

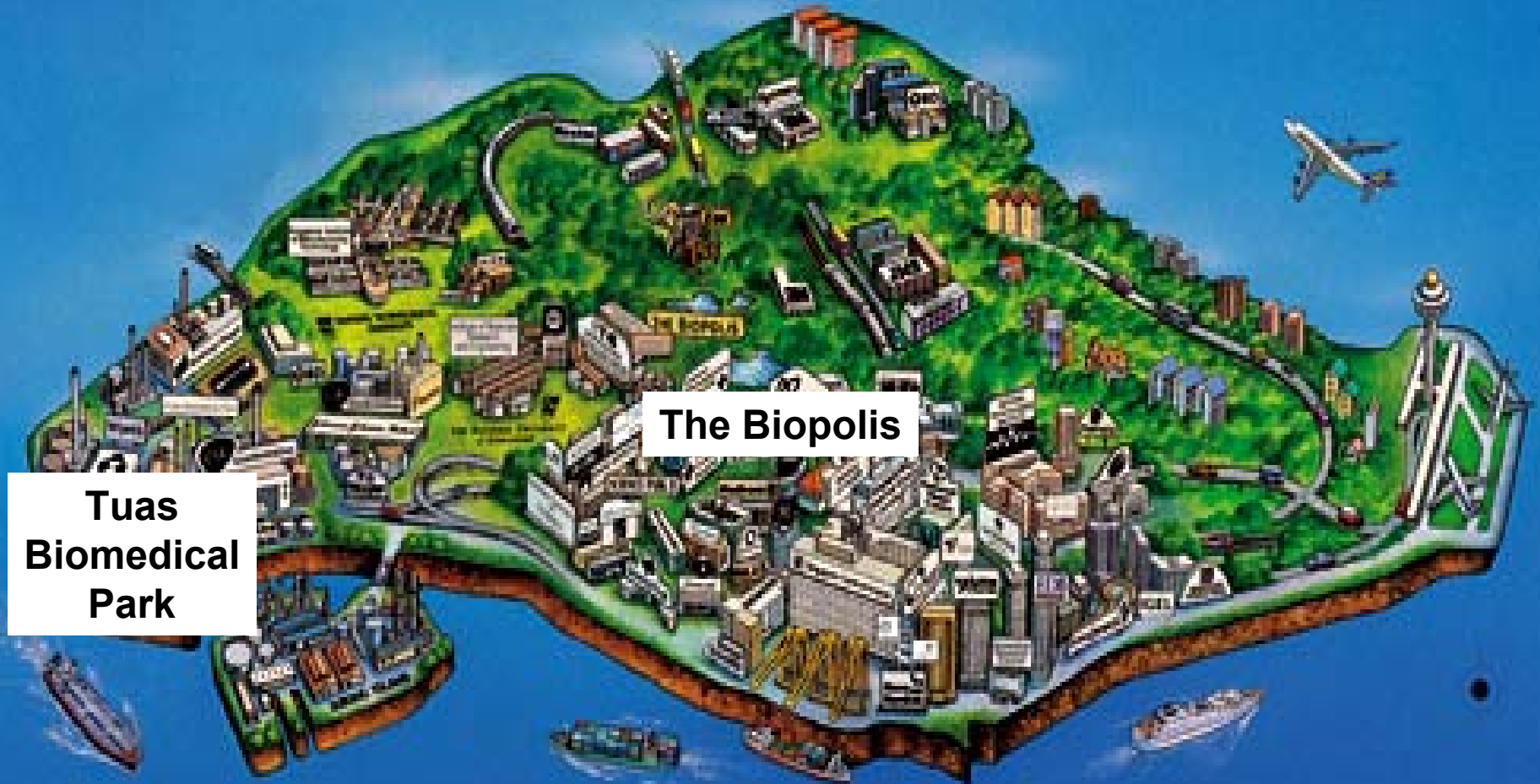
# **Bioinformatics Institute**

*Opening New Frontiers in Biology*

**Dr. Guna Rajagopal**

**Developing Biomedical Applications on the IA64**

# Global City, Island Nation



**Tuas  
Biomedical  
Park**

**The Biopolis**

**To establish Singapore as a global biomedical hub**



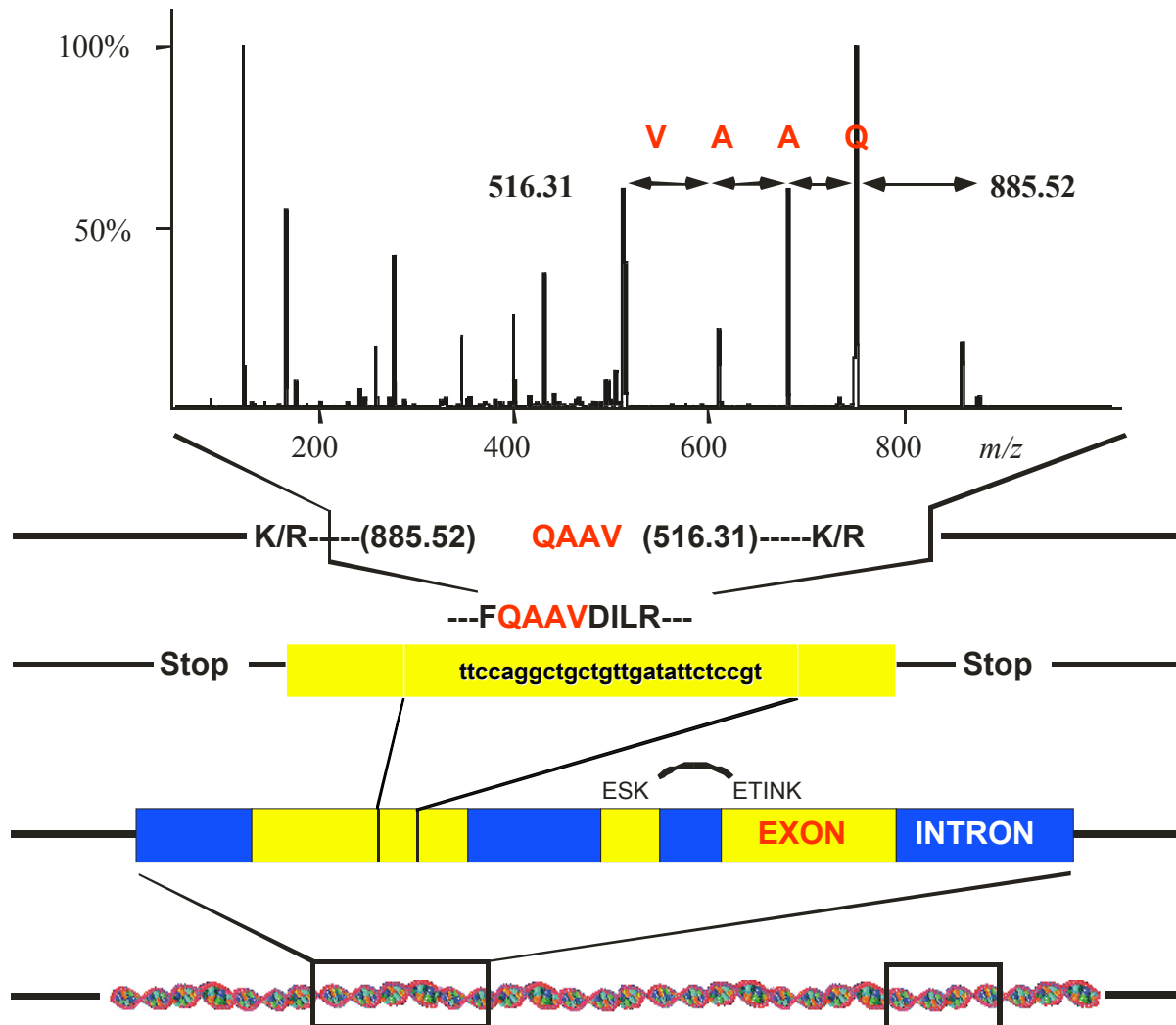
**BII**

**BII@Biopolis**

# Applications: Proteomics

**Work done by: Dr. Arun Krishnan, Mr. Atif Shahab  
Mr. Choon Peng, Mr. Lai LF & IT team.**

# Identifying proteins in large genome databases



tandem mass spectrum

peptide sequence tag

DB query

exon

gene

genome

***Trash*** : A code that is used in proteomics to analyze high-throughput mass spectrometry data. (in collaboration with Dr. Newman Sze of the Genome Institute of Singapore)

- Code originally from Cornell Proteomics Group
- Ported from PV-WAVE to C, single CPU application.
- Ported application compiles for 32 bit and 64 bit architectures.
- Floating point ops especially FFT's a key element determining code performance.
- Reduced the execution time from more than 1 hour per spectrum to 5 minutes on Itanium 1.
- High throughput analysis on Itanium cluster.

# Applications: Bioinformatics

**Work done by:**

**Scientists:** Dr. Arun Krishnan, Dr. Li Kuo Bin  
& team members.

**IT :** Lai LF, Stephen Wong and their team.

- Completed porting of popular bioinformatics software on the Itanium platform: **NCBI BLAST, ClustalW-MPI, EMBOSS, T-Coffee, Wise**, etc.
- Collaboration with UCSD/SDSC via PRAGMA on **Encyclopedia of Life (EOL)** comparative genomics project involving grid based pipeline/workflow for protein structure and functional annotation. Itanium 1 clusters initially.
- **Exhaustive Tandem Repeats Search Algorithm** in whole genome analysis, implemented on IA64.
- Itanium is used as the primary machine for debugging and testing the implementation. Results accepted for publication in *Bioinformatics (2004)*.



# Tandem Repeats Search

## ● Tandem Repeats

- Two adjacent approximate copies of a sequence of nucleotides: e.g.
  - AAAT AAGT
  - TAAA GAAA TAAA
- Frequently occur in genomes
- Unusually long repeats have been related to diseases, e.g. Alzheimer's

- Exhaustive Tandem Repeats Search program in C, memory intensive.
- Exhaustively find tandem repeats in whole genomes
  - Mostly character-based comparisons, and **integer arithmetic**
  - Some floating-point arithmetic

- Organism: *Archaeoglobus fulgidus*

- About 2 Mbp

- Search all tandem repeats of pattern size between 2 and 500

- Compilers:

- UltraSPARC: Sun Forte

- Athlon: gcc

- Intel: Intel Compiler 7.1

**Hardware:**

Unbranded Athlon XP 2000 1.67GHz, Linux

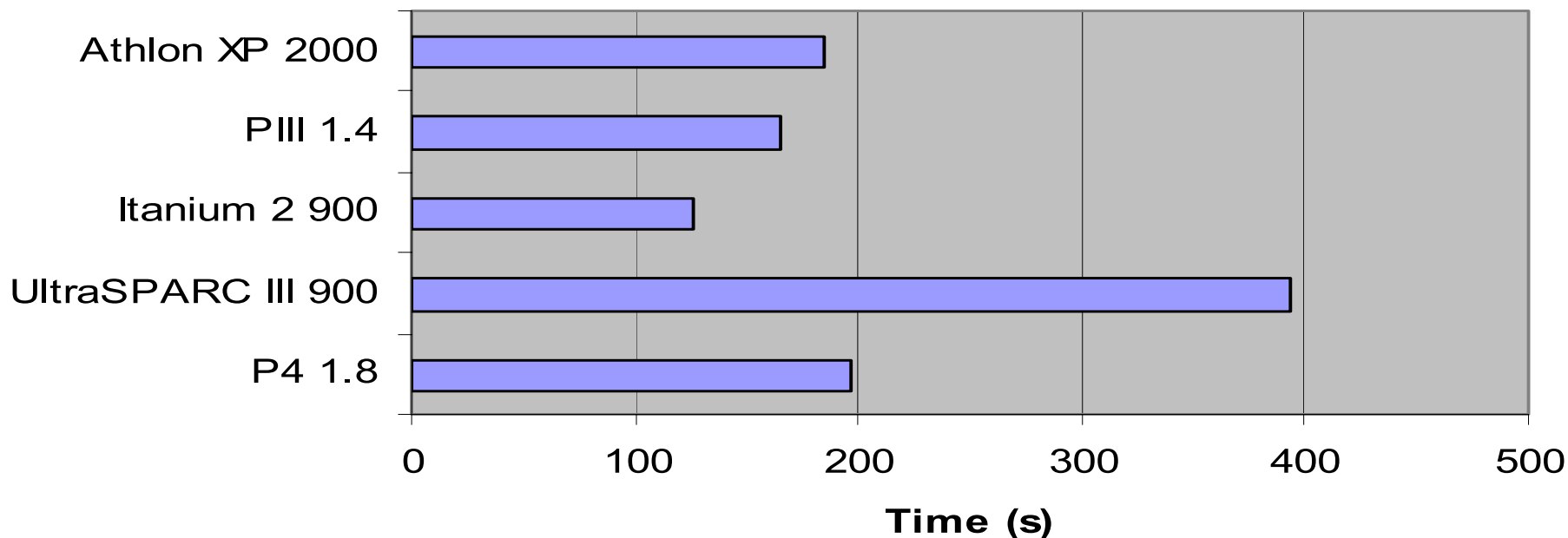
HP DL360-G2 Pentium III 1.4GHz, Linux

HP rx2600 Itanium 2 900MHz, Linux

Sun Fire v880 UltraSPARC III 900MHz, Solaris

Dell Optiplex GX260 Pentium 4 1.8GHz, Linux

## Performance



# Applications: Molecular Dynamics

Work done by: **Dr. Chandra Varma** & team members  
With help form the IT team.

- **GROMACS** is used to perform molecular dynamics simulation for biological systems with hundreds to millions of particles.
- Tested both Parallel and non-Parallel version of GROMACS on the Itanium.
- **CHARMM** - Chemistry at HARvard Molecular Mechanics
- Have ported and tested CHARMM on the Itanium Platform, both single and multiple-CPU version.

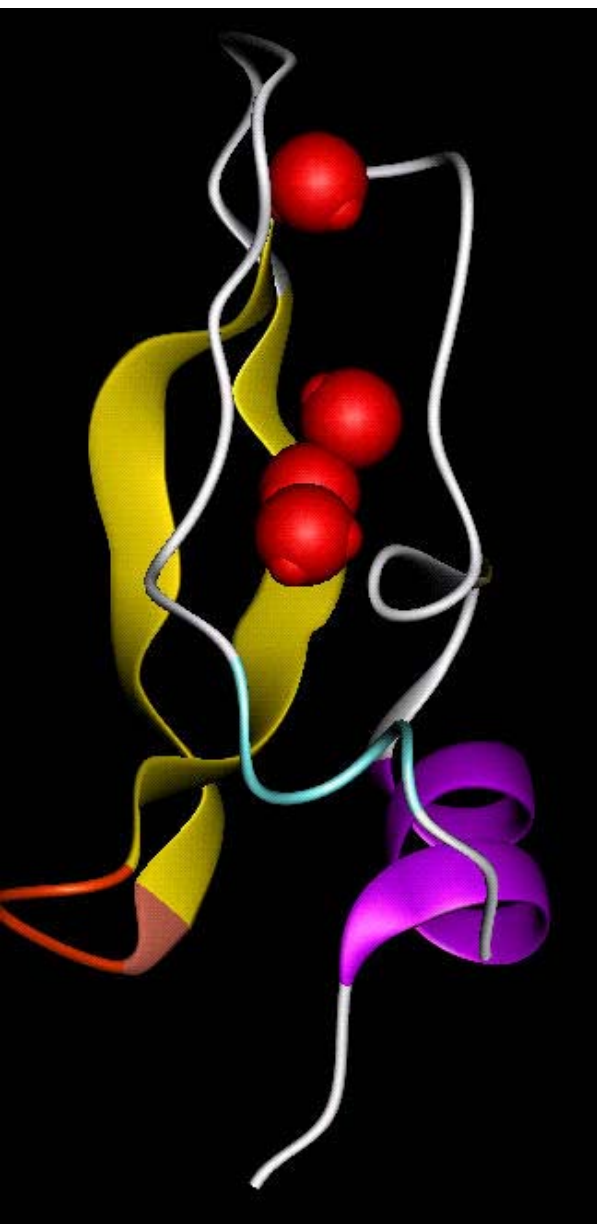
# *ThermoDynamics of buried water in proteins*

## CHARMM on IA64

*Using various simulation revealing surprising insights into how water molecules deep inside proteins modulate structural dynamic and functional aspects of proteins. (Work to be published in J. Phy. Chem. B)*

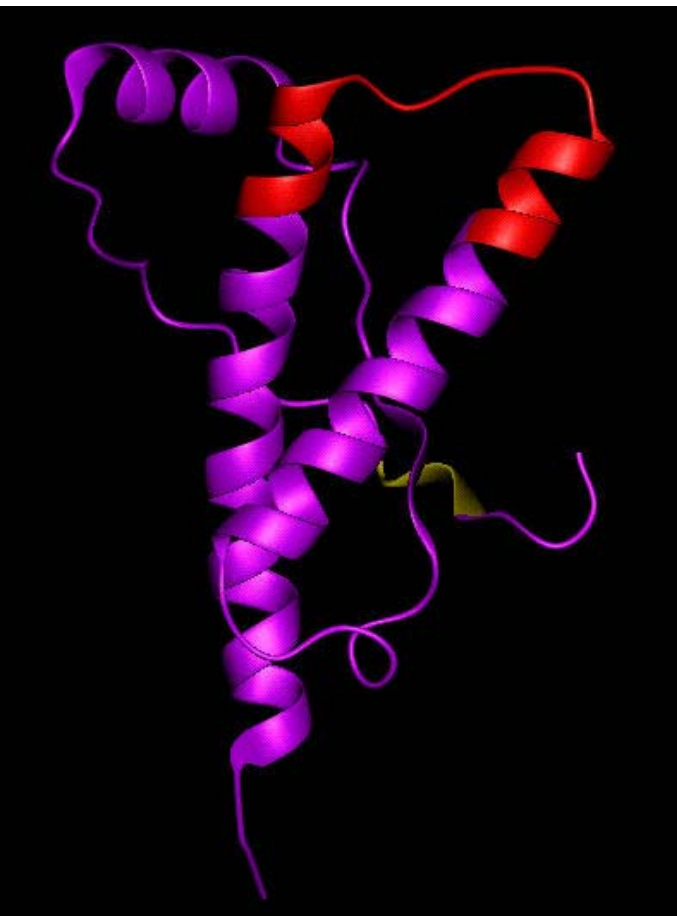
- *Relevance in drug design & development of drug resistance mechanisms*
- *Protein Engineering*
- *Biomaterials*

*The crystal structure of BPTI with 4 buried waters shown as spheres*



# Dynamics of prion proteins

## GROMACS on IA64

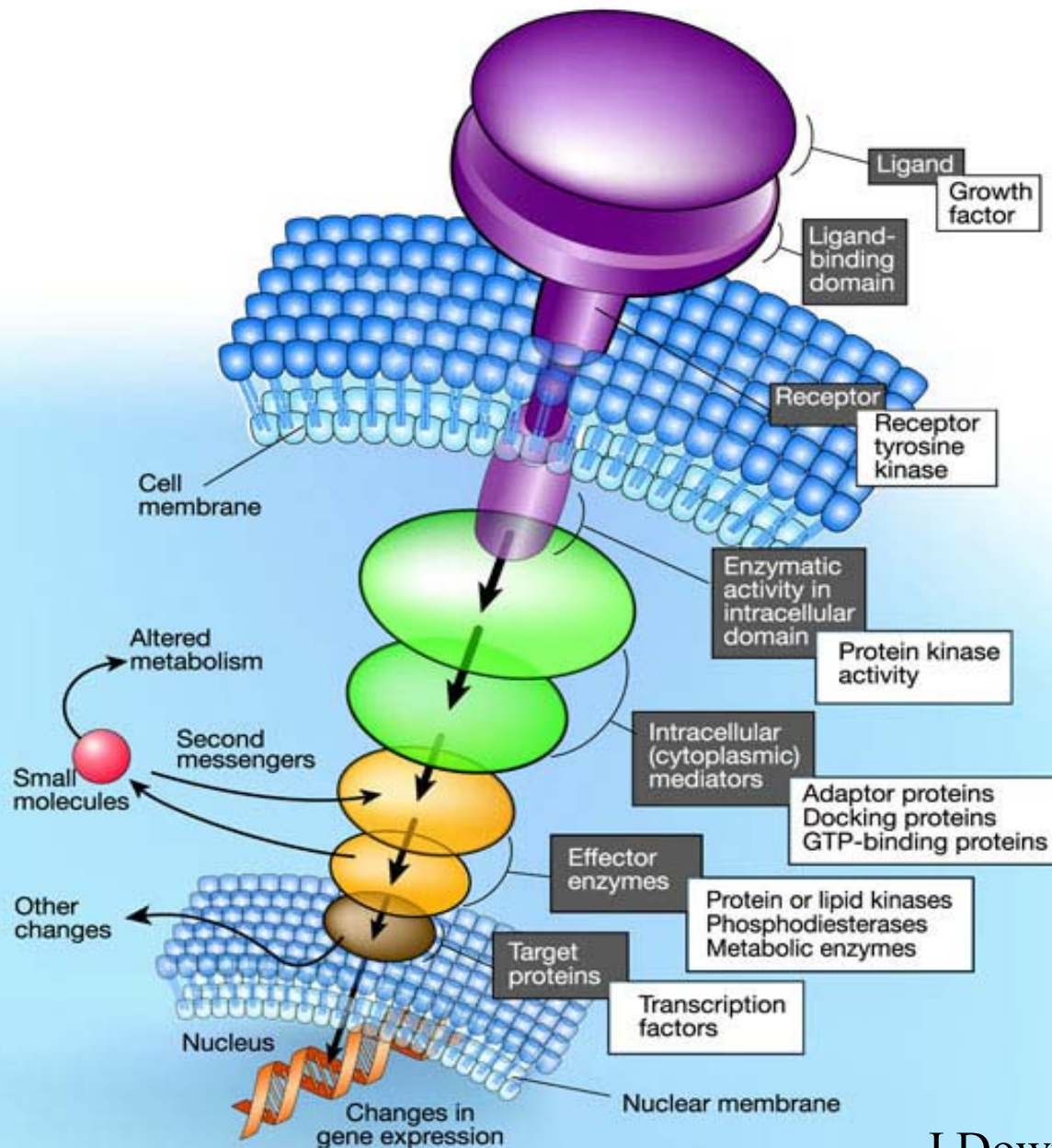


*Simulations combined with crystallography and mutagenesis are revealing important insights into how the normal prions undergo conformational changes that make them infectious.*

*Red & yellow are two regions that have been identified in the prion from sheep a joint project involving the National Institute of Medical Research and Univ of York, UK (Prof. Guy Dodson).*

# Applications: Cell Signaling

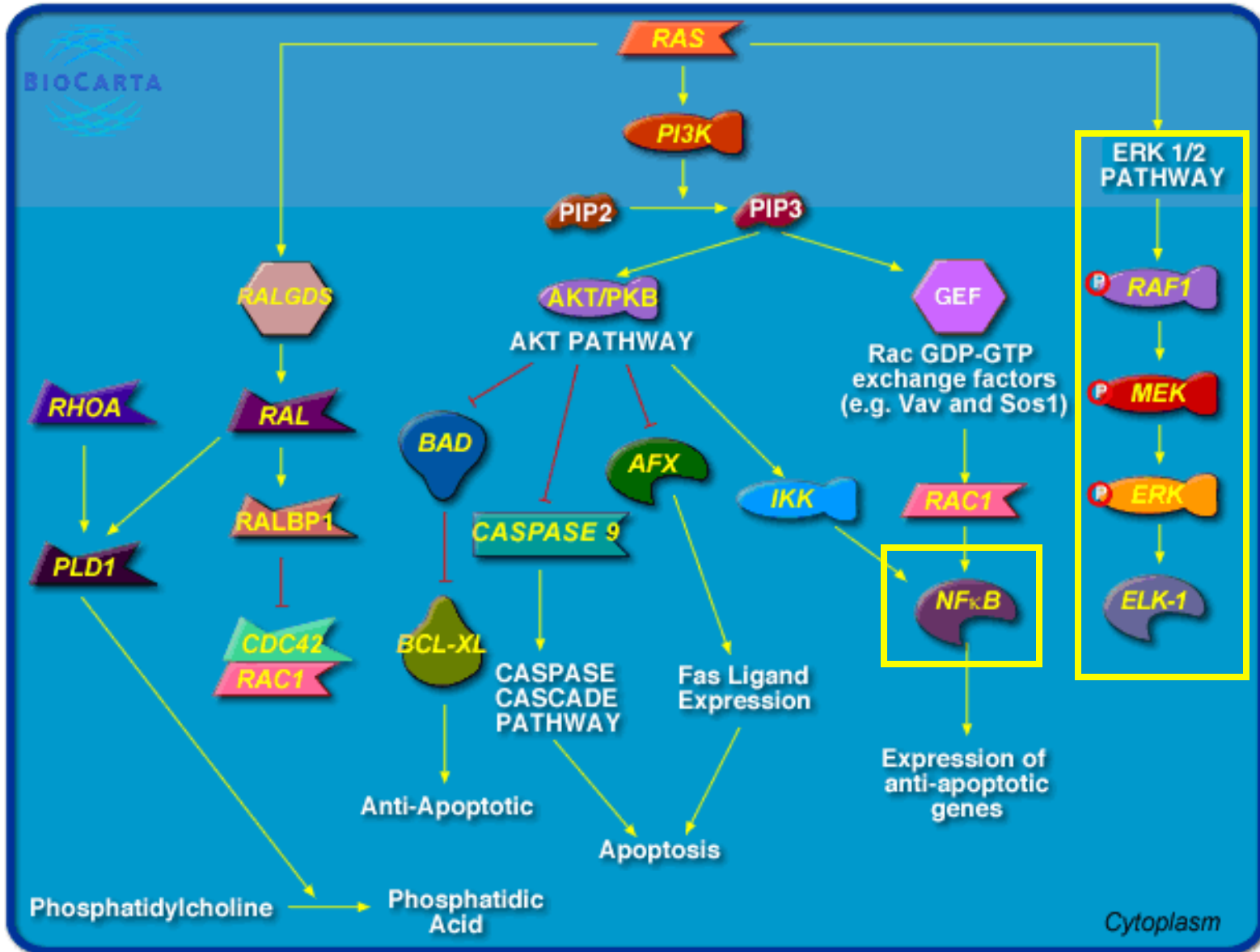
Work done by: **Dr. Guna Rajagopal & team.**



Modeling stochastic chemical kinetics on the IA64.

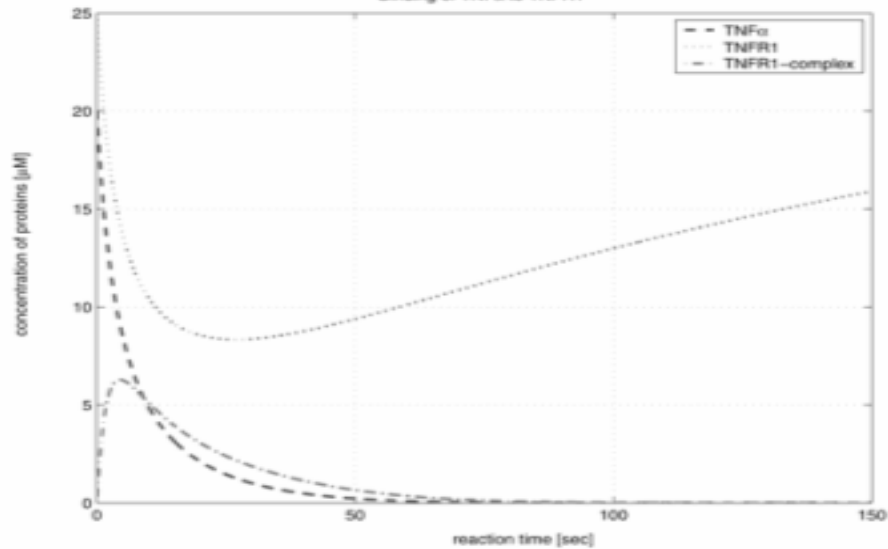
Goal: To model the flow of information from signal reception to biochemical response

# Ras Signaling Pathway

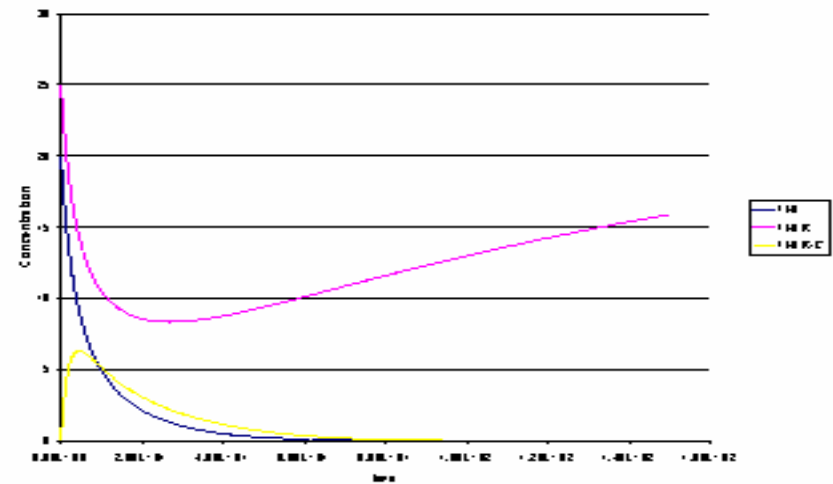


# TNF $\alpha$ mediated NF $\kappa$ B signaling

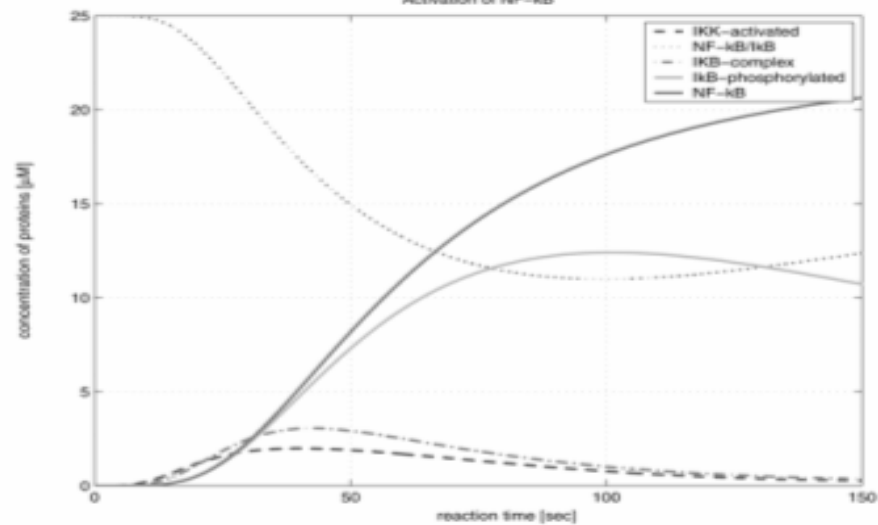
Binding of TNF $\alpha$  to TNFR1



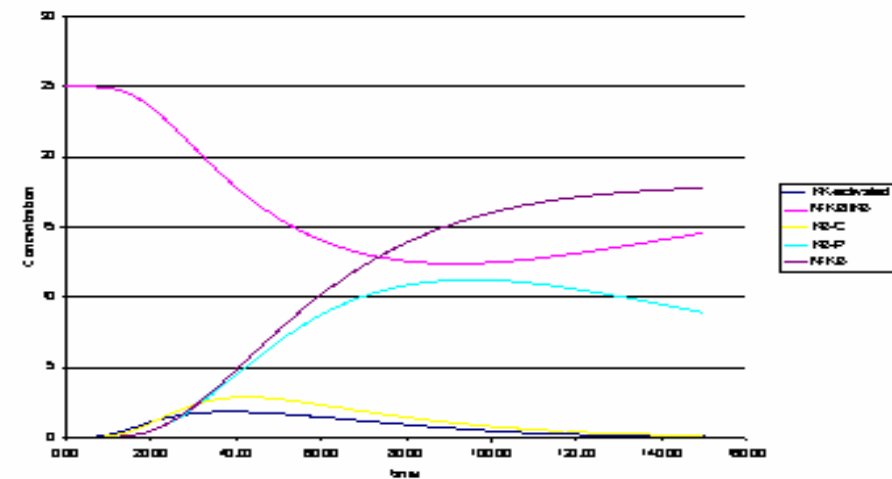
binding of I $\kappa$ B to I $\kappa$ B $\beta$



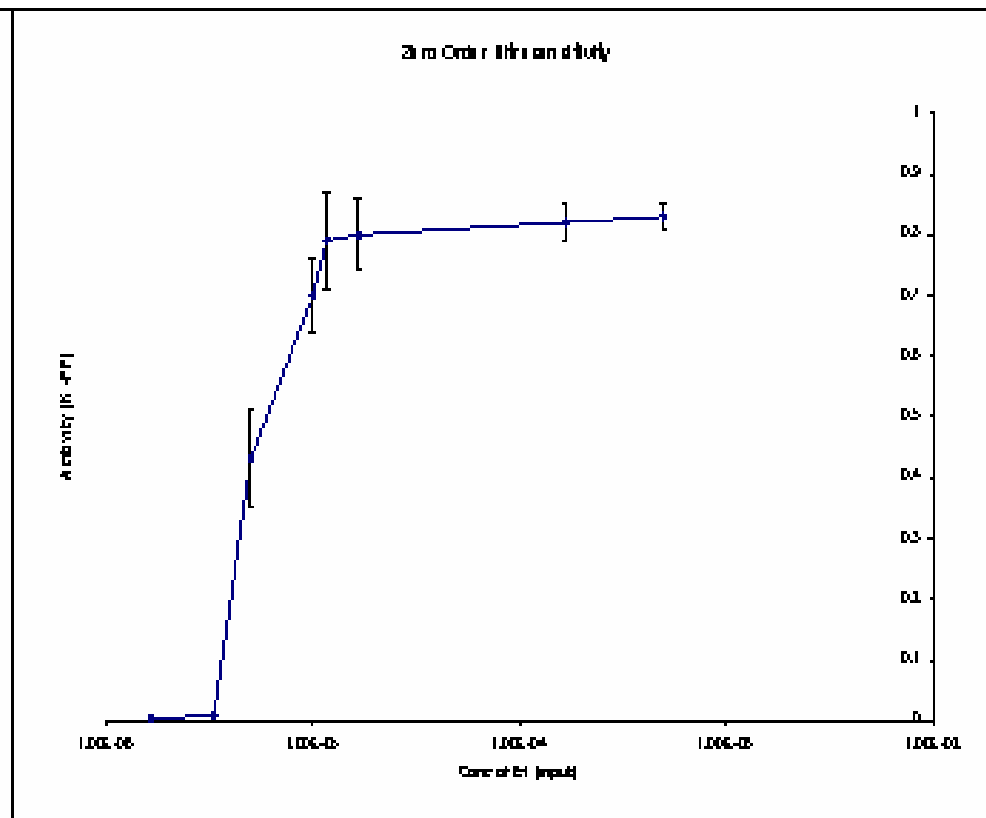
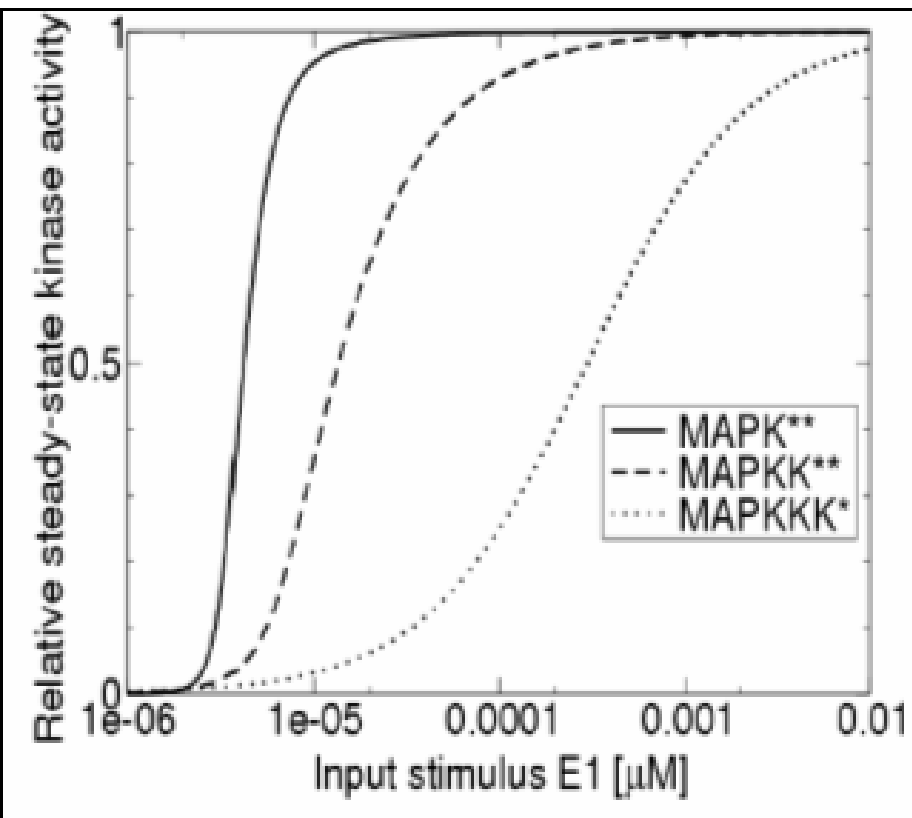
Activation of NF- $\kappa$ B



activation of NF $\kappa$ B



# MAPK Pathway - Zero Order Ultrasensitivity



To be presented at the Int Systems Biology Conf (ISBC) Nov 2004

# Systems Development Work on IA64

**Work done by:**

**Mr. Lai LF, Mr. Stephen Wong, Mr. Atif Shahab and  
their teams.**

- **Tested various compilers on IA64**
  - GCC
  - Intel's ECC
  - Open64 compiler
- **Tested Vtune on IA64**
  - NOTE: Patching the kernel is not always an option. Kernel version for Quardics patch and the kernel version required for Vtune had a conflict. This limited the range of testing that could be performed.
- **Studied qprof, PAPI, perfmon and oprofile.**

# Parallel Filesystems

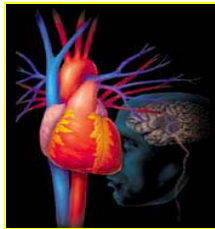
- Benchmarking and testing of various parallel filesystems on the Itanium platform
- Such filesystems include PVFS, OpenGFS, Lustre, etc

## Oracle RAC

- Benchmarking and testing of Oracle RAC on the Itanium platform
- Clustered filesystem is running off HP EVA 5000
- Comparing with MySQL Cluster

# The Road Ahead ...

Modeling the cardioelectrical activity with a 3D whole-heart electrophysiological



Cellware: The first global grid based tool for modeling and simulating cellular pathways

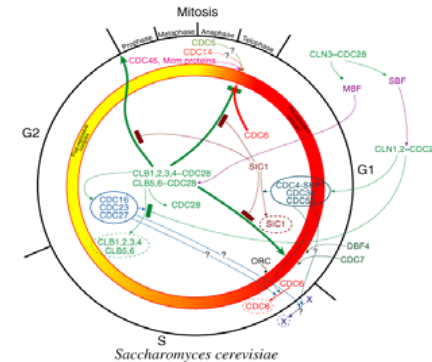
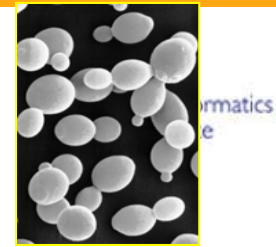
**Cellware**—a multi-algorithmic software for computational systems biology

Pawan Dhar\*, Tan Chee Meng, Sandeep Somani, Li Ye, Anand Sairam, Mandar Chitre, Zhu Hao and Kishore Sakharkar

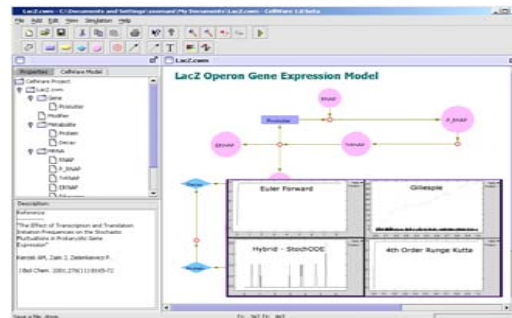
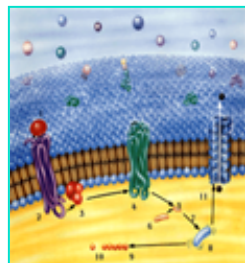
**Grid Cellware: The first Grid-enabled tool for modeling and simulating cellular processes**

Pawan Dhar, Tan Chee Meng, Sandeep Somani, Li Ye, Kishore Sakharkar, Arun Krishnan, Azmi B.M.Ridwan, Mandar Chitre, Zhu Hao

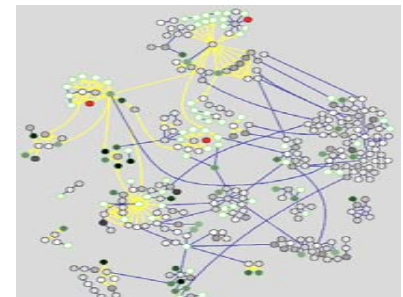
Discovering novel regulatory loops in cell cycle through modeling and reverse engineering



Modeling Signaling processes during development



Computational Modeling of metabolic networks



Dr. Pawan Dhar, Dr. Andrew Grovechey and team

# Neuroinformatics R&D roadmap

Prof. Novinski & team

