J Mol Model (2011) 17:133–150
DOI 10.1007/s00894-010-0694-8

ORIGINAL PAPER

Binding Selectivity of RecA to a single stranded DNA, a computational approach

Claudio Carra · Francis A. Cucinotta

Received: 2 December 2009 / Accepted: 23 February 2010 / Published online: 13 April 2010

Abstract Homologous recombination (HR) is the major DNA double strand break repair pathway which maintains the genomic integrity. It is fundamental for the survivability and functionality of all organisms. One of the initial steps in HR is the formation of the nucleoprotein filament composed by a single stranded DNA chain surrounded by the recombinases protein. The filament orchestrates the search for an undamaged homologue, as a template for the repair process. Our theoretical study was aimed at elucidating the selectivity of the interaction between a monomer of the recombinases enzyme in the *Escherichia coli*, EcRecA, the bacterial homologue of human Rad51, with a series of oligonucleotides of nine bases length. The complex, equilibrated for 20 ns with Langevian dynamics, was inserted in a periodic box with a 8 Å buffer of water molecules explicitly described by the TIP3P model. The absolute binding free energies are calculated in an implicit solvent using the Poisson-Boltzmann (PB) and the generalized Born (GB) solvent accessible surface area, using the MM-PB(GB)SA model. The solute entropic contribution is also calculated by normal mode analysis. The results underline how a significant contribution of the binding free energy is due to the interaction with the Arg196, a critical amino acid for the activity of the enzyme. The study revealed how the binding affinity of EcRecA is significantly higher toward dT_9 rather than dA_9, as expected from the experimental results.

Keywords Binding · EcRecA · Homologous recombination · MM-PBSA · Molecular dynamics

C. Carra (✉)
Universities Space Research Association,
2101 NASA Parkway,
Houston, TX 77058, USA
e-mail: claudio.carra-1@nasa.gov

F. A. Cucinotta
NASA-JSC Space Radiation Health Project,
2101 NASA Parkway,
Houston, TX 77058, USA
e-mail: francis.a.cucinotta@nasa.gov

Introduction

Improving the understanding of DNA repair mechanisms and pathways has not only an academic interest but it can have a strong influence on our understanding of disease, by helping to clarify how pathologies can be treated or diagnosed more effectively. Deficiencies in DNA repair [1] often contribute to developmental abnormalities, increased rates of cancer, and tissue and organ degeneration. DNA lesions are induced from endogenous sources such as free radicals generated by metabolic processes to environmental agents like ultraviolet lights (UV), chemical agents, or ionizing radiations. Of the various classes of strand-breaks and base damages, the double strand break (DSB) is the most genotoxic [2]. The DSB phenomena is also relevant in human space exploration and radiation protection on Earth [3]. Our interest is mostly oriented toward the homologous recombination (HR), which is the primary DSB repair pathway that keeps the genomic integrity by using an undamaged homologue, a sister-chromatid, as a template for the repair process [4]. The dominant step of this process involves the DNA recombinase proteins, RecA [5] in bacteria, RadA in archaea [6], and Rad51 in eukaryotes [7], which promote an ATP-mediated DNA strand-exchange reaction.

A molecular framework to model HR is *via* a three-strand exchange reaction between linear double stranded DNA, dsDNA, and circular single-stranded DNA, ssDNA,

promoted by the recombinases enzymes. Biochemical studies have revealed that the reaction proceeds in two distinguishable steps. The first phase consists in the *presynaptic* polymerization of RecA protein on ssDNA forming a helical nucleoprotein filament assembled in the 5′-3′ direction [8]. The second is the *synaptic* homologous alignment of nucleoprotein filament formed by RecA-ssDNA with a naked dsDNA filament. In the final step, a unidirectional strand exchange, promoted by ATP hydrolysis, brings to a stretching of heteroduplex DNA and to a dissociation of the RecA.

Each RecA monomer is formed by three domains, whose X-Ray structure was determined for the first time by Story et al. [9]. The residues 1–33 (the N domain) form an α-helix and β-strand which have the function to pack against the neighboring subunit, including the majority of the monomer interface. The central part of the enzyme, core domain, residues 34–268, is formed by eight stranded mixed α-helices and β-sheet and that is the area where the ATP binding site is located. The position of the loops binding the ssDNA in the *presynaptic* filament, L1 and L2 has recently been elucidated after the X-ray structure was solved by Chen et al. [10]. The final part of the protein is represented by the C-terminal domain, residues 269–352, which is rather far from the core of the enzyme and it is responsible for the formation of the *synaptic* filament [11, 12]. An interesting aspect of RecA is the high degree of conservation that the L1 and L2 have in prokaryotic cells. In a sequence of 64 bacteria the RecAs have 11 of the residues from 193 to 212 almost identical, and 6 are highly conserved, thus 17 over 20 residues are either identical or chemical conserved [13, 14]. Moreover, the eukaryotic homologues Rad51 and Dmc1 have an analogue function to RecA, and reveal a homology sequence, especially in within the 230 aminoacids in the core of the structure [15]. Despite the similarities, some of the residues in L2 of RecA like Arg196 and Lys198 are not present in the corresponding region of the Rad51 and Dmc1 [16]. It has been proposed that part of the functions provided by the missing aminoacids can be restored by other proteins not yet discovered [17]. Since the degree of conservation in the family of the recombinase enzymes is very pronounced, we centered our study on the *Escherichia coli* RecA for which the crystal structure of the binding complex has been determined.

We aim to model by classical dynamics, the interaction of the monomer of the EcRecA with a sequence of oligonucleotydes of 9 bases length, dT_9, dA_9, dC_9, and dG_9, to evaluate the charge in binding free energy and the modification of the complex hydrogen bonding network present in the complex. To shed light on the binding mechanism we looked at several aspects of the enzyme interacting with the ssDNA fragment. The mobility of the complexes is analyzed by comparing the averaged RMSD values as a function of residue number. The study is extended to the analysis of the correlated and uncorrelated motions revealed by the cross correlation matrix plotted for the systems. We restricted the configuration space by performing a principal component analysis (PCA) on the trajectory of molecular dynamics. This approach is aimed to complement the previous results by providing a more accurate description of the distinct states for the complexes, compatibly with the inevitable error that a short sampling will imply [18]. Our interest, however, is not only in the conformational analysis of the complexes, but also in the estimation of the binding free energy.

For this purpose, we adopted the MM-PB(GB)SA protocol, using the Poisson-Boltzmann (PB) and generalized Born (GB) solvent accessible surface area methods. In order to calculate the absolute binding free energy, we calculated the solute entropy by normal mode analysis. The results revealed how the main contribution of the binding free energy is due to the interaction with the Arg196, critical amino acid to maintain the activity of the enzyme [17]. The model showed how the free binding energy of EcRecA is higher for dT_9 than for dA_9, and dC_9 in agreement with experimental evidence.

Materials and methods

The structure of the of EcRecA monomer binding dT_9, **T**, was built by taking part of the crystal structure determined by Chen et al. [10] (pdb id 3mcu). The terminal C was removed since its role is to recognize the homologue dsDNA once the *presinaptic* filament is formed [11, 12], thus it is assumed to be irrelevant for the binding process. The terminal N was also disregarded because its role is to allow the protein fusion by polymerization [10]. The ATP residue was inserted in the same geometry of the $ADP{\cdot}AlF_4$ present in the crystal structure. The library from Carlson et al. [19] was used to model the nucleotide. The geometry of **T** was used as a template to generate the initial structure of the RecA interacting with dA_9, dC_9, and dG_9 (**A**, **C** and **G** respectively), in which the fragment dT_9 was replaced with the new sequence of nucleotides. Several Na^+ ions were added to reach the neutrality of the complex. The molecular dynamics calculations and the data analysis were carried out with the AMBER 10 [20] package using the parmbsc0 force field [21]. The solute was modeled in a periodic box with a 8 Å buffer of water molecules explicitly described by the TIP3P [22] model. The particle mesh Ewald (PME) method [23] was used to treat the long range electrostatic interactions.

The equilibration of the system was performed according to the following protocol. Initially the geometry of the

system went through an optimization process with 4000 cycles, the first 1000 by steepest descent and the rest with a conjugate gradient method, with the complex constrained, to relax the solvent. Then a further optimization of 9000 cycles with no constraints on the whole system was carried out to obtain the final relaxed geometry. The first equilibration was executed with a weak restraint on the complex for 50 ps at constant volume, constantly increasing the temperature from 0 to 300 K. The equilibration continued to for 100 ps at a constant pressure of 1 atm, by keeping the temperature constant with the Langevin temperature equilibration scheme [24] using a collision frequency of 1.0 ps^{-1}. Under these conditions the restrains were gradually removed. Finally, the production run was performed without restrains for 20 ns. During the simulations, hydrogen stretching motions were removed using SHAKE bond constraints [25], allowing a longer sampling time step of 2 fs. The Molecular Mechanics Poisson Boltzmann and Generalized Born Surface Area method, MM-BP(GB)SA, was used to calculate the free binging energies of the complex. According to the MM-BP(GB)SA protocol, the binding free energy is estimated by taking the average energy difference between the complex, $\overline{G}_{comp}$, and the reactants, $\overline{G}_{rec} + \overline{G}_{lig}$.

$$\Delta G_{bind} = \overline{G}_{comp} - \left(\overline{G}_{rec} + \overline{G}_{lig}\right) \quad (1)$$

Where the average free energy, G, for the complex, RecA-ATP-$Mg^{2+}$$dN_9$, receptor, RecA-ATP-$Mg^{2+}$, and ligand, dN_9, is composed by:

$$\overline{G} = E_{MM} + G_{PB/GB} + G_{SA} - T\Delta S^{(s)} \quad (2)$$

where E_{MM} is the molecular mechanics interaction energy, in "gas phase", within the system, $G_{PB/GB}$ is the electrostatic energy component calculated with the Poisson-Boltzmann (PB) [26], or generalized Born (GB) [27] methods. One of the advantage of using the MM-PB(GB)SA method is that the "nonphysical" annihilation [28, 29] or decupling [30, 31] of the species alone in solution or bounded to a substrate is not required anymore. Moreover, contrarily to the umbrella sampling method, it is not necessarily to model the partially unbound states. The MM-PB(GB)SA method was extensively mentioned in the literature as a model capable to well reproduce qualitatively the binding free energy of such systems [32–34]. We used different GB methods to compare the results of different protocols to our systems. We used the protocol of Howkins et al. [35], IGB1, the modified GB methods, GB^{OBC}, as in the approach of Onufriev et al. [36], IGB2, a modification of IGB2 [37], IGB5, and the modification of Mongan et al. [38], IGB7. For all the GB^{OBC} approaches, we used the bondi and the mbondi2 radii definitions, IGB2′, IGB5′ and IGB7′ respectively, whereas in IGB2″ we used PBradii, Table 2. The ΔG_{SA} represents the non-polar contribution to the solvation free energy which is determined with solvent-accessible-surface-area-dependent terms (SA) [39] approach. The term $T\Delta S^{(s)}$ is the conformational entropy change of the solute which was calculated but normal mode analysis. The grid size used to solve the Poisson-Boltzmann equation was 0.5 Å, and the values of interior dielectric constant and exterior dielectric constant were set to 1 and 80, respectively. The gas phase and the solvation free energies were calculated over 400 snapshots taken at 20 ps interval from the last 8 ns of the MD trajectories. To model the experimental conditions, the bulk of the salt concentration was set 50 mM.

We calculated the correlation matrix for the complex, displayed as a two dimension correlation map to investigate motion between regions in the protein, as domain-domain communication [40–42], or between ligand and receptor. A positive value shows that the atoms are moving in the same direction, whereas a negative value indicates an anticorrelated motion. We used ptraj applet in Amber code to generate the correlation matrix and Matlab to generate the 3D plots.

The coordinate trajectory matrix, in a MD calculation, has the dimensions of (3N, n) where N is the number of atoms and n the number of snapshots used in the analysis. The corresponding conformational space has a high dimensionality and a simple description or visualization is impracticable. We used the principal component analysis (PCA), methods to reduce the number of dimensions of the conformation space, to allow the depiction of extreme structures and the major fluctuations of motions which results correlated. This protocol projects the multidimensional conformational space onto a new set of axis which maximizes the variance of the projection along orthogonal directions. Following this procedure, the resulting projection has a low-dimensional representation of the spatial relationship between conformations. The analysis starts by diagonalizing the covariance matrix $\sigma_{m,w}$ whose individual element is:

$$\left(\sigma_{m,w}\right)_{i,j} = \left\langle \left(y_i - \langle y_i \rangle\right)\left(y_j - \langle y_j \rangle\right)\right\rangle. \quad (3)$$

The MD trajectory is projected onto the main essential directions, corresponding to the larger eigenvectors. At this point the visualization of the extreme structures and the major fluctuations of the correlated motions is possible. Because of the memory limitations we considered only the Cα and P atoms in the projection of the MD trajectories, onto the first two principal eigenvectors. The eigenvectors and the energy landscapes are visualized with Matlab. The PCA [43–48] analysis was carried out with PCAZIP software [49], and the collective dynamic modes are plotted using the porcupine method developed by Tai at al. [50].

Results

The model was constructed on the crystal structure of the active complex RecA-ATP-p(dT) determined by Chen et al. [10], (PDB id 3cmu). The geometry revealed that each triplet set of dT base interacts with three adjacent RecA monomers, however the binding process propagates *via* a monomer initiation [51–53] and continues *via* a co-operative binding mechanism [51, 53, 54]. As a consequence, it is appropriate to design a model formed by a single monomer to study the geometry of the complex at the incipient formation of the nucleoprotein filament. In our simulation we considered the monomers of the experimental crystal structure of EcRecA interacting with a sequence of nine bases, dT_9, **T**. The 9-mer was then replaced with dA_9, dC_9, and dG_9, kept in the same orientation of dT_9, obtaining the complexes **A**, **C**, and **G** respectively.

For all trajectories, the root mean square deviation, rmsd, of the backbone atoms in function of the residue number is shown in Fig. 1. The loops, L1, and L2, (residues 157–164 and 194–209) and the 9-mers, dN_9 are the residues directly involved in the binding process. For the four complexes the core part of the enzyme shows a rather high rigidity with a value of rmsd below the unity, underlying how the geometry change of the binding moieties does not deviate sensibly from the original X-Ray structure. However a noticeable change in the rmsd in the binding loops regions is observed, with values from 4 to 6 Å and ~3 Å for L1 and L2 respectively. The complex **T** shows en extended stability in the L1 and L2 regions, and in the core part of the enzyme, with relatively low RMSD values. **A**, on the other side, reveals the highest values in rmsd: ~6 Å for L1 and ~3 Å for L2 respectively. **C**, unexpectedly, appears to have very little movement, with RMS values very low in L1 and in most of the core part of the enzyme. On the other end, **G** present a rather wide range of movements, with a rmsd value of ~5 Å in the L1 region and ~2.5 Å in the L2 binding site. The fluctuation of the core part of the enzyme is in range with the other complexes.

In order to investigate motion between different region in the protein, as domain-domain communication [40–42], or between ligand and receptor, we calculated the correlation matrix for the binding complexes, displayed as a two dimension correlation map. A positive value (in red) shows atoms moving in the same direction, whereas a negative value (in blue) indicates an anticorrelated motion.

Our data display the correlation values of the Cα atoms for the protein and P atoms for the nucleotide, Fig. 2. The complex **T** shows areas of mild correlation and anticorrelation, with values changing from -0.3 to 0.3. The movement of the residues belonging to L1 is anticorrelated, underlying the low bonding character of the fragment. The interaction of the enzyme with dT_9 can be seen by looking at the top and right border of the graph. dT_9 shows a barely noticeable correlation with L2, and a sporadic small correlation motion with respect to the rest of the protein. The smallest values in the correlation matrix found in **T** are due to a minor change in the geometry of the complex during MD simulation, suggesting how the geometry of crystal structure is close to corresponding modeled structure in solution. The change of the ligand leads to an increase in the absolute values of the correlation matrix elements for **A** and **T**, underlining how the resulting geometries change during the simulation. The complex **C** shows a mild region of correlation and anticorrelation, indicating how the initial guess is already close to the final structure. In particular, the correlation map shows a significant increase in the correlation and anti-correlation areas to values closer to ±0.5 for **A** and **G**, including extended areas with values close to the unit. Those regions are almost uncorrelated with the binding loops of the RecA, confirming how the binding geometry does not change significantly from the X-ray structure.

Because of the size of the systems considered and the potential flexibility that the binding units might show, we extended our analysis to a study based on the principal component analysis (PCA), often called essential dynamics [48] applied to protein dynamics [50]. The PCA technique was extensively used [48, 55–58], despite the issues raised

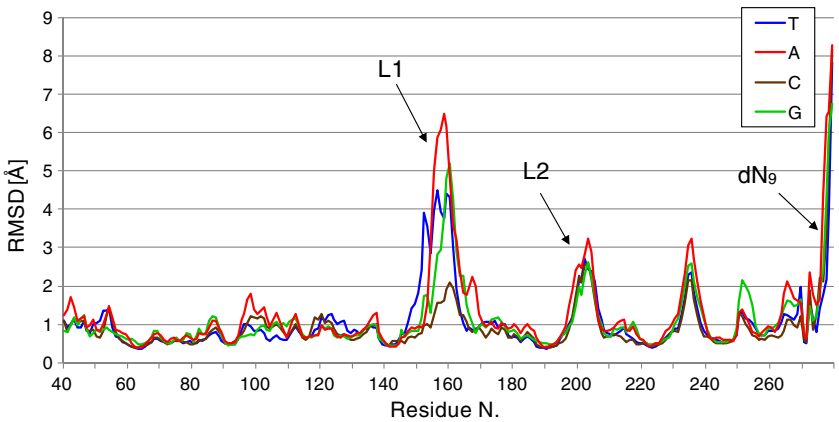

Fig. 1 RMSD value per residue averaged over 20 ns for the complexes considered

on the effective accuracy [59]. With this approach, the high dimensional space spanned by 3N-6 degrees of freedom (where N is the number of atoms) is highly reduced. The eigenvalues resulting from the principal component projection represent the variation of the original set of data along the corresponding eigenvector. These eigenvalues are normally sorted in terms of their magnitude, and their accumulative sum gives an indication of the quality of the representation for a given number of dimensions. As a limiting case, if all the eigenvectors are considered, the original space is correctly represented in the projection subspace. The representations of the two principal component eigenvectors, relative to Cα and P atoms, for RecA-dN_9 are shown in Fig. 3. The eigenvectors ev1 and ev2 relative to RecA are shown in red, while the vectors for the dN_9 are shown in blue. The binding loops are represented in marine blue and green for L1 and L2 respectively. The main component of the motion of **T**, ev1, is dominated by a concerted motion of the monomers toward dT_9. In particular, L2 has a strong component toward the binding regions. L1, on the contrary, seems showing a minor binding character, with vectors pointing far from dT_9.

The components of ev2 are lower in magnitude, but still they underline the motion of the RecA toward dT_9. Changes in the ligand alter the values of ev1 and ev2 but the general pattern results preserved. The common tendency in binding the dN_9 fragment is maintained, noticeable by the motion of the protein toward the bases. As a general picture we can identify two main motions for all the systems: a combined rotation of dN_9 toward the monomer in particular in ev1 for **T**, **A** and **C**, and a general motion of the core of the enzyme toward the dN_9 moiety. In a more

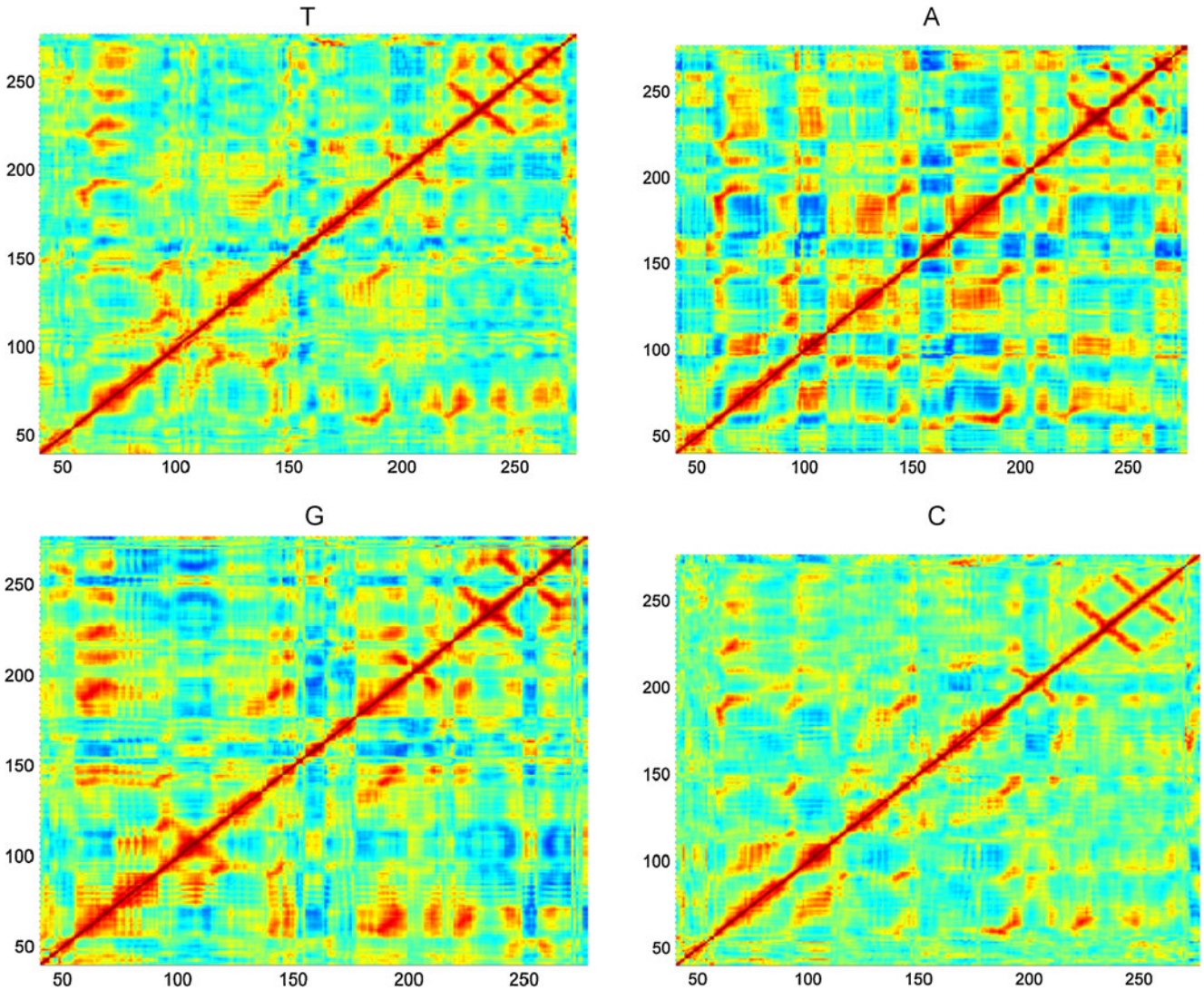

Fig. 2 Results for the cross correlation matrix. Positive values (in red) show atoms moving in the same directions, whereas negative values (in blue) indicate anti correlated motions

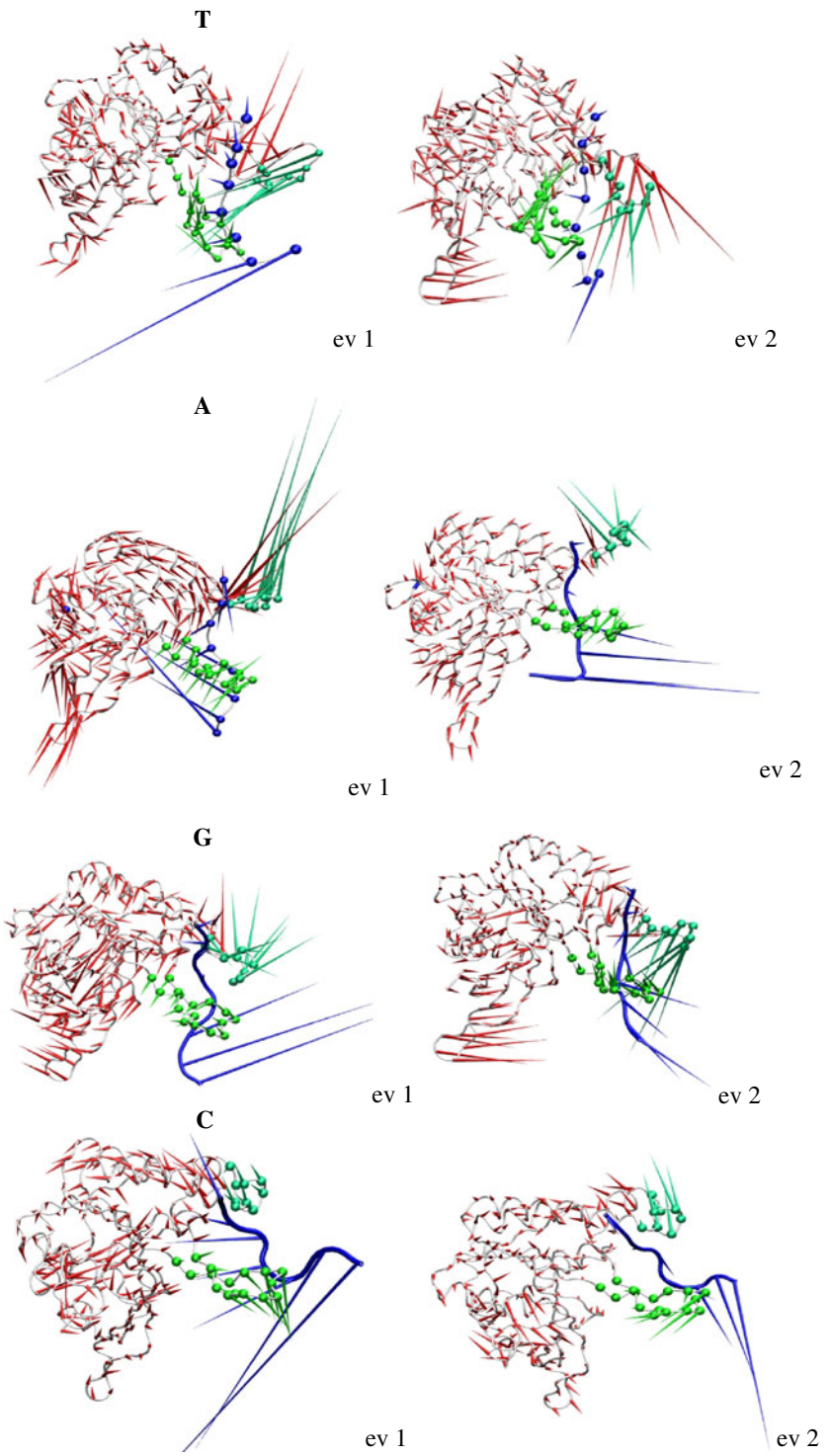

Fig. 3 The first and second component of the principal component analysis plotted for Cα and P atoms. The vectors are generated with the porcupine program

specific case, **C** shows a strong component for ev1 in the L1 and L2 regions toward dC_9. **A** and **G**, however, have the ev1 component on L1 and L2 pointing not directly on dN_9 fragment, underlining a minor binding tendency. The components in ev2 show a binding propensity in **T** and **G**, while in **A** and **C** the components of the eigenvector ev2 describe a dendency to separate the dN_9 from L1 and L2.

According to the energy landscape model [60], the binding coupling is often followed by conformational changes related to the biological functions of proteins and, as such determine the nature of the energy landscape [61–66]. In this model a selection of conformations for the binding process in biomolecules, implies that the flexibility of the binding partners and the binding mechanism can be

established by the dynamic equilibrium of the complex conformational states [67].

The two dimensional projections of the trajectories of the systems considered onto the first and second principal components is shown in Fig. 4. This technique allows to visualize the shape of the potential energy surface in spite the inevitable distortion introduced by the projection. The model was performed using only the Cα atoms of the protein residues and P atom for the dN_9 fragment, thus it describes the interaction with the ssDNA only partially. Nevertheless, it is relevant to observe groups of local minimum along the trajectories which are more or less pronounced in function of the nature of the ligand. It is interesting to remark how the interpolation of the values of the projection of the trajectory onto the two main principal eigenvectors along the simulation, black line, crosses the funnel areas shown by the histogram.

The numbers have not been scaled to reproduce relative energies because our statistics from MD simulation is insufficient to obtain a quantitative accuracy; however the results show how the basins structure is altered by changing the nature of the ligand. For **T** there are several small basins separated by low barriers showing the exploration of the conformational space along the simulation. The first basin, encountered at the initial 4 ns of the simulation, is located in proximity of the point with ev1, ev2 coordinates (-20, 15). The system rapidly evolves though a shallow valley, for about 4ns, to reach the most dominant basin at coordinates (10, 0) for all the remaining time of the simulation. A similar behavior is found for **C**, where from

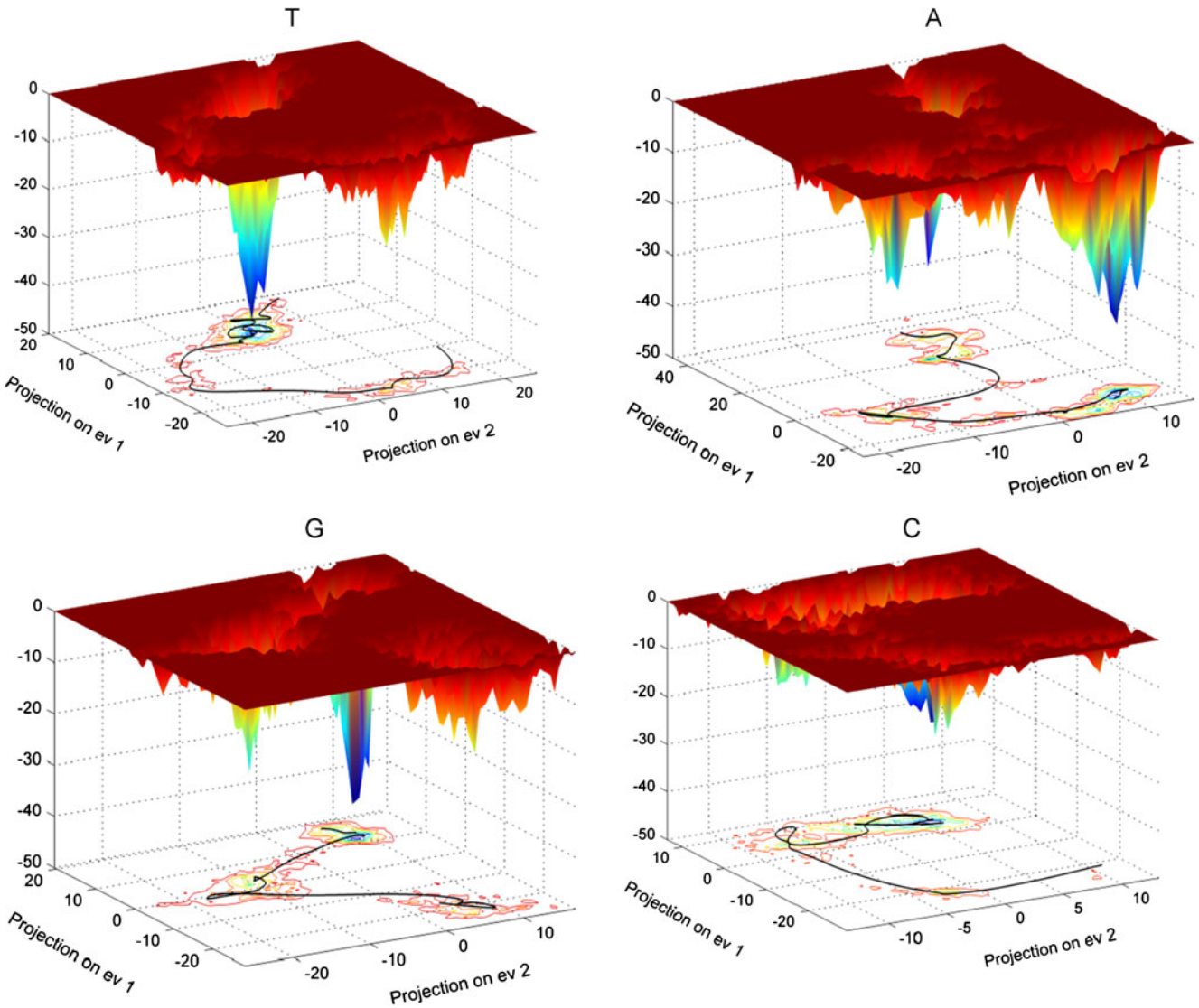

Fig. 4 The "minimum energy envelope" mapping the potential energy surface of the complexes. The black line represents the interpolation of the values of the projection of the trajectory on the first two principal eigenvectors

the initial point (-20, 10) the complex reaches in 4 ns the final state at (10, 5), through a deeper valley. Both final basins are the deepest compared to the other complexes. The binding structures **A** and **G** have common differences respect the previous cases. Both reveal a complex structure of basins, more separated rather then connected by valleys as previously encountered. In addition, the basin reached at the end of the simulation lies to a higher minimum. The shape of the energy envelope graph appear correlated to the binding properties, where the highest is the binding energy the deeper and pronounces is the final minimum samples for the same numbers of bins.

The input for the MM-PB(GB)SA calculations were extracted from the last 8 ns of the MD trajectories, from 400 adjacent snapshots, at 20 ps time interval. Since the correlation time for decay of fluctuation of the free energy is about 1 ps, extracting the snapshots at a time significantly longer should lead to independent series of sample points. Under these conditions, the standard error in the mean value, (sem), decreases with the square-root of the number of points, thus having a number of sample big enough it is possible to estimate the value for the binding free energy. The study conducted by Kollman and Case [68, 69], revealed how a number of snapshots from 100 to 200 is sufficient to calculate the binding energy with a good numerical accuracy. The values for the binding free energy, calculated by MM-PB(GB)SA, oscillate in an expected range respect the average value.

The entropy of the solute, necessarily to obtain the absolute binding free energy, has been calculated by normal mode analysis based on a harmonic approximation.

In our study we used several methods to calculate binding free energy. To the MM-PBSA method, Table 1, we compared several MM-GBSA approaches with different effective Born radii formulations, Table 2. To verify the consistency of the results, we plot the MM-PBSA and MM-GBSA values (for this case we considered the IGB2″), for each snapshot for the last 8 ns of the trajectory, Fig. 5. In the discussion we consider the MM-PBSA energies as a reference to evaluate the relative stability of the complexes.

Table 1 reports the mean decomposition values of the MM-PBSA energies relative to RecA binding the different dN_9, including the contribution of the binding entropy. The two major attractive contributions are the gas phase Coulombic energy, H_{elec}, and van der Waals energy, H_{vdW}, whose sum is labeled as H_{gas}. In general the antibonding components are due to the polar solvation values, G_{BP}, while the non-polar part, G_{np}, gives a minor contribution. The sum of both quantities is labeled as G_{solv}. The polar solvation part of the free energy is calculated by solving the Poisson-Boltzmann (PB) equation using radii from the PARSE parameter set [70], while G_{np} is calculate by the solvent accessible surface area dependent method, SA. The overall MM-PBSA energy, $G_{gas+solv}$, represents the sum of H_{ele} and G_{solv}. The absolute binding free energy, G_{tot}, includes the solute binding entropy.

The contributions to the overall $G_{gas+solv}$ is similar for the complexes having a binding character, **T**, **C** and **G**, contrarily to **A** which shows a different composition of $G_{gas+solv}$. The highest binding contribution of H_{elec} is found in **T** with a value of -209.2 kcal mol^{-1}, and similar values are encountered for **G** and **C**, -147.9 kcal mol^{-1} and -189.8 kcal mol^{-1} respectively, in contrast with **A** which has a very unfavorable Coulombinc energy contribution term, +183.7 kcal mol^{-1}. H_{vdw} energy component has negative values for all complexes, with the highest term for **G** and **T**, -104.1 kcal mol^{-1}

Table 1 MM-PBSA energy components of the complexes calculated for 800 equidistant snapshots extracted from the last 8 ns of the same trajectory

Contrib	T		A		G		C	
	Mean	sem[h]	Mean	sem	Mean	sem	Mean	sem
H_{elec}[a]	-209.2	6.4	183.7	6.7	-147.9	6.4	-189.8	6.7
H_{vdW}[b]	-101.7	1.9	-63.9	1.8	-104.1	1.6	-85.9	1.6
H_{gas}[c]	-310.9	6.8	119.9	7.0	-252.0	6.8	-275.7	7.0
G_{np}[d]	-13.6	0.1	-9.9	0.1	-14.3	0.1	-11.6	0.1
G_{PB}[e]	249.4	5.8	-128.3	6.4	221.1	5.5	225.3	6.1
G_{solv}[f]	235.9	5.7	-138.2	6.4	206.8	5.5	213.7	6.1
$G_{gas+solv}$[g]	-75.1	3.2	-18.3	3.4	-45.2	3.4	-62.0	3.2
$TS_{tot}^{(s)}$[i]	-39.4	2.7	-37.9	2.8	-39.1	2.4	-42.6	2.1
G_{tot}[l]	-35.7	4.2	19.6	4.4	-6.1	4.2	-19.4	3.9

The values are averaged over 400 snapshots, 20 ps time interval, of the last 8 ns trajectories. The standard state is assumed to be at 50 mM. The units are in kcal mol^{-1} . [(a)] H_{elec} coulombic energy; [(b)] H_{vdW} van der Waals energy; [(c)] $H_{gas} = H_{elec} + H_{vdW}$; [(d)] G_{np} non-polar solvation free energy calculated by MM-PBSA; [(e)] G_{PB} polar solvation free energy; [(f)] $G_{solv} = G_{np} + G_{PB}$; [(g)] $G_{gas+solv} = H_{gas} + G_{solv}$; [(h)] Standard error of mean values; [(i)] $TS_{tot}^{(s)}$ total solute entropy contribution; [(l)] $G_{tot} = G_{gas+solv.} - TS_{tot}^{(s)}$.

Table 2 Comparison of MM-GBSA method calculations for the RecA-dN_9, extracted from the last 8 ns of the trajectory

Method[a]	T			A			G			C		
	mean	sem[(b)]	Δ[(c)]	mean	sem	Δ	mean	sem	Δ	mean	sem	Δ
GB1	-103.1	2.9	-28.1	-44.0	3.2	-25.6	-87.2	3.2	-42.0	-77.4	3.1	-15.4
GB2	-76.3	2.9	-1.2	-18.8	3.2	-0.4	-52.3	3.1	-7.1	-54.6	3.1	7.4
GB5	-205.0	3.8	-130.0	-71.7	3.6	-53.4	-151.1	4.0	-105.9	-148.2	3.7	-86.2
GB7	-200.2	7.2	-125.1	-198.0	6.7	-179.6	-216.2	6.2	-171.0	-209.5	6.7	-147.5
GB2′	-84.1	2.9	-9.0	-22.9	1.6	-4.5	-57.8	3.2	-12.6	-60.8	3.1	1.2
GB5′	-229.6	4.2	-154.5	-90.6	6.8	-72.3	-177.8	4.4	-132.5	-169.8	4.1	-107.8
GB7′	-241.7	7.0	-166.6	-216.2	0.1	-197.8	-251.1	6.0	-205.9	-244.8	6.5	-182.8
GB2″	-81.8	2.9	-6.7	-21.4	3.2	-3.0	-56.7	3.2	-11.5	-58.6	3.1	3.4

The values are averaged over 400 snapshots, 20 ps time interval, of the last 8 ns trajectories. The standard state is assumed to be at 50 mM. The units are in kcal mol^{-1}. [(a)] Several MM-GBSA methods, see text for details; [(b)] Standard error of mean values, [(c)] Difference respect the corresponding MM-PBSA value reported in Table 1.

and -101.7 kcal mol^{-1} respectively, and significantly smaller for **A**, -63.9 kcal mol^{-1}. The values of G_{np} for all systems is between -9.9 kcal mol^{-1} and -14.3 kcal mol^{-1}. The polar contribution, however, helps the binding only in **A**, -128.3 kcal mol^{-1}, while for the other cases the terms are largely positive, 235.9 kcal mol^{-1}, 213.7 kcal mol^{-1}, and 206.8 kcal mol^{-1} for **T**, **C** and **G** respectively. The strongest binding character is encountered for **T**, G_{tot} is -35.7 kcal mol^{-1}, despite the unfavorable entropic contribution of -39.4 kcal mol^{-1}. The data come from a well equilibrated structure with a linear regression slope of 10^{-3} kcal·mol^{-1}·ps^{-1}, Fig. 5, comparable to previously published results [33].

The binding of RecA with dA_9 reveals a completely different feature. Even if the $G_{gas+sol}$ is negative, -18.3 kcal mol^{-1}, the binding entropy, -37.9 kcal mol^{-1}, make the process unfavorable. The convergence for **A** is achieved with a higher slope value, $-1.44 \cdot 10^{-2}$ kcal·mol^{-1}·ps^{-1}. This result is in full accord with a series of experiments, for instance Wittung et al. showed that the binding enthalpy of RecA to ssDNA is a function of the base sequence, with a net preference of the p(dT) rather than p(dA)[71]. The absence of cofactor was also considered and the binding affinity are still in the same order, with a strong preference of p(dT) [72]. The data presented by Bugreeva et al. [73, 74] confirm that the RecA affinity to dT_{20} is about 240 time higher than to dA_{20}. Experimental results on the affinity of RecA to p(dC) are much more elusive, and the data reported in the literature are controversial but still confirming an intermediate affinity of RecA for p(dC) respect p(dT) and p(dA) [71, 75, 76]. Consistently, our results on the complex **C** reveal an intermediate binding character, with a G_{tot} of -19.4 kcal mol^{-1}, despite the strong entropic factor, -42.6 kcal mol^{-1}. Experimental studies on the interaction of RecA with p(dG) are rare and they are mainly relative to the ATP hydrolysis rather than on the binding affinity [76]. Thus, the comparison of **G** with the other systems is based only on the theoretical prediction. Our simulation shows that there is a mild binding, -6.1 kcal mol^{-1}, and a similar entropic contribution to **T**, -39.1 kcal mol^{-1}. The convergence achieved in **C** is very good with a slope of $2.5 \cdot 10^{-3}$ kcal·mol^{-1}·ps^{-1}, whereas G shows and intermediate value of $2.5 \cdot 10^{-2}$ kcal·mol^{-1}·ps^{-1}, but still in an acceptable range.

Contrary to the PB method, in the GB approach each atom is modeled as a sphere with a radius referred as the effective Born radii. In order to increase the speed of computation, several methods have been proposed to evaluate the Born radii. The modified CFA method introduced by Hawkins et al. [35], denoted as IGB1, was one of the first introduced in the Amber code. This method, however, showed some problems related to the buried atoms. A rescaling approach was introduced to estimate more accurately the Born radii, with the GB^{OBC} methods, namely IGB2 and IGB5. The most recent model developed on the GB^{OBC} technique is the IGB7. In our calculations we compared all the methods available, considering two different radius definitions, bondi, IGB2, 5, 7, and mbondi2, IGB2′, 5′, 7′. In IGB2″ we used PBradii. All data summarized in Table 2 report the mean values for MM-GBSA energies, sem, and the energy difference with respect to the corresponding MM-PBSA results listed in Table 1.

The MM-GBSA results are very sensitive to the method used, even if the order of binding is preserved among the considered species. The methods IGB5 and IGB7, regardless off the radii definition, give an unrealistic result with overestimation of the binging free energy. IGB1 and IGB2, however, predict a binding free energy comparable with the PB method. The corresponding sem results are close with values between 3 and 7. IGB1 has a general tendency to overestimate the binding energy, in particular in **G** where the overbinding character is -42 kcal mol^{-1}. For the other species, the

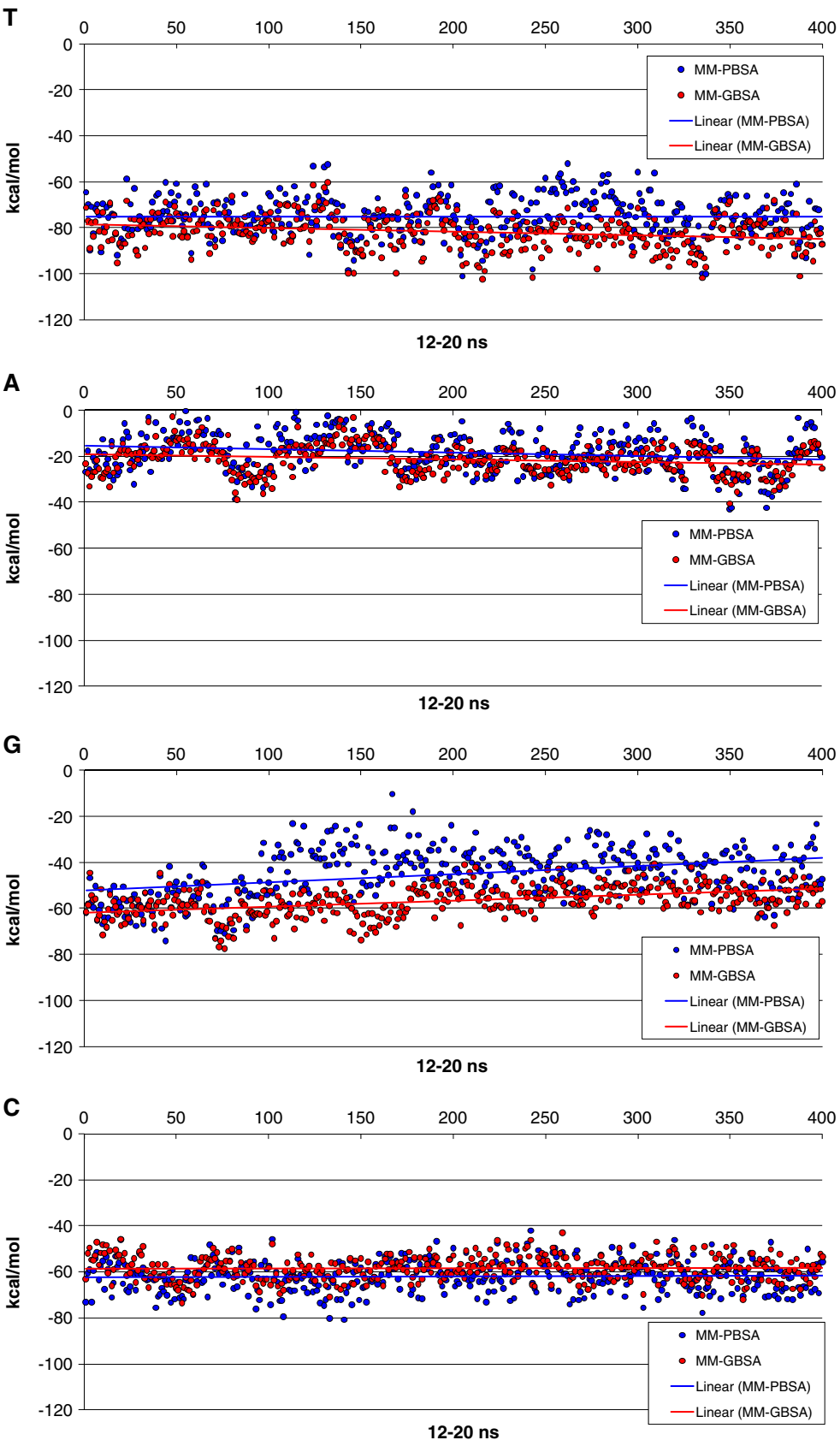

Fig. 5 MM-PB(GB)SA binding free energies for **T**, **A**, **G** and **C**, calculated in the last 8 ns of the trajectory. Each snapshot is taken at 20 ps time step

overbinding character is less pronounced, -28.1 kcal mol^{-1} for **T**, and -25.6 kcal mol^{-1} and -15.4 kcal mol^{-1} for **A** and **C** respectively. On the contrary, IGB2 shows a small underbinding character by only +7.4 kcal mol^{-1} for **C**, while **G** results over bonded by -7.1 kcal mol^{-1}. The corresponding values for **A** and **T** are very close to the PB results. The change of the radii definition effects the results. In IGB2— there is an overbinding tendency but not so pronounced as in IGB1, except for **C** where the binding free energy is higher by 1.2 kcal mol^{-1}. The IGB2—method provides values that are very close to IGB2—confirm a general overbinding, which is not uncommon [77], but still in much better agreement than IGB1. Because of the good agreement with the PB results, we used IGB2″ to analyze the residue contribution to the binding free energy, Fig. 6.

The MM-GBSA protocol allows to decompose the binding free energy in terms of contribution from each residue in a given structure. The data calculated with the IGB2″ method, Fig. 6, reports three components for each residue in the complex for the highest 20 contributions to the binding energy, $G_{gas+sol}$, the Coulombic interaction summed with the solvation free energy (coul + GB), the non polar (NP) contribution, and the van der Waals (vdW) energy. The results shown in Fig. 6 reveal few intriguing common features. First of all only about 15 aminoacids contribute to the overall binding free energy with values from -10 to -2 kcal mol^{-1}, confirming the complexity of the binding process and its strong cooperative character. For all cases arginine, positively charged, gives the major binding contribution, in particular in position 196 and 169. The former amino acid is located in the L2 binding site and it contributes by -9.8 kcal mol^{-1} to $G_{gas+sol}$ in **T**, -5.6 kcal mol^{-1} in **G**, -2.0 kcal mol^{-1} in **C**, and -8.4 kcal in **A**. The position 169, close to L1, is considered to have a marginal effect in the binding process. The calculations, however, revealed an unexpected strong role of Arg169 in the interaction with the oligomer, with contributions between -6 kcal mol^{-1} to -4 kcal mol^{-1}. Another aminoacid commonly involved in the interaction with dN_9 with a significant binding character is the aspargine in position 213. Minor contributions to the oveall $G_{gas+solv}$ are given by Thr210, Ile199, Met164, Gly211, Gly121, and Ala168. There are, moreover, few minor antibonding contributions due to the negatively charged ATP, by less than 2 kcal mol^{-1}, and to the Mg^{2+} ion, close to 1 kcal mol^{-1}.

To evaluate which domain in the complex is more relevant to the $G_{gas+solv}$ values, we grouped the IGB2″ contributions to the relevant part of the complex as L1, L2, the oligomers, and ATP, Table 3. The analysis shows that L1 has a minor contribution to the binding, with -7.9 kcal mol^{-1} for **G** and -5.0 kcal mol^{-1} for **T**, and almost a negligible portion for **A** and **C**. The role of Mg^{2+} and ATP is minor with a weight of less than a kcal mol^{-1}. L2, on the other side, has the most significant contribution to the $G_{gas+solv}$ with values from -23.3 kcal mol^{-1} for **T** to -8.6 kcal mol^{-1} for **A**. The affinity of the oligomer is sensitive to the nature of the nt used in the model. For instance, dT_9 contributes by -13.7 kcal mol^{-1} while dA_9 has an antibonding character of +13.7 kcal mol^{-1}.

The geometries of the averaged structures over the last 4 ns are reported in Fig. 7. L1 and L2 are shown in light blue and violet cartoon respectively. The labeled aminoacids are the ones located at a distance for which a Coulomb interaction with the oligomer is expected, dotted line. The ATP and Mg^{2+} are in gray. In **T**, the complex series of interaction between RecA and the oligomer delineated by the X-Ray structure [10] is globally maintained along the simulations except for some minor changes. In L1, Ser172 and Arg176 lose the bond with the dT_9 fragment in favor of the oxygen of the peptide group in Met164, the nitrogen of the peptide moiety of Gly165, and the amino terminal of Arg169. These results are consistent with the MM-GBSA energy per residue reported in Fig. 6. The NH fragment of Gly196 interacts with the back-bone of the oligomer, but not with a specific base justifying the presence of this interaction for all the series considered. Similarly, the peptidic NH interaction found in Gly211, Gly212, and Asn213 with the phosphate of dT_9 is encountered in all the complexes examined. In **A**, the number of effective bonding interactions has significantly decreased, where only few animoacids contribute to the binding free energy. Arg196 and Arg169 still have a major role in the binding, as well as Thr210. L1 shows a low affinity, as clearly moved far from the oligomer. This can be an interesting factor describing the low affinity of RecA with p(dA) observed experimentally. The complex **G**, consistent with the previous cases, has Arg196 and Arg169 as major components to $G_{gas+solv}$, together with Gly211, Asn213, and Gly212. The complex **C**, contrarily to the previous cases, has Asn213 and Ile199 providing the strongest binding contribution, whereas Arg196 contributes only marginally to the overall binding.

Discussion

To model the interaction between RecA and ssDNA the position of the binding loops, L1 and L2 (residues 157–164 and 195–209 respectively) has to be known. Chen et al. [10], were able to solve the X-Ray structure relative to the orientation of the loops, and describe their interaction with the sequence of ssDNA (pdb id 3cmu). The network of hydrogen bonds is rather complex: each nucleotide triplet interacts with three consecutive RecA monomers. The first nucleotide interacts with the closest RecA and with the next monomer proceeding in direction 5′ of the ssDNA. The

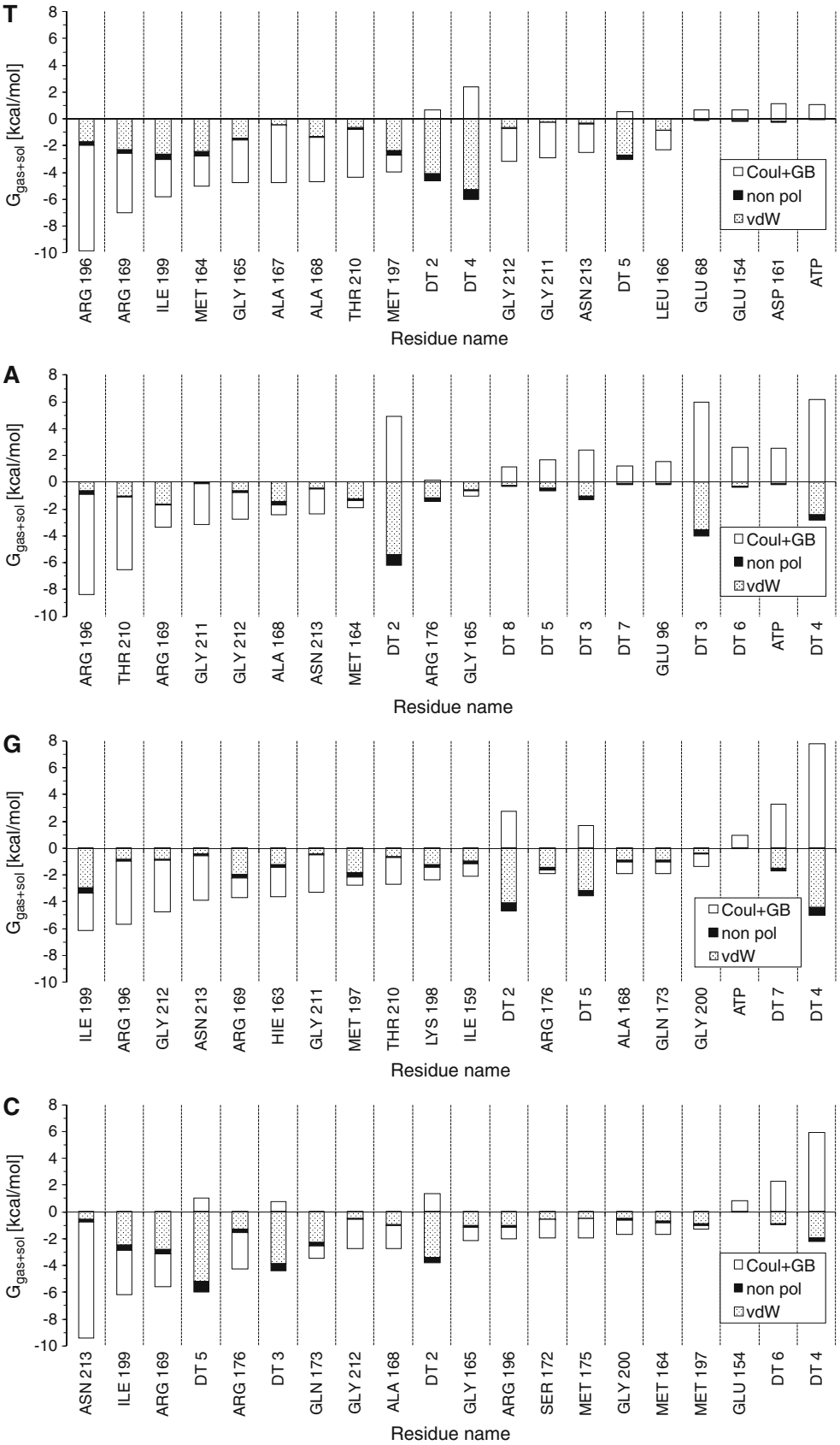

Fig. 6 Residue decomposition of the binding free energy calculated using the IGB2" method for **T**, **A**, **G** and **C**, calculated in the last 8 ns of the trajectory. The values show the contributions of the Coulombic interaction plus the salvation free energy (coul + GB), the non polar (NP), and van der Waals (vdW) contriutions

Table 3 Components of the binding free energy calculated with the IGB2″ model summed for each binding loops and, ATP, Mg^{2+} and dN_9

Contrib[a]	T	A	G	C
L1	-5.0	-2.2	-7.9	-0.8
L2	-23.6	-8.6	-21.6	-13.1
ATP	1.0	2.5	1.0	0.6
Mg^{2+}	-0.3	-0.2	-0.3	-0.3
dN_9	-13.7	13.7	1.1	-6.3
tot[b]	-41.5	5.2	-27.7	-19.8
Δ[c]	40.3	26.6	29.0	38.7

[(a)] Mean contributions to the MM-GBSA energy, using the IGB2" protocol, in kcal mol^{-1}, restricted to significant part of the complex. The values are averaged over 400 snapshots, 20 ps time interval, of the last 8 ns trajectories. The standard state is assumed to be at 50 mM. [(b)] sum of energy of the components considered. [(c)] Difference respect the corresponding IGB2" calculated for the entire complex, reported in Table 2.

second nucleotide interacts only with the closest RecA monomer. The third nucleotide is bound to the closest RecA and with the following monomer in direction 3′ of the filament. The hydrogen bonds between the backbone of the nucleotide triplet and the RecA monomers where depicted, involving Met197, for the first RecA, Asn213 Gly211 Gly212 from the closest RecA monomer and Ser172 and Arg176 with the following RecA monomer. The bases also are located to a position that might suggest an interaction with the RecA monomers. The first base has a van der Waals contact with Met197, Ile199, Gly200 of the first RecA, while the last base is close to Lys198, Ile199, and Thr208 of the third RecA monomer. Also the closest RecA, in the central position, is far from any base. Although the work of Chen depicted the structure of the RecA-p(dT) complex, many questions remain on the activity controlled by a growing network of functional interactions [52, 78–85]. The kinetic rates are particularly hard to determine, except for the work of Ha et al. [86] who provides relative values measured using a single molecule fluorescence assay.

Up to now, the mechanism of formation of the nucleoprotein filament, despite numerous years of study, is still unclear [53, 78, 87–89]. It was particularly elusive to determine which is the binding unit involved in the formation of the nucleoprotein filament [51, 90] since the oligomeric form of RecA is sensitive to the condition in which the protein is studied [91–96]. It has been observed that RecA must disassemble from the inactive form prior to reassembly on the nucleoprotein filament [89, 97]. The work of Knight et al. [90] indicates that the disassembly has not to continue up to the monomeric form, but even the dimeric RecA can have the role as nucleating unit for the assembly onto ssDNA. Often, however, the monomeric form of RecA can indeed start the nucleoprotein formation [16, 98]. The process of nucleation has a high degree of co-operativity [51, 53], even if the oligomeric form of the protein takes place in the assembly [90]. Several mechanisms were proposed to elucidate the kinetics of the formation of the nucleoprotein filament, by considering the presence of various reactions and functional states in equilibrium [80, 84, 85, 99–103]. Despite the efforts spent in understanding the RecA kinetics, a clear model capable of properly describing the RecA binding rate to ssDNA is also missing.

A strong contribution on the kinetics of the RecA binding was brought by Joo et al. [51], where, for the first time using very sophisticated single molecule FRET technique [104–106], the dissociation and binding unit and the corresponding rates were measured in proximity of both ends of the filament but near the ss-sdDNA junction.

Numerous studies have been conducted on the RecA activity in function of different mutants [17], however measurements of binding free energies for RecA to the corresponding mutants are absent in the literature. Camerini-Otero et al. [17] reveiled that, by an extensive site-directed mutagenesis of the L2 chain of EcRecA, there is a significant possibility that only a minor part of the loop is indeed involved in the reactivity of the enzyme [10]. This picture of RecA binding with ssDNA is consistent with a recent concept of interaction energy "hot spots", where only a resticted set of amino acids located at the binding interface make a dominant contribution to the binding free energy [107], leading to a distinction between functional epitope (hot spot residues) and structural epitope (all residues that participate in the interface) [108].

The lack of accurate experimental evaluation of the standard binding free energy prevents us from validating our results in a more quantitative manner; however a qualitative description was addressed. With this prospective, our purpose is to try to understand in more detail some of the crucial aspects of the binding process by using molecular dynamics (MD) simulations.

The main principle governing ligand receptor interactions are reasonably well known; however methods which allow a good estimate of the binding free energy of large ligands are still unavailable despite the unquestionable importance in bimolecular reactions [109–112]. In many cases physics-based approaches such as free energy perturbation based on molecular dynamics calculation (FEP/MD) are the most accurate for relative binding energy evaluations. The observable of relevance is the standard binding free energy, ΔG^0, often referred to in theoretical studies as "absolute" binding free energy, to distinguish it from the relative binding free energy calculated by FEP. Despite the ambiguity in the nomenclature, in order to calculate the ΔG^0, it is necessary to develop a mathematical model capable in order to extend the microscopic properties of the molecular model

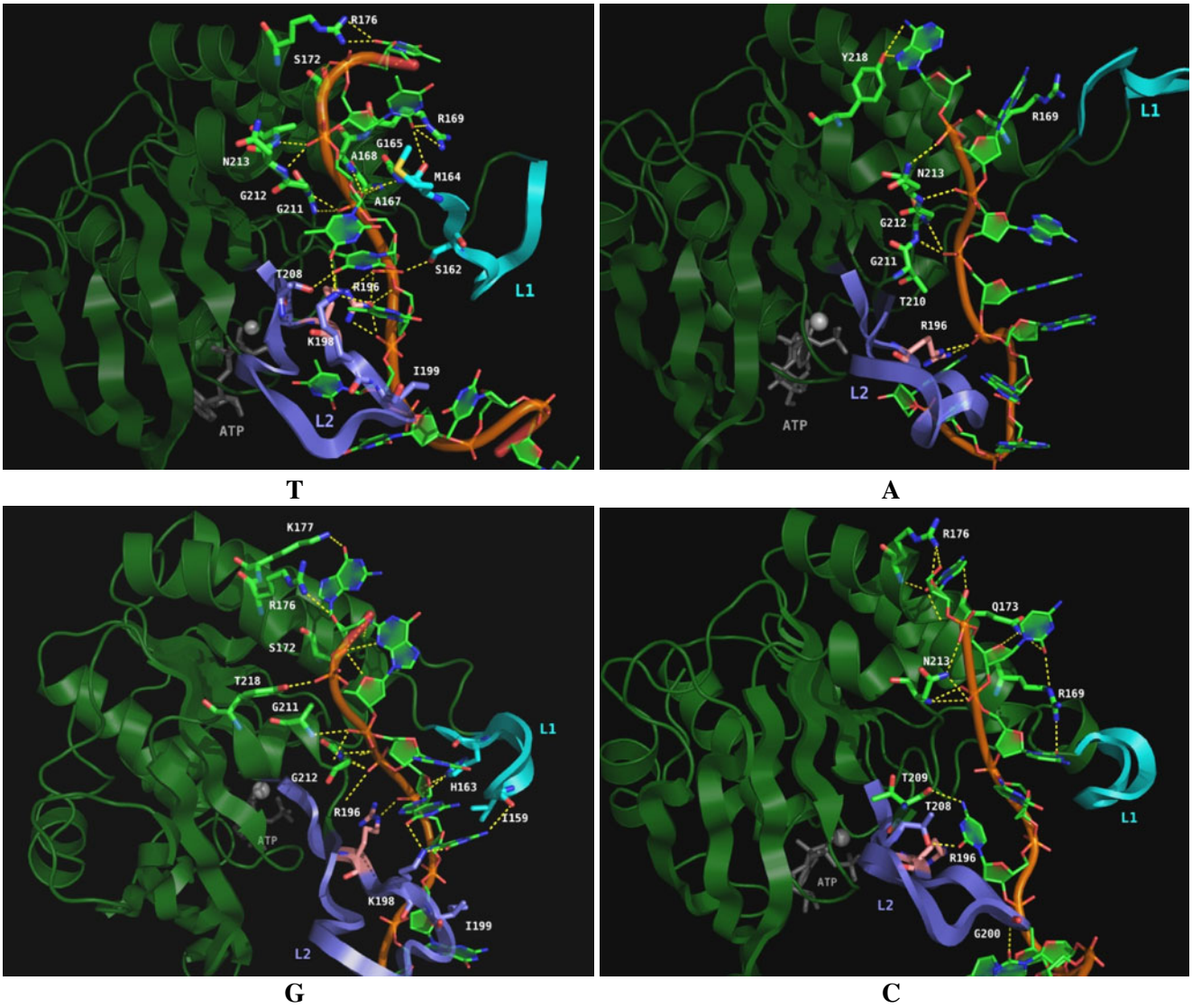

Fig. 7 Averaged structures of the last 4 ns of the trajectory for **T**, **A**, **C** and **G**. The binding loops L1 and L2 are shown in light blue and violet respectively. The polar contacts are represented in yellow, and in pink the ARG196 residue

to derive the value of the experimental measured quantity. Encouraging results were recently achieved by Roux [113, 114] and Gilson [115, 116], where the estimation were shown to have a reasonable accuracy. However their approach is designed for small ligands, significantly smaller than dN_9 considered in this work. Nevertheless, there are other methods which provided a good estimation of the binding free energy, such as the above mentioned FEP [117, 118], thermodynamical integration (TI) [119], and molecular mechanics generalized Born/Piosson-Bolzmann surface area (MM-GBSA/PBSA) [33, 120]. More recently, the MM-PBSA method has been well developed [27, 121] and applied to HIV reverse transcriptase [122], avidin [123], neuraminidase [124], cathespin D [125], growth factor receptor binding protein 2 [126], metallo-protease [127, 128] hystone deacetylase [129], DNA glycosilase [34], and protein-protein interface [130].

As above mentioned, the monomeric form of RecA initiates preferentially the formation of the nucleoprotein filament. This assumption supports our choice of the monomer of EcRecA as a model system to study the interaction with a ssDNA homopolymer. The size of the ligand, dN_9, prevents us from using the above mentioned methods. As an alternative, we estimated the absolute binding free energy by the MM-PG(GB)SA protocol, in combination wiyh a normal mode analysis for the solute binding entropy. This approach, regardless of the use of single or multiple trajectories, gives an approximate value of the standard binding free energy [114]. Because of these limitations, our study is centered on a qualitative evaluation of the binding free energies, which will permit us to describe the most relevant aspects of the binding process of RecA. Our model predicts, in full agreement

with experiment, the highest binding free energy for **T**, and **C**, and the weakest for **A** and **G**, confirming the highest affinity of RecA with homopolymers composed by pyridine residues, p(dT), p(dC), than homopolimers composed by purine residues, p(dG) and p(dA).

These results are in agreement with the binding enthalpy values measured by Wittung et al. [71] by isothermal titration calorimetry. The thermodynamic data showed that the enthalpy of interaction between RecA and $(dT)_{36}$ is 4 kcal mol^{-1} higher than the value measured for $(dA)_{36}$. The binding with $p(dC)_{36}$ gives values closer to $p(dA)_{36}$, consistently with our predictions, Table 1. A similar agreement is found also with the work of Bugreeva et al. [74] which shows by monitoring the ATP hydrolysis a higher affinity of RecA toward $(dT)_{20}$ versus a much lower affinity for $(dA)_{20}$, and intermediate activity with $d(C)_{20}$.

Even though experimentally, consistently with our results, the highest affinity of RecA is for p(dT), data on the binding and ATP hydrolysis relative to RacA interacting with p(dC) and p(dG) are still inconclusive [72, 75, 76, 131].

It has been reported that the nucleoside triphosphate cofactor, ATP, can induce a transient dissociation of the RecA filament [72, 87–89, 132], hence ATP plays a major role to mediate the interaction between the RecA and the single stranded DNA. According to the model, ATP has a small antibonding contribution, about a kcal mol^{-1} for all the complexes, which is partially balanced by the interaction component of its counterions, Mg^{2+}, which stabilizes the complex by a fraction of kcal mol^{-1}.

In our model the aminoacids involved in the interaction with the ssDNA are always in the binding loops or very close to them, but no contribution coming for other sides of the protein has been observed. This result although intuitively expected is not always verified; in fact few cases showed how the position of the binding units can be far from the interacting area of the enzyme with the substrate [33].

The results derived from the correlation analysis are fully consistent with the thermo chemistry. The correlation matrix, Fig. 2, shows how **T** and **C** have a smaller covariance values respect **A** and **C**, consistently with a higher binding affinity. The PCA analysis also predicts binding motion in L2 for **T** and **C**, contrarily to **A** and **G**. The energy landscape locates a deeper and localized basing for **T** and **C**, while for **A** and **G** the reduced potential energy surface is more characterized by the presence of several basins, confirming the lower binding tendency or RecA versus dG_9 and dA_9.

Conclusions

The goal of this study is to shed light on the modality of the interaction of the recombinases protein RecA with a series of ssDNA oligomers. To achieve this goal, a series of calculations aimed to estimate the absolute free binding energy were performed using the MM-PB(GB)SA protocol. The technique is based the Poisson-Boltzmann (PB) and the generalized Born (GB) solvent accessible surface area methods. The solute entropic contribution is also included and estimated by normal mode analysis. The results show how a strong contribution of the binding free energy is due to a minor number of aminoacids, in particular with Arg196, critical residue for the functionality of the enzyme. The data confirm how the free binding energy of RecA is significantly higher for dT_9 than for dA_9, in agreement with experimental results. Moreover, the data confirms the more controversial data still pointing at a higher affinity of RecA for dC rather than dG.

Even if our study is a qualitative approach to describe the binding interaction of the RecA monomer with a model sequences of ssDNA, the agreement with the experimental findings made the results useful to underline the major residue contributions to the binding free energy, and the geometry modification in function of the change of the ligand.

Acknowledgments We gratefully acknowledge support for this work from the NASA Space Radiation Risk Assessment Project.

References

1. Vessey CJ, Norbury CJ, Hickson ID (1999) Genetic disorders associated with cancer predisposition and genomic instability. Prog Nucleic Acid Res Mol Biol 63:189–221
2. Wood RD, Mitchell M, Sgouros J, Lindahl T (2001) Human DNA repair genes. Science 291:1284–1289
3. Durante M, Cucinotta FA (2008) Heavy ion carcinogenesis and human space exploration. Nat Rev Cancer 8:465–472
4. Scharer OD (2003) Chemistry and biology of DNA repair. Angew Chem Int Ed Engl 42:2946–2974
5. Lusetti SL, Cox MM (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. Annu Rev Biochem 71:71–100
6. Seitz EM, Brockman JP, Sandler SJ, Clark AJ, Kowalczykowski SC (1998) RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. Genes Dev 12:1248–1253
7. Ogawa T, Yu X, Shinohara A, Egelman EH (1993) Similarity of the yeast RAD51 filament to the bacterial RecA filament. Science 259:1896–1899
8. Register JC 3rd, Griffith J (1985) The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange. J Biol Chem 260:12308–12312
9. Story RM, Steitz TA (1992) Structure of the recA protein-ADP complex. Nature 355:374–376
10. Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature 453:489–484
11. Aihara H, Ito Y, Kurumizaka H, Terada T, Yokoyama S, Shibata T (1997) An interaction between a specified surface of the C-terminal domain of RecA protein and double-stranded DNA for homologous pairing. J Mol Biol 274:213–221

12. Kurumizaka H, Aihara H, Ikawa S, Kashima T, Bazemore LR, Kawasaki K, Sarai A, Radding CM, Shibata T (1996) A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. J Biol Chem 271:33515–33524
13. Roca AI, Cox MM (1997) RecA protein: structure, function, and role in recombinational DNA repair. Prog Nucleic Acid Res Mol Biol 56:129–223
14. Karlin S, Brocchieri L (1996) Evolutionary conservation of RecA genes in relation to protein structure and function. J Bacteriol 178:1881–1894
15. Brendel V, Brocchieri L, Sandler SJ, Clark AJ, Karlin S (1997) Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. J Mol Evol 44:528–541
16. Story RM, Bishop DK, Kleckner N, Steitz TA (1993) Structural relationship of bacterial RecA proteins to recombination proteins from bacteriophage T4 and yeast. Science 259:1892–1896
17. Hortnagel K, Voloshin ON, Kinal HH, Ma N, Schaffer-Judge C, Camerini-Otero RD (1999) Saturation mutagenesis of the E. coli RecA loop L2 homologous DNA pairing region reveals residues essential for recombination and recombinational repair. J Mol Biol 286:1097–1106
18. Balsera MA, Wriggers W, Oono Y, Shulten K (1996) Principal component analysis and long time protein dynamics. J Phys Chem 100:2567–2572
19. Meagher KL, Redman LT, Carlson HA (2003) Development of polyphosphate parameters for use with the AMBER force field. J Comput Chem 24:1016–1025
20. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2008) Amber 9. University of California, San Francisco
21. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE 3rd, Laughton CA, Orozco M (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophys J 92:3817–3829
22. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
23. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
24. Pastor RW, Brooks BR, Szabo A (1988) Mol Phys 65:1409–1419
25. Miyamoto S, Kollman PA (1992) J Comput Chem 13:952–962
26. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268:1144–1149
27. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE 3rd (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
28. Jorgensen WL (1989) Free energy calculations: a breakthrough for modeling organic chemistry in solution. Acc Chem Res 22:184–189
29. William LJ, Buckner JK, Stephane B, Julian T-R (1988) Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. J Chem Phys 89:3742–3746
30. Hermans J, Wang L (1997) Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. J Am Chem Soc 119:2707–2714
31. Roux B, Nina M, Pomes R, Smith JC (1996) Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. Biophys J 71:670–681
32. Reyes CM, Kollman PA (2000) Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. J Mol Biol 297:1145–1158
33. Gohlke H, Kiel C, Case DA (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. J Mol Biol 330:891–913
34. Olufsen M, Smalas AO, Brandsdal BO (2008) Electrostatic interactions play an essential role in DNA repair and cold-adaptation of uracil DNA glycosylase. J Mol Model 14:201–213
35. Hawkins GD, Cramer CJ, Truhlar DG (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 100:19824–19839
36. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized Born modelsuitable for macromolecules. J Phys Chem B 104:3712–3720
37. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 55:383–394
38. Mongan J, Simmerling C, McCammon JA, Case D, Onufriev A (2006) Generalized Born with a simple, robust molecular volume correction. J Chem Theory Comput 3:156–169
39. Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 98:1978–1988
40. Ichiye T, Karplus M (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 11:205–217
41. Rod TH, Radkiewicz JL, Brooks CL 3rd (2003) Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proc Natl Acad Sci USA 100:6980–6985
42. Radkiewicz JL, Charles CL (2000) protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. J Am Chem Soc 122:225–231
43. Prompers JJ, Bruschweiler R (2002) Dynamic and structural analysis of isotropically distributed molecular ensembles. Proteins 46:177–189
44. Prompers JJ, Bruschweiler R (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. J Am Chem Soc 124:4522–4534
45. Hayward S, Kitao A, Hirata F, Go N (1993) Effect of solvent on collective motions in globular protein. J Mol Biol 234:1207–1217
46. Amadei A, Linssen AB, de Groot BL, van Aalten DM, Berendsen HJ (1996) An efficient method for sampling the essential subspace of proteins. J Biomol Struct Dyn 13:615–625
47. van Aalten DM, Amadei A, Linssen AB, Eijsink VG, Vriend G, Berendsen HJ (1995) The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. Proteins 22:45–54
48. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. Proteins 17:412–425
49. Meyer T, Ferrer-Costa C, Pérez A, Rueda M, Bidon-Chanal A, Luque FJ, Laughton CA, Orozco M (2006) Essential dynamics: a tool for efficient trajectory compression and management. J Chem Theory Comput 2:251–258
50. Tai K, Shen T, Borjesson U, Philippopoulos M, McCammon JA (2001) Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. Biophys J 81:715–724

51. Joo C, McKinney SA, Nakamura M, Rasnik I, Myong S, Ha T (2006) Real-time observation of RecA filament dynamics with single monomer resolution. Cell 126:515–527
52. Galletto R, Amitani I, Baskin RJ, Kowalczykowski SC (2006) Direct observation of individual RecA filaments assembling on single DNA molecules. Nature 443:875–878
53. Menetski JP, Kowalczykowski SC (1985) Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulation by nucleotide cofactors. J Mol Biol 181:281–295
54. McGhee JD, von Hippel PH (1974) Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. J Mol Biol 86:469–489
55. Abagyan R, Argos P (1992) Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 225:519–532
56. Barrett CP, Noble ME (2005) Molecular motions of human cyclin-dependent kinase 2. J Biol Chem 280:13993–14005
57. Troyer JM, Cohen FE (1995) Protein conformational landscapes: energy minimization and clustering of a long molecular dynamics trajectory. Proteins 23:97–110
58. Caves LS, Evanseck JD, Karplus M (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 7:649–666
59. Becker OR (1998) J Comput Chem 19:1255–1267
60. Becker MO (1997) Quantitative visualization of a molecular potential energy "funnel". J Mol Struct THEOCHEM398–399:507–516
61. Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. Q Rev Biophys 38:405–410
62. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations. Proc Natl Acad Sci USA 103:11844–11849
63. Levy Y, Cho SS, Onuchic JN, Wolynes PG (2005) A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. J Mol Biol 346:1121–1145
64. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. Proc Natl Acad Sci USA 101:511–516
65. Papoian GA, Wolynes PG (2003) The physics and bioinformatics of binding and folding-an energy landscape perspective. Biopolymers 68:333–349
66. Miller DW, Dill KA (1997) Ligand binding to proteins: the binding landscape model. Protein Sci 6:2166–2179
67. Wang J, Verkhivker GM (2003) Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. Phys Rev Lett 90:188181
68. Huo S, Massova I, Kollman PA (2002) Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J Comput Chem 23:15–27
69. Tsui V, Case DA (2001) Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. J Phys Chem B 105:11314–11325
70. Sitkoff D, Sharp KA, Honig B (1994) Accurate Calculation of hydration free energies using macroscopic solvent models. J Phys Chem 98:1978–1988
71. Wittung P, Ellouze C, Maraboeuf F, Takahashi M, Norden B (1997) Thermochemical and kinetic evidence for nucleotide-sequence-dependent RecA-DNA interactions. Eur J Biochem 245:715–719
72. Cazenave C, Chabbert M, Toulme JJ, Helene C (1984) Absorption and fluorescence studies of the binding of the recA gene product from E. coli to single-stranded and double-stranded DNA. Ionic strength dependence. Biochim Biophys Acta 781:7–13
73. Bugreeva IP, Bugreev DV, Nevinskii GA (2005) Physico-chemical basis of RecA nucleo-protein filament formation on single. Mol Biol (Mosk) 39:984–998
74. Bugreeva IP, Bugreev DV, Nevinsky GA (2005) Formation of nucleoprotein RecA filament on single-stranded DNA. Analysis by stepwise increase in ligand complexity. Febs J 272:2734–2745
75. McEntee K, Weinstock GM, Lehman IR (1981) Binding of the recA protein of Escherichia coli to single- and double-stranded DNA. J Biol Chem 256:8835–8844
76. Amaratunga M, Benight AS (1988) DNA sequence dependence of ATP hydrolysis by RecA protein. Biochem Biophys Res Commun 157:127–133
77. Gohlke H, Case DA (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 25:238–250
78. Defais M, Phez E, Johnson NP (2003) Kinetic mechanism for the formation of the presynaptic complex of the bacterial recombinase RecA. J Biol Chem 278:3545–3551
79. Cox MM (2007) Regulation of bacterial RecA protein function. Crit Rev Biochem Mol Biol 42:41–63
80. Kowalczykowski SC, Paul LS, Lonberg N, Newport JW, McSwiggen JA, von Hippel PH (1986) Cooperative and noncooperative binding of protein ligands to nucleic acid lattices: experimental approaches to the determination of thermodynamic parameters. Biochemistry 25:1226–1240
81. Cazaux C, Blanchet JS, Dupuis D, Villani G, Defais M, Johnson NP (1998) Investigation of the secondary DNA-binding site of the bacterial recombinase RecA. J Biol Chem 273:28799–28804
82. Gourves AS, Tanguy Le Gac N, Villani G, Boehmer PE, Johnson NP (2000) Equilibrium binding of single-stranded DNA with herpes simplex virus type I-coded single-stranded DNA-binding protein, ICP8. J Biol Chem 275:10864–10869
83. De Zutter JK, Knight KL (1999) The hRad51 and RecA proteins show significant differences in cooperative binding to single-stranded DNA. J Mol Biol 293:769–780
84. Chabbert M, Cazenave C, Helene C (1987) Kinetic studies of recA protein binding to a fluorescent single-stranded polynucleotide. Biochemistry 26:2218–2225
85. Morrical SW, Cox MM (1985) Light scattering studies of the recA protein of Escherichia coli: relationship between free recA filaments and the recA X ssDNA complex. Biochemistry 24:760–767
86. McKinney SA, Joo C, Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophys J 91:1941–1951
87. Takahashi M, Kubista M, Norden B (1989) Binding stoichiometry and structure of RecA-DNA complexes studied by flow linear dichroism and fluorescence spectroscopy. Evidence for multiple heterogeneous DNA co-ordination. J Mol Biol 205:137–147
88. Cox MM, Soltis DA, Lehman IR, DeBrosse C, Benkovic SJ (1983) ADP-mediated dissociation of stable complexes of recA protein and single-stranded DNA. J Biol Chem 258:2586–2592
89. Lee JW, Cox MM (1990) Inhibition of recA protein promoted ATP hydrolysis. 1. ATP gamma S and ADP are antagonistic inhibitors. Biochemistry 29:7666–7676
90. Forget AL, Kudron MM, McGrew DA, Calmann MA, Schiffer CA, Knight KL (2006) RecA dimers serve as a functional unit for assembly of active nucleoprotein filaments. Biochemistry 45:13537–13542
91. Brenner SL, Zlotnick A, Griffith JD (1988) RecA protein self-assembly. Multiple discrete aggregation states. J Mol Biol 204:959–972
92. Brenner SL, Zlotnick A, Stafford WF 3rd (1990) RecA protein self-assembly. II. Analytical equilibrium ultracentrifugation

studies of the entropy-driven self-association of RecA. J Mol Biol 216:949–964
93. Wilson DH, Benight AS (1990) Kinetic analysis of the pre-equilibrium steps in the self-assembly of RecA protein from Escherichia coli. J Biol Chem 265:7351–7359
94. DiCapua E, Schnarr M, Ruigrok RW, Lindner P, Timmins PA (1990) Complexes of RecA protein in solution. A study by small angle neutron scattering. J Mol Biol 214:557–570
95. Heuser J, Griffith J (1989) Visualization of RecA protein and its complexes with DNA by quick-freeze/deep-etch electron microscopy. J Mol Biol 210:473–484
96. Ruigrok RW, DiCapua E (1991) On the polymerization state of recA in the absence of DNA. Biochimie 73:191–198
97. Yu X, Egelman EH (1992) Structural data suggest that the active and inactive forms of the RecA filament are not simply interconvertible. J Mol Biol 227:334–346
98. Masui R, Mikawa T, Kato R, Kuramitsu S (1998) Characterization of the oligomeric states of RecA protein: monomeric RecA protein can form a nucleoprotein filament. Biochemistry 37:14788–14797
99. Haruta N, Yu X, Yang S, Egelman EH, Cox MM (2003) A DNA pairing-enhanced conformation of bacterial RecA proteins. J Biol Chem 278:52710–52723
100. Bugreeva IP, Bugreev DV, Nevinskii GA (2005) Physico-chemical basis of RecA nucleo-protein filament formation on single. Mol Biol (Mosk) 39:984–998
101. Bar-Ziv R, Libchaber A (2001) Effects of DNA sequence and structure on binding of RecA to single-stranded DNA. Proc Natl Acad Sci USA 98:9068–9073
102. Silver MS, Fersht AR (1982) Direct observation of complexes formed between recA protein and a fluorescent single-stranded deoxyribonucleic acid derivative. Biochemistry 21:6066–6072
103. Gourves AS, Defais M, Johnson NP (2001) Equilibrium binding of single-stranded DNA to the secondary DNA binding site of the bacterial recombinase RecA. J Biol Chem 276:9613–9619
104. Ha T (2001) Single-molecule fluorescence methods for the study of nucleic acids. Curr Opin Struct Biol 11:287–292
105. Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, Weiss S (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. Proc Natl Acad Sci USA 93:6264–6268
106. Weiss S (1999) Fluorescence spectroscopy of single biomolecules. Science 283:1676–1683
107. Lafont V, Schaefer M, Stote RH, Altschuh D, Dejaegere A (2007) Protein-protein recognition and interaction hot spots in an antigen-antibody complex: free energy decomposition identifies “efficient amino acids”. Proteins 67:418–434
108. Cunningham BC, Wells JA (1993) Comparison of a structural and a functional epitope. J Mol Biol 234:554–563
109. Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 18:431–492
110. Straatsma TP, McCammon JA Computational Alchemy. Annu Rev Phys Chem 43:407–435
111. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417
112. Simonson T, Archontis G, Karplus M (2002) Free energy simulations come of age: protein-ligand recognition. Acc Chem Res 35:430–437
113. Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113:2234–2246
114. Wang J, Deng Y, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. Biophys J 91:2798–2814
115. Mardis KL, Luo R, Gilson MK (2001) Interpreting trends in the binding of cyclic ureas to HIV-1 protease. J Mol Biol 309: 507–517
116. Chen W, Chang CE, Gilson MK (2004) Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. Biophys J 87:3035–3049
117. Miyamoto S, Kollman PA (1993) Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. Proteins 16:226–245
118. Luzhkov VB, Almlof M, Nervall M, Aqvist J (2006) Computational study of the binding affinity and selectivity of the bacterial ammonium transporter AmtB. Biochemistry 45:10807–10814
119. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl 41:2644–2676
120. Srinivasan J, Miller J, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of RNA hairpin loops and helices. J Biomol Struct Dyn 16:671–682
121. Massova I, Kollman PA (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov Des 18:113–135
122. Zhou Z, Madrid M, Evanseck JD, Madura JD (2005) Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase. J Am Chem Soc 127:17253–17260
123. Kuhn B, Kollman PA (2000) Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J Med Chem 43:3786–3791
124. Masukawa KM, Kollman PA, Kuntz ID (2003) Investigation of neuraminidase-substrate recognition using molecular dynamics and free energy calculations. J Med Chem 46:5628–5637
125. Huo S, Wang J, Cieplak P, Kollman PA, Kuntz ID (2002) Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. J Med Chem 45:1412–1419
126. Wang W, Lim WA, Jakalian A, Wang J, Wang J, Luo R, Bayly CI, Kollman PA (2001) An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. J Am Chem Soc 123:3986–3994
127. Donini OA, Kollman PA (2000) Calculation and prediction of binding free energies for the matrix metalloproteinases. J Med Chem 43:4180–4188
128. Hou TJ, Guo SL, Xu XJ (2002) Predictions of binding of a diverse set of ligands to gelatinase-A by a Combination of molecular dynamics and continuum solvent models. J Phys Chem B 106:5527–5535
129. Yan C, Xiu Z, Li X, Li S, Hao C, Teng H (2008) Comparative molecular dynamics simulations of histone deacetylase-like protein: binding modes and free energy analysis to hydroxamic acid inhibitors. Proteins 73:134–149
130. Wong S, Amaro RE, McCammon JA (2009) MM-PBSA captures key role of intercalating water molecules at a protein-protein interface. J Chem Theory Comput 5:422–429
131. Tracy RB, Kowalczykowski SC (1996) In vitro selection of preferred DNA pairing sequences by the Escherichia coli RecA protein. Genes Dev 10:1890–1903
132. Menetski JPK, Stephen C (1985) Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulation by nucleotide cofactors. J Mol Biol 181:281–295