J. Indian Inst. Sci., Jan.–Feb. 1993, **73**, 83–91.

RecA protein-promoted homologous pairing and strand exchange *in vitro*: A model to study the molecular mechanism(s) of gene targeting

K. MUNIYAPPA
Department of Biochemistry and Centre for Genetic Engineering, Indian Institute of Science, Bangalore 560 012, India.

Received on July 25, 1992.

Abstract

Homologous genetic recombination is ubiquitous. The insight(s) into the cascade of events that culminate in the production of heteroduplex DNA emanated from the studies done, *in vitro*, using the *E. coli* RecA protein paradigm. Our pioneering studies to understand the role of structural features of the *E. coli* and eukaryotic chromosomes on the mechanistics of homologous recombination, *in vitro*, and to evolve strategies for robust gene targeting has been discussed

Key words: Homologous genetic recombination, RecA protein, chromatin template, eukaryotic paradigm.

1. Introduction

The year 1992 marks the 27th anniversary of the epochal 1965 paper of Clark and Margulies which reported the first *Escherichia coli recA* mutants of homologous genetic recombination. Clark and Margulies account coincided with the discovery of yet another gene, *recG*, both were defined by the deficiency they caused in conjugational recombination[1, 2]. Subsequent studies in several laboratories showed that *recA* plays a crucial role in the related processes of recombinational repair, mutagenesis and in the cellular SOS response to DNA damage. The present essay will be concerned mainly with the RecA protein paradigm which has been so crucial not only to the success of initial experiments, but continues to supply ingenious methodology and conceptual leads. Biochemical analyses of homologous genetic recombination, although began a little over a decade ago, with the *E. coli* system, has reached a high level of sophistication in a short span of time. The ultimate goal of reconstituting a cell-free system, at least with the naked DNA substrates, is becoming a reality. I shall begin by outlining the initial developments and a brief outline of the current literature and then go on to sketch a summary of research accomplishments in this area from my laboratory.

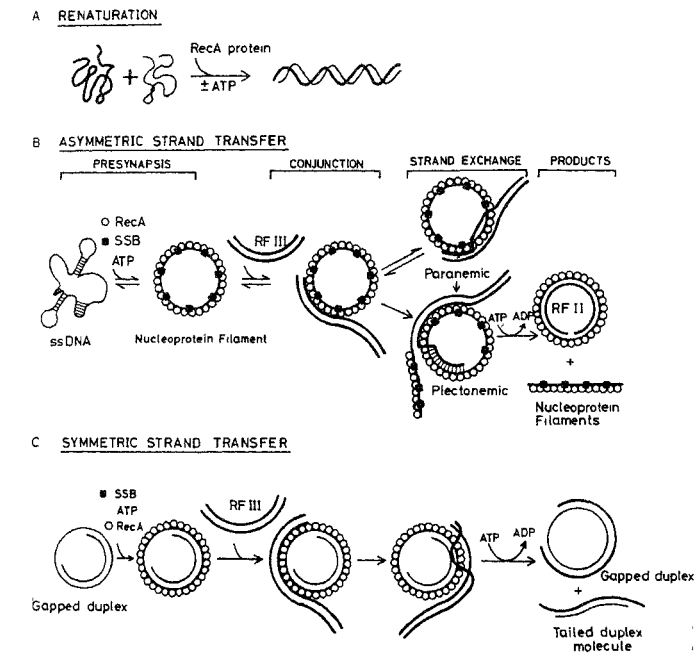

FIG. 1. Recombination reactions promoted by RecA protein. A: renaturation of complementary single strands; B: asymmetric or non-reciprocal strand exchange between ssDNA and linear duplex DNA; C: symmetric or reciprocal strand exchange.

The *recA* gene was cloned in 1976 by McEntee and its product was subsequently purified independently in several laboratories. The purified RecA protein promotes DNA strand-transfer reactions, *in vitro*, believed to mimic key steps in homologous genetic recombination, *in vivo*. The mechanistics of these recombination-like reactions promoted by RecA protein have been studied in great detail and pertinent literature is covered in recent reviews[3–6]

2. Biochemistry of homologous pairing and strand exchange promoted by RecA protein

RecA protein-promoted recombination-like reactions have been studied *in vitro* using the pairs of DNA substrates illustrated in Fig. 1. At catalytic amounts, RecA protein

promotes renaturation of complementary single strands which was among the first recombination reaction characterized (Fig. 1). However, the kinetics of renaturation catalyzed by RecA protein seems to differ from the reaction promoted by the class of single-stranded DNA-binding proteins. RecA protein-catalyzed renaturation reaction follows first-order kinetics. The renaturation activity of RecA protein is especially important in the context of its strand invasion activity and in view of the fact that it is an integral component of all models of homologous genetic recombination.

DNA strand transfer between circular single-stranded DNA (ssDNA) and linear duplex DNA promoted by RecA protein is the most extensively studied *in vitro* reaction (Fig. 1b). This combination of substrates has served as a useful model for elucidating the formal steps of homologous genetic recombination. This process consists of multistep reactions thought to be promoted by the products of more than 15 genes, *in vivo*, and is characterized by a number of kinetically distinct steps[7, 8]. The first step, termed *presynapsis*, is the cooperative binding of RecA protein to ssDNA to produce an highly ordered helical nucleoprotein filament. This important step which has been characterized by the good old biochemistry was not predicted *a priori* by any one of the models of genetic recombination. The nucleoprotein filament, which contains 1 monomer of RecA protein per 3-4 nucleotides, is an instrumental intermediate which directs all activities of RecA protein[9,10].

The formation of helical nucleoprotein filaments is abetted by single-stranded binding protein[9, 10]. The reaction requires the continuous presence of single-stranded DNA binding protein and its distribution in the nucleoprotein filament follows a regular repeating pattern[11,12]. Two important features of the nucleoprotein filament comprising RecA protein–DNA are worth emphasizing: its helical configuration and the DNA in the filament is stretched 1.5 times of that of B-form DNA resulting in 5.1 Å axial rise per nucleotide. Such an unprecedented property has never been documented for a DNA-binding protein[10, 13].

The homologous pairing and strand exchange reaction promoted by RecA protein is initiated by the binding of RecA protein to ssDNA. Within the confines of the *in vitro* reaction time schedule, and at physiological conditions, RecA protein binds preferentially to ssDNA over duplex DNA. This low affinity of RecA protein binding to duplex DNA has been attributed to a kinetic rather than thermodynamic barrier[14]. The active nucleoprotein filament promotes conjunction of DNA molecules, with naked duplex DNA, in homology-independent manner. The search for homologous sequences in the duplex DNA by the nucleoprotein filament is facilitated by the concentration of DNA into large networks, also referred to as coaggregates[3,4]. Aligment or *synapsis* of homologous sequences produce a joint molecule in which the two interacting partners are joined by the hydrogen bonds of paired bases. The formation of a joint molecule can occur anywhere along the length of the naked duplex DNA even in the absence of a nick or break in the region of homology. Such a side-by-side structure, with no intertwining of strands, is called a paranemic joint molecule[3]. The presence of a free end in the region of homology directs the reaction towards the formation of plectonemic joint molecules[3]. The precise mechanism(s) of this 'genome wide search' for homologous sequences in DNA and the conversion of

paranemic into plectonemic joint molecules is unknown. Following homologous alignment, the strand exchange ensues which is directional starting from 3′ end into the nucleoprotein filament. The rate of strand exchange is slow, when compared to other processes related to DNA metabolism, proceeding at 2-10 bp/s but it could be accelerated by single-stranded binding protein[3,4]. RecA protein-promoted strand exchange passes through heterologous insertions as long as 50–100 bp. The products of this reaction, the nicked circular heteroduplex and a displaced linear single strand, are not naked DNA molecules, in the model under consideration, but rather nucleoprotein filaments when sufficient amounts of RecA protein and single-stranded binding protein are included in the reaction. A number of assays have been devised to monitor the progress of this reaction and relevant information could be found in some of the recent reviews[3,6].

3. Strand exchange in relation to reciprocal transfer

The description given above on the formation of asymmetric heteroduplex DNA from single-stranded DNA and linear duplex DNA has provided information about the stages and the mechanism of homologous pairing promoted by RecA protein, although, these substrates are not the common ones found in the *in vivo* context. An assumption made early on was that the fundamental mechanism(s) that govern the formation of heteroduplex DNA either with three- or four-strand exchange is the same[3–6]. Therefore, the conclusion drawn from the asymmetric strand exchange, which is simpler to work with, was extrapolated to the symmetric reaction. The support for this comes from the similarity of rates and polarity of strand exchange, hydrolysis of ATP and the deposition of RecA protein on the DNA strands during the reaction[3–6].

The pair of substrates that have been used to analyze four-strand interactions include circular duplex molecule with a gap in one strand and a linear duplex molecule with a blunt end or DNA molecule with a tail (Fig. 1c). The reaction begins the same way as three-stranded reaction but the exchange continues into the duplex region and becomes symmetric producing, *in vitro*, the classical recombination intermediate postulated by Holliday. A crucial step in homologous genetic recombination is the resolution of Holliday recombination intermediates to mature heteroduplex products into either 'patch' or 'splice' type of recombinants. This pair of substrates has been utilized to identify a putative activity in *E. coli* cell-free extracts[15]. The presence of an endonuclease, coded by *ruvC*, specific for Holliday junction was identified which produced both 'patch' and 'splice' recombinant DNA molecules[16]. *ruvC* endonuclease is the first bonafide enzyme identified which has the intrinsic Holliday intermediate resolution activity. More importantly, defect in the resolution of Holliday junction has been correlated phenotypically to *ruvC* mutants of *E. coli*[16,17].

Rec BCD enzyme is a multifunctional enzyme required for DNA repair and genetic recombination; its precise role, however, has remained elusive[18]. To obtain an insight into its role, the research groups of Kowalczykowski and West supplemented the 4-strand reaction with RecBCD enzyme which led to the accumulation of a 'patch' type

of recombinant product suggesting the existence of an alternate pathway of resolution of Holliday intermediate in *E. coli*[19, 20]. An additional role, which was postulated several years ago, for RecBCD enzyme in homologous genetic recombination has been confirmed[21]. Consistent with a variety of models, a chi-dependent formation of joint molecules in the presence of RecA protein and SSB has been demonstrated[22]. These observations are compatible with the genetic data implicating RecBCD enzyme and chi as fundamental to the chi-stimulated recombination events *in vivo*[23].

4. Homologous pairing between chromatin templates and nucleoprotein filaments comprises protein–ssDNA

The DNA in both pro- and eukaryotes is contained in a condensed state and packaged by a multitude of structural and regulatory proteins and various degrees of supercoiling into different orders of chromatin to give rise to chromosomes. Despite this fundamental knowledge, all the molecular models of homologous genetic recombination and much important work dealing with the studies of homologous pairing and strand exchange, *in vitro*, have assumed these to be 'invisible' to the enzymatic machinery. We have therefore pioneered studies to address the issue(s) of how the nucleoprotein filaments of recA protein–ssDNA search for homology, establish synapsis and subsequently execute exchange of strands between duplex DNA molecules where one of the interacting partners is complexed with structural or regulatory proteins. Understanding of these events, *in vitro*, we believe will provide insights as to the context *in vivo* where homologous pairing and strand exchange should occur with duplex DNA associated with a multitude of scaffolding and regulatory proteins.

5. Prokaryotic paradigm

E. coli chromosome, like eukaryotic chromosomes, does not comprise DNA alone. The proteins that are found to be in association with DNA include HU protein, integration host factor, repressors, transcriptional activators and terminators. In addition, the *E. coli* chromosome is organized into distinct superhelical domains. These features have been believed to keep the DNA, *in vivo*, in a condensed state. Among the best studied *E. coli* chromosomal proteins are the histone-like proteins which share partial sequence homology with eukaryotic histones[24]. These are basic proteins and one of them, HU protein, is the most abundant form with about 60,000 monomers per actively dividing cell. HU protein is a heterodimer composed of two subunits, HU-1 and HU-2, and binds to DNA as a tetramer covering 60 base pairs of duplex DNA-producing nucleosome-like structure(s)[24].

RecA protein promotes homologous pairing of ssDNA with duplex DNA producing two types of synaptic structures; paranemic structures which are formed in the absence of a free end in the region of homology whereas the plectonemic joints are initiated when the duplex DNA, in the region of homology, harbors a free end[3]. We observed that the ability of RecA protein to promote homologous pairing with linear duplex DNA complexed with HU protein (nucleosome cores) was found to be

differentially affected. The formation of paranemic joint molecules was not affected whereas the formation of plectonemic joint molecules was inhibited from the start of the reaction. The formation of paranemic joint molecules between nucleoprotein filaments comprises RecA protein–circular single-stranded DNA and closed circular duplex DNA is believed to generate positive supercoiling in the duplex DNA. We found that the positively supercoiled DNA was inert in the formation of joint molecules but could be converted into an active substrate, *in situ*, by the action of wheat-germ topoisomerase I[25,26]. Consistent with these observations we also observed that unwinding of duplex DNA, which precedes homologous pairing, introduces torsional stress further suggesting a requirement for topoisomerase activity in homologous genetic recombination[27]. Although the detailed molecular events underlying homologous genetic recombination in *E. coli* remain to be elucidated, in view of these observations, one should consider duplex DNA in the context of supercoiling and nucleosome structures.

The nucleoprotein filament comprising recA protein–ssDNA directs all the subsequent steps in homologous pairing and strand exchange reactions promoted by RecA protein. The discovery that the binding of RecA protein to DNA results in the unprecedented stretching of the latter presented an intriguing conundrum[10]. To investigate this paradoxical difference between the spacing of bases in the nucleoprotein filament *versus* the target duplex DNA and the mechanism(s) of 'genome wide search', we explored the effect of heterologous contacts on conformation of DNA, and *vice versa*[28]. In the presence of wheat-germ topoisomerase I, nucleoprotein filaments induced a rapid, limited reduction in the linking number of heterologous circular duplex DNA. This limited unwinding of heterologous duplex DNA, termed heterologous unwinding, was detected within 30 s and reached a steady state within a few minutes. Nucleoprotein filaments that were formed in the presence of ATPγ S and separated from free RecA protein by gel filtration also generated a ladder of topoisomers upon incubation with relaxed duplex DNA and topoisomerase. The inhibition of heterologous contacts by 60mM NaCl or 5mM ADP resulted in a corresponding degree in heterologous unwinding. In reciprocal fashion the stability or number of heterologous contacts with nucleoprotein filaments was inversely related to the linking number of circular duplex DNA[28]. These observations show that heterologous contacts with nucleoprotein filament cause a limited unwinding of duplex DNA, and conversely that the ability of DNA to unwind stabiliszes transient heterologous contacts.

6. Eukaryotic paradigm

High on the wish-list of investigators studying mammalian cells is the ability efficiently to mutate any cloned gene whose function *in vivo* is unknown. The general strategy for achieving this goal is to target a mutation to a desired locus on the chromosome is by applying the principles of homologous genetic recombination. In recent years, homologous recombination has received increased attention because of the powerful tool that targeted recombination between a transfected DNA molecule and a cognate

chromosomal sequence promises to bring to molecular genetics. A major difficulty with approaches involving targeted recombination is the efficiency with which mammalian cells carry out homologous recombination. By contrast to fungi, bacteria and some parasites, mammalian cells efficiently integrate transfected DNA non-homologously into genomic loci thereby decreasing the frequency of the occurrence of positive recombinants. Efforts to develop a robust targeting system should examine the DNA at the locus of interest in an 'open' state. Indeed it has been illustrated that the state of chromatin determines the accessibility of DNA to various enzymatic processes.

To investigate the molecular mechanisms of gene targeting we chose the *Escherichia coli* RecA protein as a model since the system is well defined at the genetic and molecular levels. With this system we have focused on three related issues: (i) how does chromatin influence homologous pairing and strand exchange of duplex DNA, (ii) what parameters of eukaryotic chromatin determine the efficiency of 'gene targeting', and (iii) why mammalian cells show a greater propensity to promote non-homologous recombination over homologous recombination whereas some eukaryotes such as yeast and trypanosomes are mutants for non-homologous recombination.

To understand the molecular basis of gene targeting we studied the interaction of nucleoprotein filaments comprising single-stranded DNA and RecA protein with chromatin templates reconstituted from linear duplex DNA with histones. We observed that for the chromatin templates with histone/DNA mass ratios of 0·8 and 1·6 the efficiency of homologous pairing was indistinguishable from that of naked duplex DNA but strand exchange was repressed[29]. In contrast, the chromatin templates, with a histone/DNA mass ratios of 9·0 supported neither homologous pairing nor strand exchange. The addition of histone H1, in stoichiometric amounts, to chromatin templates quells homologous pairing. Furthermore, the pairing of chromatin templates with nucleoprotein filaments of RecA protein–single-stranded DNA proceeded without the production of detectable networks of DNA, suggesting that coaggregates are unlikely to be the intermediates in homologous pairing[29]. These observations have initiated an understanding of the involvement of chromatin structure in homologous genetic recombination with several implications for evolving strategies for efficient 'gene targeting' and for formulating models of genetic recombination.

References

1. CLARK, A. J AND MARGULIES, A. D. — Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12, *Proc. Natn. Acad. Sci. USA*, 1965, **53**, 451–459.
2. STORM, P. K., HOEKSTRA, W. P. M., DEHAAN, P. G. AND VERHOEF, C. — Genetic recombination in *Escherichia coli*: Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12, *Mutat. Res.*, 1971, **13**, 9–17.
3. RADDING, C. M. — Helical interactions in homologous pairing and strand exchange driven by RecA protein, *J. Biol. Chem.*, 1991, **266**, 5355–5358.
4. KOWALCZYKOWSKI, S. — Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange, *A. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 539–575.

5. COX, M. M. AND LEHMAN, I. R. Enzymes of general recombination, *A. Rev. Biochem.*, 1987, **56**, 229–262.

6. GRIFFITH, J. AND HARRIS, L. D DNA strand exchanges, *CRC Crit. Rev. Biochem.* (Suppl.), 1988, **23**, S43–S86.

7. MAHAJAN, S. K. Pathways of homologous recombination in *Escherichia coli*. In *Genetic recombination* (Kucherlapati, R. and Smith, G. R., eds), 1988, pp 87–140, American Society for Microbiology, Washington DC.

8. CLARK, A. J. AND LOW, K. B. Pathways and systems of homologous recombination in *Escherichia coli*. In *The recombination of genetic material* (Low, K., ed.), 1988, pp 155–215, Academic Press.

9. MUNIYAPPA, K., SHANER, S L., TSANG, S. S. AND RADDING, C. M. Mechanism of the concerted action of recA protein and helix-destabilizing proteins in homologous recombination, *Proc. Natn. Acad. Sci. USA*, 1984, **81**, 2757–2761.

10. FLORY, J., TSANG, S. S. AND MUNIYAPPA, K. Isolation and visualization of active presynaptic filaments of recA protein and single-stranded DNA, *Proc. Natn. Acad. Sci. USA*, 1984, **81**, 7026–7030.

11. MORRICAL, S. W. AND COX, M. M. Stabilization of recA protein–ssDNA complexes by the single-stranded binding protein of *Escherichia coli*, *Biochemistry*, 1990, **29**, 837–843

12. MUNIYAPPA, K., WILLIAMS, K , CHASE, J. W. AND RADDING, C. M. Active nucleoprotein filaments of single-stranded binding protein and RecA protein on single-stranded DNA have a regular repeating structure, *Nucleic Acids Res.*, 1990, **18**, 3967–3973.

13. STASIAK, A AND EGELMAN, E. H. Visualization of recombination reactions. In *Genetic recombination* (Kucherlapati, R. and Smith, G. R., eds), 1988, pp 265–308, American Society for Microbiology, Washington DC.

14. PUGH, B. F. AND COX, M. M. Stable binding of recA protein to duplex DNA, *J. Biol. Chem.*, 1987, **262**, 1326–1336.

15. CONNOLLY, B. AND WEST, S. C. Genetic recombination in *E. coli*: Holliday junctions made by RecA protein are resolved by fractionated cell-free extracts, *Proc. Natn. Acad. Sci. USA*, 1990, **87**, 8476–8480.

16. CONNOLLY, B., PARSONS, C. A., BENSON, F. E., DUNDERDALE, H. J., SHARPLES, G. J., LLOYD, R. G. AND WEST, S. C. Resolution of Holliday junctions *in vitro* requires the *E. coli ruvC* gene product, *Proc. Natn. Acad. Sci. USA*, 1991, **81**, 6063–6067.

17. DUNDERDALE, H. J., BENSON, F. E., PARSONS, C. A., SHARPLES, G. J., LLOYD, R. G. AND WEST, S C. Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC protein, *Nature*, 1991, **354**, 506–510.

18. SMITH, G. R. Homologous recombination in prokaryotes, *Microbiol. Rev.*, 1988, **52**, 1–28.

19. ROMAN, L. J., DIXON, D. A., AND KOWALCZYKOWSKI, S. C. RecBCD-dependent joint molecule formation promoted by the *E. coli* RecA and SSB proteins, *Proc. Natn. Acad. Sci. USA*, 1991, **88**, 3367–3371.

20. MÜLLER, B., BOEHMER, P. E., EMMERSON, P. T. AND WEST, S. C. Action of RecBCD enzyme on Holliday structures made by RecA, *J. Biol. Chem.*, 1991, **266**, 19028–19033.

21. STAHL, F. Special sites in generalized recombination, *A. Rev. Genet.*, 1979, **13**, 7–24

22. DIXON, D A. AND KOWALCZYKOWSKI, S. C. Homologous pairing *in vitro* stimulated by the recombination hotspot, Chi, *Cell*, 1991, **66**, 361–371.

23. SMITH, G. R. Conjugational recombination in *E. coli:* Myths and mechanisms, *Cell*, 1991, **64**, 19–27.

24. SCHMID, M. B. More than just 'histone-like proteins', *Cell*, 1991, **63**, 451–453

25. RAMDAS, J., MYTHILI, E. AND MUNIYAPPA, K. RecA protein promoted homologous pairing *in vitro*: Pairing between linear duplex DNA bound to HU protein (nucleosome cores) and nucleoprotein filaments of RecA protein–ssDNA, *J. Biol. Chem.*, 1989, **264**, 17395–17400.

26. MUNIYAPPA, K., RAMDAS, J., MYTHILI, E. AND GALANDE, S. Homologous pairing between nucleosome cores on a linear duplex DNA and nucleoprotein filaments of RecA protein–ssDNA, *Biochimie*, 1991, **73**, 187–190.

27. MUNIYAPPA, K , RAMDAS, J., MYTHILI, E. AND GALANDE, S. Evidence for template supercoiling during homologous pairing promoted by *E. coli* RecA protein, *DAE Symp. Adv. Mol. Biol.*, 1990, pp 107–121, BARC, Bombay.

28. ROULD, E., MUNIYAPPA, K. AND RADDING, C. M. Unwinding of heterologous DNA by RecA protein during the search for homologous sequences, *J. Mol. Biol.*, 1992, **226**, 127–139.

29. RAMDAS, J , MYTHILI, E AND MUNIYAPPA, K. Nucleosomes on linear duplex DNA allow homologous pairing but prevent strand exchange promoted by RecA protein, *Proc. Natn. Acad. Sci. USA*, 1991, **88**, 1344–1348.

202/19 v4 28 4 93 YCM