Research article

Open Access

PURE: A webserver for the prediction of domains in unassigned regions in proteins

Chilamakuri CS Reddy^{1,2}, Khader Shameer¹, Bernard O Offmann^{*2} and Ramanathan Sowdhamini^{*1}

Address: ¹National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bellary Road, Bangalore 560 065, India and ²Laboratoire de Biochimie et Genetique Moleculaire, Universite de La Reunion, 15 avenue Rene Cassin, BP 7151, 97715 Saint Denis Messag Cedex 09, La Reunion, France

 $\label{eq:email: Chilamakuri CS Reddy - csreddy@ncbs.res.in; Khader Shameer - shameer@ncbs.res.in; Bernard O Offmann* - bernard.offmann@univreunion.fr; Ramanathan Sowdhamini* - mini@ncbs.res.in = for the shame of the shame o$

Received: 5 February 2008 Accepted: 14 June 2008

* Corresponding authors

Published: 14 June 2008

BMC Bioinformatics 2008, 9:281 doi:10.1186/1471-2105-9-281

This article is available from: http://www.biomedcentral.com/1471-2105/9/281

© 2008 Reddy et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<u>http://creativecommons.org/licenses/by/2.0</u>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Protein domains are the structural and functional units of proteins. The ability to parse proteins into different domains is important for effective classification, understanding of protein structure, function, and evolution and is hence biologically relevant. Several computational methods are available to identify domains in the sequence. Domain finding algorithms often employ stringent thresholds to recognize sequence domains. Identification of additional domains can be tedious involving intense computation and manual intervention but can lead to better understanding of overall biological function. In this context, the problem of identifying new domains in the unassigned regions of a protein sequence assumes a crucial importance.

Results: We had earlier demonstrated that accumulation of domain information of sequence homologues can substantially aid prediction of new domains. In this paper, we propose a computationally intensive, multi-step bioinformatics protocol as a web server named as **PURE** (**P**rediction of **U**nassigned **RE**gions in proteins) for the detailed examination of stretches of unassigned regions in proteins. Query sequence is processed using different automated filtering steps based on length, presence of coiled-coil regions, transmembrane regions, homologous sequences and percentage of secondary structure content. Later, the filtered sequence segments and their sequence homologues are fed to PSI-BLAST, cd-hit and Hmmpfam. Data from the various programs are integrated and information regarding the probable domains predicted from the sequence is reported.

Conclusion: We have implemented PURE protocol as a web server for rapid and comprehensive analysis of unassigned regions in the proteins. This server integrates data from different programs and provides information about the domains encoded in the unassigned regions.

Background

Protein domains are the structural and functional units of proteins and represent one of the most useful levels to understand protein function. Analysis of proteins at the level of domain families has had a profound impact on the study of individual proteins. Protein domain discovery using various computational approaches has been progressing steadily over the past 35 years [1]. They can be defined using multiple criteria, or combinations of criteria, including evolutionary conservation, discrete functionality, and the ability to fold independently [2]. The identification of protein domains within a polypeptide chain can be achieved in several ways. Methods applied by classification databases such as the Dali domain dictionary [3], CATH [4], SCOP [5], DIAL [6] employ structural data to locate and assign domains. Identification of domains at the sequence level most often relies on the detection of global and local sequence alignments between a given target sequence and domain sequences found in databases such as Pfam [7]. However, difficulties in elucidating the domain content of a given sequence arise when searching the target sequence against sequence databases resulting in a lack of significant matches. For example, Mycoplasma genitalium is a small genome with 483 proteins but only 386 protein sequences have known Pfam hits with 56% residue coverage. In such situations, there is a need to further explore other methods for domain assignment from sequence. Though, similar approaches of integrating multiple, sensitive database searches to detect distant homologues has been reported as a successful method to establish remote homology [8], we have recently shown that it is possible to enhance prediction of domains by 25% through indirect connections, namely consulting the domain architecture of sequence homologues [9]. In this paper, we report the availability of a bioinformatics protocol, as a web server called PURE (Prediction of unassigned regions in proteins), which will enhance the domain predictions. PURE protocol utilizes the concept of intermediate sequence search (ISS) [10] to assign functional domain to a given unassigned region with the help of connecting sequences. Indirect connections between the query and distantly related domain is established through a powerful procedure using PSI-BLAST hits which are individually routed through a rigorous Hmmpfam search against Pfam database [11].

Implementation

PURE server: description and features

PURE server processes the query sequence in FASTA format using a computationally intensive bioinformatics protocol. A detailed flow-chart of the protocol is provided in Figure 1. The web interface is developed in HTML and Java script. The wrapper scripts for external programs like Globplot [12], Disopred2 [13] Pepcoils [14], Tmap [14], psipred [15], Scanprosite [16], PSI-BLAST [17], cd-hit [18] and Hmmpfam [19], while core programs that integrate results, CGI programs and automated e-mail programs are coded in Perl. Multiple alignment visualization of PSI-BLAST output is enabled using Mview [20]. The graphics that integrate the results are generated using Bio::graphics module from Bioperl [21].

Input options

PURE server is designed to accept a single sequence in FASTA format at a time (Figure 2). Users can submit the sequence as a file using the upload option or copy paste option. Users should also submit a valid, non-commercial email address to the server. PURE server will send the result URL to the email address. Options are provided to select E-value for the PSI-BLAST search and Hmmpfam [16] search against various sequence databases like NCBI-NR [22] and SwissProt [23]. Options are also provided to change the clustered sequence space obtained from PSI-BLAST search using different threshold value for cd-hit. User can also choose the filtering option.

Output details

PURE Server examines unassigned regions for the presence of disordered regions, coiled coils, transmembrane helices, appropriate extent of predicted secondary structural content and presence of homologous sequences before the assignment of probable structural domains. These are also provided as links to the URL where the output is stored in. Output can be mainly divided into two:

A. Consensus output: provides a summary of the overall results (Figure 3). Consensus result page provides a complete overview of the PURE results for the query sequence. The First table in the Consensus result provides the details about the domain assigned, if any, using PURE protocols. This table provides the name of the domain associated to the unassigned region, domain frequency and the direct link to the Pfam database [7] of the identified domain. The table is followed by a graphical representation of PURE Results indicating domain boundaries and the region of similarity between the unassigned region and the associated Pfam domains. This figure is dynamically generated after processing complete results from PURE run for the query sequence. Further, the Bio::graphics [19] based picture provides a complete overview of the query sequence (query), transmembrane regions (Tmap), coiled coil regions (Pepcoil), extracted sequence segments (sequence_segments) after removal of transmembrane and coiled coil regions from the query sequence and the domains assigned by PURE. Apart from the above mentioned information, consensus results page also provides a link to the detailed results page of the PURE results.

B. Detailed output: This output is divided into 10 sections for a successful PURE run. Each of the files provided in detailed output gives the background details about the final result file (Figure 4).

1. Disorder prediction: This section shows the predicted disordered regions in the query sequence. We have employed Globplot [12] and Disopred2 [13] programs





PREDICTION OF UNASSIGNED REGIONS IN PROTEINS					
Input Method 1 - Upload Sequence: Enter your valid email:	Input Method 2 - Paste Sequence of Unassigned Regions in FASTA format: Enter your valid email: Enter Query Name of UR: Enter Amino Acid Sequence Here: [Click here for Example Run]				
PSI-BLAST Database: PSI-BLAST E-Value: 0.001 CD-HIT Threshold Value: 0.4 HMMPFAM E-Value: 0.01 HMMSEARCH Cut-off: 50	Switch-off TMAP/COIL Filter:				
Figure 2	CD-HIT Threshold Value: 0.4 v HMMPFAM E-Value: 0.01 v HMMSEARCH Cut-off: 50 v PURE Clear Form				
Screenshot of input options from PURE server.					

PURE

for the prediction of disordered regions in the protein sequence.

2. Pepcoils results: This file is generated for the prediction of coiled coils by Pepcoils (EMBOSS) [14] program that works around COILS [24] algorithm. Details about the coiled coils identified from the query sequence are available in this file. Another file with the query sequence parsed for coiled coil regions (such regions are substituted with '=') is also provided.

3. Tmap results: This file is generated to record probable transmembrane helices in the query sequence as identified by Tmap (EMBOSS) [14] program. Another file with transmembrane regions identified from query sequence substituted with 'x' is also provided for better understanding of the presence of transmembrane regions.

4. Integrated filter results: Both COILS and Tmap files are processed using a Perl program. This program integrates the query sequence to provide a modified query sequence

file, with coiled coils and transmembrane regions substituted with '=' and 'x' respectively. By default, such regions are not considered for further analysis. Further, the sequence is split into segments based on the presence of transmembrane regions and coiled coil regions. Only sequence segments having ≥ 30 residues are considered for further analysis, in order to avoid spurious hits in subsequent PSI-BLAST jobs.

5. Examine the sequence segments extracted from query sequence after filtering using Pepcoils and Tmap. Sequence is split into different fragments based on the presence of coiled coil and transmembrane regions, each fragment should have at least 30 residues to be considered for further analysis.

6. Psipred results: We have used psipred [15] program for secondary structure prediction, sequence segments with \geq 15% secondary structural content considered for further analysis.

PURE

PREDICTION OF UNASSIGNED REGIONS IN PROTEINS

Home	Server	Неђ	Group	
PURE Server : Cor	ncise Results fo	or O8EXW3.903-97	9	
		a Qolar (150505) (
Indirect Domain A	ssociations of I	Unassigned Region	n Domain Frequency	rcy Pfam Link
TPR_1			3/12	TPR 1
TPR_2			9/12	<u>TPR 2</u>
Carabian Damasa		E Dlt-		
	tation of PUR			
Query			400	500 600
Score =0	Que	ry Sequence		
Sequence_Segment_1	—— Sequ	uence Segme	nt extracted aft	fter filtering steps
PD:TPR_2 CS:ZP_00107	619	CS : Connect	ting Sequence	e ZP_00107619 [NCBI - Accession ID]
Score =0.0042		PD : Pfam D	omain : TPR_2	_2 is the predicted domain using the connecting sequence ZP_00107619
PD:TPR_2 CS:NP_76897	'8			
Score =0.00041 PD:TPR 11CS:ZP 01109	234			
Score =0.008	201			
PD:TPR_2 CS:ZP_01109	234			
Score =0.00023				
PU: IPR_2[CS:2P_0116:			omains [PD] p	predicted from the Query sequence using different Connecting Sequency
PD:TPR_2 CS:NP_97154	12			,
Score =0.0088				
PD:TPR_1 CS:NP_7146	52			
Score =0.0013 PD-TPD 210S-NP 7146	52			
Score =4.8e-05				
PD:TPR_1 CS:YP_6600	78			
Score =0.0062				
PD:TPR_2ICS:YP_6600	78			
PD:TPR_2[CS:YP_797	390			
Score =3.3e-05				
PD:TPR_2 CS:XP_001	437139			
Score =0.0075				
CI-1-1	3.1	DIDEC	OFFE	
Click here for the det	ailed results fron	n FUKE Server for y	our query : Q8EXW3	13.303-373

Figure 3 Screenshot of concise output from PURE server.

7. PSI-BLAST result file: Sequence segments which passed the filtering criteria above are fed to PSI-BLAST [17] for similarity search. Mview based visualization of PSI-BLAST results are also provided for better insight into the PSI-BLAST results.

8. Scanprosite results: Prosite [16] scan results are supplied as a supplement to our method. Scanprosite scans query sequence for the occurrence of patterns, profiles and rules (motifs) stored in the Prosite database [25].

9. Hmmpfam Results: The results obtained by running Hmmpfam on homologous sequences and the assignment of Pfam domains for homologues are provided here. Only representative sequences, as recognized after clustering by cd-hit [18] are considered for the search for domains.

10. Domains assigned using PURE: This is the major output file that integrates the PSI-BLAST and Hmmpfam search output and provide the probable domains identified using PURE protocol for the unassigned regions in

the query. Bio::graphics [21] based image is provided for the better overview of the detailed result.

11. Concluding remarks: The concluding mark is derived at the end of each run after analyzing all the files that are generated. Various output possibilities and concluding remarks are discussed in Table 1.

Results and Discussion

In cases where the PURE server is not effective with default parameters, the user can manoeuvre the search directions by changing certain parameters. However, the minimum length of 30 residues for an unassigned region (UR) is not recommended to go lower, in order to avoid sub-optimal alignments and false positives. By default, unassigned region will be split whenever transmembrane helices (TM) or coiled coils regions are encountered; however, there is a possibility to override this option. Likewise, the E-value thresholds of PSI-BLAST runs and cd-hit thresholds for representative homologous sequences can be modified according to the choice of the user. It is also possible to increase the number of homologues identified

PURE					
PREDICTION O	of Unassigned	REGIONS I	N PROTEINS		

HRE : Revelor for QUEXNO 3.992 for 1 kik Inter Constant attack for an aner. HER : Dockal of QUEXNO 3.992 for Query attack in the second of Probine Second	Home	Server 1	Help	Сгоцр	
kk. km fat constant sets for your and. KRE - Detailed Ought File: High Anabris Resetts Detailed The VRE - WRE - WRE - WRE - WRE AND Excepted for Disorder Predicts Sets Form any encode the andreer regions may	PURE : Results for (8EXW3_903-979.fas	sta		
Balance Instrument and and a langer and and a langer and langer and a langer and a langer and a	Clinte Lang Conservation				
DHRE - Detailed Ougen File: High Analysis (Desk level is advance regions on garge pageness and retry lie FUSE real Scheme for GOBINED COUNTY - 1 FUECUSA state (Eds. Under Scheme County - 1 FUECUSA state (Eds. Under FUECUSA state (Eds. Under F	Click here for consensu	3 output for your query.	2		
<pre>thield Advance Network Second Public Advance Adva</pre>	PURE - Detailed Out	put Files:			
Description of Production PUREE in using GabAPD and DisposeD 200 Davids PURDE read Schemet and Construction of Confere Conference in the State PEPCORE Science in Science in Conference in Science in Science in Science in Conference in Science i	Detailed Analysis Re	<u>sults:</u>			
bar : Der ange rennene hen dannalt-regional from ganz regenere and recey the FUEE res How here discussed in a Cold Cold Regional Result: FCOLDS so que: (Cold Region MERCOLLS COTTON 1) FCOLDS so que: (Cold Region MERCOLLS COTTON 1) FCOLDS so que: (Cold Region HORIZON HINE) Marting of Transmontance Regions Result: Marting of Transmontance Regions Result: FCOLDS so que: (Cold Region HORIZON HINE) Marting of Transmontance Regions Result: FCOLDS so que: (Cold Cold Cold Region HORIZON HINE) Marting of Transmontance Regions Result: FCOLDS so que: (Cold Cold Cold Cold Region HORIZON HINE) FCOLDS to the Cold Cold Cold Cold Regions HORIZON HINE (FCOLD & TMAP) FCOLDS to the Cold Cold Cold Cold Regions HORIZON HINE HINE HINE HINE HINE HINE HINE HIN	1. Disorder Prediction	- PURE is using Glob	Plot and Disopred2	for Disorder Pr	ediction
Scheme Sc	Note : User may remo	ve the disorder region	ns from query sequ	ence and retry	he PURE run
Part Instrumental Procession of Califordian Resolut: Procession of Califordian Advanced Wine "(Califordian Advanced Wine" (Califordian Advanced Technology (Califordian Advanced Wine" (Califordian Advanced Wine Wine Wine Wine Wine Wine Wine Wine	[Click here for GLOBP	LOT OUTPUT]			
PPFCOILS you! Collections of Collections (Control 1) Process you: Collections of Collections without of the (Cole have the PEPCOILS CONTROL 1) That-Productions of Collections without of the (Cole have the PEPCOILS CONTROL 1) That-Productions of Collections o	Click here for DISOPI	<u>ED OUTPUT J</u>			
EFCOLS organ : Check here for PERCOLS CUTTOR 1: TAM- Production of Transmonkrame Regions Result: MAP organ : Check here for FIRACOUTTOR 1: MAP or FIRA	2. PEPCOILS - Predi	tion of Coiled Coil R	legions Results:		
EFOOLS oppie: Code Coal regions a should be dreed by Effect (End End of EFOOLS OUTPOT _ II) MAP - dreed, "Instancembers regions and built dreed by COUTPOT _ II) MAP - dreed, "Instancembers regions and built dreed by COUTPOT _ II) Integrated Output - Colled Coals and Transmerbrane regions substituted with "* and Y: The Inste for the Instance of Coal and Coale regions we glid the acqueres is to different fragments, sequence fragments abund have at least 10 reaches for fBRA Sequence Sequents extracted from query sequence after fibring using EFFCOLS & TMAP: Sequence Sequents extracted from query sequence after fibring using EFFCOLS & TMAP: Sequence Sequents extracted from query sequence after fibring using EFFCOLS & TMAP: Sequence Sequents extracted from query sequence after fibring using EFFCOLS & TMAP: Sequence Sequents extracted from query sequence after fibring using EFFCOLS & TMAP: Sequence Sequents extracted from query sequence after fibring using EFFCOLS & TMAP: Sequence Sequents is Coole and Sequence To a sequence from the Sequence To a sequence from the Sequence To a sequence for sequence To a sequence To a sequence To a sequence for sequence To a sequence To a sequence for sequence To a sequence for sequence To a se	PEPCOILS output : [C	ick here for PEPCOIL	<u>.s output - I]</u>		
<pre>ThMP. Publicing of Transmolvem Regions Regions Results: Move ongo (Concentration Regions Regions</pre>	PEPCOILS output : Co	iled Coil regions substit	tuted with '=' [Click	t here for PEPCO	DILS OUTPUT - II]
MAP orgen [: Smaller for TMAP 2017F1]] Mar 2 data (in the intermediance region where the the sequence in the different fragments, sequence fragments should have at least 10 resolves for fills and for the properties of TMAP and Cale region we give the the sequence in the different fragments, sequence fragments should have at least 10 resolves for fills and for the properties of TMAP and Cale region we give the sequence in the different fragments, sequence fragments should have at least 10 resolves for fills and for the properties of TMAP and Cale region we give the sequence in the different fragments, sequence fragments should have at least 10 resolves for fills and for the properties of the different fragments, sequence fragments with 2 - 15% according structural context are used for further analysis. FOURTED Context for the properties of the fills and the PEPCOLIS & TMAP Rm : for the fills for sequence generation the different for PERCATS of the for terment	3. TMAP - Prediction	of Transmembrane H	Regions Results:		
MAP orger. Transmitherator regions substituted with 's' and 's: The hard for the laterated () force file.] Inter state of the laterated () force file.] Inter	TMAP output [Click he	re for TMAP OUTPU	<u>T - I]</u>		
Integrate Cale Cale and Taxameehear expises substituted with "s' and "	TMAP output : Transm	embrane regions substit	tuted with 'x' [Click	here for TMAP	<u>OUTPUT - II]</u>
Number of the integrated Output Bill is Note: Stated of the presence of TM and Color regions wight the sequence in the different frequence, preparence should have at least 10 reachang for fifther sequence in the different frequence, preparence should have at least 10 reachang for fifther sequence in the different frequence, preparence sequence in the different frequence, preparence sequence in the different frequence, preparence sequence in the different frequence sequence in the different of the preparence sequence in the different frequence sequence in the different of the preparence in the different of the preparence sequence in the different of the preparence in the different of the preparence sequence in the preparence sequence in the different of the preparence sequence in the prepa	4. Integrated Output	Coiled Coils and Tra	ansmembrane reg	ions substitute	l with '=' and 'x':
tote : Stands or the presence of TM and Colle region we split the sequence in to different fragment, sequence fragments thends have at least 30 residues for fifth we split the sequence sequent 1: [Calch lens for Sequence Sequents 1] FURPERD Results Represe Sequent 1: [Calch lens for Sequence Sequence 1] FURPERD Results Represe Sequent 1: [Calch lens for DETERING or the represent 1] FURPERD Results Represe Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence Sequence 1: [Calch lens for FORTERING or the represent 1] FURPERD Results Represe Sequence Sequence Sequence Sequence 1] Represe Sequence Sequence Sequence Sequence Sequence 1] Represe Sequence Sequence Sequence Sequence Sequence 1] Represe Sequence Sequence Sequence Sequence Sequence Sequence 1] Represe Sequence	[Click here for the Inter	rated Output file] :	-		
<pre>segures to the set of the segures requests after filtering using PEPCOLS & TMAP: In the segures reguest is '[26k here for Sequence Segures1]' FIFLED Results for requested and the set of the segures and year. Segures Segures 1: [26k here for FSEPEID on the for segures1] FIFLED Results for the set of the sequence segures at '16k here for FSEPEID at the for segures 1: [36k here for FSEPEID at the for Segures 1: [36k here for FSEPEID at the for Segures 1: [36k here for FSEPEID at the for Segures 1: [36k here for FSEPEID at the for Segures 1: [36k here for FSEPEID at the for Segures 1: [36k here for FSEPEID at the for Segures 1: [</pre>	Note : Based on the p	esence of TM and Co	ils region we split	the sequence in	to different fragments, sequence fragments should have at least 30 residues for füllthe
Sequence seguent 1: [Cick here for Sequence seguent 1] FIFERED Results for sequence seguent seturated for fEFCOLLS & TMAP Res: Expence seguent 1: [Cick here for Sequence seguent 1] FIFERED Results for sequence seguent seturated for fEFCOLLS & TMAP Res: Expence Seguent 1: [Cick here for FEFEFED out op for results 1] FIFERED Results FIFERED Results FIFERED RESULTS FIFERED RESULTS FIFERED RESULTS FIFERED RESULTS FIFERED RESULTS FIFERED RE	analysis.				
<pre>separate segurat 1: [Cack lears for Segurate Segurate 1] FSUFER Decalls for sequence segurates sequence segurates control of ser USA of ALAP Rue: to choose segurates segurates segurates sequence segurates control of sequence 1] FSUFEACT Revals FSUFER Revals FSUFE</pre>	5. Sequence Segment	s extracted from quer	ry sequence after	filtering using	PEPCOILS & TMAP :
PUPRED Results for sequence segments extracted after PEPCOLLS & TANAP Run: terpence Sequence 1: [Cack here for PEREED out one for segment :] PEILLAST Results ends for Sequence Sequence 1: [Cack here for PEREED out one for segment :] out Ungane the "O Y marker PEIL BLAST Here for PEREED To set one for segment :] PEILLAST Results Outgoing the "O Y marker PEIL BLAST Here for PERCETTE out one for segment :] PEILLAST Results Outgoing the "O Y marker PEIL BLAST Here are in MView form of its [Wiew outgoing of PEIELAST for regnent :] PEILLAST Results Outgoing the "O Y marker PEIL BLAST Here are in MView form of its [Wiew outgoing of PEIELAST for regnent :] PEILLAST Results Outgoing the "O Y Here PEILLAST HER after CD- HIT [Sequence Segment 1] = 48 PEILLAST Results OUTGOING THE TO Y HERE AND THE INFORMATION (INFORMATION	Sequence segment 1 : [Click here for Sequence	e Segment 1]		
<pre>international sector in the international sector international activated for further analysis. expense Segment 1: [Cick here for PSEPED out put for segment 1] PROFINE Results expense Segment 1: [Cick here for PSEPED out put for segment 1] and Urape him = 95: Visualer PSI-PLAST contrast for segment 1] and Urape him = 95: Visualer PSI-PLAST contrast for segment 1] and Urape him = 95: Visualer PSI-PLAST contrast for segment 1] CO-HIT Result: monologies form TSI-PLAST Him after CD-HIT [Sequence Segment 1] = 48 MONOPMA texts for expense Segment 1: [Cick here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] COSTAM Texts for expense 2 (Cick here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for expense 2 (Cick here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for expense 2 (Cick here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for expense 2 (Cick here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for POST (Cick Here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for POST (Cick Here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for POST (Cick Here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for POST (Cick Here for DOSTAM contrast for 2 (Cick here for DOSTAM contrast for expenses 1] Cick Here for DOSTAM contrast for expenses 1</pre>	6 PSTPRED Desertes	for sequence com	te extracted ofter	PEPCOTCA	TMAP Bun ·
sequence Seguent 1: [Citck here for PSDFRED out out for segment 1] PSELLAST Revalu: multi for Sequence Seguent 1: [Citck here for PSDFLAST out put for segment 1] Out There are a sequence seguent 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: Out Durate the 1 sequence segment 1: [Citck here for PSDFLAST out put for segment 1] POSTE Revalu: POSTE Revalu: POSTE	Note : Only sequences	with $>= 15\%$ second	lary structural con	tent are used fo	r further analysis.
<pre>square: Seguret 1: [Cack here for FSTPRED: out patifor regener.]] FST.JLAST Result: FST.JLAST Res</pre>					
PS-FILAST Result: ende for Sequence 1: [Cick here for PSELAST one pat for resument :]] to Urapien the - 20. Fundame PSTELAST Results in NA/Wew forman 1: [NA/Wew compan of PSEELAST for resument :]] TFOURT Result: CO-DITT	Sequence Segment 1 :	Click here for PSIPRE	D out put for segm	ent : 1]	
<pre>ends for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1: [LtChere endpets of PSUBLAST for sequence 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for Sequence Sequence 1: [Clock here for PSUBLAST cost performance 1:]] POSITE Reads: Total PSUBLAST the for PSUBLAST Here device 1: [PSUBLAST cost performance 1:]] POSITE Reads: Total Partner for PSUBLAST Here device 1: [PSUBLAST cost performance 1:]] POSITE Reads: Total PSUBLAST PSUBLAST</pre>	7. PSI-BLAST Resul	s:			
tal Unique his - 89 · Vinaiñe PSI-ELAST Rends in MView forma 1 : [MView output of PSIELAST for segment _]] PROFIT Rends: rote pattern for Sequence Segmen 1 : [Cick here for PROSTE out out for segment _]] CD-HIT Rends: DMOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 2 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 2 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 2 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 2 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1453-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74654, 1452-721.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-1551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-1551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-1551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-1551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-1551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context SNP : 74694, 1510-1551.htm] MOPPAM Rends for query 1 [Cick here for IBMOPPAM context	Results for Sequence S	gment 1 : [Click here f	for PSIBLAST out	put for segment :	1]
PROSITE Result: To the presenter Sequent 0: [Cick here for PROSITE out out for regment .] CD-HI Result: CD-HI Result: CD-HI Result: CD-HI Result: CD-HI Comparison of the com	Total Unique hits = 89	Visualise PSI-BLAST	Results in MView	format 1 : [MVie	w output of PSIBLAST for segment : 1]
Concent patherms for Sequence Seguent 1: [Cick here for PROSTE out put for regnent:1] CD-HIT Reach: CO-HIT REACH:	8. PROSITE Results				
CD-HIT Results: cmologree from PSI-RLAST Hits after CD-HIT [Sequence Segment 1] = 48 IMMEPVAN Results IMMEPVAN Result	Prosite patterns for Sec	uence Segment 1 : [Cliv	ck here for PROSI	TE out put for se	gment : 1]
<pre>SUD-III Kendit: moniogues for DPS-BLRAST His after CD-HTT [Sequence Segment 1] = 48</pre>					
Montregate is contracted and and contract (SP 2011) (Section 2012)	9. CD-HIT Results:	BI AST Hite after CD.	HTT (Sequence Se	ment 11 = 48	
<pre>iBMDPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 714652.1652-721 hzm] ADPPAAM Reads for query 2 { Cack here for IBADGPAAM detext of NP: 714552.1652.71 hzm] ADPPAAM Reads for query 3 { Cack here for IBADGPAAM detext of NP: 714552.1652.71 hzm] ADPPAAM Reads for query 5 { Cack here for IBADGPAAM detext of NP: 713572.1622.70 hzm] ADPPAAM Reads for query 5 { Cack here for IBADGPAAM detext of NP: 713572.1622.70 hzm] ADPPAAM Reads for query 5 { Cack here for IBADGPAAM detext of NP: 713572.1622.70 hzm] ADPPAAM Reads for query 5 { Cack here for IBADGPAAM detext of NP: 71573.1622.70 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71573.1728.771 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71573.1728.771 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71573.1128.771 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71573.1128.771 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71573.1128.771 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71573.1128.771 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71574.1165.7721 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71574.1165.7721 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71574.1165.7721 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71574.1165.7721 hzm] ADPPAAM Reads for query 1 { Cack here for IBADGPAAM detext of NP: 71577.1173.11731.11731.11731.11731.1174.1174</pre>	Homologues Hom F.SI-	SEAST THIS ALL OF	IIII [Dequence De	ginene 1j = 40	
- Graphical Representation of PURE Result:	HMMPFAM Results for qu HMMPFAM Results for qu	17 41 (31ck here for HMM 175 (31ck here for HMM 176 (31ck here for HMM 177 (31ck here for HMM 178 (31ck here for HMM 179 (31ck here for HMM) 179 (31ck here for HMM) 170 (31ck here for HMM) 170 (31ck here for HMM) 170 (31ck here for HMM) 171 (31ck here for HMM) 172 (31ck here for H	GPFAM output of YP GPFAM output of YP GPFAM output of YP GPFAM output of YP GPFAM output of YP MFFAM output of ZP MFFAM output of ZP	660078.1295.261 43376.1660.733 867588.1620.659 7971305.1728.797 797390.12.73 hm 422981.1672.742 00961161.1671- 769493.1510.58 01612033.1665- 3934421.667.77 01623122.1106 630940.1762.85 01109224.1287- 720573.1057- 165394.1287- 1016212.1106 650940.1762.85 01109224.1287- 1016213.1664- 0109224.1287- 10160155.1282- 001460197.189 01132186.1643- 10	hema] 244 hema] 244 hema] 244 hema] 244 hema] 243 hema] 243 hema] 240 hema] 240 hema] 240 hema] 240 hema] 25 hema] 26 hema] 27 hema h 27 hem
100 100 <td>12. Graphical Represe</td> <td>ntation of PURE Bass</td> <td>ult:</td> <td></td> <td></td>	12. Graphical Represe	ntation of PURE Bass	ult:		
hery Second Second Sec	<	200	300	400	500 600 F
charges_categor_1. charges_categor_2. conset	Query Score =0				
<pre>index rest inter inter rest inter inter rest inter inter inte</pre>	Score =0 PD:TPP 21CS:7P 00407010				
works woods works woods woods works woods works woods w	Score =0.0042 PD:TPR_2ICS:NP_768978				
Server successfully processed your sequence	Score =0.00041 PD:TPR_1 CS:ZP_01109234				
Constant Constant Const	Score =0.008 PD:TPR_2 CS:ZP_01109234				
Sore = 00039 Sore = 00039 Sore = 00030 Sore = 00030 Sore = 000300 Sore = 000300	Score =0.00023 PD:TPR_2 CS:ZP_01165594				
Sover = 0,	Score =0.0019 PD:TPR_2ICS:NP_971542				
Some %0.0013 Sin J. 163 (1997) Sin J. 163 (1997)	Score =0.0088 PD:TPR_1 CS:NP_714662				
Second Se	Score #0.0013 PD:TPR_2 CS:NP_714662				
Text Environment approximate approximate from Course and the first approximate approximate approximate approximate from Course approximate approximate approximate from Course approximate	PD:TPR_1ICS:YP_660078				
Terrer 21 (15379: 279390 POTER 21 (15379: 2004) POTER 21 (15379: 2004) Sorrer 10, 2007 Sorrer 10, 2007	PD:TPR_21CS:YP_660078				
P01785_2155.95_00437139 Sours = 0.0075 . Cancheding Remarks: R.E. Server successfully processed your sequence R.E. Server successfully processed your sequence R.E. Server successfully processed your sequence 	PD:TPR_21CS:YP_797390				
Concluding Remarks: TRE Server successfully processed your sequence TRE Server successfully processed your sequence = 1	PD:TPR_2[CS:XP_001437:	39			
, Concluding Remarks: IRE Server successfully processed your sequence to Windows of Servence Serversted from Overs Serverse = 1		CONTRACTOR OF A	a a ta ta ta ta ta	C1 X 13 P1 13	
tol Number of Seguence Segment extracted from Query Seguence = 1	12. Concluding Remar PURE Server successful	cs: y processed your seaue:	nce		
tea avantos os sequence segurente extracted nom Query Sequence = 1	Total Number of Sequen	e Segment extracted fr	om Query Sequence	e = 1	



Table I: Co	oncluding rema	rks from PURE	server based	on various outputs
-------------	----------------	---------------	--------------	--------------------

Case	Output details	Concluding remarks
I	Output Files : Globplot, Disopred2 and Pepcoils or Globplot, Disopred2 and Tmap or Globplot,, Disopred2, Pepcoils and Tmap.	Query sequence is divided into sequence segments based on Pepcoils and Tmap output. After excluding the sequence portions encoded by transmembrane and coiled-coils, the unassigned region could be associated with coiled coils or TM helices and the rest of the unassigned regions are insignificantly small and therefore structure association terminates.
2	Output Files : Pepcoils, Tmap and psipred	psipred results suggest very little secondary structure, therefore the unassigned region is to be viewed as unstructured and procedure terminates.
3	Output Files : Pepcoils, Tmap, psipred and PSI-LAST hits (less homologues)	PSIBLAST did not provide enough homologues and terminates. Here, the user could relax thresholds or choose a different database and retry. If there is enough secondary structural content, this unassigned region could be a potential new domain that is somewhat species-specific and hence not evolutionarily conserved.
4	Output files: Pepcoils, Tmap, psipred, PSI-BLAST and Hmmpfam out put (no hits in Hmmpfam search)	The indirect HMM runs and the consensus did not yield any hit to a Pfam domain. The user can relax cd-hit and get more homologues included in the search or relax the HMM threshold and retry. Or else, if there is enough predicted secondary structural content and the unassigned region is evolutionarily conserved, this could point to a potential new and novel domain.
5	Output files: Pepcoils, Tmap, psipred, PSI-BLAST and Hmmpfam out put (hits in Hmmpfam search with partial domain assignment)	The indirect HMM runs point to partial assignment to a pre-existing Pfam domain family. This could be due to suboptimal or spurious alignment or point to the presence of discontinuous domains in the query.
6	Output files: Pepcoils, Tmap, psipred, PSI-BLAST and Hmmpfam out put (hits in Hmmpfam search with complete domain assignment)	The indirect HMM runs point to assignment to a pre-existing Pfam domain family through one or more homologues. This could be due to distant relationships where the homologues act as intermediates or due to borderline E-values. New connection to an old domain!
7	Output files: Pepcoils, Tmap, psipred, PSI-BLAST and Hmmpfam out put (hits in Hmmpfam search against UR itself)	Direct HMM assignment to a pre-existing Pfam family. This could be due to newer Pfam domain entries or slight differences in thresholds to E-values. Connection to a fairly new domain.

and simultaneously decrease the cd-hit clustering threshold so that fewer representatives are examined for the domain architecture. This strategy is recommended to improve the sampling of sequence space.

We have tested the reliability of the PURE algorithm as well as the performance of the server using various datasets. PURE protocol was successfully employed to annotate 67 adenylyl cyclase proteins with unassigned regions [24] and in *Mycoplasma gallisepticum* genome analysis where 82 new domains are added to 72 proteins out of which 48 proteins are earlier completely unassigned (unpublished data).

Conclusion

Due to advancement in high-throughput techniques, sequence data are generating at a rapid pace, but the biochemical validation of available sequences is still a challenge. Efficient computational methods can be employed to overcome the lacunae in high-throughput sequence data generation and function annotation. PURE can be used as an efficient and computationally intensive protocol to successfully annotate unassigned regions in sequences.

Availability and requirements

Project name: PURE – Prediction of domains in the protein unassigned regions

Project home page: [26]

A sample consensus page, with above explained features, is available: [27]

Pre-computed results for selected unassigned regions from earlier analysis [9] are available: [28]

Operating system(s): Platform independent (web server)

Programming language: HTML, Perl, CGI, Java script

License: Free for academics, Authorization license needed for commercial usage (Please contact the corresponding author for more details)

Any restrictions to use by non-academics: license needed

Authors' contributions

RS and BOO conceived and helped in the implementation of the webserver idea. CCSR and KS have developed the scripts and tools for the webserver. CCSR and KS wrote the manuscript; RS and BOO provided critical inputs to improve the manuscript.

Acknowledgements

RS was a senior research fellow of the Wellcome trust, U.K. by which KS's stay could be supported. CSR is supported by a PhD scholarship from Conseil Régional de La Réunion. BO is partially funded by PRES Ocean Indien from Université de La Reunion. We would also like to thank NCBS for infrastructural support. The authors would like to thank Prof. N. Srinivasan and Dr. Alexandre de Brevern for helpful suggestions.

References

- Copley RR, Doerks T, Letunic I, Bork P: Protein domain analysis in the era of complete genomes. FEBS Lett 2002/03/26 edition. 2002, 513(1):129-134.
- Holm L, Sander C: Parser for protein folding units. Proteins 1994/ 07/01 edition. 1994, 19(3):256-268.
- Dietmann S, Holm L: Identification of homology in protein structure classification. Nat Struct Biol 2001/10/31 edition. 2001, 8(11):953-957.
- Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM: CATH--a hierarchic classification of protein domain structures. Structure 1997/08/15 edition. 1997, 5(8):1093-1108.
- Murzin AG, Brenner SE, Hubbard T, Chothia C: SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995/04/07 edition. 1995, 247(4):536-540.
- Sowdhamini R, Blundell TL: An automatic method involving cluster analysis of secondary structures for the identification of domains in proteins. Protein Sci 1995/03/01 edition. 1995, 4(3):506-520.
- Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R: Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res 1998/02/21 edition. 1998, 26(1):320-322.
- Salamov AA, Suwa M, Orengo CA, Swindells MB: Combining sensitive database searches with multiple intermediates to detect distant homologues. *Protein Eng* 1999/04/09 edition. 1999, 12(2):95-100.
- Reddy CS, Manonmani A, Babu M, Sowdhamini R: Enhanced structure prediction of gene products containing class III adenylyl cyclase domains. In Silico Biol 2007/02/06 edition. 2006, 6(5):351-362.
- Park J, Teichmann SA, Hubbard T, Chothia C: Intermediate sequences increase the detection of homology between sequences. *J Mol Biol* 1997/11/21 edition. 1997, 273(1):349-354.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR: The Pfam protein families database. Nucleic Acids Res 2003/12/19 edition. 2004, 32(Database issue):D138-41.
- Linding R, Russell RB, Neduva V, Gibson TJ: GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 2003/06/26 edition. 2003, 31(13):3701-3708.
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT: Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004/03/17 edition. 2004, 337(3):635-645.
- Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000/05/29 edition. 2000, 16(6):276-277.
- McGuffin LJ, Bryson K, Jones DT: The PSIPRED protein structure prediction server. *Bioinformatics* 2000/06/27 edition. 2000, 16(4):404-405.
- de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N: ScanProsite: detection of PROSITE signature matches and ProRule-associated func-

tional and structural residues in proteins. Nucleic Acids Res 2006/07/18 edition. 2006, 34(Web Server issue):W362-5.

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997/09/01 edition. 1997, 25(17):3389-3402.
- Li W, Godzik A: Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006/05/30 edition. 2006, 22(13):1658-1659.
- Eddy SR: Profile hidden Markov models. Bioinformatics 1999/01/ 27 edition. 1998, 14(9):755-763.
- Brown NP, Leroy C, Sander C: MView: a web-compatible database search or multiple alignment viewer. Bioinformatics 1998/ 06/20 edition. 1998, 14(4):380-381.
- Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BJ, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E: The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* 2002/10/09 edition. 2002, 12(10):1611-1618.
- Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E: Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2006/12/16 edition. 2007, 35(Database issue):D5-12.
- 23. Bairoch A, Apweiler R: The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 1999/12/11 edition. 2000, 28(1):45-48.
- Lupas A, Van Dyke M, Stock J: Predicting coiled coils from protein sequences. Science 1991/05/24 edition. 1991, 252(5010):1162-1164.
- Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ: The PROSITE database. Nucleic Acids Res 2005/12/31 edition. 2006, 34(Database issue):D227-30.
- 26. **PURE server home page** [http://caps.ncbs.res.in/PURE]
- 27. Sample consensus page [http://caps.ncbs.res.in/PURE/ PURE Results/PURE Consensus/
- TueAug21095436IST2007 consensus.html]
- 28. Pre-computed results [http://caps.ncbs.res.in/PURE/ example result.html]

