

# **Structural and Biophysical Characterization of Human Pyruvate Dehydrogenase Multi-Enzyme Complex**

## **Dissertation**

for the award of the degree

“Doctor rerum naturalium (Dr. rer. nat.)”

of the Georg-August-Universität-Göttingen

within the doctoral program Biomolecules: Structure-Function-Dynamics!

of the Göttingen Graduate School for Neuroscience, Biophysics and Molecular

Biosciences (GGNB)

submitted by

**Sabin Prajapati**

born in Kathmandu, Nepal

Göttingen 2016

### **Members of the Thesis Committee:**

Prof. Dr. Kai Tittmann, Reviewer 1	Department of Molecular Enzymology Georg-August-Universität-Göttingen
Prof. Dr. Peter Rehling, Reviewer 2	Institute of Cell Biochemistry Georg-August-Universität-Göttingen
Associate Prof. Dr. Iwan Schaap	Institute of Biological Chemistry, Biophysics and Bioengineering Heriot-Watt University, Edinburgh

### **Further members of the Examination Board:**

Prof. Dr. Ralf Ficner	Department of Molecular Structural Biology Georg-August-Universität-Göttingen
Prof. Dr. Holger Stark	Department of Structural Dynamics Max Planck Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Jörg Enderlein	III. Institute of Physics Georg-August-Universität-Göttingen

**Date of oral examination: 29.11.2016**

## **Affidavit**

I hereby, declare that this PhD thesis “Structural and Biophysical Characterization of Human Pyruvate Dehydrogenase Complex” has been written independently with no other aids or sources than quoted.

Sabin Prajapati

18<sup>th</sup> October, 2016

Göttingen, Germany

## Contents

<b>Acknowledgements</b>	1
<b>List of Abbreviations</b>	3
<b>List of Figures</b>	7
<b>List of Tables</b>	10
<b>Abstract</b>	11
<b>1. Introduction</b>	13
1.1. Multi-enzyme complexes	13
1.2. General overview of pyruvate dehydrogenase complex	14
1.3. Peripheral subunits of human pyruvate dehydrogenase enzyme complex	17
1.3.1. Human pyruvate dehydrogenase ( <i>hE1</i> )	17
1.3.1.a. Structure and mechanism	17
1.3.1.b. Binding of <i>hE1</i> to its cognate binding site at <i>hE2</i>	19
1.3.2. Human dihydrolipoamide dehydrogenase ( <i>hE3</i> )	20
1.3.2.a. Structure and mechanism	20
1.3.2.b. Binding of <i>hE3</i> to its cognate binding site at E3BP	23
1.4. Core subunits of human pyruvate dehydrogenase enzyme complex	25
1.4.1. Human dihydrolipoamide acetyltransferase ( <i>hE2</i> )	25
1.4.1.a. Lipoyl domains	27
1.4.1.b. Linkers	30
1.4.1.c. E1 binding domain (E1BD)	31
1.4.1.d. Inner catalytic (IC) domain	32
1.4.2. Human dihydrolipoamide dehydrogenase binding protein (E3BP)	34
1.5. Organization of the human PDH complex	36

1.6. Conformational plasticity of PDHc core	37
1.7. Regulation of the activity of human PDH complex and health implications	39
1.8. Motivation	42
<b>2. Materials and Methods</b>	<b>43</b>
2.1. Materials	43
2.1.1. Chemicals	43
2.1.2. Instruments	45
2.1.3. Commodities	48
2.1.4. Primers	50
2.1.5. Media compositions	50
2.1.6. Software and web resources	52
2.2. Methods	53
2.2.1. Molecular biology	53
2.2.1.1. Determination of DNA concentration	53
2.2.1.2. Polymerase chain reaction (PCR)	53
2.2.1.3. Agarose gel electrophoresis (AGE)	54
2.2.1.4. Restriction digestion	54
2.2.1.5. Ligation	55
2.2.1.6. Bacterial transformation	55
2.2.1.7. DNA sequencing	55
2.2.2. Protein chemistry	56
2.2.2.1. Determination of protein concentration	56
2.2.2.2. Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)	56
2.2.2.3. Over expression of proteins	57

2.2.2.3.1. High cell density fermentation for <i>hE1</i>	57
2.2.2.3.2. Auto-induction of <i>hE3</i> , <i>hE2</i> , <i>hE2/E3BP</i> , <i>KD-hE2</i> , <i>eGFP-hE2</i> and <i>SNAP-hE2</i> (all in pET28a expression vector)	58
2.2.2.4. Purification of proteins	59
2.2.2.4.1. Purification of <i>hE1</i>	59
2.2.2.4.2. Purification of <i>hE3</i>	59
2.2.2.4.3. Purification of <i>hE2</i> , <i>hE2/E3BP</i> , <i>KD-hE2</i> , <i>eGFP-hE2</i> and <i>SNAP-hE2</i>	60
2.2.2.5. Proteolysis of <i>hE2</i> and <i>hE2/E3BP</i> core	61
2.2.2.6. In-vitro reconstitution of <i>hPDHc</i> , <i>hE1-hE2/E3BP</i> sub-complex and <i>hE3-hE2/E3BP</i> sub-complex	62
2.2.3. Activity assays	62
2.2.3.1. DCPIP assay	62
2.2.3.2. PDHc assay	64
2.2.4. FAD saturation test	66
2.2.5. Isothermal titration calorimetry (ITC)	66
2.2.5.1. Background	66
2.2.5.2. Experiment	68
2.2.6. Small angle X-ray scattering (SAXS)	68
2.2.7. Atomic force microscopy (AFM)	69
2.2.7.1. Background	69
2.2.7.2. Experiment	71
2.2.8. Electron microscopy (EM)	72
2.2.8.1. Negative staining electron microscopy (neg-EM)	72
2.2.8.2. Single molecule electron cryo-microscopy (Cryo-EM)	73
2.2.9. Direct stochastic optical reconstruction microscopy (dSTORM)	74

2.2.9.1. Background	74
2.2.9.2. Experiment	75
2.2.9.2.1. Sample preparation	75
2.2.9.2.2. dSTORM measurement	76
2.2.10. Stimulated emission depletion microscopy (STED)	76
2.2.11. Mass spectrometry (MS)	77
2.2.11.1. MS for protein band identification	77
2.2.11.2. Cross linking mass spectrometry (XL-MS)	77
2.2.11.2.1. Sample preparation	77
2.2.11.2.2. LC-MS	78
2.2.11.2.3. Quantification of crosslinks	79
2.2.11.3. Native mass spectrometry (native-MS)	79
<b>3. Results</b>	<b>80</b>
3.1. Quality check for <i>hE1</i> , <i>hE2/E3BP</i> and <i>hE3</i>	80
3.2. Proteolysis of <i>hE2</i> and <i>hE2/E3BP</i> core	83
3.3. In-vitro reconstitution of <i>hPDHc</i> and sub-complexes	85
3.4. Steady state kinetics of <i>hPDHc</i>	86
3.5. ITC-binding experiment of <i>hE2/E3BP</i> core with <i>hE3</i>	88
3.6. Size estimation of <i>hPDHc</i> core and its variants by SAXS	89
3.7. Size and stiffness estimation of <i>hPDHc</i> core and its variants by AFM	90
3.8. dSTORM for single molecule analysis	92
3.9. Estimation of size variation of <i>hPDHc</i> core by STED	95
3.10. Electron microscopy	96
3.10.1. Truncated cores	96

3.10.2. <i>hE2/E3BP</i> cores	99
3.10.3. <i>hPDHc</i> and sub-complexes	103
3.11. Mass spectrometry	104
3.11.1. Cross-linking mass spectrometry	104
3.11.1.1. Crosslinks within the core-forming domain	105
3.11.1.2. Inter lipoyl domain interaction	109
3.11.1.3. Interactions of lipoyl domains with <i>hE1</i> and <i>hE3</i> proteins	111
3.11.1.4. Position of <i>hE3</i> protein	112
3.11.1.5. Change in lipoyl arm dynamics in CoA bound state	114
3.11.2. Native MS of <i>hE2/E3BP</i> cores and <i>hPDHc</i>	116
<b>4. Discussion</b>	120
4.1. Structure of <i>hE2/E3BP</i> core and its dynamics	120
4.2. Organization of the <i>hPDHc</i> core and its implications on its function	123
4.3. Functionally nonequivalent lipoyl domains and their mode of networking	127
4.4. Effect of CoA binding to the <i>hPDHc</i> : a substrate dictates large conformational changes	128
4.5. Active site layers in <i>hPDHc</i>	133
<b>5. Summary</b>	136
<b>6. Outlook</b>	138
<b>7. References</b>	140
<b>8. Appendix</b>	159
<b>CURRICULUM VITAE</b>	184



## **Acknowledgments**

I would like to express my deepest gratitude to Prof. Dr. Kai Tittmann, who supervised me over the course of this PhD thesis. His scientific aptitude always inspired me to think deeper and challenge myself to question my conclusions regarding any scientific observations I made. I am also thankful to him for being ever supportive in any administrative or personal problems, like a perfect boss.

I am also thankful to Prof. Dr. Peter Rehling and Associate Prof. Dr. Iwan Schaap, for accepting to be 2<sup>nd</sup> and 3<sup>rd</sup> PhD thesis committee members, actively participating in the committee meetings and guiding the project to a positive direction.

Next, I would like to thank Prof. Dr. Holger Stark, Prof. Dr. Jörg Enderlein and Prof. Dr. Ralf Ficner for kindly accepting to be members of examination board, despite their busy schedules.

I am indebted to Dr. Kathrin S. Tittmann and Dr. Florian Brodhun, former post doctorates from my lab, for guiding me at various stages of the project and help me optimize some of the protein overexpression and purification protocols.

As this thesis work was the culmination of efforts from many laboratories, I would like to thank following collaborators for providing their valuable time in conducting experiments, data analysis and also sharing me their expertise on their methods:

Dr. Mitja Platen for AFM and data analysis.

Dr. Oliva Saldanha for SAXS and data analysis.

Dr. David Haselbach for electron microscopy and data analysis.

Dr. Carla Schmidt for crosslinking MS, native MS and their data analysis.

Dr. Qui Van and Simon Stein for dSTORM and data analysis.

Dr. Nickels Jensen for STED and data analysis.

Dr. Oliver Valerius for validating E2, E3BP and their proteolytically derived truncated versions with MS.

Dr. Achim Dickmanns for ThermoFluor assay.

I would also like to acknowledge Christian Sommereisen for accepting to be my student in a scientific project and finishing his BSc thesis successfully.

I would like to thank Prof. Dr. Ralf Ficner and Prof. Dr. Ivo Feussner for kindly allowing me to use their lab facilities.

I would like to thank Dr. Piotr Neumann for all the discussions we had regarding protein crystallization and structures. Also, for testing some of the crystals I got in protein crystallization trials.

Also, I would like to thank all the past and current lab members of Department of Molecular Enzymology for assisting me in routine lab works and also organizing time-to-time social activities. I would like to specially thank Fabian N.R. von Pappenheim and Oliver Kupski for proof reading the draft of this thesis in a short notice and ShaoBo Dai for his advices in editing protein structures in PyMol.

Last but not the least, I cannot thank enough my family members for all their love and care. And I thank my wife, Nidhi Subhashini, for her constant support.

Finally, I would like to dedicate this thesis to all the victims of a deadly earthquake on 25<sup>th</sup> April 2015 in Nepal.

**List of abbreviations**

Å	Angstrom
Ala	Alanine
<i>A. vinelandii</i>	<i>Azotobacter vinelandii</i>
AFM	Atomic force microscopy
BisTris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BSA	Bovien serum albumin
<i>bsPDHc</i>	<i>Bacillus stearothermophilus</i> pyruvate dehydrogenase complex
<i>B. stearothermophilus</i>	<i>Bacillus stearothermophilus</i>
C	Cysteine
° C	degree Celsius
CoA	Coenzyme A
cv	Column volume
DCPIP	2,6-Dichlorophenolindophenol
dSTORM	Direct stochastic optical reconstruction microscopy
DTE	Dithioerythritol
DTT	Dithiothreitol
E1	Pyruvate dehydrogenase
E2	Dihydrolipoamide acetyltransferase
E3	Dihydrolipoamide dehydrogenase
E1BD	Pyruvate dehydrogenase (E1) binding domain
E3BD	Dihydrolipoamide dehydrogenase (E3) binding domain

List of abbreviations

E3BP	Dihydrolipoamide dehydrogenase (E3) binding protein
ecPDHc	<i>Escherichia coli</i> pyruvate dehydrogenase complex
<i>E. coli</i>	<i>Escherichia coli</i>
EC	Enzyme Commission number
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
F	Phenyl alanine
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
Fig.	Figure
FMN	Flavin adenine mononucleotide
GraFix	Gradient fixation
<i>hE1</i>	Human pyruvate dehydrogenase
<i>hE2</i>	Human dihydrolipoamide acetyltransferase
<i>hE3</i>	Human dihydrolipoamide dehydrogenase
<i>hPDHc</i>	Human pyruvate dehydrogenase complex
HEPES	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HEThDP	Hydroxyethyl thiamine diphosphate
His	Histidine
hr	hour
I	Isoleucine
ITC	Isothermal titration calorimetry
K	Lysine
KDa	Kilo Dalton

List of abbreviations

LB	Lysogeny broth
LD	Lipoyl domain
MEA	2-mercaptoethylamine
MES	2-(N-morpholino)ethanesulfonic acid
mM	millimolar
ms	millisecond
µm	micrometer
µM	micromolar
µmoles	micromoles
µs	microsecond
min	minute
MS	Mass spectrometry
NAD	Nicotineamide adenine dinucleotide
nm	nanometer
nM	nanomolar
ns	nanosecond
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PDHc	Pyruvate dehydrogenase complex (generic)
PDK	Pyruvate dehydrogenase kinase
PDP	Pyruvate dehydrogenase phosphatase
PEG	Polyethylene glycol
pmoles	picomoles
PMSF	Phenyl methane sulfonyl fluoride

psi	Pound-force per square inch
Pro	Proline
R	Arginine
RPM	Revolutions per minute
SAM	S-adenosyl methionine
SANS	Small angle neutron scattering
SAXS	Small angle X-ray scattering
SBD	Subunit binding domain
SDS	Sodium dodecyl sulfate
sec	second
Ser	Serine
SiR	Silicon rhodamine
STED	Stimulated emission depletion
TEMED	N,N,N',N'- Tetramethylethylenediamine
ThDP	Thiamine diphosphate
Val	Valine
v/v	Volume/volume
w/v	Mass/volume

**List of figures**

Fig.1: Overall reaction scheme of PDHc.	15
Fig. 2: Crystal structure of human E1 and its catalytic mechanism.	18
Fig. 3: Binding of E1 $\beta$ chains (magenta) with E2-SBD (cyan) in <i>bs</i> PDHc (PDB 1W85).	20
Fig. 4: Physiological reaction of E3.	22
Fig. 5: Crystal structure of <i>h</i> E3 with bound FAD and NADH (PDB code 1ZMD).	23
Fig. 6: Crystal structure of <i>h</i> E3-E3BD complex (PDB code: 1zy8).	24
Fig. 7: Cartoon representation of modular architecture of human dihydrolipoamide acetyltransferase ( <i>h</i> E2).	26
Fig. 8: Comparing structures of LD1 and LD2 of <i>h</i> E2.	27
Fig. 9: Comparing linker regions of <i>h</i> E2.	31
Fig. 10: Sequence alignment between E1BD and E3BD of human PDHc with SBD of <i>B. stearrowthermophilus</i> PDHc.	32
Fig. 11: A general model of gating mechanism for entry of acylated-lipoamide in a dihydrolipoamide acyltransferases (E2) of $\alpha$ -keto acid dehydrogenase complexes.	34
Fig. 12: Cartoon representation of modular architecture of human E3BP and sequences of associated inter domain linkers.	35
Fig. 13: Breathing motion of PDHc core.	39
Fig.14: Schematic of an ITC calorimeter.	67
Fig. 15: Schematic of an AFM set up.	70
Fig. 16: STORM concept.	75
Fig. 17: A representative progress curve of DCPIP assay.	80
Fig. 18: FAD saturation test of <i>h</i> E3.	81
Fig. 19: An ITC-binding experiment between CoA and <i>h</i> E2/E3BP core.	82
Fig. 20: Sequence alignment of PDHc E2 of <i>Azotobacter vinelandii</i> ( <i>Av</i> E2) with <i>h</i> E2 and E3BP.	83
Fig. 21: Proteolytic treatment of <i>h</i> PDHc cores with thermolysin.	84
Fig. 22: In-vitro reconstitution of the <i>h</i> PDHc and sub-complexes from their component subunits.	85
Fig. 23: PDHc reaction progress curves (left) and their respective Michaelis-Menten plots (right).	87

Fig. 24: An ITC-binding experiment between <i>hE3</i> and <i>hE2/E3BP</i> core.	88
Fig. 25: A typical SAXS curve of scattering intensity ( <i>I</i> ) decay over momentum transfer ( <i>s</i> ).	89
Fig. 26: An example of AFM scans of <i>hPDHc</i> .	92
Fig. 27a: dSTORM images of the eGFP- <i>hE2</i> core bound to Alexa647 coupled anti-GFP nanobody.	93
Fig. 27b: Variation in sizes of the eGFP- <i>hE2</i> core complexed with Alexa 647 coupled anti-GFP nanobody.	94
Fig. 28: Images obtained by scanning same region by confocal microscope and STED microscope and the size variation upon CoA binding.	95
Fig. 29: Negative staining EM of proteolytically obtained truncated <i>hE2</i> cores.	96
Fig. 30: Cryo-EM density maps of truncated <i>hE2</i> core in the absence (grey) and presence (green) of CoA at their 3-fold axis.	97
Fig. 31: Negative staining EM micrograph of KD- <i>hE2</i> (+ GraFix + 2 mM CoA).	98
Fig. 32: Comparing EM density maps of <i>hE2/E3BP</i> core with truncated E2 core reported by Zhou and colleagues (2008) and fitting 60 E2 monomers to the hybrid EM map of <i>hE2/E3BP</i> core.	100
Fig. 33: Superposition of atomic models of <i>hE2</i> and E3BP core forming domains.	101
Fig. 34: Comparing structural model of <i>hE2</i> with those from bacterial PDHc E2s and reported truncated <i>hE2</i> .	102
Fig. 35: Negative staining EM micrograph of <i>hE1-hE2/E3BP</i> , <i>hE3-hE2/E3BP</i> and <i>hPDHc</i> .	103
Fig. 36: Negative staining EM micrograph of <i>hE2/E3BP</i> and <i>hE3-hE2/E3BP</i> .	104
Fig. 37: Crosslinking map of <i>hPDHc</i> .	105
Fig. 38: Inter-chain crosslinks in <i>hE2</i> and E3BP core forming IC domains.	108
Fig. 39: Crosslinks between distant lysine residues in E3BP IC domain.	109
Fig. 40: Crosslinks between lipoyl domains.	110
Fig. 41: Surface sampling by lipoyl domains of <i>hE2</i> and E3BP at <i>hE1</i> and <i>hE3</i> .	112
Fig. 42: Native MS spectra of <i>hE2/E3BP</i> core and <i>hPDHc</i> .	117
Fig. 43: Examples of some of the inter trimer associations possible in an isolated <i>hE2/E3BP</i> core and when it's reconstituted with <i>hE1</i> and <i>hE3</i> proteins.	119



Fig. 44: Dynamics along the length of E3BP.	122
Fig. 45: CoA binding to E2.	126
Fig. 46: CoA induced <i>h</i> PDHc core compaction and comparison with EM.	130
Fig. 47: Net effect of CoA binding to <i>h</i> PDHc.	132
Fig. 48: Docking of <i>h</i> E3 in a pentagonal opening of <i>h</i> E2/E3BP core.	134
Fig. 49: Radial separation of active sites in <i>h</i> PDHc.	135
Fig. 50: 12 % SDS-PAGE analysis of overexpression of various proteins used during this thesis work.	162
Fig. 51: Chromatograms of <i>h</i> E1 and <i>h</i> E3 purifications.	163
Fig. 52: SDS-PAGE analysis of the purification of <i>h</i> E1 and <i>h</i> E3 proteins.	164
Fig. 53: SDS-PAGE analysis of the purification of <i>h</i> E2/E3BP core.	165
Fig. 54: SDS-PAGE analysis of the purification of <i>h</i> E2 core.	165
Fig. 55: SDS-PAGE analysis of the purification of KD- <i>h</i> E2 core.	166
Fig. 56: SDS-PAGE analysis of the purification of SNAP- <i>h</i> E2 core.	166
Fig. 57: SDS-PAGE analysis of the purification of eGFP- <i>h</i> E2 core.	167
Fig. 58: An ITC-binding experiment between <i>h</i> E1 and <i>h</i> E2/E3BP core.	167
Fig. 59: Surficial details of <i>h</i> PDHc cores observed occasionally during AFM study.	168
Fig. 60: Negative staining EM images of truncated <i>h</i> E2/E3BP.	170

**List of tables**

Table 1: Subunits of PDHc with their respective cofactors and substrates.	16
Table 2: Comparison of number of lipoyl domains in well-defined E2 subunits between different organisms.	26
Table 3: Parameters of ITC experiment.	68
Table 4: Steady state kinetics parameters of reconstituted <i>h</i> PDHc with associated standard errors.	86
Table 5: Radius of gyration ( $R_g$ ) and maximum diameter ( $D_{max}$ ) estimates for human PDH complex cores and its variants.	90
Table 6: Heights of human PDH complex and different core variants.	91
Table 7: Stiffness of human PDH complex and different core variants.	91
Table 8: All the crosslinks in <i>h</i> E2 and E3BP core forming IC domains in the presence and absence of CoA.	107
Table 9: All the crosslinks between lipoyl domains of <i>h</i> PDHc with <i>h</i> E1 and <i>h</i> E3.	111
Table 10: All the crosslinks between inner hairpin loops with <i>h</i> E1 and <i>h</i> E3.	113
Table 11: Quantification of crosslinks between <i>h</i> E3 and inner hairpin loop of <i>h</i> E2 and E3BP.	114
Table 12: Quantification of all the crosslinks of lipoyl domains with SBD and lysine residues near the lipoyl entrance site at <i>h</i> E1, <i>h</i> E3 and the core.	115
Table 13: List of all the possible logical combinations of subunits that have similar masses to the one assigned for resolved peaks in the native MS spectra.	118
Appendix H: Crosslinks in <i>h</i> PDHc as detected in crosslinking MS	171
Appendix I: Crosslinks in <i>h</i> E2/E3BP core as detected in crosslinking MS	178

**Abstract**

Pyruvate dehydrogenase multi-enzyme complex (PDHc) is an assembly of multiple copies of four different proteins. Together they carry out the oxidative decarboxylation of pyruvate and generate acetyl-CoA and NADH, which are components of Krebs cycle, energy production and fatty acid biosynthesis in cells. Out of the four different subunits of PDHc, three are known to have distinct active sites and are found in PDHc from all organisms. The three catalytically important components are termed E1 (pyruvate dehydrogenase), E2 (dihydrolipoamide acetyltransferase) and E3 (dihydrolipoamide dehydrogenase). Exclusively in eukaryotic PDHc, an additional component, E3BP (E3 binding protein), is present whose role has been proposed to be for structural support of the PDHc assembly

This enzyme complex, which has already been studied for half a century, is a textbook prototype for substrate channelling between remotely located active sites in a multi-enzyme system. The individual active sites are spatially separated by at least few nanometers, and are coupled by highly flexible lipoyl arms of E2 and E3BP. The lipoyl arms consist of one or more lipoyl domains, each carrying covalently linked lipoamide groups. These lipoamide “swinging arms” need to visit all three active sites at E1, E2 and E3 in a sequential manner in order to complete a reaction cycle of PDHc. By structural design, the E2 core in prokaryotes and E2/E3BP core in eukaryotes make 24meric cube or 60meric pentagonal dodecahedron from which lipoyl arms and binding domains of E1 and E3 emanate outward.

In this PhD thesis, we focussed our attention on human PDHc and could elucidate the structural architecture of the core in details not achieved before. Firstly, we were able to calculate structural models for human E2/E3BP (*hE2/E3BP*) core from the cryo-EM density map at a resolution of  $\sim 6$  Å. Our data revealed that the published pseudo-atomic model of human E2 was in part erroneous. By integrating crosslinking MS data in our E2/E3BP structural model, we predict a hitherto unknown mode of structural dynamics that act along the length of core subunits, at least for E3BP. This mode is different from the ‘breathing motion that acts at the inter-trimer bridges orthogonal to the length of the subunits.

We proved that *hE2/E3BP* core most likely consists 40 *hE2* and 20 E3BP subunits, which can bind to 20 human E3 dimers (*hE3*). Furthermore, only *hE2* can bind to substrate coenzyme A (CoA). We observed that, unlike in prokaryotic PDHc, in *hPDHc*, only *hE1* appear to form outer shell while *hE3* can fluctuate between the outer shell and the core cavities. Also, in native MS experiments, we detected different

trimer arrangements of core subunits in isolated *hE2/E3BP* core versus fully assembled *hPDHc*, namely 2 *hE2* - 1 *E3BP* type and 1 *hE2* - 2 *E3BP* type, respectively. All of these observations indicate that in *hPDHc*, *hE2* and *E3BP* are not equally distributed in the core but rather with local patches. Wherever *hE2-hE1* is in excess, *E1/E2* reactions might be preferred. And at the patches where *E3BP-hE3* is dominant, *E3*-catalyzed regeneration of the lipamide cofactor mostly occur.

Another key discovery made during this thesis work was the large conformational changes in the lipoyl arms when *CoA* binds to *hE2*. By quantifying crosslinks detected in crosslinking MS, we could show that when *CoA* substrate is bound, the preferred destination of lipoyl domains are the core surface and *hE3*. Together with increased crosstalk between lipoyl domain of *E3BP* and *hE2*, it appears that *CoA* binding primes the *hPDHc* for specifically *E2* and *E3* reactions. To our knowledge, this is the first instance where substrate binding in any *PDHc* could be shown to impact the conformational landscape of lipoyl arm dynamics. In addition, the orientation of lipoyl domains while they come into proximity to each other seems to be conserved. This was an unexpected observation due to the very high flexibility of the lipoyl arms and the distances they have to travel in order to visit all the various active sites. These findings all indicate that movement of lipoyl arms to couple active sites in *PDHc* is not a multiple random coupling mechanism alone but is also impacted by substrate binding and catalysis synchronizing their movements for subsequent steps in the multi-step *PDHc* reaction.

**Keywords:** pyruvate dehydrogenase complex, architecture, *CoA*, movement of lipoyl arms

## 1. Introduction

### 1.1. Multi-enzyme complexes

To support life, two things are of utmost importance for a cell. First one is to contain important chemical compounds and biomolecules inside its protective space, thereby, enabling them to efficiently harness energy from sunlight and chemicals from its surroundings. Second one is to copy its genomic material and transfer to its daughter cells. To do both however, it needs to perform a myriad of chemical reactions, which could be catabolic or anabolic in nature. Evidence suggests that the responsibility of catalyzing these reactions have been delegated to proteins by ribonucleic acids (Cech 2012). Nature might have found proteins to be versatile as they have more diverse functional groups than nucleic acids, which arm them to tackle wide array of biochemical reactions.

These biocatalysts, which are known as 'enzymes', sometimes are part of a process, which needs several chemical transformations of a biomolecule (Reed 1966). Multiple enzymes activities are then required to finish the job, and occasionally they come together to associate into higher oligomers termed 'multi-enzyme complex'. One well-known example is fatty acid synthase (FAS) during biosynthesis of fatty acids. Depending on organisms, this whole process requires 7-8 completely different enzymatic steps. In prokaryotes and plants, a separate mono functional enzyme catalyzes each step, with their relative abundance directing the reaction (Anselmi, Grininger et al. 2010, Bukhari, Jakob et al. 2014). But in fungi, they are part of a large ensemble (fFAS), which is around 2.6 MDa in size. All the enzymatic activities are encoded in two peptide chains ( $\alpha$  and  $\beta$ ), 6 copies from each of them forming this colossal hetero-dodecameric enzyme machine. In metazoan, the idea remains the same for bringing active sites to close vicinity, but they are rather contained in, single peptide chain. Its homodimer (540 KDa) forms the fatty acid synthesizing scaffold (Leibundgut, Maier et al. 2008).

One intriguing question that arises here is, what could be the relative advantage of such assembly? The more complex an organism, the more tendencies there is to find such aggregations. The first advantage could be that important chemical intermediates won't diffuse into the cellular milieu till the final product is formed, making the whole pathway much more efficient (Reed 1966). This holds true even if the catalytic activity of the member enzymes is unaltered, since the active sites are optimally fixed in space. Other advantages could be the protection of intermediates from enzymes belonging to unrelated pathways and also easier regulations.

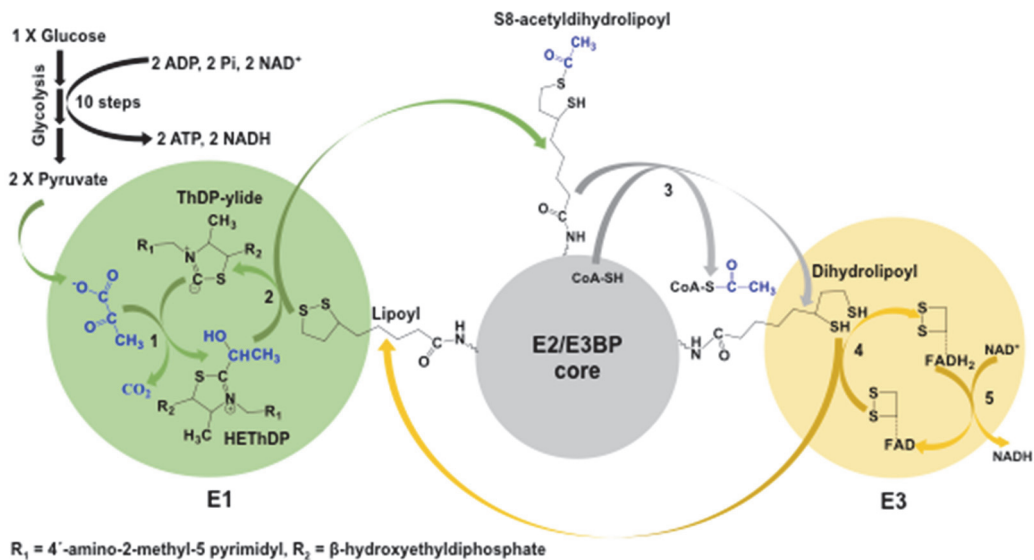
Quite often, the active sites of multi-enzyme complexes are at considerable distance from each other's. Use of extended appendages comprising flexible linkers, domains and prosthetic groups as means to couple distant active sites and channel substrates, seem to be a common feature in multi-enzyme complexes. In the FAS described above, the phosphopantetheine arm of acyl carrier protein does the job of carrying intermediates. In the pyruvate dehydrogenase complex (PDHc) which is the center of attention of this thesis, a lipoic acid covalently linked to lipoyl domain of a highly flexible swinging arm does the similar job (Reed and Hackert 1990).

## 1.2. General overview of pyruvate dehydrogenase complex

Pyruvate dehydrogenase complex (PDHc) is an enzyme complex dedicated to catalyze oxidative decarboxylation of pyruvate (Korkes, Delcampillo et al. 1951, Reed 1966). This process involves several steps, which are catalyzed at the active sites of three enzymes namely, pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and dihydrolipoamide dehydrogenase (E3) (see Fig. 1). The final products of the overall reaction are acetyl-CoA and reduced nicotinamide adenine dinucleotide (NADH). The acetyl-CoA contains a high-energy thioester bond between acetyl group and thiol group of CoA, which makes it an excellent acetyl donor to other biomolecules (Tittmann 2009). It is also funneled to the Krebs's cycle for complete combustion of acetyl group to produce further energy rich compounds like NADH and reduced flavin adenine dinucleotide (FADH<sub>2</sub>). These energy rich products are fed into the electron transport chain to maintain proton gradient across lipid membranes, which is a prerequisite for adenosine tri-phosphate (ATP) synthesis in a cell (Watt, Montgomery et al. 2010, Sazanov 2015). All in all, this complex channels the glycolytic end product i.e. pyruvate to cellular respiration.

PDHc is present in all organisms, which depend on aerobic respiration. It is localized in the cytoplasm of prokaryotic cells whereas, in eukaryotes, it is localized in the mitochondria and plastids (Reed 1966, Reid, Lyttle et al. 1975). Another important difference between prokaryotic and eukaryotic PDHc is their regulation. While activity of both of them is affected by the metabolic state of a cell indicated by a redox state and CoA/acetyl-CoA ratio, only eukaryotic PDHc is regulated by posttranslational modifications (Patel and Korotchikina 2006). Site-specific phosphorylation and de-phosphorylation of E1 component by pyruvate dehydrogenase kinase (PDK) and pyruvate dehydrogenase phosphatase (PDP) respectively, to inactivate and

reactivate PDHc, is one classic example (Harris, Bowker-Kinley et al. 2002, Patel and Korotchkina 2006).



**Fig.1: Overall reaction scheme of PDHc.**

The pyruvate molecule is decarboxylated at E1 (green sphere) and the remaining two-carbon moiety is covalently attached to E1 bound ThDP. A highly mobile lipoyl group from E2 (probably also E3BP) with a dithiolane ring, comes to the active site of E1 and takes away the acetyl group. Then, acetyl-lipoyl moves to the active site of E2 (grey sphere), which has an acetyl group. Here, acetyl-lipoyl donates the acetyl group to a CoA molecule. The product acetyl-CoA leaves PDHc while the reduced dihydrolipoyl has to go to active site of the E3 (yellow sphere) to get oxidized back to active lipoyl state. During this oxidation, the reducing equivalent moves from dihydrolipoyl to an intrinsic disulfide to the FAD cofactor and finally to NAD<sup>+</sup> to yield NADH. This completes a cycle for the movement of lipoyl-group, and is ready for next round of reaction.

Architecturally, the central component of PDHc is the E2 subunit, which oligomerizes into 24mer cube in gram-negative bacteria and plant and 60mer pentagonal dodecahedron in gram-positive bacteria (Reed and Hackert 1990). In mammals, nematodes and yeast, E2 also has a companion in the E3-binding protein (E3BP), together they form a dodecahedron, similar as in gram-positive bacteria (Stoops, Cheng et al. 1997, Vijayakrishnan, Kelly et al. 2010). There are however exceptions such as in the case of *Thermoplasma acidophilum*, where the E2s can arrange into 42meric oligomer showing mixed characteristics of both the cubic and dodecahedral symmetries (Marrott, Marshall et al. 2014). These structures formed by E2 alone or with E3BP are called 'cores'. Multiple copies of E1 and E3 proteins decorate these cores from outside by binding to their respective binding sites. In prokaryotes, E1 and E3 proteins compete with each other to bind to the common binding site at E2, which

is termed as subunit binding domain (SBD). In eukaryotes, E1 protein binds to the E1 binding domain (E1BD) of E2 and E3 binds to the E3 binding domain (E3BD) of E3BP (Harris, BowkerKinley et al. 1997).

Towards the N-terminus of E2 and E3BP, there are varying numbers of lipoyl domains, depending on source organism or organelle. In those lipoyl domains, a conserved lysine residue is covalently linked to lipoic acid to form a 14 Å long lipoamide. The rotational flexibility of the lipoyl group coupled with the flexibility of alanine and proline rich inter domain linkers; creates a highly flexible arm known as a 'lipoyl arm' (Reed and Hackert 1990). These arms are the real work force behind substrate channeling as they act as spatial intermediate mover to connect the distant active sites of E1, E2 and E3, which are at least 5 nm away from each other (Wagenknecht, Grassucci et al. 1991, Zhou, McCarthy et al. 2001).

Subunits	Cofactors	Substrates
Pyruvate dehydrogenase (E1)	Mg <sup>2+</sup> and ThDP	Pyruvate, lipoamide
Dihydrolipoamide acetyltransferase (E2)	Lipoic acid	S-8 acetyl-dihydrolipoamide and CoA
Dihydrolipoamide dehydrogenase binding protein (E3BP)	Lipoic acid	S-8 acetyl-dihydrolipoamide and CoA
Dihydrolipoamide dehydrogenase (E3)	FAD	NAD <sup>+</sup>

**Table 1: Subunits of PDHc with their respective cofactors and substrates.**

While E1, E2 and E3 are enzymatic components of all PDHc types, E3BP is regarded to have lost its acetyltransferase activity. However, its involvement in overall catalysis of human PDHc will be a subject of discussion in later chapters. Overall, PDHc utilizes magnesium ions and vitamin derived cofactors such as ThDP, lipoate and FAD in processing pyruvate to acetyl-CoA, which also involves CoA and NAD<sup>+</sup>. In addition to these subunits, the eukaryotic PDHc also consist of regulatory proteins PDK and PDP, which are lesser in copy number and are bound to lipoyl domains of E2 and E3BP.



### 1.3. Peripheral subunits of human pyruvate dehydrogenase enzyme complex

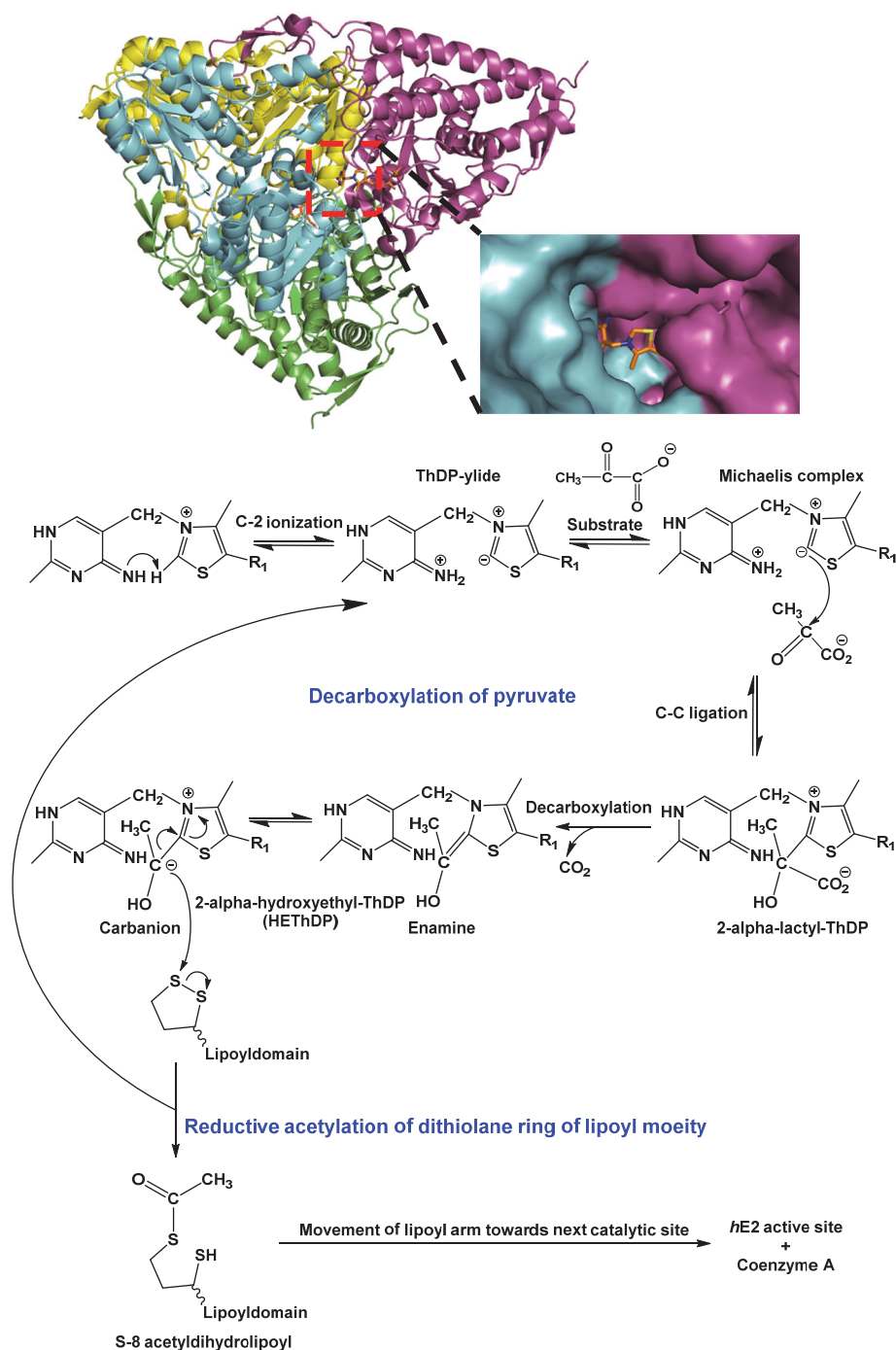
#### 1.3.1. Human pyruvate dehydrogenase (hE1)

Human pyruvate dehydrogenase (EC 1.2.4.1) is a decarboxylase consisting of loosely bound thiamine diphosphate (ThDP) and magnesium ion as cofactors. It is a heart shaped dimer of  $\alpha\beta$  hetero dimers (Ciszak, Korotchkina et al. 2003). In humans, the  $\beta$  chain is encoded by the *PDHB* gene (chromosome 3) while  $\alpha$  chain has two isoforms expressed in somatic cells by *PDHA1* (chromosome X) and in the testes by intron less *PDHA2* (chromosome 4) (Brown, Brown et al. 1990, Dahl, Brown et al. 1990).

##### 1.3.1.a. Structure and mechanism

It catalyzes the decarboxylation of pyruvate and the reductive acetylation of lipoamide (see Fig. 1, step 1) (Reed 1966). It possesses two active sites, one at each  $\alpha\beta$  dimer. The active site consists of ThDP and  $Mg^{2+}$  bound via several hydrogen bonds and metal ligation. The  $\alpha$  subunit is responsible for binding a  $Mg^{2+}$  and the diphosphate part of ThDP whereas the pyrimidine portion is bound to  $\beta$  chain, thereby forming a catalytic site between these two chains (Ciszak, Korotchkina et al. 2003).

Similar to other ThDP dependent enzymes, the reaction starts with an interaction of conserved glutamate (Glu-89,  $\beta$  chain) with N1' in the pyrimidine ring which increases the nucleophilicity of N4' group promoting deprotonation of C2 carbon (Kern, Kern et al. 1997, Fang, Nixon et al. 1998). This generates a C2-carbanion called ylide that attacks the carbonyl group of pyruvate forming tetrahedral lactyl-ThDP (L-ThDP) intermediate. The decarboxylation of L-ThDP results in hydroxyethyl-ThDP intermediate (a protonated form of enamine/2- $\alpha$ -carbanion tautomer). At this point, the lipoyl group from E2 or/and E3BP with oxidized dithiolane ring arrives at the active site. The S-8 of the lipoyl group is attacked by HEThDP carbanion while S-6 is protonated. It's been proposed that conserved histidine residues perform the role of general acid-base catalyst to protonate the lipoyl S-6 and deprotonate  $\alpha$ -OH of the HEThDP intermediate. The ThDP returns back into the ylide form, competent again for processing another pyruvate molecule, once the product from first round is cleared (see Fig. 2).



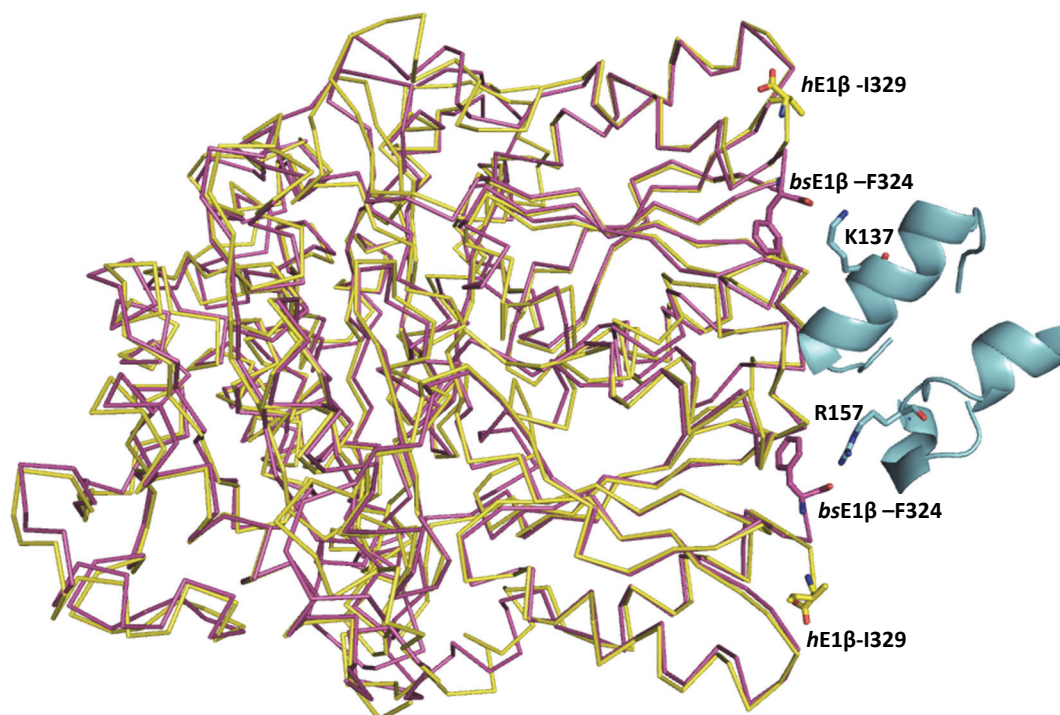
**Fig. 2: Crystal structure of human E1 and its catalytic mechanism.**

A heart shaped *hE1* molecule is a hetero dimer of  $\alpha\beta$  dimers. An active site pocket of one of the dimers (cyan-pink) is enlarged to show ThDP (PDB 1NI4). A C-2 ionization in a thiazolium ring of ThDP (1'-4'-imino-pyrimidine tautomer) is assisted by N4' of pyrimidine resulting in a biologically active ylide moiety. The substrate 'pyruvate' enters the active site pocket where its carbonyl group is attacked by C-2 carbanion. A decarboxylation step produces HEThDP intermediate, whose carbanion attacks the dithiolane ring of incoming lipoyl-group, producing S-8-acetyldihydrolipoyl, completing the reaction at E1. The acetyldihydrolipoyl group then moves to active site of E2 for further processing.

The first step of PDHc reaction, which is catalyzed by E1, is the rate-limiting step for PDHc. For example: the rate of reductive acetylation contributes to only 1 % of the rate of *E. coli* PDHc overall reaction, while the second major step i.e. formation of acetyl-CoA contributes to around 70 %, and the final step of re-oxidizing dihydrolipoyl group is at least 7 times faster than the overall reaction (Reed 1966). Similarly, in reconstituted human PDHc, the  $k_{cat}$  for overall reaction is  $\sim 50-70 \text{ s}^{-1}$  per active site of E1 while the rate constants for the steps such as the C2-ThDP ionization, formation of L-ThDP and subsequent decarboxylation step are much slower, with rate constants of  $51 \text{ s}^{-1}$ ,  $2.3 \text{ s}^{-1}$  and  $5.1 \text{ s}^{-1}$ , respectively (Seifert, Golbik et al. 2006). It has been proposed that once E1 proteins are organized at the exterior of the E2 core (prokaryote) or E2/E3BP core (eukaryote), the 3-D arrangement would enhance E1 activity probably by an increase in favorable interactions between lipoyl groups and HEThDP intermediates at the E1 active site, with an aid of lipoyl arm (Reed 1966).

### 1.3.1.b. Binding of *hE1* to its cognate binding site at *hE2*

*hE1* is bound to the E1BD of *hE2* in a one to one fashion and shows tight association with a low  $K_D$  of 9.47 nM. This involves a key electrostatic interaction between surficial *hE1* $\beta$ -D289 residue and *hE2*-K276 (Patel, Korotchkina et al. 2009). This pair is equivalent to a key interaction pair E1 $\beta$ -E285 with E2-R136 of the *B. stearothermophilus* PDDHc), for which the crystal structure of E1-SBD sub-complex has been published (Frank, Pratap et al. 2005). Two positively charged residues K137 and R157 of *bsPDHc* E2 are involved in making salt bridges with C-terminal F324 residue of *bE1* $\beta$  chains. But unlike in bacterial counterpart, the C-terminal I329 residue of *hE1*  $\beta$  chain does not have any suitable partner at E1BD of *hE2* to make salt bridges. The equivalent residues of these two positively charged residues of *bsPDHc* E2-SBD in E3BP protein of human PDHc, makes interactions with several residues of *hE3* (Ciszak, Makal et al. 2006, Patel, Korotchkina et al. 2009). The absence of interaction of these residues with C-terminal residues of *hE1*  $\beta$  chains is thought to confer monospecific binding between *hE1* and *hE2*-E1BD. It is noteworthy to mention again that in mammalian PDHc, separate subunit binding domains are present in E2 and E3BP to bind E1 and E3, respectively. But in bacterial PDHc, the same subunit-binding domain of E2 can bind to E1 and E3 proteins in a mutually exclusive manner, possibly via common interaction partners K137 and R157 at *bsPDHc* E2-SBD.



**Fig. 3: Binding of E1  $\beta$  chains (magenta) with E2-SBD (cyan) in *bsPDHc* (PDB 1W85).**

The *hE1*  $\beta$  chains (PDB 1NI4, yellow) are superimposed with the *bsE1*  $\beta$  chains. The two C-termini of *hE1*  $\beta$  chains (I329) seem to be too far from making meaningful contact with *hE2*-E1BD whereas the equivalent residues in *bsE1*  $\beta$  chains i.e. F324 can interact with K137 and R157 of *bsE2*-SBD. Adapted from (Patel, Korotchkina et al. 2009)

### 1.3.2. Human dihydrolipoamide dehydrogenase (*hE3*)

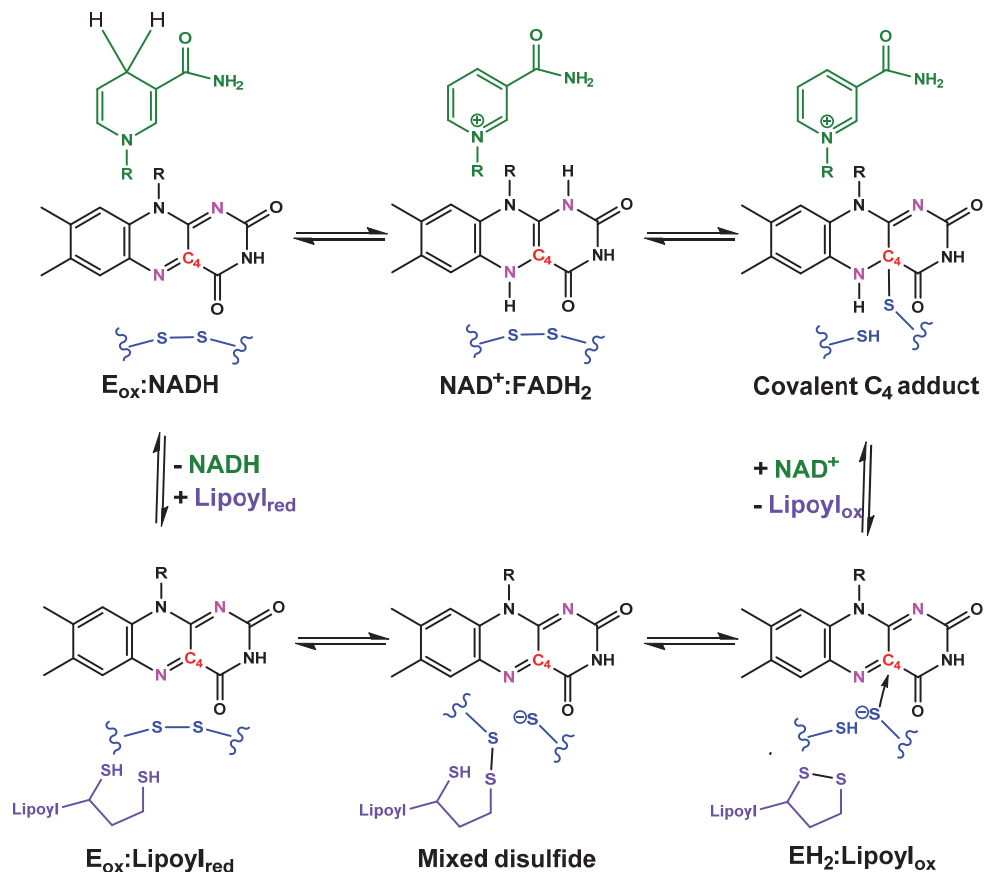
This enzyme is not only an exclusive partner to PDHc but also shares its utility with other related  $\alpha$ -keto acid dehydrogenase complexes like  $\alpha$ -ketoglutarate dehydrogenase complex and branched chain  $\alpha$ -keto acid dehydrogenase complex, as well as glycine cleavage system (Carothers, Pons et al. 1989). It is encoded by *DLD* gene (chromosome 7) in human. Its primary function in all of those complexes is to oxidize dihydrolipoamide residue (see Fig. 1) and then, reduce  $\text{NAD}^+$  to NADH (Patel, Nemeria et al. 2014).

#### 1.3.2.a. Structure and mechanism

*hE3* (EC 1.8.1.4) is oxido-reductase with tightly bound FAD and a redox active C80-C85 disulfide bridge, analogous to C45-C50 of *E. coli* E3. It is a homodimer with each monomer having one active site that consists of separate FAD and NAD binding domains (Brautigam, Chuang et al. 2005).

Several works have been carried out to understand the reaction mechanism of E3 in detail. With series of steady and pre-steady state kinetic experiments in E3 and their variants (Koike, Shah et al. 1960, Argyrou and Blanchard 2001, Argyrou, Blanchard et al. 2002, Argyrou, Sun et al. 2003), important active site residues, intermediates and charge-transfer complexes (see Fig. 4) have been characterized. Earlier, JK Reed, and colleagues (Reed 1973), showed in rat E3 that the reaction involves ping pong mechanism, meaning the first substrate dihydrolipoamide residue leaves after its oxidation and only then second substrate  $\text{NAD}^+$  would get involved.

In all E3 enzymes characterized so far, the reaction it catalyzes can be divided into reductive half and oxidative half reactions, which alternate between each other in every reaction cycle (Carothers, Pons et al. 1989). In the reductive half reaction, the reduced dihydrolipoyl-E2/E3BP binds to oxidized E3-FAD ( $\text{E}_{\text{ox}}$ ) and forms a mixed disulfide with the redox active disulfide bridge. The oxidized lipoyl-E2/E3BP residue moves away from E3, leaving two electrons reduced enzyme ( $\text{EH}_2$ ) whose disulfide is reduced but not the FAD. In an oxidative half reaction,  $\text{NAD}^+$  comes to its binding pocket and is thought to induce one of the thiol of disulfide to form a covalent link with C4 carbon of heterocyclic isoalloxazine ring of FAD (covalent C4a adduct), thereby reducing its N5 atom. The second thiol then attacks the first one regenerating disulfide, that breaks the covalent bond with FAD and in a process reduce N1 atom. The  $\text{FADH}_2$  thus formed, is rapidly oxidized by  $\text{NAD}^+$ . The final output is an oxidized enzyme ( $\text{E}_{\text{ox}}$ ), disulfide and NADH. The oxidized enzyme is then ready for the next reaction cycle.

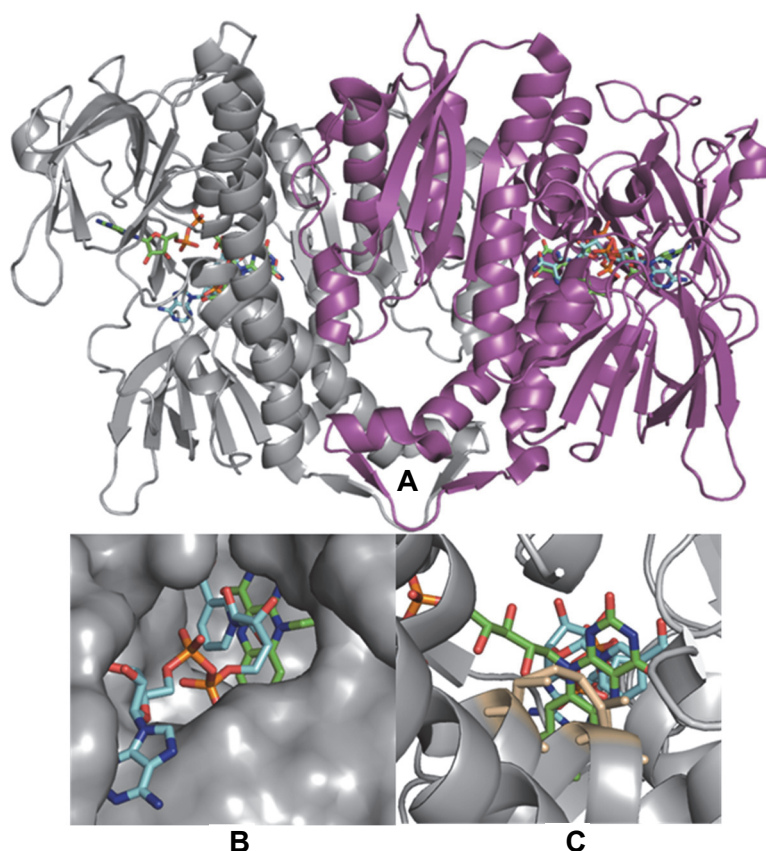


**Fig. 4: Physiological reaction of E3.**

At the start, FAD cofactor (black, with N1 and N5 in pink and C4 in red) and intrinsic disulfide bridge (blue) are in the oxidized state. A reduced lipoyl group (purple) visits its active site after acetyl transfer to CoA in the E2 active site. The reaction proceeds with a thiolate being generated at the lipoyl group, which attacks an intrinsic disulfide bridge, producing a mixed disulfide intermediate. Next, the lipoyl moiety is reduced and leaves E3 while thiolate is generated at one of the cysteines of the intrinsic disulfide. Once the NAD<sup>+</sup> substrate enters its binding pocket, the thiolate makes a covalent adduct at C4 with reduction of N5. Then, the intrinsic disulfide bridge is formed again followed by reduction of N1. The fully reduced FAD then reduces the NAD<sup>+</sup> and the reaction cycle is complete with the clearance of NADH. As the enzyme now is in oxidized state, next reaction cycle can begin with another incoming reduced lipoyl moiety.

In 2005, structures of *hE3* in the presence of FAD and NAD<sup>+</sup>/NADH were solved (Brautigam, Chuang et al. 2005) (see Fig. 5). It showed the nicotinamide base of NADH and not NAD<sup>+</sup> facing the isoalloxazine ring of FAD, which was interesting because the hydride transfer from N5 of FADH<sub>2</sub> to C4 atom of nicotinamide base of NAD<sup>+</sup> would what produces NADH in the biological reaction. They explained that NAD<sup>+</sup> would only adopt the conformation suiting to face FAD isoalloxazine ring once the negatively charged charge-transfer complex between C85 and FAD is formed (see Fig. 4). Furthermore, as crystallization was performed with the oxidized protein

(FAD and C80-C85), this possibly inclined the hydrophobic and neutral NAD binding pocket to better support neutral NADH.



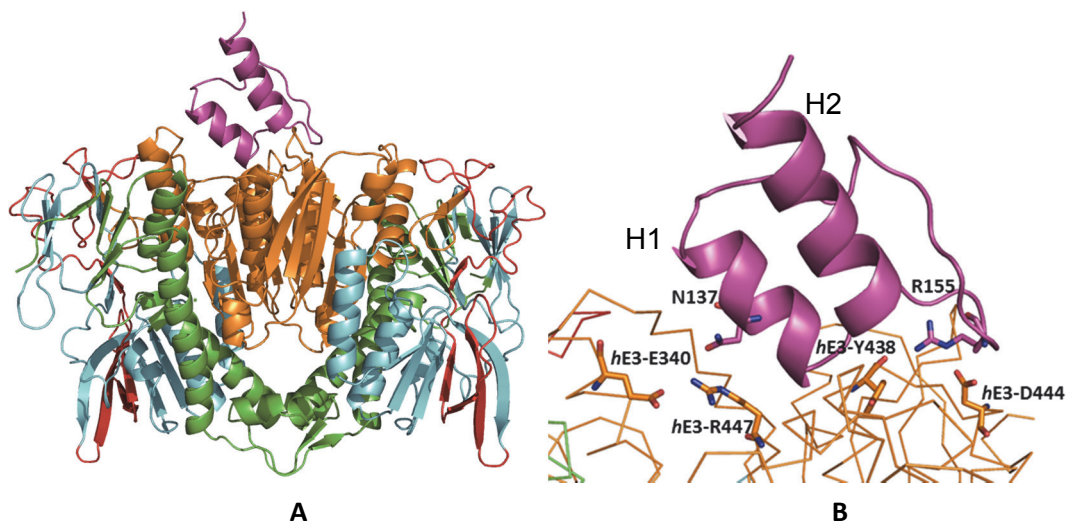
**Fig. 5: Crystal structure of *hE3* with bound FAD and NADH (PDB code 1ZMD).**

(A) The overall structure of *hE3* has a sandwich topology with the active sites lying almost at the opposite ends. The two monomers are shown in cartoons of grey and magenta colors with carbon atoms of FAD in green and NADH in light blue respectively. (B) A large binding pocket could be seen from outside for NAD<sup>+</sup> substrate. The nicotinamide ring of NADH is facing the isoalloxazine ring of FAD. (C) Opposite to the binding pocket of NAD<sup>+</sup>, a small space is available for dihydrolipoyl-E2 to enter, which would encounter C80-C85 disulfide (wheat) for electron transfer.

### 1.3.2.b. Binding of *hE3* to its cognate binding site at E3BP

A crystal structure of *hE3* bound to its binding site E3BD, a small domain belonging to E3BP, is available (Ciszak, Makal et al. 2006). E3BD makes contact with the *hE3* at the interface of its two subunits such that another E3BD cannot bind to *hE3* due to steric hindrance. Much like *bsE2*-SBD (K137 and R157) interacting with the *bsE1* (C-terminal residues of  $\beta$  chains) (see Fig. 3), the equivalent residues in E3BD (N137 and R155) have interacting partners at *hE3*, unlike in the case of *hE1* where equivalent residues of E1BD seemed to be partnerless (Patel, Korotchkina et al. 2009). Furthermore, I157 of E3BD makes hydrophobic interaction with Y439 of *hE3*

and combining with other hydrophobic interactions, give monospecificity of binding to E3BD. The *hE2*-E1BD lacks similar residues at equivalent positions, possibly disallowing it to bind *hE3* (Ciszak, Makal et al. 2006, Patel, Korotchkina et al. 2009). The *hE3*-E3BD is different from the *bsE3*-SBD complex, as both helices (H1 and H2) of E3BD are involved in binding to *hE3* with residues like K160 and E161 of H2. In *bsE3*-SBD however, H2 does not form any interactions with E3 protein (Mande, Sarfaty et al. 1996, Brautigam, Wynn et al. 2006).



**Fig. 6: Crystal structure of *hE3*-E3BD complex (PDB code: 1zy8).**

A) An E3BD (pink) binds to an interface between two *hE3* monomers, which is formed by C-terminal interface domain of each monomer (brown). NAD, FAD and central domains of *hE3* are shown in cyan, green and red, respectively. (B) Some of the important interactions between residues of E3BD and *hE3*. The residues N137 and R155 of E3BD make several interactions with *hE3* residues. The equivalent residues of E1BD are assumed not to play any role in binding to *hE1* while they are important for *bsE1* binding to *bsE2*-SBD.

If two E3BD attempt to bind to *hE3* at the same time, the rigid loop between H1 and H2 of E3BD is supposed to clash (Ciszak, Makal et al. 2006). However, they can bind to *hE3* with two possible orientations at the interface between *hE3* subunits. This supports one E3BP to one *hE3* binding model from (Brautigam, Wynn et al. 2009) whereas other groups have reported 2:1 stoichiometry (Smolle, Prior et al. 2006, Vijayakrishnan, Callow et al. 2011). Either way, the binding was shown to be very tight, with  $K_d$  of 0.78 nM calculated via isothermal titration calorimetry (Brautigam, Wynn et al. 2006) and 5.2 nM via surface plasmon resonance (Patel, Korotchkina et al. 2009).

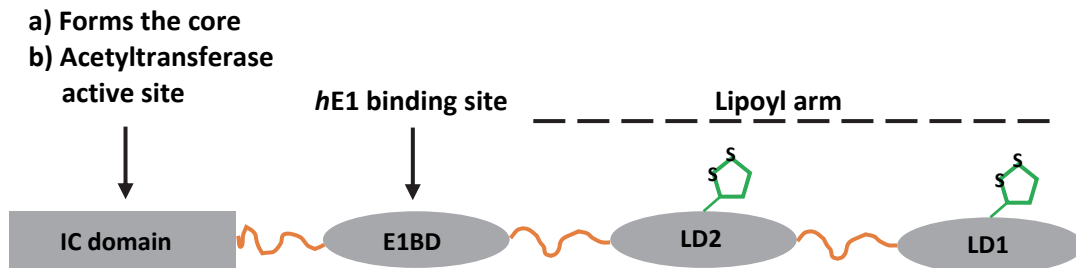


## **1.4. Core subunits of human pyruvate dehydrogenase enzyme complex**

### **1.4.1. Human dihydrolipoamide acetyltransferase (*hE2*)**

*hE2* (EC 2.3.1.12) is an acetyl transferase which is encoded by *DLAT* gene located in the chromosome 11 in human. In all life forms, they cluster together to give rise to 24meric cuboids or 60meric pentagonal dodecahedra (Reed 1966). Structurally it consists of two N-terminal lipoyl domains (LD) followed by one E1 binding domain (E1BD) and core forming C-terminal inner catalytic domain (IC), all interconnected via highly flexible linkers (see Fig. 7) (Perham 1991). The number of IC domain and E1BD is one in all characterized PDHc from various organisms. But the number of lipoyl domains can vary between species within same biological kingdom (Perham 1991, Yu, Du et al. 2012) (see table 2).

Several groups have contributed in structural and functional characterization of domains and linkers of *hE2* spanning several decades. The structure of these domains was solved mostly in isolation and not together as a functional unit. In the following sub-chapters, all the domains of *hE2* are individually discussed with suitable comparison to their orthologs, especially from bacterial PDHc, which are characterized well.



**Fig. 7: Cartoon representation of modular architecture of human dihydrolipoamide acetyltransferase (*hE2*).**

The IC domains of several E2s together with similar IC domains of E3BPs form a central core of *hPDHc*. E1 protein binds to the E1BD, which is tethered to the exterior of the core by a fairly long linker. The lipoyl domains (LD1 and LD2) each carry a lipoyl group (green). A highly flexible Ala-Pro rich linker similar to that between E1BD and LD1 connects these lipoyl domains. (all linkers are in orange)

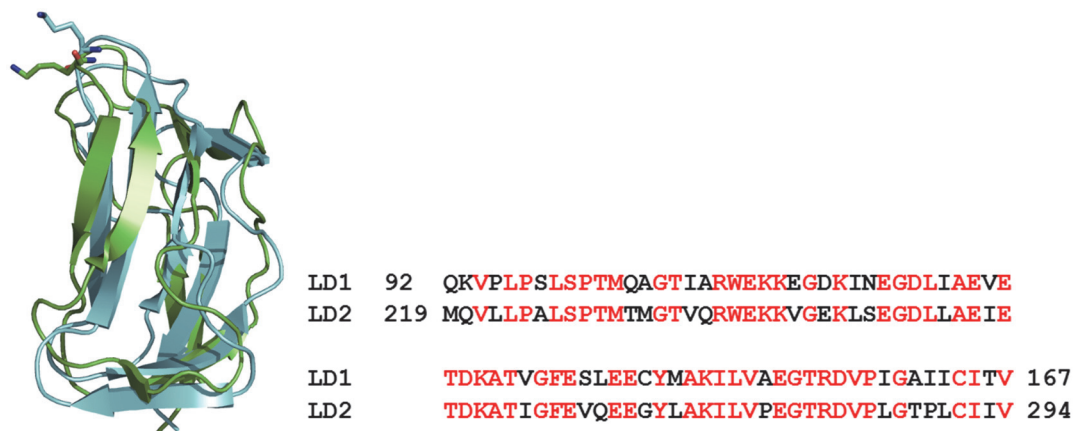
E2 source	Number of lipoyl domains
<b>Mammals</b>	2
<b>Yeasts</b>	1
<b>*Plants</b>	
<i>Z. mays</i>	1
<i>S. tuberosum</i> and <i>A. thaliana</i>	Two isoforms of E2 with 1 and 2 lipoyl domains
<b>Bacteria</b>	
<i>E. coli</i>	3
<i>E. faecalis</i>	2
<i>B. stearothermophilus</i>	1

**Table 2: Comparison of number of lipoyl domains in well-defined E2 subunits between different organisms.**

Although the number of lipoyl domains varies greatly between organisms, it appears that the trend does not follow hierarchy in taxonomy. The number of lipoyl domains does not correlate with the activity of E2. Furthermore, extra lipoyl domains could be deleted (genetically or proteolytically) without significantly reducing acetyl transferase activity of E2, suggesting functional redundancy. \*Plant E2 mentioned in the table is mitochondrial in origin. Plastids also have their own PDHc, but are not well described.

### 1.4.1.a. Lipoyl domains

*hE2* has two lipoyl domains with similar primary sequences and structures (PDB 2DNE and 1 FYC) (Howard, Fuller et al. 1998). The lipoyl domain situated at the N-terminus part of *hE2* is termed as lipoyl domain 1 (LD1) and the inner lipoyl domain situated between LD1 and E1BD is called lipoyl domain 2 (LD2). They are globular domains of around 75 amino acid residues with two sets of anti-parallel beta sheets each containing four  $\beta$  strands (PDB 2DNE and 1 FYC). They are roughly 35 Å long and 18 Å broad. The conserved lysine residue, which is covalently linked to lipoic acid via amide linkage, lies at the loop region between  $\beta$  strand 4 and 5. These domains share structural similarity to lipoyl domains of E2 component of bacterial PDHc, even though the primary sequence differs by more than 50 % (Dardel, Davis et al. 1993, Green, Laue et al. 1995). Furthermore, the lipoyl domains of other related multi enzyme complexes such as alpha-ketoglutarate dehydrogenase complex and branched-chain alpha-ketoacid dehydrogenase complex also appear to share similar protein fold (Berg, Vervoort et al. 1996, Ricaud, Howard et al. 1996, Huang, Chang et al. 2002).



**Fig. 8: Comparing structures of LD1 and LD2 of *hE2*.**

Both the primary sequences and structure of LD1 (green, PDB 2DNE) and LD2 (cyan, PDB 1FYC) look similar. When their structures are superimposed, the lipoylation sites (K114 for LD1 and K241 for LD2) occupy similar space (represented as sticks at the top). The lipoamide group lies at the tip of lipoyl domain making it lot easier to enter the narrow active sites of *hE1*, *hE2* and *hE3*.

Lipoic acid is a derivative of octanoic acid, in which sulfur atoms are added to C-6 and C-8, with a possibility of ring closure giving rise to 1-2 dithiolane ring. The reduced form of lipoate is dihydrolipoate, which is biologically important as it can carry carbon

chain. Deprotonation of any of the sulfur atom of dihydrolipoate forms highly nucleophilic thiolate ion. This can attack carbonyl carbons to form thioester bond. The pKa of thiols in dihydrolipoate is around 10.7 (Radda 1966) which suggests that at physiological pH, the generation of thiolate ion would be rare. But since lipoate group is covalently linked to protein, a suitable environment with strategically placed basic amino acid sidechain could accelerate the deprotonation of thiol to generate thiolate.

In humans, lipoylation of proteins occur in two steps. First, an octanoyl transferase transfers the octanoyl group from octanoyl-acyl carrier protein to  $\epsilon$ -amino group of conserved lysine residue of lipoyl domain. In this process, an amide linkage is formed between lysine residue and octanoate. In the second step, lipoyl synthase enzyme replaces the hydrogen from C-6 and C-8 position of octanoyl group with sulfur atoms using two S-adenosyl methionine (SAM) molecules (Jordan and Cronan 1997, Zhao, Miller et al. 2003). In bacteria, there is also another pathway for lipoylation where free lipoic acid is added directly to lipoyl domain by lipoate protein ligase (Morris, Reed et al. 1994).

One of the main advantages for an enzyme using lipoyl group as an acyl carrier in a multi-enzyme complex, beside its long reach of 14 Å, is its rotational flexibility. The lipoyl moiety was shown via electron paramagnetic resonance (EPR) spectroscopy to have a small rotational correlation time of 0.2 ns in *E. coli* PDHc (Ambrose and Perham 1976) and 0.18 ns in *A. vinelandii* PDHc (Ambrose and Perham 1976). It was proposed that it can rotate freely and rapidly in all direction from its anchoring point at peptide chain. Adding to that the flexible inter-domain linkers, we get extensions emanating from the core that are virtually free to move in all directions.

Free lipoate and lipoamide easily can serve as a substrate for E2 and E3, which means the recognition of lipoyl group is enough for catalysis in those enzymes. But in the case of E1, these molecules are poor substrates with  $k_{cat}/K_m$  reduced by 20000 fold in comparison to lipoylated lipoyl domain as a substrate (Graham, Packman et al. 1989). This suggests lipoyl group is not the only entity, which is recognized at bacterial E1, but interactions between residues, possibly at the contact region between E1 and lipoyl domain, are also important.

However, one thing that is yet to be explained is the mechanism of acetyl and electron pair transfer between lipoyl domains. In a study where 90% of E1 active sites in *E. coli*. PDHc were inactivated using thiamine thiazolone pyrophosphate (an analog of ThDP), all the lipoyl domains were still found to be acetylated (Collins and Reed 1977). Even after adding excess of normal E1s, the overall activity of PDHc did not

increase suggesting the E1s are tightly bound and cannot possibly hop around to acetylate distant lipoyl domains. Similarly, in a theoretical model explaining experimental observations of change in activity vs E1:E2 molar ratio, it was found that one E1 protein must be serving at least 12 E2 chains in octahedral *E. coli* PDHc (Bates, Danson et al. 1977). These studies confirmed that there must be an extensive inter lipoyl acetyl transfer network, as it's physically impossible for single E1 to serve so many E2 chains at once. This networking is thought to aid in overall activity even when the PDHc core is not saturated with E1 and E3 proteins or when the pyruvate level is depleted. But the mechanism for such transfer is not explained in any detail. Also no data is available regarding the favorable orientation of such lipoyl domains coming into contact.

This also brings us to another intriguing fact about lipoyl domain, which is its copy number in E2 from various organisms (see Table 2). It is still an open question why nature chose variable number of lipoyl domains when just a single copy can do the job. Genetic experiments carried out in *E. coli* PDHc demonstrated that the three lipoyl domains of its E2 could be mutated or deleted in any order, keeping at least one normal domain without altering the viability of the cell culture (Guest, Lewis et al. 1985, Allen, Perham et al. 1989). Prior to these comprehensive experiments, protease and lipamidase were used to decrease active lipoyl content in *E. coli* PDHc, where it was shown that even when 50 % active lipoyl content is destroyed, it does not significantly reduce overall activity of PDHc and related  $\alpha$ -ketoglutarate dehydrogenase complex. Computational models were built to best explain these observations, which again pointed toward extensive networking between lipoyl domains to transfer acetyl group or electron pairs. This mechanism was termed as multiple random coupling (MRC) and is assumed to be one of the bases of substrate channeling in all PDHc (Stepp, Bleile et al. 1981, Stepp, Pettit et al. 1981, Hackert, Oliver et al. 1983, Hackert, Oliver et al. 1983). Similar observations have also been made on mammalian PDHc, when a collagenase was used to selectively remove lipoyl domains from PDHc-E2 (Rahmatullah, Radke et al. 1990). If at least one lipoyl domain was left for E2, it could support 100 % PDHc activity. But, if both lipoyl domains were removed, only < 20 % activity would remain, possibly due to the presence of lipoyl domain from E3BP. However, there are also E2s in nature with only one lipoyl domain and thus, MRC could also be a mere consequence of the presence of multiple lipoyl domains in a single E2 chain rather than an absolute need for PDHc function.

### 1.4.1.b. Linkers

There are two linkers associated with lipoyl domains in human PDHc. An outer linker connects LD1 and LD2 whereas inner linker connects LD2 with E1BD. A relatively short linker connects E1BD to the core (see Fig. 9). The importance of such linkers was highlighted in a study of an engineered *E. coli* PDHc E2 with only one lipoyl domain. Shortening of the linker region between lipoyl domain and SBD reduced the overall activity, which was suggestive of the critical role of its length in helping lipoyl domain to reach active sites of E1, E2 and E3 (Miles, Guest et al. 1988). What is characteristic regarding all these linkers are that they are extremely rich in proline and alanine residues (~ 40 % of total constituents). This gives them lots of flexibility, which is thought to be important for movement of lipoyl domains to all the catalytic sites. The direct proof for this came from the proton NMR experiment where a histidine residue was introduced via point mutation in the linker region between lipoyl domain and SBD of similar construct used in experiment mentioned above (Texter, Radford et al. 1988). The mutated version produced two sharp resonances in  $^1\text{H}$  NMR spectra at the aromatic region, which was absent for wild type PDHc. The chemical shifts of those sharp peaks (7.94 and 7.05 ppm) are characteristics of C-2 and C-4 protons of histidine residue. An apparent pKa of 6.4 was calculated for the side chain of this histidine. All in all, this made it clear that the newly added histidine is exposed to solvent and was a part of conformationally plastic region of E2.

Outer linker: 168 GKPEDIEAFKNYTLDSSAAPTPQAAPAPTPAATASPPTPSAQAP  
 GSSYPFH 218  
 Inner linker: 295 EKEADISAFADYRPTTEVTDLKPQVPPPTPPPVAAPPTPQPLAP  
 TPSAPCPATPAGPKGRVFS 358  
 Linker between  
 Core and E1BD: 394 PSKVAPAPAAVVPPTGPGMAPVPTGVFTDIP 425

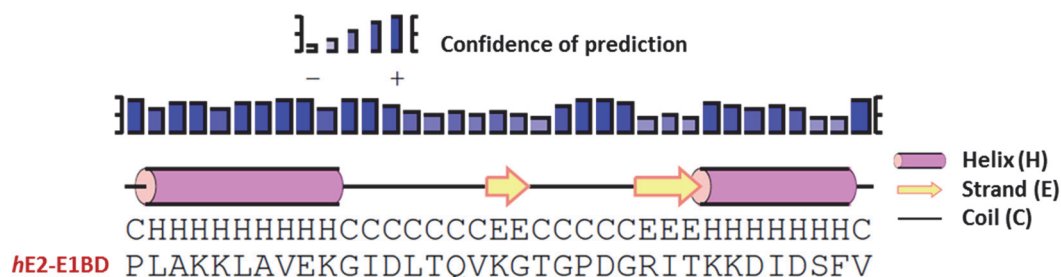
**Fig. 9: Comparing linker regions of *hE2*.**

All three linker regions constitute high number of alanine and proline residues. The peptide chains selected here are between globular domains (LD1, LD2 and IC domain from human PDHc and E1BD by comparing to SBD from *bsPDHc*).

**1.4.1.c. E1 binding domain (E1BD)**

It is a very small domain (35 amino acid residues) making contact with E1  $\beta$  chains (refer Section 1.3.1.b.). The structure of E1BD has not been solved for human PDHc and thus comparative study with bacterial PDHc E2-SBD and E3BD from human E3BP is useful. Secondary structure prediction shows that they all might share similar structure (see Fig. 10). E1BD and E3BD diverged from SBD of prokaryotic PDHc during the course of evolution and lost ability to bind to E1 and E3 in a mutually exclusive manner. But instead they developed monospecificity towards these peripheral proteins. Albeit inconclusive, the orientation of E1BD could dictate where the two active sites of E1 proteins face, towards core or away from it. This may influence feasibility for lipoyl domains to visit those active sites (Zhou, McCarthy et al. 2001). It is important to note that E1 proteins like many other ThDP dependent enzymes, show half of the sites reactivity meaning only one active site remain operational at a given time (Schroder-Tittmann, Meyer et al. 2013).

<i>hE2</i> -E1BD	359	PLAKKLAVEK <b>G</b> IDLTQ <b>V</b> KGTG <b>P</b> DGRIT <b>K</b> KD <b>I</b> DSF <b>V</b>	393
<i>hE2</i> -E3BD	186	PAARNILEKHS <b>L</b> DAS <b>Q</b> GTATG <b>P</b> RGIF <b>T</b> KEDALKLV <b>Q</b> LK <b>Q</b>	224
<i>bsE2</i> -SBD	132	PSV <b>R</b> KY <b>A</b> REK <b>G</b> V <b>D</b> IRLV <b>Q</b> GT <b>G</b> K <b>N</b> GRVL <b>K</b> ED <b>I</b> DA <b>F</b> L	167



**Fig. 10: Sequence alignment between E1BD and E3BD of human PDHc with SBD of *B. stearotherophilus* PDHc.**

Although the sequences have < 20 % identity, there are ~ 60 % of similar residues (Red-identical, Blue-same in two sequence and Green-similar in all). The secondary structure prediction of E1BD by PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred>) yields two helices with high confidence, which is comparable to crystal structures of *hE2*-E3BD and *bsE2*-SBD. They also have two helices ( $\alpha$ 1 and  $\alpha$ 2) joined by a long linker.

#### 1.4.1.d. Inner catalytic (IC) domain

Also known as core forming domain, it is responsible for oligomerization of *hE2* proteins. In human PDHc, such aggregation provides a pentagonal dodecahedron framework onto which *hE1* and *hE3* proteins join in and altogether, form a huge assembly with molecular weight of ~ 10 MDa. This dodecahedral core is formed by twenty trimers made up of *hE2* and E3BP, which share similar modular architecture of E2. Such trimers are located at the vertices of the dodecahedron resulting in twelve pentagonal empty openings.

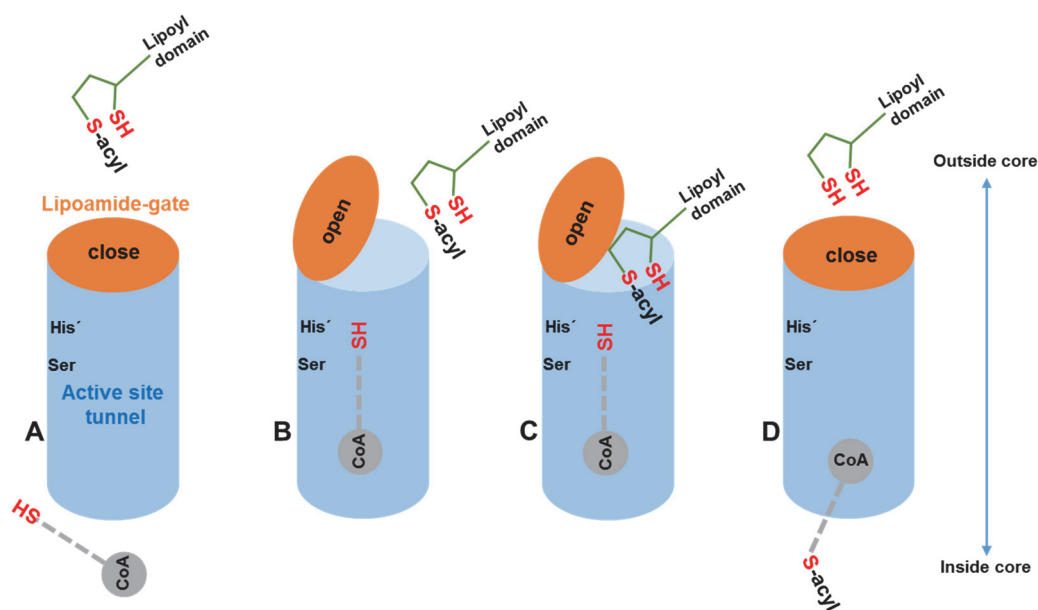
In all the available structural models of the PDHc core from various organisms, the internal interactions in the monomer are similar and the N-terminal arm makes several hydrophobic and hydrogen bond interactions with the body of another monomer within a trimer (Mattevi, Obmolova et al. 1992, Yu, Hiromasa et al. 2008, Wang, Nemeria et al. 2014). The C-terminal residues of monomers linking two trimers make contact with each other at their hydrophobic pockets, like a 'ball and socket joint' creating higher order structure from trimers. Furthermore, it was predicted that minor changes in the positioning of amino acid residues at the contact point should suffice in changing the symmetry of PDHc cores between cuboidal in gm-negative bacteria and dodecahedral in gm-positive bacteria and mammals (Izard, Aevarsson et al. 1999,



Yu, Hiromasa et al. 2008). Following Euclidean geometric principle, a dodecahedral core could be constructed from cubic one, by first increasing the distance between centers of the core to trimers, such that the edges of cube and dodecahedron become same and then rotating the trimer to make monomeric components face the edges of the dodecahedron. This concept was validated using examples from crystal structures of cubic core of PDHc of *A. vinelandii* and dodecahedral one from *B. stearrowthermophilus* (Izard, Aevvarsson et al. 1999).

Irrespective of the symmetry though, each monomeric unit possess one catalytic site where an acetyl group is transferred from S8-acetyldehydrolipoyl group to CoA. This catalytic site is a long tunnel (~ 30 Å), at the interface between two monomers of a trimer unit. The CoA substrate first enters inside region of this tunnel probably via pentagonal openings, whereas the S8-acetyldehydrolipoyl group arrives from outside the core (Mattevi, Obmolova et al. 1992, Wang, Nemeria et al. 2014).

While comparing crystal structures of binary complexes of *A. vinelandii* PDHc core with CoA (IN conformation) and dihydrolipoamide, the thiols from both molecules, were shown to be at close proximity to catalytic histidine residue (Mattevi, Obmolova et al. 1993). This residue is part of a conserved **DHRXXDG** motif present in all acetyl transferases and is thought to play a role of general acid-base catalyst in deprotonation of CoA thiol to form thiolate anion that attacks carbonyl in S8-acetyldehydrolipoyl group. However, in a same study, a crystal structure of a tertiary complex between *A. vinelandii* PDHc core, CoA and dihydrolipoamide was also determined, in which thiol of CoA (OUT conformation) was pushed to ~ 12 Å away from active site residues. In a separate study on dihydrolipoamide acetyl transferase (E2) from bovine branched chain  $\alpha$ -ketoacid dehydrogenase complex (Kato, Wynn et al. 2006), it was shown that in the presence of CoA, the affinity of the core to the dihydrolipoamide or dihydrolipoamide-lipoyl domain is significantly increased from undetectable to 48  $\mu$ M and 11  $\mu$ M, respectively. Based on this, they proposed a gating mechanism in which, CoA binding triggers the opening of an entrance gate for acetyl-dihydrolipoamide (see Fig. 11). After acetyl transfer, once the dihydrolipoamide leaves, the gate shifts back to its closed conformation. Conforming to this model, a residue (D228) proposed to be crucial for stabilizing open state was mutated. Afterwards, dihydrolipoamide could not bind to the core even in the presence of CoA. Similarly, when a residue (L293) important for stabilizing closed state was mutated; dihydrolipoamide could bind both in the presence and absence of CoA.

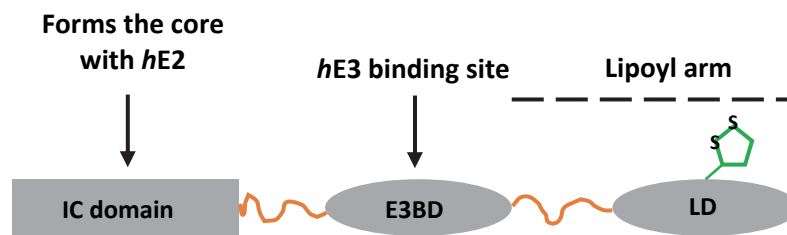


**Fig. 11: A general model of gating mechanism for entry of acylated-lipoamide in a dihydrolipoamide acyltransferases (E2) of  $\alpha$ -keto acid dehydrogenase complexes.**

In the absence of substrates, the gate remains closed (A). When CoA enters a long active site tunnel from inside of the core, the gate at the other end opens (B). The acylated lipoyl moiety can then enter the active site from outside the core (C). Upon acyl transfer, the dihydrolipoyl-group leaves the active site, and the gate closes again (D). The conserved histidine residue (His') is used as a general acid-base catalyst, which is contributed from another monomer of a same trimeric unit. Another well-conserved residue is a serine (Ser), which stabilizes tetrahedral intermediate formed after attack of thiolate of CoA on carbonyl of acyl-lipoyl group.

#### 1.4.2. Human dihydrolipoamide dehydrogenase binding protein (E3BP)

E3BP, formerly known as protein X, is only present in eukaryotic PDHc (Demarcucci and Lindsay 1985, Jilka, Rahmatullah et al. 1986). It is involved in forming the core with E2 and binding to E3 protein (Gopalakrishnan, Rahmatullah et al. 1989, Rahmatullah, Gopalakrishnan et al. 1989, Roche, Rahmatullah et al. 1989). It has similar modular architecture to *hE2*, with only one lipoyl domain (Neagle, Demarcucci et al. 1989) and E3 binding domain (E3BD) instead of E1BD. Although the structure of lipoyl domain of E3BP has not been solved, it should be similar to LD1 and LD2 of *hE2*, since they share high sequence similarity of  $\sim 45\%$  (Harris, BowkerKinley et al. 1997). Flexible linkers much like in the case of *hE2* connect domains in E3BP. The structure of E3BD-*hE3* is known and is already discussed in section 1.3.2.a.



Linker between

LD and E3BD: 133 EGEDWKHVEI**PKDVGPPPPVSKPSEPRPSPEPQISIP**  
VKKEHI**PGTLRFRLS** 185

Linker between

Core and E3BD: 223 **KQTGKITESRPTPAPTATPTAPSP**LQATAGPSYPRPV  
**IPPVSTPGQPN**AVGTFTEIPA 280

**Fig. 12: Cartoon representation of modular architecture of human E3BP and sequences of associated inter domain linkers.**

Similar to *hE2*, E3BP possess core forming IC domain but is catalytically inactive. E3BD lies between IC domain and N-terminal lipoyl domain (LD). The inter domain linkers (orange) are rich in proline and alanine with an exception being the linker between LD and E3BD which does not have any alanine residues. Nevertheless, the high proline content means that it should still be flexible.

The C-terminal part is a core-forming domain, which shares 50 % sequence identity with IC domain of *hE2* (Harris, BowkerKinley et al. 1997). While *hE2* can be recombinantly overexpressed and purified in the absence of E3BP, yielding 60meric dodecahedral core, successful overexpression of E3BP in the absence of *hE2* has not been reported thus far. An artificial *hE2* core however cannot form fully functional PDH complex, as it cannot bind to E3 proteins (Yang, Song et al. 1997).

E3BP is not expected to have acetyltransferase activity as the conserved histidine in DGRXXDG motif present in all acetyl transferase, is mutated to serine residue (Harris, BowkerKinley et al. 1997). But it can still take part in the PDHc overall reaction since its lipoyl domain is shown to be lipoylated and also reductively acetylated in the presence of pyruvate (Demarcucci, Hodgson et al. 1986). Also, selective removal of lipoyl domain of E3BP did not abolish the overall PDH activity (~ 80 % remaining) in a bovine PDHc, which meant its function, can be easily overtaken by lipoyl domains of E2 (Neagle and Lindsay 1991). However, as mentioned in section 1.4.1.a, the role reversal for lipoyl domains of E3BP and E2 might not be fine for both ways, as when only lipoyl domains of E3BP were left functional, it could only sustain 10 % - 20 %

overall PDHc activity in mammalian PDHc. Thus, MRC mechanism (refer Section 1.4.1.a.) explained for E2 might also be extended to the E3BP lipoyl domains.

### **1.5. Organization of the human PDH complex**

Until the discovery of E3BP protein, it was imagined that eukaryotic PDHc core like in prokaryotic version, is also organized solely by E2 proteins. After electron microscopic investigation of eukaryotic PDHc (Hayakawa, Kanzaki et al. 1969, Reed 1974), it was established that the core of eukaryotic PDH complex has similar dodecahedral symmetry to that from gram-positive bacteria (Henderson, Perham et al. 1979). The copy number of E2 was assumed to be 60, with 20 trimers occupying 20 vertices of pentagonal dodecahedron.

Once E3BP (Protein X) was proven to be an quintessential part of eukaryotic PDH complex, due to its affinity to aggregate with E2 component and also capacity to bind to E3 proteins, a discussion started for the localization of E3BP in PDH complex. The first model to emerge was 60+12 'addition model' where 12 E3BP monomers were put at the 12 pentagonal openings of the 60meric E2 core (Gopalakrishnan, Rahmatullah et al. 1989, Maeng, Yazdi et al. 1994, Stoops, Cheng et al. 1997). Later, it was found mainly via sedimentation equilibrium studies that the E2/E3BP core was smaller in molecular weight than artificial E2 core lacking E3BP (Hiromasa, Fujisawa et al. 2004, Vijayakrishnan, Kelly et al. 2010). Also, the electron microscopic structure of human PDHc core lacked any density at the pentagonal openings supposed to be the location of E3BP (Vijayakrishnan, Kelly et al. 2010). This meant the native E2/E3BP core does not follow 60+12 model. Rather it appeared that E3BP proteins would substitute equivalent number of the E2 proteins in the core ('substitution model').

Problem then arose in assigning the copy number of E3BP and E2 in the core, as several groups published conflicting results. Some studies show 12 E3BP proteins to be present in eukaryotic PDHc core along with 48 E2 proteins (48+12 substitution model) (Hiromasa, Fujisawa et al. 2004) while others have reported stoichiometry of 40 E2 and 20 E3BP proteins (40+20 substitution model) (Brautigam, Wynn et al. 2009). One study even demonstrated the discrepancy in the stoichiometry of E2 to E3BP between native mammalian PDHc core (48+12 model) and recombinantly expressed core (40+20) (Vijayakrishnan, Callow et al. 2011).

The first reason why stoichiometry of the core components is important is that based on stoichiometry, location of E3BP could be better resolved. For example, if 40+20 model is true, then, it can be imagined that all trimer elements are same, consisting 2 E2 and 1 E3BP each. Second important reason why this is important is that the copy number of *hE2* and E3BP also dictate number of *hE1* and *hE3* that can bind to the core.

While there is a consensus in one *hE2* binding to one *hE1* via its E1BD (Brautigam, Wynn et al. 2009), E3BP binding to *hE3* is not straightforward. Although in the crystal structure of E3BD-*hE3* sub-complex, one E3BD is seen to bind with only one *hE3* protein (Ciszak, Makal et al. 2006), possibility of 2 E3BPs binding to one *hE3* in solution has been proposed by other groups using small angle X-ray scattering (SAXS), isothermal titration calorimetry (ITC) and analytical ultracentrifugation (AUC) techniques (Smolle, Prior et al. 2006, Vijayakrishnan, Callow et al. 2011). However, using similar techniques, another group got different result, one that supports the crystal structure (Brautigam, Wynn et al. 2006, Brautigam, Wynn et al. 2009).

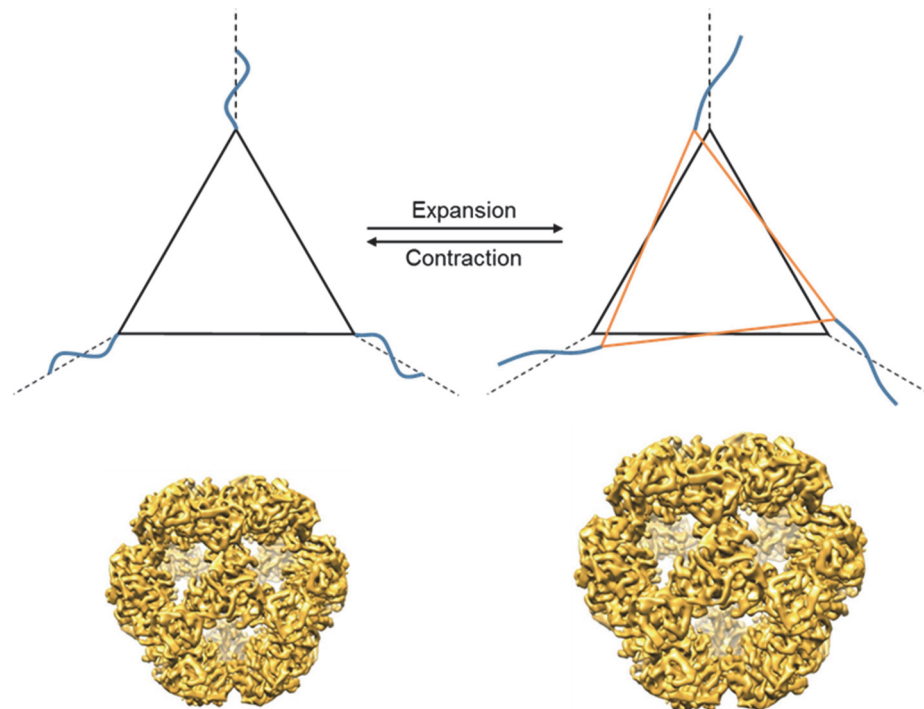
From the electron microscopic studies, the E1 and E3 proteins have been localized at the periphery of the core in bacterial PDHc (Milne, Shi et al. 2002, Milne, Wu et al. 2006) anchored by the linkers. However, for mammalian PDHc, only E1 have been observed at the periphery (Wagenknecht, Grassucci et al. 1991, Zhou, McCarthy et al. 2001) and E3's position is in debate. Both, models for E3 proteins localized near the pentagonal openings of the core (Stoops, Cheng et al. 1997, Zhou, McCarthy et al. 2001) or distant from the core like E1 (Hiromasa, Fujisawa et al. 2004), are available in the literature.

## 1.6. Conformational plasticity of PDHc core

The functioning of a protein also depends in its dynamics. Very fast thermal fluctuation at pico-second scale, as well as domain movements and associated global conformational changes occurring at slower millisecond regime, all contribute to the flexibility of protein (Hammes 1964, Mccammon, Gelin et al. 1977, Hay and Scrutton 2012). Just by looking at the reaction scheme of PDHc (see Fig. 1), it is self-explanatory that the lipoyl arm must have certain reach and be extraordinarily flexible for lipoyl domains to visit three different active sites (details in section 1.4.1.b.).

Another important dynamic feature of PDHc complex is the size variation of its core, which was first observed during electron microscopic study of yeast PDHc (Zhou, Liao

et al. 2001). It was found that the size of yeast PDHc core could vary by  $\sim 4$  nm between smallest and largest cores in a same population, which corresponds to 1.4 nm changes in length of inter trimer bridges. It was proposed that such size variation is possible due to relaxation and stretching of inter trimer linkage (see Fig. 13). This phenomenon, which is termed as 'breathing', has also been observed in human PDHc cores (Yu, Hiromasa et al. 2008). Based on the size distribution and treating the cores as harmonic oscillator, the effective spring constant of such cores has been computed to be 0.015 N/m (Kong, Ming et al. 2003) which makes them roughly 10 times softer than other regular globular proteins (Zaccai 2000). This inherent softness allows the spontaneous expansion and contraction of the inter trimer linkage with small thermal energy. Although, speculations have been made on possible role of such dynamics in substrate channeling and product clearance, no connection have been experimentally established.



**Fig. 13: Breathing motion of PDHc core.**

Adapted from (Kong, Ming et al. 2003). A trimer of PDHc core component (black triangle) with S-shaped inter trimer bridge (blue) along the edges of a dodecahedron (dashed line) rotates spontaneously (orange triangle) such that the bridges are radially stretched following the trimer. It causes the global expansion of the core, which is as much as 20 % of the original size. In a reversible manner, it relaxes and the cycle of expansion and contraction continues which is assumed to be thermally driven process.

### 1.7. Regulation of the activity of human PDH complex and health implications

As PDH complex lies at the junction between glycolysis and Krebs's cycle, almost 50 % of the food source is broken down and passed through PDH complex in a respiring cell. Hence, it is imperative to check and balance its activity, as the demand for energy can dramatically change between dormant and active states of a cell and organism or their healthy and diseased states. Both prokaryotic and eukaryotic PDH complexes are prone to inactivation by its products NADH and acetyl CoA. These ensure that there is at least a metabolic check to decrease PDH activity when energy requirement is less.

Another way to regulate the PDHc activity is reversible phosphorylation of pyruvate dehydrogenase (E1) component. The phosphorylation is done by pyruvate dehydrogenase kinase (PDK), which inhibits E1 while de-phosphorylation is performed by pyruvate dehydrogenase phosphatase (PDP), which relieves the

inhibition (Harris, Bowker-Kinley et al. 2002, Patel and Korotchkina 2006). There are several organisms (both prokaryote and eukaryote), which possess PDK/PDP, and also putative phosphorylation sites at E1, yet, there are no evidences of E1 phosphorylation. For example, among fungi, *S. cerevisiae* lacks this phosphorylation mechanism but is present in *N. crassa* (Patel and Korotchkina 2006).

In humans, there are three phosphorylation sites at E1 alpha subunits, which are Ser-264 (site 1), Ser-271 (site 2) and Ser-203 (site 3). There are four known isoforms of PDK (1, 2, 3 and 4) and two isoforms of PDP (1 and 2). Site 1 and site 2 can be phosphorylated by all the PDKs but site 3 is exclusively phosphorylated by PDK1 (Korotchkina and Patel 2001). Phosphorylation of site 1 (most important one) or its mutation to glutamate and aspartate, severely impair pyruvate binding and reductive acetylation of lipoyl-E2 (Seifert, Ciszak et al. 2007, Kato, Wynn et al. 2008). The site 1 is located near the substrate channel. As such, addition of negatively charged bulky group like phosphate would hinder the entry of pyruvate and lipoyl-*hE2*. The phosphorylation of other sites is supposed to affect ThDP binding to *hE1* (Korotchkina and Patel 2001). PDP 1 and 2, when necessary relieve inactivation by dephosphorylating the phosphorylated sites.

Adding to phosphorylation/dephosphorylation, which are known for decades, recent studies have revealed much more complexity in *hPDHc* regulation. Acetyl coenzyme-A acetyl transferase 1 (ACAT1) and Sirtuin 3 (SIRT3) have been found to be involved in lysine acetylation/deacetylation of PDP and *hE1*, affecting the ability of PDP and PDK to stay bound to *hPDHc* (Fan, Shan et al. 2014). Sirtuin 4 (SIRT4) has been demonstrated to delipoylate the *hE2*-lipoyl domains of *hPDHc*, which is supposed to play role in glutamine metabolism (Mathias, Greco et al. 2014).

Recent novel findings in regulatory pathways of *hPDHc* bear several health implications, especially in cancer biology. All of the negative regulations mentioned above are enhanced in tumor cells to stop carbon flow to cellular respiration. This, in turn, decrease production of reactive oxygen species, a side product of electron transport chain and an activator of apoptotic pathways (Chernyak 1997). Activation of PDP or deactivation of PDK was shown to maintain oncogene induced cell senescence, and also abolish established tumors (Kaplun, Zheng et al. 2013). Hence, engineering drugs that will target pathways of PDK/PDP inactivation/activation might provide a simpler treatment to cancer patients in the future. So far, few compounds like dichloroacetate (potent inhibitor of PDK) have been explored in pre-clinical in-vitro trials to be effective against various forms of tumor, but a clinical trial have not been



carried out yet, due to its neurotoxicity and other severe side effects (Michelakis, Webster et al. 2008, Michelakis, Sutendra et al. 2010).

## 1.8. Motivation

The first priority of this thesis work was to obtain high-resolution structure of the *h*PDHc core via cryo-EM. The only such model published so far for *h*PDHc core was of truncated E2 core with cryo-EM map at mere 9 Å resolution. The core was shown to be dynamic as it undergoes rhythmic expansion and contraction and the motion termed as breathing. Although it was predicted that this motion has role in PDHc catalysis, especially in substrate channeling, no proof have ever been documented.

One of its substrates, 'coenzyme A', was used to see if its binding induces any change in the conformations of E2 and E3BP. The idea stemmed from our multiple observations during routine laboratory work that CoA stabilize the integrity of the *h*PDHc core and also *E. coli* PDHc. Various biophysical techniques such as atomic force microscopy (AFM), small angle X-ray scattering (SAXS) and direct stochastic optical reconstruction microscopy (dSTORM) were employed to see if size distribution of the core changes upon CoA binding.

Cross-linking mass spectrometry was used to directly track movement of lipoyl domains in the functional *h*PDHc in CoA bound and unbound state. This would give snap shots of where and how lipoyl arm moves. Is the lipoyl arm movement stochastic as predicted by multiple random coupling mechanism or do key events in catalysis somehow dictate which active site lipoyl domain should preferentially visit? An answer for this might stimulate further discussion on its mode of function.

Also, the stoichiometry between *h*E3 and E3BP in the sub-complex as well as between E2 and E3BP in a core needed further elucidation. An ITC method was used to discriminate between 40:20 model versus 48:12 model for the *h*E2 and E3BP stoichiometry in the core. Also, 1 E3BP binding to 1 *h*E3 dimer versus a monomer models were evaluated. Neg-staining EM and native-MS were used to complement ITC in understanding the arrangement of core subunits and peripheral proteins.

An ultimate goal for me during this thesis work was to model human PDH complex architecture and its inherent dynamics using multiple biophysical approaches and to probe their respective influence in its function.

## 2. Materials and methods

### 2.1. Materials

#### 2.1.1. Chemicals

<b>Product</b>	<b>Supplier</b>
Acetic acid	Carl Roth, Karlsruhe
Acrylamide (30 %)	Carl Roth, Karlsruhe
Agar	AppliChem, Darmstadt
Agarose	AppliChem, Darmstadt
Ammonium hydrogen carbonate	Carl Roth, Karlsruhe
Ammonium chloride	AppliChem, Darmstadt
Ammonium sulfate	AppliChem, Darmstadt
Antifoam 204	Sigma-Aldrich, Munich
BisTris	AppliChem, Darmstadt
Bromophenol blue, sodium salt	AppliChem, Darmstadt
Carbenicillin, disodium salt	AppliChem, Darmstadt
Chloramphenicol	AppliChem, Darmstadt
Coenzyme A trilithium salt dihydrate	AppliChem, Darmstadt
Coomassie-Brilliant-Blue R-250	AppliChem, Darmstadt
Dimethylsulfoxide	Sigma-Aldrich, Munich
Dithioerythritol (DTE)	AppliChem, Darmstadt
Dithiothreitol (DTT)	AppliChem, Darmstadt
2,6-Dichlorophenolindophenol (DCPIP)	Merck KgaA
Ethanol	Nordhauser Spirituosen GmbH
Ethidium bromide	Carl Roth, Karlsruhe
Ethylenediaminetetraacetic acid (EDTA)	AppliChem, Darmstadt
Ethanol	Sigma-Aldrich, Munich
Flavin adenine dinucleotide (FAD) disodium salt 2-hydrate	Sigma-Aldrich, Munich
Flavin adenine mononucleotide (FMN) monosodium salt 2-hydrate	Sigma-Aldrich, Munich
D-Glucose	Carl Roth, Karlsruhe
Glutaraldehyde	Sigma-Aldrich, Munich
Glycerol (anhydrous)	AppliChem, Darmstadt

Glycerol (87 %)	AppliChem, Darmstadt
Glycine	AppliChem, Darmstadt
4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid	AppliChem, Darmstadt
Hydrochloric acid (37 %)	Th.Geyer GmbH and CoKG
Imidazole	AppliChem, Darmstadt
IPTG	AppliChem, Darmstadt
Isopropanol	AppliChem, Darmstadt
Kanamycin sulfate	Carl Roth, Karlsruhe
D-Lactose 1-hydrate	AppliChem, Darmstadt
(+/-)- $\alpha$ -Lipoic acid	Sigma-Aldrich, Munich
Magnesium acetate tetrahydrate	AppliChem, Darmstadt
Magnesium chloride, hexahydrate	Carl Roth, Karlsruhe
Magnesium sulfate, hydrate	Carl Roth, Karlsruhe
$\beta$ -Mercaptoethanol	Carl Roth, Karlsruhe
2-(N-morpholino)ethanesulfonic acid (MES)	AppliChem, Darmstadt
Methanol	Sigma-Aldrich, Munich
Nicotineamide adenine dinucleotide (NAD <sup>+</sup> )	AppliChem, Darmstadt
Phenyl methane sulfonyl fluoride (PMSF)	AppliChem, Darmstadt
Phosphoric acid (85 %)	Carl Roth, Karlsruhe
Polyethylenglycol (PEG)-3350	Hampton Research Corp, CA
Polyethylenglycol (PEG)-6000	CALBIOCHEM, San Diego
Polyethylenglycol (PEG)-6000	Carl Roth, Karlsruhe
Potassium chloride	Carl Roth, Karlsruhe
Potassium dihydrogen phosphate	Carl Roth, Karlsruhe
Potassium hydroxide	AppliChem, Darmstadt
Sodium dodecyl sulfate (SDS)	AppliChem, Darmstadt
Sodium chloride	AppliChem, Darmstadt
Sodium hydroxide	AppliChem, Darmstadt
D-Sucrose	AppliChem, Darmstadt
SYPRO Orange protein stain	BioRad GmbH, Munich
N,N,N',N'- Tetramethylethylenediamine	Carl Roth, Karlsruhe
Thiamine hydrochloride	AppliChem, Darmstadt

Thiamine diphosphate	Sigma-Aldrich, Munich
Tris	AppliChem, Darmstadt
Trypton	Carl Roth, Karlsruhe
Yeast Extract	Carl Roth, Karlsruhe
Yeast Extract for fermentation	Ohly GmbH, Hamburg

### 2.1.2. Instruments

<b>Product</b>	<b>Supplier</b>
<b><i>E. coli</i> culture</b>	
Autoclave	Zirbus, Bad Grund
Biofermenter, Biostat C	Sartorius AG, Göttingen
Incubation shaker, Unitron	Infors AG, Bottmingen
Laminar flow Prettl-Telstar Bio-II-A	Telstar, Terrasa, Spain
<b>Cell Disruption</b>	
Microfluidizer, M-110S	Microfluidics (Newton, MA, USA)
Mortar Grinder RM 200	Retsch GmbH, Haan
<b>Centrifuge and its accessories</b>	
Avanti™ HP-30I	Beckmann Coulter GmbH, Krefeld
Rotor JA-10	Beckmann Coulter GmbH, Krefeld
Rotor JA-30.50 Ti	Beckmann Coulter GmbH, Krefeld
Avanti™ J-20 XPIJA-20	Beckmann Coulter GmbH, Krefeld
Rotor JLA-8.1000	Beckmann Coulter GmbH, Krefeld
Centrifuge bottle assemblies (polypropylene)	Beckmann Coulter GmbH, Krefeld
Eppendorf 5810 R	Eppendorf AG, Wasseling-Berzdorf
Rotor A-4-81	Eppendorf AG, Wasseling-Berzdorf
Micro 200	Hettich GmbH & Co. KG

Rotor 2424B	Hettich GmbH & Co. KG
Qik spin	Edwards Group Pty. Ltd., Australia
Optima™ L-90K ultracentrifuge	Beckmann Coulter GmbH, Krefeld
Rotor SW40 Ti	Beckmann Coulter GmbH, Krefeld
Rotor SW60 class GH	Beckmann Coulter GmbH, Krefeld
Centrifuge tubes for SW 40 and SW 60	Seton Scientific, Petaluma, USA
Universal 320 R	Hettich GmbH & Co. KG
Rotor 1420 A/B	Hettich GmbH & Co. KG
Rotor 1617 A	Hettich GmbH & Co. KG
Rotor 1620 A	Hettich GmbH & Co. KG

### Liquid chromatography

Äkta prime plus	GE Healthcare Europe, Munich
Äkta purifier	GE Healthcare Europe, Munich
HiPrep™ 26/10 desalting	GE Healthcare Europe, Munich
HisTrap™ FF 5 ml	GE Healthcare Europe, Munich
Superdex™ 75 HiLoad™ 16/60 prep grade	GE Healthcare Europe, Munich
Superdex™ 200 HiLoad™ 16/60 prep grade	GE Healthcare Europe, Munich
Superloop (50 ml, 150 ml)	GE Healthcare Europe, Munich
Frac-920 fraction collector	GE Healthcare Europe, Munich

### UV-VIS spectroscopy

NanoDrop	Thermo Scientific, USA
UV-Vis spectrometer, V-630	Jasco GmbH, Gross-Umstade
UV-Vis spectrometer, V-650	Jasco GmbH, Gross-Umstade
Precision cuvettes, suprasil	Hellma GmbH & Co. KG, Mühlheim

### Isothermal Titration Calorimetry

iTC200 Microcalorimeter	MicroCal, Northampton, USA
-------------------------	----------------------------

**Electron Microscopy**

(Department of Structural Dynamics,  
Max Planck Institute for Biophysical  
Chemistry, Göttingen)

TEM CM200 FEG	Philips
TEM Titan Krios	FEI Company
Vitrobot	FEI Company
Copper EM grids	Plano

**Atomic Force Microscopy**

(Third Institute of Physics, Georg-  
August-University Göttingen)

Asylum MFP3D	Oxford Instruments Asylum research, Wiesbaden
Cypher S	Oxford Instruments Asylum research, Wiesbaden
Nanotec AFM	Nanotec Electronica, Madrid, Spain
BL-AC40TS-C2 micro cantilever	Olympus
BL-RC150VB micro cantilever	Olympus
N'-(3-[Trimethoxysilyl]-propyl) diethylenetriamine (DETA)	SigmaAldrich, Munich

**Small Angle X-ray Scattering (SAXS)**

(Research Group: Nanoscale Imaging  
of Cellular Dynamics, Georg-August-  
University Göttingen )

Pilatus 100K detector	Dectris, Baden, Switzerland
Hecus S3-MICROpix camera system	SWAXS version, Hecus X-ray Systems
iMOXS X-ray source	IfG - Institute for Scientific Instruments GmbH
Glass capillaries	Hilgenberg GmbH

**Direct Stochastic Optical  
Reconstruction Microscopy  
(dSTORM)**

8-well imaging chamber	Sarstedt, Nümbrecht
------------------------	---------------------

**Mass spectrometry**

Orbitrap Fusion Tribrid Mass Spectrometer	Thermo Scientific
G1 mass spectrometer	Waters

**2.1.3. Commodities****Product****Supplier****Enzymes**

DNase	AppliChem, Darmstadt
FastAP (Thermosensitive alkaline phosphatase)	Thermo Fisher Scientific, Braunschweig
Lysozyme	AppliChem, Darmstadt
Phusion® high fidelity DNA polymerase	Thermo Fisher Scientific, Braunschweig
Restriction endonucleases	Thermo Fisher Scientific, Braunschweig
T4-DNA-Ligase	Thermo Fisher Scientific, Braunschweig
Trypsin (sequencing grade modified, porcine)	Promega Corporation, Mannheim
Thermolysin	SigmaAldrich, Munich

**Bacterial strains**

<i>E. coli</i> BL21 (DE3)	<i>In-vitrogen</i> ™, Karlsruhe
<i>E. coli</i> BL21 Star™ (DE3)	<i>In-vitrogen</i> ™, Karlsruhe
<i>E. coli</i> DH5α	<i>In-vitrogen</i> ™, Karlsruhe
<i>E. coli</i> SoluBL21 (DE3)	<i>In-vitrogen</i> ™, Karlsruhe
<i>E. coli</i> Top 10	<i>In-vitrogen</i> ™, Karlsruhe
XL1-Blue	Stratagene, Heidelberg
M15	QIAGEN

**Plasmids and vectors**

pET28a	Novagen, Schwalbach/Ts.
--------	-------------------------



pET28a- <i>hE2</i>	GeneArt, ThermoFisher Scientific
pET28a- <i>hE3</i>	GeneArt, ThermoFisher Scientific, Braunschweig
pET28a-His-eGFP- <i>hE2</i>	GeneArt, ThermoFisher Scientific, Braunschweig
pET28a-KD- <i>hE2</i> (catalytic domain)	Dr. Kathrin S. Tittmann
pET28a-SNAP- <i>hE2</i>	SNAP construct kindly received from Prof. Stefan Jakobs, MPIBPC
pET28a-His- <i>hE2</i>	Sub-cloning by self
pPDHE2/E3BP	Prof. Mulchand Patel
pQE-9-6HE1alphaE1beta	Prof. Mulchand Patel

### **Kits and solutions**

GC-buffer	Thermo Fisher Scientific, Braunschweig
HF-buffer	Thermo Fisher Scientific, Braunschweig
dNTP mix (10 mM)	Thermo Fisher Scientific, Braunschweig
NucleoSpin™ Plasmid kit	Macherey Nagel, Düren
NucleoSpin™ Gel and PCR clean up kit	Macherey Nagel, Düren
Bradford reagent (5x)	SERVA Electrophoresis GmbH, Heidelberg

### **Protein and DNA standards**

Gene Ruler™ 1 kb DNA-Ladder	Thermo Fisher Scientific, Braunschweig
PageRuler Unstained Protein Ladder	Thermo Fisher Scientific, Braunschweig
Prestained Protein Molecular Weight Marker	Thermo Fisher Scientific, Braunschweig
Unstained Protein Molecular Weight Marker	Thermo Fisher Scientific, Braunschweig

### **Miscellaneous**

Balance Kern EW	Kern, Balingen-Frommern
-----------------	-------------------------

Balance Kern ABJ	Kern, Balingen-Frommern
C1000 thermal cycler	BioRad laboratories GmbH, Munich
CFX96™ Optical Reaction Module	BioRad laboratories GmbH, Munich
Filters (0.2 µm and 0.4 µM)	Sartorius, Göttingen
Gradient Master 108	BioComp, Canada
Nitrocellulose membrane	PROTRAN BA 83, Whatman GmbH
pH meter FiveEasy	Mettler-Toledo, Giessen
SDS gel electrophoresis SE250	Hoefer, Holliston, USA
Spin-X® UF concentrator, 20 mL and 5 mL (MWCO 50 and 100 KDa)	Corning GmbH, Kaiserslautern
Vortex-Genie 2	Scientific Industries, USA
ZelluTrans/Roth® dialysis membrane (MWCO 3.5 KDa)	Carl Roth, Karlsruhe
Water bath Isotemp 202	Thermo Scientific, Schwerte
DNA loading dye (6 x)	Fermentas

#### 2.1.4. Primers

All the oligonucleotides were purchased from Sigma-Aldrich (Munich, Germany).

Name	Sequence (5'-3')
FP_Mut_NdeI_hE2	GCTATCCGCCGCACATGCAGGTTCTGC
RP_Mut_NdeI_hE2	GCAGAACCTGCATGTGCCGGCGGATAGC
FP_NdeI_polyHIS_hE2	GCATTAACATATGGGTTCACTGCCTCCGC
RP_XhoI_polyHIS_hE2	GGTACTCGAGTTACAGCAGCATGGTGAT CG
FP_SNAPtag	CGATGCTCTATAGCCATGGTGAGGTAGGGCG
RP_SNAPtag	CGCAGCATATGCTTGTACAGCTCGTCCATGC
T7 promotor primer	TAATACGACTCACTATAGGG
T7 terminator primer	GCTAGTTATTGCTCAGCGG

#### 2.1.5. Media compositions

**Lysogeny Broth (LB) Medium**                      **Per 200 ml**  
Modified from (Bertani 1951)

Tryptone	2 g
Yeast	1 g
NaCl	1 g
<b>Auto Induction Medium</b>	<b>Per 1 L</b>

Modified from (Studier 2005)

50 x 52

Glycerol	250 g
D-Glucose	25 g
D-Lactose	100
20 x NPS	
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	66 g
KH <sub>2</sub> SO <sub>4</sub>	136 g
Na <sub>2</sub> HPO <sub>4</sub>	142 g

#### **TB medium**

Tryptone	10 g
Yeast Extract	2 g

#### **SOC media**

Tryptone	2 % w/v
Yeast extract	0.5 % w/v
NaCl	0.05 % w/v
KCl	2.5 mM
MgCl <sub>2</sub>	10 mM
Glucose	2 % w/v

#### **LB-agar plates**

1.5 % agar was added to LB media and heated to dissolve agar. The solution was cooled down and supplemented with required antibiotics. The mix was poured to the petri dishes before it could solidify.

#### **Antibiotics concentration**

Kanamycin 25-50 µg/ml

Carbenicillin 100 µg/ml

Chloramphenicol 35 µg/ml

**2.1.6. Software and web resources**

Adobe Photoshop CS6	Adobe
ATSAS 2.6.0	<a href="https://www.embl-hamburg.de/biosaxs/software.html">https://www.embl-hamburg.de/biosaxs/software.html</a>
UCSF Chimera	<a href="https://www.cgl.ucsf.edu/chimera/">https://www.cgl.ucsf.edu/chimera/</a>
Chromas 1.45	McCarthy, C., Griffith University, Australia
Clustal Omega	<a href="http://www.ebi.ac.uk/Tools/msa/clustalo/">http://www.ebi.ac.uk/Tools/msa/clustalo/</a>
CS ChemBioDraw Ultra 14.0	CambridgeSoft, UK
ExPasy, ProtParam	<a href="http://web.expasy.org/protparam/">http://web.expasy.org/protparam/</a>
Gene runner V.3.05	Hastings Software, Inc.
MicroCal Analysis	Malvern
Microsoft Office 2013	Mocrosoft
Origin 7	OriginLab
Proteome-Discoverer 1.0.0	Thermo Fisher Scientific
PSIPRED	<a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a>
PyMol v0.99	DeLano Scientific LLC
Robetta	<a href="http://rosetta.bakerlab.org/">http://rosetta.bakerlab.org/</a>
SigmaPlot 11	Systat Software, Inc
SUMMIT_1_0	Taverner, Hernandez et al. 2008
UNIPROT	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>
WSxM 5.0 Develop 7.0	WSxM solutions (Horcas, Fernandez et al. 2007)

## 2.2. Methods

### 2.2.1. Molecular biology

#### 2.2.1.1. Determination of DNA concentration

The concentration and purity of DNA samples were determined using NanoDrop instrument by simultaneously measuring the absorbance at 260 nm, 280 nm and 230 nm. The concentration of the double stranded (ds) DNA is calculated from the relation  $A_{260} = 1 = 50 \mu\text{g/ml ds DNA}$ . The significant absorbance at 280 nm and 230 nm indicate contaminations of RNA, protein and other reagents used during purification protocol. The isolated DNA was deemed to be free of such contaminant when the ratio between  $A_{260}/A_{280}$  and  $A_{260}/A_{230}$  were  $\sim 1.8$  and 2.0-2.2, respectively.

#### 2.2.1.2. Polymerase chain reaction (PCR)

A typical PCR mix had a total volume of 50  $\mu\text{L}$ , which was equally divided into five Eppendorf tubes. The reaction was run with a temperature gradient such that the tubes will cover maximum, minimum and intermediate temperatures. After optimizing temperature and reagent concentration, the total volume was scaled up to 25  $\mu\text{L}$  per aliquot.

Following were the parameters used for PCR:

S.No.	Temperature (° C)	Time (m:s)	goto	loops
1	98	00:30	-	-
2	98	00:10	-	-
3	62-72	00:20	-	-
4	72	00:30	2	29
5	72	05:00	-	-
6	4	Pause		

Following were the components of reaction mixture:

Reagents	Volume ( $\mu\text{L}$ )	$C_{\text{end}}$
H <sub>2</sub> O	31	
5 x HF buffer	10	1 x
dNTPs (10 mM of each nucleotide-5'-triphosphate)	1	200 $\mu\text{M}$
Forward primer (10 $\mu\text{M}$ )	2.5	0.5 $\mu\text{M}$
Reverse primer (10 $\mu\text{M}$ )	2.5	0.5 $\mu\text{M}$
Template DNA (10 ng/ $\mu\text{L}$ )	1	0.2 ng/ $\mu\text{L}$
DMSO	1.5	3 % w/v
Phusion DNA Polymerase (2 U/ml)	0.5	0.02 U/ml
<b>Total</b>	<b>50</b>	

#### 2.2.1.3. Agarose gel electrophoresis (AGE)

Agarose gel electrophoresis was performed with a 1 % agarose gel and TAE running buffer (40 mM Tris, 1 mM EDTA, 20 mM acetate, pH 8.5). Sample aliquots were mixed with loading dye (6 x DNA loading dye, Fermentas) and pipetted onto the wells, alongside DNA ladder (Gene Ruler™ 1 kb DNA-Ladder, Fermentas). The electrophoresis was carried out at 100 V, 60 mA and 150 W condition for 40 minutes. The region of interest in the gel were cut and soaked into ethidium bromide solution (2 mg/ml) for 10 min. The DNA bands were observed under UV light in a gel chamber (Department of Plant Biochemistry). The print out of the gel was made and if required, DNA bands of further interest were cut out before disposing the gel.

#### 2.2.1.4. Restriction digestion

DNA was digested by appropriate restriction endonucleases as described in the manufacturer's manual (Thermo Fisher Scientific).

### **2.2.1.5. Ligation**

AGE was performed after restriction digestion to isolate DNA bands of interest. Then, ligation reaction was performed using T4 DNA ligase according to manufacturer's manual (Thermo Fisher Scientific).

### **2.2.1.6. Bacterial transformation**

Post ligation mixtures were used to transform chemo-competent *E. coli* strains amenable for cloning purposes (DH5 $\alpha$ , Top 10 or XL1-Blue). For expression of recombinant proteins, chemocompetent strains BL21 (DE3), BL21Star<sup>TM</sup> (DE3) or SoluBL21 (DE3) were prepared (Inoue, Nojima et al. 1990).

Shortly, the chemo-competent cells were thawed from -80 °C freezer on ice for 10-20 min. 10 ng plasmid DNA was added to 50  $\mu$ L cells and incubated on ice for at least 5 min. The mixture was incubated in a water bath at 42 °C for 45 sec. Subsequently, it was cooled down on ice for 2 min. SOC media (500  $\mu$ L - 1 ml) pre-incubated at 37 °C was added to the transformed cell mix and allowed to grow at 37 °C and 750 RPM for 30 min – 45 min in a thermomixer. 100  $\mu$ L of this culture was spread onto one half of LB-agar plate (supplemented with desired antibiotics for screening). Rest of the culture was briefly centrifuged and most of the solution from the top was removed. The pelleted cells were re-suspended in remaining 50-100  $\mu$ L liquid and spread onto another half of the LB-agar plate and incubated overnight at 37 °C (Froger and Hall 2007).

The colonies obtained were used in inoculating pre-culture (200 ml LB media) for protein over expression. However, for cloning purpose, they were picked to inoculate 20 mL LB media and the grown cells were used to isolate plasmids according to manufacturer's manual (NucleoSpin<sup>TM</sup> Plasmid kit).

### **2.2.1.7. DNA sequencing**

For sequencing, 600-800 ng DNA were mixed with the corresponding sequencing primers (20 pM) and provided to SeqLab (SeqLab Sequence Laboratories GmbH, Göttingen, Germany). The result of the sequencing was analyzed using Chromas 1.45 to validate the insert sequences and their mutations.

## **2.2.2. Protein chemistry**

### **2.2.2.1. Determination of protein concentration**

A commercially available 5 x Bradford reagent (SERVA Electrophoresis GmbH, Heidelberg, Germany) was diluted in double distilled water to make 1 x solution. 10  $\mu$ L protein solution was added to 990  $\mu$ L of 1 x Bradford reagent, incubated at room temperature for 5 min and absorbance at 595 nm was noted. Protocol was modified from (Bradford 1976).

First, a standard curve of absorbance at 595 nm vs protein concentration was generated for BSA (0.1 mg/ml to 0.9 mg/ml). This curve was used to calculate the protein concentration in the query sample. A fresh 1 x Bradford reagent and the calibration curve was prepared every two days to ensure accuracy.

### **2.2.2.2. Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)**

SDS-PAGE was performed based on a modified protocol from Laemmli 1970, Weiner, Weber et al. 1972. The gels were casted such that  $\sim 1/4^{\text{th}}$  part from the top will be stacking gel (5 % acrylamide/bis-acrylamide mix, 0.1 % w/v SDS, 125 mM Tris, pH 6.8, polymerized with 0.03 % w/v APS and 0.2 % v/v TEMED) and remaining part will be separating gel (12 % acrylamide/bis-acrylamide mix, 0.1 % w/v SDS, 375 mM Tris, pH 8.8, polymerized with 0.03 % w/v APS and 0.1 % v/v TEMED). Protein samples were mixed with SDS loading buffer (100 mM Tris-HCl, pH 6.8, 0.02 % w/v bromophenol blue, 25 % w/v glycerol, 2 % w/v SDS and 2 % v/v  $\beta$ -mercaptoethanol) and heated at 95  $^{\circ}$ C for 5 min. The denatured samples were loaded into the wells. As a standard, a protein molecular weight marker was always co-electrophoresed. Electrophoresis was carried out in SDS running buffer (25 mM Tris, 200 mM glycine, 0.1 % w/v SDS, pH 8.3) at 30 mA, 275 V and 150 W conditions for 40 min. After completion of electrophoresis, the gels were stained by an incubation step in Coomassie staining solution (0.25 % w/v Coomassie Brilliant Blue R-250, 30 % ethanol, 6 % acetic acid) for 20 min. The de-staining of the gels was done by shaking them in a de-staining solution (30 % v/v methanol and 10 % v/v acetic solution in water).



### 2.2.2.3. Over expression of proteins

#### 2.2.2.3.1. High cell density fermentation for *hE1*

*E. coli* (M15 strain) cells harboring pQE-9-6HE1 $\alpha$ /E1 $\beta$  plasmid (Korotchkina, Tucker et al. 1995) were grown on a LB-agar plate containing 25  $\mu$ g/mL kanamycin and 100  $\mu$ g/mL carbenicillin overnight at 37 °C. Single colonies were picked and inoculated into each of the two 1 L chicane flasks containing autoclaved LB medium (200 mL). These primary cultures were kept shaking at 200 RPM overnight at 30 °C. Next day, the cultures were pooled together and centrifuged at 5000 x g at 5 °C for 20 minutes. Thus obtained cell pellet was gently re-suspended in 50 ml fresh LB medium (autoclaved).

Separately, 5.6 L of yeast extract [300 g yeast extract (Ohly, GmbH) + 3 g NH<sub>4</sub>Cl + 1 mL anti-foam] was autoclaved in a BioStat C biofermenter (Sartorius AG, Göttingen). Using the feed tubes, 250 ml K<sub>2</sub>HPO<sub>4</sub> (26.4 % w/v), 100 ml glucose (30 % w/v) and 50 ml MgSO<sub>4</sub> (8.16 % w/v) were added. From the septum, 6 ml each of kanamycin (25 mg/ml) and carbenicillin (100 mg/ml) were injected. The pH was kept constant at 7 by connecting the fermentation chamber to 10 % w/v H<sub>3</sub>PO<sub>4</sub> and 10 % w/v NaOH solutions. The temperature and partial oxygen pressure (pO<sub>2</sub>) were set at 37 °C and 30 % respectively.

Once the temperature and pO<sub>2</sub> set conditions are reached, the re-suspended primary culture was added to the fermentation chamber via feed tube. The stirring speed of the culture (200 to 1500 RPM) and the airflow (5-25 L/min) were maintained to keep pO<sub>2</sub> at 30 %. To ensure optimal growth of the bacteria every hour in the beginning and later every 30 min, optical density at 600 nm (OD<sub>600</sub>) was determined. At the same time, its glucose level was monitored by Midi Test Glucose strip (Macharey-Nagel). Once, it's confirmed that glucose is completely consumed; feeding process was started with a flow rate of 15 %. The autoclaved feed contained 600 g of yeast extract (Ohly, GmbH) and 500 mL glycerol dissolved in 2 L deionized water. Additionally, corresponding antibiotics (2 ml kanamycin and 4 ml carbenicillin) were also added.

When the OD<sub>600</sub> reached 40, temperature was lowered to 25 °C (in small steps of 3 - 5 °C). Once the temperature reached 25 °C, 8 ml thiamine hydrochloride (100 mg/ml) was added. Finally, protein over expression was induced by adding 1 mM IPTG. Next day, the culture was harvested by centrifugation step at 4800 RPM (JLA 8.1000 rotor), 6 °C for 30 min. The cells were separated into 300 g portions and stored in a plastic bag at -80 °C. The samples for SDS-PAGE analysis were made from cells just before IPTG addition and post-harvest, to confirm successful *hE1* over expression.

This protocol was adapted from the one developed in the lab by Dr. Kathrin S. Tittmann and described in a publication (Seifert, Golbik et al. 2006).

#### **2.2.2.3.2. Auto-induction of *hE3*, *hE2*, *hE2/E3BP*, *KD-hE2*, *eGFP-hE2* and *SNAP-hE2* (all in pET28a expression vector)**

A plasmid containing desired gene was transformed into *E.coli* strains designed for over expression purpose (refer Section 2.2.1.6). Single colonies were picked and inoculated into each of the two 1 L chicane flasks containing autoclaved LB medium (200 mL) and desired\* antibiotics. These primary cultures were kept shaking at 200 RPM overnight at 37 °C.

Separately, a main culture, which is 465 ml of TB media, was added to 2 L chicane flask and autoclaved. Desired antibiotics\*, 10 ml 50 x 5052 solution, 25 ml 20 x NPS solution and 500 µL MgSO<sub>4</sub> (1 M) were added. For *hE3* and *hDPHc* cores (except *KD-hE2* core), 500 µL FMN (30 mg/ml) and 750 µL lipoic acid (0.25 M) were further supplemented, respectively. Then, respective primary cultures were added to main cultures to reach OD<sub>600</sub> of ~ 0.1. It was then grown at 37 °C while shaking at 200 RPM. Once, the OD<sub>600</sub> reached 0.6-0.8, the temperature was lowered to 16 °C and allowed to grow for at least two days. The cells were harvested by centrifuging at 9000 RPM (1620 A rotor, Hettich GmbH & Co. KG), 6 °C for 30 min and stored at -80 °C. The samples for SDS-PAGE analysis were made from cells just before temperature switch and post-harvest, to confirm successful protein over expression.

\*For overexpression from plasmids pET28a-*hE3*, pET28a-*hE2*, pET28a-*KD-hE2*, pET28a-*SNAP-hE2* and pET28a-*eGFP-hE2*, 50 µg/ml kanamycin was added to the media. Whereas, for overexpression from pPDHE2/E3BP, 35 mg/ml chloramphenicol was added to the media.

#### **2.2.2.4. Purification of proteins**

##### **2.2.2.4.1. Purification of *hE1***

100 g cell pellet was re-suspended in cold 400 ml lysis buffer (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.0, 300 mM KCl, 20 mM Imidazole and 1 mM DTE) supplemented with 1 mM PMSF. Lysozyme (0.25 g) was added and the cell suspension was stirred in a cold room for minimum of 30 min or until the cells suspension became homogenous. Then, DNase I (5 mg) and MgCl<sub>2</sub> (C<sub>end</sub> of 3 mM) were added and the cell suspension was stirred in a cold room for another 30 min. This suspension was passed through fluidizer (2-3 passages) at 70-80 psi for complete cell lysis. The lysed cell suspension was centrifuged at 45000 x g, 6 °C for 30 min and the clear supernatant was decanted in a beaker.

A HisTrap™ FF column was equilibrated with 5 column volumes (cv) of lysis buffer and then, cell lysis supernatant was injected to the column. After this, the column was washed with 3 cv of lysis buffer. Now, the bound *hE1* protein was selectively eluted by applying a gradient of 0 % - 100 % over 10 cv of elution buffer (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.0, 300 mM KCl, 300 mM Imidazole). SDS-PAGE was performed for elution fractions and the pure looking fractions were pooled and concentrated via centrifugation-based concentrator (MWCO 100 KDa) to reach volume < 1.5 ml.

The concentrated protein was injected to the Superdex™ 75 HiLoad™ 16/60 prep grade column pre-equilibrated with 3 cv of *hE1*-buffer (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.6, 300 mM KCl). The eluted protein fractions were again subjected to SDS-PAGE analysis and the pure fractions were pooled together and concentrated like earlier. The protein concentration was estimated by Bradford assay (2.2.2.1.) The pure *hE1* protein was stored on ice for further usage.

This protocol was adapted from Seifert, Golbik et al. 2006.

##### **2.2.2.4.2. Purification of *hE3***

5 g - 7.5 g of cell pellet was re-suspended in cold 20 ml lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 20 mM Imidazole, 200 mM NaCl) supplemented with 1 mM PMSF. Spatula tip lysozyme was added and the cell suspension was stirred in a cold room for minimum of 30 min. Then, spatula tip DNase I and MgCl<sub>2</sub> (C<sub>end</sub> of 3 mM) were added and the cell suspension was stirred in a cold room for another 30 min. The step for complete cell lysis using fluidizer is same as for *hE1* protein (refer Section 2.2.2.3.1).

The purification step using His-trap column is same as for *hE1* protein except the buffers. For lysis, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 200 mM NaCl, 20 mM imidazole and for elution 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 200 mM NaCl, 100 mM imidazole were used. SDS-PAGE was performed for elution fractions and the pure looking fractions were pooled and concentrated via centrifugation-based concentrator (MWCO 50 KDa) to reach volume of 2 ml. The protein concentration was estimated by Bradford assay (2.2.2.1.). FAD (5 times protein concentration) was added and the protein was kept overnight on ice.

Next day, the concentrated protein supplemented with FAD was injected to the Superdex™ 200 HiLoad™ 16/60 prep grade column pre-equilibrated with 3 cv of *hE3*-buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8, 200 mM NaCl). The eluted protein fractions were again subjected to SDS-PAGE analysis and the pure fractions were pooled together and concentrated like earlier. The pure *hE3* protein was stored on ice for further usage.

#### **2.2.2.4.3. Purification of *hE2*, *hE2/E3BP*, *KD-hE2*, *eGFP-hE2* and *SNAP-hE2***

Cells (35 g – 45 g) were mixed with 100 g alumina and 20 ml of lysis buffer (supplemented with 1 mM PMSF) and grinded manually in a mortar using a pestle, to reach a dough like consistency. Then, the mortar and pestle was mounted to Mortar Grinder RM 200, Retsch GmbH and the cells were grinded for 45 min. Upon reaching 35 min, lysozyme and DNase I (both spatula tip) was added along with 80 ml of lysis buffer (supplemented with 1 mM PMSF).

The cell suspension was centrifuged at 9000 RPM (1620 A rotor, Hettich GmbH & Co. KG), 6 °C for 30 min to remove alumina and cell debris partially. The supernatant was again centrifuged at 45000 x g, 6 °C for 30 min to remove all the cell debris. The supernatant was brought to room temperature and 0.1 % w/v protamine sulfate was added and stirred at room temperature for 30 min. Then, it was centrifuged at 15000 x g, 20 °C for 20 min. The clear supernatant was incubated on ice for 10 min and cold PEG-6000 (50 % stock) was added to reach first cut off and stirred for 30 min in a cold room. It was centrifuged at 15000 x g, 4 °C for 20 min. PEG-6000 (50 % stock) was added to the supernatant to reach second cut off, stirred for 30 min in a cold room and centrifuged at 15000 x g, 4 °C for 20 min. The supernatant this time was slowly decanted and the pellet was washed with 5 ml of final buffer to remove excess PEG. The walls of centrifuge tube were dried with a tissue paper. Depending on the size of the pellet, 400 µL to 1 ml of final buffer was added and the pellet was re-suspended by gentle shaking.

When the protein is completely solubilized, it is added on the top of 10 % - 30 % linear sucrose gradient in a SW 40 rotor tubes and centrifuged at 20000 RPM - 25000 RPM, 4 °C for 16 hr. The sucrose solutions were prepared in the \*final buffer and the linear gradient was made by Gradient Master 108 according to the manufacturer's manual (BioComp, Canada). Once the ultracentrifugation is finished, the gradient was fractionated from the top with 400 µL fraction size using micropipette. These fractions were subjected to SDS-PAGE analysis and the pure fractions were pooled together. PEG-6000 (50 % stock) was added to the pure protein pool until protein precipitation was visible (normally at 6 % - 8 %). It was centrifuged at 15000 x g, 4 °C for 20 min and the protein pellet was solubilized in final buffer. In order to remove the insoluble part, a further centrifugation step was performed at 15000 RPM (1420 B, Hettich GmbH & Co. KG), 4 °C for 15 min. The supernatant was transferred to fresh Eppendorf tube and kept on ice until further usage.

\*Lysis buffer for *hE2*, *hE2/E3BP*, *eGFP-hE2* and *SNAP-hE2* is 50 mM HEPES, pH 7.5, 150 mM KCl, 20 mM Mg(OAc)<sub>2</sub>, 30 mM NH<sub>4</sub>Cl and 1 mM DTE whereas for *KD-hE2* is 50 mM MES, pH 6, 300 mM KCl, 20 mM Mg(OAc)<sub>2</sub>, 30 mM NH<sub>4</sub>Cl and 5 mM DTE. Final buffer here means lysis buffer without Mg(OAc)<sub>2</sub> and NH<sub>4</sub>Cl.

\*The first and second PEG cut off were 4 % and 6 %, respectively for *hE2*, *hE2/E3BP*, *eGFP-hE2* and *SNAP-hE2* whereas for *KD-hE2*, they are 2 % and 4 %, respectively.

#### **2.2.2.5. Proteolysis of *hE2* and *hE2/E3BP* core**

Thermolysin protease was added to *hE2* and *hE2/E3BP* to a mass ratio of 1:400 and 1:100, respectively. They were then incubated at 20 °C for 60 min in a thermomixer with shaking at 700 RPM. 5 mM EDTA was added to quench the reaction and the protein mix was loaded on top of 10 % - 30 % linear sucrose gradient. The sucrose gradient ultracentrifugation and protein precipitation part is same as in section 2.2.2.3.3. Exception being the final buffer used which is 50 mM MES, pH 6, 150 mM KCl for *hE2* and 50 mM Bis-Tris, pH 6.5, 150 mM KCl for *hE2/E3BP*.

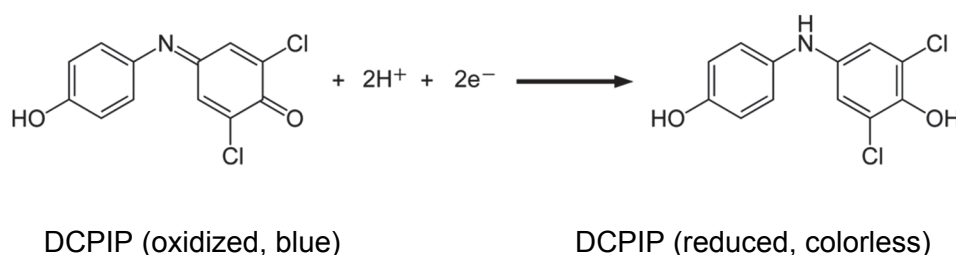
### 2.2.2.6. In-vitro reconstitution of *hPDHc*, *hE1-hE2/E3BP* sub-complex and *hE3-hE2/E3BP* sub-complex

Purified *hE2/E3BP* core was mixed with *hE1* or *hE3* or both, depending on target complex. Assuming each core consist 40 *hE2* and 20 *E3BP* proteins, *hE1* and *hE3* were added such that they would be in molar excess (3 - 5 times) compared to their binding sites at *hE2* and *E3BP*, respectively. After incubating in a cold room for 15 min, the protein mix was loaded onto 10 % - 30 % linear sucrose gradient. The gradient was prepared in a high salt buffer (50 mM HEPES pH 7.6, 300 mM KCl and 1 mM DTE) if the target complex consists *hE1*. Else, low salt buffer was used which consists 150 mM KCl instead of 300 mM. To reconstitute holo-enzyme complex, cofactors and substrates such as CoA, ThDP, MgCl<sub>2</sub> and NAD<sup>+</sup> were added to the protein mix and buffers, with final concentrations of 2 mM, 0.1 mM, 1 mM and 0.2 mM. The sucrose gradient ultracentrifugation and protein precipitation part is same as in section 2.2.2.3.3.

### 2.2.3. Activity assays

#### 2.2.3.1. DCPIP assay

This assay was carried out to investigate enzymatic activity of *hE1* protein. An artificial electron acceptor namely 2,6-dichlorophenolindophenol (DCPIP) which is dark blue in color was used to report the HEThDP intermediate formation. Once the DCPIP receives two electrons from this intermediate, it turns colorless. This process can thus be followed as depletion of absorbance at 600 nm ( $A_{600}$ ).



Initial rate of  $A_{600}$  reduction can be correlated with pyruvate processing using following equation:

$$\text{Specific activity, } A_{\text{spec}} \left( \frac{\text{mU}}{\text{mg}} \right) = \frac{(-\Delta \text{Abs}/\text{min} * 10^9)}{\epsilon_{600} * C_{\text{enz}} * V_{\text{enz}} * d}$$

Where, 1 unit (U) = micromoles of DCPIP reduced/min,  $-\Delta \text{Abs}/\text{min}$  = slope of the progress curve (absorbance vs time),  $\epsilon_{600}$  = molar extinction coefficient of DCPIP at 600 nm =  $17700 \text{ M}^{-1}\text{cm}^{-1}$ ,  $C_{\text{enz}}$  = concentration of *hE1* (stock) mg/ml,  $V_{\text{enz}}$  = volume of *hE1* added in  $\mu\text{L}$  and  $d$  = path length of the cuvette = 1 cm.

The protocol was adapted from (Nemeria, Yan et al. 2001) and the experimental parameters were as follows:

Stock solution	Final concentration
5 x <i>hE1</i> buffer (0.5 M $\text{KH}_2\text{PO}_4$ , 1.5 M KCl, pH 7.6)	0.1 M KPP, 0.3 M KCl, pH 7.6
0.1 M pyruvate	10 mM
0.01 M ThDP + 0.1 M $\text{MgCl}_2$	0.2 mM ThDP + 2 mM $\text{MgCl}_2$
Enzyme ( <i>hE1</i> )	0.4 mg/mL
DCPIP (10 mM)	0.1 mM

UV/Vis spectrophotometer V-650 (Jasco GmbH)	Parameters
Wavelength	600 nm
Data pitch	5 sec
Temperature	30 °C
Time	1000 sec

890  $\mu\text{L}$  pre-reaction mixture was made with 200  $\mu\text{L}$  of 5 x *hE1* buffer, 20  $\mu\text{L}$  ThDP +  $\text{MgCl}_2$  mix, required volume of *hE1* and water in a 1 ml cuvette. It was incubated on ice for 5 min. Then, 10  $\mu\text{L}$  DCPIP was added, quickly mixed by pipetting and absorbance monitored at 600 nm for 600 sec. The reaction was started by adding 100  $\mu\text{L}$  pyruvate and absorbance monitored for further 400 sec. The slope from 620 sec - 650 sec was noted down for calculating specific activity of *hE1*.

### 2.2.3.2. PDHc assay

The overall activity of an in-vitro reconstituted human PDH complex (assumed to be 40hE1:40hE2:20E3BP:20hE3) and its dependency on substrate concentrations were studied using this assay. The protocol was adapted from (Korotchkina and Patel 2001), where basically reduction of NAD<sup>+</sup>, which represents the final step of PDH reaction, is monitored. The formation of NADH can be spectrophotometrically observed at 340 nm. The rate of increase in absorbance corresponds to specific activity of hPDHc by following equation:

$$\text{Specific activity, } A_{\text{spec}} \left( \frac{\text{U}}{\text{mg}} \right) = \frac{(-\Delta \text{Abs}/\text{min} * 10^6)}{\epsilon_{600} * C_{\text{enz}} * V_{\text{enz}} * d}$$

Where, 1 unit (U) = micromoles of NADH produced/min,  $-\Delta \text{Abs}/\text{min}$  = slope of the progress curve (absorbance vs time, 0-25 sec),  $\epsilon_{600}$  = molar extinction coefficient of DCPIP at 600 nm = 6220 M<sup>-1</sup>cm<sup>-1</sup>,  $C_{\text{enz}}$  = concentration of hPDHc (stock) mg/ml,  $V_{\text{enz}}$  = volume of hPDHc added in  $\mu\text{L}$  and  $d$  = path length of the cuvette = 1 cm.

Similarly, reaction velocity ( $v$ ) can be calculated as,

$$\text{reaction velocity (V)} = \frac{(-\Delta \text{Abs}/\text{min} * 10^6)}{\epsilon_{600} * d}$$

For only checking activity, 100  $\mu\text{L}$  buffer, 705.05  $\mu\text{L}$  H<sub>2</sub>O, 20  $\mu\text{L}$  L-cysteine, 20  $\mu\text{L}$  ThDP + Mg Cl<sub>2</sub> mix, 31.2  $\mu\text{L}$  CoA, 100  $\mu\text{L}$  NAD<sup>+</sup> and 3.75  $\mu\text{L}$  hPDHc were added in a 1 mL cuvette. It was incubated at 37 °C for 3 min and used as a blank. Then, 20  $\mu\text{L}$  pyruvate was added and mixed quickly by pipetting. A measurement was started at 340 nm for 300 seconds. From the absorbance vs time curve, the slope from 0 sec to 25 sec was noted down to calculate specific activity.



The experimental parameters were as follows:

<b>Stock solution</b>	<b>Final concentration</b>
PDH buffer (0.5 M KH <sub>2</sub> PO <sub>4</sub> , pH 7.5)	50 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.5
<i>h</i> PDHc (0.5 mg/ml)	1.874 mg/ml (represents 1 µg of <i>h</i> E1 in original protocol)
0.01 M ThDP + 0.1 M MgCl <sub>2</sub>	0.2 mM ThDP + 2 mM MgCl <sub>2</sub>
L-cysteine (200 mM)	4 mM
CoA (5 mM)	0.156 mM
NAD <sup>+</sup> (20 mM)	2 mM
Pyruvate (100 mM)	2 mM
<b>UV/Vis spectrophotometer V-650 (Jasco GmbH)</b>	<b>Parameters</b>
Wavelength	340 nm
Data pitch	1 sec
Temperature	37 °C
Time	300 sec

For steady state kinetics purpose, concentrations of all the constituents in a reaction mixture were kept same, except for a substrate being investigated. Variable concentrations of pyruvate (6.25, 12.5, 25, 50, 100, 250, 500, 1000 and 2000 µM), CoA (5, 10, 25, 50, 100, 250 and 500 µM) and NAD<sup>+</sup> (10, 25, 50, 100, 250, 500, 1000 and 2000 µM) were used in order to calculate  $V_{\max}$  and  $K_M$  values using Michaelis-Menten equation.

$$V = \frac{V_{\max} * [S]}{K_M + [S]}$$

Where,  $V$  = reaction velocity (µM/min) =  $A_{\text{spec}} * \text{enzyme amount in } \mu\text{g}$ ,  $V_{\max}$  = maximum reaction velocity,  $[S]$  = substrate concentration and  $K_M$  = Michaelis constant.

### 2.2.4. FAD saturation test

The proportion of FAD bound active site in *hE3* protein was estimated for each batch of purified protein. First of all, an absorbance spectrum from 300 nm to 700 nm for 1 mg/ml *hE3* in 1 cm quartz cuvette was taken at room temperature. The protein was transferred to an Eppendorf tube and heated at 95 °C for 10 min for precipitation, releasing bound FAD to solution. The coagulated *hE3* was centrifuged at 15000 RPM (1420 B, Hettich GmbH & Co. KG), 20 °C for 10 minutes. The yellow supernatant was transferred back to cuvette and fresh spectra taken for the same wavelength range. The absorbance at 450 nm ( $A_{450}$ ) which is a  $\lambda_{\max}$  for free FAD, was used to calculate its concentration ( $[FAD]_{\text{free}}$ ). It must be same to *hE3* monomer concentration (19.058  $\mu\text{M}$ ) in case of 100 % saturation, as each monomer can bind to one FAD.

$[FAD]_{\text{free}} = A_{450} / \epsilon_{450}$ , where  $\epsilon_{450} = 11300 \text{ M}^{-1}\text{cm}^{-1}$  is molar extinction coefficient for free FAD at 450 nm.

### 2.2.5. Isothermal titration calorimetry (ITC)

#### 2.2.5.1. Background

Within a community of techniques, which can calculate binding constants of biomolecular interaction, ITC holds a special place. Till date, it is an only method that can provide information on association constant ( $K_a$ ) and thermodynamic parameters such as binding enthalpy ( $\Delta H$ ), Gibb's free energy of binding ( $\Delta G$ ) and binding entropy ( $\Delta S$ ), in a single experiment. Furthermore, by doing measurements at multiple temperature, heat capacity of binding ( $\Delta C_p$ ) can also be calculated, which aids in dissecting the nature of binding interfaces (Falconer, Penkova et al. 2010).

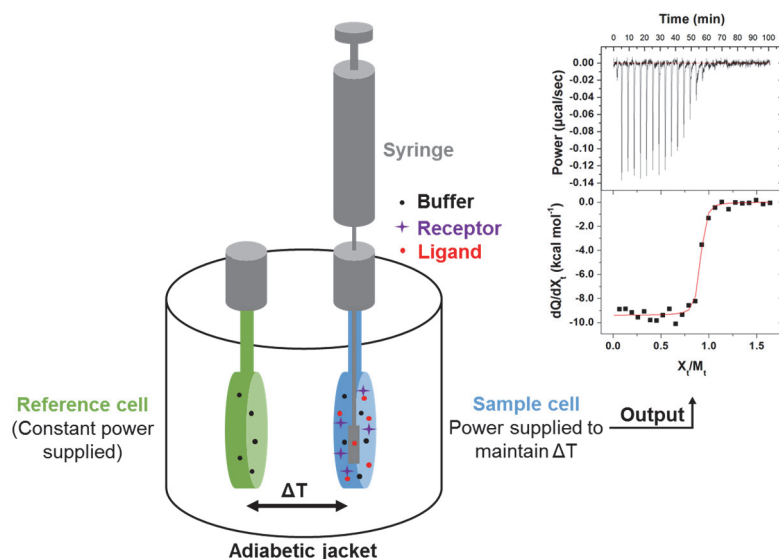
A receptor solution is kept in a temperature controlled sample cell which is coupled to a reference cell, usually filled with water or buffer, by a thermopile/thermocouple circuits (Fig. 14). Both of these cells are kept inside an adiabatic jacket. The thermopile is sensitive enough to detect tiny changes in temperature (nanojoules) between the two cells (Pierce, Raman et al. 1999). A ligand solution is then injected to the sample cell in aliquots, which upon binding to receptor molecules, most likely, releases or absorbs heat. This will bring a tiny change in temperature in the sample cell, which is allowed to equilibrate with the reference cell by using external power. By integrating the power with respect to the time required for equilibration of temperature between sample and reference cell for a series of injections,  $\Delta H$  is calculated. As the observed heat signal can be treated as an indicator of ligand-receptor complex

formation,  $K_a$  can be calculated from the plot between heat signal per mole of added ligand to the molar ratio between ligand and receptor from the following equation (Wiseman, Williston et al. 1989), assuming 1:1 binding (stoichiometry,  $N = 1$ ):

$$\frac{1}{V} * \frac{dQ}{dX_t} = \Delta H \left( \frac{1}{2} + \frac{1-(1+r)/2 - X_r/2}{\{(X_r)^2 - 2X_r(1-r) + (1+r)^2\}^{1/2}} \right)$$

Where,  $X_t$  is total ligand concentration in the volume  $V$ ,  $Q$  is the heat signal,  $\Delta H$  is the enthalpy of binding and  $M_t$  is the total receptor concentration. The parameters  $r$  and  $X_r$  are given as  $1/r = c = N * M_t * K_a$  and  $X_r = X_t / M_t$ .

Once the values for  $\Delta H$  and  $K_a$  are known,  $\Delta G$  and  $\Delta S$  can be calculated by simply using general thermodynamic equations,  $\Delta G = -RT \ln K_a = \Delta H - T \Delta S$ . To the plot of  $dQ/dX_t$  to  $X_r$  from experimental data, the above-mentioned equation was fitted, assuming 1 ligand binds to only 1 receptor molecule. This resulted in the stoichiometry value for the best fitting of obtained data.



**Fig.14: Schematic of an ITC calorimeter.**

A constant temperature difference is maintained between sample and reference cells during the experiment. Upon injection of ligand to the sample cell containing receptor molecules, receptor-ligand complexes are formed which might be exothermic or endothermic process, thereby subtly changing the temperature of the cell. External power is used to bring back the sample cell temperature to original level, which is the raw data. For each injection, the power required to maintain the set temperature when integrated with respect to equilibration time, yields the heat of interaction for each injection. The plot between heat of interaction per mole of ligand to the molar ratio between ligand and receptor at each injection is the final output.

### 2.2.5.2. Experiment

ITC binding experiments were conducted principally, to investigate CoA and *hE3* binding to *hE2/E3BP* core. The enzymes to be used were kept inside separate dialysis membrane chamber (ZelluTrans/Roth® dialysis membrane, MWCO 3.5 KDa), and kept together in a beaker containing 1 L of 50 mM HEPES, 150 mM KCl, pH 7.5, with gentle stirring overnight in a cold room. The CoA solution was prepared in the post-dialysis buffer. An iTC200 Microcalorimeter was kept at 25 °C (*hE3-hE2/E3BP*) or 30 °C (*CoA-hE2/E3BP*) and the sample cell was filled with dialyzed ~ 0.5-2 µM *hE2/E3BP* core solution. A dummy injection of 0.1 µL followed by 24 injections of 1.5 µL of ligands i.e. CoA (2 mM) or *hE3* (~140-210 µM) were made to the sample cell with constant stirring at 500 RPM. A control experiment was also done where sample cell was filled with post-dialysis buffer keeping all the parameters and ligand concentration same. This control data set containing heat of dilutions was subtracted from the experimental one and resulting data set was analyzed further via MicroCal Analysis software for base line correction and data fitting.

**Table 3: Parameters of ITC experiment**

Experimental parameters	Total injections	Temperature	Reference power	Stirring speed	Feedback mode
	25	25 °C / 30 °C	7 µCal/sec	500 RPM	High
Injection parameters	Volume of a dummy injection	Volume per injection	Duration	Spacing	Filter period
	0.1 µL	1.5 µL	3 sec	180 / 240 sec	2 sec

### 2.2.6. Small angle X-ray scattering (SAXS)

SAXS has already been exploited previously for estimation of the size of *hPDHc* cores or sub-complexes in solution (Hiromasa, Fujisawa et al. 2004, Vijayakrishnan, Kelly et al. 2010). Similar experiments in an in-house SAXS setup (8 KeV, CuK $\alpha$  radiation at 1.54 Å, rotating anode, Pilatus 100k pixel detector) with Kratky camera setup (Kratky and Stabinger 1984, Bergmann, Orthaber et al. 2000) at the department of Nanoscale Imaging of Cellular Dynamics, Georg-August-University Göttingen (Prof. Dr. Sarah Koester) were performed to quantify the possible effect of CoA binding on *hPDHc* cores. Firstly, the purified proteins were dialyzed against their respective

buffers (refer Section 2.2.5.). Then, ~ 100  $\mu\text{L}$  of post-dialysis buffer ( $\pm$  2 mM CoA) was pipetted inside a quartz-glass capillary (Hilgenberg) and sealed using vacuum grease before mounting into the setup. Measurements were carried out for 3-4 hours in  $4 \times 10^{-1}$  mbar vacuum condition with 1 mm x 400  $\mu\text{m}$  beam size and flux of roughly  $10^{-7}$  photons per sec. A separate data set was created every 10 min. Once the measurement is done, the buffer ( $\pm$  CoA) was removed and a protein sample (800 nM full length cores or 500 nM truncated cores,  $\pm$  CoA) was added to the same capillary. It was again sealed with vacuum grease and the same measurement was carried out.

The preliminary data analysis was done with the help of Dr. Oliva Saldanha (Nanoscale Imaging of Cellular Dynamics). First, the raw data containing scattering intensities were radially integrated using a homemade MATLAB script. The data from buffer ( $\pm$  CoA) was subtracted from that of protein sample ( $\pm$  CoA). The final intensity profile was processed via PRIMUS program (ATSAS package, EMBL, Hamburg) to calculate radius of gyration ( $R_g$ ) and maximum size ( $D_{\text{max}}$ ) of the sample molecule, using pair-distance distribution function (PDDF) (Petoukhov, Franke et al. 2012). Additionally,  $R_g$  was also estimated by Guinier approximation method to validate the quality of sample (Blanchet and Svergun 2013).

## **2.2.7. Atomic force microscopy (AFM)**

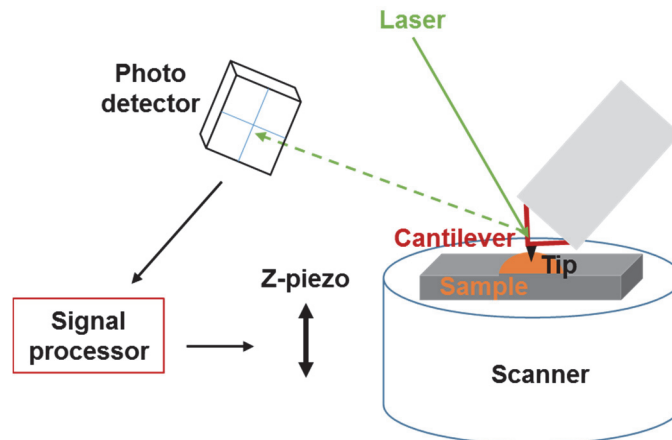
### **2.2.7.1. Background**

Although, the mechanical investigation of *h*PDHc by AFM was proposed for more than a decade (Zhou, Liao et al. 2001), a dedicated effort is still lacking. Using AFM, we attempted complement data on sizes of *h*PDHc cores from SAXS and force spectroscopic experiments to get information on their apparent stiffness. Furthermore, we tried to see changes in morphology and mechanical properties of such cores and whole complex, upon binding to CoA.

AFM, in short, is a type of scanning probe microscopy, where a fine tip of a micro cantilever is used to raster scans the sample surface (see Fig. 15). A height resolution of 1 nm is routinely achievable with this technique. Ideally, a sample (in solution or air) is settled on a plain glass surface or other materials with uniform surface like mica. The surfaces used, can be chemically modified (add metal ions or spacers) to suit the experiment. A laser beam is targeted on the back of a cantilever, which is reflected, to a center of a position sensitive photo diode, which acts as a reporter of cantilever

deflection. While scanning the surface of a sample at X-Y plane, the cantilever is bent because of the non-uniformity in heights of a surface due to sample particles. A piezo controller moves the sample surface or cantilever in Z direction so that a set distance/force between surface and the tip of the cantilever is maintained (contact mode-AFM). Since the piezo movement, which is proportional to the height of the sample surface, is recorded throughout scanning, a topographical image of a surface can be constructed (Butt, Cappella et al. 2005).

However, to probe a fragile specimen like a protein complex, contact mode-AFM is not suitable, as the tip may destroy the sample (vice-versa) during scanning. To circumvent this problem, amplitude modulation (tapping mode) is employed where a cantilever is oscillated near its resonance frequency (several KHz) using modulating piezo. When a tip comes into close proximity to sample surface, the amplitude of the oscillation decreases. Analogous to contact mode, this time, the set amplitude is restored by moving cantilever in Z direction and a topographical image is constructed. Since, the sample comes into contact of a tip during peaks of the oscillation only, the damage to the sample is limited, and this is achieved without compromising on lateral resolution (Putman, Vanderwerf et al. 1994).



**Fig. 15: Schematic of an AFM set up.**

A laser beam is focused onto a cantilever and the reflected beam is detected by a photo-diode. While the tip scans the sample surface, the cantilever bends according to the sample topography, changing the laser spot's position on the photo diode. A signal processor will process this information. It then directs Z-piezo to move sample surface, so as to keep the force between sample and tip constant. In tapping mode, cantilever oscillates at set amplitude near its resonance frequency and this amplitude is kept constant while scanning.

### 2.2.7.2. Experiment

Microscope glass cover slips were kept in Teflon rack, placed in a glass container and immersed in KOH solution (10 g KOH in 10 ml water and 200 ml ethanol). The immersed cover slips were sonicated for 5 minutes. Subsequently, the surfaces were rinsed and sonicated in deionized and filtered water three times for 5 minutes. The container was refilled with deionized water and 200  $\mu$ L of N'-(3-[Trimethoxysilyl]-propyl) diethylenetriamine (DETA) and 20  $\mu$ L acetic acid were added. After sonication for 5 minutes the surfaces were cleaned by three washing steps as described above. Finally, the cover slips were dried at 120 °C for about 1 hour and stored until usage (max. for one week) (Eghiaian and Schaap 2011). In case of mica, office tape was used to remove the top surface of the mica revealing the fresh new surface each time.

Clean cover slips were fixed on the microscopy slides using vacuum grease. 50-100  $\mu$ L protein solutions (*h*PDHc core or complex, 50-500 pM) was pipetted onto the cover slip and the cores were allowed to settle on the surface for 15 min. Finally, the microscopy slide was mounted into the AFM instrument before experiment. At 30 min interval, 50  $\mu$ L fresh buffer were added to the sample drop. BL-RC150VB cantilever (manufacturing label: spring constant 0.03 N/m and resonant frequency of 37 KHz) was used at room temperature in tapping mode for imaging. Usually, a set point of  $\sim$  150 mV was kept and scanned with a scan size of 1  $\mu$ M x 1  $\mu$ M with 256 x 256 lines and points at scan rate of 1-2 Hz. The data collected were first analyzed by WSxM 5.0 software to flatten all the images (Flatten plus tool). Then, height information at each pixel for single particle was processed with home written MATLAB script. The biggest and smallest 2.5 % heights were ignored to remove artefacts.

For stiffness measurement, a spring constant (*k*) of a cantilever was calculated from the thermal noise spectrum by fitting Lorentz function to the first peak (Hutter and Bechhoefer 1993). In a force curve generated by pressing a tip at non-deformable surface like glass, a linear fit of the extension curve represents the sensitivity (nm/V). Using Hook's law (Force,  $F = kx$ ), the deflection in volt can thus be converted to force. Now, the force vs distance curve for the sample can be obtained and also the effective spring constant for the sample.

AFM experiments and their analysis were done at the Third Institute of Physics, Georg-August-University Göttingen (Prof. Dr. Iwan Schaap), together by Dr. Mitja Platen and myself.

### 2.2.8. Electron microscopy (EM)

**Sample preparation:** In a SW 60 rotor tubes, a 10 % - 30 % linear sucrose gradient was prepared via Gradient Master 108 according to the manufacturer's manual (BioComp, Canada). For a + CoA condition, 2 mM CoA was added to both the sucrose solutions prior to gradient making. For sparse inter molecular crosslinking between nearby lysine residues of a sample protein by glutaraldehyde (GraFix), (Kastner, Fischer et al. 2008), 0.02 % glutaraldehyde was added to 30 % sucrose solution before making gradient. In case of *h*PDHc and its sub-complexes, NAD<sup>+</sup>, ThDP and MgCl<sub>2</sub> was added to a final concentration of 2 mM, 0.2 mM, and 2 mM respectively, to both protein mix and sucrose solutions before gradient making. The sucrose solutions were prepared in a buffer according to target protein (refer Section 2.2.2.3.). Next, 200 µL, 0.25 µM human PDH complex, sub-complex or cores, were loaded onto the sucrose gradient and centrifuged at 20000 - 25000 RPM, 4 °C for 16 hr. The gradient was then fractionated using a pipette (200 µL fraction size) and dot blot was performed as following:

1 µL from each fraction were pipetted to the nitrocellulose membrane (PROTRAN BA 83, Whatman GmbH) in a row and allowed to dry. It was then immersed in the Amido black stain (0.1 % w/v amido black in 25 % isopropanol, 10 % acetic acid) for 5 min. The membrane was then washed with water until the clear blots of protein drops were visible. For a homogenous sample, usually most of the proteins would migrate to middle of the gradient and thus, the fractions belonging to those region will have darker blots. The dot blot was visually inspected and the best fractions from the middle of the gradient were chosen for electron microscopy. The excess glutaraldehyde was quenched by adding 10 mM aspartate (in the same buffer as sucrose gradient).

All the electron microscopy measurements were carried out at the Department of Structural Dynamics, MPIBPC, Göttingen (Prof. Dr. Holger Stark).

#### 2.2.8.1. Negative staining electron microscopy (neg-EM)

The protein particles from a chosen fraction were then bound to a thin carbon film for 30-60 sec. Protein soaked film was transferred to an electron microscopic grid covered with a perforated carbon film. Excess liquid was removed from the grid using a soft paper. The protein-containing surface of the grid was faced towards saturated uranyl formate solution (in water) for staining. After 1 min-2 min staining, excess liquid was again removed using a soft paper and air-dried. All these steps were performed



over ice. Now, images were recorded at a magnification of 88,000-fold on a 4k x 4k CCD camera (TVIPS GmbH) using two-fold pixel binning (2,45 Å/pixel) in a Phillips CM200 FEG electron microscope (Philips/FEI) operated at 160 kV acceleration voltage. The raw images were edited using Adobe Photoshop CS6 software.

#### **2.2.8.2. Single molecule electron cryo-microscopy (Cryo-EM)**

Good looking samples (checked via neg-staining EM) was buffer exchanged to remove sucrose from the GraFix step and vitrified on continuous carbon foil on a quantifoil grid (3.5/1) in a FEI Vitrobot mark IV at a blot force of 13 and a blot time of 8.5 s.

Roughly 2250 micrographs were recorded on a Titan Krios microscope equipped with a Cs corrector at 300kV, 59000x (pixel size: 1.27 Å/px) with a dose of 41 e/Å<sup>2</sup>. Then, 212468 particles were selected using the software john Henry (custom written in the Stark lab) and an initial model was constructed using angular reconstitution method in cow eyes (custom written in Stark lab). False positively picked particles were removed from this initial model by 2D classification. Remaining good particles were subjected to 3D classification in Relion (Scheres 2012), imposing icosahedral symmetry and the best class was refined. The final map consisted of 29788 particles with a resolution of ~ 6.3 Å.

Model building was performed using Robetta (robetta.bakerlab.org). C-terminal 233 amino acid residues of hE2 and E3BP were added in the server for structure prediction. After few hours, it predicted the secondary structure and a template (PDB 3B8K). Structure prediction took couple of days, and it calculated 5 structure coordinates for each protein sequence in a PDB file format. The monomers were assembled in UCSF Chimera and molecular dynamics flexible fitting (MDFF) was performed to fit the model to the map. Real space refinement was performed for thus obtained final model via Phenix (Afonine, Grosse-Kunstleve et al. 2013).

The cryo-EM measurements, data analysis and fitting of model into the obtained EM density maps were kindly done by Dr. David Haselbach from Department of Structural Dynamics, MPIBPC, Göttingen. Visualization of EM density map was done in UCSF Chimera (Pettersen, Goddard et al. 2004) and the structural models were compared using PyMol.

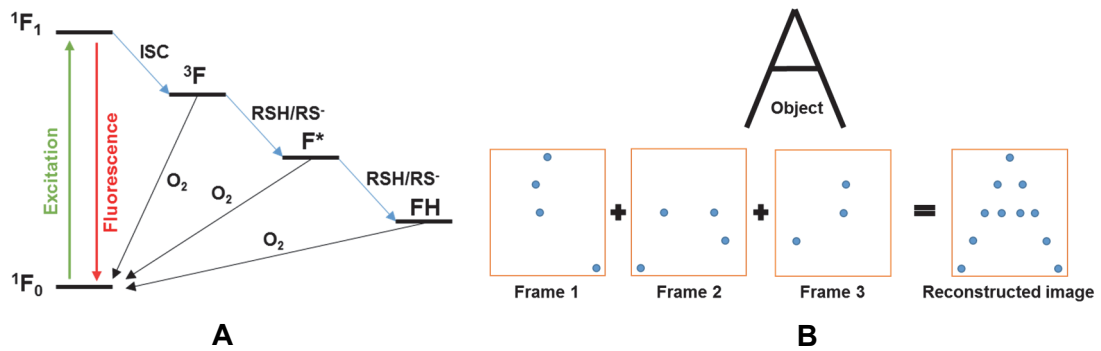
## 2.2.9. Direct stochastic optical reconstruction microscopy (dSTORM)

### 2.2.9.1. Background

Conventional optical microscopy methods are limited to a lateral resolution of roughly 200 nm as described by Ernst Abbe ( $d = \lambda / 2n \sin \theta$ ). Technically, resolution ( $d$ ) can be increased by using smaller wavelength ( $\lambda$ ) like X-ray and UV or improving the numerical aperture,  $n \sin \theta$  (using immersion oil). But light with lower wavelength show lesser contrast, while their prolong exposure causes damage to valuable sample. As most of the experiments are carried out in air or buffer/water, the numerical aperture term also is difficult to increase after some limit (1.4-1.6).

Various ingenious ways using photo physics-chemistry of fluorophores have been described in past few decades, to cross this diffraction barrier. It can be achieved in principle, by discerning the signals from fluorophores at spatial, spectral or time domains. Stochastic optical reconstruction microscopy (STORM) is one of such super resolution microscopy techniques. By separating spatially overlapping fluorophores temporally, an accurate localization of fluorophore's within diffraction-limited region is possible in order to reach super resolution (Huang, Bates et al. 2009). Photo switchable dyes, which can be activated from dark state to fluorescent state and excited to emit fluorescent signal, by light with different wavelengths, are employed in this method. An intense excitation light is used to bring all the fluorophore into stable but reversible 'off' state. Next, cycles of milder activation and excitation light will allow only a subset of fluorophores to go to 'on' state, emit fluorescent signals and go back to 'off' state, resolving their respective locations. Data from many cycles when stacked together, provides a super resolution image of a sample (van de Linde, Loschberger et al. 2011). The positions of the fluorophores can be laterally resolved up to ~ 20 nm via STORM (Rust, Bates et al. 2006).

In direct-STORM (dSTORM), a photo switchable dye is kept in an oxidizing and reducing (ROX) buffer system. When they are excited with a laser of suitable wavelength, they enter dark states. As they require oxygen to revert back to fluorescent state, ROX buffer system ensures their slow return. Hence, at a time, only few fluorophores will stochastically return to fluorescent state, whose signals could be detected and fluorophores localized (van de Linde, Loschberger et al. 2011).



**Fig. 16: STORM concept.**

(A) A photo switchable dye cycles between ground ( $^1F_0$ ) and excited ( $^1F_1$ ) singlet states by absorbing light of appropriate wavelength and emitting fluorescence photons. Fraction of the population of excited state can undergo inter system crossing (ISC) to go to triplet state ( $^3F$ ). In the presence of molecular oxygen, it can return to ground state by producing singlet oxygen. If a buffer contains oxygen scavenging system and reducing agents like 2-mercaptoethylamine (MEA), the triplet state can further hop to radical anion ( $F^*$ ) and completely reduced (FH) states. As the intermediary states ( $^3F$ ,  $F^*$  and FH) takes several seconds to return back to ground state, very few fluorophores at a given time will be in a state to give fluorescence signal. Figure adapted from (van de Linde, Loschberger et al. 2011). (B) Reconstruction of an image is possible by piling up multiple frames, each with a subset of localized fluorophores.

Mostly, such techniques have been used successfully in cell biology to observe cells organelles, cytoskeletons as well as membrane proteins (Rust, Bates et al. 2006). However, it remains a challenge to study the biophysical features of single molecule, for example protein's conformational changes. Here, we attempted to observe changes in size of eGFP-*hE2* core upon binding to coenzyme A using dSTORM in collaboration with Prof. Dr. Jörg Enderlein (III. Institute of Physics, Georg-August-Universität-Göttingen).

## 2.2.9.2. Experiment

### 2.2.9.2.1. Sample preparation

Purified eGFP-*hE2* core ( $\sim 1000$  pmoles) was mixed with anti-GFP-nanobody ( $\sim 3000$  pmoles) labeled with Alexa 647 dye (provided by Dr. Qui Van). The volume was brought to 200  $\mu\text{L}$  by adding a buffer (50 mM HEPES, 150 mM KCl, 5 mM DTE, pH 7.5). In case for + CoA condition, 2 mM CoA was added to both the protein mix and the buffer. After 30 min of incubation in a cold room, they were loaded on top of a 10 % - 30 % linear sucrose gradient (0 % - 0.02 % glutaraldehyde) in a SW 60 rotor tubes (the top contained 400  $\mu\text{L}$  cushion of 10 % sucrose without glutaraldehyde), which was prepared via Gradient Master 108 according to the manufacturer's manual

(BioComp, Canada). For a + CoA condition, 2 mM CoA was added to both 10 % and 30 % sucrose solutions before making a gradient. The ultracentrifugation was performed at 20000 RPM, 4 °C for 16 hr. The gradient was then fractionated by pipette (200 µL fractions) from the top and dot blot was performed (see 2.2.8). The best two fractions were pooled together and 10 mM aspartate (pH adjusted) was added to quench excess glutaraldehyde.

The core-dye complex (50-60 nM) was then mixed with BSA (20 % in PBS) in a v/v ratio of 1:2 to 1:5 in total volume of 20-50 µL. This mix was added to 8-well imaging chamber (Sarstedt, Nümbrecht) and the droplets were incubated for 10-15 min at room temperature to allow the complex to settle on the chamber surface. The droplet was then removed gently using pipette and the spot was rinsed with UV-bleached PBS. The chamber was finally filled with 500 µL freshly prepared imaging buffer. Preparation of imaging buffer: 50 µL enzyme stock + 400 µL glucose stock + 50 µL MEA stock (final Volume 500 µL, 100mM MEA final concentration)

#### STOCK SOLUTIONS:

A) Enzyme stock solution (50 mL) contained 100 µL Catalase, 200 µL TCEP (1 M), 25 mL Glycerine, 22.5 mL water, 1.25 mL KCl (1 M), 1 mL Tris-HCl pH 7.5 (1 M), 50 mg Glucose oxidase. Aliquots were stored at -20 °C.

B) Glucose stock solution (50 mL) contained 5 g Glucose, 45 mL water, 5 mL Glycerine. Aliquots were stored at -20 °C.

C) 2-mercaptoethylamine-HCl (MEA) stock solution (1 M)

#### **2.2.9.2.2. dSTORM measurement**

The dSTORM measurement was started at 33% laser power (continuous wave), gain 200, exposure 20 ms (~1000-5000 frames). Dr. Qui Van and Simon.C. Stein did the imaging and data analysis in the lab of Prof. Dr. Jörg Enderlein (Biophysics/complex system, III. Institute of Physics, Georg-August-Universität-Göttingen).

#### **2.2.10. Stimulated emission depletion microscopy (STED)**

STED is also a super resolution microscopy, but unlike dSTORM, high resolution is achieved by only allowing fluorophores at small space (less than diffraction limited region) to fluoresce. A pulse of excitation laser is followed by a STED pulse (doughnut

shaped hollow laser beam) to scan the sample, where all the fluorophores except those inside the inner ring of the STED laser are bleached (Klar, Jakobs et al. 2000).

Purified SNAP-*hE2* sample in PBS buffer (10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 137 NaCl, 2.7 mM KCl, pH 7.4) was labelled with silicon rhodamine (SiR) dye ( $\lambda_{\text{excitation}} = 650 \text{ nm}$ ,  $\lambda_{\text{emission}} = 670 \text{ nm}$ ). SNAP tag is basically a small protein (a mutant of the DNA repair protein O6-alkylguanine-DNA alkyltransferase) of ~ 20 KDa that specifically reacts with benzylguanine (BG) derivatives carrying a fluorophore such as SiR (Keppler, Gendreizig et al. 2003). This approach was used to label each monomer of *hE2* core and STED was performed in solution at room temperature.

Labeling and STED measurements, as well as data analysis for size estimations were performed by Dr. Jensen Nickels (Prof. Dr. Stefan Jakobs, Department of NanoBiophotonics, MPIBPC, Göttingen).

### **2.2.11. Mass spectrometry (MS)**

#### **2.2.11.1. MS for protein band identification**

The proteolytically derived truncated versions of *hE2* and *hE2/E3BP* were run in 12 % SDS-PAGE gels. They were stained and de-stained with fresh staining and detaining solutions (refer Section 2.2.2.2.) to avoid contamination. Then, they were washed with water to remove de-staining solution. The pure protein bands were cut and transferred to fresh Eppendorf tubes and provided to Dr. Oliver Valerius (Mass Spectrometry-Yeast group, Georg-August-Universität-Göttingen) who performed tryptic digestion and the mass spectrometric analysis of those protein bands (Shevchenko, Tomas et al. 2006).

#### **2.2.11.2. Cross-linking mass spectrometry (XL-MS)**

##### **2.2.11.2.1. Sample preparation:**

20  $\mu\text{L}$  *hE2/E3BP* core (~ 18 mg/mL) or 50  $\mu\text{L}$  *hPDHc* (3 mg/mL) were incubated on ice for 5-10 min with 2 mM CoA. Then, 5  $\mu\text{L}$  BS3 (5 mM) was added to these solutions and also to the control (protein without CoA) for crosslinking and incubated for 1 hour at 25 °C in a thermomixer operating at 200 RPM. Ethanol precipitation was performed for the cross-linked protein samples and the pellet was resuspended in 1% RapiGest surfactant (Waters) followed by tryptic digestion according to manufacturer's protocol

(Waters). The peptides were then dried in a speed vacuum and dissolved in 55  $\mu\text{L}$  30 % (v/v) acetonitrile (ACN) / 0.1 % (v/v) formic acid solution. Then, the cross-linked peptides were separated from non-cross-linked ones, by passing through the gel filtration column (Superdex peptide 3.2/300, GE Healthcare). The cross linked peptides were eluted via 30 % (v/v) acetonitrile (ACN) / 0.1 % (v/v) formic acid solution at a flow rate of 50  $\mu\text{L}/\text{min}$  with 50  $\mu\text{L}$  fraction size. The interesting fractions were dried using vacuum centrifugation.

#### **2.2.11.2.2. LC-MS**

The dried samples were re-dissolved in 2% (v/v) ACN / 0.05 % TFA solution and were separated by nano-flow reversed-phase liquid chromatography (Dionex UltiMate 3000 RSLC, Thermo scientific). The mobile phases used were mobile phase A: 0.1 % (v/v) formic acid and mobile phase B: 80 % (v/v) acetonitrile, 0.08% (v/v) formic acid. Then, the peptides were loaded onto a trap column (Reprosil C18, 100  $\mu\text{m}$  I.D., particle size 5  $\mu\text{m}$ ; Dr. Maisch GmbH, prepared in-house) and separated on an analytical C18 capillary column (Reprosil C18, 75 $\mu\text{m}$  I.D., particle size 1.9  $\mu\text{m}$ , 27-28 cm; Dr. Maisch GmbH, prepared in-house), with a gradient of 5-90 % (v/v) mobile phase B over 75 min at flow rate of 300 nL/min. A set up consisted direct elution of thus separated peptides into an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo scientific).

Mass spectrometry was operated in data-dependent mode with following parameters: spray voltage: 2.0 kV, capillary temperature: 275°C and normalized collision energy: 30 % at an activation of  $q = 0.25$  and an activation time of 30 ms. Survey full scan MS spectra were acquired in the orbitrap ( $m/z$  350–1550) with a resolution of 120,000 and an automatic gain control (AGC) target at 500,000. The top 20 most intense ions were selected for HCD MS/MS fragmentation in the orbitrap at an AGC target of 30,000 and with a first  $m/z$  of 110. Previously selected ions within previous 30 s were dynamically excluded for 20 s. Ions with charge states 3-8 were only selected. Internal calibration of the orbitrap was performed using the lock mass option (lock mass:  $m/z$  445.120025)(Olsen, de Godoy et al. 2005).

Raw data files were converted into 'mgf' file format using pXtract software tool (<http://pfind.ict.ac.cn/software.html>). By searching against a reduced database containing Atg18, potential cross-links were recognized with pLink search engine (Yang, Wu et al. 2012),. Search parameters were as follows: fragmentation: HCD, enzyme: Trypsin, variable modifications: oxidation (methionine) and carbamidomethylation (cysteine) and cross-linker: BS3. Finally, confirmation of

spectra of possible crosslinked di-peptides was done manually. Only those crosslinks, which have at least one good spectrum, were considered in the final output.

#### **2.2.11.2.3. Quantification of crosslinks**

Quantification of selected crosslinks was performed to find out if the frequency of a crosslink changes between presence and absence of CoA. For every ion intensity extracted for crosslink at particular state, an ion chromatogram for the same crosslink in another state was extracted and the mean value of intensities were calculated for both states.

#### **2.2.11.3. Native mass spectrometry (native-MS)**

*h*PDHc and *h*E2/E3BP cores in their respective buffers were exchanged against 200 mM ammonium acetate (pH 7) using Micro Bio-spin 6 columns (Bio Rad). Spectra were acquired on a G1 mass spectrometer (Waters) modified especially for high masses (Sobott, Hernandez et al. 2002) using in-house prepared gold-coated glass capillaries (Hernandez and Robinson 2007). Parameters used for data acquisitions are capillary voltage 1.7 kV, cone voltage 150 V, extractor 40 V, trap collision energy 20 V, transfer collision energy 15 V, source backing pressure 7–10 mbar and trap gas 0.5 mL/min. Thus obtained spectra were processed and complexes were assigned using MassLynx (Waters) and Massign software (Morgner and Robinson 2012).

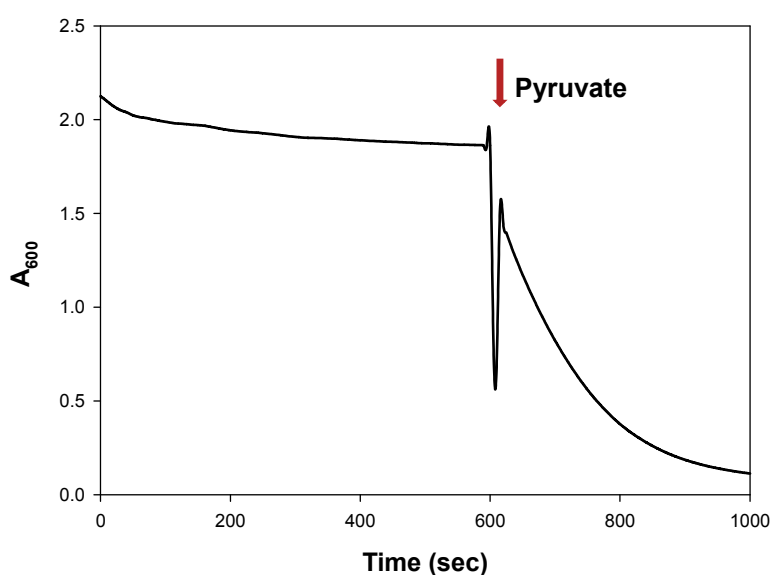
Cross-linking and native mass spectrometry experiments and data analysis were done by Dr. Carla Schimdt (HALOmem – membrane proteins and dynamics, Martin Luther University Halle-Wittenberg) in the lab of Prof. Carol Robinson, University of Oxford and Prof. Henning Urlaub, MPIBPC.

### 3. Results

All the results from molecular biology, protein expressions and protein purifications are in Appendix A and B, if not described in this section.

#### 3.1. Quality check for *hE1*, *hE2/E3BP* and *hE3*

DCCPIP assay was performed for *hE1* protein to check if it was as active as described in the literature. In general, the activity of *hE1* depended on the batch of over expression and purity. A protein band (near 25 KDa) sometimes contaminated the protein (see Fig. 52), whose identity has been established to be *E. coli* FKBP20 by mass spectrometry (PhD thesis of Franziska Seifert). The most active protein batches had specific activity of 70-90 mU/mg which is slightly more than reported value of 40-60 mU/mg (Seifert, Golbik et al. 2006) while other batches had specific activity in the reported range or lower. The protein batches with lower specific activity were discarded.

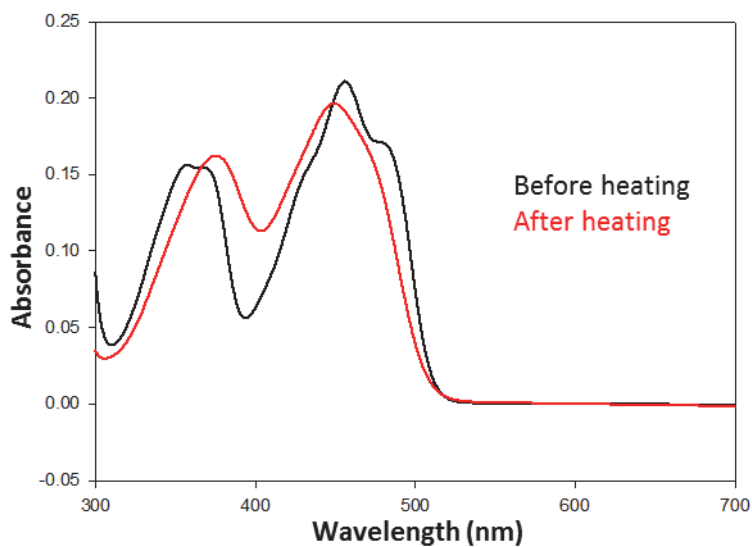


**Fig. 17: A representative progress curve of DCPIP assay.**

*hE1* ( $C_{\text{end}} = 0.4$  mg/ml) was added to a reaction mixture containing 200  $\mu\text{L}$  of 5 x buffer (0.5 M  $\text{KH}_2\text{PO}_4$ , 1.5 M KCl, pH 7.6), 20  $\mu\text{L}$  of 0.01 M ThDP + 0.1 M  $\text{MgSO}_4$  mix and required volume of water to reach 890  $\mu\text{L}$ , in a 1 mL cuvette (path length of 1 cm). It was incubated at room temperature for 5 min, then 10  $\mu\text{L}$  DCPIP was added. Time course measurement was started immediately at 600 nm at 30  $^\circ\text{C}$ . After 600 sec, 100  $\mu\text{L}$  of 0.1 M pyruvate was added, mixed rapidly with pipette and depletion of absorbance measured for 400 sec. A slope from 620-650 sec was used to calculate the specific activity. (Details in section 2.2.3.1.)



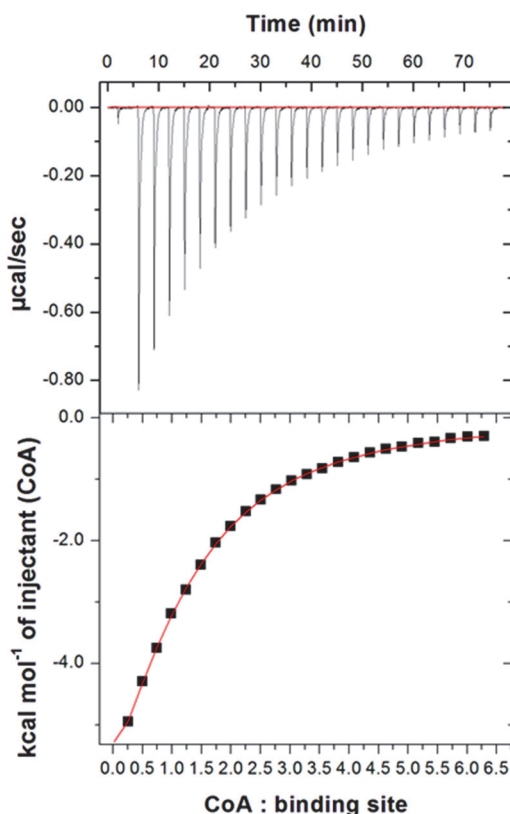
For *hE3* protein, a fraction of active site bound to FAD cofactor was calculated. By heat denaturing the protein, FAD was released to the solution. The concentration of released FAD can be calculated, as the molar extinction coefficient of free FAD is known ( $\epsilon_{450} = 11300 \text{ M}^{-1}\text{cm}^{-1}$ ). In case of 100 % saturation, the concentration of free FAD must be twice of the concentration of used *hE3* (homodimer), as each monomer can bind to one FAD. The purified protein batches generally showed FAD saturation of > 80 %, which were deemed good enough for further usage.



**Fig. 18: FAD saturation test of *hE3*.**

A spectrum of 1 mL *hE3* (1 mg/mL) was taken from 300-700 nm at room temperature (black). The protein was transferred to an Eppendorf tube and heated at 95 °C for 10 min. Denatured protein was removed via centrifugation and a spectra of clear supernatant was taken (Red). The absorbance at 450 nm for the post-heating spectra was noted and used to calculate concentration of free FAD and subsequently, total FAD saturation level as mentioned in section 2.2.3.3.

For *hE2/E3BP* core, it was checked if it could bind to CoA. An ITC-binding experiment was performed according to section 2.2.5, where by heat uptake/release upon CoA binding could be used to verify the binding of CoA to the core.



**Fig. 19: An ITC-binding experiment between CoA and *hE2/E3BP* core.**

An ITC data was obtained after titration of 2 mM CoA to  $\sim 1.5 \mu\text{M}$  *hE2/E3BP* core in 50 mM HEPES, 150 mM KCl, pH 7.5 at 30 °C with stirring speed of 500 RPM. The upper panel shows the heat of binding for each injection. Each peak is integrated, concentration normalized and the heat of dilution from control experiment is subtracted to give a plot in a bottom panel, which is fitted according to 'one set of sites' model fitting in MicroCal Analysis software (refer Section 2.2.5.). Three independent experiments were performed from two batches of protein purification.

The binding of CoA to the core subunits was exothermic in nature. The fitting curve was not sigmoidal, probably, due to high dissociation constant ( $K_D$ ) of 88.5  $\mu\text{M}$  (SE = 14.12  $\mu\text{M}$ ). The stoichiometry (N) between CoA and potential binding sites reached 0.989 (SE = 0.08) if only *hE2* was considered to be able to bind CoA. When E3BP was also considered to bind CoA, it decreased to 0.66 (SE = 0.05). The concentration of binding sites was calculated, considering each core consists of 40 *hE2* and 20 E3BP subunits (refer Section 3.5.) Since both *hE2* and E3BP are considered to contain one putative CoA binding site at their respective core forming domains, the data on hand suggests E3BP might not bind CoA.

However, the basis for E3BP not binding CoA needs to be explored further. In the crystal structure of E2 from *A. vinelandii* PDHc, many residues are involved in interactions with CoA as it is a big molecule with four important structural parts

(adenine ring, Ribose-5'phosphate, pyrophosphate and pantetheine arm). When the sequence of E2 from *A. vinelandii* PDHc is compared with human E2 and E3BP, two residues Ala534 and Ala585 important for binding to CoA stand out, as they are similar in *hE2* (Val542, Ser593) but substituted by bulkier residues in E3BP (Lys394, Phe445). These residues could be an important discriminant in CoA binding ability between E2 and E3BP. Near C-terminus, further three residues show such substitutions, but they are only involved with IN conformation of CoA (Mattevi, Obmolova et al. 1993), when the pantetheine arm is in extended form and thus might be important during catalysis only (eg: catalytic histidine) .

```

AvE2 432 - VPHVTQFESADITELEAFRVAQKAVAQKAGVKLTVLPILLKACAYLLKELPDFNSSLAPS
hE2 442 - IPHYLSIDVNMGEVLLVRKELNK-ILEGRSKISVNDFIIKASALACLKVPEANSSWMDT
E3BP 297 - VPAYATADCDLGAVLKVRQDL----VKDDIKVSVNDFIIKAAAVTLKQMPDVNVSWDGE

AvE2      GQALIRKKYVHIGFAVDTPDGLLVPIRVNDQKSLIQLAEEAELAEKARSKKLGADAMQ
hE2      V--IRQNHVVDVSVAVSTPAGLITPIVFNNAHIKGVETIANDVVSLATKAREGKLPHEFQ
E3BP     G--PKQLPFIDISVAVATDKGLLTPIIKDAAAKGIQEIADSVKALSKKARDGKLLPEEYQ

AvE2      GACFTISSLGHIGGTAFTPIVNAPEVAAILGVSKAS--MQPVWDG--KAFQPRMLPLSL
hE2      GGFTISNLGMFGIKNFSAIINPPQACILAIGASEDKLV PADNEKG---FDVASMMSVTL
E3BP     GGSFSISNLGMFGIDEFTAVINPPQACILAVGGRFRPVLKLTDEEENAKLQQRQLITVTM

AvE2      SYDHRVINGAAAAAFTKRLGDLLADIRAILL - 638
hE2      SCDHRVVDGAVGAQWLAEFRKYLEKPI TMLL - 547
E3BP     SSDSRVVDDELATRFLKSFKANLENPIRLA - 501

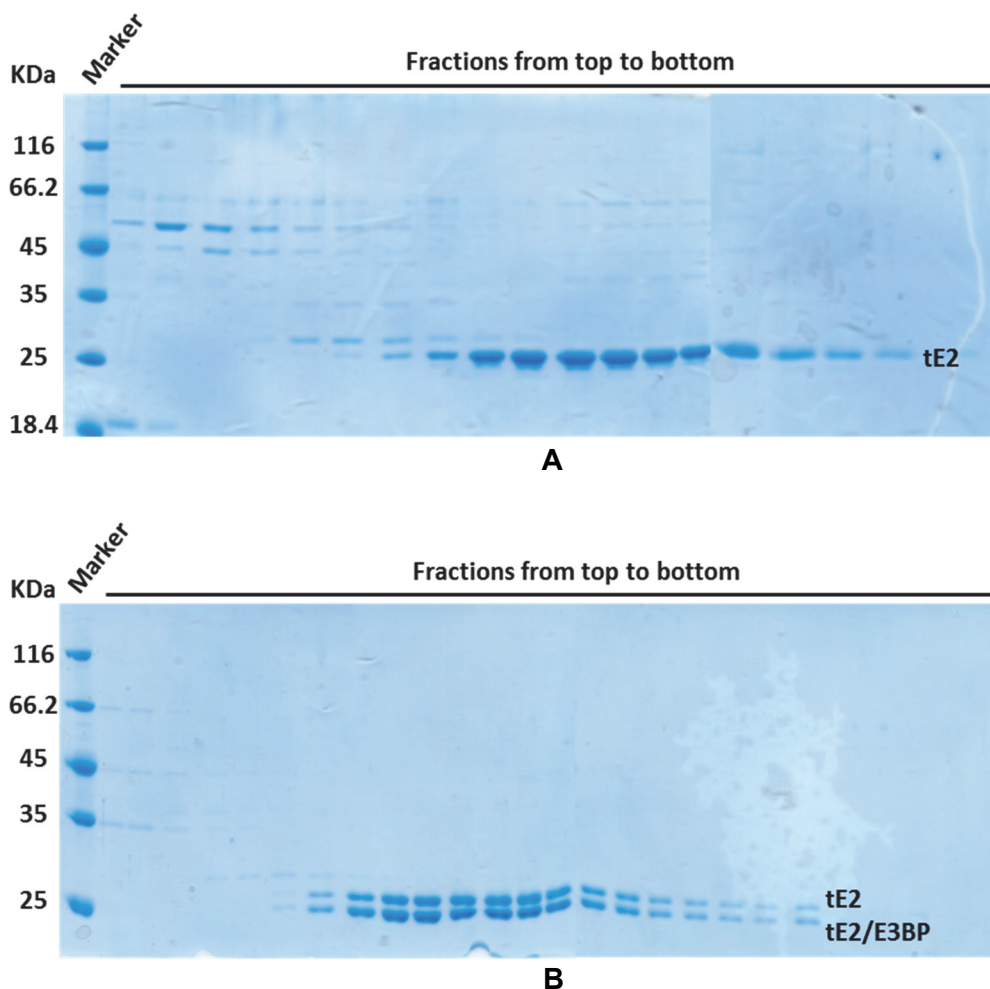
```

**Fig. 20: Sequence alignment of PDHc E2 of *Azotobacter vinelandii* (AvE2) with hE2 and E3BP.**

All the amino acyl residues involved in interaction with CoA (IN and OUT conformations) are colored, based on the crystal structure of AvE2 complexed with CoA and dihydrolipoamide (Mattevi, Obmolova et al. 1993) Among those residues, the ones which are similar between the two E2s but significantly different in E3BP are shown in green. Near C-terminus, further three residues show such substitutions (shown in blue). Sequence alignment was performed with Clustal Omega.

### 3.2. Proteolysis of hE2 and hE2/E3BP core

By incubating *hE2* and *hE2/E3BP* cores with thermolysin protease and then separating the proteolysed fraction from the stable portion of the protein using sucrose gradient ultra-centrifugation, it was possible to isolate intact truncated cores.



**Fig. 21: Proteolytic treatment of *hPDHc* cores with thermolysin.**

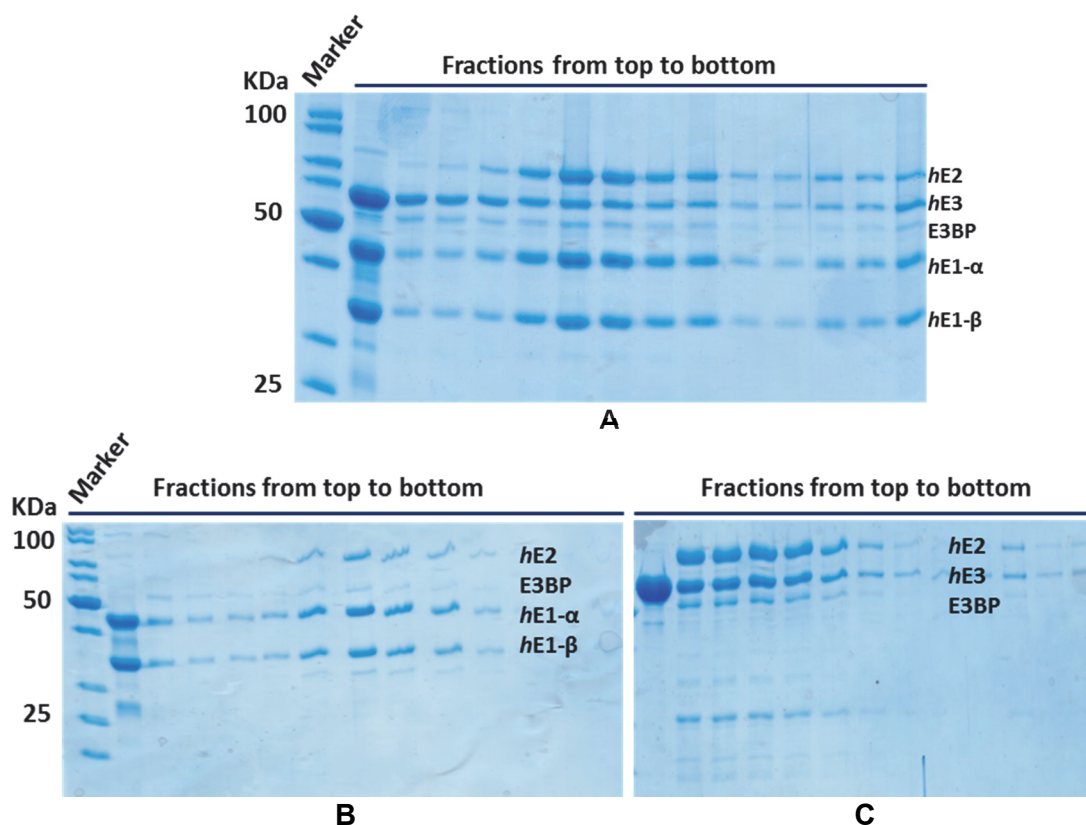
Purified *hE2* (A) and *hE2/E3BP* (B) were incubated with thermolysin in a mass ratio of 400:1 and 100:1, respectively, at 20 °C for 60 min in a thermomixer with shaking at 700 RPM. After adding 5 mM EDTA, they were loaded in 10 % - 30 % linear sucrose gradient and centrifuged at 25000 RPM, 4 °C for 16 hr. The fractions were collected from the top to bottom manually and SDS-PAGE was performed. (Details in section 2.2.2.4.)

Single band for truncated *hE2* and two bands for *tE2/E3BP*, all close to 25 KDa, were visible for fractions belonging to the middle of the sucrose gradient. A mass spectrometric analysis (refer Section 2.2.11.1.) was carried out for these protein bands, which identified them to belong to the core forming inner catalytic domains of *hE2* and E3BP (see Appendix E). The apparent molecular weight of ~ 25 KDa also suggested that all the flexible outer parts (linkers + SBD + lipoyl domains) were proteolytically cleaved leaving inner core intact. However, only ~ 1/3<sup>rd</sup> of the cores were recovered. Additionally, brief centrifugation was required every day to remove precipitates, suggesting these cores were unstable. Hence, the protein yield from this

process was only suitable for those experiments that require tiny amount of protein like electron microscopy or atomic force microscopy.

### 3.3. In-vitro reconstitution of *h*PDHc and sub-complexes

The purified components of *h*PDHc (*h*E2/E3BP core, *h*E1 and *h*E3) were mixed and 10 % - 30 % linear sucrose gradient ultracentrifugation was performed (refer Section 2.2.2.5.). Most of the reconstituted complex migrated to middle of the gradient and could be isolated from excess *h*E1 and *h*E3 proteins (see Fig. 22). Only fresh samples (< 5 days old) were used for further experiments, since they precipitated with time. At least half of the used core concentration could be recovered as whole complex and sub-complexes assuming all the binding sites in *h*E2 and E3BP are bound to *h*E1 and *h*E3.



**Fig. 22: In-vitro reconstitution of the *h*PDHc and sub-complexes from their component subunits.**

To *h*E2/E3BP core, *h*E1 or *h*E3 or both were added in 3-5 times molar excess of their respective binding sites, in reconstitution conditions mentioned in section 2.2.2.5. After sucrose gradient ultracentrifugation, whole complex (A), *h*E1-subcomplex (B) and *h*E3-subcomplex (C) were isolated from purer fractions (near middle of the gradient) which were judged by 12 % SDS-PAGE analysis.

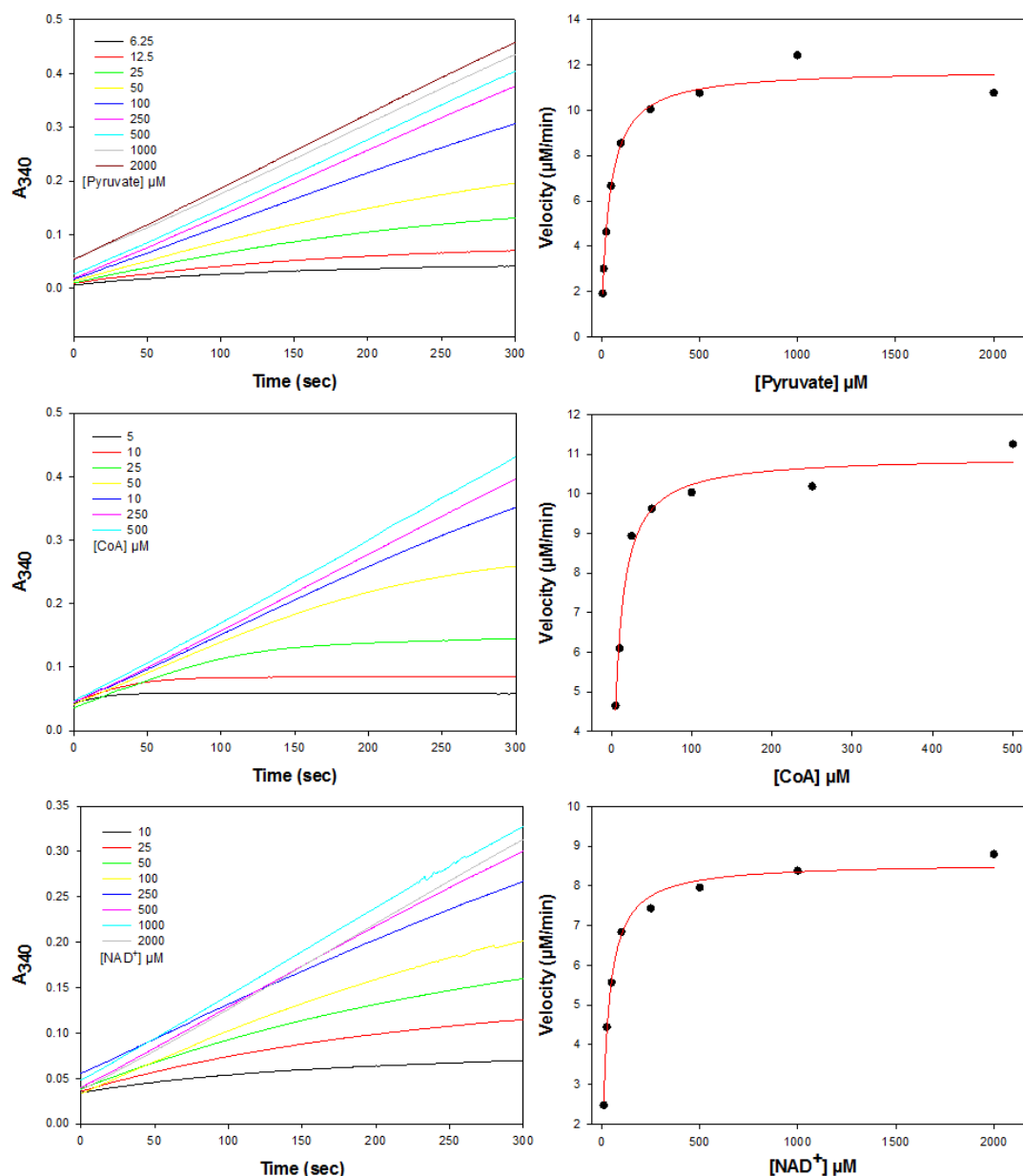
### 3.4. Steady state kinetics of *h*PDHc

A steady state kinetics study of in-vitro reconstituted *h*PDHc was performed with an assay explained in section 2.2.3.2. Varying concentrations of its substrates (pyruvate, CoA and NAD<sup>+</sup>) were used and reaction velocities ( $\mu\text{M}/\text{min}$ ) were calculated from a slope (0-25 sec) of absorbance vs time plots. A hyperbolic curve (acc. to Michaelis-Menten equation) was fitted to the plot of reaction velocities to substrate concentrations, in order to estimate  $V_{\text{max}}$  and  $K_M$ , which are listed in Table 4. In general, the  $K_M$  values were comparable to those mentioned for PDHc from eukaryotic sources (Lazo and Sols 1980, Kresze and Ronft 1981). But  $A_{\text{spec}}$  (5.41 U/mg) was lower than reported for *h*PDHc i.e. 29 U/mg of *h*E1 (Korotchikina and Patel 2001). If we assume that all the binding sites of *h*E1 are occupied in the reconstituted *h*PDHc, the  $A_{\text{spec}}$  will be  $\sim 10$  U/mg of *h*E1 (i.e. one third of the reported value).

Substrate	$K_M$ ( $\mu\text{M}$ )	$V_{\text{max}}$ ( $\mu\text{M}/\text{min}$ )	Max. $A_{\text{spec}}$ (U/mg)
Pyruvate	40.9 +/- 10.36	11.4 +/- 0.75	
CoA	7.02 +/- 3.5	10.96 +/- 2.45	$\sim 5.41$ +/- 0.55
NAD <sup>+</sup>	28.8 +/- 6.27	8.1 +/- 0.6	

**Table 4: Steady state kinetics parameters of reconstituted *h*PDHc with associated standard errors.**

The affinity of substrates to *h*PDHc ( $K_M$ ) and the maximum velocity ( $V_{\text{max}}$ ) were calculated by fitting a hyperbolic function on the plot between substrates concentration and velocity according to Michaelis-Menten equation. The *h*PDHc lost activity gradually (experiments performed in different days with same stock showed different  $V_{\text{max}}$ ).

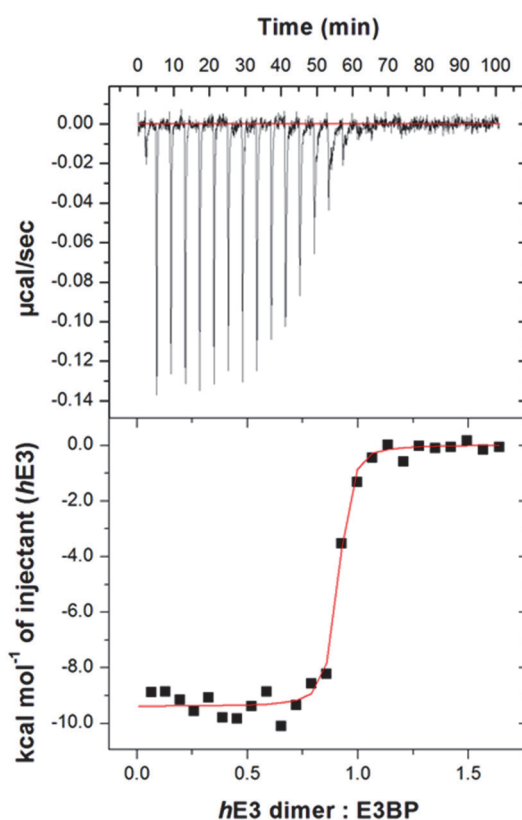


**Fig. 23: PDHc reaction progress curves (left) and their respective Michaelis-Menten plots (right).**

In a standard assay, 1 mL reaction mixture consisted of 50 mM  $\text{KH}_2\text{PO}_4$ , pH 7.5, 2 mM  $\text{MgCl}_2$ , 0.2 mM ThDP, 4 mM cysteine, 156  $\mu\text{M}$  CoA, 2 mM pyruvate, 2 mM  $\text{NAD}^+$  and 1.875  $\mu\text{g}$  *h*PDHc. For studying substrate dependence, assays were performed at 37 °C with varying pyruvate, CoA or  $\text{NAD}^+$  concentrations, keeping concentrations of all the constituents of standard assay fixed. The progress of the reaction was monitored spectrophotometrically, with NADH production corresponding to increase in absorbance at 340 nm ( $A_{340}$ ). By fitting a hyperbolic function (SigmaPlot 11.0 software) to reaction velocity vs substrate plots, steady state parameters like  $V_{\text{max}}$  and  $K_{\text{M}}$  were calculated.

### 3.5. ITC-binding experiment of *hE2/E3BP* core with *hE3*

In an ITC-binding experiment, it could be confirmed that nearly 20 *hE3* proteins bind to a *hE2/E3BP* core at full saturation. Two models of different stoichiometry between *hE2* and E3BP which are in a debate, were tested (refer Section 1.5.). For 40:20 model (40 *hE2* and 20 E3BP), stoichiometry of 1.04 (SE = 0.05) was estimated between *hE3* dimer and E3BP. Whereas, for 48:12 model, stoichiometry of 1.89 (SE = 0.05) was estimated, which is highly unlikely as one E3BP can only bind to one *hE3* dimer (at max.). This means, a *hE2/E3BP* core consists of nearly 20 E3BP subunits, and possibly forms 20 trimers with *hE2* in a core with 1:2 stoichiometry if distributed homogenously. The binding of *hE3* to the core was too tight ( $K_D = 4.5$  nM, SE = 0.95) for accurate estimation.



**Fig. 24: An ITC-binding experiment between *hE3* and *hE2/E3BP* core.**

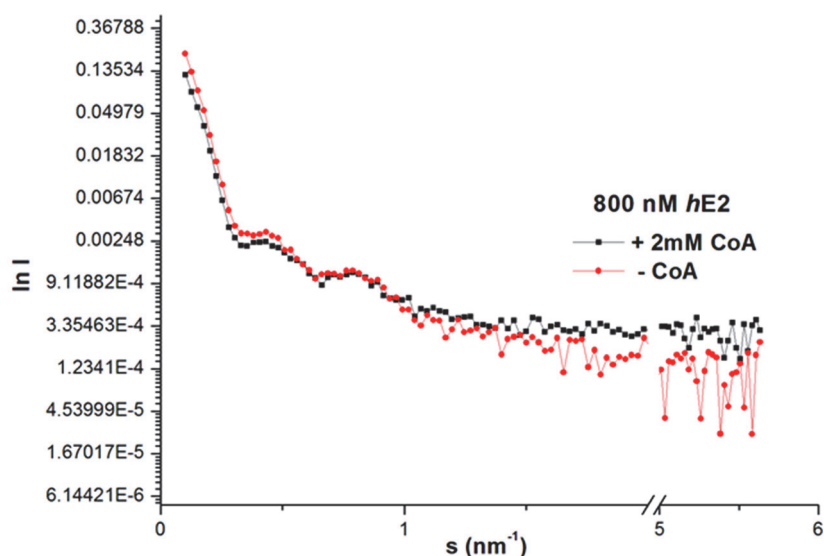
ITC data were obtained after titration of 140-210  $\mu\text{M}$  *hE3* to  $\sim 0.5$ -1.5  $\mu\text{M}$  *hE2/E3BP* core in 50 mM HEPES, 150 mM KCl, pH 7.5 at 25 °C with stirring speed of 500 RPM. The upper panel shows the heat of binding for each injection. Each peak is integrated, concentration normalized and the heat of dilution from control experiment is subtracted to give the plot in a bottom panel, which is fitted according to 'one set of sites' model fitting in MicroCal Analysis software (Details in 2.2.5.). The concentration of E3BP was calculated considering 40:20 *hE2*:E3BP model (in this figure) and 48:12 model. Five independent experiments were performed from three batches of protein purification.



### 3.6. Size estimation of *h*PDHc core and its variants by SAXS

Previously we observed improved *h*PDHc core sample quality in the presence of CoA (MSc thesis of Sabin Prajapati) during EM study. Thus we performed SAXS with *h*PDHc core variants to see if CoA binding translates into any large conformational changes of the cores. Such changes would be then quantified as a change in  $R_g$  and  $D_{max}$  values.

These size estimates obtained for *h*PDHc cores and its variants are similar to the reported values (Hiromasa, Fujisawa et al. 2004, Vijayakrishnan, Kelly et al. 2010). In all the measurements, these parameters decreased upon CoA binding ( $\Delta D_{max} \sim 4\text{-}5$  nm), which suggest the significant compaction of *h*PDHc cores (see Table 5). Interestingly, the change in size is similar to the difference in size between smallest and largest yeast PDHc cores during breathing process (Zhou, Liao et al. 2001). However, the compaction also might be the result of increased lipoyl arm internalization (refer Section 1.4.1.d.).



**Fig. 25: A typical SAXS curve of scattering intensity ( $I$ ) decay over momentum transfer ( $s$ ).**

In general, an addition of CoA to the core, made the data smoother (at higher  $s$ ). For a big particle like *h*PDHc cores, very few points could be recorded at Guinear region (low  $s$ ), hence, pair distance distribution function was used to calculate both  $R_g$  and  $D_{max}$ . The protein concentrations were chosen according to similar experiments published before (Vijayakrishnan, Kelly et al. 2010) and CoA concentration was chosen to saturate the core (approx.  $> 10 \times K_D$ , refer Section 3.1).

Sample (N)	- CoA		+ 2 mM CoA		$\Delta D_{\max}$ (nm)
	$R_g$ (nm)	$D_{\max}$ (nm)	$R_g$ (nm)	$D_{\max}$ (nm)	
<i>hE2/E3BP</i> core (3)	12.8	41.5	11.9	36.7	4.8
<i>hE2</i> core (1)	13.1	42	12.8	38	4
<i>tE2</i> core (1)	12.3	30.5	10.6	26	4.5
SNAP- <i>hE2</i> (1)	12.3	43	12.2	38.5	4.5

**Table 5: Radius of gyration ( $R_g$ ) and maximum diameter ( $D_{\max}$ ) estimates for human PDH complex cores and its variants.**

The measurements were done in the presence and absence of 2 mM CoA. Upon binding to CoA, all the cores showed marked reduction in both  $R_g$  and  $D_{\max}$ . As a quantification of size reduction, difference in  $D_{\max}$  ( $\Delta D_{\max}$ ) is used.

### 3.7. Size and stiffness estimation of *hPDHc* core and its variants by AFM

The sizes estimated for all *hPDHc* and cores by AFM, were significantly lesser (~ 40 %) than expected. We assume that the PDHc is inherently soft and spreads over solid surfaces of glass and mica used in AFM. Even at the low scanning forces, these particles showed same tendency, which affected our attempts in understanding the effect of CoA binding to the size of these particles. The apparent spring constant which is a measure of stiffness were smaller (5-10 times) for all the cores and complexes studied here (see Table 7) than normal globular proteins (Zaccai 2000). PDHc cores have been predicted to be very soft with apparent spring constant of 0.015 N/m based on its breathing motion (Kong, Ming et al. 2003). This explains why they all tend to collapse on the solid surface.

The heights and spring constant of *h*PDHc and core variants are listed in following tables:

Sample	Avg. height (nm)	Standard error	Total number (N)
<i>h</i> E2 core	28.1	0.33	215
<i>h</i> E2/ <i>E</i> 3BP core	24.8	0.3	206
tE2 core	22.6	0.17	70
SNAP- <i>h</i> E2 core	19.2	0.1	556
eGFP- <i>h</i> E2 core	25.7	0.21	80
<i>h</i> PDHc (GraFix)	29.6	0.88	37

**Table 6: Heights of human PDH complex and different core variants.**

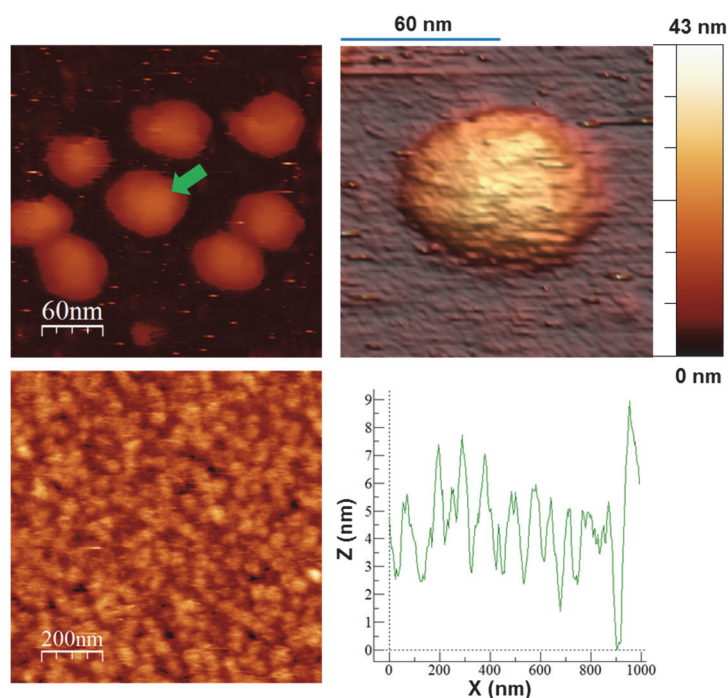
The sizes of the cores all deviate by ~ 40 % if compared to values obtained by SAXS. The truncated cores resisted this tendency but still appear to be ~ 25 % smaller than the expected size in solution. Addition of *h*E1 and *h*E3 proteins to form reconstituted *h*DPHc did not seem to stop this, as they are also significantly smaller than sizes estimated by electron microscopy (see Fig. 35).

Sample	App. spring constant (N/m)	Standard error	Total number (N)
tE2 core	0.023	0.004	6
tE2 core + CoA	0.032	0.002	15
KD- <i>h</i> E2 core	0.035	0.014	6
KD- <i>h</i> E2 core + CoA	0.012	0.001	11
<i>h</i> PDHc (GraFix)	0.072	0.02	12

**Table 7: Stiffness of human PDH complex and different core variants**

As the *h*PDHc and its cores drifted a lot while doing force spectroscopy, only few reliable measurements could be recorded (reason for low N). Force curves from only those particles that did not drift were taken into account. A scanning was performed after taking each force curve to see if the particle probed was still there in the previous location.

While scanning the particles, interesting areas were zoomed in, while changing scanning rate and number of lines and points per scan. However, the intricate surface details of the complex and core (For example: pentagonal holes of the core or lipoyl arms) were not observed consistently (see Appendix D).



**Fig. 26: An example of AFM scans of *hPDHc*.**

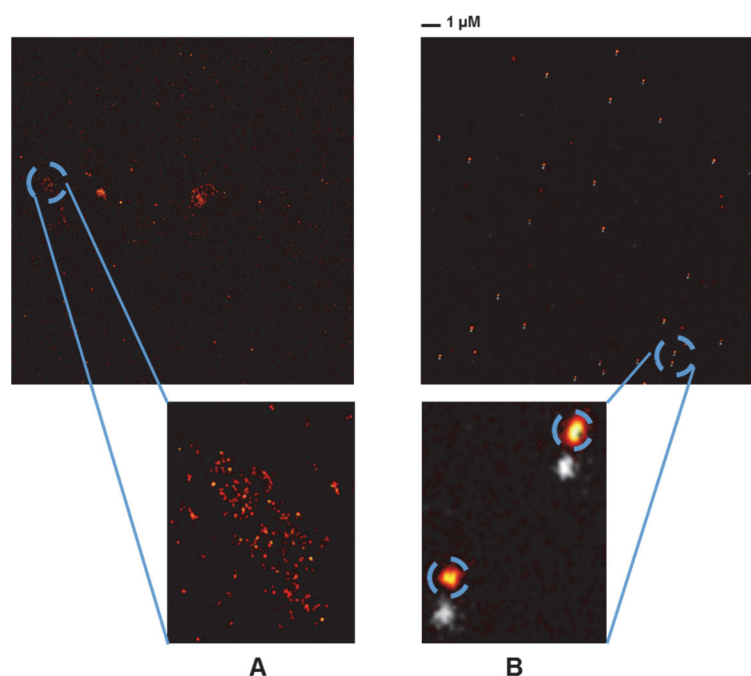
The *hPDHc* sample that underwent GraFix step, could be scanned as homogeneously distributed particles (Top row). After raster scanning the protein settled in glass/mica surface (top-left), a good-looking area was focused (green arrow). The scan rate was usually decreased for scanning selected area (top-right). However, without this step, the complex seemed to fall apart into its constituent subunits, with smaller *hE1* and *hE3* (5 nm), occupying majority of the surface (bottom row).

### 3.8. dSTORM for single molecule analysis

Although the maximum resolution reported for super resolution fluorescence microscopy is only in a range of tens of nanometers, we attempted to observe CoA induced size shrinkage of the core (~ 4-5 nm). The idea was to collect dSTORM images of many homogeneously distributed eGFP-E2 cores which were complexed with anti-GFP-nanobody\*Alexa 647 dye conjugate. During sample preparation optimization, it was very important to remove as much unbound nanobody-dye conjugate as possible to reduce the background.

Our first strategy to achieve this was to use Spin-X<sup>®</sup> UF concentrator (MWCO 50 KDa), so that any unbound nanobodies are filtered out. However, this did not prove to be successful approach as the background from unbound nanobody-dye conjugate could not be reduced by much and also, the complex appeared to aggregate (see Fig. 27a A). Second strategy was to perform linear sucrose gradient ultracentrifugation so that the whole complex migrates to the middle of the gradient leaving excess unbound nanobody-dye conjugates at the top of the gradient. BSA was added to the complex

to minimize local aggregation of complex on the surface while imaging. This was partially successful because homogeneously distributed complex could be observed with significantly reduced background. But the signal from the Alexa 647 dye was very weak in comparison to eGFP, which could be the result of separation of nanobody-dye conjugate from eGFP-*hE2* core during ultracentrifugation. In a third round of optimization, only parameter that was kept different from second strategy was to use GraFix (refer Section 2.2.9.2.). This method of fixing protein complexes in a gradient of chemical cross-linkers while migrating along a sucrose gradient is used in electron microscopy field. It was useful while doing AFM as well (refer Section 3.7.). This time, we could observe homogeneously distributed complex particles, minimum background and good signal from Alexa 647 dye (see Fig. 27a B).

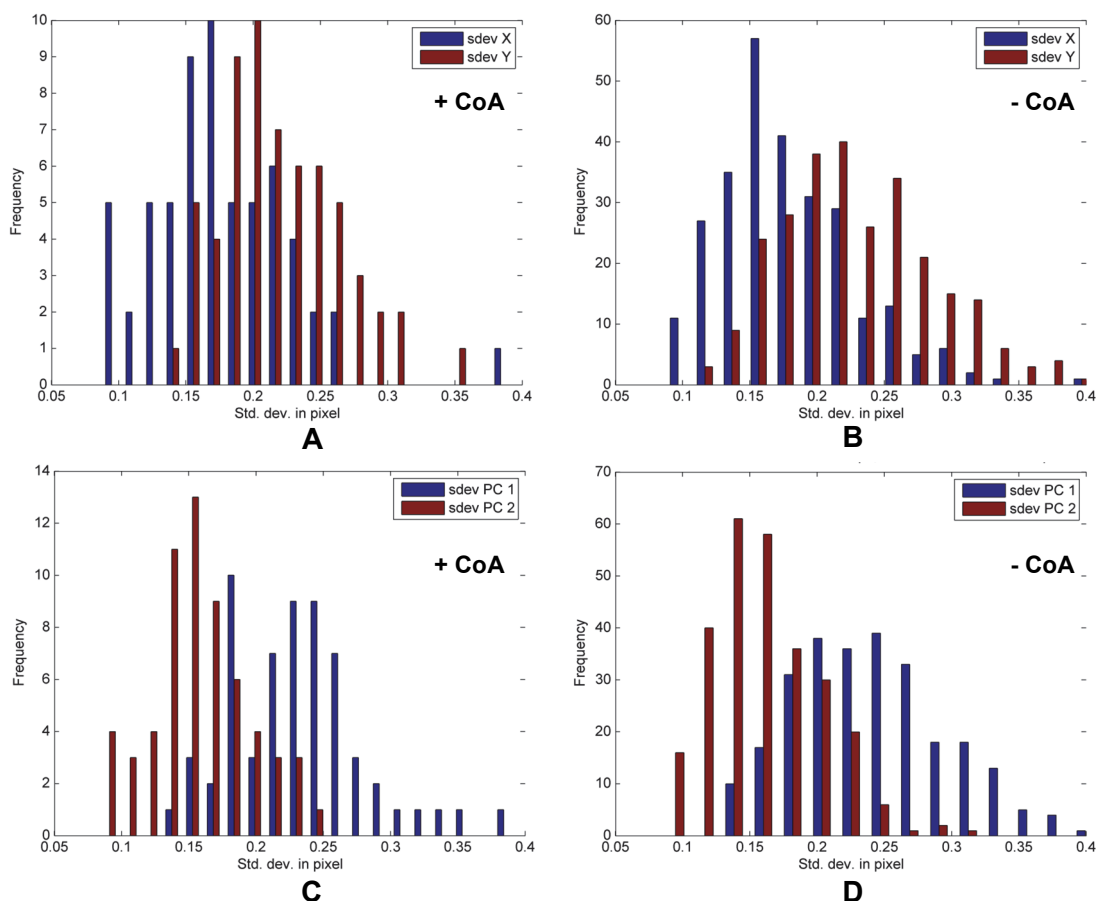


**Fig. 27a: dSTORM images of the eGFP-*hE2* core bound to Alexa647 coupled anti-GFP nanobody.**

A) The first strategy to use centrifugation-based concentrator to remove excess dye conjugated nanobody did not work, as evident by fluorescence signals from large clusters close to each other signaling protein aggregation. B) Post GraFix and adding BSA to reduce unspecific clustering at the coverslip surface, individual cores can be homogeneously separated (grey - eGFP signal, Red – Alexa 647 signal; signals are not overlapping due to chromatic aberration).

Next, the size variation was checked along X and Y-axes, in which we observed the particles were larger at Y-axis, suggesting the cores were not radially symmetrical (Fig. 27b A and B). Following AFM data, it is most likely that the cores spread over the cover slip surface changing its geometry from spherical to ellipsoidal. Another major contributor for this might be drift during measurement. Secondly, similar analysis was also performed at principle component 1 (longer

axis) and principle component 2 (shorter axis) of the particles (Fig. 27a C and D). Here, the sizes of the cores at PC1 was slightly smaller for + CoA sample, but, at the given distributions and asymmetry of the cores, the observed effect of CoA binding was inconclusive. Possibly, the GraFix step improved the integrity of the eGFP-*hE2* core, but at the cost of dynamics involved in CoA binding.

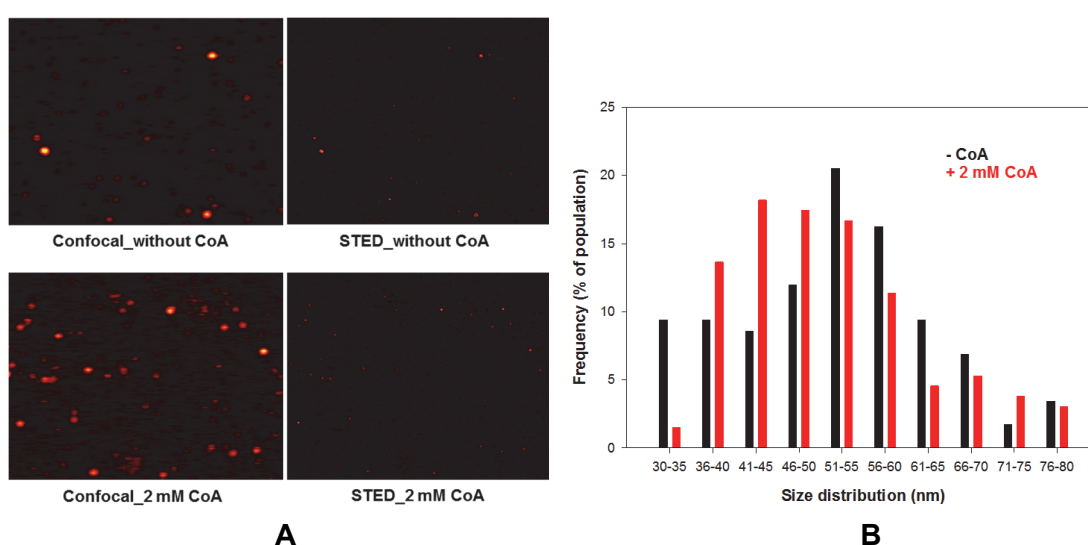


**Fig. 27b: Variation in sizes of the eGFP-*hE2* core complexed with Alexa 647 coupled anti-GFP nanobody.**

A & B) The size variation in standard deviation in pixel is plotted against the number of particles (frequency) in presence and absence of 2 mM CoA conditions at X and Y axes, respectively. The median of standard deviations in X-axis was 0.211 pixels for both conditions. However, in Y-axis it was higher with 0.211 and 0.224 pixels for + and – CoA states, respectively. C & D) The size variation in standard deviation in pixel is plotted against the number of particles (frequency) in presence and absence of 2 mM CoA conditions at PC1 and PC2, respectively. In PC1, the median of standard deviations were 0.232 and 0.238 pixels for + and – CoA states, respectively. While in PC2, they had similar values of 0.155 and 0.156 pixels. Thus, the particles look slightly smaller in + CoA state but due to asymmetry and a broad variation in sizes, it is not conclusive. (The figures and the statistical data were kindly provided by Simon Stein, III. Institute of Physics, Georg-August-Universität-Göttingen)

### 3.9. Estimation of size variation of *h*PDHc core by STED

Another method that was used to observe size variation in *h*PDHc cores was STED microscopy. We used N-terminally SNAP tagged *h*E2 construct for this experiment. Silicon Rhodamine (SiR) dye was self-incorporated into SNAP and the fluorescent signals could be detected near 660 nm (Excitation at 645 nm). The lateral size variation of the particles was large, although majority lie between 36-65 nm. Large sizes (80-100 nm) were possibly due to inter core oligomerization and thus omitted from the final analysis. Overall, the size distribution profile shifts towards smaller size groups upon addition of 2 mM CoA (see Fig. 28).



**Fig. 28: Images obtained by scanning same region by confocal microscope and STED microscope and the size variation upon CoA binding.**

A) In an images obtained by STED microscopy, SiR coupled SNAP tagged *h*E2 cores appear smaller than in images obtained by normal confocal microscopy as the attainable resolution is better in STED. B) In a size distribution profile of the cores calculated from the STED images, upon binding to CoA, the cores appear to be getting smaller. (N = 117 for – CoA and 132 for + CoA condition)

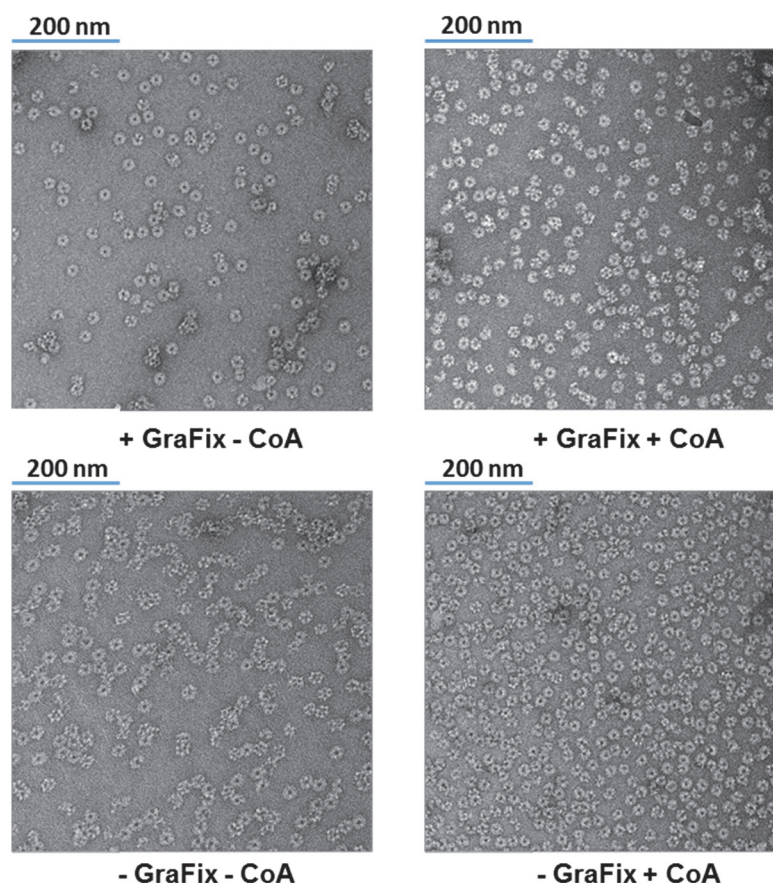
A considerable population of SNAP-*h*E2 cores at size range beyond 43 nm ( $D_{max}$  calculated via SAXS) indicates the particles collapsed on glass coverslip surface during STED microscopy. This effect was also problematic during AFM and dSTORM. The increased overall apparent sizes of the cores might be due to increase in size at lateral position (XY plane) with complementary decrease in height (Z axis). Finally, while these data do complement SAXS in demonstrating the CoA induced PDHc core compaction, due to collapsing nature of these cores on solid surface; the effect of CoA

binding shown here with STED should be taken qualitatively only. The shift in size distribution upon CoA binding was perhaps visible in STED unlike in dSTORM, since the sample did not undergo GraFix.

### 3.10. Electron microscopy

#### 3.10.1. Truncated cores

Prior to cryo-EM, proteolytically obtained truncated PDHc core samples were tested if they were intact and homogeneously distributed by negative staining EM. Typically, the samples that did not undergo GraFix step tend to aggregate more with only slight improvement when 1-2 mM CoA was added. In contrast, by performing GraFix, these particles could be stabilized and homogeneously distributed to a level deemed good enough to carry out cryo-EM measurements.

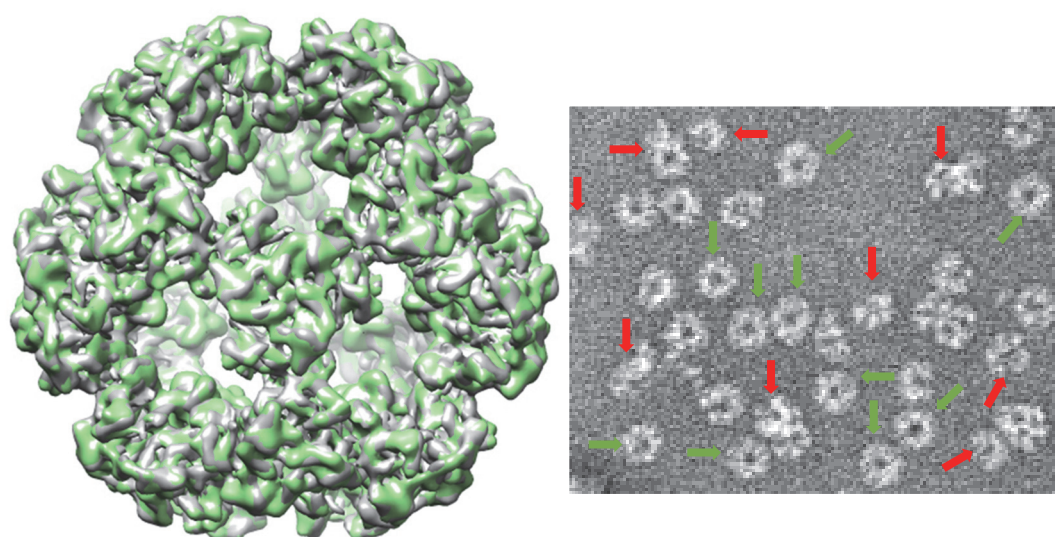


**Fig. 29: Negative staining EM of proteolytically obtained truncated *hE2* cores.**

Without GraFix, these cores either completely aggregated or tend to associate close to each other. In the GraFixed sample, the core particles appear to be distributed better, although fraction of the population did form aggregates or higher order associations. Same was the case with truncated *hE2/E3BP* cores (see Appendix G).



From the good-looking batch of truncated *hE2* cores, cryo-EM was performed in both presence and absence of CoA. The maximum resolution that could be achieved with these two samples were  $\sim 7 \text{ \AA}$ , with no striking differences in density maps (see Fig. 30). Furthermore, with in the same population, no apparent variation in sizes was detected, which meant fixation with glutaraldehyde in GraFix step stopped the breathing process. Nevertheless, this endeavor was considered a success, as the EM density was better resolved than the best resolution described in literature ( $\sim 9 \text{ \AA}$ ) for any human PDH complex core or its variants (Yu, Hiromasa et al. 2008). However, we also observed many broken particles in the negative staining EM micrograph, which could be the result of proteolysis. Since an unspecific protease ‘thermolysin’ was used to cleave flexible lipoyl arms of *hPDHc* cores, it must have also digested parts belonging to core forming domain, at least in few E2 chains.

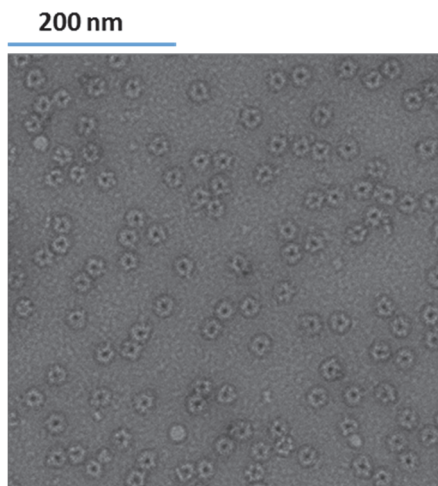


**Fig. 30: Cryo-EM density maps of truncated *hE2* core in the absence (grey) and presence (green) of CoA at their 3 fold axis.**

The EM densities at substrate bound and unbound state of the cores are similar at  $7 \text{ \AA}$  resolution. Pseudo-atomic modeling was not done with these densities as the re-inspection of the EM micrographs revealed that the core population in a sample was a mixture of intact (green arrows) and broken (red arrows) cores. The fraction of the particles in broken states or aggregating with other particles was high.

Thus, alternative way of generating core forming IC domain of *hE2* was considered. One of the option was to clone this domain alone and express and purify as lipoyl arm free construct. Fortunately, this construct (KD-*hE2*) was already available in our lab (cloning done by Dr. Kathrin Schroeder Tittmann), but it was shown to be completely

insoluble. Similar problems were also expressed in literatures (Yu, Hiromasa et al. 2008) sharing a concern that the solubility of the core-forming domain is the real bottleneck for crystallization trials. Also, while performing proteolysis of full-length *hE2* and *hE2/E3BP*, few minutes after addition of thermolysin, the solution would turn turbid signifying the insolubility of naked core forming domains at pH 7.5 (buffer for *hPDHc* cores). After performing ThermoFluor assay of thus obtained truncated cores to check their stability (see Appendix F), the pH of the buffer was lowered to pH 6-6.5 which improved the stability. Using this knowledge, 50 mM MES pH 6, 150 mM KCl, 1 mM DTE was used as a purification buffer for KD-*hE2* and indeed it could yield sparingly soluble cores (enough for EM investigation). With higher salt and reducing agent concentrations to 300 mM KCl and 5 mM DTE and adding 2 mM CoA, these KD-*hE2* core samples looked homogeneously pure and intact, at least in negative staining EM.



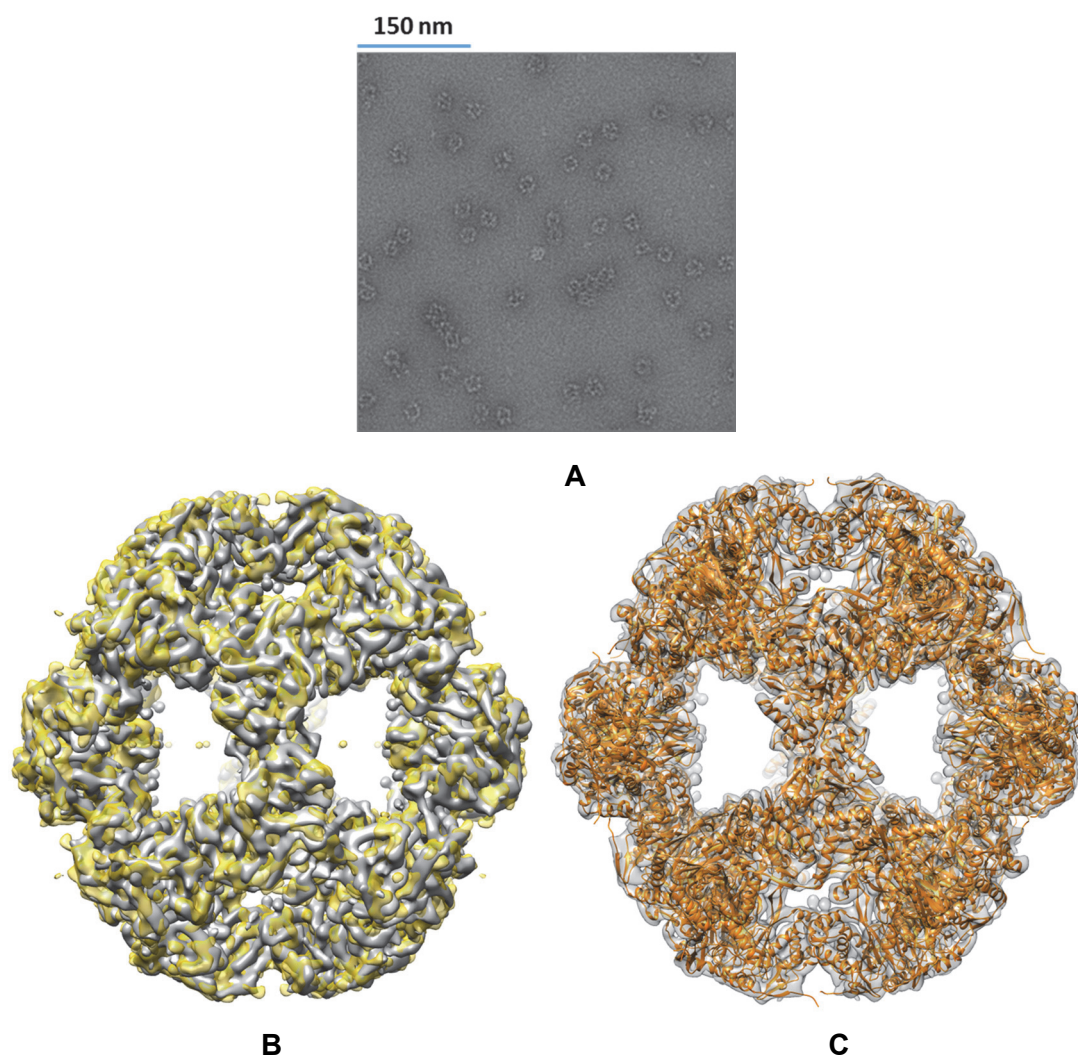
**Fig. 31: Negative staining EM micrograph of KD-*hE2* (+ GraFix + 2 mM CoA).**

The core particles looked homogeneously distributed and much more intact than proteolytically obtained *hPDHc* cores. Their tendency to associate with each other could be greatly reduced by using high salt and high reducing agent concentrations (300 mM KCl, 5 mM DTE in 50 mM MES pH 6).

However, in a parallel experiment, we could also obtain high quality sample for full length *hE2/E3BP* cores as well and thus, cryo-EM measurement was performed for it instead and is discussed in the following section.

### 3.10.2. *hE2/E3BP* cores

The best cryo-EM density map reported for full length human PDHc cores are  $\sim 15$  Å for *hE2* core (Yu, Hiromasa et al. 2008) and  $\sim 18$  Å for *hE2/E3BP* core (Vijayakrishnan, Kelly et al. 2010). They are bettered only by truncated *hE2* core with a resolution of  $\sim 9$  Å (Yu, Hiromasa et al. 2008), and a pseudo-atomic model of E2 core forming domain was reconstructed from this density map. We performed cryo-EM measurement with full-length *hE2/E3BP* and could attain a much better resolution of 6.3 Å. At this resolution, densities from *hE2* and E3BP cannot be differentiated, mainly because they share 50 % sequence similarity in their core-forming domain and are expected to have similar structure. Nevertheless, atomic models for these subunits were calculated from the hybrid EM density map. The initial homology models for them were calculated by online Robetta server which used PDB 3B8K as a template which were fitted to the density map and models refined (refer Section 2.2.8.2.).

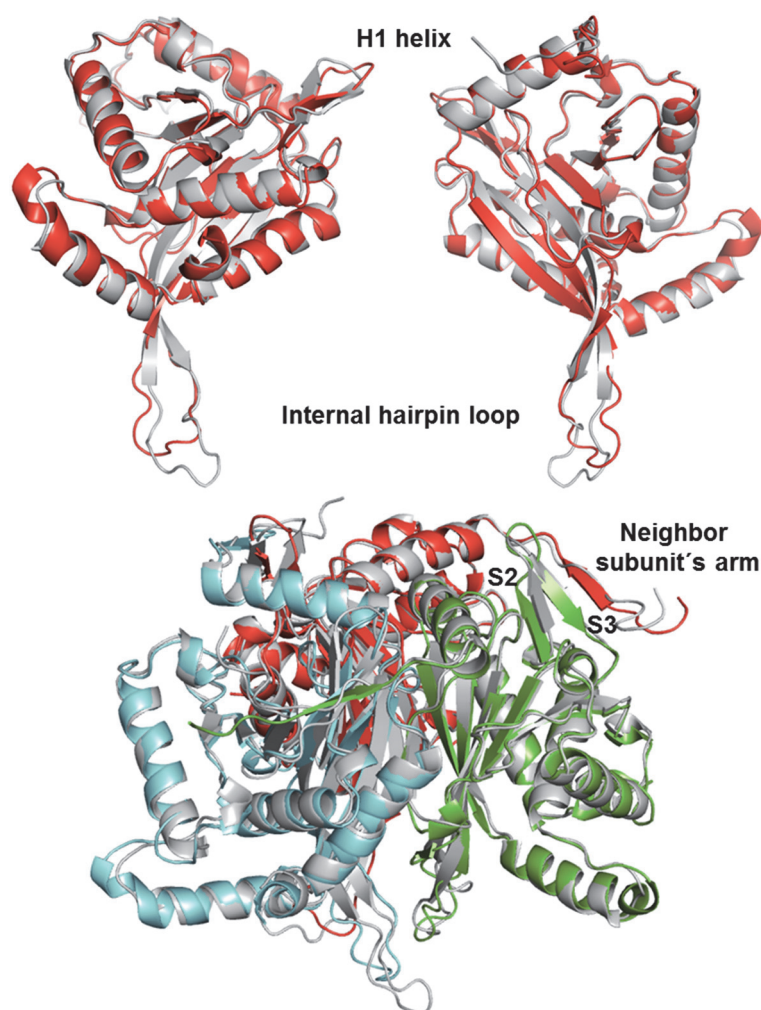


**Fig. 32: Comparing EM density maps of *hE2/E3BP* core with truncated E2 core from Yu, Hiromasa et al. 2008 and fitting 60 E2 monomers to the hybrid EM map of *hE2/E3BP* core.**

A) Negative staining EM micrograph of *hE2/E3BP* core (+GraFix+CoA) B) The cryo-EM density map of *hE2/E3BP* core (grey) at 2 fold axis is compared with that from truncated E2 core (yellow). They look similar, signifying the presence of E3BP in the core do not alter the overall shape. The improvement in resolution from 8.8 Å (truncated E2 core) to 6.3 Å (*hE2/E3BP*) meant that at least the densities from helices were much sharper. C) To the EM density map of *hE2/E3BP*, 60 E2 monomers are fitted. It can be done with E3BP as well, since the final pseudo atomic models for *hE2* and E3BP core forming domains are similar and at this resolution, they cannot be differentiated.

The calculated atomic models for *hE2* and E3BP core forming domains were understandably similar due to their high sequence similarity. Their structures were similar to crystal structure of *A. vinelandii* and *E. coli* truncated E2 structure but showed differences to only reported truncated human E2 model (see Fig. 34). Interestingly, the differences were exactly at the regions where it was shown to be

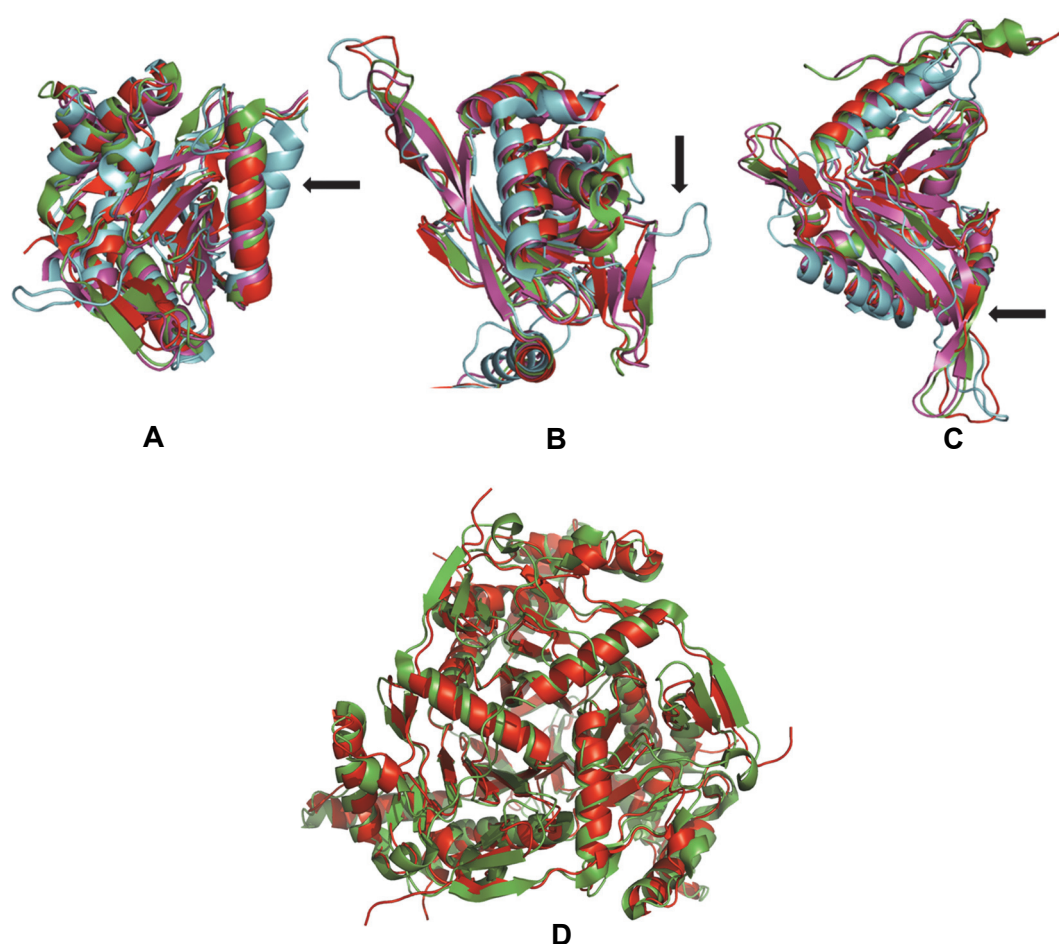
unique for human E2 when compared to bacterial versions. Since, we calculated our models from the best EM density map for PDHc core, we expect our models to be more accurate. Furthermore, recently it was shown via MD simulation that the truncated human E2 pseudo atomic model might be unstable and inaccurate (Hezaveh, Zeng et al. 2016), and recommended corrections. Taken all together, it appears that the core forming domains of human E2 and E3BP are indistinguishable from prokaryotic PDHc E2s.



**Fig. 33: Superposition of atomic models of *hE2* and E3BP core forming domains.**

The two opposing views of E2 (red) and E3BP (grey) monomers (upper panel) show that much of the secondary structural elements are conserved between the two. Only noticeable differences between the two models were in loop regions. The H1 helix and internal hairpin loop are the structural elements, which are characteristic surficial and innermost parts of the core. In the trimers of *hE2* and E3BP (lower panel), S2 and S3  $\beta$  strands can be seen making  $\beta$  sheet interaction with external arm of neighbor subunit of a related trimer unit, in a counter clockwise fashion.

Our structural model for *hE2* was also similar to bacterial E2 at the trimeric level (see Fig. 33), which meant the difference between 24meric and 60meric PDHc cores must lay at the orientation of inter trimer bridges (refer Section 1.4.1.d.). Keeping in mind that these bridges aren't properly resolved in our EM density map in comparison to the crystal structure of bacterial E2 cores ( $< 3 \text{ \AA}$ ), the direct comparison of the orientations of the inter trimer bridges are not carried out here.



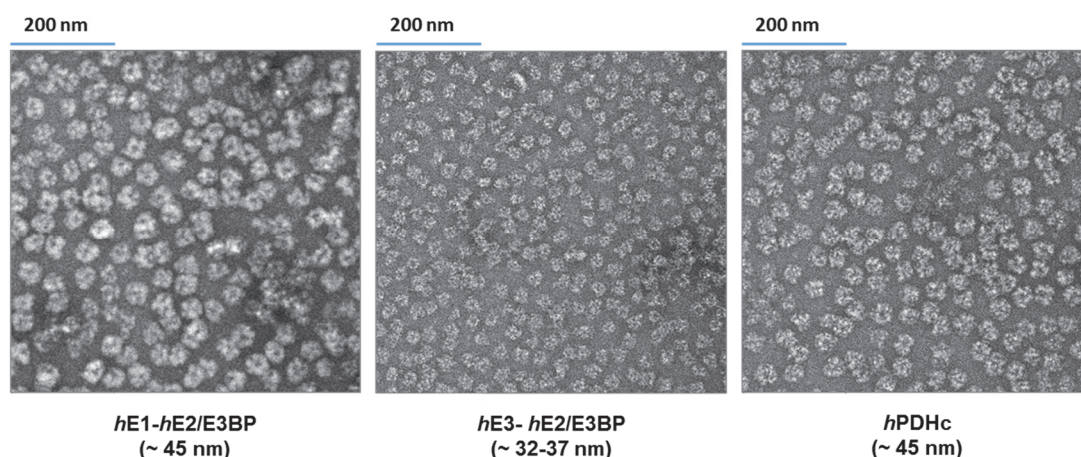
**Fig. 34: Comparing structural model of *hE2* with those from bacterial PDHc E2s and reported truncated *hE2*.**

Our *hE2* model (red) shows nice similarity to the E2 crystal structure from *E. coli* (green) and *A. vinelandii* (purple) while the truncated *hE2* monomer (blue) deviates from all the other models, specifically at three regions (black arrows). A) The arrow shows the H1 helix, which has different orientation for truncated *hE2* model. B) In a disordered region of the truncated *hE2* model, both the bacterial E2 structure and our *hE2* structure show a beta strand. C) In an inner hairpin loop region, bacterial E2s have long beta strands. While our *hE2* structure could show some ordered region at this part (small beta strand), the truncated *hE2* structure appears to have only a long loop. All these three differences were reported to be distinctive feature for human E2, but our model challenges this notion. D) Furthermore, at a trimeric level also, our *hE2* structure bears similarity with bacterial E2s (*hE2* trimer in red and *E. coli* PDHc E2 trimer in green)

### 3.10.3. *h*PDHc and sub-complexes

One of the main aim of the thesis was structural investigation of *h*PDHc and sub-complexes. This is important because a complete structure of PDH complex from any organismal sources have remained elusive. During the present work, these large protein assemblies were prepared by in-vitro reconstitution (refer Section 3.3). In the negative staining EM, we observed different size distribution of these complexes. While *h*PDHc and *hE1-hE2/E3BP* were roughly 45 nm, *hE3-hE2/E3BP* had a size variation of ~ 32-37 nm.

This was unlike sub-complexes from *B. stearotherophilus* where both E1 and E3 forms the outer shell near the same space (Milne, Shi et al. 2002, Milne, Wu et al. 2006), with the annular gap between core and peripheral proteins differing by ~ 1.5 nm only. Since, the same SBD is used for binding E1 and E3 in *bs*PDHc, albeit in mutually exclusive manner, such observation was not surprising. However, in human PDHc, since *hE2* and E3BP binds to *hE1* and *hE3* respectively, it does not need to be true.



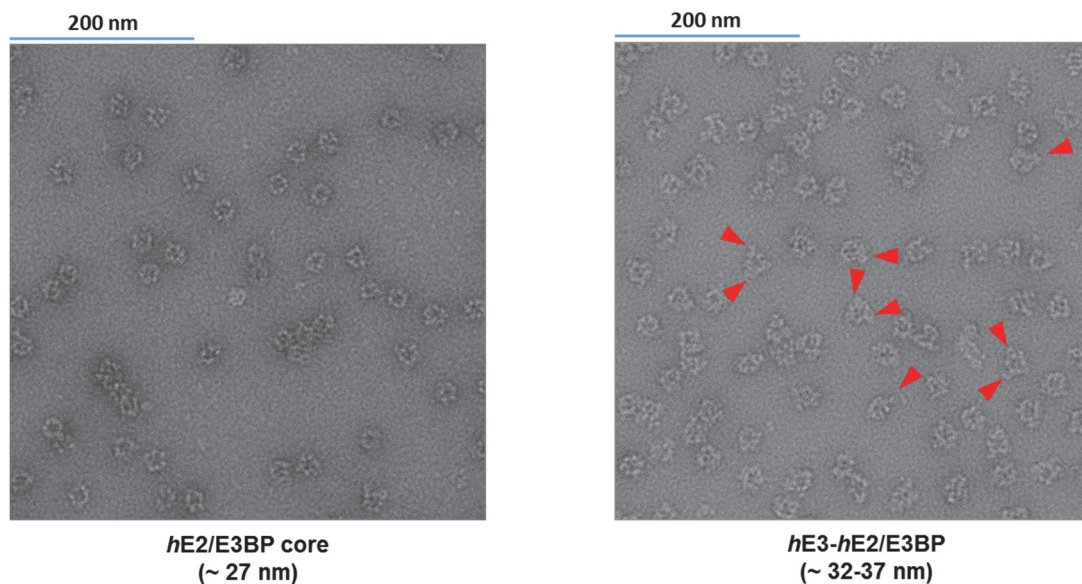
**Fig. 35: Negative staining EM micrograph of *hE1-hE2/E3BP*, *hE3-hE2/E3BP* and *h*PDHc.**

The *hE3* sub-complex appeared to be much smaller than *hE1* sub-complex or *h*PDHc (by ~ 10 nm). The sizes were calculated manually by using ruler. The presence or absence of cofactors such as CoA, ThDP or NAD<sup>+</sup> had no influence in this behavior.

On a closer inspection, *hE3* proteins densities had inhomogeneous distribution on the core surface (see Fig. 36), forming big patches at few spots while absent in the rest of the area of core surface. Even though 20 *hE3* proteins could be detected to bind to *hE2/E3BP* core via ITC (refer Section 3.5.) and a high resolution cryo-EM map could be generated from this core (refer Section 3.10.2.), the ability of *hE3* to cluster

together and thus, also perhaps E3BP, provides previously unidentified landscape for *hPDHc*. Also the smaller sizes of *hE3*-subcomplexes could be result of such distribution.

Although, a model for two E3BPs at 2 fold axis of the core binding to one *hE3* protein exists (Smolle, Prior et al. 2006), it meant that only 10 or 6 *hE3* would bind to the core. But it has already been disproven using ITC method during this thesis work (refer Section 3.5.).



**Fig. 36: Negative staining EM micrograph of *hE2/E3BP* and *hE3-hE2/E3BP*.**

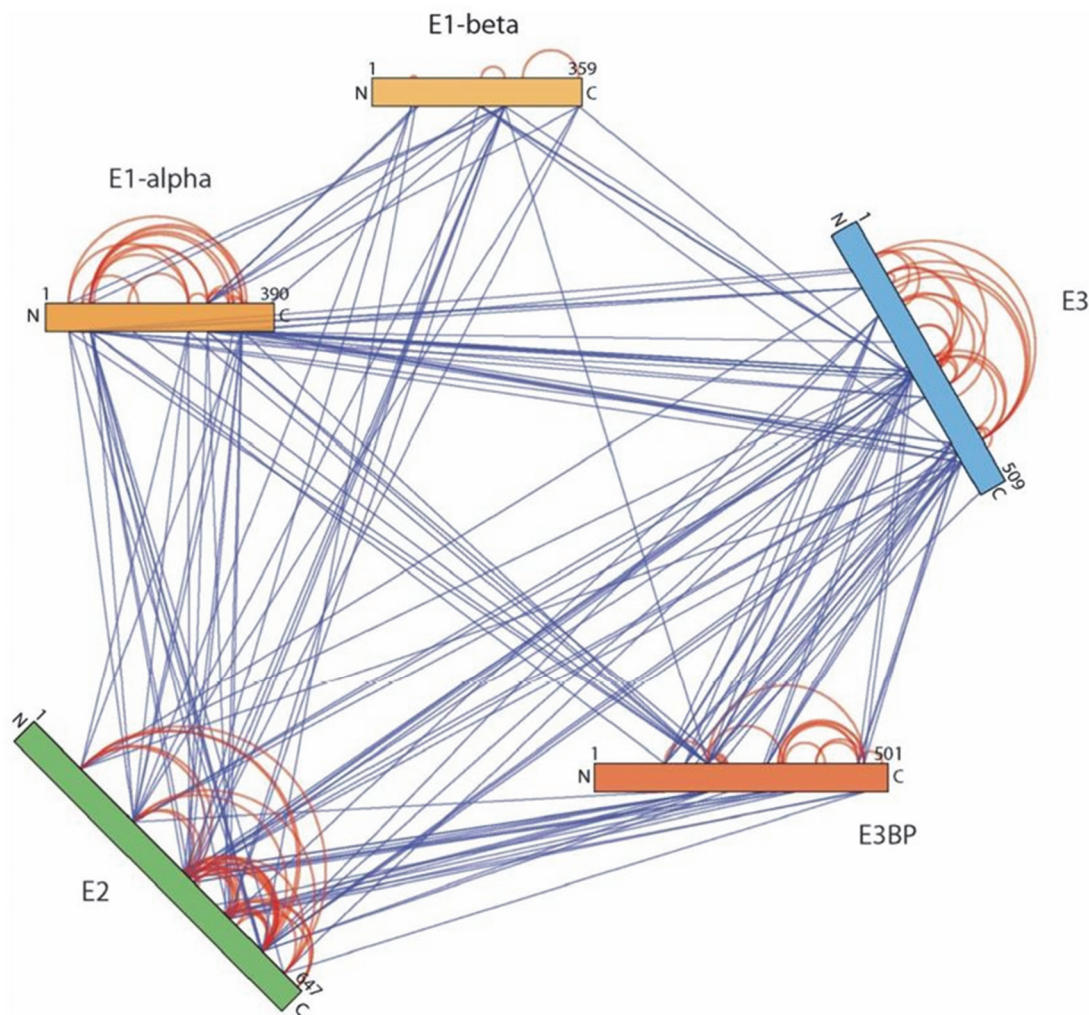
The *hE3* sub-complex is slightly bigger than *hE2/E3BP* core. Inhomogeneous distribution of peripheral densities could be clearly visible in the *hE3* sub-complex (marked by red arrows). The varying sizes of the sub-complex might be due to presence of extra peripheral densities at few orientations of the particles.

### 3.11. Mass spectrometry

#### 3.11.1. Cross-linking mass spectrometry

In order to study the dynamics of *hPDHc*, cross linking MS was employed using BS3 as a cross-linker. This should give us crosslinks between C $\alpha$  of lysine residues that are roughly 27 Å apart (Rappsilber 2011). As expected, many crosslinks were detected between various subunits of *hPDHc*. They were carefully sorted out to remove crosslinks involving linker regions (full set of crosslinks in Appendix H). Unless mentioned specifically, the crosslinks described hereafter would only involve *hE1* and *hE3* proteins and IC domains, SBD and lipoyl domains of *hE2* and *E3BP*.





**Fig. 37: Crosslinking map of *hPDHc*.**

All the subunits could make multiple crosslinks with each other, which meant they are at close proximity to each other, at least for some time. The subunits here are shown as rectangular bars with their N and C terminus at opposite ends. (This Fig. was kindly provided by Dr. Carla Schmidt)

By classifying these crosslinks based on domains where crosslinking partners are located, many important details about *hPDHc* functioning could be revealed and will be described in the following sections:

### 3.11.1.1. Crosslinks within the core-forming domain

Based on the structure of *hE2/E3BP* core we obtained, all the lysine residues that are C-terminal to residue number 415 for *hE2* and 269 for *E3BP* were considered to belong to the core forming domains. In total, we could detect nearly 36 different cross-linkings in the core-forming domain of *hE2* and *E3BP* that had at least one acceptable

spectrum (see Table 8). Out of those, five crosslinks could be identified as inter-*hE2* (K490-K'490 and K604-K'604) or inter-E3BP (K450-K'450, K460-K'460, K488-K'488 and K491-K'491) chains (see Fig. 38). Except crosslinks between *hE2* and E3BP (10 in total) and aforementioned crosslinks, rest could not be determined as inter or intra chain. Hence, precaution was taken to take such crosslinks as both inter and intra subunit, such that option yielding a distance near 27 Å or smaller between crosslinkers, were considered likelier. For all the inter subunit crosslinks, it was considered to be both intra trimer and inter trimer. Again, option yielding a distance near 27 Å or smaller between crosslinking partners was considered likelier.

#### Crosslinks between lysine residues of *hE2* IC domain

1 <sup>st</sup> crosslinking partner	2 <sup>nd</sup> crosslinking partner + CoA	2 <sup>nd</sup> crosslinking partner - CoA
462	473, 547, 604	473, 547, 637
466	473, 637	473, 547, 637
473	462, 466, 547, 552, 637	462, 466, 547, 552
490	490	490
547	462, 473, 552	462, 466, 473, 552
552	473, 547	473, 547
604	462, 604	604
637	466, 473	462, 466

#### Crosslinks between lysine residues of E3BP IC domain

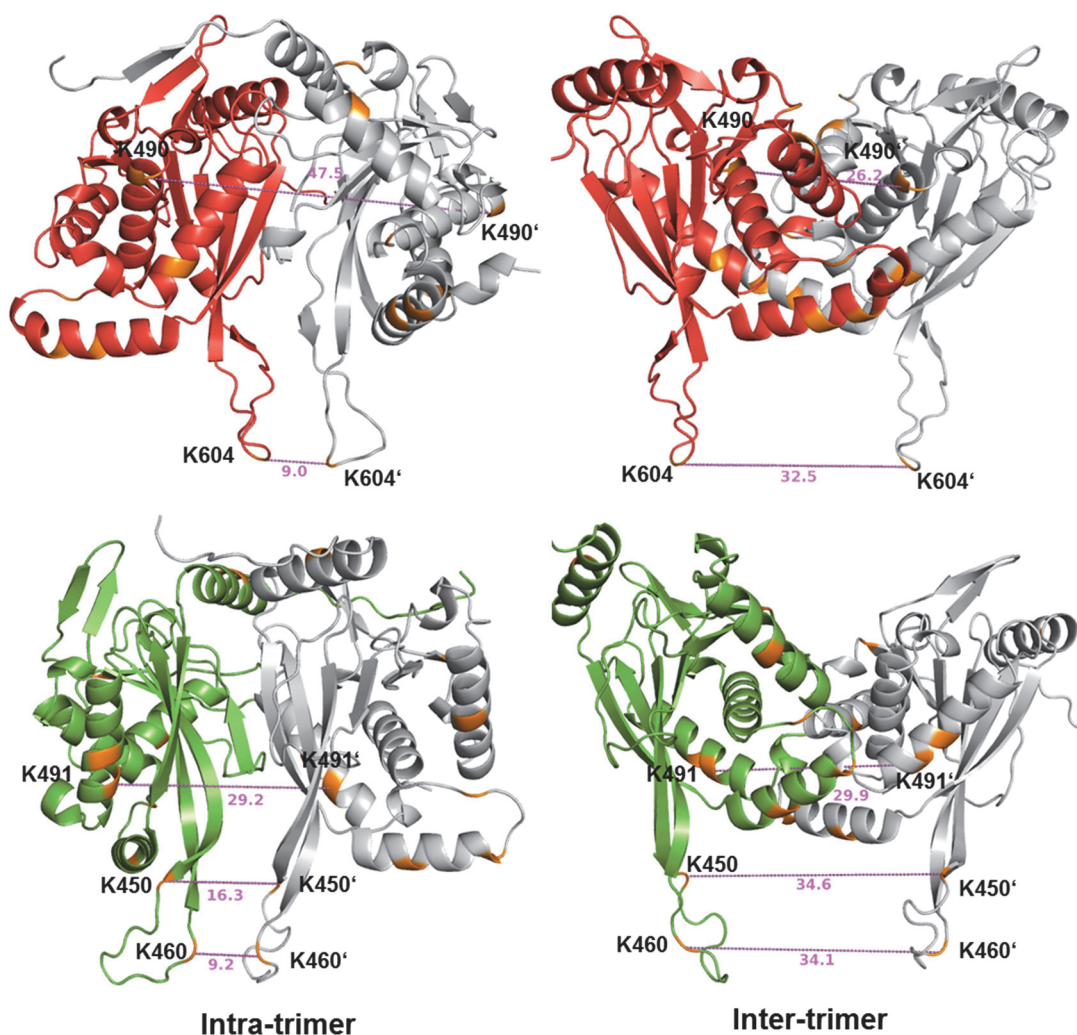
289	460	
314	321, 450, 460	450, 460
321	314, 450, 460	450, 460
325	394	394
379	450	460
394	325	325
450	314, 321, 379, 450, 491	314, 321, 491
460	289, 314, 321, 460	314, 321, 379, 460
488	488	
491	450	450, 491

**Crosslinks between lysine residues of E3BP and *hE2* IC domains**

E3BP	<i>hE2</i> + CoA	<i>hE2</i> - CoA
295	547	440
314	462	462
341	473	473
450	466, 473	473
460	604	462, 604
488	462	
491	547	

**Table 8: All the crosslinks in *hE2* and E3BP core forming IC domains in the presence and absence of CoA.**

Although crosslinks could be sorted out between inter *hE2* and E3BP, those only involving either *hE2* or E3BP could not be sorted out to occur intra chain or inter chain. Exceptions are crosslinks between lysine residues at same positions that ought to be inter chain (red). Still, no crosslinks can be assigned with full certainty to occur between chains in a same or different trimers.

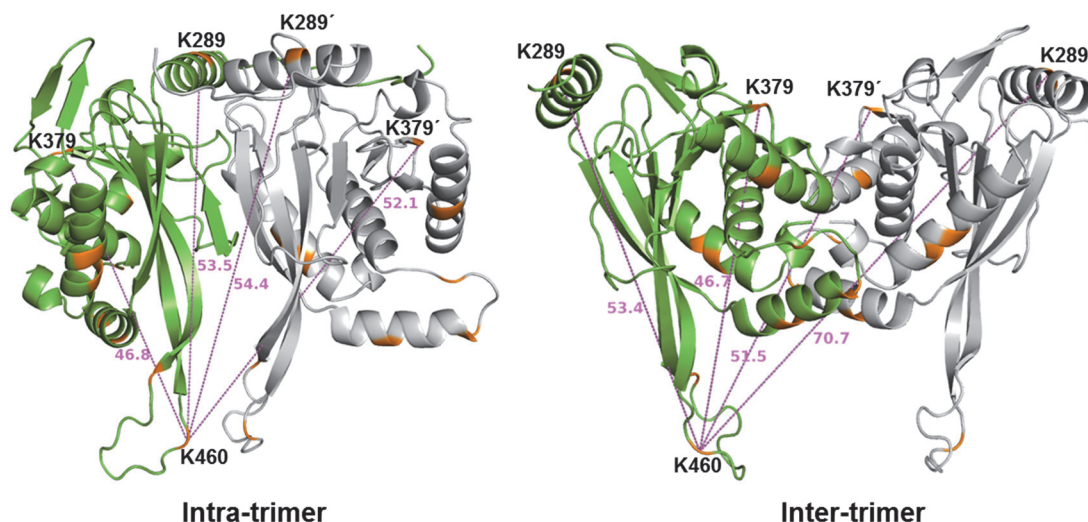


**Fig. 38: Inter-chain crosslinks in *hE2* and *E3BP* core forming IC domains.**

*hE2* chains (red and grey : upper panel) and *E3BP* chains (green and grey : lower panel) are shown as intra-trimer dimer or inter-trimer dimer with all the crosslinking lysine residues (C $\alpha$ ) in orange. Confirmed inter chain crosslinks are highlighted by dashed lines in magenta with distance between C $\alpha$  of crosslink partners given in Angstrom. As an example, crosslink between K490-K490' in *hE2* chains is much likelier to be inter trimer where crosslink distance will be only 26.2 Å whereas it is 47.5 Å for inter-trimer case.

Most of the crosslinks mentioned in Table 8 have a crosslinking distance of < 35 Å when analyzed considering intra or inter chains as well as intra or inter trimer. This was deemed acceptable considering protein as well as cross linker's dynamics. The exceptions were K450-K379 (~ 47 Å), K460-K289 (~ 53 Å) and K460-K79 (~ 38 Å) for *E3BP*-*E3BP* crosslinks and K295-K547 (~ 35 Å) for *E3BP*- *hE2* crosslinks. Among them, the first two crosslinking distances are very large and occur between *E3BP*'s surficial residues and inner hairpin loop. Overall, the core forming domains of *hE2* and *E3BP* seem to be dynamic and moreover, four crosslinks between lysine residues of

E3BP were observed that are possible only if there is a large-scale conformational changes.



**Fig. 39: Crosslinks between distant lysine residues in E3BP IC domain.**

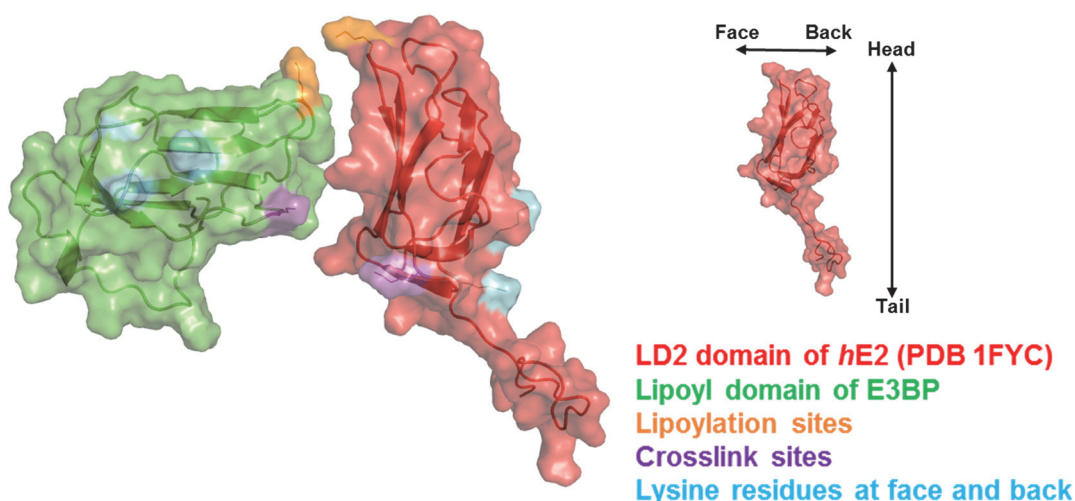
The two crosslinks K460-K379 and K460-K289 are shown here as both intra and inter chain in origin for both intra trimer and inter trimer situations. The two E3BP dimers are shown in green and grey monomers for both intra and inter trimer associations. The crosslinks are shown as dashed line in magenta with distance between C $\alpha$  of crosslink partners given in Angstrom. All the lysine residues that make at least one crosslink are shown in orange. After comparing each possible situation, it seems that these crosslinks are intra chain in origin which have the shortest crosslink distances of ~ 47 Å for K460-K379 and ~ 53 Å for K460-K289.

### 3.11.1.2. Inter lipoyl domain interaction

One of the key step in PDHc reaction which have never been observed physically but have been shown extensively via biochemical and computer simulation studies (refer Section 1.4.1.a.) is the interaction between lipoyl domains for acetyl as well as reducing equivalent transfer. Luckily, in the crosslinking MS, we did observe crosslinking between lipoyl domain of E3BP with those from *hE2*. The lysine residues that are lipoylated cannot make crosslinks with other lysine residues and we did not observe it either. Still, there are many other lysine residues at the surface of these domains, which should be able to make crosslinks when they come to close proximity to each other. Interestingly, we observed crosslinks between K-120 of E3BP lipoyl domains with K-114 and K-241 of *hE2*. We also observed K-120 of E3BP interacting with K-78 of another E3BP subunit for which intra chain crosslinking was excluded as this lysine is structurally at equivalent position to K-114 and K-241 of LD1 and LD2 of *hE2* and thus most likely is the interacting partner to K-120 of another E3BP lipoyl

domain. Same set of crosslinks were also present in the crosslink MS data from *hE2/E3BP* core (see Appendix I).

When the structure of LD1 (PDB 2DNE), LD2 (PDB 1FYC) and the homology model of lipoyl domain of E3BP (calculated by Robetta online server) was oriented such that the lysine residues that are lipoylation sites come close to each other (must happen for acetyl or electron pair transfer) as well as the crosslink partners, it appears that these lipoyl domains interact with each other at only one orientation (see Fig. 40). Other orientations do not look probable, especially because even when there are bunch of lysine residues at the back portion of all these lipoyl domains as well as face of E3BP lipoyl domain they did not make any crosslinks. As the lipoyl arm is supposed to be very flexible and moving around stochastically, this observed restrain in how the lipoyl domains would actually come into each other's contact should challenge the notion of unrestrained mobility of lipoyl arms.



**Fig. 40: Crosslinks between lipoyl domains.**

The orientation of lipoyl domains shown here (face of LD2 and head of lipoyl domain of E3BP) seems to be the only plausible orientation to explain the observed crosslinks. K-114 and K-241 of LD1 and LD2 and K-78 of E3BP lipoyl domain are all at equivalent position shown here in the purple in the structure of LD2. Other orientations involving back or face of E3BP lipoyl domain and back of LD1 and LD2 should have encouraged crosslinks between other lysine residues (shown in cyan) than the observed ones.

### 3.11.1.3. Interactions of lipoyl domains with *hE1* and *hE3* proteins

The same lysine residues of lipoyl domains of *hE2* and E3BP that were making crosslinks with each other were also the ones making crosslinks with *hE1* and *hE3* proteins. In the first glance, it appears that all the lipoyl domains of *hE2* and E3BP can make crosslinks with both the peripheral enzymes (see Table 9). But the numbers of different crosslinks they can make with these enzymes are different. The number of different crosslinks can also be used as a measure of space explored by these lipoyl domains on the surface of *hE1* and *hE3*. In that regard, the lipoyl domain of E3BP seems restricted to explore *hE1* (see Fig. 41). The lipoyl domains of *hE2* however can explore the surface of *hE3* albeit lesser than lipoyl domain of E3BP. If the space explored by these lipoyl domains at *hE1* and *hE3* is directly related to the activity, then it can be assumed that the lipoyl domains of *hE2* can participate in *hE1* and *hE3* reactions with similar efficiency while the lipoyl domain of E3BP has much higher preference for *hE3* reaction.

#### Crosslinks between *hE1* lysine residues with lipoyl domains

Lipoyl domains	<i>hE1</i> + CoA	<i>hE1</i> - CoA
E3BP-120	$\alpha$ -40, $\alpha$ -77	$\alpha$ -77
<i>hE2</i> -114	$\alpha$ -63, $\alpha$ -244, $\alpha$ -321, $\beta$ -227	$\alpha$ -244, $\alpha$ -336
<i>hE2</i> -241	$\alpha$ -77, $\alpha$ -83, $\alpha$ -147, $\alpha$ -244, $\alpha$ -321, $\beta$ -184, $\beta$ -227, $\beta$ -354	$\alpha$ -40, $\alpha$ -83, $\alpha$ -244, $\alpha$ -321, $\alpha$ -336, $\alpha$ -344, $\beta$ -68, $\beta$ -227, $\beta$ -354

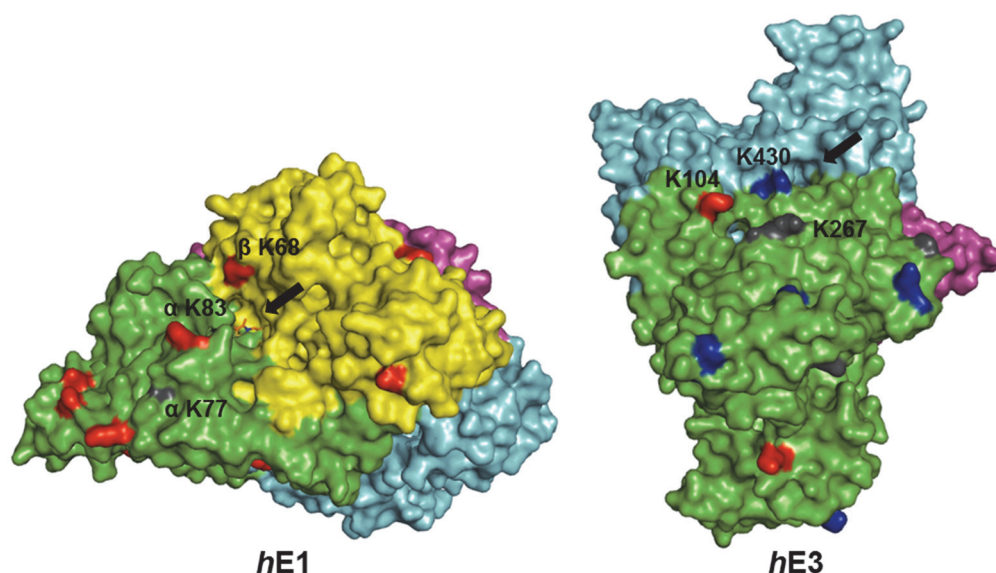
#### Crosslinks between *hE3* lysine residues with lipoyl domains

Lipoyl domains	<i>hE3</i> + CoA	<i>hE3</i> - CoA
E3BP-120	159, 267, 273, 277, 284, 410, 430	159, 267, 277, 410, 430, 445
<i>hE2</i> -114	104, 267, 445	104, 445
<i>hE2</i> -241	104, 267, 445	277, 320

**Table 9: All the crosslinks between lipoyl domains of *hPDHc* with *hE1* and *hE3*.**

Understandably, the variety of crosslinks between lipoyl domains of *hE2* with *hE1* and lipoyl domain of E3BP with *hE3* is more as *hE1* binds to *hE2* and *hE3* binds to E3BP, respectively. Still, it appears that lipoyl domains of *hE2* can make considerable number of crosslinks to *hE3* in comparison to lipoyl domain of E3BP with *hE1*. Additionally, the pattern did not change significantly upon binding of core to CoA.

The dichotomy in relative preferences to visit *hE1* and *hE3* possibly also has functional consequences, especially in regard to responsibility shared by them in PDHc reaction. Furthermore, the functional redundancy shown between lipoyl domains of prokaryotic PDHc E2 might not have conserved in the lipoyl domains of eukaryotic E2 and E3BP during the course of evolution.



**Fig. 41: Surface sampling by lipoyl domains of *hE2* and E3BP at *hE1* and *hE3*.**

The structure of *hE1* (PDB 1ZY8) and *hE3* (PDB 1NI4) are shown with all the lysine residues that crosslinks with lipoyl domains (entrance for lipoyl moiety shown by black arrows). Exclusive crosslinks with *hE2*-lipoyl domains are in red, exclusive crosslinks with E3BP-lipoyl domain are in blue and common crosslinks are in grey. Only those lysine residues that are close to the lipoyl entrance sites at *hE1* and *hE3* are labelled.

#### 3.11.1.4. Position of *hE3* protein

Another observation we made with crosslinking MS was the ability of *hE1* and *hE3* to make crosslinks with all the other components of *hPDHc* and their domains (see Fig. 37). This was consistent with the negative stain EM results as both of these proteins could be seen at peripheral region to the core.

Interestingly, we also observed few crosslinks between *hE3* with inner hairpin loops of *hE2* and E3BP (see Table 10). As these loops are situated at the inner cavity of the dodecahedral *hPDHc* core, the only way *hE3* can make crosslinks with them is if it snugs inside the cavity via pentagonal openings of the core. Additionally, one crosslink was also observed for *hE1*  $\alpha$  subunit with inner hairpin loop of *hE2*, but it might be a false positive or a result of degradation/dissociation of *hE1*, as the full *hE1* ( $\sim 8 \times 10$  nm) is too big to enter pentagonal openings ( $\sim 6.5$  nm).



**Crosslinks between *hE3* with inner hairpin loop of  
*hE2* and E3BP**

Inner hairpin loop	<i>hE3</i> + CoA	<i>hE3</i> - CoA
E3BP-450		267, 430, 445
E3BP-460	277	277, 430
<i>hE2</i> -604	277	

**Crosslinks between *hE1* with inner hairpin loop of  
*hE2* and E3BP**

Inner hairpin loop	<i>hE1</i> + CoA	<i>hE1</i> - CoA
<i>hE2</i> -604	$\alpha$ -77	

**Table 10: All the crosslinks between inner hairpin loops with *hE1* and *hE3*.**

In general, high quality MS spectra could be obtained for crosslinks between inner hairpin loops of *hE2* and E3BP with peripheral proteins in the absence of CoA. While only one crosslink was found in the case of *hE1* (only  $\alpha$ -subunit), several were found for *hE3*, indicating higher probability for *hE3* to enter the pentagonal openings of the core.

Furthermore, the crosslinking pattern shows that upon binding of CoA to the core, the ability of *hE3* to enter the inner cavity of the core is reduced. Perhaps, the CoA binding affects the breathing such that the pentagonal openings do not stretch to accommodate *hE3* protein. To validate this pattern, we performed quantification of the crosslinks between *hE3* and inner hairpin loop in CoA bound and unbound states (see Table 11), which further confirmed that these crosslinks are observed more frequently when CoA is absent.

Inner hairpin loop	<i>hE3</i>	CoA: no CoA	no CoA : CoA
E3BP-450	267	0.85	1.3
E3BP-450	430	0.81	1.23
E3BP-450	445	1.04	1.1
E3BP-460	277	1.21	0.9
E3BP-460	430	0.69	2
<i>hE2</i> -604	277	0.54	1.9

**Table 11: Quantification of crosslinks between *hE3* and inner hairpin loop of *hE2* and E3BP.**

All the crosslinks included lysine residues from inner hairpin loop of E3BP except for one (*hE2*-604). All the mentioned crosslinks are provided with the ratio of their respective ion intensities during MS in presence and absence of CoA. As an example, the ratio greater than one in 'CoA: No CoA' means the corresponding crosslinks are abundant in the presence of CoA in comparison to its absence.

### 3.11.1.5. Change in lipoyl arm dynamics in CoA bound state

Another key observation we could make from quantification of crosslinks were that the lipoyl domains of E3BP and *hE2* visit active site of *hE3* more frequently after CoA binding (see Table 12). For example, the crosslink between lipoyl domain of E3BP and *hE3*-267, which is a lysine residue, near the lipoyl entrance site of *hE3* increased by 41 folds in the presence of CoA. Lipoyl domains of *hE2* also appear to go near lipoyl entrance site of *hE3* (< 2 folds) but not as much as lipoyl domain of E3BP. Similarly, the cross talk between lipoyl domain of E3BP and LD2 of *hE2* increases by ~ 3 folds. The approach of *hE2* lipoyl domains to the core's lipoyl entrance site also increased by 1.5-2.5 folds but not E3BP lipoyl domains. All of these changes in lipoyl domains crosslinking pattern indicate heightened preference to E2 and E3 reactions when CoA substrate is present. On the contrary, these lipoyl domains did not show large increase in inclination to visit active site of *hE1* compared to with *hE3*, suggesting CoA binding has lesser influence on steps happening before its binding in the PDHc reaction. The moderate increase in crosslinks of lipoyl domains with E1BD and E3BD in the presence of CoA indicates lipoyl domains occupy spaces closer to the core more frequently, possibly on their way to visit E2 and E3 active sites.

### Crosslinks between lipoyl domains

Lipoyl domain- E3BP	Lipoyl domain- <i>hE2</i>	CoA: no CoA	no CoA: CoA
120	114	0.92	1
120	241	2.94	0.75

### Crosslinks between lipoyl domains and lysine residues near lipoyl entrance site at *hE1*

Lipoyl domain	<i>hE1</i>	CoA: no CoA	no CoA: CoA
<i>hE2</i> -241	$\beta$ -68	2.24	0.45
<i>hE2</i> -241	$\alpha$ -83	0.41	2.44

### Crosslinks between lipoyl domains and lysine residues near lipoyl entrance site at *hE3*

Lipoyl domain	<i>hE3</i>	CoA: no CoA	no CoA: CoA
<i>hE2</i> -114	267	1.3	0.77
<i>hE2</i> -241	104	1.92	0.81
<i>hE2</i> -241	277	1.1	0.91
E3BP-120	267	40.77	0.06
E3BP-120	430	0.83	1.58

### Crosslinks between lipoyl domains and lysine residues near lipoyl entrance site at core

Lipoyl domain	core	CoA: no CoA	no CoA: CoA
<i>hE2</i> -114	<i>hE2</i> -547	1.03	1.3
<i>hE2</i> -114	<i>hE2</i> -552	2.48	0.4
<i>hE2</i> -241	<i>hE2</i> -547	1.54	1.21
E3BP-120	E3BP-488	1.05	0.95

### Crosslinks between lipoyl domains and subunit binding domains of *hE2* and E3BP

Lipoyl domain	SBD	CoA: no CoA	no CoA: CoA
---------------	-----	-------------	-------------

<i>hE2-114</i>	<i>hE2-367</i>	1.5	0.73
<i>hE2-241</i>	<i>hE2-362</i>	3.23	0.5
<i>hE2-241</i>	<i>hE2-367</i>	1.24	0.91
<i>hE2-241</i>	E3BP-194	1.76	0.84
E3BP-120	<i>hE2-363</i>	1.4	0.78
E3BP-120	E3BP-194	2.1	1.1

**Table 12: Quantification of all the crosslinks of lipoyl domains with SBD and lysine residues near the lipoyl entrance site at *hE1*, *hE3* and the core.**

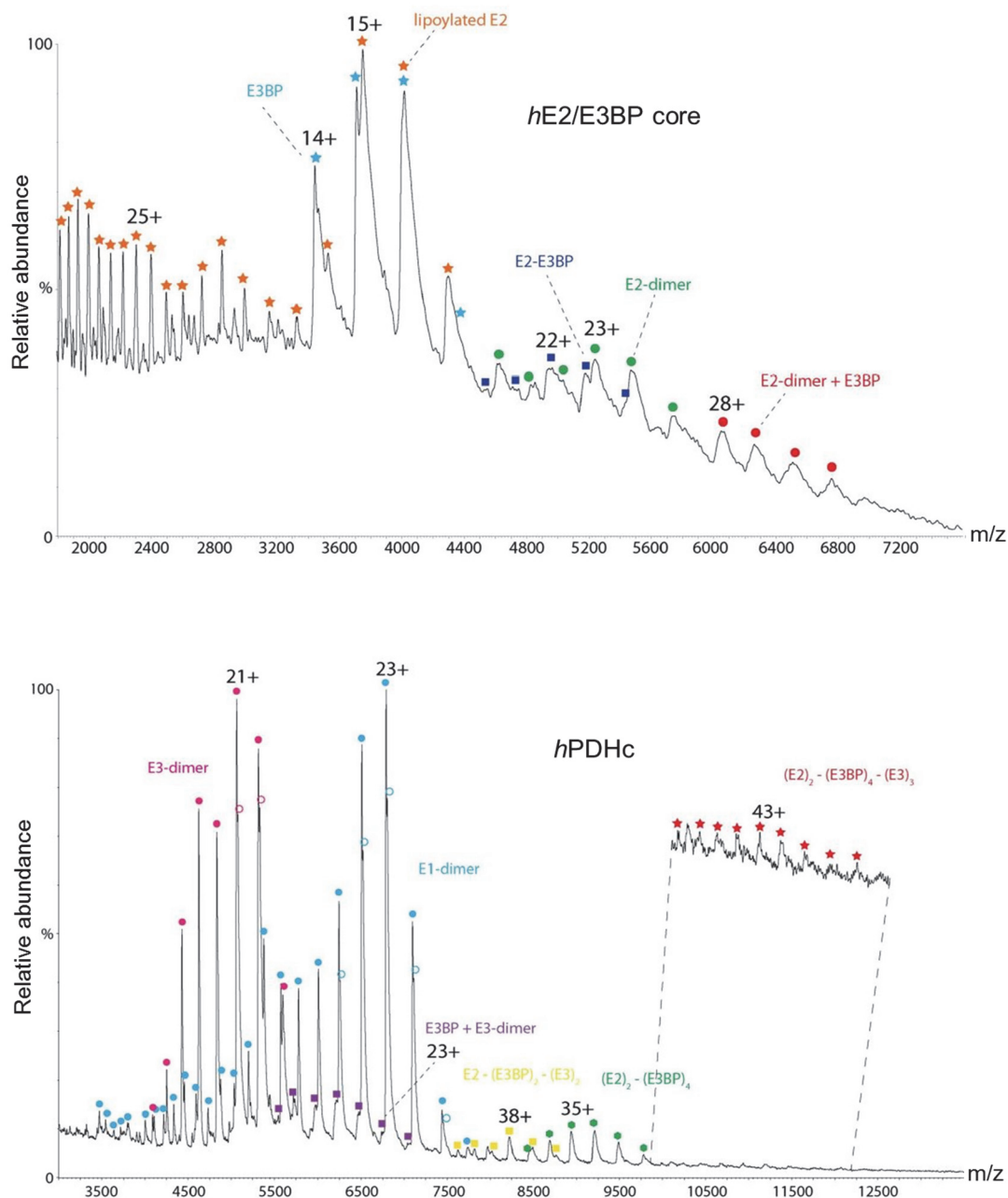
All the mentioned crosslinks are provided with the ratio of their respective ion intensities during MS in presence and absence of CoA. As an example, the ratio greater than one in 'CoA: No CoA' means the corresponding crosslinks are abundant in presence of CoA in comparison to its absence.

### 3.11.2. Native MS of *hE2/E3BP* cores and *hPDHc*

As the stoichiometry between subunits of PDHc from any organismal sources has not been accurately estimated, we performed native mass spectrometry in order to experimentally demonstrate all the possible ways of inter subunit associations. We performed these experiments with *hE2/E3BP* core and fully reconstituted *hPDHc*, but did not get proper spectra to assign masses to the whole complex or core. However, it was possible to assign peaks in the obtained MS spectra to smaller sub-complexes and single proteins (see Fig. 42).

As there might be many different combinations of subunits, which would yield masses close to the one assigned to a peak, we used SUMMIT software to discard all the combinations that we thought did not make sense. For example, a combination including many subunits of *hE1*  $\alpha$  subunits alone to give a certain mass will be considered highly unlikely if for the similar mass another combination in which both  $\alpha$  and  $\beta$  subunits are present in equal number. We also gave a window of 500-750 Da as a possible error to accommodate as many combinations as possible. Also, the lipoyl domain of E3BP was considered to be both lipoylated or un-lipoylated and the *hE2* chain to be bi-lipoylated, mono-lipoylated and un-lipoylated, with corresponding masses of lipoate added to the peptide chain mass, in order to cover whole range of possibilities. At the end, the presence or absence of ThDP,  $Mg^{2+}$ , FAD, lipoate and CoA did not change the logical combinations of *hPDHc* subunits we arrived at. These

combinations for each masses assigned to a peak resolved in the native MS spectra are given in Table 13.



**Fig. 42: Native MS spectra of *hE2/E3BP* core and *hPDHc*.**

Relative abundance of charged species are plotted against their respective mass by charge ratio, in a typical native mass spectra. The *hE2/E3BP* core and *hPDHc* could not be detected as an intact unit but rather disintegrated during MS, thereby leaving footprints of their corresponding subunits and their oligomers. (These figures were kindly provided by Dr. Carla Schmidt)

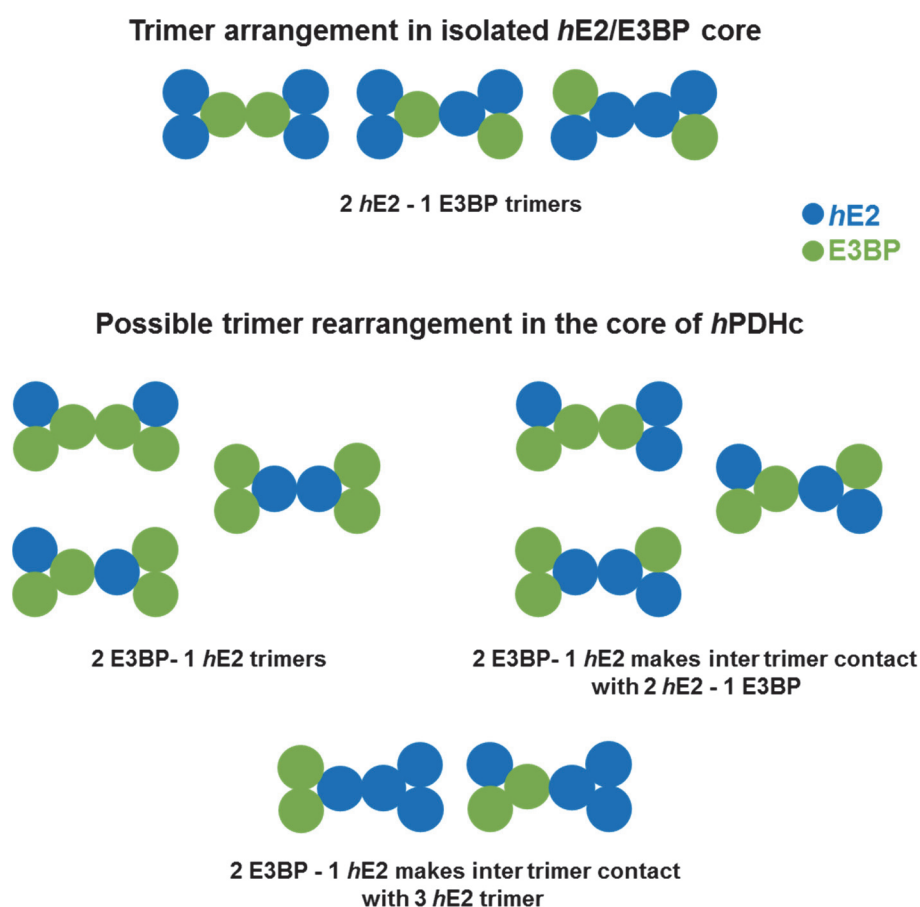
The most striking difference in spectra between the core and the whole complex was spotted in the trimer unit formation of the core. While in the isolated core, a peak for a trimer unit could be assigned for 2 *hE2* - 1 E3BP, in *hPDHc*, it was replaced by 2 E3BP - 1 *hE2* type of trimer. As the core of *hPDHc* consists of 40 *hE2* and 20 E3BP (refer Section 3.5), in order to balance the stoichiometry, for every 2 E3BP – 1 *hE2* kind of trimer, there must exist one trimer with all *hE2* subunits. Since we did not detect 3-*hE2* type of trimer for *hPDHc*, it cannot be concluded with absolute certainty that the isolated core will only contain a trimer with 2 *hE2* - 1 E3BP composition. However, it can be predicted that the different trimer arrangements detected in an isolated core and reconstituted complex might be dominant species in the corresponding system.

<b><i>hE2</i>/E3BP core</b>		<b>Human PDH complex</b>	
Peak (Da)	Likely candidates (+/- 750 Da)	Peak (Da)	Likely candidates (+/- 500 Da)
169422	2 <i>hE2</i> - 1 E3BP	470420	6 E3BP- 3 <i>hE2</i> or 4 E3BP – 2 <i>hE2</i> – 3 <i>hE3</i>
120705	2 <i>hE2</i>	312938	4 E3BP – 2 <i>hE2</i> or 2 E3BP – 1 <i>hE2</i> – 1 <i>hE1</i>
108845	1 <i>hE2</i> - 1E3BP	263147	2 E3BP – 1 <i>hE2</i> – 1 <i>hE3</i>
60113	1 <i>hE2</i>	156157 and 156447	2 E3BP – 1 <i>hE2</i>
48251	1 E3BP	154996	1 E3BP - 1 <i>hE3</i>
		106311, 106695 and 104577	1 <i>hE3</i>

**Table 13: List of all the possible logical combinations of subunits that have similar masses to the one assigned for resolved peaks in the native MS spectra.**

In total, 5 peaks for the *hE2*/E3BP core and 9 for *hPDHc* could be assigned masses and their logical combinations of subunits are tabulated (SUMMIT software). Some peaks with similar masses yielded the same set of combinations. The protein subunits mentioned in the table can be both with and without their corresponding cofactors, as it did not change the combination. (*hE1* = heterodimer of  $\alpha\beta$  dimer and *hE3* = homodimer)

The detection of 1 *hE2* - 2 E3BP kind of trimer in at least a reconstituted complex means that a population of PDH complex containing E3BPs might be an inhomogeneous mixture of cores with different permutations and combinations of three types of trimers which are 2 *hE2*-1 E3BP, 1 *hE2* - 2 E3BP and 3 *hE2*. As now there are more ways in which an inter trimer bridge can be constructed between trimer units (see Fig. 43), it adds another layer of complexity to the architecture of this enzyme complex. Furthermore, two or three such trimers could be detected in a single peak, which further supports inhomogeneous distribution of E3BP and *hE2* proteins in a core surface (refer Section 3.10.3.).



**Fig. 43: Examples of some of the inter trimer associations possible in an isolated *hE2*/E3BP core and when it's reconstituted with *hE1* and *hE3* proteins.**

Although only one type of trimer between *hE2* and E3BP was detected in native MS, each for an isolated core and reconstituted *hPDHc*, the fixed stoichiometry between *hE2* and E3BP of 40:20 means that there must exist trimers with other arrangements between these subunits. Trimers containing various combinations of *hE2* (blue) and E3BP (green) monomers are shown here in making higher order trimer-trimer frameworks. Taken together, as the variation in the inter trimer association is already large, if subsequent trimer units are added to any given trimer-trimer skeleton to build a dodecahedron, the final product will have even larger variation.

## 4. Discussion

### 4.1. Structure of *hE2/E3BP* core and its dynamics

Pseudo-atomic models for *hE2* and *E3BP* monomers could be calculated from cryo-EM map of full-length *hE2/E3BP* core at 6.3 Å (see Fig. 33). They are similar to crystal structure of truncated PDHc cores from *A. vinelandii* and *E. coli* (Mattevi, Obmolova et al. 1992, Wang, Nemeria et al. 2014) (see Fig. 34). However, it showed differences with the only available pseudo-atomic model of human tE2 core (Yu, Hiromasa et al. 2008), exactly at the regions (such as H1 helix and internal hairpin loop) where it showed differences to PDHc core from bacterial sources.

Structurally, both *hE2* and *E3BP* subunit's core forming IC domains share similar architecture (see Fig. 33). A small  $\beta$  strand at the N-terminus of H1 helix could be observed to form anti-parallel beta sheet with S3 and S2  $\beta$  strands of a neighboring monomer in a counter clockwise fashion, which might be important for trimer unit formation. This was also observed in crystal structure of *E. coli* tE2 core but not in *A. vinelandii* tE2 core, where equivalent region has an unstructured loop. In a human tE2 pseudo atomic model, N-terminus  $\beta$  strand are absent and instead flexible loop is present. In a similar manner, our models consist of an internal hairpin loop that is much more ordered than human tE2 model (see Fig. 34C). This loop was speculated to be disordered compared to E2 from prokaryotic sources, in order to interact specifically with equivalent hairpin loop of *E3BP*. But this hypothesis can be partially discredited, as these hairpin loop regions in our *hE2* and *E3BP* models consist a  $\beta$  strand that is absent in human tE2 model. But it's very small compared to long  $\beta$  strands present at the same area of prokaryotic PDHc E2 (see Fig. 34B). Consistent with this, the H1 helix that forms the outer face of the trimer unit in our models bear more resemblance with prokaryotic E2s than human tE2 model, where the orientation of this helix is slightly rotated along the core surface (see Fig. 34A).

Although, in a present work, structures of *hE2* and *E3BP* were calculated from much improved EM density map, the discrepancy with the human tE2 model can also be due to following two factors: 1) our models were derived from full length *hE2/E3BP* core while human tE2 model was from truncated *hE2* core (EM density map at  $\sim 9$  Å) and 2) our sample preparation included GraFix step and as well as addition of 2 mM CoA, to stabilize the cores. Nevertheless, during the course of this thesis work, Hezaveh and colleagues independently showed via MD simulation that the atomic model of human tE2 core requires correction as it was deemed unstable and inaccurate at the same regions explained above (Hezaveh, Zeng et al. 2016). Our

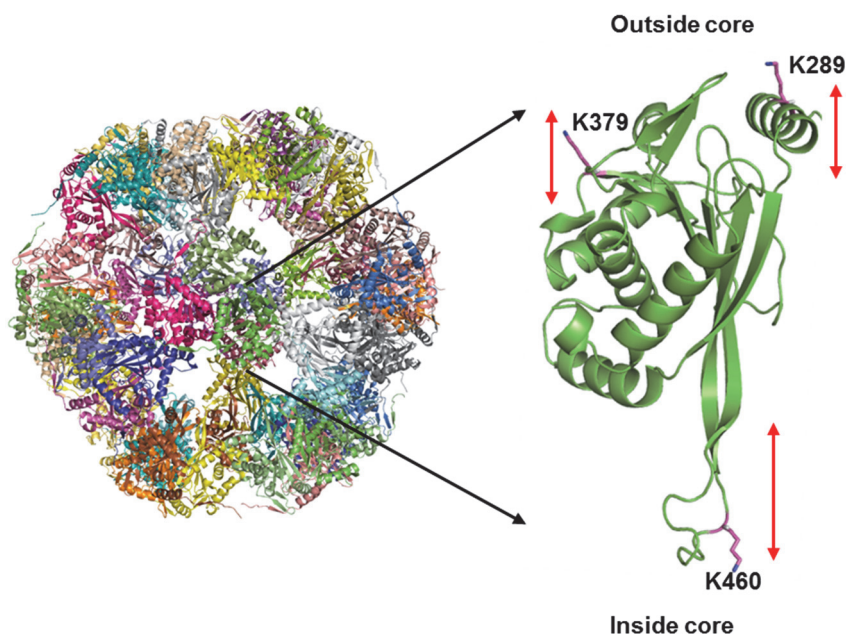


atomic models of *hE2* and *E3BP* thus provide experimental back up to the MD simulated models and thus, the human *tE2* model needs appropriate revision.

Taken all together, *hE2* and *E3BP* core forming IC domains appear structurally similar to that of bacterial PDHc E2, even though they organize into 24meric cube while eukaryotic version organizes into 60meric dodecahedron. The difference might lie at the inter trimer bridges where it was hypothesized to have subtle differences in orientation of C-terminal helices (Izard, Aevarsson et al. 1999). The direct comparison of orientations of these bridges in human PDHc core and bacterial versions was refrained at the current resolution of cryo-EM map, thus providing scope for improving resolution via cryo-EM or X-ray crystallography in the future.

Nevertheless, the obtained pseudo atomic models of *hE2* and *E3BP* were used to visualize cross-links detected in the crosslinking MS of *hPDHc*, in order to understand dynamics within the core. The crosslinked lysine residues by BS3 cross-linkers provided us with maps of conformational fluctuations in the *hPDHc* core. At priory, we expected it to be highly flexible in order to accommodate breathing process. The benchmark for distance between crosslinking lysine residues was taken as  $\sim 27 \text{ \AA}$  between two C $\alpha$  atoms, based on crystal structure of RNA polymerase II (Rappsilber 2011). This distance accounts for length of BS3 (11.4  $\text{\AA}$ ), lysine side chain (6-6.5  $\text{\AA}$ ) and possible error in crystal structure.

Four crosslinks were considered to represent large dynamics of the core: K450-K379 ( $\sim 47 \text{ \AA}$ ), K460-K289 ( $\sim 54 \text{ \AA}$ ) and K460-K79 ( $\sim 38 \text{ \AA}$ ) for *E3BP-E3BP* crosslinks and K295-K547 ( $\sim 35 \text{ \AA}$ ) for *E3BP-hE2*. The latter two were considered to be normal, as small movement of domains could yield such crosslinks while the first two crosslinks have very high crosslinking distances, which might be false positives. However, since K379 and K289 residues are located near the surface of the core while K450 and K460 are at the inner hairpin loop region of *E3BP* (see Fig. 44), at least a possibility of false positive due to the inter *hPDHc* particle interaction or aggregation can be safely ignored. On the contrary, it might be an indicator of a high flexibility of the hairpin loop to partially flip upward towards the core surface. Additionally, the secondary structural elements to which those surficial lysine residues are part of (For e.g. H1 helix - K289), might also move towards the central cavity of the core. This provides us a previously unidentified mode of dynamics in the core where at least an *E3BP* monomer could be stretched and compressed along the length of the monomer. Equivalent crosslinks in *hE2* was not observed primarily because H1 helix of *hE2* core forming domain lacks any lysine residues.



**Fig. 44: Dynamics along the length of E3BP.**

A representative E3BP monomer is focused on a 60meric pentagonal dodecahedral *hPDHc* core, such that surficial part is on the top and the innermost part in the cavity on the bottom. K379 and K289 are located at surficial structural elements and K460 is situated at the inner hairpin loop. These structural elements can come to close proximity to allow crosslinking. This is only possible if these structural elements can fluctuate along the length of the monomer (red arrows), which is orthogonal to inter trimer bridge motion (breathing).

These dynamics could also imply that the structural models we generated represent only one of the many major conformations of the core. However, its contribution in size variation due to breathing would largely depend on pulling and pushing of H1 helix only, as the flipping of inner hairpin loop at the cavity of dodecahedral *hPDHc* core won't effect its overall size. Furthermore, assigning crosslinks to probe breathing, a process of relaxation and contraction of inter trimer bridge was not possible with this approach. This is because it's impossible to credit an inter chain crosslinking to inter trimer units as the distance between crosslinking partners were usually smaller when crosslinking was considered to occur between two subunits in a same trimer, and thus more feasible. Nonetheless, this cannot exclude existence of inter trimer crosslinks completely.

#### 4.2. Organization of the *h*PDHc core and its implications on its function

One of the two models that's under debate for the stoichiometry between *h*E2 and E3BP subunits in the *h*PDHc core are 40:20 and 48:12 substitution model. In an absence of direct evidences such as structure at atomic resolution or absolute quantification MS, the best bet have been to quantify the number of *h*E3 proteins that bind to the core at full saturation. To do this, techniques such as ITC, SAXS and analytical ultracentrifugation (AUC) were employed by several groups. But interestingly, they arrived at different results.

To understand this conundrum however, another contradiction needs to be addressed first. And it is about the mode of binding of E3BP to *h*E3 protein. According to the crystal structure of *h*E3 with E3BD, only one E3BD can make contact with a *h*E3 dimer, albeit with two different orientations, but never with both in a same complex. This means one E3BP will bind to only one *h*E3 protein, and for the two models in question, *h*PDHc core will bind either 20 or 12 *h*E3 proteins. Brautigam and colleagues showed via ITC and AUC that 20 *h*E3 proteins binds to *h*PDHc core (Brautigam, Wynn et al. 2009). Problem arose when other groups used similar methods plus SAXS and showed only around 10 or 6 *h*E3 proteins bind to *h*PDHc core at full occupancy, and argued against the crystal structure (Vijayakrishnan, Callow et al. 2011). According to them, two E3BP subunits can bind to a *h*E3 protein in solution. Furthermore, recombinantly expressed cores bind to 10 *h*E3s while native cores bind to only 6, possibly showing lesser number of E3BPs present in a native system.

Without scrutinizing the methods used by those authors to purify proteins, especially the *h*PDHc core in those studies, we had observed before that E3BP proteins are extremely labile (MSc thesis of Sabin Prajapati). When harsh cell disruption techniques like fluidizer or French press are used followed by column chromatography (ion exchange and gel filtration), where protein has to spend considerable amount of time in a column matrix, the fraction of E3BP proteins leave the core. It was also evident during the follow up negative staining EM that most of the cores obtained such way were partially broken. Moreover, native PDHc cores were obtained from native bovine PDHc in earlier works by performing discontinuous sucrose gradient ultracentrifugation at very high salt concentration of 2 M NaCl, which must have also removed fraction of E3BP from the complex along with E1 and E3 proteins (Vijayakrishnan, Callow et al. 2011). Possibly, this is the reason why they could only find 6 E3 proteins binding to such cores.

In contrast to these, the *h*PDHc core that was recombinantly expressed during this thesis work was purified in a gentler way. Firstly, electric mortar was used to mechanically grind *E. coli* cells (with desired protein overexpressed). Performing PEG and protamine sulfate precipitations followed by linear sucrose gradient ultracentrifugation, the *h*PDHc core were purified. The cores obtained this way were far more reliable as nearly 20 E3 proteins could bind to the core in at least 3 different purification batches, using ITC (see Fig. 24). The binding affinity was very tight ( $K_d < 10$  nm). The negative staining EM of these cores showed far lesser number of broken particles and also the best ever EM map of any PDHc core could be calculated via cryo EM (see Fig. 32).

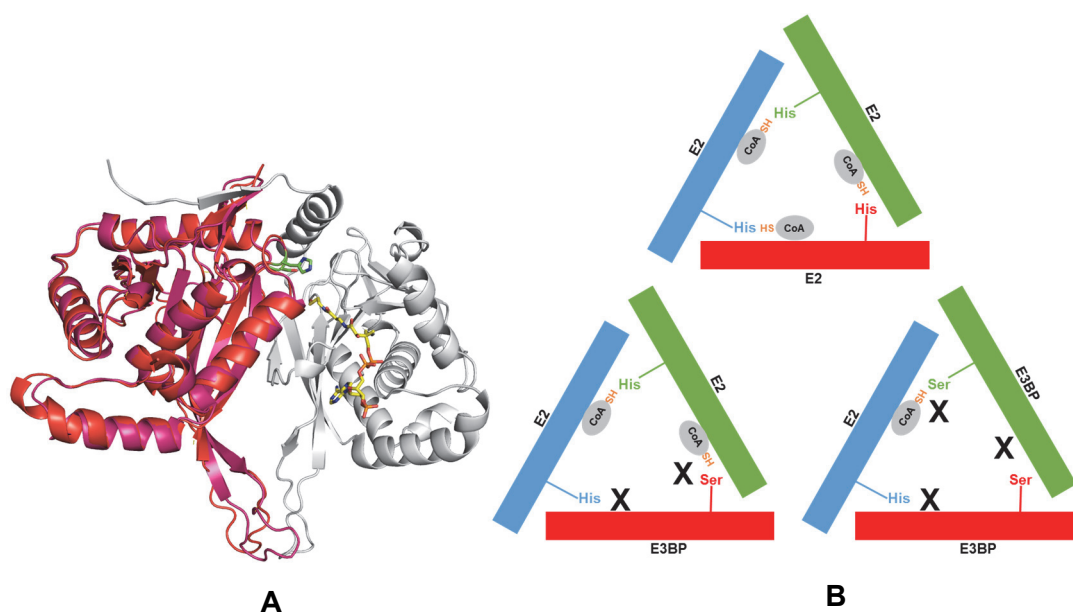
The fitness of the *h*E2/E3BP was judged by its ability to bind CoA, also via ITC. Past experiments (in the same lab by Kathrin S. Tittmann and MSc thesis of Sabin Prajapati) on an artificial *h*PDHc core that was devoid of E3BP, showed that such cores could bind to nearly 60 CoA molecules, meaning each *h*E2 subunit binding to one CoA molecule. In contrast, the *h*E2/E3BP cores used in this thesis work, could only bind around 40 CoA molecules (see Fig. 19). This is interesting because these cores contain 40 *h*E2 and 20 E3BP subunits, which possibly indicate that E3BP do not bind to CoA. Although it is assumed that E3BP do not possess any acetyl transferase activity like *h*E2, principally due to substitution of catalytic histidine with serine, its ability to bind substrate 'CoA' has not been investigated before. Although the binding isotherm in ITC is not a desirable sigmoidal but rather hyperbolic (for both *h*E2/E3BP and *h*E2) which is observed for low affinity system ( $K_d$  of  $\sim 88$   $\mu$ M and 40  $\mu$ M respectively for *h*E2/E3BP and *h*E2), the calculation of stoichiometry may not be as inaccurate as supposed (Turnbull and Daranas 2003). Also, the data obtained for stoichiometry between CoA and the cores were highly reproducible.

As the primary structure of the core forming IC domain of *h*E2 and E3BP are similar (sequence similarity of 50 %), E3BP is supposed to have similar tertiary structure as *h*E2 (Hezaveh, Zeng et al. 2016). At monomeric or trimer level, it was shown that both cubic and pentagonal dodecahedral PDHc cores are similar with the slight difference in the orientation of the trimer being the key in formation of either the cube or pentagonal dodecahedron. This means *h*E2 and E3BP can be directly compared with the available crystal structure of bacterial E2 cores from *A. vinelandii* and *E. coli* at monomeric or trimeric level. Doing so, one thing which is quite apparent is that the catalytic histidine of *h*E2 or its substitution serine in E3BP, lie at the active site of neighboring monomer in a same trimer (see Fig. 45A). A CoA molecule bind to the long active site tunnel from inside of the core, with its thiol group at close proximity to

this catalytic histidine (*hE2*) or serine in E3BP. If E3BP is unable to bind to CoA, only 40 sites remain in a core which can bind to CoA. But since, E3BP lacks catalytic histidine residue as well, which is situated at the active site of neighboring monomer, even if that neighbor is *hE2*, acetyl transfer reaction cannot occur there.

Hence, in a trimer containing 3 *hE2* subunits like in bacterial PDHc, all the active sites for acetyl transfer are operational. Whereas, for eukaryotic PDHc, a trimer containing 2 *hE2* and 1 E3BP has only one functional active site out of three while those containing 1 *hE2* and 2 E3BP have no functional active sites (see Fig. 45B).

As there is no experimental report on the distribution of 20 E3BP subunits in a *hPDHc* core, it is not confirmed if all the 20 trimers are same (2 *hE2* + 1E3BP type) or a mixture of 3 *hE2*, 2 *hE2* + 1 E3BP, 1 *hE2* +2 E3BP and 3 E3BP types of trimers. Based on negative staining EM of *hE3-hE2/E3BP* sub-complex (see Fig. 36), it seems *hE3* proteins are clustered together in few spaces around the core and not uniformly distributed. This indicates that corresponding E3BP subunits must also be clustered together rather than homogenously distributed in the core. Separately, via native MS (Table 13) it could be shown that 1 *hE2* + 2 E3BP type of trimers are present in in-vitro reconstituted *hPDHc* but not for isolated *hE2/E3BP* core. Such trimers could be also detected in a group of as much as 2 to 3. It raises the question if the E3BP subunits change their distribution in a core once *hE3* proteins are around i.e. uniform distribution to clustering.



**Fig. 45: CoA binding to E2.**

(A) *hE2* dimer within same trimer (red and grey as monomers) is superposed to an E3BP monomer (pink). A CoA molecule (yellow, coordinates from PDB 1EAD) is bound to grey *hE2* monomer. A catalytic histidine (green) is contributed from a neighboring *hE2* monomer (red) while at the same position E3BP contributes serine. (B) A cartoon representation of different scenarios in a trimer containing different number of E3BP (black cross marks denote nonfunctional active site/binding site). When no E3BP is present in a trimer, all the active sites of E2 are functional. Adding just one E3BP in a trimer, brings down operational sites to mere one, as E3BP neither binds CoA nor contributes catalytic histidine to its neighbor. Adding two or more E3BP in a trimer, kills all the potential active sites.

Nevertheless, such clustering behavior of E3BP in a PDH complex might have positive functional ramification. As described before, presence of 1 *hE2* + 2 E3BP or 3 E3BP kind of trimers would completely abolish the activity of acetyl transferase site in that particular trimer, but the number of such trimers is limited by the copy number of E3BP in a core. It means for every trimer with 1 *hE2* + 2 E3BP composition, somewhere in a core there must exist a 3 *hE2* kind of trimer, so as to balance the overall ratio between *hE2* and E3BP to 2:1 (40:20). Similarly, to balance 3 E3BP kind of trimer (if it exists), there must exist 2 trimers of 3 *hE2* type.

Hence, in *hPDHc*, there exists possibility for separate islands dedicated for segregating E1/E2 reactions and E3 reaction. A cluster where *hE2* is in majority can focus mainly on E1 and E2 reaction whereas a cluster where E3BP is in majority can perform E3 reaction. This is unlike bacterial PDHc where there is no E3BP and all three reactions are to be performed by E2 lipoyl domains. However, a recent study on *E. coli* PDHc suggested a mechanism of inter chain acetyl transfer where by, an E2 monomer provides an active site for acetyl transferase reaction, a lipoyl domain

from neighboring E2 visits this active site after going through E1 reaction, and probably third E2 is involved in E3 reaction (Song and Jordan 2012). Similar separation of duties might have been conserved for PDHc from higher organisms that have E3BP.

### **4.3. Functionally nonequivalent lipoyl domains and their mode of networking**

Although, lipoyl domain of eukaryotic E3BP is acetylated during PDHc reaction (Demarcucci, Hodgson et al. 1986), it might not be its principal responsibility to visit catalytic site of E1. Rather, reduced dihydrolipoyl-E2 (after E1 and E2 reaction) might exchange the reducing equivalent with lipoyl domain of E3BP. The oxidized lipoyl-E2 can go back to active site of E1 and reduced lipoyl-E3BP can go to active site of E3 for re-oxidation. Extensive networking between lipoyl domains for exchanging acetyl group and reducing equivalent is the hallmark of multiple random coupling mechanism (MRC) model for all PDHc (Stepp, Pettit et al. 1981, Hackert, Oliver et al. 1983, Hackert, Oliver et al. 1983).

During crosslinking mass spectrometry experiment on in-vitro reconstituted *h*PDHc, we observed low number of crosslinks between lipoyl domain of E3BP and *h*E1. In comparison, both the lipoyl domains of *h*E2 showed extensive crosslinks with *h*E3. And, as expected, the lipoyl domains of *h*E2 made numerous crosslinks with *h*E1 and lipoyl domain of E3BP did the same with *h*E3 (see Fig. 41). All in all, the lipoyl domains appear redundant if only their capacity to come close to *h*E1 and *h*E3 is considered. However, the lipoyl domains of *h*E2 and E3BP show certain dichotomy on how much space they can explore on the surface of *h*E1 and *h*E3. Based on crosslinking MS data, it can be assumed that lipoyl domains of *h*E2 are free enough to take over the function of lipoyl domain of E3BP, but the lipoyl domain of E3BP cannot do reverse without compromising the overall activity. In the follow up to previous discussion (refer Section 5.2.), this provides another basis for a model where *h*E2 lipoyl domains focus mainly on E1 and E2 reaction while E3BP lipoyl domain focuses on E3 reaction. This is very much in line with the two experiments carried out to specifically remove lipoyl domains of mammalian PDHc, either from E3BP or E2, at a time. When a lipoyl domain of E3BP was removed, nearly 80 % of the overall activity was retained (Neagle and Lindsay 1991). But, when the lipoyl domains from E2 were proteolytically cleaved, only 10-20 % of the activity was left (Rahmatullah, Radke et al. 1990).

We also observed crosslinks between lipoyl domain of E3BP (K-120) with lipoyl domains of *hE2* (K-114 and K-241). As expected, the lysine residues, which are the lipoylation sites, do not make any crosslinks with other lysine residues. When the available structure of lipoyl domains of *hE2* is compared to the homology model of lipoyl domain of E3BP, such that crosslinking lysine residues are near to each other while lipoylation sites come to close proximity, it appears that these lipoyl domains come into contact with each other at certain orientation only i.e. head of lipoyl-E3BP with a face of lipoyl-E2 (see Fig. 40). Any other orientations look less probable as no other crosslinks were observed even though multiple lysine residues are present at other parts of lipoyl domains. Since, movement of lipoyl arms is assumed to be highly random, this provides one of the first ever experiment-derived restraint to this randomness. Furthermore, this restraint appear to hold up for lipoyl domains of *hE2* and E3BP even when they are moving around at different active sites, meaning how so ever the lipoyl arms move, there is only one dominant orientation for lipoyl domain interaction.

And if a possible dichotomy in responsibility for lipoyl domains of E3BP and *hE2* is also added to the overall picture, the whole functioning of *hPDHc* now looks a product of careful planning and not chaos. Hence, this demands a further interrogation on what else about PDH complex is not random.

#### **4.4. Effect of CoA binding to the *hPDHc*: a substrate dictates large conformational changes**

As already mentioned before, the investigation of CoA binding to *hPDHc* started after observing an increased integrity of *hPDHc* cores while doing EM and AFM experiments (MSc thesis of Sabin Prajapati). As the crystal structure of CoA bound E2 core of *A. vinelandii* PDHc (only available CoA–core complex in literature) do not deviate significantly from unbound version (Mattevi, Obmolova et al. 1992, Mattevi, Obmolova et al. 1993), the origin of this stability begged explanation. We had two hypotheses in the beginning of the investigation. First one was the interference with breathing motion of a core by CoA, possibly directing the cores to become more compact. Second one was increased internalization of lipoyl domains based on a report of increased affinity of lipoyl domains to the core in the presence of CoA (Kato, Wynn et al. 2006).

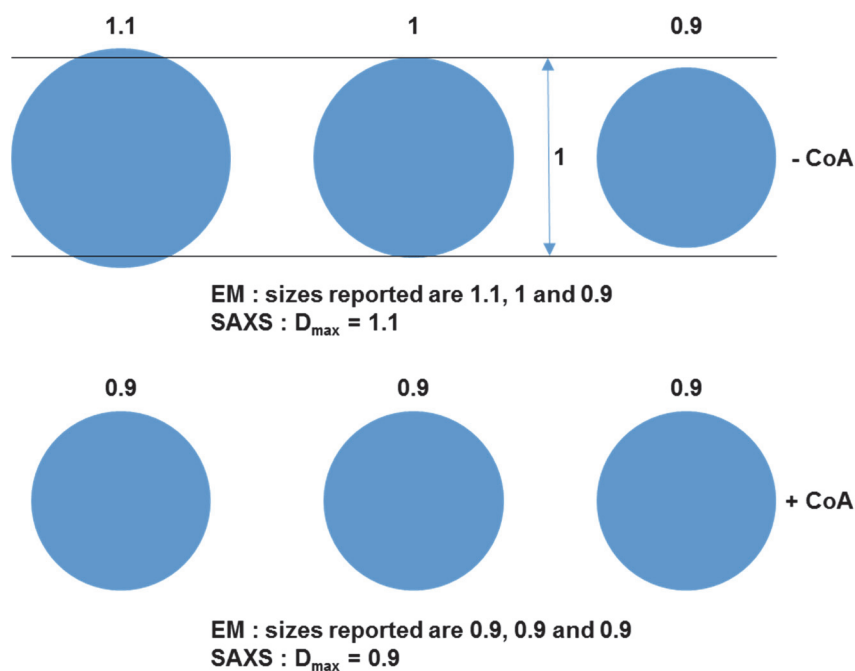
The first hypothesis could have been effectively tested if we had achieved good quality sample for electron microscopy. But on my hands, the *hPDHc* cores and its variants



did not survive until EM step in a good state, especially without CoA (see Fig. 29). Since, their stability increased post GraFix, major attention was given to such samples and at the end we were successful in obtaining the best ever cryo-EM map for *h*PDHc core at  $\sim 6$  Å. Unfortunately, this was possible at the cost of core dynamics as the particles after chemical crosslinking by glutaraldehyde (GraFix) did not show size variation as expected for breathing PDHc core.

Nevertheless, SAXS was performed on *h*PDHc cores and its variants, in the presence and absence of saturating concentration of CoA (2 mM). All the investigated cores showed CoA induced compaction in their size by 4-5 nm of  $D_{\max}$  (refer Section 3.6). Similar effect was also observed during STED experiments when the CoA binding changed the size distribution of SNAP-*h*E2 cores towards lower size groups (refer Section 3.9.).

If the size of the post-GraFix core (+ CoA) is  $\sim 25$  nm based on neg-staining EM micrographs and cryo-EM derived model (only core forming part is visible), and if it's assumed that CoA stops breathing by locking smallest size conformation, the same sample in absence of CoA should have  $\sim 29$ -30 nm size considering 20 % size variation during breathing. This is same as the value obtained for truncated E2 core by SAXS (see Table 5) and small angle neutron scattering (SANS) (Vijayakrishnan, Kelly et al. 2010). Thus, it looks logical to explain the CoA induced increase in *h*PDHc core stability on the basis of core compaction, possibly due to stopping of breathing motion.



**Fig. 46: CoA induced *h*PDHc core compaction and comparison with EM.**

A truncated *h*PDHc core during breathing can be imagined to have a difference in sizes (1.1, 1 and 0.9) between the largest and the smallest particles of nearly 20 % of the average size. For this sample, when SAXS is performed, a  $D_{\max}$  of 1.1 will be calculated. In the presence of CoA, if breathing stops and the cores are locked at smallest size conformation, SAXS will report  $D_{\max}$  of 0.9, which will make the difference between the  $D_{\max}$  values as 20 % of the average size of – CoA state. Adapted from (Zhou, Liao et al. 2001)

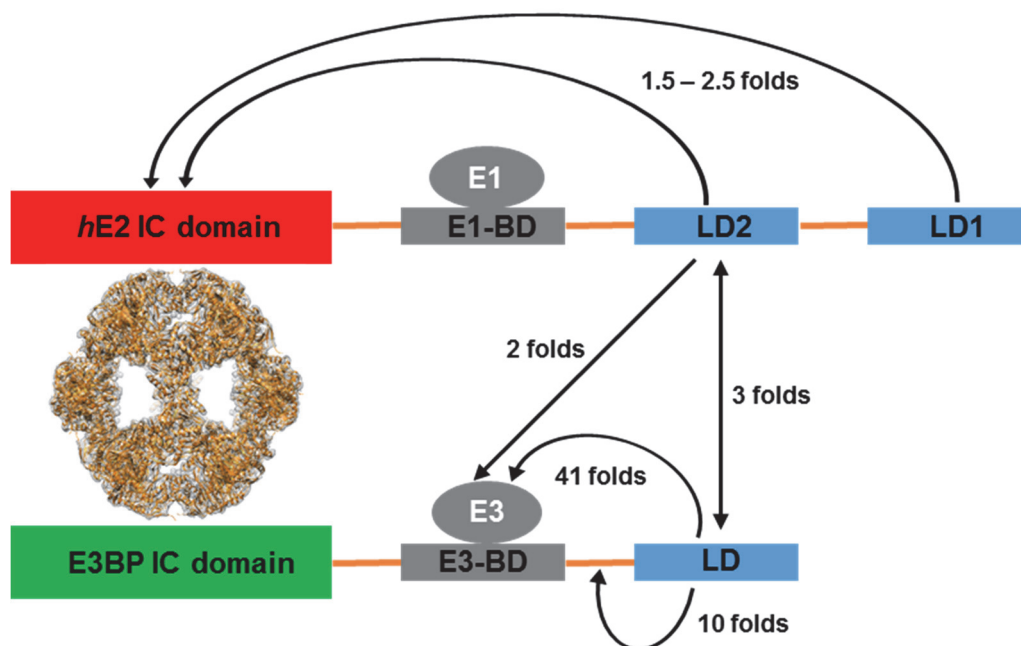
As fitting as it seems, this model though lacks one crucial parameter and that is lipoyl arm. Since highly flexible lipoyl arms are invisible in EM studies, the size difference between largest and smallest core particles during breathing, in afore mentioned literature is completely ignoring them. But during SAXS, the size estimate is made out of the full-length core and not just core-forming domain alone. Although same level of compaction was observed for proteolytically derived truncated E2 core, only single data set could be measured and additionally, they were deemed unsuitable for structural studies after conducting EM experiments due to overwhelming presence of broken particles (see Figure 30 and Table 5). Hence, other techniques were explored to see if lipoyl arms are internalized towards cores upon binding to CoA, which should also result in core size compaction observed in SAXS.

Quantification of crosslinks made by all three lipoyl domains of *h*E2 and E3BP in the presence and absence of CoA (see Table 12) detected in crosslinking MS revealed that upon binding to CoA, the lipoyl domains of *h*E2 make as much as 2.5 folds more crosslinks with the core forming domain's residues. This clearly indicates the increased movement of lipoyl arms towards the inner core and would contribute to

CoA induced size compaction, at least partially. However, the lipoyl domain of E3BP does not significantly change its crosslinking to the core. Its interaction with linker region between E3BD and E3BP lipoyl domain though increased by 10 folds, possibly indicating that E3BP-lipoyl arm also folds back when CoA is around.

Furthermore, inter-lipoyl domain crosslinks also increased by 3 folds, signifying increased crosstalk between lipoyl domains upon CoA binding. And the crosslink between lipoyl domain of E3BP with K-267 of *hE3* (near lipoyl entrance site) amplified by 41 folds and *hE2-LD2* with K-104 of *hE3* (another lysine near lipoyl entrance site) increased roughly by 2 times. Similar observations could not be made in regards to *hE1* protein and lipoyl domains of E3BP, but for *hE2-LD2*, it was a mixed result as crosslinks with lysine residues near active site i.e.  $\alpha$ -K-83 increased by 2 folds while with  $\beta$ -K-68 decreased by half.

Summing up these observations, it can be stated that upon binding of CoA to the *hPDHc* core, more lipoyl arms of *hE2* folds back towards the inner core, possibly for acetyl transfer reaction at *hE2* catalytic site. The mobility of lipoyl domain of E3BP increases such that they can increase cross talk with *hE2-LD2* (for potential exchange of reducing equivalent) and also visits E3 active site much more frequently. Thus, it appears that the CoA binding to *hE2* primes the PDH complex for the subsequent events in the PDHc multi-step reaction, which are E2 and E3 reactions. And at least from the data on hand, it can be ruled out that CoA binding to the core would change *hE2*-lipoyl domains movement to *hE1* as significantly as lipoyl domain of E3BP to *hE3* (Table 12). This is a key finding as for the first time, a substrate of PDHc is observed to dictate such a dramatic change in conformational landscape of lipoyl arms with an explainable functional implication.



**Fig. 47: Net effect of CoA binding to *hPDHc*.**

The lipoyl domains of *hE2* can make more crosslinks (1.5-2.5 folds) with the core forming IC domain once the CoA binds, providing alternative explanation to CoA induced core compaction. One of the functional consequences of this could be increased mobility of lipoyl arm of E3BP, which appears to be now virtually free to move everywhere. In CoA bound state of *hPDHc*, E3BP lipoyl domain significantly makes more contact with *hE3* active site region, E3-BD, linker between E3BD and LD of E3BP and LD2 of *hE2*. (Lipoyl domain movement is shown with black arrows, domains of *hE2* and E3BP and *hE1* and *hE3* proteins in colored solid bodies and the linkers between domains in orange).

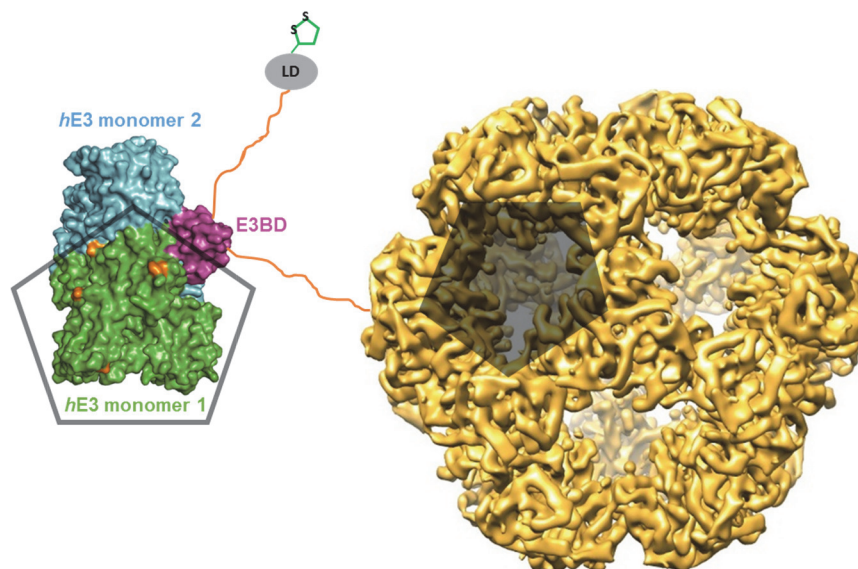
So far from SAXS and STED, we observed CoA induced size compaction, which could be due to stopping of breathing, internalization of lipoyl arms or both. From mass spectrometric analysis, it could be understood that internalization of lipoyl arms of *hE2* will contribute in CoA induced size compaction of *hPDHc* core, at least partially. Although the contribution of breathing on CoA induced size compaction could not be conclusively estimated, using dSTORM or related techniques to check the fluorescence signal intensity vs radial position in the particle should provide a direct glimpse of lipoyl arm internalization. If CoA increases the frequency of lipoyl arm internalization, then the fluorescence signals from N-terminal eGFP or organic dye bound anti-eGFP nanobody should be significantly higher near the central part of the core. But if the intensity remains unchanged after CoA addition, then the observed size compaction could be attributed to stopping of breathing. However, the mixed scenario cannot be out ruled. Due to time constraint of the thesis, such analysis could not be performed with the dSTORM data obtained for eGFP-*hE2* cores.

#### 4.5. Active site layers in *hPDHc*

With the confusion about the position and number of E3BP in *hPDHc* core, previous works have positioned varying number of *hE3* proteins at either outside the core or at the pentagonal openings (Stoops, Cheng et al. 1997, Zhou, McCarthy et al. 2001, Hiromasa, Fujisawa et al. 2004). The position of *hE1* protein is however not disputed as they have been clearly shown to form a cage from outside the core (Wagenknecht, Grassucci et al. 1991, Zhou, McCarthy et al. 2001). In the present work, we observed that *hE3* protein might not form a regular outer shell like *hE1*, as the size of *hE3:hE2/E3BP* sub-complex was much smaller than *hE1:hE2/E3BP* sub-complex or *hPDHc* (refer Section 4.2.). This was in contrast to similar sub-complexes in *Bacillus Stearothermophilus* where both E1 and E3 can form similar outer shells (Milne, Shi et al. 2002, Milne, Wu et al. 2006). Furthermore, while comparing electron micrographs of *hE3:hE2/E3BP* sub-complex and *hE2/E3BP*, we observed *hE3*'s inhomogeneous distribution at the core surface (see Fig. 36).

Another key finding regarding location of *hE3* came from crosslinking mass spectrometry. In addition to making crosslinks with the peripheral domains of *hE2* and E3BP as well as *hE1* and surficial core residues, *hE3* could also make crosslinks with lysine residues of inner hairpin loop of *hE2* and E3BP. Even if we take account of high flexibility of this hairpin loop to flip upward towards core surface (refer Section 4.1.), these crosslinks would only be possible if *hE3* protein partially moves inside the core via pentagonal openings. The lysine residues of *hE3* that makes crosslinks (K267, K277, K430 and K445) with the innermost part of core, lie at a same face of *hE3* monomer. As *hE3* is a homo dimer, possibly a monomer of *hE3* would get inside the core with E3BD acting as a hinge (see Fig. 48). The other *hE3* monomer could simply lie at the core surface.

The dimension of *hE3* dimer is  $\sim 6 \times 9$  nm while the widest size of the pentagonal openings in *hE2/E3BP* core model is  $\sim 6.5$  nm. This roughly allows a monomer to snug into the pentagonal openings of the core via its smallest edge. However, during breathing, the size of the openings can be imagined to increase subtly enabling easier accommodation of *hE3*. This does not appear feasible for *hE1* protein ( $\sim 8 \times 10$  nm), which is bulkier than *hE3*, though a single crosslinking could be identified between *hE1* ( $\alpha$ -K77) and inner hairpin loop of *hE2* (see Table 10).



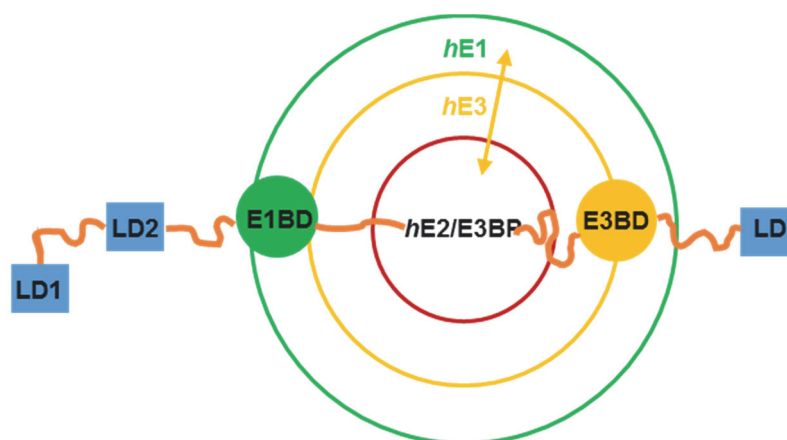
**Fig. 48: Docking of *hE3* in a pentagonal opening of *hE2/E3BP* core.**

A black shadow represents one of the twelve pentagonal openings in *hE2/E3BP* core. A linker connects the core with the E3BD (pink) while another linker connects E3BD to N-terminal lipoyl domain (LD). E3BD associates with the *hE3* at the dimer interface, such that one monomer (green) can enter the pentagonal opening (grey). Inner hairpin loop of *hE2* and E3BP core forming domains can make crosslinks with four different lysine residues in *hE3* (shown in orange). The structure of *hE3*:E3BD is adapted from PDB 1ZY8.

Thus, the observations made with negative staining EM in combination with crosslinking MS, reveals that the position of *hE3* might not be fixed in space but rather dynamic. This accommodates conflicting findings of previously mentioned works and in addition, gives an indication of three-dimensional separation of *hE1*, *hE2* and *hE3* catalytic sites. A working model can be thought of where *hE1* active sites are located at the exterior of a sphere and *hE2* active sites at the interior, with *hE3* proteins free to fluctuate between these layers (see Fig. 49). But it also must be taken into consideration that in an *hPDHc*, E3BP-*hE3* and *hE2-hE1* forms islands in a core where one of the two entities would be in majority (refer Section 4.2). While the inhomogeneous distribution of E3BP and *hE2* in a core can be explained in context of possible separation of E1/E2 reactions with E3 reaction, the need for *hE3* to move between outer shell and core cavity needs further clarification.

One possibility could be that after acetyl transfer reaction at the core, the lipoyl domains of *hE2* will find *hE3* immediately for its re-oxidation before going to *hE1* for fresh round of reaction (see Fig. 1). However, this would make sense only for lipoyl domains of *hE2* located at islands where *hE2* are dominantly distributed. In the islands where E3BP are in majority, acetyl transfer reaction does not occur. And perhaps

these are the regions where *hE3* is at periphery where mainly lipoyl domain of E3BP would visit for re-oxidation of lipoamide cofactor.



**Fig. 49: Radial separation of active sites in *hPDHc*.**

Active sites of *hE1* (green circle) and *hE2* (brown circle) are well separated in *hPDHc*, whereas *hE3*'s position (golden circle) appears to vary between *hE1* and *hE2* active site layers. The presence of *hE3* at the core surface might enhance re-oxidation of lipoyl domains of *hE2* whereas distant *hE3* might have a role in re-oxidation of lipoyl domain of E3BP. However, these situations might not be exclusive as all the lipoyl domains of *hE2* and E3BP are able to visit all the active sites and they are able to sustain the overall activity in the absence of each other, albeit with different efficiency.

## 5. Summary

Our structural models for *hE2* and E3BP, which we calculated from the best ever cryo-EM density map of human PDHc core till date, are similar to those from prokaryotic PDHc E2 at monomeric and trimeric levels. These models deviate from the observation made regarding differences between human tE2 core pseudo-atomic structure with *A. vinelandii* PDHc E2 crystal structure (Yu, Hiromasa et al. 2008). Furthermore, we have now provided experimental backup to the recent MD simulation based claims of Hezaveh and colleagues (Hezaveh, Zeng et al. 2016) about inaccuracy of human tE2 core structure. In addition to inter trimer breathing motion in PDHc core described in 2001 by Zhou and colleagues (Zhou, Liao et al. 2001), we have identified another potential mode of dynamics which operates along the length of core subunits, at least in E3BP (surface $\leftrightarrow$ cavity).

We have also verified 40:20 stoichiometries between *hE2* and E3BP in *hPDHc* core with only *hE2* being able to bind CoA. Since, E3BP is catalytically incompetent as well, the burden of CoA binding and acetyl transfer reaction lies solely on *hE2*. During native MS experiments, we observed trimers of core subunits in *hPDHc* consisting of 1 *hE2* and 2 E3BP. This meant that somewhere in the same core, 3 *hE2* type of trimer should also be present to balance the overall stoichiometry, creating islands where either *hE2* or E3BP are dominant members. Consistently, in a negative staining EM, we observed inhomogeneous distribution of *hE3* densities on the core surface. And crosslinking MS result showed a population of *hE3* proteins entering the cavity of the core. Taken altogether, we arrive at a model where the *hPDHc* is heterogenous in its arrangement of *hE2* and E3BP. They have distinct patches where E3BP-*hE3* or *hE2*-*hE1* are in majority. Additionally, *hE1* and *hE3* also appear to be radially separated in the *hPDHc*, with *hE1* forming a robust outer shell while *hE3* moving between outer shell and core cavity. We think this is all nature's way to segregate E1/E2 reaction from E3 reaction. The reduced ability of lipoyl domain of E3BP to sample space on *hE1* and *hE2* further supports this model.

Although our crosslink MS data backs the multiple random coupling (MRC) model, as all the lipoyl domains in principal could visit all the domains of *hE2* and E3BP as well as peripheral proteins, there is a certain twist. Upon quantification of crosslinks, we discovered that once the core binds its substrate 'CoA', the lipoyl domains appear to concentrate more on subsequent steps of PDHc reaction, which are E2 and E3 reactions. Thus, PDHc function might not always be the product of random movement of lipoyl arms. On the contrary, key steps of its reaction such as substrate binding



drives lipoyl domains toward active sites where next catalytic steps are supposed to happen.

The notion of ultra-flexibility of lipoyl arms also needs a minor revision. We observed only one orientation for the crosstalk between lipoyl domains (head of E3BP lipoyl domain with face of *hE2* lipoyl domains). Since all the lipoyl domains also have bunch of surficial lysine residues at other parts, many different crosslinks must have been detected favoring other orientations, if the lipoyl arms possessed unrestrained translational and rotational flexibility. Hence, despite the flexibility required for lipoyl arms to visit all three active sites and also to communicate with each other, there must be some restrictions in how they move. It's also interesting as it implies that lipoyl domains should always end up in same orientation regardless of their position in the *hPDHc* (near *hE1*, *hE3*, core forming domain or stretched away from all of them).

From SAXS and STED, we could observe reduction in sizes of all *hPDHc* cores and its variants. The size shrinkage falls in the range of difference between smallest and largest size groups during breathing motion of eukaryotic PDHc cores. While it's tempting to attribute CoA induced core shrinkage to the stopping of breathing motion, the contribution from lipoyl arms internalization has to be accounted for. Especially because lipoyl domains could crosslink with inner components of *hPDHc* many fold higher in the presence of CoA. Perhaps, the information of CoA binding is transduced to distant lipoyl arms by changing dynamics of core-forming domain, prompting these long appendages to move inward such that lipoyl domains are prepared for E2 and E3 reactions.

## 6. Outlook

One of the main discoveries made during this work was inability of E3BP to bind CoA. As this has major functional repercussion in how *h*PDHc functions (refer Section 4.2.), it must be investigated why it cannot bind CoA. The two residues predicted as potential discriminant between *h*E2 and E3BP's CoA binding abilities (refer Section 3.1.) should be studied further. Substitutions of amino acids in those positions by the equivalent residues found in *h*E2 should make E3BP able to bind CoA. In the reverse way, mutating these residues in *h*E2 will render it incapable of binding the substrate.

It will also be an interesting experiment to swap inner linkers between *h*E2 and E3BP. If these linkers play any role for positioning *h*E1 at distal part while allowing *h*E3 to move between *h*E1's outer shell and the core cavity (refer Section 4.5.), the swapping will create a reverse situation for these peripheral proteins. Will then the whole complex still hold onto its activity? And what if both *h*E2 and E3BP have same inner linker, either of *h*E2 or E3BP? Will the competition between *h*E1 and *h*E3 now to occupy similar space compromise the activity in human PDHc? In the case of prokaryotic PDHc, where both peripheral proteins bind to the same subunit-binding domain of E2, in a mutually exclusive manner, this competition for space does not seem to matter.

Regarding structural study of the PDH complex, next step should be to purify native PDHc from mammalian sources (bovine/porcine tissues or HeLa cell) and check if they also show inhomogeneous distribution of peripheral proteins. If not, then cryo-EM should be performed to reach at least moderately resolved density map, as at present, there is no structural model for any PDHc. Native MS should be continued and aim to fly a larger part of the core and complex in a spectrometer. Perhaps it will bring to us other surprises as when it detected 2 E3BP – 1 *h*E2 kind of trimer in the *h*PDHc.

As we also discovered that CoA binding changed the preferences of lipoyl domains to focus on E2 and E3 reactions, it is only natural to try the similar set of crosslinking MS experiments in the presence of pyruvate and its analogs. Perhaps, these lipoyl domains would show increased crosslinks with *h*E1 then. A step ahead would be to use quenched flow instrument where holo-PDHc is mixed with pyruvate rapidly and the reaction quenched with desired chemical cross-linkers in a time bound manner. It will be helpful to understand how this enzyme complex works by tracking movements of each lipoyl domains individually, while in action.

We ventured into a field of super resolution fluorescence microscopy to directly demonstrate dynamics in a large protein complex at molecular level. It was partially successful as we could observe shift in size distribution of the core in STED and dSTORM. Sample preparation part was successfully optimized for dSTORM as we could observe homogeneously distributed cores with minimal background. But since the *h*PDHc cores were very soft (refer Section 3.7), they tend to spread over the solid surface like glass. To avoid this, new techniques like cryo-STORM is desirable where sample can be frozen rapidly in a buffer solution. To make data analysis easier, sparse labelling can be done to PDHc cores by using lesser concentrations of dye conjugated nanobody such that only few lipoyl arms are fluorophore labelled.

## 7. References

Afonine, P. V., R. W. Grosse-Kunstleve, P. D. Adams and A. Urzhumtsev (2013). "Bulk-solvent and overall scaling revisited: faster calculations, improved results." Acta Crystallographica Section D-Biological Crystallography **69**: 625-634.

Allen, A. G., R. N. Perham, N. Allison, J. S. Miles and J. R. Guest (1989). "Reductive Acetylation of Tandemly Repeated Lipoyl Domains in the Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli Is Random Order." Journal of Molecular Biology **208**(4): 623-633.

Ambrose, M. C. and R. N. Perham (1976). "Spin-Label Study of Mobility of Enzyme-Bound Lipoic Acid in Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli." Biochemical Journal **155**(2): 429-432.

Anselmi, C., M. Grninger, P. Gipson and J. D. Faraldo-Gomez (2010). "Mechanism of Substrate Shuttling by the Acyl-Carrier Protein within the Fatty Acid Mega-Synthase." Journal of the American Chemical Society **132**(35): 12357-12364.

Argyrou, A. and J. S. Blanchard (2001). "Mycobacterium tuberculosis lipoamide dehydrogenase is encoded by Rv0462 and not by the lpdA or lpdB genes." Biochemistry **40**(38): 11353-11363.

Argyrou, A., J. S. Blanchard and B. A. Palfey (2002). "The lipoamide dehydrogenase from Mycobacterium tuberculosis permits the direct observation of flavin intermediates in catalysis." Biochemistry **41**(49): 14580-14590.

Argyrou, A., G. Sun, B. A. Palfey and J. S. Blanchard (2003). "Catalysis of diaphorase reactions by Mycobacterium tuberculosis lipoamide dehydrogenase occurs at the EH4 level." Biochemistry **42**(7): 2218-2228.

Bates, D. L., M. J. Danson, G. Hale, E. A. Hooper and R. N. Perham (1977). "Self-Assembly and Catalytic Activity of Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli." Nature **268**(5618): 313-316.

Berg, A., J. Vervoort and A. deKok (1996). "Solution structure of the lipoyl domain of the 2-oxyglutarate dehydrogenase complex from *Azotobacter vinelandii*." Journal of Molecular Biology **261**(3): 432-442.

Bergmann, A., D. Orthaber, G. Scherf and O. Glatter (2000). "Improvement of SAXS measurements on Kratky slit systems by Gobel mirrors and imaging-plate detectors." Journal of Applied Crystallography **33**(2): 869-875.

Bertani, G. (1951). "Studies on Lysogenesis .1. The Mode of Phage Liberation by Lysogenic *Escherichia-Coli*." Journal of Bacteriology **62**(3): 293-300.

Blanchet, C. E. and D. I. Svergun (2013). "Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution." Annual Review of Physical Chemistry, Vol 64 **64**: 37-54.

Bradford, M. M. (1976). "Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding." Analytical Biochemistry **72**(1-2): 248-254.

Brautigam, C. A., J. L. Chuang, D. R. Tomchick, M. Machius and D. T. Chuang (2005). "Crystal structure of human dihydrolipoamide dehydrogenase: NAD(+)/NADH binding and the structural basis of disease-causing mutations." Journal of Molecular Biology **350**(3): 543-552.

Brautigam, C. A., R. M. Wynn, J. L. Chuang and D. T. Chuang (2009). "Subunit and Catalytic Component Stoichiometries of an in Vitro Reconstituted Human Pyruvate Dehydrogenase Complex." Journal of Biological Chemistry **284**(19): 13086-13098.

Brautigam, C. A., R. M. Wynn, J. L. Chuang, M. Machius, D. R. Tomchick and D. T. Chuang (2006). "Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex." Structure **14**(3): 611-621.

Brown, G. K., R. M. Brown and H. H. M. Dahl (1990). "The Human Pyruvate-Dehydrogenase E1-Alpha Genes - Location, Gene Organization and Expression." Genetical Research **55**(2): 125-125.

Bukhari, H. S. T., R. P. Jakob and T. Maier (2014). "Evolutionary Origins of the Multienzyme Architecture of Giant Fungal Fatty Acid Synthase." Structure **22**(12): 1775-1785.

Butt, H. J., B. Cappella and M. Kappl (2005). "Force measurements with the atomic force microscope: Technique, interpretation and applications." Surface Science Reports **59**(1-6): 1-152.

Carothers, D. J., G. Pons and M. S. Patel (1989). "Dihydrolipoamide Dehydrogenase - Functional Similarities and Divergent Evolution of the Pyridine Nucleotide-Disulfide Oxidoreductases." Archives of Biochemistry and Biophysics **268**(2): 409-425.

Cech, T. R. (2012). "The RNA worlds in context." Cold Spring Harb Perspect Biol **4**(7): a006742.

Chernyak, B. V. (1997). "Redox regulation of the mitochondrial permeability transition pore." Biosci Rep **17**(3): 293-302.

Ciszak, E. M., L. G. Korotchkina, P. M. Dominiak, S. Sidhu and M. S. Patel (2003). "Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase." Journal of Biological Chemistry **278**(23): 21240-21246.

Ciszak, E. M., A. Makal, Y. S. Hong, A. K. Vettaikorumakankau, L. G. Korotchkina and M. S. Patel (2006). "How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex." Journal of Biological Chemistry **281**(1): 648-655.

Collins, J. H. and L. J. Reed (1977). "Acyl Group and Electron Pair Relay System - Network of Interacting Lipoyl Moieties in Pyruvate and Alpha-Ketoglutarate Dehydrogenase Complexes from Escherichia-Coli." Proceedings of the National Academy of Sciences of the United States of America **74**(10): 4223-4227.

Dahl, H. H. M., R. M. Brown, W. M. Hutchison, C. Maragos and G. K. Brown (1990). "A Testis-Specific Form of the Human Pyruvate-Dehydrogenase E1-Alpha Subunit Is Coded for by an Intronless Gene on Chromosome-4." Genomics **8**(2): 225-232.

Dardel, F., A. L. Davis, E. D. Laue and R. N. Perham (1993). "3-Dimensional Structure of the Lipoyl Domain from Bacillus-Stearothermophilus Pyruvate-Dehydrogenase Multienzyme Complex." Journal of Molecular Biology **229**(4): 1037-1048.

Demarcucci, O. and J. G. Lindsay (1985). "Component-X - an Immunologically Distinct Polypeptide Associated with Mammalian Pyruvate-Dehydrogenase Multi-Enzyme Complex." European Journal of Biochemistry **149**(3): 641-648.

Demarcucci, O. G. L., J. A. Hodgson and J. G. Lindsay (1986). "The Mr-50000 Polypeptide of Mammalian Pyruvate-Dehydrogenase Complex Participates in the Acetylation Reactions." European Journal of Biochemistry **158**(3): 587-594.

Eghiaian, F. and I. A. Schaap (2011). "Structural and dynamic characterization of biochemical processes by atomic force microscopy." Methods Mol Biol **778**: 71-95.

Falconer, R. J., A. Penkova, I. Jelesarov and B. M. Collins (2010). "Survey of the year 2008: applications of isothermal titration calorimetry." Journal of Molecular Recognition **23**(5): 395-413.

Fan, J., C. L. Shan, H. B. Kang, S. Elf, J. X. Xie, M. Tucker, T. L. Gu, M. Aguiar, S. Lonning, H. B. Chen, M. Mohammadi, L. M. P. Britton, B. A. Garcia, M. Aleckovic, Y. B. Kang, S. Kaluz, N. Devi, E. G. Van Meir, T. Hitosugi, J. H. Seo, S. Lonial, M. Gaddh, M. Arellano, H. J. Khoury, F. R. Khuri, T. J. Boggon, S. M. Kang and J. Chen (2014). "Tyr Phosphorylation of PDP1 Toggles Recruitment between ACAT1 and SIRT3 to Regulate the Pyruvate Dehydrogenase Complex." Molecular Cell **53**(4): 534-548.

Fang, R., P. F. Nixon and R. G. Duggleby (1998). "Identification of the catalytic glutamate in the E1 component of human pyruvate dehydrogenase." Febs Letters **437**(3): 273-277.

Frank, R. A. W., J. V. Pratap, X. Y. Pei, R. N. Perham and B. F. Luisi (2005). "The molecular origins of specificity in the assembly of a multienzyme complex." Structure **13**(8): 1119-1130.

Froger, A. and J. E. Hall (2007). "Transformation of plasmid DNA into E. coli using the heat shock method." J Vis Exp(6): 253.

Gopalakrishnan, S., M. Rahmatullah, G. A. Radke, S. Powersgreenwood and T. E. Roche (1989). "Role of Protein-X in the Function of the Mammalian Pyruvate-Dehydrogenase Complex." Biochemical and Biophysical Research Communications **160**(2): 715-721.

Graham, L. D., L. C. Packman and R. N. Perham (1989). "Kinetics and Specificity of Reductive Acylation of Lipoyl Domains from 2-Oxo Acid Dehydrogenase Multienzyme Complexes." Biochemistry **28**(4): 1574-1581.

Green, J. D. F., E. D. Laue, R. N. Perham, S. T. Ali and J. R. Guest (1995). "3-Dimensional Structure of a Lipoyl Domain from the Dihydrolipoyl Acetyltransferase Component of the Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli." Journal of Molecular Biology **248**(2): 328-343.

Guest, J. R., H. M. Lewis, L. D. Graham, L. C. Packman and R. N. Perham (1985). "Genetic Reconstruction and Functional-Analysis of the Repeating Lipoyl Domains in the Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli." Journal of Molecular Biology **185**(4): 743-754.

Hackert, M. L., R. M. Oliver and L. J. Reed (1983). "A Computer-Model Analysis of the Active-Site Coupling Mechanism in the Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **80**(10): 2907-2911.

Hackert, M. L., R. M. Oliver and L. J. Reed (1983). "Evidence for a Multiple Random Coupling Mechanism in the Alpha-Ketoglutarate Dehydrogenase Multienzyme Complex of Escherichia-Coli - a Computer-Model Analysis." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **80**(8): 2226-2230.



- Hammes, G. G. (1964). "Mechanism of Enzyme Catalysis." Nature **204**(495): 342-&.
- Harris, R. A., M. M. Bowker-Kinley, B. L. Huang and P. F. Wu (2002). "Regulation of the activity of the pyruvate dehydrogenase complex." Advances in Enzyme Regulation, Vol 42, Proceedings **42**: 249-259.
- Harris, R. A., M. M. BowkerKinley, P. F. Wu, J. J. Jeng and K. M. Popov (1997). "Dihydrolipoamide dehydrogenase-binding protein of the human pyruvate dehydrogenase complex - DNA-derived amino acid sequence, expression, and reconstitution of the pyruvate dehydrogenase complex." Journal of Biological Chemistry **272**(32): 19746-19751.
- Hay, S. and N. S. Scrutton (2012). "Good vibrations in enzyme-catalysed reactions." Nature Chemistry **4**(3): 161-168.
- Hayakawa, T., T. Kanzaki, T. Kitamura, Y. Fukuyoshi, Y. Sakurai, K. Koike, T. Suematsu and M. Koike (1969). "Mammalian Alpha-Keto Acid Dehydrogenase Complexes .V. Resolution and Reconstitution Studies of Pig Heart Pyruvate Dehydrogenase Complex." Journal of Biological Chemistry **244**(13): 3660.
- Henderson, C. E., R. N. Perham and J. T. Finch (1979). "Structure and Symmetry of B-Stearothermophilus Pyruvate-Dehydrogenase Multi-Enzyme Complex and Implications for Eucaryote Evolution." Cell **17**(1): 85-93.
- Hernandez, H. and C. V. Robinson (2007). "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry." Nature Protocols **2**(3): 715-726.
- Hezaveh, S., A. P. Zeng and U. Jandt (2016). "Human Pyruvate Dehydrogenase Complex E2 and E3BP Core Subunits: New Models and Insights from Molecular Dynamics Simulations." Journal of Physical Chemistry B **120**(19): 4399-4409.
- Hiromasa, Y., T. Fujisawa, Y. Aso and T. E. Roche (2004). "Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 components." Journal of Biological Chemistry **279**(8): 6921-6933.

Horcas, I., R. Fernandez, J. M. Gomez-Rodriguez, J. Colchero, J. Gomez-Herrero and A. M. Baro (2007). "WSXM: a software for scanning probe microscopy and a tool for nanotechnology." Rev Sci Instrum **78**(1): 013705.

Howard, M. J., C. Fuller, R. W. Broadhurst, R. N. Perham, J. G. Tang, J. Quinn, A. G. Diamond and S. J. Yeaman (1998). "Three-dimensional structure of the major autoantigen in primary biliary cirrhosis." Gastroenterology **115**(1): 139-146.

Huang, B., M. Bates and X. W. Zhuang (2009). "Super-Resolution Fluorescence Microscopy." Annual Review of Biochemistry **78**: 993-1016.

Huang, T. H., C. F. Chang, H. T. Chou, J. L. Chuang and D. T. Chuang (2002). "Structure and dynamics of the lipoyl acid-bearing domain of the E2 component of human mitochondria branched-chain A-ketoacid dehydrogenase (BCKD)." Biophysical Journal **82**(1): 439a-439a.

Hutter, J. L. and J. Bechhoefer (1993). "Calibration of Atomic-Force Microscope Tips." Review of Scientific Instruments **64**(7): 1868-1873.

Inoue, H., H. Nojima and H. Okayama (1990). "High efficiency transformation of Escherichia coli with plasmids." Gene **96**(1): 23-28.

Izard, T., A. Aevarsson, M. D. Allen, A. H. Westphal, R. N. Perham, A. de Kok and W. G. J. Hol (1999). "Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes." Proceedings of the National Academy of Sciences of the United States of America **96**(4): 1240-1245.

Jilka, J. M., M. Rahmatullah, M. Kazemi and T. E. Roche (1986). "Properties of a Newly Characterized Protein of the Bovine Kidney Pyruvate-Dehydrogenase Complex." Journal of Biological Chemistry **261**(4): 1858-1867.

- Jordan, S. W. and J. E. Cronan (1997). "A new metabolic link - The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in *Escherichia coli* and mitochondria." *Journal of Biological Chemistry* **272**(29): 17903-17906.
- Kaplon, J., L. Zheng, K. Meissl, B. Chaneton, V. A. Selivanov, G. Mackay, S. H. van der Burg, E. M. E. Verdegaaal, M. Cascante, T. Shlomi, E. Gottlieb and D. S. Peeper (2013). "A key role for mitochondrial gatekeeper pyruvate dehydrogenase in oncogene-induced senescence." *Nature* **498**(7452): 109-+.
- Kastner, B., N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. L. Hrmann and H. Stark (2008). "GraFix: sample preparation for single-particle electron cryomicroscopy." *Nature Methods* **5**(1): 53-55.
- Kato, M., R. M. Wynn, J. L. Chuang, C. A. Brautigam, M. Custorio and D. T. Chuang (2006). "A synchronized substrate-gating mechanism revealed by cubic-core structure of the bovine branched-chain alpha-ketoacid dehydrogenase complex." *Embo Journal* **25**(24): 5983-5994.
- Kato, M., R. M. Wynn, J. L. Chuang, S. C. Tso, M. Machius, J. Li and D. T. Chuang (2008). "Structural Basis for Inactivation of the Human Pyruvate Dehydrogenase Complex by Phosphorylation: Role of Disordered Phosphorylation Loops." *Structure* **16**(12): 1849-1859.
- Keppler, A., S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson (2003). "A general method for the covalent labeling of fusion proteins with small molecules in vivo." *Nature Biotechnology* **21**(1): 86-89.
- Kern, D., G. Kern, H. Neef, K. Tittmann, M. KillenbergJabs, C. Wikner, G. Schneider and G. Hubner (1997). "How thiamine diphosphate is activated in enzymes." *Science* **275**(5296): 67-70.
- Klar, T. A., S. Jakobs, M. Dyba, A. Egnér and S. W. Hell (2000). "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission." *Proceedings of the National Academy of Sciences of the United States of America* **97**(15): 8206-8210.

Koike, M., P. C. Shah and L. J. Reed (1960). "Alpha-Keto Acid Dehydrogenation Complexes .3. Purification and Properties of Dihydrolipoic Dehydrogenase of Escherichia-Coli." Journal of Biological Chemistry **235**(7): 1939-1943.

Kong, Y. F., D. M. Ming, Y. H. Wu, J. K. Stoops, Z. H. Zhou and J. P. Ma (2003). "Conformational flexibility of pyruvate dehydrogenase complexes: A computational analysis by quantized elastic deformational model." Journal of Molecular Biology **330**(1): 129-135.

Korkes, S., A. Delcampillo, I. C. Gunsalus and S. Ochoa (1951). "Enzymatic Synthesis of Citric Acid .4. Pyruvate as Acetyl Donor." Journal of Biological Chemistry **193**(2): 721-735.

Korotchkina, L. G. and M. S. Patel (2001). "Probing the mechanism of inactivation of human pyruvate dehydrogenase by phosphorylation of three sites." Journal of Biological Chemistry **276**(8): 5731-5738.

Korotchkina, L. G., M. M. Tucker, T. J. Thekkumkara, K. T. Madhusudhan, G. Pons, H. J. Kim and M. S. Patel (1995). "Overexpression and Characterization of Human Tetrameric Pyruvate-Dehydrogenase and Its Individual Subunits." Protein Expression and Purification **6**(1): 79-90.

Kratky, O. and H. Stabinger (1984). "X-Ray Small-Angle Camera with Block-Collimation System an Instrument of Colloid Research." Colloid and Polymer Science **262**(5): 345-360.

Kresze, G. B. and H. Ronft (1981). "Pyruvate-Dehydrogenase Complex from Bakers-Yeast .1. Purification and Some Kinetic and Regulatory Properties." European Journal of Biochemistry **119**(3): 573-579.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4." Nature **227**(5259): 680-&.

Lazo, P. A. and A. Sols (1980). "Pyruvate-Dehydrogenase Complex of Ascites Tumor - Activation by Amp and Other Properties of Potential Significance in Metabolic-Regulation." Biochemical Journal **190**(3): 705-710.

Leibundgut, M., T. Maier, S. Jenni and N. Ban (2008). "The multienzyme architecture of eukaryotic fatty acid synthases." Current Opinion in Structural Biology **18**(6): 714-725.

Maeng, C. Y., M. A. Yazdi, X. D. Niu, H. Y. Lee and L. J. Reed (1994). "Expression, Purification, and Characterization of the Dihydrolipoamide Dehydrogenase-Binding Protein of the Pyruvate-Dehydrogenase Complex from *Saccharomyces-Cerevisiae*." Biochemistry **33**(46): 13801-13807.

Mande, S. S., S. Sarfaty, M. D. Allen, R. N. Perham and W. G. J. Hol (1996). "Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: Dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase." Structure **4**(3): 277-286.

Marrott, N. L., J. J. T. Marshall, D. I. Svergun, S. J. Crennell, D. W. Hough, J. M. H. van den Elsen and M. J. Danson (2014). "Why are the 2-oxoacid dehydrogenase complexes so large? Generation of an active trimeric complex." Biochemical Journal **463**: 405-412.

Mathias, R. A., T. M. Greco, A. Oberstein, H. G. Budayeva, R. Chakrabarti, E. A. Rowland, Y. B. Kang, T. Shenk and I. M. Cristea (2014). "Sirtuin 4 Is a Lipoamidase Regulating Pyruvate Dehydrogenase Complex Activity." Cell **159**(7): 1615-1625.

Mattevi, A., G. Obmolova, K. H. Kalk, A. Teplyakov and W. G. J. Hol (1993). "Crystallographic Analysis of Substrate Binding and Catalysis in Dihydrolipoyl Transacetylase (E2p)." Biochemistry **32**(15): 3887-3901.

Mattevi, A., G. Obmolova, E. Schulze, K. H. Kalk, A. H. Westphal, A. Dekok and W. G. J. Hol (1992). "Atomic-Structure of the Cubic Core of the Pyruvate-Dehydrogenase Multienzyme Complex." Science **255**(5051): 1544-1550.

Mccammon, J. A., B. R. Gelin and M. Karplus (1977). "Dynamics of Folded Proteins." Nature **267**(5612): 585-590.

Michelakis, E. D., G. Sutendra, P. Dromparis, L. Webster, A. Haromy, E. Niven, C. Maguire, T. L. Gammer, J. R. Mackey, D. Fulton, B. Abdulkarim, M. S. McMurtry and K. C. Petruk (2010). "Metabolic Modulation of Glioblastoma with Dichloroacetate." Science Translational Medicine **2**(31).

Michelakis, E. D., L. Webster and J. R. Mackey (2008). "Dichloroacetate (DCA) as a potential metabolic-targeting therapy for cancer." British Journal of Cancer **99**(7): 989-994.

Miles, J. S., J. R. Guest, S. E. Radford and R. N. Perham (1988). "Investigation of the Mechanism of Active-Site Coupling in the Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli by Protein Engineering." Journal of Molecular Biology **202**(1): 97-106.

Milne, J. L. S., D. Shi, P. B. Rosenthal, J. S. Sunshine, G. J. Domingo, X. W. Wu, B. R. Brooks, R. N. Perham, R. Henderson and S. Subramaniam (2002). "Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine." Embo Journal **21**(21): 5587-5598.

Milne, J. L. S., X. W. Wu, M. J. Borgnia, J. S. Lengyel, B. R. Brooks, D. Shi, R. N. Perham and S. Subramaniam (2006). "Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy." Journal of Biological Chemistry **281**(7): 4364-4370.

Morgner, N. and C. V. Robinson (2012). "Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies." Analytical Chemistry **84**(6): 2939-2948.

Morris, T. W., K. E. Reed and J. E. Cronan (1994). "Identification of the Gene Encoding Lipoate-Protein Ligase-a of Escherichia-Coli - Molecular-Cloning and Characterization of the Lpla Gene and Gene-Product." Journal of Biological Chemistry **269**(23): 16091-16100.

Neagle, J., O. Demarcucci, B. Dunbar and J. G. Lindsay (1989). "Component-X of Mammalian Pyruvate-Dehydrogenase Complex - Structural and Functional-Relationship to the Lipoate Acetyltransferase (E2) Component." Febs Letters **253**(1-2): 11-15.

Neagle, J. C. and J. G. Lindsay (1991). "Selective Proteolysis of the Protein-X Subunit of the Bovine Heart Pyruvate-Dehydrogenase Complex - Effects on Dihydrolipoamide Dehydrogenase (E3) Affinity and Enzymatic-Properties of the Complex." Biochemical Journal **278**: 423-427.

Nemeria, N., Y. Yan, Z. Zhang, A. M. Brown, P. Arjunan, W. Furey, J. R. Guest and F. Jordan (2001). "Inhibition of the Escherichia coli pyruvate dehydrogenase complex E1 subunit and its tyrosine 177 variants by thiamin 2-thiazolone and thiamin 2-thiothiazolone diphosphates - Evidence for reversible tight-binding inhibition." Journal of Biological Chemistry **276**(49): 45969-45978.

Olsen, J. V., L. M. F. de Godoy, G. Q. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning and M. Mann (2005). "Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap." Molecular & Cellular Proteomics **4**(12): 2010-2021.

Patel, M. S. and L. G. Korotchkina (2006). "Regulation of the pyruvate dehydrogenase complex." Biochemical Society Transactions **34**: 217-222.

Patel, M. S., L. G. Korotchkina and S. Sidhu (2009). "Interaction of E1 and E3 components with the core proteins of the human pyruvate dehydrogenase complex." Journal of Molecular Catalysis B-Enzymatic **61**(1-2): 2-6.

Patel, M. S., N. S. Nemeria, W. Furey and F. Jordan (2014). "The Pyruvate Dehydrogenase Complexes: Structure-based Function and Regulation." Journal of Biological Chemistry **289**(24): 16615-16623.

Perham, R. N. (1991). "Domains, Motifs, and Linkers in 2-Oxo Acid Dehydrogenase Multienzyme Complexes - a Paradigm in the Design of a Multifunctional Protein." Biochemistry **30**(35): 8501-8512.

Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev and D. I. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." Journal of Applied Crystallography **45**: 342-350.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Pierce, M. M., C. S. Raman and B. T. Nall (1999). "Isothermal titration calorimetry of protein-protein interactions." Methods-a Companion to Methods in Enzymology **19**(2): 213-221.

Putman, C. A. J., K. O. Vanderwerf, B. G. Degrooth, N. F. Vanhulst and J. Greve (1994). "Tapping Mode Atomic-Force Microscopy in Liquid." Applied Physics Letters **64**(18): 2454-2456.

Radda, G. K. (1966). "Chemistry of Flavins and Flavoproteins .2. Inhibition of Photoreduction of Flavin Nucleotides and Analogues." Biochimica Et Biophysica Acta **112**(3): 448-&.

Rahmatullah, M., S. Gopalakrishnan, P. C. Andrews, C. L. Chang, G. A. Radke and T. E. Roche (1989). "Subunit Associations in the Mammalian Pyruvate-Dehydrogenase Complex - Structure and Role of Protein-X and the Pyruvate-Dehydrogenase Component Binding Domain of the Dihydrolipoyl Transacetylase Component." Journal of Biological Chemistry **264**(4): 2221-2227.

Rahmatullah, M., G. A. Radke, P. C. Andrews and T. E. Roche (1990). "Changes in the Core of the Mammalian-Pyruvate Dehydrogenase Complex Upon Selective Removal of the Lipoyl Domain from the Transacetylase Component but Not from the Protein-X Component." Journal of Biological Chemistry **265**(24): 14512-14517.



Rappsilber, J. (2011). "The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes." Journal of Structural Biology **173**(3): 530-540.

Reed, J. K. (1973). "Studies on Kinetic Mechanism of Lipoamide Dehydrogenase from Rat-Liver Mitochondria." Journal of Biological Chemistry **248**(13): 4834-4839.

Reed, L. J. (1966). "Macromolecular Organization of Enzyme Systems." Annual Review of Biochemistry **35**: 57-&.

Reed, L. J. (1974). "Multienzyme Complexes." Accounts of Chemical Research **7**(2): 40-46.

Reed, L. J. and M. L. Hackert (1990). "Structure-Function-Relationships in Dihydrolipoamide Acyltransferases." Journal of Biological Chemistry **265**(16): 8971-8974.

Reid, E. E., C. R. Lyttle, D. T. Canvin and D. T. Dennis (1975). "Pyruvate-Dehydrogenase Complex Activity in Proplastids and Mitochondria of Developing Castor Bean Endosperm." Biochemical and Biophysical Research Communications **62**(1): 42-47.

Ricaud, P. M., M. J. Howard, E. L. Roberts, R. W. Broadhurst and R. N. Perham (1996). "Three-dimensional structure of the lipoyl domain from the dihydrolipoyl succinyltransferase component of the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*." Journal of Molecular Biology **264**(1): 179-190.

Roche, T. E., M. Rahmatullah, S. L. Powersgreenwood, G. A. Radke, S. Gopalakrishnan and C. L. Chang (1989). "The Lipoyl-Containing Components of the Mammalian Pyruvate-Dehydrogenase Complex - Structural Comparison and Subdomain Roles." Annals of the New York Academy of Sciences **573**: 66-75.

Rust, M. J., M. Bates and X. Zhuang (2006). "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)." Nat Methods **3**(10): 793-795.

Prajapati, S. (2013). "Structure and dynamics of the core of human pyruvate dehydrogenase multi-enzyme complex." MSc thesis (IMPRS for Molecular Biology graduate program)

Sazanov, L. A. (2015). "A giant molecular proton pump: structure and mechanism of respiratory complex I." Nature Reviews Molecular Cell Biology **16**(6): 375-388.

Scheres, S. H. W. (2012). "RELION: Implementation of a Bayesian approach to cryo-EM structure determination." Journal of Structural Biology **180**(3): 519-530.

Schroder-Tittmann, K., D. Meyer, J. Arens, C. Wechsler, M. Tietzel, R. Golbik and K. Tittmann (2013). "Alternating Sites Reactivity Is a Common Feature of Thiamin Diphosphate-Dependent Enzymes As Evidenced by Isothermal Titration Calorimetry Studies of Substrate Binding." Biochemistry **52**(15): 2505-2507.

Seifert, F. (2010) Katalyse und Regulation der E1-Komponente des humanen Pyruvat-Dehydrogenase-Komplexes auf molekularer Ebene (2010). PhD thesis.

Seifert, F., E. Ciszak, L. Korotchkina, R. Golbik, M. Spinka, P. Dominiak, S. Sidhu, J. Brauer, M. S. Patel and K. Tittmann (2007). "Phosphorylation of serine 264 impedes active site accessibility in the E1 component of the human pyruvate dehydrogenase multienzyme complex." Biochemistry **46**(21): 6277-6287.

Seifert, F., R. Golbik, J. Brauer, H. Lilie, K. Schroder-Tittmann, E. Hinze, L. G. Korotchkina, M. S. Patel and K. Tittmann (2006). "Direct kinetic evidence for half-of-the-sites reactivity in the E1 component of the human pyruvate dehydrogenase multienzyme complex through alternating sites cofactor activation." Biochemistry **45**(42): 12775-12785.

Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen and M. Mann (2006). "In-gel digestion for mass spectrometric characterization of proteins and proteomes." Nat Protoc **1**(6): 2856-2860.

Smolle, M., A. E. Prior, A. E. Brown, A. Cooper, O. Byron and J. G. Lindsay (2006). "A new level of architectural complexity in the human pyruvate dehydrogenase complex." Journal of Biological Chemistry **281**(28): 19772-19780.

Sobott, F., H. Hernandez, M. G. McCammon, M. A. Tito and C. V. Robinson (2002). "A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies." Analytical Chemistry **74**(6): 1402-1407.

Song, J. Y. and F. Jordan (2012). "Interchain Acetyl Transfer in the E2 Component of Bacterial Pyruvate Dehydrogenase Suggests a Model with Different Roles for Each Chain in a Trimer of the Homooligomeric Component." Biochemistry **51**(13): 2795-2803.

Stepp, L. R., D. M. Bleile, D. K. Mcrorie, F. H. Pettit and L. J. Reed (1981). "Use of Trypsin and Lipoamidase to Study the Role of Lipoic Acid Moieties in the Pyruvate and Alpha-Ketoglutarate Dehydrogenase Complexes of Escherichia-Coli." Biochemistry **20**(16): 4555-4560.

Stepp, L. R., F. H. Pettit and L. J. Reed (1981). "Use of Trypsin and Lipoamidase to Probe the Catalytic Mechanism of Alpha-Keto Acid Dehydrogenase Complexes." Federation Proceedings **40**(6): 1795-1795.

Stoops, J. K., R. H. Cheng, M. A. Yazdi, C. Y. Maeng, J. P. Schroeter, U. Klueppelberg, S. J. Kolodziej, T. S. Baker and L. J. Reed (1997). "On the unique structural organization of the Saccharomyces cerevisiae pyruvate dehydrogenase complex." Journal of Biological Chemistry **272**(9): 5757-5764.

Studier, F. W. (2005). "Protein production by auto-induction in high-density shaking cultures." Protein Expression and Purification **41**(1): 207-234.

Taverner, T., H. Hernandez, M. Sharon, B. T. Ruotolo, D. Matak-Vinkovic, D. Devos, R. B. Russell and C. V. Robinson (2008). "Subunit architecture of intact protein complexes from mass spectrometry and homology modeling." Acc Chem Res **41**(5): 617-627.

Texter, F. L., S. E. Radford, E. D. Laue, R. N. Perham, J. S. Miles and J. R. Guest (1988). "Site-Directed Mutagenesis and H-1-Nmr Spectroscopy of an Interdomain Segment in the Pyruvate-Dehydrogenase Multienzyme Complex of Escherichia-Coli." Biochemistry **27**(1): 289-296.

Tittmann, K. (2009). "Reaction mechanisms of thiamin diphosphate enzymes: redox reactions." Febs Journal **276**(9): 2454-2468.

Turnbull, W. B. and A. H. Daranas (2003). "On the value of c: Can low affinity systems be studied by isothermal titration calorimetry?" Journal of the American Chemical Society **125**(48): 14859-14866.

van de Linde, S., A. Loschberger, T. Klein, M. Heidebreder, S. Wolter, M. Heilemann and M. Sauer (2011). "Direct stochastic optical reconstruction microscopy with standard fluorescent probes." Nat Protoc **6**(7): 991-1009.

Vijayakrishnan, S., P. Callow, M. A. Nutley, D. P. McGow, D. Gilbert, P. Kropholler, A. Cooper, O. Byron and J. G. Lindsay (2011). "Variation in the organization and subunit composition of the mammalian pyruvate dehydrogenase complex E2/E3BP core assembly." Biochemical Journal **437**: 565-574.

Vijayakrishnan, S., S. M. Kelly, R. J. C. Gilbert, P. Callow, D. Bhella, T. Forsyth, J. G. Lindsay and O. Byron (2010). "Solution Structure and Characterisation of the Human Pyruvate Dehydrogenase Complex Core Assembly." Journal of Molecular Biology **399**(1): 71-93.

Wagenknecht, T., R. Grassucci, G. A. Radke and T. E. Roche (1991). "Cryoelectron Microscopy of Mammalian Pyruvate-Dehydrogenase Complex." Journal of Biological Chemistry **266**(36): 24650-24656.

Wang, J. J., N. S. Nemeria, K. Chandrasekhar, S. Kumaran, P. Arjunan, S. Reynolds, G. Calero, R. Brukh, L. Kakalis, W. Furey and F. Jordan (2014). "Structure and Function of the Catalytic Domain of the Dihydrolipoyl Acetyltransferase Component in Escherichia coli Pyruvate Dehydrogenase Complex." Journal of Biological Chemistry **289**(22): 15215-15230.

Watt, I. N., M. G. Montgomery, M. J. Runswick, A. G. W. Leslie and J. E. Walker (2010). "Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria." Proceedings of the National Academy of Sciences of the United States of America **107**(39): 16823-16827.

Weiner, A. M., K. Weber and T. Platt (1972). "Amino-Terminal Sequence Analysis of Proteins Purified on a Nanomole Scale by Gel-Electrophoresis." Journal of Biological Chemistry **247**(10): 3242-&.

Wiseman, T., S. Williston, J. F. Brandts and L. N. Lin (1989). "Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter." Analytical Biochemistry **179**(1): 131-137.

Yang, B., Y. J. Wu, M. Zhu, S. B. Fan, J. Z. Lin, K. Zhang, S. Li, H. Chi, Y. X. Li, H. F. Chen, S. K. Luo, Y. H. Ding, L. H. Wang, Z. Q. Hao, L. Y. Xiu, S. Chen, K. Q. Ye, S. M. He and M. Q. Dong (2012). "Identification of cross-linked peptides from complex samples." Nature Methods **9**(9).

Yang, D. Q., J. S. Song, T. Wagenknecht and T. E. Roche (1997). "Assembly and full functionality of recombinantly expressed dihydrolipoyl acetyltransferase component of the human pyruvate dehydrogenase complex." Journal of Biological Chemistry **272**(10): 6361-6369.

Yu, H. L., X. Q. Du, F. X. Zhang, F. Zhang, Y. Hu, S. C. Liu, X. N. Jiang, G. D. Wang and D. Liu (2012). "A mutation in the E2 subunit of the mitochondrial pyruvate dehydrogenase complex in Arabidopsis reduces plant organ size and enhances the accumulation of amino acids and intermediate products of the TCA Cycle." Planta **236**(2): 387-399.

Yu, X. K., Y. Hiromasa, H. Tsen, J. K. Stoops, T. E. Roche and Z. H. Zhou (2008). "Structures of the human pyruvate dehydrogenase complex cores: A highly conserved catalytic center with flexible N-terminal domains." Structure **16**(1): 104-114.

Zaccai, G. (2000). "Biochemistry - How soft is a protein? A protein dynamics force constant measured by neutron scattering." Science **288**(5471): 1604-1607.

Zhao, X., J. R. Miller, Y. F. Jiang, M. A. Marletta and J. E. Cronan (2003). "Assembly of the covalent linkage between lipoic acid and its cognate enzymes." Chemistry & Biology **10**(12): 1293-1302.

Zhou, Z. H., W. C. Liao, R. H. Cheng, J. E. Lawson, D. B. McCarthy, L. J. Reed and J. K. Stoops (2001). "Direct evidence for the size and conformational variability of the pyruvate dehydrogenase complex revealed by three-dimensional electron microscopy - The "breathing" core and its functional relationship to protein dynamics." Journal of Biological Chemistry **276**(24): 21704-21713.

Zhou, Z. H., D. B. McCarthy, C. M. O'Connor, L. J. Reed and J. K. Stoops (2001). "The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes." Proceedings of the National Academy of Sciences of the United States of America **98**(26): 14802-14807.

## 8. Appendix

### A. Generation of 6xHis-*hE2* and SNAP-*hE2* construct

First of all, a site directed mutagenesis PCR was performed with pET28a-*hE2* (GeneArt, ThermoFisher Scientific) in order to remove a Nde I restriction site in the sequence of *hE2* using following two primers.

FP\_Mut\_NdeI\_*hE2*: 5'-GCTATCCGCCGCACATGCAGGTTCTGC-3'

RP\_Mut\_NdeI\_*hE2*: 5'-GCAGAACCTGCATGTGCGGCGGATAGC-3'

Then using following two primers, a linear sequence for *hE2* from pET28a-*hE2* (GeneArt, ThermoFisher Scientific) was generated with Nde I site at its 5' and Xho I site at its 3' end with STOP codon at immediate 5' to Xho I site.

FP\_NdeI\_polyHIS\_*hE2*: 5'- GCATTAACA**CATATG**GGTTCACTGCCTCCGC – 3'

Overlap **Nde I** 5'-sequence E2

RP\_XhoI\_polyHIS\_*hE2*: 5'- GGTTA**CTCGAGTTA**CAGCAGCAGGTGATCG – 3'

**Xho I** **Stop** end of E2 sequence

Now, Nde I and Xho I restriction endonucleases were used to cleave the obtained construct and an empty pET28a vector, and ligated to obtain N-terminally poly histidine tagged 6xHis-*hE2* construct.

Sequence of pET28a-6xHis-*hE2* is as follows:

MGSSHHHHHSSGLVPRGSHMGSLPPHQKVPLPSLSPTMQAGTIARWEKKEGDK  
 INEGDLIAEVETDKATVGFESLEECYMAKILVAEGTRDVPIGAIICITVGKPEDIEAFK  
 NYTLDSSAAPTPQAAPAPTPAATASPPTPSAQAPGSSYPPHMVLLPALSPTMTM  
 GTVQRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAKILVPEGTRDVPLGT  
 PLCIIVEKEADISAFADYRPTEVTDLKPQVPPPTPPPVAAPPTPQPLAPTPSAPCP  
 ATPAGPKGR  
 VFVSPLAKKLAVEKGIDLTQVKGTGPDGRITKKDIDSFVPSKVAPAPAAVVPPTGPG  
 MAPVPTGVFTDIPISNIRRVIAQRLMQSKQTIPHYLSIDVNMGEVLLVRKELNKILE  
 GRSKISVNDFIKASALACLKVPEANSSWMDTVIRQNHVVDVSVAVSTPAGLITPIVF  
 NAHIKGVETIANDVVSLATKAREGKLPHEFQGGTFTISNLGMFGIKNFSAINPPQA  
 CILAIGASE  
 DKLVPADNEKGFVDVASMMSVTLSCDHRVVDGAVGAQWLAEFRKYLEKPITMLL

Using following two primers, NcoI and NdeI restriction sites were added to start and end of SNAP sequence from a plasmid received from the lab of Prof. Steffen Jakobs.

FP\_SNAPtag: 5'- CGATGCTCTATAG**CCATGG**TGAGGTAGGGCG – 3'

Random sequence    **Nco I**    SNAP seq. start

RP\_SNAPtag: 5'- CGCAG**CATATG**CTTGTACAGCTCGTCCATGC – 3'

Random seq    **Nde I**    SNAP seq end

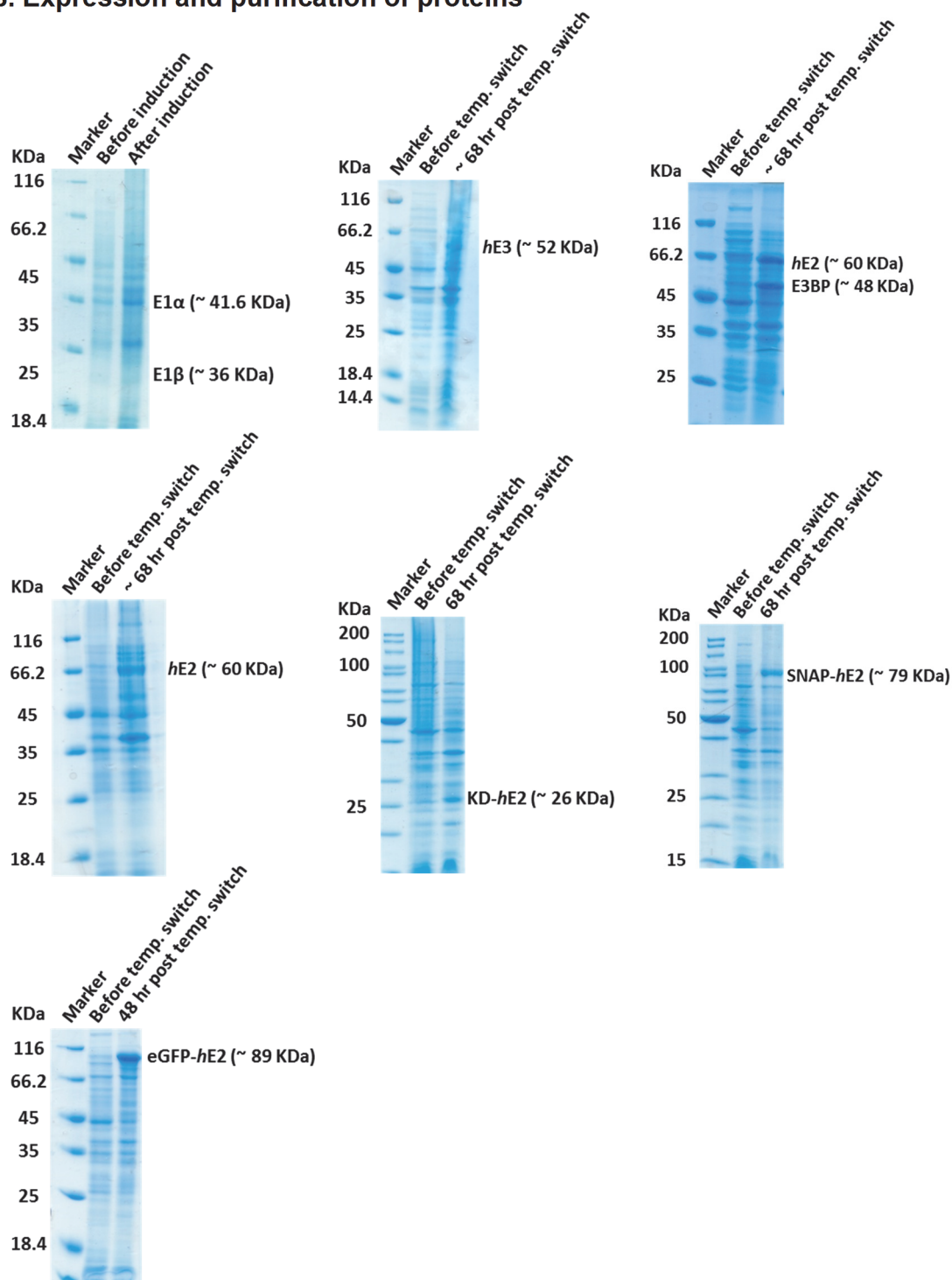
Using the corresponding restriction sites endonucleases on SNAP tag generated this way and pET28a-6xHis-*hE2* construct (sub-cloned using NdeI and XhoI restriction sites), SNAP-*hE2* was created where SNAP tag is N-terminally tagged to *hE2*.



Sequence of **SNAP-hE2** is as follows:

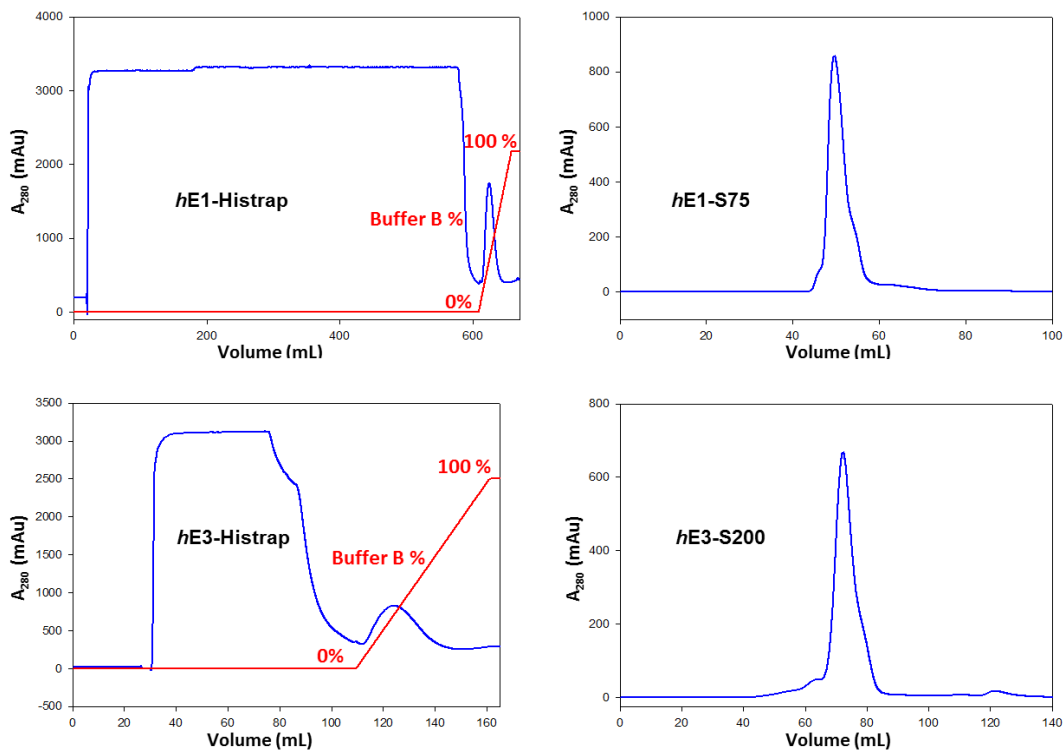
MDKDCEMKRTTLDSP LGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGG  
PEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGE  
VISYQQLAALAGNPAATAAVKTALSGNPVPIIPCHRVS SSGAVGGYEGGLAVKE  
WLLAHEGHRLGKPGLGHMGS LPPHQVPLPSLSPTMQAGTIARWKKKEGDKINE  
GDLIAEVETDKATVGFESLEECYMAKILVAEGTRDVPIGAIICITVGKPEDIEAFKNYT  
LDSSAAPTQQAAPAPTPAATASPPTPSAQAPGSSYP PHMQVLLPALSPTMTMGTV  
QRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAKILVPEGTRDVPLGTPLCII  
VEKEADISAFADYRPTVETDLKPQVPPPTPPPVA AVPPTPQPLAPTPSAPCPATPA  
GPKGRV FVSPLAKKLAVEKGIDLTQVKGTGPDGRITK KDIDSFVPSKVAPAPAAVVP  
PTGPGMAPVPTGVFTDIPISNIRRVIAQRLMQSKQTIPHYYSIDVNMGEVLLVRKEL  
NKILEGRSKISVNDFIKASALACLKVPEANSSWMDTVIRQNHVVDVSVAVSTPAGLI  
TPIVFNAHIKGVETIANDV VSLATKAREGKLQPHEFQGGTFTISNLGMFGIKNFSAIIN  
PPQACILAIGASEDKLVPADNEKGF DVASMMSVTLSCDHRVVDGAVGAQWLAEFR  
KYLEKPITMLL STOP

## B. Expression and purification of proteins



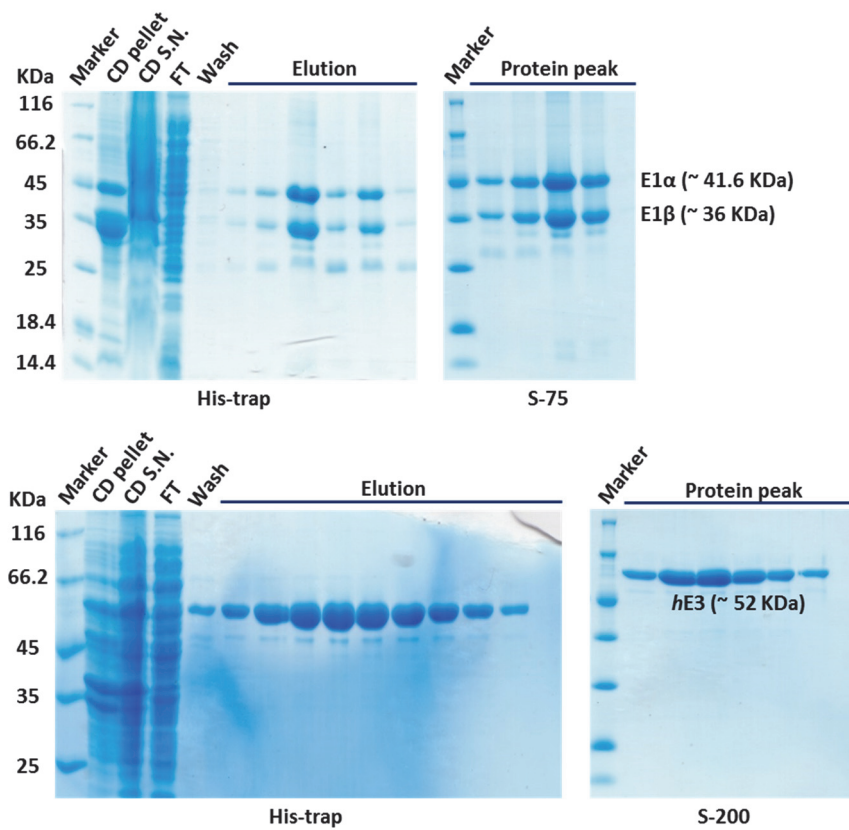
**Fig. 50: SDS-PAGE analysis of overexpression of various proteins used during this thesis work.**

*hE1* was over expressed using bio fermenter whereas auto-induction method in an incubation shaker was used for over expressing all the other proteins. The bands of respective proteins can be seen at the lane of post overexpression close to corresponding molecular weights.



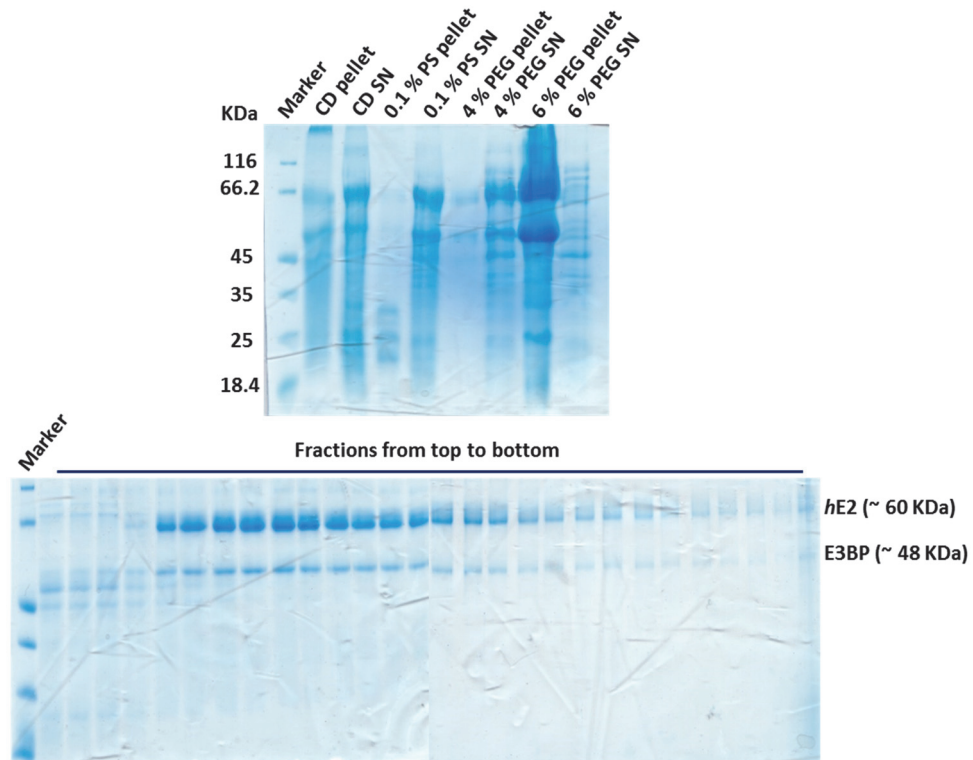
**Fig. 51: Chromatograms of *hE1* and *hE3* purifications.**

The *hE1* and *hE3* proteins both were purified in His-trap columns using gradient of respective elution buffers. It was followed by size exclusion chromatography (S75 for *hE1* and S200 for *hE3*) after which the proteins were judged by SDS-PAGE analysis if the batch was pure.



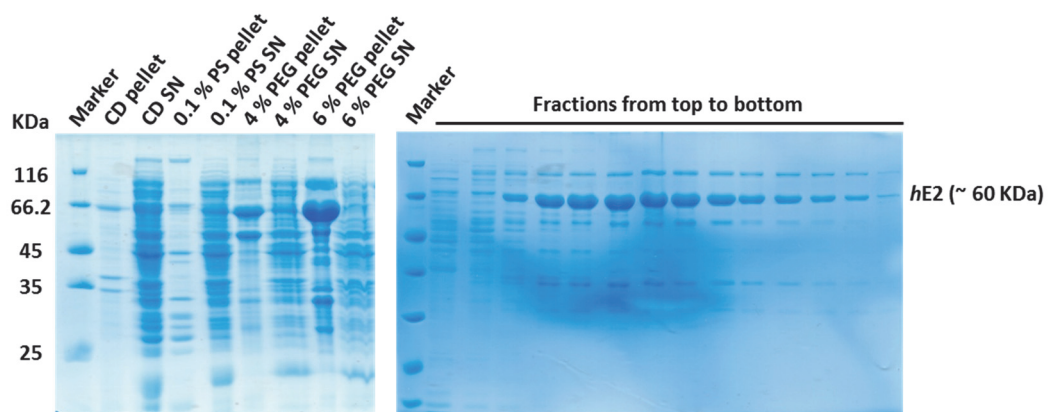
**Fig. 52: SDS-PAGE analysis of the purification of *hE1* and *hE3* proteins.**

In general, both these proteins were seen to be pure after gel filtration step. The final yield of *hE1* and *hE3* purification are 0.6 mg and 3 mg per 1 g cell pellet used. (CD-cell disruption, S.N. – supernatant, FT - flow through)



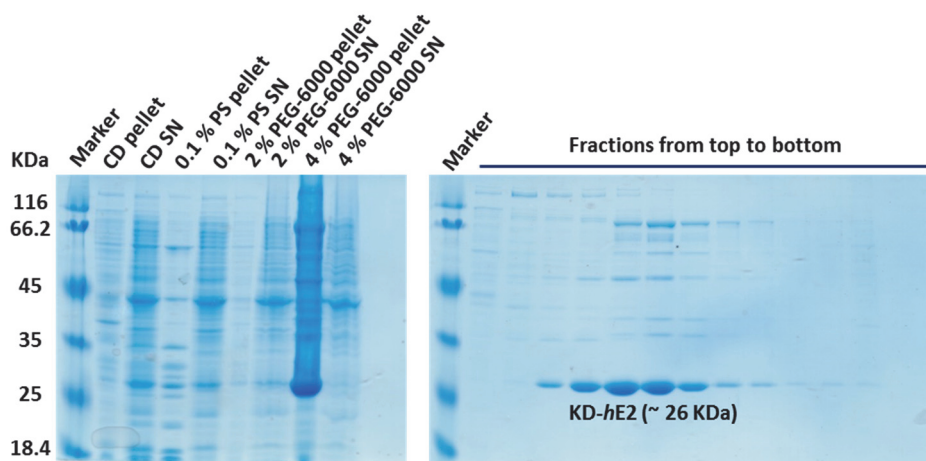
**Fig. 53: SDS-PAGE analysis of the purification of *hE2/E3BP* core.**

After cell disruption, the lysate underwent 0.1 % protamine sulfate and PEG 6000 (4 % → 6 %) induced precipitations. Post-sucrose gradient ultracentrifugation, the fractions representing middle of the bottom gel were selected and pooled. (CD-cell disruption, S.N. - supernatant, PS - protamine sulfate, PEG – PEG6000)



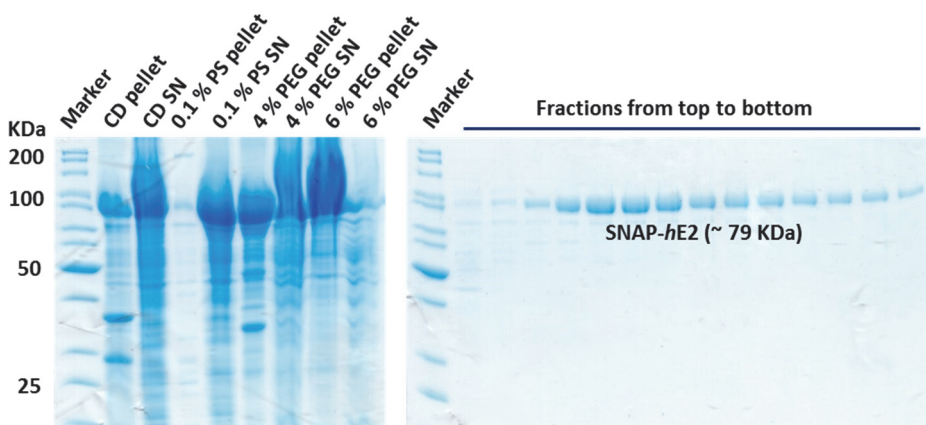
**Fig. 54: SDS-PAGE analysis of the purification of *hE2* core.**

After cell disruption, the lysate underwent 0.1 % protamine sulfate and PEG 6000 (4 % → 6 %) induced precipitations. Post-sucrose gradient ultracentrifugation, the fractions representing middle of the gel were selected and pooled. (CD-cell disruption, S.N. - supernatant, PS - protamine sulfate, PEG – PEG6000)



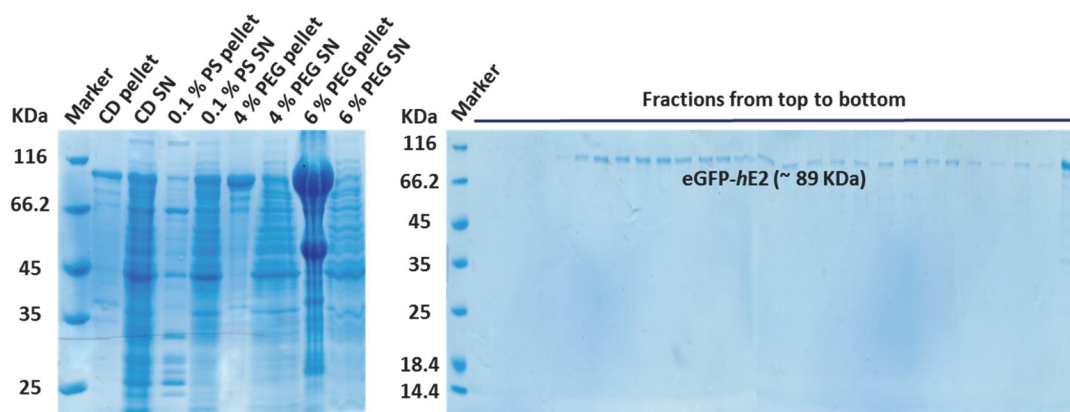
**Fig. 55: SDS-PAGE analysis of the purification of KD-*hE2* core.**

After cell disruption, the lysate underwent 0.1 % protamine sulfate and PEG 6000 (2 % → 4 %) induced precipitations. Post-sucrose gradient ultracentrifugation, the fractions representing middle of the gel were selected and pooled. (CD-cell disruption, S.N. - supernatant, PS - protamine sulfate)



**Fig. 56: SDS-PAGE analysis of the purification of SNAP-*hE2* core.**

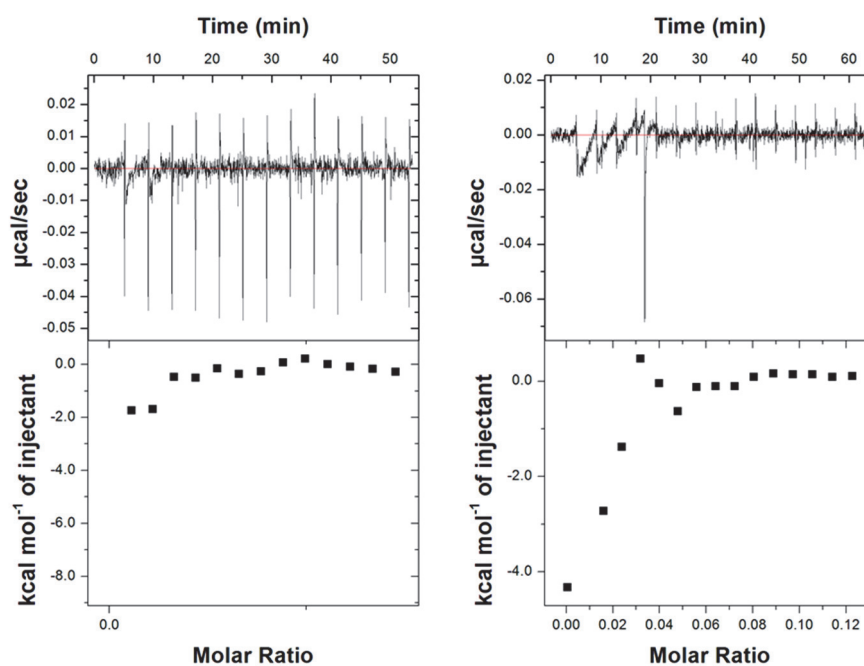
After cell disruption, the lysate underwent 0.1 % protamine sulfate and PEG 6000 (4 % → 6 %) induced precipitations. Post-sucrose gradient ultracentrifugation, the fractions representing middle of the gel were selected and pooled. (CD-cell disruption, S.N. - supernatant, PS - protamine sulfate, PEG – PEG6000)



**Fig. 57: SDS-PAGE analysis of the purification of eGFP-*hE2* core.**

After cell disruption, the lysate underwent 0.1 % protamine sulfate and PEG 6000 (4 %  $\rightarrow$  6 %) induced precipitations. Post-sucrose gradient ultracentrifugation, the fractions representing middle of the gel were selected and pooled. (CD-cell disruption, S.N. - supernatant, PS - protamine sulfate, PEG – PEG6000)

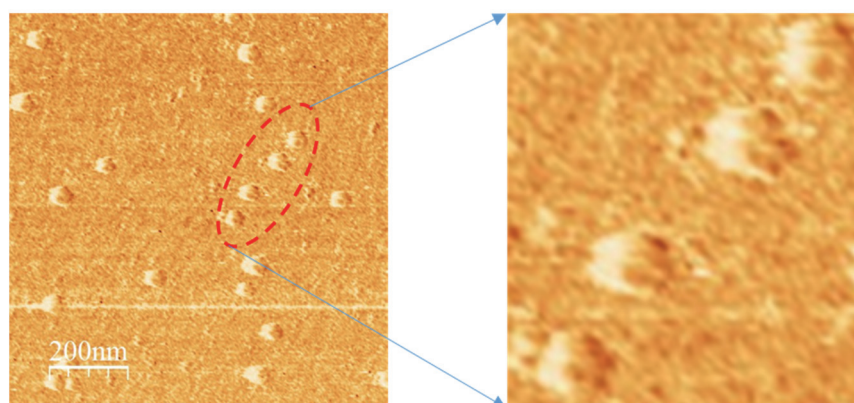
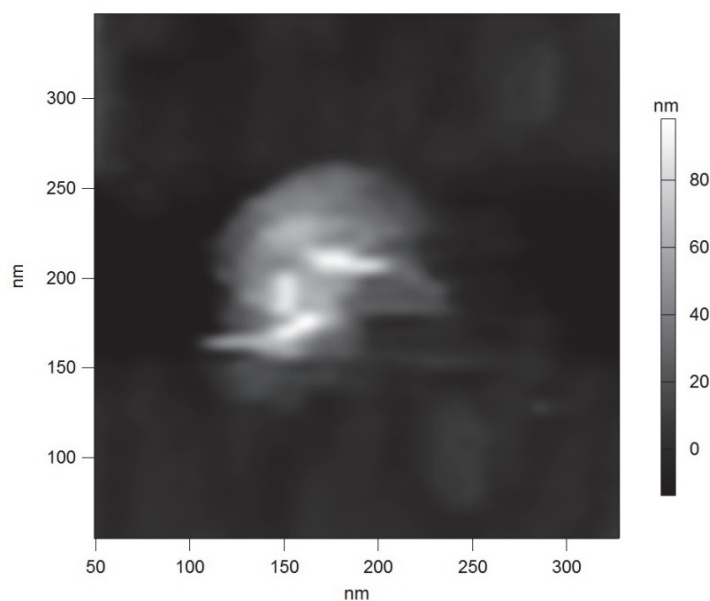
### C. ITC binding experiment of *hE1* with *hE2/E3BP* core



**Fig. 58: An ITC-binding experiment between *hE1* and *hE2/E3BP* core.**

An ITC data was obtained after titration of 147.4  $\mu\text{M}$  *hE1* to 1  $\mu\text{M}$  and 3.44  $\mu\text{M}$  *hE2/E3BP* core in 50 mM  $\text{KH}_2\text{PO}_4$ , 300 mM KCl, pH 7.5 at 25  $^\circ\text{C}$  with stirring speed of 500 RPM. The upper panel shows the heat of binding for each injection. Since *hE1* precipitated severely in this ITC condition, no meaningful data could be obtained.

### D. Intricate surface details of *h*PDHc cores observed in AFM



**Fig. 59: Surficial details of *h*PDHc cores observed occasionally during AFM study.**

In the top Fig., features resembling three lipoyl arms can be seen. These arm like features had higher heights than rest of the particle. In the bottom figures, depressions on KD-*h*E2 cores can be seen when zoomed in (at 5 fold axis only one such hole should be visible, while on 3 fold or 2 fold axis more than one such holes should be visible). However these observations were not reproducible.



### **E. Mass spectrometric confirmation of proteolytic processing of full length *hE2* and *E3BP* into their IC domains (by Dr. Oliver Valerius)**

Sequence of tE2 based on peptides detected in MS (~ 243 a.a., 26.4 KDa)

VVPPTGPGMAPVPTGVFTDIPISNIRRVIAQRLMQSKQTIPHYYSIDVNMGEVLLV  
 RKELNKILEGRSKISVNDFFIIKASALACKLVPEANSSWMDTVIRQNHVVDVSAVST  
 PAGLITPIVFNAHIKGVETIANDVVSLATKAREGKLQPHEFQGGTFTISNLGMFGIKN  
 FSAIINPPQACILAIGASEDKLVPADNEKGFVDVASMMSVTLSCDHRVVDGAVGAQW  
 LAEFRKYLEKPITMLL

Sequence of tE3BP based on peptides detected in MS (~ 205 a.a., 22.2 KDa)

STVPHAYATADCDLGAVLKVRQDLVKDDIKVSVNDFIIKAAAVTLKQMPDVNVSWD  
 GEGPKQLPFIDISVAVATDKGLLTPIIKDAAAKGIQEIADSVKALSKKARDGKLLPEEY  
 QG  
 GSFISISNLGMFGIDEFTAVINPPQACILAVGRFRPVLKLTEDEEGNAKLQRQLITVTM  
 S  
 SDSRVVDEDELATRFLKSFKANLENPIRLA

### **F. ThermoFluor assay of truncated *hPDHc* cores**

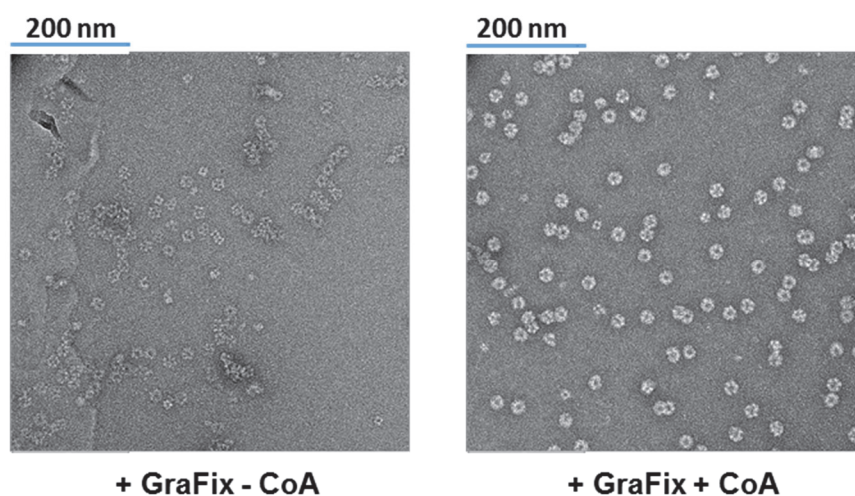
In order to check the stability of the truncated *hPDHc* cores generated by proteolysis, ThermoFluor assay was performed. In this assay, the protein (5  $\mu$ M tE2 or tE2/E3BP monomers) was kept at different buffer (100 mM) and pH conditions and mixed with 5 x SYPRO Orange dye. With increasing temperature, the protein gets denatured and the hydrophobic regions will be exposed. This dye binds to these exposed hydrophobic regions and gives fluorescence signals. Thus, fluorescent signal obtained at particular temperature and particular solution condition can be correlated with level of protein denaturation and the melting curve is generated.

Many buffer conditions and pH ranges were screened from 20-95 °C with stepwise temperature increment of 1 K / 30 s in a C1000 thermal cycler. The fluorescence signals were monitored using CFX96™ Optical Reaction Module (excitation: 515-535, detection: 560-580 nm). The assay was performed in the facility of 'Department of Molecular Structural Biology, Georg-August-Universität-Göttingen'.

MES (pH 4.6-7.4), BisTris (pH 5.2-8.0), Imidazole (pH 5.4-8.2), Phosphate (pH 5.8-8.6), HEPES (pH 6.0-8.8) and Tris (pH 6.6-9.4) were used for screening with pH-step 0.4 for each buffer component.

The proteins were found to be similarly stable at many conditions but the optimum buffer and pH was chosen to be best in MES and BisTris 6-6.5.

### G. Negative staining electron microscopy of truncated *hE2/E3BP* cores



**Fig. 60: Negative staining EM images of truncated *hE2/E3BP*.**

Roughly 50 pmoles *tE2/E3BP* cores were used for GraFix and the best fraction from dot blot was used for negative staining EM. The sample containing CoA could be clearly seen homogenously distributed. Without CoA though, most of these cores aggregated.

## H. Crosslinks in hPDHc as detected in crosslinking MS

The numbering of amino acyl residues is based on Uniprot protein database (<http://www.uniprot.org/>). The crosslinks observed in + CoA condition are shown in blue whereas - CoA condition is in black.

In the presence of 2 mM CoA				In the absence of CoA			
Protein 1	Residue 1	Protein 2	Residue 2	Protein 1	Residue 1	Protein 2	Residue 2
E3BP	120	E3BP	78				
E3BP	120	E3BP	173				
E3BP	120	E3BP	488	E3BP	120	E3BP	194
E3BP	155	E3BP	173				
E3BP	155	E3BP	227				
E3BP	194	E3BP	155				
E3BP	194	E3BP	173				
E3BP	194	E3BP	213	E3BP	194	E3BP	194
E3BP	194	E3BP	218	E3BP	194	E3BP	213
E3BP	194	E3BP	223	E3BP	194	E3BP	223
E3BP	194	E3BP	227	E3BP	194	E3BP	227
E3BP	194	E3BP	227	E3BP	194	E3BP	321
E3BP	194	E3BP	450	E3BP	194	E3BP	460
E3BP	213	E3BP	227	E3BP	194	E3BP	460
E3BP	218	E3BP	227	E3BP	218	E3BP	227
E3BP	227	E3BP	194				
E3BP	227	E3BP	218				
E3BP	314	E3BP	321	E3BP	314	E3BP	450
E3BP	314	E3BP	450	E3BP	314	E3BP	460
E3BP	314	E3BP	460	E3BP	314	E3BP	460
E3BP	325	E3BP	394	E3BP	325	E3BP	394
E3BP	341	E3BP	194				
E3BP	450	E3BP	321	E3BP	450	E3BP	321
E3BP	450	E3BP	379				
E3BP	450	E3BP	450				
E3BP	450	E3BP	491	E3BP	450	E3BP	491
E3BP	460	E3BP	289				
E3BP	460	E3BP	321	E3BP	460	E3BP	321
E3BP	460	E3BP	460	E3BP	460	E3BP	379
				E3BP	460	E3BP	460
E3BP	460	E3BP	488	E3BP	491	E3BP	491
E3BP	488	E3BP	488	E3BP	491	E3BP	491
E3BP	120	E1-alpha	40	E3BP	120	E1-alpha	77
E3BP	120	E1-alpha	77				
E3BP	155	E1-alpha	40				
E3BP	155	E1-alpha	277				
E3BP	194	E1-alpha	77	E3BP	194	E1-alpha	40
				E3BP	194	E1-alpha	77
E3BP	194	E1-alpha	244	E3BP	194	E1-alpha	149
E3BP	194	E1-alpha	277	E3BP	194	E1-alpha	244
E3BP	194	E1-alpha	336	E3BP	194	E1-alpha	277
E3BP	289	E1-alpha	40	E3BP	218	E1-alpha	277
E3BP	120	E3	159	E3BP	120	E3	159
E3BP	120	E3	267	E3BP	120	E3	267
E3BP	120	E3	273				
E3BP	120	E3	277	E3BP	120	E3	277

E3BP	120	E3	284	E3BP	120	E3	410
E3BP	120	E3	410	E3BP	120	E3	430
E3BP	120	E3	430	E3BP	120	E3	445
E3BP	155	E3	159	E3BP	155	E3	159
E3BP	155	E3	277	E3BP	155	E3	277
E3BP	155	E3	410	E3BP	155	E3	410
E3BP	173	E3	445	E3BP	155	E3	430
E3BP	194	E3	277	E3BP	173	E3	445
E3BP	194	E3	430	E3BP	194	E3	159
E3BP	194	E3	445	E3BP	194	E3	267
E3BP	194	E3	505	E3BP	194	E3	277
E3BP	213	E3	445	E3BP	194	E3	410
E3BP	223	E3	445	E3BP	194	E3	430
E3BP	227	E3	277	E3BP	194	E3	445
E3BP	227	E3	445	E3BP	194	E3	505
E3BP	289	E3	267	E3BP	194	E3	445
E3BP	289	E3	277	E3BP	218	E3	445
E3BP	295	E3	277	E3BP	289	E3	267
E3BP	460	E3	277	E3BP	450	E3	267
E3BP	120	E2	114	E3BP	450	E3	430
E3BP	120	E2	241	E3BP	450	E3	445
E3BP	120	E2	363	E3BP	460	E3	277
E3BP	120	E2	462	E3BP	460	E3	430
E3BP	155	E2	241	E3BP	120	E2	241
E3BP	194	E2	241	E3BP	172	E2	462
E3BP	194	E2	362	E3BP	194	E2	241
E3BP	194	E2	363	E3BP	194	E2	362
E3BP	194	E2	368	E3BP	194	E2	363
E3BP	194	E2	376	E3BP	194	E2	376
E3BP	194	E2	387	E3BP	194	E2	387
E3BP	194	E2	473	E3BP	194	E2	462
E3BP	194	E2	552	E3BP	194	E2	473
E3BP	227	E2	241	E3BP	194	E2	552
E3BP	227	E2	376	E3BP	289	E2	387
E3BP	289	E2	387	E3BP	295	E2	440
E3BP	295	E2	362	E3BP	314	E2	462
E3BP	295	E2	376	E3BP	341	E2	376
E3BP	295	E2	547	E3BP	341	E2	473
E3BP	314	E2	462	E3BP	450	E2	473
E3BP	341	E2	473	E3BP	460	E2	462
E3BP	450	E2	387	E3BP	460	E2	604
E3BP	450	E2	466	E3BP	194	E1-beta	227
E3BP	450	E2	473	E3BP	488	E1-beta	229
E3BP	460	E2	604	E1-alpha	40	E1-alpha	40
E3BP	488	E2	363	E1-alpha	40	E1-alpha	142
E3BP	488	E2	362				
E3BP	488	E2	376				
E3BP	488	E2	462				
E3BP	491	E2	547				
E3BP	194	E1-beta	227				
E3BP	488	E1-beta	229				
E1-alpha	40	E1-alpha	40				
E1-alpha	40	E1-alpha	142				

E1-alpha	63	E1-alpha	83	E1-alpha	77	E1-alpha	63
E1-alpha	77	E1-alpha	77	E1-alpha	77	E1-alpha	77
E1-alpha	77	E1-alpha	83				
E1-alpha	77	E1-alpha	85				
E1-alpha	77	E1-alpha	244	E1-alpha	77	E1-alpha	244
E1-alpha	77	E1-alpha	313	E1-alpha	77	E1-alpha	313
E1-alpha	77	E1-alpha	321	E1-alpha	77	E1-alpha	321
E1-alpha	83	E1-alpha	40				
E1-alpha	83	E1-alpha	142				
E1-alpha	149	E1-alpha	244				
				E1-alpha	158	E1-alpha	77
E1-alpha	158	E1-alpha	83				
E1-alpha	158	E1-alpha	182				
E1-alpha	158	E1-alpha	277	E1-alpha	158	E1-alpha	277
E1-alpha	244	E1-alpha	40	E1-alpha	244	E1-alpha	40
E1-alpha	244	E1-alpha	63				
E1-alpha	244	E1-alpha	142	E1-alpha	244	E1-alpha	142
E1-alpha	244	E1-alpha	244	E1-alpha	244	E1-alpha	244
E1-alpha	277	E1-alpha	77	E1-alpha	277	E1-alpha	77
E1-alpha	277	E1-alpha	83	E1-alpha	277	E1-alpha	83
E1-alpha	277	E1-alpha	85				
E1-alpha	277	E1-alpha	244	E1-alpha	277	E1-alpha	244
E1-alpha	277	E1-alpha	277	E1-alpha	277	E1-alpha	277
E1-alpha	277	E1-alpha	313	E1-alpha	277	E1-alpha	313
E1-alpha	277	E1-alpha	321	E1-alpha	277	E1-alpha	321
E1-alpha	321	E1-alpha	40				
E1-alpha	321	E1-alpha	142				
				E1-alpha	336	E1-alpha	40
E1-alpha	336	E1-alpha	77	E1-alpha	336	E1-alpha	77
E1-alpha	336	E1-alpha	277	E1-alpha	336	E1-alpha	277
E1-alpha	336	E1-alpha	313	E1-alpha	336	E1-alpha	313
E1-alpha	336	E1-alpha	321	E1-alpha	336	E1-alpha	321
E1-alpha	344	E1-alpha	40				
E1-alpha	344	E1-alpha	77	E1-alpha	344	E1-alpha	77
E1-alpha	344	E1-alpha	336	E1-alpha	344	E1-alpha	336
				E1-alpha	40	E3	66
				E1-alpha	40	E3	284
E1-alpha	40	E3	277				
E1-alpha	40	E3	320				
E1-alpha	40	E3	445				
E1-alpha	63	E3	430				
				E1-alpha	77	E3	104
E1-alpha	77	E3	159				
E1-alpha	77	E3	320				
E1-alpha	77	E3	410	E1-alpha	77	E3	410
E1-alpha	77	E3	445	E1-alpha	77	E3	445
E1-alpha	83	E3	159				
E1-alpha	83	E3	445				
E1-alpha	158	E3	277				
E1-alpha	244	E3	267				
E1-alpha	244	E3	277				
E1-alpha	244	E3	445				
E1-alpha	277	E3	122				
				E1-alpha	277	E3	211
				E1-alpha	277	E3	267
				E1-alpha	277	E3	277
				E1-alpha	277	E3	430
				E1-alpha	313	E3	430
E1-alpha	277	E3	277				
				E1-alpha	336	E3	267
E1-alpha	321	E3	277	E1-alpha	336	E3	430
E1-alpha	321	E3	320	E1-alpha	344	E3	277
				E1-alpha	344	E3	445
E1-alpha	336	E3	430	E1-alpha	40	E2	241
E1-alpha	40	E2	362				

E1-alpha	40	E2	376				
E1-alpha	40	E2	387				
E1-alpha	40	E2	547	E1-alpha	40	E2	547
E1-alpha	63	E2	114				
E1-alpha	77	E2	241	E1-alpha	63	E2	552
E1-alpha	77	E2	376	E1-alpha	77	E2	362
E1-alpha	77	E2	462	E1-alpha	77	E2	376
E1-alpha	77	E2	552	E1-alpha	77	E2	473
E1-alpha	77	E2	604				
E1-alpha	83	E2	241	E1-alpha	83	E2	241
E1-alpha	147	E2	241				
E1-alpha	244	E2	114	E1-alpha	244	E2	114
E1-alpha	244	E2	241	E1-alpha	244	E2	241
E1-alpha	244	E2	363				
E1-alpha	244	E2	462				
E1-alpha	277	E2	376	E1-alpha	277	E2	376
E1-alpha	277	E2	387				
E1-alpha	277	E2	462				
E1-alpha	277	E2	473	E1-alpha	277	E2	473
E1-alpha	321	E2	114				
E1-alpha	321	E2	241	E1-alpha	321	E2	241
				E1-alpha	336	E2	114
E1-alpha	336	E2	362	E1-alpha	336	E2	241
E1-alpha	336	E2	376				
E1-alpha	336	E2	473	E1-alpha	336	E2	363
E1-alpha	385	E2	387	E1-alpha	336	E2	376
E1-alpha	40	E1-beta	227	E1-alpha	336	E2	462
E1-alpha	77	E1-beta	229	E1-alpha	336	E2	473
E1-alpha	77	E1-beta	258	E1-alpha	344	E2	241
E1-alpha	83	E1-beta	68				
E1-alpha	83	E1-beta	354	E1-alpha	40	E1-beta	227
E1-alpha	85	E1-beta	227	E1-alpha	77	E1-beta	227
E1-alpha	85	E1-beta	336				
E1-alpha	244	E1-beta	227	E1-alpha	83	E1-beta	68
E1-alpha	277	E1-beta	76				
E1-alpha	277	E1-beta	68	E1-alpha	277	E1-beta	75
				E1-alpha	277	E1-beta	188
				E1-alpha	277	E1-beta	227
				E1-alpha	277	E1-beta	355
				E1-alpha	321	E1-beta	354
				E1-alpha	385	E1-beta	227
				E3	66	E3	430
E3	66	E3	445				
E3	72	E3	166				
E3	72	E3	320	E3	72	E3	320
E3	104	E3	122	E3	104	E3	122
E3	104	E3	267	E3	104	E3	267
				E3	104	E3	277
E3	104	E3	284				
E3	104	E3	430	E3	104	E3	430
E3	122	E3	430	E3	122	E3	430
E3	132	E3	122				
E3	132	E3	430	E3	132	E3	430
E3	143	E3	66	E3	143	E3	66
E3	159	E3	159				
E3	159	E3	267				
E3	159	E3	277	E3	159	E3	277
E3	159	E3	445				
E3	166	E3	66	E3	166	E3	66

E3	267	E3	267	E3	215	E3	284
E3	277	E3	211	E3	267	E3	284
E3	277	E3	267				
E3	277	E3	277	E3	277	E3	277
E3	284	E3	211	E3	284	E3	211
E3	288	E3	211	E3	288	E3	211
				E3	288	E3	277
E3	320	E3	155	E3	320	E3	155
E3	320	E3	211				
E3	320	E3	277				
E3	346	E3	159	E3	346	E3	159
				E3	410	E3	277
E3	410	E3	420	E3	410	E3	420
				E3	410	E3	430
E3	410	E3	445	E3	410	E3	445
E3	417	E3	430				
E3	417	E3	445				
E3	430	E3	122	E3	430	E3	122
E3	430	E3	267	E3	430	E3	267
E3	430	E3	277	E3	430	E3	277
E3	430	E3	430				
E3	430	E3	445				
E3	445	E3	267				
E3	445	E3	273				
E3	445	E3	445				
				E3	66	E2	547
E3	104	E2	114	E3	104	E2	114
E3	104	E2	241				
E3	104	E2	462				
E3	159	E2	362				
				E3	159	E2	473
E3	159	E2	376				
E3	159	E2	552				
E3	267	E2	114				
E3	267	E2	241				
E3	267	E2	362				
E3	267	E2	376	E3	267	E2	376
				E3	277	E2	241
E3	277	E2	368	E3	277	E2	368
E3	277	E2	376	E3	277	E2	376
E3	277	E2	387	E3	277	E2	376
E3	277	E2	466	E3	277	E2	387
E3	277	E2	490				
E3	277	E2	552	E3	277	E2	552
E3	277	E2	604				
E3	284	E2	376				
				E3	320	E2	241
E3	320	E2	362	E3	320	E2	376
E3	320	E2	376	E3	410	E2	363
				E3	410	E2	376
E3	430	E2	368				
E3	445	E2	114	E3	445	E2	114
E3	445	E2	241				
E3	445	E2	362				
E3	445	E2	363				
E3	445	E2	376	E3	445	E2	376
E3	445	E2	387				
E3	445	E2	462				
E3	277	E1-beta	80				
				E3	277	E1-beta	184
E3	277	E1-beta	227	E3	277	E1-beta	188
				E3	284	E1-beta	227
E3	430	E1-beta	184				

E3	430	E1-beta	188	E3	430	E1-beta	188
E2	362	E2	114	E2	362	E2	114
E2	362	E2	241	E2	362	E2	241
E2	362	E2	362	E2	362	E2	362
E2	362	E2	363	E2	362	E2	363
E2	362	E2	462	E2	362	E2	462
E2	363	E2	363				
E2	368	E2	362	E2	368	E2	362
E2	368	E2	387	E2	368	E2	387
E2	368	E2	473				
E2	376	E2	114	E2	376	E2	114
E2	376	E2	241	E2	376	E2	241
E2	376	E2	362				
E2	376	E2	363	E2	376	E2	363
E2	376	E2	368	E2	376	E2	368
E2	376	E2	376	E2	376	E2	376
E2	376	E2	387	E2	376	E2	387
E2	376	E2	462	E2	376	E2	462
E2	376	E2	473	E2	376	E2	473
E2	387	E2	362				
E2	387	E2	363	E2	387	E2	363
E2	387	E2	387				
				E2	462	E2	637
E2	466	E2	362				
E2	466	E2	637	E2	466	E2	637
E2	473	E2	362	E2	473	E2	362
E2	473	E2	363				
E2	473	E2	387	E2	473	E2	387
E2	473	E2	462	E2	473	E2	462
E2	473	E2	466	E2	473	E2	466
E2	490	E2	376	E2	490	E2	376
E2	490	E2	490	E2	490	E2	490
E2	547	E2	114	E2	547	E2	114
E2	547	E2	241	E2	547	E2	241
E2	547	E2	363	E2	547	E2	363
E2	547	E2	368				
E2	547	E2	387	E2	547	E2	387
E2	547	E2	462	E2	547	E2	462
				E2	547	E2	466
E2	547	E2	473	E2	547	E2	473
				E2	552	E2	114
E2	552	E2	362				
E2	552	E2	363	E2	552	E2	363
E2	552	E2	368	E2	552	E2	368
E2	552	E2	376	E2	552	E2	376
E2	552	E2	387				
E2	552	E2	473	E2	552	E2	473
E2	552	E2	547	E2	552	E2	547
				E2	604	E2	114
E2	604	E2	462	E2	604	E2	241
E2	604	E2	604	E2	604	E2	604
E2	637	E2	473				
E2	114	E1-beta	227				
				E2	241	E1-beta	68
E2	241	E1-beta	184				
E2	241	E1-beta	227	E2	241	E1-beta	227
E2	241	E1-beta	354	E2	241	E1-beta	354
E2	363	E1-beta	227	E2	363	E1-beta	227
E2	363	E1-beta	354				
				E2	376	E1-beta	68
E2	376	E1-beta	258	E2	376	E1-beta	227
E2	376	E1-beta	354				
E2	376	E1-beta	355				
E2	387	E1-beta	227	E2	376	E1-beta	355
E2	387	E1-beta	354	E2	387	E1-beta	227



E2	462	E1-beta	354	E2	473	E1-beta	184
E2	473	E1-beta	184	E2	473	E1-beta	227
E2	473	E1-beta	227	E2	473	E1-beta	258
E2	473	E1-beta	258	E2	552	E1-beta	75
E2	552	E1-beta	188				
E1-beta	52	E1-beta	80	E1-beta	76	E1-beta	68
E1-beta	68	E1-beta	76				
E1-beta	173	E1-beta	227	E1-beta	188	E1-beta	227
E1-beta	184	E1-beta	227	E1-beta	227	E1-beta	227
E1-beta	227	E1-beta	76				
E1-beta	227	E1-beta	229	E1-beta	354	E1-beta	258
E1-beta	258	E1-beta	354				
E1-beta	354	E1-beta	229	E1-beta	354	E1-beta	258

## I. Crosslinks in *hE2/E3BP* core as detected in crosslinking MS

Three different types of crosslinkings were obtained for *hE2/E3BP* core which were inter E2 monomers, inter E3BP monomers and between *hE2* and E3BP monomers. Spectral counts (number of spectra identified) were used as a way to discriminate between presence and absence of CoA states. Although many crosslinks were abundant or absent in either of the two states, a coherent effect could not be modelled. Thus, they were instead used primarily for proof checking some of the interesting crosslinks observed in *hPDHc*.

<b>E2-E2</b>		<b>with CoA</b>	<b>without CoA</b>
<b>Pos1</b>	<b>Pos2</b>	<b>spectral counts</b>	<b>spectral counts</b>
114	114		2
362	114	4	10
362	241		15
362	363	2	
362	462	11	22
362	637		2
368	241		10
368	473		9
368	362	9	
368	363	2	
368	387	1	
368	473	9	
376	114	16	16
376	241	21	23
376	362	12	12
376	363	9	7
376	368		3
376	376	11	16
376	387	18	16
376	462	21	
376	466	12	11
376	473	28	37
376	641		18
387	362	9	
387	363	4	
387	387	2	
387	462	6	5
387	466		2
396	368		3
396	376	21	28
462	363	1	
466	114	5	
466	241	10	
466	362	7	7
466	363	3	
466	387		1
466	462	6	
466	637	17	15
473	362	11	22
473	387	11	13
473	462	482	651
473	466	251	301
473	473	9	4
490	637	20	26
547	114		13

547	241		18
547	363	2	3
547	368	6	3
547	376	29	
547	387	5	5
547	462	12	7
547	466		6
547	473	60	92
552	241	23	35
552	363		3
552	363		3
552	368		3
552	376	55	58
552	387	12	8
552	466	5	
552	473	17	15
552	552	3	
552	637	10	15
604	241		11
604	376	5	5
604	462	6	4
604	473	4	10
604	552	1	2
604	604	1	
637	466		37
637	473	9	

<b>E3BP-E3BP</b>		<b>with CoA</b>	<b>without CoA</b>
<b>Pos1</b>	<b>Pos2</b>	<b>spectral counts</b>	<b>spectral counts</b>
79	194	1	2
120	120	4	6
120	173	12	12
120	213	13	10
120	227		5
120	321		5
120	321		5
139	120		5
145	120		9
145	173	14	2
145	194		2
155	173		17
155	194		4
155	223	8	
155	321		3
172	488		4
173	223	10	8
173	227	7	6
194	120		30
194	145	5	4
194	155		17
194	172	13	7
194	173	26	
194	194	28	30
194	213	108	147
194	218	5	
194	223	92	107
194	227	40	38
194	321	10	9
194	325	4	5
194	334	1	
194	379		10
194	384	3	
194	394		9
194	450		20
194	460	21	25

194	488		24
194	491	5	4
213	173	6	5
213	223	29	
213	227		3
213	289		10
295	120	6	4
295	173	14	13
295	194	25	29
295	213	5	9
295	223	4	
295	295		1
295	450		6
295	460	9	8
295	491	2	
			7
314	450	52	59
314	460	125	161
321	321		2
321	173	3	
325	194	4	13
325	394	90	92
325	460	4	4
341	194	8	
341	213	6	6
341	379		6
341	384	4	5
341	460		5
356	194		2
379	120	3	4
379	173	7	7
379	213	6	5
379	321	3	
379	491	4	
384	213	3	
394	491	2	
450	120		7
450	145	1	
450	155	8	12
450	173	10	14
450	213	5	11
450	223	2	3
450	227	6	3
450	321	40	31
450	334	2	
450	379	4	4
450	394	3	
450	450	4	3
450	460	11	10
450	488		56
450	491	62	66
460	120	5	7
460	155	10	11
460	172		3
460	173	9	9
460	213	7	7
460	223		4
460	227	5	
460	321	13	16
460	325	2	
460	379	6	5
460	384	4	4
460	394	2	5
460	460	13	14
460	488		15
460	491	6	7
488	289		25

488	488		7
491	120		3
491	213	2	

<b>hE2-E3BP</b>		<b>with CoA</b>	<b>without CoA</b>
<b>hE2</b>	<b>E3BP</b>	<b>spectral counts</b>	<b>spectral counts</b>
114	155		13
114	173	5	4
114	194		22
114	460		10
114	488		11
241	120		18
241	223	8	6
241	295		13
241	314	4	
241	325		6
241	379		7
241	394	4	
362	120	6	4
362	213	4	
362	394	3	
362	450	6	7
362	460	4	7
363	120	2	1
363	173	2	
363	194	10	6
363	213	2	
363	289	3	
363	295	3	
363	341	1	
363	384	1	
363	460	6	5
363	488	5	4
368	120	20	
368	194	5	4
376	120	15	11
376	145	2	
376	173		7
376	194	33	37
376	213	5	9
376	223	4	
376	227	6	5
376	295		9
376	321	8	4
376	325	1	6
376	341	3	5
376	384	3	
376	394	7	9
376	460	6	8
386	488	18	10
376	491	2	
387	120	7	
387	155		3
387	173	1	
387	194	7	6
387	213	3	
387	289	7	8

387	295	4	2
387	325	1	
387	341	3	2
387	394	2	
387	450	9	5
387	460	9	8
387	488	11	
440	295	7	6
462	120	16	14
462	145	5	7
462	155	19	14
462	172	5	2
462	173	7	11
462	194	31	37
462	213	5	7
462	314		15
462	341		4
462	450	10	
462	460	14	9
466	120	5	6
466	194	17	16
466	213	5	2
466	223	4	
466	450	9	10
466	460	8	12
466	488		14
473	120	5	14
473	145	2	
473	155	14	14
473	194	24	50
473	213		9
473	223	7	
473	227	5	
473	295	20	4
473	341	6	6
473	379	7	2
473	384	6	
473	394	4	1
473	450	8	10
473	460	19	12
473	491	4	6
547	173	5	5
547	194	20	19
547	491	11	12
552	173		7
552	194		28
552	213		11
552	223	2	
552	227	4	8
552	295	5	2
552	341	3	5
552	379		5
552	384	5	
552	450	10	11
552	460	6	
552	491	5	4
604	194		5
604	450	3	4

604	460	17	38
637	120	4	6
637	194		16
637	213	1	2

## **Curriculum Vitae**

### **Sabin Prajapati**

Department of Molecular Enzymology  
Georg-August-Universität-Göttingen  
Email: sprajap@gwdg.de

#### Education:

MSc. Molecular Biology (2013), IMPRS, Georg-August-Universität-Göttingen

BSc. Biochemistry (2010), Pokhara University, Nepal

#### Scholarships:

2011 - 2012: Stipend of the Excellence Foundation for the promotion of the Max Planck Society.

2005 - 2009: Scholarship for the academic excellence. Biochemistry Department, Universal Science College affiliated to Pokhara University, Nepal.

#### Publication:

“Characterization of Choreospondias axillaries (Lapsi) fruit protease.” Prajapati et al. International journal of life sciences, 2010.

Link: <http://www.nepjol.info/index.php/IJLS/article/view/2386/2152>

#### Research experiences:

02.05.2013 – 18.10.2016, “Structural and biophysical characterization of human pyruvate dehydrogenase multi-enzyme complex.” PhD Thesis under supervision of Prof. Dr. Kai Tittmann.

01.10.2012 – 31.03.2013, “Structure and dynamics of the core of human pyruvate dehydrogenase multi-enzyme complex.” MSc. Thesis under supervision of Prof. Dr. Kai Tittmann.



02.05.2012 - 22.06.2012, "DNA-catalyzed labeling of RNA." under supervision of Prof. Dr. Claudia Höbartner.

27.02.2012 - 30.04.2012, "Expression, purification and functional analyses of all three catalytic subunits of *h*PDHc." under supervision of Prof. Dr. Kai Tittmann.

03.01.2012-24.02.2012, "Screening of alignment media for the residual dipolar coupling (RDC) based study of recognition dynamics between ubiquitin and Dsk2." under supervision of Prof. Dr. Christian Griesinger.

01.05.2009 – 29.10.2009, "Characterization of *Choreospondias axillaries* (Lapsi) fruit protease." B.Sc. Thesis under supervision of Prof. Dr. Vishwanath P. Agrawal.

#### Research skills:

PCR, protein overexpression and purifications, high cell density biofermentation, steady state and pre-steady state enzyme kinetics, CD spectroscopy, isothermal titration calorimetry, negative staining electron microscopy, nuclear magnetic resonance spectroscopy, small angle X-ray scattering and atomic force microscopy,

Participated in various method courses in the field of biophysics like cryo-EM, X-ray crystallography, NMR, EPR, mass spectrometry, FRET, analytical ultracentrifugation and in cell biology like RNA-microarray, quantitative PCR, in-situ hybridization in frog embryo, mouse fibroblast and embryonic stem cell culture and confocal microscopy.

#### Contributions in scientific meetings:

Poster presentation: Why stop breathing? Kyoto winter school (2016), HeKKSaGon, Japan

Talk: Structural studies of the human pyruvate dehydrogenase multi-enzyme complex (*h*PDHc). SFB 860 retreat (2015)

Poster presentation and a talk: Understanding the breathing phenomenon observed in an icosahedral core of pyruvate dehydrogenase multi-enzyme complex. PhD retreat of GGNB Biomolecules program, 2014 (Best poster award)