OPEN ACCESS Freely available online

Current Synthetic and Systems Biology

Research Article

In Silico Study of STAT3 and its Role in MMP Mediated Metastasis

Sunil Kanti Mondal[1*], Samyabrata Mukhoty[1]

[1]*Department of Biotechnology, The University of Burdwan, India*

ABSTRACT

Metastasis is one of the major reasons for cancer associated deaths. Constitutive activation of Signal Transducer and Activator of Transcription 3(STAT3) by a variety of the factors, such as Cytokines and Growth factors within the cytoplasm, from where it is transported into the nucleus and promote the transcription of various anti-apoptotic and metastatic genes like MMP2, MMP7, MMP9, as a transcription factor and leads to several different forms of Cancers. Multiple Sequence Alignment and Jalview analysis reveals that the human STAT3 is highly conserved, with no changes along the entire length of the human STAT3 protein and maintaining close resemblance with the selected organisms. The aim of our systematic study is to identify the residues of STAT3 interacts with its nuclear transporter protein Importin and the promoter of MMP genes. So that specific inhibitor can be designed, this will block the identified interacting sites of STAT3. Furthermore, finding for interacting proteins of STAT3, MMP2, MMP7 and MMP9 from String database and differential expression studies of STAT3 and the respective MMP2, MMP7 and MMP9 in different organs led us to conclude that specific drug molecules has to be designed to block the interacting sites of STAT3 to overcome the metastatic spread of cancers by these MMPs. The Virtual screening studies by targeting the amino acid residues of STAT3 responsible for interacting with the promoter of MMPs led us to conclude that the small molecule Deferoxamine is able to be fruitful in minimizing the cancer associated metastatic spread.

Keywords: Metastasis; Matrix metalloproteinase; Signal Transducer and Activator of Transcription 3; molecular docking; drugs; virtual screening.

INTRODUCTION

Cancer is currently the second leading cause of death globally after Cardiovascular disease, with an approximation of 9.6 million deaths, and the ratio is one in six deaths, in 2018, while a significant increase of the death rates is due to the cancer associated metastatic spread. The alarming increase in the death rates for cancers have got the attraction of the researchers across the world for designing effective treatments to mitigate the rate of deaths associated with various forms of cancers. The Signal Transducer and Activator of Transcription 3 (STAT3) is a transcription factor, known for up regulating various anti-apoptotic and cell cycle progression genes. STAT3 of human is 770 amino acids long.

Previous researchers have long been reported that STAT3 can be activated by a variety of factors and is observed to remain constitutively activated in multiple cancers, including Prostate cancer (Abdulghani et al., 2008; Doncow et al.,2017), Breast cancer (Sonnenblick et al.,2012), Head and neck cancer (Geiger et al.,2016), Renal cell carcinoma (S. Li et al.,2013), Liver cancer (He et al.,2010), Pancreatic ductal Adenocarcinoma (Fukuda et al.,2011), Lung cancer (Chang et al.,2012), Acute Myeloid Lymphoma (Redell et al.,2011) and various other types of cancers.

STAT3 can be activated by a variety of factors, such as, certain tyrosine kinases that are being termed as Receptor Tyrosine Kinases (RTK), which gets activated when certain growth factors like Vesicular Endothelial Growth Factor (VEGF), Epidermal Growth Factor binds to their respective receptors. These RTKs

***Corresponding Author:** Sunil Kanti Mondal, Department of Biotechnology, The University of Burdwan, India;Tel: + 9064081134; Email: skmondal@biotech.buruniv.ac.in

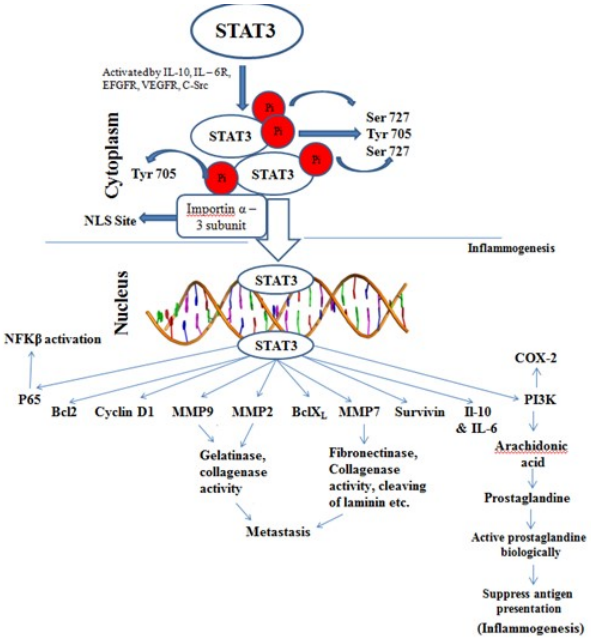
Received date: August 24, 2020; **Accepted date:** August 31, 2021; **Published date:** September 13, 2021

Citation: Mondal SK (2021) In Silico Study of STAT3 and its Role in MMP Mediated Metastasis. Curr Synthetic Sys Biol.Vol.9. p244.

phosphorylate the tyrosine and serine residues of STAT3, that is present on site 705 and 727 respectively (Figure:1) of the trans activation domain of STAT3 (Sakaguchi et al.,2012; Schuringa et al.,2000; Sgrignani et al.,2018). Besides STAT3 can also be activated by Janus Tyrosine Kinases (JAKs) and certain other tyrosine kinases like C-Src etc. (Lau et al.,2019).

Once unphosphorylated STAT3 becomes phosphorylated, the TAD of two STAT3, containing the phophorylating Tyr residues tend to dimerize (Yu et al.,2007). The STAT3 homodimer is then transported into the nucleus by importin, typically by binding with the Nuclear Localizing Sequence(NLS) binding motifs or Armadilo (ARM) repeats of the Importin alpha-3 subunit (Lui et al.,2005), eventually binds through it's DNA binding Domain into the promoter region of the target genes, bringing about the induction of the transcription of respective genes, including c-myc, bcl-xL and other anti-apoptotic and cell cycle progression genes (Figure:1) (Yu et al.,2007; Lee et al., 2009; Grivennikov et al.,2009; Huynh et al.,2019).

Figure 1: Flowchart representing the Activation of STAT3, it's nuclear transportation and upregulation of the transcription of anti-apoptotic and metastatic genes. The activation of IL-10 and IL-6 exerts positive feedback loop for STAT3 activation.



Although it acts as a transcription factor for a set of anti-apoptotic and cell cycle progression genes, STAT3 homodimer is also being reported in enhancing the transcription of three matrix-metalloprotinase (MMP) genes(Figure:1), namely, MMP2 (Xin et al.,2004; Zhang et al.,2015; Kamran et al.,2013), MMP7 (Fukuda et al.,2011) and MMP9 (Jia et al.,2017; Zhang et al., 2015), which are associated with cancer associated metastatic spread.

The MMPs are being termed as Zinc dependent endopeptidases (Verma et al.,2007) and is typically synthesized in the latent pro-enzyme form in normal cellular conditions, in which the Pro domain (positioned adjacent to the signal sequence) interacts with the zinc binding residues, that is known to be present within the Catalytic domain of respective MMPs. This interaction is mediated by the help of conserved Cysteine residue and brings about the blocking of zinc binding to the zinc binding motif within the active site of MMPs, resulting in the inhibition of cleavage of the substrates of respective MMPs. This Cysteine residue is known as Cysteine switch (Rosenblum et al., 2007). However, the inflammatory cell has long been reported to have the poteintiality in producing large amount of Reactive Oxygen Species (ROS), which in it's turn may exhibit the MMP activation, typically by oxidizing the cysteine residue present within the pro-domain (Weiss et al.,1985; Kessenbrock et al., 2011).Upon the activation, the MMPs cleave certain proteins, specific for each of the classes of MMPs, for example, MMP2 and MMP9 are being termed as Gelatinase, because of the capability of degrading the Gelatin (Devarajan et al.,1992). Furthermore, MMP2 is also known for the activation and cleavage of Transforming Growth Factor-beta (TGF-beta), the activation of which facilitate the Epithelial to Mesenchymal Transition(EMT), which is one of the hallmarks in cancer (Gialeli et al.,2011). On the other hand, MMP7 is being reported for cleaving a number of substrates, including Laminin, Fibronectin, collagen type IV, and also have the tendency of cleaving the MMP2 and MMP9 to active them from their latent form (Edmen et al.,2011; Yokoyoma et al.,2008). Therefore, these MMPs can cleave a number of proteins of the Extracellular membrane, bringing about the metastatic spread of cancers from the primary site to the other parts of the body.

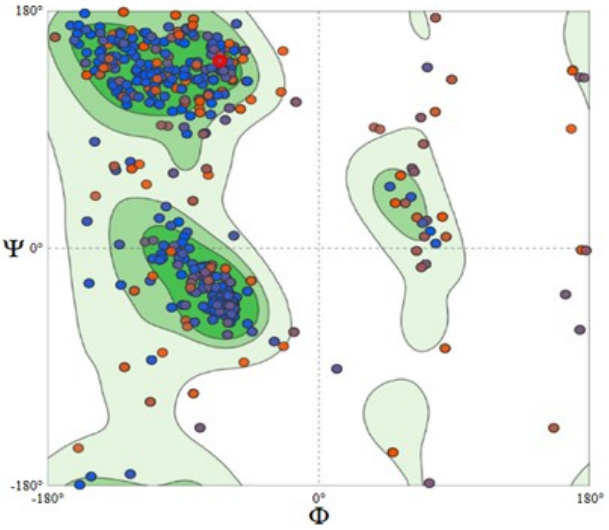
The significant increase in the rate of cancer affected patients urgently demands for effective therapeutics necessary to combat cancers. Previous researchers have mentioned about the activity of STAT3 as a transcription factor in regulating the expressions of anti-apoptotic and metastatic genes. But, the detailed analysis still needs to be done regarding the conservations of STAT3 in the amino acid sequence level, the combinatorial expression levels of STAT3 and the MMPs mentioned above and also the identification of the residues of STAT3 that interacts with the promoter region of MMPs. In this article, we have investigated the transportation process of STAT3 from cytoplasm into the nucleus upon binding to the alpha subunit of Importin and identified the amino acids of STAT3 responsible for the interaction with Importin. Furthermore, we have also identified the interacting amino acid residues of STAT3 responsible for binding to the promoter regions of these genes, so that effective drug molecules can be designed in near future to overcome the cancer related metastatic spread. We have also identified the mutable positions within the STAT3 amino acid sequences among the selected organisms, although the STAT3 of human appears to be highly conserved with zero observed mutations. Furthermore, we have conducted gene ontology studies to identify the biological functionalities of these MMPs and also identified the conserved domains of MMP2, MMP7 and MMP9 by Conserved domain analysis. The molecular docking study helps us to identify the amino acid residues of STAT3 necessary for interacting with the promoters of MMP2, MMP7 and MMP9. Lastly, we have gone through Virtual screening test by targeting the sites of STAT3, necessary for the interaction with MMP promoters in search for effective drug molecules to block these interacting sites.

STAT3 modelled structure

Amino acids are joined with one another, firstly in a linear fashion for forming the primary structure of a protein. Each of the amino acids in a protein structure are joined in such a way that the carboxy terminus (-COOH) of the first amino acid is joined with the Amino group (-NH3) of the incoming second amino acid by formation of a peptide (CONH2) bond with the exclusion of one molecule of water(H2O). These peptide bonds are observed to be rigid and planner and the central alpha carbon of each amino acid is maintained within the main chain by two rotable bonds. The torsion angle of these two bonds, also known as their di-hedral angles are being termed as Phi(between N and Calpha) and Psi(between Calpha and C). These torsion angles is known for providing the flexibility to the backbone of a protein to adopt a certain fold.

Ramachandran plot is a way of visualizing the energetically allowed and disallowed regions for the di-hedral bond angles phi against psi of amino acid residues contained within a protein structure. Here, after constructing the modelling for STAT3 homodimer in Swissmodel, the required Ramachandran plot is also obtained, by the help of which, we can conclude about the kind of the confirmations obtained by the protein that has been modelled.

Supplementary Figure1: Ramachandran plot showing the allowed tortional angles of amino acids of the modelled STAT3 homo-dimer.



MATERIALS AND METHODS

Retrieval of the Sequences

At first the amino acid sequence of Signal Transducer and Activator of Transcription3 (STAT3) for human is collected from National Centre of Biotechnology Information (NCBI) database https://blast.ncbi.nlm.nih.gov/). After the retrieval of the protein sequence of human STAT3 (accession: P40763), the retrieved sequence is given as query and the homologus sequence are obtained using BLASTP 2.3.32 (developed and maintained by National Centre for Biotechnology Information; and freely accessible from https://blast.ncbi.nlm.nih.gov/Blast.cgi). The selection criteria for homologus sequences are based on e-values (threshold e-value<0.01) in BLASTP search (Atschul et al.,1997) and only one sequence from each organisms is collected. The resultant homologus sequences for 25 organisms are selected and saved in FASTA format, which are subjected to further processing.

Multiple sequence alignment Analysis

Aligning of the homologus sequences provide necessary information about the mutations occurred along the length of the conserved equivalent regions, which in turn proved to be fruitful in defining the evolutionary relationship among the organisms of interest. Protein sequence alignment is now a day considered as one of the most important steps in bioinformatic and biomedical research fields. The retrieved sequences are subjected to undergo multiple sequence alignment by Clustal Omega tool (freely accessible from https://www.ebi.ac.uk/Tools/msa/clustalo/) and the result is further viewed by using Jalview option to obtain well defined normalized sequence logos.

Promoter analysis for Transcription Factor binding

We have analyzed the promoter region of MMP2 (ENSG00000087245), MMP7 (ENSG00000137673) and MMP9 (ENSG00000100985) genes by taking the FASTA format of the gene sequences upto 2000 flanking sequences in 5'UTR and 200 flanking sequences in 3'UTR of the coding region of the MMPs mentioned above from Human Ensembl database (https://asia.ensembl.org/index.html, Date- 09-06-2020, Time- 6.25 p.m.). The obtained sequences are uploaded in the GP-MINER (Lee et al., 2012) web based tool (http://gpminer.mbc.nctu.edu.tw/search.php) for the identification of the STAT3 transcription factor binding site and nucleotide character pattern of the binding site.

Designing of the promoter structure

After getting the nucleotide sequence of the STAT3 binding site within the promoter region of MMP2, MMP7 and MMP9 by the help of GP-MINER, we have taken the nucleotide sequences located upstream to the annotated TSS of MMP2, MMP7 and MMP9 genes. From the multiple STAT3 binding sites, we have considered only one sequence, whose nucleotide character pattern exactly matches with that of the nucleotide sequences among the predicted STAT3 binding site on promoters in the 5'-3' direction of the flanking sequence of MMP2, MMP7 and MMP9 individually. We have further extended the nucleotide sequence of the selected region of MMP promoters and a three dimensional DNA structure of twelve bases length is designed for each of the MMP promoter regions by using 3D-DART online web-server (https://milou.science.uu.nl/services/3DDART/). We have chosen the following parameters for designing of the DNA structure- Number of repeats=1; Nucleic acid type= B-DNA; Modelling mode= Local and we have used the customized base pair steps and nucleotide parameters for the construction of these three dimensional DNA structures.

Conserved Domain analysis of MMPs

We have identified the conserved domains present within MMP2, MMP7, MMP9 and STAT3 by ScanProsite tool, which is accessible freely from https://prosite.expasy.org/scanprosite/ (Date- 09-06-2020, Time- 6.40 p.m..)

We have also gone through NCBI Conserved Domain Search tool (search against database: pfam v32.0 – 17919 PSSMs) (accessible freely from https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for better validation of our result.

Interacting protein analysis

We have identified the interacting proteins for STAT3, MMP2, MMP7 and MMP9 in String Database, which is a database for functional protein association networks to identify the variety of factors interacting with STAT3, MMP2, MMP7 and MMP9.

Molecular Docking study

First, due to the unavailability of any PDB structure for human STAT3 homo-dimer, we have the structure of STAT3 homo-dimer with the help of Swissmodel (https://swissmodel.expasy.org/interactive) by directly putting the obtained FASTA sequence of STAT3 protein (https://swissmodel.expasy.org/interactive/4wFCXw/models/). The aim of doing so is to maintain the parity of our analysis with that of the biological system. After the modelling is done, we have chosen only one template (1bg1.1.B –STAT3B/DNA complex), the sequence of which is>99% identical with the input sequence. The resultant PDB file for this structure is saved along with the Ramachandran Plot given for this structural model (Supplementary Figure1). The Ramachandran plot favoured region for this structure is 88.06%, which indicates the fully allowed region of tortion angles. While, the Ramachandran outliers(those amino acids that exhibit non-favorable tortional angles) is 5.02%, which should not exceed >10% for a protein structure be treated as good (Balaji et al.,2006) and that of the bad bonds are 5 out of 9502, as indicated by Ramachandran plot generated in Swissmodel for STAT3 homo-dimer protein structure.

Next, based on the previously published literatures, two types of molecular docking study is done-

Protein-Protein molecular docking study

We have chosen Importin-alpha3, as it is being reported to bring about the nuclear transport of STAT3 from cytoplasm (L. Lui et al.,2005). The associated information related to this protein and the FASTA sequence. The molecular docking between modelled STAT3 homodimer and Importin-alpha3 is done using HDOCK online server, which is freely accessible from http://hdock.phys.hust.edu.cn/ (Yan et al.,2017). The resultant PDB file for this receptor-ligand interaction is assessed by UCSF Chimera (https://www.cgl.ucsf.edu/chimera/) for finding the contacts between receptor-ligand interactions including information regarding hydrogen bond formation.

Protein-DNA molecular docking study

The protein-DNA molecular docking study is done between the STAT3 homodimer and MMP2, MMP7, MMP9 promoter DNA individually by using the HDOCK online server. The interaction between the receptor-ligand is identified using Protein Ligand Interaction Profiler (PLIP), which is freely accessible from https://projects.biotec.tu-dresden.de/plip-web/plip.

We have also conducted a series of molecular docking by choosing four drugs, namely Galiellalactone, all of which has been reported to block the DNA Binding Domain by previous researchers (Bharadwaj et al.,2016).

Virtual Screening study for STAT3

We have also conducted Structure Based Virtual Screening (SBVS) test using e-LEA3D (Li et al.,2017)by putting the constructed model of STAT3 homo-dimer to obtain the list of small molecules from different online resources that might act as the potential inhibitors of STAT3 in order to block the amino acids residues of this transcription factor that are responsible for the binding to MMP promoters. After putting the modelled STAT3 structure in the server, we have chosen the target amino acid residue of interest and put the three letter code of the chosen amino acid along with the specific position within the STAT3 structure and also the name of the chain in which the particular residue is present. After that, we have chosen docking with PLANTS option, which offers the modules for SBVS computations. After receiving the list of the drugs, we have conducted another series of molecular docking studies taking the listed drugs individually and the modelled STAT3 structure to check whether any of these listed drugs interacts with the amino acid residues of STAT3 that are found to be responsible for interacting with the promoters of MMP2, MMP7 and MMP9 through our docking studies mentioned above.

Expression Analysis

We have also analysed the differential gene expression levels of STAT3, MMP2, MMP7 and MMP9 in various organs from Human Ensembl database (accessible freely from https://asia.ensembl.org/index.html, Date- 09-06-2020, Time- 07.30 p.m.). Before that, we have observed the occurrence of STAT3 mediated cancers from Disgenet database (accessible freely from https://www.disgenet.org/rdf, Date- 09-06-2020, Time- 07.55 p.m.). We have collected the expression levels of these genes in selected organs by filtering the data and selecting the organs of our interest. From that, we have prepared a heatmap representation of the selected expression data for these four genes through R statistical package (Voorrips et al.,2002).

RESULT AND DISCUSSION

Multiple Sequence Alignment and Jalview Analysis

Multiple sequence alignment (MSA) is a process to bring the equivalent portions of the sequences in the same position. In this study, we have taken the sequence of STAT3 from 25

different selected organisms. The aim of this analysis is to observe the conservations of amino acid sequences and the sequence similarities of STAT3 for Human with that of the other selected organisms. The result is then viewed by Jalview, and observed the changes in the amino acid residues among the selected organisms in all 20 mutable positions, although the amino acid sequences of STAT3 among the selected organisms seemed to be highly conserved. As a result, the non-conservations within the conserved sequences might reflect the evolutionary relatedness among the selected organisms, and also their divergence from each other (Figure:2).

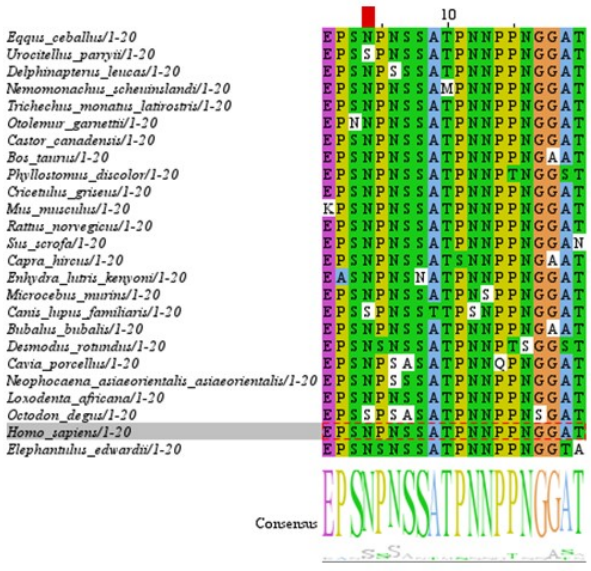
The position 16 is mostly occupied by Glutamic acid, which is a negatively charged amino acid, except only for Mus musculus, in which Glutamic acid is substituted by positively charged Lysine.

The position 192 is almost occupied by polar and hydrophilic amino acid Asparagine, except for Delphinopterus leucas, Neophocaena asiaeorientalis asiaeorientalis Octodon degus and Cavia porcellus, in which Asparagine is replaced by another polar and hydrophilic amino acid Serine.

The position 699 is almost occupied by Proline, which is a polar amino acid, except for Phyllostomus discolor and Desmodus rotundus, in which Proline is replaced by Threonine, which is also a polar and hydrophilic amino acid(Figure:2).

The position 748 is almost occupied by non-polar Alanine, except for Phyllostomus discolor, Desmodus rotundus, in which Alanine is replaced by polar-hydrophilic amino acid Serine and Elephantulus edwardii, in which Alanine is likely to be replaced by Threonine, which is a polar-hydrophilic amino acid.

Figure 2: JalView representation of Multiple Sequence Alignment of mutable positions of STAT3 in 25 organisms. Amino acid positions within STAT3 from human are as follows: 16, 36, 113, 130, 132, 192, 194, 216, 227, 234, 336, 485, 664, 695, 699, 720, 741, 743 and 748; displayed as successive sites of jalview.



From the above analysis, it can be concluded that the STAT3 for Human is highly conserved with no observed mutations along the entire length of Human STAT3, which is also reflected by the height of the sequence logos in Jalview. Amino acids of STAT3 from human maintaining parity with most of the selected organisms in the respective positions those are mutable for other very few organisms.

Promoter Analysis

Transcription factors have known to interact with the promoter region, typically by binding with several consensus nucleotide sequences located upstream to the coding region of DNA, which is being considered as the Transcription Start Site (TSS) with the help of the DNA binding domain of the transcription factor. STAT3, being a transcription factor also exhibit similar mechanisms and ultimately binds with several nucleotides located upstream to the TSS.

We have gone through the promoter analysis study by the help of GP-Miner software, which is an online software. The aim of doing this is to identify specific STAT3 binding consensus nucleotides on MMP2, MMP7 and MMP9 genes. We have to put the FASTA formats of corresponding genes, whose STAT3 binding sites are to be identified. We have collected the FASTA format of the gene sequence ofMMP2, MMP7, MMP9 individually from Ensembl database, where the coding regions of the DNA are clearly marked. We have taken 2000 flanking sequences in 5' UTR , located upstream to the TSS and 200 flanking sequences in 3'UTR, downstream to the TSS from Ensembl (https://asia.ensembl.org/index.html, Date-09-06-2020, Time- 6.25 p.m.). The obtained FASTA formats of the gene sequences of MMP2, MMP7, MMP9 are then putted into the GP-Miner software individually.

STAT3 transcription factor recognizes specific TTCCC consensus residues, on the promoter region of each of these MMPs, located within the positive DNA strand as indicated by GP-Miner. For each of the three gene sequences taken, we have observed the presence of this consensus residues with some other variable residues located upstream to the annotated TSS. We have considered only one out of several STAT3 binding sites, which is located upstream adjacent region of the annotated TSS and for which besides the consensus nucleotides (TTCCC), the positions of the variable nucleotides also matches exactly with the transcription factor binding sites for STAT3 indicated by GP-Miner. The STAT3 binding sites located upstream to the TSS and has the nucleotide character pattern gaaTTCCC, tctTTCCC and actTTCCC for MMP2, MMP7 and MMP9, respectively (Figure:3).

Figure 3: Selected STAT3 binding sites on MMP2, MMP7 and MMP9 gene promoter, as indicated by GP-Miner. For MMP2 and MMP7 obtained 1.000 as Core and Matrix Score 1.000 and for MMP9 1.000 and 0.995 as Core score and Matrix score respectively. Position and sequence of the STAT3 binding sites (+ strand) are indicated at the above and bottom of the MMP specific bar of length 1200 nucleotides.

From the promoter analysis of MMP2, MMP7 and MMP9, it can be concluded that the transcription factor STAT3 binds to the promoter region of these peptides and brings about the upregulation of the transcription of these MMP genes. Besides, it can also be concluded that this transcription factor results in the more efficient transcription of MMP2 gene upon binding to the transcription factor binding site than MMP7 and MMP9, because the transcription factor binding site for STAT3 in MMP2 gene is located in the closest upstream proximity (figure 3) of the annotated TSS than that of the MMP7 and MMP9, as reflected by their nucleotide sequence level.

Conserved Domain analysis

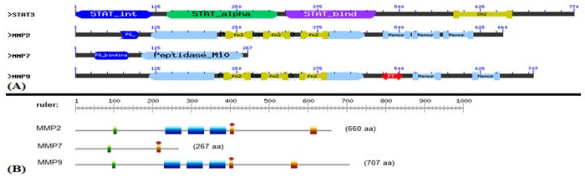
NCBI Conseved Domain Search tool identifies the presence of four conserved domains along the entire length of STAT3 protein. These are, STAT3 int domain(2-120) for various protein protein interactions, Coiled-coil domain(139-318), SH2 domain(554-715) for interaction with the tyrosine kinases ,which is required for STAT3 activation and DNA-binding domain(321-484) for binding to the promoter regions of the genes as transcription factor(Figure:4A)took placet is being observed that, only a single amino acid changes took place within the DNA Binding Domain, at positions of 336(Pro - Ser) and three amino acid changes took place within the SH2 domain, at the positions of 663(Asp - Ser), 695(Pro - Gln) and 699(Pro - Thr) for another set of organisms, as revealed by JalView analysis (Figure:2). Although, the physico-chemical properties of the amino acids for all of the above mentioned positions appears to be same in comparison with the conserved amino acids for that positions.

We have also identified the conserved domain of MMP2, MMP7 and MMP9 through the same search methods. MMP2 is observed to contain a single putative peptydoglycan binding domain (70-97), consisted of three alpha helices, a single Peptidase M_10 domain (118-446) for cleaving the peptides, which in turn is consisted of three fibronectin type II (FN2) repeats (226-274, 284-332 and 342-390) (Figure: 4), and a Hemopexin (HX) like repeat (466-660), as indicated by NCBI Conserved Domain Search tool. On the other hand, ScanProsite identifies the presence of three FN2 repeats (233-274, 291-332 and 349-390) for MMP2. Also, the presence of Cysteine switch (100-107), Active site (400-409) and Hemopexin domain (606-621) is identified by ScanProsite tool (Figure: 4).

Similar conserved domains have also identified for MMP9; i.e., the Peptidase M_10 domain(115-444),consisting three FN2 repeats(223-271, 281-329 and 347-388) and Hemopexin domain(514-704), as identified by NCBI Conserved Domain search tool(Figure: 4A). No peptydoglycan binding site is observed in case of MMP9, as indicated by NCBI Conserved Domain search tool. Similarly, ScanProsite identifies the presence of Cysteine switch (97-104), FN2 repeats (230-271, 283-329 and 347-388) and Hemopexin domain(556-571) for MMP9. Also, the Active site (398-407) of MMP9 is identified by ScanProsite (Figure: 4).

The NCBI Conserved Domain search tool identifies the presence of Peptidase_M10 domain (103-259) for cleaving the peptides, but no FN2 repeats are found within this domain. Besides, the putative peptydoglycan binding domain(31-82) is also present in case of MMP7, as indicated by NCBI Conserved Domain search tool(Figure: 4A). ScanProsite also confirms the absence of FN2 repeats and Hemopexin domain in case of MMP7. Although the presence of Cysteine switch (85-92) and Active sites (211-220) are identified for MMP7 by ScanProsite tool(Figure: 4).

Figure 4: analysis of the conserved domains of STAT3, MMP2, MMP7, MMP9 by NCBI Conserved Domain search tool. 4(B): positions of Cystiene switch (green), FN2 repeats(blue) , Zinc binding domain (yellow with pointed red at the top) and Active site (yellow) of MMP2, MMP7 and MMP9 by Scanprosite tool.

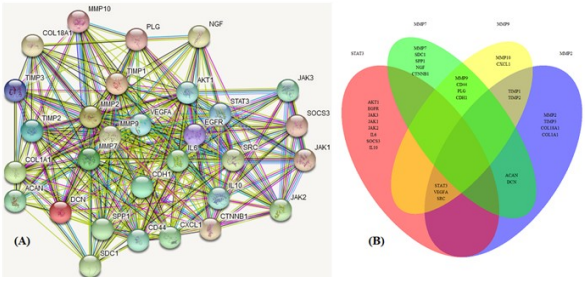


From all of the above analysis, it can be concluded that the MMP2 and MMP9 are functionally similar, as the presence of identical conserved domains is observed for MMP2 and MMP9, which also gives the justification in putting them under the same class of metalloproteinase.(Introduction e reference somet bola achhe). On the other hand, MMP7 is functionally dissimilar from that of the MMP2 and MMP9 as it lacks the FN2 repeats and Hemopexin domains. The absence of FN2 repeats and Hemopexin domain might be a cause that MMP7 is not observed to exhibit the substrate specificity in comparison to MMP2 and MMP9.

Interacting Protein network Analysis

We have searched for the interacting networks in String database for STAT3, MMP2, MMP7 and MMP9 for better understanding about the activation mechanism of metalloproteinase genes. We have found that the transcription factor STAT3 can be activated by a variety of the factors/protein, including Jak1, Jak2, Jak3, IL-10, IL-6, Src, VEGFA and EFGR, as indicated by the String database. Besides, String database clearly indicates that STAT3 establishes direct interacting networks with MMP2 and MMP9. Also STAT3 directly interacts with VEGFA, which in turn interacts with MMP2 and MMP9, bringing about the activation of metastatic cascade as indicated by String database (accessible freely from https://string-db.org/cgi/network.pl? Date-09-06-2020, Time-07.05 p.m.).

Figure 5: Interacting proteins for (A) STAT3, MMP2, MMP7 and MMP9 obtained from String Database and (B) Venn Diagram to understand the common interacting set of proteins.



MMP7 can interact with MMP9 directly and MMP7 also interacts with CD44 receptor, which is a receptor for Hyaluronic Acid (interacts with cell-cell and cell-matrix) and cell surface proteoglycan Syndecan (SDC1), which establishes the link between cytoskeleton and interstitial matrix, as indicated by String database (Figure: 5).

With reference to the Venn diagram (Figure: 5b) we can conclude as follows: MMP 2 and MMP 7 interacts with Aggrecan (ACAN) and DECORIN (DCN). Aggrecan is the major proteoglycan in the articular cartilage. The catabolism of ACAN by MMPs is said to cause cartilage tumors. It is also said to interact with DECORIN (DCN). MMP 7 and MMP 9 interacts with CD 44, PLG and CDH1 out of which CD44 and CDH1 are important in the field of cancer. CD44 is a cell surface hyaluron receptor that can cause enhanced metastasis in breast cancer. It is also said to interact and bind with MMP 9 at cell surface to initiate metastasis. It also contributes to prostate cancer. CDH1 also called E-CADHERIN is said to may cause metastasis if its expression is lost. It plays an important role in maintaining tissue skeleton. Loss of expression of CDH1 can cause ovarian cancer. MMP 2 and MMP 9 interacts with TIMP1 and TIMP 2.their expressions are seen in human prostate cancer and colorectal cancer and other forms too.

Apart from this, MMP 2, MMP 7, MMP 9 exclusively interact with certain proteins that might have the role in metastasis. MMP 2 interacts with TIMP 3, COLI8A1, COLIA1; MMP7 interacts with SDC 1, SPP1, NGF, CTNNB1 and MMP9 interacts with MMP 10, CXCL1.

Molecular Docking

Binding between Biological macromolecules (Protein, DNA, RNA etc) are mediated by some important intermolecular interactions. These intermolecular interactions between the biological macromolecules can be grouped into several non-covalent interactions, including Hydrogen bonding (interaction of hydrogen atoms with basic lone pair of electrons) hydrophobic interactions, electrostatic interactions, pie-cataion stacking, vanderWaals interactions, salt-bridge interactions etc.. Although, none of these non-covalent interactions are stronger than covalent interactions or ionic bond individually, but a number of these non-covalent interactions are known to act in concert to provide the stability to the interacting molecules. Out of these non-covalent interactions, Hydrogen bond holds the most importance and is responsible for holding together the DNA, Protein and other important biological macromolecules. Pie stacking is another type of important interactions that are known to occur between aromatic protein side chains of amino acids with other biological macromolecules.

We have conducted several molecular docking experiments by using HDOCK online server, in which we observed several poses by which the transcription factor STAT3 binds with the promoter region of MMP2, MMP7 and MMP9 by formation of a STAT3 homodimer. On the other hand, we have also conducted protein-protein molecular docking experiment between STAT3 homodimer and Importin-alpha3 as the later is known to mediate the entrance of STAT3 into the nucleus from the cytoplasm upon direct binding with STAT3 homodimer by recognizing specific amino acids of the dimer, known as Nuclear Localizing(NLS) Sequence(Lui L., et al.,2005).We have also observed that, Importin-alpha3 contains two NLS binding sites, known as Major and minor NLS binding sites respectively, as indicated by Uniprot (https://www.uniprot.org/uniprot/O00629, Date-09-06-2020, 7.23 p.m.).

These two NLS binding sites of Importin are known for their involvement in the recognition of NLS motifs of the proteins to be transported. The major NLS binding site of Importin is known to be positioned from Trp137-Arg229, while the minor NLS binding site is known to be positioned from Arg306-Asn394 as indicated by Uniprot. Besides, it is also indicated that several Trp and Asn residues must be present within the NLS binding motif of Importin-alpha3, which will mediate the key binding with the NLS site of the protein to be transported into the nucleus, as indicated by Uniprot (https://www.uniprot.org/uniprot/O00629, Date-09-06-2020, 7.23 p.m.). Previous researchers have also indicated the presence of several Armadilo Repeats (ARM repeats) within Importin-alpha, which may act as major NLS binding sites (K. Melen et al.,2003). We have analyzed the domains of Importin-alpha3 by ScanProsite (https://prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=1579002159588.scan.gz) and found three ARM repeats for Importin-alpha3, positioned in Gly114-Ala158, Asn157-Ile184 and Gly284-Val322. Besides, Uniprot indicated the presence of ten such ARM repeats along the entire sequence of Importin-alpha3. We have also observed the formation of several non-covalent interactions that are known to play the important role in binding of the biological macromolecules by UCSF Chimera (for protein-protein interaction) and Protein-ligand Interaction (PLIP) Profiler (for protein-DNA interactions).

In our first docking study, we have taken STAT3 dimer as receptor and Importin-alpha3 as receptor in HDOCK online server to observe the binding between these two protein, necessary for the transportation of STAT3 from cytoplasm into the nucleus. We have obtain 10 models for the interaction between these two selected protein , from which we have taken the model1, as it shows the higher negative binding energy for the interaction between the proteins considered.

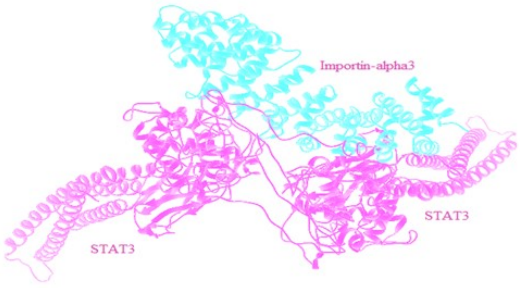
Several hydrogen bonds are observed to formed between Importin-alpha3 and STAT3 homodimer as indicated by Chimera. The Lys461 forms hydrogen bond with Asn401 of STAT3, Asn451 of Importin-alpha3, which is known to be

positioned within the ARM10 of Importin-alpha3 (as indicated by Uniprot) forms hydrogen bond with Val136 of STAT3, Asn411 of Importin-alpha3, positioned within the ARM9 of Importin-alpha3 (as indicated by Uniprot) forms hydrogen bond with Asn257 of STAT3, Gln385 of Importin-alpha3, positioned within the ARM8 of Importin-alpha3 (as indicated by Uniprot) and which is also predicted to be positioned within the NLS binding motif of Importin-alpha3 (as indicated by Uniprot) forms hydrogen bond with Asn400 of STAT3 and His252 of Importin-alpha3, which is known to be positioned within the ARM repeats of Importin-alpha3 (as predicted by ScanProsite and Uniprot) forms hydrogen bond with Lys615 of STAT3. The binding energy for this model is observed to be -269.93 with the RMSD value of 169.45.

Table 1: Amino acid residues of STAT3 homodimer, that interacts with Importin-alpha3.

Name of the Interaction	Hydrogen Bonds	
Stat3 homodimer to Importin-alpha3	Amino acid position	Residues
	136B	Val
	257B	Asn
	400B	Asn
	401B	Asn
	615D	Lys

Figure 6: Molecular interaction (Visualization through UCSF Chimera) between STAT3 homodimer and Importin-alpha3.

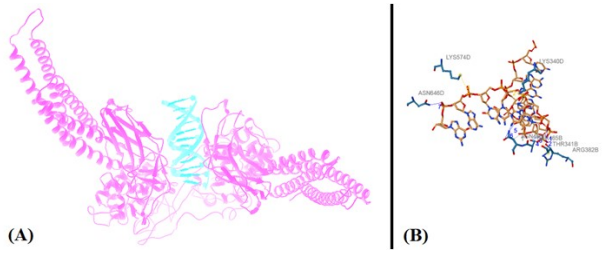


From the above analysis, it can be concluded that, the ARM8, ARM9 and ARM10 repeats of Importin-alpha3 mediate the key interactions with STAT3 homodimer for the nuclear transport of the later.

We have conducted the subsequent docking studies by taking STAT3 homodimer and the promoter region of MMP2, MMP7 and MMP9 individually as STAT3 homodimer is reported to act as potential transcription factor for MMP2 (T. Xie et al.,2004), MMP7 (A. Fukuda et al.,2011) and MMP9 (Z-H. Jia et al.,2017) and the expression of these MMPs have been upregulated by STAT3 homodimer in several types of cancer upon binding with the promoter regions of the former. We have constructed the DNA structure by taking the nucleotide located upstream to the coding regions of the DNA, which exactly matches with the character pattern of nucleotide located upstream to TSS within the promoters of MMP2, MP7 and MMP9, as given by GP-Miner software for the STAT3 binding sites. The DNA structures have been made individually for MMP2, MMP7 and MMP9 by using 3D-DART software and the protein-DNA molecular docking is done by using HDOCK software. Out of the several models given by HDOCK, We have chosen the model1 as it gives the higher binding energy for the protein-DNA interactions. We input the complex in PLIP analytical tool to visualize and extract the information regarding molecular interaction.

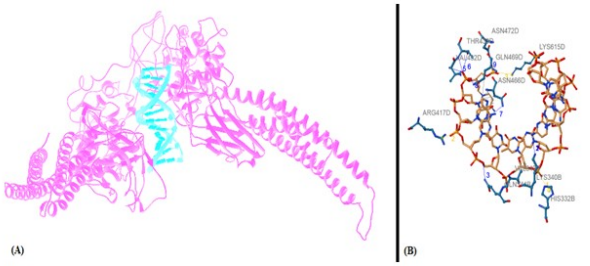
Firstly, we have docked STAT3 homodimer with MMP2 promoter DNA. Several hydrogen bonds are known to be formed between the promoter DNA with the residues of Thr341, Ser465, Asn466 of STAT3, which strictly falls under the DNA Binding Domain (DBD) of STAT3, with an additional hydrogen bond between the MMP2 promoter and Asn646 residue of STAT3, as indicated by PLIP. Besides, one hydrophobic interaction is also found between the MMP2 promoter DNA with Asn466 of STAT3 and salt bridge interactions are also observed between the MMP2 promoter DNA with the residues of Lys340, Arg382 of STAT3(as indicated by PLIP), which are present within the DBD of STAT3 and an additional salt-bridge interaction with the Lys574 of STAT3 outside the DBD, as indicated by the conserved domain analysis of STAT3(Figure: 7). The binding energy is predicted to be -278.38 with the RMSD value of 138.82 for this interaction for model1.

Figure 7: Molecular interaction (Visualization through UCSF Chimera on left) and interaction sites of STAT3(visualization through PLIP on right) homodimer with MMP2 promoter DNA.



Next, we have docked STAT3 homodimer with the promoter DNA of MMP7. Formations of several hydrogen bonds are observed between the promoter DNA of MMP7 with the residues of Lys340, Gln344, Val432, Thr433, Asn466 and Asn472, as indicated by PLIP. Also, hydrophobic interactions are also observed between the promoter DNA of MMP7 with the residues of Val343, Asn466 and Gln469, and salt-bridge interactions between the promoter DNA with the residues His332, Arg417 and Lys615 of STAT3, as indicated by PLIP (Figure:8). All of the interacting residues of STAT3 mentioned above, falls under the DBD of STAT3, except for Lys615, as reflected through the conserved domain analysis of STAT3.The binding energy for this interaction in model1 is predicted to be -273.23, which is slightly lesser than that of the binding between STAT3 with MMP2 promoter and the RMSD value is 148.81. This clearly indicates that STAT3 has a higher binding affinity for MMP2 promoter region than that of the MMP7 promoter.

Figure 8: Molecular interaction (Visualization through UCSF Chimera on left) and interaction sites of STAT3 (visualization through PLIP on right) homodimer with promoter DNA of MMP7.



After that, we have docked STAT3 dimer with MMP9 DNA promoter. We have observed the formation of several hydrogen bonds between the promoter DNA with the residues of Lys340, Val432, Thr433, Ser465, Asn466 and Gln469 of STAT3, as indicated by PLIP. Besides, hydrophobic interaction is also observed between the Promoter DNA of MMP9 with Asn466 of STAT3 and salt-bridge interactions are also observed between the promoter DNA with the residues His332, Arg382, Arg417 and Lys615 of STAT3, as indicated by PLIP(Figure: 9). All of the interacting residues of STAT3 mentioned above strictly falls under the DBD of STAT3, except for Lys615, as reflected through our conserved domain analysis of STAT3.The binding energy for this interaction is predicted to be -277.64, which is greater than that of the interaction between the MMP7 promoter and STAT3, with the RMSD value of 145.63 for model1.

Figure 9: Molecular interaction (Visualization through UCSF Chimera on left) and interaction sites of STAT3 homodimer (visualization through PLIP on right) with MMP9 promoter DNA.

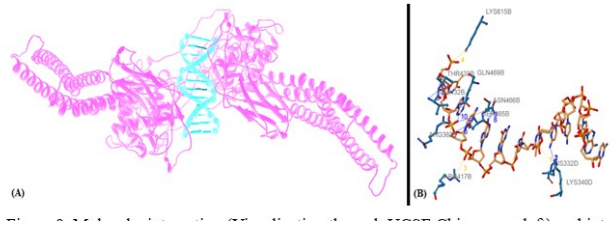


Table 2: Amino acid residues of STAT3 homodimer, that interacts with the promoter DNA of MMP2, MMP7 and MMP9. The brackets indicate the number of times the amino acid appears in specific interactions.

Name of the Interaction And Interacting Molecules	Hydrophobic Interaction		Hydrogen Bonds		Salt-Bridge	
	Amino Acid Position	Residues	Amino Acid Position	Residues	Amino Acid Positions	Residues
STAT3	466B	Asn	341B(2)	Thr	340D(2)	Lys
Homodimer to					465B(2)	Ser
MMP2					466B(2)	Asn
					646D	Asn
STAT3	343B	Val			340B	Lys
Homodimer to					340B	Lys
MMP7	466D	Asn			344B	Gln
	469D	Gln			432D	Val
					433D(2)	Thr
					466D	Asn
					469D	Gln
					472D	Asn
STAT3	466B	Asn			340D(2)	Lys
Homodimer to					432B	Val
MMP9					433B(2)	Thr
					465B(2)	Ser
					466B(2)	Asn
					469B	Gln
STAT3	466B	Asn	341B(2)	Thr	340D(2)	Lys
Homodimer to					465B(2)	Ser
MMP2					466B(2)	Asn
					646D	Asn
STAT3	343B	Val			340B	Lys
Homodimer to					340B	Lys
MMP7	466D	Asn			344B	Gln
	469D	Gln			432D	Val
					433D(2)	Thr
					466D	Asn
					469D	Gln
					472D	Asn
STAT3	466B	Asn			340D(2)	Lys
Homodimer to					432B	Val
MMP9					433B(2)	Thr
					465B(2)	Ser
					466B(2)	Asn
					469B	Gln

Name of the Interaction And Interacting Molecules	Hydrophobic Interaction		Hydrogen Bonds		Salt-Bridge	
	Amino Acid Position	Residues	Amino Acid Position	Residues	Amino Acid Positions	Residues
STAT3	466B	Asn	341B(2)	Thr	340D(2)	Lys
Homodimer to					465B(2)	Ser
MMP2					466B(2)	Asn
					646D	Asn
STAT3	343B	Val			340B	Lys
Homodimer to					340B	Lys
MMP7	466D	Asn			344B	Gln
	469D	Gln			432D	Val
					433D(2)	Thr
					466D	Asn
					469D	Gln
					472D	Asn
STAT3	466B	Asn			340D(2)	Lys
Homodimer to					432B	Val
MMP9					433B(2)	Thr
					465B(2)	Ser
					466B(2)	Asn
					469B	Gln

We have then conducted another molecular docking study by taking four selected drugs, namely, Galiellactone, STX-0119, Ins 3-548 and MMPP , all of which is reported to block the DNA Binding Domain of STAT3, necessary for interaction with the MMP promoters (Bharadwaj et al.,2016). The aim of doing so is to observe whether these drugs lower the binding affinity of STAT3 towards MMP2, MMP7 and MMP9 promoters or not, upon binding with STAT3 prior to it's binding with the promoters of MMP genes.

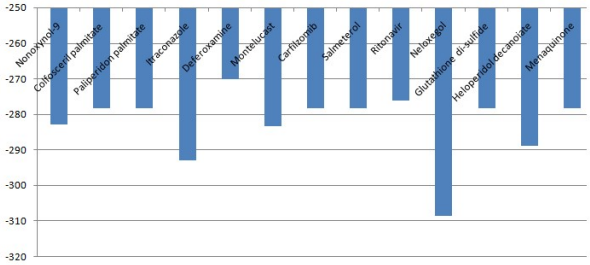
Table 3: Amino acid residues of STAT3 homodimer, that interacts with the promoter DNA of MMP2, MMP7 and MMP9 upon the introduction of drug Deferoxamione. The brackets indicate the number of times the amino acid appears in specific interactions.

Name of interaction	Hydrophobic interaction		Hydrogen bond		Salt bridge interaction	
	Position	Amino acid	Position	Amino acid	Position	Amino acid
STAT3 to Deferoxamine	NA		342D	Gly	414B	Arg
			344D(2)	Gln		

to MMP2			387D	Leu		
			389D	Thr		
			412D	Thr		
			415B(2)	Glu		
			417B(2)	Arg		
			465B	Ser		
			469B	Gln		
STAT3 to Deferoxamine to MMP7	412D	Thr	341B	Thr	382B	Arg
			387D	Leu		
			412D	Thr	414B	Arg
			415B(2)	Glu		
			465B	Ser	417B	Arg
			469B	Gln		
STAT3 to Deferoxamine to MMP9	NA		344D	Gln	414B	Arg
			382B	Arg	417B	Arg
			387D	Leu		
			389D	Thr		
			412D	Thr		
			415B(3)	Glu		
			417B(2)	Arg		
			465B(2)	Ser		
			469B	Gln		

We have observed that all of these selected drugs, except for Galiellactone binds to the SH2 domain by , the help of hydrophobic interaction at the sites of Phe621, Gln638, Leu670 and Lys679 of STAT3, instead of binding with the DNA Binding Domain. The drug InS3-548 forms hydrophobic interactions with the residues of Glu638, Gln644 and Phe710 present within the SH2 domain of STAT3 and no hydrogen bonds or Salt-bridge interactions are found with STAT3 for this drug, as indicated by PLIP. Also, the drug MMPP forms hydrophobic interactions with the residues of Phe606, Phe621, Trp623, Tyr657 and Pro669 of STAT3 present within the SH2 domain of STAT3. Besides, this drug forms hydrogen bonds with the residues of STAT3 at the positions of Leu607, Thr622, Leu670 and salt-bridge interaction with the residue of Lys679, which also falls within the SH2 domain of STAT3, as indicated by PLIP. Furthermore, the drug STX-0119 forms hydrophobic interactions with the residues of Asn646, Glu690 and Phe710, as indicated by PLIP. Also the hydrogen bonds are observed between STX-0119 with the residues Gln644, Asn646 and Asn647, as indicated by PLIP. All of these above mentioned interacting residues of STAT3 are present within it's SH2 domain. On the other hand, Galieallactone binds do the DBD of STAT3 by forming hydrogen bonds with the residues of Val323, Gln326, His457 and hydrophobic interactions with the residues of Leu459 and Pro487 of the DBD of STAT3, with additional hydrophobic interactions with the residues of Lys244, falls outside the DBD of STAT3, as indicated by STAT3. It is observed that none of these drugs are able to lower the binding affinity of STAT3 to the promoter of MMP2 and formation of all the identical interactions still persist between STAT3 and MMP2 promoter, even when each of these drugs have been introduced for binding with STAT3, prior to the promoter binding of the latter. Moreover, the binding affinity of STAT3 toward the promoter of MMP2 observed to increase upon the binding of STX-0119 with STAT3 (Binding energy increased from -278.38 to -287.37) prior to the binding of STAT3 to MMP2 promoter DNA.

Figure 10: Binding energy between STAT3 and MMP2 promoter upon the inclusion of drugs selected from virtual screening.



Lastly, we have gone through the virtual screening analysis based on the structure of STAT3 transcription factor by the help of e-LEA3D web server (accessible freely from https://chemoinfo.ipmc.cnrs.fr/TMP/tmp.904248/VISU/visu.html) to identify the novel and diverse molecular scaffolds that may act as the potential inhibitor of STAT3 mediated MMP transcriptional activation. After that, we have performed molecular docking by taking each drug from the list provided by the e-LEA3D server individually. We have observed that, none of these small molecules except Deferoxamine have the potential of lowering the binding affinity of STAT3 towards MMP2, MMP7 and MMP9 promoters, even when they gets bind to STAT3 prior to the promoter binding of STAT3. Besides, it has been also observed that Deferoxamine has much higher affinity towards the promoters of MMP2 (Binding energy: -218.67 with RMSD value: 7.83), MMP7 (Binding energy: -225.61 with RMSD value: 7.70) and MMP9 (Binding energy: -222.86 with RMSD value: 7.84) than that of the STAT3 (Binding energy: -129.21 with RMSD value: 123.06). This indicates that Deferoxamine will more likely bind with the MMP promoters instead of binding with STAT3. Furthermore, it has been also observed that, the binding affinity of STAT3 complexed with this drug towards the promoters of MMP2 (Binding energy: -270.24, with RMSD value: 136.26), MMP7 (Binding energy: -277.63 with RMSD value: 134.73) and MMP9 (Binding energy: -273.50 with RMSD value: 136.72) does not lower to a significant level and numerous interactions between amino acid residues of STAT3

DBD and the promoters of MMP2, MMP7 and MMP9 are still observed. On the other hand, when Deferoxamine binds to the respective MMP promoters prior to the binding of STAT3 to the same MMP promoters effects the binding affinity between STAT3 and MMP promoters and the number of essential non-covalent interactions are also reduced. For example, the binding affinity of STAT3 towards the promoter of MMP2 lowers when the latter is complexed with the drug(Binding energy: -255.38 with RMSD value: 159.83). no hydrophobic interactions are now observed between the amino acid residues of STAT3 and MMP2 promoter. Only one hydrogen bond is observed between the amino acid residues of STAT3 DBD and MMP2 promoter at the site of His332 and two additional hydrogen bonds are also observed between STAT3 and MMP2 promoter at the sites of Arg688 and Glu690, both of which fall outside the DNA Binding Domain of STAT3(as indicated by PLIP). Also, the Salt-bridge interactions between STAT3 and MMP2 promoters are observed at the sites of Lys573, Lys574, Lys707 and Lys709, all of which falls outside the DBD of STAT3(as indicated by PLIP). Only one salt bridge interaction between His332 of STAT3 with MMP2 promoter is observed to occur within the DBD, as indicated by PLIP. Besides, upon introducing Deferoxamine to MMP7 prior to the binding of STAT3 to the MMP7 promoter, the affinity of STAT3 towards MMP7 promoter also gets lowered(Binding energy: -261.95 with RMSD value: 150.32). only one hydrophobic interaction is observed between STAT3 and MMP7 promoter at the site of Tyr686, which falls outside the DBD of STAT3, as indicated by PLIP. Besides, most of the hydrogen bonds are observed between STAT3 and MMP7 promoter at the sites of Leu577, Ala578, Gln644, Lys685 and Arg688 of STAT3, all of which falls outside the DBD of STAT3(as indicated by PLIP). Only one hydrogen bond occurs between STAT3 and MMP7 promoter at the sites of Asn466, that falls under the DBD of STAT3, as indicated by PLIP. Furthermore, only one salt-bridge interaction is observed between STAT3 and MMP7 promoter at the site of Lys517, which also falls outside the DBD of STAT3.lastly, Deferoxamine also lowers the binding affinity of STAT3 toward MMP9 promoter, when it is complexed with MMP9 promoter prior to the binding of STAT3 with the same promoter(Binding energy: -244.02 with RMSD value: 138.33). the hydrophobic interactions are observed between STAT3 and MMP9 promoter at the sites of Leu387 and Asn390, both of which falls within the DBD of STAT3 and hydrogen bonds between STAT3 and MMP9 promoter at the sites of Gly388, Thr389, Asn390 and Thr391, allof which falls within the DBD of STAT3. But the number of total non-covalent interactions gets lowered in comparison to the interaction between only STAT3 to MMP9 promoter, which provide strong support for the alteration of binding affinity of STAT3 towards MMP9 promoter. The binding of the promoters of MMP2, MMP7 and MMP9 outside the DBD of STAT3 is due to the binding of Deferoxamine to respective MMP promoters, prior to STAT3 binding and it might results in the alteration of STAT3 mediated transcriptional upregulation of MMP2, MMP7 and MMP9 genes.

From all of the above analysis, it may be concluded that STAT3 upon forming homodimer acts as a transcription factor for MMP2, MMP7 and MMP9 by directly binding to the promoter region of MMP2, MMP7 and MMP9 gene, and thereby upregulating the expression of the MMPs in several cancers. Although, the binding affinity among these MMPs for STAT3 transcription factor varies and STAT3 binds to the MMP2 promoter with highest affinity among these MMPs, indicating that STAT3 upregulate the transcription of MMP2 more than that of the MMP7 or MMP9. So, it is very clear that STAT3 must have possess DNA binding domain for mediating the interaction with the promoters. Additionally, STAT3 must also have contain NLS site for binding with Importin-alpha3, which facilitate the entrance of STAT3 into the nucleus. Therefore, by designing specific drug molecules against the promoter regions of the MMPs mentioned above, will result in the blockade of the promoter region and therefore, will result in the lowering of the expression level of the MMPs mentioned above by preventing the binding of STAT3 to the promoter region of these MMPs. As a result cancer associated metastasis risk can also be mitigated. The virtual screening and docking analysis by taking Deferoxamine led us to the conclusion that Deferoxamine has a higher affinity for MMP promoters and upon binding of this drug to the promoters of MMP2, MMP7 and MMP9 may hamper the interaction between STAT3 and the MMP promoters, typically by inhibiting the interactions of the promoters with the amino acid residues of STAT3 DNA Binding Domain. Therefore, the drug molecules will have to be designed in accordance with the chemical structure of Deferoxamine to inhibit the MMP associated metastatic spread in cancers.

Differential Expressions Analysis

Analysis of the expression levels is helpful in providing the functionalities of the genes in specific organs and under specific conditions. The aim of this analysis is to observe whether the increased activation of STAT3 in specific organs coincides with the increased concentration of the considered MMPs or not, and also to what extent. We have gone through the list of Cancers mediated due to the constitutive activation of STAT3 from Disgenet database (Accessible freely from https://www.disgenet.org/rdf, Date- 09-06-2020, Time- 7.55 p.m.) We have taken the already reported expression levels of STAT3, MMP2, MMP7 and MMP9 in specific organs obtained from three different experiments of Wang et al.,2019; 32 Uhlen's lab and Hallstrom et al.,2014 provided in the Ensembl search engine (accessible freely from https://asia.ensembl.org/index.html,Date- 09-06-2020, Time- 7.30 p.m.).

Upon the analysis of the differential expression levels, we are able to observe that elevated expression of STAT3 (elevated expression in organs like, Brain, Lymph node, Liver, Stomach, Kidney, Lung, Pancreas, Prostate, Esophagus , Colon and Rectum) correlate with the elevated expression of MMP2 in Esophagus only. Whereas, expression of MMP2 is also elevated in Gallbladder, as indicated by Wang et al.,2019.

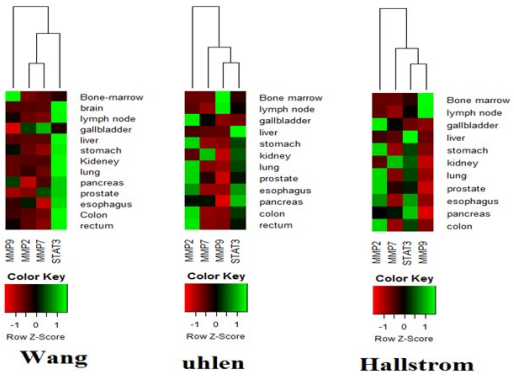
On the other hand, the elevated expression level of STAT3 exhibits similarity with the elevated expression level of MMP7 in Prostate, as indicated by Wang et al.,2019. Besides, MMP7

expression is also elevated in Gallbladder, which correlates with the elevated expression level of MMP2.

The expression level of STAT3 is elevated in Pancreas also, which exhibits similar fold of expression with the elevated levels of MMP9, as indicated by Wang et al.,2019.

When we analyzed the expression levels of STAT3, MMP2, MMP7 and MMP9 in the same organs considered earlier by taking another experiment (32 Uhlen's lab), we observed that, the elevated STAT3 expression exhibit similar folds of expression with the elevated levels of MMP2 genes in various organs, including Stomach, Lung, Prostate, Esophagus, Colon and Rectum. While, that of the elevated expression levels of STAT3 exhibit similar folds of expression with the elevated expression levels of MMP7 in Kidney only. Furthermore, we have also observed that the levels of MMP9 in organs like Bone marrow and Lymph node, although the expression levels of STAT3 seemed to be in normal range for these two organs.

Figure 11: Expression levels of STAT3, MMP2, MMP7 and MMP9 in the form of heat map by taking three different experiments (Wang et al.,2019; 32 Uhlen's lab; Hallstrom et al., 2014), The heat map is prepared using R-Program.



Upon analyzing the expression levels of STAT3, MMP2, MMP7 and MMP9 in different organs by taking another experiment (Hallstrom et al.,2014), we have observed that the elevated expression of STAT3(elevated in Liver, Stomach, Kidney, Lung, Prostate, Esophagus, Pancreas and Colon) exhibit similar folds of expression with the elevated expression levels of MMP2 in various organs , including Lung, Prostate, Esophagus and Colon. Besides, the elevated levels of STAT3 in Kidney and Pancreas also corresponds with the elevated expressions of MMP7 in those organs. While, the expression levels of MMP9 is elevated in Bone marrow and Lymph node independently from that of the expression of STAT3 (Figure:11).

From all of the above analysis, it can be concluded that, the elevated levels of STAT3 more likely corresponds with the elevated expressions of MMP2 than that of the MMP7 and MMP9, as indicated in the above experiments, possibly by binding to the promoter region of the MMP2 gene by the DNA binding domain of STAT3 with more affinity than that of the MMP7 and MMP9. As a result, tumerigenesis and constitutive activation of the transcription factor STAT3 in these organs might cause metastatic spread by cleaving certain component proteins of extra-cellular matrix with the help of MMP2 in these organs. Besides, it can also be concluded that the levels of MMP9 is much more in Bone marrow and Lymph node, in which tumerigenesis and constitutive activation of STAT3 might result in the metastatic spread by cleaving certain proteins of the extra-cellular matrix with the help of MMP9. As a result, specific drug molecules have to be designed against individual MMPs depending upon the organ specificity.

CONCLUSION

The Interactive protein network analysis by String database takes us to some important protein networks involved in the activation of metastatic cascade. We would like to conclude that the STAT3 homodimer binds with the Importin-alpha3 subunit and makes it's entry into the nucleus from the cytoplasm. Upon being entered into the nucleus, it directly binds with the promoter regions of MMP2, MMP7 and MMP9 individually through it's DNA binding domain by recognizing specific consensus nucleotide sequences of the respective gene promoters and results in the transcriptional activation of the genes mentioned above, although the affinity of STAT3 transcription factor towards the promoter regions of MMP2, MMP7 and MMP9 varies and the highest affinity is shown by STAT3 transcription factor towards the promoter of MMP2, as indicated by the Binding energy obtained from H-DOCK server and also upon analyzing the differential expression analysis. We can further conclude that, the domain structures of the protein products of MMP2 and MMP9 genes are almost similar and consisted of identical conserved domains, which exhibit significance dissimilarities with that of the protein product of MMP7 gene. This may be the reason for the difference in acting against different substrates.

Therefore, by designing of specific molecules, which will eventually bind with the promoter regions of MMP2, MMP7 and MMP9, specifically to the binding sites of STAT3 transcription factor may provide the fruitful inhibition of metastatic spread caused by these endopeptidases in cancers caused due to the constitutive activation of STAT3. Also by blocking the active sites (obtained from NCBI conserved Domain search tool and Scanprosite analysis) of the MMP2, MMP7 and MMP9 protein may provide another rational of mitigating the cancer associated metastatic spread. Furthermore, the designing of these drug molecules also have to be specific for different organs, in which different MMPs regulate the cancer associated metastatic spread. Furthermore, another alternative approach could have been taken for designing of drug molecules, which will hamper the interaction between STAT3 and Importin, necessary for the nuclear transportation of the former, prior to the binding with the promoters of Matrix Metalloproteinase genes. We have also found that the drug Deferoxamine has the potential of hampering the STAT3 mediated transcriptional upregulation of MMP2, MMP7 and MMP9 genes, when this drug binds to STAT3 prior to the binding of STAT3 to the promoters of MMP2, MMP7 and MMP9 genes. Deferoxamine is found to interact with amino acid residues positioned within the DNA Binding Domain of STAT3, eventually lowers the affinity of STAT3 towards the promoter of MMP genes. Therefore, special attention should be

given to this drug molecule during the further processing of small molecules, necessary for the inhibition of cancer associated metastatic spread by MMP2, MMP7 and MMP9.

REFERENCES

1. Abdulghani J, Gu L, Dagvadorj A. Stat3 promotes metastatic progression of prostate cancer. Am J Pathol. 2008;172(6):1717-1728.
2. Chang L, Wang S, Wu I. Impaired dendritic cell maturation and IL-10 production following H. pylori stimulation in gastric cancer patients. Appl Microbiol Biotechnol. 2012;96:211-220.
3. Devarajan P, Johnston JJ, Ginsberg SS, Van Wart HE, Berliner N. Structure and expression of neutrophil gelatinase cDNA. Identity with type IV collagenase from HT1080 cells. J Biol Chem. 1992;267(35): 25228-25232.
4. Don-Doncow N, Marginean F, Coleman I. Expression of STAT3 in Prostate Cancer Metastases. Eur Urol. 2017;71(3):313-316.
5. Edman K, Furber M, Hemsley P. The discovery of MMP7 inhibitors exploiting a novel selectivity trigger. ChemMedChem. 2011;6(5): 769-773.
6. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389-3402.
7. Ferguson SD, Srinivasan VM, Heimberger AB. The role of STAT3 in tumor-mediated immune suppression. J Neurooncol. 2015;123(3): 385-394.
8. Fiorentino DF, Bond MW, Mosmann TR. Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. J Exp Med. 1989;170(6):2081-2095.
9. Fitch W. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. Systematic Zoology. 1971;20(4):406-416.
10. Fukuda A, Wang SC, Morris JP, Folias AE, Liou A, Kim GE, Akira S, Boucher KM, Firpo MA, Mulvihill SJ, Hebrok M, et al. Stat3 and MMP7 contribute to pancreatic ductal adenocarcinoma initiation and progression. Cancer Cell. 2011;19(4):441-455.