Article

3D structure prediction of replication factor C subunits (RFC) and their interactome in *Arabidopsis thaliana*

Mohamed Ragab Abdel Gawwad[1], **Jasmin Šutković**[1], **Emina Zahirović**[1], **Faruk Berat Akcesme**[1], **Betul Akcesme**[1], **Lizhi Zhang**[2]

[1]Genetics and Bioengineering department, International University of Sarajevo, Ilidza, 71220 Bosnia and Herzegovina

[2]Department of Molecular Genetics, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210, USA

E-mail: mragab@ius.edu.ba

Received 17 January 2013; *Accepted 30 January 2013*; *Published online 1 June 2013*

Abstract

DNA stress can causes potentially spontaneous genome damage during DNA replication process. Proteins involved in this process are DNA-dependent ATPases, required for replication and repair. In this study the 3-D structure of RFC protein subunits in *Arabidopsis thaliana*: RFC1, RFC2, RFC3, RFC4 and RFC5 are predicted and confirmed by Ramachadran plot. The amino acid sequences are highly similar to the sequences of the homologous human RFC 140-, 37-, 36-, 40-, and 38 kDa subunits, respectively, and also show amino acid sequence similarity to functionally homologous proteins from *E. coli*. All five subunits show conserved regions characteristic of ATP/GTP-binding proteins and have significant degree of similarity among each other. The segments of conserved amino acid sequences that define a family of related proteins have been identified. RFC1 is identical to CDC44, a gene identified as a cell division cycle gene encoding a protein involved in DNA metabolism. Subcellular localization and interactions of each protein RFC protein subunit is determined. It subsequently became clear that RFC proteins and their interactome have functions in cell cycle regulation and/or DNA replication and repair processes. In addition, AtRFC subunits are controlling the biosynthesis of salicylic and salicylic acid-mediated defense responses in Arabidopsis.

Keywords DNA; RFC protein; 3-D structure; structure prediction; interactome; *Arabidopsis thaliana*.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Arabidopsis thaliana is widely used for studying plant sciences, including genetics, evolution, population genetics, and plant development. Although *A. thaliana* has little direct significance for agriculture, it has several traits that make it a useful model for understanding the genetic-, cellular-, and molecular biology of flowering plants. The small size of its genome makes *A. thaliana* useful for genetic mapping and sequencing with about 157 mega base pairs and five chromosomes. *A. thaliana* has one of the smallest genomes among plants. It was the first plant genome to be sequenced, completed in year 2000 by the Arabidopsis Genome

Initiative. The most up-to-date version of the *A. thaliana* genome is maintained by the Arabidopsis Information Resource (The Arabidopsis Genome Innovative, 2000).

Post-genomic research, such as metabolomics, has also provided useful insights to the metabolism of this species and how environmental perturbations can affect metabolic processes (Alberts et al., 2002). DNA replication depends on the coordinated action of numerous multiprotein complexes (Martinez-Antonio, 2011; Rahman et al., 2013). At the simplest level, it requires an initiator to establish the site of replication initiation, a helicase to unwind DNA, a polymerase to synthesize new DNA, and machinery to process the Okazaki fragments generated during discontinuous synthesis. Much is known about the DNA replication machinery in e.g. *S.cerevisiae* and animal model systems, but relatively little is known about the apparatus in plants. To understand plant DNA replication components it is necessary to combine already published experimental information and new bioinformatics analysis of genomic sequence data and therefore to examine the core of DNA replication machinery in the model plants of Arabidopsis species. In recent years there has been increased interest in plant DNA replication and repair machinery and in using plants as models for understanding the same processes in eukaryotes (Shultz et al., 2007). The DNA replication process represents a source of DNA stress that causes potentially spontaneous genome damage (Takahashi et al., 2010). DNA damage and replication errors might originate from DNA stress provoked either by exogenous and/or endogenous sources. The latter includes the replication process itself that necessitates the cooperation of many different proteins in a highly complex manner. Errors arisen during replication are preferentially repaired through homologous recombination between the replicated sister chromatids that lay in close proximity thanks to cohesion. The process of cohesion establishment is intimately connected with DNA replication. Cohesion depends on acetyltransferase (Eco1/Ctf7), which acetylates two lysine residues on the ATPase head domain of SMC3. Eco1/Ctf7 interacts physically and genetically with the PCNA and the RFC is observed to travel along the DNA with replication fork, suggesting that replication fork progression and sister chromatid cohesion are coupled events (Takahashi et al., 2010; Ashwin et al., 2011).

This model is supported by the observation that cohesion defects are caused by mutations in replisome components, such as DNA polymerase a-binding protein Ctf4, the Chl1 helicase and RFC components, which includes RFC2, RFC3, RFC4, and RFC5 protein subunits. Due to their sessile nature and need of sunlight, plants are particularly exposed to environmental genotoxins, which lead directly or indirectly, via generation of ROS, to DNA lesions including single and double strand breaks. Double strand breaks are particularly critical lesions, because if unrepaired they lead to major karyotypic instability and cell death (Kozak et al., 2009). RFC, a clamp-loader complex consisting of five different subunits, binds DNA at the template-primer junctions and displaces polymerase to terminate DNA primer synthesis. The binding of RFC to DNA creates a loading site for recruiting the DNA sliding clamp PCNA. RFC is an AAA+-type ATPase that requires ATP hydrolysis for opening and closing PCNA around DNA during DNA replication, repair, and recombination. Genetic studies in *A. thaliana* have identified many DNA replication- and repair related proteins are involved in the regulation of transcriptional gene silencing (Liu et al., 2010). Homotrimeric PCNA is considered an essential component in recruiting the general mediators for chromatin imprinting of epigenetic markers (Takashi et al., 2011). RFC can bind to DNA template-primer junction and load the PCNA clamp onto DNA with the assistance of ATP. PCNA loading recruits DNA polymerase to the site of DNA synthesis. Five subunits of RFC were identified as one large subunit RFC140/RFC1 and four small subunits RFC37/RFC2, RFC36/RFC3, RFC40/RFC4 and RFC38/RFC5, and have been found in all eukaryotes. RFC plays an essential role in DNA replication, DNA damage repair and check-point control during cell cycle progression. Recently, three RFC-like complexes (RLCs), namely, Rad24 RLC, Ctf18-RLC and Elg1-RLC, have been identified. Each RLC is made up of the four small subunits of the archetypal RFC, but the large subunit, RFC1, is replaced with an RFC related

protein (Takashi et al., 2011).

Replication factor C (RFC), which is composed of five subunits, is an important factor involved in DNA replication and repair mechanisms. In DNA replication and repair processes, replication factor C (RFC) plays an important role as a clamp loader to load a clamp protein, PCNA, on the DNA strand in an ATP-dependent manner. After loading PCNA at a primer-template junction, DNA polymerase δ and ε bind to the protein-DNA complex and are tethered to the junction through protein-protein interaction. All five RFC subunits are found in all eukaryotes (Cullmann et al., 1995; Bunz et al., 1993; Li and Burgers, 1994; Noskov et al., 1994; Gary and Burgers, 1995; Gray and MacNeill, 2000).

RFC1 is the largest subunit of the RFC complex in Arabidopsis. Mutations in RFC1 lead to developmental defects and earlier flowering (Takashi et al., 2011). RFC1 is an important mediator of transcriptional gene silencing (TGS) DNA replication, DNA repair, homologues replication, and telomere length regulation. TGS controls the expression of transposable elements and of endogenous genes containing promoter repeats, and it is associated with increased DNA methylation.

RFC2 is a cellular component of replication factor C complex, found in *Arabidopsis thaliana* structure and twenty plants more. It is involved in biological ATP catabolic processes and in DNA replication. Molecular function of RFC2 is ATP-, DNA- and nucleotide binding mainly resulting in ATPase activity, DNA clamp loader activity, nucleoside-triphosphatase activity (Vladimir et al., 2008).

RFC3 is one of the small subunits of the RFC complex, originally purified from the HeLa cells, that is essential for the in vitro replication of Simian virus 40. RFC3 and other subunits, compromising the function of the protein complex, and leading to cell proliferation defects in the leaves and roots of Arabidopsis. Partial dysfunction in rfc3-1 leads to smaller plant size due to the reduced number of cells, suggesting defects in replication. Therefore, RFC3 plays an essential role in the process of cell proliferation. Also RFC3 is involved in ATP-, DNA- and nucleotide binding (Majka and Burgers, 2004).

RFC4 is located in DNA replication factor C complex, in nucleolus. It is involved in biological processes such as ATP catabolic process, embryo development ending in seed dormancy, and metabolic process. A molecular function of RFC4 is ATP-, DNA- and nucleotide binding, ATPase activity, DNA clamp loader activity, nucleoside-triphosphatase activity (Majka and Burgers, 2004).

RFC5 is a protein encoded by the RFC5 gene in humans and it is a part of the small subunits of the RFC complex originally. In *Arabidopsis thaliana* it encodes a protein with high homology to the RFC3 of yeast and other eukaryotes. RFC5 biological functions are mainly ATP catabolic process, metabolic process, and negative regulation of defense response. This protein is also involved in the processes of ATP-, DNA- and nucleotide binding, having the molecular functions of ATPase activity, DNA clamp loader activity and nucleoside-triphosphatase activity (Liu et al., 2010).

2 Materials and Methods

2.1 RFC family sequences modeling and multiple sequence alignment

The FASTA sequences of RFC family proteins (RFC1, RFC2, RFC3, RFC4 and RFC5) were obtained from NCBI databases. The Gene ID codes from The Arabidopsis Information Resource (TAIR) database are shown in Table 1.

The multiple sequence alignment searches for RFC proteins were performed with the ClustalW2 program. Default parameters were applied and aligned sequences were executed using this software.

2.2 Phylogenetic tree construction

Phylogenetic tree was constructed using Maximum likelihood method in Robust Phylogenetic Analysis software, where phylogenetic relationships between amino acids of AtRFC subunits were reconstructed and

analyzed (Dereeper et al., 2008).

Table 1 RFC subunits and their gene ID codes from TAIR

RFC proteins	Gene IDs
RFC1	AT5G22010
RFC2	AT1G63160
RFC3	AT1G77470
RFC4	AT1G21690
RFC5	AT5G27740

2.3 3-D structure prediction and conformation

Due to the missing information about the 3-D structure of AtRFC subunits in Protein Data Bank, Pdb files had been generated based on homology moduling using CPH models 3.2 servers (Nielsen et al., 2010). The 3-D structures for all AtRFCs were confirmed by Ramachadran plots using RAMPAGE program (Lovell et al., 2002).

2.4 Domain search and interaction prediction

The protein's genetically mobile domain annotations were obtained using *Pfam* database (Finn et al., 2010). Arabidopsis Interactions Viewer was used for the protein- protein interaction in order to find the partners of AtRFC subunits (Arabidopsis interaction viewer). Subcellular localization for each protein was predicted in WoLF PSORT program (Horton et al., 2007).

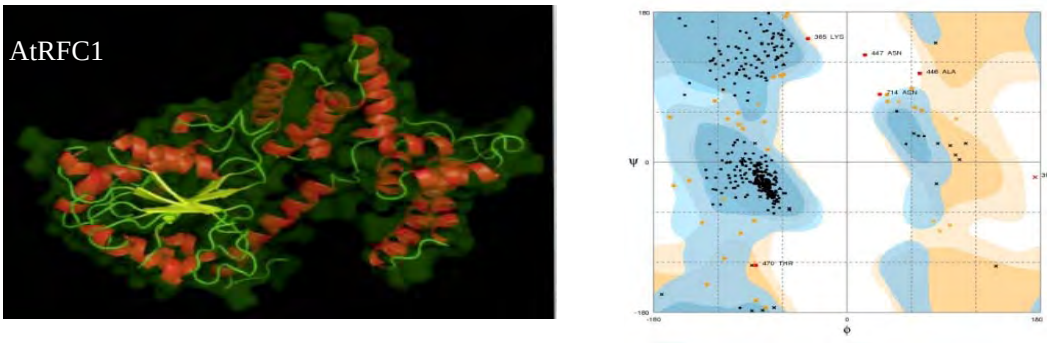
3 Result and Discussion

3.1 3-D structure prediction

The three-dimensional (3-D) structure details of proteins are of major importance in providing insights into their molecular functions. Replication factor C (RFC) is a five-subunit DNA polymerase accessory protein that functions as a structure-specific, DNA-dependent ATPase in *A.thaliana*. 3-D structure of factor C subunits; AtRFC1, AtRFC2, AtRFC3, AtRFC4and AtRFC5 were predicted. Their 3-D structure of the five subunits showed highly structural similarities. Additionally outputs have been confirmed by Ramachadran plots.

All five subunits show conserved regions characteristic of ATP/GTP-binding proteins and also have a significant degree of similarity among each other (Fig. 1).

3-D structure **Ramachadran Plot**



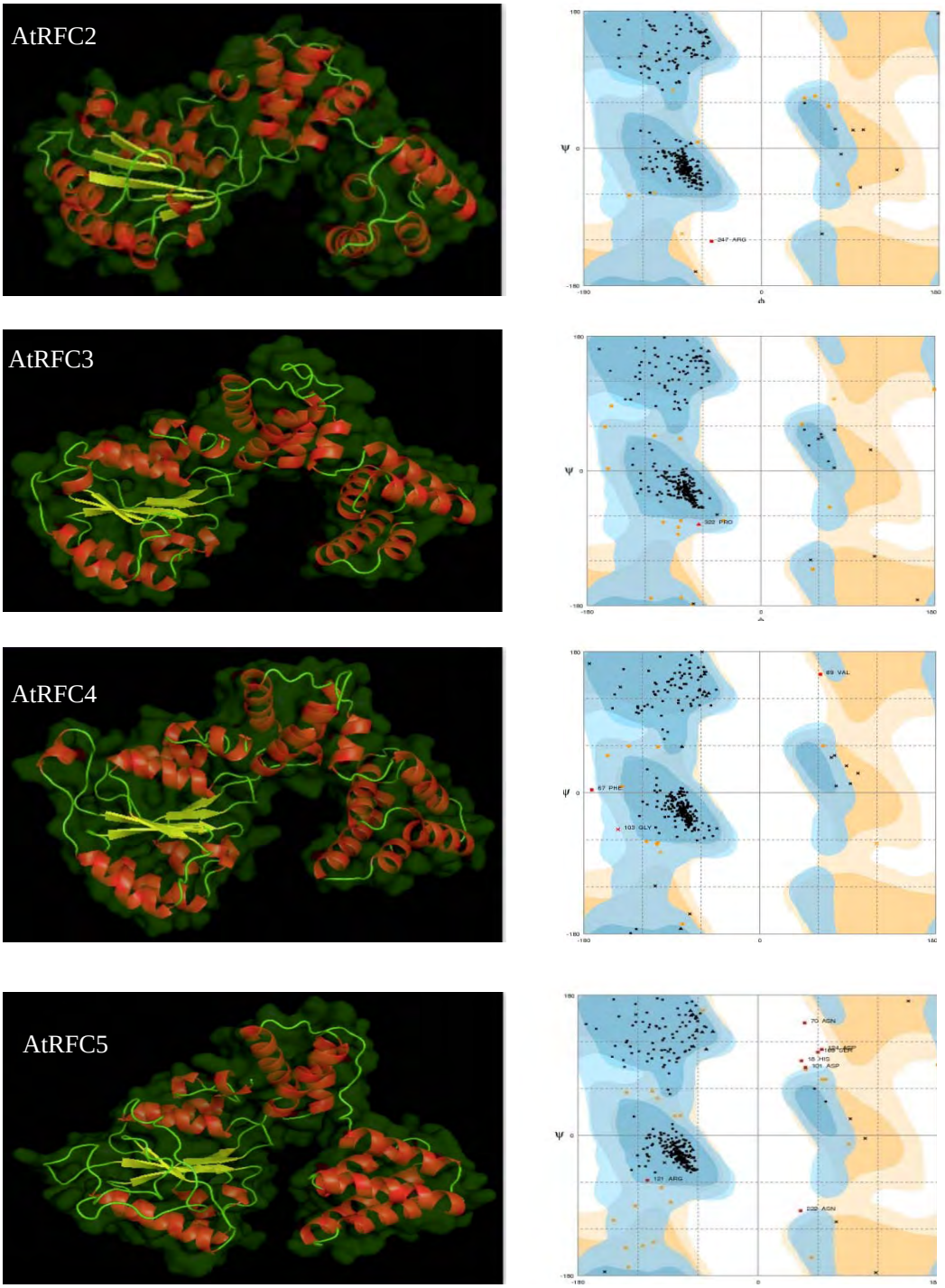


Fig. 1 Predicted 3-D structure of RFC homologous proteins.

3.2 Multiple sequence alignment and phylogenetic tree

Aligned sequences showed substantially significant homology, and constructed phylogenetic tree represents strong relationship between proteins in this family which play their major role as DNA-dependent ATPase in replication and repair systems (Fig. 2).

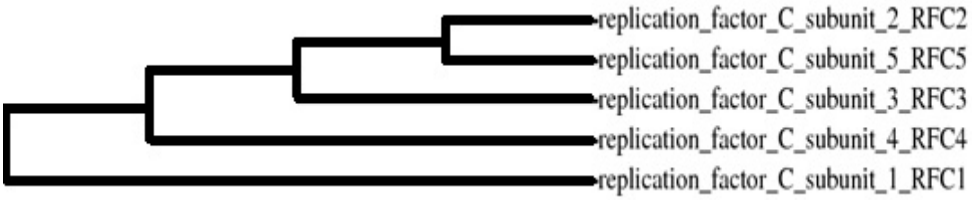


Fig. 2 Phylogenetic tree of AtRFC subunits in *A. Thaliana*.

The protein sequence of RFC1, RFC2, RFC3, RFC4, and RFC5 were obtained from the NCBI sequence database. The amino acid sequences are highly similar to the sequences of the homologous human RFC 140-, 37-, 36-, 40-, and 38 kDa subunits, respectively, and also show amino acid sequence similarity to functionally homologous proteins from *E. coli* and S. *cerevisiae.* The sequence alignment scores are shown in Table 2.

Table 2 ClustaW2 alignment and similarities among AtRFC subunits.

SEQA	NAME	LENGTH	SEQB	NAME	LENGTH	SCORE
1	ATRFC1	956	2	ATRFC2	333	19.0
1	ATRFC1	956	3	ATRFC3	369	21.0
1	ATRFC1	956	4	ATRFC4	332	22.0
1	ATRFC1	956	5	ATRFC5	354	16.0
2	ATRFC2	333	3	ATRFC3	369	35.0
2	ATRFC2	333	4	ATRFC4	332	37.0
2	ATRFC2	333	5	ATRFC5	354	23.0
3	ATRFC3	369	4	ATRFC4	332	34.0
3	ATRFC3	369	5	ATRFC5	354	23.0
4	ATRFC4	332	5	ATRFC5	354	23.0

The phylogenic relationships among the five subunits revealed that AtRFC4 and AtRFC1 subunits had been diverged from the two homologous (AtRFC2, AtRFC5) and AtRFC3 with reserved evolutionary relationships (Fig. 2). This confirms the similarity in function of AtRFC subunits.

The alignment of AtRFC subunits showed important conserved regions; RFC box I (ligase homology) and boxes II (ATP/GTP binding region) and III (Phosphate binding loop), which are conserved in all RFC subunits. The most conserved motif is RFC box III, forming a phosphate-binding loop (P loop, also known as Walker A) with the consensus sequence GxxxxGK(S/T). This loop usually contains additional glycines and prolines and has the consensus sequence PHUUUYGPPGTGKT(S/T), where U stands for a bulky aliphatic residue such as I, L, V or M. Substituting the third Gly for Asp in RFC3-1 mutant may affect the interaction between RFC3 and other subunits. The conserved regions are shown in Fig. 3 (Horton et al., 2007).


AtRFC1 MSDIRKWFMKAHEKGNGSAPKSTSSKAGPVKNAAETAPIKSEQASEDLETADRRKTSKYF 60
AtRFC2 --
AtRFC3 --
AtRFC4 --
AtRFC5 --
AtRFC1 GKDKTKVKDEKEVEAIPAKRKLKTESDDLVKPRPRKVTKVVDDDDDDFDVPISRKTRDTT 120
AtRFC2 --
AtRFC3 --
AtRFC4 --
AtRFC5 --
AtRFC1 PSKKLKSGSGRGIASKTVDNDDDDDGEDKETPLKSAGRGRGGRAAPGASTGGRGRGGGRG 180
AtRFC2 --
AtRFC3 --
AtRFC4 --
AtRFC5 --
AtRFC1 GFMNFGERKDPPHKGEKEVPEGTPDCLAGLTFVISGTLDSLEREEAEDLIKRHGGRITGS 240
AtRFC2 --
AtRFC3 --
AtRFC4 --
AtRFC5 --
AtRFC1 VSKKTTYLLCDEDIGGRKSEKAKELGTKFLTEDGLFDIIRSSKPVKKSLPERSNKGTEKI 300
AtRFC2 --
AtRFC3 --
AtRFC4 --
AtRFC5 --
AtRFC1 CAPPKTSPQKEETRGKPLAKSSPKKVPPAKGKNKIIETSLPWTEKYRPKVPNEIVGNQSL 360
AtRFC2 -MASSSSTSTGDGYNE----------------------------PWVEKYRPSKVVDIVGNEDA 35
AtRFC3 -MTELTSAMDIDVDEIQPRKPINKGKDVVGFGPPPQSKATPWVEKYRPQSLDDVAAHRDI 59
AtRFC4 -MAPVLQSSQ----------------------------------PWVEKYRPKQVKDVAHQEEV 29
AtRFC5 ------------------------------------MLWVDKYRPKSLDKVIVHEDI 21
AtRFC1 VTQLHNWLSHWHDQFGGTGSKGKGKKLNDAGSKKAVLLSGTPGIGKTTSAKLVSQMLGFQ 420
AtRFC2 VSRLQ----------------------VIARDGNMPNLILSGPPGTGKTTSILALAHELLGT 75
AtRFC3 IDTID----------------------RLTNENKLPHLLLYGPPGTGKTSTILAVARKLYGP 99
AtRFC4 VRVLT----------------------NTLQTADCPHMLFYGPPGTGKTTTALAIAHQLFG- 68
AtRFC5 AQKLK----------------------KLVSEQDCPHLLFYGPSGSGKKTLIMALLKQIYGA 61
AtRFC1 AVEVNASDSRGKANSNIAKGIGGSNANSVKELVNNEAMAANFD---------------- 463
AtRFC2 NYKEAVLELNAS----DDRGID-VVRNKIKMFAQKKVT---------------------- 108
AtRFC3 KYRNMILELNAS----DDRGID-VVRQQIQDFASTQSFS---------------------- 133
AtRFC4 -----VLELNAS----DDRGIN-VVRTKIKDFAAAVAGSNHR------------------- 100
AtRFC5 SAEKVKVENRAWKVDAGSRTID-LELTTLSSTNHVELTPSDAGFQDRYIVQEIIKEMAKN 120
AtRFC1 -----RSRHPKTVLIMDEVDGMSAGDRGGVADLIASIKISKIPIICICNDRYSQKLKSLV 518
AtRFC2 -----LPPGRHKVVILDEADSMTSGAQQALRRTIEIYSNSTR--FALACNTSAKIIEPIQ 161
AtRFC3 -----LGKSSVKLVLLDEADAMTKDAQFALRRVIEKYTKSTR--FALIGNHVNKIIPALQ 186
AtRFC4 --QSGYPCPSFKIIILDEADSMTEDAQNALRRTMETYSKVTR--FFFICNYISRIIEPLA 156
AtRFC5 RPIDTKGKKGYKVLVLNEVDKLSREAQHSLRRTMEKYSSSCR--LILCCNSSSKVTEAIK 178
AtRFC1 NYCLPLNYRKPTKQQMAKRLMHIAKAEGLEINEIALEELAERVNGDIRLAVNQLQYMSLS 578
AtRFC2 SRCALVRFSRLSDQQILGRLLVVVAAEKVPYVPEGLEAIIFTADGDMRQALNNLQAT--- 218
AtRFC3 SRCTRFRFAPLDGVHMSQRLKHVIEAERLVVSDCGLAALVRLSNGDMRKALNILQSTHM- 245
AtRFC4 SRCAKFRFKPLSEEVMSNRILHICNEEGLSLDGEALSTLSSISQGDLRRAITYLQSA--- 213
AtRFC5 SRCLNVRINAPSQEEIVKVLEFVAKKESLQLPQGFAARIAEKSNRSLRRAILSLETCR-- 236
AtRFC1 MSVIKYDDIRQRLLSSAKDEDISPFTAVDKLFGYNGGKLRMDERIDLSMSDPDLVPLLIQ 638
AtRFC2 ----------------------------------FSGF-SFVNQENVFKVCDQPHPLHVKNIV 246
AtRFC3 --------------------------ASKEITEEESKQITEEDVYLCTGNPLPKDIEQIS 279
AtRFC4 ------------------------------TRLFGSTITSTDLLNVSGVVPLEVVNKLF 242
AtRFC5 ----------------------------------VQNYPFTGNQVISPMDWEEYVAEIATDMMK 266
AtRFC1 ENYLNYRPSGKDEAKRMDLLARAAESIADGDIINVQIRRYRQWQLSQSCCVASSILPASL 698
AtRFC2 RNVL---------ESKFDIACDGLKQLYD---------------------------------- 266
AtRFC3 HWLL---------NKPFDECYKDVSEIKTR---------------------------------- 300
AtRFC4 TACK---------SGDFDIANKEVDNIVA---------------------------------- 262
AtRFC5 EQSP---------KKLFQVRGKVYELLVN---------------------------------- 286
AtRFC1 LHGSREVLEQGERNFNRFGGWLGKNSTAGKNRRLMEDLHVHVLASRESSAGRETLRVDYL 758
AtRFC2 --
AtRFC3 --
AtRFC4 --
AtRFC5 --
AtRFC1 PLLLSRLTSPLQTLPKDEAVSEVVDFMNSYSISQEDFDTILELGKFKGRENPMEGVPPPV 818
AtRFC2 -----------LGYSPTDIITTLFRIIKN-----------------------YDMAEYL 291
AtRFC3 -----------KGLAIVDIVKEITLFIFK-----------------------IKMPSAV 325
AtRFC4 -----------EGYPASQIINQLFDIVAEA---------------------DSDITDMQ 289
AtRFC5 -----------CIPPEVILKRLLHELLK-----------------------KLDSEL 309
AtRFC1 KAALTKKYNEMNKTRMVRVADMVQLPGVKKAPKKRIAAMLEPTVDSLRDEDGEPLADNEE 878
AtRFC2 KLEFMKETGFAHMRICDGVGSYLQLCGLLAK-LSIVRETAKAP---------------- 333
AtRFC3 RVQLINDLADIEYRLSFGCNDKLQLGAIISTFTHARSIIVGAAK---------------- 369
AtRFC4 KAKICKCLAETDKRLVDGADEYLQLLDVASSTICALSEMAQDF---------------- 332
AtRFC5 KLEVCHWAAYYEHRMRLGQKAIFHIEAFVAKFMSIYKNFLISTFG---------------- 354

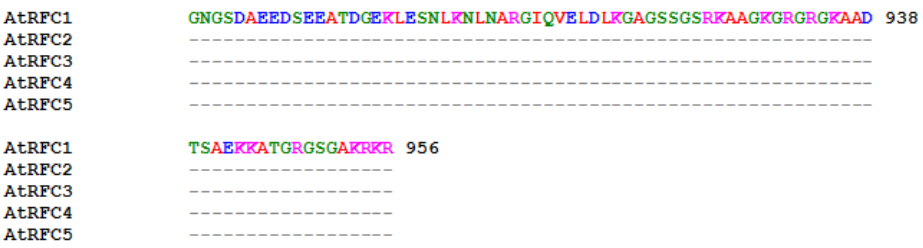


Fig. 3 Alignments of AtRFC subunits protein by ClustalW2 software showing Box I, Box II and Box III respectively.

3.3 Domain search in *Pfam* database

Pfam domain search shows that RFC1 contains significant BRCT domain, which is found predominantly in proteins involved in cell cycle checkpoint functions, responsive to DNA damage reference and one AAA-ATPases domain which is associated with diverse cellular activities. BRCT domain extends from 204 to 282 amino acid residues, while AAA-ATPases domain extends from 392 to 530 residues.

Analysis showed that RFC2, RFC3a, RFC4, RFC5 contain significant AAA domains which extend from 47 to182 residues, 71 to 200 residues, 41 to 167 and 33 to 192 residues respectively.

3.4 Subcellular localization

The Subcellular localizations of RFCs were analyzed with using WoLF PSORT program. Results showed good and significant reliabilities that RFC1 is located in nucleus; RFC2, RFC4 and RFC5 are in cytoplasm whereas RFC3 is in cytoskeleton. TAIR software confirmed that RFC4 localization is in nucleus whereas others protein localization are unknown (Table 3).

Table 3 AtRFC family localization results

Protein	Subcellular localization
AtRFC1	Nucleus
AtRFC2	Cytoplasm
AtRFC3	Cytoskeleton
	Cytoplasm
AtRFC4	Cytoplasm
	Chloroplast
	Nucleus
	Vacuole
	Endoplasmic reticulum
AtRFC5	Cytoplasm
	Nucleus
	Mitochondria

3.5 Interactions

According to data obtained from the Arabidopsis Interactions Viewer we found many interactions belonging to all five subunits of AtRFC in different subcellular localizations, but mainly the partners were localized in

nucleus. The interactome network of AtRFC subunits showed strongly interaction with AtPCNA1, AtPCNA2, CTF18 and RAD17 (Fig. 4). The common partner for RFC1/2/4 is AtPCNA whereas RFC2/3/4/5 interact strongly with CTF18 protein (Table 4). On the other hand we found that all AtRFC subunits interact with at least one of their AtRFC subunits.

Table 4 Interactome of AtRFC subunits

AtRFCs	Interactoms	Interolog Confidence	Interactome functions
RFC1	ATPCNA1	High	Proliferating cellular nuclear antigen 1
RFC1	ATRECQ4A	High	ATP-dependent DNA helicase
RFC1	ATRFC4	High	Rplication factroc C 4 ATPase family
RFC1	ATPCNA2	High	Proliferating cellilar nuclear antigen 2
RFC1	ATMSH2	High	Mismatch repair - MUTS homolog 2
RFC1	ATHDA19	High	Histone deacetylase 1
RFC1	ATRFC3	High	Rplication factroc C 3ATPase family
RFC1	ATRAD17	medium	Radiation sensitive 17
RFC2	ATCTF18	High	P-loop containing nucleoside triphosphate hydrolases
RFC2	ATRFC4	High	Rplication factroc C 4 ATPase family
RFC2	ATRFC5	High	Rplication factroc C 5 ATPase family
RFC2	ATRAD17	High	Radiation sensitive 17
RFC3	ATCTF18	High	P-loop containing nucleoside triphosphate hydrolases
RFC3	ATPCNA1	Medium	Proliferating cellular nuclear antigen 1
RFC3	ATRFC4	High	Rplication factroc C 4 ATPase family
RFC3	ATRFC2	High	Rplication factroc C 2 ATPase family
RFC3	ATPCNA2	High	proliferating cellular nuclear antigen 2
RFC3	AtRFC5	Medium	Rplication factroc C 5 ATPase family
RFC3	ATRAD17	High	Radiation sensitive 17
RFC4	ATCTF18	High	P-loop containing nucleoside triphosphate hydrolases
RFC4	ATPCNA1	High	Proliferating cellular nuclear antigen 1
RFC4	ATHDA19	Medium	Histone deacetylase 1
RFC4	ATPCNA2	High	Proliferating celltlar nuclear antigen 2
RFC4	ATRAD17	High	Radiation sensitive 17
RFC5	ATCTF18	High	P-loop containing nucleoside triphosphate hydrolases
RFC5	ATPCNA1	High	Proliferating cellular nuclear antigen 1
RFC5	ATRFC4	High	Rplication factroc C 4 ATPase family
RFC5	ATRFC1	High	Rplication factroc C 1 ATPase family
RFC5	ATRAD17	High	Radiation sensitive 17

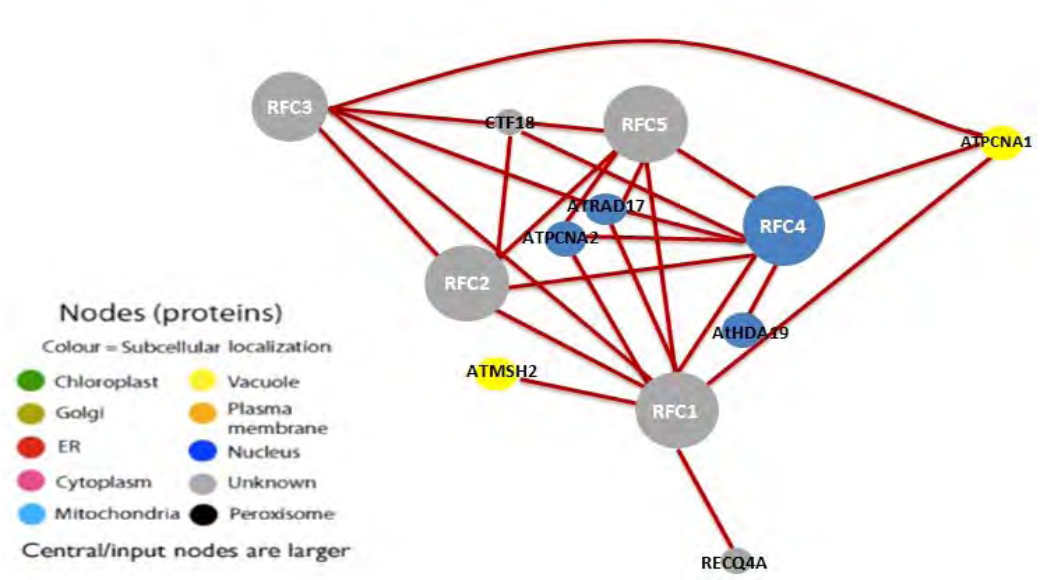


Fig. 4 Interactome of AtRFC subunits as obtained from Arabidopsis Interactions Viewer.

AtRFC1 is identical to CDC44 (Pai et al., 1989), identified as a cell division cycle gene encoding a protein involved in DNA metabolism. Due to their similarity, CDC44and RFC1 interact genetically with the gene encoding proliferating cell nuclear antigen, confirming previous biochemical evidence of their functional interaction in DNA replication. AtRFC1 strongly interacts with AtPCNA1 which is linked to a "sliding clamp" that functions as a mobile platform for the docking of an impressive array of enzymes responsible for the replication and repair of DNA (Yao et al., 2006). AtRFC1 also interacts with AtMSH2 which is involving in the repair of DNA replication errors (Leonard et al., 2003). The network of interactome for AtRFC1 showed an interesting dimerization with AtRFC3 and AtRFC4 respectively. The latter RFC subunits regulate cell proliferation and pathogen resistance in Arabidopsis (Depeiges et al., 2005). The unique property of AtRFC1 is the interaction with RECQ4 (ATP-dependent DNA helicase Q-like 4). The prominent functions of RECQ4 is the processing of double-holliday junctions (dHJs) that occur as intermediates during replication, DNA repair, or recombination and dissolve them in a manner which prevents deleterious crossover recombination (Johnson-Schlitz and Engels, 2006; Wu et al., 2005; Wu et al., 2006). Furthermore, RFC1 strongly correlates with HDA19, required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in Arabidopsis. Loss of HDA19 activity increases the salicylic acid content , expressinf the genes required for accumulation of salicylic acid as well as *pathogenesis related* (*PR*) genes, resulting in enhanced resistance to *Pseudomonas syringae* (Choi et al., 2012).

AtRFC2 highly interacts with AtRFC4 and AtRFC5, which regulates cell proliferation and pathogen resistance in Arabidopsis (Noskov et al., 2008). AtRFC1, AtRFC2 and AtRFC4 interacts with CTF18, which is playing an important role in ATP binding, ATP catabolic process, ATPase activity, metabolic process, nucleoside-triphosphatase activity, nucleotide binding, replication fork and sister chromatid cohesion (Ghaemmaghami et al., 2003; Geisler-Lee et al., 2007).

The interactome partners of AtRFC3 are somehow similar to that of AtRFC1 and AtRFC2, with addition of AtRAD17, which encodes a homolog to yeast RAD17. AtRAD17 interacts with all RFC sububits , involved in the regulation of DNA damage repair and homologous recombination (Heitzeberg et al., 2004).

The possible proteins interaction with AtRFC4 are numerous containing three AtRFC subunits; AtRFC1,

AtRFC2 and AtRFC3. This conforms the strong interaction of AtRFC4 with AtPCNA1 and AtPCNA2, playing regulatory roles in ATP catabolic process, ATPase activity and the machinery enzymes responsible for the replication and repair of DNA (Ghaemmaghami et al., 2003; Geisler-Lee et al, 2007).

AtRFC5 interaction with AtRFC1/2/3 and AtRFC4 confirms the connectivity to CTF18, AtPCNA1 and AtRAD17. AtRFC5 interacts with with all RFC subunits, opening the door for the RFC coordination mechanism for regulation of DNA repair, DNA replication and DNA irradiation protection.

Measurements of the intracellular levels of the RFC subunits have been determined in a global analysis of protein expression in yeast (Geisler-Lee et al, 2007). The current Bioinformatics study confirms that subcomplexes of RFC, RFC2/3/4/5 and RFC2/5 (Yao et al., 2006) are capable of unloading PCNA clamps. Furthermore, due to the strong interaction of all RFC subunits interact with radiation sensitive 17 protein (RAD17), a notable new function could be assigned RFC family, which is the regulation of DNA damage repair. Due to the strong interaction of all RFC subunits interact with radiation sensitive 17 protein (RAD17), a notable new function could be assigned to our proteins family, which is the regulation of DNA damage repair.

4 Conclusion

Replication factor C (RFC) is a multi protein complex consisting of one large and four small subunits. RFC subunits have an associated ATPase activity that is stimulated by the binding of RFC to DNA and is further stimulated by proliferating cell nuclear antigen (PCNA). The interactome of all five AtRFC subunits assigned new functions to this subfamily, especially for DNA UV-protection, metabolic processes and controlling the biosynthesis of salicylic and salicylic acid-mediated defense responses in Arabidopsis.

Abbreviations

RFC: replication factor C
RFC1: Replication factor C subunit 1 protein
RFC2: Replication factor C subunit 2 protein
RFC3: Replication factor C subunit 3 protein
RFC4: Replication factor C subunit 4 protein
RFC5: Replication factor C subunit 5 protein
TAIR: Arabidopsis Information Resource
PCNA: the proliferating cell nuclear antigen
SMC3: structural maintenance of chromosome
ROS: reactive oxidative species
SSB: single strand breaks
DSB: double strand breaks
RLC: RFC-like complexes
TGS: Transcriptional gene silencing
RAD17:Radio Sensitive 17
CTF18:P loop containing nucleotide phosphate
HR: Homologous recombination
ClustalW2: multiple sequence alignment program, version 2
MSA: Multiple Sequence Alignment
Pymol: program to obtain 3D structure of the target proteins
SMART: Simple Modular Architecture Research Tool

WoLF PSORT: Subcellular localization prediction for each protein
PDB: Protein data bank

Acknowledgment
Authors thank the board of trustees, International University of Sarajevo for financial and moral support to realize this work.

References

Alberts B, Johnson A, Lewis J. et al. 2002. DNA replication, repair and recombination. In: Molecular Biology of the Cell (4th ed.). Garland Science, New York, USA

Arabidopsis interaction viewer. http://bar.utoronto.ca/interactions/cgi bin/arabidopsis_interactions_viewer.cgi

Bunz F, Kobayashiand R, Stillman B. 1993. cDNAs encoding the large subunit of human replication factor C. Proceedings of the National Academy of Sciences of USA, 90: 11014-11018

Choi SM, Song HR, Han SK, et al. 2012. HDA19 is required for the repression of salicylic acid biosynthesis and salicylic acid-mediated defense responses in Arabidopsis. Plant Journal, 71(1): 135-146

Cullmann G, Fien K, Kobayashi R, et al. 1995. Characterization of the five replication factor C genes of *Saccharomyces cerevisiae*. Molecular and Cellular Biology, 15(9): 4661-4671

Depeiges A, Farget S, Degroote F, et al. 2005. A new transgene assay to study microsatellite instability in wild-type and mismatch-repair defective plant progenies. Plant Science, 168: 939-947

Dereeper A, Guignon V, Blanc G, et al. 2008. Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Research, 36: 465-469

Elizabeth AG, Michael AA, David R, et al. 1994. CDC44: a putative nucleotide-binding protein required for cell cycle progression that has homology to subunits of replication factor C. Molecular and Cellular Biology, 14(1): 255-267

Finn RD, Mistry J, Tate J, et al. 2010. The Pfam protein families database: Bateman. Nucleic Acids Research, 38: 211-222

Ganpudi AL, Schroeder DF. 2011. UV Damaged DNA Repair & Tolerance in Plants. In: Selected Topics in DNA Repair (Chen CC, ed.). Intechopen, USA

Gary SL, Burgers MJ. 1995. Identification of the fifth subunit of Saccharomyces cerevisiae replication factor C. Nucleic Acids Research, 23(24): 4986-4991

Geisler-Lee J, O'Toole N, Ammar R, et al. 2007. A predicted interactome for Arabidopsis. Plant Physiology, 145: 317-329

Ghaemmaghami S, Huh WK, Bower K, et al. 2003. Reduced polymorphism in domains involved in protein-protein interactions. Nature, 425(6959): 737-741

Gray FC, MacNeill SA. 2000. The Schizosaccharomyces pombe rfc3_ gene encodes a homologue of the human hRFC36 and *S. cerevisiae* Rfc3 subunits of replication factor C. Current Genetics, 37: 159-167

Heitzeberg F, Chen IP, Hartung F, et al. 2004. The Rad17 homologue of Arabidopsis is involved in the regulation of DNA damage repair and homologous recombination. Plant Journal, 38(6): 954-968

Horton P, Park KJ, Obayashi T, et al. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Research, 35: 585-587

Johnson-Schlitz D, Engels WR. 2006. Template disruptions and failure of double Holliday junction dissolution during double-strand break repair in *Drosophila BLM* mutants. Proceedings of the National Academy of Sciences of USA, 103(45): 16840–16845

Kozak J, West CE, White C, et al. 2009. Rapid repair of DNA double strand breaks in Arabidopsis thaliana is dependent on proteins involved in chromosome structure maintenance. DNA Repair, 8(3): 413-419

Leonard JM, Bollmann SR, Hays JB. 2003. Reduction of stability of *Arabidopsis* genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function. Plant Physiology, 133: 328-338

Li X, Burgers PMJ. 1994. Cloning and characterization of the essential Saccharornyces cerevisiue RFC4 gene encoding the 37 kDa subunit of replication factor C. Journal of Biological Chemistry, 269: 21880-21884

Liu Q, Wang J, Miki D, et al. 2010. DNA replication factor C1 mediates genomic Stability and transcriptional gene silencing in Arabidopsis. Plant Cell, 22(7): 2336-2352

Lovell SC, Davis IW, Arendall III WB, et al. 2002. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 50: 437-450

Majka J, Burgers PM. 2004. The PCNA-RFC families of DNA clamps and clamp loaders. Progress in Nucleic Acid Research and Molecular Biology, 78: 227-260

Martinez-Antonio A. 2011. *Escherichia coli* transcriptional regulatory network. Network Biology, 1(1): 21-33

Mossi R, Hübscher U. 1998. Clamping down on clamps and clamp loaders – the eukaryotic replication factor C. Europe Journal of Biochemistry, 254: 209-216

Nielsen M, Lundegaard C, Lund O, et al. 2010. CPHmodels-3.0--Remote homology modeling using structure guided sequence profiles. Nucleic Acids Research, 38: 578-581

Noskov V, et al. 1994. The RFC2 gene encoding a subunit of replication factor C of *Saccharomyces cerevisiae*. Nucleic Acids Research, 22(9): 1527-1535

Noskov VN, Araki H, Sugino A. 2008. The RFC2 gene, encoding the third-largest subunit of the replication factor C complex, is required for an S-phase checkpoint in *Saccharomyces cerevisiae*. Molecular and Cellular Biology, 18(8): 4914-4923

Pai EF, Kabsch W, Krengel U, et al. 1989. Structure of the guanine-nucleotide-binding domain of the Ha-rasoncogene product p21 in the triphosphate conformation. Nature, 341: 209-214

Rahman KMT, Islam MdF, Banik RS, et al. 2013. Changes in protein interaction networks between normal and cancer conditions: Total chaos or ordered disorder? Network Biology, 3(1): 15-28

Shultz RW, Tatineni VM, Hanley-Bowdoin L, et al. 2007. Genome-wide analysis of the core DNA Replication machinery in the higher plants Arabidopsis and rice. Plant Physiology, 144(4): 1697-1714

Takahashi N, et al. 2010. The MCM-binding protein ETG1 aids sister chromatid cohesion required for postreplicative homologous recombination repair. PLoS Genetics, 6(1): 1-13

Takashi Y, Kobayashi Y, Tanaka K, et al. 2011. Arabidopsis replication protein A 70a is required for DNA damage response and telomere length homeostasis. Plant Cell Physiology, 50(11): 1965-1976

The Arabidopsis Genome Innovative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, 408: 796-815

Wu L Bachrati CZ, Ou J, et al. 2006. BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. Proceedings of the National Academy of Sciences of USA, 103(45): 4068-4073

Wu L, Chan KL, Ralf C, et al. 2005. The HRDC domain of BLM is required for the dissolution of double Holliday junctions. EMBO Journal, 24(14): 2679-2687

Xia ST, Xiao LT, Gannon P, et al. 2010. RFC3 regulates cell proliferation and pathogen resistance in Arabidopsis. Plant Signaling and Behavior, 5(2): 168-170

Yao NY, Johnson A, Bowman G, et al. 2006. Mechanism of PCNA clamp opening by RFC. Journal of Biological Chemistry, 281(25): 17528-17539