 75.74, c = 88.04 α = β = 90, γ = 120
Resolution[1] (Å)	20.0 – 2.60 (2.74 – 2.60)	88.04 – 2.10 (2.21 – 2.10)
Wavelength[1] (Å)	0.8726	0.9800
Completeness[1] (%)	98.8 (99.8)	99.8 (100.0)
Multiplicity[1]	3.5 (3.5)	10.2 (10.5)
Unique reflections[1]	12049 (1755)	9205 (1291)
<I>/s(I)1	11.7 (3.1)	25.8 (4.8)
Rmeas[1,2] (%)	12.9 (47.7)	7.4 (49.2)
Wilson B factor (Å^2)	36.0	40.0
Refinement		
Rfactor/Rfree[3] (%,#)	19.2/25.1 (974)	23.9/25.6 (918)
r.m.s.d. bonds/angles[4] (Å,°)	0.008 (0.022) / 1.032 (1.951)	0.011 (0.022) / 1.186 (1.949)
Average B (Å^2)	44.2	55.7

[1] Values in parenthesis are for the highest resolution shell.

[2] The value of the multiplicity-weighted merging R factor between equivalent measurements of the same reflection, $R = \Sigma_{hkl} \sqrt{\{n_{hkl}/(n_{hkl} - 1\}} \Sigma_j |I_{hkl,j} - <I_{hkl}>| / \Sigma_{hkl} \Sigma\, I_{hkl,j}$.

[3] Crystallographic R_{factor}, $R_{free} = \Sigma\, ||Fo| - k|Fc|| / \Sigma\, |Fo|$. The size of the test set used for the calculation of R_{free} is in parentheses.

[4] Root mean square deviation from the standard values are given with target values in parentheses.

Figure Legends

Figure 1. Sequence alignment of lactococcal phage closest relatives of ORF34$_{p2}$. Except for ORF34$_{p2}$, sequences are labelled with the name of the corresponding *Lactococcus phage*. Residue backgrounds are grayed out by similarity starting from a low similarity cut-off of 70%. Numbers and secondary elements above the sequences correspond to ORF34p2. β–strands are numbered according to the text. Triangles and stars under the sequences indicate the conserved DLFGG and LPF motifs, respectively.

Figure 2. Expression of ORF34$_{p2}$ during the phage p2 infection of *L. lactis subsp. cremoris MG1363*. The IEF assay was performed on IPG strips pH 4-7, then migrated in gradient 10-20% SDS-PAGE, and stained with Colloidal Coomassie Brilliant Blue. (A) 2-DE gel of intracellular proteins of L. lactis MG1363 extracted from a sample taken 10 minutes after phage p2 infection. The box represents the zoom area of panel B. (B) Zoom areas from 2-DE gels obtained with intracellular proteins extracted from non-infected cells (NI) and after 5, 10, 15, and 20 minutes following the beginning of the phage infection. Arrows indicate the position of the SSB protein.

Figure 3. Electrophoretic Mobility Shift Assays. (A) DNA retardation gel assay of the indicated concentrations of ORF34$_{p2}$ on single- (lanes 2-5) and double-strand DNA (lanes 8-11). Lanes 1 and 7, controls without protein; lanes 6 and 12, binding of EcoSSB (0.5 μM) on single- and double-strand DNA. (B) Agarose gel electrophoresis of M13 ssDNA-ORF34$_{p2}$ complexes at increasing DNA:ORF34$_{p2}$ monomer molar ratios (R, at the bottom). Lane 1 corresponds to protein-free ssDNA.

Figure 4. Crystal structure ORF34$_{p2}$. (A) Stereo view of the ORF34$_{p2}$ monomer. A ribbon representation of the main-chain is colored from blue (N-terminus) to to red (C-terminus). β-strands are numbered. Side-chain atoms of positively charged and aromatic residues present in the possible ssDNA binding site are represented as sticks. Those among them that are conserved in lactococcal SSB genes are labelled. (B) Structural comparison of the ORF34$_{p2}$ dimer and the archetypical EcoSSB dimer. Dimers are shown in the same orientation after being superposed. The 45_1-45_2 β-hairpin and L_{23} loop of one of the monomers are represented in each case for reference. (C) Pseudo-D_2 symmetry of the ORF34$_{p2}$ tetramer. The dyad axes are represented as green bars capped with black spheres. Monomers are labelled a-d. (D) Electrostatic potential of the ORF34p2 tetramer mapped onto its solvent-accessible surface from negative (-0.5 V, red) to positive (0.5 V, blue). The tetramer is represented in the same orientation as in panel C. This figure was prepared with the program CHIMERA. Electrostatic calculations were performed with APBS using the PARSE force-field in water with an implicit solvent model including 150 mM NaCl.

Figure 5. AFM images of ORF34$_{p2}$-DNA complexes. Image of M13 ssDNA without ORF34$_{p2}$: deposition with magnesium (A) and spermidine (B-a: typical view, B-b: case of a partially extended M13 ssDNA - Image size: 250 nm). From (C) to (F), increasing concentrations of ORF34$_{p2}$ were used to titrate M13 ssDNA (0.5 nM): 30 nM (C), 50 nM (D), 100 nM (E) and 200 nM (F). (G) Representative 3D image of a ORF34$_{p2}$-saturated M13 ssDNA (Image size: 300 nm). (H) Collage of AFM images showing representative examples of the 170 ORF34$_{p2}$-ssDNA(linear) complexes analyzed. Contour length distribution (I) from these ORF34$_{p2}$-ssDNA(linear) complexes.

Figure 6. Strand Exchange Stimulation by ORF34$_{p2}$. (A) DNA strand exchange as a function of RecA concentration (lanes 2-8). Lane 1, control without protein. (B) Effect of ORF34$_{p2}$ on RecA-mediated strand exchange. RecA at limiting concentrations (1 μM, lane 2) was stimulated by increasing concentrations of ORF34$_{p2}$ (0.25-5 μM, lanes 3-8). Lane 1, control without ORF34$_{p2}$.

Supplementary Figure 1. Surface Plasmon Resonance. Sensorgram of SSB-ssDNA on SA chip. Reponses Units (RU) are represented in the Y axis, while time (s) is represented in the X axis.
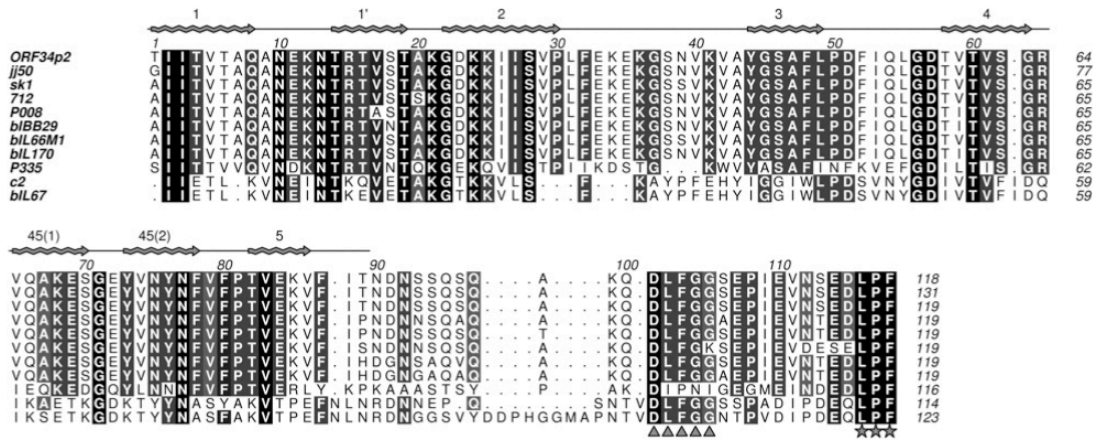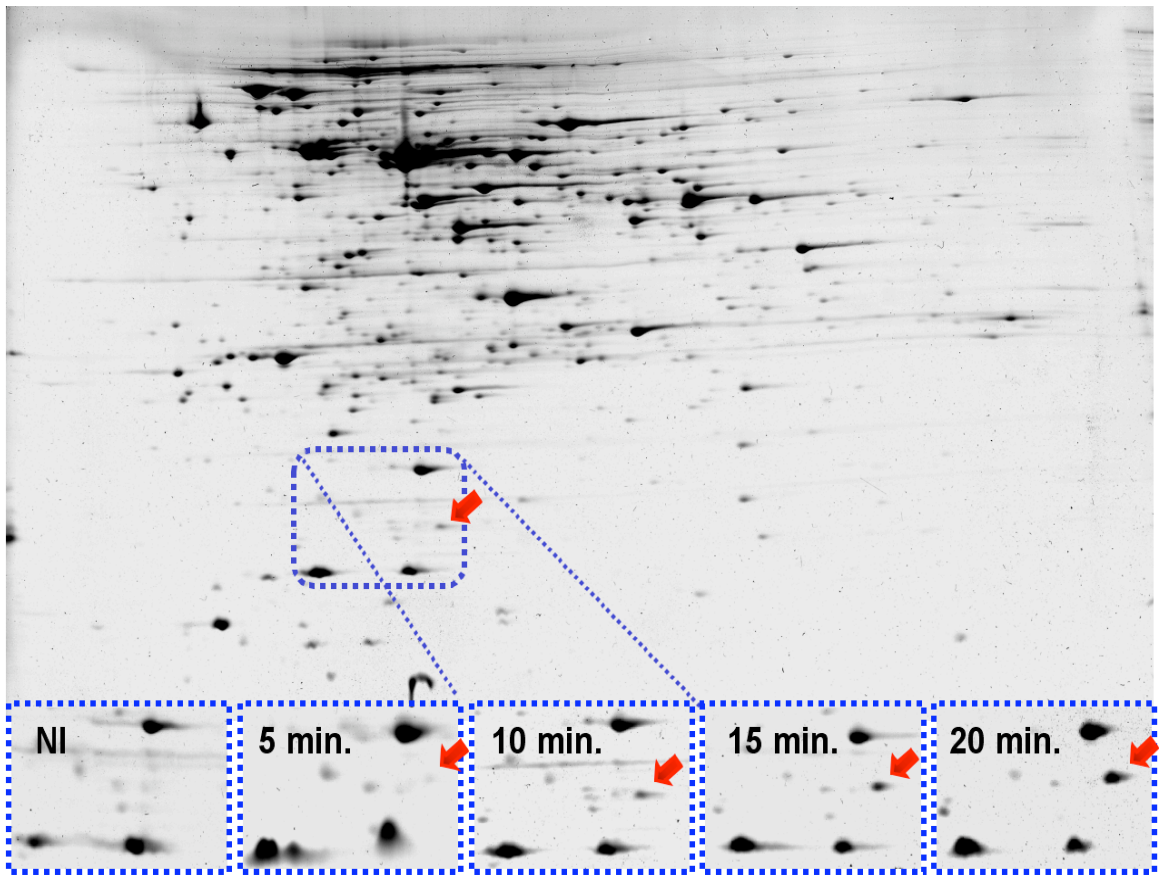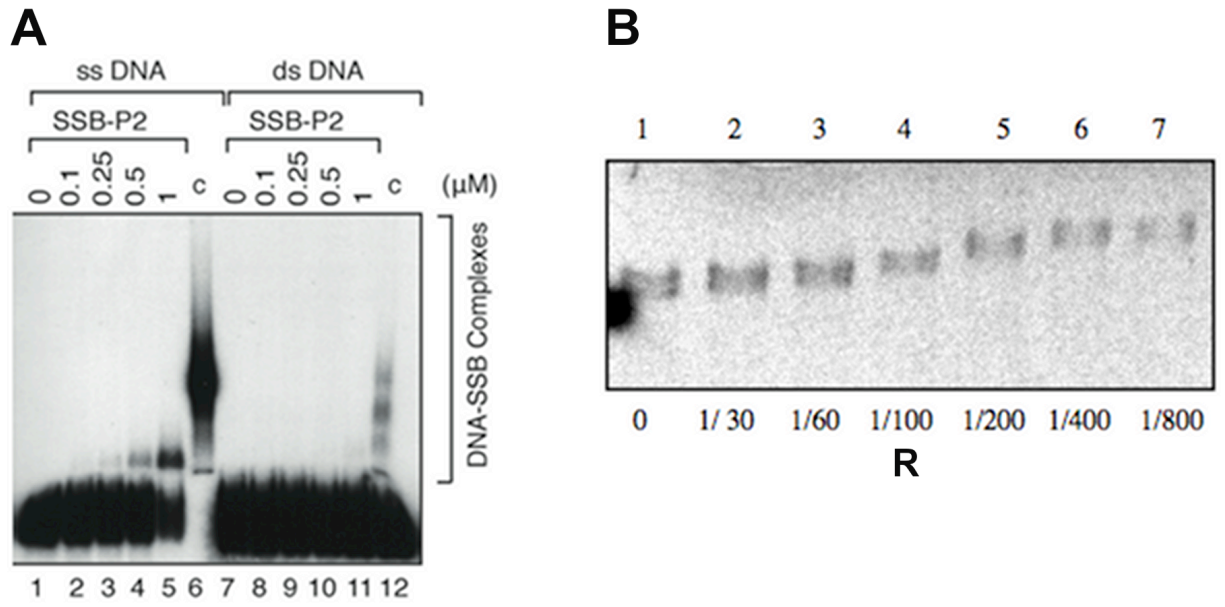
Figures

Figure 1

Figure 2

Figure 3

Figure 4

A

B

ORF34_{p2}   *Eco*SSB

C   D
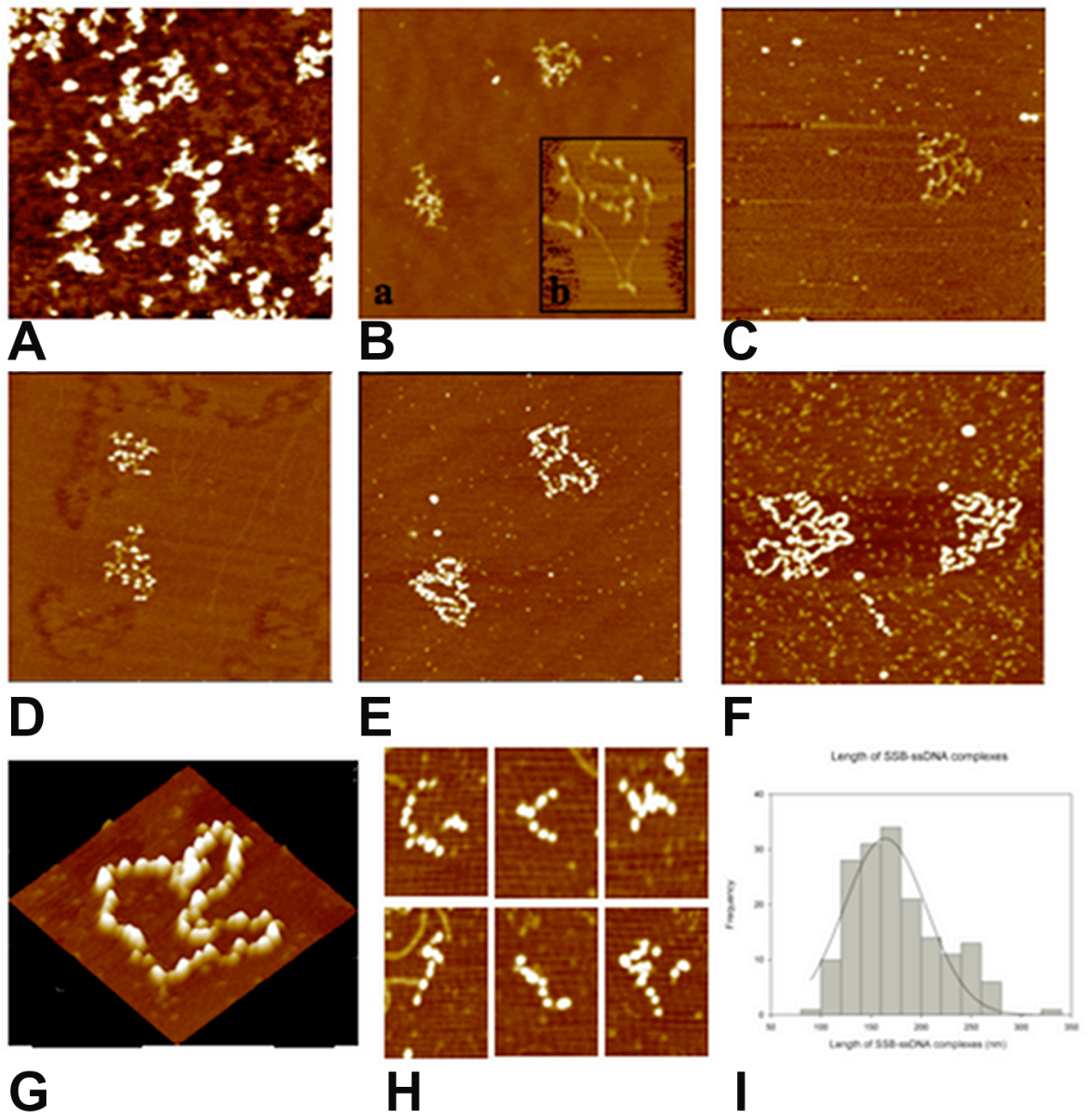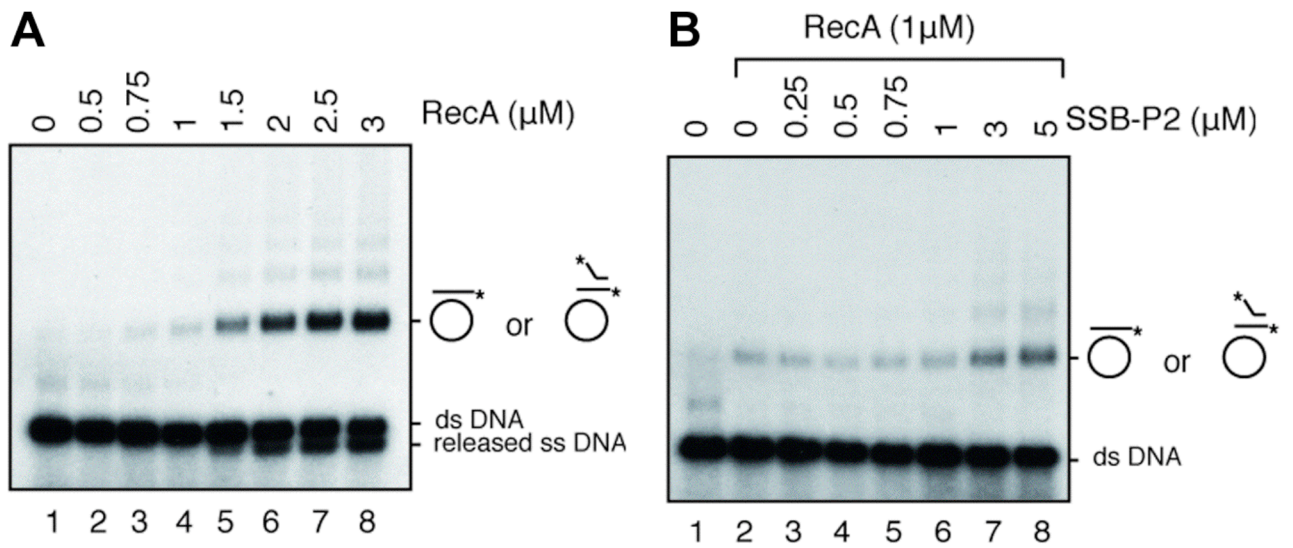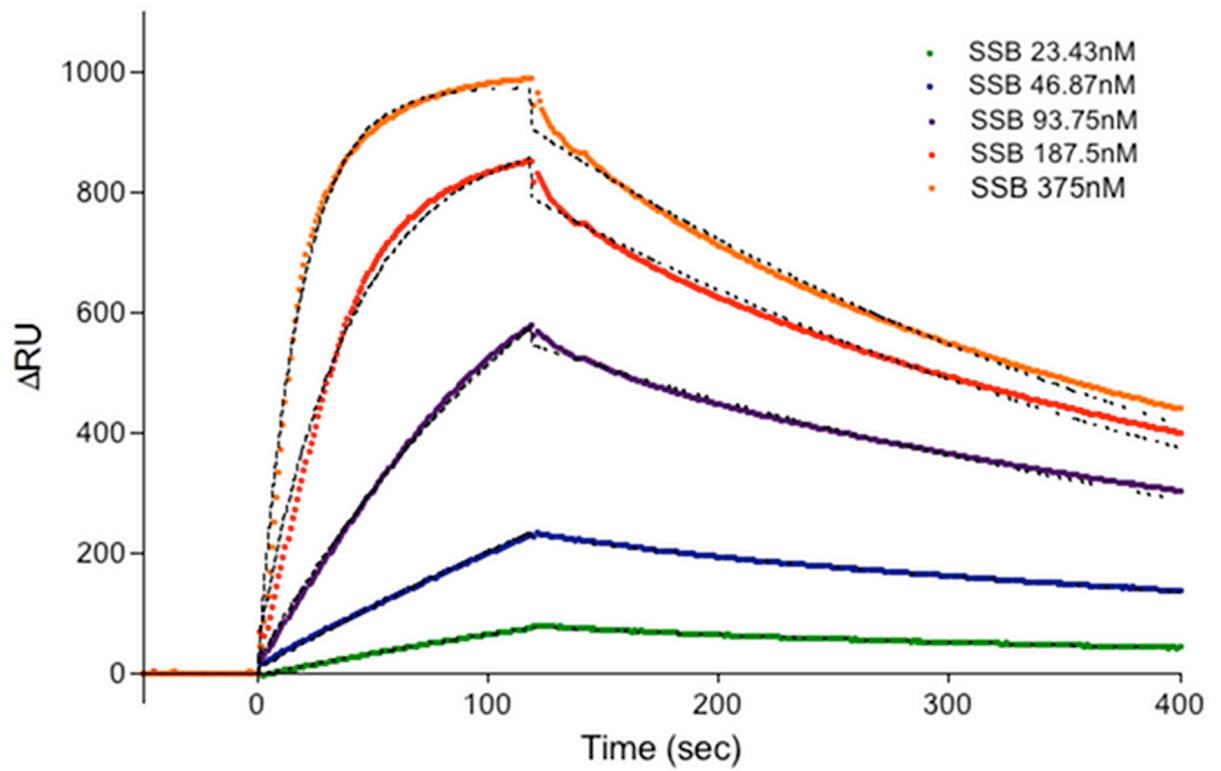

157
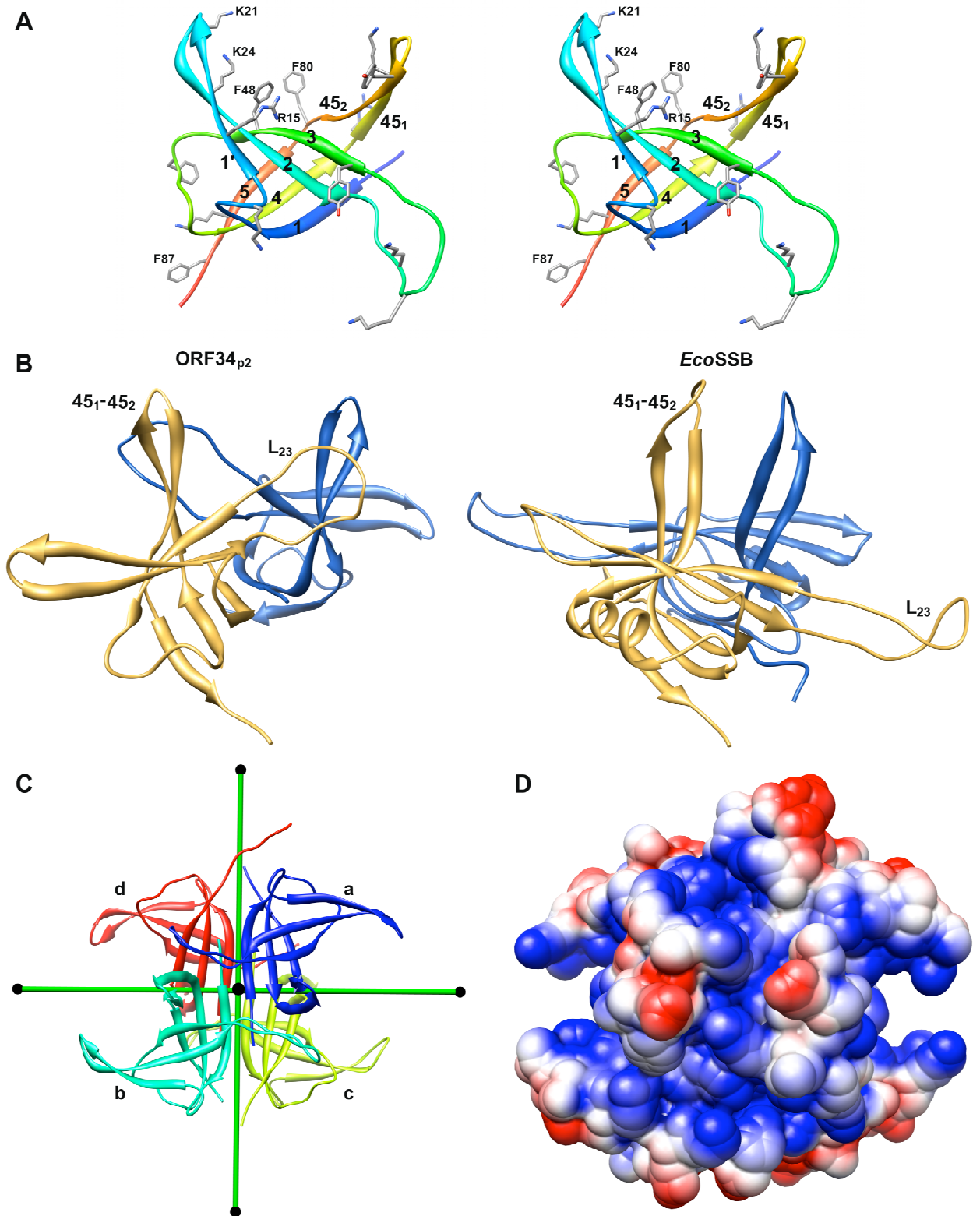
Figure 5

Figure 6

Supplementary Figure 1

Chapter 3: Perspectives

My thesis work, realized through a comprehensive approach utilizing genomic, proteomic, biochemistry, structural and molecular biology techniques, has been focused on the study of the structure and function relationships of proteins, like Sak and SSB, which are involved in the processing of the genomes of phages infecting the diary bacterium *Lactococcus Lactis.*

With regard to Sak3 from phage p2, on the basis of the results reported in the previous chapters, it clearly appears that a more accurate investigation of the helicase/topoisomerase activity and of the 3D structure might help to highlight additional functions of the protein as well as its involvement in processes, which have not yet been elucidated. Moreover, the involvement of both Sak wild type and Sak3 mutant in the inhibition of the AbiK mechanism may possibly be explained by a proteomic approach, through the evaluation of their expression levels at different phases of phage infection cycle. These studies might allow identifying the precise moment at which Sak can play its role during phage replication and DNA maturation processes. In parallel, biochemical functional studies in the presence of purified Sak protein and AbiK proteic product may help to the definition of the reaction(s) and function(s) involving these proteins. A similar experimental approach could be utilized also with Sak4 of phage ϕ31, a Sak protein not similar to Sak from phage UL36, and Sak3 from phage p2 (Bouchard and Moineau, 2004). Sak4, in fact, shows different properties that might aid to shed light on the functional properties of this category of proteins. Interaction studies with other putative functional protein partners could also be carried out with SSB of phage p2. The study of the proteins that are encoded in its recombinant cassette (Fig.10, Bouchard and Moineau, 2004) and their interaction with Sak proteins, might lead, on one side to the identification of a network of functionally related proteins responsible of the single steps of the recombination process, and on the other to the characterization of the molecular mechanisms that these same proteins have with the lactococcal antiviral system AbiK.

Bibliography:

Ackermann H W. 2007. Phages examined in the electron microscope. Arch Virol 152,227-243.

Ackermann H W. 1999. Tailed bacteriophages. The order Caudovirales. Adv. Virus Res. 51, 135-201.

Ackermann H W. and DuBow. 1987. Viruses of Prokaryotes 1 and 2. CRC Press, Boca Raton.

Aihara H, Ito Y, Kurumizaka H, Terada T, Yokoyama S and Shibata T. 1997. An interaction between a specified surface of the C terminal domain of RecA protein and double-stranded DNA for homologous pairing. J Mol Biol 274:213-221.

Allison GE and Klaenhammer TR. 1998. Phage resistance mechanisms in Lactic Acid Bacteria. Int Dairy J. 8:207-226.

Ayora S, Missich R, Mesa P, Lurz R, Yang S, Egelman E H and Alonso J C. 2002. Homologous-pairing Activity of the Bacillus subtilis Bacteriophage SPP1 Replication Protein G35P*. JBS 277: 3596-35979.

Ayora S, Stasiak A and Alonso J C. 1999. The Bacillus subtilis bacteriophage SPP1 delivers and activates the G40P DNA helicase upon interacting with G38P-bound replication origin. J Mol. Biol. 288:71-85.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 313:1709-1712.

Behnke D, Malke H. 1978. Bacteriophage interference in Streptococcus pyogenes I. Characterization of prophage-host systems interfering with the virulent phage A25. Virology. 85:118-128.

Bell C E. 2005. MicroReview: Structure and mechanism of Escherichia coli RecA ATPase. Mol.Microb. 58(2),358-366.

Benson F E, Stasiak A and West S C. 1994. Purification and characterization of the human Rad51 protein, an omologue of E.coli RecA. EMBO J. 13:5764-5771.

Bleuit Jill S, Hang Xu, Yujie Ma, Tongsheng Wang, Jie Liu and Scott W. Morrical. 2001. Colloquium: Mediator proteins orchestrate enzyme-ssDNA assembly during T4 recombination-dependent DNA replication and repair. PNAS 98: 8298-8305.

Bianco Piero R, Robert B.Tracy and Stephen C. Kowalczykowski. 1998. DNA strand exchange proteins: a biochemical and physical comparison. Frontiers in Bioscience 3, d570-603

Bidnenko E, Ehrlich SD, Chopin MC. 1995. Phage operon involved in sensibility to the Lactococcus lactis abortive infection mechanism Abi D1. J Bacteriolo 177:3824-3829.

Bidnenko E, Chopin MC, Ehrlich SD, Anba J. 2002. Lactococcus lactis AbiD1 abortive infection efficiency is drastically increased by a phage protein. FEMS Microbiol Lett. 214:283-287.

Bishop D K. 1994. RecA homoloug Dmc1 and Rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. Cell 79, 1081-1092.

Bochkareva Elena, Visar Belegu, Sergey Korolev and Alexey Bochkarev. 2001. Structure of major single-stranded DNA –binding domain of replication protein A suggests a dynamic mechanism for DNA binding. The EMBO journal. 20:612-618.

Bouchard J D, Dion E, Bissonnette F and Moineau S. 2002. Characterization of the two-component abortive phage infection mechanism AbiT from Lactococcus lactis. J Bacteriol. 184:6325-6332.

Bouchard J D and Moineau S. 2000. Homologous Recombination between a Lactococcal Bacteriophage and the Chromosome of its Host Strain. Virology 270:65-75.

Bouchard J D and Moineau S. 2004. Lactococcal Phage Genes Involved in Sensibility to AbiK and their Relation to Single-Strand Annealing Proteins. J. Bact. 186:3649-3652.

Boucher I, Emond E, Dion E, Montpetit D and Moineau S. 2000. Microbiological and molecular impacts of AbiK on the lytic cycle of *Lactococcus lactis* phage of the 936 and P335 species. Microbiology 146:445-453.

Boucher I and Moineau S. 2001. Phages of Lactococcus Lactis: An ecological and economical equilibrium. Recent. Research Dev. Virol. 3:243-256.

Bradley D. 1967. Ultrastructure of Bacteriophages and Bacteriocins. Bacteriological Reviews. 31:230-241.

Bleuit J S, Xu X, Ma Y, Wang T, Liu J and Morrical S W. 2001. Colloquium: Mediator proteins orchestrate enzyme-ssDNA assembly during T4 recombination-dependent DNA replication and repair. PNAS 98: 8298-8305

Brussow H. 2001. Phages of Dairy Bacteria. Annu. Rev. Microbiol. 55 :283-303.

Brussow H. and Desiere F. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insight from dairy phages. Molec. Microbiol. 39 :213-222.

Brussow H and Hendrix RW. 2002. MiniReview: Phage Genomics: Small is Beautiful. Cell. 108,13-16.

Burger K J and Trautner T A. 1978. Specific labelling of replication SPP1 DNA: analysis of viral DNA synthesis and identification of phage DNA-genes. Mol. Gen. Genet. 166: 277-285.

Calendar R. (edited by), The Bacteriophages 2nd edition, Oxford University Press .

Campbell A. 1994. Comparative molecular biology of lamboid phages. Ann. Rev. Microbiol. 48:193-222.

Carr FJ, Chill D, Maida N. 2002. The lactic acid bacteria: a literature survey. Crit Rev Microbiol. 28(4):281-370.

Casjens S, Hatfull G and Hendrix R. 1992. Evolution of dsDNA tailed-bacteriophage genomes. Semin. Virol. 3,383-397.

Clark AJ and Margulies AD. 1965. Isolation and characterization of recombinant-deficient mutants of Escherichia coli K12. Proc. Natl Acad. Sci; USA 53,451-459.

Chai S, Lurz R and Alonso J C. 1995. The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. J. Mol. Biol. 252, 386-398.

Chen Z, Yang H and Pavletich N P. 2008. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature 453,22.

Chopin M-C, Chopin A and Bidnenko E. 2005. phage abortive infection in lactococci: variations on a theme. Current Opinion in Microbiology. 8:473-479.

Chopin A, Deveau H, Ehrlich S D, Moineau S and Chopin M-C. 2007. KSY1, a lactococcal hage with a T7-like trascription. Virology 365,1-9.

Daly C, Fitzgerald GF and Davis R. 1996. biotechnology of lactic acid bacteria with special reference to bacteriophage resistance. Antoine Van Leeuwenhoek. 70:99-110.

Chowdhury R, Biswas SK, Das J. 1989. Abortive replication of choleraphage f149 in *Vibrio cholerae* biotype E1 Tor. J Virol 63:392-397.

Desiere F, Lucchini S, Brüssow H. 1999. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate cos-site Streptococcus thermophilus bacteriophage Sfi21. Virology. 260(2):244-53

Deveau H, Labrie S J, Chopin M-C and Moineau S. 2006. Biodiversity and Classification of Lactococcal Phages. App. Env. Microbiol. 72:4338-4346.

Dinsmore P.K, and Klaenhammer T R. 1997. Molecular characterization of a genomic region in a *Lactococcus* bacteriophage that is involved in its sensibility to the phage defense mechanism AbiA. J. Bacteriol. 179: 2949-2957.

Dinsmore P.K, O'Sullivan DJ, Klaenhammer T R. 1998. A leucine repeat motif in AbiA is required for resistance of *Lactococcus lactis* to phages representing three species. Gene. 212:5-11.

Djodjevic G M, and Klaenhammer T R. 1997. Genes and gene expression in *Lactococcus* bacteriophages. Int Dairy J. 7,489-509.

Domingues S, Chopin A, Ehrlich SD, Chopin MC. 2004. A phage protein confers resistance to the lactococcal abortive infection mechanism AbiP. J Bacteriol.186(10):3278-81.

Durmaz E and Klaenhammer T R. 2000. Genetic analysis of chromosomal regions of *Lactococcus lactis* acquired by recombinant lytic phages. Appl. Environ; Microbiol; 66:895-903.

Durmaz E, Klaenhammer TR. 2007. Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis. J Bacteriol. 189(4):1417-25.

Durmaz E, Klaenhammer TR. 2000. Genetic Analysis of Chromosomal regions of *Lactococcus lactis* Acquired by Recombinant Lytic Phages. Appl. Env. Microbiol. 66,895-903.

Edmond E, Holler B J, Boucher I, Vandenbergh P A, Vedamuthu E R, Kondo J K and Moineau S. 1997. Phenotypic and genetic characterization of the phage abortive infection mechenism AbiK from *Lactococcus lactis*. Appl. Environ. Microbiol. 63:1274-1283.

Egelman EH, Yu X, Wild R, Hingorani MM, Patel SS. 1995. Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexamerichelicases. Proc Natl Acad Sci U S A. 92(9):3869-73.

Egelman EH, and Stasiak A. 1993. Electron microscopy of RecA-DNA complexes: two different states, their functional significance and relation to the solved crystal structure. Micron. 24:309-324.

Forde A, Fitzgerals GF. 1999. Bacteriophage defence systems in lactic acid bacteria. Antoine Van Leeuwenhoek. 76:89-113.

Formosa T and Alberts B M. 1986. DNA synthesis dependent on genetic recombination: chracterization of a reaction catalyzed by purified T4 proteins. Cell 47, 793-806.

Fortier LC, Bouchard JD, Moineau S. 2005. Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. J Bact 187:3721-3730.

Fortier L C, Bransi A and Moineau S. 2006. Genome Sequence and Global Gene Expression of Q54, a New Phage Species Linking the 936 and c2 Phage Species of *Lactococcus lactis*. J Bact 188: 6101-6114.

Ganesan A T, AndersenJ J, Luh J and Effron M. 1976. Microbiology 1976 (Schlessinger D ed). Pp. 319-325., American Society of Microbiology, Washington D. C.

Garneau J E, tremblay D M, Moineau S. 2008. Characterization of 1706, a virulent phage from *Lactococcus lactis* with similarities to prophages from other Firmicutes. Virology, 373,298-309.

Gasior Stephen L, Heidi Olivares, Uy Ear, Danielle M.Hari, Ralph Weichselbaum and Douglas K.Bishop. 2001. Colloquium: Assembly of RecA-like recombinases: Distinct roles for mediator proteins in mitosis and meiosis. PNAS 98, 8411-8418.

George J W, Stohr B A, Tomso D J, Kreuzer K N. Colloquium: the tight linkage between DNA replication and double-strand break repair in bacteriophage T4. PNAS. 98,8290-8297.

Gething M J. 1997. Protein folding. The difference with prokaryotes. Nature 388, 329.

Gravey PA, Hill C and Fitzgerald GF. 1996. The lactococcal plsmid pNP40 encodes a third bacteriophage resistance mechanism, one which affects phage DNA penetration. Applied and Environmental Microbiology. 62:676-679.

Haaber J, Moineau S, Fortier L-C, Hammer K. 2008. AbiV, a novel antiphage abortive infection mechanism on the chromosome of *Lactococcus lactis* subsp. *Cremonis* MG1363. Appl. Env. Microbiol. 74: 6528-6537.

Haaber J, , Rousseau G M, Hammer K, Moineau S. 2009. Identification and characterization of the phage gene sav, involved in sensibility to the lactococcal Abortive Infection Mechanism AbiV. Appl. Env. Microbiol. 75: 2484-2494.

Harris L D and Griffith J D. 1989. UvsY protein of bacteriophage T4 is an accessory protein for in vitro catalysis of strand exchange. J Biol Chem. 264,7814-7820.

Harmon FG and Kowalczykowski SC. 1998. RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. Genes Dev. 12,1134-1144.

Hemphill HE, Whiteley HR. 1975. Bacteriophages of *Bacillus subtilis*. Bacteriol Rev. 39:257-315.

Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. roc Natl Acad Sci U S A. 96(5):2192-7

Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J.(1994)
CLUSTAL W: improving the sensitivity of progressivemultiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.

Hill C, Miller L A, and Klaenhammer T R. 1991. In vivo genetic exchange of a functional domain from a type IIA methylase between lactococcal plasmid pTR2030 and a virulent bacteriophage. J. Bacteriol. 173,4363-4370.

Horvath P, Coute-Monvoisin A-C, Romero D A, Boyaval P, Fremaux C, Barrangou R. 2008. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. J Food Microbiology. Doi:10.1016/j.ijfoodmicro.2008.05.30.

Iyer M.Lakshminarayan, Eugene V Koonin and L Aravind (2002). Classification and evolutionary history of the single-strand annealing proteins RecT, Red beta, ERF and RAD52. BMC Genomics 3:8.

Jarvis A W, FitzgeraldG F, mata M, Mercenier A, Neve H, Powell I A, ronda C, Saxelin M and Teuber M. 1991. Species and type phages of lactococcal bacteriophages. Intervirology. 32: 2-9.

Josephsen J, nyengaard NR, Vogensen FK and Madsen A. 1996. Plasmid-derived type II restriction-modification systems from *Lactococcus lactis*. Patent WO9625503.

Kagawa W, Kagawa A, saito K, Ikawa S, Shibata T, Kurumizaka H and Yokoyama S.2008. Identification of a Second DNA Binding Site in the human Rad52 Protein. JBC 283,24264-24273.

Klein H. 2000. A radical solution to death. Nature Genetics. 25, 132-134.

Koti J S, Morais M C, Rajagopal R, Owen BAL, McMurray CT and Anderson DL. 2008. DNA packaging motor assembly intermediate of Bacteriophage φ29. JMB 381,1114-1132.

Kuzminov A. 1999. Recombinational rtepair of DNA damage in E.coli and Bacteriophage λ. Microbial Mol Biol. Rev. 63,751-813.

Labrie S and Moineau S. 2002. Complete Genomic Sequence of Bacteriophage ul36: Demostration of phage Heterogeneity within the P335 quasi-species of lactococcal phages. Virology. 296:308-320.

Labrie S and Moineau S 2000. Multiplex PCR for the detection and the identification of lactococcal bacteriophages. Appl Environ. Microbiol. 66(3):987-94

Lansing M. Prescott, J P Harley, D A Klein. 1993. Microbiology. Second Editino. Edited by Wm. Brown Communications, Inc.

Lohman T M and Ferrari M E .1994. Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. .Annu. Rev. Biochem. 63,527-570.

Lohman T M and Overman L.B. 1985. Two binding modes in Escherichia coli single strand binding protein-single stranded DNA complexes:modulation by NaCl concentration. J Biol Chem 260: 3594-3603

Lucchini S, desiere F and Brussow H 1999. Comparative genomics of *Streptococccus thermophilus* phage species supports a modular evolution theory. J Virol 73:8647-8656.

Malkov VA, and Camerini-Otero RD. 1995.. Photo cross-links between single-stranded DNA and Escherichia coli RecA protein map to loops L1 (amino acid residues 157-164) and L2 (amino acid residues 195-209). J Biol Chem 270: 30230-30233.

Mahony J , Deveau H , Mc Grath S ,Ventura M , Canchaya C , Moineau S , Fitzgerald GF & Van Sinderen D. 2006. Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species
FEMS Microbiology Letters261:253-261.

Madera C, Monjardin C and Suarez JE. 2004. Milk contamination and resistance to processing conditions determine the fate of Lactococcus lactis bacteriophages in dairies. Appl. Environ. Microbiol. 70:7365-7371.

Masson J Y, Davies A A, Hajibagheri N, Van Dick E, Benson F E, Stasiak A Z, Stasiak A and West S C. 1999. The meiosis-specific recombinase hDmc1 forms ring structures ans interacts with hRad51. EMBO J. 18,6552-6560.

McEachern MJ and Haber JE. 2006. Break-induced replication and recombinational telomere elongation in yeast. Annu. Rev. Biochem. 75, 111-135.

McKay LL. 1983. Functional properties of plasmids in lactic streptococci. Antoine Van Leeuwenhoek. 49:259-274.

Mcllwraith M J, Hall D R, Stasiak A Z, Stasiak A, Wigley D B and West S C. 2001. RadA protein from *Archaeoglobus fulgidus* forms rings, nucleoprotein filaments and catalyses homologous recombination. Nucleic Acids res. 29,4509-4517.

Mcllwraith M J, Van Dyck E, Masson J Y, Stasiak A Z, Stasiak A, Wigley D B and West S C. 2000. Recostruction of the strand invasion step of double-strand break repair using human Rad51 Rad52 and RPA proteins. J Mol Biol 304,151-164.

167

Missich R, Weise F, Chai S, Lurz R, Pedré X and Alonso J C. 1997. The replisome organizer (G38P) of Bacillus subtilis bacteriophage SPP1 forms specialized nucleoprotein complexes with two discrete distant regions of the SPP1 genome.J. Mol. Biol. 270:50-64.

Moineau S. 1999. Applications of phage resistance in lactic acid bacteria. Antoine Van Leeuwenhoek. 76:377-382.

Moineau S, Fortier j, Ackermann HW, Pandian S. 1992. Characterization of lactococcal bacteriophages from Quebec cheese plants. Can. J. Microbiol. 38:875-882.

Moineau S and Levesque C. 2005. Control of bacteriophages in industrial fermentations. P286-296. In E Kutter and A Sulakvelidze (ed). Bacteriophages: biology and applications. CRC Press. Boca Raton, Fla.

Moineau S, Pandian S, Klaenhammer T R. 1994. Evolution of a lytic bacteriophage via DNA acquisition from the Lactococcus lactis chromosome. Appl. Environ. Microbiol. 60:1832-1841.

Mosig G. 1983 in Bacteriophage T4 eds. Mathews C K kutter E M, Mosig G and Berget P M (Am. Soc. Microbiol. Washington, DC) pp. 120-130.

Mortensen Uffe H, Christian Bendixen, Ivana Sunjevaric and Rodney Rothstein (1996). DNA strand annealing is promoted by yeast Rad52 protein. Proc.Natl.Acad.Sci.USA 93:10729-10734.

O'Connor L, Coffey A, Daly C, Fitzgerald GF. 1996. AbiG a genotipically novel abortive infection mechanism encoded by plasmid pCI750 of *lactococcus lactis* ssp. *Cremoris* UC653. Appl. Environ Microbiol 62:3075-3082.

Ortega ME, Gaussier H and Catalano CE. 2007. The DNA maturation domain of gpA, the DNA packaging motor protein bacteriophage lambda, contains an ATPase site associated with endonuclease activity. JMB 373, 851-865.

Ostergaard S, Brondsted L, Vogensen F. 2001. identification of a Replication Protein and Repeats essential for DNA Replication of the temperate Lactococcal Bacteriophage TP901-1. Appl. Env.Microbiol. 67,774-781.

Passy S I, Yu X, Li Z, Redding C M, Masson J Y, West S C and Egelman E H. 1999. human Dmc1 protein binds DNA as an octameric ring. Proc Natl. Acad Sci. U.S.A. 96,10684-10688.

Paques F and Haber JE. 1999. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Mol. Biol.Rev. 63,349-404.

Patel S.S and Picha K M. 2000. Structure and Function of hexameric helicases. Annu; Rev. Biochem. 69:651-97.

Perugino G, Valenti A, D'Amaro A, Rossi M and Ciaramella M. 2009. Reverse gyrase and genome stability in hyperthermophilic organisms. 37,69-73.

Ploquin M, Bransi A, Paquet E R, Stasiak A Z, Stasiak A, Yu X, Cieslinska A M, Egelman E H, Moineau S and Masson J-Y. 2008. Functional and structural basis for bacteriophage homolog of Human RAD52. Current Biology 18,1142-1146.

Poteete A R and Fenton A C. 1984. λ red-dependent growth and recombination pf phage P22. Virology 134:161-167.

Poteete A R and Fenton A C. 1993. Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages λ and P22. Genetics 134:1013-1021.

Prévots F, Tolou S, Delpech B, Kaghad M, Daloyau M. 1998. Nucleotide sequence and analysis of the new chromosomal abortive infection gene abiN of *lactococcus lactis* subsp *cremoris* S114. FEMS Microbiol Lett 159:331-336.

Prévots F, Ritzenthaler P. 1998. Complete sequence of the new lactococcal abortive phage resistance gene AbiO. J Dairy Sci 81: 1483-1485.

Proux C, van Sinderen D, Suarez J, Garcia P, Ladero V, Fitzgerald GF, Desiere F, Brüssow H. 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. J Bacteriol. 184(21):6026-36.

San Martín MC, Gruss C, Carazo JM. 1997. Six molecules of SV40 large T antigen assemble in a propeller-shaped particle around a channel. J Mol Biol. 268(1):15-20.

San Martín C, Radermacher M, Wolpensinger B. 1998. Three-dimensional reconstructions from cryoelectron microscopy images reveal an intimate complex between helicase DnaB and its loading partner DnaC. Current Biol. 6:501-509.

Savvides S. N, Raghunathan S, Futteree K, Kozlov A G, Lohman T M and Wksman G. 2004. The C-terminal domain of full-length E.coli SSB is disordered even when bound to DNA. Protein Science 13,1942-1947.

Sharan SK, Kuznetsov SG. 2007. Resolving RAD51C function in late stages of homologous recombination. Cell Div. 4:2-15

Snyder L.1995. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? Mol Microbiol 15:415-420.

Shinohara A, Shinohara M, Ohta T, Matsuda S, Ogawa T. (1998) Rad52 forms ring structures and co-operates with RPA in single strand DNA annealing. Genes Cells 3: 145-156.

Singleton Martin R, Lois M Wentzell, Yilun Liu, Stephen C. West and Dale B Wigley. 2002. Structure of the single-strand annealing domain of human RAD52 protein. PNAS 99:13492-13497.

Stasiak A Z, Larquet E, Stasiak A, Muller S, Engel A, Van Dyck E, West S C, and Egelman E. (2000). The humen Rad52 protein exists as a heptameric ring. Current Biology. 10, 337-340.

Story R M and Steitz T A. 1992. Structure of the RecA protein-ADP complex. Nature 355:374-376.

Story R M, Weber I T and Steitz TA. 1992. The structure of the *E.coli* RecA protein monomer and polymer. Nature 355:318-325.

Sung P, Klein H. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Molecular cell biology.Nature Review. 7: 739-750.

Symington LS. 2002. Role of Rad52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol. Mol. Biol; Rev. 66,630-670.

Szczepanska Agnieska K, Elena Bidnenko, Danuta plochocka, Stephen McGovern, S.Dusko Ehrlich, Jacek Bardowski, Patrice Polard, Marie-Christine Chopin. 2007. A distinct single-stranded DNA-binding protein encoded by the Lactococcus lactis bacteriophage bIL67. Virology 363:104-112.

Teuber M. 1995. The genus Lactococcus. In the genera of Lactic Acid Bacteria. Ed. BJ Wood, WH Holzapfel, 2:173-234. London: Blackie Academic and Professional.

Theobald L. Douglas, Rachel M. Mitton-Fry, and Deborah S. Wuttke. 2003. Nucleic Acid Recognition by OB-Fold Proteins. Annu. Rev. Biophys. Biomol. Struct. 32: 115-33.

Tuteja N and Tuteja R. 2004. Unraveling DNA helicases: motif, structure, mechanism and function. Eur. J Biochem. 271, 1849-1863.

VanLoock MS, Yu X, Yang S, Lai AL, Low C, Campbell MJ, and Egelman EH. 2003a. ATP-mediated conformational changes in the RecA filament. Structure. 11:187_196.

Van Regenmortel M H V. 1990. Virus species, a much overlooked, but essential concept in virus classification. Inervirology 31:241-254.

Von Hippel P H and Dalagoutte E. 2001. A general model for nucleic acid helicases and their "coupling" within macromolecular machines. Cell 104,177-190.

Weisberg RA, Enquist LW, Foeller C, Landy A.. 1983. Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. J Mol Biol. 170:319-42.

Weise F, Chai S, Luder G and Alonso J C. 1994. Nucleotide sequence and complementation studies of the gene 35 region of the Bacillus subtilis bacteriophage SPP1. Virology. 202: 1046-1049.

Wold M.S. 1997. Replication Protein A: A Heterotrimeric, Single-strand-DNA-Binding Protein Required for Eukaryotic DNA Metabolism. Annu. Rev. Biochem. 66: 61-92.

Yang S, Yu X, Seitz E M, Kowalczywski S C and Egelman E H. 2001. Archaeal RadA protein binds DNA as both helical filaments and octameric rings. J Mol 314, 1077-1085.

Yang S, Yu X, VanLoock MS, Jezewska MJ, Bujalowski W, Egelman EH. 2002. Flexibility of the rings: structural asymmetry in the DnaB hexameric helicase. J Mol Biol. 321(5):839-49.

Yassa D S, Chou K M and Morrical S W. 1997. Biochimie 79,275-285.

Ye J, Osborne AR, Groll M and Rapoport TA. 2004. RecA-like motor ATPases- lessons from structures. Biochem Biophys Acta. 1659:1-18.

Yu X, Jacobs SA, West SC, Ogawa T and Egelman E. 2001. Colloquium: Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. PNAS 98:8419-8424.

Remerciements

Je voudrais tout d'abord remercier les membres du jury qui ont accepté de juger ce travail et de se déplacer à Parme: Sylvain Moineau et Douwe Van Sinderen.
Je remercie également l'Université Franco-Italienne qui a financé mes voyages et mon séjour à Marseille.

J'adresse toute ma gratitude à Christian Cambillau et Mariella Tegoni pour leur soutien, leur générosité (illimitée), leurs exemples dans la vie professionnelle et quotidienne et leurs conseils (je garde encore la Camomille que Mariella m'a donnée pour me calmer et de très bons souvenirs des après-midi s passes au Biacore). Merci pour m'avoir permis de travailler à l'AFMB et d'avoir pu faire cette expérience fantastique.

Merci à Roberto Ramoni et Stefano Grolli qui m'ont permis de faire cette thèse en co-tutelle et m'ont toujours soutenu en alimentant ma passion pour la recherche. Merci aussi pour tous les bons conseils, les discussions scientifiques, les opportunités que vous m'avez données.

Merci à nos collaborateurs, en particulier à Patrick Bron et J-Y Masson qui ont enrichi nos résultats et à Claudio Rivetti pour m'avoir initiée a la microscopie à force atomique (AFM) avec patience et avec de bons conseils.

Je remercie également Université Franco-Italienne qui a financé mes voyages et mon séjour a Marseille.

Merci à Elisa et Virna qui, avec discrétion, m'ont été très proches dans me période en France et qui ont toujours le bon conseil et la bonne façon de regarder les choses.
Merci à tous les « habitants » de la cuisine du labo à Parme qu'on partage avec la section de Physiologie Vétérinaire : merci pour la bonne ambiance pendant le repas et pour les conseils culinaires et d'économie domestique (une des spécialisations de mon doctorat!).

Merci à toutes les personnes de l'AFMB qui ont collaboré avec moi à ce sujet : Valérie pour tous les conseils sur les techniques les plus différentes, Silvia pour la partie de cristallogenèse (cela mérite encore du café italien !!), Miguel pour la partie de cristallographie et Giuliano pour l'aide au Wyatt. À travers la qualité de votre travail et votre disponibilité vous m'avez fait apprécier et rendre compte de la valeur du travail en collaboration.

Je remercie très chaleureusement tous les gens du labo AFMB. J'adresse en particulier toute mon affection à Amandine pour l'amitié, la disponibilité et la patience quand je ne parlais pas un seul mot de français, à Julie qui a été ma complice au travail et dans la vie en France (les soirées a la patinoire, au cinéma ou devant un plat de spaghetti !!), à Laure, ma voisine de bureau pour les éclates , les sorties et aussi pour la soirée « carbonara ». Merci à Stephanie, toujour disponible et souriante, à Adeline et David pour les bons conseils sur le doctorat et pour toutes les fois qu'on a ri au labo. Merci a Marion pour le voyage en Italy, à Catarina pour les bons moments de la dernière période en France et merci à toutes les filles pour les soirées très amusantes !

Merci à mes amis italiens qui m'ont soutenu « a distance » : à Nicola, Michela, Alice pour les déplacements chez moi à la Ciotat, à Diana, Ennia, Deborah pour les conversationnes « philosophiques » au téléphone, à Susanna, Sara et Alessandra pour les mails d'encouragements , à Elena pour la communication en France (et pour l'aide avec le français) et à Lorena pour l'échange ade loin des idées sur le GREST…je garde tout cela dans mon cœur !

Et ma dernière pensée (mais c'est la premiere !) s'adresse à tous ceux qui ont suivi mon parcours et m'ont soutenu depuis toujours, en particulier les personnes les plus importantes dans ma vie : mes parents, ma sœur et Matteo. Merci pour les voyages ensemble (la Provence n'est pas mal, c'est vrai ?!), pour les week ends (trop courts), pour les appels (très longs), pour m'avoir motivé aussi dans les moments plus difficiles…Vous étiez toujours là pour moi…si je suis arrivé jusqu'ici, c'est vraiment grâce à vous…merci de tout mon cœur !

Ringraziamenti

In Italiano posso concedermi qualche parola in più…

Vorrei per primi ringraziare i membri della giuria che hanno accetttato di giudicare questo lavoro e di recarsi a Parma : Sylvain Moineau et Douwe Van Sinderen.
Ringrazio in ugual maniera l'Università Italo-Francese che ha finanziato i miei viaggi ed i miei soggiorni a Marsiglia.

Rivolgo tutta la mia gratitudine a Christian Cambillau e Mariella Tegoni per il loro sostegno, la loro generosità (illimitata), i loro esempi nella vita professionale e quotidiana, ed i loro consigli (conservo ancora la Camomilla che Mariella mi ha regalato per essere più tranquilla e dei bei ricordi dei pomeriggi passati al Biacore). Grazie per avermi permesso di lavorare all'AFMB e per avermi dato la possibilità di fare questa fantastica esperienza: lavorare con voi e con il vostro gruppo mi ha arricchito ed aperto la mente.

Grazie a Roberto Ramoni e Stefano Grolli che mi hanno permesso di fare questa tesi in co-tutela e mi hanno sempre sostenuto alimentando la mia passione per la ricerca. Grazie anche per tutti i buoni consigli, le discussioni scientifiche, le opportunità che mi avete dato.

Grazie ai nostri collaboratori, in particolare a Patrick Bron e J-Y Masson che hanno arricchito con il loro lavoro i nostri risultati ed a Claudio Rivetti per avermi iniziato alla microscopia a forza atomica con pazienza e buoni consigli .

Grazie ad Elisa e Virna che, con discrezione, mi sono state sempre vicine nei mie periodi in Francia ed hanno sempre indirizzato il mio sguardo verso la corretta visione delle cose dandomi buoni consigli nei momenti più duri . Grazie a Cinzia per l'affetto ed il minuzioso lavoro di segreteria.
Grazie in generale a tutta la (piccola) Sezione di Biochimica Veterinaria perché il lavoro con voi si tinge sempre dei colori dell'amicizia e della disponibilità.
Grazie agli « abitanti » della cucina del laboratorio di Parma, che dividiamo con la Sezione di Fisiologia Veterinaria, per la bella atmosfera durante i pasti e per i consigli culinari e di economia domestica (una delle specializzazioni del mio dottorato !).

Grazie a tutte le persone dell'AFMB che hanno collaborato al mio progetto : Valérie per tutti i consigli su ogni tecnica, Silvia per la parte di cristallogenesi (questo merita ancora del caffé italiano !), Miguel per la parte di cristallografia e Giuliano per l'aiuto al Wyatt. Attraverso la qualità del vostro lavoro e la vostra disponibilità mi avete fatto apprezzare e rendere conto del valore del lavoro di squadra.

Vorrei poi volgere un caloroso ringraziamento a tutte le persone del laboratorio AFMB. Rivolgo in particolare tutto il mio affetto a Amandine per l'amicizia, la disponibilità e la pazienza quando ancora non parlavo una parola di francese, a Julie che é stata mia complice al lavoro e nella vita in Francia (serate a pattinare sul ghiaccio, al cinema o davanti ad un piatto di spaghetti !!), a Laure, la mia vicina di scrivania per le risate, le uscite e la serata « carbonara ». Grazie e Stephanie, sempre disponibile e sorridente, ad Adeline e David per i buoni consigli sul dottorato e per tutte le volte che abbiamo riso in laboratorio. Grazie a Marion per i piacevoli viaggi in Italia ed a Catarina per i bei momenti dell'ultimo periodo in Francia e grazie a tutte le ragazze per le serate divertenti in compagnia !

Grazie ai miei amici italiani che mi hanno sostenuto « a distanza » : a Nicola, Michela, Alice per la loro presenza costante e i loro spostamenti a La Ciotat, a Diana, Ennia e Deborah per le lunghe conversazioni « filosofiche » al telefono, a Susanna, Sara ed Alessandra per le mail d'incoraggiamento, a Elena per la comunicazione tra Gaionesi in Francia (e per l'aiuto con il francese) a Lorena per lo scambio di idee sul GREST a distanza ed a Franca e Claudio per il sostegno e per i buonissimi tortelli che mi hanno fatto sentire a casa durante le mie colazioni….vi ho tutti nel mio cuore !

Il mio ultimo pensiero (ma é il primo !) và a tutti coloro che hanno seguito il mio percorso e mi hanno sostenuto da sempre, in particolare le persone più importanti della mia vita : i miei genitori, mia