Role of Mutually eXclusive Exons (MXEs) in the functional shift of isoforms: A case study of human Pyruvate Kinase M (PKM)

Shabduli Arjun Sawant

A dissertation submitted for the partial fulfilment of MS degree in Science

Indian Institute of Science Education and Research, Mohali

April, 2021

Certificate of Examination

This is to certify that the dissertation titled "Role of Mutually eXclusive Exons (MXEs) in the functional shift of isoforms: A case study of human Pyruvate Kinase M (hPKM) submitted by Ms. Shabduli Arjun Sawant (Reg. No. MP18030) for the partial fulfilment of MS degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Shravan K. Mishra	Dr. Rajesh Ramachandran	Dr. Shashi B. Pandit (Supervisor)

Dated: April 28, 2021

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Shashi Bhushan Pandit at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Shabduli Arjun Sawant

(Candidate)

Dated: April 28, 2021

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Shashi B. Pandit

(Supervisor)

Dated: April 28, 2021

Acknowledgements

This thesis is dedicated to my family

I would like to express my heartfelt gratitude to my supervisor Dr. Shashi Bhushan Pandit for entrusting me with the freedom to drive my own project. He has been a very supportive guide throughout the course of my thesis and has never failed to give helpful suggestions and insights from time to time. I would like to thank the past and present members of the lab for being so accommodative and providing me with a scientifically enriching atmosphere at work. I would especially like to thank Deeksha and Paras for their indispensable feedback and comments on my work and motivating me when things didn't work out.

I would like to thank my batch of MP18 for being my family far from home. I would like to extend my special thanks to Tejal for always being just a call away, Anusha, Azeez, Pallavi and Sanjeev for being there for me during the good and bad times.

I am immensely grateful to my parents, my sister and my grandparents for making me a strong and confident woman. I owe to my parents all that I am and will ever be. I would also like to express my heartfelt gratitude to late Dr. NM Sunder for his invaluable thoughts and guidance and Suchitra for all her love and support.

Lastly, I would like to thank Varun for always being there for me in every way possible. He has been encouraging and supportive of all my dreams and decisions and has been a constant source of inspiration.

This work has been funded by IISER Mohali.

List of Figures

List of Tables

Abbreviations

ANM - Anisotropic Network Model

AS - Alternative Splicing

BSA - Buried Surface Area

FBP - Fructose-1,6-bisphosphate

MXEs - Mutually eXclusive Exons

NMA - Normal Mode Analysis

PCA - Principal Component Analysis

PEP - Phosphoenolpyruvate

PKM - Pyruvate Kinase M

PKM1 - M1 isoform of PKM

PKM2 - M2 isoform of PKM

PPI - Protein-Protein Interaction

RIN - Residue Interaction Network

RMSD - Root Mean Squared Deviation

Contents

Abstract

The process of alternative splicing (AS) is known to account for a major source of human proteome diversity. Among various known types of AS events, Mutually eXclusive Exon (MXE) splicing results in isoforms having one (or more) out of two (or more) exons mutually eliminated in a coordinated manner. MXE events are known to generate highly diverse function protein variants from the same gene. These splicing events can lead to proteins with similar length and scaffold but highly specific functions.

It has been of great interest to understand how change(s) in region(s) of a protein can significantly alter the function of protein. In order to gain insights into structural changes in isoforms generated from MXEs events, we considered human Pyruvate Kinase M (PKM) as a model system because tertiary structures are known for both isoforms (PKM1 and PKM2). It is known that PKM1 is a constitute enzyme and PKM2 shows allostery on binding various effectors. In the present study, we have systematically analyzed the origin of allosteric behavior in monomeric structures by extensive analyses of structural features of both PKM1 and PKM2. Our analyses showed that differences in the inherent dynamics of loop, which is a region encoded by the mutually exclusive exons, in PKM1 and PKM2 could affect oligomerization as well as affect allosteric transitions. The analyses of allosteric paths suggest that FBP-mediated allostery is greatly enhanced in PKM2 whereas in PKM1 the path has low significance.

Chapter I

Introduction

1. Introduction

Eukaryotic pre-mRNA undergoes a process called splicing wherein the intronic regions are removed to combine together the exons, thus forming a continuous mature mRNA transcript, which is then translated to a protein. Splicing of pre-mRNA occurs when a set of proteins and ribonucleoproteins specifically recognize the splice site sequence at the exon-intron boundaries. Alternative splicing on the other hand, occurs when two or more such splice sites compete, resulting in different splice variants (proteins) from the same gene. *(Srebrow & Kornblihtt, 2006)* It therefore accounts for protein diversity in eukaryotes by allowing a single gene to generate multiple mRNA transcripts which can be translated into distinct proteins. Alternative splicing can be of different types depending on how the set of exonic sequences are combined together into a mature transcript [Figure 1].

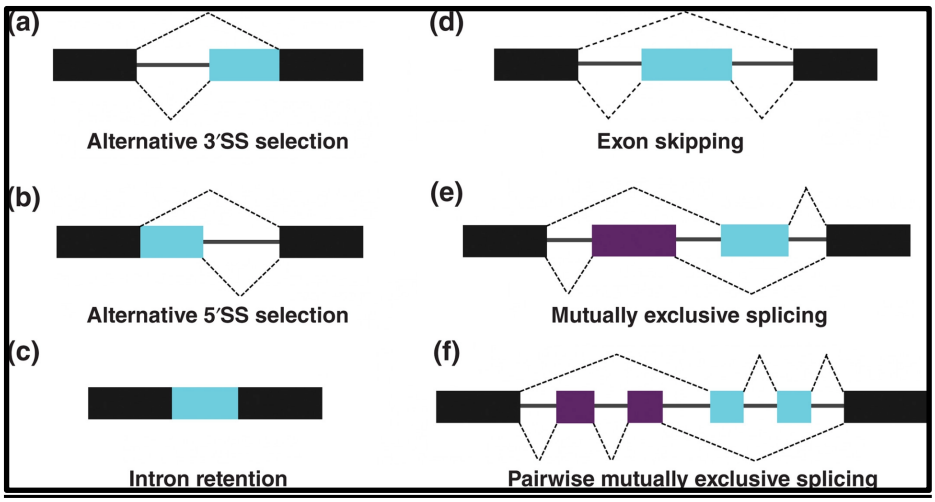

Figure 1: Types of Alternative Splicing (adapted from Jin et al., 2017)

Mutually exclusive exon (MXE) splicing is a specific type of splicing wherein only one from amongst the set of tandemly arranged exons is included or spliced in the mature transcript. *(Jin et al., 2018).* The term "mutually exclusive" implies that two (or more) exons do not co-occur in any of the transcripts. A recent high throughput study predicted that the human genome contains 1399 high confidence MXEs of which 1116 were found to belong to a cluster of two exons. *(Hatje et al., 2017).* In addition to their arrangement

in tandem, they are also found to exhibit a high sequence similarity, suggesting that they have resulted from an exon duplication event. *(Kondrashov & Koonin, 2001)*. Despite sharing a high sequence similarity, they are usually functionally non-redundant, their inclusion in the transcript being highly regulated. *(Hatje et al., 2017)* Deregulation of MXEs have been implicated in diseases like Timothy syndrome (mutation in CACNA1C gene) *(Splawski et al., 2004),* cardiomyopathy (defect in SLC25A3) *(Mayr et al, 2011)* and cancer (mutations in pyruvate kinase PKM) *(Chen et al., 2019 & David et al., 2010)*. Exon duplication coupled with MXE splicing presents an interesting strategy for generating molecular and functional diversity, and hence we wanted to understand how these changes are manifested at the level of protein structure of the resulting isoforms. However, studying the implications of such MXE splicing on the protein structure is limited by the fact that not all splice variants are translated to functional proteins. Although various groups have studied the functional implications of alternative splicing based on computational predictions *(Menon et al., 2011, Li et al., 2016 & Mishra et al., 2020);* another problem is finding such an MXE spliced isoform pair for which the structures of both the isoforms have been experimentally determined. We wanted to study the cases wherein MXE splicing led to a functional switch in the isoforms while retaining the length of the isoforms.

We made use of our in-house database, NextRAPdb to find examples of such isoform pairs from the human proteome. This is when we came across the isoforms of Pyruvate Kinase M (PKM, muscle isoform), a metabolic enzyme catalyzing the last step of glycolysis, thereby committing the cell to undergo oxidative phosphorylation. It catalyzes the transfer of a phosphate from phosphoenolpyruvate (PEP) to ADP generating pyruvate and ATP, in an irreversible reaction. Pyruvate kinase also regulates the supply of phospho metabolites, and its down regulation leads to accumulation of the phospho metabolite pool preceding the final glycolytic reaction to be channelized into biosynthetic processes. As a result, the regulation of pyruvate kinase becomes important in controlling tumor growth. *(Dombrauckas et al., 2005)*. MXE splicing of the human *pkm* gene gives rise to the non-allosteric M1 isoform (PKM1) encompassing exon 9 and the allosteric M2 isoform (PKM2) comprising exon 10, both of the exact same length of 531 amino acids, differing from each other only at 22 residue positions (~96% sequence identity) in the

alternate exon region. Pyruvate kinase monomer is composed of four domains namely, the N-terminal domain, A domain, a mobile B domain and the effector binding C domain [Figure 2]. It is known to catalyse the reaction only upon forming a functional tetramer of D2 symmetry with the A domains of adjacent monomers forming one interface (A-A') and the C domains forming the other interface (C-C') [Figure 2]. The reaction substrates, Phosphoenolpyruvate and ADP bind in a cleft formed between the A and the B domains.

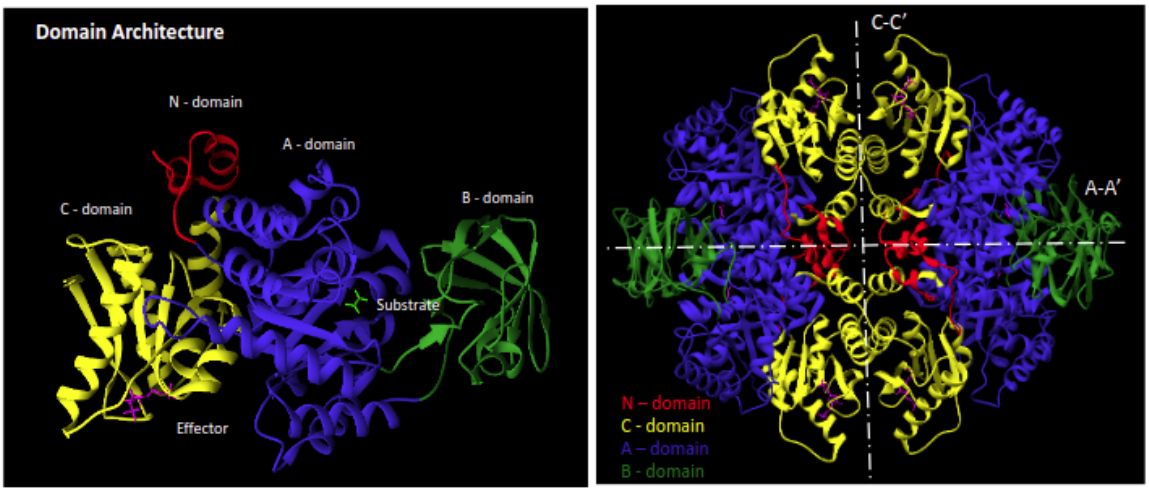

Figure 2: Human Pyruvate Kinase M Domain Architecture and Tetramer Symmetry

The non-allosteric PKM1 is found to always exist as a tetramer and is the constitutive isoform in heart, muscles and brain. On the contrary, PKM2 is the fetal isoform and is found to be in mixed populations of monomer, dimer and tetramer, and is allosterically regulated by Fructose-1,6-bisphosphate (FBP), amino acids *(Yuan et al., 2018)*, SAICAR *(Yan et al., 2016)* and T3. *(Morgan et al., 2013)* FBP is a strong allosteric regulator of PKM2 (exhibits K-type allostery) and reduces the Km value of PEP binding to its binding site 40 Å away in addition to aiding the formation of a functional tetramer.

Regulation of PKM2 by FBP is crucial for maintaining metabolic homeostasis. Recent studies have shown that dimeric and monomeric PKM2 can be translocated to the nucleus and undergo HIFα-mediated transactivation and enhance cell proliferation. *(Chen*

et al., 2019) Studies have also shown the up-regulation of the M2 isoform in many cancers and its role in cancer metabolism. *(Christofk et al., 2008)*. Both PKM1 and PKM2 catalyze the same reaction but differ in their mode of regulation. In the present work, we investigate tertiary structures of PKM to understand how structural dynamics accompanying MXE leads to drastic shift in protein function from non-allosteric to allosteric (or vice-versa).

Allostery refers to the communication that occurs between spatially separated, non-overlapping sites upon binding of an effector to a site distinct from the catalytic site, thereby regulating a protein's function. It was in 1904, when *Bohr, Hasselbalch & Krogh* first reported their observation of such a phenomenon wherein CO_2 influenced the binding of O_2 to hemoglobin. This further led to the proposition of the two classic models of allostery, the MWC model by *Monod, Wyman & Changeux, 1965* and the KNF model by *Koshland, Nemethy & Filmer, 1966,* which suggested conformational change to be a signature of allosteric transition. The understanding of the concept of allostery underwent a paradigm shift when *Cooper & Dryden 1984,* proposed their model of allostery without a conformational change, and that it could also be an outcome of changes in thermal fluctuations in the protein upon effector binding. This new perspective of "dynamic allostery" served as a stepping stone (foundation) for new concepts and theories *(Nussinov et al, 1999 & Ranganathan et al., 2003)* in the quest for unravelling the mechanistic details of the second secret of life. *(Fenton et al., 2008)* The current view of allostery proposed by *Nussinov et al., 2004* suggests that all dynamic proteins can have the potential to be allosteric by redistribution of the underlying conformational ensemble in presence of ligands or mutations, which can enable such a shift. Although FBP and amino acid mediated allostery is quite well studied in case of PKM2 and the R and T transitions of the tetramer have been explained by the seesaw model *(Wang et al., 2014)* and the dock-rock-lock model *(Morgan et al., 2013)*; the structural basis of allostery in an individual monomer is not well understood.

In this study, we have employed several computational approaches to address two key questions; firstly, what structural changes occur as a result of MXE splicing, at the level of monomer/tetramer in PKM1 and PKM2; and, distinguishing structural feature to understand allostery exhibit within a monomer. This will provide insights into structural changes accompanying sequence variation responsible for functional shift in PKM.

Chapter II

Methodology

2.1 RMSD Calculation

2.2 Calculation of Interfacial Energy and Area

2.3 Network Analysis

2.4 Principal Component Analysis (PCA)

2.5 Normal Mode Analysis (NMA): Anisotropic Network Model (ANM)

2.6 OHM: Calculation of Allosteric Paths

2. Methodology

As a broad question, we wanted to study the role of structural dynamics in the function of nearly identical length splice variants (isoforms) of human proteins. To construct a starting dataset to work with, we filtered out protein isoform pairs from our in-house database, NextRAPdb. Isoform pairs were chosen based on the criteria that the two isoforms should have been spliced from the same gene and their protein length shouldn't vary more than ± 5 amino acid residues. Next, we looked for known functional implications with respect to dysfunction, disorder, and disease, as a result of such alternative splicing giving rise to the respective isoforms. From amongst the filtered examples, the isoforms of the metabolic enzyme Pyruvate Kinase M (PKM) presented a really interesting case. PKM isoforms M1 and M2 are a result of a mutually exclusive splicing event of the human *pkm* gene such that the transcript of the former includes exon 9 while that of the latter includes exon 10. Apart from differing by the mutually exclusive exons, the two isoforms are of the exact same lengths and share a very high sequence identity of 96 %. We found that the structures of both the PKM1 and PKM2 isoforms had been determined at a good resolution (2.85 Å and 2.90 Å respectively) in the PDB. We examine the role exon 9 and exon 10 in altering the functional regulation of PKM1 and PKM2 respectively.

A systematic approach for computational structural analyses of PKM1 and PKM2 was employed to understand the role of structural dynamics in altered functional regulation, as described in the following flow diagram. [Figure 3]

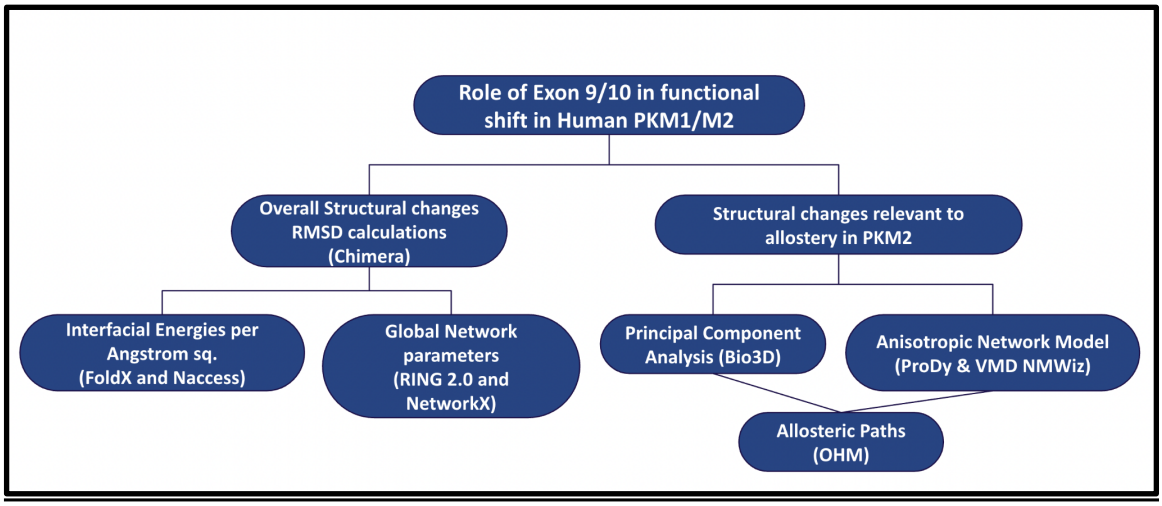

Figure 3: Flowchart of Methodology

2.1 RMSD Calculation

To calculate conformation change in proteins, we used Root Mean Square Deviation (RMSD) after optimal superposition of two structures. A basic structural comparison of representative PDBs of PKM1 and PKM2 was performed by superimposition of C-alpha atoms and RMSD was calculated using Chimera *(Pettersen et al., 2004)*, which is the known standard method for structural comparison. The calculation of the RMSD involves two main steps: (i) alignment and (ii) optimal superposition. It aligns two proteins by establishing a one-to one correspondence between equivalent atoms by superimposing the two structures. This optimal superimposition is achieved by rotating and translating one structure with reference to the other so as to minimize the weighted sum of squares of distances between equivalent atoms. *(Coutsias et al., 2019)*

$$RMSD = \sqrt{(1/N)\sum_{i=1}^{N} \delta_i^2}$$

The global RMSD was calculated by superimposition of full length, comparable structures (w.r.t. bound-state) of PKM1 *(3SRF)* and PKM2 *(3SRD)*. In order to compute local RMSD we superimposed only this region of interest for calculation of RMSD. For computing the orientation of loop, we extended this loop by ±5 residues, considering as a stem region and superposed this whole region.

2.2 Calculation of Interfacial Energy and Area

PKM is functional in its tetrameric state, which comprises two distinct interfaces i.e. the substrate binding A-A' interface and the C-C' interface mostly formed by the respective alternate exons in the M1 and M2 isoforms [Figure 3 & 4]. The exon's contribution to formation of one entire interface prompted us to compare the interfacial energies of the whole tetrameric complexes of both PKM1 and PKM2. Interfacial energies were calculated using tetramers of *3SRF, 3SRD and 1ZJH*, representative structures of PKM1 (substrate bound structure), PKM2 holo (substrate and allosteric activator (FBP) bound structure) and PKM2 apo respectively, using FoldX's AnalyseComplex. *(Durme et al., 2011)* The interfacial buried surface area (BSA) was calculated using Naccess with default parameters. *(Thornton and Hubbard, 1992-6)* To calculate the accessible area buried between 2 chains, we performed three separate calculations as follows:

For eg., If one wants to find the area buried between a complex AB, the one for chain A, second for chain B and a third time for the AB complex. Then one can calculate the BSA between A and B by subtracting the Surface Area (SA) of AB from the sum of SAs of A and B as follows

$$BSA = (SA\ of\ A + SA\ of\ B) - SA\ of\ AB$$

2.3 Network Analysis

Global network properties were calculated for all-atom networks of PKM1 and PKM2. We used monomers of *3SRF, 1ZJH, 3BJT, 3SRH* and *3SRD* as representatives of M1 holo (+substrate), M2 apo, M2 holo (+FBP), M2 holo (+substrate) and M2 holo (+substrate, +FBP) respectively. RING 2.0 *(Piovesan et al., 2016)* was used to get all atom interactions in these monomers with residue separation of 1 and one interaction per residue pair. The list of edges from RING 2.0 was used to construct a network with a distance cut-off of 3.4 Å and global network parameters were calculated, using NetworkX in Python. The following graph parameters were calculated as shown below:

- *Average clustering coefficient (C)* = $1/n \sum_{v \in G} c_v$
- *Diameter, Radius and,*
- *Average shortest path length (a)* = $\sum_{s,t \in V} d(s,t) \div n(n-1)$

2.4 Principal Component Analysis (PCA)

Principal Component Analysis is a multivariate statistical technique applied to systematically reduce the number of dimensions needed to describe protein dynamics through a decomposition process that filters observed motions from the largest to smallest spatial scales. PCA is a linear transform that extracts the most important elements in the data using a covariance or a correlation matrix. An eigenvalue decomposition (EVD) of the C-matrix leads to a complete set of orthogonal collective modes (eigenvectors), each with a corresponding eigenvalue (variance) that characterizes a portion of the motion, where larger eigenvalues describe motions on larger spatial scales. When the original data is projected onto an eigenvector, the result is called a principal component (PC). *(David et al., 2014).* We wanted to capture protein dynamics relevant to allostery in case of PKM2 from its many structures crystallized with different ligand molecules. Therefore, we subjected a total of 24 wild-type PKM2 structures along with one PKM1 structure to a PCA using Bio3D, which is an R package. *(Grant et al., 2006)* Monomers with non-terminal missing residues were completed using Schrödinger.

2.5 Normal Mode Analysis (NMA): Anisotropic Network Model(ANM)

While a PCA can be performed on any high dimensional dataset, Normal Mode Analysis derived from an Elastic Network Model (ENM) provides an alternative method to probe for large-scale protein dynamics. Atilgan et al. suggest that fluctuation dynamics of proteins can be modelled as those of elastic networks, the nodes being the residues and the springs being the corresponding inter-residue potentials. Although the Gaussian network model provides close to experimental information on the magnitude of fluctuations, it does not account for their direction and therefore considers them to be

isotropic. Because the directional fluctuations predicted by NMA can give useful insights into biologically relevant mechanisms like allostery, *(Atilgan et al, 2001)* we performed a NMA derived from an Anisotropic Network Model using ProDy, *(Bakan et al., 2011)* at default parameters for monomers of *3SRF* (+substrate) and *3BJT* (+substrate, -FBP) as representatives of PKM1 and PKM2 respectively. The overlap (RMSIP - Root mean square inner product) of the first 10 modes of both PKM1 and PKM2 was calculated using the "calcSubspaceOverlap" function from the Comparison Functions in Prody. Other PKM2 structures used were *1ZJH* (apo) and *3SRD* (+substrate, +FBP).

2.6 OHM: Calculation of Allosteric Paths

From the insights derived from both PCA and NMA on underlying protein dynamics, we sought to find plausible coupling between the predicted fluctuating elements and role in allosteric transmission. Predicting a single allosteric path with confidence can be challenging from amongst the multiple available communication paths between the specified source and sink.

Wang et al., (2020) have given an efficient methodology for predicting allosteric paths in protein structures in a robust and exhaustive manner. The algorithm uses information from the all-atom contact matrix of the protein structure which is further used for probabilistic calculations involving perturbation propagation for finding the most likely path between the user-defined source and sink residues with confidence. This perturbation propagation algorithm with default parameters was used to predict the allosteric paths as it accounts for structural, network and statistical aspects thereby predicting the most likely allosteric path with confidence.

Chapter III

Results & Discussion

3. Results

We obtained human PKM1 and PKM2 structures, which do not have any mutations and are either unbound or bound to ligands from Protein Data Bank. The structures having missing residues were modelled based on another structure using Schrödinger. The list of PDB structures used in the present work is tabulated in Table 1. First, we compared tertiary structures (tetrameric form) of PKM1 and PKM2 to find conformational changes due to MXE event. The superposition of tetramers of PKM1 and PKM2 superposed well with a global RMSD of 0.9 Å. On analyzing the superposed structures, we identified three regions having high local RMSD [Figure 4] that include FBP binding site, Exon 9/10 loop region and B-domain. It is well established that B-domain undergoes conformational change with open and close states for releasing product and binding substrate respectively. There are multiple residue changes in region of exon 9/10 that may contribute to RMSD and the loop in FBP binding region undergoes conformational change to accommodate FBP activator. Apart from these, rest structure is highly similar in PKM1 and PKM2. This suggests that apart from local perturbations, exon swapping does not result in any major conformational variations.

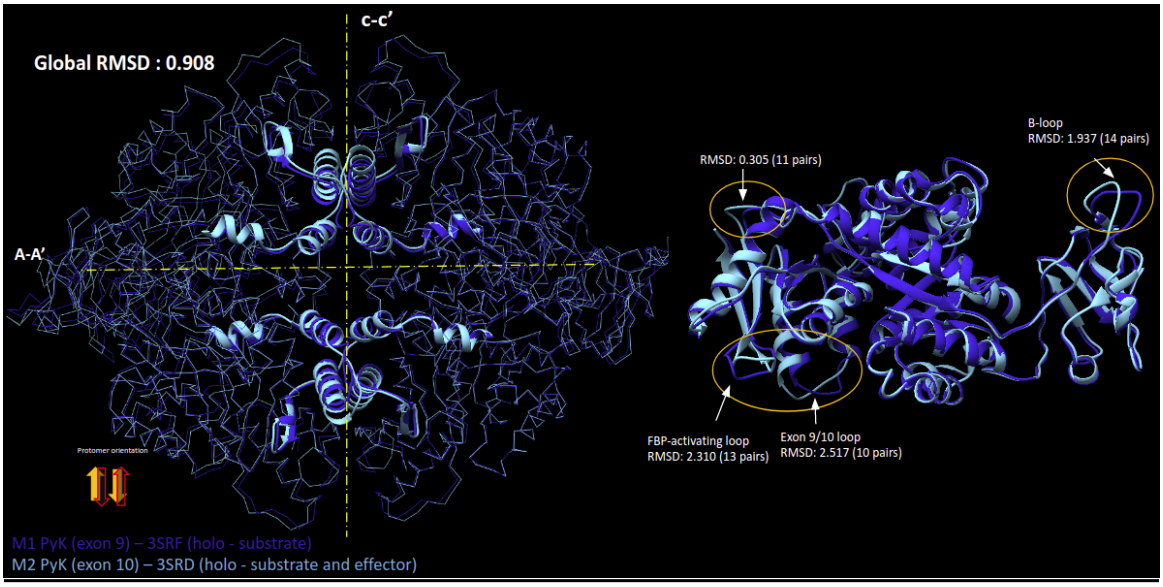

Figure 4: Superposed structures of PKM1 and PKM2

Table 1: Information of PKM PDB ids used in the study

PKM1/2	PDB ID	Ligands
PKM1	3SRF	PO4, GOL, PYR, K, MG
PKM2	1ZJH	-
PKM2	3GQY	DZG, FBP, TLA, UNX
PKM2	3GR4	ADP, DYY, FBP, TLA, UNX
PKM2	3H6O	FBP, D8G, UNX
PKM2	3ME3	3SZ, FBP, SO4, UNX
PKM2	3SRD	FBP, GOL, K, MG, OXL
PKM2	3SRH	PO4
PKM2	3U2Z	07T, FBP, UNX
PKM2	4B2D	FBP, MG, SER
PKM2	4G1N	NZT, OXL, MG
PKM2	4JPG	FBP, 1OX
PKM2	5X1V	FBP, 7XX
PKM2	5X1W	FBP, 7Y0
PKM2	6GG3	ALA, PO4
PKM2	6GG4	PHE, PO4, K
PKM2	6GG5	TRP, PO4, K
PKM2	6GG6	SER,PO4, MG, K
PKM2	6JFB	SER,PO4
PKM2	6NU1	FBP, CYS, OXL, K, MG
PKM2	6TTF	FBP, LZ2
PKM2	6TTQ	FBP, NXH
PKM2	6V74	FBP, ASN, 144, GOL, OXL, K, CL, MG
PKM2	6V75	FBP, ASP,PEG, GOL, OXL, K, CL, MG
PKM2	6V76	FBP, VAL, GOL, OXL, K, CL, MG

Since we did not observe large conformational changes between PKM1 and PKM2, we compared the protein-protein interaction (PPI) interface strength to find whether exon swap results in a stable tetrameric structure for either isoforms. There are two distinct symmetry related PPI interfaces referred to as A-A' and C-C' [Figure 4]. It is a well-established fact that a stable tetramer in the it's correct conformation is necessary for optimal enzyme activity. Additionally, the differences in populations of oligomeric states in case of PKM1 being constitutively tetrameric while PKM2 existing as monomer, dimer and tetramer further prompted us to probe the properties of interfaces formed by tetramers of the respective PKM isoforms.

As described in methods, we computed buried surface area and interfacial energies of the PPI interfaces (A-A' and C-C') formed in hPKM1 (holo) and hPKM2 (both apo and holo). Since buried surface area varied in different structures, we compared the normalized interfacial energy by their respective buried surface area to compute interfacial energies/$Å^2$. These results are tabulated in Table 2. As can be seen, we did not find any significant difference in normalized interfacial energies between hPKM1 and hPKM2 PPI interfaces. This clearly suggests that both isoforms are able for form a stable interface. Surprisingly, the C-C' interfaces were also found to be energetically equivalent despite being mostly formed by the residues from the alternate exons. It is worthwhile to note that hPKM2 structures are bound to FBP, which may indicate that tetramer interfaces are already formed and we may not be able to see differences in interfacial energies.

Table 2: Interfacial energy per Angstrom sq. buried surface area

PDB	Interface	Energy (Kcal/mol)	Buried Surface Area (Å^2)	Energy/Å^2
M1 holo	A-A'	-43.496	6028.5	-0.007
M2 holo	A-A'	-56.303	5974.2	-0.009
M2 apo	A-A'	-42.554	5457.8	-0.008
M1 holo	A-A'	-50.597	5895.9	-0.009
M2 holo	A-A'	-51.730	6054.7	-0.008
M2 apo	A-A'	-42.975	5457.8	-0.008
M1 holo	C-C'	-17.247	2937.5	-0.006
M2 holo	C-C'	-20.785	2676.6	-0.008
M2 apo	C-C'	-19.963	2852.5	-0.007
M1 holo	C-C'	-16.218	2863.3	-0.006
M2 holo	C-C'	-18.814	2730.6	-0.007
M2 apo	C-C'	-19.892	2852.5	-0.007

Having observed no significant changes in the RMSD and interfacial energies between PKM1 and PKM2 structures, we sought to find whether underlying residue interaction network undergo any change due to MXE events. We obtained all atom residue interaction network (RIN) for both isoforms considering them as monomer (see methods) and calculated global network properties such as average clustering coefficient, diameter, radius, and average shortest path length [Table 3]. The visual comparison of RIN [Figure 5] showed no drastic differences and these seem to form a cohesive connected component. A similar observation is seen from comparison of global properties [Table 3]. This suggests that the network representation and parameters used in the present analysis cannot assist in understanding functional shift in the isoforms of human pyruvate kinase M.

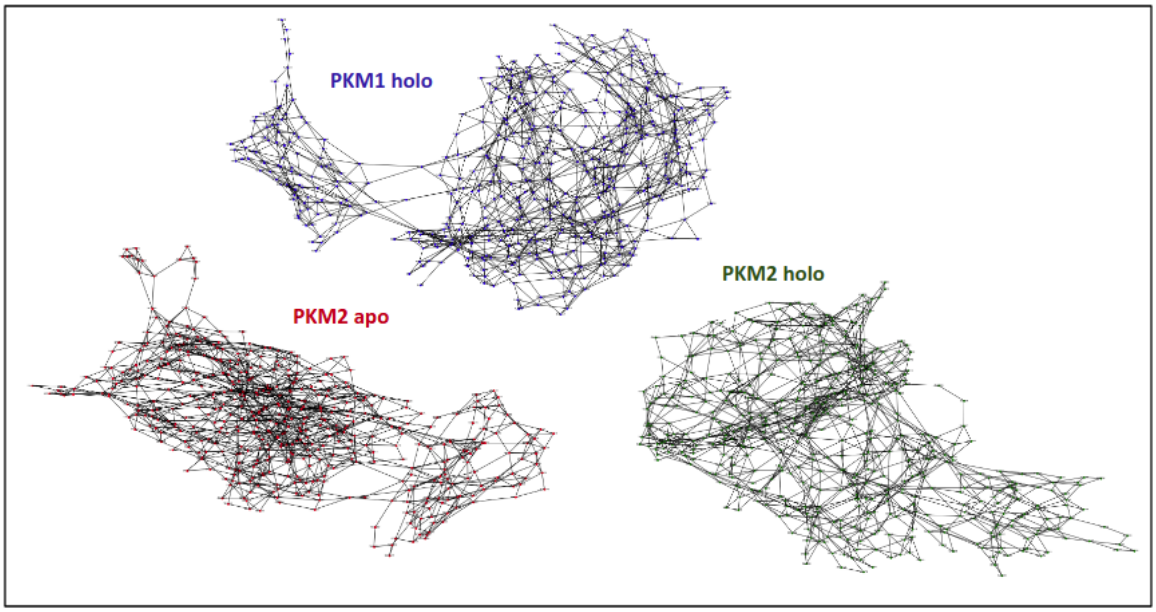

Figure 5: All atom contact network of PKM1 and PKM2

Table 3: Global network properties of PKM1 and PKM2

All atom network (Dist. cutoff - 3.4 Å)	**M1 holo (+substrate)**	**M2 apo**	**M2 holo (+substrate, -FBP)**	**M2 holo (+substrate, +FBP)**
Avg. clustering coeff.	0.26	0.28	0.27	0.25
Diameter	19	19	19	18
Radius	10	10	10	10
Avg. shortest path len	7.54	7.38	7.35	7.49

The above analyses could not identify how exon changes lead to change in allosteric behavior on hPKM isoforms. This prompted us to investigate change in dynamics of exon 9/10 region to find whether it can play any role in allostery.

In this section of study, we analyzed isoform structures to identify characteristic internal dynamics using Normal Mode Analysis and find region of proteins, which show large conformational variation using Principal Component Analysis (PCA). The work of *Macpherson et al., (2019)* used AlloHubMat to predict hubs of allosterically important residues and showed that these were evolutionarily conserved; while *Tang & Fenton (2017)* showed that a large fraction of protein contributed to the allosteric regulation, by alanine scanning mutagenesis of the liver isoform of Pyruvate kinase. Despite these studies along with several others, the structural basis of FBP-mediated allostery at the level of an individual monomer is not very clearly understood. Here, we have used a different approach to finding the mechanistic details of allostery in PKM2 at the level of monomer and also tried to understand absence of allostery in PKM1 despite sharing high sequence and structural similarity with the former. Although only one human PKM1 structure has been determined till now, there exist many structures of the human PKM2 isoform that have been solved under different conditions and ligand-bound states. We subjected monomers of 24 wild-type PKM2 structures in different bound-states and the lone PKM1 structure to clustering using Principal Component Analysis (PCA) using Bio3D. As can be seen, ~85% of variation could be explained by first two principal components (PCs) and ~91% of variation is captured by 3 PCs. Further the projection of structures on first two PCs showed that these cluster into three distinct groups across Principal Component (PC) 1 as shown in [Figure 6]. The red cluster includes the PKM2 structures with the B domain in an open (inactive) conformation while the green cluster is composed of PKM2 structures having their B domains in a closed (active) conformation. The blue Alanine bound PKM2 structure falls in between the two groups with B domain in a half closed position. It is known that Alanine is inhibitor of PKM2 catalytic activity. Based on PCA, it can be suggested that Alanine bound structures is stuck between open and close state that may explain its inhibitory activity. Interestingly, substrate bound PKM1 was found to cluster along with the PKM2 active state conformers. This suggests that structural conformations of hPKM1 and active hPKM2 are similar.

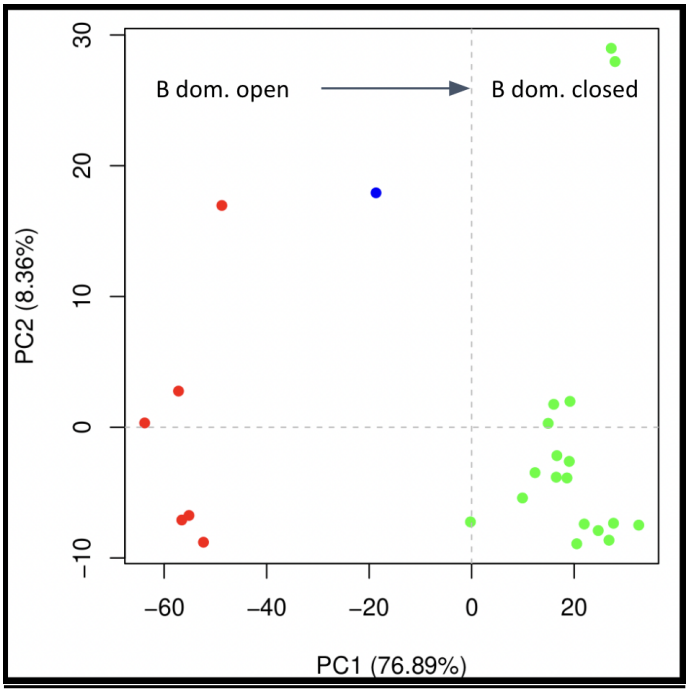

Figure 6: A plot of PC1 v/s PC2 showing three distinct clusters

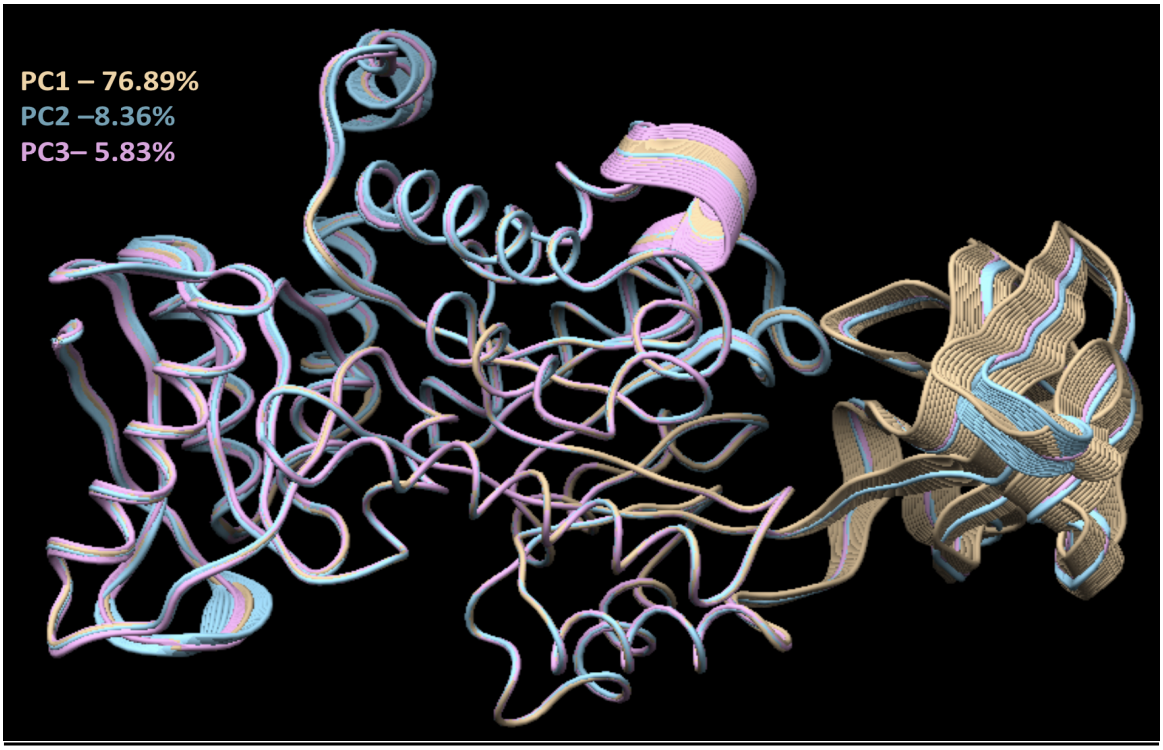

Figure 7: Contribution to structural variance by the first three PCs

We further projected the first three PCs on tertiary structures to find regions, which contribute, to various PCs [Figure 7]. As is evident, the B-domain contributes maximally to variation along PC1 that is known to undergo conformation changes during enzymatic function. This B domain movement is known to be common to almost all Pyruvate Kinases wherein the closing of B domain on the active site cleft is crucial for catalysis. *(Schormann et al., 2019)* Although the catalytic mechanism is common to all pyruvate kinases, they are known to show evolutionarily divergent allosteric strategies. *(Morgan et al., 2014)*. The variation captured in PC2 from two distant loops was therefore interesting, although its contribution was very less, 8.36% as compared to PC1. The loops showing variance are found to be the FBP-activating loop from the C domain and the B-loop from the B domain, which also had a high local RMSD value as seen from our structural analysis. Therefore, we thought that the variance captured by PC2, although small, might be significant in the context of FBP-mediated allostery in human PKM2.

Although MD simulation data is quite insightful when it comes to probing the molecular motions of a protein by sampling the conformational space available to it; can become computationally expensive with increase in the size of the protein. Coarse-grained elastic network models with simplified force fields have been shown to effectively capture these equilibrium fluctuations. *(Bahar et al., 1999)*. Thus to further understand the differences between PKM1 and PKM2 with respect to their conformational flexibility and to gain insights into allostery in PKM2, we performed a Normal Mode Analysis (NMA) with an Anisotropic Network Model (ANM) in ProDy *(Bakan et al., 2011)*. NMA essentially describes the slow motions adopted by a protein around an equilibrium position and these states have shown to be biologically and functionally relevant in several instances. *(Eyal et al., 2011, Ayyildiz et al., 2020 & Hu 2021)* Substrate bound PKM1 and PKM2 (*3SRF and 3BJT - apo for FBP*) were subjected to NMA with ANM using ProDy with default parameters. The first 10 modes of PKM1 and PKM2 were found to be comparable on a one-on-one basis from RMSIP values, showing a perfect subspace overlap of 1 in each case. The first two modes did not show any significant differences in the dynamics of PKM1 and PKM2 [Figure 8]. Interestingly, the residue-wise squared fluctuation plot showed differences in the overall trend of the peaks upon comparing PKM1 with PKM2 [Figure 9].

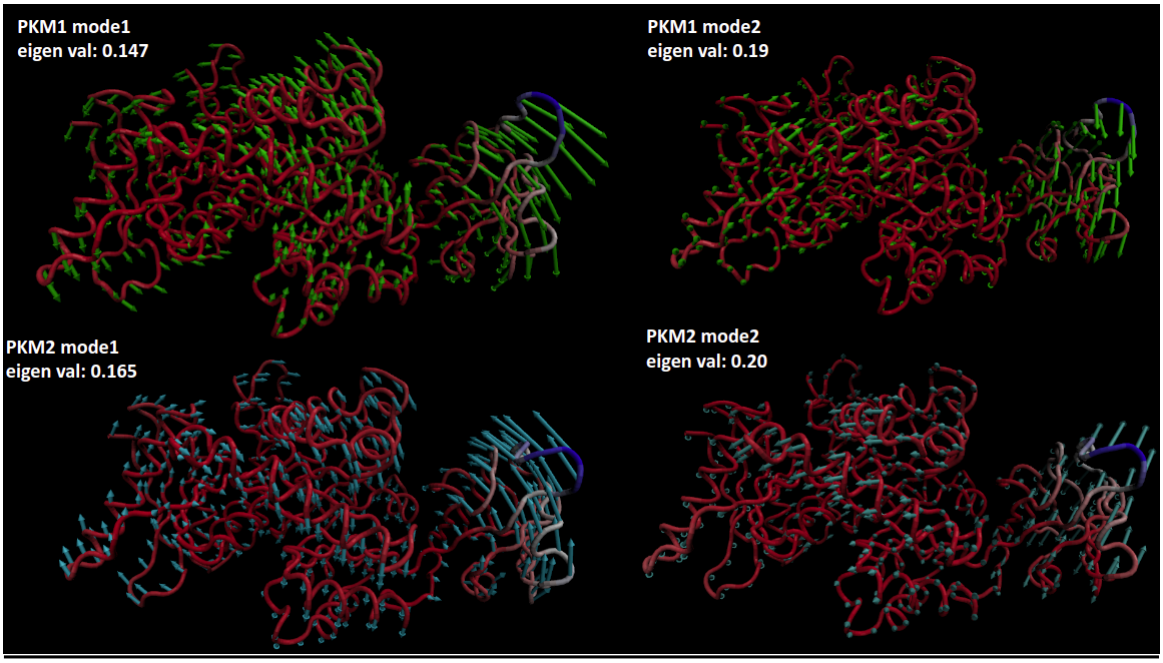

Figure 8: Normal modes of PKM1 and PKM2

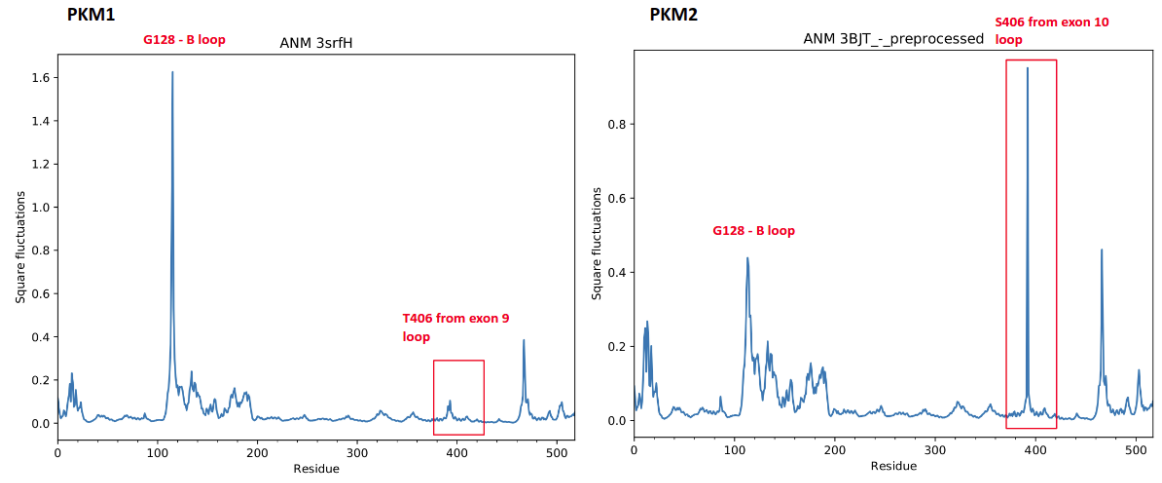

Figure 9: Plots of residue-wise squared fluctuations in PKM1 and PKM2

While the residue with the highest fluctuation value was G128 from the B-loop, in case of PKM1, S406 (T406 in M1) from the Exon-loop was found to have the highest fluctuation value in case of PKM2. Although, to a lesser extent, G128 also showed a peak in case of PKM2 however, the same was not true in case of T406 in PKM1, which was found to be

quite rigid. Importantly, T/S406 is a part of exon 9/10. To further understand how fluctuation values of these two residues changed with change in the bound-state of PKM2, we performed NMA using ANM for monomers of PKM2 apo *(1ZJH)*, PKM2 holo (+substrate, +FBP - *3SRD)*. From the plots showing residue wise squared fluctuations, we observed that going from PKM2 apo to PKM2 holo the squared fluctuation value of S406 from the Exon loop decreases from 1.1 to 0.3 (but is still greater than 0.1 in PKM1). Conversely the squared fluctuation value of G128 from the B-loop increases from 0.35 to 0.7 (but is still lesser than 1.6 in PKM1). [Table 4] This suggested that these residues might be important players in the FBP-mediated allostery and oligomerization of PKM2.

Table 4: Fluctuation values of residues from the B-loop, Exon-loop and FBP-activating loop

Sq. fluctuations	M1 holo (+substrate)	M2 apo	M2 holo (+substrate, -FBP)	M2 holo (+substrate +FBP)
G128	1.623	0.252	0.416	0.731
S406	0.106	1.084	0.952	0.391
G518	0.099	0.174	0.089	0.056

Drawing insights from the information of per residue contribution to structural variance captured by PC2, we decided to check whether these residues were physically connected via allosteric paths. OHM's perturbation propagation algorithm for finding allosteric paths was the method of choice because of its exhaustive nature of predicting the most likely allosteric path in static structures. The choice of source (allosteric site) and sink (active site) residues was key as it determined the validity of the generated paths. We knew from literature that B domain closure is crucial to catalysis and this was also captured in our

PCA and ANM results. From the residue-wise squared fluctuation values given by ANM, we were able to select residues of interest in the context of allostery. Additionally, *Dombrauckas et al., (2005)* had proposed a role of the FBP activating loop and the significant difference in its backbone and side-chain orientation in PKM2 as compared to PKM1. We therefore hypothesized that the movement of the FBP-activating loop could be important for B-domain closure and stabilizing the exon loop fluctuation in PKM2. Taking into account all of the above we sought to find allosteric paths between three pairs of source-sink residues; being G518-G128, S406-G128 and G518-S406. We used PKM2 apo *(1ZJH)*, PKM2 holo (+substrate, +FBP *3SRD*), PKM2 holo (+substrate, -FBP *3BJT*). Although PKM1 is non-allosteric we still decided to check if there was a path available connecting the respective source and sink residues. In all the three source and sink pairs we found residues connecting the allosteric site with the active site [Figure 10].

Overall, we observed that the weight of the paths connecting the source and sink gradually decreased as PKM2 attained a tetramer, favourable state for catalysis, from M2 apo to M2 holo. [Table 5] PKM1 despite being non-allosteric still showed allosteric paths with the path weights being comparable with those found in case of PKM2 holo.

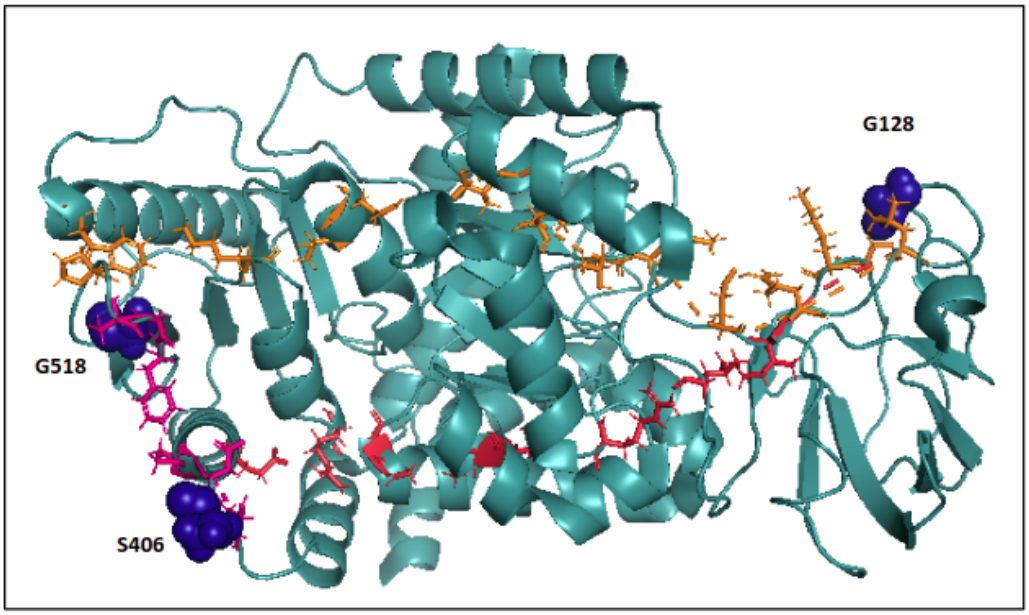

Figure 10: Allosteric Paths

Table 5: Allosteric Path weights signifying their likelihood in PKM2 and PKM1

	S406-G128 (path wt.)	S518-G128 (path wt.)	G518-S406 (path wt.)
PKM2 apo	0.13	0.13	0.13
PKM2 (+subs, -FBP)	0.07	0.07	0.11
PKM2 (+subs, +FBP)	0.02	0.02	0.11
PKM1 (+subs)	0.02	0.02	0.03

4. Discussion

In the present analysis, we investigated the effect of MXEs on the structure-function dynamics hPKM isoforms (M1 and M2). The comparison of structures and residue interaction networks s of PKM1 and PKM2 were unable to capture any significant differences at the global level of these proteins that can provide insight into their function shifts. Subsequently, we employed PCA to multiple structures of PKM and identified regions of proteins responsible for large conformational change, which were mostly B-domain. This is already known to contribute in enzyme function. Interestingly, we could identify relatively lower contribution of FBP binding loop and exon region fluctuations in PC modes. Further, ANM based study provide residue-wise fluctuations to slower modes of protein and we identified residues G128 and S406 having differing fluctuations between PKM1 and PKM2. This possibly could be allosterically coupled structural elements. The change in squared fluctuation values of G128 and S406 with a change in the bound-state of PKM2 further suggested their importance in the context of allostery. Allosteric paths found using these unusual non-binding site residues as source and sink were found to show an interesting trend in the path weights, again, depending on the bound-state of PKM2. The likelihood of allosteric paths was found to decrease gradually with increase in the likelihood of attaining the tetrameric state i.e. upon going from an unbound to a completely bound state. The weights of allosteric paths found in the fully bound state of PKM2 tetramer were comparable to those of PKM1. Therefore, this gives a plausible explanation as to how FBP-mediated allostery could be affecting PEP binding and oligomerization in PKM2 and absence of this effect is not seen in the case of PKM1.

Our results from ANM suggest that the exon 10 loop dynamics resulting from the high squared fluctuation value of S406, does not allow the formation of a stable C-C' interface and therefore a functional tetramer of PKM2, under normal circumstances. We think that FBP-binding, which directly affects the conformation of the FBP-activating loop as suggested by *Dombrauckas et al.,(2005);* might be having an effect on stabilizing the exon 10 loop as observed by the decrease in its squared fluctuation and the weight of allosteric path in between G518 and S406, in the completely bound PKM2. This FBP-mediated stabilization might be allowing the formation of a functional tetramer.

Additionally, our results also suggest the plausible roles of S406 from the exon 10 loop and G518 from the FBP-activating loop in B domain closure by affecting B-loop dynamics via G128. The missing peak of 406 residue position in PKM1 from exon 9 loop along with the low weights of allosteric paths comparable to fully-bound PKM2, found in it also suggest that because it has already attained the tetrameric state, the allosteric coupling might not be required despite there being a path physically connecting these distant sites.

Therefore, our analyses present a probable explanation for the functional shift in isoforms by change in the underlying loop dynamics of the PKM isoforms as a result of the MXE splicing. The allosteric paths provide evidence for physical connectedness between distant loops in both the isoforms and also suggests a plausible mechanism of the effect of FBP-mediated allostery on enzyme oligomerization.

Bibliography

- Bohr KCH, Krogh A. (1904) Ueber einen in biologischer beziehung wichtigen Einfluss, den die kohlensäurespannung des blutes auf dessen sauerstoffbindung übt. Skandinavisches Archiv Für Physiologie 16: 402–412.
- Monod J, Wyman J, Changeux JP (1965) On the Nature of Allosteric Transitions: A Plausible Model. J Mol Biol 12: 88–118.
- Koshland DE Jr., Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5: 365–385.
- Cooper A, Dryden DT (1984) Allostery without conformational change. A plausible model. Eur Biophys J 11: 103–109.
- Ma, B., Kumar, S., Tsai, C. J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein engineering*, *12*(9), 713–720.
- Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins.
- Fenton A. W. (2008). Allostery: an illustrated definition for the 'second secret of life'. Trends in biochemical sciences, 33(9), 420–425.
- Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? Proteins. 2004 Nov 15;57(3):433-43.
- David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol. 2014;1084:193-226.
- Grant, B.J., A.P.D.C Rodrigues, K.M. Elsawy, A.J. Mccammon, and L.S.D. Caves. 2006. “Bio3d: an R Package for the Comparative Analysis of Protein Structures.” *Bioinformatics* 22: 2695–2696.
- Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J. 2001 Jan;80(1):505-15.
- Dokholyan, N. (2016) Controlling Allosteric Networks in Proteins. *Chemical Reviews* 116, 11, 6463–6487
- Bakan A, Meireles LM, Bahar I *ProDy*: Protein Dynamics Inferred from Theory and Experiments *Bioinformatics* **2011** 27(11):1575-1577.

- Wang, J., Jain, A., McDonald, L. R., Gambogi, C., Lee, A. L., & Dokholyan, N. V. (2020). Mapping allosteric communications within individual proteins. *Nature communications*, *11*(1), 3862.
- Srebrow A, Kornblihtt AR. The connection between splicing and cancer. J Cell Sci. 2006 Jul 1;119(Pt 13):2635-41.
- Jin Y, Dong H, Shi Y, Bian L. Mutually exclusive alternative splicing of pre-mRNAs. Wiley Interdiscip Rev RNA. 2018 May;9(3):e1468.
- Kondrashov, F. A., & Koonin, E. V. (2001). Origin of alternative splicing by tandem exon duplication. *Human Molecular Genetics*, 10, 2661–2669.
- Hatje, K., Rahman, R. U., Vidal, R. O., Simm, D., Hammesfahr, B., Bansal, V., … Kollmar, M. (2017). The landscape of human mutually exclusive splicing. *Molecular Systems Biology*, 13, 959.
- Menon R, Roy A, Mukherjee S, Belkin S, Zhang Y, Omenn GS. Functional implications of structural predictions for alternative splice proteins expressed in Her2/neu-induced breast cancers. J Proteome Res. 2011 Dec 2;10(12):5503-11.
- Li HD, Omenn GS, Guan Y. A proteogenomic approach to understand splice isoform functions through sequence and expression-based computational modeling. Brief Bioinform. 2016 Nov;17(6):1024-1031.
- Mishra SK, Muthye V, Kandoi G. Computational Methods for Predicting Functions at the mRNA Isoform Level. Int J Mol Sci. 2020 Aug 8;21(16):5686.
- Dombrauckas JD, Santarsiero BD, Mesecar AD, (2005) Structural Basis for Tumor Pyruvate Kinase M2 Allosteric Regulation and Catalysis. Biochemistry 2005 *44* (27), 9417-9429
- Yuan M, McNae IW, Chen Y, Blackburn EA, Wear MA, Michels PAM, Fothergill-Gilmore LA, Hupp T, Walkinshaw MD. An allostatic mechanism for M2 pyruvate kinase as an amino-acid sensor. Biochem J. 2018 May 31;475(10):1821-1837.
- Yan M, Chakravarthy S, Tokuda JM, Pollack L, Bowman GD, Lee YS. Succinyl-5-aminoimidazole-4-carboxamide-1-ribose 5'-Phosphate (SAICAR) Activates Pyruvate Kinase Isoform M2 (PKM2) in Its Dimeric Form. Biochemistry. 2016 Aug 23;55(33):4731-6.

- Morgan HP, O'Reilly FJ, Wear MA, O'Neill JR, Fothergill-Gilmore LA, Hupp T, Walkinshaw MD. M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. Proc Natl Acad Sci U S A. 2013 Apr 9;110(15):5881-6.
- Chen TJ, Wang HJ, Liu JS, Cheng HH, Hsu SC, Wu MC, Lu CH, Wu YF, Wu JW, Liu YY, Kung HJ, Wang WC. Mutations in the PKM2 exon-10 region are associated with reduced allostery and increased nuclear translocation. Commun Biol. 2019 Mar 15;2:105.
- Wang P, Sun C, Zhu T, Xu Y. Structural insight into mechanisms for dynamic regulation of PKM2. Protein Cell. 2015 Apr;6(4):275-287.
- Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature. 2008 Mar 13;452(7184):230-3.
- UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004 Oct;25(13):1605-12.
- Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. "A graphical interface for the FoldX forcefield." *Bioinformatics*. 2011;27(12):1711-2
- NACCESS is (c) S. Hubbard and J. Thornton 1992-6
- Macpherson, J. A., Theisen, A., Masino, L., Fets, L., Driscoll, P. C., Encheva, V., Snijders, A. P., Martin, S. R., Kleinjung, J., Barran, P. E., Fraternali, F., & Anastasiou, D. (2019). Functional cross-talk between allosteric effects of activating and inhibiting ligands underlies PKM2 regulation. *eLife*, *8*, e45068.
- Tang Q, Fenton AW. Whole-protein alanine-scanning mutagenesis of allostery: A large percentage of a protein can contribute to mechanism. Hum Mutat. 2017 Sep;38(9):1132-1143.
- Schormann N, Hayden KL, Lee P, Banerjee S, Chattopadhyay D. An overview of structure, function, and regulation of pyruvate kinases. Protein Sci. 2019 Oct;28(10):1771-1784.

- Morgan HP, Zhong W, McNae IW, Michels PA, Fothergill-Gilmore LA, Walkinshaw MD. Structures of pyruvate kinases display evolutionarily divergent allosteric strategies. R Soc Open Sci. 2014 Sep 24;1(1):140120.
- Eyal E, Dutta A, Bahar I. Cooperative dynamics of proteins unraveled by network models. Wiley Interdiscip Rev Comput Mol Sci. 2011 May-Jun;1(3):426-439.
- Ayyildiz M, Celiker S, Ozhelvaci F, Akten ED. Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets. Front Mol Biosci. 2020 May 14;7:88.
- Hu G. Identification of Allosteric Effects in Proteins by Elastic Network Models. Methods Mol Biol. 2021;2253:21-35.
- Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, Napolitano C, Schwartz PJ, Joseph RM, Condouris K, Tager-Flusberg H, Priori SG, Sanguinetti MC, Keating MT. Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell. 2004 Oct 1;119(1):19-31.
- David CJ, Chen M, Assanah M, Canoll P, Manley JL. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. Nature. 2010 Jan 21;463(7279):364-8.
- Coutsias EA, Wester MJ. RMSD and Symmetry. J Comput Chem. 2019 Jun 5;40(15):1496-1508.
- Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. J Mol Biol. 1999 Jan 22;285(3):1023-37.
- Piovesan D, Minervini G, Tosatto SC. The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res. 2016 Jul 8;44(W1):W367-74.