Supplementary material 0: RecA-induced Homologous Recombination

Recombinases such as RecA (prokaryotes) or Rad51 (eukaryotes) are involved in the repair of DNA double strand breaks and in the incorporation or the exchange of genetic material. Homologous recombination has promising applications in gene therapy [1]. Rad51 has also been associated with cancer development [2–4]. All known recombinases can polymerize on DNA. They form helical filaments with similar geometric characteristics [5] that induce homologous recombination via closely related mechanisms.

The crystal structures of RecA nucleofilaments recently solved by Chen et al. [6] address two intermediate stages of the filament during homologous recombination. The first structure (PDB code 3CMW) features slightly less than a helical turn of the filament polymerized on a single-stranded DNA (ssDNA), in the presence of a non-hydrolyzable ATP analog. It corresponds to what is often referred to as "the active form" of the filament, which is active for ATP hydrolysis, for inducing LexA auto-cleavage and for DNA sequence recognition and strand exchange. The second structure (3CMX), where RecA is bound to a double stranded DNA, corresponds to the result of strand exchange. Between these two stages, genomic double-stranded DNA is searched for homology, a homologous portion of this DNA is incorporated into the filament, the two paired strands of the incorporated DNA separate, the complementary strand forms new Watson-Crick interactions with the single strand initially present in the filament and its former Watson-Crick partner is displaced. The whole process, pictured in Figure S0-1, corresponds to a minimum of four kinetic steps, starting from the encounter of two homologous regions of ssDNA and dsDNA [7,8]. The present study investigates, at the structural level, the initial phase of the reaction where the genomic dsDNA first interacts with the RecA/ssDNA filament and where sequence homology is searched for (arrow in Figure S0-1).

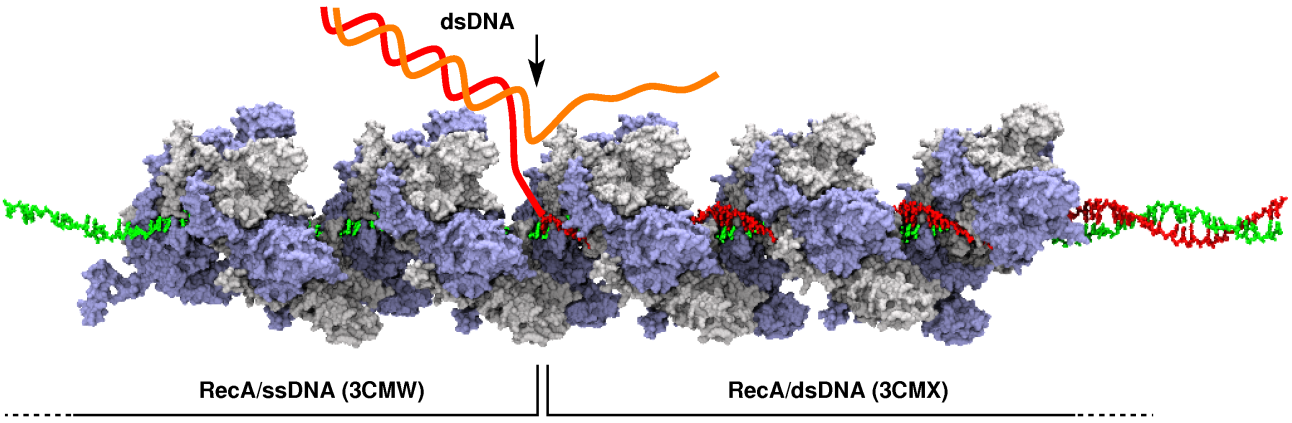

Figure 1: Schematic description of RecA-induced DNA strand exchange. A total of 32 RecA monomers (5 helical turns, with 6.2 monomers/turn) are represented in surface mode and alternatively colored in white and blue. The left part (RecA/ssDNA, 3CMW) corresponds to the filament state before strand exchange has taken place. The represented view has been constructed by multiplying two central monomers (8 times) and six central ssDNA bases (green, 10 times) of the crystal structure with PDB code 3CMW. The right part was built in a similar way from the RecA/DNA structure of PDB code 3CMX. It corresponds to the filament state after strand exchange, with the double stranded DNA represented in green and red. The incoming dsDNA is shown in orange for the homologous strand and red for the complementary strand, in a schematic representation. The arrow represents the stage of the reaction which is investigated in this study.

References

[1] Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., and Smithies, O. (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells.. *Nature,* **330**(6148), 576–578.

[2] Howard-Flanders, P., West, S. C., and Stasiak, A. (1984) Role of RecA protein spiral filaments in genetic recombination. *Nature,* **309**, 215–219.

[3] Kuzminov, A. (2001) DNA replication meets genetic exchange: chromosomal damage and its repair by homologous recombination. *Proc. Nat. Acad. Sci. U.S.A.,* **98**, 8461–8468.

[4] Prévost, C. and Takahashi, M. (2004) Geometry of the DNA strands within the RecA nucleofilament: role in homologous recombination. *Q. Rev. Biophys.,* **36**, 429–453.

[5] Egelman, E. H. (2001) Does a stretched DNA structure dictate the helical geometry of RecA-like filaments?. *J. Mol. Biol.,* **309**, 539–542.

[6] Chen, Z., Yang, H., and Pavletich, N. P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature,* **453**, 489–484.

[7] Bazemore, L. R., Folta-Stogniew, E., Takahashi, M., and Radding, C. M. (1997) RecA tests homology at both pairing and strand exchange. *Proc. Nat. Acad. Sci. U.S.A.,* **94**, 11863–11868.

[8] Xiao, J., Lee, A. M., and Singleton, S. F. (2006) Construction and evaluation of a kinetic scheme for RecA-mediated DNA strand exchange. *Biopolymers,* **81**, 473–496.

Supplementary material 1: BioSpring and underlying methods

Description of the BioSpring molecular simulation engine used for interactive simulations.

In this section we describe the approach underlying the *BioSpring* simulation engine. As indicated by its name, our approach is based on a spring network model, also referred to as elastic network [1].

Using experimental structural data (e.g. from the PDB), we first build a spring network model of our system from its particle positions. When two particles are closer than an arbitrary distance, a spring is created between these particles. Common cut-off values are between 7 and 15 Å and are abundantly documented and discussed in the literature [2]. The initial equilibrium spring length is called e. For all springs, a stiffness factor k is chosen either to match with experimental results (0.6 kcal.mol^{-1}.A^{-2}), or to match the expected rigidity of the studied structure. Then, at each time step of the simulation and for each particle p belonging to a selected dynamic particle set ($P_{dynamic}$), we compute the force applied on p, considering all the particles p' connected to p by a spring. If the distance $d_{pp'}$ is greater than the equilibrium distance $e_{pp'}$, an attractive force is applied on p, otherwise a repulsive one is applied.

$$\vec{F}_{\text{spring}}(p \in P_{\text{dynamic}}) = \sum_{p' \in \text{neighbor}(P)} k\left(d_{pp'} - e_{pp'}\right)\vec{u}_{pp'} \tag{1}$$

In the simulation setup, a subset of particles can be set as static (P_{static}), but continues to interact with dynamic particles ($P_{dynamic}$).

In the original spring network model, particles are considered as points with no radius. We augmented this model to take into account steric and electrostatic interactions between the particles in the spring network. We thus explicitly consider non bonded van der Waals and Coulomb interactions. The van der Waals forces F_{vdw} applied to a dynamic particle ($P_{dynamic}$) are computed taking into account all interactions between all particles (P_{all}=$P_{dynamic}$∪P_{static}) of the system. As described in the equations given below, forces are computed using the following approximation for the van der Waals interactions.

$$\vec{F}_{\text{vdw}}(p \in P_{\text{dynamic}}) = \sum_{p' \in P_{\text{all}}} \vec{u}_{pp'} 4\epsilon_{pp'} \left[\left(\frac{\sigma_{pp'}}{9d_{pp'}}\right)^9 - \left(\frac{\sigma_{pp'}}{7d_{pp'}}\right)^7\right] \tag{2}$$

The Coulomb forces $F_{coulomb}$ operate between all charged particles ($P_{charged}$⊂ P_{all}).

$$\vec{F}_{\text{coulomb}}(p \in P_{\text{dynamic}}) = \sum_{p' \in P_{\text{charged}}} -\frac{q_p q_{p'} \vec{u}_{pp'}}{4\Pi\epsilon_0 d_{pp'}^2} \tag{3}$$

Moreover, instead of or complementarily to the above pairwise interactions, spring network particles can be immersed into a potential field induced by the static target molecule. This is an extension of our *MyPal* software [3]. In the present work, we use Poisson-Boltzmann type electrostatic potential fields computed using the *APBS* software [4]. The potential forces F_{elect} act on the dynamic particles and originate from the potential map. They are defined by computing the gradient of the potential. In the following equation, we consider particle p belonging to the spatial cell $C_{i,j,k}$ of the potential grid, of size *delta*, and $E_{i,j,k}$, the value of the potential in this cell. We define the gradient as the mean of the difference between the $E_{i,j,k}$ potential and the potentials of the six adjacent cells, two for each axis. This method of computing the gradient reduces the bias related to the discretization of the grid.

$$\vec{F}_{\text{elec}}(p \in C_{i,j,k}) = \begin{bmatrix} \frac{(E_{i,j,k}-E_{i-1,j,k})+(E_{i+1,j,k}-E_{i,j,k})}{2.\text{delta}_x} q_p \\ \frac{(E_{i,j,k}-E_{i,j-1,k})+(E_{i,j+1,k}-E_{i,j,k})}{2.\text{delta}_y} q_p \\ \frac{(E_{i,j,k}-E_{i,j,k-1})+(E_{i,j,k+1}-E_{i,j,k})}{2.\text{delta}_z} q_p \end{bmatrix} \tag{4}$$

Finally, these forces are summed with an external force provided by the user through the graphical interface during the simulation.

$$\vec{F}(p \in P_{\text{dynamic}}) = F_{\text{spring}}(p) + \vec{F}_{\text{vdw}}(p) + \vec{F}_{\text{coulomb}}(p) + F_{\text{elec}}(p) + \vec{F}_{\text{user}}(p) \tag{5}$$

It should be stressed that the *BioSpring* approach is computationally cheap. The potential map is computed offline prior to the interactive simulation and computation of pairwise interactions between particles are optimized using a 3D grid, which allows us to compute the particle neighbor list with a constant complexity for each particle at each time step.

Hardware setup and typical configuration

All interactive *BioSpring* simulations were performed on a 3 GHz dual quad-core MacPro Apple computer. The *BioSpring* application itself is parallelized, although performance is not an issue for the systems presented here. Furthermore a multi-core workstation makes it possible to run all parts of the application on the same machine without performance issues. This includes visualization, haptic device server and calculation tasks. The molecular scene display was rendered using an NVIDIA Quadro FX 5600 graphics card. Two devices were simultaneously used for user interaction and tactile feedback: a Phantom Omni haptic device and a higher precision Phantom Premium 1.5A, both by Sensable Technologies. Scene navigation was achieved with a Spaceball device providing six degrees-of-freedom.

Overall, hardware requirements are modest. In our experience, this approach is viable using the smaller and more affordable haptic device, providing 3D positions and handling 3D directional force feedback. The *BioSpring* application can also easily be run in a desktop context, with minimal spatial requirements. In our experiments, the typical configuration was to mount the haptic devices, Spaceball, mouse and keyboard in front of a big screen. Figure S1-1A illustrates such a typical setup.

Interacting with the RecA models during a simulation

In order to interact with a *BioSpring* simulation in progress, the user manipulates a haptic device, in particular for selecting and moving particles in 3D space. Such a device with three degrees of freedom is intuitive and efficient for interacting with a complex three-dimensional object. Furthermore, the immediate haptic feedback when a particle is actually picked significantly improves the user experience and greatly helps to immerse the user in the molecular scene. The interactive simulation becomes intuitive and is comparable to dextrous manipulations such as those carried out in daily life.

In the particular case of the RecA nucleofilament, a complex manipulation had to be achieved and two haptic devices were needed. Figure S1-1B illustrates the task that was to be carried out by the user, controlling dsDNA (red and orange) with one hand, and RecA loops (blue) with the other. Figure S1-1C shows snapshots from a typical experiment.

A total of 22 simulations were performed on the system, first to adjust the spring cutoff value used to build the elastic network, which was finally set to 9 Å, then to sample different pathways that result from the interactive choices. The loop atoms where the force was applied varied at each test, together with the number of loops that were displaced (from one to three loops per simulation) and the direction and amplitude of the applied force, controlled via the haptic device. Analysis of the results involved monitoring the final energy, calculating the number of ssDNA/dsDNA contacts and evaluating the range of deformations for each loop.

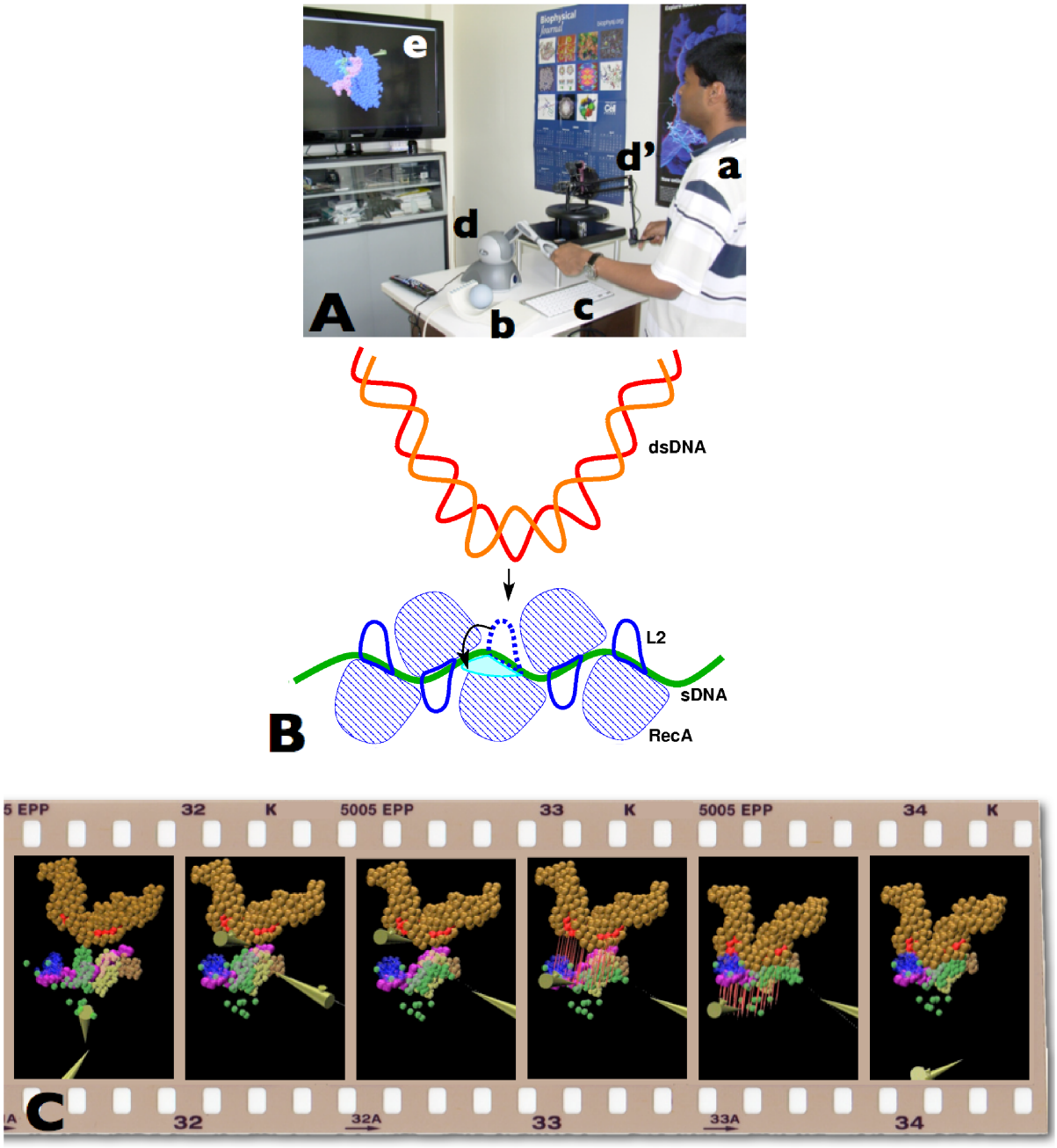

Figure 1: A. Typical hardware setup for an interactive exploration experiment based on the *BioSpring* simulation engine. The user (a) can navigate the scene with a Spaceball device (b). Further control can be achieved via a keyboard and a mouse (c). Interaction with the calculation and force feedback from *BioSpring* occur via two haptic devices (d,d'). The scene is shown on a large screen (e) with an avatar representing the user-driven tool (cone), and a visual rendering of the scene. B. Schematic representation of the nucleofilament assembly task. C. Sequential snapshots representing an interactive experiment.

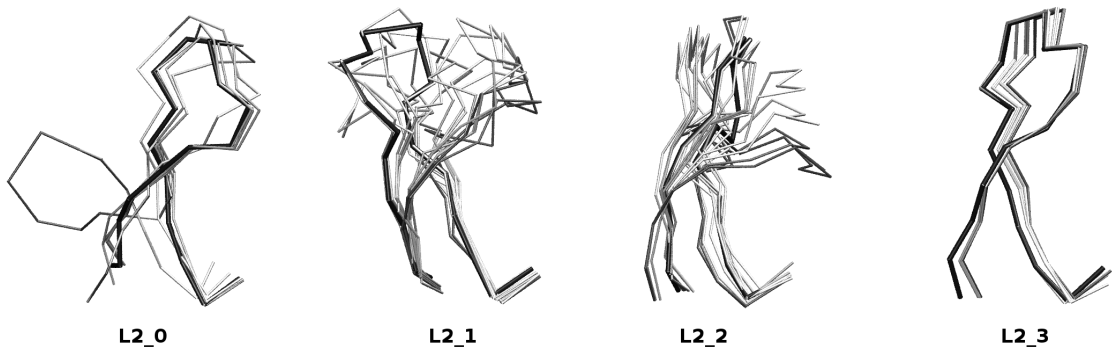

Figure 2: Superposition of loop conformations obtained during the interactive simulations. The grey scale indicates the number NC of dsDNA/ssDNA contacts (see Material and Methods), from 0 (white) to 25 (black).

Figure S1-2 displays the superposition of loop positions resulting from the final structural relaxation. It emphasizes that interactive simulation allows broad exploration of conceivable spatial positions. Furthermore, it can be noted that L2 loops from monomers two and three of the helical turn (L2_1 and L2_2) are more displaced than the others. This might suggest that at least two loops oppose a close approach of dsDNA and need to be displaced. When this is done, the number of ssDNA/dsDNA contacts can surpass that obtained by simple rigid body docking performed without L2 loops. Another observation is that loop positions corresponding to the highest number of contacts occupy spatially defined regions (from one to two regions for each loop, see Figure S1-2). This suggests (i) the essential role of a major motion of L2 loops during recombination and (ii) the existence of L2 loop positions that encourage accessibility of the incoming DNA.

References

[1] Chennubhotla, C., Rader, A. J., Yang, L.-W., and Bahar, I. (2005) Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Physical biology,* **2**(4), S173–80.

[2] Jeong, J. I., Jang, Y., and Kim, M. K. (Jan, 2006) A connection rule for alpha-carbon coarse-grained elastic network models using chemical bond information. *J. Mol. Graph. Model.,* **24**(4), 296–306.

[3] Delalande, O., Ferey, N., Laurent, B., Gueroult, M., Hartmann, B., and Baaden, M. (2010) Multi-resolution approach for interactively locating functionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device. *Pac Symp Biocomput,* pp. 205–15.

[4] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.,* **98**(18), 10037–41.

Supplementary material 2: Coarse grain DNA/DNA rigid body docking

The coarse grain DNA representation and force field used in the ATTRACT docking program [1,2] have been designed and tested for assembling protein/DNA systems [3]. We assess here the performance of this method and force field for coarse grain DNA/DNA rigid body docking.

The test consisted in re-assembling the single strand and the double strand parts of a DNA triple helix. The structure of this system comes from a model of triple helix dodecamer of sequence (dT.dA)×dT derived from NMR data [4]. The triplex was split into a Watson-Crick paired double-stranded component called *TA12* and a single strand *T3*, originally situated in the major groove of the duplex where it interacts *via* its Hoogsteen edge, in an anti-parallel orientation with respect to the (dT) strand of *TA12*.

A docking simulation was performed with ATTRACT (see Material and Methods) and each partner was kept rigid, with an internal conformation identical to that in the complex.

Analysis of the DNA/DNA docking test was based on the interaction energy *vs* root mean square deviation (RMSD) criteria generally used to evaluate the performance of docking simulations [3]. The analysis also involved clustering the solutions based on a RMSD threshold of 1 Å.

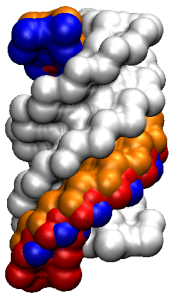

Figure 1: Result of a docking simulation between the *TA12* (grey) and *T3* (orange) components of a triple helix. The DNA strands are represented in surface mode constructed from the reduced representation. The red and blue strands correspond to predicted positions of strand *T3*, respectively ranked 1 and 2 based on interaction energy.

As can be seen in Figure S2-1 for the two highest-ranked clusters of solutions (in terms of interaction energy), the calculation successfully recovered geometries close to the correct triple helix geometry. The four best-ranked predictions, found within a tight 0.5 RT energy interval, all correctly present strand *T3* in the major groove of the duplex, with slight deviations that reflect adjustments due to the coarse graining. For example, among these solutions, exact phasing of the base pairs along the longitudinal axis was only observed for the structure ranked second. In this case, the deviation of 3.2 Å with respect to the reference structure arises from a slight rotation of the strand around the axis, which repells it from the nearby backbone of the complementary $(dA)_{12}$ strand. Alternatively, in the predictions ranked 1 (RMSD 6.1 Å), 3 (5.2 Å) and 4 (11.7 Å), the strand presents a longitudinal sliding of one (ranks 1, 3) or two (rank 4) base pairs along the DNA. Note that one base pair sliding corresponds to a deviation of 3.4 Å from the reference structure.

References

[1] Zacharias, M. (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.,* **12**, 1271–1282.

[2] Saladin, A., Fiorucci, S., Poulain, P., Prévost, C., and Zacharias, M. (2009) PTools: an opensource molecular docking library. *BMC Struct. Biol.,* **9**, 27–37.

[3] Poulain, P., Saladin, A., Hartmann, B., and Prévost, C. (2008) Insights on protein-DNA recognition by coarse grain modelling. *J. Comput. Chem.,* **29**, 2582–2592.

[4] Bornet, O. and Lancelot, G. (1995) Solution structure of a selectively 13C-labeled intramolecular DNA triplex. *J. Biomol. Struct. Dyn.,* **12**, 803–814.

Supplementary material 3: Interaction sites on the RecA/ssDNA surface: influence of the dsDNA structure

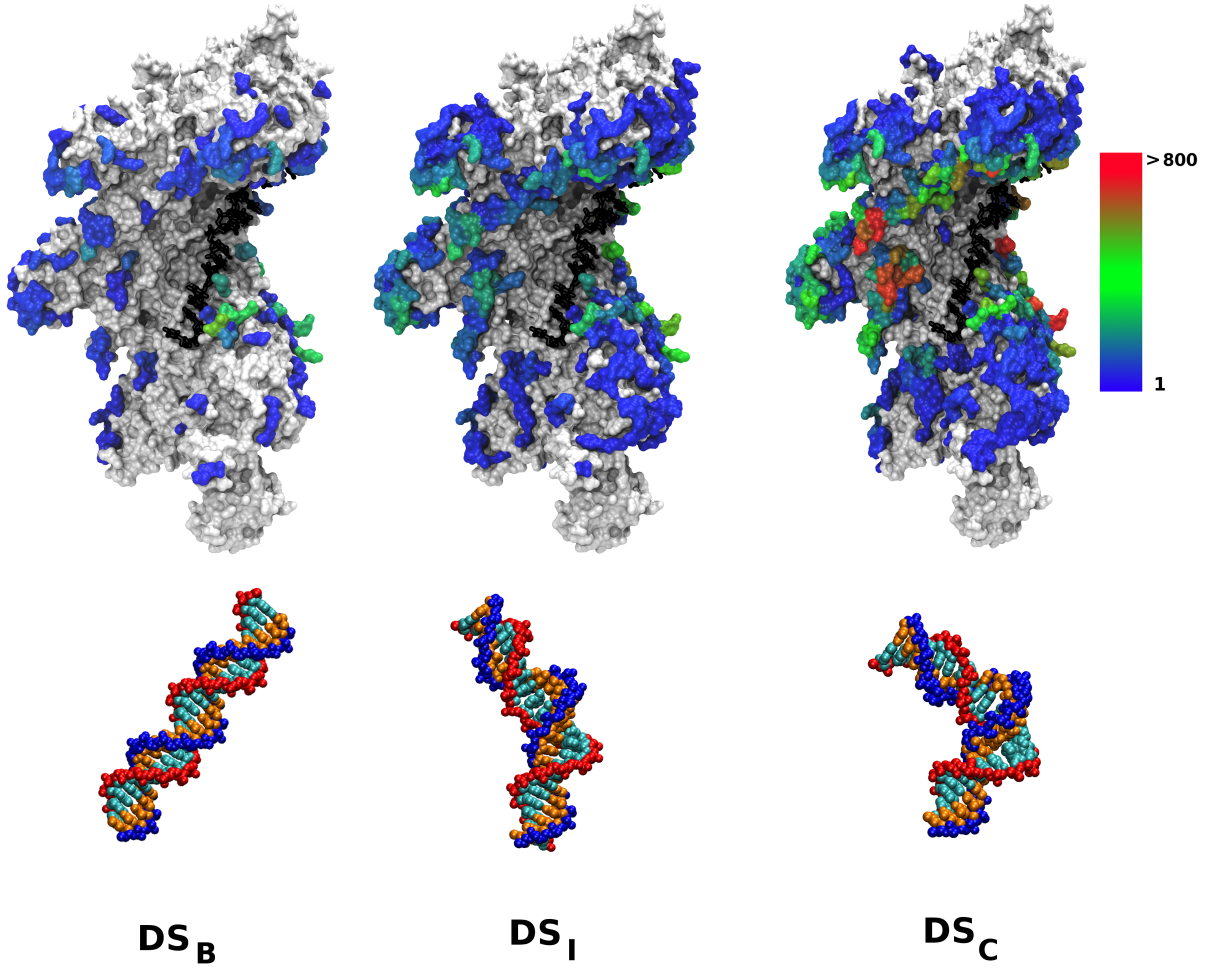

Figure 1: (top) Location of the dsDNA interaction sites on the filament/ssDNA surface, when dsDNA is the highly curved DS_C structure (left, from Figure 1 of the main article), an intermediary curved structure called DS_I (middle) or the straight DS_B structure (right). The DS_I structure was built from the structure of SRY-bound DNA (PDB code 1HRY), using the procedure described for DS_C in Material and Methods "DNA structures ". Analysis of the DS_B (resp. DS_I; DS_C) results was performed on a set of 6,485 (resp. 12,367; 27,401) predictions with interaction energies lower than -3.5 RT, where at least one contact is found between the central nine base pairs (8 to 16) and the filament. The protein filament is represented in surface mode constructed from the atomic representation. The ssDNA strand (black) is represented at atomic resolution in a stick representation. Amino acids are colored according to the number of times they are contacted by the dsDNA strands, cumulated over the set of predictions. The maximum value of this number is 514 for DS_B, 577 for DS_I and 1,142 for DS_C. White/grey patches correspond to contact-free amino acids. Amino acids are colored likewise in all three views, according to a color scale from blue (once) to red (more than 800 times). (bottom) The stuctures of DS_C, DS_I and DS_B are represented from left to right in van der Waals representation, with the purines in orange, the pyrimidines in cyan and the sugar-posphate backbones in red and blue.

Supplementary material 4: Details of the dsDNA/ssDNA/L2 interactions in the model system

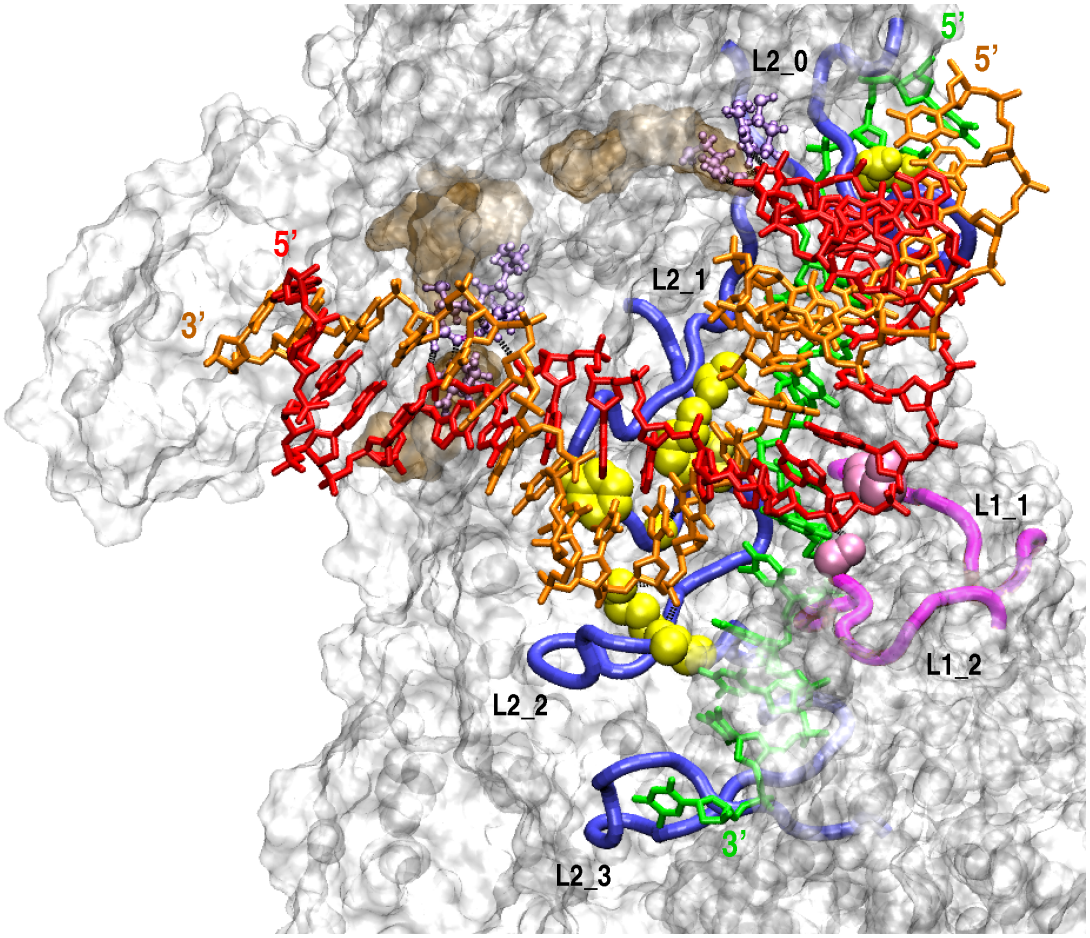

Figure 1: Details of the dsDNA/ssDNA/L2 interactions. The protein is represented as white and transparent surface. Specific amino acid side chains interacting with the dsDNA are represented in van der Waals mode in yellow for L2 residues (from top to bottom, Asn205, Thr1208, Lys1198, Asn2213, Phe1203, Gly1200, Lys2198, Ile2199) and pink for L1 residues (Met2164, Ser3162). The L1 loops participating in the dsDNA stabilization are represented as mauve tubes. The solvent accessible surface of protein regions 233 to 243, 1240 to 1243, experimentally identified as a DNA-contacted region [1], is shown in brown. Residues Lys245, Glu257 and Arg243 (from top to bottom, right to left), which hook the 3'-side of the complementary strand, are displayed in CPK (purple). The corresponding hooking region at the 5'-side contains residues Lys232, Lys1245, Gln1257 and Arg1243.

References

[1] Rehrauer, W. M. and Kowalczykowski, S. C. (1996) The DNA binding site(s) of the Escherichia coli RecA protein. *J. Biol. Chem.*, **271**, 11996–12002.