

Structural and Functional Studies on VC0395_0300 Protein from *Vibrio cholerae*

THESIS

Submitted in partial fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

Bandekar Divya Ramesh

Under the Supervision of

Dr. Sumit Biswas



BITS Pilani
K. K. Birla Goa Campus



BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2017

CERTIFICATE

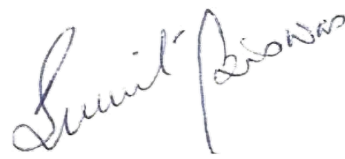
This is to certify that the thesis entitled “Structural and Functional Studies on VC0395_0300 Protein from *Vibrio cholerae*” submitted by Bandekar Divya Ramesh, ID No. 2012PHXF0008G for award of PhD degree of the institute embodies original work done by her under my supervision.

Signature in full of the supervisor:

Name in Capital Block Letters: **Dr. SUMIT BISWAS**

Designation: **ASSISTANT PROFESSOR**

(DEPARTMENT OF BIOLOGICAL SCIENCES)

A handwritten signature in black ink, reading "Sumit Biswas", written in a cursive style.

Date: 06/04/2018

ABSTRACT

The bacterial secondary messenger, c-di-GMP, has been known to play an important role in various signalling pathways affecting biofilm formation, motility, virulence, cell cycle regulation and secondary metabolite synthesis. It modulates a plethora of functions in bacteria at various levels of gene regulation ranging from transcription to translation. The level of c-di-GMP regulates the transition between sessile and motile lifestyles. Besides lifestyle transition, c-di-GMP affects an array of pgv phenotypes like heterocyst formation and lipid metabolism.

The intracellular concentration of c-di-GMP is regulated, in response to environmental stimuli, primarily by the opposing action of diguanylate cyclase (DGC) and phosphodiesterase (PDE). DGC encoded by GGD(/E)EF domain proteins utilise 2 GTP molecules to synthesize c-di-GMP. On the other hand, PDE encoded by EAL domain proteins breakdown c-di-GMP to pGpG or GMP. Moreover, c-di-GMP also acts as an allosteric activator of cellulose synthase and aids in the production of cellulose and thereby, promotes the formation of an extracellular matrix of biofilms. Most of the GGD(/E)EF domains are found in association with other regulatory domains like PAS that help it to identify environmental signals. These regulatory domains are generally located upstream of the GGD(/E)EF domain at the N-terminal. Although GGD(/E)EF domains have been well studied in other organisms, not much is known about its structure and function in *V. cholerae*.

In order to examine the role of GGD(/E)EF domain in *V. cholerae*, the putative protein VC0395_0300 (Sebox3) with GGEEF domain was expressed as recombinant proteins fused to Glutathione-S-Transferase (GST) tag. These recombinant proteins were biophysically characterized to analyze the *in vitro* behaviour of the proteins. *In vitro* diguanylate activity of the proteins was demonstrated by HPLC in presence of GTP. Smaller truncates were also created by shortening the N-terminal of the full-length Sebox3 protein, in order to enhance the stability and yield of the proteins. Consequently, biofilm formation and motility assay of the constructs were also checked and found to exhibit true activity.

Crystallization of the Sebox3 was attempted but didn't yield diffractable crystals. The homology model exhibited that the GGEEF signature sequence was highly conserved in nature. *In silico* analysis showed the mechanism of GTP substrate binding to the active site of GGEEF domain of the Sebox3 protein.

In conclusion, the major aspects of the thesis are as follows:

- Deciphering function and physicochemical features of GGEEF domain protein VC0395_0300 from *V. cholerae*
- Construction and characterization of truncates of Sebox3: Sebox31 and Sebox32
- Elucidation of the function of the truncates with respect to the Sebox3
- Prediction of the structure of the Sebox3 protein

ACKNOWLEDGEMENT

It gives me immense joy to express my gratitude to all those who have helped me in the completion of the thesis. Although the following work is an individual work, it would not have been possible without the efforts of a lot of people.

I would like to express my deep gratitude to my guide, Dr. Sumit Biswas for giving me an opportunity to work in his research group. I thank him immensely for his continuous guidance, help, encouragement and patience at every stage of the research work. His infectious enthusiasm, scientific knowledge and excellent writing skills have deeply influenced me throughout my career at BITS. I am highly indebted to him for all the support over the years.

I would like to thank Prof. Souvik Bhattacharyya (Vice Chancellor, BITS Pilani), Prof. Raghurama G. (Director, BITS, Pilani, K. K. Birla Goa Campus), Prof. S. K. Verma (Dean, Academic Research Ph.D. Programme), Prof. Prasanta Kumar Das (Associate Dean, Academic Research Division), Prof. Sunil Bhand (Dean, Sponsored Research and Consulting) for all the help.

I would like to extend my gratitude to my Doctoral Advisory Committee members, Prof. Dibakar Chakrabarty and Prof. Angshuman Sarkar, Department of Biological Sciences, for their valuable guidance, advice and co-operation throughout the thesis completion. I would also like to thank Prof. Judith Braganca, HOD, Department of Biological Sciences, Prof. Utpal Roy, Convenor and members of the Doctoral Research Committee for their support.

I would like to express my acknowledgement to Dr. Saugata Hazra, Mrs. Mousami Hazra, Dr. Swati Mohapatra, Department of Biotechnology, IIT, Roorkee for their assistance with SEM analysis, CD spectroscopy and ITC analysis. I would like to thank Prof. D. Velmurugan and Dr. K. Gunasekaran, Centre of Advanced Study in Crystallography and Biophysics, Madras University, Chennai for their help and laboratory facilities for protein purification. My sincere gratitude to Dr. Ravindra Makde, Dr. Biplab Ghosh and Dr. Ashwani Kumar for helping me with data collection at Protein Crystallography Beamline (PX-BL-21), Raja Ramanna Centre for Advanced Technology (RRCAT), Indore. I would also like to thank Prof. Pinak Chakrabarti, Bose Institute, Kolkata for extending out his laboratory facilities.

I thank past and current members of our research group - Dr. Ram Kothandan, Om Prakash Chouhan, Subhasish Sahoo and Malvika Sudhakar for their support and friendship. My special thanks to all the past and current research scholars for their timely help especially Dr. Bhakti Salgaonkar, Dr. Kshipra Naik, Dr. Meghnath Prabhu, Dr. Priyadarshini Parekh, Dr. Vilas Desai, Dr. Chandrasekhar Chanda, Dr. Maria Celisa Santimano, Ms. Akriti Rastogi, Mr. Rajesh Pasumarthi, Ms. Geethanjali Ravindran, Ms. Nupur Kumar, Ms. Chitra Nair, Ms. Deepthi Das, Mr. Mansoor Shaik, Mr. Uday Kumar Padidela and Mr. Makhan Kumar.

I owe my deepest gratitude to Ms. Ansie Martin, Ms. Ramya Ramchandran, Mr. Rakesh Manuka and Mr. Ajay Ghosh Chalasani for their unconditional support and for being there for me whenever I needed them. I would like to thank Mrs. Kamna Upadhyay, Mr. Mahadeo Shetkar and Mr. Mahaling Lamani for the timely help during the entire study.

I am thankful to various funding agencies like BRNS for providing Project Fellowship (August 2012 – March 2015), DBT-CTEP for travel grant, Centre for International Cooperation in Science (CICS) support and International Union of Crystallography for Young Scientist Award for attending the 16th International Conference on the Crystallization of Biological Macromolecules, 2016, Prague, Czech Republic and BITS-Pilani, K.K. Birla, Goa Campus for institute fellowship (April 2015 – July 2017).

I am indebted to my parents, family and loved ones for their unconditional love and support and for tolerating me when I was being difficult. I would like to thank my father Mr. Ramesh Bandekar and Mrs. Ranjana Bandekar who trusted me and stood by me through thick and thin. I would like to wholeheartedly thank my sister Deepa for being my pillar of strength and making me realize that there is more to life than bench work. Lastly, I would like to thank the Almighty for giving me the courage to complete this research work successfully.



Bandekar Divya Ramesh

Dedicated to
my beloved parents

CONTENTS

Abstract.....	i
Acknowledgements	iii
List of Figures.....	x
List of Tables	xiii
List of Abbreviations	xiv
List of Symbols	xix

Chapter 1: Introduction and literature review

1.1 Introduction.....	1
1.1.1 Cholera.....	1
1.1.2 Historical record on cholera.....	2
1.1.2.1 Cholera epidemics in Bangladesh	5
1.1.2.2 Haiti cholera outbreak in 2010	5
1.2 Genus <i>Vibrio</i>	6
1.2.1 <i>Vibrio cholerae</i> – the opportunistic pathogen	8
1.2.2 <i>V. cholerae</i> pathogenesis	9
1.2.3 Genetic makeup of <i>V. cholerae</i>	10
1.3 Cholera Toxin (CT)	11
1.4 Current approaches to control/tackle cholera	14
1.4.1 Rehydration therapy.....	14
1.4.2 Antibiotics.....	14
1.4.3 Cholera vaccines	15
1.5 Biofilms - a survival strategy in bacteria	18
1.5.1 Biofilm formation regulation	19
1.5.2 Significance of biofilms	19
1.5.3 Biofilm formation steps	20
1.6 Cyclic di-GMP	21
1.6.1 Synthesis of c-di-GMP	22
1.6.2 Degradation of c-di-GMP	24
1.6.3 Functions of c-di-GMP	24
1.6.3.1 Biofilm formation.....	25
1.6.3.2 Motility.....	26

1.6.3.3 Virulence	27
1.7 GGD(/E)EF domains	29
1.7.1 GGD(/E)EF domain.....	29
1.7.2 EAL domain.....	30
1.7.3 Universality of GGD(/E)EF domain.....	31
1.7.4 Plurality of GGD(/E)EF-EAL domains	32
1.7.5 Coupling of regulatory domains with GGD(/E)EF and EAL domains	33
1.8 VC0395_0300 protein from <i>V. cholerae</i>	34
1.9 Gaps in existing research	35
1.10 Aim of the study	35
1.11 References.....	36

Chapter 2: Biophysical Characterization of Sebox3 Protein

2.1 Introduction.....	48
2.2 Materials and methods	49
2.2.1 Materials	49
2.2.2 Cloning of Sebox3 construct	49
2.2.3 Protein expression.....	51
2.2.4 Protein purification	52
2.2.5 Western Blot Analysis	52
2.2.6 Cleavage of GST tag.....	53
2.2.7 Determination of oligomeric status	53
2.2.7.1 Size exclusion chromatography.....	53
2.2.7.2 HPLC.....	54
2.2.8 Fluorescence spectroscopy	54
2.2.8.1 Thermal denaturation.....	55
2.2.8.2 Chemical denaturation.....	55
2.2.8.3 Quenching of fluorescence	55
2.2.9 HPLC-Diguanylate cyclase reaction	56
2.2.10 Transverse Urea Gel Electrophoresis (TUGE)	57
2.2.11 DTNB (5, 5'-dithio-bis-[2-nitrobenzoic acid]).....	57
2.2.12 Limited proteolysis.....	58

2.3 Results and Discussion	58
2.3.1 Sebox3 cloned into pGEX-6P1 in <i>E. coli</i> DH5a	58
2.3.2 Sebox3 protein expressed in soluble fraction	60
2.3.3 Protein purified as GST-Sebox3 fusion protein.....	61
2.3.4 Removal of GST tag using PreScission protease	63
2.3.5 Tagless Sebox3 protein exists as a monomer in solution	63
2.3.6 Unfolding studies on Sebox3 protein provides information on local environment of Trp residues	65
2.3.6.1 Unfolding studies on Sebox3 protein by thermal denaturation.....	65
2.3.6.2 Guanidine hydrochloride denatured and unfolded the protein.....	66
2.3.6.3 Quenching with acrylamide, KI and CsCl	68
2.3.7 Sebox3 shows diguanylate cyclase activity	70
2.3.8 Transverse urea gel electrophoresis (TUGE)	71
2.3.9 Detection of free thiol in Sebox3 protein	72
2.3.10 Stable domains appear on treatment with proteolytic enzymes	73
2.4 Conclusion	74
2.5 References.....	76

Chapter 3: Construction of Truncates of Sebox3: Sebox31 and Sebox32; and their Functional Characterization

3.1 Introduction.....	80
3.2 Materials and Methods.....	81
3.2.1 Domain organization	81
3.2.2 Plasmid construction and cloning	82
3.2.3 Protein expression, purification and GST tag removal.....	83
3.2.4 Assays for formation of biofilms	83
3.2.4.1 Crystal Violet Assays	83
3.2.4.2 Scanning Electron Microscopy (SEM).....	84
3.2.5 Congo Red Assay	85
3.2.5.1 Congo Red uptake by cells	85
3.2.5.2 Quantitative estimation of Congo Red Binding	85
3.2.6 Motility Assays	85
3.2.6.1 Soft agar plate assay	85

3.2.6.2 TTC assay	86
3.3 Results and Discussion	86
3.3.1 Domain Organization	86
3.3.2 Secondary structure prediction and protein solvation	87
3.3.3 Cloning, expression and purification of Sebox31 and Sebox32	98
3.3.4 Biofilm formation by Sebox proteins	99
3.3.4.1 Crystal Violet assay	99
3.3.4.2 SEM of biofilms	100
3.3.5 Congo Red binding	104
3.3.6 Inverse effect of diguanylate on motility	105
3.3.6.1 Motility on plates	106
3.3.6.2 TTC assay	106
3.4 Conclusion	107
3.5 References	109

Chapter 4: Comparative characterization of Sebox3 versus Sebox31 and Sebox32

4.1 Introduction	113
4.2 Methods	114
4.2.1 Diguanylate cyclase assays	114
4.2.2 Circular Dichroism (CD) spectroscopy	114
4.2.3 Fluorescence spectroscopy	114
4.2.3.1 Chemical denaturation	115
4.2.3.2 Quenching of fluorescence	115
4.2.4 Analytical gel filtration chromatography	115
4.3 Results and Discussion	115
4.3.1 Diguanylate cyclase activity of N-terminal truncates	115
4.3.2 Comparison of secondary structure of constructs	117
4.3.3 Thermal stability and binding of ligand	119
4.3.4 Oligomeric status of constructs	121
4.3.5 Quenching of native and denatured Sebox31 protein and Sebox32 proteins	122
4.4 Conclusion	124
4.5 References	125

Chapter 5: The Structure of Sebox3 protein

5.1 Introduction.....	128
5.2 Materials and Methods.....	130
5.2.1 Protein expression and concentration.....	130
5.2.2 Crystallization trials	131
5.2.3 Diffraction at Synchrotron beamline.....	136
5.2.4 Homology modelling of Sebox3	136
5.2.5 Molecular docking of GTP at active site of Sebox3	137
5.3 Results and Discussion	138
5.3.1 Concentration of Sebox3 protein	138
5.3.2 Crystallization of Sebox3 protein.....	139
5.3.3 Weak diffraction of Sebox3 crystals	140
5.3.4 Alignment of Sebox3 with other diguanylate cyclases	141
5.3.5 Homology model of Sebox3	143
5.3.6 Docking of the GTP molecule to the active site of Sebox3	146
5.4 Conclusion	147
5.5 References.....	148
Chapter Conclusions and Summary	151
Future plan of work.....	153
Appendices.....	154
A Reagents preparation.....	154
B List of license numbers	156
C List of Publications and Conference Presentations	157
D List of Workshops attended	158
E Biography of the Candidate	159
F Biography of the Supervisor.....	160

List of Figures

1.1	Scanning Electron Microscopy (SEM) image of <i>V. cholerae</i> bacterium showing monotrichous flagellum and pili bundle.....	1
1.2	The world map by World Health Organization shows the geographical distribution of cholera outbreaks between 2010 and 2015 indicating the persistence of the disease predominantly in the Indian sub-continent and African countries.....	4
1.3	The Haiti cholera outbreak followed by the massive earthquake killed at least 250 people and sickened more than 3,000 people	6
1.4	Classification of the toxinogenic <i>Vibrio cholerae</i> based on its antigens.....	8
1.5	<i>V. cholerae</i> life cycle in aquatic bodies as well as in the human body	9
1.6	The large chromosome I and small chromosome II in <i>Vibrio cholerae</i>	11
1.7	Structure of cholera toxin (CT) showing the A subunit with the five B subunits....	12
1.8	Mode of action of cholera toxin (CT).....	13
1.9	Commercially available oral vaccine manufactured by Shantha Biotech Company	16
1.10	Smooth and Rugose morphology exhibited by different strains of <i>V. cholerae</i>	19
1.11	Steps involved in biofilm formation and the role of c-di-GMP in its regulation.....	20
1.12	Synthesis of c-di-GMP from two GTP molecules	23
1.13	Role of c-di-GMP binding effectors at different levels gene transcription	25
1.14	The regulatory network leading to exopolysaccharide secretion and biofilm formation in <i>Salmonella</i>	26
1.15	The concentration of the intracellular c-di-GMP determines the upregulation and downregulation of the motility and biofilm	27
1.16	Role of diffusible signaling factor (DSF) from <i>Xanthomonas campestris</i> in regulation of virulence	28
1.17	Crystal structure of GGDEF domain from <i>Thermotoga maritima</i>	29
1.18	The GGDEF domain is omnipresent in different bacteria	32
2.1	Design of the plasmid map	50
2.2	Cell viability of <i>E. coli</i> host cells due to toxicity of GGD(E)EF protein	59
2.3	SDS-PAGE profile of pilot scale expression of VC0395_0300 in <i>E.coli</i>	59
2.4	Expression profile of GST-tagged protein.....	62

2.5	SDS-PAGE profile of tagless Sebox3 protein	63
2.6	Analytical size exclusion chromatography for determination of molecular weight of Sebox3	64
2.7	Thermal denaturation ranging from 20°C-90°C for Sebox3 protein	66
2.8	Guanidine hydrochloride-induced unfolding by utilizing Gdn.HCl concentrations ranging from 0 M – 6 M	67
2.9	Stern - Volmer plots for quenching of tryptophan fluorescence of Sebox3 in acrylamide, KI and CsCl.....	69
2.10	The Sebox3 protein with active GGEEF domain is a diguanylate cyclase.....	70
2.11	Transverse Urea Gel Electrophoresis (TUGE) of native Sebox3 protein	71
2.12	Limited proteolysis of Sebox3 protein	74
3.1	Interproscan results showing GGDEF domain in the sequence of Sebox3 protein with PAS sensory domain at the N-terminal	87
3.2	Consensus secondary structure prediction	89
3.3	Protein Solvation properties of the entire amino acid sequence of the Sebox3 protein shows the position of buried and exposed amino acid	95
3.4	PCR amplification of sebox31 and sebox32 on 1% agarose gel	98
3.5	Crystal Violet staining of the strains for biofilm estimation.....	99
3.6	SEM images of biofilms	101
3.7	Congo Red Assay: <i>E.coli</i> strains with Sebox3, Sebox31, Sebox32, QrgB and pGEX-6P1	104
3.8	Motility on soft agar of Sebox3, Sebox31, Sebox32, QrgB and pGEX-6P1.....	106
3.9	Motility in tetrazolium tubes.....	107
4.1	Diguanylate cyclase activity of Sebox proteins in HPLC.....	116
4.2	CD spectrum of the protein, protein with GTP and Protein with c-di-GMP	118
4.3	Thermal shift assay using CD spectra of the protein from 20°C to 90°C, protein with GTP and protein with c-di-GMP	120
4.4	Analytical gel filtration of Sebox31 and Sebox32 proteins along with standard protein markers carried out using a Superdex200 column in GE Healthcare AKTA Prime system.....	121
4.5	Quenching of Sebox proteins with acrylyamide, CsCl amd KI.....	122
5.1	Phase diagram showing steps involved in protein crystallization	129
5.2	SDS – PAGE showing retentate and filtrate collected at different stages of Sebox3 protein concentration	138

5.3	The images of different conditions in hanging drop vapour diffusion method	139
5.4	Image of Sebox3 protein crystals.....	140
5.5	Diffraction image obtained for Sebox3 with the detector at 110 mm.....	141
5.6	Multiple Sequence Alignment by ClustalW showing conserved amino acids and secondary structure.....	142
5.7	Remarkable structural similarities between the model and the template.....	144
5.8	Homology model of the C terminal domain of Sebox3	145
5.9	Ramachandran Plot of the modelled structure of the C terminal domain of Sebox3	146
5.10	Docking of GTP against the active site of Sebox3	147

List of Tables

1.1	Different species of <i>Vibrio</i> associated with human diseases	7
1.2	Summary of the main types of cholera vaccines	18
1.3	Historical timeline of discovery of c-di-GMP and its regulatory elements	22
2.1	The sequence of primers used for PCR amplification	51
2.2	List of cysteine residues in the Sebox3 protein	73
3.1	List of strains used in the study.....	81
3.2	Congo Red uptake from Standard solution (40 µg/ml) is measured at 490nm.....	104
5.1	Different concentrations of PEG 8000, PEG 6000 and Ammonium sulphate used for checking zone of protein precipitation.....	132
5.2	Conditions from Hampton Research Crystal Screen Formulations were also used to determine the ideal condition for protein crystal formation.....	133
5.3	The initial grid design of a 24 – well culture plate for Sebox3 protein crystallization with buffers like citrate, MES, MOPS, HEPES and Tris – HCl ranging from pH 4.0 – pH 8.0 were used.....	134
5.4	The grid design of a 24 – well culture plate for Sebox3 protein crystallization.....	135

List of Abbreviations

AC: Accessible Cysteine

ADP: Adenosine diphosphate

ADT: AutoDock Tool

AMS: ammonium sulfate

Arnt: Aryl hydrocarbon receptor nuclear translocator protein

BSA: Bovine Serum Albumin

cAMP: Cyclic adenosine monophosphate

CD: Circular Dichroism

c-di-GMP: bis-(3'-5')-cyclic dimeric guanosine monophosphate

CPD: Critical Point Dryer

CR: Congo Red

CT: Cholera Toxin

CTAB: Cetyl Trimethyl Ammonium Bromide

CtxB: B subunit of the cholera toxin

DGC: diguanylate cyclase

DLS: Dynamic Light Scattering

DSF: Diffusible Signaling Factor

DTNB: 5, 5'-dithio-bis-[2-nitrobenzoic acid]

DTT: Dithiothreitol

EAL: Glutamic acid – Alanine – Luecine

ECL: Enhanced Chemiluminescence

EDTA: Ethylenaminetetraaceticacid

EPS: extracellular polysaccharides

ESPrict: Easy Sequencing in Postscript

ExPASy: Expert Protein Analysis System

FE SEM: Field Emission Scanning Electron Microscope

GAF: cGMP-specific phosphodiesterases, adenylates, FhlA

GdnHCl: Guanidine hydrochloride

GGD(/E)EF: Glycine – Glycine – Aspartic Acid – (/Glutamic acid) – Glutamic Acid –

GM1: Monosialotetrahexosylganglioside

GMP: Guanosine monophosphate

GROMACS: Groningen Machine for Chemical Simulations

GST: Glutathione-S-Transferase

GTP: Guanosine-5'-triphosphate

HAMP: Histidine kinases, Adenyl cyclases, Methyl-accepting proteins and Phosphatases

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HgR: Mercury Resistance

HIV: Human Immunodeficiency Virus

HPLC: High Pressure Liquid Chromatography

HTH: helix-turn-helix DNA-binding domain

I site: Inhibitory site

IDT: Integrated DNA technologies

IPTG: Isopropyl β -D-1-thiogalactopyranoside

IR: Isomorphous Replacement

ITC: Isothermal Titration Calorimetry

IV: Intravenous

kV: kilovolt

LB: Luria Bertani

MAD: Multi-wavelength Anomalous Dispersion

MES: 2-(N-morpholino)ethanesulfonic acid

MOPS: 3-(N-morpholino)propanesulfonic acid

MR: Molecular Replacement

NCBI: National Centre for Biotechnology Information

NMR: Nuclear Magnetic Resonance

NMWCO: Nominal Molecular Weight Cut Off

OD: Optical Density

ORS: Oral Rehydration Solution

PAGE: Polyacrylamide Gel Electrophoresis

PAS: Per – Arnt – Sim

PBS: Phosphate – Buffered Saline

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PDE: phosphodiesterase

PEG: polyethylene glycol

Per: Period circardian protein

pGpG: 5'-phosphoguanlyl-(3'-5')- guanosine

Phenylalanine

PPi: Inorganic Pyrophosphate

PSI-BLAST: Position-Specific Iterative Basic Local Alignment Search Tool

rBS: recombinant B Subunit

RCSB: Research Collaboratory for Structural Bioinformatics

REC: Receiver domain

RRCAT: Raja Ramanna Centre for Advanced Technology

RT: Retention Time

SAD: Single-wavelength Anomalous Dispersion

SDS: Sodium dodecyl sulfate

SEM: Scanning Electron Microscopy

Sim: Single-minded protein

SMART: Simple Modular Architecture Research Tool

SPR: Surface Plasma Resonance

TCP: Toxin-coregulated pilus

TE: Tris EDTA

TEMED: Tetramethylethylenediamine

TFA: Trifluoroacetic acid

TMP – SMX: trimethoprim sulfamethoxazole

TNB: 2-nitro-5-mercaptopbenzoic acid

TTC: Triphenyl Tetrazolium Chloride

TUGE: Transverse Urea Gel Electrophoresis

UV: Ultraviolet

VPI: Vibrio Pathogenicity Island

w/v: weight/volume

WC: Whole Cell

WHO: World Health Organization

List of Symbols

K_{cat} : 1st order rate constant

R_f : Retention factor

ϵ : Molar Extinction Coefficient

CHAPTER 1

Introduction and Literature Review

1.1 Introduction

1.1.1 Cholera

Cholera is a deadly diarrhoeal disease caused by a Gram-negative bacterium *Vibrio cholerae* (Figure 1.1). This water-borne disease, transmitted by the fecal-oral route is endemic in southern Asia, parts of Africa and Latin America. In such places, seasonal outbreaks are common, owing to poverty and poor sanitation conditions. For assessment of public health significance, two vital properties of *V. cholerae* are considered. These include cholera toxin (CT) production and possession of O1 or O139 antigen. Cholera toxin is one of the main reasons for causing diarrhoea and therefore, it has also been used as a molecular marker for detection of the virulent form of *V. cholerae* [1]. Apart from CT, the presence of toxin-coregulated pilus (TCP) and ToxR, a regulatory protein which controls both CT and TCP expression, plays a vital role in causing cholera. [2]. Therefore, it can be said that cholera pathogenesis involves the combined effect of a number of pathogenic factors by toxigenic *V. cholerae*.

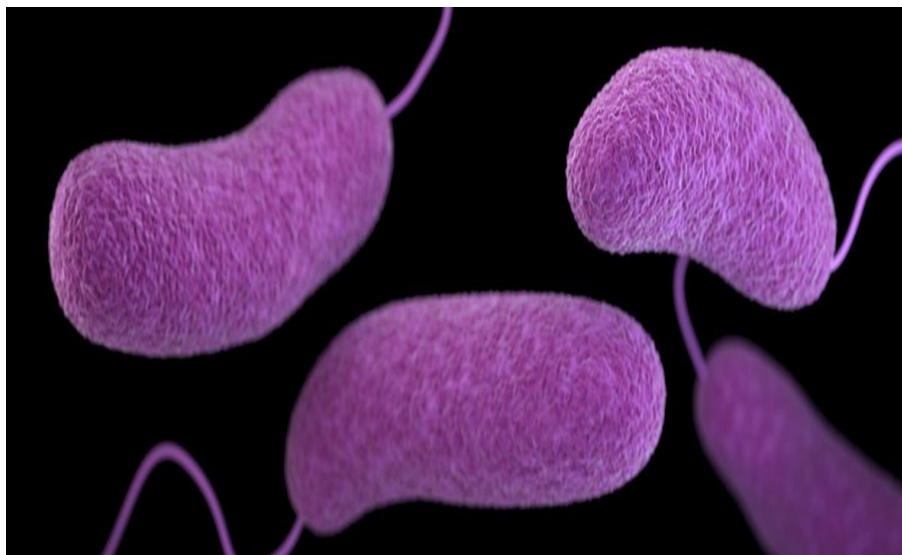


Figure 1.1: Scanning Electron Microscopy (SEM) image of *V. cholerae* showing bacterium, monotrichous flagellum and pili bundle.

Source: <https://www.cdc.gov/vibrio/images/vibrio.jpg>

1.1.2 Historical record on cholera

The word cholera is derived from the Greek word 'khole' meaning bile. This disease which was earlier endemic, is now capable of causing epidemic and pandemics. One of the earliest records about cholera has been from Hippocrates and Galen who described a malady that had symptoms similar to cholera.

In the Indian subcontinent, cholera has been an endemic disease for centuries with the first cholera pandemic occurring in 1817 outside the Indian subcontinent and reaching as far as Europe through the various trade routes during the early 1830s [3]. Eventually, cholera spread from the Indian subcontinent to almost the entire world causing 7 pandemics in the past 185 years [4]–[6].

a) First pandemic of cholera (1817 – 1823)

It originated in 1817 in River Ganga, India and spread to Southeast Asia, Eastern Africa and the Mediterranean coast by trade routes and colonization. By 1820, cholera had spread to Java, Indonesia with a death toll of 100, 000.

b) Second pandemic of cholera (1829-1849)

This pandemic began in India spreading to Poland and Finland through Russia. In 1830; 22,000 people died of cholera in England. Between 1832 and 1833, the disease had spread from Europe to North America. By 1848, another 52,000 people were killed due to another cholera outbreak in England.

c) Third pandemic of cholera (1852 – 1859)

This pandemic originated in India and affected large parts of Asia, Europe, Africa and North America claiming at least 23,000 lives in England alone.

In 1854, John Snow, an English physician was the first to show that contaminated water was the source of a cholera epidemic in London. Due to his pioneering work in cholera epidemiology, he showed that cholera transmission could be stopped by simply preventing the source of contaminated drinking water.

d) Fourth pandemic of cholera (1863 – 1879)

This pandemic again originated in India and was carried by the Indian pilgrims travelling to Mecca affecting around 30,000 in the Middle East. Around the same time, approximately 90,000 people were killed in Russia due to cholera.

e) Fifth pandemic of cholera (1881 – 1896)

This time also the source was the Bengal region of India and cholera spread across Asia, parts of Europe, Africa and South America. By 1883, Robert Koch had identified *V. cholerae* as the causal organism of cholera during an outbreak in Egypt. In 1892, Waldemar Haffkine had developed a cholera vaccine.

f) Sixth pandemic of cholera (1899 – 1923)

This pandemic claimed 800,000 lives in India and by 1923, cholera cases had reduced with incidences restricted to the Indian subcontinent.

g) Seventh pandemic of cholera (1961 – 1975)

This pandemic originated in Indonesia and unlike the previous six pandemics, it was caused by the El Tor strain and not the classical strain.

However, the exact mechanism by which *V. cholerae* causes diarrhoea was not known. It was Robert Koch again who first suggested that the symptoms of cholera were due to a ‘poison’ from *V. cholerae*.

It was only in 1959, that Shambhu Nath De discovered this poison - the Cholera Toxin (CT) which binds to the GM1 ganglioside of the small intestine of the infected individual and causes acute diarrhoea [7], [8]. In the same year, Dutta and colleagues demonstrated that cell-free *V. cholerae* culture filtrate was capable of eliciting watery stools in infant rabbits, thereby supporting the discovery made by S.N. De [9].

In 1992, a new strain of *V. cholerae* O139 (Bengal) emerged leading to endemics in at least 10 countries. In the recent years, development of new strains and climate change has caused cholera outbreaks in different parts of the world with the notable ones being in Bangladesh and Haiti (Figure 1.2).

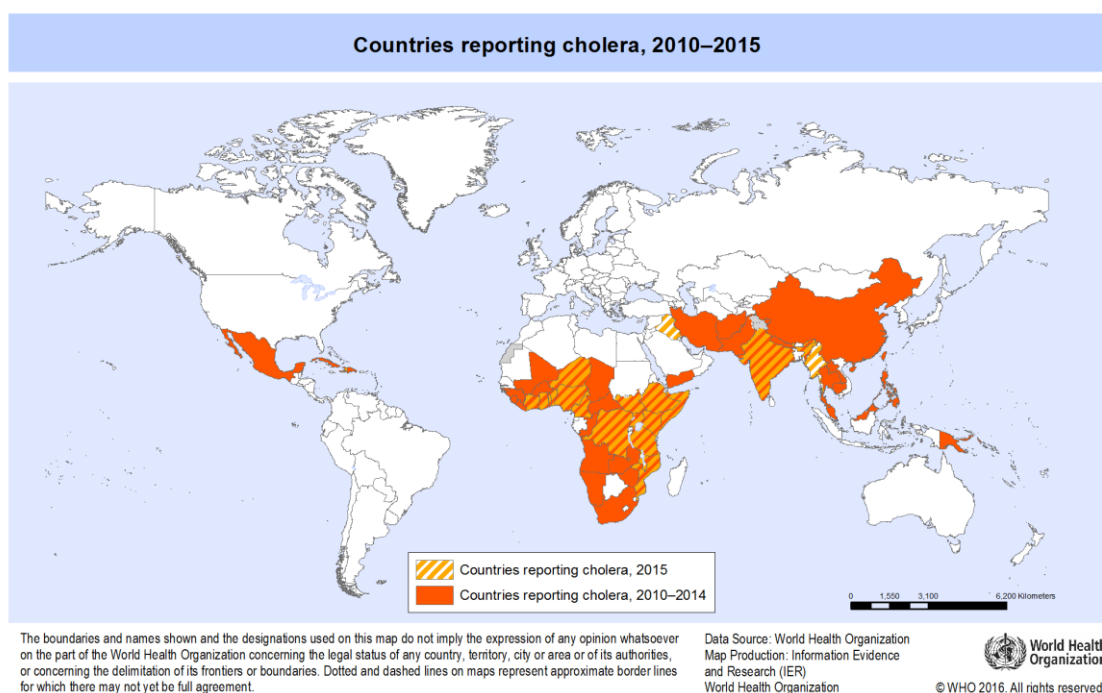


Figure 1.2: The world map shows the geographical distribution of cholera outbreaks between 2010 and 2015 indicating the persistence of the disease predominantly in the Indian sub-continent and African countries (World Health Organization).

Source: http://gamapsserver.who.int/mapLibrary/Files/Maps/Global_Cholera_2010_2015.png

1.1.2.1 Cholera epidemics in Bangladesh

In Bangladesh, at least two incidences of cholera are seen every year. Cholera epidemics in Bangladesh have been associated with cyclones, floods, natural calamities and also with the biannual seasonal change. The 2007 cholera outbreak saw the most severe dehydrating diarrhoeal outbreak as compared to the previous flood-associated 2003 and 1998 epidemics. Most of the cases were associated with the *V. cholerae* non-O1 strain, also called the Bengal O139 strain, which had lead to the severe watery stools in afflicted patients [10]. Along with O1, a non-O1 serotype O139 Bengal also caused major outbreaks in India and Bangladesh. Cholera epidemics caused by *V. cholerae* O139 serotype are more prominent during the warmer seasons of the year [11], [12].

1.1.2.2 Haiti cholera outbreak in 2010

The 2010 earthquake-associated cholera outbreak led to the death of 4672 reported deaths and hospitalization of thousands of people. After the major earthquake on 12th January 2010, thousands of people got displaced and due to lack of potable water, people, especially from rural areas, were forced to consume river water. This lead to the spread of the disease in rural Artibonite, 100 km north of the capital of Port-au-Prince (Figure 1.3) [13].

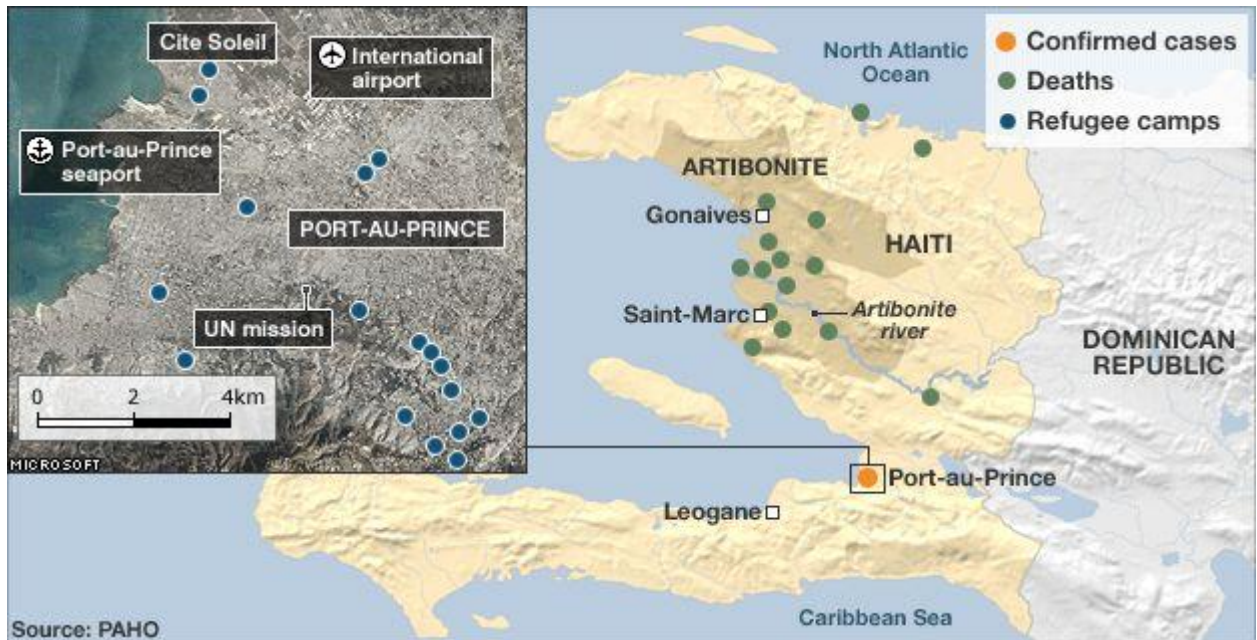


Figure 1.3: The Haiti cholera outbreak followed by the massive earthquake killed at least 250 people and sickened more than 3,000 people.

Source: https://i.chef1.bbci.co.uk/news/624/media/images/49647000/gif/_49647326_haiti_cholera_624map.gif

1.2 Genus *Vibrio*

The genus name *Vibrio* is based on the fact that the organism appears as if it is vibrating when it moves. There are several species in the genus *Vibrio*, of which *V. cholerae*, *V. vulnificus* and *V. parahaemolyticus* are the most important pathogens for humans causing diarrhoeal illnesses (Table 1.1). Infections caused by *V. vulnificus* vary from mild gastroenteritis and wound infections to severe necrotic infections and fatal septicaemia [14]

Table 1.1: Different species of *Vibrio* associated with human diseases

Species of <i>Vibrio</i>	Source of Infection	Diseases caused in humans
<i>V. alginolyticus</i>	Seawater	Wound infection, external otitis
<i>V. cholerae</i>	Water, food	Gastroenteritis
<i>V. cincinnatiensis</i>	Unknown	Bacteremia, meningitis
<i>V. fluvialis</i>	Seafood	Gastroenteritis, wound infection and bacteremia
<i>V. furnissii</i>	Seawater	Gastroenteritis
<i>V. harveyi</i>	Seawater	Wound infection
<i>V. metschnikovii</i>	Infected animals	Bacteremia
<i>V. mimicus</i>	Fresh water	Gastroenteritis, wound infection and bacteremia
<i>V. parahaemolyticus</i>	Shellfish, seawater	Gastroenteritis, wound infection and bacteremia
<i>V. vulnificus</i>	Shellfish, seawater	Bacteremia, wound infection and cellulitis

Among all the species of *Vibrio*, most of the diarrhoeal diseases due to the production of heat-stable like toxin and cholera toxin has been attributed to *V. mimicus*, *V. cholerae* and *V. parahaemolyticus* [15], [16]. There are a couple of unknown *Vibrios* that have been isolated from marine water and marine organisms and are confused with those that infect humans.

1.2.1 *Vibrio cholerae* - the opportunistic pathogen

V. cholerae is a Gram-negative, comma-shaped, motile bacillus with a single polar flagellum. It is a natural inhabitant of aquatic ecosystems and is well known for forming biofilms on abiotic surfaces [17]. This ability has enabled the organism to survive in between cholera outbreaks in both freshwater and marine environments.

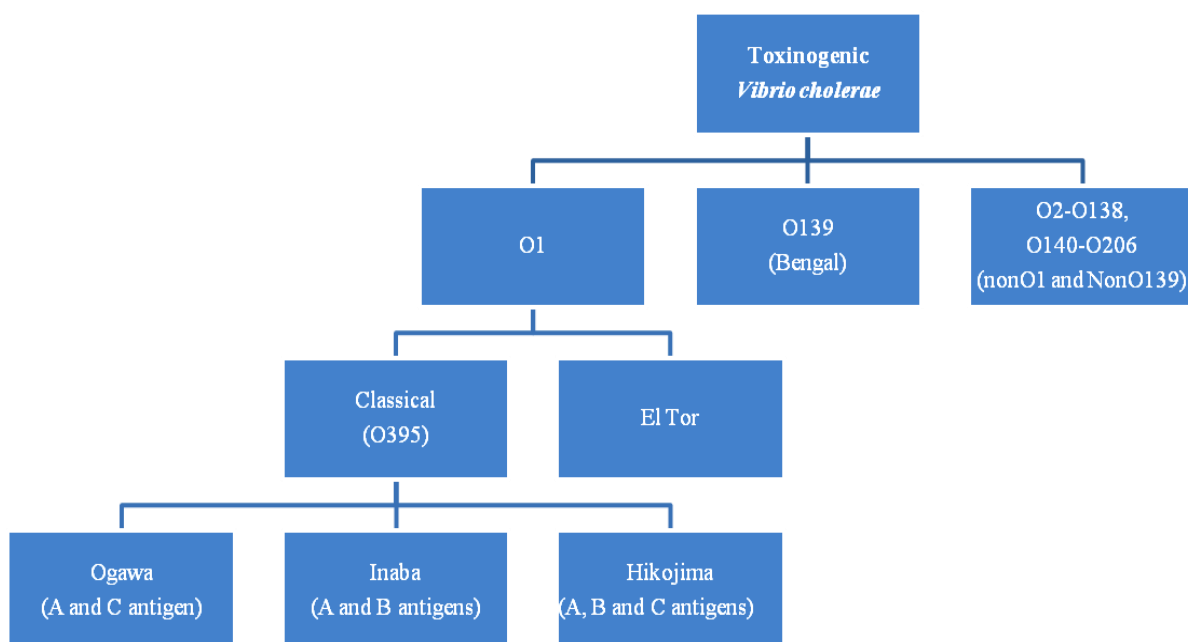


Figure 1.4: Classification of the toxinogenic *Vibrio cholerae* based on its antigens.

Based on the presence of somatic antigens (O antigen), *V. cholerae* is classified into serovars with at least 155 known serovars. Amongst these, O1 and O139 serovars are known to colonize and cause majority of the cholera epidemics. The *V. cholerae* O1 can be further classified into 2 biotypes based on susceptibility to bacteriophages and biochemical characterization, that is, classical and El Tor. Each of these, in turn, have two major serotypes, Ogawa and Inaba, and the Hikojima serotype which has been rarely reported (Figure 1.4) [18]–[20].

Moreover, there are at least 154 known serotypes of non-O1 vibrios. These vibrios are very similar to the O1 vibrios but do not agglutinate with polyvalent O:1 antiserum. Non-O1 serotypes of *V. cholerae* gained significance only after the 1993 cholera outbreak in Bangladesh and India caused by the O139 Bengal strain.

1.2.2 *V. cholerae* pathogenesis

V. cholerae completes its life cycle in water bodies as well as in the human intestine (Figure 1.5). It enters into the human body through the fecal-oral route by consumption of contaminated food or water. Cholera is also transmitted by eating uncooked raw fish or shellfish. Most of the consumed organisms will get killed due to the acidic conditions in the stomach while, a few that manage to survive, colonize and thrive on the mucosal cells of the small intestine with the help of the toxin co-regulated pilus (TCP). The *tcpA* gene for TCP, a type IV pilus, is located in the *Vibrio* pathogenicity island (VPI). Apart from the *tcpA* gene, the complete assembly of TCP is dependent on gene products of *tcp* and *acf* gene clusters located on TCP pathogenicity [21].

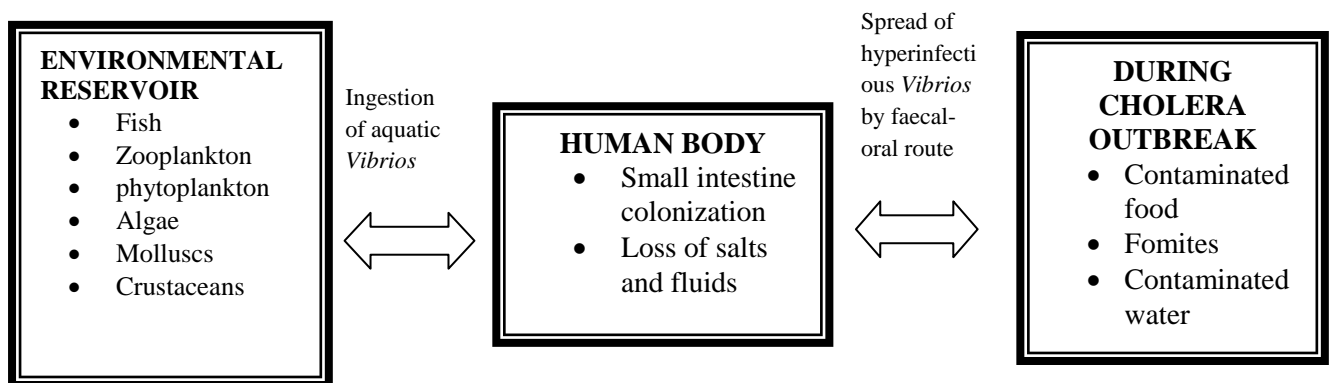


Figure 1.5: *V. cholerae* life cycle in aquatic bodies as well as in the human body.

1.2.3 Genetic Makeup of *V. cholerae*

Unlike other bacteria, the entire genome of *V. cholerae* is divided into two chromosomes: the larger chromosome I (2.96 Mb) and smaller chromosome II (1.07 Mb) [22]. The chromosome I carries majority of the genes required for survival of the organism like DNA replication, transcription, translation and cell wall synthesis and; pathogenesis genes like toxins, adhesions and surface antigens. Whereas, the chromosome II carries a large number of hypothetical genes, a gene capture system (the integron island) and host addiction genes. It is believed that the small chromosome may have existed as a megaplasmid and was engulfed by the ancestral *Vibrio* species (Figure 1.6) [23]. This distribution of genes on two chromosomes enables the organism to survive in various environmental conditions and adapt to different lifestyles. Due to differential expression of genes on the two chromosomes, *V. cholerae* can live in aquatic reservoirs as well as in human hosts. The expression of genes on the chromosome II is enhanced during colonization of colon as compared to when the organism is growing in aerobic conditions [24].

Apart from these genes, all pathogenic strains of *V. cholerae* have the virulence gene cluster called the *Vibrio* pathogenicity island (VPI). The VPI consists of a phage-like integrase gene (*int*), phage-like attachment (*att*) sites and a transposase-like gene (*vpiT*). The VPI is inserted site specifically in the pathogenic *V. cholerae* downstream of a tRNA-like locus (*ssrA*).

In classical and El Tor *V. cholerae*, the VPI encodes for virulent proteins like the TcpP, TcpH and ToxT.

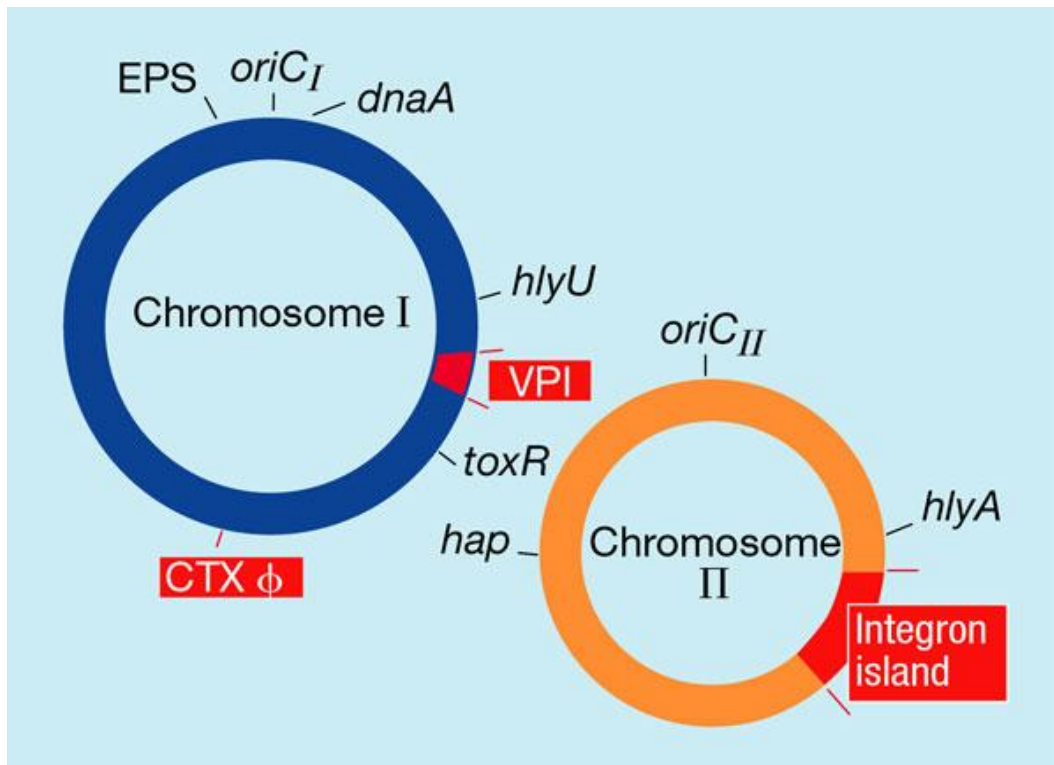


Figure 1.6: The large chromosome I and small chromosome II in *Vibrio cholerae*. EPS – exopolysaccharide, VPI – *Vibrio* pathogenicity island.

Source: Waldor et al, 2000, Reproduced with permission under the Nature Publishing Group License.

1.3 Cholera toxin (CT)

The cholera toxin (CT) is a potent enterotoxin and mostly produced in association with TCP. This toxin is primarily responsible for the rice water stools. The production of cholera toxin depends on the *ctxA* and *ctxB* genes located on the lysogenic phage genome (CTX Φ) and not on the *V. cholerae* chromosomal genome. CT protein is made up of a large centrally located A subunit and 5 smaller identical B subunits at the periphery (Figure 1.7) [25], [26].

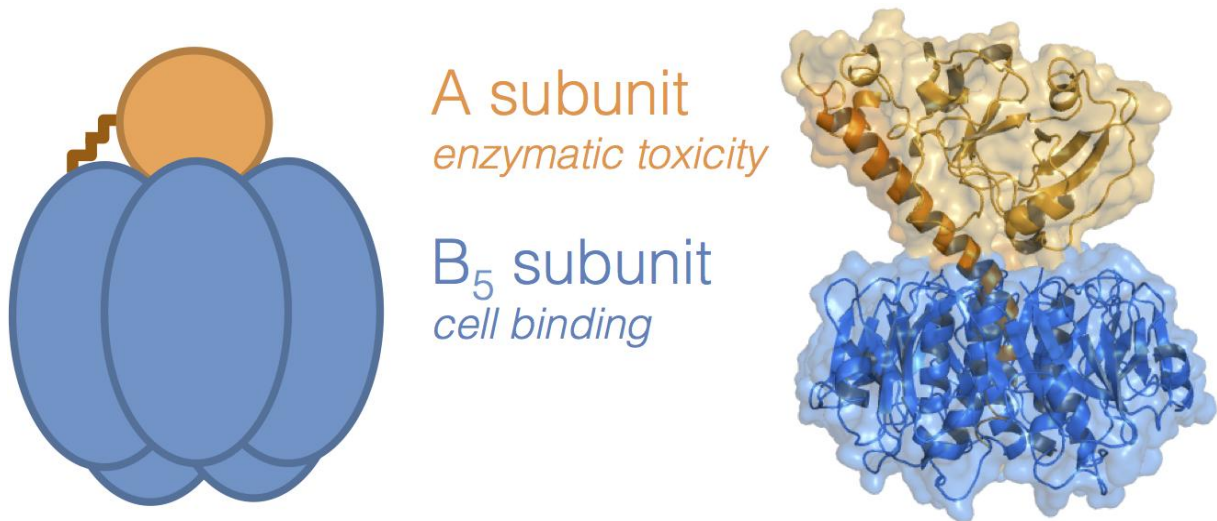


Figure 1.7: Structure of cholera toxin (CT) showing the A subunit with the five B subunits.

Source: <https://sites.tufts.edu/quorumsensing/files/2014/11/structureofcholera toxin1.png>

The B subunit binds the toxin to the eukaryotic cell receptor. The A subunit of CT increases intracellular cAMP (cyclic adenosine monophosphate) concentration eventually causing diarrhoea.

Consequently, the A subunit leads to ADP-ribosylation of G protein, and constitutive activation of adenylate cyclase raising the cAMP levels and increasing the accumulation of excessive water and electrolytes into the lumen of the small intestine (Figure 1.8). This would eventually lead to loss of body fluids and electrolytes via profuse watery diarrhoea and vomiting causing excessive dehydration, low blood pressure and muscle cramps. Some patients may also experience acute renal failure and coma leading to death. The patient can die within a few hours if the water and electrolytes are not replaced immediately [17], [18].

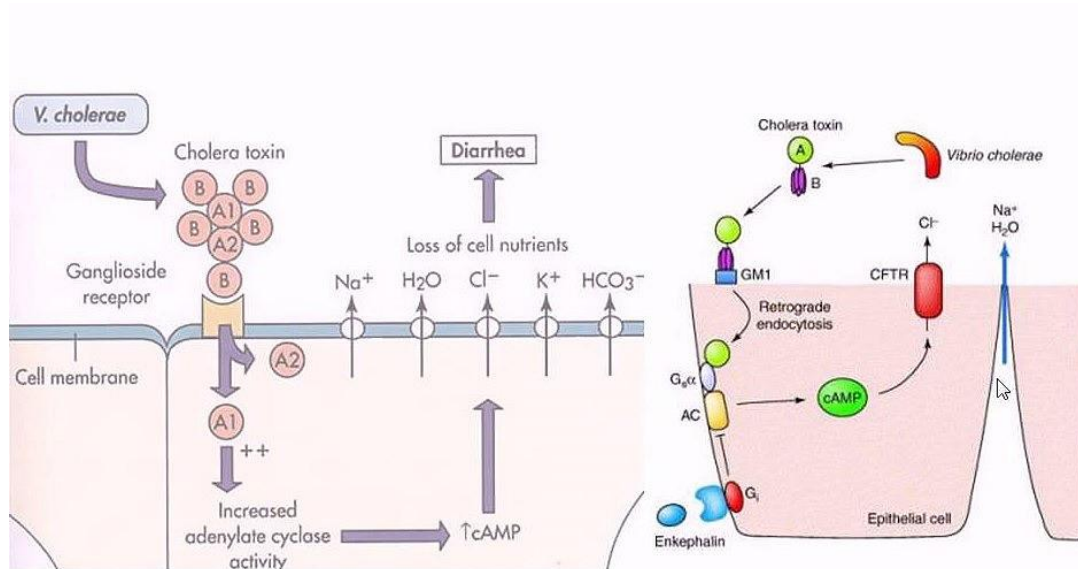


Figure 1.8: Mode of action of cholera toxin (CT).

Source:http://getlinkyoutube.com/wpcontent/uploads/2016/10/1476436570_maxresdefault.jpg

Replenishment of body fluids by Oral Rehydration Solution (ORS) is the only way to prevent dehydration. However, Intravenous (IV) fluid replacement therapy may be necessitated if the patient's condition is critical. Ultimately, the body fluids lost from the patient's body aids the organism to move from one individual to another through the aquatic environment [26]. Within aquatic systems, *V. cholerae* picks up cues from the surrounding environment and regulates differential expression of genes. This, in turn, is dependent upon the neighbouring populations of *V. cholerae* which by quorum sensing, allows the organism to flourish by biofilm formation. Quorum sensing is cell-population dependent gene regulation process and is achieved by detection of extracellular signalling molecules [27].

1.4 Current approaches to control cholera

As cholera is transmitted by the fecal-oral route, the best way to prevent it would be to educate people on proper sanitation and maintenance of food and water hygiene. Potable clean drinking water free of seepage from septic tanks or fecal matter would ensure that the disease is contained.

1.4.1 Rehydration therapy

The 1st step towards treatment of cholera is to rehydrate the affected individual either by oral rehydration solution (ORS) or by intravenous drips to maintain the functionality of vital organs. According to World Health Organization (WHO), ORS is composed of glucose, sodium citrate and salts like sodium chloride and potassium chloride.

1.4.2 Antibiotics

Antibiotics are prescribed, as the 2nd line of treatment, to patients who have lost a lot of fluids during the rehydration therapy. In most cases, antibiotics like doxycycline and azithromycin are preferred. Treatment with tetracycline has been more effective than furazolidone, chloramphenicol and sulfaguanidine. However, a single dose of 300 mg of doxycycline has shown the same effect as the treatment with tetracycline for 3 days. Apart from doxycycline, other antibiotics like erythromycin, ciprofloxacin and trimethoprim-sulfamethoxazole (TMP-SMX) have also been administered [28], [29]. Most of these might reduce the scale of infection but are highly ineffective at completely eradicating the disease. Over the years, due to the rampant use of antibiotics during cholera outbreaks, *V. cholerae* has developed resistance to majority of the antibiotics. The acquired antibiotic resistance can be attributed to the accumulation of mutations, introns, plasmids and conjugative elements, and is mainly seen in areas where cholera is endemic [30]–[32].

1.4.3 Cholera vaccines

Cholera vaccine is generally recommended for people travelling to places where cholera is endemic. Depending on the type of vaccine, it provides protection ranging from a few days to few months. These cholera vaccines are of the following 3 three types:

- a) parenteral phenol - inactivated vaccine (Dukoral)
- b) oral killed whole cell - cholera toxin recombinant B subunit vaccine (WC-rBS)
- c) oral live attenuated *Vibrio cholerae* vaccine (CVD 103-HgR)

a) Parenteral vaccines:

Parenteral cholera vaccines are killed or modified whole cells, antigens or cholera subunit that is administered intravenously, intramuscular or subcutaneously into the body. An example of a parenteral vaccine is, Dukoral which is an oral suspension of inactivated *V. cholerae* and a non-toxic component of the cholera toxin. This vaccine made from phenol-killed O1 organisms (classical and El Tor), provides protection (more than 50%) only for 3-6 months against cholera caused by *V. cholera* O1 only and does not cover cholera caused by *V. cholerae* O139.

If extended protection is required, then a booster dose needs to be administered after every 6 months. Also, the efficacy of the vaccine is low in children and will only enhance already existing immune responses. Nonetheless, cholera vaccine has a number of side effects like nausea, fever, headache as well as pain and inflammation at the site of injection [33]. Due to so many adverse effects, it cannot be given to infants and pregnant women. There is also a possibility of suffering an allergic reaction after administration of the vaccine. And most importantly, these vaccines are not very effective in curbing cholera outbreaks and do not disrupt the transmission of *V. cholerae* within a community [34]. Due to its low efficacy and a large number of drawbacks, it is not a recommended vaccine for cholera [35].

b) Oral killed whole cell - cholera toxin recombinant B subunit vaccine (WC-rBS):

This vaccine is made of killed whole cells of *V. cholerae* O1 (classical and El Tor strain) and nontoxic B subunit of the cholera toxin (CtxB).



Figure 1.9: Commercially available oral vaccine manufactured by Shantha Biotech Company.

Source: <https://www.historyofvaccines.org/sites/default/files/inline-images/shanchol.jpg>

This vaccine provides equal protection against mild as well as severe cholera. Although it offers a high level of protection, its efficacy begins to come down gradually after 6 months of vaccination. Booster doses are recommended after a period of 6 months (infants) and 2 years (adults) [36]–[38]. WC-rBS resulted in a reduction of hospital admissions due to diarrhoea in Bangladesh by 50% and even reduced mortality by 45%. In Peru, WC-rBS led to a 2-fold increase in vibriocidal antibody with the booster dose being required after a period of 1 year. Thus, it can be concluded that it is safer and efficient than the parenteral vaccine provided that there is the timely administration of booster dose [39].

c) Oral live attenuated *Vibrio cholerae* vaccine (CVD 103-HgR):

This vaccine has been synthesized from *V. cholerae* O1 classical strain which cannot produce the enzymatically active CtxA subunit of the cholera toxin and has the mercury resistance gene (*hgR*) that allows the identification of the vaccine strain [40], [41]. This vaccine is immunogenic in patients and causes mild discomfort like abdominal pain, cramping, nausea and diarrhoea. However, its effect in immunosuppressed and immunocompromised individuals (HIV patients) is not well known. As this vaccine is consumed with aspartame as a sweetener, it cannot be used by patients with phenylketonuria. In studies conducted with volunteers, it has been seen that the single dose of oral live attenuated vaccine is three times more effective than a single dose of oral killed vaccine. However, CVD 103-HgR could not protect against cholera in areas where the disease is endemic [42]-[43]. Like all other cholera vaccines, this vaccine also needs to be followed by a booster dose after the first vaccination to provide immunity up to 2 years (Table 1.2). As this vaccine contains live cells, it should not be consumed along with antibiotics. Concurrent use of chloroquine can also decrease the efficacy of the vaccine [34], [44].

Currently, rehydration therapy is the only way to contain the adverse effects of cholera. There is no single treatment for complete eradication of cholera. Most of the *V. cholerae* strains have become resistant to the antibiotics and offer very little protection against the disease. In addition to this, the effect of cholera vaccines last only for a limited period of time and does not cover cholera caused by all strains of *V. cholerae*. Therefore, novel strategies are needed to curb it as the current solutions are ineffective.

Table 1.2: Summary of the main types of cholera vaccines.

Vaccine	Dosage	Route of administration	Time to immunity	Requirement of booster dose	Vaccine efficiency (%)	Side effects
Parenteral phenol - inactivated vaccine	2 doses	Intramuscular or intradermal	6 days after 2nd dose	6 months	30-50%	High
Oral killed whole cell-cholera toxin recombinant B subunit vaccine (WC-rBS)	3 doses	Oral	7 days after 3rd dose	6 months-2 years	50-85%	Low
Oral live attenuated <i>Vibrio cholerae</i> vaccine (CVD 103-HgR)	1 dose	Oral	Within 8 days of dose	6 months	62-100%	Low

1.5 Biofilms - a survival strategy in Bacteria

In aquatic ecosystems, bacteria prefer to live as specialized structures known as biofilms. These biofilms form a protective layer against shear and unfavourable conditions. They are made up of multiple biopolymers including glycolipids, extracellular DNA, polysaccharides and proteins with cells embedded in it [45]. These biofilms provide a niche for the bacteria to proliferate and also allow them to attach themselves to abiotic and biotic surfaces. This attachment is mitigated by TCP, pili, fimbria, accessory colonization factor and a couple of proteins such as hemagglutinin and enzymes [26].

1.5.1 Biofilm formation regulation

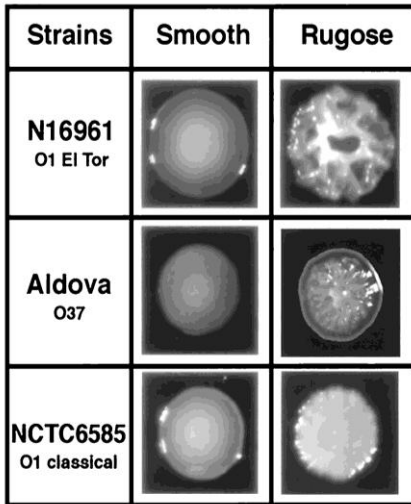


Figure 1.10: Smooth and Rugose morphology exhibited by different strains of *V. cholerae*.

Source: Ali et al, 2002, Reproduced with permission from ASM under the Copyright Clearance Center's RightLink

Most bacteria prefer to live in communities in the form of surface biofilms or as pellicles at liquid-air interface. Depending upon environmental cues, the bacteria can get dislodged from the biofilm matrix and give rise to new biofilms elsewhere. Although the composition of the biofilm matrix varies depending on the microorganisms present, nutrient availability, temperature and the shear

forces, it is predominantly made up of polysaccharides. Consequently, the presence of polysaccharides can influence two morphologies: rugose (R) and smooth (S) form (Figure 1.10). Rugose

colonies have a dry, wrinkled, raised surface due to the presence of polysaccharides to otherwise smooth colony morphology.

Rugose forms can form highly resistant biofilms that are capable of tolerating different forms of stress and provide an advantage to the wild-type cells. Also, from an immunological point of view, rugose forms are more resistant to serum killing as compared to the smooth forms thereby helping them survive and flourish in the harsh conditions [46]–[48].

1.5.2 Significance of Biofilms

Biofilms enable the organisms within it to survive in a stable, protective sheath away from the environmental stress [49]. These organisms are more resistant to physical stress, osmotic pressure, pesticides, oxidative stress and antimicrobial agents. This is one of the reasons why biofilms by pathogenic organisms are persistent in chronic infections [50], [51].

Apart from providing a stable environment, biofilms also act as a source for the formation of new biofilms in another place [52]–[54].

1.5.3 Biofilm formation steps

Biofilm formation generally involves five basic stages that may vary depending upon the type of organism.

These steps include:

- 1) reversible attachment,
- 2) irreversible attachment,
- 3) formation of microcolonies,
- 4) biofilm formation
- 5) biofilm dispersal



Figure 1.11: Steps involved in biofilm formation and the role of c-di-GMP in its regulation.

Source: Toyofuku et al, 2015, Reproduced with permission from Francis and Taylor Group under the Copyright Clearance Center's RightLink

During reversible attachment, cells attached to the biotic or abiotic surface may revert back to the motile form. However, in the case of irreversible attachment, bacteria orient itself for attachment to the surface and lose its swimming motility. As cells accumulate together, they form microcolonies either by the division of the attached cells or by the movement of nearby adherent cells by twitching motility.

These microcolonies grow into macrocolonies and biofilm by the sheath of exopolysaccharides, extracellular DNA and proteins which protect it from the harsh external environment. During favourable conditions, individual cells emerge from the biofilms and get dispersed by water to distant places (Figure 1.11) [55]–[57].

Another feature that is commonly associated with biofilm formation is pellicle formation. A pellicle is a thin layer of cells growing at the liquid-air interface. Pellicle formation starts with bacterial attachment to a surface at the liquid-air interface followed by monolayer formation [58]–[61]. This monolayer develops into larger aggregates depending upon the availability of oxygen. Dispersal of aggregates leads to a supply of a large number of bacteria which can aid disease transmission [62].

1.6 Cyclic-di-GMP

The bacterial secondary messenger, bis-(3'-5')-cyclic dimeric guanosine monophosphate (cyclic di-GMP or c-di-GMP), is globally present in most bacteria and regulates a lot of cellular activities like the expression of virulence factors and transition between motile to planktonic form in bacteria. Cyclic-di-GMP was first identified as an activator of the cellulose synthetase complex in *Komagataeibacter xylinus* (formerly *Gluconacetobacter xylinum*) 30 years ago (Table 1.3) [63]. Cyclic-di-GMP also regulates the production of extracellular polysaccharides (EPS), a key component of bacterial biofilms. In general, high intracellular concentrations of c-di-GMP promote EPS biosynthesis, the production of the adhesin(s), as well as biofilm and rugose colony formation while repressing motility and the expression of virulence factors, and vice versa.

Table 1.3: Historical timeline of the discovery of c-di-GMP and its regulatory elements.

Time	Event
~ 220 BC, Qin dynasty in China	Reportedly the first use of the Kombucha “tea mushroom,” a symbiotic culture of yeast and acetobacteria which produces a thick cellulose pellicle
1946	First studies of bacterial cellulose synthesis at The Hebrew University
1987	Discovery of c-di-GMP, its chemical synthesis, proof that c-di-GMP is the true activator of cellulose synthase
1995	Discovery that c-di-GMP suppresses replication of cancer cells
1995	Characterization of GGDEF domain in the <i>C. crescentus</i> response regulator PleD
1998	Characterization of DGC and c-di-GMP PDE genes (published in <i>Journal of Bacteriology</i>)
1998	Characterization of the EAL domain protein BvgR in <i>Bordetella pertussis</i> , alignment of the EAL domains
1999	Description of the HD-GYP domain, proposal of a c-di-GMP-related novel signal transduction system
1999	Characterization of the GGDEF-containing response regulators PleD and CslR
2000	Involvement of Adra, a transmembrane protein with a C-terminal GGDEF domain, in intercellular adhesion
2000	Involvement of the HD-GYP domain protein RpfG in regulation of pathogenicity in <i>X. campestris</i>
2000	The COG database identifies GGDEF, EAL, and HD-GYP domain genes in most bacteria but not in archaea
2001	Genetic proof that the GGDEF domain has DGC activity
2001	Detailed description of the GGDEF, EAL, and HD-GYP domains as components of bacterial signal transduction
2001	Binding of oxygen to its PAS domain regulates activity of the c-di-GMP PDE from <i>G. xylinus</i>
2004	Crystal structure of the GGDEF domain, experimental proof of its DGC activity, identification of the allosteric I site for feedback inhibition
2004	Proposal that c-di-GMP is a universal second messenger
2004	c-di-GMP involvement in pathogenesis of <i>Yersinia pestis</i> and <i>Vibrio cholerae</i>
2004	c-di-GMP and transition from sessility to motility
2005	GGDEF-catalyzed c-di-GMP biosynthesis in various bacterial phyla
2005	Experimental proof of the PDE activity of the EAL domain
2005	Biofilm dispersal by c-di-GMP
2006	Description of the c-di-GMP-binding PilZ domain
2006	Description of global c-di-GMP network regulation by the stress sigma factor RpoS in <i>E. coli</i>
2006–2007	Experimental proof that the PilZ domain binds c-di-GMP
2006–2007	Characterization of GGDEF-EAL domain proteins in which both domains are enzymatically active
2007	Description of immunostimulating activity of c-di-GMP
2008	Discovery of a c-di-GMP-sensing riboswitch
2008–2010	Description of global c-di-GMP network regulation by the RNA-binding protein CsrA and the quorum sensing system
2009	Crystal structure of the EAL domain
2010	Discovery of the second c-di-GMP-sensing riboswitch
2011	Molecular mechanism of regulation of LapG proteolytic activity through the c-di-GMP receptor LapD
2011–2012	Identification and structural characterization of the first eukaryotic c-di-GMP receptor
2012	Discovery of a c-di-GMP signaling system in the eukaryote <i>Dictyostelium</i> , a social amoeba

Source: Romling et al. 2013, Reproduced with permission from ASM under the Copyright Clearance Center’s RightLink

1.6.1 Synthesis

Intracellular c-di-GMP pools are controlled through the activity of two classes of enzymes.

Diguanylate cyclases (DGCs) synthesize c-di-GMP while specific phosphodiesterases (PDEs) degrade it.

Cyclic di-GMP is usually found as an intercalated dimer in the presence of monovalent ions. Moreover, c-di-GMP is in a monomer/dimer equilibrium resulting in low intracellular concentration. Due to its low concentration, preformed c-di-GMP is of very little use during cellular signalling.

Synthesis of c-di-GMP is catalyzed from 2 GTP molecules by diguanylate cyclase GGD/(E)EF domain which can homodimerize temporarily (Figure 1.12).

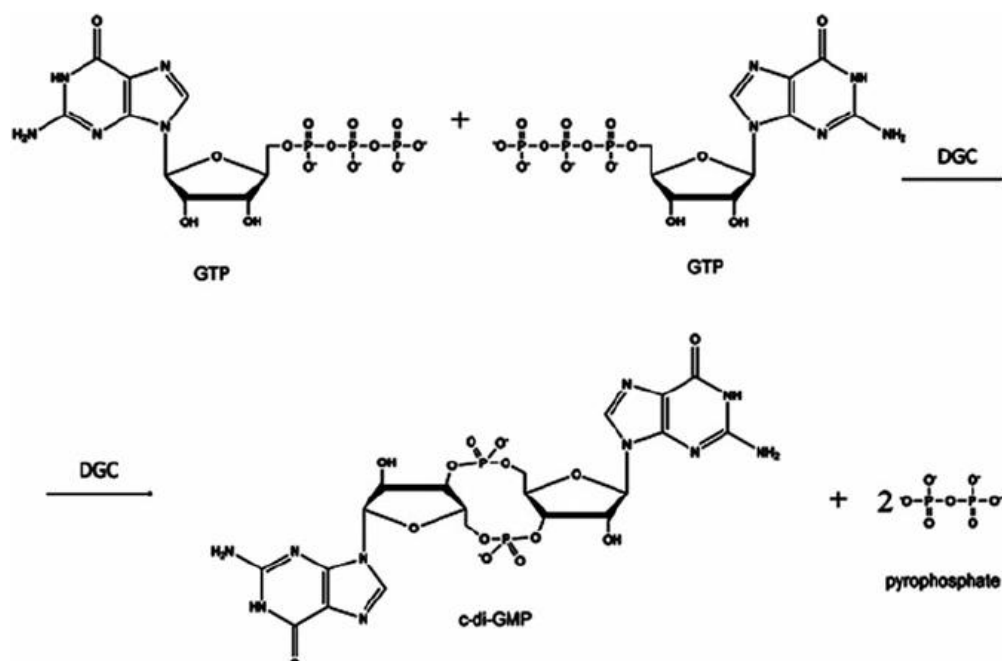


Figure 1.12: Synthesis of c-di-GMP from two GTP molecules.

Source: Shchokolova et al, 2015, Reproduced with permission from Francis and Taylor Group under the Copyright Clearance Center's RightLink

The phosphodiester bond formation for c-di-GMP synthesis requires the 2 GTP molecules to lie anti-parallel to each other. As the K_{cat} of DGC is considerably low (ranging between $k_{\text{cat}} = 1 \text{ min}^{-1}$ to 65 min^{-1}), increasing the concentration of GTP molecules does not necessarily increase the activity of DGC. The low K_{cat} could be attributed to the GGD/(E)EF rearrangement to facilitate product or substrate exchange [65], [66].

1.6.2 Degradation of c-di-GMP

Cyclic di-GMP is degraded by the phosphodiesterase (PDE) activity of EAL or HD-GYP protein domain. Degradation of c-di-GMP leads to the formation of linear 5'-pGpG or two GMP and generally requires Mn^{2+} or Mg^{2+} for its catalytic activity. But this rate of c-di-GMP degradation is much lower than its rate of synthesis. Although PDEs are considerably active as monomers, most of the active PDEs have shown to be in a dimeric or oligomeric state. Eventually, at low intracellular concentration of c-di-GMP, the biofilm is reduced and; motility and virulence are enhanced.

1.6.3 Functions of c-di-GMP

Cyclic di-GMP regulates a plethora of functions in bacteria at various levels of gene regulation ranging from transcription to translation (Figure 1.13). Depending on the level of c-di-GMP, it regulates the transition between sessile and motile lifestyles. Aside from lifestyle transition, c-di-GMP affects an array of phenotypes like antibiotic production in Streptomycetes, heterocyst formation in cyanobacteria and lipid metabolism in mycobacteria.

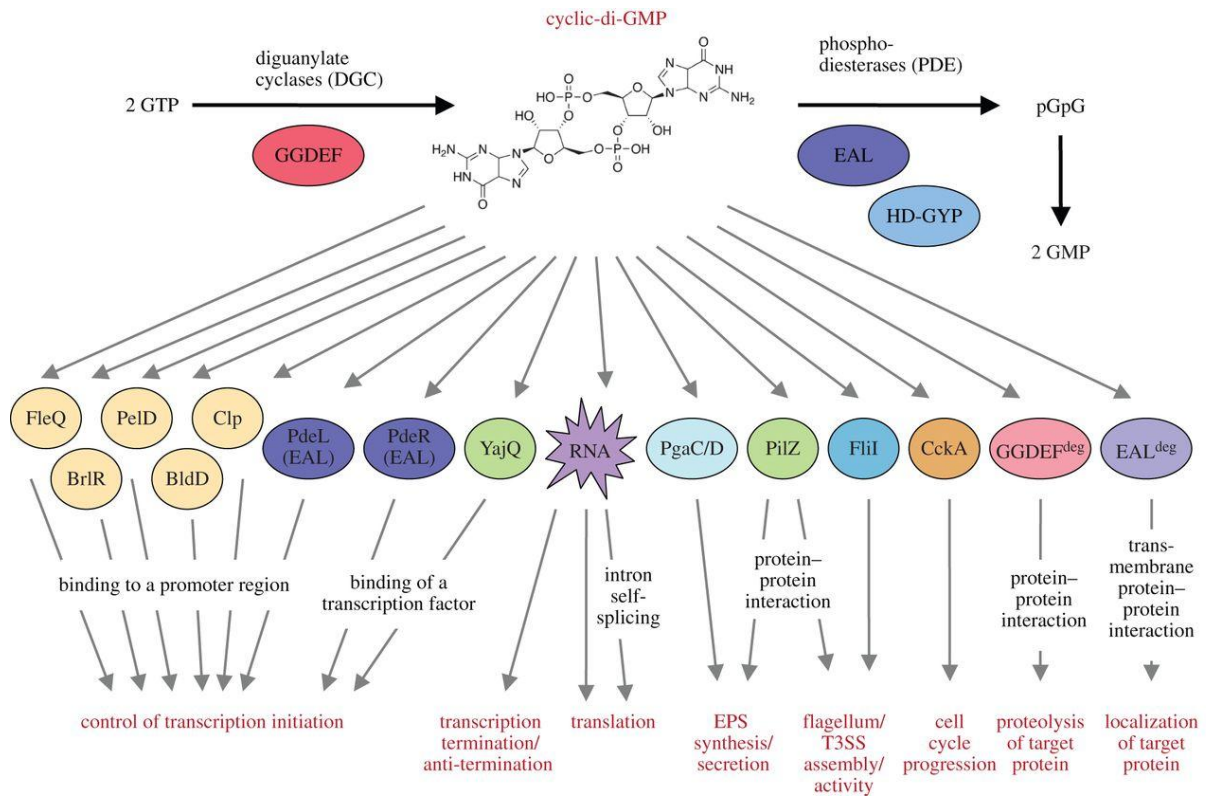


Figure 1.13: Role of c-di-GMP binding effectors at different levels gene transcription.

Source: Hengge, 2016, Reproduced with permission from Royal Society Publishing under the Copyright Clearance Center’s RightLink

1.6.3.1 Biofilm formation

High levels of c-di-GMP lead to biofilm formation in bacterial species enhancing community behaviour and adhering properties. Biofilm formation due to high levels of c-di-GMP is a common feature of bacterial species with community behaviours and adhering properties [67], [68]. For example, the diguanylate cyclase AdrA from *Salmonella typhimurium* allosterically controls cellulose production by binding to the PilZ domain of the cellulose synthase BcsA.

Along with cellulose synthesis, c-di-GMP controls the expression of other biofilm enhancing genes like the transcriptional activator *csgD* gene [69]–[72], *bapA* gene encoding for surface protein BapA [73] and *yih* gene encoding for O-antigen capsule [74].

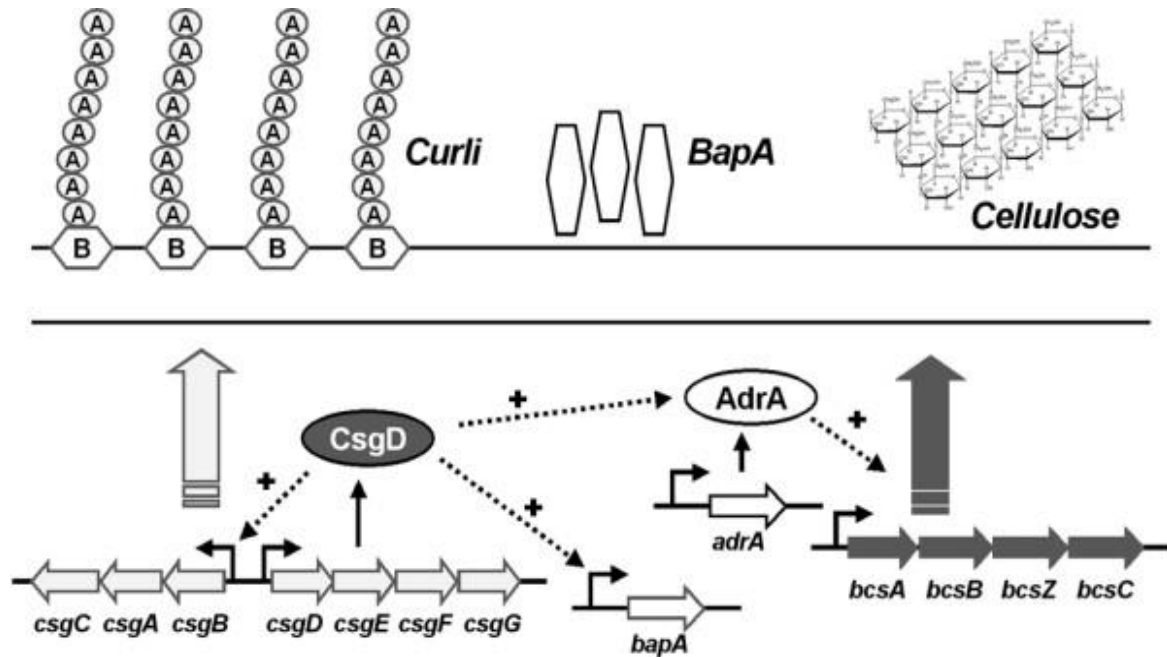


Figure 1.14: The regulatory network leading to exopolysaccharide secretion and biofilm formation in *Salmonella*.

Source: http://www.biologyonline.org/user_files/Image/Microbiology/curl,%20cellulose%20and%20BapA%20f01.jpg

1.6.3.2 Motility

High level of intracellular c-di-GMP is known to inhibit twitching motility and flagellar synthesis in numerous pathogens such as *S. typhimurium* and *Pseudomonas aeruginosa* [75], [76]. Different types of bacterial motilities like swimming, swarming and twitching, are regulated by c-di-GMP. Swimming motility in liquid media and swarming motility over solid surfaces are driven by the flagella.

However, twitching motility is regulated by the extension and retraction of Type IV pili (Tfp). Flagellar Type IV pili movement has been shown to be under the control of c-di-GMP.

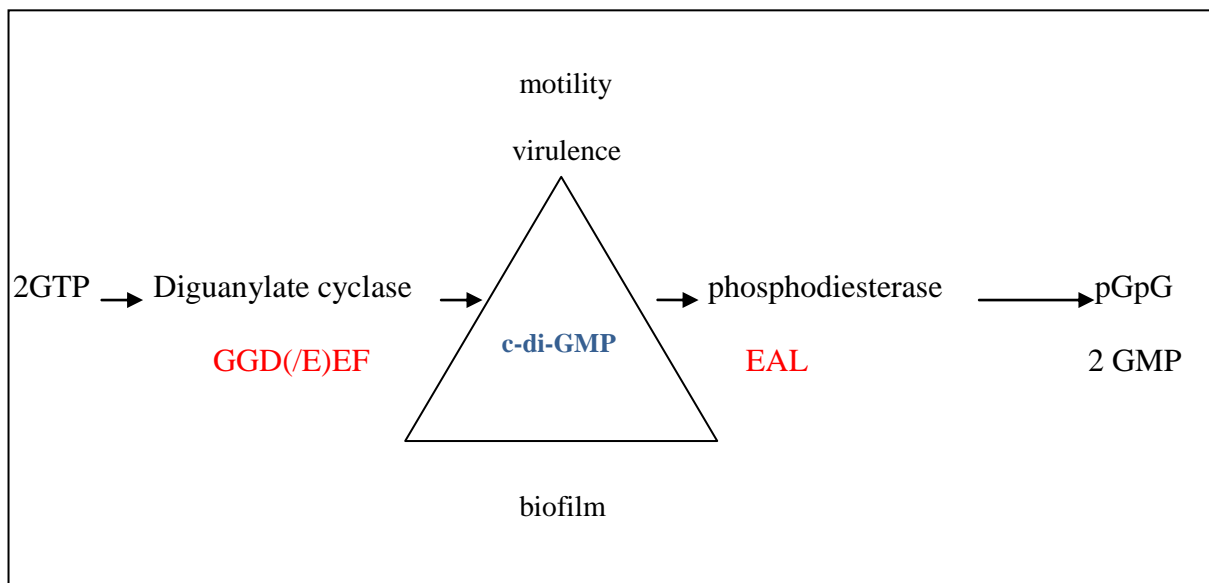


Figure 1.15: The concentration of the intracellular c-di-GMP determines the upregulation and downregulation of the motility and biofilm.

Suppression of flagellar motility by c-di-GMP has been demonstrated in *V. cholerae* and *S. typhimurium*, while transcriptional regulation of several proteins involved in flagellar synthesis is seen in several other bacterial strains [77].

1.6.3.3 Virulence due to c-di-GMP

Cyclic di-GMP signalling can be seen in a wide range of pathogens like *Escherichia coli*, *P. aeruginosa*, *V. cholerae*, *S. typhimurium*, *Brucella melitensis* and *Xanthomonas*. For survival of *Salmonella* in porcine cells, the GGD(E)EF-EAL domain protein STM1703 with c-di-GMP phosphodiesterase activity is required.

In *V. cholerae*, the activity of phosphodiesterase VieA lowers the c-di-GMP concentration thereby activating cholera toxin production. VieA, transcriptionally activates ToxT which in

turn activates *ctxAB* gene encoding for cholera toxin [78], [79]. In virulent phenotypes of *P. aeruginosa*, c-di-GMP signaling is prominently seen during chronic infections [80].

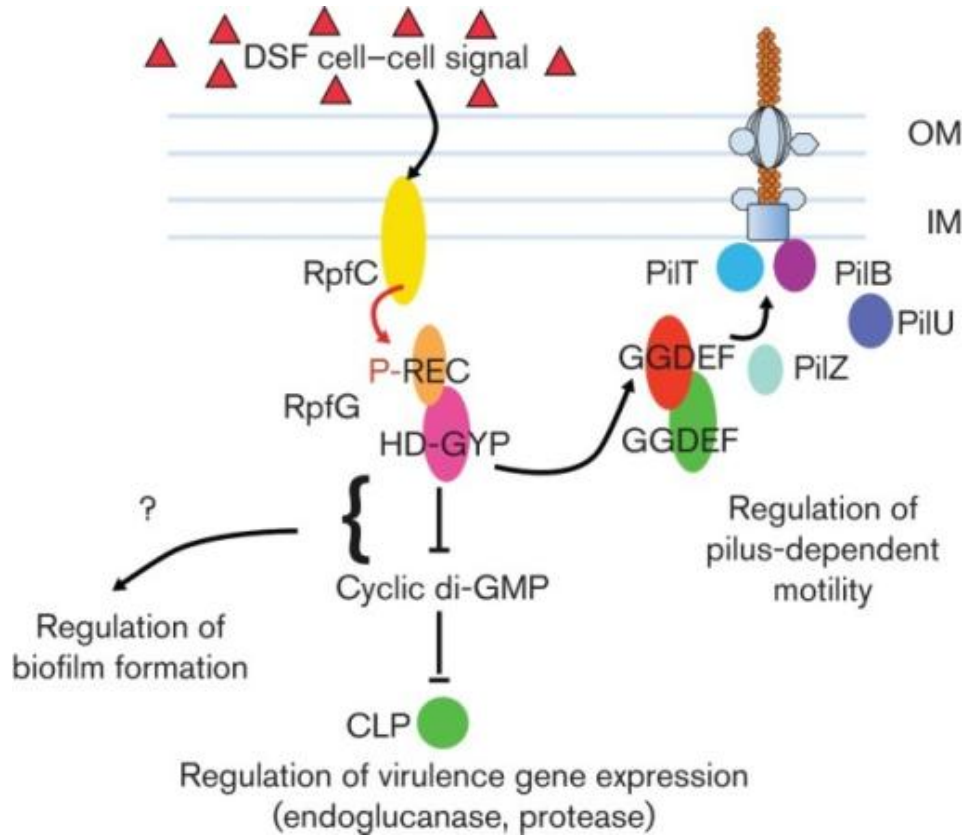


Figure 1.16: Role of diffusible signaling factor (DSF) from *Xanthomonas campestris* in the regulation of virulence.

Source: Ryan, 2013, Reproduced with permission from Microbiology Society under the Copyright Clearance Center's RightLink

In plant pathogen *Xanthomonas campestris* pathovar *campestris* (Xcc), expression of RpfG, a HDGYP domain protein and its interaction with diffusible signaling factor (DSF) regulates Xcc virulence traits like the production of extracellular polysaccharide, extracellular enzymes and motility (Figure 1.16) [81]. In *Bordetella pertussis*, EAL domain protein (BvgR) encoded by the gene *bvgR* controls the expression of virulence factors by repressing gene expression [82].

1.7 GGD(/E)EF domains

1.7.1 GGD(/E)EF domain

GGD(/E)EF (Gly-Gly-Asp/Glu-Glu-Phe) domain proteins are associated with the diguanylate cyclase activity and use two molecules of GTP to produce c-di-GMP with the release of two phosphates (PPi) (Figure 1.17) [83].

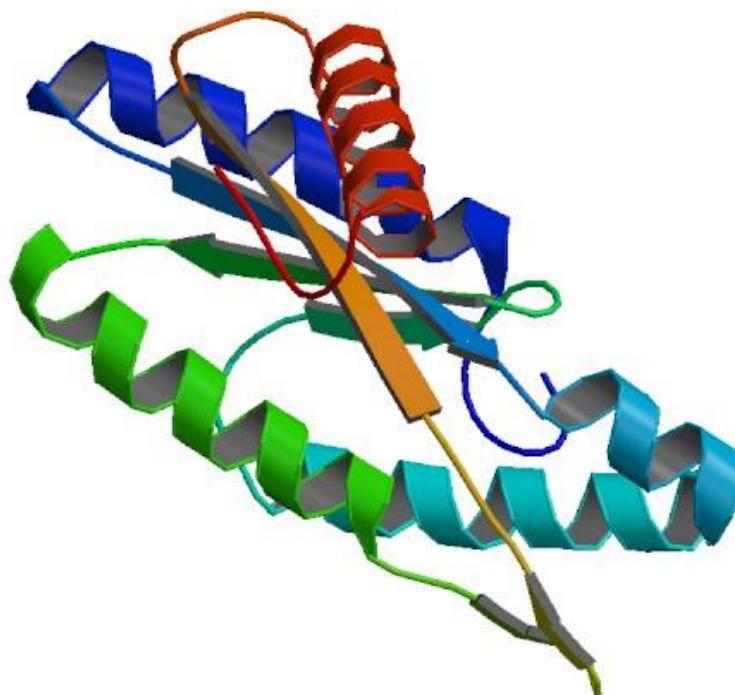


Figure 1.17: Crystal structure of GGDEF domain from *Thermotoga maritima*

Source: http://www.rcsb.org/pdb/images/4URG_bio_r_500.jpg

Usually, GGD(/E)EF domains are approximately 170 amino acids long [84]. The GGD(/E)EF motif in the domain plays a significant role in the activity of the protein and mutation in any amino acid residue may result in loss of activity of protein [85]. It is generally activated by N-terminal signalling domain like the PAS domain and is inactivated by N-terminal inhibitory site (I-site) with the RXXD motif. The active site (A-site) of the GGD(/E)EF protein domain lies downstream of the I-site and is known to bind to GTP.

Analysis of this site has revealed that the amino acids in the GGD(/E)EF motif bind to β and γ phosphates of GTP and to the guanine base. The GGD(/E)EF domain is also allosterically inhibited when c-di-GMP (product) binds to it to prevent excessive GTP consumption [86]. The mechanism of feedback inhibition involves c-di-GMP binding to the I-site that is four to nine amino acids away from the catalytic A-site. This way the amount of c-di-GMP produced by the diguanylate cyclase is regulated by non competitive product inhibition [87], [88].

In vitro assays have demonstrated that GGD(/E)EF protein needs to be in a homodimeric form to convert two GTP into c-di-GMP. Also, structural studies in GGD(/E)EF domain PelD from *Caulobacter crescentus* has shown the presence of Mg^{2+} ions indicating its role in the catalytic mechanism [89].

1.7.2 EAL domain

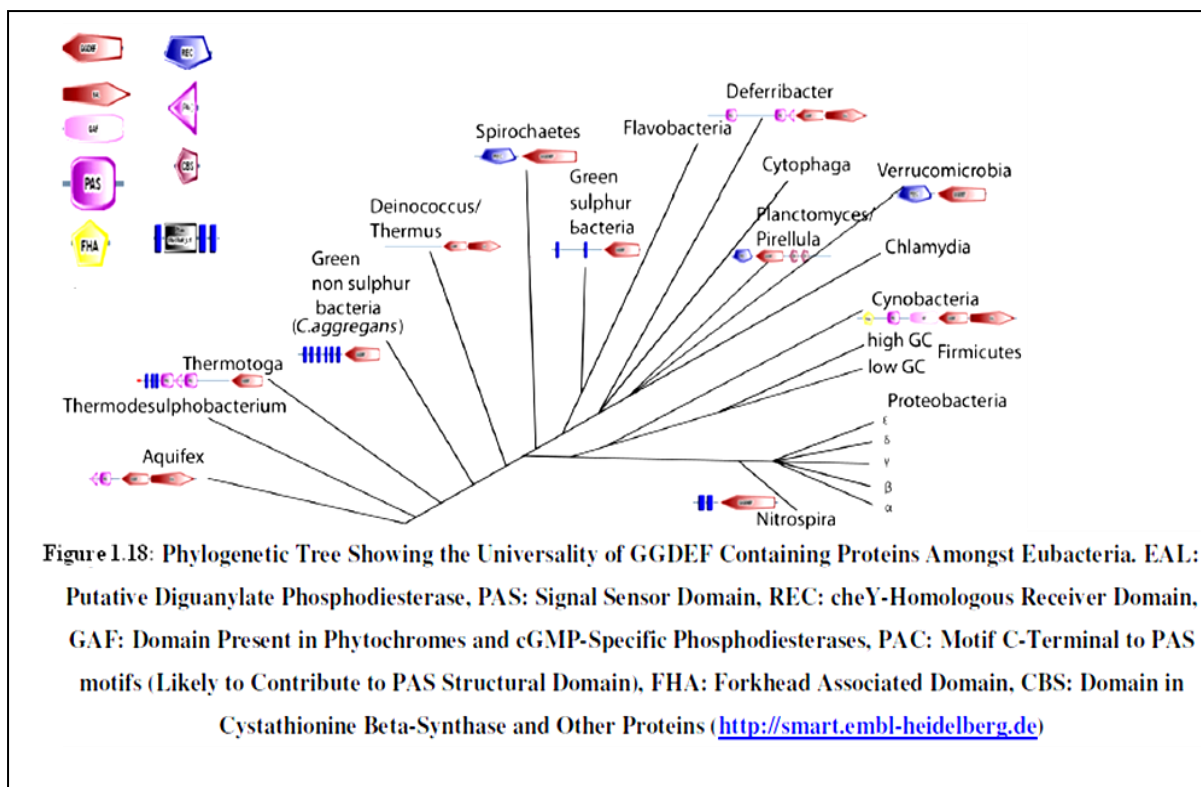
The EAL protein domain has the conserved EAL (Glu-Ala-Leu) motif and is associated with phosphodiesterase (PDE) activity. In general, the EAL domain requires the presence of divalent ions like Mg^{2+} or Mn^{2+} for activity, but it is strongly inhibited by Ca^{2+} or Zn^{2+} . The mechanism of action involves binding of the glutamic acid of the EAL motif to Mg^{2+} ions for the enzymatic activity. Therefore, mutations, particularly in the EAL motif may lead to loss of enzymatic activity in some EAL domain proteins.

Phosphodiesterase activity of EAL or HD-GYP domain results in the breakdown of c-di-GMP into two molecules of Guanosine monophosphate (GMP). HD-GYP domains are a subfamily of the HD family metal-dependent phosphohydrolases and were first studied in the plant pathogen *X. campestris* (Xcc) [81], [86]. As they occur less frequently than the EAL domains, it can be suggested that EAL domain proteins contribute to the majority of the

phosphodiesterase activity in bacteria. Studies in *S. enterica* (protein YhjH) and *V. cholerae* (protein VieA) have shown that overexpression of the EAL domain led to the reduction of the intracellular level of c-di-GMP by phosphodiesterase activity [79], [90].

1.7.3 Universality of GGD(E)EF domain

GGD(E)EF domains are found across the different members of the Domain Bacteria (Figure 1.18). The first known report of GGD(E)EF domain was seen in *Komagataeibacter xylinus* (formerly *Gluconacetobacter xylinum*) where its diguanylate cyclase activity was involved in allosteric activation of cellulose synthase by the production of c-di-GMP [63]. Later, another GGD(E)EF protein PleD was extensively studied in *Caulobacter crescentus* where it was found along with other response regulator domain. Similarly, AdrA is a transmembrane GGD(E)EF protein domain which on expression induces the production of cellulose in *S. typhimurium* [91]. Lately, the role of GGD(E)EF protein DgcA, in the production of slime, has also been documented in a eukaryotic amoeba *Dictyostelium discoideum* [92].



Source: Bandekar, 2013, Reproduced from GGD(E)EF domain: a ubiquitous eubacterial domain confers cell toxicity.

1.7.4 Plurality of GGD(E)EF-EAL domains

GGD(E)EF and EAL domain proteins are extensively found in bacterial genomes [64], [93], [94]. Generally, any given bacterial genome consists of more than one GGD(E)EF and EAL domain protein which questions the specificity of the c-diGMP signaling pathways. The genome of *S. typhimurium* codes for 20 GGD(E)EF/EAL domain proteins; 5 GGD(E)EF, 8 EAL domain and 7 GGD(E)EF-EAL domain proteins.

Whereas, *E. coli* K-12 has 12 GGD(/E)EF, 12 EAL and 7 GGD(/E)EF-EAL domain proteins. In *P. aeruginosa* genome, there are 17 GGD(/E)EF, 5 EAL and 16 GGD(/E)EF-EAL domain proteins. On the other hand, the *V. cholerae* genome has 31 GGD(/E)EF, 22 EAL and 10 GGD(/E)EF-EAL domain proteins [95].

The study of DGCs and PDEs in *K. xylinus* showed the presence of conserved domains, GGD(/E)EF and EAL. Amongst the proteins containing GGD(/E)EF and EAL domains, most contain both the domains. In general, the GGD(/E)EF-EAL domains are located at the C-terminus with regulatory domains like the PAS at the N-terminus. Mostly, both the GGD(/E)EF and EAL domains are enzymatically active with only a few cases where either one domain is enzymatic active. In such cases, the enzymatically inactive domain may have a regulatory function [64], [96]–[98]. Also, there could be a possibility where none of the two domains is active [99].

1.7.5 Coupling of regulatory domains with GGD(/E)EF and EAL domains

There are a large number of sensory and signal transduction domains such as PAS, GAF, HAMP REC, and HTH domains associated with GGD(/E)EF and EAL domains. Most of these lie upstream at the N-terminal end and enable them to identify primary signalling molecules like electrons, amino acids, oxygen, photons, nutrient starvation and antibiotics [83], [93].

For example, PAS (Per-Arnt-Sim), the conserved protein domain is involved in sensing oxygen and redox via binding of cofactors like flavin and heme or light. Another regulatory protein REC (or CheY) is a chemotaxis response protein domain that participates in signaling phosphorelays [100]. In *P. aeruginosa*, along with the GGD(/E)EF-EAL domain protein

FimX, response regulator proteins like the N-terminal PAS signaling domains and chemotaxis response regulator CheY are also found [89], [101]. The GGD(/E)EF domain PleD of the *C. crescentus* protein contains a C-terminal GGD(/E)EF domain and two N-terminal receiver domains. The proteins control cell transition from the swarmer (motile form) to the stalked form (biofilm form).

1.8 VC0395_0300 protein from *V. cholerae*

Bis-(3'-5') cyclic dimeric guanosine monophosphate (c-di-GMP) is a bacterial secondary messenger controlling diverse bacterial phenotypes mostly known to be involved in the transition from free-living, motile to biofilm lifestyle in Gram-negative bacteria like *V. cholerae*. Cyclic-di-GMP also regulates a wide range of cellular processes at the transcriptional, translational, or post-translational level. This includes the synthesis of virulence factors and toxins, the production of adhesins and biofilm matrix components, the regulation of different forms of cell motility, as well as cell cycle progression.

V. cholerae is known to thrive in an aquatic environment by forming robust biofilms which are quite resistant to shear and chemical stress. These biofilms are composed of cells embedded in sheaths of polysaccharide formed to the high levels of c-di-GMP. In most bacteria, the synthesis is regulated by the activity of diguanylate cyclase encoded by the GGD(/E)EF protein domain. VC0395_0300 is a putative protein from *V. cholerae* which has the conserved GGEEF motif in the protein domain. It also has an N-terminal PAS domain which facilitates the identification of external primary signals like sensing oxygen.

1.9 Gaps in existing research

a) Cholera is a disease of great epidemic potential. Short-lived, protective immunity develops in adults in the endemic areas due to repeated exposures to infection, but long-term protection and treatment are rare. Most of the cholera vaccines provide short-term protection and require administration of timely booster dose. Also, a single vaccine is not effective against a plethora of new *V. cholerae* strains.

b) *V. cholerae* survives in the aquatic environment by formation of biofilms. As c-di-GMP signaling plays a critical role in promoting biofilm formation, inhibition of effectors of c-di-GMP signaling systems like GGD(/E)EF would offer an attractive approach to interfere with biofilm formation.

c) Although GGD(/E)EF domains have been implicated with cyclic di-GMP turnover for a long time in many organisms, information regarding the precise biochemical role and structure of these domains is lacking. The modulation of c-di-GMP by GGD(/E)EF proteins is still poorly understood in *V. cholerae*. Since the structure of GGEEF domain from *V. cholerae* has not been solved, biophysical studies of this protein will provide us valuable insights into the structure and function of GGEEF proteins and thereby help in curbing sporadic *V. cholerae* infections.

1.10 Aim of the study

In view of the mentioned gaps in research, the following research objectives were proposed:

1. Overexpression and purification of wild-type protein VC0395_0300.
2. Biophysical characterization of wild-type protein VC0395_0300.
3. Crystallization and structural studies of wild-type protein VC0395_0300.
4. Elucidation of the GGD(/E)EF domain of wild-type protein VC0395_0300.

1.11 References

- [1] J. B. Kaper, J. G. Morris, and M. M. Levine, “Cholera,” *Clinical Microbiology Reviews*, vol. 8, no. 1. pp. 48–86, 1995.
- [2] D. A. Herrington, R. H. Hall, G. Losonsky, J. J. Mekalanos, R. K. Taylor, and M. M. Levine, “Toxin, toxin-coregulated pili, and the toxR regulon are essential for *Vibrio cholerae* pathogenesis in humans.,” *J. Exp. Med.*, vol. 168, no. 4, pp. 1487–1492, 1988.
- [3] C. T. Codeco and F. C. Coelho, “Trends in cholera epidemiology,” *PLoS Medicine*, vol. 3, no. 1. pp. 16–17, 2006.
- [4] A. Huq *et al.*, “Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh,” *Appl. Environ. Microbiol.*, vol. 71, no. 8, pp. 4645–4654, 2005.
- [5] S. M. Faruque, M. J. Albert, and J. J. Mekalanos, “Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*.,” *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 4, pp. 1301–14, 1998.
- [6] M. J. Albert, “Epidemiology & molecular biology of *Vibrio cholerae* O139 Bengal.,” *Indian J. Med. Res.*, vol. 104, pp. 14–27, 1996.
- [7] J. Holmgren, I. Lönnroth, J. Månsson, and L. Svennerholm, “Interaction of cholera toxin and membrane GM1 ganglioside of small intestine.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 72, no. 7, pp. 2520–4, 1975.
- [8] K. Bharati and N. K. Ganguly, “Cholera toxin: A paradigm of a multifunctional protein,” *Indian Journal of Medical Research*, vol. 133, no. 2. pp. 179–187, 2011.
- [9] S. N. De *et al.*, “Enterotoxicity of bacteria-free culture-filtrate of *Vibrio cholerae*.,” *Nature*, vol. 183, no. 4674, pp. 1533–4, 1959.

- [10] F. Chowdhury *et al.*, “Impact of rapid urbanization on the rates of infection by *Vibrio cholerae* O1 and enterotoxigenic *Escherichia coli* in Dhaka, Bangladesh,” *PLoS Negl. Trop. Dis.*, vol. 5, no. 4, 2011.
- [11] S. M. Faruque *et al.*, “Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 5, pp. 1702–1707, 2005.
- [12] A. K. Siddique *et al.*, “Cholera epidemics in Bangladesh: 1985-1991,” *J. Diarrhoeal Dis. Res.*, vol. 10, pp. 79–86, 1992.
- [13] E. J. Barzilay *et al.*, “Cholera surveillance during the Haiti epidemic--the first 2 years,” *N. Engl. J. Med.*, vol. 368, no. 7, pp. 599–609, 2013.
- [14] M. K. Kay *et al.*, “*Vibrio mimicus* infection associated with crayfish consumption, Spokane, Washington, 2010,” *J Food Prot*, vol. 75, no. 4, pp. 762–764, 2012.
- [15] A. Hossah, K. Haider, M. I. Huq, and A. Zamad, “Studies on the factors affecting the haemolysin production of *Vibrio mimicus* isolated from clinical and environmental sources,” *Trans. R. Soc. Trop. Med. Hyg.*, vol. 82, no. 2, pp. 337–339, 1988.
- [16] M. Blokesch and G. K. Schoolnik, “Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs,” *PLoS Pathog.*, vol. 3, no. 6, pp. 0733–0742, 2007.
- [17] K. Penrose, M. C. De Castro, J. Werema, and E. T. Ryan, “Informal urban settlements and cholera risk in Dar es Salaam, Tanzania,” *PLoS Negl. Trop. Dis.*, vol. 4, no. 3, 2010.
- [18] J. Sepúlveda, J. L. Valdespino, and L. García-García, “Cholera in Mexico: The paradoxical benefits of the last pandemic,” *International Journal of Infectious Diseases*, vol. 10, no. 1. pp. 4–13, 2006.
- [19] J. B. Harris, R. C. LaRocque, F. Qadri, E. T. Ryan, and S. B. Calderwood, “Cholera,” in *The Lancet*, 2012, vol. 379, no. 9835, pp. 2466–2476.

- [20] D. A. Sack, R. B. Sack, G. B. Nair, and A. Siddique, "Cholera," *Lancet*, vol. 363, no. 9404, pp. 223–233, 2004.
- [21] G. D. Pearson, a Woods, S. L. Chiang, and J. J. Mekalanos, "CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 8, pp. 3750–4, 1993.
- [22] M. Trucksis, J. Michalski, Y. K. Deng, and J. B. Kaper, "The *Vibrio cholerae* genome contains two unique circular chromosomes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 24, pp. 14464–9, Nov. 1998.
- [23] C. M. Fraser *et al.*, "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*," *Nature*, vol. 406, no. 6795, pp. 477–483, Aug. 2000.
- [24] Q. Xu, M. Dziejman, and J. J. Mekalanos, "Determination of the transcriptome of *Vibrio cholerae* during intractintestinal growth and midexponential phase in vitro," *Proc. Natl. Acad. Sci.*, vol. 100, no. 3, pp. 1286–1291, Feb. 2003.
- [25] J. J. Lospalluto and R. A. Finkelstein, "Chemical and physical properties of cholera exo-enterotoxin (choleraegen) and its spontaneously formed toxoid (choleraegenoid)," *BBA - Protein Struct.*, vol. 257, no. 1, pp. 158–166, 1972.
- [26] G. Liu *et al.*, "Resistance of the cholera vaccine candidate IEM108 against CTX infection," *Vaccine*, vol. 24, no. 11, pp. 1749–1755, 2006.
- [27] B. K. Hammer and B. L. Bassler, "Quorum sensing controls biofilm formation in *Vibrio cholerae*," *Mol. Microbiol.*, vol. 50, no. 1, pp. 101–114, 2003.
- [28] S. Das, S. Choudhry, R. Saha, V. G. Ramachandran, K. Kaur, and B. L. Sarkar, "Emergence of multiple drug resistance *Vibrio cholerae* O1 in East Delhi," *J. Infect. Dev. Ctries.*, vol. 5, no. 4, pp. 294–298, 2011.
- [29] R. C. Charles and E. T. Ryan, "Cholera in the 21st century.," *Curr. Opin. Infect. Dis.*, vol. 24, no. 5, pp. 472–477, 2011.

- [30] J. T. Weber *et al.*, “Epidemic cholera in Ecuador: multidrug-resistance and transmission by water and seafood,” *Epidemiol. Infect.*, vol. 112, no. 1, pp. 1–11, Feb. 1994.
- [31] K. J. Towner, N. J. Pearson, F. S. Mhalu, and F. O’Grady, “Resistance to antimicrobial agents of *Vibrio cholerae* E1 Tor strains isolated during the fourth cholera epidemic in the United Republic of Tanzania.,” *Bull. World Health Organ.*, vol. 58, no. 5, pp. 747–b51, 1980.
- [32] B. Sharifi-Mood and M. Metanat, “Diagnosis, Clinical Management, Prevention, and Control of Cholera; A Review Study,” *Int. J. Infect.*, vol. 1, no. 1, Mar. 2014.
- [33] M. M. Levine and N. F. Pierce, “Immunity and Vaccine Development,” in *Cholera*, Boston, MA: Springer US, 1992, pp. 285–327.
- [34] J. D. Clemens, S. Shin, B. K. Sah, and D. A. Sack, “Cholera vaccines,” in *Vaccines: Sixth Edition*, 2012, pp. 141–152.
- [35] A. S. Azman, L. C. Ivers, D. Legros, F. J. Luquero, and E. D. Mintz, “Safe water, sanitation, hygiene, and a cholera vaccine,” *The Lancet*, vol. 387, no. 10013. p. 28, 2016.
- [36] A. M. Khatib *et al.*, “Effectiveness of an oral cholera vaccine in Zanzibar: Findings from a mass vaccination campaign and observational cohort study,” *Lancet Infect. Dis.*, vol. 12, no. 11, pp. 837–844, 2012.
- [37] V. D. Thiem *et al.*, “Long-term effectiveness against cholera of oral killed whole-cell vaccine produced in Vietnam,” *Vaccine*, vol. 24, no. 20, pp. 4297–4303, 2006.
- [38] A. L. Lopez, M. L. A. Gonzales, J. G. Aldaba, and G. B. Nair, “Killed oral cholera vaccines: history, development and implementation challenges,” *Ther. Adv. Vaccines*, vol. 2, no. 5, pp. 123–136, 2014.

- [39] R. E. Black, M. M. Levine, M. L. Clements, C. R. Young, A. M. Svennerholm, and J. Holmgren, "Protective efficacy in humans of killed whole-vibrio oral cholera vaccine with and without the B subunit of cholera toxin," *Infect. Immun.*, vol. 55, no. 5, pp. 1116–1120, 1987.
- [40] J. F. Viret, G. Dietrich, and D. Favre, "Biosafety aspects of the recombinant live oral *Vibrio cholerae* vaccine strain CVD 103-HgR," in *Vaccine*, 2004, vol. 22, no. 19, pp. 2457–2469.
- [41] M. Levine *et al.*, "Safety, immunogenicity, and efficacy of recombinant live oral cholera vaccines, CVD 103 and CVD 103-HgR," *Lancet*, vol. 332, no. 8609, pp. 467–470, 1988.
- [42] Suharyono *et al.*, "Safety and immunogenicity of single-dose live oral cholera vaccine CVD 103-HgR in 5-9-year-old Indonesian children," *Lancet*, vol. 340, no. 8821, pp. 689–694, 1992.
- [43] R. Lagos *et al.*, "Tolerance and immunogenicity of an oral dose of CVD 103-HgR, a live attenuated *Vibrio cholerae* 01 strain: a double-blind study of Chilean adults," *Rev. Med. Chil.*, vol. 121, no. 8, pp. 857–63, 1993.
- [44] D. N. Taylor *et al.*, "Expanded safety and immunogenicity of a bivalent, oral, attenuated cholera vaccine, CVD 103-HgR plus CVD 111, in United States military personnel stationed in Panama," *Infect. Immun.*, vol. 67, no. 4, pp. 2030–2034, 1999.
- [45] E. E. Mann and D. J. Wozniak, "*Pseudomonas* biofilm matrix composition and niche biology," *FEMS Microbiology Reviews*, vol. 36, no. 4, pp. 893–916, 2012.
- [46] T. Ramamurthy *et al.*, "*Vibrio mimicus* with multiple toxin types isolated from human and environmental sources," *J. Med. Microbiol.*, vol. 40, no. 3, pp. 194–196, 1994.

- [47] F. L. Singleton, R. Attwell, S. Jangi, and R. R. Colwell, "Effects of temperature and salinity on *Vibrio cholerae* growth," *Appl. Environ. Microbiol.*, vol. 44, no. 5, pp. 1047–1058, 1982.
- [48] J. M. Janda, C. Powers, R. G. Bryant, and S. L. Abbott, "Current perspectives on the epidemiology and pathogenesis of clinically significant *Vibrio spp.*," *Clinical Microbiology Reviews*, vol. 1, no. 3, pp. 245–267, 1988.
- [49] J. W. Costerton, P. S. Stewart, and E. P. Greenberg, "Bacterial biofilms: a common cause of persistent infections.," *Science*, vol. 284, no. 5418, pp. 1318–22, 1999.
- [50] T. F. C. Mah and G. A. O'Toole, "Mechanisms of biofilm resistance to antimicrobial agents," *Trends in Microbiology*, vol. 9, no. 1, pp. 34–39, 2001.
- [51] C. A. Fux, J. W. Costerton, P. S. Stewart, and P. Stoodley, "Survival strategies of infectious biofilms," *Trends in Microbiology*, vol. 13, no. 1, pp. 34–40, 2005.
- [52] E. A. Kaissi and A. Montratos, "Preparation and properties of *Vibrio cholerae* antifimbrial antibody," *J. Appl. Bacteriol.*, vol. 58, no. 2, pp. 221–229, 1985.
- [53] L. Hall-Stoodley and P. Stoodley, "Biofilm formation and dispersal and the transmission of human pathogens," *Trends in Microbiology*, vol. 13, no. 1, pp. 7–10, 2005.
- [54] S. N. Wai, Y. Mizunoe, and S. I. Yoshida, "How *Vibrio cholerae* survive during starvation," *FEMS Microbiology Letters*, vol. 180, no. 2, pp. 123–131, 1999.
- [55] J. Liao and K. Sauer, "The MerR-like transcriptional regulator BrIR contributes to *Pseudomonas aeruginosa* biofilm tolerance," *J. Bacteriol.*, vol. 194, no. 18, pp. 4823–4836, 2012.
- [56] T. Tolker-Nielsen and S. Molin, "Spatial Organization of Microbial Biofilm Communities," *Microb Ecol*, vol. 40, no. 2, pp. 75–84, 2000.

- [57] J. S. Webb *et al.*, “Cell death in *Pseudomonas aeruginosa* biofilm development,” *J. Bacteriol.*, vol. 185, no. 15, pp. 4585–4592, 2003.
- [58] J. L. Enos-Berlage, Z. T. Guvener, C. E. Keenan, and L. L. McCarter, “Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*,” *Mol. Microbiol.*, vol. 55, no. 4, pp. 1160–1182, 2005.
- [59] C. L. Darnell, E. A. Hussa, and K. L. Visick, “The putative hybrid sensor kinase SypF coordinates biofilm formation in *Vibrio fischeri* by acting upstream of two response regulators, SypG and VpsR,” *J. Bacteriol.*, vol. 190, no. 14, pp. 4941–4950, 2008.
- [60] D. C. Old and J. P. Duguid, “Selective outgrowth of fimbriate bacteria in static liquid medium,” *J. Bacteriol.*, vol. 103, no. 2, pp. 447–456, 1970.
- [61] S. L. Kuchma, N. J. Delalez, L. M. Filkins, E. A. Snavelly, J. P. Armitage, and G. A. O’Toole, “Cyclic di-GMP-mediated repression of swarming motility by *Pseudomonas aeruginosa* PA14 Requires the MotAB stator,” *J. Bacteriol.*, vol. 197, no. 3, pp. 420–430, 2015.
- [62] J. G. Morris and R. E. Black, “Cholera and other vibrioses in the United States,” *N. Eng. J. Med.*, vol. 312, no. 6, pp. 343–50, 1985.
- [63] P. Ross *et al.*, “Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid,” *Nature*, vol. 325, no. 6101, pp. 279–281, 1987.
- [64] U. Romling, M. Y. Galperin, and M. Gomelsky, “Cyclic di-GMP: the first 25 years of a universal bacterial second messenger,” *Microbiol. Mol. Biol. Rev.*, vol. 77, no. 1, pp. 1–52, 2013.
- [65] M. Tarnawski, T. R. M. Barends, and I. Schlichting, “Structural analysis of an oxygen-regulated diguanylate cyclase,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 71, pp. 2158–2177, 2015.

- [66] J. Zhou *et al.*, “Potent suppression of c-di-GMP synthesis via I-site allosteric inhibition of diguanylate cyclases with c-di-GMP,” *Bioorganic Med. Chem.*, vol. 21, no. 14, pp. 4396–4404, 2013.
- [67] P. V. Krasteva *et al.*, “*Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP,” *Science*, vol. 48, no. Suppl 2, pp. 1–6, 2010.
- [68] U. Römling, “Characterization of the rdar morphotype, a multicellular behaviour in *Enterobacteriaceae*,” *Cellular and Molecular Life Sciences*, vol. 62, no. 11, pp. 1234–1246, 2005.
- [69] N. Grantcharova, V. Peters, C. Monteiro, K. Zakikhany, and U. Römling, “Bistable expression of CsgD in biofilm development of *Salmonella enterica* serovar Typhimurium,” *J. Bacteriol.*, vol. 192, no. 2, pp. 456–466, 2010.
- [70] S. Da Re and J.-M. Ghigo, “A CsgD-independent pathway for cellulose production and biofilm formation in *Escherichia coli*,” *J. Bacteriol.*, vol. 188, no. 8, pp. 3073–3087, 2006.
- [71] K. Zakikhany, C. R. Harrington, M. Nimtz, J. C. D. Hinton, and U. Römling, “Unphosphorylated CsgD controls biofilm formation in *Salmonella enterica* serovar Typhimurium,” *Mol. Microbiol.*, vol. 77, no. 3, pp. 771–786, 2010.
- [72] E. Brombacher, A. Baratto, C. Dorel, and P. Landini, “Gene expression regulation by the curli activator CsgD protein: Modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion,” *J. Bacteriol.*, vol. 188, no. 6, pp. 2027–2037, 2006.
- [73] C. Latasa *et al.*, “BapA, a large secreted protein required for biofilm formation and host colonization of *Salmonella enterica* serovar Enteritidis,” *Mol. Microbiol.*, vol. 58, no. 5, pp. 1322–1339, 2005.

- [74] J. M. Marshall and J. S. Gunn, “The O-antigen capsule of *Salmonella enterica* serovar Typhimurium facilitates serum resistance and surface expression of FliC,” *Infect. Immun.*, vol. 83, no. 10, pp. 3946–3959, 2015.
- [75] J. W. Hickman and C. S. Harwood, “Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor,” *Mol. Microbiol.*, vol. 69, no. 2, pp. 376–389, 2008.
- [76] V. Zorraquino *et al.*, “Coordinated cyclic-Di-GMP repression of *Salmonella* motility through *ycgR* and cellulose,” *J. Bacteriol.*, vol. 195, no. 3, pp. 417–428, 2013.
- [77] M. E. Anyan *et al.*, “Type IV pili interactions promote intercellular association and moderate swarming of *Pseudomonas aeruginosa*,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 50, p. 18013–18018, 2014.
- [78] A. D. Tischler, S. H. Lee, and A. Camilli, “The *Vibrio cholerae* *vieSAB* locus encodes a pathway contributing to cholera toxin production,” *J. Bacteriol.*, vol. 184, no. 15, pp. 4104–4113, 2002.
- [79] R. Tamayo, A. D. Tischler, and A. Camilli, “The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase,” *J. Biol. Chem.*, vol. 280, no. 39, pp. 33324–33330, 2005.
- [80] M. Valentini and A. Filloux, “Biofilms and Cyclic di-GMP (c-di-GMP) signaling: Lessons from *Pseudomonas aeruginosa* and other bacteria,” *Journal of Biological Chemistry*, vol. 291, no. 24, pp. 12547–12555, 2016.
- [81] Y. Fouhy, J. F. Lucey, R. P. Ryan, and J. M. Dow, “Cell-cell signaling, cyclic di-GMP turnover and regulation of virulence in *Xanthomonas campestris*,” *Research in Microbiology*, vol. 157, no. 10, pp. 899–904, 2006.

- [82] T. J. Merkel, P. E. Boucher, S. Stibitz, and V. K. Grippe, “Analysis of bvgR Expression in *Bordetella pertussis*,” *J. Bacteriol.*, vol. 185, no. 23, pp. 6902–6912, 2003.
- [83] C. Chan *et al.*, “Structural basis of activity and allosteric control of diguanylate cyclase,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 49, pp. 17084–17089, 2004.
- [84] A. J. Schmidt, D. A. Ryjenkov, and M. Gomelsky, “The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains,” *J. Bacteriol.*, vol. 187, no. 14, pp. 4774–4781, 2005.
- [85] D. A. Ryjenkov, M. Tarutina, O. V. Moskvina, and M. Gomelsky, “Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain,” *J. Bacteriol.*, vol. 187, no. 5, pp. 1792–1798, 2005.
- [86] C. Y. Yang, K. H. Chin, M. L. C. Chuah, Z. X. Liang, A. H. J. Wang, and S. H. Chou, “The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)₂ at the active site,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 12, pp. 997–1008, 2011.
- [87] R. Hengge, “Principles of c-di-GMP signalling in bacteria,” *Nat. Rev. Microbiol.*, vol. 7, no. 4, pp. 263–273, 2009.
- [88] H. Sondermann, N. J. Shikuma, and F. H. Yildiz, “Truncated form of the title: Mechanism of c-di-GMP signaling,” *Current Opinion in Microbiology*, vol. 15, no. 2, pp. 140–146, Apr-2012.
- [89] L. S. Marmont, J. C. Whitney, H. Robinson, K. M. Colvin, M. R. Parsek, and P. L. Howell, “Expression, purification, crystallization and preliminary X-ray analysis of *Pseudomonas aeruginosa* PelD,” *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 68, no. 2, pp. 181–184, 2012.

- [90] N. De, M. V. A. S. Navarro, R. V. Raghavan, and H. Sondermann, “Determinants for the Activation and Autoinhibition of the Diguanylate Cyclase Response Regulator WspR,” *J. Mol. Biol.*, vol. 393, no. 3, pp. 619–633, 2009.
- [91] B. García, C. Latasa, C. Solano, F. G. Portillo, C. Gamazo, and I. Lasa, “Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation,” *Mol. Microbiol.*, vol. 54, no. 1, pp. 264–277, Aug. 2004.
- [92] Z. Chen and P. Schaap, “The prokaryote messenger c-di-GMP triggers stalk cell differentiation in *Dictyostelium*,” *Nature*, vol. 488, no. 7413, pp. 680–683, 2012.
- [93] M. Y. Galperin, A. N. Nikolskaya, and E. V. Koonin, “Novel domains of the prokaryotic two-component signal transduction systems,” *FEMS Microbiology Letters*, vol. 203, no. 1, pp. 11–21, 2001.
- [94] J. G. Malone, R. Williams, M. Christen, U. Jenal, A. J. Spiers, and P. B. Rainey, “The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain,” *Microbiology*, vol. 153, no. 4, pp. 980–994, 2007.
- [95] K. Lewis, “Riddle of biofilm resistance,” *Antimicrobial Agents and Chemotherapy*, vol. 45, no. 4, pp. 999–1007, 2001.
- [96] C. Wei *et al.*, “A systematic analysis of the role of GGDEF-EAL domain proteins in virulence and motility in *Xanthomonas oryzae* pv. *oryzicola*,” *Sci. Rep.*, vol. 6, no. 1, p. 23769, 2016.
- [97] F. Zappacosta *et al.*, “Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody,” *Protein Sci.*, vol. 5, no. 5, pp. 802–13, 1996.

- [98] S. Lindenberg, G. Klauck, C. Pesavento, E. Klauck, and R. Hengge, “The EAL domain protein YciR acts as a trigger enzyme in a c-di-GMP signalling cascade in *E. coli* biofilm control.,” *EMBO J.*, vol. 32, no. 14, pp. 2001–14, 2013.
- [99] T. Schirmer and U. Jenal, “Structural and mechanistic determinants of c-di-GMP signalling.,” *Nat. Rev. Microbiol.*, vol. 7, no. 10, pp. 724–735, 2009.
- [100] S. An, J. Wu, and L. H. Zhang, “Modulation of *Pseudomonas aeruginosa* biofilm dispersal by a cyclic-di-GMP phosphodiesterase with a putative hypoxia-sensing domain,” *Appl. Environ. Microbiol.*, vol. 76, no. 24, pp. 8160–8173, 2010.
- [101] M. V. A. S. Navarro, N. De, N. Bae, Q. Wang, and H. Sondermann, “Structural Analysis of the GGDEF-EAL Domain-Containing c-di-GMP Receptor FimX,” *Structure*, vol. 17, no. 8, pp. 1104–1116, 2009.

Chapter 2
Biophysical Characterization of Sebox3
Protein

2.1 Introduction

Vibrio cholerae, the causal organism for diarrhoeal disease cholera, exhibits a dual mode of survival. It can be motile in human hosts and also form biofilms on abiotic surfaces in aquatic bodies [1]–[3]. These biofilms are very robust and provide a protective niche in which the organisms can survive. When the conditions are favourable, the organisms can multiply and spread out of the biofilm to give rise to new biofilms or infect [4]. The matrix of these biofilms is composed of extracellular DNA, proteins, lipids and polysaccharides like cellulose [5]–[8]. Large amounts of cellulose are synthesized when c-di-GMP allosterically binds to cellulose synthase [9], [10]. This c-di-GMP is in turn, produced by diguanylate cyclases encoded by GGD(E)EF domain proteins.

Although there are 53 GGDEF and/or EAL domains in the *V. cholerae* genome (31 GGDEF domains, 12 EAL domains, and 10 GGDEF-EAL domains), not much is known about the mechanism of action of diguanylate cyclase from *Vibrio cholerae*. Therefore, VC0395_0300, a putative GGD(E)EF protein from the chromosome I of *Vibrio cholerae* classical strain O395, serotype O1 and biotype Ogawa was selected for the study. According to National Centre for Biotechnology Information (NCBI) gene annotation (NCBI Reference Sequence: NC_009456.1) and domain search by SMART, VC0395_0300 is a sensory box GGD(E)EF domain proteins from *V. cholerae* and belongs to a family of previously unknown function. The full-length 321 amino acid-long protein has a GGEEF motif with an N-terminal PAS domain and is referred to as Sebox3 throughout the text. Biophysical characterization would provide ample information about the structural features of the protein and enable us to understand how the protein functions.

2.2 Materials and methods

2.2.1 Materials

DNA Primers were custom synthesized by Integrated DNA technologies (IDT). *Escherichia coli* strains BL21 (DE3) and DH5 α were obtained from Novagen. Luria Bertani broth, Miller from HiMedia Laboratories was used for cultivation and maintenance of all the recombinant strains. Enzymes like DNA polymerase, restriction enzymes were procured from Thermo Fischer Scientific. QiAquick DNA purification kit from Qiagen was used. GeneJET Plasmid Midiprep Kit was purchased from Thermo Scientific. PreScission protease was acquired from GE Healthcare. Solvents used for liquid chromatography were bought from Merck LiChrosolv and were of High Pressure Liquid Chromatography (HPLC) grade. All the reagents used in the study were of molecular grade and procured from Merck India.

2.2.2 Cloning of Sebox3 construct

The genomic DNA from *Vibrio cholerae* classical strain O395 was extracted and purified from cells grown overnight in 5 ml of LB broth. These cells were then lysed in lysis buffer (10 mM Tris-HCl, 1 mM EDTA (Ethylenediaminetetraacetic acid) - pH 8.0 and 10% SDS (Sodium dodecyl sulfate) and incubated with Proteinase K - 100 mg/mL in 0.5% SDS at 37°C for 1 hour. After the incubation period, 1% CTAB (Cetyl Trimethyl Ammonium Bromide) in 0.7 M NaCl was added and incubated at 65°C for 10 min. RNaseA treatment was given to the extracted DNA at 37°C for 1 hour to remove RNA. Phenol-chloroform extraction was followed by isopropanol precipitation of DNA. Any trace of salt was removed by washing with 70% ethanol. Finally, the genomic DNA was air-dried and dissolved in TE (Tris-EDTA) buffer.

The concentration was checked and purity was quantified by calculating the ratio of optical densities at 260 nm and 280 nm and stored in -20°C for further use as a template in PCR reactions[11].

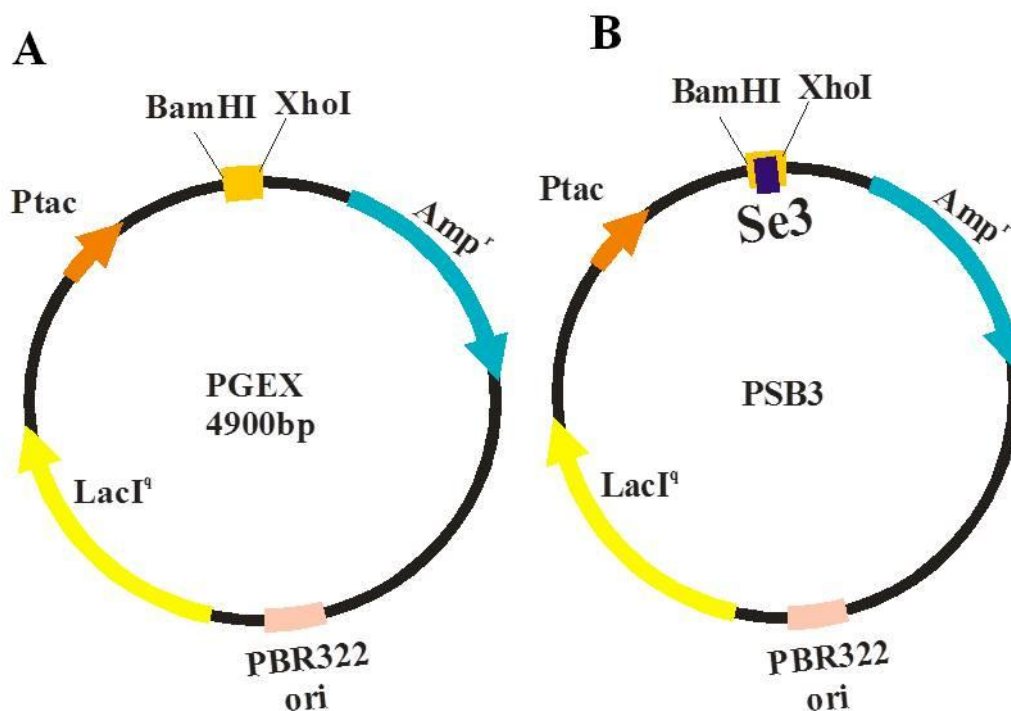


Figure 2.1: Design of the plasmid map. A: Plasmid map showing of pGEX-6P1 vector showing Bam HI and XhoI restriction site. B: Construct map of Sebox3 showing gene inserted between the Bam HI and XhoI site.

The 966 bp long gene encoding VC0395_0300 from *Vibrio cholerae* 0395 was amplified by PCR using Thermo Scientific DreamTaq DNA polymerase. The amplified DNA fragment was purified by using QiAquick kit from Qiagen using manufacturer's instructions. BamHI and XhoI (Thermo Scientific) were used for restriction digestion of the amplicon. This amplified fragment was cloned into a similarly digested pGEX-6P1 (GE Healthcare) that allows the cytoplasmic expression of Glutathione S transferase (GST) fusion protein with a PreScission protease cleavage site at the N-terminal. The purified restriction digested DNA fragments were ligated by T4 DNA ligase into the restriction digested plasmid to obtain the Sebox3 construct (**Figure 2.1**). This was transformed into DH5 α and subsequently in BL21

host cells. The sequence of the cloned DNA fragment was checked by nucleotide sequencing for incorporation of any unrequired mutations. The primers used for PCR amplification are listed in the below Table 2.1.

Table 2.1. The sequence of primers with restriction enzymes (BamHI in Sebox1A and XhoI in Sebox3A) underlined within the primer sequences.

<u>Sebox 1A</u> (Forward primer)	5'AATACT <u>GGATCC</u> ATGAAAAATTGGCTGTGTCAGGCAGTG 3'
<u>Sebox 3A</u> (Reverse primer)	5'AATACT <u>TCTCGAGT</u> TATTCTGTGGATTGGCGATAGATACA 3'

2.2.3 Protein expression

Initially, the recombinant Sebox3 protein was expressed in *E.coli* BL21 (DE3). A single colony of transformed cells from a Luria Bertani (LB) agar plate with ampicillin was used to inoculate 10 ml of LB medium with 100 µg/mL ampicillin and was allowed to grow at 37°C with continuous shaking at 150 rpm. One percent of this culture was used as the inoculum for a 50 mL medium which was grown to saturation and used to inoculate a further bulk culture of 500 mL volume, which was allowed to reach an OD₆₀₀ of 0.6. The expression of the recombinant protein was induced by the addition of different IPTG concentrations (ranging from 0.1 to 1 mM IPTG) and at different temperatures (ranging from 16 to 25°C) for different time spans. Cells were pelleted by centrifugation, washed and suspended in a homogenization buffer [50 mM Tris-HCl (pH7.4), 150 mM NaCl, 1 mM DTT and 0.5 mM EDTA] [12], [13].

The cell suspension was lysed by sonication using 30-second pulses of low amplitude (interrupted by 5-minute intervals where the cells were kept in ice) and centrifuged at 12000

rpm to separate the insoluble fractions. The total protein concentration of the lysate was determined using Bradford reagent [14] at each step.

2.2.4 Protein purification

For purification purposes, the media volume was scaled up to 2 litres and 0.05 mM IPTG final concentration was used for induction for 8 hours at 16°C and 180 rpm. The protein obtained in the supernatant was loaded into glutathione agarose resin column (GE Healthcare Life Sciences), which had been washed and equilibrated in equilibration buffer [50 mM Tris-HCl, 150 mM NaCl, 1 mM DTT and 0.5 mM EDTA (pH 7.4)], for 1 hour at 4°C. Unbound fractions that did not bind were collected for further analysis. After binding of the target protein, the column was rinsed with 10 bed volumes of wash buffer [50 mM Tris-HCl, 150 mM NaCl, 0.5 mM reduced glutathione (pH 8.0)] and finally eluted with 10 column volumes of freshly prepared elution buffer [50 mM Tris-HCl and 10 mM reduced glutathione (pH 8.0)]. The purity of the protein was checked on a 12.5% Tris-glycine SDS-PAGE without the addition of β -mercaptoethanol. Fractions containing the purified protein were collected together and dialyzed against a 50 mM Tris-HCl buffer (pH8.0), to completely remove any traces of glutathione. Protein was quantified with Bradford assay and spectrophotometric methods utilizing the extinction coefficient of $24660 \text{ M}^{-1}\text{cm}^{-1}$ at 280 nm and a theoretical mass of 37.18 kDa, as provided by ExpASY (<http://expasy.org/protparam/>).

2.2.5 Western Blot Analysis

The purified protein in the SDS gel was transferred to an Immobilon-P Transfer Membrane (Merck Millipore) in a tank transfer assembly using 25 mM Tris-HCl (pH 8.3), 150 mM glycine and 10% methanol. Blocking of the membrane was carried out for 2 hours with 50 mM Tris-HCl (pH 7.4), 0.1% Tween-20 and 5% low-fat milk. The anti-GST antibody (GE

Healthcare) in the same buffer was introduced at a dilution of 1:3000 and incubated for 10 hours.

The membrane was washed with a wash buffer (Tris-buffered saline – 50 mM Tris and 150 mM NaCl) and incubated with rabbit I_gG-Horse Radish Peroxidase conjugated antibody at a dilution of 1:6,000. After sufficient washing, the membrane was developed using enhanced chemiluminescence kit (Thermo Pierce ECL Western Blotting Substrate) before being exposed to a photographic X-ray film.

2.2.6 Cleavage of GST tag

Removal of the tag was essentiated by the fact that the GST tag is comparable in size to the Sebox3 protein. Therefore, the tagged protein was subject to overnight digestion at 4°C with 10 units/mg of PreScission Protease (GE Healthcare Life Sciences) in solution. After digestion, the entire digestion mixture containing GST tag, GST-tagged Sebox3 protein and tagless Sebox3 protein was reloaded into an equilibrated GST Sepharose column. The tagless protein was collected from the flow through of the column. Subsequent washing and elution of the column resulted in purified Sebox3 bereft of GST tag. Since the yield of the tagless protein was a little low when compared to the tagged one, the tagless protein was concentrated at 4°C using a spin column concentrator [(NMWCO= 10 kDa), Amicon Ultra, Millipore].

2.2.7 Determination of oligomeric status

2.2.7.1 Size exclusion chromatography

Analytical size-exclusion chromatography was performed using a Superdex-200 column (GE Healthcare), equilibrated with a buffer containing 50 mM Tris-HCl and 150 mM NaCl (pH 7.4). The column was attached to an ÄKTA prime purifier system (GE Healthcare). Standard

gel filtration markers (Sigma Aldrich) having a range from 12.4 kDa to 200 kDa were run in the column to estimate standard retention times. The tagless protein sample was injected at a flow rate of 1.0 mL/min into the column. The retention volumes for each of the samples were measured and used to calculate the retention factor R_f , defined by the following equation:

$$R_f = (V_e - V_o)/(V_t - V_o)$$

where V_e = elution volume, V_o = void volume (calculated based on the retention time of the blue dextran standard), and V_t = geometric bead volume for the column. The collected fractions were then evaluated by 12.5% SDS-PAGE electrophoresis [15].

2.2.7.2 HPLC

Oligomeric state in a solution of GST-tagged Sebox3 protein was determined by HPLC (Agilent Infinity 1260 system) using a reverse phase HPLC C-18 column (Waters Xterra - 3.5 μ m X 4.6 mm X 250 mm) connected to an Agilent Infinity 1260 system. The mobile phase for the separation was solvent A: 5% acetonitrile in water with 0.1% TFA and solvent B: 95% acetonitrile in water with 0.1% TFA. 20 μ L of the 7 μ M protein sample in 50 mM Tris-HCl (pH8.0) was separated at a flow rate of 0.8mL/min with a gradient of 0-80% in 40 min.

2.2.8 Fluorescence Spectroscopy

Fluorescence spectroscopy is used as a device to study protein folding and unfolding. In proteins, three amino acids, that is, tryptophan, tyrosine and phenylalanine contribute to intrinsic fluorescence. Usually, only tryptophan fluorescence is preferred as it has the highest quantum yield amongst the three amino acids and gives out strong fluorescence signal. The Sebox3 protein has tryptophans at the W4 and W172 position. To study the intrinsic fluorescence of the two tryptophans in the protein sequence, fluorescence measurements were carried out using a Jasco FP8200 spectrofluorometer. Fluorescence spectra were measured

keeping both the excitation and emission band passes at 5 nm. The tryptophans were selectively excited at the wavelength of 295 nm and the emission spectra obtained between 305 nm to 400 nm were noted.

2.2.8.1 Thermal denaturation

For thermal denaturation, fluorescence spectra were collected with the spectrofluorometer equipped with ESCY IC201 temperature controller water bath. Temperature-induced denaturation of the protein was carried out by increasing the temperature from 25°C-90°C. The emission readings were collected after incubating the protein samples for 5 min in cuvettes surrounded by a constant temperature-controlled water jacket.

2.2.8.2 Chemical denaturation

For the denaturation study, a series of freshly prepared solutions of guanidine hydrochloride (GdnHCl) (Merck) having concentrations in the range of 0.5 M to 5 M in buffer [50 mM Tris-HCl and 150 mM NaCl (pH 7.4)] were prepared and the proteins were added to a final concentration of 7 μ M in a quartz cuvette with 1 cm pathlength. Proteins were incubated separately in GdnHCl containing buffer overnight at 25°C.

Equal volumes of buffer were added to the same volume of GdnHCl solutions and these mixtures were used as blanks [16]. All the fluorescence spectra were blank corrected for any background fluorescence intensity using measurements of buffers.

2.2.8.3 Quenching of fluorescence

In the quenching experiments, the protein was incubated with acrylamide (Ultrapure), KI and CsCl (Merck). Aliquots of the 5 M quencher stock solution was added to the 7 μ M protein

sample to obtain the desired quencher concentration ranging from (0.2 M - 1.0 M). The samples were excited at 295 nm in order to ensure only tryptophan fluorescence and emission was recorded in the range of 305-400 nm. The results obtained were analyzed by the Stern-Volmer equation,

$$F_0/F=1+K_{SV}\cdot[Q]$$

where F_0 and F are fluorescence intensities in the absence and presence of quencher, respectively; K_{SV} is Stern-Volmer constant and Q is the concentration of the quencher [17], [18].

2.2.9 HPLC-Diguanylate cyclase reaction

A reaction mixture containing 5 μ M protein, 50 mM Tris-HCl (pH 8.0), and 10 mM $MgCl_2$ was pre-warmed at 37°C for 10 min before initiating the reaction. 50 μ M of GTP was added to the reaction mixture and incubated at 37°C for 45 min. The reaction was terminated by addition of one-fourth the volume of 0.5 M EDTA and by boiling the entire reaction mixture for 5 min [19]. Due care was taken to remove any and all traces of the residual protein in the final solution.

Subsequently, the separation was achieved on a reverse phase HPLC C18 column (Waters Xterra - 3.5 μ m X 4.6 mm X 250 mm) connected to an Agilent Infinity 1260 system. 20 μ L of the sample was separated using a gradient of mobile phase consisting of solvent A with 100% methanol and solvent B with 10 mM tributylamine and 15 mM acetic acid in water: methanol (97:3) [20].

The sample was injected into the column maintained at a flow rate of 0.4 mL/min and the following gradient was used: 0% to 10% solvent B in 15 min, 10-20% solvent B from 15-20 min, 20-30% solvent B from 20-25 min, 30-50% solvent B from 25-30 min, 50-90% solvent

B from 30-35 min, 90-95% solvent B from 35-40 min and 95-0% solvent B from 40-45 min. Commercially available GTP and c-di-GMP (Sigma Aldrich) were used to calibrate the C18 column.

2.2.10 Transverse Urea Gel Electrophoresis (TUGE)

The transverse urea gel was made with the help of a gradient maker with gel buffer (30% acrylamide solution, 1.5 M Tris-HCl, 10% SDS, 10% ammonium per sulphate, TEMED and water) and 0 M urea in the 1st mixing chamber and gel buffer and 8 M urea in the second chamber. The gel was overlaid with isopropanol and allowed to polymerize for 1 hour.

The spacer was clamped together to give rise to a trough on top of the gel where the protein sample was loaded. The resultant 0 M - 8 M urea polyacrylamide gradient gel was run transverse to the direction of electrophoresis [21].

2.2.11 DTNB (5, 5'-dithio-bis-[2-nitrobenzoic acid]) Reaction

Quantification of free sulfhydryl groups in protein was carried out by DTNB reaction. For the reaction three samples were taken: a native protein sample, native protein reduced with 10 mM DTT and native protein denatured by 6 M GdnHCl. The buffer, buffer with 10 mM DTT and buffer with 6 M GdnHCl were used as blanks respectively. Aliquots from the 10 mM DTNB stock were added to the samples to obtain a final DTNB concentration of 0.3 mM. These samples were incubated for 1 hour before taking the reading on a double beam Shimadzu UV-visible 2450 spectrophotometer. The amount of yellow-coloured 2-nitro-5-mercaptobenzoic acid (TNB) released due to the presence of free sulfhydryl was measured at 412 nm.

The molar extinction coefficient of VC0395_0300 at 280 nm (ϵ_{280}) was $24660 \text{ M}^{-1}\text{cm}^{-1}$ as found by the ProtParam program [22]. The molar extinction coefficient at 412 nm (ϵ_{412}) of

TNB was $13700 \text{ M}^{-1}\text{cm}^{-1}$ in phosphate buffer with GdnHCl and $14150 \text{ M}^{-1}\text{cm}^{-1}$ in phosphate buffer, respectively. Accessible cysteine residues were calculated according to the formula:

$$\text{Accessible cysteine (AC)} = (\text{OD}_{412}/\epsilon_{412}) / (\text{OD}_{280}/\epsilon_{280}),$$

where, AC represents the number of accessible cysteine residues per monomer, OD_{412} is the absorbance of protein measured at 412 nm in presence of DTNB, ϵ_{412} is the molar extinction coefficient of TNB molecule at 412 nm, OD_{280} is the absorbance of protein measured at 280 nm and ϵ_{280} is the molar extinction coefficient of protein at 280 nm.

2.2.12 Limited Proteolysis

In order to probe the possible cleavage sites in the Sebox3 protein, 4 μg of protein in 50 mM Tris-HCl (pH 7.4) was digested by incubating at 25°C with 16 ng of trypsin and chymotrypsin respectively. Aliquots of the reaction mixture were removed and terminated at different time intervals of 0, 2, 5, 15, 30, 60, 90, 120 and 150 min.

The reaction was stopped by addition of 5X protein loading dye followed by boiling the sample for 2 min. The products of proteolytic cleavage were analyzed on a 15% Tris-Tricine SDS PAGE [23], [24].

2.3 Results and Discussion

2.3.1 Sebox3 cloned into pGEX-6P1 in *E. coli* DH5 α

Low cell viability has been reported for GGD(/E)EF proteins previously expressed from various sources [19]. The same was evident for the Sebox3 clone as well, and transformation efficiency was extremely low, especially when the plasmid 6P1-VC0395_300 was expressed in BL21 (DE3) cells. Also, earlier efforts to generate full-length His-tagged protein using pET28a and pET23a vector were unsuccessful (Figure 2.2). Therefore, the gene encoding the protein was cloned into the pGEX-6P1 vector. Glutathione-S-transferase (GST) tag was used

as it enhances the solubility of the protein and offers the use of mild elution conditions which preserves the activity of protein [25].

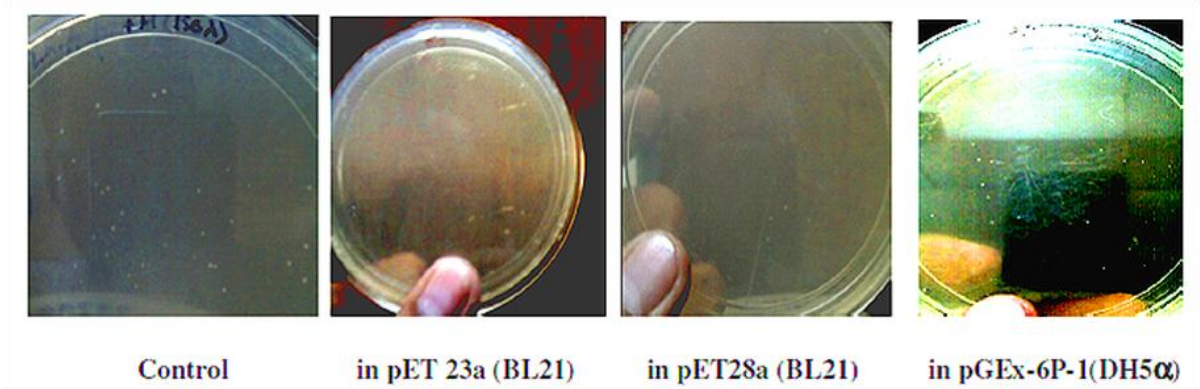


Figure 2.2: Cell viability of *E. coli* host cells due to the toxicity of GGD/(E)EF protein (left to right). Control plate (left) without the Sebox3 plasmid showed the maximum number of colonies. The *sebox3* gene cloned in pGEX-6P-1 and transformed in DH5 α showed the better cell viability as compared to the other constructs.

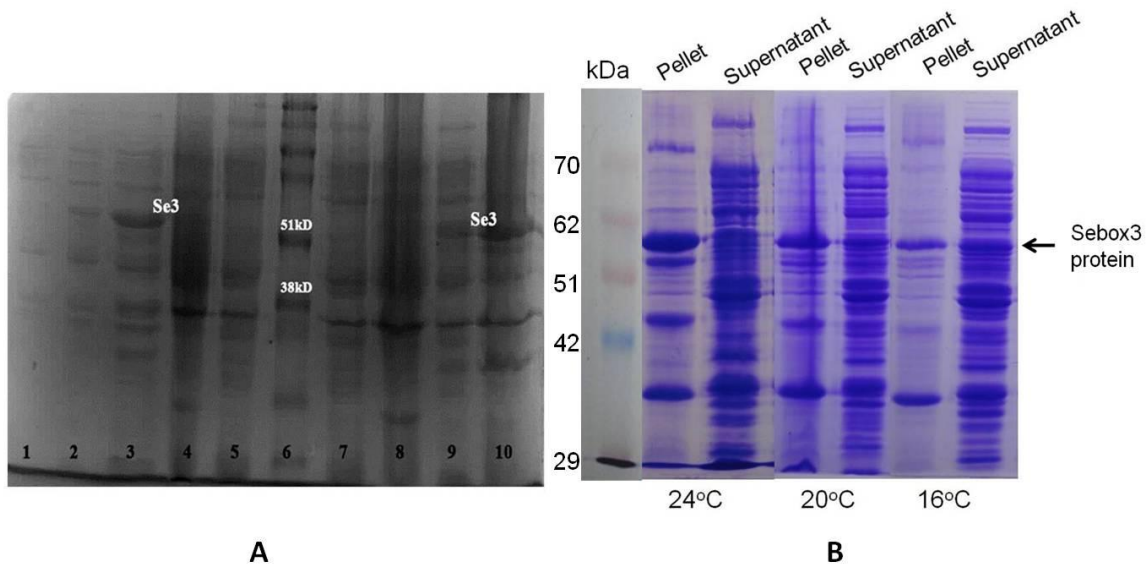


Figure 2.3: SDS-PAGE profile of pilot scale expression of VC0395_0300 in *E. coli*. Lanes 3, 4: Pellet and sup fractions of Sebox3 (abbreviated as Sebox3) in BL21, 6: ECL Plex Rainbow marker showing 51 kDa and 38 kDa protein bands; 9, 10: Pellet and sup fractions of Sebox3 in DH5 α . B) Small-scale expression standardization of Sebox3 in DH5 α at different temperatures.

2.3.2 Sebox3 protein expressed in soluble fraction

Generally, *E.coli* BL21 (DE3) strain is used as a host for recombinant protein expression. However, in this case, it was observed that the expressed Sebox3 protein was toxic to the host cells. As a result of which, only a few cells expressing the induced recombinant Sebox3 protein survived. Even in these cells, most of the protein was seen in the insoluble fraction. Any effort to enhance the solubility of the protein by modifying the culture conditions did not yield any positive result. According to previous literature [19], GGD(E)EF proteins were specifically toxic to *E.coli* BL21 strains and hence, protein expression was carried out in *E.coli* DH5 α cells (Figure 2.2A).

The yield of the protein in soluble fractions improved sufficiently when the cloning strain *E.coli* DH5 α was used for protein expression. However, a large fraction of the Sebox3 protein was still in the insoluble fraction (Figure 2.3).

Various parameters like induction temperature, IPTG concentration and duration of induction were tried to check for the possibility of improving solubility in the cell-free supernatant. Additionally, a combination of protease inhibitor cocktails was added during cell lysis to prevent the action of various proteases produced in the strain. Lower incubation period and, high induction temperature and IPTG concentrations did not aid in sufficient expression of the Sebox3 protein. Based on the study of these parameters, it was inferred that there was better yield in the soluble fractions after induction at 16°C with an IPTG concentration of 0.05 mM for 8 hours (Figure 2.2B). Low level of protein expression was obtained as seen in the SDS PAGE gel and very little of the total expressed protein was in the soluble fraction. Permutations and change in conditions did not show any remarkable improvement in the yield of the soluble fractions.

2.3.3 Protein purified as GST-Sebox3 fusion protein

The native Sebox3 protein with an amino-terminal GST was purified on GE Healthcare Glutathione SepharoseTM 4B affinity column from *E.coli* DH5 α cell lysate. Expression of the GST fusion protein was confirmed by immunoblotting and the signals emitted using ECL was captured on an X-ray film (Figure 2.4D). The purity of the protein was assessed on a 12.5% SDS-PAGE gel by Coomassie blue staining (Figure 2.4A and 2.4B). Also, it was noticed that the affinity-purified GST-tagged protein was accompanied by smaller bands. These bands did not appear when the sample loading buffer used, was devoid of β - mercaptoethanol indicating that the denaturation of the protein caused the appearance of these bands (Figure 2.4C).

Based on the purity and concentration, the eluted fractions were subjected to dialysis using 50mM Tris-HCl (pH 8.0) to thoroughly remove all traces of reduced glutathione. The concentration of protein was estimated by Bradford using Bovine Serum Albumin (BSA) as a standard and A_{280} .

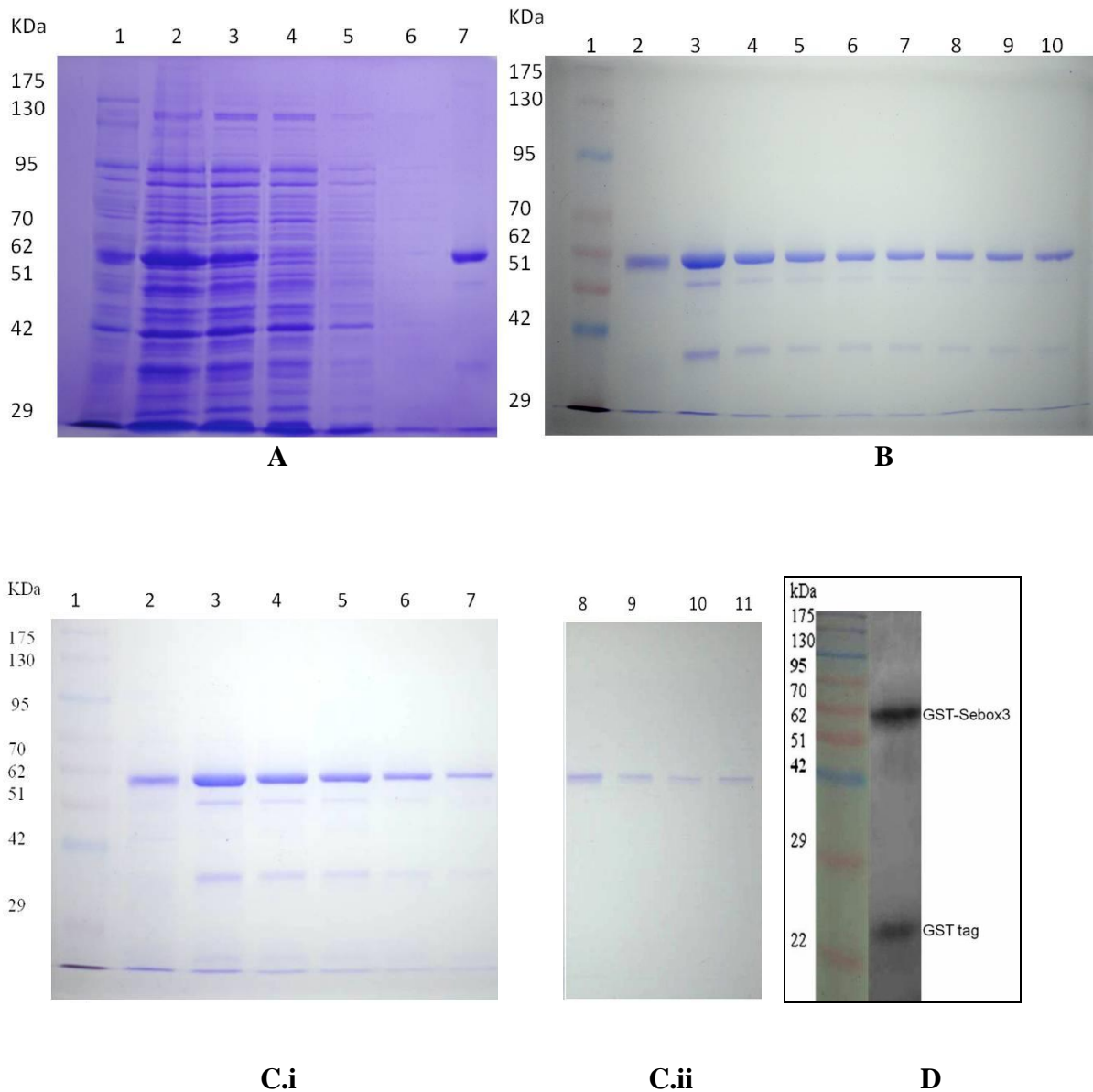


Figure 2.4: Expression profile of GST-tagged protein. A) lane 1: NEX-GEN PinkADD protein ladder, lane 2: crude extract, lane 3: supernatant, lane 4: flowthrough, lane 5, 6: column wash and lane 7: eluted GST-tagged Sebox3 protein. B) lane 1: Protein marker and lane 2-10 GST-tagged Sebox3 protein eluted with 10mM reduced glutathione (pH8.0). C.i) SDS-PAGE profile with β -mercaptoethanol; lane 1: protein marker, lane 2-7: GST-tagged Sebox3 protein with β -mercaptoethanol, C.ii) SDS-PAGE profile of Sebox3 without β -mercaptoethanol; lane 8-11: GST-tagged Sebox3 protein without β -mercaptoethanol. D) Western Blot showing anti-GST antibody using HRP (Horse Radish Peroxide)-conjugated secondary antibody.

2.3.4 Removal of GST tag using PreScission protease

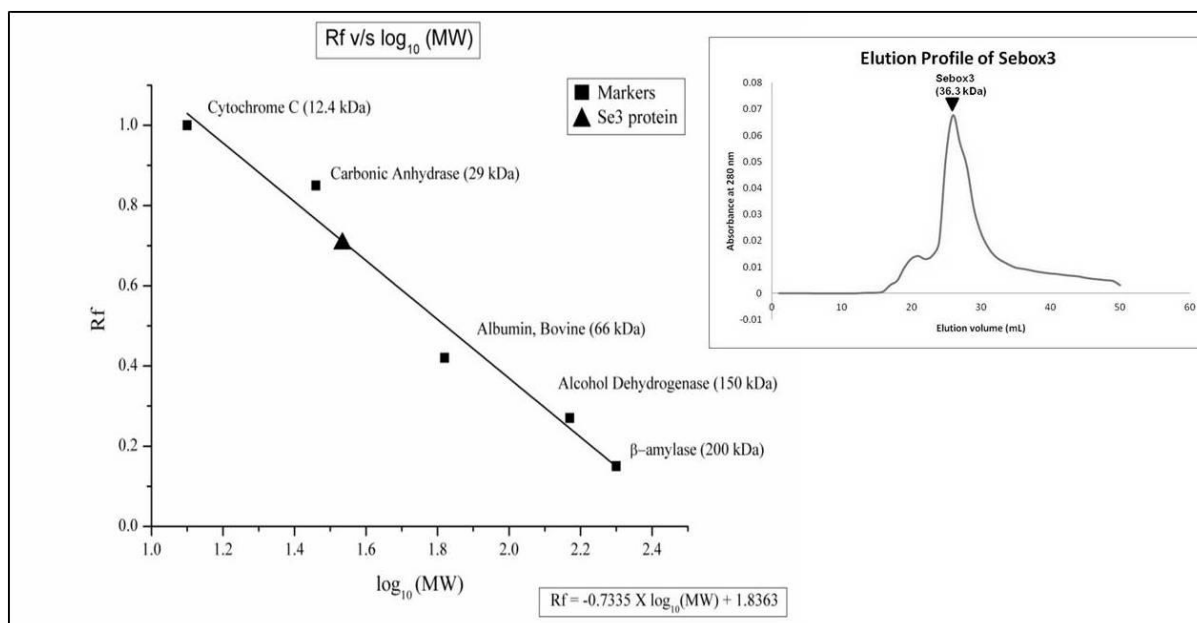
For cleavage of the GST tag from the protein, PreScission protease (10 units/mg tagged protein) was used which specifically identifies the cleavage site between the N-terminal GST tag and Sebox3 (Figure 2.5). Finally, both the tagged and untagged Sebox proteins were overexpressed, purified, quantified and were used for further biophysical studies.



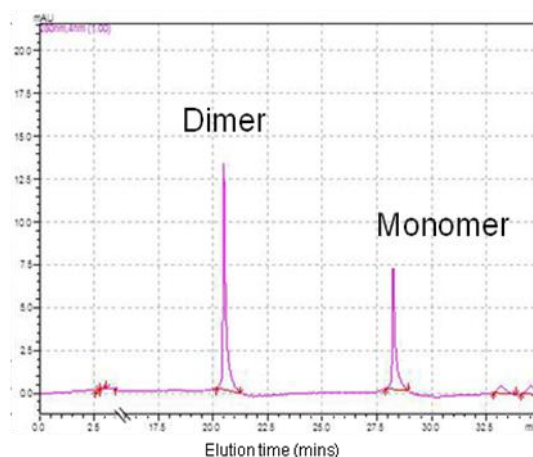
Figure 2.5: SDS-PAGE profile of tagless Sebox3 protein. Lane 1: NEX-GEN-PinkADD Prestained Protein Ladder, lane 2 and lane 5: empty, lane 3 and lane 4: tagless Sebox3 protein and GST tag and lane 6 – lane 9 purified tagless Sebox3 protein.

2.3.5 Tagless Sebox3 protein exists as a monomer in solution

The oligomeric status of protein was determined by analytical size-exclusion chromatography. The Superdex 200 column was calibrated using gel filtration markers kit consisting of albumin β amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum (66 kDa), carbonic anhydrase (29 kDa) and cytochrome c (12.4 kDa) (Figure 2.6A). The molecular weight of Sebox3 by gel filtration chromatography was found to be 36.3 kDa and is similar to the theoretical mass of 37.18 kDa obtained by Expasy ProtParam tool (section 2.2.4).



A



B

Figure 2.6: A) Analytical size exclusion chromatography for determination of molecular weight of Sebox3. The standard curve shows the plot of R_f values vs gel filtration markers. Inset: elution profile of tagless Sebox3 on Superdex 200 column. B) HPLC separation of GST tagged protein showing dimer-monomer equilibrium.

The freshly dialyzed GST-Sebox3 protein was separated on a C18 column with a gradient of 0-80% of solvent A: 5% acetonitrile in water with 0.1% TFA and solvent B: 95% acetonitrile in water with 0.1% TFA. The appearance of 2 peaks indicates the presence of monomers and dimers in solution (Figure 2.6B). The GST tag, by itself, is a dimeric protein and any protein fused to it is also likely to be a dimer in solution. If the fused protein interferes with the

dimerization then it can exist as a monomer. Thus, it can be predicted that the Sebox3 protein exists as a monomer without the GST-tag and, in a dimer-monomer in solution when fused to the GST-tag.

2.3.6 Unfolding of protein provides information on local environment of Trp residues

Tryptophan is an intrinsic fluorophore/fluorescent probe and its fluorescence acts as a tool to facilitate the understanding of protein folding to some extent. Trp has a characteristic emission at 295 nm and based on the location and the environment of the Trp residues on/within the protein, the fluorescence emission intensity would vary.

A protein will be biologically active only if it is folded properly. *In vivo* protein folding is a dynamic process which is difficult to monitor. Therefore, another contrary approach would be to study protein unfolding to understand the features that would contribute to the structural stability of the folded native protein. In order to examine the intricate details of protein folding, the protein needs to be unfolded first using thermal and chemical denaturation and then refolded *in vitro*.

However, protein unfolding cannot be always extrapolated to protein refolding as there are a lot of factors that influence *in vivo* protein folding. As it is difficult to replicate the exact conditions of *in vivo* protein folding, most of the studies have been limited to protein unfolding.

2.3.6.1 Unfolding studies on Sebox3 protein by thermal denaturation

At each temperature, the protein was incubated and allowed to reach equilibrium. Due to the heat absorbed, the weak interactions like hydrogen bonding, electrostatic interactions and hydrophobic interactions got disrupted. This lead to partial or complete unfolding of the protein thereby affecting the tryptophan fluorescence.

Only the fluorescence intensity decreases with increasing temperature with no apparent change in λ_{\max} for tryptophan fluorescence of GST-tagged and tagless Sebox3. The protein samples precipitated in the cell holder beyond the temperature of 70°C indicating complete denaturation (Figure 2.7).

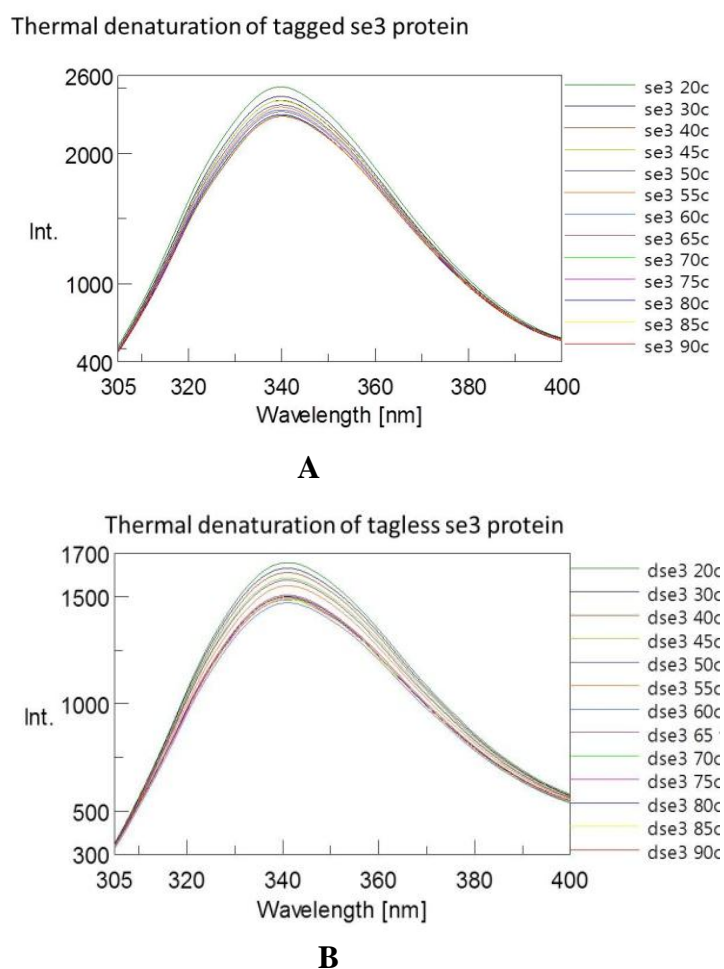


Figure 2.7: Thermal denaturation ranging from 20°C-90°C for A) Tagged Sebox3 protein B) Tagless Sebox3 protein.

2.3.6.2 Guanidine hydrochloride denatured and unfolded the protein

Guanidine hydrochloride is a strongly charged denaturant that binds to the protein and disrupts its highly ordered native structure. It was used to examine whether it could lead to the complete unfolding of the Sebox3 protein.

With increasing concentration of GdnHCl, the equilibrium shifted from native folded state to unfolded state of the protein. This, in turn, led to an increase in fluorescence intensity of the protein. The fluorescence signal (the emission wavelength and intensity) varies depending upon the local environment of the tryptophan residue which, in turn, also varies according to protein folding or unfolding. In the native folded protein, tryptophans are usually buried within the hydrophobic region resulting in low fluorescence signal. However, in a hydrophilic environment, fluorescent signal increases due to exposure of tryptophan to the solvent. Thus, the local environment of the tryptophan residue can be inferred from the changes in the fluorescence signal as a function of GdnHCl concentration. In the Sebox3 protein, two tryptophan residues are present at the 4th position and 172nd position in the polypeptide chain. The λ_{\max} of the tryptophan red-shifted as the protein unfolded. This could be due to exposure of the second tryptophan to the external polar environment.

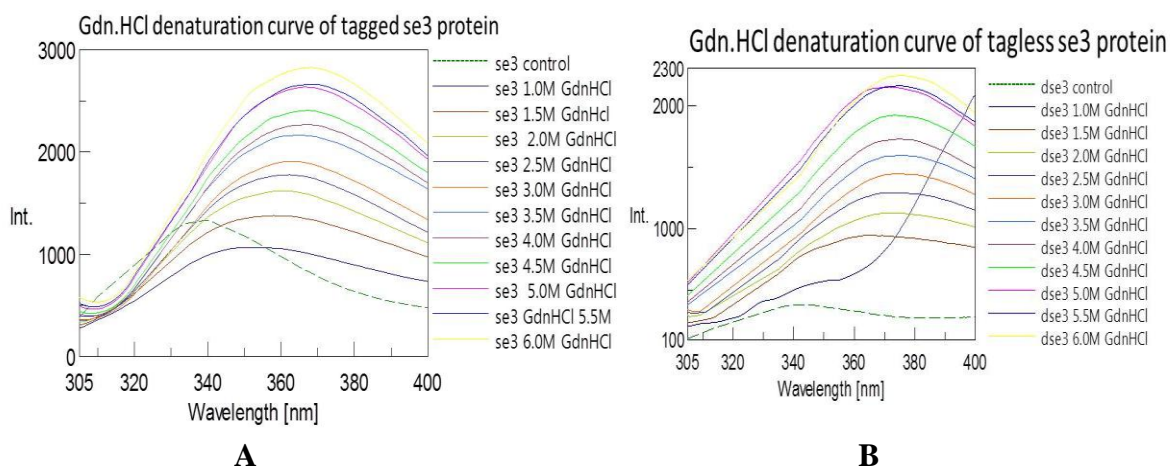


Figure 2.8: Guanidine hydrochloride-induced unfolding by utilizing Gdn.HCl concentrations ranging from 0 M-6 M for A) Tagged Sebox3 protein B) Tagless Sebox3 protein.

The difference in λ_{\max} of Trp fluorescence of native and denatured protein indicated that one of the buried tryptophan residues may be buried in the native folded protein and was gradually exposed to the solvent. This is accompanied by a red shift in λ_{\max} from 340 nm to 367.5 nm for GST-tagged protein and 341 nm to 375.5 nm for tagless protein (Figure 2.8). By

comparison, the shift in λ_{\max} for tryptophan fluorescence of GST-tagged and tagless Sebox3 on treatment with 6 M Gdn.HCl can be seen.

2.3.6.3 Quenching with acrylamide, KI and CsCl

Quenching of tryptophan fluorescence helps in finding the polarity of the local environment of the tryptophan as well as the solvent accessibility. Quenching of fluorescence was performed using neutral, negative and positive quenchers, that is, acrylamide, KI and CsCl, respectively. As acrylamide is small, uncharged and polar, it can easily react with any uncharged residue. In contrast, charged quenchers like CsCl and KI can bind and quench only when the oppositely charged residues are present in the local environment of the intrinsic fluorescent probe.

It was noted that acrylamide quenched tryptophan fluorescence in the protein more efficiently than KI while, CsCl was not so effective. This could be because acrylamide is a neutral quencher and can easily access the buried tryptophan residue and quench its fluorescence. On the contrary, KI has a strong negative charge which prevents it from diffusing into the hydrophobic core of the folded proteins. Fluorescence quenching by KI is, therefore, a good standard to suggest the presence of tryptophan residues on the protein surface.

The Stern-Volmer constant (K_{sv}) values were deduced from the slope of the Stern-Volmer plot. The regression analysis of all the linear plots shows that they are interrelated.

The following equation was used for calculation of K_{sv}

$$F_0/F=1+ K_{sv} \cdot [Q]$$

High values of (K_{sv}) of 1.978 and 0.827 for acrylamide and KI support the fact that there is at least one tryptophan residue which is exposed in the native protein.

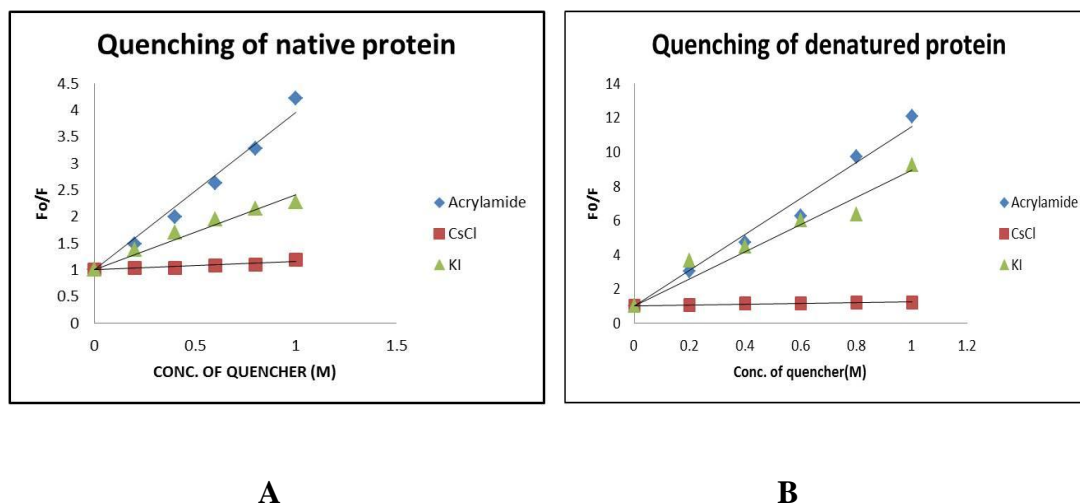


Figure 2.9: Stern-Volmer plots for quenching of tryptophan fluorescence of A) native Sebox3 protein and B) denatured Sebox3 protein in acrylamide, KI and CsCl is shown. Protein samples were incubated in increasing concentrations of quencher (up to 1 M). The excitation wavelength was 295 nm (for W). Straight lines have been drawn by linear regression of data points.

CsCl, on the other hand, is unable to quench tryptophan fluorescence as evidenced by a relatively low K_{sv} value of 0.0599 (Figure 2.9). This stipulates that the tryptophan residue is mostly unavailable to the quencher due to its location within a possible positively charged region of the protein. Quenching reaction in the case of charged quenchers like positively charged cesium and negatively charged iodide depended heavily on solvent accessibility and charge of the residues adjacent to the tryptophan. The utilisation of protein denaturants like GdnHCl enhances protein unfolding and increases the possibility of finding the fluorophore in the protein, thereby, increasing the rate of quenching. Hence, tryptophan quenching can also be used as a tool to study protein unfolding.

The quenching results are in agreement with the unfolding studies which suggest that one of the Trp residues lies on the surface and the other one lies within the hydrophobic core of the Sebox3 protein.

2.3.7 Sebox3 shows diguanylate cyclase activity

To determine the enzymatic activity of the Sebox3 protein, the purified protein was incubated *in vitro* with the substrate (GTP) and checked for the synthesis of c-di-GMP. The reaction mixture containing 5 μ M protein, 50 mM Tris-HCl (pH 8.0), and 10 mM MgCl₂ was pre-warmed at 37°C for 10 min before initiating the reaction. 50 μ M of GTP was added to the reaction mixture and incubated at 37°C for 45 min. The reaction was terminated by addition of 0.5 M EDTA and by boiling the entire reaction mixture for 5 min. Standard samples of GTP, c-di-GMP and GTP with c-di-GMP were also run to check their peak profiles in HPLC.

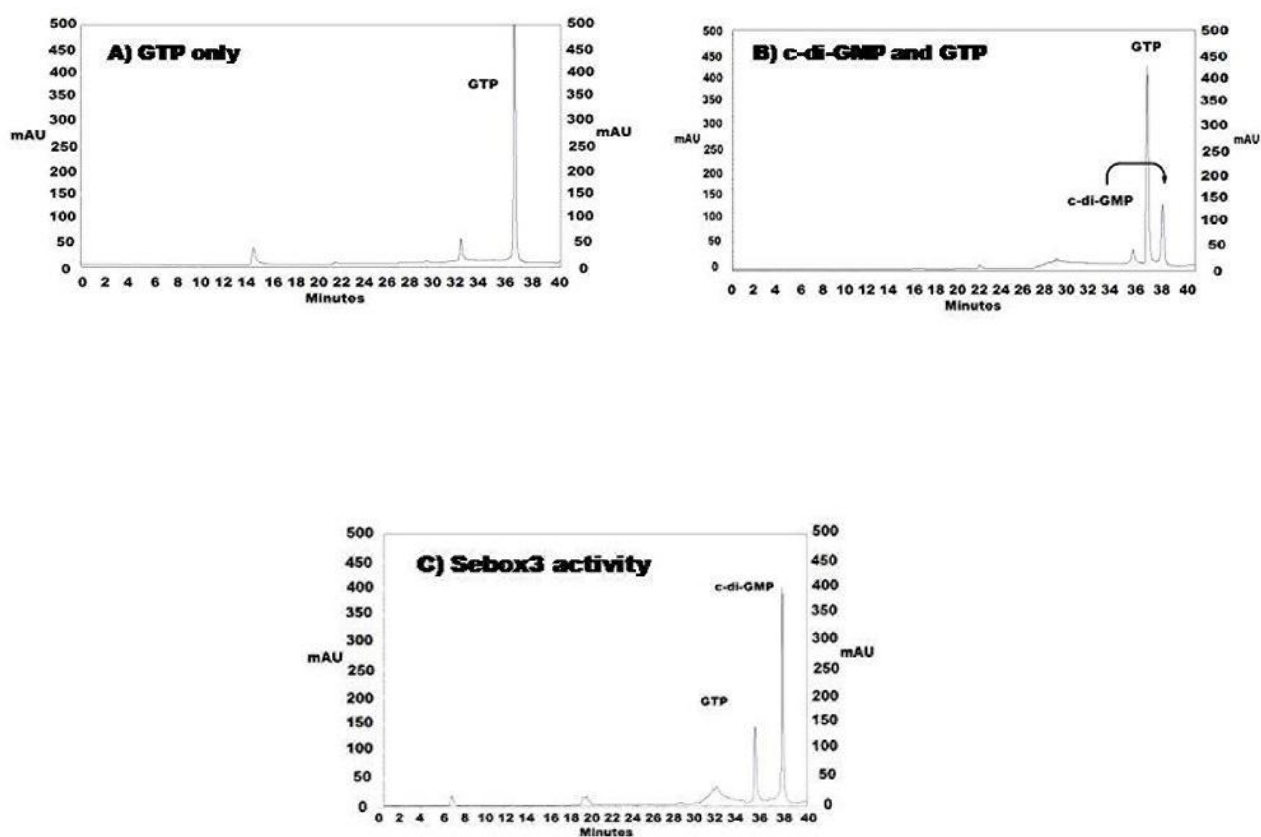


Figure 2.10: The Sebox3 protein with active GGEEF domain is a diguanylate cyclase: A) RP-HPLC profile of the GTP standard. B) GTP and c-di-GMP standards show retention time of 36.1 min and 38 min respectively. C) The diguanylate cyclase (DGC) reaction mixtures analyzed on C-18 column shows the utilization of GTP for the formation of c-di-GMP peak. 5 μ M protein was incubated with 50 μ M of GTP at 37°C as described in the experimental procedure. This shows that the GGEEF domain of the Sebox3 is active and has diguanylate cyclase activity.

The retention time (RT) of 36.1 min and 38 min was observed for GTP and c-di-GMP respectively (Figure 2.10A and 2.10B). In the chromatogram for Sebox3 protein sample, peaks were seen in the same position as the standard/control samples (Figure 2.10C). Thus, it can be concluded that Sebox3 protein has diguanylate cyclase activity.

2.3.8 Transverse Urea Gel Electrophoresis (TUGE)

Urea is used as a denaturing agent as it is a non-ionic reagent and does not interfere with the mobility of the molecules particularly proteins while running the gel electrophoresis. In this gel, the concentration of urea varies across the gel horizontally but remains constant vertically. This means that at any given time during the run, the molecules will be subject to the same. As the protein molecules move in the gel, if it comes across the denaturing condition then the protein will unfold retarding the migration of the protein. However, if the denaturant concentration is not very high, there will be an equilibrium between the reversible unfolded and folded state. Thus, the unfolded molecules will have a compact smaller size enabling it to move faster in the gel.

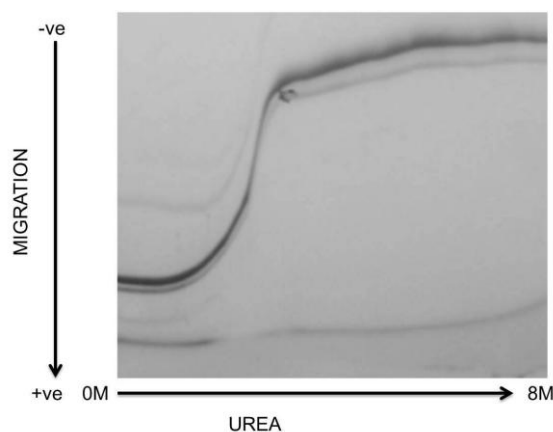


Figure 2.11: Transverse Urea Gel Electrophoresis (TUGE): TUGE gel of native Sebox3 protein denatured with urea (0 M – 8 M). sigmoidal shape of migration shows native protein in folded conformation followed by a reversible region (S-shaped) and the denatured unfolded state of protein at high concentrations of the protein.

In Figure 2.11, the concentration of the urea varied from 0 M to 8 M from left to right of the gel. In low concentrations of urea, it was observed that the Sebox3 protein retains its native folded state due to which it can rapidly migrate vertically towards the positive electrode. At moderate concentrations of urea, the protein rapidly unfolds in a reversible transition leading to a single point of inflection. At high concentrations of urea, the Sebox3 protein unfolds fairly. As a result of which, there is a lot of retardation in the migration of the Sebox3 protein. From this, it can be concluded that Sebox3 protein denatured by urea undergoes two-state transition: native and unfolded. The midpoint of unfolding transition appears to be 3 M urea and there are also no intermediates.

2.3.9 Detection of free thiol groups in Sebox3 protein

DTNB reaction was carried out to quantify the number of thiol groups in the given protein. The Sebox3 protein consists of 8 cysteine residues. DTNB reacts with free (-SH) groups to give a yellow-coloured product [2-nitro-5-thiobenzoic acid (NTB)] which has λ_{max} at 412 nm.

Three different conditions, that is, native protein, DTT-treated and GdnHCl-treated were used for accessing the number of free cysteine residues. The molar extinction coefficient at 412 nm (ϵ_{412}) of TNB was $13700 \text{ M}^{-1}\text{cm}^{-1}$ in phosphate buffer with GdnHCl and $14150 \text{ M}^{-1}\text{cm}^{-1}$ in phosphate buffer, respectively. Accessible cysteine residues were calculated according to the formula:

$$\text{Accessible cysteine (AC)} = (\text{OD}_{412}/\epsilon_{412}) / (\text{OD}_{280}/\epsilon_{280}),$$

Due to the inherent nature of the native protein, the value of DTNB reaction was 2.53. This shows that 3 cysteine residues were exposed to the solvent and the remaining sulfhydryl groups were either involved in disulphide bond formation or buried within the protein core. Next, the disulphide bonds of the thiol groups in the protein were reduced with 10 mM DTT.

The value of 4.85 indicates that the number of accessible cysteine residues after the breaking of disulphide bonds were 5. In case of treatment with 6 M Gdn.HCl, the protein was denatured completely and all the 8 cysteine residues (7.75) were available for the DTNB reaction.

Table 2.2 List of cysteine residues in the Sebox3 protein.

No. of cysteine residues	State of protein
2	native protein
5	after reduction with 10 mM DTT
8	after treatment with 6 M Gdn.HCl

2.3.10 Stable domains appear on treatment with proteolytic enzymes

It is well-known that proteases act on specific active sites on a given protein like the loops or flexible sites leaving the inaccessible compact regions intact. Thus, the Sebox3 protein was subjected to limited proteolysis by trypsin, chymotrypsin and papain in order to find the presence of stable domains in it. Partial proteolysis with trypsin, chymotrypsin and papain respectively resulted in the appearance of a fragment of ~8 kDa at 15 min which was stable during the entire duration of the digestion and was checked on the Tris-Tricine SDS PAGE gel (Figure 2.12).

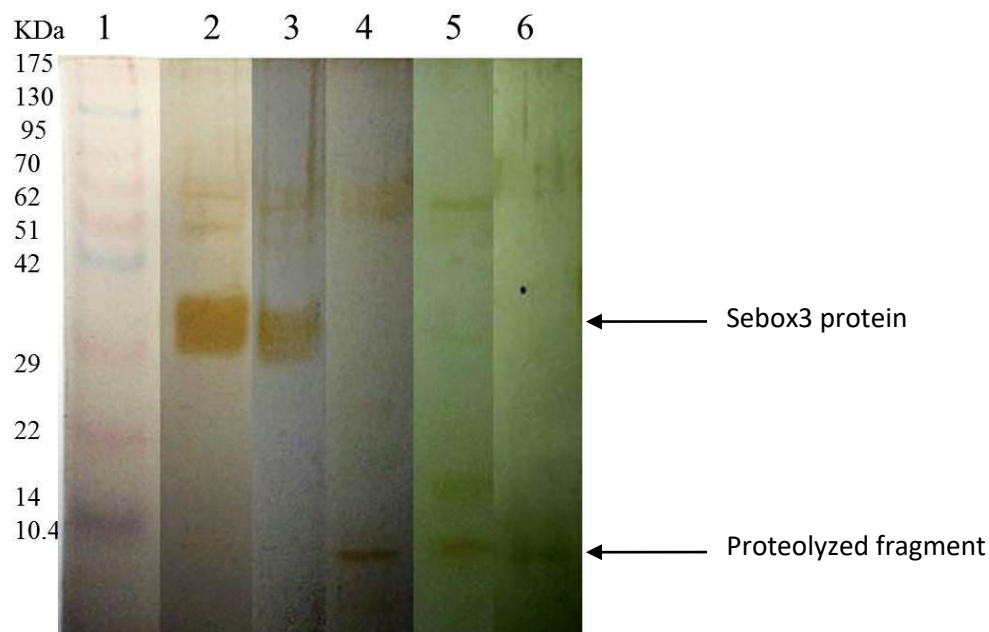


Figure 2.12: Limited proteolysis of Sebox3 protein: Lane1: NEX-GEN PinkADD protein ladder, lane 2: Sebox3 protein incubated at 25°C for 150 min, lane 3: Sebox3 protein incubated overnight at 25°C, lane 4: protein sample treated with chymotrypsin, lane 5: protein treated with trypsin and lane 6: protein treated with papain.

The proteolytic susceptibility of the Sebox3 protein indicates the presence of floppy sites within the protein. Finally, it can be concluded that proteolytic digestion shows the existence of a well-folded compact stable protein domain in the Sebox3 protein.

2.4 Conclusion

In this study, the GGEEF domain with an N-terminal PAS domain from *V. cholerae* O395 was cloned, expressed and purified as Sebox3 protein. It has been noted that GGD(E)EF domains are toxic to *E. coli* (BL21) and more prominently in pET vectors. Therefore, the pGEX-6P1 vector in DH5 α was used for soluble expression and purification of the Sebox3 protein. Proteins expressed in pGEX-6P1 are synthesized as GST fusion proteins. The major advantage of having the GST tag is that it enhances the solubility of the protein and offers mild elution conditions. However, GST exists as a dimer and due to its huge size of 26 kDa, it is necessary to remove the GST tag.

The GST tag leads to dimerization of the fused protein, as such; the protein existed as a heterogeneous population of dimer and monomer in solution, as seen in the HPLC chromatogram profile. Subsequently, the GST tag was removed to maintain homogeneity of protein. After the removal of the GST tag, by PreScission protease, the monomer was checked for oligomeric status using gel filtration. The tagless Sebox3 protein was found to be a monomer in solution with a molecular weight of 36.3 kDa.

This was followed by the study of protein folding and unfolding by checking tryptophan fluorescence. Protein was denatured using thermal and chemical methods like GdnHCl. During thermal denaturation, there was very little shift in the fluorescence signal irrespective of whether the Sebox3 protein was GST-tagged or tagless. Beyond the temperature of 70°C, the protein lost its ability to refold due to precipitation of the protein. Unfolding by strong chemical denaturant like GdnHCl led to the loss of the natively folded structure which resulted in a change in the fluorescence signal.

Quenching of tryptophan fluorescence was carried out using acrylamide (neutral), CsCl (positively charged) and KI (negatively charged). Both native tagless Sebox3 protein as well as the denatured tagless protein showed drastic variations. Since the denatured protein was unfolded it gave way to the charged quenchers. Acrylamide was found to be the most effective quencher followed by KI. In HPLC, the RT of 36.1 min and 38 min was observed for GTP and c-di-GMP respectively (Figure 2.10A and 2.10B). In the chromatogram for Sebox3 protein sample, peaks were seen in the same position as the standard/control samples. The protein shows its characteristic diguanylate cyclase activity as confirmed by HPLC assays.

In TUGE gels, the retarding effect of the unfolding of the protein due to urea denaturation on protein migration pattern under constant voltage can be seen. While three of the eight

cysteines in the protein were freely accessible, reduction of free sulfhydryl groups exposed two more, whereas complete denaturation alone revealed all eight. Limited proteolysis of Sebox3 pointed to the presence of a stable domain of approximately 8 kDa, which was confirmed by digestion with three different proteolytic enzymes.

In conclusion, Sebox3 is an enzymatically active diguanylate cyclase which exists in two-state, native and unfolded state during the protein denaturation process.

2.5 References

- [1] E. J. Nelson, J. B. Harris, J. G. Morris, S. B. Calderwood, and A. Camilli, “Cholera transmission: the host, pathogen and bacteriophage dynamic.,” *Nat. Rev. Microbiol.*, vol. 7, no. 10, pp. 693–702, 2009.
- [2] J. Reidl and K. E. Klose, “*Vibrio cholerae* and cholera: Out of the water and into the host,” *FEMS Microbiol. Rev.*, vol. 26, no. 2, pp. 125–139, 2002.
- [3] R. C. Charles and E. T. Ryan, “Cholera in the 21st century.,” *Curr. Opin. Infect. Dis.*, vol. 24, no. 5, pp. 472–477, 2011.
- [4] O. Lieleg, M. Caldara, R. Baumgärtel, and K. Ribbeck, “Mechanical robustness of *Pseudomonas aeruginosa* biofilms.,” *Soft Matter*, vol. 7, no. 7, pp. 3307–3314, 2011.
- [5] H. Flemming and J. Wingender, “The biofilm matrix,” *Nat. Rev. Microbiol.*, vol. 8, no. 9, pp. 623–33, 2010.
- [6] H. C. Flemming, T. R. Neu, and D. J. Wozniak, “The EPS matrix: The ‘House of Biofilm Cells,’” *J. Bacteriol.*, vol. 189, no. 22, pp. 7945–7947, 2007.
- [7] I. W. Sutherland, “The biofilm matrix - An immobilized but dynamic microbial environment,” *Trends Microbiol.*, vol. 9, no. 5, pp. 222–227, 2001.

- [8] E. E. Mann and D. J. Wozniak, “*Pseudomonas* biofilm matrix composition and niche biology,” *FEMS Microbiol. Rev.*, vol. 36, no. 4, pp. 893–916, 2012.
- [9] P. Ross *et al.*, “Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid,” *Nature*, vol. 325, no. 6101, pp. 279–281, 1987.
- [10] C. Chan *et al.*, “Structural basis of activity and allosteric control of diguanylate cyclase,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 49, pp. 17084–17089, 2004.
- [11] N. A. Bhuiyan *et al.*, “Genetic characterization of *Vibrio cholerae* O1 strains isolated in Zambia during 1996–2004 possessing the unique VSP-II region of El Tor variant,” *Epidemiol. Infect.*, vol. 140, no. 3, pp. 510–518, 2012.
- [12] S. Harper and D. W. Speicher, “Expression and purification of GST fusion proteins,” *Curr. Protoc. Protein Sci.*, no. SUPPL. 52, 2008.
- [13] N. K. Vinckier, A. Chworos, and S. M. Parsons, “Improved isolation of proteins tagged with glutathione S-transferase,” *Protein Expr. Purif.*, vol. 75, no. 2, pp. 161–164, 2011.
- [14] M. Bradford, “Rapid and Sensitive Method for Quantification of Microgram Quantities of Protein utilizing principle of Protein-Dye-Binding,” *Anal. Biochem.*, vol. 72, no. 1–2, pp. 248–254, 1976.
- [15] A. a S. Takeda *et al.*, “Biophysical characterization of the recombinant importin- α from *Neurospora crassa*,” *Protein Pept. Lett.*, vol. 20, no. 1, pp. 8–16, 2013.
- [16] J. T. Vivian and P. R. Callis, “Mechanisms of Tryptophan Fluorescence Shifts in Proteins,” *Biophys. J.*, vol. 80, no. 5, pp. 2093–2109, 2001.
- [17] S. R. Phillips, L. J. Wilson, and R. F. Borkman, “Acrylamide and iodide fluorescence

- quenching as a structural probe of tryptophan microenvironment in bovine lens crystallins.," *Curr. Eye Res.*, vol. 5, no. 8, pp. 611–619, 1986.
- [18] M. R. Eftink and C. A. Ghiron, "Fluorescence quenching studies with proteins.," *Anal. Biochem.*, vol. 114, no. 2, pp. 199–227, 1981.
- [19] D. A. Ryjenkov, M. Tarutina, O. V. Moskvina, and M. Gomelsky, "Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain," *J. Bacteriol.*, vol. 187, no. 5, pp. 1792–1798, 2005.
- [20] L. M. Ruiz, M. Castro, A. Barriga, C. A. Jerez, and N. Guiliani, "The extremophile *Acidithiobacillus ferrooxidans* possesses a c-di-GMP signalling pathway that could play a significant role during bioleaching of minerals," *Lett. Appl. Microbiol.*, vol. 54, no. 2, pp. 133–139, 2012.
- [21] D. P. Goldenberg and T. E. Creighton, "Gel electrophoresis in studies of protein conformation and folding.," *Anal. Biochem.*, vol. 138, no. 1, pp. 1–18, 1984.
- [22] T. Schirmer and U. Jenal, "Structural and mechanistic determinants of c-di-GMP signalling.," *Nat. Rev. Microbiol.*, vol. 7, no. 10, pp. 724–735, 2009.
- [23] F. Zappacosta *et al.*, "Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody.," *Protein Sci.*, vol. 5, no. 5, pp. 802–13, 1996.
- [24] A. Fontana, P. P. De Laureto, B. Spolaore, E. Frare, P. Picotti, and M. Zambonin, "Probing protein structure by limited proteolysis," in *Acta Biochim. Pol.*, vol. 51, no. 2, pp. 299–321, 2004.
- [25] C. Dian *et al.*, "Strategies for the purification and on-column cleavage of glutathione-

S-transferase fusion target proteins,” *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 769, no. 1, pp. 133–144, 2002.

CHAPTER 3

Construction of Truncates of Sebox3: Sebox31 and Sebox32; and their Functional Characterization

3.1 Introduction

The determination of solubility and stability of a recombinant protein is a challenging task and generally requires knowledge of the behaviour of proteins. For protein structure determination, huge amounts of homogenous purified soluble protein are required and therefore having constructs of the same protein would facilitate the process of obtaining stable soluble protein suitable for structural characterization of these proteins. Removal of disordered regions from the full-length protein significantly enhances the possibility of obtaining single crystals. This is because the loops/disordered regions make the protein unstable and insoluble lowering the overall yield of the protein. Also, it lowers the possibility of nucleation by precipitating the protein. The function and interdependence of individual domains within a protein can also be studied by isolating the domains and checking their activity individually. Different approaches can be selected in making constructs like secondary structure prediction, protein solvation properties, protein domain prediction, protein stability and even crystallization properties of the protein [1]. Although simultaneous screening of a large number of constructs is time-consuming, it enables one to identify a protein construct that can be successfully used in structural characterization [2].

In this chapter, truncated versions (Sebox31 and Sebox32) of the full-length protein Sebox3 protein with the GGD(/E)EF domain were created. Such constructions would usually lead to the production of stable and soluble recombinant proteins with increased chances of protein crystallization due to the absence of the disordered regions. Consequently, it would provide valuable information on the catalytic mechanism of the GGD(/E)EF domain protein.

The GGD(/E)EF domain containing proteins with diguanylate cyclase activity are known for their role in biofilm formation. These proteins encode for diguanylate cyclase activity thereby increasing the level of c-di-GMP by conversion of GTP molecules. The resultant c-di-GMP has been known

to prominently influence characteristics like biofilm formation, virulence and motility of bacteria. In this study, the role of c-di-GMP on the strains containing the full-length and truncated GGEEF protein was analysed. Finally, the truncated constructs were checked for the retention of the activity in spite of the deletions in the full-length protein.

3.2 Materials and Methods

3.2.1 Domain organisation

Sebox3 (aminoacids 1-321), Sebox31 (aminoacids 130-321) and Sebox32 (aminoacids 161-321) protein were created based on the secondary structure prediction and protein solvation properties obtained from GeneSilico Metaserver 2.0 (Figure 3.2 and Figure 3.3). This server integrates protein secondary structure from various prediction methods into a single consensus sequence.

Table 3.1: List of strains used in the study

Strain Name	Relevant Characteristics	Source of Reference
pGEX-6P1	Tac promoter, <i>lacI^q</i> gene, PreScission Pro cleavage site	[4]
OrgB	<u>pCMW75</u> + diguanylate cyclase gene <i>Vibrio harveyi</i>	[5]
Sebox3	pGEX-6p1 + <i>sebox3</i> gene	In this study
Sebox31	pGEX-6p1 + <i>sebox3</i> ₁₃₀₋₃₂₁ fragment	In this study
Sebox32	pGEX-6p1 + <i>sebox32</i> ₁₆₁₋₃₂₁ fragment	In this study

3.2.2 Plasmid construction and cloning

The *Vibrio cholerae* O395 genomic DNA was isolated and purified with minor modifications according to the method previously described [6]. This genomic DNA was used as a template for subsequent PCR amplifications. The DNA corresponding to the putative GGEEF domain (VC0395_0300) from *Vibrio cholerae* chromosome I, amino acids 1-321 (Sebox3), was amplified and cloned into the pGEX6P1 vector (GE Healthcare) [as described in section 2.2.2]. The truncated versions of VC0395_0300, amino acids 130-321 (Sebox31) and amino acids 161-321 (Sebox32) were PCR amplified with Pfu DNA polymerase (Thermo Scientific) and cloned in the same vector. The gene-specific forward (F) and reverse (R) primers for the PCR reaction were designed based on the nucleotide sequence of the genes in the *Vibrio cholerae* genome.

The following primers were used: Sebox3 – Sebox1AF (Bam HI, Xho I) 5'AATACTGGATCCATGAAAAATTGGCTGTGTCAGGCAGTG 3' and Sebox3AR 5'AATACT CTCGAGTTATTCTGTGGATTGGCGATAGATACA 3', Sebox31F - 5'ATACGCGGATCCATGACAGGCGAAGTGATTGGGCTGATT 3' and Sebox32F - 5' ATACGCGGATCCATGTCTTTAACTCAGCTGTGTAATCGG 3'. The names of the restriction enzyme are within parentheses, and the underlined portions correspond to the sequence in the primers.

The PCR amplicons were purified using QiAquick kit from Qiagen, digested using Bam HI and Xho I and cloned into multiple cloning sites of the vector pGEX-6P-1 (GE Healthcare) and transformed into *E. coli* DH5 α . The recombinant plasmid was confirmed by double digestion with Bam HI and Xho I and subsequently propagated in *E. coli* BL21 (DE3) Novagen.

3.2.3 Protein expression, purification and GST tag removal

The protein was purified, expressed and quantified according to the protocol mentioned in section 2.2.3, 2.2.4 and 2.2.5 respectively.

3.2.4 Assays for formation of biofilms

3.2.4.1 Crystal Violet Assays

Crystal violet assay was performed to estimate the biofilm formed by the cultures of pGEX-6P1 (negative control), QrgB (positive control) and the Sebox3, Sebox31 and Sebox32. For each experiment, a single colony was inoculated in 5 mL Luria Bertani (LB) broth and grown at 37°C with overnight shaking. From the overnight cultures, 1% of inoculum was transferred into fresh 5 mL LB broth in glass tubes (18 X 150 mm) with 100 µg/mL ampicillin and 0.05 mM IPTG and incubated for 12 hours on shaking at 37°C. After the incubation period, all the tubes were kept stationary for 7 days without any disturbance to allow biofilm formation at the air-liquid interface on the inner surface of the borosilicate tubes. Three replicates for each culture under each condition were performed.

For crystal violet staining, the liquid culture was drained from the tubes and the attached bacterial cells were washed with distilled water and stained with 0.2% crystal violet at room temperature for 5 min. The tubes were then carefully washed with distilled water to remove any traces of unbound crystal violet and were dried at room temperature. Subsequently, the crystal violet bound biofilm was dissolved in 4 mL of 75% ethanol from 10 min and OD at 570 nm was measured using a Shimadzu UV-visible spectrophotometer UV-2450.

3.2.4.2 Scanning electron microscopy (SEM):

For SEM analysis, pGEX-6P1 (negative control), QrgB (positive control), Sebox3, Sebox31 and Sebox32 strains were grown overnight on shaking in 5 mL LB broth at 37°C and 1% of the overnight grown inoculum was introduced into fresh media next day and allowed to grow till saturation. Sterile glass coverslips (20 mm X 20 mm) were gently inserted into the static cultures under sterile conditions. To enhance biofilm on the coverslips at the air-liquid interface, the culture tubes were incubated without any disturbance at 37°C. After a period of 7 days, the coverslips with the bound biofilm were gently removed and washed with sterile PBS (pH 7.4) to remove any unattached cells. Prior to observation, the biofilm formed was fixed with 2.5% glutaraldehyde for 20 min. Next, the coverslips with the fixed biofilm were rinsed gently with PBS.

The fixed biofilm was then gradually dehydrated in a graded ethanol series (20 to 100% for 20 min at each concentration). Thereafter, the coverslips were trimmed and dried further in Critical Point Dryer (CPD) using liquid carbon dioxide (EMITECH, K850). The CPD chamber with the coverslips was cooled to a temperature of 15°C during CO₂ exchange into the chamber and the slowly heated to 35°C. Carbon dioxide was slowly vented out of the system to enable controlled drop of pressure at the end of the critical drying process. The dried coverslips were mounted onto stubs with the aid of adhesive, double-sided, conductive carbon tapes, and sputter coated with gold under vacuum. Field Emission Scanning Electron Microscope (FE-SEM QUANTA 200 FEG, Netherlands) was utilized to visualize the surface of the biofilms formed by the cultures.

All the samples were examined by the secondary electron emission mode and high vacuum mode with an accelerating voltage of 15 kV. Images were captures at magnifications of 5000X, 7000X and 10000X.

3.2.5 Congo Red Assay

3.2.5.1 Congo Red uptake by cells

All the cultures were grown in 5 mL Luria Bertani (LB) broth overnight on shaking with 100 µg/mL of antibiotic. These cultures were inoculated into fresh LB media and allowed to grow at 37°C till the OD₆₀₀ of 0.5 was reached. Each culture was streaked onto LB agar plates containing 5 µg/mL Congo Red (CR) and 0.05 mM IPTG). Plates were incubated for 4 days at 37°C [11–15]. The red colour of the streaked culture indicated uptake of Congo Red dye from the bacterial media.

3.2.5.2 Quantitative estimation of Congo Red binding

For quantitative estimation of Congo Red, the uptake of Congo Red (CR) is determined using a modified protocol [16]. All the 5 strains were incubated in 5 mL of LB containing the required antibiotic and 0.05 mM IPTG for 24 hours at 37°C at 120 rpm. After 24 h, 1 mL of bacterial culture was centrifuged and resuspended in 1 mL of 40 µg/mL Congo Red in 1% tryptone and incubated for 2 h at 37°C at 180 rpm.

The bacteria with bound Congo Red were centrifuged at 6,000 rpm for 10 min and the amount of Congo Red remaining in the supernatant was estimated by measuring the OD₄₉₀ of the supernatant. One percent tryptone solution was used for blank correction. The readings were compared with the OD₄₉₀ of the Congo Red standard solution to determine the amount of bound CR.

3.2.6 Motility Assay

3.2.6.1 Soft agar plate assay

LB soft-agar motility plate with 0.35% (w/v) agar supplemented with 0.05 mM IPTG (isopropyl-β-D-thiogalactopyranoside) was used to determine the motility of bacterial strains [17]. 5 µL of culture from overnight grown LB broth was spot inoculated on the plates and incubated for 16

hours at 37°C. The diameter of the migration zone from the spot of inoculation was measured in mm after 18 h of incubation at 37°C [18–20].

3.2.6.2 TTC assay

Motility test was performed using LB agar medium containing triphenyl tetrazolium chloride (TTC) as adapted from [21–23]. TTC is a colourless water-soluble compound in its oxidized form. During growth of bacteria in media containing TTC, the compound gets absorbed by the bacterial cells and gets reduced to an insoluble red coloured compound.

Experimentally, a single colony was stab inoculated into a test tube containing LB agar with TTC incubated at 37°C for 48 hours. Red coloured formazan radiating away from the line of stab indicated bacterial growth.

3.3 Results and Discussion

3.3.1 Domain Organization

Based on the secondary structure prediction, it can be speculated that the full-length Sebox3 protein (aminoacids 1-321) has flexible loops in the N-terminal region. Therefore, Sebox31 (aminoacids 130-321) and Sebox32 (aminoacids 161-321) proteins with shortened N-terminal were created retaining the functional part of the full-length protein.

InterProScan groups proteins into families based on the functional analysis of the proteins. It uses signatures from many different databases like PFAM, PROSITE, TIGRFAM, PRODOM, HAMAP, PRINTS, PIR, SUPERFAMILY, SIGNAL P, SMART, TMHMM, PANTHER, PROFILE and GENE3D and annotates protein sequences into functional groups. InterProScan clubs member databases which utilize various methodologies to obtain information of well-characterized proteins and produce a protein signature. For example, Pfam classifies protein sequences based on divergent

domains, PROSITE focuses on functional sites and the function of TIGRFAM revolves around building Hidden Markov Models of functionally similar proteins.

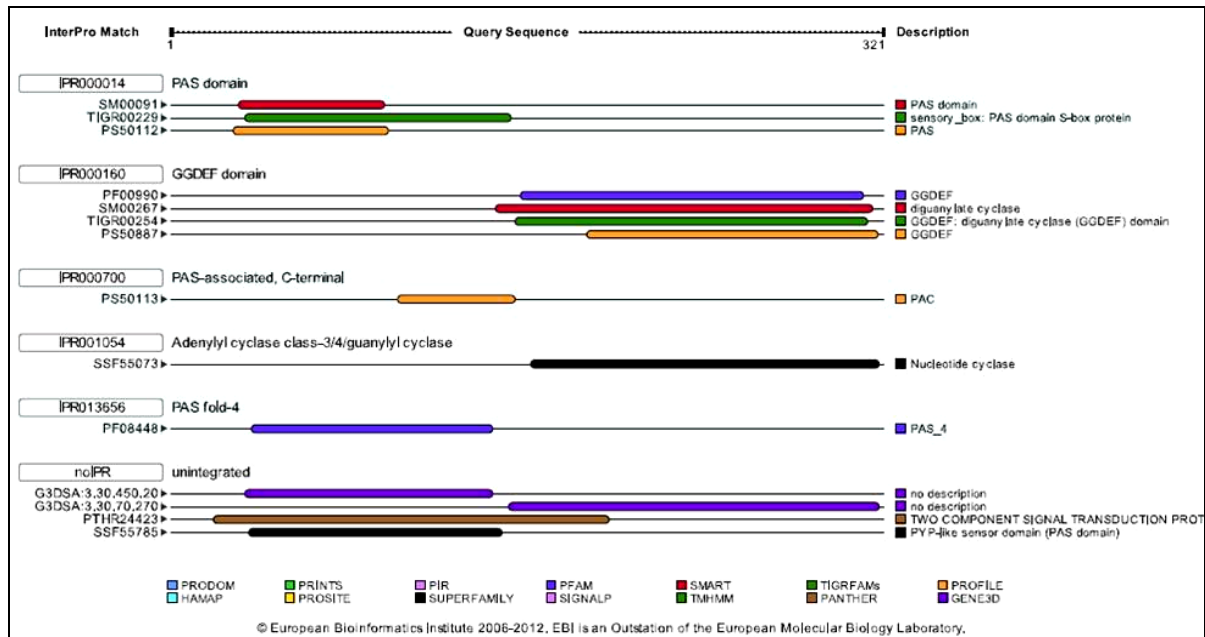


Figure 3.1: Interproscan results showing GGDEF domain in the sequence of the Sebox3 protein with PAS sensory domain at the N-terminal.

Although most of the GGDEF proteins are present as multi-domain proteins in conjugation with regulatory domains like sensory PAS domain, some are also present individually. InterProScan distinctly shows the presence of two domains, that is, GGDEF and PAS domain. From the InterProScan, it can be seen that the GGDEF domain with diguanylate cyclase activity is predominantly located at the C-terminal of the full-length protein whereas; the sensory domain is located at the N-terminal (Figure 3.1). The constructs of the Sebox3 proteins were created based on the consensus secondary structure prediction and protein solvation properties (Figure 3.2 and Figure 3.3).

3.3.2 Secondary structure prediction and protein solvation

GeneSilico Metaserver 2.0 provides the secondary structure of the protein based on the consensus results of prediction programs like sspro4, sspal, jnet, proteus, *SPARROW, sspreed, SPARROW,

sable, prof, nnssp, netsurfp, ssp, pasfinder, raptorxss, psspred, spineX, spine, psipred and soprano.

In the figure 3.2, H denotes helical, E denotes extended and – denotes other conformation.

SECONDARY STRUCTURE PREDICTION

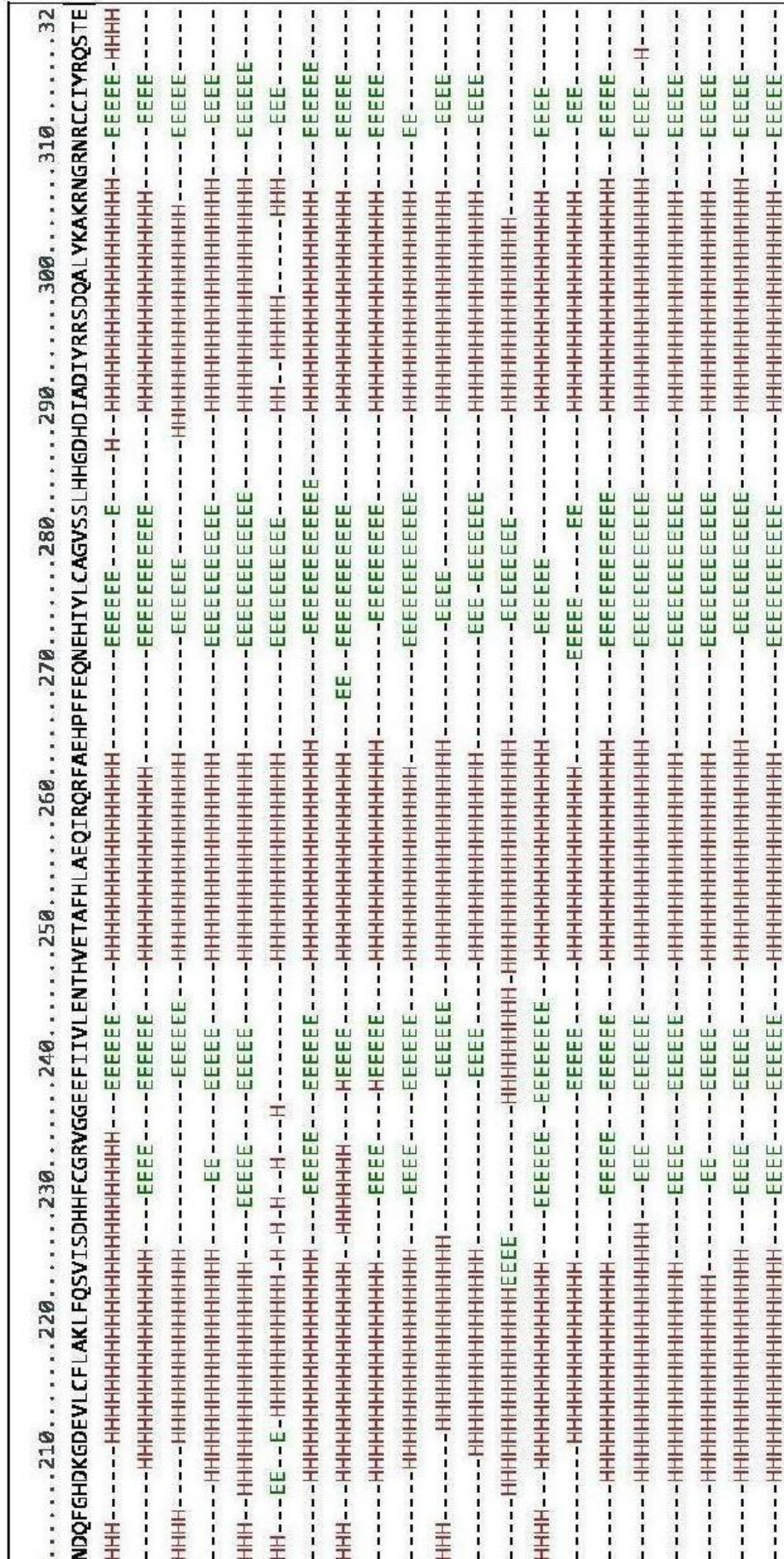
```

.....110.....120.....130.....140.....150.....160.....170.....180.....190.....200
HEERATAKSNGLVRIYRAVKHPILHRVTGEVIGLIGVSTDITDIVELREQLYQLANTDSL TQLCNRKLNADFRRAFAKAKRLRQPLSCISIDIDNFKLI
--EEEE--EEHHHEEEEE--EEEEEE--HHHHHHHHHHHHHHHHHH-- -- -- --HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
HHHHHH--HHHH--EE--HHHHHHHHHHHHHHHH--HHHHHH--HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
EEEEEE--EEEEEE--E--EEEEEE--HHHHHHHHHHHHHHHH-- -- -- --HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
--EEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
HHHHHH--HHHHHH--EEEE--HHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
HHHHHH--HHHHHHHH-- -- -- --EEEEHHHHHHHHHHHHHHHH-- -- -- --HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
HHHHHHHHHHHHHHHHHHHH-- -- -- --EEEE--HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
--EEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--
EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--EEEEEE--

```

B

Figure 3.2 (I): B Consensus secondary structure prediction using GeneSilico Metaserver 2.0 showing amino acids 101-200. The N-terminal region between aminoacids 1-129 is rich in flexible regions, therefore during synthesis of constructs this region was eliminated.



C

Figure 3.2 (D): C Consensus secondary structure prediction using GeneSilico Metaserver 2.0 showing amino acids 200 - 321. The N-terminal region between aminoacids 1-129 is rich in flexible regions, therefore during synthesis of constructs this region was eliminated.

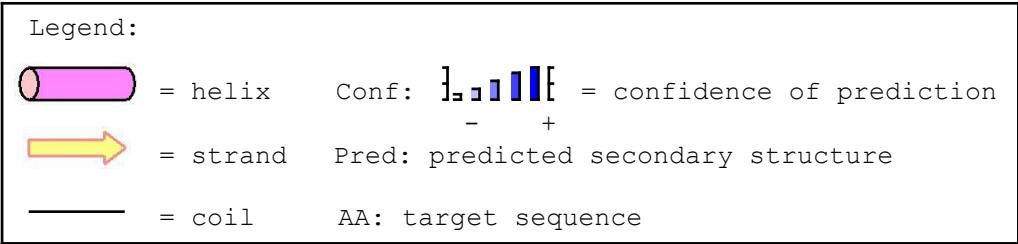
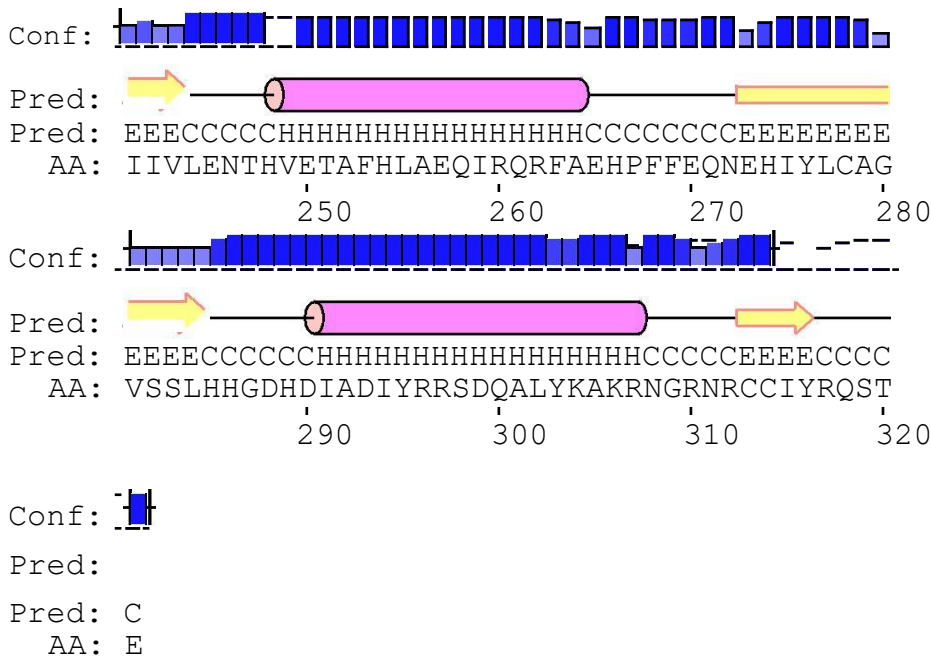
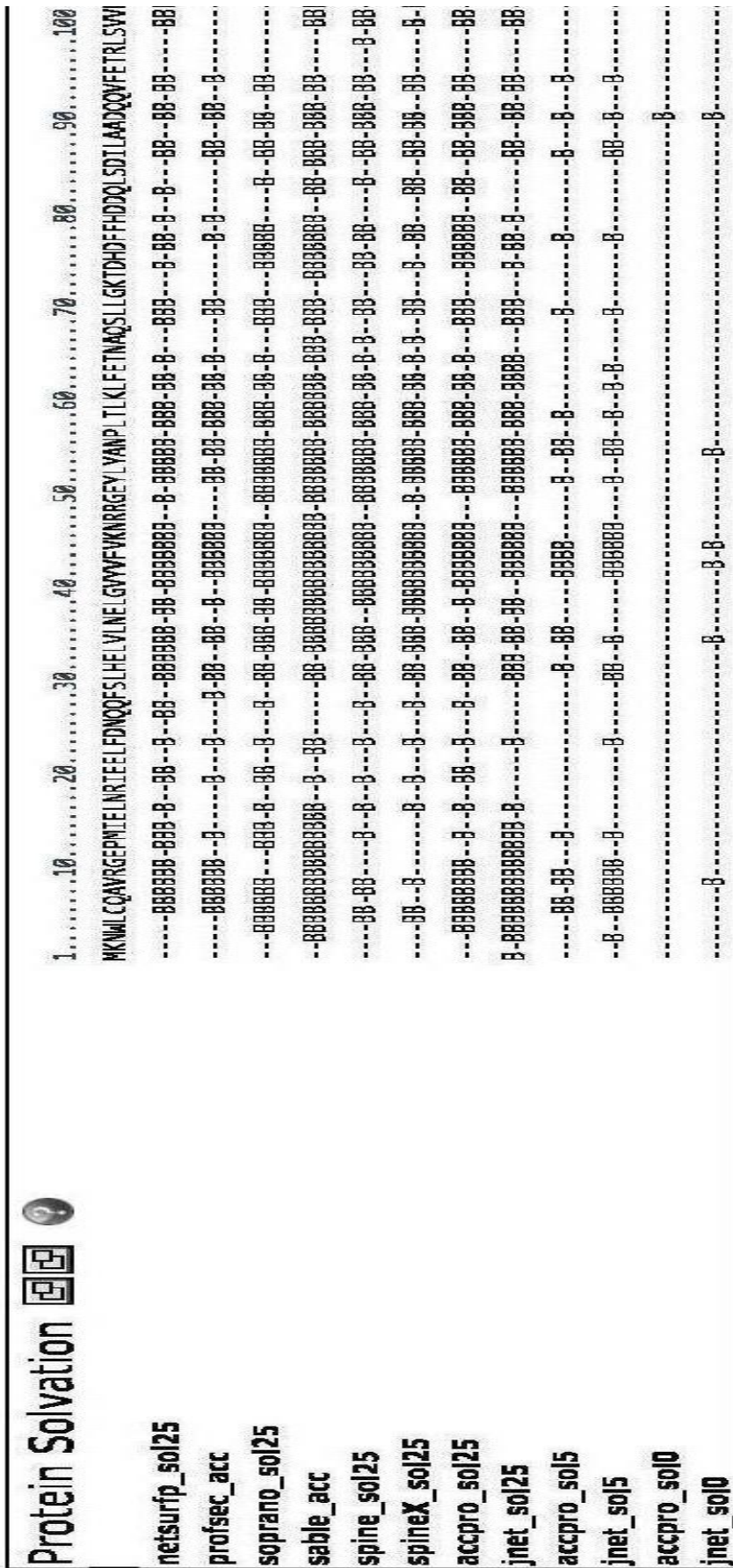


Figure 3.2 (II): B) Secondary structure prediction of Sebox3 protein (amino acid 241-321) by Psipred, which uses results obtained from the PSI-BLAST algorithm.

From the consensus secondary structure, it appears that there are small stretches of disordered regions occurring at the N-terminal of the full-length Sebox3 protein. This information is useful identifying regions in the gene that can be expressed in soluble form in bacterial expression systems and aid in successful crystallization of the required protein. When the full-length Sebox3 protein was expressed in *E. coli* strains, its yield was low as compared to the Sebox31 and Sebox32 constructs. This data supports the fact that proteins with disorder/unstructured region are generally unstable and degrade easily. It can be concluded that due to the removal of considerable size of the N-terminal in the protein constructs, they are stable, soluble and produce a higher yield.

Protein Solvation prediction was carried out to determine the exposure of 321 individual amino acids. The position of the 2nd tryptophan residue (W) in the Sebox3 protein was of prime importance. According to the protein solvation prediction, the tryptophan at the 4th position was solvent exposed whereas, the tryptophan at the 172nd position appeared to be buried within the hydrophobic core of the protein.



A

Figure 3.3: A Protein Solvation properties of the Sebox3 protein from amino acids 1 - 100 shows the position of buried and exposed amino acid.

3.3.3 Cloning, expression and purification of Sebox31 and Sebox32

The gene fragment of sebox 31 and sebox32 was PCR amplified by using Pfu polymerase (Figure 3.4). These PCR amplicons were gel purified before proceeding for restriction digestion with Bam HI and Xho I.

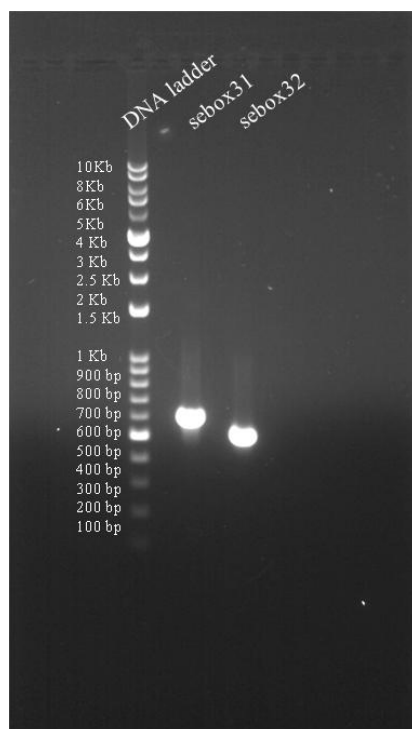


Figure 3.4: PCR amplification of sebox31 and sebox32 on 1% agarose gel. From left to right: Puregene Genetix NEX-GEN 10 Kb DNA ladder, sebox31 PCR product and sebox32 PCR product.

The gene encoding the GGEEF protein was cloned into the pGEX-6P1 vector. Glutathione-S-transferase (GST) tag was used as it enhances the solubility of the protein and offers the use of mild elution conditions which preserves the activity of the protein. The purity of the protein was assessed on a 12.5% SDS-PAGE gel stained by Coomassie. Based on the purity and concentration, the eluted fractions were subjected to dialysis using 50 mM Tris-HCl (pH

8.0) to thoroughly remove all traces of reduced glutathione. The concentration of protein was estimated by Bradford and A_{280} .

For cleavage of the GST tag from the protein, PreScission protease was used which specifically identifies the cleavage site between the N-terminal GST tag and Sebox proteins. Finally, both the tagged and untagged Sebox proteins were overexpressed, purified, quantified and were used for further biophysical studies.

3.3.4 Biofilm formation by Sebox proteins

3.3.4.1 Crystal Violet assay

Biofilm formation by Sebox proteins was analyzed by growing the recombinant strains in both liquid and solid media. The growth of the surface pellicle was monitored at one-day intervals and quantified by recording the OD at 570 nm.

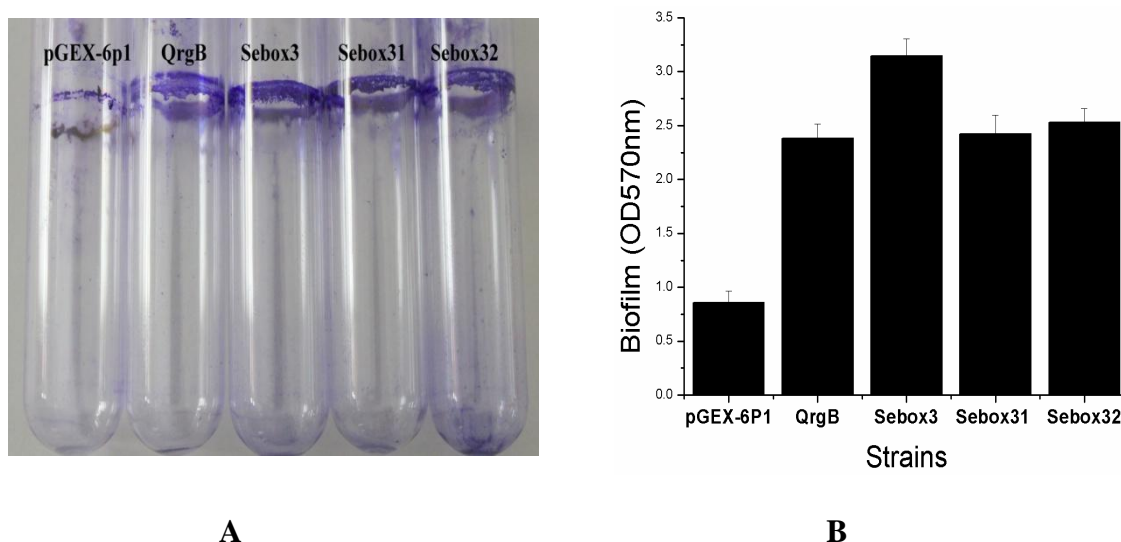


Figure 3.5: Crystal Violet staining of the strains for biofilm estimation: A) Biofilm formed in borosilicate tubes by static cultures of Sebox3, QrgB and pGEX-6P1. The cultures were drained and the pellicle was stained with 0.2% crystal violet and photographed. B) The crystal violet bound biofilm was dissolved in 75% ethanol and estimated spectrophotometrically at 570 nm.

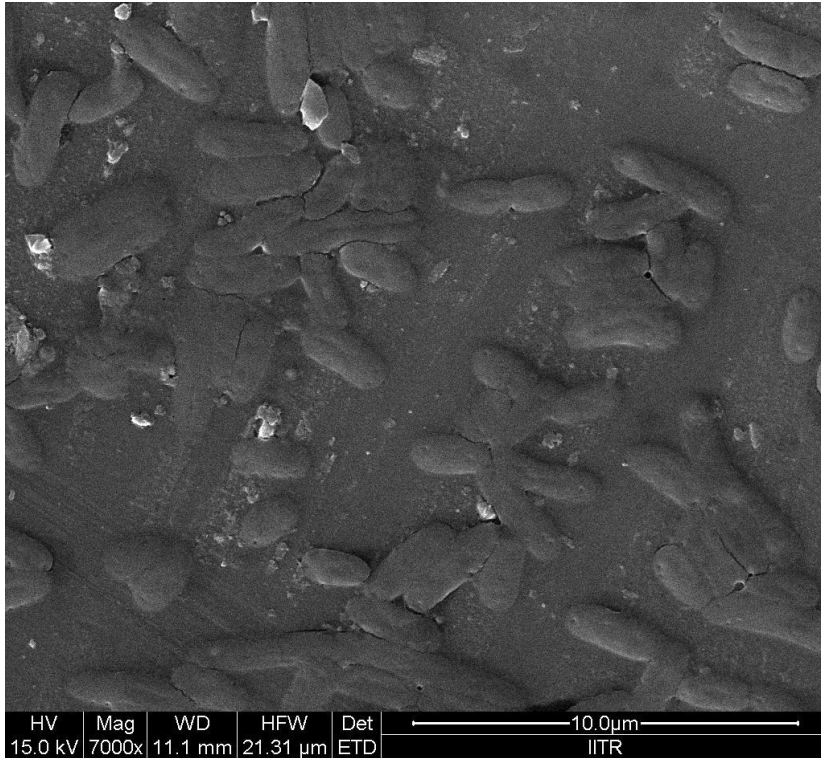
Visually, the comparison of the five growth cultures revealed similar pellicle thickness for QrgB, Sebox3, Sebox31 and Sebox32, while the negative control pGEX-6P1 had

significantly less thickness (Figure 3.4A). During the crystal violet staining procedure, the dye binds to the peptidoglycan of the bacterial cell wall and also to negatively charged molecules in the biofilm.

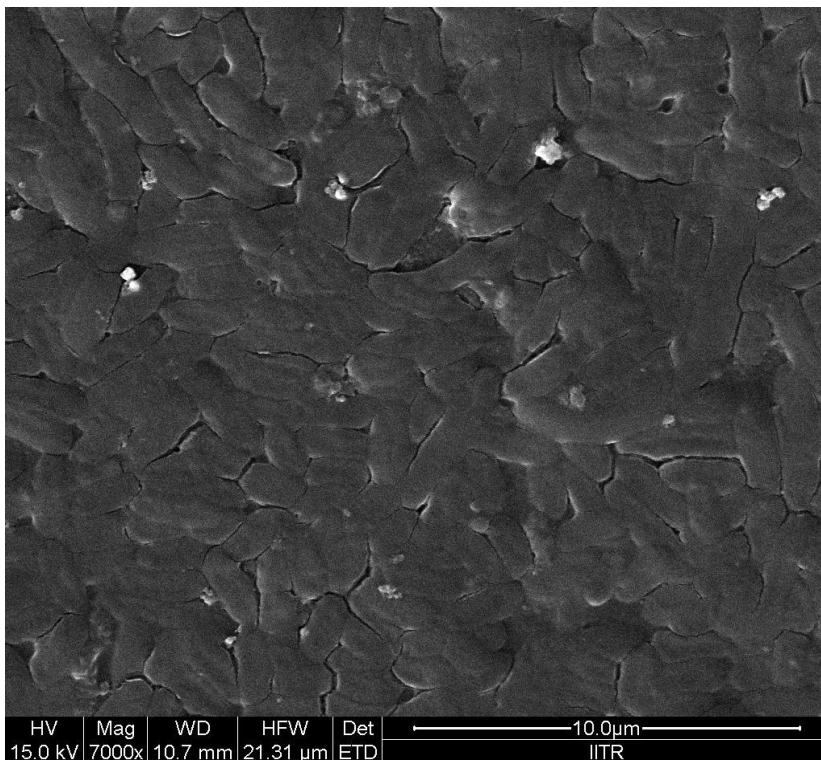
Utilization of crystal violet is cost-effective and due to its ease of solubility, it has been widely used for staining biofilms. Therefore, if more cells were attached to the glass tube then more crystal violet will be absorbed by the biofilm. Thus, the intensity of the crystal violet-dissolved solution is directly proportional to the amount of biofilm formed in the respective tubes (Figure 3.4B). Sebox3 strain with the full-length protein showed the highest biofilm forming activity. The Sebox31 and Sebox32 also retained their biofilm forming ability.

3.3.4.2 SEM of biofilms

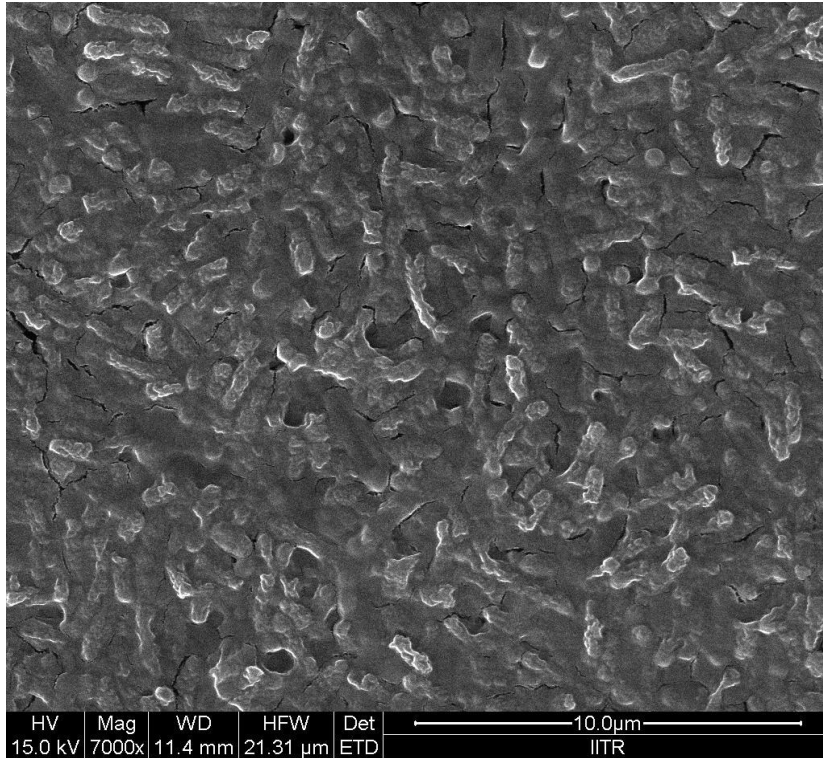
All the cultures have shown exopolysaccharide secretion which predominantly helps in the maintenance of the microbial matrix. From the SEM images, it is clearly evident that the cells form a dense network of cells (Figure 3.5). This is aided by cellulose produced by the allosteric effect of the c-di-GMP, in turn, produced by GGEEF proteins in all the cultures except pGEX-6P1.



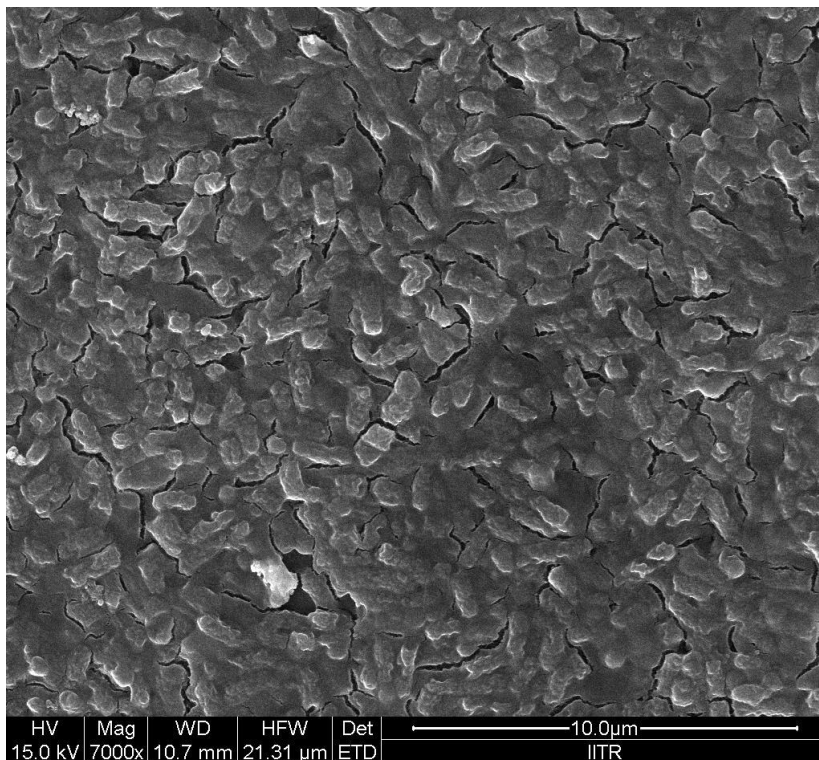
A pGEX-6P1



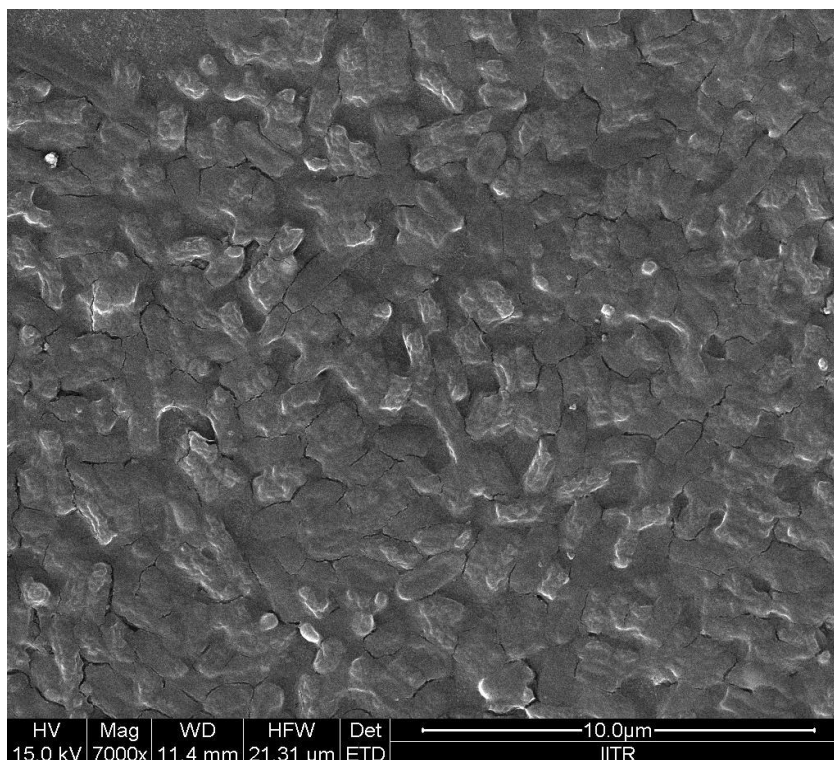
B QrgB



C Sebox3



D Sebox31



E Sebox32

Figure 3.6: SEM images of biofilms: A) pGEX-6P1, B) QrgB, C) Sebox3, D) Sebox31 and E) Sebox32 were grown statically on 20 mm glass coverslips placed vertically in test tubes containing 5 mL of Luria Bertani (LB) broth supplemented with 0.05 mM IPTG.

After the incubation period, the biofilm formed on the coverslip was fixed and imaged by FEI Quanta SEM. Due to *in vivo* DGC activity of the GGEEF domain, more amount of cellulose is produced and biofilm formation is enhanced. When the GGEEF domain is functionally active it catalyzes the conversion of GTP molecules to c-di-GMP. This bacterial secondary messenger binds to cellulose synthase and allosterically activates it to induce the production of cellulose. In the biofilm images of strains of positive control – QrgB, Sebox3, Sebox31 and Sebox32, a multilayered cluster of cells can be seen.

Apart from this, the density of cells is way to high in these strains as compared to the negative control-pGEX6P1. Extracellular matrix-like structures could be observed

predominantly in the image of Sebox3 strain. Thus, it has been confirmed that c-di-GMP molecules control bacterial phenotypic activity like biofilm formation.

3.3.5 Congo Red binding



Figure 3.7: Congo Red Assay: *E.coli* strains with Sebox3, Sebox31, Sebox32, QrgB and pGEX-6P1 were streaked on Congo Red agar plates and incubated at 37°C. All cultures except for pGEX-6P1 acquired a red colour due to binding of Congo Red to cellulose and curli fimbriae produced in these cultures.

Table 3.2: Congo Red uptake from Standard solution (40 µg/mL) is measured at 490 nm.

	Standard Congo Red Reading	pGEX-6P1	QrgB	Se3	Se31	Se32
unbound Congo Red (µg/mL)	40	36.8929	34.942	31.0862	32.8842	35.5023
bound Congo Red (µg/mL)	0	3.1071	5.058	8.9138	7.1158	4.4977

Congo Red binds to β – glucans in cellulose and curli fimbriae to give rdar (red, dry and rough) phenotype. Together cellulose and curli fimbria form an important component of the extracellular matrix of the bacterial biofilm. The gene for cellulose synthesis is located on the *bcsABZC* operon. While *bcsA* encodes the cellulose synthase enzyme, the exact functions of

the other genes remain to be determined. Cyclic-di-GMP produced by DGC like AdrA allosterically regulates cellulose synthase. The operon *csgBA(C)* encodes for curli fimbriae. Cellulose and curli fimbria biosynthesis are commonly regulated, either through *adrA* or directly by *csgD*, a transcriptional activator.

All the cultures were checked for their ability to bind to Congo Red on the production of cellulose. Except for pGEX-6P1, all the cultures showed bright pink colour indicating cellulose production and binding to Congo Red (Figure 3.6). This is indicative of the diguanylate cyclase (DGC) activity of the GGEEF domain in Sebox proteins. DGC activity leads to increased level of c-di-GMP which positively modulates cellulose synthase and leads to cellulose production [24–26]. The logical inference was, therefore that the biofilm formation by Sebox proteins-infused strains involved cellulose production, which was possibly triggered by the synthesis of c-di-GMP by the GGEEF domain of Sebox proteins.

3.3.6 Inverse effect of diguanylate on motility

From the results of the plate assay, it can also be seen that Sebox protein overexpression leads to the inhibition of swarming motility in the strains expressing Sebox3, Sebox31 and Sebox32 (Figure 3.7).

3.3.6.1 Motility on plates

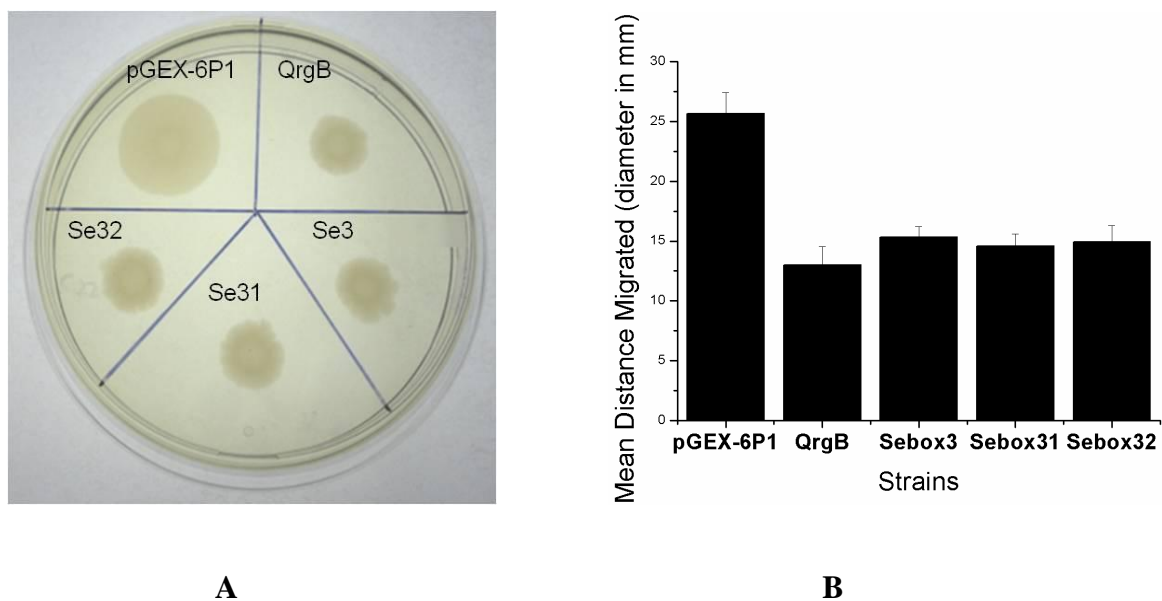


Figure 3.8: Motility on soft agar: A) Sebox3, Sebox31, Sebox32, QrgB and pGEX-6P1 were spot inoculated on the soft LB agar plates and incubated at 37°C for 6 days. At the end of the incubation period, the diameter of the zone of migration was measured from the point of inoculation. B) The graph shows the diameter of migration of the culture from the point of inoculation.

The motility test of the strains on soft agar media plates was carried out to check the *in vivo* inhibitory activity of the GGEEF domain containing strains. It was observed that the zone of migration for the GGEEF domain containing strains was drastically low as compared to the pGEX-6P1. This indicates that the GGEEF domain is functionally active and reinforces the fact that high level of expression of diguanylate cyclase (encoded by the GGEEF domain) inhibits the motility of the strains by increasing the level of c-di-GMP. Under the influence of high levels of c-di-GMP, the motility is drastically reduced.

3.3.6.2 TTC assay

In the TTC assay, the migration of the bacteria from the zone of inoculation by the needle can be visualized by the diffusion of the red coloured dye in TTC assays. As the bacteria migrate they convert the colourless soluble TTC to insoluble red-coloured formazan (Figure 3.8).

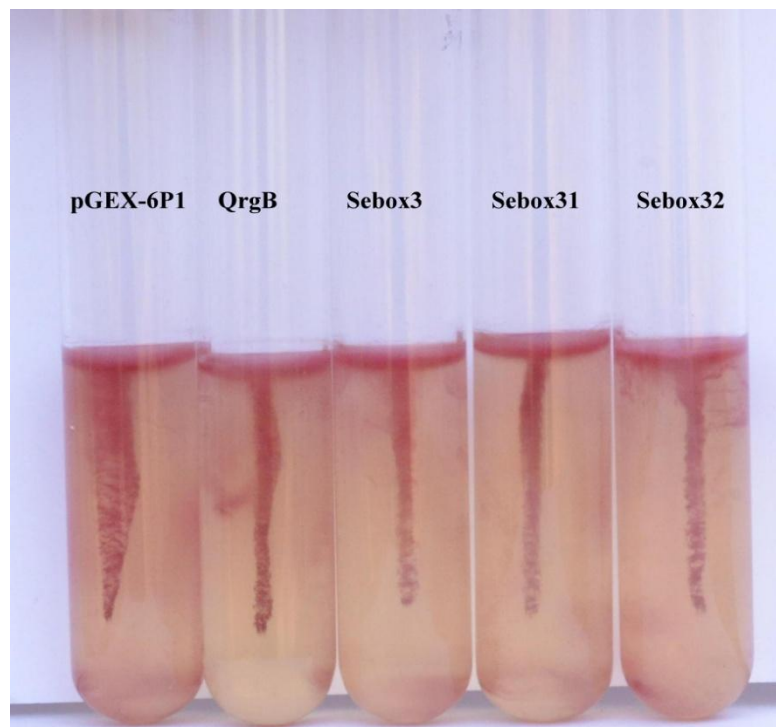


Figure 3.9: Motility in tetrazolium tubes: Zone of migration from the point of inoculation is seen only in the case of pGEX-6P1. Motility in all the remaining strains is retarded due to the *in vivo* inhibitory action of DGC.

After incubation at 37°C for two days without shaking, it was observed that the Sebox3, Sebox31 and Sebox32 did not show any migration from the initial stab line, in a pattern very similar to that of the QrgB containing strain. However, pGEX-6P1 showed a sizeable migration from the stab line, pointing to their motility. Motility is normally inhibited in biofilm forming bacteria and the motility results support the results from the biofilm assay.

3.4 Conclusion

The bacterial secondary messenger, c-di-GMP controls vital phenotypic activities like biofilm formation and motility of the bacterial cells. At high level of expression, it enhances biofilm formation and inhibits motility. It is produced from GTP molecules by the diguanylate activity of the GGEEF domain-containing proteins. These protein domains are usually found

in association with other domains indicating that these other domains could have a regulatory role in its function.

In this study, we have described the construction, expression and purification of soluble Sebox31 and Sebox32 proteins with the GGEEF motif from *V. cholerae*. Like most other GGD(/E)F proteins, Sebox proteins have demonstrated biofilm forming abilities as evidenced by the crystal violet assay and confirmed by SEM imaging. The negative control, (pGEX-6P1) shows poor biofilm formation with a lesser volume of structure as compared to the strains containing the GGD(E/)EF domain.

The GGD(/E)EF domain proteins with diguanylate cyclase activity are known to lead to biofilm formation in the host strain. Congo Red can be used to measure cellulose and other biofilm-related material like curli fimbriae. The uptake of bright red colour in all the Sebox strains indicates the presence of exopolysaccharide secretion which holds the cells together in the extracellular matrix of the biofilm. In bacteria, cellulose production has been known to be initiated by allosteric activation of cellulose synthase by c-di-GMP. Due to the diguanylate cyclase activity associated with GGD(/E)EF, high levels of c-di-GMP is produced which allosterically activates cellulose synthase to produce cellulose. Thus, it can be concluded that Sebox proteins could trigger the production of exopolysaccharides and hence, aid biofilm formation.

Inversely, motility of the biofilm-forming strains was reduced, a result which was confirmed both by swarming and swimming motility analyses. Swarming aids the bacteria to attach to the substratum during biofilm formation whereas, swimming motility reduces biofilm formation. In case of the Sebox3 protein carrying strains, swimming motility was drastically

reduced as indicated by lesser dispersion of red-coloured reduced tetrazolium product from the point of stab in the soft agar medium. The results were validated by the diameter of the swarm in the plate culture as well. Hypothetically, increasing levels of c-di-GMP prevented motility by repressing the transcription regulation of flagellar genes or by interacting with flagellar motor proteins.

These results indicate that high level of diguanylate cyclase and in turn, high level of intracellular c-di-GMP lead to the synthesis of cell aggregation and surface attachment factors which invariably led to the three dimensional multi-structured biofilm formation.

3.5 References

- [1] G. A. Malawski *et al.*, “Identifying protein construct variants with increased crystallization propensity - a case study.,” *Protein Sci.*, vol. 15, no. 12, pp. 2718–28, 2006.
- [2] R. M. Rasia *et al.*, “Parallel screening and optimization of protein constructs for structural studies,” *Protein Sci.*, vol. 18, no. 2, pp. 434–439, 2009.
- [3] M. Boettner, B. Prinz, C. Holz, U. Stahl, and C. Lang, “High-throughput screening for expression of heterologous proteins in the yeast *Pichia pastoris*,” *J. Biotechnol.*, vol. 99, no. 1, pp. 51–62, 2002.
- [4] W. G. Kaelin *et al.*, “Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties,” *Cell*, vol. 70, no. 2, pp. 351–364, 1992.
- [5] C. M. Waters, W. Lu, J. D. Rabinowitz, and B. L. Bassler, “Quorum sensing controls biofilm formation in *Vibrio cholerae* through modulation of cyclic Di-GMP levels and repression of vpsT,” *J. Bacteriol.*, vol. 190, no. 7, pp. 2527–2536, 2008.

- [6] B. K. Hammer and B. L. Bassler, “Distinct sensory pathways in *Vibrio cholerae* El Tor and classical biotypes modulate cyclic dimeric GMP levels to control biofilm formation,” *J. Bacteriol.*, vol. 91, no. 1, pp. 169–177, 2009.
- [7] S. Harper and D. W. Speicher, “Expression and purification of GST fusion proteins,” *Current Protocols in Protein Science*, no. SUPPL. 52. 2008.
- [8] N. K. Vinckier, A. Chworos, and S. M. Parsons, “Improved isolation of proteins tagged with glutathione S-transferase,” *Protein Expr. Purif.*, vol. 75, no. 2, pp. 161–164, 2011.
- [9] M. M. Bradford, “A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding,” *Anal. Biochem.*, vol. 72, no. 1–2, pp. 248–254, 1976.
- [10] A. Boyd and A. M. Chakrabarty, “*Pseudomonas aeruginosa* biofilms: role of the alginate exopolysaccharide,” *J. Ind. Microbiol.*, vol. 15, no. 3, pp. 162–168, 1995.
- [11] L. Rajeev *et al.*, “Identification of a cyclic-di-GMP-modulating response regulator that impacts biofilm formation in a model sulfate reducing bacterium,” *Front. Microbiol.*, vol. 5, no. 7, 2014.
- [12] X. Zogaj, M. Nimtz, M. Rohde, W. Bokranz, and U. Römling, “The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix,” *Mol. Microbiol.*, vol. 39, no. 6, pp. 1452–1463, 2001.
- [13] S. Da Re and J.-M. Ghigo, “A CsgD-independent pathway for cellulose production and biofilm formation in *Escherichia coli*,” *J. Bacteriol.*, vol. 188, no. 8, pp. 3073–3087, 2006.

- [14] N. Liu, T. Pak, and E. M. Boon, "Characterization of a diguanylate cyclase from *Shewanella woodyi* with cyclase and phosphodiesterase activities.," *Mol. Biosyst.*, vol. 6, no. 9, pp. 1561–4, 2010.
- [15] L. M. Ruiz, M. Castro, A. Barriga, C. A. Jerez, and N. Guiliani, "The extremophile *Acidithiobacillus ferrooxidans* possesses a c-di-GMP signalling pathway that could play a significant role during bioleaching of minerals," *Lett. Appl. Microbiol.*, vol. 54, no. 2, pp. 133–139, 2012.
- [16] A. J. Spiers, J. Bohannon, S. M. Gehrig, and P. B. Rainey, "Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose," *Mol. Microbiol.*, vol. 50, no. 1, pp. 15–27, 2003.
- [17] J. Fernández, F. Sisti, D. Bottero, M. E. Gaillard, and D. Hozbor, "Constitutive expression of bvgR-repressed factors is not detrimental to the *Bordetella bronchiseptica*-host interaction," *Res. Microbiol.*, vol. 156, no. 8, pp. 843–850, 2005.
- [18] F. Sisti, D. G. Ha, G. A. O'Toole, D. Hozbor, and J. Fernández, "Cyclic-di-GMP signalling regulates motility and biofilm formation in *Bordetella bronchiseptica*," *Microbiol. (United Kingdom)*, vol. 159, no. PART 5, pp. 869–879, 2013.
- [19] S. L. Kuchma, N. J. Delalez, L. M. Filkins, E. A. Snavely, J. P. Armitage, and G. A. O'Toole, "Cyclic di-GMP-mediated repression of swarming motility by *Pseudomonas aeruginosa* PA14 Requires the MotAB stator," *J. Bacteriol.*, vol. 197, no. 3, pp. 420–430, 2015.
- [20] D. Srivastava, M. L. Hsieh, A. Khataokar, M. B. Neiditch, and C. M. Waters, "Cyclic di-GMP inhibits *Vibrio cholerae* motility by repressing induction of transcription and inducing extracellular polysaccharide production," *Mol. Microbiol.*, vol. 90, no. 6, pp.

1262–1276, 2013.

- [21] R. J. Ball and W. Sellers, “Improved motility medium.,” *Appl. Microbiol.*, vol. 14, no. 4, pp. 670–673, 1966.
- [22] S. An, J. Wu, and L. H. Zhang, “Modulation of *Pseudomonas aeruginosa* biofilm dispersal by a cyclic-di-gmp phosphodiesterase with a putative hypoxia-sensing domain,” *Appl. Environ. Microbiol.*, vol. 76, no. 24, pp. 8160–8173, 2010.
- [23] O. P. Chouhan, D. Bandekar, M. Hazra, A. Baghudana, S. Hazra, and S. Biswas, “Effect of site-directed mutagenesis at the GGEEF domain of the biofilm forming GGEEF protein from *Vibrio cholerae*.,” *AMB Express*, vol. 6, no. 1, p. 2, 2016.
- [24] P. Ross *et al.*, “Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid,” *Nature*, vol. 325, no. 6101. pp. 279–281, 1987.
- [25] U. Romling, M. Y. Galperin, and M. Gomelsky, “Cyclic di-GMP: the first 25 years of a universal bacterial second messenger.,” *Microbiol. Mol. Biol. Rev.*, vol. 77, no. 1, pp. 1–52, 2013.
- [26] R. Tamayo, J. T. Pratt, and A. Camilli, “Roles of cyclic diguanylate in the regulation of bacterial pathogenesis.,” *Annu. Rev. Microbiol.*, vol. 61, pp. 131–48, 2007.

CHAPTER 4

Comparative Characterization of Sebox3 versus Sebox31 and Sebox32

4.1 Introduction:

Biophysical characterization of proteins aids in the study of the molecular behaviour of proteins *in vitro*. Information regarding the shape, conformation, size, polarity of the proteins and their interactions with other biomolecules can be acquired by carrying out an array of biophysical techniques [1]–[4]. Most importantly, in the absence of a three-dimensional structure, these techniques help to define individual domains, motifs and folds within a protein and enhance the understanding of the function of the proteins. Furthermore, any biophysical characterization supplements the concrete details of the molecular structure obtained from X-ray crystallography or high-resolution NMR [5], [6]. Classical techniques like CD and fluorescence spectroscopy have been used to provide information about the conformation of proteins in solution under varying conditions [7]–[13]. Although these techniques provide rapid and sensitive details, the information provided is limited to a specific aspect of the protein, compared to X-ray crystallography and high resolution NMR [14]. Yet these biophysical techniques are popular as they show behaviour of proteins in solution by using small volumes of sample by non-destructive methods. Other techniques like gel filtration chromatography, surface plasma resonance (SPR), Isothermal Titration Calorimetry (ITC) [15], [16], Dynamic light scattering (DLS) [17]–[19] and chemical cross-linking can also be used to extensively characterize proteins [20], [21].

It helps us to understand how the activity of even a single biomolecule can control the behaviour or survival of the host organism in which it is present. For example, in this study, it is seen that the expression of recombinant GGD(/E)EF domain proteins influence the lifestyle switch between the motile and sessile form in the host *E. coli*. The *in vivo* study of Sebox3, Sebox31 and Sebox32 proteins provides valuable information about the activity of the GGEEF domain within the *E. coli* host strain.

In spite of the varying lengths, all the three proteins with GGEEF motif have been shown to retain their characteristic diguanylate cyclase activity. In this study, *in vitro* experiments were carried out to analyze how changes in structure-function relationships of the smaller constructs, could be attributed to deletions in the N-terminal of the full-length Sebox3 protein.

4.2 Methods

4.2.1 Diguanylate cyclase assays

The protocol stated in Chapter 2, section 2.2.10 was used for assessing the diguanylate activity of Sebox31 and Sebox32 proteins.

4.2.2 Circular Dichroism (CD) spectroscopy

Circular dichroism (CD) measurements were performed on a Jasco J-1500 CD spectrometer equipped with a Peltier temperature controller for monitoring the temperature between 20°C-90°C. Conformational changes in the secondary structure of protein were monitored in the region between 190 nm to 250 nm with a protein concentration of 7 μ M in a quartz cuvette (Jasco) with a path length of 1 mm.

The scanning speed, bandwidth and data pitch were set to 100 nm/min, 1.0 nm and 1.0 nm, respectively. Three readings of the spectral scans were taken and averaged to get the final complete spectra.

4.2.3 Fluorescence spectroscopy

To study the intrinsic fluorescence of the tryptophans in the Sebox31 and Sebox32 proteins, fluorescence measurements were carried out using a Jasco FP8200 spectrofluorometer. Fluorescence spectra were measured keeping both the excitation and emission band passes at

5 nm. The tryptophan fluorescence was measured at the wavelength of 295 nm and the emission spectra obtained between 305 nm to 400 nm was noted.

4.2.3.1 Chemical denaturation

For the chemical denaturation study with guanidine hydrochloride, the protocol mentioned in Chapter 2, section 2.2.8.2 has been followed with necessary adaptations.

4.2.3.2 Quenching of fluorescence

In the quenching experiments, the same standard quenchers, that is, acrylamide, KI and CsCl were used and the protocol mentioned in Chapter 2, section 2.2.8.3 were used.

4.2.4 Analytical gel filtration chromatography

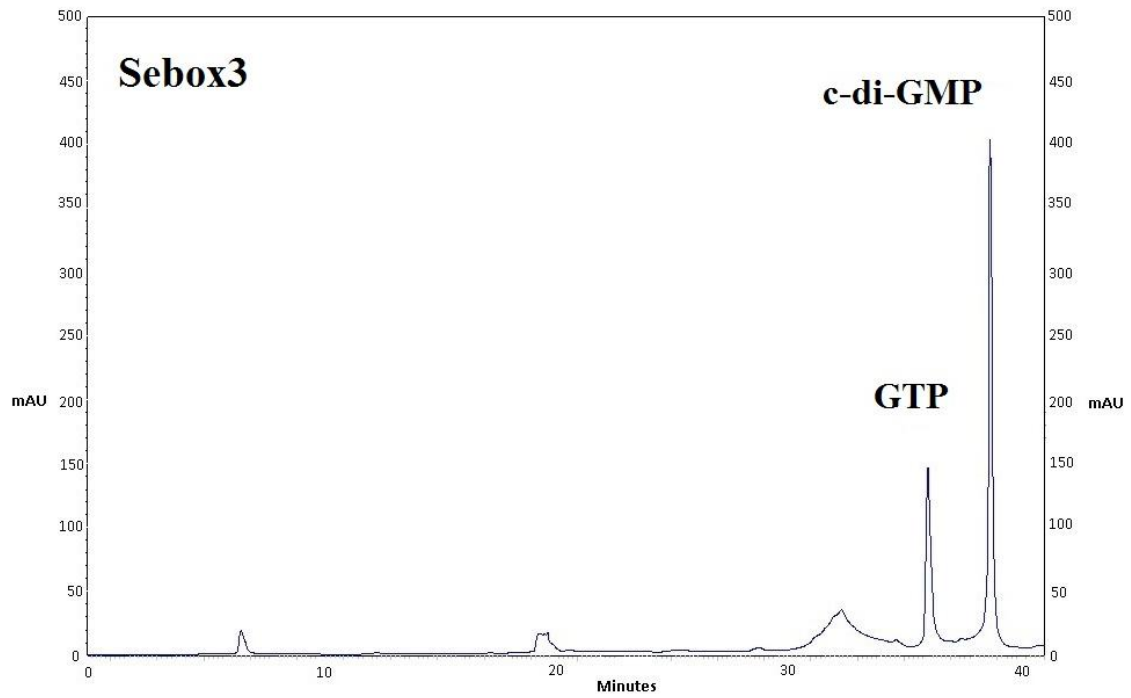
Analytical gel filtration chromatography was performed using the protocol mentioned in Chapter 2, section 2.2.7.1.

4.3 Results and Discussion

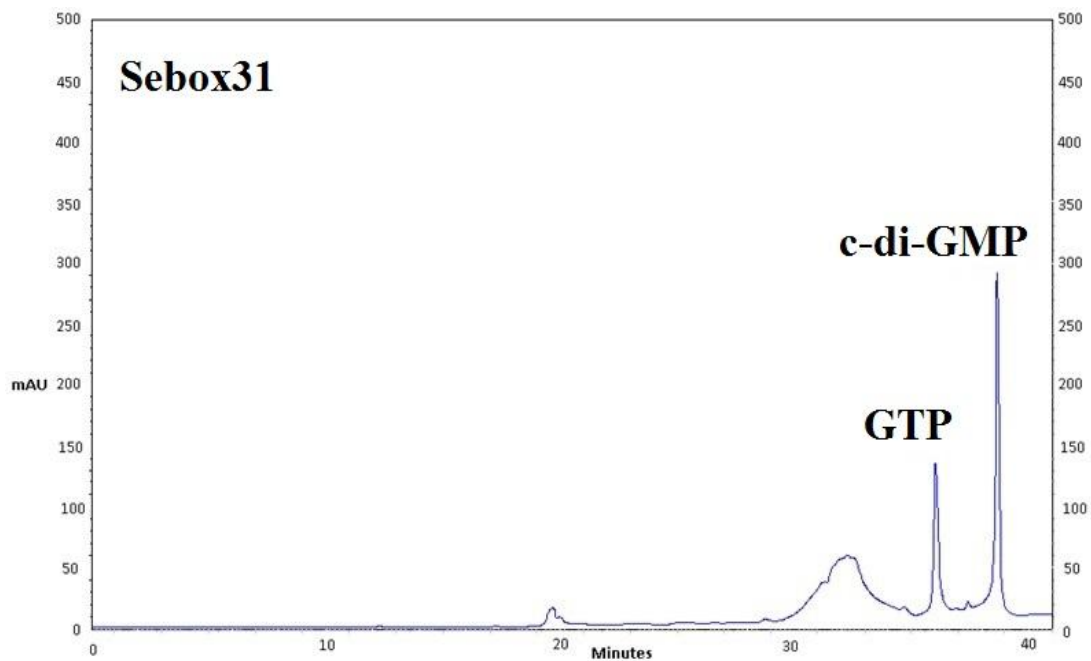
4.3.1 Diguanylate cyclase activity of N-terminal truncates

Despite the deletions at the N-terminal of the full-length Sebox3 protein, the diguanylate activity is still retained by the smaller constructs. The same retention time (as observed for Sebox3 sample) of 36.1 min and 38 min for GTP and c-di-GMP respectively was seen in the case of Sebox31 and Sebox32 (Figure 4.3). However, Sebox31 showed a slight decrease in the level of activity. However, Sebox32 had comparable levels of diguanylate cyclase activity to Sebox3. Comparing with the chromatogram for Sebox3 protein sample, peaks were seen in the same position as the standard/control samples.

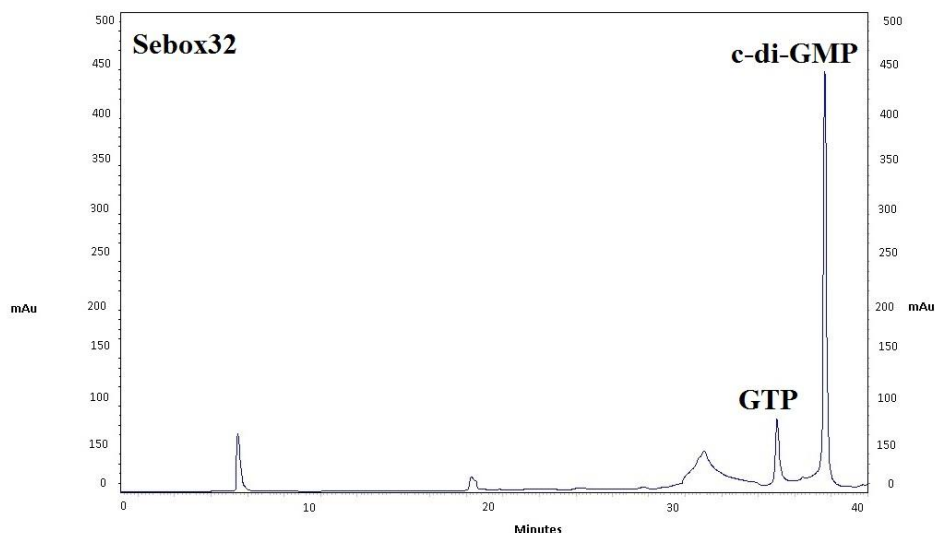
Thus, it can be concluded that Sebox31 and Sebox32 still possess diguanylate cyclase activity, despite the deletions at the N terminal.



A



B

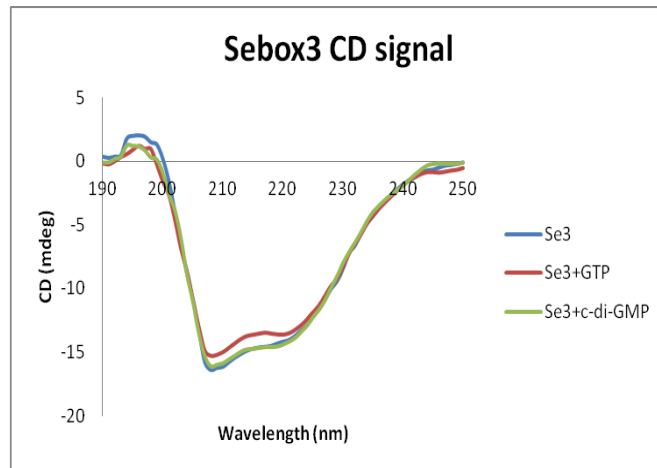


C

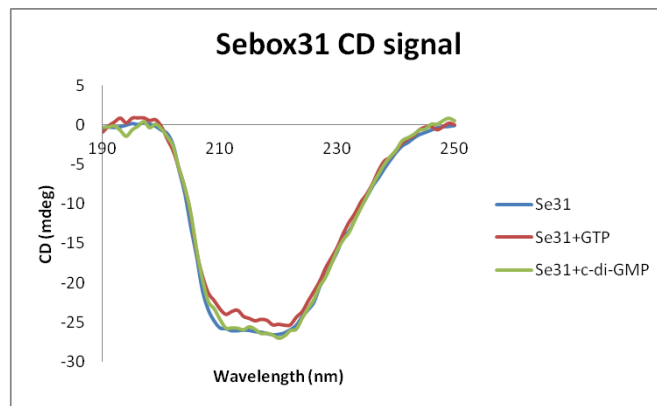
Figure 4.1: Diguanylate cyclase activity of Sebox proteins in HPLC. A) Sebox3 protein, B) Sebox31 protein and C) Sebox32 protein.

4.3.2 Comparison of secondary structure of constructs

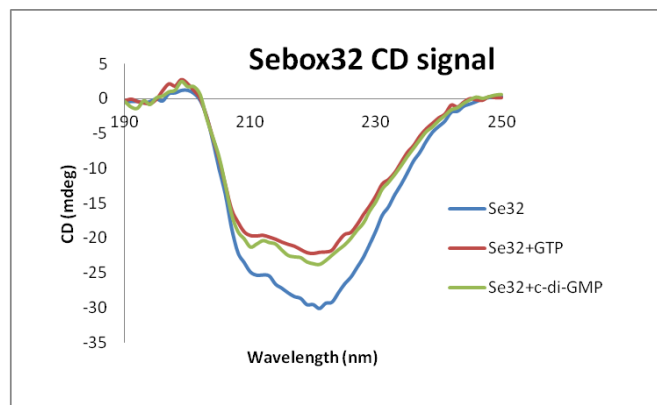
For secondary structure estimation, circular dichroism (CD) spectroscopic measurements were taken from a range of 190 nm – 250 nm to check for the difference in structure between Sebox3 and its smaller constructs Sebox31 and Sebox32. The spectra of the proteins are characterized by a positive peak at 192 nm and negative peaks at 208 nm and 222 nm which generally indicates an α -helix (Figure 4.1). The positive peak at 192 nm and minimum at 208 nm can be attributed to $\pi\pi^*$ transition in peptide bond whereas, the minimum at 222 nm is due to $n\pi^*$ transition. There is an overall change in spectra of the proteins indicating that there is a slight change in the secondary structure of the protein. In Sebox31 protein, the magnitude of the positive peak at 192 nm is smaller indicating a transition from the α -helical form. The low peak at 192 nm is observed due to the low concentration of the proteins and also due to the use of Tris-HCl buffer. Generally, phosphate buffer is used for CD spectroscopy analysis but due to the low concentration of the protein samples and instability of the proteins in phosphate buffer, buffer exchange of the protein was not carried out.



A



B



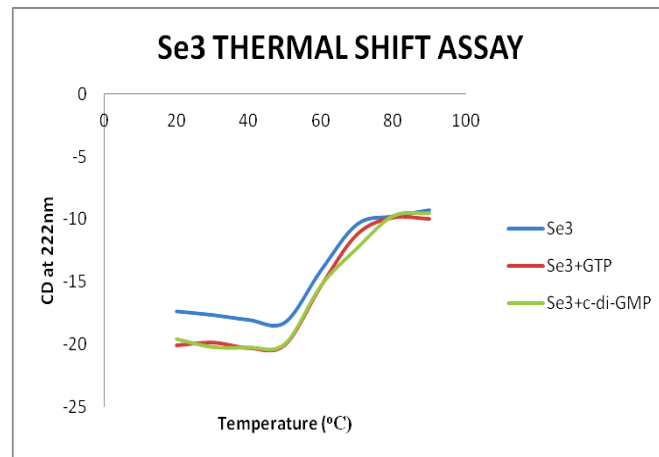
C

Figure 4.2: CD spectrum of the protein, protein with GTP and Protein with c-di-GMP. A: Sebox3 CD signal, B: Sebox31 CD signal and C: Sebox32 signal.

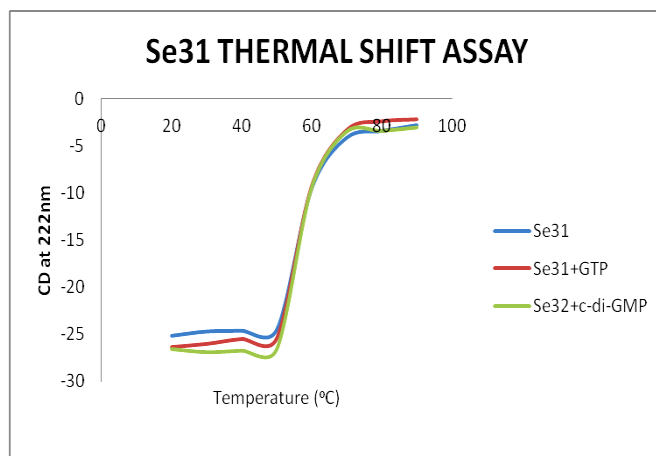
4.3.3 Thermal stability and binding of ligand

The proteins were preincubated for a fixed period at a particular temperature ranging from 20°C to 90°C to check for a change in T_m of protein due to binding of ligand. Generally, ligands bind to the protein and stabilize the protein. A high T_m is obtained if the natively folded protein is stabilized by the ligand binding and conversely, a lower T_m is obtained if the unfolded state is stabilized. The protein stabilized by ligand binding would denature at a higher temperature. In case of Sebox3 and Sebox32 a small change in T_m was elicited. The For Sebox3, the T_m changes from 55°C to 60°C (Figure 4.2A) and for Sebox32 the T_m changes from 57°C to 60°C (Figure 4.2C).

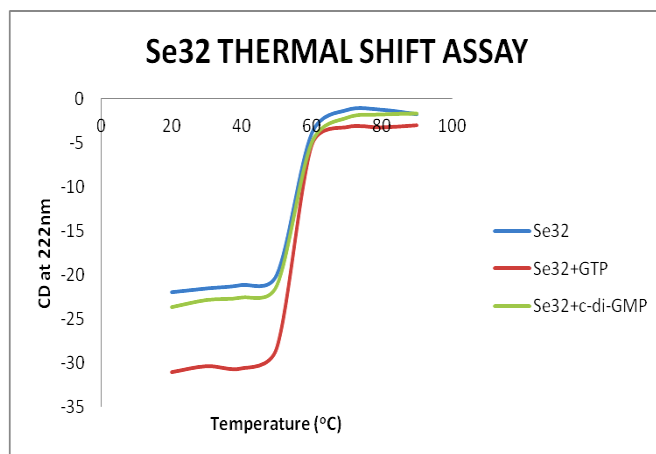
In the Sebox31 protein, the substrate GTP binds to the protein (as evidenced by its diguanylate activity) but does not bring about a change in the T_m of the protein (Figure 4.2B).



A



B



C

Figure 4.3: Thermal shift assay using CD spectra of the protein from 20°C to 90°C, protein with GTP and protein with c-di-GMP. A: Sebox3 CD sig, B: Sebox31 CD signal and C: Sebox32 signal.

4.3.4 Oligomeric status of constructs

The tagless Sebox31 and Sebox32 were run on a Superdex200 column equilibrated with 50 mM Tris-HCl and 150 mM NaCl (pH 7.4). A single peak was obtained for Sebox31 as well as Sebox32 protein indicating that both the proteins exist as monomers in solution. The molecular weights of Sebox31 and Sebox32 were found to be 22.0 kDa and 17.6 kDa respectively (Figure 4.4).

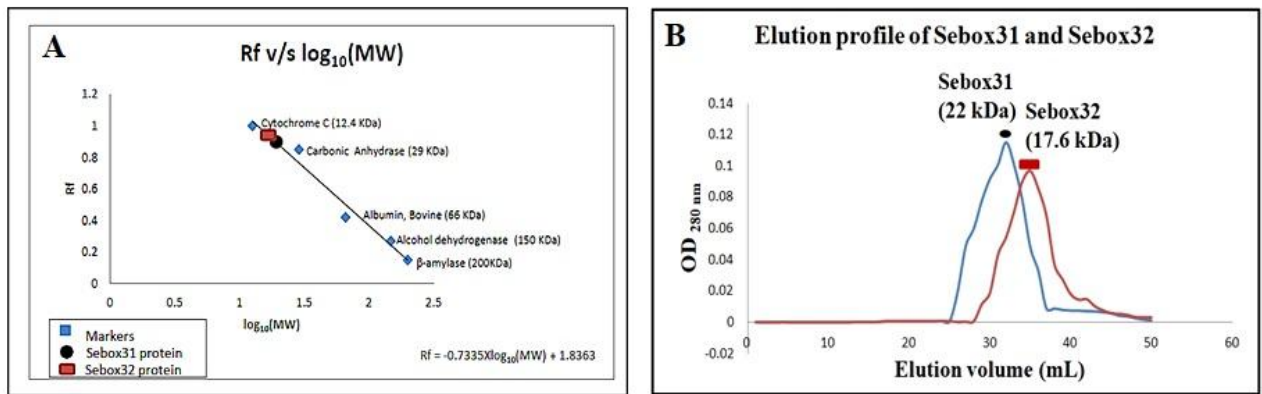
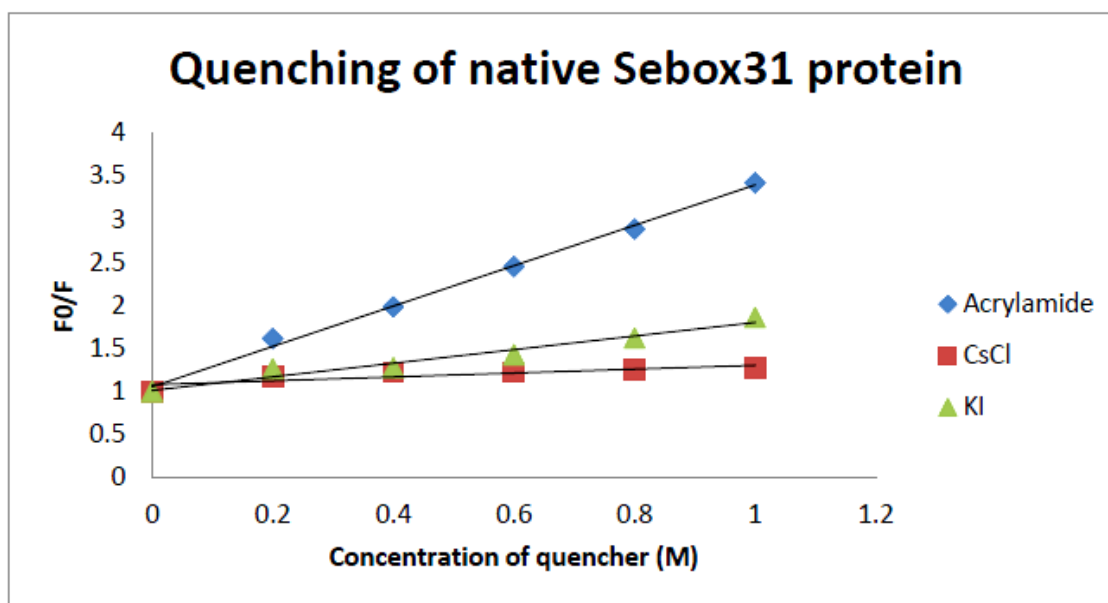


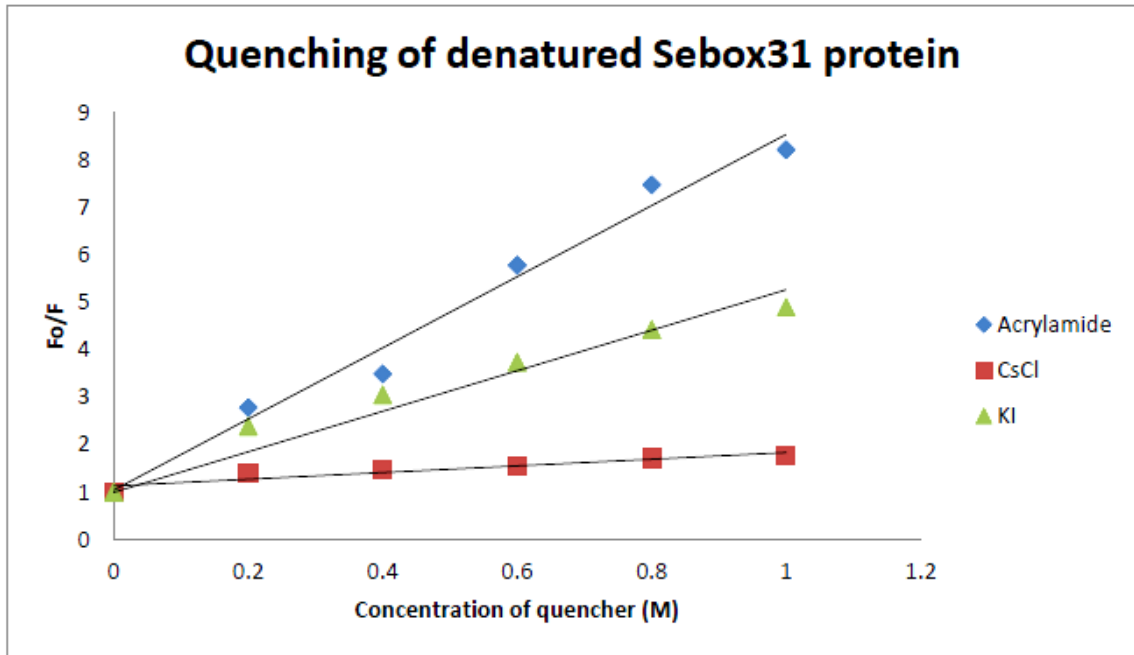
Figure 4.4: A) Analytical gel filtration of Sebox31 and Sebox32 protein along with standard protein markers carried out using a Superdex200 column in GE Healthcare AKTA Prime system. B) The elution profile shows the elution profile of Sebox31 protein (22 kDa) and Sebox32 protein (17.6 kDa).

4.3.5 Quenching of native and denatured Sebox31 protein and Sebox32 proteins

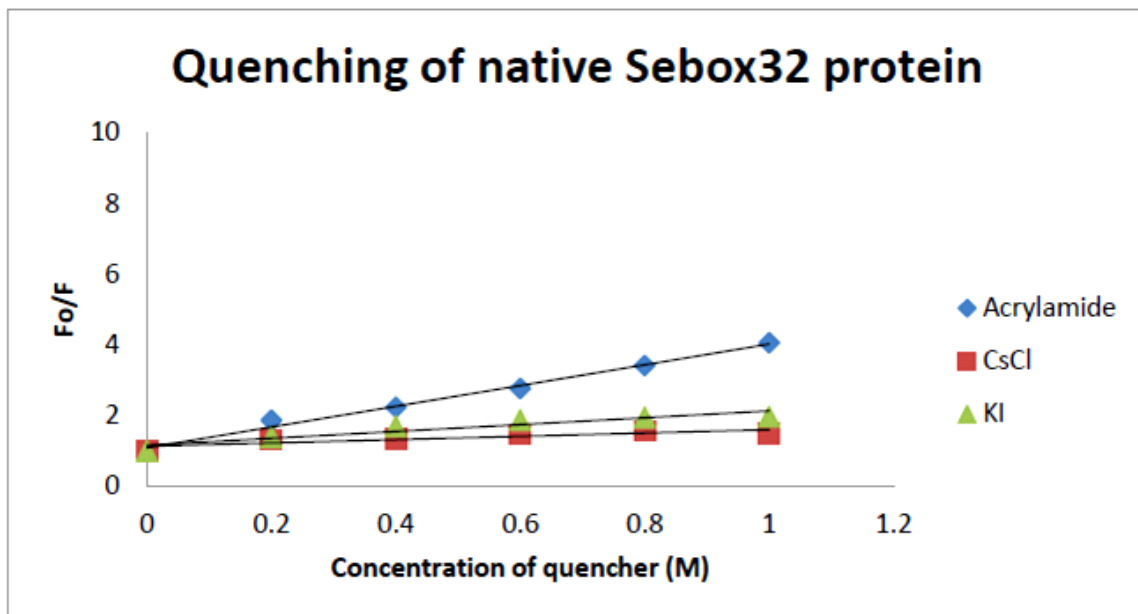
Quenching experiment was carried to check the quenching potential of the standard quenchers on the internal fluorescent probes. Acrylamide is the neutral quencher, CsCl is the positive quencher and KI is the negative quencher. Both the native proteins have an emission maximum at 337 nm (Figure 4.5). For the denaturation study, GdnHCl was used. Due to denaturation, the tryptophan was exposed to the external solvent environment as ascertained by an increase in fluorescence signal. When the denatured protein with the exposed tryptophan was quenched, there was a drastic change in the fluorescence. This was because the quencher could easily reach the immediate environment and lower the fluorescent intensity.



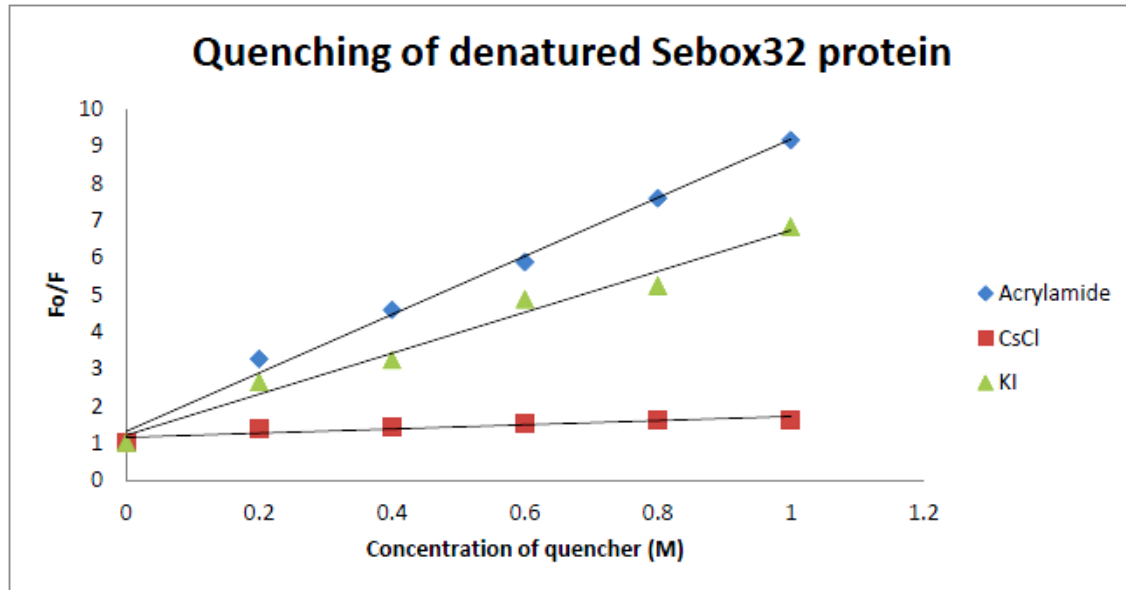
A



B



C



D

Figure 4.5: Quenching of Sebox proteins with acrylamide, CsCl and KI. Quenching with A) native Sebox31, B) denatured Sebox31, C) native Sebox31 and D) denatured Sebox32 proteins.

4.4 Conclusion

In this study, the secondary structure of the protein sequences was determined using the CD spectra. It was seen that the Sebox family of proteins were predominantly α -helical, a hypothesis that held true for the truncated versions as well. The change in spectra from the apo to the truncated forms was due to deletion of the N-terminal sequences in Sebox31 (between 1-129 amino acids) and Sebox32 (between 1-159 amino acids). Thus, it can be concluded that the conservation of the GGEEF domain in all the 3 proteins (Sebox3, Sebox31 and Sebox32) was necessary for the activity of the proteins. In the thermal shift assay, only Sebox3 and Sebox32 were stabilized by the binding of the substrate GTP and hence these proteins denatured at a higher T_m as compared to Sebox31.

Both the Sebox31 and Sebox32 proteins showed diguanylate cyclase activity. Among the two constructs, Sebox32 showed more activity as compared to the Sebox31 protein. The molecular weights of the Sebox31 and Sebox32 protein were 22 kDa and 17.6 kDa respectively. Like Sebox3 protein, these proteins were also monomeric in solution. The quenching experiments illustrate that acrylamide was the most effective quencher of native and denatured Sebox31 and Sebox32 proteins, pointing to the neutrality of the surrounding environment. This was followed by charged ionic quenchers like KI and CsCl. In the samples denatured by GdnHCl, the quenching was more effective due to the exposure of tryptophan to solvent.

4.5 References

- [1] K. E. Neet and J. C. Lee, "Biophysical characterization of proteins in the post-genomic era of proteomics.," *Mol. Cell. Proteomics*, vol. 1, no. 6, pp. 415–20, 2002.
- [2] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, and K. Dunker, "Intrinsic disorder and functional proteomics.," *Biophys. J.*, vol. 92, no. 5, pp. 1439–56, 2007.
- [3] P. Nordlund *et al.*, "Protein production and purification," *Nat. Methods*, vol. 5, no. 2, pp. 135–146, 2008.
- [4] N. D. Denslow, P. T. Wingfield, and K. Rose, "Overview of the characterization of recombinant proteins.," *Curr. Protoc. Protein Sci.*, vol. Chapter 7, p. Unit 7.1, 2001.
- [5] A. a S. Takeda *et al.*, "Biophysical characterization of the recombinant importin- α from *Neurospora crassa*," *Protein Pept. Lett.*, vol. 20, no. 1, pp. 8–16, 2013.
- [6] D. N. Langelaan, P. Ngweniform, and J. K. Rainey, "Biophysical characterization of G-protein coupled receptor-peptide ligand binding.," *Biochem. Cell Biol.*, vol. 89, no. 2, pp. 98–105, 2011.

- [7] N. J. Greenfield, "Methods to estimate the conformation of proteins and polypeptides from circular dichroism data.," *Anal. Biochem.*, vol. 235, no. 1, pp. 1–10, 1996.
- [8] N. J. Greenfield, "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions," *Nat. Protoc.*, vol. 1, no. 6, pp. 2527–2535, 2007.
- [9] D. Chakraborty and M. Bhattacharyya, "Deferiprone (L1) induced conformation change of hemoglobin: A fluorescence and CD spectroscopic study.," *Mol. Cell. Biochem.*, vol. 204, no. 1–2, pp. 17–20, 2000.
- [10] C. A. Royer, "Probing protein folding and conformational transitions with fluorescence," *Chem. Rev.*, vol. 106, no. 5, pp. 1769–1784, 2006.
- [11] J. T. Yang, C. S. C. Wu, and H. M. Martinez, "Calculation of Protein Conformation from Circular Dichroism," *Methods Enzymol.*, vol. 130, no. C, pp. 208–269, 1986.
- [12] N. J. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure.," *Nat. Protoc.*, vol. 1, no. 6, pp. 2876–90, 2006.
- [13] N. Sreerama and R. W. Woody, "Computation and Analysis of Protein Circular Dichroism Spectra," *Methods Enzymol.*, vol. 383, pp. 318–351, 2004.
- [14] K. Inomata *et al.*, "High-resolution multi-dimensional NMR spectroscopy of proteins in human cells.," *Nature*, vol. 458, no. 7234, pp. 106–109, 2009.
- [15] G. Bains and E. Freire, "Calorimetric determination of cooperative interactions in high affinity binding processes.," *Anal. Biochem.*, vol. 192, no. 1, pp. 203–206, 1991.
- [16] A. Velazquez-Campoy and E. Freire, "ITC in the post-genomic era...? Priceless," in *Biophysical Chemistry*, 2005, vol. 115, no. 2–3 Spec. Iss., pp. 115–124.
- [17] D. Arzenšek, R. Podgornik, and D. Kuzman, "Dynamic light scattering and application to proteins in solutions," in *Seminar*, 2010, pp. 1–18.

- [18] B. Lorber, F. Fischer, M. Bailly, H. Roy, and D. Kern, “Protein analysis by dynamic light scattering: Methods and techniques for students,” *Biochem. Mol. Biol. Educ.*, vol. 40, no. 6, pp. 372–382, 2012.
- [19] A. Proteau, R. Shi, and M. Cygler, “Application of dynamic light scattering in protein crystallization.,” *Curr. Protoc. Protein Sci.*, vol. Chapter 17, no. August, p. Unit 17.10, 2010.
- [20] F. H. Niesen *et al.*, “An approach to quality management in structural biology: Biophysical selection of proteins for successful crystallization,” *J. Struct. Biol.*, vol. 162, no. 3, pp. 451–459, 2008.
- [21] S. Yokoyama, “Protein expression systems for structural genomics and proteomics,” *Curr. Opin. Chem. Biol.*, vol. 7, no. 1, pp. 39–43, 2003.

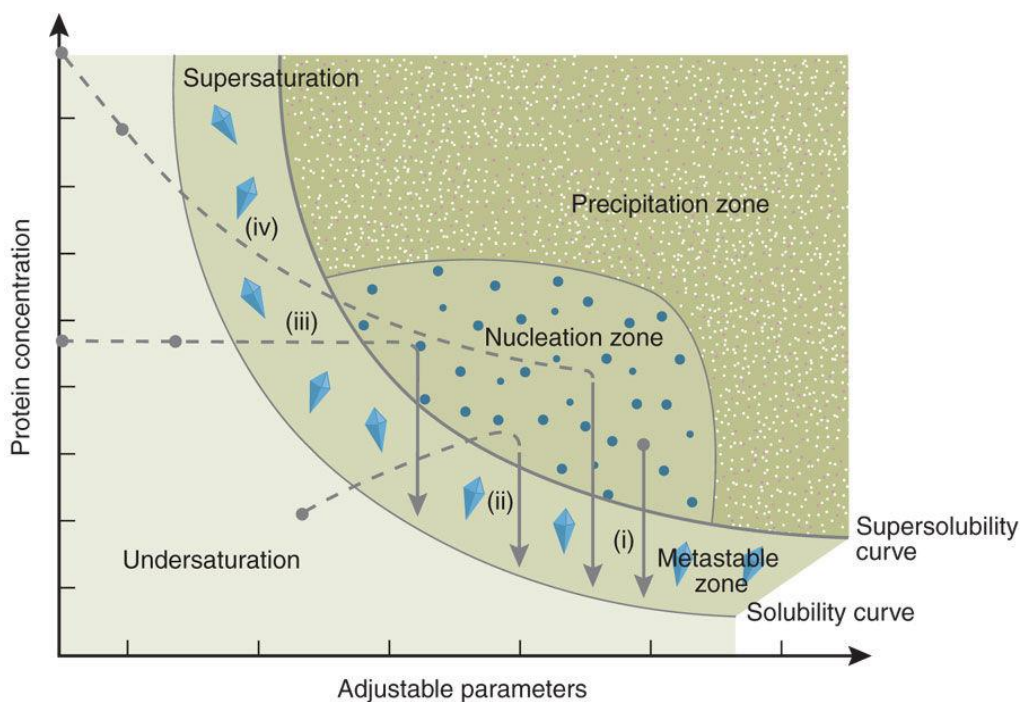
CHAPTER 5

The Structure of Sebox3 protein

5.1 Introduction

X-ray crystallography has been widely used for solving the structure of biological macromolecules like proteins and DNA and has helped to understand the role of these macromolecules in various metabolic processes. This technique provides definitive information at the atomic level as to how the proteins interact with other proteins, ligands or DNA and participate in regulatory and signaling pathways [1]–[4]. Furthermore, it helps identify prospective target/binding sites for the design of new drugs. Three-dimensional protein structure determination by X-ray crystallography basically involves four steps: 1) cloning, expression and purification of the target protein [5], 2) optimization of crystallization conditions for obtaining diffractable crystals, 3) diffraction data collection, 4) structure solving, refinement and analysis of the model.

With the advent of improved techniques for data collection like highly responsive X-ray detectors, synchrotron beam lines and cryomounting of crystals, the major challenge that remains is to find the right crystallization condition for obtaining highly diffractable crystals [6], [7]. Protein crystallization will take place when the protein solubility is reduced and the unit cells align in an ordered manner to allow crystal growth to proceed. Protein crystal growth is best explained by the phase diagram (Figure 5.1). According to the solubility curve, there are two regions: 1) undersaturation region and supersaturation region. The supersaturation region is further classified as precipitation zone, nucleation (labile) zone and metastable zone [6].



Source: Chayen et al, 2008, Reproduced with permission under the Nature Publishing Group License.

Figure 5.1: Phase diagram showing steps involved in protein crystallization.

In a typical crystallization experiment, the idea is to drive out the protein from the undersaturation region to supersaturation region by lowering its solubility and increasing protein concentration in the solution. In the supersaturation region, if the protein aggregates rapidly then it will form an amorphous precipitate in the precipitation zone. However, in the nucleation zone, if it precipitates gradually in an ordered fashion then it would give rise to critical nucleus around which other molecules would attach and grow. While in the metastable zone, the nuclei will grow to form the crystals.

Till date, there is no well-defined sequential procedure to obtain diffractable crystals. Thus, a large number of conditions need to be checked to initiate protein crystallization. Some of the parameters that affect crystallization include protein purity, concentration and stability, buffer type and concentration, pH, temperature, precipitant type and concentration, additives, heavy metals, sample handling and method used for setting drops.

After obtaining protein crystals, they are X-ray diffracted and the diffraction pattern is recorded. Generally, methods like Molecular Replacement (MR), Isomorphous Replacement (IR), Single-wavelength Anomalous Dispersion (SAD) or Multiwavelength Anomalous Dispersion (MAD) are used for solving the phase problem [8], [9]. Subsequently, the model is built based on the electron density map of the atomic coordinates. Finally, refinement ensures that the atomic model agrees with the diffraction pattern and there are no unaccounted regions in the electron density map.

In this chapter, crystallization of the Sebox3 protein would be discussed and thereafter, the Sebox3 protein model would be evaluated.

5.2 Materials and Methods

5.2.1 Protein expression and concentration

The construction of the Sebox3 clone has been described earlier in section 2.2.2. The Sebox3 protein was expressed and purified according to the protocol mentioned in Section 2.2.3, 2.2.4. Before the final concentration, the protein was dialyzed in a stepwise manner, to gradually reduce the salt concentration from 150 mM NaCl to 100 mM NaCl and finally, into a buffer containing 50 mM Tris-HCl (pH 7.4) and 50 mM NaCl with 5% glycerol added to stabilize and store the protein.

Thereafter, Amicon Ultra 0.5 mL centrifugal filter with a Molecular Weight Cut-Off (MWCO) of 10 kDa was used to concentrate the protein. At a time, 400 μ L of protein sample was added to the top of the spin column and centrifuged at 2,500 rpm at 4°C. The protein was checked for precipitation at regular intervals. Finally, the concentrated sample was recovered by reverse spinning the retentate at 2,000 rpm for 2 min at 4°C and collected in a sterile vial.

This concentrate was centrifuged and filtered through a 0.22 μm sterile syringe filter membrane immediately before setting crystallization drops to remove any traces of protein precipitate.

5.2.2 Crystallization trials

The ideal protein concentration for setting up crystallization trays was determined by pre-crystallization examination of various conditions. For crystallization of the Sebox3 protein, the hanging drop vapour diffusion technique was used. In this method, 2 μL of the protein sample and equal volume of crystallization reagent from the well reservoir was mixed on a clean siliconized glass coverslip and inverted over a well of a Corning 24 well culture plate containing 1 mL of the reservoir solution. The edge of the well was sealed using silicone vacuum grease and incubated at 12°C in Rumed cooled crystallization incubator. Initially, crystals were screened at 20°C but it led to rapid precipitation of the protein, therefore the temperature of 12°C was eventually used.

Buffer optimization: The zone of precipitation was estimated by mixing increasing concentration of polyethylene glycol (PEG) and ammonium sulphate (AMS) with the protein sample. Based on the PEG and AMS concentrations, at which the protein precipitation occurred, the crystallization screening grid was curated with a stock protein concentration of 7 mg/mL (Table 5.1). Thereafter, reagents from Hampton Research Crystal Screen Formulations were also used to determine the ideal condition for protein crystal formation (Table 5.2). The reservoir solution/mother liquor was buffered between pH 4.0 to pH 8.0. Buffers like citrate (pH 4.0, pH 5.0), MES (pH 6.0), MOPS (pH 6.5), HEPES (pH 7.0) and Tris-HCl (pH 8.0) were used to examine the various buffering systems for the possibility of crystal formation (Table 5.3).

Buffer optimization was extensively carried out to find out the range at which the protein would show desirable precipitation by equilibration with 1 mL of reservoir buffer.

Table 5.1: Different concentrations of PEG 8000, PEG 6000 and Ammonium sulphate used for checking zone of protein precipitation.

PEG 8000 (%)	PEG 6000 (%)	Ammonium sulphate (M)
5	5	0.8
10	10	1.6
20	20	2.4
40	40	3.0

Table 5.2: Conditions from Hampton Research Crystal Screen Formulations showing different salt, buffer and precipitant concentration.

Reagents	Salt	Buffer	Precipitant
1	0.2 M Calcium chloride dihydrate	0.1 M Sodium acetate trihydrate (pH 4.6)	30% v/v (+/-)-2-Methyl-2,4-pentanediol
2			0.4 M Potassium sodium tartrate tetrahydrate
3			0.4 M Ammonium phosphate monobasic
4		0.1 M Tris hydrochloride (pH 8.5)	2.0 M Ammonium sulfate
5	0.2 M Sodium citrate tribasic dihydrate	0.1 M HEPES sodium (pH 7.5)	30% v/v (+/-)-2-Methyl-2,4-pentanediol
6	0.2 M Magnesium chloride hexahydrate	0.1 M Tris hydrochloride (pH 8.5)	30% w/v Polyethylene glycol 4,000
7		0.1 M Sodium cacodylate trihydrate (pH 6.5)	1.4 M Sodium acetate trihydrate
8	0.2 M Sodium citrate tribasic dihydrate	0.1 M Sodium cacodylate trihydrate (pH 6.5)	30% v/v 2-Propanol
9	0.2 M Ammonium acetate	0.1 M Sodium citrate tribasic dehydrate (pH 5.6)	30% w/v Polyethylene glycol 4,000
10	0.2 M Ammonium acetate	0.1 M Sodium acetate trihydrate (pH 4.6)	30% w/v Polyethylene glycol 4,000

Table 5.3: The initial grid design of a 24-well culture plate for Sebox3 protein crystallization with buffers like citrate, MES, MOPS, HEPES and Tris-HCl ranging from pH 4.0-pH 8.0 were used.

0.1 M Citrate (pH 4.0) + 25% PEG (1:1)	0.1 M Citrate (pH 5.0) + 25% PEG (1:1)	0.1 M MES (pH 6.0) + 25% PEG (1:1)	0.1 M MOPS (pH 6.5) + 25% PEG (1:1)	0.1 M HEPES (pH 7.0) + 25% PEG (1:1)	0.1 M Tris-HCl (pH 8.0) + 25% PEG (1:1)
0.1 M Citrate (pH 4.0) + 25% PEG (2:1)	0.1 M Citrate (pH 5.0) + 25% PEG (2:1)	0.1 M MES (pH 6.0) + 25% PEG (2:1)	0.1 M MOPS (pH 6.5) + 25% PEG (2:1)	0.1 M HEPES (pH 7.0) + 25% PEG (2:1)	0.1 M Tris-HCl (pH 8.0) + 25% PEG (2:1)
0.1 M Citrate (pH 4.0) + 25% PEG (3:1)	0.1 M Citrate (pH 5.0) + 25% PEG (3:1)	0.1 M MES (pH 6.0) + 25% PEG (3:1)	0.1 M MOPS (pH 6.5) + 25% PEG (3:1)	0.1 M HEPES (pH 7.0) + 25% PEG (3:1)	0.1 M Tris-HCl (pH 8.0) + 25% PEG (3:1)
0.1 M Citrate (pH 4.0) + 1.5 M Ammonium sulphate (1:1)	0.1 M Citrate (pH 5.0) + 1.5 M Ammonium sulphate (1:1)	0.1 M MES (pH 6.0) + 1.5 M Ammonium sulphate (1:1)	0.1 M MOPS (pH 6.5) + 1.8 M Ammonium sulphate (1:1)	0.1 M HEPES (pH 7.0) + 1.8 M Ammonium sulphate (1:1)	0.1 M Tris-HCl (pH 8.0) + 1.8 M Ammonium sulphate (1:1)

Based on the initial hits in the crystallization screen, the conditions were narrowed down to a pH range of 6.0 to 7.5 (Table 5.3). The protein: buffer ratios of 1:1, 2:1 and 3:1 were also varied to test for the concentration of protein as a variable. After the zone of precipitation test, 25% (w/v) PEG, 35% (w/v) PEG, 1.5 M AMS and 1.8 M AMS were used as precipitants in the screen. Any crystal formed was picked up with a crystal loop and quickly transferred into liquid nitrogen.

Table 5.4: The grid design of a 24-well culture plate for Sebox3 protein crystallization. Buffers like MES, MOPS and HEPES ranging between pH 6.0 - pH 7.5 were used.

0.1 M MES (pH 6.0) + 35% PEG (1:1)	0.1 M MOPS (pH 6.5) + 35% PEG (1:1)	0.1 M HEPES (pH 7.5) + 35% PEG (1:1)	0.1 M MES (pH 6.0) + 25% PEG (1:1)	0.1 M MOPS (pH 6.5) + 25% PEG (1:1)	0.1 M HEPES (pH 7.5) + 25% PEG (1:1)
0.1M MES (pH 6.0) + 35% PEG (2:1)	0.1M MOPS (pH 6.5) + 35% PEG (2:1)	0.1M HEPES (pH 7.5) + 35% PEG (2:1)	0.1M MES (pH 6.0) + 25% PEG (2:1)	0.1 M MOPS (pH 6.5) + 25% PEG (2:1)	0.1 M HEPES (pH 7.5) + 25% PEG (2:1)
0.1 M MES (pH 6.0) + 35% PEG (3:1)	0.1M MOPS (pH 6.5) + 35% PEG (3:1)	0.1 M HEPES (pH 7.5) + 35% PEG (3:1)	0.1 M MES (pH 6.0) + 25% PEG (3:1)	0.1 M MOPS (pH 6.5) + 25% PEG (3:1)	0.1M HEPES (pH 7.5) + 25% PEG (3:1)
0.1 M MES (pH 6.0) + 1.5 M Ammoni um sulphate (1:1)	0.1 M MOPS (pH 6.5) + 1.5 M Ammoni um sulfate (1:1)	0.1 M HEPES (pH 7.5) + 1.5 M Ammonium sulfate (1:1)	0.1 M MES (pH 6.0) + 1.8 M Ammonium sulfate (1:1)	0.1 M MOPS (pH 6.5) + 1.8 M Ammonium sulfate (1:1)	0.1 M HEPES (pH 7.5) + 1.8 M Ammonium sulfate (1:1)

5.2.3 Diffraction at Synchrotron beamline

Crystals were diffracted at the Indus II beamline at Raja Ramanna Centre for Advanced Technology at Indore, India. Crystals were transferred in liquid nitrogen filled cryo cans by a 36-hour journey by train.

5.2.4 Homology modelling of Sebox3

Homology modelling or comparative modelling of a protein refers to building up of a three-dimensional atomic-resolution model of the target protein. This is based on the primary sequence of the target protein and an experimental three-dimensional structure of a structurally and a functionally related protein called “template”. The methodology of homology modelling could be divided into following steps. The first step includes searching of variously related sequences to select a perfect template representing a homologous protein of Sebox3 both structurally and functionally. In order to perform this task, a protein BLAST search was conducted on the PDB Database (from RCSB) using the online version available with NCBI [10]. The BLAST query revealed that the sequence homology of Sebox3 never exceeded 20%, which was not optimal to develop a good quality model. For further analysis, the sequences were imported into the CLUSTALW server [11].

Results from CLUSTALW showed that PleD, a response regulator from *Caulobacter vibrioides* (CvPleD), (PDB ID 2V0N) [12] too displayed less than 20% identity when the complete sequence of Sebox3 was considered. However, considering the catalytic domain (C-terminal domain) alone increases the identity to almost 35% (identity: 34.7%, strongly similar: 22.4%), which was sufficient to generate a model.

Subsequently, we decided to develop a homology model of the Sebox3 catalytic domain, which contains the GGEEF active site based on the C-terminal domain of CvPleD as a model. This structure was used as the template for homology modelling by Swiss-Model server [13]. The molecular visualization program Coot [14] was used to modify the position of amino acid residues of the GGEEF active site (core catalytic region). The initial model was further improved by energy minimization using the GROMACS 4.5 [15]–[17] software package. After optimization, the minimized model was validated using PDBsum. The evaluation parameters considered were standard bond length, bond angle, Ramachandran plot etc. [18], [19].

5.2.5 Molecular docking of GTP at active site of Sebox3

Molecular docking for this study was carried out by using the program AutoDock Vina 1.1.2 [20]. AutoDock Vina utilizes its global search algorithm to find the best binding pose after a thorough docking calculation. The coordinates of GTP and Mg^{2+} were constructed with the help of unrelated structure from PDB database (RCSB) [21]. The AutoDock tool (ADT) had been extensively used to prepare all the structures and set up the docking protocol [22]. The developed homology model of Sebox3 was used as a rigid receptor. All the docking calculations were run to produce 10 docking poses for the ligand. The docked pose of the ligand with higher binding affinity and biologically relevant conformation was considered for further analysis.

5.3 Results and Discussion

5.3.1 Concentration of Sebox3 protein

Sebox3 protein was concentrated to 7 mg/mL using 0.5 mL spin column concentrator and checked for concentration, homogeneity and degradation during the concentration process on a SDS-PAGE gel (Figure 5.2).

Beyond the concentration of 7 mg/mL, the protein precipitated and hence, further concentration was not carried out. Protein concentration served as an important parameter for setting up hanging drops by vapour diffusion in the crystallization screen. If the protein concentration was low then the crystallization drop remained clear and if the protein concentration was too high then it rapidly led to the formation of an amorphous precipitate (Figure 5.3A). Both the conditions, that is, clear and amorphous precipitate never led to crystal formation. By utilizing the right conditions the concentrated protein sample can be efficiently used and the probability of obtaining crystals can be amplified.

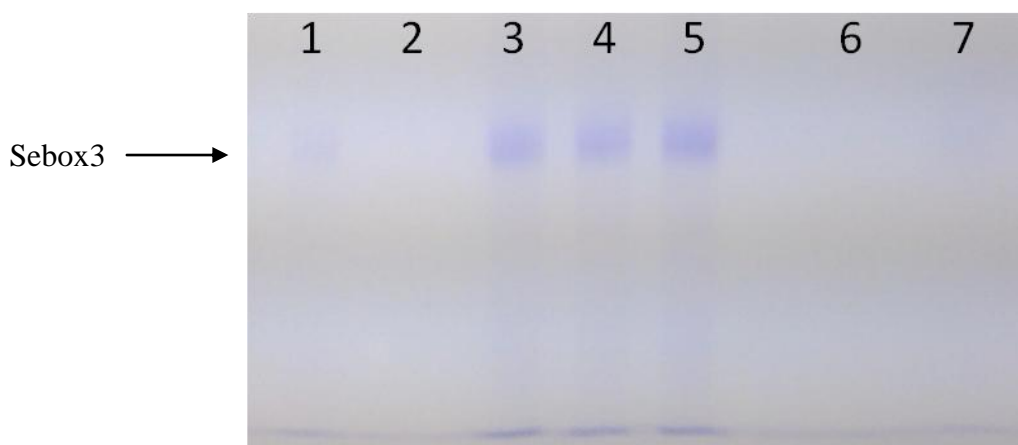


Figure 5.2: SDS-PAGE showing retentate and filtrate collected at different stages of Sebox3 protein concentration. Lane 1: Initial before concentration Sebox3 protein, lane 2: NA, lane 3: retentate 1, lane 4: retentate 2, lane 5: retentate 3, lane6: filtrate 1 and lane 7: filtrate 2.

5.3.2 Crystallization of Sebox3 protein

The initial hits were obtained by varying the type of precipitant, concentration of precipitant, concentration of protein and temperature. From the initial trials, the optimum protein concentration for crystallization was found to be 5 mg/mL. At this concentration, almost half of the drops in a crystallization culture plate were clear and the remaining half of the drops had white precipitate. The crystallization screens were initially incubated at 4°C and finally at 12°C by the hanging drop vapour diffusion method. Almost 70% of drops set at a high concentration of protein (7 mg/mL) precipitated immediately.

Subsequently, a lower concentration (5 mg/mL) was used for testing the crystallization conditions. This could be considered as a sign of enhanced stability at low concentrations of protein. In the initial crystallization screen, heavy precipitation was observed in the condition with 0.1 M MOPS (pH 7.0), 25% (w/v) PEG and 3:1 protein:reservoir solution ratio (Figure 5.3A). For the next set of crystallization trials, buffers with pH ranging from 6.0 to 7.5 like MES, MOPS and HEPES were used.

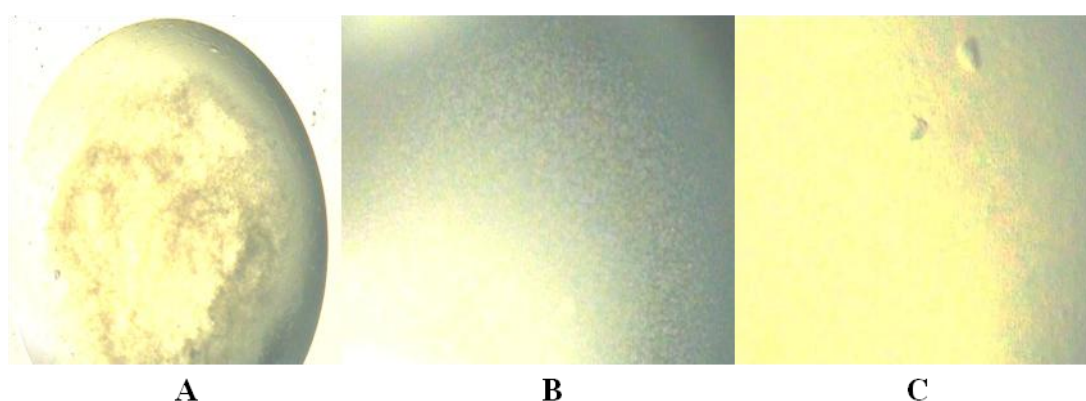


Figure 5.3: The images of different conditions in hanging drop vapour diffusion method. A) Amorphous precipitate, B) White crystalline precipitate, C) Microcrystal.

White crystalline precipitate was obtained in condition with 0.1 M HEPES (pH 7.0), 25% (w/v) PEG and 3:1 protein:reservoir solution ratio (Figure 5.3B). Microcrystals were

observed in condition with 0.1 M HEPES (pH 7.0), 25% (w/v) PEG and 2:1 protein:reservoir solution ratio after incubation of 14 days at 12°C by (Figure 5.3C). Better crystals between 40-80 μm was obtained in the final crystallization condition with 0.1 M HEPES (pH 7.0), 30% (w/v) PEG and 1:1 protein: reservoir solution ratio after incubation of 25 days at 12°C (Figure 5.4)

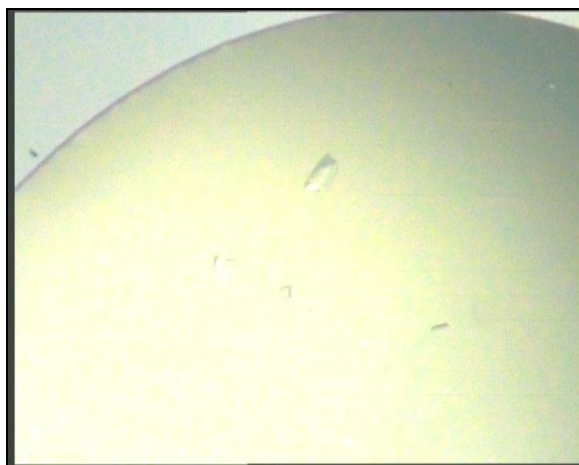


Figure 5.4: Image of Crystals of Sebox3 protein

5.3.3 Weak diffraction of Sebox3 crystals

The crystals of Sebox3 were carried in liquid nitrogen and diffracted at the beamline 23 of the Indus-II synchrotron at RRCAT, Indore. Though the crystals looked morphologically unchanged, the effect of 36 hours of travel seemed to have an effect, with distinct ice rings visible in the diffraction images. Nevertheless, the diffraction was not strong enough, with few spots in the low resolution zones (Figure 5.5). Soaking crystals in various cryo-protectants did not result in any further improvements. Lack of viable diffraction data prompted us to look into the modelled structure of Sebox3 (next sections)

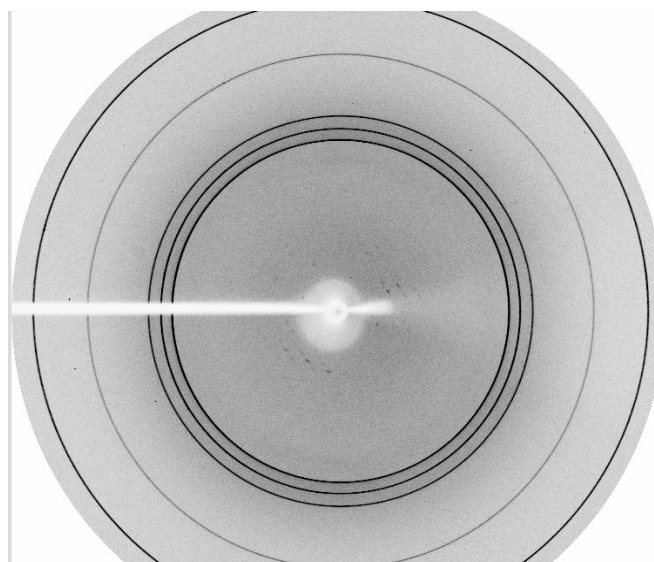


Figure 5.5: Diffraction image obtained for Sebox3 with the detector at 110 mm.

5.3.4 Alignment of Sebox3 with other diguanylate cyclases

The sequence of proteins (in FASTA format) with a solved crystal structure was downloaded from the RSCB Protein Data Bank (<https://www.rcsb.org/pdb/home/home.do>) and aligned together by the multiple sequence alignment tool (ClustalW). ESPript 3.0 analysis shows that the GGEEF motif is highly conserved in all the aligned sequences and hence proves its necessity in its functionality as a diguanylate cyclase (Figure 5.6).

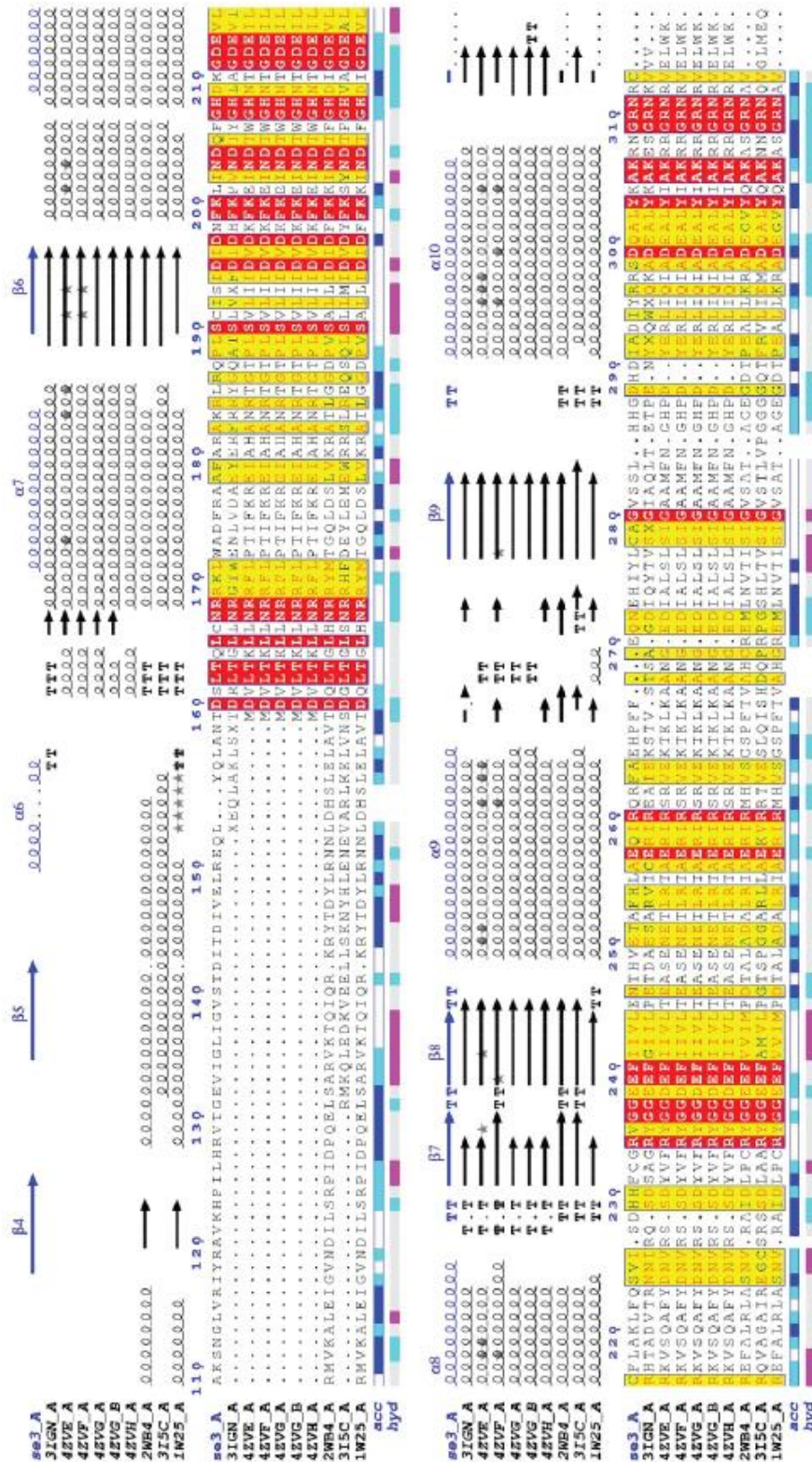


Figure 5.6: Multiple Sequence Alignment by ClustalW showing conserved amino acids and secondary structure.

5.3.5 Homology model of Sebox3

The full-length protein of Sebox3 is a 321 amino acid long sequence with two distinct domains – the N terminal domain with the PAS signature sequence, and the C terminal domain harbouring the GGEEF residues, with a probable EAL domain in reverse (vide previous mention). The full sequence though, has low homology with other diguanylate cyclases. Only when we considered the truncated sequence without the N-terminal domain, it was possible to attain appreciable homology with a previously available structure of the CvPleD. The residues (286-454) of CvPleD and Sebox3 catalytic domain (amino acid residues 154-321) displayed about 35% homology.

Hence, CvPleD (PDB ID 2V0N) served as the template for the development of the homology model. The final structure of the modelled Sebox3 showed a remarkable conservation of the GGEEF architecture that is present in the diguanylate cyclases. The structural alignment of the model and the template CvPleD highlights the similarity in the tertiary structure of the two despite the 35% sequence similarity (Figure 5.7).

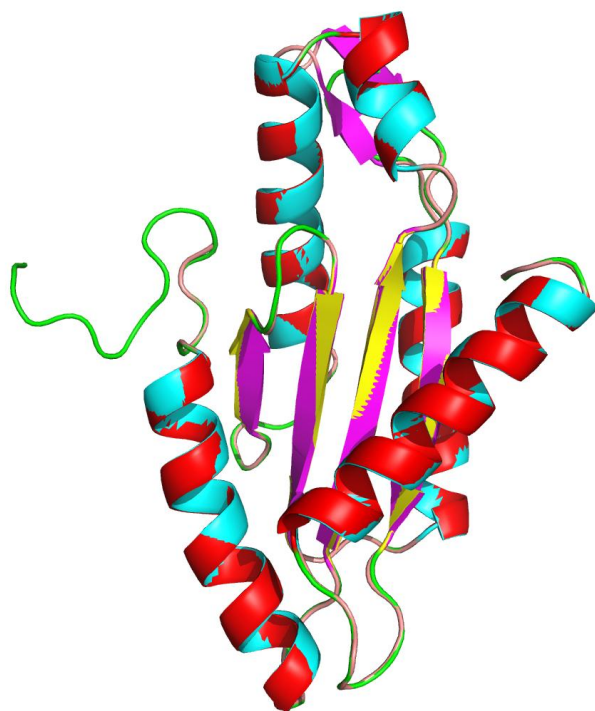


Figure 5.7: Remarkable structural similarities between the model and the template.

Refinement of the initial structure by optimization and molecular dynamics simulations yielded the structure shown here as Figure 5.8. The active site showed the presence of 5 important alpha helices and 3 beta sheets in the structure, the hallmarks of the diguanylate cyclase motif. The rest of the active site consists of turns and loops, the GGEEF sequence being part of one of these.

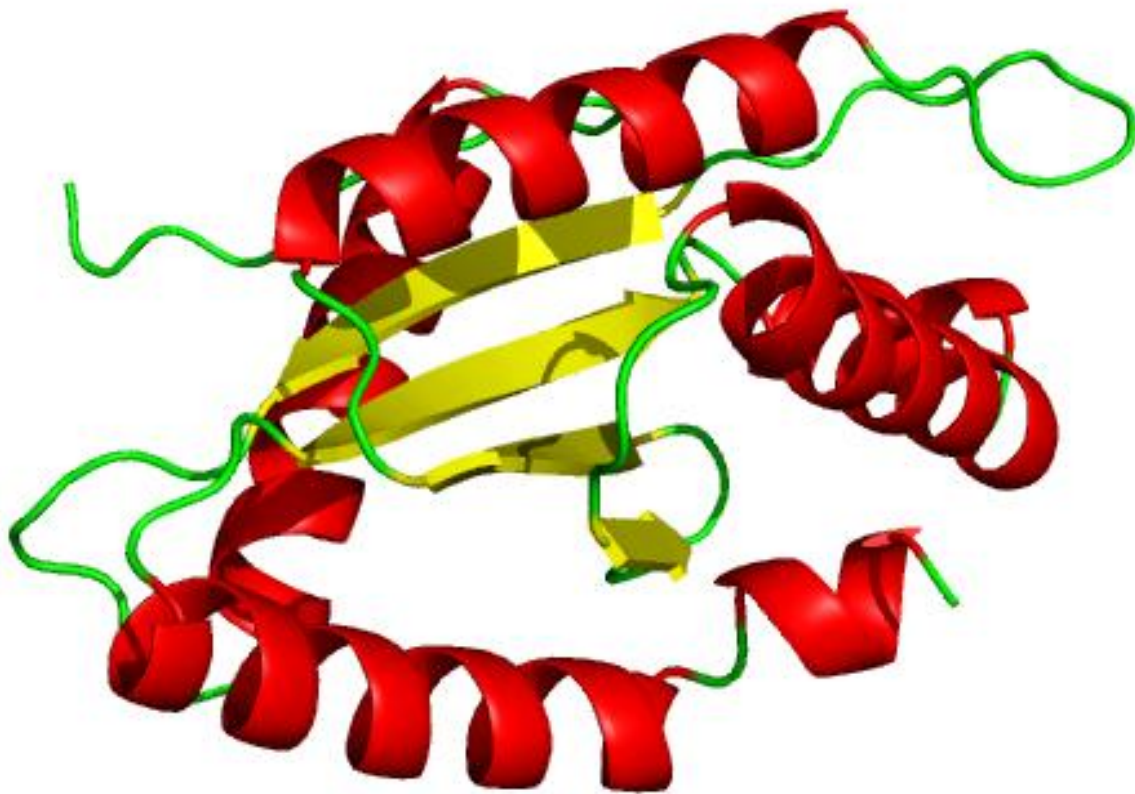


Figure 5.8: Homology model of the C terminal domain of Sebox3.

The PDBsum evaluation showed that the energetic and other properties were comparable with those of the template. Other parameters like the bond angles, bond lengths, side chain conformations were within the limits of a well-defined model. The Ramachandran plot (Figure 5.9) of the model showed a score of 91.8% residues were in the most favoured region, 4.1% in the additional allowed regions and 2.1% in generously allowed regions. Moreover, only three residues were in the disallowed region, proving it to be an acceptable homology model of Sebox3.

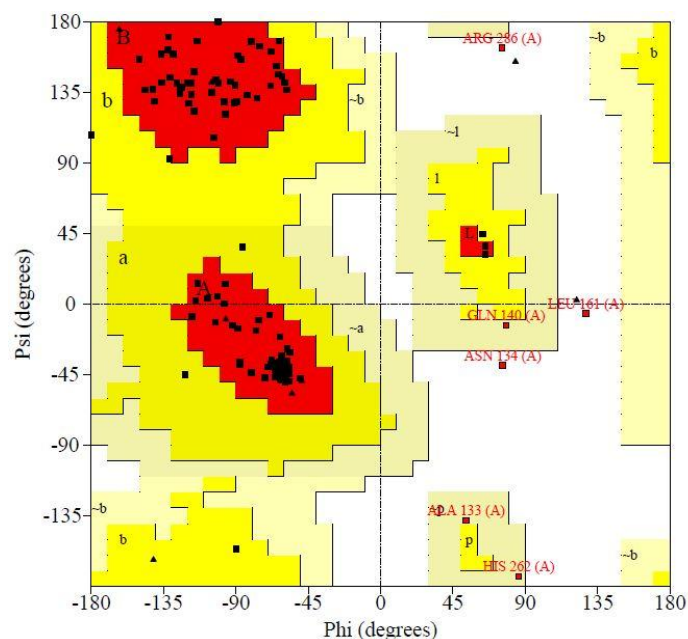


Figure 5.9: Ramachandran Plot of the modelled structure of the C terminal domain of Sebox3

5.3.6 Docking of the GTP molecule to the active site of Sebox3

There are multiple nonbonding interactions that play critical and important roles in the functioning of protein-ligand interactions. Among the GGEEF domain residues, the GTP can be clearly seen to interact with the second G of the motif (Figure 5.10). This residue (Glu238) from the GGEEF loop forms hydrogen bonds with the oxygen atoms of alpha and beta phosphates. Other residues like Asp195 and Lys306 also interacted with oxygens of the gamma phosphate of GTP. Additionally, the purine base of the GTP oriented slightly towards the negatively charged glutamates as well. This conclusively demonstrated that the modelled structure with the diguanylate cyclase active site had a GTP binding activity, which had been demonstrated by the HPLC assay and other experiments.

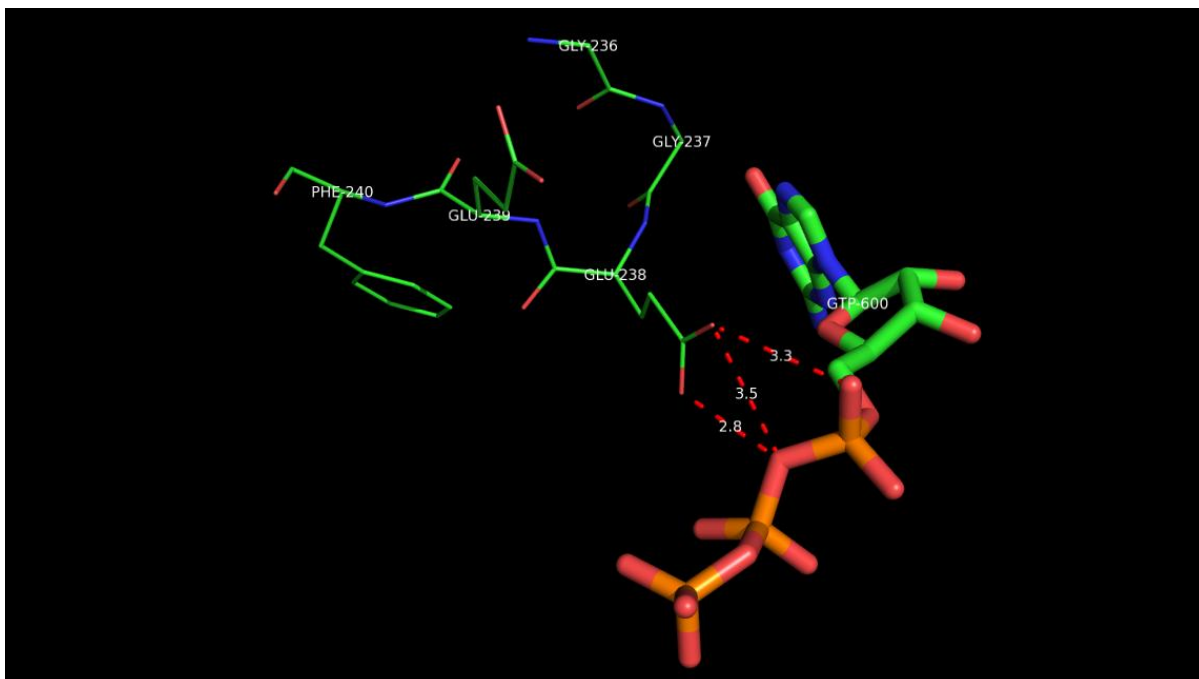


Figure 5.10: Docking of GTP against the active site of Sebox3.

5.4 Conclusion

The protein was purified to homogeneity as seen in the SDS-PAGE gel stained by Coomassie. The protein was dialysed and the concentration of the NaCl was gradually brought down from 150 mM to 50 mM in a buffer containing 50 mM Tris-HCl (pH 7.4). The protein was concentrated to 7 mg/mL; however, only 5 mg/mL was used for setting the drops as high concentration led to immediate precipitation of protein.

During the initial screening, the trials were carried out with buffers ranging in pH from 4.0 to 8.0. Finally, crystals were obtained in the condition with 0.1 M HEPES (pH 7.0), 30% (w/v) PEG and 1:1 protein:reservoir solution ratio after incubation of 25 days at 12°C. Crystals though did not diffract well enough to attempt structure solution.

Hence, attempts were made to model the structure of Sebox3 by *in silico* methods. *In silico* analysis showed that the GGEEF domain architecture was conserved in the homology model derived by considering the sequence similarity of the C terminal domain against the closest diguanylate cyclase structure of CvPleD. The docking experiments showed the interactions of the engineered GTP molecule to the second G of the active site of Sebox3. This explained the diguanylate cyclase activity of the protein as demonstrated by the c-di-GMP production assay and the other indirect methods. The docking of GTP with Sebox3 protein needs to be validated by Molecular Dynamic Simulation and is currently underway.

5.5 References

- [1] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures," *FEBS J.*, vol. 275, no. 1. pp. 1–21, 2008.
- [2] G. Rhodes, "An Overview of Protein Crystallography," *Crystallogr. Made Cryst. Clear*, pp. 7–30, 2006.
- [3] A. Ilari and C. Savino, "*Protein Structure Determination by X-ray Crystallography*", *Methods Mol Biol.*, vol. 452, pp. 63-87, 2008.
- [4] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination," *FEBS J.*, vol. 280, no. 22, pp. 5705–5736, 2013.
- [5] C. W. Goulding and L. J. Perry, "Protein production in *Escherichia coli* for structural studies by X-ray crystallography," *J. Struct. Biol.*, vol. 142, no. 1, pp. 133–143, 2003.
- [6] N. E. Chayen and E. Saridakis, "Protein crystallization: from purified protein to diffraction-quality crystal.," *Nat. Methods*, vol. 5, no. 2, pp. 147–153, 2008.
- [7] R. Henderson, "Cryo-Protection of Protein Crystals against Radiation Damage in Electron and X-Ray Diffraction," *Proc. R. Soc. B Biol. Sci.*, vol. 241, no. 1300, pp. 6–

- 8, 1990.
- [8] A. González, “A comparison of SAD and two-wavelength MAD phasing for radiation-damaged Se-MET crystals,” in *J. Synchrotron Rad.*, 2007, vol. 14, no. 1, pp. 43–50.
- [9] W. I. Karain, G. P. Bourenkov, H. Blume, and H. D. Bartunik, “Automated mounting, centering and screening of crystals for high-throughput protein crystallography,” in *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2002, vol. 58, no. 10 II, pp. 1519–1522.
- [10] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool.,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–10, 1990.
- [11] J. D. Thompson, D. G. Higgins, and T. J. Gibson, “CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice,” *Nucleic Acids Res.*, vol. 22, no. 22, pp. 4673–4680, 1994.
- [12] P. Wassmann *et al.*, “Structure of BeF₃--Modified Response Regulator PleD: Implications for Diguanylate Cyclase Activation, Catalysis, and Feedback Inhibition,” *Structure*, vol. 15, no. 8, pp. 915–927, 2007.
- [13] N. Guex, M. C. Peitsch, and T. Schwede, “Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective,” *Electrophoresis*, vol. 30, no. SUPPL. 1, 2009.
- [14] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, “Features and development of Coot,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 4, pp. 486–501, 2010.
- [15] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 43–56, 1995.
- [16] M. J. Abraham and J. E. Gready, “Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5,” *J. Comput. Chem.*,

- vol. 32, no. 9, pp. 2031–2040, 2011.
- [17] S. Pronk *et al.*, “Gromacs 4.5: A high throughput and highly parallel open source molecular simulation toolkit,” *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.
- [18] R. A. Laskowski, “PDBsum new things,” *Nucleic Acids Res.*, vol. 37, no. SUPPL. 1, 2009.
- [19] R. A. Laskowski, “PDBsum: summaries and analyses of PDB structures.,” *Nucleic Acids Res.*, vol. 29, no. 1, pp. 221–2, 2001.
- [20] O. Trott and A. J. Olson, “AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading,” *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–61, 2010.
- [21] E. Sabini, S. Hazra, S. Ort, M. Konrad, and A. Lavie, “Structural Basis for Substrate Promiscuity of dCK,” *J. Mol. Biol.*, vol. 378, no. 3, pp. 607–621, 2008.
- [22] G. Morris and R. Huey, “AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility,” *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–2791, 2009.

Chapter Conclusions and Summary

V. cholerae, the causal organism for cholera, still continues to threaten the lives of people throughout the world. The bacterial secondary messenger c-di-GMP plays a significant role in regulating the expression of genes involved in the shift from motile to sessile form of the host organism. GGD(/E)EF domains through the action of diguanylate cyclase modulate the *in vivo* turnover of c-di-GMP. In the pursuit of understanding the role of c-di-GMP signaling in *V. cholerae*, GGD(/E)EF domains have gained popularity in the recent past. In this study, putative GGEEF protein domain from *V. cholerae* classical strain was studied and analyzed for its structural and functional features. The outcomes of this study were summarized as mentioned below:

Chapter 1 (Introduction and Literature Review):

In this chapter, cholera, its transmission, control and the causative agent – *V. cholerae*, have been elaboratively discussed. It throws light on how *V. cholerae* forms biofilm by the differential expression of genes that result in the formation of extracellular matrix. The intricate details of c-di-GMP signaling mechanism and the role of GGD(/E)EF and EAL domain proteins in biofilm formation have also been deliberated. The structure of other GGD(/E)EF domain proteins from other sources was also discussed.

Chapter 2 (Biophysical Characterization of Sebox3 Protein):

The recombinant Sebox3 protein was cloned and expressed in the pGEX-6P1 vector in DH5 α as the protein was toxic to the generally utilized expression host BL21 (DE3). Additionally, the use of other vectors like pET vectors did not aid in the expression as well as the solubility of the protein. Due to its huge size and dimeric nature, the removal of the GST tag was necessary before carrying out the subsequent experiments. From the fluorimetric study, it is

evident that the local environment of the tryptophan residues is positively charged. Furthermore, the level of free thiol groups in the protein throws light on the conformational rigidity of the protein. The Sebox3 protein was found to have a functionally active GGEEF domain as shown by the conversion of the substrate GTP to c-di-GMP. When the protein was subjected to limited proteolysis with proteolytic enzymes, a stable 8 kDa fragment appeared indicating the presence of flexible regions in the protein.

Chapter 3 (Construction of truncates of Sebox3: Sebox31 and Sebox32; and functional characterization):

This chapter dealt with the construction of Sebox31 (130 – 321 amino acids) and Sebox32 (161 – 321 amino acids). These truncates were designed to enhance the yield and the stability of the Sebox3 protein. Both the truncates had higher yield and stability as compared to the Sebox3 protein, with Sebox32 showing the highest stability. The biofilm forming ability as well as the retarded motility in soft agar was retained in both Sebox31 and Sebox32. Sebox3 showed highest biofilm forming ability as exemplified in the crystal violet assay. In the SEM analysis, the cells of the biofilm were held together by exopolysaccharides which maintained the three-dimensional structure of biofilm. The motility tests of the strains showed that the swarming ability of the strains with GGEEF domain was retarded. When the GGEEF domain was functional, it synthesized c-di-GMP by the associated diguanylate activity. High turnover of intracellular c-di-GMP affected the behaviour of the host cell ultimately leading to biofilm formation and repression of cell motility. Thus, it can be deduced that the GGEEF protein domain was active *in vivo* in these strains.

Chapter 4 (Comparative characterization of Sebox3 versus Sebox31 and Sebox32):

In this chapter, the changes in the secondary structure of the Sebox31 and Sebox32 were checked. Due to the truncation, there was considerable change in the secondary structure of

the proteins. By the thermal shift assay, it was observed that the binding of the ligand GTP stabilized the protein by increasing the T_m of the protein. The shortening of the N-terminal in the protein did not affect the functionality of the truncates. All the three proteins showed diguanylate cyclase activity. It also did not affect the oligomeric status of the proteins, with the proteins remaining monomeric in solution. The positive quencher, potassium iodide, still remained the major quencher of intrinsic fluorescence amongst all the other quenchers.

Chapter 5 (The Structure of Sebox3):

The Sebox3 protein was concentrated and crystallization trials were carried out in pH ranging from pH 4.0 – pH 8.0. Classical protein precipitating agents like polyethylene glycol and ammonium sulfate in Hampton Crystal Screens were used. Proteins crystals were observed in the condition with 0.1 M HEPES (pH 7.0), 30% (w/v) PEG and 1:1 protein:reservoir solution ratio after incubation for 25 days at 12°C. As the protein crystals did not diffract well, the protein structural model was elucidated by *in silico* methods. The homology model was made based on the template of CvPleD and showed 5 α -helices and 3 β -sheets.

Future scope of work

- Introduction of *sebox3* gene in wild-type *V. cholerae* 0395, its functional characterization and comparison with sebox gene in *E. coli* DH5 α
- Assessment and comparison of phosphodiesterase activity of Sebox3, Sebox31 and Sebox32
- Cloning, expression and purification of PAS domain; and its effect on catalytic activity of Sebox proteins

Appendices

Appendix A

Reagents preparation

- 1) 5X Protein Loading Dye
 - Tris – HCl 0.6 M
 - Glycerol 25% (v/v)
 - β – mercaptoethanol 14.4 mM
 - SDS 2% (w/v)
 - bromophenol blue 0.1% (w/v)
 - pH 6.8

- 2) Staining Solution
 - Coomassie Brilliant Blue R250 0.1%
 - Acetic acid 10% (v/v)
 - Methanol in water 45% (w/v)

- 3) Destaining Solution
 - Acetic acid 10% (v/v)
 - Methanol in water 10% (v/v)

- 4) 5X TBE buffer
 - Tris base 54g
 - Boric Acid 27.5g
 - EDTA 0.5M (pH 8.0) 20 mL
 - Make up the volume to 1 litre.

- 5) SDS PAGE Running Buffer
 - Tris Base 25 mM
 - Glycine 192 mM
 - SDS 0.1%

- 6) STET Buffer
Sucrose 8%
Triton X-100 5%
EDTA 50 mM
Tris- HCl (pH 8.0) 50 mM)
- 7) Alkaline Lysis Solution I
Glucose 50 mM
Tris-HCl (pH 8.0) 25 mM
EDTA (pH 8.0) 10 mM
- 8) Alkaline Lysis Solution II
NaOH 0.2 N
SDS 1% (w/v)
- 9) Alkaline Lysis Solution III
5 M Potassium acetate 60 mL
Glacial acetic acid 11.5 mL
Deionized water 28.5 mL
- 10) TE Buffer
Tris - HCl 10 mM
EDTA 1.0 mM
(pH 8.0)

Appendix B

List of license numbers

- 1) Figure 1.6 License Number 4165180622865
- 2) Figure 5.1 License Number 4153570982008
- 3) Figure 1.10 ASM, Creative Commons Attribution 4.0 International license
- 4) Figure 1.11 Francis and Taylor Group, Creative Commons Attribution 4.0 International license
- 5) Table 1.3 ASM, Creative Commons Attribution 4.0 International license
- 6) Figure 1.12 Francis and Taylor Group, Creative Commons Attribution 4.0 International license
- 7) Figure 1.13 Royal Society Publishing, Creative Commons Attribution 4.0 International license
- 8) Figure 1.16 Microbiology Society, Creative Commons Attribution 4.0 International license
- 9) Figure 1.18 ASM, Creative Commons Attribution 4.0 International license

Appendix C

List of Publications and Conference Presentations

Publications:

- 1) **D. Bandekar**, O. P. Chouhan, S. Mohapatra, M. Hazra, S. Hazra, and S. Biswas, “Putative protein VC0395_0300 from *Vibrio cholerae* is a diguanylate cyclase with a role in biofilm formation,” *Microbiol. Res.*, vol. 202, pp. 61–70, 2017 (SCI Impact Factor: 3.037, SJR: 1.09, H Index: 55).
- 2) O. P. Chouhan, **D. Bandekar**, M. Hazra, A. Baghudana, S. Hazra, and S. Biswas, “Effect of site-directed mutagenesis at the GGEEF domain of the biofilm forming GGEEF protein from *Vibrio cholerae*,” *AMB Express*, vol. 6, no. 1, p. 2, 2016 (SCI Impact Factor: 1.825, SJR: 0.65, H Index: 19). **Author Contribution:** Part of the results reported in this paper are included in Chapter 2 of this thesis with a major contribution by DB.
- 3) **D. Bandekar**, D. Singh and S. Biswas, “GGD(/E)EF domain: a ubiquitous eubacterial domain confers cell toxicity,” *Int. J. Biotech. Res.*, vol. 3, no. 1, p. 73-80, 2013.

Conference Proceedings:

- 1) **Bandekar D**, Biswas S. Screening and Optimization of Crystallization Conditions of VC0395_300 protein. Paper presented at: National Seminar on Crystallography and International Workshop on Application of X-ray diffraction for Drug Discovery; 2013 Nov 21-23; JNU, New Delhi.
- 2) **Bandekar D**, Biswas S. Deciphering the structural and functional aspects of GGEEF protein domain involved in lifestyle switch of *Vibrio cholerae*. Poster session presented at: 16th International Conference on the Crystallization of Biological Macromolecules; 2016 Jul 2-7; Prague, Czech Republic: Materials Structure in Chemistry, Biology, Physics and Technology, Vol 23:2; 2016. p. 164.

Appendix D

List of Workshops Attended:

1. “Workshop on Scientific Writing and Effective Communication” from 5th - 6th January 2015 at Goa University, Goa.
2. “International Workshop on Drug Development and Neglected Tropical Diseases” from 22nd - 27th February 2015 at University of Madras, Guindy Campus, Chennai.
3. “National Workshop on Structure Based Drug Designing” from 28th February - 1st March 2015 at Biophysics and Bioinformatics Infrastructure Facility, University of Madras, Guindy Campus, Chennai.
4. DBT BIRAC Workshop on “Bio-Entrepreneurship, Grant-Writing & Intellectual Property Management” from 18th - 19th February 2016 at BITS - Pilani, K. K. Birla Goa Campus, Goa.
5. Workshop on “Basics and Applications of Molecular Dynamics” from 19th - 21st November 2016 at Department of Chemical Engineering, BITS - Pilani, K. K. Birla Goa Campus, Goa.

Appendix E

Biography of the Candidate

Bandekar Divya Ramesh is a Research Scholar in the Vibrio and Structural Analysis Lab, of Dr. Sumit Biwas' group, Department of Biological Sciences, BITS – Pilani, K. K. Birla Goa Campus, Goa, India. She completed her B.Sc. in Biotechnology from Dhempe College of Arts and Science, Goa University in 2007. She received her M.Sc. in Biotechnology from St. Aloysius College, Mangalore University in 2009. She has cleared Graduate Aptitude Test in Engineering (GATE) 2009 with 95.86 percentile. Subsequently, she worked as Assistant Professor in Parvatibai Chowgule College of Arts and Science till 2012. She joined BITS – Pilani, K. K. Birla Goa Campus as a project JRF in August 2012 in the BRNS sponsored project entitled “Structure elucidation of VC0395_0300 from *Vibrio cholerae* (leading to an alternate method of checking cholera)”. She later received an institute fellowship for the completion of her Ph.D. studies.

She was awarded a travel grant by Department of Biotechnology, Conference, Travel, Exhibition and Popular Lectures (DBT, CTEP), Government of India, (Proposal code: DBT/CTEP/02/201600483), Centre for International Co-operation in Science (CICS) support (DO/Lr./TF-I/2016-17) and International Union of Crystallography Young Scientist Award for presenting a poster at the 16th International Conference on the Crystallization of Biological Macromolecules at Prague, Czech Republic from 2nd – 7th July 2016. She has published three research papers in International Journals.

Appendix F

Biography of the Supervisor

Dr. Sumit Biswas completed his Ph.D. in Bose Institute, Kolkata, under the supervision of Prof. Pinak Chakrabarti, as a CSIR fellow in 2008. His doctoral work elucidated the interfaces of protein-nucleic acid interactions, as well as the structure determination of two very important proteins. He went on to work as a DBT Research Associate in the DBT initiative, “Setting up of National Facility on Interactive Graphysics Computer System for Biomolecular Modelling, Molecular Dynamics & Structures” till 2009. Dr. Biswas joined BITS, Pilani, K K Birla Goa Campus as a faculty in 2009. He has since been involved as the Principal Investigator of four research projects funded by BRNS, DAE, DBT and DST, as well as the co-Investigator of a UGC project. His work in the institute involves the molecular mechanism and biology of the *Vibrio* life cycle, bioinformatics of con-coding RNA and protein-nucleic acid interactions, and therapeutic biology of natural products. Dr. Sumit Biswas has 16 publications in reputed journals and several conference publications to his name. He is also working on a book on Biophysics sanctioned by Prentice Hall of India.

Dr. Biswas has acted as the convenor for 4 meetings/symposia and workshops funded by DST, DBT, DSTE, Goa and BRNS. He is a life member of the Indian Crystallography Association, and a member of CholdInet (a WHO initiative for cholera research) and the Proteomics Society of India. He has received several awards and honours, the most recent being the prestigious EMBL Scholarship for presenting paper at EMBL Conference on Cancer Genomics, held at Heidelberg. Besides, he has delivered invited talks at different international conferences as well as institutes of repute like IIT, Kharagpur, IIT-BHU, etc. He has been actively involved as a reviewer of international journals from OUP, Elsevier, etc., as well as a question setter for DBT.

Presently, he has three registered Ph.D. students under his tutelage and numerous thesis dissertation and project students working with him. He has also been the certified Radiological Safety Officer for the Institute, and is instrumental in setting up biosafety practices in BITS.