Computational Biology @ C-DAC

Technology Enablers

BRAF Features

- BRAF (Bioinformatics Resources and Applications Facility) is based on Supercomputing Clusters dedicated for life sciences and has aggregate capacity of 20 TF
- FPGA based inhouse developed RCS cards as accelerators
- High memory servers for memory intensive applications like NGS

ANVAYA - The Bio-Logical connection

- Anvaya is a workflows environment for high throughput comparative genomics
- It provides a set of 13 pre-defined workflows for frequently used pipeline in genome annotation and comparative genomics ranging from EST assembly and annotation to phylogenetic reconstruction and RNAseq analysis
- Easy to use Client supported on Windows/Linux which communicates with pre-installed Anvaya server at BRAF, C-DAC

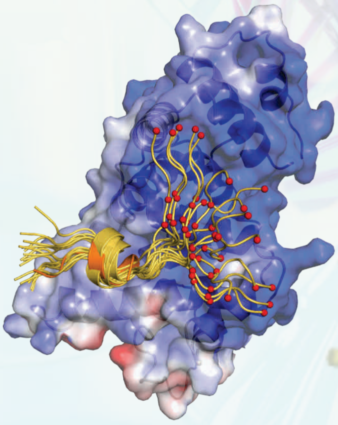
Cloud Computing

- Private Cloud virtualized resources with custom installed Bioinformatics tools
- Re-engineered bioinformatics workflows on map-reduce cloud framework
- HDFS support for data intensive distributed applications

Analytics Tools

- **Lambda:** Big Data Analytics Dashboard for Structural Biology
- **GAMUT:** Genomics BigData Management Tool
- **SUM:** Supercomputing User Management
- **TANGO:** High Throughput Conformation Generation and Optimization
- **ANVAYA:** A workflows environment for Automated Genome Analysis
- **DPICT:** A Parallel Visualization engine for Biomolecular Trajectories

Human Health



Cancer Research

- RNA-seq analysis to unravel the effect of pre-operative progesterone in operable breast cancer
- Identification and quantification of novel splice iso forms
- Obtaining digital gene expression counts reflective of transcript abundance

...in collaboration with Tata Memorial Hospital, Mumbai

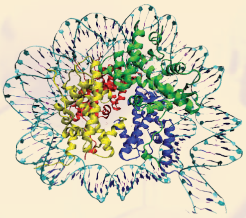
- Molecular simulation studies on nucleosome complex having histone variants
- Molecular Docking and simulation studies of anti-cancer metal complex

...in collaboration with ACTREC, Mumbai & Chemistry Dept., Pune University

- Drug repurposing studies on KRas protein
- Entire FDA approved drug database used for study
- Simulations for few of best docked complex were carried out

...in collaboration with National Cancer Institute (NCI), USA

In-House Research

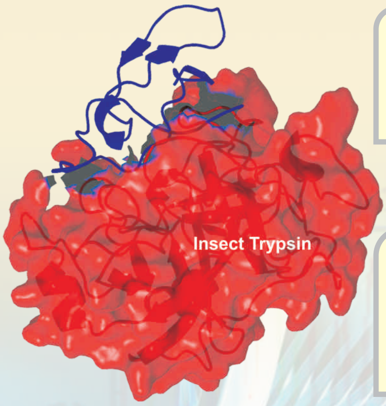


- Studies on binding affinity of p53 mutants with DNA RAS activation mechanism studies
- Study on development of new antisense modification for gene silencing

GPCR Research

- Homology modelling, docking and simulations of OA1 GPCR with IIT Madras studying aggregation and activation mechanism for β2AR

Plant / Microbial Genomics



Plant Inhibitor (IRD)

- Modeling network of gene responses to abiotic stress in rice
- Development of gene regulatory network analysis algorithms/tools under HPC environment using parallel computing

...in collaboration with Indian Agricultural Statistics Research Institute, Delhi

- Understanding the molecular interactions of plant (capsicum annum) proteinase inhibitors against insect (helicoverpa armigera) gut proteinase
- Patent: "Method for Effective Management of Helicoverpa Armigera" (Publication Number: WO/2013/102937)

...in collaboration with National Chemical Laboratory, Pune

Comparative genomics of Prokaryotic pathogens

- **Mycobacteria-**
  Comparative analysis of cell wall components of 21 Mycobacterium genomes list minimal functional gene set of cell wall components
- **Salmonella-**
  Identification of species-specific biomarkers among 16 Salmonella spp

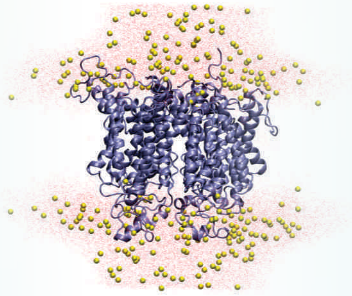
Animal Health

- Annotation of whole genomes of several Salmonella strains responsible for infection across various animal hosts
- Study of salmonella pathogenicity islands, biomarkers, role of horizontal gene transfer and metabolic networks
- Identifying SNPs specific for chicken lines, based on NGS data
- Analyzing differentially regulated genes during viral infection using RNA-Seq data
- Integratively analyze data to identify correlations between genomic regions, gene expression features, and further aspects such as gene function and phenotypic traits
- Building genome scale metabolic network model of Mycobacterium leprae
- Differential metabolite utilization of leprosy bacillus using producibility analysis

...in collaboration with Roslin Institute, Pirbright Institute, University of Surrey, UK

- Homology modelling of Ion channel protein
- Multiple simulations of Ion channel in membrane
- Docking of Multiple ligand molecules on ion channel

...in collaboration with Lupin Pharma Ltd, Pune



Salmonella
Unraveling Host Specificity

प्रगत संगणन विकास केंद्र
CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING
HPC: Medical and Bioinformatics Applications Group
सी-डैक इनोवेशन पार्क, स. न. 34/ब/1, पंचवटी, पाषाण, पुणे – 411008, भारत
C-DAC Innovation Park, S. No. 34/B/1, Panchavati, Pashan, Pune - 411008, India
Tel : 020-25503272 Email : braf-help@cdac.in

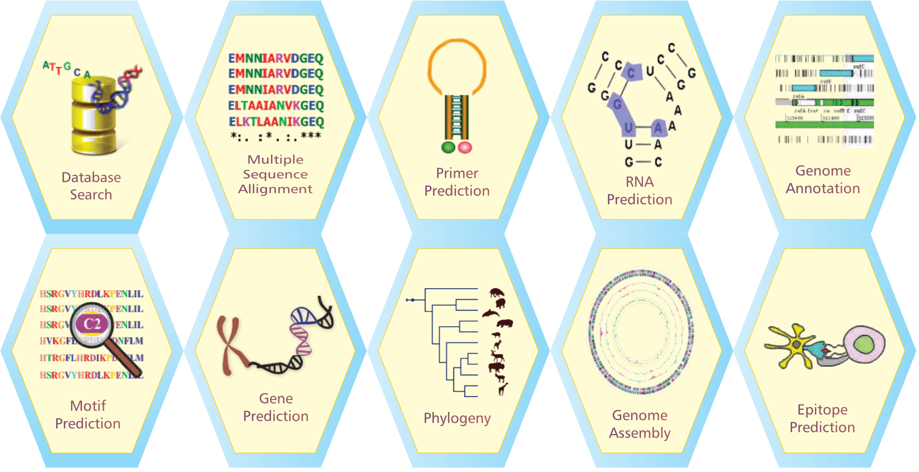
A Workflows Environment for Automated Genome Analysis

Anvaya

Anvaya is a software application, which provides a workflow environment for high throughput Comparative Genomics. It consists of an interface to Bioinformatics tools and databases, that are loosely coupled together in a coordinated system to execute a set of analyses tools in series or in parallel.

Need for Anvaya

- ⌘ High throughput sequencing technologies are generating large volumes of raw data
- ⌘ Manually running multiple tools for analysis is a big challenge
- ⌘ Anvaya enables handling of genomics data at optimum level with a very user friendly interface



Features

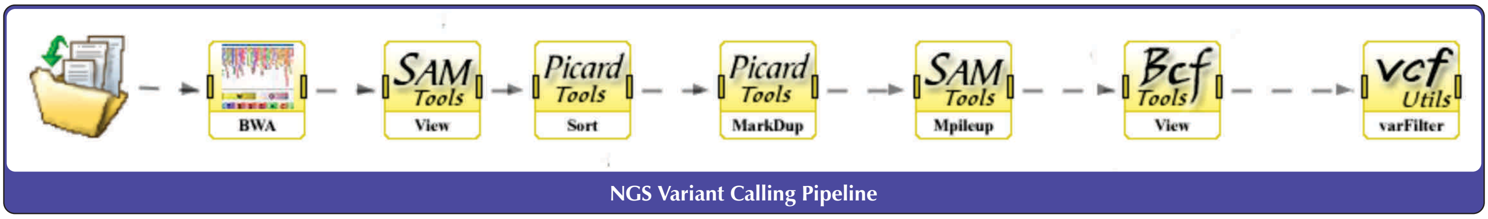
- ⌘ Powerful automated environments
- ⌘ Custom Tools to provide novel functionalities
- ⌘ Drag-drop facility to build custom workflows
- ⌘ Rules Engine which adds intelligence to control tools connectivity
- ⌘ Can handle multiple genomes / proteomes at a time
- ⌘ Can use a back-end parallel supercomputer for rapid execution

Anvaya Technology

Server: Java EE 7, JAX-RS, JWT, JPA2/Hibernate, Pbs/Torque Server, Wildfly Server, MySQL

Client: Java, Swings, Retrofit

NGS Pipeline



NGS Variant Calling Pipeline

Deployment

Deployed in Agriculture Research Institutes:

- National Agri-Food Biotechnology Institute (NABI), Mohali
- Indian Agricultural Statistics Research Institute (IASRI), New Delhi

प्रगत संगणन विकास केंद्र
CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING
HPC: Medical and Bioinformatics Applications Group
सी-डैक इनोवेशन पार्क, स. न. 34/ब/1, पंचवटी, पाषाण, पुणे – 411008, भारत
C-DAC Innovation Park, S. No. 34/B/1, Panchavati, Pashan, Pune - 411008, India
Tel : 020-25503272 Email : braf-help@cdac.in

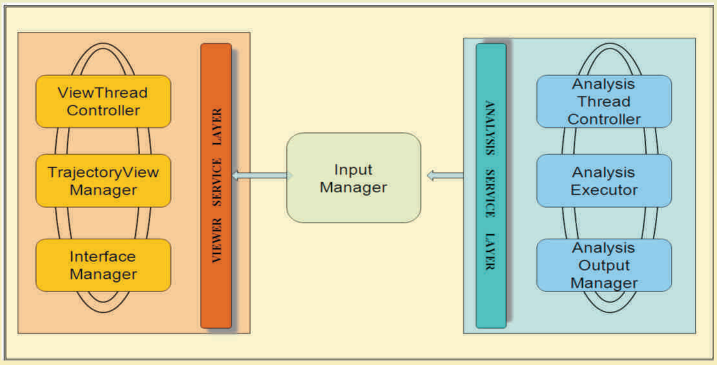
dPICT

A parallel visualization engine for bimolecular trajectories

Overview

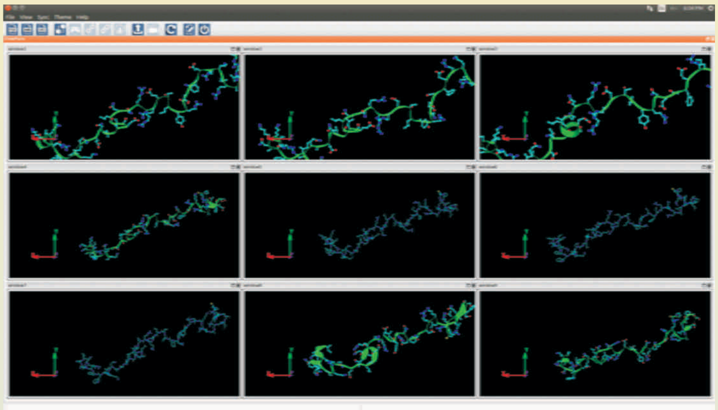
- A tool to elucidate the visualization of huge molecular dynamics trajectories simultaneously for better understanding of the simulation data
- Supports visualization of nine molecules simultaneously
- Different rendering options for biomolecules like ribbon, cartoon, ball and stick can be viewed
- Works in synchronous manner, where in nine trajectories may be handled simultaneously to perform certain operations
- Widely used file formats of PDB, AMBER and GROMACS are supported

Tool Architecture



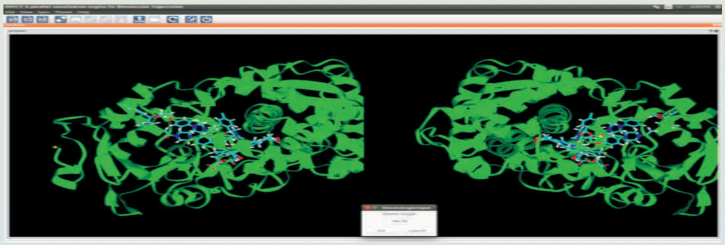
DPICT Features : Multiple windows

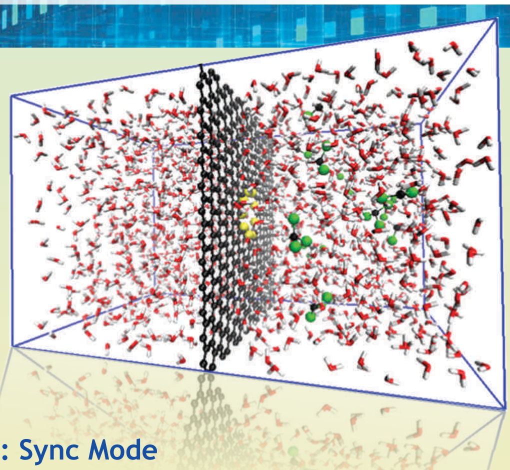
- Loading nine biomolecular structures for simultaneous visualization
- Use: Replica exchange and enhanced molecular dynamics techniques



DPICT Features : Stereo view

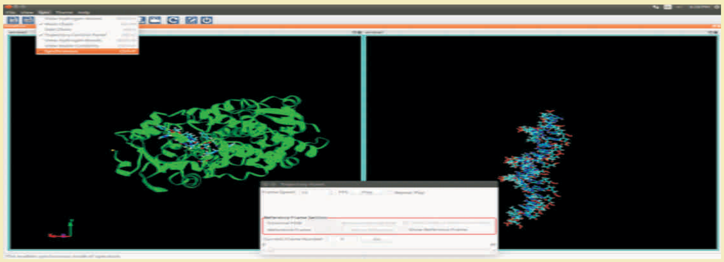
- Visualizing the molecule in stereo mode with an additional handle to alter the stereo angle
- Use: Viewing docked pose of ligand in drug-bound simulations





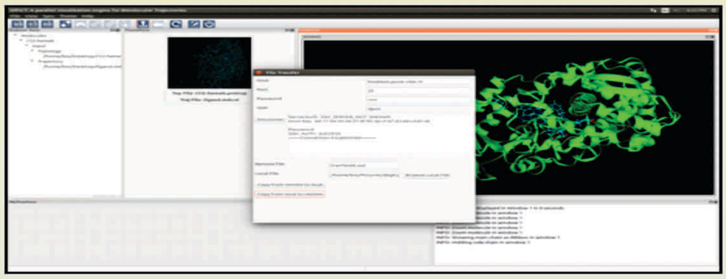
DPICT Features : Sync Mode

- Performing operations on different trajectories in the synchronous mode
- Use: Observe identical parameters in different trajectories in a single go



DPICT Features : File Transfer

- SSH feature enables the users to handle the transfer of large files from remote to local and vice-versa
- Use: Data transfer from remote HPC Clusters



Technologies Used

- DPICT has been implemented in C++
- It uses OpenGL based opensource code for rendering of structure
- User interface has been developed using WX Widgets
- Multi-threading has been incorporated to improve the performance of the tool

प्रगत संगणन विकास केंद्र
CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING
HPC: Medical and Bioinformatics Applications Group
सी-डैक इनोवेशन पार्क, स. न. 34/ब/1, पंचवटी, पाषाण, पुणे – 411008, भारत
C-DAC Innovation Park, S. No. 34/B/1, Panchavati, Pashan, Pune - 411008, India
Tel : 020-25503272 Email : braf-help@cdac.in