INVITED REVIEW

Photoreactive DNA as a Tool to study Replication Protein A Functioning in DNA Replication and Repair

Nadejda I. Rechkunova\textsuperscript{1,2}, Olga I. Lavrik\textsuperscript{1,2*}

\textsuperscript{1}Institute of Chemical Biology and Fundamental Medicine, Lavrentiev av. 8, Novosibirsk, 630090, Russia;
\textsuperscript{2}Department of Natural Sciences, Novosibirsk State University, 1 Pirogov Street, Novosibirsk, 630090, Russia

Corresponding author’s email: lavrik@niboch.nsc.ru (Olga Lavrik)
Abstract

Replication protein A (RPA), eukaryotic single-stranded DNA-binding protein, is a key player in multiple processes of DNA metabolism including DNA replication, recombination and DNA repair. Human RPA composed of subunits of 70-, 32-, and 14-kDa binds ssDNA with high affinity and interacts specifically with multiple proteins. The RPA heterotrimer binds ssDNA in several modes, with occlusion lengths of 8–10, 13–22 and 30 nucleotides corresponding to global, transitional and elongated conformations of protein. Varying the structure of photoreactive DNA, the intermediates of different stages of DNA replication or DNA repair were designed and applied to identify positioning of the RPA subunits on the specific DNA structures. Using this approach, RPA interactions with various types of DNA structures attributed to replication and DNA repair intermediates were examined. This review is dedicated to blessed memory of Prof. Alain Favre who contributed to the development of photoreactive nucleotide derivatives and their application for the study of protein-nucleic acids interactions.
INTRODUCTION

Photocrosslinking is a useful technique to study interaction of protein and protein complexes with nucleic acids. The method of affinity modification with the use of reactive DNA structures is one of the most informative approaches to study dynamic protein–DNA systems, because it allows crosslinking even unstable intermediate complexes which play important roles in the dynamic molecular machines carrying key cellular processes, e.g. DNA replication, transcription and repair [1, 2]. Photoreactive dNMP (NMP) residues can be introduced into DNA in the course of DNA transactions by the activity of DNA (or RNA) polymerases to design photoreactive intermediates of the process under investigation [1, 3]. Alternatively, photoreactive groups can be introduced into DNA via solid phase synthesis of oligonucleotides. The various photoreactive groups were synthetized and introduced into nucleotide moieties to provide variations of the efficiency of photocrosslinking. Wide range of photoreactive base-substituted derivatives of dNTPs and oligonucleotides has been synthesized and used to design photoreactive intermediates of DNA metabolism pathways [1].

Human RPA is a heterotrimeric protein, composed of 70-, 32- and 14-kDa subunits which harbor six oligonucleotide/oligosaccharide binding folds (OB-folds; indicated A through F). The DNA binding OB-folds have been referred as DNA binding domains (DBDs). The functions of this protein are based on its DNA-binding activity and specific protein–protein interactions. The major ssDNA binding activity of RPA is associated with p70 [4]. This subunit contains fore DNA binding domains: two central copies of ssDNA-binding motifs (domains A and B) and a putative zinc finger motif (domain C). The central DNA-binding domains are necessary and sufficient for interactions with ssDNA; however, domain C and part of the N-terminus of the p70 (domain F) are needed for optimal ssDNA binding activity [5, 6]. The fifth DNA binding domain of RPA, domain D, resides in the p32 subunit [7]. RPA binds ssDNA with a defined polarity: in the crystal structure ssDNA binds to the A and B domains in a 5’→3’ direction where the 5’-end associated with domain A and the 3’-end associated with domain B [8]. The third RPA subunit, p14, also contains OB domain (E) [9] which possesses weak DNA-binding activity [10] and is most likely involved in the formation of a RPA heterotrimer [9].
The three RPA subunits form a stable complex; trimerization is mediated by the p70C, p32D, and p14E domains, which together form the trimerization core [11]. Flexible linkers connect all other RPA regions to the trimerization core, providing RPA with high conformational mobility and the capability of using a broad range of DNA sequences as ligands [12].

It is now accepted that the mechanism of ssDNA binding by RPA includes sequential binding events [11]. First, binding initially involves an unstable recognition site of 8–10 nt with the high-affinity DBD A and DBD B domains on the 5′-side of the occluded ssDNA; it is designated ‘compact (or globular) conformation’ or 8–10 nt binding mode. Second, this step is followed by the weaker binding of DBD C, on the 3′-side, leading to an intermediate or ‘elongated contracted’ or ‘transient’ (13–22 nt) binding mode [13, 14]. Finally binding of DBD D on the 3′-side forms a stable ‘elongated extended’ complex characterized by a 30 nt long occluded binding site (30 nt binding mode). Electron microscopy data have shown that conformational changes of RPA accompanied binding process [15, 16]. When ssDNA binds to domains A and B, the effective concentration of DNA is raised so that the lower affinity sites of domains C and D can bind [17]. Recent studies have suggested that the RPA-ssDNA complex is relatively dynamic where not all DBDs are stably bound to the DNA and microscopic dissociation of individual DNA binding domains occurs [18–21]. Thus, although RPA binds to ssDNA with very high affinity [20, 22–24], it can be displaced by DNA binding proteins with much lower DNA binding affinity. This fact is important for RPA functioning in the dynamic cell systems.

The mechanism of RPA interaction with DNA is investigating during several decades by different approaches. This review focuses on advances of photoaffinity labeling technique in detecting the RPA binding modes described above and in proving the polarity of ssDNA binding.

**STRUCTURE AND PHOTOCHEMICAL CHARACTERISTICS OF PHOTOREACTIVE dNTP ANALOGUES, SYNTHESIS OF PHOTOREACTIVE DNA**

The most important characteristic of photoreactive groups used to modify biological molecules is their absorption spectrum, i.e. their ability to be excited by near ultraviolet light, to avoid stimulating the inherent photoreactivity of nucleic acids and proteins, and to avoid degradation and inactivation caused by ultraviolet light. Another property of photoreagents is efficiency of crosslinking to target molecules. Azido group modified nucleotides can be activated at wavelengths >300 nm and produce the highly reactive nitrene which covalently bound to
molecules in close proximity. Therefore, azido derivatives of nucleotides, in particular aryl azido substituted, are widely used for study of the protein-nucleic acid interactions.

A set of base-substituted dNTP analogues containing different aryl azido groups was synthetized and used as the substrates of DNA polymerases to produce photoreactive DNA for following photoaffinity modification of DNA polymerases and other DNA binding proteins, e.g. RPA. Structural formulas, designations and full names of the dNTP derivatives are shown in Fig. 1.

Substituents in the aromatic ring effect on the photochemical characteristics of a reagent and its biochemical features. Substrate properties of dCTP (Fig. 1, a) and dUTP (Fig. 1, b) analogues in primer elongation catalysed by DNA polymerases and efficiency of biopolymer modification were intensively analysed [25–31]. Among these compounds, the dCTP analogue substituted at the exo-amino group in the 4th position of the pyrimidine ring with a FAP group (FAP-dCTP on Fig. 1, a) provided the combination of good substrate properties in DNA synthesis catalysed by DNA polymerases with effective cross-linking to proteins when introduced into DNA [31]. This analogue was used for protein modification both in reconstituted systems and in cellular extracts [32–37].

Aryl azido substituents are large and their attachment to the base via linker changes a size of nucleotide residue and can induce distortion in DNA structure and influence on the protein affinity to DNA. Furthermore, in this case it is unclear whether crosslinked protein locates at the indicated position of DNA or at adjacent site and contacts photoreactive group which is available due the extended linker. This problem may be solved by using DNA bearing zero length crosslinker, such as halopyrimidines (Fig. 2). The substitution of 5-bromo-2′-deoxyuridine (BrdU) for thymidine in DNA has been used for long time to investigate DNA-protein interactions via photocross-linking because excitation can be achieved using ultraviolet light >300 nm [38] and the Van der Waals radius of bromine (1.95 Å) is similar to the radius of a methyl group (2.0 Å), making bromouracil comparable to thymine in size. As a consequence, BrdU-substituted DNAs show specific binding to proteins with affinities similar to those of the non-substituted DNA. Although the iodine atom of 5-iodo-2′-deoxyuridine (IdU) has slightly larger van der Waals radius (2.15 Å) than the corresponding methyl group, it only marginally affects nucleic acid structure and protein binding. Furthermore, IdU possess more intense absorption at longer wavelengths than BrdU that allows reducing unwanted photo-damage from excitation of other light-absorbing
chromophores in the DNA-protein complex. For this reason, IdU has become more popular in experiments on cross-linking studies of DNA-protein interactions. It should be noted that both BrdU and IdU undergo crosslinking reactions only with amino acid side chains that are in close contact to the photoreactive nucleobase and exhibit a strong preference for aromatic side chains [39].

Another zero length chromophore to study nucleic acid-protein and nucleic acid-nucleic acid interactions is 4-thiouridine (4-SU). Its structure is similar to uridine, with only the 4-keto oxygen replaced by a sulfur atom (Fig. 2). In comparing the Van der Waals radii, the sulfur is only 0.45 Å larger than oxygen [40]. This small impact in structure variation and high photoreactivity with amino acid residues and nucleic bases make 4-thiouracil an excellent chromophore to study nucleic acid-protein and nucleic acid-nucleic acid interactions. In addition, the substitution with sulfur at the 4-position centers the main absorption band around 330 nm, compared with uracil’s absorption band centered at 260 nm [40]. The ability to excite 4-SU at wavelengths above 300 nm allows minimizing unwanted additional photochemistry or photodamage to the system. The photoreaction of this moiety in the structure of tRNA was discovered in 1969 by Alain Favre [41, 42]. For more than 40 years, Dr. Favre deeply contributed to the study of 4-SU photochemistry, biochemical properties and its application for analyzes of the structures and functioning of protein–nucleic acid complexes. In particular, this photoreactive probe was used to study interactions between tRNA and aminoacyl-tRNA synthetases [43, 44], protein-nucleic acids and nucleic acids-nucleic acids interactions within human ribosome [45, 46], as well as interaction of RPA with ssDNA [47–49].

MODES OF RPA BINDING TO DNA

Photoreactive DNAs were applied to reveal the arrangement of RPA subunits on different DNA structures. Initially, using partial DNA duplex with the 19 nt single-stranded extension and photoreactive dUMP analogue at the 3’-end of the primer, preferable labeling the p32 subunit was demonstrated [29]. In contrast, only limited crosslinking of p70 and no crosslinking of p14 were observed. Interaction of the p32 subunit of RPA with nascent DNA in replicating SV40 chromosomes has also been observed indicating that the p32 contacts early intermediates produced by DNA polymerase α-primase [50, 51]. Further, using several template-primer systems differing in the size of the single-stranded template tail (31 to 4 nt), it was shown that the pattern of p32 and
p70 RPA subunit labeling, and consequently their interaction with the template-primer junction, is strongly dependent on the template extension length at a particular RPA concentration [13, 47, 52], as well as on the ratio of RPA to DNA template concentration [13]. When template extension was shortened up to 14 nt or less, the p70 subunit of RPA was predominantly photocrosslinked to the 3′-end of the photoreactive primer [13]. A model of changes in the RPA configuration modulated by the length of the template extension in the course of nascent DNA synthesis in the replication fork was suggested (Fig. 3, a). These data approved the changes of RPA structure depending on the length of ssDNA template.

To examine RPA subunit arrangements when bound to the ssDNA in gap structures, two sizes of single-stranded DNA platforms, 9 and 30 nt, flanked with the photoreactive group at the 3′ or 5′ margin were used [53, 54]. The ability of RPA subunits to be crosslinked to the 3′ or 5′ photoreactive DNAs indicates their location in the close proximity to the 3′- or 5′-end of the gap in the RPA–DNA complex. The pattern of RPA subunit crosslinking strongly depended on the location of the photoreactive group and on the gap size (Fig. 3, b). When bound to a 9 nt single-stranded part of gap using the high affinity A and B DNA binding domains of p70, RPA adopts the compact conformation in which the p32 subunit crosslinking to the 3′-end of upstream oligonucleotide is much less in comparison with crosslinking of p70 subunit. In that case p70 can be crosslinked to the both 3′- and 5′-ends of DNA, therefore suggesting that p70 is located near both ends of gapped DNA in globular protein conformation. When RPA binds in the stable elongated conformation, 30 nt of ssDNA is occluded and p70 is located near the 5′-end of oligonucleotide providing downstream orientation of p32 near the 3′-end of the upstream oligonucleotide along the polarity of RPA binding to extended ssDNA platform. The same subunit orientation can be realized when RPA binds DNA in the stable elongated conformation when 30 nt of ssDNA is occluded by RPA in the case of extended DNA duplex (Fig. 3, a). The data demonstrate that p14 is far from 3′- or 5′-ends of the DNA gap or masked by other subunits. In addition because RPA can be crosslinked to the 3′-end of primer under some conditions these studies suggest that p70 is located in close proximity to both ends of the gap, at least transiently. It is interesting that the same modes of RPA binding were demonstrated by using photoreactive aryl azido groups attached via extended linkers to nucleotide moiety or with the zero length 4-SU cross-linker differ in photoreactivity [47].

This article is protected by copyright. All rights reserved
In summary, the pattern of RPA subunit labeling in DNA structures with gap or template extension is determined by the polar RPA binding to ssDNA [55] when p70 subunit initially resides at the end of ssDNA platform and then extends in the direction of the 3′-end. Subunit orientation near the 3′-end of DNA structures is predetermined by the size of ssDNA platform that modulates changes in RPA conformation which appears to be nearly the same with gap and template extension. Polarity of RPA binding in elongated conformation has been suggested to be realized in nucleotide excision repair (NER) for positioning of the excision nucleases XPG and EXCC1-XPF on the DNA [56]. However, it should be noted, that 30 nt binding mode can be realized only in the case of low ratio of RPA protein concentration to DNA. Otherwise, RPA binding modes with (8-10) and (13-22) nt binding sites contribute to cooperative binding several RPA molecules along ssDNA. It is these low-affinity interactions can play important role in dynamic processes of DNA replication and repair [57].

The same DNA structures, namely, partial DNA duplexes with the 3′- or 5′-recessed photoreactive end or DNA duplexes with 30 nt gap containing the photoreactive group at either the 3′ or 5′ terminus, were used to analyze an interaction of replication factor C (RFC) with DNA intermediates of replication or repair. RFC, a multiprotein complex composed of five subunits, RFC140, RFC40, RFC38, RFC37 and RFC36, named according to their molecular masses, is the loader of sliding clamp factor, proliferating cell nuclear antigen (PCNA), which is essential for processive synthesis of DNA by eukaryotic DNA polymerases δ and ε [58–60]. DNA primers containing a photoreactive group at their 5′-end were able to crosslink with the largest RFC subunit, RFC140, on both primer-templates and DNA gap structures, whereas 3′-end photoreactive primers crosslinked only with RFC140 within the DNA gap structure [61]. Addition of RPA to the reaction mixture resulted in the crosslinking of RPA subunits and inhibited crosslinking of RFC140 using 3′- but not 5′-photoreactive primers at the gap. PCNA was neither crosslinked with any DNA used nor influenced on the RFC140 labeling. Crosslinking of RFC140 to primers with 5′-end photoreactive group suggests the interaction of RFC with the 5′-end of a nascent DNA primer during DNA replication. This interaction may be important for the switch of DNA replication from initiation to elongation. A model for the DNA polymerase switch during eukaryotic DNA replication was proposed on the basis of these data [61].

Modification of both the 70- and 32-kDa subunits by photoaffinity labeling indicates that RPA can bind the primer-template junction of partial duplex DNAs by interacting with the 3′-end
of the primer. The following experiments were aimed to investigate conformational changes accompanying the RPA binding to partial DNA duplexes and assigned protein domains playing important roles in such interactions. Limited proteolysis combined with immunoblotting identification of the proteolytic fragments was used to address whether the binding of RPA to partial DNA duplexes containing 5′-protruding ends shows some peculiarities in comparison to the ssDNA which binding leads to a significant conformational changes in heterotrimer [62]. It was shown that the conformation of the RPA bound to partial DNA duplexes significantly differs from that of RPA bound to ssDNA of comparable length and of free RPA in solution [63]. Proteolytic digestion of RPA crosslinked to photoreactive 3′-end of the primer followed with immunoblotting allow determining that domains located in the central part of the RPA32 subunit (amino acids 39–180) and the C-terminal part of the RPA70 subunit (amino acids 432–616) are involved in interactions with the 3′-end of the primer in partial DNA duplexes [63]. These data are consistent with the interpretation that RPA binds the primer–template junction via contacts with the RPA trimerization core. Furthermore, using crosslinking experiments with RPA deletion mutants, p70ABC·p32D·p14, p70AB, p70C·p32D·p14 (trimerization core), and p32D·p14, it was shown that in the absence of two major DBDs, p70A and p70B, the RPA trimerization core was able to recognize correctly the primer–template junction and adopt an orientation similar to that in native RPA [64]. In contrast, the RPA dimerization core, consisting of p32D· and p14, was not detectably crosslinked with such DNA structures. Therefore, p70 seems to be the predominant subunit to bind single-stranded DNA and this interaction positions the p32 subunit to the 3′-end of the primer [52]. Dimeric RPA complex p70·p32 lacking p14 was able to bind single-stranded DNA, but its binding mode and affinity differed from those of the heterotrimeric complex. Moreover, in this complex p32 only minimally recognized the 3′-end of a primer in a primer–template junction [65]. These data speak in favor of key role of the RPA trimerization core in proper orientation of RPA relatively to primer–template junction in the course of DNA replication.

As it was mentioned above, no cross-linking of the small subunit of RPA, p14, to arylazido-containing DNA was observed. However, when 30 nt oligonucleotides containing 4-thiodeoxyuridine in different positions as zero-length photoreagent were used, covalent adducts which can be attributed to p14 crosslinked to such DNA were observed [66]. Indeed, further photoaffinity labelling experiments using 31 nt oligonucleotides with the 4-thiothymine residues introduced at a defined site coupled with the identification of cross-linked targets using specific
antibodies to certain RPA subunits, revealed that in the elongated extended RPA conformation p14 closely contacts the 3'-end positioned nucleotides and yields a covalent adduct with 4-SdT [49]. The proposed scheme of elongated complex was in good agreement with sequential 5′→3′ binding of RPA where the first DBDs A and B located to the 5′-side start to interact with the DNA, followed along the DNA by DBD C and DBD D, and finally DBD E from p14 RPA binds to the DNA progressing to the 30 nt high-affinity binding mode (Fig. 4). Thus, 4-thiothymidine containing DNA can be useful to determine RPA subunits positioning on the certain sites of ssDNA. Furthermore, such DNA probes can be used to trap single-stranded DNA binding proteins in cellular extracts, which provides in combination with mass spectrometry a powerful tool for proteomic investigations [67].

DNA structures containing single-stranded platforms are common intermediates of cellular processes of DNA metabolism and also occur in some specific regions of the genomic DNA. One of the specific types of ssDNA is a part of telomeric DNA which locates at the ends of each chromosome and contains G-rich termini as relatively short single-stranded 3′ overhangs designated G-overhangs [68]. RPA was shown to be present at the telomeric ends of chromosomes with maximum association in the S phase and to play an essential role in telomere maintenance [69]. Further, it was shown under near-physiological in vitro conditions that human RPA is able to bind and unfold G-quadruplex structures formed from a 21 nt human telomeric sequence [48]. Analyses by native gel electrophoresis, cross-linking and fluorescence resonance energy transfer indicated the formation of both 1:1 and 2:1 complexes in which G-quadruplexes are unfolded. In addition, quadruplex opening by RPA is much faster than observed with the complementary DNA, demonstrating that this protein efficiently unfolds G-quartets.

A two-step mechanism accounting for the binding of RPA to G-quadruplexes was proposed [48]. In this model, RPA initially takes advantage of the natural G4 breathing that exposes ssDNA at its extremities. RPA then breaks Hoogsteen hydrogen bonds that stabilize the guanine quartets. Destabilization of a region of the G4 structure makes possible the binding of a second molecule of RPA. It was also demonstrated that unfolding of telomeric G4 by RPA is strongly impaired by small G4-stabilizing ligands [70]. In addition, photo-crosslinking experiments elucidated the polar positioning of the RPA protein subunits along the unfolded G4. RPA1 and RPA2, but not RPA3, interact with the telomeric G4 and they are arranged from 5′ toward 3′. RPA can bind on each side of the G4 but it unwinds the G4 only from 5′ toward 3′, with
the same directionality as for duplex unfolding [71]. The 5' to 3' unfolding directionality may be explained in terms of the 5' to 3' orientation of RPA subunits along single-stranded DNA. Furthermore, RPA was shown to prevent the formation of G-quadruplex structures at lagging-strand telomeres to promote shelterin association and facilitate telomerase action at telomeres [72]. These and other data [73] point to the involvement of RPA in regulation of telomere maintenance.

**RPA POSITIONING ON THE NER DNA INTERMEDIATES**

RPA is one of the inherent participants of the NER process which is responsible for removing from DNA UV-induced pyrimidine dimers and a wide range of the chemically diverse bulky lesions distorting DNA duplex. Repair process follows through sequential stages including damage detection, helix opening, dual incision of the damaged strand 5' and 3' to the lesion, release of the 24-32 nt oligonucleotide, gap filling DNA synthesis, and ligation. In this process, RPA plays an integral role in damage recognition preceding the dual incision of a damaged strand, and then again in post-excision repair synthesis [20, 74, 75]. Moreover, RPA, along with the TFIH, was found to be bound with excised damaged DNA fragments [76].

One of the earliest applications of photoaffinity labeling to study RPA interaction with the NER intermediates was the experiments using model 24 bp oligonucleotides containing a single 1,3-d(GTG)-cisplatin modification and 5-iodo-2'-deoxyuridine (5-IdU) as a crosslinking reagent [77]. The 70 kDa subunit of RPA was crosslinked with high efficiency to cisplatin-modified DNA probes carrying 5-IdU. Crosslinking efficiency depended on the presence of the DNA lesion and 5-IdU positions. Examination of the crosslinking efficiency in dependence on the position of the 5-IdU revealed that the major contact points for p70 located to the 5'-side of the DNA lesion on the damaged strand. The interaction of XPA, the main RPA partner in the NER process, with the same DNA structures was also analyzed. Under the conditions tested XPA alone formed a protein–DNA crosslink only with low efficiency, exhibiting no preference for any 5-IdU-position on the DNA substrate. In samples containing both RPA and XPA it was only RPA that underwent significant crosslinking. The presence of XPA affected neither the specificity nor the crosslinking yield compared to RPA alone.

Further, this approach was extended and applied to analyze RPA and XPA interactions with DNA intermediates during different stages of the NER process. A set of 48-mer oligonucleotides bearing fluorescein residue as the lesion and/or 5I-dU at a definite position as the
photoreactive group was used to create DNA duplex and 15 nt bubble DNA which mimic intermediates of the initial and pre-incision stages of the NER process, respectively [78, 79]. The size of bubble equal to 15 nt is in agreement with the size of bubble opened by TFIIH in the damaged DNA duplex [80]. Crosslinking experiments with both DNA duplex and bubbled DNA display higher yields of DNA-protein adducts of RPA and XPA with undamaged strand. The data on DNA duplex modification disagree with previous results [77] claiming RPA and XPA crosslinks mainly to 5I-dUMP residues in the damaged strand. This difference is likely due to the use of different lengths and/or sequences of DNA duplexes and/or different types of lesions. Each can influence DNA secondary structure, resulting in changes of the DNA-protein binding surface. It should be noted that protein crosslinking to the cisplatin-damaged DNA could occur via direct crosslinking to cisplatin which was shown to act as an effective photoinduced cross-linking agent [81] and could increase the level of crosslinking to damaged strand. RPA crosslinking to undamaged strand was also demonstrated using DNA duplex with cholesterol lesion [82].

The model of RPA and XPA positioning on DNA within the pre-incision NER complex was proposed based on crosslinks combined with the EMSA and nuclease footprinting experiments (Fig. 5). It was assumed [79] that precise positioning RPA bound to undamaged strand and XPA located on the 5'-side from a lesion near to ss/ds junction in the damaged bubble can provide proper orientation of the XPF-ERCC1 and XPG endonucleases which perform dual incision of the damaged DNA strand.

From the NER factors only RPA participates during all stages of the repair process. Using 60-mer DNA duplexes containing a 31 nt flap with the fluorescein residue mimicking the bulky lesion and a 26 or 10 nt gap, RPA and XPA interactions with intermediates of the NER late stages were analyzed [83]. The structure containing a damaged flap and a 26 nt gap imitates the DNA intermediate arising in the NER process after damaged strand cleavage by XPF-ERCC1; DNA with the same flap and a 10 nt gap can be attributed to the intermediate of the subsequent partial gap filling [84]. Crosslinking with 5-I-dUMP containing flap-gap26 DNA revealed that RPA interacts with the ssDNA track in the gap and does not utilize the flap [83]. Based on this study, it may be supposed that, after XPF-ERCC1 incision, RPA remains bound with the undamaged strand in the gap created and can be translocated on the flap when gap is shortened during the re-synthesis step (Fig. 6). It should be noted that RPA was shown to bind effectively long flap within the DNA structure lacking a gap [85].
CONCLUSIONS

Photochemical reactions play an essential role in cellular metabolism. UV irradiation damages genomic DNA and can also induce cross-links within DNA-protein complexes. An ability of nucleic acids to form covalent adducts with amino acid residues located in close proximity was utilized in photoaffinity labeling method for investigation of protein-nucleic acids interactions. This approach was developing during several decades with both natural chromophores (e.g. 4-thiothymine incorporated into the DNA or nucleic bases themselves) and chemically synthesized nucleotide analogues. Identification of the formed covalent bonds can help to define the interface between the DNA and the protein, providing a better understanding of the interactions that occur in the DNA-protein complex. For more than 40 years, Dr. Alain Favre contributed to the development of photoreactive nucleotide derivatives and their application for the study of protein-nucleic acids interactions.

This review summarizes data on the application of photoreactive DNA probes as a tool to analyze the interaction of RPA with the DNA structures mimicking intermediates of replication and repair processes. The obtained data were important to propose molecular mechanisms of protein- nucleic acid interactions in DNA replication, repair and telomere maintenance. Cross-linking technique combined with mass-spectrometry analysis as well as with structural data seems to be very promising approach for the discovery of new proteins interacting with specific DNA structures in cellular extracts of different sources and for deeper understanding of the architecture of multiprotein complexes operating in cellular DNA metabolism.

Acknowledgements The authors are thankful to Dr. Jean Cadet for the invitation to publish this review and to Dr. Yuliya Krasikova for the assistant in the manuscript preparation. The work was supported by the Russian Science Foundation (project #19-14-00107).
References


Nadejda I. Rechkunova obtained her Ph.D. degree in biochemistry at Novosibirsk Institute of Bioorganic Chemistry (now Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, ICBFM SB RAS). Her postdoctoral studies were devoted to development of new approaches for the selective labeling of DNA polymerases and replication protein A by photoreactive DNA intermediates of DNA replication and repair. Now she is a senior researcher in the Laboratory of Bioorganic Chemistry of Enzymes at ICBFM SB RAS. Current research activities of Dr. Rechkunova focus on the compositions and DNA-protein and protein-protein interactions in nucleotide excision repair complexes and their regulation by poly(ADP-ribosyl)ation.
**Prof. Olga I. Lavrik** obtained her Ph.D. degree in bioorganic chemistry at the Institute of Bioorganic Chemistry of the Russian Academy of Sciences (Moscow). She is head of the Laboratory of Bioorganic Chemistry of Enzymes at ICBFM SB RAS, professor in the Department of Molecular Biology at the Novosibirsk State University and Corresponding Member of the Russian Academy of Sciences. Her main research interests focus on the DNA repair mechanisms and their contributions to human health and longevity. She is author/co-author of 450 peer-reviewed articles and book chapters. Several times during 1990-2003 years, she was invited as visiting professor of the University of Pierre and Marie Curie in Paris where she was working with Prof. Alain Favre in his laboratory at the Institute Jacques Monod on the study of protein-nucleic acids interactions using photoreactive tRNA and DNA. This collaboration resulted in more than 20 joint publications.
Figure legends

Fig. 1. Aryl azido substituted analogues of dCTP (a) and dUTP (b). Top – overall structures of the substituted dNTPs, where $R_1$ indicates substitutes attached to the exo-amino group in the 4th position of the cytosine, $R_2$ – substitutes at the 5N-position of the uracyl. Below – structural formulas of $R_1$ and $R_2$, designations and full names of the analogues.

Fig. 2. Structures of zero length photoreactive cross-linker.

Fig. 3. RPA binding to ssDNA platform in the partial DNA duplexes with the 5′-protruding strand (a) and a gap (b). Contacts with the 3′-end of the primer are indicated by the dashed red circle. The RPA structure adopts the globular, transient and elongated conformation according to the size of the ssDNA platform.

Fig. 4. Schematic models of the RPA binding to ssDNA containing 4-SdT substitution in the indicated positions under different binding modes. In the 30 nt binding mode complex, small RPA subunit p14 (DBD E) can be positioned near the 3′-side of ssDNA and contact DNA directly.

Fig. 5. XPA and RPA location on the open DNA duplex provides correct topography of the NER pre-incision complex. RPA bound to undamaged strand and XPA located on the 5′-side from a lesion near to ss/ds junction in the damaged bubble recruit the XPG and XPF-ERCC1 endonucleases, respectively, to proper positions to perform dual incision of the damaged DNA strand.

Fig. 6. Proposed location of the RPA and XPA proteins in the pre- and post-excision complexes. After XPF-ERCC1 incision, RPA remains bound with the undamaged strand in the gap created. RPA can be translocated on the flap when gap is shortened during the re-synthesis step.
**R₁**

- FAB-dCTP - exo-N-[2-(2,3,5,6-tetrafluoro-4-azidobenzoyl)-aminoethyl]-2'-deoxycytidine-5'-triphosphate

- NAB-dCTP - exo-N-[2-(2-nitro-5-azidobenzoyl)-aminoethyl]-2'-deoxycytidine-5'-triphosphate

**R₂**

- FAP-dUTP - 5-{N-{N-(4-azido-2,5-difluoro-3-chlorophiridine-6-yl)-3-aminopropionyl}-trans-3-aminopropenyl-1'}-2'-deoxyuridine-5'-triphosphate

- FAB-dUTP - 5-{N-{(2,3,5,6-tetrafluoro-4-azidobenzoyl)-trans-3-aminopropenyl-1'}-2'-deoxyuridine-5'-triphosphate

- NAB-4-dUTP - 5-{N-(2-nitro-5-azidobenzoyl)-trans-3-aminopropenyl-1'}-2'-deoxyuridine-5'-triphosphate

- NAB-12-dUTP - 5-{N-[N-(2-nitro-5-azidobenzoyl)-7-aminohexanoyl]-trans-3-aminopropenyl-1'}-2'-deoxyuridine-5'-triphosphate
BrdU - 5-bromo-2′-deoxyuridine-5′-monophosphate

IdU - 5-iodo-2′-deoxyuridine-5′-monophosphate

4SU - 4-thio-2′-deoxyuridine-5′-monophosphate

4ST - 4-thio-2′-deoxyuridine-5′-monophosphate
8-10 nt binding mode
globular conformation

13-22 nt binding mode
transient conformation

30 nt binding mode
elongated conformation
8-10 nt binding mode

13-22 nt binding mode

30 nt binding mode
bubble
gap ~ 25-30nt
gap ~ 10nt