Study of the recognition of G-quadruplex DNA by human ORC protein

Afaf Sobhi Mohamed Mahmoud Eladl

DNA replication starts from multiple chromosomal loci called replication origins. Origin recognition complex (ORC) binds to a replication origin in eukaryotic DNA and recruit other replication factors. ORC composed of six subunits (ORC1-6) is highly conserved in all eukaryotes and plays important role in the initiation of DNA replication. It is well-known that the ORC of *Saccharomyces cerevisiae* recognizes an origin through sequence-specifically binding to autonomously replicating sequences. However, human ORC (hORC) binds to a replication origin without sequence specificity and how hORC recognizes the origin remains unknown.

Previous genome-wide studies revealed that guanine (G)-rich sequences, forming G-quadruplex (G4) structure, are present in most replication origins of fly, mouse, and human cells. So far, it was revealed that the deletion of G-rich sequences causes functional impairment of replication origins and the insertion of G-rich sequences creates a new replication origin. These results suggested that the G4 structure plays a critical role in the initiation of DNA replication.

Previously, we found that the region comprising residues 413–511 of human ORC subunit 1, hORC1$^{413–511}$, is responsible for binding to G-rich DNAs, which form a G4 structure in the absence of hORC1$^{413–511}$. Here, we investigated the interaction of hORC1$^{413-511}$ with various DNAs derived from human *c-myc* promoter and telomere region, having G4 structures. mtPu22 and mtPu19, were revealed to fold only into a parallel-type G4 structure while teloDNA was reported to be the (3+1)-type G4 structure.

In this study, we showed the binding affinity to G4 relative to double-stranded DNAs (dsDNAs) using fluorescence anisotropy (FA). Then, we demonstrated the conformational changes of hORC1$^{413–511}$ and G4-DNA in response to their interacting binding by circular dichroism (CD) and nuclear magnetic resonance (NMR). Moreover, the binding sites for that interaction had been identified.

In chapter 1, firstly, a general introduction about ORC protein and its structure were given. Then, the diverse physiological and pathological roles of hORC1 were introduced. Finally, the background and the aim of this study were described.

In chapter 2, bindings of hORC1$^{413–511}$ to G-rich DNAs were characterized by monitoring the change in the FA of fluorescein (FAM)-labeled G-rich DNAs upon titration with hORC1$^{413–511}$. Binding curves indicated that hORC1$^{413–511}$ can bind to all the studied G4-DNAs stronger than its binding to dsDNA.We further identified the necessity of the G4 structure for the binding. Data showed that G4 structure of DNA is critical for strong binding of hORC1$^{413–511}$.Then, the recorded CD spectra of G-rich DNAs showed that G- rich DNAs retain the G4 structure even after binding with hORC1$^{413–511}$ and the binding of hORC1$^{413–511}$ does not unfold the G4 structure. These results were confirmed by the retention of the G4 structure in each complex using NMR-detected hydrogen-deuterium (H/D) exchange experiments. By the analysis of NMR chemical shift perturbation (CSP), we revealed that the external G-tetrad planes of the G4 structures are the primary binding sites for hORC1$^{413–511}$.

In chapter 3, first, we showed the resonance assignment and data deposition which was indicated by backbone $^{1}H_{N}$, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and $^{13}C'$ chemical shifts of hORC1$^{413-511}$. Then, the structural characteristics of the hORC1$^{413-511}$ in its free were investigated. CD and NMR studies indicated that hORC1$^{413-511}$ is intrinsically disordered containing a short helical region in the free form.

The program TALOS+ was used to predict the secondary structure of hORC1$^{413-511}$ from the assigned backbone $^{1}H_{N}$, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and $^{13}C'$ chemical shifts. The correlation observed between the predicted secondary structure by NMR data suggests confidence in the

obtained resonance assignment. Finally, the predicted three dimensional model of hORC1 for alpha fold structure was shown.

In chapter 4, the structural characteristics of the hORC1$^{413\text{-}511}$ in its complexed forms with G4-DNAs were investigated. NMR studies indicated that hORC1$^{413\text{-}511}$ retains the same structural properties upon complex formation. The CSP analysis indicated that the basic residues, arginine and lysine, and the polar residues, serine and threonine, are involved in the G4-DNA binding. Interestingly, the results revealed that hORC1$^{413\text{-}511}$ binds to both parallel- and 3+1-type G4-DNAs using the same residues, thereby in the same manner.

In chapter 5, a general conclusion as to the research background, results outcomes, and importance of the study was given.

In this study we demonstrated the conformational change of hORC1$^{413–511}$ upon binding G4-DNAs sequences. The identified mode through which hORC1$^{413–511}$ interacts with the guanine residues located in the external G-tetrads may rationalize the fold-independent interaction. For G4 binding, we assumed a stacking interaction between the guanine bases and an aromatic ring of F511, which is an only aromatic residue of hORC1$^{413–511}$ and/or the guanidium groups of the arginine residues of hORC1$^{413–511}$.

For hORC1$^{413–511}$ binding, we revealed that both electrostatic interaction and hydrogen bond formation are involved. Our studies suggested that hORC1 uses its intrinsically disordered G4-binding region to play a crucial role in the recognition of the parallel-type and 3+1-type G4 structures at the replication origins.

In conclusion, our results indicate a possible role of various G4 structures in replication origins recognition by hORC1$^{413–511}$, which is a hot topic of study due its direct implication in the cancer diseases.