Phasing of RecA monomers on quasi-random DNA sequences

Alexander A. Volodin[a,*], Elena A. Smirnova[a], Tatjana N. Bocharova[a], R. Daniel Camerini-Otero[b]

[a] *Institute of Molecular Genetics of the Russian Academy of Sciences, Kurchatov sq., 2, 123182 Moscow, Russia*
[b] *Genetics and Biochemistry Branch, NIDDK, National Institutes of Health, Bethesda, MD 20892-1810, USA*

Received 7 December 2002; revised 9 May 2003; accepted 13 May 2003

First published online 27 May 2003

Edited by Takashi Gojobori

Abstract **We show that some arbitrarily chosen DNA sequences have the ability to influence the positioning of RecA monomers in RecA–DNA complexes. The preferential phase of binding of RecA monomers is shown to depend on the DNA sequence and its nucleotide composition. A simple rearrangement of bases in a limited DNA stretch influences the phasing of RecA monomers. On the other hand, that some features of DNA sequences interfere with the phasing on specific DNA sites demonstrates the existence of mechanisms for both positive and negative regulation of phasing on natural DNAs. The possible role of phasing of RecA monomers on DNA is discussed.**

Key words: RecA protein; Presynaptic complex; RecA monomer positioning; Homologous recombination; DNA strand exchange

1. Introduction

RecA protein-promoted DNA strand exchange is the central step in the process of homologous recombination in *Escherichia coli* and, as appreciated recently, an important process in DNA repair and replication (see for review [1,2]). RecA protein homologs with similar functions have been found in eukaryotes and archaebacteria [3,4].

The key intermediate in strand exchange is the RecA–single-stranded DNA (ssDNA) complex (presynaptic complex) – the nucleoprotein filament that is formed in the presence of adenosine triphosphate (ATP) or some other nucleotide cofactors. While current models of the strand exchange reaction are based on the assumption that there is a structural and functional equivalence of the nucleotides in the presynaptic complex [5,6], a more careful consideration of the problem suggests that this assumption is probably not true. In the presynaptic complex each RecA monomer interacts with three nucleotides of ssDNA. That is, any nucleotide in the ssDNA may be situated in one of three nucleotide binding sites in a given RecA monomer. Since there is no indication of a repetition either in the RecA protein amino acid sequence [7] or in its three-dimensional (3-D) structure [8], the microenvironment about these sites cannot be physically identical. Each of the three individual nucleotides bound to one of three different binding sites in a RecA monomer presumably finds itself in a physically distinct environment. The question of interest to us here is whether this difference plays a role in RecA-promoted reactions.

Theoretically, there are three possible phases for the binding of RecA monomers to any given DNA and these phases differ in the placement of the RecA monomers relative to the DNA sequence. Earlier we revealed that in complexes of RecA protein with fluorescent dye-tagged oligonucleotides there is preferential binding of RecA monomers in one of the three possible 'phases' and we called this phenomenon the phasing of RecA monomers [9]. Recently we have demonstrated that a specific DNA sequence – the chi site (a recombinational hotspot from *E. coli*) is capable of influencing phasing [10]. These findings attest to the fact that the different microenvironments of the nucleotides bound in the three different sites of a RecA monomer can at least influence one RecA-promoted reaction, its binding to ssDNA.

Since the nucleotide binding centers of a RecA monomer differ in their affinity for different bases, thermodynamically favorable modes of RecA monomer arrangements (phasing) should form for RecA complexes on a wide range of non-uniform DNA sequences. The aim of the present study is to validate this hypothesis. We have asked whether the capability to phase RecA monomers is inherent in an arbitrary DNA sequence. We suggest that such properties of DNA sequences may underlie mechanisms of regulation of homologous recombination or other RecA-promoted processes.

2. Materials and methods

RecA protein purification and oligonucleotide synthesis methods were described earlier [11].

A number of quasi-random DNA sequences, eight bases in length, were generated. Each of the sequences contained an equal number of all four bases. That is, the gross composition of each of the sequences was $A_2C_2G_2T_2$. The sequences chosen for this study contained no A-nucleotides in the first and the last positions. This arrangement guarantees a distinct separation of the quasi-random insertion from the A-nucleotides spacer (see Table 1).

We also took into consideration the homology of a chosen sequence to that of the chi site (GCTGGTGG). The theoretically maximum extent of homology of our sequence to the chi sequence is 62.5% (five bases out of a total of eight). For the present study in order to eliminate the occasional identity of our sequences with that of the chi site, sequences were chosen that contained not more than three nucleotides coinciding with the sequence of the chi site. That is, the extent of homology between the chosen sequence and chi site did not exceed 37.5%. The selected octamers were inserted in different

*Corresponding author. Fax: (7)-095-1960221.
E-mail address: volodin@img.ras.ru (A.A. Volodin).

Abbreviations: ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; DMS, dimethyl sulfate

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)00572-6

positions along the oligonucleotides labeled with fluorescent dye (tetramethylrhodamine) at the 3′-end (see Fig. 1A and Table 1).

RecA protein complexes with ss oligonucleotides were formed in the presence of adenosine-5′-*O*-(3-thiotriphosphate) (ATP-γ-S). The experimental approach was described earlier [9,10]. To estimate the reproducibility of the experimental approach, the experiments with some of the series of oligonucleotides were repeated several times. The modification profiles were reproducible and the standard deviations of the peak to peak amplitude ratios calculated for these experiments did not exceed 15% and in most cases were much lower.

3. Results

As it was demonstrated earlier, in complexes of RecA protein with ssDNA the guanine bases react with dimethyl sulfate (DMS) with the same efficiency as in free ssDNA. In contrast to that, the reactivity of guanine toward DMS in RecA protein complexes with double-stranded DNA (dsDNA) is substantially increased [9,12]. This effect was used in our previous studies to demonstrate that RecA monomers are phased in presynaptic complexes with fluorescently labeled and chi-containing oligonucleotides [9,10]. The outline of the experimental procedure is presented in Fig. 1. After formation of the presynaptic complex with an oligonucleotide containing a quasi-random insertion, a short oligonucleotide complementary to the control sequence of the former was added. Annealing of the oligonucleotides resulted in the formation of dsDNA in this region of the complex and, consequently, the appearance of a strong signal from the G bases modified after treatment with DMS. The generated modification pattern was shown earlier to depend on the positioning of RecA monomers [9,10].

Presented in Fig. 2 are examples of modification profiles obtained for an insertion chosen as described above and an insertion of a uniform DNA sequence (eight C bases). The phasing of RecA monomers manifests itself as a periodic variation of the modification efficiencies of the G residues in the oligonucleotide control regions (G1–G4 residues in Figs. 1 and 2). The modification pattern depends on the position of the insertion along the oligonucleotides (scans 1–4 in Fig. 2A). The efficiencies of the modification of G1 and G2 change in a coordinated manner in agreement with the 'in phase' (three nucleotides spacing) location of these bases. On the other hand, the efficiencies of the modification of G3 and G4 located 'out of phase' with G1 and G2 and with each other do not show such coordinated changes. In agreement with the periodicity of these changes, the overall modification pattern for all positions G1 through G4 reverts after three steps of the displacement of the insertion (compare profiles 1 and 4 in Fig. 2A).

To confirm that the random sequence is the origin of the phasing in Fig. 2A, we show that the insertion of a uniform sequence instead of the quasi-random gave rise to a modification pattern that was not dependent on the position of the insertion (Fig. 2B).

To conveniently compare the phasing induced by the various insertions, the ratios of the amplitudes of the signals of the modification of the G3 and G2 residues (see Fig. 2) were plotted versus the position of the insertion along the oligonu-

Table 1
List of oligonucleotides used in the formation of complexes with RecA protein

Series	Oligonucleotides
I. Chi site insertion	1. TACAAAAGCTGGTGGAA**CGACGGCCAGT**AAA
	2. TACAAAGCTGGTGGAAA**CGACGGCCAGT**AAA
	3. TACAAGCTGGTGGAAAA**CGACGGCCAGT**AAA
	4. TACAGCTGGTGGAAAAA**CGACGGCCAGT**AAA
II. Random 1 insertion	1. TACAAAAGTCAGATCAA**CGACGGCCAGT**AAA
	2. TACAAAGTCAGATCAAA**CGACGGCCAGT**AAA
	3. TACAAGTCAGATCAAAA**CGACGGCCAGT**AAA
	4. TACAGTCAGATCAAAAA**CGACGGCCAGT**AAA
III. Random 2 insertion	1. TACAAAACATGCGATAA**CGACGGCCAGT**AAA
	2. TACAAACATGCGATAAA**CGACGGCCAGT**AAA
	3. TACAACATGCGATAAAA**CGACGGCCAGT**AAA
	4. TACACATGCGATAAAAA**CGACGGCCAGT**AAA
IV. Random 3 insertion	1. TACAAAATACGTAGCAA**CGACGGCCAGT**AAA
	2. TACAAATACGTAGCAAA**CGACGGCCAGT**AAA
	3. TACAATACGTAGCAAAA**CGACGGCCAGT**AAA
	4. TACATACGTAGCAAAAA**CGACGGCCAGT**AAA
V. No G, T insertion	1. TACAAAACACACAATAA**CGACGGCCAGT**AAA
	2. TACAAACACACAATAAA**CGACGGCCAGT**AAA
	3. TACAACACACAATAAAA**CGACGGCCAGT**AAA
	4. TACACACACAATAAAAA**CGACGGCCAGT**AAA
VI. Uniform (C_8) insertion	1. TACAAAACCCCCCCCAA**CGACGGCCAGT**AAA
	2. TACAAACCCCCCCCAAA**CGACGGCCAGT**AAA
	3. TACAACCCCCCCCAAAA**CGACGGCCAGT**AAA
	4. TACACCCCCCCCAAAAA**CGACGGCCAGT**AAA
VII. T-spacer	1. TACTTTTGCTGGTGGTT**CGACGGCCAGT**AAA
	2. TACTTTGCTGGTGGTTT**CGACGGCCAGT**AAA
	3. TACTTGCTGGTGGTTTT**CGACGGCCAGT**AAA
	4. TACTGCTGGTGGTTTTT**CGACGGCCAGT**AAA
VIII. C-spacer	1. TACCCCCGCTGGTGGCC**CGACGGCCAGT**AAA
	2. TACCCCGCTGGTGGCCC**CGACGGCCAGT**AAA
	3. TACCCGCTGGTGGCCCC**CGACGGCCAGT**AAA
	4. TACCGCTGGTGGCCCCC**CGACGGCCAGT**AAA

The quasi-random insertions and control sequences are in bold, the quasi-random insertions are underlined.

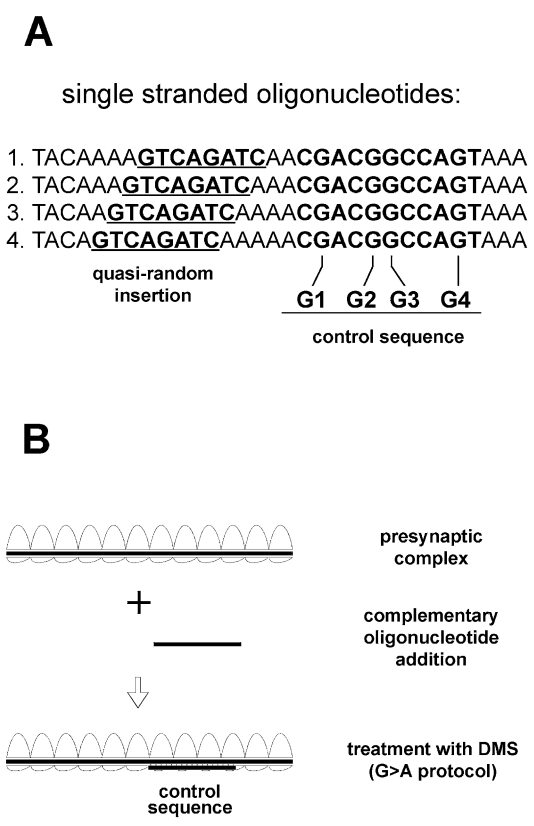

Fig. 1. Outline of the experimental procedure to study the phasing of RecA monomers on oligonucleotides with quasi-random insertions. A: An example of an oligonucleotide sequence that was used for the formation of presynaptic complexes with RecA protein. Underlined are the oligonucleotide regions containing the quasi-random insertions and control sequences. G residues in the control sequence are numbered. B: The formation of RecA protein complexes with oligonucleotides for chemical modification experiments. After formation of presynaptic complexes of RecA protein with the dye-labeled oligonucleotides, short unlabeled oligonucleotides complementary to the control sequence-containing regions were added. The resulting complexes were subjected to treatment with DMS.

cleotides. The data are presented in Fig. 3. In this kind of plot, the phase and period can be seen at glance while the amplitude of the change in the ratios may be considered a measure of the phasing efficiency.

Presented in Fig. 3A are plots for the insertions of the chi site, GCTGGTGG, and the three quasi-random sequences. Since in the absence of phasing the plot should be essentially flat (see below, Fig. 3B), the data in the figure demonstrate that all these sequences are able to influence the phasing of RecA monomers.

In the majority of cases, both in the complexes and in free oligonucleotides (see also [9,10]), modification of G2 base is much more efficient than the modification of G3 and the G3/G2 ratio is less than one. But in the complexes with RecA, at some positions of the insertions the modification efficiencies become comparable or the ratio is even reversed and exceeds one (Fig. 2A, plot 2). These situations correspond to the local maxima in the plots presented in Fig. 3. All three quasi-random DNA insertions presented in Fig. 3A (plots 2, 3 and 4) give rise to local maxima at different positions. This finding indicates that these three sequences tend to generate preferential binding of RecA monomers in different phases.

The range of variation of the G3/G2 ratio also changes depending on the insertion sequence. For some quasi-random insertions this range is comparable to that in the case of phasing on the chi site (compare plots 1 and 4 in Fig. 3A), but there is also the case when the plot is almost flat, demonstrating that the corresponding sequence cannot phase RecA monomers efficiently (Fig. 3A, plot 3). Thus, the phasing efficiencies differ for the various insertions and chi is among those sequences with the strongest capacity to phase RecA monomers.

Although the examples shown in Fig. 3A all have the same period of three nucleotides, examining the positions of the maxima reveals that they represent all three possible different phases of RecA binding with respect to position of the insertion. Thus, the plots demonstrate that the phase of the binding of RecA depends on the particular sequence of the insertion even though the nucleotide content of all the insertions is the same. Simple rearrangement of bases inside the insertion influences both the phase of RecA monomer binding and the efficiency of the phasing. That is, although the period is an inherent property of the RecA monomer, the arrangement of bases inside the insertion dictates the phase of the binding of RecA monomers.

To estimate the effect of base composition on the phasing of RecA monomers, some bases were excluded in one of the quasi-random sequences. Presented in Fig. 3B is the case when G and T bases in a quasi-random sequence were substituted by C and A bases, respectively, giving rise to the CACACAAC sequence (plot 2). As is obvious from the significant drop of the amplitude of the variation of the G3/G2 ratio, such substitutions decrease the efficiency of phasing. In the extreme case of a uniform DNA insert (CCCCCCCC), that is, when all three G, T and A bases were substituted by C bases, no phasing was observed (Fig. 3B, plot 3).

In the examples described above the DNA sequences were inserted in different positions within the A-nucleotides spacers (see Fig. 1). To check the influence of surrounding DNA sequences on phasing, the chi site was flanked with different spacers (Table 1, Series VII and VIII). As presented in Fig. 3C, using Ts instead of As as spacers conserves the general appearance of the plot. On the other hand, when chi was flanked with a C-nucleotides spacer three-base periodicity was not observed. We propose that this could be the result of base pairing between the G bases of the chi site and the C bases track of the spacer in the ssDNA. Such a partially duplex DNA would interact with RecA in a different manner than in a presynaptic complex. Another possible explanation is an interaction of RecA with the C track in the linker region that could interfere with phasing originating from chi.

These suggestions remain to be verified but, in any case, the result obtained demonstrates that a DNA sequence can contain not only peculiarities that phase RecA but also some features capable of quenching the phasing originating on neighboring DNA sequences. This property could provide a mechanism for a negative control of the phasing of RecA monomers in the presynaptic complex.

4. Discussion

As we demonstrated earlier for the case of the chi site, a nucleotide sequence as short as eight nucleotides can be capable of influencing the phasing of RecA monomers in complexes with ssDNA [10]. Our present results confirm this ob-

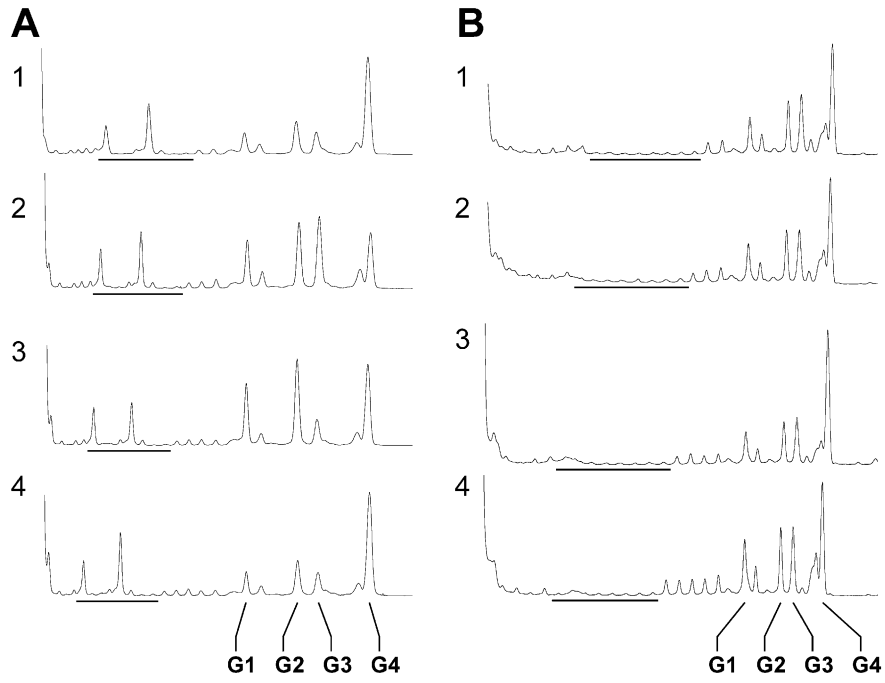

Fig. 2. DMS modification of oligonucleotide–RecA complexes reveals phasing of RecA monomers. Presented are the results of DMS treatment of the RecA protein complexes with oligonucleotides containing quasi-random (A) and uniform (B) insertions (Series II and VI of Table 1, respectively). In each scan the regions corresponding to quasi-random or uniform insertions are underlined. Electrophoretic bands corresponding to G residues in the control sequence region are numbered.

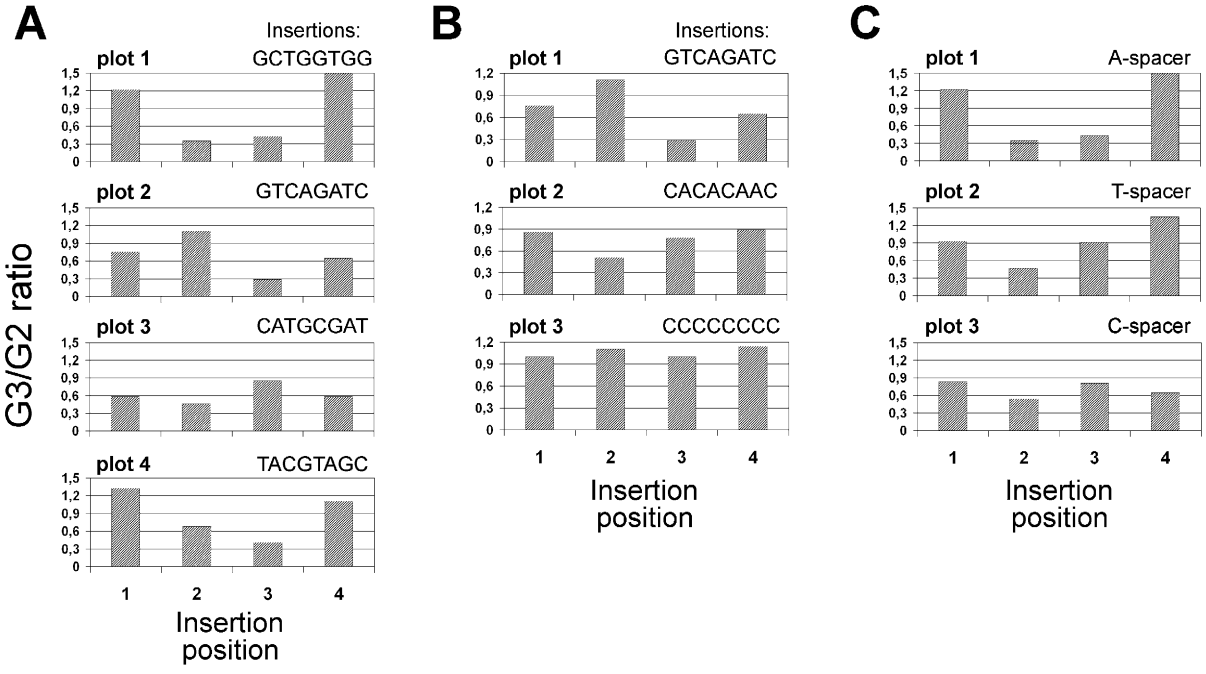

Fig. 3. Phasing of RecA on the chi site and on different quasi-random octamer sequences. Presented are the ratios of the G3 to G2 signal amplitudes as a function of the insertion position along the oligonucleotides (values 1 to 4 on the abscissa correspond to the positions of the insertion presented in Fig. 1A and Table 1). A: Phasing on the insertions of the chi site (1) and quasi-random sequences: GTCAGATC (2), CATGCGAT (3) and TACGTAGC (4). B: Influence of base composition on phasing. Presented are the results for phasing on (1) GTCAGATC sequence (the same as (2) in A), (2) the sequence with excluded G and T bases – CACACAAC (obtained by substituting T for A and G for C bases in the GTCAGATC sequence) and (3) a uniform sequence – CCCCCCCC. C: Influence of the spacer composition. The chi site was flanked by a (1) A-spacer, (2) T-spacer, and (3) C-spacer.

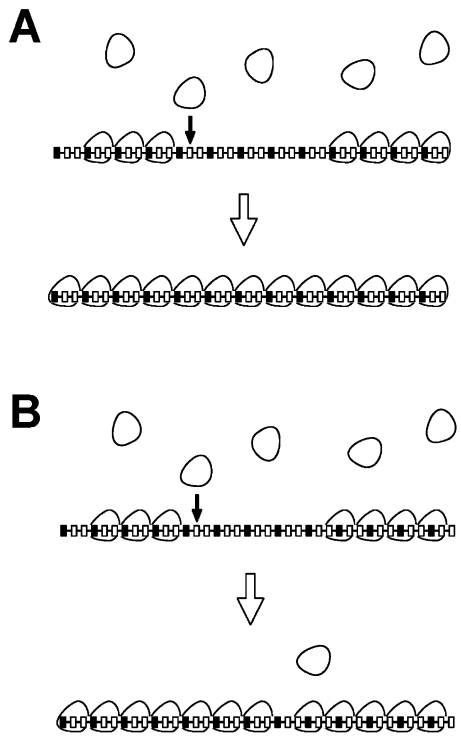

Fig. 4. Cooperative loading of RecA monomers onto ssDNA. A: Phased nucleation events result in a perfect RecA–ssDNA cofilament formation. B: Nucleations 'out of phase' result in formation of defected filaments. Presented is one of two possible modes of 'out of phase' nucleation. Linked rectangles designate nucleotides; each third is filled to highlight one of the possible phases of RecA monomer binding.

servation and demonstrate that natural DNA contains a number of features that influence the phasing of RecA monomers.

An influence of base content and sequence on the interaction of RecA with DNA has been reported previously [13–15]. Some preferential binding of RecA to DNA regions enriched with G and T bases has been demonstrated. The chi site was shown to be among the sequences with the strongest affinity for RecA and to stimulate strand exchange [15]. In agreement with these observations, in this study eliminating G and/or T bases from the sequences explored resulted in a drop of the efficiency of phasing.

The efficiency of binding RecA protein to ssDNA has been shown to be very sensitive to minute changes in the DNA sequence [16]. In the present work we document another aspect of the effect of the sequence of DNA on the binding of RecA to ssDNA. Inserting different DNA octamers of the same base content at varying positions along an oligonucleotide is physically equivalent to a simple rearrangement of the nucleotides in a limited DNA stretch inside the oligonucleotide. The results presented here demonstrate that such rearrangements influence the positioning of RecA monomers in the complex.

The phenomenon of phasing of RecA reveals a new possible influence of DNA sequence on RecA-promoted reactions. We propose several possible roles of the phasing of RecA. Earlier we proposed [10] that the ability of RecA monomers to be phased on certain DNA sequences and with the nucleotide bias found in the coding regions of the *E. coli* genome [17–19] may reflect a mechanism for phasing RecA monomers relative to the protein coding frame. Such a phasing of RecA monomers relative to open reading frames (ORFs) could be important in the discrimination of point mutations or some other DNA lesions depending on their positions in the codons of protein coding DNA regions.

Another possibility is regulating the process of RecA protein loading on ssDNA. The functionally active state of the RecA protein is its closely packed cofilament with ssDNA. In this complex each RecA monomer interacts with three nucleotides of ssDNA and forms a regular structure active in strand exchange reaction and in some other processes [2]. Presumably, disturbances of this regularity could unfavorably affect the activity of the presynaptic complex. Meanwhile, in the absence of a special mechanism, such disturbances must occur with a high probability each time growing clusters of RecA initiating at different points in the same ssDNA molecule come into contact. Indeed, a gap between such clusters can be filled perfectly with additional RecA monomers only in the case when the length of this gap is a multiple of three nucleotides (see Fig. 4). To prevent such an unfavorable arrangement of the growing clusters the positions of possible nucleation points must be preferentially located 'in phase', with a period of three nucleotides. In this case the processive expansion of separately nucleated clusters results in the formation of perfect RecA – ssDNA cofilaments (Fig. 4A). According to the data presented here DNA sequences can contain signals to position the nucleation points in phase with respect to each other.

We suppose that the above are not the only possible suggestions about the role for the phasing of RecA monomers. Due to the plurality and complexity of the processes in which RecA protein is a participant it is difficult to foresee all the possible consequences of this phenomenon. Nevertheless, we suggest that the DNA sequence-dependent positioning of RecA monomers may be an important structural and functional feature of the recombinational protein–DNA cofilaments that merits further investigation.

Acknowledgements: We acknowledge Dr. Oleg Voloshin for useful discussions. We are also grateful to Linda Robinson for technical help. The work was supported by the Russian Foundation for Basic Research (Grant 01-04-49690-a).

References

[1] Camerini-Otero, R.D. and Hsieh, P. (1995) Annu. Rev. Genet. 29, 509–552.
[2] Kuzminov, A. (2001) Proc. Natl. Acad. Sci. USA 98, 8461–8468.
[3] Seitz, E.M. and Kowalczykowski, S.C. (2000) Mol. Microbiol. 37, 555–560.
[4] Bianco I, P.R., Tracy, R.B. and Kowalczykowski, S.C. (1998) Front. Biosci. 3, 570–603.
[5] Howard-Flanders, P., West, S.C. and Stasiak, A. (1984) Nature 309, 215–219.
[6] Zhurkin, V.B., Raghunathan, G., Ulyanov, N.B., Camerini-Otero, R.D. and Jernigan, R.L. (1994) J. Mol. Biol. 239, 181–200.
[7] Sancar, A., Stachelek, C., Konigsberg, W. and Rupp, W.D. (1980) Proc. Natl. Acad. Sci. USA 77, 2611–2615.
[8] Story, R.M., Weber, I.T. and Steitz, T.A. (1992) Nature 355, 318–325.
[9] Volodin, A.A., Smirnova, H.A. and Bocharova, T.N. (1997) FEBS Lett. 407, 325–328.
[10] Volodin, A.A. and Camerini-Otero, R.D. (2002) J. Biol. Chem. 277, 1614–1618.

[11] Volodin, A.A., Smirnova, H.A. and Bocharova, T.N. (1994) FEBS Lett. 349, 65–68.
[12] Di Capua, E. and Muller, B. (1987) EMBO J. 6, 2493–2498.
[13] Wittung, P., Ellouze, C., Maraboeuf, F., Takahashi, M. and Norden, B. (1997) Eur. J. Biochem. 245, 715–719.
[14] Biet, E., Sun, J. and Dutreix, M. (1999) Nucleic Acids Res. 27, 596–600.
[15] Tracy, R.B. and Kowalczykowski, S.C. (1996) Genes Dev. 10, 1890–1903.
[16] Bar-Ziv, R. and Libchaber, A. (2001) Proc. Natl. Acad. Sci. USA 98, 9068–9073.
[17] Tracy, R.B., Chedin, F. and Kowalczykowski, S.C. (1997) Cell 90, 205–206.
[18] Colbert, T., Taylor, A.F. and Smith, G.R. (1998) Trends Genet. 14, 485–488.
[19] Biaudet, V., El Karoui, M. and Gruss, A. (1998) Mol. Microbiol. 29, 666–669.