Chapter 1:

Integrative Bioinformatics Study of KISS1 in Oncogenesis and Metastatic Pathways

INTRODUCTION

Cancer is a multifaceted and heterogeneous, which is characterised by progressive genetic, epigenetic, and phenotypic alterations that enable cells to acquire abnormal growth and survival advantages (Akshata Desai, 2012; Fultang et al., 2021). Despite remarkable advances in early detection and therapeutic strategies, cancer remains one of the leading causes of mortality worldwide. The complexity of cancer not only lies in uncontrolled cell proliferation but also in the dynamic reprogramming of cellular behaviour that allows tumour cells to adapt, survive, and disseminate.

The conceptual framework which was proposed by Hanahan and Weinberg, known as the Hallmarks of Cancer, provides a comprehensive model to understand cancer progression. As initially described in 2000 and later expanded in 2011 and 2022, these hallmarks include sustained proliferative signalling, evasion of growth suppressors, resistance to cell death, replicative immortality, induction of angiogenesis, activation of invasion and metastasis, deregulated metabolism, immune evasion, genome instability, and tumour-promoting inflammation. Hence, among these hallmarks, activation of invasion and metastasis represents the most lethal feature of cancer, as metastatic disease is responsible for the vast majority of cancer-related deaths (Hanahan & Weinberg, 2011).

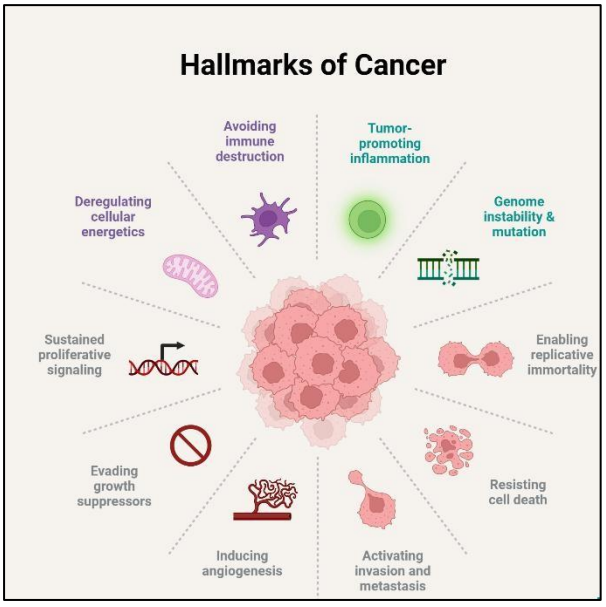


Figure 1.1: *Different Hallmarks of Cancer as stated by Hanahan and Weinberg, 2011(Image Rendered using BioRender)*

Metastasis is a highly orchestrated and multicellular process that includes epithelial-mesenchymal transitions, degradation of the extracellular matrix, intravasation, survival in the bloodstream, and colonization of other body parts (Luo & Ge, 2020). All such steps are under the regulation of intricate networks of molecules. Eliminating such networks generates occurrences of aggressiveness and resistivity to therapy. It has become essential to focus on the molecules that hinder and promote such events for arriving at effective therapies for cancers (Usman et al., 2021).In this context, metastasis suppressor genes have been identified as key regulatory molecules in the process of cancer progression. Unlike conventional tumor suppressors, metastasis suppressors act by repressing metastatic spread without altering the growth of the primary tumor. Established metastasis suppressors include E-Cadherin, NM23, BRMS1, KAI1, and *KISS1* (Jiang et al., 2005; Prabhu et al., 2013). Among these, the unique aspect of the *KISS1* gene is its mode of action in the metastasis process. The gene predominantly acts at a relatively later stage of the metastasis cascade in preventing the outgrowth of secondary metastases.

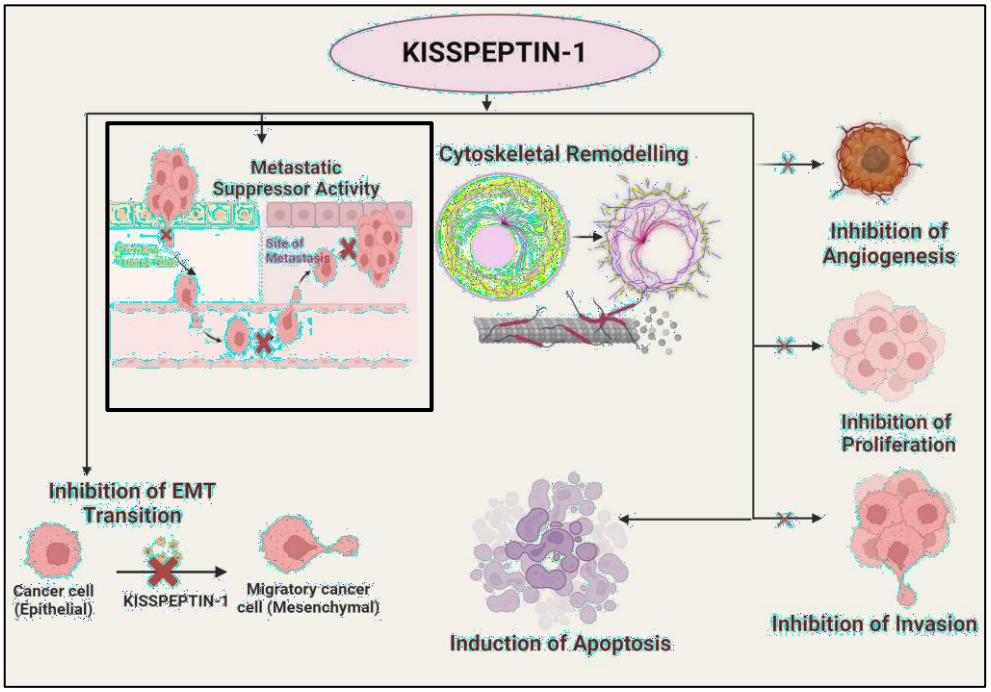


Figure 1.2: *Different Role of* Kisspeptin-1 *in cancers. (Image Rendered using BioRender)*

KISS1 is the parental metastasis suppressor gene that codes for a 145-amino acids long pre-peptide. The proteolysis of this pre-peptide leads to generation of different peptides known as kisspeptins, such as Kisspeptin-54, Kisspeptin-14, Kisspeptin-13, and Kisspeptin-10. (Gottsch et al., 2009; Motti & Meccariello, 2019). Out of all these peptides, Kisspeptin-10 is the shortest biologically active fragment, responsible for activation of *KISS1R*. Originally discovered in studies in reproductive endocrinology, the *KISS1* and *KISS1R* system is currently appreciated to be multifaceted and plays several other roles in physiology and pathological processes that include metabolism, neuronal communication, blood vessel regulation, and cancer biology. Notably in cancer, kisspeptin action has been demonstrated to

affect several features of cancer that have been originally proposed by Hanahan and Weinberg, which include inhibition of invasion and metastasis, regulation of proliferative signaling, control of apoptosis, and angiogenesis. Despite these roles, the function of kisspeptin in cancer is not always consistently tumor suppressive. Under different types of cancer, various expressions, intracellular pathways, and microenvironmental conditions. These scenarios suggest the complex regulatory mechanism involved in the *KISS1/KISS1R* system, thus requiring more intensive study.

Transcriptional regulation is one of the hallmark acquired by cancer cells, and it involves the action of several transcription factors that act like master regulators of various programs such as proliferation, survival, differentiation, metabolism, evasion of the immune system, and occurrence of metastasis. Some of the most prominent factors that are implicated in the oncogenic processes are SP1, c-Myc, CDX2, HDAC2, FLI1, and GATA2 (Cho et al., 2009; Kokura et al., 2001; Ramirez Moreno et al., 2021). Notably, a number of these transcription factors have already been identified to regulate *KISS1* gene expression, while *KISS1* itself mediates the transcriptional activity regulation of these transcription factors in a rather intricate regulatory feedback phenomenon. For instance, SP1 and AP-2α regulate the transcriptional activity of *KISS1*, while c-Myc, CDX2, and FLI1 regulate the proliferative and metabolic, differentiation and metastasis, and the survival and invasiveness processes, respectively, through the mediation of *KISS1*. GATA2, on the other hand, is involved in the regulation of cellular differentiation and transcriptional stability (Cho et al., 2009; Scheiber et al., 2014).

Though quite a lot of experimental evidence has been accumulated concerning the patterns of expression and connections of *KISS1* with transcription factors, its structural and molecular aspects are still poorly understood. The major part of experimental research has been dedicated to correlation analysis of expression patterns and has not been covering an essential gap concerning the molecular interaction of *KISS1* with transcription machinery and DNA(Jones et al., 1998; Shan et al., 2017).

Contributing to this complexity is the fact that *KISS1* is known to be an intrinsically disordered protein. Intrinsic disorder is referred to as the lack of tertiary structure in proteins under physiological conditions. Intrinsically disordered proteins are known to play crucial roles in cellular signaling and the regulation of transcription due to their high adaptability and potential to interact with various target proteins. Such characteristics of *KISS1* not only justify its role as a regulatory protein but also explain why structural analyses of the protein would be difficult. Bioinformatics integrates sequence analysis, structural modelling, binding residue prediction, Protein-DNA docking, and molecular dynamics simulations to explore biomolecular interactions at an atomic resolution. Protein-DNA docking has provided information on the binding affinity, interacting residues, and interaction energies, while MD simulations have revealed time-dependent stability and flexibility, number of hydrogen bonds, and

conformational transitions and energetic landscapes. These approaches are particularly powerful for intrinsically disordered and regulatory proteins such as KISS1.

As such, within the context of cancer biology and hallmarks of cancer, bioinformatics investigation of *KISS1* DNA-Protein interactions presents a first opportunity to decode how transcriptional programs orchestrating metastasis, proliferation, apoptosis, and cellular plasticity are structurally regulated. Understanding these interactions may show how *KISS1* selectively stabilizes tumor-suppressive transcriptional programs while disrupting oncogenic networks. Regarding the therapeutic aspect, it can be proposed that *KISS1* is potentially a switch molecule able to control various characteristics of cancers concurrently. Through regulations in transcription factor levels, it is possible that *KISS1* may control invasion, metastasis, proliferation, apoptosis, as well as chromatin remodelling. The study targeting this regulatory molecule may offer comprehensive management of cancers because it focuses on system biology rather than pathways.

Consequently, research on the mechanism of KISS1 on transcription regulation using modern bioinformatics techniques can help fill the gap that exists between the structure of molecules in cancer research and their function. This is not only beneficial in the understanding of the mechanism of cancer but can also help in the discovery of intervention targets in the regulation of metastasis in cancer.

In summary, cancer progression, as described by Hanahan and Weinberg, is driven by coordinated acquisition of hallmark capabilities that are ultimately controlled by transcriptional and structural regulatory mechanisms. KISS1 emerges as a critical metastasis-suppressor and transcriptional modulator within this framework. A comprehensive bioinformatics analysis of *KISS1* DNA-protein interactions provide an integrated platform to unravel the molecular architecture underlying cancer hallmarks and offers promising avenues for the development of next-generation anti-metastatic therapeutic strategies applicable across diverse cancer types.

HYPOTHESIS

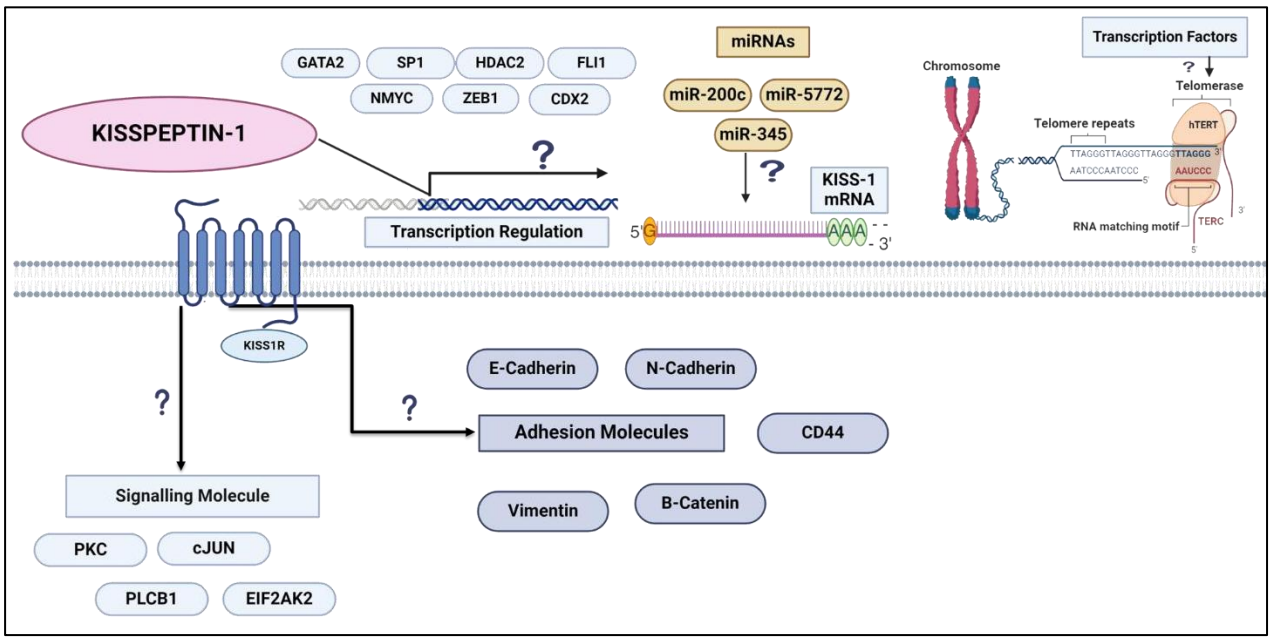


Figure 1.3: *Hypothesis for identifying molecular mechanism underlying with the role of Kisspeptin-1 in various cancers.*

Aggressive cancers may show downregulation of metastasis-associated genes through complex and poorly understood mechanisms. We hypothesized that the activation of Kisspeptin-1 (KISS1) upregulates the metastasis-suppressing pathways by influencing transcription factors, signaling molecules, adhesion molecules, and microRNAs. The anti-metastatic action of Kisspeptin-1 may act through transcriptional and post-transcriptional pathways.

MATERIALS AND METHODS

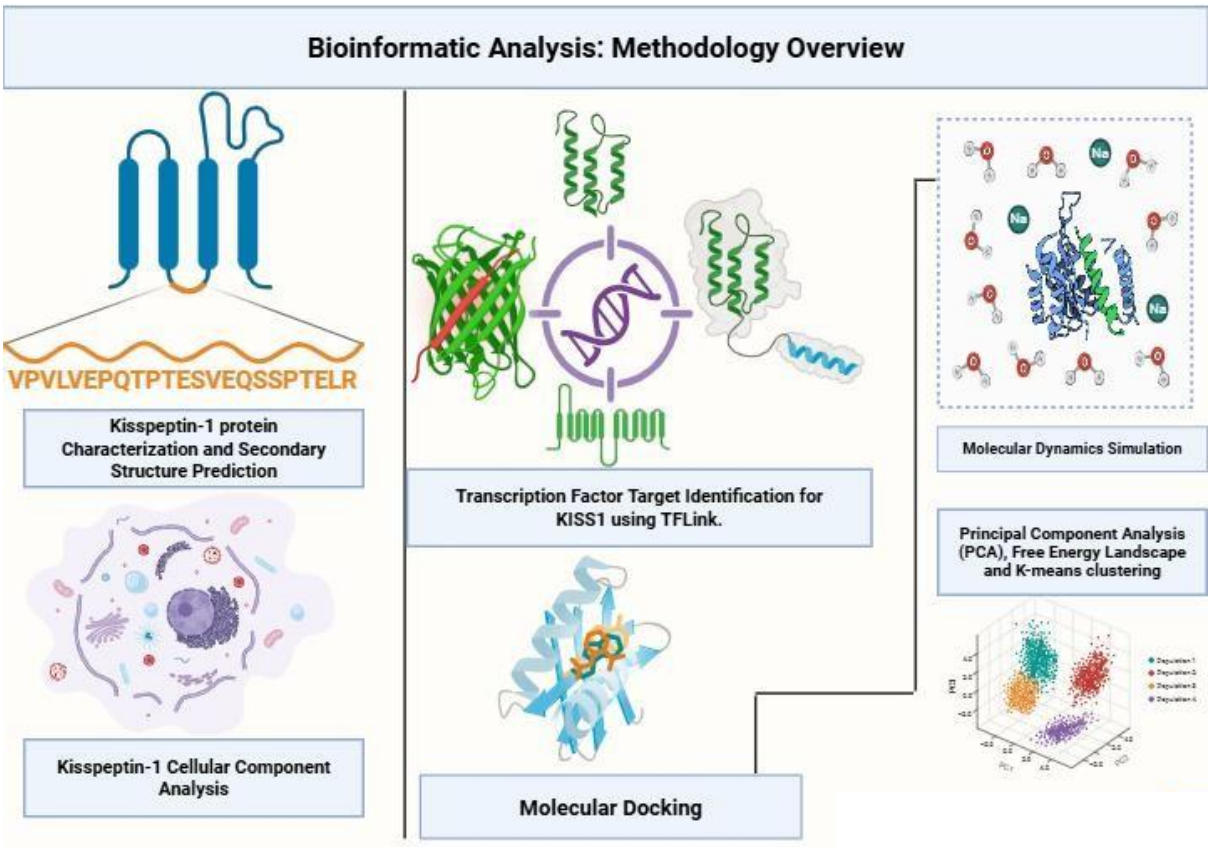


***Figure 1.4**: Overall Methodology involved with Bioinformatics Analysis for the role of Kisspeptin-1 in varied cancers.*

1.3.1 in-silico Analysis for nucleotide and protein sequence of KISS1

The nucleotide sequence of the human gene *KISS1* was isolated from the National Centre for Biotechnology Information (NCBI) database using the accession number NG_032151.1. The DNA sequence obtained was translated to its corresponding amino acid sequence utilizing the ExPASy Translate Tool. The translated amino acid sequences were further analysed to determine the correct open reading frame (ORF) by utilizing the NCBI ORF Finder to determine the correct region coding for the protein.

The presence of signal peptides in the KISS1 protein sequence was also analyzed using the SignalP 5.0 online tool, which predicts signal peptide sequences and cleavage sites using neural network models. This was critical in understanding whether the protein undergoes a secretory pathway or an intracellular pathway.

1.3.2 Physicochemical Characterization of KISS1 Protein

The physicochemical properties for the KISS1 protein were calculated using the tool named ProtParam available on the ExPASy server (https://www.expasy.org/). The parameters calculated include the number of each amino acid, molecular weight, theoretical isoelectric point (pI), number of positively and negatively charged residues, the instability index, aliphatic index, and the Grand Average of Hydropathicity (GRAVY). The calculated physicochemical properties were useful for determining the stability, solubility, as well as biochemical properties of the KISS1 protein.

1.3.3 Secondary Structure Prediction and Subcellular Localization

The secondary structure elements for the KISS1 protein were predicted using the PSIPRED server (https://bioinf.cs.ucl.ac.uk/psipred/). The server predicts alpha-helix, beta-strands, and coils for proteins using profiles generated by PSI-BLAST. The server offered insights to the structure and flexibility exhibited by the protein.

The PROTTER (https://protter.ethz.ch/start/) online tool was employed to display the predicted signal peptide and structural characteristics of the KISS1 protein. The subcellular location of the KISS1 protein was predicted using the WoLF PSORT online tool (https://wolfpsort.hgc.jp/). It predicts the possible cellular compartments of proteins based on amino acid composition and sequence motifs.

1.3.4 Identification of KISS1-Associated Transcription Factors

For finding associated transcription factors for the *KISS1* gene in Homo sapiens, TFLink gateway was employed. TFLink is a database of interaction data for pairs of transcription factors and target genes, limited to interactions that are confirmed through laboratory experiments. Network graphing made possible through TFLink was helpful for picking out important transcription factors having possible regulations at the *KISS1* gene.

1.3.5 Protein 3D Structure Prediction, Refinement, and Validation

Protein sequences of selected transcription factors and KISS1 were retrieved from the UniProt database with the following accession numbers:

- KISS1: Q15726
- CDX2: Q99626
- SP1: P08047
- FLI1: Q01543
- HDAC2: Q92769
- GATA2: P15976

- NMYC: P04198

Three-dimensional structures of these proteins were generated using the SWISS-MODEL homology modeling server. The *KISS1* DNA sequence was retrieved from NCBI and modeled using BIOVIA Discovery Studio.

The generated protein models were further refined using the GalaxyRefine server to improve side-chain conformations and overall structural quality (https://galaxy.seoklab.org). Model validation was performed using the PROCHECK server, where Ramachandran plot analysis was used to assess stereochemical quality. The percentage of residues in favored, allowed, and disallowed regions was recorded to ensure structural reliability (https://saves.mbi.ucla.edu/).

1.3.6 Binding Residue Prediction and Molecular Docking

The binding residues involved in KISS1-transcription factor interactions were predicted using COACH, which predicts binding site residues based on multiple algorithms. The predicted residues were active sites in the docking process.

The protein-DNA docking calculations were carried out using the HADDOCK v2.4 web server (https://wenmr.science.uu.nl/haddock2.4/). Docking calculations were done for *KISS1* DNA and all the transcription factors SP1, HDAC2, CDX2, FLI1, GATA2, and MYCN separately. HADDOCK clustering calculations were done based on the scores of dockings, van der Waals energy, electrostatic energy, desolvation energy, and buried surface area. The top-scoring cluster for all complexes was taken for further molecular dynamics studies.

1.3.7 Molecular Dynamics Simulation

The simulations were carried out using the Desmond v4.3 simulation package for molecular dynamics (MD). The docked complex was embedded inside an orthorhombic simulation box, with explicit SPC water molecules. The simulation box was expanded by 10 Å around the outermost atoms of the docked complex. To counter the system and replicate the biological ionic conditions, Na^+ and Cl^- ions were added to the system to obtain the final ionic strength of 0.15 M. The energy minimization was done using the OPLS3 force field to eliminate the steric clashes.

The production phase of the simulation was done for 150 ns in the NPT ensemble, controlling the temperature at 300K using the Nose Hoover chain thermostat and the pressure at 1 atm using the Martyna-Tobias-Klein barostat. All other parameters will retain the default setting.

1.3.8 MD Trajectory Analysis

RMSD, RMSF, and Radius of Gyration

Trajectory analysis was carried out using the Simulation Interaction Diagram in the Desmond tool. Root mean square deviation (RMSD) was employed to check the structural stability of the molecule, whereas root mean square fluctuation (RMSF) values were used for residue-level flexibility.

Hydrogen bond analysis was performed by utilizing a tool called Visual Molecular Dynamics (VMD). The data file (.out.cms) and trajectory file (.dtr) obtained from running simulations were imported into VMD. Hydrogen bond analysis was performed by using a plugin available in VMD with the required cut-off distances and angles. Data regarding hydrogen bond retention at all simulation steps was collected.

The radius of gyration (Rg) analysis was carried out using a Python script written specifically for this purpose, using the libraries ProDy, NumPy, and Matplotlib. The trajectory file (.dcd), as well as the structure file (.pdb), generated using VMD, was processed for obtaining the values of Rg for each frame, which were then written into CSV files.

1.3.9 Principal Component Analysis, Free Energy Landscape, and Clustering

Principal Component Analysis (PCA) was carried out using ProDy v2.0 based on C-alpha atom trajectories. The first two principal components (PC1 and PC2) were used to visualize dominant conformational motions.

Free Energy Landscapes (FEL) were generated by applying Boltzmann inversion to PC1–PC2 projections. Low-energy conformations were represented in blue, while high-energy unstable conformations were represented in yellow.

k-means clustering was performed on PCA data to identify dominant conformational clusters. Each cluster represented a unique structural population sampled during the MD simulation.

RESULTS

1.4.1 in-silico Characterization of KISS1

The human gene *KISS1* was traced using the NCBI nucleotide database with accession number NG_032151.1. The genomic fragment that was completely sequenced was 726 bp long, with an ORF of 453 bp, which codes for a polypeptide of 131 amino acids. The existence of an ORF within this gene confirms that it is a protein-coding gene with possible biological relevance. (Figure 1.5a) Analysis of the signal peptide suggested the possible presence of a signal sequence from residues 1-20, which could mediate secretory or membrane functions. Additionally, there was the conserved cleavage/amidation motif -GKR- found at positions 83-85, which suggests possible post-translational modification into bioactive peptides as in the case of Kisspeptin peptides. Despite the ORF representing 131 amino acids, the length used in the processing, based on variations in annotation, was 138 amino acids, as determined by the usage in the analysis of the ProtParam tool. (Figure 1.5b)

a

b

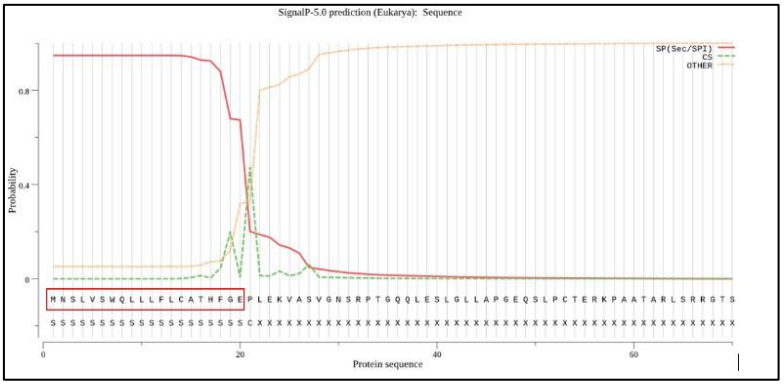


Figure 1.5: *(a) The nucleotide sequence of* KISS1 *and its in silico translation. The putative signal peptide is underlined and conserved Kisspeptin-10 is represented in rectangular box. (b) Predicted signal peptide in the amino acid sequence of* KISS1 *protein*

1.4.2 Physicochemical Properties of the KISS1 Protein

The physicochemical characterization of KISS1 was done through the use of ProtParam. The protein has a molecular weight of 14,704.57 Da, with an estimated isoelectric point (pI) of 10.17. This suggests it is a strongly basic protein. This is also further supported in the composition of the protein in relation to charge. The protein has 15 positively charged amino acids (Arg + Lys) against only 9 negatively charged amino acids (Asp + Glu), suggesting it has high affinity for interaction with negatively charged molecules like DNA or phospholipid biologic membranes. The atomic composition for this protein was calculated to be C641H1024N198O194S3, which largely depicts a sulfur-deficient composition, implying that there are few cysteine residues. The paucity of cysteine residues is supported by the extinction coefficients calculated for cystine reduction and cystine oxidation (12,490 and 12,615 M^-1cm^-1, respectively), which are rather low and imply that disulfide bonds are not largely responsible for protein structure. The absorbance at 0.1% concentration was estimated to be in the range of 0.849-0.858, ensuring a moderate UV absorbance (Table 1.1).

Table 1.1: Physicochemical properties of KISS1 protein.

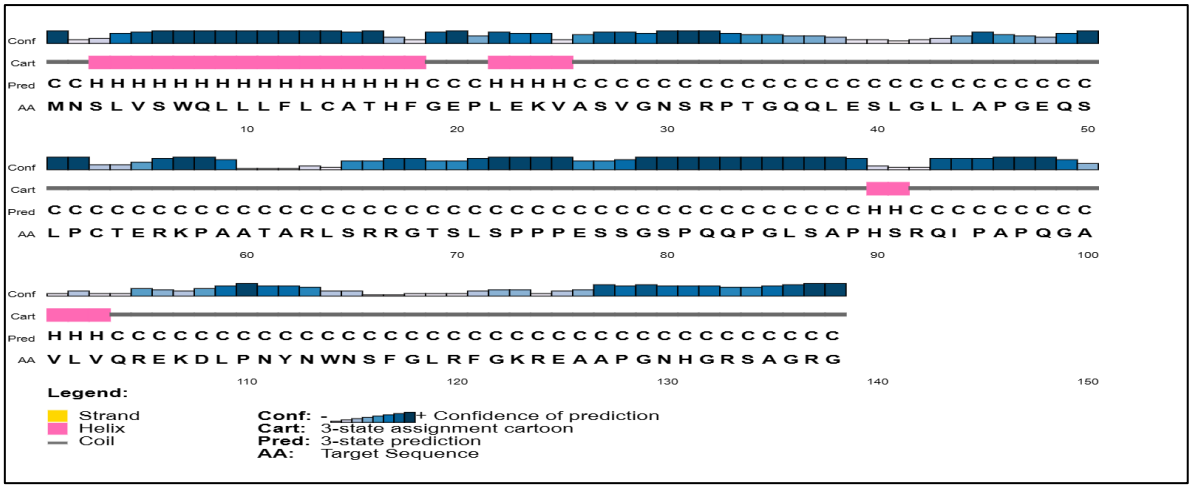
Property	Value
Number of amino acids	138
Molecular weight	14704.57
Theoretical pI	10.17
Total negatively charged residues (Asp + Glu)	9
Total positively charged residues (Arg + Lys)	15
Atomic composition (C, H, N, O, S)	C641H1024N198O194S3
Extinction coefficient (assuming cystines)	12615
Abs 0.1% (=1 g/l)	0.858
Extinction coefficient (assuming reduced Cys)	12490
Abs 0.1% (=1 g/l)	0.849
Estimated half-life (mammalian reticulocytes)	30 hours
Estimated half-life (yeast)	>20 hours
Estimated half-life (Escherichia coli)	>10 hours
Instability index	60.57
Aliphatic index	70.07
Grand average of hydropathicity (GRAVY)	-0.567

The half-life predicted for the KISS1 protein was considerably over 30 hours for mammalian reticulocytes, over 20 hours for yeast, and over 10 hours for *Escherichia coli*, indicating stability within

both prokaryotic and eukaryotic cells, with the suggestion that it does not turnover easily once translated. Despite the correct estimation of the half-life of this protein, the Instability Index was revealed to be 60.57; therefore, the protein can be said to be unstable in *in vitro* conditions. The aliphatic index of 70.07 and the negative GRAVY value of -0.567 represent moderate thermal stability and hydrophilic and flexible nature of the KISS1 protein. Since the Kisspeptin-1 protein has been revealed to be disordered in nature and the Instability Index value clearly reveals it to be unstable in *in vitro* conditions, it can be said that this protein may display disordered behaviour. PSIPRED predictive software used for secondary structure predictions for which for a protein of 138 amino acids, the sequence possessed a helix of 19%, a random coil portion that made it an intrinsically disordered protein with a random coil portion that provided conformational flexibility (Buchan et al., 2013) The software PROTTER can also analyze a signal peptide that spans from amino acid positions 1 to 20 and post-translational modifications Y112 and F121 with variants and site numbers 36, 81, and 115.

1.4.3 Secondary Structure Prediction and Intrinsic Disorder Analysis

a



b

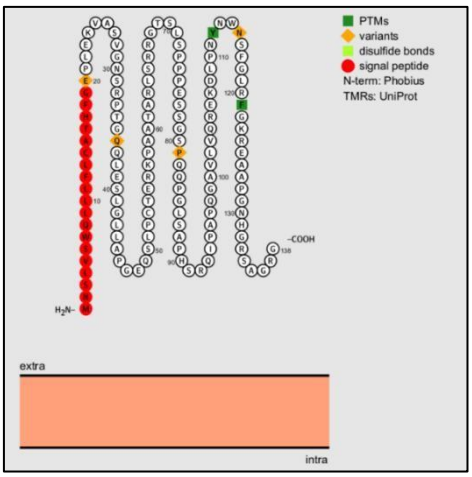


***Figure 1.6**: (a)Secondary structure prediction using PSIPRED software indicating the helical regions, random coils, and motifs in the structure of* KISS1 *protein. (b) PROTTER visualization for signal peptide sequence, post translational modifications and variants*

The secondary structure prediction of KISS1 using PSIPRED yields a pattern where the random coil dominance is observed at roughly 81% of the total sequence, while α-helical segments were captured by only about 19% residues. Interestingly, β-sheets were highly negligible (Figure 1.6a).

This structural distribution classifies KISS1 as an intrinsically disordered protein. Intrinsically disordered proteins have been associated with regulatory proteins involved in transcription, signaling, and protein-protein interactions. The high random coil area suggests a very high degree of conformational flexibility in KISS1 that structurally accommodates multiple molecular partners upon binding. It is functionally beneficial for a metastasis suppressor protein to be able to dynamically interact with transcription factors, chromatin regulators, and signaling mediators; such is the potential benefit provided by its intrinsic disorder, supporting the multifunctional regulatory role of KISS1 in cancer biology (Figure 1.6b).

1.4.4 Subcellular Localization Prediction

The Wolf PSORT analysis indicated a complex localization factor for KISS1. The PSG score indicated a weak signal peptide, which suggested that KISS1 might not be a classical secreted protein, even though it had a hydrophobic region at the beginning. Von Heijne's rule estimated a moderate probability for processing at residues 21-22.

The ALOM algorithm showed a high possibility of peripheral membrane binding, but no obvious transmembrane helices could be found. This result supports the idea that the *KISS1* gene possibly binds to membranes by a non-trans membranous process, for example, electrostatic bonds.

MTOP showed a preference for an orientation promoting N-terminal localization inside the cell, consistent with the hypothesis of membrane insertion rather than secretion. MITDISC prediction yielded a highly negative score (-5.19) rejecting mitochondrial localization.

In contrast, the analyses of NUCDISC and NNCN strongly supported nuclear localization. NNCN predicted nuclear localization with a reliability of 94.1%, indicating that KISS1 was highly likely to function within the nucleus. Similarly, k-NN classification reinforced dual extracellular and nuclear localization probabilities of 44.4% each, pointing toward multifunctional spatial distribution.

Collectively, these results also illustrate that KISS1 follows a dynamic localization pattern, with predominant nuclear association and potential membrane interaction, with only limited extracellular presence, in accordance with its dual function as a secreted peptide precursor and intracellular regulatory protein (Table 1.2).

Table 1.2: List of parameters with different methods for Subcellular Localization Parameters.

Method	Full Form	Parameter/Score	Value	Explanation
PSG	Protein Signal Peptide Prediction	N-region length	0	Length of the nonpolar (N) region in the signal peptide
		N-region positive charge	0	Number of positively charged amino acids in the N-region
		N-region negative charge	0	Number of negatively charged amino acids in the N-region
		H-region length	19	Length of the hydrophobic (H) region in the signal peptide
		H-region peak value	10.30	Hydrophobicity score at the peak of the H-region
		PSG score	5.90	Prediction score for the presence of a signal peptide
GvH	von Heijne's Method for Signal Sequence Recognition	GvH score (threshold: -2.1)	-0.59	Score indicating the likelihood of a signal sequence according to GvH
		Possible cleavage site	21-22	Position where the signal peptide is predicted to be cleaved
ALOM	Amino Acid Limitation Organelle Membrane	Initial position for calculation	22	Position from which transmembrane segments (TMSs) are calculated
		Tentative number of TMSs (threshold 0.5)	0	Number of tentative transmembrane segments based on a certain threshold
		Number of TMSs (fixed)	0	Final number of transmembrane segments after calculation
		Peripheral Likelihood	10.82 (at 38)	Likelihood of a peripheral membrane protein based on ALOM
		ALOM score	10.82	Prediction score for transmembrane helices based on ALOM

MTOP	Membrane Topology Prediction	Center position for calculation	10	Position from which membrane topology is calculated
		Charge difference	-1.5	Difference in charge between cytoplasmic (C) and non-cytoplasmic (N) sides
		N >= C: N-terminal side will be inside	Inside	Prediction of whether the N-terminal side of the protein will be inside
MITDISC	Mitochondrial Targeting Sequence Discrimination	R content	0	Content of arginine (R) residues in the protein sequence
		Hyd Moment (75)	4.91	Hydrophobic moment score at the 75th percentile
		Hyd Moment (95)	2.25	Hydrophobic moment score at the 95th percentile
		G content	1	Content of glycine (G) residues in the protein sequence
		D/E content	1	Content of aspartic acid (D) and glutamic acid (E) residues
		S/T content	3	Content of serine (S) and threonine (T) residues
		Score	-5.19	Prediction score for mitochondrial targeting sequence discrimination
Gavel	Prediction of Cleavage Sites for Mitochondrial Presequence	Cleavage site motif	Not found	Presence or absence of a specific motif for mitochondrial presequence
NUCDISC	Nuclear Localization Signal Discrimination	Content of basic residues	10.9%	Percentage of basic residues (lysine, arginine, histidine) in the sequence
		NLS Score	-0.47	Prediction score for nuclear localization signal presence
NNCN	Cytoplasmic/Nuclear Discrimination	Prediction	Nuclear	Prediction of cytoplasmic/nuclear

				localization based on Reinhardt's method
		Reliability	94.1	Confidence level of the NNCN prediction
COIL	Coiled-Coil Region Detection	Total coiled-coil residues	0	Number of residues predicted to form coiled-coil structures
k-NN Prediction	k-Nearest Neighbors Prediction	k = 9/23		Parameters used for the k-nearest neighbors (k-NN) prediction
		Extracellular	44.4%	Percentage likelihood of extracellular localization
		Nuclear	44.4%	Percentage likelihood of nuclear localization
		Mitochondrial	11.1%	Percentage likelihood of mitochondrial localization

1.4.5 Identification and Modelling of Transcription Factors

Transcription factors that are kisspeptin-associated were revealed using the TFLink gateway, which is a bioinformatics database that brings together the interaction between the transcription factors and their target genes that have been validated through high and low-throughput experiments. The TFLink gateway can therefore efficiently be applied to the construction of maps of the regulation of the genome that is involved in the regulation of Kisspeptin-induced metastasis suppression. This research applied the TFLink gateway to reveal the transcription factors that are associated with the *KISS1* gene in Homo sapiens. Network visualization and filter functions within the TFLink interface facilitated the selection of transcription factors that either directly or indirectly modulate the expression of the *KISS1* gene. On the basis of confidence in interaction data, biological plausibility, and their established role in cancer progression and metastasis, a total of six transcription factors—SP1, NMYC, FLI1, GATA2, HDAC2, and CDX2—were selected for further investigation. These TFs are established as essential modulators of critical biological processes like cell cycle regulation, differentiation, chromatin alterations, migration, and invasions, which are pivotal in cancer metastasis and progression (Figure 1.7).

The identification of these transcription factors allowed for a structured approach to downstream protein-DNA docking analyses and molecular dynamics simulations, to better enable a bioinformatics analysis of the structural/functional rationale in KISS1-regulated transcription in cancer.

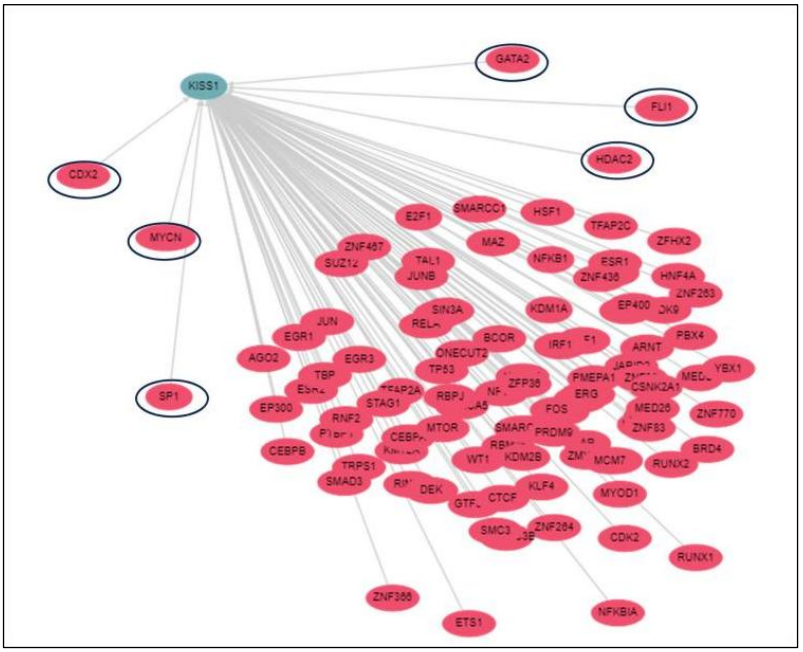


Figure 1.7: *Identification of different transcription factors using TFLink.*

The retrieved sequences of the targeted transcription factors: SP1 , NMYC, CDX2, FLI1,,GATA2 and HDAC2 along with *KISS1* gene were used to make modelled structure, the confidence score for the respective structures were 97%, 97.3%, 100%, 99.5%, 99% and 99.7% in accordance with the number of residues in the most favoured and additionally allowed region after refining were increased to more than 90% for each of the structures and validated using PROCHECK-SAVES server. The binding residues as mentioned were selected for the respective targeted transcription factors which on further docking with *KISS1* gene (Table 1.3).

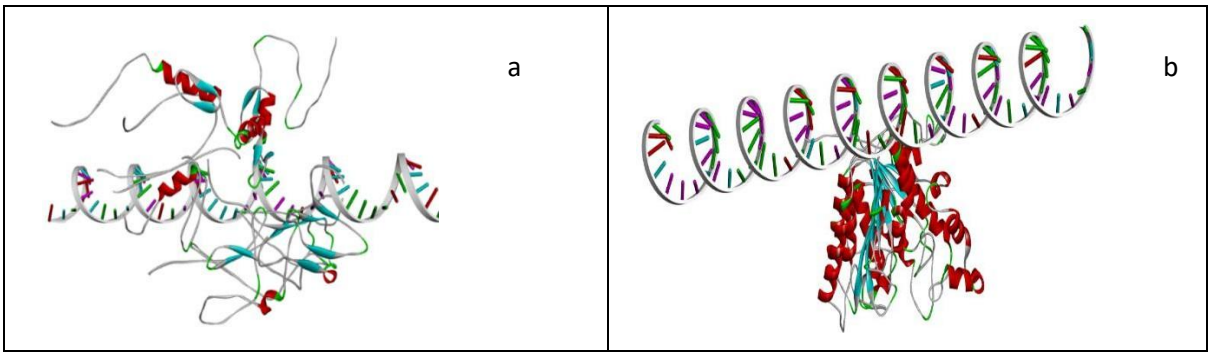
Table 1.3: Different transcription factors along with the modelled confidence scores, binding residues, and PROCHECK score.

Transcription Factor	UniProt ID	Confidence Score (%)	Binding Residues	Favoured + Allowed Regions (%) (PROCHECK)
SP1	P08047	97.0	658, 663, 676, 680	>90%
NMYC	P04198	97.3	265, 363, 367, 394, 395	>90%
CDX2	Q99626	100.0	312, 316, 317, 319, 320, 321	>90%
FLI1	Q01543	99.5	358, 361, 371, 375, 376, 379, 380, 384, 388, 393, 394, 395, 396	>90%

GATA2	P23769	99.0	100, 142, 150, 151, 177, 179, 206, 265, 272, 302, 304	>90%
HDAC2	Q92769	99.7	383, 386, 387, 390, 391, 394, 398, 418, 419	>90%

1.4.6 Protein–DNA Docking of KISS1–TF Complexes

The docking scores of the *KISS1*-transcription factor complexes, the binding affinities in terms of the interaction scores which was calculated using parameters like the Van der Waals energy, electrostatic energy, desolvation energy, and the buried surface area. The *KISS1*-CDX2 complex had the highest favourable binding interaction, thus the lowest HADDOCK score of -144.2 ± 0.0 and a Z-score of -2.6, indicating a highly favourable binding reaction. In addition, the complex had the highest buried surface area of 2526.5 Å², implying a larger surface area of interaction for the two proteins. The *KISS1*-FLI1 complex had the lowest favorable binding interaction, hence the highest HADDOCK score of -99.5 and a positive Z-score of 0.5, indicating a least favorable binding reaction. The other two complexes, namely the *KISS1*-GATA2 and the *KISS1*-HDAC2, had moderate levels of binding affinities, as the HADDOCK scores of the former and the latter complex indicated, at -116.0 and -118.4, respectively. The latter complex had the highest electrostatic energy of -316.4. The other complex had a score of -114.8 and a Z-score of -1.0. The remaining complex had a HADDOCK score of -102.6. (Table 1.4) (Figure 1.8 a-f)



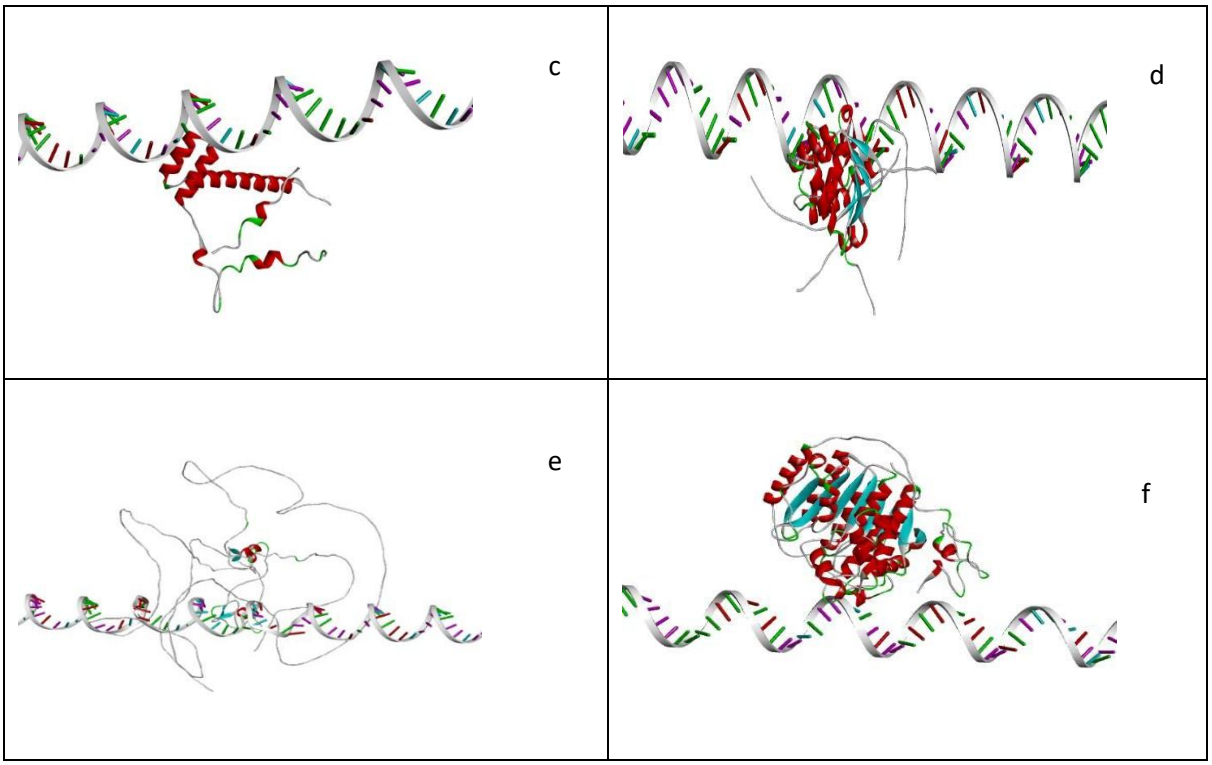


Figure 1.8 (a-f): *Protein-DNA docking analysis of KISS1-DNA structure with different transcriptional factors SP1, NMYC, CDX2, FLI1, GATA2, HDAC2 using HADDOCK server.*

Table 1.4: Protein-DNA docking scores via HADDOCK energies

Docked Complex	HADDOCK Score	Z-score	Van der Waals Energy	Electrostatic Energy	Desolvation Energy	Buried Surface Area (Å²)
KISS1–CDX2	−144.2 ± 0.0	−2.6	−88.0 ± 0.0	−226.1	−11.0	2526.5
KISS1–FLI1	−99.5	0.5	−69.2	−116.6	−7.0	2035.8
KISS1–GATA2	−116.0	−1.9	−74.3	−192.5	−3.3	2108.8
KISS1–HDAC2	−118.4	−2.3	48.5	−316.4	−6.6	2169.7
KISS1–NMYC	−114.8	−1.0	−82.3	−103.5	−11.7	2238.4
KISS1–SP1	−102.6	−2.5	−78.6	−71.4	−12.4	2137.1

1.4.7. Molecular Dynamics Simulations

1.4.7.1 Molecular Dynamics Simulation Analysis of Transcription Factor–*KISS1* Complexes

The dynamic stability, conformation characteristics, and interaction properties of *KISS1* and its six partners (SP1, NMYC, CDX2, FLI1, GATA2, HDAC2) were explored through 150ns-level molecular dynamics simulations. These partners were chosen since *KISS1* is known to interact closely with these transcription factors. The analysis of several structure-related parameters like Root Mean Square Deviation (RMSD), radius of gyration (Rg), occupancy of the H-bond, and root mean square fluctuation (RMSF) has been done.

1.4.7.2 Root Mean Square Deviation (RMSD)

The protein backbone RMSD profiles documented a distinct pattern in the complexes concerning stability. *KISS1*-SP1 complex equilibrated rapidly during the first 20 ns, achieved a mean stabilisation around ~26 Å and continued fluctuating within a relatively narrow range afterwards. This shows a binding conformation that is both stable and persistent.

Similarly, the stabilization of the *KISS1*-NMYC complex was achieved after approximately 30 ns at 28–30 Å; however, during the simulation, some moderate oscillations were obtained that reflected a stable but also dynamically flexible interaction.

In the case of the *KISS*-CDX2*1* complex, a gradual increase in RMSD to ~24–30 Å was observed in the first 30 ns, followed by its stable maintenance until the end of the simulation; this indeed suggests moderate stability with minor conformational adjustments. In contrast, *KISS1*- FLI1 and *KISS1*-GATA2 complexes demonstrated higher RMSD values (~30–32 Å) and larger fluctuation amplitudes, indicating higher conformational mobility and more frequent structural rearrangements. Of all, the RMSD value of the *KISS1*-HDAC2 complex was most erratic, reflecting fluctuations without any tangible plateau over the course of the simulation, thereby reflecting lack of binding stability or lack of stability of binding (Figure 1.9).

Ligand RMSD plots were consistent with the above findings. *KISS1*-SP1 ligand RMSD ranged between 20-24 Å after initial equilibration, with *KISS1*-NMYC ligand RMSD undergoing large fluctuations, with values up to ~52 Å, thereby reflecting ligand instability. The *KISS1*-CDX2 ligand RMSD converged after 80 ns, while that for *KISS1*-HDAC2 ligand was inconsistent.

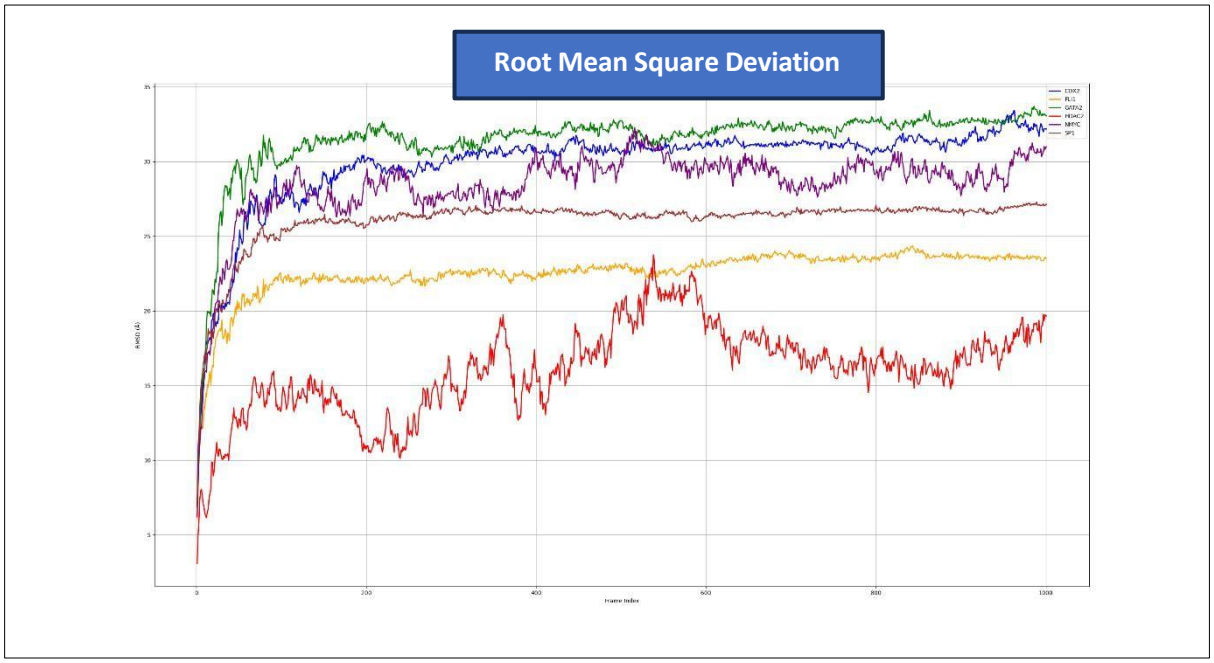


Figure 1.9: RMSD values of docked complex across different frames.

1.4.7.3 Radius of Gyration (Rg)

Rg analysis helped understand the compactness of each complex. *KISS1*-SP1 and *KISS1*-NMYC had low and unchanged Rg values (around 20.1 to 21.0 Å), depicting compact packing. The Rg value for *KISS1*-CDX2 is slightly higher but stable, indicating moderately compact structures. For *KISS1*-FLI1 and *KISS1*-GATA2, the Rg value is higher (~23.5 Å), signifying fewer compact structures. *KISS1* HDAC2 expression levels fluctuated with variable Rg values (Figure 1.10).

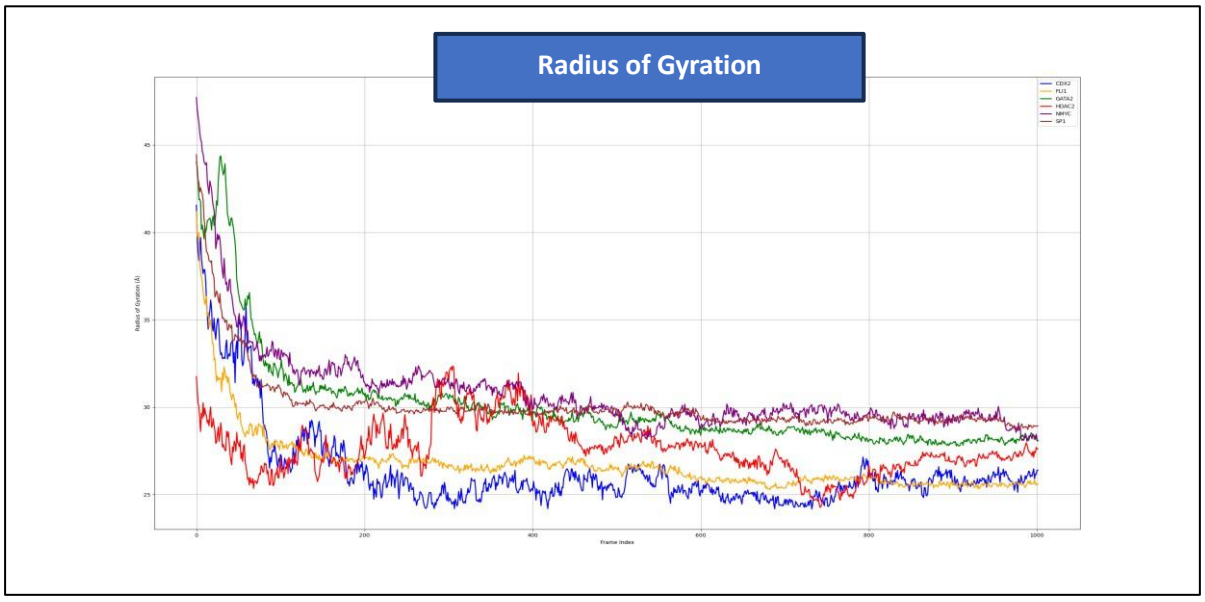


Figure 1.10: Radius of Gyration of all the docked complexes.

1.4.7.4 Hydrogen Bond Analysis

The occupancy of the hydrogen bond plays a crucial role in the stabilization of the interaction. KISS1-NMYC and KISS1- HDAC2 had the maximum average number of hydrogen bonds occupied (7–9 and 6-8 respectively), even though the hydrogen bonds in the former pair were less persistent in time. The pairing of SP1-KISS1 made consistent hydrogen bonds of 5-7, representing a well-balanced interaction. The interaction of CDX2-KISS1 had moderate hydrogen bonds of 4-6, whereas in FLI1-KISS1 and GATA2-KISS1, the number of hydrogens. Such results show how hydrogen bonding plays an unequal role in the stability of different complexes, where bond persistence is given greater importance than its number (Figure 1.11).

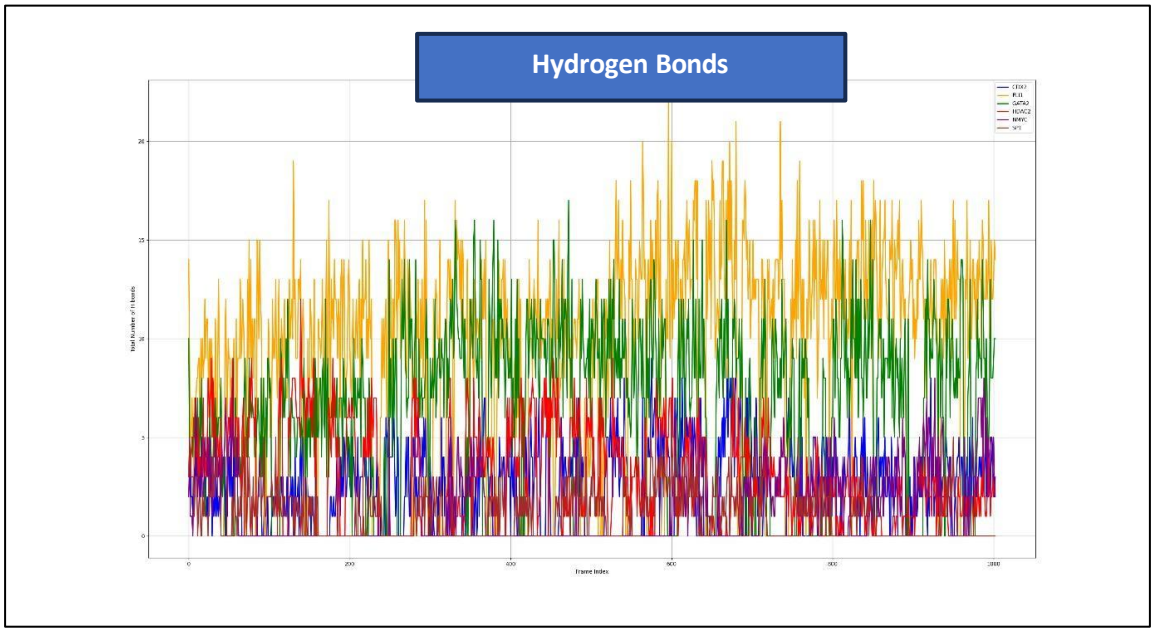


Figure 1.11: Hydrogen Bonds of all the docked complexes showing occupancy.

1.4.7.5 Root Mean Square Fluctuation (RMSF)

Root mean square fluctuation (RMSF) analysis was carried out to assess the flexibility of individual residues in the *KISS1*–transcription factor complexes. The results revealed the existence of variation in the dynamic characteristics of the different transcription factors studied. The RMSF values for the DNA binding residues of SP1, NMYC, and HDAC2 revealed low flexibility with values in the range of 0.8-1.6 Å, implying high rigidity in the structural form. Such rigidity results in the presence of persistent intermolecular interactions that lead to the adoption of a stable transcription factor complex form.

On the other hand, the *KISS1*-FLI1 and *KISS1*-GATA2 complexes had high RMSF scores of approximately 3.2 Å. This can be attributed to a higher level of flexibility within the structures. On the other hand, the *KISS1*- CDX2 complex had a relatively intermediate RMSF score. This can indicate that the complex can act as a stable binder while at the same time displaying a high level of malleability that can make it adaptive (Figure 1.12).

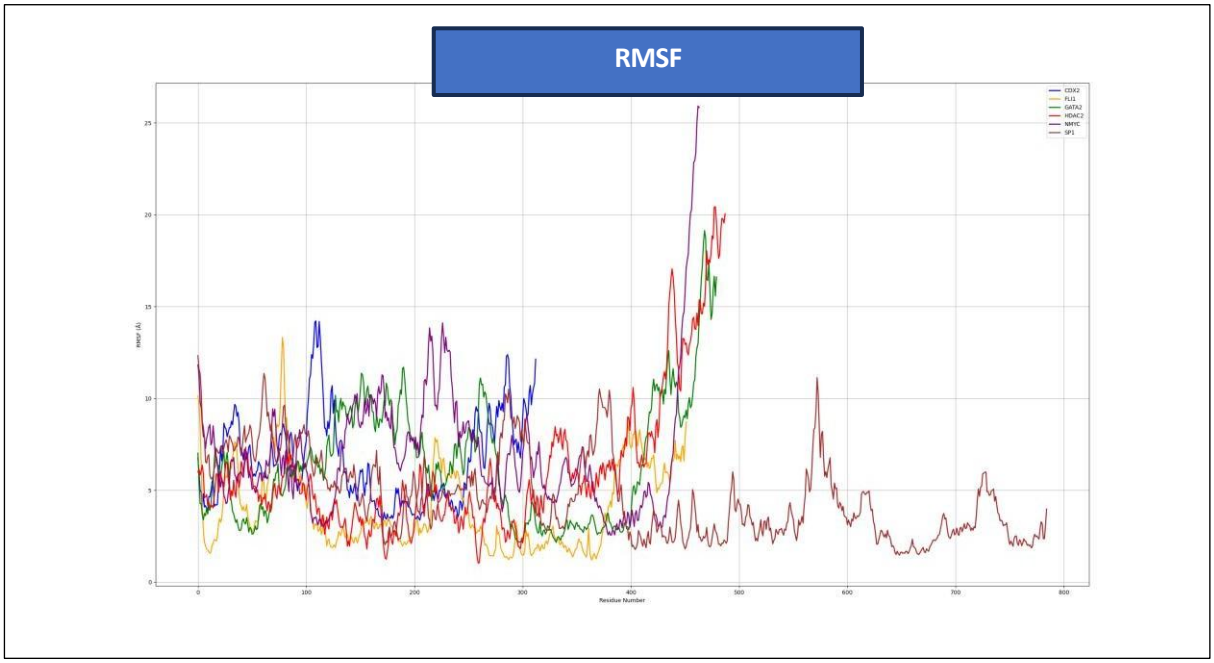


Figure 1.12: *RMSF for all the docked complex across residues numbers.*

1.4.7.6 Post-Simulation Protein–Ligand RMSD Behavior

After full equilibration, FLI1–KISS1 and GATA2–KISS1 displayed stable ligand RMSD profiles, suggesting strong binding once conformational adjustments were completed. CDX2–KISS1 also stabilized after initial rearrangements.

Conversely, HDAC2–KISS1 and NMYC–KISS1 complexes showed declining or fluctuating ligand RMSD without clear stabilization, indicating weak and inconsistent binding behaviour.

SP1–KISS1 maintained moderate ligand RMSD stability throughout the simulation.

1.4.8 Principal Component Analysis of Transcription Factor–KISS1 Complexes

Principal Component Analysis (PCA), also known as essential dynamics analysis, was carried out for 150-ns molecular dynamic paths of all six-transcription factor/KISS1 protein complexes (SP1, NMYC, CDX2, FLI1, GATA2, and HDAC2) to analyze their principal dynamic modes. PCA is a method of decomposing a multi-dimensional matrix of atomic fluctuations into a limited number of orthogonal eigenvectors or principal components. This helps in understanding dynamic modes.

The primary contributors in all systems were the first two principal components (PC1 and PC2), and hence, these two components were used in the conformational space and sampling of all complexes (Figure 1.13 a-f).

a. SP1–KISS1 Complex

In the PCA analysis of the SP1-KISS1 complex, it was found that its projection onto PC1 was predominantly elongated and had little spread along PC2. This implies that the main movement of the complex is along a single collective vector, and its variation along other perpendicular vectors is small.

This indicates limited conformational freedom with high structural stability over the simulation period. The lack of distantly separated clusters further implies that the SP1-KISS1 complex is confined to one conformational basin with minimal fluctuations about the equilibrium state. This finding is consistent with the RMSD and K-means analysis that identified the SP1-KISS1 complex as structurally stable with high dynamic restraint.

The small amplitude of essential motions implies that the interaction between SP1 and KISS1 is mediated through a rigid and well-optimized binding interface.

b. NMYC–KISS1 Complex

However, for the NMYC-KISS1 complex, the data had a wide and continuous distribution over PC1 and PC2, taking the shape of a U-distribution. This is indicative of a full sampling of conformations without settling on one particular state.

This lack of distinct clusters implies that NMYC/KISS1 undertakes a gradual transition among a series of closely related structural configurations rather than transitioning between well-separated structural states. This is a characteristic of high flexibility and dynamic instability.

It also corresponds with the RMSD and Clustering Analysis that showed variability in the ligand placement and low stabilization of the NMYC-KISS1 interface. The result of PCA, therefore, emphasizes NMYC as the most dynamic and plastic transcription factor among the proteins examined that can adapt various structural conformations even while bound with KISS1.

c. CDX2–KISS1 Complex

The CDX2-KISS1 data set in the PCA plot showed a semi-compact distribution that had one predominant conformational cluster together with a second smaller cluster. These clusters are slightly differentiated along the first principal component, which indicated the coexistence of two metastable conformational states.

The major cluster reflected the most prominent fraction of simulation frames and hence appears as the most favoured structure of the protein complex. The second cluster might reflect an alternative binding mode of proteins, which temporarily exists within the protein interface region.

This implies that the flexibility of CDX2-KISS1 is intermediate, as there is structural integrity with the ability for occasional transitions. This is based on the flexibility profile, which is a balance for rigidity and flexibility, as expected from CDX2's strong docking affinity with moderate MD stability.

d. FLI1–KISS1 Complex

FLI1–KISS1 complex presents a quite compact PCA distribution close to the origin with little spread both along PC1 and PC2. This profile is characteristic of highly restricted collective motion consistent with strong structural rigidity.

Confining the conformations within narrow areas suggests that FLI1–KISS1 remains locked in one dominant conformational state for most of the simulation time. That rigidity supports the observation of stable values in the calculation of ligand RMSD and low RMSF in residues in the interface.

While the earlier clustering results suggested multiple local conformations, PCA indicated that such variations fall under a very small essential dynamic space, hence suggesting micro-adjustments rather than transitions. Thus, PCA identifies FLI1–KISS1 as one of the most conformationally stable and rigid complexes among all systems studied.

e. GATA2–KISS1 Complex

On projecting the PCA of the GATA2/KISS1 complex, it was found to have a significantly scattered distribution in both PC1 and PC2. This greatly implies that a vast range of conformational space is covered and that no particular structural well is favoured.

This highly diffusive pattern is indicative of high plasticity and flexibility. GATA2-KISS1 is constantly switching between different conformations with low barriers between the states.

Such an interaction promotes K-means clustering and FEL findings, which suggested the existence of shallow energy minima for this complex. The PCA results above describe the GATA2-KISS1 complex as a dynamically adaptable and structurally unstable interaction.

The ability that this provides for GATA2 to potentially control KISS1 gene transcription through multiple binding modes could be context-dependent.

f. HDAC2–KISS1 Complex

The PCA plot for the HDAC2-KISS1 complex showed that there are two well-separated regions in the first principal component, which implies that the complex exists in two metastable conformational states. There is a clear separation between the two clusters, which suggests that the complex prefers to switch between two favourable structures.

The observed bimodal distribution is indicative of the conformation switch of the HDAC2–KISS1 complex between two preferred binding conformations. The coexistence of two clusters validates the

existence of fluctuation in the RMSD values of the complex, which further confirmed the presence of two energy wells in the FEL analysis.

The results of PCA suggest that the flexibility of the HDAC2–KISS1 system is moderate with defined transitions and not diffusion, as required for its enzymatic function as a regulator.

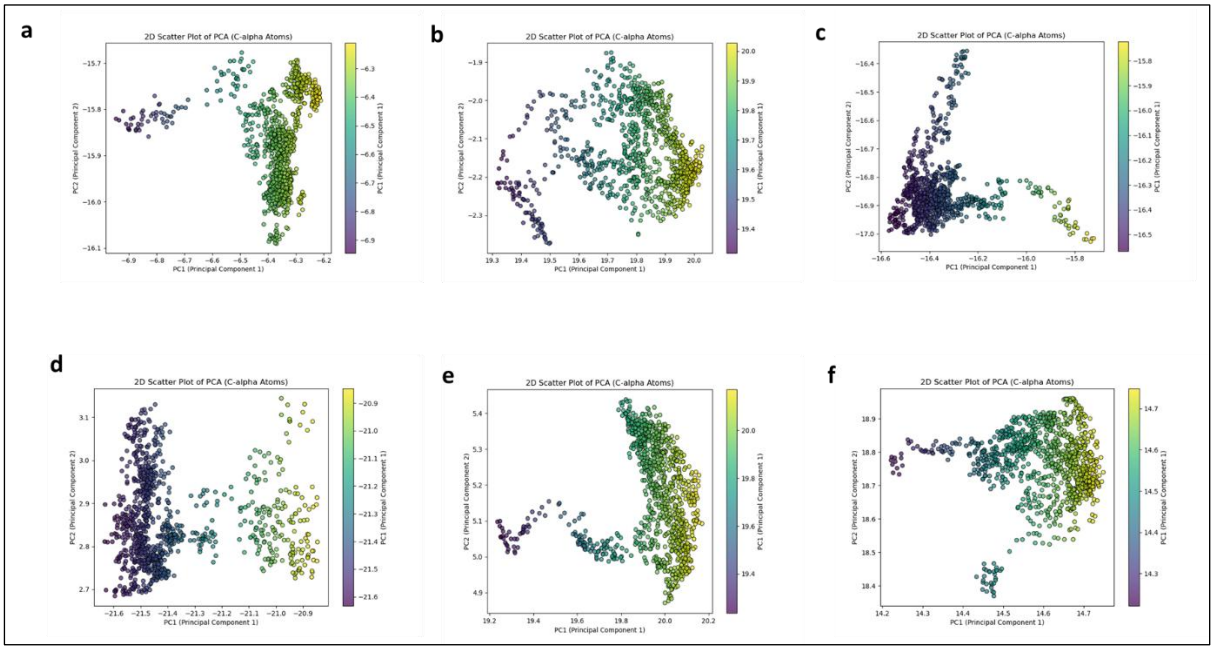


Figure 1.13: *PCA for C-alpha atoms for the following docked complexes in order: a) CDX2-KISS1, b) FLI1-KISS1, c) GATA2-KISS1, d) HDAC2-KISS1, e) NMYC-KISS1 and f) SP1-KISS1.*

1.4.9 Free Energy Landscape Analysis of Transcription Factor–KISS1 Complexes

For characterization of the thermodynamic stability and favored conformations of the transcription factor-KISS1 complexes, the Free Energy Landscape (FEL) technique was used. The free energy landscapes were generated based on the first two principal components (PC1 and PC2), which were obtained from principal component analysis of the 150 ns simulation trajectories. It enables the visualization of the free energy distribution of the explored conformations and the extraction of preferred, metastable, and transitional structure states.

The topology of each FEL yields direct information on conformational stability and flexibility for the respective protein-protein complexes, since the number, depth, and positions of energy minima define the conformational properties of the system (Figure 1.14 a-f).

a) SP1–KISS1 Complex

FEL of SP1-KISS1 complex had one clearly visible, deep energy minimum, which was enclosed by densely packed contour lines. FEL with densely packed contour lines around the

minimum indicated that most of its simulation structures were trapped in a small region of the conformational space. The presence of secondary minima indicates the possibility of energetically unfavourable conformations. The FEL of the complex reveals the overall stable and favourable conformation of the SP1-KISS1 complex. Additionally, the reduced conformation space represents the strong interaction of the complex, indicating the efficiency of SP1 as a reliable transcription factor for the gene *KISS1*.

b) NMYC–KISS1 Complex

In contrast, the FEL of the NMYC–KISS1 complex had a broad, shallow, and fragmented energy surface. A number of low-depth minima were scattered throughout the landscape, none of which acted like a dominant basin of the conformational population. These minima are shallow, indicating a weak energetic stabilization of any one conformation, with small energy barriers between them, suggesting frequent interconversion. This energy landscape reflects highly dynamic and flexible binding behavior, wherein NMYC–KISS1 continuously interconverts among multiple structurally similar states. Such an energy profile is typical for transient regulatory interactions, thus supporting the idea that NMYC interacts with *KISS1* in a very adaptable and context-dependent mode rather than via a single stable binding mode.

c) CDX2–KISS1 Complex

The FEL of the CDX2-KISS1 complex had one prominent deep minimum and a secondary basin of a medium depth. The prominent minimum included the vast number of the frames that constituted the trajectory. The existence of the secondary basin indicates that the system CDX2-*KISS1* can exist in another metastable state temporarily. The existence of such a high energy barrier between the two basins ensured that even with possible transitions, the overall stability of the system was maintained.

This FEL topology corresponds to a trade-off between rigidity and flexibility, suggesting that CDX2-KISS1 is overall quite stable while preserving a certain degree of conformational mobility. This is perfectly in line with its high docking affinity and moderate MD stability.

d) FLI1–KISS1 Complex

The complex of FLI1/KISS1 had the most limited FEL range. The topography was characterized by one extremely deep minimum with very sharp edges, with not much else present around it. This reflects an incredibly strong thermodynamic stability and a very low conformational diversity. They are fixed in only one main structural basin, implying a rigid and stable binding mode. A condensed form of the FEL of this textually complex protein clearly identifies FLI1–KISS1 as being among the most thermodynamically stable complexes in this analysis. This rigidity of the energy landscape supports the existence of FLI1 as a stable transcription factor for KISS1.

e) GATA2–KISS1 Complex

The FEL profile for the GATA2-KISS1 complex had a strikingly different topology. The energy landscape was broad, fuzzy, and consisted of many shallow basins of similar depth. There was no prominent global minimum; hence, there are no favoured conformations. The energy wells were small, implying continuous transition among various metastable conformations.

Such a FEL profile is a clear indicator of conformational flexibility and a lack of thermodynamic constraints. A dynamic state of various binding modes therefore prevails within the complex GATA2-KISS1.

f) HDAC2–KISS1 Complex

The FEL of the HDAC2-KISS1 complex had two distinct deep minima that were separated by a well-defined energy barrier. These two basins indicate that there are two thermodynamically stable conformational structures that are cycled between in the simulation runs.

The coexistence of dual minima also implied the possibility of conformational switching, in which the complex would alternate between two favourable binding poses. However, neither basin was overwhelmingly dominated, suggesting the importance of the two poses in the biological function.

This bimodal FEL is consistent with the function of regulatory proteins that have structural transitions for modulating their activities. For the HDAC2-KISS1 pair, this notion of switching can be connected with different interaction modes for either transcriptional suppression or chromatin modification.

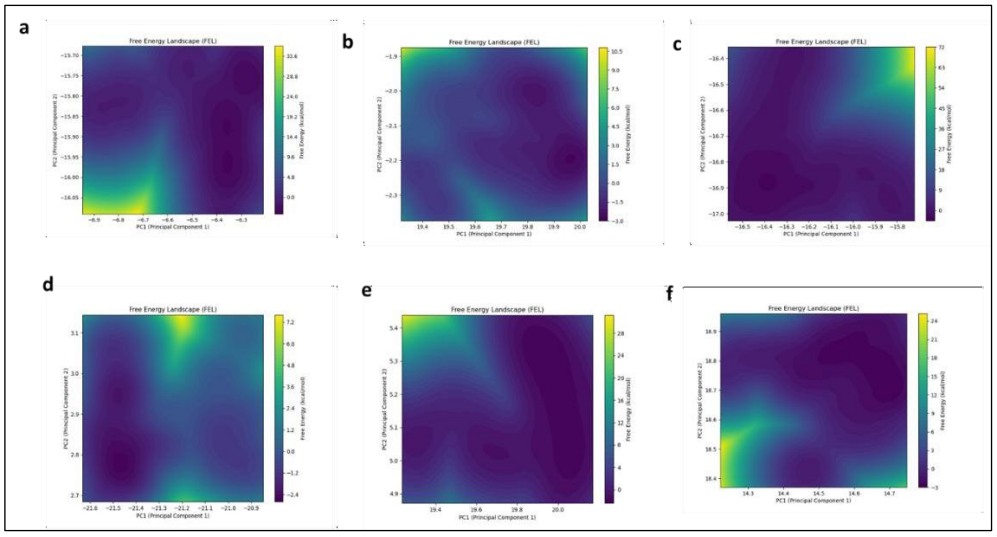


Figure 1.14: *Free Energy Landscape for the following docked complexes in order: a) CDX2-KISS1, b) FLI1-KISS1, c) GATA2-KISS1, d) HDAC2-KISS1, e) NMYC-KISS1 and f) SP1-KISS1.*

1.4.10 K-Means Clustering Analysis of Transcription Factor–KISS1 Complexes

K-means clustering was used to assess the convergence in structure and the structural variety across the six-transcription factor-KISS1 protein complexes in the 150 ns MD simulations. K-means clustering is an unsupervised machine learning algorithm that was used to classify MD runs into similar clusters based on the backbone RMSD, thus giving information on the major structural states, stability, and transition dynamics.

a. SP1–KISS1 Complex

Throughout the course of the simulation, the SP1–KISS1 complex showed a high degree of conformational convergence. In total, 85.5% of the sampled frames were confined within two dominant clusters, with the primary cluster alone accounting for approximately 62% of the total trajectory. These findings evidence a strong tendency toward clustering in SP1–KISS1 and suggest that the protein rapidly adopted a stable binding conformation, maintained for the majority of the simulation time. The secondary cluster represented a closely related conformational state with minor rearrangements at the backbone level, suggesting limited structural transitions rather than complete dissociation or large-scale reorientation.

The compact distribution of frames within these clusters reflects a well-defined binding mode and supports RMSD and Rg observations that indicated the early stabilization of the SP1-KISS1 complex. The presence of only one dominating structural basin ensures SP1 interacts with KISS1 not only in a structurally consistent but also energetically favorable way. Therefore, this may bolster the dependability of SP1 as a steady regulatory associate for KISS1 (Figure 1.15a).

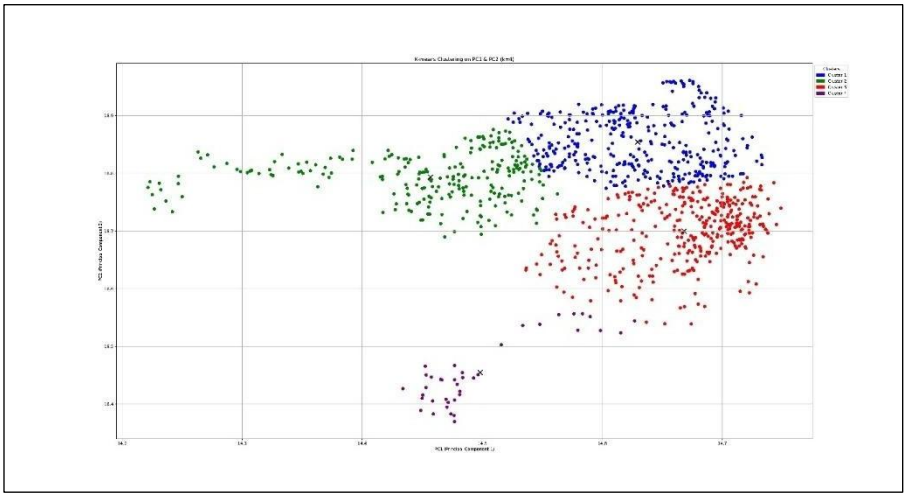


Figure 1.15 a): *K-Means clustering for* SP1-KISS1 *complex*

b. NMYC–KISS1 Complex

The NMYC-KISS1 complex showed a high level of clustering convergence too, whereby more than 85% of the frames fell into the first two clusters. However, unlike the SP1-KISS1 complex, the proportion of the two clusters showed a relatively equal proportion, which reflected high interconversion between the metastable states of the two structures. This reflected the fact that although there is a high affinity between NMYC and KISS1, there are dynamic changes at the interface.

The existence of two prominent conformational basins means that NMYC-KISS1 could display several functional binding modes. This adaptability can be ascribed not only to the flexibility of NMYC, as already noticed in PCA, but also supports its postulate as being a kind of transcription factor that can adapt its structure through constraint satisfaction in response to cellular regulation signals. Even as the complex formed, its oscillation between clusters indicates less inflexible binding than seen in the case of SP1-KISS1 (Figure 1.15b).

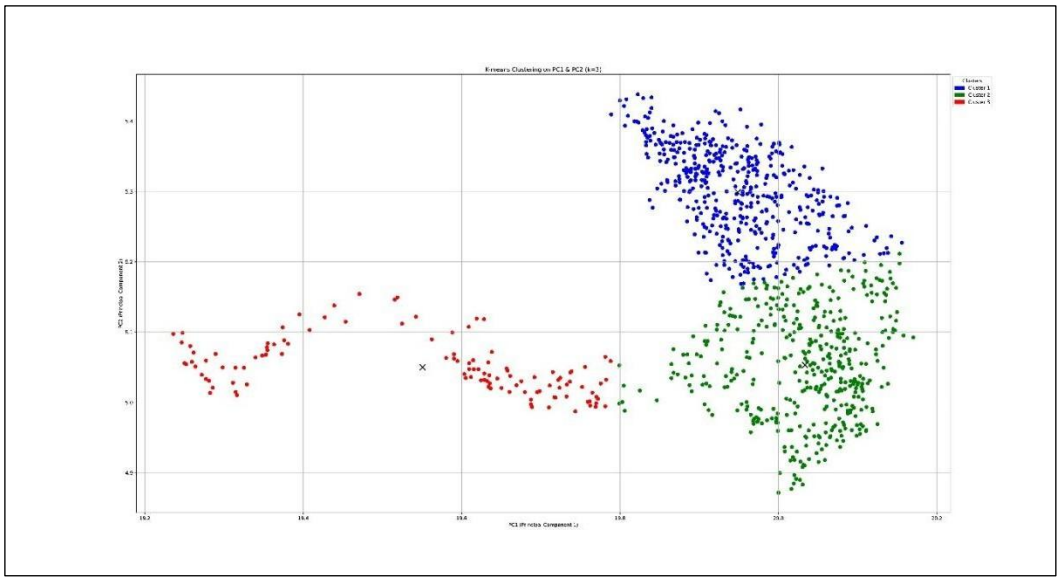


Figure 1.15 b): *K-Means clustering for* NMYC-KISS1 *complex*

c. CDX2–KISS1 Complex

The complex CDX2-KISS1 had high clustering convergence, similar to systems SP1 and NMYC. Around 88% of the frames were associated with two major clusters, with the main cluster containing close to 70% of the data. This revealed that the complex CDX2-KISS1 exists in one stable conformational state, mostly transitioning to a second state. This dominance of a particular cluster helps validate the results of docking, which showed a higher binding affinity and buried surface area for the CDX2-KISS1 protein. A low number of clusters implies a characteristic structural stability of the formed complex after binding of CDX2 and KISS1, as there is less conformation change. This stability implies a high level of authenticity of the transcriptional regulator function of CDX2 (Figure 1.15c).

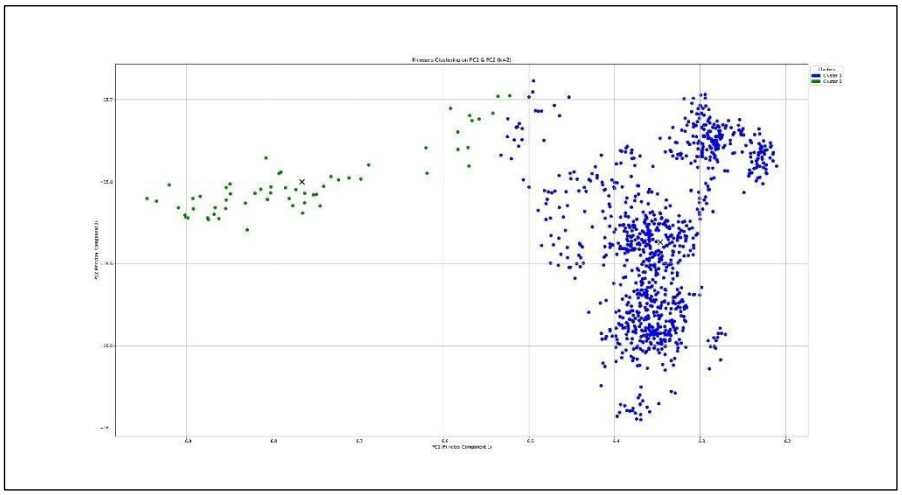


Figure 1.15 c): *K-Means clustering for* NMYC-KISS1 *complex*

d. FLI1–KISS1 Complex

In comparison with the preceding complexes, the number of conformational states in the FLI1/KISS1 system proved to be significantly higher. The trajectory occupied five clusters; no cluster accounted for more than 40% of the entire number of frames. The distribution over the clusters signifies the high flexibility of the system with abundant conformation changes.

Although RMSD analysis had previously indicated stable protein RMSD values, further analysis using K-means clustering indicated that FLI1-KISS1 undergoes small but continuous conformational changes within its binding interface. Such transitions are probably indicative of local secondary structural rearrangements and flexibility rather than complete dissociation events.

The large number of clusters indicates a dynamic equilibrium among various binding conformations, implying that the interaction of FLI1 with KISS1 occurs through a versatile binding interface. Such flexibility in binding conformations would allow the regulation of KISS1 transcription by FLI1 in different cellular contexts (Figure 1.15d).

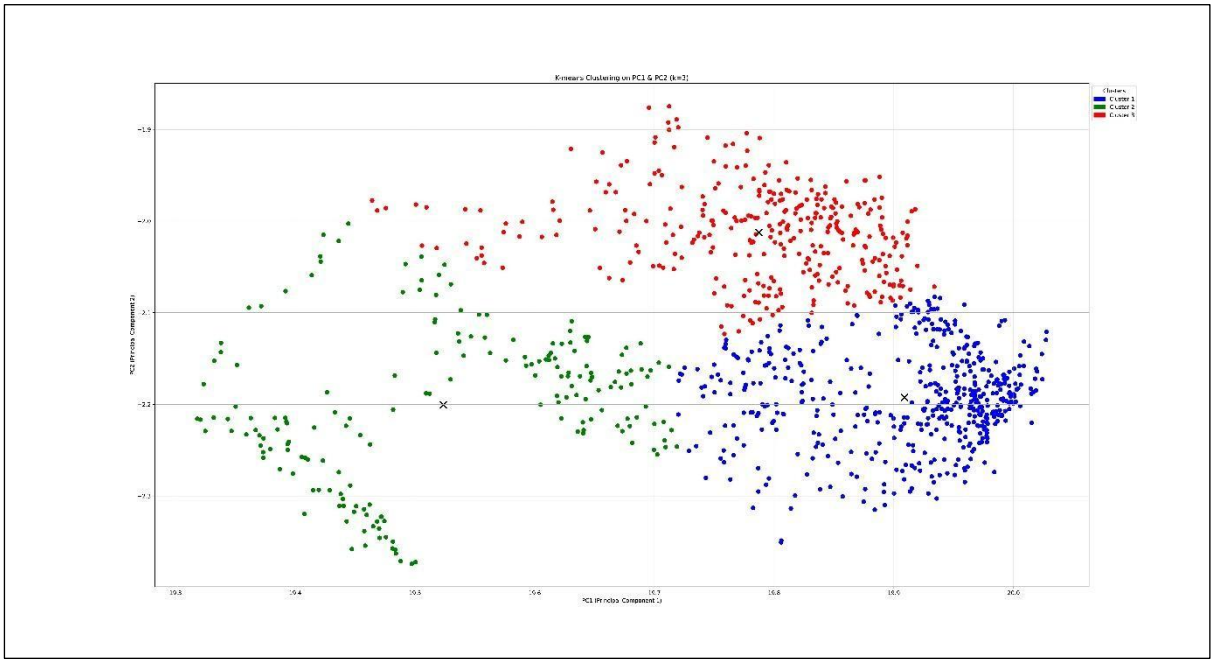


Figure 1.15 d): *K-Means clustering for* FLI1-KISS1 *complex*

e. GATA2–KISS1 Complex

The GATA2/KISS1 complex has one of the most variable clustering behavior in the series of systems studied. The clustering of the MD trajectory produced five clusters with nearly equal distribution, with no cluster exceeding 35% of the distribution in the entire frames. This suggests high variability and structural fluctuations in the system during the simulation.

The presence of such dispersion indicates that the GATA2-KISS1 interface does not have a single, energetically favored binding structure. This observation is in line with the results obtained from the PCA and FEL studies, which revealed extensive sampling of the configuration space and shallow potential minima for the GATA2 protein. This heterogeneity in structure that is apparent in clustering suggests that the interaction between GATA2 and KISS1 is very adaptable or context-dependent. Although this adaptability reduces rigidity in binding, it increases versatility in regulation and makes it possible for GATA2 to regulate KISS1 transcription in different binding contexts (Figure 1.15e).

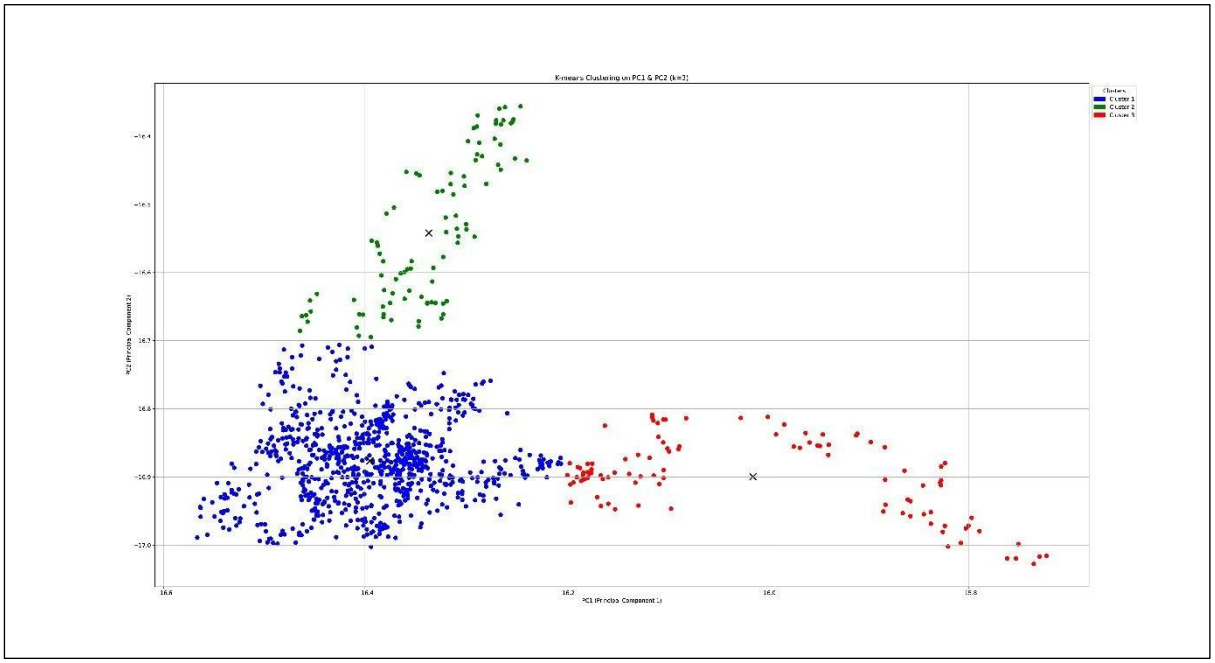


Figure 1.15 e): *K-Means clustering for* FLI1-KISS1 *complex*

f. HDAC2–KISS1 Complex

The HDAC2-KISS1 complex had intermediate clustering characteristics. Around 70% of the snapshots were bounded within the two major clusters; the rest were scattered in small clusters. This suggests the structure has mild convergence with periodic transitions between the major structures.

The fact that there are two prominent clusters indicates that there are fluctuations between two stable binding states for the HDAC2-KISS1 complex that could represent different binding angles/orientations. Yet the scattered points represent instability in the structures against the SP1, NMYC, and CDX2 complexes. Such a process helps to support the findings of RMSD and FEL that HDAC2–KISS1 was identified as a system with moderate stability but one that transitions dynamically. The ability of HDAC2 to exist in several configurations could be linked with the function of the enzyme, which may require flexibility (Figure 1.15f).

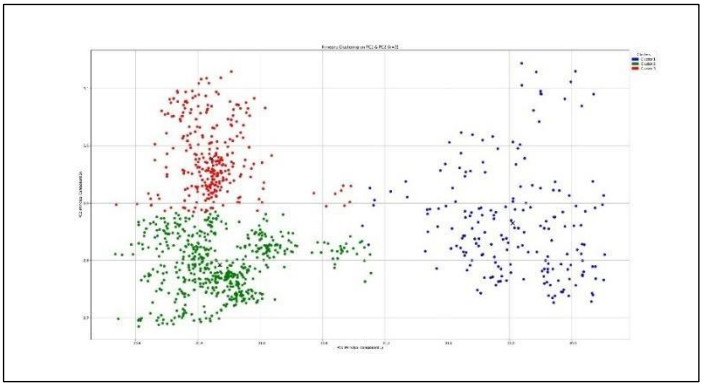


Figure 1.15 f): *K-Means clustering for* FLI1-KISS1 *complex*

DISCUSSION

The current work offers a complete framework using computational analysis and molecular dynamics simulation to understand the structure, physicochemical properties, dynamic behaviour, and regulatory aspects of human *KISS1* gene with its binding protein network of chosen transcription factors. By using analysis in terms of genetic sequence, physicochemical properties, secondary structure prediction, protein location, protein-DNA docking simulation, molecular dynamics simulation, and trajectory analysis such as RMSD, RMSF, PCA, Free Energy Landscape, and K-means clustering, this work attains multi-dimension insights in to KISS1 gene function as a transcriptionally regulated gene encoding metastasis suppressor(Rather et al., 2017b).

The role of Kisspeptin-1 as a metastasis suppressor has been acknowledged for a long time; the molecular processes that determine the transcription and the biological activity of this gene are still unclear. The aforementioned study bridges this research gap by providing a structural basis for the description of the regulation of the gene corresponding to the biological activity of Kisspeptin-1 through the transcription factors and the structural characteristics of the *KISS1* gene (Biran et al., 2008). Analysis of the *KISS1* gene sequence confirmed the presence of a defined ORF that encoded a relatively small protein with a predicted signal peptide and conserved cleavage and amidation site. These characteristics are in complete agreement with the known processing of this gene product into bioactive kisspeptin peptides. Beyond its recognized function as a precursor to bioactive peptides, however, physicochemical properties of KISS1 strongly support the hypothesis that the full-length protein itself has regulatory potential. The high theoretical isoelectric point and predominance of positively charged residues confer a strong electrostatic affinity on KISS1 for negatively charged biomolecules including DNA, RNA, and chromatin-associated proteins. This electrostatic bias is a hallmark of nuclear regulatory proteins and transcriptional modulators (Hsu et al., 2025; Yang et al., 2022).

The hydrophilic nature of KISS1, as indicated by its negative GRAVY score, further supports its solubility and accessibility for molecular interactions. Although the instability index classifies KISS1 as unstable, this parameter should be interpreted in the context of intrinsic disorder rather than structural fragility. Indeed, many intrinsically disordered proteins score high on instability indices despite being biologically stable and functionally essential. The long-predicted half-life of KISS1 in mammalian systems further confirms that it is not prone to rapid degradation, reinforcing that its instability index reflects conformational flexibility rather than biological instability. Secondary structure analysis demonstrated that KISS1 is dominated by random coil regions, with only a small fraction forming α-helices and negligible β-sheet content. This structural organization firmly classifies KISS1 as an intrinsically disordered protein. Intrinsically disordered proteins are increasingly recognized as key

regulatory molecules in transcription, signalling, and chromatin remodelling due to their ability to adopt multiple conformations and bind to diverse partners with high specificity but low affinity. This property allows them to act as dynamic regulatory hubs rather than static structural entities (Shah, Mohan, et al., 2025a).

As regards metastasis suppression, this kind of intrinsic disorder is highly helpful. Metastasis suppressor proteins are required to process various signals from diverse pathways, interact with several transcription factors and regulator complexes, and react to environmental cues. It appears that this intrinsic disorder in KISS1 helps this protein to function as a multifunctional regulator (Shah, Mohan, et al., 2025a). Subcellular localization analysis further supported this view. While the identification of a signal peptide should, in principle, imply involvement in either secretory or membrane-associated routes, relatively weak signal peptide prediction scores argue that KISS1 is not efficiently secreted via conventional secretion mechanisms. Instead, association with peripheral membranes was predicted in the absence of definitive transmembrane helices, suggesting transient interactions of KISS1 with membranes via electrostatic or lipid-mediated means. This type of behaviour is well documented among various classes of regulatory proteins known to shuttle between cellular compartments (Cho et al., 2009; Gan et al., 2021).

Most importantly, nuclear localization predictions reached a high degree of reliability that strongly supported a nuclear regulatory role for *KISS1*. A mechanistic basis for such involvement in transcriptional regulation was provided by this nuclear localization potential. Based on the observation of nuclear localization of KISS1, one may also suggest that it participates in more levels of regulation by performing intracellular regulatory function and being a precursor of extracellular signaling peptides.

Further, the identification of SP1, NMYC, CDX2, FLI1, GATA2, and HDAC2 as *KISS1*-associated transcription factors underscores the complexity of the regulation of *KISS1*. Such transcription factors are responsible for modulating various cellular processes such as proliferation, differentiation, chromatin remodelling, stemness, and oncogenic transformation in various cancer types. Furthermore, the regulation of *KISS1* appears to be governed by a finely tuned regulatory mechanism involving oncogenic and tumor suppressor transcription factors. Structural modelling and PROCHECK analysis confirmed the excellent stereochemistry of all the three-dimensional structures of the transcription factors used in this study. This indicates the validity of the results obtained from the following docking simulation analysis. Identification of biologically active binding residues further underscored the physiological validity of the results obtained from the docking simulation analysis (Cho et al., 2009; Gan et al., 2021).

Protein-DNA docking showed clear distinctions in binding preferences of the TFs. CDX2 had the best predicted interaction with *KISS1*, as shown by the lowest HADDOCK score and the highest buried

surface area, suggesting it as a major transcriptional activator/stabilizer of *KISS1* expression, mainly in differentiation-related contexts. HDAC2 gave the highest electrostatic contribution among all TFs, consistent with its role in chromatin remodelling and transcriptional repression. The SP1, NMYC, FLI1, and GATA2 proteins showed average ensembles of docking, indicating auxiliary or context-dependent regulation (Rather et al., 2017a).

While docking together allows only a static view of the interaction potential, such an interaction has to be tested in molecular dynamics simulations to assess whether they would indeed remain stable under physiological conditions. The 150 ns MD simulations performed here showed that dynamic adaptability, and not static affinity, dictates transcription factor–*KISS1* interactions. RMSD and Rg calculations revealed that SP1-KISS1 and CDX2-KISS1, NMYC-KISS1, and HDAC2-KISS1 complexes had stable structures, whilst FLI1-KISS1 and GATA2-KISS1 complexes had stable RMSD for the proteins, but relatively flexible ligands. This suggested stable complex structures for SP1-KISS1, CDX2-KISS1, NMYC-KISS1. Analysis of hydrogen bonds showed that complex stability is a function not only of hydrogen bonds, but also of their lifetime. RMSF calculations also showed that complex stability is strongly related to the rigidity of residues at the protein-protein interface. These data confirm that binding of the transcription factor to KISS1 is not strictly binding, but rather amounts to a range of possibilities from stiff stabilizing to highly flexible adapting (Iseki et al., 2017).

The Principal Component Analysis revealed more complex information for the relevant dynamic motions. FLI1-KISS1 and SP1-KISS1 had a limited conformational space, indicating rigid and stable binding. The CDX2-KISS1 and HDAC2-KISS1 pairs had an intermediate level of flexibility with well-defined metastable conformation, while GATA2-KISS1 and NMYC-KISS1 had high conformational plasticity. These PCA patterns reveal that the regulation of KISS1 is mediated by transcription factors in different dynamic modes. While some transcription factors are practicing rigid transcriptional control in *KISS1* expression, others are facilitating signal-responsive regulation of *KISS1*. This dissimilarity in dynamic control could be critical in regulating *KISS1* in different contexts (Dashti et al., 2020).

Free Energy Landscape calculations gave thermodynamic support for these observations of dynamics. FLI1-KISS1 and SP1-KISS1 have a very stable energy landscape with a very low minimum energy point. CDX2-KISS1 and HDAC2-KISS1 are moderately stable with a secondary minimum, although GATA2-KISS1 and NMYC-KISS1 are shallow and fragmented energy structures. The presence of two wells in the energy landscape of HDAC2-KISS1 indicates conformational phase transitions exhibited by HDAC2 as an epigenetic regulator (Shan et al., 2017).

These thermodynamic characteristics emphasize that the *KISS1*-TFs interaction is not simply dependent on binding affinity but also on topology in their energy landscapes. The *KISS1* expression maintained through stable TFs has profound implications compared with that mediated through flexible TFs. The

K-means clustering analysis also demonstrated these findings by estimating convergence along the trajectories. The clusters of SP1-KISS1, CDX2-KISS1, and NMYC-KISS1 occupied the major clusters based on their structural similarity, whereas FLI1-KISS1 and GATA2-KISS1 were found in various clusters, suggesting regular switching, and HDAC2-KISS1 occupies intermediate clusters based on the topology of their energy landscapes. Taken together, the docking simulation results and the MD simulation and PCA/FEL and clustering analyses results provide strong evidence that the *KISS1* gene regulation system has a complex and dynamic transcription factor network. The versatility in the transcription regulation system may be necessary for it to function as a metastasis suppressor gene. The metastasis suppressor has to be amenable to different transcription regulation systems. The *KISS1* gene has this capacity with its structural characteristics. (Ciaramella et al., 2018) From a biological perspective, stable transcription factors such as SP1, FLI1, and CDX2 may maintain basal *KISS1* expression required for metastasis suppression, while flexible regulators such as GATA2 and NMYC may fine-tune expression in response to oncogenic or microenvironmental cues. HDAC2-mediated epigenetic regulation may further modulate *KISS1* expression through chromatin remodelling, providing an additional layer of transcriptional control (Biran et al., 2008; Rather et al., 2017b).

The intrinsically disordered property of KISS1 makes it an efficient integrator of these multiple regulatory cues. It does not function as a rigid protein that has a single binding surface but as a conformational ensemble that has the ability to vary its structure to suit the various regulatory proteins. The reason for the context-dependent expression of KISS1 in various types and stages of cancers might also lie in its adaptability. Notably, the results of this research study also emphasize that the decrease in KISS1 expression in aggressive cancers could be due to reasons other than gene mutations, which could be due to dysregulation of its regulatory pathways. The imbalance in expression of certain transcription factors, chromatin modifications, and instability in the structure of certain complexes could functionally contribute to the reduction of KISS1 expression. Hence, there could be a possibility of targeting its regulation pathways for its re-expression in cancers.

In conclusion, this research has ascertained the role of KISS1 as a disordered nuclear-prone regulatory protein with transcription processes modulated by a thermodynamically diverse transcription factor network. The results of the combination of both static and dynamic computational models portray the involvement of the transcription processes of the *KISS1* gene as a complex multi-state process rather than a simple binary interaction. This integrated model, in addition to facilitating an understanding of KISS1 biology, will, in future research, ensure experimental verification. Results will hopefully lead to the production of new therapeutic approaches to reverse the suppression of *KISS1* gene expression and metastatic suppressor function in highly aggressive cancers.

CONCLUSION

The present study provides a comprehensive computational and molecular dynamics–based investigation into the structural, physicochemical, dynamic, and regulatory characteristics of the human *KISS1* gene and its interaction network with key transcription factors. By integrating sequence analysis, physicochemical profiling, secondary structure prediction, subcellular localization, protein–DNA docking, long-timescale molecular dynamics simulations, and advanced trajectory analyses including RMSD, RMSF, PCA, Free Energy Landscape, and K-means clustering, this work establishes a detailed mechanistic framework for understanding the transcriptional regulation and functional behaviour of KISS1. Importantly, this study supports the concept that loss of *KISS1* expression in aggressive cancers may result not only from genetic alterations but also from disruption of its transcriptional regulatory network. Altered transcription factor expression, epigenetic repression, or destabilization of regulatory complexes may collectively contribute to *KISS1* silencing. Therefore, therapeutic strategies aimed at restoring *KISS1* expression should consider targeting its regulatory network rather than the gene alone.

Overall, this study establishes KISS1 as an intrinsically disordered, nuclear-prone regulatory protein whose transcription is governed by a dynamic, thermodynamically diverse, and biologically adaptive transcription factor network. The integrated computational framework presented here provides a robust structural and mechanistic foundation for understanding KISS1-mediated metastasis suppression and offers valuable insights for further analysis especially for Triple-Negative Breast cancer, Glioblastoma and Ovarian cancer.

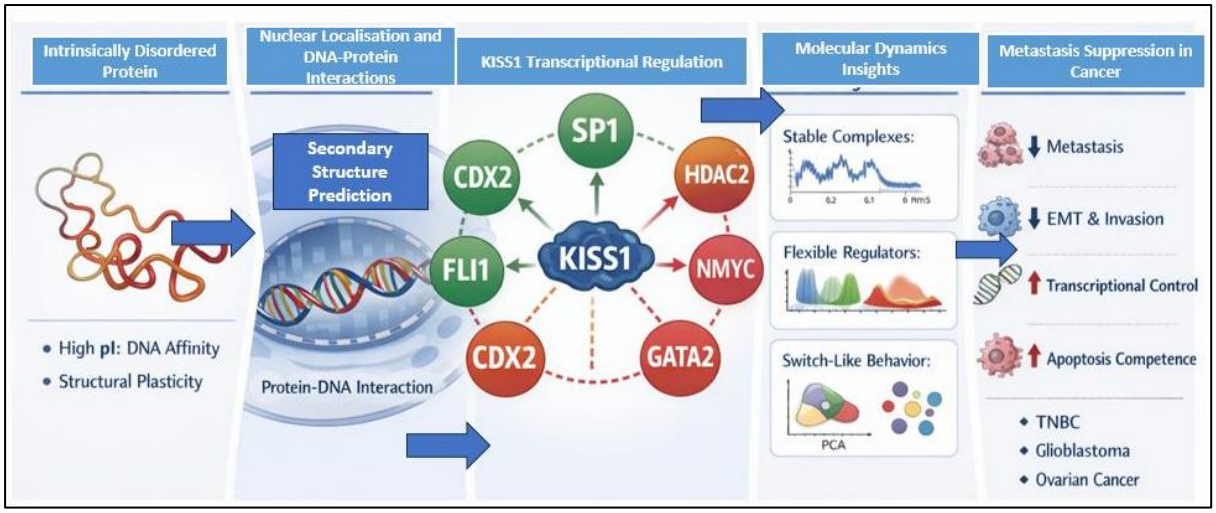


Figure 1.16: Conclusion summarizing the computational framework and regulatory role of KISS1 in cancer