List of Figures:

List of Tables

sss

***In silico* Identification of novel ligand enhances Myelopoiesis to prevent infection caused after Chemotherapy.**

Jyoti Parmar , M.Tech Bioinformatics

Delhi Technological University Delhi

ABSTRACT

Myelopoiesis is required in order to prevent infection during bone marrow transplantation and to build enhanced immune systems in patients. In this study our aim is to enhance HSC myelogenesis and minimise erythropoiesis, this can be done with the help of regulation of various transcription factor. PU.1 is a member of the ETS family of Transcription Factors and it plays an important role in the adaptive immune system. Transcription factor c-Jun binds with PU.1 β3/β4 and leads to Myelopoiesis. However, GATA-1 competes for the binding site with c-Jun to bind with PU.1, its binding leading to Erythropoiesis. However, the molecular mechanism underlying this interaction is not known and a lot of inconsistencies exist in literature. Therefore, the PU.1 binding motifs of GATA-1 and c-Jun should be docked with PU.1. This can reveal the mechanism of PU.1 binding with GATA-1 and c-Jun which will help in identification of most hotspot residues which take part in interaction. It has been reported in papers that the GATA-1 mediated repression of PU.1 can be abolished if the physical interaction between the two proteins is modified. Therefore, we need to design a PU.1 β3/β4 region Mimetic Peptide Library to find a suitable antagonist that would specifically bind to GATA-1, inhibiting it to bind to PU.1 and allowing c-Jun to bind with PU.1, and lead to Myelopoiesis.

Key words: HSC (Hematopoietic Stem Cell) , PU.1 ,GATA-1, c-JUN

2.0 INTRODUCTION:

Stem Cell has the remarkable potential to develop in many different cell types in the body during early life and growth. They can divide without limit to replenish other cells and serve as a sort of internal repair system in many tissues.

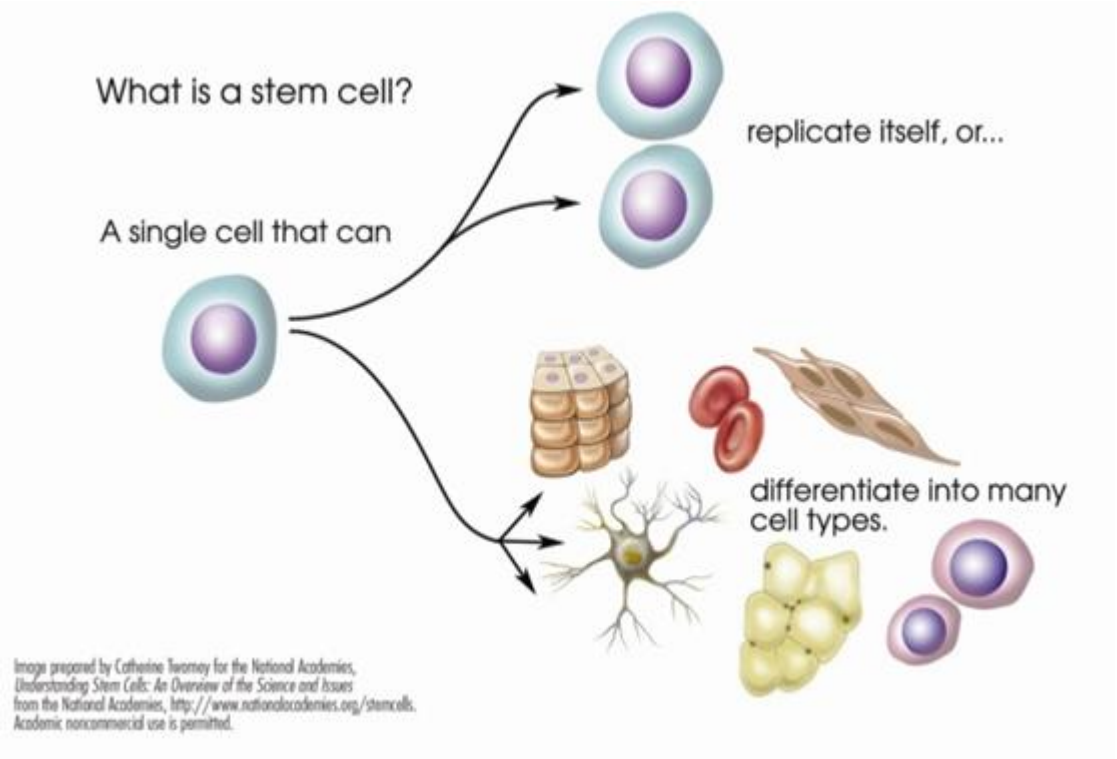


Figure 1: When a stem cell divides, each new cell has the potential either to remain a stem cell or become another type of cell with a more specialized function, such as a muscle cell, a red blood cell, or a brain cell.

Stem cells are distinguished from other cell types by two important characteristics i.e.Self renewing and differentiation.

The broad objective of this research is to study the development of adaptive and innate immune system which is coordinated by a complex regulatory network of transcription factors in a cell intrinsic manner to control the development of myeloid lineage through PU.1, c-Jun and GATA-1

The development of adaptive immune system is regulated by PU.1, c-Jun and GATA1, and since it is activated during cell proliferation, differentiation and different lineage commitment of HSC.

The specific aims of this study were:

- To find potential PU.1 mutation that would have strongest binding toward c-Jun and GATA1.
- To predict the binding affinity of GATA1 and c-Jun with PU.1 and hotspot residues play important role in interactions.
- To generate a potential lead molecule through extensive mutagenesis study that is specifically designed to abrogate PU.1-GATA1 interaction and allow PU.1, c-Jun interaction.

3.0 REVIEW OF LITERATURE

Hematopoiesis is a highly orchestrated multi-step process which is regulated by various important transcription factors that involves the proliferation, differentiation and maturation of a very small population of self renewing, pluripotent hematopietic stem cells into various specialized and distinct blood type (Burda et.al 2010). PU.1 and GATA-1 are often described as 'master regulators' largely based on their hierarchical position during hematopoiesis as well as their ability to induce cellular trans-differentiation when expressed transgenically.

PU.1 acts as a weak transactivator (Tenen, D. G e.t.al (1997)), suggesting that it requires co-activators to achieve activation function through physical interactions. The DNA-binding domain of PU.1 has been found to interact physically with many proteins, including myeloid regulators such as AML-1B and CyEBPa (Petrovick, et.al.(1998)).It has been demonstrated that the ß3/ß4 region (amino acids 243–254) of PU.1 in the Ets domain interacts with a number of proteins (N.W.-a., Y. Koyama, G.B., S. Tetradis, J. Tsukada, Y.-T. Ro, D.G.T., and P.E.A.,), including c-Jun, which acts as a critical co-activator of PU.1 transactivation of myeloid promoters (Behre e.t.al. (1999)). Repression of PU.1 function is a critical step for erythroid cell differentiation, because overexpression of PU.1 in erythroblast blocks erythroid cell differentiation (Moreau-Gachelin e.t.al. (1996)).

Work from laboratories of A.Skoultchi, D.Tenen and T.Garf showed that PU.1 and GATA-1 proteins physically interact with each other within hematopoietic cells. It was in subsequent studies that elucidated how this process is accomplished and, more importantly, the molecular consequences of this protein-protein interaction during normal and pathological hematopoiesis (Rekhtman et.al 1999; Zhang et.al 1999; Nerlov et.al 2000).

The initial findings that GATA-1 and PU.1 proteins are components of a minimal gene regulatory network regulating the erythroid cell fate choice came from their unnatural co-expression in erythroleukemic cells (MEL cells) of mice that harbor the friend virus integration within the URE of the Spi-1 gene. This viral integration caused the aberrant and constitutive expression of PU.1 throughout the erythroid compartment resulting in the erythroleukemic pathology (Moreau-Gachelin F et.al). These findings were subsequently corroborated by experiments with PU.1-overexpressing transgenic mice that develop erythroleukemias at a high rate (Moreau-Gachelin F and Wendling et.al 1996).

Initially it was thought that PU.1 prevents binding of GATA-1 to DNA as shown by electrophoretic mobility shift assays (Zhang P and Zhang X et.al 2000) However Chromatin immunoprecipitation studies showed the co-occupancy of both PU.1 and GATA-1 proteins to DNA (Rekhtman et.al 2003) and that this co-occupancy was dependent on the presence of both a GATA-1 cis-binding site and the Ets domain on the PU.1 protein. Interestingly, the N-terminal portion of PU.1 does not affect the occupancy of PU.1 and GATA-1 on its DNA-binding site (Rekhtman et.al 2003) yet is required to bind with the retinoblastoma protein (pRb) for repression of GATA-1 (Rekhtman N et.al 2003). Data of several laboratories

collectively show that PU.1 inhibits GATA-1 on DNA (Stopka T, Amanatullah DF et.al 2005) although effects of PU.1-binding GATA-1 outside DNA cannot be excluded. Several key studies have shown that dysregulation of either PU.1 or GATA-1 activity can contribute to leukemogenesis (Rosenbauer F, Wagner K, et.al 2004 ; Wechsler J, Greene M et.al 2002; Li Z, Godinho FJ, Klusmann JH et.al 2005; Shimizu R, Kobayashi Eet.al 2009).

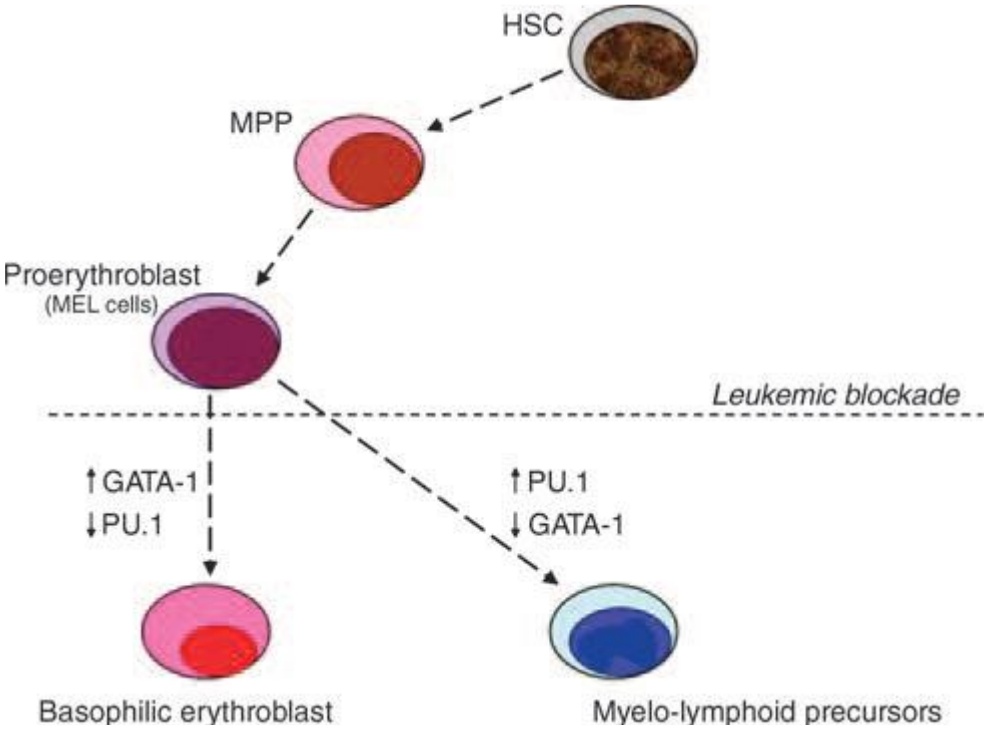


Figure 2: Scheme of erythroleukemia and its differentiation. Upon differentiation from multi-potential progenitor (MPP) the erythroid development may be blocked at the level of proerythroblasts as exemplified by MEL cell model. Manipulation of either PU.1 or GATA- 1 levels (indicated by arrows) can relieve this block and restore differentiation potential toward erythroid (basophilic erythroblast) or mixed population of myelo-lymphoid cells with nuclear abnormalities

Several Studies collectively show that manipulating the relative expression of transcription factor levels (for example, inhibition of GATA-1 or activation of PU.1 (or vice versa), in erythroleukemias may represent an efficient tool for inducing, and rescuing, leukemic blasts cells into a differentiated cell state (Papetti M, Skoultchi et.al 2007; Burda P, Curik N, Kokavec J, Basova P, Mikulenkova D, Skoultchi AI et al.2009). Regulation of both PU.1 and GATA-1 expression levels are crucial for the correct cell fate determination of multipotential progenitors and that inappropriate regulation of these primary determinants can cause a block of cell differentiation and expansion of a clonal population of leukemic cells.

GATA-1 represses PU.1 function, in which GATA-1 interaction with PU.1 blocks the ability of the PU.1 coactivator c-Jun to bind to the same small region that GATA-1 binds but does not inhibit PU.1 DNA binding (Zhang P, Behre G, Pan J, et al. 1999)

Yeast two-hybrid screen and *in vitro* glutathione *S*-transferase (GST) pull-down and *in vivo* immunoprecipitation assays that the DNA-binding domain of PU.1 interacts with the zinc finger region of GATA-2 and the same highly conserved region of GATA-1 (Lee, M. E.,

Temizer, D. H., Clifford, J. A. & Quertermous, T.(1991)). These interactions result in an inhibition of PU.1 transactivation function as a result of blocking c-Jun binding to PU.1, as both c-Jun and GATA proteins interact with the same region of PU.1, the ß3/ß4 region.

Table PU.1 blocks erythroid cell differentiation

Cell line	With/without induction	Benzidine-positive cells (%)
$ZnSO_4$		
K562/pC18 (vector)*	−	35.0
	+	32.5
K562/mPU.1 no. 3*	−	46.4
	+	6.7
K562/mPU.1 no. 7*	−	35.0
	+	8.9
β-Estradiol induction		
G1ER/M†	+	52.0
	−	0.5
G1ER/mPU.1 no. 1†	+	6.8
	−	0.5
G1ER/mPU.1 no. 2†	+	5.6
	−	0.5

Figure 3: PU.1 blocks K562 cell differentiation, as assessed by benzidine-positive cells as an indicator. PU.1 blocks G1ER cell differentiation. (Table adopted from Pu Zhang, Xiaobo Zhang, Atsushi Iwama, Channing Yu, Kent A. Smith, Beatrice U. Mueller, Salaija Narravula, Bruce E. Torbett, Stuart H. Orkin, and Daniel G. Tenen PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding)

Previously, it was demonstrated that GATA-1 interacts with the ß3/ß4 region of the PU.1 ETS domain and inhibits PU.1 transactivation by blocking the PU.1 coactivator c-Jun from binding to this same region. However, the mechanism of how PU.1 inhibits GATA-1 function was unclear. Zhang and Zhang et.al demonstrated that the mechanism of PU.1 inhibition of GATA-1 function differs from that of GATA-1 inhibition of PU.1 function and involves a different region of the PU.1 protein.

PU.1 and GATA-1 inhibit each other's function through protein–protein interactions by different mechanisms. The left panel shows that GATA-1 inhibits PU.1 transactivation of PU.1 target genes by blocking the binding of the PU.1 co activator c-Jun to the β3/β4 region of PU.1. The right panel shows that the N-terminal region of PU.1 blocks GATA-1 function by binding to the GATA-1 C-finger and inhibiting GATA-1 binding to DNA. Figure 4 and figure 5 can describe this very effectively.

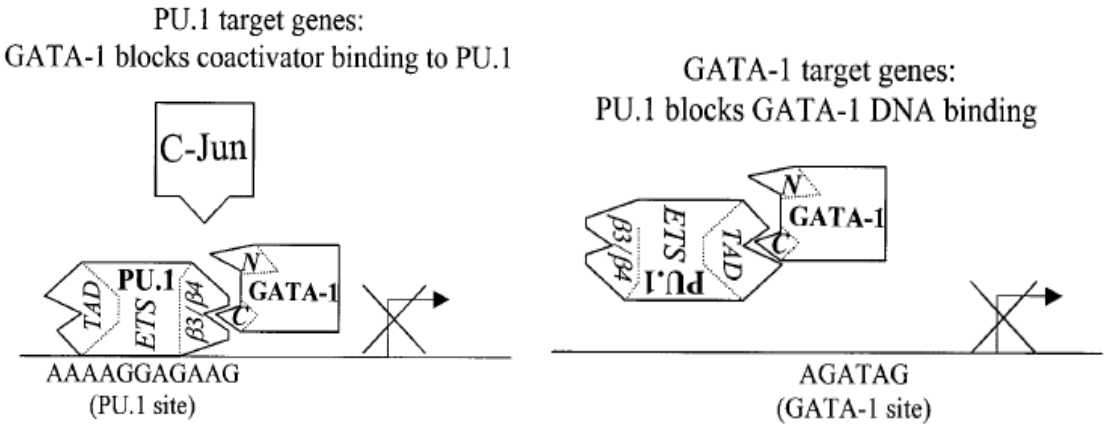


Figure 4: PU.1 and GATA-1 inhibit each other's function through protein–protein interactions by different mechanisms. The top panel shows that GATA-1 inhibits PU.1 transactivation of PU.1 target genes (such as CD11b46) by blocking the binding of the PU.1 coactivator c-Jun to the b3/b4 region of PU.1. The bottom panel shows that the N-terminal region of PU.1 blocks GATA-1 function by binding to the GATA-1 C-finger and inhibiting GATA-1 binding to DNA.

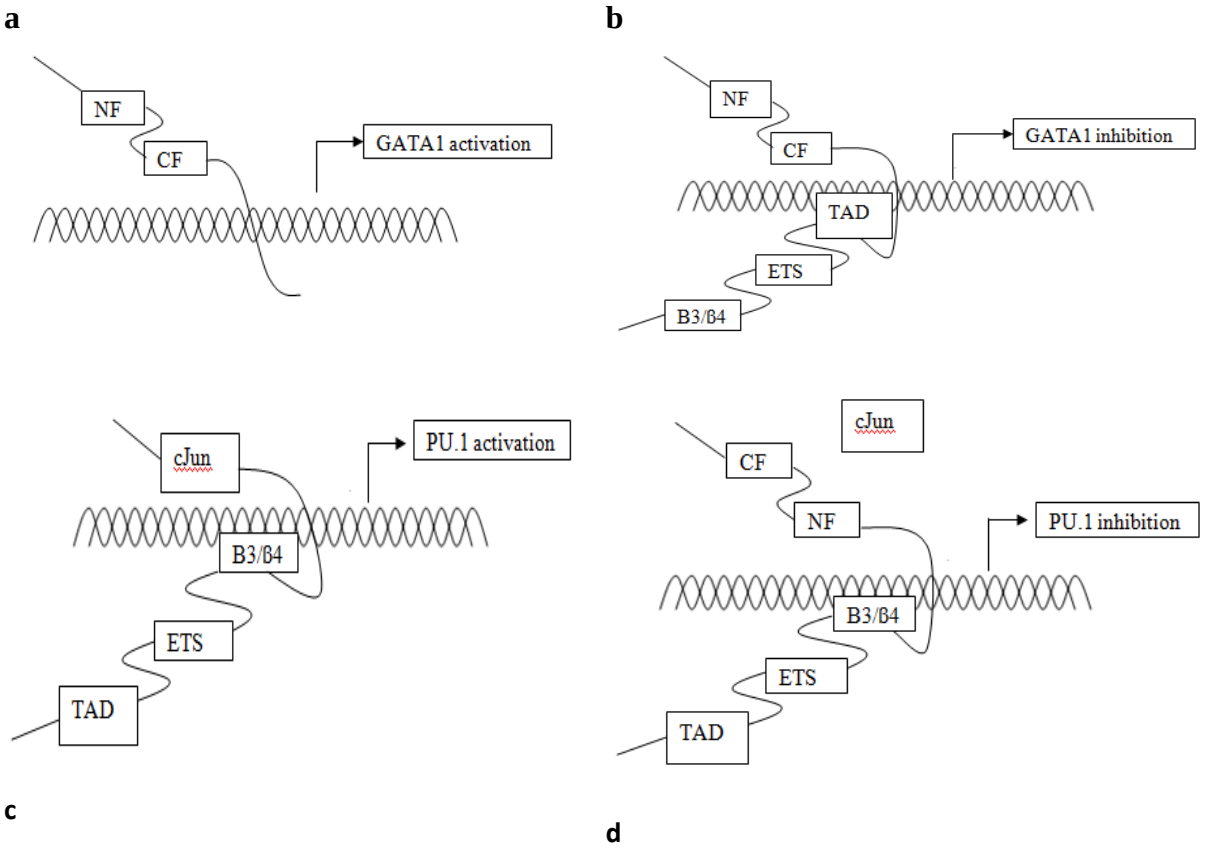


Figure 5: Inhibition of GATA1 via PU.1 or PU.1 via GATA1 binding.

3.01 Hematopoietic Stem Cells: Molecular Diversification and developmental interrelationships

Blood cells are continuously produced throughout our lifetime from rare pluripotent bone marrow stem cells, called hematopoietic stem cells (HSCs). HSCs are endowed with two characteristics: They give rise to additional HSCs through self-renewal and also undergo differentiation to progenitor cells that become variably committed to different hematopoietic lineages (Weissman 2000).

It is estimated that in normal humans there are approximately 50 million HSCs, some of which can generate up to 1013 mature blood cells over a normal lifespan. Proliferation and differentiation are not necessarily strictly coupled, and in the most primitive HSCs this renders a capacity for self-renewal, the third cardinal property of all stem cell types. Self-regeneration is critical for HSCs because they are constantly subjected to physiological stresses that stimulate their recruitment along maturational pathways that ultimately result in their depletion, for example under conditions of hypoxia to increase red blood cell numbers or during infections to amplify granulocytes and macrophages. Self renewal, at least at the population level, thus ensures that sufficient numbers of stem cells are available to meet the demands of hematopoiesis over a normal adult lifespan.

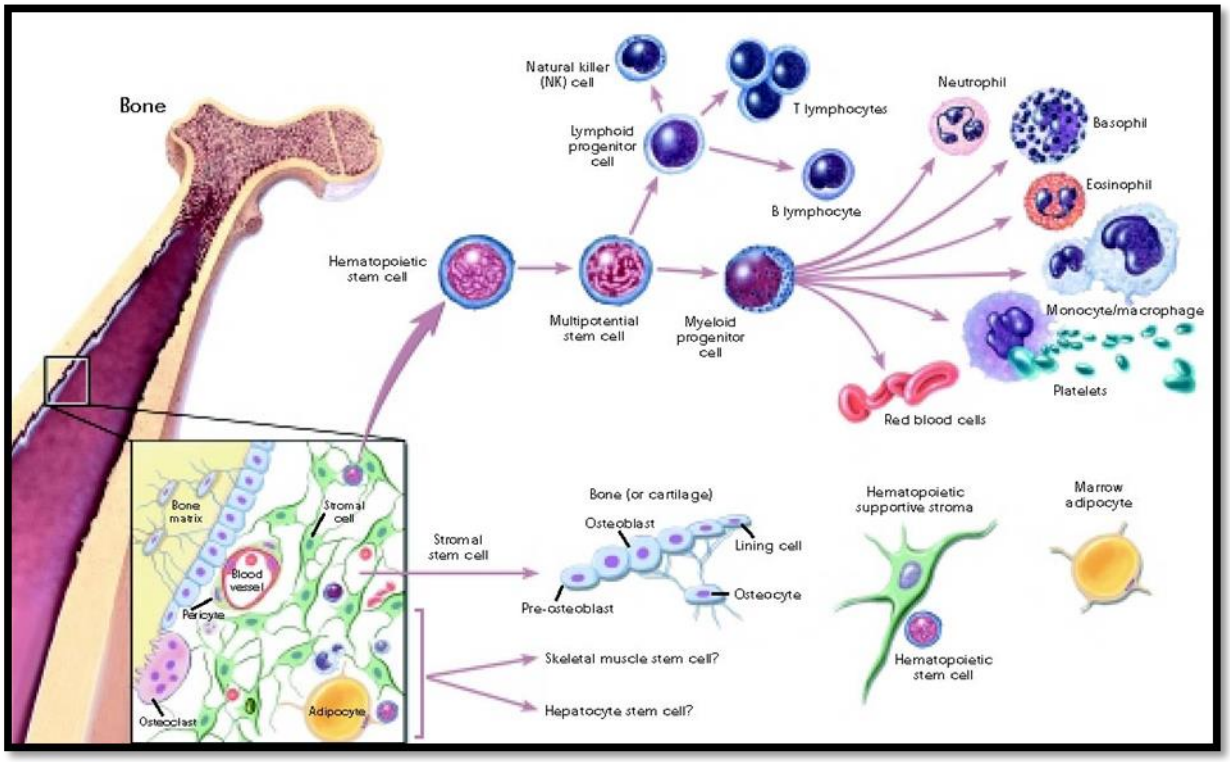


Figure 6: Bone marrow contains two kinds of stem cells. One population, called hematopoietic stem cells, forms all the type of blood cells in the body i.e. red blood cells, B lymphocytes, T lymphocytes, natural killer cells, neutrophils, basophils, eosinophils, monocytes, and macrophages. A second population, called bone marrow stromal stem cells (also called mesenchymal stem cells or skeletal stem cells) forms give rise to a variety of cell types: bone cells (osteocytes), cartilage cells (chondrocytes), fat cells (adipocytes), and other kinds of connective tissue cells such as those in tendons.

Less than 0.1% of the hematopoietic cells in BM are pluripotent stem cells capable of long-term proliferation and self renewal. The term lineage specification is used to define pluripotent, to characterize the sequence of decision steps that finally leads to lineage commitment, manifested by the acquisition of particular lineage specific features. The term differentiation refers to the actual process of acquisition of these lineage specific features and does also include the loss of self-renewal ability. Lineage potential describes the general ability of a stem or progenitor cell to produce a certain number of different types of mature functional cells. In contrast, the lineage contribution of a particular cell describes the cell types actually produced in a particular differentiation process, and may therefore comprise only a part of the cell's lineage potential.

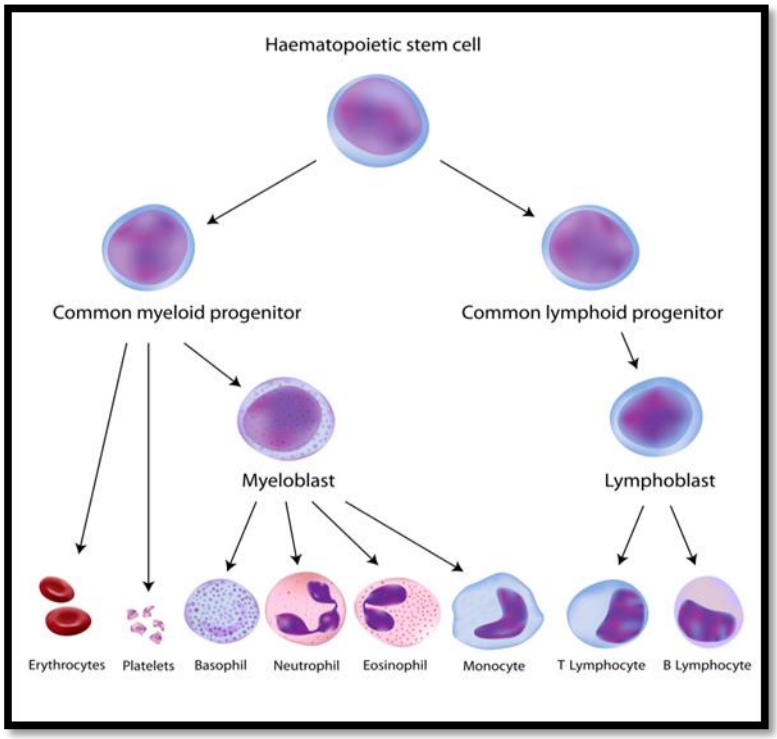


Figure 7: Hematopoietic stem cells, forms all the type of blood cells in the body i.e. red blood cells, B lymphocytes, T lymphocytes, natural killer cells, neutrophils, basophils, eosinophils, monocytes, and macrophages.

3.01.2 Site of Hematopoiesis:

Age of Animal	Site of hematopoiesis
Embryo	Yolk Sac then liver
3rd to 7th Month	Spleen
4th and 5th Month	Marrow cavity-esp. granulocytes and platelets
7th Month	Marrow cavity - erythrocytes
Birth	Mostly bone marrow; spleen and liver when needed
Birth to maturity	Number of active sites in bone marrow decreases but retain ability for hematopoiesis
Adult	Bone marrow of skull , ribs , sternum, vernal column , pelvis, proximal ends of femurs

3.01.3 Hematopoietic Hierarchy:

The hematopoietic system is commonly perceived as a hierarchy with the stem cell at the origin giving rise to lineage restricted progenitors and finally to terminally differentiated end cells. There are various cell surface markers involved in the hematopoietic hierarchy. The figure shows the comparison between cell surface marker of human and mouse.

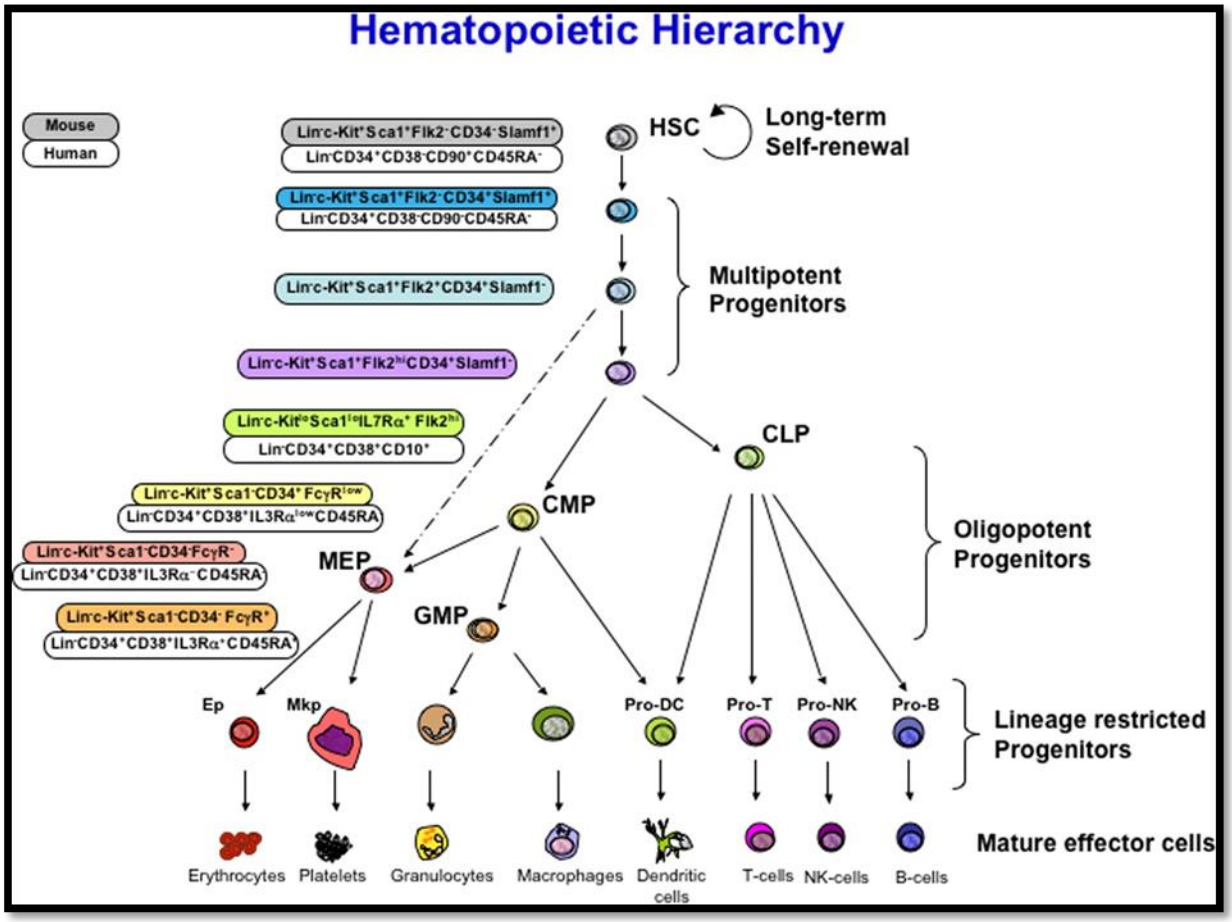


Figure 8: Comparison between human and mouse cell surface marker in the hematopoietic stem cell hierarchy. HSC (Hematopoietic Stem Cell) , CMP (Common Myeloid Progenitors) , CLP (Common Lymphoid progenitors) EMP (Myeloid Erythrocyte Progenitors), GMP (Granulocytes Myeloid progenitors).

3.02 The role of transcription factors and their interplay in the development of different hematopoietic lineages

3.04.1 Function of myeloid transcription factors:

Proliferation and survival of hematopoietic progenitor cells is regulated by growth factor receptors and their signaling intermediates. However, a network of transcription factors regulates the expression of a cell type-specific pattern of genes and directs the cells down the path from stem cells and early precursors to fully differentiated cells of the various lymphoid and myeloid lineages.

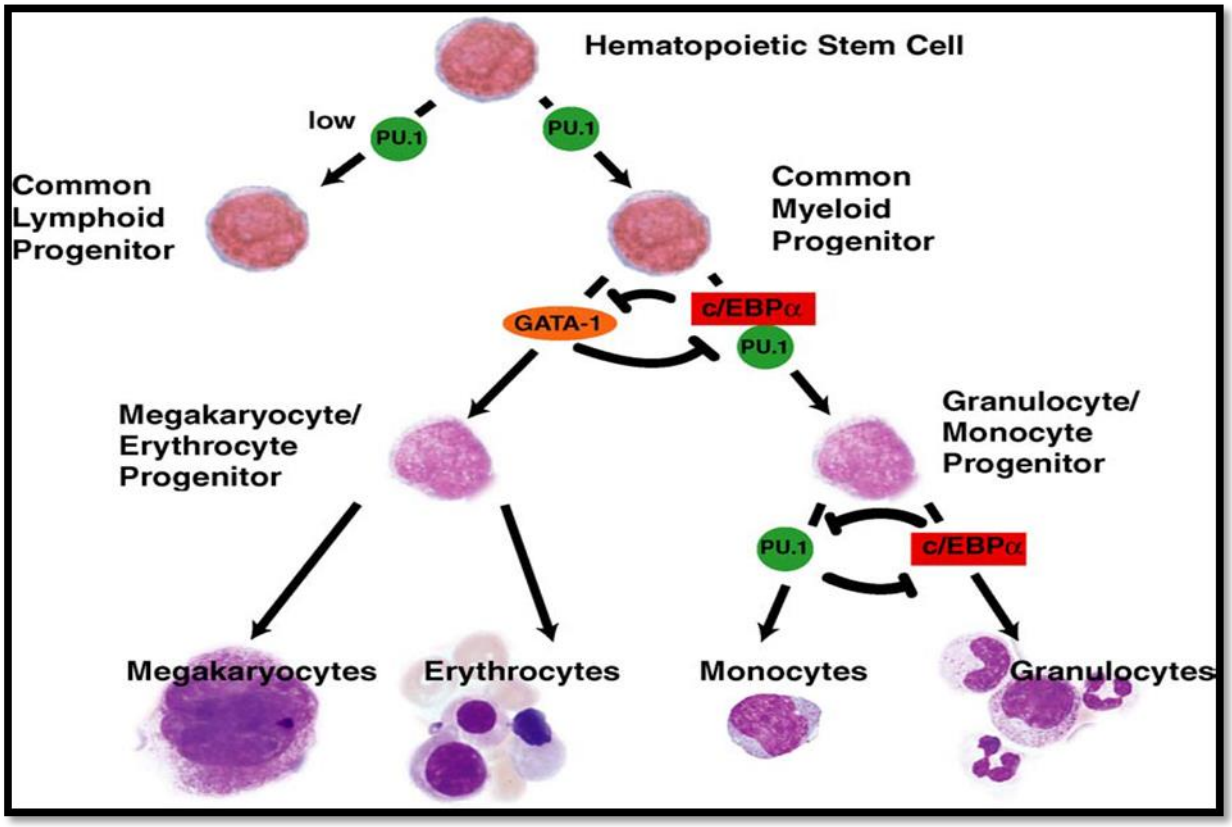


Figure 9: Network of transcription factors PU.1, GATA-1, c/EBP α regulates the expression of a cell type-specific pattern of genes and directs the cells down the path from stem cells and early precursors to fully differentiated cells of the various lymphoid and myeloid lineages.

4.0 Materials and Methodology

4.0.1 Target Protein Identification

Proteins or transcription factor which play important role in the determination of Erythropoieisis versus Myelopoieisis fate of HSCs were found from previous research papers. Master Regulator of hematopoeisis is PU.1, GATA1 and c-JUN. PU.1ia a member of Ets family transcription factor involved in the process of lineage specification of HSC as it supports development of myeloid and lymphoid cells. Zinc finger factor GATA1 required for the differentiation and maturation of erythroid/megakaryocytic cells. GATA binding protein 1 (globin transcription factor 1); transcriptional activator which serves as a general switch factor for erythroid development. It binds to DNA sites with the consensus sequence [AT]GATA[AG] within regulatory regions of globin genes and of other genes expressed in erythroid cells (413 aa).

Direct interaction can take place between GATA-1 and PU.1 and that the regions responsible for this interaction are the interaction are: C-terminal zinc finger of GATA-1 (CF) and the Ets domain of PU.1. Expression of PU.1 and GATA1 appears to be mutually exclusive, suggesting an antagonistic effect of these two transcription factors. GATA1 inhibits PU.1 by preventing it from interacting with its cofactor c-Jun, while PU.1 inhibits GATA1 by preventing its binding to DNA.Inhibition of PU.1 activity by GATA-1 appears to involve the transcription factor c-Jun. c-Jun a critical co-activator of PU.1 trans-activation of myeloid promoters. Due to this reason we select PU.1, GATA1 and c-Jun.

4.0.1.1 Screening of Scrambled peptide against c-Jun and GATA-1

We have generated 499 scrambled peptide sequences extracted from β3/β4 region on PU.1.Virtual screening has been defined as the "automatically evaluating very large libraries of small molecules" using computer software. We Screen the scrambled peptides firstly against cJun (PDB ID: 1JNM) via Pyrx and select only those peptides which have binding energy lesser than wild type. Pyrx is a virtual Screening Software for Computational Drug Discovery that can be used to screen libraries of compounds against potential drug targets. Further, peptides having lowest binding energy than wild type against GATA1 (PDB ID: 2GAT) were selected.

4.0.2 Proteins Three dimensional structure

The 3D structures of PU.1, GATA1 and c-Jun were obtained from PDB database and their PDB IDs are 1PUE, 2GAT and 1JNM respectively. All the three PDBs contain the DNA chains, and PU.1 and c-Jun have the dimer form.

4.0.3 DNA PROTEIN INTERACTING RESIDUES:

Table 1: DNA Protein interacting residues

S.No.	Protein (PDB Id)	Organism	Type of structure	Resolution	References
1	PU.1 (1PUE)	Mus Musculus	X-Ray diffraction	2.10 Å	http://www.rcsb.org/pdb/files/1PUE.pdb
2	GATA-1 (2GAT)	Chicken	NMR	NA	http://www.rcsb.org/pdb/files/2GAT.pdb
3	c-JUN (1JNM)	Homo Sapien	X-Ray diffraction	2.20 Å	http://www.rcsb.org/pdb/files/1JNM.pdb

4.0.3.1 Transcription Factor PU.1,

Total residues 171-259 i.e. 88 residues
41% helical (3 helices; 37 residues)
17% beta Sheet (4 strands; 16 residues)

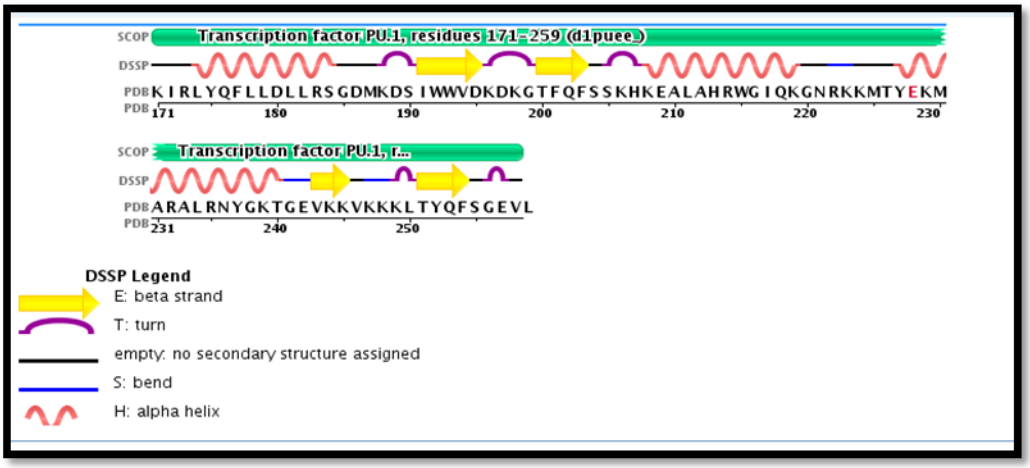


Figure 50: Secondary Structure of PU.1 obtained from PDB

4.0.3.2 Transcription Factor GATA-1:

The zinc-finger protein GATA-1 is normally expressed at a low level in multipotential progenitors and at higher levels in erythroid precursors, megakaryocytes, mast cells, and eosinophils (Orkin 1992). Knock-out studies in mice demonstrate that GATA-1 is essential for the maturation of both erythroid and megakaryocyte precursors, but not for their initial generation (Weiss et al. 1994; Fujiwara et al. 1996; Shivadasani et al. 1997).

D2gata_Erythroid transcription Factor GATA-1
Total Residues : 66
19% helical (2 helices; 13 residues)
9% beta sheet(2 strands;6 residues)

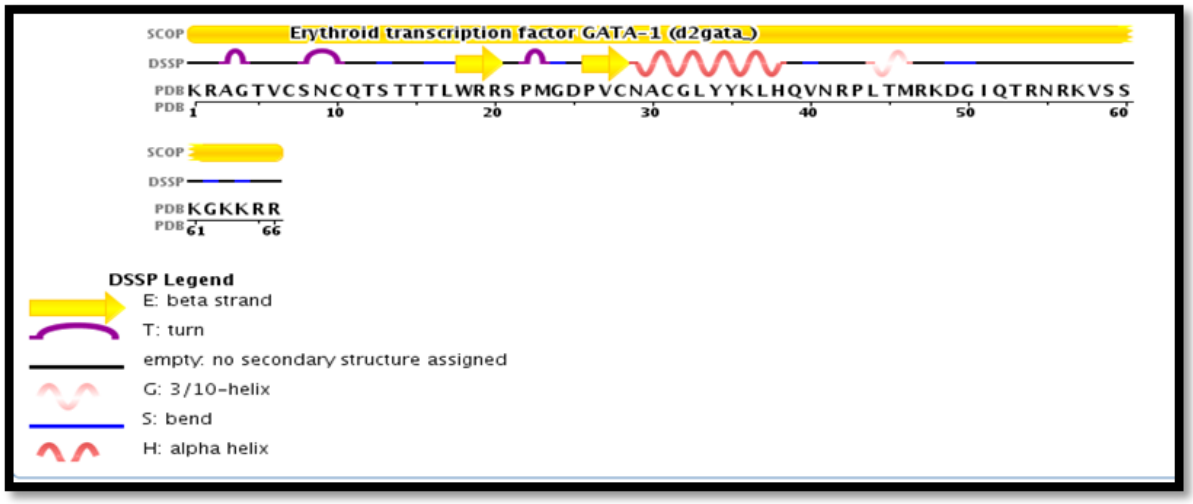


Figure 11: Secondary Structure of GATA-1 obtained from PDB

4.0.3.2 Transcription Factor c-JUN

Amino acid sequences "254 to 285(Length= 32) "which are reported to make interaction with PU.1

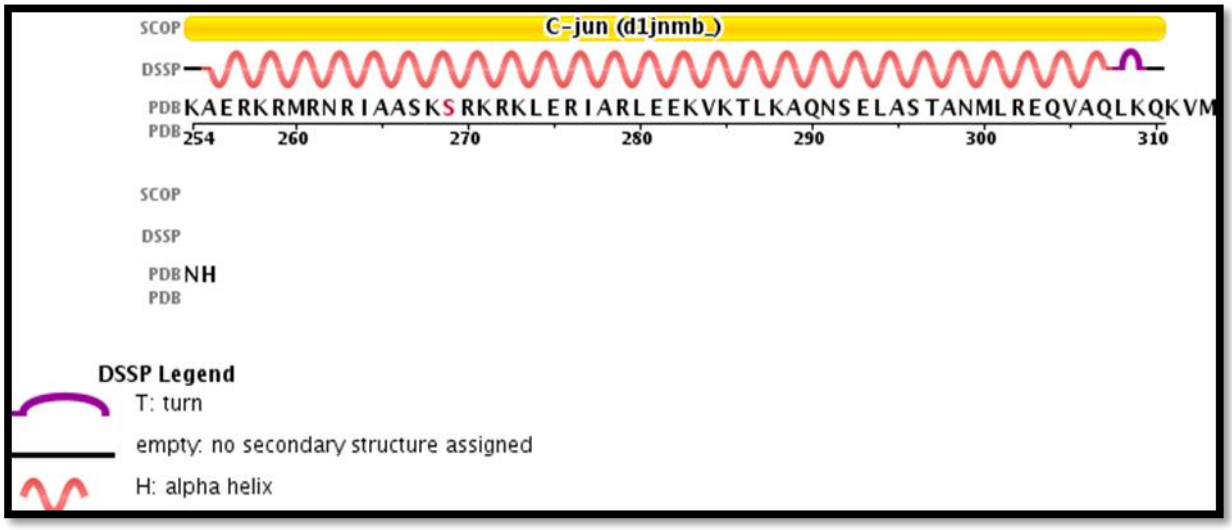


Figure 12: Secondary Structure of c-Jun obtained from PDB

4.0.4 Families of Transcription Factors

Table 2: List of Transcription Factors involved in HSC and their families, Super class and Class.

S.No.	Transcription Factor	Family	Super Class	Class	References
1	PU.1	ETS domain	Helix-turn-Helix	Tryptophan clusters	Stegmaier et.al
2	GATA-1	GATA-Factors	Zinc-coordinating DNA-binding domains	Diverse Cys4 zinc fingers	Matys et.al
3	c-JUN	AP-1(like) components ; includes (c-Fos/c-Jun)	Basic Domains	Leucine zipper factors (bZIP	Transfac database
4	C/EBP-α	C/EBP like factors	Basic Domains	Leucine zipper factors (bZIP)	Matys et.al

4.0.5 ETS Domain:

Transcription Factor PU.1 is a protein that in Human is encoded by the *SP1* gene. This gene encodes an ETS-domain transcription Factor that activates gene expression during myeloid and B-lymphoid cell development.

All ETS Family members are identified through a highly conserved DNA binding domain , the ETS domain which is a winged helis turn helix structure that binds to DNA sites with a central **GGA(A/T)** DNA sequence and this domain is also involved in protein- protein interactions.

Table 3: Sub family and their family members of Transcription factor

Sub Family	Mammalian Family members	Invertebrate orthologs	References
ELF	ELF1,ELF2(NERF) , ELF4(MEF)		
ELG	GABP-α	ELG	Gutierrez-Hartman A, Duval DL, Bradford AP (2007). "ETS transcription factors in endocrine systems". *Trends Endocrinol Metab* **18** (4): 150–8
ERG	ERG , FLI1 , FEV		
ERF	ERF(PE2), ETV3(PE1)		
ESE	ELF3(ESE1/ESX), ELF(ESE2), ESE3(EHF)		
ETS	ETS1,ETS2	POINTED	
PDEF	SPDEF(PDEF/PSE)		
PEA3	ETV4(PEA3/E1AF), ETV5(ERM), ETV1 (ER81)		
ER71	ETV2(ER71)		
SPI	SPI1(**PU.1**), SPIB, SPIC		
TCF	ELK1, ELK4(SAP1), ELK3(NET/SAP2)	LIN	
TEL	ETV6(TEL), ETV7(TEL2)	YAN	

4.0.6 Functional Domain of PU.1:

PU.1 domain assumes a tight globular structure with three α-helices and a four β-stranded antiparallel β-sheet enclosing a hydrophobic core.

The PU.1 domain contacts DNA from three sites: the recognition helix (α3) interacts with the GGAA core sequence in the major groove, while contacts with the phosphate backbone on either side of this site are made in the minor groove by two loops. Therefore, the PU.1 ETS domain binds DNA by a loop helix- loop motif. One loop is formed between β-strands 3 and 4 (a "wing") and the other is a loop in the position of the turn in the HTH motif (α2-turn-α3).

4.0.6.1PU.1 contains several functional domains.

The Ets domain corresponds to the DNA binding domain, and recognizes sequences harboring the core GGAA motif. This domain is also involved in protein–protein interactions between PU.1 and other factors such as GATA-1, Runx1 or C/EBP. At the structural level, the Ets domain exhibits a winged helix-turn-helix motif with a loop-helix-loop architecture, and interacts with a 10-base-pair region on duplex DNA (Kodandapani et al., 1996).

Other domains include an N-terminal acidic domain and a glutamine-rich domain, involved in transcriptional activation, and a PEST region, involved in protein–protein interactions, particularly with the IRF4/8 transcription factors.

(A) Primary structure and important functional domains, as well as the positions of serines that can be phosphorylated. Involvements of particular domains in interactions with other proteins or with DNA are indicated in the diagram

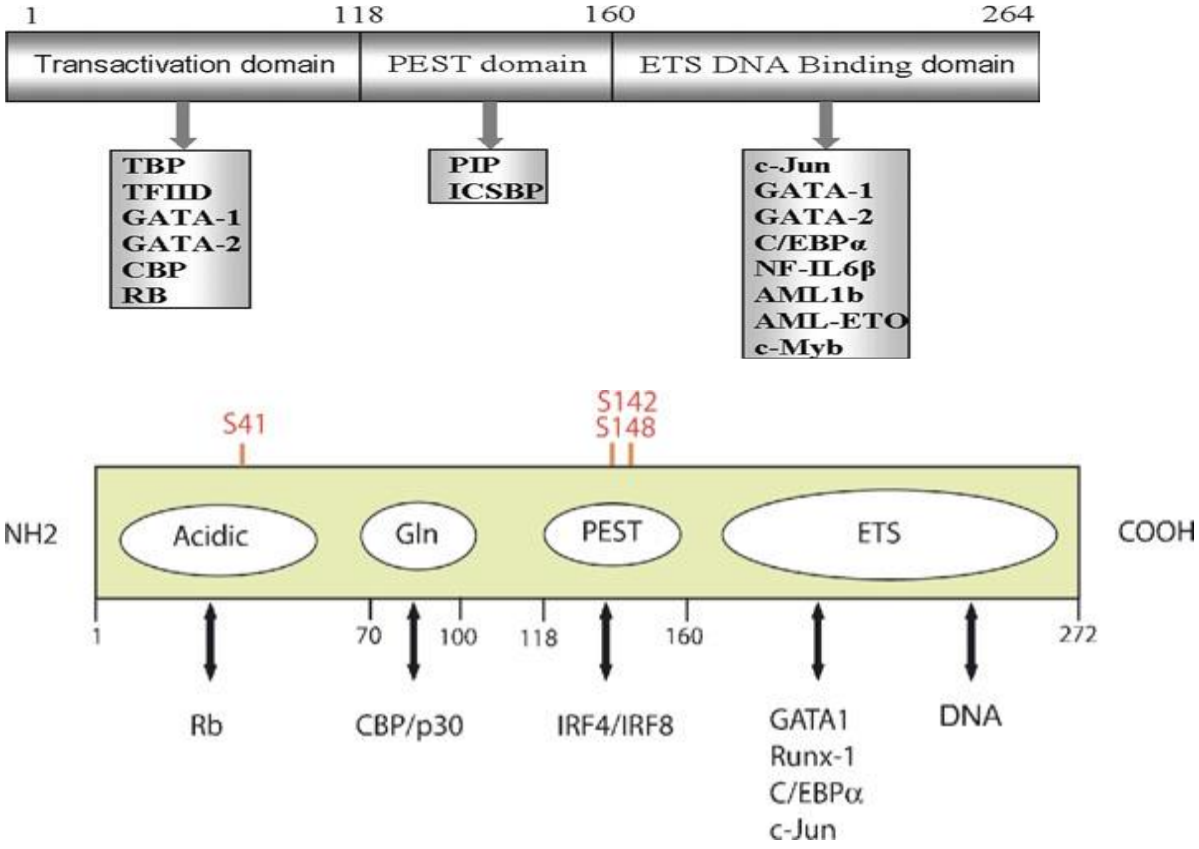


Figure 13: Functional domain of PU.1. PU.1 has major three functional domain 1) Tran activation domain 2) PEST Domain and 3) ETS DNA Binding Domain.

(B) An expanded view of the ETS DNA-binding domain encoded by exon 5 is shown. The approximate location of the α-helices and anti-parallel β-sheets within the ETS domain are indicated.

The C terminus ETS domain codes for a DNA-binding domain that recognizes the sequence 5-GGAA3.

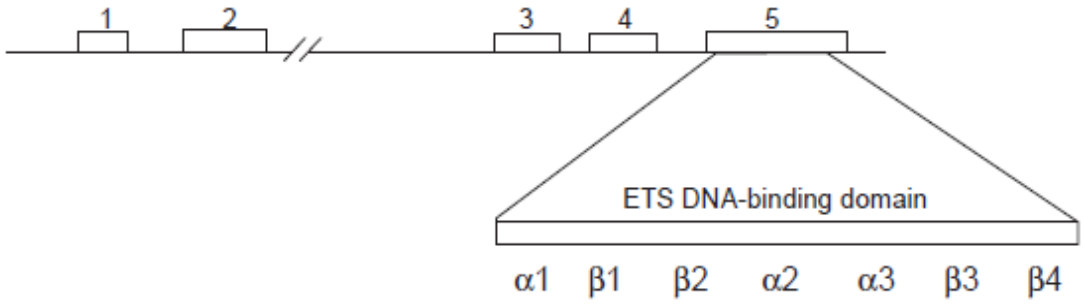


Figure 16: An expanded view of the ETS DNA-binding domain encoded by EXON 5

4.0.6.2 Genomic Organization of the PU.1 Gene.

The amino acid sequence of the ETS DNA-binding domain is displayed along with the location of the three α-helices and the four anti-parallel β-sheets.

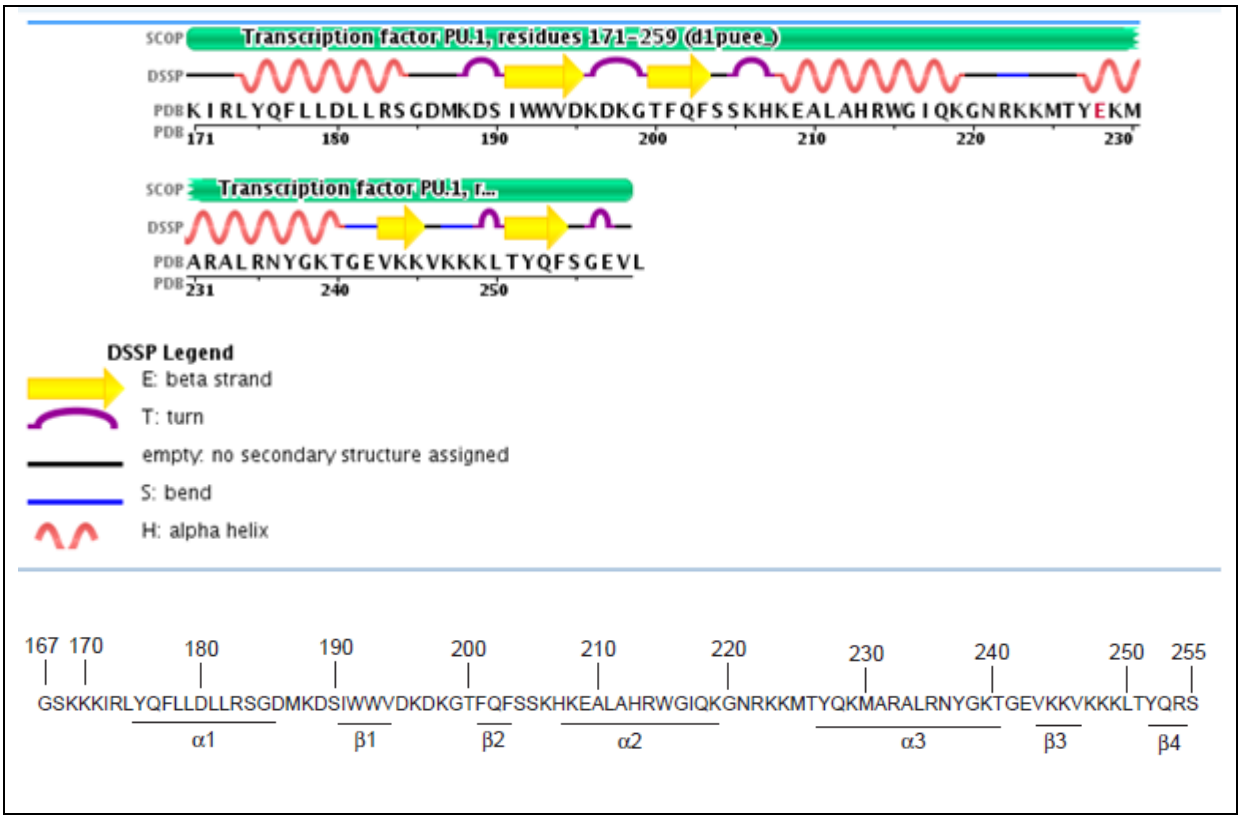


Figure 17: Genomic organization of PU.1 gene and amino acid sequence of ETS DNA-binding domain

4.0.7 PU.1 residues showing Interaction with GATA-1

High interaction	Reference
Leu172, Tyr173, His205, Glu207, Lys227, Met228, and Lys245, together with the side chain indole NH of Trp213 and the side chain amide group of Asn219	Liew CW, Rand KD, Simpson RJ, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba M, Nerlov C, Poulsen FM, Mackay JP.

4.0.8 FUNCTIONAL DOMAINS OF GATA1

Three functional domains have been identified within the GATA1 protein 1) N-terminal activation domain; 2) N-terminal zinc finger (N-finger); 3) C-terminal zinc finger (C-finger).

The N-finger contributes to the stabilization and specificity of DNA binding. The N-finger mediates the formation of complexes with cofactors. The C-finger is essential for GATA1 function, since it is responsible for the recognition of the GATA consensus sequence and consequent binding to DNA**.**

4.0.9 PU.1 blocks GATA-1 binding to DNA

GATA-1 interacts with the β3/β4 region of the PU.1 ETS domain and inhibits PU.1 transactivation by blocking the PU.1 co-activator c-Jun from binding to this same region.

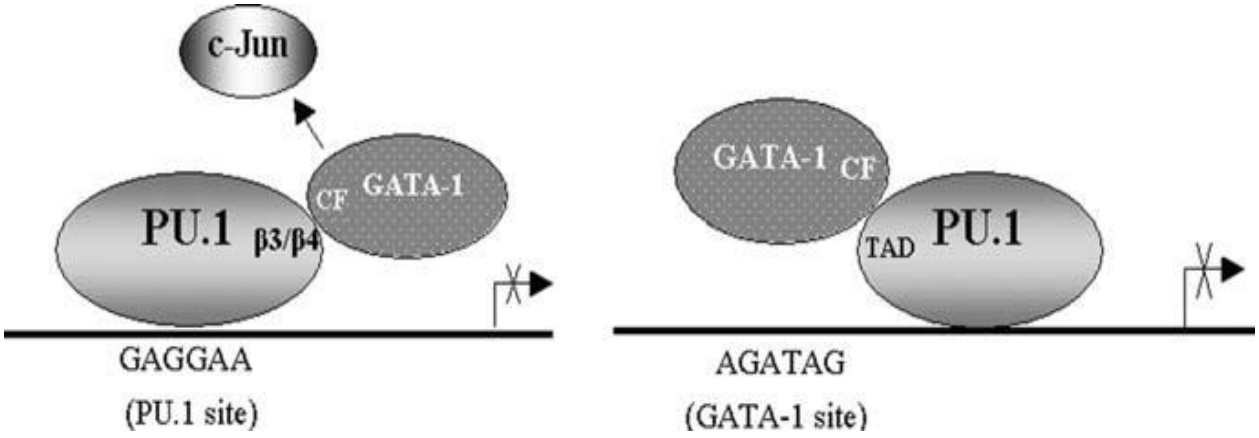


Figure 18: Cross antagonism between PU.1 and GATA-1. GATA-1 represses PU.1 function by interacting through its C-terminal zinc finger with the β3/β4 region of PU.1 and displacing its co activator c-Jun. On the other hand, PU.1 represses GATA-1 function by interacting through its trans-activation domain with the C-terminal zinc finger of GATA-1 and thereby inhibiting its binding to the cognate DNA sequence.

4.0.10 GATA-1 residues showing Interaction with PU.1

High interaction	Moderate Interaction	Lesser Interaction	Reference
GATA-1 C -Terminal end of -α helix in CF i.e. **Gly283, Leu284, Leu288, His289, Gln290, Arg293**	GATA-1 N-terminal i.e. **Ile242, Met248, Ser251**	GATA-1 C -Terminal end of -α helix in CF i.e. **Ala309, Ser310**	Liew et.al M, Nerlov C, Poulsen FM, Mackay JP.

4.0.11 PU.1 residues showing Interaction with GATA-1

High interaction	Reference
Leu172, Tyr173, His205, Glu207, Lys227, Met228, and Lys245, together with the side chain indole NH of **Trp213** and the side chain amide group of **Asn219**	Liew CW, Rand KD, Simpson RJ, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba

4.0.12 Pre-Docking Preparation of Protein

All the PDBs were cleaned with the help of various software tools. Removal of water molecules done via AutoDock Vina, Addition of Polar Hydrogens done via AutoDock Vina, Removal of ligand and phosphate groups from their active site, and Addition of Kollman Charges, Removal of repeated subunits from PDBs such as 1PUE and 1JNM.

All the PDBs are cross check via Pymol that they supposed to not having any repeated subunits or any ligand and phosphate group in their active site. PyMol is an open source visualizer tool available for use in structural biology for 3D structures and for generating molecular graphics. The Py portion of the software's name refers to the fact that it extends, and extensible by the Python programming language. PyMol can produce high-quality 3D images of small molecules and biological macromolecules such as proteins. It is also used to demonstrate the high degree of sequence and structural similarities. These all PDBs of macromolecule were treated to be completely rigid for all docking studies to reduce the extensive computational costs. Binding pockets was predicted by CASTp or Computer Atlas of Surface Topology of Proteins. It is a tool in bioinformatics which is online resource for identifying some of the geometric properties of protein like locating, delineating and measuring concave surface regions on 3D structures of proteins obtained from PDB. Protein protein docking and Protein peptide docking and virtual screening were performed using AutoDock Vina.

4.0.13 Extraction of β3/β4 region of PU.1

Transcription factor c-Jun binds with β3/β4 region of PU.1 and leads to Myelopoiesis. GATA-1 competes for the binding site with c-Jun to bind to PU.1, its binding leading to Erythropoiesis. The extraction of β3/β4 region from PDB (1PUE) was done via Pymol and Notepad, as this part is responsible for binding to both GATA-1 and c-JUN.

4.0.14 Generation of scrambled library of β3/β4 region of PU.1

The β3/β4 region of PU.1 is 13mer peptide sequence. Peptide library tool of genscript.com was used to generate scrambled peptide of β3/β4 region. GenScript provides high-quality peptide synthesis service with a 95% success rate. We took β3/β4 region (13 mer) of PU.1 and generate scrambled peptides via peptide_library tool of genscript.com

- the number of sequences = 579
- the size of library must be <= 50
- length of the inputted sequence = 13 [NYGKTGEVKKVKK] ;
- the Molecular Weight = 1694.93802

This tool will give us only the peptide sequences of whole 500 scrambled peptides which are generated. The Stability of generated peptides was predicted by PoPMuSiC. It is a web server for the estimation of protein stability changes upon mutation and sequence optimality.

4.0.15 Orientation of PU.1 against c-Jun and GATA1

I found the orientation of PU.1 binding with both c-Jun and GATA1 for checking our results of docking. This orientation was found via superimposition of PU.1 on both the GATA1 and on cJun using Pymol.

4.0.16 Peptide 3D Structure Modelling and Energy Minimization

Three dimensional structures of scrambled peptide sequences which are generated via peptide_library tool of genscript.com with the help of PEP-FOLD tool at mobile@RPBS site. PEP-FOLD is a *de novo* approach aimed at predicting peptide structures from amino acid sequences. It uses hidden Markov Model in which greedy algorithm and a coarse-grained force field is used for prediction of 3D structures.Energy minimization of all peptides was done with the help of SPDBviewer. Swiss-PdbViewer analyze the structure as well as interaction between different molecules, locate and determine a model's error and generate images.

4.0.17 Intermolecular H-Bonds interaction

I found those residues which makes interactions with the other chain of same protein in case of protein which exist in the form of dimer under the natural conditions, of all the PDBs for the purpose of removing the steric hindrance or the possibility of breaking the bonds or interactions of intermolecular HBonds after binding with our peptide. These residues are found via both Pymol and Discovery studio.

4.0.18. Docking of wild type (β3/β4 region of PU.1) with c-Jun and GATA1

In this case wild type was used in the form of reference for checking our virtual screening results. Docking of wild type (β3/β4 region of PU.1) with cJun (1JNM PDB) was done via AutoDock Vina with the grid of centre_x = 3.8057, centre_y = -5.6020, centre_z=13.8560, size_x = 32.3736, size_y = 39.8905, size_z = 59.1722 with spacing 0.375A.

Docking of wild type (β3/β4 region of PU.1) with GATA1 (2GAT PDB) was done via AutoDock Vina with the grid of centre_x = 7.8340, centre_y = 54.4930,centre_z =42.0176, size_x = 17.1276, size_y = 17.9244, size_z = 20.8277 with spacing 0.375A.

Intermolecular HBonds were seen via Discovery studio, Ligplot or Pymol visualizer software, Polar contact via Pymol, Dessolvation energy, Vander waal energy and electrostatic energy were calculated using FastContact tool, contact maps via COCOMAPS tool or Pymol, RMSD using DSvisualizer and SASA(Solvent Accessible Surface Area of protein) using Pymol.

FastConatct is a server that estimates the direct electrostatic and desolvation interaction free energy between two peptides in unit of kcal/mol. It also reports residue contact free energies that rapidly highlight the hotspots of the interaction and evaluates the vander Waals interaction using CHARMm. COCOMAPS(bioCOmplexes Contact MAPS) is a web application to easily and effectively analyse and visualize the interface in biological complexes (such as protein-protein, protein-DNA and protein-RNA complexes), by making use of intermolecular contact maps.

4.0.19 Virtual screening of scrambled library against c-Jun and GATA-1

Pyrx uses pdbqt format for both protein and peptides, so firstly convert all the protein and peptides in pdbqt format. Virtual screening was performed at centre_x = 3.8057, centre_y = -5.6020, centre_z = 13.8560, size_x = 32.3736, size_y = 39.8905, size_z = 59.1722 with spacing 0.375A, of PU.1 scrambled library against cJun via Pyrx. Pyrx gave result in the form of Binding energy with Rank, and output of peptides will be stored under the macromolecular folder. Output file contain all the possible poses in which a particular peptide bind with the desired protein. All the poses of peptides will be open in running Pyrx window and than save one by one in PDB format.

Now the Intermolecular HBonds were seen via Discovery studio, Ligplot or Pymol visualizer software, Polar contact via Pymol, Dessolvation energy, Vander waal energy and electrostatic energy were calculated using FastContact tool, contact maps via COCOMAPS tool or Pymol, RMSD using DSvisualizer and SASA calculate using Pymol.

Further selection of peptides for screening against GATA1 done on the basis of Binding energy, Intermolecular HBonds, Dessolvation energy, Electrostatic energy, Vander waal energy etc. Select those peptides which have the lowest Binding energy, Dessolvation energy, Electrostatic energy, Vander waal energy as compared to wild type and also have reported interacting residues for further screening against GATA1(2GAT PDB).

Pyrx uses pdbqt format for both protein and peptides, so firstly convert all the protein and peptides in pdbqt format. Virtual screening was performed at centre_x = 7.8340, centre_y = 54.4930, centre_z = 42.0176, size_x = 17.1276, size_y = 17.9244, size_z = 20.8277 with spacing 0.375A, of selected PU.1 scrambled library against GATA1 via Pyrx. Pyrx gave result in the form of Binding energy with Rank, and output of peptides will be stored under the macromolecular folder. Output file contain all the possible poses in which a particular peptide bind with the desired protein. All the poses of peptides will be open in running Pyrx window and than save one by one in PDB format.

Now the Intermolecular HBonds were seen via Discovery studio, Ligplot, or Pymol visualizer software, Polar contact via Pymol, Dessolvation energy, Vander waal energy and electrostatic energy were calculated using FastContact tool, contact maps via COCOMAPS tool or Pymol, RMSD using DSvisualizer and SASA calculate using Pymol.

5.0 RESULTS

Sequences of PU.1, c-Jun and GATA1 protein are given below and multiple sequence alignment was done for finding the conserved residues or motifs in the sequences of all these sequences but we could not found any similarity.

1) Sequence of c-Jun

>gi|209724|gb|AAA42416.1| jun oncogene, partial [Avian sarcoma virus]

GSGLYPSLAGVGEQQGQGGDTPPGAEQSRAATGHAGLAPGPALAVPPLRGLCSMSA
KMEPTFYEDALNAS

FAPPESGGYGYNNADILTSPDVGLLKLASPELERLIIQSSNGLITTTPTPTQFLCPKNVT
DEQEGFAEGFVRALAELHNQNTLPSVTSAAQPVSGGMAPVSSMAGGGSFNTSLHSE
PPVYANLSNFNPNALNSAPNYNANRMGYAPQHHINPQMPVQHPRLQALKEEPQTVP
EMPGETPPLFPIDMESQERIKAERKRMRNRIAASKSRKRKLERIARLEEKVKTLKAQN
SELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF

Figure17: Amino acid sequence information of cJun transcription factor

Amino acid sequences which are reported to make interaction with PU.1 is represented using red color font

2) Sequence of PU.1

>gi|124028517|ref|NP_003111.2| transcription factor PU.1 isoform 2 [Homo sapiens]

MLQACKMEGFPLVPPPSEDLVPYDTDLYQRQTHEYYPYLSSDGESHSDHYWDFHPH
HVHSEFESFAENNF

TELQSVQPPQLQQLYRHMELEQMHVLDTPMVPPHPSLGHQVSYLPRMCLQYPSLSP
AQPSSDEEEGERQS

PPLEVSDGEADGLEPGPGLLPGETGSKKKIRLYQFLLDLLRSGDMKDSIWWVDKDK
GTFQFSSKHKEALA

HRWGIQKGNRKKMTYQKMARALRNYGKTGEVKKVKKKLTYQFSGEVLGRGGLAE
RRHPPH

Figure18: Amino acid sequence information of PU.1 transcription factor

Amino acid sequences which are reported to make interaction with GATA1 and c-Jun is represented in red color font

3) Sequence of GATA-1

>gi|45382623|ref|NP_990795.1| erythroid transcription factor [Gallus gallus]

MEFVALGGPDAGSPTPFPDEAGAFLGLGGGERTEAGGLLASYPPSGRVSLVPWADT
GTLGTPQWVPPATQ

MEPPHYLELLQPPRGSPPHPSSGPLLPLSSGPPPCEARECVNCGATATPLWRRDGTGH
YLCNACGLYHRLNGQNRPLIRPKKRLLVSKRAGTVCSNCQTSTTTLWRRSPMGDPV
CNACGLYYKLHQVNRPLTMRKDGIQTRNRKVSSKGKKRRPPGGGNPSATAGGGAP
MGGGGDPSMPPPPPPPAAAPPQSDALYALGPVVLSGHFLPFGNSGGFFGGGAGGYT
APPGLSPQI

Figure19: Amino acid sequence information of GATA1 transcription factor

Amino acid sequences which are reported to make interaction with PU.1 and compete with binding of cJun to PU.1 is represented in red color font

4) Multiple Sequence Alignment using Clustal-W

```
gi|45382623|ref|NP_990795.1|      ----
MEFVALGGPDAGSPTPFPDEAGAFLGLGGGERTEAGGLLASYPPSG 46

gi|209724|gb|AAA42416.1|
GSGLYPSLAGVGEQQGQGGDTPPGAEQSRAATGHAGLAPGPALAVPPLRG 50

gi|124028517|ref|NP_003111.2|     -------
MLQACKMEGFPLVPPPSEDLVPYDTDLYQRQTHEYYPYLSSDG 43

                                  :     *    *        .    .   . . *

gi|45382623|ref|NP_990795.1|      RVSLVPWADTGTLG---
TPQWVPPATQMEPPHYLELLQPPRG---SPPHP 90

gi|209724|gb|AAA42416.1|
LCSMSAKMEPTFYEDALNASFAPPESGGYGYNNADILTSPDVGLLKLASP 100

gi|124028517|ref|NP_003111.2|     ESHSDHYWD-------
FHPHHVHSEFESFAENNFTELQSVQP-------P 79

                                  :       . . .     :   * .        *

gi|45382623|ref|NP_990795.1|
SSGPLLPLSSGPPPCEARECVNCGATATPLWRRDGTGHYLCNACGLYHRL 140

gi|209724|gb|AAA42416.1|
ELERLIIQSSNGLITTTPTPTQFLCPKNVTDEQEGFAEGFVRALAELHNQ 150

gi|124028517|ref|NP_003111.2|     QLQQLYRHMELEQMHVLDTPMVPPHPS-----
LGHQVSYLPRMCLQYPSL 124
```

```
                    . *  .        .       : .

gi|45382623|ref|NP_990795.1|      N------
GQNRPLIRPKKRLLVSKRAGTVCSNCQTSTTTLWRRSPMGDPV 184

gi|209724|gb|AAA42416.1|
NTLPSVTSAAQPVSGGMAPVSSMAGGGSFNTSLHSEPPVYANLSNFNPNA 200

gi|124028517|ref|NP_003111.2|     S--------------
PAQPSSDEEEGERQSPPLEVSDGEADGLEPGPGLL 160

                    .              .  . . .    .

gi|45382623|ref|NP_990795.1|      CNACG-------
LYYKLHQVNRPLTMRKDGIQTRNRKVSSKGKKRRPPGG 227

gi|209724|gb|AAA42416.1|
LNSAPNYNANRMGYAPQHHINPQMPVQHPRLQALKEEPQTVPEMPGETPP 250

gi|124028517|ref|NP_003111.2|     PGETG-----------
SKKKIRLYQFLLDLLRSGDMKDSIWWVDKDKGTF 199

                    .         ::    .   ::: . : .

gi|45382623|ref|NP_990795.1|
GNPSATAGGGAPMGGGGDPSMPPPPPPPAAAPPQSDALYALGPVVLSGHF 277

gi|209724|gb|AAA42416.1|
LFPIDMESQERIKAERKRMRNRIAASKSRKRKLERIARLEEKVKTLKAQN 300

gi|124028517|ref|NP_003111.2|
QFSSKHKEALAHRWGIQKGNRKKMTYQKMARALRNYGKTGEVKKVKKKLT 249

                    .              .      . .     . .

gi|45382623|ref|NP_990795.1|      LPFGNSGGFFGGGAGGYTAPPG----------LSPQI--- 304

gi|209724|gb|AAA42416.1|
SELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF 340

gi|124028517|ref|NP_003111.2|     YQFSG-EVLGRGGLAERRHPPH---------- 270

                    :..  :    .
```

Figure20: Multiple sequence alignment of GATA1, cJun and PU.1, this MSA was computed using CLUSTALW2.

First one has amino acid sequence of GATA1, second has cJun and third has PU.1 amino acid sequence. Amino acid sequence of GATA1 which has been reported to make interaction with PU.1 and show competitive binding with cJun is represented in red color font, amino acid

sequence of cJun which reported to bind with PU.1 is represented in blue color font while sequence information of PU.1 which has reported to make interaction with GATA1 and cJun is represented in pink color font.

5.0.1 Orientation of PU.1 against c-Jun and GATA-1

The main goal is to predict how a Ligand may bind but not whether it can bind means binding affinity and stability of Ligand with receptor is important. Binding Affinity can also be termed as strength of association between two molecules i.e. Ligand and receptor in our case (Ligand is PU.1 and receptor is c-Jun and GATA-1). Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using. The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g., agonism v/s antagonism). Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs. So that for finding the correct orientation of PU.1 withGATA1 and c-Jun in complex form we firstly done the docking of PU.1 wild type (β3/β4 region) with GATA1 and then with c-Jun for further testing the scrambled peptides.

5.02 Docking of PU.1 wild type (β3/β4 region) with GATA1 (2GAT)

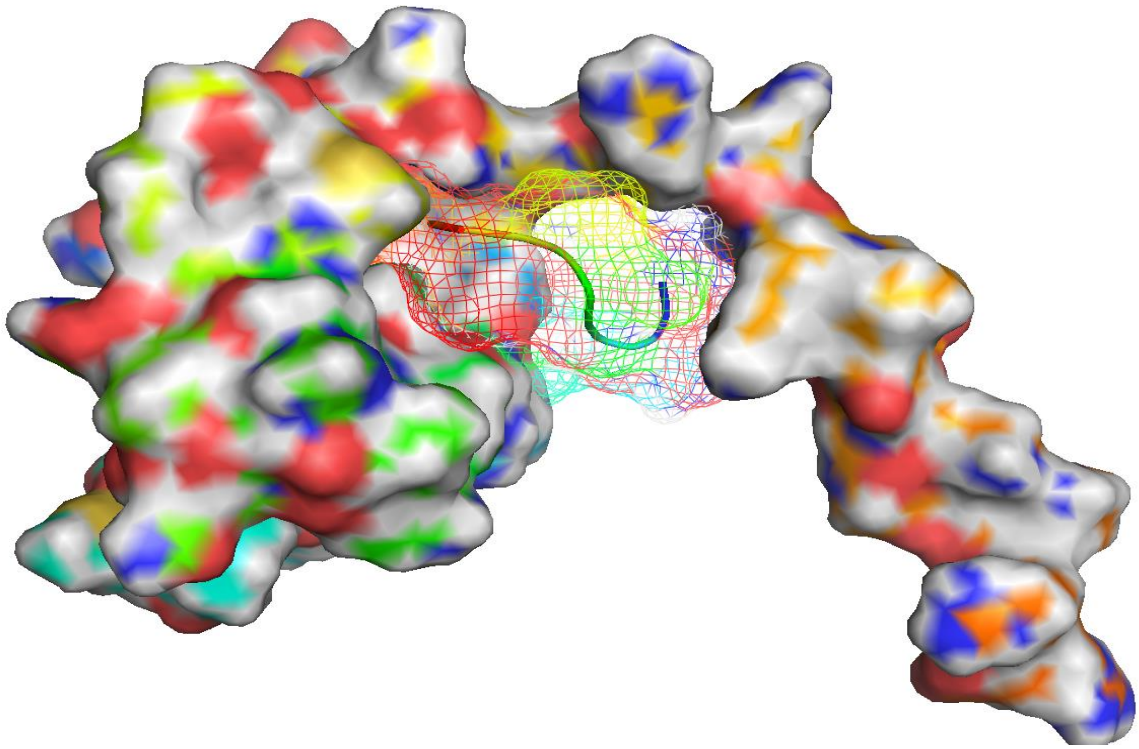


Figure 21: PU.1 (β3/β4 region) captured in active site of GATA-1. PU.1 is represented in mesh like network while GATA-1 is represented with colored solid surface

5.03. Docking of PU.1 wild type (β3/β4 region) with c-Jun (1JNM)

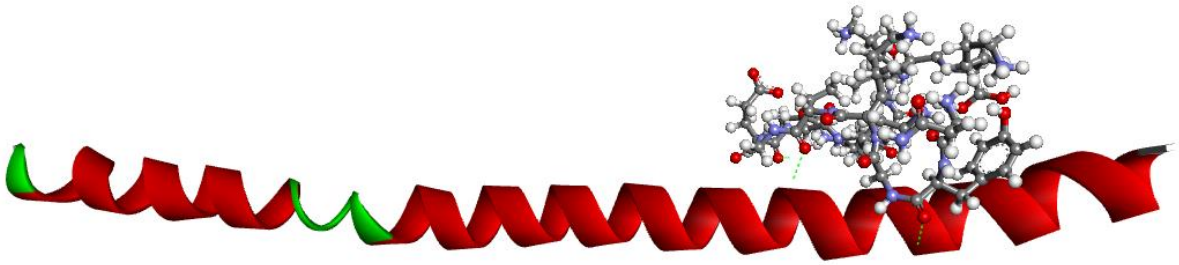


Figure22: Red color represents monomer unit of c-Jun while multicolored ball and stick model used for representing PU.1 (β3/β4)

5.04. Docking result of wild type PU.1 with c-JUN

Ligand	Binding Affinity	rmsd/ub	rmsd/lb
jnm_254-285_o_wild_model1	-3.3	0	0
jnm_254-285_o_wild_model1	-3.1	10.98	3.419
jnm_254-285_o_wild_model1	-3.1	16.138	10.458
jnm_254-285_o_wild_model1	-3.1	10.003	3.518
jnm_254-285_o_wild_model1	-3.1	23.149	14.729
jnm_254-285_o_wild_model1	-3	9.628	4.28
jnm_254-285_o_wild_model1	-2.9	13.159	4.567
jnm_254-285_o_wild_model1	-2.9	8.347	4.119
jnm_254-285_o_wild_model1	-2.8	24.835	17.744

We did control docking in which 254-285 residues of c-Jun (1JNM) were docked with PU.1 (ß3/ß4 region).

Interacting Reported Residues in papers: 254 to 285; Total residue: 32. We obtained

KAERKRMRNRIAASKSRKRKLERIARLEEKVK

Interacting residues obtained after docking:

Name	Symbol	Distance
B:ARG272:HH11 - :THR5:O	R	2.43671
B:ASN262:HD21 - :LYS13:O1	N	2.46611
B:LYS268:HZ3 - :TYR2:O	K	1.83358
B:SER269:HG - :GLU7:O	S	2.19185

Reported residue: KAERKRMRNRIAASKSRKRKLERIARLEEKVK

Interacting residue after docking: KAERKRMRNRIAASKSRKRKLERIARLEEKVK

Red color shows the c-Jun residues showing interaction with PU.1

5.05 Intermolecular Hydrogen bonds

Hydrogen bond is the electromagnetic attractive interaction between polar molecules, in which hydrogen (H) is bound to a highly electronegative atom, such as nitrogen (N), oxygen (O) or fluorine (F). The hydrogen bond (5 to 30 kJ/mole) is stronger than a vander Waals interaction, but weaker than covalent or ionic bonds. Intermolecular hydrogen bonding is responsible for the high boiling point of water (100°C). Intra-molecular hydrogen bonding is partly responsible for the secondary and tertiary structure of proteins and nucleic acids. Weak hydrogen bonds are useful in the rational design of drugs. These H-Bonds were found for the purpose of removing the steric hindrance or the possibility of breaking the intermolecular bonds or interactions between protein and DNA after binding with our desired peptide.

5.05.1 Hydrogen bonds with DNA

This part contain those residues which makes interactions with the DNA of all the PDBs, for the purpose of removing the steric hindrance or the possibility of breaking the bonds or interactions between protein and DNA after binding with our desired peptide.

5.05.1.0 2GAT with DNA

We calculated first amino acid residues which make contact with the DNA in 2GAT of GATA1 protein.

Table 4: Shows the residues of both protein and DNA which participate in the formation of intermolecular H-Bonds they were found using DS visualizer.

Sr.No.	Atom 1	Atom 2	Distance
1.	A:ARG2:HH12	B:DC105:OP2	1.84991
2.	A:ARG2:HH22	B:DC105:OP2	2.10056
3.	A:ARG42:HH12	C:DT123:OP1	2.4412
4.	A:ARG54:HH22	B:DC113:OP1	2.19801
5.	A:LYS57:HZ1	B:DT109:O2	2.40056
6.	A:ARG19:HH12	B:DG107:N7	2.10283
7.	A:LYS57:HZ2	B:DT109:O2	2.13808

5.05.1.03 Pictorial representation of 2GAT with DNA

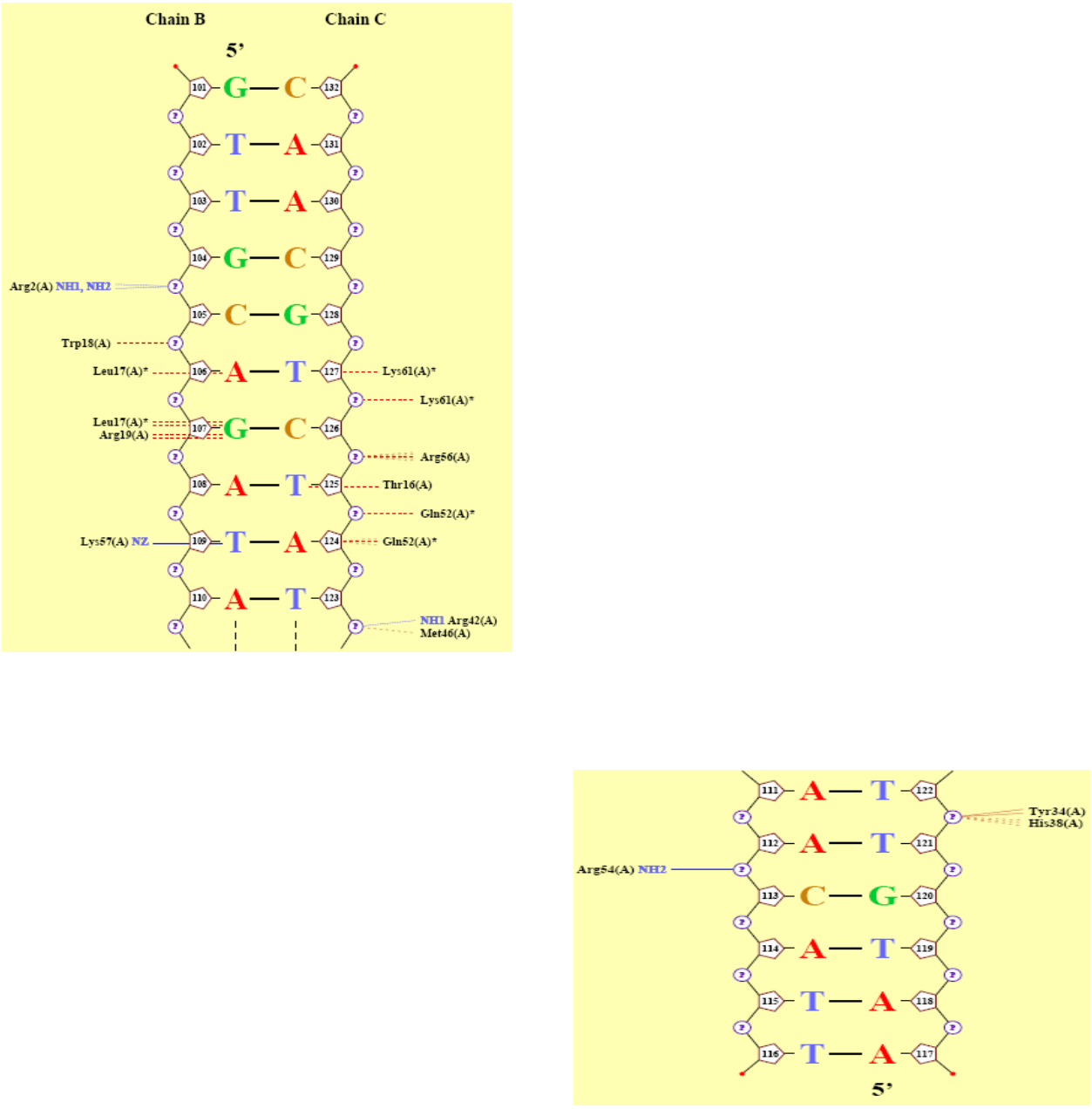


Figure 23: 2GAT with DNA

Here chain A represents protein chain while B and C represents DNA chain, ring represents Backbone sugar and base-number, circle represents phosphate group, blue straight line represents Hydrogen bond to DNA, red dashed line represents Non-bonded contact to DNA (<3.5A), while ice circle represents the Water molecule and number via PDBSUM.

5.05.1.04 Pictorial representation of 1JNM with DNA

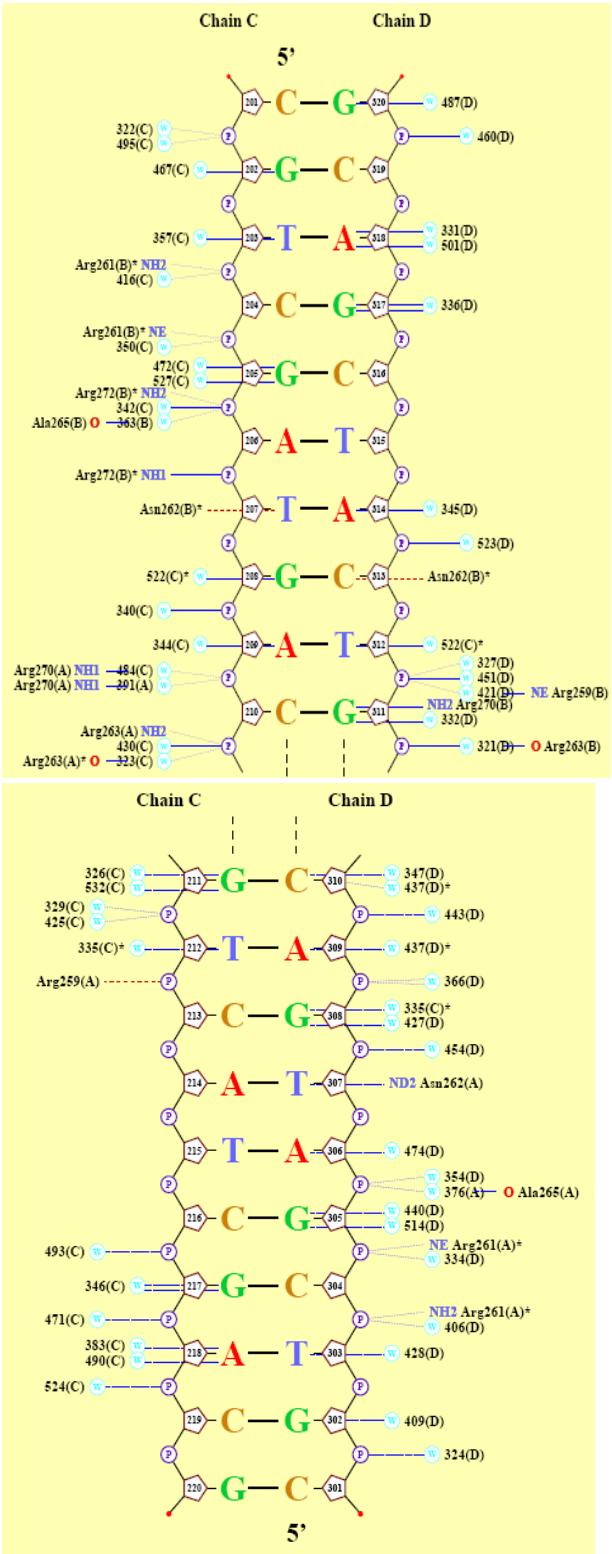


Figure 24: 1JNM with DNA. Here A and B represents protein chain while, while C and D represents DNA chain, Ring represents Backbone sugar and base-number, circle represents phosphate group, blue straight line represents Hydrogen bond to DNA, red dashed line represents Nonbonded contact to DNA (<3.5A), while ice circle represents the Water molecule and number via PDBSUM.

5.05.1.05 PUE with DNA

We calculated first the amino acid residues which make contact with the DNA in 1PUE pdb of PU.1 protein.

Table 5: Shows the residues of both protein and DNA which participate in the formation of intermolecular HBonds they were found using DSvisualizer.

Sr.No.	Atom 1	Atom 2	Distance
1.	E:LEU174:N	B:DC22:OP1	2.95628
2.	E:SER205:OG	A:DA5:OP1	2.6557
3.	E:TRP215:NE1	B:DT23:OP2	3.00042
4.	E:LYS219:NZ	B:DC22:O3	3.11827
5.	F:LYS219:NZ	B:DT23:OP1	2.62778
6.	E:ARG222:NE	A:DG14:OP1	2.85055
7.	E:ARG223:N	B:DT24:OP1	2.9427
8.	E:LYS227:OH	C:DG6:OP2	3.19493
9.	E:LYS229:NZ	B:DT24:OP2	2.68753
10.	E:ARG232:NH1	C:DG9:O6	2.6313
11.	F:ARG232:NH1	C:DG9:N7	2.52408
12.	E:ARG235:NH2	A:DG8:N7	2.82999
13.	E:TYR237:OH	B:DC22:OP2	2.35643

5.05.1.06 Pictorial representation of 1PUE with DNA

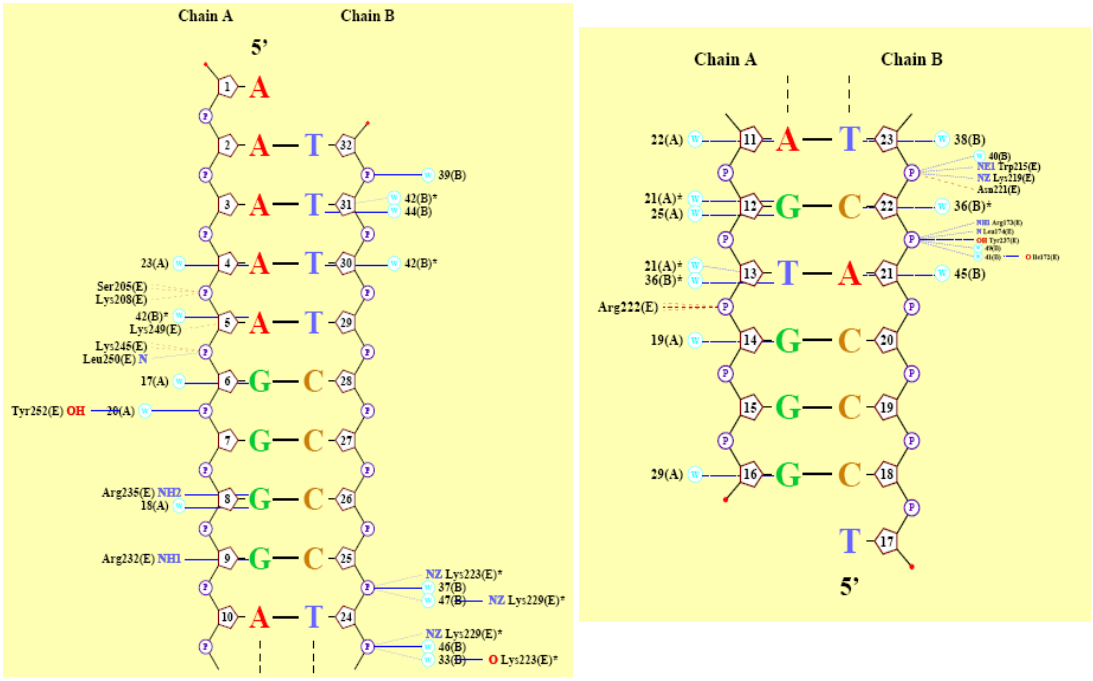
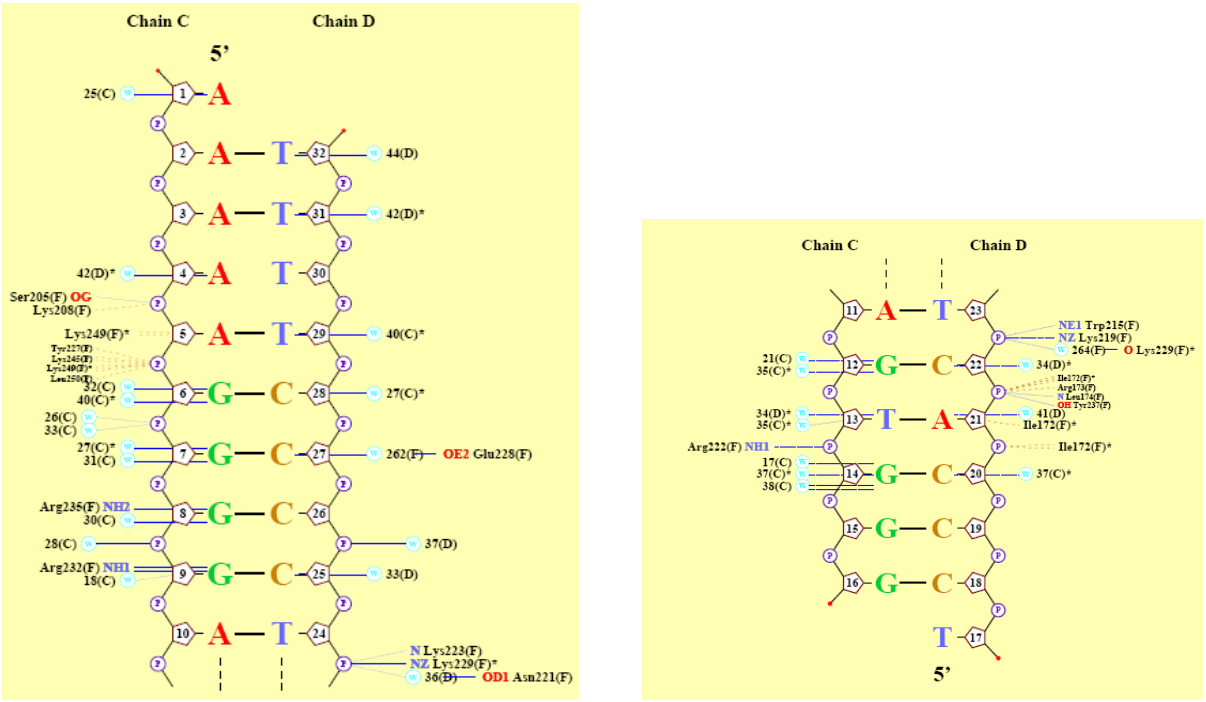


Figure 25: 1PUE with DNA. Here A and B represents DNA chain while, while E and F represents protein chain, Ring represents Backbone sugar and base-number, circle represents phosphate group, blue straight line represents Hydrogen bond to DNA, red dashed line represents Nonbonded contact to DNA (<3.5A), while ice circle represents the Water molecule and number via PDBSUM.



Here C and D represents DNA chain while, while E and F represents protein chain, Ring represents Backbone sugar and base-number, circle represents phosphate group, blue straight line represents Hydrogen bond to DNA, red dashed line represents Nonbonded contact to DNA (<3.5A), while ice circle represents the Water molecule and number via PDBSUM.

5.06 HYDROGEN BONDS WITHIN PROTEIN CHAIN

5.06.1 JNM inter molecular interaction between two protein chains

DS Visualizer can determine those atoms which makes interactions with the other chain of same protein. I found those residues which makes interactions with the other chain of same protein in case of protein which exist in the form of dimer under the natural conditions, of all the PDBs.

Table 6: Shows the residues of two protein chains which participate in the formation of intermolecular HBonds they were found using DSvisualizer.

Sr.No.	Atom 1	Atom 2	Distance
1.	A:ARG276:NH2	B:GLU281:OE1	2.36188
2.	A:ASN291:ND2	B:ASN291:OD1	2.35255
3.	B:ARG276:NH1	A:GLU281:OE2	3.12336
4.	B:ASN291:ND2	A:GLN290:OE1	2.84221

As reported in research papers it has been found there are two chain of c-Jun, chain A and Chain B. Atoms of both chain make bonded as well as non-bonded intermolecular interaction. We calculated the hydrogen bond distance between two atoms which are given in table. According to electrostatic force field distance between two atoms is inversely proportional to the force. we analysed that ASN291 residue have minimum distance, hence it can be said that this residue is showing strong intermolecular interaction.

5.06.2. H-bonds between wild type of PU.1 with c-Jun and GATA1

All the hydrogen bonds which were found after cJun docking with wild type PU.1 are given in table, or the residue number Lys12 in atom 2 column is already a reported residue of PU.1 to make hydrogen bond with GATA1. It is also known that the GATA1 and cJun used a competitive binding with the PU.1 β3/β4 region, so we supposed that they both bind with the similar residue of PU.1.

5.06.2 H-bonds between wild type of PU.1 with c-Jun

Table 7: Shows the residues of c-Jun and PU.1 which participate in the formation of intermolecular HBonds they were found using DSvisualizer.

Times	Sr.No.	Atom 1	Atom 2	B.E.	Distance
1st	1.	B:SER267:OG	:LYS12:H22	-3.6	2.34621
	2.	B:SER267:O	:LYS12:HZ3	-3.5	2.27793
	3.	B:SER267:O	:LYS12:HZ2	-3.6	2.19032
	4.	B:ARG279:HH12	:LYS12:O	-3.5	2.2682
	5.	B:LYS273:HZ3	:LYS12:O	-3.5	2.00134
2nd	1	B:ARG276:HE	:LYS12:O	-3.4	2.35822
	2	B:ARG276:HH22	:LYS12:O	-3.4	2.19736
	3	B:ARG272:HE	:LYS12:O	-3.4	2.214
	4.	B:ARG272:HH22	:LYS12:O	-3.4	2.23247
	5.	B:ARG276:HH22	:LYS12:O	-3.4	1.96431
3rd	1.	B:ARG276:HH22	:LYS12:O	-3.9	2.131
	2.	B:ARG276:HE	:LYS12:O	-3.9	2.0696
	3.	B:SER269:OG	:LYS12:HZ2	-3.9	2.13122
	4.	B:SER269:O	:LYS12:HZ3	-3.9	2.24022
	5	B:ARG263:O	:LYS12:HZ1	-3.9	2.20988
	6.	B:SER269:HG	:LYS12:O	-3.9	2.39537
4th	1.	B:ARG272:O	:LYS12:HZ2	-3.9	2.48703
	2.	B:ARG270:HE	:TYR2:O	-3.9	2.38847
5th	1.	B:ASN262:OD1	:LYS12:H	-3.9	1.90552
	2.	B:ASN262:HD22	:LYS12:O	-3.9	2.36164
	3.	B:ASN262:OD1	:LYS12:H	-3.9	1.89311
	4.	B:ASN262:HD22	:LYS12:O	-3.9	2.15535
	5.	B:ASN262:OD1	:LYS12:H	-3.9	2.07379

All the hydrogen bonds which were found after GATA1 docking with wild type PU.1 are given in table or the residue number Lys12 in atom 2 column and the residue number Ser59 is already a reported residue of PU.1 and GATA1 to make hydrogen bond.

5.06.4 H-bonds between wild type of PU.1 with GATA1

Table 8: Shows the residues of GATA1 and PU.1 which participate in the formation of intermolecular H-Bonds they were found using DS visualizer.

Times	**Sr. No.**	**Atom 1**	**Atom 2**	**B.E.**	**Distance**
1st	1.	A:SER59:HG	:LYS10:O	-3.6	2.028
	2.	A:LYS61:HZ1	:LYS12:O	-3.6	2.014
	3.	A:VAL58:H	:LYS12:O	-3.6	2.091
	4.	A:SER59:H	:GLY6:O	-3.6	2.165
	5.	A:SER59:HG	:GLY6:O	-3.6	2.410
	6.	A:SER59:O	:LYS10:HZ2	-3.6	2.465
	7.	A:SER59:OG	:LYS10:HZ2	-3.6	2.403
	8.	A:SER59:HG	:VAL11:O	-3.6	2.147
	9.	A:SER59:OG	:LYS13:O1	-3.6	3.164
	10.	A:SER59:HG	:THR5:O	-3.6	2.372
	11.	A:SER59:HG	:THR5:OG1	-3.5	2.148
	12.	A:SER60:OG	:LYS12:H	-3.5	1.991
	13.	A:SER59:OG	:LYS13:HZ2	-3.5	2.363
2nd	1.	A:ARG56:HH11	:LYS12:O	-3.5	2.357
	2.	A:SER59:H	:GLU7:O	-3.5	2.224
	3.	A:SER59:H	:GLU7:O	-3.5	1.927
	4.	A:SER59:HG	:GLU7:O	-3.5	2.274
	5.	A:SER59:H	:GLU7:O	-3.5	2.236
	6.	A:SER59:HG	:GLU7:O	-3.5	2.093

	7.	A:SER59:OG	:LYS10:H	-3.5	2.062
	8.	A:ARG56:O	:LYS12:HZ3	-3.5	2.163
	9.	A:ASN55:O	:LYS12:HZ3	-3.5	2.479
	10.	A:SER59:H	:ASN1:O	-3.5	2.075
	11.	A:SER59:HG	:ASN1:O	-3.5	2.448
	12.	A:SER59:O	:GLU7:O	-3.5	2.347
	13.	A:SER59:H	:LYS10:O	-3.5	2.153
	14.	A:SER59:HG	:LYS10:O	-3.5	2.353
	15.	A:SER59:O	:ASN1:H3	-3.5	2.239
	16.	A:SER59:OG	:GLY3:H	-3.5	2.121
	17.	A:SER59:O	:TYR2:H	-3.5	1.949
3rd	1.	A:SER59:H	:GLY6:O	-3.9	2.457
	2.	A:SER59:H	:VAL11:O	-3.9	2.374
	3.	A:SER59:HG	:GLY6:O	-3.9	1.988
	4.	A:SER59:HG	:LYS9:O	-3.9	2.117
	5.	A:SER59:H	:VAL8:O	-3.9	2.483
	6.	A:SER59:H	:LYS10:O	-3.9	2.432
	7.	A:SER59:HG	:VAL8:O	-3.9	1.917
	8.	A:SER59:H	:GLY6:O	-3.9	1.877
	9.	A:SER59:HG	:GLY6:O	-3.9	2.324
	10.	A:SER59:H	:GLY6:O	-3.9	2.171
	11.	A:SER59:H	:LYS12:O	-3.9	2.33608
	12.	A:SER59:HG	:LYS12:O	-3.9	1.97124
4th	1.	A:SER59:H	:GLU7:OE2	-3.6	1.95867
	2.	A:SER59:HG	:GLU7:OE2	-3.6	2.34224
	3.	A:SER59:OG	:GLU7:H	-3.6	2.24027

	4.	A:SER59:O	:GLU6:H	-3.6	2.1118
	5.	A:SER59:OG	:THR5:HG1	-3.6	2.0031
	6.	A:SER59:O	:ASN1:H1	-3.6	2.21596
	7.	A:SER59:OG	:ASN1:H2	-3.6	2.23565
	8.	A:SER59:OG	:LYS9:HZ3	-3.6	2.43325
	9.	A:SER59:OG	:GLY3:H	-3.6	2.00982
	10.	A:ASN55:O	:LYS12:HZ3	-3.6	2.49038
	11.	A:SER60:OG	:LYS13:HZ3	-3.6	2.22762
	12.	A:SER59:OG	:LYS9:HZ3	-3.6	2.31477
	13.	A:SER59:OG	:GLU7:H	-3.6	2.1706
	14.	A:SER59:HG	:LYS9:N	-3.6	2.34787
	15.	A:SER59:H	:VAL8:O	-3.6	2.36833
	16.	A:SER59:HG	:VAL8:O	-3.6	2.43998
	17.	A:SER59:HG	:GLY3:N	-3.6	2.33814
	18.	A:SER59:OG	:GLY3:H	-3.6	2.1048
	19.	A:SER59:O	:ASN1:H2	-3.6	1.99203
	20.	A:SER59:OG	:ASN1:H2	-3.6	2.45933
5^{th}	1.	A:SER59:H	:GLY6:O	-3.9	2.41166
	2.	A:SER59:HG	:GLY6:O	-3.9	1.99121
	3.	A:SER59:OG	:VAL8:H	-3.9	2.0857
	4.	A:SER59:OG	:ASN1:HD22	-3.9	2.02744
	5.	A:SER59:OG	:LYS9:H	-3.9	2.23862
	6.	A:SER59:O	:LYS9:HZ3	-3.9	2.37205
	7.	A:SER59:OG	:LYS9:HZ3	-3.9	2.1108
	8.	A:SER59:O	:LYS10:HZ2	-3.9	2.32274
	9.	A:SER59:H	:GLY6:O	-3.9	2.17832

	10.	A:SER59:HG	:GLY6:O	-3.9	2.1247
	11.	A:SER59:OG	:LYS9:HZ3	-3.9	2.20893
	12.	A:ARG56:HH11	:LYS12:O	-3.9	2.4767
	13.	A:SER59:H	:LYS9:O	-3.9	2.08065
	14.	A:SER59:HG	:LYS9:O	-3.9	2.30081
	15.	A:SER59:OG	:GLU7:H	-3.9	2.10812
	16.	A:SER59:H	:GLU:7	-3.9	2.23953
	17.	A:SER59:OG	:LYS10:H	-3.9	2.467
	18.	A:SER59:O	:LYS12:HZ2	-3.9	2.31446
	19.	A:SER59:H	:GLU7:OE1	-3.9	2.24435
	20.	A:SER59:H	:GLU7:OE2	-3.9	2.44877
	21.	A:SER59:HG	:GLU7:OE1	-3.9	2.06452
	22.	A:SER59:OG	:TYR2:H	-3.9	2.14926
	23.	A:SER59:OG	:ASN1:H3	-3.9	2.03082
	24.	A:SER59:H	:GLY6:O	-3.9	2.34949
	25.	A:SER59:HG	:GLY6:O	-3.9	2.0166
	26.	A:ARG56:O	:LYS12:HZ1	-3.9	2.48106
	27.	A:SER59:HG	:VAL8:O	-3.9	2.23729
	28.	A:SER59:OG	:LYS12:H	-3.9	2.12221
	29.	A:SER59:H	:GLU7:O	-3.9	2.28443
	30.	A:SER59:OG	:VAL11:H	-3.9	1.96444
	31.	A:SER59:OG	:LYS12:H	-3.9	2.30107
	32.	A:SER59:OG	:LYS12:HZ3	-3.9	2.49021

5.06.5 Bump table of wild type of PU.1 with c-Jun and GATA1

If each grid point was flagged if it is too close to an atom in the target (known as "bump") and if it is close enough to touch the target (known as "contact"). The Bump tell us about all the residues which are close enough to our grid or to target.

Table 9: Bump found after docking of cJun and PU.1 of cJun and PU.1 docking complex using DSvisualizer.

Times	**Sr.No.**	**Atom 1**	**Atom 2**	**B.E.**	**Distance**
1st	1.	B:SER267:HG	:LYS9:HZ3	-3.6	1.554
	2.	B:ARG261:HE	:THR5:HG1	-3.6	1.824
	3.	B:ALA265:CB	:GLU7:H	-3.6	2.210
2nd	1.	B:ASN262:HD21	:TYR2:HH	-3.4	1.264
	2.	B:SER269:HG	:LYS9:H	-3.4	1.673
3rd	1.	B:ARG261:HH22	:TYR2:HH	-3.9	1.643
	2.	B:ASN262:CB	:THR5:HG1	-3.9	2.306
	3.	B:LYS258:HZ3	:LYS9:HZ3	-3.9	1.568
	4.	B:SER269:HG	:LYS4:H	-3.9	1.971
	5.	B:ARG279:NH1	:TYR2:HH	-3.9	1.206
	6.	B:ARG279:HH11	:TYR2:HH	-3.9	1.669
4th	No Bump found				
5th	1.	B:ASN262:ND2	:THR5:HG1	-3.9	2.048
	2.	B:ASN262:ND2	:THR5:HG1	-3.9	1.537

Table 10: Bump found after docking of GATA1 and PU.1 of c-Jun and PU.1 docking complex using DS visualizer.

Times	Sr.No.	Atom 1	Atom 2	B.E.	Distance
1st	1.	A:VAL58:CB	:LYS9:H	-3.6	2.244
2nd	1.	A:SER59:HG	:LYS13:HZ3	-3.5	1.554
	2.	A:SER59:HG	:LYS9:H	-3.5	1.602
	3.	A:LYS63:HZ3	:LYS9:HZ1	-3.5	1.737
	4.	A:SER59:HG	:LYS9:H	-3.5	1.359
	5.	A:SER59:HG	:LYS9:H	-3.5	1.702
	6.	A:SER59:HG	:LYS9:H	-3.5	1.660
3rd	1.	A:SER59:HG	:VAL8:H	-3.9	1.626
	2.	A:SER59:HG	:GLU7:N	-3.9	2.153
	3.	A:SER60:HG	:LYS10:HZ2	-3.9	1.659
4th	1.	A:SER59:HG	:GLU7:H	-3.6	1.687
	2.	A:VAL58:H	:LYS13:HZ2	-3.6	1.826
	3.	A:SER59:HG	:LYS9:HZ3	-3.6	1.634
	4.	A:SER59:HG	:GLY3:H	-3.6	1.826
	5.	A:SER59:HG	:GLU7:H	-3.6	1.263
	6.	A:SER59:HG	:GLY3:H	-3.6	1.376
	7.	A:VAL58:H	:LYS13:HZ2	-3.6	1.826
	8.	A:SER59:HG	:GLU7:H	-3.6	1.687
5th	1.	A:SER59:HG	:VAL8:H	-3.9	1.391
	2.	A:SER59:HG	:GLY6:C	-3.9	2.210
	3.	A:SER59:H	:GLU7:CD	-3.9	2.208
	4.	A:SER59:HG	:ASN1:H3	-3.9	1.584
	5.	A:VAL58:H	:GLY3:H	-3.9	1.724

5.06 SASA table of c-Jun complexes

This was used to compute the accessible surface traced out by a probe molecule. The SASA should be lowest in complex form as compared to the SASA which was comes after doing sum of target protein and the ligand SASA value.

Table 11: SASA of wild type (PU.1) and c-Jun complexes

Pyrx run	SASA
1st	7193.87875
2nd	7149.284
3rd	7056.03175
4th	7010.458
5th	7065.938
Average	**7095.11812**

Table 12: SASA of wild type (PU.1) and GATA1 complexes

Pyrx run	SASA
1st	7339.772
2nd	7385.35557
3rd	7393.2525
4th	7368.5895
5th	7377.59125
Average	**7372.91218**

5.07 Table of GATA1 and c-Jun complex information with PU.1

Table 13: Energy calculations of wild type (PU.1) and c-Jun complexes

	SASA	Electrostatic energy (kcal/mol)	Vander waals energy (kcal/mol)	Dessolvation energy (kcal/mol)	Binding energy
GATA1	7372.91218	4.18336447	-284.266	0.105213057	-3.66
c-Jun	7095.11812	0	-337.565	0	-3.91

5.08 Contact maps of PU.1 with c-JUN

A protein contact map represents the distance between all possible amino acid residue pairs of a three-dimensional protein structure using a binary two-dimensional matrix. Contact maps provide a more reduced representation of a protein structure than its full 3D atomic coordinates. These contact maps are made using COCOMAPS tool.

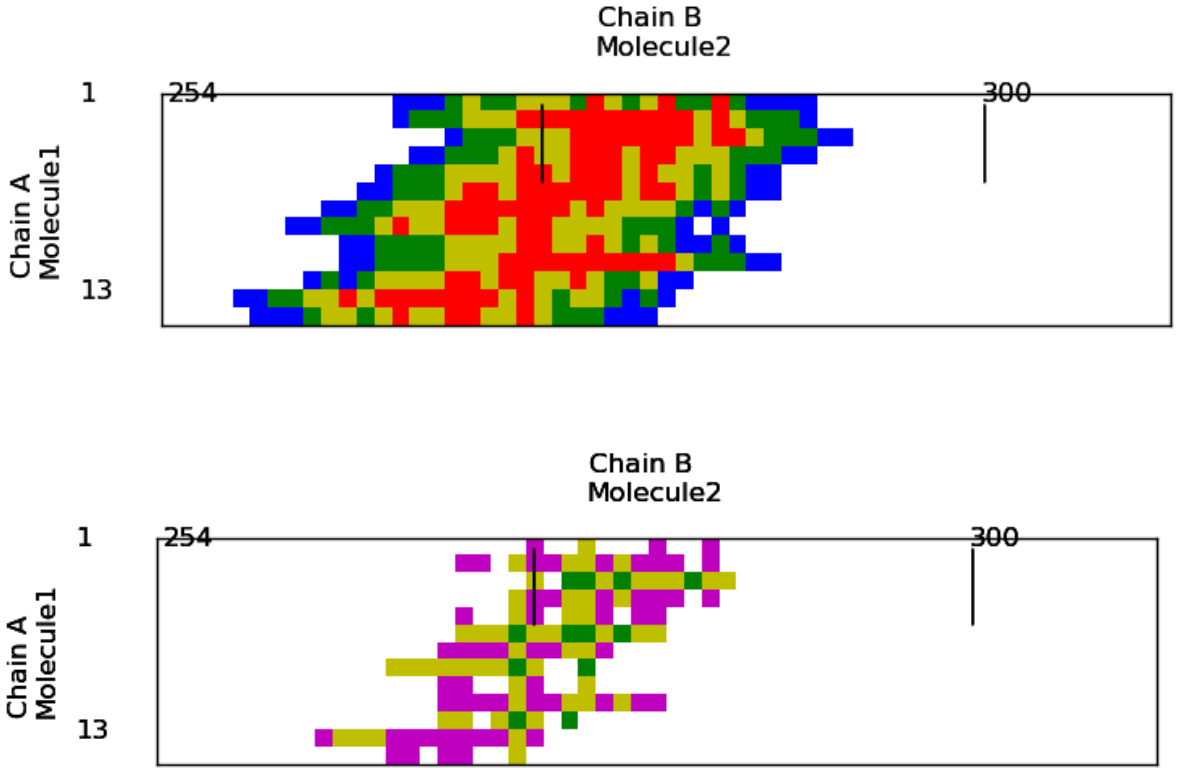


Figure 26: Contact map of wild type of PU.1 and cJun.

Here pink color represents the Hydrophilic Hydrophilic interaction, yellow color represents Hydrophilic Hydrophobic interactions while green color represents Hydrophobic Hydrophobic interactions.

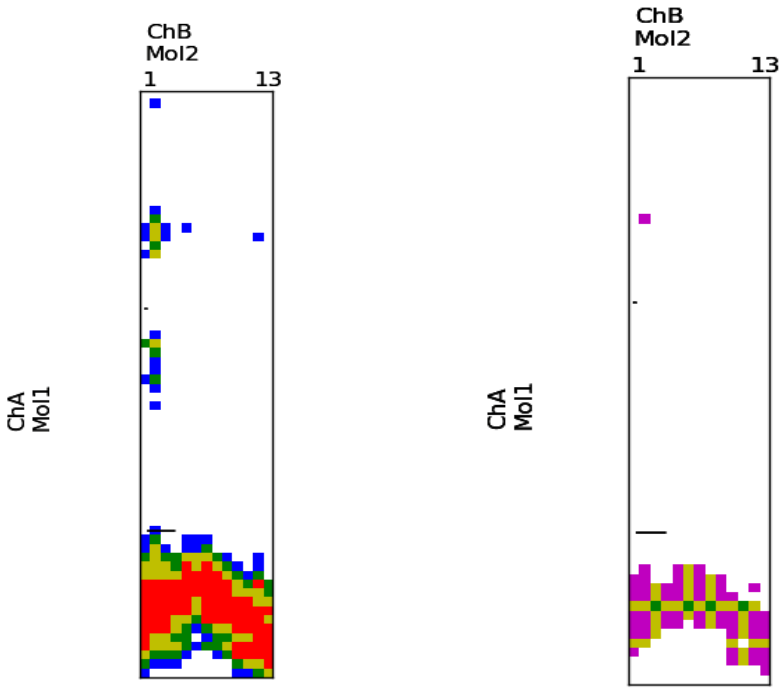


Figure 27: Contact map of wild type of PU.1 and GATA1.

Here pink color represents the Hydrophilic Hydrophilic interaction; yellow color represents Hydrophilic Hydrophobic interactions while green color represents Hydrophobic Hydrophobic interactions.

5.0.9 VIRTUAL SCREENING RESULT

I was done Virtual screening of 500 mimetic peptides against cJun using Pyrx to obtain the one peptide which has the lowest binding affinity with cJun while highest with GATA1.

Table 14: This table shows the binding energy of mimetic peptide against GATA-1 and c-Jun using Pyrx

Peptide Sequence	B.E.(GATA1)	B.E(c-Jun)	Peptide Sequences	B.E(GATA1)	B.E (c-Jun)
NYGKTGEVKKVKK	-3.66	-3.91	VGKKTEGKNKYVK	-4.1	-3.6
GYNKKGKVEKVTK	-4	-3.9	KVKKTGEVYKNKG	-3.7	-4.7
VKYGKKTNKGEVK	-3.5	-3.9	GKKEGKKTKVYVN	-3.6	-4.5
VKTGGKVKKNEKY	-4	-3.8	GKNEVKYTKGKVK	-3.5	-4.3
YVKKGGKNKVTKE	-3.7	-3.3	KKVEVGYKTGKNK	-3.5	-4.3
EKGNYKGVKKTKV	-3	-3.4	YKVKGVKTENKGK	-3.5	-4.2
KEYKGVNKKVKGT	-3.6	-3.9	ENVTKKKGVGKKY	-3.5	-4.1
GKKKTYGKVKENV	-3.3	-2.6	GKKKTKVNGEVKY	-3.5	-4.1
TKKVKKGVNEKGY	-3.4	-3.4	KKGVVKTGNEKKY	-3.5	-4.1
KKTNEKKVGKYGV	-3.9	-3.9	KKGKKYVEGKTNV	-3.5	-4.1
EKYKTVKNGKGVK	-3.3	-3.9	EVKKGKKVYNKGT	-3.5	-4.1
GKKKEVYNKKGTV	-3.9	-3.9	GKKEVVKGKKNYT	-4	-3.6
VVKYEKKGTKKNG	-3.4	-3.7	KGYKVEKVKKTNG	-4	-3.4
TGKGKVVENKKYK	-3.8	-3.7	YKKTVVKGENKGK	-3.9	-3.3
YENGGKTKVKVKK	-3.7	-3.9	KVKGKVKYENTGK	-3.9	-3.3
GVVKKKKENTKYG	-3.6	-3.6	YVKGKNVKGTKKE	-3.8	-3.2
NGYEKGVTKVKKK	-2.9	-3.9	VGKNKYKEKGVTK	-3.8	-3.1
GNGEKVVKKYKKT	-3.2	-3.9	VNEKGYVKTKKKG	-3.7	-3.1
KTEKKKKGNVVYG	-3.5	-3.7	VKKKGGKEYTKNV	-3.7	-3.1
KKGNYTEGVKKKV	-3.9	-3.8	TKVKGKKEYGVNK	-3.7	-3
VGVENKKGYKKKT	-3.8	-2.8	KKVKYNKGTGEKV	-3.8	-3
VYNKKKKTGGKEV	-3.3	-3.7	TGEYKVGKKVKKN	-3.7	-2.9
TKEKKVNVGGKKY	-3.4	-3.9	GEVNKVKYTKKGK	-3.7	-2.9
KVYTKKGNGVKKE	-3.3	-3.8	YKKGGETKVKNVK	-3.7	-2.8
KVKVNEKKGYTKG	-4.2	-3.2	TKKKGEVVKKNYG	-3.6	-2.8
KVKNVGKKKGYET	-3.6	-3.4	TVEGVKKYKKNGK	-3.6	-2.8
KGVVTKEKGKKNY	-3.4	-3.7	YKGEVKTGKNVKK	-3.5	-2.7
KKVKTVKGNYKEG	-3.6	-3.8	NKKGVGKTKKVYE	-3.5	-2.7
KNVKKKGKETVYG	-3.4	-3.8	EVVKKGKTKGYKN	-3.5	-2.7
YGKKGEKKVNKTV	-2.8	-3.7	KTVKKGKVYENGK	-3.5	-4.1
EYGVKKNTVGKKK	-3.4	-3.5	VNGTYKKEKGVKK	-3.5	-4
TGNKKVKEYKVKG	-3	-3.7	KNGKKVYVTGKEK	-3.5	-3.8
NVKKKVKGGEKTY	-3.9	-3.7	VKNKGYTKVKGEK	-3.5	-3.8
ETKVNKVKKKGGY	-3.4	-3.4	YVEGKTVNKKKGK	-3.4	-3.8
VKNKKKVTGKEGY	-2.8	-3.8	KTNKVYEKKKGGV	-3.4	-3.8
GYEKVKKGVKTKN	-2.7	-3.7	GVKKNYEVTKKGK	-3.4	-3.7
VKKYGNKKVTKGE	-3.7	-3.4	GEGVVKKKYNKKT	-3.3	-3.7

KKEVYNKTKKGGV	-2.9	-3.6	GETKYKVGNKKVK	-3.3	-3.6
EVKKYGTGVKNKK	-3.2	-3.9	VKEVYKGKTNGKK	-4.5	-3
VEKKVYTKKKGNG	-3.4	-3.8	KYGKNVTEGKVKK	-4.4	-3
KGNYVKKGTVKKE	-3.7	-3.4	EYKTGVKKKGVKN	-4.4	-3
TYEKNVVGKKKGK	-3.7	-3.7	YVKKGVNEKGTKK	-2.6	-2.9
VVNGTKEKGYKKK	-4	-3.8	GKVEKTKNKYKGV	-3.4	-2.9
VYKKKEKGTKGNV	-3.2	-3.5	KNEKKGKVTKYGV	-3.2	-2.8
TKVNKKKEVGYKG	-3.4	-3.4	KKKGVNYVTKGEK	-3.2	-2.8
KYVNKEGKKKTGV	-3.1	-3.8	GKKTEVYKVGNKK	-3.1	-2.8
EGTGKKKKNVYVK	-3.6	-3.8	KNKGTGYKKKVEV	-3.3	-2.8
KKVGKTVNEKGYK	-3.9	-3.3	VVGKYKTNKEKGK	-3.8	-3.2
KKNKEYGTKGVKV	-3.7	-3.6	TKKGKNVKGYEVK	-3.8	-3
EYGTNVKKKKKVG	-3.6	-3.8	KTYNVGKKEGVKK	-3.8	-3
KNYKETKVKKVGG	-3.8	-3.8	YVEKGTKNGKVKK	-3.7	-2.9
NGTGYKEVKVKKK	-3.5	-3	YKNGVKVTKKKGE	-3.7	-2.8
GGEKKKKTVKYVN	-3.9	-3.9	KEVGKKKGVYKTN	-3.6	-2.7
GKVVGKNEKTKYK	-3.7	-3.3	KEGVKKGKYTVKN	-3.6	-2.7
KGGKEVKNVTKKY	-3.3	-3.4	GYKKKVKNETKVG	-3.5	-2.7
KKNKGKTVKEVGY	-3.4	-2.8	KEKVVKGNYTGKK	-3.5	-2.7
KENKKGKVTYGVK	-3.4	-3	TGVYNKEKKVKKG	-4.4	-4.1
KGGNTVKKYKVEK	-3.4	-2.6	GVNYKTEKKKKVG	-4.4	-4.3
GVTEKKVNGKYKK	-3.1	-2.9	KVYKKNEKGTKGV	-4.4	-4.5
VTENKGYKKKVGK	-4.5	-3.6	YEKVKTGKVKNGK	-3.4	-3.9
VKTKEVNGKYKKG	-3.6	-4.4	KVEYGTKVKKGNK	-3.4	-4.4
YKKKVTNEKGGKV	-3	-4.3	VYKGETKKGVKKN	-3.6	-3.8
YVKGKKKNTGEKV	-4.1	-3.8	KGNKKTYEVKKVG	-3.8	-4.1
VKKKNKVGKYGTE	-3	-3.6	VKKNKTEVKKGGY	-4.1	-3.8
VKKENKYGTKGVK	-3.2	-4.1	YGKKKKGEVNTVK	-3.7	-3.6
KVKKNGKTGKYEV	-3.5	-4.4	KYNGVKGEVTKKK	-4.5	-4.5
KVKKNYKGEVGKT	-3	-3.9	GKYVGEKKNKTKV	-2.9	-3.4
TVKGNKEKYKVKG	-4	-4.3	GNKVKKGETYKVK	-3.2	-3.4
KKKEVTGGVYNKK	-3.1	-3.6	KVKVGEKNYTGKK	-3.9	-4.1
KVKGKNKKVYTGE	-3.7	-4.1	TKKVGEYNKGVKK	-3.1	-4.6
EKGTKKNKVVKYG	-3.8	-3.8	GKGVTKNKKKYEV	-3.4	-3.8
KKTEGVNKYVKGK	-3.6	-3.6	GGTVKKVKYKENK	-4.4	-3.1
KGTYKKNVKKVEG	-4.4	-4.9	VKEKNKGYKTKVG	-3.5	-3.2
KVGKGKVETYKKN	-3.8	-5	YGENKVKKKTKGV	-4.8	-4.1
KYKGKNTGVEKKV	-3.2	-4.4	KETVKYKGKKNVG	-4	-4.2
NEKKVTKVYKGGK	-2.8	-4.8	TVNKKYKGGEKVK	-4	-4.8
EGGVKKVNKYKTK	-4.6	-3.8	GKTVGKKKKVYNE	-3.4	-4.2
GKEKKGNKKYTVV	-3.4	-3.3	KKGGVVTKKNYKE	-3.7	-4.3
KKGVNYKKGVTKE	-3.4	-4.1	KKGVKVTKKNYGE	-3.5	-3.9
VKVKTEYKGKKNG	-3.4	-3.3	YGKVKKTNGVKKE	-3.4	-4.6
KKGVENKYKVTGK	-3.6	-4.7	GTGKKKYEVVNKK	-3.4	-3.8
EKKKTKVKVGYGN	-3.4	-3.6	GKKKVEYVTNGKK	-3.4	-3.7
EKYNVKGKKTKVG	-3.7	-4.3	KKKGNKGKTVVEY	-3.4	-4.2

KKGVKENYTGKVK	-4.1	-3.9	KKYKNTGKVVGEK	-3.4	-4.8
KNVVGTKGKYEKK	-3.9	-4.2	VGETKYKVGNKKK	-3.3	-4.1
NGKVYGVKKKETK	-4	-4	KGVKTGKVNKKYE	-3.3	-4.9
KGKKKEVYTNVGK	-3.2	-3.5	KYVGEGKKNTKKV	-3.8	-4.3
YGGKKEKTKKVVN	-2.8	-4	VVKGTKNKYKGKE	-3.8	-4.1
YGTVKKEKKNVKG	-3.5	-3.1	VKKGKGNKKYVTE	-3.8	-4
KVGKKYEVTGKNK	-3.9	-3.8	KKKGVGTEKYVKN	-3.7	-4.7
KKVYGKETGKVNK	-3.4	-4.2	KKYKVGKEGNVKT	-3.7	-4.7
EKTGKVKKGVKYN	-3.7	-4.6	KNEKGVKTKGKYV	-3.7	-4.9
EGKGKVKKTNKVY	-3.2	-3.5	KYTKGVEVKGKKN	-3.7	-4.7
KGYKVNKKTKGEV	-4	-3.6	EKTKGKVGKYVKN	-3.5	-4.1
KVYKTGKKKEGVN	-3.4	-3.6	KKEGVKVKGYTKN	-3.5	-3.6
GKGKTYNKKKVVE	-3.3	-3.5	VKTNEKGKGKYKV	-3.8	-3.7
KEGKVKKNVTYGK	-4.1	-3.4	VEYVKKKGKKNGT	-3.5	-4.3
KVENYKKKGGVKT	-4.6	-3.3	KKKVEVGGYKTKN	-3	-4.9
GYNVKKKKKGEVT	-4.4	-3.3	VKKKVKKEGYTNG	-3.4	-4.4
EYGKVKVKTNGKK	-4.1	-3.3	GEKVKKGKVNKTY	-3.2	-4.3
KVKGKGKKVTYNE	-3.1	-3.3	GGNKKKEVTYVKK	-3.7	-4.6
KTKVGKYEKGNVK	-4.4	-3.3	GKKVKKGYNKVTE	-3.7	-3.8
KKKGYGVKVENKT	-3.7	-3.5	KKYVNGTKEKVKG	-3.3	-4.2
KGVTYKVKEGNKK	-3.3	-3.3	NVYKGTGKKEVKK	-3.4	-4.6
TKEKGKNVKYGVK	-3.5	-3.3	KNYGKTKGVVKEK	-2.2	-4.2
KKGKKKEVNYGVT	-3.3	-3.2	KGVKYTKVNKKEG	-3.4	-4.5
VGKKKVKNGYKTE	-4	-3.2	KKEKYTVKGVNGK	-3.8	-3.7
KEYKGVKGNTKKV	-3.8	-3.1	KYVVGEKKKNGTK	-3.8	-3.2
NTVVKKKKKGYEG	-3.3	-3.1	TENKGYKKKVVGK	-2.7	-4.2
VEGKGVYKKNKKT	-3.4	-3.8	TYNVKGKKVGKEK	-3.9	-3.5
KGYKVNEKVKGKT	-3.7	-4	KKKTNYEGKVVKG	-3.8	-3.1
GVKKKYTVNGEKK	-2.8	-4.8	GKKKKNETVGKVY	-3.2	-4
NVKVKYTEGKKGK	-3.4	-3	GTNKKEVKGKKVY	-3.7	-3.4
GVTKKYGKKNVKE	-3.7	-3.8	ETKVKKKKGYVNG	-4.6	-4.2
YGETKGVKVKNKK	-3.1	-5.2	KKENGKKKVYVGT	-3.2	-4
NKGKVVGKKTKEY	-3.2	-3.9	KKKKVKYNEGGTV	-3.6	-3.3
TGKNKVKKGKEYV	-4.2	-4.3	KTKNYVKVKGGEK	-4.3	-4.5
KKGNYETKGKVVK	-4.7	-3.9	KKYEKGKNGTKVV	-3.5	-3.6
EKGKVTKVYNKKG	-3.5	-4.9	YEGGVKKNVTKKK	-3.6	-4.4
GKYKKTENKVGVK	-3.7	-4	KGTKVKNKVEYKG	-3.1	-3.5
KKEKTYGGNVKVK	-3.6	-4.4	KKVKTVKNGEKYG	-3.6	-4.5
KVEKGKKVKTYGN	-3.3	-3.5	VEGKYKGKKNKTV	-3	-3.9
YKEKNKKGVGKTV	-3.5	-3.9	TKKYGKKGEKNVV	-3.2	-4
GVEKKGKNTYKKV	-3	-4.5	KKVVKTNGEYKKG	-3.4	-4.2
TNGKGKKVVYKKE	-3	-4.5	KYKVKNGVETKKG	-4.1	-4.2
NGKEKVKVKTKYG	-4.2	-4.5	GKKVEGKYKNVKT	-3.8	-4.5
GKVEKKGKKTYNV	-3.5	-4.1	NYGKTGEVKKVKK	-3.7	-3.3

Those which are selected from the above table on the basis of criteria of Binding affinity (kcal/mol), poses, and binding residues are further screens against GATA1 for finding the highest binding affinity. Our aim is to select those residues which showed lowest binding energy with c-Jun and Highest binding energy with GATA-1,So we calculated binding energy difference between GATA-1 and c-Jun in such a way that binding energy of c-Jun is low and GATA-1 is high.

Sequences	B.E Difference (B.E_GATA-1 - B.E_c-Jun)	Sequences	B.E Difference (B.E_GATA-1 - B.E_c-Jun)
VKEVYKGKTNGKK	-1.5	GYKKKVKNETKVG	-0.8
KYGKNVTEGKVKK	-1.4	KEKVVKGNYTGKK	-0.8
EYKTGVKKKGVKN	-1.4	TKVKGKKEYGVNK	-0.7
GGTVKKVKYKENK	-1.3	YGENKVKKKTKGV	-0.7
KVENYKKKGGVKT	-1.3	GKKKTYGKVKENV	-0.7
GYNVKKKKKGEVT	-1.1	KEGKVKKNVTYGK	-0.7
KTKVGKYEKGNVK	-1.1	KEYKGVKGNTKKV	-0.7
VGVENKKGYKKKT	-1	VGKNKYKEKGVTK	-0.7
KVKVNEKKGYTKG	-1	KKKTNYEGKVVKG	-0.7
YKKGGETKVKNVK	-0.9	KKVGKTVNEKGYK	-0.6
YKNGVKVTKKKGE	-0.9	KKNKGKTVKEVGY	-0.6
VTENKGYKKKVGK	-0.9	KGYKVEKVKKTNG	-0.6
KEVGKKKGVYKTN	-0.9	YKKTVVKGENKGK	-0.6
KEGVKKGKYTVKN	-0.9	KVKGKVKYENTGK	-0.6
KKGNYETKGKVVK	-0.8	VNEKGYVKTKKKG	-0.6
TGEYKVGKKVKKN	-0.8	VKKKGGKEYTKNV	-0.6
GEVNKVKYTKKGK	-0.8	YVKGKNVKGTKKE	-0.6
TKKKGEVVKKNYG	-0.8	VVGKYKTNKEKGK	-0.6
TVEGVKKYKKNGK	-0.8	KYVVGEKKKNGTK	-0.6
YVEKGTKNGKVKK	-0.8	NGTGYKEVKVKKK	-0.5
KGGNTVKKYKVEK	-0.8	GKVEKTKNKYKGV	-0.5
EGGVKKVNKYKTK	-0.8	KNKGTGYKKKVEV	-0.5
EYGKVKVKTNGKK	-0.8	VGKKTEGKNKYVK	-0.5
VGKKKVKNGYKTE	-0.8	YVKKGGKNKVTKE	-0.4
KKVKYNKGTGEKV	-0.8	GKVVGKNEKTKYK	-0.4
YKGEVKTGKNVKK	-0.8	KNEKKGKVTKYGV	-0.4
NKKGVGKTKKVYE	-0.8	KKKGVNYVTKGEK	-0.4
EVVKKGKTKGYKN	-0.8	NYGKTGEVKKVKK	-0.4
TKKGKNVKGYEVK	-0.8	YGETKGVKVKNKK	2.1
KTYNVGKKEGVKK	-0.8		

5.09.1 Graphical representation of virtual screening result:

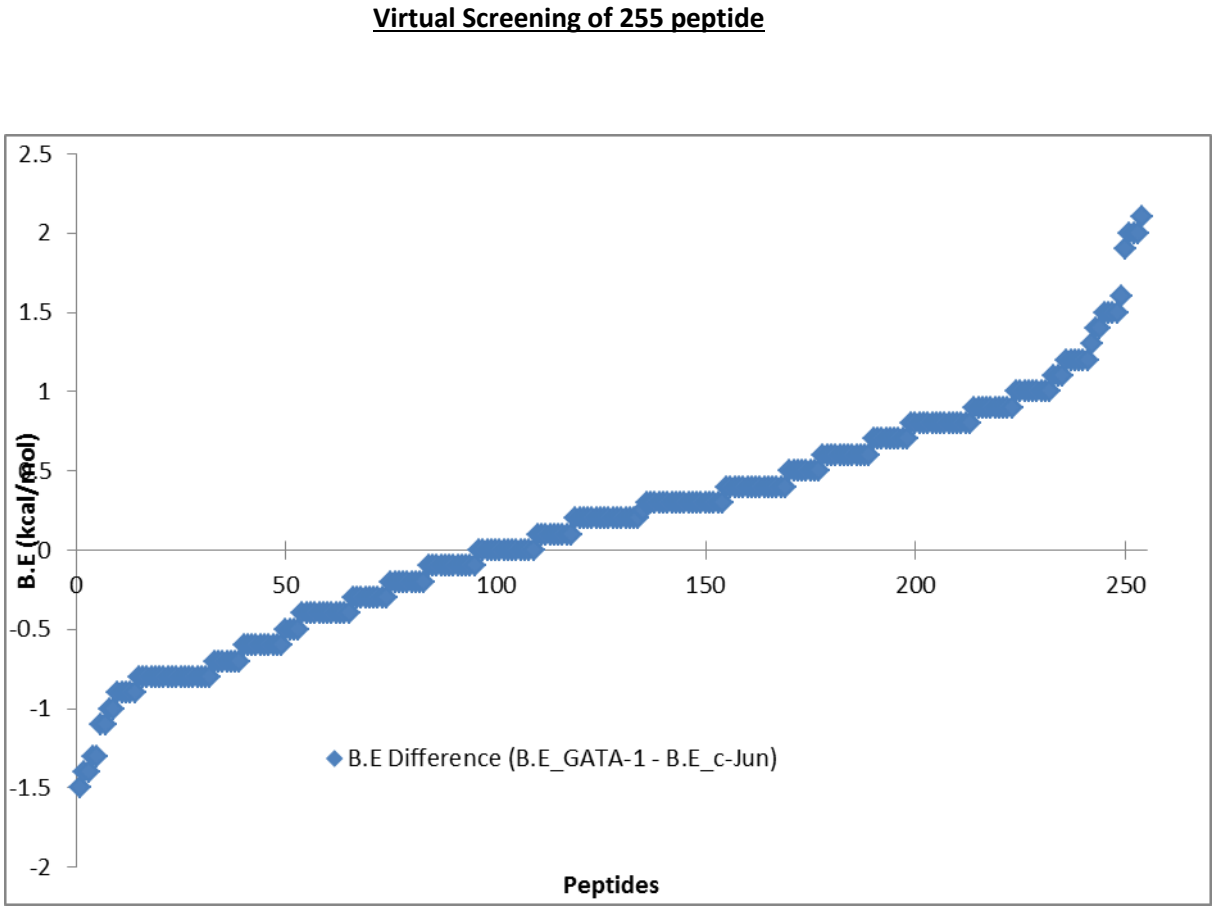


Figure 28: figure shows the difference in binding energy of GATA-1 and c-Jun. Negative value of this difference shows that selected mimetic peptide has high binding affinity towards GATA-1 and less binding affinity towards c-Jun. We then identified those peptides which showed significant difference in binding affinity and then we did colour coding which represent strong and weak interaction.

Table15: Sequences based on the difference between the B.E of GATA-1 minus B.E of c-Jun.

Sequences	B.E Difference (B.E_GATA-1 - B.E_c-Jun)	B.E. (GATA1)	B.E. (cJun)
VKEVYKGKTNGKK	-1.5	-4.5	-3
KYGKNVTEGKVKK	-1.4	-4.4	-3
EYKTGVKKKGVKN	-1.4	-4.4	-3
GGTVKKVKYKENK	-1.3	-4.4	-3.1
KVKVNEKKGYTKG	-1	-4.2	-3.2
VTENKGYKKKVGK	-0.9	-4.5	-3.6
KKGNYETKGKVVK	-0.8	-4.7	-3.9
EGGVKKVNKYKTK	-0.8	-4.6	-3.8
EYGKVKVKTNGKK	-0.8	-4.1	-3.3
VGKKKVKNGYKTE	-0.8	-4	-3.2
YGENKVKKKTKGV	-0.7	-4.8	-4.1
KEGKVKKNVTYGK	-0.7	-4.1	-3.4
KGYKVEKVKKTNG	-0.6	-4	-3.4
VGKKTEGKNKYVK	-0.5	-4.1	-3.6
NYGKTGEVKKVKK	-0.4	-3.7	-3.3
YGETKGVKVKNKK	2.1	-3.1	-5.2

In this Table sequences are arranged based on the difference between the B.E of GATA-1 minus B.E of c-Jun. Colour coding is done to highlight given mimetic peptide residue. Green colour represent those peptide which have highest binding affinity towards GATA-1 and lowest binding affinity towards c-Jun. Here we considered -1 as reference difference. Pink colour shows those residues which have lowest binding affinity towards GATA-1 and Highest Binding Affinity towards c-Jun.

Based on these values we identified peptide <KVKVNEKKGYTKG> as control having high difference between binding energy of GATA-1 and c-Jun i.e. it has lowest binding affinity with c-Jun and highest with GATA-1. So we calculated Contact Map, intermolecular Hydrogen bond distance and Energy calculation.

Table 16: Sequences based on the difference between the B.E of GATA-1 minus B.E of c-Jun

Sequences	B.E Difference (B.E_GATA-1 - B.E_c-Jun)	B.E. (GATA1)	B.E. (cJun)
VKEVYKGKTNGKK	-1.5	-4.5	-3
KYGKNVTEGKVKK	-1.4	-4.4	-3
EYKTGVKKKGVKN	-1.4	-4.4	-3
GGTVKKVKYKENK	-1.3	-4.4	-3.1
KVKVNEKKGYTKG	-1	-4.2	-3.2
NYGKTGEVKKVKK	-0.4	-3.7	-3.3
YGETKGVKVKNKK	2.1	-3.1	-5.2

Here we examined that VKEVYKGKTNGKK mimetic peptide coded in green colour have maximum binding energy difference having highest binding energy (-4.5 kcal/mol) with GATA-1 and lowest binding energy (-3kcal/mol) with c-Jun. that YGETKGVKVKNKK mimetic peptide coded in red colour have minimum binding energy difference having highest binding energy (-3.1 kcal/mol) with GATA-1 and maximum binding energy (-5.2 kcal/mol) with c-Jun.

5.0.10Result analysis of peptide <KVKVNEKKGYTKG>

We found that the **<KVKVNEKKGYTKG>** mimetic peptide shows the lowest binding affinity with cJun such as -3.2 and highest with GATA1 such as -4.2.

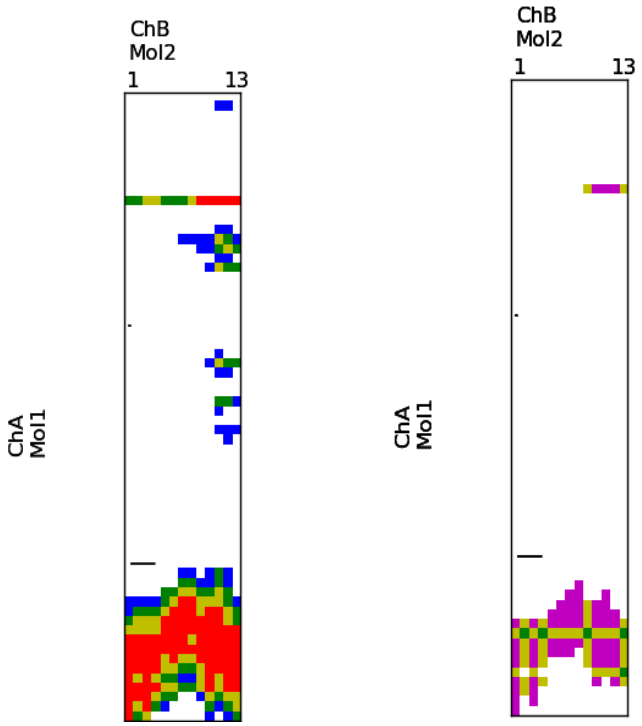


Figure 29: Contact map of **<KVKVNEKKGYTKG>** mimetic peptide of PU.1 and GATA1.

Here pink color represents the Hydrophilic Hydrophilic interaction, yellow color represents Hydrophilic Hydrophobic interactions while green color represents Hydrophobic Hydrophobic interactions.

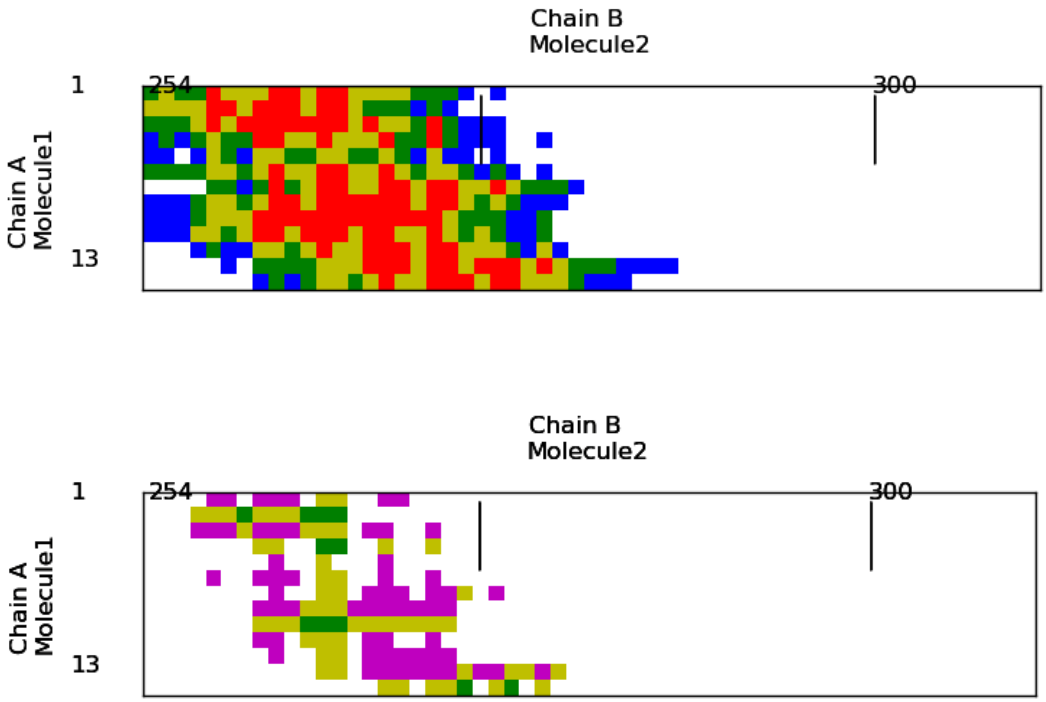


Figure 30: Contact map of **<KVKVNEKKGYTKG>** mimetic peptide of PU.1 and c-Jun.

Here pink color represents the Hydrophilic Hydrophilic interaction; yellow color represents Hydrophilic Hydrophobic interactions while green color represents Hydrophobic Hydrophobic interactions.

Most of the residues of cJun which was found to make hydrogen bonds with **<KVKVNEKKGYTKG>** mimetic peptide of PU.1 were already reported above to make hydrogen bonds with wild type of PU.1 peptide.

Table 17: Hydrogen bonds found in between the **<KVKVNEKKGYTKG>** mimetic peptide with cJun

Sr.No.	**Atom 1**	**Atom 2**	**Distance**
1.	B:ASN262:HD22	:VAL2:O	2.26895
2.	B:SER269:HG	:LYS7:O	2.35886
3.	B:ARG272:HH11	:GLY9:O	2.02258
4.	B:ARG276:HE	:GLY13:O1	2.36528
5.	B:ARG276:HH22	:GLY13:O1	2.34617
6.	B:ASN262:OD1	:LYS1:HZ2	2.46484
7.	B:ALA265:O	:GLY9:H	2.4806
8.	B:ARG261:HE1	:LYS1:O	2.40806
9.	B:ARG261:HH22	:LYS1:O	2.41047
10.	B:ASN262:HD21	:GLU6:O	1.87865
11.	B:LYS268:HZ3	:LYS3:O	2.06537
12.	B:SER269:HG	:LYS8:O	2.21672
13.	B:ARG272:HH11	:VAL4:O	2.22828
14.	B:ARG279:HH21	:TYR10:OH	2.3422

Most of the residues of GATA1 which was found to make hydrogen bonds with **<KVKVNEKKGYTKG>**mimetic peptide of PU.1 were already reported above to make hydrogen bonds with wild type of PU.1 peptide.

Table 18: Hydrogen bonds found in between the **<KVKVNEKKGYTKG>** mimetic peptide with GATA1

Sr.No.	**Atom 1**	**Atom 2**	**Distance**
1.	A:ARG56:H	:LYS7:O	2.33249
2.	A:ARG56:HH11	:GLY9:O	2.24618
3.	A:LYS57:HZ2	:LYS12:O	2.01568
4.	A:SER59:HG	:VAL4:O	2.29119
5.	A:SER59:OG	:VAL4:H	2.07474
6.	A:SER59:O	:LYS3:H	2.46122
7.	A:SER59:OG	:LYS3:H	2.05645
8.	A:LYS61:O	:LYS1:HZ3	2.16458
9.	A:ASN55:OD1	:LYS8:HZ2	2.36427
10.	A:ASN55:HD21	:ASN5:O	2.28073
11.	A:SER59:H	:LYS12:O	1.98396
12.	A:SER59:HG	:LYS12:O	2.28169
13.	A:ASN55:O	:ASN5:HD22	2.24063
14.	A:SER60:O	:TYR10:HH	1.82549

Table 19: Energy calculations of **<KVKVNEKKGYTKG>** mimetic peptide of PU.1 with cJun and GATA1 in complex form

	SASA	**Electrostatic energy (kcal/mol)**	**Vander waals energy (kcal/mol)**	**Dessolvation energy (kcal/mol)**	**Binding energy**
GATA1	7348.3275	2.69188842	-278.635	-0.0230944	-4.2
cJun	7098.984	0	-331.988	0	-3.2

6. DISCUSSION AND FUTURE PERSPECTIVE

PU.1, GATA1 and c-Jun plays an important role in lineage commitment of HSCs into erythroid and myeloid, leading to the depression of diseased state in the cell. Their structural and functional studied leads to the development of novel therapeutic agents that can bind with the GATA1 and c-Jun protein and control their effects. Nevertheless, the formation of scrambled library of 500 compounds and checking them one by one for their effect against GATA1 and c-Jun protein is a stressful task due to time consumption and cost. So we have used mimetic peptides of PU.1 β3/β4 region and identify the level of their binding with GATA1 and c-Jun in Pyrx to obtained novel mimetic peptide which has affinity toward GATA1 and lowest binding affinity towards c-Jun for enhancing the myelopoiesis instead of erythropoiesis. Our study identified <KVKVNEKKGYTKG> mimetic peptide bind with highest binding affinity with GATA1 and with lowest binding affinity with c-Jun.Statistic result predicts effect of mutation from beta strand helix to beta strand and conformational change in peptides leads to high binding affinity with GATA1 and lowest with c-Jun.

Hydrogen bond, bump and RMSD analysis using "DS visualizer". Electrostatic, van der Waals (CHARMm19) and dessolvation energy calculated via "FastContact". Contact maps via "cocomaps", SASA using "Pymol". Intermolecular hydrogen bondings were identified by DS 2.0 and Polar contacts were analysed by Pymol. Energy minimization of all the ligands is done via SPDB Viewer.

Based on screening new low-micromolar inhibitors might be suitable as leads for further inhibitor ion development efforts against GATA1, based on the fact their size and chemical properties are appropriate to classify them as drug like compounds. The models of these protein-peptide complexes found strategies for future development of selective GATA1 inhibitors.

10 REFERENCES

- Behre, G., Whitmarsh, A. J., Coghlan, M. P., Hoang, T., Carpenter, C. L., Zhang, D.-E., Davis, R. J. & Tenen, D. G. (1999) *J. Biol. Chem.* **274,** 4939–4946

- Burda P, Curik N, Kokavec J, Basova P, Mikulenkova D, Skoultchi AI et al. PU.1 activation relieves GATA-1-mediated repression of Cebpa and Cbfb during leukemia differentiation. Mol Cancer Res 2009; 7: 1693–1703

- Lee, M. E., Temizer, D. H., Clifford, J. A. & Quertermous, T. (1991) *J. Biol. Chem.* **266,** 16188–16192.

- Liew CW, Rand KD, Simpson RJ, Yung WW, Mansfield RE, Crossley M, Proetorius-Ibba M, Nerlov C, Poulsen FM, Mackay JP.Gutierrez-Hartman A, Duval DL, Bradford AP (2007). "ETS transcription factors in endocrine systems". *Trends Endocrinol Metab* **18** (4): 150–8

- Li Z, Godinho FJ, Klusmann JH, Garriga-Canut M, Yu C, Orkin SH. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. Nat Genet 2005; 37: 613–619.

- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E (2006). "TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes". *Nucleic Acids Res.* 34

- Moreau-Gachelin F, Tavitian A, Tambourin P. Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. Nature 1988; 331: 277–280

- Moreau-Gachelin, F., Wendling, F., Molina, T., Denis, N., Titeux, M., Grimber, G., Briand, P., Vainchenker, W. & Tavitian, A. (1996) *Mol. Cell. Biol.* **16,** 2453–2463.

- Moreau-Gachelin F, Wendling F, Molina T, Denis N, Titeux M, Grimber G et al. Spi-1/PU.1 transgenic mice develop multistep erythroleukemias. Mol Cell Biol 1996; 16: 2453–2463.

- Nerlov C, Querfurth E, Kulessa H, Graf T. GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1- dependent transcription. Blood 2000; 95: 2543–2551.

- Papetti M, Skoultchi AI. Reprogramming leukemia cells to terminal differentiation and growth arrest by RNA interference of PU.1. Mol Cancer Res 2007; 5: 1053–106

- P Burda, P Laslo and T Stopka The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. Leukemia (2010) 24, 1249–1257;

- Petrovick, M. S., Hiebert, S. W., Friedman, A. D., Hetherington, C. J., Tenen, D. G. & Zhang, D.-E. (1998) *Mol. Cell. Biol.* **18,**3915–3925.

- Rekhtman N, Choe KS, Matushansky I, Murray S, Stopka T, Skoultchi AI. PU.1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation. Mol Cell Biol 2003; 23: 7460–7474

- Rekhtman N, Radparvar F, Evans T, Skoultchi AI. Direct interaction of hematopoietic transcription factors PU.1 and GATA- 1: functional antagonism in erythroid cells. Genes Dev 1999; 13: 1398–1411.

- Rosenbauer F, Owens BM, Yu L, Tumang JR, Steidl U, Kutok JL et al. Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. Nat Genet 2006; 38: 27–37.

- Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. Nat Genet 2004; 36: 624–630.

- Shimizu R, Kobayashi E, Engel JD, Yamamoto M. Induction of hyperproliferative fetal megakaryopoiesis by an N-terminally truncated GATA1 mutant. Genes Cells 2009; 14: 1119–1131.

- Stegmaier P, Kel AE, Wingender E (2004). "Systematic DNA-binding domain classification of transcription factors". *Genome informatics. International Conference on Genome Informatics* 15 (2): 276–86.

- Stopka T, Amanatullah DF, Papetti M, Skoultchi AI. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. EMBO J 2005; 24: 3712–3723.

- Tenen, D. G., Hromas, R., Licht, J. D. & Zhang, D.-E. (1997) *Blood* **90,** 489–519

- Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat Genet 2002; 32: 148–152.

- Zhang P, Behre G, Pan J, et al. Negative crosstalk between hematopoietic regulators: GATA proteins repress PU.1. Proc Natl Acad Sci U S A. 1999;96:8705-8710.

- Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, Radomska HS et al. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc Natl Acad Sci USA 1999; 96: 8705–8710.

- Zhang P, Zhang X, Iwama A, Yu C, Smith KA, Mueller BU et al. PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood 2000; 96: 2641–2648