

SPICE & DAS registry

bridging from genome to structure

Andreas Prlić



- ID history
- Genomic sequence
- Transcript information
- Exon information
- Protein information
- Export protein data

Chromosome 17
74,183,768 - 74,289,900

- View of Chromosome 17
- Graphical view
- Graphical overview
- Export information about region
- Export sequence as FASTA
- Export EMBL file
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Ensembl Archive

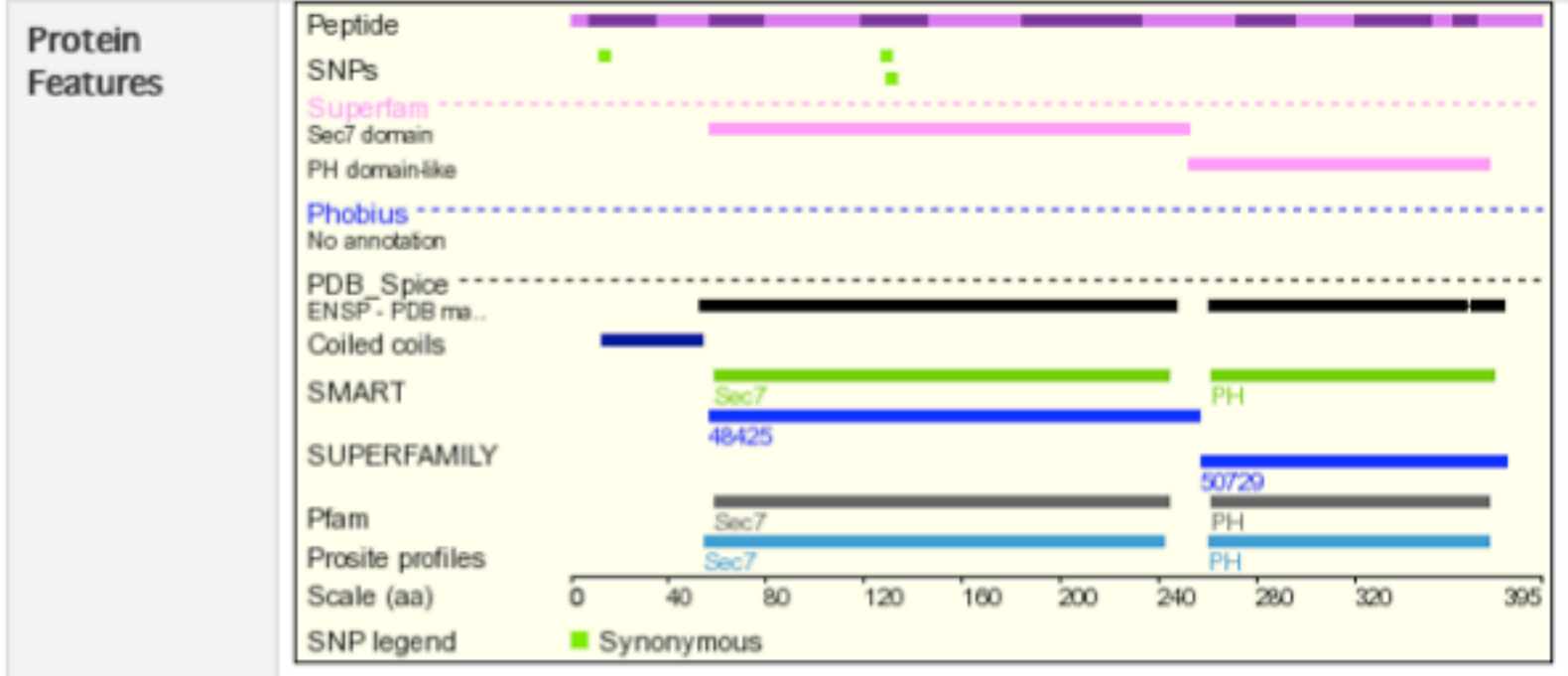
- View previous release of page in Archive!
- Stable Archive! link for this page



with available aligned cDNAs to annotate UTRs (For more information see V.Curwen et al. , Genome Res. 2004 14:942-50.)

InterPro [IPR000904](#) SEC7-like - [[View other genes with this domain](#)]
[IPR001849](#) Pleckstrin-like - [[View other genes with this domain](#)]

Protein Family [ENSF00000001295](#) : CYTOHESIN
 This cluster contains 4 Ensembl gene member(s) in this species.



Protein Sequence

```

MEEEDSYVPSDLTAIEPQILENIDRDKQELLADIQRLKDEIAEVANEIENLGSSTEEIKNM
QANKQVMGKXKINPDKKQIQFLIENDLQAQTCEDIAQFLYKGEGLNKTARIGDYLGEPD
IINIQVLHAFVILKEITDNLVQALDQFLMSTRLPGEAOKIDRNPTEAFAQRQCCHNGVT
QSTDTCVLSTAIINLWNTSLQGNWKKKPTVERFIAPNRGINDGGDLPEELLNLYEIK
NEPFKIPEDDGHDLTHTTFINPDRIGWLLKLGSRVKTWKRRWFILTDNCLYTYEYTPDKEP
RGIIPLENLSIRIVEDSKKPNCFELYIPDKKQVVKACKTEADGRVVEIGHKFTQSGLRPP
EKIEWIKCIKRAISDDFFYIHLAARAKKKVSSIKRH
  
```

Show the following features:

Number residues:

Refresh

- ID history
- Genomic sequence
- Transcript information
- Exon information
- Protein information
- Export protein data

Chromosome 17
74,183,768 - 74,289,900

- View of Chromosome 17
- Graphical view
- Graphical overview
- Export information about region
- Export sequence as FASTA
- Export EMBL file
- Export Gene info in region
- Export SNP info in region
- Export Vega info in region

Ensembl Archive

- ! View previous release of page in Archive!
- ! Stable Archive! link for this page



with available aligned cDNAs to annotate UTRs (For more information see V.Curwen et al. , Genome Res. 2004 14:942-50.)

InterPro
[IPR000904](#) SEC7-like - [[View other genes with this domain](#)]
[IPR001849](#) Pleckstrin-like - [[View other genes with this domain](#)]

Protein Family
[ENSF00000001295](#) : CYTOHESIN
 This cluster contains 4 Ensembl gene member(s) in this species.

Protein Features

Peptide
 SNPs
 Superfam
 Sec7 domain
 PH domain-like
 Phobius
 No annotation
 PDB_Spice
 ENSP - PDB ma...
 Coiled coils
 SMART
 SUPERFAMILY
 Pfam
 Prosite profiles
 Scale (aa) 0 40 80 120 160
 SNP legend
 ■ Synonymous

Protein Sequence

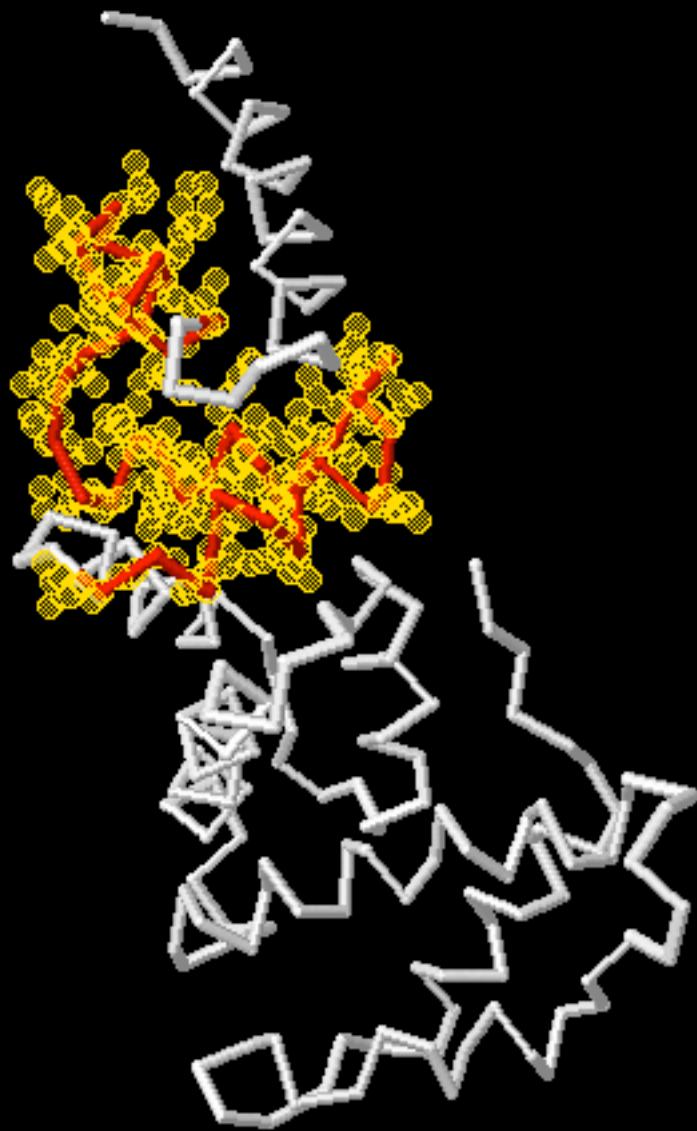
```

MEEDSYVPSDLTAIEPQLENIIDKQYELLADIQRLKDEIAEVAHEIENLGSTY
QHNKQVANGKXKINQPKKQIQFLIENDLIANTCEDIAQFLYKGEGLNKTATIGD
EENIQVLAHAFVELKEFTDLMLVQALQFLNSFRLPGEAQKIDRPTAEFAQRVCOQ
QSTDTCYVLSFAIIPMLNLSLGGHNVKDKPTVERFIAMNQGINDGGDLPEELLNMI
NEPPKIPEDDGHDLNHTTINPDDEIGWLLKLGGRVKTWKRRWFILTDNCLYYEYV
DGIIPLENLSIREVEDSKKPNCFELYIPDKKQVIKACKTEADGRVVEGHHFTQS
EKKEEWKCKIKRAIISDPPFYEMLAARKKKVSSTKRH
  
```

Show the following features:
 Number residues:

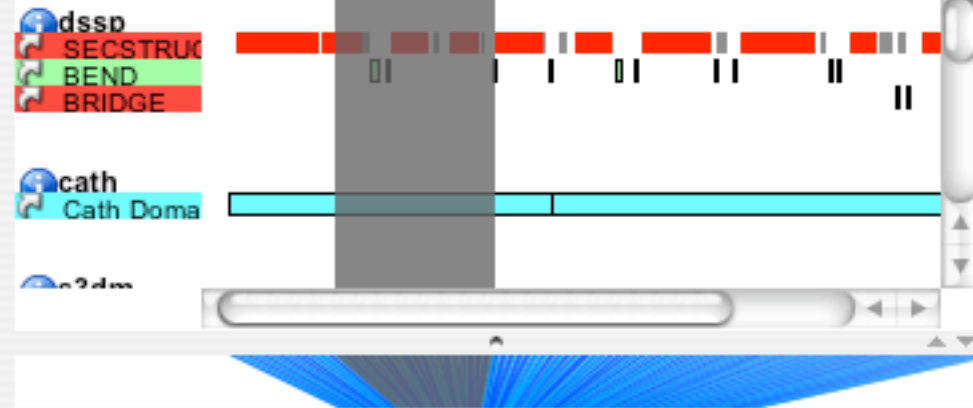
PDB: 1pbv. via UniProt Q99418 PDB: ARNO
 ID: ENSP - PDB mapping to 1pbv.
 TYPE: ENSP - PDB mapping
 METHOD: Compara, MSD-Mapping of UniProt to
 LINK:
<http://das.sanger.ac.uk/registry/showspice.jsp?pc>
 NOTE: PDB: 1pbv. 85%id 49%coverage via UniProt Q99418 PDB:ARNO

also works in contig-view



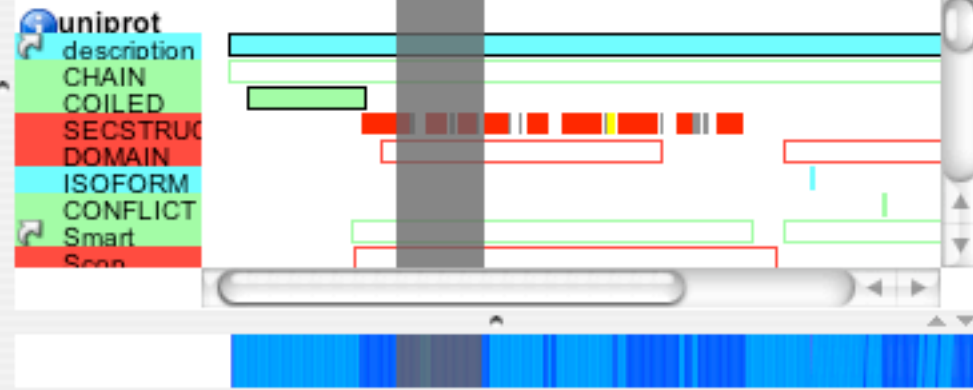
PDB

1PBV



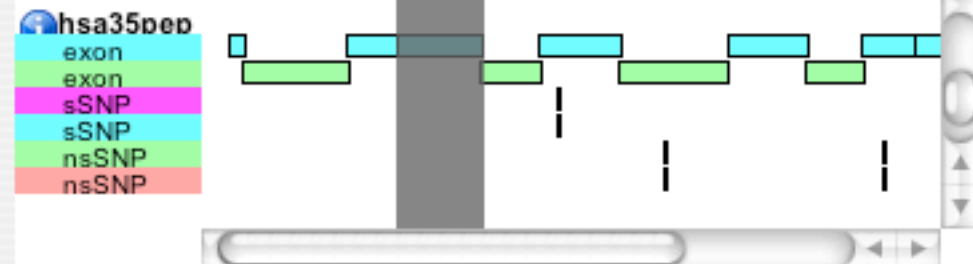
UniProt

Q99418



ENSP

ENSP00000314566



SISYPHUS

Submit Query

Structural alignments for proteins with non-trivial relationships



Sisyphus: in greek: crafty

Authors:

Antonina Andreeva, Andreas Prlic, Tim Hubbard, Alexey Murzin

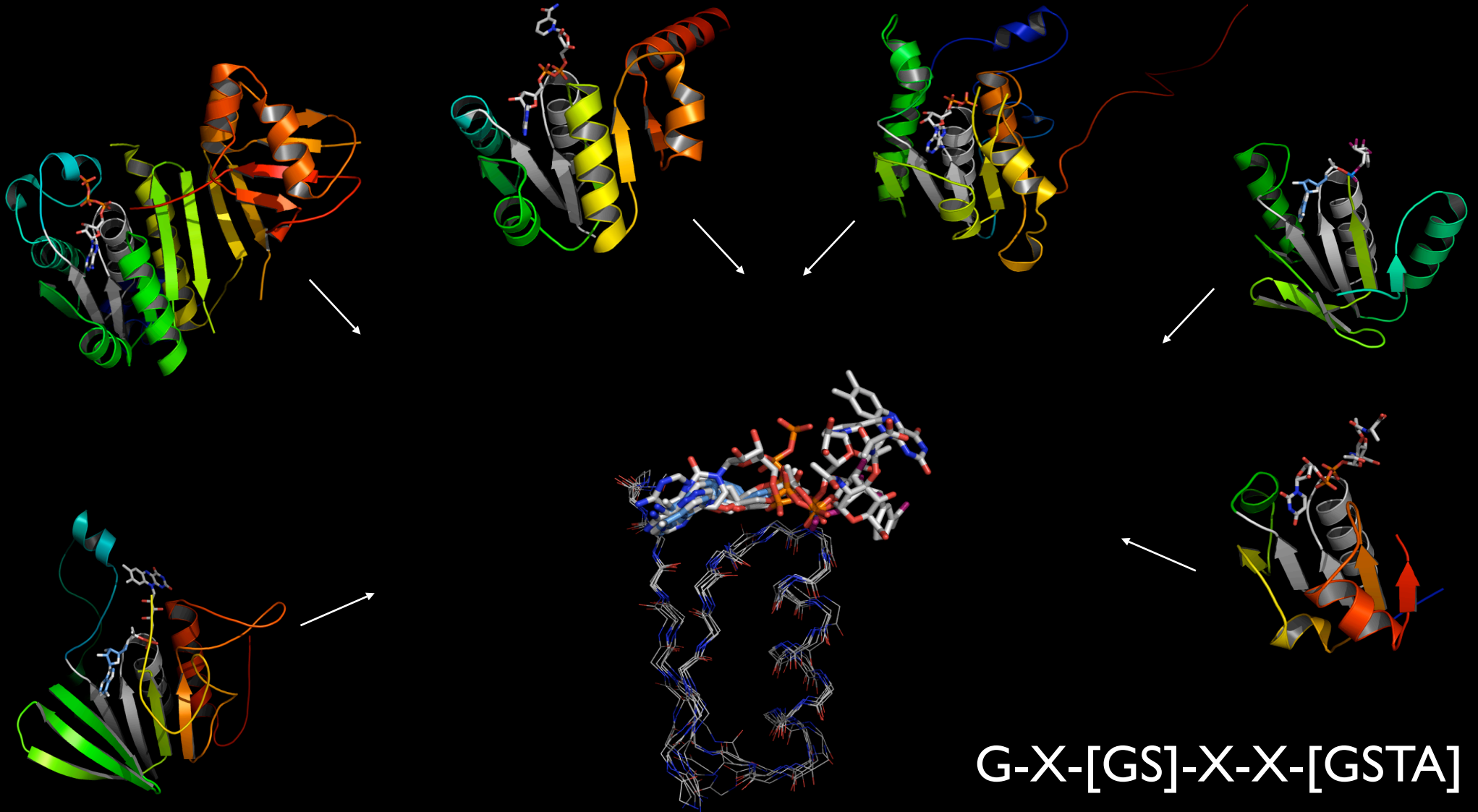
```

SATVGIFVDagsraenvknNGTAHFLEHLAFkgtqnr-----pgQGIELeienig---shLNAYTsr----eNTVYYAKSLq--edIPKAVDI
TTHIAIALEgvswsap--dYFVALATQAIIVGnwdraigtgtn-----spSPLAVaasqngslansYMSFStsyadsgLWGMYIVTDSnehnVRLIVNE
FSALGLYIDagsrfegrnlKGCTHILDRLAFkstehv-----egRAMAEtlellg---gnYQCTSSr----eNLMYQASVFN--gdV GKMLQL
LFHIQIGFEGlpidhp--dIYALATLQTLGgggsfsagppgkgmysrlyTHVLNqy---yfvencVAFNhsysdsgIFGISLSCIP--gaAPQAVEV
TCTVGWVIDagsryeseknNGAGYFVEHLAFkgtknr-----pgNALEKevesmg---ahLNAYStr----eHTAYYIKALs--kdLPKAVEL
LAHVAIAVEgpgwahp--dNVALQVANAIIGHydctygggah-----lsSPLASiaatn-klcqsFQTFNicyadtgLLGAHFVCDhms--IDDMMFV
TCTVGWVIDagsryeseknNGAGYFLEHLAFkgtknr-----pgNALEKevesmg---ahLNAYSSr----eHTAYYIKALs--kdVPKAVEL
LAHVAIAVEgpgwahp--dLVALQVANAIIGHydrtyggglh-----ssSPLASiavtn-klcqsFQTFNicysetgLFQGFYFVCDrms--IDDMMFV
TASVGVVFGsgaanenpynNGVSNLWKNIFL-----sKENSavaakeg---laLSSNISr----dFQSYIVSSLp--gsTDKSLDF
KAWISLAVEgepvnsn--nYFVAKLAAQIFGsynafepasrl-----ggIKLLDnigey-qlcdnFNHFSlsykdsGLWGFSTATRNvt-mIDDLIHF
ISTLAVKVHggstryat--kDGVAHLLNRFNFgntntr-----saLKLVRResellg---gtFKSTLdr----eYITLKATFLk--ddLPYYVNA
DSVAAIGIPvn---ka--sLAQYEVLANYLtsal-----SELGLi-----SSAKLdkftdggLFTLFVRDQ--dsavVSSNIKK
LVHAAIAVESaaigga--eANAFSVLQHVLG-----ANPHVkrngnp---fdVSAFNasysdsgLFGFYTISQaay--AGQVIKA
GSTIGVFIKagsryenssnLGTSHLLRLASSlttkga-----ssFKITRgieavg---gkLSVestr----eNMAYTVECLr--ddVEILMEF
ASRIGLFIKagsryenssnLGTSHLLRLASSlttkga-----ssFKITRgieavg---gkLSVestr----eNMAYTVECLr--ddVDILMEF
LVHAALVAESaaigsa--eANAFSVLQHVLGagphvkragsna-----tSSLYQavakgvhgpfcv---asysds---FEYTISQa--gdVVIKA

```



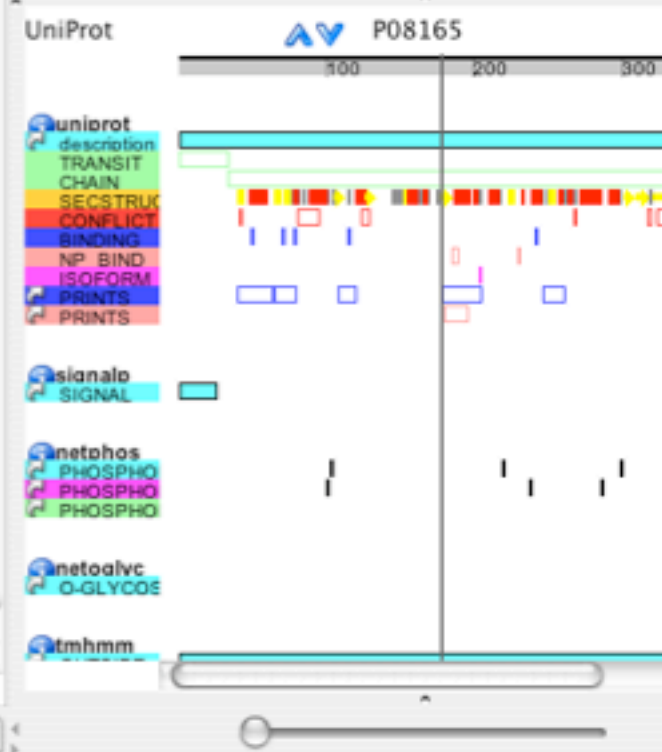
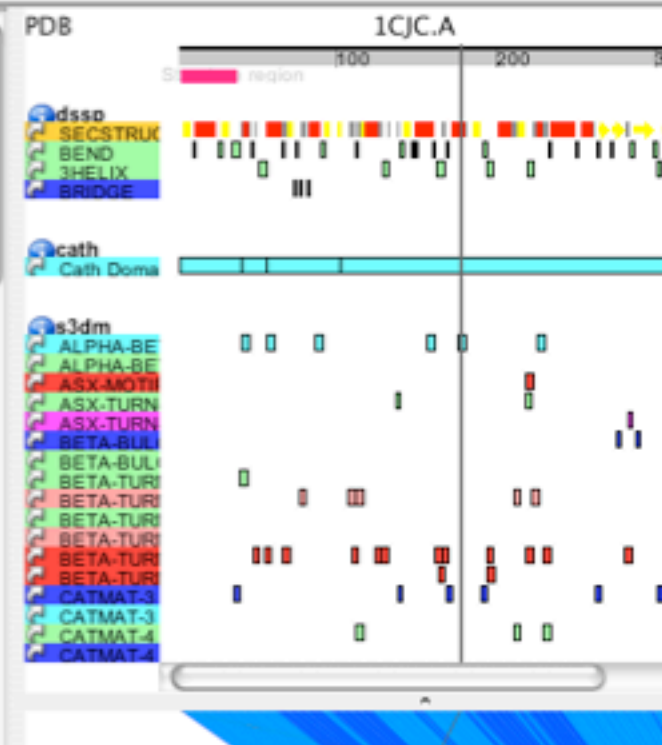
Nucleotide binding motif

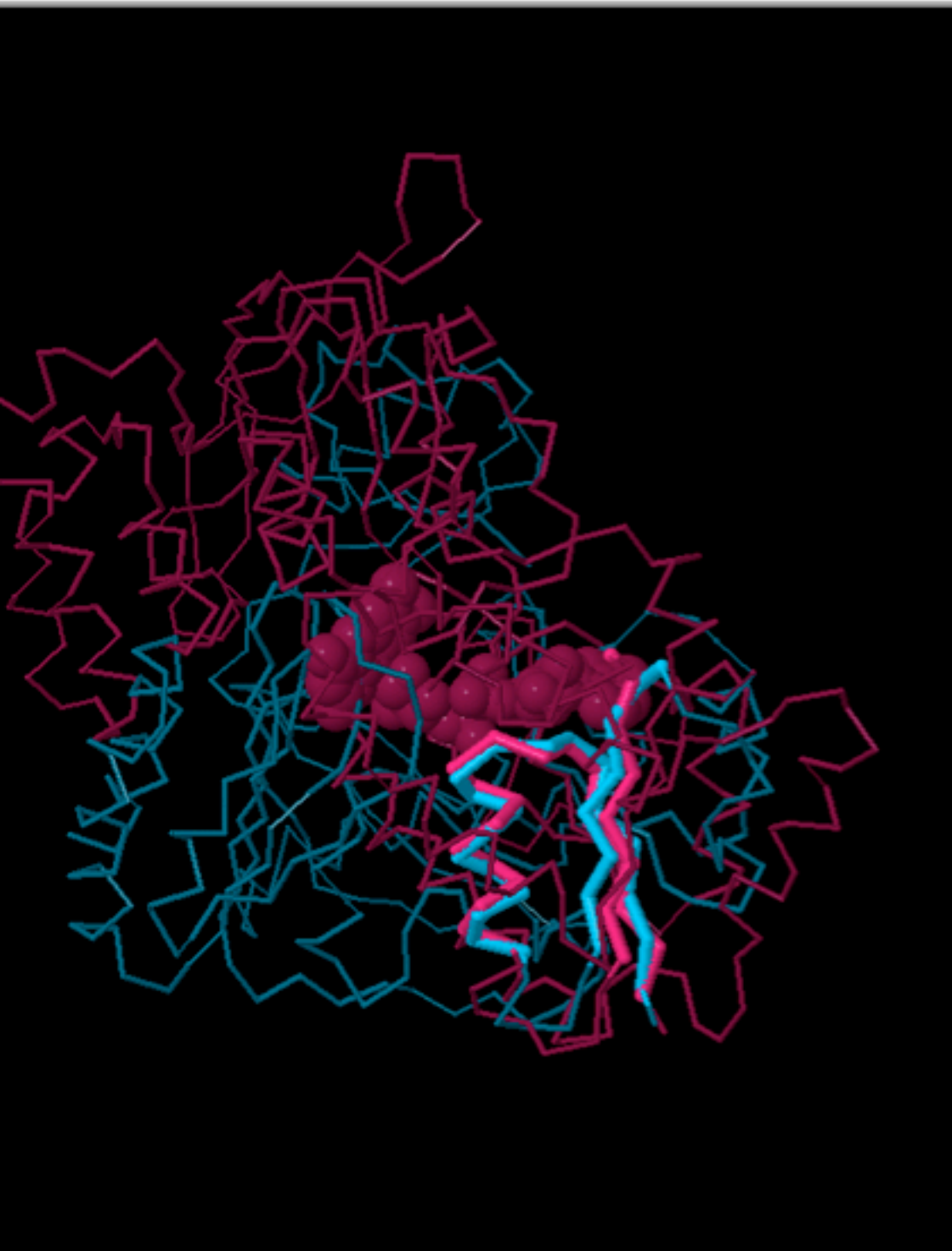


G-X-[GS]-X-X-[GSTA]



- 1 1f8w.A_A150:168
- 2 1e0d.A_A6:24_A3
- 3 1b6r.A_A2:20_A2
- 4 1eyz.A_A13:31_A
- 5 1jw9.B_B32:50_B5
- 6 1yov.B_B49:67_B7
- 7 1djn.A_A390:408
- 8 1ps9.A_A374:392
- 9 1cjc.A_A7:25_A33
- 10 1lqt.A_A4:22_A3
- 11 1gt8.A_A188:200
- 12 1i8t.A_A2:20_A2
- 13 1c0i.A_A1005:10
- 14 1an9.A_A1:19_A
- 15 1ps9.A_A495:51
- 16 1cjc.A_A146:164
- 17 1lqt.A_A148:166
- 18 1gt8.A_A333:35
- 19 1coy_12:30_36:4
- 20 1mxt.A_A11:29_
- 21 1cf3.A_A20:38_A
- 22 1gpe.A_A25:43_
- 23 1ju2.A_A27:45_A
- 24 1kdg.B_B217:235
- 25 1pbe_3:21_27:34
- 26 1iut_3:21_27:34
- 27 2gb0.B_B4:22_B2
- 28 1ng3.A_A5:23_A
- 29 1foh.A_A8:26_A3
- 30 1b37.B_B5:23_B3
- 31 1f8r.A_A34:52_A
- 32 1reo.A_A34:52_A
- 33 1oj9.A_A5:23_A2





- 1 1f8w.A_A150:168
- 2 1e0d.A_A6:24_A3
- 3 1b6r.A_A2:20_A2
- 4 1eyz.A_A13:31_A
- 5 1jw9.B_B32:50_B5
- 6 1yov.B_B49:67_B7
- 7 1djn.A_A390:408
- 8 1ps9.A_A374:392
- 9 1cjc.A_A7:25_A33
- 10 1lqt.A_A4:22_A3
- 11 1gt8.A_A188:200
- 12 1i8t.A_A2:20_A2
- 13 1c0i.A_A1005:10
- 14 1an9.A_A1:19_A
- 15 1ps9.A_A495:51
- 16 1cjc.A_A146:164
- 17 1lqt.A_A148:166
- 18 1gt8.A_A333:351
- 19 1coy_12:30_36:4
- 20 1mxt.A_A11:29_
- 21 1cf3.A_A20:38_A
- 22 1gpe.A_A25:43_
- 23 1ju2.A_A27:45_A
- 24 1kdg.B_B217:235
- 25 1pbe_3:21_27:34
- 26 1iut_3:21_27:34
- 27 2gb0.B_B4:22_B2
- 28 1ng3.A_A5:23_A
- 29 1foh.A_A8:26_A3
- 30 1b37.B_B5:23_B3
- 31 1f8r.A_A34:52_A
- 32 1reo.A_A34:52_A
- 33 1oj9.A_A5:23_A2



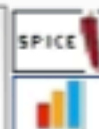
3D structure evaluation

First Model [ALL Models](#) | Category filter: ALL [TBM](#) [FM](#) | Sort targets by: name, [target size](#) | [Help](#)

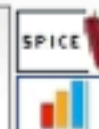
T0283 (TBM)
TABLES
(1-97)



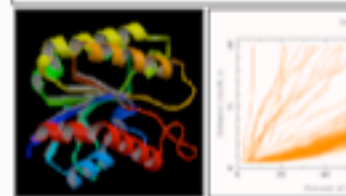
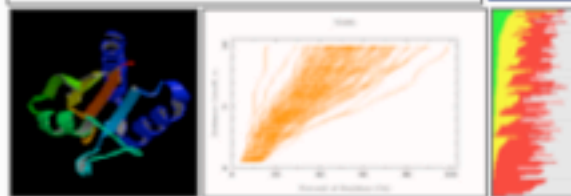
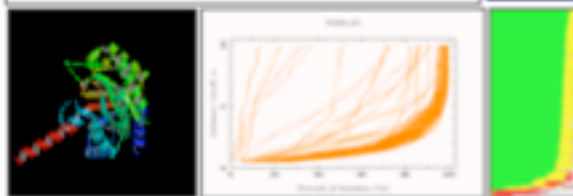
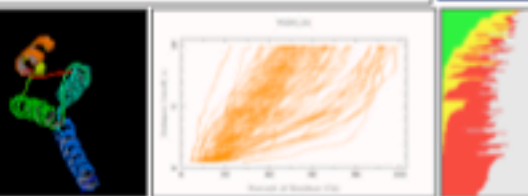
T0284 (TBM)
TABLES
(4-253)



T0285 (TBM)
TABLES
(7-105)



T0286 (TBM)
TABLES
(3-204)



T0287 (FM)
TABLES
(21-181)



T0288 (TBM)
TABLES
(1-26, 29-88)



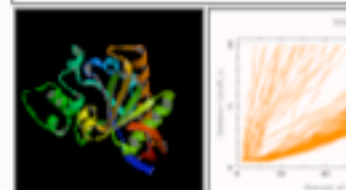
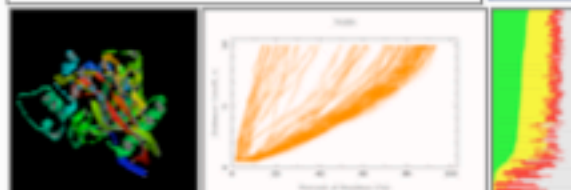
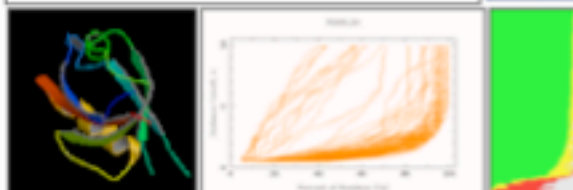
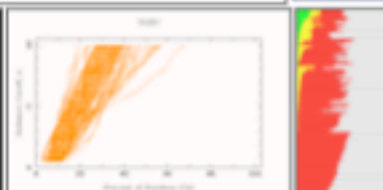
T0289 (TBM)
TABLES
(4-310)



T0289_D1 (TBM)
TABLES
(4-223, 298-310)



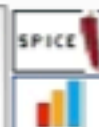
Structure
not
released
yet



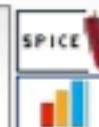
T0289_D2 (TBM)
TABLES
(224-297)



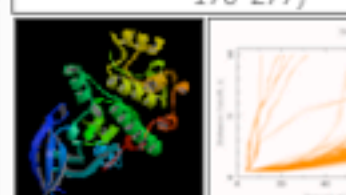
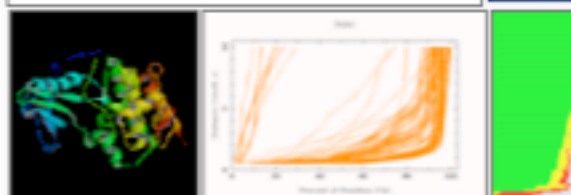
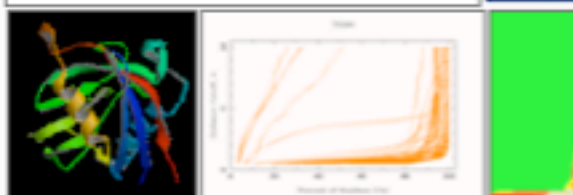
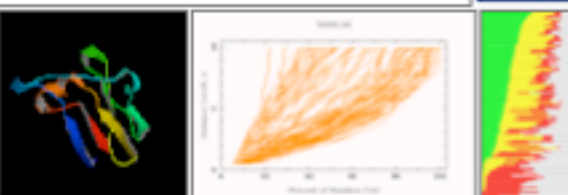
T0290 (TBM)
TABLES
(1-173)

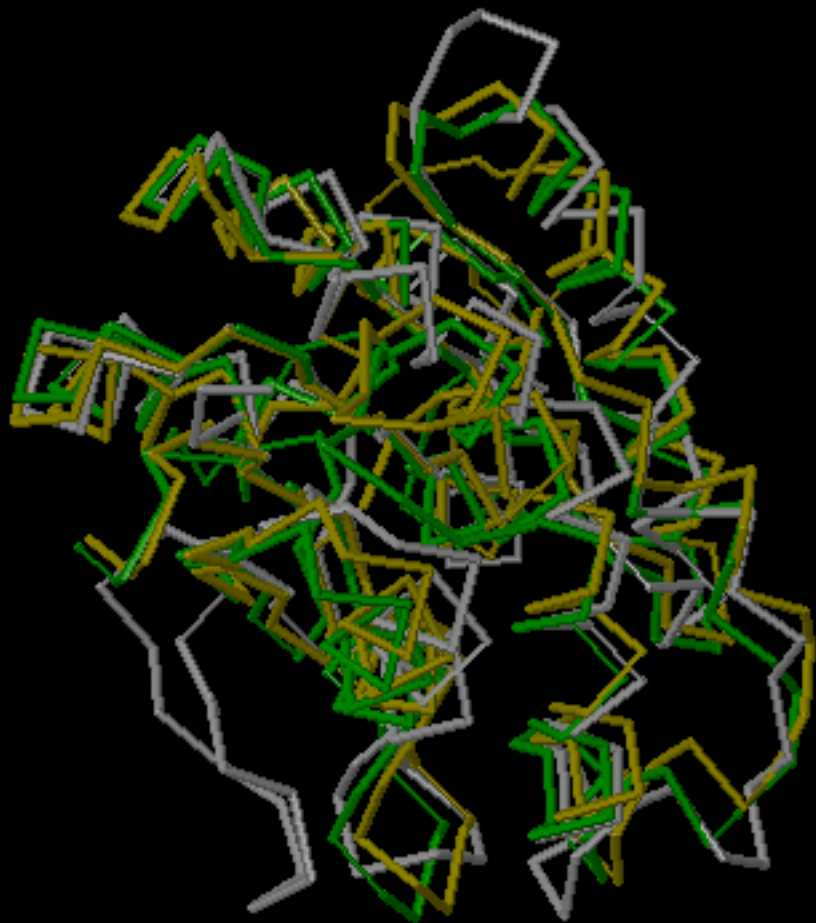


T0291 (TBM)
TABLES
(1-6, 15-176, 193-299, 305-310)



T0292 (TBM)
TABLES
(1-43, 49-74, 79-129, 176-277)





- 1 T0286
- 2 T0286TS556_5
- 3 T0286TS556_1
- 4 T0286TS556_2
- 5 T0286TS556_4
- 6 T0286TS556_3
- 7 T0286TS137_3
- 8 T0286TS020_2
- 9 T0286TS038_2
- 10 T0286TS275_
- 11 T0286TS125_
- 12 T0286TS675_
- 13 T0286TS658_
- 14 T0286TS248_
- 15 T0286TS136_
- 16 T0286TS125_
- 17 T0286TS675_
- 18 T0286TS675_
- 19 T0286TS664_
- 20 T0286TS568_
- 21 T0286TS274_
- 22 T0286TS710_
- 23 T0286TS136_
- 24 T0286TS038_

PDB T0286TS556_4.

100

lga dep
 high-distan
 medium-dis
 low-distan

lga indep
 medium-dis
 low-distan

dal featur
 DAL0
 DAL4
 DAL1

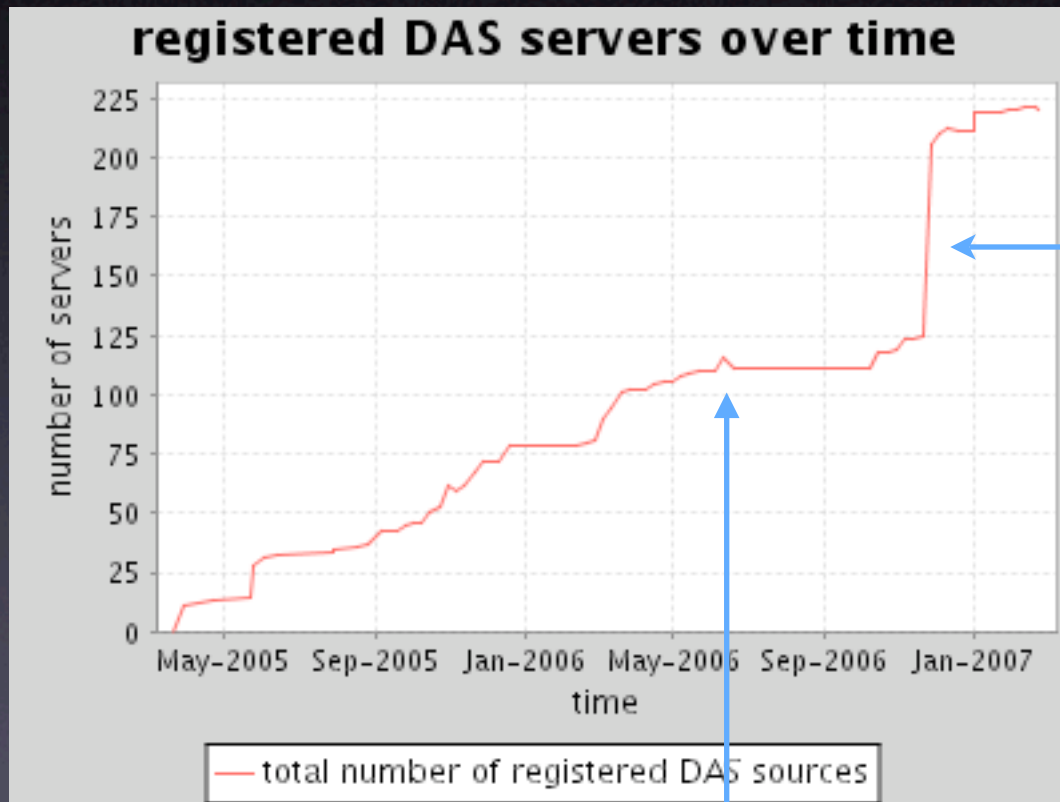
Shows:

- Targets
- Predictions
- closest know structures

support for
 different
 alignments

enter RASMOL like command...

DAS registration server



new Ensembl
DAS servers

removed dead servers

Web site,

DAS & SOAP style interfaces

<http://www.dasregistry.org>

Growth

- ENSEMBL will release many more DAS sources (tomorrow!)
- several 100s
- TODO: make sure data is well described and can be found

Coordinate Systems

- Authority

e.g. UniProt, NCBI

- Type

e.g. Protein sequence, Chromosome

- Organism (*optional*)

- (Assembly) Version (*optional*)

DAS usage

- das.sanger.ac.uk
(www.dasregistry.org, + DAS servers)
- 100.000 - 400.000 hits / week

Some new ideas

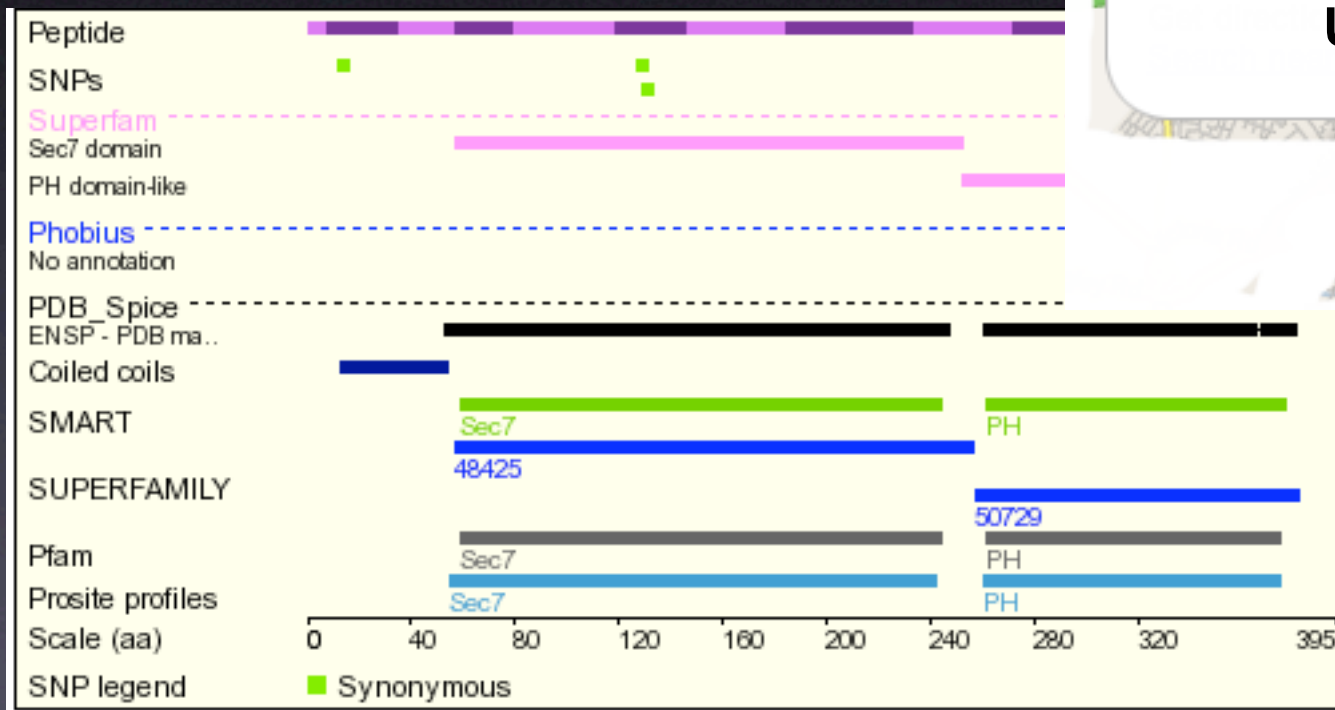
- Biological applications
- Social networking
- Technical improvements

Biological applications

- make the current clients better
- more applications needed
- new sequencing technologies / variation data

User rating

Did you find this
DAS source
useful?



★★★★☆ (38 customer reviews)

[4]

- User - rating of DAS servers?

pre-compiled views on the data

- look at the e! “groups”
- e.g. show all transmembrane / signal peptide predictions
- let users create these groups and share them with other users

Bridge to SOAP

- there is a big Web Service community out there
- add support for “come back later” response
 - calculate data on the fly
- e.g. for SPICE: trigger alignment searches, work with non-public data

application talks to application

- E.g SPICE - Jalview:
 - Jalview :“User selected residue 12 in UniProt P00123”
 - Spice -> shows selection
 - Spice -> “User selected active site annotation”
 - Jalview -> displays where it is in the alignment

Writeback

- write “my annotation” back to a server
- optional: keep information private or share with community
- -> user rating of individual annotations
- DAS/2

Mailing lists

- spice-das@sanger.ac.uk
- das_registry_announce@sanger.ac.uk
- <http://lists.sanger.ac.uk>

Acknowledgments

- T. Down, T. Hubbard
- eFamily Project
- R. Finn, E. Kulesha, A. Kahari, A. Andreeva
- Sanger/Ensembl Web Team. J Clements, H. R. Hotz
- Jmol, Biojava, MSD, everybody who sets up DAS servers