SUMMARY

Gelsolin is a six domain actin modulatory protein which performs actin depolymerization, nucleation and capping function. In available literature, gelsolin is not known to bind any DNA or RNA, and thus regulate any activity. This thesis details how an aptamer was evolved rationally and tweaked to tightly and specifically bind full-length gelsolin. As of now, activation of gelsolin is known to be governed by calcium, low pH, temperature and is deactivated by PIP_2 [11, 24, 30, 114]. Gelsolin is known to exist in three different isoforms namely cytoplasmic, plasma and gelsolin-3 which arise from alternate splicing of same gene. The physiological relevance of gelsolin is well documented in literature, the cytoplasmic and gelsolin-3 are responsible for actin cytoskeleton remodelling, where as plasma gelsolin found in blood plasma plays a very important role in maintaining the viscosity of blood. Plasma gelsolin plays an integral role in Extracellular Actin Scavenging System (EASS) which is responsible for clearance of actin released from the dead and decaying cells [15]. This released actin is captured by circulating gelsolin resulting in gelsolin-actin complex which in-turn binds to vitamin D-binding protein and the resultant complex is cleared from circulation in liver. Under healthy conditions, the levels of gelsolin are constantly replenished, but in cases of severe trauma or injury there is an increased influx of actin from the damaged cells which rapidly binds to gelsolin and is cleared from the system [125]. This leads to a sudden fall in gelsolin levels, this observation was also made in ICU patients where it was observed that decreases in levels of gelsolin below 25% of circulating value leads to certain death of the patient [16]. Though the criticality of gelsolin is well documented, it has been observed that there is discrepancy in the gelsolin estimation [16]. This discrepancy observed is not limited to the technique used, but also the levels of gelsolin in normal people. Thus, there is a dire need of a reliable procedure for accurate, affordable and accelerated estimation kit.

SUMMARY

In the current study, we rationally screened and identified the aptamers capable of binding to recombinant human gelsolin through SELEX method **(Figure 3.1)**. Out of 10^{18}, four unique sequences were identified namely L26F, L24F, L16F and 10.10R which were capable of binding to Ca^{2+} activated gelsolin protein exclusively [105]. These sequences were 75-77 bases long and their sequences were revealed through DNA sequencing **(Table 3.1)**. The uniqueness of the sequences was ascertained by BLAST search with genomic nucleotide database where no significant hits were observed in any of the organism. The ultimate aim of this study is to establish a reliable technique for quantification of gelsolin, thus, these sequences were subjected to microtiter binding assays, where we could observe that out of all the screened aptamers, L26F displayed best performance and binding **(Figures 3.3 A and B)**. This performance of L26F was not limited to individual concentration of aptamers but also across the range of concentrations. When the microtiter assay was performed with saturating concentrations of gelsolin it was observed that L26F, L24F and L16F followed a similar pattern and reached saturation at a concentration of ~6 pmoles of aptamer per well while the binding of 10.10R was comparatively lower and reached saturation at much higher concentration, ~12 pmoles per well of immobilized aptamer **(Figure 3.3 C)**. Furthermore, the assays were performed using truncated versions of gelsolin where we observed that aptamers can bind to G1-G3 and G4-G6 minimal versions but with comparatively lower affinity than full length gelsolin; while no detectable binding was observed in case of bonsai version 28-161,G1 domain **(Figure 3.3 C)**. These results strongly suggested that the binding interface of gelsolin and aptamers lies in the G2-G6 domain of gelsolin protein. This observation was further corroborated by the series of *in silico* experiments, where ESP and cleft analyses of full length activated gelsolin revealed three clefts as ideal sites of aptamers binding. These three clefts identified were C1 (g3-g4 linker), C2 (G2-G3

domain) and C3 (G4-G6), these are named as cleft volume decreasing from C1 to C3 **(Figure 3.5)**. Hence, using a bi-pronged approach of *in vitro* and *in silico* analysis, we established that G2-G6 domains are critical to aptamer binding owing to the presence of highly positively charged clefts. An interesting by-product of this study was that using this aptamers we could successfully purify gelsolin from unadulterated blood plasma, the yield of gelsolin recovered was as high as 80% versus the 20-40% yield obtained from the conventional techniques **(Figure 3.6)**.

Once it was established that aptamers bind strongly to gelsolin molecule, we started to explore the reason behind this high binding affinity. Taking the *in silico* approach first we attempted modelling of aptamers sequence using the mFold server which showed presence of unique hairpin stem or loop structures arising from the back bone of the ssDNA **(Figure 3.7)**. However, this information was not sufficient and needed validation for which large amount of stable aptamer-gelsolin complexes were required. For this, we screened for osmolytes which could stabilize or enhance the complex formation, for this we undertook a literature survey and selected various osmolytes and performed microtitre binding assays in their presence. In this case, we observed that the sucrose had highest effect on stability of the protein-aptamer complex, as evident from the 1.6 fold high absorbance observed as compared to control **(Figure 3.8 A)**. In case of positive or promoting osmolytes, sucrose shows better performance followed by TMAO, betaine and trehalose. The other osmolytes screened were found to be disruptive in nature as evident from relatively lower absorbance in assays, these osmolytes included glycerol, sorbitol, glycine, mannitol, KCl and NaCl. These experiments were done in range of concentration (10 to 200 mM) and their effect was nearly independent of concentration used **(Figure 3.8)**.

The aptamers and gelsolin binding was efficient and strong, this lead to numerous unanswered questions about the conformation of aptamers binding to gelsolin, the stoichoiometry of binding, as well as the shape changes in gelsolin as a result of this interaction. We attempted to answer these questions using various biophysical techniques. First in these experiments were BLI experiments, through which we estimated the strength of binding of gelsolin to aptamers. The observations made here were in accordance with those made in mictotitre binding assays. Here also, L26F outperformed other aptamers with lowest K_D ~ 5.4 nM followed by L24F and L16F (~ 6 and 8 nM, respectively), the weakest binding was shown by 10.10R (~9 nM) **(Figure 4.1, Table 4.1)**. In case of the N- and C- terminal halves of gelsolin, a decrement in K_D values was observed as high as 4-6 and 30-70 folds, respectively for all the aptamers **(Table 4.1)**. These observations further strengthened the observation that G2-G6 domains are central to binding with the presence of critical g3-g4 linker for binding of gelsolin to aptamers with high affinity.

Till now, it is clear that there is a strong binding between aptamers and gelsolin *but physico-chemical factors responsible for this strong binding remained unclear? Was it dependent on the sequence or the structure or both attained by the aptamer(s)*? During literature survey, we came across numerous reports of aptamers capable of forming G-quadruplexes in solution. G-quadruplex are stable structures of nucleic acids which are formed due to guanidine tetrads capable of forming one, two or four strands through Hoogsteen base pairs and are stabilized by cations [126-129]. G-quaruplex formation can be predicted through two web servers, G4IPDB and QGRS mapper, both of these servers showed that L26F, L24F and L16F are capable of forming G-quadruplexes but 10.10R is incapable in doing so [118, 130]. *In vitro*, G-quadruplexes were identified by the presence of characteristic positive peaks at 260 and 210 nm and a negative peak at 240 nm in a CD

spectra [131]. As predicted by the *in silico* analyses, we observed that L26F, L24F and L16F form parallel G-quadruplexes but 10.10R does not show that property **(Figure 4.2,** left panel). Moreover, these peaks were further pronounced in presence of K^+ ions as reported earlier supporting the role of K+ ions in formation of G-quadruplexes [126]. This observation was also confirmed by the ^1H NMR spectra collected for the aptamers L26F and 10.10R where we saw the characteristic chemical shift (δ) for Hoogsteen base pairs arising between 10.6 and 12 ppm in case of L26F but not in 10.10R **(Figure 4.2,** right panel). This G-quadruplex structure of the aptamers was maintained even post complex formation with gelsolin as observed through the differential CD experiments **(Figure 4.3)**.

G-quadruplex formation raised a question that what is important for aptamers-gelsolin interaction, is it sequence, structure or both that resulted in this high affinity interaction? The aptamers we screened contain a core sequence with two sequences flanking on either sides, thus we generated mutants by cutting half (y_aptamer) and full flanking regions (x_aptamer) of each aptamer, and finally by mutating middle guanidine by adenine residues of each triplet G **(Table 4.3)**. The CD experiments showed that the core sequence was capable of forming G-quadruplexes even in absence of flanking regions but in case of G→A mutants no G-quadruplex formation was observed **(Figure 4.4,** left panel). However, in case of microtitre binding assays significant differences were observed, for the sequences with no flanking region only 20-25% of binding was observed, where as sequences with partial sequence nearly 60% activity was observed and in case of aptamers with G→A mutation no binding was observed **(Figure 4.4,** right panel). For this experiment we chose TBA as a reference, an aptamers capable of forming parallel G-quadruplex but not binding to gelsolin. This experiment helped us to conclude that though the core sequence is responsible for the highly selective binding of aptamers

to gelsolin, it is the flanking region that provides the high affinity and act as an anchoring agent. So far, we have reasoned the high affinity of the aptamers for gelsolin, but still we need purified and stable aptamer gelsolin complex for structural studies. So, we used size exclusion chromatography to purify the stable complex, here we got a shifted peak towards higher molecular weight for complex. Indirect calculations of molecular weights from the elution volume suggested a 1:1 stoichiometry of aptamer-gelsolin complex which further supports the results of equimolar stochiometry in BLI experiments. Here, it is important to mention that in BLI either of the partners was immobilized whereas in SEC the interaction was studied in solution; thus, the molecular weight determined in SEC confirmed that the aptamer gelsolin complex is a 1:1 complex (**Table 4.4**).

CD data collected for understanding the G-quadruplex did not just provide valuable insight into the mechanism of gelsolin-aptamer interaction but also provided guidelines for *in silico* modelling of tertiary structures of the aptamers. Therefore, template based modelling of the GQ forming core sequence of the aptamers was done using Chimera 1.8.1 visualization software which employs amber99bsc01FF force field, while the flanking regions were modelled using 3D dart server (**Figure 4.6**, upper panel). Both these models were then ligated using the Chimera 1.8.1 software to obtain the model for full length aptamers (**Figure 4.6**, lower panel). Though these models were obtained for the aptamers, they still needed validation which was provided by SAXS data collected for gelsolin and aptamers in isolation as well as in complex.

In order to get structural/shape insight, SAXS data was collected at Bio-SAXS beamline ESRF, Grenoble, France, under identical conditions using HPLC purified aptamers and SEC purified gelsolin and gelsolin aptamers complexes. The data was reduced, processed and analysed as described previously. The intensity plot showed the scattering species to be pre dominantly monomer in solution, the R_G and D_{max} calculated

for gelsolin in presence and absence of Ca^{2+} ions was in agreement with the data reported earlier (**Figure 4.7A**)[19]. Interestingly in case of Kratky plots we observed that the Ca^{2+} activated gelsolin and aptamers are inherently disordered proteins but the complex is a globular in nature (**Figure 4.7C**). Using the DATMOW program, the molecular weights of the scattering species were calculated; the molecular weight observed here was in concurrence with the estimation made through SEC and BLI further confirming the 1:1 binding stoichiometry of the complex **(Table 4.4** and **4.5)**.

The SAXS data strongly suggested that the gelsolin and aptamer bind to one another such that there are no large scale changes in the dimensions of Ca^{2+} activated gelsolin and aptamer bound gelsolin. This was evident from the fact that the D_{max} of Ca^{2+} activated gelsolin was 15.3 nm which increased marginally to 17.6±0.37 nm upon binding with aptamers **(Table 4.5)**. This value was significantly lower than 30 nm D_{max} which would have been observed if the binding was end to end which indicated that gelsolin aptamer are closely bound or coiled around each another. In order to visualize the solution shape of complex, we performed the *ab initio* modelling of SAXS data. Total 10 models were generated and averaged, the resultant global shape was chosen for representation, the shapes for gelsolin in presence and absence of Ca^{2+} was as observed as reported earlier [1]. In case of aptamer-gelsolin complex, we could observe a more rigid and dense shape against the sparse shape observed for Ca^{2+} activated gelsolin in isolation (**Figure 4.9**). For the validation of the observed shapes we superimposed the SAXS volumes with the structure of complexes obtained through *in silico* docking. The docked complexes were generated through NP dock server for all the aptamers individually binding to gelsolin in 1:1 stoichiometry, total 3 energy minimized structures were obtained. Of these three structures the structures which showed the best fitting with the SAXS volume was chosen for representation (**Figure 4.10**). The SAXS data validated the

in silico structure obtained from docking, confirming their presence in solution as well **(Figure 4.10)**. These docking structures ultimately confirmed that the binding interface for aptamers is within G2-G6 domains with g3-g4 linker or C2 cleft central for establishing the gelsolin aptamer complex. This observation was reflected even in the activity assay performed for gelsolin bound to aptamers, where neither the polymerisation or depolymerisation activity of pyrene labelled actin was observed, these two functions are the native functions of gelsolin spread across the G2-G6 domains of the protein **(Figure 4.11)**. In both biophysical and biochemical data, it was observed that G1 domain is excluded in gelsolin aptamer interaction. The exclusive binding to the g3-g4 linker of gelsolin strengthens the fact that aptamer bind specifically to gelsolin and not to any other member of the gelsolin superfamily as this linker is present exclusively in gelsolin but not in any other members of gelsolin superfamily.

This overall study is unique where we have attempted to characterize SELEX screened aptamers using various biochemical and biophysical techniques. This work can prove to be a model study or pave a way for other researchers to characterize and visualize the binding partners (aptamer-protein) without using any cumbersome or high resolution techniques. The final goal of this study was to apply the information in hand to compose an affordable, accurate and accelerated kit for gelsolin estimation which we could develop successfully. Considering the indicator to be role of plasma gelsolin levels of expectant mothers and the duration of gestation, clinical studies were done first in collaboration of Dr. Rashmi Bagga PGI, Chandigarh. Subsequently, the trials have been done at expanded locations in Rajasthan, Madhya Pradesh and Gujarat with the help of an agreement with Oniosome Healthcare limited, Mohali. The data analysis indicates that plasma gelsolin levels increase substantially during the transition of first to second trimester of human gestation. Failure to notice this increase in the samples studied

reflected in preterm birth of the babies. More studies and prototype finalization steps are underway to commercially launch a kit based on this thesis. We hope to save some lives at birth via our work.

Figure 5.1: *The affordable, accurate and accelerated (AAA) plasma gelsolin estimation kit to be marketed as 'Kompal', is shown in the figure above.*