Author's personal copy

doi:10.1016/j.jmb.2007.09.074 *J. Mol. Biol.* (2007) **374**, 865–876

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

COMMUNICATION

Structure of the Full-length Human RPA14/32 Complex Gives Insights into the Mechanism of DNA Binding and Complex Formation

Xiaoyi Deng[1,2], Jeff E. Habel[3], Venkataramen Kabaleeswaran[3] Edward H. Snell[4], Marc S. Wold[5] and Gloria E. O. Borgstahl[1]*

[1]*The Eppley Institute for Research in Cancer and Allied Diseases, 987696 Nebraska Medical Center, Omaha NE 68198-7696, USA*

[2]*Department of Biochemistry and Molecular Biology 987696 Nebraska Medical Center, Omaha NE 68198-7696, USA*

[3]*The University of Toledo Department of Chemistry 2801 W. Bancroft Street Toledo, OH 43606, USA*

[4]*Hauptman-Woodward Medical Research Institute, Buffalo NY; Department of Structural Biology, SUNY at Buffalo Buffalo, NY 14203, USA*

[5]*Department of Biochemistry University of Iowa College of Medicine, 3107 MERF Iowa City Iowa 52242-2600, USA*

Received 19 May 2007; received in revised form 22 September 2007; accepted 26 September 2007
Available online
2 October 2007

Edited by R. Huber

Replication protein A (RPA) is the ubiquitous, eukaryotic single-stranded DNA (ssDNA) binding protein and is essential for DNA replication, recombination, and repair. Here, crystal structures of the soluble RPA heterodimer, composed of the RPA14 and RPA32 subunits, have been determined for the full-length protein in multiple crystal forms. In all crystals, the electron density for the N-terminal (residues 1–42) and C-terminal (residues 175–270) regions of RPA32 is weak and of poor quality indicating that these regions are disordered and/or assume multiple positions in the crystals. Hence, the RPA32 N terminus, that is hyperphosphorylated in a cell-cycle-dependent manner and in response to DNA damaging agents, appears to be inherently disordered in the unphosphorylated state. The C-terminal, winged helix-loop-helix, protein–protein interaction domain adopts several conformations perhaps to facilitate its interaction with various proteins. Although the ordered regions of RPA14/32 resemble the previously solved protease-resistant core crystal structure, the quaternary structures between the heterodimers are quite different. Thus, the four-helix bundle quaternary assembly noted in the original core structure is unlikely to be related to the quaternary structure of the intact heterotrimer. An organic ligand binding site between subunits RPA14 and RPA32 was identified to bind dioxane. Comparison of the ssDNA binding surfaces of RPA70 with RPA14/32 showed that the lower affinity of RPA14/32 can be attributed to a shallower binding crevice with reduced positive electrostatic charge.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: replication protein A; X-ray crystallography; helix bundle; subunit interaction; OB-fold

**Corresponding author.* E-mail address: gborgstahl@unmc.edu.

Present addresses: J. E. Habel, Scripps Florida, Drug Discovery, 5353 Parkside Drive, Jupiter, FL 33458, USA; V. Kabaleeswaran, Department of Biochemistry & Molecular Biology, The Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, IL 60064, USA.

Abbreviations used: RPA, replication protein A; RPA70, 70 kDa subunit; RPA32, 32 kDa subunit; RPA14, 14 kDa subunit; DBD, DNA binding domain; OB-fold, oligosaccharide/oligonucleotide binding-fold; wHLH, winged helix-loop-helix domain; MAD, multiple-wavelength anomalous dispersion.

0022-2836/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.

Author's personal copy

Replication protein A (RPA) plays important roles in DNA replication, recombination, and repair by binding and stabilizing single-stranded (ss)DNA generated during DNA metabolism.[1,2] RPA is a high affinity, ssDNA binding protein composed of three subunits that have been named for their molecular weight as RPA70, RPA32, and RPA14 (Figure 1(a)) and is conserved in all eukaryotic cells.[1–4] Recently, RPA was found to be involved in the cell-cycle signaling pathway by regulating the function of ATR–ATRIP complex, and the ssDNA–RPA complex may function as a DNA damage signal to recruit down-stream proteins involved in DNA repair or cell apoptosis.[5,6] RPA exists in two soluble forms: heterotrimer and RPA14/32 heterodimer.[7,8] Interestingly when protein expression levels were analyzed in *Saccharomyces cerevisiae* using a GFP-tagging method, there was significantly more RPA32 expressed as compared to RPA70 or RPA14 (6080, 4100, and 4280 molecules per cell, respectively; the

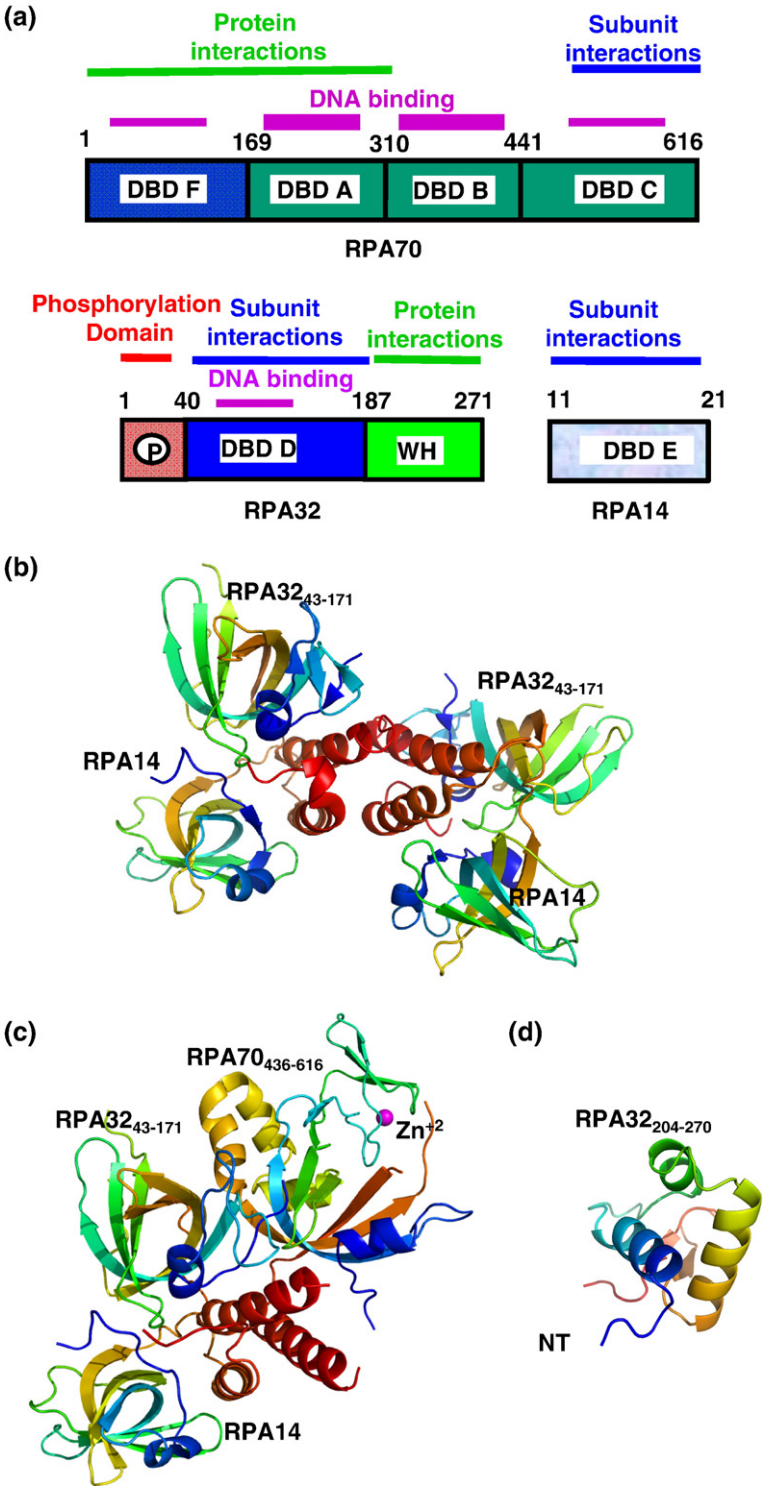

Figure 1. Structural information for RPA. (a) Schematic of RPA heterotrimer[19] with DBDs, regions that are phosphorylated and that interact with proteins and/or DNA are labeled. (b) The RPA14/$32_{43\text{-}171}$ crystal structure; including RPA32 residues 45–110, 117–171 and RPA14 residues 3–116 (PDB ID, 1QUQ).[13] Note, the four-helix bundle (in red) is located at the interface between labeled heterodimers. (c) The RPA14/$32_{43–171}$/$70_{436–616}$ crystals structure (PDB ID, 1L1O).[26] (d) NMR structure of the RPA32 winged helix-loop-helix domain, including RPA32 residues 204–270 (PDB ID, 1DPU).[33] Ribbon diagrams were made with Pymol [http://pymol.sourceforge.net/] and rainbow-colored blue to red from the N to C-terminal ends.

Author's personal copy

RPA70 level was measured in triplicate with error± 500).[9] The heterotrimer is thought to be the biologically important form of RPA in the cell; however, the RPA14/32 heterodimer has been observed *in vitro*,[8,10] in cells undergoing apoptosis[11] and may be important in binding telomere sequences.[12] Here several crystal structures of RPA14/32 have been solved and analyzed.

RPA can bind both ssDNA and proteins involved in DNA metabolism.[2] There are four DNA binding domains (DBDs) located in RPA70 and one located in RPA32 (Figure 1(a)). RPA70 has the highest ssDNA binding affinity among the three subunits. RPA14 is involved in heterotrimer formation and, although it contains a similar fold, DNA binding has not been demonstrated in humans.[1,2,13] A recent study by Gao *et al* has shown that in yeast the RPA32 and RPA14 subunits interact with telomere sequences.[12] RPA32 contains a weak DBD,[14] an N-terminal phosphorylation domain,[15–17] and a C-terminal protein–protein interaction domain.[2,16] There are six to nine N-terminal Ser and Thr residues on RPA32 which can be phosphorylated specifically during the cell cycle and in response to DNA damage.[18,19] Phosphorylated RPA32 serves as a marker for damaged DNA and is thought to reduce association of RPA with replication centers.[20] RPA32 is involved in several DNA repair pathways and specifically binds several repair proteins, including XPA in nucleotide excision repair,[21] Rad52 in double strand break repair,[22] and UNG2 in base excision repair.[23]

A wealth of structural data on RPA has been provided by X-ray crystallography and solution NMR. The crystal structures of $RPA70_{181–422}$ alone and in a complex with a short oligonucleotide, showed how the oligosaccharide/oligonucleotide binding-folds (OB-folds) of DBDs A and B bind ssDNA with high affinity.[24,25] The crystal structure of $RPA14/32_{43–171}$, containing a fragment of RPA32 that was resistant to proteolysis, showed that RPA14 and the RPA32 core were structurally homologous, both containing an OB-fold with a conserved N-terminal extension and a C-terminal α-helix.[13] The loop between β strands 3 and 4 (L34 loop) of the OB-fold of $RPA32_{43–171}$ was disordered and not included in the model. Two C-terminal α-helices were identified to form a four-helix bundle interface between RPA14/32 heterodimers (Figure 1(b), red helices). The crystal structure of the trimerization core, $RPA14/32_{43–171}/70_{436–616}$, revealed the structure of zinc-containing RPA70 DBD-C and the complex was mediated by three C-terminal helices arranged in parallel (Figure 1(c), red helices).[26] These observations that the OB-folds in RPA tend to form aligned helical bundles is a central part of the current hypothetical mechanism for heterotrimer formation and the conformational changes associated with ssDNA binding.[26,27] A structure of the RPA32 C terminus, including residues 172 to 270, in complex with DNA repair factors such as UNG2 was determined by NMR and showed that the C terminus has a winged helix-loop-helix domain (wHLH) (Figure 1(d)).[28] The N-terminal portion of this fragment, including residues 172–204, was unstructured in solution. Solution NMR heteronuclear single quantum coherence (HSQC) data on full-length RPA14/32 indicated that residues 1–40 were intrinsically disordered and suggest that the wHLH domain tumbles independently of $RPA14/32_{43–171}$.[28] The crystal structures of full-length RPA14/32 reported here shed light on the structural features of the full-length RPA14/32 and how they relate to function.

Structure of full-length RPA14/32

Native X-ray diffraction data were collected from two crystal forms of full-length RPA14/32 which belonged to orthorhombic and tetragonal space groups and contained four and two molecules of RPA14/32 heterodimer in the crystallographic asymmetric unit, respectively (Table 1). Crystallographic phases were solved by molecular replacement using the previously solved $RPA14/32_{43–171}$ core structure (PDB ID, 1QUQ) as a search model. Additional molecular replacement calculations were attempted including the RPA32 C-terminal wHLH domain[28] (PDB ID, 1DPU; Figure 1(d)) but without success. Then, portions of the 1QUQ search model (~10%) were systematically omitted from the phase calculation and these non-biased electron density maps were visually inspected to ensure that the molecular replacement calculations were correct. The maps from both the orthorhombic and tetragonal crystals had excellent electron density in the central core region and relatively poor electron density for the N and C-terminal regions (residues 1–40 and 177–270, respectively; Figure 2(a)). The orthorhombic structure was of higher data quality (2.0 Å resolution); therefore it was completely remodeled, including refitting the side-chains for correct hydrogen bonding, rebuilding the ends of the core search model and extending the ends by two and six more residues on the N and C termini of RPA32, respectively; the RPA14 subunit was nearly completely modeled (chain F in the crystallographic asymmetric unit), and solvent molecules (244 water and 4 dioxane) were modeled. The 6XHis tag was disordered in all crystals. The remaining density for the missing N and C-terminal regions was weak and not continuous. To ensure that the crystals contained fully intact RPA14/32, the crystals were dissolved, and SDS–PAGE analysis indicated that the RPA14/32 in the crystal was intact (Figure 2(b)). All attempts to improve crystal quality failed (see Supplementary Data). The coordinates were refined against the diffraction data to a final crystallographic R value of 23.2% with R_{free} 25.5% with reasonable stereochemistry (Table 1).

Selenylmethionine crystals were grown and multiple-wavelength anomalous dispersion (MAD) diffraction data collected in order to obtain experimental phases without model bias. The hexagonal crystal had a suspiciously low solvent content that could indicate twinning so the diffraction data was interrogated using the Merohedral Crystal

Author's personal copy

Table 1. Crystallographic data and refinement statistics

A. *Data collection statistics*					
Crystal form	Tetragonal	Orthorhombic	Hexagonal		
Temperature (K)	100	100	100		
X-ray source	Rigaku-FRE	SSRL BL7-1	APS IMCA-CAT BL17-ID		
Space group	$P4_1$	$P2_12_12_1$	$P6_5$		
			Peak	Edge	Remote
Wavelength (Å)	1.54	1.08	0.9794728	0.979609	0.999879
Cell dimensions (Å)					
a	96.7	81.5	63.4		
b	96.7	139.8	63.4		
c	126.3	171.8	272.7		
Resolution (Å)	3.0	2.0	2.5		
No. observations	32,072	124,950	24,160	24,253	24,393
Redundancy	3.1 (3.0)	3.5 (2.8)	6.0 (5.9)	6.0 (5.9)	6.0 (5.9)
Completeness (%)	95.8 (99.5)	100 (98.7)	95.7 (97.1)	88.3 (88.6)	95.7 (97.2)
$I/\sigma(I)$	11.2 (2.9)	19 (2.7)	16 (3.8)	15 (2.1)	24 (3.5)
R_{sym} (%)	9.3 (40.1)	5.5 (40)	9.8 (53.3)	8.7 (56.4)	8.4 (58)
B. *Refinement statistics*					
PDB ID	2Z6K	2PI2	2PQA		
No. RPA14/32 in ASU	2	4	2		
Solvent (%)	60	55	24		
R_{cryst} (%)	24.9	23.2	22.8		
R_{free} (%)	29.0	25.5	27.9		
No. of atoms	3823	7883	3779		
No. of water molecules	0	244	0		
rms deviations from ideality					
Bond lengths (Å)	0.008	0.011	0.012		
Bond angles (deg.)	1.2	1.37	1.51		
Ramachandran plot					
Most favored (%)	81.0	89.7	87.3		
Additionally allowed	18.2	9.5	11.9		
Generously allowed	0.8	0.8	0.8		

Human RPA14/32 was expressed and purified in the presence of a cocktail of protease inhibitors as described.[7] Selenomethionyl RPA14/32 was prepared using a methionine pathway inhibition method.[42] The proteins were concentrated using an Amicon Ultra centrifugal filter and stored at −80 ° C in 10 mM Hepes (pH 7.5), 200 mM KCl, and 10 mM dithiothreitol (DTT) with native protein at 20 mg/ml and selenomethionyl protein at 7.8 mg/ml. Crystals of RPA14/32 were grown by hanging drop vapor diffusion with 4 μl drops composed of an equal volume of protein and reservoir solution. The tetragonal crystals grew in 0.95 M potassium sodium tartrate, 0.1 M Mes (pH 5.9), and 10 mM DTT. The orthorhombic and hexagonal crystals were grown as described.[7] Crystals were cryo-protected by immersion in reservoir solution with 30% (v/v) glycerol for a few seconds and/or by coating with paratone-N oil and then cryo-cooled in a 100 K N_2 gas stream. Diffraction data for the orthorhombic crystal were collected Stanford Synchrotron Radiation Laboratory beamline 7-1 as described.[7] Diffraction data for the tetragonal crystal was collected in-house with a Rigaku-FRE X-ray generator, R-axis IV++ detector and Xstream 2000 low temperature system. Diffraction data were integrated and intensities scaled with the HKL2000 package.[43] PDB entry 1QUQ was used as the search model to solve the tetragonal and orthorhombic crystal forms by molecular replacement using Phaser[44] as implemented in the CCP4 package.[45] Phases were improved with DM.[46] The MAD data for the hexagonal crystal were collected at the Advance Photon Source (APS) beamline IMCA17 BL17-ID at three wavelengths, integrated with PROW[47] and intensities were scaled with the HKL2000 package. SHELX[48] was applied to find 16 selenium sites and four more sites were identified from Fourier map analysis using SHARP.[49] These 20 Se sites were applied to calculate the initial MAD phasing information by using SOLVE/RESOLVE[50] with 10% solvent flattening (FOM = 0.51). Restrained crystallographic refinement were undertaken using Refmac.[51]

Values for highest resolution shell are in parentheses. $R_{sym} = \sum_{hkl}|I_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum I_{hkl}$. Solvent content was calculated based on the Matthews method.[52] $R_{factor} = \sum||F_o| - |F_c|| / \sum|F_o|$, where F_o and F_c are the observed and calculated structure factor, respectively. R_{free} is equal to *R*-factor for a randomly selected 5% subset of reflections not used in the refinement.[53] Ramachandran analysis was performed with PROCHECK.[54]

ASU, asymmetric unit.

Twinning Server and no evidence for twinning was found.[29] The positions of 20 out of a possible 32 selenium positions were identified using direct and Fourier methods, MAD phases and an electron density map were calculated. All 20 selenium positions were located in the RPA14/$32_{43–171}$ core structure (see anomalous difference map in Supplementary Data, Figure S1). The missing 12 selenium positions were not observed in the anomalous difference electron density maps and were located in disordered regions. The MAD-phased Fourier electron density of the core structure was well ordered and, again, was relatively poor for the N and C-terminal regions (residues 1–41 and 173–270, respectively; see Fourier map in Supplementary Data, Figure S2). Modeling included refitting the side-chains for correct hydrogen bonding, rebuilding the ends of the core search model and extending these ends by one to two more residues. An additional structural feature was built into the hexagonal crystal structure that did not exist in any other crystals. The L34 loop (Figure 2(c)) for RPA32 was built (chain C in the crystallographic asymmetric unit). This loop is not well conserved in eukaryotes and may include a species specific protein–protein interaction site for human partner proteins (see

Author's personal copy

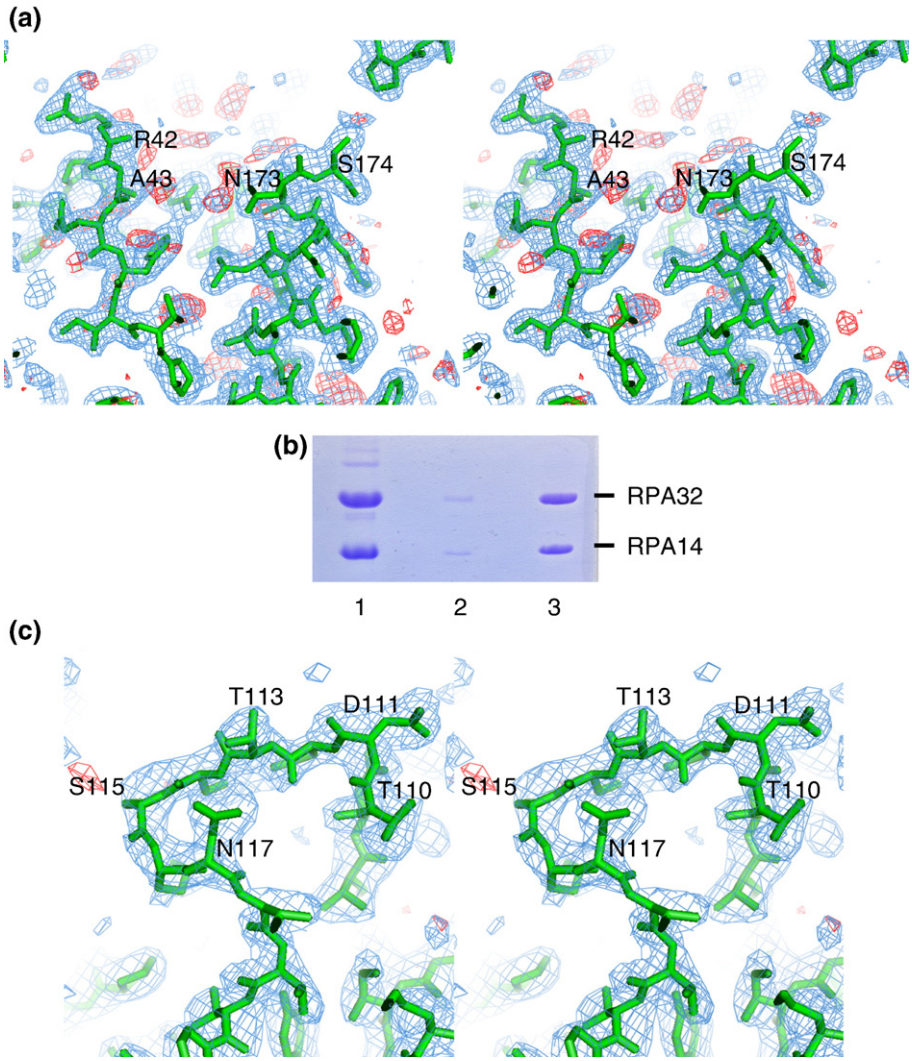

Figure 2. Analysis of full-length RPA14/32 crystals. (a) Region of RPA32 where disorder starts from the orthorhombic crystal. (b) SDS-PAGE analysis of the RPA14/32 crystal (lanes: 1, RPA14/32 full-length crystal; 2, mother liquor that surrounded crystal; 3, purified RPA14/32 full-length protein). (c) Well-ordered L34 loop of RPA32 from the hexagonal crystal that was disordered in 1QUQ and the other full-length crystal structures. Stereo diagrams include green final atomic models and Fourier maps where blue is $2F_o$–F_c displayed at 1σ and red is F_o–F_c electron density displayed at 3σ.

sequence alignments by Millership *et al.*[3] and Ishibashi *et al.*[4]). Thus, by combining results from the orthorhombic crystal with the hexagonal crystal the complete secondary structure for RPA32 residues 41 to 176 and RPA14 residues 2 to 120 were defined (Figure 3(a)).

It was concluded that the N and C-terminal regions of RPA32 are disordered when they are not phosphorylated or bound by partner proteins. In the heterodimer, both the N and C-terminal domains have no specific interactions with the protein core and are connected to the core by flexible linkers. A relatively large fraction of the protein is disordered, approximately 37%, and it is remarkable that the crystals diffract so well. To assist our interpretation of these data we analyzed the sequence of RPA14 and RPA32 using PONDR software that predicts naturally disordered regions[30–32] and compared them to the inherent motion of the refined atomic model (this information is contained in the refined isotropic *B*-values). Overall PONDR predicted the positions where the disordered N and C-terminal regions begin accurately and the apparent motion in the refined crystal structure matches the PONDR prediction very well (see Supplementary Data, Figure S4). PONDR also predicts that the RPA32 C-terminal wHLH to be intrinsically well-ordered (Supplementary Data, Figure S4, part A). The failure of molecular replacement calculations with the NMR-determined RPA32 C-terminal wHLH domain (Figure 1(d)) could be caused by several factors, including a somewhat different structure for the RPA32 C-terminal domain in the intact protein, or that the wHLH domain in full-length protein adopts multiple positions smearing the electron density signal to background levels. It is not possible to distinguish between these two possibilities; although NMR data on the wHLH domain alone[28] and PONDR analysis support the latter conclusion. The N-terminal domain is

Author's personal copy

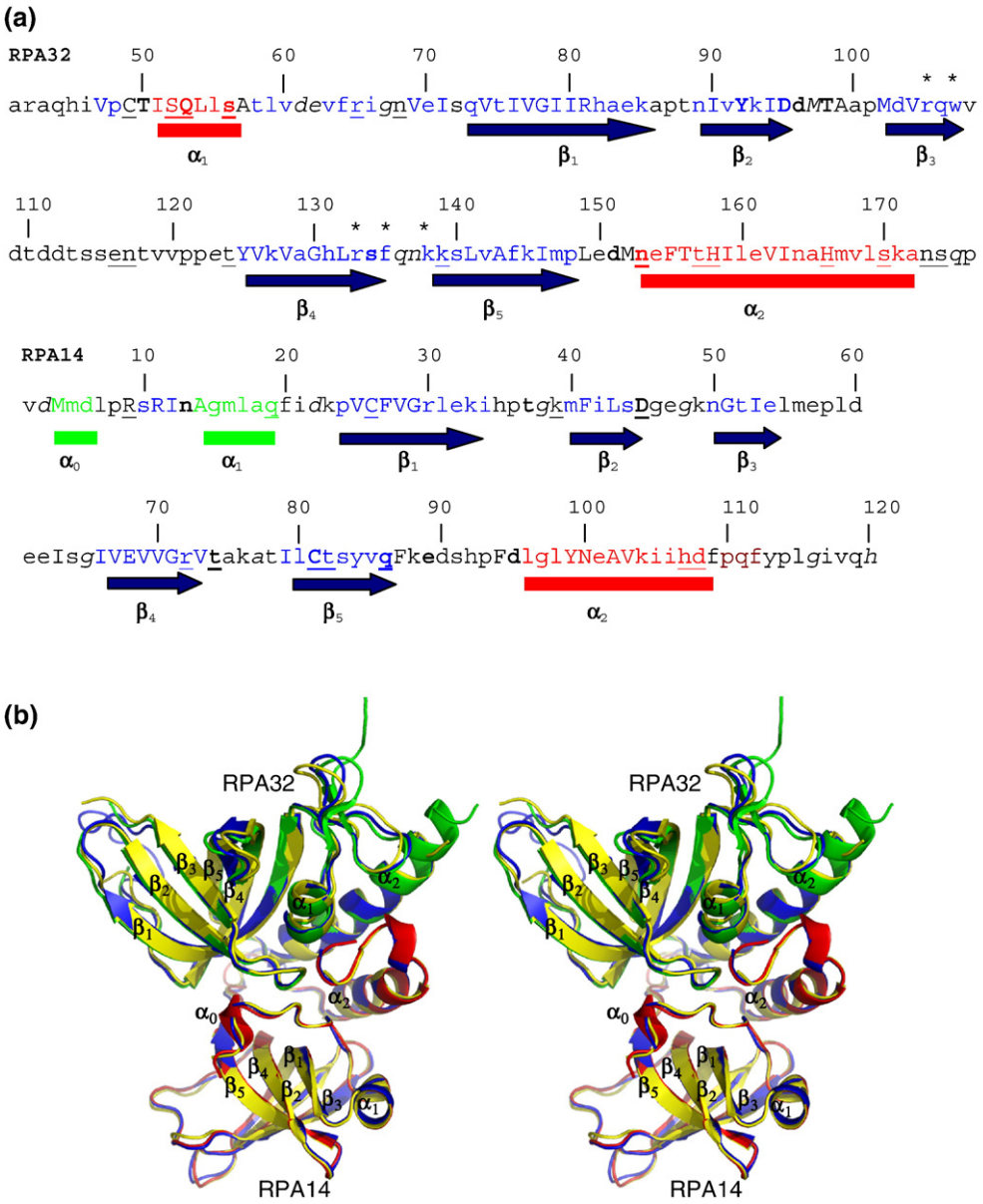

Figure 3. Sequence and structure summary of full-length RPA14/32 crystals (a) Annotated protein sequence and secondary structure analysis using JOY software[40] compiled from orthorhombic and hexagonal crystals. The Figure is organized as follows. Line 1, residue number. Line 2, an asterisk indicates putative ssDNA binding residues that were identified by the comparison in Figure 6. Line 3, JOY annotation of the protein sequence, where α-helices are red, β-strands are blue, and 3_{10}-helices are green; solvent-accessible residues are in lower case, solvent-inaccessible residues are in upper case, residues making hydrogen bonds to main-chain amides are in bold, hydrogen bond to main-chain carbonyl is underlined; and positive phi torsion angle is in italic. Line 4, secondary structure: α-helices have a red rectangle, 3_{10}-helices have a green rectangle, β-strands have a blue arrow. Line 5, major secondary structure elements are named as defined in 1QUQ structure. (b) Orthorhombic heterodimer (RPA14 red, RPA32 green) superimposed onto the hexagonal crystal structure (blue) and RPA14/$32_{43–171}$ (yellow). Common residues between the RPA14 subunits were superimposed using with RMS differences of 0.57 Å and 0.59 Å, respectively. RPA14 was superimposed with Lsqkab, and the RMSD of the resulting RPA32 coordinates were then calculated by Molman2.

probably intrinsically disordered. These observations are in good agreement with earlier interpretations of NMR data on full-length RPA14/32.[33] This would allow the two domains to be independent of each other and may facilitate interactions with various kinases and DNA repair proteins. Intrinsically disordered proteins are known to often participate in high specificity/low affinity binding. Disordered regions may facilitate binding to a wide variety of structurally distinct substrates[34] and protein phosphorylation is thought to predominantly occur within intrinsically disordered protein regions.[35]

Orientation of the heterodimer subunits

To examine any differences between the RPA14/$32_{43–171}$, orthorhombic full-length RPA14/32 and

Author's personal copy

hexagonal full-length RPA14/32 quaternary structures, the RPA14 subunits were superimposed (RMSD 0.6 Å) and then the RPA32 subunits were examined (Figure 3(b)). After superposition of RPA14, examination of the RPA32 structures shows that they are very similar with C^{α} differences of 0.90 Å on average. In this analysis the subunit–subunit interface contacts are essentially the same with average C^{α} differences of 0.5 Å. The largest differences in structure were at the RPA32 surface loops that differed by 2 Å to 4 Å (Figure 3(b)). The hexagonal crystal structure also included the L34 loop (Figure 3(b), blue line, behind strands β_1, β_2, β_3). We conclude that the orientation of the RPA14 subunit with the RPA32 subunit is the same in the RPA14/$32_{43–171}$ proteolytic core structure and in both of the full-length RPA14/32 structures.

Higher-order RPA14/32 interactions and implications

Historically, the first structure that gave information on the subunit–subunit interactions that hold the RPA heterotrimer together was the RPA14/$32_{43–171}$ crystal structure. A four-helix bundle was observed between RPA14/$32_{43–171}$ heterodimers and it was proposed that a similar helical bundle was involved in the trimerization mechanism of the RPA heterotrimer.[13] This four-helix bundle was formed by duplicating two parallel helices (helices α_2 of RPA14 and α_2 of RPA32;Figure 3(a) and (b)) with a non-crystallographic 2-fold axis of symmetry (Figure 1(b), red helices). Then, the crystal structure of the RPA trimerization core RPA14/$32_{43–171}$/$70_{436–616}$ was solved and it contained a parallel three-helix bundle comprised of RPA14 α_2 and RPA32 α_2 helices plus the RPA70 α_3 helix (Figure 1(c), red helices).[26] As a result of these two structures, a molecular mechanism involving the coming together of parallel helical bundles was developed for how the RPA heterotrimer quaternary structure forms and changes in the different ssDNA binding modes of full-length RPA.[27] We postulated that if the heterodimer four-helix bundle was indicative of this mechanism, this structure should be conserved in all crystal forms of the full-length heterodimer.

To compare the RPA14 α_2 and RPA32 α_2 helices of full-length protein with RPA14/$32_{43–171}$, ribbon drawings were made with the orientation of one RPA14/32 heterodimer set so that the relative positions of the α_2 helices would be easily observed (left side RPA14/32 has same orientation in Figure 1(b) and Figure 4). The orientation of the four helices are similar between the tetragonal and the orthorhombic crystal structures (Figures 4(a) and (b)), while the helices do not form a bundle at all, but are stacked in the hexagonal crystal structure (Figure 4(c)). The four helices do not form a four-helix bundle structure in any the three full-length crystal structures. In fact, the dimer–dimer interface is completely different compared to the orientation in the RPA14/$32_{43–171}$ structure (Figure 1(b)) in all three structures (see also Supplementary Data, Tables S2, S3, and S4 that list the residues involved in the dimer–dimer interface).

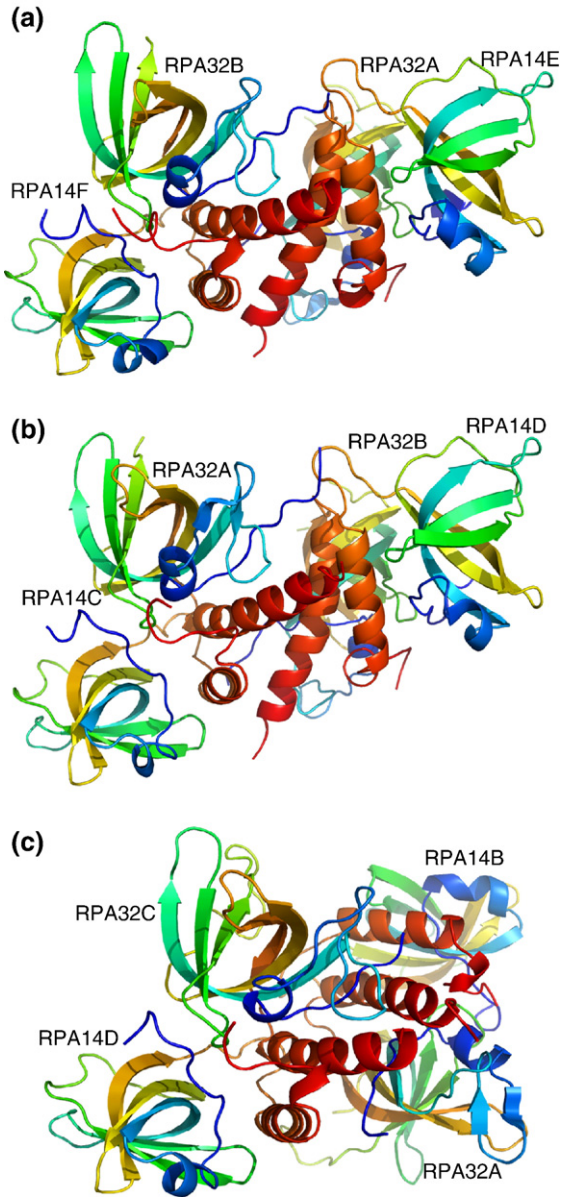

Figure 4. Comparison of the dimer of heterodimers quaternary structure in the full-length RPA14/32 crystals: (a) orthorhombic, (b) tetragonal, and (c) hexagonal crystal forms. Each subunit is rainbow-colored. The RPA14/32 on the left and has approximately same orientation in each panel and in Figure 1(b). In each crystal, the dimer-of-heterodimers is related by non-crystallographic symmetry. Subunit names and corresponding chain IDs in the Protein Data Bank entry are labeled. Note, the four-helix bundle seen in 1QUQ was not observed in the full-length crystal structures.

To compare the orientation of the RPA14 α_2 and RPA32 α_2 helices of full-length RPA14/32 with the RPA70 α_3 helix from the trimerization core RPA14/$32_{43–171}$/$70_{436–616}$, the C^{α} positions of RPA14/32 dimers were superimposed and only the helices displayed (Figure 5). As was originally noted by Bocharev, the RPA70 α_3 helix from the trimerization

Author's personal copy

core lies in between the RPA14 α_2 and RPA32$_{43–171}$ α_2 helices (Figure 5(a), gray helices). The orientation of the RPA70$_{436–616}$ α_3 helix (gray) is at a 15–25 degree angle and antiparallel to the RPA14 α_2 and RPA32$_{43–171}$ α_2 helices of the heterodimer core (Figure 5(a), red and green helices). In similar comparisons, the orientation of the RPA70$_{436–616}$ α_3 helix is at a 75–85 degree angle with the RPA14 α_2 and RPA32 α_2 helices of full-length RPA14/32 from the orthorhombic and tetragonal crystal forms (Figure 5(b)) and bears no relationship to the helices from the hexagonal crystal from (Figure 5(c)). Thus, the helical bundle quaternary structure does not form consistently and may be an artifact of the crystal or the deletion mutation. On the other hand, we believe that these differences in quaternary structure may reflect the actual structural repertoire of the RPA heterotrimer and the relative locations of the non-superimposed RPA14/32 α_2 helices in Figure 5 may represent alternate locations for the RPA70 α_3 helix. Also, these structures could predict the locations of C-terminal helices of DBD-B and DBD-C relative to RPA14/32 α_2 helices in the sequential binding model for RPA (see Figure 2 of Bochkarev & Bochkareva[27]). Future work would include point mutation studies to deconvolute these possible quaternary structures of the RPA heterotrimer.

The RPA heterotrimer quaternary structure is remarkably stable and does not dissociate in 6 M urea.[36] This level of stability suggests that a significant amount of surface area is buried when the heterotrimer complex is formed. The full-length RPA14/32 structures all form a more extensive heterodimer–heterodimer interface (buried surface areas of 1546 Å^2 for orthorhombic, 1780 Å^2 for hexagonal) relative to the RPA14/32$_{43–171}$ structure (1126 Å^2 for RPA14/32$_{43–171}$). Thus the higher buried surface area observed with the full-length RPA14/32 dimer crystals suggests that this interface may represent possible interfaces available to the heterotrimer in solution.

Identification of an organic ligand binding site

Crystallization of the full-length RPA14/32 heterodimer was facilitated by the addition of organic solvents.[7] These solvents included acetonitrile, ethanol, isopropanol and dioxane. Since the orthorhombic crystal form diffracted to 2.0 Å resolution, ordered solvent molecules were easily observed in the well-ordered core regions of the protein. In this crystal a dioxane binding site was also clearly observed in the electron density (Figure 6(a)). Organic solvent was also visible in this location in the hexagonal crystals (data not shown). The dioxane binds at the subunit–subunit interface of the heterodimer and is nestled between residues Q86 and K88 of RPA14 and E123, Y125, and L149 of RPA32. Dioxane forms hydrogen bonds with main-chain amide nitrogen atoms of Y125 and K88. The electrostatic surface at the dioxane binding site is negatively charged with a small positively charged surface created by K88 (Figure 6(b), top view,

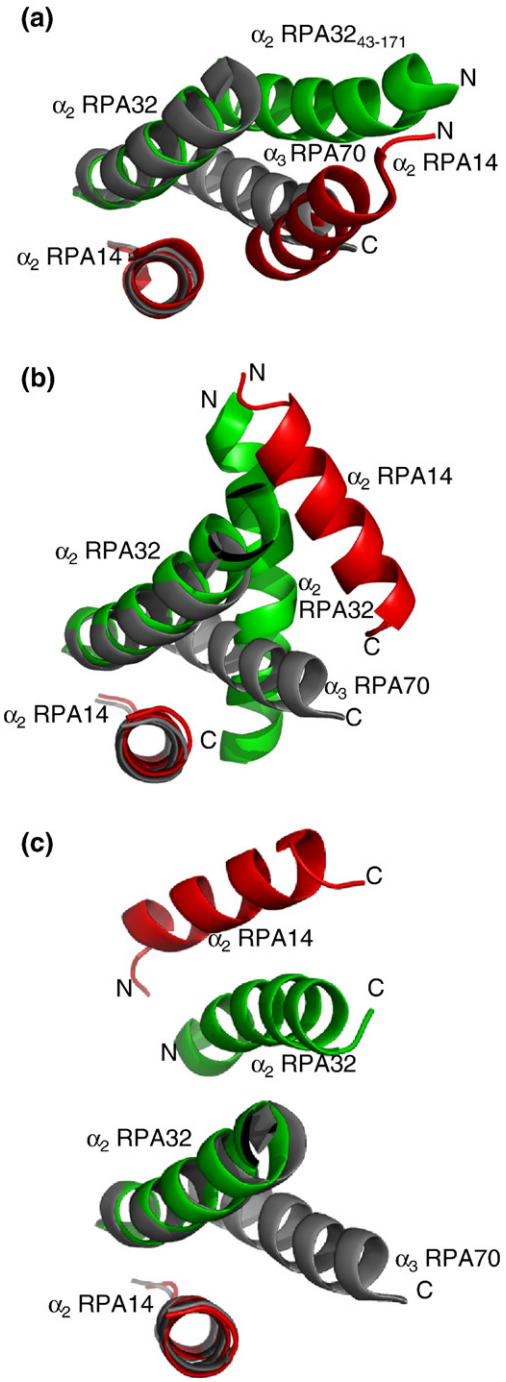

Figure 5. Comparison of the quaternary structure of full-length RPA14/32 structures with the RPA trimerization core structure.[26] The C-terminal helices of the RPA trimerization core (all three helices are gray) are displayed with corresponding helices from the following RPA14/32 crystal structures (RPA14 α_2 helices are red and RPA32 α_2 helices are green): (a) RPA14/32$_{43–171}$ (PDB ID, 1QUQ; RMSD 0.42 Å), (b) orthorhombic crystal forms (RMSD 0.36 Å) (c) full-length hexagonal crystal (RMSD 0.42 Å). For superposition, the entire RPA14/32$_{43–171}$ component of the trimerization core was superimposed with the corresponding residues in the RPA14/32 crystal structures.

Author's personal copy

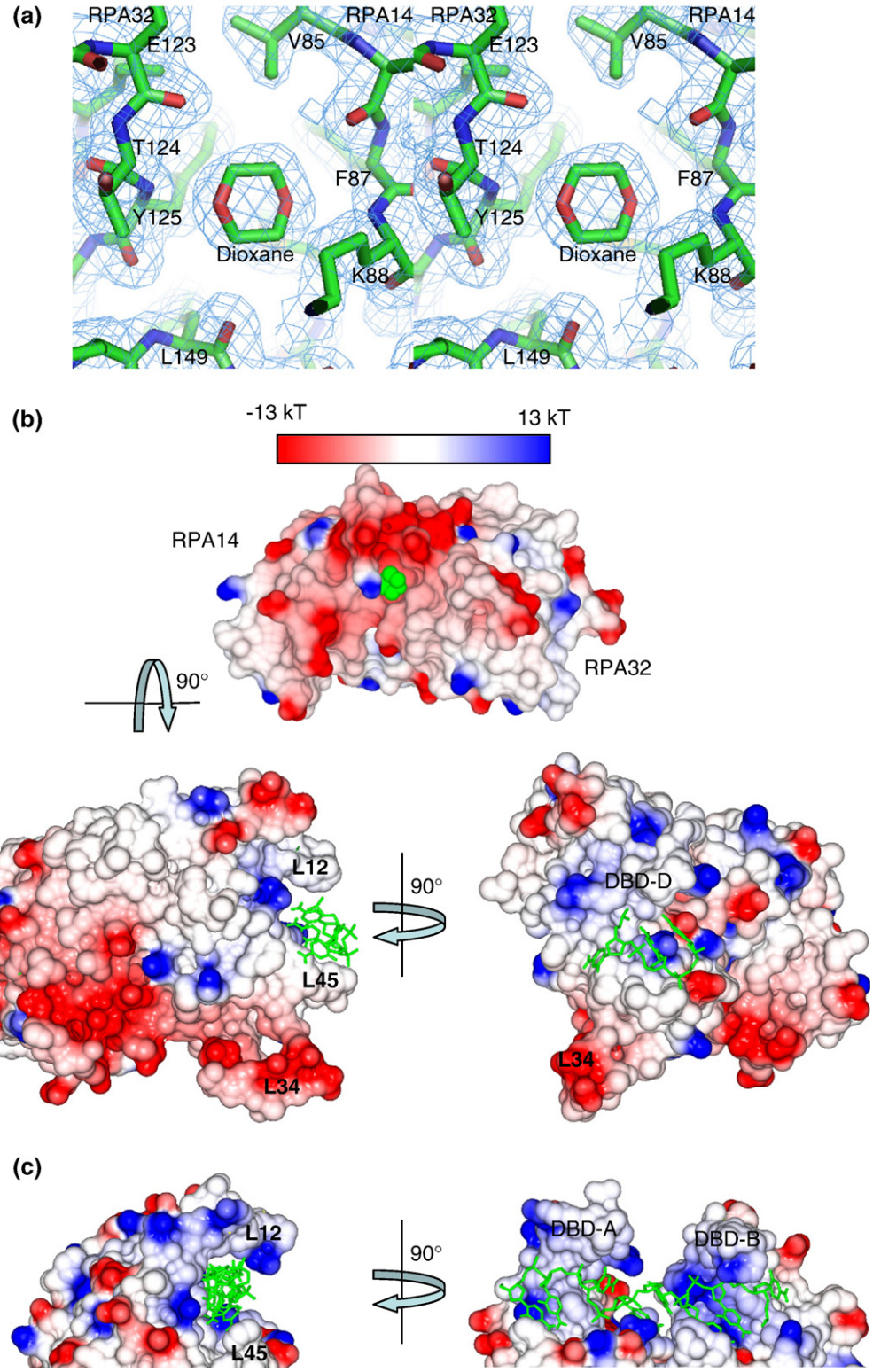

Figure 6. Surface analysis of RPA binding sites. (a) Electron density for the dioxane binding site from the orthorhombic crystal. Stereo diagram includes green final atomic models (carbon green, oxygen red, nitrogen blue) and Fourier maps where blue is $2F_o–F_c$ displayed at 1σ. (b) Three orthogonal views of the electrostatic surface of RPA14/32 from the hexagonal crystals. The dioxane binding site (green spheres) is shown in the top view. A putative dC_4 oligonucleotide (green lines) binding site is show in the bottom views. The oligonucleotide was positioned by superimposing the OB-fold from RPA70A (PDB ID, 1JMC) onto RPA32. (c) Electrostatic surface and oligonucleotide binding site of RPA70AB (PDB ID, 1JMC).[24] Electrostatic surfaces were calculated with CCP4mg.[41]

Author's personal copy

dioxane atoms are green). Molecular modeling indicates that pyrimidine bases could also fit into the dioxane site (G.E.O. B., unpublished results). The physiological relevance of this binding site is currently uncertain but it could represent a previously unidentified regulatory site on RPA. Preliminary data indicate that 5% (v/v) dioxane significantly lowers the affinity of RPA for ssDNA (A. Prakash, unpublished results) and has no detectable effect on the subunit composition of RPA (X. Deng, unpublished results). It is possible that this organic ligand binding site on RPA could be of pharmaceutical/experimental use and perhaps this binding site could be exploited in the future to find/design molecules that specifically affect the function of RPA. The specific regulatory effects of this binding site on RPA function will require further study.

Putative ssDNA binding site

RPA heterotrimer binds ssDNA with several DNA binding domains (Figure 1(a)) and with high affinity. RPA70 DBDs A and B constitute the high affinity ssDNA-binding core and bind with only 10^1-fold lower affinity than the intact heterotrimer.[37,38] RPA14/32 DBD-D has weak ssDNA binding affinity that is approximately 10^6-fold lower than the heterotrimer and 10^3-fold lower than RPA70 DBD-A. Interestingly RPA14/32 may be involved in binding telomeric sequences.[12] How structure relates to ssDNA binding affinity was apparent when the electrostatic surfaces of RPA70 were compared with RPA14/32 (Figure 6(b) and (c)). Oligonucleotide binding to RPA14/32 was predicted by superimposing the OB-folds of RPA32 DBD-D onto RPA70 DBD-A with its corresponding ssDNA.[39] For RPA70 DBDs A and B, the loops between β strands 1 and 2 (L12 loop) and β strands 4 and 5 (L45 loop) form a deep ssDNA binding pocket that has an overall positively charged surface (Figure 6(c)). The L12 and L45 loops on RPA32 DBD-D are shorter making a binding pocket that is relatively shallow, narrow and has less positive charge on its surface (Figure 6(b)). The relatively narrow, and shallow ssDNA binding cleft and weak electrostatic interaction cause RPA14/32 to have millimolar ssDNA binding affinity (O. Dickson & M.S.W., unpublished results). The dramatically higher affinity of RPA70 DBD-A and B can be attributed to containing two OB-folds, deeper binding pocket that electrostatically complements the negatively charged phosphate backbone of the ssDNA. Interestingly, the loop between β strands 3 and 4 (L34 loop) that was observed and modeled in the hexagonal crystals is located below the DNA binding site (Figure 6(b)), and it is heavily, negatively charged. The L34 loop may play a role in regulating protein–DNA/protein–protein interactions.

Database accession numbers

Structure factors and atomic coordinates for the orthorhombic and hexagonal RPA14/32 crystals have been deposited in the RCSB Protein Data Bank with accession codes 2Z6K, 2PI2 and 2PQA, respectively.

Acknowledgements

We thank George Sheldrick, Andy Howard, and Tom Terwilliger and for technical advice and assistance. We thank Aishwarya Prakash for helpful discussion and technical assistance. This work was supported by the U. S. Army Medical Research and Materiel Command under DAMD17-98-1-8251, the American Cancer Society RSG-02-162-01-GMC, NASA grant NAG8-1983, the Ohio Cancer Research Associates, the DeArce Memorial Endowment Fund and the University of Toledo. M.S.W. was supported by NIH grant GM044721.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.09.074

References

1. Iftode, C., Daniely, Y. & Borowiec, J. A. (1999). Replication protein A (RPA): the eukaryotic SSB. *Crit. Rev. Biochem. Mol. Biol.* **34**, 141–180.
2. Wold, M. S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92.
3. Millership, J. J., Cai, X. & Zhu, G. (2004). Functional characterization of replication protein A2 (RPA2) from *Cryptosporidium parvum*. *Microbiology*, **150**, 1197–1205.
4. Ishibashi, T., Kimura, S., Furukawa, T., Hatanaka, M., Hashimoto, J. & Sakaguchi, K. (2001). Two types of replication protein A 70 kDa subunit in rice, *Oryza sativa*: molecular cloning, characterization, and cellular and tissue distribution. *Gene*, **272**, 335–343.
5. Wu, X., Yang, Z., Liu, Y. & Zou, Y. (2005). Preferential localization of hyperphosphorylated replication protein A to double-strand break repair and checkpoint complexes upon DNA damage. *Biochem. J.* **391**, 473–480.
6. Namiki, Y. & Zou, L. (2006). ATRIP associates with replication protein A-coated ssDNA through multiple interactions. *Proc. Natl Acad. Sci. USA*, **103**, 580–585.
7. Habel, J. E., Ohren, J. F. & Borgstahl, G. E. O. (2001). Dynamic light scattering analysis of full-length, human RPA14/32 dimer: purification, crystallization and self-association. *Acta Crystallog. sect. D*, **57**, 254–259.
8. Henricksen, L. A., Umbricht, C. B. & Wold, M. S. (1994). Recombinant replication protein A: expression, complex formation and function characterization. *J. Biol. Chem.* **269**, 11121–11132.
9. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N. *et al.* (2003). Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.

Author's personal copy

10. Treuner, K., Findeisen, M., Strausfeld, U. & Knippers, R. (1999). Phosphorylation of replication protein A middle subunit (RPA32) leads to a disassembly of the RPA heterotrimer. *J. Biol. Chem.* **274**, 15556–15561.
11. Treuner, K., Okuyama, A., Knippers, R. & Fackelmayer, F. O. (1999). Hyperphosphorylation of replication protein A middle subunit (RPA32) in apoptosis. *Nucl. Acids Res.* **27**, 1499–1504.
12. Gao, H., Cervantes, R. B., Mandell, E. K., Otero, J. H. & Lundblad, V. (2007). RPA-like proteins mediate yeast telomere function. *Nature Struct. Mol. Biol.* **14**, 208–214.
13. Bochkarev, A., Bochkareva, E., Frappier, L. & Edwards, A. M. (1999). The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *EMBO J.* **18**, 4498–4504.
14. Philipova, D., Mullen, J. R., Maniar, H. S., Lu, J., Gu, C. & Brill, S. J. (1996). A hierarchy of SSB protomers in replication protein A. *Genes Dev.* **10**, 2222–2233.
15. Brush, G. S., Clifford, D. M., Marinco, S. M. & Bartrand, A. J. (2001). Replication protein A is sequentially phosphorylated during meiosis. *Nucl. Acids Res.* **29**, 4808–4817.
16. Lee, S. H. & Kim, D. K. (1995). The role od the 34-kDa subunit of human replication protein A in simian virus 40 DNA replication in vitro. *J. Biol. Chem.* **270**, 12801–12807.
17. Henricksen, L. A., Carter, T., Dutta, A. & Wold, M. S. (1996). Phosphorylation of human replication protein A by the DNA-dependent protein kinase is involved in the modulation of DNA replication. *Nucl. Acids Res.* **24**, 3107–3112.
18. Nuss, J. E., Patrick, S. M., Oakley, G. G., Alter, G. M., Robison, J. G., Dixon, K. & Turchi, J. J. (2005). DNA damage induced hyperphosphorylation of replication protein A. 1. Identification of novel sites of phosphorylation in response to DNA damage. *Biochemistry*, **44**, 8428–8437.
19. Binz, S. K., Sheehan, A. M. & Wold, M. S. (2004). Replication protein A phosphorylation and the cellular response to DNA damage. *DNA Rep. (Amst.)*, **3**, 1015–1024.
20. Vassin, V. M., Wold, M. S. & Borowiec, J. A. (2004). Replication protein A (RPA) phosphorylation prevents RPA association with replication centers. *Mol. Cell. Biol.* **24**, 1930–1943.
21. He, Z., Henricksen, L. A., Wold, M. S. & Ingles, C. J. (1995). RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature*, **374**, 566–569.
22. Park, M. S., Ludwig, D. L., Stigger, E. & Lee, S. H. (1996). Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells. *J. Biol. Chem.* **271**, 18996–19000.
23. Nagelhus, T. A., Haug, T., Singh, K., Keshav, K. F., Skorpen, F., Otterlei, M. *et al.* (1997). A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to XPA interacts with the C-terminal part of the 34-kDa subunit of replication protein A. *J. Biol. Chem.* **272**, 6561–6566.
24. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). Crystal structure of the DNA binding domain of replication protein A bound to DNA. *Nature*, **385**, 176–181.
25. 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.
26. Bochkareva, E., Korolev, S., Lees-Miller, S. P. & 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.
27. 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.
28. Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J. *et al.* (2000). Structural basis for the recognition of DNA repair proteins UNG2, XPA and Rad52 by replication factor RPA. *Cell*, **103**, 449–456.
29. Yeates, T. O. (1997). Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344–358.
30. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Ser. Workshop Genome Inform.* **10**, 30–40.
31. Romero, Obradovic & Dunker, K. (1997). Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform. Ser. Workshop Genome Inform.* **8**, 110–124.
32. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Struct. Funct. Genet.* **42**, 38–48.
33. Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J. *et al.* (2000). Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell*, **103**, 449–456.
34. Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Bruce, J. E., Garner, E. C., Brown, C. J. *et al.* (2001). Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.* **10**, 560–571.
35. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucl. Acids Res.* **32**, 1037–1049.
36. Fairman, M. P. & Stillman, B. (1988). Cellular factors required for multiple stages of SV40 DNA replication *in vitro*. *EMBO J.* **7**, 1211–1218.
37. Arunkumar, A. I., Stauffer, M. E., Bochkareva, E., Bochkarev, A. & Chazin, W. J. (2003). Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains. *J. Biol. Chem.* **278**, 41077–41082.
38. Wyka, I. M., Dhar, K., Binz, S. K. & Wold, M. S. (2003). Replication protein A interactions with DNA: differential binding of the core domains and analysis of the DNA interaction surface. *Biochemistry*, **42**, 12909–12918.
39. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature*, **385**, 176–181.
40. Mizuguchi, K., Deane, C. M., Blundell, T. L., Johnson, M. S. & Overington, J. P. (1998). JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **14**, 617–623.
41. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P. *et al.* (2004). Developments in the CCP4 molecular-graphics project. *Acta Crystallog. sect D*, **60**, 2288–2294.
42. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures

Author's personal copy

of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
43. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**.
44. Lu, Z., DiBlasio-Smith, E., Grant, K., Warne, N., LaVallie, E., Collins-Racie, L. *et al.* (1996). Histidine patch thioredoxins: mutant forms of thioredoxin with metal chelating affinity that provide for convenient purifications of thioredoxin fusion proteins. *J. Biol. Chem.* **271**, 5059–5065.
45. Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–767.
46. Cowtan, K. (1998). Modifies phased translation functions and their application to molecular fragment location. *Acta Crystallog. sect. D*, **54**, 750–756.
47. Bourgeois, D. (1999). New processing tools for weak and/or spatially overlapped macromolecular diffraction patterns. *Acta Crystallog. sect. D*, **55**, 1733–1741.
48. Sheldrick, G. M. (1990). Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallog. sect. A*, **46**, 467–473.
49. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallog. sect. D*, **59**, 2023–2030.
50. Terwilliger, T. C. & Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Crystallog. sect. D*, **55**, 849–861.
51. Murshudov, G. N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
52. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
53. Brunger, A. T. & Rice, L. M. (1997). Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol.* **277**, 243–269.
54. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.