RESULTS

Results

4.1) Results for Benzo(a)pyrene

Up-regulated and down-regulated genes were separately extracted from T3DB. T3DB is a comprehensive database that contains approximately 3678 toxic compounds and their targets. For BaP, T3DB had 1722 target genes which are getting upregulated and 2369 target genes which are getting downregulated.

4.1.1) Network generation:

For network generation, upregulated genes and downregulated genes were subjected separately into STRING-db and then with the help of Cytoscape software, both the networks were merged. The upregulated network generated had 1676 nodes and 7113 edges. The average node degree was 8.49 and average clustering coefficient was 0.398. The PPI enrichment p-value was <1.0e-16.

For downregulated proteins, STRING-db generated a network which had approximately 2563 nodes and 11,101 edges. STRING-dbis capable of generating network of upto 2000 nodes only. The total number of down-regulated proteins that were extracted from T3DB was 2369. To generate the complete interactome of down-regulated proteins, two networks with approximately 1200 proteins were generated by STRING-db and were then merged together using the merge tool of cytoscape

software.The merged network was then again merged with the network of up-regulated proteins to generate a combined interactome of proteins hampered by BaP (Fig 9).

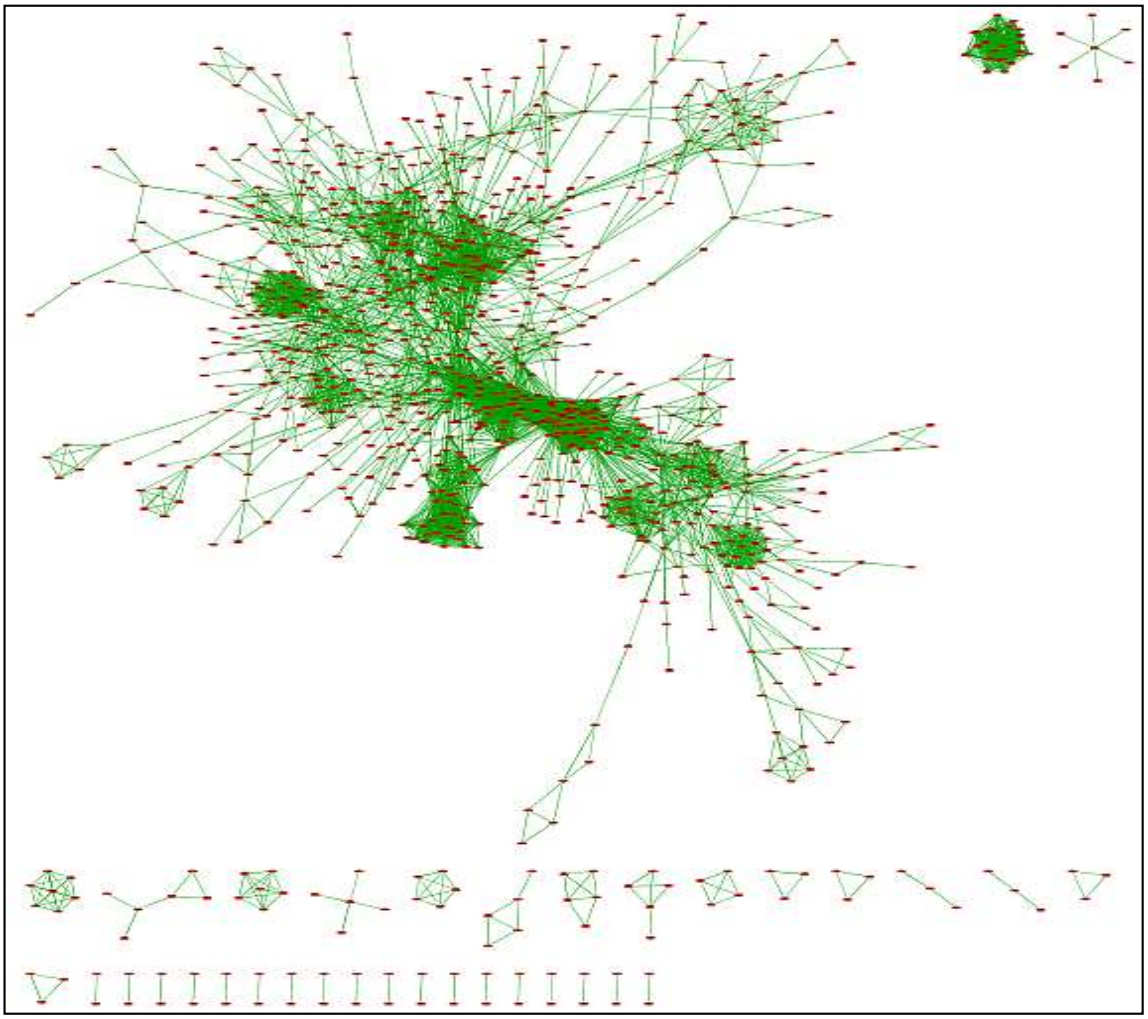


Fig 9:BaP rewired PPIN (merged network generated by cytoscape). The PPIN comprises of upregulated and downregulated proteins having 2058 nodes and 13850 edges.

4.1.2) Network analysis results:

Network analysis was done using Network Analyser tool of cytoscape software. The network was analyzed based on the various topological properties like clustering coefficient, node degree distribution, shortest path length etc. The clustering coefficient of the network was 0.559, the characteristic pathlength was 4.615, average number of neighbors was 13.460 and network density was 0.007 and the diameter of the network was 13.

The length of paths of individual proteins are biologically significant. It is the average steps taken through the shortest paths to efficiently transfer the information. Shortest path length is the shortest distance between any two communicating nodes to pass the information completely. The characteristic path length in the network was 4.615. Fig 10 depicts the shortest path length distribution within the network. Highest peak has a frequency of approximately 960,000 at pathlength of 4 while the smallest peak has a frequency of less than 25000 at path length of 10.

The average clustering coefficient graph (fig 11) depicts the average of clustering coefficients of all the nodes present in the network. It tells about the tendency of the nodes to form clusters with the functionally related proteins in a network. Power law ($y= ax^{-b}$) has been applied to the graph in which a=1.276, b= -0.249, $r^2 = 0.285$.

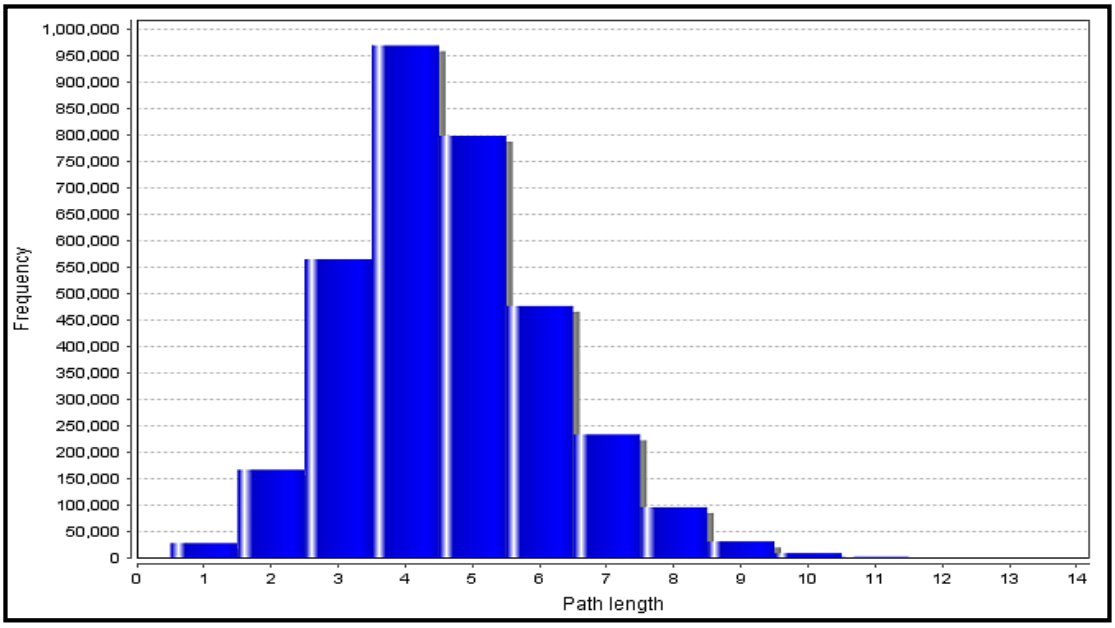


Fig 10: Shortest pathlength distribution. 960,000 interactions have the shortest path length of 4.

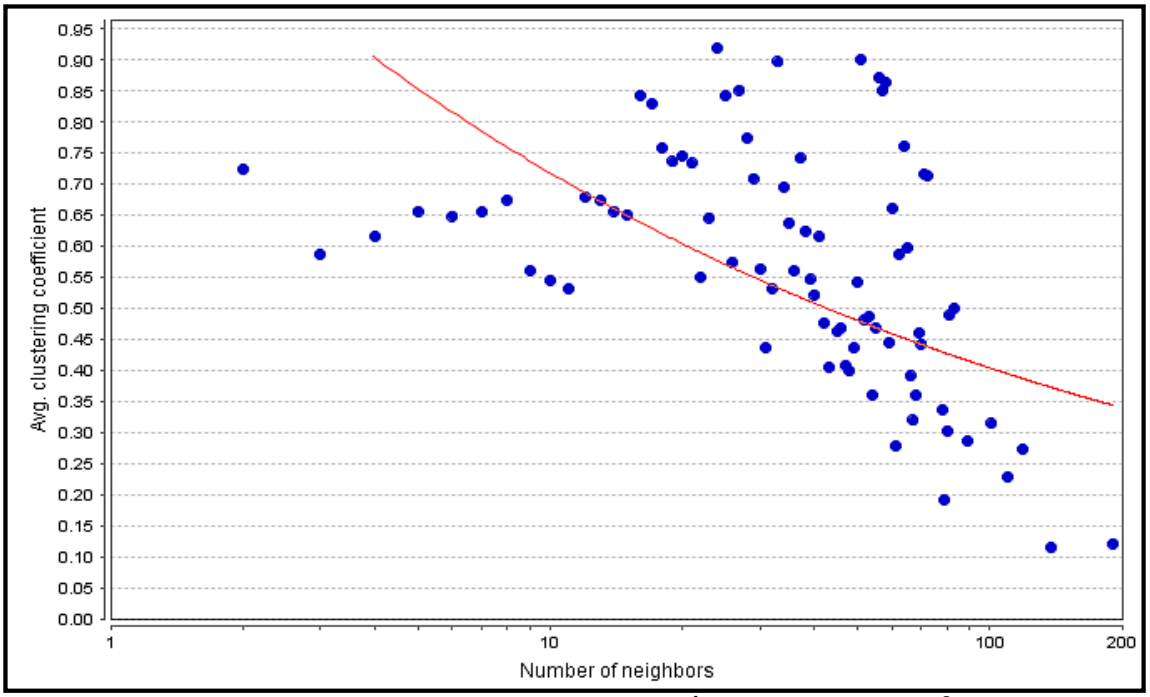


Fig 11: Avg. clustering coefficient (power law: $y = ax^{-b}$; a=1.276, b= 0.249, $r^2 = 0.285$)

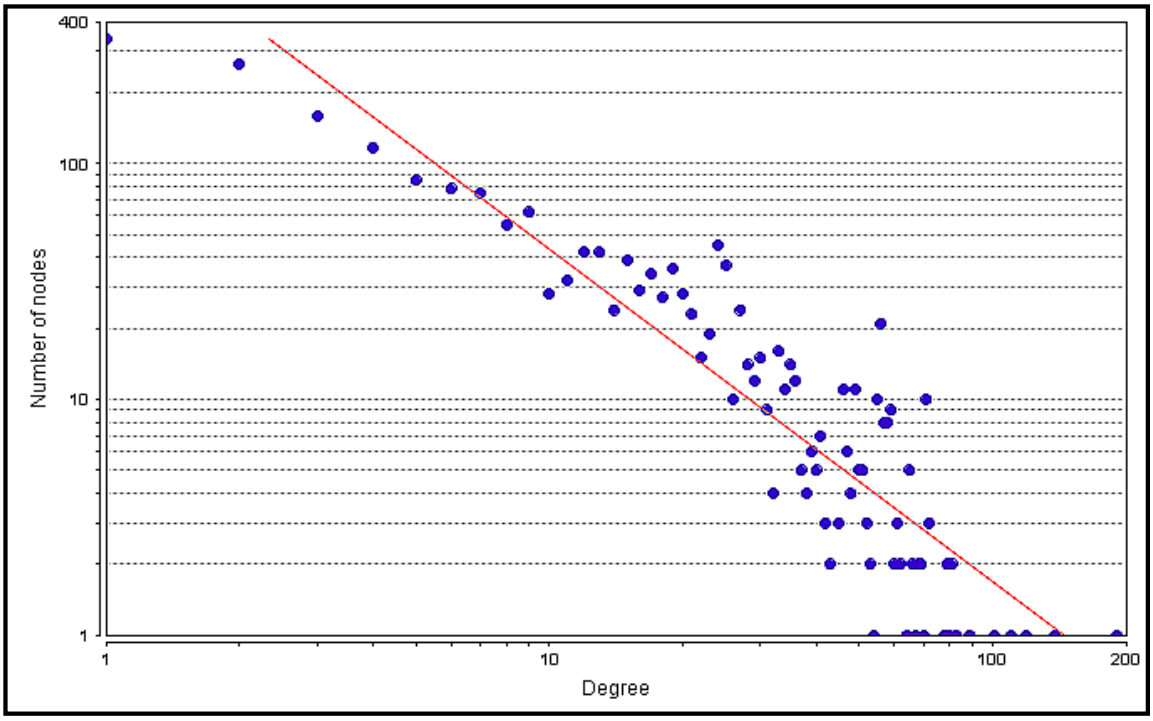


Fig 12: Node degree distribution (power law: $y= ax^{-b}$; a=1106.5, b= 1.408, $r^2 = 0.824$)

Fig 12 is a graph for node degree distribution which is to analyse the type of network. With the help of node degree distribution, skewness of the curve and all the degrees are adjusted around the average value. The graph has been fitted using power law where R-squared value was calculated to be 0.824 and b was 1.408. Another topological property of networks is neighborhood connectivity. Fig 13 represents the neighborhood connectivity distribution curve fitted with power law having R-squared value of0.590.

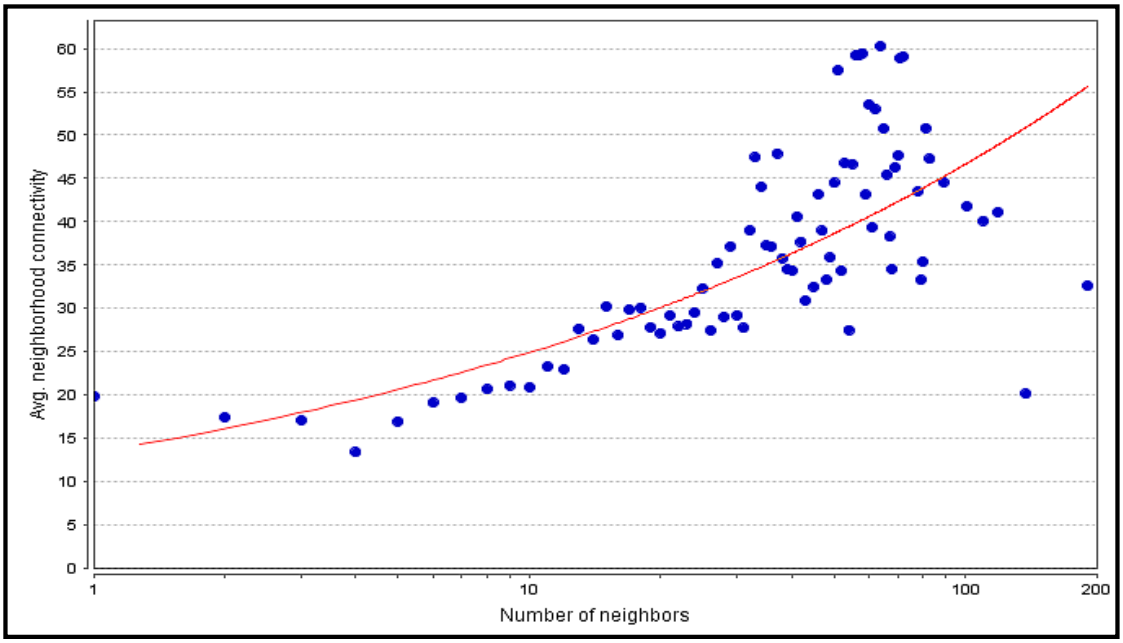


Fig13: Neighborhood connectivity distribution (power law: $y = ax^{-b}$; a= 13.316, b= 0.272, r^2 = 0.590)

Annexure 1 contains the topological analysis of all the proteins present in the merged BaP rewired PPIN.

4.1.3) Cluster generation and GO enrichment analysis

MCODE was used to generate clusters to remove the noise from the network. Annexure 3 is the list of clusters generated and the finally selected seed proteins (the proteins in bold). Seed proteins identified from each cluster were again subjected to STRING database for PPIN generation (fig 14). The confidence level was maintained at 0.9 and 50-50 connectors were added in the first and the second shells. Table 1 shows the topological properties of the finally selected proteins (seed proteins). The most important topological properties on the basis of which further screening of the most

probable biomolecule was done were degree, clustering coefficient, betweenness centrality and finally bottleneck score. The proteins were first sorted based on degree and then those proteins were selected that had high degree and clustering coefficient less than 0.5. The next criterion for sorting the proteins was high betweenness centrality followed by bottleneck scores. To further screen the proteins, median was applied as the data generated was discrete in nature. For BaP, 38 proteins finally emerged out to be the most potent biomolecular targets with high bottleneck score, high degree and clustering coefficient less than 0.5 as shown in table 1.

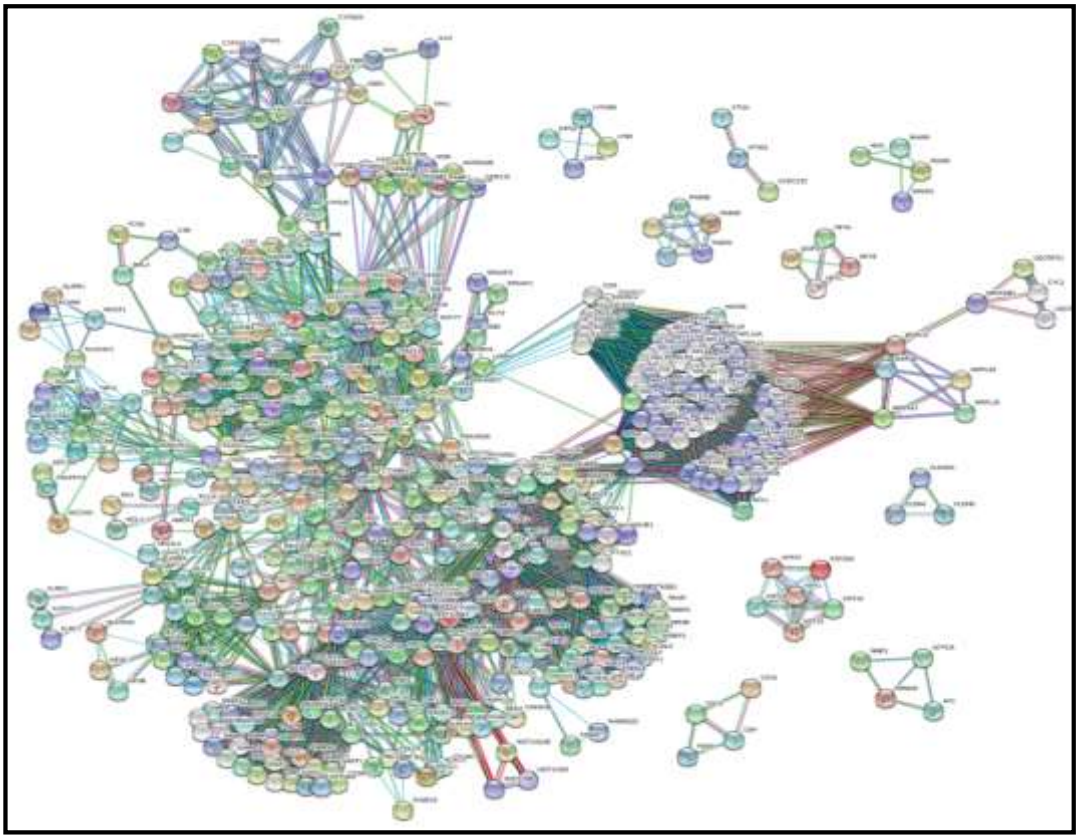


Fig 14: PPIN generation from seed proteins identified from clusters generated by MCODE

GO enrichment was done using ClueGO which helps in finding genes/proteins that are functionally linked together and helps in enhanced interpretation of the data. Fig 15 shows the pathways and the number of genes involved in those enriched pathways. The pie chart below (fig 16) shows that most of the proteins/genes that have got enriched below to various phases of cell cycle regulation.

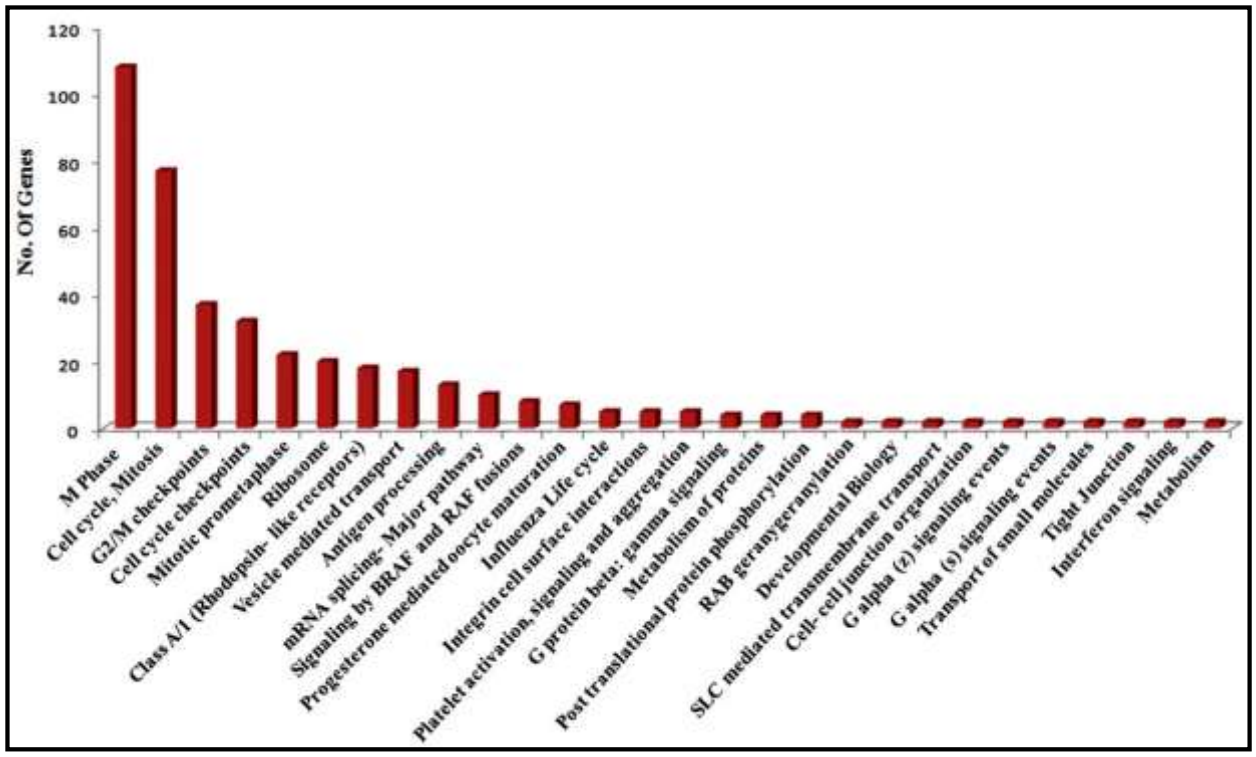


Fig 15: Chart for number of genes involved in the GO enriched pathways

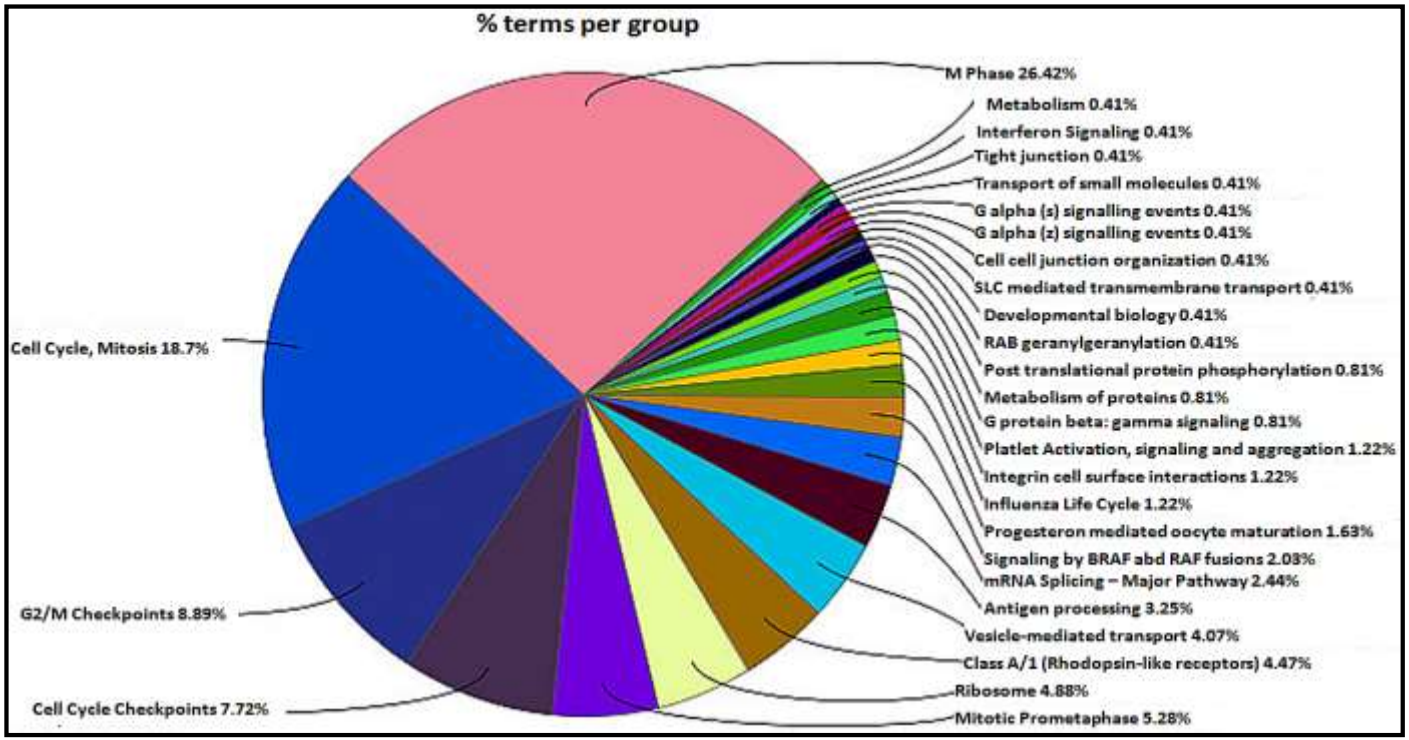


Fig 16: ClueGO results of GO functional enrichment of key proteins

Table 1: Final sorted proteins along with their network topological analysis

Name	Betweenness	BottleNeck	Closeness	ClusteringCoefficient	Degree
CXCL8	37245.68	113	215.1167	0.40621	39
HSP90AA1	25603.75	69	217.4333	0.1954	30
HSPA8	15835.38	31	191.8833	0.37374	45
CTNNB1	16598.02	27	202.0833	0.13547	34
PTGS2	12040.96	22	159.55	0.39286	8
JUN	15855.92	21	200.7833	0.22984	32
POLR2B	5114.551	21	184.8619	0.40426	48
EGFR	11690.42	17	199.4333	0.2381	21
ALOX15B	4879.058	14	146.6619	0.46429	8
PLK1	5846.966	13	209.2833	0.3697	45
TUBB	6692.258	12	174.5333	0.46199	19
HLA-E	6008.815	12	158.0881	0.43841	24
GGH	5026.219	12	159.3619	0.40952	15
QSOX1	2714.159	12	174.3786	0.45873	36
APP	9332.196	11	197.4833	0.29861	64
GNG11	9143.955	11	190.3333	0.36078	47
HIST2H2BE	2540.818	10	199.5167	0.38655	35
ESR1	7749.228	9	202.7667	0.29667	25
AKT1	7198.725	7	194.2	0.18301	18
CYP2B6	6283.716	7	130.6952	0.46154	13
TUBA1A	4199.112	7	173.4	0.40351	19
IRF7	3836.475	7	176.3381	0.45333	25
HIST1H2BD	2540.818	7	199.5167	0.38655	35
CDH1	946.825	6	166.1667	0.25275	14
POLR2A	8166.016	5	191.2119	0.3418	53
CYP2E1	6169.535	5	131.9286	0.34615	13
ITGAX	2379.19	5	147.1048	0.46667	10
MNAT1	2376.592	5	170.2857	0.32164	19
HIST1H2AD	2134.95	5	192.85	0.41129	32
ITGA1	1585.011	5	159.0333	0.29091	11
PSMC4	4311.267	4	196.2167	0.46667	30

PSMC5	4027.439	4	195.7167	0.46059	29
HIST1H2BK	3668.499	4	201.0167	0.33855	38
NOS2	3433.019	4	153.2119	0.26667	6
HRAS	3376.991	4	184.85	0.22794	17
HIST1H2BO	2540.818	4	199.5167	0.38655	35
LAMA3	2328.637	4	159.5929	0.33333	6
AKR1B10	1506.571	4	116.7536	0.4	5

Summary of Modulation and Enrichment analysis of BaP

- 65 clusters got generated.
- 411 seed proteins were obtained from the clusters.
- Enrichment analysis showed that most of the proteins were involved in cell cycle regulatory pathways.

4.1.4)Molecular docking simulation analysis

Molecular docking simulation was used to finally pinpoint the most important biomolecular target of BaP.Total 38 molecules were docked and were arranged in decreasing binding energy values. On docking simulations, QSOX1 showed the highest binding efficiency for BaP (-8.01 Kcal/mol) followed by PTGS2, NOS2 and ESR1 with -7.82 Kcal/mol, -7.09 Kcal/mol and -7.08Kcal/mol binding energies respectively. The table (Table 2) is the list of docked proteins along with their binding energies and Ki values. Fig 17 is the pictorial representation of the top 3 docked biomolecular targets of BaP. Fig 18 is the collective graphical representation of the

impact of BaP on system level protein interaction network with their respective key regulatory proteins along with the enriched pathways.

Table 2: Docking results of final proteins selected from BaP rewired PPIN

S.No	Name	Ligand	Binding energy (Kcal/mol)	Ki	Binding residues	H-Bond	Distance
1	**QSOX1**	BaP	**-8.01**	1.34 uM	Pro404 Leu407 Trp408 Phe411 Phe447 Cys452 His455 Phe456 Val484 Leu488	A:LEU407:N - A:PRO404:O A:TRP408:N - A:PRO404:O A:PHE411:N - A:LEU407:O A:PHE411:N - A:TRP408:O A:LEU488:N - A:VAL484:O A:TRP408:CD1 - A:PRO404:O	3.31929 3.21556 2.84181 3.20787 3.09837 3.5322
2	**PTGS2 (COX2)**	BaP	**-7.82**	1.84 uM	Ala199 Ala202 Gln203 Thr206 His207 Phe210 Asn382 Tyr385 His386 Trp387 His388 Leu390 Leu391	A:THR206:N - A:ALA202:O A:THR206:OG1 - A:ALA202:O A:TRP387:N - A:TYR385:O A:HIS388:ND1 - A:HIS388:NE2:B A:HIS388:N:B - A:TRP387:O A:LEU390:N - A:TRP387:O	3.01797 2.69521 2.89137 1.22008 2.26883 3.01943
3	**NOS2**	BaP	**-7.09**	6.37 uM	Trp194 Ala197 Cys200 Leu209 Ser242 Phe369 Asn370 Gly371 Tyr489	D:TRP194:NE1 - D:CYS200:SG D:ALA197:N - D:TRP194:O D:CYS200:N - D:ALA197:O D:ASN370:ND2 - D:GLY371:O D:SER242:CA - D:ASN370:O D:CYS200:SG - :UNK1 D:CYS200:SG - :UNK1	3.61459 2.87458 3.24962 3.3139 3.52924 3.99886 4.11644

4	ESR1	BaP	-7.08	6.49 uM	Leu346 Ala350 Glu353 Leu387 Met388 Leu391 Arg394 Phe404 Met421 Ile424 Leu428 His524	A:ALA350:N - A:LEU346:O A:MET388:N - A:LEU384:O A:MET388:N - A:GLU385:O A:GLY390:N - A:LEU387:O A:LEU391:N - A:LEU387:O A:ARG394:NH2 - A:PHE404:O A:HIS524:ND1 - A:HIS524:O	2.97454 2.9999 3.00627 3.08902 2.88658 3.06961 2.72008
5	CYP2E1	BaP	-6.87	9.16 uM	Ala299 Thr303 Thr304 Thr307 Gln358 Leu363 Val364 Pro429 Phe430 Cys437 Ala443 Leu447	A:THR301:N - A:ALA299:O A:THR303:OG1 - A:ALA299:O A:THR307:OG1 - A:THR303:O A:THR303:OG1 - :UNK1 A:THR303:OG1 - :UNK1	3.15039 3.14884 2.56984 3.32776 3.18553
6	AKT1	BaP	-6.82	9.97 uM	Arg15 Lys20 Thr21 Trp22 Pro68 Arg69	A:ARG15:NH1:B - :UNK1 A:ARG15:NE - A:LYS20:O A:ARG15:NH1:B - A:LYS20:O A:ARG15:NH2:B - A:LYS20:O	3.98741 3.10778 3.19239 3.07423
7	CYP2B6	BaP	-6.45	18.69 uM	Thr302 Thr303 Thr306 Gln357 Asp361 Leu362 Leu363 Leu396 Pro428 Phe429 Cys436 Ala442	A:THR306:OG1 - A:THR302:O A:ASP361:N - A:GLN357:O A:LEU362:N - A:SER360:O A:MET365:N - A:LEU363:O A:GLY366:N - A:LEU363:O A:LEU363:N - :UNK1	2.68713 3.0493 3.19138 3.19738 2.89141 3.91878

8	PSMC4	BaP	-5.92	45.44 uM	Asp166 Gly168 Met170 Gly209 Cys210 Gly211 Ile344 Thr347 Ile348 Gly372 Ala373 Asn376	K:GLY168:N - K:ASP166:O K:THR347:OG1 - K:LEU343:O K:THR347:OG1 - K:ILE344:O K:ILE348:N - K:ILE344:O K:ILE348:N - K:PHE345:O K:ILE348:N - K:THR347:OG1 K:SER350:OG - K:THR347:O K:ALA373:N - K:GLY209:O K:ALA373:N - K:SER371:O	3.02228 2.64764 2.54605 2.84086 2.9878 2.97982 2.48498 3.00543 2.89215
9	ALOX15B	BaP	-5.78	58.16 uM	Phe184 Ala188 Phe192 Leu415 Ala416 Leu419 Leu420 Leu605 Ala606 Leu609 Leu610	A:GLU418:H - A:LEU415:O A:LEU419:H - A:LEU415:O A:LEU610:H - A:ALA606:O	2.34574 2.20884 2.40142
10	GGH	BaP	-5.7	65.91 uM	Arg79 Gly122 Glu123 Cys124	A:CYS124:N - A:CYS124:N:B A:CYS124:N - A:CYS124:O:B A:CYS124:N:B - A:GLU123:O A:CYS124:N:B - A:CYS124:O A:CYS124:SG:B - A:CYS124:SG A:LEU125:N - A:CYS124:O:B A:CYS124:N - :UNK1 A:CYS124:N - :UNK1 A:CYS124:N:B - :UNK1 A:CYS124:N:B - :UNK1 A:CYS124:N:B - :UNK1	0.034233 2.83887 2.25589 2.81042 2.1872 2.24857 3.40935 3.72364 3.3954 4.15631 3.72771
11	HIST2H2BE	BaP	-5.2	155.01 uM	Ile30 Phe34 Pro163 Leu166 Gln167 Ile170 Ile180 Ile184	C:HIS165:N - C:PRO163:O C:LEU166:N - C:PRO163:O C:GLN167:N - C:PRO163:O	3.14576 3.01296 2.86164
12	TUBB	BaP	-5.16	164.79 uM	Gln11 Cys12 Gln15 Val171 Ile204 Asn206 Leu209 Tyr224 Leu227 Asn228 Val231	B:ASN228:ND2 - B:TYR224:O B:CYS12:SG - :UNK1 B:CYS12:SG - :UNK1 B:CYS12:SG - :UNK1	2.70811 3.24276 3.67464 3.57764

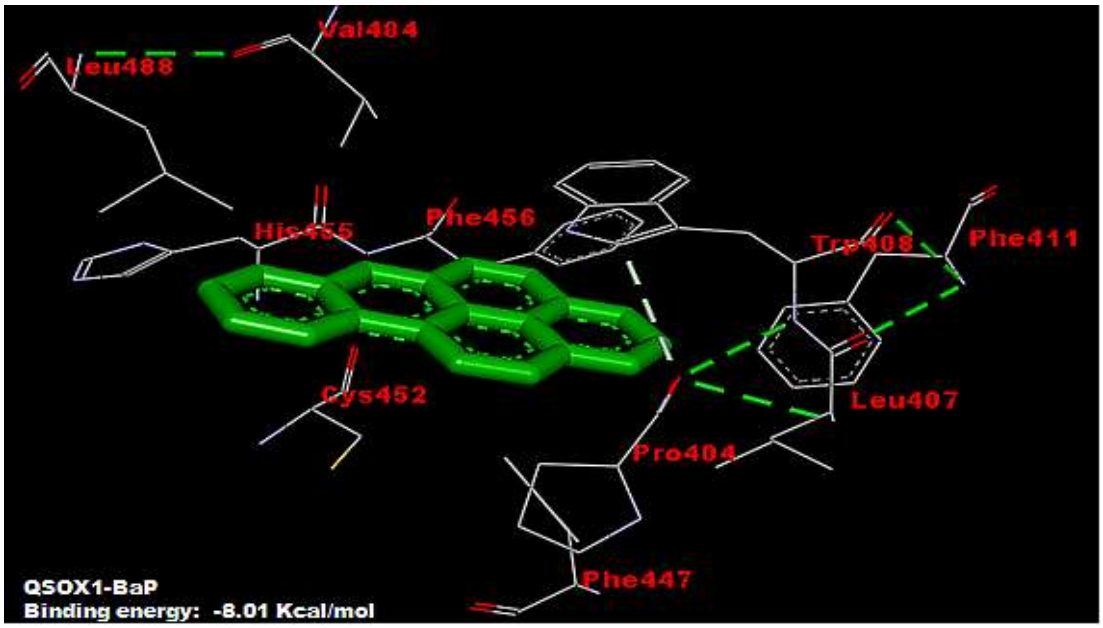
13	PSMC5	BaP	-5.09	185.49 uM	Pro144 Met150 Ile151 Leu198 Ile328 Ile331 His332 Lys335 Gly356 Lys360	J:ILE331:N - J:ILE328:O J:HIS332:N - J:ILE328:O J:LYS335:N - J:HIS332:O J:LYS360:NZ - :UNK1 J:LYS360:NZ - :UNK1	2.9943 2.99262 3.04049 3.51503 3.55091
14	MNAT1	BaP	-5.07	193.75 uM	Met22 Val23 Asn24 His28 Thr29 Leu30 Val35 Leu38 Phe39 Gly44 Asn45 Cys46 Pro47 Phe58	H:LEU30:N - H:MET22:O H:PHE39:N - H:VAL35:O H:ARG59:N - H:VAL23:O	2.88877 2.90719 2.969
15	HSP90AA1	BaP	-5.03	204.50 uM	Leu48 Asn51 Ser52 Ala55 Ile91 Asp93 Met98 Asn106 Leu107 Phe138 Thr184 Val186	A:SER52:N - A:LEU48:O A:SER52:N - A:ILE49:O A:SER52:OG - A:ASP93:OD2 A:ALA55:N - A:ASN51:O A:ASN106:ND2 - A:ASP102:O A:THR184:OG1 - A:THR152:OG1	3.06082 3.11456 2.8652 3.06749 3.21119 2.90778
16	HRAS	BaP	-4.88	264.49 uM	Gly15 Ala18 Phe28 Val29 Asp30 Glu31 Asn116 Lys117 Asp119 Ser145 Ala146 Lys147	A:ASN116:ND2 - A:VAL14:O A:LYS117:N - A:ASN116:OD1 A:LYS117:N - A:THR144:O A:ASP119:N - A:ASP119:OD1 A:THR144:OG1 - A:ASN116:OD1 A:SER145:OG - A:ASP119:OD1 A:LYS147:N - A:SER145:OG A:THR148:N - A:SER145:OG	2.9456 3.02206 3.0571 2.8577 2.85236 2.7634 3.14647 3.14295

17	LAMA3	BaP	-4.87	268.26 uM	His152 Val153 Ala154 Ser232 Leu233 Ile234 Phe243 Thr253 Gln298	:LEU233:H - :ALA154:O :THR253:OG1 - :LEU249:O :THR253:OG1 - :ARG250:O	1.8038 3.13673 2.48714
18	HIST1H2BK	BaP	-4.8	301.13 uM	Val63 Ile66 Phe67 Ile70 Ala71 Ile91	D:PHE67:N - D:VAL63:O D:ARG69:N - D:ILE66:O D:ILE70:N - D:ILE66:O D:ALA71:N - D:PHE67:O	3.02251 3.14917 2.76846 2.78248
19	HIST1H2BO	BaP	-4.76	325.05 uM	Tyr41 Val42 Phe66 Val67 Ile70 Phe71 Leu103	:ILE70:H - :PHE66:O :PHE71:H - :VAL67:O	2.04361 2.02236
20	PLK1	BaP	-4.7	359.01 uM	Lys413 Trp414 Val415 Asn533 Phe534 Phe535 His538 Lys540	A:PHE534:N - A:THR539:O A:PHE535:N - A:SER519:O A:TRP414:N - :UNK1 A:TRP414:NE1 - :UNK1	3.11226 3.05907 3.97532 4.07583
21	APP	BaP	-4.56	453.54 uM	His44 Met45 Asn46 Val47 Val76 Tyr77 Pro78 Glu79 Leu80	A:VAL47:N - A:MET45:O A:GLN48:N - A:ASN46:OD1 A:ASN49:N - A:ASN46:OD1 A:TYR77:OH - A:MET45:O	3.16246 2.94005 3.1866 2.60936
22	POLR2A	BaP	-4.44	558.60 uM	Thr489 Thr490 Asn493 Ala494 Val538 Gln539 Gly773 Ala774 Lys775	A:TYR492:N - A:THR489:O A:ASN493:N - A:THR489:O A:ASN493:N - A:ASN493:OD1	3.11314 2.38715 3.17336

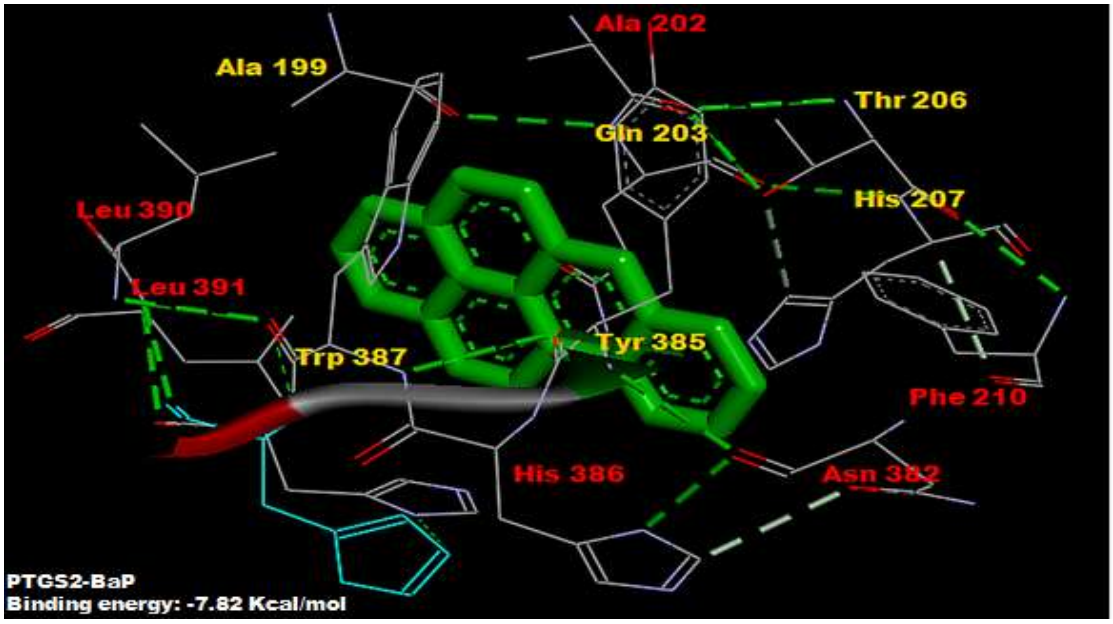
23	IRF7	BaP	-4.2	839.87 uM	Thr71 Arg73 Glu101 Pro140 Gln162 Val163 Thr164 Arg174 Leu175 Pro177	A:ARG73:N - A:GLN162:O A:ARG73:NH2 - A:GLU101:OE2 A:LEU175:N - A:VAL163:O A:ARG174:NE - :UNK1 A:ARG174:NE - :UNK1 A:ARG174:NH2 - :UNK1	3.10222 3.13119 3.09036 3.0746 4.06399 3.79654
24	TUBA1A	BaP	-4.19	854.96 uM	Gln11 Ala12 Gln15 Ser140 Gly142 Gly143 Ile171 Asn206 Tyr224 Asn228	H:SER140:OG - H:SER140:O H:SER140:OG - :UNK1 H:TYR224:OH - :UNK1 H:TYR224:OH - :UNK1	3.17793 3.74854 3.71284 3.89172
25	AKR1B10	BaP	-4.11	972.03 uM	Trp21 Lys22 Pro24 Val48 Tyr49 Asp217 Pro219	A:GLN50:N - A:VAL48:O A:ASN51:N - A:TYR49:O	3.13566 3.01469
26	CXCL8	BaP	-4.11	964.23 uM	His18 Pro19 Lys20 Arg60 Val61 Lys64	D:LYS20:N - D:HIS18:O D:VAL61:N - D:TRP57:O D:LYS64:N - D:ARG60:O D:LYS64:NZ - :UNK1	3.01247 3.02725 2.94629 3.9367
27	HSPA8	BaP	-4.08	1.02 mM	Gly230 Lys271 Arg272 Ser275 Gly339 Ser340 Arg342 Ile343	A:LYS271:NZ - A:GLY230:O A:LYS271:NZ - A:ASP234:OD1 A:ARG272:NE - A:GLU268:OE1 A:SER275:N - A:ARG272:O A:SER340:N - A:LEU200:O A:SER340:OG - A:GLY201:O A:ARG342:NH1 - A:ASP366:OD2	3.01822 2.83873 2.89543 3.02904 3.15856 2.76197 2.78272
28	ITGA1	BaP	-4.05	1.08 mM	Lys300 Thr307 Glu308 Phe311 Phe312 Asn313 Ile323 Arg330	A:LYS309:N - A:THR307:O A:PHE311:N - A:THR307:O A:ARG330:NH1 - A:LYS309:O A:ASN313:N - :UNK1 A:ASN313:N - :UNK1	3.19446 2.87875 3.11426 3.85053 3.85758

29	HIST1H2AD	BaP	-4.01	1.15 mM	Leu55 Leu58 Thr59 Ile62 Leu63	C:THR59:N - C:LEU55:O C:THR59:OG1 - C:LEU55:O C:ILE62:N - C:LEU58:O C:ILE62:N - C:THR59:O C:LEU63:N - C:THR59:O C:THR59:OG1 - :UNK1 C:THR59:OG1 - :UNK1 C:THR59:OG1 - :UNK1	2.87453 2.89841 3.00939 3.15014 2.93671 3.00563 3.54203 3.56045
30	POLR2B	BaP	-4	1.17 mM	Phe1159 Ile1167 Ala1168 Pro1169 Arg1170	B:PRO1169:CD - B:ILE1167:O B:ARG1170:N - :UNK1	3.66115 3.62667
31	GNG11	BaP	-3.8	1.63 mM	Glu39 Ile40 Tyr43	:TYR43:H - :GLU39:O :ILE44:H - :ILE40:O :TYR43:H - :GLU39:O	2.01642 2.05099 2.01642
32	ITGAX	BaP	-3.71	1.92 mM	Ala736 Asp737 Ala738 Gln739 Arg740 Tyr741	A:GLN739:NE2 - A:TYR741:O A:GLN739:NE2 - :UNK1 A:ARG740:N - :UNK1 A:ARG740:N - :UNK1	2.61208 3.77116 3.6276 3.67686
33	CTNNB1	BaP	-3.69	1.98 mM	Leu279 Val283 Leu286 Thr298 Leu301 Gln302 Ala305	A:THR298:OG1 - A:LEU294:O A:GLN302:N - A:THR298:O A:GLN302:N - A:ASP299:O A:ALA305:N - A:LEU301:O A:THR298:OG1 - :UNK1	2.81009 3.00803 3.06266 2.95985 3.93597
34	HLA-E	BaP	-3.57	2.43 mM	Arg79 Thr80 Gly83 Tyr84 Lys146	A:THR80:N - A:VAL76:O A:THR80:N - A:ASN77:O A:THR80:OG1 - A:VAL76:O A:GLY83:N - A:ARG79:O A:GLY83:N - A:THR80:O A:TYR84:N - A:THR80:O A:TYR84:N - A:LEU81:O A:LYS146:NZ - A:TYR84:OH A:THR80:OG1 - :UNK1 A:THR80:OG1 - :UNK1 A:THR80:OG1 - :UNK1	2.85324 3.18202 2.82884 3.14672 2.98344 2.96049 3.11681 3.12476 3.75066 4.10035 3.08187
35	CDH1	BaP	-3.4	3.19 mM	Thr141 Tyr142 Asn143 Leu196 Gln197	C:TYR142:N - C:THR141:OG1 C:ASN143:N - C:ASN143:OD1 C:GLY198:N - C:LEU196:O C:GLU199:N - C:LEU196:O C:TYR142:N - :UNK1	3.12833 3.07176 3.05958 3.1935 3.65885
36	JUN	BaP	-3.18	4.63 mM	Arg272 Glu275 Arg276 Arg279 Leu280	A:GLU275:N - A:ARG272:O A:ARG276:N - A:ARG272:O A:ARG279:N - A:GLU275:O	3.16704 2.96858 3.01565

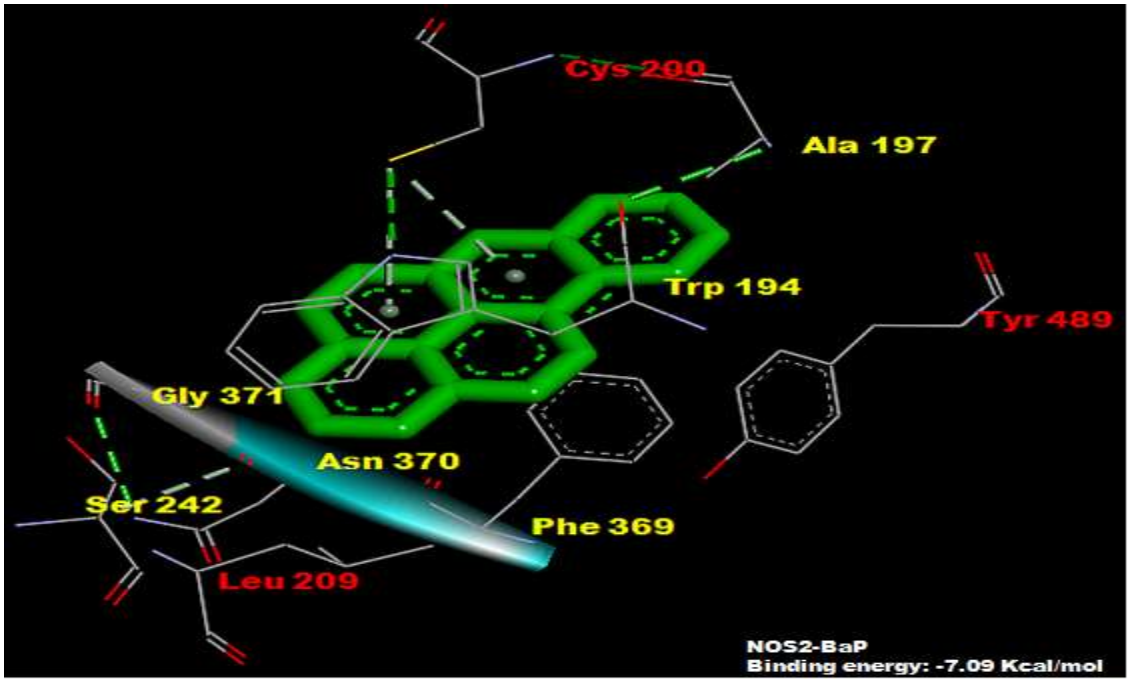
37	EGFR	BaP	-3.14	5.01 mM	Ile316 Gly317 Ile318 Asp323 Ser324 Leu325 Asp344 His346	A:ASP323:N - A:ASP323:OD1	3.1527
38	HIST1H2BD	BaP	-3.07	5.59 mM	Val42 Tyr43 Leu46 Lys47 His50 Pro51 Asp52 Thr53 Gly54 Ile55 Met60	:LEU46:H - :VAL42:O :LYS47:H - :TYR43:O :HIS50:H - :LEU46:O :THR53:H - :HIS50:O	1.95811 2.03921 1.6828 1.8177



a)



b)



c)

Fig 17: Pictorial representation of top 3 proteins interacting with BaPalong with their binding energies

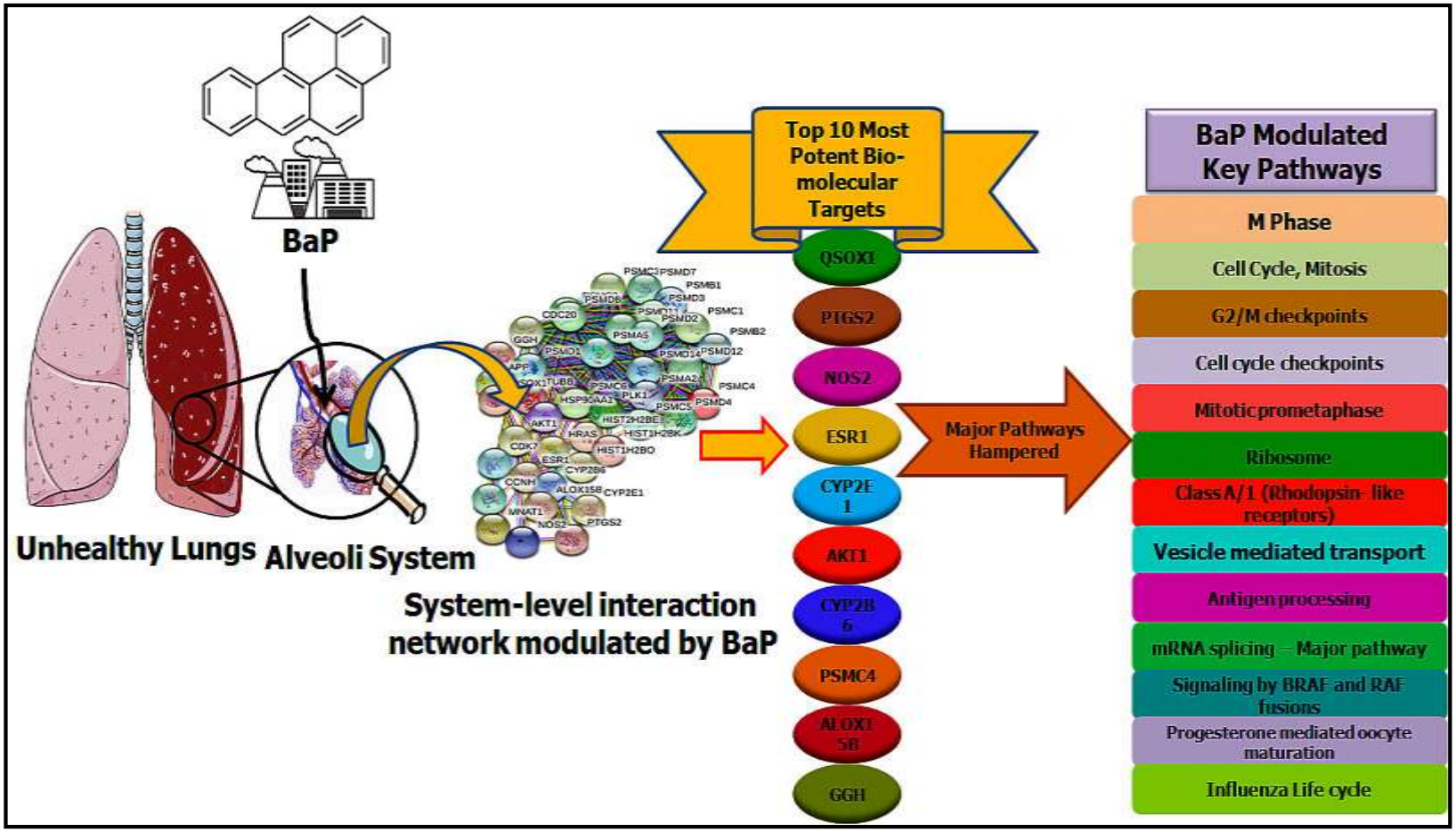


Fig 18: ComprehensiveGraphical representation of BaP's impact on systems level PIN with their respective key proteins, number of associated and enriched pathways.

4.1.5) Cell cycle regulatory model

For mathematical modeling of cell cycle, the model was first created using cell designer. SBML squeezer was used for generating the kinetics of reactions. For kinetics of some reactions, biomodel BIOMD0000000941 (Gerard and Goldbeter, 2010) was also referred. The concentration of BaP was taken from the work of Agarwal T (2009). Fig 19 is the pictorial representation of the cell cycle model developed for BaP. For the ease of handling of data we simplified the cell cycle model (Fig 20) and then further used that for our analysis purpose. In modeling the impact of BaP on cell cycle, we have used PTGS2 (second biomolecular target of BaP) as QSOX1 (first biomolecular target of BaP) and NOS2 (third biomolecular target of BaP) are not directly related with cell cycle regulation machinery. QSOX1 and NOS2 have been reported to have important role in development of cancer (Vannini *et al*., 2015 and Knutsvik *et al*., 2016).

PTGS2/COX2 acts as a regulator of cell cycle regulatory machinery by increasing the phosphorylation of p27 kip1 which plays a key role in G1-S transition of cell cycle. BaP upregulates COX2 gene (T3DB), increasing its expression. Increased levels of COX2 increases the phosphorylation rate of p27 increases the cell proliferation rate and cells start dividing rapidly causing cancer. Due to changes in the concentration of available p27, fluctuations in the concentrations and oscillations of Cyclin E CDK2 complex has been observed (table 3 and table 4). Furthermore, there are various research articles present which prove that high concentration of PTGS2 can be used as a potent

biomarker in various cancers like small cell lung cancer and colorectal cancer. Detailed table of change in concentrations with time has been presented in annexure 5.

For analysis, COPASI software was used. Initially the time course analysis was done without BaP and normal graph was obtained (fig 21). The concentration changes in all the molecules were recorded and then time course analysis was done in presence of BaP and variations from normal graph were obtained (fig 22a and 22b)

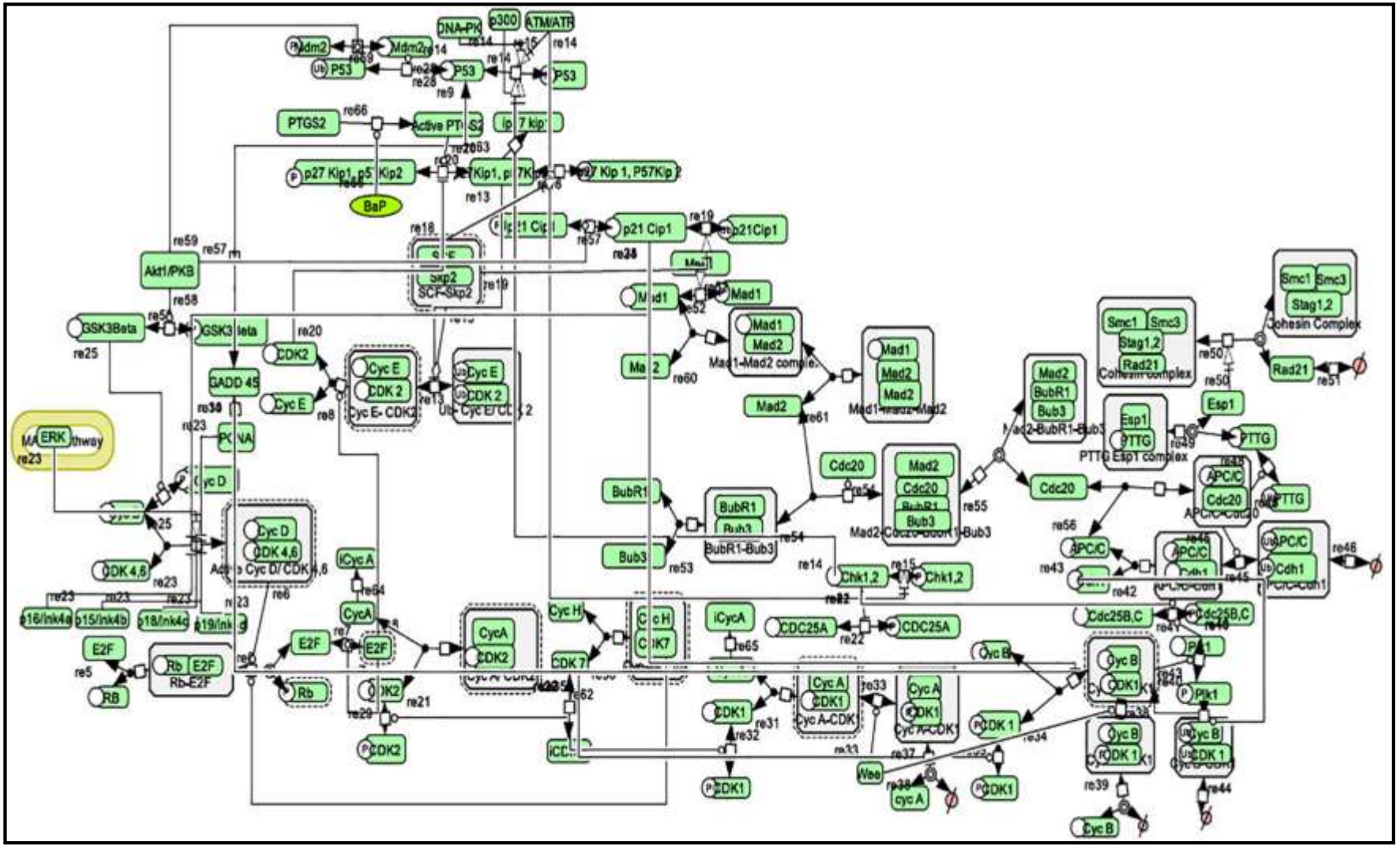


Fig19: Cell cycle designed by cell designer for study of the impact of BaP on cell cycle regulatory machinery

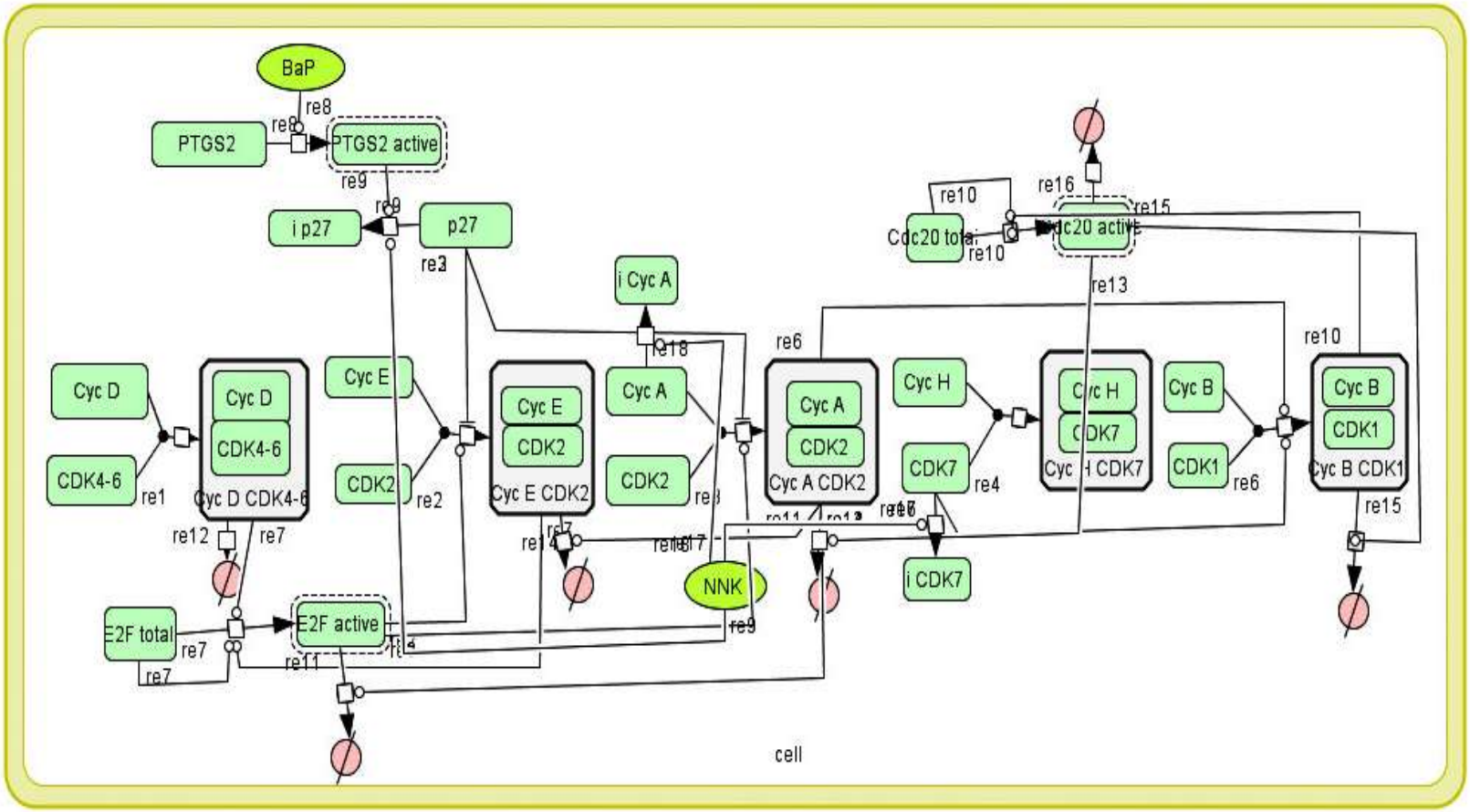


Fig20: Simplified cell cycle designed by cell designer for study of the impact of BaP on cell cycle regulatory machinery

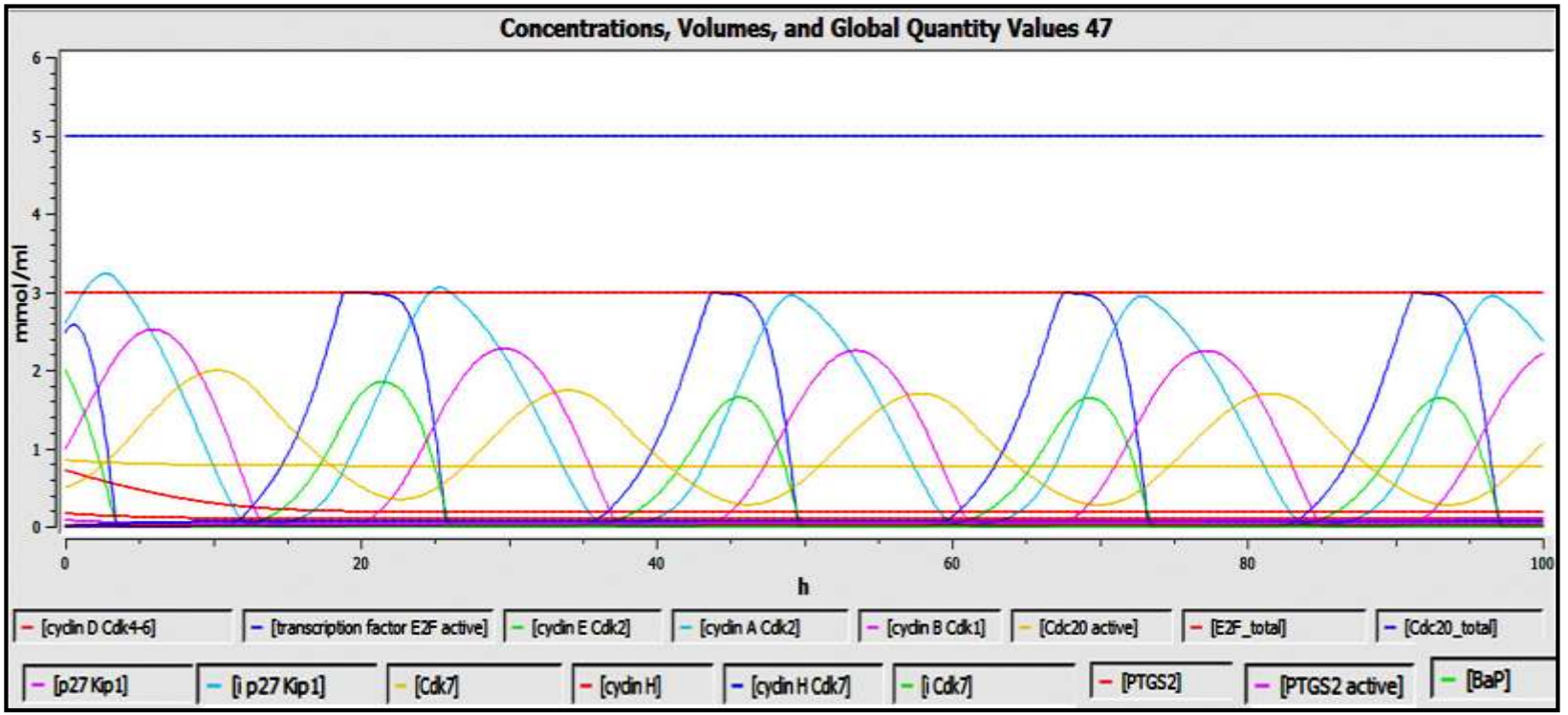


Fig 21:Cell cycle graph in normal condition when BaP and NNK are absent

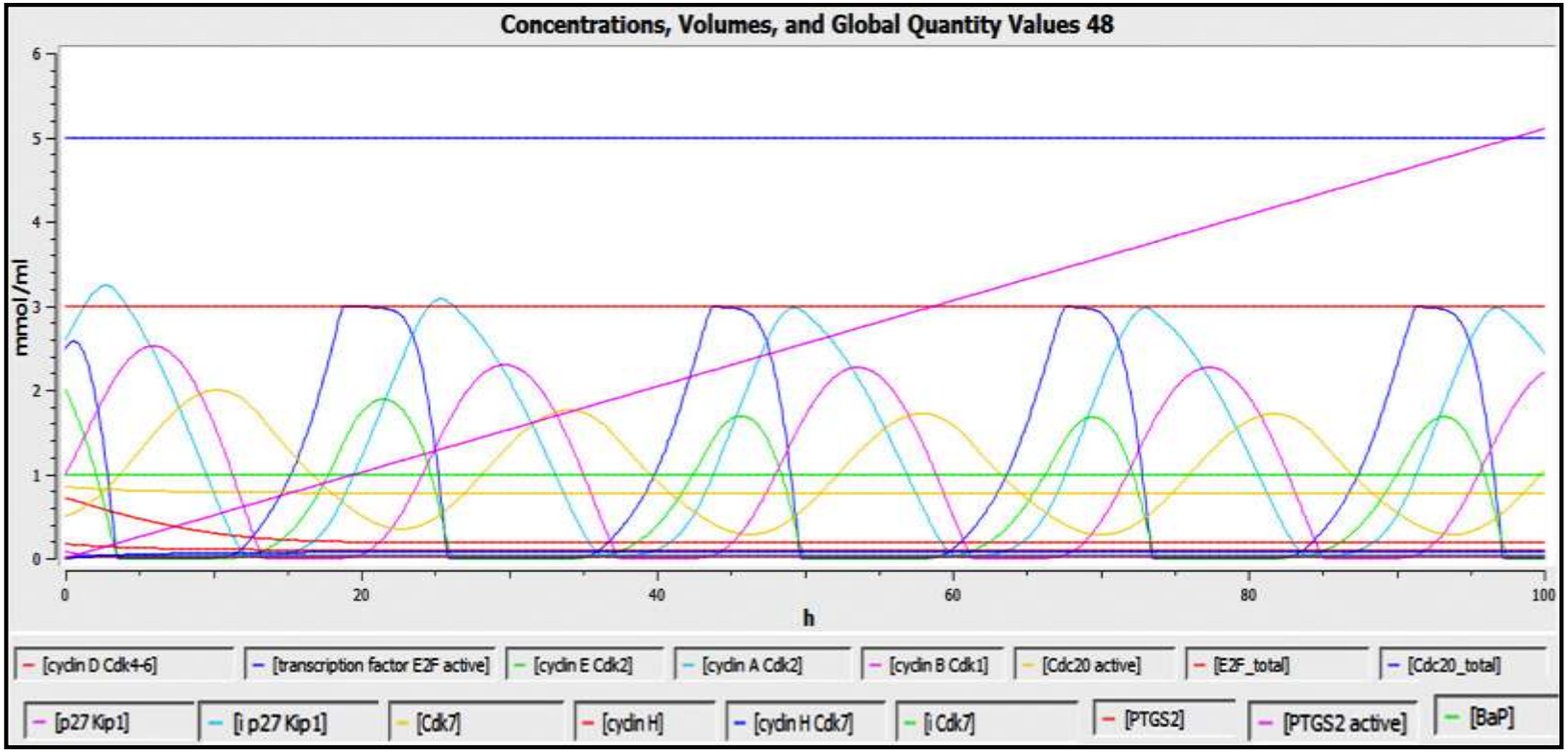


Fig 22a: BaP inhibited graph: A steep rise in the concentration of PTGS2 active can be easily noticed.

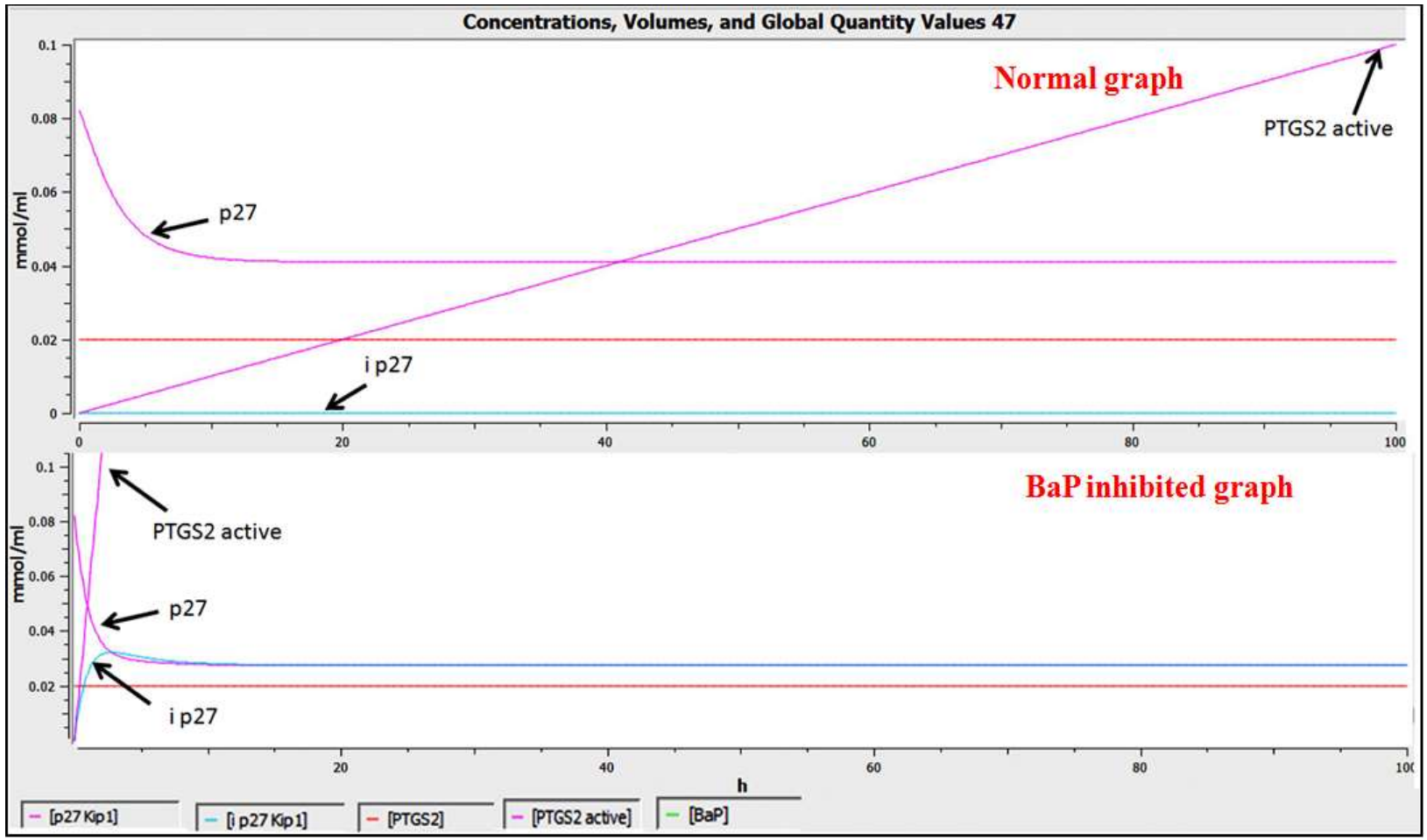


Fig 22b: Zoomed in comparative graph of BaP inhibited PTGS2 and p27

Table 3: Concentration based comparisonof species when PTGS2 is upregulated by BaP

Species	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited
	0 hrs		5 hrs		10 hrs		15 hrs		20 hrs		25 hrs	
PTGS2 active	0	0	0.005	0.255	0.01	0.51	0.015	0.765	0.02	1.02	0.025	1.275
i p27	0	0	0	0.0305	0	0.028	0	0.0275	0	0.0273	0	0.0273
cyc E CDK2	2	2	0.0015	0.0009	0.0025	0.0021	0.2745	0.2806	1.6895	1.7272	0.7512	0.8026
cyc A CDK2	2.6	2.6	2.7538	2.7588	0.7833	0.7864	0.058	0.058	1.1831	1.1934	3.0369	3.0602
cyc B CDK1	1	1	2.4661	2.4688	1.6007	1.6043	0.0001	0.0001	0.0441	0.0451	1.3012	1.3089
E2F active	2.4855	2.4855	0.0024	0.0024	0.0076	0.0075	0.9027	0.9056	2.9882	2.9885	1.2349	1.3001
Cdc20 active	0.5	0.5	1.2705	1.2709	1.9928	1.9952	1.3053	1.3084	1.5096	0.5115	0.47712	0.4803

Table 4: Time based comparison of species when PTGS2 is upregulated by BaP

Species		Conc. at first max (mmol/ml)	time (hrs)	Conc. at first min (mmol/ml)	time (hrs)	Conc. at second max (mmol/ml)	time (hrs)	Conc. at second min (mmol/ml)	time (hrs)	
PTGS2 active	Normal	Continuous increase in concentration with maximum of 0.1mmol/ml at 100 hrs								
	Inhibited	Continuous increase in concentration with maximum of 5.1mmol/ml at 100 hrs								
I p27	Normal	Throughout at 0 mmol/ml								
	Inhibited	0.0321	2.8	Gradual decrease						
Cyc A CDK2	Normal	3.2371	2.7	0.0154	13	3.0558	25.4	0.0131	37	
	Inhibited	3.2393	2.7	0.0154	13	3.0837	25.4	0.0133	37	
Cyc B CDK1	Normal	2.5194	6	4.65E-05	13.4	2.276	29.7	3.83E-05	37.5	(curve was almost flat from 13.4 hrs to 17 hrs)
	Inhibited	2.2782	8	5.08E-05	13.3	2.2926	29.4	3.53E-05	37.5	curve was almost flat from 13.4 hrs to 14.6 hrs
Cyc E CDK2	Normal	1.8501	21.5	0.0015	5.7	1.6514	45.5	0.00209	27	(curve was almost flat from 4.7hrsto 7hrs)
	Inhibited	1.8901	21.5	0.0009	4.6	1.6884	45.7	0.0014	27	Curve was almost flat between 4 hrs to 6 hrs
E2F active	Normal	2.5786	0.6	0.0024	4.3	2.9926	19	0.0008	26.5	(curve was almost flat from 4.1 to 5hrs)
	Inhibited	2.5789	0.6	0.0024	4.3	2.9931	18.9	0.0008	26.5	(curve was almost flat from 3.8 to 5.2hrs)
Cdc20 active	Normal	1.9954	10.3	0.3474	22.7	1.7401	34.1	0.2799	46.3	
	Inhibited	1.9979	10.3	0.3494	22.7	1.7604	34.1	0.2823	46.3	

4.1.6) Protection by carbon nanoparticles

Carbon nanoparticles were designed using nanotube modeler (Fig 23) and then with the help of Material studio, the structures were optimized. They were then docked with BaP(Fig 24) and with the adsorption load was calculated using Blend of material studio (fig 25). The binding energy of BaP was highest with MWCNT (-11.86 Kcal/mol) and also has the highest adsorption capacity of 11 molecules/NT. BaP has least binding energy with fullerene (-3.9 Kcal/mol) and also the least adsorption capacity (4 molecules/NT) as shown in table 5.

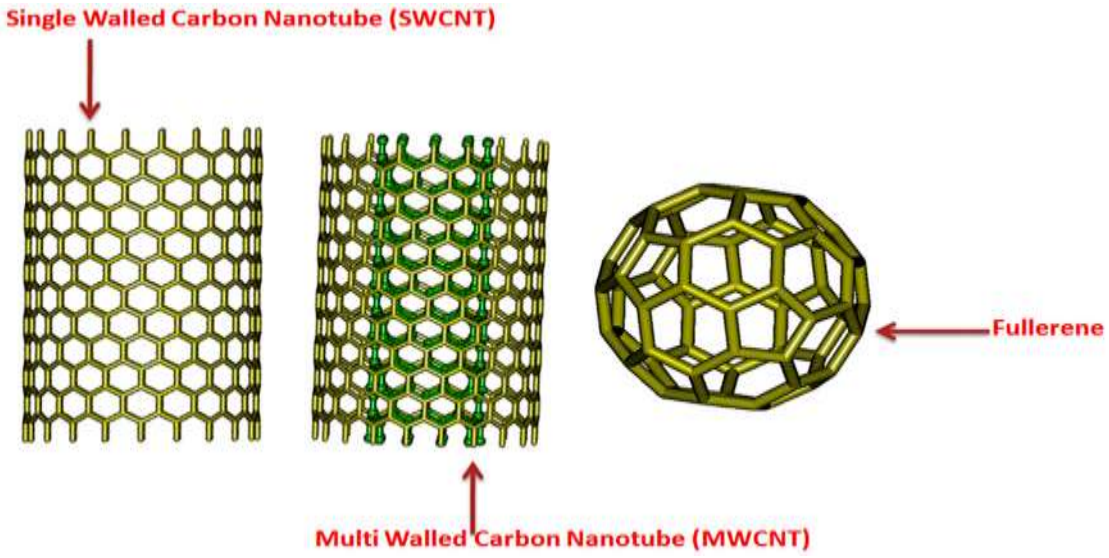
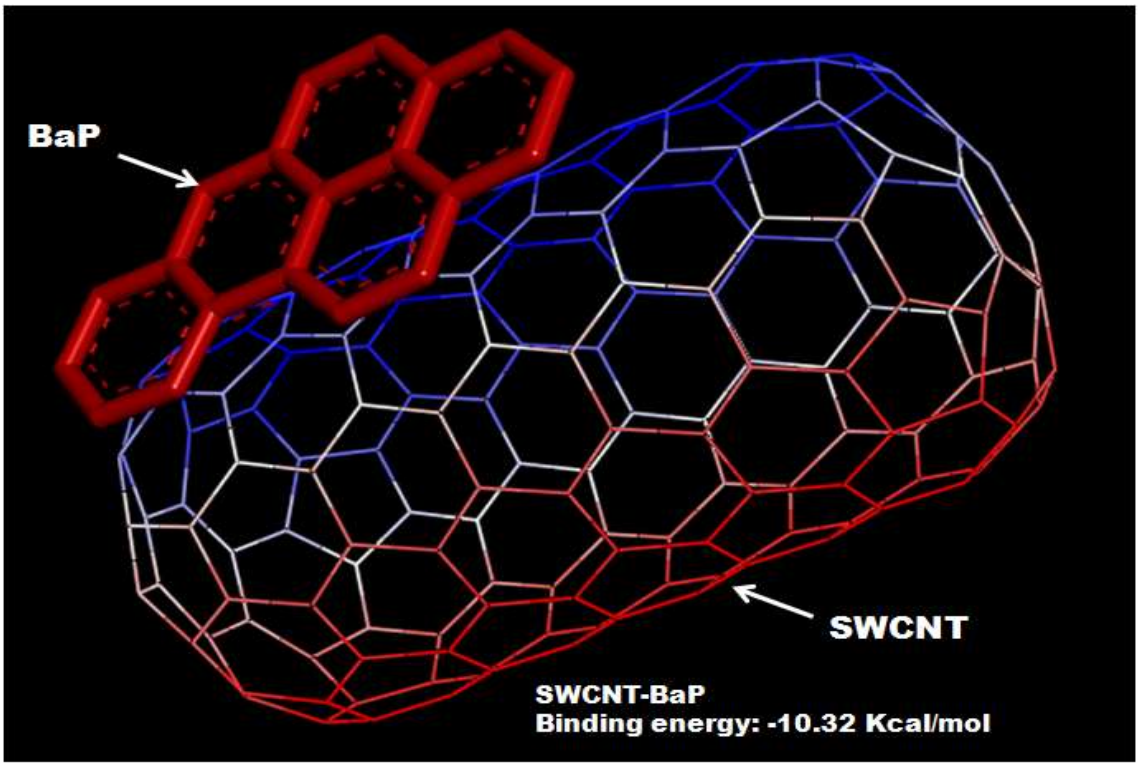
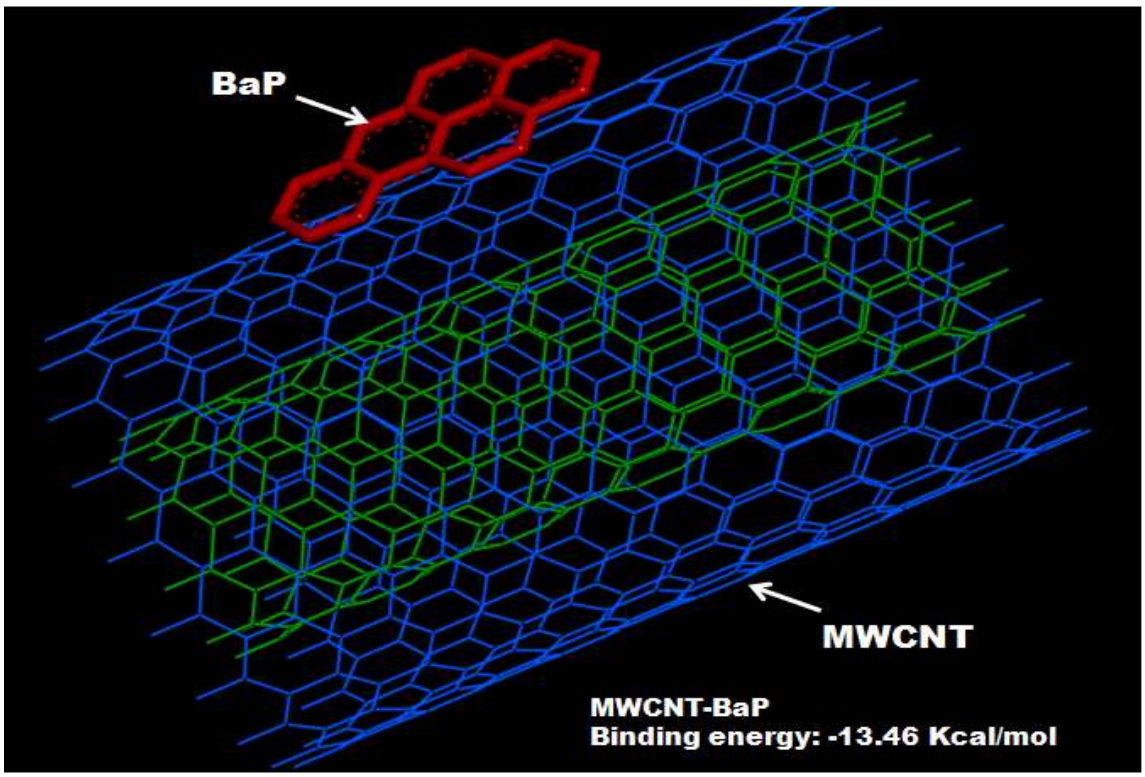


Fig 23: Structures of carbon nanotubes (SWCNT and MWCNT) and Fullerene


BaP
SWCNT
SWCNT-BaP
Binding energy: -10.32 Kcal/mol


BaP
MWCNT
MWCNT-BaP
Binding energy: -13.46 Kcal/mol

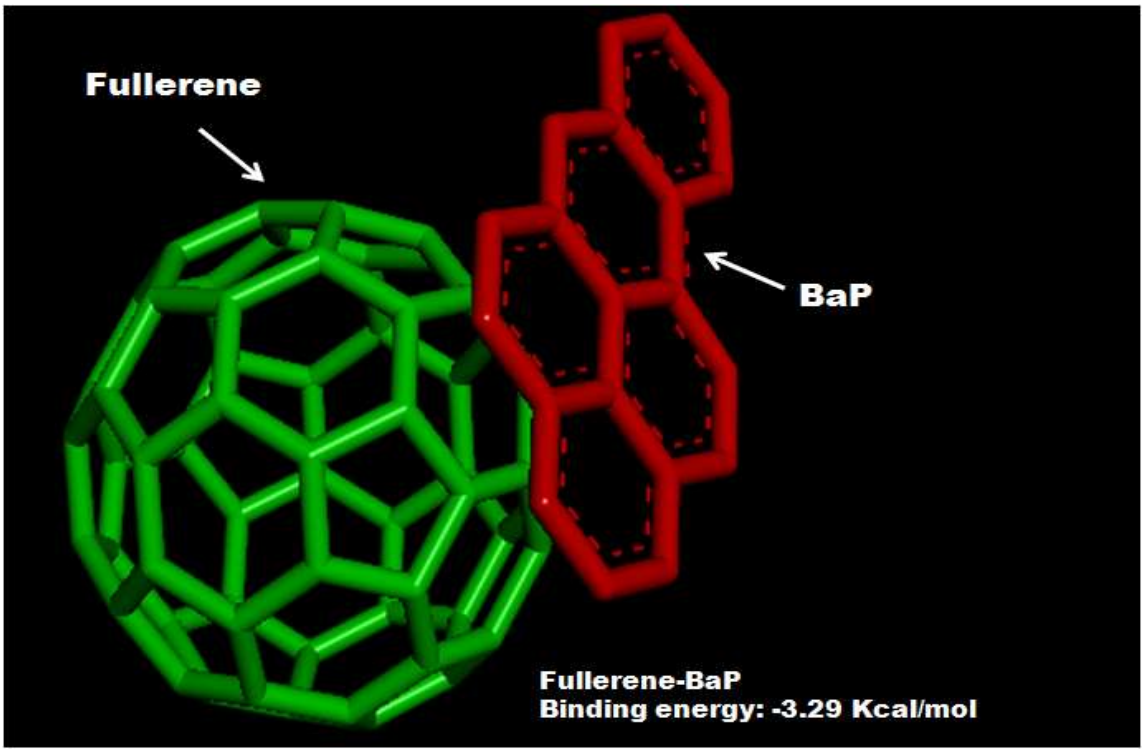
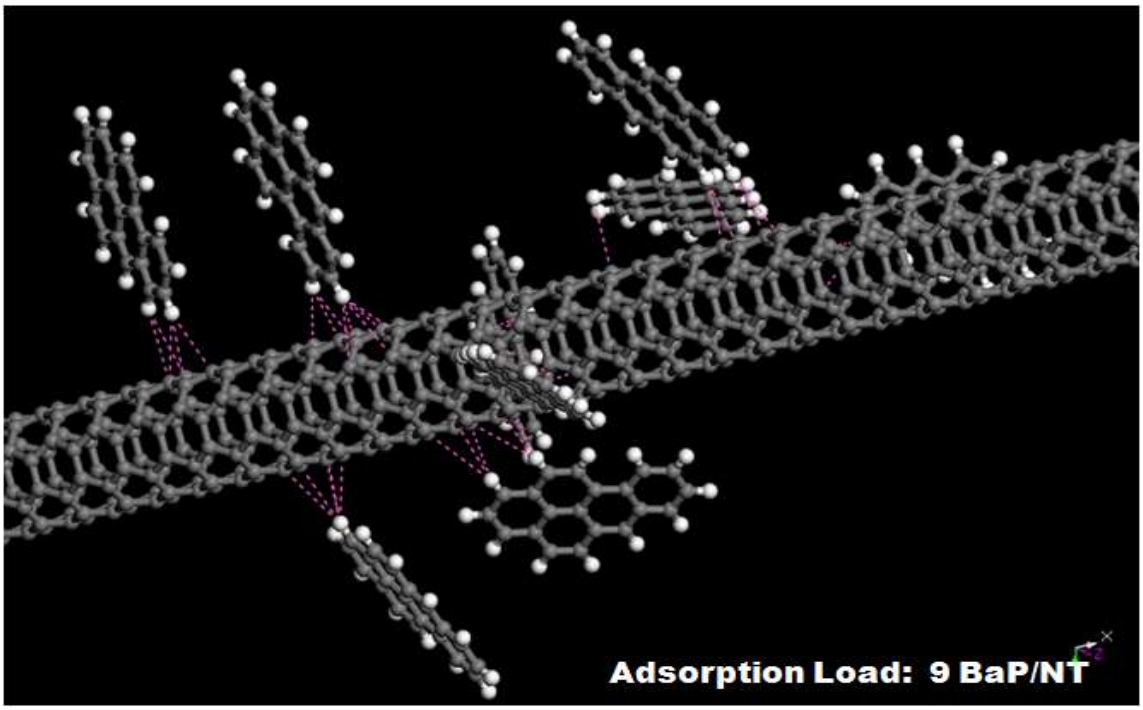
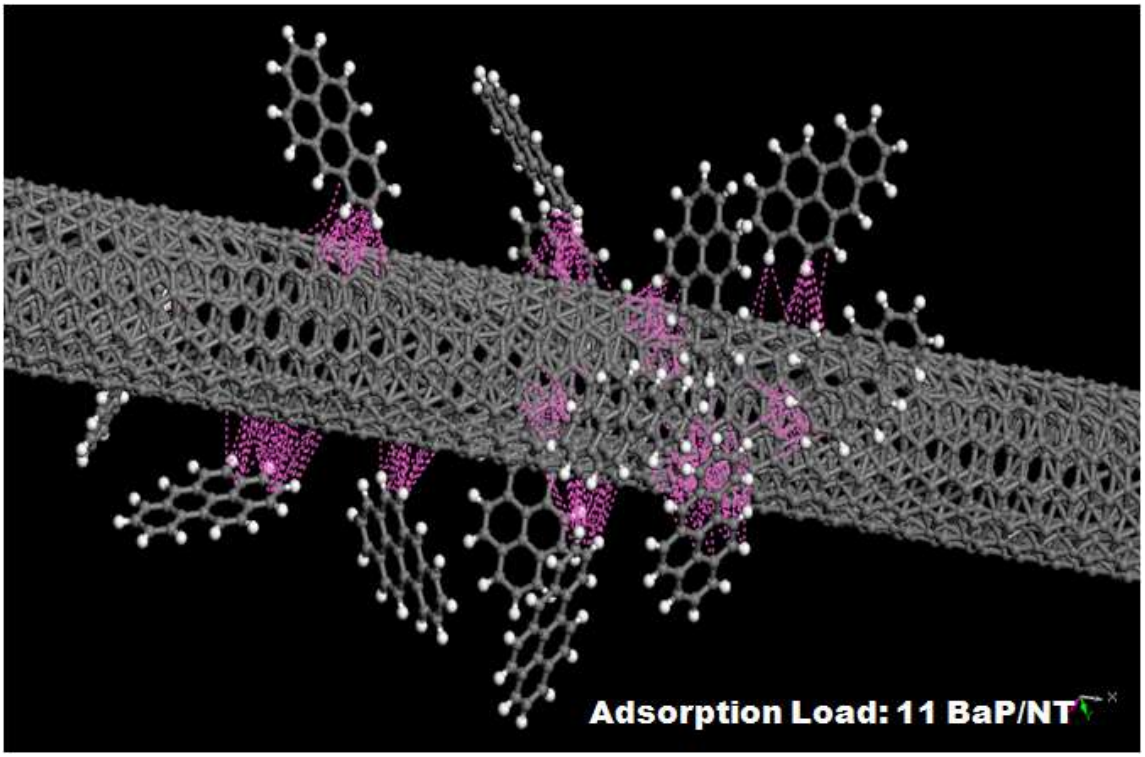


Fig 24: BaP interacting with SWCNT, MWCNT and Fullerene. The binding energies of SWCNT, MWCNT and Fullerene were-10.32 Kcal/mol, -13.46 Kcal/mol and -3.29 Kcal/mol respectively. MWCNT showed the maximum binding energy against BaP.

Adsorption Load/ Adsorption Capacity of BaP on SWCNT, MWCNT and Fullerene





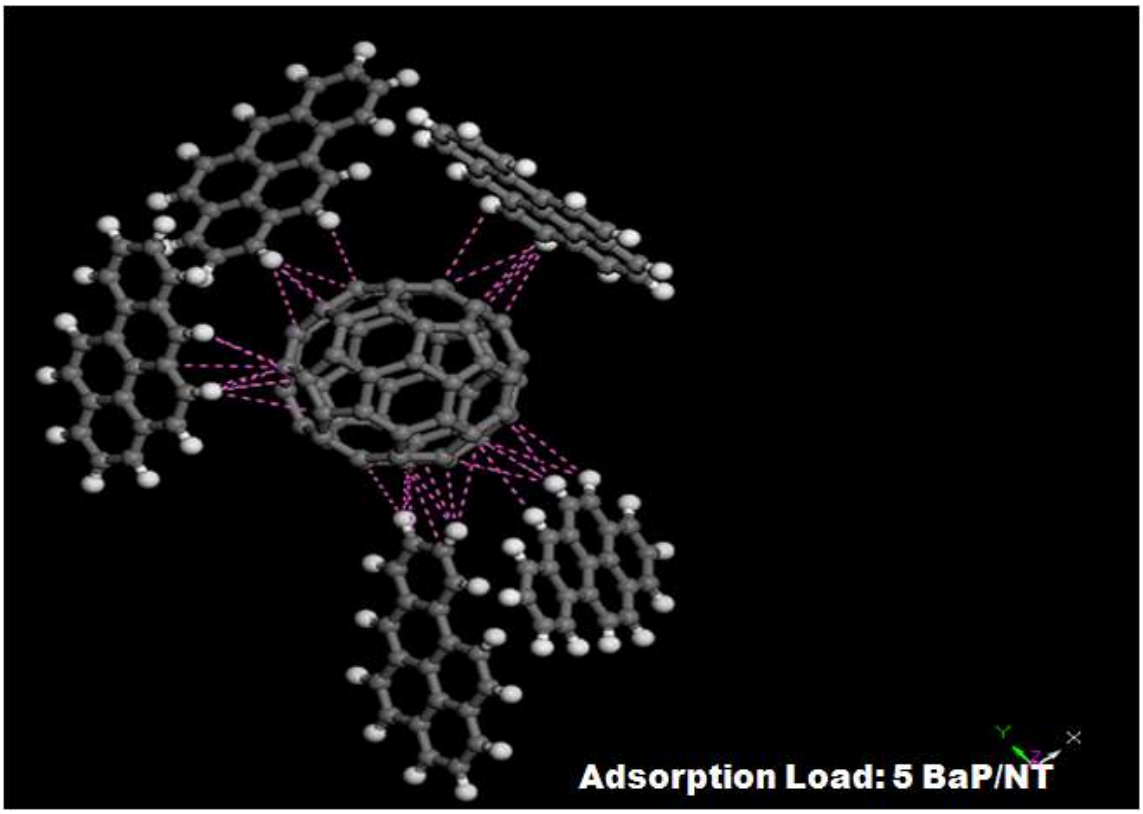


Fig 25: Pictorial representation of BaP getting adsorbed on SWCNT, MWCNT and Fullerene

Table 5) Comparative chart of binding energy and adsorption load of BaP on SWCNT, MWCNT and Fullerene along with the binding energy of QSOX1

Carbon Nanomaterial	Ligand	Binding Energy (Kcal/Mol)	Adsorption Load (Molecules/NT)
SWCNT	BaP	-10.32	9
MWCNT	BaP	-13.46	11
Fullerene	BaP	-3.29	5
QSOX1	BaP	-8.01	-

Important findings for BaP

- *BaP upregulates 1722 genes and downregulates 2363 genes.*
- *The combined network of BaP rewired PPIN had a clustering coefficient of 0.559, Characteristic pathlength was 4.165. Average number of neighbors were 13.460, suggesting that the network is a real-world network.*
- *65 clusters were generated using MCODE plugin of cytoscape software which had 411 seed proteins.*
- *38 most probable biomolecular targets were identified based on the various topological peoperties which were used as screening criteria.*
- *On enrichment analysis, most of the proteins were found to be involved in the cell cycle regulatory pathways.*
- *On molecular docking analysis of the 38 most probable biomolecular targets, QSOX1, PTGS2 and NOS2 were the top 3 biomolecules with highest binding energies of -8.01 Kcal/mol, -7.82 Kcal/mol and -7.09 Kcal/mol respectively.*
- *On molecular docking analysis and adsorption load analysis, MWCNTs showed highest binding energy of -13.46 Kcal/mol and an adsorption load of 11 molecules/NT followed by SWCNT with -10.32 Kcal/mol binding energy and 9 molecules/NT adsorption load. Fullerenes showed minimum binding energy and adsorption capacity against BaP.*

4.2) Results for NNK:

Approximately 1320 research articles were scrutinized and 544 genes/proteins have been identified which show changes on interaction with NNK.

4.2.1) Network generation using STRING database:

Protein-protein interaction network for 544 NNK hampered proteins was generated using STRING-db software(fig 26). The network had 534 nodes and 2909 edges. The confidence score was kept at the highest of 0.9 and 50-50 connectors were added in both the shells. The average node degree score calculated by the server was 10.09 and the average clustering coefficient of whole interactome came out to be 0.501 with PPI enrichment value $< 1x\ 10^{-16}$.

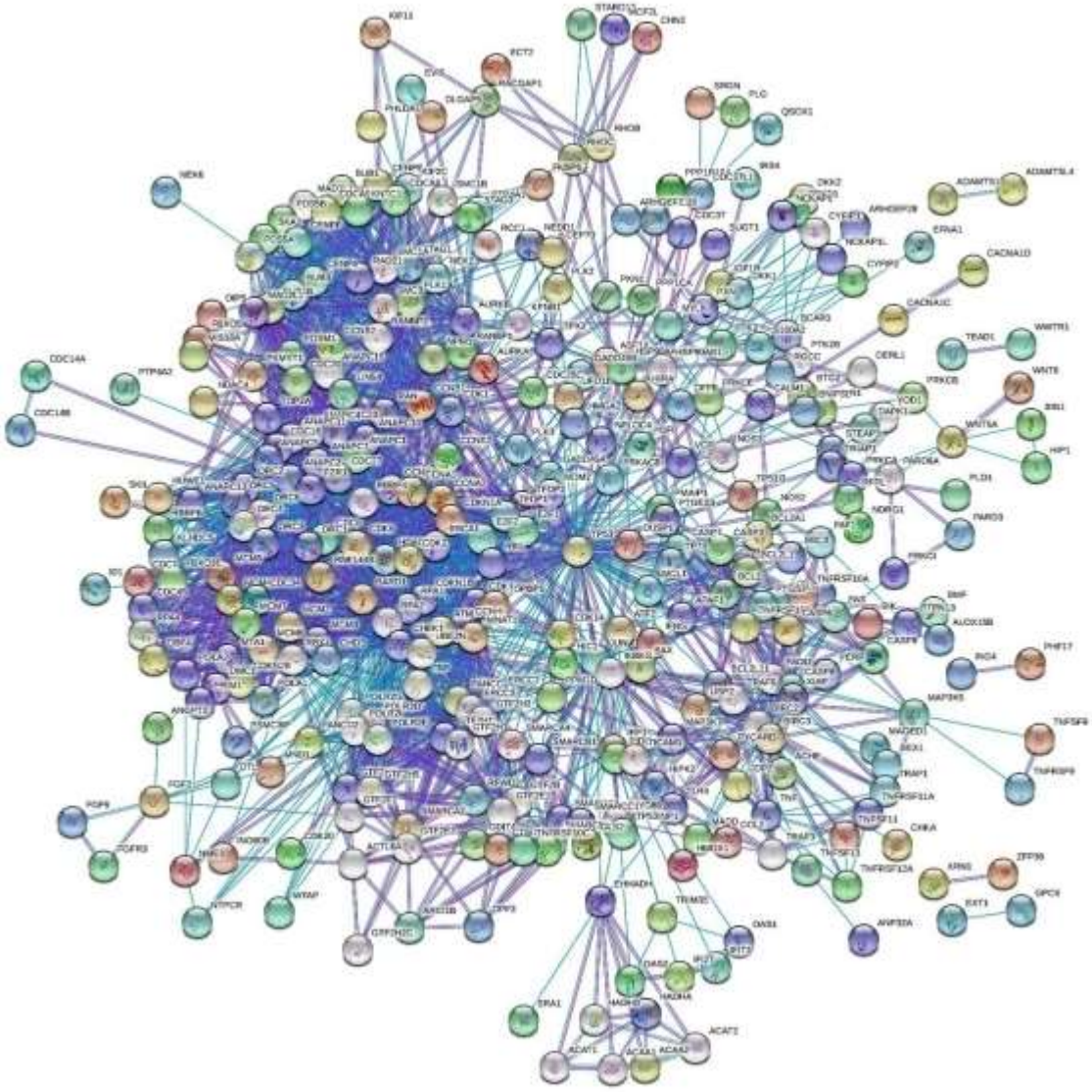


Fig 26: NNK rewired PPIN

4.2.2) Network analysis:

The network generated by STRING database was imported to Cytoscape software for the analysis of the topological properties using Network analyzer, an in-built app of Cytoscape. The topological properties analyzed were characteristic path

length distribution, node degree distribution, average neighborhood connectivity distribution and average clustering coefficient. Fig 27 represents the graph of shortest path length distribution. Shortest paths are the minimum steps that are required to transfer any information. The path length distribution calculated by network analyzer was 2.971. In the graph below, maximum frequency is allotted to 3 path length. Fig28 is the graph of node degree distribution following the power law. Node degree distribution is an important property of any network that helps in understanding the structure of the network.

Neighbourhood connectivity distribution is the average number of nodes connected with the neighboring nodes. The average number of neighbors present in the network was 15.940. The distribution curve was fitted with power law where $y = 19{,}838x^{-0.137}$ (R-squared 0.253). Fig 29 represents the neighborhood connectivity distribution curve.

The clustering coefficient of a network is the tendency of network to be divided into clusters. The average clustering coefficient in this network (Fig 30) was found to be 0.597. The curve was fitted with power law $y = 1.326\ x^{-0.277}$ (R-squared = 0.329).

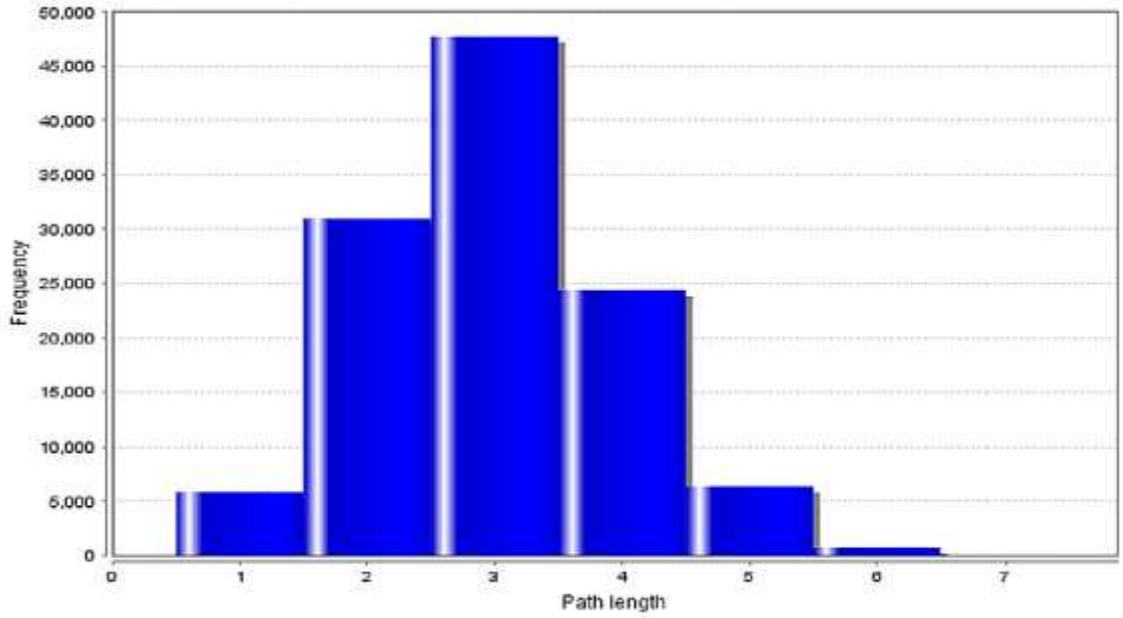


Fig 27: Shortest path length distribution. Approximately 47,000 nodes have the shortest path length of 3.

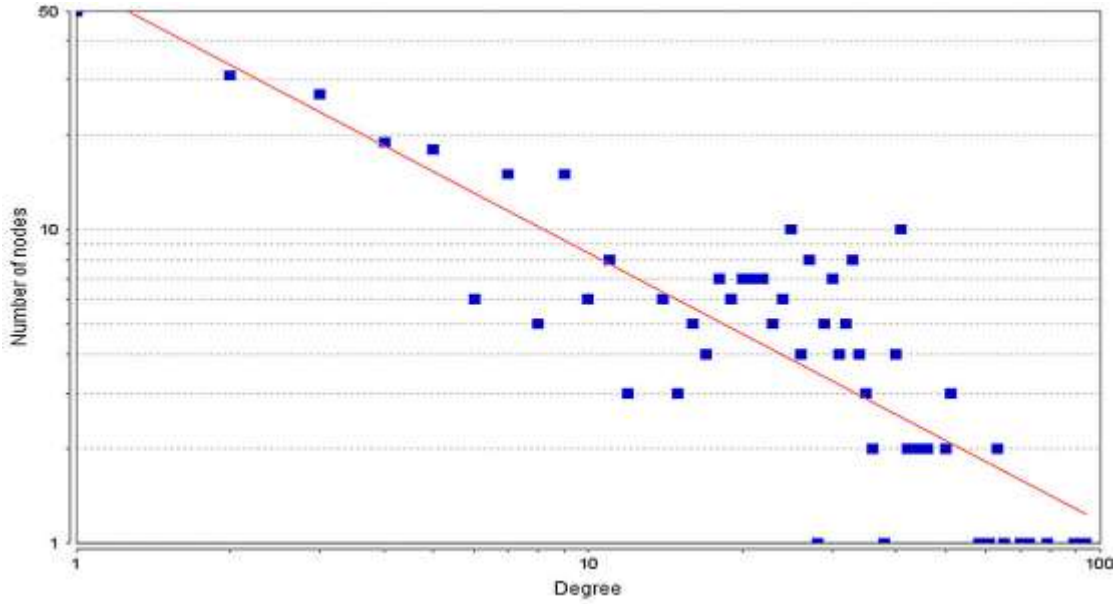


Fig 28: Node degree distribution (power law: $y = ax^{-b}$; a= 61.323, b= 0.861, r^2 = 0.824)

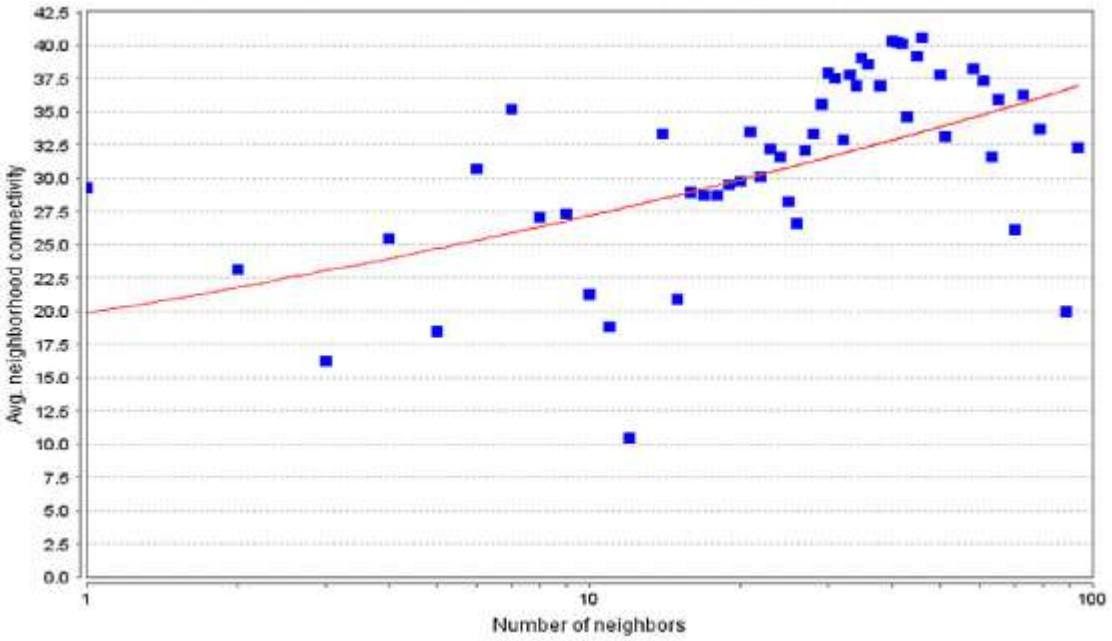


Fig 29: Neighborhood connectivity distribution (power law: $y= ax^{-b}$; a= 19.838, b= 0.137, $r^2 = 0.590$)

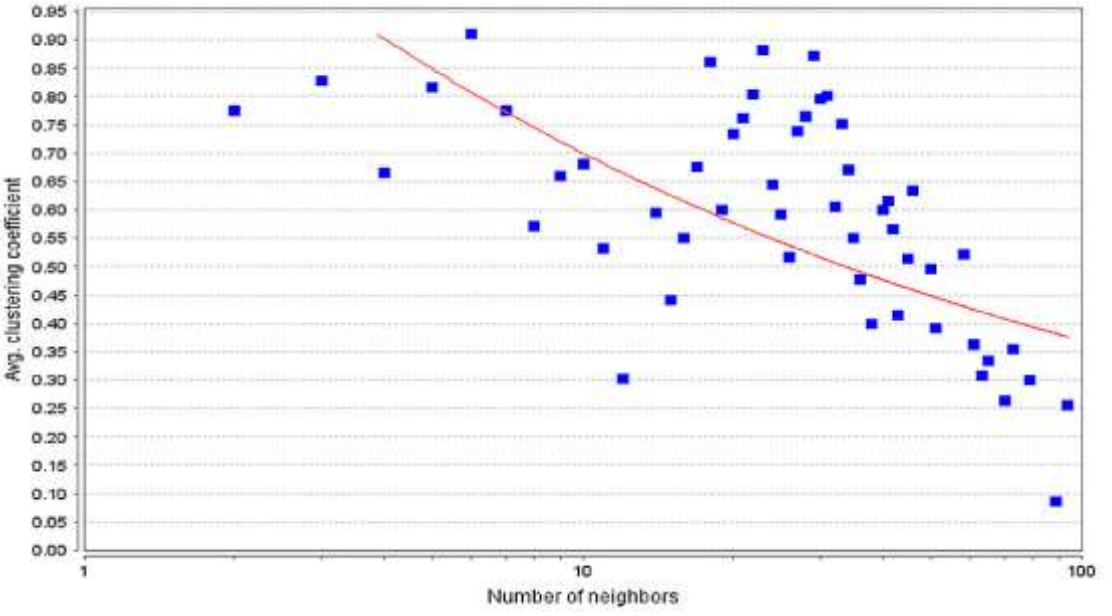


Fig 30: Avg. clustering coefficient (power law: $y= ax^{-b}$; a=1.326, b= 0.277, $r^2 = 0.285$)

4.2.3)Modulation and GO enrichment analysis

MCODE was used to generate clusters to remove the noise from the network. Annexure4 is the list of clusters generated and the finally selected seed proteins (the proteins in bold). A total of 19 clusters were generated by MCODE which had 115 seed proteins in them. Seed proteins identified from each cluster were again subjected to STRING database for PPIN generation. The confidence level was maintained at 0.9 and 50-50 connectors were added in the first and the second shells. Table4 shows the topological properties of the finally selected proteins (seed proteins). The most important topological properties on the basis of which further screening of the most probable biomolecule was done were degree, clustering coefficient, betweenness centrality and finally bottleneck score. The proteins were first sorted based on degree and then those proteins were selected that had high degree and clustering coefficient less than 0.5. The next criterion for sorting the proteins was high betweenness centrality followed by bottleneck scores. To further screen the proteins, median was applied as the data generated was discrete in nature. For NNK, 21 proteins finally emerged out to be the most potent biomolecular targets with high bottleneck score, high degree and clustering coefficient less than 0.5 as shown in table 6.

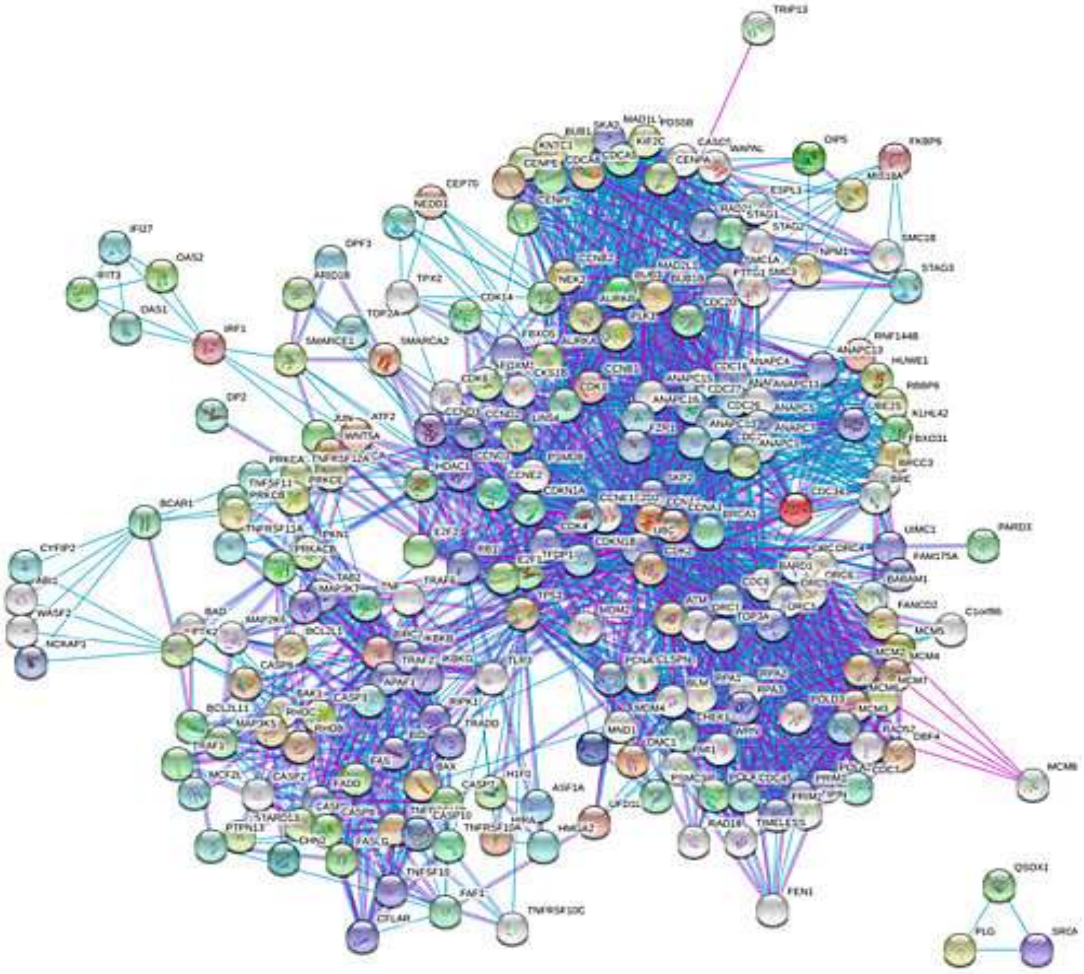


Fig 31: PPIN construction using seed proteins generated from the modulation of the NNK rewired PPIN

Table 6: Topological analysis of the final proteins selected from NNK seed rewired PPIN

Name	Betweenness	Bottleneck	Closeness	ClusteringCoefficient	Degree
CHEK1	835.45994	29	116.61667	0.36501	52
TP53	8007.14223	27	126.41667	0.19394	55
BRCA1	2686.48081	23	127.45	0.26346	65
CDK1	3705.28949	19	140.41667	0.32157	85

CDK4	935.26901	14	112.66667	0.42521	35
HSP90AA1	3657.32532	13	101.66667	0.27273	22
RPA2	1523.40238	9	125.78333	0.33978	64
ATM	1803.73545	8	115.36667	0.33718	40
TFDP1	426.33754	6	111.78333	0.46702	34
CDKN1B	1218.04331	4	120.41667	0.44245	50
CASP8	1842.98595	4	90.11667	0.44853	17
PYCARD	796	3	62.18333	0.33333	3
CCNA1	894.78122	3	125.2	0.39548	60
CCNB1	1208.39103	3	127.26667	0.40665	69
RPA1	1657.28981	2	126.45	0.33269	65
CDK2	1887.40725	2	134.2	0.34035	76
CHEK2	126.13416	2	95.95	0.3619	15
BID	577.03969	2	90.78333	0.4269	19
RB1	608.84911	2	107.66667	0.44138	30
PLK1	1284.13385	2	121.25	0.49545	56
CDK7	488.15015	2	108.78333	0.49733	34

GO enrichment was done using ClueGO plugin of cytoscape software, which helps in finding genes/proteins that are functionally linked together and helps in enhanced interpretation of the data. The pie chart below (fig 32) and bar chart (fig 33) shows that most of the proteins/genes (approximately 46%) that have got enriched belong to various phases of cell cycle regulation.

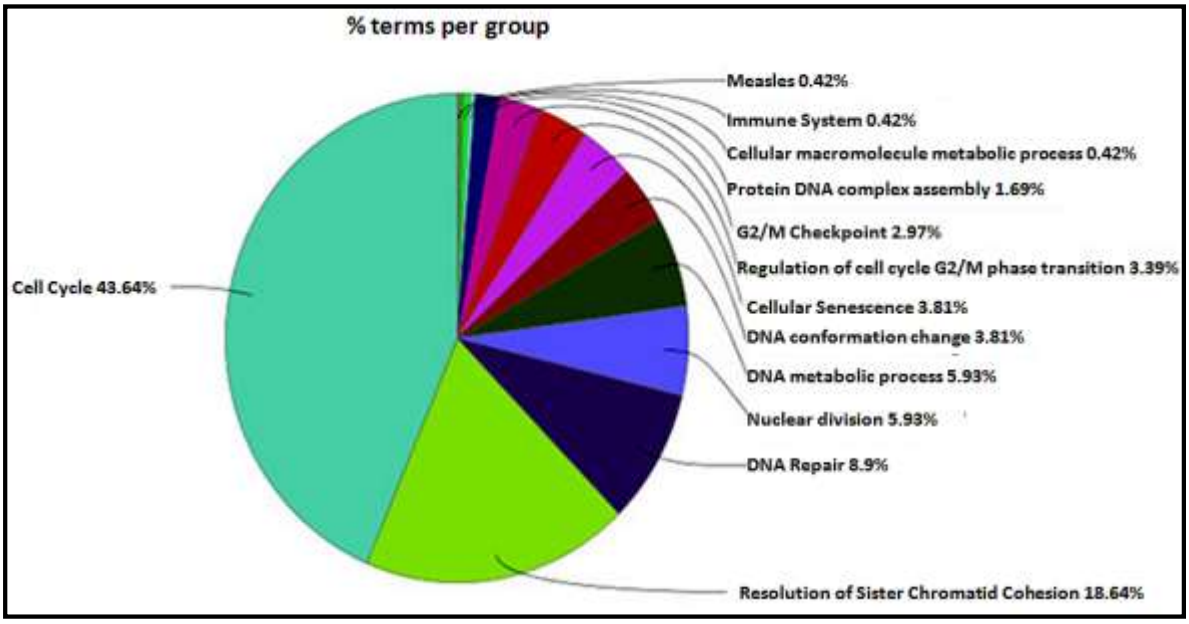


Fig 32: ClueGO results of GO functional enrichment of key proteins for NNK

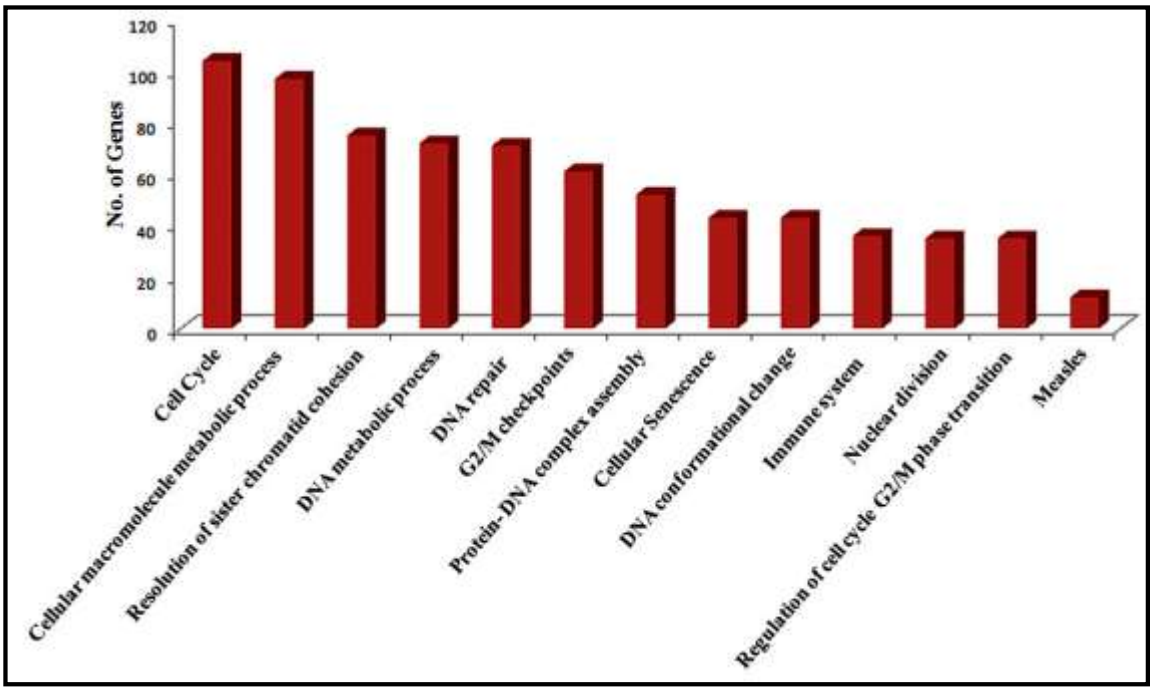


Fig 33: Number of genes involved in the pathways enriched by GO analysis.

Summary of Modulation and Enrichment analysis of NNK

- 19 clusters got generated.
- 115 seed proteins were obtained from the clusters.
- Enrichment analysis showed that most of the proteins were involved in cell cycle regulatory pathways.

4.2.4) Molecular docking simulation

Molecular docking simulation was used to pinpoint the most important biomolecular target of NNK. Total 21 molecules were docked and were arranged in decreasing binding energy values. On docking simulations, CDK7 showed the highest binding efficiency for NNK (-5.93 Kcal/mol) followed by CCNA1, CDKN1B and CASP8 with -5.60 Kcal/mol, -5.42 Kcal/mol and -5.35 Kcal/mol binding energies respectively. CCNA1 is a connector protein while remaining others are seed proteins. In total there are 5 connectors protein in the final selected protein list. The table (Table 7) is the list of docked proteins along with their binding energies and Ki values. Fig 34 is the pictorial representation of the top 3 docked biomolecular targets of NNK.

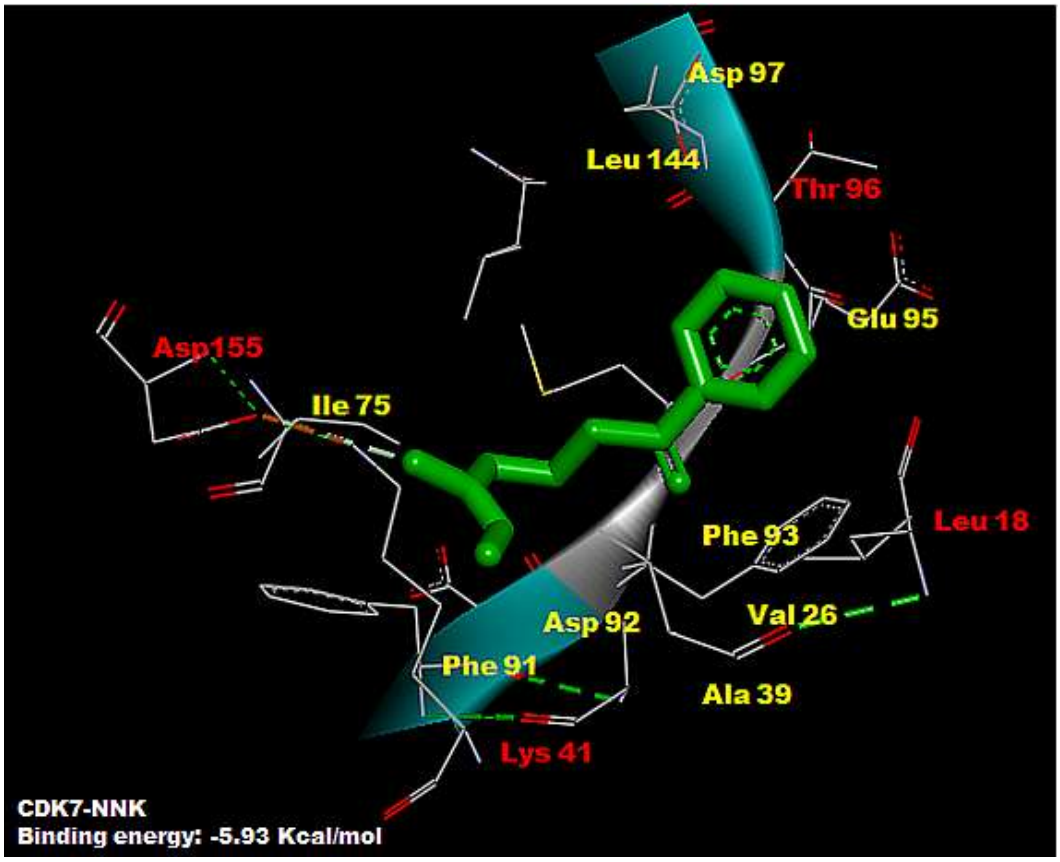
Fig 35 is the collective graphical representation of the impact of NNK on system level protein interaction network with their respective key regulatory proteins along with the enriched pathways.

Table 7: Docking results of final proteins selected from NNK rewired PPIN

S.No.	Protein	Ligand	Binding Energy (Kcal/ Mol)	Ki	Binding Residues	H-Bond	Distanc
1.	**CDK7**	NNK	**-5.93**	45.31 uM	Leu18, Val26, Ala39, Lys41, Ile75, Phe91, Asp92, Phe93, Met94, Glu95, Thr96, Asp97, Leu144, Asp155	CDK7:MET94:N -:NNK:O7	2.86224
2.	**CCNA1** (connector)	NNK	**-5.60**	79.09 uM	Cys97, Gly98, Gln99, Gly100, Val164, Asp165, Thr166, Gly167, Thr168, Leu169, Lys170, Leu173, Tyr218	:GLY98:H - :NNK:O7 :GLY167:H - :NNK:N10 :THR168:H - :ASP165:O :LYS170:H - :NNK:N14 :LYS170:H - :NNK:O15 :NNK:N14 - :THR168:O	1.83311 1.98057 2.11393 2.36928 1.99807 3.02246
3.	**CDKN1B**	NNK	**-5.42**	106.27 uM	His573, Lys574, Pro575, Leu576, Glu581, Trp582, Gln583, Glu584	CDKN1B:GLN583:N - :NNK:N14 CDKN1B:GLN583:N - :NNK:O15 :NNK:N14 - CDKN1B:GLN583:O	2.84311 2.79833 2.96186
4.	CASP8	NNK	-5.35	119.75 uM	Lys2224, Tyr2226, Gln2227, Asp2308, Gly2350, Lys2351, Pro2352, Asp2398, Arg2471, Lys2472	CASP8:LYS2351:HZ1 - :NNK:N10 CASP8:ARG2471:HH21 - :NNK:N2 CASP8:ARG2471:HH21 - :NNK:N14 CASP8:ARG2471:HH21 - :NNK:O15 CASP8:LYS2472:HZ1 - :NNK:O7	1.9527 2.48624 2.19123 2.25672 1.82255
5.	CHEK2 (connector)	NNK	-5.35	120.25 uM	Ser228, Gly229, Ala230, Cys231, Gly232, Val234, Lys249, Leu301, Thr367, Asp368	CHEK2:CYS231:N - :NNK:N10 CHEK2:GLY232:N - :NNK:N10 CHEK2:LYS249:NZ - :NNK:N14 CHEK2:ASP368:N - :NNK:O15	3.11295 3.11412 3.20168 2.86381
6.	PLK1	NNK	-5.21	152.98 uM	Lys413, Trp414, Val415, Asp416, Leu490, Asn533, Lys540	PLK1:TRP414:N - :NNK:O7 PLK1:ASP416:N - :NNK:N14 PLK1:ASP416:N - :NNK:O15 PLK1:ASN533:ND2 - :NNK:N10	2.97424 2.94596 2.71702 3.10323
7.	BID (connector)	NNK	-5.13	174.27 uM	Leu21, Phe24, Gly25, Gln28, Leu39, Asp40, Leu42, Gly43, Arg86, Ala89, Arg90, Phe173	:NNK:N14 - BID:GLN28:OE1 BID:PHE24:HA - :NNK:O15 BID:ARG86:HA - :NNK:N10 :NNK:C3 - BID:GLN28:OE1 :NNK:C1 - BID:LEU39:O :NNK:C11 - BID:ARG86:O :NNK:O15 - BID:PHE24 :NNK:N14 - BID:GLN28:OE1 BID:PHE24:HA - :NNK:O15 BID:ARG86:HA - :NNK:N10 :NNK:C3 - BID:GLN28:OE1 :NNK:C1 - BID:LEU39:O :NNK:C11 - BID:ARG86:O :NNK:O15 - BID:PHE24	3.29101 2.94776 2.82375 3.41587 3.00214 3.32997 3.70686 3.29101 2.94776 2.82375 3.41587 3.00214 3.32997 3.70686
8.	HSP90AA1	NNK	-5.10	183.46 uM	Leu48, Asn51, Ser52, Ala55, Asp93, Ile96, Gly97, Met98, Asn106, Phe138, Thr184, Val186	HSP90AA1:ASN51:ND2 - :NNK:N14 HSP90AA1:ASN51:ND2 - :NNK:O15 HSP90AA1:THR184:OG1 - :NNK:O7	3.12471 2.99653 2.69827

9.	BRCA1	NNK	-5.08	187.68 uM	Val1654, Ser1655, Gly1656, Leu1657, Thr1658, Pro1659, Phe1662, Thr1700, Leu1701, Lys1702	BRCA1:SER1655:OG - :NNK:N14 BRCA1:GLY1656:N - :NNK:O7 BRCA1:LEU1657:N - :NNK:O7 BRCA1:LEU1701:N - :NNK:O15 BRCA1:LYS1702:N - :NNK:O15 :NNK:N14 - BRCA1:SER1655:OG	3.19252 2.75464 2.77131 3.00037 2.75018 3.19252
10	CDK1	NNK	-5.00	217.06 uM	Lys88, Leu91, Asp92, Ile94, Pro95, Pro96, Glu196, Lys200	CDK1:LYS200:NZ - :NNK:N14 CDK1:LYS200:NZ - :NNK:O15 :NNK:N14 - CDK1:ILE94:O	3.0124 2.96559 2.89466
11	CDK2	NNK	-4.90	255.02 uM	Val29, Glu81, Phe82, Leu83, His84, Ile135, Asn136, Thr137	CDK2:PHE82:N - :NNK:N10 CDK2:HIS84:N- :NNK:N14 CDK2:HIS84:N - :NNK:O15 CDK2:HIS84:ND1 - :NNK:N14 :NNK:O15 - CDK2:ILE135:O	2.88593 2.95506 2.76842 3.18956 2.91961
12	CCNB1	NNK	-4.86	274.97 uM	Ile253, Lys256, Tyr257, Glu285, Leu289, Phe294, Gly295, Leu296, Gly297	CCNB1:TYR257:N - :NNK:N10 CCNB1:LEU296:N - : NNK:N14 CCNB1:LEU296:N - : NNK:O15 CCNB1:GLY297:N - : NNK:O15 : NNK:N14 - CCNB1:PHE294:O	3.14377 2.86319 2.81795 2.91169 3.10697
13	CHEK1	NNK	-4.80	303.50 uM	Val23, Val37, Ile39, Glu55, Asn59, Leu82, Phe149	CHEK1:ILE39:N - :NNK:N10 CHEK1:ASN59:ND2 - :NNK:N14 CHEK1:ASN59:ND2 - :NNK:O15 CHEK1:PHE149:N - :NNK:O15 :NNK:N14 - CHEK1:GLU55:OE2 :NNK:O15 - CHEK1:GLU55:OE2 :NNK:O15 - CHEK1:PHE149:N	2.8265 2.71952 3.02158 3.15867 3.05383 2.67183 3.15867
14	RPA2 (connector)	NNK	-4.70	358.90 uM	Cys49, Thr50, Ile76, Val77, Asp96, Met97, Tyr125, Phe155, His158, Ile159	RPA2:VAL77:N - :NNK:O15 :NNK:O15 - RPA2:CYS49:O :NNK:O15 - RPA2:VAL77:O :NNK:O15 - RPA2:HIS158:NE2	3.09031 3.06724 3.18087 2.98064
15	ATM	NNK	-4.48	523.49 uM	Thr2059, Ala2062, Gly2063, Ile2065, Gln2066, Gln2069, Leu2077, Tyr2080, Leu2081, Leu2084, Glu2094, Leu2095, Leu2098	ATM:GLN2066:N - :NNK:O7	3.06063
16	CDK4	NNK	-4.35	650.98 uM	Val44, Leu54, Pro55, Thr58, Val59, Val62, Ala63, Arg66, Val82, Ile92, Val94	CDK4:PRO55:CD - :NNK:O15 CDK4:VAL59:CA - :NNK:O7	2.9883 2.92285
17	TFDP1	NNK	-4.20	833.96 uM	Val264, Phe285, Asn286, Phe287, Phe291	TFDP1:PHE287:N - :NNK:O15 :NNK:O15 - TFDP1:PHE287:O	2.78538 3.08744
18	TP53	NNK	-4.14	927.72	Gln136, Leu137, Ala138, His179, Cys182, Asp184, Asn239, Cys275,	TP53:LEU137:N - :NNK:O15 TP53:ASN239:ND2 - :NNK:N14 TP53:ASN239:ND2 - :NNK:O15 :NNK:O15 - TP53:CYS275:O	2.93164 3.07617 2.87469 2.82627

					Ala276		
19	RB1	NNK	-4.02	1.14 mM	Val434, Gly435, Gln436, Cys438, Asn505, Leu506, Asp507, Ser508, Gly509, Thr510	RB1:GLN436:HN - :NNK:O15 RB1:SER508:HN - :NNK:O7 RB1:GLY509:HN - :NNK:O7	1.93726 2.36625 1.90514
20	RPA1 (connector)	NNK	-3.67	2.03 mM	Val375, Asn402, Pro403, Ala408, Tyr409, Arg412, Gly413	RPA1:ARG412:NH1 - :NNK:N14 :NNK:O15 - RPA1:ALA408:O	3.06161 2.78707



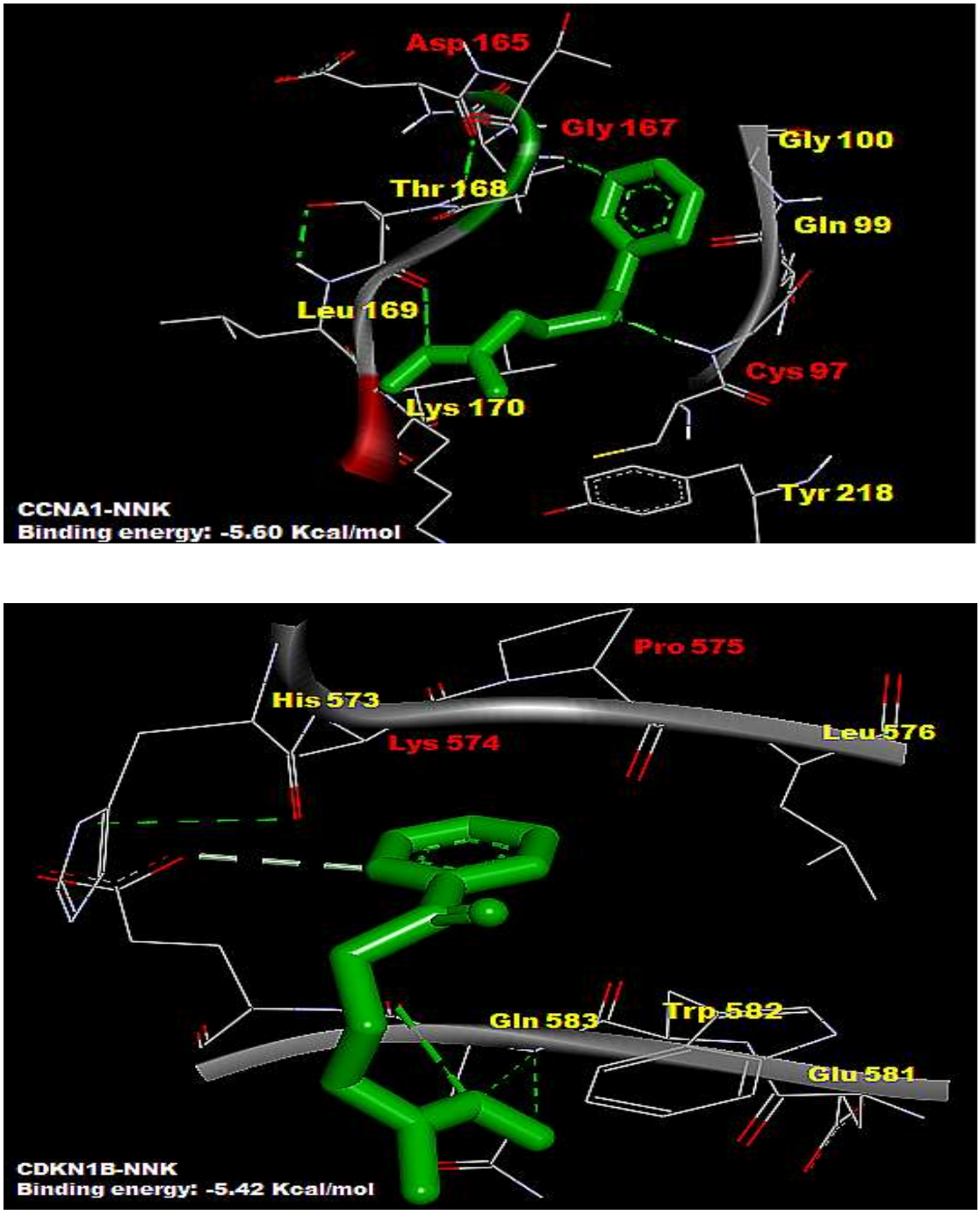


Fig 34: Binding interactions of top 3 proteins with NNK.

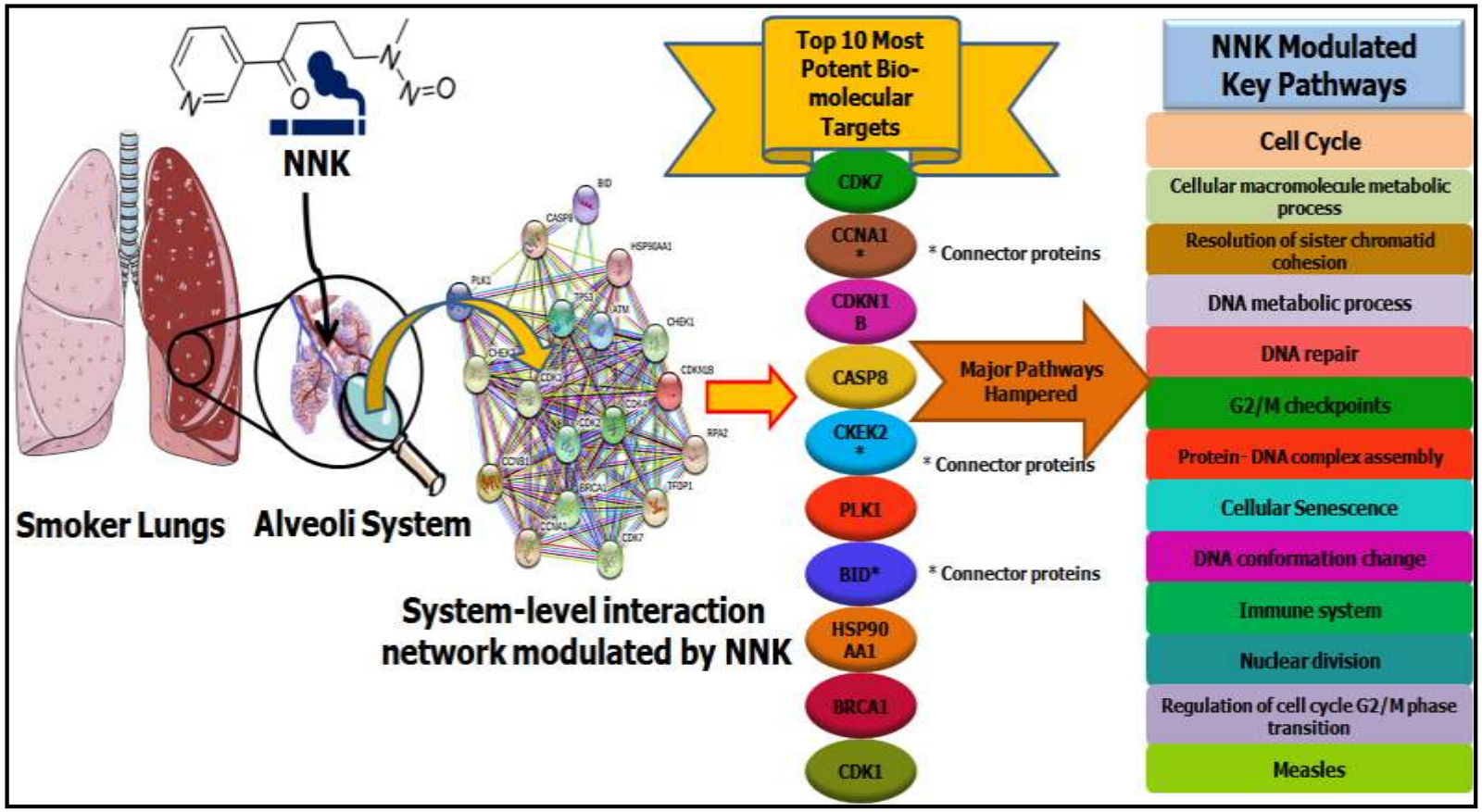


Fig 35: Comprehensive graphical representation of the impact of NNK on system level PIN with their respective key regulatory proteins along with the enriched pathways.

4.2.5) Mathematical modeling

For mathematical modeling of cell cycle, the model was first created using cell designer. SBML squeezer was used for generating the kinetics of reactions. For kinetics of some reactions, biomodel BIOMD0000000941 (Gerard and Goldbeter, 2010) was also referred. The concentration of NNK used in our study is based on the studies conducted by Schick and Glantz (2007). Fig 36 is the pictorial representation of the cell cycle model developed for NNK. In modeling the impact of NNK on cell cycle, we have used we have used top 3 biomolecular targets of NNK which play important role in the regulation of cell cycle.

For analysis, COPASI software was used. Initially the time course analysis was done without NNK and normal graph was obtained (fig 37 a). Fig 37b is the graphical representation of the fluctuations observed in the concentrations of major cyclin -CDK complexes of cell cycle in normal conditions when NNK is absent.

To study the changes in the oscillations of various components of the cell cycle that have got perturbed NNK, we have simplified our model (fig 20) and have focused only on the main components of cell cycle that are in close connection with the biomolecular targets under study.

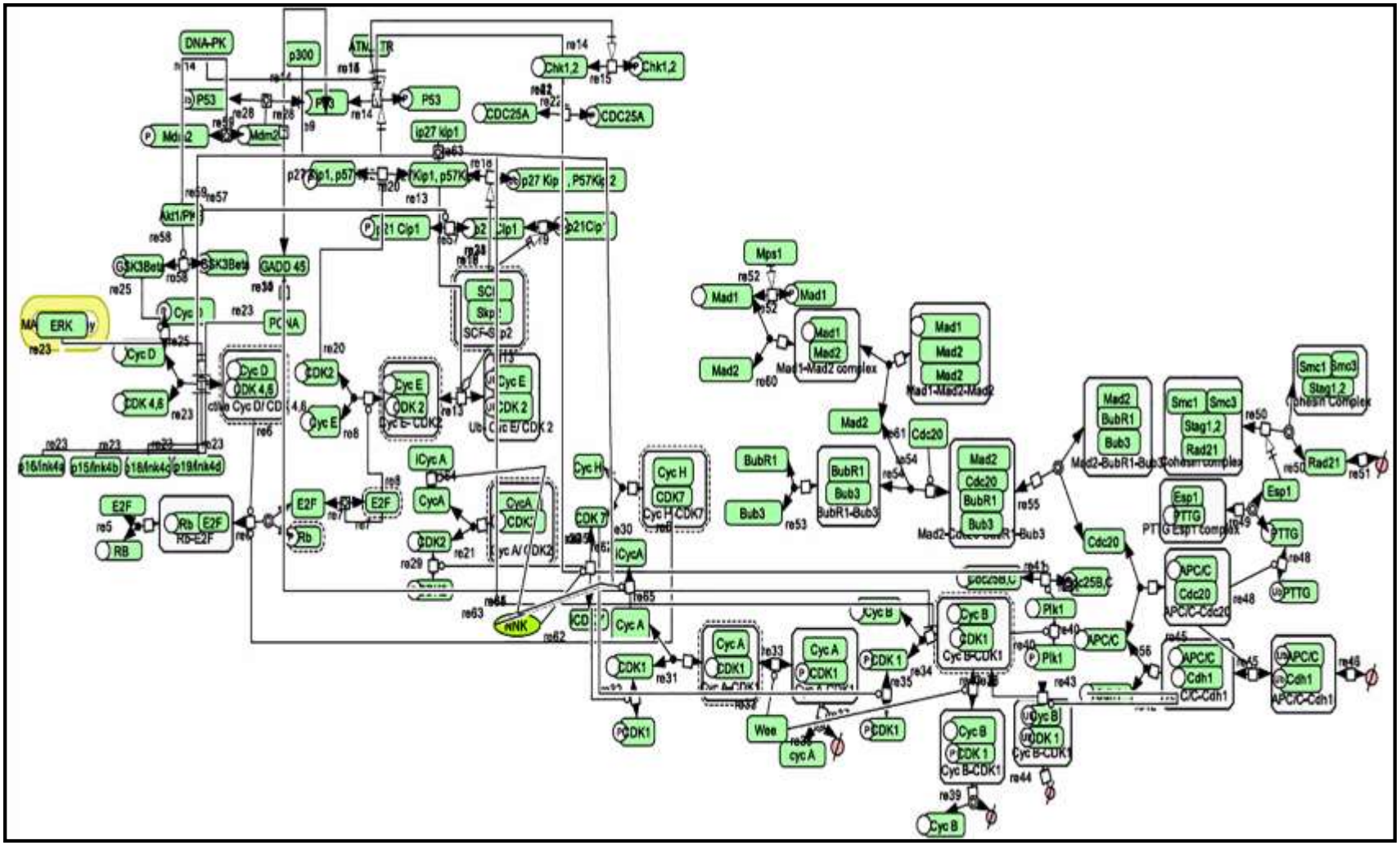


Fig 36: Cell Cycle designed using Cell Designer for the study of impact of NNK on cell cycle regulatory machinery

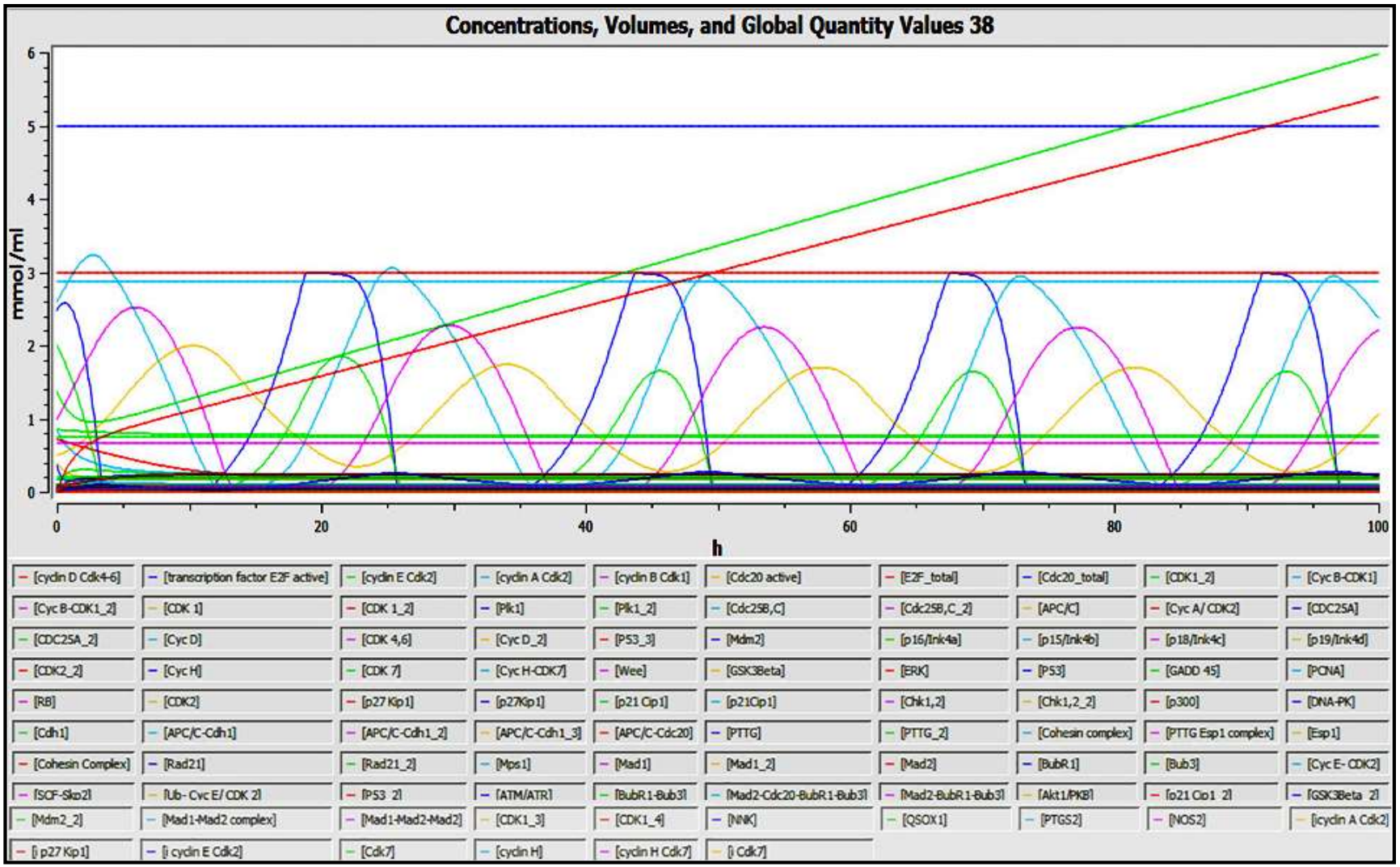


Fig 37a: Graph of normal concentration changes in cell cycle regulatory biomolecules when NNK is not present.

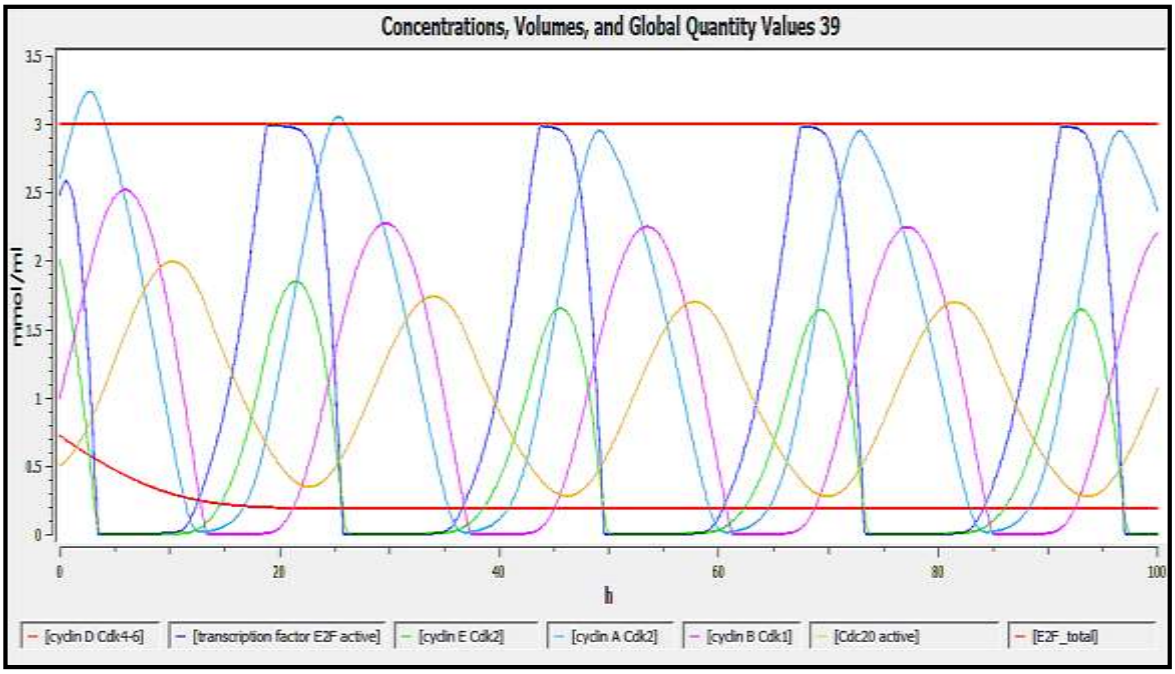


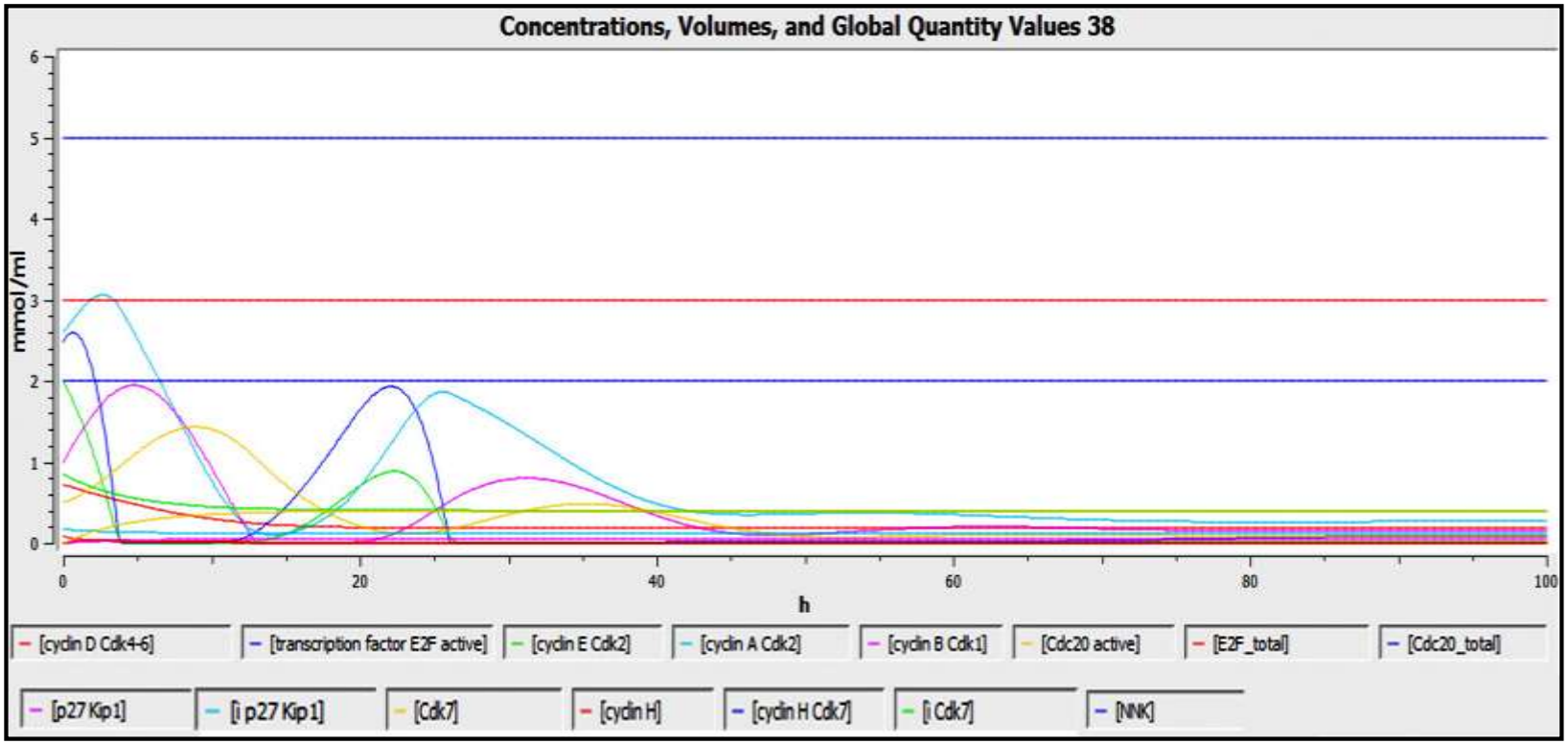
Fig 37b: Normal graph of all the CDK-cyclin complexes of cell cycle without NNK inhibition.

CDK7 showed the highest binding efficiency towards NNK. During the cell cycle progression, the level of CDK7-cyclin H complexes does not show much variation in concentrations unlike other CDK-cyclin complexes. CDK7 on being inhibited by NNK, caused fluctuations in the concentrations of CDK7-Cyclin H complex along with changes in all the important components of the cycle (fig 38). CDK7 plays an important role in the regulation of Cyclin B-CDK1 which is further regulates the activation of CDC20. CDC20 plays a role in the degradation of cyclinB-CDK1 and CyclinA-CDK2. CyclinA-CDK2 further helps in degradation of CyclinE-CDK2 and in the inactivation of E2F. CyclinE-CDK2 helps in the activation of E2F and E2F plays a role in the activation of CyclinA-CDK2 and CyclinE-CDK2. Hence inhibition of CDK7 has caused variations in the all important components of the cell cycle regulation. Table 8 depicts

the changes in concentration of all the major species of cell cycle when CDK7 gets inhibited by NNK.

Another biomolecule to be modeled was CCNA1 (Cyclin A1). NNK inhibits cyclin A causes changes in the concentrations of cyclin E CDK2, cyclin B CDK1, E2F etc. Changes in the concentrations and oscillations of all the impacted species has been shown in table 9 and table 10. Detailed changes in concentrations are listed in Annexure 6 and graph 39 represents the fluctuations.

CDKN1B (p27 Kip 1) is the third biomolecule being inhibited by NNK. P27 is also known as tumor suppressor protein. Inhibition of p27 leads to increased levels of cell proliferation and ultimately to cancer. Fig 40 isthe graphs which clearly depict the changes in concentrations of p27. Table 11and table 12enlist the changes in the concentrations and fluctuations in the oscillations when NNK inhibited p27. The details of all the time based fluctuations in all the species are presented in appendix 6.



38) Graph of cell cycle regulation when NNK inhibits CDK7. Extreme changes in the concentration changes can be observed when compared with uninhibited graph.

Table 8: Comparison of concentrations of various species when CDK7 is inhibited by NNK

Species	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited
	0		5		10		15		20		25	
cyc E CDK2	2	2	0.0015	0.0008	0.0025	0.0015	0.2745	0.1111	1.6895	0.7049	0.7512	0.4634
cyc A CDK2	2.6	2.6	2.7538	2.5181	0.7833	0.7668	0.058	0.1176	1.1831	0.699	3.0369	1.8391
cyc B CDK1	1	1	2.4661	1.9408	1.6007	0.9231	0.0001	0.0002	0.0441	0.0144	1.3012	0.3944
E2F active	2.4855	2.4855	0.0024	0.0027	0.0076	0.0079	0.9027	0.4475	2.9882	1.6404	1.2349	0.9494
Cdc20 active	0.5	0.5	1.2705	1.1192	1.9928	1.4019	1.3053	0.7152	1.5096	0.2086	0.47712	0.1365

Table 9: Comparison of concentrations of various species when CCNA1 (Cyclin A) is inhibited by NNK

Species	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited
	0		5		10		15		20		25	
Cyc A CDK2	2.6	2.6	2.7538	2.4815	0.7833	0.4687	0.058	0.1371	1.1831	1.3847	3.0369	2.9561
Cyc B CDK1	1	1	2.4661	2.3476	1.6007	1.3263	0.0001	0.0004	0.0441	0.1127	1.3012	1.4756
Cyc E CDK2	2	2	0.0015	0.0008	0.0025	0.0028	0.2745	0.38	1.6895	1.7757	0.7512	0.7463
E2F active	2.4855	2.4855	0.0024	0.0027	0.0076	0.0195	0.9027	1.0899	2.9882	2.9847	1.2349	1.1689
Cdc20	0.5	0.5	1.2705	1.2358	1.9928	1.8231	1.3053	1.0833	1.5096	0.3935	0.47712	0.5253

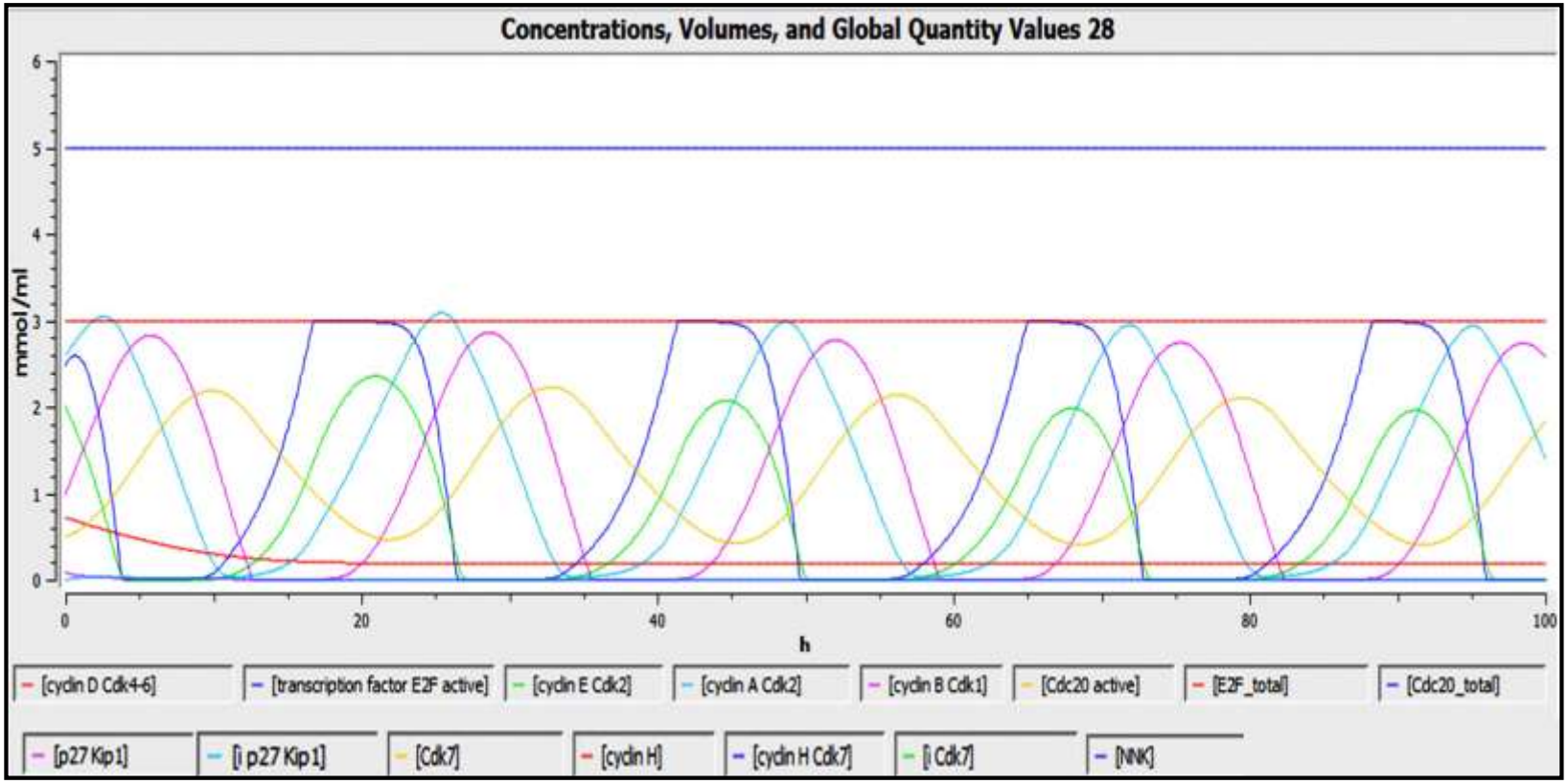


Fig 39: Graph of Cyclin A of cell cycle regulation when NNK is present

Table 10: Time based analysis of CCNA1 inhibited graph

Species		Conc at first max (mmol/ml)	time (hrs)	Conc at first min (mmol/ml)	time (hrs)	Conc at second max (mmol/ml)	time (hrs)	Conc at second min (mmol/ml)	time (hrs)	Duration of flat curve
Cyc A CDK2	Normal	3.2371	2.7	0.0154	13	3.0558	25.4	0.0131	37	
	Inhibited	3.0546	2.6	0.048	12.5	2.9571	24.9	0.052	36.4	
Cyc B CDK1	Normal	2.5194	6	4.65E-05	13.4	2.276	29.7	3.83E-05	37.5	(curve was flat between 13.4 hrs to 17 hrs)
	Inhibited	2.3705	5.6	0.0001	13.1	2.1977	29.3	0.0001	36.9	(curve was flat from 13.1 hrsto 16.1hrs)
Cyc E CDK2	Normal	1.8501	21.5	0.0015	5.7	1.6514	45.5	0.00209	27	(curve was flat from 4.7hrsto 7hrs)
	Inhibited	1.8797	21.1	0.0008	5	1.3567	46	8.65E-05	26.9	(curve was flat from 4.7 to 6.3 hrs)
E2F	Normal	2.5786	0.6	0.0024	4.3	2.9926	19	0.0008	26.5	(curve was flat from 4.1 to 5hrs)
	Inhibited	2.5909	0.6	0.0027	4.2	2.9911	18.7	0.0008	26.5	(curve was flat from 4 to 4.8 hrs)
Cdc20	Normal	1.9954	10.3	0.3474	22.7	1.7401	34.1	0.2799	46.3	
	Inhibited	1.8246	9.8	0.3143	21.9	1.6451	33.2	0.2264	45.8	

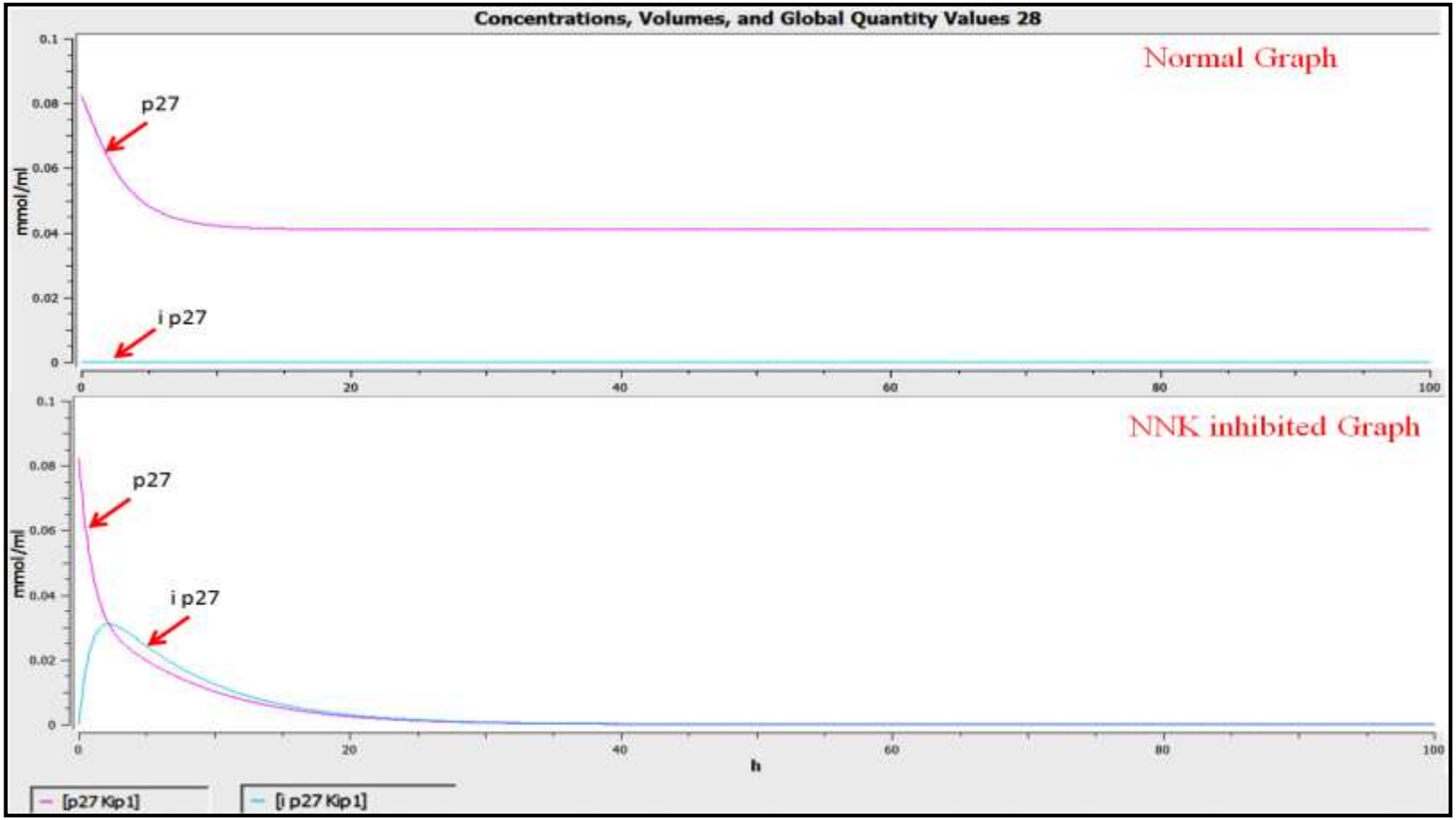


Fig 40:.Comparative chart of p27 and inactivated p27 in absence of NNK and in presence of NNK.

Table 11: Comparison of concentrations of various species when CCNA1 (Cyclin A) is inhibited by NNK

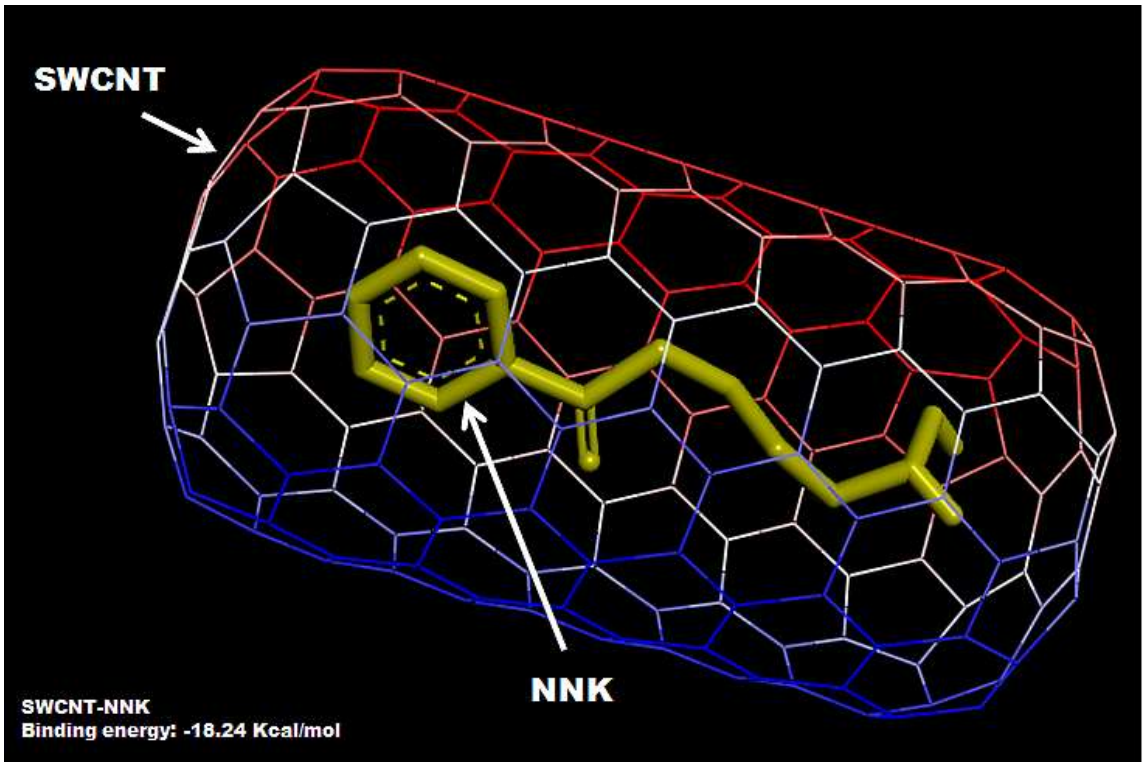
Species	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited	Normal	Inhibited
	0 hrs		5 hrs		10 hrs		15 hrs		20 hrs		25 hrs	
p27	0.082	0.082	0.0482	0.0196	0.0421	0.0102	0.0412	0.005	0.041	0.0024	0.041	0.0011
I p27	0	0	0	0.0242	0	0.0124	0	0.0061	0	0.0029	0	0.0013
Cyc A CDK2	2.6	2.6	2.7538	2.4815	0.7833	0.4687	0.058	0.1372	1.1831	1.3847	3.0369	2.9561
Cyc B CDK1	1	1	2.4661	2.3476	1.6007	1.3263	0.0001	0.0003	0.0441	0.1127	1.3012	1.4756
Cyc E CDK2	2	2	0.0015	0.0008	0.0025	0.0028	0.2745	0.38	1.6895	1.7757	0.7512	0.7463
E2F active	2.4855	2.4855	0.0024	0.0027	0.0076	0.0195	0.9027	1.0899	2.9882	2.9847	1.2349	1.1689
Cdc 20	0.5	0.5	1.2705	1.2359	1.9928	1.8231	1.3053	1.0833	1.5096	0.3935	0.47712	0.5253

Table 12: Time based analysis of CDKN1B inhibited graph

Species		Conc at first max (mmol/ml)	time (hrs)	Conc at first min (mmol/ml)	time (hrs)	Conc at second max (mmol/ml)	time (hrs)	Conc at second min (mmol/ml)	time (hrs)	
p27	Normal	0.082	0	0.041	18.4	Gradual decrease				
	Inhibited	0.082	0	Gradual decrease						
ip27	Normal	throughout 0 mmol/ml								
	Inhibited	0.031	2.2	Gradual decrease						
Cyc E CDK2	Normal	1.8501	21.5	0.0015	5.7	1.6514	45.5	0.00209	27	(curve was flat from 4.7hrsto 7hrs)
	Inhibited	1.8797	21.1	0.0008	5	1.3567	46	8.65E-05	26.9	(curve was flat from 4.7 to 6.3 hrs)
E2F	Normal	2.5786	0.6	0.0024	4.3	2.9926	19	0.0008	26.5	(curve was flat from 4.1 to 5hrs)
	Inhibited	2.5909	0.6	0.0027	4.2	2.9911	18.7	0.0008	26.5	(curve was flat from 4 to 4.8 hrs)

4.2.6) Protection by nanoparticles

Carbon nanoparticles were designed using nanotube modeler (Fig 23) and then with the help of Material studio, the structures were optimized. They were then docked with NNK(Fig 41) and with the adsorption load was calculated using Blend of material studio (fig 42). The binding energy of NNK was highest with SWCNT (-18.24 Kcal/mol) and MWCNT showed the highest adsorption capacity of 12 molecules/NT for NNK. NNK has least binding energy with fullerene (-3.09 Kcal/mol) and alsothe least adsorption capacity (6 molecules/NT) as shown in table 13.



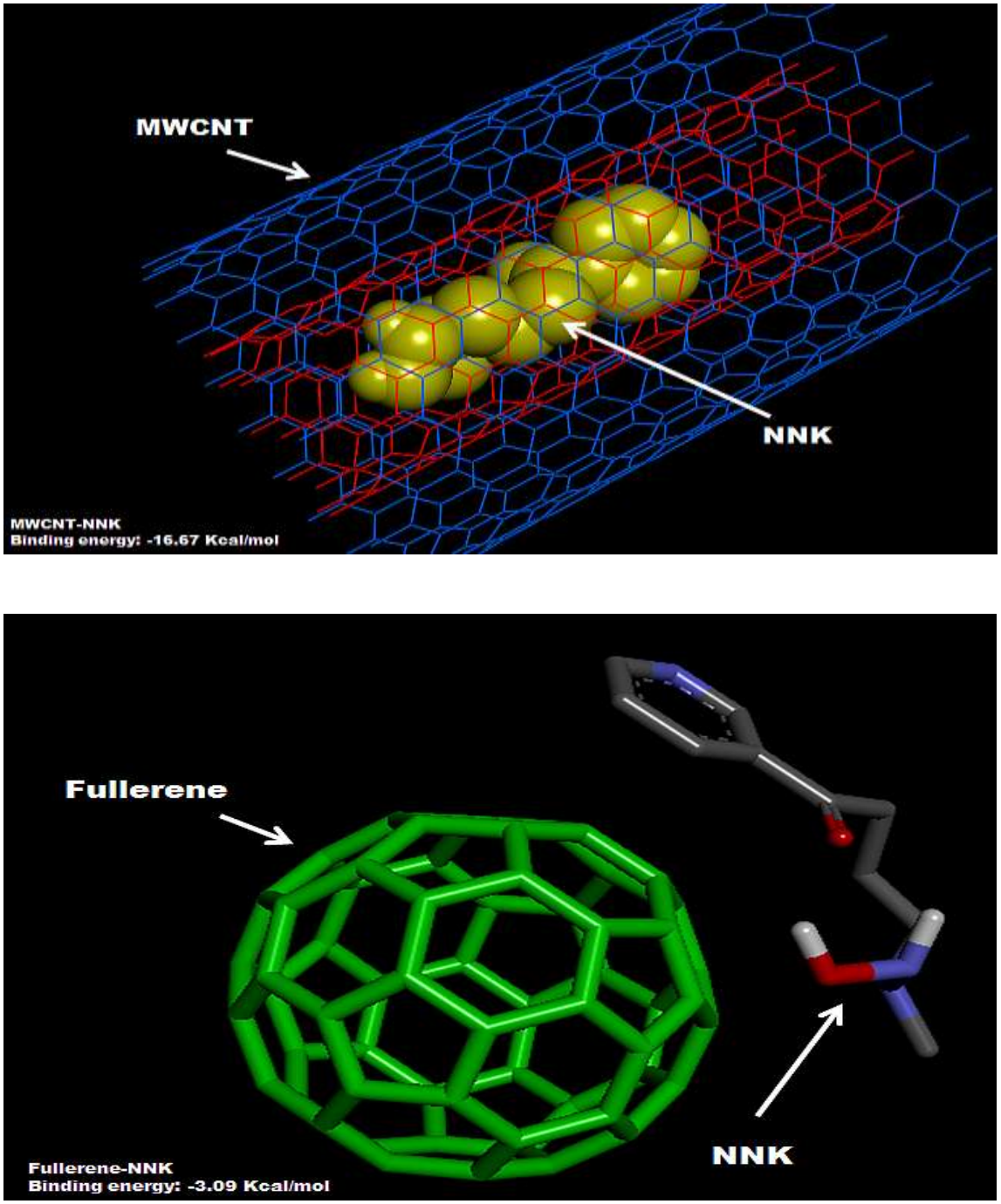
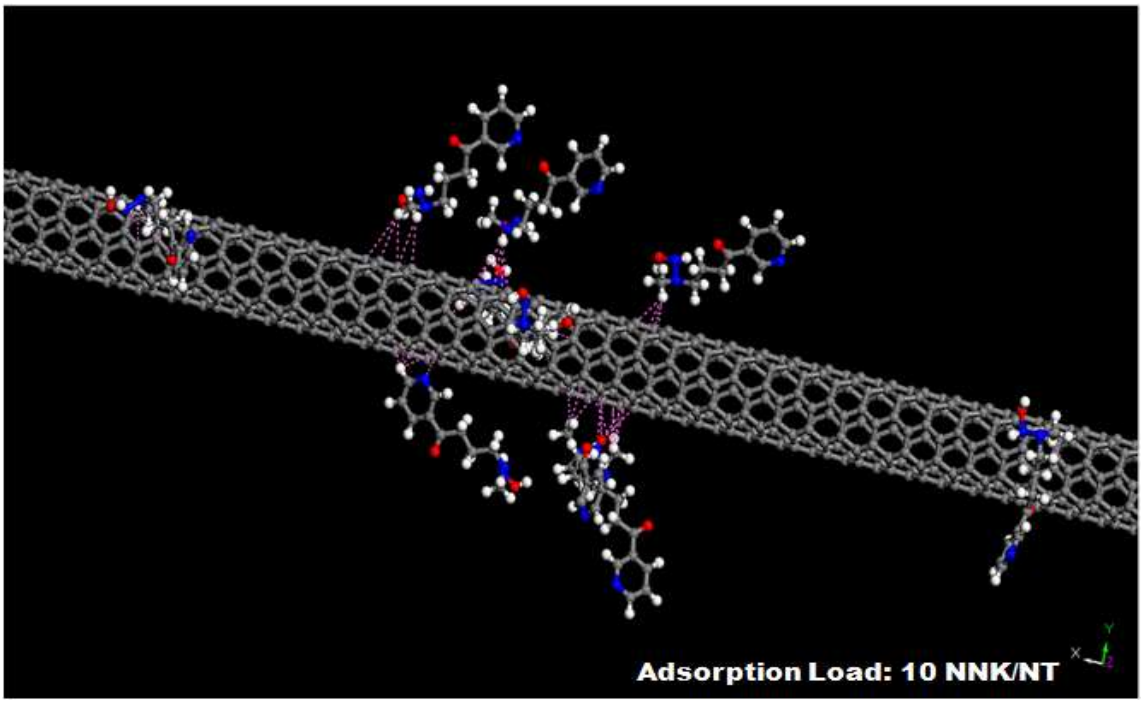
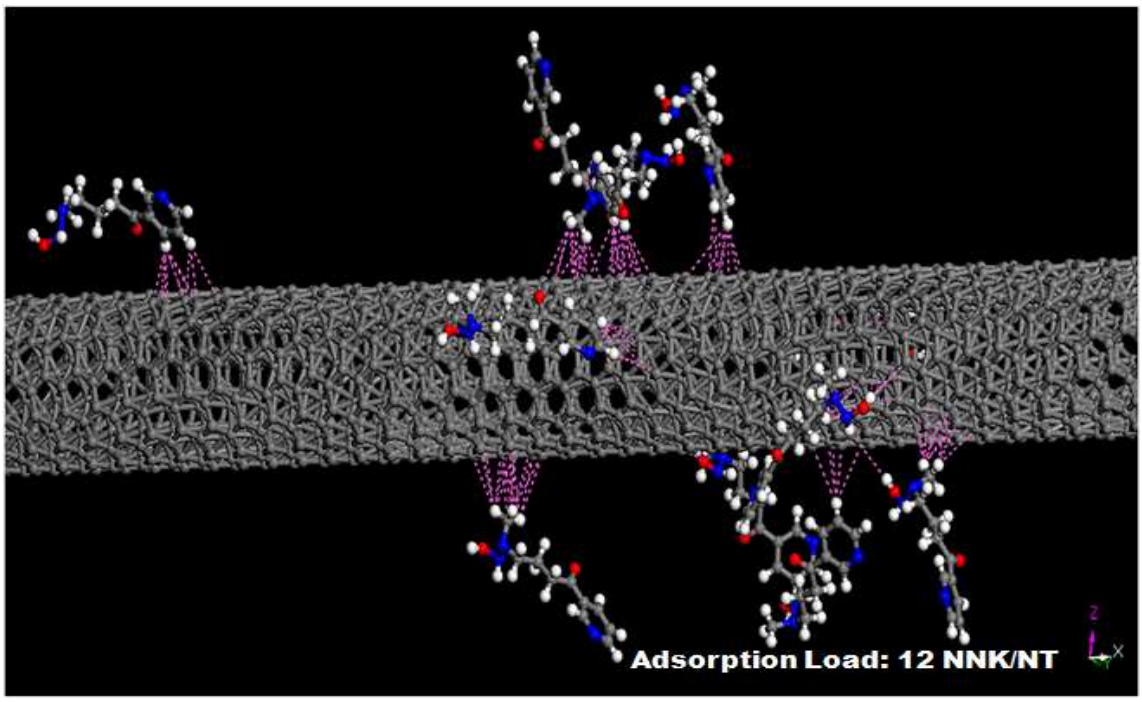


Fig 41: NNK dockedon SWCNT, MWCNT and fullerene. The binding energy of SWCNT with NNK was -18.24 Kcal/mol. MWCNT had binding energy of -16.67 Kcal/mol whereas Fullerene showed the binding energy of -3.09 Kcal/mol when docked with NNK.

Adsorption capacity/adsorption load analysis





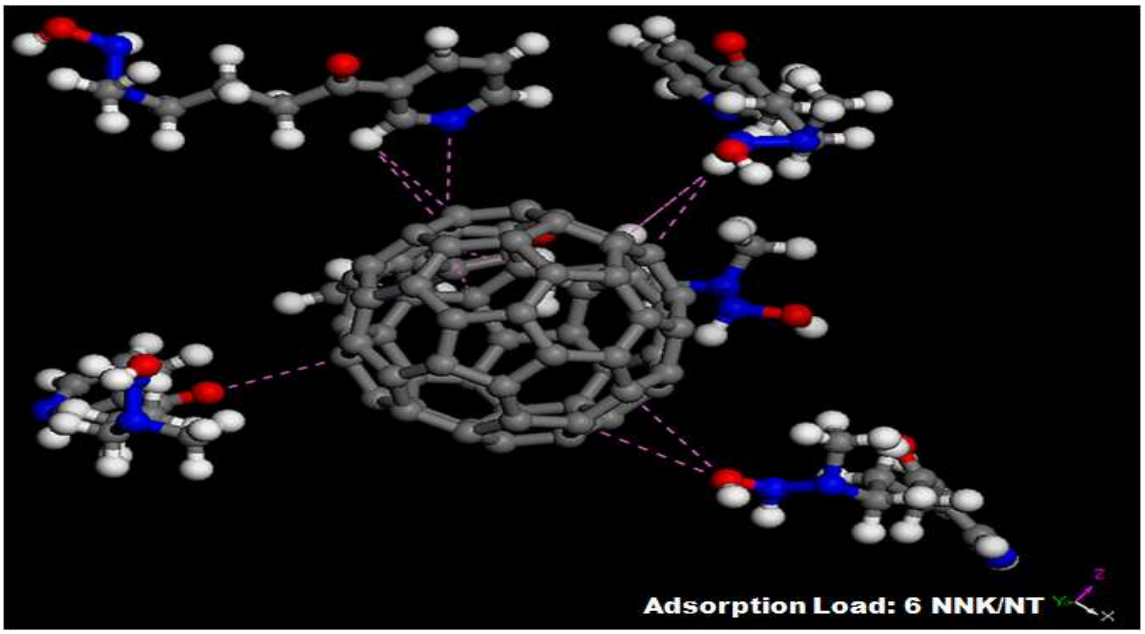


Fig 42:Adsorption of NNK on SWCNT, MWCNT and Fullerene. SWCNT adsorbed 10 molecules of NNK per nanotube. MWCNT adsorbed 12 molecules of NNK per nanotube while fullerene showed a minimum adsorption of 6 molecules of NNK on its surface.

Comparative study of binding energy and adsorption load of NNK on SWCNT, MWCNT and fullerenes is represented in table 13 below.

Table 13: Comparative chart of binding energy and adsorption load of NNK against SWCNT, MWCNT and Fullerene along with the binding energy of CDK7

Carbon Nanomaterial/ Target Protein	Ligand	Binding Energy (Kcal/Mol)	Adsorption Load (Molecules/NT)
SWCNT	NNK	-18.24	10
MWCNT	NNK	-16.67	12
Fullerene	NNK	-3.09	6
CDK7	NNK	-5.93	-

Important findings for NNK:

- *NNK hampers the functioning of 544 proteins.*
- *The NNK rewired PPIN had a clustering coefficient of 0.597, Characteristic pathlength was 3. Average number of neighbors were 15.940, suggesting that the network is a small world network.*
- *19 clusters were generated using MCODE plugin of cytoscape software which had 115 seed proteins.*
- *21 most probable biomolecular targets were identified based on the various topological properties which were used as screening criteria.*
- *On enrichment analysis, most of the proteins were found to be involved in the cell cycle regulatory pathways.*
- *On molecular docking analysis of the 21 most probable biomolecular targets, CDK7, CCNA1 and CDKN1B were the top 3 biomolecules with highest binding energies of -5.93 Kcal/mol, -5.60 Kcal/mol and -5.42 Kcal/mol respectively.*
- *On molecular docking analysis and adsorption load analysis, SWCNTs showed highest binding energy of -18.24 Kcal/mol and an adsorption load of 10 molecules/NT followed by MWCNT with -16.67 Kcal/mol binding energy and 12 molecules/NT adsorption load. Fullerenes showed minimum binding energy of -3.09 Kcal/mol and adsorption capacity of 6 molecules/NT against NNK.*

DISCUSSION

Discussion

Cancer is the second deadliest disease across the globe with approximately 9.6 million deaths in 2018 (WHO). WHO also has reported that 1 in every 6 deaths is due to cancer and the disease is more prevalent in low to middle income countries. 1,42,670 deaths were estimated to happen due to lung cancer in USA in 2019, which is approximately 27% of all cancer deaths in USA (Lung cancer foundation of America). Environmental carcinogens play an important role in cancer initiation and progression by causing multiple disruptions in the cellular processes (Parsa N., 2012). Along with environmental carcinogens, cigarette smoke is also a major contributor in causing cancer and share some major carcinogens with environment like PAHs and some metals.

International Agency for Research on Cancer has classified various chemicals present in cigarette smoke into categories and has placed BaP and NNK into group 1, which means that there are various evidences which confirm that the substance is carcinogenic to humans. BaPis an aromatic hydrocarbon with five fused rings which forms bulky adducts by hampering the functioning of DNA polymerase. It is formed by incomplete combustion of organic matter at high temperatures between the ranges of 300°C to 600°C.

NNK is a potent carcinogen present in tobacco products which damages the DNA and reduces the mass of mitochondria (Xue *et al*., 2014 and Chen *et al*., 2011). NNK, like BaP is involved in the formation of DNA adducts for inducing mutations. NNK is responsible for various chromosomal aberrations which accumulate and ultimately lead to the formation of tumor.

In this study we have used the tools of systems biology to pinpoint the most important biomolecular targerts of BaP and NNK and to find the inherent mechanism how these environmental carcinogens perturb the cell cycle machinery. As all the pathways are constantly interacting with each other and are tightly regulated, any change in a single protein can change the harmony of whole system. By performing *in silico* biomolecular interaction analysis, we have tried to find the proteins that are potent biomolecular targets of BaP and NNK and how BaP and NNK hamper the functioning of these biomolecular targets and hence perturb the whole machinery. For the analysis we have used various tools of systems biology. Systems biology is an interdisciplinary field which helps in filling the gap between the in vitro and in vivo biological systems by analyzing the data that has been generated and uncovering the functional aspects of all the genes and proteins present in the system (Koutsogiannouli *et al*. 2013). Information revealed by systems biology help in the analysis of complex relationship between huge networks of various pathways (Arora D and Singh A., 2018). Systems biology works on the principles of graph theory. Graph theory is the study of graphs and its elements using mathematical notations which was first proposed by Leonhard Euler

in 1736. Graphs or networks are used commonly to represent the binary relationships between various bioentities (Kourtrouli M., 2020). Graph theory finds its immense use in biology in genomics, in understanding the perturbations in various pathways underlying any disease, in neurological studies etc.

In the current study, we have tried to find the most important biomolecular targets of BaP and NNK which are known environmental carcinogens. We used T3DB (The Toxin and Toxin Target database) for extracting the genes/proteins which get hampered by the action of BaP from which 1722 upregulated genes and around 2370 downregulating genes were identified. To find the most potent biomolecular targets of NNK, we did literature survey. We used pubmed to find research articles on NNK using various keywords like NNK, human cell cycle, cancer, cell progression etc. Approximately 1320 research papers were searched, and 544 genes have been identified that get hampered due to the action of NNK.

Once the genes were identified, STRING database was used to generate the protein-protein interaction networks for BaP and NNK separately. For BaP, separate networks for upregulated and downregulated genes were constructed which were later merged using the merge tool of the cytoscape software, an opensource software platform which offers many plugins for analysis and visualization of the networks (Sebatian S and Shamsir MS, 2019). The PPIN for upregulated proteins comprised of 1676 nodes which were connected by 7113 edges with an average node degree of 8.49 and an

average clustering coefficient of 0.398. The PPI enrichment p-value was < $1.0e^{-16}$. Similarly, the PPIN generated for down-regulated proteins had approximately 2563 nodes which were held together by approximately 11,100 edges. During the construction of the network by STRING database, the confidence score was kept highest at 0.9 which means that the there are less chances of false positives in the network. This is a probability score that predicts the existence of links between two enzymes within same metabolic KEGG pathways.

Another criterion that was added during the construction of networks was inclusion of 50-50 connectors in first and second shell respectively. First shell connectors are the direct connections with the submitted proteins whereas the second shell connectors are those which have indirect relations with the submitted proteins but are in direct connection with the proteins of the first shell. Addition of connectors facilitate in providing an opportunity to find some new unreported proteins that may be holding importance role in any given pathway (RS Wang and J Lascalzo, 2018).

For the construction of PPIN using STRING database, the active interaction sources that were chosen were experiments, databases, co-expression, neighborhood, gene fusion and co-occurance. All these interaction sources were chosen to get maximum number of interactions possible which have highest confidence score of 0.9.

The protein-protein interaction network developed for NNK comprised of 534 nodes which were 2910 edges which had an average node degree score of 10.09 and an average clustering coefficient of the network was 0.501. The network had a PPI enrichment index of $< 1x\ 10^{-16}$.

The networks were further analyzed using the network analyzer tool of cytoscape. It helps in calculating the topological parameters like average neighbors of nodes, diameter of network, clustering coefficient, shortest pathlength etc. Analysis all these parameters help in understanding the type of network that has been generated and also gives information regarding each and every node present in the network. Node degree helps in analyzing the number of connections present in a node. Higher the number of connections of any node means that more information is getting passed from that.

Network analysis of BaP rewired PPIN

For the complete analysis of the BaP rewired protein interactome, all the up-regulated and down-regulated networks were merged using merge tool of cytoscape software and were then analyzed using network analyzer tool. On analysis, the network thus generated had average clustering coefficient of 0.599(fig 11).It is the average of clustering coefficients of all the nodes present in the network. High scores of average clustering coefficients indicate the small world properties (Watts DJ and Strogatz SH 2011).Such networks also represent the properties of scale free, real world networks

(Buchanan M *et al.*, 2010). The characteristic pathlength of the network was 4.615 which means that there are 4.615 average number of edges that fall in the shortest paths between all the given nodes (Dubitzky *et al* 2013) which can be clearly visualized from fig 10 of shortest pathlength distribution in which the highest peak is obtained at 4 units with a frequency of more than 950,000.

The density of the network was 0.007. Network densities give the information about the compactness of the network. Highly dense network systems tend to produce enormous redundant information. Low density networks present less redundant information and are considered to be more efficient in finding solutions (Stokman, 2001). Low density networks also offer more stable and reliable transfer of information as compared with a high density network (Alsaqour *et al* 2012). As the density of our network is low, this means that the transfer of information is quite stable and reliable. The diameter of the network was 13, which is usually used to measure the effect of removal of any node or edge from a network. It is a property of scale free networks which shows an increase on slightest removal of nodes or edges (Eskin *et al.*, 2005).

Further analysis of BaP rewired PPIN node degree distribution and average node connectivity graphs were obtained. Node degree distribution (Fig 12) helps in finding the number of nodes having a particular number of connections. Power law ($y=ax^{-b}$) was applied and on fitting the node degree distribution curve, degree exponent (b) was calculated to be 1.408. Degree distribution is one of the prominent topological

characteristic of any biological network. The measure of any network being scale free depends on this degree exponent value b which ranges from 2 to 3 in case of strong to very strong scale free networks. Networks which have a value b<2 are considered to be weak or weakest scale free networks. In real world we have more networks having b< 2 than the networks with b> 2. Such networks have slightly different from the later ones but still follow the small world properties (Broido AD and Clauset A. 2019). Such networks also has a key property that the average degree grows with the size of the network with no expenses incurred and the nodes are free to interact globally (Seyed-allaei H *et al.,* 2006). Neighborhood connectivity distribution curve helps in finding the average of all the neighbourhoodconnectivities of all the nodes. The curve is also fitted using power law in which the exponential calculated was 0.272 (Fig 13). If there is a decreasing trend it means that edges are present between low connection nodes and high connection nodes.

Once the topological analysis of whole network was studies, the topological analysis of individual nodes was done by using cytohubba plug-in of the cytoscape software. Cytohubba helps in attaining all the centrality measures of all the nodes present in network. Out of various centrality measures, we gave emphasis to degree followed by clustering coefficient, betweenness centrality and finally bottleneck scores (Table 1 Annexure1).

The network generated had noise in it which is evident from many nodes that have 0 degrees which meant that they have no connections and have no interaction with other nodes. To reduce the noise, modules were generated which not only reduces noise but also creates the clusters of functionally related nodes (Ma X *et al.* 2017). MCODE was used to make clusters and from the clusters, seed proteins were identified. Seed proteins are those proteins which were present in the initial gene set. For BaP rewired PPIN, 65 clusters were formed which had around 420 seed proteins (table1).This significantly reduced the number of proteins that are not only involved in the interaction with BaP but are also playing crucial role in the biological processes.

We again subjected these seed proteins to generate network and this time also we considered adding connectors which gave us an opportunity to find some non-reported proteins. Then various topological properties like degree, clustering coefficient, betweenness and bottleneck were analyzed using cytoscape software.

First, we sorted the list of proteins based on the clustering coefficient(Pearson correlation coefficient) an important topological feature of the selection of these possible key proteins, those proteins are less than 0.5 are selected and assumed to classified as date hubs. These date proteins are important in PIN but holding lesser no. of connection so chances of binding of drug or other ligands will be higher than party proteins (party proteins, co-efficient more than 0.5 and show high degree of co-expression with

interaction nodes/partner are assumed to interact at the same time with their interaction nodes/partners) (Andorf *et al* 2013).

Then we chose bottleneck as another important criteria as proteins with higher bottleneck emerge out to be hub proteins in any network thus playing a pivotal role in maintaining the integrity of the network. Bottlenecks are significant indicators to figure out essential proteins or genes. These are the dynamic components of the interaction network. The bottlenecks proteins control the flow of biological information within the network, and their disruption can break the entire network into small components or subgraphs. These proteins also measure the number of shortest paths and are supposed to have highest betweenness scores. Therefore, nodes with the highest betweenness control most of the information flow in the network, representing the critical points of the network. We thus, call these nodes the ‘‘bottlenecks’’ of the network. Here, we focus on bottlenecks in protein networks. We find that, in the regulatory network, where there is a clear concept of information flow, protein bottlenecks indeed have a much higher tendency to be essential genes (Haiyuan *et al* 2007, Rosado *et al* 2011). Here we have conveniently avoided proteins with low bottleneck score because even if they are strong targets of BaP, their interaction with BaP will not be able to create an impact on whole network. By doing this step we chose the proteins with high degree, clustering coefficient not more than 0.5 and high bottleneck score.

As our data is a discrete data, to further refine the set and find a cut off value, we applied median and selected top proteins on the basis of preferential docking energies. The calculated median score was 4, hence the proteins till bottleneck score 4 were selected.Finally 38 proteins were selected for further screening process of molecular docking simulations.

Apart from reducing the noise from the initial data using modulation method, GO enrichment analysis of the final seed proteins obtained from clusters was also performed.ClueGO was used to perform GO enrichment analysis. It is done to find all the pathway and functionally related genes (Zheng and Wang, 2008). In the gene set of seed proteins obtained after modulation, GO enrichment analysis result (Fig 15 and Fig 16), shows that most of the genes are related to thecell cycle regulatory machinery. From this result we can easily deduce that most of the genes that get hampered by BaP belong to various phases of cell cycle regulatory pathways.

Network analysis of NNK Rewired PPIN

For the topological analysis of network generated by NNK using STRING db, network analyzer tool of cytoscape software was used. The basic topological properties of the network that were analyzed characteristic path length, clustering coefficient, average number of neighbors and network density.The characteristic pathlength calculated was 2.971. This means that maximum information is being passed from the group of nodes that are 3 unit path length away. Characteristic path length is the average

of all the path lengths. The clustering coefficient of the network or global clustering coefficient was calculated to be 0.560. This means that overall nodes in the network have fairly good tendency of making clusters. High scores of average clustering coefficients indicate the small world properties (Watts and Strogatz2011). Such networks also represent the properties of scale free, real world networks (Buchanan*et al.*, 2010).

The density of the network was calculated to be 0.042. Network density depicts about the closeness or connectedness of nodes in a network whose value ranges from 0 to 1. The nodes with no edges have a value of 0 while fully connected nodes have a value of 1. The value obtained for NNK rewired PPIN depicts that the network is sparsely populated which is a real world property. This also indicates that the network is stable and the transfer of information is reliable (Alsaqour *et al* 2012) and less redundant in comparison to dense networks (Stokman, 2001).

Another topological property is network diameter. It tells about the largest distance between two nodes (Crescenzi *et al* 2013). The value of network diameter was calculated to be 8 which helps in understanding that the longest length for complete transfer of information between two nodes in the network is of 8 edges. The shortest pathlength histogram (fig 27) shows that the path length of unit 3 has the highest frequency of more than 45000. This means that the shortest distance between two nodes for complete transfer of information is 3 (Dubitzky *et al* 2013).

Further analysis of NNK rewired PPIN node degree distribution and average node connectivity graphs were obtained. Node degree distribution (Fig 28) helps in finding the number of nodes having a particular number of connections. Power law ($y=ax^{-b}$) was applied and on fitting the node degree distribution curve, degree exponent (b) was calculated to be -0.861. Neighborhood connectivity distribution curve helps in finding the average of all the neighbourhoodconnectivities of all the nodes. The curve is also fitted using power law in which the exponential calculated was 0.137(Fig 29). If there is a decreasing trend it means that edges are present between low connection nodes and high connection nodes.

Once the topological analysis of whole network was studies, the topological analysis of individual nodes was done by using cytohubba plug-in of the cytoscape software. Cytohubba helps in attaining all the centrality measures of all the nodes present in network. Out of various centrality measures, we gave emphasis to degree followed by clustering coefficient, betweenness centrality and finally bottleneck scores (Annexure 2).

The network generated had noise in it which is evident from many nodes that have 0 degrees which meant that they have no connections and have no interaction with other nodes. To reduce the noise, modules were generated which not only reduces noise but also creates the clusters of functionally related nodes (Ma*et al.* 2017). MCODE was used to make clusters (Annexure 4) and from the clusters, seed proteins were identified.

Seed proteins are those proteins which were present in the initial gene set. For NNK rewired PPIN, 19 clusters were formed which had around 115 seed proteins. This significantly reduced the number of proteins that are not only involved in the interaction with NNK but are also playing crucial role in the biological processes. The clusters generated were ranked and scored. High score depicts that the cluster is tightly knit and has high density.

We again subjected these seed proteins to generate network and this time also we considered adding connectors which gave us an opportunity to find some non-reported proteins. Then various topological properties like degree, clustering coefficient, betweenness and bottleneck were analysed using cytoscape software.

First, we sorted the list of proteins based on the clustering coefficient(Pearson correlation coefficient) an important topological feature of the selection of these possible key proteins, those proteins are less than 0.5 are selected and assumed to classified as date hubs. These date proteins are important in PIN but holding lesser no. of connection so chances of binding of drug or other ligands will be higher than party proteins (party proteins, co-efficient more than 0.5 and show high degree of co-expression with interaction nodes/partner are assumed to interact at the same time with their interaction nodes/partners) (Andorf*et al* 2013).

Then we chose bottleneck as another important criteria as proteins with higher bottleneck emerge out to be hub proteins in any network thus playing a pivotal role in maintaining the integrity of the network. Bottlenecks are significant indicators to figure out essential proteins or genes. These are the dynamic components of the interaction network. The bottlenecks proteins control the flow of biological information within the network, and their disruption can break the entire network into small components or subgraphs. These proteins also measure the number of shortest paths and are supposed to have highest betweenness scores. Therefore, nodes with the highest betweenness control most of the information flow in the network, representing the critical points of the network. We thus, call these nodes the ‘‘bottlenecks’’ of the network. Here, we focus on bottlenecks in protein networks. We find that, in the regulatory network, where there is a clear concept of information flow, protein bottlenecks indeed have a much higher tendency to be essential genes (Haiyuan *et al* 2007, Rosado *et al* 2011). Here we have conveniently avoided proteins with low bottleneck score because even if they are strong targets of NNK, their interaction with NNK will not be able to create an impact on whole network. By doing this step we chose the proteins with high degree, clustering coefficient not more than 0.5 and high bottleneck score.

As our data is a discrete data, to further refine the set and find a cut off value, we applied median and selected top proteins on the basis of preferential docking energies. The calculated median score was 2, hence the proteins till bottleneck score 2 were

selected. Finally21 proteins were selected for further screening process of molecular docking simulations.

Apart from reducing the noise from the initial data using modulation method, GO enrichment analysis of the final seed proteins obtained from clusters was also performed. ClueGO was used to perform GO enrichment analysis. It is done to find all the pathway and functionally related genes (Zheng and Wang, 2008). In the gene set of seed proteins obtained after modulation, GO enrichment analysis result (Fig 32 and fig 33), shows that most of the genes are related to the cell cycle regulatory machinery. From this result we can easily deduce that most of the genes that get hampered by NNK belong to various phases of cell cycle regulatory pathways.

Molecular Docking Analysis

Molecular docking simulation is an *in silico* approach of finding the intermolecular framework of two interacting molecules. It helps in finding the interaction patterns and orientations of a ligand with a macromolecule in a 3D environment. With the help of docking the strength of the bound molecules can be predicted with the help of binding affinity scores. It is also a potent tool in structure based drug discoveries (Meng *et al* 2012).

Molecular Docking of BaP:

Molecular docking simulations help in finding the binding efficiencies of the selected proteins with BaP. This helped in finding the proteins with highest binding efficiency which further helped in pin pointing the best biomolecular targets of BaP. Molecular docking was performed using Autodock suite 1.5.6 on a standalone HP machine. Docking results have shown that BaP has the best binding affinity for QSOX1, followed by PTGS2 and NOS2. Table 2 shows the docking results of all the finally selected proteins with BaP.

QSOX1 is Sulfhydryl oxidase 1 (UniProtKB - O00391), which helps in formation of disulphide bonds in various extracellular proteins (Chakravarthi *et al.,* 2007 and Alon *et al*. 2012). It also helps in cell to cell adhesion and in cell migration (Ilani *et al* 2013 and Javitt *et al*., 2019). BaP on interaction with QSOX1, upregulates it (Lizarraga *et al.,* 2012) and over expressed QSOX1 is reported to play an important role in tumor cell invasion and also helps in migration of tumor cells (Lake and Feigal 2014 and Borges *et al.,* 2015).

PTGS2 (Prostaglandin G/H synthase 2) (UniProtKB - P35354) is the second biomolecule with highest binding affinity. BaP has been reported to upregulate PTGS2 (Degner *et al.,* 2007 and Sparfel *et al.,* 2010). It is also known as COX2. Upregulation of PTGS2 has been reported to cause cancer by hampering the functioning of p27 and p21 (Toyoshima *et al.,* 2002; Baldi *et al* 2004 and Sobolewski *et al.,* 2010). There are

studies that prove that high concentration of COX2 along with low concentration of p27 and p21 are potent biomarkers and are prognostic targets for various cancers like SCLC, colorectal cancer etc (Mineo et al 2010).

NOS2 is the third biomolecule which we have considered because of the binding affinity ir showed towards BaP. NOS2 is Nitric oxide synthase, inducible (UniProtKB - P35228) which on interaction with BaP, gets upregualted (Sohn et al., 2008 and Lizarraga *et al.,* 2012). NOS2 has been reported as an emerging biomolecule which helps in progression of cancer (Thomas and Wink, 2017). High expressions of NOS2 have been reported in many tumors (Vannini *et al.,* 2015).

Molecular docking of NNK:

Molecular docking of finally selected 21 proteins was done with NNK in which CDK7 showed highest binding affinity of -5.93 Kcal/mol. CDK7 is also known as tumor suppressor protein (Zhong *et al.,* 2019). Its interaction with NNK causes perturbations in the cell cycle. CDK7 has been reported to play roles in cell cycle regulation as well as in transcription (Matthew *et al.,* 2002 and Fischer 2005).

Second molecule is CCNA1 (Cyclin A1) with binding affinity of -5.60 Kcal/mol. It was earlier not present in the seed proteins extracted from the clusters generated by MCODE. It is a connector which was incorporated in the network developed post MCODE seed selection step. The presence of CCNA1 in the top 3 biomolecular

targets,provides a hope of finding a novel target of NNK which has not been reported earlier. CCNA1 plays an important role in the cell cycle progression and any changes in the concentrations of CCNA1 will impact the cell cycle regulatory machinery.

Third molecular target of NNK was CDKN1B whose binding affinity is -5.42 Kcal/mol. It is also known as p27Kip1. It is an important regulator of cell cycle regulatory machinery. Mutated forms of CDKN1B have been reported in some types of tumors (Chang *et al*., 2004 and Belletti&Baldassarre., 2015). Any perturbations in the concentrations of CDKN1B would lead severe changes in the cell cycle regulation.

Biokinetics of cell cycle model

Mathematical modeling of highly complex biological processes has becoming essential to make a deep understanding of the system. It provides an opportunity of understanding how various components inside a system are interacting with each other. With the help of mathematical modeling, patterns for variability inside a cell can be studied (Mura *et al* 2019). In this study, we have designed a model of cell cycle regulatory machinery with the help of KEGG pathways (map045110). We designed the cell cycle using cell designer and then used SBML squeezer for the generation of kinetics. We have adopted the base kinetics from the model of Gerard. We have also used COPASI software for time course analysis of the model.

Mathematical modeling of BaP:

The mathematical model was designed for interaction of BaP with PTGS2(fig20). We have not modeled QSOX1 and NOS2 as they are not directly related with the perturbations in the cell cycle regulatory machinery.

Fig 21 is the graph obtained for the cell cycle when BaP was absent. When BaP was introduced into the model, changes in the concentrations were observed, not only in PTGS2 but also in the concentrations of active cyclin E CDK2 complex and p27 kip1. This happened because when PTGS2 gets upregulated, p27 gets hampered (Toyoshima *et al.,* 2002; Baldi *et al* 2004 and Sobolewski *et al.,* 2010) which in return hampers the functioning of cyclin E CDK2 (Fig 22a and 22b) (Table 3). Apart from change in concentrations, there are fluctuations observed in the oscillating patterns of main regulatory protein complexes (Table 4). From these fluctuations we can say that when PTGS2 gets overexpressed due to the presence of BaP, the precise timing of formation and degradation of the CDK-cyclin complexes have got hampered. Changes in the oscillation patterns of Cyc E CDK2 may lead to the progression of the cell cycle with faulty DNA whereas changes in the oscillations of Cyc B CDK1 complex may cause progression of cell into M phase with improper separation of chromosomes.

Mathematical modeling of NNK:

Mathematical models were developed for all the three top biomolecular targets of NNK namely CDK7, CCNA1 and CDKN1B (Fig 20 and fig 36).

Fig 37a is the graph obtained when NNK was not introduced in the cell cycle. When NNK was introduced in the system, functioning of CDK7 got hampered which further hampered the functioning of Cyclin ACDK2 and Cyclin B CDK1. As Cyc A CDK2 and Cyc B CDK1 are involved in the activation of CDC20as well as in the formation of complex with cyclin E and cyclin A respectively, massive fluctuations were observed in the graph obtained (Fig 38). Moreover CDC20 has role in synthesis of CDK-cyclin complexes, hence whole graph has shown changes in concentrations. From the observations shown in table 8, we can conveniently say that CDK7 can be seen as a very important cyclin in cell cycle as deviation in this can perturb the whole cycle.

Fig 39 is the graph obtained when NNK interacts with cyclin A. On interaction with Cyclin A, fluctuations have been observed in the concentrations of Cyclin A CDK1 complex as well as there is an increased degradation of complex(table 9). From this we can say that inhibiting Cyc A can lead to the problems associated with the transition of cell cycle from S to G2 phase and also in the transition to M phase. As cyclin A CDK2 complex helps in the activation of the Cyclin B CDK1, we can observe the changes in the oscillation patterns of Cyc B CDK1 complex(table 10).This can result in improper

separation of chromosomes during the mitosis. This will result in increased cell cycle proliferation which will ultimately lead to cancer.

On inhibition with NNK, there is a drastic drop in the concentration of p27. Apart from this, we can also observe that the time required for reaching half of its value has also decreased. The inhibition product also shows an initial rise in the concentration but soon it also starts decreasing gradually(fig 40). As a consequence of such a drastic drop in the concentration of available p27, the regulation of Cyc E CDK2 complex is also getting hampered due to which slight changes in the concentration as well as fluctuations in the oscillating patterns of Cyc E CDK2 complex was observed (Table 11 and Table 12). The changes in the oscillating patterns may lead to problems associated with DNA replication and other S phase related proteins. As a result, the cell cycle may get hampered leading to severe consequences like cancer.

Protection by carbon nanoparticles

Nanoparticles due to their size dependent properties are at the leading edge of nanotechnology. These are being extensively used for scientific advances as well as in commercial applications. Nanoparticles provide convenient surfaces for the assembly of molecules which makes them excellent carriers of drugs inside a system (Patra *et al*., 2018).Carbon based nanoparticles are also finding their extensive use in the drug delivery methods due to their high aspect ratio and bioavailability and nano fluid nature

(Aboofazeli. 2010). Apart from being used as suitable drug carriers, carbon nanoparticles are also potent scavengers of pollutants owing to their great adsorption capacities (Mauter and Elimelech 2008 and Khin *et al* 2012). Taking advantage of these properties of carbon based nanoparticles, we have tried to check the binding affinities and adsorption capacities of single walled carbon nanotubes, multiwalled carbon nanotubes and fullerenes against BaP and NNK inside biological systems using genetic based algorithms. The scavenging potential of the nanoparticles towards BaP and NNK are attributed because of the presence of charges on their surfaces along with their high surface area to volume ratio. Also there are π- π overlapping orbitals present on the surface of the CNTS which interact with the free electron cloud of the benzene rings present in BaP and NNK. This interaction is favourable as this reduces the electrostatic repulsion leading to minimization of the energy of whole system.

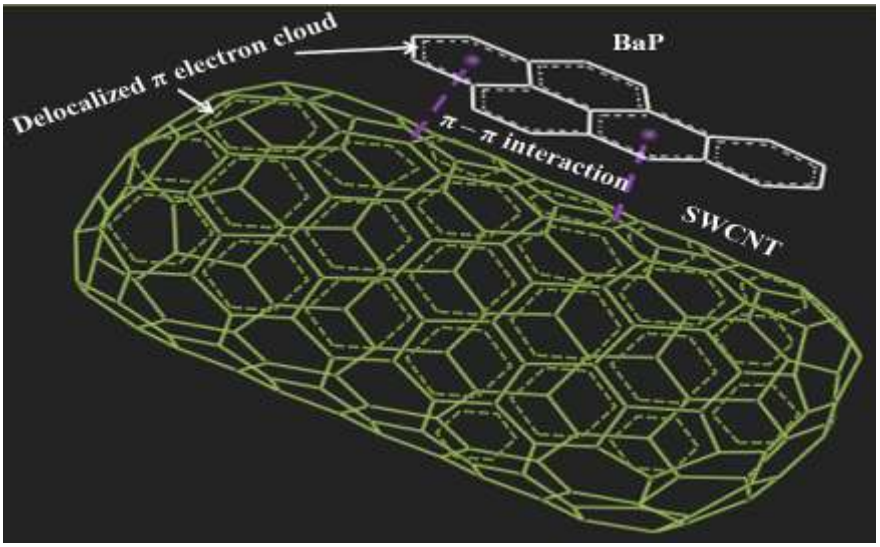


Fig 43: π-π interactions between BaP/NNK and CNTs reduces the electrostatic repulsion leading to BaP/NNK adsorption on CNTs

On performing the docking analysis, we have found that fullerenes are not capable of showing good binding energy as well as adsorption load due to its small surface area. The molecular docking results show that MWCNTs have highest binding affinity (-13.46 Kcal/mol) for BaP and has also shown highest adsorption capacity (11 molecules/NT) towards BaP. SWCNTs have lesser binding affinity (-10.32 Kcal/mol) and lesser adsorption capacity (9 molecules/NT) for BaP. For NNK, SWCNT has the highest binding affinity (-18.24 Kcal/mol) while the adsorption capacity (10 molecules/NT) followed by MWCNT which showed a binding affinity of -16.67 Kcal/mol and the highest adsorption capacity of 12 molecules/NT. In both the cases, fullerene molecules have showed the least binding affinity as well as adsorption capacities of -3.29 Kcal/moland 5 molecules/NT and -3.09 Kcal/mol and 6 molecules/NT for BaP and NNK respectively. From this we can easily say that MWCNTs are better options for scavenging BaPwhile both SWCNTs and MWCNTs are better option against NNK.