A model for the RecA protein of *Mycoplasma synoviae*

Marbella Maria Fonseca¹*, Frank J.B. Alarcon²*, Ana Tereza de Vasconcelos² and Lucymara Fassarella Agnez-Lima¹

¹Departamento de Biologia Celular e Genética, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil.
²Laboratório Nacional de Computação Científica, Petrópolis, RJ, Brazil.

**Abstract**

In this work, we predict a structural model for the RecA protein from *M. synoviae* (MsRecA) by theoretical homology modeling and evaluate the occurrence of polymorphisms in this protein within several isolates of this species. The structural model suggested for MsRecA conserves the main domains present in MtRecA and EcRecA. The L1 and L2 regions showed six and three amino acid substitutions, respectively, which apparently do not affect the conformation and function of MsRecA. The C-terminal domain is shorter than that found in EcRecA and MtRecA, which may increase its capacity to bind dsDNA and displace SSB, compensating the absence of recombination initiation enzymes. The MS59 isolate RecA sequence showed one polymorphism which does not affect its functions since these belong to the same physical-chemical group.

**Key words:** DNA repair, recombination, RecA, *Mycoplasma synoviae*.

Received: May 4, 2006; Accepted: November 6, 2006.

Mycoplasmas are prokaryotes distinguished, phenotypically, from other bacteria due to their small size and their complete lack of a cell wall. Currently, there are more than 190 species of mycoplasmas described, several with medical and veterinary importance since they are pathogens of humans and domesticated animals (Maniloff, 1996; Razin et al., 1998).

One of the major characteristics that allow mycoplasmas to establish chronic infection is their great genomic flexibility. The chromosomes of these organisms are dynamic structures that frequently go through rearrangements, insertions, deletions, and inversions of genes or whole genomic segments. The majority of those chromosome rearrangements are usually induced by homologous recombination, mediated by the RecA protein (Rocha and Blanchard, 2002).

The RecA protein is the key component of bacterial homologous recombination. It is involved in the initial steps of this process, where it is responsible for the invasion and exchange of DNA filaments. This protein is also involved in recombinational repair and has a coprotease activity important for the induction of the SOS response (Cox, 2003). The *M. synoviae recA* gene is 990 nucleotides long and encodes a protein of 329 amino acids (MsRecA).

In *Escherichia coli*, the RecA protein (EcRecA) has 352 amino acids and its monomers are packed head-to-tail with an extensive monomer-monomer interface (Story et al., 1992). RecA presents three structural domains. The N-terminal domain comprises the residues 1-33 which are responsible for forming the monomer-monomer interface. The central core (residues 34-268) contains the ATP binding site, represented in the protein by the highly conserved signature motif GPRESSGKT. This region also has two loops, L1 (residues 157-164) and L2 (residues 195-209), thought to be the DNA-binding sites of the RecA protein (Malkov and Camerini-Otero, 1995; Lusetti and Cox, 2002). The residues 269-328 form the C-terminal domain, which is projected away from the body of the filament and, according to crystal packing, contacts a monomer of a neighboring filament. The terminal residues 329-352 are mostly highly negatively charged and form a tail that appears disordered in the crystal structure. The presence of these residues may regulate the direct binding of RecA to DNA through electrostatic repulsion of the DNA phosphate chain (Tateishi et al., 1992; Eggler et al., 2003). Since RecA is highly conserved, the aspects described above for EcRecA could be used for studies comprising other organisms.
Studies conducted by Kleven et al. (1975) showed that there is considerable variation among populations of the same *M. synoviae* isolate in their ability to produce clinical disease. This finding is corroborated by the work of Citti and Rosengarten (1997), as well as that of Narat et al. (1998), who verified that the hemagglutination-positive phenotype of *M. synoviae* induces experimental acute synovitis infection in chickens more frequently than does its hemagglutination-negative phenotype.

The presence of numerous combinations of antigenic phenotypes is due to the high oscillation frequency of each individual gene. This variation can be related to the on and off expression state and to the ability to produce genomic variation, either by size (insertion and deletion) or through nucleotide substitution (Robertson and Meyer, 1992; Lysnyansky et al., 2001). Among the already sequenced small-size genomes (less than 1 Mb), mycoplasmas exhibit high density of large repetitive regions (Rocha et al., 1999). The source of these repetitive regions is related to recombination processes, which in bacteria are generally associated with RecA action, as suggested by Cox and Roca (1997).

Considering the importance of the RecA protein for genome variation in *Mycoplasma*, the present work aimed to generate a theoretical structural model for the RecA from *M. synoviae* (MsRecA) and to evaluate the occurrence of polymorphisms of this protein in several isolates of this species.

The model for MsRecA was constructed by using the homology molecular modeling method based on templates from *Mycobacterium tuberculosis* (PDB access: 1g19, 1mo3) and *Escherichia coli* (PDB access: 2reb, 1u94), whose protein structures were experimentally resolved (Story et al., 1992; Datta et al., 2000; Datta et al., 2003).

Atomic coordinates of all bacterial RecA as available at RCSB (Berman et al., 2000) were used to define a structural framework for MsRecA models. All models and their parts were calculated using the program MODELLER 8v1 (Sali and Blundell, 1993; Marti-Renom et al., 2000) by the previous calculation of pairwise and multiple alignments between the template structures/sequences and the target sequence (RecA sequence from *M. synoviae*). To obtain the final model presented in this study, models that matched the lowest objective function score were chosen by both visual and stereochemical inspection. Stereochemical validation was performed by using the package PROCHECK v.3.5.4 (Morris et al., 1992; Laskowski et al., 1993).

The genomic DNA samples from six *M. synoviae* isolates, named MS35, MS59, MS443, MS541, MS542 and MSWVN, kindly provided by Dr. Laurimar Fiorentin (EMBRAPA, Brazil), were amplified by PCR (Mastercycler Gradient - Eppendorf), using primer pairs designed for the recA gene (Forward primer: TTAGGAGCAAAATGA TAGA; Reverse primer: TTATAATTTATATTCTTTTA). Cycling was as follows: after an initial denaturation step at 95 °C for 3 min, the reactions were subjected to 39 cycles of denaturation at 94 °C for 40 s, annealing at 45 °C for 90 s, and polymerization at 72 °C for 90 s. After that, an aliquot was analyzed by electrophoresis through a 0.7% agarose gel.

Positive samples were purified, and the resulting DNA was cloned in pGEM-T Easy vector (Promega) and used to transform *E. coli* DH10B-competent cells prepared by electroporation using standard procedures (Taketo, 1988).

DNA minipreparations of MsRecA clones (Birnboim, 1988) were sequenced using the enzymatic sequencing technique adapted from Sanger et al. (1977), through PCR with DYEEnamic ET - Dye Terminator cycle sequencing kit for Mega BACE (GE Healthcare, Life Sciences), following manufacturer's instructions.

Several independent clones of each isolate were sequenced and the consensus sequences were assembled using Phred Phrap software (Ewing and Green, 1998) and analyzed using Consed (Gordon et al., 1998). Other bioinformatics tools were also utilized, such as BLAST, ORF Finder and ClustalW 1.8 (protein weight matrix: GONNET).

Bacterial recombination is a complex pathway involving more than 40 enzymes, and is divided into four steps: initiation, strand exchange, migration of the Holliday junction and its resolution, with the possible participation of many alternative enzymes in each step (Amundsen and Smith, 2003). *M. synoviae* has a reduced recombinational apparatus, with only seven of the classical enzymes. The initiation pathways RecBCD and RecFOR are incomplete (Carvalho et al., 2005). These complexes are involved with RecA loading and stabilization in alternative pathways (Amundsen and Smith, 2003). RecA protein plays a pivotal role in this process. The mechanism of action of this enzyme includes: polymerization on ssDNA, forming a nucleoprotein filament; subsequently, dsDNA is bound and the search for segments homologous to the ssDNA is initiated; finally, homologous strands are exchanged, a new DNA duplex is formed, and ssDNA is displaced. In addition, RecA presents a coprotease activity that is important for LexA and UmuD cleavage during SOS response induction (Lusetti and Cox, 2002; Prévoist and Takahashi, 2004).

In this work, we calculated structural models for the *Mycoplasma synoviae* MsRecA protein using protein homology molecular modeling approaches with EcRecA and MtRecA as templates, whose crystal structures were determined experimentally (Story et al., 1992; Datta et al., 2000; Datta et al., 2003). Using the *E. coli* RecA protein as a reference, the percentage of identical amino acids residues in bacterial homologous proteins was found to range from 49% for *Mycoplasma pulmonis* to 100% for *Shigella flexneri*. Even though *M. tuberculosis* RecA is only 62% identical to *E. coli* RecA (Table 1), the structures of RecA proteins obtained were very similar (Lusetti and Cox, 2002).
The structure suggested for MsRecA (Figure 1) shows a structural pattern similar to the ones described for MtRecA and EcRecA. Table 1 shows the identity and strong similarity of RecA proteins from mycoplasmas, *E. coli* and *M. tuberculosis*. Despite the fact that *M. synoviae* (Mollicutes class) is phylogenetically closer to *M. tuberculosis* (Actinobacteria class) than to *E. coli* (Gammaproteobacteria class) (Doolittle, 1999), the amino acid sequence identity is higher between RecA proteins from *E. coli* (51%) than *M. tuberculosis* (49%) (Table 1), which may be a consequence of the high mutation frequency typical of mycoplasmas (Rocha and Blanchard, 2002). RecA from *Mycoplasma* shows the highest degree of variation in the C-terminal domain. However, the crucial amino acids involved in the monomer-monomer interface, are well conserved in MsRecA, such as the residues 97, 117-128 and 214-227 (Lusetti and Cox, 2002), likewise, the ATP binding sites (residues 67-74, 90-100 and 140-150), DNA binding sites (L1 and L2 regions) and the amino acid threonine (T) at position 73, which interacts with Mg$^{2+}$, are present (Story et al., 1992).

Six amino acids substitutions were observed at the loop L1 region of MsRecA when compared to EcRecA, some of them were also found in other *Mycoplasma* species. A significant substitution from Lys$^{152}$ to Glu$^{152}$ results in the inhibition of DNA repair activity (Nastri and Knight, 1994). However, in MsRecA, this deficiency is suppressed by a second substitution, Gly$^{160}$ to Lys$^{160}$. Substitutions at positions 155, 156, 159, 160 and 162 should not result in a reduction in MsRecA activity, since they occur between amino acids belonging to the same family.

Studies conducted by Hörtmanegel et al. (1999), using L2 EcRecA mutants had shown that positions 199 and 202 tolerate a number of substitutions of varying chemical character with no adverse effect on either RecA activity. The third substitution in the L2 region of MsRecA, from Glu$^{197}$ to Met$^{197}$, although occurring between amino acids from different physicochemical groups, also should not alter RecA activities, since a large number of mutations are tolerated at this position.

Another difference observed in the MsRecA model is the C-terminal domain, which is shorter than that found in EcRecA and MtRecA. Among bacterial RecAs, this region exhibits the lowest sequence conservation in relation to N-terminal and central regions. However, the deletion of the final residues of the C-terminal domain is observed in *Mycoplasma* species (data not shown). In *E. coli*, this do-

### Table 1 - Identity and similarity of amino acids between RecA proteins analyzed through CLUSTAL W.

<table>
<thead>
<tr>
<th>Species compared</th>
<th>Complete RecA protein</th>
<th>RecA C-terminal domain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identity$^a$</td>
<td>Strong similarity$^b$</td>
</tr>
<tr>
<td>EcRecA x MsRecA</td>
<td>62%</td>
<td>17%</td>
</tr>
<tr>
<td>EcRecA x MsRecA</td>
<td>51%</td>
<td>22%</td>
</tr>
<tr>
<td>MtRecA x MsRecA</td>
<td>49%</td>
<td>22%</td>
</tr>
<tr>
<td>MsRecA x MtRecA</td>
<td>67%</td>
<td>13%</td>
</tr>
<tr>
<td>MsRecA x MpRecA</td>
<td>41%</td>
<td>27%</td>
</tr>
</tbody>
</table>

Ec = *Escherichia coli*; Mt = *Mycobacterium tuberculosis*; Ms = *Mycoplasma synoviae*; Mb - *Mycoplasma hypopneumoniae*; Mp = *Mycoplasma pneumoniae*. $^a$Identity = means that the residues are identical in all sequences in the alignment. $^b$Strong similarity = means that conserved substitutions have been observed in a certain position of the alignment.

**Figure 1** - Molecular structure of MsRecA: A = N-terminal (red), central (green) and C-terminal (blue) regions of MsRecA; B = Alignment of RecA protein and structural domains of MsRecA: regions of β-sheets and α-helices are highlighted in dark and light gray respectively; N-terminal and C-terminal domains are underlined; L1 region is in the continuous box, the Lys161 is in bold; L2 region is in the discontinuous box.
main (residues 270-352) folds into three \( \alpha \)-helices and two \( \beta \)-sheets and can be related to the initial docking site for dsDNA during the recombination reaction (Aihara et al., 1997). Furthermore, this domain is also associated with coprotease activity (Liu et al., 1993). Yu et al. (2001) have observed that the C-terminal domain movement relative to the core domain may be responsible for the active or inactive state of the RecA filament. Recently, the X-ray crystal structure of the uncomplexed EcRecA protein has been determined in three new crystal forms at resolutions of 1.9 Å, 2.0 Å, and 2.6 Å. These new structures show significant variation in the orientation and conformation of the C-terminal domain, as well as in the inter-filament packing interactions, suggesting that this conformational variability of the C-terminal domain of RecA protein is important for many aspects of RecA functions (Xing and Bell, 2004).

Studies with EcRecA mutants, presenting deletions at the last 24 residues of the C-terminal domain (C-terminus), have shown that these mutants have an enhanced capacity to bind dsDNA \textit{in vitro} and to displace single-strand DNA binding protein (SSB) and, also, that these activities are dependent on the Mg\(^{2+}\) concentration. It has been proposed that the negatively charged C-terminus of RecA regulates the direct binding of RecA to dsDNA by electrostatically repelling the phosphate backbone of the DNA, since this C-terminal region has six negatively charged amino acids. In addition, the initiation proteins (such as RecBCD and RecFOR) might interact directly with the RecA protein during the filament nucleation process, altering RecA conformation so that the C-terminus is no longer inhibitory. These data sets suggest that the DNA binding regulation ability of RecA and/or its ability to compete with SSB would allow RecA to gain access to SSB-coated DNA only at the appropriate time, such as after a replication fork stalls, and that the C-terminus acts as a regulatory switch, modulating the access of double-stranded DNA to the presynaptic filament, thereby inhibiting homologous DNA pairing and strand exchange at low magnesium ion concentrations (Egglér et al., 2003; Lusetti et al., 2003a, 2003b). In MsRecA, the absence of the last 25 amino acids from the C-terminal domain (a region rich in negatively charged amino acids) could have implications for its functions since regulatory mechanisms, such as those described for EcRecA, can be affected. Considering these aspects, it is possible that the deletion observed in MsRecA can enhance its capacity to displace SSB and bind to dsDNA, compensating the absence of initiation enzymes such as RecBCD and RecFOR complexes.

MsRecA presents a similar distribution in electrostatic potential, showing a clear electric polarization (Figure 2), as described for other bacterial RecA proteins (Rajan and Bell, 2004). This can contribute to RecA monomer interactions that are important for RecA polymer formation, RecA stability, and folding or specific protein-protein recognition.

In this work, we also investigated the occurrence of RecA polymorphism in six isolates from \textit{M. synoviae}. Only one difference was observed in the RecA from the isolate MS59, which presents an amino acid substitution of Val for Ile at position 195 (data not shown). This change should not affect the L2 function, since both amino acids belong to the same physicochemical group. With these isolate samples it was not possible to detect significant differences in RecA proteins, as increased sampling is necessary for this kind of analysis, which aims to generate data that can be used for comparisons between recombination and pathogenicity aspects.

The high frequency of antigenic variation in lipoproteins, such as haemagglutinins, has been described as an important aspect of pathogenesis of mycoplasmas, since they are found on the cellular surface, an important aspect for bacterial pathogens that need to escape from immunological responses to survive (Chambaud et al., 1999). Moreover, these regions are constant targets of recombination events. In \textit{M. synoviae} these proteins are codified by the vlhA gene, formed by a promoter region, by part of a conserved 5’ coding sequence that occurs as a single copy, and by a remainder of the coding sequence that occurs as multiple copies. In the strain sequenced by the Brazilian Genome Project, the haemagglutinin genes are within a DNA segment of approximately 70 kb, containing a single gene and about 70 pseudogenes (Vasconcelos et al., 2005). Post-translational cleavage of the VlhA protein generates an amino-terminal part (the lipoprotein MSPB) and a car-

![Figure 2](image_url)
boxyl-terminal part (MSPA), (Noormohammadi et al., 2000; Allen et al., 2005). Recombination events between the 5′vlhA gene and pseudogenes in the genome generate changes in antigenic determinants in the carboxyl two-thirds of the MSPB molecule, as well as in MSPA, resulting in changes in the domains involved in the binding of M. synoviae to erythrocytes (Noormohammadi et al., 2000; Bencina, 2002).

The recombinational apparatus in M. synoviae is reduced and the molecular mechanism of vlhA recombination is still not well understood. Furthermore, classical site specific proteins (such as integrases and recombinases) were not found in this genome (Vasconcelos et al., 2005) and new mechanisms of illegitimate recombination, such as those involving RecA systems and restriction enzymes, have been described (Kusano et al., 2003; Handa and Kobayashi, 2005). It is, therefore, possible that the RecA-dependent recombination mechanism in M. synoviae is involved in vlhA antigenic variation and that structural and polymorphism analyses, as described in this study, can be useful to understanding the recombination mechanisms in M. synoviae.

Acknowledgments
We would like to thank Dr. Laurimar Fiorentin (EMBRAPA, Brazil, in memoriam) for having kindly supplied the DNA samples of M. synoviae isolates and Mr. Glenn Haves (from the American Language Program of the University of Georgia) for editing this manuscript. This work received financial support from CNPq (Brazil) LFAL, and a CNPq fellowship.

Abbreviations
EcRecA: RecA from Escherichia coli.
MtRecA - RecA from Mycobacterium tuberculosis.
MsRecA - RecA from Mycoplasma synoviae.
PDB - Protein Data Bank.
RCSB - Research Collaboratory for Structural Bioinformatics.

References


Internet Resources


Protein Data Bank (PDB), http://www.rcsb.org/pdb (March 15, 2005).