Correlated motion of protein subdomains and large-scale conformational flexibility of RecA protein filament

2012 J. Phys.: Conf. Ser. 340 012094

(http://iopscience.iop.org/1742-6596/340/1/012094)

Correlated motion of protein subdomains and large-scale conformational flexibility of RecA protein filament

Garmay Yu.[1], Shvetsov A.[1], Karelov D.[1], Lebedev D.[1], Radulescu A.[2], Petukhov M.[1,3], Isaev-Ivanov V.[1]

[1] Petersburg Nuclear Physics Institute, Russia

[2]Jülich Center for Neutron Science, Germany

[3]Research and Education Centre "Biophysics", St. Petersburg State Polytecnic University, St. Petersburg, Russia.

alexxy@omrb.pnpi.spb.ru

Abstract. Based on X-ray crystallographic data available at Protein Data Bank, we have built molecular dynamics (MD) models of homologous recombinases RecA from *E. coli* and *D. radiodurans*. Functional form of RecA enzyme, which is known to be a long helical filament, was approximated by a trimer, simulated in periodic water box. The MD trajectories were analyzed in terms of large-scale conformational motions that could be detectable by neutron and X-ray scattering techniques. The analysis revealed that large-scale RecA monomer dynamics can be described in terms of relative motions of 7 subdomains. Motion of C-terminal domain was the major contributor to the overall dynamics of protein. Principal component analysis (PCA) of the MD trajectories in the atom coordinate space showed that rotation of C-domain is correlated with the conformational changes in the central domain and N-terminal domain, that forms the monomer-monomer interface. Thus, even though C-terminal domain is relatively far from the interface, its orientation is correlated with large-scale filament conformation. PCA of the trajectories in the main chain dihedral angle coordinate space implicates a co-existence of a several different large-scale conformations of the modeled trimer. In order to clarify the relationship of independent domain orientation with large-scale filament conformation, we have performed analysis of independent domain motion and its implications on the filament geometry.

1. Introduction

There are wide range of supramolecular protein machines that are important for proper functioning of the living beings. Their function proceeds from their conformational flexibility, so that an understanding of the nature of the motion of a molecular machine is obviously important for understanding of its function. Recent advantages in modelling of the protein dynamics allows building

realistic models of the protein mobility at the atomic level. Although the current state of molecular modelling and limited computational resources put limitations on the system size and modelling time, making it impossible to simulate real supramolecular systems on their characteristic time-scale, it seems however possible to simulate the motion of their essential building blocks and try to approximate the large-scale motion/conformation from the motion of the smaller parts of the system.

Homologous recombination protein RecA plays a key role in DNA repair following radiation damage in bacteria. A number of X-ray crystallographic structures of RecA proteins exist [1-4], including complexes of *E.coli* RecA pentamer with ss- and ds-DNA [5]. Although the structure and function of a RecA and its analogues (such as RadA in archae and Rad51 in higher eucariotes) are very well studied, the molecular mechanisms for the enzymatic activity of these types of proteins is yet to be elucidated in detail [6].

Protein mobility on the subdomain scale, such as C-terminal domain and L2 loop, are thought to play an important role in ATP hydrolysis and DNA binding, respectively [6]. In all RecA isoforms where this process has been described formation of a presynaptic complex upon binding of ATP and either single- or double-stranded DNA brings about a significant increase in the helix pitch and a moderate decrease in the cross-sectional gyration radius of the protein filament. This change in the filament geometry is required to accommodate the stretched form of the DNA [7], which seems to be an essential intermediate in the process of homologous pairing and strand exchange. Similar large-scale conformational flexibility of the filament can be observed under conditions of high ionic strength in the absence of DNA, thus being inherent property of the enzyme, and is accomplished largely by N-terminal domain motion [8]. There are indications, however, that the transition between the two conformational states may be more complex and involve the motion of the C-terminal domain and the unstructured regions of RecA monomer, particularly the unstructured C-terminal tail [8;9]. Whether these or other possible conformational changes also facilitate homological pairing of DNA are unclear.

Deinococcus radiodurans is the most radiation resistant species currently known (LD_{37} = 5.5 kGy, compared to 0.17 kGy for *E. coli*). Compared to RecA protein from *E. coli,* its analog from *D. radiodurans* has higher affinity toward double-stranded DNA [10] and seems to exhibit a higher large-scale flexibility [11]. These unusual properties, along with its role in DNA repair following damage by ionizing radiation makes it an interesting object to compare with *E. coli* protein.

In this work we use molecular dynamics to analyze the conformational motion of RecA proteins from *E. coli* and *D. radiodurans* to elucidate on the possible mechanisms of the filament stretching associated with DNA binding.

2. Materials and Methods

2.1. Molecular Dynamics

Regularized models of X-ray crystallographic structures of RecA proteins were prepared based on the crystal structure of RecA from *E.coli* (PDB: 2REB) [3] using standard protocols of ICM-Pro software package (Molsoft LLC, USA) [12]. Two loops absent in the crystal structure (residues number 156-165 and 194-210) were rebuild using Monte-Carlo energy minimization as implemented in loop modelling tools of ICM-Pro. During the energy minimization only the standard torsion angles (φ, ψ, ω and χ_i) of amino acids were allowed to vary. ICM's default set of energy parameters (ECEPP/3 potential) for van-der-Waals, electrostatic, torsion energy interactions and hydrogen bonding were used in the calculations.

Molecular dynamics simulations (MD) were done with GROMACS version 4.0, the widely used software complex for molecular modelling and dynamics of proteins [13] using multiprocessor cluster of NRC Kurchatov Institute (Moscow, Russia). MD modelling involved several standard steps including (a) creation of the protein topology file and preparation of input data for GROMACS using pdb2gmx tool, (b) construction of hydration model for the RecA proteins in a periodic water box with the minimal distance to water box border of 10Å, (c) thermodynamic equilibration of hydrated

proteins and surrounding solvent, and (d) MD simulations at a constant temperature using the G53A6 parameter set and SPC water model [14]. The thermodynamically equilibrated system was used to simulate MD at 37°C with duration of 10 ns and MD step of 2 fs. Every 10 ps of MD time states of the model system were recorded for a future analysis. Neighbor searching was performed every 10 steps. The PME-Switch algorithm was used for electrostatic interactions with a cut-off of 1.5 nm as implemented in GROMACS. A single cut-off of 1.7 nm was used for Van der Waals interactions. Temperature coupling was done with the V-rescale algorithm [15]. Pressure coupling was done with the Parrinello-Rahman algorithm [16;17].

2.2. Rigid domain search

The task of determining rigid domains in a molecular dynamics trajectory analysis requires a measure of quality of a molecule segmentation. A molecular dynamics trajectory is an array of *N* frames, $T=\{T_i\}_{i=1}^{N}$. Each frame contains a set of *n* atomic coordinates $T_i=\{T_{ij}\}_{j=1}^{n}$. To determine whether a structure is "hard" or "soft", we can attempt to fit it to some reference structure *ref* with an LSQ fitting algorithm for each frame of the trajectory. Denoting the fitted structures with tilde, $T\to\tilde{T}$, we then calculate the RMSD $r(\tilde{T})=\frac{1}{Nn}\sum_{i,j}\|ref_j-\tilde{T}_{ij}\|^2$ that quantifies the changes of the structure over time. It should be noted that *r(T)* depends on the choice of the reference structure. Obviously, for any set of structures *T* the reference that minimizes *r(T)* is its average, consequently it is possible to use some kind of iterative algorithm to find the best reference structure, using the mean of the previous iteration as the reference for the current one. However, it seems reasonable to use some frame of the trajectory as the reference structure for fitting, and use the mean of the fitted structures for calculating RMSD. This measure does not provide the superposition with the minimal RMSD, but generally it is smaller for more similar structures, so that we used it for computation speed.

The principal points of the algorithm are:

1. define the number of domains *D;*
2. provide initial segmentation, $S=\{S_i\}_{i=1}^{n}$, where *n* is number of atoms in structure, and S_i is the index of the domain that i^{th} atom is reffered to, and
3. do global minimization of the function $F(S)=\sum_{k=1}^{D} r(\tilde{U}_k), U_k=\{U_{ki}\}_{i=1}^{N}, \{U_{kij}=T_{ij}\}_{j=1}^{n}\big|_{S_j=k}$, which is the sum of RMSD for all domains.

Global optimization was performed using simulated annealing algorithm realization provided by Gnu Scientific Library. Pairwise structure fitting was done using Kabsch algorithm [18].

2.3. Principal component analysis

PCA is the commonly used method of dimension reduction, that we use for describing and quantifying of the conformational changes of the modelled structures during MD simulation. The well-known method maps multidimensional input data into a lower-dimension space by a linear transformation, that diagonalizes the covariance matrix of the data. So it represents the data in the set of non-correlated variables that explains maximum possible (within linear transforms) part of the data variance.

We performed PCA using GROMACS molecular simulation package for analysis in atomic coordinate space, and Modular Data Processing library for analysis in the main chain dihedral angle coordinate space.

2.4. Filament construction

Filament was constructed using co-orientation of the central rigid domains of the neighboring monomers of the trimer. We evaluate the translation vector and the rotation matrix of the most identical domains of the neighbouring monomers using Kabsch algorithm and apply it to produce the coordinates of the monomers in the long filament.

2.5. SANS experiment and simulation

Samples of RecA protein from *D. radiodurans* in D_2O buffer containing 5 mM $MgCl_2$ and 20 mM TrisHCl (pH7.4) were prepared as described in [20]. SANS measurements were performed at 15°C on KWS-1 spectrometer (Juelich Centre for neutron science outstation at FRM-2 reactor, Muenich, Germany) using two detector positions in the range of momentum transter 0.006 - 0.2 Å^{-1}. Simulated SANS spectra were calculated using Debye formula, $I(q)=\int_0^{r_{max}} p(r)\frac{\sin(rq)}{rq}dr$, where the pair distance distribution *p(r)* was calculated by Monte-carlo simulations. Instrument resolution function was assumed to be a Gaussian with 10% FWHM. Fitting the data to scattering by long uniform helix was performed as described in [20].

3. Results and discussion

SANS spectrum of RecA from *D. radiodurans* (figure 1) exhibits 1/Q relationship in the low-Q region and a diffraction maximum near the 0.1 Å^{-1}, characteristic for the helical filaments formed by RecA proteins [19;20]. While the experimental data show a reasonable agreement with scattering by a long uniform helix [20], there is a significant systematic deviation of the model curve near the diffraction maximum (figure 0, inset), evident from high χ^2 value for the fit. Such a discrepancy was previously indicated for this RecA protein [11], but not observed for RecA proteins from *E. coli* and *P. aeruginosa* [8;20]. This could indicate that compared to the latter enzymes, RecA from *D. radiodurans* either has a much lower conformational mobility, so that the monomer structure is not averaged out by the thermal motion, making the uniform body model inadequate, or has a substantial polydispersity in the filament structure.

The molecular dynamics simulations of 10 ns in length were performed for *E. coli* and *D. radiodurans* RecA trimers with and without ATP. To visualize the relative mobility of the protein parts, C_α atoms mobility maps of the center monomer of the trimers was obtained. Color maps, shown in figure 2, reflect the deviation of the C_α atoms of the monomers fitted to the initial structure relative to their initial position. As expected, both proteins exhibit high mobility in the regions of L1 and L2 loops and C-terminal domain. Mobility of C-terminal domain was significantly higher in RecA protein from *D. radiodurans* as compared with the *E. coli* protein, and was restricted by binding of ATP (figure 2).

Further analyzing the trimer motion of *D. radiodurans* RecA we attempted to pinpoint the correlated motions in the protein monomer. We used PCA in the coordinate space of the carbon backbone atoms of the structures obtained during MD simulation fitted to the initial structure. The first principle component of the central monomer of the trimer clearly reveals correlated motion in all three main domains of RecA protein, comprising C-terminal domain and N-terminal α-helix motion and β-sheet deformation in the central domain (figure 3).

Mobility of N-terminal α-helix that forms the intermonomer interface with the two helices of the central domain of the adjacent monomer is directly linked to the large-scale conformational changes of the RecA polymer, particularly to the increase in the helical pitch of the filament when the presynaptic complex is formed [5;8]. Results of our MD simulation imply that N-terminal domain mobility is a part of a global conformational motion in the protein monomer that also affects orientation of C-terminal domain through allosteric interaction. Motion of C-terminal domain of RecA protein upon presynaptic complex formation has been established experimentally [9], which could imply that the conformations seen in MD model of the protein trimer are stabilized by cofactors (ATP and DNA) binding.

For further characterization of the monomer dynamics, we have searched for the rigid parts of the monomer. We found that RecA monomer can be divided into seven regions which motion should be described independently: two parts of N-terminal domain, entire C-terminal domain, rigid part comprising most of the central domain, loop L1, loop L2, and the small region adjacent to it. Figure 4 shows the positions of C_α atoms of the monomer for all frames of the trajectory fitted by

central domain. The fact that most of the central domain shows little flexibility allows us to use this part of the protein to extrapolate filament geometry from the trimer structure. The co-orientation (i.e. translation vector and rotation matrix) of the rigid parts of two central domains of the second and third monomers of the trimer was used for construction of the long filament. This extrapolation allowed for an estimate of the main parameters of the large-scale geometry of the filament: helical pitch and gyration radius. These two parameters are observable experimentally and could by measured by SANS. There appears to be a substantial variability in these parameters, particularly, even though in the absence of cofactors the protein is expected in the inactive form (helical pitch ~6.3 nm), a secondary maximum near 8 nm is observed, which is close to the active form geometry (figure 5).

To investigate the relationship between domain motion and large-scale filament conformation, we have performed analysis of the relative domain motion. We performed PCA of atomic coordinates of C_α atoms of each domain of the frames which were aligned by the positions of C_α atoms of the central domain. This procedure characterizes motion of each domain relative to the central domain by several principal components. We observe a strong correlation of the first principle motion of N-terminal α-helix with both the calculated cross-sectional gyration radius and the pitch of the protein filament extrapolated from the trimer (figure 6 a,b). We indeed would expect an effect of N-terminal domain motion on the filament geometry as it is a part of the intermonomer interface. At the same time, the first principle motion of C-terminal domain is also correlated with the filament pitch (figure 6 c). Projection of the C-terminus motion on the first principle axis shows two distinct clusters of values, which give rise to the two peaks in the helical pitch histogram. This may suggest that the existence of two distinct states of the filament, active and inactive, is determined by the allowed conformation of C-terminal domain, rather than by the character of motion of N-terminal α-helix.

To isolate quasi-stable states of the protein in the MD trajectory we performed PCA analysis in the space of polypeptide chain torsion angles. The vectors used for analysis were comprised of sine and cosine values of the main chain dihedral angles. For all three monomers of the trimer the projection of the trajectory on the first two principle axes had similar shape (figure 7 a,b,c). The corresponding population maps plotted in terms of free energy (figure 7 d,e,f) show three shallow energy minima, which suggests tree preferred states of the protein trimer. We went on to simulate neutron scattering experiment by calculating small-angle scattering spectra of the protein 50-mer, reconstructed from the trimer on the assumption of uniform geometry.

Spectra were calculated for each frame of the trajectory (10 ps step), and subsequently averaged over 1 ns time around each energy minimum shown in figure 7. Any single frame of the 10 ns trajectory provided a poorer fit as compared to the uniform helix (best-fit frame shown in figure 8, dashed line, χ^2~53). Thus we could not find any "frozen" conformation of this RecA protein that would explain the data better than the uniform helix model. On the other hand, comparison of the curves simulated based on the energy profile (figure 8, solid lines) with the experimental spectrum, obtained for RecA protein from *D. radiodurans* (figure 8, squares) also shows poor fit of the three-state model to the data.

Such a discrepancy likely points to a higher flexibility of the modelled trimer as compared to a long filament and could also indicate essential inhomogeneity of the real filaments. The latter is supported by the elevated scattering intensity near Q=0.07 $Å^{-1}$ that corresponds to the helical pitch values of ~80 Å, which, according of the MD simulation, is characteristic for one of the two clusters of conformational states of the trimer (figure 6c).

4. Conclusions

Molecular dynamics model of the trimer of RecA from *D. radiodurans* reveals pronounced motion of the C-terminal domain that is allosterically coupled with the motion of N-terminal domain and conformational changes in the β-sheet structure of the central domain.

The positional relationship of the rigid domains of the monomers in the trimer allows to use their co-orientation to extrapolate the large-scale geometry of the protein filament from the trimer

structure. Analysis of the relative trajectories of individual domains suggests that motion of N-terminal α-helix is strongly correlated with the filament pitch and cross-sectional gyration radius. Motion of C-terminal domain is associated with the change in the filament pitch characteristic for the transition between active and inactive protein forms, which may point to the guiding role of C-terminal domain dynamics over large-scale filament conformation.

The extrapolated filament geometry shows high variability along the MD trajectory that can not be directly reconciled with SANS data, that implies higher flexibility of the modeled trimer as compared to a long filament and could also indicate essential inhomogeneity of the real filaments. This calls for simulation of larger RecA oligomers to inquire into long-range correlations of the inter-monomer co-orientation.

5. Acknowledgements

This work was supported by the Russian Academy of Science program 'Physics of Elementary Particles' (subprogram "Neutron physics", direction "Studies of structure, dynamics and non-ordinary properties of matter by neutron scattering"), Russian Foundation for Basic Research (grant № 07-04-01548-a and 11-04-01229-a) and by grants of the Ministry of Education and Science of the Russian Federation (RNP 2.2.1.1.4663 and state contract 02.740.11.5223).

References

[1] Datta S, Prabu MM, Vaze MB, Ganesh N, Chandra NR, Muniyappa K & Vijayan M 2000 *Nucleic Acids Res.* **28** pp. 4964-4973
[2] Rajan R & Bell CE 2004 *J. Mol. Biol.* **344** pp. 951-963
[3] Story RM, Weber IT & Steitz TA 1992 *Nature* **355** pp. 318-325
[4] Xing X & Bell CE 2004 *J. Mol. Biol.* **342** pp. 1471-1485
[5] Chen Z, Yang H & Pavletich NP 2008 *Nature* **453** pp. 489-484
[6] Bell CE 2005 *Mol. Microbiol.* **58** pp. 358-366
[7] Nishinaka T, Ito Y, Yokoyama S & Shibata T 1997 *Proc. Natl. Acad. Sci. U.S.A.* **94** pp. 6623-6628
[8] Petukhov M, Lebedev D, Shalguev V, Islamov A, Kuklin A, Lanzov V & Isaev-Ivanov V 2006 *Proteins* **65** pp. 296-304
[9] Yu X, Jacobs SA, West SC, Ogawa T & Egelman EH 2001 *Proc. Natl. Acad. Sci. U.S.A.* **98** pp. 8419-8424
[10] Battista JR 1997 *Annu. Rev. Microbiol.* **51** pp. 203-224
[11] Karelov DV, Lebedev DV, Suslov AV, Shalguev VI, Kuklin AI, Islamov AK, Lauter H, Lanzov VA & Isaev-Ivanov VV 2008 *Journal of Physics: Condensed Matter* **20** p. 104215
[12] Abagyan R & Totrov M 1994 *J. Mol. Biol.* **235** pp. 983-1002
[13] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE & Berendsen HJC 2005 *J Comput Chem* **26** pp. 1701-1718
[14] Oostenbrink C, Villa A, Mark AE & van Gunsteren WF 2004 *J Comput Chem* **25** pp. 1656-1676
[15] Bussi G, Donadio D & Parrinello M 2007 *J Chem Phys* **126** p. 014101
[16] Parrinello M & Rahman A 1981 *Journal of Applied Physics* **52** pp. 7182-7190
[17] Nose S 1984 *The Journal of Chemical Physics* **81** pp. 511-519
[18] Kabsch W 1976 *Acta Crystallographica Section A* **32** pp. 922-923
[19] DiCapua E, Schnarr M, Ruigrok RW, Lindner P, Timmins PA. 1990 *J Mol Biol* 214:557-570.
[20] Lebedev DV, Baitin DM, Islamov A, Kuklin AI, Shalguev V, Lanzov VA, Isaev-Ivanov VV.

2003 *FEBS Lett* 537:182-186.

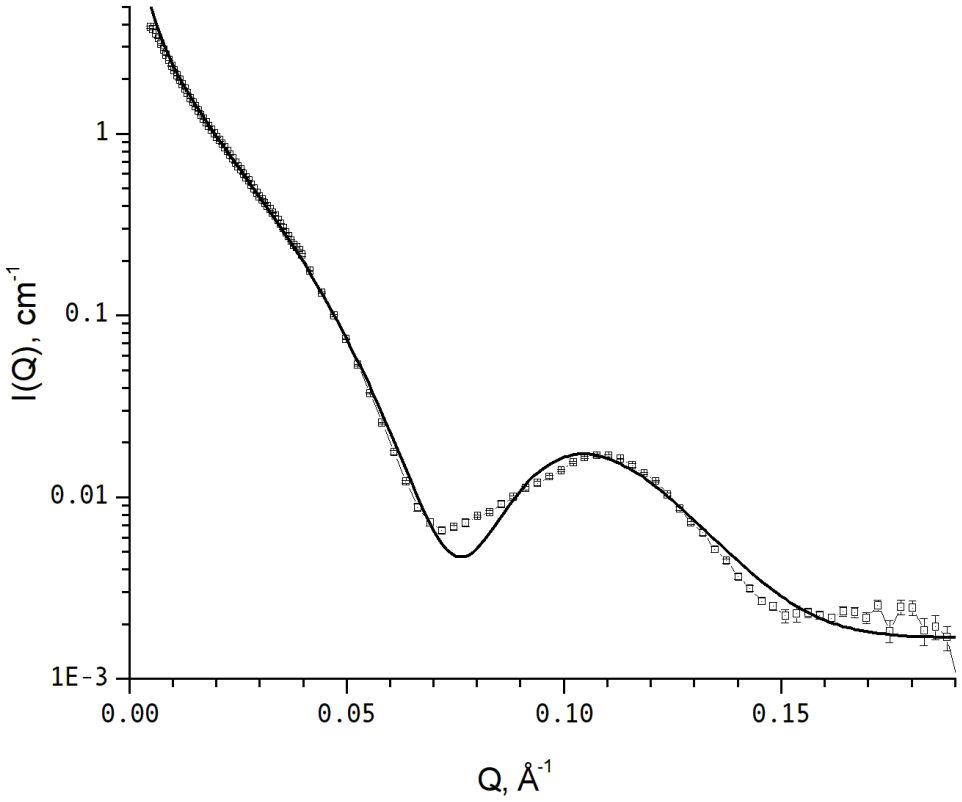
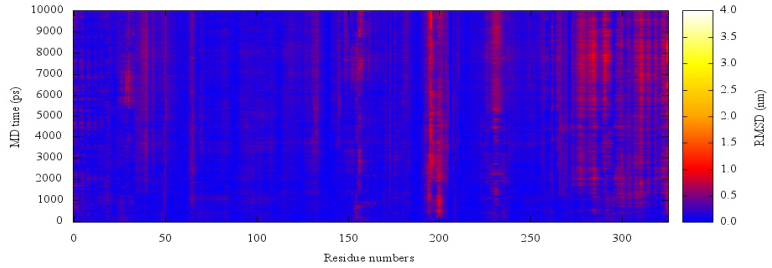
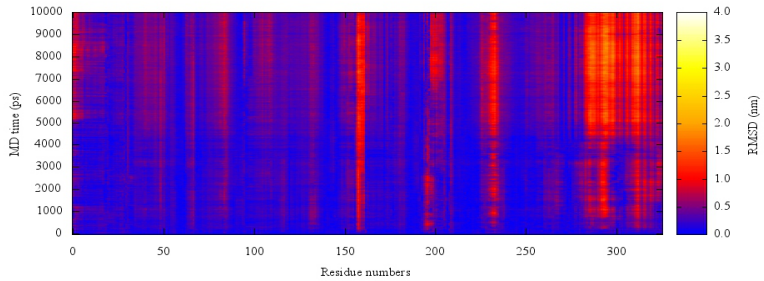


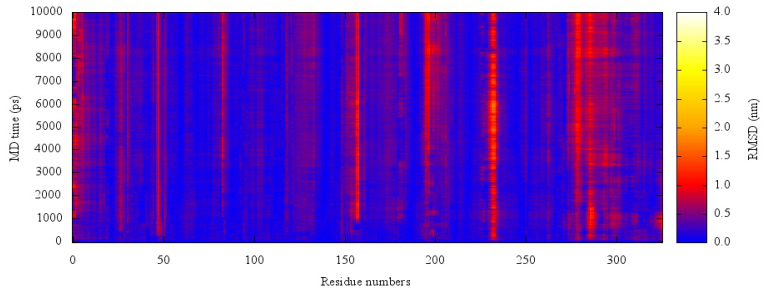
Figure 1: SANS spectrum of RecA self-polymer. Solid line shows the fit to scattering by a long uniform helix with the pitch of 67 Å and the cross-sectional gyration radius of 39.5 Å, χ^2=30.6.



a)



b)



c)

Figure 2. Mobility maps from MD simulations for a) *E. coli* protein; b) *D. radiodurans* protein and c) *D. radiodurans* protein complex with ATP.

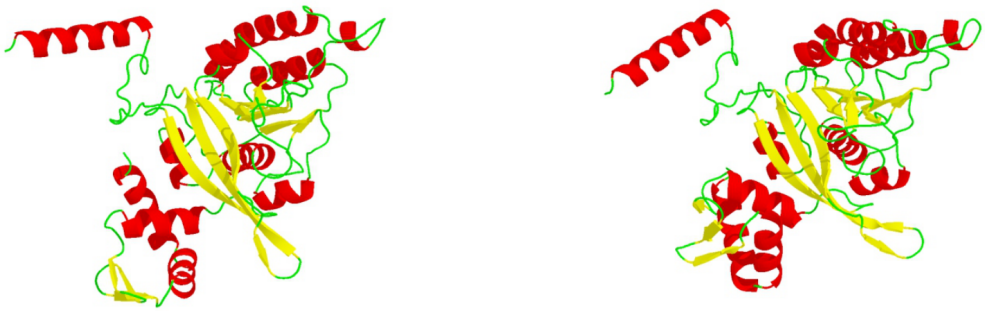


Figure 3. Extremal projections of trajectory of the middle monomer of a trimer of RecA from *D.radiodurans* on the first principal axis.

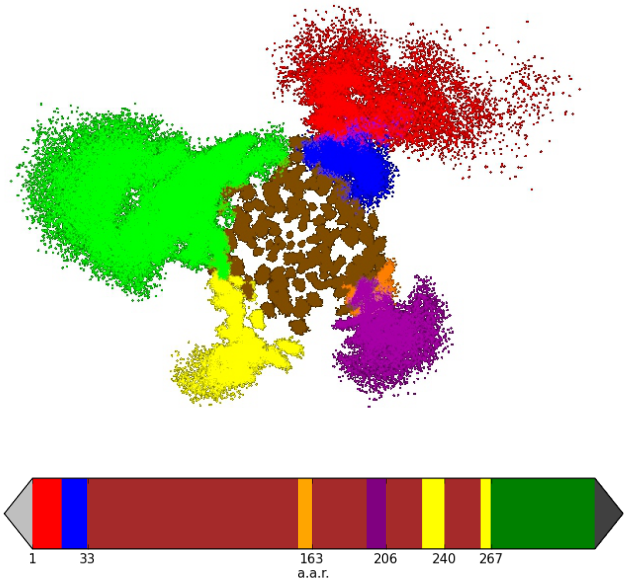
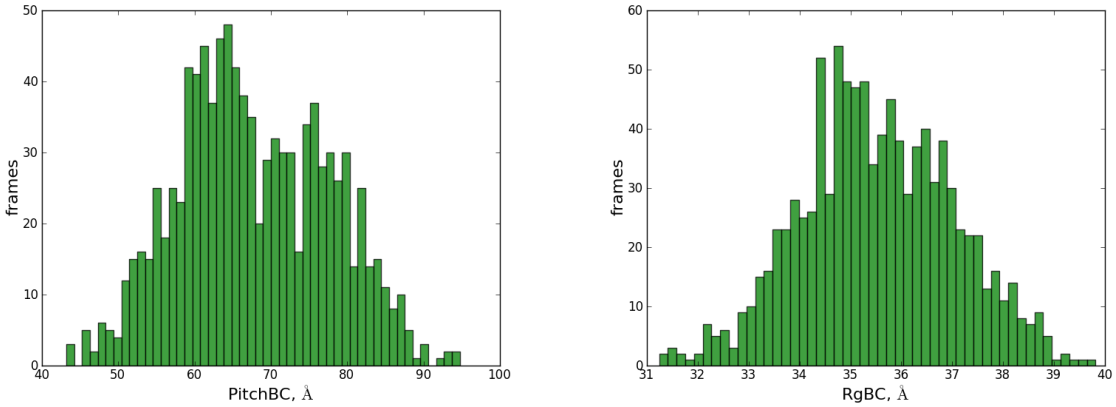


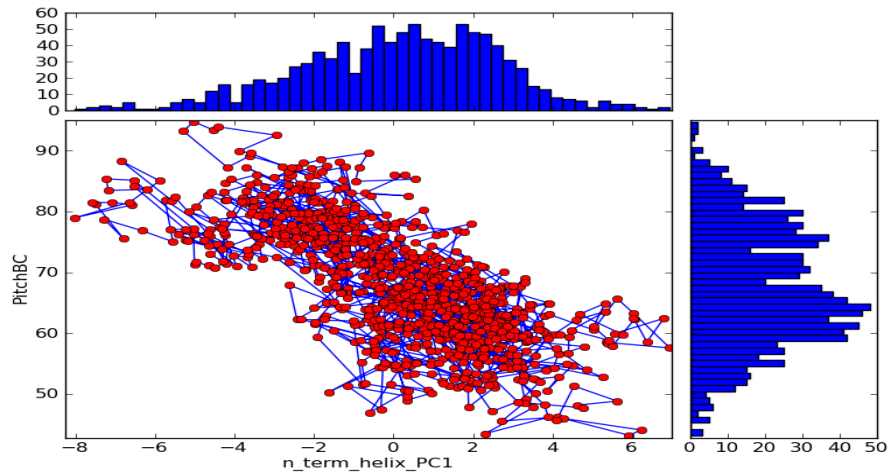
Figure 4. Relative domain motion in RecA protein from *D. radiodurans*. The bar below represents what parts of amino-acid sequence correspond to the respective domains.

Color	residue numbers
Red	1-17
Blue	18-32
Yellow	227-239, 261-266
Orange	155-162
Purple	195-205
Green	267-327
Brown	remaining residues

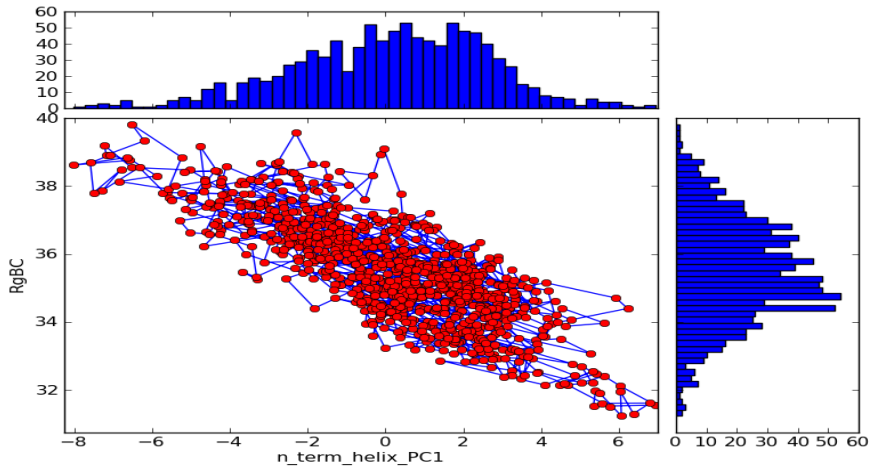


a. b.

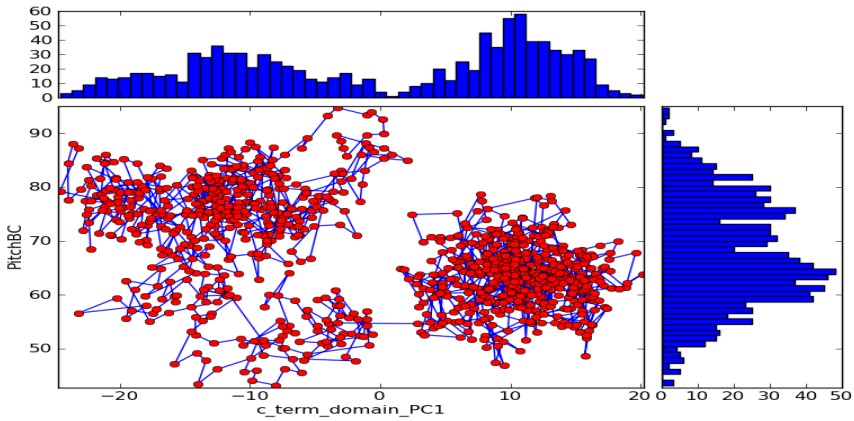
Figure 5. Distribution of the geometric parameters of the filament, - helical pitch (a) and cross-sectional gyration radius (b) based on the MD trajectory.



a.



b.



c.

Figure 6. Relationship between filament geometry and domain orientation:
a) N-terminal helix vs pitch (Å);
b) N-terminal helix vs gyration radius (Å) ; c) C-terminal domain vs pitch (Å)

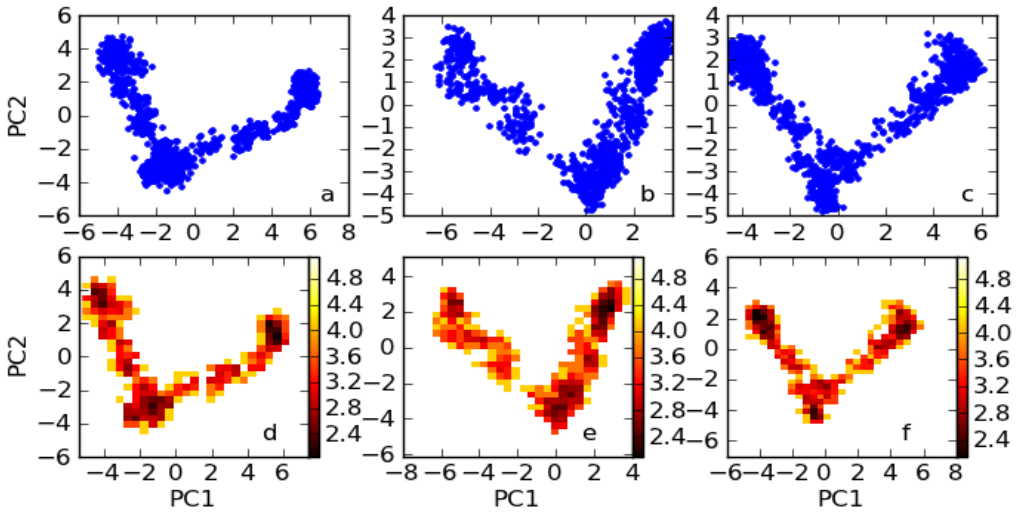


Figure 7. Projections of the monomers trajectory of *D. radiodurans* RecA monomers on the plane of the first two principal axes in dihidral angle space (a,b,c), and the corresponding estimated free energy maps (d,e,f). Color scale in kcal/mol.

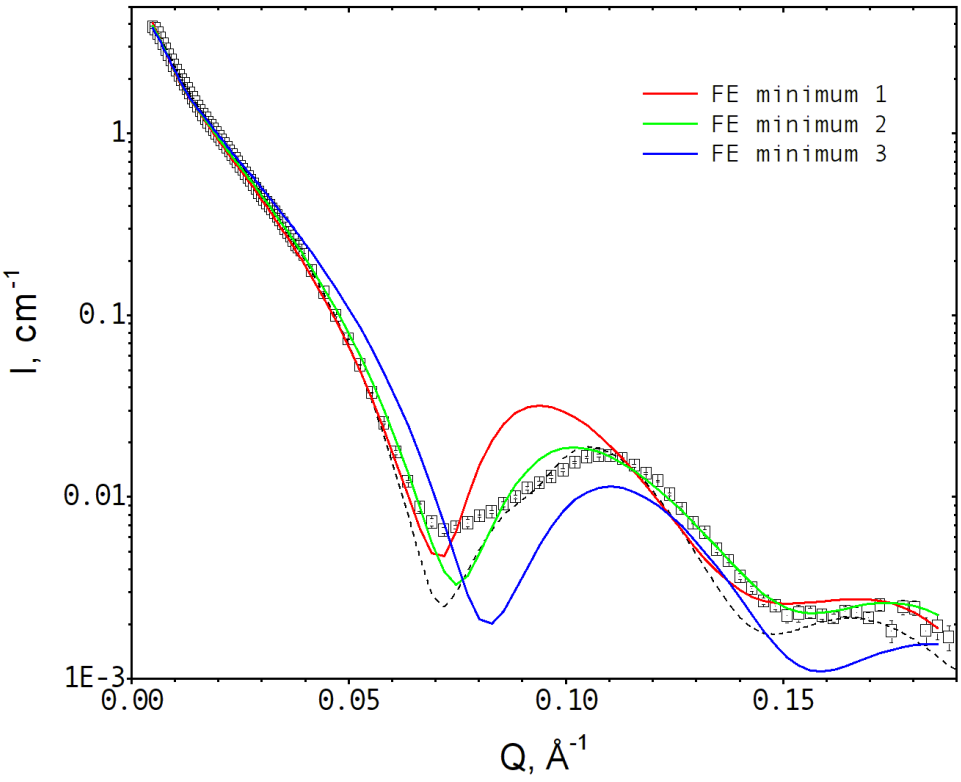


Figure 8. Simulated SANS spectra of RecA from *D. radiodurans*. Lines show the average of 100 spectra calculated for the structures yielded by MD around each of the three local minima of the free energy shown in figure 7. Open squares represent the experimental curve shown in figure 1. Dashed line is the best fit of the scattering spectrum to a single frame of the MD trajectory.