MATERIALSA NDMETHODS

Materials and Methods

3.1) Identification of biomolecular targets

Proteins that get hampered by the activity of BaP were extracted out using the Toxin and Toxin Target Database (T3DB) (Fig 3a) (Lim *et al.,* 2010 and Wishart *et al.,* 2015),also known as the "Toxic Exposome Database". Currently it manages the records of approximately 3680 toxins along with their target associations and is linked with "Human Metabolome Database" and "DrugBank". Approximately 1722 genes have been reported that have got upregulated and around 2369 genes have been downregulated.

For retrieval of the genes hampered by NNK, PubMed (Fig 3b) was used. PubMed is a database with more than thirty million biomedical research articles. With the help of keywords like NNK, environmental carcinogens, cell cycle etc., 1320 research articles were scrutinized from which 544 genes/ proteins were obtained that get hampered due to NNK.

a b

Fig 3: Databases used to retrieve genes for BaP and NNK

3.2) Protein-Protein interaction network

Protein- protein interaction networks are generated to find the interaction between various proteins that are reported to get up-regulated or down-regulated by the interaction of environmental carcinogens (BaP and NNK). To generate the network, STRING database (version 10.0 and version 11.0) (https://string-db.org/)was used (Szklarczyk *et al*. 2015 and Szklarczyk *et al*. 2019). STRING database generates the protein- protein interaction network based on the data generated by various high through put experiments, co-expression, automated text-mining data from various literature sources like pub-med and genomic context predictions. For experimental data, BIND, GRID, IntAct, DIP, HPRD, PID and MINT databases are used. STRING extracts the curated data from GO, Reactome, KEGG, BioCyc and Biocarta databases. For text mining, STRING database uses scientific articles from PubMed and OMIM and for protein based co-expression data, Proteome HD database is used (Szklarczyk *et al.* 2019). In both the networks, 50-50 connectors were added to both first and second shells. The confidence for "minimum required interaction score" was kept at highest level of 0.9. Cytoscape was used to merge upregulated and downregulated networks of BaP generated by STRING software. Both STRING and Cytoscape (Fig4) were used in the generation of networks for BaP and NNK.

Fig 4: Softwares used for generating and analyzing of BaP and NNK rewired PPINs

3.3) Network Analysis

Once the network has been generated by STRING database, cytoscape software (version 3.6.1) was used to analyse the network (Shannon*et al.,* 2003). Creating a network provides information of various interactions between different genes and proteins. To get an indepth knowledge about the network and the proteins involved in it, its analysis was done. The analysis of the network was done based on the topological properties like degree, average degree distribution, average clustering coefficient, shortest path length, betweennesscentrality,bottleneck etc. The basic analysis of the network is done using the inbuilt network analyzer app provided by Cytoscape. For the indepth analysis of each node, CytoHubba plugin of Cytoscape software is used (Chin *et al.,* 2014). Analysis of topological properties also helps in reducing the noise present in the network.

3.4) Modulation and Pathway enrichment

Clusters or modules are the groups of closely interacting proteins that help each other in passing the information to accomplish a function inside a system. Modulation is a process of reducing the noise that is present in the network. It provides an opportunity to find subgraphs present within the main network that perform same function. Molecular Complex Detection (MCODE) (Fig 5a), a plug-in provided by Cytoscape, was used to find clusters. It helps in finding highly interlinked clusters in the network (Bader and Hogue, 2003).

To remove the noise from the main network, seed proteins were detected from the clusters generated by MCODE and a new network of seed proteins along with 50-50 connectors in both the shells was generated using STRING database. Topological analysis was done using CytoHubba plug-in of Cytoscape and most important proteins were extracted out based on their degree, clustering coefficients, betweenness score and bottleneck score. To further refine the list, Median of the bottleneck scores of selected proteins was calculated and all the proteins till calculated median score were preferred for further analysis.

Pathway enrichment was done using ClueGo(version 2.5.1) (Fig 5b) plug-in of Cytoscape software. This is used for the analysis of the interrelated functional groups within the network. For analysis, threshold p-value was kept < 0.05 with Bonferroni correction method and along with two-sided hypergeometric test (Bindea*et al.*, 2009).

3.5) Molecular Docking Analysis

Molecular Docking was performed on all the selectedproteins extracted after applying median. Docking simulation helped in understanding the binding patterns of the environmental carcinogens (BaP and NNK) on their respective proteins with the help of binding energy score and the inhibition constant (Ki) value. For molecular docking Autodock 4.0(Fig 5c) from MGL suite (Morris*et al.* 1998) was used and all the molecular docking simulations were performed on CPU 1.4 GHz, AMD E1-6015 APU

processor, 4 GB RAM on Hewlett-Packard (HP) machine. All the docking parameters applied in this study were based on a study published by Dhasmana A *et al*in 2016.

The *.pdb* structures were extracted from RCSB PDB database (Berman, 2000) and Uniprot server (The UniProt Consortium, 2019) and I-TASSER server (Yang and Zhang, 2015) was used for those proteins whose *pdb* structures were not available. Further refining of such structures was done usingModRefiner server(Xu and Zhang 2011) and Rampage Ramachandran Plot Analysis software (Lovell *et al* 2002).

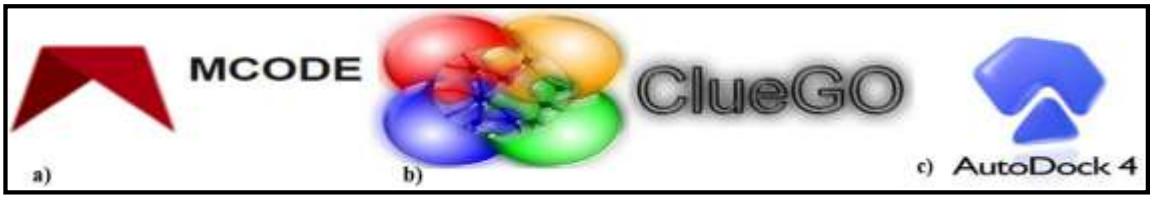

Fig5: For finding the most potent biomolecular targets for BaP and NNK

3.6)Construction of human cell cycle model and time course simulation

Human Cell cycle for BaP interactions and for NNK interactions were mainly designed based on the KEGG pathways using Cell Designer version 4.4(fig 6a)(Funahashi, 2003). For the complete formation of human cells for BaP and NNK respectively, biomodels from Biomodel database were used (Novère, *et al.* 2006 and Li, *et al* 2010). When the models were complete formed, SBMLsqueezer v2.1(Fig 6b) was used to for incorporation of desired kinetics (Drager*et al.*, 2008). Concentrations for all the proteins were extracted out from Proteomics DB (Schmidt *et al*. 2017)and Protein Abundance Database (Wang *et al.,* 2015) and the model was run for time course analysis using COPASI (Fig 6c) (Hoops*et al* 2006).

Fig6: For designing cell cycle model and time course analysis

3.7) Protection by carbon-based nanoparticles

Single walled carbon nanotubes, multiple walled carbon nanotubes and fullerenes were designed using nanotube modeler(Fig 7a)which generates the *xyz*coordinates of the nanotubes (Melchor and Dobado, 2004). The parameters of the designed single walled carbon nanotubes (SWCNTs) were bond length 1.421nm, tube length 2.5nm, whereas the parameters of multi walled carbon nanotubes (MWCNTs) were with bond length 1.421nm, two wall layers and tube length 2.5 nm, while the parameter of fullerene was with diameter of 1nm.Genetic based algorithm was used to calculate the binding energies of SWCNTS, MWCNTS and Fullerenes against BaP and NNK. BIOVIA Materials Studio BLENDS(Fig 7b) was used to calculate the binding energy and the adsorption potential of SWCNT, MWCNT and Fullerene with BaP and NNK.

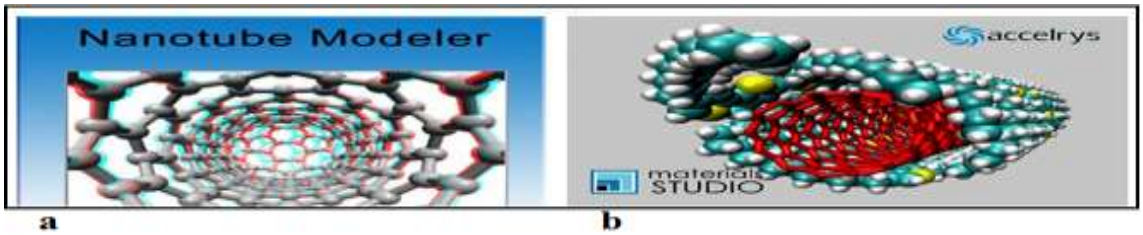

Fig 7: For designing nanoparticles and analyzing their scavenging potential against BaP and NNK

Figure 8 represents the flowchart of the steps followed to find the most probable biomolecular targets of BaP and NNK followed by the investigation of the most suitable

carbon nanoparticle as a scavenger against BaP and NNK. The methodology used has already been published (Dhasmana et al. 2020).

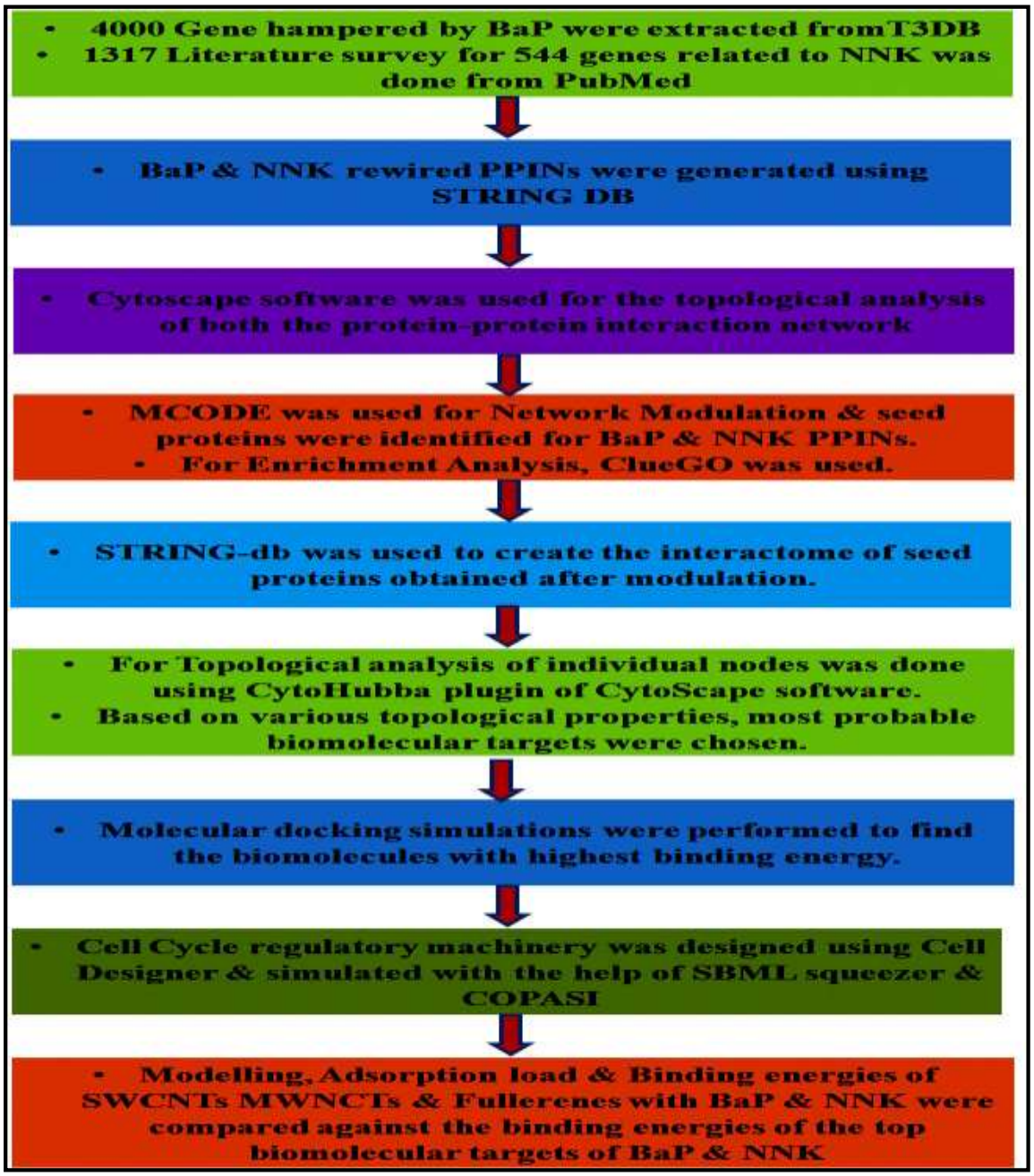

Fig 8) Flowchart of the methodology adopted for finding the most potent biomolecular targets of BaP and NNK using systems biology approach and for finding the most suitable scavenging agent against BaP and NNK.