J Mol Model (2012) 18:2761–2783
DOI 10.1007/s00894-011-1288-9

ORIGINAL PAPER

# Accurate prediction of the binding free energy and analysis of the mechanism of the interaction of replication protein A (RPA) with ssDNA

**Claudio Carra · Francis A. Cucinotta**

Received: 14 March 2011 / Accepted: 19 October 2011 / Published online: 25 November 2011


**Abstract** The eukaryotic replication protein A (RPA) has several pivotal functions in the cell metabolism, such as chromosomal replication, prevention of hairpin formation, DNA repair and recombination, and signaling after DNA damage. Moreover, RPA seems to have a crucial role in organizing the sequential assembly of DNA processing proteins along single stranded DNA (ssDNA). The strong RPA affinity for ssDNA, $K_A$ between $10^{-9}-10^{-10}$ M, is characterized by a low cooperativity with minor variation for changes on the nucleotide sequence. Recently, new data on RPA interactions was reported, including the binding free energy of the complex RPA70AB with $dC_8$ and $dC_5$, which has been estimated to be $-10\pm0.4$ kcal mol$^{-1}$ and $-7\pm1$ kcal mol$^{-1}$, respectively. In view of these results we performed a study based on molecular dynamics aimed to reproduce the absolute binding free energy of RPA70AB with the $dC_5$ and $dC_8$ oligonucleotides. We used several tools to analyze the binding free energy, rigidity, and time evolution of the complex. The results obtained by MM-PBSA method, with the use of ligand free geometry as a reference for the receptor in the separate trajectory approach, are in excellent agreement with the experimental data, with ±4 kcal mol$^{-1}$ error. This result shows that the MM-PB(GB)SA methods can provide accurate quantitative estimates of the binding free energy for interacting complexes when appropriate geometries are used for the receptor, ligand and complex. The decomposition of the MM-GBSA energy for each residue in the receptor allowed us to correlate the change of the affinity of the mutated protein with the $\Delta G_{gas+sol}$ contribution of the residue considered in the mutation. The agreement with experiment is optimal and a strong change in the binding free energy can be considered as the dominant factor in the loss for the binding affinity resulting from mutation.

**Keywords** Amber · Binding · MM-PBSA · Molecular dynamics · Replication protein · RPA

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

## Introduction

The eukaryotic replication protein A (RPA), identified as a heterotrimeric ssDNA-binding protein, is involved in the chromosomal replication, repair and recombination pathways in eukaryotic cells, and more recently shown to participate in signaling following DNA damage [1–9]. One of the functions of RPA is to protect the single stranded DNA, ssDNA, from nucleases and prevent ssDNA from the hairpin formation. Moreover, RPA has a pivotal role in organizing the sequential assembly of DNA processing proteins along the ssDNA [10, 11]. The human homologue subunits of RPA are named according to their respective molecular weight, RPA70, RPA32 and RPA14 [3, 8, 12, 13]. The X-ray structure of RPA shows six domains that assume the oligonucleotide/oligosaccharide-binding (OB) fold geometry found in other known ssDNA-binding proteins, SSBs [2, 14, 15]. RPA has four ssDNA binding domains, denoted as A (amino acids 181–290), B (amino acids 300–420), C (amino acids 436–616), and D (amino acids 43–170), in order of

decreasing affinity. The first three are located in the subunit RPA70, the remaining binding domain, D, is found in RPA32 subunit. The binding modes depend on the length of the substrate, with 8–10, 12–23 and 28–30 nucleotides, nt, respectively [16–18]. Also, the N terminal of RPA70 has some affinity with ssDNA, however its contribution to the overall binding is predicted to be mostly regulatory under specific conditions [19–21]. The presence of three main binding modes adopted by RPA to bind ssDNA, suggests that three different conformations can be assumed by the protein in solution. This hypothesis was reconfirmed by scanning transmission electron microscopy and gel filtration which lead to the confirmation that all three conformations coexist, likely in equilibrium, in solution [22].

RPA70A has an affinity for a series of proteins, including the papilloma virus E1 helicase [23], the SV40T antigen [23, 24], XPA [25], and the human Rad51 recombinase [26]. Mutations on Rad51 show that the interaction between the N-terminus of Rad51 with RPA70A plays a role in the RPA displacement from the ssDNA in the formation of the Rad51-ssDNA nucleoprotein filament [26]. Several DNA damage signaling and processing proteins revealed an interaction with RPA: Rad17 [27, 28], Rad9 [29], ATRIP [9, 30, 31], 53BPI [32], BRCA2 [33], Mre11-Rad50-Nbs1 [34], and nucleolin [35, 36].

RPA has a strong affinity (association constant ($K_A$) between $10^{-9}$–$10^{-10}$ M [37]) for ssDNA with a defined 5′→3′ polarity [38, 39]. The binding shows a low cooperativity and its strength is mostly a function of the length of the substrate, with minor variation for changes on the nucleotide sequence and experimental conditions [37, 40–42]. The individual affinity of each subunit for the ssDNA is rather small, with greatest values measured for RPA70A being $K_D$ ~2 μM. In the following step, RPA70B, separated by RPA70A by a short linkage, binds the ssDNA, with an overall affinity enhanced by several orders of magnitude [43]. In the next step, the RPA70AB is joined by RPA70 C, as shown by a protection of the linker 420–430 from proteolytic cleavage [44]. The bonding progresses involving all the remaining subunit, reaching a high affinity binding mode that form a stable complex with a ~30nt mer length [16, 17, 38, 39, 43, 45].

The X-ray structure of the entire RPA-ssDNA complex, including the subunit orientation, is not known, even with some interesting attempts to shed light on the overall structure [46, 47].

The difficulty in solving the complete structure is due to the profound conformational changes RPA goes through binding to ssDNA. Experimentally it has been observed that the structure of the free RPA70AB unveils a higher separation of the domains 70A and 70B with respect to the active RPA interacting with ssDNA [2, 16, 22, 48–51]. The linker (amino acids 291–299) between the two domains is expected to be flexible, as independently confirmed by NMR studies on RPA70AB in solution [43].

The X-ray structure of RPA70AB binding $dC_8$ was determined [51], and it shows a series of complex base specific hydrogen bonds between the residues located in L12 and L45 of the protein and the oligomer. This result is rather puzzling in view of the very low binding selectivity that RPA has for the base sequence of ssDNA [43]. The non-specific binding can be justified as a dynamic remodeling of the binding sites [52].

The interaction of RPA with double stranded DNA, dsDNA, is not fully clarified, and no X-ray structures are available to date, however the affinity of RPA for dsDNA is found to be significantly lower, about three orders of magnitude with respect to ssDNA, with some variations due to experimental conditions [8, 53]. The preferential binding behavior of RPA is observed being toward a damaged dsDNA rather than an undamaged chain, and the affinity is a function of the damaged DNA structure; in particular the binding is more favorable where disruptive damages lead to a ssDNA. [54–56].

To investigate the binding mechanism, we examined several aspects of the protein interacting with ssDNA. The mobility of the complexes is analyzed by comparing the averaged root mean square deviation (RMSD) values as a function of residue number. The study is extended to the analysis of the correlated and uncorrelated motions shown by the cross correlation matrix plotted for our systems. We restricted the configuration space by performing a Principal Component Analysis (PCA) on the trajectory of the molecular dynamics. This approach is intend 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 [57]. Our interest, however, is not only in the conformational analysis of the complexes, but also the estimation of the binding free energy. Although the main aspects guiding the interaction of ligand with the receptor are reasonably well understood, methods to estimate the binding free energy, $\Delta G^0$, of large ligands are still unavailable despite their dominant role in bimolecular simulations [58–61]. We focus our attention on the molecular mechanics Generalized Born/Poisson-Boltzmann surface area (MM-GBSA/PBSA) approach [62, 63].

The goal of this work is to understand the interaction of the RPA70AB protein with $dC_8$ and $dC_5$ for which the binding free energy is experimentally available. We also consider the possibility to assign the S-S bond between Cys200 and Cys289, which are overstretched in the X-ray structure of the free protein. The results reveal how the use of the separate trajectories improves the accuracy of the prediction of the binding free energy. Harmonic and quasi-harmonic approximations are used to estimate the binding

entropy. In addition, if the geometry of the receptor is taken from the X-ray structure of the free RPA70AB, the agreement with experiment is remarkably improved. In particular, if the geometry containing the S-S bond between Cys200 and Cys289 is considered, the comparison with the experiment is within 4 kcal $mol^{-1}$ for both complexes with $dC_5$ and $dC_8$. The solvent effect is also analyzed by comparing the change of binding free energy in function of different solute concentrations.

## Materials and methods

The geometry of the of RPA70AB interacting with $dC_8$, **1**, was constructed by considering part of the crystal structure determined by Bochkarev et al. [51] (pdb id 1jmc). The same geometry was used to build the complex of RPA70AB with a smaller oligomer of 5 nt, **2**, interacting with RPA70A. To calculate the binding free energy using a separate trajectories method, we used the geometry of RPA70AB from the complex without the ligand, **3**, and the X-ray structure of the receptor reported by Bochkareva et al. [16] (pdb id 1fgu) obtained in absence of the ligand, **4**. The latter geometry contains a higher number of residues, thus a few amino acids not present in the former complex were removed to maintain the same number of atoms. The two crystal structures, however, revealed an inconsistency on the sulfur distance between two cysteine residues in positionS 200 and 289. Before starting the simulation, the X-ray structures were optimized using the WHAT IF series of programs [64]. In the geometry of the free RPA70AB the S-S separation is 4.2 Å, despite the WHAT IF refinement, contrary to the structure obtained with the ligand where the sulfur bond is unequivocally present. To our knowledge, additional information on this issue are absent in the literature. To investigate this apparent ambiguity, we performed two series of separated simulations considering the presence of the cysteine bond, structure **1**, **2**, **3**, and **4**, and the absence of the cysteine bond, **1′**, **2′**, **3′**, and **4′**, Fig. 1.

The molecular dynamics simulations and the data analysis were carried out with the AMBER 10 [65] package using the parmbsc0 force field [66]. The solute was modeled in a periodic box with a 8 Å buffer of water molecules explicitly described by the TIP3P [67] model. The particle mesh Ewald (PME) method [68] was used to treat the long range electrostatic interactions.

The equilibration of the system was performed as follows: an initial optimization of 20,000 cycles, the first 10,000 by steepest descent then followed by the conjugate gradient method. The complex is constrained to relax the solvent. Then a further optimization of 30,000 cycles with no constraints on the whole system was carried out to lead to a final relaxed geometry. The first equilibration was carried

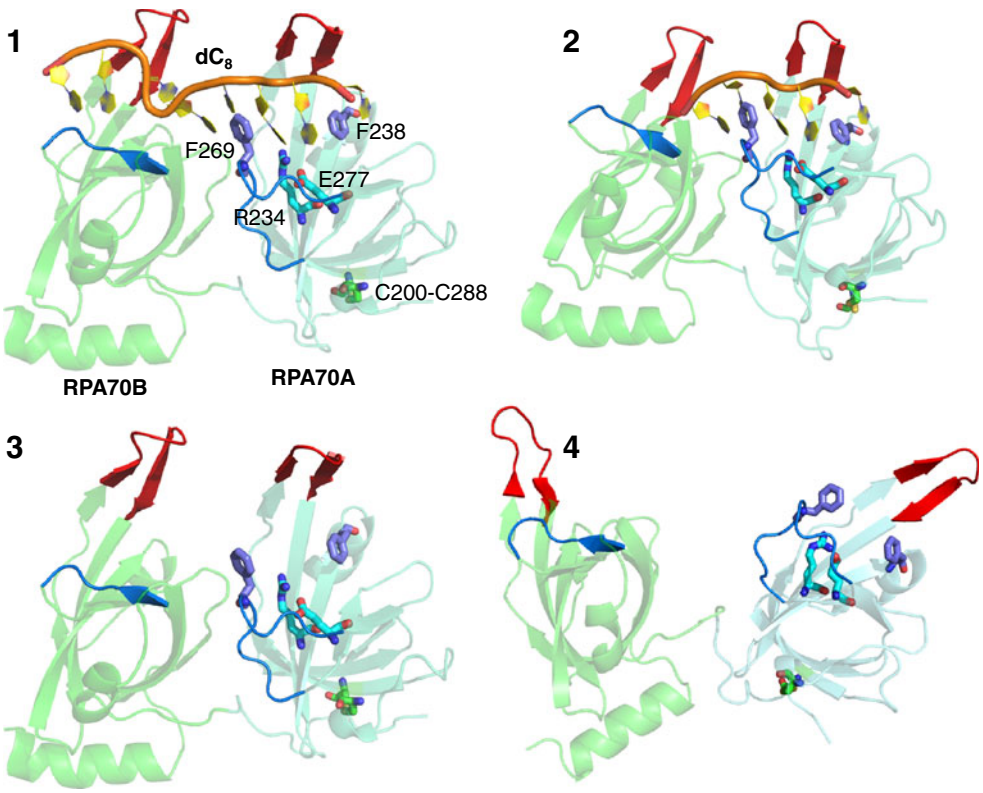


**Fig. 1** The picture of the X-ray structure for the system examined. The RPA70AB in complex with $dC_8$, **1**, and with $dC_5$, **2**. The free RPA70AB from the complex structure, **3**, and from the experimental geometry, **4**. When the sulfur bonds between CYS200 and CYS289 is not considered, the corresponding structures are labeled **1′**, **2′**, **3′**, **4′**. The binding loops are colored in red for L12 and in blue for L45 for both subunits A and B. The hydrophobic mutations are represented in violet, F234 and F269, and the hydrophilic mutations are shown in cyan, F238 and E277

out with a weak restraint on the complex for 100 ps at constant volume, constantly increasing the temperature from 0 to 300 K. The equilibration continued for 200 ps at a constant pressure of 1 atm, by keeping the temperature constant with the Langevin temperature equilibration scheme [69] using a collision frequency of 1.0 $ps^{-1}$. Under these conditions the restraints were gradually removed. The production run was carried out without restraints for 24 ns. During the MD calculation, hydrogen stretching motions were removed using SHAKE bond constraints [70], allowing a longer time step of 2 fs without introducing any instability. Free binding energies of the complex were calculated with the molecular mechanics Poisson Boltzmann and generalized Born surface area method, MM-BP(GB)SA. The binding free energy is calculated by taking the average energy difference between the complex, $\overline{G}_{comp}$, and the reactants, $\overline{G}_{rec} + \overline{G}_{lig}$,

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

where the average free energy, $\overline{G}$, for the complex, RPA70AB-$dC_n$, receptor, RPA70AB, and ligand, $dC_n$, 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 component of the electrostatic energy calculated with the Poisson-Boltzmann (PB) [71], or generalized Born (GB) [72] method. One of the advantages of using the MM-PB(GB)SA method is that the "nonphysical" annihilation [73, 74] or decoupling [75, 76] of the species alone in solution or bounded to a substrate is not required anymore. Moreover, it is not necessary to model the partially unbound states as demanded using umbrella sampling. It has already been shown that MM-PB(GB)SA method was able to qualitatively reproduce well the binding free energy of such systems [63, 77, 78]. We used several GB methods to compare the adaptability of different protocols to our systems. We consider the method of Onufriev et al. [79], IGB2, with mbondi radii definition. The $\Delta G_{SA}$ represents the non-polar contribution to the solvation free energy which is determined with solvent-accessible-surface-area-dependent terms (SA) [80] approach. The term $T\Delta S^{(s)}$ is the conformational entropy change of the solute. 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 get closer to the experimental conditions, the concentration of the salt in the bulk of the solution was set at 50 mM.

The solute binding entropy was calculated by normal mode, NM, analysis for the single trajectory approach, where the standard state is assumed to be at 1 M [81]. For comparison, the quasi harmonic, QH, approximation was used by analyzing the trajectories with ptraj module in Amber 10 by extracting Cα and P atoms for the complex, Cα atoms for the receptor and P atoms for the ligand. With the single trajectory approach, the solute binding entropy is converged and the results are similar to the NM predictions (data not shown). On separate trajectories method, there is no convergence in the QH entropy estimation (data not shown), supposedly due to a cumulating error in the covariance matrix [82]. Due to the lack of software to estimate correctly the conformational entropy by QH method for systems involving DNA, we used the NM analysis also in the separate trajectory case. In order to investigate motion between different regions in the protein, as domain-domain communication [83–85] or between ligand and receptor, we calculated the correlation matrix for the binding complex, displayed as a two dimension correlation map. A positive value shows that the atoms are moving in the same direction, whereas a negative value indicates an anti-correlated motion. We used ptraj applet in the Amber code, to generate the correlation matrix and Matlab to generate the 3D plots.

In MD calculations, the coordinate trajectory matrix has the dimensions of (3 N, *n*) where N is the number of atoms and *n* the number of snapshots selected for the analysis. The resulting conformational space has a high dimensionality and does not allow a simple description or a visualization analysis. The use of the principal component analysis (PCA), methods that reduces the dimensionality of the conformation space, can allow for the depiction of extreme complex structures and the major fluctuations of the correlated motions. This method projects the multidimensional conformational space onto a new set of axis which maximizes the variance of the projection along orthogonal directions. Such a projection enables low-dimensional representation of the spatial relationship between conformations. With this approach, the analysis starts by diagonalization of the covariance matrix $\sigma_{m,w}$ whose individual element is:

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

By projecting the MD trajectory onto the main essential directions, corresponding to the larger eigenvectors, one can visualize the extreme structures and the major fluctuations of the correlated motions. The histograms obtained by the projection of the MD trajectories, where we considered only the Cα and P atoms because of the memory limitations, onto the first two principal eigenvectors are visualized with Matlab. The PCA [86–91] analysis was carried out with PCAZIP software [92], and the collective dynamic modes are plotted using the porcupine method developed by Tai et al. [93].

## Results

### Structure flexibility

The model of the complexes of RPA70AB interacting with $dC_8$ and $dC_5$, **1** and **2**, was constructed on the crystal structure of the active complex determined by Bochkarev et al. [51] (pdb id 1jmc). Because of the ambiguity on the effective existence of the cysteine bond between the residues 200 and 289 for the free protein, we also considered a series of complexes without the S-S bond, **1′** and **2′**. In the separate trajectory calculations, the geometries chosen for the receptor are derived from complex, **3**, and gathered from the X-ray structure of the free RPA70AB, Bochkareva et al. [16] (pdb id 1fgu), **4**. Consistent with the geometries of the complex, the receptors without the cysteine bond are also examined, **3′** and **4′**, respectively.

The root mean square deviation, rmsd, of the Cα atoms in function of the sampling time (Fig. 2a and b), 24 ns, and the root mean square fluctuation, rmsf, of the backbone atoms, in function of the residue number (Fig. 2c and d), are shown. The binding loops, L12, and L45 (residues 212–220 and 264–274 for RPA70A, and residues 332–342 and 282–388 for RPA70B respectively), are directly involved in the binding process. For the four complexes, **1**, **2**, **1′** and **2′**, the core part of the protein shows a rather high rigidity with a value of rmsf below the 2 Å, underlying how the geometry change of the binding moieties does not deviate substantially from the X-ray structure. Even the ligands, $dC_8$ and $dC_5$ reveal a small tendency to change from the initial geometry. The presence of the cysteine bond seems to have a minor effect on the dynamics of the complexes. Additionally, even the binding loop regions show a strong stability, **1** and **1′**, Fig. 2c and d. Shortening the ligand to a $dC_5$, **2** and **2′**, does not alter sensibly the rmsf values in both the subunits, with the exception of the L45 binding regions, with a rmsf over 4 Å, and L12 close to 3.5 Å for **2′**. The rmsd in function of time is very similar for all complexes with a value close to 2 Å, Fig. 2a and b. The receptor modeled in absence of the ligand behaves very differently from the corresponding complex. If only the receptor from the structure of the complex is modeled, **3** and **3′**, the rmsd values are almost unaltered, with values between 2 and 3 Å for **3** and between 3 and 4 Å for the last 7 ns for **3′**. The strongest changes are observed in the region of the binding loops, where their orientation along the simulation was mainly dictated by the presence of the ligand. If for the simulation the geometry is taken from the X-ray structure of the free RPA70AB, **4** and **4′**, the flexibility of the protein is highly enhanced. The experimental structure already predicts a higher distance between the 2 subunit A and B [16] with respect to the corresponding structure derived from the complex. Under the assumption that the free protein is more flexible [16, 43], the simulations show higher values of rmsd, close to 7, 8 Å for **4** and close to 4 Å for **4′** (Fig. 2a–d).

The correlation between the mobility of the complex with the average NMR chemical shift measured for the first subunit RPA70A interacting with $dC_5$ is shown in Fig. 3. Experimentally [43], the average $^{1}H$ and $^{15}N$ NMR chemical shift is mildly a function of the nature of the base sequence of a pentamer, $dN_5$, binding RPA70A. The result suggests a low selectivity of the protein for ssDNA, hypothesis supported by the similar value of the binding constant $K_D$ of ~2 μM measured with different sequences of nucleotides. In Fig. 3 we reported the RPA70A interacting with $dC_5$. The top part of the figure, A, shows the averages rmsd over 24 ns for the residues from 183 to 291, for **1** and **2**, and RPA70A-$dC_5$, **2b**. The bottom part of the figure, B, from the work of Arunkumar et al. [43], shows the chemical shift between free RPA70A and the corresponding complex with $dC_5$. There is indeed a similarity between the graphs, in particular in proximity of the binding loops L12 and L45, the higher differences are highly noticeable. In the first part of the protein, from residues 183 to 205, there is not a significant change in the chemical shift, contrary to a modest change in the rmsf that peaks at 3.5 Å. Between residues 210 and 220 there is a remarkable matching between the NMR signal and geometric change. Comparably, the region between residues 264 to 280 has similar behavior. A good agreement is also encountered in the region around the residue 237 even if with a minor rmsf value.

### Correlated and anti-correlated motions between various structural elements

In order to investigate motion between different regions in the protein, or between ligand and receptor, we calculated the correlation matrix for the binding complexes, which are displayed as two dimension maps. A positive value (in red) shows that the atoms are moving in the same direction, whereas a negative value (in blue) indicates an anti-correlated motion.

The correlation values for the Cα atoms for the protein and P atoms for the nucleotide are reported in Fig. 4. Complex **1** shows two main areas of strong correlation and anti-correlation in correspondence to the RPA subunits, underlying how the binding motions of RPA70A and RPA70B are strongly anti-correlated. The linker, residues 300–315, has a substantial non-correlated pattern, green area. In this respect, we can still argue the presence of a domain-domain communication which leads to an anti-correlated motion of binding for the two subunits. A similar behavior is encountered if $dC_5$ is used as a ligand, **2**, even

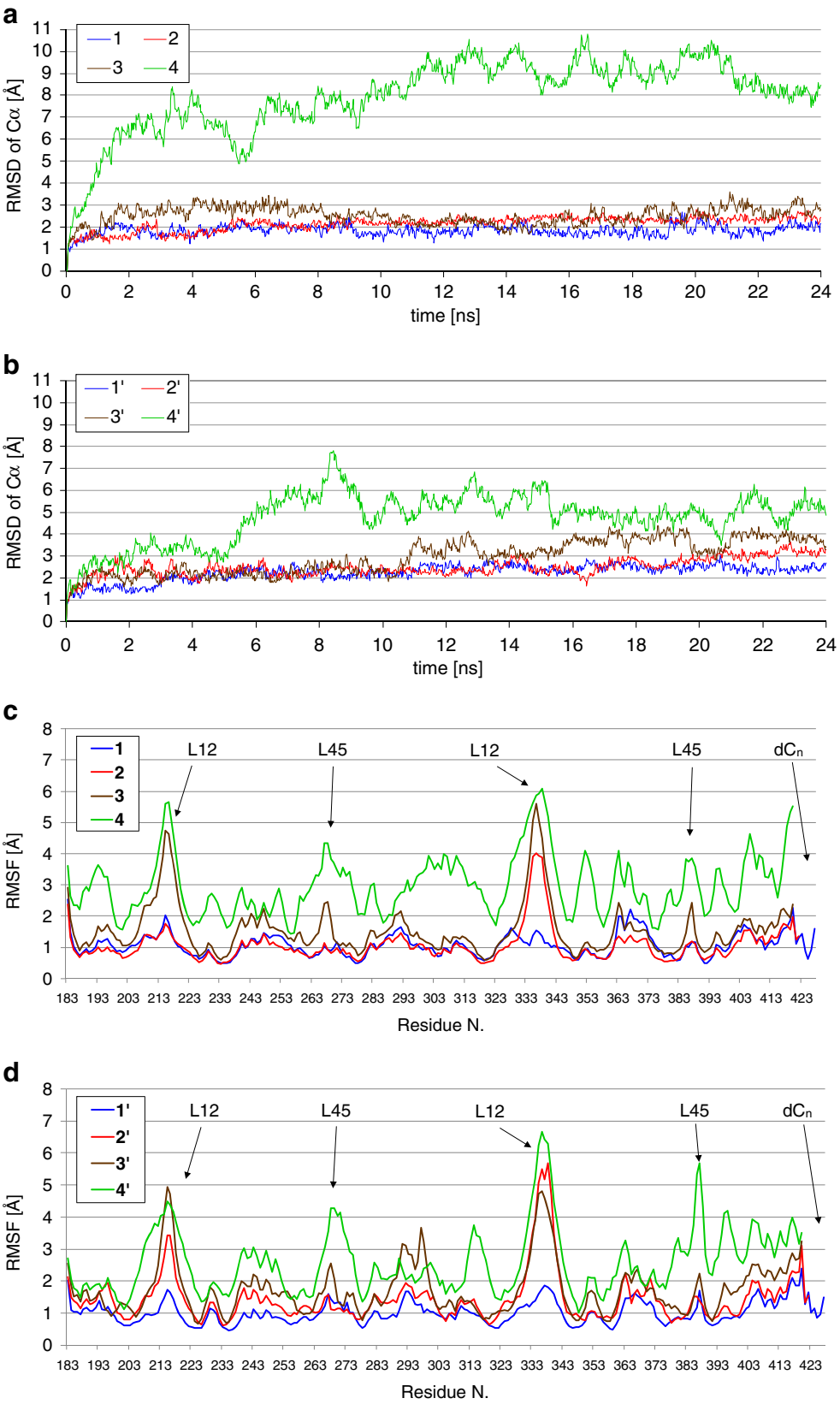


**Fig. 2** Root mean square deviation, RMSD, for Cα and P atoms as a function of the simulation time (24 ns) for the complexes and the receptors, (**a** and **b**). Cα and P root mean square fluctuation, RMSF, per residue (**c** and d) averaged over the simulation time. The notation $dC_n$ refers to $dC_8$ for **1** and **1′** and $dC_5$ for **2** and **2′**

though the anti-correlation region for the RPA70B has smaller values, more green areas, than **1**. The above observation would suggest that the anti-correlation motion is induced by the presence of the ligand, and it is not

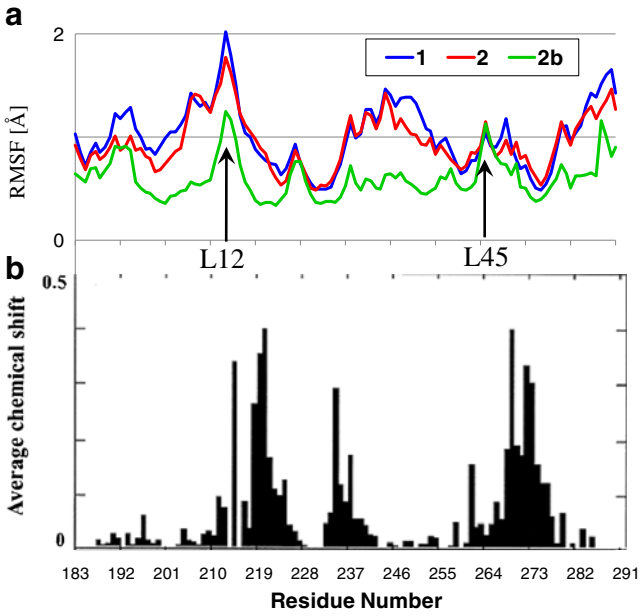


**Fig. 3** Comparison between the RMSF (**a**) per residue in the RPA70A region for the system **1** and **2**. The species 2b in RPA70A alone, was separately modeled for the same time scale. The bottom part shows the chemical shift measured by NMR (**b**). The regions relative to the binding loops, L12 and L45, are labeled

peculiar of the two subunits linked together. However, in the total absence of the ligand, **3**, the trajectory evolves toward a motion which resembles the one encountered in **1**. The difference is that the correlation and anti-correlation motion are enhanced and there is a higher mix of opposite motions in both domains. This might indicate that the dimer, in the geometry of the complex, still has some domain-domain communication, however, additional opposite movements limiting the effect is observed if the geometry analyzed is of free RPA, **4**. The correlation and anti-correlation areas in the maps are strongly enhanced, even if regions of weaker opposite movement are still present. The elimination of the sulfur bond between the two cysteines in position 200 and 289, right hand side of the picture, reveals an increased mobility of the complex. The general pattern for each case is preserved, but there is an increase in minor and intense local opposite movements, in particular in the case of the free protein, **4′**. This might indicate a disruption of the domain-domain communication, as clearly visible on the left hand side of Fig. 4.

## Principal component analysis

Due to the size of the systems considered and the potential flexibility of the binding units, we extended our analysis to a study based on the principal component analysis (PCA), often called essential dynamics [91] when applied to protein dynamics [91, 93–97]. With this approach, the high dimensional space spanned by 3 N-6 degrees of freedom (where N is the number of atoms) is substantially 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 cumulative 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 the complexes are shown in Fig. 5. The eigenvectors ev1 and ev2 relative to RPA70AB are shown in yellow, while the vectors for the $dC_8$ and $dC_5$ are represented in blue. The binding loops are depicted in cyan and green for L12 and L45, respectively. The main component of the motion of **1**, ev1, is dominated by a concerted motion of the two subunits toward $dC_8$. In particular, L12 has a strong component toward the binding regions. L45, on the contrary, shows only a minor binding character, with vectors pointing far from the $dC_8$ fragment.

The components of ev2 are lower in magnitude, but still they underline the motion of the RPA toward the $dC_8$. In **1′** the components of the two principal eigenvectors are mildly altered, however the binding motions of RPA towards the ligand is still present. Complexes **2** and **2′** reveal different dynamics of the subunit RPA70A, which appear to move far from the ligand. However, the subunit RPA70B, where the binding interaction is present, has the two main components of the first two eigenvectors which are very similar to **1**. Complex **2′**, on the other hand, behaves in an unexpected way, where only the L45 of RPA70B shows a movement towards the ligand while L12 has a motion in the opposite direction. The second principal component eigenvectors, for **2** and **2′**, describe a concerted motion toward the ligand, similar to ev1 found for **1** and **1′**.

## Energy landscape analysis

According to the energy landscape model [98], the folding and 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 [99–104]. 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 [105].

The two dimensional projections of the trajectories of the systems considered on the first and second principal component is shown in Fig. 6. This technique helps visualize the shape of the potential energy surface in spite of the inevitable distortion introduced by the projection.

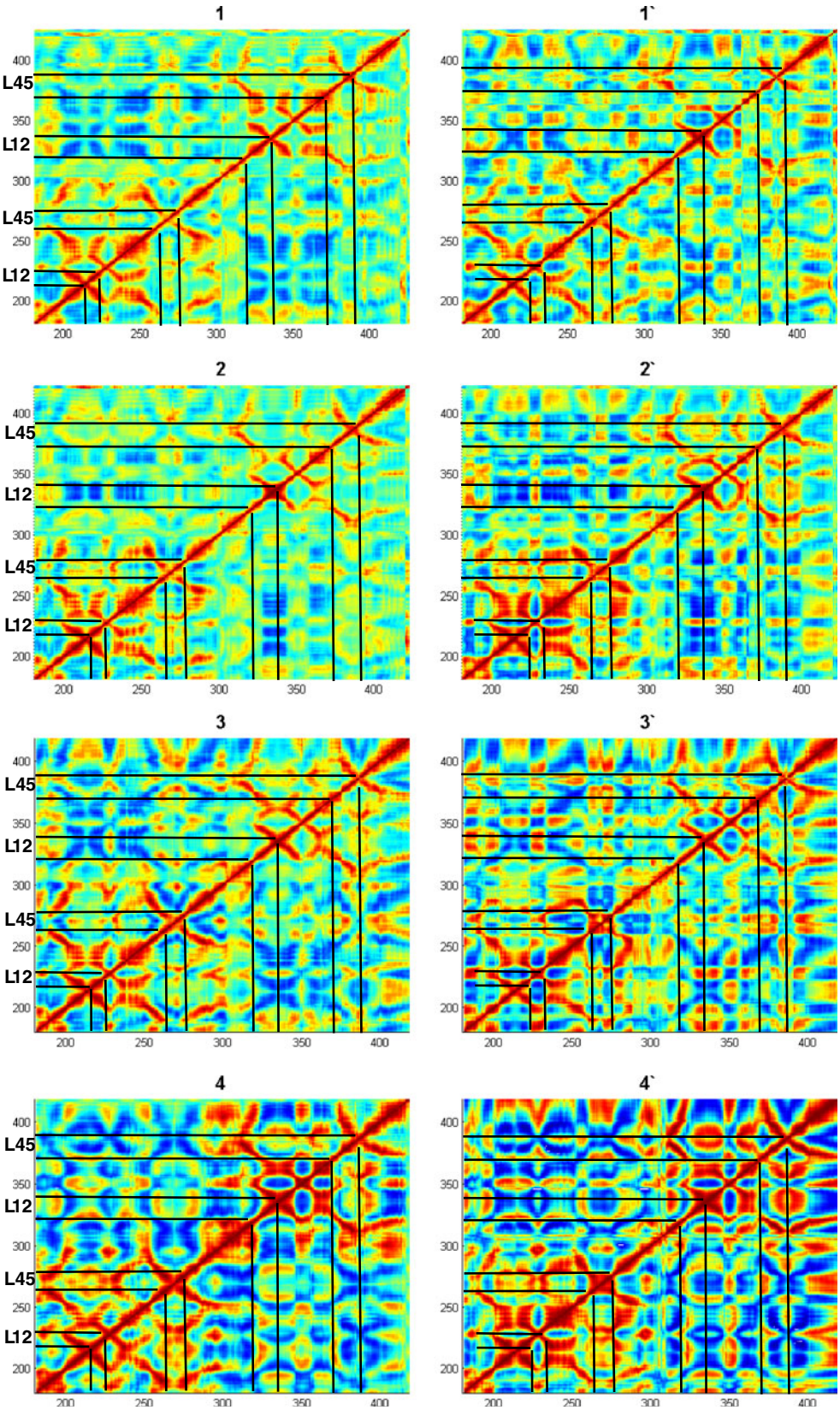


**Fig. 4** Cross correlation matrix relative to the Cα and P atoms for the complex and single receptors. Positive correlations in descending order are shown in red, orange and yellow. Anti-correlated motions are shown in cyan and blue. Light green shows regions with negligible. The binding loops regions L12 and L45 for both subunits are underlined by black lines

The model was performed using only the Cα atoms of the protein residues and P atoms for the ligand, thus the binding interaction is described 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. In addition, we plotted the

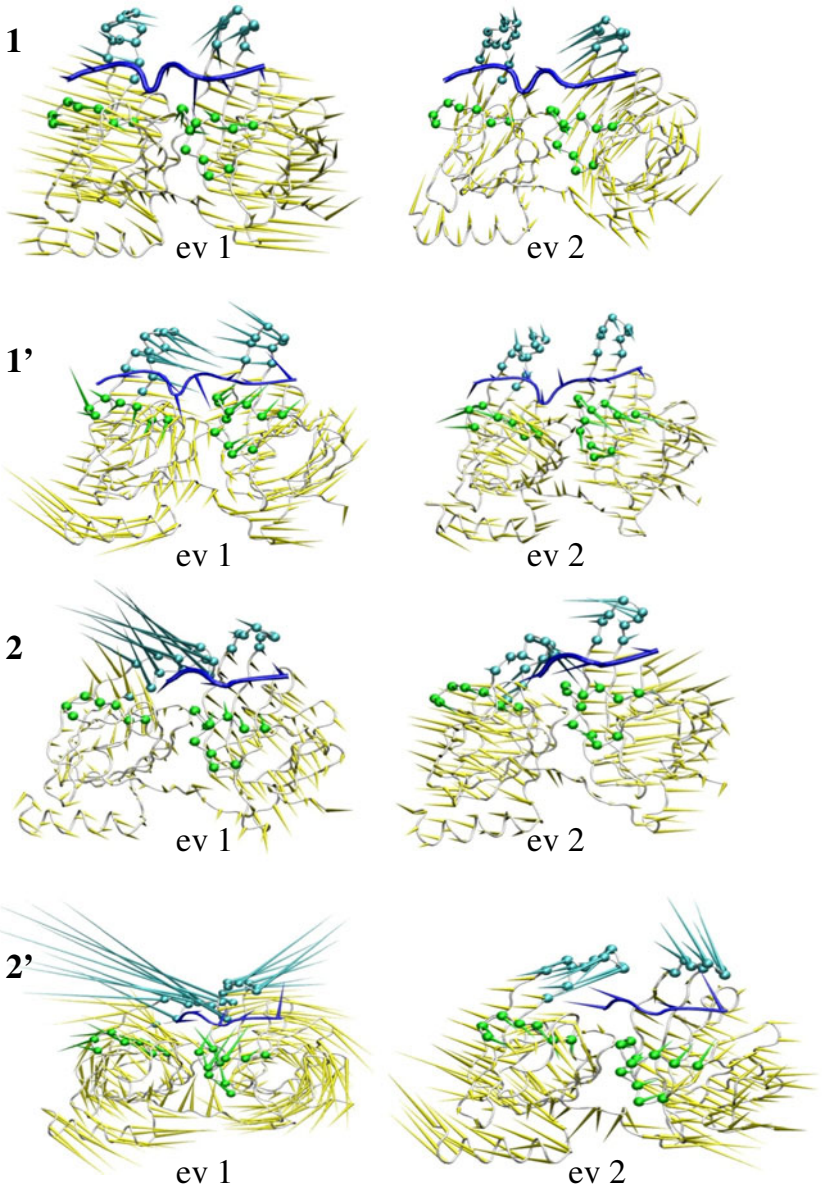


**Fig. 5** Representation of the first two principal eigenvectors from the PCA analysis on **1**, **1′**, **2** and **2′** mapped onto the Cα atoms for the receptors and P atoms for the ligands. The binding loops are represented in cyan and green dots for L12 and L45 respectively. The ligand is represented in blue

interpolation of the values of the projection of the trajectory onto the two main principal eigenvectors along the simulation (black line in Fig. 6) which crosses the funnel areas shown by the histogram.

The numbers have not been scaled in any way to reflect relative energies because our statistics from direct MD is insufficient to achieve quantitative accuracy. The results reveal that the effect of changing the ligand modifies the basins' structure. For **1** there are several minor basins separated by low barriers showing the exploration of the conformational space along the simulation. The first series of basins, encountered at the initial 6 ns of the simulation, is located in proximity of the point with ev1, ev2 coordinates (−20,0). The system evolves through a shallow valley for about 6 ns to reach the most dominant basin at coordinate (10,0) for all the remaining time of the simulation. A similar behavior is found for **1′**, where from the initial point (15,5) the complex reaches through a small valley in ~3 ns and an intermediate state with coordinates (0,10), to evolve toward the final state at (−10,0) where the most dominant basin is found. Both final basins are the deepest compared to the other complexes, revealing a strong binding tendency. RPA70AB binding $dC_5$, **2** and **2′**, have a different energy envelop profile with respect to the previous cases. Both show a complex structure of basins less prominent but more separated rather than connected by valleys as previously encountered. In addition, the basin reached at the end of the simulation lies at a higher minimum. The shape of the energy envelope appears correlated to the binding properties, where the higher the binding energy the

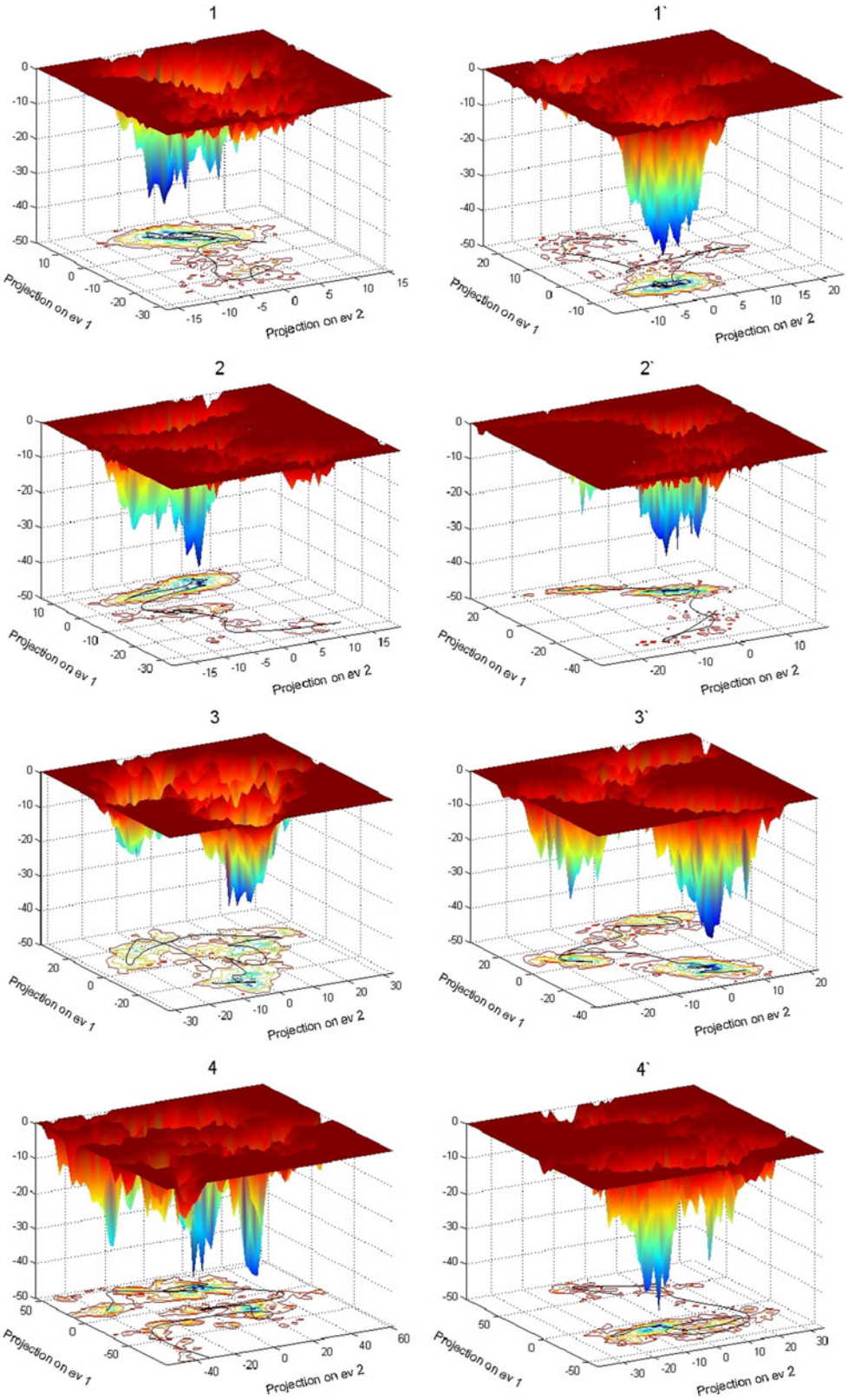


**Fig. 6** Histograms for the complexes and receptors representing the "minimum energy envelope" obtained from two-dimensional projections of the trajectories onto the first two principal eigenvectors. The black line represents the interpolation of the histogram values mapped on the domain surface

deeper and pronounced is the minimum envelop region. The complex structure in absence of ligand, **3** and **3′** lead to an even more pronounced and scattered series of basins, indicating that the structure changes frequently along the trajectory. There is indeed a final region encountered in the basin at coordinates (0,10) for **3** and (−20,0) for **3′** which underline a certain stability. If in the simulation the X-ray structure of the free protein is used, **4** and **4′**, the number of

small and scattered basins increase with respect to **3** and **3′**, revealing how the two subunits, RPA70A and RPA70B, have independent motions.

## Binding free energy calculations

Another aspect of importance relative to the binding properties is the evaluation of the absolute or standard binding free energy. In order to perform this computation, the last 8 ns of the MD trajectories, from 400 adjacent snapshots at 20 ps time interval, were extracted and used as input for the calculation of the gas phase and solvation free energy, $\Delta G_{gas+solv}$, by the MM-PB(GB)SA protocols. Since the correlation time for decay of the 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 large number of samples it is possible to estimate properly the value for the binding energy. The study conducted by Kollman and Case [106, 107], revealed how a number of snapshots from 100 to 200 is sufficient to calculate the binding energy with good accuracy. The values for the binding $\Delta G_{gas+solv}$, oscillate to some degree with respect to the average value but still in an expected range, Fig. 7.

The entropy of the solute, necessary to obtain the absolute binding free energy, has been calculated by normal mode analysis based within the harmonic approximation. The $\Delta G_{gas+solv}$ values calculated on a single trajectory by the MM-PBSA and MM-GBSA methods, (PB)$G_{tot}$ and (GB)$G_{tot}$ respectively, are reported in Table 1. The mean values of the energy decomposition, including the binding entropy, are also shown. The binding character is over estimated which is a common feature of the single trajectory MM-PB(GB)SA approach. Complex **1** has a (PB)$G_{tot}$ of −79.8 kcal mol$^{-1}$, with an unfavorable entropic contribution of −43.5 kcal mol$^{-1}$. The data are derived from a well equilibrated structure with a slope of the linear regression of 9.8 $10^{-3}$ kcal mol$^{-1}$ ps$^{-1}$, Fig. 7. The (GB)$G_{tot}$ value, not surprisingly, overestimate the binding by 3.5 kcal mol$^{-1}$ value averaged over a series of data with a slope of the liner regression of 9.4 $10^{-3}$ kcal mol$^{-1}$ ps$^{-1}$, comparable to previously published results [63]. The loss of the sulfur bond between Cys200 and Cys289, **1′**, reduces the binding energy by more than 15 kcal mol$^{-1}$ with a value of −64.6 kcal mol$^{-1}$ and −73.1 kcal mol$^{-1}$ for (PB)$G_{tot}$ and (GB)$G_{tot}$ respectively. The slope of the linear regression of the series of data used to determine the free energy indicates a slightly better equilibration, with value of the slope of the linear regression of 6.0 $10^{-4}$ kcal mol$^{-1}$ ps$^{-1}$, and 5.0 $10^{-4}$ kcal mol$^{-1}$ ps$^{-1}$, for PB and GB respectively, Fig. 7. The change in entropy, however, is rather constant, with a drop of 2.3 kcal mol$^{-1}$ with respect to **1**. Reducing the size of the ligand, $dC_5$, decreases by almost half the binding free energy. Complex **2** has a (PB)$G_{tot}$ value of −36.2 kcal mol$^{-1}$, 5.6 kcal mol$^{-1}$ higher than **2′**. The (GB)$G_{tot}$ estimations are higher, with a value of −37.5 kcal mol$^{-1}$ for **2** and −35.6 kcal mol$^{-1}$ for **2′**. In this case, the loss of the sulfur bond, consistent with the previous case, reduces the binding free energy, despite its location in the subunit B far from the ligand located in the subunit A. Complexes **2** and **2′** seem less equilibrated, where the slope of the linear regression line for (PB)$G_{tot}$ value for **2** and **2′** are 1.8 $10^{-2}$ kcal mol$^{-1}$ ps$^{-1}$ and 2.4 $10^{-2}$ kcal mol$^{-1}$ ps$^{-1}$ respectively, similar to the values for the (GB)$G_{tot}$ energies.

The contributions to the MM-PB(GB)SA energies, including the corresponding standard error of the mean, sem, are listed in Table 1. The two major contributions with a bonding character 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 contributions come from the polar solvation component, $G_{PB}$, while the non-polar part, $G_{np}$, gives a minor contribution. The sum of both quantities is labeled as $G_{solv}$. The polar contribution to the solvation free energy is calculated by solving the Poisson-Boltzmann (PB) equation [71] using radii from the PARSE parameter set [108], or with the generalized Born (GB) method [72] as described by Onufriev et al. [79], IGB2.

The presence of the sulfur bond between Cys200 and Cys289 in **1** changes the binding property quite substantially, Table 1. The value of $H_{elec}$ is reduced by 159 kcal mol$^{-1}$ while the $H_{vdw}$ has only a mild increase, ~4 kcal mol$^{-1}$. The non-polar contributions to the binding free energy, $G_{np}$, Table 1, has the same value for both complexes, -18.5 kcal mol$^{-1}$. The polar part of the solvation energy, calculated by PB and GB, $G_{PB}$ and $G_{GB}$, is very different for the two complexes, but both methods predict an increase of 167 kcal mol$^{-1}$ for **1′**. The change of the solute entropy is minor, 2.3 kcal mol$^{-1}$, thus the difference in binding free energy is mainly due to the Coulombic energy and to the polar component of the solvation energy. The decrease of $H_{ele}$ for **1′** is substantially high, 198 kcal mol$^{-1}$, whereas the change in $H_{vdw}$ is within a few kcal mol$^{-1}$. The non-polar contribution is the same for both complexes. The polar contribution, however, consistent to what was observed for **1** and **1′**, increased by 177 kcal mol$^{-1}$ for **1′**, independently of the method used. The change of solute entropy is less than one kcal mol$^{-1}$. The reduction in the size of the ligand from $dC_8$ to $dC_5$, **2** and **2′**, significantly decreases the binding free energy and the corresponding components, Table 1. The general trend is however maintained, with a reduction of $H_{elec}$ by 100 kcal mol$^{-1}$ from **2** to **2′**, a moderate reduction of $H_{vdW}$, and $G_{np}$ of 5 kcal mol$^{-1}$ and 1 kcal mol$^{-1}$ respectively. The increase of the solvation energy is 122 kcal mol$^{-1}$ for $G_{PB}$ and 108 kcal mol$^{-1}$ for $G_{GB}$.

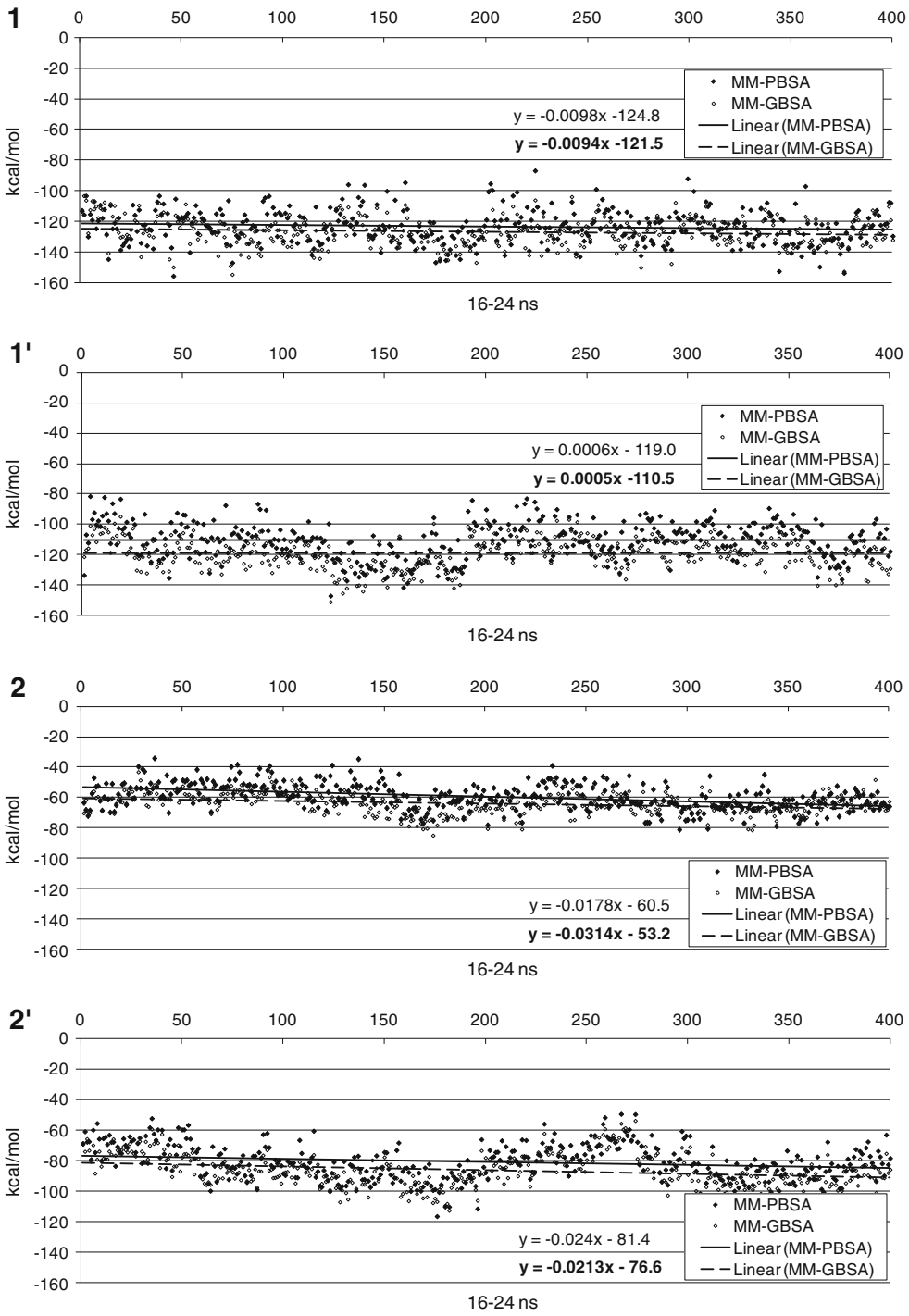


**Fig. 7** MM-PB(GB)SA binding free energies for **1**, **1'**, **2** and **2'**, calculated in the last 8 ns on the single trajectory. Each snapshot has 20 ps step

Single and separate trajectories approach

Single trajectory method gives results which are far from the experimental estimation of the binding free energy. This problem is common in the literature and it seems due to the strong change of conformation that the receptor has before the formation of the complex. In order to circumvent this limitation, we adopted the separate trajectory approach, where the ligand and the receptor are modeled independently, mimicking the expected structure of the interacting species before the formation of the complex. For this purpose we chose two different geometries for the receptor. The first structure was obtained by simply removing the ligand from the experimental geometry of the complex, **3** and **3′**, the second one was the structure determined without the ligand, the free RPA70AB, **4** and **4′**. Table 2 reports the energies calculated using **1** and **2** as a reference for the complex values, **3** and **4** for the receptor and $dC_5$ and $dC_8$ for the

**Table 1** MM-PB(GB)SA energy components of the complexes calculated for 800 equidistant snapshots extracted from the last 8 ns of the same trajectory

| Contrib | **1** | | **1′** | | **2** | | **2′** | |
|---|---|---|---|---|---|---|---|---|
| | Mean | sem[h] | Mean | sem | Mean | sem | Mean | sem |
| $H_{elec}$[a] | −885.0 | 6.6 | −1,044.7 | 5.7 | −494.1 | 8.3 | −594.4 | 7.8 |
| $H_{vdW}$[b] | −128.3 | 1.8 | −124.0 | 1.6 | −73.2 | 1.8 | −77.6 | 1.8 |
| $H_{gas}$[c] | −1,013.3 | 7.0 | −1,168.7 | 6.3 | −567.3 | 8.7 | −672.1 | 8.3 |
| $G_{np}$[d] | −18.5 | 0.1 | −18.5 | 0.1 | −11.3 | 0.1 | −12.5 | 0.1 |
| $G_{PB}$[e] | 908.5 | 6.0 | 1,076.7 | 5.3 | 510.8 | 7.6 | 622.7 | 7.3 |
| $G_{solv}$[f] | 890.0 | 5.9 | 1,058.3 | 5.3 | 499.5 | 7.6 | 610.2 | 7.3 |
| $G_{gas+solv}$[g] | −123.3 | 3.4 | −110.4 | 3.5 | −67.8 | 3.2 | −61.8 | 3.5 |
| $G_{np}$ | −18.5 | 0.1 | −18.5 | 0.1 | −11.3 | 0.1 | −12.5 | 0.1 |
| $G_{GB}$ | 905.0 | 6.0 | 1,068.2 | 5.3 | 509.5 | 7.6 | 617.7 | 7.4 |
| $G_{solv}$’ | 886.5 | 6.0 | 1,049.8 | 5.2 | 498.2 | 7.6 | 605.3 | 7.3 |
| $G_{gas+solv}$’ | −126.8 | 3.1 | −118.9 | 3.4 | −69.1 | 3.2 | −66.8 | 3.4 |
| $TS^{(s)}_{tot}$[i] | −43.5 | 3.6 | −45.8 | 3.7 | −31.6 | 3.2 | −31.2 | 3.2 |
| (PB)$G_{tot}$[l] | −79.8 | 4.9 | −64.6 | 5.1 | −36.2 | 4.5 | −30.6 | 4.7 |
| (GB)$G_{tot}$ | −83.3 | 4.8 | −73.1 | 5.0 | −37.5 | 4.5 | −35.6 | 4.6 |

The standard state is assumed to be at 1 M. The units are in kcal $mol^{-1}$. Ionic strength is set at 50 mM. [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; [e] $G_{PB}$: polar solvation free energy calculated by solving the Poisson-Boltzmann equation PB, in italic by solving the Generalized Born equation, GB; [f] $G_{solv}=G_{np}+G_{PB}$; [g] $G_{gas+solv}=H_{gas}+G_{solv}$; [h] Standard error of mean values; The MM-GBSA value are shown in italic; [i] $TS_{tot}^{(s)}$: total solute entropy contribution; [l] $G_{tot}=G_{gas+solv.}-TS_{tot}^{(s)}$

ligands, where each species are equilibrated individually. The introduction of the separate trajectory method increases substantially the accuracy on the prediction of the binding free energy. The overbounded value for **1** is reduced when

**Table 2** MM-PBSA energy components of complexes **1–3**, **1–4**, **2–3** and **2–4** calculated for 800 equidistant snapshots extracted from the last 8 ns of separate trajectories

| Contrib | **1–3** | | **1–4** | | **2–3** | | **2–4** | |
|---|---|---|---|---|---|---|---|---|
| | Mean | sem[i] | Mean | sem | Mean | sem | Mean | sem |
| $H_{elec}$[a] | –669.0 | 7.0 | −691.5 | 6.8 | −400.2 | 7.8 | −422.7 | 7.7 |
| $H_{vdW}$[b] | −108.3 | 1.8 | −166.2 | 1.8 | −87.4 | 1.7 | −145.3 | 1.7 |
| $H_{int}$[c] | 5.3 | 3.0 | 28.7 | 3.0 | 0.9 | 3.1 | 24.3 | 3.1 |
| $H_{gas}$[d] | −772.0 | 7.3 | −829.0 | 7.1 | −486.7 | 8.0 | −543.7 | 7.9 |
| $G_{np}$[e] | −17.2 | 0.1 | −25.8 | 0.1 | −14.9 | 0.1 | −23.5 | 0.1 |
| $G_{PB}$[f] | 693.2 | 6.3 | 749.3 | 6.3 | 441.7 | 7.1 | 497.9 | 7.1 |
| $G_{solv}$[g] | 676.0 | 6.3 | 723.5 | 6.3 | 426.9 | 7.0 | 474.4 | 7.0 |
| $G_{gas+solv}$[h] | −96.0 | 3.4 | −105.4 | 3.4 | −59.8 | 3.3 | −69.3 | 3.3 |
| $G_{np}$’ | −17.2 | 0.1 | −25.8 | 0.1 | −14.9 | 0.1 | −23.5 | 0.1 |
| $G_{PB}$’ | 683.4 | 6.3 | 753.3 | 6.3 | 424.8 | 7.0 | 494.6 | 7.0 |
| $G_{solv}$’ | 666.3 | 6.3 | 727.5 | 6.2 | 409.9 | 7.0 | 471.1 | 7.0 |
| $G_{gas+solv}$’ | −105.7 | 3.3 | −101.5 | 3.2 | −76.8 | 3.3 | −72.6 | 3.2 |
| $TS^{(s)}_{tot}$[l] | −28.3 | 3.8 | −49.0 | 3.7 | −17.0 | 3.2 | −37.7 | 3.7 |
| (PB)$G_{tot}$[m] | −62.4 | 7.4 | −27.7 | 7.3 | −42.0 | 4.6 | −7.3 | 4.8 |
| (GB)$G_{tot}$ | −72.2 | 5.0 | −23.8 | 4.9 | −58.9 | 4.7 | −10.5 | 4.9 |

The standard state is assumed to be at 1 M. The units are in kcal $mol^{-1}$. Ionic strength is set at 50 mM. [a] $H_{elec}$: coulombic energy; [b] $H_{vdW}$: van der Waals energy; [c] $H_{int}$: internal energy; [d] $H_{gas}=H_{elec}+H_{vdW}+H_{int}$; [e] $G_{np}$ non-polar solvation free energy; [f] $G_{PB}$: polar solvation free energy calculated by solving the Poisson-Boltzmann equation PB, in italic by solving the generalized Born equation, GB; [g] $G_{solv}=G_{np}+G_{PB}$; [h] $G_{gas+solv}=H_{gas}+G_{solv}$; [i] Standard error of mean values; The MM-GBSA value are shown in italic; [l] $TS_{tot}^{(s)}$: total solute entropy contribution; [m] $G_{tot}=G_{gas+solv.}-TS_{tot}^{(s)}$

the energy of the complex is compared with **3**, (PB)$G_{tot}$ decreases by 17.4 kcal mol$^{-1}$ and (GB)$G_{tot}$ decreases by 11.1 kcal mol$^{-1}$. A similar behavior is observed for **1′**, Table 3, (PB)$G_{tot}$ decreases by 1.0 kcal mol$^{-1}$ but (GB)$G_{tot}$ decreases by 13.9 kcal mol$^{-1}$. The energy change in the separate trajectory protocol is due to the conformational adaptation of the receptor in the complex [77]. The change in entropy is actually in favor of a stronger binding, with a change of $T\Delta S$ of 15 and 6.1 kcal mol$^{-1}$ in **1–3** and **1′–3′** respectively. When the geometry of the receptor is taken from the crystal structure of the free RPA70AB, the correction in the binding free energy is more pronounced, structures **1–4** and **1′–4′**. In the former case, the energy decreases by 52.1 kcal mol$^{-1}$, (PB)$G_{tot}$, and 59.5 kcal mol$^{-1}$, (GB)$G_{tot}$, for **1** and by 33.5 kcal mol$^{-1}$, (PB)$G_{tot}$, and 38.0 kcal mol$^{-1}$, (GB)$G_{tot}$, for **1′**. The combination **1–4** is the closest to the experimental value estimated to be −10±0.4 kcal mol$^{-1}$ with an overestimation of the binding free energy of ~17 kcal mol$^{-1}$ by (PB)$G_{tot}$ and ~14 kcal mol$^{-1}$ by (GB)$G_{tot}$, with a much better agreement than the single trajectory approach. The error is substantial but very much in line with the accuracy expected with this method. The big size of the ligand, including its high flexibility, might represent an issue in estimation of the adaptation free energy. For the smaller ligand, however, the description reaches a surprising level of accuracy. The estimate (PB)$G_{tot}$ energy for **2–4**,-7.3 kcal mol$^{-1}$, is in excellent agreement with the experimental value, -7± 1 kcal mol$^{-1}$, Table 2. Even the (GB)$G_{tot}$ energy is in very good agreement, with a value of −10.5 kcal mol$^{-1}$. The absence of the S-S bond makes the complex **1′-4′** overbounded, (PB)$G_{tot}$ value of −31.1 kcal mol$^{-1}$ and (GB)$G_{tot}$ value of −35.1 kcal mol$^{-1}$, and the complex **2′-4′** unstable, with a positive bonding free energy, Table 3. Due to the very good agreement of structures **1–4** and **2–4** with the experimental value, we used these complexes to compare different MM-GBSA methods to verify if a better agreement can be reached.

Contrary to the PB method, in the GB approach each atom is represented as a sphere with a radius called an effective Born radii. In order to increase the computational efficiency, several methods have been proposed to determine the Born radii. The modified CFA method introduced by Hawkins et al. [109], IGB1, was one of the first implemented in Amber. This method, however, gave some issues on the buried atoms. A rescaling approach was introduced to estimate more accurately the Born radii, with the GB$^{OBC}$ methods, namely IGB2, approach used in the data shown in Tables 1, 2 and 3. In addition we extended our calculations using two different radius definitions for IGB2, bondi and mbondi2, IGB2' and IGB2" respectively. Also the IGB1 protocol is considered as a comparison. The data are summarized in Table 4, indicating the mean values for the contributions of each MM-GBSA method, and the standard error of the mean, sem.

**Table 3** MM-PBSA energy components of complexes **1′-3′**, **1′-4′**, **2′-3′** and **2′-4′** calculated for 800 equidistant snapshots extracted from the last 8 ns of separate trajectories

| Contrib | **1′–3′** | | **1′–4′** | | **2′–3′** | | **2′–4′** | |
|---|---|---|---|---|---|---|---|---|
| | Mean | sem[i] | Mean | sem | Mean | sem | Mean | sem |
| $H_{elec}$[a] | −886.1 | 7.4 | −822.5 | 6.6 | −565.8 | 8.4 | −502.1 | 7.7 |
| $H_{vdW}$[b] | −135.1 | 1.7 | −154.4 | 1.7 | −85.6 | 1.8 | −104.9 | 1.8 |
| $H_{int}$[c] | −0.2 | 3.4 | −2.9 | 3.4 | 8.5 | 3.3 | 5.7 | 3.2 |
| $H_{gas}$[d] | −1,021.4 | 7.7 | −979.8 | 7.2 | −642.9 | 8.7 | −601.3 | 8.2 |
| $G_{np}$[e] | −42.2 | 0.1 | −47.6 | 0.1 | −13.2 | 0.1 | −18.7 | 0.1 |
| $G_{PB}$[f] | 960.5 | 6.7 | 943.3 | 6.0 | 600.1 | 7.7 | 583.0 | 7.1 |
| $G_{solv}$[g] | 918.3 | 6.6 | 895.7 | 5.9 | 587.0 | 7.6 | 564.3 | 7.0 |
| $G_{gas+solv}$[h] | −103.1 | 3.6 | −84.1 | 3.5 | −55.9 | 3.6 | −36.9 | 3.5 |
| $G_{np}$' | −18.7 | 0.1 | −24.1 | 0.1 | −13.2 | 0.1 | −18.7 | 0.1 |
| $G_{PB}$' | 941.4 | 6.7 | 915.7 | 6.0 | 595.4 | 7.7 | 569.7 | 7.1 |
| $G_{solv}$' | 922.8 | 6.6 | 891.6 | 5.9 | 582.2 | 7.6 | 551.0 | 7.0 |
| $G_{gas+solv}$' | −98.7 | 3.5 | −88.2 | 3.3 | −60.7 | 3.4 | −50.2 | 3.3 |
| $TS^{(s)}_{tot}$[l] | −39.7 | 4.0 | −56.0 | 3.0 | −27.7 | 4.0 | −44.0 | 3.0 |
| (PB)$G_{tot}$[m] | −63.6 | 5.4 | −31.1 | 4.6 | −19.8 | 5.4 | 12.7 | 4.6 |
| (GB)$G_{tot}$ | −59.2 | 5.3 | −35.1 | 4.5 | −24.6 | 5.3 | −0.5 | 4.5 |

The standard state is assumed to be at 1 M. The units are in kcal mol$^{-1}$. Ionic strength is set at 50 mM. [a] $H_{elec}$: coulombic energy; [b] $H_{vdW}$: van der Waals energy; [c] $H_{int}$: internal energy; [d] $H_{gas}=H_{elec}+H_{vdW}+H_{int}$; [e] $G_{np}$ non-polar solvation free energy; [f] $G_{PB}$: polar solvation free energy calculated by solving the Poisson-Boltzmann equation PB, in italic by solving the generalized Born equation, GB; [g] $G_{solv}=G_{np}+G_{PB}$; [h] $G_{gas+solv}=H_{gas}+G_{solv}$; [i] Standard error of mean values; The MM-GBSA value are shown in italic; [l] $TS_{tot}^{(s)}$: total solute entropy contribution; [m] $G_{tot}=G_{gas+solv}$. $-TS_{tot}^{(s)}$

**Table 4** MM-GBSA energy components calculated by IGB1, IGB2, IGB2' and IGB2" for complexes **1–4** and **2–4**

| **1-4** Contrib | IGB1 | | IGB2 | | IGB2' | | IGB2" | |
|---|---|---|---|---|---|---|---|---|
| | Mean | sem[d] | Mean | sem | Mean | sem | Mean | sem |
| $G_{GB}$[a] | 697.0 | 6.1 | 756.6 | 6.3 | 766.9 | 6.3 | 750.3 | 6.2 |
| $G_{solv}$[b] | 671.0 | 6.1 | 730.6 | 6.2 | 741.0 | 6.3 | 724.4 | 6.2 |
| $G_{gas+solv}$[c] | −161.1 | 3.2 | −101.5 | 3.2 | −91.1 | 3.1 | −107.8 | 3.1 |
| $TS^{(s)}_{tot}$[e] | −49.0 | 3.7 | −49.0 | 3.7 | −49.0 | 3.7 | −49.0 | 3.7 |
| $(GB)G_{tot}$[f] | −82.9 | 4.9 | −23.3 | 4.9 | −13.0 | 4.8 | −29.6 | 4.9 |
| **2-4** Contrib | IGB1 | | IGB2 | | IGB2' | | IGB2" | |
| | Mean | sem[h] | Mean | sem | Mean | sem | Mean | sem |
| $G_{GB}$ | 442.2 | 6.8 | 497.9 | 7.0 | 504.5 | 7.0 | 493.4 | 6.9 |
| $G_{solv}$ | 418.6 | 6.8 | 474.3 | 7.0 | 480.9 | 7.0 | 469.7 | 6.9 |
| $G_{gas+solv}$ | −128.2 | 3.2 | −72.5 | 3.2 | −65.9 | 3.1 | −77.1 | 3.2 |
| $TS^{(s)}_{tot}$ | −37.7 | 3.7 | −37.7 | 3.7 | −37.7 | 3.7 | −37.7 | 3.7 |
| $(GB)G_{tot}$ | −65.7 | 4.9 | −10.0 | 4.9 | −3.4 | 4.8 | −14.6 | 4.9 |

The standard state is assumed to be at 1 M. The units are in kcal $mol^{-1}$ . Ionic strength is set at 50 mM. The terms $H_{elec}$, $H_{vdW}$ $H_{int}$, $H_{gas}$, and $G_{np}$ are omitted since they are reported in Table 2. [a] $G_{GB}$: polar solvation free energy calculated by solving the generalized Born equation, GB; [b] $G_{solv}=G_{np}+G_{PB}$; [c] $G_{gas+solv}=H_{gas}+G_{solv}$; [d] Standard error of mean values; The MM-GBSA value are shown in italic; [e] $TS_{tot}^{(s)}$: total solute entropy contribution; [f] $G_{tot}=G_{gas+solv.}-TS_{tot}^{(s)}$

The results obtained with the IGB1 approximation are far from the experimental values and comparable with the single trajectory approach. The data from IGB2" protocol slightly overbinds both complexes by a few kcal $mol^{-1}$, however the IGB2' theory estimates the binding free energy with an error of only 4 kcal $mol^{-1}$ for both complexes. This result is very interesting since often the MM-PB(GB)SA results show a strong overbound character of the interacting structures [110]. The use of separate trajectories approach, where the individual geometries are experimentally determined, seems to be the key method to use in order to have optimal prediction of the binding free energy.

The MM-GBSA protocol allows to decompose the $\Delta G_{gas+sol}$ energy in terms of contributions from each residue of a structure considered. The data calculated with the IGB2 method, Fig. 8, shows the components for the highest 20 contributions to the $\Delta G_{gas+sol}$ energy. The Coulombic interaction summed with the solvation free energy (coul+GB), the non-polar contribution (np), and the van der Waals (vdW) energy for each residue considered are represented in Fig. 8. About 15 aminoacids contribute to the overall binding energy, with $\Delta G_{gas+sol}$ values from −12 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 382 and 210. The former is located in the L45 binding side of RPA70B subunit; the amino acid contributes by more than 12 kcal $mol^{-1}$ to $\Delta G_{gas+sol}$ in both **1** and **1′**. The latter position is in L12 in the subunit RPA70A, which is particularly dominant in **2** and **2′**, with a $\Delta G_{gas+sol}$ value close to −12 kcal $mol^{-1}$, while in **1** and **1′** the value is lower, 8 kcal $mol^{-1}$. In **1** and **1′** other Arg have a dominant role in the binding, in position 335, 216, 234, and 339. Other amino acids show a strong binding character, such as tryptophan in position 361, with a $\Delta G_{gas+sol}$ value close to −8 kcal $mol^{-1}$, Asparagine in position 214, $\Delta G_{gas+sol}$ value close to −5 kcal $mol^{-1}$, and unexpectedly the non-polar Phenylalanine 269 in L45 of RPA70A and in position 238, in L12 of RPA70B, which contribute by −6 kcal $mol^{-1}$ and −4 kcal $mol^{-1}$. Minor contribution, close to 2 kcal $mol^{-1}$, are given by Trp212, Phe386, Ile332, and Asn214. When the ligand used is $dC_5$, **2** and **2′**, a different set of amino acids interact with the ssDNA. The strongest contribution is due to ARG210 with $\Delta G_{gas+sol}$ values of −9 kcal $mol^{-1}$ and −12 kcal $mol^{-1}$ in **2** and **2′** respectively, position with relevant binding contribution already found for **1** and **1′**. Arg126 has a similar contribution for both complexes with a $\Delta G_{gas+sol}$ value of −8 kcal $mol^{-1}$. Lower binding properties are observed for Arg234, Asn214, Phe269, and Phe238 with $\Delta G_{gas+sol}$ values between 6 and 4 kcal $mol^{-1}$.

## Solvent effect

The energetics of the electrostatic interaction depends on the screening property of the surrounding environment, which is correlated to the ionic strength. In order to monitor this effect we performed a series of MM-PBSA calculation on the separate trajectory for complexes **1–4** and **2–4**, Table 5, for which we encountered the optimal agreement with experiment. The polar contributions to the solvation free energy were computed at 0, 50, 100, 200, 500 and 1000 mM. We observed a mild change in the binding free

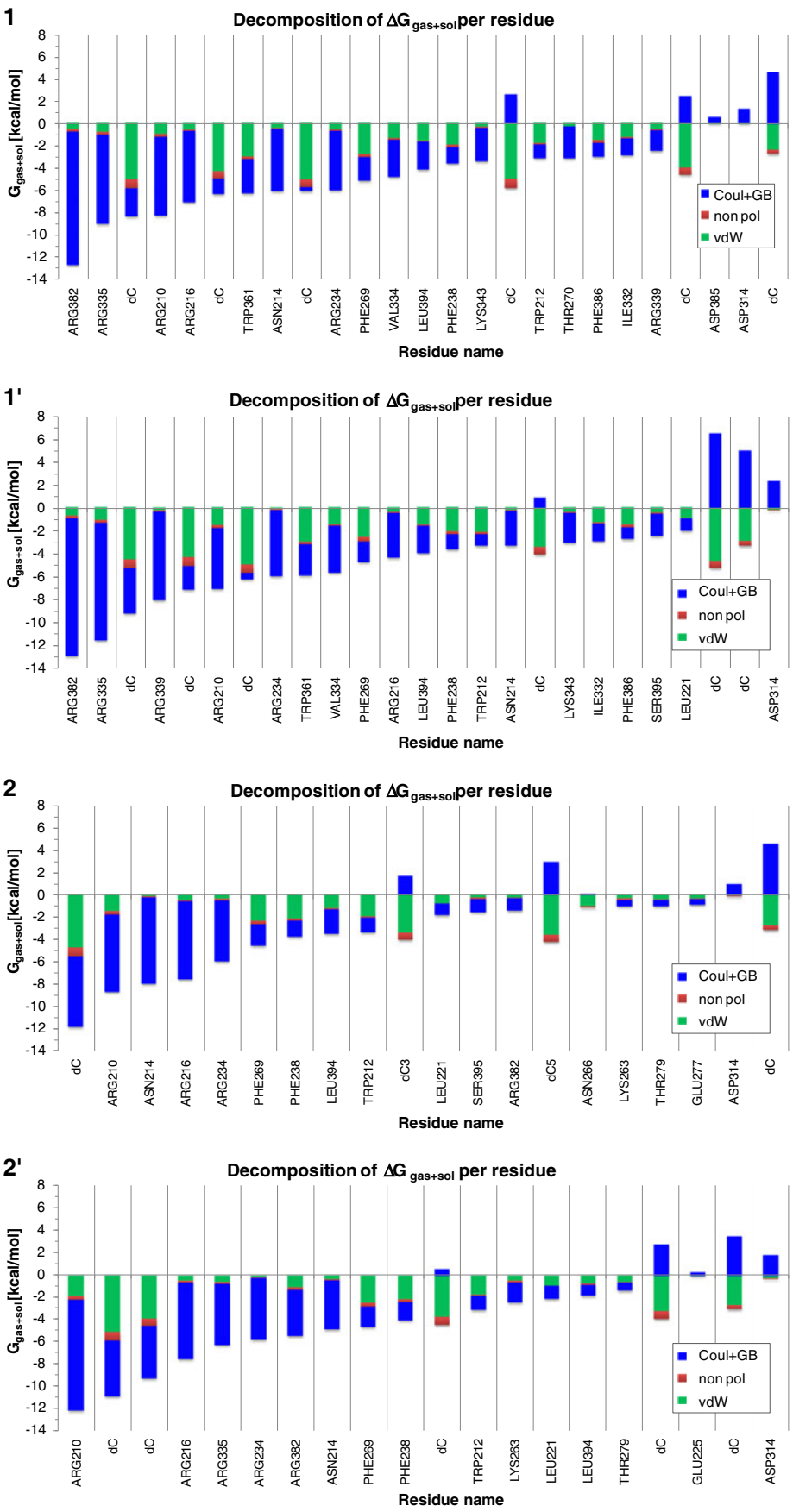


**Fig. 8** decomposition per residues of the binding free energies calculated using the IGB2. To estimate the MM-GBSA energies 400 equidistant snapshots were extracted in the last 8 ns of the trajectory. The values show the contributions of the Coulombic interaction plus the solvation free energy (coul + GB) the non polar (NP) and van der Waals (vdW). The number in parenthesis shows to which monomer the residue belongs

**Table 5** MM-GBSA energy components calculated by, IGB2, for complexes **1–4** and **2–4**, in the function of the change of the ionic strength

| **1–4**[a] | $G_{gas+solv}$[b] | | (GB)$G_{tot}$[c] | |
|---|---|---|---|---|
| Salt concentration | Mean | sem[d] | Mean | sem |
| 0 | −100.8 | 3.2 | −22.6 | 4.9 |
| 50 | −101.5 | 3.2 | −23.3 | 4.9 |
| 100 | −100.9 | 3.2 | −22.7 | 4.9 |
| 200 | −100.1 | 3.2 | −21.9 | 4.9 |
| 500 | −98.7 | 3.2 | −20.5 | 4.9 |
| 1000 | −97.5 | 3.2 | −19.3 | 4.9 |
| **2-4** | $G_{gas+solv}$[b] | | (GB)$G_{tot}$[c] | |
| Salt concentration[a] | Mean | sem[h] | Mean | sem |
| 0 | −65.5 | 3.2 | −3.5 | 4.9 |
| 50 | −65.9 | 2.2 | −3.9 | 4.9 |
| 100 | −65.6 | 3.2 | −3.6 | 4.9 |
| 200 | −65.2 | 3.2 | −3.2 | 4.9 |
| 500 | −64.5 | 3.2 | −2.5 | 4.9 |
| 1000 | −63.9 | 3.2 | −1.9 | 4.9 |

[(b)] Binding free energy, by IGB2, in function of different salt concentrations, in mM, for complexes **1–4** and **2–4**. The standard state is assumed to be at 1 M. The units are in kcal $mol^{-1}$. [(b)] $G_{gas+solv}=H_{elec}+H_{vdW}+H_{int}+G_{solv}$; [(c)] $G_{tot}=G_{gas+solv}-TS_{tot}^{(s)}$, where $TS_{tot}^{(s)}$: total solute entropy contribution; [(d)] Standard error of mean values

energy of about 4 kcal $mol^{-1}$ for **1–4** and 2 kcal $mol^{-1}$ for **2–4**, from the lowest salt concentration to the highest. The result reveals how the binding free energy for these highly charged systems is not very sensitive to the ionic strength of the solution.

## Discussion

In this study, we have demonstrated that the MM-PBSA method can accurately calculate the binding free energy between RPA70AB and a short oligomer. Several experimental results are available in the literature on the RPA affinity with a ssDNA sequence. In particular, the work of Arunkumar et al. [43] showed the binding activity for the single and combined A and B RPA70 domains in function of different nt sequences, including the oligo $dC_8$ used for the crystallization. The data revealed a low specificity of RPA70AB for ssDNA, and equivalent binding rate constant for $dC_8$ and $dC_{10}$, confirming that $dC_8$ is the minimum binding unit for RPA70AB. Schubert et al. [111] calculated the binding free energy by surface plasmon resonance (SPR) and fluorescence correlation spectroscopy (FCS), with the assumption that a single equilibrium reaction occurs in the temperature range of the measurements. These results can be used to validate theoretical models based on molecular dynamics aimed to describe the binding affinity of RPA70AB.

Due to the sensitivity of the binding parameters in terms of the nature of ssDNA, experimental conditions and sequence of RPA, we attempted to reproduce the experimental results relative to the only complex solved by X-ray crystallography, the RPA70AB interacting with the oligomer $dC_8$, pdb code 1jmc [51]. Measurements conducted by NMR titration and fluorescence spectroscopy [43], confirmed a $K_D$ value for RPA70AB binding to $dC_8$ equal to 52±16 nM, and 2–10 μM if $dC_5$ is used as a ligand instead. The data measured for a different sequence of oligonucleotides, but with the same length, are very similar, proving the low RPA70AB selectivity for ligands with a length between 5 and 8 nt. The measurements are consistent with a previous study [112] where the value of $K_D$ for $dC_8$ was estimated to be 15 nM by electrophoretic mobility assay.

The binding free energy is related to the equilibrium association and dissociation constants, $K_A$ and $K_D$, by

$$\Delta G = -RTInK_A = RTInK_D, \quad (4)$$

where R is the gas constant and T the temperature in K. This allows us to estimate the experimental binding free energy values of −10.0±0.4 kcal $mol^{-1}$ for $dC_8$ and −7± 1 kcal $mol^{-1}$ for $dC_5$. The use of MM-PB(GB)SA methods, combined with the normal mode analysis for the solute entropy calculation, allows to estimate the absolute, or standard, binding free energy of the complexes and to compare the results with experiments.

The first issue was to evaluate whether the S-S bond between Cys200 and Cys289 was actually present in the complex, as there was seen a puzzling long S-S distance, more than 4 Å, measured in the X-ray structure of the free RPA70AB. The same issue was not noticeable in the geometry of the complex for which the corresponding S-S distance is in the expected length range. The distinction gave rise to two sets of structures for which we performed the study. The major question was to find the most accurate method to predict the binding free energy. The problem is particularly challenging, in fact, even though the fundamental principles governing ligand binding are reasonably well understood, an accurate method to calculate the absolute, or standard, binding free energy for complexes with a ligand of large size is still unavailable. Much progress has been achieved by Roux et al. [113] who proposed protocols based on the alchemical free energy perturbation (FEP) method where the ligand is progressively "decoupled" from the surroundings, and based on the potential mean force (PMF) where the binding constant is computed via a protein-ligand interaction. If the protocols are applied correctly, the error estimation is of the order of kcal $mol^{-1}$. Another very promising technique is the M2 method proposed by Gilson et al. [114], which is based on the traditional MM-PBSA approach, but includes an accurate

strategy to search for stable conformations that allow the code to reproduce binding energies. These methods are designed for relatively small ligands, which is a common scenario in drug design. The size of $dC_8$ and $dC_5$ prevents us from using the above mentioned techniques. As an alternative, we estimated the absolute binding free energy by the MM-PB(GB)SA protocol. This approach, regardless of the use of single or separate trajectories, gives an approximate value of the standard binding free energy [115]. More recently, the MM-PBSA method has been developed [72, 116] and applied to HIV reverse transcriptase [117], avidin [118], neuraminidase [119], cathespin D [120], growth factor receptor binding protein 2 [121], metallo-protease [122, 123] hystone deacetylase [124], DNA glycosilase [78], and protein-protein interface [125]. Despite the high flexibility of the system investigated, the MM-PB(GB)SA is considered a suitable method [126–131].

For very flexible molecules, the NM analysis is applicable in very short time scales, in the fs region, where the dynamics is dominated by bond vibrations. For longer periods of time, ns, conformational motions take place and should be included in the estimation of the entropy. In the latter case, the QH method is an appropriate theory to use to estimate the solute conformational entropy which is still to date the most difficult component to estimate in the binding free energy. Several studies reported little success in estimating the conformational entropy by QH analysis, and all cases encountered strong convergence problems [81, 132]. Interestingly Grünberg et al. [82] proposed different approaches based on the protocol used by Gohlke and Case [81], in which they solved the issue of poor convergence due to accumulating inaccuracies in calculating the covariance matrix. Their code, however, does not support DNA structures. More recently, efforts [133–136] introduced interesting developments, however the missing direct comparison with experiment leaves the problem unresolved. For ligands of moderate size the conformational entropy can be satisfactorily described by NM analysis, as observed for a series of oligomers [137–141], thus this theory can still be suitable to estimate the solute binding entropy. Until a systematic study which underlines the limitations of the NM analysis for ssDNA fragments by comparison with experiment is made, the approach used in this work can still be considered an appropriate attempt to tackle the issue.

The MM-PB(GB)SA results, calculated from the separate trajectory method, are surprisingly close to the experiment, in spite of the use of the NM analysis to estimate the binding entropy. As a peculiar aspect, the choice of the initial structure of the receptors seems to be critical. When the receptor geometry is the same as the one found in the complex, the improvement of the binding energy is modest, however, when the geometry used comes from the free RPA70AB, the agreement is optimal. For **1–4** we have optimal results, ~13 kcal $mol^{-1}$ overbinding, but for **2–4** the prediction matches the experiment, Tables 2 and 4. The reduced size of the ligand, $dC_5$ vs. $dC_8$, helps to improve the accuracy. The other combinations of receptor geometries and ligand length reduce the accuracy, as shown by the limit case **2′−4′** where the interaction results are repulsive. As a conclusion we predict that the S-S bond between Cys200 and Cys289 is indeed present in the structure and the overestimation in the experimental length can be explained as an error in refining the density map derived from the X-ray measurement.

Because of the encouraging prediction of the binding free energy, we tested several approaches to calculate the solvation free energy, Table 4. The IGB1 method gave poor results, with an estimation of the binding free energy comparable to the single trajectory approach. IGB2, previously used, is listed in the table as a comparison. IGB2 with bondi radii definition, IGB2′, gives an excellent agreement with the experiment. The binding free energy for **1–4** are over bound by only 3 kcal $mol^{-1}$, whereas for complexes **2–4** the binding free energy is underbound by 4 kcal $mol^{-1}$. The good prediction of the binding free energy shows how the MM-PB(GB)SA method can be extremely accurate if the appropriate structures are used in the modeling. It is true, however, that it is rather rare to find the X-ray structure of a complex and the corresponding geometries of the individual fragments, but ideally, this seems to be the most suitable approach to optimize the agreement with experiment.

The MM-GBSA technique allows to decompose the $G_{gas+solv}$ term calculated for each residue, Fig. 8. This approach can contribute to establish a correlation between change of affinity by point mutation and the corresponding $G_{gas+solv}$ value calculated for the same residue. For comparison we considered system **1** which is the closest to the experimental work based on RPA70 mutations. Studies on the change of the binding constant for mutations of RPA70 were conducted by Wold et al. [18, 45, 142]. The individual hydrophilic point mutations R234A and R382A, for instance, reduced the relative binding ratio, $K_A$(mutant)/$K_A$(wild type) to 0.027 and 0.040, respectively. The data on the top of Fig. 8 shows how the two residues contributed strongly to the binding free energy, with a $G_{gas+sol}$ value of −7 kcal $mol^{-1}$ and −12 kcal $mol^{-1}$, respectively. The loss in affinity experimentally determined can be associated to a reduction in binding free energy, caused by the loss of polar contacts, as underlined by the strong polar contribution that the $G_{gas+solv}$ has for the two residues. The results underline the role of the strength of residue interaction with the ssDNA as a component to estimate the binding. It has to be noted that, however, mutations

may lead to conformational changes with respect to the wild type complex, thus the analysis of the energy contribution by residue has only a qualitative prediction power. Although this approach prevents an accurate prediction of the activity of mutants, it can suggest new point mutations on the basis of energy contributions. The simple analysis of the crystal structure might be insufficient to suggest new mutations or understand the change of affinity of the current ones. For instance, the mutations K263A and E277A, located close to the binding sites, show a smaller change in the relative binding ratio, values of 0.18 and 0.10 respectively, inconsistent with what the X-ray structure may suggest. The experimental findings, however, are in perfect agreement with a lower $\Delta G_{gas+sol}$ contribution ~2 kcal mol$^{-1}$ for both mutations, (data not shown because below the threshold chosen for the representation). The polarity of the residue is not the only component to alter the affinity. The double point mutations F238A-F269A and W361A-F386A, for instance, reduce the relative binding by 0.23 and 0.15 respectively. This noticeable reduction is not as strong as the previous one because of weaker interactions involved in the binding, as shown by the corresponding $G_{gas+solv}$ component shown in Fig. 8. Due to the non polarity of the residues, the energy contribution has mostly a van der Waals character rather than a Coulomb+GB component.

The theoretical results confirm that different types of interactions might have a role in the affinity prediction for mutants. A model that allows for an estimate of the energy contribution by residues to the binding is an important tool to predict the design of new mutations with desired affinity properties. Results might be obtained with less accuracy if only the geometry of the complex is considered.

The root mean-square deviation (rmsd) and fluctuation (rmsf) relative to Cα and P atoms over the course of the simulation can be used as a measure of the conformational stability of the structures examined. The expected conformational flexibility of the free RPA, **4** and **4′**, is observed in the MD simulation, green line in Fig. 2 (a–c). The values of rmsd for **4** and **4′**, Fig. 2 (a–b), increase significantly after the first 10 ns, with values from 7 Å to 10 Å, as a possible consequence for the relaxation from the initial structure. However, the trajectory results stabilized along the remaining simulation time. The corresponding ligand free structures derived from the geometry of complex, **3** and **3′**, show a much higher rigidity, with a rmsd values below 3 Å and 4 Å respectively. The lack of flexibility observed for **3** and **3′**, compared with **4** and **4′**, suggests that intermolecular forces between the subunits RPA70A and RPA70B are the cause of a stabilization of the RPA dimer. The geometries of complexes **1**, **1′** and **2**, **2′** show a low fluctuation with rmsd values below 3 Å. The result is intriguing because the conformational stability of the complex seems to be independent of the length of the ligands, in particular the vicinity of the subunits 70A and 70B seems to be sufficient to preserve the dimer conformation in the bound state. The rmsf results depict which residue has the highest mobility averaged along the whole simulation time. All the domains in the geometries of complex, **1** and **1′**, show a high stability, with rmsf values below 2 Å. When the ligand is reduced, **2** and **2′**, its contact with RPA70AB is limited to the subunit RPA70A. In this condition, the binding loops L12 and in RPA70B show a high mobility, while the loop L45 preserves a strong rigidity. When the free RPA70AB in the complex geometry, **3** and **3′**, is considered, only the L12 in the domains 70A and 70B shows a high mobility. We can conclude that the L12 is the sub domain with the highest flexibility, and even if L45 has a strong binding character, its orientation is not affected by the presence of the ligand once it has reached the binding conformation. When the geometry of the free RPA70 is considered, few changes in most of the domains are encountered, however, consistent with the previous observation, L12 is still the most flexible residue. We can speculate that the high flexibility of L12 might help to recognize and start the binding process.

## Conclusions

The goal of this study is to shed light on the modality of the interaction of the RPA70AB protein with $dC_8$ and $dC_5$ for which the binding free energy has been experimentally determined. We also addressed the ambiguity in assigning an S-S bond between Cys200 and Cys289 where results appear overstretched in the X-ray structure. To address this issue a series of calculations aimed to estimate the absolute binding free energy were performed using the MM-PB(GB)SA protocol adopting the single and multi trajectory approach. The technique is based on 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 reveal how the use of the separate trajectories, in particular when the ligand free structure of RPA is considered for the receptor geometry, enhances the performance of the simulations. In addition, if the geometry of the receptor is taken for the X-ray structure of the free RPA70AB, the agreement with experiment is remarkably improved. In particular, if the geometry containing the S-S bond between Cys200 and Cys289 is considered, the error with the experiment is reduced to 4 kcal mol$^{-1}$ for both of the complexes. The decomposition of the MM-GBSA energy for different residues allows us to correlate the change of the affinity of the mutated protein with the $\Delta G_{gas+sol}$ contribution of the residue considered. The agreement with experiment is optimal and a strong change in the binding free energy can be considered as a

dominant factor in the loss of the binding affinity encountered in the mutants. This work also underlines the accuracy of the MM-PB(GB)SA method if suitable geometries are taken in the model, even though we are aware that often the geometry of the complex and the corresponding structure of the free receptor and ligand may not be available.

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

## References

1. Binz SK, Sheehan AM, Wold MS (2004) Replication protein A phosphorylation and the cellular response to DNA damage. DNA Repair (Amst) 3:1015–1024
2. Bochkarev A, Bochkareva E (2004) From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. Curr Opin Struct Biol 14:36–42
3. Iftode C, Daniely Y, Borowiec JA (1999) Replication protein A (RPA): the eukaryotic SSB. Crit Rev Biochem Mol Biol 34:141–180
4. Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 74:283–315
5. Machida YJ, Hamlin JL, Dutta A (2005) Right place, right time, and only once: replication initiation in metazoans. Cell 123:13–24
6. Shechter D, Costanzo V, Gautier J (2004) ATR and ATM regulate the timing of DNA replication origin firing. Nat Cell Biol 6:648–655
7. Stauffer ME, Chazin WJ (2004) Structural mechanisms of DNA replication, repair, and recombination. J Biol Chem 279:30915–30918
8. Wold MS (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem 66:61–92
9. Zou L, Elledge SJ (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300:1542–1548
10. Kowalczykowski SC (2000) Some assembly required. Nat Struct Biol 7:1087–1089
11. Yuzhakov A, Kelman Z, Hurwitz J, O'Donnell M (1999) Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. EMBO J 18:6189–6199
12. Fanning E, Klimovich V, Nager AR (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. Nucleic Acids Res 34:4126–4137
13. Mer G, Bochkarev A, Chazin WJ, Edwards AM (2000) Three-dimensional structure and function of replication protein A. Cold Spring Harb Symp Quant Biol 65:193–200
14. Murzin AG (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J 12:861–867
15. Theobald DL, Mitton-Fry RM, Wuttke DS (2003) Nucleic acid recognition by OB-fold proteins. Annu Rev Biophys Biomol Struct 32:115–133
16. Bochkareva E, Belegu V, Korolev S, Bochkarev A (2001) Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding. EMBO J 20:612–618
17. Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A (2002) Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. EMBO J 21:1855–1863
18. Bastin-Shanower SA, Brill SJ (2001) Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding. J Biol Chem 276:36446–36453
19. Daughdrill GW, Ackerman J, Isern NG, Botuyan MV, Arrowsmith C, Wold MS, Lowry DF (2001) The weak interdomain coupling observed in the 70 kDa subunit of human replication protein A is unaffected by ssDNA binding. Nucleic Acids Res 29:3270–3276
20. Binz SK, Lao Y, Lowry DF, Wold MS (2003) The phosphorylation domain of the 32-kDa subunit of replication protein A (RPA) modulates RPA-DNA interactions. Evidence for an intersubunit interaction. J Biol Chem 278:35584–35591
21. Bochkareva E, Kaustov L, Ayed A, Yi GS, Lu Y, Pineda-Lucena A, Liao JC, Okorokov AL, Milner J, Arrowsmith CH, Bochkarev A (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc Natl Acad Sci USA 102:15412–15417
22. Blackwell LJ, Borowiec JA, Mastrangelo IA (1996) Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase. Mol Cell Biol 16:4798–4807
23. Loo YM, Melendy T (2004) Recruitment of replication protein A by the papillomavirus E1 protein and modulation by single-stranded DNA. J Virol 78:1605–1615
24. Park CJ, Lee JH, Choi BS (2005) Solution structure of the DNA-binding domain of RPA from Saccharomyces cerevisiae and its interaction with single-stranded DNA and SV40 T antigen. Nucleic Acids Res 33:4172–4181
25. Daughdrill GW, Buchko GW, Botuyan MV, Arrowsmith C, Wold MS, Kennedy MA, Lowry DF (2003) Chemical shift changes provide evidence for overlapping single-stranded DNA- and XPA-binding sites on the 70 kDa subunit of human replication protein A. Nucleic Acids Res 31:4176–4183
26. Stauffer ME, Chazin WJ (2004) Physical interaction between replication protein A and Rad51 promotes exchange on single-stranded DNA. J Biol Chem 279:25638–25645
27. Ellison V, Stillman B (2003) Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5′ recessed DNA. PLoS Biol 1:E33
28. Zou L, Liu D, Elledge SJ (2003) Replication protein A-mediated recruitment and activation of Rad17 complexes. Proc Natl Acad Sci U S A 100:13827–13832
29. Wu X, Shell SM, Zou Y (2005) Interaction and colocalization of Rad9/Rad1/Hus1 checkpoint complex with replication protein A in human cells. Oncogene 24:4728–4735
30. Ball HL, Myers JS, Cortez D (2005) ATRIP binding to replication protein A-single-stranded DNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation. Mol Biol Cell 16:2372–2381
31. Namiki Y, Zou L (2006) ATRIP associates with replication protein A-coated ssDNA through multiple interactions. Proc Natl Acad Sci U S A 103:580–585
32. Yoo E, Kim BU, Lee SY, Cho CH, Chung JH, Lee CH (2005) 53BP1 is associated with replication protein A and is required for RPA2 hyperphosphorylation following DNA damage. Oncogene 24:5423–5430
33. Wong JM, Ionescu D, Ingles CJ (2003) Interaction between BRCA2 and replication protein A is compromised by a cancer-predisposing mutation in BRCA2. Oncogene 22:28–33
34. Robison JG, Elliott J, Dixon K, Oakley GG (2004) Replication protein A and the Mre11.Rad50.Nbs1 complex co-localize and interact at sites of stalled replication forks. J Biol Chem 279:34802–34810

35. Daniely Y, Borowiec JA (2000) Formation of a complex between nucleolin and replication protein A after cell stress prevents initiation of DNA replication. J Cell Biol 149:799–810
36. Kim K, Dimitrova DD, Carta KM, Saxena A, Daras M, Borowiec JA (2005) Novel checkpoint response to genotoxic stress mediated by nucleolin-replication protein a complex formation. Mol Cell Biol 25:2463–2474
37. Kim C, Paulus BF, Wold MS (1994) Interactions of human replication protein A with oligonucleotides. Biochemistry 33:14197–14206
38. Iftode C, Borowiec JA (2000) 5′ –>3′ molecular polarity of human replication protein A (hRPA) binding to pseudo-origin DNA substrates. Biochemistry 39:11970–11981
39. de Laat WL, Appeldoorn E, Sugasawa K, Weterings E, Jaspers NG, Hoeijmakers JH (1998) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. Genes Dev 12:2598–2609
40. Liu Y, Yang Z, Utzat CD, Liu Y, Geacintov NE, Basu AK, Zou Y (2005) Interactions of human replication protein A with single-stranded DNA adducts. Biochem J 385:519–526
41. Kim C, Wold MS (1995) Recombinant human replication protein A binds to polynucleotides with low cooperativity. Biochemistry 34:2058–2064
42. Kim C, Snyder RO, Wold MS (1992) Binding properties of replication protein A from human and yeast cells. Mol Cell Biol 12:3050–3059
43. Arunkumar AI, Stauffer ME, Bochkareva E, Bochkarev A, Chazin WJ (2003) Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains. J Biol Chem 278:41077–41082
44. Gomes XV, Henricksen LA, Wold MS (1996) Proteolytic mapping of human replication protein A: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA. Biochemistry 35:5586–5595
45. Wyka IM, Dhar K, Binz SK, Wold MS (2003) Replication protein A interactions with DNA: differential binding of the core domains and analysis of the DNA interaction surface. Biochemistry 42:12909–12918
46. Pretto DI, Tsutakawa S, Brosey CA, Castillo A, Chagot ME, Smith JA, Tainer JA, Chazin WJ (2010) Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering. Biochemistry 49:2880–2889
47. Cai L, Roginskaya M, Qu Y, Yang Z, Xu Y, Zou Y (2007) Structural characterization of human RPA sequential binding to single-stranded DNA using ssDNA as a molecular ruler. Biochemistry 46:8226–8233
48. Blackwell LJ, Borowiec JA (1994) Human replication protein A binds single-stranded DNA in two distinct complexes. Mol Cell Biol 14:3993–4001
49. Lavrik OI, Kolpashchikov DM, Weisshart K, Nasheuer HP, Khodyreva SN, Favre A (1999) RPA subunit arrangement near the 3′-end of the primer is modulated by the length of the template strand and cooperative protein interactions. Nucleic Acids Res 27:4235–4240
50. Treuner K, Ramsperger U, Knippers R (1996) Replication protein A induces the unwinding of long double-stranded DNA regions. J Mol Biol 259:104–112
51. Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature 385:176–181
52. Bhattacharya S, Botuyan MV, Hsu F, Shan X, Arunkumar AI, Arrowsmith CH, Edwards AM, Chazin WJ (2002) Characterization of binding-induced changes in dynamics suggests a model for sequence-nonspecific binding of ssDNA by replication protein A. Protein Sci 11:2316–2325
53. Lao Y, Lee CG, Wold MS (1999) Replication protein A interactions with DNA. 2. Characterization of double-stranded DNA-binding/helix-destabilization activities and the role of the zinc-finger domain in DNA interactions. Biochemistry 38:3974–3984
54. Burns JL, Guzder SN, Sung P, Prakash S, Prakash L (1996) An affinity of human replication protein A for ultraviolet-damaged DNA. J Biol Chem 271:11607–11610
55. Patrick SM, Turchi JJ (1998) Human replication protein A preferentially binds cisplatin-damaged duplex DNA in vitro. Biochemistry 37:8808–8815
56. Patrick SM, Turchi JJ (1999) Replication protein A (RPA) binding to duplex cisplatin-damaged DNA is mediated through the generation of single-stranded DNA. J Biol Chem 274:14972–14978
57. Balsera MA, Wriggers W, Oono Y, Shulten K (1996) Principal component analysis and long time protein dynamics. J Phys Chem 100:2567–2572
58. Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 18:431–492
59. Straatsma TP, McCammon JA (1992) Computational alchemy. Annu Rev Phys Chem 43:407–435
60. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417
61. Simonson T, Archontis G, Karplus M (2002) Free energy simulations come of age: protein-ligand recognition. Acc Chem Res 35:430–437
62. 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
63. 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
64. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8(52–56):29
65. Case DA, Darden TA, Cheatham TE, 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. 10. University of California, San Francisco
66. 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
67. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, K ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
68. 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
69. Pastor RW, Brooks BR, Szabo A (1988) Mol Phys 65:1409–1419
70. Miyamoto S, Kollman PA (1992) J Comput Chem 13:952–962
71. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268:1144–1149
72. 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 (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
73. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J (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
74. Jorgensen WL (1989) Free energy calculations: a breakthrough for modeling organic chemistry in solution. Acc Chem Res 22:184–189
75. 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
76. Roux B, Nina M, Pomès R, Smith JC (1996) Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. 71:670–681
77. 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
78. 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
79. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized born modelsuitable for macromolecules. J Phys Chem B 104:3712–3720
80. Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 98:1978–1988
81. 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
82. Grunberg R, Nilges M, Leckner J (2006) Flexibility and conformational entropy in protein-protein binding. Structure 14:683–693
83. 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
84. Rod TH, Radkiewicz JL, Brooks CL (2003) Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proc Natl Acad Sci U S A 100:6980–6985
85. Radkiewicz JL, Charles CL (2000) Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase. J Am Chem Soc 122:225–231
86. Prompers JJ, Bruschweiler R (2002) Dynamic and structural analysis of isotropically distributed molecular ensembles. Proteins 46:177–189
87. 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
88. Hayward S, Kitao A, Hirata F, Go N (1993) Effect of solvent on collective motions in globular protein. J Mol Biol 234:1207–1217
89. 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
90. 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
91. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. Proteins 17:412–425
92. 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
93. 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
94. Abagyan R, Argos P (1992) Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 225:519–532
95. Barrett CP, Noble ME (2005) Molecular motions of human cyclin-dependent kinase 2. J Biol Chem 280:13993–14005
96. Troyer JM, Cohen FE (1995) Protein conformational landscapes: energy minimization and clustering of a long molecular dynamics trajectory. Proteins 23:97–110
97. Caves LS, Evanseck JD, Karplus M (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 7:649–666
98. Becker MO (1997) Quantitative visualization of a molecular potential energy "funnel". J Mol Struct (THEOCHEM) 398–399:507–516
99. Wolynes PG (2005) Recent successes of the energy landscape theory of protein folding and function. Q Rev Biophys 38:405–410
100. 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 U S A 103:11844–11849
101. 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
102. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. Proc Natl Acad Sci U S A 101:511–516
103. Papoian GA, Wolynes PG (2003) The physics and bioinformatics of binding and folding-an energy landscape perspective. Biopolymers 68:333–349
104. Miller DW, Dill KA (1997) Ligand binding to proteins: the binding landscape model. Protein Sci 6:2166–2179
105. Wang J, Verkhivker GM (2003) Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. Phys Rev Lett 90:188181
106. 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
107. 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
108. Sitkoff D, Sharp KA, Honig B (1994) J Phys Chem 98:1978
109. 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
110. Woo HJ, Roux B (2005) Calculation of absolute protein-ligand binding free energy from computer simulations. Proc Natl Acad Sci U S A 102:6825–6830
111. Schubert F, Zettl H, Hafner W, Krauss G, Krausch G (2003) Comparative thermodynamic analysis of DNA–protein interactions using surface plasmon resonance and fluorescence correlation spectroscopy. Biochemistry 42:10288–10294
112. Pfuetzner RA, Bochkarev A, Frappier L, Edwards AM (1997) Replication protein A. Characterization and crystallization of the DNA binding domain. J Biol Chem 272:430–434
113. Roux B, Deng Y (2009) J Phys Chem 113:2234–2246
114. Gilson MK, Chang CE (2004) J Am Chem Soc 126:13156–13164

115. Wang J, Deng Y, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. Biophys J 91:2798–2814
116. Massova I, Kollman PA (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov 18:113–135
117. 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
118. 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
119. 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
120. 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
121. 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
122. Donini OA, Kollman PA (2000) Calculation and prediction of binding free energies for the matrix metalloproteinases. J Med Chem 43:4180–4188
123. 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
124. 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
125. 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
126. Li L, Uversky VN, Dunker AK, Meroueh SO (2007) A computational investigation of allostery in the catabolite activator protein. J Am Chem Soc 129:15668–15676
127. Xiong B, Burk DL, Shen J, Luo X, Liu H, Shen J, Berghuis AM (2008) The type IA topoisomerase catalytic cycle: a normal mode analysis and molecular dynamics simulation. Proteins 71:1984–1994
128. Lee J, Kim JS, Seok C (2010)Cooperativity and specificity of Cys2His2 zinc finger protein-DNA interactions: a molecular dynamics simulation study. J Phys Chem B 114:7662–7671
129. Hao GF, Yang GF, Zhan CG (2010)Computational mutation scanning and drug resistance mechanisms of HIV-1 protease inhibitors. J Phys Chem B 114:9663–9676
130. Zeng J, Li W, Zhao Y, Liu G, Tang Y, Jiang H (2008) Insights into ligand selectivity in estrogen receptor isoforms: molecular dynamics simulations and binding free energy calculations. J Phys Chem B 112:2719–2726
131. Bucher D, Grant BJ, Markwick PR, McCammon JA (2011) Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. PLoS Comput Biol 7: e1002034
132. Hsu ST, Peter C, van Gunsteren WF, Bonvin AM (2005) Entropy calculation of HIV-1 Env gp120, its receptor CD4, and their complex: an analysis of configurational entropy changes upon complexation. Biophys J 88:15–24
133. Nguyen PH (2009) Estimating configurational entropy of complex molecules: a novel variable transformation approach. Chem Phys Lett 468:90–93
134. Hensen U, Lange OF, Grubmuller H (2010) Estimating absolute configurational entropies of macromolecules: the minimally coupled subspace approach. PLoS One 5:e9179
135. Baron R, Hunenberger PH, McCammon JA (2009) Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: correction terms and convergence properties. J Chem Theory Comput 5:3150–3160
136. Wereszczynski J, Andricioaei I (2010) Conformational and solvent entropy contributions to the thermal response of nucleic acid-based nanothermometers. J Phys Chem B 114:2076–2082
137. Tidor B, Irikura KK, Brooks BR, Karplus M (1983) Dynamics of DNA oligomers. J Biomol Struct Dyn 1:231–252
138. Irikura KK, Tidor B, Brooks BR, Karplus M (1985) Transition from B to Z DNA: contribution of internal fluctuations to the configurational entropy difference. Science 229:571–572
139. Garcia AE, Soumpasis DM (1989) Harmonic vibrations and thermodynamic stability of a DNA oligomer in monovalent salt solution. Proc Natl Acad Sci U S A 86:3160–3164
140. Kottalam J, Case DA (1990) Langevin modes of macromolecules: applications to crambin and DNA hexamers. Biopolymers 29:1409–1421
141. Ha Duong T, Zakrzewska K (1997) Influence of drug binding on DNA flexibility: a normal mode analysis. J Biomol Struct Dyn 14:691–701
142. Lao Y, Gomes XV, Ren Y, Taylor JS, Wold MS (2000) Replication protein A interactions with DNA. III. molecular basis of recognition of damaged DNA. Biochemistry 39:850–859