2013

# A new structural framework for integrating replication protein A into DNA processing machinery

Chris A. Brosey
*Washington University School of Medicine in St. Louis*

Chunli Yan
*Georgia State University*

Susan E. Tsutakawa
*Lawrence Berkeley National Laboratory*

William T. Heller
*Oak Ridge National Laboratory*

Robert P. Rambo
*Lawrence Berkeley National Laboratory*

***See next page for additional authors***

*Published online 8 January 2013*

*Nucleic Acids Research, 2013, Vol. 41, No. 4* **2313–2327**
*doi:10.1093/nar/gks1332*

# A new structural framework for integrating replication protein A into DNA processing machinery

**Chris A. Brosey[1], Chunli Yan[2], Susan E. Tsutakawa[3], William T. Heller[4,5], Robert P. Rambo[3], John A. Tainer[3,6,7], Ivaylo Ivanov[2,*] and Walter J. Chazin[1,8,*]**

[1]Department of Biochemistry, Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA, [2]Department of Chemistry, Georgia State University, Atlanta, GA 30302-4098, USA, [3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [4]Center for Structural Molecular Biology, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA, [5]Biology & Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA, [6]Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA, [7]The Skaggs Institute for Chemical Biology, La Jolla, CA 92037, USA and [8]Department of Chemistry, Vanderbilt University, Nashville, TN 37232, USA

Received September 5, 2012; Revised November 26, 2012; Accepted November 27, 2012

## ABSTRACT

**By coupling the protection and organization of single-stranded DNA (ssDNA) with recruitment and alignment of DNA processing factors, replication protein A (RPA) lies at the heart of dynamic multi-protein DNA processing machinery. Nevertheless, how RPA coordinates biochemical functions of its eight domains remains unknown. We examined the structural biochemistry of RPA's DNA-binding activity, combining small-angle X-ray and neutron scattering with all-atom molecular dynamics simulations to investigate the architecture of RPA's DNA-binding core. The scattering data reveal compaction promoted by DNA binding; DNA-free RPA exists in an ensemble of states with inter-domain mobility and becomes progressively more condensed and less dynamic on binding ssDNA. Our results contrast with previous models proposing RPA initially binds ssDNA in a condensed state and becomes more extended as it fully engages the substrate. Moreover, the consensus view that RPA engages ssDNA in initial, intermediate and final stages conflicts with our data revealing that RPA undergoes two (not three) transitions as it binds ssDNA with no evidence for a discrete intermediate state. These results form a framework for understanding how RPA integrates the ssDNA substrate into DNA processing machinery, provides substrate access to its binding partners and promotes the progression and selection of DNA processing pathways.**

## INTRODUCTION

Replication protein A (RPA) is a modular multi-domain protein that functions in a wide range of DNA processing pathways required to maintain and propagate the genome of all living organisms. RPA functions by interfacing with dynamic multi-protein machinery and acts as a central hub that links many DNA transactions. RPA provides the primary single-stranded DNA (ssDNA) binding activity in eukaryotes and also serves as a scaffold and coordinator of DNA processing machinery (1,2). Binding of ssDNA is critical for shielding DNA strands from endonuclease activity and preventing the formation of disruptive secondary structures. RPA couples this activity to the recruitment of DNA processing factors, thereby providing a platform for organization of DNA processing machinery and managing access to the DNA substrate. Conjectures have been made about how protein interactions couple to the DNA binding activity of RPA (2), but uncertainty remains about how this might occur, and there is no structural framework to work from for intact RPA.

Despite its central importance in DNA processing machinery, little information is available on the physical basis for the coordination of RPA functions. This is in large part because its modular nature poses a significant

*To whom correspondence should be addressed. Tel: 1+ 615 936 2210; Fax: 1+ 615 936 2211; Email: walter.chazin@vanderbilt.edu
Correspondence may also be addressed to Ivaylo Ivanov. Tel: 1+ 404 413 5529; Fax: 1+ 404 413 5505; Email: iivanov@gsu.edu
Present address:
Chris A. Brosey, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA.

challenge for current structural methods. RPA is a heterotrimer of RPA70, RPA32 and RPA14 subunits organized into five structural modules connected by flexible linkers (70N, 70AB, 70C/32D/14, 32N, 32C; see Figure 1). The trimer core has one domain from each subunit, from which extend the remaining modules. Nuclear magnetic resonance (NMR) experiments on intact RPA have shown that the other modules are structurally independent of the core (3). The dynamic independence of the structural modules makes techniques, such as X-ray crystallography challenging to apply to full-length RPA and unlikely to fully capture the functionally relevant ensemble in solution.

The binding of ssDNA by RPA has been studied for >20 years, and it is generally held that RPA has three discrete DNA-binding modes (2,4,5). Four domains (70A, 70B, 70C and 32D) are known to engage ssDNA with progressively higher affinity and 5′ 3′ polarity, respectively. An initial 8 10 nucleotide (nt)-binding mode involves the tandem domains 70A and 70B, which are connected by a short 10 residue flexible linker. A poorly characterized intermediate binding mode has been suggested, encompassing an excluded site size of 12 23 nt. In addition to 70AB, this mode is presumed to engage 70C. The final binding mode engages a second domain from the trimer core, 32D, and occludes up to 30 nt of ssDNA (6,7).

As DNA processing proceeds, RPA must navigate between its different DNA-binding states. The differences in the number of domains directly contacting the DNA in the three ssDNA binding modes are expected to result in significant differences in the spatial organization of the DNA-binding apparatus of RPA. X-ray diffraction, NMR and scattering studies of isolated 70AB have provided insight into the initial ssDNA-binding mode (8 11). Binding of 8 10 nt of ssDNA aligns and compacts the domains, although the complex remains dynamic, presumably as a consequence of torsional motion between the two domains (11). The extent to which isolated 70AB typifies the action of intact RPA during DNA binding is not known. Here, we combine small angle X-ray and neutron scattering (SAXS and SANS) experiments with all-atom molecular dynamics (MD) simulations to determine the effects of ssDNA binding on the architecture of RPA. These results provide new perspectives on how coupling of protein and DNA-binding activities drive changes in the architecture and progression of DNA processing machinery.

## MATERIALS AND METHODS

### Materials

The pET15b vector containing human RPA DNA-binding core (RPA-DBC, $RPA70_{181\ 616}/32_{43\ 171}$/14, RPA70ABC/32D/14) was a kind gift of A. Bochkarev. Thrombin cleavable, 6×-His fusion tags precede the N-termini of the 70ABC and 14 subunits. Active thrombin was purchased from CALBIOCHEM. All ssDNA substrates-$dCCAC_7$, $dCCAC_{17}$, $dCCAC_{21}$, $dCCAC_{24}$ and $dCCAC_{27}$ (d10, d20, d24, d27 and d30, respectively), as well as fluorescently modified $dC_{10}$, $dC_{20}$ and $dC_{30}$ oligonucleotides were purchased from Integrated DNA Technologies (IDT) with standard desalting purification and were resuspended in sterile distilled water before use (sequences were validated by electrospray ionization mass spectrometry (ESI-MS) analysis performed by IDT). Full-length RPA used in these studies was prepared as described (3) and provided as a kind gift by Dr S. Michael Shell.

### Expression and purification of recombinant RPA-DBC

Expression and purification of the RPA-DBC have been described previously by the Bochkarev laboratory during its initial biochemical description of the construct (12,13). Modifications to this protocol, along with characterization of monodispersity by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) are provided in Supplementary Materials.

### Fluorescence anisotropy ssDNA-binding assays

The ssDNA-binding activity of RPA-DBC was assessed by a rise in fluorescence anisotropy as increasing amounts of protein were added to polycytidine substrates labelled at their 5′-ends with 6-carboxyfluorescein (5′-FAM-$dC_{10}$, -$dC_{20}$, -$dC_{30}$). Triplicate serial dilutions of protein (0 0.5 μM) were prepared in 384-well plates with 20 mM HEPES KOH (pH 7.5), 200 mM NaCl and 10 mM β-mercaptoethanol, then mixed with fluorescently labelled ssDNA (final concentration 25 nM). Polarized fluorescent intensities were measured with a Spectramax M5 plate reader (Molecular Machines) at excitation and emission wavelengths of 492 and 520 nm, respectively, for 100 s (1 reading/s) and averaged. Dissociation constants ($K_d$) were calculated by fitting the data to a simple two-state binding model in KaleidaGraph (v. 3.51).

### Preparation of RPA-DBC/ssDNA complexes for SAXS

Purified RPA-DBC was concentrated to 7 12 mg/ml, combined with 1.5 2-fold molar excess ssDNA substrate (d10, d20, d24, d27 or d30) and incubated on ice for 6 18 h (d10, d20) or at room temperature for 30 min (d24, d27, d30). To remove excess ssDNA and ensure homogeneous complexes, samples (300 μl) were injected onto a Superdex 200 HR 10/30 gel filtration column (GE Healthcare) equilibrated overnight in 20 mM HEPES KOH (pH 7.5), 200 mM NaCl, 5 mM dithiothreotol (DTT) and 2% glycerol. Samples for RPA-DBC, RPA-DBC/d10 and RPA-DBC/d20 eluted as single peaks. Samples for RPA-DBC/d24, RPA-DBC/d27 and RPA-DBC/d30 eluted with high-molecular weight shoulders, but conservative fractionation (290 μl fractions) of each elution profile allowed each 2 3 ml peak width to be completely isolated from multiply DNA-bound species or free ssDNA. The presence of both protein and DNA in each fraction was confirmed by ultraviolet absorbance readings. Calculations based on the $K_d$ values determined for ssDNA binding to RPA-DBC and the concentrations of protein and DNA used for sample preparation indicated that the amount of free protein was <3% for the 10mer complex and ≤1% for the longer ssDNA substrates. Experimental extinction coefficients calculated

from each gel filtration fraction used for SAXS confirmed 1:1 stoichiometry for the purified complexes.

### Small-angle X-ray scattering data collection and analysis

SAXS data were collected at the SIBYLS beamline 12.3.1 at the Advanced Light Source, Lawrence Berkeley National Laboratory. Scattering measurements were performed on 20 μl samples at 15°C using a Hamilton robot for loading samples from a 96-well plate into a helium-purged sample chamber (14,15). Data were collected on the original gel filtration fractions from each SEC run, as well as concentration series from fractions sampled from the later eluting half of each SEC elution peak (~2 8-fold concentration). Fractions before the SEC void volume were used for buffer subtraction of the original 1× gel filtration fractions, and concentrator eluates were used for buffer subtraction of each concentration series (2 8×).

SAXS experiments were performed using an X-ray beam from a multilayer monochromator of 12 keV ( 1.3 Å) covering the following momentum transfer range: $0.011\ \text{Å}^{-1} < q < 0.322\ \text{Å}^{-1}$, where $q$ is defined as $q = 4\pi \sin(\theta/2)/$ with scattering angle θ and wavelength . The multilayer monochromator provides increased X-ray flux, allowing stronger signals for lower protein concentrations. Sequential exposures (0.5, 0.5, 2, 5 or 6 and 0.5 s) were taken, and data were monitored for radiation-dependent aggregation. All SAXS data were collected using the MarCCD 165 detector in fast frame transfer mode and reduced via normalization to the incident beam intensity. Standard procedures were used for processing the data and are described in detail in the Supplementary Materials. SAXS data from this publication have been submitted to the BIOSIS database (http://bioisis.net).

SANS experiments were performed using the Bio-SANS instrument at the High Flux Isotope Reactor of Oak Ridge National Laboratory (16), as described in the Supplementary Materials.

### Interpreting RPA-DBC P(r) distributions

A model of free RPA-DBC was constructed from crystal structures of 70AB (PDB ID: 1FGU) and the trimer core (PDB ID:). Residues for the B C linker were added in PyMOL (17) and joined using the Modeller (9v4) interface in Chimera (18). Multiple inter-domain arrangements were generated from these starting models using BILBO-MD (19), from which models were selected that matched the experimental $D_{max}$ value of free RPA-DBC. Inter-domain distances were assessed in PyMOL and compared with features in the experimental P(r) distribution. Assignment of P(r) features to specific inter-domain distances were confirmed by simulating P(r) distances from the models (using the FoXS server and GNOM) and examining the impact of removing a given domain from the model (Supplementary Figure S3). A similar process was used to investigate P(r) distributions of the 10-nt and 20-nt complexes. The 10-nt model was generated in a manner similar to the free protein, with substitution of DNA-bound coordinates for 70AB (1JMC), while the 20-nt model was taken from the trajectory of the compact 20-nt MD simulation.

### All-atom molecular dynamics simulations

Crystal structures for 70AB, 70AB/dC8 and RPA70C/32D/14 were obtained from the RCSB Protein Data Bank [PDB ID: 1FGU (9), 1JMC (8) and 1L1O (13)]. Five models were constructed: (i) DNA-free RPA-DBC; (ii) RPA-DBC/dC10; (iii) RPA-DBC/dC20 compact; (iv) RPA-DBC/dC20 extended; and (v) RPA-DBC/dC30 (two independent simulations performed). The length of the ssDNA was adjusted to correspond to the initial (dC10) and final (dC30) modes of binding ssDNA, and one of the putative intermediate states (dC20). The DNA-free RPA-DBC model was built by connecting 70AB (9) and RPA70C/32D/14 (13). The RPA-DBC/dC10 complex was created directly from the 70AB/$dC_8$ (8) and RPA70C/32D/14 crystal structure (13). For the 20mer models, two limiting cases were considered: the first engaged only the RPA70A, B and C domains with nine bases positioned between the 70AB and RPA70C/32D/14 ('extended'). In the compact RPA-DBC/dC20 (and also in the RPA-DBC/dC30 complex) all four domains, 70A, 70B, 70C and 32D, together directly contact ssDNA. For the RPA-DBC/dC20 and RPA-DBC/dC30 complexes, the RPA70C and 32D domains were assumed to bind ssDNA in a binding mode identical to the crystal structure of 70AB (8). The RPA70B domains were aligned to the RPA70C and 32D domains, respectively, by using homologous structural elements, and the position of the ssDNA fragment was identified based on this alignment. The residues Phe532, Tyr581 from RPA70C and residues Trp107, Phe135 from 32D were found in position to stack with the ssDNA bases. Missing loops and the linker between 70B and 70C domains of RPA were built with the program Modeller 9v4 (20).

Hydrogen atoms were introduced using the tLeap module of AMBER 11 (21). To accelerate sampling of the conformational ensemble for each of the systems, we carried out the simulation without inclusion of explicit water using a modified Generalized Born implicit solvent (GBIS) model (22,23), which dramatically reduces the number of degrees of freedom in the system and speeds up the sampling of domain motions. All systems were minimized for 5000 steps with backbone atoms fixed, followed by 5000 steps of minimization with harmonic restraints to remove unfavourable contacts. The systems were then gradually brought up to 300 K and run for 50 ps while keeping the protein backbone restrained. The ssDNA substrates were 'fixed' into each DNA-binding domain during the equilibration period only by defining distance restraints that maintained canonical base-stacking interactions between aromatic residues and DNA bases based on the initial X-ray crystal structures (8). The equilibration was continued for another 50 ns, and the harmonic restraints were gradually released. Production runs were continued for 140 ns for the RPA-DBC system and 200 ns for all ssDNA-containing systems. Two independent simulations for RPA-DBC/

dC30 system were carried out. All simulations used an integration time step of 2 fs with the SHAKE algorithm being applied to fix the bonds between hydrogen atoms and heavy atoms in the systems. The r-RESPA multiple time step method (24) was adopted with a 2 fs time step for bonded, 2 fs for short-range non-bonded and 4 fs for long-range electrostatic interactions. The cut-off for non-bonded interactions and computation of the effective Born radii was set to 18 Å. Dielectric constants of 1 (interior) and 78.5 (exterior) were used in all GBIS-MD simulations. Ionic strength was set to 0.15 M to mimic physiological conditions. All simulations were performed using the NAMD 2.8 code (25,26) with the AMBER Parm99SB parameter set (27) containing the basic force field for nucleic acids and proteins, as well as the refined parameters for backbone dihedrals for protein (SB) and nucleic acids dihedrals (BSC0) on Hopper II, a Cray XE6 system at the National Energy Research Scientific Computing Center.

### Analysis of MD simulations

Computation of scattering profiles from the simulation and comparison to the experimental data used the FoXS code (28). For each of the five models, we computed theoretical scattering profiles for all conformations in the trajectory (last 100 000 frames or 200 ns, except for the DNA-free model with 140 ns and 70 000 frames). The first 50 ns were excluded as equilibration. The theoretical scattering curves were then averaged over the entire ensemble, and these average profiles were compared with the experimental scattering data. This approach of calculating the X-ray scattering from the entire trajectory is distinct from minimal ensemble searches, which typically select only three to five models based on ensemble fits to the data. Averaging over the entire conformational ensemble is expected to better represent solution-phase scattering compared to using a single conformer. To further characterize the structural ensemble, root mean square deviation (rmsd)-based clustering was performed using PTRAJ utility in AMBER 11 and based on a previously reported clustering algorithm (29). The 100 000 frames taken from the trajectory were clustered according to the RMSD, using a cut-off of 6 Å. Data were further analysed using the PTRAJ utility and custom VMD TCL scripts. The 2D histogram figures were generated using Origin 8.0.

## RESULTS

The DNA-binding apparatus of RPA consists of the tandem domains 70AB and the trimer core (Figure 1). As the other three RPA domains (70N, 32N, 32C) are structurally independent (3,11,30,31), we used a construct containing only the DNA-binding core (70ABC/32D/14, RPA-DBC) (12,13) (Figure 1C and D). This construct facilitates interpretation of the scattering data because the extra non-DNA binding domains would add complexity to the analysis without improving the information content in regard to the binding of ssDNA. The long linkers to the 32C and 70N domains are flexible, and 32N is dynamically

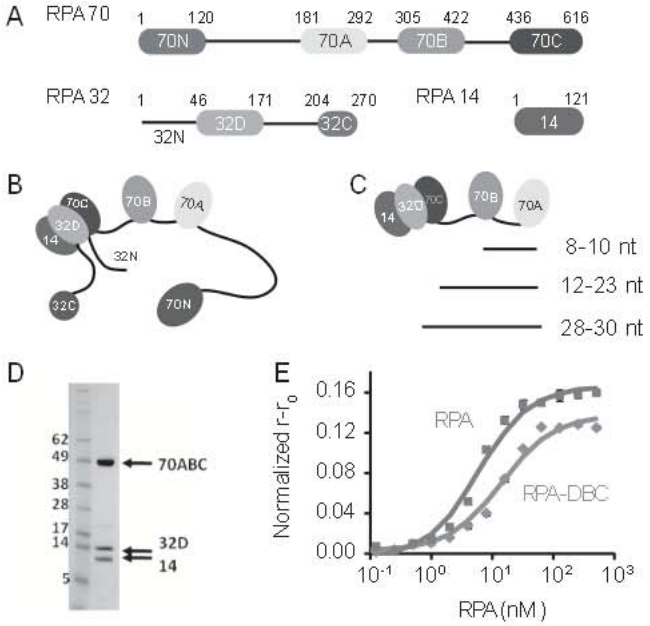


**Figure 1.** Modular, multi-domain RPA has an independent DNA-binding core. (**A**) Domain organization of the RPA70, RPA32 and RPA14 subunits. (**B**) The RPA trimer interface involves domains 70C, 32D and 14. Flexible linkers connect the remaining domains. (**C**) Three proposed modes of RPA binding of ssDNA: the initial 8–10 nt, intermediate 12–23 nt and final 28–30 nt modes engage domains 70A 70B, 70A 70B 70C and 70A 70B 70C 32D, respectively. (**D**) Sodium dodecyl sulphate polyacrylamide gel electrophoresis of purified RPA-DBC, with molecular weight standards (left lane) in kDa. (**E**) Binding curves from fluorescence anisotropy assays of RPA (squares) ($K_d$ 5.8 ± 0.4 nM) and RPA-DBC (diamonds) ($K_d$ 14.2 ± 0.2 nM) binding to $dC_{30}$. Solid lines are fits to a two-state binding model.

disordered; thus, the architecture of the isolated DBC in the absence and presence of ssDNA is expected to be the same as that in full-length RPA.

To test that this model system accurately reproduces the behaviour of full-length RPA, ssDNA-binding affinities were measured using a fluorescence anisotropy assay. To ensure binding of 70A at the 5′-end of each substrate, poly-dC substrates were used with a single adenine at position 3, as reported previously (11). As anticipated, the affinities of RPA and RPA-DBC were similar (Figure 1E) and consistent with those reported previously (32). Subsequent SEC-MALS analysis confirmed the monodispersity of RPA-DBC alone and in complex with its ssDNA substrates (Supplementary Figure S1) and established optimal solution conditions for examining this system by small-angle scattering (33).

The effect of binding ssDNA on the architecture of RPA was investigated using both SAXS and SANS, powerful low-resolution techniques for characterizing the structure of proteins and protein complexes in solution (34–36). SAXS data were acquired on isolated RPA-DBC and complexes with 10-, 20-, 24-, 27- and 30-nt substrates, whereas complementary SANS contrast variation experiments were performed only on the 30-nt complex (Supplementary Figure S2). Guinier analysis verified an absence of sample aggregation in all cases (Supplementary Figures S2 and S6). Joint analysis of the