CHAPTER 6

Unveiling the Protein-Protein interactions between GRB2 and LMTK3 that induce integrin β_1 during breast cancer progression and metastasis:

an *in silico* study

Unveiling the Protein-Protein interactions between GRB2 and LMTK3 that induce integrin β_1 during breast cancer progression and metastasis: an *in silico* study

6.1. Abstract

Recently, LMTK3 found to bind with growth factor receptor-bound protein 2 (GRB2) and accelerates the binding of GRB2 with son of sevenless (SOS), which in turn activate RAS and Cell division control protein 42 homolog (CDC42) protein, CDC42 then stimulates the binding of serum response factors (SRF) to integrin promoter, leading to the transcriptional activation of integrin β1 that ultimately results in breast cancer progression. Therefore, it becomes crucial to understand the interactions between LMTK3 and GRB2 at molecular level to control the breast cancer progression. Our objective is to study the Protein-Protein interactions between GRB2 and LMTK3 (133-411 amino acids) domain. We have docked GRB2 with LMTK3 domain using ClusPro server and studied the interactions profile using PDBsum server. Using Molecular dynamics simulation we analyzed the stability and total binding free energy (BFE) of the complex. Online servers (KFC, PredHS, Robetta, DrugScorePPI server) were used to identify probable hotspots at GRB2-LMTK3 interface**.** Total BFE of the complex was found to be -51.86 kcal mol^{-1} that substantiates the stability of the complex. The hotspot residues in GRB2 (Chain A) and LMTK3 (Chain B) were Arg21A, Tyr160A, Arg215A, Asn216A, Tyr143B, Arg167B, Tyr297B, Arg302B, Trp304B. Identification of interacting residues and hotspot residues at GRB2-LMTK3 interface may aid in designing the potential inhibitors to disrupt their interactions, thereby controlling the breast cancer progression and invasion.

6.2. Introduction

Apart from ERα positive breast cancer, a recent report has highlighted the association of LMTK3 in motility and invasion of triple negative breast cancer through growth factor receptor-bound protein 2 (GRB2) mediated induction of integrins [9]. Integrins are composed of two chains α and β, wherein the binding of integrin (α_5, β_1) to the collagen and fibronectin, the components of extracellular matrix (ECM) causes the motility and

invasion of cancer cells. Numerous studies have suggested the abundance of altered α_5 and β_1 integrins subunits to be often associated with tumourogenesis and metastasis which is accountable for the increase in disease progression and decreases the patient survival rate [100-104]. GRB2 is an adaptor protein which is expressed ubiquitously in healthy conditions and overexpressed in tissue samples and breast cancer cell lines [105, 331]. GRB2 has been reported to directly associate with receptor tyrosine kinase (RTK), activating downstream RAS guanosine triphosphosphatase (GTPase) and other extracellular signals that regulate kinase and mitogen-activated protein kinase (ERK/MAPK) [106, 107, 108] Recent study reveals the interaction between LMTK3 with GRB2 to enhance the binding of GRB2 with son of sevenless (SOS) that subsequently leads to the activation of RAS and Cell division control protein 42 homolog (CDC42). CDC42, in turn, stimulates the serum response factors (SRF) activity which then binds to integrin promoter, thus inducing the transcriptional activation of integrin [9]. Considering these events, understanding the interaction between GRB2 and LMTK3 at the molecular level is helpful in providing the insights to control this signaling pathway.

The characterization of the protein-protein interaction (PPI) sites has become an essential step towards identifying drug targets in order to design potential drugs to obstruct the protein-protein interactions that bind two or more proteins together [332-337]. The specificity of drug action may be achieved better in PPIs, rather than targeting active sites or co-factor binding sites as active sites are well conserved in evolution [338, 339]. Amino acid residues present at PPI interact with each other, where some of these residues contribute highly to stabilizing energy of the protein-protein complex, provide specificity at their binding sites [258], and thus these residues are termed as hots pots. Identifying these hot spot residues within the protein-protein interfaces can help us to understand protein-protein interactions better and may also to modulate protein-protein binding [264]. In literature, the hot spot residues are typically defined as residues when mutated to alanine show an increase in the binding free energy of at least 4.0 kcal mol^{-1} [257]. However, we see that this threshold value of 4.0 kcal mol^{-1} is not strictly followed. Some authors have defined the hot spots are those upon alanine

mutation results in significant increase in binding free energy (ΔΔG) of 1.5 kcal mol^{-1}, but others have considered the ΔΔG binding increase of at least 2.0 kcal mol^{-1} [263, 340, 255, 256]. Conversely, Null-spots (NS) corresponds to residues with ΔΔG binding are lower than 2.0 kcal mol^{-1} when mutated to alanine and null-spots exist in the surrounding regions of the hotspots and protect them from solvent exposure [257]**.** Hot spot residues exist in clusters and are well conserved and more buried in comparison to other interface residues in the protein-protein complex [258, 259,260, 261, 157]. Tyr, Arg and Trp amino acids have a greater tendency in being a hotspot [257], while Leu, Thr, Ser and Val are less likely to act as a hotspot. Similarly, Asp and Asn have been observed as hotspots more frequently than Glu and Gln [263,257]. Identification of hotspots proves helpful in studying protein dimer and also aid in the identification of probable binding sites for other binding partners [341]. Therefore identifying the hotspots residues at GRB2-LMTK3 interfaces can be helpful in better understanding the protein-protein interactions and may also help us to modulate interacting interface area.

Experimental Alanine scanning mutagenesis (ASM) have been used extensively to identify hotspot residues at protein-protein interfaces [342]. However, this method is time-consuming and expensive. So, we have used computational methods which are freely available web-based services including KFC [265**],** PredHS [266**],** Robetta [267,268] and $DrugScore^{PPI}$ server [269] to determine the probable hotspots at the GRB2-LMTK3 interface. Predicting the hot spot residues using a single method might give inaccurate results, so these four methods were used to improve the accuracy of the result.

In the present study, the structure of GRB2 (PDB ID: 1gri) taken from Protein Data Bank [343,344] was docked using ClusPro server [230] to the model structure of the functional kinase domain of LMTK3 (133-411 amino acid) obtained from I-TASSER [99]. Then the GRB2-LMTK3 complex was subjected to Molecular Dynamics (MD) simulation. MD simulation is considered to be one of the feasible tools to obtain the dynamic information in protein-protein interfaces as protein-protein interactions are dynamic in nature and adopt different conformations. MD simulation allows the transient pockets and buried hot spot residues to emerge on the protein surfaces and

these transient areas and hotspots could be targeted with small molecules [345-350]. From MD trajectory the highly populated lowest energy conformer obtained from RMS clustering was selected for protein-protein interaction study using PDBsum server [252] and also to identify hotspots across GRB2-LMTK3 interface. The conformers obtained from MD simulations were then used to calculate the total binding free energy contribution using molecular mechanics-generalized born surface area (MM-GBSA) method [191, 186, 351, 188]. To calculate the total binding free energy we used MM-PBSA.py script [192] in AMBER12 software tool [122]. We found GRB2-LMTK3 complex to be significantly stable from its calculated negative total binding free energy.

Our findings provide significant insights into the probable interface area, bonded and non-bonded interactions, interface residues and hotspot residues across GRB2-LMTK3 complex. This information may be helpful in targeting the interacting interface across the GRB2-LMTK3 complex in order to control the breast cancer motility and invasion.

6.3. Materials and Method

6.3.1. Modelling and validation of protein structures

The x-ray determined structure of GRB2 (PDB ID: 1gri) was taken from Protein Data Bank. In the GRB2 structure, the coordinates of amino acid residue in the range 28-33 were missing. Using RaptorX [352-354] software we predicted the missing residue coordinates and then joined the sequence with GRB2 (PDB ID: 1gri) structure by using Matchmaker tools, followed by copy/combine function under Model Panel command of UCSF Chimera package alpha v.1.12 [180]. The amino acid sequence of human LMTK3 (1460 amino acids, accession number: Q96Q04) was retrieved from the UniProtKB Swiss-Prot database [296]. The protein kinase domain of LMTK3 with 279 amino acids (133-411) identified as a functional kinase domain referred from UniProtKB database was chosen as our desired sequence for the further study. The prediction of the 3-D structure of LMTK3 domain and the validation of the generated structure has been discussed in Chapter 4 under section 4.4.1.

6.3.2. Protein-Protein Docking of GRB2 and LMTK3 domain

Using ClusPro, a web-based molecular docking server, we docked GRB2 with the modelled structure of $LMTK3_{133-411}$domain and obtained GRB2-LMTK3 complex model. ClusPro is considered to be the best docking server according to the most recent three Critical Assessment of PRediction of Interactions (CAPRI) evaluation meetings, in 2009 [355]**,** 2013 [356] and 2016 [357]. ClusPro server uses the direct docking method for two interacting proteins. ClusPro server carries out three computational steps based on (i) rigid-body docking by sampling billions of conformations; (ii) clustering of 1000 lowest energy structures based on root-mean-square deviation (RMSD), in order to find the largest clusters that represent the most likely models of the complex; and (iii) refining the selected structures using energy minimization [230]. Finally, the server generated ten complex models defined by centers of highly populated clusters of low-energy docked structure. The highly populated complex with lowest binding energy was subjected to MD simulation.

6.3.3. Molecular dynamics simulations of GRB2-LMTK3 complex

Molecular dynamics simulations have been performed on GRB2-LMTK3 complex using the Particle Mesh Ewald Molecular Dynamics (PMEMD) [305] module of AMBER12 software package [122]. GRB2-LMTK3 complex system was simulated with AMBER ff99SB force field [306] of the AMBER12 software package. The complex system was neutralized with 7 Na+ ions. Then the complex structure was solvated using TIP3P [167] water box with a distance of 10 Å around the solute. The initial minimization was done holding the restraints (harmonic constraints with a force constant of 30 kcal $mol^{-1}Å^2$) over the solute for 500 steps using steepest descent algorithm followed by another 500 steps with a conjugate gradient process. The second minimization was carried out without any restraints for another 2000 steps. Then heating dynamics was applied to the GRB2-LMTK3 complexes with the gradual increase in the temperature from 0 to 300 K at constant volume (NVT). Then, the entire system was optimized and equilibrated using standard equilibration protocol at constant pressure (NPT). All simulations were carried out using periodic boundary conditions under isothermal and isobaric conditions (T = 300 K; P = 1 atm). For controlling the

temperature Berendsen thermostat [166] was used. The shake [165] algorithm was used to restrain all the bonds at the time step of 2 fs.

All the MD trajectories of GRB2-LMTK3 system were analyzed using PTRAJ algorithm [358]. For visualization of the protein structures UCSF Chimera package alpha v.1.12 [180] and VMD v.1.9.3 [179] has been used. The stability of GRB2-LMTK3 complex was analyzed by measuring the Root Mean Square Deviation (RMSD). Hydrogen bond analysis was also examined for GRB2-LMTK3 complex on all possible hydrogen donors (HD) and hydrogen acceptors (HA). The criterions were set for hydrogen bond occupancy, bond distance, and the bond angle formed (HA–H–HD) between hydrogen donor (HD) and hydrogen acceptor (HA) atoms.

6.3.4 Hot spot residue identification at GRB2-LMTK3 interface

From MD trajectory, the highly populated lowest energy conformer of GRB2-LMTK3 was selected to study the interaction profile using PDBsum server, and the same conformer was used to identify the probable interface hot spot residues at GRB2-LMTK3 interface. Hot spot residues at GRB2-LMTK3 interface have been identified using different computational methods which are freely available online servers including KFC2, PredHS, Robetta, and DrugScore^{PPI} server. KFC2 server is a machine learning based tool that utilizes *in silico* alanine scanning mutagenesis, considering hydrogen bonds, atomic contacts and residue sizes for hot spot identification [256]. PredHS server is a structure-based hot spot prediction method which predicts hot spot residues using algorithms based on structural neighborhoods (Euclidian and Voronoi), and then selects optimal features using random forest and sequential backward elimination algorithms [266]. In addition, the predicted hotspot residues were further confirmed with computational alanine scanning mutagenesis using Robetta and DrugScore^{PPI} server.

6.3.5. Binding free energy calculation

The ensemble conformations of GRB2-LMTK3 complex were generated using molecular dynamics simulation, from which GRB2 and LMTK3 structures were extracted to calculate the binding free energy using MM-GBSA method. MM-GBSA

analysis was performed with the MM-PBSA.py script in AMBER12. The total binding free energy from MM-GBSA calculation incorporates explicit solvation model with the calculations of electrostatic contribution to the solvation and the non-polar contribution to estimate the binding free energy $\Delta G_{bind.}$ The overall equations of binding free energy calculations are described in **Chapter 3**, section: 3.5.1

6.4. Results and Discussions

6.4.1. Protein-protein interaction at GRB2-LMTK3 interface

From ClusPro server, we selected the highly populated clusters of low-energy docked complex with the binding energy value of -940 kcal mol^{-1}. The resultant complex structure was subjected to MD simulation. From MD simulation, the highly populated lowest energy conformer of GRB2-LMTK3 complex structure was identified and the same was submitted to PDBsum server to study the intermolecular interactions. From the GRB2-LMTK3 complex we see that both the SH3 and SH2 domain of GRB2 interacted with $LMTK3_{133-411}$ domain as shown in **Figure 6.1**. As earlier reports have also suggested that the SRC Homology 3 (SH3) and SH2 domains are directly associated with receptor tyrosine kinase [106-108]. We also observed that the SH2 domain of GRB2 interacting with the C-lobe of the LMTK3 (**Figure 6.2**). In case of well-studied kinases such as abl and src, the SH2 domain interact with the catalytic C-lobe of the kinase domain [359,360]. These reports support our interaction study between GRB2 and LMTK3. We also noticed that the activation loop (284-313 res) of LMTK3 [361] interacting with GRB2 as shown in **Figure 6.3.** In literature it has mentioned that the activation loop of kinases can mimic the protein substrate by folding and interacting with the substrate binding site. In our study, activation loop of LMTK3 found to be interacting with GRB2 as its substrate. Activation loop of the kinases is a flexible sub-region of the activation segment and consists of 20–35 amino acid residues that start after the DFG motif and end before the APE, ALE or SPE sequence [362, 71].

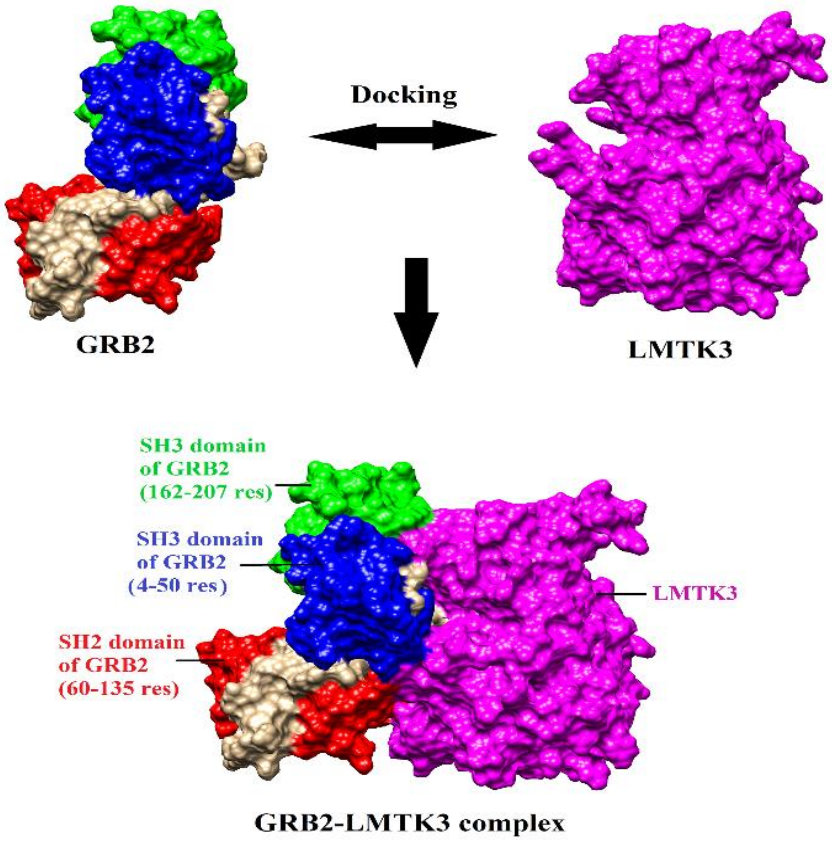

Figure 6.1. *GRB2-LMTK3 docked complex from ClusPro*

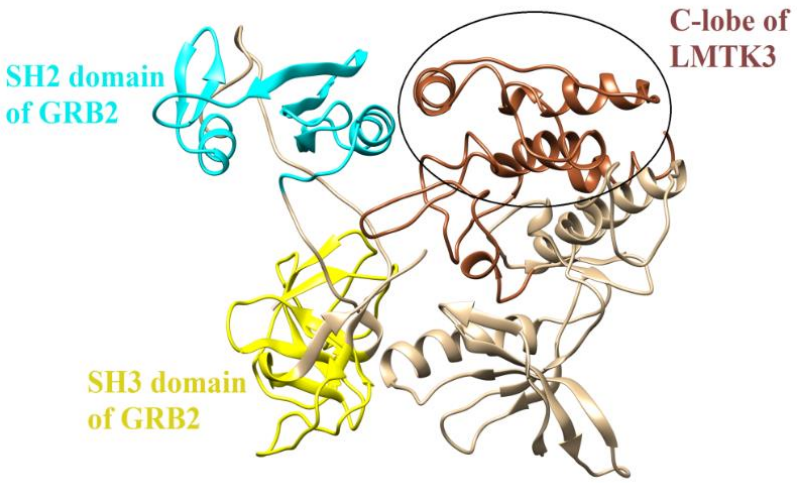

Figure 6.2. *Representation of GRB2-LMTK3 complex, where SH2 domain is interacting with the C-lobe of LMTK3 domain*

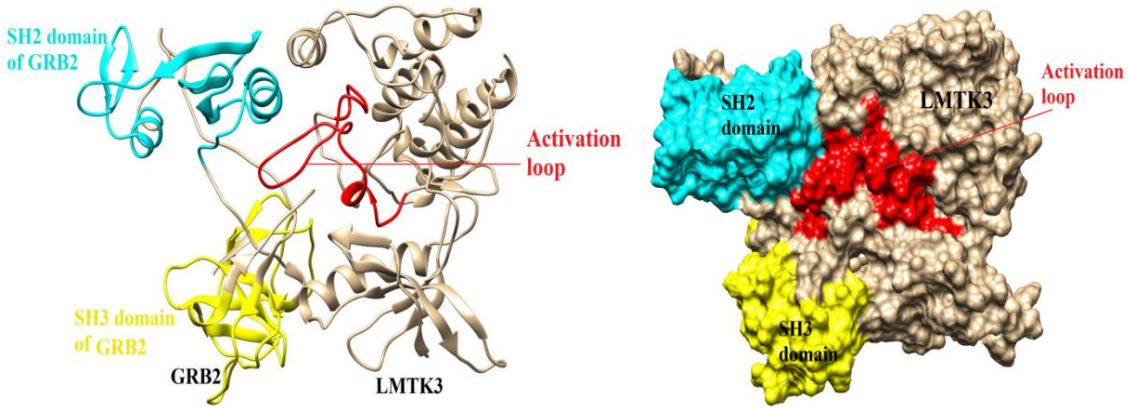

Figure 6.3. *Activation loop (in red) of LMTK3 domain interacting with the GRB2*

From the PDBsum results we see that, in GRB2-LMTK3 complex the total number of probable interface residues to be 32 (in GRB2) and 30 (in LMTK3 domain) as shown in **Figure 6.4** and the interface area for each protein involved in the interaction was observed to be 1766 $Å^2$ (in GRB2) and 1814 $Å^2$ (in LMTK3 domain). The interface contact area of GRB2 and LMTK3 are within the standard size of the protein-protein interaction interfaces that is 1200–2000 A^2 [363]. The interaction profile was summarized in **Table 6.1. Figure 6.5** depicts all the interacting residues at GRB2-LMTK3 interface, residues that are involved in the formation of hydrogen bonds and salt bridges and the residues involved in the formation of non-bonded contacts shown in Table **2** and Table **3**.

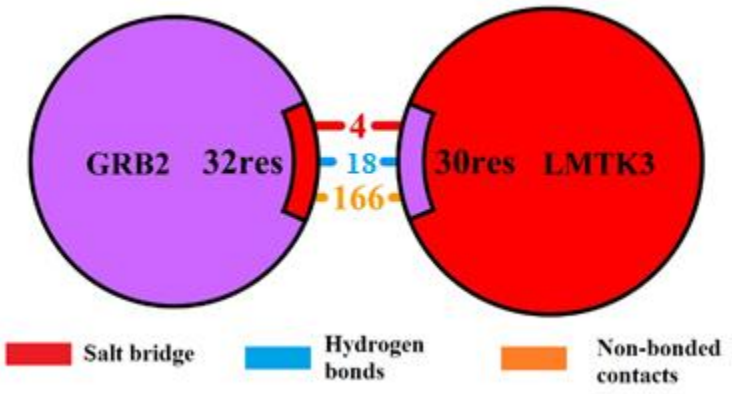

Figure 6.4. *Number of interface residues, bonded and non-bonded contacts involved in GRB2-LMTK3 interface*

Table 6.1. Summary of interface statistics for GRB2-LMTK3 complex determined from PDBsum server

GRB2-LMTK3 complex obtained from ClusPro server					
Chains	**No. of interface residues**	**Interface area ($Å^2$)**	**No. of salt bridges**	**No. of hydrogen bonds**	**No. of non-bonded contacts**
GRB2	32	1766	4	18	166
LMTK3	30	1814			

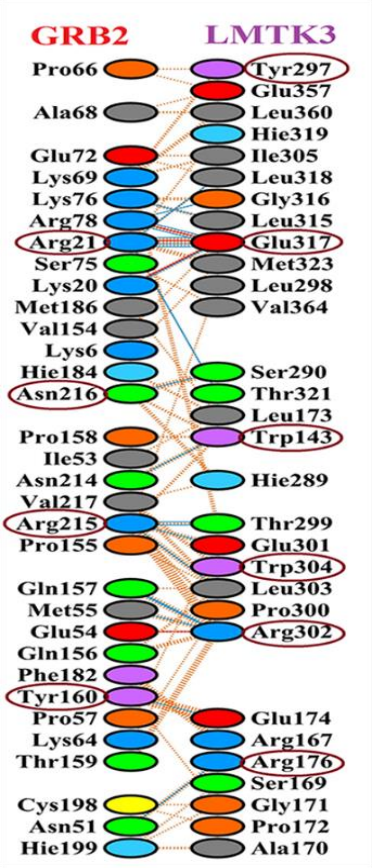

Figure 6.5. *Interface residues and their interactions at GRB2-LMTK3 interface*

Table 6.2. Total number of H-bonds involved in GRB2-LMTK3 interface identified from PDBsum.

No. of H-bonds	(GRB2) Chain A				($LMTK3_{133-411}$) Chain B				H-bond distance
	Chain	Atom no.	Atom name	Residue name and no.	Chain	Atom no.	Atom name	Residue name and no.	
1	A	310	NZ	LYS20	B	6471	OE2	GLU317	2.67
2	A	332	NH1	ARG21	B	6473	O	GLU317	2.84
3	A	332	NH1	ARG21	B	6470	OE1	GLU317	2.85
4	A	335	NH2	ARG21	B	6473	O	GLU317	3.26
5	A	335	NH2	ARG21	B	6492	O	LEU318	3.18
6	A	335	NH2	ARG21	B	6527	OG1	THR321	2.98
7	A	808	OD1	ASN51	B	4169	NH1	ARG176	3.00
8	A	885	O	MET55	B	6217	NH2	ARG302	2.74
9	A	1238	NZ	LYS76	B	6451	O	LEU315	2.98
10	A	1277	NH1	ARG78	B	6471	OE2	GLU317	2.88
11	A	1280	NH2	ARG78	B	6458	O	GLY316	2.88
12	A	1280	NH2	ARG78	B	6471	OE2	GLU317	2.94
13	A	2536	N	GLN157	B	6221	O	ARG302	3.12
14	A	2552	O	GLN157	B	6214	NH1	ARG302	2.98
15	A	2594	OH	TYR160	B	4132	OE1	GLU174	2.84
16	A	3430	O	ASN214	B	3645	NE1	TRP143	2.68
17	A	3447	NH1	ARG215	B	6168	O	THR299	2.96
18	A	3447	NH1	ARG215	B	6197	O	GLU301	2.90

Table 6. 3. Number of salt bridges at GRB2-LMTK3 interface obtained from PDBsum

No. of Salt bridges	GRB2 (Chain A)				$LMTK3_{133-411}$ (Chain B)				Salt bridge distance
	Chain	Atom no.	Atom name	Residue name and no.	Chain	Atom no.	Atom name	Residue name and no.	
1	A	310	NZ	LYS20	B	6471	OE2	GLU317	2.67
2	A	332	NH1	ARG21	B	6470	OE1	GLU317	2.85
3	A	886	OE2	GLU54	B	6217	NH2	ARG302	3.56
4	A	1277	NH1	ARG78	B	6471	OE2	GLU317	2.88

6.4.2. Molecular dynamics simulation analysis

From the molecular dynamics, the analysis of stability and convergence of the modelled structure of GRB2-LMTK3 complex has been studied, as a function of time. From the energy plots (**Figure 6.6**), we infer that the modelled complex structure has attained equilibration.

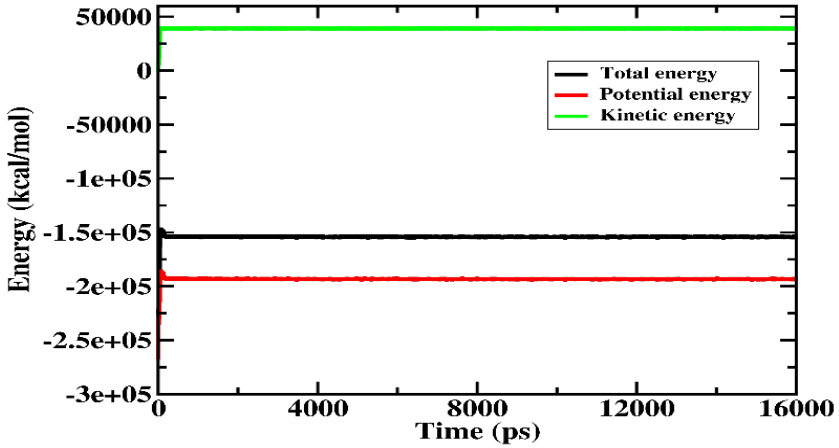

Figure 6.6. *Potential energy, kinetic energy and total energy of the GRB2-LMTK3 complex as a function of time*

For RMSD analysis, the equilibrated complex structure was used as a reference structure. In the complex structure, the RMSD value settles well around 2.5 Å after 8

ns. The stability of the complex structure can be seen from RMSD analysis as shown in **Figure 6.7**. The snapshots of GRB2-LMTK3 complex at different intervals of simulation time were shown in **Figure 6.8**.

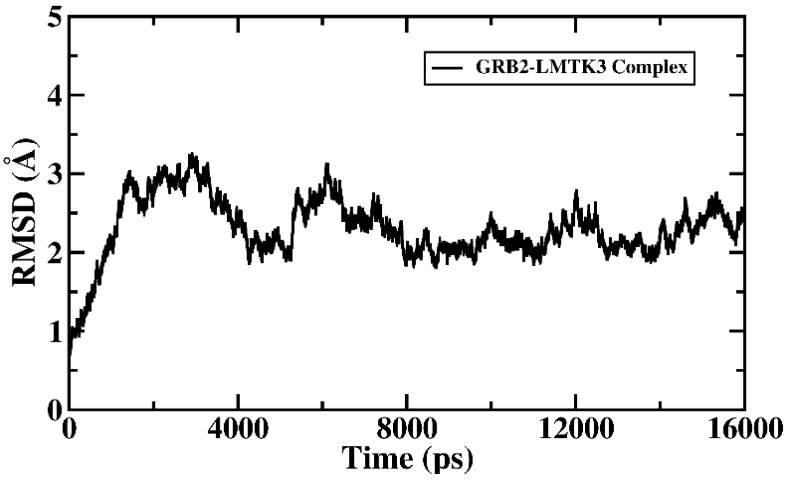

Figure 6.7. *RMSD of GRB2-LMTK3 complex as a function of simulation time.*

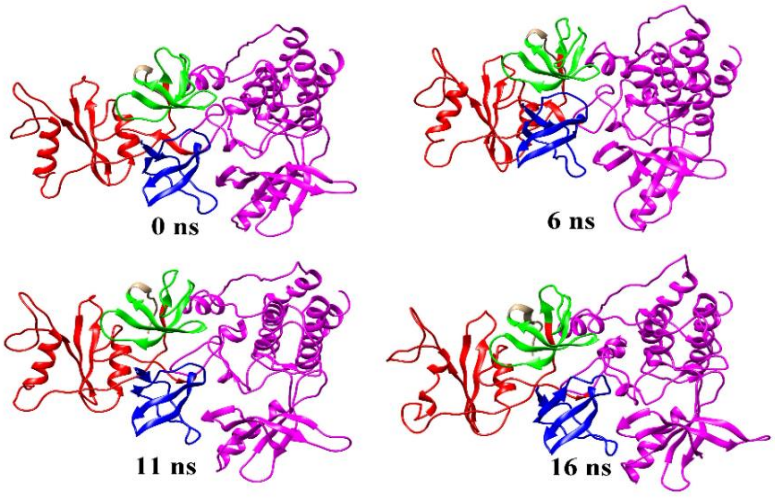

Figure 6.8*. Snapshots of GRB2-LMTK3 complex at different time intervals of MD simulation*

We also analyzed the number of intermolecular hydrogen bonds between GRB2 and LMTK3 as a function of time. Since the hydrogen bonds play a crucial role in stabilizing the protein complexes, therefore we analyzed the hydrogen bonds between GRB2 and LMTK3. The hydrogen bond analysis was carried out from the MD simulation trajectory as shown in **Figure 6.9**. The hydrogen bonds obtained are within the ideal range (2.5 Å – 3.2 Å) [364,365] Maximum hydrogen bonds are 20 with an average of 15-20

hydrogen bonds formed during the course of MD simulation. The occupancy of the hydrogen bond formation of GRB2-LMTK3 complex along with their respective bond distances and bond angles were presented in **Table 6.4**. Thus we infer the intermolecular hydrogen bonds stabilizing the GRB2-LMTK3 complex in the dynamics system.

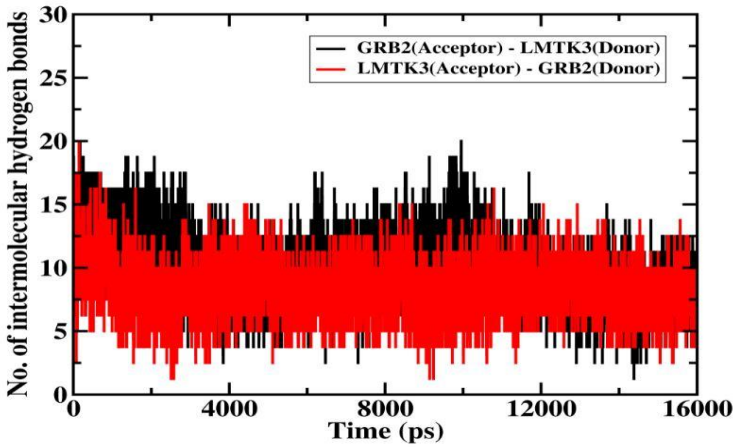

Figure 6.9. *Number of inter-molecular hydrogen bonds in GRB2-LMTK3 complex as a function of simulation time period. Here we have considered inter-molecular hydrogen bonds involved in two cases: GRB2 as acceptor and LMTK3 as donor (depicted in black) and the other one, GRB2 as donor and LMTK3 as acceptor (depicted in red).*

Table 6.4. Intermolecular H-bond occupancy of GRB2-LMTK3 complex from MD simulation.

Proteins	#Acceptor	DonorH	Donor	Frames	Frac	AvgDist	Avg Ang
LMTK3 as an Acceptor And GRB2 as donor	ARG_387@O	GLN_157@H	GLN_157@N	4669	0.81	2.85	164.66
	THR_384@O	ARG_215@HH11	ARG_215@NH1	3602	0.62	2.82	153.38
	GLU_402@O	ARG_21@HH21	ARG_21@NH2	2301	0.40	2.81	154.7
	GLU_402@O	ARG_21@HH11	ARG_21@NH1	1685	0.29	2.82	151.76
	TRP_228@O	TYR_160@HH	TYR_160@OH	1602	0.27	2.82	163.95
	ASP_411@OD1	ARG_21@HH11	ARG_21@NH1	1184	0.20	2.81	155.84
	GLU_386@O	ARG_215@HH21	ARG_215@NH2	1170	0.20	2.83	145.86
	PRO_257@O	ASN_51@HD21	ASN_51@ND2	1047	0.18	2.85	153.40
	ASP_411@OD1	ARG_21@HH21	ARG_21@NH2	947	0.16	2.81	153.31
	SER_254@HG	ARG_178@HH22	ARG_178@NH2	817	0.14	2.87	149.93
	SER_375@OG	ASN_216@HD22	ASN_216@ND2	802	0.13	2.89	153.32
	GLU_386@O	ARG_215@HH11	ARG_215@NH1	786	0.13	2.83	143.50
	GLU_259@OE1	TYR_160@HH	TYR_160@OH	781	0.13	2.69	154.22
	HIE_374@O	ASN_216@HD21	ASN_216@ND2	670	0.11	2.86	148.96

	GLY_401@O	LYS_76@HZ2	LYS_76@NZ	663	0.11	2.8382	154.70
	GLU_402@OE1	ARG_21@HH22	ARG_21@NH2	628	0.10	2.7883	160.46
	LEU_388@HD21	GLN_156@HE22	GLN_156@NE2	496	0.08	2.8219	151.05
	LEU_388@O	ARG_215@HH21	ARG_215@NH2	486	0.08	2.8864	144.84
	GLU_402@OE2	ARG_21@HH22	ARG_21@NH2	450	0.07	2.794	160.56
	THR_406@OG1	ARG_21@HH22	ARG_21@NH2	436	0.07	2.8879	152.17
	GLU_402@OE2	ARG_21@HH11	ARG_21@NH1	409	0.071	2.7914	156.99
	LEU_388@HD22	GLN_156@HE22	GLN_156@NE2	377	0.06	2.817	151.38
	ARG_387@HH22	GLU_54@HG3	GLU_54@CG	369	0.06	2.8548	142.37
	GLU_402@OE1	ARG_21@HH11	ARG_21@NH1	369	0.06	2.7866	157.09
GRB2 as an Acceptor and LMTK3 as donor	MET_55@O	ARG_387@HH12	ARG_387@NH1	4768	0.83	2.8259	158.87
	ASP_181@OD2	SER_254@HG	SER_254@OG	4638	0.80	2.6686	163.33
	PHE_182@O	ARG_252@HH22	ARG_252@NH2	4132	0.71	2.7761	153.11
	ASP_181@OD1	ARG_252@HH21	ARG_252@NH2	3765	0.65	2.7864	147.79
	GLN_157@O	ARG_387@HH21	ARG_387@NH2	3073	0.53	2.8334	153.19
	GLU_71@OE1	ARG_450@HH11	ARG_450@NH1	1969	0.34	2.7941	160.49
	GLU_71@OE2	ARG_450@HE	ARG_450@NE	1607	0.28	2.8307	161.42
	GLU_54@HG3	ARG_387@HH22	ARG_387@NH2	1040	0.18	2.8488	148.87
	GLU_71@OE2	ARG_450@HH11	ARG_450@NH1	936	0.16	2.8124	157.14
	ASN_51@OD1	ARG_261@HH21	ARG_261@NH2	834	0.14	2.8185	156.80
	PHE_182@O	ARG_252@HH12	ARG_252@NH1	724	0.12	2.8735	143.94
	ASN_216@OD1	TRP_228@HE1	TRP_228@NE1	695	0.12	2.8553	156.75
	GLU_71@OE1	ARG_450@HE	ARG_450@NE	649	0.11	2.8301	161.76
	MET_55@O	ARG_387@HH22	ARG_387@NH2	640	0.11	2.8726	150.86
	GLU_71@OE1	ARG_450@HH22	ARG_450@NH2	623	0.10	2.8205	161.46
	GLU_71@OE2	ARG_450@HH12	ARG_450@NH1	601	0.10	2.8031	162.95
	ASP_181@OD2	ARG_252@HH22	ARG_252@NH2	576	0.10	2.7753	157.45
	ASP_104@OD1	ARG_450@HH11	ARG_450@NH1	563	0.09	2.7907	157.46
	ASN_214@O	TRP_228@HE1	TRP_228@NE1	556	0.09	2.8772	156.80
	GLU_54@OE2	ARG_387@HH22	ARG_387@NH2	519	0.09	2.8439	145.86
	GLU_71@OE1	ARG_450@HH12	ARG_450@NH1	514	0.08	2.7936	159.52
	ARG_215@HH21	LEU_388@H	LEU_388@N	384	0.06	2.8681	142.49
	ASN_51@O	ARG_261@HH21	ARG_261@NH2	376	0.06	2.8447	154.82
	THR_159@OG1	ARG_387@HH21	ARG_387@NH2	364	0.06	2.8248	146.48
	ASP_104@OD1	ARG_450@HE	ARG_450@NE	347	0.06	2.8522	153.49

6.4.3. Hot Spot residues at GRB2-LMTK3 interface

Using four computational methods (KFC2 server, PredHS, Robetta and DrugScore^{PPI} server), the probable hot spot residues in GRB2 (Chain A) and LMTK3 (Chain B) were identified. KFC and PredHS server list out the interface hot spot and non-hot spot residues. Robetta and DrugScore^{PPI} server use ASM method to identify hot spots. Here, it mutates the individual interacting interface residues of GRB2 and LMTK3 to alanine, and then calculate the change in binding free energy (ΔΔG) of the GRB2-LMTK3 complex. We noticed that there is a significant change in binding free energy of alanine mutated protein, those residues when mutated to alanine changes the binding free energy values of GRB2-LMTK3 complex to more than 2 kcal mol^{-1} are considered as hot spots and those having binding free energy value less than 2 kcal mol^{-1} considered as null-spots. Thus, we ensured that our predicted interface hot spots may contribute significant binding free energy to the structural stability of the GRB2-LMTK3 complex. The hot spot residues at GRB2-LMTK3 interface to be Arg21A, Tyr160A, Arg215A, Asn216A, Tyr143B, Arg167B, Tyr297B, Arg302B and Trp304B. Our predicted hot spots contain mostly Tyr, Arg and Trp amino acids and in literature these residues found to have greater tendency in being a hot spot. The hot spots are encircled in **Figure. 6.5**. **Figure 6.10** represents the hot spot and null-spot residues at GRB2 and LMTK3 interface as spheres. Hot spots residues are known to be enriched in forming H-bonding and salt bridges [258, 365,366]. The results obtained from all the four methods are summarised in **Table 6.5**. Our predicted hot spots are also involved in the formation of H-bonds and salt bridges as shown in **Table 6.2** and **Table 6.3.**

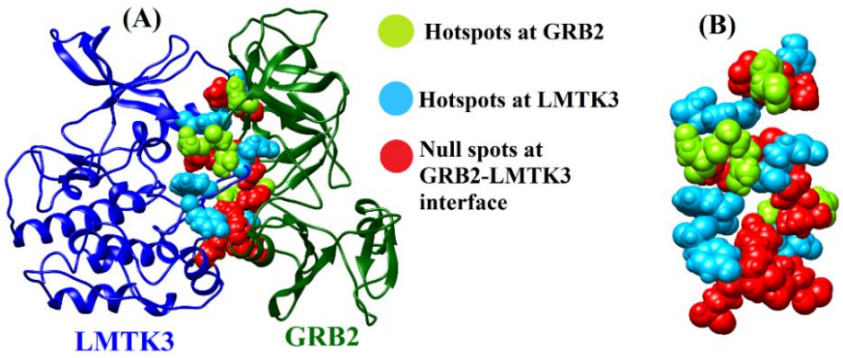

Figure 6.10 *(A) Representation of hot spot and null spot residues at GRB2 and LMTK3 interface. Red color representing the null-spot residues. Whereas the light blue and green color depicted as hot spots. (B) Representation of the interaction of hot spot (green and blue) residues in a form of spheres without secondary structure*

Table 6.5. Comparison of results from six different computational methods used to detect probable hotspot residues at GRB2-LMTK3 interface

Proteins	Residue Name	PredHS server	KFC server	Alanine scanning mutagenesis ΔΔG kcal mol^{-1}	
				Robetta server	Drug score server
GRB2	**Arg21A**	H	H	4.10	2.26
	Lys69 A	NH	NH	1.46	1.04
	Glu72 A	H	NH	1.39	0.45
	Lys76 A	NH	H	0.63	1.14
	Val154 A	NH	NH	0.84	0.86
	Gln156 A	NH	NH	1.05	0.90
	Gln157 A	NH	H	1.72	0.35
	Tyr160 A	H	H	3.14	2.41
	Arg215 A	H	H	4.20	2.41
	Asn216 A	H	H	2.25	3.41
	Val217 A	H	H	1.09	1.25
LMTK3	**Typ143B**	H	H	4.68	2.56
	Arg176 B	H	H	2.43	2.01
	Leu173 B	NH	H	1.17	0.87
	Glu174 B	H	H	1.85	0.56
	Tyr 297 B	H	H	3.88	1.50
	Arg302 B	H	H	3.34	2.68
	Trp304 B	H	H	3.87	1.36
	Glu317 B	H	H	2.33	0.29
	Leu318B	H	NH	0.48	0.88
	Leu360 B	NH	H	1.27	0.87

H: Hot spot; NH: Non hot spot (it may be the null spot residues)

6.4.4. Binding free energy using MM-GBSA method

The total binding free energy calculation for GRB2-LMTK3 complex was performed by using the MM-PBSA.py script in AMBER12. The estimated total binding free energy

and the component energies for GRB2-LMTK3 is shown in **Table 6.6**. The total binding free energy value of the complex was found to be -51.86 kcal mol^{-1}. The negative total binding free energy value infers the favorable energy contribution towards the formation of GRB2-LMTK3 complex in solvated system and provides structural stability but it should be noted that our result does not equal to the actual binding free energy since we did not estimate the (unfavorable) entropy contribution to binding.

Table 6.6 Binding free energy calculation of GRB2-LMTK3 complex using MM-GBSA method

Energy components	Complex (GRB2-LMTK3) kcal mol^{-1}	Receptor (LMTK3) kcal mol^{-1}	Ligand (GRB2) kcal mol^{-1}	$\Delta G_{binding}$ kcal mol^{-1} (complex-receptor-ligand)
VDWAALS	-3967.3921	-2234.0304	-1603.8141	-129.5476
EEL	-34932.3476	-18834.4469	-15787.6786	-310.2221
E_{GB}	-5323.8013	-2764.1319	-2966.1774	406.5079
E_{CAVITY}	183.667	113.2998	88.9664	-18.5995
G_{gas}	-38899.7398	-21068.4773	-17391.4927	-439.7698
G_{solv}	-5140.1346	-2650.8321	-2877.2110	387.9085
TOTAL	-44039.8744	-23719.3094	-20268.7037	-51.8613

VDWAALS = van der Waals contribution from molecular mechanics (MM); **EEL** = electrostatic energy; $\mathbf{E_{GB}}$ = the electrostatic contribution to the solvation free energy calculated by GB respectively; $\mathbf{E_{SURF/ECAVITY/ENPOLAR}}$ = nonpolar contribution to the solvation free energy calculated by an empirical model; $\mathbf{G_{gas}}$ = average interaction energy of complex, receptor and ligand in gas phase; $\mathbf{G_{solv}}$ = average interaction energy of complex, receptor and ligand in solvent; $\mathbf{\Delta G_{binding}}$ = final estimated binding free energy calculated from the terms above. (kcal mol^{-1})

6.5. Conclusions

In this study, we have demonstrated the transient protein-protein interactions in GRB2-LMTK3 complex that induces the transcriptional activation of integrin β_1. Our findings present the probable interactions, interface area, hydrogen bonds and non-bonded contacts involved in modelled structure of GRB2-LMTK3complex.We have also

identified the probable interface hot spot residues across GRB2-LMTK3 complex using four different computational methods (KFC2 server, PredHS, Robetta and DrugScore^{PPI} server). MD simulation study corroborates the stability of the complex. And the total estimated MM-GBSA binding free energy value was found to be -51.86 kcal mol^{-1} and this value indicates the extent of stability of GRB2-LMTK3 complex. The predicted interface hotspot residues at GRB2 (Chain A) and LMTK3 (chain B) were found to Arg21A, Tyr160A, Arg215A, Asn216A, Tyr143B, Arg167B, Tyr297B, Arg302B, Trp304B. The information about probable interface area and hotspot residues at GRB2-LMTK3 interface could be useful in designing potential PPI inhibitors for therapeutic target that could obstruct the interacting interface between GRB2 and LMTK3 and thereby controlling breast cancer progression and metastasis.