Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Web Server**

IWS: Integrated web server for protein sequence and structure analysis

Khader Shameer[1] and Ramanathan Sowdhamini[1]*

[1]National Centre for Biological Sciences (TIFR), GKVK Campus, Bellary Road, Bangalore 560 065, India;
Ramanathan Sowdhamini* - E-mail: mini@ncbs.res.in; * Corresponding author

received October 24, 2007; accepted October 29, 2007; published online November 01, 2007

Abstract:
Rapid increase in protein sequence information from genome sequencing projects demand the intervention of bioinformatics tools to recognize interesting gene-products and associated function. Often, multiple algorithms need to be employed to improve accuracy in predictions and several structure prediction algorithms are on the public domain. Here, we report the availability of an Integrated Web-server as a bioinformatics online package dedicated for *in-silico* analysis of protein sequence and structure data (IWS). IWS provides web interface to both in-house and widely accepted programs from major bioinformatics groups, organized as 10 different modules. IWS also provides interactive images for Analysis Work Flow, which will provide transparency to the user to carry out analysis by moving across modules seamlessly and to perform their predictions in a rapid manner.

Availability: IWS is available from the URL: http://caps.ncbs.res.in/iws

Key Words: protein sequence; structure analysis

Background:
Bioinformatics is now in a transition state – "from a data-centric science to knowledge based science", analysis and extraction of relevant information from huge amount of data from various high-throughput experiments remains as the major challenge in bioinformatics. Rapid increase in data generation has left us with many genes and proteins as 'unknown' or 'hypothetical' ones. As it is impossible to validate all the sequence data by means of biochemical experiments for confirmation of the likely associations, bioinformatics approaches can play an important role as a filter for recognizing potential gene products that can represent new fold or a novel function. Computational approaches enable the recognition of putative gene products of a family and to rationally design mutation experiments. Along with rapid incoming data, the availability of various resources to analyze the data has also increased. IWS is a compilation of in-house databases, web servers and web interface for various programs related to protein sequence and structure analyses clustered as ten modules. IWS is an easy-to-use web server, which will enable the novice as well as the expert users to carry out protein sequence and structure analysis rapidly and easily.

Implementation:
IWS provides various tools and database related to protein sequence and structure analysis classified into 10 different modules. Detailed information about the available modules, various tools, URL and its applications are given in Table 1 (supplementary material). IWS provides the tools and database under the following 10 different modules: Database and Servers, Sequence Retrieval and Search, Alignment, Sequence Analysis, Secondary Structure Prediction, Structure Analysis, Protein Modeling and Structure Validation, Sequence-Structure analysis, Phylogeny and Fold Recognition. Some of the major programs and databases available from IWS are PSI-BLAST **[1]**, CASCADE PSI-BLAST **[2]**, PHYLIP **[3]**, SEQPLOT, JOY **[4]**, MODIP **[5]**, SCANMOT **[6]**, MODELLER **[7]**, HARMONY **[8]**, PASS2 **[9]**, DSDBASE **[10]** etc. More than 40 bioinformatics resources for protein sequence and structure analysis is available from IWS. Figure 1 illustrates a flowchart that explains about different databases and tools available from IWS. IWS is running on a CentOS-Apache server. Front-end of IWS is developed using HTML, Perl script, CGI script, and Java scripts. Back-end is a combination of different programs developed using different languages like FORTRAN, C library (GD), C++, and Perl.

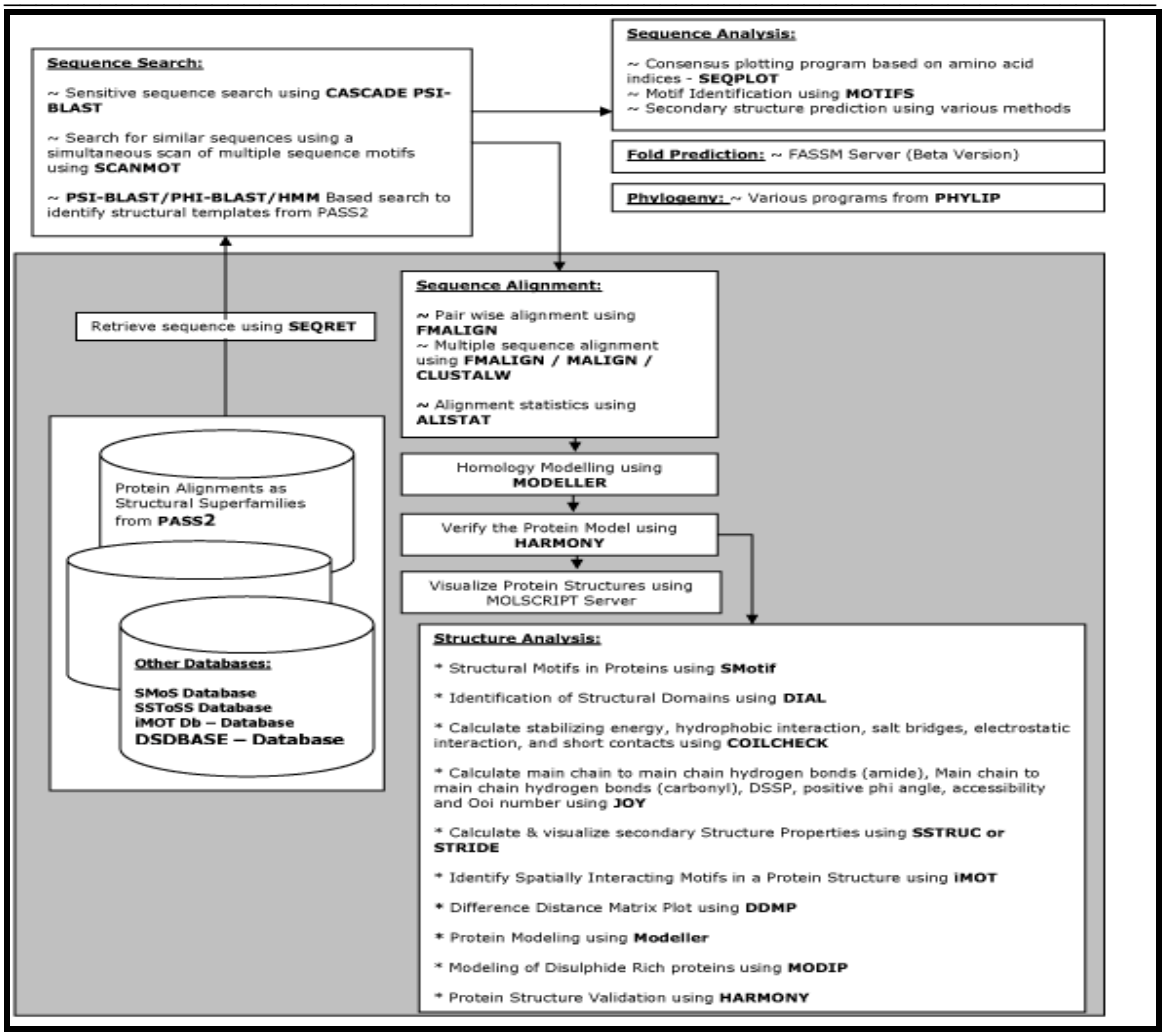

Figure 1: Schematic flow-chart about different databases and tools available from IWS

Input-output options:

IWS accepts sequences in FASTA, PIR, and Phylip format for protein sequence analysis and PDB file format for structure analysis. IWS generates different output for different programs. For instance, we have projected the results of the run on an 'unknown protein' sequence (from Drosophila ambigua (gi: 3676414) from NCBI Protein database) at the URL: http://caps.ncbs.res.in/iws/unknown/example1.html.

Caveats and future directions:

IWS can be accessible using any standard web-browsers (IE, Mozilla, Opera, Firefox etc.). We are planning to upgrade IWS version 2 as cluster-based server. We will continue to integrate new tools as they become available from the lab.

Conclusion:

In this article, we have explained about the availability of a new web server, NCBS-IWS an integrated web server for protein sequence and structure analysis. IWS will be a useful resource for research and academic communities interested in protein sequence and structure analysis.

Acknowledgement:

R.S. was a Senior Research Fellow of the Wellcome Trust; U.K. K.S. was briefly supported by the Wellcome Trust Grant. We thank Mr. G. Pugalenthi for useful discussions. We also thank NCBS (TIFR) for financial and infrastructural support.

References:

[01] S. F. Altschul, *et al.*, *Nucleic Acids Res.*, 25: 3389 (1997) [PMID: 9254694]

[02] S. Sandhya, *et al.*, *J Biomol Struct Dyn.*, 3:283 (2005) [PMID: 16218755]

[03] J. Felsenstein, *Annu Rev Genet.*, 22: 521 (1988) [PMID: 3071258]

[04] K. Mizuguchi, *et al.*, *Bioinformatics*, 14: 617 (1998) [PMID: 9730927]

[05] R. Sowdhamini, *et al*., *Prot. Engng.*, 3: 95 (1989) [PMID: 2594728]

[06] S. Chakrabarti, *et al*., *BMC Bioinformatics*, 5: 167 (2004) [PMID: 15509307]

[07] A. Sali, and T. L. Blundell, *J Mol Biol.*, 234: 779 (1993) [PMID: 8254673]

[08] G. Pugalenthi, *et al*., *Nucleic Acids Res*., 34: W231 (2006) [PMID: 16844999]

[09] A. Bhaduri, *et al*., *BMC Bioinformatics*, 5: 35 (2004) [PMID: 15059245]

[10] A. Vinayagam, *et al*., *Nucleic Acids Res*., 32: D200 (2004) [PMID: 14681394]

Edited by P. Kangueane

Citation: Shameer & Sowdhamini, Bioinformation 2(3): 86-90 (2007)

Supplementary material:

No	Tool/Database	URL	Application
Module 1 : Database & Server			
1	Database & Server	http://caps.ncbs.res.in/iws/dbs.html	List of selected database and web servers related to protein sequence and structure analysis.
Module 2: Sequence Search			
2	CASCADE PSI-BLAST	http://caps.ncbs.res.in/iws/cascade.html	PSI-BLAST for many 'generations', initiating searches from new homologues
3	SCANMOT	http://caps.ncbs.res.in/scanmot/scanmot.html	Search for similar sequences using a simultaneous scan of multiple sequence motifs
4	PSI-BLAST Search PASS2 Database	http://caps.ncbs.res.in/iws/psib_pass2.html	PSI-BLAST Search using PASS2 Database
5	PHI-BLAST Search against PASS2 Database	http://caps.ncbs.res.in/iws/phib_pass2.html	PHI-BLAST Search using PASS2 Database
6	HMM Search against PASS2	http://caps.ncbs.res.in/iws/hmm_pass2.html	HMM Search against PASS2 Database to identify Structural Templates
7	SEQRET	http://caps.ncbs.res.in/iws/seqret.html	Retrieve Protein Sequence from Swiss-Prot/NR/TrEMBL using user Query / Accession ID
Module 3: Alignment			
8	FMALIGN	http://caps.ncbs.res.in/iws/FMAmulti.html	Fixed Motif ALIGNment [FMALIGN] – Pair-wise & Multiple Alignment
9	MALIGN	http://caps.ncbs.res.in/iws/malign.html	Multiple Sequence Alignment using MALIGN
10	MALIGN Search against PASS2	http://caps.ncbs.res.in/iws/malign_pass2.html	MALIGN Search against PASS2 Database to identify Structural templates for protein modeling
11	ALISTAT	http://caps.ncbs.res.in/iws/alistat_ali.html	Alignment Statistics using ALISTAT
12	iMOT from Alignments	http://caps.ncbs.res.in/iws/imot_ali.html	Find iMOT(Interacting MOTifs) from an alignment using iMOT Algorithm
Module 4: Sequence Analysis			
13	SEQPLOT	http://caps.ncbs.res.in/iws/seqplot.html	Consensus Sequence plotting program based on various amino acid indices
14	DIAL Sequence	http://caps.ncbs.res.in/iws/dial_seq.html	Identify Structural Domains from Protein Sequence using DIAL (Domain Identification ALgorithm)
15	iMOT Conserved Regions	http://caps.ncbs.res.in/iws/imot_cr.html	Identify Sequentially conserved regions in Proteins sequence using iMOT Algorithm
16	iMOT Spatially Interacting Motifs	http://caps.ncbs.res.in/iws/imot_seq.html	Identify Spatially Interacting Motifs in Proteins sequence using iMOT Algorithm
17	MOTIFS	http://caps.ncbs.res.in/iws/motifs.html	Identify sequence motifs from Protein Sequence using MOTIFS program
18	SEQREPORT	http://caps.ncbs.res.in/iws/seqreport.php	Amino acid composition, Molecular weight, Molar absorption coefficient, Protein iso-electric point with pK values, Charge at different pH, sequence as 3 letters amino acid code and Classification of amino acid residues
19	Secondary Structure Analysis	http://caps.ncbs.res.in/iws/secstr.html	Access Various Secondary Structure Servers through a single Window of NCBS-IWS
Module 5: Phylogeny			
20	PROTPARS	http://caps.ncbs.res.in/iws/protpars.html	PROTPARS - Estimate phylogenies from protein sequences (PHYLIP)
21	PROTDIST	http://caps.ncbs.res.in/iws/protdist.html	PROTDIST - Computes a distance measure for protein sequences (PHYLIP)
22	PROML	http://caps.ncbs.res.in/iws/proml.html	PROML - Estimates phylogenies from protein amino acid sequences by maximum likelihood (PHYLIP)
23	PROMLK	http://caps.ncbs.res.in/iws/promlk.html	PROMLK - Same as PROML but assumes a molecular clock (PHYLIP)
24	SEQBOOT	http://caps.ncbs.res.in/iws/seqboot.html	SEQBOOT - Reads in a data set, and produces multiple data sets from it by bootstrap re-sampling

			(PHYLIP)
25	CONSENSE	http://caps.ncbs.res.in/iws/consense.html	CONSENSE - Computes consensus trees by the majority-rule consensus tree method (PHYLIP)
26	NEIGHBOR	http://caps.ncbs.res.in/iws/neighbor.html	NEIGHBOR - Neighbor-Joining method of Saitou and Nei (1987) and the UPGMA method of clustering (PHYLIP)
27	KITSCH	http://caps.ncbs.res.in/iws/kitsch.html	KITSCH - Fitch-Margoliash and Least Squares Methods with Evolutionary Clock (PHYLIP)
Module 6: Secondary Structure Prediction			
28	Integrated web-interface for Secondary Structure Prediction	http://caps.ncbs.res.in/iws/secstr.html	Jpred, BetaTurnv1.1 - Prediction of Beta-turn By Support Vector Machine, SPLIT - Membrane Protein Secondary Structure Prediction, HMMTOP - Transmembrane Helices & Topology, Sosui - Classification and Secondary Structure Prediction of Membrane Proteins
Module 7: Sequence-Structure Analysis			
29	SSTOSS Database	http://caps.ncbs.res.in/SSTOSS/passlist.htm	SSToSS : A database of Sequence Structural Templates of Single member Superfamilies
30	JOY	http://caps.ncbs.res.in/iws/joy_tem.html	Calculate main-chain to main-chain hydrogen bonds (amide), Main-chain to main-chain hydrogen bonds (carbonyl), DSSP, positive phi angle, accessibility and Ooi number using JOY Program
Module 8: Structure Analysis			
31	SMotif	http://caps.ncbs.res.in/SMotif	SMotif - Structural Motifs in Proteins
32	DIAL	http://caps.ncbs.res.in/iws/dial_struc.html	Identify Structural Domains using DIAL (Domain Identification ALgorithm)
33	COILCHECK	http://caps.ncbs.res.in/iws/coilcheck.html	Calculate Stabilizing Energy, Hydrophobic Interaction, Salt Bridges, Electrostatic Interaction, and Short Contacts using COILCHECK
34	STRIDE	http://caps.ncbs.res.in/iws/stride.html	Calculate & Visualise Secondory Structure Properties using STRIDE
35	SSTRUC	http://caps.ncbs.res.in/iws/sstruc.html	Secondory Structural Patterns using SSTRUC
36	HBOND	http://caps.ncbs.res.in/iws/hbond.html	Hydrogen bonding Properties using HBOND
37	C-Alpha Calculation	http://caps.ncbs.res.in/iws/cadistance.html	Calculate C-alpha distance between two atoms of a PDB file
38	atm2seq	http://caps.ncbs.res.in/iws/atm2ali.html	Extract Sequence Information from a PDB file using atm2seq
39	JOY Structural features	http://caps.ncbs.res.in/iws/joy_tem.html	Generate JOY Structural features output for a PDB File
40	iMOTdb	http://caps.ncbs.res.in/imotdb	Database of Spatially Interacting Motifs in Proteins
41	DDMP	http://caps.ncbs.res.in/iws/ddmatrix.html	Difference Distance Matrix using DDMP
42	SMoS Database	http://caps.ncbs.res.in/SMoS	Structural Motifs of Superfamilies Database - provides information on structural motifs or templates of aligned protein domain superfamilies like PASS2 and CAMPASS
43	MOLSCRIPT	http://caps.ncbs.res.in/iws/molscript.html	Generate MOLSCRIPT Images of Macromolecules
Module 9: Fold Prediction			
44	FASSM Server (Beta-Version)	http://caps.ncbs.res.in/iws/fassm.html	Neural Network based FASSM (Enhanced Function Association in whole genome analysis using Sequence and Structural Motifs.) Server for fold prediction
Module 10: Protein Modeling & Structure Validation			
45	MODIP	http://caps.ncbs.res.in/iws/modip.html	MODIP(Modeling Of DIsulphide bonds in Proteins)
46	DSDBASE Search	http://caps.ncbs.res.in/dsdbase//accessmts.html	Search for Disulphide rich proteins using DSDBASE Search Tool
47	MODELLER* (License Required)	http://caps.ncbs.res.in/iws/protmod.html	Build Proteins using MODELLER
48	HARMONY	http://caps.ncbs.res.in/harmony	Validate Protein Models using HARMONY

Table 1: Detailed list of Tools with name, URL and application grouped in to 10 different modules