

***CDFD***

**Annual Report  
April 2006 to March 2007**

*(English version)*



***CDFD***

**Center for DNA Fingerprinting and Diagnostics  
Hyderabad – 500 076**

# CONTENTS

1	From the Director's Desk
2	Mandate
3	<b>Services</b> DNA Fingerprinting Diagnostics  <b>Research</b> 1. Laboratory of Molecular Genetics 2. Laboratory of Genomics and Profiling Applications 3. Laboratory of Fungal Pathogenesis 4. Laboratory of Molecular Virology 5. Laboratory of Immunology 6. Laboratory of Bacterial Genetics 7. Laboratory of Computational Biology 8. Laboratory of Molecular Cell Biology 9. Laboratory of Structural Biology 10. Laboratory of Mammalian Genetics 11. Laboratory of Molecular Oncology 12. Laboratory of Cancer Biology 13. Laboratory of Computational & Functional Genomics 14. Laboratory of Transcription 15. Other Scientific services/facilities a) National genomics and transcriptomics facility b) Bioinformatics services c) Instrumentation services
4	Publications
5	Human Resource Development
6	Lectures, Meetings, Workshops, and Important events
7	Senior staff and officers of CDFD
8	Deputations Abroad of CDFD Personnel
9	Committees of the Institute
10	Budget and Finance
11	Auditor's Report

## **CDFD MANTRA**

“We dedicate ourselves to achieve world class excellence in basic research and simultaneously endeavour to transfer the benefits of modern biology to every section of society. We view our population not as a problem, but as an advantage, which is unique to India. We believe we can use our extraordinarily large pool of genetic diversity as a “genetic playground” to address a number of questions; questions which will continue to accumulate and will demand answers, as the new millennium is bombarded with increasing arrays of nucleotide sequences as a consequence of global genome projects. Compassion, when coupled with science, can realize its highest ideal, viz., improving the quality of life of the average citizen. This is the lifeline of our activities at CDFD. We are prepared to face the challenging tasks ahead of us with our exceptional human resources. We are confident that we will meet and surpass the expectations and responsibilities reposed in us by the creation of the CDFD”.

**FROM THE DIRECTOR'S DESK**

***From the Director's desk***

)

### **Director's Message**

On behalf of my colleagues and myself, I am happy to present the Annual Report of CDFD for the year 2006-07. This Centre uniquely combines two kinds of activities, the first being that of service provision in the areas of DNA profiling for law-enforcement agencies and of diagnostic tests for genetic disorders, and the second that of frontier-level research in various disciplines of molecular biology, in such a way that each supports and in turn is enriched by the other. I am sure that this symbiosis will be in clear evidence to the reader in much of the work that is described in the Report.

Given both its small faculty strength (more or less that of a modest-sized University department) and the fact that it is still relatively young, the Centre has achieved an impressive record of publications in international peer-reviewed journals and of patents in this reporting year. Additionally, several awards and honors have come its way, including the Young Scientist Medal of the Indian National Science Academy (Dr P Prachee), Innovative Young Biotechnologist Award (Drs N Madhusudhan Reddy and Rupinder Kaur), Innovative Award for Young Scientist at Bio-Asia 2007 meet (Dr N Madhusudhan Reddy), and the Scopus Young Scientist Award (Dr S K Manna).

From the details of the work undertaken by the Centre that are given elsewhere in the Report, I give below a few highlights, not necessarily exhaustive. The DNA fingerprinting services were provided to a large number of states in the country, and two prominent cases that were handled by the Centre during this period were the identification of skeletal remains of children and women recovered from Nithari village in Noida, Uttar Pradesh, and that of victims of the Samjautha Express train blast case. Diagnostics services comprising biochemical, cytogenetic, and molecular investigations for genetic disorders were offered to about 3300 patients this year.

Following the establishment in the CDFD of the Centre of Excellence on Genetics and Genomics of silkworms, the RNAi strategy has been successfully used to generate transgenic virus-resistant silkworms (*Bombyx mori*), and the immune transcriptome of both *Bombyx mori* and the Indian tasar silkworm (*Antheraea mylitta*) have been constructed and analyzed. In work on bacterial transcription, a novel mechanism of environmentally regulated promoter inhibition was identified in which a regulator protein, in the presence of its co-effector ligand, traps *Escherichia coli* RNA polymerase at the promoter in a dormant, reversible Rip Van Winkle complex. Further, the roles of different domains of the bacterial transcription termination factor Rho were established as was the mechanism of inhibition of Rho by the phage-encoded protein Psi. A genetic linkage map of basmati rice based on microsatellite markers was constructed and assays were developed and

employed for detection and quantification of adulteration in export batches of basmati rice.

In computational and structural biology, a new algorithm for prediction of global protein-protein interaction networks in bacteria was developed. Molecular dynamics simulations were undertaken to determine the mutational basis of congenital glaucoma, and the likely role of simple sequence repeats in pathogenicity of the plague bacterium *Yersinia pestis* was determined. As part of attempts to develop new methods for improved prediction and modeling of protein structures, it was found that a combination of solvent accessibility and secondary structural state information of dipeptides and amino acids provided the most suitable correlation for fold-specific discriminatory potential in proteins. The mechanistic differences between the chaperonins of *E. coli* and *Mycobacterium tuberculosis* were elucidated, and the crystal structure of the YefM antitoxin from the latter was determined. Bioinformatic analyses of the genomes of mycobacteria and the malaria parasite *Plasmodium* have been carried out in order to delineate operons and regulons in the former and to provide for improved annotation of genes in the latter. Many of these studies were undertaken with the aid of the powerful computational facilities available in the SUN-Centre of Excellence in CDFD.

In the area of pathogen biology, genomic analysis was undertaken to establish evolutionary relationships between individual isolates of both *M. tuberculosis* and *Helicobacter pylori*. Several genes putatively involved in iron and siderophore metabolism and transport in *M. tuberculosis* have been characterized. It has also been shown that, as part of the host response to infection, the B7.1 and B7.2 co-stimulatory molecules on macrophages play a crucial role in activation of macrophage effector responses through the mediation of the NF-kappa B family of transcription factors, and that free radicals can modulate T cell responses by modulation of the IL-12 signaling pathway. Mechanisms of nucleo-cytoplasmic shuttling of protein complexes in HIV-infected cells were investigated, and a new research group has been set up for studies on the pathogenic yeast *Candida glabrata*.

In the tumor biology field, GATA6 has for the first time been identified as a transcription factor that is implicated in pancreatic cancer, based on studies of gene copy number alterations and genome-wide expression profiling in biopsy samples. Differences in the frequency of Wnt pathway activation were observed between old and young patients with colorectal cancer. The plant-derived glycoside oleandrin was shown to induce apoptosis in tumor cells but not normal cells, and the mechanism of oleandrin action was characterized. Studies on nuclear transport of Ras-associated tumor suppressor proteins demonstrated for the first time that arginine-rich nuclear targeting signals in different proteins exhibit different nuclear transport receptor binding specificities to mediate the transport of these proteins to distinct subcellular compartments. In exploring the role of changes in genome epigenetic status in cancer, DNMT3L was identified as a candidate gene which shows promoter DNA methylation changes during

cancer development. In a related study, a putative imprinting control region for the *Neuronatin* gene in mouse was identified by biochemical experiments.

In all of the work described in this Report, I must sincerely acknowledge the contributions of and co-operation from my colleagues in the scientific, technical, and administrative cadres as well as from students and staff working in various projects at the Centre. The CDFD has been functioning under several constraints, including those related to operating from two campuses and severe space limitation for lab operations, and its achievements described here are all the more creditable for this reason. We have also benefitted immensely during the year from the advice, support, and encouragement from the officers of the Department of Biotechnology, and members of the Governing Council and the Research Area Panels – Scientific Advisory Committee of CDFD. We will continue to strive to greater heights in both our research and service activities in the years ahead.

March 31, 2007

**Dr. J Gowrishankar**

**MANDATE**



## **Mandate**

**The objectives for which the Centre for DNA Fingerprinting and Diagnostics (CDFD) was established as enumerated in the EFC documents are:**

1. To establish DNA diagnostic methods for detecting genetic disorders and to develop probes for such detection;
2. To use DNA fingerprinting techniques for the authentication of plant and animal cell material, cell lines and to develop new probes where necessary for such purposes;
3. To provide training in DNA fingerprinting techniques and offer consultancy services to medical institutions, public health agencies and industry in the country;
4. To undertake basic, applied and development R & D work;
5. To collaborate with foreign research institutions and laboratories and other international organizations; and establish affiliation with recognized universities and institutions;
6. To acquire or transfer technical know-how from/to entrepreneurs & industries and, to register patents, designs & technical know-how in the interest of the Centre;
7. To carryout DNA profile and related analysis in civil cases like paternity disputes, immigration, and exchange of new-borns in hospitals, for various agencies including private parties, on appropriate payment;
8. To provide DNA fingerprinting and related analysis and facilities to crime investigation agencies;
9. To assist police personnel, forensic scientists, lawyers and the judiciary in understanding the evidential value of the DNA profile analysis and related techniques in crime investigation and family matters.

# RESEARCH

*DNA FINGERPRINTING AND DIAGNOSTICS SERVICES*



***DNA FINGERPRINTING SERVICES***

<b>Service Co-ordinator</b>	J Nagaraju	Staff Scientist
<b>Other Members</b>	N Madhusudan Reddy	Staff Scientist
	Varsha	Staff Scientist
	SPR Prasad	Senior Technical Examiner
	V N Sailaja	Technical Officer Gr.I
	Ch V Goud	Technical Examiner
	D S Negi	Technical Examiner
	B Ramesh Babu	Technical Assistant
	Md Mahfooz Alam	Junior Assistant-I
	G Rajalingam	Attendant

**Objectives:**

- ◆ To provide DNA fingerprinting services in cases forwarded by law-enforcing agencies and judiciary of State and Federal Governments, relating to murder, rape, paternity, maternity, immigration, child-swapping, body identification, wildlife identity, seed purity testing, strain identification, etc;
- ◆ To develop human resources skilled in DNA fingerprinting, to cater to the needs of State and Federal Government agencies;
- ◆ To impart periodical training to manpower involved in DNA fingerprinting sponsored by State and Federal Government agencies;
- ◆ To provide advisory services to State and Federal Government agencies in establishing DNA fingerprinting facility; and
- ◆ To create DNA marker databases of different caste populations of India.

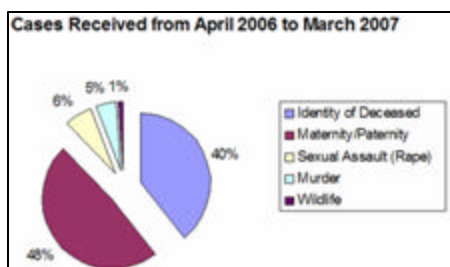
**Details of progress made in the current reporting year (April 1, 2006-March 31, 2007)**

I. Cases received for DNA fingerprinting examination

Paternity/maternity disputes	: 83
Identification of deceased	: 68
Sexual assault	: 11
Murder	: 08
Wildlife poaching	: 02
<b>Total</b>	<b>: 172</b>

**I. Summary of the state-wise breakup of DNA Fingerprinting cases**

State	Deceased	Paternity	Rape	Murder	Wildlife	Total
ANDHRA PRADESH	1	6		1		8
ASSAM		1				1
BIHAR	1					1
CHANDIGARH	1					1
CHATTISGARH	3	3	1		1	8
DELHI	5		1	1		7
GOA	1	2				3
HARYANA	3	5		1		9
HIMACHAL PRADESH	1					1
JAMMU & KASHMIR		1				1
JHARKHAND		2				2
KARNATAKA	10	25	1	2	1	39
KERALA	1	6				7
MADHYA PRADESH	6	7	3	1		17
MAHARASHTRA	4	7	1	1		13
NEPAL	1					1
ORISSA		1				1
PONDICHERRY		3				3
PUNJAB	4	4	1			9
RAJASTHAN	21	6	3	1		31
TAMILNADU		1				1
TRIPURA		1				1
UTTAR PRADESH	3	1				4
UTTARANCHAL	1					1
WEST BENGAL	1	1				2
<b>TOTAL</b>	<b>68</b>	<b>83</b>	<b>11</b>	<b>8</b>	<b>2</b>	<b>172</b>



## II. A few prominent cases

- 1) Identification and individualization (by DNA profiling) of skeletal remains recovered from Nithari village, Noida, Uttar Pradesh, forwarded by Central Bureau of Investigation, New Delhi.
- 2) Samjoutha Express Train Blast Case forwarded by Railway Police, Haryana.

### III. Deposition of evidence in Hon'ble Courts

During the reporting period, the DNA experts defended their reports in about 24 cases in various Hon'ble Courts throughout the Country.

### IV. Training provided

Training on DNA fingerprinting examination has been provided to the scientists from:

- 1) The State Forensic Science Laboratory, Sagar, Madhya Pradesh
- 2) The State Forensic Science Laboratory, Jaipur, Rajasthan
- 3) Rajiv Gandhi Centre for Biotechnology, Tiruvananthapuram, Kerala

### V. Lectures delivered

Frequent lectures have been delivered for the benefit of the following law enforcing officers:

- 1) Trainee investigating officers from Central Detective Training School, Ramanthapur, Hyderabad
- 2) IPS Officers from Sardar Vallabhabhai Patel National Police Academy, Hyderabad
- 3) International Police Officers from 22 different countries, coordinated by National Crime Records Bureau (NCRB), New Delhi.

### VI. Workshop conducted

The Press Institute of India, New Delhi in collaboration with DBT and CDFD organized a workshop on "Forensic Science and Biotechnology for Journalists" during 30-31<sup>st</sup> March, 2007 at CDFD.

## DIAGNOSTICS SERVICES

<b>Service Co-ordinator</b>	A Radha Rama Devi	Consultant	
<b>Other Members</b>	R Angalena	Technical Officer 1	
	S M Naushad	Technical Officer 1	
	Usha Rani Dutta	Technical Assistant (On study leave)	
	Jamal Md Nurul Jain	Technical Assistant	
	Ram Prakash Singh	Tradesman	
	C Krishna Prasad	Tradesman	
	R Sudheer Kumar	Tradesman	
	V Murali Mohan	Attendant	
	G Srinivas	Attendant	
	P Rajitha	Project Associate	
	<b>Collaborators</b>	Gayatri Ramakrishna	CDFD, Hyderabad
		Muralidharan Bhashyam	CDFD, Hyderabad
		Nagarajaram	CDFD, Hyderabad
Usha Naik		Niloufer Hospital, Hyderabad	
Suma Prasad		Sridevi Nursing Home, Hyderabad	

### Objectives:

1. Providing services for diagnosis, counseling and management of genetic disorders
2. Early intervention programs
3. Research on orphan diseases including multi-locus investigations to study complex multi-factorial disorders
4. Providing social awareness in early detection and prevention of genetic disorders
5. Training the medical and paramedical students in genetic testing

### Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)

#### I. Services provided

Mode of investigation	Total cases	Positives
<i>Biochemical</i>		
Proband	2241	284
High risk pregnancy screening	165	10
Prenatal Lysosomal enzymes	2	0
<i>Cytogenetic</i>		
Proband	810	60
Prenatal	70	0
<i>Molecular</i>		
Proband	456	108
Prenatal	15	7

**Disease-wise segregation of positive cases**

Disease	Number of positives
<b>Biochemical</b>	
<i>Aminoacid disorders (N=655)</i>	<b>121</b>
Phenylketonuria	1
Hyper phenylalaninemia	2
Maple syrup urine disease	3
Leigh's	1
Hyperalaninemia	17
Hyperglycinemia	17
Non Ketotic Hyperglycinemia	2
Citrullinemia	1
Hyperornithinemia	3
Tyrosinemia	6
Leucinemia	1
Hyper Lysinemia	3
Glutamic acidemia	62
Homocysteinemia (N=615)	<b>133</b>
Mild (15-30)	90
Moderate (30-100)	39
Severe (>100)	4
<i>Lysosomal storage disorders (N=64)</i>	24
Metachromatic Leukodystrophy	2
Gauchers	4
GMI-Gangliosidosis	3
Mucopolysaccharidosis (N=8)	MPS Type 1:2
Pompe's	4
Tay Sach's	2
Sandhoff	6
I-Cell disease	1
<i>Newborn screening (N=185)</i>	1 (Congenital hypothyroidism)
Galactosemia	2
Biotinidase deficiency (N=110)	3
<i>High risk pregnancy screening</i>	
Triple Marker screening (N=113)	6 (High risk for Down's)
First Trimester screening (N=52)	2 (High risk for Down's)

Disease	Number of positives
<b>Cytogenetic</b>	<b>60</b>
Downs syndrome	35
Turner	10
Sex Reversals	6
Edwards	1
Marker chromosomes	2
Triplex	2
Del (8) p21	1
Dup 1q	1
Inv Y	1
Fanconi	1
<i>Prenatal (N=70)</i>	0
<b>Molecular</b>	<b>108</b>
Duchenne muscular dystrophy	11
Spinal muscular atrophy	6
Fragile X syndrome	2
Spino cerebellar ataxia-2	1
Spino cerebellar ataxia-7	1
Friedrich's ataxia	3
Huntington disease	5
Connexin	1
SRY	1
Thalassemia	26 (Hetero:16, Compound hetero:2, Homo:8)
MTHFR (N=148)	26 (Hetero:21 and Hom $\alpha$ : 5)
Cystic Fibrosis (N=42)	17 (Hetero:6, Compound hetero:4, Homo:7)
MTB PCR	4
Factor V	4 (Hetero:4)
<i>Prenatal</i>	7
Thalassemia	2 homo and 2 hetero
Spinal muscular atrophy	2
Duchenne Muscular Dystrophy	1

During the current reporting year, 2241 cases were referred for biochemical evaluation and 12.7% cases were confirmed to have inborn errors of metabolism. Hyperhomocysteinemia was the most prevalent inborn error (21.6% in referred cases) followed by glutamic academia, hyperalaninemia and hyperglycinemia. Early diagnosis, prompt clinical intervention and management helped the families of the affected children. Among the lysosomal storage disorders, mucopolysaccharidosis was the most prevalent followed by Sandhoff, Gaucher and Pompe's. There has been an increased number of referrals for high risk pregnancy screening due to awareness created both in physicians and public. The screen positives were confirmed to be normal by karyotyping despite high risk for Down's. About 810 cases were referred for cytogenetic evaluation and 7.4% were confirmed to have chromosomal anomalies. Downs syndrome was the most prevalent followed by Turner's syndrome. Seventy high-risk families opted for prenatal diagnosis and all were normal. 456 cases were referred for molecular diagnosis and 23.7% were positive, the most important



being patients with thalassemia, cystic fibrosis and Duchenne muscular dystrophy in the order of prevalence. Out of 15 cases referred for prenatal diagnosis, seven were positive. Genetic testing at this tertiary testing center will be thus useful in the management and prevention of burden of genetic diseases in the population.

## **II. Research**

Extodermal dysplasia : In the current year molecular analysis was carried out on one of the orphan disorders, ectodermal dysplasia, by sequencing of Ectodysplasin-A (EDA-1) and Ectodysplasin-A receptor (EDAR) genes (in collaboration with Molecular Oncology group). Out of four cases examined, two had novel disease-causing mutations in the EDA-1 gene (resulting in X-linked hypohydrotic ectodermal dysplasia (HED)), one had a novel mutation in the EDAR gene and the fourth had the G382S mutation in the EDAR gene (both resulting in autosomal recessive HED). Pedigree analysis correlated well with the genotype in terms of mode of inheritance. Molecular analysis of Maple syrup urine disease (MSUD) was initiated by sequencing branched-chain alpha-keto acid dehydrogenase complex (BCKD) E3 subunit.

### **Assessment of paternal risk factors in recurrent pregnancy loss**

The current study was undertaken to investigate whether paternal risk factors for recurrent pregnancy loss are associated with DNA damage and chromatin abnormalities. Two different cell populations were detected in the semen sample of cases and in normal controls using flow cytometry. The results showed increased percentage of sperm with immature chromatin i.e.  $30.21 \pm 13.49$  in the patients compared to  $19.18 \pm 17.19$  in controls ( $P < 0.02$ ). The mean comet tail length in cases was  $64.99 \pm 6.98 \mu\text{m}$  and in controls was  $54.11 \pm 7.34 \mu\text{m}$  ( $P < 0.001$ ). Both FACS analysis and comet assay indicate significant DNA damage in male partners with recurrent pregnancy loss.

### **Prenatal detection of deletion-duplication of chromosome 3 arising from meiotic recombination of familial pericentric inversion**

A clinical and cytogenetic finding of prenatally diagnosed recombinant chromosome 3 arising from a paternal pericentric inversion was studied in collaboration with the Royal Children's Hospital, Melbourne, Victoria, Australia. The recombinant chromosome 3 leading to deletion of 3p and duplication of 3q is very rare; there are six postnatal reports in the literature and there are no reports of prenatal diagnosis. Amniocentesis was performed for abnormal ultrasound findings viz. ventriculomegaly, lumbosacral meningocele and lemon shaped head. The karyotype from cultured amniocytes demonstrated a karyotype of 46,XY, add(3)(p25). Parental karyotype revealed a pericentric inversion of chromosome 3 in the father with breakpoints p25 and q21 [46,XY, inv(3)(p25q21)]. The abnormal chromosome found in amniotic fluid cells was a recombinant chromosome 3 arising from paternal meiotic

recombination between normal and inverted 3. Hence, the karyotype of the fetus was interpreted as 46,XY, rec(3)dup(3q)inv(3)(p25q21)pat. The deletion 3p and duplication 3q was retrospectively confirmed by quantitative multiplex ligation-dependent probe amplification (MLPA) assay for subtelomeres. The pregnancy was terminated soon after amniocentesis. The aborted fetus was confirmed to have abnormalities that were detected through ultrasonography and other dysmorphic features superimposed to that of deletion-duplication syndrome of chromosome 3. This is the first report of prenatal diagnosis of deletion-duplication syndrome of chromosome 3. The prenatal findings of distorted head shape, meningocele and ventriculomegaly may serve as markers for deletion-duplication syndrome of chromosome 3. The ultrasound detection of neural tube defects, especially in association with additional markers might be a strong diagnostic clue for fetal chromosomal aneuploidy.

#### **National database on prevalent genetic disorders in India: Development, curation and services**

In collaboration with Bioinformatics division, a national database on prevalent genetic disorders in India is being developed. This will be helpful in establishing the nation-wide incidence of these genetic disorders. A module has been developed that can be incorporated by the genetic testing laboratories in the country for the easy transfer of data from other centers.

#### **External quality control for six parameters**

Since 2003, the Diagnostic s lab was enrolled with QCQAP program with CDC, Atlanta, USA for the newborn screening program and for the fourth consecutive year 100% satisfactory results were obtained for all the six parameters, namely, neonatal TSH, 17-hydroxy progesterone, immuno reactive trypsinogen, total galactose, biotinidase, and galactose-1-phosphate uridyl transferase. This year we have accredited TLC based quantitation of phenylalanine, leucine, methionine, tyrosine, valine and citrulline. Based on this, ICMR has identified CDFD as the quality control center in the multi-center project on newborn screening in the country

#### **Publications**

- 1 Devi AR, Gopikrishna M, Ratheesh R, Savithri G, Swarnalata G and Bashyam MD (2006) Farber lipogranulomatosis-clinical and molecular genetic analysis reveals a novel mutation in an Indian family. *Journal of Human Genetics*51:811-814.
- 2 Naushad SM, Jamal NJ, Angalena R, Prasad CK, and Devi AR (2007) Hyperhomocysteinemia and the compound heterozygous state for methylene tetrahydrofolate reductase are independent risk factors for deep vein thrombosis among South Indians. *Blood Coagulation & Fibrinolysis* 18:113-117.

**CDFD RESEARCH**



**Laboratory of Molecular Genetics**  
**Centre of Excellence on Genetics and Genomics of Silkmoths**  
**and**  
**Basmati rice genetics and genomics**

<b>Principal Investigator</b>	J Nagaraju	Staff Scientist
<b>Ph D student</b>	N Mrinal	Senior Research Fellow
	V L N Reddy	Senior Research Fellow
	Archana Gandhe	Senior Research Fellow
	Sunil Archak	Senior Research Fellow
	Jayendra Shukla	Senior Research Fellow
	K P Arunkumar	Senior Research Fellow
	Jyoti Singh	Junior Research Fellow
<b>Other members</b>	K Sriramana	Post doctoral fellow
	V Satish	Research Associate
	Edupalli V Subbaiah	Research Associate
	G Janardhan	Project Assistant
	S H John	Project Assistant
	P Sravan Kumar	Project Assistant
	G Janardhan	Project Assistant
	M Eshwar	Project Assistant

**Project 1: Development of RNAi-based baculovirus resistant transgenic silkmoths**

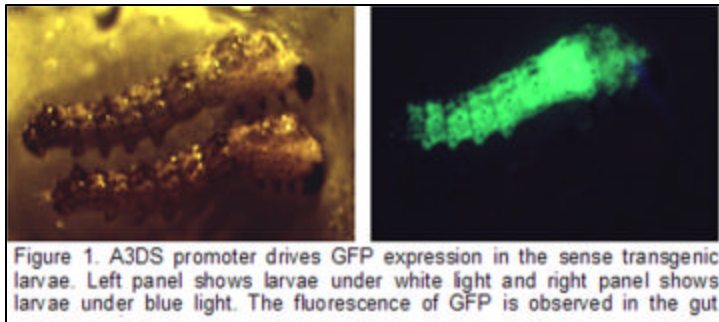
**Objectives:**

To generate transgenic silkworms resistant to *Bombyx mori* nucleopolyhedrosis virus (BmNPV) using RNAi strategy

**Summary of the work done until the beginning of this reporting year**

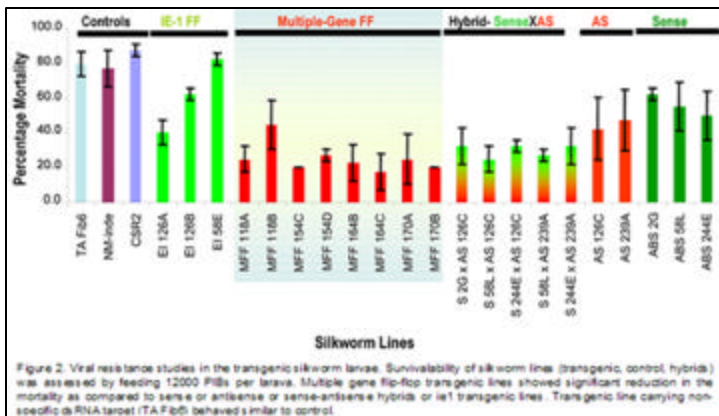
After the molecular analysis confirmed the transgenicity of the two silkworm lines (126A and 126B) for baculoviral ie-1 gene, the BmNPV infection assays were carried out to evaluate these transgenic lines for their ability to withstand viral infection. Reduction in the mortality of infected larvae was observed to the tune of 40% in per os infected larvae as compared to non-transgenic control lines, which succumbed to viral infection. Research was also initiated to construct vectors that carry dsRNA encoding fragments for multiple essential viral genes. The transgenic lines were developed using these vectors and the ability of the transgenic lines to combat viral infection was tested using viral infection assay.

**Details of progress made in the current reporting year (April 1, 2006 – 31, March 2007):**



After confirming the transgenicity by reporter assays (Fig. 1), viral infection assays were performed to ascertain the efficiency of transgenic lines that contained dsRNA encoding fragments for multiple viral genes in

combating viral infection. Additional assays were performed to validate the result of multigenic transgenic on virus proliferation by carrying out viral infection assays at various developmental stages. Infection was achieved either by pupal injection of free virus or by per os infection of 3<sup>rd</sup> or 4<sup>th</sup> instar larvae. The latter is a natural way of baculovirus infection of silkworm larvae. Twenty each of transgenic and control pupae were injected with a viral dose of 1200 budded virus (BV)/larva. The survival was measured based on the number of moths emerged from infected pupae. Proliferation of virus was evident in control lines where GFP



expressed by the virus was all pervasive, but not in transgenic lines. When fourth instar larvae were infected per os with 12000 PIBs per larva, almost all control larvae succumbed to infection as compared to ~ 25% mortality in multigenic flip-flop transgenic lines (Fig.2).

The transgenic line carrying non-specific dsRNA and water injected silkworms recorded ~ 90% mortality. Confirmation of viral induced death was carried out by symptomatic observation as well as presence of polyhedral inclusion bodies. Viral proliferation was also assayed by the western blot analysis using anti-GP64 antibody. Multi-flip-flop transgenic lines performed better than other lines (sense or antisense lines, or hybrids or ie-1FF lines) suggesting that multigene targeting of baculovirus essential genes is an efficient way of incorporating baculovirus resistance in silkworms.

### Details of technologies developed and transferred:

The transgenic lines are being transferred to APSSRDI, Hindupur and Central Silk Board under MTA for large scale controlled trials, after necessary clearance from RCGM and GEAC.

### Project 2: Comparative and functional genomics of silkmoths

#### Objectives:

To generate microsatellite and EST data, construction of genetic maps, and comparative and functional genomic analysis of genes involved in silk protein production, immune response, sex-specific functions as well as microRNA and their targets.

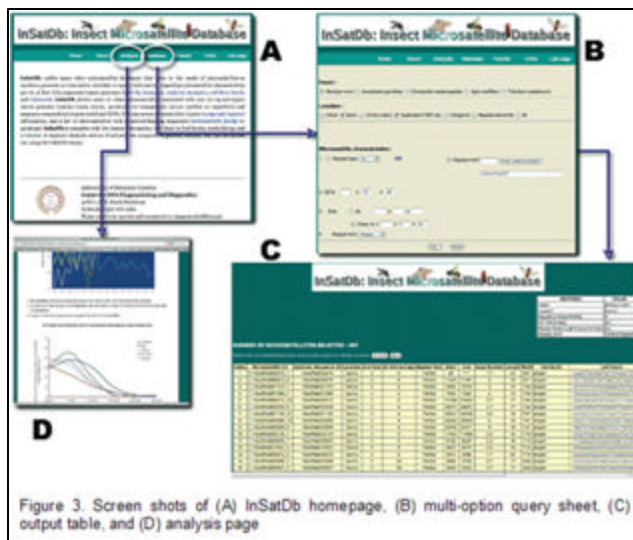
#### Summary of the work done until the beginning of this reporting year

Silkworm microsatellite database, SilkSatDb was developed. Framework linkage map and a Z-chromosome map of *Bombyx mori* were generated.

#### Details of progress made in the current reporting year (April 1, 2006 – 31, March 2007):

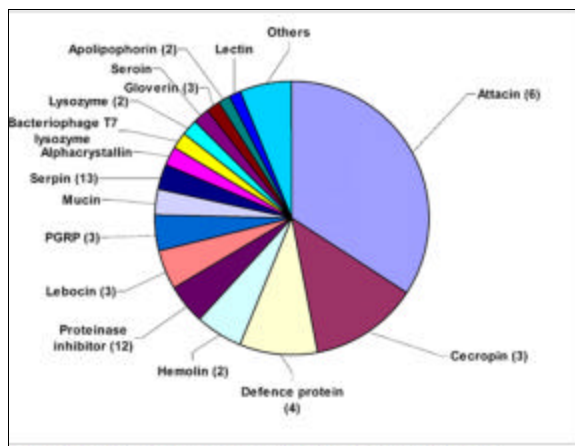
Consequent to the availability of whole genome sequence of many insects (Silkworm,

Mosquito, Honey bee, *Drosophila*, *Tribolium*), microsatellite loci from these insects were extracted computationally and annotated for comparative analysis. The information has been made available to the researchers as an open access database InSatDb, hosted on CDFD server (Fig. 3). Microsatellite markers developed and experimentally confirmed at the Centre were included in generating a high density linkage

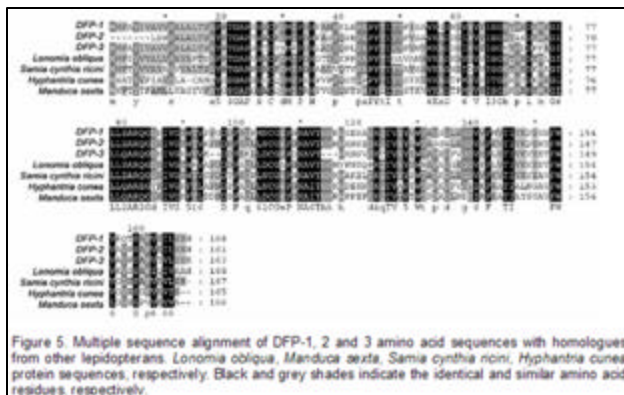


map by the international consortium. Many of these microsatellite loci were also tested for the flanking region similarities among different insect species. The results show that in lepidopterans, a large number of microsatellites are associated with transposable elements

and the high mutation rate in the regions may lead to high likelihood of null alleles. Immune transcriptome of *Antheraea mylitta* (Indian tasar silkmoth) was constructed and analysed with a view to unravel the potential immune-related genes and



pathways. Immune challenged tissue produced 719 unique ESTs from a total of 1412 sequences (Fig. 4). Four novel defence proteins (DFPs) were identified and annotated by PSI-BLAST. Three showed similarity to extracellular matrix proteins from vertebrates implicated in innate immunity (Fig. 5), while the fourth was similar to, yet distinct from, the anti-microbial protein, Cecropin. Several unknown proteins were also identified, some of which were predicted to have immune-related functions.



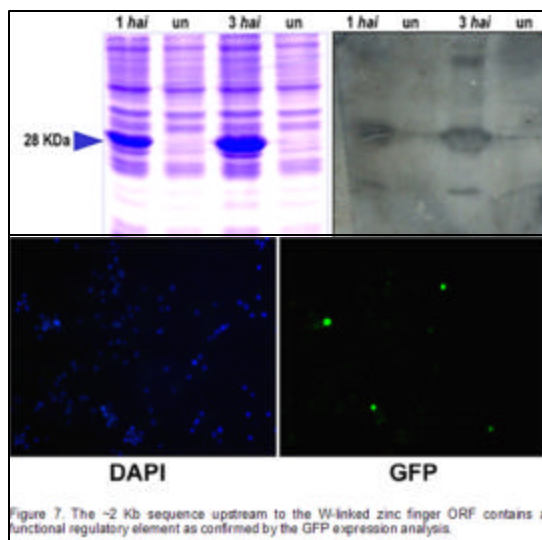
### Project 3: Sex determination mechanism and genetic sexing of *Bombyx mori*

#### Objectives:

To model sex determination mechanism in the female-heterogametic system of *Bombyx mori*

#### Summary of the work done until the beginning of this reporting year:

A female specific gene coding for zinc-finger motif containing protein was putatively placed as an upstream regulator of the sex determination pathway. One W-specific and two autosomal copies of this specific gene were discovered and the gene structure was deduced based on RACE and bioinformatic approaches.



#### Details of progress in the current reporting year (April 1, 2006 – 31, March 2007):

A W-linked C-x8-C-x5-C-x3 Zinc-finger domain-containing gene was identified through differential display of total RNA extracted from 36hrs old female and male embryos. The ubiquitous female-specific expression of this gene in all the tissues starting from early embryonic development and in all developmental stages suggests its possible role as the primary sex-determining signal.

Also, two putative genes containing the C-x8-C-x5-C-x3 domain were mapped on to the 25<sup>th</sup> chromosome. The W-linked CCCH-type zinc finger gene is being functionally characterized. The cDNA of the W-linked zinc finger is cloned and expressed in *E. coli* and the size of the protein was confirmed to be ~28 Kda, inclusive of the 6x-His tag (Fig. 6). The ~2 Kb stretch upstream to the W-linked zinc finger encoding gene was cloned, sequenced and analyzed for the presence of a promoter and two



regions were short listed as putative promoters. The expression analysis with 2 Kb upstream sequence cassette conferred GFP expression in silkworm cells (Fig. 7).

#### **Project 4: Molecular characterisation of immune response proteins of silkmoths**

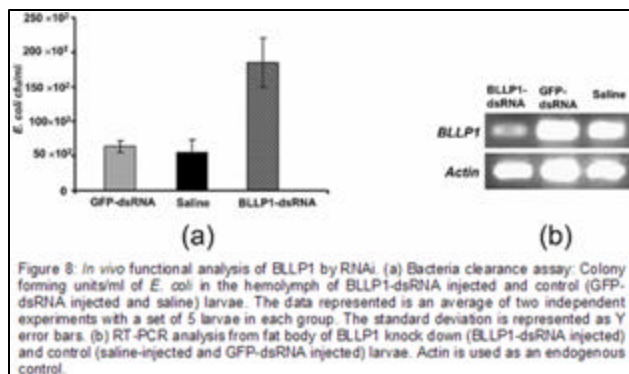
##### **Objectives**

Genetic, biochemical and functional characterisation of lepidopteran immune response pathway.

##### **Summary of the work done until the beginning of this reporting year:**

Immune transcriptome analysis of *B. mori* and wild silkworm *Antheraea mylitta*, led to identification of a new class of anti-microbial proteins. These proteins were designated as lysozyme-like proteins (LLPs) owing to their partial similarity with lysozymes. However, lack of characteristic catalytic amino acid residues essential for muramidase activity in LLPs puts them functionally apart from classical lysozymes. Two LLPs, one from *B. mori* (BLLP1) and the other from *A. mylitta* (ALLP1) expressed in a recombinant system, exhibited a broad-spectrum anti-bacterial action. Further investigation of the antibacterial mechanism revealed that BLLP1 is bacteriostatic rather than bactericidal against *E. coli* and *Micrococcus luteus*.

##### **Details of progress made in the current reporting year (April 1, 2006 – 31, March 2007):**



Previous reports have shown that immunity is mediated by secretion in the hemolymph of recognition proteins and effector anti-microbial proteins upon immune challenge. Substantial increase in hemolymph bacterial load was observed in *B. mori* upon RNA interference mediated in vivo knockdown of BLLP1. Hence

clearance of *E. coli* from hemolymph in dsRNA mediated BLLP1 knockdown insects as compared to larvae injected with non-specific dsRNA (GFP-dsRNA) or saline prior to *E. coli* challenge was studied. Figure 8a indicates that *E. coli* load was 3-4 fold higher in BLLP1 knockdown larvae as compared to the control larvae. The BLLP1-dsRNA induced knockdown of BLLP1 was confirmed at molecular level by RT-PCR and is shown in Figure 8b. This indicates the involvement of BLLP1 in immunity and point towards its occurrence in hemolymph as predicted from the signal peptide studies. Anti-bacterial mechanism of this protein was demonstrated to be due to peptidoglycan binding rather than via peptidoglycan hydrolysis or membrane permeabilization as observed with lysozymes and most other anti-microbial peptides. This is the first report on identification and functional analysis of novel, non-catalytic lysozyme-like family of anti-bacterial proteins that are functionally distinct from

classical lysozymes. The present analysis holds promise for functional annotation of similar proteins from other organisms.

## Basmati rice genetics and genomics

### Project 1: QTL mapping in Basmati rice

#### Objectives:

To map important Basmati traits like aroma, kernel length, kernel breadth and kernel elongation

#### Summary of the work done until the beginning of this reporting year:

In order to construct Basmati rice linkage map, a traditional Basmati variety Basmati370 and a semi dwarf variety Jaya were chosen as parental strains. The parental strains were screened with 493 rice microsatellite loci, of which 257 loci showed polymorphism. F<sub>1</sub> hybrid was confirmed by using two microsatellite markers RM263 and RM339.

#### Details of progress in the current reporting year (April 1, 2006– 31, March 2007):



F<sub>1</sub> progeny was advanced to F<sub>2</sub> to be used as mapping population (see Fig. 9 for grain characteristics and a segregating marker). 150 F<sub>2</sub> plants were randomly selected from 6000 plants. Data on quantitative and qualitative traits of parents, F<sub>1</sub> and F<sub>2</sub>

progeny were recorded. All the quantitative and qualitative traits (plant height, panicle number, panicle length, number of filled and chaffy grains, per panicle, spikelet no./panicle, spikelet fertility, 1000 seed weight and yield/plant and kernel length, kernel breadth, alkali spreading value, amylose content (%), aroma, grain chalkiness, kernel length after cooking, kernel elongation ratio) showed normal distribution except spikelet fertility. Transgressive segregants were observed in all traits. Screening of mapping population with polymorphic microsatellite markers is being carried out.

### Project 2: Detection and quantification of adulteration in Basmati samples

**Objectives:** To develop microsatellite-based assay to detect and quantify adulteration in Basmati samples

#### Summary of the work done until the beginning of this reporting year:

Basmati rice varieties are known for their extra long grains, and pleasant and distinct aroma. Traditional Basmati rice cultivars confined to Indo-Gangetic regions of the Indian sub-continent, are often reported to be adulterated in the export market. Surprisingly, there is no commercial scale technology to reliably detect adulteration. Unlike multiplex PCR, real-time PCR and microarray based methodologies employed for the traceability of genetically modified organisms in foods, detection of adulteration in Basmati is tricky as the adulterant is not really a foreign material. This genetic homogeneity of the adulteration renders technologies such as real-time PCR non-applicable necessitating employment of non-genic markers.

**Details of progress in the current reporting year (April 1, 2006– 31, March 2007):**

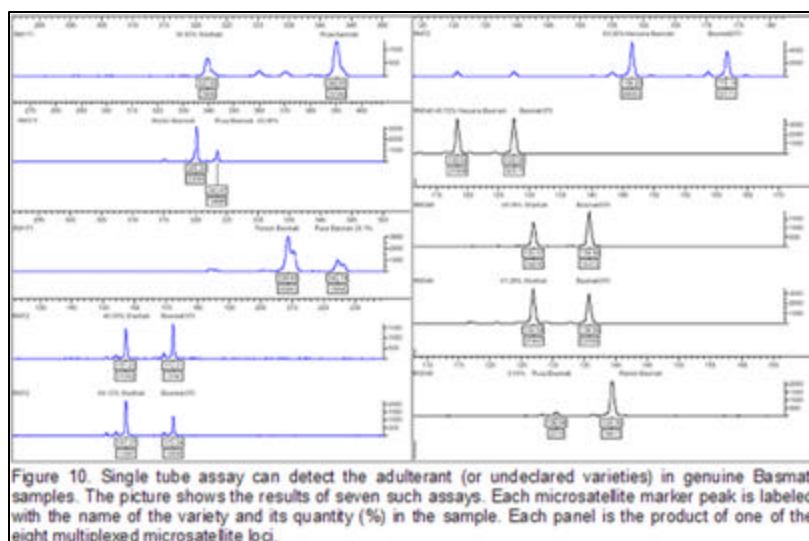


Figure 10. Single tube assay can detect the adulterant (or undeclared varieties) in genuine Basmati samples. The picture shows the results of seven such assays. Each microsatellite marker peak is labeled with the name of the variety and its quantity (%) in the sample. Each panel is the product of one of the eight multiplexed microsatellite loci.

A novel methodology founded on capillary electrophoresis based multiplex microsatellite marker assay for detection as well as quantification of adulteration in Basmati rice samples was developed. The single-tube assay multiplexes eight microsatellite loci

to generate variety-specific allele profiles that can detect adulteration from 1% upwards (Fig. 10). The protocol also incorporates a quantitative-competitive PCR based analysis for quantification of adulteration. Accuracy of quantification has been shown to be  $\pm 1.5\%$ .

**Publications**

- 1 Arunkumar KP and Nagaraju J (2006) Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. **PLoS ONE** 1: e1110.
- 2 Gandhe AS, Arunkumar KP, John SH and Nagaraju J (2006) Analysis of bacteria-challenged wild silkworm, *Antheraea mylitta* (Lepidoptera) transcriptome reveals potential immune genes. **BMC Genomics** 7:184.

- 3 Archak S, Meduri E, Kumar PS and Nagaraju J (2007) InSatDb: a microsatellite database of fully sequenced insect genomes. ***Nucleic Acids Research*** 35:D36 - D39.
- 4 Meglecz E, Anderson SJ, Bourguet D, Butcher R, Caldas A, CasseLundhagen A, d'Acier AC, Dawson DA, Faure N, Fauvelot C, Franck P, Harper G, Keyghobadi N, Kluetsch C, Muthulakshmi M, Nagaraju J, Patt A, Petenian F, Silvain JF and Wilcock HR (2007) Microsatellite flanking region similarities among different loci within insect species. ***Insect Molecular Biology*** 16:175-185.
- 5 Archak S, Reddy VLN and Nagaraju J (2007) High-throughput multiplex microsatellite marker assay for detection and quantification of adulteration in Basmati rice (*Oryza sativa*). ***Electrophoresis*** (InPress).
- 6 Gandhe AS, Gude J and Nagaraju J (2007) Immune upregulation of novel anti-bacterial proteins from silkmoths (Lepidoptera) that resemble lysozymes but lack muramidase activity. ***Insect Biochemistry and Molecular Biology*** (InPress).

*Laboratory of Genomics and Profiling Applications*

<b>Principal Investigator</b>	N Madhusudan Reddy	Staff Scientist
<i>Other Members</i>	<i>Anita Kumari</i>	<i>Project Assistant</i>
	<i>Deepika Dashyam</i>	<i>Project Assistant</i>
	<i>K Hanuma Kumar</i>	<i>Project Assistant</i>

**Objectives:**

- 1 Development, standardization and validation of DNA markers for genetic fidelity testing of tissue culture raised plants and phylogenetic studies
- 2 Referral Centre for detection of genetically modified foods employing DNA-based markers
- 3 Development of novel strategies/methodologies for enrichment of human DNA from mixtures of human and non-human DNAs for DNA profiling-based human identification
- 4 To study the human genetic diversity among various caste and tribal populations of India

**Project 1: Referral Centre for Genetic Fidelity Testing of Tissue Culture Raised Plants Employing DNA Markers.**

This is a new activity.

The National consultation group on certification of tissue culture raised plants sponsored by Department of Biotechnology has proposed that plant certification is essential for import/export material as per the rules/procedures laid down in the Plant Quarantine Order 2003, whose primary goal is to ensure production of virus-free plants of assured quality raised through tissue culture. The need for identification of more region-based centres, especially around the parts of the country and also at locations where there is a hub of tissue culture units, was recognized in view of the large volume of material to be tested, for which Department of Biotechnology has been notified as the Certification Agency under the National Certification System for Tissue Culture Raised Plants (NCSTCP).

In view of the above recommendations/proposals, a 'Referral Centre for the Genetic Fidelity Testing of Tissue Raised Culture Plants Employing DNA Markers' is established at CDFD and is proposed to undertake the true-to-type testing of important crop plants like banana, potato, sugarcane, black pepper and vanilla. The project entails providing services in the genetic fidelity testing of tissue culture raised micropropagules and development, standardization and validation of microsatellite or Simple Sequence Repeats (SSR) markers,

Inter-SSR (ISSR) markers and Amplified Fragment Length Polymorphism (AFLP) techniques to test true-to-typeness of above micropropagated plants. Further, it is proposed to conduct periodic proficiency tests to verify sensitivity, specificity, reproducibility and conformity of the test results.

**Project 2: Referral Centre for Detection of Genetically Modified Foods Employing DNA-based Markers .**

This is a new activity.

The advent of genetic engineering in agriculture and food production has impacted, not only on the environment and biodiversity, but also the human health. A thorough Biosafety assessment requires not only evaluation of environmental impacts of Genetically Modified Organisms (GMOs), but also assessment of the risks that genetically engineered foods pose for the health of the consumers. The need to monitor and verify the presence and the amount of GMOs in agricultural crops and in products derived thereof has generated a demand for analytical methods capable of detecting, identifying and quantifying either the DNA introduced or the proteins expressed in transgenic plants, because these components are considered as the fundamental constituents.

The methods presently employed by various laboratories worldwide for GMO detection and quantification differ in their reliability, robustness and reproducibility and have different levels of cost, complexity and speed. Also, there is no one universal method that is applicable to all circumstances. In order to obtain comparable results, and hence giving consumers and producer's confidence in the testing methods, an urgent need was felt for internationally validated methods, which could serve as reference methods. A 'Referral Centre for Detection of Genetically Modified (GM) Foods Employing DNA-based Markers' is established at CDFD. As part of this project, we propose to undertake the GM food detection services employing the protocols developed earlier at CDFD in cotton and rice (whole or crushed seeds). In the research and development component of the project, we plan to develop, standardize, validate and employ DNA-based markers for GM food detection for various crops/varieties containing specific transgenes as envisaged by the Joint Working Group of the Department of Biotechnology and Ministry of Health & Family Welfare.

Project 3: "Human epigenetics to the rescue of Human Identification Process: Enriching human DNA from DNA mixtures employing antibodies directed against 5-methylcytosine followed by whole genome amplification" (as part of Innovative Young Biotechnologist Award (IYBA) - 2006).

This is a new activity.

Mass fatality incidents like natural disasters (earthquakes, tsunamis, flooding and cyclones), accidental disasters (aircraft crashes, train crashes and derailments, and building fires) or intentioned terrorism (direct attacks on significant objectives and bombing of populated

areas, including suicide affects and deployments of chemical and biological weapons) results in a very large number of casualties. In such scenarios, forensic identification of victims is essential not only for humanitarian reasons but also for civil or criminal investigative needs. In such situations, DNA testing plays an important role in human remains identification in mass fatality incidents throughout the world. However, degradation and contamination of DNA extracted from bone and teeth samples with non-human DNA could make that process difficult.

Owing to its tropical environment, which results in robust microbial infestation, obtaining good quality human DNA for body identification by DNA profiling has not been very successful in India. Therefore newer methodologies are needed to circumvent these problems. In the present project, it is proposed to selectively enrich the human DNA from the mixture of microbial DNAs by employing antibodies directed against 5-methyl cytosine (5-mC), an epigenetic modification found in humans, to specifically 'pull down' human DNA sequences containing 5-mC, followed by whole genome amplification to amplify such DNA sequences. Subsequently, human short tandem repeat (STR)-containing fragments would be 'captured' and isolated employing biotinylated tetra-nucleotide repeat-containing oligos ("probe") and streptavidin-coated beads. The 'captured' DNAs would then be subjected to STR/mini-STR analysis to unambiguously identify the body parts/ skeletal remains of deceased persons.

**Project 4: "To study the genetic diversity among various caste and tribal populations of India" (as part of Max Planck India Visiting Scientist Fellowship).**

This is a new activity.

The diverse caste and tribal populations in India offer a rich source of human biologic variation, whose study would aid in our understanding and knowledge of social stratification, human evolution and migration patterns. In India, various studies have been undertaken to assess the genetic diversity among various population groups. But thus far, no comprehensive comparison of both uni-and bi-parental genetic diversities has been made in Andhra Pradesh and Madhya Pradesh populations, employing a large set of DNA markers. It is proposed in the present project to study and catalog the human diversity in 6-10 different castes in the states of A.P. and M.P. The DNA polymorphism in these communities would be studied using mtDNA, autosomal and Y-chromosomal genetic markers. The data obtained from the studies will be compared with published data from other population groups and regions from rest of the country and the world to ascertain male and female migration rates during the evolution of native A.P. and M.P. populations. Such studies would shed light on male/female migration patterns and to address questions about patrilocality v/s matrilocality in the course of human pre-history and evolution.

Some of the specific hypotheses that we wish to test with the genetic data generated in this project are: (a). to understand as to what extent was the caste system an indigenous development vs. something that was brought in by a migration. Earlier attempts by various groups to address this question have been inconclusive. Therefore the present project proposes to address this by studying a large number of DNA markers developed in Prof. Mark Stoneking's lab at Leipzig, Germany. (b). another interesting hypothesis that is proposed to be tested in the present proposal is to determine to what extent was the spread of agriculture across India the result of demic vs. cultural diffusion. Previous studies had addressed this problem by characterizing only a limited number of samples. In order to better understand this, it is proposed to sample a statistically significant number of agricultural and hunting-gathering tribal populations from India and study the distribution of uni- and bi-parental genetic markers in them. The knowledge gained by the proposed human genetic variation studies in different populations of South and Central India would provide new insights in our understanding of human genetic variability.



**Laboratory of Fungal Pathogenesis**  
**Understanding the pathobiology of an opportunistic human pathogen, *Candida glabrata***

**Principal Investigator**      Rupinder Kaur      Staff  
scientist

*Candida spp.* are important opportunistic pathogens in humans, primarily causing mucosal infections which in the immuno-compromised host can breach the mucosal barrier leading to life threatening systemic infections. The major clinical risk factors that predispose patients to *Candida* infections are a severe underlying disease (AIDS or leukemia), impaired phagocytic function (granulocytopenia or neutropenia) and exogenous factors (such as pre-existing use of broad spectrum antibiotics, implantation of prosthetic material like catheter, IV drug use, trauma, abdominal surgery *etc.*).

Although some pre-disposition in the patients appears to be an important factor in most of the *Candida* infections reported to date, questions remain unanswered regarding the intrinsic properties that make commensal yeasts potential pathogens. Survival *in-vivo* is a complex multifactorial process requiring the co-ordination of several responses, a few being the ability to survive in the nutrient poor host environment, the ability to evade the host immune response and the ability to develop resistance to anti-fungal drugs. Our lab is trying to better understand the pathogenesis of a human opportunistic pathogen, *Candida glabrata*, by focusing on some of these responses.

**Project 1: Functional genomic analysis of *C. glabrata*-macrophage interaction .**

This is a new activity.

To study the molecular interactions of *C. glabrata* with macrophages which constitute the first line of host defense against this pathogenic yeast, we are carrying out a genomic analysis of host-pathogen interactions in *C. glabrata* using an *in-vitro* human macrophage model system *via* a modified version of signature tagged mutagenesis (STM) approach. This project involves screening of a *C. glabrata* Tn7 insertional mutant library (20,000 mutants representing ½ of *C. glabrata* genome) for altered intracellular survival profiles by a high throughput strategy whereby a pool of 96 mutants will be screened at once. This genetic screen will help us identify genes required for survival/replication of *C. glabrata* in human macrophages and will thus yield insights into the kind of nutritional and anti-microbial

stresses, *C. glabrata* encounters in the intracellular milieu of macrophages and the strategies, this yeast has evolved to counteract/evade the anti-microbial arsenal of macrophages.

### **Project 2: Innate resistance of *C. glabrata* to fluconazole**

This is a new activity.

A major clinical challenge in treating *C. glabrata* infections is the innate resistance of this yeast to fluconazole which is the most commonly used anti-fungal drug. One of the first step towards developing new combinatorial anti-fungal agents is a better understanding of *C. glabrata*'s response to fluconazole. To gain insights into the intrinsic resistance of *C. glabrata*, a project in the lab is focused on a genome-wide screen for mutants that lose viability in the presence of prolonged fluconazole stress. We hope that this screen will help us identify the novel mechanisms that *C. glabrata* has developed to survive the fluconazole exposure.

### **Project 3: Transcriptional regulation of a multi-gene yapsin family**

This is a new activity

Our recent transcriptional profiling analysis of macrophage internalized *C. glabrata* cells in an in-vitro system (murine macrophage like cell line J774A.1) shows an up-regulation of seven members of a multi-gene family that codes for a family of cell-surface associated, glycosylphosphatidylinositol (GPI)-linked aspartyl proteases (yapsins). A part of our research is centered around a comprehensive analysis of differential transcriptional regulation of this yapsin gene family in response to different environmental cues. Using a combination of molecular biology, genetics and biochemical approaches, we will identify and characterize the regulatory regions in yapsin gene promoters that are responsible for the differential expression of yapsin genes. The knowledge gained from this analysis will shed light into the mechanism that *C. glabrata* has evolved to regulate the expression and activity of its family of eleven extracellular proteases and how do these protease contribute to its virulence.

## *Laboratory of Virology*

### *Molecular Pathogenesis of lentiviruses*

<b>Principal Investigator</b>	S Mahalingam	Staff Scientist
<b>Ph D Students</b>	P R Kumar	Senior Research Fellow (till September 2006)
	P K Singhal	Senior Research Fellow
	M K Subba Rao	Senior Research Fellow
	Gita Kumari	Senior Research Fellow
	B Neelima	Junior Research Fellow
<b>Other Members</b>	Dr. P Srilakshmi	Post doctoral Fellow
	Dr.Y Padward	Post doctoral Fellow (from Feb 2007)
	Dr Janaki Ram	Post doctoral Fellow (DBT) (from March 2007)
	Ch Vijayakumar	Project Assistant
	S Chary	Project Assistant (from Feb 2007)

#### **Objectives:**

1. Understanding the mechanism of human immunodeficiency virus pathogenesis
2. Functional analysis of Nucleolar GTPases
3. Understanding the function of Ras effector proteins

#### **Summary of work done until the beginning of this reporting year**

Transport of the lentiviral genome into the nucleus required phosphorylation of components in the preintegration complex by virion-associated host cellular kinases. In this study, we showed that ERK-2/MAPK is associated with simian immunodeficiency virus (SIV) virions and regulated the nuclear transport of Vpx and virus replication in non-proliferating target cells by phosphorylating Vpx. Suppression of the virion-associated ERK-2 activity by MAPK pathway inhibitors impaired both Vpx nuclear import and viral infectivity without affecting virus particle maturation and release. In addition, mutation analysis indicated that the inactivation of Vpx phosphorylation precluded nuclear import and reduced virus replication in macrophage cultures, even when functional integrase and Gag matrix proteins implicated in viral preintegration complex nuclear import are present. In this study, we also showed that co-localization of Vpx with Gag precursor in the cytoplasm is a prerequisite for Vpx incorporation into virus particles. Substitution of hydrophobic Leu-74 and Ile-75 with serines in the helical domain abrogated Vpx nuclear import, and its incorporation into virus particles, despite its localization in the cytoplasm, suggested that the structural integrity of helical

domains is critical for Vpx functions. Taken together, these studies demonstrated that the host cell MAPK signal transduction pathway regulated an early step in SIV infection.

We reported that the fission yeast nucleoporin Nup124p is required for the nuclear import of retroviral HIV-1 Vpr. Nuclear import of HIV-1 Vpr was impaired in nup124 null mutant strains and cells became resistant to Vpr's cell-killing activity. On the basis of protein domain similarity, the human nucleoporin Nup153 was identified as a putative homolog of Nup124p. We demonstrate that in vitro-translated Nup124p and Nup153 coimmunoprecipitate HIV-1 Vpr. We provide evidence that both nucleoporins share a unique N-terminal domain, Nup124p(AA264-454) and Nup153(AA448-634) that is absolutely essential for Vpr nuclear translocation. Our results established a unique relationship between two analogous nucleoporins Nup124p and Nup153 wherein the function of a common domain in Vpr nuclear transport is conserved.

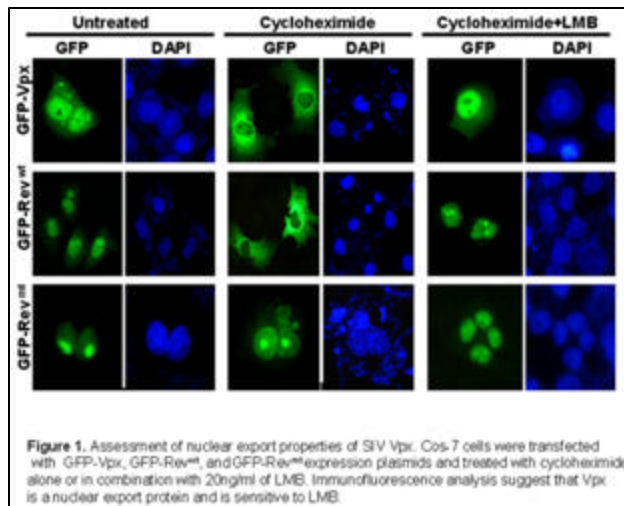
Grn1p from fission yeast and GNL3L from human cells, two putative GTPases from the novel HSR1\_MMR1 GTP-binding protein subfamily with circularly permuted G-motifs play a critical role in maintaining normal cell growth. Deletion of Grn1 resulted in a severe growth defect, a marked reduction in mature rRNA species with a concomitant accumulation of the 35S pre-rRNA transcript, and failure to export the ribosomal protein Rpl25a from the nucleolus. Deleting any of the Grn1p G-domain motifs resulted in a null phenotype and nuclear/nucleolar localization consistent with the lack of nucleolar export of preribosomes accompanied by a distortion of nucleolar structure. Heterologous expression of GNL3L in a Deltagr1 mutant restored processing of 35S pre-rRNA, nuclear export of Rpl25a and cell growth to wild-type levels. Genetic complementation in yeast and siRNA knockdown in HeLa cells confirmed the homologous proteins Grn1p and GNL3L are required for growth. Failure of two similar HSR1\_MMR1 putative nucleolar GTPases, Nucleostemin (NS), or the dose-dependent response of breast tumor autoantigen NGP-1, to rescue deltagr1 implied the highly specific roles of Grn1p or GNL3L in nucleolar events. Our analysis uncovered an important role for Grn1p/GNL3L within this unique group of nucleolar GTPases.

#### **Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)**

The following significant findings were made by the group in the current year.

**SIV Vpx shuttles between nucleus and cytoplasm.** GFP fusions containing different Vpx deletion fragments expressed in either HeLa or Cos-7 cells were found to localize differentially. This indicated the presence of NES in SIVsm Vpx. GFP-Vpx1 -63 fusion proteins were probably shuttling into and out of the nucleus and appeared cytoplasmic due to longer dwell time. We therefore studied the effect of a well-known nuclear export inhibitor, LMB,

and/or translational inhibitor, cycloheximide, on the nuclear accumulation of GFP-Vpx fusion proteins. LMB is known to block nuclear export due to a covalent modification at a cysteine residue in the central conserved domain of an export receptor, chromosomal region maintenance 1 (CRM-1). GFP-Vpx was localized in cytoplasm in the presence of cycloheximide alone and in nucleus with both cycloheximide and LMB (Fig. 1). This suggests



**Figure 1.** Assessment of nuclear export properties of SIV Vpx. Cos-7 cells were transfected with GFP-Vpx, GFP-Rev<sup>wt</sup>, and GFP-Rev<sup>mt</sup> expression plasmids and treated with cycloheximide alone or in combination with 20ng/ml of LMB. Immunofluorescence analysis suggest that Vpx is a nuclear export protein and is sensitive to LMB.

the presence of both nuclear import signal and nuclear export signal (NES) in Vpx. As a positive control, we examined the effect of the same doses of cycloheximide and LMB on shuttling of a GFP-HIV-1-Rev fusion protein and observed a similar pattern of localization. This assay was further validated with the observation of unchanged nuclear localization for the nuclear export-defective Rev, GFP-Rev<sup>mt</sup>, in the presence of cycloheximide

with or without LMB (Fig. 1). In summary, these data provide evidence that Vpx is a nucleocytoplasmic shuttling protein and that its export from nucleus is sensitive to LMB.

### **Nuclear export of SIVsm(Pbj1.9) Vpx can complement effector domain function of HIV-1 Rev.**

A transient transfection system with reporter plasmid pDM128 derived from the *env* region of HIV-1 was used to characterize the nuclear export of Vpx. The transcript produced by pDM128 harbors a single intron containing a CAT coding sequence which is excised when the RNA is spliced. Cells transfected with pDM128 alone expressed the spliced transcripts in the cytoplasm and yielded very low levels of CAT activity. In contrast, cotransfection with a functional HIV-1 Rev expression vector (pRev<sup>wt</sup>) permitted the unspliced transcripts to enter the cytoplasm, thus increasing the CAT activity. Our results indicate the low levels of CAT activity in Rev<sup>mt</sup>-transfected cells compared with Rev<sup>wt</sup>-transfected cells, suggesting that mutation in the Rev effector domain blocks the nuclear export of CAT mRNA (unspliced RNA) to the cytoplasm. A fusion protein was generated by fusing full-length Vpx to the carboxyl terminus of HIV-1 Rev<sup>mt</sup> to assay the activity of Vpx nuclear export signal in the above-mentioned system. Interestingly, Vpx could complement the defects of HIV-1 Rev effector domain and increase the CAT activity in the Rev<sup>mt</sup>-Vpx-transfected cell. These results provide evidence for the presence of a functional NES in Vpx and further suggest that Vpx NES is able to complement export activity of heterologous proteins.

**Signal sequence within residues 41 to 63 of Vpx is essential for nuclear export.** Deletion analysis of Vpx showed that amino acid residues 1 to 40 of Vpx were able to transport the heterologous protein GFP to the nucleus, whereas residues 1 to 63 localized it to the cytoplasm. It appears that GFP-Vpx1-63 shuttles into and out of the nucleus. To this end, GFP fusion constructs containing Vpx1-40 and Vpx1-63 were transfected into Cos-7 cells and treated with cycloheximide in the presence or absence of LMB. Our results indicate more nuclear accumulation of the GFP-Vpx1-63 protein upon cycloheximide and LMB treatment compared to the results seen with untreated as well as cycloheximide-treated cells. Identical doses of the indicated drugs did not alter the nuclear localization of GFP-Vpx1-40. These results suggest the possibility that export signal resides between residues 41 and 63 of Vpx.

Nuclear export activity of Vpx mutant protein was examined with Cos-7 cells in the presence of cycloheximide with or without LMB by indirect immunofluorescence using anti-Vpx monoclonal antibody. Cycloheximide (with or without LMB) did not alter the nuclear localization of Vpx protein containing a mutation at tryptophan 49 (W49S) alone or in combination with tryptophan residues at positions 53 and 56 (W49, 53, 56S). In addition, replacement of tryptophan residues abrogated the nuclear export activity of Vpx1-63 as well. These data provide evidence for the presence of nuclear export signal within residues 41 to 63 of Vpx. Surprisingly, a similar pattern of localization for mutant Vpx protein with tyrosine residues exchanged to alanine (Y66, 69, 71A) was observed. Cycloheximide treatment did not alter the nuclear localization of these mutant proteins. Collectively, these data suggest that tryptophan residues within domain 41 to 63 may play a critical role in export of Vpx from nucleus.

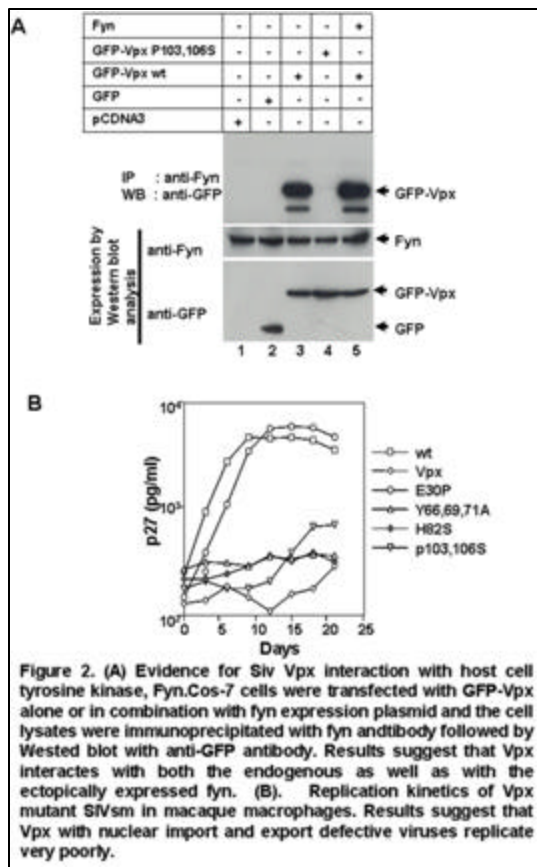
**Vpx interacts with Fyn SH3 domain through its C-terminal proline-rich domain.** Proline-rich motifs are known to interact with SH3 domains of Src-like tyrosine kinases as well as of proteins that are involved in signal transduction pathways. To understand the role of the C-terminal proline-rich domain (RPGP<sub>7</sub>GLA) in Vpx function, we first tested whether Vpx interacts with any host Src-like kinases. SH3 domains of Fyn, Src, Hck, and full-length Crk were used as an affinity matrix to identify the Vpx-interacting partner in an in vitro GST pulldown assay. Equal amounts of GST, GST-Fyn-SH3, GST-Src-SH3, GST-Hck-SH3, and GST-Crk bound with glutathione-Sepharose beads were incubated with in vitro-translated GFP or GFP-Vpx, and the bound proteins were analyzed by autoradiography. Vpx was found to specifically associate with Fyn SH3 domain. The specificity of the interaction between the Vpx and Fyn SH3 domains was further confirmed by competition experiments with SH3 and SH2 domain peptides. Cos-7 cell lysates expressing Vpx protein were incubated with glutathione-agarose-bound Fyn SH3/SH2 domains in the presence or absence of peptides

corresponding to the SH3 or SH2 domain of Fyn and probed with anti-Vpx monoclonal antibody. Vpx interaction was observed only with the SH3 domain and not with the SH2 domain. Furthermore, interaction between Vpx and SH3 domain was selectively blocked by SH3 domain peptide but not by SH2 peptide.

Cos-7 cell lysates expressing Vpx protein were incubated with glutathione-agarose-bound Fyn SH3/SH2 domains in the presence or absence of peptides corresponding to the SH3 or SH2 domain of Fyn and probed with anti-Vpx monoclonal antibody. Vpx interaction was observed only with the SH3 domain and not with the SH2 domain. Furthermore, interaction between Vpx and SH3 domain was selectively blocked by SH3 domain peptide but not by SH2 peptide.

Involvement of C-terminal proline-rich motif (RPGP<sub>7</sub>-GLA) in the interaction of Vpx with the Fyn SH3 domain was next examined. Cos-7 cell lysates containing wild-type or mutant Vpx proteins were incubated with glutathione-agarose bead-bound Fyn SH3/SH2 domain fusion proteins. Our results indicate that exchange of proline residues at position 103 and 106 with serine in the C-terminal proline-rich domain completely blocked Vpx binding with the Fyn SH3 domain. In contrast, exchange of the glutamic acid residue at position 30 with proline as well as tyrosine residues at positions 66, 69, and 71 with alanine resulted in a wild-type interaction. Together, these data suggest that the C-terminal proline-rich motif plays a critical role in Vpx

interaction with SH3 domain of Fyn.



We performed coimmunoprecipitation assays in which the interaction between Vpx and Fyn was carried out *in vivo* (Fig. 2A). Cell lysates from GFP and GFP fusions containing wild-type as well as P103 and 106S mutant Vpx expression vectors were transfected in Cos-7 cells and immunoprecipitated with anti-Fyn polyclonal antibody followed by Western blot analysis using anti-GFP monoclonal antibody. Vpx specifically interacts and coprecipitates with endogenous Fyn kinase, and the exchange of proline 103 and 106 with serine in Vpx prevented this interaction with Fyn. Vpx interacts with both endogenous and ectopically expressed Fyn (Figure 2A). Interestingly, more binding of Vpx was noticed in cells

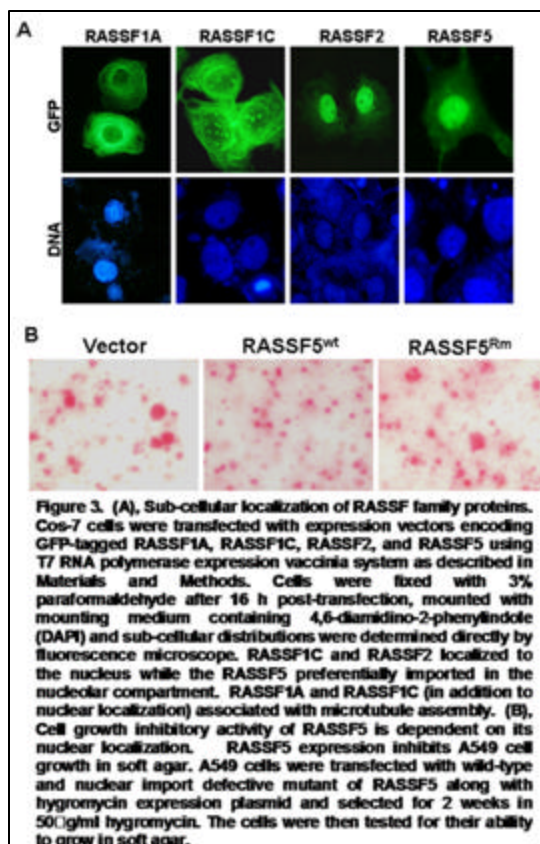
cotransfected with Vpx and Fyn; this correlates with Fyn expression in the cotransfected cells. These results reconfirmed the specific binding of Vpx to the cellular tyrosine kinase Fyn.

**Fyn kinase modulates nuclear export of Vpx.** All the potential phosphorylation residues such as serines, threonines, and tyrosines in Vpx were exchanged by site-directed mutagenesis to determine the role of phosphorylation in export of Vpx from nucleus. The phosphorylation status and expression status of all Vpx mutant proteins were ascertained by expressing them using a vaccinia virus expression system and labeling with  $^{32}\text{P}$  in Cos-7 cells. The lysates from the labeled cells were immunoprecipitated with anti-Vpx monoclonal antibodies and separated by SDS-15% PAGE. Equal levels of phosphorylation were found for all the Vpx mutants in the absence of kinase inhibitors. Analysis of Vpx mutants in the presence of various kinase inhibitors indicated that phosphorylation of STm (both serine and threonine residues were mutated) was completely inhibited by the tyrosine kinase inhibitor PP2 and tyrosine mutant (Y66, Y69, Y71) by the MAPK pathway inhibitor hypericin despite equal amounts of mutant proteins, as observed in the transfected cells. These results suggest that Vpx is phosphorylated by two different cellular kinases, MAPK-ERK-2 and Fyn. Using immunofluorescence analysis, Vpx-specific signal was observed for all the mutants in Cos-7 cells. Cytoplasmic localization for serine as well as threonine mutants and wild-type nuclear localization for tyrosine mutants confirms the role for serine and threonine residues in Vpx import into the nucleus. We next investigated whether export of Vpx from nucleus is dependent on Fyn-mediated phosphorylation. To this end, nuclear export of Vpx was determined in the presence of PP2 (inhibitor of Fyn kinase activity) alone or in combination with the translational inhibitor cycloheximide. Vpx localizes into the nucleus in both PP2-treated and untreated cells. Interestingly, PP2 inhibits the export of Vpx from nucleus. This inhibition was specific, since an inactive Fyn tyrosine kinase inhibitor, PP3, or a MAPK pathway inhibitor, PD98059, did not alter the cytoplasmic localization of Vpx in cells treated with cycloheximide. In addition, treatment with cycloheximide and LMB (alone or in combination) did not change the nuclear localization of Vpx Y66, Y69, 71A mutant protein. These data suggest that Fyn-mediated phosphorylation may be important for export of Vpx from nucleus.

**Nuclear export activity of Vpx modulates virus replication in macrophages.** To understand the importance of the nuclear export and virion incorporation properties of Vpx for optimal virus replication in nondividing cells, SIVsm(PBj1.9) proviral clones with the following characteristics were selected: (a) clones with wild-type Vpx virion incorporation and defective Vpx nuclear import and (b) clones with wild-type Vpx nuclear import and defective nuclear export as well as virion incorporation functions. All vpxmutant viruses replicated as efficiently



as the wild type and to high titers in macaque PBMCs but not in terminally differentiated monocyte-derived macaque macrophage cultures. As expected, mutant Vpx proteins that failed to localize in the nucleus were severely impaired in their ability to support virus replication in macrophages (Fig. 2B). For example, impairment of virus replication in macrophages was observed for Vpx H82S and P103, 106S mutants that are packaged into virus particles at levels similar to wild-type levels but failed to localize to the nucleus. Importantly, failure to replicate in macrophages was also observed for Y66, 69, 71A mutants despite maintaining wild-type nuclear localization (Fig. 2B). These Vpx mutant proteins failed to be exported into the cytoplasm and therefore could not be packaged into virus particles. PBj1.9 mutant viruses with defective nuclear export or virion incorporation of Vpx were found to replicate poorly in macrophages. These results suggest that nuclear export plays a critical role in translocation of Vpx into the cytoplasm for its packaging into virus particles. These findings support the notion that virion-associated Vpx with wild-type nuclear import and export is critical for efficient virus replication in nondividing target cells.

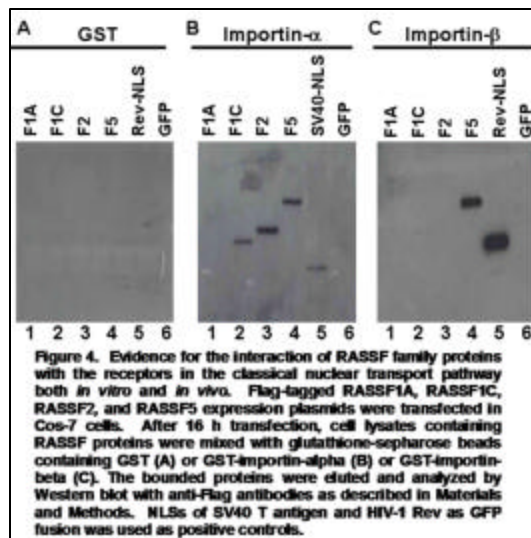


### Nuclear transport of ras associated tumor suppressor proteins

Ras proteins regulate a wide range of biological processes by interacting with a variety of effector proteins. In addition to the known role in tumorigenesis, activated form of Ras also exhibits growth-inhibitory effects by yet unknown mechanisms. Several Ras effector proteins identified as mediators of apoptosis and cell cycle arrest also exhibit properties normally associated with tumor suppressor proteins. Here, we show that Ras effector, RASSF5/NORE-1 binds strongly to K-Ras but weakly to both N-Ras and H-Ras. RASSF5 was found to localize both in nucleus and nucleolus in contrast to other Ras effector proteins, RASSF1C and RASSF2, which are localized in the nucleus excluded from

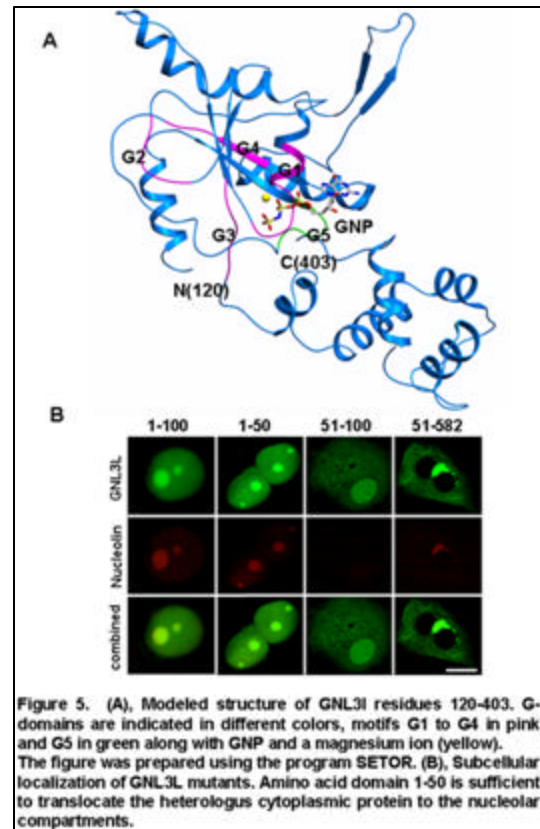
nucleolus (Fig. 3A). A 50-amino acid transferable arginine-rich nucleolar localization signal (NoLS) identified in RASSF5 is capable of interacting with importin-beta and transport the cargo into the nucleolus. Surprisingly, similar arginine-rich signals identified in RASSF1C and RASSF2 interact with importin-alpha and transport the heterologous cytoplasmic proteins to the nucleus. Interestingly, mutation of arginine residues within these nuclear targeting signals

prevented interaction of Ras effector proteins with respective transport receptors and abolished their nuclear translocation. These results provide evidence for the first time that arginine-rich signals are able to recognize different nuclear import receptors and transport the RASSF proteins into distinct sub-cellular compartments (Fig. 4). In addition, our data suggest



that the nuclear localization of RASSF5 is critical for its cell growth control activity (Fig. 3B). Together, these data suggest that the transport of Ras effector superfamily proteins into nucleus/nucleolus may play a vital role in modulating RAS mediated cell proliferation during tumorigenesis.

A novel lysine-rich domain and GTP binding motifs regulate the nucleolar retention of human guanine nucleotide binding protein, GNL3L. A variety of G-proteins and GTPases are known to be involved in nucleolar function. We describe here a new evolutionarily conserved putative human GTPase, guanine nucleotide binding protein-like 3-like (GNL3L). Genes encoding proteins related to GNL3L are present in bacteria and yeast to metazoa and suggests its critical role in development. Conserved domain search analysis revealed that the GNL3L contains a circularly permuted G-motif described by a G5-G4-G1-G2-G3 pattern similar to the HSR1/MMR1 GTP-binding protein subfamily. Highly conserved and critical residues were identified from a three-dimensional structural model obtained for GNL3L using the crystal structure of an Ylqf GTPase from *Bacillus subtilis* (Fig. 5A). We demonstrate here that GNL3L is transported into the nucleolus by a novel lysine-rich nucleolar localization signal (NoLS) residing within 1-50 amino acid residues. NoLS identified here is necessary and sufficient to target the heterologous proteins to the nucleolus (Fig. 5B). We show for the



first time that the lysine-rich targeting signal interacts with the nuclear transport receptor, importin-beta and transports GNL3L into the nucleolus. Interestingly, depletion of intracellular GTP blocks GNL3L accumulation into the nucleolar compartment. Furthermore, mutations within the G-domains alter the GTP binding ability of GNL3L and abrogate wild-type nucleolar retention even in the presence of functional NoLS, suggesting that the efficient nucleolar retention of GNL3L involves activities of both basic NoLS and GTP-binding domains. Collectively, these data suggest that GNL3L is composed of distinct modules, each of which plays a specific role in molecular interactions for its nucleolar retention and subsequent function(s) within the nucleolus.

### ***Publications:***

1. Du X, Rao MR, Chen XQ, Wu W, Mahalingam S and Balasundaram D (2006) The homologous putative GTPases Grn1p from fission yeast and the human GNL3L are required for growth and play a role in processing of nucleolar pre-rRNA. ***Molecular Biology of the Cell*** 1:460-474.
2. Rao MR, Kumari G, Balasundaram D, Sankaranarayanan R, and Mahalingam S (2006) A novel lysine-rich domain and GTP binding motifs regulates the nucleolar retention of human guanine nucleotide binding protein, GNL3L. ***Journal of Molecular Biology*** 364:637-654.
3. Singhal PK, Kumar PR, Rao MR, Kyasani M and Mahalingam S (2006) Simian Immunodeficiency Virus Vpx is imported into the nucleus via importin alpha dependent and independent pathways. ***Journal of Virology*** 80:526-536.
4. Singhal PK, Kumar PR, Rao MR and Mahalingam S (2006) Nuclear export of simian Immunodeficiency Virus Vpx protein. ***Journal of Virology*** 80:12271-12278.
5. Kumari G, Singhal PK, Rao MR, Mahalingam S (2007) Nuclear transport of Ras-associated tumor suppressor proteins: Different transport receptor binding specificities for arginine-rich nuclear targeting signals. ***Journal of Molecular Biology*** (InPress).



**Laboratory of Immunology**  
**Understanding the role of 1,2,4, thiadiazilidine - and cardiac glycoside on anti-inflammatory and anti-tumorigenic responses**

<b>Principal Investigator</b>	S K Manna	Staff Scientist
<b>Ph D Students</b>	A Sarkar	Senior Research Fellow
	G Charitha	Junior Research Fellow
	Maikho Thoh	Junior Research Fellow
<b>Other Member</b>	PB Raghavendra	Guest Worker
<b>Collaborator</b>	GT Ramesh	TSU, Texas, USA

*Objectives:*

1. Regulation of Interleukin-8 receptors in neutrophils and macrophages
2. Detection of novel cell death pathway induced by small molecules and its exploitation for design of chemotherapeutic agents for combination therapy in cancers.

**Summary of work done until the beginning of this reporting year**

Considering the role of interleukin-8 (IL-8) in a large number of acute and chronic inflammatory diseases, an understanding of the regulation of IL-8-mediated biological responses is important. A tridecapeptide inhibits most forms of inflammation by an unknown mechanism. In the present study, we found that alpha-melanocyte-stimulating hormone ( $\alpha$ -MSH) inhibited several IL-8-induced biological responses in macrophages and neutrophils. It downregulated receptors for IL-8, but not for TNF, IL-4, IL-13, or TRAIL in neutrophils. Level of neutrophil elastase, specific serine protease, but not cathepsin G or proteinase 3 increased in  $\alpha$ -MSH-treated cells and restoration of IL-8Rs by specific neutrophil elastase or serine protease inhibitor suggested the involvement of elastase in downregulation of IL-8Rs induced by  $\alpha$ -MSH. Oleandrin (a cardiac glycoside) inhibited IL-8-, formyl peptide (FMLP)-, EGF-, or nerve growth factor (NGF)-, but not IL-1- or TNF-induced NF- $\kappa$ B activation in macrophages. Oleandrin inhibited the binding of IL-8, EGF, or NGF, but not IL-1 or TNF. It decreased almost 79% IL-8 binding without altering affinity towards IL-8 receptors and this inhibition of IL-8 binding was observed in isolated membrane. Phospholipids significantly protected oleandrin-mediated inhibition of IL-8 binding thereby restoring IL-8-induced NF- $\kappa$ B activation. Oleandrin altered the membrane fluidity as detected by microviscosity parameter and a decrease in diphenylhexatriene (a lipid binding fluorophore) binding in a dose-dependent manner. Overall, our results suggested that oleandrin inhibits IL-8-mediated biological responses in diverse cell types by modulating IL-8Rs through altering membrane fluidity and microviscosity. These studies indicated that oleandrin might help to regulate IL-8-mediated biological responses

involved in inflammation, metastasis, and neovascularization. The compound, 5-(4-methoxyarylimino)-2-N-(3,4-dichlorophenyl)-3-oxo-1,2,4-thiadiazolidine (P<sub>3</sub>-25) is known to possess anti-bacterial, anti-fungal, anti-tubercular, and local anaesthetic activities. We studied the anti-tumorogenic activity of P<sub>3</sub>-25 and the role of NF- $\kappa$ B in this process. In constitutive NF- $\kappa$ B-expressing cells, treatment with P<sub>3</sub>-25 inhibited expression of NF- $\kappa$ B-dependent reporter gene, adhesion molecules, and cyclooxygenase. It decreased phosphorylation of p65 by inhibiting upstream kinases, such as protein kinase A and casein kinase II. By itself P<sub>3</sub>-25 induced apoptosis in NF- $\kappa$ B-expressing and doxorubicin-resistant breast cancer cells and in presence of other chemotherapeutic agents it potentiated apoptosis. Overall, our results suggested that P<sub>3</sub>-25 exerts anti-tumorogenic activity by inhibiting phosphorylation of p65, the transcriptional active subunit of NF- $\kappa$ B through inhibiting its upstream kinases, and potentiates apoptosis mediated by chemotherapeutic agents. These results have also suggested novel approaches for design of anti-cancer drugs for combination chemotherapy.

#### **Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)**

##### **A. Oleandrin induces apoptosis in human, but not in murine cells: dephosphorylation of Akt, expression of FasL, and alteration of membrane fluidity.**

A common practice has been to evaluate the efficacy of any compound as drug in cell-based *in vitro* system followed by *in vivo* murine model prior to clinical trial in human. Cardiac glycosides are very effective to kill human cells, but not murine cells. In this report, we describe the comparative molecular mechanism of the cardiac glycoside oleandrin's action in human and murine cells. Treatment with oleandrin facilitated nuclear translocation of FKHR in human, but not murine cells, by dephosphorylating Akt. It activated MAPK and JNK in human, but not in murine cells and also induced expression of FasL leading to apoptosis in human cells as detected by assaying caspase activation, PARP cleavage, nuclear fragmentation, and annexin staining. Oleandrin interacted with human plasma membrane as evaluated by HPLC, altered its fluidity as detected by DPH binding, inhibited Na<sup>+</sup>/K<sup>+</sup>-ATPase activity, and increased intracellular free Ca<sup>2+</sup> level followed by calcineurin activity only in human, but not in murine cells. Results suggest that human plasma membrane might be different than murine, which interact with oleandrin that disturb Na<sup>+</sup>/K<sup>+</sup>-ATPase pump resulting in the calcification followed by induction of Ca<sup>2+</sup>-dependent cellular responses such as apoptosis.



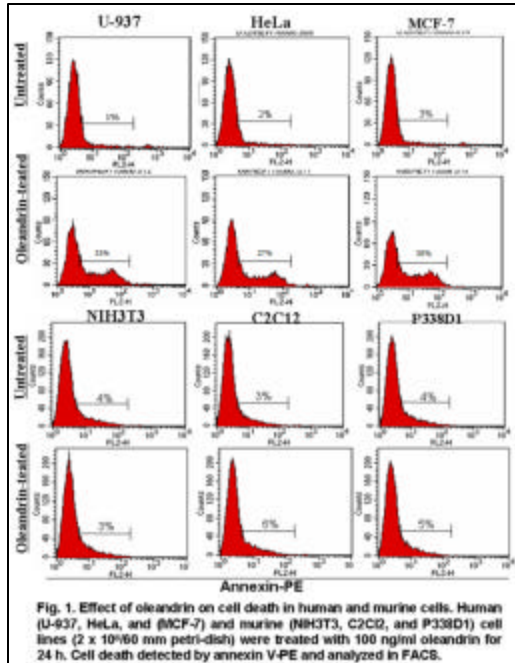


Fig. 1. Effect of oleandrin on cell death in human and murine cells. Human (U-937, HeLa, and MCF-7) and murine (NIH3T3, C2C12, and P338D1) cell lines ( $2 \times 10^6$  80 mm petri-dish) were treated with 100 ng/ml oleandrin for 24 h. Cell death detected by annexin V-PE and analyzed in FACS.

measured. A gradual decrease in the amount of phospho-Akt was observed in oleandrin incubated U-937, but not NIH3T3 cells with increasing time (Fig. 2A). As Akt dephosphorylation is required for nuclear translocation of forkhead transcription factor (FKHR), the amounts of FKHR were measured from cytoplasm and nucleus from oleandrin-treated U-937 and NIH3T3 cells. Oleandrin treatment led to gradual decrease in amounts of FKHR in the cytoplasmic extracts with increasing times and that was corresponding with the gradual increase in the amounts of FKHR in the nuclear extracts in U-937, but not in NIH3T3 cells (Fig. 2B). These results suggest that oleandrin induces translocation of FKHR from cytoplasm to nucleus in human cells. The amount of FasL increased with time of oleandrin treatment in U-937 but not in NIH3T3 cells as detected by Western blot (Fig. 2C) and RT-PCR (Fig. 2D).

Oleandrin induces apoptosis in human but not murine cells. U-937, HeLa, MCF-7, NIH3T3, C2C12, and P338D1 cells were treated with oleandrin (100 ng/ml) for 24 h and live and dead cells were detected by annexin V-PE in FACS. The Fig.1 shows that oleandrin induces cell death by 33%, 27%, and 36% in U-937, HeLa, and MCF-7 cells respectively. But other murine cells like NIH3T3, C2C12, and P338D1 remain unaffected.

Oleandrin decreases amount of phospho-Akt and induces nuclear translocation of FKHR and FasL expression in U-937 cells. U-937 and NIH3T3 cells were treated with oleandrin for different time periods and phospho-Akt was

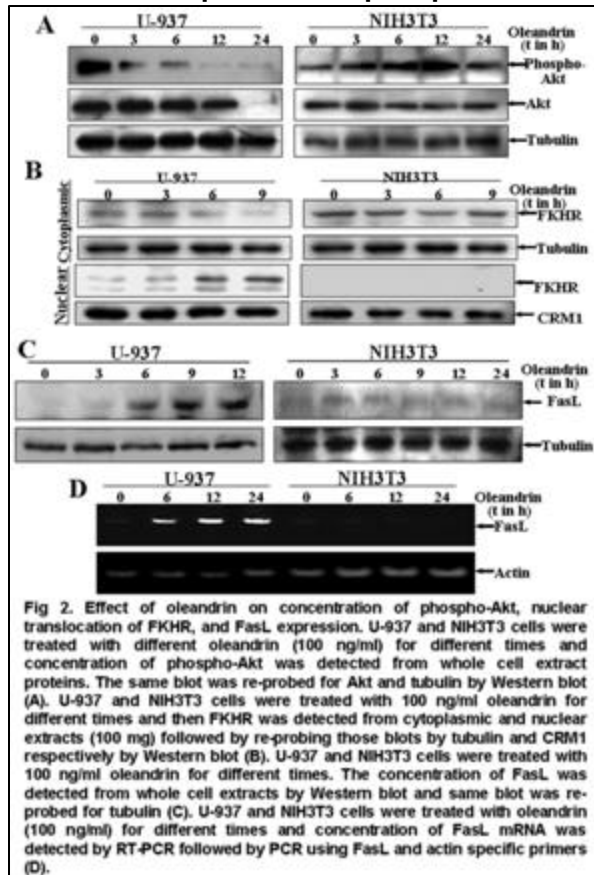
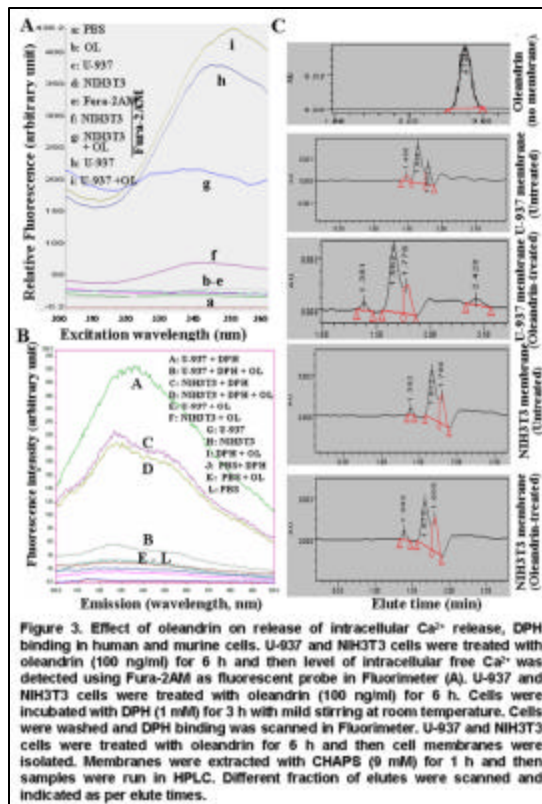


Fig 2. Effect of oleandrin on concentration of phospho-Akt, nuclear translocation of FKHR, and FasL expression. U-937 and NIH3T3 cells were treated with different oleandrin (100 ng/ml) for different times and concentration of phospho-Akt was detected from whole cell extract proteins. The same blot was re-probed for Akt and tubulin by Western blot (A). U-937 and NIH3T3 cells were treated with 100 ng/ml oleandrin for different times and then FKHR was detected from cytoplasmic and nuclear extracts (100 mg) followed by re-probing those blots by tubulin and CRM1 respectively by Western blot (B). U-937 and NIH3T3 cells were treated with 100 ng/ml oleandrin for different times. The concentration of FasL was detected from whole cell extracts by Western blot and same blot was re-probed for tubulin (C). U-937 and NIH3T3 cells were treated with oleandrin (100 ng/ml) for different times and concentration of FasL mRNA was detected by RT-PCR followed by PCR using FasL and actin specific primers (D).

**Oleandrin increases intracellular free Ca<sup>2+</sup>, alters membrane fluidity and interacts with plasmamembrane.** Oleandrin treatment enhanced intracellular Ca<sup>2+</sup> level in U-937 cells as detected by Fura-2AM fluorescent probe. In NIH3T3, oleandrin did not increase intracellular



free Ca<sup>2+</sup> level (Fig. 3A). Fluorescence spectrum obtained on DPH binding decreased in oleandrin-treated U-937 cells. Remarkably no alteration of DPH binding was observed in NIH3T3 cells (Fig. 3B). These results suggest that oleandrin alters membrane fluidity only in human cells. Interaction of oleandrin with cell membrane was evaluated by HPLC. Both human and murine cells were treated with oleandrin, membranes were isolated, and extracted with CHAPS. These extracts were run through a HPLC column. Average retention times for oleandrin alone (without membrane extract) and oleandrin treated U-937 cells membrane extract were 2.6 and 2.4 respectively whereas no peaks were observed in the extracts of cell membrane from U937 (untreated), NIH3T3 (untreated) and NIH3T3

(oleandrin-treated) cells (Fig. 3C).

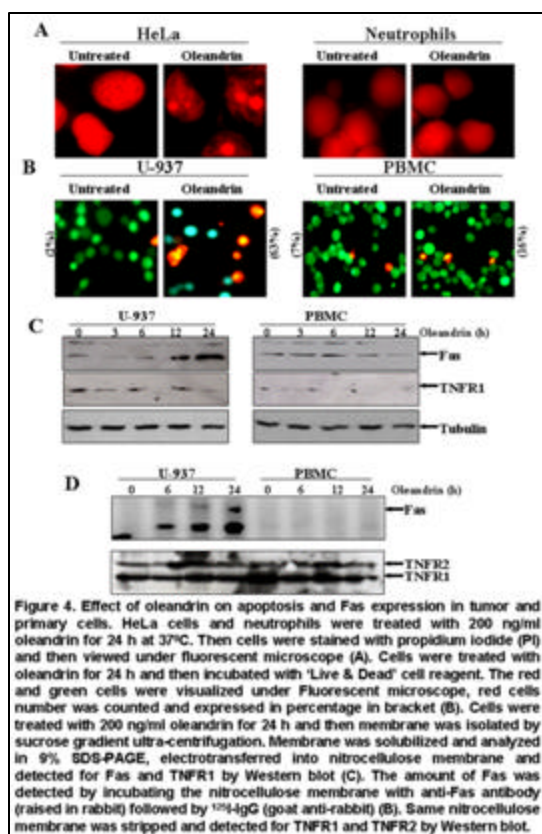
We report here for the first time that oleandrin-mediated cell death is occurring through alteration of membrane fluidity due to binding of oleandrin with membrane in human tumor cells. Oleandrin does not interact with murine membrane and thus bypass oleandrin-mediated cell death. This study might restrict us to choose the proper *in vivo* model to test the efficacy of any drug.

### **B. Oleandrin potentially induces apoptosis in tumor cells, but not in primary cells**

Chemotherapeutic agent is characterized by its concentration in tumor cells with slight side effects. Oleandrin, a polyphenolic cardioglycoside and known to induce apoptosis in tumor cells. No report is available on its efficacy in primary cells. In this report we are providing the evidences that oleandrin induces apoptosis in tumor cells potently but not in primary cells like peripheral blood mononuclear cells (PBMC) and neutrophils isolated from fresh human blood as detected by cytotoxicity, caspase activation, lipid peroxidation, reactive oxygen intermediates (ROI) generation, and DNA fragmentation. Oleandrin inhibited NF- $\kappa$ B activation in tumor cells, but not in primary cells. Oleandrin failed to inhibit apoptosis in NF- $\kappa$ B



overexpressed tumor cells. Primary cells when downregulated with NF- $\kappa$ B failed to induce apoptosis by oleandrin. Oleandrin induced Fas, thereby induced apoptosis in tumor cells, but not in primary cells. Overall, these results suggest that oleandrin mediates apoptosis in tumor



cells by inducing Fas, but not in primary cells indicating its potential anti-cancer property with no or slight side effect.

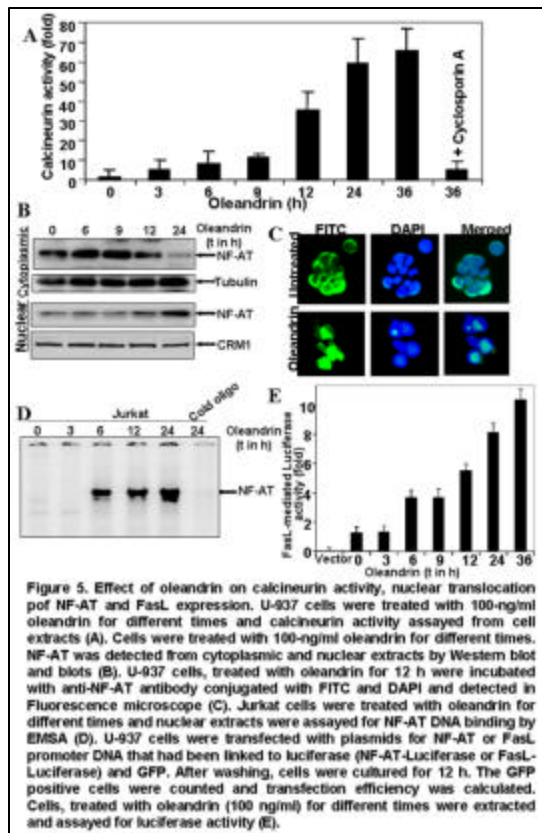
**Oleandrin induces apoptosis in tumor cells, but not in primary cells.** Apoptosis is reflected in DNA and nuclear fragmentation. Propidium iodide stained cells showed the nuclear fragmentation, as oleandrin induced it in HeLa cells, but not in neutrophils (Fig. 4A). Oleandrin induced 63% cell death as shown by the number of red cells in U-937 cells and 16% in case of PBMC as detected by 'Live & Dead' assay (Fig. 4B) indicating that the tumor but not primary cells are affected by oleandrin.

**Oleandrin up-regulates Fas, but not anti-TNFR1 in U-937 cells, but not in PBMC.** To detect the level of Fas and TNFR1 in oleandrin treated U-937 and PBMC, cells were treated

with oleandrin (0.2  $\mu$ g/ml) for different times at 37°C. Cell membrane then isolated, solubilized, and 100  $\mu$ g membrane soluble proteins were analyzed for Fas and TNFR1 by Western blot. Oleandrin induced Fas at 12 and 24 h in U937 cells, but not in PBMC, whereas the level of TNFR1 remained unchanged in U-937 cells or PBMC (Fig. 4C). The solubilized membrane protein (100  $\mu$ g) were analyzed in 9% SDS-PAGE and transferred to nitrocellulose membrane. Then incubated with anti-Fas antibody followed by <sup>125</sup>I-IgG (goat anti-rabbit). Then the blot was exposed and scanned in PhosphorImager (Fuji, Japan). The result in Fig. 4D reveals that the level of Fas binding is increased gradually with time of oleandrin treatment in U-937 cells, but not in PBMC further suggesting that up-regulation of Fas in tumor cells (but not in primary cells) might cause cell death.

### C. Cardiac glycoside induces cell death via FasL by activating calcineurin and NF-AT, but apoptosis initially proceeds through activation of caspases

Decrease in caspase activity is a common phenomenon in drug resistance. For effective therapeutic intervention, detection of such agents, which affects other pathway independent of caspases to promote cell death, might be important. Oleandrin, a polyphenolic glycoside



**Figure 5.** Effect of oleandrin on calcineurin activity, nuclear translocation of NF-AT and FasL expression. U-937 cells were treated with 100-ng/ml oleandrin for different times and calcineurin activity assayed from cell extracts (A). Cells were treated with 100-ng/ml oleandrin for different times. NF-AT was detected from cytoplasmic and nuclear extracts by Western blot and blots (B). U-937 cells, treated with oleandrin for 12 h were incubated with anti-NF-AT antibody conjugated with FITC and DAPI and detected in Fluorescence microscope (C). Jurkat cells were treated with oleandrin for different times and nuclear extracts were assayed for NF-AT DNA binding by EMSA (D). U-937 cells were transfected with plasmids for NF-AT or FasL promoter DNA that had been linked to luciferase (NF-AT-Luciferase or FasL-Luciferase) and GFP. After washing, cells were cultured for 12 h. The GFP positive cells were counted and transfection efficiency was calculated. Cells, treated with oleandrin (100 ng/ml) for different times were extracted and assayed for luciferase activity (E).

induced cell death through activation of caspases in a variety of human tumour cells. In this report we provide evidence that besides caspases activation, oleandrin interacts with plasma membrane, changes fluidity of the membrane, disrupts Na<sup>+</sup>/K<sup>+</sup>-ATPase pump, enhances intracellular free Ca<sup>2+</sup> and thereby activates calcineurin (Fig. 5A). Calcineurin, in turns, activates nuclear transcription factor NF-AT (Fig. 5, B-D) and its dependent genes such as FasL (Fig. 5E), which induces cell death as a late response of oleandrin. Cell death at early stages is mediated by caspases where inhibitors partially protected oleandrin-mediated cell death in vector-transfected cells, but almost completely in Bcl-xL-overexpressed cells. Overall, our data suggest that oleandrin might be an important therapeutic molecule in

case of tumors where cell death pathway occurs due to deregulation of caspase-mediated pathway

#### D. Thiadiazolidine derivatives induce cell death by arresting cell cycle and FasL expression via Akt/FKHR pathway

The 1,2,4-thiadiazolidine derivatives have been shown to induce cell death but the exact mechanism is not known. In this study we have obtained evidence that dichlorophenyl form of thiadiazolidine (designated as P<sub>3</sub>-25) is a strong inducer of cell death in comparison with other forms. P<sub>3</sub>-25 potentially arrested cell cycle at G1 phase and down regulated the amounts of cyclin D1 and cyclin E without interfering with levels of p16 and p27. It decreased c-Myc level and thereby inhibited DNA binding ability of Myc-Mad complex. P<sub>3</sub>-25 dephosphorylated retinoblastoma and Akt thereby facilitated nuclear translocation of forkhead transcription factor, which in turn, expressed its dependent gene FasL. Activated FasL inhibited cell proliferation and induced cell death. Overall our results suggest that P<sub>3</sub>-25 derivative exerts anti-tumor activities by decreasing Myc-mediated response and increasing FasL expression, which may have role for designing such drugs for tumor therapy.

*Publications:*

1. Manna SK (2006) Alpha-MSH and IL-8-induced biological responses. ***Modern Aspects of Immunobiology*** 19:32-33.
2. Sarkar S, Wise KC, Manna SK, Ramesh V, Yamauchi K, Thomas RL, Wilson BL, Kulkarni AD, Pellis NR and Ramesh GT (2006) Activation of activator protein-1 in mouse brain regions exposed to stimulated microgravity. ***In Vitro Cell Developmental Biology– Animal*** 42:96-99.
3. Sreenivasan Y, Raghavendra PB and Manna SK (2006) Oleandrin-mediated expression of Fas potentiates apoptosis in tumor cells. ***Journal of Clinical Immunology*** 26:308-322.
4. Manna SK, Manna P and Sarkar A (2007) Inhibition of RelA phosphorylation sensitizes chemotherapeutic agents-mediated apoptosis in constitutive NF- $\kappa$ B-expressing and chemoresistant cells. ***Cell Death & Differentiation*** 14:158-170.
5. Raghavendra PB, Sreenivasan Y, Ramesh GT and Manna SK (2007) Cardiac glycoside induces cell death via FasL by activating calcineurin and NF-AT, but apoptosis initially proceeds through activation of caspases. ***Apoptosis*** 12:307-318.
6. Raghavendra PB, Sreenivasan Y and Manna SK (2007) Oleandrin induces apoptosis in human, but not in murine cells: dephosphorylation of Akt, expression of FasL and alteration of membrane fluidity. ***Molecular Immunology*** 44:2292-2302.
7. Manna SK, Aggarwal RA, Sethi G, Aggarwal BB and Ramesh GT (2007) Morin (3,5,7,2',4'-pentahydroxyflavone) abolishes NF- $\kappa$ B activation induced by various carcinogens and inflammatory stimuli, leading to suppression of NF- $\kappa$ B-regulated gene expression and upregulation of apoptosis. ***Clinical Cancer Research*** (InPress).



## Laboratory of Bacterial Genetics

### Studies on transcription termination and amino acid and ion-transport

<b>Principal Investigator</b>	J Gowrishankar	Staff Scientist (and Director)
<b>Ph.D Students</b>	Rakesh Laishram	Senior Research Fellow
	Syeda Aisha Haneea	Senior Research Fellow
	Shivalika Saxena	Junior Research Fellow
	Carmelita N Marbaniang	Junior Research Fellow (since July 2006)
<b>Other Members</b>	V K Mishra	Staff Scientist III
	K Anupama	Staff Scientist III
	T S Shaffiqu	Technical Assistant
	Abhijit A Sardesai	Postdoctoral Fellow
	J Krishna Leela	Postdoctoral Fellow

#### Objectives:

- 1 To test the model of and mechanisms mediating Rloop formation from nascent untranslated transcripts
- 2 To study the mechanism of ArgP-mediated transcriptional regulation of the arginine exporter ArgO
- 3 To investigate an unusual phenomenon of K<sup>+</sup> toxicity in *hns trx* double mutant strains

#### Summary of work done until the beginning of this reporting year

Based on genetic studies in our lab with *rho* and *nusG* mutants defective in factor-dependent transcription termination (Harinarayanan and Gowrishankar, J. Mol. Biol., 2003, 332: 31-46; Gowrishankar and Harinarayanan, Mol. Microbiol., 2004, 54: 598-603), we had proposed a model that *rho* and *nusG* mutants suffer increased chromosomal R-loops by the re-annealing of nascent untranslated transcripts (that are expected to occur with increased frequency in the mutants) to the DNA upstream of the elongating RNA polymerase, and that such R-loops are inimical to the health of the cells because they block movement of succeeding RNA polymerase molecules and/or of the replication fork. (An R-loop is a structure in which one strand of double-stranded DNA is hybridized with RNA, with the complementary DNA strand concomitantly displaced.). Accordingly, we had suggested that, given the toxicity associated with R-loops in the chromosome, Rho's action in mediating the premature termination of untranslated transcripts in bacteria is indeed its primary function. The novel phenotypes of

*rho* and *nusG* mutants identified by us included (i) lethality with ColE1-like plasmids such as pACYC184 whose replication is R-loop-dependent (for which we have proposed that the lethality is caused by uncontrolled replication of the plasmids consequent to a titration effect of increased chromosomal R-loops in the mutants); (ii) suppression of the plasmid-mediated lethality phenotype by overexpression of known R-loop removing enzymes such as RNase H1 or RecG from *E. coli* and UvsW from phage T4; and (iii) synthetic lethality of *rho/nusG* mutations on the one hand with *recG/rnhA* mutations on the other (the latter encoding RNase H1 and RecG, respectively). We had also obtained genetic evidence that the endoribonuclease RNase E is required to promote R-loop formation in *rho* and *nusG* mutants, and we have proposed that the absence of a fully functional RNase E is associated with a reduced propensity of endonucleolytic cleavage of the nascent untranslated transcripts and consequently with a reduced propensity of R-loop formation in *rho* and *nusG* mutants. Overexpression of phage T7 protein kinase, which is known to phosphorylate RNase E and inactivate it, also led to suppression of R-loop-associated phenotypes of the *rho* and *nusG* mutants.

In our lab, the *argP* gene product was identified as a transcriptional regulator of an anonymous open-reading frame *yggA* (or *argO*) that was deduced to encode an arginine exporter, designated by us as ArgO (Nandineni and Gowrishankar, J. Bacteriol., 2004, 186: 3539-3546). Transcription of *argO* (*yggA*) was induced by arginine (Arg) in the presence of ArgP, whereas lysine (Lys) supplementation strongly inhibited its transcription. Primer-extension analysis experiments demonstrated that the start-site of *argO* transcription is at an A residue 28 bases upstream of the predicted translation start site of the structural gene. The ArgP protein (with N-terminal His-tag) was overexpressed and purified and shown to be a dimer in solution. Electrophoretic mobility shift assay (EMSA) experiments established that ArgP binds to *argO* with nanomolar affinity in the absence or presence of Arg or Lys.

We have earlier shown (Sardesai and Gowrishankar, J. Bacteriol., 2001, 183: 86-93) that transcription of the *kdp* operon, which encodes an inducible high-affinity K<sup>+</sup>-uptake system, is substantially reduced in strains mutant for thioredoxin or thioredoxin reductase (encoded by *trxA* and *trxB*, respectively) as well as for the abundant nucleoid protein H-NS (enclosed by *hns*), but the mechanism underlying these effects are not known. Subsequently, we observed an unusual phenomenon of K<sup>+</sup> toxicity in these double mutants (*trxA/trxB hns*) whereby their growth is severely impaired in media whose K<sup>+</sup> concentration is >10 mM. Suppressors of this phenotype were obtained and shown to be transposon insertions in *ahpC*, *ptsP*, *rpoS*, or *metL*.

**Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)**

The following significant findings were made by the group in the current year:

### **1. Chromosomal R-loops from nascent untranslated transcripts in bacteria**

In the current year, the following findings were obtained:

- a. The Gterminal half of RNase E is known to provide a scaffold for assembly of a multiprotein complex called the degradosome which is involved in degradation of mRNA. Two other proteins in the degradosome are polynucleotide phosphorylase (PNPase) and an RNA helicase RhlB, but mutation in neither of the cognate genes (*pnp* and *rhlB*, respectively) was able to suppress *nusG//rho*-pACYC184 lethality.
- b. Even moderate overexpression of RNase E conferred growth inhibition in *rho* and *nusG* mutants (but not in the wild-type strain), and the growth inhibition was suppressed by overexpression of UvsW. These results lend support to the notion that RNase E action potentiates R-loops toxicity in *rho* and *nusG* mutants.
- c. A direct test to determine R-loop formation from nascent untranslated transcripts in *rho* and *nusG* mutants was undertaken at two genetic loci *lacZ* and *trpE*, following the introduction of premature stop-codon mutations in the two ORFs. The basis of this test was that R-loop formation will lead to the displacement as single-stranded DNA of the non-template DNA strand which will now become susceptible to bisulphite-mediated conversion of C residues to uracils. The results obtained so far indicate that indeed C→T mutations occur on the non-template strand at high frequency under these conditions, and that their occurrence is dependent on (i) presence of *rho/nusG* mutation, (ii) transcription of the *lacZ* and *trpE* genes, as well as (iii) absence of translation downstream of the stop-codon mutations in the genes.

### **2. ArgP-mediated transcriptional regulation of ArgO**

In the current year, we have elucidated the mechanism of ArgP-mediated transcriptional regulation of *argO*. The model of ArgP-mediated regulation of *argO* transcription which emerges from the results is the following. In the absence of ArgP, RNA polymerase (RNAP) engages with the *argO* promoter to form a relatively unstable binary complex, resulting in just a low basal level of productive *argO* transcription. The ArgP dimer has a common binding site(s) for its ligands Arg and Lys, and is able to bind the sequence between approximately -85 and -20 of *argO* to form a stable binary complex in both its liganded and unliganded states. However, the binding of unliganded ArgP and of RNAP to the *argO* regulatory region appear to be mutually exclusive, and it is only ArgP in its ligand-bound form which is able to co-operatively bind along with RNAP at *argO* to establish formation of a stable open complex. Whereas the open complex established by RNAP through the mediation of Arg-bound ArgP is proficient for productive transcription, that mediated by Lys-bound ArgP is trapped in a paused and inactive state. The latter, however, can be chased into the productive state by competitive replacement of the Lys ligand which is bound to ArgP by Arg.

We therefore propose that the *argO* case is the first known instance of an environmental signal regulating bacterial transcription initiation at the final stage of promoter escape (ie., after open complex formation) by RNAP. Mechanistically, one possibility is that the RNAP complex at *argO* in the presence of ArgP and Lys exists in a dominant Rip Van Winkle state after melting of the promoter duplex at the –10 region, and that it is blocked in the step of “DNA scrunching”, which has recently been shown to be an obligatory prerequisite both for abortive initiation as well as for promoter clearance during productive transcription. We also propose that, in *E. coli* cells, the ternary complex remains assembled and poised at the *argO* promoter at all times to respond, positively or negatively, to instantaneous changes in the ratio of intracellular Arg to Lys concentrations.

### **3. K<sup>+</sup> toxicity in E. coli *trx hns* double mutants**

In the current year, we have obtained several lines of genetic evidence that at least one mechanism of suppression of K<sup>+</sup> toxicity in *trx hns* double mutants operates through the loss of RpoS activity, achieved either directly or through a pathway linking OxyR and RpoS. Thus, it is known that the OxyR regulon is activated by hydrogen peroxide or by mutations in *ahpC*, and that one of the members of this regulon OxyS (a non-coding RNA) binds the Hfq protein to lead to inhibition of RpoS translation. Thus, we found that K<sup>+</sup> toxicity in the *trx hns* double mutants could be reversed by mutations in *rpoS* or *ahpC*, or by addition of hydrogen peroxide to the medium. A gain-of-function mutation in *oxyR* (*oxyR2*) that renders the protein constitutively active was also effective in suppressing the K<sup>+</sup> toxicity phenotype, whereas a null mutation in *oxyR* was synthetic lethal in the *trx hns* double mutants. Taken together, these results indicate that reduction of RpoS function serves to alleviate the K<sup>+</sup> toxicity phenotype, the mechanism(s) for which remains to be ascertained.

#### ***Publications:***

1. Laishram RS and Gowrishankar J (2007) Environmental regulation operating at the promoter clearance step of bacterial transcription. ***Genes & Development***.

#### ***Patents:***

1. Gowrishankar J and Harinarayanan R. A method of altering levels of plasmids. Application No. 10/266,510: US Patent No. 7,176,028 B2 issued February 13, 2007.
2. Gowrishankar J and Harinarayanan R. A method of altering levels of plasmids. Application No. 11/377,380: Notice of Allowance dated 12 January, 2007 issued by the US Patent Office.



## Laboratory of Computational Biology

<b>Principal Investigator:</b>	H A Nagarajaram	Staff Scientist
<b>Other Members:</b>	V B Sreenu	Senior Research Fellow (Until June, 2006)
	Sridhar M Achary	Senior Research Fellow
	Pankaj Kumar	Senior Research Fellow
	Md Tabrez Anwar Shamim	Senior Research Fellow
	Mohammad Anwaruddin	Project Assistant
	Suresh Babu	Project Assistant

### **1. Dynamics-Function correlation studies on the Wild-type and PCG associated Mutant forms of Human CYP1b1**

**Objective:** To study the effect of deleterious mutations on the collective and correlated motions of human CYP1b1 protein.

#### **Summary of the work done until the beginning of this reporting year**

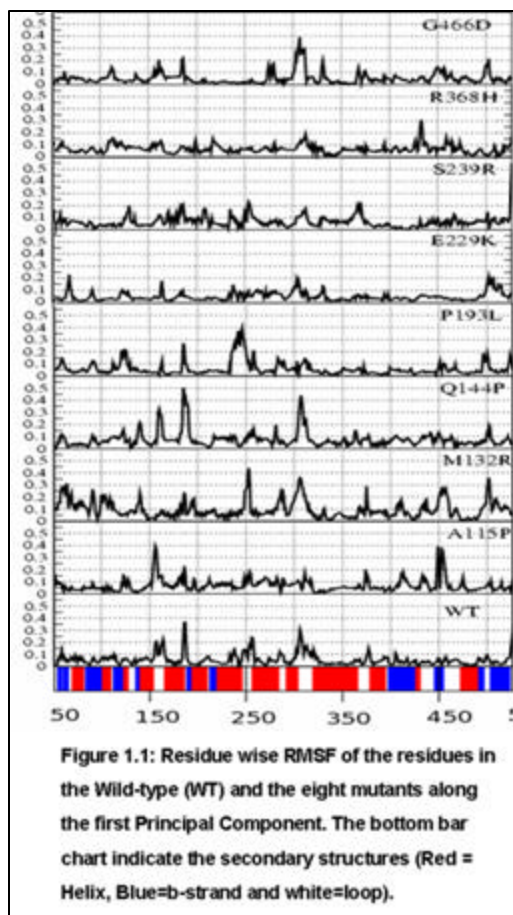
Molecular dynamics simulations of 30ns duration were carried out on the wild-type (WT) and the eight different PCG disease mutants (MTs) of the Human CYP1b1 protein. Analysis of time evolution as well as time averaged values of various structural properties, especially of those of the functionally important regions: the heme binding region, substrate binding region and substrate access channel, gave some insights into the possible structural characteristics of disease mutant and the wild-type forms of the protein. In a nutshell, as compared to the wild-type the core regions in the mutant structures are associated with subtle but significant changes and that the functionally important regions seem to adopt such structures that are not conducive for the wild-type like functionality.

#### **Details of progress in the current reporting year (April 1, 2006– 31, March 2007):**

In continuation of the above work we carried out principal component analysis (PCA) also referred to as Essential Dynamics to get an insight into the large-scale, functionally significant correlated motions in the wild-type and the disease mutants.

The principal components (PCs) were calculated by diagonalizing the Ca positional covariance matrix generated from the trajectories. The eigenvectors thus generated were arranged in the descending order of their corresponding eigenvalues. Figure 1.1 shows the plot of eigenvalues of WT and MTs which reveal that first few eigenvectors sufficiently

describe the total fluctuations in the dynamics. In all the structures, the first 3 eigenvectors have eigenvalues  $\geq 1$ . The Total positional fluctuation for the first 3 eigenvectors in the WT and the MTs viz., A115P, M132R, Q144P, P193L, E229K, S239R, R368H, G466D, are respectively 45.2%, 53.2, 61.1, 57.9, 49.6, 46.2, 48.1, 46.4 and 44.2.



2D projection of the trajectories along the first and second PCs gave the characteristic semi-circle shaped graph, indicative of cosine contents. Though this feature has been consistently observed in protein simulations, it has been suggested to consider this carefully while analyzing the first PC, giving due consideration to the random diffusion component in the dynamics. We are further investigating into this aspect.

In order to identify the distribution of fluctuations in the protein sequence along the PCs, we examined the RMSF fluctuations. In the WT only the loop regions show large fluctuations along the first PC direction (Fig.1.2). Also included in this is the F/G loop region, which was shown to be involved in the substrate access. All the mutants also show fluctuations in these regions and in addition have fluctuations in many other

regions. E229K, Q144P, R368H and G466D do not show fluctuations in the F/G loop region. Thus in these mutants the F/G loop region does not exhibit the flexibility as observed in the WT. Mutant E229K shows very reduced fluctuations, observed only in the terminal loops of the molecule, indicating a relatively rigid nature of the molecule. Mutant M132R shows the highest flexibility among all the molecules.

Table 1.1 shows the % of residues showing RMSF higher than  $1\text{\AA}$  in the Functionally Important Regions (FIRs) along the first PC. SAC region commonly shows fluctuations along this PC. In the WT, the HBR region shows no fluctuations. In other mutations except P193L, the HBR region shows considerable fluctuations. The SBR region, responsible for substrate binding, is also expected to be relatively least fluctuating. The WT shows small fluctuation (5%) in this region. This is probably required for optimum docking and binding of the

substrate in the binding pocket. Some of the Mutants (M132R, P193L, E229K, S239R and G466D) show higher fluctuations in SBR, while mutants A115P, Q144P and R368H show no fluctuations in SBR. E229K when analyzed with respect to the individual secondary structures, showed the least number of structures with high fluctuations. But when analyzed for the entire FIR regions, it showed considerable increase in fluctuations than the WT. P193L showed the highest fluctuations in the SAC region, indicating marked changes in the

**Table 1.1**  
**Percentage of residues in various FIRs showing RMSF>1Å**

	HBR (24)	SBR (17)
WT	0	6
A115P	4	0
M132R	4	24
Q144P	4	0
P193L	0	12
E229K	8	18
S239R	13	18
R368H	13	0
G466D	8	12

dynamics of this region. Thus this study revealed the effects of the mutants in terms of the differences in collective motions between the WT and Mutants. Further analysis is underway.

## **2. Knowledge based approach for protein fold-recognition**

### **Objectives:**

To develop a new fold-recognition tool with a sensitivity and specificity better than the existing tools

### **Summary of the work done until the beginning of this reporting year**

We initiated a new project on development of a protein fold-recognition procedure having sensitivity and specificity better than the existing methods. The underlying principle of the new method stems from the fact that the number of protein folds in the universe is limited and that each protein adopts one of these known folds and therefore the protein fold-recognition can essentially be regarded as a classification problem where protein sequences can be classified into different fold-categories. For this the most important requirement is a set of easily computable protein fold specific features for each fold. Analysis of dipeptide distribution in various folds revealed that the each fold can be characterized by the presence/absence as well as preferences or otherwise of a unique set of dipeptides. Although this was an interesting discovery, the same information was of less use for prediction purposes due to the fact that some folds are thinly populated and therefore the observed dipeptide frequencies could simply be the artifact of the number of available protein sequences. We therefore decided to use the entire information of available dipeptides for class discrimination by means of the supervised machine learning algorithm-support vector machine (SVM).

### **Details of progress in the current reporting year (April 1, 2006– March 31, 2007)**

We have developed a SVM-based method for protein fold-recognition using secondary structural state and solvent accessibility state information of amino acids and dipeptides as

discriminatory features. These features have not been exploited so far for protein fold-recognition to the best of our knowledge. The accuracy obtained using our method is strikingly higher than the other existing methods.

### The data sets used for training and testing SVM

The training and the test data sets used in this study contains 302 and 379 proteins, respectively (<http://crd.lbl.gov/~cding/protein/>) and these according to SCOP classification belong to 27 different folding classes representing all major structural classes: all  $\alpha$ , all  $\beta$ ,  $\alpha/\beta$ ,  $\alpha + \beta$ , and small proteins. These are the datasets used by several authors for benchmarking their fold-recognition tools. No two proteins in the training set have more than 35% sequence identity to each other and similarly none of the proteins in the test set has greater than 25% sequence identity to the proteins in the training set.

Table 2.1. Various fold-discriminatory features used along with their notations and dimensions

S. No.	Feature	Feature Notation	Dimensions
<b>Sequence Features</b>			
1	Amino acid composition	AAC	20
2	First order amino acid pair (dipeptide) composition	DI	400
3	Second order amino acid pair (1-gap dipeptide) composition	1GAP	400
<b>Structural Features</b>			
4	Secondary structural state frequencies of amino acids	SSS_AA	60
5	Secondary structural state frequencies of dipeptides	SSS_DI	1200
6	Secondary structural state frequencies of 1-gap dipeptides	SSS_GAP	1200
7	Solvent accessibility state frequencies of amino acids	SAS_AA	60
8	Solvent accessibility state frequencies of dipeptides	SAS_DI	1200
9	Solvent accessibility state frequencies of 1-gap dipeptides	SAS_GAP	1200

this study (Table 2.1).

### Fold-discriminatory features used

Machine learning methods such as Neural Network (NN) and Support Vector Machine (SVM) require inputs of fixed length. This necessitates a strategy for encapsulating the global information of proteins of variable length in a fixed length vector. Amino acid and amino acid pair based features provide us a way to do so. We extracted both sequence and structure-based features for

### Training and Testing using SVM

Support vector machines (SVM) were basically designed for binary classification. To extend it to multi-class classification, many methods have been developed. Currently there are two kinds of methods, one is binary classification based method, which constructs and combines several binary classifiers, and the other is referred to as 'All-together method', which directly considers all data in one big optimization formulation. Since protein-fold recognition is a typical multi-class problem we used multi-class methods viz., All-together method: Crammer and Singer method and the two binary classification based methods: "one versus all," & "one versus one". To the best of our knowledge ours is the first study to use Crammer and Singer multi-class method for protein fold classification.

## Performance Measure

Overall accuracy was used as a measure to evaluate the performance of our approach. Overall accuracy is the most commonly used parameter for assessing the global performance of multi-class problems. Overall accuracy (Q) is defined as the number of instances correctly predicted over the total number of instances in the test set:

$$Q = \frac{\sum_i z_{ii}}{N} \times 100$$

Where, N is the total number of proteins (instances) in the test set, and  $z_{ii}$  are the true positives.

## Results obtained on a test set

We first analyzed the fold-discriminatory potential of individual features (Table 2.1). Both

Features	One versus all			One versus one			Cramer and Singer		
	Q	C	$\Gamma$	Q	C	$\gamma$	Q	C	$\gamma$
AAC	<b>47.2</b>	2 <sup>1</sup>	2 <sup>1</sup>	44.9	2 <sup>2</sup>	2 <sup>1</sup>	46.7	2 <sup>2</sup>	2 <sup>1</sup>
DI	<b>49.1</b>	2 <sup>2</sup>	2 <sup>6</sup>	46.4	2 <sup>4</sup>	2 <sup>7</sup>	45.6	2 <sup>3</sup>	2 <sup>6</sup>
1GAP	<b>49.6</b>	2 <sup>1</sup>	2 <sup>7</sup>	49.3	2 <sup>3</sup>	2 <sup>7</sup>	47.2	2 <sup>3</sup>	2 <sup>7</sup>
SSS_AA	<u>52.8</u>	2 <sup>2</sup>	2 <sup>2</sup>	<b>64.4</b>	2 <sup>3</sup>	2 <sup>3</sup>	<u>59.6</u>	2 <sup>3</sup>	2 <sup>3</sup>
SSS_DI	<b>58</b>	2 <sup>1</sup>	2 <sup>7</sup>	57.3	2 <sup>3</sup>	2 <sup>6</sup>	54.6	2 <sup>3</sup>	2 <sup>10</sup>
SSS_GAP	<b>61.5</b>	2 <sup>2</sup>	2 <sup>7</sup>	58	2 <sup>3</sup>	2 <sup>6</sup>	56.2	2 <sup>3</sup>	2 <sup>11</sup>
SAS_AA	<b>57.8</b>	2 <sup>2</sup>	2 <sup>2</sup>	57.3	2 <sup>2</sup>	2 <sup>5</sup>	56.2	2 <sup>0</sup>	2 <sup>2</sup>
SAS_DI	<b>58.3</b>	2 <sup>2</sup>	2 <sup>6</sup>	54.4	2 <sup>4</sup>	2 <sup>10</sup>	58.6	2 <sup>1</sup>	2 <sup>6</sup>
SAS_GAP	<b>57.3</b>	2 <sup>3</sup>	2 <sup>7</sup>	54.9	2 <sup>3</sup>	2 <sup>6</sup>	55.4	2 <sup>0</sup>	2 <sup>7</sup>

Table 2.2: Best overall prediction accuracy (Q) and their corresponding C and  $\gamma$  values for individual features using three multi-class methods. Best accuracies across a row is shown as undefined, depicting the best feature for a particular multi-class method and the best accuracy across a row is shown as bold, depicting the best multi-class method for a particular feature.

sequence and structure-based features were individually used for learning the protein fold classifier. Prediction accuracies of each feature for all three multi-class methods were calculated and shown in Table 2.2. It can be seen that out of nine individual features used in this study, secondary structural state frequencies of amino acids (SSS\_AA) gives the best prediction accuracy for all three multi-class methods with overall accuracy of 64.4%.

This accuracy itself is higher than the best accuracy (62%) reported in literature so far. As

Comb	Descriptor	Constituent features	Dimensionality
Combo1	Structural frequencies of amino acids	SSS_AA + SAS_AA	120
Combo2	Structural frequencies of first order amino acid pairs	SSS_DI + SAS_DI	2400
Combo3	Structural frequencies of second order amino acid pairs	SSS_GAP + SAS_GAP	2400
Combo4	Secondary structural state frequencies of amino acids and first & second order amino acid pairs	SSS_AA + SSS_DI + SSS_GAP SAS_AA + SAS_DI + SAS_GAP	2460
Combo5	Solvent accessibility state frequencies of amino acids and first & second order amino acid pairs	SSS_AA + SAS_AA + SSS_DI + SAS_DI	2520
Combo6	Structural frequencies of amino acids and first order amino acid pairs	SSS_AA + SAS_AA + SSS_GAP + SAS_GAP	2520
Combo7	Structural frequencies of amino acids and second order amino acid pairs	SSS_DI + SAS_DI + SSS_GAP + SAS_GAP	4800
Combo8	Structural frequencies first and second order amino acid pairs	SSS_AA + SAS_AA + SSS_DI + SAS_DI + SSS_GAP + SAS_GAP	4920

shown in Table 2.2, all structural features perform better than the sequence-based features for all the three multi-class methods employed. The maximum overall accuracy achieved by any sequence-based feature is 49.6% for 1-gapped dipeptide composition (1GAP), which is less than the accuracy obtained by any structural features used in this study. This study, therefore, further



strengthens the fact that structural features have better discriminatory potential for protein fold classification than the sequence-based features.

We further used the combination of features to improve the accuracy of prediction. As structural features have better fold-discriminatory potential than the sequence-based features (Table 2.3), we used only combinations of the structural features. Different combinations of structural features were made in order to define and encapsulate overall

Table 2.4: Best overall prediction accuracy (Q) and their corresponding C and  $\gamma$  values for combination of features using three multi-class methods

*One versus all One versus one Cramer and Singer*

Features	Q	C	$\gamma$	Q	C	$\gamma$	Q	C	$\gamma$
Combo1	64.9	2 <sup>1</sup>	2 <sup>-3</sup>	<u>66</u>	2 <sup>2</sup>	2 <sup>-4</sup>	64.4	2 <sup>5</sup>	2 <sup>-3</sup>
Combo2	<b>61.7</b>	2 <sup>2</sup>	2 <sup>-9</sup>	55.4	2 <sup>3</sup>	2 <sup>-9</sup>	58.8	2 <sup>2</sup>	2 <sup>-10</sup>
Combo3	<b>64.1</b>	2 <sup>2</sup>	2 <sup>-8</sup>	58.8	2 <sup>3</sup>	2 <sup>-10</sup>	62.3	2 <sup>-3</sup>	2 <sup>-8</sup>
Combo4	<b>64.4</b>	2 <sup>2</sup>	2 <sup>-9</sup>	59.4	2 <sup>2</sup>	2 <sup>-9</sup>	60.7	2 <sup>2</sup>	2 <sup>-11</sup>
Combo5	<b>63.3</b>	2 <sup>2</sup>	2 <sup>-10</sup>	55.9	2 <sup>4</sup>	2 <sup>-11</sup>	62.8	2 <sup>-3</sup>	2 <sup>-11</sup>
Combo6	<b>63.1</b>	2 <sup>2</sup>	2 <sup>-9</sup>	58.3	2 <sup>3</sup>	2 <sup>-10</sup>	61.5	2 <sup>2</sup>	2 <sup>-10</sup>
Combo7	<b><u>66</u></b>	2 <sup>2</sup>	2 <sup>-11</sup>	61.2	2 <sup>3</sup>	2 <sup>-10</sup>	<u>65.7</u>	2 <sup>-3</sup>	2 <sup>-11</sup>
Combo8	<b><u>66</u></b>	2 <sup>2</sup>	2 <sup>-10</sup>	55.9	2 <sup>4</sup>	2 <sup>-11</sup>	64.6	2 <sup>-3</sup>	2 <sup>-11</sup>
Combo9	<b>65.4</b>	2 <sup>2</sup>	2 <sup>-10</sup>	57	2 <sup>4</sup>	2 <sup>-11</sup>	64.9	2 <sup>-3</sup>	2 <sup>-11</sup>

Best accuracies across a column is shown as underlined, depicting the best combo feature for a particular multi-class method and the best accuracy across a row is shown as bold, depicting the best multi-class method for a particular combo feature

structural features of amino acids and amino acid pairs. The best overall prediction accuracy for each combo feature is shown in Table 2.4. Combo1, Combo7 and Combo8 are the best feature combinations, which achieved maximum overall accuracy of 66% (shown as bold & underlined in Table 2.4). Combo1, which corresponds to structural properties (secondary structural state and solvent accessibility state frequencies) of amino acids, achieved an overall accuracy of

66% for one versus one multi-class method. Combo7, which corresponds to structural properties (secondary structural state and solvent accessibility state frequencies) of amino acids and second order amino acid pairs, achieved an overall accuracy of 66% for one versus all multi-class method. Combo8, which corresponds to structural properties (secondary structural state and solvent accessibility state frequencies) of first and second order amino acid pairs, achieved an overall accuracy of 66% for one versus all multi-class method. So the combination of features improves the accuracy of fold classification. Currently we are trying to increase the size of the training set by including close homologues in an effort to increase the accuracy further.

### 3 Analysis of microsatellites in prokaryotic genomes.

#### Objectives

Analysis of microsatellite frequencies, abundance and polymorphism in fully sequenced prokaryotic genomes.

#### Summary of the work done until the beginning of this reporting year

Cross-genome comparisons of *M. tuberculosis* H37Rv (MTH), *M. tuberculosis* CDC1551 (MTC) and *M. bovis* (MB) revealed tract length variations in some of the microsatellites. Furthermore, it was also discovered that tract length variations, at several loci, were associated frame-shifts in the coding regions. While some frame-shifts caused fission/fusion

of ORFs, some resulted in premature termination of ORFs and numerous others caused length changes in the ORFs.

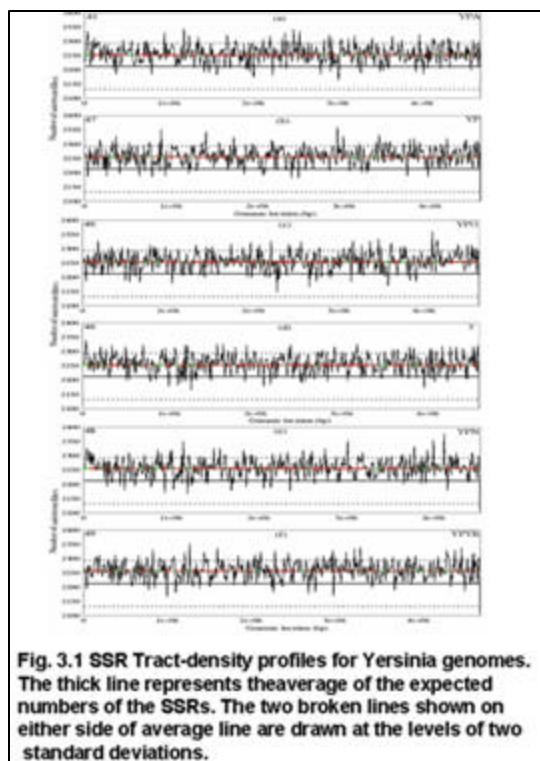
We also analysed *Neisseria* genomes for microsatellite distributions and polymorphisms. An analysis of these genomes revealed an even distribution of microsatellite tracts throughout these genomes. Upon comparing the equivalent microsatellites occurring in the coding regions of the three genomes we discovered nearly 120 microsatellite tracts showing polymorphism. Most (99 out of 119) of the polymorphic tracts are the mononucleotide tracts. In most of these cases microsatellite polymorphism led to frameshifts in the ORFs. The affected ORFs were annotated as periplasmic protein, transpoter, restriction endonuclease, adhesin protein (Maf), transcriptional regulator, lipoprotein, membrane protein, penicillin binding protein, bacteriocin resistance protein, methy-transferase, sigma factor, kinases, pilin glycosylation protein, hydratase, oxidoreductase, phage associated proteins and hypothetical protein.

#### **Details of progress in the current reporting year (April 1, 2006 – March 31, 2007)**

**(a) Studies on *Yersinia* genomes:** *Yersinia pestis* is a gram negative bacterium which seems to have diverged from the enteric pathogen, *Yersinia pseudotuberculosis* very recently. Whereas *Yersinia pseudotuberculosis* is a soil and water born pathogen, *Yersinia pestis* is a dangerous pathogen and in humans, this pathogen causes bubonic, septicemic and pneumonic plagues. The bacterium also exhibits wide range of hosts such as rodents, flea etc. Flea gut is necessary for the colonization of these bacteria. However they also can survive in different hosts having different environments. How they survive in different host environments and what genetic factors are responsible for their variable host adaptations, has not been completely understood. As SSRs are known for imparting variability in pathogenic

genomes, it was tempting to carryout an in-depth analysis of their distribution, enrichment and polymorphism in all the known genomes of this pathogen.

The six genomes analysed in this study are as follows: *Yersinia pestis* CO92 (YPO), *Yersinia pestis* KIM (Y), *Yersinia pestis* 91001 (YP), *Yersinia pestis* Antiqua (YPA), *Yersinia pestis* Nepal516 (YPN) and *Yersinia pseudotuberculosis* (YPTB). Our analysis revealed that every genome harbors as many as one million SSRs constituted by 68% of the total nucleotides present in the genome. The



**Fig. 3.1** SSR Tract-density profiles for *Yersinia* genomes. The thick line represents the average of the expected numbers of the SSRs. The two broken lines shown on either side of average line are drawn at the levels of two standard deviations.

observed tract density is 2250 SSRs per 10kb which is higher than the expected tract density of 2212 indicating general enrichment of SSRs in the genomes. SSR density profiles of the *Yersinia* genomes (Fig. 3.1) further revealed some regions that are significantly enriched by SSRs. The numbers of such SSR enriched regions vary from 41 in YPA to 49 in YPTB. These regions by virtue of being abundant in SSRs can be called as hypervariable regions and are expected to be more prone for mutations as compared to the rest of the genomes. Very interestingly many of these regions are conserved across the *Yersinia* genomes. These regions harbour ORFs whose functions have been assigned as ABC type transporter protein, membrane protein, siderophores, transferases and different types of kinases, haemolysin, heat shock protein and hypothetical proteins.

Table 3.1 gives the number of different types of SSR types observed and expected in all the six genomes. As also observed in the other bacterial genomes (for example, mycobacterial genomes) mononucleotide tracts are the most abundant type in the *Yersinia* genomes with their numbers varying in the range of 196-252 tracts per kb of genome and hexas are the least abundant ranging from 0.4 tracts per kb. The number of SSR tracts seems to be inversely correlated to their lengths, in other words, short arrays are more abundant than the long arrays. The arrays longer than 12bp are nearly absent in all the genomes indicating general scarcity of long tracts in the *Yersinia* genomes. The infrequently occurring long tracts are mostly populated in the non-coding regions.

Mono, tri, tetra, penta and hexa SSRs are found more number of times than expected indicating their enrichment whereas di tracts are underrepresented. Generally prokaryotic genomes show enrichment of only tri and hexa SSRs and underrepresentation of mono, di, tetra and penta SSRs. It is believed that selection against frameshifts in coding regions is the reason for underrepresentation of mono, tetra and penta tracts. In the light of this it is interesting to note that *Yersinia* genomes are distinct as they show unexpected enrichment of mono, tetra and penta tracts indicating their potential preference for frameshift mutations in their coding regions.

Although the genomes are slightly AT rich, selective bias towards enrichment of certain sequences is seen. For example, among the mono tracts all the four sequences viz., A, T, G and C tracts are enriched and among the di tracts only GC repeat sequence is enriched.

Furthermore cross-genome comparisons of the genomes yielded 391 SSRs showing length variation (referred to as PSSRs) across the six genomes of which the 193 were found in the coding regions. Of 391 SSRs 183 were showing length variation between YPTB and YP genomes. Therefore they can be considered as the examples of inter-species SSR polymorphism. The remaining examples, 208 in total, can be considered as the examples of intra-species polymorphism



RT	YFN					YPO					Y					YFN					YPTB						
	Total	PIL 30	PNC 16	RPK 50	RPK 50	Total	PIL 55	PNC 19	RPK 50	RPK 50	Total	PIL 51	PNC 16	RPK 51	RPK 51	Total	PIL 42	PNC 16	RPK 49	RPK 49	Total	PIL 54	PNC 16	RPK 49	RPK 49		
Micro Nucleotide Repeat																											
2	600329 + (187183)	38	15	50	129	503652 + (102544)	50	17	50	129	584181 + (158424)	51	17	51	129	602916 + (174296)	43	15	48	129	613959 + (162514)	54	16	49	129		
3	164337 + (116000)	26	17	50	35	161229 + (117240)	50	19	50	35	161269 + (117440)	51	19	51	35	150321 + (155181)	44	17	48	35	166615 + (112298)	55	18	49	35		
4	49658 + (41265)	39	19	51	19	47786 + (40362)	50	22	50	19	48473 + (40421)	51	22	51	19	47299 + (39950)	44	20	49	19	49224 + (41862)	54	21	49	19		
5	16150 + (19483)	38	25	50	3	16069 + (10258)	49	27	50	3	16069 + (10247)	50	27	51	3	15603 + (10104)	43	25	48	3	16070 + (10670)	53	26	48	3		
6	5137 + (2646)	41	30	50	1	5054 + (2598)	50	31	50	1	5051 + (2587)	49	31	51	1	4902 + (2552)	42	30	48	1	5200 + (2673)	53	32	50	1		
7	1539 + (866)	39	37	50	0.33	1524 + (853)	50	38	49	0.33	1529 + (855)	50	38	51	0.33	1514 + (847)	44	37	44	0.33	1611 + (876)	54	37	50	0.33		
≥2	423 + (227)	38	44	49	0.09	434 + (223)	46	47	51	0.09	422 + (222)	51	46	54	0.09	408 + (220)	45	43	43	0.09	472 + (230)	51	44	48	0.09		
BI Nucleotide Repeat																											
2	126660 + (149130)	38	17	50	26	124166 + (147500)	50	19	50	26	122702 + (145036)	50	20	51	26	120946 + (143814)	43	17	48	26	126659 + (150466)	54	19	49	26		
3	5312 + (9374)	26	25	50	1.3	6201 + (9167)	50	26	49	1.3	6187 + (9177)	49	27	51	1.3	6073 + (9052)	43	25	47	1.3	6415 + (9487)	54	26	49	1.3		
4	394 + (591)	44	29	46	0.07	399 + (577)	53	33	50	0.07	395 + (576)	47	31	55	0.07	347 + (571)	43	33	52	0.07	348 + (599)	54	35	50	0.07		
5	20 + (37)	25	45	36	0.01	19 + (37)	68	58	53	0.01	18 + (37)	44	50	44	0.01	18 + (36)	32	42	37	0.01	24 + (38)	50	46	31	0.01		
≥5	3 + (3)	0	67	100	0.001	4 + (3)	0	75	100	0.001	4 + (3)	0	75	100	0.001	4 + (3)	25	75	0	0.001	3 + (3)	33	66	100	0.001		
TI Nucleotide Repeat																											
2	83449 + (20002)	28	15	50	14	82195 + (20871)	50	18	50	14	82035 + (20490)	50	18	51	14	81059 + (20281)	43	15	47	14	84039 + (20477)	54	17	49	14		
3	2019 + (796)	40	12	50	0.44	1966 + (772)	53	15	50	0.44	1996 + (773)	49	15	49	0.44	1959 + (762)	42	13	46	0.44	2108 + (797)	57	14	50	0.44		
4	62 + (13)	29	5	52	0.01	62 + (12)	41	8	47	0.01	61 + (12)	56	3	42	0.01	60 + (12)	42	3	40	0.01	65 + (13)	43	6	52	0.01		
≥4	3 + (1)	0	0	38	0.001	4 + (1)	25	0	0	0.001	4 + (1)	50	0	67	0.001	4 + (1)	50	0	0	0.001	6 + (1)	33	0	60	0.001		
Tera Nucleotide Repeat																											
2	13737 + (12826)	38	22	50	3	13390 + (12562)	50	25	50	3	13417 + (12965)	50	24	51	3	13145 + (12382)	43	23	49	3	13936 + (12999)	54	24	49	3		
3	98 + (51)	45	39	50	0.01	93 + (50)	58	36	52	0.01	94 + (50)	41	36	50	0.01	94 + (49)	43	40	42	0.01	98 + (51)	54	31	51	0.01		
≥3	2 + (1)	50	100	0	0.001	2 + (1)	100	100	0	0.001	3 + (1)	67	67	100	0.001	3 + (1)	66	100	0	0.001	5 + (1)	40	100	0	0.001		
Penta Nucleotide Repeat																											
2	3694 + (2445)	38	21	51	0.8	3515 + (2371)	49	23	51	0.8	3526 + (2377)	50	23	52	0.8	3436 + (2320)	43	22	47	0.8	3662 + (2465)	53	23	50	0.8		
≥2	3 + (6)	33	33	50	0.001	3 + (6)	0	33	0	0.001	4 + (6)	25	50	50	0.001	5 + (7)	60	60	50	0.01	10 + (4)	60	70	100	0.002		
Hexa Nucleotide Repeat																											
2	1723 + (852)	36	14	53	0.4	1690 + (834)	48	17	50	0.4	1679 + (836)	48	17	50	0.4	1671 + (822)	42	14	44	0.4	1781 + (862)	5	16	48	0.4		
≥2	12 + (13)	42	67	75	0.002	10 + (13)	50	56	25	0.002	6 + (13)	33	50	75	0.002	7 + (13)	14	57	67	0.01	14 + (17)	43	58	20	0.002		

Table 3.1 - Observed and expected numbers of SSRs found in *Yersenia* genomes (*Y.pestis Antiqua* (YPA); *Y.pestis biovar Mediaevalis* (YP); *Y.pestis CO92* (YPO); *Y.pestis KIM* (Y); *Y.pestis Nepal516* (YPN); *Y.pseudotuberculosis* IP32953 (YPTB)). The +/- sign indicates the over-representation and under-representation respectively. RT = Repeat times, PIL = Percentage in leading strand, PNC = Percentage in Non-coding region, PIR = Percentage in Replication direction, RPK = Repeat per Kb.

Of the PSSRs a large majority (291 out of 391 i.e., 74%) are mono nucleotide tracts of which about 60% are A/T tracts. In non-coding regions 152 (77%) mono nucleotide PSSR are present and out of which 109 (72%) are of A|T type whereas 43 (28%) are of G|C type. In coding region the number of mono nucleotide PSSR is 139 (72%) and out of which 64 (46%) are of A|T type and 75 (54%) are of G|C type. With regard to the extent of length variation, almost every PSSR (369) has undergone only one repeat copy number variation, in other words, indel of only one repeat unit.



**(b) Development of a computational tool for identification and extraction of imperfect SSRs from genome sequences.**

In all the studies conducted so far, we focused only the perfect SSRs. A look at any genome reveals a number of SSRs with imperfections (those harbouring substitutions and indels as compared to their perfect counter-parts). Since analysis is preceded by identification and extraction of imperfect SSRs from fully sequenced genomes and that our survey of existing tools revealed several lacunae in them, we resorted to develop a new tool called Imperfect Microsatellite Extractor (IMEx). IMEx uses a simple yet novel algorithm to find both perfect as well as imperfect microsatellites from genome sequences. A representative flow-chart is shown in Fig. 3.2.

**IMEx Algorithm:**

We define a sequence at a given locus as microsatellite if that sequence can be expressed as a tandem repeat of a motif of 1-6 bp size. The repeating motif can at every iteration harbor upto 'k' number of point mutations (substitutions or indels of nucleotides). IMEx algorithm uses this definition and works as a two-step procedure: (a) Identification of microsatellite nucleation sites which are nothing but the loci where a repeat motif is repeated twice either tandemly (type I nucleation site) or after certain intervening nucleotides (type II nucleation sites). in both cases the repeat motif is without any imperfection (i.e., k=0) and (b) Extension of the nucleation sites on both sides in the steps of the repeating motif (with imperfections less than 'k' value) as long as one of the termination criteria is satisfied: (a) The number of imperfections (inclusive of substitutions

and a maximum of one indel) between the consensus and the perfect repeat motif is more than the limit (denoted by “*k*” parameter set by the user) and (b) The percentage of imperfection is more than the limit set by the user (denoted by “*p*” parameter). The percentage imperfection is calculated as follows:

$$\% \text{ Imperfection} = \left( \frac{\text{no. of point mutations in the observed tract}}{\text{total no. of bases in the equivalent perfect tract}} \right) * 100$$

The user can set a value for ‘*k*’ between 0 and *m* where *m*=repeat motif size(consensus size). Once the termination criteria are satisfied, only those candidate microsatellites that are more than the minimum repeat number of that repeat size set by the user (denoted by “*n*” parameter) are reported.

Input to the program consists of a sequence file and the following parameters: (a) No. of edit operations/motif (*k*); (b) percentage imperfection for the entire tract (*p*); (c) minimum repeat number (*n*); (d) coding information file. The program generates two files, one of which gives a summary table describing the microsatellite tracts along with their information that includes tract size, no. of iterations, % imperfection, nucleotide composition and coding/non-coding information. The second file contains the alignment of each repeat with its consensus sequence. These two files are produced both in html form as well as in text formats. The text files produced can be downloaded and used for further studies. In html outputs, the files are linked so that on clicking a repeat will display its corresponding alignment in a separate html page. An option for designing primers for the microsatellite regions found has also been provided.

IMEx is developed in standard “C” language. A web-server called “IMEx-web” has been created using the IMEx program at the back-end. IMExweb is a web-based, fully automated server which provides a list of known genomes belonging to prokaryotes (430 genomes) and viruses (1809 genomes) from which the user can select a genome of his interest for extraction of microsatellites satisfying his criteria of imperfection, SSR type, SSR motif sequence etc. The web-server as well as the IMEx program can be accessed freely from [www.cdfd.org.in/imex](http://www.cdfd.org.in/imex) or <http://203.197.254.154/IMEX/>.

#### ***Publications:***

- 1 Achary MS, Reddy ABM, Chakrabarti S, Panicker SG, Mandal AK, Ahmed N, Balasubramanian D, Hasnain SE and Nagarajaram HA (2006) Disease-causing mutations in proteins: structural analysis of the CYP1B1 mutations causing primary congenital glaucoma in humans. *Biophysical Journal* 91:4329-4339.

2. Sreenu VB, Kumar P, Nagaraju J and Nagarajaram HA (2007) Simple sequence repeats in mycobacterial genomes ***Journal of Biosciences*** 32:1-15.
3. Suresh BM and Nagarajaram HA (2007) IMEx: Imperfect Microsatellite Extractor. ***Bioinformatics*** (InPress).

## Laboratory of Molecular and Cellular Biology

<b>Principal Investigator</b>	Seyed E Hasnain*	
<b>Ph D Students</b>	Krishnaveni M	Senior Research Fellow
	Sheeba Rasheedi	Senior Research Fellow
	Aisha Farhana	Senior Research Fellow
	Smanla Tundup	Senior Research Fellow
	Nasreena Bashir	Senior Research Fellow
	Yusuf Akther	Senior Research Fellow
	Sandeep Kumar	Senior Research Fellow
<b>Other members</b>	Bandi Aruna	Project Associate
	Leena Bashyam	Project Assistant
	Niteen Pathak	Technical officer
<b>Collaborators</b>	Niyaz Ahmad	CDFD, Hyderabad
	K V A Ramaiah,	University of Hyderabad
	Nasreen Z Ehtesham	NIN, Hyderabad
	Sangita Mukhopadhyay	CDFD, Hyderabad
	Shekar C Mande	CDFD, Hyderabad
	Anil K Tyagi	University of Delhi

\* Present Address : University of Hyderabad, Hyderabad

Our laboratory focuses on (A) Understanding of the transcriptional apparatus during baculovirus infection of *in vitro* cultured insect cells and the ensuing apoptotic response of these cells (B) The other major focus of our group is molecular epidemiology, dissemination dynamics and infection biology of infectious pathogens, notably, *Mycobacterium tuberculosis*.

### A. Dissecting the baculovirus transcription regulatory apparatus

#### **Objectives:**

1. To study the role of *cis*-acting elements, trans-acting factors and the RNA polymerase involved in hyperactivation of transcription from the very late polyhedrin gene promoter of *Autographa californica* nuclear polyhedrosis virus (AdMNPV).
2. Molecular cross talks between Baculovirus P35 with host apoptotic machinery
3. Cloning and characterization of S9 P53 homolog.

#### **Summary of work done until the beginning of this reporting year**

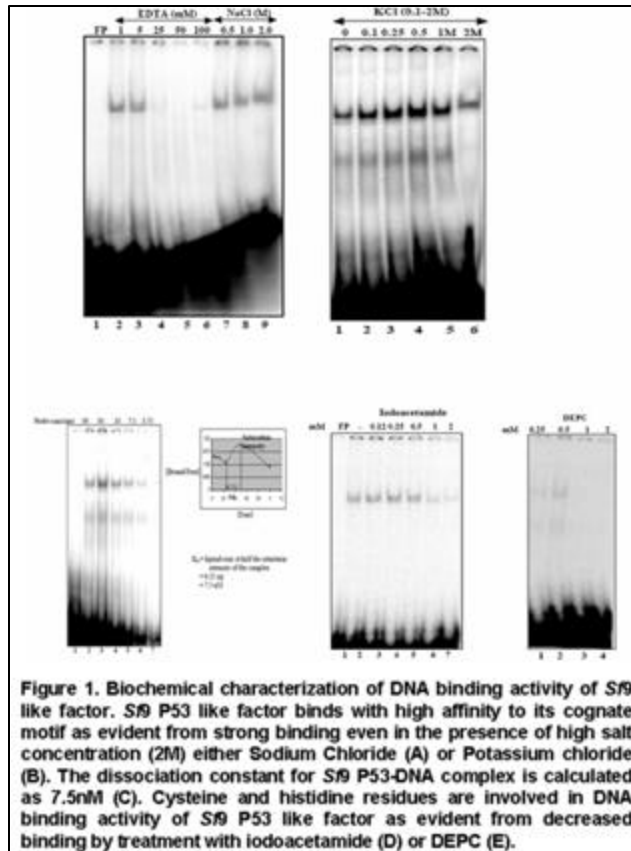
The host RNA polymerase transcribes early genes of the baculovirus whereas the polymerase utilized in the late and very late phase is a viral encoded  $\alpha$ -amanitin resistant RNA polymerase. This polymerase recognizes promoters with a TAAG motif. Promoter

recognition is an integral characteristic of the purified enzyme. This viral RNA polymerase has been purified and four of its subunits have been identified till date. The four subunits are LEF-4, LEF-8, LEF-9 and P47. LEF-8 and LEF-9 are hypothesized to constitute the catalytic core of RNA polymerase since they possess amino acid sequence motifs homologous to other polymerases. For the *in-vitro* reconstitution of viral RNA polymerase activity, purified forms of its subunits are necessary. The presence of purified subunits will also help in understanding the interaction between the subunits and their interaction with other factors and the promoter. LEF4 and P47 were purified as GST-tagged proteins to homogeneity. We find that through GST-pull down method LEF4 interacts with LEF9 and also P47 interacts with LEF9. These interactions involve different domains since there is no competition. GST pull down assays also suggests that there may be no direct interaction between LEF4 and P47.

**Transcriptional cross talk between host and viral factors during apoptosis:** P53 protein, the central molecule of the apoptosis pathway, is often targeted and inactivated by several animal viruses as a prelude to productive infection. Of late, P53 homologues have been identified from different invertebrates including *Drosophila*, *C. elegans*, Squid and Clams. We have reported the identification of a P53 like protein in *Spodoptera frugiperda* (Sf9) insect cells, which is activated during oxidative stress, caused by exposure to UV-B or H<sub>2</sub>O<sub>2</sub> and binds to P53 consensus DNA binding motifs as well as other P53 sequence motifs. Sf9 P53-like factor is apparently similar to murine and *Drosophila* P53 in terms of molecular size, which is around 50-60kDa as was evident from UV cross-linking and SouthWestern analyses. Expression of baculovirus P35 appears to down regulate the activity of Sf9 P53-like factor, which was evident from electrophoretic mobility shift assays. The binding of P53 to its cognate motifs is inhibited in the presence of recombinant P35 and this inhibition is specific to P53 and not AP-1 or NF $\kappa$ B which are also down regulated upon baculovirus infection. The rP35 mediated DNA binding inhibition is limited to the transcriptionally active tetrameric form of P53. The P53 target gene expression is also reversed in the presence of P35 protein. *In vitro* transcription confirms the molecular cross talks between the viral P35 and the host P53-like transcription factor. The identification of P53 dependent apoptotic pathway in Sf cells adds a new facet to our understanding of the mechanisms of baculovirus P35 mediated abrogation of apoptosis under oxidative stress.

**Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)**

Having shown *Spodoptera frugiperda* insect cells as a model for oxidative stress, we investigated the presence of P53 dependent apoptotic pathway in Sf9 cells. In the present work, we report the biochemical characterization of the DNA binding activity of Sf9 P53 like factor, which is activated upon oxidative stress in Sf9.



**Sf9 P53 like factor binds to cognate motifs with high affinity:**

To check the resistance of the Sf9 P53-DNA complex to salt (NaCl) induced dissociation, EMSA with Sf9 nuclear extract was carried out in the presence of increasing concentration of NaCl (0.5M-2M) (Figure 1A, lanes 7-9). The Sf9 P53 complex remains unaffected and is steady even up to a concentration of 2M NaCl demonstrating that ionic interactions do not have a major role in the complex formation. Similarly when EMSA was carried out in the presence of increasing concentration of KCl up to 2M (Figure 1B, lanes 1-6), the binding was stable and in fact increased at lower concentration (up

to 0.5M). To determine the dissociation constant of Sf9 P53 DNA binding to consensus P53, EMSA was carried out with 5µg of Sf9 nuclear extract and decreasing concentration of probe (1-0.625ng) (Figure 1C). The dissociation constant of a ligand is defined as half the concentration at which it reaches saturation. The complex intensity decreases as a function of probe dilution (Figure 1C, lanes 2-6). From the graph of probe vs P53- DNA complex intensity (Figure 1C), it is clear that complex saturates at 0.5ng equivalent to 7.5nmoles. Hence the  $K_d$  for Sf9 P53-DNA binding is 7.5nmoles. These assays demonstrate that Sf9 P53 like factor binds with high affinity to its cognate motif.

**Sf9 P53 like factor requires divalent metal ions for DNA binding:** To look into the requirement of a divalent metal ion in the DNA binding activity of Sf9 P53, EMSA for Sf9 P53 was carried out using 5µg of Sf9 nuclear extract in the presence of increasing concentration of divalent metal ion chelator, EDTA up to 250mM (Figure 1A, lanes 2-6). At concentration of 25mM EDTA, the binding of Sf9 P53 to cognate motif was completely abolished, demonstrating that divalent ions are essential for DNA binding activity of Sf9 P53 like factor.

**Cysteine and histidine residues are involved in DNA binding activity of Sf9 P53 like factor:** To demonstrate the role of cysteine and histidine residues in the DNA binding activity of Sf9 P53, EMSA was carried out with 5µg of Sf9 nuclear extract by modifying either cysteine (Iodoacetamide, Figure 1D) or histidine (DEPC, Figure 1E). Iodoacetamide causes

acetylation of sulfhydryl groups of cysteine residues thus blocking the availability of sulfhydryl group for DNA protein interaction. The binding of *Sf* P53 like factor to its cognate motif decreases (Figure 1D lanes 2 to 6) as a function of concentration of iodoacetamide (0.2mM – 2mM). A decrease in the DNA protein complex is evident in the presence of 1mM (Figure 1D, lane 6), 2mM (Figure 1D, lane 7) Iodoacetamide, demonstrating the role of cysteine residues in the DNA protein interaction. Similarly by treatment of *Sf* nuclear extract with DEPC (0.25-2mM), which modifies the histidine residues by alkylation, the *Sf* P53 interaction with DNA is abolished (Figure 1E) even at minimal concentration of DEPC used, i.e. 0.25mM (Figure 1E, lane 1), also pointing the involvement of histidine residues in *Sf* P53-DNA interaction. Taken together, both cysteine and histidine residues of *Sf* P53 are involved in interaction with DNA.

In conclusion, we report the biochemical characterization of *Sf* P53 like factor in terms of its DNA binding activity, which appears identical to the reported vertebrate P53. The molecular size, cognate element recognition, divalent metal ion requirement, the involvement of cysteine and histidine residues and the dissociation constant suggest the *Sf* P53 like factor to be the authentic *Sf* P53 homolog. Additionally, the *Sf* P53 like factor has been purified by conventional protein purification strategies although in minute quantity for the N-terminal sequencing of the protein. Cloning of the gene corresponding to this factor is underway which will conclusively demonstrate the P53 homolog in *Sf* insect cells.

## **B. Molecular pathogenesis and infection biology of *Mycobacterium tuberculosis***

### **Objectives:**

1. Identification and characterization of genes involved in siderophore mediated iron uptake in *Mycobacterium tuberculosis* and their role in virulence
2. Characterization of genes involved in the virulence and persistence of *M. tuberculosis*
3. Identification and characterization of novel antigens of *M. tuberculosis* vis-a vis their immune response

### **Summary of work done until the beginning of this reporting year**

The success of *Mycobacterium tuberculosis* (*M. tb*) as an efficient pathogen lies in its ability to persist within humans for long periods in a clinically latent state, in hostile intracellular environment including microaerophilic and extreme acidic conditions, without causing any disease symptoms. Fumarate Nitrate Reductase Regulator (FNR) family proteins are global regulators of low oxygen conditions and regulate target genes carrying the FNR box elements. By relative positional entropy method using database Predict Regulon Server (<http://210.212.212.6/prindex.html>) we predicted the presence of 60 FNR boxes across the complete (3919 ORFs) *M. tb* genome. We also pooled FNR box elements present in homologues of *M. tb* in other organisms like *E. coli*, *Pseudomonas*, *Bacillus*, *Paracoccus*, *Rhodospirillum* etc. Many of the predicted ORFs were previously reported to be induced



under hypoxia in various microarray data. To validate our computer prediction, *M. tb* ORF Rv3676, a putative FNR regulator, was selected and subjected to *in silico* analysis to theoretically confirm its likely role as a member of the FNR family of transcription regulators. Recombinant protein coded by ORF Rv3676 was finally demonstrated to bind to some of the predicted FNR box elements using electrophoretic mobility shift assay. Our results will additionally aid in our understanding of gene regulation in hypoxia and the role of such gene in *M. tb* pathogenesis.

About 10% of the coding capacity of the *Mycobacterium tuberculosis* genome is devoted to the PE/PPE family of genes scattered throughout the genome. We have identified 28 PE/PPE operons which are organized within the *M. tb* genome in such a way that most PE members are upstream to PPE members. One example of such a gene arrangement is the PPE gene Rv2430c, earlier shown by us to code for a highly antigenic protein eliciting strong B-cell responses in TB patients [Choudhary, R.K., Mukhopadhyay, S., Chakhaiyar, P., Sharma, N., Murthy, K.J.R., Katoch V.M. and Hasnain, S.E. (2003) PPE antigen Rv2430c of *Mycobacterium tuberculosis* induces a strong B cell response. *Infect. Immun.* 71, 6338-6343], situated downstream to PE gene Rv2431c. Rv2431c and Rv2430c are transcribed as an operon. Expression of either rRv2431c or rRv2430c alone in *E. coli* limited their localization to the inclusion bodies. However, when they were co-expressed, both the proteins appeared in the soluble fraction. These two proteins interact with each other and form oligomers when alone, however, when present together they exist as heteromer.

**Details of progress made in the current reporting year (April 1, 2006 - March 31, 2007)**

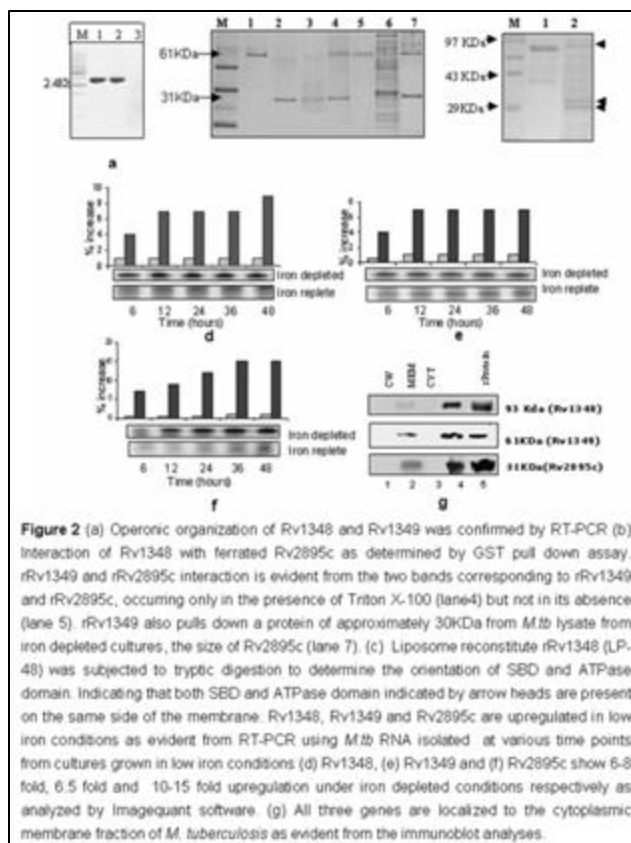
**1. Trafficking of iron from host to pathogen involves interplay between *Mycobacterium tuberculosis* exporter protein Rv1348 and importer Rv1349 and Rv2895c proteins**

Despite a reasonable understanding of mycobacterial iron uptake system and their role as virulence determinants, mechanistic and biochemical principles involved in this process are not well investigated. We report the role of three iron-regulated proteins namely, Rv1348, Rv1349 and Rv2895c as modulators of intracellular siderophore concentrations and consequent iron uptake. *In silico* analysis suggests that Rv1348 and Rv1349 have regions corresponding to ATP binding cassettes with typical Walker A and Walker B signature motifs. Whereas, Rv2895c and the N-terminal region of Rv1348 showed regions of homology to the siderophore interacting proteins. Reverse transcriptase PCR from total mRNA of *M. tuberculosis* reveals that Rv1348 and Rv1349 are transcribed as a part of same operon and that all the three genes are upregulated in low iron conditions. Also, their membrane localization is indicative of their likely role in membrane transport. The N-terminal siderophore binding domain within Rv1348 selectively binds to non-ferrated siderophores, whereas, rRv2895c shows comparatively higher affinity towards ferrated siderophores

(180 $\mu$ M) than non-ferrated (127 $\mu$ M) siderophore. Besides, our pull down assay points to an interaction between rRv1348 and Rv2895c, possibly involving the permease domain. Liposomal reconstitution experiments demonstrate that recombinant Rv1348 is involved in the siderophore export whereas the two-component Rv1349-Rv2895c system serves as an importer of ferrated siderophore, both of which require ATP hydrolysis. Knockout of Rv1348 homologue *M. smegmatis*, *msmeg6554* showed phenotypic changes suggesting disruption of siderophore export which could be restored by complementation by Rv1348. Our data demonstrate, for the first time, the functional characterization of Rv1348 and Rv1349-Rv2895c as an exporter-importer system functioning together to maintain the fine balance of siderophores and thus iron inside the cells.

Computational analyses predict Rv1348, Rv1349 as transporters and Rv2895c as siderophore binding protein: Rv1348 and Rv1349 are operonic: The sequence homology and the motif scanning of proteins coded by *Rv1348* and *Rv1349* showed the presence of the nucleotide binding, Walker A and B and ABC signature domains, besides the six and five membrane spanning segments, respectively (Bribant et al., 2000). Rv1348 sequence also

harbors an additional N terminal domain with similarity to the periplasmic substrate-binding proteins of other ABC transporters, more specifically with those involved in siderophore uptake. The siderophore interacting motif (SIM) is present in Rv2895c (Schubert et al., 1999). The sequence analysis of the two Rv1348 and Rv1349 as well as RT-PCR shows that Rv1348 and Rv1349 are operonic as evident from amplification of 2.4Kb fragment (Figure 2A). To determine the orientation of SBD and ATPase domain tryptic digestion of liposome reconstituted eRv1348 (LP-48) was carried out which indicate that the two domains the N-terminal SBD and the C-terminal ATPase domain



are on the same side of the membrane (Figure 2C).

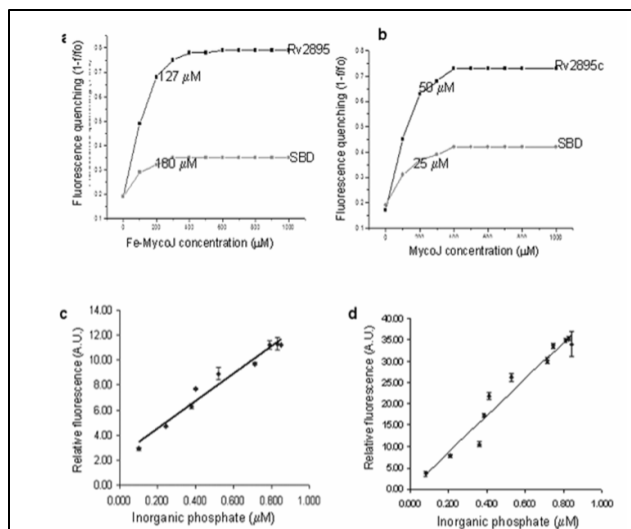
**Immunoblotting and RT-PCR indicates that all three proteins are membrane localized and upregulated in low iron conditions:** RT-PCR using *M. tb* RNA from control and iron

depleted cultures and immunoblot analyses with antibodies raised against individual proteins indicated that the proteins are cytoplasmic membrane localized. In iron replete conditions the level of expression of the three proteins is basal, as is evident from faint bands in RT-PCR (Figure 1D,E,F lower panel) and immunoblot (Figure 2G, lane 2). The expression is however enhanced many folds as the iron depletion stress increases (Figure 2D, E, F upper panel, Figure 2G, lane 4), suggesting their role in iron uptake and supporting their function in membrane specific siderophore transport.

Fluorimetric analyses suggests higher affinity of SBD to ferrated siderophores whereas Rv2895c shows higher affinity to non-ferrated siderophores; To address the role of SBD and Rv2895c in siderophore export or import, the substrate specificities and the dissociation constant ( $K_D$ ) of the recombinant proteins were analyzed fluorimetrically.  $K_D$  for SBD and Rv2895c gave a value of  $25.5 \mu\text{M}$  for non ferrated siderophore and  $50 \mu\text{M}$  for ferrated siderophore, pointing that SBD selectively binds to non ferrated siderophores (Figure 2B). However, Rv2895c shows relatively higher binding affinity to ferrated ( $180 \mu\text{M}$ ) than non ferrated ( $127 \mu\text{M}$ ) siderophores (Figure 3A). This indicates that Rv2895c is primarily involved in import of ferrated siderophores *viz. a viz.*, helps in facilitating export of non ferrated

siderophores through a ligand exchange mechanism.

Pull down assay indicates that Rv1349 interacts with Rv2895c through the permease domain: Rv2895c, which has only siderophore interacting domain serves as a periplasmic binding domain of Rv1349, was indicated by their interaction between the rRv1349-GST with FeMycoJ bound rRv2895c by GST pull down assay. This is evident from the Figure 1B, lane 4 and lane 7 by the presence of band corresponding to Rv2895c upon rRv1349-GST pull down confirming the interaction between the two proteins. The interaction was abrogated in the absence of TritonX-100 in the buffer

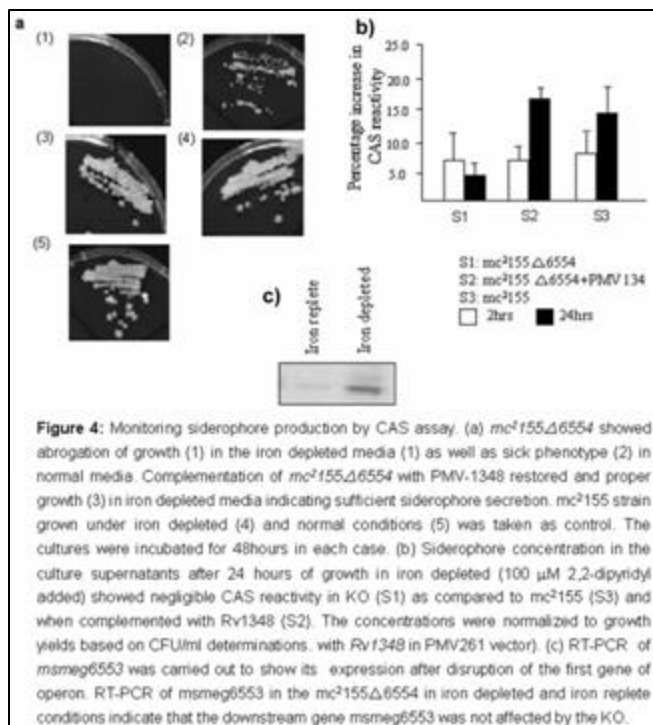


**Figure 3:** Binding affinity of Rv2895c and SBD to ferrated and non ferrated siderophores : The fluorescence quenching of Rv2895c and SBD proteins upon binding of the ligands FeMycoJ (a) and MycoJ (b) at various concentrations was monitored. The graph shows the binding of one representative analysis. Dissociation constant representing the half maximum of the emission quenching saturation was calculated as mean of three independent binding analyses.  $K_D$  values for substrate binding calculated using equation  $1-f/f_0$ , are listed. Liposome based siderophore export and import assay monitored in terms of relative increase in fluorescence as a function of ATPase activity. (c) In LP-48 liposomes, MycoJ export was monitored in the supernatant. (d) In LP-49 liposomes, the transport was initiated by incubating the liposomes with Fe-MycoJ bound Rv2895c. Export and import were analyzed by taking aliquots after every 15min for two hours. Fluorescence intensity was monitored to estimate the concentration of FeMycoJ and inorganic phosphate estimation indicates the intraliposomal ATP hydrolysis.

suggesting that the two proteins interact via the transmembrane permease domain (Figure 2B, lane 5).

***In vitro* and *in vivo* studies reveal that Rv1348 is involved in the export of non-ferrated siderophores whereas Rv1349-Rv2895c mediates the import of ferrated siderophores:** Liposome based transport assays showed a time dependent increase in the fluorescence of the LP-48 supernatant indicating the release of MycoJ from the liposome. This is accompanied by a concomitant increase in intraliposomal ATPase activity as evident from the increase in inorganic phosphate release, suggesting an energy driven export of Myco-J by Rv1348 (Figure 3C). This correlates well with the presence of both the domains on the same side of membrane as shown by tryptic digestion of LP-48 (Figure 2C). However, in case of LP-49, Fe-MycoJ internalization occurred only upon incubation with Rv2895c bound form indicating the import property of Rv1349-Rv2895c combination (Figure 3D). A linear relationship between export (LP-48) or import (LP-49) of Fe-MycoJ to an increase in intraliposomal ATPase activity was observed. The release of inorganic phosphate reflects the ATPase activity of respective proteins (Figure 3C,D). These liposomal experiments indicate that while rRv1348 is involved in siderophore export, a two component rRv139-rRv2895c system serves as an exporter of ferrated siderophores.

*M. smegmatis* KO for Rv1348 homologue renders it incapable of siderophore uptake, that could be restored by Rv1348 complementation: To convincingly demonstrate the involvement of Rv1348 in siderophore uptake *invivo*, *M. smegmatis* KO for Rv1348



homologue *msmeg6554* (*mc*<sup>2</sup>Δ6554) was constructed. The downstream gene *msmeg6553* (homologous to *M. tuberculosis* Rv1349, expected to be involved in siderophore uptake, was analyzed by RT-PCR to show that there is no polar effect. The presence of the band corresponding to *msmeg6553* transcription in *mc*<sup>2</sup>Δ6554 strain in low iron conditions (Figure 4C) indicates that the disruption of the upstream *msmeg6554* has no effect on the expression of the downstream gene, *msmeg6553*.

The KO strain, *mc*<sup>2</sup>Δ6554, shows a

sick phenotype with slow growth (Figure 4A.1,2) in iron rich agar plates and low CAS reactivity when grown in iron depleted media (Figure 4B.S1) indicating inability of the strain to secrete sufficient concentration of siderophores into the media. This abrogation of CAS reactivity is consequent to the disruption of the *msmeg6554* gene. The supernatant from *mc<sup>2</sup>Δ6554* cultures, which showed reduction of blue color quenching signifying low production of CAS reactive substance after 24 hours of growth (Figure 4B.S1). However, complementation of *mc<sup>2</sup>155Δ6554* with plasmid harboring Rv1348 restored normal growth of the strain in low iron conditions (Figure 4A) as well as increase in the concentration of CAS reactive substance in the culture supernatant after 24 hours of growth (Figure 4B.S2) comparable to *mc<sup>2</sup>155*. This suggests that *M. tuberculosis* Rv1348 could be involved in the release of siderophores.

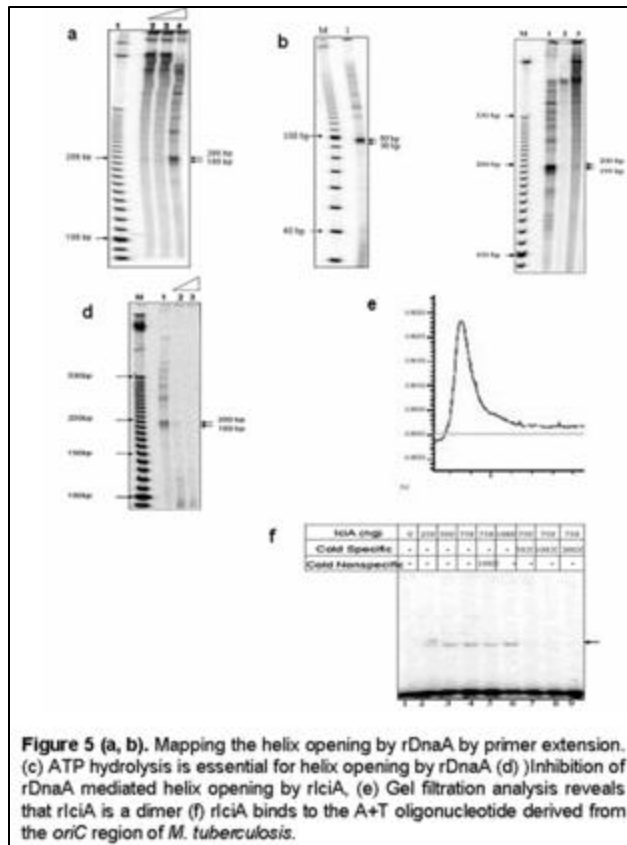
## 2. Characterization of genes involved in DNA replication

Replication in eubacteria is initiated when DnaA, an initiator protein, binds to DnaA boxes located within the origin of replication (*oriC*). Initiation of replication from *oriC* can be divided into three stages namely, initiation, open complex formation and pre priming. Initiation of replication proceeds with the binding of DnaA protein to *oriC* and leads to opening of 13-mer regions, which is followed by entry of DnaB helicase to form the prepriming complex. There are five DnaA-binding sites in *oriC* region of *E. coli*, referred to as R boxes, to which both active ATP-DnaA and inactive ADP-DnaA bind with equal affinity. There are additional initiator-binding sites in *oriC* region referred to as I sites to which DnaA-ATP can bind. DnaA protein binds with nearly equal affinity to ATP and ADP. For opening of the DNA duplex, multiple DnaA protein complexed with ATP bind to *oriC* and melt the DNA unwinding element (DUE). ADP bound form of DnaA is highly inactive for replication initiation, forming an important level of regulation at the origin.

IciA (Inhibitor of *Chromosome* initiation) of *E. coli* blocks initiation at a very early stage *in vitro* by binding specifically to A+T rich region of *oriC*. Binding of IciA blocks the opening of A+T rich region mediated by DnaA and HU (or IHF) protein and this inhibition of strand opening by IciA does not affect binding of DnaA and IHF (or HU) protein to their respective binding sites. IciA is also implicated in binding to A+T rich regions in plasmid replication origins and the copy number of the F plasmid is increased in *iciA* deletion mutant and shows higher binding preference for curved DNA. IciA as a transcription regulator controls *nrd* gene encoding ribonucleoside diphosphate reductase, as an activator of *dnaA* gene and as a regulator of arginine exporter encoded by *yggA* gene.

*M. tb* maintains itself in two physiologically distinct growth states – an active replicative state and a non-replicative persistent state. The genetic elements responsible for the replication process in *M. tb*, specifically its initiation and regulation, are not known. Given the clinical significance of persistence within the macrophages, it is important to identify and characterize the events involved in *M. tb* replication initiation and the negative effectors of replication initiation.

**Open complex is formed near the A+T rich repeat:** At the 3' end of *dnaA* gene there is a



single stretch of A+T nucleotide repeat which could be the site for helix opening. For open complex formation DnaA protein is sufficient and requires higher temperature for both its formation and stability. We used potassium permanganate ( $KMnO_4$ ) footprinting to monitor in vitro opening of the DNA helix. At physiological pH,  $KMnO_4$  selectively oxidizes unpaired pyrimidines, especially thymine residues, in single stranded DNA and in helically distorted duplex DNA. Oxidized pyrimidines prevent primer extension by the Klenow fragment of DNA polymerase I beyond the modified residues.

For the helix-opening assay increasing amounts of DnaA protein (25-75ng) were incubated with supercoiled pUCoriMtb. The products were fractionated on sequencing gel (Figure 5B). The helix opening was observed only in the presence of 75ng of rDnaA protein (lane 3) but not when 25ng (lane1) or 50ng (lane2) of rDnaA was used. The extension products were then separated on a standard (6% or 15%) urea sequencing gel (Figure 5A and B). In primer extension we sequenced only the template upstream of *dnaN* gene and hence all the distances are measured from start point of *dnaN* gene. Primer oriR1 annealed between positions – 292 to – 320 of template strand and primer oriR2 annealed between positions – 402 to – 420 of the template strand. Analysis of the sequencing product obtained with primer oriR1 (Figure 5C, D) show two bands of size 200bp and 199bp (Figure 5A, lane 4)

corresponding to thymidine residues of – 501 and – 500 from the start of *dnaN* gene. To further confirm the exact location of the oxidized thymidine another primer oriR2 was used. When primer oriR2 was used once again two bands of size 99bp and 98bp (Figure 5B, lane1), corresponding to positions – 500 and – 501 could be seen. *E. coli* origin contains three tandem repeats of an A+T rich 13-mers and four 9-mer dnaA boxes where DnaA protein binds and opens the DNA for initiation of DNA synthesis. *M. tb* *oriC*, however, is very complex as it contains 13 non perfect dnaA boxes and lacks a distinct A+T nucleotide repeat thereby explaining the inability to experimentally to map the precise nucleotide involved in duplex opening. Our results provide evidence for the first time that in *M. tb* the duplex opening occurs near position – 500 (from start of *dnaN* gene) which lies in this A+T rich nucleotide repeat.

**ATPase activity is essential for open complex formation:** Having mapped the precise nucleotide, within the *oriC* region of *M. tb*, involved in opening of the duplex DNA we investigated the requirement of ATP hydrolysis in this process and also whether other hydrolysable and poorly hydrolysable analogue of ATP could provide the necessary energy to drive this process. The *E. coli* DnaA protein has a very weak intrinsic ATPase activity but in *M. tb* intrinsic ATPase activity of DnaA promotes rapid oligomerization of DnaA on *oriC* and both ATP binding and ATP hydrolysis are required for rapid oligomerization of DnaA on *oriC*. We therefore carried out helix opening reaction with 5mM of ATP, ADP and ATP?S (Lithium salt). After oxidation with 8mM KMnO<sub>4</sub> and primer extension, helix opening was observed only in the presence of 5mM ATP (Figure 5C). Only when 75ng of rDnaA is incubated with 5mM ATP (Figure 5C, lane 1), helix opening is apparent but not when 75ng of rDnaA is incubated with 5mM ADP (lane 2) or with 5mM ATP?S (lane 3). Helix opening once again begins at positions – 500 and – 501 (taken from start of *dnaN* gene). These results directly support the role of ATP in helix opening, which is the first step of replication initiation.

**Inhibition of helix opening by IciA:** IciA in addition to other function is a known inhibitor of chromosome replication initiation. We used a putative IciA (*ORF Rv1985c*) of *M. tb*, which showed 35.8% identity in 285 amino acid overlap to *iciA* of *E. coli*. Analysis of secondary structure through FUGUE server also demonstrated that both IciA of *E. coli* and *Rv1985c* are identical. To determine the effect, if any, on inhibition of open complex formation by IciA of *M. tb*, experiments were designed to carry out the helix opening reaction in increasing concentration of recombinant purified IciA protein. For monitoring inhibition of helix opening 75ng of rDnaA protein was used as this amount was sufficient for helix opening, using assay condition described earlier (Figure 5A) helix opening is evident from the appearance of 200 bp and 199 bp band in primer extension reaction (Figure 5D, lane 1). However, in the presence of 50ng of IciA protein (lane 2) a very faint band could be seen which completely disappears as the concentration of IciA is increased to 75ng (lane 3). In this reaction IciA

was present before the adding of DnaA protein, but when IciA was included 8min after incubation at 37°C to allow the formation of open complex, there was no inhibition. These results demonstrate that IciA protein can block open complex formation only before the 13-mer region is opened and not after the opening of the 13-mer region has already occurred. IciA protein therefore apparently has no inhibitory role once the helix opening stage is crossed.

**Electrophoretic Mobility Shift Assays:** Having identified the nucleotides involved in helix opening *in-vitro*, as seen by *oriC* unwinding assays, oligonucleotides corresponding to this region was used to determine DNA-protein interaction involving IciA. Electrophoretic mobility shift assay were carried out using an A+T rich *oriC* element and huge excess of poly(dI-dC). Results clearly show that IciA protein binds specifically to A+T rich region (Figure 5F). That this binding is specific is clearly evident from homologous and heterologous cold competition assays. Even in 100 fold molar excess of non-specific competitor DNA, the DNA-protein complex is present (Figure 5F, lane 5). Whereas the DNA-protein complex completely disappears in presence of 50 fold molar excess of specific competitor DNA (lane 7). These results demonstrate that IciA inhibits helix opening by specifically binding to A+T rich region on the *oriC*.

**Gel Filtration Chromatography reveals rIciA is a dimer:** IciA belongs to a lysR family of prokaryotic transcription regulators and the members of this family are either dimer or tetramer. Gelfiltration results (Figure 5E) clearly demonstrate that *M. tb* IciA is a complete dimer with no monomeric or tetrameric forms. This is reflected in the form of single a peak, with a molecular of around 69.12 KDa. The estimated molecular mass of dimer is 71.28 KDa, which is in close agreement with the molecular mass determined experimentally. These results clearly demonstrate that mIciA also behaves as a dimer.

### **3. Identification and Characterization of novel antigens of *M. tb* vis-a vis their immune response**

Secretory proteins of *Mycobacterium tuberculosis* have been shown to be major immunomodulators of host immune response and promising vaccine candidates. The conserved hypothetical protein Rv2626c was earlier shown to be a secretory protein present in the *M. tb* culture filtrates, upregulated upon infection in mice and in hypoxic conditions and elicited humoral immune response in TB patients. We describe the ability of recombinant Rv2626c in eliciting Th1 type adaptive immune response in murine macrophage cell lines. rRv2626c induces murine macrophage cell lines to secrete Nitric oxide (NO) mediated by expression of iNOS. Significant induction of pro-inflammatory cytokine, TNF- $\alpha$ , was evident

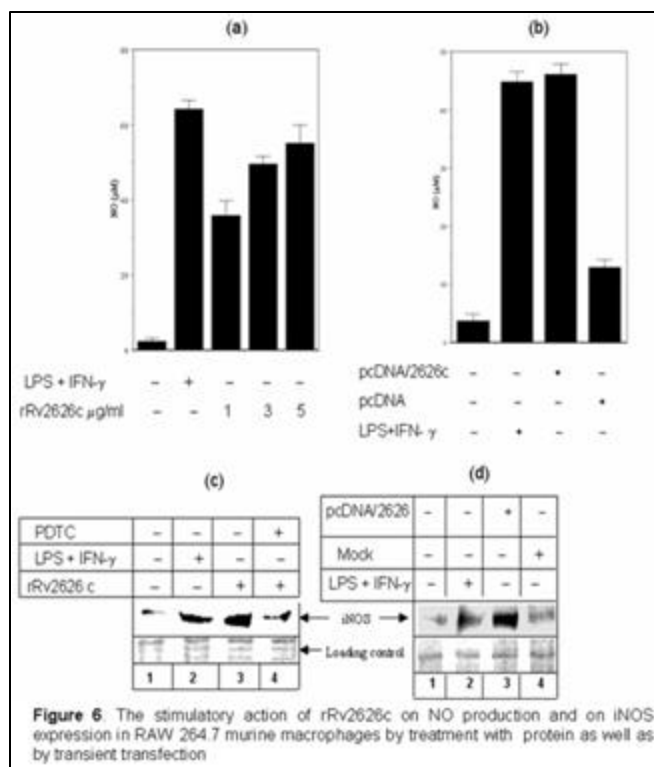


in *in vitro* cultured murine macrophage cell lines upon stimulation with rRv2626c. The secretion of NO and pro-inflammatory cytokines in response to Rv2626c is mediated by transcription factors like NF- $\kappa$ B and AP-1, and this was evident from the administration of specific pharmacological inhibitor. That Rv2626c could modulate macrophage functions eliciting innate and adaptive immune response has implications of using this hypothetical protein in providing protective immunity against *M. tuberculosis* infection.

**Recombinant Rv2626 (rRv2626c) induces nitrite and iNOS expression in RAW 264.7 murine macrophages:**

In the initial experiments, ability of rRv2626c on nitrite production *via de novo* expression of iNOS in RAW264.7 cells was assessed. The cells were stimulated with different concentrations of rRv2626c as well as with the potent NO activator LPS plus IFN- $\gamma$ . The addition of rRv2626c significantly increases production of NO in RAW 264.7 cells in a dose dependent manner. The addition of 0.3  $\mu$ g of rRv2626c was sufficient enough to increase NO to 1.4 fold as compared to the cells co-treated with LPS and IFN- $\gamma$ , the potent NO activators as shown in Figure 6A. The same effect was seen in another murine macrophage cell line J774. The stimulation with rRv2626c was able to increase the nitrite release from cells co stimulated with LPS and IFN- $\gamma$  in a synergetic manner. Heat inactivated

rRv2626c markedly decreased NO production equal to cells that were left unstimulated. Increased NO production by RV2626c was correlated well with the expression pattern of the inducible nitric oxide (iNOS) protein, known to activate NO production (Figure 6C). It could be seen that rRv2626c was able to increase iNOS expression by about 3 fold, but in the presence of NF- $\kappa$ B transcription factor inhibitor PDTC rRv2626c was not able to induce the expression of the iNOS indicating that NF- $\kappa$ B is important for Rv2626c-mediated induction of iNOS gene. Similar results were seen in another



murine macrophage cell line J774 indicating that the effect of rRv2626c is independent of cell line used for our study. The observed increase in NO or iNOS expression was

specifically due to the rRv2626c and not due to any LPS contamination as similar results were observed in transiently transfected RAW 264.7 cells with *pcDNA-2626c* construct as compared to expression vector control. *pcDNA-2626c* transfected cells were able to increase the NO to 10 fold as compared to the vector control as shown in Figure 6B; even at the level of iNOS expression also the *pcDNA-2626c* transfected cells were able to increase iNOS expression to 8 fold as compared to the vector control and 10 fold as compared to the unstimulated cells as shown in Figure 6D indicating that rRv2626c is potent enough to increase NO by activating iNOS expression equal to the level induced by known NO activators LPS plus IFN- $\gamma$ ? This indicates that the protein coded by rRv2626c plays a significant role in the host pathogen interaction at the immune interface.

**Publications:**

- 1 Achary MS, Reddy ABM, Chakrabarti S, Panicker SG, Mandal AK, Ahmed N, Balasubramanian D, Hasnain SE and Nagarajaram HA (2006) Disease-causing mutations in proteins: structural analysis of the CYP1B1 mutations causing primary congenital glaucoma in humans. ***Biophysical Journal*** 91:4329-4339.
- 2 Akif M, Akhter Y, Hasnain SE and Mande SC (2006) Crystallization and preliminary X-ray crystallographic studies of *Mycobacterium tuberculosis* CRP/FNR family transcription regulator. ***Acta Crystallographica F*** 62:873-875.
- 3 Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Loch C and Supply P (2006) Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India suggests an ancient focus of tuberculosis in South Asia. ***Emerging Infectious Diseases*** 12:367-374.
- 4 Kazim SN, Sarin SK, Sharma BC, Khan LA and Hasnain SE (2006) Characterization of naturally occurring and Lamivudine-Induced surface gene mutants of hepatitis B virus in patients with chronic hepatitis B in India. ***Intervirology*** 49:152-160.
- 5 Kenchappa P, Duggirala A, Ahmed N, Pathengay A, Das T Hasnain SE and Sharma S (2006) Fluorescent amplified fragment length polymorphism (FAFLP) genotyping demonstrates the role of biofilm-producing methicillin-resistant periocular *Staphylococcus epidermidis* strains in postoperative endophthalmitis. ***BMC Ophthalmology*** 6:1.

- 6 Khan N, Rahim SS, Boddupalli CS, Ghousunnissa S, Padma S, Pathak N, Thiagarajan D, Hasnain SE and Mukhopadhyay S (2006) Hydrogen peroxide inhibits IL-12 p40 induction in macrophages by inhibiting c-rel translocation to the nucleus through activation of calmodulin protein. **Blood** 107:1513-1520.
- 7 Qamra R, Prakash P, Aruna B, Hasnain SE and Mande SC (2006) The 2.15Å crystal structure of *M. tuberculosis* chorismate mutase reveals unexpected gene duplication and suggests a role in host-pathogen interactions. **Biochemistry** 45:6997-7005.
- 8 Rao KR, Ahmed N, Sriramula S, Sechi LA and Hasnain SE (2006) Rapid identification of *Mycobacterium tuberculosis* Beijing genotypes on the basis of the mycobacterial Interspersed repetitive Unit locus 26 signature. **Journal of Clinical Microbiology** 44:274-277.
- 9 Sitalaxmi T, Kashyap VK, Guha S, Hima Bindu G, Hasnain SE and Trivedi R (2006) Genetic structure of Indian populations based on fifteen autosomal microsatellite loci. **BMC Genetics** 7:28.
- 10 Tundup S, Akhter Y, Thiagarajan D and Hasnain SE (2006) Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. **FEBS Letters** 580:1285-1293
- 11 Boddupalli CS, Ghosh S, Rahim SS, Nair S, Ehtesham NZ, Hasnain SE and Mukhopadhyay S (2007) Nitric oxide inhibits interleukin-12 p40 through p38 MAPK-mediated regulation of calmodulin and c-rel. **Free Radical Biology and Medicine** 42:686-697.
- 12 Khan N, Ghousunnissa S, Jegadeeswaran SM, Thiagarajan D, Hasnain SE and Mukhopadhyay S (2007) Anti-B7-1/B7-2 antibody elicits innate-effector responses in macrophages through NF-κB-dependent pathway. **International Immunology** 19:477-486 (InPress).
- 13 Bashyam MD and Hasnain SE (2007) Array-based comparative Genomic Hybridization: applications in cancer and tuberculosis. Bioarrays, Ed. K Appasani, **Humana press** (InPress).

- 14 Banerjee S, Nandyala AK, Raviprasad P, Ahmed N and Hasnain SE (2007) Iron dependent Iron binding activity of Mycobacterium tuberculosis aconitase. **Journal of Bacteriology** (InPress).
- 15 Mukhopadhyay S, Nair S and Hasnain SE (2007) Nitric oxide: Friendly rivalry in tuberculosis. **Current Signal Transduction Therapy** (InPress).
- 16 Rasheedi S, Aruna R, Nasreen ZE and Hasnain SE (2007) Biochemical characterization of Sf9 Sp family-like protein factors reveals interesting features. **Archives of Virology** (InPress).
- 17 Rasheedi S, Ghosh S, Suragani M, Tuteja N, Sopory SK, Hasnain, S E, Nasreen, Z E (2007) Pisum sativum Contains A Factor With Strong Homology to CIF5B. **Gene** (InPress).
- 18 Savitri S, Aparna D, Prashanth K, Justine KP, Ahmed N, Prashan K, Das T and Hasnain S E (2007) High-resolution genome profiling differentiated *Staphylococcus epidermidis* strains from patients with ocular infections and normal individuals: A Pilot study. **Journal of Clinical Microbiology** (InPress).
- 19 Yusuf A, Smanla T and Hasnain SE (2007) Novel Biochemical Properties of a CRP/FNR Family Transcription Factor from *Mycobacterium tuberculosis*. **International Journal of Medical Microbiology** (InPress).

## Understanding how *Mycobacterium tuberculosis* interferes with macrophage-signaling cascades to suppress anti-mycobacterial protective immune responses

<b>Principal Investigator</b>	Sangita Mukhopadhyay	Staff Scientist
<b>Ph D Students</b>	Nooruddin Khan	Senior Research Fellow
	Shiny Nair	Senior Research Fellow
	Kaisar Alam	Senior Research Fellow
	Khalid Hussain Bhat	Junior Research Fellow
	G Sreejit	Junior Research Fellow
<b>Other Members</b>	A R Poongothai	Project Associate
	C S Bodupalli	Project Assistant
	Sheikh Ghousunnissa	Project assistant
	Shaik Tabrej Hussain	Project Assistant
<b>Collaborators</b>	Seyed E Hasnain	Univ of Hyderabad, Hyderabad
	Sekhar C Mande	CDFD, Hyderabad
	Nasreen J Ehtesham	NIN, Hyderabad
	K J R Murthy and V Valluri	Mahavir Hosp, Hyderabad

### Objectives

1. Examining the signaling pathways involved in the regulation of IL-12 by free radicals
2. Investigating the role of B7 family proteins in regulating macrophage innate-effector functions.

3. Examining the effect of mycobacterial heat shock protein 60 in modulating macrophage signaling and anti-PPD T cell responses

### Summary of work done until the beginning of this reporting year

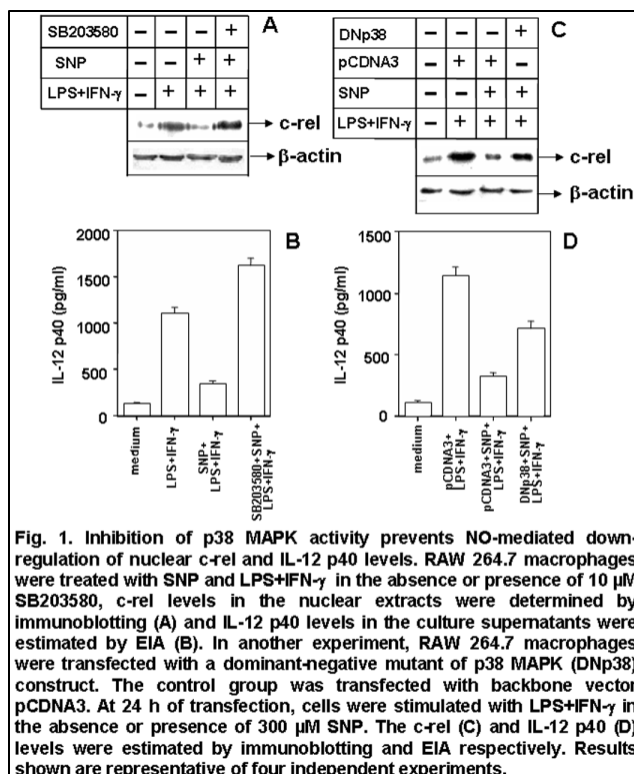
Higher levels of nitric oxide (NO) are produced by macrophages during infection with *Mycobacterium tuberculosis*. NO though shows an ability to inhibit mycobacterial infection by inducing innate cytotoxic response, it may become a problem in later phase when it converts a protective Th1 response to a subversive Th2 response mainly by inhibiting IL-12 cytokine, thus acting as a potential regulator of Th1/Th2 response. Therefore, an understanding of the signaling pathways involved in the regulation of IL-12 by NO may help in the selective modulation of Th1/Th2 to regulate mycobacterial infection. We reported earlier that NO inhibited IL-12 induction in macrophages by downregulating the nuclear c-rel transcription factor. A detailed signaling pathway by which NO inhibits IL-12 may be helpful to design new therapeutic agent(s) for treatment of tuberculosis. A study to understand the upstream signaling pathways involved in the regulation of c-rel by NO revealed that NO increased p38 MAPK phosphorylation. Further studies revealed a cross-talk between p38 MAPK and

IL-12 p40 signaling involving the c-rel. We have then checked whether NO involves the p38 MAPK pathway to subsequently inhibit c-rel transcription factor and IL-12 p40 transcription.

## Details of progress made in the current reporting year (April 1, 2006-March 31, 2007)

### 1. Examining the signaling pathways involved in the regulation of IL-12 by free radicals

NO targets the p38 MAPK to downregulate nuclear c-rel level and IL-12 p40 in activated macrophages: Since NO activates p38 MAPK and both nuclear c-rel and IL-12 p40 levels are lower in NO-exposed macrophages, it is likely that NO targets the upstream p38 MAPK to subsequently down-regulate nuclear c-rel level and IL-12 p40 induction. The SNP-treated RAW 264.7 macrophages were therefore, stimulated with LPS+IFN- $\gamma$  in the absence or



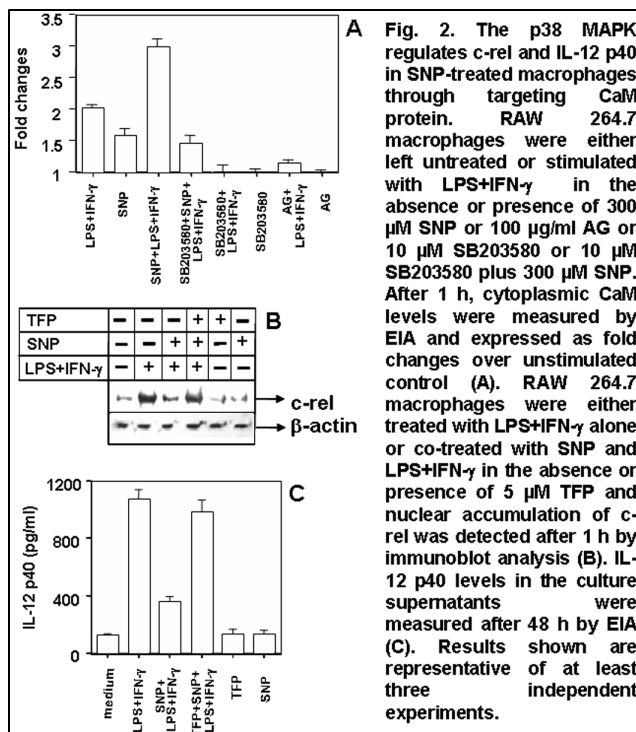
presence of 10  $\mu$ M SB203580. Data shown in Fig. 1A reveal that treatment with SB203580 augmented nuclear c-rel level in the group treated with LPS+IFN- $\gamma$  and SNP. Although c-rel was increased, p65 NF- $\kappa$ B level was not affected by SB203580 indicating that regulation of IL-12 p40 by p38 MAPK is probably mediated through c-rel but not p65 NF- $\kappa$ B in SNP-treated macrophages. This up-regulation of nuclear c-rel level was correlated with an enhancement of IL-12 p40 (Fig. 1B). This study indicates that p38 MAPK activity is targeted by NO causing inhibition of nuclear c-rel level and subsequently resulting in

suppression of IL-12 p40 induction.

To further confirm the role of p38 MAPK in NO-mediated inhibition of IL-12 p40, we next used dominant-negative mutant of p38 (DNP38). It could be seen that the c-rel level was upregulated by transfection with DNP38 in spite of SNP exposure (Fig. 1C). This enhancement in nuclear c-rel level in the DNP38-transfected group resulted in concomitant up-regulation of IL-12 p40 when compared with the group transfected with the backbone vector pCDNA3 (Fig. 1D) These results further support the view that activated p38 MAPK

negatively regulates nuclear c-rel level and this consequently plays an important role in NO-mediated IL-12 p40 inhibition in LPS+IFN- $\gamma$  activated RAW 264.7 macrophages.

The p38 MAPK targets CaM protein to negatively regulate nuclear c-rel level and IL-12 p40 in SNP-treated RAW 264.7 macrophages: Recently, we have shown a role of calmodulin (CaM) in the regulation of c-rel and IL-12 p40 in activated macrophages during oxidative stress. CaM binds to c-rel in the cytosol preventing its transport to the nucleus causing decreased nuclear c-rel level leading to decreased IL-12 p40 induction. Therefore, it is possible that p38 MAPK inhibits c-rel vis-a-vis IL-12 p40 in SNP-treated macrophages by up-regulating CaM. Thus the cytoplasmic CaM levels were examined in activated macrophages treated with SNP in the absence or presence of SB203580. Stimulation with LPS+IFN- $\gamma$  led to almost two-fold increase in CaM level (Fig. 2A, bar 1) over the control. Upon combined stimulation with SNP and LPS+IFN- $\gamma$ , CaM level was significantly increased (Fig. 2A, bar 3)



and SB203580 treatment reduced CaM level (Fig. 2A, compare bar 4 with bar 3;  $P < 0.001$ ). In LPS+IFN- $\gamma$  activated macrophages, inhibition of NO production by AG also decreased CaM level (Fig. 2A, compare bar 7 with bar 1;  $P < 0.001$ ). Interestingly, CaM induction was upregulated by SNP treatment alone (Fig. 2A, bar 2) indicating that activation of p38 MAPK was positively correlated with increased CaM level. As expected, SB203580 reduced CaM level in LPS+IFN- $\gamma$  -activated macrophages (Fig. 2A, compare bar 5 with bar 1;  $P < 0.001$ ). However, SB203580 or AG

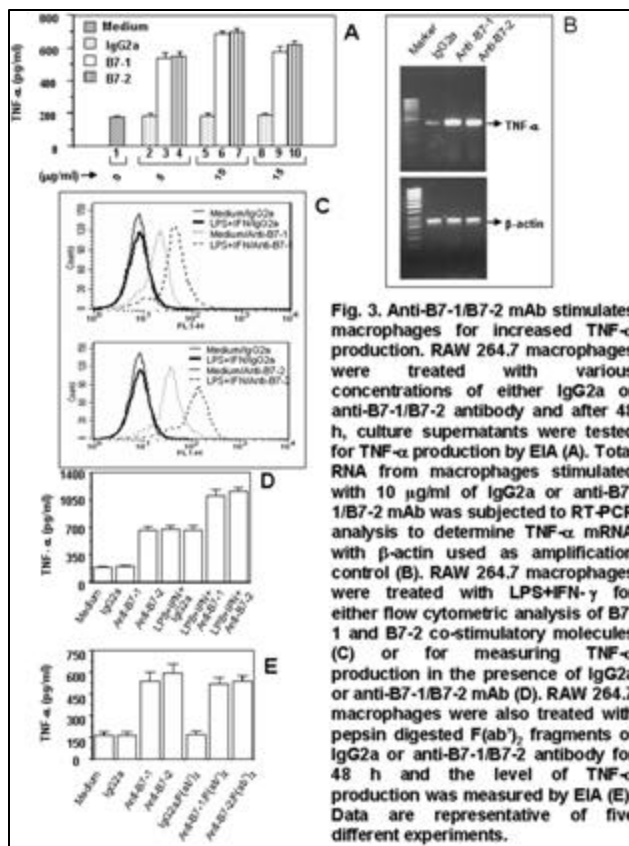
alone did not affect CaM level (Fig. 2A, bar 6 and bar 8) with respect to the control.

In SNP-treated macrophages, we further confirmed involvement of CaM in down-regulating c-rel vis-a-vis IL-12 p40 using TFP, a known pharmacological inhibitor of CaM activity. It was observed that TFP increased nuclear c-rel level in the LPS+IFN- $\gamma$ -activated macrophages exposed to SNP (Fig. 2B, compare lane 4 with lane 3). Increased c-rel in the TFP-treated group was positively correlated with increased IL-12 p40 (Fig. 2C, compare bar 4 with bar 3;

P < 0.001). TFP alone had no effect on c-rel (Fig. 2B, lane 5) and IL-12 p40 (Fig. 2C, bar 5) in unstimulated cells. TFP upregulated c-rel and IL-12 levels also in LPS-activated macrophages treated with SNP (data not shown) indicating that CaM was involved in the NO-mediated down-regulation of c-rel and IL-12 p40 irrespective of the stimulator used to activate macrophages. These results confirm a direct role of CaM in NO-mediated down-regulation of c-rel and IL-12 p40 levels in RAW 264.7 macrophages.

## 2. Investigating the role of B7 family proteins in regulating macrophage innate-effector functions

In this study, we showed that signaling through the B7-1/B7-2 molecules results in macrophage activation for various innate-effector functions like induction of TNF- $\alpha$ , bactericidal activity etc. indicating that the B7 family proteins not only provide co-stimulatory stimulations to T cells for eliciting optimum T cell proliferation but also are crucial to control



the macrophage-innate responses. We have further studied the upstream signaling pathways involved in the regulation of TNF- $\alpha$ , bactericidal activity induced during signaling through the B7 proteins.

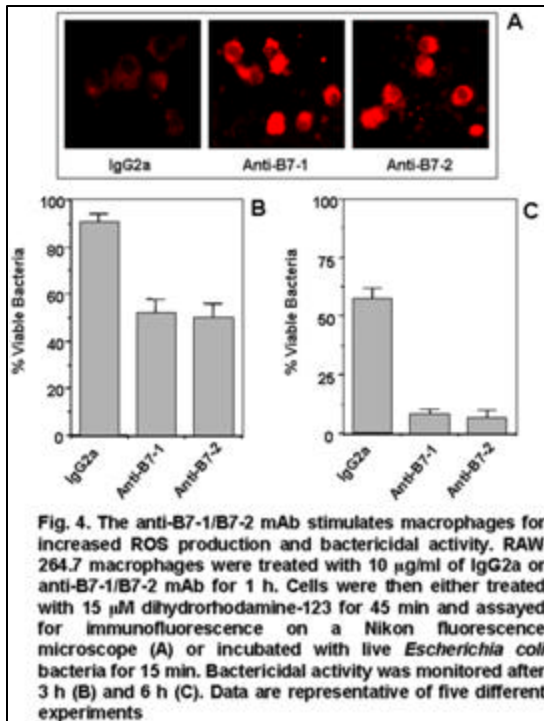
### Induction of TNF- $\alpha$ is higher in macrophages treated with anti-B7-1/B7-2 mAb:

We first examined whether murine macrophages treated *in vitro* with mAb to either B7-1 or B7-2, produced higher levels of TNF- $\alpha$  in the cultures. Accordingly, RAW 264.7 macrophages were cultured with various concentrations (5, 10 and 15  $\mu$ g/ml) of either isotype-matched control antibody (IgG2a isotype) or anti-B7-1 mAb or anti-B7-2 mAb and

after 48 h culture supernatants were tested for the levels of TNF- $\alpha$  produced by EIA. As opposed to the isotype control group, treatment with anti-B7-1 mAb resulted in increased production of TNF- $\alpha$  (Fig. 3A). Similarly, the anti-B7-2 mAb also increased TNF- $\alpha$  production (Fig. 3A). Both anti-B7-1 mAb and anti-B7-2 mAb individually stimulated murine macrophages to preferentially express mRNA transcripts for TNF- $\alpha$  as compared to the



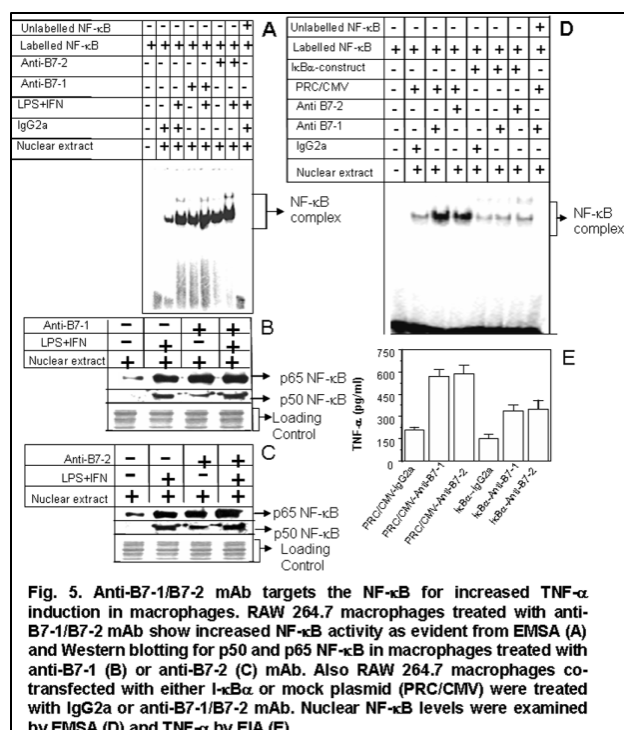
IgG2a control group (Fig. 3B). We next examined whether induction profile of TNF- $\alpha$  was directly proportional to the number of B7-1/B7-2 molecules. Therefore, macrophages were stimulated with LPS+IFN- $\gamma$  to increase B7-1 and B7-2 levels as detected by flow cytometry (Fig. 3C). It could be seen that the TNF- $\alpha$  production was more in the LPS+IFN- $\gamma$  activated group (Fig. 3D) than those treated with anti-B7-1/B7-2 mAb alone. It was found that the activity of anti-B7-1/B7-2 mAb on the production of TNF- $\alpha$  could be abrogated by heat but not by polymyxin B or anti-CD14- plus anti-TLR4 antibody treatment ruling out contribution of



species (ROS) production and microbicidal function: TNF- $\alpha$  can augment ROS production in macrophages. It was evident that the anti-B7-1 mAb and the anti-B7-2 mAb increased intracellular ROS levels in RAW 264.7 macrophages as compared to the IgG2a control antibody (Fig. 4A). Induction of ROS is critically important for macrophage-mediated microbicidal activity. The microbicidal function of macrophages treated with either anti-B7-1 mAb or

endotoxin to the macrophage stimulatory property of anti-B7-1/B7-2 mAb. It is known that signaling through Fc receptors can activate macrophages for higher production of TNF- $\alpha$ . To rule out this possibility, we next treated anti-B7-1 and anti-B7-2 antibodies with pepsin to purify F(ab')<sub>2</sub> fragments. It was found that the F(ab')<sub>2</sub> fragments activated macrophages for increased TNF- $\alpha$  induction (Fig. 3E) in a similar manner as observed with the respective intact antibody (Fig. 3A). It was noticed that anti-B7-1/B7-2 mAb had little effect on the induction profile of IL-10 and IL-12 cytokines.

Macrophages treated with anti-B7-1/B7-2 mAb show enhanced reactive oxygen



anti-B7-2 mAb was therefore directly examined. It was observed that the bactericidal ability of both anti-B7-1- and anti-B7-2-treated macrophages was significantly higher than that of IgG2a-treated macrophages when examined after 3 h (Fig. 4B) and 6 h (Fig. 4C) of incubation.

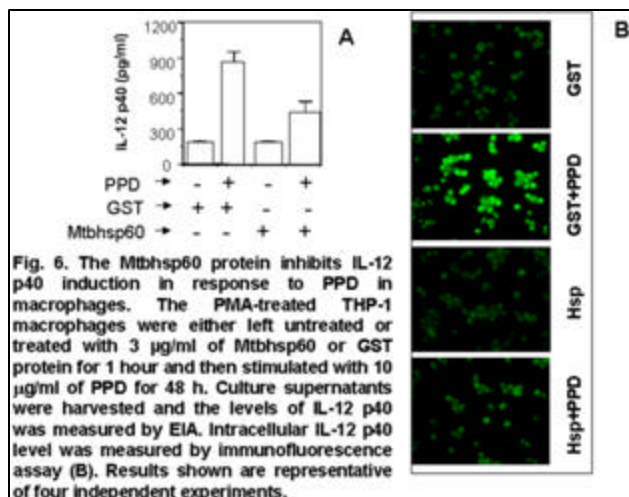
**The anti-B7-1/B7-2 mAb activated p50 and p65 NF- $\kappa$ B which play important roles in the induction of TNF- $\alpha$  in anti-B7-1/B7-2 mAb-treated macrophages:** The NF- $\kappa$ B family of transcriptional regulators is important for induction of TNF- $\alpha$  as well as microbicidal activity. Therefore, NF- $\kappa$ B induction was measured by EMSA using a consensus NF- $\kappa$ B deoxy-oligonucleotide probe labeled with [ $\gamma$ - $^{32}$ P]-ATP. As compared to the IgG2a control group, intensity of the DNA-protein complex corresponding to the NF- $\kappa$ B proteins was apparently increased upon treatment of macrophages with either B7-1 mAb (Fig. 5A, compare lane 4 with lane 2) or B7-2 mAb (Fig. 5A, compare lane 6 with lane 2). As expected, LPS+IFN- $\gamma$  treatment increased intensity of the complex (Fig. 5A, compare lanes 5 and 7 with 3). Since p50 and p65 NF- $\kappa$ B play important role in TNF- $\alpha$  gene activation, we next examined expression profile of p50 and p65 NF- $\kappa$ B. The Western blotting experiment also revealed that treatment with anti-B7-1 mAb (Fig. 5B) and anti-B7-2 mAb (Fig. 5C), as compared to the respective control populations, resulted in an increase in p50 and p65 NF- $\kappa$ B proteins demonstrating anti-B7-1/anti-B7-2 mAb treatment increased NF- $\kappa$ B activity as evident from EMSA and Western blotting. Further experiment using the phosphorylation-defective I $\kappa$ B $\alpha$  plasmid construct to inhibit NF- $\kappa$ B (Fig. 5D) revealed that activation of NF- $\kappa$ B complex by anti-B7-1/B7-2 mAb, was actually responsible for increased TNF- $\alpha$  induction in macrophages (Fig. 5E).

### **3. Examining the effect of mycobacterial heat shock protein 60 in modulating macrophage signaling of anti-PPD T cell responses**

The IL-10 and IL-12 cytokines from macrophages can directly regulate T cell effector phenotype of T helper 1 (Th1) or Th2 type. Earlier, we have reported that IL-10 can inhibit IL-12 p40 induction by targeting the rel proteins. The mycobacterial heat shock protein (Hsp) 60 has recently been shown to influence the immune responses and their over-expression during infection suggests that this might offer a survival advantage to mycobacteria within the host. Since Hsp60 protein can influence the Th2 responses, we examined whether anti-PPD T cell response is modulated by the mycobacterial Hsp60 protein (MtbHsp60). To understand whether MtbHsp60 modulates anti-PPD T cell response we collected PBMC

samples from TB-infected patients from local TB Hospital of Hyderabad, India, and used for *in vitro* antigen proliferation assay using PPD as recall antigen with 3 µg/ml of either

MtbHsp60 or unrelated recombinant glutathione-S-transferase (GST) protein. The anti-PPD Th1 and Th2 cytokines were examined measuring IFN-γ as the Th1 cytokine and IL-5 as the Th2 as described previously. When compared with GST-control group, it could be seen that the anti-PPD T-cell response in the MtbHsp60-treated group was poorer and biased towards the Th2 type with an increase in IL-5 and a decrease in IFN-γ response. Further studies revealed that MtbHsp60 protein targets the macrophage to skew the anti-PPD T-cell response towards the Th2.



the anti-PPD T cells responses were skewed to Th2 type by MtbHsp60 when macrophages were used as antigen-presenting cells indicating that MtbHsp60 might influence the PPD-induced IL-12 p40 signaling in macrophages to establish a Th2 response. The results shown in Fig. 6, A and B clearly reveal that mycobacterial Hsp60 inhibits PPD-induced IL-12 p40 in a dose dependent manner. The IL-12 p40 is mainly regulated by various rel transcription factors like p50 and p65 NF-κB and c-rel. While studying the transcription factor(s) involved in the downregulation of IL-12 p40 by MtbHsp60, it was observed that MtbHsp60 protein mainly inhibited the nuclear c-rel transcription factor but not the p50- or the p65 NF-κB in PPD-activated macrophages. To confirm whether c-rel played a critical role in Hsp60 protein-mediated inhibition of IL-12 p40, we over-expressed c-rel transcription factor in the macrophages co-treated with PPD and MtbHsp60. The results reveal that transfection with c-rel could significantly restore IL-12 p40 induction in response to PPD in macrophages even being treated with MtbHsp60 indicating that mycobacterial Hsp60 protein targets the c-rel

transcription factor to downregulate IL-12 p40 induction in PPD-stimulated macrophages. The detailed upstream signaling mechanism involved in the MtbHsp60-mediated downregulation of c-rel and IL-12 is under investigation.

### ***Publications***

- 1 Khan N, Rahim SS, Boddupalli CS, Ghousunnissa S, Padma S, Pathak N, Thiagarajan D, Hasnain SE and Mukhopadhyay S (2006) Hydrogen peroxide inhibits IL-12 p40 induction in macrophages by inhibiting c-rel translocation to the nucleus through activation of calmodulin protein. ***Blood*** 107:1513-1520.
- 2 Khan N, Ghousunnissa S, Jegadeeswaran SM, Thiagarajan D, Hasnain SE and Mukhopadhyay S (2007) Anti-B7-1/B7-2 antibody elicits innate-effector responses in macrophages through NF- $\kappa$ B-dependent pathway. ***International Immunology*** 19:477-486.
- 3 Boddupalli CS, Ghosh S, Rahim SS, Nair S, Ehtesham NZ, Hasnain SE and Mukhopadhyay S (2007) Nitric oxide inhibits interleukin-12 p40 through p38 MAPK-mediated regulation of calmodulin and c-rel. ***Free Radical Biology and Medicine*** (In press).
- 4 Mukhopadhyay S, Nair S and Hasnain SE (2007) Nitric oxide: Friendly rivalry in tuberculosis. ***Current Signal Transduction Therapy*** (In Press).

## Genomics and biology of important bacteria I pathogens

<b>Principal Investigator</b>	Niyaz Ahmed	Staff Scientist
<b>Other Members</b>	Ayesha Alvi	DST Fast Track Young Scientist
	S Manjulata Devi	Project Assistant
	M Abid Hussain	ICMR SRF
	Mohammed Rizwan	Post-Doctoral Fellow
<b>Collaborators</b>	Seyed E. Hasnain	University of Hyderabad
	Francis Megraud	INSERM, France
	C M Habibullah	Deccan Medical College, Hyderabad
	Leonardo A Sechi	Univ. of Sassari, Sassari, Italy
	Rudy A Hartskeerl	KIT, Amsterdam, The Netherlands
	Anil K Tyagi	Univ of Delhi, New Delhi
	Akash Ranjan	CDFD, Hyderabad
Sangita Mukhopadhyay	CDFD, Hyderabad	

### Objectives

1. Analysis of origins and current geographic diversity of strains of *H. pylori*
2. Comparative genomics of strains from different patient populations and analysis of genes responsible for chronic gastric inflammation, pathogen survivability/adaptability and phylogeographic variability.
3. Identification of genes/proteins of clinical significance for the development of diagnostics

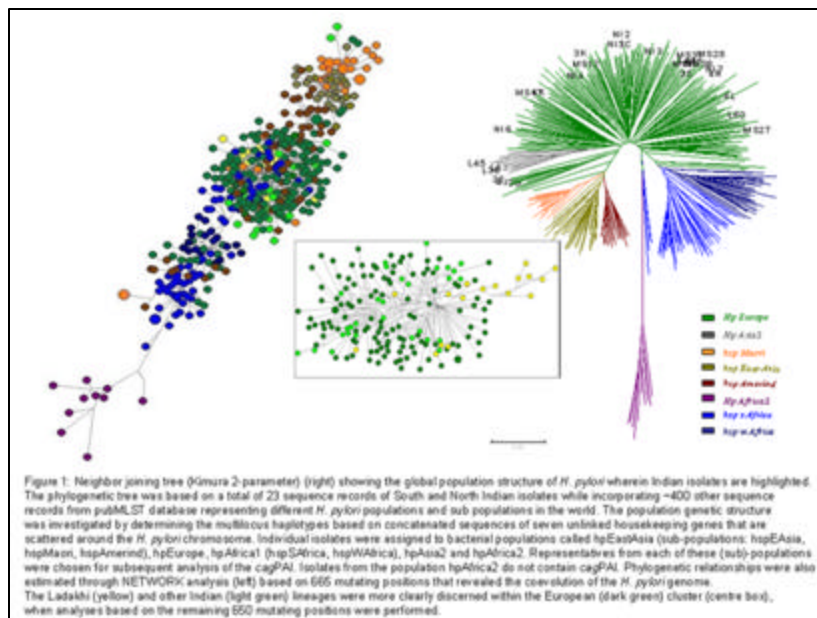
### Summary of work done until the beginning of this reporting year

Last year we reported research progress pertaining to several projects encompassing *Helicobacter pylori* evolution, development of new diagnostic markers for the detection of *H. pylori* and *M. avium paratuberculosis* (MAP) and the development of molecular epidemiological tools for the identification and characterization of strains of Leptospire and other opportunistic bacteria. We presented findings related to the evolution of ancestral and modern lineages of *H. pylori* in Peru. Also, we discussed in vivo evolution of *H. pylori* in a single patient observed for 10 years and discussed implications on understanding acquisition and optimization of virulence in pathogenic bacteria when analyzed as a population and as an individual strain. The diagnostic potentials of isocitrate dehydrogenase of *H. pylori* and the heparin binding haemagglutinin of MAP were systematically analyzed using recombinant protein antigens against patient sera in different ELISA methods. We also highlighted the importance of molecular genotyping methods for the individualization of Leptospiral serovars

representing various different epidemics and sporadic cases and introduced our own in house MLST method for such purposes.

## Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)

### 1. Predominantly European origins of *Helicobacter pylori* in India

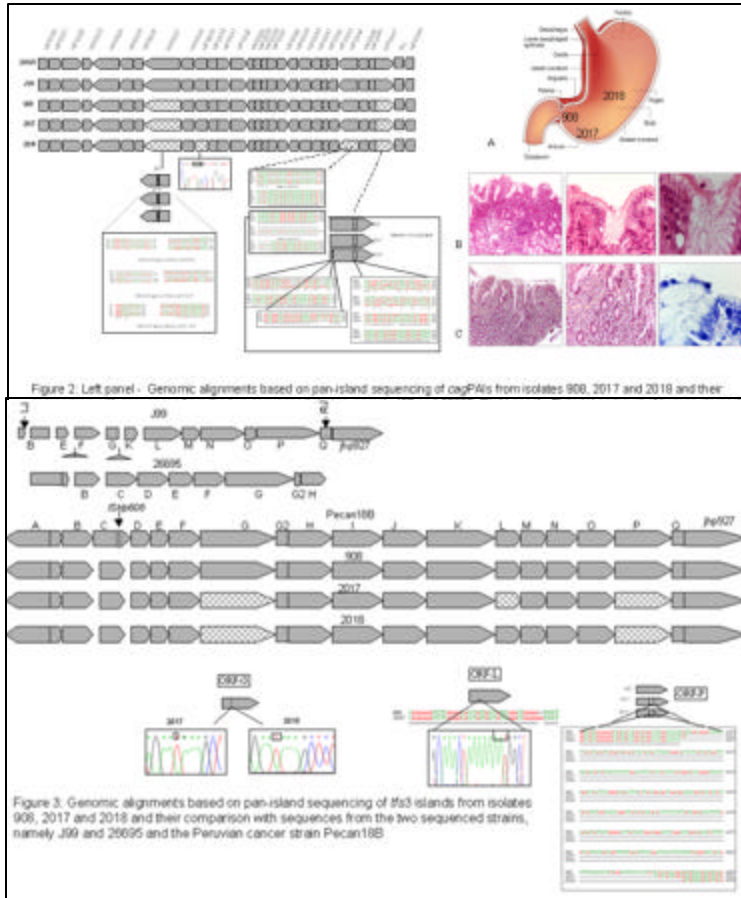


The human gastric pathogen *H. pylori* is co-evolved with its host and therefore, origins and expansion of multiple populations and sub populations of *H. pylori* mirror ancient human migrations. Ancient origins of *H. pylori* in the vast Indian subcontinent are debatable. It is not

clear how different waves of human migrations in South Asia shaped the population structure of *H. pylori*. We tried to address these issues through mapping genetic origins of present day *H. pylori* in India and their genomic comparison with hundreds of isolates from different geographic regions. We attempted to dissect genetic identity of strains by multilocus sequence typing (MLST) of the 7 housekeeping genes (*atpA*, *efp*, *ureI*, *ppa*, *mutY*, *trpC*, *ypcC*) and phylogeographic analysis of haplotypes using MEGA and NETWORK software while incorporating DNA sequences and genotyping data of whole *cag* pathogenicity-islands (*cagPAI*). The distribution of *cagPAI* genes within these strains was analyzed by using PCR and the geographic type of *cagA* phosphorylation motif EPIYA was determined by gene sequencing. All the isolates analyzed revealed European ancestry and belonged to *H. pylori* sub-population, hpEurope (Fig. 1). The *cagPAI* harbored by Indian strains revealed European features upon PCR based analysis and whole PAI sequencing. These observations suggest that *H. pylori* strains in India share ancient origins with their European counterparts. Further, non-existence of other sub-populations such as hpAfrica and hpEast Asia, at least in our collection of isolates, suggest that the hpEurope strains enjoyed a special fitness advantage in Indian stomachs, possibly conferred by 'western' type *cagPAIs* to out-compete any endogenous strains. These results also might support



speculations related to gene flow in India through Indo-Aryan nomads bringing Neolithic practices and languages from the Fertile Crescent.



## 2. Genome dynamics of *H. pylori* observed for over a decade in a single host: different evolutions of type IV secretion systems in the antrum and corpus regions of stomach

*H. pylori* corresponds to a highly diverse bacterium, due to frequently occurring mutations and recombinations. Genotypes of specific virulence genes such as *cagA* and *vacA* have been used for a long time as stable entities to link with severity of the *H. pylori* induced gastric diseases. In this study the two genomic islands, namely the *cagPAI*

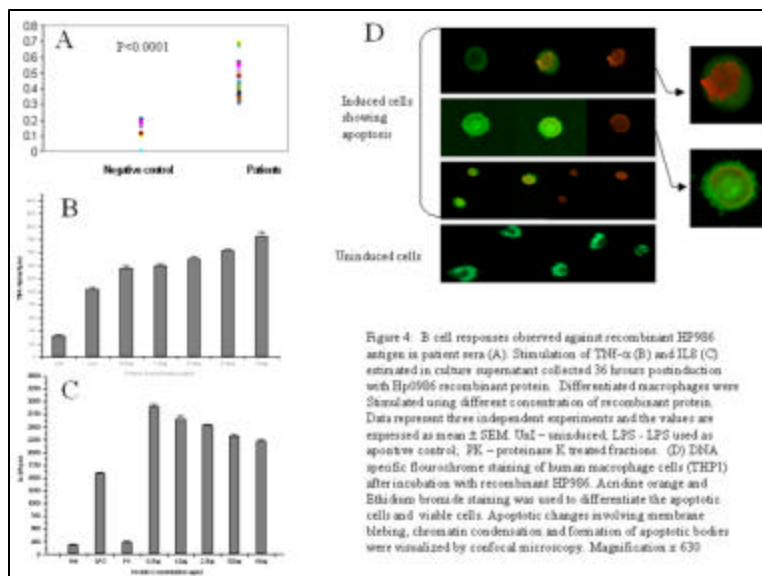
(40kb) and *fts3* (16kb) were completely sequenced and analyzed in isolates obtained at inclusion (one subclone) and after a 10-year period (two subclones) in a patient who had a duodenal ulcer relapse following the failure of eradication therapy. Our results indicate evolution of *cagA*, *cagE* and *cag7* genes of the *cagPAI* (Fig. 2); and ORFs G, P and L in *fts3* (Fig. 3), which possibly lead to inactivation or pseudogenization of these genes. Interestingly, no significant reduction in severity of gastroduodenal pathology (Fig. 2) was found although apparently benign sub-clones were recovered from the patient. These results strengthen the possibility of the involvement of many more virulence factors in the pathology of duodenal ulcers than just the *CagA* and *VacA*. This also points to an obvious difficulty in correlating the continuously evolving virulence factors such as the *cagPAI* genes with the disease characteristics that appear to remain stable.

### 3. Genes from core and flexible genome compartments – putative roles in virulence

Isocitrate dehydrogenase (ICD) of *H. pylori* has been the target of interest in our laboratory for quite some time and we have previously observed that it induces strong B cell responses in gastritis and ulcer disease patients. We continued our efforts further ahead to know the proinflammatory potentials of this housekeeping protein. We found that ICD potentially induces signaling molecules IL-1B and IL-8 and therefore could possibly be linked to the development of chronic gastric inflammation and might have roles in immunomodulation (data not shown). Further work is under progress.

The plasticity region of *H. pylori* genome comprises strain specific loci. We performed genotyping and functional biology of two such loci (JHP940 and HP986) that were previously found to be functionally unknown but present in disease associated strains from many different countries. We found their geographic prevalence to be independent of the presence of the *cagA* gene and the disease status. Cloning, expression and purification of the JHP940 revealed a novel, ~36kDa protein in biologically active form. The HP986 was also expressed and purified to homogeneity. Just like JHP940, the HP986 protein also does not have any

database homology and was therefore, accepted as a novel protein. Both the recombinant proteins induced strong and significant levels of TNF- $\alpha$  and IL-8 in human macrophage cell lines (Fig. 4). Also, HP986 significantly induced NF- $\kappa$ B and therefore, it can be assumed that induction of IL-8 was possibly mediated



through NF- $\kappa$ B. Induction of the proinflammatory cytokines points to the putative role of the plasticity region proteins (which were otherwise assumed as dispensable) in chronic gastric inflammation and various other outcomes of *H. pylori* infection including gastric cancer. In addition, HP986 induced extremely significant B cell responses in gastritis and duodenal ulcer patients as compared to control cases and therefore could be tapped for diagnostic development (Fig. 4). We also found that HP986 significantly induces apoptosis of macrophages after 24 hours of induction with the recombinant protein (Fig. 4). We are in the



process of deducing various signaling pathways involved in induction of apoptosis by HP986. Nonetheless, it appears that this protein might constitute a survival mechanism for

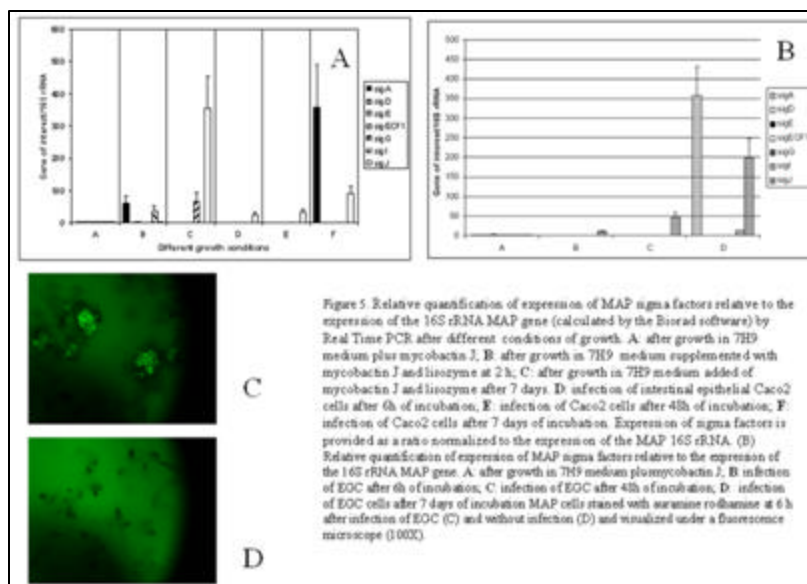
*H. pylori* where immune evasion through killing of macrophages could be the strategy to prolong the inflammatory process leading to extraordinary chronicity of the pathology.

#### 4. Novel MLST method for genotyping of Leptospiral agents

*Leptospira* are the parasitic bacterial organisms associated with a broad range of mammalian hosts and are responsible for severe cases of human Leptospirosis. The epidemiology of leptospirosis is complex and dynamic. Multiple serovars have been identified, each adapted to one or more animal hosts. Adaptation is a dynamic process that changes the spatial and temporal distribution of serovars and clinical manifestations in different hosts. Serotyping based on repertoire of surface antigens is an ambiguous and artificial system of classification of *Leptospira* agents. Molecular typing methods for the identification of pathogenic leptospires up to individual genome species level have been highly sought after since the decipherment of whole genome sequences. Only a few resources exist for microbial genotypic data based on individual techniques such as Multiple Locus Sequence Typing (MLST), but unfortunately no such databases are existent for leptospires.

We report development of a robust MLST method for genotyping of *Leptospira*. Genotyping based on DNA sequence identity of 4 housekeeping genes and 2 candidate genes was analyzed in a set of 120 strains including 41 reference strains representing different geographical areas and from different sources. Of the six selected genes, *adk*, *icdA* and

*secY* were significantly more variable whereas the LipL32 and LipL41 coding genes and the *rrs2* gene were moderately variable. The phylogenetic tree clustered the isolates according to the genome-based species.



The main advantages of MLST over other typing methods for leptospire include reproducibility, robustness, consistency and portability. The genetic relatedness of the leptospire can be better studied by the MLST approach and can be used for molecular epidemiological and evolutionary studies and population genetics.

## **5. Genomics and functional biology of opportunistic mycobacteria**

### **A) Molecular pathogenesis of *M. avium* subsp. *paratuberculosis* infections**

(Work carried out in collaboration with LA Sechi, University of Sassari, Italy)

*Mycobacterium avium* subsp. *paratuberculosis* (MAP) is the etiological agent of Johne's disease (JD), a chronic gastroenteritis of ruminants and other animals, including primates. Increasing evidences have been recently collected on the association of MAP to Crohn's disease, a human chronic granulomatous gastrointestinal disease with strong similarities with JD. The aim of the present study was a first evaluation of global gene regulation in MAP, which has not been addressed previously, despite the availability of MAP genome sequence. For this purpose, we investigated: (i) the presence of sigma factors and their relationship to sigma factors of other mycobacteria (*M. avium* subsp. *avium*, *M. tuberculosis*, *M. bovis*, *M. leprae* and *M. smegmatis*), and (ii) their expression during different growth conditions and *in vitro* infection of intestinal epithelial Caco-2 cells. MAP genome contains 19 putative sigma factor, but only twelve belong to gene families common to other mycobacteria. Gene expression was evaluated with Real-Time PCR during growth in 7H9 medium and mycobactin J, in 7H9 medium plus mycobactin J and lysozyme, and during infection of Caco-2 cells: very different expression patterns were highlighted and, on the whole, only 7 factors were expressed (Fig. 5). Comparative sequence analysis and expression data highlighted some peculiarities of MAP with respect to other mycobacteria, and to *M. tuberculosis* in particular. SigJ was upregulated during Caco2 cell infection in contrast to *M. tuberculosis*. Even if only few sigma factors were expressed in the three conditions tested, the high number of MAP sigma factors suggests a noteworthy flexibility of this pathogen. Thus, this first report on expression of MAP sigma factors opens the way to an extensive characterization of global gene regulation, fundamental to understand its strategies of survival and mechanisms of infections.

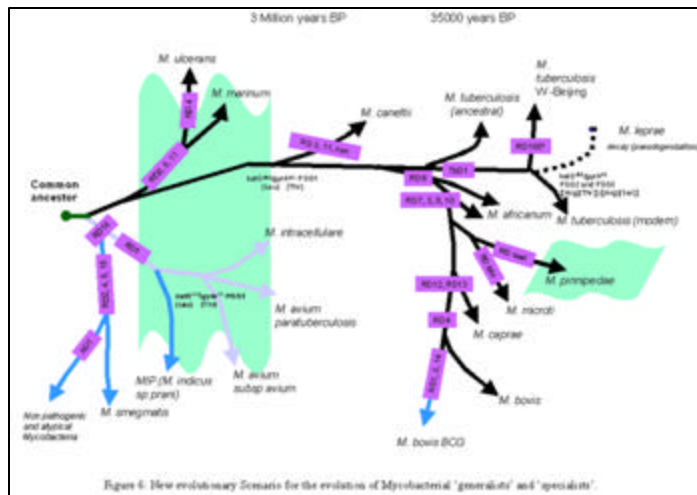
Recent evidence indicates that enteric glia are actively involved in the control of gastrointestinal functions. An active role for enteric glia is being envisaged towards neuronal and neuroimmune communication in the gastrointestinal tract, particularly during inflammation. MAP enters the host through Peyer's patches and intestinal epithelium and sampled by intestinal dendritic cells; they survive inside macrophages. Nothing is known about the role of enteric glial cells during MAP infection. We attempted to establish the role

of enteric glial cells during infection with MAP in Crohn's disease. Map adhesion experiments on enteric glial cells were performed as well as expression analysis of Map sigma factors during infection (Fig. 5). In this study, we found a high affinity of MAP to enteric glial cells and we analyzed the expression of MAP sigma factors under different

conditions of Growth. The fact that MAP showed a high affinity to the glial cells raises concerns about the complicated etiology of the Crohn's disease. Elucidation of the mechanisms whereby inflammation alters enteric neural control of gut functions may lead to novel treatments for inflammatory bowel diseases (IBD).

**B) Comparative genomics of *Mycobacterium indicus pranii* (formerly known as *Mycobacterium w*) portrays new evolutionary scenario for the genus *Mycobacterium***

Mycobacterial pathogens have been one of the most successful groups of infective organisms in human history. Although, genomic alterations are key to virulence optimization



and adaptation, there is a need to understand dynamics of mycobacterial virulence over time. Comparative genome analyses including subtractive hybridization (1), BAC based genome mapping (2) and microarray based approaches (3) have been employed to understand the variability among members of the *Mycobacterium*

*tuberculosis* complex and also to dissect the intricate mechanisms of ancient and contemporary dissemination and evolution of mycobacteria. Lately, several genomic regions of deletion (RDs) have been identified and characterized in *Mycobacterium spp* which are shown to be associated with the attenuation of virulence. Also, these deletions are an important evolutionary paradigm in mycobacteria. Considering the broadly conserved nature of mycobacteria, it has been hypothesized that a single strain may not have sufficiently divergent genetic material to qualify for a species status and, therefore, speciation most likely arises due to specific deletion events. We performed whole genome comparisons together with molecular phylogenetic analyses of pathogenic and non-pathogenic mycobacteria by fluorescent amplified fragment length polymorphism (FAFLP) analysis, enterobacterial repetitive intergenic consensus (ERIC) based genotyping and gene sequencing of candidate orthologues corresponding to *rrn16*, *hsp65*, *sodA*, *rpoB*, *gyrB*, *recA*, *its* and the 32kDa protein gene. These analyses revealed that the soil derived, 'generalist', *Mycobacterium indicus pranii* (MIP), has been the predecessor of *Mycobacterium avium intracellulare* complex (MAIC). Further, our data revealed that MAIC bacilli in their early life history might have shared a common aquatic phase with the early pathogenic forms, well before the latter diverged as 'specialists' (Fig. 6). This evolutionary paradigm may help to understand the acquisition and optimization of virulence in mycobacteria and what determines the boundaries of virulence in such organisms.

## Publications

- 1 Ahmed N, Devi SM, Valverde M de LA, Vijayachari P, Machangu RS, Ellis WA and Hartskeerl RA (2006). Multilocus sequence typing method for identification and

- genotypic classification of pathogenic leptospira species. ***Annals of Clinical Microbiology and Antimicrobials***5:28.
- 2 Devi SM, Ahmed I, Khan AA, Rahman SA, Alvi A, Sechi LA and Ahmed N (2006) Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type cag-pathogenicity island. ***BMC Genomics***7:191.
  - 3 Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Loch C and Supply P (2006) Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India suggests an ancient focus of tuberculosis in South Asia. ***Emerging Infectious Diseases***12:367-374.
  - 4 Rao KR, Ahmed N, Sriramula S, Sechi LA and Hasnain SE (2006) Rapid identification of *Mycobacterium tuberculosis* Beijing genotypes on the basis of the mycobacterial Interspersed repetitive Unit locus 26 signature. ***Journal of Clinical Microbiology*** 44:274-277.
  - 5 Sechi LA, Gazouli M, Sieswerda L, Mollicotti P, Ahmed N, Ikonomopoulos J, Scanu AM, Paccagnini D and Zanetti S (2006). Relationship between Crohn's disease, infection with *Mycobacterium avium* subspecies paratuberculosis and Slc11a1 gene polymorphisms in Sardinian patients. ***World Journal of Gastroenterology*** 12:7161-7164.
  - 6 Zanetti S, Mollicotti P, Cannas S, Ortu S, Ahmed N and Sechi LA (2006). *In vitro* activities of antimycobacterial agents against *Mycobacterium avium* subsp. paratuberculosis linked to Crohn's disease and paratuberculosis. ***Annals of Clinical Microbiology and Antimicrobials***5:27.
  - 7 Akhtar Y, Ahmed I, Devi SM and Ahmed N (2007). The co-evolved *Helicobacter pylori* and gastric cancer: Trinity of bacterial virulence, host susceptibility and lifestyle. ***Infectious Agents and Cancer***2:2.
  - 8 Ahmed N, Majeed AA, Ahmed I, Hussain MA, Alvi A, Devi SM, Rizwan M, Ranjan A, Sechi LA, and Megraud F (2007) genoBASE pylori: A genotype search tool and database of human gastric pathogen *Helicobacter pylori*. ***Infection Genetics and Evolution*** (InPress).

- 9 Banerjee S, Nandyala AK, Raviprasad P, Ahmed N and Hasnain SE (2007) Iron dependent Iron binding activity of Mycobacterium tuberculosis aconitase. ***Journal of Bacteriology*** (InPress).
- 10 Salih B, Abasiyanik F, Ahmed N (2007). A preliminary study on the genetic profile of cag pathogenicity island and other virulent gene loci of Helicobacter pylori strains from Turkey. ***Infection Genetics and Evolution*** (InPress).
- 11 Sechi LA, Ruhel A, Ahmed N, Usai D, Paccagnini D, Felis G and Zannetti S (2007) *Mycobacterium avium* subspecies paratuberculosis is able to infect and persist in enteric glial cells. ***World Journal of Gastroenterology*** (InPress).



**Laboratory of Structural Biology**  
**Structural and biochemical characterization of some**  
***M. tuberculosis* Proteins**

<b>Principal Investigator</b>	Shekhar C. Mande	Staff Scientist
<b>Ph D Students</b>	Kshama Goyal	Senior Research Fellow (Till Jan 07)
	Mohd Akif	Senior Research Fellow
	C M Santosh Kumar	Senior Research Fellow
	N Madhav Rao	Senior Research Fellow
	Debashree Basu	Senior Research Fellow
	Pramod Kumar	Junior Research Fellow
	Shubhada Hegde	Junior Research Fellow
<b>Other Members</b>	A Sheeba	Technical Assistant
	Soghra Haq	Project Associate
	Y Sailu	Senior Project Assistant (Till Aug 06)
	P Manimaran	Senior Project Assistant
	T N Jayashri	DBT Post Doctoral Fellow
	Ashutosh Shukla	Project Assistant
	Roshna Gomez	Project Assistant
<b>Collaborators</b>	Sayed E. Hasnain	University of Hyderabad
	Ranjan Sen	CDFD, Hyderabad
	Anil Tyagi	University of Delhi South Campus)
	Abhijit A. Sardesai	CDFD, hyderabad

**Objectives:**

- 1 Identify important proteins of *Mycobacterium tuberculosis* for crystallographic and biochemical analysis
- 2 Express the chosen proteins and characterize them biochemically
- 3 Attempt to crystallize the chosen proteins

**Summary of work done until the beginning of this reporting year**

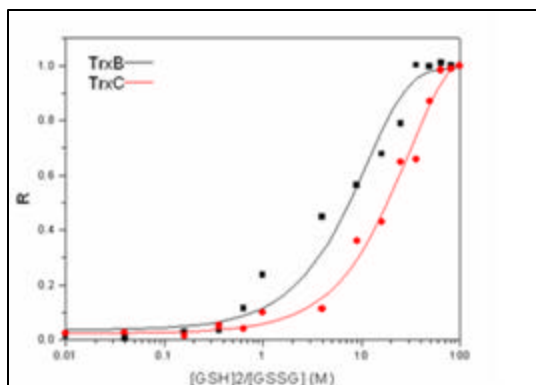
Three broad categories of proteins from *Mycobacterium tuberculosis* were chosen for biochemical and structural work. The three categories are:

Redox proteins:	<ul style="list-style-type: none"> <li>• Alkylhydroperoxide reductase (AhpC)</li> <li>• Thioredoxins (A, B and C)</li> <li>• Thioredoxin reductase</li> </ul>
Heat shock proteins:	<ul style="list-style-type: none"> <li>• Chaperonin-60 family (Cpn60.1, Cpn60.2 and</li> </ul>



	Cpn10) <ul style="list-style-type: none"> <li>Heat shock protein 70 family (Hsp70, Hsp40)</li> </ul>
Other protein including proteins involved in transcription processes:	<ul style="list-style-type: none"> <li>Transcription cleavage factor GreA</li> <li>Transcription termination factor Rho</li> <li>Sigma factors E and H</li> <li>Anti-sigma factors E and H</li> <li>HrcA and HspR: transcriptional repressors of the Hsp60 and Hsp70 operons respectively</li> </ul>

- The alkylhydroperoxidase reductase (AhpC) protein was purified and characterized biochemically. Reaction mechanism of this protein was proposed based on site-directed mutants. Crystals of this protein were grown, and diffraction data were collected. The crystals were found to be highly twinned.
- Chaperonin-60 gene was cloned and the protein was purified to homogeneity. Crystals of Chaperonin60.2 were grown and data were collected upto 3.2Å resolution. The crystal structure of the protein was reported at this resolution.



**Figure 1. Measurement of redox potential of *M. tuberculosis* TrxB and TrxC.** The fraction of reduced protein as a function of redox equilibrium buffer was measured by intrinsic Trp fluorescence. Redox potential was calculated using Nernst's equation using the value of  $K_{eq}$  obtained. These measurements showed that the redox potential of TrxB was  $-290$  mV while that of TrxC was  $-310$  mV.

- Thioredoxin reductase was cloned, purified to homogeneity and crystals were grown. Diffraction data upto 3Å resolution were collected. The crystal structure was determined and reported at this resolution.
- Crystals of *M. tuberculosis* chorismate mutase were grown and the crystal structure was determined to 2.1Å resolution.

**Details of progress made in the current reporting year (April 1, 2006 - March 31, 2007)**

**Thioredoxins and Thioredoxin reductase:**

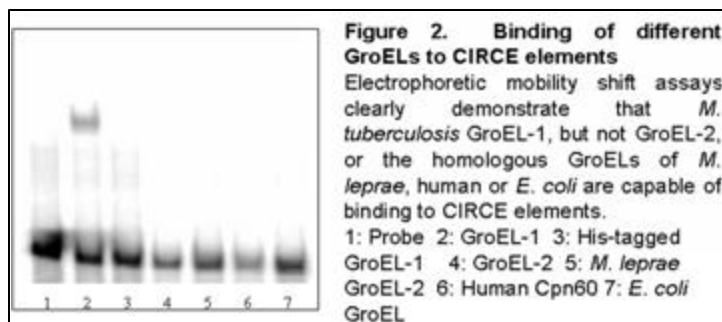
During the period under review, the three thioredoxins (Trx's A, B and C) encoded by the Rv1470, Rv1471, and Rv3914 were cloned and expressed in *E. coli*. All the three proteins

were purified to homogeneity. Size exclusion chromatography showed that TrxA, TrxB and TrxC are all dimeric proteins. Insulin reduction assays showed that all the three proteins were capable of reducing insulin in the presence of a non-physiological reductant, DTT. TrxA was seen to be much slower in this assay than the other two Trx's. Further it was also

tested if all the three Trx's are substrates of TrxR by reduction of insulin in the presence of physiological reductants, NADPH and TrxR. It was observed that TrxB and TrxC are the substrates of TrxR, however, TrxA is not a substrate of TrxR. Multiple sequence alignments showed that although all the three Trx's possess the conserved CXXC motif, TrxA does not possess a Trp residue preceding this motif, which is highly conserved among all thioredoxins. Moreover, some of the other residues known to be important for the activity of Trx's, such as Asp26 (*E. coli* nomenclature), were also replaced in TrxA. Thus, subtle sequence variations were ascribed to the lack of functional activity of TrxA. RT-PCR showed that TrxB and TrxC are expressed at varying levels under different oxidative conditions, while TrxA is not expressed. These results suggest that *trxA* might be a pseudogene, while TrxB and TrxC might be the canonical thioredoxins.

The redox potential of TrxB and TrxC measured through fluorescence is  $-290\text{mV}$  and  $-310\text{mV}$  respectively (Fig. 1). Between the two thioredoxins, TrxC is not as readily reduced by TrxR, which might be explained by its lower redox potential. Since thioredoxins are the primary reductants in the cytoplasm, our results suggest that the overall redox potential of the *M. tuberculosis* cytoplasm might be lower than that of other bacteria, thereby offering a possible explanation to why *M. tuberculosis* is more resistant to oxidative killing.

**Chaperonins:** We had demonstrated earlier that *M. tuberculosis* GroELs are dimeric in nature, and that they do not exhibit canonical chaperonin-like properties. These results seemed to suggest that they might have lost their chaperoning ability. This hypothesis was tested in the period under review using heterologous complementation and biochemical studies. Our results suggest that the loss of chaperoning ability of the mycobacterial GroEL-1 is indeed due to its inability to form canonical tetradecamers. A chimeric *groEL-1* ORF bearing the DNA sequence corresponding to the equatorial domain of *E. coli groEL*, unlike the unmodified *groEL-1* ORF, is able to complement the loss of cell viability phenotype conferred by a conditional allele of *groEL*, in *E. coli*. The corresponding polypeptide *in vitro* is capable of existing principally in a tetradecameric state, a canonical feature of *E. coli* GroEL. Complementary studies show that an *E. coli* GroEL variant, displaying properties similar to that of *Mtb* GroEL-1 can be obtained provided it bears the amino acid sequence corresponding to *Mtb* GroEL-1 equatorial domain. Analyses of various *groEL* alleles



obtained via DNA shuffling and error prone PCR starting with the *Mtb groEL-1* and *groEL-2* DNA as templates suggest that

the basic property of a functional GroEL molecule appears to be oligomerization, to facilitate formation of the cavity for sequestration of substrate polypeptides. The substrate -recognizing apical domain may be fairly plastic and is capable of tolerating deletions or insertions. These experiments provide direct evidence for the importance of oligomerization in biologically relevant GroEL function.

The inability of *M. tuberculosis* GroELs to act as molecular chaperones is intriguing, especially since the genes encoding them are not known to be pseudogenes. Considering the view that the fundamental property of a molecular chaperone is that of substrate binding with broad specificity, we explored if GroEL-1 has retained the ability to bind different substrates. Interestingly we observed that GroEL -1 is capable of binding to DNA without any sequence specificity (Figure 2). The affinity of DNA recognition by GroEL -1 is sufficiently

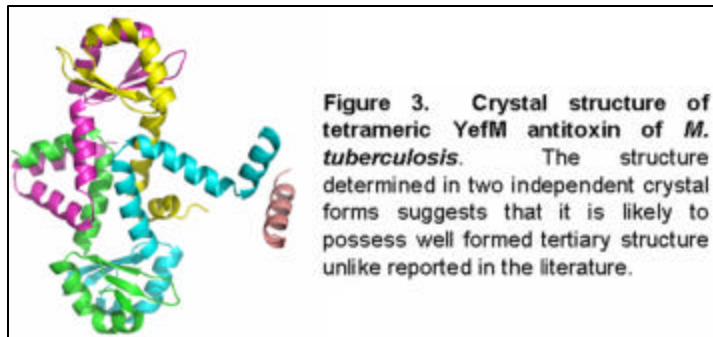
high in the range of 100-200 nM suggesting that the protein has naturally evolved to bind DNA. Testing a range of oligonucleotide sequences of varying lengths

Crystallographic data	Native	Selfet
Unit cell parameters		
a (Å)	65.0	65.1
b(Å)	64.6	65.1
c (Å)	83.5	83.0
Spacegroup	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P4 <sub>2</sub> 2 <sub>1</sub> 2
Resolution range (Å )	41.77 - 2.50	51.23 - 3.5
Completeness(%)	97.7(100)	100(100)
Mean I/σ(I)	22.3	33.3
Rmerge(%)	5.9	8.1(26.4)
Multiplicity	25.6	7.2
Anamolous completeness(%)		100(100)
Anamolous multiplicity		15(15)
No. of unique reflections	11837	

Number of protein atoms	2515
Number of subunits per asymmetric unit	4
Number of atoms per asymmetric unit	2563
Number of waters per asymmetric unit	59
Resolution range (Å)	19.98-2.5
R <sub>int</sub> (%)	20.7
R <sub>free</sub> (%)	25.4
Rmsd from ideal values	0.014
bond distances(Å)	1.604
bond angles(deg)	
Average B-factor(Å <sup>2</sup> )	
overall	42.0
main chain	41.1
side chain	42.9
Water	33.6
sulfate ion	47.3
Prochek Ramachandran plot (%)	
core	95.5
allowed	4.5
Molprobability	
clashscore(all atoms)	1.19(100th percentile)
clashscore(B<40)	0.84(94th percentile)
rotamer outliers	11.57
C <sup>2</sup> deviations of > 0.25Å	5

revealed that GroEL-1 binds DNA in a sequence-non-specific manner. Treatment of DNA against Dnasel or hydroxyl-radicals suggests that GroEL -1 can effectively function as a DNA protecting agent. Moreover, atomic force microscopic studies reveal that it can condense large DNA molecules into a compact structure. Interestingly, the *M. tuberculosis* genome reveals absence of many non-specific DNA binding nucleoid-associated proteins, such as H-NS. Thus, the *groEL-1* gene of *M. tuberculosis* might have acquired a nucleoid association function through loss of oligomerization, by utilizing its fundamental property as a non-specific substrate binding protein. Since the loss of nucleoid associated proteins has been correlated to defect in biofilm formation in other bacteria, our hypothesis that GroEL -1 might

be a nucleoid associated protein offers an effective explanation to the intriguing observation that *groEL-1* mutants of *M. smegmatis* are defective in biofilm formation.



**Other projects:** During the period under review, we have been able to determine the crystal structure of YefM antitoxin of *M. tuberculosis*. The 10kDa YefM antitoxin was crystallized in two different space groups under two

different crystallization conditions (Table 1). The structure was determined through Single Anomalous Dispersion (SAD) technique using Se-Met substituted proteins. The high resolution structure has been refined at 2.5 Å resolution with the final R and R<sub>free</sub> being 0.207 and 0.254 respectively (Table 2).

Observation of its well-formed tertiary structure (Fig 3) was unanticipated since the antitoxins are reported to be intrinsically unstructured proteins, whereas they are believed to assume structure only in the presence of their cognate toxin partners. Analysis of our structure and homologous sequences, however, suggests that YefM in all species is likely to possess well-formed tertiary structure.

### ***Publications:***

- 1 Akif M, Akhter Y, Hasnain SE and Mande SC (2006) Crystallization and preliminary X-ray crystallographic studies of *Mycobacterium tuberculosis* CRP/FNR family transcription regulator. ***Acta Crystallographica F*** 62:873-875.
- 2 Goyal K, Qamra R and Mande SC (2006) Multiple gene duplication and rapid evolution in the *groEL* gene: functional implications. ***Journal of Molecular Evolution*** 63:781-787.
- 3 Qamra R, Prakash P, Aruna B, Hasnain SE and Mande SC (2006) The 2.15Å crystal structure of *M. tuberculosis* chorismate mutase reveals unexpected gene duplication and suggests a role in host-pathogen interactions. ***Biochemistry*** 45:6997-7005.

- 4 Vidyasagar M, Hasnain SE, Mande SC, Nagarajaram HA, Ranjan A, Acharya MS, Anwaruddin, Arun SK, Gyanraj Kumar, Kumar D, Priya S, Ranjan S, Reddi BR, Seshadri J, Sravan Kumar P, Swaminathan S, Umadevi P and Vindal V (2007) BioSuite: A comprehensive bioinformatics software package (A unique industry-academia collaboration) **Current Science** 92:29-38.
- 5 Goyal K, Mohanty D and Mande SC (2007) PAR-3D: A server to predict protein active site residues. **Nucleic Acids Research** (InPress).
- 6 Sailu Y, Goyal K and Mande SC (2007) Inferring genome-wide functional linkages in E. coli by combining improved genome context methods: comparison with high-throughput experimental data. **Genome Research** (InPress).

## Laboratory of Mammalian Genetics Epigenetic mechanisms underlying developmental pathways

### Principal Investigator

Sanjeev  
Khosla

Staff Scientist

### Ph D Students

*Sudhish Sharma*  
*G Gokul*  
*Devi T*  
*Divya Tej*

Senior Research Fellow  
Senior Research Fellow  
Junior Research Fellow  
Senior Research Fellow

### Other Members

*S Malathi*  
*B Gautami*  
*P Mallikarjuna*  
Neeraja C

Project Assistant (till Jan 2007)  
Project Assistant  
Project Assistant  
Project Assistant

### Collaborators

Gayatri Ramakrishna  
Rakesh Mishra  
K Nagamohan  
Aleem A Khan

CDFD, Hyderabad  
CCMB, Hyderabad  
CHG, Bangalore  
CLRD, Hyderabad

### Objectives:

Specialized chromatin structures as epigenetic imprints to distinguish parental alleles

### Summary of work done until the beginning of this reporting year

To elucidate the mechanisms that underlie the establishment and maintenance of parental-allele-specific specialised chromatin organisation we have been examining a novel imprinted gene, *neuronatin*, present on mouse chromosome 2. Previously, we reported a transcription-independent parental-allele-specific DNase I hypersensitive site within the second intron of the *neuronatin* gene. Furthermore, methylation-restricted protein binding was observed to a GC-rich motif within this intronic region. Methylation-restricted protein binding along with nucleosomal positioning on the unmethylated paternal allele suggested a role of protein binding in regulating the imprinting of the *neuronatin* locus via chromatin organization. In the last report we showed that tissue-specific histone modifications were associated with the promoter and the second intron of *neuronatin*. Interestingly, for the unmethylated paternal allele it was the histone modification profile for the intron and not the promoter which correlated with the transcription of the gene. Furthermore, in a functional assay we had shown the second intron to be a directional inhibitor of transcription.

## Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)

Yeast One-hybrid assay for identifying GC-binding proteins: We had earlier identified a 38 bp GC-rich region within the second intron of neuronatin where methylation-restricted protein binding was observed. To identify protein(s) which bind to this motif we performed yeast one-hybrid analysis using the 38 bp GC-rich sequence from within the intronic region as bait in the assay. As the proteins involved in determining imprinting status of a gene are expected to be expressed ubiquitously, a mouse testis specific cDNA library was used to identify proteins which bind to this motif. Seven of the positive clones (mentioned below) were found to be proteins with possible role in chromatin organization.

Clone Name	Gene Name	Motifs
YMH16-2	<i>Baz1a</i>	Bromodomain, PHD
11M5	<i>Phf20</i>	PHD
YMH17-1	<i>Phf7</i>	PHD
YMH25-2	<i>Pabpc2</i>	RRM domain
YMH34-1	<i>Zfp488</i>	Zn finger
YMH38-1	<i>Cbx5</i>	Chromo domain
YMH39-1	<i>Phlpp</i>	PHD

Amongst the seven, three proteins *phf7*, *phf20* and *cbx5a* were taken up for characterization initially. To confirm the interaction with GC, it was decided to over-express these proteins in the yeast *pichia pastoris* rather than *E coli* as this yeast strain is known to provide for posttranslational modifications which are similar to those found in higher mammals. In our

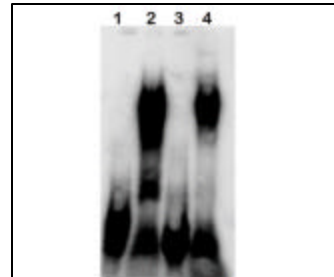


Figure 1: EMSA analysis for *phf7*, *phf7* protein, overexpressed in *pichia pastoris* was analysed for binding with either GC or methylated GC fragment. 1. GC fragment 2. GC + *phf7*; 3. Methylated GC; 4 Methylated GC + *phf7*.

preliminary experiments we have been able to clone *phf7* and *cbx5* into pPICZ $\alpha$ A vector. As the protein Since the aim was to identify proteins which show a methylation-restricted binding to the GC motif, the overexpressed proteins were used in EMSA assay to examine their DNA binding ability. As can be seen from figure 1, *phf7* showed binding to the GC fragment. However, its binding was observed to both the methylated and unmethylated probes. We are in the process of analyzing *cbx5* and *phf20* by the same assay for GC binding.

**Functional analysis of *neuronatin*'s 250 bp second intron:** We have previously shown this 250 bp intron to be a directional inhibitor of transcription in transient transfection assay in mammalian cells. To analyse its function *in vivo*, the intronic region cloned in a plasmid upstream of drosophila *miniwhite* gene was integrated at various loci in the drosophila genome using P-element transposition strategy. The aim was to examine whether the 250 bp second intron of *neuronatin* has any regulatory effect on the reporter gene (which codes



for the eye color). We have generated several drosophila lines which showed effect of intronic region on the reporter gene. To confirm that the effect is due to the intronic region and not because of the chromatin context at the integration site, the intronic regions are being flipped out of the site using *loxP* sites flanking the region. Following the confirmation, we plan to perform genetic crosses with known chromatin-affecting mutations to analyse the possible trans-acting molecules which cooperate with the cis-sequences within the intronic region in influencing the transcription of the reporter gene.

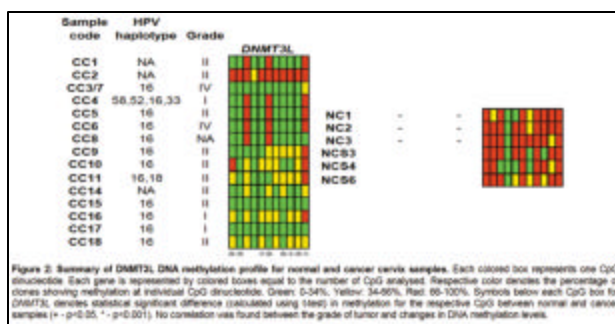
### 1) DNMT3L: epigenetic correlation with cancer

#### Summary of work done until the beginning of this reporting year

To date, no one type of DNA mutation (genetic) has been defined which can be correlated with all cancers, but in most cancers, aberrant methylation, an epigenetic modification, has been observed. Previously, we had reported comparison of DNA methylation profile of a few reprogramming genes like DNA methyltransferases, histone modifiers, etc and stem cell specific genes like *Oct4*, *Nanog*, etc. between normal and cervical cancer samples. Our pilot study had shown interesting DNA methylation profile for *DNMT3L*, a reprogramming molecule belonging to the family of DNA methyltransferases but which lacks the catalytic domain. We had observed remarkable difference between the normal and cancer cervix groups.

#### Details of progress made in the current reporting year (April 1, 2006 - March 31, 2007)

Reprogramming molecules like DNA and histone methyltransferase, histone acetylases and deactylases are the epigenetic effector molecules which can reprogram genetic information. These molecules are essential for normal development as any change to the epigenetic status of haplotypic loci in response to an environmental cue would have to be perpetuated through these reprogramming molecules. The possibility, therefore, exists that in the altered environment of cancer cells, epigenetic status of the genome undergoes changes due to deregulation of these genes. This deregulation could either be due to genetic mutations within these genes or due to changes in their epigenetic profile. We sought to explore the possibility that the reprogramming genes themselves undergo epigenetic modifications



reflecting their changed transcriptional status during cancer development. To examine this possibility, we had undertaken a pilot study, wherein, we analysed DNA methylation status for the promoters of a few reprogramming

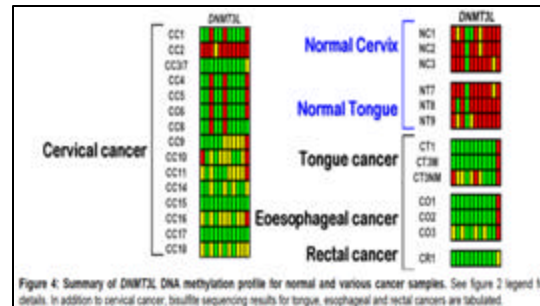
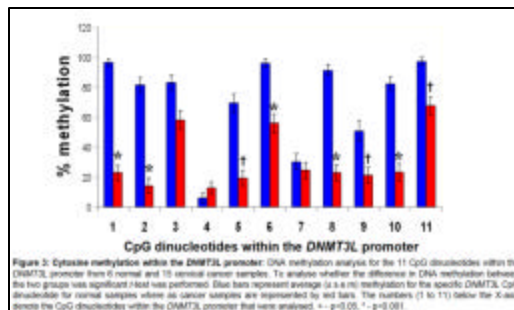


genes in a small number of normal and cancer cervix samples collected from city based cancer hospital of Hyderabad, India.

As reported earlier, we had observed DNA methylation difference between normal and cancer cervix samples for the promoter of *DNMT3L*. A nuclear reprogramming related gene, *DNMT3L* regulates the DNA methylation function of *de novo* methyltransferases, *DNMT3A* and *DNMT3B*. The difference in DNA methylation at the promoter of *DNMT3L* is significant as *DNMT3L* is a regulator of *de novo* methyltransferases and expression of this molecule in tissues where it is normally not expressed could in turn stimulate *DNMT3A/3B* activity. This could cause aberrant DNA methylation (changes in the level of methylation as well as DNA methylation at sites which are normally not methylated).

### 1 DNMT3L promoter methylation: a potential cancer biomarker

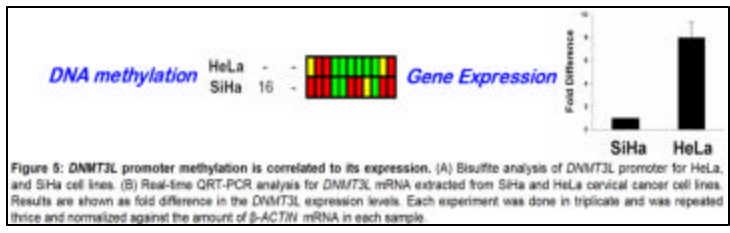
To confirm our results from the pilot study and to examine the possibility that *DNMT3L* promoter methylation could be a potential cancer biomarker we extended our analysis on 10 more cervical cancer samples. As can be seen from figure 2, 14 out of 15 (93%) cervical cancer samples had a DNA methylation profile different from the normal samples. Whereas



CC3, CC15 and CC17 showed complete loss of methylation for the *DNMT3L* promoter, we found that 11 out of the other 12 cancer samples also had a methylation profile different from normal cervix samples. Comparison of the two groups for DNA methylation using t-test showed that the difference at 8 of the 11 CpG's was significant ( $p < 0.05$ , denoted by + below each CpG for *DNMT3L*, figure 2 and 3). For 5 out of these 8 CpG's this difference was highly significant ( $p < 0.001$ , denoted by \* in figure 2 and 3). Preliminary comparison of *DNMT3L* promoter DNA methylation between normal and cancer samples from tongue, Esophageal and rectal tissues show similar differences (figure 4).

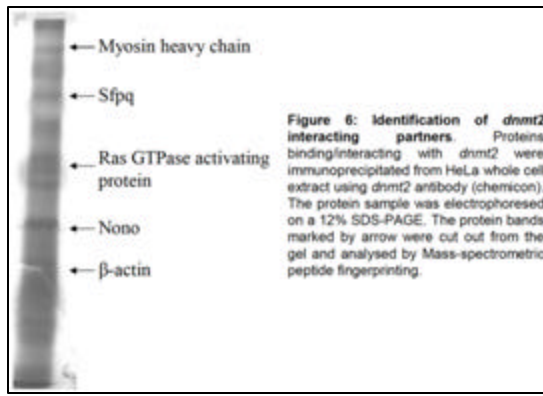
### 2) Characterisation of epigenetic signatures at the DNMT3L loci: Correlation of DNA methylation changes at the DNMT3L promoter with expression level.

The difference in the DNA methylation at the promoter of *DNMT3L* is significant as *DNMT3L* is a regulator of *de novo* methyltransferases and expression of this molecule in tissues where it is normally not expressed could in turn activate *DNMT3A*. This could cause aberrant DNA methylation (changes in the level of methylation as well as DNA methylation at sites which are normally not methylated). Hypomethylation is usually correlated with expression and hypermethylation with repression of a gene. We were unable to analyse the expression level of *DNMT3L* in the cancer and normal samples due to the limitation of specimen available but we found similar correlation of DNA methylation within *DNMT3L* promoter with



its expression in cervical cancer cell lines. As can be seen from figure 5, HeLa cell line, which had lost methylation at several CpG's within the *DNMT3L* promoter,

showed 8 fold more expression than SiHa cell line (promoter is predominantly methylated). However, the difference in DNA methylation may not be the only epigenetic modification correlated with *DNMT3L* expression. It is also important to note that a regulator of epigenetic effector molecule (*DNMT3L*) showed changed methylation pattern upon cancer development whereas the effector molecules (*DNMT3A AND DNMT3B*) did not show any change.



## 2) Dnmt2 and non-CpG methylation.

**Details of progress made in the current reporting year (April 1, 2006 - March 31, 2007)**

*Dnmt2* is a highly conserved protein of 320 aminoacids with C-terminal catalytic domain characteristic of other methyl transferases and lacking regulatory N-

terminal region which the other methyl transferases seemed to possess. Though it has all the domains specific for methyl transferases, *Dnmt2* has failed to show significant methylation *in vitro* and *in vivo* conditions. In drosophila and dictyostelium, *dnmt2* has been linked with non-CpG methylation. However, its role in mammals has been a mystery, though recently, *Dnmt2* was shown to methylate tRNA in mice. Our laboratory is interested in examining the role of *dnmt2* in mammalian cells. In particular, we would like to examine whether *dnmt2* is part of the non-CpG methylation machinery.

### Identification of *dnmt2*-interacting partners

To characterize the putative function of *dnmt2*, we decided to identify protein(s) with which *dnmt2* interacts. We have undertaken the following two different approaches to identify *dnmt2*-interacting partners: a) Yeast two-hybrid assay; b) Co-immunoprecipitation assay using *dnmt2* antibodies. The results of the yeast two-hybrid analysis were presented in the last year report. The co-immunoprecipitation assay was done on HeLa whole cell extract using *dnmt2* antibodies. The bound proteins were separated on a 12% SDS-PAGE (figure 6). A few prominent protein bands were cut out from the gel and analysed by MALDI-TOF peptide fingerprinting for their identity (done at TCGA, New Delhi). As shown in figure 6, the protein bands were identified as; Nono, SpfQ, Myosin heavy chain, Beta actin, Rho-GTPase activating protein. Interestingly, previous studies have shown the identified proteins to be interacting with one another and part of the eukaryotic transcription machinery. To confirm the interaction of *dnmt2* with the proteins identified in the co-immunoprecipitation and yeast two-hybrid analysis, mammalian two-hybrid analysis has been undertaken. We are also performing reverse co-immunoprecipitation to examine if *dnmt2* is bound to the identified proteins.

### **3) Epigenetic mapping of the human Y-chromosome (new project)**

#### **Details of progress made in the current reporting year (April 1, 2006 - March 31, 2007)**

Epigenetic modifications of the human genome have an important role in the maintenance of chromatin states distinctive to each cell type and, as a consequence, in the expression of cell type-specific genes. Abnormalities in epigenetic modifications have been demonstrated in several types of cancer and in some inherited human diseases, such as the Rett syndrome. Recent studies have also led to hypotheses linking epigenetic alterations to certain complex human diseases such as diabetes, in which there is a high rate of discordance between monozygotic twins. Despite the intense interest in the roles that epigenetic modifications might play in health and disease, the epigenetic features of the human genome are only now beginning to be characterized.

In collaboration with CCMB, Hyderabad and CHG, Bangalore, we have undertaken a pilot project to determine DNA methylation patterns as well as chromatin modification patterns (histone H3-lysine 9 di methylation and histone H3-K9 acetylation) of all genes on human Y chromosome (~320 genes). Within this project, our laboratory's focus is on determining the DNA methylation status of the promoters of 160 genes present on the Y chromosome. The analysis was done using bisulfite sequencing assay on DNA isolated from human blood. We have completed bisulfite PCR on all the 160 gene promoters and methylation profile has

already been generated for more than 100 of these gene promoters. As expected most of the analysed Y chromosome genes showed hypermethylation of their promoters in blood

DNA. Further work is in progress to collate bisulfite sequencing and histone modification Chromatin immunoprecipitation results for all the Y-chromosome genes on other samples.

***Publications:***

- 1 Khosla S, Mendiratta G and Brahmachari V (2006) Genomic imprinting in the mealybugs. ***Cytogenetics and Genome Research*** 113:41-52.
- 2 Gokul G, Gautami B, Malathi S, Sowjanya A, Poli U, Jain M, Ramakrishna G and Khosla S (2007) DNA methylation profile at the DNMT3L promoter: a potential biomarker for cervical cancer. ***Epigenetics*** (InPress).



## Laboratory of Molecular Oncology

### Understanding the biology of cancer using a genomic and molecular genetic approach and molecular analyses of disease-causing mutations in human genetic disorders

<b>Principal Investigator</b>	Murali Dharan Bashyam	Staff Scientist
<b>Ph D Students</b>	R Ratheesh	Senior Research Fellow
	M Khursheed	Junior Research Fellow
	Sreejata Chatterjee	Junior Research Fellow
<b>Other Members</b>	Sarita Ranjan	DBT Post Doc
	G R Savithri	Technical Officer
	Ajay Chaudhari	Technical Assistant
	Chandrakanth Reddy	Project Assistant
	G Purushotham	Project Assistant
	Siddharth Singh	Project Assistant
	K Viswakalyan	Project Assistant
	E Mahesh	Project Assistant
	K Madhumohan	Project Assistant
<b>Collaborators</b>	A Radha Rama Devi	CDFD, Hyderabad
	H A Nagarajaram	CDFD, Hyderabad
	Jonathan R Pollack	Stanford University, USA
	G Swarnalata	Apollo Hospitals, Hyderabad
	M Srinivasulu	MNJ Hospital, Hyderabad
	R A Sastry	NIMS, Hyderabad
	Anjaneyulu	MNJ Hospital, Hyderabad
	K Chandrashekhar Rao	Apollo Hospitals, Hyderabad
	Calambur Narsimhan	Care Hospital, Hyderabad
	V Hariram, Usha Mullapudi	Cardiac Centre, Hyderabad
	Dr Prameela	Centre for Human Genetics, Bangalore
	Anirban Maitra,df	Johns Hopkins University

#### **Project 1a: Identification of novel localized high-level amplifications and homozygous deletions in pancreatic/biliary cancer**

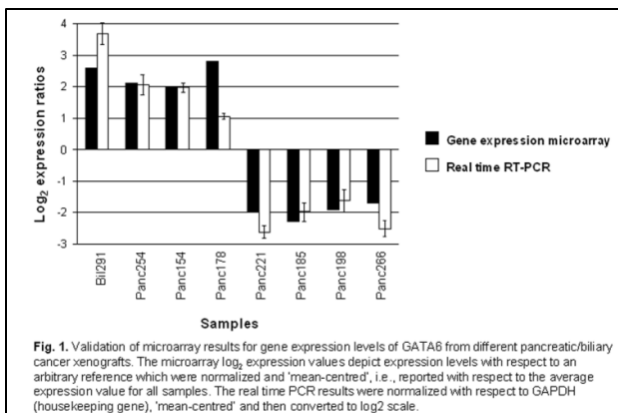
##### **Objective**

To identify and characterize novel oncogenes and tumor suppressor genes for pancreatic/biliary cancer

## Summary of work done until the beginning of this reporting year

In collaboration with Dr. Jonathan R Pollack, Department of Pathology, Stanford Univ Sch Med, and Dr Anirban Maitra, Johns Hopkins University, array comparative genomic hybridization (aCGH) performed on pancreatic/biliary cancer mouse xenografts were analyzed. We identified novel regions of localized high-level amplifications and homozygous deletions presumably harboring as yet uncharacterized oncogenes and tumour suppressor genes that may be important for pancreatic/biliary cancer initiation and progression. In addition, an analysis of gene expression data carried out using the same microarray platform on which aCGH was performed, enabled us to prioritize the 'driver' genes within amplicons for specific samples.

## Details of progress in the current reporting year (April 1, 2006 - March 31, 2007)



We have now initiated functional characterization of selected genes that lie within amplicons and deletions. One such gene is the transcriptional activator, GATA6, which has been shown to be important for the early development of pancreas. Using simple statistical approaches we endeavoured to determine transcriptional targets for GATA6. An authentic GATA6 target is

expected to exhibit elevated expression levels in samples where GATA6 is over expressed and vice versa. Using real time RT-PCR we first validated the expression level of GATA6 obtained from gene expression microarray experiments for different pancreatic/biliary cancer samples (Fig. 1). We employed the following criteria for selection of GATA6 transcriptional targets from the gene expression array data; a Pearson's correlation coefficient (R) greater than 0.65, significant variation of gene expression in the two data sets (two-tailed student's t-test, p value <0.05) and more than two-fold difference in average expression between the two data sets. We also carried out 1000 random permutations to provide an estimate of the false discovery rate (FDR) and we obtained an FDR of 15%. In a parallel approach, the same array data set for the ten samples were subjected to Supervised Analysis of Microarrays (SAM) that is a well validated statistical package to select statistically significant genes that can distinguish between two sets of data sets, with an FDR of 1.68%. Using these complementary approaches we have identified several novel genes that were hitherto unknown to be transcriptional targets for GATA6 and one or a few of these could have a

particular significance for pancreatic/biliary cancer. We are currently carrying out transcriptional assays to validate the GATA6 targets in pancreatic cancer cell lines and functional assays to determine their role in pancreatic/biliary cancer.

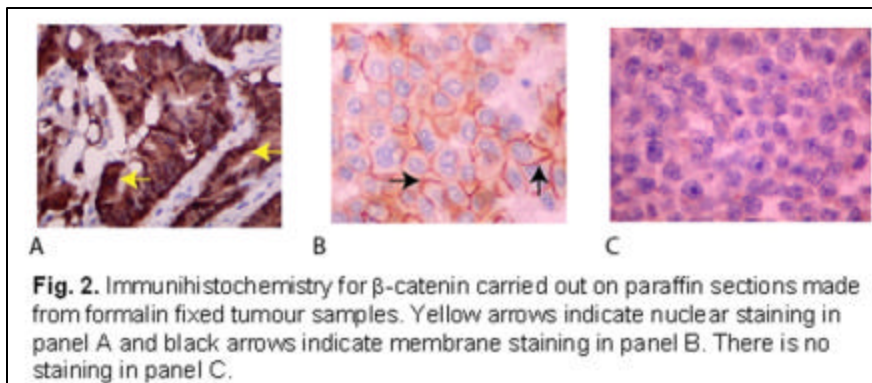
### **Project 1b: Molecular genetic analyses of sporadic colorectal cancer occurring in the young**

#### **Objectives:**

1. Screening young colorectal cancer (CRC) patients for chromosomal instability (CIN) and microsatellite instability (MSI) and comparison with older patients.
2. Identifying important copy number alterations (CNAs) as well as important genes located within the CNAs from non-familial young CRC patients and comparison with older patients.
3. Identification of pathways that might distinguish young non-familial CRC patients from elder patients.

#### **Summary of work done until the beginning of this reporting year**

In India and other Asian countries, a high incidence of CRCs has been reported to occur in the young (age below 50 yrs) as against the western population. In addition, these cancers are non-familial and do not appear to share a common etiology including socio-economic status of patients, association with specific infectious agents and/or chronic syndromes such as ulcerative colitis, etc. A majority of these young patients harbour an aggressive form of the disease with dismal prognosis; no effective treatment options are currently available. We had initiated work during the previous reporting year, with the main aim of identifying genetic pathways important for progression of CRC occurring in young colorectal cancer patients from India. We had standardized DNA isolation from tumour cells micro dissected from paraffin sections, with the help of the collaborating pathologist. Using this DNA, we were



able to standardize DNA sequencing to determine mutations in the APC gene, as well as microsatellite analyses in order



to determine the MSI status in the tumour samples.

### Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)

Our first objective was to determine whether one of the two canonical pathways known to occur in CRC viz. CIN or MSI, was active in the tumours isolated from the young CRC patients. In the current reporting year, we have included immunohistochemistry-based detection of intracellular localization of  $\beta$ -catenin, to determine the status of the canonical Wnt signaling pathway. A sample was scored positive for nuclear localization of  $\beta$ -catenin if at least 10% of tumor cells in the section exhibited immunohistochemical (IHC) staining in the nucleus.  $\beta$ -catenin was either shown to be present in the nucleus (Fig. 2A), in the membranes (Fig. 2B) or was not detected at all (Fig. 2C). So far, we have analyzed a total of

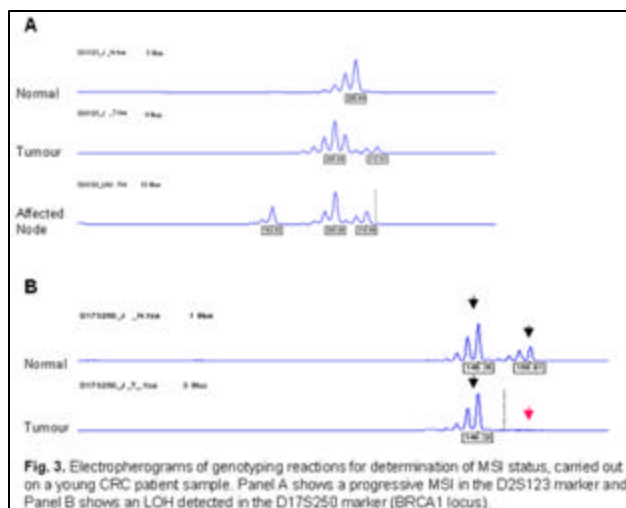


Fig. 3. Electropherograms of genotyping reactions for determination of MSI status, carried out on a young CRC patient sample. Panel A shows a progressive MSI in the D2S123 marker and Panel B shows a LOH detected in the D17S250 marker (BRCA1 locus).

20 young colorectal cancer patients and 21 older patients. At least two separate paraffin blocks were analysed for each tumour sample, wherever possible. The results reveal that only about 20% of young patients appear to have an activated Wnt signalling pathway in their tumours as compared to about 80% for the older patients, a significant difference ( $P=0.00007$ ; Fisher's exact test). We detected four novel mutations in the

MCR region of the APC gene: a 2bp deletion nt4461del2 (-TT) at codon 1488 (the resultant altered reading frame causes generation of a premature termination codon (PTC) 72 bp downstream of the deletion); a 1 bp deletion nt4551del1 (-G) at codon 1517 that results in generation of a PTC 13bp downstream of the deletion; a missense mutation p.P1369S (CCC to TCC); and a missense mutation p.D1486N (GAT to AAT). The D1486 residue is conserved in five of the seven 20 amino acid repeats (that are important for binding to  $\beta$ -catenin) and has been shown to directly interact with the 'charged button' (K435) in  $\beta$ -catenin. The K435 residue of  $\beta$ -catenin plays a crucial role in interaction with other binding partners as well, including E-Cadherin and Tcf. All the four mutations were not present in the matched normal for each sample (except for p.P1369S for which analysis of the matched

normal is ongoing at present), indicating that these were not germline mutations. Out of 11 samples analyzed that were positive for  $\beta$ -catenin nuclear staining, only 6 harbored a mutation in the MCR region of APC gene, indicating that mutations might occur in other region(s) of the APC gene, in  $\beta$ -catenin itself (or in another gene in the Wnt signaling

pathway, such as Axin), or APC might be inactivated through other mechanism(s) (e.g. methylation/deletion). There was no significant difference in the frequency of MSI-H tumors between the young and older patients (although result from only a limited number of samples is available at present), once again pointing to a distinct tumorigenesis pathway(s) operating in the young CRC patients. A representative electropherogram exhibiting a progressive Microsatellite instability and a loss of heterozygosity in the BRCA1 locus is shown in figure 3A and 3B respectively. The frequency of BRCA1 LOH reported from other studies is approximately 30% and our results revealed a similar frequency in both young and older patients. An interesting observation was a higher proportion of mucinous/signet ring cell type carcinomas in the young patients as compared to the older patients. Signet ring cell type adenocarcinoma has been shown to be a rare but highly aggressive subtype of CRC in the USA.

The data on Wnt pathway activation indicate molecular differences between sporadic CRC that occur in young and older patients, and support the worth of additional studies that we have initiated to understand the biology of the tumours occurring in the young patients. Screening for mutations in Kras and p53, determination of status of the CpG island methylator phenotype (CIMP) including screening for BRAF mutations, and analyses of genome-wide CNAs and gene expression profiles, might reveal the biological pathway(s)

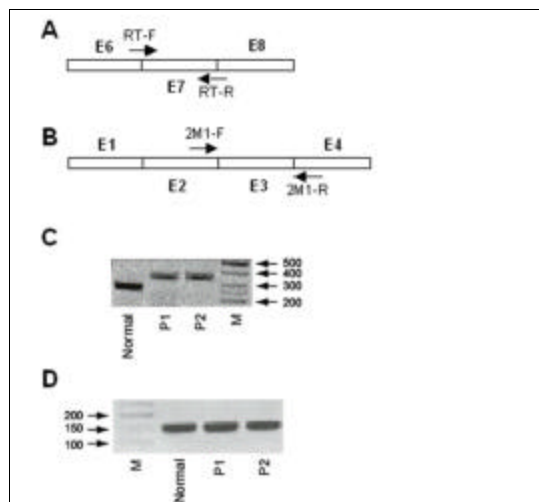
that drive(s) tumour initiation/progression in the young sporadic CRC patients.

## Project 2a: Molecular genetic analyses of Phenylketonuria (PKU)

### Objectives

Identification and characterization of disease-causing mutations in PKU

### Summary of work done until the beginning of this reporting year



**Fig. 4.** Characterization of splice mutation in PAH. Panels A and B show the location of primers used to determine whether the mutation was leading to the activation of a cryptic 3' splice site and for determination of transcript levels of PAH using real time RT-PCR, respectively. Panels C and D show the results of analysis of PAH transcripts from normal and patient samples using RT-PCR. Lanes: Normal, lymphoblasts generated from normal sample; P1, lymphoblasts isolated from Patient 1; P2, lymphoblasts isolated from Patient 2; M, 50bp DNA ladder. Panel C shows results from RT-PCR carried out using primer pair 2M1F and 2M1R for the PAH transcript. Panel D shows results from RT-PCR carried out for analysis of the GAPDH transcript from the same samples as

In our earlier work, we reported three novel mutations in the Phenylalanine Hydroxylase (PAH) gene causing Phenylketonuria (PKU) viz. an IVSII-2 A to G mutation (in three patients from the same family), an IVSI+5 G to A mutation and a p.N393ochre mutation (in two affected siblings in the same family). The first two mutations were predicted to inactivate splice sites in the PAH gene.

### Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)

Illegitimate transcription of the PAH gene occurs in lymphocytes and this observation has been used by workers previously to characterize mutations that affect transcript processing. In order to characterize the PAH splice site mutation that affects the second intron (IVSII-2 A to G) (from patient 1 and 2, both harboring the same mutation), we first generated lymphoblast cell lines from the patient blood samples (as well as from a control sample). We designed primers to determine the affect of the mutation (Fig. 4A and 4B). Often splice mutations resulting in frame shift generate a premature termination codon (PTC), and the cell triggers a process called nonsense mediated decay (NMD), wherein PTC-containing transcripts are degraded, protecting the cell from the potentially harmful affects of the truncated protein. Therefore, we first wanted to determine whether the transcript levels were reduced in the patient samples, when compared to a normal sample. We carried out real time RT-PCR using the primer pairs RT-F and RT-R (Fig. 4A) and confirmed the downregulation of the PAH transcript level in the patient as compared to the normal sample. Next, we used the NN-splice web based splice site prediction software to predict cryptic 3'-splice sites. As expected, the software predicted the authentic 3' splice site with the best score (0.94). An AG di-nucleotide located 79 bases upstream of the authentic 3' splice site had the second best score (0.67). This alternate site also fulfilled some of the criteria for an authentic 3'-splice site, for example, the presence of a pyrimidine rich sequence immediately upstream (including a uridine residue at -11 position). In order to confirm whether this AG dinucleotide was indeed being used as the 3'-splice acceptor site in the patients, we used the primer pair 2M1-F and 2M1-R (Fig. 4B) to PCR amplify the PAH transcript from the

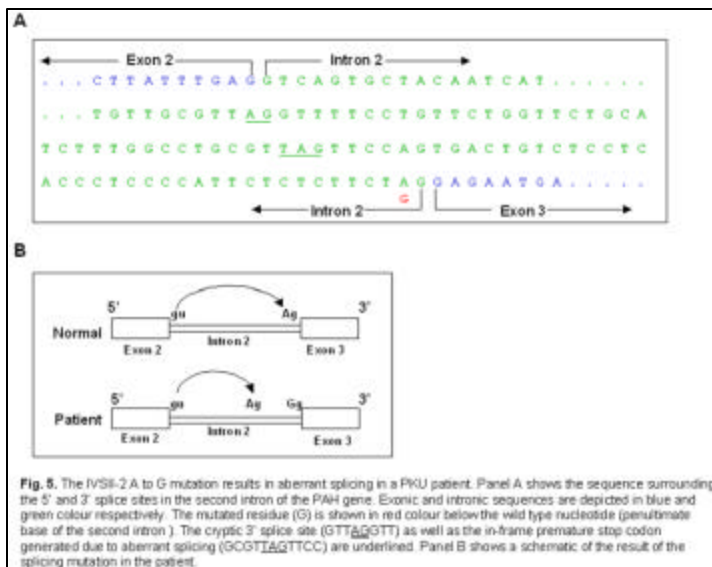


Fig. 5. The IVSII-2 A to G mutation results in aberrant splicing in a PKU patient. Panel A shows the sequence surrounding the 5' and 3' splice sites in the second intron of the PAH gene. Exonic and intronic sequences are depicted in blue and green colour respectively. The mutated residue (G) is shown in red colour below the wild type nucleotide (penultimate base of the second intron). The cryptic 3' splice site (GTTAGGTT) as well as the in-frame premature stop codon generated due to aberrant splicing (GCGTLAGTTCC) are underlined. Panel B shows a schematic of the result of the splicing mutation in the patient.

patients as well as from the normal sample. We observed that the RT-PCR product obtained from the two patients was about 100 bp more than that obtained from the normal sample (Fig. 4C) suggesting that a cryptic 3' splice site was being utilized. We also observed a reduction in the amount of the RT-PCR product in the patient samples as compared to the normal

samples, although there was no difference when a control transcript (that of GAPDH) was amplified (Fig. 4, C and D), as expected. The RT-PCR product obtained from both patient

samples was purified and subjected to automated DNA sequencing and the result revealed that the cryptic 3' splice site predicted by the NN-splice software was indeed being utilized in the patients (Fig. 5, A and B). Activation of the 3' cryptic splice site, results in a change in the reading frame leading to the generation of a PTC within the second intron itself, which may trigger NMD (Fig. 5B).

## **Project 2b: Molecular genetic analyses of Familial Hypertrophic Cardiomyopathy (FHC)**

### **Objectives:**

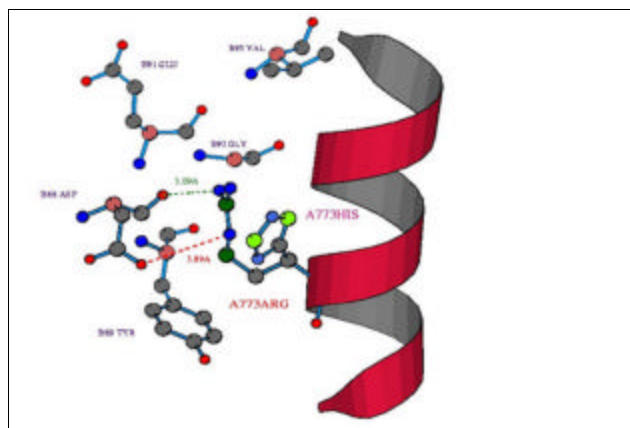
Identification and characterization of disease causing mutations in FHC

### **Summary of work done until the beginning of this reporting year**

This is a new activity.

### **Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)**

We have initiated a new study to profile mutations that cause FHC in Andhra Pradesh. So far we have collected clinical data for 64 FHC patients from our collaborating hospitals. We observed a high male:female ratio in the patients, probably due to socio-economic factors rather than a higher occurrence in males. The age varied from 20-71 yrs; it has been reported from studies performed in Europe and USA that specific mutations may be associated with an early occurrence and it would be interesting to find if such differences exist in the Indian population as well. The patients exhibited a wide heterogeneity in clinical presentation, emphasizing the variations in the disease as previously reported. Almost all patients exhibited abnormalities in their ECG. The interventricular septum (IVS) diameter varied to a great extent and interestingly, fifteen patients did not exhibit an abnormal IVS diameter. Again it is possible that specific mutations may be associated with such 'asymptomatic' patients. The IVS diameter varied from 1.2 cm to 2.7 cms. The mean IVS diameter in males was 2.4 cm, whereas in females it was 1.94, although this may not be



**Fig. 6.** Interfacial residues occurring within 5 A<sup>9</sup> region of 773Arg position in the chicken MYH7 protein. The figure depicts the probable interactions of 773Arg with the myosin essential light chain residues. Probable H-bond (green dotted line) and ionic bond (red dotted line) between the residues 773Arg of MYH7 and 88ASP of myosin essential light chain is represented. Probable side chain conformation of the mutant Histidine at the 773 position of MYH7 is also shown.

significant since we had much fewer female patients than males. In addition, the IVS diameter varied equally in different age groups. All patients exhibited symptoms including syncope/pre-syncope, angina, palpitation, etc. In order to determine the disease-causing mutation, we first

began screening the beta cardiac myosin heavy chain gene (MYH7),

which has been shown to harbour mutations in about 30-40% of FHC cases in the western population. Interestingly, four probands harboured an identical mutation (p.R787H) in the 21<sup>st</sup> exon of the MYH7 gene. The mutation was confirmed by bi-directional DNA sequencing. One hundred normal control samples did not harbour this nucleotide change, indicating that it was not a single nucleotide polymorphism.

We have carried out sequence and structure analyses of the mutant MYH7 protein, in order to understand the molecular basis for the pathogenic effect of the p.R787H mutation. The R787 residue is present in the neck/hinge region of MYH7 and is predicted to make contact with the myosin essential light chain (ELC). First we constructed a sequence profile for myosin chains i.e., the position-specific preferences of the 20 amino acid residues calculated from the multiple sequence alignment of related myosin chains. We observed that the profile score for arginine at the 787 position was 0.3, for aspartic acid -1.8, for lysine 0.3 and for histidine -0.6 indicating that any substitution that either reversed the charge or resulted in a polar amino acid was not favoured. The crystal structure of human MYH7 is not available where as that of the chicken homologue is available. The residue in the chicken protein equivalent to the human R787 is R773 and the latter is also a conserved residue. Structural analysis of the R773 interactions in the chicken homologue and mapping the same on the human myosin indicated that a H-bond is likely to be formed between R773 of normal MYH7 and the D88 residue of the ELC (Fig. 6). In addition, an ionic interaction is also predicted to occur between these two residues (Fig. 6). However, neither of these bonds is predicted to form when arginine is substituted by histidine (Fig. 6). Therefore, the changes in the interactions that occur due to the p.R787H mutation is may possibly compromise the binding of MYH7 to ELC, thereby affecting efficiency of muscle contraction.

We observed clinical heterogeneity among the affected members harbouring the p.R787H mutation. In order to determine the basis for this heterogeneity, we have begun to analyze the role of factors such as age, gender as well as status of ACE polymorphism that may play a role in modulation of the clinical presentation of the disease. Work has also been initiated to determine whether the p.R787H mutation could be a founder mutation in Andhra Pradesh, using microstellite markers located within or adjacent to the MYH7 gene.

### ***Publications***

- 1 Devi AR, Gopikrishna M, Ratheesh R, Savithri G, Swarnalata G and Bashyam MD (2006) Farber lipogranulomatosis-clinical and molecular genetic analysis reveals a novel mutation in an Indian family. ***Journal of Human Genetics***51:811-814.
1. Bashyam MD, Savithri GR, Gopikrishna M and Narsimhan C (2007) A p.R870H mutation in the beta cardiac myosin heavy chain 7 gene causes Familial Hypertrophic Cardiomyopathy in several members of an Indian family. ***Canadian Journal of Cardiology*** In Press .
2. Bashyam MD and Hasnain SE (2007) Array-based comparative Genomic Hybridization: applications in cancer and tubercubsis. Bioarrays, Ed. K Appasani, ***Humana press*** In Press.





## Laboratory of Cancer Biology

### Understanding the biology of Ras-mediated signaling events and prevalence of Human Papilloma virus in cervicovaginal samples collected from a rural set-up of Andhra Pradesh

<b>Principal Investigator</b>	Gayatri Ramakrishna	Staff Scientist
<b>Ph D Student</b>	Arvind Singh	Senior Research Fellow
<b>Other Members</b>	Santha Lakshmi	Project Assistant
	Lavanya	Project Assistant
	Pavani Sowjanya	Project Assistant
<b>Collaborators</b>	S K Sarin and Sujoy Bose (PhD student) Mediciti Group Johns Hopkins University	G B Pant Hospital, Delhi Hyderabad  USA

#### 1 Role of Ras-mediated signaling pathway in Hepatocellular carcinoma with persistent Hepatitis B Virus (HBV) infection

##### Objective:

To study the expression and functionality of Ras-p21 signaling pathway in hepatocellular carcinoma (HCC) with persistent HBV infection to understand how these regulatory mechanisms have deviated during the cancer progression.

##### Summary of work done until the beginning of this reporting year

Since it was difficult to get fresh resected tumor specimens of HCC, a retrospective study was initiated by collecting archived specimens from the G. B. Pant hospital, Delhi. The archived serological specimens were also collected from the Gastroenterology division. Matched archived wax blocks of patients with proven cases of HCC, cirrhosis and chronic liver disease were sectioned and standardization for the Ras-p21 expression by immunohistochemistry performed.

##### Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)

Although persistent infection with HBV is reported to be a major etiological risk factor for HCC, there is still a big lacuna with respect to other host and cellular components involved in progression of the liver carcinomas. Recent microarray data on HCC samples have identified changes in expression of numerous oncogenes and tumor suppressor however not one single pathway has been delineated for progression of HCC. Using *in vitro* cell culture system it has been shown the HBV-X protein can deregulate the Ras mediated signaling events and this warrants a comprehensive study to test if the same is true in the *in vivo*

human carcinoma samples with persistent HBV infection. In view of this, the present study was carried out to evaluate the deregulation in Ras mediated signaling in the HCC cases. The study was divided into a prospective and retrospective part. For the retrospective study, histopathologically proven cases of HCC (N=22), cirrhosis (N=10) and chronic liver disease (N=5) were analyzed for Ras expression by immunohistochemistry. For the prospective study, biopsy specimens of HCC (8) and chronic liver disease (N=5) were collected for evaluation of signaling pathways.

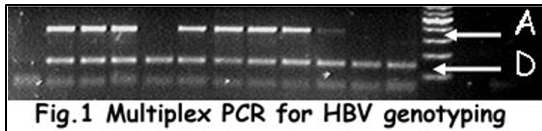


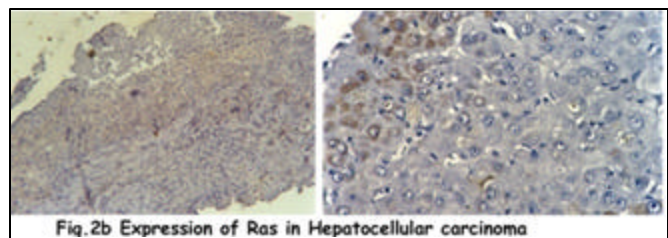
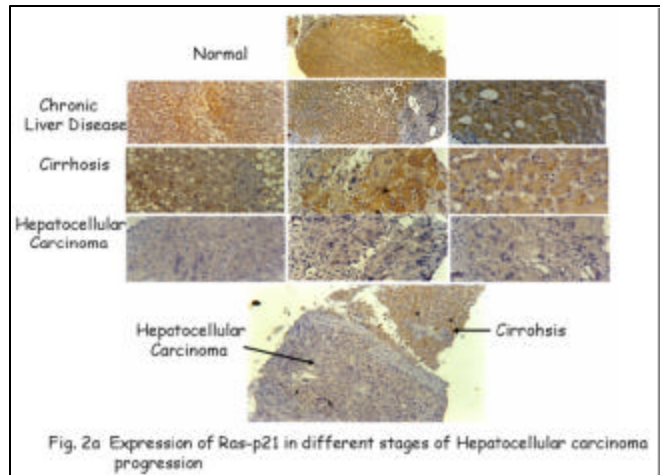
Table 1. HBV genotyping results from different liver patient groups

Cases	N	Single Genotypes						Mixed Genotypes	
		A	B	C	D	E	F	A&D	C&D
HCC	30	5	-	-	16	-	-	9	-
Cirrhosis	15	4	-	1	7	-	-	3	-
Chronic Hepatitis	10	1	-	-	8	-	-	1	-

Detailed serological study was also carried out in these cases for presence of hepatitis B virus in both the archived and fresh samples. All the HCC cases were positive for Hepatitis B virus by PCR assay. There were a few occult cases where the serology was negative for HBV surface antigen however they were

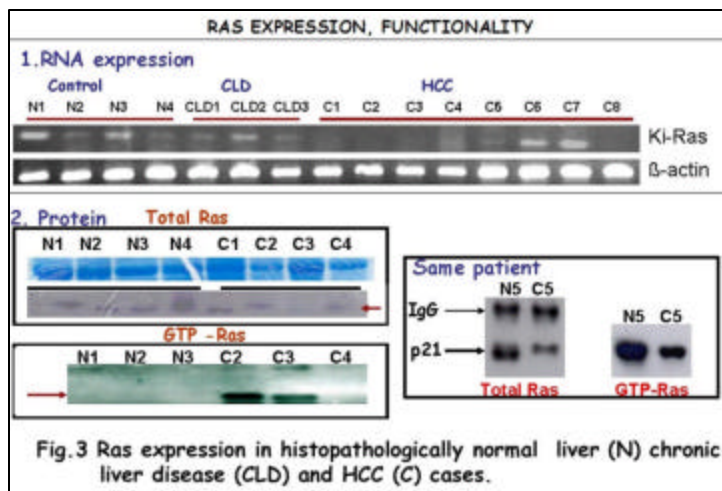
positive by the PCR diagnostics. HBV genotype D was found to be the most prevalent genotype found in the HCC cases (Fig.1, Table 1).

The chronological events towards progression to malignant carcinoma of liver involve chronic hepatitis and cirrhosis. Therefore, we evaluated the expression of Ras in all stages of liver disease progression by Immunohistochemistry (Fig. 2a). Normal liver showed an intense staining pattern for expression of Ras-p21, while the chronic hepatitis cases and cirrhotic liver showed strong to moderate expression. The HCC samples showed a very feeble expression of Ras-p21. Few cases of HCC showed distinct focal expression of Ras (Fig. 2b). The Ras expression was mostly cytosolic however membrane localization was also seen in a few non-malignant hepatocytes. Interestingly the hepatic tissue showed a distinct gradation in pattern of Ras staining with



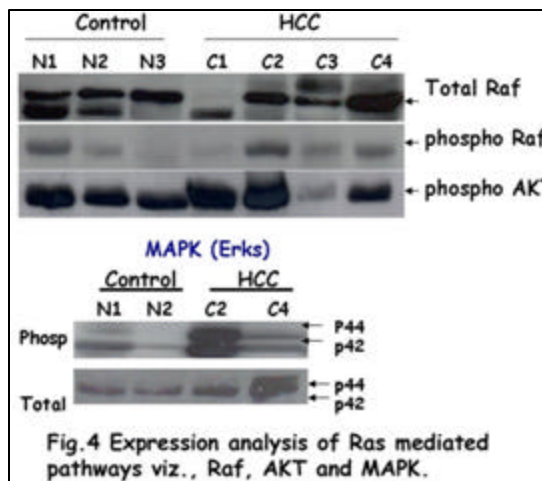
maximal staining in chronic hepatitis followed by cirrhosis and the least in carcinoma. In the biopsy samples of chronic hepatitis and resected specimens of HCC there was a significant heterogeneity in expression of Ras, while in general the levels appeared downregulated both at RNA and protein levels in the HCC cases compared to the normals (Fig. 3). For RNA and immunoblot analysis, it has to be noted that the normal and the cancer specimens were not from the same patient.

Since Ras exists in both GTP and GDP bound state, we analyzed the level of Ras-GTP by doing a GST-RBD pull down assay in the non-malignant and HCC cases. To get detectable GTP bound form in the normal samples (when Ras is not mutated) one has to use as high as 500-700µg of protein to detect the signal by the GST-RBD assay. Since the amount of



total protein obtained from the biopsy samples of HCC cases were not enough we could not use this high amount to do the GTP assay for all the samples. We performed this assay using 250µg of the total protein and found that only two samples showed a very high level of active GTP-Ras compared to the normals and incidentally

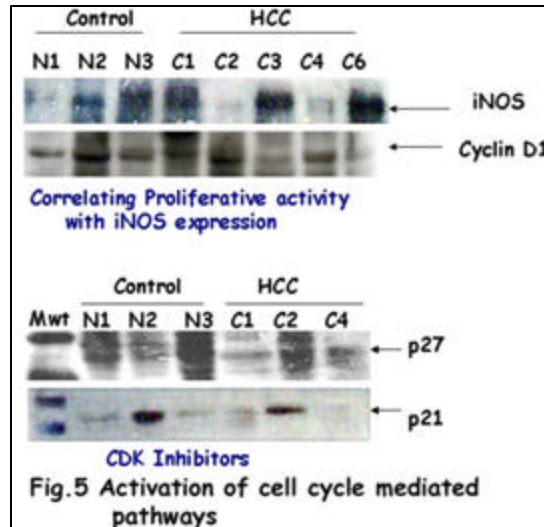
these two samples carried a mutation in K-ras in exon 1 region. However there was one



case where from the same patient, histopathologically normal and malignant tissue was obtained following resection and we could get enough protein (500µg) to do the GST-RBD assay. Interestingly this pair matched sample not only showed downregulation of the total Ras protein in the cancerous counterpart but also the functionally active GTP bound Ras (Fig. 3) was low compared to normal. No mutation in Ras was observed in the matched samples,

thereby indicating that even the GTP bound form of wild type Ras is downregulated in HCC.

We next analyzed the level of phosphorylated Raf, AKT and MAPK in the biopsy samples of hepatocellular carcinomas (Fig.4). Again we found heterogeneity with levels of activation of these three pathways. Of the four tumors analyzed by us, two consistently showed presence of higher levels of phosphorylated Raf, AKT and MAPK even though the levels of Ras appeared downregulated in these. This is indicative that at least Raf and MAPK can be activated independent of Ras. Interestingly the other two tumors which showed no



significant upregulation of Raf or AKT pathways, showed an upregulation of iNOS (Fig. 5). To correlate the activation of the Ras mediated pathways with the cell cycle proteins we analysed levels of CyclinD1 which has been shown to be overexpressed in many tumors. However at least in the samples we analyzed we did not find any change in its level. The CDK inhibitors p21 and p27 similarly showed varied expression, with p27 slightly downregulated in HCC cases compared to the controls (Fig. 5).

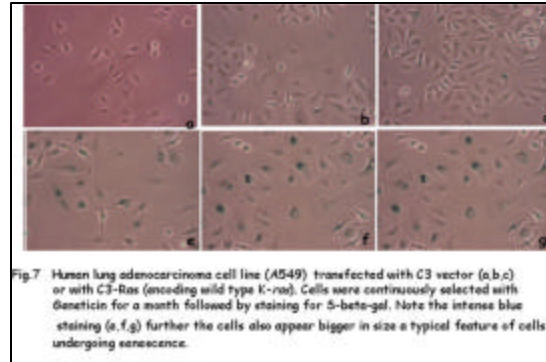
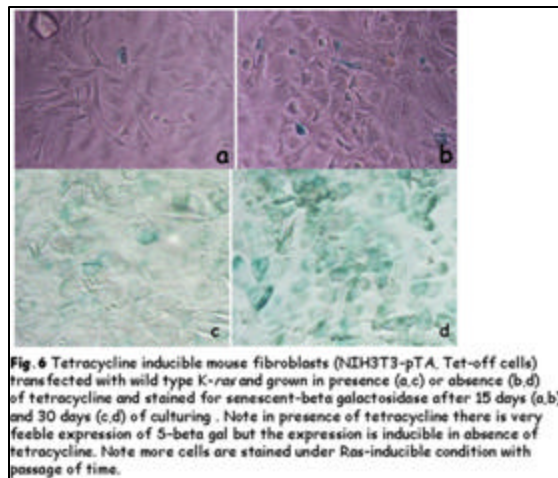
## Conclusions

1. HBV-Genotype D is the most prevalent genotype detected in the HCC samples tested by us.
2. The immunohistochemistry study is in stark contrast to other reports on different cancer types where expression of Ras is upregulated. Our observation of downregulation of Ras expression in absence of its mutation (in atleast two isoforms of Ras) is supportive of a tumor suppressive activity of the wild type form.
3. The varied protein expression profiles of downstream signaling proteins and the cell cycle regulated proteins are indicative of heterogeneity in the hepatocellular cancer samples.

## 2. Wild type K-ras can induce senescence

The induction of senescence in neoplastic cells constitutes an important pathway for tumor suppression. There are two forms of senescence described in the literature, the replicative senescence induced by telomere erosion and the oncogene induced senescence (OIS). Oncogenic *ras* induces tumor proliferation in immortalized cells however when it is introduced in primary cultures it induces OIS. However the role of endogenous wild type *ras* in inducing senescence is not known. In an ongoing work in the laboratory we could show that in certain cell types (A549 and NIH3T3) wild type K-*ras* can induce growth arrest. We

therefore, tested if wild type *ras* can induce beta-galactosidase a well known marker for senescence. Wild type *K-ras* was transfected either in (a) tetracycline inducible mouse fibroblasts (Fig. 6) or (b) in lung carcinoma cell line (A549, Fig. 7) and selected in geneticin followed by staining for senescent beta galactosidase (S-beta-gal). In both fibroblast and epithelial cells, wild type *K-ras* induced senescence.



### 3. Prevalence and distribution of high-risk Human papilloma virus (HPV) types in invasive squamous cell carcinoma of the cervix and in cervicovaginal samples collected from normal women in Andhra Pradesh

#### Objectives

To understand the genotype distribution of high risk HPV types in squamous cell carcinoma of cervix and to evaluate the feasibility of HPV testing in cervical screening programme in rural set ups.

#### Summary of work done until the beginning of this reporting year

Earlier we had completed a pilot study initiated in collaboration with both city based and rural hospital to evaluate the presence of HPV in women from Andhra Pradesh. In the pilot study we screened cervicovaginal samples (N=185) collected from women enrolled in the cervical cancer screening programme conducted in the rural community (Medchal Mandal) where nearly 10% of the samples were positive for high risk HPV types. Further we also did HPV testing and genotyping in cervical cancer specimens (N=41) obtained from the Cancer Hospital (M.N.J. Cancer hospital), where we found nearly 87.8% (N=36/41) of the squamous cell carcinomas positive for high-risk HPV types. The prevalence of the major high-risk HPV types detected in the cervical cancers was HPV 16.



**Details of progress made in the current reporting year (April 1, 2006- March 31, 2007). HPV prevalence in cervicovaginal samples collected from women residing in Medchal community**

Despite the high incidence of cancer cervix reported from India, large scale population based studies on prevalence and distribution of HPV types are very few from this region. In view of the clinical trials for HPV vaccine taking place in India it is of utmost importance to understand the prevalence of HPV in various geographical regions of India. Since the pilot study to screen women for HPV testing proved successful, a larger study was initiated in the rural community of Medchal Mandal. Till date we have screened close to 1976 women by Hybrid capture-2 assay (HC2) which is based on hybridization of RNA probes complementary to genome sequence of high risk HPV types. The RNA:DNA hybrid is captured on a plate using antibody and the final detection is based on a luminescent reaction and expressed as relative light unit (RLU). Using this assay, a total of 194 cervicovaginal

**Table 1: HPV prevalence in Medchal Mandal rural community by hybrid capture Method.**

Total number of cervicovaginal samples screened	HPV -	HPV +	Percentage HPV positive
1976	1782	194	9.8%

**Table 2. Comparison of PCR based line blot assay and Hybrid Capture 2 assay for detection of high risk HPV in cervical samples**

Hybrid Capture	Line Blot Assay		
	HPV +	HPV-	Total
HPV+	76	12	88
HPV-	12	335	347
Total	88	347	435

The overall agreement is 94.4% (411/435) and for HPV positivity is 76% (76/100)

**Table 3. Comparison between the self and vaginal collected samples for presence of high risk HPV types in cervical samples**

Self collected	Clinician collected		
	HPV +	HPV-	Total
HPV+	58	11	69
HPV-	29	335	364
Total	87	346	433

The overall agreement is 92.1% (393/433) and HPV positivity is 58.5% (58/98)

samples tested positive for high-risk HPV. The HPV positivity rate of 9.8% (194/1976) in the Medchal Mandal rural community is similar to prevalence rate as seen in other parts of the world (Table 1). At present the projected target is to screen close to 2500 women for HPV and corroborate these results with cytopathology so that HPV testing can be used an adjunct test for cancer cervix screening programmes. Besides the HC2 assay we also tested 435 samples by PCR based line blot approach. The total overall

agreement was good while the agreement for HPV positive samples was 76% (Table 2). The 12 samples which were positive by hybrid capture but negative by PCR had RLU/Co value below 5.0 indicating very low viral loads which could have got undetected by the line blot assay.

**Use of self collected samples as an alternate to clinician collected samples for testing HPV prevalence in rural set-ups:** In the pilot study we had evaluated HPV-DNA testing

using both clinician - and self-collected In the current reporting year we continued evaluation of about 435 paired samples collected from same women using two different methods of collection. The overall agreement was good and similar to pilot study. However, unlike the pilot study the percentage agreement among the HPV positive samples was better (pilot study 39% HPV positivity, present study 59%) (Table 3). This is indicative that self-collection by women can be a good alternative strategy for cervical screening programmes in rural settings with better health education programmes. It has to be however seen if similar results of feasibility of self-collection by women can be reproduced in other rural communities in India.

**Conclusions:**

Our results suggests a 10% HPV prevalence in a general screening population from the Medchal rural community of Andhra Pradesh; furthermore self collection of vaginal samples by women is a feasible alternative for cervical cancer screening programme.





**Laboratory of Computational and Functional Genomics**  
**Computational and Functional Genomics of Microbial Pathogens**

<b>Principal Investigator:</b>	Akash Ranjan	Staff Scientist
<b>PhD Students:</b>	Vaibhav Vindal	Senior Research Fellow
	P Uma Devi	Senior Research Fellow
	Jamshaid Ali	Junior Research Fellow
	Vijaykumar Muley	Junior Research Fellow
	Rohan Misra	Junior Research Fellow
<b>Other Members:</b>	Sarita Ranjan	Project Associate (Until 2006)
	K Suma	Project Assistant (Until 2006)
	K Rohini	Project Assistant
	E Ashwanth Kumar	Project Assistant
	M Jayavardhan Reddy	Project Assistant
	K Madhumathi	Project Assistant
	Siddharth Narayanan	Project Assistant
<b>Collaborators:</b>	Ashis K Das	BITS, Pilani, India
	Seyed E Hasnain	Univ. of Hyderabad, Hyderabad
	Jorg Hacker	ZINF, Univ. of Wuerzburg, Germany
	Niyaz Ahmed	CDFD, Hyderabad

**Project 1: Identification of functional linkages in mycobacteria**

**Objectives:**

Transcription regulators are important proteins that coordinate gene expression to bring together all the essential gene products which are physiologically linked. Since many of the genes are expressed as parts of an operons, we have used *in silico* approach to construct an operon model for mycobacterium genomes. The present objective of this work is

- To Identify and classify GntR-like regulators from Mycobacteria
- To construct a database of *in silico* identified operons in Mycobacteria
- To develop *in silico* approach to identify and analyse proteins responsible for given phenotype

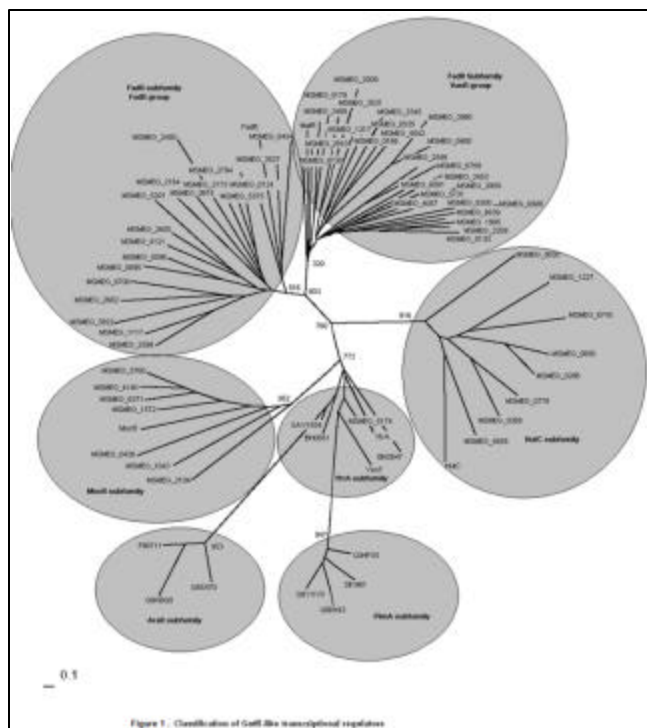
**Summary of the work done until the beginning of this reporting year**

The genome of tubercle bacillus *Mycobacterium tuberculosis* contains a large number of hypothetical and poorly characterized proteins including the proteins belonging to putative GntR family. The regulators belonging to this family show a conserved N-terminal region containing winged helix-turn-helix DNA-binding domain but have a highly diverse C-terminal domain involved in effector-binding and oligomerization. This heterogeneity leads to a further

classification of this family into sub-families, such as FadR, HutC, MocR, YtrA, AraR and PlmA. We have reported classification of five putative gntR genes as fadR-like regulators, one gene as hutC-like regulator and one as ytrA-like regulator. Interestingly we did not find any representative of other families in *M. tuberculosis*. This classification was also evident with all specific secondary structural features belonging to their family. We also observed that regulator Rv3060c shows an unusual size, which led us to demonstrate it successfully as a product of gene duplication and fusion. This study extends the genome annotation of *M. tuberculosis* and provides important leads for initiating experimental characterization of these proteins. Apart from this in silico classification, Rv0494 was structurally analyzed. We have proposed 3-D structure for Rv0494; conservation of DNA binding domain of Rv0494 and *E. coli* FadR is defined. Important residues of *E. coli* FadR that are reported to make critical contact with DNA are found conserved in Rv0494 protein sequence. The similarity between the two proteins goes above 50% in helix-turn-helix region which suggest a possibility that Rv0494 can recognize *E. coli* motif. Using the recombinant Rv0494, we carried out binding assay with known *E. coli* DNA targets of FadR. In this binding assay we also assessed the role of minor groove and major groove of DNA targets in target recognition.

**Details of progress in the current reporting year (April 1, 2006– 31, March 2007)**

**Characterization of *M. smegmatis* GntR regulators**



Using previously described approach to characterize *M. tuberculosis* GntR like regulator, we characterized *M. smegmatis* GntR-like regulators.

**Classification of putative *M. smegmatis* GntR into subfamilies**

Unrooted tree of *M. smegmatis* GntRs was constructed with classified representatives of all subfamilies. To construct the tree, multiple sequence alignment of the proteins was followed by manual adjustment according to predicted secondary structure. Each branch of constructed tree represents a subfamily. FadR subfamily is divided

into two groups, FadR group and VanR group (Figure 1).

### **FadR-like proteins of *M. smegmatis***

Of all the putative GntRs, 45 proteins were classified as the FadR-like regulators. These subfamily members are further classified into two groups' viz., FadR and VanR groups, where the C-terminal effector binding and/or oligomerization domain length is about 170 and 150 amino acid residues respectively comprising all  $\alpha$ -helices. Among all FadR-like regulators, 19 regulators were clustered as members of the FadR group while 26 for the VanR group. To study secondary structural features both the group members were dealt separately. C-terminal domain of all the members of FadR group were predicted with seven  $\alpha$ -helices except MSMEG\_2599. All the regulators showed distinguishable predicted secondary structural features specific to this subfamily. Secondary structural patterns of the regulator MSMEG\_3959 revealed an extra secondary structural element, which could be significant in studying protein family evolution. FadR-like regulators are known to be involved in the regulation of gene expression in response to oxidized substrates related to either amino acid metabolism or in various metabolic pathways such as glycolate, pyruvate, lactate, malonate or gluconate. These results provide a starting point for a detailed biochemical and genetic characterization of *M. smegmatis* FadR-like regulators.

### **HutC-like proteins of *M. smegmatis***

Contrary to the FadR-like regulators, the regulators of this subfamily consist of both  $\alpha$ -helices and  $\beta$ -sheet structures in the C-terminal domain. We identified eight GntRs as members of this subfamily. All these members showed distinguishable predicted secondary structural features specific to this subfamily. These regulators are known to acquire the same protein fold as *Escherichia coli* UbiC; hence it is also named as UbiC transcription regulator-associated (UTRA) domain. This effector-binding domain responds to various ligands like histidine (HutC), long chain fatty acids, trehalose 6-phosphate or alkylphosphonate. A range of known ligands, specific to many HutC-like regulators, will help in characterizing the classified *M. smegmatis* regulators.

### **MocR-like protein of *M. smegmatis***

Among all the putative GntR regulators, eight were classified as members of the MocR subfamily. All the eight regulators showed distinguishable predicted secondary structural features specific to this subfamily. MocR-like regulators show homology to the class I aminotransferase proteins, which requires pyridoxal 5'-phosphate (PLP) as a co-factor. All MocR-like regulators exhibit a PLP attachment site with a conserved lysine residue, which is also evident in the classified MocR-like regulators. It would thus be interesting to study the role of pyridoxal phosphate regulation in the classified regulators.

### **YtrA-like protein of *M. smegmatis***

The YtrA subfamily is the least represented GntR-like regulator in the bacterial genomes. Among all *M. smegmatis* GntR regulators, only one regulator MSMEG\_5174, showed the signatures of the YtrA subfamily member. YtrA possesses a reduced C-terminal domain with only two  $\alpha$ -helices. The average length of the putative effector binding and/or oligomerization domain is about 50 amino acids. YtrA from *B. subtilis* is an experimentally explored regulator, which is part of a large self-regulated operon. This operon consists of genes encoding the ATP binding cassette (ABC) transport systems in addition to the YtrA. It would be interesting to study whether MSMEG\_5174 has any role in modulating such an operon.

### **Operator site analysis**

We have tabulated a list of potential operator sites near the perfect palindrome sequence with conserved residues, which are found to be specific for most of the subfamily members. We did not find an operator sequence in the upstream sequences of all the remaining regulators. All the predicted sites were found to be in the upstream region from the translation start site except MSMEG\_2599. Identification of these sites is an important step to understand the GntR associated regulon or the gene regulatory network in the genome.

### **Ortholog prediction**

We have found a number of *M. smegmatis* GntR regulators that are orthologs of proteins from the other species of mycobacteria and *B. subtilis*. As orthologs typically share the same function, these regulators could serve as a model to study homologs from the other species of mycobacteria. These characterized orthologs may provide clues for initiating detailed biochemical characterization of *M. smegmatis* proteins. Many putative orthologs were experimentally known like Rv0165c that is involved in regulation of *mce1* operon; GntR, a transcriptional repressor of gluconate operon; YcbG, involved in utilization of D-glucarate and D-galactarate; YcnF, involved in utilization of gamma-aminobutyrate. However, we did not find the orthologs for all *M. smegmatis* GntRs in other pathogenic species.

### ***Ab initio* mapping of genes to bacterial phenotypes**

Microorganisms are ubiquitous, occurring at amazingly high densities in water, hot springs and soil, as well as in the stable biological niches of host organisms. Many of the latter are disease causing, with novel evolved mechanisms to cope with the physiological conditions of their host. Apart from physical conditions of life, these living forms are systematically classified into various groups depending upon variations among them. The quest for genetic basis of variation is not new. Since Gregor Mendel's time, this has been one of the central issues of biology. In the present work, we use curated databases to extract the presence or

absence of orthologous groups of proteins in 331 complete genome sequences. Phenotypic data were also extracted for many of the organisms. We then systematically mapped genes responsible for phenotype using mutual information and hypergeometric distribution.

More than 83000 orthologous groups from 331 complete genome sequences were extracted from STRING database. Phylogenetic profile were constructed and represented in Matrix Mo,g, where 'o' represents the number of organisms (331 organisms) and 'g' the number of orthologous groups (Ogs) under investigation. Presence or absence of OG in each organism was represented by 1 or 0. Phenotypic information for each organism was extracted from curated databases and literature mining. Phenotype profile for each organism is represented by 1 if organism exhibits the phenotype, otherwise 0 is assigned. The phylogenetic profile of each OG was compared with phenotype profile using Hypergeometric distribution (HD) and Mutual information (MI). Top 100 hits with highest mutual information and lowest *p-value* score were selected for further analysis. These are the probable proteins responsible for phenotype under investigation.

## **Project 2: Improved Genome Annotation of *P. falciparum* Genome**

### **Objectives:**

*P. falciparum* genome is a highly AT Rich Genome that offer great challenges for gene identification and annotation. The objectives of this project are to:

1. Develop computational approach, such as identification of novel *Plasmodium* specific substitution pattern, that can overcome this problem
2. Identify and annotate functional proteins such as membrane-associated protein encoded in genome.

### **Computational and development of Pf specific substitution matrix**

Sequence alignments are weighed both at the nucleotide and amino acid levels. However, as many mutations within the DNA are synonymous, there are chances of over estimation of divergence and hence protein alignment methods supersede nucleotide alignment methods. Most of the widely used protein homology detection methods like FASTA and BLAST make use of amino acid substitution matrices to score the alignments. The matrix consists of log-likelihood scores that reflect how likely one amino acid is substituted over the other. A positive score indicates a favored substitution over a chance event and a negative score reflects a less likely occurrence of the substitution than predicted by chance. In case of homology searches with standard matrices for compositionally biased genomes, the background frequencies vary greatly as they are calculated from the protein databases having standard compositions of the protein. The naive use of standard matrices for proteins

with non-standard composition has thus been debated for long and the need has been proposed the need to adjust the target frequencies accordingly. We have attempted to use an entirely different dataset for calculating the target and background frequencies in order to figure an exclusive substitution matrix for *Plasmodium falciparum* that reflects substitutions from a set of completely annotated proteins of *Plasmodium falciparum* and its distant orthologs. Our matrix is calculated from protein blocks obtained from our dataset based on the formalism of Henikoff. We expect that use of the substitution matrix thus generated might increase the threshold of alignment scores for database searches with *Plasmodium falciparum* leading to better annotation of the *Plasmodium falciparum* hypothetical proteins, and our results support this notion.

### **Computational prediction and analysis of transmembrane proteins in *Plasmodium falciparum***

At present, 60% of the available anti-infective agents act on the transmembrane proteins. Transmembrane proteins play a crucial role in transport, maintenance of homeostasis and cellular communication. Transmembrane proteins play a variety of roles in pathogenic organisms such as adhesion to host cell, downstream signaling in response to external environment etc. In light of this host-pathogen interaction, we have attempted to predict and estimate the transmembrane proteins of *Plasmodium falciparum*.

In the present work, all *P. falciparum* proteins were screened for transmembrane helices. Signal sequence prediction was carried out if predicted transmembrane helix is < 18 aa and is present within 50 aa from N-terminus. In case the protein show the signal sequence, transmembrane protein prediction was repeated after removing the signal sequence for all proteins that predict “membrane-out” disposition of the N-terminal region. Proteins that were not predicted as transmembrane protein upon removal of signal sequence were considered as false positive and discarded. The remaining proteins (2047) were selected for further studies as probable transmembrane proteins. Bi-directional BLAST searches were made against complete genome sequences of eukaryotic organisms and clustering of these proteins was carried out on the basis of phylogenetic profile. Clusters were used to infer lineage specific occurrence of these proteins and for mapping interactions among them. BLAST searches were made against NR and Pfam sequence profiles for functional annotation and these sequence similarity searches were used for determination of their functional role.

## **Publications**

- 1 Ranjan S, Seshadri J, Vindal V, Yellaboina S and Ranjan A (2006) iCR: a web tool to identify conserved targets of a regulatory protein across the multiple related prokaryotic species. **Nucleic Acids Research** 34: W 584-587.
- 2 Ranjan S, Gundu RK and Ranjan A (2006) MycoPeronDB: a database of computationally identified operons and transcriptional units in Mycobacteria. **BMC Bioinformatics** 7: 59.
- 3 Ranjan S, Yellaboina S and Ranjan A (2006) IdeR in mycobacteria: From target recognition to physiological function. **Critical Reviews in Microbiology** 32:69-75.
- 4 Yellaboina S, Ranjan S, Vindal V and Ranjan A (2006) Comparative analysis of iron regulated genes in mycobacteria. **FEBS Letters** 580:2567-2576.
- 5 Vidyasagar M, Hasnain SE, Mande SC, Nagarajaram HA, Ranjan A, Acharya MS, Anwaruddin, Arun SK, Gyanraj Kumar, Kumar D, Priya S, Ranjan S, Reddi BR, Seshadri J, Sravan Kumar P, Swaminathan S, Umadevi P and Vindal V (2007) BioSuite: A comprehensive bioinformatics software package (A unique industry- academia collaboration) **Current Science** 92:29-38.
- 6 Ahmed N, Majeed AA, Ahmed I, Hussain MA, Alvi A, Devi SM, Rizwan M, Ranjan A, Sechi LA, and Megraud F (2007) genoBASE pylori: A genotype search tool and database of human gastric pathogen Helicobacter pylori. **Infection Genetics and Evolution** In Press.
- 7 Vindal V, Ranjan S and Ranjan A (2007) In silico analysis and characterization of GntR family of regulators from *Mycobacterium tuberculosis*. **Tuberculosis** In Press.





**Laboratory of Transcription**  
**Mechanism of transcription termination and antitermination in**  
***Escherichia Coli***

<b>Principal Investigator</b>	Ranjan Sen	Staff Scientist
<b>Ph.D. students</b>	Anoop Cheeran	Senior Research Fellow (until March 4 <sup>th</sup> 2007)
	Bibhusita Pani	Senior Research Fellow
	Jisha Chalissery	Senior Research Fellow
	Ghazala Muteeb	Junior Research Fellow
<b>Other members</b>	Dipak Dutta	Senior Project Assistant
	Nanci R Kolli	Junior Project Assistant
	Dhanajay Joshi	Project Assistant
	Gowresh	Lab-attendant
<b>Collaborators</b>	Shekhar Mande	CDFD
	H A Nagarajaram	CDFD

**Objectives *Escherichia***

Bacterial transcription must terminate at the end of each operon. In *Escherichia.coli*, the ends of 50% of the operons each contains an intrinsic termination signal that codes for a hairpin followed by a U-rich stretch in mRNA. The other 50% of operons do not have any signature sequence and it is possible that termination of these operons depend on the factor called Rho. In certain instances, these termination signals can be overcome in response to certain type of modifications in the transcription elongation complex and the process is termed as antitermination. The mechanisms of these termination and antitermination processes are still not very clear and offer an exciting subject for study. In this laboratory, studies in the following areas have been undertaken.

- 1) Role of RNA:DNA hybrid sequences in ternary elongation complex (Ec) during termination.
- 2) Mechanism of transcription antitermination by an antiterminator protein N from bacteriophage H-19B.
- 3) Mechanism of action of transcription termination factor, Rho.
- 4) Mechanism of antagonism of Rho-dependent termination by an anti-rho factor *Psu*.

## **Summary of the work done until the beginning of this reporting year**

- 1) Using several protein footprinting techniques we have identified the region of transcription elongation complex (EC) which comes close to the C-terminal domain of the antiterminator protein N.
- 2) We isolated termination defective mutants in Rho by random mutagenesis, mapped them on the structure of Rho and characterized their in vitro properties, and correlated involvement of different regions in the process of termination.
- 3) Psi which is a 20 kd protein encoded by bacteriophage P4, known to be an anti-rho factor, was cloned, purified to homogeneity and characterized for its physical properties, and the mechanism of its anti-Rho function was elucidated.

## **Details of progress in the current reporting year (April 1, 2006 - March 31, 2007)**

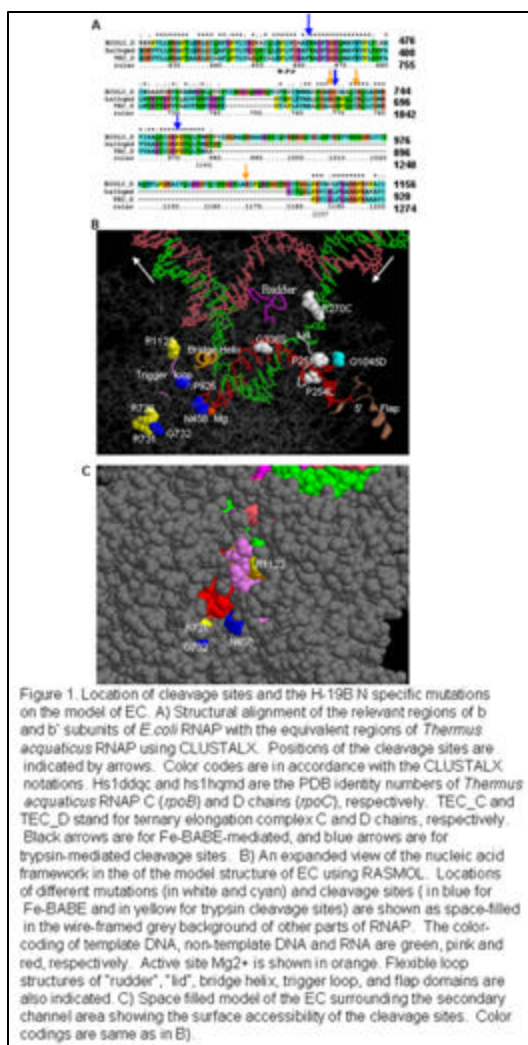
### **A) Anti-pausing activity of N-modified elongation complex and mechanism of action of N.**

The N protein of bacteriophage  $\lambda$  activates expression of the late genes by facilitating the read-through of transcription terminators present in the early genes of the phage. N is a small basic protein, which binds to a specific stem-loop structure (box B of *nut* site) of its own mRNA through the arginine rich motif (ARM) present in its Nterminal domain, and interacts with the EC through its C-terminal domain via an RNA looping mechanism. N requires several host-coded factors called Nus factors, to form a processive antitermination complex. The interacting surface of N on the EC is not yet known, knowledge of which is critical to understand the mechanism of antitermination.

Using a random mutagenesis screen, we have recently isolated and characterized *Escherichia coli* RNAP mutants specifically defective for H-19B N-mediated antitermination. These mutations are located very close to important structural elements of the EC, like the RNA exit channel, the lid, and the rudder. In this report, we have identified, by protein footprinting of the N-modified EC the regions of the EC that come physically close to the C-terminal domain of H-19B N. Specific cleavage patterns of the EC were generated using Fe-BABE-labeled N, Fe-DTT from the active site, and trypsin. We have also analyzed the effects of the binding of H-19B N on the interactions around the RNA:DNA hybrid at a class II pause site, and the interactions of the flap domain with hairpin RNA near the RNA exit channel at a class I pause site. We concluded that the C-terminal domain of H-19B N comes close to the active site  $Mg^{2+}$  and that N-induced altered interactions in the active center cleft stabilizes the RNA:DNA hybrid at a class I pause site and destabilizes the RNA hairpin-flap domain interactions at a class II pause site.

i) Mapping the interaction domain of H-19 B N.

We have earlier described RNAP mutations (P251S, P254L, R270C, G333S in  $\beta'$  and G1045D in  $\beta$  subunit) specific to H-19B N action, which are located in and around the RNA exit channel, the lid and rudder elements of the elongation complex. In order to determine the spatial relationship of the N-induced cleavage sites with the positions of those mutations,



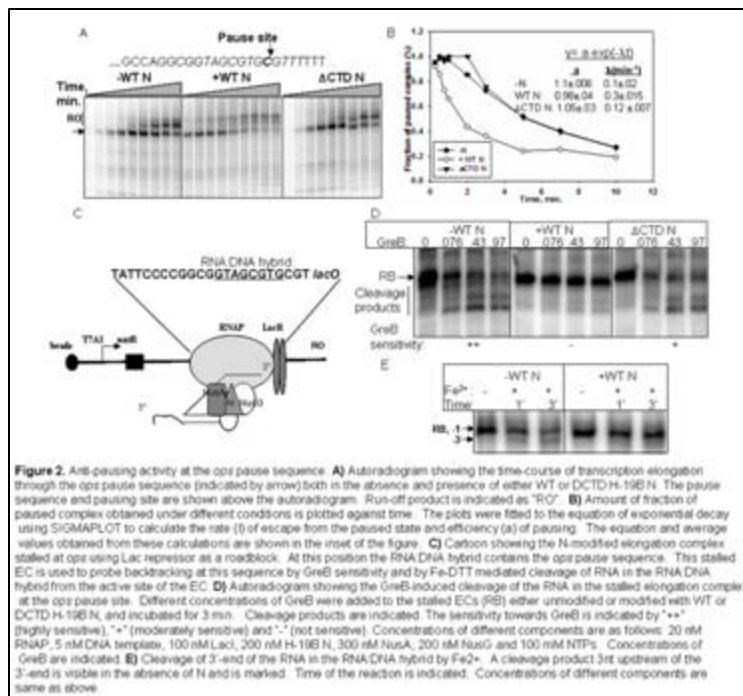
we localized both the cleavage sites and the positions of the mutations on the available model of EC, developed on the basis of the crystal structures of *Thermus aquaticus* RNA polymerase and the yeast RNA polymerase II elongation complex, and the cross-linking data on protein-nucleic acids interactions in the EC of *E. coli* RNAP. We first determined the equivalent amino acids of these mutations and the cleavage sites in the model structure by structural alignment (Figure 1A) using CLUSTALX program. Figure 1B shows the location of the cleavage sites together with H-19B N-specific mutations obtained earlier. Interestingly, the cleavage sites obtained from Fe-BABE or trypsin are located close to the active site and did not overlap with the sites of the mutations. The space-filled version of the model (Figure 1C) shows that both the trypsin cleavage sites and two of the Fe-BABE-induced cleavage sites are visible through the secondary channel and located close to the surface.

The Fe-BABE-induced cleavage sites are found to be ~30-60 Å away from the positions of the different mutations. Position G333 comes closest (~ 29 Å) to N458 and P926. Theoretically hydroxyl radical generated from Fe-BABE tagged to cysteine 107 of H-19B N can cleave the peptide backbone within a 12 Å radius. Therefore, it is unlikely that the C-terminal domain of H-19B N binds to the region defined by the mutations as the hydroxyl

radicals will not be able to travel over an average distance of  $\sim 45 \text{ \AA}$  to generate cleavages near active site  $\text{Mg}^{2+}$ . One possible explanation for this observation is that the C-terminal domain comes physically close to the active site  $\text{Mg}^{2+}$  and that this N-induced interactions near the active site exerts allosteric effects around the RNA-exit channel and the 5'-half of the RNA:DNA hybrid as defined by the location of the mutations. Alternatively, mutations near the RNA exit channel may define the interaction site of NusA present in the N-NusA complex.

ii) N prevents reversible backtracking at ops pause sequences

The antiterminators N and Q can suppress pausing during elongation. Results from protein footprinting experiments, and from earlier mutational analyses suggest that H-19B N may modulate interactions around the RNA:DNA hybrid as well as in the RNA exit channel. Therefore, we assayed the effect of H-19B N on two well-defined pause sequences, namely

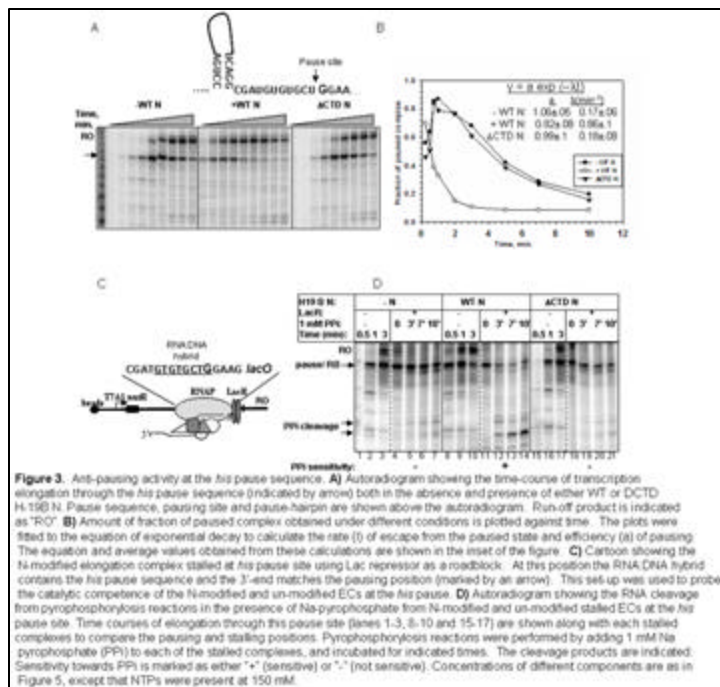


*ops* and *his* pauses, which involves altered interactions in these two regions. At the *ops* pause sequence, pausing occurs due to the backtracking of the EC, which is dependent on the sequence of the RNA:DNA hybrid. Pausing at *his* pause sequence is proposed to be mediated by interaction of a RNA-hairpin with the flap domain of  $\beta$ -subunit located near the RNA exit channel.

The sequence at the *his* pause codes for RNA that folds into a hairpin near the exit channel.

We followed the pausing kinetics through the *ops* pause sequence cloned downstream of the *nutR* site, both in the absence and presence of either WT or mutant H-19B N proteins (Figure 2A). The amount of paused complex was plotted against time (Figure 2B). In the presence of WT N, pausing efficiency (denoted as "a" in the equation of exponential decay; shown in the inset of Figure 2B) was not affected, whereas the rate of escape (denoted as "l" in the equation) from the paused state was three times faster. This result suggests that N does not prevent the EC from entering the paused state, but reduces the half-life of this

conformational state possibly by disfavoring the backtracking of EC which is an important component of this type of pausing. In order to determine whether N prevents backtracking at this sequence, the EC was stalled at this site by using Lac repressor as a road-block (Figure 2C). Backtracking of the stalled EC was monitored by GreB-induced cleavage of the RNA and by the Fe-DTT mediated cleavage of the 3'-end of the RNA from the active site of the EC. In the absence of WT N, the EC was sensitive to GreB-induced cleavage (Figure 2D) and Fe-DTT mediated cleavage of the RNA was also observed at an internal position (3 nt upstream from the 3'-end and marked as "-3" in the Figure 2E). These are the indications of the backtracking of the EC at this pause site. In the presence of WT N, sensitivity of EC for GreB was significantly reduced (Figure 2D) and Fe-DTT mediated cleavage was also not observed (Figure 2E). These two results strongly suggest that N prevents backtracking at *ops* pause site. This observation is specific to WT N, as its mutant derivative did not elicit this response.



iii) N destabilizes the flap domain-RNA hairpin interactions at *his*-site pause sequences

Next we followed the pausing kinetics through the *his* pause site cloned downstream of the *nutR* sequence (Figure 3A). The amount of paused complex was plotted as described above (Figure 3B). As in the case of the *ops* pause, the pausing efficiency at the *his* pause did not change significantly in the presence of

WT N, whereas the rate of escape from the paused state increased by about 5-fold (see the inset of Figure 3B for the rate constant values). The effect of N on the *his* pause was more pronounced than at the *ops* pause. The dwelling time in this paused state depends on the stability of the flap domain-RNA hairpin interaction. As binding of N only affected the pause half-life, it is possible that it did not affect the formation of the pause hairpin, but rather weakened the flap-hairpin interaction. It has also been shown that the flap-hairpin interaction at this pause site leads to catalytic inactivation of the EC. Therefore, if the presence of H-19B N destabilizes this interaction it will also prevent the catalytic inactivation



caused by the RNA hairpin. To test this hypothesis, the EC was stalled by Lac repressor at the pause site, NTPs were removed by washing, and the catalytic competence of the stalled EC was tested by the pyrophosphorolysis reaction (Figure 3C). Figure 3D shows that the 3'-

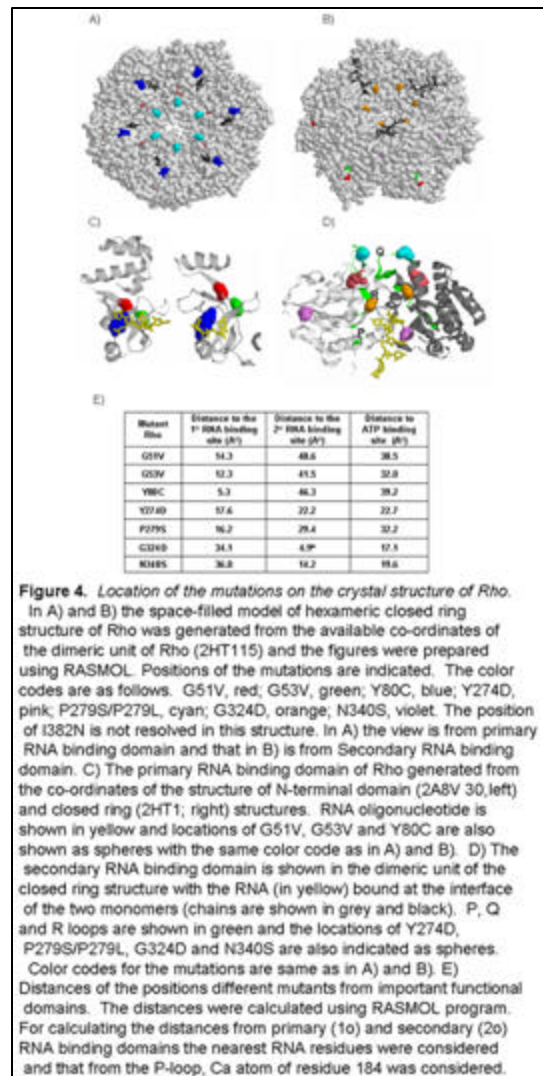
end of the RNA of this stalled EC exactly matched the pause site (compare the lanes 3 and 4, 10 and 11, 17 and 18). Formation of RNA–cleavage products (marked as “PPi cleavage”) induced by pyrophosphate was only observed in the presence of WT H-19B N. This effect was specific to WT N as it was not observed either in the absence of N or in the presence of  $\Delta$ CTD N, that does not bind to the EC. We conclude that the flap domain–RNA interaction is destabilized or significantly weakened in the N-modified EC and thereby preventing the catalytic inactivation.

### B) Transcription termination defective mutants of Rho: Role of different activities of Rho in releasing RNA from the EC.

We have finished the characterization of each of the Rho mutants (whose isolation had been described earlier) and we report here the summary of all the results, location of the mutations on the recently solved hexameric structure of Rho, and also the C-terminal conformational changes induced by a point mutation in the N-terminal domain.

#### *Location of the mutations on the crystal structure of Rho and prediction of functional defects.*

We mapped the positions of the mutations on the recently reported hexameric closed ring structure of Rho which has both the primary and secondary RNA binding sites occupied with nucleic acids (figure 4A and 4B). In general the mutations are located within or close to the previously identified important functional domains of Rho. Among them, G51V, G53V and Y80C are in the primary RNA binding domain (figure 4C). Y274D, P279S are in or close to the Q-loop (figure 4D). G324D and N340S are close to the secondary RNA binding domain and G324D is in the R-loop (figure 4D). The mutation I382N could not be located on



the structure as the C-terminal end of the closed ring structure of Rho is not resolved. Interestingly mutations in these important structural elements of Rho did not have a lethal phenotype, even though null mutants of rho are inviable. It is revealed from the structure that amino acid Y80 makes direct contact with the nucleic acid in the primary RNA binding domain (figure 4C). Hence there is a high probability that Y80C change will affect the primary RNA binding drastically. Amino acids G51 and G53 also come within 12-14 Å of the nucleic acid in the crystal structure and changes in these amino acids can also affect the primary RNA binding (see the distance calculations in figure 4E). The crystal structure revealed the binding of only two nucleotides, whereas about 10 nucleotides can be occupied in the primary RNA binding site of a monomer. Therefore it is likely that other amino acids of this domain (including G51 and G53) will take part in the primary RNA binding. Defect in the primary RNA binding due to the change in these three amino acids will subsequently affect the secondary RNA binding and the RNA release processes.

The amino acids G324 and N340 are situated close to the RNA in the secondary RNA binding site (figure 4D and 4E). It is likely that G324, located in the R-loop, will take part directly in the interaction with RNA. Therefore it can be predicted that changes in these two amino acid will cause defect in the secondary RNA binding and as well as in the ATP hydrolysis activities. This may in turn affect the processive translocation of Rho along the RNA, which will lead to termination defect.

P279S and P279L changes in the Q-loop can alter the loop conformation by extending the length of the preceding helix. The closed ring structure of Rho revealed that Q-loop is about 30 Å (figure 4E) away from the nearest RNA residue in the secondary RNA binding site. In the closed ring structure of Rho, Q-loop forms a hairpin-like structure from the disordered conformations observed in the open ring structure. This alteration may be important for attaining the active conformation of Rho and change in the loop conformation due to the mutations will affect its function. On the other hand Y274, which is located just outside this loop, may come on the pathway of the RNA passing through the dimeric interface of two Rho protomers (figure 4D). So Y274D may have similar defects as G324D and N340S changes.

*ii) Summary of different properties of the mutants.*

Different properties of the mutants are summarized in Table 1. All the termination defective mutants (except I382N) were found to be located in the previously identified functional domains such as in the primary RNA binding domain and in the secondary RNA binding domain including Q-loop and R-loops. The termination defect of the mutants G51V, Y80C and P279S could not be overcome under the most relaxed conditions that have been tested,

suggesting that primary RNA binding domain and Q-loop are the most crucial elements for RNA release activity. These mutants are also defective for most of the other functions of Rho. The termination defects of the mutants (Y274D, G324D and N340S), which are mainly defective for secondary RNA binding and as well as for the translocase activity, could be restored under relaxed *in vitro* conditions. The functional defects of most of the mutants correlate with their spatial localization in the crystal structure. We also show that mutations in the primary RNA binding domain (Y80C) can affect functions and induce conformational changes in the distal C-terminal domain which is not predictable from the structure of Rho (see below). However, we did not observe any severe *in vitro* defect in the I382N protein, which is not consistent with its *in vivo* phenotype; probably a modest *in vitro* defect can be amplified under more stringent *in vivo* conditions.

iii) Y80C change in the N-terminal region causes conformational changes in the distal C-terminal domain

Besides impairing the primary RNA binding, the Y80C mutation also caused significant defect in ATP binding. This led us to hypothesize that this change may cause conformational changes in the distal (~40 Å) C-terminal domain allosterically. To test this hypothesis we probed the conformation of WT

and mutant Rho by limited proteolytic cleavages, fluorescence anisotropy, and fluorescence quenching.

We monitored the surface accessibility of the single tryptophan residue (W381) located ~15 Å of the ATP-binding site, by fluorescence quenching technique using acrylamide as a neutral quencher. This tryptophan residue emits a fluorescence signal at 350 nm upon excitation at 295 nm. Quenching of this signal was plotted against the increasing concentration of acrylamide to obtain the quenching constant ( $K_{sv}$ ) (figure 5A). The value of  $K_{sv}$  increases as the tryptophan becomes more surface accessible. We observed that this tryptophan in the Y80C mutant is more surface accessible compared to the WT.

We used a fluorescent analogue of GTP, Tb-GTP, which is a complex of terbium chloride and GTP, to see the local conformational flexibilities at the ATP binding pocket. Here we assumed Tb-GTP will bind to the same ATP binding site as GTP is also a good substrate of Rho. We measured the anisotropy ( $r$ ) of the Tb-GTP moiety upon binding to the ATP binding pocket. The anisotropy ( $r$ ) of a fluorophore gives the measure of the rotational freedom of the

Table 1: Summary of different properties of Rho mutants:

Mutants	1 <sup>o</sup> RNA Binding	2 <sup>o</sup> RNA <sup>a</sup> Binding	ATP Binding	RNA:DNA <sup>c</sup> Unwinding
G51V	++	+/-	++	-
Y80C	-	-	<sup>b</sup>	-
Y274D	+	-	+	+/-
P279S	++	-	+	-
G324D	+	-	+	+/-
N340S	+	-	++	+/-
I382N	+	+	+++	+
WT	+	++	+++	+

All the activities of the mutants are expressed with respect to that of WT Rho.

<sup>a</sup> Binding activities are for rC<sub>10</sub> and rC<sub>25</sub> templates. +/- indicates ~100 fold reduced binding on rC<sub>25</sub> template.

<sup>b</sup> For ATP binding of Y80C the  $K_m$  value is considered.

<sup>c</sup> For the mutants Y274D, G324D and N340S the unwinding activity was only observed at 1 mM ATP.

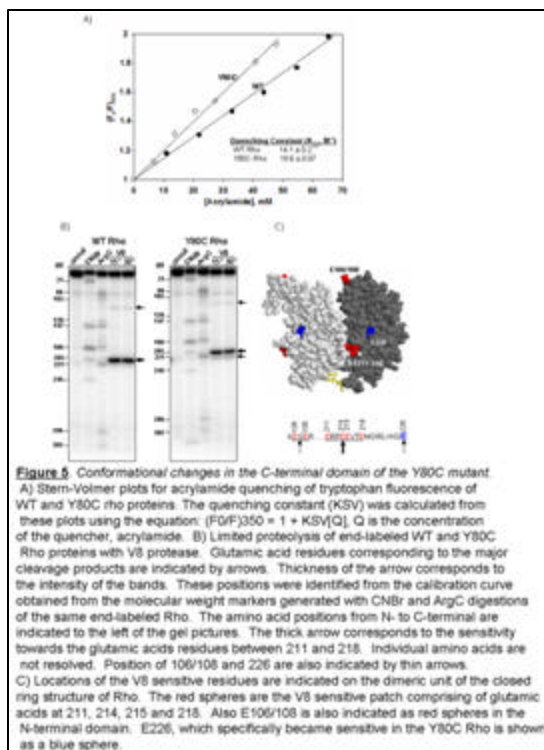


species and reports the conformational flexibility of the surroundings. Higher value of  $r$  means less rotational freedom of the fluorescent probe. Lower value of  $r$  for Tb-GTP bound to Y80C compared to that obtained for WT Rho (table 2) suggested more conformational disorders in the ATP-binding pocket.

Table 2: Fluorescence anisotropy ( $r$ ) values of Tb-GTP

Species	Anisotropy ( $r$ )
Free Tb-GTP [150 $\mu$ M: 50 $\mu$ M]	0.096 $\pm$ .005
Tb-GTP + 100 nM WT Rho	0.246 $\pm$ .007
Tb-GTP + 150 nM Y80C Rho	0.183 $\pm$ .003

In all the experiments Tb-GTP complexes were made by incubating 150  $\mu$ M Tb with 50  $\mu$ M GTP. Majority of the Tb-GTP species were in bound form in the presence of indicated amount of hexameric Rho. Anisotropies were measured at 25  $^{\circ}$ C.



In general, quenching constant and anisotropy values did not change drastically because of the Y80C mutation, which suggests that the conformational changes induced by the mutation in the Gterminal domain are more subtle. To further corroborate these data we employed limited proteolytic digestions of the WT and Y80C Rho proteins to probe the conformational changes with V8 protease, which cleaves preferably at glutamic acids residues. We observed that a cluster of surface exposed glutamic acid residues (figure 5B) near the dimeric interface of the C-terminal domain (figure 5C) of both the WT and Y80C mutant are very sensitive to this protease. Interestingly, a new band corresponding to

the Glu 226, close to this cluster (figure 5C and D), was found to become sensitive to V8 digestion specifically in Y80C mutant. This observation further supports the proposal that Y80C mutation in the primary RNA binding domain induces distinct but subtle conformational changes in the distal C-terminal domain, which might have affected the ATP binding domain and the surrounding regions.

### Publications:

- 1 Pani B, Banerjee S, Chalissery J, Abhishek M, Ramya ML, Suganthan R and Sen R (2006) Mechanism of inhibition of Rho-dependent transcription termination by bacteriophage P4 protein Psu. *Journal of Biological Chemistry*. 281: 26491-26500.

- 2 Chalissery J, Banerjee, S, Bandey I and Ranjan Sen (2007) Transcription termination defective mutants of Rho: role of different functions of Rho in releasing RNA from the elongation complex. ***Journal of Molecular Biology*** In Press.
- 3 Sen R, Chalissery J and Muteeb G (2007). Nus factors of Escherichia coli. Invited chapter of the book *Escherichia coli* and Salmonella: Cellular and Molecular Biology. ***ASM press*** In Press.

**OTHER SCIENTIFIC SERVICES/FACILITIES**



## **NATIONAL GENOMICS AND TRANSCRIPTOMICS FACILITY**

**Principal Investigator** Murali D Bashyam      Staff Scientist

**Other Members**                      Ajay Chaudhary              Technical Assistant  
   Chandrakanth Reddy      Project Assistant

### **Objectives:**

The National Genomics and Transcriptomics Facility (NGTF) is a Biotechnology services core group that provides on demand services in the areas of automated DNA sequencing, genotyping, real-time PCR and microarrays. Services currently available for CDFD groups and collaborating laboratories include:

- High throughput automated DNA sequencing and analyses
- Gene scan analysis and Genotyping including human STR profiling for forensic cases
- Real time PCR
- Microarrays

### **Details of progress made in the current reporting year (April 1, 2006- March 31, 2007)**

NGTF continues to provide services to scientists both within and outside CDFD in the areas of DNA sequencing, genotyping, real time PCR and microarrays. A total of 5.7 million nucleotides were sequenced during the reporting period. In addition, a total of 13200 genotyping reactions were performed. We also carried out a total of 105 real time PCR reactions for several laboratories within CDFD. Microarrays using the Human Exonic Evidence Based oligonucleotides (HEEBO) were procured from the Stanford Functional Genomics Facility and array-based comparative genomic hybridization as well as gene expression microarrays were standardized.



## Bioinformatics Facility and Sun Centre of Excellence in Medical Bioinformatics

<b>Service Co-ordinator</b>	Akash Ranjan	Staff Scientist
<b>Other members</b>	M Kavita Rao	Staff Scientist
	G Thanu	Technical Officer
	S Swaminathan	Technical Officer
	R K Gundu	Technical Officer (until December 2006)
	M N Pavan	Technical Assistant
	P Prashanthi	Junior Assistant
	A Imam	Project Assistant

**Collaboration** European Molecular Biology Network(EMBnet)  
SUN Microsystems, Inc, California, USA  
C-DAC, Pune, India  
University of Hyderabad, Hyderabad

### **Objectives:**

1. To provide Bioinformatics services to the Centre and the scientific community in the region
2. To maintain the CDFD website, to provide web based services and e-mail services.
3. To maintain Computer Servers, PCs, Printers and other peripheral devices.
4. To maintain the internet connectivity of the institute and institute-wide Local Area Network.

### **Details of progress made in the current reporting year (April 1, 2006 – 31, March 2007)**

Services such as Email, Internet, web were taken care of at Nacharam and Gandipet campuses. Installation, administration and maintenance of servers which provides various services, databases and computational jobs were taken care of at both the campuses.

Installation, administration and maintenance of PCs, Printers and Scanners were also done. Networking of servers and all the PCS and LAN were also administered and maintained. Also maintenance of other network components such as routers, switches, firewall were taken care. AMC for Sunserver and SGI Fuel machines were renewed.

Website updating were also done as per requirement. New Databases were added and given public access. CDFD domain name and other software licenses were renewed. An in

house package for online application for Research Scholar Program – 2006 selection was developed using freeware. We have added a number of in house developed scientific web server and databases which includes – genoBASE pylori, INSATDB, MycoperonDB, IMEX and ICR.

In addition to Nacharam, CDFD website was also configured in a Sun Blade at SUN CoE Gandipet, so that our site is always accessible should there be a failure at any one place. Changes done at one place are updated immediately at the other place also. Website was moved to a high end PC in Nacharam. CDFD domain was registered in a self-manageable DNS service; hence we can manage all the DNS records ourselves from the control panel. CDFD Domain name was renewed for 2 more years.

CDFD is one of the eight nodes of the high speed Virtual Private Network called “India Bio-Grid” and was offered a leased line along with accessories with ISDN backup. This has been renewed and is mainly used for downloading data. This is also used for Internet access when 2 Mbps leased line goes down.

High-end PCs (70 Nos.) and laptops (2 Nos.) and printers (25 Nos.) were procured and installed. To cope up with the rising networking demands several switches were procured and new PCs were connected to the existing LAN.

Symantec corporate edition was upgraded to Enterprise edition 10.0 and all the PCs were regularly scanned. In addition to this Symantec Mail security 3.0 was upgraded to 5.0 with Antispam support. This scans for all the viruses, worms etc in the incoming and outgoing mails and also completely stops SPAM.

### **Databases and Software**

The databases were downloaded and updated regularly as and when new releases became available from their respective host sites. Software required for DNA/protein sequence analysis and structure prediction was downloaded and installed in the SUN system.

### **Networking/Internet maintenance**

The ISDN backup system for the VPN was completely installed. An assessment of performance of the leased line as well as the ISDN backup line was performed and an expert report was submitted to DBT.



## **SUN-Centre of Excellence (COE)**

SUN microsystems, a leading hardware manufacturer signed an MOU with CDFD to set up one of its COEs at the bioinformatics facility. The COE will have a major focus on the Medical bioinformatics. The new setup has come up in the newly constructed building of CDFD near Gandipet. The “Centre of Excellence in Medical Bioinformatics” was inaugurated at the Gandipet campus of CDFD by His Excellency, Dr. A.P. J. Abdul Kalam, and dedicated to the service of the Nation. This is also the first SUN-CoE (SUN Centre of Excellence) in the Indian sub-continent. This Centre is a joint venture between M/s. Sun Microsystems, CDFD and the Government of Andhra Pradesh. The SUN-COE is equipped with mostly all requisite facilities and infrastructure back-bone to facilitate carrying out high impact research in Bioinformatics.

### **SUN-COE Activities**

Sun CoE provided a six months project training to M.Sc Bioinformatics, Jamia Milia Islamia University, New Delhi which was attended by 9 students of Jamia Milia Islamia University, New Delhi.

Sun CoE also participated in University of Hyderabad’s M.Tech. Bioinformatics Program and provided lecture and practical training the field of Bioinformatics 18 students of M.Tech Bioinformatics program of University of Hyderabad.

Sun CoE Coordinator, Dr. Akash Ranjan, organized an Indo German Workshop at University of Cologne Germany which was attended by leading Bioinformatics Scientist from India and Germany. He was also invited to University of Wuerzberg as Bioinformatics Expert and Resource person at a Euro Patho Genomics-sponsored Late Summer School on Structural Genomics.

SUN CoE, CDFD has entered into a memorandum of understanding with GDAC and ERNET to be a partner institute in the GARUDA National Grid Computing Initiative. In the initial proof of concept phase, we have signed a User Acceptance Test of the 10Mbps high speed connectivity. This enables our researchers to seamlessly gain access at 10Mbps to the Supercomputing Facility at C-DAC, Bangalore, and also other computational resources to be made available by Pune, Hyderabad, and Chennai CDAC centres and other 45 research and academic partner institutes spread across 17 cities in India. Recently the GARUDA SIGMA middleware has been deployed in our Grid head node which is a suite of grid tools and services required to facilitate access to GARUDA resources. Having

successfully conducted several bandwidth tests for the network fabric, we would soon involve in the development of applications on GARUDA wherein we could also explore joint experiments with other partner institutes and collaborate on the research and engineering of technologies, architectures, standards and applications in High Performance Computing and Bioinformatics related areas.

The hardware installed at our premises as part of the GARUDA National Grid Computing Initiative enables connectivity at 10Mbps via the last mile fiber link, provided by ERNET representative M/s Sify using CISCO routers, to other scientific partners of this project.

## Instrumentation Services

<b>Service co-ordinator</b>	Raghavendrachar J	Staff Scientist
<b>Other Members</b>	R N Mishra	Technical Officer I
	S Pavan Kumar	Technical Officer I
	M Laxman	Technical Assistant
	R M K Sathyanarayana	Technical Assistant
	N P Sharma	Junior Assistant-II
	Dinesh Kumar	Junior Assistant-II

### Objectives:

To maintain, repair and service all the equipments in laboratory. To provide pre-installation requirements for new instruments and to coordinate with the manufacturers / their agents in Installation and warranty service of the new instruments. Also to provide the reports on the newly arrived instruments and to follow up with the suppliers for short shipped items.

### Details of progress made in the current reporting year (April 1, 2006 – March 31, 2007)

We have performed over 76 new installations of various equipment, like Confocal Microscope, HPLC System, LPLC System, PCR Machines, Water Purification Systems, BOD Incubator, Nanodrop Spectrophotometers, Refrigerated Table Top Centrifuges, Shaking Water baths, Orbital Shaker, Tissue Homogenizer, -86°C Freezer, -20°C Freezers, Refrigerators, Micro centrifuges, Electrophoresis Apparatus etc.

We have also completed 440 work orders for repair & maintenance of various laboratory equipments. We were actively involved in reorganizing the Laboratory for additional space and also the newly leased space in the Ground Floor. We were actively involved in getting the Confocal Microscope Facility functional in the ground floor and also created the NGTF DNA Sequencing facility in the Ground floor. We have played active role getting all instruments installed in the APEDA facility in the Second Floor. In addition we were involved in organizing the audio & visual requirements for presentations in various seminars, lectures and workshops, Foundation day lectures, Distinguished Scientist Lectures held in CDFD both at Nacharam and Gandipet Auditorium. We maintained most of the equipment with maximum uptime in the Laboratory. Most of the Instruments are maintained by our Instrumentation staff, there by saving on the expensive AMCs and with very little downtime of the equipment.

Since we have been entrusted with the additional responsibility of Total maintenance of the CDFD Nacharam Campus, including Electrical, Refrigeration, EPABX & Telephones, UPS Systems, House Keeping, we are successfully monitoring these activities and also procuring the Laboratory tables & getting the partitions done for various group leaders & facilities.

## **PUBLICATIONS**



## **Research Papers Published (2006)**

- 1 Achary MS, Reddy ABM, Chakrabarti S, Panicker SG, Mandal AK, Ahmed N, Balasubramanian D, Hasnain SE and Nagarajaram HA (2006) Disease-causing mutations in proteins: structural analysis of the CYP1B1 mutations causing primary congenital glaucoma in humans. *Biophysical Journal* 91:4329-4339.
- 2 Ahmed N, Devi SM, Valverde M de LA, Vijayachari P, Machangu RS, Ellis WA and Hartskeerl RA (2006) Multilocus sequence typing method for identification and genotypic classification of pathogenic leptospira species. *Annals of Clinical Microbiology and Antimicrobials* 5:28.
- 3 Akif M, Akhter Y, Hasnain SE and Mande SC (2006) Crystallization and preliminary X-ray crystallographic studies of *Mycobacterium tuberculosis* CRP/FNR family transcription regulator. *Acta Crystallographica F* 62:873-875.
- 4 Arunkumar KP and Nagaraju J (2006) Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. *PLoS ONE* 1: e110.
- 5 Devi SM, Ahmed I, Khan AA, Rahman SA, Alvi A, Sæchi LA and Ahmed N (2006) Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type cag-pathogenicity island. *BMC Genomics* 7:191.
- 6 Devi AR, Gopikrishna M, Ratheesh R, Savithri G, Swamalata G and Bashyam MD (2006) Farberlipogranulomatosis-clinical and molecular genetic analysis reveals a novel mutation in an Indian family. *Journal of Human Genetics* 51:811-814.
- 7 Du X, Rao MR, Chen XQ, Wu W, Mahalingam S and Balasundaram D (2006) The homologous putative GTPases Grn1p from fission yeast and the human GNL3L are required for growth and play a role in processing of nucleolar pre-rRNA. *Molecular Biology of the Cell* 17:460-474.
- 8 Gandhe AS, Arunkumar KP, John SH and Nagaraju J (2006) Analysis of bacteria-challenged wild silkworm, *Antheraea mylitta* (lepidoptera) transcriptome reveals potential immune genes. *BMC Genomics* 7:184.

- 9 Goyal K, Qamra R and Mande SC (2006) Multiple gene duplication and rapid evolution in the groEL gene: functional implications. ***Journal of Molecular Evolution*** 63:781-787.
- 10 \*\*Granneman S, Lin C, Champion EA, Nandineni MR, Zorca C and Baserga SJ (2006) The nucleolar protein Esf2 interacts directly with the DExD/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. ***Nucleic Acids Research*** 34:3189-3199.  
( \*\* from work done elsewhere)
- 11 Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Loch C and Supply P (2006) Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India suggests an ancient focus of tuberculosis in South Asia. ***Emerging Infectious Diseases*** 12:367-374.
- 12 Kazim SN, Sarin SK, Sharma BC, Khan LA and Hasnain SE (2006) Characterization of naturally occurring and Lamivudine-Induced surface gene mutants of hepatitis B virus in patients with chronic hepatitis B in India. ***Intervirology*** 49:152-160.
- 13 Kenchappa P, Duggirala A, Ahmed N, Pathengay A, Das T Hasnain SE and Sharma S (2006) Fluorescent amplified fragment length polymorphism (FAFLP) genotyping demonstrates the role of biofilm-producing methicillin-resistant periocular *Staphylococcus epidermidis* strains in postoperative endophthalmitis. ***BMC Ophthalmology*** 6:1.
- 14 Khan N, Rahim SS, Boddupalli CS, Ghousunnissa S, Padma S, Pathak N, Thiagarajan D, Hasnain SE and Mukhopadhyay S (2006) Hydrogen peroxide inhibits IL-12 p40 induction in macrophages by inhibiting c-rel translocation to the nucleus through activation of calmodulin protein. ***Blood*** 107:1513-1520.
- 15 Khosla S, Mendiratta G and Brahmachari V (2006) Genomic imprinting in the mealybugs. ***Cytogenetics and Genome Research*** 113:41-52.
- 16 Manna SK (2006) Alpha-MSH and IL-8-induced biological responses. ***Modern Aspects of Immunobiology*** 19:32-33.



- 17 Pani B, Banerjee S, Chalissery J, Abhishek M, Ramya ML, Suganthan R and Sen R (2006) Mechanism of inhibition of Rho-dependent transcription termination by bacteriophage P4 protein Psu. ***Journal of Biological Chemistry***. 281: 26491-26500.
- 18 Qamra R, Prakash P, Aruna B, Hasnain SE and Mande SC (2006) The 2.15Å crystal structure of *M. tuberculosis* chorismate mutase reveals unexpected gene duplication and suggests a role in host-pathogen interactions. ***Biochemistry***45:6997-7005.
- 19 Ranjan S, Seshadri J, Vindal V, Yellaboina S and Ranjan A (2006) iCR: a web tool to identify conserved targets of a regulatory protein across the multiple related prokaryotic species. ***Nucleic Acids Research*** 34: W 584-587.
- 20 Ranjan S, Gundu RK and Ranjan A (2006) MycoPeronDB: a database of computationally identified operons and transcriptional units in Mycobacteria. ***BMC Bioinformatics*** 7: 59.
- 21 Ranjan S, Yellaboina S and Ranjan A (2006) IdeR in mycobacteria: From target recognition to physiological function. ***Critical Reviews in Microbiology*** 32:69-75.
- 22 Rao KR, Ahmed N, Sriramula S, Sechi LA and Hasnain SE (2006) Rapid identification of *Mycobacterium tuberculosis* Beijing genotypes on the basis of the mycobacterial Interspersed repetitive Unit locus 26 signature. ***Journal of Clinical Microbiology*** 44:274-277.
- 23 Rao MR, Kumari G, Balasundaram D, Sankaranarayanan R, and Mahalingam S (2006) A novel lysine-rich domain and GTP binding motifs regulates the nucleolar retention of human guanine nucleotide binding protein, GNL3L. ***Journal of Molecular Biology*** 364:637-654.
- 24 Sarkar S, Wise KC, Manna SK, Ramesh V, Yamauchi K, Thomas RL, Wilson BL, Kulkarni AD, Pellis NR and Ramesh GT (2006) Activation of activator protein-1 in mouse brain regions exposed to stimulated microgravity. ***In Vitro Cell Developmental Biology – Animal*** 42:96-99.
- 25 Sechi LA, Gazouli M, Sieswerda L, Mollicotti P, Ahmed N, Ikonopoulou J, Scanu AM, Paccagnini D and Zanetti S. (2006) Relationship between Crohn's disease, infection with *Mycobacterium avium* subspecies paratuberculosis and Slc11a1 gene polymorphisms in Sardinian patients. ***World Journal of Gastroenterology*** 12:7161-7164.

- 26 Singhal PK, Kumar PR, Rao MR, Kyasani M and Mahalingam S (2006) Simian Immunodeficiency Virus Vpx is imported into the nucleus via importin alpha dependent and independent pathways. ***Journal of Virology***80:526-536.
- 27 Singhal PK, Kumar PR, Rao MR and Mahalingam S (2006) Nuclear export of Simian Immunodeficiency Virus Vpx protein. ***Journal of Virology*** 80:12271-12278.
- 28 Sitalaxmi T, Kashyap VK, Guha S, Hima Bindu G, Hasnain SE and Trivedi R (2006) Genetic structure of Indian populations based on fifteen autosomal microsatellite loci. ***BMC Genetics*** 7:28.
- 29 Sreenivasan Y, Raghavendra PB and Manna SK (2006) Oleandrin-mediated expression of Fas potentiates apoptosis in tumor cells. ***Journal of Clinical Immunology*** 26:308-322.
- 30 Tundup S, Akhter Y, Thiagarajan D and Hasnain SE (2006) Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. ***FEBS Letters*** 580:1285-1293.
- 31 Yellaboina S, Ranjan S, Vindal V and Ranjan A (2006) Comparative analysis of iron regulated genes in mycobacteria. ***FEBS Letters***580:2567-2576.
- 32 Zanetti S, Molicotti P, Cannas S, Ortu S, Ahmed N and Sechi LA (2006). *In vitro* activities of antimycobacterial agents against *Mycobacterium avium* subsp. paratuberculosis linked to Crohn's disease and paratuberculosis. ***Annals of Clinical Microbiology and Antimicrobials***5:27.

## Research Papers Published 2007 (until 31 March)

- 1 Akhtar Y, Ahmed I, Devi SM and Ahmed N (2007). The co-evolved *Helicobacter pylori* and gastric cancer: Trinity of bacterial virulence, host susceptibility and lifestyle. ***Infectious Agents and Cancer*** 2:2.
- 2 Archak S, Meduri E, Kumar PS and Nagaraju J (2007) InSatDb: a microsatellite database of fully sequenced insect genomes. ***Nucleic Acids Research*** 35:D36-D39.
- 3 Khan N, Ghousunnissa S, Jegadeeswaran SM, Thiagarajan D, Hasnain SE and Mukhopadhyay S (2007) Anti-B7-1/B7-2 antibody elicits innate-effector responses in macrophages through NF- $\kappa$ B-dependent pathway. ***International Immunology*** 19:477-486.
- 4 Manna SK, Manna P and Sarkar A (2007) Inhibition of RelA phosphorylation sensitizes chemotherapeutic agents-mediated apoptosis in constitutive NF-kappaB-expressing and chemoresistant cells. ***Cell Death & Differentiation*** 14:158-170.
- 5 Meglec E, Anderson SJ, Bourguet D, Butcher R, Caldas A, CasseLundhagen A, d'Acier AC, Dawson DA, Faure N, Fauvelot C, Franck P, Harper G, Keyghobadi N, Kluetsch C, Muthulakshmi M, Nagaraju J, Patt A, Petenian F, Silvain JF and Wilcock HR (2007) Microsatellite flanking region similarities among different loci within insect species. ***Insect Molecular Biology*** 16:175-185.
- 6 Naushad SM, Jamal NJ, Angalena R, Prasad CK, and Devi AR (2007) Hyperhomocysteinemia and the compound heterozygous state for methylene tetrahydrofolate reductase are independent risk factors for deep vein thrombosis among South Indians. ***Blood Coagulation & Fibrinolysis*** 18:113-117. Raghavendra PB, Sreenivasan Y, Ramesh GT and Manna SK (2007) Cardiac glycoside induces cell death via FasL by activating calcineurin and NF-AT, but apoptosis initially proceeds through activation of caspases. ***Apoptosis*** 12:307-318.

- 8 Raghavendra PB, Sreenivasan Y and Manna SK (2007) Oleandrin induces apoptosis in human, but not in murine cells: dephosphorylation of Akt, expression of FasL and alteration of membrane fluidity. *Molecular Immunology* 44:2292-2302.
- 9 Sreenu VB, Kumar P, Nagaraju J and Nagarajaram HA (2007) Simple sequence repeats in mycobacterial genomes *Journal of Biosciences* 32:1-15.
- 10 Vidyasagar M, Hasnain SE, Mande SC, Nagarajaram HA, Ranjan A, Acharya MS, Anwaruddin, Arun SK, Gyanraj Kumar, Kumar D, Priya S, Ranjan S, Reddi BR, Seshadri J, Sravan Kumar P, Swaminathan S, Umadevi P and Vindal V (2007) BioSuite: A comprehensive bioinformatics software package (A unique industry-academia collaboration) *Current Science* 92:29-38.

## **RESEARCH PAPERS – IN PRESS (as on 31 March 2007)**

- 1 Ahmed N, Majeed AA, Ahmed I, Hussain MA, Alvi A, Devi SM, Rizwan M, Ranjan A, Sechi LA, and Megraud F (2007) genoBASE pylori: A genotype search tool and database of human gastric pathogen *Helicobacter pylori*. ***Infection Genetics and Evolution***.
- 2 Archak S, Reddy VLN and Nagaraju J (2007) High-throughput multiplex microsatellite marker assay for detection and quantification of adulteration in Basmati rice (*Oryza sativa*). ***Electrophoresis***
- 3 Banerjee S, Nandyala AK, Raviprasad P, Ahmed N and Hasnain SE (2007) Iron dependent Iron binding activity of *Mycobacterium tuberculosis* aconitase. ***Journal of Bacteriology***.
4. Bashyam MD and Hasnain SE (2007) Array-based comparative Genomic Hybridization: applications in cancer and tuberculosis. Bioarrays, Ed. K Appasani, ***Humana press***.
5. Bashyam MD, Savithri GR, Gopikrishna M and Narsimhan C (2007) A p.R870H mutation in the beta cardiac myosin heavy chain 7 gene causes Familial Hypertrophic Cardiomyopathy in several members of an Indian family. ***Canadian Journal of Cardiology***.
6. Boddupalli CS, Ghosh S, Rahim SS, Nair S, Ehtesham NZ, Hasnain SE and Mukhopadhyay S (2007) Nitric oxide inhibits interleukin-12 p40 through p38 MAPK-mediated regulation of calmodulin and c-rel. ***Free Radical Biology and Medicine***.
7. Chalissery J, Banerjee, S, Bandey I and Ranjan Sen (2007) Transcription termination defective mutants of Rho: role of different functions of Rho in releasing RNA from the elongation complex. ***Journal of Molecular Biology***.
8. Gandhe AS, Gude J and Nagaraju J (2007) Immune upregulation of novel anti-bacterial proteins from silkworms (Lepidoptera) that resemble lysozymes but lack muramidase activity. ***Insect Biochemistry and Molecular Biology***.
9. Gokul G, Gautami B, Malathi S, Sowjanya A, Poli U, Jain M, Ramakrishna G and Khosla S (2007) DNA methylation profile at the DNMT3L promoter: a potential biomarker for cervical cancer. ***Epigenetics***.
10. Goyal K, Mohanty D and Mande SC (2007) PAR-3D: A server to predict protein active site residues. ***Nucleic Acids Research***.

11. \*\*\* Kaur R, Ma B and Cormack BP (2007) A family of aspartyl proteases is required for virulence of *Candida glabrata*. ***Proceedings of the National Academy of Sciences, USA***.  
(\*\*\* work done from elsewhere)
12. Kumari G, Singhal PK, Rao MR, Mahalingam S (2007) Nuclear transport of Ras-associated tumor suppressor proteins: Different transport receptor binding specificities for arginine-rich nuclear targeting signals. ***Journal of Molecular Biology***.
13. Laishram RS and Gowrishankar J (2007) Environmental regulation operating at the promoter clearance step of bacterial transcription. ***Genes & Development***.
14. Manna SK, Aggarwal RA, Sethi G, Aggarwal BB and Ramesh GT (2007) Morin (3,5,7,2',4'-pentahydroxyflavone) abolishes NF- $\kappa$ B activation induced by various carcinogens and inflammatory stimuli, leading to suppression of NF- $\kappa$ B-regulated gene expression and upregulation of apoptosis. ***Clinical Cancer Research***.
15. Mukhopadhyay S, Nair S and Hasnain SE (2007) Nitric oxide: Friendly rivalry in tuberculosis. ***Current Signal Transduction Therapy***.
16. Rasheedi S, Aruna R, Nasreen ZE and Hasnain SE (2007) Biochemical characterization of Sf9 Sp family-like protein factors reveals interesting features. ***Archives of Virology***.
17. Rasheedi S, Ghosh S, Suragani M, Tuteja N, Sopory SK, Hasnain, S E, Nasreen, Z E (2007) Pisum sativum Contains A Factor With Strong Homology to CIF5B. ***Gene***.
18. Sailu Y, Goyal K and Mande SC (2007) Inferring genome-wide functional linkages in *E. coli* by combining improved genome context methods: comparison with high-throughput experimental data. ***Genome Research***.
19. Salih B, Abasiyanik F, Ahmed N (2007). A preliminary study on the genetic profile of cag pathogenicity island and other virulent gene loci of *Helicobacter pylori* strains from Turkey. ***Infection Genetics and Evolution***.
20. Savitri S, Aparna D, Prashanth K, Justine KP, Ahmed N, Prashan K, Das T and Hasnain S E (2007) High-resolution genome profiling differentiated *Staphylococcus epidermidis* strains from patients with ocular infections and normal individuals: A Pilot study. ***Journal of Clinical Microbiology***.
21. Sechi LA, Ruhel A, Ahmed N, Usai D, Paccagnini D, Felis G and Zannetti S (2007) *Mycobacterium avium* subspecies paratuberculosis is able to infect and persist in enteric glial cells. ***World Journal of Gastroenterology***.

22. Sen R, Chalissery J and Muteeb G (2007). Nus factors of *Escherichia coli*. Invited chapter of the book *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. **ASM press**.
23. Suresh BM and Nagarajaram HA (2007) IMEx: Imperfect Microsatellite Extractor. **Bioinformatics**.
24. Vindal V, Ranjan S and Ranjan A (2007) In silico analysis and characterization of GntR family of regulators from *Mycobacterium tuberculosis*. **Tuberculosis**
25. Yusuf A, Smanla T and Hasnain SE (2007) Novel Biochemical Properties of a CRP/FNR Family Transcription Factor from *Mycobacterium tuberculosis*. **International Journal of Medical Microbiology**.

**Patents :**

1. Gowrishankar J and Harinarayanan R A method of altering levels of plasmids. Application No. 10/266,510: US Patent No. 7,176,028 B2 issued February 13, 2007.
2. Gowrishankar J and Harinarayanan R A method of altering levels of plasmids. Application No. 11/377,380: Notice of Allowance dated 12 January, 2007 issued by the US Patent Office.

**Other Publications :**

1. Nandineni MR (Book Review, 2006): Annual Review of Genomics and Human Genetics Vol.6, 2005. Aravinda Chakravarti and Eric Green (eds). **Current Science** 91:1413-1415.