

**Functional Characterization of Recombinant  $\alpha$ -Crystallin of**

***Mycobacterium tuberculosis* H37Rv**

**THESIS**

Submitted in partial fulfillment

of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

by

**GAUTAM KRISHNAN**

Under the Supervision of

**PROFESSOR UTPAL ROY, Ph.D.**

Under the Co-supervision of

**Dr. SANTANU DATTA, Ph.D.**



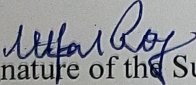
**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**2019**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**CERTIFICATE**

This is to certify that the thesis entitled “**Functional Characterization of Recombinant  $\alpha$ -Crystallin of *Mycobacterium tuberculosis* H37Rv**” and submitted by Gautam Krishnan, ID No. 2010PHXF0813G for the award of Ph.D. of the Institute embodies original work done by him under my supervision.

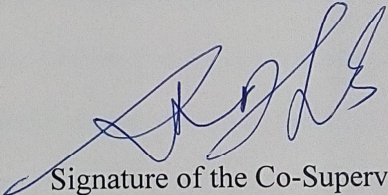
  
Signature of the Supervisor

**UTPAL ROY, Ph.D.**

Professor,

Dept. of Biological Sciences

Date: 13-6-2019

  
Signature of the Co-Supervisor

**SANTANU DATTA, Ph.D.**

Chief Scientific Officer,

Bug Works

Date: 13th June 2019

*Dedicated to*

*My Esteemed Mentor Dr. Daisaku Ikeda*

## ACKNOWLEDGEMENTS

I sincerely thank the esteemed Honourable Vice Chancellor Professor Souvik Bhattacharya, BITS Pilani and the Director, Professor G. Raghurama, ex-Director Late Professor Sanjeev Aggarwal, and former acting Director, Prof K E Raman, BITS Pilani KK Birla Goa Campus.

I am deeply indebted to my esteemed guide Dr. Utpal Roy, Professor, Department of Biological Sciences without whose involvement and timely intervention this thesis could never have seen the light of the day.

I extend my sincere thanks to my co-guide Dr. Santanu Datta who enabled me to carry out a part of my thesis work in his laboratory at Bugworks, C-CAMP, Bangalore. I wholeheartedly thank his staff, Arindam, Dr. Nainesh, Anirudh and Deepa at C-CAMP for assisting me with their technical expertise and use of their facilities.

My sincere thanks to the DRC convener Professor Dibakar Chakrabarty and past DRC conveners. I also thank my DAC members Professor Anasuya Ganguly, and Professor Vijayashree Nayak for being with me throughout this journey with practical guidance and advice.

I thank the former Dean ARD, Professor P.K. Das, the present Associate Dean AGSRD, Professor Bharat Deshpande, and staff member AGSRD, Pratap Behera. I thank the HOD of Biological Sciences, Prof. Srikanth Mutnuri, all the ex-HODs Prof. Meenal Kowshik and Prof. Judith Braganca for always believing in and morally supporting me. I thank all my dear PhD colleagues for helping me at crucial times and lab staff Kamna, Pallavi, Mahadev and Mahaling.

I also acknowledge Prof. Mridula Bose of Patel Chest Institute, New Delhi for providing *M. tb* genomic DNA, Mr Bala Madiletti and his team at CDFD (Centre for DNA fingerprinting and Diagnostics), Hyderabad for the Circular Dichroism analysis.



I deeply thank my wife Radhamani and my family for always supporting me through all my decisions. I deeply thank my parents, my mother in particular, for always encouraging me to aim for higher goals and aspirations in life.

Last, but not the least, I would like to extend my deepest gratitude to my life's esteemed mentor Dr. Daisaku Ikeda who guided and inspired me to pursue my college dream of doing research for the sake of humankind.

Gautam Krishnan

## **Abstract**

Tuberculosis is an ancient disease that has reappeared in the developed world due to rapid urbanization, immigration and is a major threat in the developing world due to poor immune and nutritional status. It is caused by *Mycobacterium tuberculosis* (*M. tb*). There is a need to develop new vaccines and diagnostics in view of the variable efficacy of the BCG vaccine in adults and the phenomenon of latent tuberculosis (TB). The HIV epidemic also plays a role in suppressing the immune system; leading to reactivation of the tubercle bacilli even in the latent form. The aim of this study was to clone, express and characterize important proteins associated with latent TB and immune evasion with the aim of generating new data for important biomarker proteins for diagnosis. In lieu of this, attempts were made to clone and express three *M. tb* proteins; superoxide dismutase A (Sod A), superoxide dismutase C (Sod C) and  $\alpha$ -crystallin (Acr), known to be upregulated in a macrophage model developed in human monocytic cell line (THP-1). In the present study, a total of five constructs were obtained; two of *sod A*, one of *sod C* and two of *acr*. While undertaking gene-cloning experiments, the success of cloning experiments depends upon a number of factors including quality of DNA, units of restriction enzymes and ligase, molar ratios of vector and insert, high efficiency of competent cells.

Traditional ligation dependent cloning methods entail use of DNA modifying enzymes that are inefficient in spite of being expensive; affecting overall cloning success. Parameter analysis was used in Microsoft Excel to propose key observations and recommendations for optimum results. Based on the above data, a broad model for prediction of cloning success was developed which included eighteen parameters. This was applied to fifteen cases out of which nine were negative and six were positive outcomes. In addition, we used different classifiers in MATLAB and Clone. R to predict

possibility of obtaining a clone. In MATLAB, the Simple Tree classifier was most optimum with accuracy of 93.3% and precision and specificity of 1.0, while in Clone. R, Random forest classifier was the best. These might be useful tools for successful planning of any cloning experiment. An approach of directional cloning using two different restriction enzymes and phosphatase was followed. This required screening of fewer than ten recombinants to get a clone. Using this strategy, three Mycobacterial genes *sod A*, *sod C* and *acr* were cloned successfully.

***sod A* construct (minus His-tag):** The *sod A* gene was PCR-amplified from genomic DNA of H37Rv and cloned into pET21a with sticky end approach. The clone of *sod A* gene was tested for production of soluble protein. Optimisation of expression conditions such as temperature, time of induction, and optical density for induction with Iso-propyl-thio-galactopyranoside (IPTG) was carried out to achieve this. It was observed that 2 mM IPTG induction overnight at 25°C produced soluble protein.

***acr* Construct 1 (C-terminal His-tag):** *Mycobacterium α-crystallin (acr)* gene was PCR-amplified from genomic DNA of *M. tb* H37Rv and cloned into the BamHI-XhoI site of pET21a. The *acr* clone was sequenced and identity of protein verified using denaturing Nickel-NTA His-tag purification. Growth condition of one clone was optimized to produce Acr in a soluble form at 25°C. The dose-dependent chaperone activity was shown by inhibition of thermal aggregation of citrate synthase (CS) at 45°C, which improved upon boiling which is reported for Acr.

***acr* Construct 2 (N-terminal His-tag):** *Mycobacterium α-crystallin (acr)* gene was cloned and expressed in *E. coli*. The recombinant Acr protein was purified by Nickel-NTA resin. The oligomeric state of Acr was confirmed by gel filtration chromatography using Sephacryl S-200 and Native-PAGE (8-16%). Chaperone activity studies were performed with insulin at different mole ratios of Acr with 2 types of samples; His-tag

elutes (**H**) and His-tag elutes with gel filtration (**G**). Polynomial graphs with equation  $ax^n + bx^{n-1} + c$  was plotted where;  $n$  is primarily 2<sup>nd</sup> or 3<sup>rd</sup> order. This could be used to predict activity. It was observed that ratio of different sizes of oligomers (9 to 24 mers) had a significant effect on chaperone activity. Using mole ratio of Acr for both **H** and **G** samples to insulin substrate and ratio of oligomers, the number of Acr molecules binding to insulin as a model substrate was determined. It was found that if 1.5% of the insulin chain is covered completely by the **G** samples, aggregation is completely inhibited as compared to 6% with **H** samples. Pre-heat treatment studies were carried out at 37°C, 60°C and 70°C. Far-Ultraviolet circular dichroism (UV-CD) analysis provided fresh insights into the role of  $\beta$  sheets and  $\alpha$  helices in activity; particularly in **H** samples; suggesting a reversible transition from  $\alpha$  helices to  $\beta$  sheets. A functional model for binding of Acr to insulin B chain was formulated which incorporated 4 types of secondary structure molecules. This could be a useful tool to analyse *in vitro* preparations of recombinant Acr and build more consensus on the structure-activity relationship; particularly in terms of oligomeric ratios.

## TABLE OF CONTENTS

S. No.	Title	Page No.
1	Acknowledgements	iv
2	Abstract	vi
3	List of Tables	x
4	List of Figures	xii
5	List of Abbreviations / Symbols	xviii
6	Introduction and Review of Literature	1
7	Gaps in Research	6
8	Aims and objectives	8
9	Chapter 1: Cloning of Mycobacterial genes <i>acr</i> , <i>sod A</i> and <i>sod C</i>	9
10	Chapter 2: Expression and purification studies of Acr, Sod A and Sod C	52
11	Chapter 3: Activity Assays of <i>Mycobacterium</i> $\alpha$ -Crystallin (Acr)	83
12	Chapter 4: Structure and activity relationship of Acr	118
13	Thesis Conclusions	155
14	Future Scope	156
15	List of References	157



## LIST OF TABLES

S. No.	Title	Page No.
	<b>Chapter 1</b>	
1.1.1.1	Plasmid yield calculation for 6 batches of 1.5 ml culture of pET21a vector by alkaline lysis - Birnboim and Doily method	27
1.1.1.2	Plasmid yield calculation for 6 batches of 5 ml culture of pET21a vector by alkaline lysis - Birnboim and Doily method	27
1.1.1.3	Plasmid yield calculation for 4 batches of 10 ml culture of pET21a vector - Macherey Nagel Kit method	28
1.1.1.4	Plasmid yield calculation for 8 batches of 25 ml culture of pET21a vector - alkaline lysis and Macherey Nagel Kit	28
1.1.1.5	Plasmid yields for different volumes of culture and ratios of lysis buffer to culture volume	29
1.1.1.6	Documentation of summary of plasmid yield calculation in Excel	30
1.1.1.7	Calculation of number of molecules to be lysed per $\mu$ l of lysis buffer	31
1.1.2	Restriction digestion summary of different batches	32
1.1.3	Summary of all cloning attempts	34
1.2.1	Summary of MATLAB Classifier analysis	35
1.5	Development of software to estimate DNA	48
	<b>Chapter 2</b>	
2.6.1	Oligomer ratio calculation of batch 1	76
2.6.2	Oligomer ratio calculation of batch 2	78
2.7	Expression levels of different <i>acr</i> constructs	81
	<b>Chapter 3</b>	
3.2.2.1	Molecular level interaction of Acr with insulin B chain at 60°C	94
3.2.2.2	Plot of Acr non-gel-filtered monomer molecules versus % insulin B chain covered	96
3.2.2.3	Plot of Acr non-gel-filtered Nonamer molecules versus % insulin B chain covered	97
3.2.2.4	Comparison of chaperone activity of different constructs	109
3.3.3	Chaperone activity of <b>H</b> sample versus Mole ratio at 60°C	110
3.3.4	Chaperone activity of <b>H</b> sample versus Mole ratio at 37°C	110

3.3.5	Chaperone activity of <b>G</b> samples, run 1, pool 1	110
3.3.6	Chaperone activity of <b>G</b> samples, run 1, pool 2	110
3.3.7	Chaperone activity of <b>G</b> samples, run 2	110
3.3.8	Re-calculation of polynomials	111
3.3.9	Re-calculation of values of inhibition for <b>H</b> samples at 60°C	112
3.3.10	Re-calculation of values of inhibition for <b>G</b> samples	113
	<b>Chapter 4</b>	
4.1.1.1	Consolidated CD profile of proportion of molecules of secondary structure of <b>H</b> samples	127
4.1.1.2	Consolidated CD profile of proportion of molecules of secondary structure of <b>G</b> samples	128
4.1.2	Comparison of secondary structure of <b>H</b> and <b>G</b> samples with predicted sequence obtained from <b>PSIPRED</b>	128
4.2	Calculations of molecular level binding of Acr to insulin B chain	130
4.3	A comparison of <b>H</b> and <b>G</b> samples in terms of 4 secondary structures to achieve 95% inhibition of insulin aggregation	143
4.4.1	Melting temperature calculation	145

## LIST OF FIGURES

S. No.	Title	Page No.
	<b>Chapter 1</b>	
I	The Vicious cycle of TB	1
1.1.1.1	Plot of plasmid yield versus ratio of lysis buffer $b_1$ to culture volume b	28
1.1.1.2	Plot of plasmid yield versus ratio of lysis buffer $b_2$ to culture volume b	29
1.1.1.3	Plot of plasmid yield versus ratio of lysis buffer $b_3$ to culture volume b	29
1.1.2.1	Plot of time for completion of restriction digestion versus no. of molecules of plasmid DNA	32
1.1.2.2	Plot of units of enzyme for restriction digestion versus no. of plasmid molecules	33
1.2.1.1	Simple Tree classifier (ROC Curve, Confusion matrix)	35
1.2.1.2	Linear Discriminant classifier (ROC Curve, Confusion matrix)	36
1.2.1.3	Logistic Regression (ROC Curve, Confusion matrix)	36
1.2.1.4	Support Vector Machines classifier (ROC Curve, Confusion matrix)	37
1.2.1.5	k-Nearest Neighbour classifier (ROC Curve, Confusion matrix)	37
1.2.1.6	Random forest classifier (ROC Curve, Confusion matrix)	38
1.2.2.1	Precision, Recall, Accuracy and AUC using Clone. R	38
1.2.2.2	Plot of effect of important parameters on Accuracy of model in Clone. R	39
1.2.2.3	Effect of units of enzyme on output in Clone. R	39
1.2.2.4	Effect of reaction volume on output in Clone. R	40
1.3.1.1	PCR of <i>acr</i> from genomic DNA / pET21a	40
1.3.1.2	Colony PCR of <i>acr</i> clones	41
1.3.1.3	Restriction digestion of <i>acr</i> clones with BamHI and XhoI	41
1.3.1.4	Nucleotide Sequence of <i>acr</i> clone #1	42
1.3.1.5	Nucleotide Sequence of <i>acr</i> clone #1 Histidine tag and stop codon.	42

1.3.2.1	PCR of <i>sod A</i> with GTG in start codon (non-directional) from H37Rv genomic DNA	43
1.3.2.2	Colony PCR cloning of recombinants of <i>sod A</i> with GTG in start codon	43
1.3.2.3	Restriction digestion of recombinant <i>sod A</i> with GTG in start codon	43
1.3.2.4	Sequencing of recombinant <i>sod A</i> with GTG in start codon	44
1.3.3.1	PCR of recombinant <i>sod C</i>	44
1.3.3.2	Colony PCR of recombinant <i>sod C</i>	44
1.3.3.3	Restriction digestion of <i>sod C</i> clones	45
1.3.3.4	Sequencing of <i>sod C</i> clone #5	45
1.3.4.1	Colony PCR of modified <i>sod A</i> with ATG start codon	46
1.3.4.2	Restriction digestion of modified <i>sod A</i> clones with ATG start codon	46
1.3.4.3	Sequencing of <i>sod A</i> clone using reverse primer T7	46
1.4.1	Restriction digestion of <i>acr</i> -pET28a clones	47
1.4.2	Colony PCR of <i>acr</i> recombinants	47
1.4.3	PCR of <i>acr</i> recombinant plasmids	48
1.5	Gel picture of sample used to test software	48
	<b>Chapter 2</b>	
2.1	Expression of <i>sod A</i> (GTG) in the start codon	61
2.4.1.1	Expression of Acr in BL21DE3 cells	62
2.4.1.2	25 ml sonicated protein	63
2.4.2	200 ml after sonification	63
2.4.3	Nickel-NTA purification of Acr 25 ml cell pellet lysed in 8M Urea	64
2.4.4	Soluble Acr expression at 0.8 O. D at 25°C and 0.05 to 0.20 mM IPTG	65
2.4.5	Nickel-NTA purification of Soluble Acr	66
2.5.1.1	Expression of <i>acr</i> -pET28a #3, #6	67

2.5.1.2	0.5 mM IPTG Expression Gel <i>acr</i> -pET28a	67
2.5.1.3	Expression of <i>acr</i> -pET28a to check for soluble protein	68
2.5.2.1	Nickel-NTA purification of <i>acr</i> -pET28a	68
2.5.2.2	Different batch of Nickel-NTA purification of <i>acr</i> -pET28a	69
2.5.3.1	Gel Filtration Chromatogram Run 1: X Axis: UV 280 nm; Y Axis: Elution time (mins.)	70
2.5.3.2	Gel Filtration Chromatogram Run 2: X Axis: UV 280 nm; Y Axis: Elution time (mins.)	70
2.5.3.3	Bio-Rad Standards Chromatogram	71
2.5.3.4	10% Reducing SDS-PAGE analysis of gel filtration Run 1	71
2.5.3.5	10% Reducing SDS PAGE analysis of gel filtration Run 2	71
2.5.4.1	Tryptic digest of recombinant Acr and matching of peptide fragments using MASCOT	73
2.5.4.2	Molecular Mass determination of recombinant Acr/clone #3	74
2.6.1	Native-PAGE of his-tag elute <b>H</b> and gel filtration elute <b>G</b>	75
2.6.2	Plot of log molecular weight of the different forms of BSA in Native-PAGE	75
2.6.3	Different batch Native-PAGE analysis	77
2.6.3.2	Different batch gel filtration analysis	77
2.6.4	Plot of Log Mol. wt. versus distance migrated	78
	<b>Chapter 3</b>	
3.1.1	Thermal aggregation assay of Acr using citrate synthase	89
3.1.2	Dose-dependent thermal aggregation assay of Acr using citrate synthase	90
3.2.1	Thermal aggregation assay of N-terminal His-tag Acr with CS	92
3.2.2.1.1	Insulin aggregation assay at 44 $\mu$ M Acr	92
3.2.2.1.2	Insulin aggregation assay at 37.5 $\mu$ M Acr	93
3.2.2.1.3	Insulin aggregation assay at 12 $\mu$ M Acr	94
3.2.2.2.1	Polynomial graph of % inhibition versus Mole ratio at 60%	95



3.2.2.1.4	Insulin aggregation at 12 $\mu$ M gel-filtered and 30 $\mu$ M non-gel-filtered Acr	95
3.2.2.1.5	Dose dependent chaperone activity of <b>H</b> Acr at 60°C	96
3.2.2.2.2	Polynomial graph of % inhibition of Acr monomer versus % of insulin B chain covered at 60°C	97
3.2.2.2.3	Polynomial graph of % inhibition of Acr nonamer versus % of insulin B chain covered at 60°C	98
3.2.2.2.4	Polynomial plot of mole ratio of <b>H</b> sample versus % inhibition of insulin	98
3.2.2.2.5	Simulation plot of monomer versus % insulin B chain covered	99
3.2.2.2.6	Simulation plot of higher ratio of 9 mers versus % insulin B chain covered	99
3.2.2.2.7	Simulation plot of lower ratio of 9 mers versus % insulin B chain covered	100
3.2.2.3.1	Inhibition of insulin aggregation using Acr (18 $\mu$ M) and insulin 0.5 mg/ml was carried out at 37°C	100
3.2.2.3.2	Inhibition of insulin aggregation using Acr (11 $\mu$ M) <b>H</b> samples, without gel filtration and insulin 1 mg/ml at 37°C	101
3.2.2.4.1	Pre-heat treatment studies and chaperone activity of <b>G</b> samples; pool 2 run 2	102
3.2.2.4.2	Chaperone activity of <b>H</b> at 37°C after pre-heat treatment (37, 60°C)	102
3.3.1.1	Chaperone activity of <b>H</b> at 37°C	103
3.3.1.2	Polynomial plot of mole ratio of <b>H</b> sample versus % inhibition of insulin at 37°C	104
3.3.1.3	Chaperone activity of <b>G</b> sample pool 1, run 1	104
3.3.1.4	Polynomial plot of mole ratio of <b>G</b> sample run 1, pool 1 versus % inhibition of insulin at 37°C	104
3.3.1.5	Chaperone activity of <b>G</b> sample run 1, pool 2	105
3.3.1.6	Polynomial plot of mole ratio of <b>G</b> sample run 1, pool 2 versus % inhibition of insulin at 37°C	105
3.3.1.7	Chaperone activity of <b>G</b> sample run 2	106
3.3.1.8	Polynomial plot of mole ratio of <b>G</b> sample run 2 versus % inhibition of insulin at 37°C	106
3.3.1.9	Overlay of activity of <b>H</b> samples at 37°C and 60°C	107
3.3.1.10	Overlay of activity of <b>H</b> samples versus activity of <b>G</b> samples at 37°C	107
3.3.2.1	Assay of prevention of insulin B chain aggregation using <b>H</b> samples	108
3.3.2.2	Assay of prevention of insulin B chain aggregation using <b>G</b> samples	109

3.3.11	Re-plotting of <b>H</b> samples in terms of molecules and % of insulin covered at 60°C	113
3.3.12	Re-plotting of <b>G</b> samples in terms of molecules and % of insulin covered at 37°C	113
	<b>Chapter 4</b>	
4.1.1 (A)	Far-UV CD analysis of <b>H</b> sample at 25°C without pre-heat treatment	124
4.1.1 (B)	Far-UV CD analysis of <b>H</b> sample at 37°C without pre-heat treatment	124
4.1.1 (C)	Far-UV CD analysis of <b>H</b> sample after pre-heat treatment for 15 mins. at 37°C	124
4.1.1 (D)	Far-UV CD analysis of <b>H</b> sample after pre-heat treatment for 15 mins. at 60°C	124
4.1.1 (E)	Far-UV CD analysis of <b>H</b> sample after pre-heat treatment for 15 mins. at 70°C	125
4.1.1 (F)	Overlay of Far-UV CD analysis of <b>H</b> samples	125
4.1.2 (A)	Far-UV CD analysis of <b>G</b> sample at 37°C without pre-heat treatment	126
4.1.2 (B)	Far-UV CD analysis of <b>G</b> sample at 37°C with pre-heat treatment	126
4.1.2 (C)	Far-UV CD analysis of <b>G</b> sample with pre-heat treatment at 60°C	126
4.1.2 (D)	Far-UV CD analysis of <b>G</b> sample with pre-heat treatment at 70°C	126
4.1.2 (E)	Overlay of Far-UV CD Analysis of <b>G</b> samples	127
4.3 (A)	Plot of molecules of $\beta$ sheets versus % inhibition for <b>G</b> samples without pre-heat treatment	132
4.3 (B)	Plot of $\beta$ sheets versus % inhibition for <b>G</b> sample (1 $\mu$ M) with pre-heat treatment	132
4.3 (C)	Plot of $\alpha$ helices versus % inhibition for <b>G</b> samples without pre-heat treatment	133
4.3 (D)	Plot of $\alpha$ helices versus % inhibition for <b>G</b> sample (1 $\mu$ M) with pre-heat treatment	133
4.3 (E)	Plot of random coils versus % inhibition for <b>G</b> samples without pre-heat treatment	134
4.3 (F)	Plot of random coils versus % inhibition for <b>G</b> sample (1 $\mu$ M) with pre-heat treatment	134
4.3 (G)	Plot of turns versus % inhibition for <b>G</b> samples without pre-heat treatment	135
4.3 (H)	Plot of turns versus % inhibition for <b>G</b> sample (1 $\mu$ M) with pre-heat treatment	136
4.3 (I)	Plot of molecules of $\beta$ sheets versus % inhibition for <b>H</b> samples	136
4.3 (J)	Plot of $\beta$ sheets for <b>H</b> sample (11 $\mu$ M) with pre-heat treatment	137

4.3 (K)	Plot of $\alpha$ helices versus % inhibition for <b>H</b> samples without pre-heat treatment	137
4.3 (L)	Plot of $\alpha$ helices versus % inhibition for <b>H</b> sample (11 $\mu$ M) with pre-heat treatment	138
4.3 (M)	Plot of turns versus % inhibition for <b>H</b> samples without pre-heat treatment	138
4.3 (N)	Plot of turns versus % inhibition for <b>H</b> sample (11 $\mu$ M) with pre-heat treatment	139
4.3 (O)	Plot of random coils versus % inhibition for <b>H</b> samples without pre-heat treatment	139
4.3 (P)	Plot of random coils versus % inhibition for <b>H</b> sample (11 $\mu$ M) with pre-heat treatment	140
4.3 (Q)	Overlay of $\beta$ sheets versus % inhibition for <b>H</b> and <b>G</b> samples without pre-heat treatment	140
4.3 (R)	Overlay of $\alpha$ helices versus % inhibition for <b>H</b> and <b>G</b> samples without pre-heat treatment	141
4.3 (S)	Overlay of random coils versus % inhibition for <b>H</b> and <b>G</b> samples without pre-heat treatment	141
4.3 (T)	Overlay of turns versus % inhibition for <b>H</b> and <b>G</b> samples without pre-heat treatment	142
4.3 (U)	Schematic Presentation of binding of Acr to insulin	143
4.4	Thermal shift assay of affinity <b>H</b> and <b>G</b> samples	145

## **LIST OF ABBREVIATIONS / SYMBOLS**

Acr:  $\alpha$ -crystallin protein

*acr*:  $\alpha$ -crystallin gene

AIDS: Acquired Immunodeficiency Syndrome

ANOVA: Analysis of Variance

AUC: Area under the Curve

BCG: Bacillus Calmette - Guérin

BSA: Bovine Serum Albumin

CART: Classification and regression tree

CFP-10: Culture filtrate protein-10

CIAP: calf intestinal alkaline phosphatase

DTT: Dithiothreitol

EDTA: Ethylene diamine tetra acetic acid

ESAT - 6: Early Secretory Antigen

FP: False Positive

FN: False Negative

GLM: Generalised Linear Model

GST: Glutathione-S-Transferase

HIV: Human Immunodeficiency Virus

IFN- $\gamma$ : Interferon Gamma

IL-1  $\beta$ : Interleukin-1

IL-12: Interleukin-12

IL-2: Interleukin-2

IPTG: Iso Propyl-thio-galactopyranoside

k-NN: k-Nearest Neighbour algorithm

LDA: Linear discriminant analysis

*M. tb*: *Mycobacterium tuberculosis*

MALDI-TOF: Matrix Assisted Laser Desorption/Ionisation - Time of Flight

MDG: Millennium Development Goals

O. D: Optical Density

PBS: Phosphate buffered saline

PCR: Polymerase Chain Reaction

pET: Plasmids for Expression by T7 Polymerase

PSIPRED: PSI-blast based secondary structure prediction

RNase A: Ribonuclease A

ROC: Receiver Operating Characteristics

SDS-PAGE: Sodium dodecyl sulphate Polyacrylamide Gel Electrophoresis

SOD A: Iron (Fe) factored superoxide dismutase

*sod A*: Iron (Fe) factored superoxide dismutase gene



SOD C: Copper- Zinc (Cu-Zn) factored superoxide dismutase

*sod C*: Copper- Zinc (Cu-Zn) factored superoxide dismutase gene

SVM: Support Vector Machines

TB: Tuberculosis

TN: True Negative

TP: True Positive

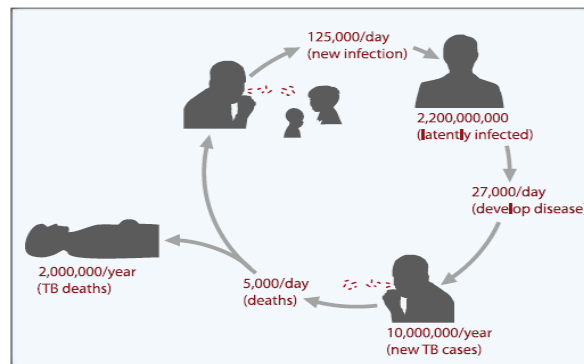
TST: Tuberculin skin test

UV-CD: Ultraviolet circular dichroism

WHO: World Health Organisation

## INTRODUCTION & REVIEW OF LITERATURE

In countries of the developing world, infectious diseases like malaria, tuberculosis (TB), diphtheria cause a high proportion of all deaths varying from as high as 50% in low income countries to 25% in middle income countries (Dye 2015). Millennium development goal (MDG) number 6 was listed as eradication of AIDS, TB and malaria, outlined as part of the MDG for the years 2000 - 2015. Towards this goal, thirty seven million lives were saved between the years 2000 and 2013 due to TB prevention, diagnosis and treatment interventions as listed in the MDG report. The global tuberculosis mortality rate fell by 45% and the prevalence rate by 41% between the years 1990 and 2013 (2015 MDG Report). However, in spite of this, the challenge to manage and cure TB across the world, remains as stiff as ever as shown in **Figure I** below (Kaufmann 2010). The World Health Organisation (WHO) has announced the goal of end TB by 2035 with goals of 95% death reduction compared to 2015, 90% reduction by 2035 in TB incidence rate compared to 2015 and Zero TB affected families facing catastrophic costs (<http://www.who.int/tb/strategy/end-tb/en/>).



**Figure I:** The Vicious Cycle of TB: The figure depicts daily numbers of infections with *Mycobacterium tuberculosis* (*M. tb*), new cases of TB, cases of death because of TB, and numbers of latently *M. tb*-infected individuals, new TB cases annually, and TB deaths annually (Kaufmann 2010).

TB is the oldest known disease to mankind. *M. tb* is an intracellular bacterium that can reside in macrophages. Originally a soil actinomycete, it jumped to humans during the period of domestication of cattle. The oldest case of TB dates 4000 years ago (Smith 2003) and 9000 years ago from human bone samples under the sea (<https://www.newscientist.com/article/dn14941-oldest-cases-of-human-tb-found-beneath-the-sea/>). Being an obligate intracellular bacterium, it resides in macrophages and causes infection when the bacterial load is high; it infects lung airways and multiplies actively causing death to the host. However, in one case out of three, it retreats into a dormant phase; in the macrophage in caseous granulomas. In latent infection, these macrophages are suppressed and the disease can be activated in case of suppression of the overall immune response like in immunocompromised individuals or HIV infected individuals. There are two key immune response proteins, Interferon gamma (IFN- $\gamma$ ) and Interleukin-12 (IL-12) that play a key role in allowing clearance of macrophages (Abbas and Lichman 2007). When the immune system of the host is compromised, it resurfaces into an active form causing disease. Almost 10% of latent TB cases are susceptible to re-infection due to a variety of factors like suppression of the immune system due to HIV co-infection, parasitic or helminthic infections. TB remains an active challenge due to the emergence of multi-drug resistance TB with the rise of HIV infection (Espindola et al. 2017).

## **LINK OF TB TO SOCIO-ECONOMIC FACTORS**

A unique feature of spread of TB has been the role of environmental factors. In the 19<sup>th</sup> century, the role of overcrowded conditions, sanitation and malnutrition was widely associated with more number of cases; particularly, in Europe. A debate over whether the disease spread through human contact in such conditions or genetically acquired was settled through the discovery of Koch's postulates by Robert Koch. The existence of the bacterium, *M. tb* in many cases proved the theory to be right. Once this was established, the focus shifted towards treatment of the disease, though it was not until the 1940's that the first antibiotics;

streptomycin, rifampicin and isoniazid were produced after the success of the infant vaccine BCG, introduced in 1921. Good treatment regimens and actions at the societal level to take care of the affected, led to complete drop of TB in the developed world by the 1980's. However, after the 1980's, with the rapid rise of the HIV epidemic, cases of TB re-emerged in the developed world. The phenomenon of global migration in the 1990's, triggered by conflicts due to the collapse of the Soviet Union and the Former Yugoslavia caused a continuous number of cases, particularly in the migrant populations (Abarca Tomas et al. 2013, Odone et al. 2015, Pareek et al. 2016, Van der Werf MJ and Jellweger JP 2016, D'Ambrosio et al. 2017). The past 20 years too, has seen a rise in immigration to the developed world from conflict ridden countries of Asia (Syria, Iraq, Afghanistan) (Krairiksh and Bruce 2011).

Traditionally, migrants carry a huge burden of the disease; in Asia and Southwest Asia, even amongst internally displaced persons (Hugo 2008). In many countries across the world, prisons too have high incidences of TB due to historical neglect by public health authorities. A study in Spain showed high incidence of TB amongst immigrants which constitute 70% of TB cases in poor neighbourhoods in Barcelona. An active program to tackle the same in Spain has shown positive results (Millet et al. 2014, Ospina et al. 2016).

## **STATE OF TB VACCINES AND DIAGNOSTICS**

The reason that studies on TB are important today is that; even after hundred years since the first vaccine; BCG, was discovered, this remains the only main vaccine with variable efficacy in children; and diminishing efficacy in adults. To address current challenges in vaccination, booster antigens to BCG are being developed. Rapid rise of multi-drug resistance, the changing face of urbanization and migration are leading to fresh new societal and economic challenges (Fletcher and Schragger 2016).

**TB diagnosis in India:** There are twenty-six standards proposed by a group of experts that included diagnostic methods of different forms of TB including pulmonary, paediatric, extra-pulmonary, TB with HIV co-infection. Genexpert is the most popular TB test in India (<https://www.tbfacts.org/genexpert/>). In the diagnostic methodology, IFN- $\gamma$  and serum tests are discouraged; lot of treatment by the private sector is unregulated and needs control and regulation, as many diagnostic kits do not meet WHO guidelines (Sreenevas et al. 2014).

**Latent TB:** The detection of latent TB biomarkers is difficult. The paradigm of latent TB has changed, with focus on a spectrum of scenarios and the possibility that some bacilli are still replicating even in dormant phase (Gideon and Flynn 2011).

**Role of Acr:** While in a latent form, TB is able to survive extreme hypoxia and stress conditions due to the expression of a series of genes under the control of DosR regulon. Of these genes, Hsp 16 (heat shock protein) expression is the highest. Hsp 16 is also known as Acr, as it belongs to the  $\alpha$ -crystallin family. The Acr protein is known to be required for log phase growth of *M. tb* and during the transition to stationary phase, at which time, it is suspected to contribute via chaperone activity to the stabilization of intracellular structures. Mycobacterial heat shock protein, Hsp 16, plays an important role in prevention of misfolding of proteins in hypoxic stress conditions (Yuan et al. 1996, 1998). It accumulates upto 25% of the total protein in bacteria and can be detected when the disease is in the latent form, where a program of genes is activated under the control of the DosR (Sherman et al. 2001). The mechanism of action of Acr in preventing thermal aggregation of substrates like citrate synthase and catalase and prevention of Dithiothreitol (DTT) induced B chain aggregation of insulin and Bovine Serum Albumin (BSA) (Chang et al. 1996, Yang et al. 1999, Gu et al. 2002, Borzova et al. 2013, Panda et al. 2017) has been studied. These studies provided insights into the mechanism of action of Acr. It is broadly agreed that the protein forms a trimer of trimers or a dodecamer and prevents misfolding of proteins (Chang et al.

1996, Fu et al. 2005, Kennaway et al. 2005 Panda et al. 2017). The exact mechanism of action of Acr in terms of oligomeric association and dissociation has given insights into how it prevents protein aggregation (Feng et al. 2002, Abulimiti and Chang 2003). These reports reveal dissociation and association of seven to nine oligomeric units for the recombinant Acr protein using urea as a denaturant. It is known to form a nonamer and dodecamer that associates and dissociates to function as a chaperone; preventing misfolding of denatured proteins in the hypoxic condition. It is also a heat shock protein known to have thermostable  $\beta$  sheets.

**Vaccine and diagnostic potential of Acr:** Acr has been studied as a potential vaccine candidate by many groups (Siddiqui et al. 2015, Wieczorek et al. 2014, Moreno-Mendieta et al. 2014). In a review (Esmail et al. 2014) have analyzed the burden of latent TB and the need to improve the current diagnostic tests; tuberculin skin test (TST) and IFN- $\gamma$  release assays. The current diagnostic tests in the pipeline include Early Secretory Antigen (ESAT-6), Culture filtrate protein-10 (CFP-10) which is being developed as a substitute for the existing TST and IFN- $\gamma$  assays (Esmail et al. 2014). Acr has known potential to be developed as a diagnostic in addition to the existing IFN- $\gamma$  assay and an alternative to the TST (Rizvi et al. 2015).

**Other important proteins:** Other important proteins known to be upregulated during infection by *M. tb* in the seven-day THP-1 model (Volpe et al. 2006) include Sod A, Sod C, Bacterioferritin, among others. The Iron (Fe) factored superoxide dismutase Sod A and the Copper-Zinc (Cu-Zn) factored superoxide dismutase Sod C have been cloned and expressed. (Zhang et al. 1991, Liao et al. 2013).

## GAPS IN RESEARCH

*M. tb* Acr, though well researched over many years, is still primarily studied through *in vitro* preparations and the substrates used may not reflect the mechanism of action and function of Acr within the host. There are reports about the dynamic association and dissociation of oligomeric forms of Acr (Feng et al. 2002, Abulimiti and Chang 2003). Most reports conclude that Acr acts a trimer of trimers or dodecamer (Chang et al. 1996, Kennaway et al. 2005, Panda et al. 2017). Recent *in silico* studies on Acr assume it to be a dodecamer (Soong et al. 2018).

However, it is difficult to predict *in vitro* batch to batch variations of recombinant Acr preparation and the biological activity of *in vitro* recombinant Acr preparations has not been correlated to secondary structure and molecular level interaction. Also, no concrete study has been conducted about the effect of pre-heat treatment of Acr on secondary structure of different types of samples before and after gel filtration. There are also no reports of measurement of melting temperature of *M. tb* Acr even though its activity is known to enhance, once it is pre-heated above 37°C and when it crosses 60°C, where a hydrophobic barrier is overcome (Mao et al. 2001). There is a requirement to analyze interactions of protein and substrate and to propose a generalized functional model for activity of *in vitro* Acr which can give insights into the further studies on the actual protein substrates that Acr acts upon. Many oligomer proteins were reported to function as high-molecular-weight aggregates and polymers and the mechanism of heat shock proteins is therefore poorly understood (Rayees 2014). By polynomial graph plotting and its analysis using the real data, we aimed to obtain fresh insights in this regard.

Hence, there is a need to study important proteins such as Sod A and Sod C associated with immune evasion and Acr associated with latent TB and characterize them in terms of

function. Considering the background information and gaps revealed from the literature survey, the following objectives have been mooted.



## **AIMS AND OBJECTIVES**

1. Cloning and expression of *M. tb sod A* and *M. tb sod C* genes.
2. Cloning, expression and purification of *M. tb* Acr.
3. Characterizing structure and function activity relationships of the recombinant *M. tb* Acr.

# CHAPTER 1

## **CHAPTER 1**

### **Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C***

#### **Objective 1**

### INTRODUCTION

Our aim was to clone and express three proteins: Acr, Sod A and Sod C, to further study about their role in *M. tb* infection in the human host. Several reports exist of cloning and expression of Acr (Valdez et al. 2002, Prenata et al. 2004, Kennaway et al. 2005, Dhepakson et al. 2008). Besides Acr, a number of proteins are known to be important for survival of *M. tb* in the human host, including the Iron (Fe) factored superoxide dismutase Sod A and the Copper-Zinc (Cu-Zn) factored superoxide dismutase Sod C (Zhang et al. 1991, Liao et al. 2013). Existing reports of Acr do not delve into a molecular level interaction of the protein with the substrates. Determination of activity units also varies with the kind of substrate used and the concentration of protein, with no clear parameters existing, to compare activity of different studies. Sod A and Sod C studies too, need to be carried out to see if they can be developed into vaccine or diagnostic candidates for TB. The purpose of cloning and expressing these genes was to carry out structure-function analysis. Attempts to clone a gene into a vector involves a variety of steps including, preparation of vector (plasmid isolation), insert preparation (target DNA), restriction digestion and finally, ligation and transformation into a bacterial host strain. There are a large number of steps and, sub-optimal results in any one can lead to failure in the end (Matsumura et al. 2015). By documenting all results numerically, a tool is needed to monitor one's work objectively, rather than subjectively and ensure a successful outcome. A report exists of a combinatorial strategy to improve success of ligation-dependent cloning (Zhang 2013).

Traditionally, cloning employs restriction endonucleases and ligases. In case of a single restriction enzyme, phosphatase treatment is required to prevent self-ligation of vector. Self-ligation of vector necessitates screening of a large number of colonies. Phosphatases are isolated from many sources; bovine, shrimp and bacteria and differ in terms of their heat lability and activity (Rani et al. 2012). However, there is always the danger of carry-over of residual phosphatase, affecting the ligation and it is difficult to inactivate with heat treatment. Using two

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

restriction enzymes can prevent self-ligation of vector. There are drawbacks to this approach as enzyme sites on the vector have to be more than twelve bases apart and the buffers chosen have to be compatible in order to ensure complete digestion (<http://lifescience.roche.com/shop/products/restriction-enzymes-technical-tips>). In spite of this, there is a possibility of having a small proportion of uncut vector left that will transform more efficiently than ligated DNA, leading to reduction in number of positive clones. There is no fool proof method to ascertain whether digestion is 100% complete, as the small proportion of uncut vector would never be visible. Some recommended solutions to these are, use of low percentage gels upto 0.6% agarose to detect uncut species, cutting out of the vector band after linearizing and use of controls of single cut enzyme to look for linearization (Oswald 2007). Cutting out the band exposes the DNA to ultra-violet (UV) radiation, reducing the efficiency of ligation reaction (Kennedy 2007).

**Review of classifiers used in the literature:** Classifiers can be used to analyze data sets accurately and build predictive models by using training set and experimental data to build models. A list of the common classifiers used is:

**Simple Tree** which is a supervised predictive model that can learn to predict discrete or continuous outputs by answering a set of simple questions based on the values of the input features it receives (<https://blog.easysol.net/machine-learning-algorithms-1/>).

**Linear Discriminant analysis (LDA)** is a method used in statistics, pattern recognition and machine learning, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification.

LDA is closely related to analysis of variance (ANOVA) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements ([https://en.wikipedia.org/wiki/Linear\\_discriminant\\_analysis](https://en.wikipedia.org/wiki/Linear_discriminant_analysis)).

**Logistic Regression** is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the Logistic Regression is a predictive analysis. It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. It tries to find a linear boundary to separate the data (<https://www.statisticssolutions.com/what-is-logistic-regression>).

**Support Vector Machine (SVM)** is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes (<https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html>).

**k-Nearest Neighbour algorithm (k-NN)** is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k-Nearest Neighbours (k is a positive integer, typically, small). If  $k = 1$ , then the object is simply assigned to the class of that single nearest neighbour ([https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)).

**Random forest algorithm** is a supervised classification / regression algorithm. As the name suggests, this algorithm creates a forest with a number of decision trees (<http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning>).

## MATERIALS AND METHODS

### 1.1 Parameter Analysis

Attempts were made to clone Mycobacterial genes into pRSETa and pET21a vectors. Numerous parameters were recorded as shown below:

**1.1.1 Insert preparation:** Yield of insert  $Y_1 = f(a_1, a_2, a_3)$ . It depends upon the following factors: Polymerase Chain Reaction (PCR) optimization ( $a_1$ ), clamp addition for restriction digestion ( $a_2$ ) and preparation of insert in terms of gene-clean versus gel-purification ( $a_3$ ).

( $a_1$ ):  $a_1$  is the basic step in preparation of the insert. It is a function of various parameters as listed below: primer concentration ( $a_{11}$ ), denaturation time ( $a_{12}$ ), extension time ( $a_{13}$ ) and melting point difference between the primer pairs ( $a_{14}$ ).  $Y_1$  increases if  $20 < a_{11} < 60$  pmols per reaction and decreases if  $a_{11} > 100$  pmols.  $Y_1$  increases if  $a_{12}$  is between 1 to 3 mins. and  $a_{13}$  is set at 1 min. for 1 kb.  $Y_1$  decreases if  $a_{14} < 5^\circ\text{C}$ .

( $a_2$ ):  $a_2$  is directly proportional to the efficiency of restriction digestion and affects the amount of insert available for ligation. This has been summarized in a study by NEB ([https://www.neb.com/media/NEBUS/Files/Chart%20image/cleavage\\_oligonucleotides\\_old.pdf](https://www.neb.com/media/NEBUS/Files/Chart%20image/cleavage_oligonucleotides_old.pdf)).

$Y_1$  changes as described in the above reference with an optimum value determined by the number and type of bases which in turn depends upon the specific structure of the enzyme.  $Y_1$  increases if  $a_2 > 3$  to 6 bases and the time of restriction digestion is 20 hrs.

( $a_3$ ):  $Y_1$  decreases if a traditional gel-purification method of cutting out the band from the gel, followed by gene-clean is used, as compared to a gene-clean method alone using commercial kits.

**1.1.2 Analysis of plasmid yields:** All plasmid analyses were done using the alkaline lysis method (Birnboim and Doily 1979) or the plasmid isolation kit of Macherey Nagel. The volume of culture used was varied from 1.5 to 100 ml. The documentation of plasmid yield was done using MS Excel analysis. The overall yield of plasmids in  $\mu\text{g}$  ( $Y_2$ ) is expressed as a function of

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

$Y_2 = f(b_0, b_1 \dots b_8)$ . The following parameters were used as factors for yield; volume of culture (b), Optical Density (O. D) of cells ( $b_0$ ), volume of lysis buffer ( $b_1$ ), volume of alkaline lysis buffer ( $b_2$ ), volume of potassium acetate buffer ( $b_3$ ), volume of phenol-chloroform ( $b_4$ ), Ethanol precipitation ( $b_5$ ), Isopropanol precipitation ( $b_6$ ), RNA removal ( $b_7$ ), final purification volume after gene-clean or gel-purification in ml ( $b_8$ ). This reflects the role each parameter plays in the yield as described below:

(b):  $Y_2$  increases if b increases and O. D ( $b_0$ ) increases till a value of 1 to 4.

( $b_1$ ): Resuspending the cell pellet in less volume of lysis buffer to get a concentrated suspension maximizes yield.  $Y_2$  increases if  $b_1$  decreases.

( $b_2$ ): In the crucial step of reversible alkaline denaturation,  $b_2$  and time of incubation plays a critical role.  $Y_2$  increases if  $b_2$  decreases and time of incubation is between 5 and 10 mins.

( $b_3$ ):  $b_3$  and the speed, time and temperature of centrifugation affects the removal of proteins and chromosomal debris, which in turn, decreases the ratio of O.  $D_{260/280}$  and reduces efficiency of the restriction digestion step.  $Y_2$  increases when speed  $> 10,000$  g.  $Y_2$  decreases if the time of spin is  $< 15$  mins. and increases when the time of spin is  $> 15$  mins.  $Y_2$  increases if the temperature of spin is between  $4^\circ\text{C}$  and  $8^\circ\text{C}$  and decreases when the temperature of spin is  $> 8^\circ\text{C}$ .

( $b_4$ ):  $b_4$  plays an important role in ensuring the adequate removal of proteins from the plasmid preparation by ensuring a good O.  $D_{260/280}$  ratio.  $Y_2$  increases if  $b_4 >$  sum of  $b_1$ ,  $b_2$  and  $b_3$  and  $Y_2$  decreases if  $b_4 <$  sum of  $b_1$ ,  $b_2$  and  $b_3$ .

( $b_5$ ): The volume of ethanol, temperature of precipitation and speed of centrifugation plays an important role in determining the yield.  $Y_2$  increases if  $b_5$  is 1 to 2 volumes of  $b_4$  and precipitation temperature is between  $-20^\circ\text{C}$  and  $-80^\circ\text{C}$  and decreases when precipitation temperature is between  $4^\circ\text{C}$  and  $8^\circ\text{C}$ .  $Y_2$  increases if speed of centrifugation  $> 10,000$  g.



(b<sub>6</sub>): The volume of isopropanol, temperature of precipitation and speed of centrifugation too affects the yield. Y<sub>2</sub> increases if b<sub>6</sub> = 0.6 times of b<sub>4</sub> and precipitation temperature is between 20°C and 25°C and decreases when the temperature < 20°C. Y<sub>2</sub> increases when speed of centrifugation > 10,000 g.

(b<sub>7</sub>): This plays an important role as the RNase A preparation must be boiled to remove residual nucleases which could reduce Y<sub>2</sub>. This is followed by a post RNase A purification step which further reduces the yield.

(b<sub>8</sub>): The final volume in which the DNA is resuspended is very important. DNA in a more concentrated form is more amenable to restriction digestion. Y<sub>2</sub> increases if b<sub>8</sub> decreases.

Tables were drawn for plasmid yield obtained at different volumes of culture. Individual values were plotted to check the trend of different optimum ratios of lysis buffer to culture volume. A table was drawn in which the ratios of lysis buffer to each plasmid molecule was theoretically calculated, based on the assumption that 1 O. D. gives  $8 \times 10^8$  cells. Further assumptions included the dry weight of 1 O. D. *E. coli* as 0.39 g/l and the amount of DNA as 3.1%. Another assumption was, the copy number of the plasmid pET21a was 20 per cell (Bolivar et al. 1977), assuming it is a pBR322 derivative, which helped quantify the actual number of plasmid molecules. With this calculation, the amount of plasmid DNA was estimated by weight and the Avogadro number of molecules were calculated. The final yield obtained was re-calculated in terms of weight by molecular weight to estimate actual number of plasmid molecules present in the DNA preparation.

**1.1.3 Restriction digestion / phosphatase treatment:** Vector digestion reaction yield is expressed as Y<sub>3</sub> = % of digestion completed; optimum value being 90 to 95%. This is defined by a single band corresponding to linear DNA as observed on gel electrophoresis. Y<sub>3</sub> = f (c<sub>1</sub>, c<sub>2</sub>, c<sub>3</sub>, c<sub>4</sub>). This is a function of time of reaction (c<sub>1</sub>), amount of DNA (c<sub>2</sub>), % of glycerol in restriction

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

enzymes ( $c_3$ ) and units of enzyme ( $c_4$ ). Each parameter affects the efficiency of restriction digestion as outlined below:

( $c_1$ ):  $c_1$  was varied from 1 hr to 12 hrs depending upon the amount of DNA used and the extent of completion of the reaction.  $Y_3$  increases if  $8 < c_1 < 12$  and decreases if  $2 < c_1 < 8$ .

( $c_2$ ):  $c_2$  was varied from 0.4  $\mu\text{g}$  to 3  $\mu\text{g}$ . Very low amount of DNA reduces efficiency of the reaction while higher amount improves efficiency.  $Y_3$  decreases if  $0.4 \mu\text{g} < c_2 < 1 \mu\text{g}$  and increases if  $1 \mu\text{g} < c_2 < 3 \mu\text{g}$ .

( $c_3$ ):  $Y_3$  decreases if  $c_3 > 10\%$  and increases if  $5\% < c_3 < 10\%$ .

( $c_4$ ):  $Y_3$  increases if  $20 < c_4 < 60$  and decreases if  $10 < c_4 < 20$ .

A series of restriction digestion reactions for single and double digestion with BamHI, EcoRI and XhoI were compiled in the form of tables and charts were plotted to observe trends in relation between amount of digestion and quantity of DNA, reaction time and units of enzyme. Another table was plotted in terms of number of plasmid molecules versus time of completion of reaction and units of enzyme.

**1.1.4 Ligation:** Success of ligation  $Y_4 = f(d_1, d_2, d_3, d_4)$ . It depends upon the following factors: amount of DNA used in vector and insert in ng ( $d_1, d_1'$ ), ligase amount in units ( $d_2$ ), amount of uncut species ( $d_3$ ) and molar ratio of insert to vector ( $d_4$ ).

( $d_1, d_1'$ ):  $d_1, d_1'$  was optimized by varying values between 100 to 400 ng.  $Y_4$  increases if  $100 < d_1, d_1' < 200$ .

( $d_2$ ):  $d_2$  for each reaction was optimized by varying values from 1 to 700.  $Y_4$  increases if  $200 < d_2 < 700$ .

( $d_3$ ):  $d_3$  affects the transformation of the ligation mix.  $Y_4$  decreases if  $d_3$  is high.

( $d_4$ ):  $Y_4$  increases if  $2 < d_4 < 10$ .

**1.1.5 Recombination success** =  $f(Y_1, Y_2, Y_3, Y_4)$ . The overall success of recombination was a function of all the intermediate steps.

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

### 1.1.6 Data for cloning

An Excel sheet was plotted with the summary of all cloning attempts. The successful attempts were indicated as output 1 and the unsuccessful attempts as 0. Analyses of key parameters that affect the success of the cloning were highlighted.

### 1.2 Classifier analysis

Fifteen data sets including eighteen parameters were fed into different classifiers in MATLAB using cross validation of 10-fold and the classifiers calculated the percent accuracy, AUC, precision, specificity and recall. The number of false positives and negatives for each classifier were also plotted to check the robustness of the prediction. Based on the results obtained, three classifiers with > 85% accuracy were used to build the predictive model in MATLAB. Similar analysis was carried out in Clone. R, using a different set of classifiers. Important parameters and their effect on accuracy were plotted.

**1.2.1 MATLAB:** Out of the raw data of all the fifteen cloning attempts, nine indicated outcome 0 (indicating no clone) and six indicated outcome 1 (indicating clone obtained). These were entered into MATLAB system and the different classifiers were checked for varying levels of accuracy, AUC, precision, recall and specificity.

**1.2.2 Clone. R:** The data of the fifteen cloning attempts was analyzed using Clone. R with four classifiers; Random forest, k-NN, SVM GLM and CART (Classification and regression tree). Variable importance was measured in terms of accuracy decrease by changing the parameters. Plots of precision, accuracy, AUC, F1 score (a weighted average of precision and recall) were done using lattice and ggplot2.

### 1.3 Protocol of cloning

**Cloning strategy of *acr*, *sod A* and *sod C* genes:** *M. tb* H37Rv genomic DNA was obtained from Professor Mridula Bose, retired Professor and former Head of Microbiology, Patel Chest Institute, New Delhi, India. The approach was directional cloning into an Isopropyl thio galacto

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

pyranoside (IPTG) inducible expression vector pET21a. The forward primer for *acr* had a BamHI site (underlined). The forward primer for *sod A* had both, EcoRI and BamHI site (underlined). The forward primer for *sod C* had an EcoRI site (underlined), while the reverse primer for all the three genes had an XhoI site (underlined) so that six histidines were added as a C-terminal tag. One construct of *sod A* had an EcoRI site (underlined) in the reverse primer. A clamp of three nucleotides was added to ensure that digestion of the PCR product takes place appropriately.

<i>acr</i> : Forward primer	5' CGC <u>GGATCC</u> ATGGCCACCACCCTTC 3'
Reverse primer	5' CCG <u>CTCGAG</u> GTTGGTGGACCGGATCTG 3'
<i>sod A</i> : Forward primer (with GTG)	5' C <u>GAATTC</u> GTGGCCGAATAC 3'
Reverse primer	5' C <u>GAATTC</u> TCAGCCGAATATCA 3'
<i>sod A</i> : Forward primer	CGC <u>GGATCC</u> ATGGCCGAATACACCTTG
Reverse primer	CCG <u>CTCGAG</u> GCCGAATATCAACCCCTT
<i>sod C</i> : Forward primer	CCG <u>GAATTC</u> ATGCCAAAGCCCGCCGAT
Reverse primer	CCG <u>CTCGAG</u> GCCGGAACCAATGACAC

**1.3.1 PCR amplification of *M. tb acr*, *sod A* and *sod C* genes:** The gene of interest was amplified by diluting template DNA to 60 ng/μl and using 2 μl of genomic DNA (120 ng), 60 pmols of primer (synthesized by Integrated DNA Technology), 2X Emerald Amp GT PCR Master Mix (Takara) containing all components, except template, primers and water. The PCR was set up in a final volume of 20 μl in the thermal cycler as follows:

***acr*:** Initial denaturation 94°C; 1.5 mins., followed by 25 cycles of denaturation 94°C; 1 min., annealing 62°C; 1 min., extension 72°C; 1 min. and final extension 72°C; 7 mins.

***sod A* (with GTG in start codon):** Initial denaturation 94°C; 1.5 mins., followed by 25 cycles of denaturation 94°C; 1 min., annealing 59°C; 1 min., extension 72°C; 1 min. and final extension 72°C; 7 mins.

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

***sod A* (with ATG in start codon):** Initial denaturation 94°C; 1.5 mins., followed by 25 cycles of denaturation 94°C; 1 min., annealing 60°C; 1 min, extension 72°C; 1 min. and final extension 72°C; 7 mins.

***sod C:*** Initial denaturation 94°C; 1.5 mins., followed by 25 cycles of denaturation 94°C; 1 min., annealing 61°C; 1 min., extension 72°C; 1 min. and final extension 72°C; 7 mins.

**1.3.2 Preparation of insert for ligation:** The PCR product was gel-purified by running it on a 2.5% low-melt agarose gel (Affymetrix) at low voltage of 40V at 4°C and then gene-cleaned using the Nucleo Spin gene-clean kit (740609.50, Macherey Nagel). The sample was subjected to double digestion with 40 units of BamHI, EcoRI and XhoI (New England Biolabs) using NEB buffer 3.1 as follows: 1 µg of gene-cleaned insert DNA was digested with XhoI for 18 hrs. followed by EcoRI / BamHI, for 4 hrs. at 37°C.

**Optimization of clamp for insert digestion:** This was based on the optimized clamps needed for restriction digestion at the ends of the DNA. The amount of digestion in the original experiment was quantified using visual estimates of autoradiographs. Based on this, with the 3-nucleotide clamp on the XhoI site, 75% digestion was reported in 20 hrs. With the 3-nucleotide clamp on the BamHI and EcoRI sites, >90% digestion was reported in 3 hrs. ([https://www.neb.com/~/media/NebUs/Files/Chart%20image/cleavage\\_oligonucleotides\\_old.pdf](https://www.neb.com/~/media/NebUs/Files/Chart%20image/cleavage_oligonucleotides_old.pdf)). The double digests were gene-cleaned using Macherey Nagel gene-clean kit. The amount of insert to be used for ligation was quantified and 25% more was used, assuming, XhoI digestion was only 75% complete.

**1.3.3 Preparation of pET21a vector for ligation:** The pET21a vector was isolated by alkaline lysis method of Birnboim and Doily using both mini-prep and midi-prep as required and the Nucleo Spin Plasmid kit from Macherey Nagel. RNase A treatment was done at a concentration of 10 µg/ml and then purified by Macherey Nagel gene-clean kit.

**1.3.4 Restriction digestion / Phosphatase treatment:** The pET21a vector was double-digested with BamHI / EcoRI and XhoI, treated with Calf intestinal alkaline phosphatase (CIAP) and then subjected to ligation to reduce background transformants. Restriction digestion was performed using varying amounts of pET21a vector (1 to 3 µg) in NEB buffer 3.1 and 20 to 60 units of enzyme. The reaction was carried out for 4-8 hrs at 37°C and digestion checked on 0.6% agarose gel. Phosphatase reaction was carried out by using CIAP 0.5 units (Takara) for 15 mins. at 37°C, followed by addition of another 0.5 units and incubating for 15 mins. The calculation of phosphatase was done assuming 3 µg of pET DNA has 2 pmol ends for vectors, digested with one enzyme and 4 pmol ends for vectors, digested with two enzymes. Recommended use of alkaline phosphatase as per pET manual is 0.05 units per pmol ends. After digestion and dephosphorylation, the enzyme was heat-inactivated at 75°C and subjected to phenol-chloroform extraction. It was then precipitated using 0.15 M NaCl and dissolved in Tris-EDTA (TE) buffer (as per manufacturer's instructions). Earlier methods of cloning have used phosphatase to reduce the background of recombinants to less than 5% of transformed colonies (Hoffman and Jendrisak 1990).

**1.3.5 Ligation:** A ligation was set up using two molar ratios of insert over vector, 2:1 and 6:1 and two hundred to seven hundred units of T4 DNA ligase (Takara) using different amounts of vector and insert in a final reaction volume of 10 µl. These are summarized as below:

**Ligation of *acr*:** vector (90 ng); insert (90 ng); reaction volume: 10 µl (successful molar ratio 6:1).

**Ligation of *sod A*:** vector (180 ng); insert (60 and 150 ng); reaction volume: 10 µl (successful molar ratio 2:1 and 6:1).

**Ligation of *sod C*:** vector (75 and 100 ng); insert (35 and 75 ng); reaction volume: 10 µl (successful molar ratio 2:1 and 6:1).

Controls of single cut vector to check ligase efficiency, vector alone to check background and supercoiled DNA to check efficiency of competent cells were used. The calculations were based on the work of Dugaiczky A. et al. (1975) cited in a manual (Bethesda Research Laboratories 1979) and also the concentration of ends available for ligation (Stephenson 2010). To optimize chances of ligation, calculations of vector and insert were done, based on the premise that, for non-identical cohesive ends  $i = N_0 M \times 10^{-3}$  (ends/ml) and  $j = j_{\lambda} (MW_{\lambda}) / (MW)^{3/2}$  and assuming that  $i$  should be three times  $j_{\text{vector}}$ . The reaction parameters are:

$i$  = total concentration of DNA termini

$j$  = effective concentration of one end of a DNA molecule in the immediate neighbourhood of the other end of the same molecule

$N_0$  = Avogadro's number and

$M$  = molar concentration of the DNA molecules

$MW$  = molecular weight

$\lambda$  = phage DNA

The ligation mix was transformed into *E. coli* DH5 $\alpha$ , made competent by Innoue's method (Sambrook and Russell 2001) and then plated on Luria-Bertani (LB) agar with 100  $\mu\text{g/ml}$  ampicillin plates to screen for recombinants.

**1.3.6 Colony PCR:** Recombinants obtained were verified by colony PCR. Single colonies were boiled in 20  $\mu\text{l}$  of autoclaved, deionized water and 8  $\mu\text{l}$  used as a template in a 28  $\mu\text{l}$  reaction using 2X Emerald PCR mix (Takara) and forward and reverse primers of *acr*, *sod A* and *sod C*. The same conditions as for amplification of insert were used and the samples run on 1.5% agarose gel. A negative control of water and positive control of H37Rv genomic DNA was used.

**1.3.7 PCR of plasmids:** The plasmids of the recombinants were isolated and PCR was carried out using the same set of primers as in **1.3.6**.

## Chapter 1: Cloning of Mycobacterial genes *acr*, *sod A* and *sod C*

**1.3.8 Restriction analysis of clones:** The recombinant plasmids that showed amplification by PCR were checked for release of insert by restriction digestion.

**1.3.9 Identification of clones:** The recombinants of *sod A*, *sod A* (with ATG), *acr* (C-terminal His-tag) and *sod C* were confirmed by Colony PCR and restriction digestion using BamHI and XhoI. The nucleotide sequence of the clone expressing the *sod A*, *sod A* (with ATG), *sod C* (C-terminal His-tag) and *Acr* (C-terminal His-tag) was determined using Big Dye Terminator chemistry and ABI 3500xL Genetic Analyzer (Chromous Biotech, Bengaluru) with T7 forward and reverse primers.

### 1.4 Cloning of *acr* into pET28a

The approach was “Directional Cloning” of the PCR amplified *acr* gene into an IPTG inducible expression vector pET28a (Novagen). The primers were designed using the *M. tb* sequence in the TB database entry Rv2031c-hsp-X ([www.tbdb.org](http://www.tbdb.org)), with NdeI in the forward primer and XhoI in the reverse primer (underlined) with stop codon, so that the 6-histidine tag would be added to the amino terminus. The primers designed were for the full-length gene:

Forward primer *acr* pET28 NdeI – 5' GGAATTCCATATGGCCACCACCCTTCCC 3'

Reverse primer *acr* pET28 XhoI – 5' CCGCTCGAGTCAGTTGGTGGACCGGATTCT 3'

**1.4.1 PCR amplification of *M. tb* *acr* gene and preparation of *acr* for ligation:** The gene of interest was amplified by diluting template DNA to 60 ng/μl. A mixture of 120 ng of genomic DNA, 60 pmols of primer (Integrated DNA Technology) and 2X EmeraldAmpGT PCR master mix (Takara) was made up to a final volume of 20 μl. The PCR was set up in the thermal cycler as follows: initial denaturation (94°C, 1.5 mins.); 25 cycles of denaturation (94°C, 1 min.); annealing (61°C, 45 sec.); extension (72°C, 45 sec.); final extension (72°C, 7 mins.).

The PCR product was gel-purified from a low-melt agarose gel (Affymetrix) by using a gene-clean kit (Macherey Nagel). 1 μg of *acr* DNA was double-digested with 50 units of NdeI



and XhoI (Takara) using 10X buffer H for 20 hrs at 37°C. The digested DNA was then gene-cleaned and used for ligation.

**1.4.2 Preparation of pET28a vector for ligation:** The pET28a vector was isolated by alkaline lysis method and purified using the gene-clean kit (Macherey Nagel). The purified vector was double-digested with NdeI and XhoI and then treated with CIAP. The dephosphorylation reaction was carried out by adding one unit of CIAP (Takara) for 30 mins. at 37°C in water bath. 50 µl of the digested, dephosphorylated vector was precipitated using 0.15 M NaCl and 2.5 volumes of ethanol. It was then dissolved in 10 mM Tris pH 8.0 / 0.1 mM EDTA buffer and quantified prior to ligation.

**1.4.3 Ligation:** Ligation was optimized based on the following parameters; molar excess of insert over vector and concentration of ends available for ligation. Ligation was set up on a 50-100 ng scale using two molar ratios of insert over vector (2:1 and 6:1) and 350 units of T4 DNA ligase (Takara) at 16°C. The final reaction volume was maintained in 10 µl. The ligation mix was transformed into *E. coli* DH5α cells, made competent by Innoue's method (Sambrook and Russell 2001) and then plated on LB agar with 50 µg/ml kanamycin plates to screen for recombinants.

**1.4.4 Identification of clones:** The recombinants were confirmed by colony-PCR of recombinant plasmids and restriction digestion using NdeI and XhoI to confirm the release of insert.

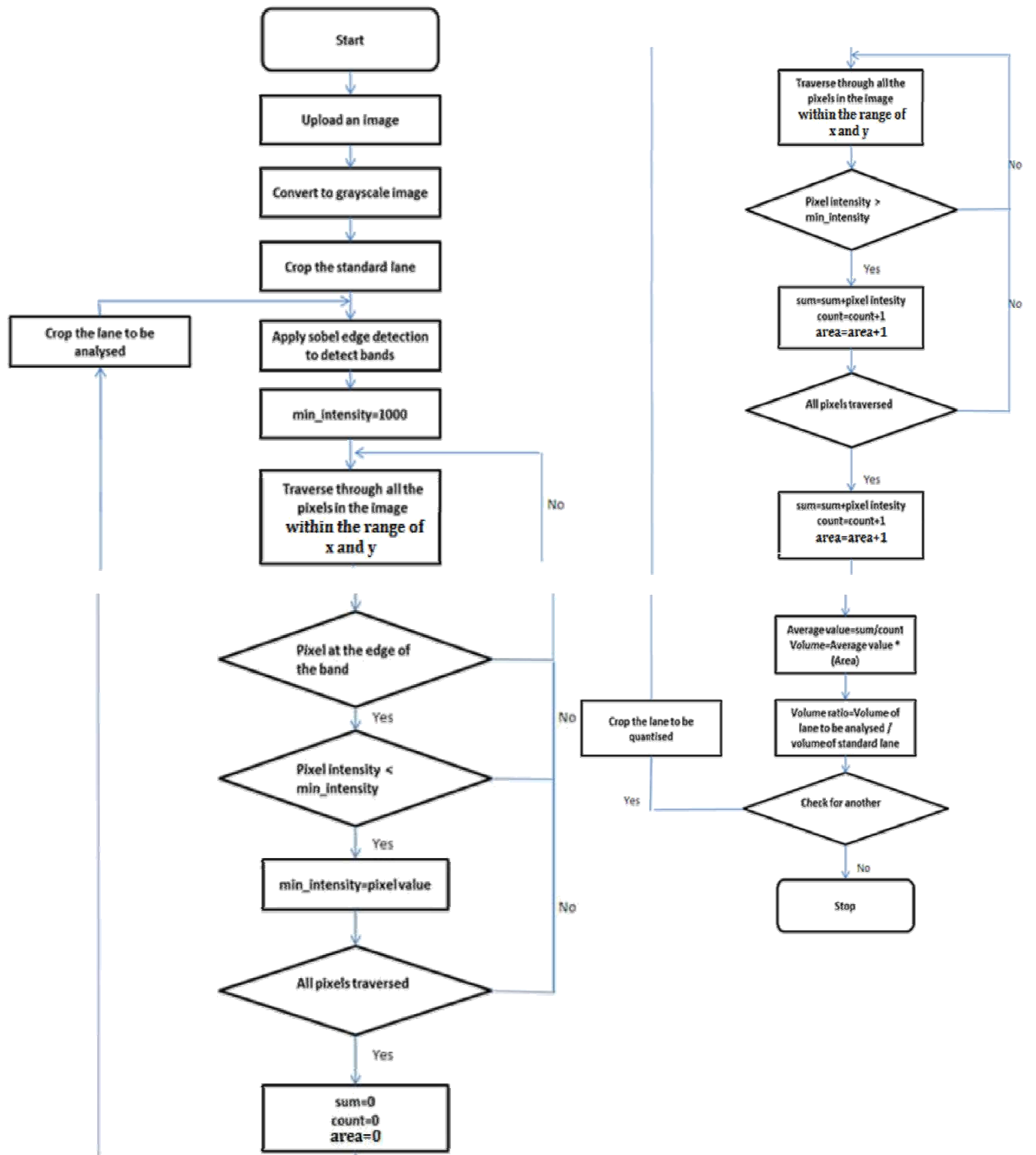
### **1.5 Development of software to quantify DNA**

The DNA quantified before ligation was verified by use of Image processing software in MATLAB by running the DNA on an agarose gel along with known standard DNA, usually, 100 bp ladder. This helped to accurately estimate the amount of DNA before setting up ligation in addition to the nanodrop method and estimation by agarose gel-electrophoresis.

### 1.5.1 Gel quantification – MATLAB code algorithm

MATLAB code involves following steps:

- Upload the image to be analyzed
- Subtract the background from the image
- Convert the image to **grayscale** image
- Crop the standard lane
- Apply **sobel** edge detection to detect bands
- Convert the grayscale image to **binary** image and store it in different variable
- Boundary pixel positions of the binary image above, store the minimum (x\_min, y\_min) and maximum (x\_max, y\_max) values of row and column; get minimum intensity value at the edges of bands
- Add all the intensities of the pixels in the band, whose intensities are greater than the minimum intensity value within the range of x and y; calculate the total area of bands in the lane
- Calculate the average value of standard lane; calculate the volume of standard lane; follow the same steps for the lanes to be analyzed
- Finally, calculate the volume ratio
- Volume ratio = Volume of lane to be analyzed ÷ Volume of standard lane
- The algorithm is presented in the flow chart below



## RESULTS

### 1.1 Analysis of parameters

**1.1.1 Use of Excel to quantify plasmid DNA yields:** The yield ranged from 1.25  $\mu\text{g}$  to 20  $\mu\text{g}$  (Table 1.1.1.1 to 1.1.1.4). The increase in yield was not in linear proportion to the volume of culture used. This could be due to the larger amount of lysis buffer needed; which contributed to this observation. It is recommended to use lysozyme for culture volumes greater than 10 ml to improve cell lysis. Any variation in this step could affect plasmid yields. Some of the other factors include, the final O. D<sub>600</sub> of the culture, as yield is reduced if O. D. enters the late stationary phase (data not recorded). It is strongly recommended to use final O. D<sub>600</sub> of 1 to 4 to get optimum yield.

**Table 1.1.1.1:** Plasmid yield calculation for 6 batches of 1.5 ml culture of pET21a vector by alkaline lysis – Birnboim and Doily method: Average yield was 1.12  $\mu\text{g}$ .

Culture volume b (1.5 ml)	Ratio of lysis buffer 1 to volume ( $b_1/b$ )	Ratio of lysis buffer 2 to volume ( $b_2/b$ )	Ratio of buffer 3 to volume ( $b_3/b$ )	Yield ( $\mu\text{g}$ )
1	0.067	0.13	0.1	0.6
2	0.067	0.13	0.1	0.1
3	0.067	0.13	0.1	1.2
4	0.067	0.13	0.1	1.2
5	0.067	0.13	0.1	1.6
6	0.067	0.13	0.1	2

**Table 1.1.1.2:** Plasmid yield calculation for 6 batches of 5 ml culture of pET21a vector by alkaline lysis – Birnboim and Doily method: Average yield was 4.25  $\mu\text{g}$ .

Culture volume b (5 ml)	Ratio of lysis buffer 1 to volume ( $b_1/b$ )	Ratio of lysis buffer 2 to volume ( $b_2/b$ )	Ratio of buffer 3 to volume ( $b_3/b$ )	Yield ( $\mu\text{g}$ )
1	0.02	0.04	0.03	2.5
2	0.02	0.04	0.03	6
3	0.02	0.04	0.03	6
4	0.02	0.04	0.03	2.5
5	0.02	0.04	0.03	2.5
6	0.02	0.04	0.03	6

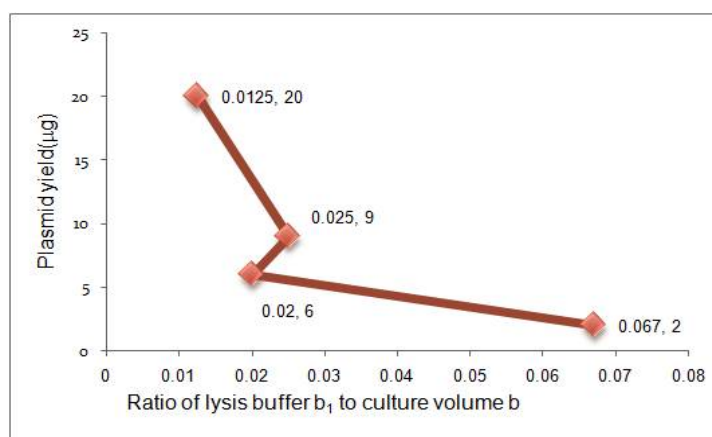
**Table 1.1.1.3:** Plasmid yield calculation for 4 batches of 10 ml culture of pET21a vector by Macherey Nagel Kit method: Average yield was 9  $\mu\text{g}$ .

Culture volume b (10 ml)	Ratio of lysis buffer 1 to volume ( $b_1 / b$ )	Ratio of lysis buffer 2 to volume ( $b_2 / b$ )	Ratio of buffer 3 to volume ( $b_3 / b$ )	Yield ( $\mu\text{g}$ )
1	0.025	0.025	0.03	9
2	0.025	0.025	0.03	9
3	0.025	0.025	0.03	9
4	0.025	0.025	0.03	9

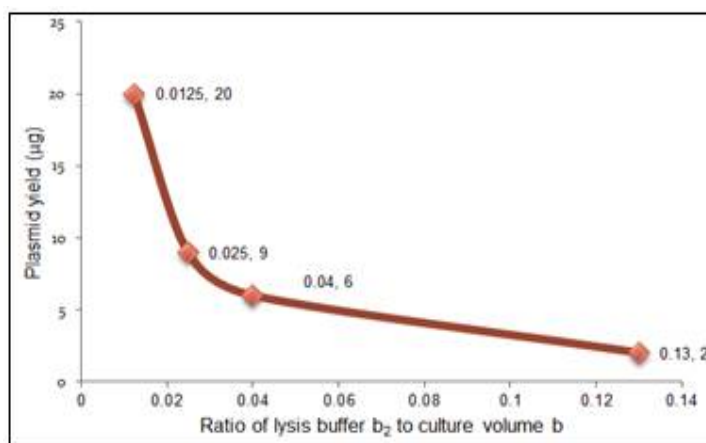
**Table 1.1.1.4:** Plasmid yield calculation for 8 batches of 25 ml culture of pET21a vector - alkaline lysis and Macherey Nagel Kit method: Average yield was 8.39  $\mu\text{g}$ .

Culture volume b (25 ml)	Ratio of lysis buffer 1 to volume ( $b_1 / b$ )	Ratio of lysis buffer 2 to volume ( $b_2 / b$ )	Ratio of buffer 3 to volume ( $b_3 / b$ )	Yield ( $\mu\text{g}$ )
1	0.016	0.032	0.024	4.4
2	0.016	0.032	0.024	4.4
3	0.016	0.032	0.024	3.6
4	0.010	0.010	0.012	16.0
5	0.033	0.067	0.050	4.5
6	0.080	0.160	0.120	10.0
7	0.040	0.040	0.048	8.76
8	0.0125	0.0125	0.015	20.0

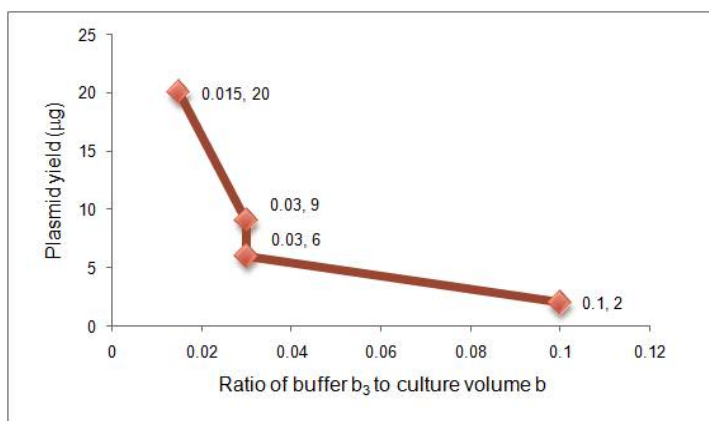
**The ratios of lysis buffer to culture volume:** The volumes of lysis buffer  $b_1$ ,  $b_2$  and  $b_3$  played a significant role in improving yields. The lower the ratio of lysis buffer to the culture volume, the better was the yield (**Figure 1.1.1.1 to 1.1.1.3; Table 1.1.1.5**).



**Figure 1.1.1.1:** Plot of plasmid yield versus ratio of lysis buffer  $b_1$  to culture volume b.



**Figure 1.1.1.2:** Plot of plasmid yield versus ratio of lysis buffer  $b_2$  to culture volume  $b$ .



**Figure 1.1.1.3:** Plot of plasmid yield versus ratio of buffer  $b_3$  to culture volume  $b$ .

**Table 1.1.1.5:** Plasmid yields for different volumes of culture and ratios of lysis buffer to culture volume.

Culture volume $b$ (ml)	Ratio of lysis buffer 1 to volume ( $b_1 / b$ )	Ratio of lysis buffer 2 to volume ( $b_2 / b$ )	Ratio of buffer 3 to volume ( $b_3 / b$ )	Yield ( $\mu\text{g}$ )
1.5	0.067	0.13	0.1	2
5	0.020	0.04	0.03	6
10	0.025	0.025	0.03	9
25	0.0125	0.0125	0.015	20

Based on all the preceding analysis, recommendations for plasmid analysis are given in **Table**

**1.1.1.6**

**Table 1.1.1.6:** A brief summary of plasmid yield calculation was documented in an Excel sheet with an example of a 1.5 ml plasmid isolation.

Y <sub>2</sub> (Yield of plasmid DNA)	
Y <sub>2</sub> = f {(b <sub>0</sub> , b <sub>1</sub> ) ..... (b <sub>8</sub> )}	
Volume of culture ml (b)	1.50
O. D of cells (b <sub>0</sub> )	1.00
Volume of lysis buffer 1 (b <sub>1</sub> )	0.10
Volume of lysis buffer 2 (b <sub>2</sub> )	0.20
Volume of buffer 3 (b <sub>3</sub> )	0.15
Volume of phenol-chloroform (b <sub>4</sub> )	0.45
Volume of solvent (ETOH)(b <sub>5</sub> ) incubation at -20°C for 16 hrs	0.45
Volume of IPA (b <sub>6</sub> ) incubation at room temperature for 30 mins. to 1 hr	0.27
RNA removal by RNase A (10 µg/ml) (b <sub>7</sub> ) incubation at -37°C for 30 to 45 mins.	0.0027
Final resuspension volume after gene-clean or gel-purification (b <sub>8</sub> ) ml	0.02
Final yield (µg)	0.67

For  $1 < b < 10$ , use ratios (b<sub>1</sub>, b<sub>2</sub> and b<sub>3</sub>) / b of 0.067, 0.133 and 0.10 respectively.

For  $10 < b < 25$ , use ratios (b<sub>1</sub>, b<sub>2</sub> and b<sub>3</sub>) / b of 0.025, 0.025 and 0.03 respectively.

For  $b > 25$ , use ratios (b<sub>1</sub>, b<sub>2</sub> and b<sub>3</sub>) / b of 0.0125, 0.0125 and 0.015 respectively.

**Figures 1.1.1.1, and 1.1.1.2 and 1.1.1.3** illustrated the above observation. However, if the culture volume crosses 10 ml, use of lysozyme is recommended to improve plasmid recovery.

The change in ratio of lysis buffer to culture volume improved the yield. This could be due to the concentrated volume to which cells are exposed. This was verified by calculating the ratio of molecules of lysis buffer per plasmid molecule (**Table 1.1.1.6, 1.1.1.7**). The volumes of lysis buffer are to be readjusted as per the Midi-prep and Maxi-prep protocols (Sambrook and Russell 2001). However, this can be done on a case by case basis depending upon O. D<sub>600</sub>.

Knowledge of the number of plasmid molecules and culture volume can help fix the amount of lysis buffer to be used.

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

**Calculation of number of molecules to be lysed per  $\mu$ l of lysis buffer:** For each culture volume and assuming final O. D of  $1 = 8 \times 10^8$  cells /ml, the total number of cells was calculated (**Table 1.1.1.7**).

**Table 1.1.1.7:** The number of plasmid molecules for pET21a were calculated by multiplying the number of cells by 20; (<https://www.qiagen.com/in/resources/technologies/plasmid-resource-center/growth%20of%20bacterial%20cultures/>), assuming it is a pBR322 derivative (Dry weight of 0.39 g/L and DNA content of 3.1% was assumed for the final calculation).

The following references were used for calculating the same <https://nebiocalculator.neb.com/#!/dsdnaamt> and <http://bionumbers.hms.harvard.edu/bionumber.aspx?id=101937>

Culture volume	No. of plasmid molecules	Volume of lysis buffer $b_1$ (ml)	Amount of molecules to be lysed per $\mu$ l
1.5	$0.98 \times 10^{14}$	0.10	$0.98 \times 10^{12}$
10	$0.65 \times 10^{15}$	0.25	$0.26 \times 10^{12}$
25	$1.635 \times 10^{15}$	0.30	$0.51 \times 10^{12}$
100	$0.65 \times 10^{16}$	4.00	$1.635 \times 10^{12}$

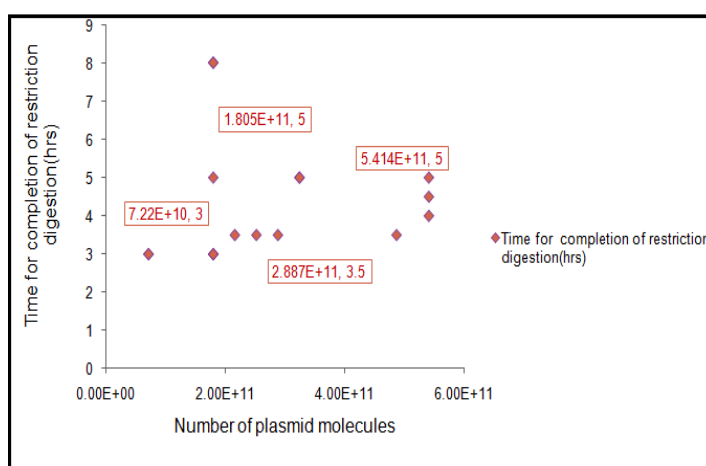
**1.1.2 Use of Excel to analyse restriction digestion data:** MS Excel analysis of seventeen batches of restriction digestion gave an insight into the approximate amount of DNA to be used for each reaction along with units of enzyme and the reaction time (**Table 1.1.2**). The average value worked out to 1.55  $\mu$ g of DNA requiring 4.5 hrs of reaction time and 41 units of restriction enzyme and a reaction volume of 100  $\mu$ l. Our results match with the recommendations given in the Novagen pET manual which used 60 units for double digestion of 3  $\mu$ g of DNA ([lifeserv.bgu.ac.il/wp/zarivach/wp-content/.../Novagen-pET-system-manual1.pdf](http://lifeserv.bgu.ac.il/wp/zarivach/wp-content/.../Novagen-pET-system-manual1.pdf)).



**Table 1.1.2:** Restriction digestion summary of different batches.

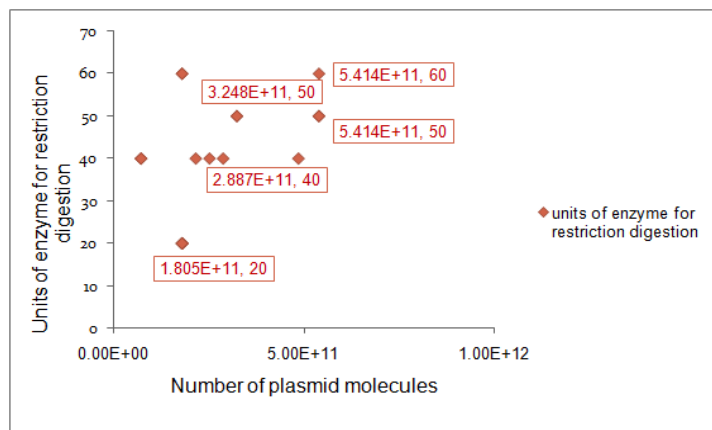
S. No.	Amount of DNA ( $\mu\text{g}$ )	Time of completion (hrs)	Units of enzyme
1	0.4	3	40
2	0.4	3	40
3	1	3	20
4	1	3	20
5	1	3	20
6	1	5	20
7	1	8	60
8	1	8	60
9	1.2	3.5	40
10	1.4	3.5	40
11	1.6	3.5	40
12	1.8	5	50
13	1.8	5	50
14	2.7	3.5	40
15	3	4	50
16	3	4.5	50
17	3	5	60

**Calculation of time required for optimum restriction digestion:** The time for completion of restriction digestion versus the number of plasmid molecules showed an average of 4.5 hrs. Number of plasmid molecules that could be digested was in the range of  $2.88 - 5.4 \times 10^{11}$  ( **Figure 1.1.2.1**).



**Figure 1.1.2.1:** Plot of time for completion of restriction digestion versus no. of molecules of plasmid DNA.

**Calculation of enzyme units required for optimum restriction digestion:** The enzyme units for completion of restriction digestion versus the number of plasmid molecules showed an average of forty units for  $2.88 - 5.4 \times 10^{11}$  (**Figure 1.1.2.2**). The number of supercoiled molecules needs to be quantified to further derive relevant conclusions.



**Figure 1.1.2.2:** Plot of units of enzyme for restriction digestion versus no. of plasmid molecules.

**1.1.3 Ligation:** Ligation attempts were carried out with varying amounts of ligase and these are summarized in **Table 1.1.3**. Optimization studies showed that addition of units of ligase without dilution and using ligase with pre-added ATP showed better results.

**1.1.4 Recombination success:** =  $f(Y_1, Y_2, Y_3, Y_4)$ . The overall success or failure of recombination was documented using an Excel sheet to highlight the different cloning attempts and summarized in **Table 1.1.3**. This documents all the attempts to clone different genes including those without directional cloning. Valuable insights were obtained from this table on the parameters that are key to successful cloning. Summary of all fifteen cloning attempts; both, non-directional and directional and with and without phosphatase treatment. The inputs were eighteen in number with values of volume of incubation of culture in ml and two plasmid isolation methods; the alkaline lysis method and the Macherey Nagel plasmid isolation kit. The DNA obtained was quantified by agarose gel analysis along with standards. The amount of vector and insert DNA was quantified when there was only a single band on agarose gel and all

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

calculations were as described in Materials and Methods. The target obtained was classified as clone obtained (1) or not obtained (0). All the different types of construct made are listed below

(Table 1.1.3).

**Table 1.1.3: Summary of all cloning attempts.**

<b>INPUTS</b>															
Incubation volume (ml)	50	100	25	30	25	25	100	100	50	50	50	50	100	50	50
plasmid isolation method	1	1	2	2	2	2	1	1	2	2	2	2	1	1	1
buffer 1 (ml) b <sub>1</sub>	2	4	0.5	1	0.25	1	2	2	0.75	0.75	0.75	0.75	2	1	1
buffer 2 (ml) b <sub>2</sub>	4	8	0.5	1	0.25	1	4	4	0.75	0.75	0.75	0.75	4	2	2
buffer 3 (ml) b <sub>3</sub>	3	6	0.6	1.2	0.3	1.2	3	3	0.9	0.9	0.9	0.9	3	1.5	1.5
vol. of TE (ml) b <sub>8</sub>	2	1	0.1	0.1	0.1	0.08	0.5	0.5	0.01	0.01	0.01	0.01	0.5	0.5	0.5
yield of plasmid (μg)	100	30	0.4	20	20	8.76	5	5	20	20	20	20	5	4.5	4.5
volume of reaction	0.05	0.1	0.05	0.05	0.05	0.02	0.05	0.05	0.1	0.1	0.1	0.1	0.05	0.1	0.1
amount of DNA (μg) c <sub>2</sub>	1.6	2	0.4	0.4	0.8	1.32	1.6	1.6	4	4	4	4	1.6	1	1
amount of enzyme c <sub>4</sub>	40	100	2	15	6	10	40	40	80	80	80	80	40	120	120
Incubation time (mins.) c <sub>1</sub>	210	300	90	90	60	60	360	360	480	480	480	480	360	360	360
amount of vector DNA (μg) (d <sub>1</sub> )	0.08	0.09	0.056	0.015	0.05	0.2	0.09	0.09	0.18	0.18	0.15	0.09	0.09	0.1	0.075
vector size (kb)	5.4	5.4	2.9	2.9	2.9	2.9	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4
amount of insert (μg) (d <sub>1</sub> )	0.04	0.025	0.03	0.048	0.02	0.08	0.09	0.09	0.06	0.15	0.055	0.055	0.09	0.035	0.075
reaction volume (ml)	0.02	0.06	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.012	0.015
molar ratio of insert to vector (d <sub>1</sub> )	6	2	11	18	2	2	6	6	2	6	2	6	6	2	6
insert size (kb)	0.624	0.624	0.434	0.624	0.624	0.624	0.723	0.723	0.624	0.624	0.434	0.434	0.723	0.723	0.723
amount of ligase units (d <sub>2</sub> )	10	200	1	1	1	30	350	350	350	350	350	350	350	700	700
<b>TARGET</b>															
Clone obtained	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1
Nature of construct		sod A with GTG					sod C	sod C	sod A (ATG)	sod A (ATG)	acr	acr	sod C	sod C	sod C
							directional vector, no CIP		directional vector with CIP						

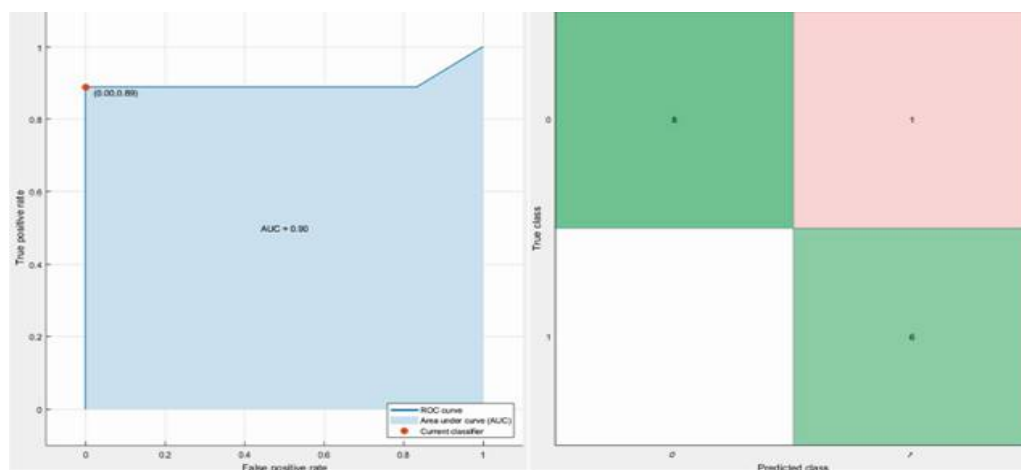
**1.2 Classifier analysis**

**1.2.1 MATLAB:** The Simple Tree classifier showed best results in terms of all the parameters, followed by the Support Vector Machine and Random forest (**Table 1.2.1; Fig 1.2.1.1 to Fig 1.2.1.6**).

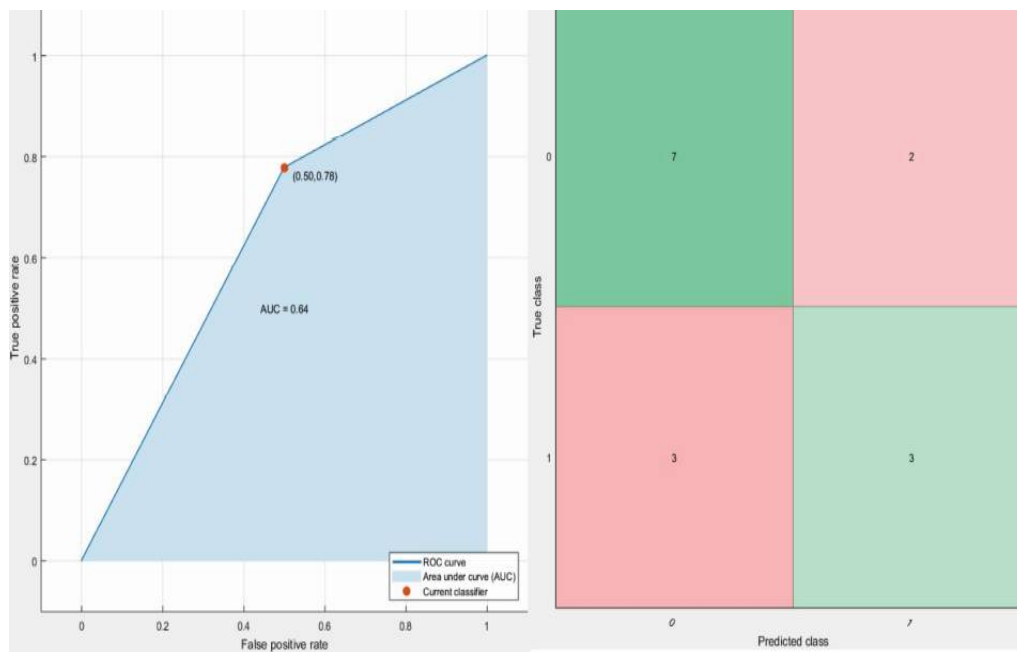
**Table 1.2.1:** Summary of MATLAB Classifier analysis

Classifier	Accuracy	TP	FN	FP	TN	AUC	Recall	Precision	Specificity
Linear discriminant	66.7	7	2	3	3	0.64	0.778	0.7	0.5
Logistic Regression	73.3	7	2	2	4	0.72	0.778	0.78	0.67
k-Nearest Neighbour	80	8	1	2	4	0.78	0.889	0.80	0.67
Support Vector Machine	86.7	8	1	1	5	0.74	0.889	0.889	0.83
Random forest	86.7	8	1	1	5	0.9	0.889	0.89	0.83
Simple Tree	93.3	8	1	0	6	0.9	0.889	1	1

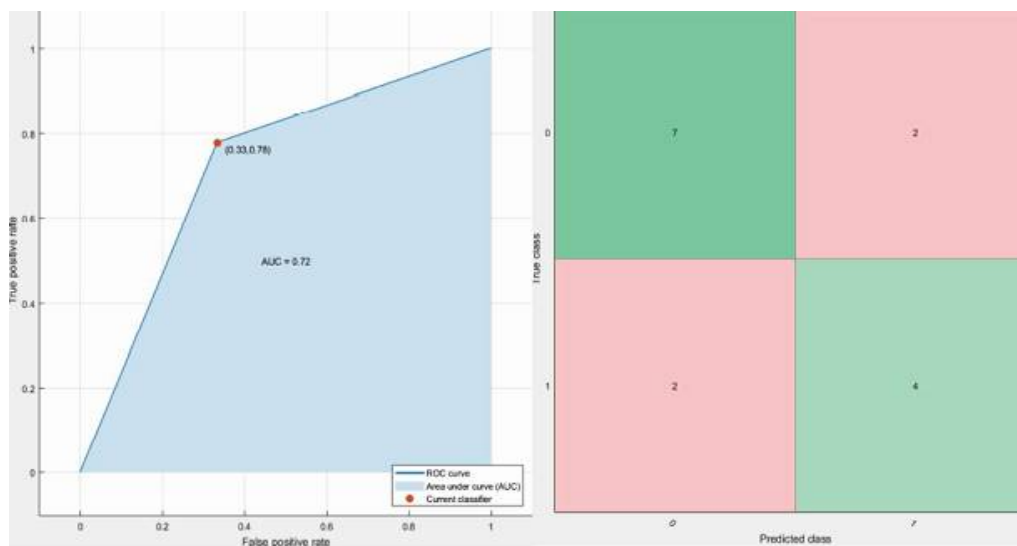
**Classifier analysis:** The number of True Positives and True Negatives was highest in Simple Tree and lowest in Linear Discriminant classifier.



**Figure 1.2.1.1:** Simple Tree classifier (ROC Curve, Confusion matrix).

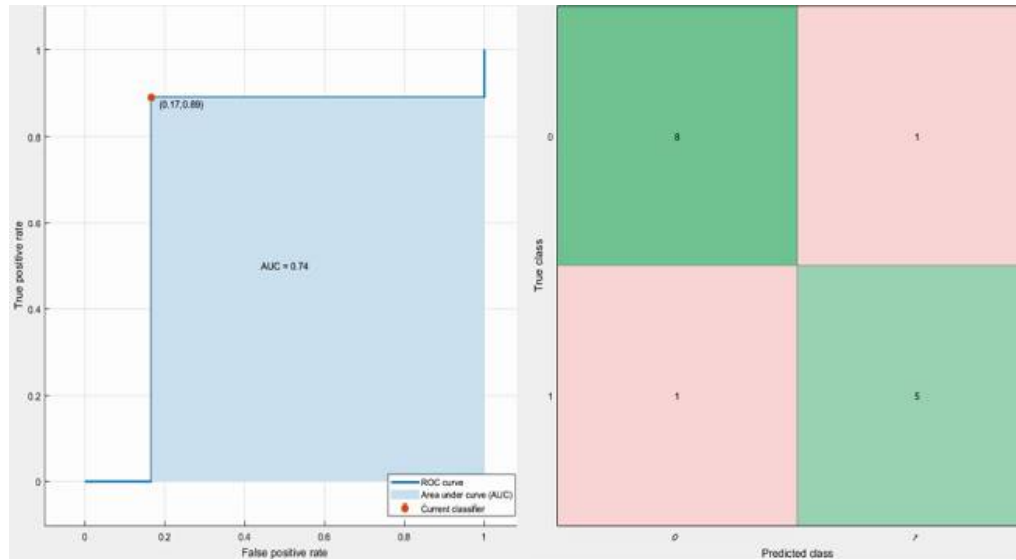


**Figure 1.2.1.2:** Linear Discriminant classifier (ROC Curve, Confusion matrix).

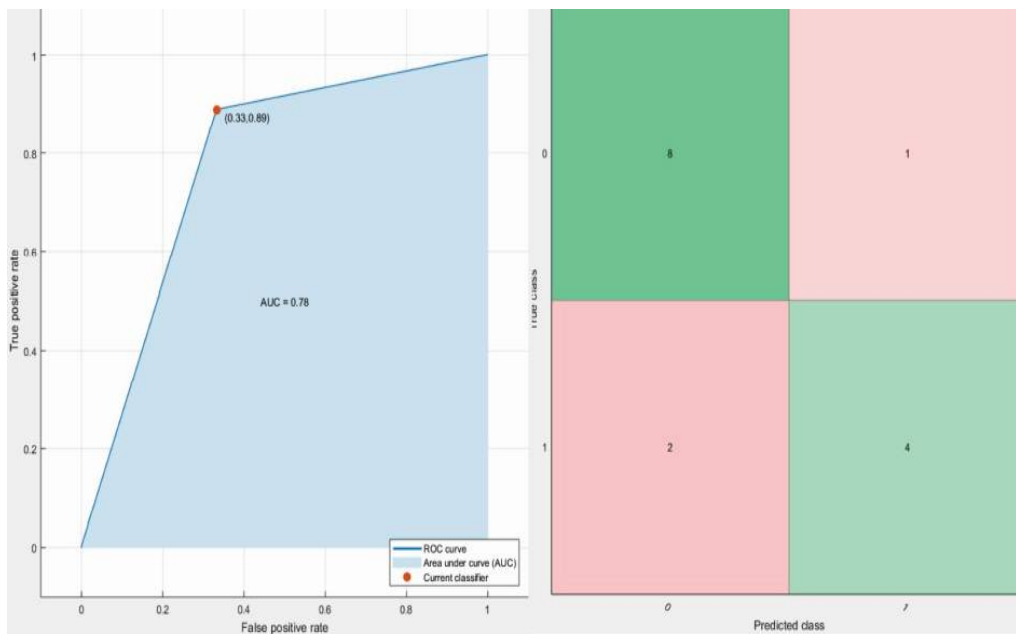


**Figure 1.2.1.3:** Logistic Regression (ROC Curve, Confusion matrix).

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*



**Figure 1.2.1.4:** Support Vector Machines classifier (ROC Curve, Confusion matrix).



**Figure 1.2.1.5:** k-nearest neighbour classifier (ROC Curve, Confusion matrix).

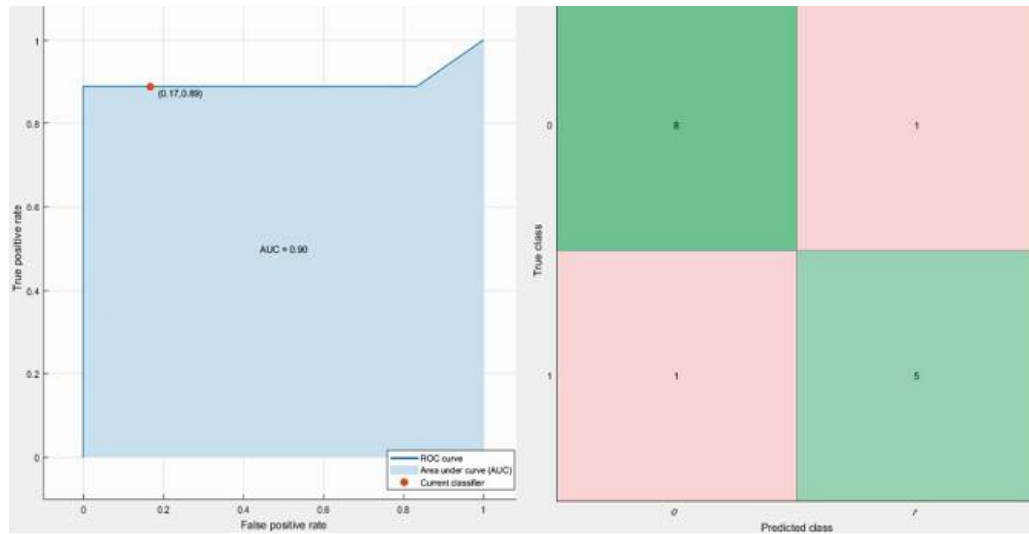


Figure 1.2.1.6: Random forest classifier (ROC Curve, Confusion matrix).

1.2.2 Clone. R: The Random forest classifier showed best results in terms of recall, precision and accuracy (Figure 1.2.2.1).

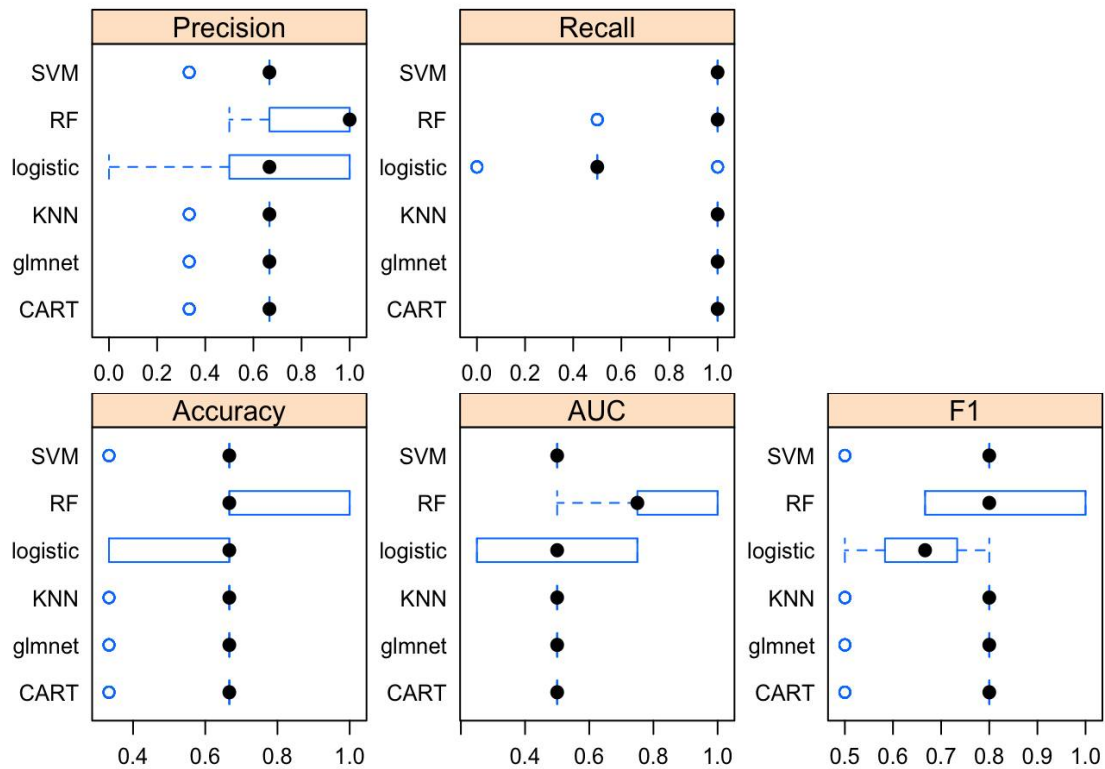
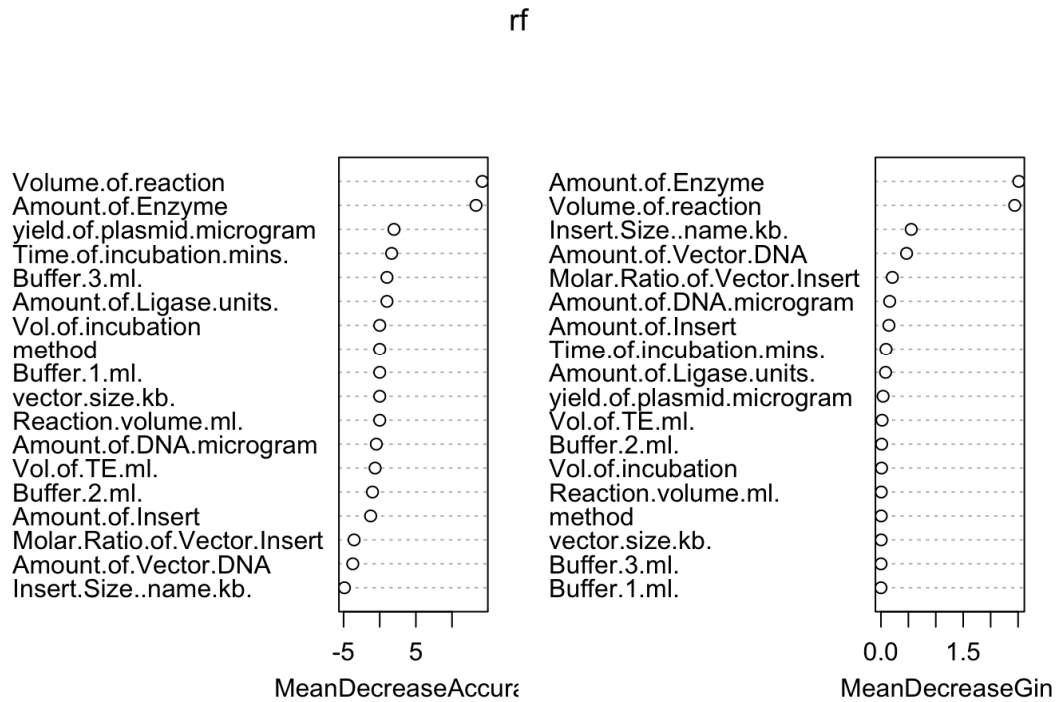
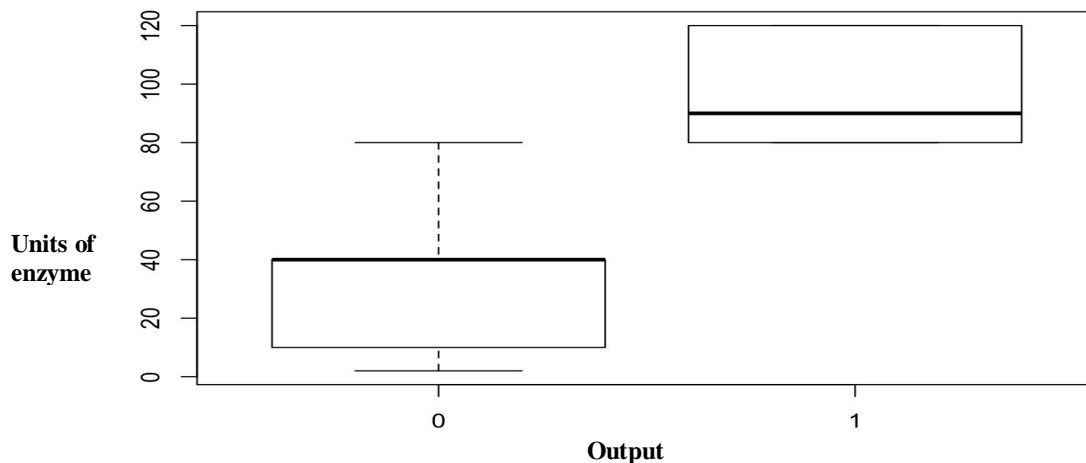


Figure 1.2.2.1 Precision, Recall, Accuracy and AUC using Clone. R

The most important parameters obtained were volume of reaction and units of enzyme (**Figure 1.2.2.2, 1.2.2.3, 1.2.2.4**).

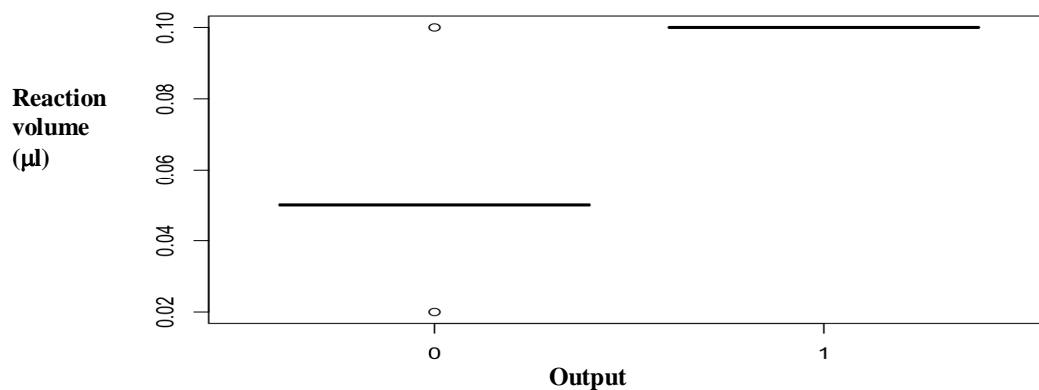


**Figure 1.2.2.2:** Plot of effect of important parameters on Accuracy of model in Clone. R



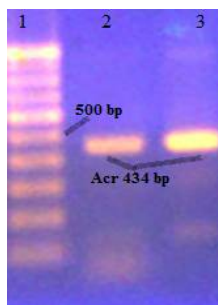
**Figure 1.2.2.3:** Effect of units of enzyme on output in Clone. R





**Figure 1.2.2.4:** Effect of reaction volume on Output in Clone. R

**1.3.** Cloning of *acr*, *sod A* (GTG), *sod C* gene (directional EcoRI / XhoI) *sod A* gene (directional BamHI / XhoI) in pET21a. The *acr* gene (directional BamHI / XhoI) was amplified from *M. tb* H37Rv using the primers designed and showed amplification of expected size of 434 base pairs and checked using 1.5% agarose gel electrophoresis (**Figure 1.3.1.1**).

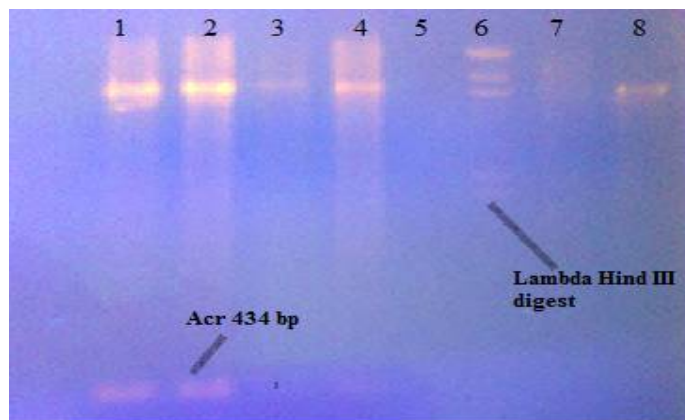


**Figure 1.3.1.1:** 1.5 % agarose gel analysis of PCR of *acr* from genomic DNA / pET21a: Lane 1:100 bp ladder; Lane 2, 3: *acr* amplified 434 bp.

The ligation mix was transformed into high efficiency DH5 $\alpha$  cells made by Innoue's method. Eight colonies were obtained in the insert plus vector plate which had mole ratio of 6:1. The vector control had one colony indicating minimization of background self-ligation and success of the directional cloning with phosphatase approach. Out of eight colonies screened for the presence of the insert, four clones (#1, 5, 7, 8) showed amplification of the *acr* gene analysed by 1% agarose gel electrophoresis (**Figure 1.3.1.2**). The non-specific bands were primer-dimers.



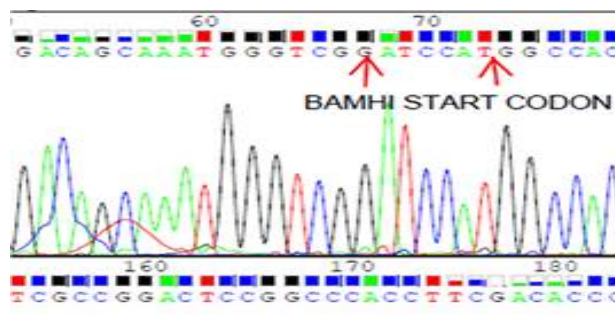
**Figure 1.3.1.2:** 1.5 % agarose gel analysis of Colony PCR of clones: Lane 1 to 8: Colony 1 to 8; Lane 9:100 bp ladder marker; Lane 10: Negative control; Lane 11: Positive Control; genomic DNA. Three of these clones (#1, 5, and 7) were screened for the presence of the insert by plasmid isolation and restriction digestion, using 40 units of BamHI and XhoI in NEB 3.1 buffer. All three clones showed release of insert of expected size of 434 base pairs (**Figure 1.3.1.3**).



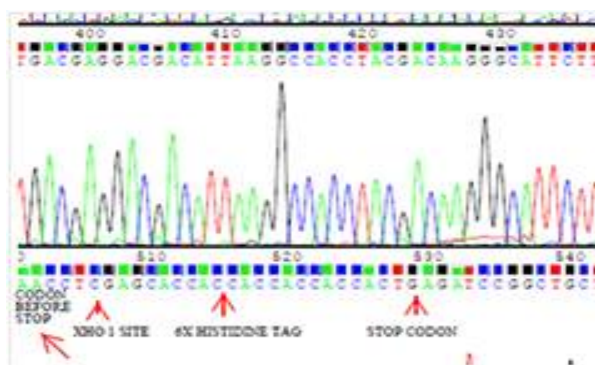
**Figure 1.3.1.3:** 1.2 % agarose gel analysis of restriction digestion of *acr* clones with BamHI and XhoI: Lane 1: Recombinant no. 1, Lane 2: Recombinant no. 5, Lane 3: Recombinant No. 7, Lane 4: Recombinant no. 8, Lane 6: Lambda DNA-HindIII Digest marker, Lane 7: Uncut vector control pET21a, Lane 8: BamHI / XhoI cut pET21a as control.

The *acr* clone was sequenced using both forward and reverse primers so as to enable us to read the entire sequence (**Figure 1.3.1.4, 1.3.1.5**). The sequencing data confirmed the presence of

BamHI and the start codon in frame with the vector. Similarly, the XhoI site and the 6X His-tag (minus the stop codon) of the *acr* gene were found to be in frame. No mutations were detected in the *acr* sequence which was found to be 100% identical to the *acr* sequence in the TB database.



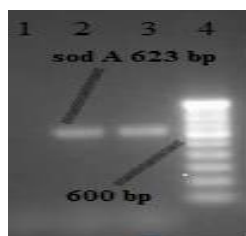
**Figure 1.3.1.4: Nucleotide Sequence of *acr* clone #1 Start codon and BamHI site.**



**Figure 1.3.1.5: Nucleotide Sequence of *acr* clone #1 Histidine tag and Stop codon.**

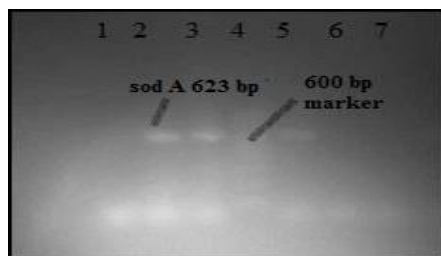
The *sod A* gene was cloned at the EcoRI site with GTG as the start codon. The gene was amplified from *M. tb* H37Rv and this showed amplification of expected size of 624 base pairs. The PCR product was digested with EcoRI and ligated into pET21a vector. This ligation mix was transformed and recombinants checked for clones by amplification of colonies. The confirmed colonies were re-tested for amplification of the desired insert from freshly isolated recombinant plasmids (**Figure 1.3.2.1**).

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*



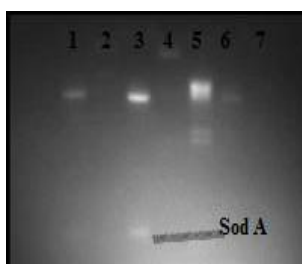
**Figure 1.3.2.1:** 1.5% agarose gel analysis of PCR of *sod A* with GTG in start codon (non-directional): Lane 1: Negative control; Lane 2, 3: *sod A* gene; Lane 4: 100 bp ladder.

The PCR of recombinant plasmids showed amplification in three out of four colonies analysed (#2, 3, 4) (**Figure 1.3.2.2**).



**Figure 1.3.2.2:** 1.5 % agarose gel analysis of colony PCR cloning of recombinants of *sod A* with GTG in start codon: Lane 1-3: Colony PCR clone #1-3; Lane 4: Marker; Lane 5: Colony PCR clone #4; Lane 6: water control; Lane 7: vector control.

Restriction digestion of two of the recombinant plasmids (#2, 4) (**Figure 1.3.2.3**).



**Figure 1.3.2.3:** 1.2 % agarose gel analysis of restriction digestion of recombinant *sod A* with GTG in start codon: Lane 1: EcoRI cut *sod A* clone #2; Lane 2: uncut clone #2; Lane 3: EcoRI cut *sod A* clone #4; Lane 4: uncut clone #4; Lane 5: Marker – Lambda DNA-HindIII Digest; Lane 6: pET21a control cut with EcoRI; Lane 7: EcoRI digested pET21a vector control.

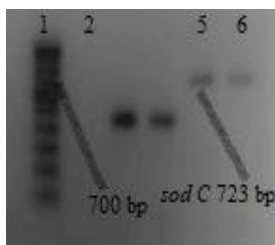
## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

The *sod A* clone was sequenced by using both forward and reverse primers. Three point mutations were found in the 624 bp sequence (**Figure 1.3.2.4**).



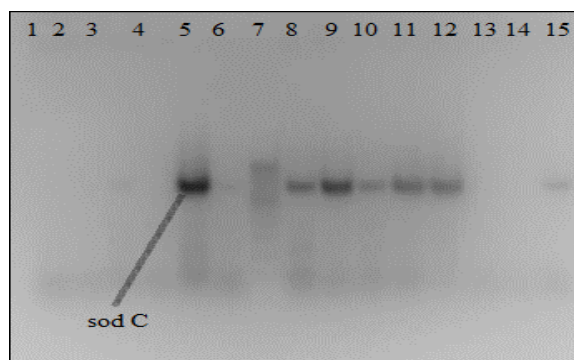
**Figure 1.3.2.4: Sequencing of recombinant *sod A* with GTG in start codon.**

**Cloning of *sod C* gene:** The *sod C* gene was cloned by directional EcoRI / XhoI cloning into pET21a. The gene was amplified from *M. tb* H37Rv and this showed amplification of expected size of 723 base pairs (**Figure 1.3.3.1**).



**Figure 1.3.3.1:** 1.5% agarose gel analysis of PCR fragment of *sod C*: Lane 1: 100 bp ladder; Lane 5, 6: *sod C* amplified gene.

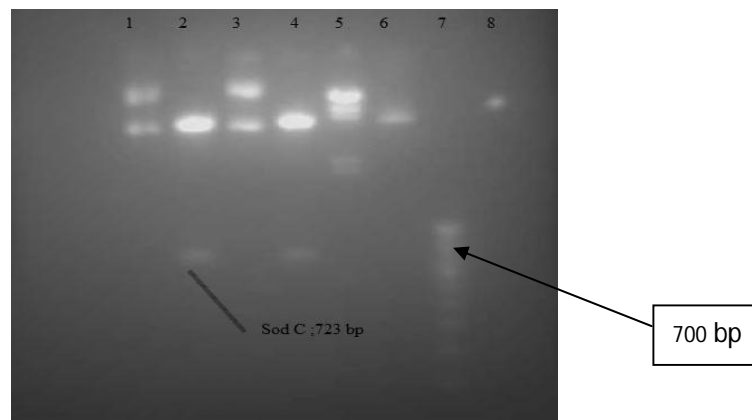
The PCR of recombinant plasmids showed amplification in six of the twelve colonies screened (#5, 7, 8, 9, 10,11) (**Figure 1.3.3.2**).



**Figure 1.3.3.2:** 1.5 % agarose gel analysis of Colony PCR of recombinant *sod C*: Lane 1 to 6: Colony #1 to 6; Lane 7: 100 bp ladder marker; Lane 8 to 13: Colony #7 to 12; Lane 14: negative control - Water; Lane 15: positive control – genomic DNA.

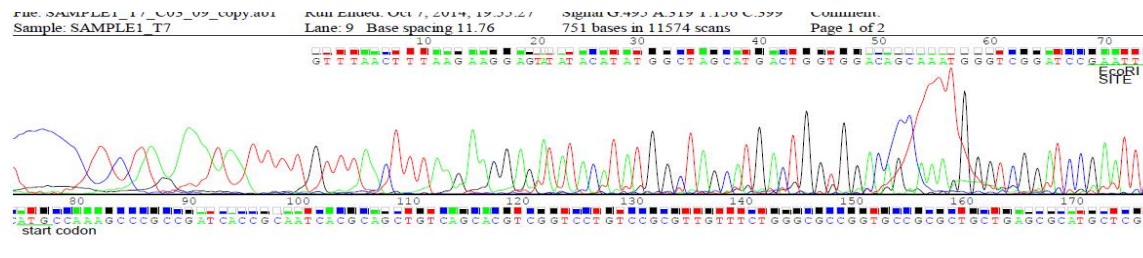
## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

Restriction digestion of the recombinants showed release of insert in two of the plasmids (#5, 8) (Figure 1.3.3.3).



**Figure 1.3.3.3:** 1.2% agarose gel analysis of Restriction digestion of *sod C* clones: Lane 1: Recombinant #1 (uncut); Lane 2: Recombinant #5 (treated with EcoRI and XhoI); Lane 3: Recombinant #8 (uncut) Lane 4: Recombinant #8 (treated with EcoRI and XhoI); Lane 5: Lambda DNA-HindIII Digest marker; Lane 6: vector control pET21a (cut with EcoRI /XhoI); Lane 7: 100 bp ladder; Lane 8: vector control pET21a (uncut).

Sequencing of the *sod C* clone #5 revealed three-point mutations in the 723 base pair sequence (Figure 1.3.3.4).



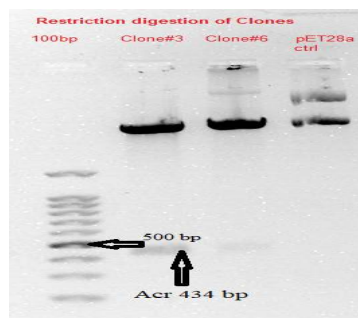
**Figure 1.3.3.4:** Sequencing of *sod C* clone #5.

The *sod A* gene (directional BamHI / XhoI) with ATG in the start codon, showed amplification of expected size of 624 base pairs (Figure 1.3.4.1).



### 1.4 Directional cloning of *acr* into NdeI and XhoI site of pET28a

The *acr* gene was amplified from the genome of *M. tb* H37Rv using the primers mentioned and an amplicon of expected size of 434 bases was obtained. The PCR-amplified *acr* gene was cloned into the pET28a vector and the construct was confirmed by restriction digestion of the plasmids (clone #3, 6) (**Figure 1.4.1**) and colony PCR of the recombinants #1 to 7 (**Figure 1.4.2**). The efficiency of directional cloning with phosphatase treatment was very high with nearly 50% of the recombinants showing the presence of the *acr* gene (#1 to 7) (**Figure 1.4.3**).



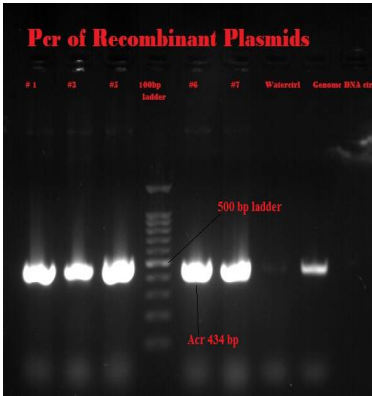
**Figure 1.4.1:** 1.2% agarose gel analysis of Restriction digestion of Clones: Lane 1: 100 bp ladder; Lane 2: Clone #3 digested with NdeI and XhoI; Lane 3: Clone #6 digested with NdeI and XhoI; Lane 4: pET28a control digested with NdeI and XhoI.



**Figure 1.4.2:** 1.5 % agarose gel analysis of Colony PCR of *acr* recombinants: Lane 1 to 4: Clone #1 to 4; Lane 5:100 bp ladder; Lane 6 to 8: Clone #5, 6, 7; Lane 9: Negative control - Water; Lane 8: Positive control - genomic DNA.



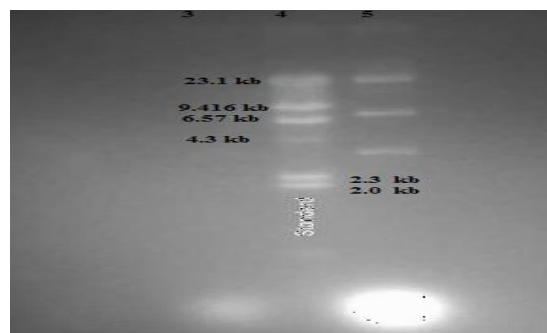
## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*



**Figure 1.4.3:** 1.2% agarose gel analysis of PCR of *acr* recombinant plasmids: Lane 1 to 3: Clone #1, 3, 5; Lane 4: 100 bp ladder; Lane 5, 6: Clone #6, 7; Lane 7: Negative control - Water; Lane 8: Positive control - genomic DNA.

**Table 1.5 Development of software to estimate DNA:** The software developed in MATLAB was used to estimate the amount of plasmid DNA run on the gel.

Lane	Area	Volume	Ratio	Mass by Software (ng/ $\mu$ l)	Mass by comparison to standard (ng/ $\mu$ l)
Standard	5472	173320	-----	500	
Lane 5	2375	39241	22.63%	28.288	33



**Figure 1.5:** Gel picture of sample used for testing software: 1% agarose gel analysis of plasmid preparation: Lane 3: pET21a plasmid (1  $\mu$ l); Lane 4: Lambda DNA-HindIII Digest; Lane 5: pET21a plasmid (4  $\mu$ l).

### DISCUSSION

Our work attempts to highlight the use of Excel analysis coupled with knowledge of molecular biology to help achieve consistency in batch preparations of vector and insert and increase the likelihood of obtaining a clone. Success was achieved with directional cloning; using two restriction enzymes and CIAP in the same reaction. In the present study, this approach was used to clone three genes; *sod A*, *sod C* and *acr*. In addition, we present a detailed parameter analysis of fifteen attempts of cloning; documenting eighteen important parameters that affect the success of the outcome. Classifiers were used to build predictive models. Binary classifiers predict output as 0 or 1 which can be used in the case of predicting success in obtaining a clone. This can be done assuming the overall success of all the intermediate steps; particularly, the ligation and later, the recombinant screening process. There are different classifiers that can be used for this. Using MATLAB and Clone. R, a predictive model could be built, using classifiers like Simple Tree, Linear discriminant, Logistic Regression / Generalised Linear model (GLM), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN) and Random forest (RF). These showed the success of the model in terms of % accuracy, area under the receiver operating characteristics (ROC), area under curve (AUC), recall, precision and specificity. This validated the eighteen parameters as an important tool to predict success of future cloning attempts.

In brief, using parameter analysis, we identified few important parameters that affect the outcome of any cloning experiment. By using classifier analysis in MATLAB and Clone. R, we identified Simple Tree and Random forest to be the best classifiers which could predict the possibility of obtaining a clone. Clone. R also identified volume of reaction and units of enzyme as the key determinants of success in obtaining a clone. As we carried out fifteen attempts, this classifier analysis model could be developed further with more cloning attempts and modified to choose the best model by adding more parameters and data points. The scope of this work can

## Chapter 1: Cloning of Mycobacterial Genes: *acr*, *sod A* and *sod C*

also be expanded to protein expression and activity analysis. All five constructs were subsequently checked for protein expression.

## **CONCLUSIONS**

Success was achieved using directional cloning using two restriction enzymes and CIAP. This approach was used to clone three genes, *sod A*, *sod C* and *acr*. Parameter analysis included eighteen important parameters that affect the outcome of any cloning experiment. Lysis buffer volumes, ligase units, and reaction volumes were critical for cloning success. The units of enzyme and volume of enzymatic reaction was found to be the most optimum for a clone to be obtained using Clone. R. Five constructs were obtained, out of which, four were sequenced. The *acr* gene showed no mutation while the *sod A* and *sod C* showed three point mutations.

# **CHAPTER 2**

## **CHAPTER 2**

### **Expression and Purification Studies of Acr, Sod A and Sod C**

#### **Objective 2**

### INTRODUCTION

Expression of cloned genes from recombinant constructs depends upon many factors including the length of the gene, the antibiotic resistance markers used as well as the conditions of expression including temperature of growth, optical density at which induction is carried out, the hours of induction, the concentration of IPTG. The addition of tags such as Glutathione-S-Transferase (GST), Histidine, Streptavidin and Maltose Binding protein makes it easier to purify proteins. (Sorensen and Mortensen 2005, Gräslund et al. 2008). The use of pET vectors has numerous advantages and few drawbacks too. The pET series has evolved to more than forty in number. The pET21a vector has ampicillin resistance marker that is known to degrade over time. An alternative is to use the more expensive carbanecillin. The use of kanamycin as an antibiotic resistance marker is preferred as the gene product is more stable and less expensive. The pET28 series of vectors has advantages of both options of N-terminal and C-terminal histidine tag and the kanamycin resistance marker (Novagen, *pET System Manual*, 11<sup>th</sup> Edition).

Expression of Mycobacterial proteins in *E. coli* is difficult due to codon usage which differs between Mycobacteria and *E. coli*. Another issue is the difference between the activity of recombinant and native protein, as in the case of Acr (Taylor et al. 2012).

*sod A* was cloned and expressed both for cytokine upregulation studies (Liao et al. 2013) as well as a booster to BCG vaccine (Jain et al. 2011). Earlier reports exist of its cloning in *M. smegmatis* as it is difficult to express in *E. coli* in a secreted form (Harth and Horwitz. 1999). Expression of *sod A* is difficult as the start codon is GTG instead of ATG which is the regular start codon. Options include to modify the codons to increase the level of expression in *E. coli* as has been shown by a report in which four Mycobacterial proteins, Antigen 85A, B, C and Superoxide dismutase were expressed at much higher levels (Lakey et al. 2000).

## Chapter 2: Expression and Purification Studies of Acr, Sod A and Sod C

Sod C was isolated from native *Mycobacterium* as well as cloned in a recombinant form (Wu et al. 1998, Orazio et al. 2001). Studies with *sod C* mutants were carried out to investigate whether knocking out the *sod C* gene increases the susceptibility of *M. tb* clearance from macrophages (Piddington et al. 2001, Dussurget et al. 2001).

Acr has been cloned and expressed in *E. coli* in the pET vectors (Chang et al. 1996, Yang et al. 1999, Valdez et al. 2002, Dhepakson et al. 2008). Reported expression levels range from 5 to 20 mg/L. In most of these constructs, the protein was produced in a soluble form enabling characterization and chaperone activity studies. The protein is active in an oligomeric form and it is well documented that the mechanism of action involves a reversible association and disassociation of different oligomers (Feng et al. 2002). Mutants too, have been constructed to study the roles of different domains in activity (Panda et al. 2017). Acr has also been expressed in soluble form along with other important proteins and tested as a mucosal booster vaccine to BCG (Copland et al. 2018).

Aim of our study was to check the expression of the constructs we made, using *E. coli* BL21DE3 and BL21DE3 star cells and optimising different conditions of expression to get soluble active protein.



## MATERIALS AND METHODS

**2.1. Expression studies of *sod A*-pET21a:** The *sod A* clone (GTG as the start codon) was tested for expression using *E. coli* BL21DE3 cells at different temperatures, O. D. of induction, time of incubation to look for soluble protein. Different temperatures; 37°C, 25°C and 16°C were attempted to optimize the best condition for soluble expression.

**2.2. Expression studies of C-terminal His-tag Sod A:** Expression of Sod A (ATG as the start codon) was tried out using *E. coli* BL21DE3 and *E. coli* BL21DE3 star cells at 37°C.

**2.3 Expression studies of C-terminal His-tag Sod C:** Expression of Sod C protein was tried out using *E. coli* BL21DE3 and *E. coli* BL21DE3 star cells at 37°C.

### **2.4. Expression study of C-terminal His-tag Acr**

**2.4.1 Small scale expression studies of Acr:** The recombinant plasmids of clone #1 and 5 were transformed into *E. coli* BL21DE3 cells and induction carried out at 37°C with 1 mM IPTG for three hours at O. D<sub>600</sub> 0.60 in 1.5 ml culture and checked for expression by SDS-PAGE. Clone #1 was used for all further studies.

**2.4.2 Large scale expression studies of Acr:** Acr was expressed in 25 ml LB by inducing with 1 mM IPTG at 37°C for six hours and later at 200 ml scale. Aliquots of 1.5 ml were spun down at 8000 g at 4°C and pellets resuspended in 1X phosphate buffered saline (PBS) and sonicated. The sonicate was spun at 13000 g for 20 mins. at 4°C. Aliquots of supernatant and pellet were loaded on 15% SDS-PAGE to check for presence of protein in supernatant as well as pellet.

**2.4.3 Denaturing Nickel-agarose purification to verify expression of Acr protein:** The 25 ml cell pellet grown in LB was lysed using 0.5 ml of lysis buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 300 mM NaCl, 8M urea pH 8.0), sonicated on ice for 6 pulses of 30 seconds each. The sonicate was spun at 13000 g for 20 mins. at 4°C and the supernatant bound to 100 µl of Nickel-NTA beads (Qiagen) at 4°C for one hour. Two washes of 1.5 ml each were carried out in equilibration buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 M NaCl, 8 M Urea pH 8.0) and equilibration buffer plus imidazole 10 mM (MP

## Chapter 2: Expression and Purification Studies of Acr, Sod A and Sod C

Biomedicals) respectively. Elution was carried out in 3 X 100 µl elution buffer of 20 mM Tris.Cl pH 7.5, 8 M Urea, 250 mM imidazole.

### 2.4.4 Optimization of soluble expression of Acr

As Acr expression was in the cell pellet, optimization of soluble expression was done by lowering the temperature to 30°C, 25°C and 16°C, varying the IPTG from 0.05 to 1 mM and varying the O. D<sub>600</sub> of induction from 0.6 to 1.0. The time of induction varied from four to sixteen hours. The final O. D. after induction was maintained between 1.5 to 2.0 to ensure enough amount of soluble protein was obtained.

### 2.4.5 Nickel-NTA agarose purification of soluble Acr protein

The optimisation of soluble Acr production was tried using different conditions of protein expression as indicated in 2.4.4. The purification conditions were also optimised by using various buffer conditions of equilibration, washing and elution. The concentration of NaCl in the load was also varied from 500 mM to 2 M in order to optimise binding conditions. The imidazole concentration in the equilibration, wash and elution also varied in order to minimize contaminant proteins of *E. coli* with His-tag from co-eluting with soluble Acr. The *acr* clone was sub-cultured into 150 ml of LB ampicillin 100 µg/ml and then inoculated into 1.5 L of LB, induced at 0.8 O. D<sub>600</sub> with 0.05 mM IPTG at 25°C for twelve hours. The cell pellet was lysed by sonication in 30 ml of 20 mM Tris pH 7.5, 750 mM NaCl, 5% glycerol, 10 mM imidazole, EDTA free protease inhibitor (Sigma Chem S-8830) and binding the sonicate supernatant to 500 µl of Nickel-NTA resin overnight at 4°C. The column was washed twice with 20 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 25 mM imidazole and 40 mM imidazole. The protein was eluted with 20 mM Tris pH 7.5, 300 mM NaCl, 300 mM imidazole. A repurification was carried out using the same protocol as above. The eluted fractions were concentrated and desalted using Amicon 10 kDa or 30 kDa membrane, cut-off and the concentrated samples run on 15% SDS-PAGE before performing the activity assays.

### 2.5 Expression study of N-terminal His-tag Acr

**2.5.1 Expression and purification studies:** Expression of the two Acr containing clones, #3 and #6 was initially carried out in 50 ml of LB kanamycin (50 µg/ml) in *E. coli* BL21DE3 cells, using 1mM IPTG induction at 37°C for three hours. Aliquots of 4 ml of uninduced and induced cultures were spun down at 8,000 g at 4°C for 10 mins. The cell pellets thus obtained, were resuspended in 200 µl 6X SDS loading dye, boiled at 85°C and loaded on 15% SDS-PAGE to check for expression of protein. The rest of the sample was sonicated using 10 mM Tris pH 7.0 / 5% glycerol, followed by centrifugation at 20,000 g at 4°C for 30 mins. Aliquots of supernatant and pellet obtained were loaded on 15% SDS-PAGE to check for the localization of protein. Scale up expression studies of the recombinant Acr was done using a freshly streaked plate of the clone. A single colony was used to inoculate a primary culture of LB broth, pH 7.0 ± 0.2. This culture was incubated for twelve hours at 37°C and then sub-cultured into secondary culture with inoculum size of 1%. This secondary culture was grown till the O. D<sub>600</sub> of cells reached 0.4 to 0.6 and then induction was carried out with 0.5-1.0 mM IPTG for three hours at 37°C ± 0.2. For the optimization of expression and purification, varying volumes of secondary culture ranging from 50 ml to 1.8 L were used.

**2.5.2 Purification of Acr using Nickel-NTA agarose (1.5 L scale):** The Nickel-NTA purification process was optimised by using varying concentrations of NaCl and imidazole in Tris 20 mM pH 7.0 / 5% glycerol. Earlier purification data from the C-Terminal His-tag Acr purification indicated better binding of Acr to Nickel-NTA at pH 7.0 in the presence of 500 – 750 M NaCl in the load sample. These conditions were utilised for end-terminal His-tag Acr purification. The optimised protocol for the same was as follows:

The *acr* clone #3 was inoculated into 15 ml of LB kanamycin (50 µg/ml), sub-cultured into 1.5 L and then induced with 0.5 - 1 mM IPTG at 0.4 - 0.6 O. D<sub>600</sub> for three hours at 37°C. The cell pellet was lysed by sonicating in 60 ml lysis buffer containing 20 mM Tris pH 7.0, 500 mM

## Chapter 2: Expression and Purification Studies of Acr, Sod A and Sod C

NaCl, 5% glycerol and 2X Sigma Fast protease inhibitor (Catalogue no. S8830). The sonicate was centrifuged at 20,000 g for 30 mins. at 4°C and the supernatant obtained was bound to 3 ml of Nickel-NTA resin (HisPur™ Ni-NTA Resin Catalogue no. 88221, Thermo Scientific) for two hours at 4°C, assuming a binding capacity of 20 mg/ml. The column was washed with 10 and 100 mM imidazole in buffer containing 20 mM Tris pH 7.0, 500 mM NaCl, 5% glycerol. The protein was pre-eluted with buffer containing 20 mM Tris pH 7.0, 300 mM NaCl, 5% glycerol and then eluted with a 3-step gradient of 300 mM, 400 mM and 500 mM imidazole. Multiple batches of Acr were purified by using scale-down versions of the above protocol. The soluble protein was estimated by Bradford's dye-binding assay at 595 nm and a binding capacity of 20 mg/ml was maintained. Multiple runs were carried out from the same batch in order to ensure high levels of purity. The protein eluted from greater than 95% purity fractions was used for N-terminal protein sequencing and determination of molecular mass. These elutes were thereafter referred to as **H** samples.

**2.5.3 Gel filtration of Acr:** The Nickel-NTA eluted protein fractions containing Acr (**H**) were dialyzed against 20 mM Tris pH 7.0, 100 mM NaCl and 5% glycerol. 1.5 ml of the dialyzed protein; equivalent to 4 mg as estimated by Bradford's dye-binding assay, was loaded twice on Sephacryl-200 Hiprep XK pre-packed 16/60 column (120 ml) at a flow rate of 0.5 ml/min. and buffer passed for a total of 1.5 column volumes using the AKTA Purifier System at C-CAMP, Bengaluru, India. The column was equilibrated using the same buffer and calibrated using BIORAD standards (Catalogue no. 151-1901) ranging in size from 1.5 kDa to 670 kDa. A plot of log molecular weight of the standards versus the retention volume was plotted and the approximate molecular weight of the eluted Acr protein estimated. A total of eight gel filtration runs were done with four batches of Acr. These elutes were thereafter designated as **G** samples. The amount of protein to be loaded in each run was based on the protein concentration of the **H**

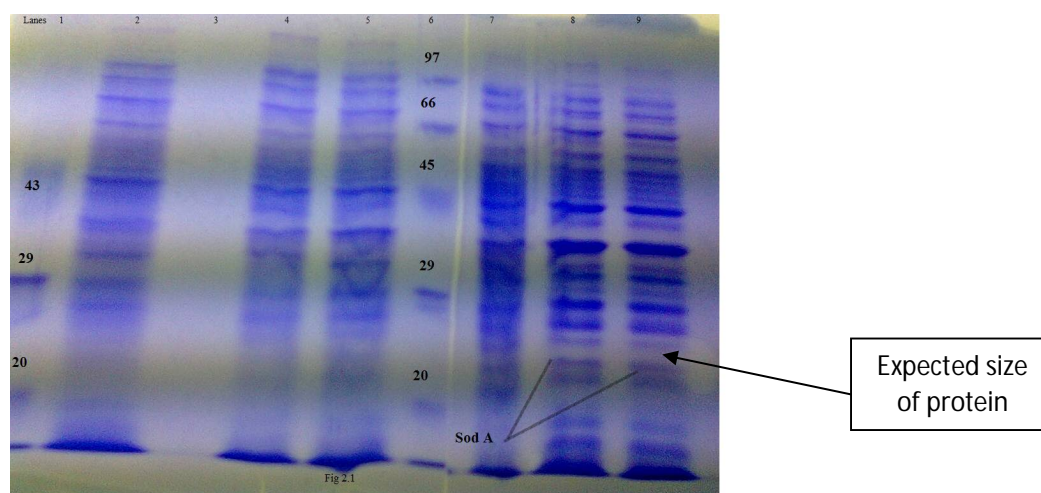
samples. The limiting factor was the volume of load as a percentage of the gel filtration column volume with the guideline being 1 – 5% for optimum purification.

**2.5.4 Identity of Expressed Protein by Matrix Assisted Laser Desorption/Ionisation - Time of Flight (MALDI-TOF):** The **H** sample band was cut out from the SDS-PAGE gel and subjected to tryptic digestion to confirm the identity of the protein by MS/MS. The tryptic digest sequences were matched with existing peptide sequences using MASCOT software. Another band was cut out and subjected to Molecular Mass Analysis by MALDI-TOF to confirm the mass of the protein. Both these studies were carried out at Sandor Life Sciences, Hyderabad, India.

**2.6 Native Polyacrylamide Gel Electrophoresis (Native-PAGE) analysis:** An 8-16% Native-PAGE gel was used to run the **H** sample and the **G** sample along with a control of Bovine Serum Albumin (BSA) to check the size of oligomers present. The distance migrated by the BSA monomer, dimer and trimer; 66, 132 and 198 kDa respectively; was plotted against log molecular weight in order to estimate the oligomeric sizes of **H** and **G** samples. A variation of  $\pm 0.05$  cm was assumed for distance migrated and  $\pm 10\%$  was assumed for calculating oligomeric size. Image J software was used to estimate the relative proportion of the oligomers.

## RESULTS

**2.1: Expression of sod A (GTG) as the start codon:** No expression of SOD A was seen at 37°C. All the different conditions of growth at lower temperatures did not show any expression except at 25°C overnight induction with 2 mM IPTG, where faint expression could be seen. As the clone did not have an affinity tag, it was decided to re-clone with ATG as the start codon to observe if the expression could be improved with the change in the start codon (**Figure 2.1**).



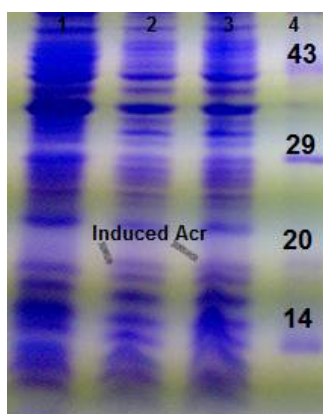
**Figure 2.1:** 15% SDS-PAGE analysis of 25°C expression of sod A with GTG as start codon:  
Lane 1: Low molecular weight marker (43, 29, 20 kDa); Lane 2: Uninduced 0 hours sample 30 ml; Lane 3: Empty; Lane 4: Uninduced 3 hrs 25°C 15ml; Lane 5: Induced 3 hrs 25°C; Lane 6: Medium range molecular weight marker (97, 66, 45, 29, 20 kDa); Lane 7: Uninduced sample: 8 µl; Lane 8: Induced sample 1 mM IPTG:12 µl; Lane 9: Induced sample 2 mM IPTG:12 µl.

**2.2: Expression studies of C-terminal His-tag sod A with ATG as start codon:** No expression was observed at 37°C in both BL21DE3 and BL21DE3 star cells. A probable reason for this could be the point mutations we observed in the sod A gene as well as leaky expression of the vector pET21a. The expression results were highly irreproducible, indicating a possibility of plasmid loss and instability, a commonly observed phenomenon in pET series of vectors.

This could also be due to the low stability of the ampicillin resistant marker, whose gene-orientation is opposite to that of the origin of replication.

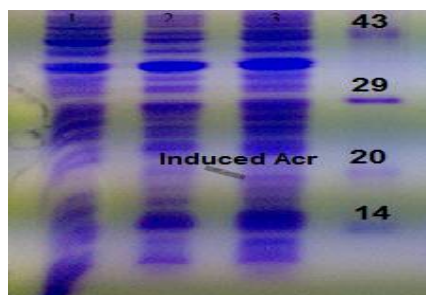
**2.3: Expression studies of C-terminal His-tag sod C:** No expression was observed at 37°C in both BL21DE3 and BL21DE3 star cells. This could also be due to the point mutations observed in the sod C gene and similar reasons as stated in 2.2.

**2.4: Expression studies of C-terminal His-tag Acr:** 1.5 ml scale expression study showed a band of 18 kDa (**Figure 2.4.1.1**) in the induced sample of both the clones (*acr* #1, 5) on SDS-PAGE. The Acr band was detected to be nearly 18 kDa instead of 16 kDa as reported earlier due to the presence of six extra His-residues and the T7 tag. The *acr* clone #1 was used for further studies. The level of expression was low with the percentage of the protein expressed less than 10%.



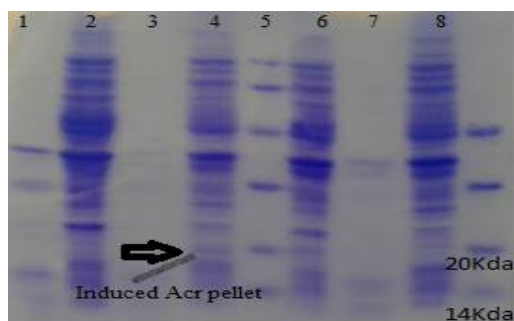
**Figures 2.4.1.1: 15% SDS-PAGE analysis of expression of Acr in BL21DE3 cells (1.5 ml scale):** Lane 1: Uninduced; Lane 2: Induced recombinant clone #1; Lane 3: Induced recombinant clone #5; Lane 4: Marker (Bangalore Genei) (43, 29, 20, 14 kDa).

At higher volume of 25 ml, protein was expressed at the same level (**Figure 2.4.1.2**) showing no significant increase.



**Figure 2.4.1.2: 15% SDS-PAGE analysis of 25 ml sonicated protein:** Lane 1: 0 hour induction; Lane 2: Uninduced; Lane 3: Induced 6h; Lane 4: Marker (Bangalore Genei) (43, 29, 20, 14 kDa).

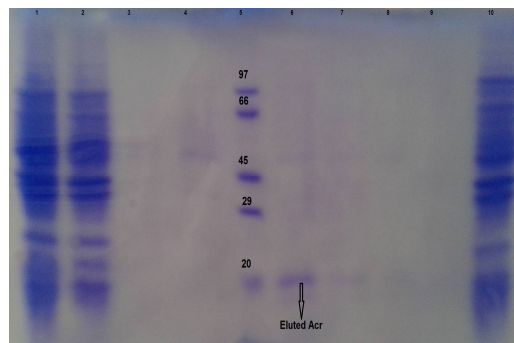
The 25 ml cell pellet was sonicated in the presence of 1X PBS and centrifuged at 13000 g to separate the supernatant and cell pellet. The protein appeared in the pellet indicating that the protein was insoluble at 37°C. The lysis of a 200 ml cell pellet at 37°C showed the localization of the protein in the insoluble fraction as indicated in the **Figure 2.4.2**. It was extremely difficult to visualise the protein even after vector controls of induced and uninduced; indicating very low levels of expression.



**Figure 2.4.2: 15% SDS-PAGE analysis of 200 ml after sonication:** Lane 1: Uninduced Acr LB supernatant; Lane 2: Uninduced Acr LB pellet; Lane 3: Induced Acr LB supernatant; Lane 4: Induced Acr LB pellet; Lane 5: Marker (Bio-Rad) (97, 66, 43, 29, 20, 14 kDa); Lane 6: pET control (minus insert) uninduced pellet; Lane 7: pET induced (minus insert) supernatant; Lane 8: pET induced (minus insert) pellet; Lane 9: Marker (43, 29, 20, 14 kDa).

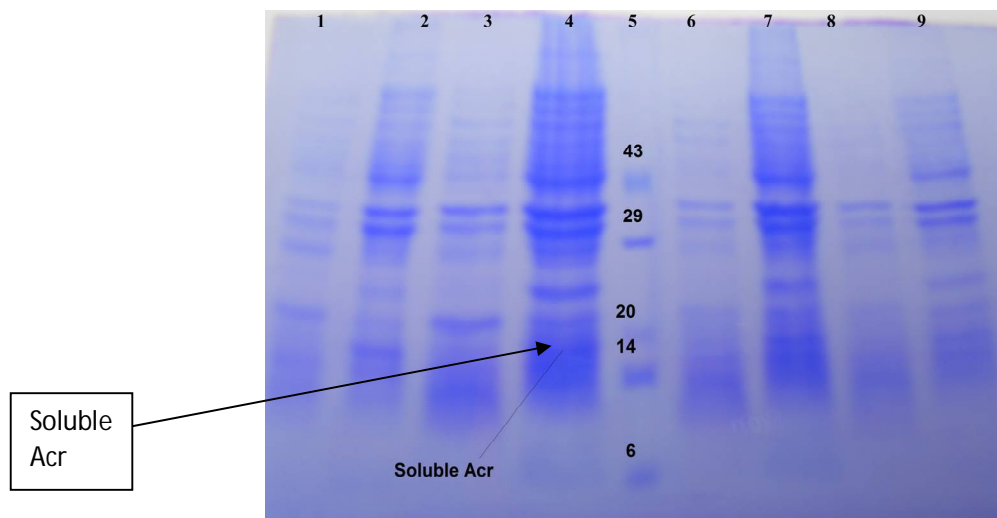


**Purification on Nickel-NTA of Acr pellet under denaturing conditions:** The 25 ml protein band, expressed at 37°C, was bound and eluted to Nickel-NTA beads under denaturing conditions. A band appeared in the eluted fraction E1 suggesting that it is the Acr protein (**Figure 2.4.3**). The molecular weight was approximately 18 kDa. This was indirect evidence of the faintly induced band being Acr. It is possible that under denaturing conditions, the protein that is expressed is completely available to bind to the Nickel-NTA. However, attempts were made to solubilise the protein as denaturing conditions may not be viable for biological activity of Acr.



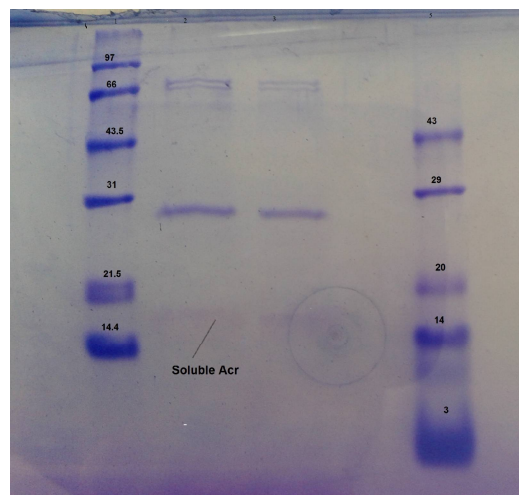
**Figure 2.4.3: 15% SDS-PAGE analysis Nickel-NTA purification of Acr 25 ml cell pellet lysed in 8M Urea:** Lane 1: Load Acr; Lane 2: Flow through; Lane 3: Wash 1; Lane 4: Wash 2; Lane 5: Marker (Bio-Rad) (97, 66, 43, 29, 20 kDa); Lane 6: Elution 1; Lane 7: Elution 2; Lane 8: Elution 3; Lane 9: whole sonicate before spin.

**Soluble expression of Acr:** Growth of *acr* clone #1 at 16°C and 30°C did not give any soluble Acr protein. However, when the 1.5 L cell pellet was grown at 25°C and induced at 0.8 O. D. and 0.05 mM IPTG, maximum soluble protein was obtained (**Figure 2.4.4**). These levels of soluble expression were too low for further studies; yet, large-scale expression studies were attempted to make enough protein for carrying out chaperone activity assays.



**Figure 2.4.4: 15% SDS-PAGE analysis of Soluble Acr expression at 25°C, 0.8 O. D, 0.05 to 0.20 mM IPTG:** Lane 1: Uninduced sonicate supernatant; Lane 2: Uninduced sonicate pellet; Lane 3: Induced sonicate supernatant; Lane 4: Sonicate pellet induced 0.05 mM IPTG; Lane 5: Marker (Bangalore Genei) (43, 29, 20, 14, 6 kDa); Lane 6: Induced sonicate supernatant; 0.1 mM IPTG; Lane 7: Induced sonicate pellet 0.1 mM IPTG; Lane 8: Induced sonicate supernatant 0.2 mM IPTG; Lane 9: Induced sonicate pellet 0.2 mM IPTG.

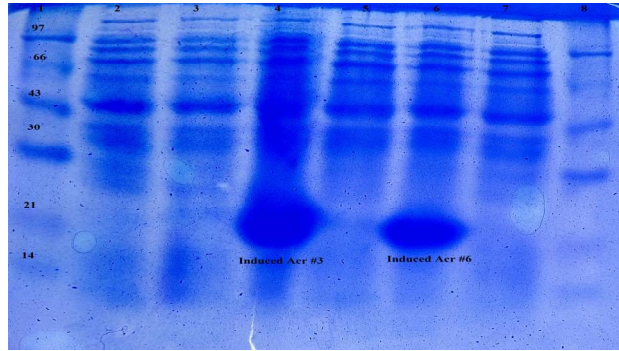
**Nickel-NTA purification of soluble ACR from 1.5 L:** Nickel-NTA purification showed the elution of an 18 kDa band from the soluble fraction along with two contaminants of higher molecular weight. The first pool of fractions E2 to E6 was repurified on Nickel-NTA column using the same conditions. The second pool of E7 and E8 fraction was used directly for thermal aggregation assay. Following the purification, the concentrated ACR sample was run on 15% SDS-PAGE (**Figure 2.4.5**).



**Figure 2.4.5: Nickel-NTA Purification of Soluble ACR:** Lane 1: Marker (Bio-Rad) (97, 66, 43, 31, 21.5, 14.4 kDa); Lane 2, 3: Purified concentrated Acr after two Nickel-NTA purifications; Lane 5: Marker (Bangalore Genei) (43, 29, 20, 14, 3 kDa).

The Acr protein appeared in the insoluble form and bound and eluted to Nickel-NTA in 8 M urea, indirectly confirming the identity of the protein. Attempts to solubilize the protein at different temperatures produced low amounts of soluble protein at 25°C and 0.05 mM IPTG. However, there was no reproducibility of batch to batch of C-terminal His-tag Acr; hence; expression was tried with the N-terminal His-tag Acr.

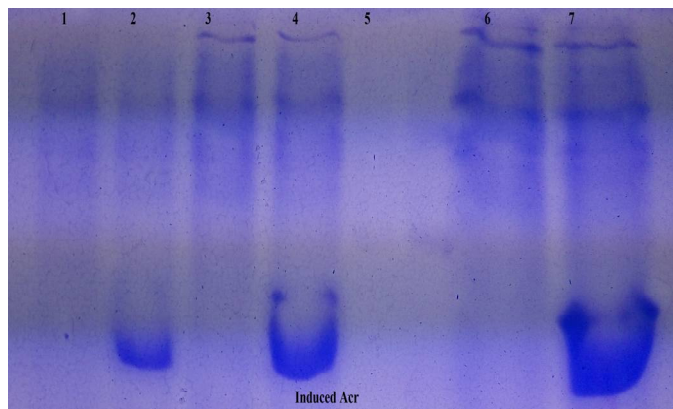
**2.5.1 Expression of N-terminal His-tag Acr:** 50 ml culture showed expression of Acr after freshly transforming *E. coli* BL21DE3 with pET28a construct and inducing with 1 mM IPTG at 0.4 to 0.6 O. D<sub>600</sub> for three hours at 37°C. Two clones labelled #3 and #6 were selected for expression. The expression level was found very high with the concentration estimated in the range of approximately 50-60 mg of recombinant protein/L. The Acr protein ran closer to the 21 kDa marker, a little higher than the expected 18 kDa. Clone #3 showed higher protein expression and was selected for expression studies subsequently (**Figure 2.5.1.1**).



**Figure 2.5.1.1: 15% SDS-PAGE analysis of Expression of *acr*-pET28a #3 and #6**

Expression gel *acr*-pET28a Lane 1: Marker (97, 66, 43, 30, 21, 14 kDa); Lane 2: uninduced vector; Lane 3: induced vector; Lane 4: induced clone #3; Lane 5: uninduced clone #3; Lane 6: induced clone #6; Lane 7: uninduced clone #6; Lane 8: Marker (97, 66, 43, 30, 21, 14 kDa).

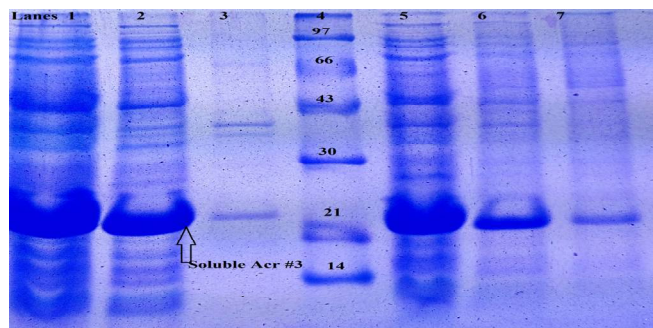
Varying the O. D. of induction from 0.4 to 0.6 produced soluble protein in the range of 40 – 60 mg/L and at IPTG concentration of 0.5 mM, expression levels were in the same range.



**Figure 2.5.1.2: 15% SDS-PAGE analysis of 0.5 mM IPTG expression of *acr*-pET28a #3:**

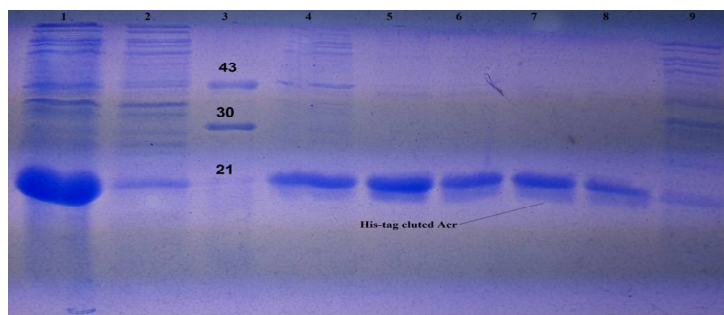
Lane 1, 3, 6: Uninduced Acr 10, 20, 40 µl; Lane 2, 4, 7: Induced Acr 10, 20, 40 µl.

**Soluble protein localisation:** The protein was found in the soluble fraction after expression at 37°C, followed by sonification and centrifugation in clones #3 and #6 (**Figure 2.5.1.3**).



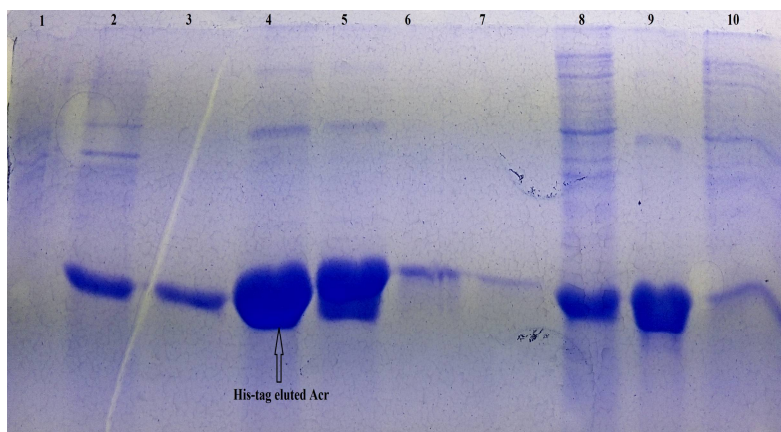
**Figure 2.5.1.3: 15% SDS-PAGE analysis of localization of supernatant/pellet of *acr*-pET28a #3, #6:** Lane 1: Whole Lysate #3; Lane 2: Sonicate Supernatant #3; Lane 3: Sonicate Pellet #3; Lane 4: Marker (97, 66, 43, 30, 21, 14 kDa); Lane 5: Whole Lysate #6; Lane 6: Sonicate Supernatant #6; Lane 7: Sonicate Pellet #6.

**2.5.2 Nickel-NTA Purification:** The soluble Acr protein bound to the Nickel-NTA column was eluted by the 3-step 300, 400, 500 mM imidazole gradient. The protein eluted with least non-specific impurities in the 500 mM imidazole fractions, indicating strong binding affinity. The washes removed all the non-specific His-tag *E. coli* contaminant proteins. The protein was greater than 95% pure on SDS-PAGE (**Figure 2.5.2.1**). Another representative run of purification showed similar result (**Figure 2.5.2.2**).



**Figure 2.5.2.1: 15% SDS-PAGE analysis of Nickel-NTA purification of Acr:** Lane 1: Load; Lane 2: Flow through; Lane 3: Marker (Bangalore Genei) (43, 30, 21, 14, 6, and 3 kDa); Lane 4: E1-E3; Lane 5: E5; Lane 6: E6; Lane 7: E7; Lane 8: E8; Lane 9: Wash 1 + Wash 2. Elution gradients E1-E2 (300 mM imidazole), E3 (400 mM imidazole), E5-E8 (500 mM imidazole).

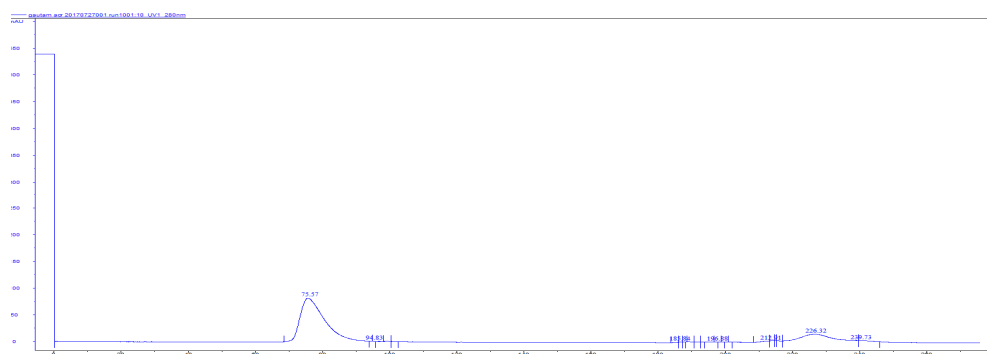




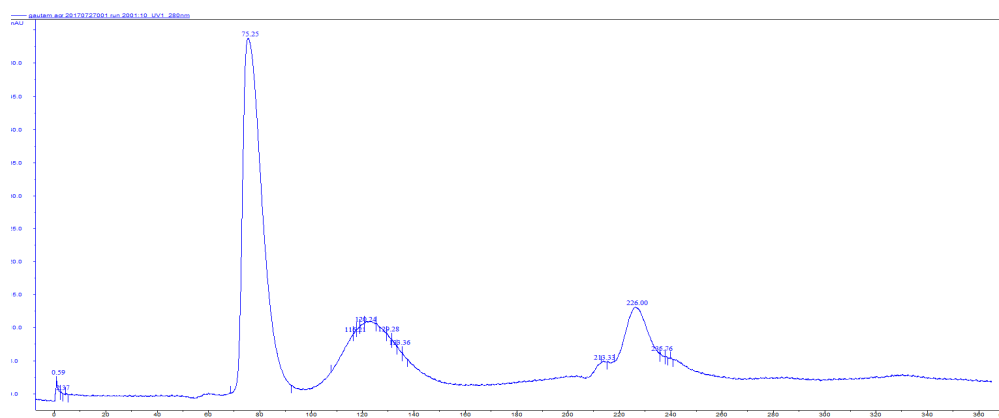
**Figure 2.5.2.2: 15% SDS-PAGE analysis of different batch of Nickel-NTA purification of *acr*-pET28a:** Lane 1: uninduced Acr; Lane 2: induced Acr; Lane 3: fractions E1-3; Lane 4: E4-7; Lane 5: E8-11; Lane 6: E12-15; Lane 7: E16-20; Lane 8: Load (Soluble Acr); Lane 9: Earlier batch final; Lane 10: Wash 1 + 2.

**Gel Filtration Purification:** The gel filtration run was carried out in the AKTA purifier system and the protein eluted in 1.5 ml fractions, labelled as A1 to A12, B12 to B1, etc. The gel filtration experiments showed the Acr peak appearing in the void volume between the size of 670 kDa and 158 kDa (relative to the standards); suggesting its probable oligomeric nature. In the first run, a major peak was observed at 37.5 ml (75 min) with a height of 80 mAU while the peak ended at 44 ml (88 min.), the position corresponding to the globulin peak of 158 kDa (**Figure 2.5.3.1**). In the second run, a peak was observed at the same time as the first run with a height of 50 mAU, while the peak ended again at 44 ml. Several smaller peaks were observed later (**Figure 2.5.3.2**). All other gel filtration runs showed a similar elution time of the Acr protein (75 min). The Sephacryl S200 Hi Prep 16/60 column was calibrated with Bio-Rad standards (Catalogue no.151-1901). All the five peaks (aggregates with thyroglobulin 670 kDa, Globulin 158 kDa, Ovalbumin 44 kDa, Myoglobin 17 kDa and vitamin B<sub>12</sub> 1.5 kDa) appeared as per the expected pattern (**Figure 2.5.3.3**). A 10% reducing SDS-PAGE was run to analyze the peaks obtained in the gel filtration runs (**Figure 2.5.3.4, 2.5.3.5**). Analyses of later batches were

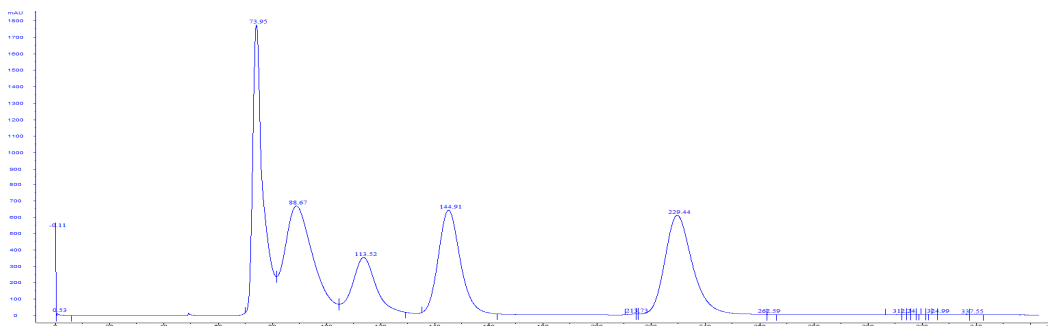
done on 15% PAGE. The peak 1 of both the runs showed the Acr band along with an upper band in the position of 36 kDa which could be a dimeric form. However, it was not clear whether the dimer appeared because of less amount of  $\beta$ -mercaptoethanol. This needs to be confirmed with further studies. Preliminary study indicated that increasing the amount of  $\beta$ -mercaptoethanol reduced the amount of dimer (data not shown). No protein was observed in the other two peaks of both the runs indicating that the protein predominantly exists in an oligomeric form as it was eluted in the void volume. Future batches of gel filtration runs were used for assays and designated as **G** samples (Run 1, pool 1 and pool 2) and **G** samples (Run 2). These samples along with **H** samples of the corresponding batch were used for activity analysis (Chapter 3).



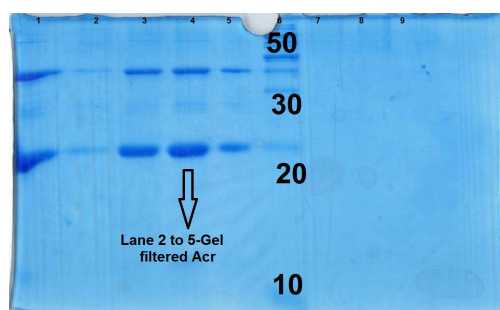
**Figure 2.5.3.1: Gel filtration chromatogram run 1: X Axis: UV 280 nm; Y Axis: Elution time (mins.).**



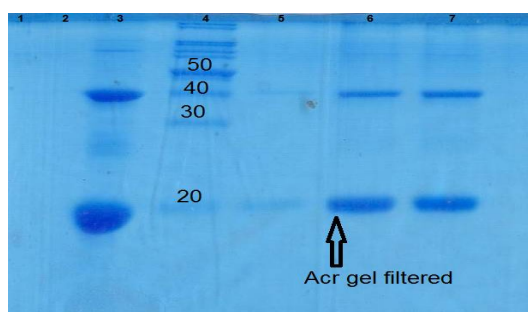
**Figure 2.5.3.2: Gel filtration chromatogram run 2: X Axis: UV 280 nm; Y Axis: Elution time (mins.).**



**Figure 2.5.3.3: Gel filtration Chromatogram of Bio-Rad Standards:** X Axis: UV 280 nm; Y Axis: Elution time (mins.) A - Aggregates + Thyroglobulin 670 kDa, 36.5 ml (73 mins.); B - Globulin 158 kDa, 44ml (88 mins.); C - Ovalbumin 44 kDa, 55 ml (110 mins.); D - Myoglobin 17 kDa, 77 ml (154 mins.); E - vitamin B<sub>12</sub> 1.5 kDa, 115 ml (230 mins.).



**Figure 2.5.3.4: 10% Reducing SDS-PAGE Analysis of Gel filtration Run 1:** Lane 1: Load; Lane 2-5: Peak 1 fractions of Run 1 (B2, B1, C1, C2); Lane 6: Marker (Puragene) (200, 150, 100, 85, 70, 60, 50, 40, 30, 20, 10 kDa); Lane 7-9; Later fractions (G2, G3, G4).



**Figure 2.5.3.5: 10% Reducing SDS PAGE analysis of gel filtration Run 2:** Lane 1: D8 Peak 2; Lane 2: Load; Lane 3: Marker (Puragene) (200, 150, 100, 85, 70, 60, 50, 40, 30, 20, 10 kDa); Lane 4-6: Run 2 Peak 1 (B2, B1, and C2); Lane 7: Peak 2 and 3 (D9, D10, G2, and G3).



**2.5.4 Sequencing:** The molecular mass was found to be 18.3 kDa corresponding to the cloned sequence 16 kDa plus an additional 2.1 kDa due to the addition of 60 bases in the N-terminal sequence. Nine peptide fragment sequences matched 100 % with Acr sequence using the MASCOT software.

K. DFDGR.S; K. DFDGR.S; K. HIQIR.S; R. LEDEMKE; R. DGQLTIKA; K. AERTEQK.D;  
R. TEQKDFDGR.S; R. SEFAYGSFVR.T; K. GILTVSVAVSEGKPTK.H

The protein score obtained was thirty-four and matched with *M. tb* H37Rv Strain ATCC 25618 (Figure 2.5.4.1 and Figure 2.5.4.2). The original Acr sequence is given below and the matched peptide fragment is highlighted.

MATTLPVQRHPRSLFPEFSELFAAFPSFAGLRPTFDTRLMRLEDEMKEGRYEV-  
RAELPGVDPDKDVDIMV RDGQLTIKAERTEQKDFDGRSEFAYGS-  
FVRTVSLPVGAEDEDDIKATYD KGILTVSVAVSEGKPTKHIQIRSTN

# Chapter 2: Expression and Purification Studies of Acr, Sod A and Sod C

09/2017

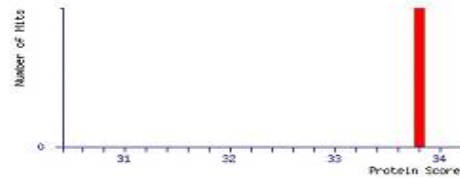
Protein Summary Report (../data/20170509/F005036.dat)

## MASCOT Search Results

User : madhurekha  
 Email : madhurekha@sandor.co.in  
 Search title :  
 MS data file : peaklist\_bits2.xml  
 Database : Acr\_bits\_recombi\_bits\_recombi\_20170428 (1 sequences; 144 residues)  
 Timestamp : 9 May 2017 at 12:45:43 GMT  
 Top Score : 34 for **P81MK1**, Alpha-crystallin OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) GI=hspx PE=1 SV=1

### Mascot Score Histogram

Protein score is  $-10 \cdot \log(P)$ , where P is the probability that the observed match is a random event.  
 Protein scores greater than 13 are significant ( $p < 0.05$ ).



### Protein Summary Report

Format As:  [Help](#)  
 Significance threshold p<:  Max. number of hits:   
 Preferred taxonomy:

### Index

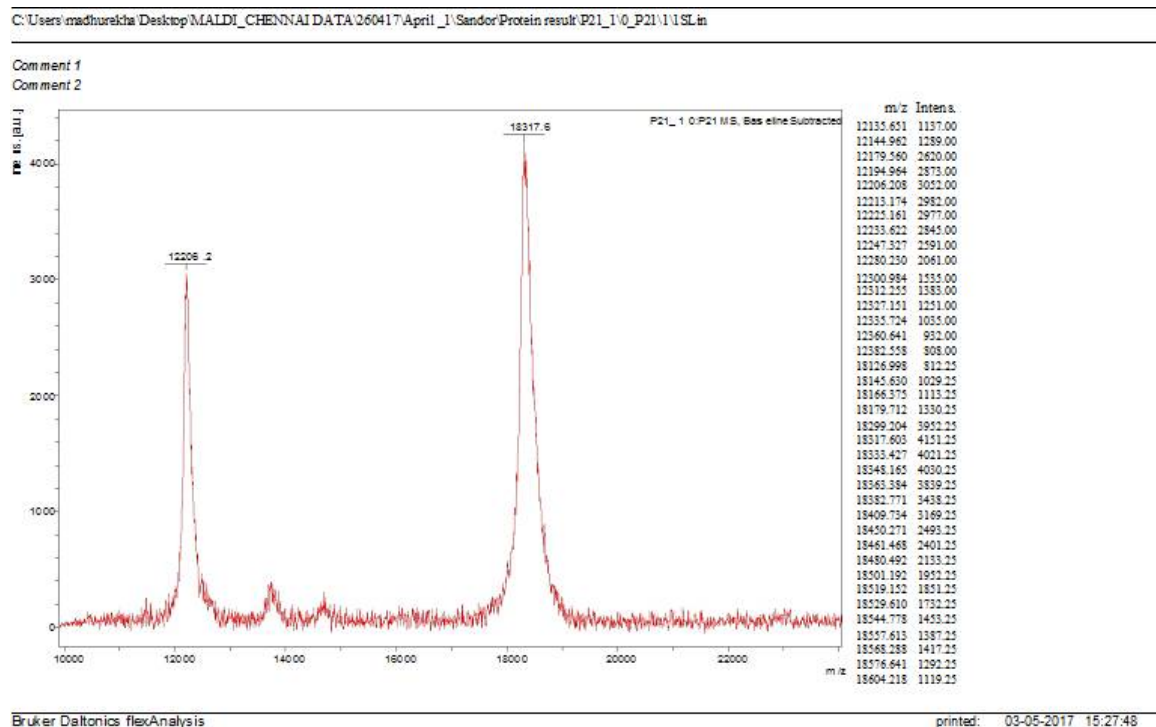
Accession	Mass	Score	Description
1. <a href="#">P81MK1</a>	16217	34	Alpha-crystallin OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) GI=hspx PE=1 SV=1

### Results List

1.	<a href="#">P81MK1</a>	Mass: 16217	Score: 34	Expect: 0.00042	Matches: 9
Alpha-crystallin OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) GI=hspx PE=1 SV=1					
Observed	Hr(expt)	Hr(calc)	Delta	Start	End Miss Peptide
608.8308	607.8235	608.2554	-0.4319	86 - 90	0 K.DFDGR.S
609.8805	608.6733	608.2554	0.4178	86 - 90	0 K.DFDGR.S
666.8237	665.8165	665.3973	0.4192	137 - 141	0 K.HIQIR.S
764.8375	763.8302	763.3422	0.4881	42 - 47	0 R.LEDHNK.E
774.8676	773.8603	773.4283	0.4320	72 - 78	0 R.DGQLTIK.A
861.8110	860.8038	860.4352	0.0686	79 - 85	1 K.AERTEQK.D
1096.0411	1095.0339	1094.4993	0.5346	82 - 90	1 R.TEQKDFDGR.S
1163.1285	1162.1213	1161.5455	0.5758	91 - 100	0 R.SEFAYGDFVR.T
1715.6747	1714.6674	1713.9513	0.7161	120 - 136	0 K.GILTVSVAVSEKPTK.H
No match to: 606.6309, 612.7325, 620.7944, 622.7819, 626.6592, 630.7980, 640.4260, 644.4280, 650.4455, 652.8769, 656.4269, 675.6072, 678.8486, 688.7974, 693.6328, 704.8080, 710.7335, 713.7141, 716.7866, 720.7647, 723.8483, 726.6305, 742.7720, 760.7230, 767.6964, 796.8510, 799.7317, 812.8333, 825.5672, 855.5107, 857.8354, 877.4792, 879.8135, 915.3592, 931.2989, 947.1418, 994.7922, 1060.5560, 1066.5597, 1074.2558, 1082.5596, 1143.9220, 1165.0465, 1198.4574, 1261.0443, 1334.8489, 1544.4989, 1665.6013, 1692.5739, 1737.7131, 1733.7446, 1769.6005, 1772.6455, 1791.5920, 1807.5834, 1822.8115, 1870.6569, 2069.0042, 2751.5652, 2764.3823, 2829.7378, 2831.2878, 2842.1479, 2856.8572, 2859.6243, 2871.5442, 2873.6406, 2874.6404, 2877.5320, 2891.5479, 2905.5220, 2913.6035					

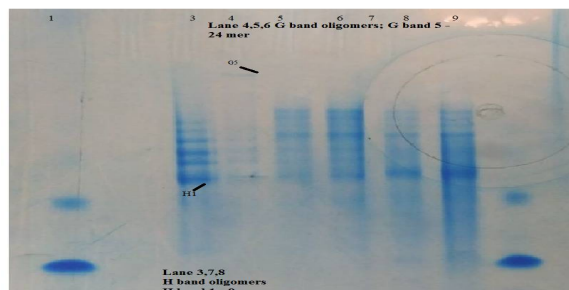
http://admin-pc/mascot/cgi/mascot\_results.pl?file=../%2Fdata%2F20170509%2F005036.dat&REPTYE=protein&sigthreshold=0.05&REPORT=55\_prefertaxo... 1/2

Figure 2.5.4.1: Tryptic digest of recombinant Acr and matching of peptide fragments using MASCOT.



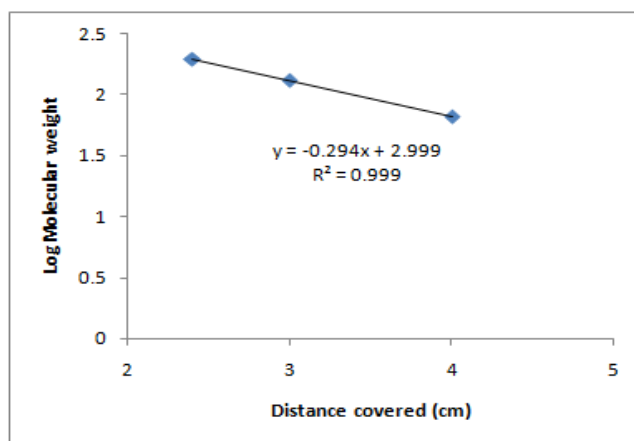
**Figure 2.5.4.2: Molecular Mass determination of recombinant Acr expressed by Clone #3**

**2.6 Native-PAGE analysis:** The non-denaturing 8-16% Tris-glycine gradient gel showed a mixture of 9 to 12 mers in case of **H** samples and 9 to 24 mers in case of **G** samples. The ratio of 9 mers was higher in the **H** samples (70%) as compared to the **G** samples where; the 24 mers ratio was higher and it appeared as a prominent band, running higher than all the other bands (**Figure 2.6.1**).



**Figure 2.6.1:** 8-16% Tris Glycine Native-PAGE analysis of H and G samples: Lane 1, 9: BSA control; Lane 3: **H**, Lane 4, 5, 6: **G**; Lane 7, 8: **H**.

A plot of log molecular weight versus distance migrated was used to estimate the size and ratio of the oligomers present (**Figure 2.6.2**). BSA (used as a marker) ran as monomer (66 kDa), dimer (132 kDa) and trimer (198 kDa).



**Figure 2.6.2:** Plot of log molecular weight of the different forms of BSA in Native-PAGE. The different sizes of BSA 66, 132 and 198 kDa were plotted against distance migrated, log molecular weight was plotted in MS Excel and the equation displayed and used to calculate oligomer size of **H** and **G** samples.

Log molecular weights obtained from **Figure 2.6.2** have been subsequently used to determine the oligomeric size of **H** and **G** samples as shown in **Table 2.6.1**. By analyzing the relative intensity of the bands, a proportion of each of the oligomeric size was estimated ranging from

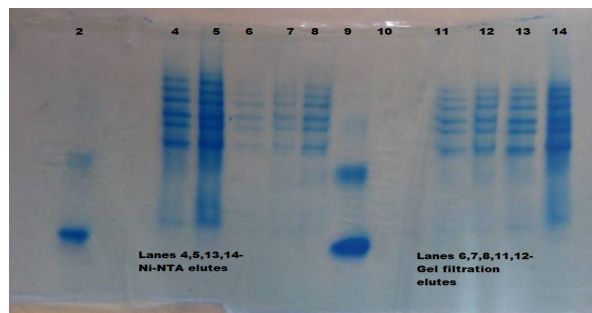
## Chapter 2: Expression and Purification Studies of Acr, Sod A and Sod C

15, 25 to 40% (**Table 2.6.1**). The relative intensity of the bands was analyzed by using the Image J software with a variation of  $\pm 10\%$ .

**Table 2.6.1: Oligomer ratio of batch 1:** Oligomer sizes were calculated, using BSA monomer, dimer and trimer as a reference (66, 132 and 198 kDa) and the log molecular weight plotted versus the distance travelled in centimetre. From this calculation, molecular weights have been extrapolated for the different bands viewed in **H** (band 1, 2, 3, 4) and **G** samples (band 1, 2, 3, 4 and 5) using the equation used in **Figure 2.6** ( $y = 0.294x + 2.999$ ). The proportion of bands was estimated using Image J software. For distance migrated, the error was  $\pm 0.05$  cm and for estimating the proportion, an assumption of  $\pm 10\%$  was made.

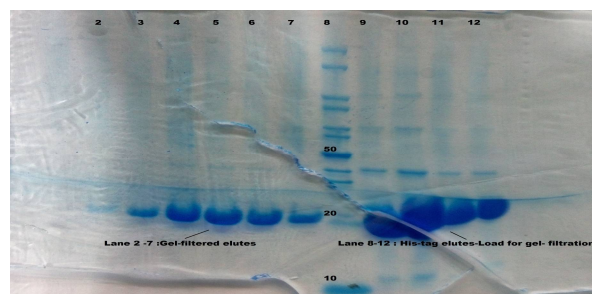
BSA (Mol. wt.)	Log Mol. wt.	distance (cm)	size (kDa) monomer 18	average oligomer size	% oligomer fraction
198	2.29	$2.4 \pm 0.24$			
132	2.12	$3 \pm 0.03$			
66	1.82	$4 \pm 0.14$			
<b>H</b> band 1	$2.21 \pm 0.04$	$2.7 \pm 0.05$	$162 \pm 15$	8.84	$70 \pm 7$
<b>H</b> band 2	$2.264 \pm 0.04$	$2.5 \pm 0.05$	$184.93 \pm 15$	10	101
<b>H</b> band 3	$2.29 \pm 0.03$	$2.35 \pm 0.05$	$204 \pm 18$	11.16	$10 \pm 1$
<b>H</b> band 4	$2.35 \pm 0.02$	$2.2 \pm 0.05$	$229.8 \pm 18$	12.48	$10 \pm 1$
				Ratio of 10 mers to 12.7	$30 \pm 3$
<b>G</b> band 1				9	$15 \pm 1.5$
<b>G</b> band 2				10	$15 \pm 1.5$
<b>G</b> band 3				11	$15 \pm 1.5$
<b>G</b> band 4				12.48	$15 \pm 1.5$
<b>G</b> band 5	2.7	1.2	436	24	$40 \pm 4$
				Ratio of 10 to 24 mers	$85 \pm 8.5$
				9 mers	$15 \pm 1.5$

**2.6.3 Purification of another representative batch of Acr by gel filtration:** The **H** and **G** samples were analysed by Native-PAGE and 15% reducing SDS-PAGE.

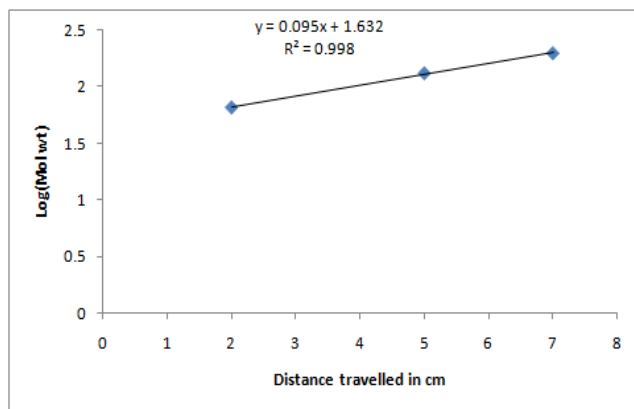


**Figure 2.6.3: 8-16% Tris-Glycine gel analysis of different batch Native-PAGE analysis:** A Native-PAGE analysis was carried out for **H** and **G** samples using 8-16% gradient Tris-Glycine gel and BSA as a standard. The oligomer size was estimated using a plot of Log Molecular Weight versus distance migrated.

There was a mixture of oligomers observed in both **H** and **G** samples with the **H** samples having a higher proportion of lower molecular weight 5 to 8 mers. The proportion of oligomers was checked by plotting Log Molecular Weight versus distance in reference to BSA standards (Figure 2.6.3; 2.6.3.2 and 2.6.4; Table 2.6.2).



**Figure 2.6.3.2: 15% SDS-PAGE analysis of different batch of gel filtration analysis:** Lane 1: **G** samples B2; Lane 2: B1; Lane 3: C1; Lane 4: C2; Lane 5: C3; Lane 6: C4; Lane 8: Marker (Puragene) (200, 150, 100, 85, 70, 60, 50, 40, 30, 20, 10 kDa); Lane 9, 10: Load of gel filtration run 10 and 30  $\mu$ l; Lane 11, 12: **H** sample (concentrated) 10 and 30  $\mu$ l.



**Figure 2.6.4:** Plot of Log Molecular weight versus distance migrated. The distance travelled by the three BSA monomers was plotted against the log molecular weight.

**Table 2.6.2: Oligomer ratio calculation of batch 2:** Oligomer sizes were calculated using BSA monomer, dimer and trimer as a reference (66, 132 and 198 kDa) and the Log Molecular Weight plotted versus the distance travelled in centimetre. From this calculation, molecular weight was extrapolated for the different bands seen in **H** (band 1 to 8) and **G** samples (band 1 to 5).

BSA (Mol. wt.)	Log Mol. wt.	Distance (cm)	Size (kDa) Monomer Acr 18; BSA 66	Oligomer size	% fraction of total
66	1.82	2			
132	2.12	5			
198	2.29	7		18	
<b>H band 1</b>		3	82	5	14
<b>H band 2</b>		5	129.93	7	9
<b>H band 3</b>		6	158.49	9	20
<b>H band 4</b>		6.5	177.82	10	15
<b>H band 5</b>		7	198.15	11	12
<b>H band 6</b>		7.5	221.3	12	8
<b>H band 7</b>		8	246.6	14	11
<b>H band 8</b>		8.5	275.4	15	13
				Ratio of 11 to 15 mers	<b>43</b>
<b>G band 1</b>		6.5	177.82	9	15
<b>G band 2</b>		7	198.15	10	15
<b>G band 3</b>		7.5	221.3	11	15
<b>G band 4</b>		8	246.6	12	15
<b>G band 5</b>		8.5	275.4	16	40
				Ratio of 11 to 15 mers	<b>75</b>

### DISCUSSION

**Sod A and Sod C:** Sod A protein minus His-tag (GTG) did not express at 37°C. Very low level of expression could be seen at 25°C. However, the amount of protein expression seen was not adequate to carry out further studies. A possible reason for this could be the GTG start codon in the gene.

The new construct of *sod A* with C-terminal His-tag did not show any expression though it was re-cloned with ATG in the start codon. This indicates a poor codon bias as the possible reason. Another reason could be the point mutations that were found in the *sod A* clone, though this needs to be verified by further investigation as it could also be due to the toxicity of the gene as we saw similar expression profiles in uninduced and induced samples. For a similar reason, no expression was observed at 37°C in both BL21DE3 and BL21DE3 star cells in ATG start codon constructs. A similar reason could be the cause of no expression in the *sod C* construct. Most studies report difficulty of cloning Mycobacterial genes in *E. coli* due high G-C content (Lakey et al. 2000; Harth and Horwitz 1999).

**Acr:** The C-terminal His-tag Acr protein showed very low level of expression. This could be due to the intrinsic problem of use of pET21a vectors which has poor stability of the ampicillin resistance gene product leading to loss of plasmid. The Acr protein appeared in the insoluble form and bound to and eluted from Nickel-NTA under denaturing conditions, indirectly confirming the identity of the protein. However, attempts to solubilize the protein at different temperatures produced low amounts of soluble protein at 25°C and 0.05 mM IPTG. This protein bound and eluted from Nickel-NTA column. However, there was no reproducibility of batch to batch of C-terminal His-tag Acr; hence expression was tried with the N-terminal His-tag Acr.

**N-terminal His-tag Acr:** The N-terminal His-tag Acr showed high level of expression in the range of 50 mg/L. Several previous reports indicate the cloning and expression of Acr (Valdez



et al. 2002, Preneta et al. 2004, Kennaway et al. 2005, Dhepakson et al. 2008). The expression levels obtained compare favourably with earlier, different constructs, reported in the previous studies as summarized in **Table 2.7**. The protein appeared in a soluble form and was purified using both Nickel-NTA and Sephacryl S-200 gel filtration column. The confirmation of the Acr protein sequence enabled further purification studies. The protein appeared as a mixture of oligomers in the gel filtration run, which was repeated eight times to confirm the same. Kennaway et al. 2005, Panda et al. 2017, have reported the occurrence of Acr as a dodecamer by gel filtration. The Native-PAGE analysis indicated a mixture of oligomers with the **H** samples having higher proportion of 9 mers and the **G** samples having higher proportion of 12 mers to 24 mers. Native-PAGE of a different batch confirmed the difference in oligomeric ratios with the **H** samples having lower molecular weight of 5 to 8 mers which were not found in the **G** samples. Earlier reports have shown a mixture of 7 to 9 mers which have a dynamic process of association and dissociation as analyzed by Urea (Feng et al. 2002).

Subsequently, all activity assay analysis was carried out with two types of samples, **H** and **G**. This was to investigate differences in activity due to difference in oligomeric ratios.

**Table 2.7: Expression levels of different *acr* Constructs.**

Construct details	Expression Level reported	Reference*
pET28a NdeI – XhoI: Full length plus N-terminal His-tag Forward primer – additional 2.1 kDa sequence N-terminal.	50 mg per L	Gautam Krishnan & Utpal Roy 2019
pET21a (BamHI-XhoI) Full length plus His-Tag C-terminal Forward primer with Shine Dalgarno sequence upstream of start, no additional amino acids	50 mg/L	Dhepakson et al. 2008
TOPO expressed in pET28a with 3 additional N-Terminal residues – Ser, Met, Ala minus His-tag	40 mg/L	Kennaway et al. 2005
The resulting plasmid pET-hsp16.3 encoded a fusion protein consisting of the entire Hsp 16.3 protein with 11 extra amino acids, including 6 histidine residues at its N-terminal	10 mg/L	Preneta et al. 2004
pET20b: no His-tag Full length-434 bp 16.3 kDa protein	10-30 mg/L	Valdez et al. 2002
pET9d Full length protein, no additional amino acids	15 mg per L	Chang et al. 1996
*As per list of references		

### CONCLUSIONS

*sod A* gene was induced at very low level at 25°C with overnight induction. The amount was too low for further studies. *sod C* gene expression was not obtained. *sod A* gene with ATG in the start codon did not express the protein.

The C-terminal *acr*-pET21a construct produced low levels of active soluble Acr. Expression of Acr was obtained in the cell pellet at 37°C, this could be indirectly confirmed as, the expressed protein could bind and elute under denaturing conditions to His-tag column. Identity of soluble Acr could not be verified, though increased chaperone activity upon boiling pointed it to be the Acr molecule. The present study indicates that the N-terminal pET28a-*acr* gene construct over-expressed the Acr protein in soluble form and the expression levels obtained may be compared favourably with different constructs reported in the previous studies as summarized in **Table 2.7**. The Acr protein was found after the gel filtration as a mixture of oligomers; with the **G** samples having a higher ratio of higher molecular weight oligomers.

# CHAPTER 3

## **CHAPTER 3**

### **Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)**

#### **Objective 2**

### INTRODUCTION

Acr activity has been measured using thermal aggregation assays with citrate synthase, alcohol dehydrogenase, catalase and DTT induced aggregation of insulin B chain. The molar concentration of Acr used, varies from 1 to 20-fold molar excess for citrate synthase (Chang et al. 1996, Valdez et al. 2002) and the mole ratio of Acr to insulin as a substrate, varies from 0.088 to 0.6 (Yang et al. 1999, Mao et al. 2001, Gu et al. 2002, Panda et al. 2017). This could be due to the different molecular weights of the substrate; citrate synthase being 85 kDa and insulin being 6 kDa as compared to the Acr size of 14 kDa. The molecular interactions of Acr and citrate synthase have been more extensively studied (Yang et al. 1999). Our aim was to test chaperone activity and compare it to existing literature and also characterize the activity mechanism using insulin as a model substrate.

### MATERIALS AND METHODS

#### 3.1 Activity Assays of C-terminal His-tag Acr: Thermal Aggregation Assay

**3.1.1:** An equimolar amount of Acr, purified from 1.5 L of culture was used for the assay. Activity of the Acr was checked using the thermal aggregation assay of Citrate synthase (CS) at a concentration of 0.613 nM at 45°C. Equimolar amounts, 2-fold and 4-fold excess concentration of soluble Acr over citrate synthase was used and absorbance was measured at 320 nm over 30 mins. by using a blank of buffer 100 mM Tris Cl; 100 mM NaCl pH 7.5. Dilutions of citrate synthase were made 1:10 of stock of 9 mg/ml porcine citrate synthase (Sigma Chem.); 15  $\mu$ l was used in the reaction volume of 250  $\mu$ l.

**3.1.2:** Another batch of soluble Acr was used to perform a dose-dependent assay using 0.85 nM citrate synthase and approximately equimolar, 2-fold and 4-fold molar excess of Acr (1.14  $\mu$ M, 2.28  $\mu$ M and 3.42  $\mu$ M respectively). Samples prepared by boiling the Acr protein at 2.28 and 3.42  $\mu$ M for 5 and 10 mins. respectively, were also assayed.

### 3.2 Assays of N-terminal His-tag Acr

#### *In vitro* chaperone activity assays of soluble Acr

Assays were carried out using both types of samples, **H** and **G**.

**3.2.1 Thermal aggregation assay using citrate synthase as substrate:** An assay was carried out using CS (SIGMA) at a concentration of 64  $\mu$ g (2.51  $\mu$ M) and 2-fold molar excess of **H** Acr 25  $\mu$ g (5.2  $\mu$ M) in a reaction volume of 300  $\mu$ l. Thermal aggregation was measured at 45°C at 320 nm over 30 mins, comparing with controls of buffer (100 mM Tris pH 7.0, 100 mM NaCl and 5% Glycerol).

**3.2.2.1 Insulin B chain aggregation Assay at 60°C:** Insulin was used as a substrate for assays as has been reported in the literature. Besides this, the citrate synthase preparation available commercially is a suspension that causes more noise with the aggregation assay. We used Insulin (Insugen R Injection, Biocon, India). This was a recombinant preparation that was relatively cheaper than the existing commercial preparation from Sigma. Insulin injection was concentrated from a stock of 0.14 mg/ml to 1.0 mg/ml using 3 kDa concentrators from Amicon (Millipore). The assay buffer used was 20 mM Tris pH 7.0, 100 mM NaCl; 5% Glycerol. Controls of insulin in buffer without dithiothreitol (DTT), insulin with 25 mM DTT were used. In all these samples, all the components except DTT were added, the O. D. values auto zeroed, baseline adjustment done from 350 nm to 370 nm and then 25 mM DTT added to start the aggregation study.

Acr was used at five different concentrations of 12, 22, 37.5, 39 and 44  $\mu$ M of **H** in three different reaction volumes of 250, 300 and 400  $\mu$ l and two different insulin concentrations of 83 and 125  $\mu$ M.

Another assay was carried out at 60°C with both **H** and **G** samples at concentrations of 12 and 30  $\mu$ M respectively using insulin at a concentration of 118  $\mu$ M.

**3.2.2.2 Polynomial analysis of activity data at 60°C:** The activity data obtained was used to plot polynomial plots of % inhibition versus mole ratio of Acr to insulin and also % inhibition versus % of insulin B chain covered. Simulation plots were also drawn assuming scenarios of lower and higher ratio of Acr nonamer to match the data obtained with real scenarios as well as monomer, 9 mers and 17 mers.

**3.2.2.3 Heat treatment assays with both H and G samples (37°C):** Recombinant Acr, after gel filtration (**G**) was assayed at 37°C at 18  $\mu$ M to observe inhibition of insulin at 166  $\mu$ M. After pre-heat treatment at 60°C for 15 mins., the assay was repeated to check for chaperone activity. Another batch of His-tag purified protein (**H**) was assayed at 37°C at 11  $\mu$ M to monitor the inhibition of insulin at 118  $\mu$ M. After pre-heat treatment at 60°C for 15 mins, the assay was repeated to check for activity.

**3.2.2.4 Heat treatment studies (repeat studies):** The heat treatment assays were repeated as follows:

**H** sample at a concentration of 11  $\mu$ M at 37°C was assayed with insulin at 118  $\mu$ M and then 15 mins pre-heat treatment studies carried out at 37°C, 60°C and 70°C. For the **G** sample, heat treatment assays were carried out at 1  $\mu$ M and 15 mins pre-heat treatment studies carried out at 37°C, 60°C and 70°C.

**3.3 Insulin assay details (37°C and 60°C) with both H and G samples:** A fresh batch of Acr was used for this set of assays to generate new data. Insulin (Insugen R Injection, Biocon, India) at two different concentrations of 118 and 167  $\mu$ M was used as the substrate in a reaction volume of 0.25 ml. Aggregation was induced by the addition of 25 mM dithiothreitol (DTT) at a temperature of 37°C and 60°C and aggregation measured over 30 mins. by monitoring increase in absorbance at 360 nm. The chaperone activity of the Acr **H** samples was checked using the DTT induced aggregation assay of 1.0 mg/ml insulin in buffer (20 mM Tris pH 7.5, 100 mM NaCl and 5% Glycerol) at 60°C at 360 nm over 20 mins.



**3.3.1 Dose dependent studies (37°C and 60°C):** The inhibition of aggregation was measured after addition of Acr at varying concentrations ranging from 5 to 35  $\mu\text{M}$  for **H** sample and 2 to 12  $\mu\text{M}$  for **G** sample at 37°C and 5.5  $\mu\text{M}$ , 11  $\mu\text{M}$ , 18.75  $\mu\text{M}$  and 37.5  $\mu\text{M}$  for **H** samples at 60°C.

The % inhibition was measured as

$$= 100 - \left[ \frac{\text{Final O. D}_{360\text{nm}} \text{ of aggregated substrate with Acr}}{\text{Final O. D}_{360\text{nm}} \text{ of aggregated substrate without Acr}} \right] \times 100$$

Curves of % inhibition versus concentration of Acr and % inhibition versus mole ratio of Acr to insulin were plotted for both **H** and **G** samples. Polynomial curves of 2<sup>nd</sup> and 3<sup>rd</sup> order were plotted to observe trends in inhibition for both the samples and used for model development.

**3.3.2 Insulin aggregation Assays (37°C):** A fresh batch of Acr was used to measure chaperone activity which was carried out by preventing DTT induced aggregation of insulin B chain concentration 118  $\mu\text{M}$  at an O. D. of 320 nm over 30 mins. at 37°C. Aggregation was started by adding 25 mM DTT and the absorbance recorded to look for inhibition of aggregation on addition of Acr. Two types of samples; **H** and **G** were used for the enzyme assays. The activity of **H** samples was checked at concentrations of 6 and 24  $\mu\text{M}$  and **G** samples at 1, 3 and 4  $\mu\text{M}$ . The assay buffer used was 20 mM Tris pH 7.0, 100 mM NaCl, 5% glycerol.

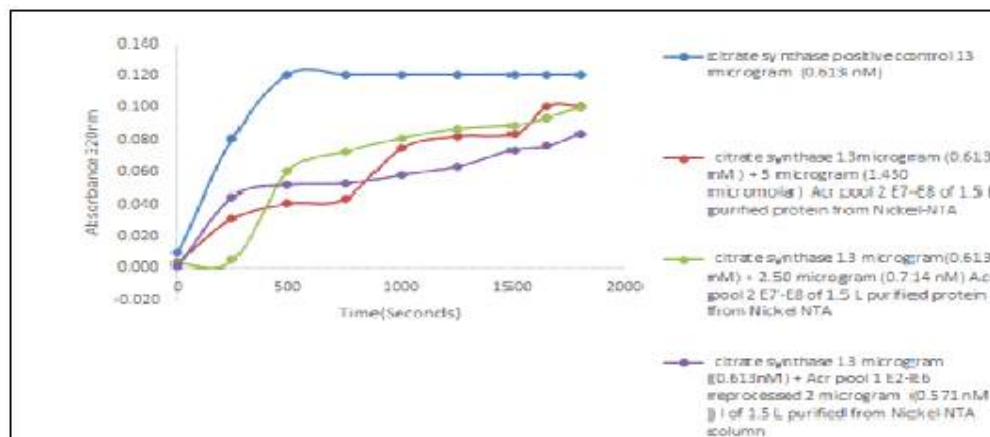
**3.3.2.2 Polynomial graph analysis:** The chaperone activity data curves were plotted in the form of polynomial graphs using MS Excel; the trend line displaying the equation and  $R^2$  values calculated using XY scatter diagrams under the chart function. Re-calculation tables were prepared using the equations generated from excel by feeding in the X values and obtaining the Y values and comparing with the expected value. Variation was noted as the deviation from the

original concentration of Acr used. Simulation plots were also drawn assuming two different ratios of Acr especially the 9 mers; one high and one low as well as monomer, 9 mers and 17 mers. This is to simulate a real scenario where there are different proportions of oligomers, in order to investigate how they would affect the coverage of the insulin B chain.

## RESULTS

### 3.1 Activity of soluble Acr by thermal aggregation assay

**3.1.1:** The activity of soluble Acr was confirmed by spectroscopic thermal aggregation assay performed at 45°C. The purified ACR from two pools showed suppression of thermal aggregation as compared to the control which showed no change over 30 mins. Pool 1 reprocessed showed enhanced inhibition of thermal aggregation (30%) as compared to the activity of pool 2 fractions (16 and 17% respectively) (**Figure 3.1.1**). The activity obtained in prevention of thermal aggregation varied from 16 to 30% for two different pools.

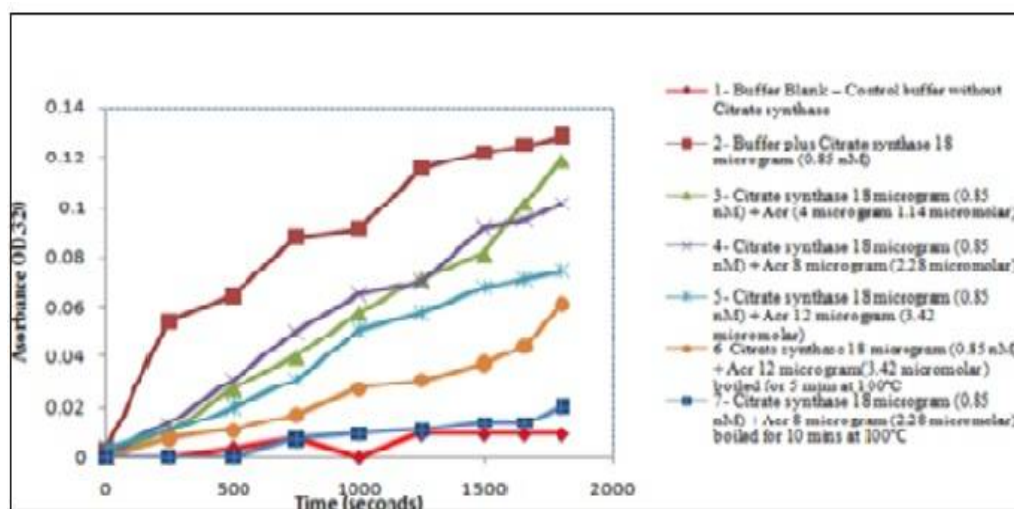


**Figure 3.1.1: Thermal aggregation assay of *acr*-pET21a using CS:** CS positive control 13  $\mu$ g (0.613 nM) thermal aggregation at 45°C; CS 13  $\mu$ g (0.613 nM ) + 5  $\mu$ g (1.430  $\mu$ M) Acr pool 2 E7-E8 of 1.5 L purified protein from Nickel-NTA; CS 13  $\mu$ g (0.613 nM) + 2.50  $\mu$ g (0.714 nM) Acr pool 2 E7-E8 of 1.5 L purified protein from Nickel-NTA; CS 13  $\mu$ g (0.613 nM) + 2 microgram (0.571 nM) reprocessed Acr pool 1 E7-E8 of 1.5 L purified protein from Nickel-NTA

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

Acr pool 2 E7-E8 of 1.5 L purified protein from Nickel-NTA; CS 13  $\mu\text{g}$  (0.613 nM) + Acr pool 1 E2-E6 reprocessed 2  $\mu\text{g}$  (0.571 nM) 1 of 1.5 L purified from Nickel-NTA column.

Another batch of Acr was tested for dose dependent inhibition of thermal aggregation of CS and it showed 7.8, 21 and 42.1% inhibition at three concentrations of Acr. Enhancement of inhibition of thermal aggregation after boiling was observed with inhibition increasing to 52% and 84% after boiling for 5 and 10 mins. respectively, as compared to 21 and 42% without boiling (**Figure 3.1.2**). This indirectly established the identity of the Acr as a heat shock protein. However, the identity of the soluble protein obtained after Nickel-NTA purification could not be verified as Acr. Henceforth, all studies were carried out with the new clone made in *acr*-pET28a



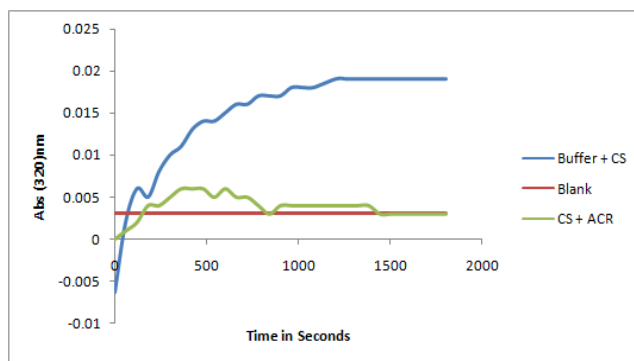
**Figure 3.1.2 Dose dependent thermal aggregation assay of Acr using CS.**

Buffer Blank: Control buffer without CS; Buffer plus CS 18  $\mu\text{g}$  (0.85 nM); CS 18  $\mu\text{g}$  (0.85 nM) plus Acr (4  $\mu\text{g}$  1.14  $\mu\text{M}$ ); CS 18  $\mu\text{g}$  (0.85 nM) plus Acr 8  $\mu\text{g}$  (2.28  $\mu\text{M}$ ); CS 18  $\mu\text{g}$  (0.85 nM) plus Acr 12  $\mu\text{g}$  (3.42  $\mu\text{M}$ ); CS 18  $\mu\text{g}$  (0.85 nM) plus Acr 12  $\mu\text{g}$  (3.42  $\mu\text{M}$ ) boiled for 5 mins at 100°C; CS 18  $\mu\text{g}$  (0.85 nM) plus Acr 8  $\mu\text{g}$  (2.28  $\mu\text{M}$ ) boiled for 10 mins. at 100°C.

**3.2 Assays of N-terminal His Acr:** The N-terminal *acr*-pET28a construct expressed the Acr protein in soluble form. This enabled us to carry out several functional and characterization studies. In gel filtration, Acr appeared as an oligomer, comparing with BIO-RAD standards. The Native-PAGE gel analysis indicated a clear difference in the pattern of oligomers with 9 mers dominant in the **H** samples as compared to the **G** samples where the 24 mers band was predominant. Mixture of 9 to 24 mers in the **G** sample indicated an association of the oligomers after gel filtration that influences chaperone activity.

**3.2.1 Thermal aggregation assay citrate synthase *in vitro* chaperone activity assay of soluble Acr:** Recombinant Acr showed 85% inhibition of aggregation of CS activity (**Figure 3.2.1**) when 2 molar excess of Acr was used over citrate synthase. The level of inhibition obtained was higher than that reported previously (**Table 3.2.1**). A similar amount of BSA showed less activity as compared to Acr (data not shown). These results compared favourably to the constructs cited in the literature which has been summarized in Table 3.2.1. However, as the constructs are made in different laboratories, unless they are all tested together and the structural features of all them compared, further conclusions cannot be drawn about the same. Also, porcine citrate synthase availability is a limitation and all future assays were done with insulin B chain.

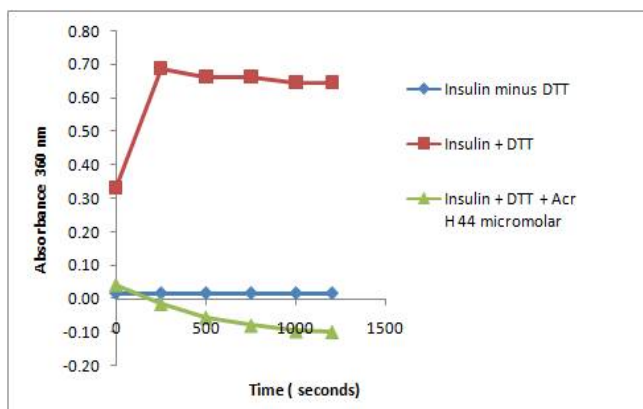
### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)



**Figure 3.2.1: Thermal aggregation assay of N-terminal His-tag Acr using CS.**

Blank: Inhibition of CS activity by Acr buffer alone, without citrate synthase; Buffer + CS: 64  $\mu\text{g}$  CS plus buffer at 45°C for 30 mins.; CS + Acr: 25  $\mu\text{g}$  of Acr was incubated with 64  $\mu\text{g}$  of citrate synthase for 30 mins. at 45°C and absorbance measured at 320 nm.

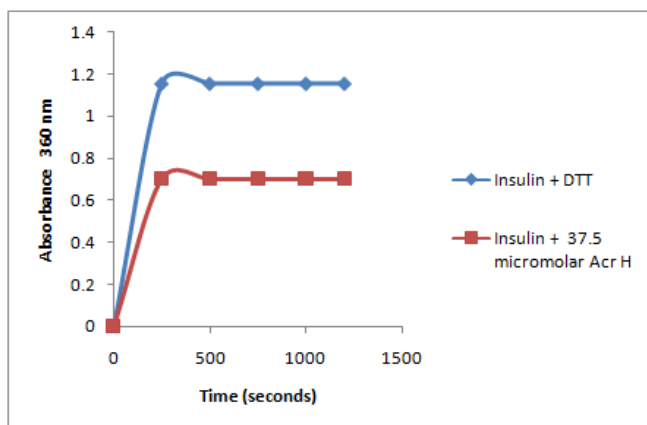
**3.2.2.1 Insulin aggregation assay of H samples of N-terminal His-tag Acr at 60°C:** The N-terminal His-tag Acr showed dose-dependent activity against insulin B chain at 60°C with 100% inhibition obtained at a mole ratio of 0.35 at a concentration of 44  $\mu\text{M}$  versus 125  $\mu\text{M}$  insulin (Figure 3.2.2.1.1).



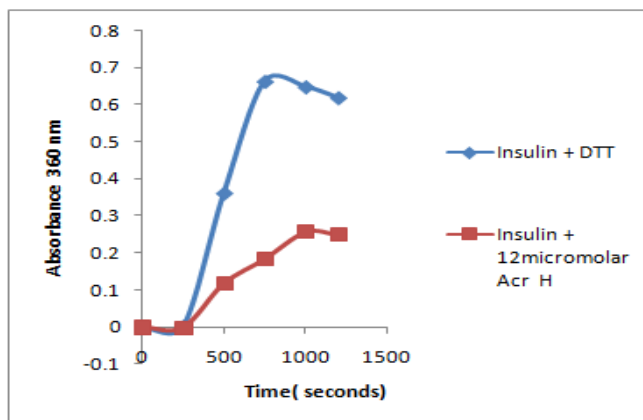
**Figure 3.2.2.1.1: Insulin aggregation assay at 44  $\mu\text{M}$  Acr:** Inhibition of insulin aggregation was carried out using 125  $\mu\text{M}$  insulin in an assay volume of 400  $\mu\text{l}$  and Acr concentration of 44  $\mu\text{M}$  along with controls of insulin minus DTT and insulin plus DTT.

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

At a concentration of 37  $\mu$ M, Acr showed 39% inhibition of insulin aggregation (**Figure 3.2.2.1.2**). At a concentration of 12  $\mu$ M, it gave an inhibition of 60% (**Figure 3.2.2.1.3**). The assays with 22  $\mu$ M Acr/83  $\mu$ M insulin and 39  $\mu$ M Acr/83  $\mu$ M insulin showed 56% and 100% inhibition respectively (data not shown). The chaperone activity of insulin at 60°C showed dose-dependence at 30, 37.5 and 44  $\mu$ M; yet, in some cases like 12  $\mu$ M, the inhibition was more than the 30  $\mu$ M sample. This could be due to the difference in assay volume and strongly suggested role of molecular level interactions which was eventually studied; revealing interesting insights into the structure-function relationship of Acr activity (Chapter 4). After initial standardisation, the assays were carried out in assay volume of 0.25 ml which was used for all molecular interaction calculations. In the dose-dependent study carried out at 60°C, 37.5  $\mu$ M of Acr was required to show complete 95% inhibition of chaperone activity.



**Figure 3.2.2.1.2: Insulin aggregation assay at 37.5  $\mu$ M Acr:** Inhibition of insulin aggregation was carried out using 125  $\mu$ M insulin in an assay volume of 250  $\mu$ l and Acr concentration of 37.5  $\mu$ M along with controls of insulin minus DTT and insulin plus DTT.



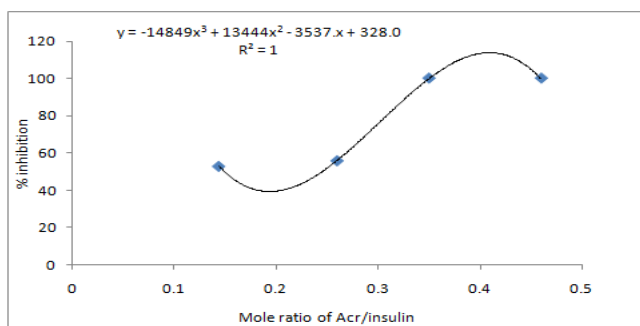
**Figure 3.2.2.1.3: Insulin aggregation assay at 12  $\mu$ M Acr:** Inhibition of insulin aggregation was carried out using 83  $\mu$ M insulin in an assay volume of 250  $\mu$ l and Acr concentration of 12  $\mu$ M along with controls of insulin minus DTT and insulin plus DTT. The Acr at four different concentrations, assayed with insulin at 60°C was converted into mole ratio of Acr v/s insulin. The Acr molecules were calculated by entering the amount ( $\mu$ g) in the website [http://molbiol.edu.ru/eng/scripts/01\\_04.html](http://molbiol.edu.ru/eng/scripts/01_04.html). The amount of Acr molecules used to inhibit insulin ranged from  $0.6 \times 10^{16}$  to  $1.5 \times 10^{16}$ , while the insulin B chain molecules ranged from  $1.8 \times 10^{15}$  to  $1.05 \times 10^{16}$ , wherein; when more than 50% of the insulin B chain was covered, complete inhibition was achieved (**Table 3.2.2.1**).

**Table 3.2.2.1:** Molecular level interaction of Acr/insulin at 60°C.

Mole ratio of Acr/insulin	% inhibition	% of insulin covered	Conc. of Acr (mg/ml)	Molecules of Acr	Molecules of 3 kDa insulin
0.144	53	14	0.2	$1.8 \times 10^{15}$	$1.26 \times 10^{16}$
0.26	56	47	0.36	$4 \times 10^{15}$	$0.85 \times 10^{16}$
0.35	100	70	0.8	$1.05 \times 10^{16}$	$1.5 \times 10^{16}$
0.46	95	94	0.7	$5.9 \times 10^{15}$	$0.63 \times 10^{16}$

**Table 3.2.2.1** was used to plot a polynomial graph as below (**Figure 3.2.2.2.1**). A 3<sup>rd</sup> order best fit was obtained as shown below.

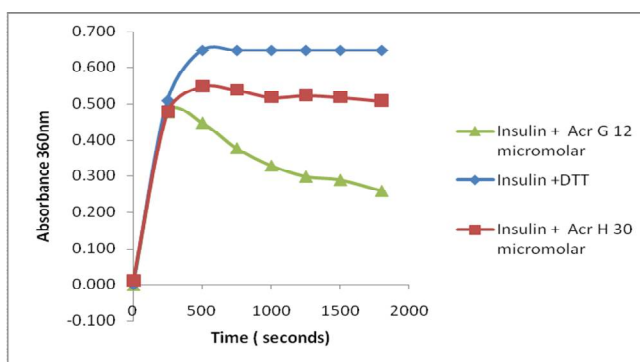
### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)



**Figure 3.2.2.2.1: Polynomial graph of % inhibition versus mole ratio of Acr/insulin at 60°C.**

The polynomial plots showed the % inhibition as a function of mole ratio of Acr/insulin as well as % of insulin B chain covered. The assumption in this calculation is that the Acr molecular weight is 18 kDa. However, if the number of molecules is calculated in terms of an oligomer, it would work out to be more. In this regard, two more figures were plotted in terms of monomer and 9 mers in terms of molecules of insulin covered and % inhibition (**Figure 3.2.2.2.1 to 3.2.2.2.7**).

**Insulin aggregation at 12  $\mu$ M G and 30  $\mu$ M H Acr (60°C):** The assay done with both **H** and **G** samples showed a difference in activity with the **G** sample showing 60% inhibition at 12  $\mu$ M as compared to 21% inhibition for the **H** sample at a concentration of 30  $\mu$ M (**Figure 3.2.2.1.4**).

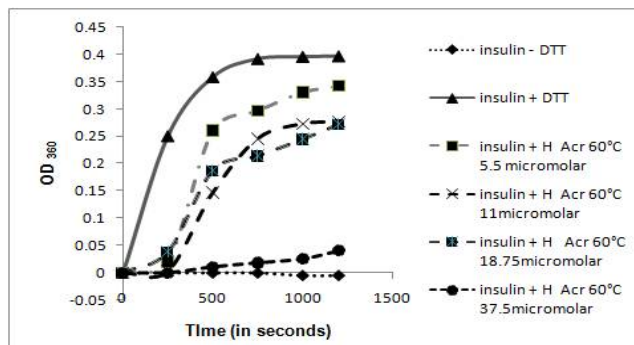


**Figure 3.2.2.1.4: Insulin aggregation at 12  $\mu$ M G and 30  $\mu$ M H Acr:** Insulin was assayed at a concentration of 167  $\mu$ M and using two concentrations of **H** (30  $\mu$ M) and **G** (12  $\mu$ M) respectively in an assay volume of 250  $\mu$ l.



### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

**Dose dependent Acr assay at 60°C:** The **H** sample showed dose-dependent inhibition of 25, 30, 60 and 95% at four different concentrations of 5.5, 11, 18.75 and 37.5  $\mu$ M respectively (Figure 3.2.2.1.5).



**Figure 3.2.2.1.5: Dose-dependent chaperone activity of H Acr at 60°C:** Insulin 1 mg/ml was used for the assay and aggregation measured at 360 nm over 20 mins. by adding 25 mM DTT. Acr was used at 4 different concentrations; 5.5, 11, 18.75 and 37.5  $\mu$ M.

Activity data of 12, 30 and 37.5  $\mu$ M of Acr non gel-filtered monomer **H** samples vs % of insulin B chain covered at 60°C was plotted for both monomer and nonamer. This helped to study differences in % of insulin B chain covered (Table 3.2.2.2 and Table 3.2.2.3).

**Table 3.2.2.2:** Plot of Acr non gel-filtered monomer molecules versus % insulin B chain covered

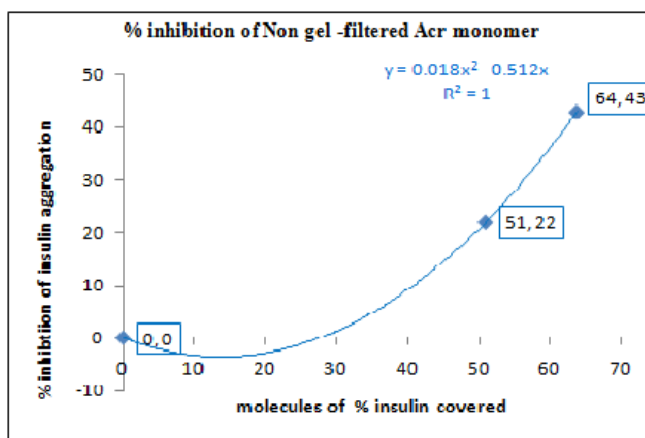
Acr non gel-filtered monomer	
% insulin covered (non gel-filtered)	% inhibition
0	0
13	22
15	40
30	60

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

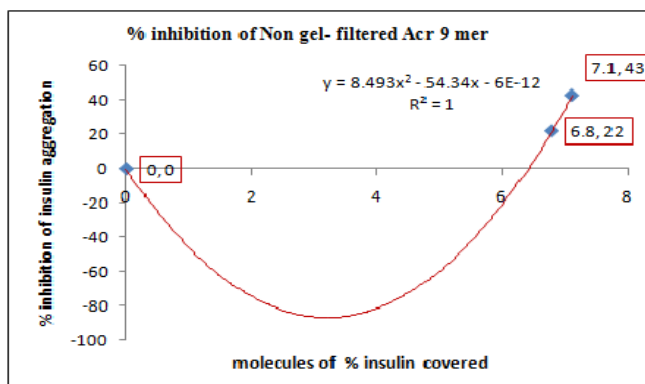
**Table 3.2.2.3:** Plot of Acr non gel-filtered nonamer molecules versus % insulin B chain covered

% insulin covered by non gel-filtered Acr 9 mer	% inhibition
0	0
6.8	22
7.1	43

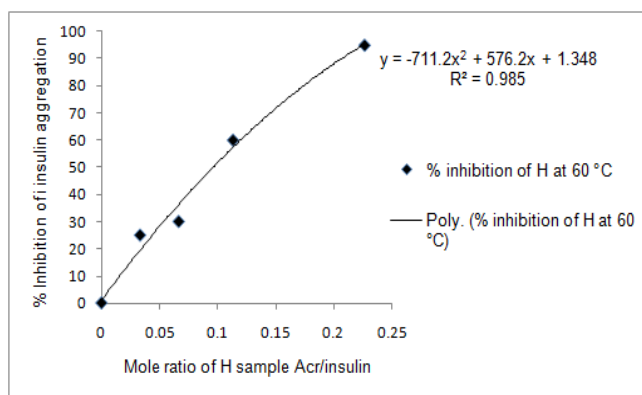
**Table 3.2.2.3** data was used to replot the mole ratio of Acr to insulin B chain in terms of % inhibition obtained and 3<sup>rd</sup> order polynomial graph plotted in MS Excel.



**Figure 3.2.2.2.2:** Polynomial graph of % inhibition of Acr monomer versus % of insulin B chain covered at 60°C: Table 3.2.2.3 data was used to replot the number of molecules that bind to insulin in terms of % of insulin B chain covered, assuming Acr to be an 18 kDa monomer and dividing the Avogadro number of molecules of Acr by the number of molecules of insulin B chain versus % inhibition. A 2<sup>nd</sup> order polynomial graph was plotted.



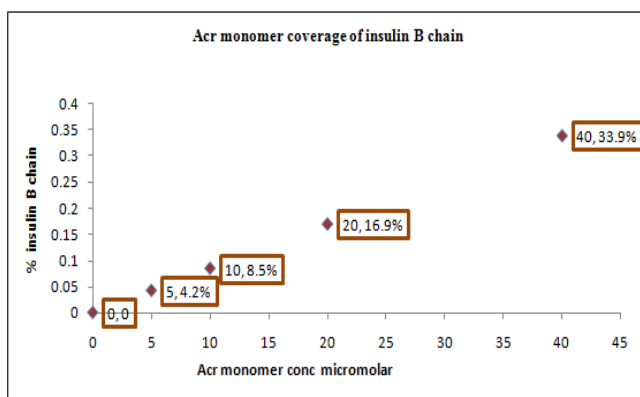
**Figure 3.2.2.2.3: Polynomial graph of % inhibition of Acr nonamer versus % of insulin B chain covered at 60°C.** Table 3.2.2.3 data was used to replot the number of molecules of Acr that bound to insulin B chain assuming Acr to be a nonamer 162 kDa and multiplying the number of molecules in **Figure 3.2.2.2** by 8.9 versus % inhibition.



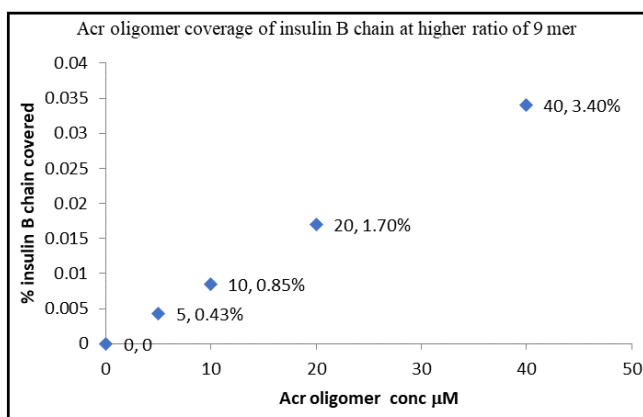
**Figure 3.2.2.2.4: Polynomial plot of mole ratio of H sample versus % inhibition of insulin** (from **Figure 3.2.2.1.5**). The % inhibition of insulin aggregation was plotted against the mole ratio of Acr to insulin at four different concentrations.

**Simulation plots of polynomials of Acr:** The simulation plot of Acr monomer coverage of insulin B chain, assuming theoretical binding at 4 molar concentrations (0.5, 10, 20 and 40  $\mu$ M) showed maximum coverage of 33.9% at 40  $\mu$ M (**Figure 3.2.2.2.5**). The simulation plots of higher ratio oligomer of 9 mers showed 3.4% coverage as compared to 2.42% coverage with

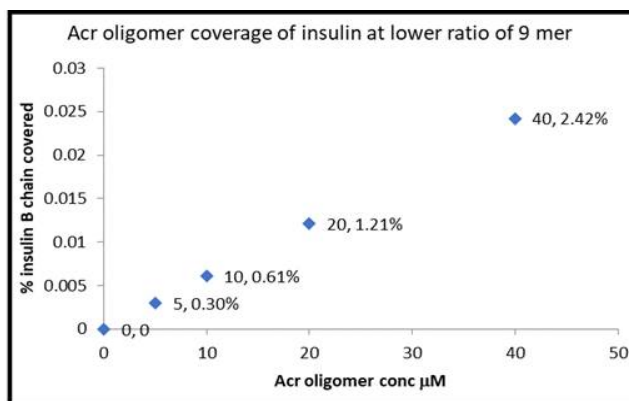
lower ratio of 9 mers indicating that higher the oligomers, less insulin is required to be covered (Figure 3.2.2.2.6 and 3.2.2.2.7).



**Figure 3.2.2.2.5: Simulation plot of monomer versus % insulin B chain covered:** A plot of simulated Acr assuming monomer versus insulin B chain covered at 118  $\mu$ M.

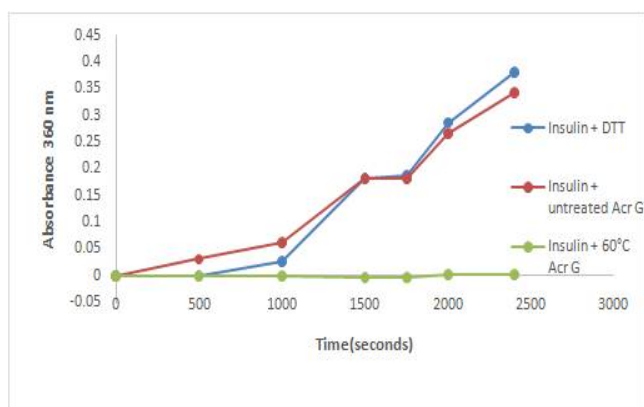


**Figure 3.2.2.2.6: Simulation plot of higher ratio of 9 mers versus % insulin B chain covered:** A plot of simulated Acr oligomer assuming 9 mers ratio of 60% and 10% of 10 to 12 mers versus % of insulin B chain covered at 118  $\mu$ M.



**Figure 3.2.2.2.7: Simulation plot of lower ratio of 9 mers versus % insulin B chain covered:** A plot of simulated Acr oligomer assuming 9 to 15 mers ratio of 20% each versus % of insulin B chain covered at 118  $\mu\text{M}$ .

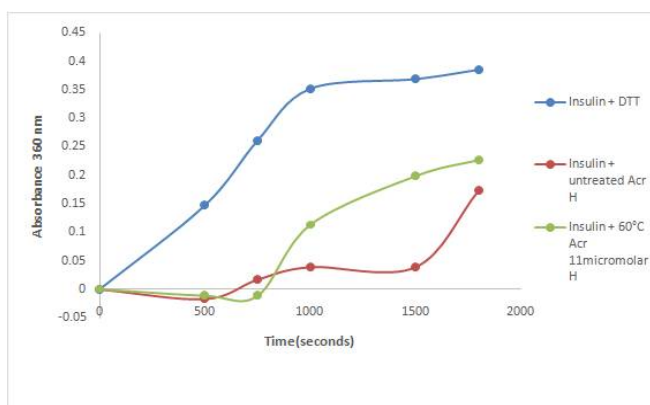
**3.2.2.3 Heat treatment assays (37°C):** The **G** sample showed complete inhibition of insulin B chain after pre-heat treatment at 60°C for 15 mins. (**Figure 3.2.2.3.1**). **H** sample failed to exhibit improvement in chaperone activity after pre-heat treatment, suggesting the need for secondary structure analysis to explain this difference (**Figure 3.2.2.3.2**).



**Figure 3.2.2.3.1:** Inhibition of insulin aggregation using Acr (18  $\mu\text{M}$ ) and insulin 0.5 mg/ml was carried out at 37°C. **G**-Acr with and without pre-heat treatment at 60°C was assayed.

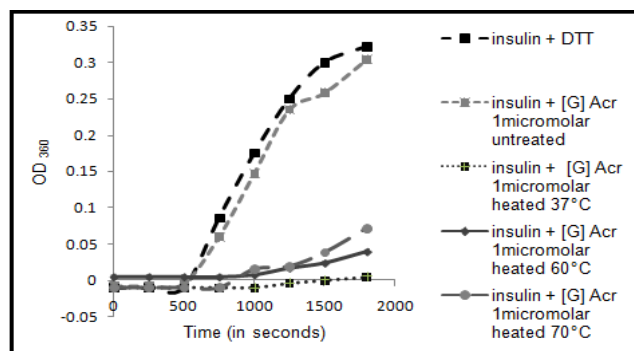
**Heat treatment analysis:** The **H** sample results indicate either an unstable secondary structure that was found to be dissociated upon heating, or the presence of impurities that interfered with

the response to heat treatments given. **G** sample results can be explained by improved binding at higher temperature due to changes in percentage of random coils and turns which possibly increased the exposure of hydrophobic surfaces of oligomeric Acr to the substrates. Both these samples were subjected to secondary structure analysis to find a correlation between structure and activity.



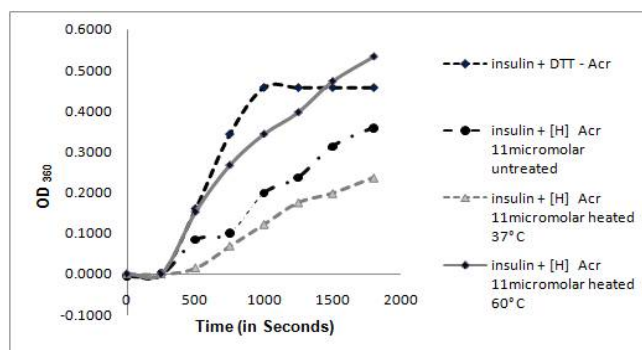
**Figure 3.2.2.3.2:** Inhibition of insulin aggregation using **H**-Acr (18  $\mu$ M) and insulin 0.5 mg/ml was carried out at 37°C. **H**-Acr with and without pre-heat treatment at 60°C was assayed.

**3.2.2.4: Heat treatment studies (repeat):** The **G** samples, after pre-heat treatment at 37°C, showed complete inhibition at 1  $\mu$ M at a mole ratio as low as 0.0084 (**Figure 3.2.2.4.1**). The **G** samples were assayed with and without pre-heat treatment at 37°C, 60°C and 70°C at a concentration of 1 $\mu$ M. The **G** samples, after pre-heat treatment at 60°C and 70°C showed 94% and 85% inhibition respectively. **G** samples showed improvement in activity in pre-heat treatment studies at 37°C, 60°C and 70°C, but in decreasing order.



**Figure 3.2.2.4.1:** Pre-heat treatment studies and chaperone activity of **G**; pool 2, run 2: Insulin 1 mg/ml was used to induce aggregation with 25 mM DTT and aggregation measured at 360 nm over 30 mins. The concentration of Acr used was 1  $\mu$ M and pre-heat treatment was carried out at 37°C, 60°C and 70°C for 15 mins.

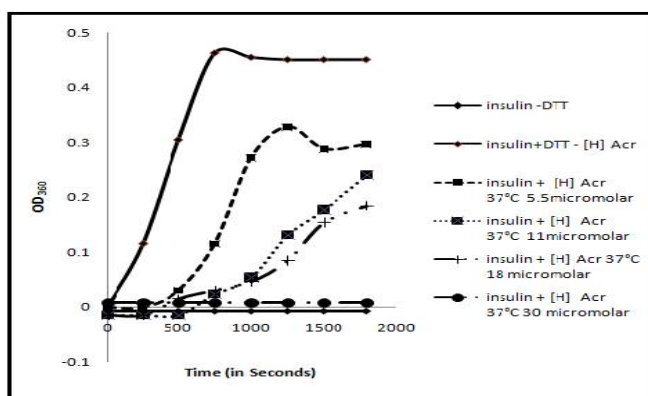
**3.2.2.4.2: Pre-heat treatment studies and chaperone activity of H samples (repeat):** Upon pre-heat treatment at 37°C, **H** samples at 11  $\mu$ M showed slight improvement in activity while, upon pre-heat treatment at 60°C, there was a decline in activity. At 70°C the samples showed similar result as at 60°C (data not shown). An interesting feature is that the activity decreased at both 60°C and 70°C. This was analyzed later by Far-UV CD studies.



**Figure 3.2.2.4.2:** Chaperone activity of **H** at 37°C; after pre-heat treatment at 37°C and 60°C: Insulin 1 mg/ml was used for the assay and aggregation measured at 360 nm over 30 mins. by adding 25 mM DTT. Acr was used at a concentration of 11  $\mu$ M, with no pre-heat treatment and with pre-heat treatment at 37°C and 60°C.

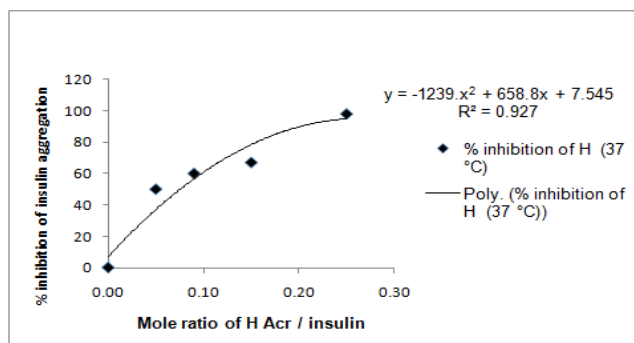
**3.3 Dose-dependent Insulin assay at 37°C and 60°C:** Activity of **H** versus **G** samples: **H** samples showed similar activity at 37°C and 60°C with marginally improved chaperone activity at 37°C (**Figure 3.3.1.1, 3.2.2.1.5**). Hence, rest of the assays were performed at 37°C.

**Insulin assay at 37°C:** Chaperone activity of Acr showed dose-dependence with 30  $\mu$ M giving 95% inhibition. The activity was similar to that at 60°C, particularly at lower mole ratios of Acr to insulin. The nature of inhibition of **H** sample at 37°C showed a change compared to that at 60°C, indicating that, at lower concentrations of Acr, the inhibition at 37°C is double at mole ratios of 0.03 to 0.10. However, at higher mole ratios of 0.115 to 0.25, the inhibition was similar. This was explained by secondary structure analysis in Chapter 4 of thesis. The **G** samples of three different pools of two runs showed better activity than **H** samples (**Figure 3.3.1.3, 3.3.1.5, 3.3.1.7**) with 12  $\mu$ M showing 95% inhibition as compared to 30  $\mu$ M of **H** samples.

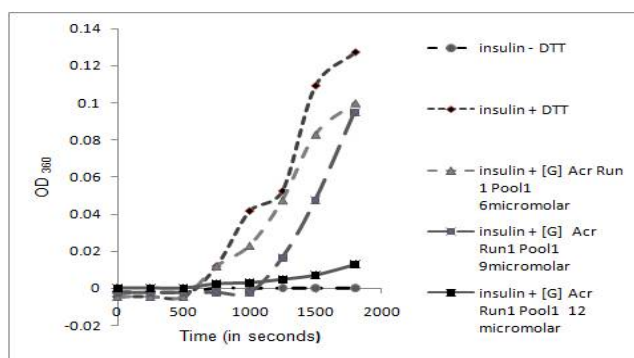


**Figure 3.3.1.1: Chaperone activity of H samples at 37°C:** Insulin 1 mg/ml (118  $\mu$ M) was used for the assay and aggregation was measured at 360 nm over 30 mins. by adding 25 mM DTT. Acr was used at four different concentrations of 5.5, 11, 18 and 30  $\mu$ M.

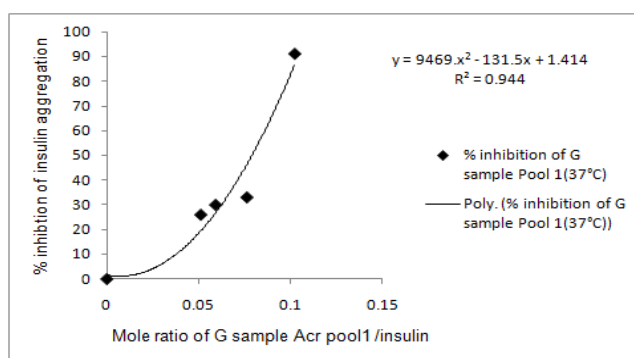




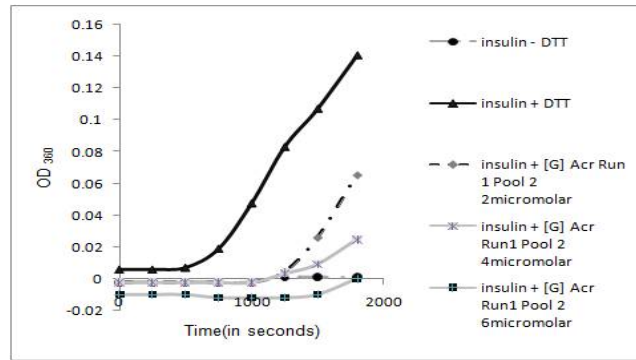
**Figure 3.3.1.2:** Polynomial plot of mole ratio of **H** sample versus % inhibition of insulin at 37°C (from **Figure 3.3.1.1**).



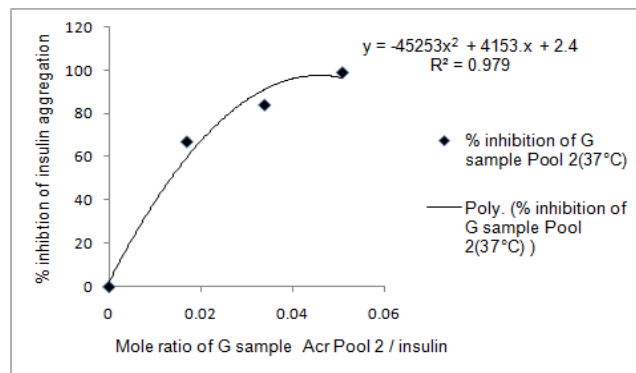
**Figure 3.3.1.3:** Chaperone activity of **G** samples pool 1, run 1: Insulin 1 mg/ml was used for the assay and aggregation measured at 360 nm over 30 mins. by adding 25 mM DTT. Acr was used at concentrations of 6, 9 and 12  $\mu$ M.



**Figure 3.3.1.4:** Polynomial plot of mole ratio of **G** sample run 1, pool 1 versus % inhibition of insulin at 37°C (from **Figure 3.3.1.3**).



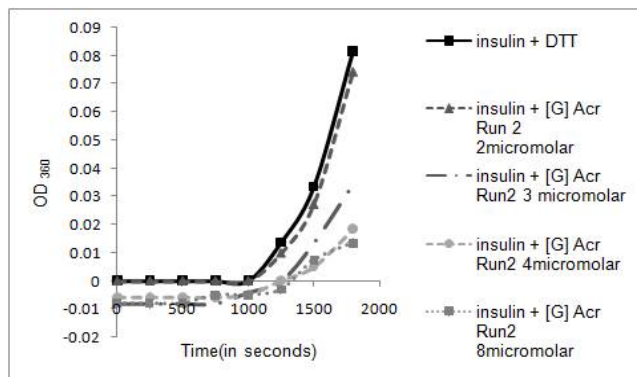
**Figure 3.3.1.5:** Chaperone activity of G sample run 1 pool 2: Insulin 1 mg/ml was used for the assay and aggregation measured at 360 nm over 30 mins. by adding 25 mM DTT. Acr was used at concentrations of 2, 4 and 6  $\mu$ M.



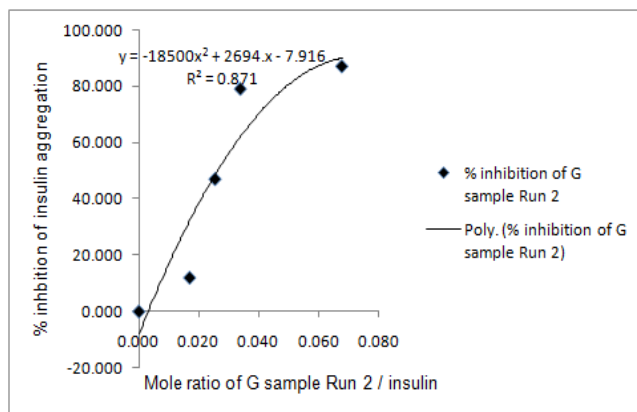
**Figure 3.3.1.6:** Polynomial plot of mole ratio of G sample run 1, pool 2 versus % inhibition of insulin at 37°C (from **Figure 3.3.1.5**).

**Polynomial curves of activity:** The activity was plotted in terms of mole ratio of Acr to insulin in the form of polynomial graphs for H samples at 60°C and 37°C (**Figure 3.2.2.1.5** and **Figure 3.3.1.1**; **Table 3.3.3** and **Table 3.3.4**) and for G samples (**Figure 3.3.1.3**, **3.3.1.5** and **3.3.1.7**; **Table 3.3.5** to **3.3.7**). Both **Figure 3.2.2.1.5** and **3.3.1.1** indicated a 2<sup>nd</sup> order polynomial curve with a higher coefficient for the  $x^2$  term in case of 37°C samples. The equations obtained were  $y = (-711x^2 + 576x + 1.348)$  and  $y = (-1239x^2 + 658.8x + 7.545)$ . A 3<sup>rd</sup> order polynomial trendline was plotted for **Figure 3.3.1.2** to check if % inhibition could be predicted from the equation,  $y = (23217x^3 - 9753x^2 + 1377x + 0.496)$ . Activity tables of H samples at 37°C and 60°C were

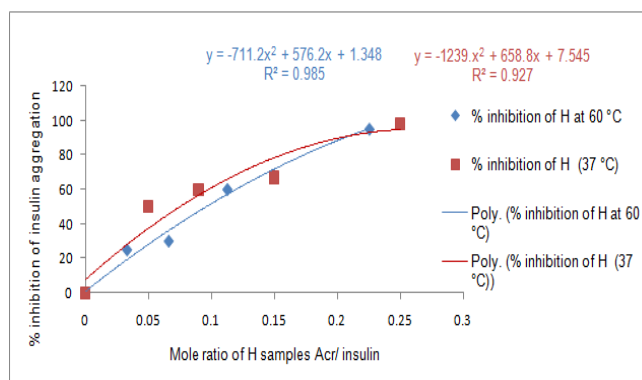
overlaid (Figure 3.3.1.9). Activity table of H sample at 37°C and two pools of G samples were overlaid (Figure 3.3.1.10). There was a clear shift to the left and changes in the nature of the polynomial best fits.



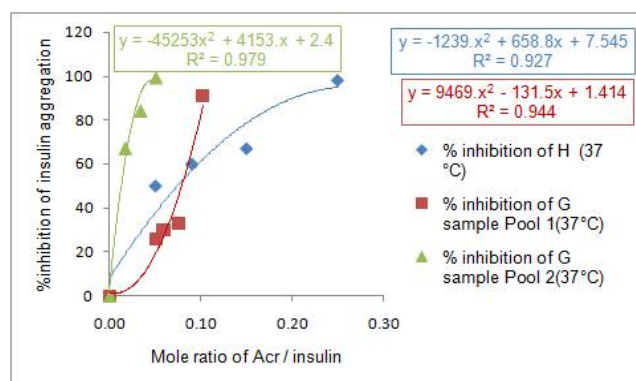
**Figure 3.3.1.7: Chaperone activity of G sample run 2:** Insulin 1 mg/ml was used for the assay and aggregation measured at 360 nm over 30 mins by adding 25 mM DTT. Acr was used at concentrations of 2, 3, 4 and 8  $\mu$ M.



**Figure 3.3.1.8: Polynomial plot of mole ratio of G sample run 2 versus % inhibition of insulin at 37°C (from Figure 3.3.1.7).**



**Figure 3.3.1.9:** Overlay of activity of **H** samples at 37°C and 60°C: overlay of **Figure 3.3.1.1** and **Figure 3.2.2.1.5** inhibition at 60°C: ◆ inhibition at 37°C: ■

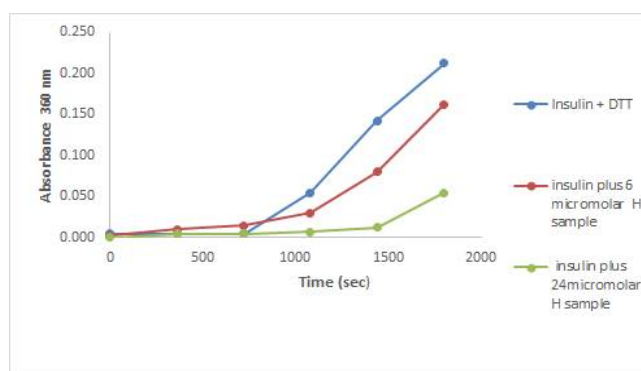


**Figure 3.3.1.10:** Overlay of activity of **H** samples versus activity of **G** samples at 37°C: inhibition at **H** sample (37°C): ◆ inhibition of **G** sample Pool 1 (37°C): ■ inhibition of **G** sample Pool 2 (37°C): ▲

**3.3.2.1** The polynomial graphs plotted, were used to re-calculate % inhibition with the equations obtained. This is to test the possibility of using polynomial graphs to check batches of Acr for chaperone activity. In some cases, for example, **H** samples (37°C and 60°C data) and **G** sample (run 2), the re-calculated values obtained were within 1 to 20% of the predicted values (**Table 3.3.8**). However, for the other two **G** sample pools, the polynomial curves could not be used to predict % inhibition especially in case of the 3<sup>rd</sup> order graphs while for the 2<sup>nd</sup> order polynomials, this could be done. This means that, this analysis had to be replotted using the ratio

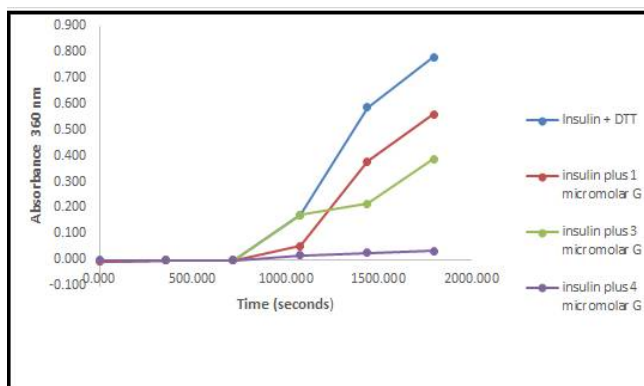
of oligomers and the number of molecules actually found, in order to strengthen the possibility of using polynomial curves to predict binding of Acr to insulin B chain. Replotting of the curves in terms of number of molecules and % of insulin chain covered showed that; the % inhibition could be re-estimated (Table 3.3.9 and Table 3.3.10, Figure 3.3.11 and Figure 3.3.12). These polynomial equations can be correlated to ratio of oligomers present. This indicates that, estimating the amount of chaperone activity can be done using polynomial graphs and a hypothesis built up based on data of **H** and **G** samples. As the **G** samples were proven to have a higher proportion of oligomers by Native-PAGE, the activity improved. This could also be a function of the concentration of the soluble Acr protein and the dynamic association and reassociation of different amounts of 9 mers to 24 mers. Later batches of Acr preparation showed 6 and 8 mers in **H** samples that were not found in **G** samples and showed better activity for **G** samples as compared to **H** (Figure 3.3.2.1, 3.3.2.2). A detailed investigation needs to be carried out in order to understand how the nature of oligomers affects the type of polynomial graphs obtained.

**3.3.2.2 Different batch of N-terminal His-tag Acr:** For a different batch of Acr at 37°C, chaperone activity of **H** samples showed nearly 100% inhibition at a concentration of 24  $\mu$ M while, **G** samples showed 95% inhibition at a concentration of 6  $\mu$ M (Figure 3.3.2.1, 3.3.2.2).



**Figure 3.3.2.1:** Assay of prevention of insulin B chain aggregation using **H** samples: insulin plus DTT; insulin plus 6  $\mu$ M **H** sample; insulin plus 24  $\mu$ M **G** sample.

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)



**Figure 3.3.2.2:** Assay of prevention of insulin B chain aggregation using **G** samples insulin plus DTT; insulin plus 1, 3 and 4  $\mu$ M Acr.

**Table 3.2.2.4:** Comparison of chaperone activity of different Acr constructs.

Clone Construct Details	Chaperone Assay and Activity	Acr concentration ( $\mu$ M)	Mole ratio of Acr: Substrate	% inhibition	Reference*
pET28a NdeI-XhoI: Full length plus N-terminus His-tag Forward primer – additional 2.1 kDa sequence N- terminal	1. Citrate synthase 2.5 $\mu$ M at 45°C	5	2	85	Gautam Krishnan & Utpal Roy 2019
pET28a NdeI-XhoI: Full length plus N-terminal His-tag Forward primer – additional 2.1 kDa sequence N- terminal	2. Insulin 166 $\mu$ M at 60°C	5.5	0.033	25	Gautam Krishnan & Utpal Roy 2019
		11	0.066	30	
		18.75	0.11	60	
		37.5	0.22	>95	
pET20b - No His-tag Full length – 434 bp -16.3 kDa protein	Citrate synthase 15 $\mu$ M at 45°C; 5 to 15-fold Mole excess Acr	75	5	20	Valdez et al. 2002
pET- Hsp16.3	Insulin 66 $\mu$ M at 30°C	12.5	0.189	60	Yang et al. 1999
pET - 9d	Citrate Synthase 1 $\mu$ M at 39.5°C; 0.5 to 8-fold Mole excess Acr	2	2	75	Chang et al. 1996
pET- Hsp16.3	1. Insulin 83 $\mu$ M at 60°C	50	0.6	>95	Gu et al. 2002
pET- Hsp16.3	2. Citrate Synthase 0.8 $\mu$ M at 40°C	6.25	8	20	Gu et al. 2002
pET28b - Hsp 16	Insulin 50 $\mu$ M at 25°C	4.38	0.088	13	Panda et al. 2017
pET28a with 3 additional N-Terminal residues Ser-Met-Ala minus His-tag	Not reported	Kennaway et al. 2005			
pET21a Full length plus His-tag C-terminal - Forward primer with Shine Dalgarno sequence upstream of start, no additional amino acids	Not reported				
* as per list of references					

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

**Table 3.3.3:** Chaperone activity of **H** sample versus Mole ratio at 60°C: Plot of mole ratio of **H** sample versus % inhibition of insulin; (from **Figure 3.2.2.1.5**). Polynomial plotted in MS Excel with XY scatter function in charts along with trendline with equation and R<sup>2</sup> value

<b>H samples</b>		<b>G samples</b>	
<b>Table 3.3.3</b>	<b>60°C</b>	<b>Table 3.3.5</b>	
<b>mole ratio</b>	<b>% inhibition</b>	<b>mole ratio</b>	<b>% inhibition</b>
0	0	0	0
0.033	25	0.051	26
0.066	30	0.059	30
0.113	60	0.076	33
0.226	95	0.102	91
<b>Table 3.3.4</b>	<b>37°C</b>	<b>Table 3.3.6</b>	
<b>mole ratio</b>	<b>% inhibition</b>	<b>mole ratio</b>	<b>% inhibition</b>
0	0	0	0
0.05	50	0.017	67
0.09	60	0.034	84
0.15	67	0.051	99
0.25	98	<b>Table 3.3.7</b>	
		<b>mole ratio</b>	<b>% inhibition</b>
		0	0
		0.017	12
		0.025	47
		0.034	79
		0.068	87

**Table 3.3.4:** Chaperone activity of **H** versus Mole ratio at 37°C: Plot of mole ratio of **H** sample versus % inhibition of insulin (from **Figure 3.3.1.1**). Polynomial plotted in MS Excel with XY scatter function in charts along with trendline with equation and R<sup>2</sup> value.

**Table 3.3.5:** Chaperone activity of **G** samples run 1, pool 1: Plot of **G** sample run 1, pool 1 versus % inhibition of insulin at 37°C (from **Figure 3.3.1.3**). Polynomial plotted in MS Excel with XY scatter function in charts along with trendline with equation and R<sup>2</sup> value.

**Table 3.3.6:** Chaperone activity of **G** samples run 1, pool 2: Plot of **G** sample run 1, pool 2 versus % inhibition of insulin at 37°C (from **Figure 3.3.1.5**). Polynomial plotted in MS Excel with XY scatter function in charts along with trendline with equation and R<sup>2</sup> value.

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

**Table 3.3.7:** Chaperone activity of **G** samples, run 2: Plot of **G** samples run 2 versus % inhibition of insulin (from **Figure 3.3.1.7**). Polynomial plotted in MS Excel with XY scatter function along with trendline with equation and  $R^2$  value.

**Polynomial re-calculation Plots for H and G samples:** For **H** samples at 60°C, mole ratios of 5.5, 11, 18 and 37  $\mu\text{M}$  as a fraction of 167  $\mu\text{M}$  insulin were entered. For **H** samples at 37°C, mole ratios of 5.5, 11, 18 and 30  $\mu\text{M}$  as a fraction of 118  $\mu\text{M}$  insulin were entered and the y values calculated. % variation was calculated by comparing with the predicted value (**Table 3.3.8** and **3.3.9**). Similar re-calculation was done for **G** samples where; 2, 4, and 6  $\mu\text{M}$  as a fraction of 118  $\mu\text{M}$  insulin x values were entered into the polynomial equation and the y values determined.

**Table 3.3.8: Re-calculation of polynomials:** The polynomial plots of **Tables 3.3.3, 3.3.4** and **3.3.6** were used to recalculate the % inhibition by entering the x values in the equation.

<b>H</b> samples at 60°C	Conc. ( $\mu\text{M}$ )	Mole ratio (x)	% inhibition (y)	Expected	Variation (%)
Equation	5.5	0.03	19.7	25	21
$y = (711x^2 + 576 + 1.348)$	11	0.07	36.4	30	21
	18.75	0.11	57.4	60	4
	37.5	0.23	95.2	95	0.23
<b>H</b> samples at 37°C	Conc. ( $\mu\text{M}$ )	Mole ratio (x)	% inhibition (y)	Expected	Variation (%)
Equation	5.5	0.047	45.8	50	8
$y = (23217x^3 - 9753x^2 + 1377x + 0.496)$	11	0.093	62.9	60	5
	18	0.153	66.0	67	1
	30	0.254	101.7	98	4
<b>G</b> samples run1, pool 2	Conc. ( $\mu\text{M}$ )	Mole ratio (x)	% inhibition (y)	Expected	Variation (%)
Equation	2	0.017	59.8	67	11
$y = (-45253x^2 + 4153x + 2.4)$	4	0.034	91.2	84	9
	6	0.051	96.6	99	2



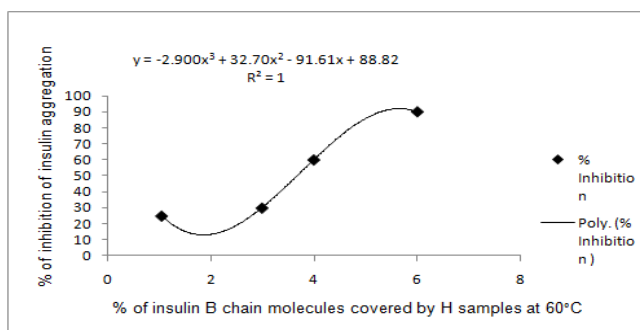
**Table 3.3.9: Re-calculation of values of inhibition for H samples at 60°C:** The values of 5.5, 11, 18 and 37  $\mu$ M as a fraction of 167  $\mu$ M were entered as x values in the new polynomial equation from **Figure 3.3.11** and the y values obtained have been compared with the expected values.

<b>% of insulin B chain covered</b>	<b>% inhibition</b>	<b>Re-calculated %</b>
1.06	25	24.99
3	30	29.97
4	60	59.97
6	90	94.97

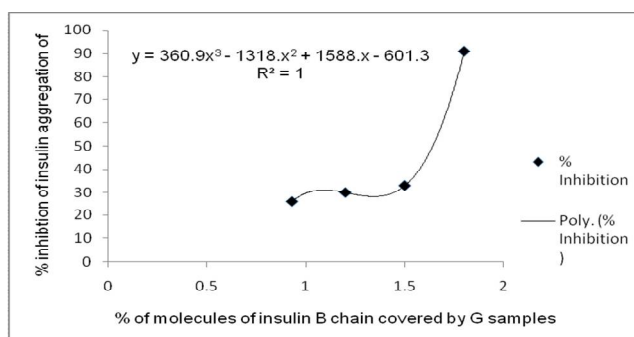
**Polynomial graph analysis:** The polynomial graphs plotted from **Figure 3.3.3.1 to 3.3.3.9** were used to re-calculate % inhibition with the equations obtained. In some cases, for example, **H** samples (37°C and 60°C data) and **G** sample (run 2), the recalculated values obtained were within 1 to 20% variation of the predicted values (**Table 3.3.8**). However, for the other two **G** sample pools, the polynomial curve could not be used to predict % inhibition. This means, that these analyses had to be replotted using the ratio of oligomers and the molecules actually found, in order to strengthen the possibility of using polynomial curves to predict binding of Acr to insulin B chain. Replotting of the curves in terms of number of molecules and % of insulin chain covered; showed that, the % inhibition could be re-estimated (**Tables 3.3.10, 3.3.12, Fig 3.3.11 and 3.3.12**). These polynomial equations can be well correlated to the ratio of oligomers present in Acr samples.

**Table 3.3.10: Re-calculation of values of inhibition for G samples:** The new polynomial equation obtained in **Figure 3.3.11** was used to enter the x values and the y values were obtained and compared with the expected values.

% of insulin B chain covered	% inhibition	Re-calculated %
0.93	26	25.9
1.2	30	30.0
1.5	33	33.2
1.8	91	91.5



**Figure 3.3.11:** Re-plotting of **H** samples in terms of molecules and % of insulin covered at 60°C: The **H** samples used were replotted by doing molecular calculations as per **Table 1** and the polynomial curve replotted in terms of % of molecules of insulin covered.



**Figure 3.3.12:** Re-plotting of **G** samples in terms of molecules and % of insulin covered at 37°C: **G** samples at four concentrations of Acr 2, 4, 6 and 8 μM were re-plotted in terms of molecules of insulin covered.

## DISCUSSION

**The C-terminal *acr*-pET21a construct:** To verify activity of expressed C-terminal His-tag Acr, a thermal aggregation assay was developed using citrate synthase and the inhibition of thermal aggregation (Valdez et al. 2002). The fraction of soluble protein showed activity in a dose-dependent manner and boiled samples enhanced chaperone activity, similar to previous findings of heat-enhanced chaperone activity at 40°C and 85°C (Chang et al. 1996, Yang et al. 1999). Though indirect evidence of the protein being Acr was established through the activity assay; particularly, the increase in activity after boiling, the yields of protein were too low for further study and characterization. Henceforth, the *acr*-pET28a construct was used for all further studies.

**The N-terminal *acr*-pET28a construct: CS activity:** The thermal aggregation activity of 85% inhibition at 2-fold molar excess of Acr:CS is 4-fold higher than the 5-fold molar excess reported for Acr:CS by (Valdez et al. 2002). The molecular chaperone activity exhibited by the clone 3 was higher than 20% inhibition, reported at an 8-fold molar excess of Acr over CS by (Gu et al. 2002). It is also equivalent to the report of 75% inhibition at 2-fold molar excess of Acr:CS by (Chang et al. 1996) though, the assay was done in a different buffer system and at a lower temperature of 40°C.

**Insulin aggregation activity:** The insulin activity at 37°C and 60°C obtained higher inhibition (> 95%) at 0.22 molar ratio of Acr to insulin as compared to the 95% inhibition reported for Acr to insulin at 0.66 molar ratio (Gu et al. 2002). This is better than the maximum inhibition of 60%, reported, and at the mole ratio of 0.189 of Acr to insulin; with pre-heating of Acr at 85°C (Yang et al. 1999), though the assay was performed at 30°C. The recombinant Acr protein expressed by the present construct shows comparatively higher activity at 60°C than that recorded by (Yang et al. 1999). This finding is also supported by (Mao et al. 2001) who reported

an enhanced chaperone activity 60°C onwards. The inhibition of insulin B chain aggregation is also better than that of 13% inhibition at a mole ratio of 0.0876 reported at 25°C (Panda et al. 2017).

Earlier reports have shown Acr to be a nonameric 158 kDa in the form of trimer of trimers (Chang et al. 1996), 210 kDa (Kennaway et al. 2005), 193 kDa (Panda et al. 2017, Fu et al. 2005). Previously, the process of dissociation and association has been extensively studied using urea with a wild type and a mutant Acr (Feng et al. 2002). They reported that it requires mild denaturant of 4M urea to find the intermediates before association into the oligomer. Reports have shown the importance of the nonamer as a pre-requisite to dissociate for showing chaperone activity; the generally accepted mechanism is a trimer of trimers aggregating to form the nonamer (Chang et al. 1996). Our results indicated varying amounts of 9 mer to 12 mers in the **H** samples and 9 mers to 24 mers in the **G** samples. We could infer that dissociation from a higher molecular weight oligomer improves chaperone activity. The 3-fold higher activity of **G** samples compared to **H** samples could be explained by the proportion of oligomers.

**Further characterization studies:** Far-UV characterization studies were performed for characterization of secondary structure and establishing correlation between structure and activity. This was to explain the difference in activity between **H** and **G** samples.

It has been reported in the literature about the occurrence of Acr as a nonamer or dodecamer (Chang et al. 1996, Kennaway et al. 2005). Our results support previous reports that monomers join together to form oligomers, as trimers of trimers that are able to bind effectively to inhibit insulin B chain aggregation. Simulation plots showed that the ratio of oligomers plays an important role in covering the insulin B chain. It also means that a higher ratio of oligomers would help better binding of insulin. Our data showed this with better activity seen in **G** samples versus **H** samples. This is the basic mechanism by which heat shock proteins can prevent unfolding of proteins. Polynomial graphs have been used to analyse activity of proteins

### Chapter 3: Activity Assays of *Mycobacterium* $\alpha$ -Crystallin (Acr)

involving oligomers (Yeow and Clayton 2007). Many oligomeric proteins function as high-molecular-weight aggregates and polymers and the mechanism of heat shock proteins is poorly understood (Rayees 2014). This approach of using polynomials can give insights into the molecular mechanism of oligomers binding to substrates.

## CONCLUSIONS

The C-terminal His-tag Acr showed dose-dependent chaperone activity of **H** sample in the citrate synthase assay at 45°C which enhanced upon boiling for five to ten minutes.

N-terminal His-tag Acr: **H** sample Acr showed dose-dependent activity seen with 95% inhibition at 37.5  $\mu$ M in the insulin assay at 60°C. The **G** sample showed better activity than the **H** sample.

Dose-dependent activity of **H** sample is seen in the insulin assay at 37°C with 95% inhibition at 30  $\mu$ M of **H** sample Acr and 95% inhibition at 12  $\mu$ M of **G** sample. **G** sample showed better activity than **H** sample.

Pre-heat treatment analysis: The **H** sample improved activity only after pre-heat treatment at 37°C. The **G** sample improved activity after pre-heat treatment at 37°C, 60°C and 70°C.

Polynomial curves of mole ratio of Acr to insulin vs % inhibition were used to predict binding of Acr to insulin B chain especially based on oligomer size using simulation plots of theoretical versus real data.

# CHAPTER 4

## **CHAPTER 4**

### **Structure and Activity Relationship of Acr**

#### **Objective 3**



### INTRODUCTION

Acr is a well characterized protein in terms of chaperone activity. Nearly all existing reports describe the purification step of gel filtration before performing enzyme assays. There are reports of dose-dependent activity against different substrates including Far-UV Circular Dichroism studies. It has been reported that the Far-UV CD spectrum of *M. tb* Acr, at 25°C, exhibited a minimum of 217 nm, indicating a possibility of high number of  $\beta$  sheets. The denaturation studies using 1M and 2M Guanidine Hydrochloride exhibited disorder in secondary structure (Yang et al. 1999). It has been reported that Fourier transform infrared (FTIR) spectroscopy at 25°C exhibited predominance in  $\beta$  sheets (67%). This is one of two reports that provide a complete break-up of all the different conformations, ranging from helices to turns, in Acr. The report indicated that structural changes are more from 25°C to 40°C than from 40°C to 60°C and the disorder increased from 8 to 10% due to the increase in random coils. It also predicted that the highly thermostable  $\beta$  sheets (67%) help in chaperone activity of Acr (Gu et al. 2002). However, there are not many more consolidated reports of the different types of secondary structure and their effect on chaperone activity. Besides, reports of correlation of pre-heat treatment to activity have been explained in terms of secondary structure studies (Chang et al. 1996, Yang et al. 1999, Mao et al. 2001, Gu et al. 2002). All reported studies so far are done with gel-filtered samples and no reports exist of difference in secondary structure prior to gel filtration. We attempted to analyse Far-UV CD data of **H** and **G** samples, with and without pre-heat treatment to observe if there is any correlation between structure and activity.

#### 4. MATERIALS AND METHODS:

**4.1.1 CD Analysis:** Far-UV CD analysis of both **H** and **G** samples was carried out at protein concentrations of 0.5 mg/ml and 0.6 mg/ml respectively, at The Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India, in the range of 190 to 260 nm using a Jasco Spectropolarimeter and 1mm cuvette with samples prepared in 10 mM Sodium Phosphate pH 7.0. The **H** samples were analyzed at 25°C and 37°C. The same were analyzed after pre-heat treatment for 15 mins, at 37°C, 60°C and 70°C. The **G** samples were analyzed at 37°C. The same were analyzed after pre-heat treatment for 15 mins, at 37°C, 60°C and 70°C.

**4.1.2 Comparison with theoretical prediction:** We compared our CD data with the predicted secondary structure analysis using PSIPRED.

**4.2 Molecular interaction calculation:** This was done using the website [http://molbiol.edu.ru/eng/scripts/01\\_04.html](http://molbiol.edu.ru/eng/scripts/01_04.html); where the following assumptions were made (i) reaction volume of 0.25 ml, (ii) insulin molecular weight of 3 kDa (assuming half of the molecule needs to be inhibited), (iii) oligomeric ratio. The secondary structure calculations were done from the % of each secondary structure as a function of total number of molecules obtained from the CD data. The mass balance was calculated by totalling the number of molecules of the four types of secondary structures from the CD plots and assuming the oligomeric ratio obtained from Native-PAGE analysis, divided by the total theoretically expected and this was expressed by the following formula:

$$\text{Number of molecules of Acr} = \frac{\text{amount in } \mu\text{g} \times \text{proportion of oligomers as per Native-PAGE}}{\text{Molecular weight}}$$

**4.3: Calculations of molecular level binding of Acr to insulin B chain in terms of secondary structure:** Graphs were plotted for **G** samples in terms of  $\alpha$  helices, random coils and turns.

## Chapter 4: Structure and Activity Relationship of Acr

Graphs of secondary structure analyses for **H** samples were plotted in order to explain activity and heat treatment data.

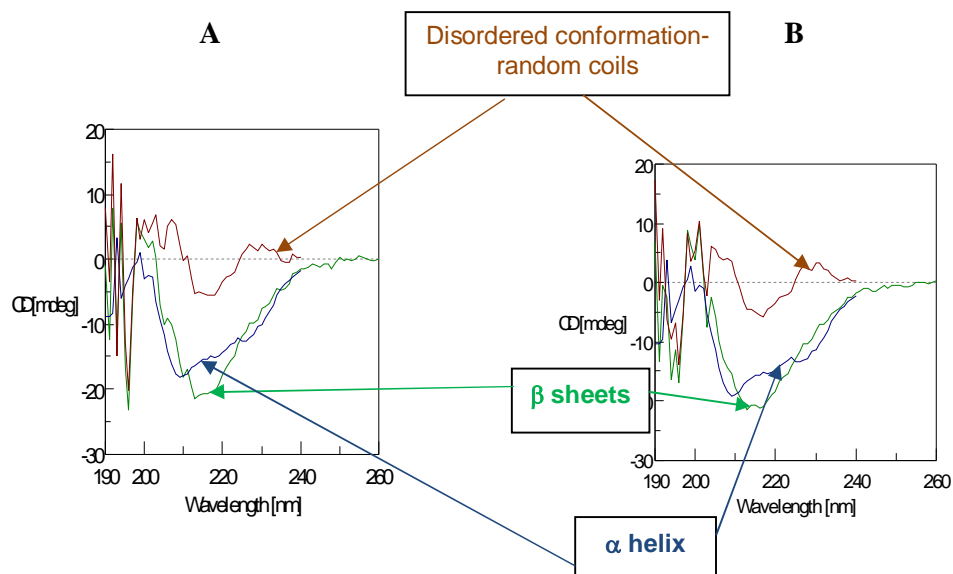
**4.4:** Calculation of  $T_m$  of *M. tb* Acr The His-tag **H** and His-tag plus gel purified Acr **G** was subject to a thermal shift assay using Sypro Orange as per the following protocol. The protein was added at a concentration of 2.5  $\mu$ M in a 50 mM Tris pH 8.0 reaction volume of 20  $\mu$ l and 20X Sypro Orange dye (Sigma Aldrich) was added and the  $T_m$  experiment carried out by ramping up the temperature from 25°C to 90°C at a ramp rate from 100% to 2%. The run was approximately done for 40 mins. in a 384 well plate for RT (Reverse Transcriptase) PCR machine (HT 7900 series from Applied Biosystems). Raw data was extracted into excel and the  $V_{50}$  is calculated using Boltzmann sigmoidal which provided the intercept on x axis giving the melting temperature ( $T_m$ ) of the protein.

### RESULTS

**4.1 Far-UV CD studies:** Far-UV CD analysis showed that the **H** and **G** samples showed different proportions of  $\alpha$  helices,  $\beta$  sheets, turns and random coils **Figure 4.1.1 (A to F), 4.1.2 (A to E); (Table 4.1.1.1, Table 4.1.1.2)**. For the **H** samples, the  $\beta$  sheets increased from 11% to 30 - 62%, after pre-heat treatment. The **H** samples without pre-heat treatment showed higher fraction of  $\alpha$  helices (22 - 26%) which decreased to 3.4 - 14% after pre-heat treatment. The fraction of turns was found to be 30 - 35% and decreased to 4.3 - 15% after pre-heat treatment. The fraction of random coils was found to be 35 - 36% and changed to 18.8 - 46% after pre-heat treatment.

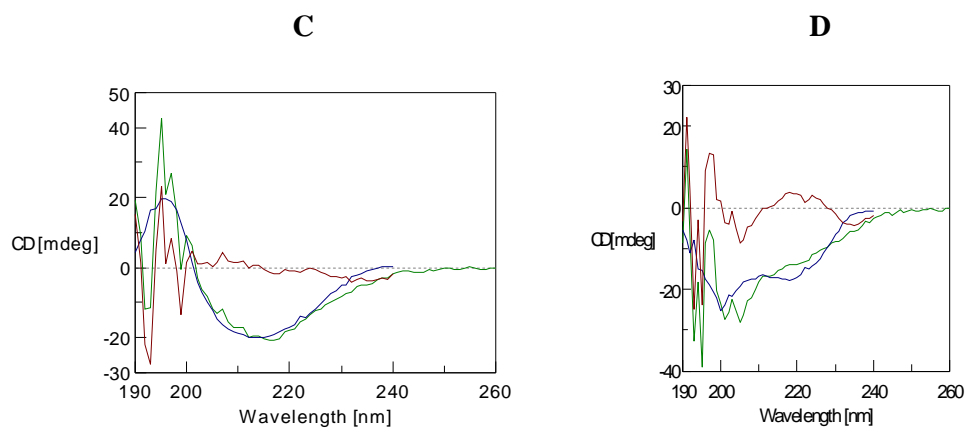
For **G** samples, the  $\beta$  sheets changed from 45% to 25 - 55% after pre-heat treatment. The fraction of  $\alpha$  helices was 9.6% and changed to 3.9 - 14% after pre-heat treatment. The fraction of turns was found to be 10.5% and changed to between 0% - 14% after pre-heat treatment. The fraction of random coils was found to be 34.4% and increased to 40 - 46% after pre-heat treatment.

**Far-UV CD data analysis:** This was generated under the same conditions at which the activity assays were carried out. The nature of inhibition of **H** sample at 37°C showed a change compared to that at 60°C, indicating that, at lower concentrations of Acr, the inhibition at 37°C is double at mole ratios of 0.03 to 0.10. However, at higher mole ratios of 0.115 to 0.25, the inhibition was similar. This could be a reflection of the higher amount of  $\beta$  sheets formed at 60°C (48.5%). However, the amount of  $\alpha$  helices could have played a role here, as they are reduced to 3.7% from 26% when heated to 60°C but this is compensated by higher % of  $\beta$  sheets. This is an important observation that needs further investigation and indicates a shift in conformation from  $\alpha$  helices to  $\beta$  sheets upon exposure to heat. Whether a similar phenomenon happens with Acr *in vivo* is a matter of investigation.



**Figure 4.1.1 (A):** Far-UV CD analysis of **H** sample at 25°C without pre-heat treatment

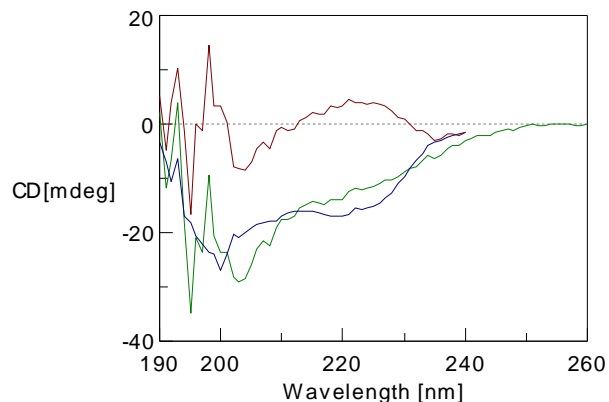
**Figure 4.1.1 (B):** Far-UV CD analysis of **H** sample at 37°C without pre-heat treatment



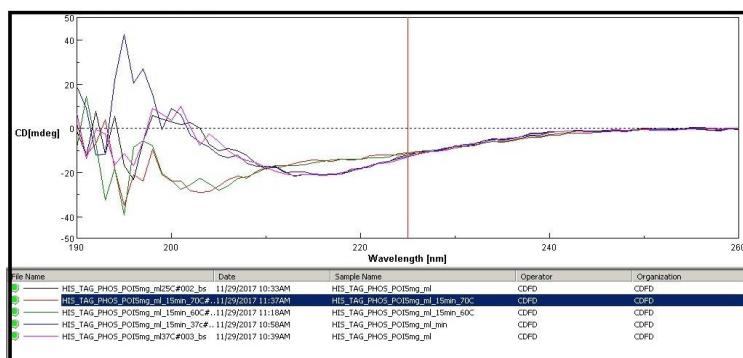
**Figure 4.1.1 (C):** Far-UV CD analysis of **H** sample after pre-heat treatment for 15 mins at 37°C.

**Figure 4.1.1 (D):** Far-UV CD analysis of **H** sample after pre-heat treatment for 15 mins at 60°C.

E



**Figure 4.1.1 (E):** Far-UV CD analysis of **H** sample after pre-heat treatment for 15 mins. at 70°C.

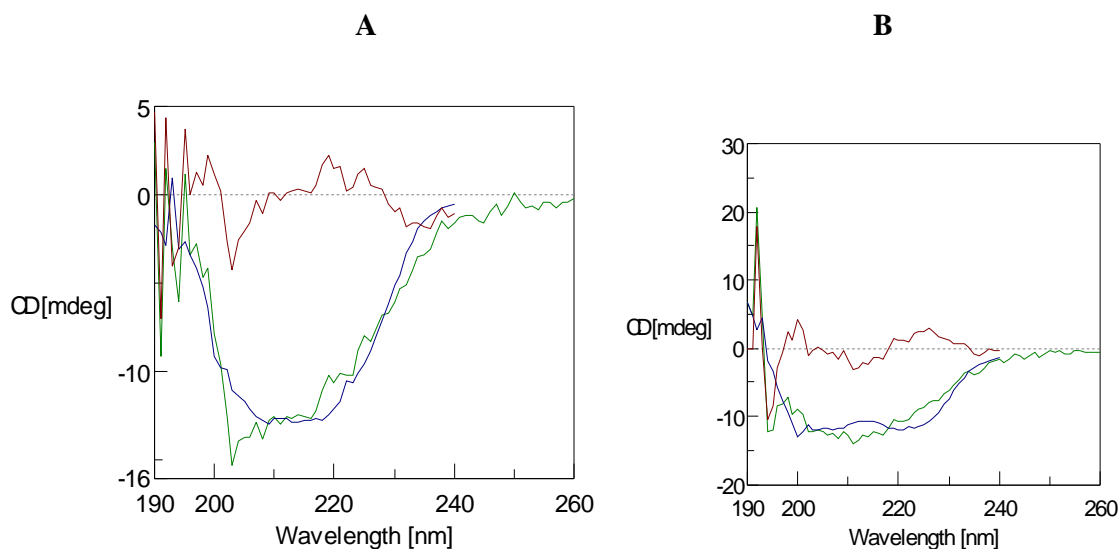


**Figure 4.1.1 (F):** Overlay of Far-UV CD analysis of **H** samples: Black: 25°C without pre-heat treatment; Purple: 37°C without heat treatment; Blue: 37°C with pre-heat treatment; Green: 60°C with pre-heat treatment; Red: 70°C with pre-heat treatment.

The **G** samples at 37°C, without pre-heat treatment, showed improved activity as compared to **H** sample at 37°C, without pre-heat treatment. The fraction of  $\beta$  sheets in the **G** samples was estimated to be 45% as compared to 2.5% in the **H** samples. **G** samples showed better activity even though the amount of  $\alpha$  helices (9.6%) was less than in the **H** samples (26%) (**Table 4.1.1.1**).

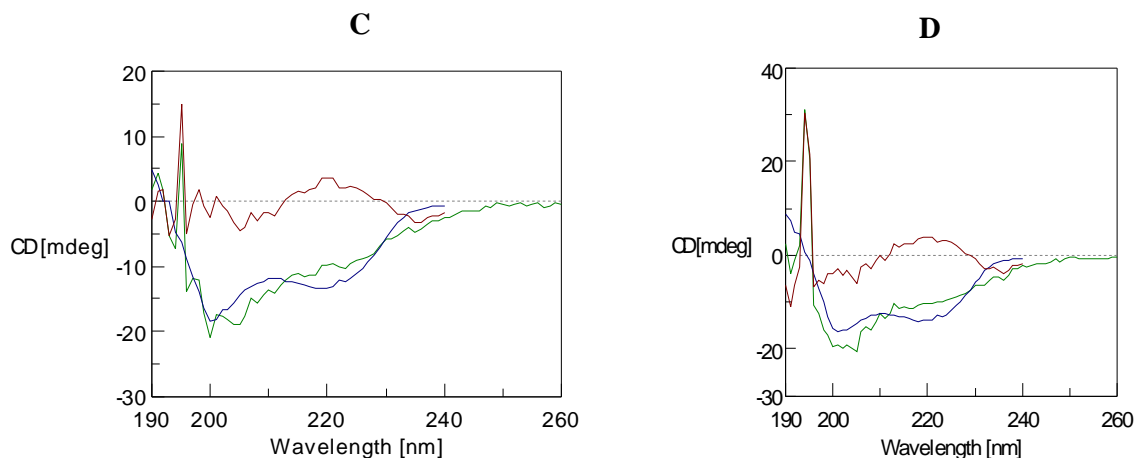
## Chapter 4: Structure and Activity Relationship of Acr

**G** sample results can be explained by improved binding at higher temperature due to changes in random coils and turns and increase in exposure to hydrophobic surfaces of oligomeric Acr. These results suggest an optimum range of 25% to 45% of  $\beta$  sheets as suitable for chaperone activity, which warrants further study of effect of pre-heating on availability of surfaces for binding to insulin.



**Figure 4.1.2 (A):** Far-UV CD analysis of **G** sample at 37°C without pre-heat treatment

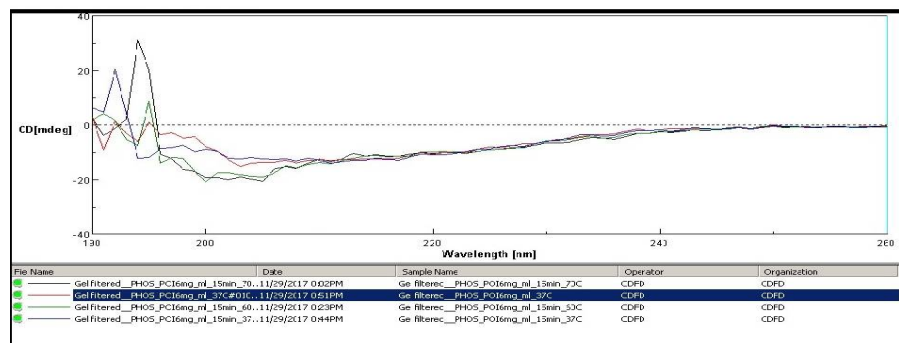
**Figure 4.1.2 (B):** Far-UV CD analysis of **G** sample at 37°C with pre-heat treatment



**Figure 4.1.2 (C):** Far-UV CD analysis of **G** sample with pre-heat treatment at 60°C

**Figure 4.1.2 (D):** Far-UV CD analysis of **G** sample with pre-heat treatment at 70°C

## Chapter 4: Structure and Activity Relationship of Acr



**Figure 4.1.2 (E):** Overlay of Far-UV CD Analysis of **G** samples: Red: 37°C without pre-heat treatment; Blue: 37°C with pre-heat treatment; Green: 60°C with pre-heat treatment; Brown: 70°C with pre-heat treatment.

**Table 4.1.1.1:** Consolidated CD profile showing proportion of molecules of secondary structure of **H** samples.

<b>CD of (H) samples at 25°C</b>	<b>% of total</b>	<b>CD of (H) samples at 60°C with pre-heat treatment</b>	<b>% of total</b>
$\alpha$ helices	22.7%	$\alpha$ helices	3.7%
$\beta$ sheets	11.5%	$\beta$ sheets	48.5%
Turns	30.3%	Turns	7.9%
Random coils	35.5%	Random coils	39.9%
<b>CD of (H) samples at 37°C</b>	<b>% of total</b>	<b>CD of (H) samples at 70°C with pre-heat treatment</b>	<b>% of total</b>
$\alpha$ helices	26.0%	$\alpha$ helices	8.4%
$\beta$ sheets	2.5%	$\beta$ sheets	30.0%
Turns	35.3%	Turns	15.4%
Random coils	36.2%	Random coils	46.1%
<b>CD of (H) samples at 37°C with pre-heat treatment</b>	<b>% of total</b>		
$\alpha$ helices	14.0%		
$\beta$ sheets	62.9%		
Turns	4.3%		
Random coils	18.8%		



## Chapter 4: Structure and Activity Relationship of Acr

**Table 4.1.1.2:** Consolidated CD profile of proportion of molecules of secondary structure of **G** samples

CD of G samples at 37°C without pre-heat treatment	% of total	CD of G samples at 70°C with pre-heat treatment	% of total
$\alpha$ helices	9.6%	$\alpha$ helices	5.4%
$\beta$ sheets	45.5%	$\beta$ sheets	54.6%
Turns	10.5%	Turns	0.0%
Random coils	34.4%	Random coils	40.0%
CD of G samples at 37°C with pre-heat treatment	% of total	CD of G samples at 60°C with pre-heat treatment	% of total
$\alpha$ helices	14.3%	$\alpha$ helices	3.9%
$\beta$ sheets	25.0%	$\beta$ sheets	51.1%
Turns	14.0%	Turns	2.9%
Random coils	46.6%	Random coils	42.1%

### 4.1.2 Comparison with theoretical prediction:

**Table 4.1.2:** Comparison of secondary structures of **H** and **G** samples with predicted sequence obtained from PSIPRED: The N-terminal sequence of Acr was used to determine the theoretical secondary structure followed by comparison with CD profile of **H** and **G** samples as summarised in **Table 4.1.1.2** and **Table 4.1.2.1** respectively.

Amino acids (162)	helix	$\beta$ sheets	coils and turns
theoretical (by PSIPRED)	3%	36%	61%
<b>G</b> samples	9.6%	45.5%	44.9%
<b>H</b> samples	26%	2.5%	71.5%

**4.2 Molecular interaction calculation:** Based on the chaperone activity data at 37°C from Chapter 3 for both **H** and **G** samples, mathematical calculations revealed that, at 12  $\mu$ M, **G** sample showed approximately 95% inhibition of 118  $\mu$ M insulin. If we correlate with the Native-PAGE data of the **G** sample, we have 54  $\mu$ g (a total of 21.6  $\mu$ g of the 24 mers and 8  $\mu$ g of the 9, 10, 11 and 12 mers each). In terms of number of molecules, for a reaction volume of 0.25 ml, this works out to  $4.3 \times 10^{13}$ ,  $2.98 \times 10^{13}$ ,  $2.78 \times 10^{13}$ ,  $2.48 \times 10^{13}$ ,  $2.28 \times 10^{13}$

#### Chapter 4: Structure and Activity Relationship of Acr

molecules, or, a total of  $1.47 \times 10^{14}$  molecules. This corresponds to a total of 1.5 % of the total molecules of insulin ( $8 \times 10^{15}$ ) assuming, half of the 6 kDa B chain is the target to be inhibited. By covering 1.5% of the B chain of insulin, 95% inhibition was achieved. The same method was used to calculate the number of molecules of Acr that bound to insulin in the study of **G** samples after pre-heat treatment.

Before pre-heat treatment, the number of molecules of Acr of 1  $\mu$ M sample was  $2.7 \times 10^{12}$  molecules of 24 mers,  $2.5 \times 10^{12}$  molecules of 9 mers,  $2.28 \times 10^{12}$  molecules of 10 mers,  $2.05 \times 10^{12}$  molecules of 11 mers and  $1.88 \times 10^{12}$  molecules of 12 mers, or a total of  $1.14 \times 10^{13}$  molecules. This still covered only 0.14% of the total molecules of the B chain. However, when pre-heat treated at 37°C, 60°C and 70°C, the chaperone activity improved significantly, showing similar inhibition as if the number of molecules was 13-fold higher. Based on the chaperone activity data at 37°C, mathematical calculations revealed that, at 12  $\mu$ M, **G** sample showed approximately 95% inhibition of 118  $\mu$ M insulin. If we correlate with the Native-PAGE data of the **G** sample, we have 54  $\mu$ g (a total of 21.6  $\mu$ g of the 24 mers and 8  $\mu$ g of the 9, 10, 11 and 12 mers each). In terms of number of molecules, for a reaction volume of 0.25 ml, this works out to  $3.011 \times 10^{13}$ ,  $2.98 \times 10^{13}$ ,  $2.78 \times 10^{13}$ ,  $2.48 \times 10^{13}$ ,  $2.28 \times 10^{13}$  molecules, or, a total of  $1.35 \times 10^{14}$  molecules. This corresponds to a total of 1.5% of the total molecules of insulin ( $8 \times 10^{15}$ ) assuming, half of the 6 kDa B chain is the target to be inhibited. By covering 1.5% of the B chain of insulin, 95% inhibition was achieved. The same method was used to calculate the number of molecules of Acr that bound to insulin in the study of **G** samples after heat treatment. Before heat treatment, the number of molecules of Acr of 1  $\mu$ M sample was  $2.49 \times 10^{12}$  molecules of 24 mers,  $2.5 \times 10^{12}$  molecules of 9 mers,  $2.28 \times 10^{12}$  molecules of 10 mers,  $2.05 \times 10^{12}$  molecules of 11 mers and  $1.88 \times 10^{12}$  molecules of 12 mers, or a total of  $1.14 \times 10^{13}$  molecules. This still covered only 0.14% of the total molecules of the B chain. However, when pre-heat treated at 37°C, 60°C and 70°C, the chaperone activity improved significantly, showing

## Chapter 4: Structure and Activity Relationship of Acr

similar inhibition as if the number of molecules was 13-fold higher. These results are summarized in Table 4.2.

**Table 4.2: Calculations of molecular level binding of Acr to insulin B chain:** Calculations of molecular level interaction of Acr 30  $\mu$ M of **H**, 12  $\mu$ M of **G** and **G** 1  $\mu$ M with pre-heat treatment are based on the following assumptions: reaction volume 0.25 ml; insulin molecular weight of B chain 3 kDa and Acr molecular weight 18 kDa. The total amount of Acr in the reaction was recalculated based on the Native-PAGE data and the ratio of oligomers assuming an error estimation of  $\pm 10$  %.

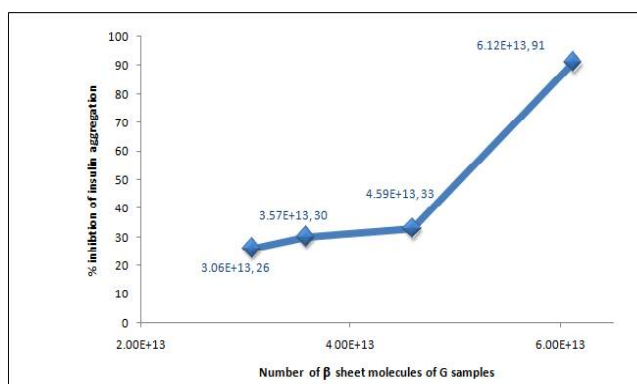
	Mol. wt (kDa)	Amount ( $\mu$ g)	molecules of insulin B chain alone	% of insulin B chain covered
Insulin	6	174	$8.73 \times 10^{15}$	
Acr	18			
<b>G</b> (12 $\mu$ M)	18	54	$1.8 \times 10^{15}$	
40% of 24 mers	288	21.6	$3.011 \times 10^{13} \pm 0.43 \times 10^{13}$	
15% of 9 mers	162	8.1	$2.98 \times 10^{13} \pm 0.298 \times 10^{13}$	
15% of 10 mers	180	8.1	$2.7 \times 10^{13} \pm 0.27 \times 10^{13}$	
15% of 11 mers	198	8.1	$2.48 \times 10^{13} \pm 0.248 \times 10^{13}$	
15% of 12 mers	216	8.1	$2.28 \times 10^{13} \pm 0.228 \times 10^{13}$	
Total Acr molecules			$1.35 \times 10^{14}$	1.54
<b>H</b> (30 $\mu$ M)		135	$4.52 \times 10^{15} \pm 0.452 \times 10^{15}$	
70% of 9 mers	162	94.5	$3.5 \times 10^{14} \pm 0.35 \times 10^{14}$	
10% of 10 mers	180	13.5	$4.39 \times 10^{13} \pm 0.439 \times 10^{13}$	
10% of 11 mers	198	13.5	$4.13 \times 10^{13} \pm 0.413 \times 10^{13}$	
10% of 12 mers	216	13.5	$3.76 \times 10^{13} \pm 0.376 \times 10^{13}$	
Total Acr molecules			$4.73 \times 10^{14} \pm 0.473 \times 10^{14}$	6
<b>G</b> (1 $\mu$ M) with pre-heat treatment	18	4.5	$1.50 \times 10^{14}$	
40% of 24 mers	288	1.8	$2.49 \times 10^{12} \pm 0.370 \times 10^{12}$	
15% of 9 mers	162	0.68	$2.53 \times 10^{12} \pm 0.253 \times 10^{12}$	
15% of 10 mers	180	0.68	$2.28 \times 10^{12} \pm 0.228 \times 10^{12}$	
15% of 11 mers	198	0.68	$2.07 \times 10^{12} \pm 0.207 \times 10^{12}$	
15% of 12 mers	216	0.68	$1.88 \times 10^{12} \pm 0.188 \times 10^{12}$	
Total Acr molecules			$1.24-1.125 \times 10^{13} + 0.124 \times 10^{13}$	0.14 - 0.15

**4.3 Calculations of molecular level binding of Acr to insulin B chain in terms of secondary structure:** The plots of  $\beta$  sheets versus % inhibition for **G** samples revealed an interesting pattern;  $6 - 7 \times 10^{13}$  molecules were required for complete inhibition of insulin [Figure 4.3 (A) and 4.3 (B)]. Graphs were plotted for **G** samples in terms of  $\alpha$  helices, random coils and turns [Figure 4.3 (C) to 4.3 (J)]. The number of molecules of  $\alpha$  helices that bound to insulin was greater than  $1.4 \times 10^{13}$ . The number of molecules of random coils that bound to insulin was greater than  $5 \times 10^{13}$ . The number of molecules of turns that bound to insulin was greater than  $1.5 \times 10^{13}$ .

Graphs of secondary structure analyses for **H** samples were plotted [Figure 4.3 (K) to 4.3 (T)] in order to explain activity and heat treatment data. The number of  $\beta$  sheets required for complete inhibition was  $1.195 \times 10^{13}$  molecules. A similar analysis was carried out for  $\alpha$  helices, random coils and turns, which showed an interesting pattern. For  $\alpha$  helices in **H** samples, it required about  $1.2 \times 10^{14}$  molecules to completely inhibit insulin B chain aggregation. The number of random coils required to ensure 95% inhibition of insulin was  $1 \times 10^{14}$ . The plot of turns showed that  $1.68 \times 10^{14}$  molecules of **H** were needed to completely inhibit the insulin B chain aggregation.

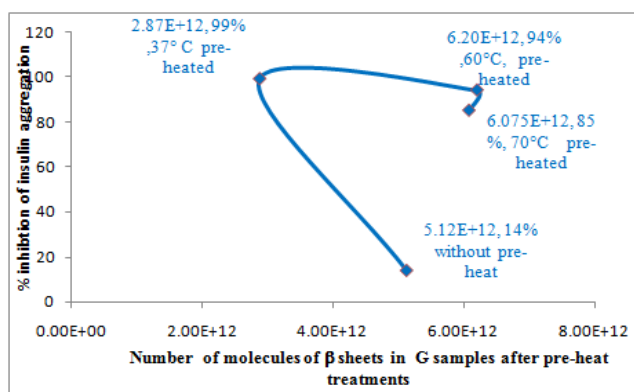
The percentage of  $\beta$  sheets (45.5%) obtained from the CD analysis (Table 4.1.1.2) was multiplied by the total number of Acr molecules at 4 different concentrations 6, 7, 9 and 12  $\mu\text{M}$  to estimate the actual number of  $\beta$  sheets interacting with insulin. The total amount of Acr in the reaction was recalculated based on the Native-PAGE data and the ratio of oligomers (40% of 24 mers and 15% of 9, 10, 11 and 12 mers) (Table 2.6.1). This calculation was repeated for all **G** samples.

## Chapter 4: Structure and Activity Relationship of Acr



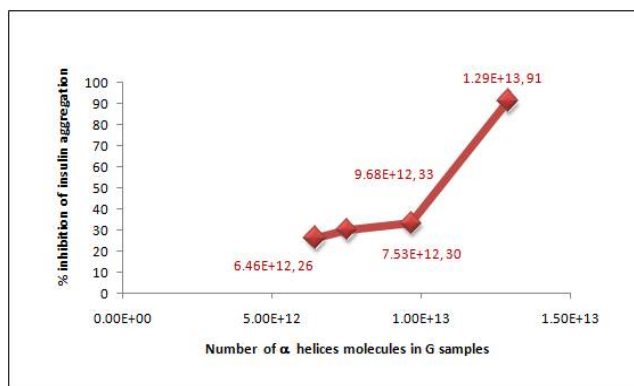
**Figure 4.3 (A):** Plot of molecules of  $\beta$  sheets versus % inhibition for **G** samples without pre-heat treatment.

The % of  $\beta$  sheets for 1  $\mu\text{M}$  Acr was calculated as 45.5% at 37°C without pre-heat treatment and 25.0, 51.1 and 54.6% respectively for 37°C, 60°C and 70°C respectively (Table 4.1.1.2) and multiplied by the number of molecules.

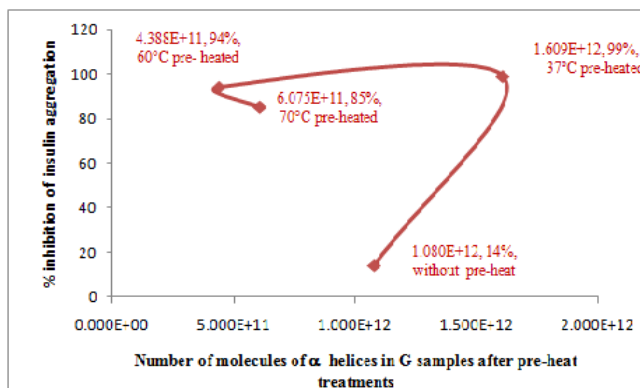


**Figure 4.3 (B):** Plot of  $\beta$  sheets versus % inhibition for **G** sample (1  $\mu\text{M}$ ) with pre-heat treatment.

The % of  $\alpha$  helices for **G** samples was 9.6% (Table 4.1.1.2) and this was multiplied with the four different concentrations of Acr 6, 7, 9 and 12  $\mu\text{M}$  to estimate the actual number of  $\alpha$  helices interacting with insulin.

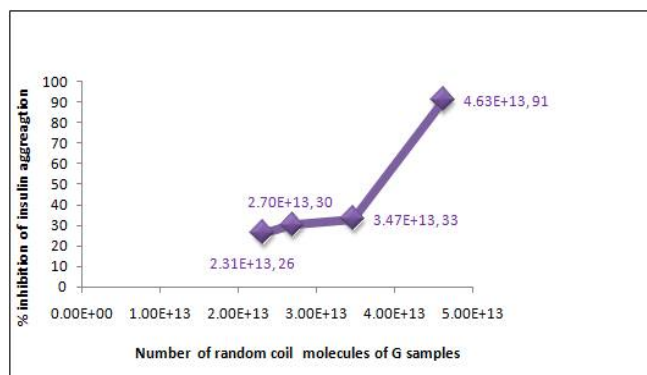


**Figure 4.3 (C)** Plot of  $\alpha$  helices versus % inhibition for **G** samples without pre-heat treatment. The % of  $\alpha$  helices for 1  $\mu$ M Acr sample was calculated to be 9.6% for 37°C without pre-heat treatment and 14.3, 3.9 and 5.4% for 37°C, 60°C and 70°C respectively (**Table 4.1.1.2**) and multiplied by the number of molecules.



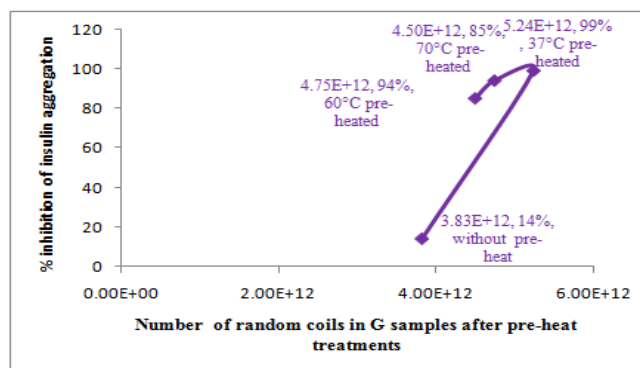
**Figure 4.3 (D)**: Plot of  $\alpha$  helices versus % inhibition for **G** sample (1  $\mu$ M) with pre-heat treatment.

The % of random coils of **G** samples (34.4%) (**Table 4.1.1.2**) was multiplied with four different concentrations of Acr 6, 7, 9 and 12  $\mu$ M to estimate the actual number of random coils interacting with insulin.



**Figure 4.3 (E):** Plot of random coils versus % inhibition for **G** samples without pre-heat treatment.

The % of random coils for 1  $\mu\text{M}$  Acr sample was calculated as 34.4% for 37°C without pre-heat treatment and 46, 40 and 42% for 37°C, 60°C and 70°C respectively (**Table 4.1.1.2**) and multiplied by the number of molecules.



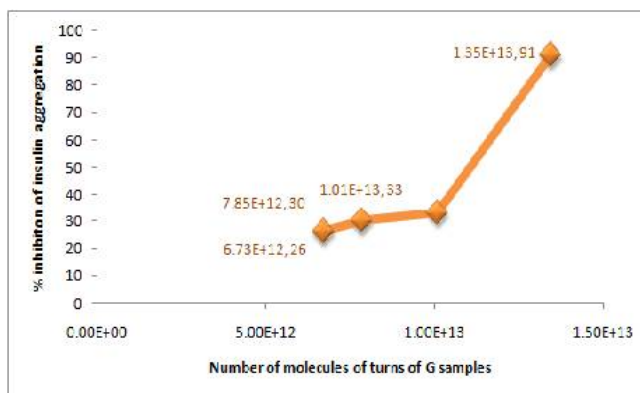
**Figure 4.3 (F):** Plot of random coils versus % inhibition for **G** sample (1  $\mu\text{M}$ ) with pre-heat treatment.

**Molecular level binding of Acr (with pre-heat treatment) to insulin B chain in terms of secondary structure:** For **G** samples, for 1  $\mu\text{M}$  Acr with pre-heat treatment,  $2.8 - 7 \times 10^{12}$  molecules of  $\beta$  sheets,  $4.4 \times 10^{11} - 1.6 \times 10^{12}$  molecules of  $\alpha$  helices,  $4.56 - 5.3 \times 10^{12}$  molecules of random coils and  $3.3 \times 10^{11} - 1.59 \times 10^{12}$  molecules of turns were sufficient for complete inhibition [**Figure 4.3 (B), 4.3 (D), 4.3 (F) and 4.3 (H)**]. In this instance, a clear pattern of reduction of activity was recorded with reduction of random coils. An interesting

## Chapter 4: Structure and Activity Relationship of Acr

point observed here, is a trend of decrease of chaperone activity with decrease in turns. However, even at 0% turns, 85% inhibition was obtained with the pre-heated sample at 70°C. Pre-heat treatment of **H** samples at 11  $\mu\text{M}$ , at 37°C, showed that,  $1 \times 10^{14}$  molecules of  $\beta$  sheets,  $2.44 \times 10^{13}$  molecules of  $\alpha$  helices,  $7.4 \times 10^{12}$  molecules of turns and  $3.27 \times 10^{13}$  molecules of random coils were required for 48% inhibition of the insulin B chain aggregation [**Figure 4.3 (L), 4.3 (P), 4.3 (N) and 4.3 (O)**].

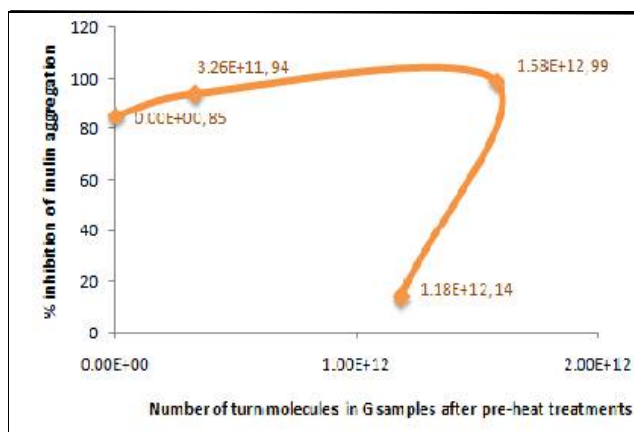
The % of turns of **G** samples (10.5%) (**Table 4.1.1.2**) was multiplied with four different concentrations of Acr 6, 7, 9 and 12  $\mu\text{M}$  to estimate the actual number of turns, obtained by multiplying with the number of molecules interacting with insulin.



**Figure 4.3 (G):** Plot of turns versus % inhibition for **G** samples without pre-heat treatment.

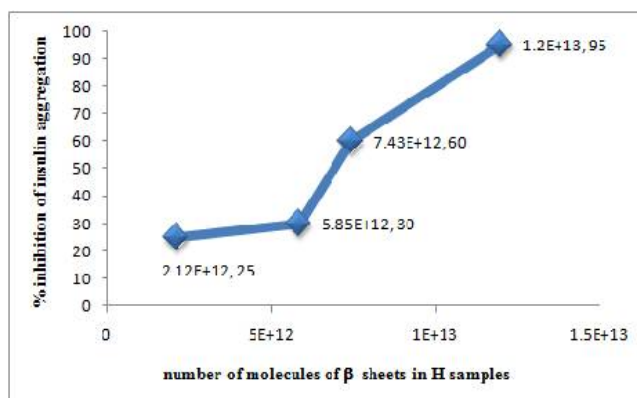
The % of turns for 1  $\mu\text{M}$  Acr sample was calculated to be 10.5% for 37°C without pre-heat treatment and 14, 2.9 and 0% for 37°C, 60°C and 70°C respectively (**Table 4.1.1.2**) and multiplied by the number of molecules.





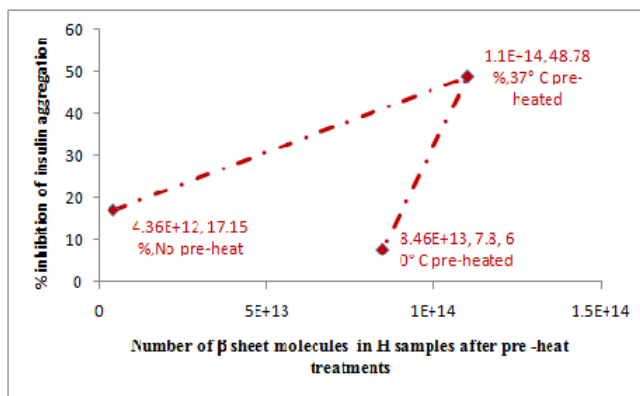
**Figure 4.3 (H):** Plot of turns versus % inhibition for **G** sample (1  $\mu\text{M}$ ) with pre-heat treatment.

The % of  $\beta$  sheets (2.5%) (Table 4.1.1.1) was multiplied by the total number of Acr molecules at four different concentrations 5.5, 11, 18 and 30  $\mu\text{M}$  to estimate the actual number of  $\beta$  sheets interacting with insulin. The total amount of Acr in the reaction was recalculated based on Native-PAGE data and the ratio of oligomers (70% of 9 mers and 10% of 10, 11 and 12 mers) (Table 2.6.1). This process was repeated for all **H** samples.



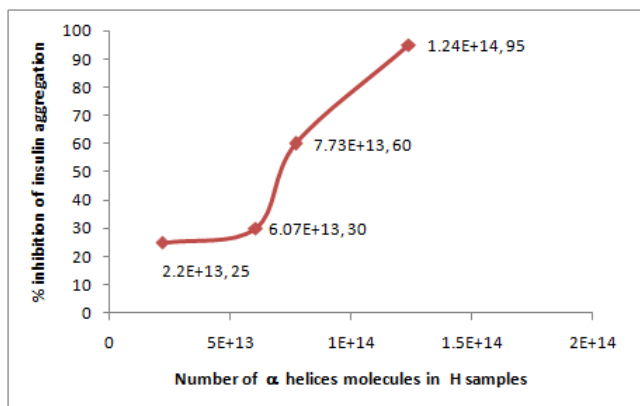
**Figure 4.3 (I):** Plot of molecules of  $\beta$  sheets versus % inhibition for **H** samples.

The % of  $\beta$  sheets for 11  $\mu\text{M}$  of **H** samples was calculated as 2.5% for 37°C without pre-heat treatment and 62.5 and 48.5% for 37°C, and 60°C respectively (Table 4.1.1.1) and multiplied by the number of molecules.



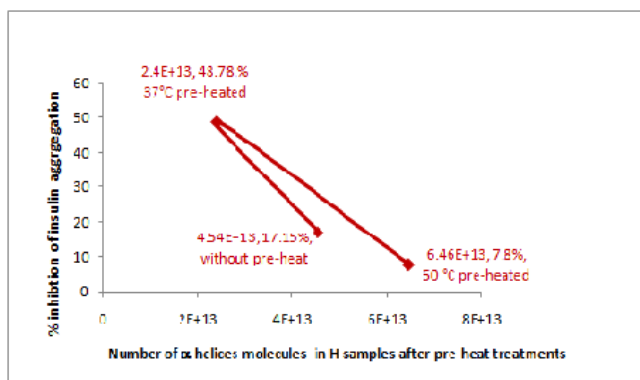
**Figure 4.3 (J):** Plot of  $\beta$  sheets for **H** sample (11  $\mu$ M) with pre-heat treatment.

The % of  $\alpha$  helices for the **H** samples was 26% (Table 4.1.1.1) and this was multiplied with the four different concentrations of Acr 5.5, 11, 18 and 30  $\mu$ M to estimate the actual number of  $\alpha$  helices interacting with insulin.



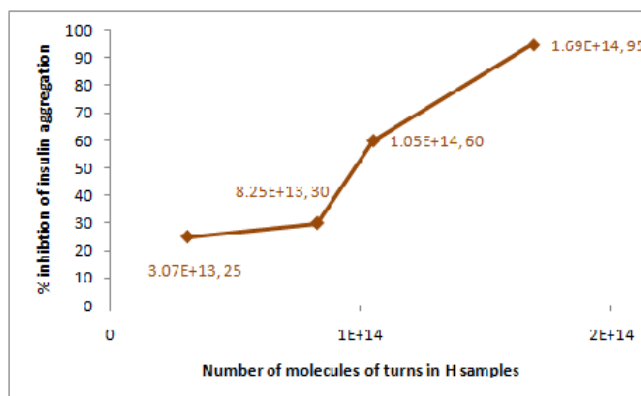
**Figure 4.3 (K):** Plot of  $\alpha$  helices versus % inhibition for **H** samples without pre-heat treatment.

The % of  $\alpha$  helices for 11  $\mu$ M of **H** samples was calculated as 26% for 37°C without pre-heat treatment and 14 and 3.7% for 37°C and 60°C respectively (Table 4.1.1.1) and multiplied by the number of molecules.



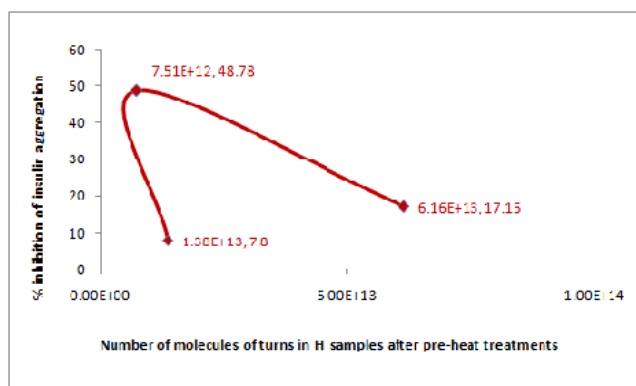
**Figure 4.3 (L):** Plot of  $\alpha$  helices versus % inhibition for **H** sample (11  $\mu$ M) with pre-heat treatment.

The % of turns for the **H** samples was 35.3% (**Table 4.1.1.1**) and this was multiplied with four different concentrations of Acr 5.5, 11, 18 and 30  $\mu$ M to estimate the actual number of turns interacting with insulin.

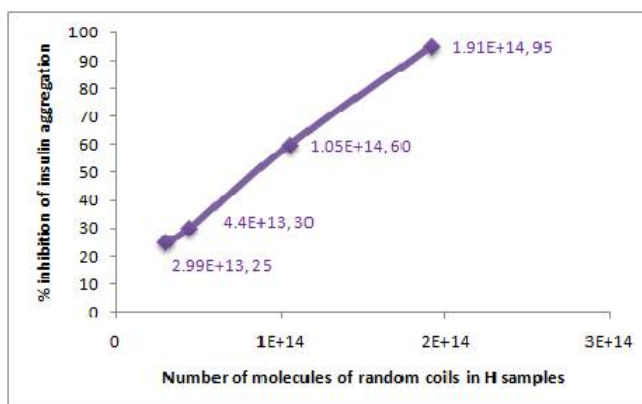


**Figure 4.3 (M):** Plot of turns versus % inhibition for **H** samples without pre-heat treatment.

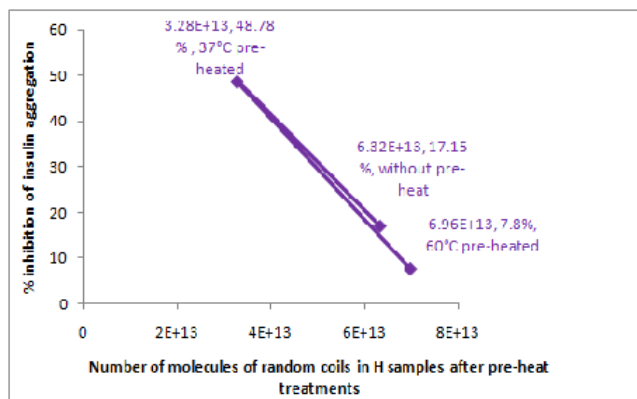
The % of turns for 11  $\mu$ M of **H** samples was calculated as 35.3% for 37°C without pre-heat treatment and 4.3 and 7.9% for 37°C and 60°C respectively (**Table 4.1.1.1**) and multiplied by the number of molecules.



**Figure 4.3 (N):** Plot of turns versus % inhibition for **H** sample (11  $\mu\text{M}$ ) with pre-heat treatment. The % of random coils for the **H** samples was 36.2% (**Table 4.1.1.1**) and this was multiplied with four different concentrations of Acr 5.5, 11, 18 and 30  $\mu\text{M}$  to estimate the actual number of turns interacting with insulin.

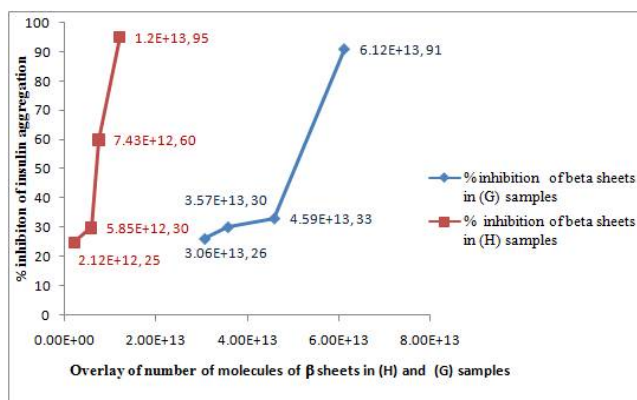


**Figure 4.3 (O):** Plot of random coils versus % inhibition for **H** samples without pre-heat treatment. The % of random coils for 11  $\mu\text{M}$  of **H** samples was calculated as 35.3% for 37°C without pre-heat treatment and 18.8 and 39.9% for 37°C and 60°C respectively (**Table 4.1.1.1**) and multiplied by the number of molecules.



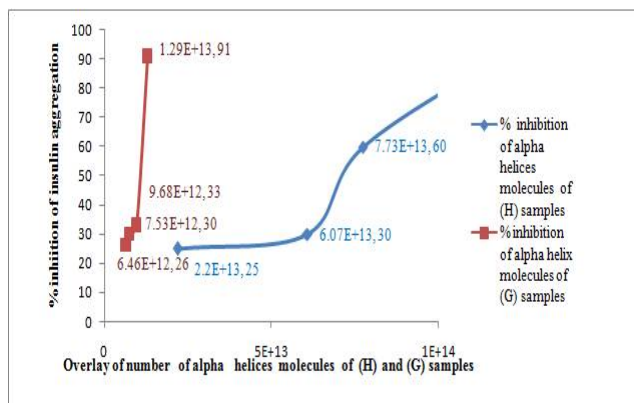
**Figure 4.3 (P):** Plot of random coils versus % inhibition for **H** sample (11  $\mu\text{M}$ ) with pre-heat treatment.

**Figures 4.3.1 and 4.3.9** were overlaid to compare the  $\beta$  sheets of **H** and **G** at four different concentrations of 5.5, 11, 18 and 30  $\mu\text{M}$  and 6, 7, 9 and 12  $\mu\text{M}$  respectively.



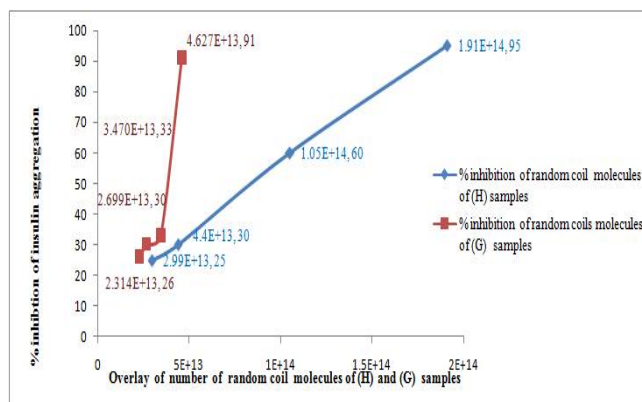
**Figure 4.3 (Q):** Overlay of  $\beta$  sheets versus % inhibition for **H** and **G** samples without pre-heat treatment.

**Figures 4.3.3 and 4.3.11** were overlaid to compare the  $\beta$  sheets of **H** and **G** at four different concentrations of 5.5, 11, 18 and 30  $\mu\text{M}$  and 6, 7, 9 and 12  $\mu\text{M}$  respectively.



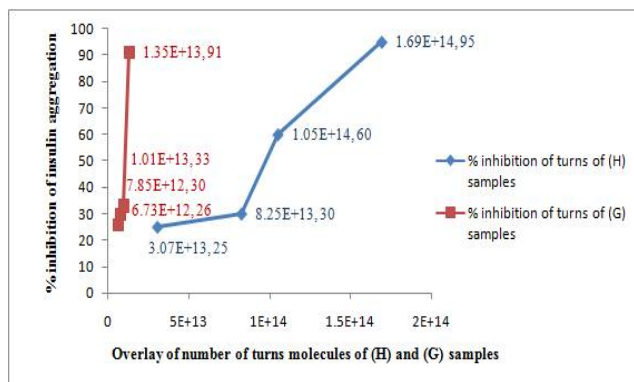
**Figure 4.3 (R):** Overlay of  $\alpha$  helices versus % inhibition for **H** and **G** samples without pre-heat treatment.

**Figures 4.3.5** and **4.3.15** were overlaid to compare the random coils of **H** and **G** at four different concentrations of 5.5, 11, 18 and 30  $\mu\text{M}$  and 6, 7, 9 and 12  $\mu\text{M}$  respectively.



**Figure 4.3 (S):** Overlay of random coils versus % inhibition for **H** and **G** samples without pre-heat treatment.

**Figures 4.3.7** and **4.3.13** were overlaid to compare the turns of **H** and **G** at four different concentrations of 5.5, 11, 18 and 30  $\mu\text{M}$  and 6, 7, 9 and 12  $\mu\text{M}$  respectively.



**Figure 4.3 (T):** Overlay of turns versus % inhibition for **H** and **G** samples without pre-heat treatment.

**General observations of secondary structure molecules required to inhibit insulin:** The following observations were recorded in terms of number of secondary structure molecules required to completely inhibit insulin (**Table 4.3**); the amount of  $\beta$  sheets required in **G** samples was 5.5 times more than in **H** samples whereas the amount of  $\alpha$  helices was ten times less than in **H** samples.

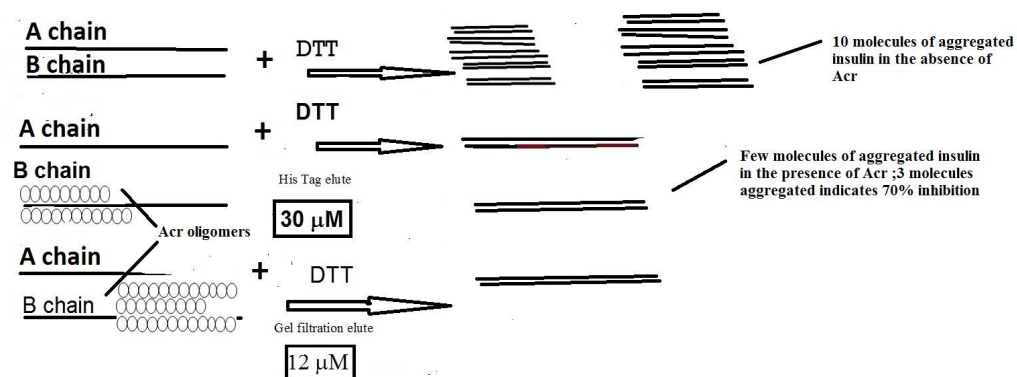
## Chapter 4: Structure and Activity Relationship of Acr

**Table 4.3:** A comparison of **H** and **G** samples in terms of 4 secondary structures to achieve 95% inhibition of insulin aggregation: The number of molecules of four secondary structures required to achieve 95% inhibition was compared for **H** and **G** samples and the total number of molecules (mass balance) compared with the theoretical number of molecules expected by Native-PAGE and the amount used for assay.

Secondary conformations and other parameters	Molecules in H samples	Molecules in G samples	Ratios G/H
$\beta$ sheets	$1.1 \times 10^{13}$	$6.1 \times 10^{13}$	5.55
$\alpha$ helices	$1.2 \times 10^{14}$	$1.29 \times 10^{13}$	0.12
Turns	$1.6 \times 10^{14}$	$1.28 \times 10^{13}$	0.10
Random coils	$1.9 \times 10^{14}$	$4.63 \times 10^{13}$	0.08
Total molecules	$4.81 \times 10^{14}$	$1.33 \times 10^{14}$	0.22
Actual expected by Native-PAGE	$4.54 \times 10^{14}$	$1.47 \times 10^{14}$	0.32
Mass balance	106 %	91%	--

A schematic diagram has been drawn to explain the molecular level interaction of Acr with insulin B chain with the two representative concentrations of **H** and **G** of 12 and 30  $\mu\text{M}$  respectively. The diagram depicts that how the higher amount of oligomers helps probably in improved coverage of the insulin B chains.

**Figure 4.3 (U):** Schematic presentation of binding of Acr to insulin





## Chapter 4: Structure and Activity Relationship of Acr

Proposed formulae: Chaperone activity of *M. tb* Acr on insulin

$$A = f \{R_1, \dots, R_n, M, C, \beta, \alpha, T, R_c, \Delta\}$$

A = Activity. It is defined as > 95% inhibition of DTT induced aggregation of insulin B chain

f = function of

R = ratio of oligomers in sample ( $R_1$  – 9 mers,  $R_2$  – 10 mers,  $R_3$  – 11 mers,  $R_4$  – 12 mers,  $R_5$  – 24 mers)

M = number of molecules of Acr, bound to 3 kDa of insulin

C = % of insulin B chain covered

$\beta$  = molecules of  $\beta$  sheets in the sample, where,  $5 - 6 \times 10^{13} < \beta < 1 - 2 \times 10^{14}$

$\alpha$  = molecules of  $\alpha$  helices in the sample, where,  $10^{13} < \alpha < 10^{14}$

T = molecules of turns, where,  $1.5 \times 10^{13} < T < 1.5 \times 10^{14}$

$R_c$  = molecules of random coils, where,  $5 \times 10^{13} < R_c < 2 \times 10^{14}$

$\Delta$  = heating of samples, where,  $37^\circ\text{C} < \Delta < 70^\circ\text{C}$

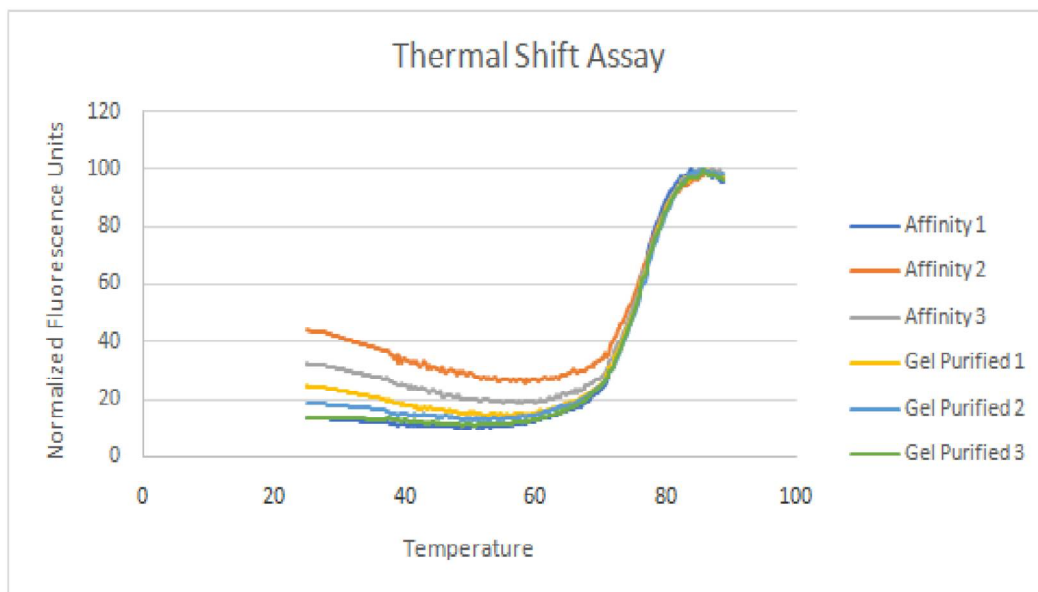
### 4.4 $T_m$ calculation

The assays showed a  $T_m$  value of  $76^\circ\text{C}$  for both His-tag and gel filtered samples. However, the initial values showed a higher value for the His-tag samples as compared to the gel filtered samples with a difference of nearly 17% at  $25^\circ\text{C}$ . This difference dropped to 6% at  $60^\circ\text{C}$  and after  $60^\circ\text{C}$ , the gel filtered fluorescence sample showed a higher value, though it did not reflect in any change in the  $T_m$ .

If we compare the  $T_m$  data with the Far-UV CD data, we observe a few interesting observations. Firstly at  $25^\circ\text{C}$  and  $37^\circ\text{C}$ , there is a clear difference in binding to the Sypro Orange, with the gel filtered samples showing nearly 17% less fluorescence at  $25^\circ\text{C}$ , 14% less fluorescence at  $37^\circ\text{C}$ . The Far-UV data shows that for His-tag samples, alone there are less  $\beta$  sheets at  $25^\circ\text{C}$  and  $37^\circ\text{C}$  and higher  $\beta$  sheets than gel filtered while pre-heating at  $37^\circ\text{C}$ . This could be the reason for the

## Chapter 4: Structure and Activity Relationship of Acr

higher initial fluorescence observed before 60°C. However, after 60°C, there is hardly any difference between the His-tag and the gel-filtered samples in terms of fluorescence.



**Figure 4.4 Thermal shift Assay of Affinity H and G samples.**

**Table 4.4.1: Melting temperature (°C) calculation.**

Sample type	Melting temperature(°C)	Average	STDEV
Affinity 1	75.22	76.0067	0.722
Affinity 2	76.64		
Affinity 3	76.16		
Gel Purified 1	75.97	75.9167	0.235
Gel Purified 2	76.12		
Gel Purified 3	75.66		

### DISCUSSION

**Use of Far-UV data to explain results of heat treatment studies:** It has been reported that mild heating of *M. tb* Acr at 30°C increases hydrophobic surfaces (Yang et al. 1999) and it was proposed that, for human  $\alpha$ -crystallin, hydrophobic binding sites are provided by  $\beta$  sheets in the oligomer and also by co-association with the C-terminal of a monomer (Van Montfort et al. 2001).

The improvement in activity at 37°C for **H** sample could be due to an increase in  $\beta$  sheets from 2.5 to 62%, which corroborates the activity profile obtained. Interestingly, at 60°C, though the number of turns increased, the activity dropped, strengthening the theory that the  $\beta$  sheets associate weakly and reversibly at high temperature but needs to be corroborated with more data points. The **H** sample, at 60°C and 70°C, initially showed an inhibition that was higher than the unheated sample, but the increase of the O. D. in the later part, indicated either dissociation or unstable nature of  $\beta$  sheets (**Figure 3.2.2.4.2**). It can be conjectured that a mixture of  $\beta$  sheets,  $\alpha$  helices, random coils and turns, might have contributed to chaperone activity, with, number of  $\beta$  sheets playing a significant role in **G** samples, due to the higher molecular weight oligomer.

**Comparison with Existing References:** It has been reported that the Far-UV CD spectrum of *M. tb* Acr, at 25°C, exhibited a minimum of 217 nm, indicating a possibility of high number of  $\beta$  sheets. The denaturation studies using 1M and 2M Guanidine Hydrochloride exhibited disorder in secondary structure (Yang et al. 1999). It has been reported that FTIR spectroscopy at 25°C exhibited predominance in  $\beta$  sheets (67%). This is one of two reports that provide a complete break-up of all the different conformations, ranging from helices to turns, in Acr. The report indicated that structural changes are more from 25°C to 40°C than from 40°C to 60°C and the disorder increased from 8 to 10% due to the increase in random coils. It also predicted that the highly thermostable  $\beta$  sheets (67%) help in chaperone activity of Acr (Gu et al. 2002). We

## Chapter 4: Structure and Activity Relationship of Acr

report a decrease in  $\beta$  sheets from 11.5 to 2.5% for **H** samples on heating from 25°C to 37°C and then, an increase to 62.9% upon pre-heat treatment at 37°C for 15 mins. which enhances the chaperone activity (**Figure 3.2.2.4.2**). Contrary to this, pre-heat treatment at 60°C and 70°C led to decrease in chaperone activity. It can be conjectured that the reduction in  $\beta$  sheets from 62.9% to 48.5% at 60°C and to 30% at 70°C could have contributed to this phenomenon.

Secondary structure variations using Far-UV CD from 25°C to 65°C have been recorded and  $\beta$  sheets were found to be dominant in Acr (Fu et al. 2005). Another report at the same temperature range showed decreased  $\beta$  sheets and increased random coils. The same report showed a shift in Far-UV CD spectra towards the left (Fu et al. 2003). We have recorded that, for **G** samples, there is an increase in random coils on heating [**Figure 4.3 (D)**]. Also, there was a decrease in  $\beta$  sheets from  $5 \times 10^{12}$  to  $2.85 \times 10^{12}$  on pre-heat treatment at 37°C. However, on pre-heat treatment at 60°C and 70°C, the  $\beta$  sheets increased to  $5.8 \times 10^{12}$  and  $6.2 \times 10^{12}$  respectively. It can be conjectured that stronger association of higher molecular weight oligomers could have contributed to this. Yet another report recorded Far-UV CD spectra of Acr at 25°C. This revealed a predominance of  $\beta$  sheets (Mao et al. 2001). Far-UV CD analysis at different temperatures from 37°C to 60°C have shown an alteration in secondary structure with a shift towards the left of the Far-UV CD spectra (Chen et al. 2005). We too have recorded the same for the **G** samples [**Figure 4.1.2 (E)**]. Far-UV CD studies of 34%  $\beta$  sheets and 16%  $\alpha$  helices at 25°C have also been recorded. They reported the break-up of different features of secondary structure (Panda et al. 2017). In brief, we recorded Far-UV CD data of **G** samples that are in agreement with existing reports with the predominance of  $\beta$  sheets and the increase in random coils upon heating. Our data for **H** samples indicate a bigger role for  $\alpha$  helices in chaperone activity and a weak and transient formation of  $\beta$  sheets that dissociates upon heating. This highlights the important role of gel filtration as a method to obtain optimum activity.

**Comparison with theoretical secondary structure:** This clearly supports the idea of a transitory  $\alpha$  helix in the **H** sample which converts into  $\beta$  sheets after gel filtration. This has been reported for other proteins, especially ovalbumin (Hu et al. 2000). We too observed the same for both **H** and **G** samples (Table 4.1.1.1 and 4.1.1.2).

**General observations of secondary structure molecules required to inhibit insulin:**

The amount of turns in **G** samples was ten times less than in **H** samples whereas the amount of random coils in **G** samples was five times less than in **H** samples. The mass balance obtained showed 106% for **H** samples and 91% for **G** samples [Table 4.3, Fig 4.3 (Q, R, S, T)]. This is the ratio of total molecules obtained from secondary structure data and the molecules theoretically expected from Native-PAGE analysis, based on the oligomer ratio.

**Determination of  $T_m$ :** The similar values of  $T_m$  obtained indicate similar secondary structure for both **H** and **G** samples. However, the initial binding of samples to the fluorescent dyes was higher for the non gel filtered samples as compared to the gel filtered samples, which could be due to the higher oligomeric status of the gel filtered samples which is an indirect confirmation of the oligomeric state. This initial lag in fluorescence was probably due to the difference in binding affinity matches with the previous report (Wong and Raleigh 2016). Sypro Orange was used to study dimer interactions by varying salt concentrations and studying the effect on dimer stability (Purohit et al. 2017). However, no such studies on  $T_m$  were done in the past with *M. tb* Acr, though effect of heating at different temperatures ranging from 37°C to 100°C (Mao et al. 2001) are available. Most of the structural studies were carried out to determine the role of different domains and many mutants generated especially in the N-terminal and C-terminal domains (Panda et al. 2017). This is the first report of the measurement of  $T_m$  of Acr. This information can be further used to study protein-ligand interactions especially Acr and insulin and see if the  $T_m$  data can be used to explain differences in chaperone activity that we observed and also help understand the mechanism of *in vitro* chaperone activity of Acr better.

## Chapter 4: Structure and Activity Relationship of Acr

**T<sub>m</sub> data analysis:** In the Far-UV data, too the His-tag samples show less  $\beta$  sheets at 60°C as compared to gel filtered samples which can be correlated with the T<sub>m</sub> data. It has been reported that at 60°C, an energy barrier is overcome, allowing hydrophobic surfaces to become exposed (Mao et al. 2001) and this seems to be the case from the T<sub>m</sub> data, where above 60°C, the fluorescence increases and becomes similar for both His-tag and gel filtered samples. The significance of this data needs to be studied with further and needs to be investigated in terms of more *in vitro* batches of Acr and a clear hypothesis proposed.

Reports exist on the association / dissociation studies of recombinant Acr (Chang et al. 1996, Yang et al. 1999, Fu et al. 2005, Feng et al. 2002, Abulimiti and Chang 2003). They have elucidated the spontaneous association and dissociation of the nonamer form of Acr. Other reports predicted that one molecule of nonamer Acr could bind to 6 molecules of insulin B chain (Yang et al. 1999).

### CONCLUSIONS

We have described here, a method of estimating oligomer ratio for Acr based on the ratio of 9 mers to 24 mers. It also helps to calculate the percentage of molecules of insulin covered and gives a mathematical tool to predict chaperone activity using polynomial graphs and equations. We have analyzed the chaperone activity of soluble *in vitro* Acr in terms of ratio of oligomer present that can accurately predict extent of binding to insulin B chain. This can explain 2 to 3-fold differences in chaperone activity. This information has been utilised to plot polynomials that can predict percent inhibition in terms of percent of insulin B chain covered that is far more accurate than the earlier polynomial curves we plotted. We also used mole ratio of Acr to insulin for all analyses, a parameter that has been rarely used in existing reports of Acr. A consolidated Far-UV CD secondary structure analysis with both, **H** and **G** samples and heat treatment studies at three different temperatures explained the variations in activity. The transition from  $\alpha$  helices to  $\beta$  sheets in the **H** and **G** samples is a new finding. The presence of more oligomers speeds up prevention of misfolding of protein substrates, within the macrophage. The calculation of  $T_m$  of **H** and **G** samples is also a new finding. The initial difference in fluorescence values correlated well with the Far-UV CD data obtained. This happens inside the host, especially, when the *M. tb* is in latent form inside the macrophage in a hypoxic condition. This new functional model of analyzing *in vitro* soluble recombinant Acr preparations can add valuable insights into its mechanism of action *in vivo*.

## **SUMMARY (CHAPTER 1)**

1. Success was achieved with directional cloning, using 2 restriction enzymes and calf intestinal alkaline phosphatase (CIAP). This approach was used to clone three genes, *sod A*, *sod C* and *acr*.
2. Parameter analysis identified 18 important parameters that affect the outcome of any cloning experiment.
3. Classifier analysis in MATLAB and Clone. R identified Simple Tree and Random forest to be the best classifiers which could predict the possibility of obtaining a clone.
4. Five constructs obtained were taken forward for expression studies.



## SUMMARY (CHAPTER 2)

1. *sod A* gene was induced at very low level at 25°C with overnight induction. The amount was too low for further studies.
2. *sod C* gene expression was not obtained.
3. *sod A* gene with ATG in the start codon did not express the protein.
4. The C-terminal *acr*-pET21a construct produced active soluble Acr from low levels of expression of Acr was obtained. Acr expression was obtained in the cell pellet at 37°C but there was a strong likelihood that this was Acr as the expressed protein could bind and elute under denaturing conditions to His-tag column. Identity of soluble Acr could not be verified, though increased chaperone activity upon boiling indicated it to be Acr.
5. In the present study, the N-terminal pET28a-*acr* gene construct overexpressed the Acr protein in soluble form. The expression levels obtained in the present study compare favourably with different constructs reported in the previous studies as summarized in **Table 2.7**.

## SUMMARY (CHAPTER 3)

1. C-terminal His-tag Acr: Citrate synthase assay at 45°C: Dose-dependent chaperone activity of **H** sample was seen which enhanced upon boiling for five to ten minutes.
2. N-terminal His-tag Acr: Citrate synthase assay at 45°C: **H** sample inhibited 85% of citrate synthase at a concentration of 5  $\mu$ M.
3. Insulin assay at 60°C: Dose-dependent activity was seen with 95% inhibition at 37.5  $\mu$ M of **H** sample Acr. **G** sample showed better activity than **H**.
4. Insulin assay at 37°C: Dose dependent activity seen with 95% inhibition at 30 mM of **H** sample Acr and 95% inhibition at 12  $\mu$ M of **G** samples. **G** sample showed better activity than **H**.
5. Heat treatment analysis at 37°C: The **H** sample improved activity only at 37°C pre-heating results.
6. The **G** sample improved activity after pre-heat treatment at 37°C, 60°C and 70°C.
7. Polynomial curves of mole ratio of Acr to insulin vs % inhibition was used to predict binding of Acr to insulin B chain especially based on oligomer size using simulation plots of theoretical versus real data.

## **SUMMARY (CHAPTER 4)**

1. Far-UV CD data. The **G** samples showed a higher  $\beta$  sheets than **H** samples which had a higher  $\alpha$  helices at 37°C than **G**.
2. On pre-heat treatment, the  $\beta$  sheets of **H** samples increased while the  $\alpha$  helices found reduced.
3. Comparison with theoretical secondary structure: the theoretical value matches more closely with **G** sample than **H** sample.
4.  $T_m$  calculations:  $T_m$  was found to be 76°C for both **H** and **G** samples. The initial binding of the fluorescent dye was higher for the non gel-filtered samples as compared to the gel-filtered samples.
5. Molecular interaction studies: **H** samples required a higher number of molecules to inhibit the same amount of insulin than **G** samples.

## THESIS CONCLUSIONS

To conclude in brief; starting with genomic DNA of *M. tb*, we could clone, express and characterize *M. tb* Acr, a very important protein for diagnosis of latent TB. A functional model to measure chaperone activity *in vitro* was developed as an outcome of the thesis for an important protein like *M. tb* Acr of latent TB. The use of polynomial graphs can be a useful tool for analyzing protein substrate interactions of *M. tb* Acr with potential *in vivo* substrates which could be done both theoretically as well as experimentally. The measurement of  $T_m$  has potential to study protein ligand interactions and draw future insights into mechanism of action of Acr.

This research could add value to the potential quest for new diagnostics especially the future studies of structure-activity relationships. This is especially important in view of the ever-increasing challenge of TB and in line with the goal of WHO; to eradicate TB by 2035.

## **FUTURE SCOPE**

The *M. tb* pET28a Acr construct can be used to raise antibodies and develop an immuno-histochemical diagnostic to detect presence of Acr in tissue samples. It could be further developed as a diagnostic for latent TB. The study could be first carried out in cattle and later in human populations.

Further structure activity relationship data could be generated with more batches of Acr, to strengthen the functional model of Acr activity by Native-PAGE analysis, Far / Near-UV CD data and Proton-NMR.

The structure activity relationship can be extrapolated to *in vivo* substrates of Acr in the host.

## REFERENCES

- Abarca Tomás B., Pell C., Bueno Cavanillas A., Guillen Solvas J., Pool R., and Roura M. Tuberculosis in Migrant Populations. (2013) A Systematic Review of the Qualitative Literature. PLoS ONE 8(12): e82440. doi: 10.1371/journal.pone.0082440.
- Abbas A. K. & Lichtman A. H. & Pillai S. (2007). Cellular Molecular Immunology 6<sup>th</sup> edition. pp 268-301. Saunders/Elsevier, Philadelphia, U. S. A.
- Abulimiti A. and Chang Z. (2003)  $\alpha$ -Crystallin Promotes Assembly of a Trimeric Form of *Mycobacterium tuberculosis* Hsp16.3 in a Cell Free System Biochemistry; (Moscow), 68(3) :269-274. Translated from Biokhimiya, 68(3):328-335. doi:10.1023/A:1023098015504
- Birnboim H. C. and Doily J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, Nucleic Acids Research; 7(6):1513-1523. doi:10.1093/nar/7.6.1513
- Bolivar F., Rodriguez R. L., Greene P. J., et al. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene;2(2): 95-113. doi:10.1016/0378-1119(77)90000-2
- Borzova V. A., Markossian K. A., Kara D. A., Chebotareva N. A., Makeeva V. F., et al. (2013) Quantification of Anti-Aggregation Activity of Chaperones: A Test-System Based on Dithiothreitol-Induced Aggregation of Bovine Serum Albumin. PLoS ONE; 8(9): e74367. doi: 10.1371/journal.pone.0074367
- Chen X., Fu X., Ma Y. and Chang Z. (2005) Chaperone-Like Activity of *Mycobacterium tuberculosis* Hsp16.3 Does Not Require Its Intact (Native) Structures; Biochemistry (Moscow), 70(8): 913-919. doi:10.1007/s10541-005-0202-4.
- Chang Z., Primm T. P., Jakana J., Lee I. H., Serysheva I., Chiu W., Gilbert H. F. and Quijcho F. A. (1996) *Mycobacterium tuberculosis* 16-kDa Antigen (Hsp16.3) Functions as an Oligomeric Structure *in Vitro* to Suppress Thermal Aggregation. Journal of Biological Chemistry; 271(12): 7218–7223. doi:10.1074/jbc.271.12.7218

Copland A., Diogo G. R., Hart P., Harris S., Tran A. C., Paul M. J., Singh M., Cutting S.M. and Reljic R. (2018) Mucosal Delivery of Fusion Proteins with *Bacillus subtilis* Spores Enhances Protection against Tuberculosis by Bacillus Calmette-Guérin. *Frontiers in Immunology*; 9:346. doi:10.3389/fimmu.2018.00346

D'Ambrosio L. D., Centis R., Dara M., Solovic I., Sulis G., Zumla A. and Migliori G. B. (2017) European policies in the management of tuberculosis among migrants. *International Journal of Infectious Diseases*; 56:85–89. doi: 10.1016/j.ijid.2016.11.002

D'Orazio M., Folcarelli S., Mariani F., Colizzi V., Rotilio G. and Battistoni. A. (2001) Lipid modification of the Cu, Zn superoxide dismutase from *Mycobacterium tuberculosis*. *Biochemical Journal* 359(1): 17–22. doi:10.1042/bj3590017

Dhepakson P., Luengchaichaweng A., Pudprom S., Thongthai P., Balachandra K. and Sawanpunyalert P. (2008) Construction and production of 16 kDa antigen from *Mycobacterium tuberculosis* for the development of TB diagnostic test. *Bulletin of Chiang Mai University, Associated Medical Sciences*; 41(3):205-213.

Dussurget O., Stewart G., Neyrolles O., Pescher P., Young D., and Marchal G. (2001) Role of *Mycobacterium tuberculosis* Copper-Zinc Superoxide Dismutase. *Infection and Immunity*; 69 (1): 529–533. doi: 10.1128/IAI.69.1.529–533

Dye C. (2014) “After 2015: infectious diseases in a new era of health and development”. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences* ; 369(1645): dx.doi.org/10/1098/rstb.2013.0426

Esmail H., Barry C. E., Young D. B. and Wilkinson R. J. (2014) The ongoing challenge of latent tuberculosis. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences*; 369(1645): 20130437 doi:10.1098/rstb.2013.0437

Espindolaa A. L, Varughese M., Laskowski M., Shoukat A., Heffernan J. M. and Moghadas S. M. (2017) Strategies for halting the rise of multidrug resistant TB epidemics: assessing the

effect of early case detection and isolation. *International Health*; 9(2):80–90. doi:10.1093/inthealth/ihw059.

Feng X., Huang S., Fu X., Abulimiti A. and Chang Z. (2002) The reassembling process of the nonameric *Mycobacterium tuberculosis* small heat shock protein Hsp16.3 occurs via a stepwise mechanism. *Biochemical Journal*; 363(2):329–334. doi :10.1042/bj3630329

Fletcher H. A. and Schrager L. (2016) TB vaccine development and the End TB Strategy: importance and current status. *Transactions of The Royal Society of Tropical Medicine and Hygiene*; 110(4): 212–218. doi:10.1093/trstmh/trw016

Focus (Bethesda Research Laboratories) (1979) Ligation Theory and Practice 2:2 & 3, Maryland, U. S. A.

Fu X., Zhang H., Zhang X., Cao Y., Jiao W., Liu C., Song Y., Abulimiti A. and Chang Z. (2005) A Dual Role for the N-terminal Region of *Mycobacterium tuberculosis* Hsp 16.3 in Self-Oligomerisation and Binding Denaturing Substrate Proteins. *Journal of Biological Chemistry*; 280(8): 6337–6348. doi: 10.1074/jbc.M406319200

Fu X., Li W., Mao Q. and Chang Z. (2003) Disulfide bonds convert small heat shock protein Hsp16.3 from a chaperone to a non-chaperone: implications for the evolution of cysteine in molecular chaperones. *Biochem Biophys Res Commun*; 308(3):627–635. doi:10.1016/S0006-291X(03)01450-5

Gideon H. P. and Flynn J. L. (2011) Latent tuberculosis: what the host “sees”? *Immunological Research*; 50(2-3): 202–212. doi:10.1007/s12026-011-8229-7

Gräslund S., Nordlund P., Weigelt J., Hallberg B. M., Bray J. et al. (2008) Protein production and purification. *Nature Methods*; 5(2): 135–46. doi: 10.1038/nmeth.f.202

Gu L., Abulimiti A., Li W. and Chang Z. (2002) Monodisperse Hsp16.3 nonamer exhibits dynamic dissociation and reassociation, with the nonamer dissociation prerequisite for



chaperone-like activity. *Journal of Molecular Biology*; 319(2): 517–526. doi:10.1016/S0022-2836(02)00311-X

Harth G. and Horwitz M. A. (1999) Export of Recombinant *Mycobacterium tuberculosis* Superoxide Dismutase Is Dependent upon Both Information in the Protein and Mycobacterial Export Machinery: A MODEL FOR STUDYING EXPORT OF LEADERLESS PROTEINS BY PATHOGENIC MYCOBACTERIA. *Journal of Biological Chemistry*; 274(7):4281–4292. doi:10.1074/jbc.274.7. 4281

Hoffman L. M. and Jendrisak J. (1990) Heat-labile phosphatase simplifies the preparation of dephosphorylated vector DNA. *Gene*; 88(1):97–99. doi.10.1016/0378-1119 (90)90064-X

Hu H. Y. and Du H. N. (2000) Alpha-to-beta structural transformation of ovalbumin: heat and pH effects. *Journal of Protein Chemistry*; 19(3):177-83.

Hugo G. J. (2008) Migration and health Situation Report on International Migration in East and South- East Asia (Bangkok, International Organization for Migration, Regional Office for Southeast Asia). Publisher: International Organization for Migration, Regional Office for Southeast Asia

Jain R., Dey B., Khera A., Srivastav P., Gupta U. D., Katoch V. M., Ramanathan V. D. and Tyagi A. K. (2011) Over-expression of superoxide dismutase obliterates the protective effect of BCG against tuberculosis by modulating innate and adaptive immune responses. *Vaccine*; 29(45): 8118– 8125. doi: 10.1016/j.vaccine.2011.08.029

Kaufmann S. H. E. (2010) Future vaccination strategies against tuberculosis thinking outside the box. *Immunity*; 2010. 33(4):567- 577, doi: 10.1106/j.immuni.2010.09.0153

Kennaway C. K., Benesch J. L. P, Gohlke U., Wang L., Robinson C. V., Orlova E. V, Salbi H. R. and Keep N. H. (2005) Dodecameric Structure of the Small Heat Shock Protein Acr1 from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*; 280(39):33419-33425 doi: 10.1074/jbc.M504263200

Kennedy S. (2007) Blog Article “10 Tips For Better DNA Gel Extraction Results”. Nucleic Acid Purification and Analysis. <https://bitesizebio.com/13506/10-tips-for-better-dna-gel-extraction-results/>

Krairiksh N. and Bruce A. (2011) Situation Report on International Migration in South and South-West Asia. UN ESCAP (Economic and Social Commission for Asia and the Pacific) and IOM (International Organisation for Migration) [https://www.unescap.org/sites/default/sites/SDD\\_PUB\\_Sit-Rep-book-pdf](https://www.unescap.org/sites/default/sites/SDD_PUB_Sit-Rep-book-pdf)

Liao D., Fan Q., and Bao L. (2013) The Role of Superoxide Dismutase in the Survival Of *Mycobacterium tuberculosis* in Macrophages Japanese Journal of Infectious Diseases; 66(6): 480-488. doi: 10.7883/yoken.66.480

Lakey D. L., Voladri R. K. R., Edwards K. M., Hager C., Samten B., Wallis R. S., Barnes P. F. and Kernodle D. S. (2000) Enhanced Production of Recombinant *Mycobacterium tuberculosis* Antigens in *Escherichia coli* by Replacement of Low-Usage Codons. Infection and Immunity; 68(1):233-238. doi: 10.1128/IAI.68.1.233-238.2000

Mao Q., Ke D., Feng X. and Chang Z. (2001) Preheat Treatment for *Mycobacterium tuberculosis* Hsp16.3: Correlation between a Structural Phase Change at 60°C and a Dramatic Increase in Chaperone-like Activity. Biochemical Biophysical Research Communication; 284(4):942–947. doi:10.1006/bbrc.2001.5074

Matsumura I. (2015) Why Johnny can't clone: Common pitfalls and not so common solutions. BioTechniques; 59(3): IV-XIII doi:10.2144/000114324

Millet J. P., Orcau A., de Olalla P. G., Martín V., Moreno A., Fina L., del Baño L., Simón P., Masdeu Eva and Caylà J. A. (2014) Chapter 1.3. The current status of TB in the world: the influence of poverty, prisons, HIV, immigration and control programmes. Book chapter: The Art and Science of Tuberculosis Vaccine Development Vaccination. Nor N. M., Acosta A. and Sarmiento M. E. Oxford University Press. 2<sup>nd</sup> Edition. Malaysia.

Moreno-Mendieta S. A, Guillen D., Espitia G. C., Hernández-Pando R. D, Sanchez S. and Rodríguez-Sanoja R. (2014) A novel antigen-carrier system: The *Mycobacterium tuberculosis* Acr protein carried by raw starch microparticles. *International Journal of Pharmaceutics*; 474: 241–248. <http://dx.doi.org/10.1016/j.ijpharm.2014.07.041>

Novagen, pET System Manual, TB055 11th Edition 01/06

Novagen (1999) TB055 10th Edition Rev.B 0403 May, 2003.

Odone A., Tillman T., Sandgren A., Williams G., Rechel B., Ingleby D., Noori T., Mladovsky P. and McKee M. (2015) Tuberculosis among migrant populations in the European Union and the European Economic Area. *European Journal of Public Health*; 25(3): 506–512. doi: 10.1093/eurpub/cku208.

Ospina J. E., Orcau À., Millet J. P., Ros M., Gil S. and Caylà J. A. (2016) Epidemiology of Tuberculosis in Immigrants in a Large City with Large-Scale Immigration (1991-2013). *PLoS ONE*; 11(10): e0164736. doi:10.1371/journal.pone.0164736

Oswald N. Blog Article “5 Tips on Vector Preparation for Gene Cloning” (2007) *Nucleic Acid Purification and Analysis*. <https://bitesizebio.com/13500/cloning-tips-vector-prep/>

Panda A. K., Chakraborty A., Nandi S. K., Kaushik A. and Biswas A. (2017) The C-terminal Extension of *Mycobacterium tuberculosis* Hsp16.3 Regulates its Oligomerization Subunit Exchange Dynamics and Chaperone Function *FEBS Journal*; 284(2):277-300 doi: 10.1111/febs.13975.

Pareek M., Greenaway C., Noori T., Munoz J. and Zenner D. (2016) The impact of migration on tuberculosis epidemiology and control in high-income countries: a review *BMC Medicine*; 14:48 doi: 10.1186/s12916-016-0595-5.

Preneta R., Papavinasasundaram K. G., Cozzone A. J. and Duclos B. (2004) Autophosphorylation of the 16 kDa and 70kDa antigens (Hsp16.3 and Hsp70) of *Mycobacterium tuberculosis*. *Microbiology*; 150:2135-2142 doi: 10.1099/mic.0.26789-0

Purohit A., England J. K., Douma L. G., Tondnevis F., Bloom L. B., and Levitus M. (2017) Electrostatic Interactions at the Dimer Interface Stabilize the *E. coli*  $\beta$  Sliding Clamp. *Biophysical Journal*; 113:794–804. doi: 10.1016/j.bpj.2017.06.05

Piddington D. L., Fang F. C., Laessig T., Cooper A. M, Orme I. M. and Buchmeier N. A. (2001) Cu, Zn Superoxide Dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. *Infection and Immunity*; 69(8): 4980–4987. doi:10.1128/IAI.69.8.4980-4987.2001

Rizvi N., Ajai S., Ali S., Yadav M., Hussain S. R. and Kumar V. (2015) Expression of Alpha-Crystallin Protein in Osteoarticular Tuberculosis in Latent Phase: A Review of Evidences. *International Journal of Biomedical Research*; 6(7): 445-451. doi:10.7439/ijbr.v6i7.2232

Rani K., Datt S. and Rana R. Brief review on alkaline phosphatases-an overview. (2012) *International Journal of Microbiology and Bioinformatics*; 2(1):1- 4.

Rayees M. Mechanistic insights into cytosolic molecular chaperones in protein unfolding and disaggregation, Thesis (2014) Originally published at: University of Lausanne Posted at the University of Lausanne Open Archive <http://serval.unil.ch> Document URN: urn: nbn:ch: serval-BIB\_EA49B7E040F21.

Sambrook. J. R. & Russell D. W. (2001) *Molecular Cloning a Laboratory Manual Volume 1 Plasmids and Their Usefulness in Molecular cloning. The Innoue Method for Preparation and Transformation of E. coli Ultra competent cells.* 21: p1.112, 3rd edition, Cold Spring Harbor Press.Cold Spring Harbor, U.S.A

Sambrook J. R. & Russell D.W. (2001) *Molecular Cloning A Laboratory Manual Volume 1 Plasmids and Their Usefulness in Molecular cloning. Dephosphorylation of Plasmid DNA Protocol.* 20: p1.93, 3rd edition, Cold Spring Harbor Press. Cold Spring Harbor, U.S.A

Sherman D., Voskuil M., Schnappinger D., Liao R., Harrell M and Schoolnik G. (2001) Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding  $\alpha$ -crystallin. Proceedings of Natural Academy of Sciences; 98(13):7534-7539. doi:10.1073\_pnas.261577598.

Siddiqui K. F., Amir M., Khan N., Rama Krishna G., Sheikh J. A., Rajagopal K. and Agrewala J. N. (2015) Prime-boost vaccination strategy with bacillus Calmette–Guérin (BCG) and liposomized alpha-crystallin protein 1 reinvigorates BCG potency. Clinical & Experimental Immunology; 181 (2) 286–296.doi:10.1111/cei.12634

Smith I. (2003). *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. Clinical Microbiology Reviews; 16(3):463–496. doi:10.1128/CMR.16.3.463-496.2003

Sreenivas A., Rade K., Sachdeva K. S., Ghedia M., Parmar M., Ramachandran R. and Shepherd J. (2014) Standards for TB care in India. World Health Organisation. [http://www.searo.who.int/india/mediacentre/events/2014/stci\\_book.pdf](http://www.searo.who.int/india/mediacentre/events/2014/stci_book.pdf)

Soong J. X., Lim T. S. and Choong Y. S. (2018) The structural insights of 16.3 kDa heat shock protein (HSP16.3) from *Mycobacterium tuberculosis* via *in silico* study. Molecular Simulation; 44(2):117-127. doi:10.1080/08927022.2017.1346254

Sorensen H. P. and Mortensen K. K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microbial Cell Factories 2005, 4:1 doi:10.1186/1475-2859-4-1.

Stephenson F. H. (2010) Calculations for Molecular Biology and Biotechnology Recombinant DNA. pp 203-207.2nd Edition Elsevier, U.S.A

Taylor J., Wiczorek A., Keyser A. R., Grover A., Flinkstrom R., Karl R. K., Bielefeldt-Ohmann H., Dobos K. M. and Izzo A. (2012) HspX-mediated protection against tuberculosis depends on its chaperoning of a mycobacterial molecule. Immunology and Cell Biology; 90(10): 945–954. doi:10.1038/icb.2012.34

Valdez M. M., Clark J. I., Wu G. J. and Muchowski P. J. (2002) Functional similarities between the small heat shock proteins *Mycobacterium tuberculosis* HSP 16.3 and human alphaB-crystallin. *European Journal of Biochemistry*; 269(7): 1806-1813. doi:10.1046/j.1432-1033.2002.02812.x

Van der Werf M. J. and Zellweger J. P. (2016). Impact of migration on tuberculosis epidemiology and control in the EU/EEA. *Euro Surveill*; 21(12). doi: 10.2807/1560-7917. ES.2016.21.12.30174.

Van Montfort R. L., Basha E., Friedrich K. L., Slingsby C. and Vierling E. (2001) Crystal structure of a eukaryotic small heat shock protein. *Nature Structural Biology*; 8(12):1025-1030. doi: 10.1038/nsb722

Volpe E., Cappelli G., Grassi M., et al. (2006) Gene expression profiling of human macrophages at late time of infection with *Mycobacterium tuberculosis*. *Immunology*; 118(4):449 – 460. doi:10.1111/j.1365-2567.2006.02378.x

Wu C. H., Tsai-Wu J. J., Huang Y. T., Lin C. Y., Liou G. G., Lee F. J. (1998) Identification and subcellular localization of a novel Cu, Zn superoxide dismutase of *Mycobacterium tuberculosis*. *FEBS Letters*; 439(1-2):192-196 doi: 10.1016/S0014-5793(98)01373-8

Wieczorek A. E., Troudt J. L., Knabenbauer P., Taylor J., Pavlicek R. L., Karls R., Hess A., Davidson R. M., Strong M., et al. (2014) HspX vaccination and role in virulence in the guinea pig model of tuberculosis *Pathogens and Disease*; 71(3): 315–325. doi: 10.1111/2049-632X.12147

Wong Amy G., and Raleigh D. P. (2016) Dye SYPRO orange binds to amylin amyloid fibrils but not pre-fibrillar intermediates. *Protein Science*; 25(10):1834—1840. doi: 10.1002/pro.2992

Yang H., Huang S., Dai H., Gong Y., Zheng C. and Chang Z. (1999) The *Mycobacterium tuberculosis* small heat shock protein Hsp16.3 exposes hydrophobic surfaces at mild conditions:

Conformational flexibility and molecular chaperone activity *Protein Science*; 8(1):174–179. doi: 10.1110/ps.8.1.174

Yeow E. K. L. and Clayton A. H. A. (2007) Enumeration of Oligomerization States of Membrane Proteins in Living Cells by Homo-FRET Spectroscopy and Microscopy: Theory and Application. *Biophysical Journal*; 92(9):3098-3104. doi:10.1529/biophysj.106.099424

Yuan Y., Crane D. D., and Barry C. E. (1996) Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: function of the mycobacterial alpha-crystallin homolog. *Journal of Bacteriology*; 178(15):4484-4492. doi:10.1128/jb.178.15.4484-4492.1996

Yuan Y., Crane D. D., Simpson R. M., Zhu Y., Hickey M. J., Sherman D. R., and Barry C. E. (1998) The 16-kDa  $\alpha$ -crystallin (Acr) protein of *Mycobacterium tuberculosis* is required for growth in macrophages. *Proceedings of National Academy of Sciences U S A*; 95(16):9578–9583. doi:10.1073/pnas.95.16.9578

Zhang G. (2013) A New Overview on the Old Topic: The Theoretical Analysis of “Combinatorial Strategy” for DNA Recombination. *American Journal of Biomedical Research*; 1(4) :108-111. doi:10.12691/ajbr-1-4-6

Zhang Y., Lathigra R., Garbe T., Catty D. and Young D. (1991) Genetic analysis of superoxide dismutase, the 23 kilodalton antigen of *Mycobacterium tuberculosis*. *Molecular Microbiology*; 5 (2): 381-391. doi:10.1111/j.1365-2958.1991.tb02120.x.

### **List of url Links**

(<http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning>).

(<http://bionumbers.hms.harvard.edu/bionumber.aspx?id=101937>).

(<http://lifeserv.bgu.ac.il/wp/zarivach/wp-content/.../Novagen-pET-system-manual-1.pdf>).

([http://molbiol.edu.ru/eng/scripts/01\\_04.html](http://molbiol.edu.ru/eng/scripts/01_04.html)).

([http://www.un.org/millenniumgoals/2015\\_MDG\\_Report/pdf/MDG%202015%20Summary%20web\\_english.pdf](http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20Summary%20web_english.pdf)).

(<http://www.who.int/tb/strategy/end-tb/en/>).

(<https://nebiocalculator.neb.com/#!/dsdnaamt>).

(<https://blog.easysol.net/machine-learning-algorithms-1>).

([https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)).

([https://en.wikipedia.org/wiki/Linear\\_discriminant\\_analysis](https://en.wikipedia.org/wiki/Linear_discriminant_analysis)).

(<https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html>).

(<https://www.qiagen.com/in/resources/technologies/plasmid-resource->).

(<https://www.statisticsolutions.com/what-is-logistic-regression>).

(<https://www.newscientist.com/article/dn14941-oldest-cases-of-human-tb-found-beneath-the-sea/>).



## List of Publications and Presentations

### Publication details:

1. Gautam Krishnan and UTPAL ROY (2019): Gautam Krishnan and Utpal Roy, “A New Functional Model for Prediction of Chaperone Activity of the Recombinant *M. tb* Acr ( $\alpha$ -Crystallin) Using Insulin as Substrate,” *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 2019, Article ID 2532045, 18 pages, 2019. <https://doi.org/10.1155/2019/2532045>. Scopus-indexed and SCI-indexed
2. Gautam Krishnan and UTPAL ROY (2019): Role of Molecular interactions and oligomerisation in Chaperone Activity of Recombinant *Mycobacterium tuberculosis* Acr. *Iranian Journal of Biotechnology (In Press)*. Available Online from 17 March 2019. Scopus-indexed and SCI-indexed
3. Krishnan G. and Roy U. Prediction of recombinant *Mycobacterium tuberculosis*  $\alpha$ -crystallin oligomer chaperone activity using polynomial graphs. [**Under Review**]. *F1000 Research* **2018, 7:1801**

### Conference Publications:

1. Title of paper “Cloning and expression of antigenic proteins of pathogenic organisms: special focus on Mycobacterial family” by Dr Utpal Roy, Gautam Krishnan, Sam C., Joy B., Natesh, Siddharth T., Rohit Shivkumar presented at NATCON 2014 – Feb 7<sup>th</sup>, 2015, Mumbai
2. Conference presentation: Key note speaker “A functional model for predicting *in vitro* chaperone activity of *M. tb* Acr” at 3<sup>rd</sup> International Conference of Clinical Microbiology, Kyoto, Japan, May 1<sup>st</sup> - 2<sup>nd</sup>, 2019

## **Brief Biography of the Candidate**

Name: Gautam Krishnan

ID No: 2010PHXF0813G

Educational Qualification: M. Sc (Biological Sciences), B. Pharm 1989-1994, BITS Pilani

Work / Research Experience: He worked in the pharmaceutical and biotechnology industry for 14 years at Astra Zeneca India Pvt. Limited, Bangalore (January 1995 – July 2001), Biocon India Pvt Ltd, Bangalore (July 2001 – July 2006), Goodwin Biotechnology India, Goa (September 2006 – May 2009). At Astra, he worked on Drug Discovery programs for malaria & antibiotics with a team of 5 personnel, facilitated screening of 20,000 compounds for an Anti-Malarial Drug Discovery program and in developing Enzymatic Assays for Anti-Malarial Drug Discovery program and Anti-Bacterial program. At Biocon, he was part of an R and D team that launched insulin and is credited with 2 international and 2 Indian patents. At Goodwin Biotechnology India, Goa, he helped establish a new lab in Goa, India, the subsidiary of Florida-based, Goodwin Biotechnology Inc. He undertook two and half-month training at Goodwin Biotechnology USA in the fall of 2006 to acquire experience in c-GMP.

Teaching Experience: He has worked as lecturer in Dhempe College, Goa and Visiting faculty at BITS Pilani KK Birla Goa campus from December 2009 to December 2014, where he has taught theory courses like General Biology for five semesters, Bioconversion Technology for three semesters and Recombinant DNA Technology for one semester. He was also Contributory Teacher, Goa University Department of Biotechnology teaching theory courses like Food Biotechnology, Environmental Biotechnology, Biosafety and IPR, Bio entrepreneurship, among others and guided four students for M. Sc Dissertation.

## **Brief Biography of the Supervisor**

**Name: Dr Utpal Roy**, Designation: Professor, Department of Biological Sciences

Present Postal Address: DR. UTPAL ROY,

Professor, Department of Biological sciences, Bits Pilani K. K. Birla

Goa Campus, Zuari Nagar, NH 17- B Bye Pass, Goa. 403726.

Educational Qualification: Ph.D. Microbiology, National Dairy Research Institute (NDRI)  
Karnal

Year of award: 1997

(ICAR) Thesis title: “Studies on Genetics and Biochemical Aspects of the Antifungal Substances Produced by Lactic Acid Bacteria.” Supervisor: Dr. Neelakantan Subramaniyan, Senior Principal Scientist; Co-Supervisor: Dr V.K. Batish.

Work Experience: Dr. Utpal Roy has worked in different capacities at BITS Pilani from 1997 to 2003 and BITS Pilani KK Birla Goa campus 2004 to current in different positions starting from Lecturer, Assistant, Associate Professor to his current position of Professor. He has 20 years of teaching and research experience. He has 45 publications in National and International Journals. He has taught courses like Microbiology, Genetic Engineering Technology, Experimental Techniques for both undergraduate and ME Biotechnology students. He was the Group Leader and Head of Department at BITS Pilani KK Birla Goa Campus from 2004 till 2013. He also carried out various administrative roles including as Doctoral Advisory Committee member and Departmental Research Committee (DRC) Chairman of the Department of Biological Sciences. He was instrumental in setting up the Department of Biological Sciences at the newly established BITS Pilani KK Birla Goa campus which now has 18 faculty plus 40 plus Ph.D. Scholars.

## **Brief Biography of the Co-Supervisor**

**Name: Dr Santanu Datta**, Designation: Chief Scientific Officer, Bugworks Research

Present Postal Address: Dr. Santanu Datta,

Centre for Cellular & Molecular Platforms National Centre for Biological Sciences,

TIFR GKVK Campus, Bellary Road, Bangalore 560065.

Educational Qualification: Ph.D. in Biophysics, Calcutta University

Year of award: 1981

Post Doctoral Experience in IISc Bangalore, Baylor College of Medicine Houston, and Karolinska Institute Stockholm. Guided 2 students for Ph.D. completion

Mentored over 50 students for Post graduate dissertation. Official Ph.D. guide of Calcutta University. Adjunct Professor in Pune and Calcutta University

Work Experience:

More than thirty years of experience in molecular biology, genetic engineering, biochemistry and systems biology. Ph.D. examiner and reviewer of papers in international journals. Mentor for setting up SME in the biotech area

Current Role:

CSO of Bugworks Research a SME located in Bangalore. Our focus in the field developing novel antibacterials using *in silico* models, wet lab validations and novel animal models.

Website: <http://www.bugworksresearch.com/>

PI in a FP7- EU grant and a Welcome Trust funded grant on Tuberculosis. More than twenty-five years' experience in *in silico* metabolic pathways modelling and anti-infective drug discovery.

Patents: 14

Publications: 38