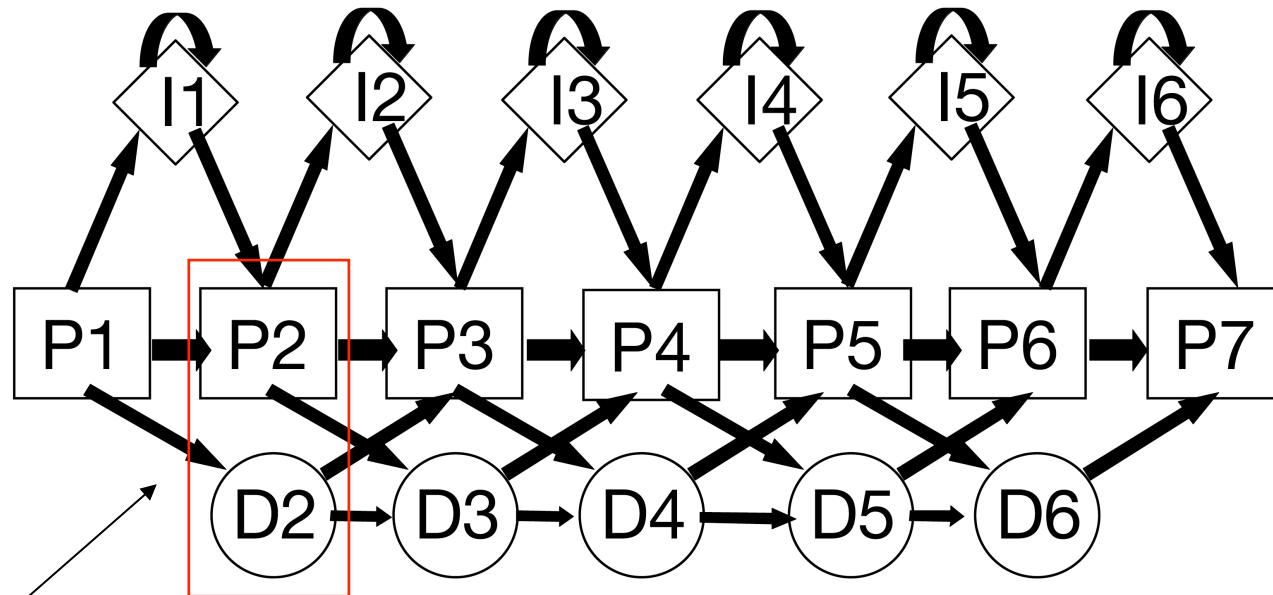


Profile HMM's



All P/D pairs must
be visited once

$$\begin{aligned}
 & L_1 - Y_2 A_3 V_4 R_5 - I_6 \\
 & P_1 D_2 P_3 P_4 I_4 P_5 D_6 P_7
 \end{aligned}$$

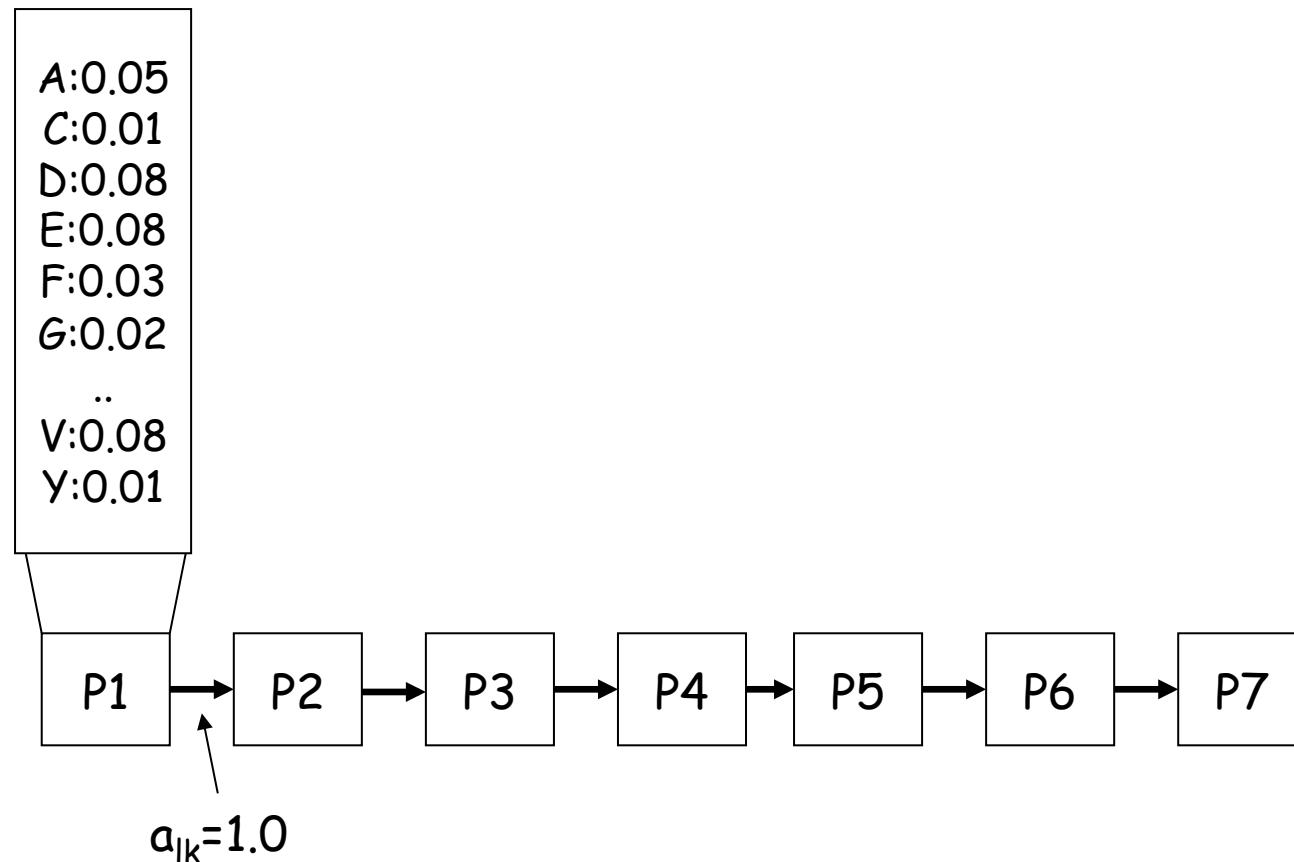
Profile HMM

- Un-gapped profile HMM is just a sequence profile



Profile HMM

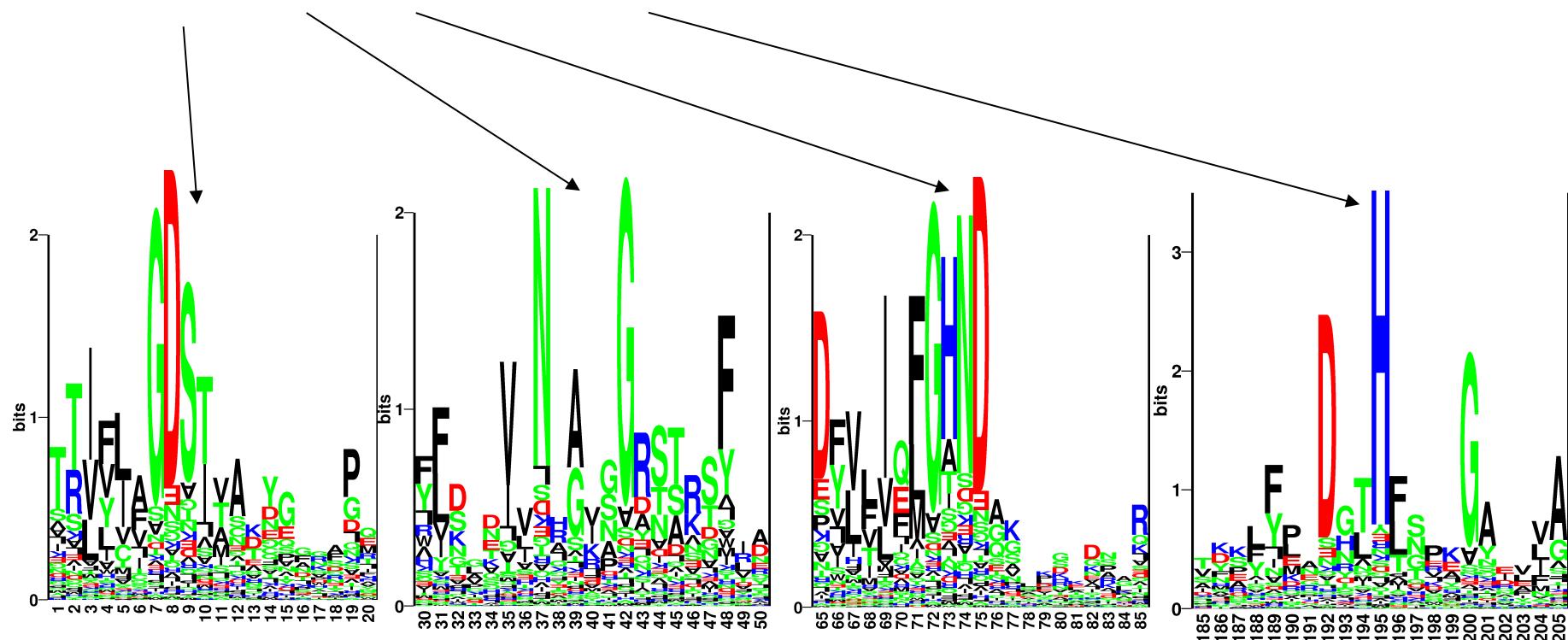
- Un-gapped profile HMM is just a sequence profile



CENTER FOR BIOLOGICAL
CALCULUS AND ANALYSIS CBS

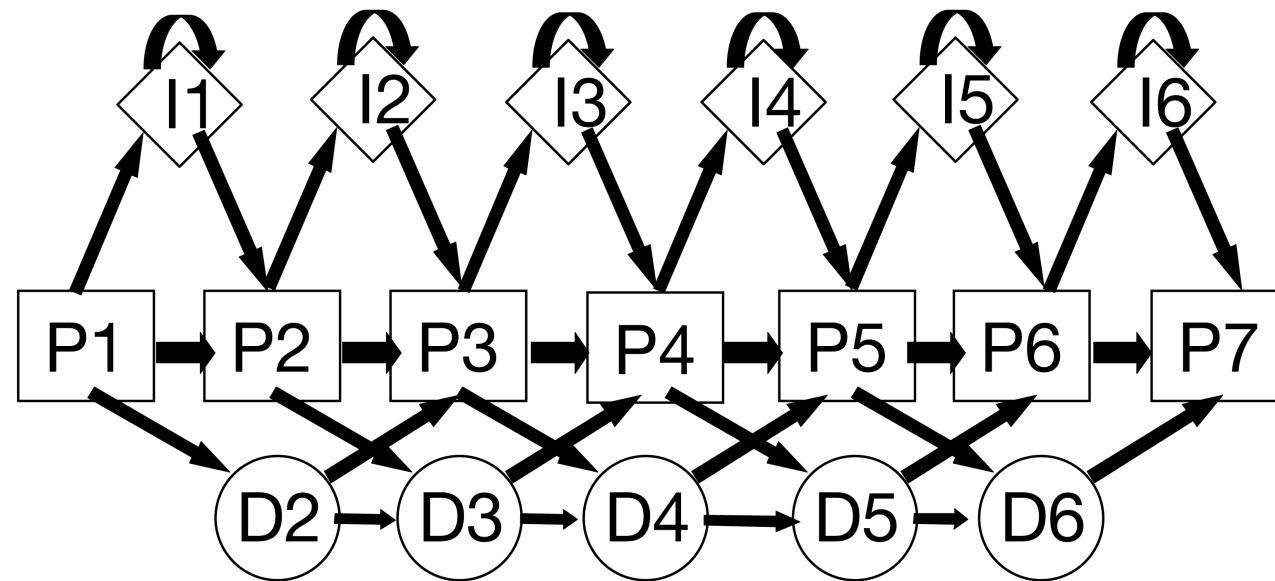
Example. Where is the active site?

- Sequence profiles might show you where to look!
 - The active site could be around
 - S9, G42, N74, and H195



Profile HMM

- Profile HMM (deletions and insertions)



Profile HMM (deletions and insertions)

QUERY	HAMDIRCYHSGG	PLHL	GEI	EDFNGQSCIVCPWHKYKITLATGE	GLY	QSINPKDPS
Q8K2P6	HAMDIRCYHSGG	PLHL	GEI	EDFNGQSCIVCPWHKYKITLATGE	GLY	QSINPKDPS
Q8TAC1	HAMDIRCYHSGG	PLHL	GDI	EDFDGRPCIVCPWHKYKITLATGE	GLY	QSINPKDPS
Q07947	FAVQDTCTHGDW	ALSE	GYI	DGD-----	VVECTLHFGKFCVRTGK	VKAL-----PA
P0A185	YATDNLCTHGSA	RMSD	GYI	EGRE-----	IECPLHQGRFDVCTGK	ALC-----APV
P0A186	YATDNLCTHGSA	RMSD	GYI	EGRE-----	IECPLHQGRFDVCTGK	ALC-----APV
Q51493	YATDNLCTHGAA	RMSD	GFI	EGRE-----	IECPLHQGRFDVCTGR	ALC-----APV
A5W4F0	FAVQDTCTHGDW	ALSD	GYI	DGD-----	IVECTLHFGKFCVRTGK	VKAL-----PA
P0C620	FAVQDTCTHGDW	ALSD	GYI	DGD-----	IVECTLHFGKFCVRTGK	VKAL-----PA
P08086	FAVQDTCTHGDW	ALSD	GYI	DGD-----	IVECTLHFGKFCVRTGK	VKAL-----PA
Q52440	FATQDQCTHGEW	SLSE	GGY	LDGD-----	VVECSLHMGKFCVRTGK	-----V
Q7N4V8	FAVDDRCSHGNA	SISE	GYI	ED-----	NATVECPLHTASFCLRTGK	ALCL-----PA
P37332	FATQDRCTHGDW	SLSDG	GYI	EGD-----	VVECSLHMGKFCVRTGK	-----V
A7ZPY3	YAINDRCSHGNA	SMSE	GYI	EDD-----	ATVECPLHAASFCLKTGK	ALCL-----PA
P0ABW1	YAINDRCSHGNA	SMSE	GYI	EDD-----	ATVECPLHAASFCLKTGK	ALCL-----PA
A8A346	YAINDRCSHGNA	SMSE	GYI	EDD-----	ATVECPLHAASFCLKTGK	ALCL-----PA
P0ABW0	YAINDRCSHGNA	SMSE	GYI	EDD-----	ATVECPLHAASFCLKTGK	ALCL-----PA
P0ABW2	YAINDRCSHGNA	SMSE	GYI	EDD-----	ATVECPLHAASFCLKTGK	ALCL-----PA
Q3YZ13	YAINDRCSHGNA	SMSE	GYI	EDD-----	ATVECPLHAASFCLKTGK	ALCL-----PA
Q06458	YALDNLEPGSEANVLSR	GLI	GDAGGEPIVISPLYKQRIRLRDG	-----	-----	-----



Core



Insertion



Deletion

The HMMer program

- HMMer is a open source program suite for profile HMM for biological sequence analysis
- Used to make the Pfam database of protein families
 - <http://pfam.sanger.ac.uk/>

A HMMer example

- Example from the CASP8 competition
- What is the best PDB template for building a model for the sequence T0391

```
>T0391 rieske ferredoxin, mouse, 157 residues
SDPEISEQDEEKKKYTSVCVGREEDIRKSERMTAVVHDREVVIFYHKGEYHAMDIRCYHS
GGPLHLGEIEDFNGQSCIVCPWHKYKITLATGEGLYQSINPKDPSAKPKWCSKGVKQRIH
TVKVDNGNIYVTLSKEPFKCDSDYYATGEFKVIQSSS
```

A HMMer example

- What is the best PDB template for building a model for the sequence T0391
- Use Blast
 - No hits
- Use Psi-Blast
 - No hits
- Use Hmmer

A HMMer example

- Use Hmmer
 - Make multiple alignment using Blast
 - Make model using
 - hmmbuild
 - hmmcalibrate
 - Find PDB template using
 - hmmsearch

A HMMer example

- Make multiple alignment using Blast

```
blastpgp -j 3 -e 0.001 -m 6 -i T0391.fsa -d sp -b  
10000000 -v 10000000 > T0391.fsa.blastout
```
 - Make Stockholm format

```
# STOCKHOLM 1.0  
QUERY DPEISEQDEEKKKYTSVCVGREEDIRKS-ERMTAVVHD-RE--V-V-IF--Y-H-KGE-Y  
Q8K2P6 DPEISEQDEEKKKYTSVCVGREEDIRKS-ERMTAVVHD-RE--V-V-IF--Y-H-KGE-Y  
Q8TAC1 ----SAQDPEKREYSSVCVGREDDIKKS-ERMTAVVHD-RE--V-V-IF--Y-H-KGE-Y
```
 - Build HMM

```
hmmbuild T0391.hmm T0391.fsa.blastout.sto
```
 - (Calibrate HMM (to estimate correct p-values)
hmmpcalibrate T0391.hmm) In older version of HMMER
 - Search for templates in PDB
hmmpsearch T0391.hmm pdb > T0391.out
-

A HMMer example

Sequence	Description	Score	E-value	N
2E4Q.A	mol:aa ELECTRON TRANSPORT	163.7	6.7e-45	1
2E4P.B	mol:aa ELECTRON TRANSPORT	163.7	6.7e-45	1
2E4P.A	mol:aa ELECTRON TRANSPORT	163.7	6.7e-45	1
2E4Q.C	mol:aa ELECTRON TRANSPORT	163.7	6.7e-45	1
2YVJ.B	mol:aa OXIDOREDUCTASE/ELECTRON TRANSPORT	163.7	6.7e-45	1
1FQT.A	mol:aa OXIDOREDUCTASE	160.9	4.5e-44	1
1FQT.B	mol:aa OXIDOREDUCTASE	160.9	4.5e-44	1
2QPZ.A	mol:aa METAL BINDING PROTEIN	137.3	5.6e-37	1
2Q3W.A	mol:aa ELECTRON TRANSPORT	116.2	1.3e-30	1
1VM9.A	mol:aa ELECTRON TRANSPORT	116.2	1.3e-30	1

Validation. CE structural alignment

CE 2E4Q A 3D89 A (run on IRIX machines at CBS)

Structure Alignment Calculator, version 1.01, last modified: May 25, 2000.

CE Algorithm, version 1.00, 1998.

Chain 1: /usr/cbs/bio/src/ce_distr/data.cbs/pdb2e4q.ent:A (Size=109)
Chain 2: /usr/cbs/bio/src/ce_distr/data.cbs/pdb3d89.ent:A (Size=157)

Alignment length = 101 Rmsd = 2.03A Z-Score = 5.5 Gaps = 20 (19.8%)
CPU = 1s Sequence identities = 18.1%

Chain 1: 2 TFTKACSVDEVPPGEALQVSHDAQKVAIFNVDGEFFATQDQCTHGEWSLSEGGYLDG----DVVECSLHM
Chain 2: 16 TSVCVGREEDIRKSERTAVVHDREVVIFYHKGEYHAMDIRCYHSGGPLH-LGEIEDFNGQSCIVCPWHK

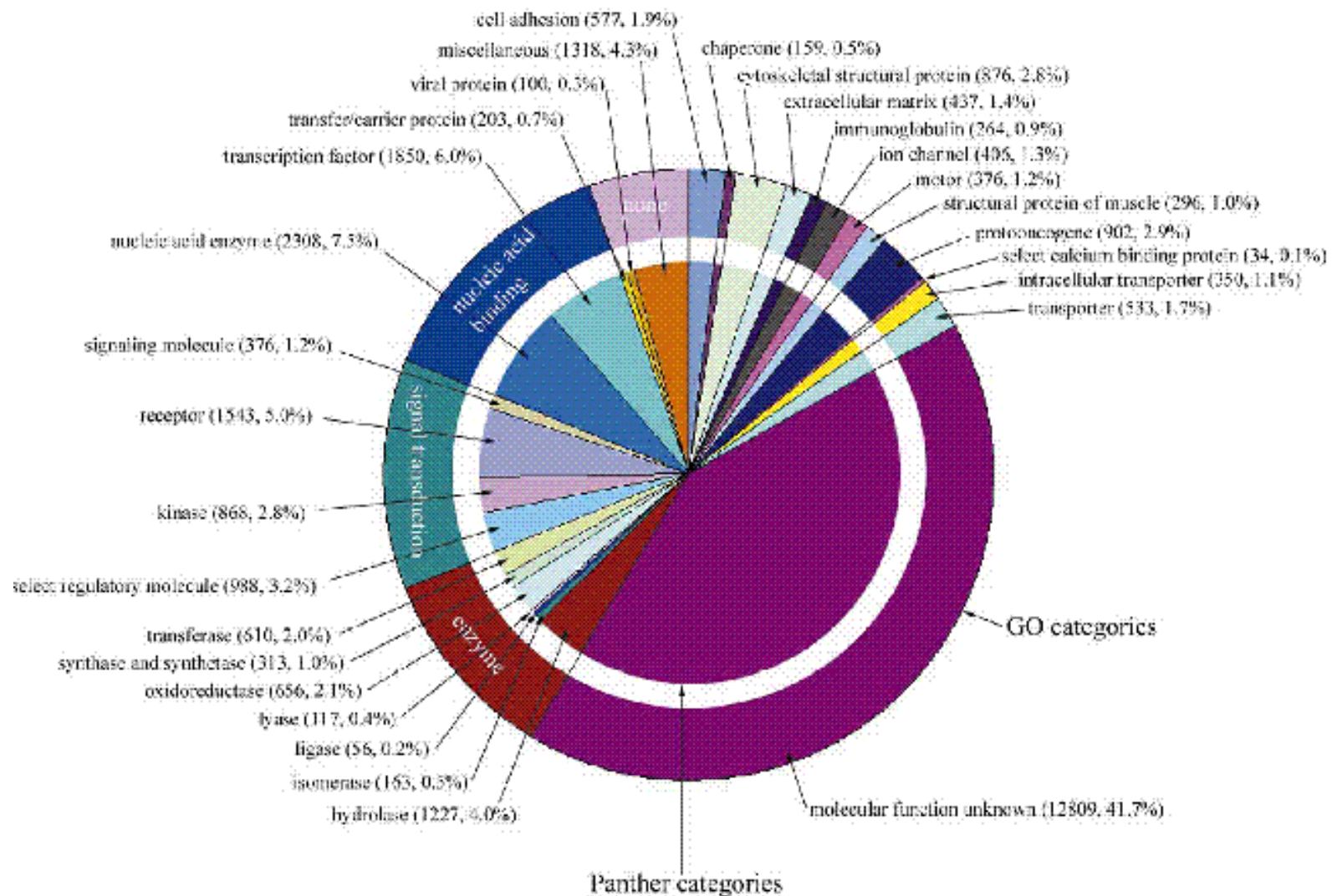
Chain 1: 68 GKFCVRTGKVKS-----PPPC-----EPLKVYPIRIEGRDVLVDFSRAALH
Chain 2: 85 YKITLATGEGLYQSINPKDPSAKPKWCSKGVKQRIHTVKVDNGNIYVTL-SKEPF

Protein homology modeling

Background. Why protein modeling?

- Because it works!
 - Close to 50% of all new sequences can be homology modeled
 - Experimental effort to determine protein structure is very large and costly
 - The gap between the size of the protein sequence data and protein structure data is large and increasing
-

Homology modeling and the human genome



How can we do it?

- Identify template(s) - initial alignment
 - Can give you the protein function
- Improve alignment
 - Can give you the active site
 - Backbone generation
- Loop modeling
 - Most difficult part
- Side chains
- Refinement
- Validation

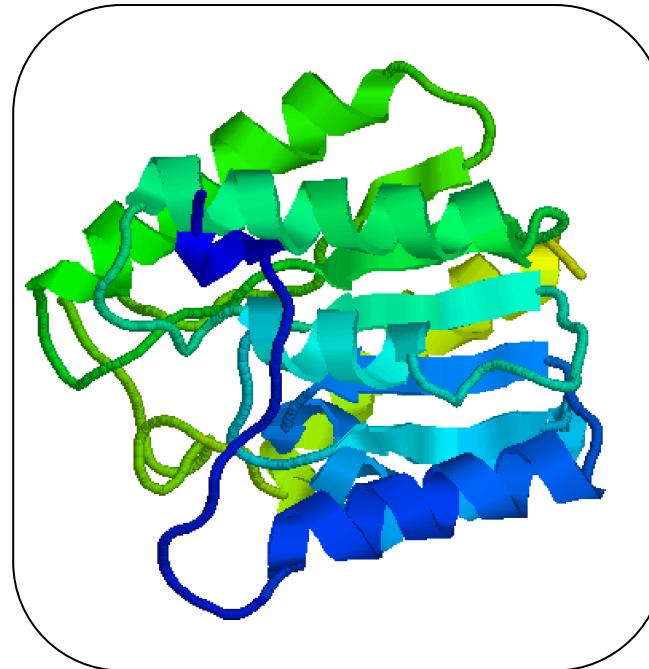
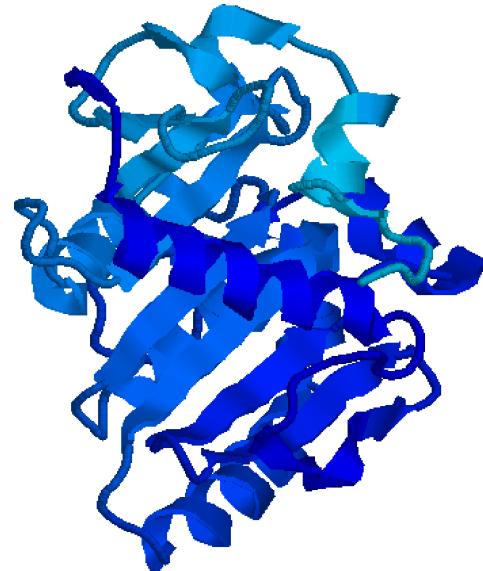
How to do it?

- Identify fold (template) for modeling
 - Find the structure in the PDB database that resembles your query protein the most
- Can be used to predict function
 - and maybe active sites
- Align protein sequence to template
 - Simple alignment methods
 - Sequence profiles
- Model side chains and loops

How to do it?

>1K7C.A

TTVYLAGDSTMAKNGGGSGTNGWGEYLASYLSATVVNDAVAGRSARSYTREGRFENIADV
VTAGDYVIVEFGHNDGGSLSTDNGRTDCSGTGAEVCYSVYDGVNELITFPAYLENAAKL
FTAKGAKVILSSQTPNNPWETGTFVNSPTRFVEYAAEVAGVEYVDHWSYVDSIYETL
GNATVNSYFPIDHTHTSPAGAEVVAEFLKAVVCTGTSLKSVLTTTSFEGTCL



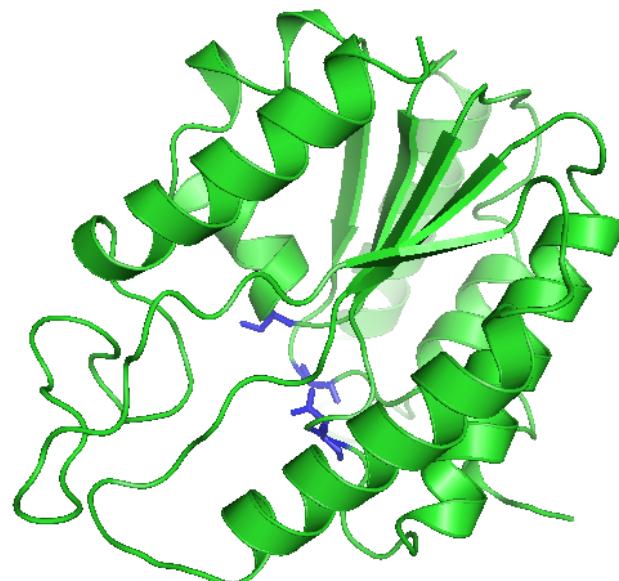
How to do it?

>1K7C.A

TTVYLAGDSTMAKNGGGSGTNGWGEYLASYLSATVVNDAVAGRSARSYTREGRFENIADV
VTAGDYVIVEFGHNDGGSLSTDNGRTDCSGTGAEVCYSVYDGVNTELTFPAYLENAAKL
FTAKGAKVILSSQTPNNPWETGTFVNSPTRFVEYAAEVAGVEYVDHWSYVDSIYETL
GNATVNSYFPIDHTHTSPAGAEVVAEAFLKAVVCTGTSLKSVLTTTSFEGTCL

1K7C . A TTVYLAGD**S**TMAKNGGGSGTNGWGEYLASYLSATVVNDAV**A**GRSARSYTREGRFENIADV**V**TAGDYVIVEFGH**N**
S G N

1WAB . _ EVVFFIGD**S**LVQLMHQCE---IWRELFS---PLHALNFGIG**G**DSTQHV**L**W--RLENG**E**LEHIRPKIVVVVWGT**N**



Is it really impossible?

Protein homology modeling is only possible
if %id greater than 30-50%

WRONG

Why %id is so bad!!

1200 models sharing 25-95% sequence identity with the submitted sequences (www.expasy.ch/swissmod)

Probabilities of SWISS-MODEL accuracy for target-template identity classes.

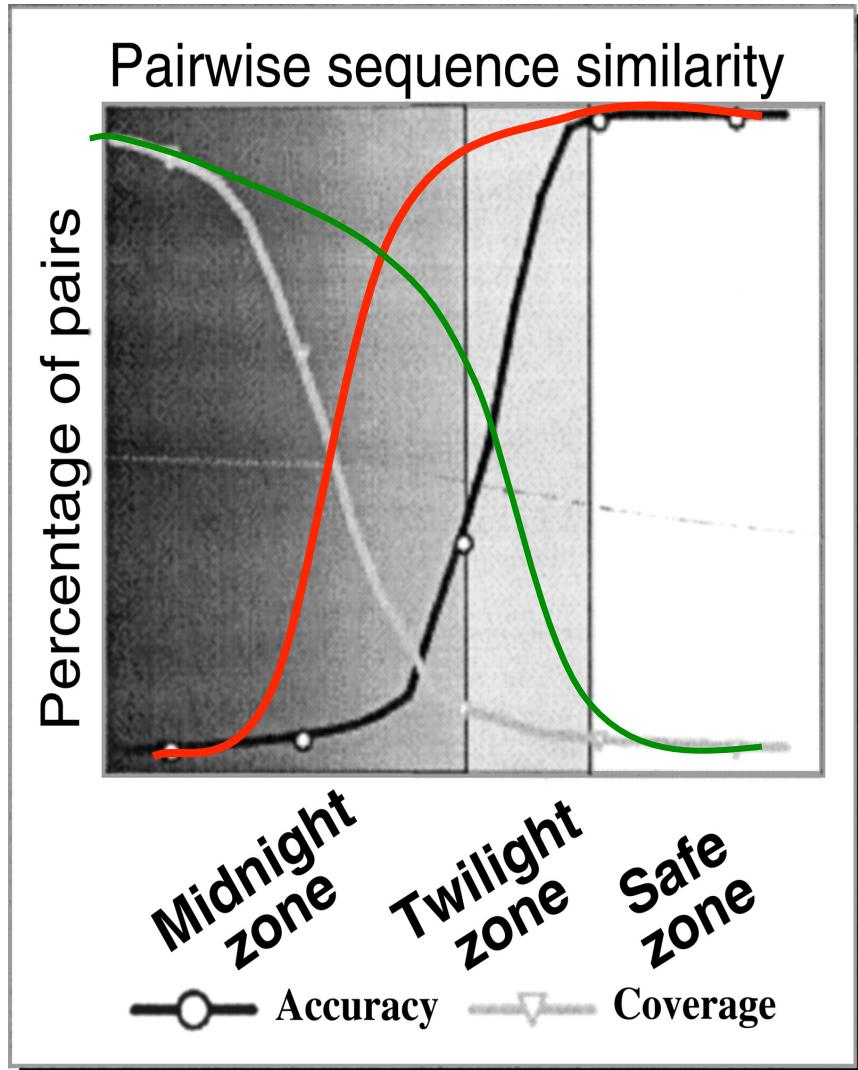
Percent sequence identity ^a	Total number of models ^b	Percent models with rmsd lower than 1 Å	Percent models with rmsd lower than 2 Å	Percent models with rmsd lower than 3 Å	Percent models with rmsd lower than 4 Å	Percent models with rmsd lower than 5 Å	Percent models with rmsd higher than 5 Å
25-29	125	0	10	30	46	67	33
30-39	222	0	18	45	66	77	23
40-49	156	9	44	63	78	91	9
50-59	155	18	55	79	86	91	9
60-69	145	38	72	85	91	92	8
70-79	137	42	71	82	85	88	12
80-89	173	45	79	86	94	95	5
90-95	88	59	78	83	86	91	9

a: Range of sequence identity between target and template sequence.

b: Total number of models in any given class of sequence identity. The table summarises 1201 model – control structure pairs.

c: Probability in percent that a model, sharing X% sequence identity with its template, deviates by 1 Å or less from the corresponding experimental control structure. The following columns provide these probabilities for other rms deviations.

Identification of fold



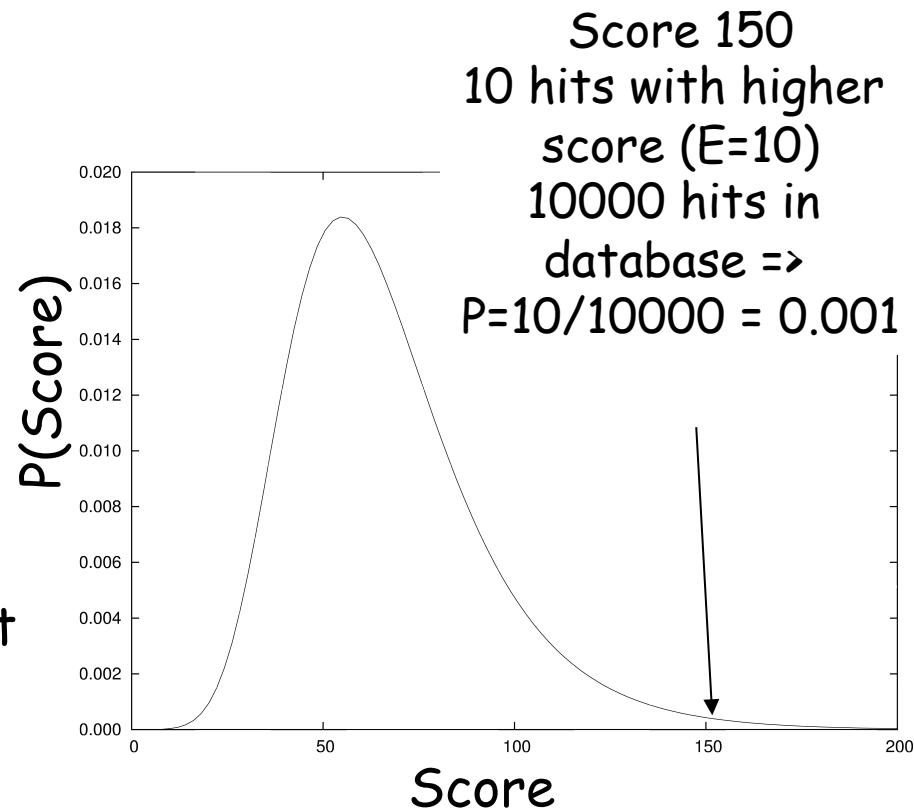
- If sequence similarity is high, proteins share structure (Safe zone)
- If sequence similarity is low, proteins may share structure (Twilight zone)
- Most proteins do not have a high sequence homologous partner

Identification of correct fold

- % ID is a poor measure
 - Many evolutionary related proteins share low sequence homology
 - A short alignment of 5 amino acids can share 100% id, what does this mean?
 - Alignment score even worse
 - Many sequences will score high against every thing (hydrophobic stretches)
 - P-value or E-value more reliable
-

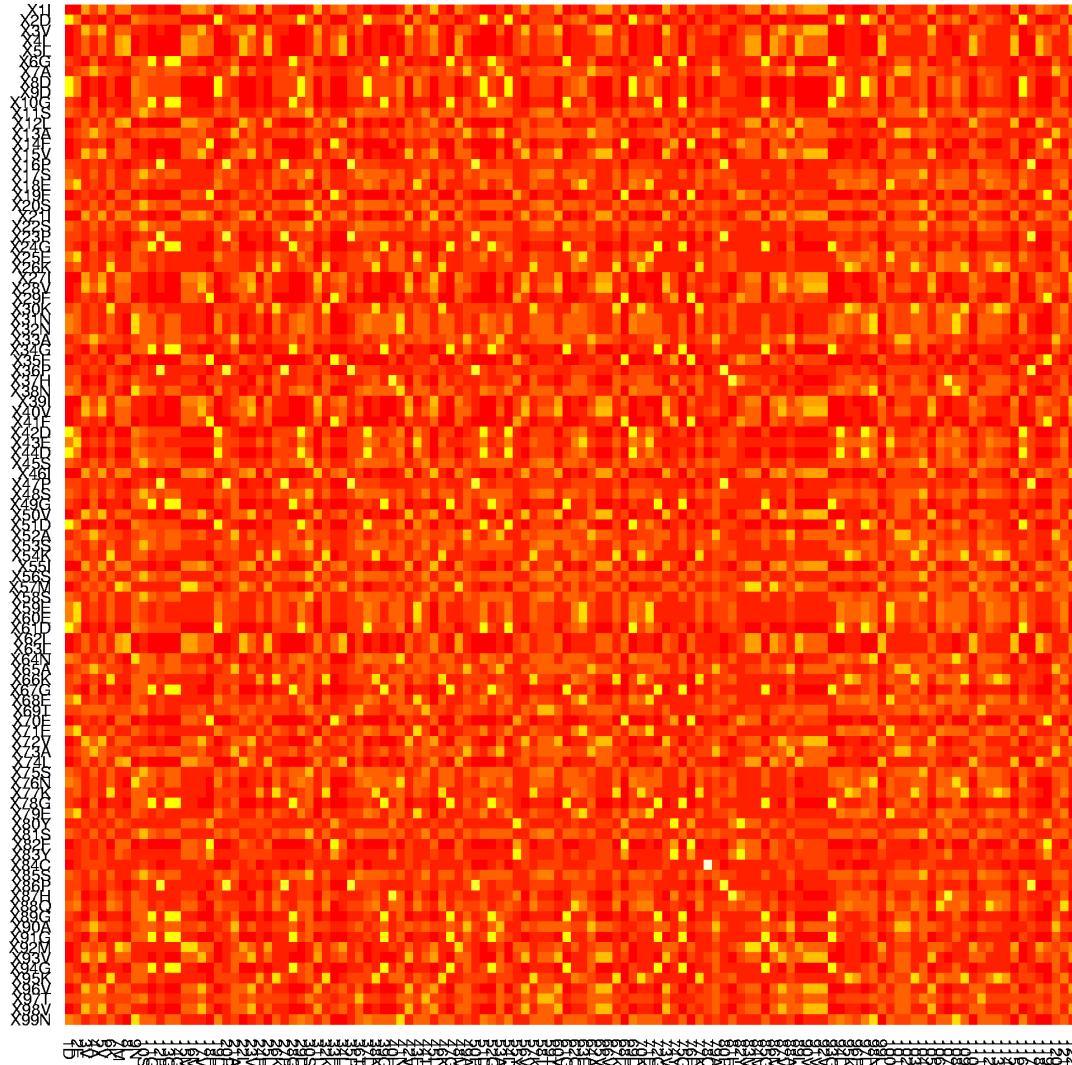
What are P and E values?

- E-value
 - Number of expected hits in database with score higher than match
 - Depends on database size
- P-value
 - Probability that a random hit will have score higher than match
 - Database size independent



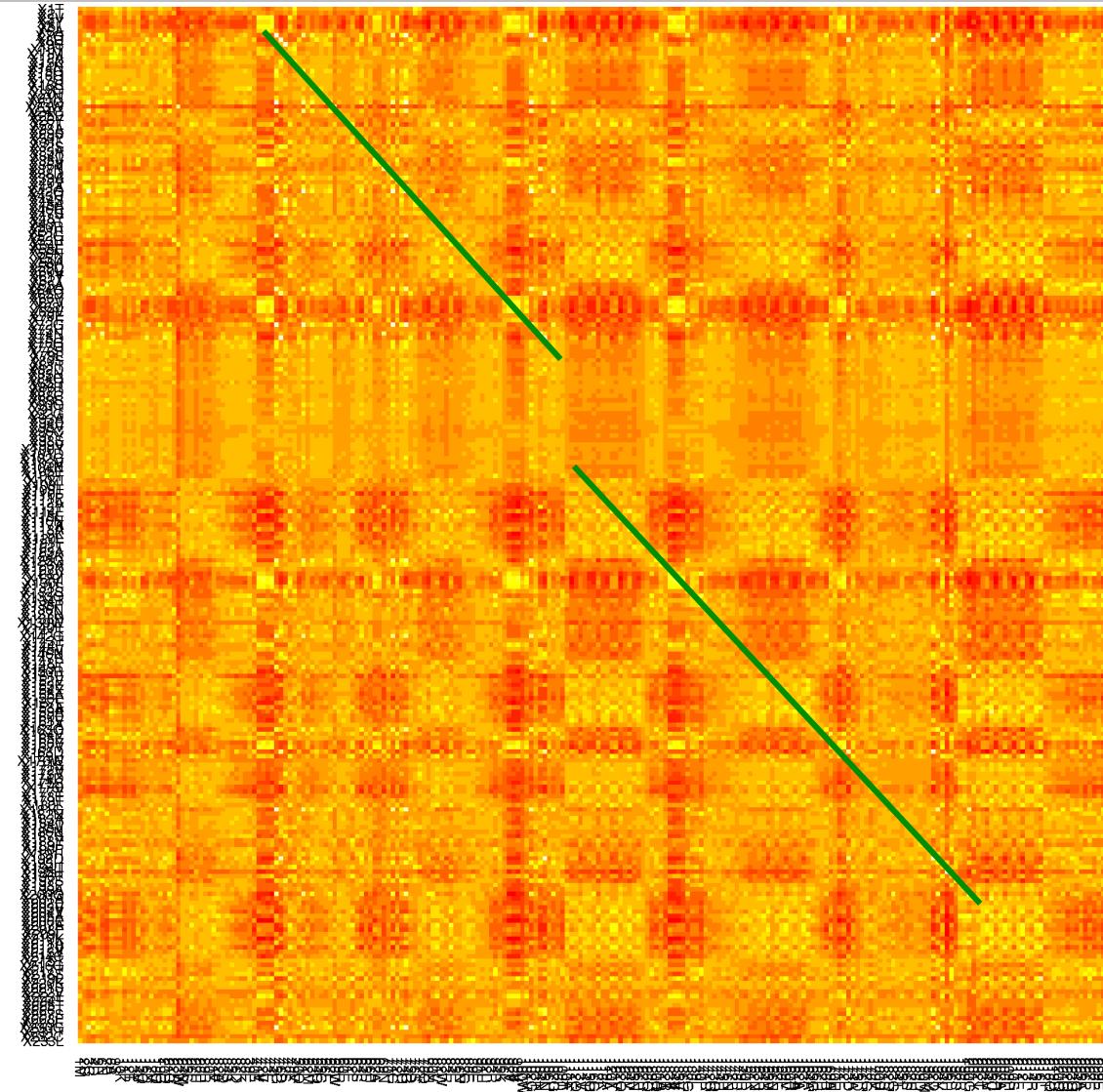
When Blast fails!

1K7A.A



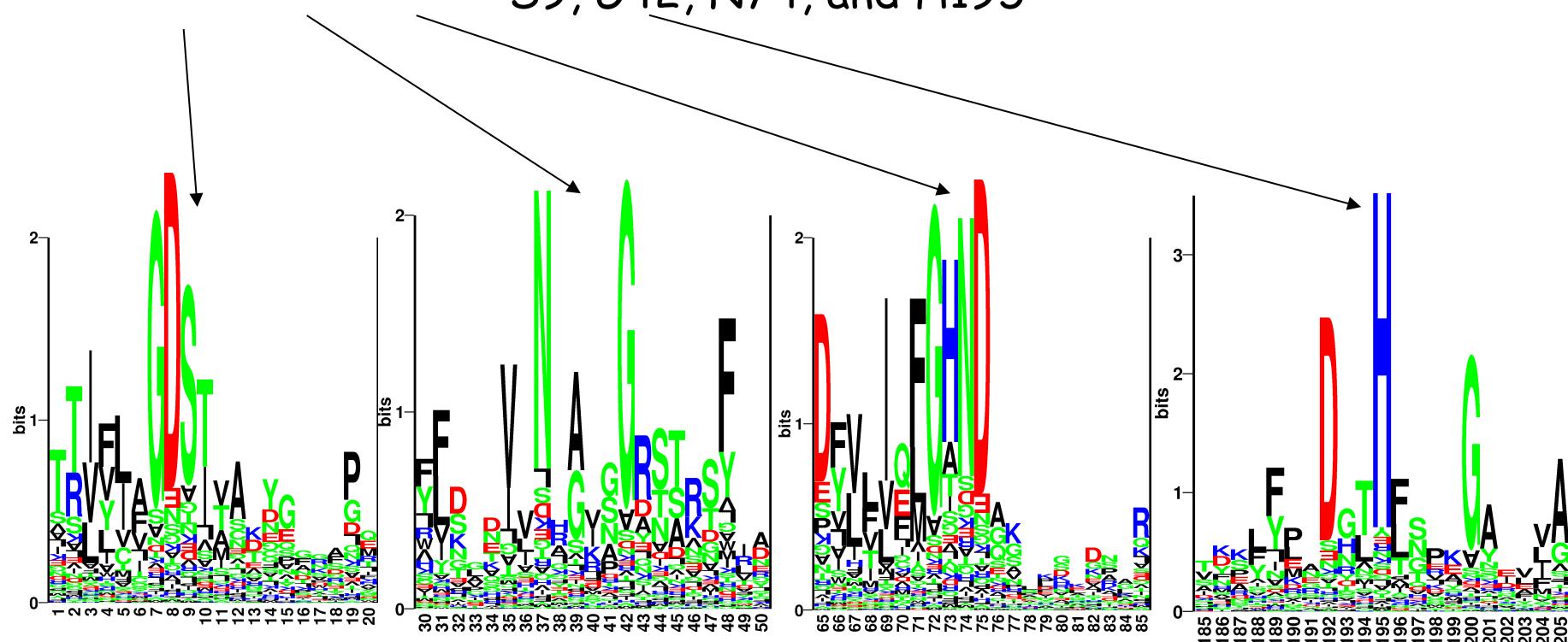
1WAB._

Profile-profile scoring matrix



Example. Where is the active site?

- Sequence profiles might show you where to look!
 - The active site could be around
 - S9, G42, N74, and H195



Example. Where is the active site?

Align using sequence profiles

ALN 1K7C.A 1WAB. RMSD = 5.29522. 14% ID

1K7C.A TVYLAGD**S**TMAKNGGGSGTNGGEYLASYLSATVVNDAVA**G**RSARSYTREGRFENIADVVTAGDYVIVEFGH**N**DGGSLSTDN
1WAB. _ EVVFIGD**S**LVQLMHQCE---IWRELFS---PLHALNFGIGG**G**DSTQHVLW--RLENGELEHIRPKIVVVVWGT**N**NHG-----

1K7C.A GRTDCSGTGAEVCYSVYDGNETILTFPAYLEAAKLFTAK--GAKVILSSQTPNNPWETGTFVNSPTRFVEYael-AAEVA
1WAB. _ -----HTAEQVTGGIKAIQLVNERQPQARVVVLGLLPRGQ-HPNPLREKNRRVNELVRAALAGHP

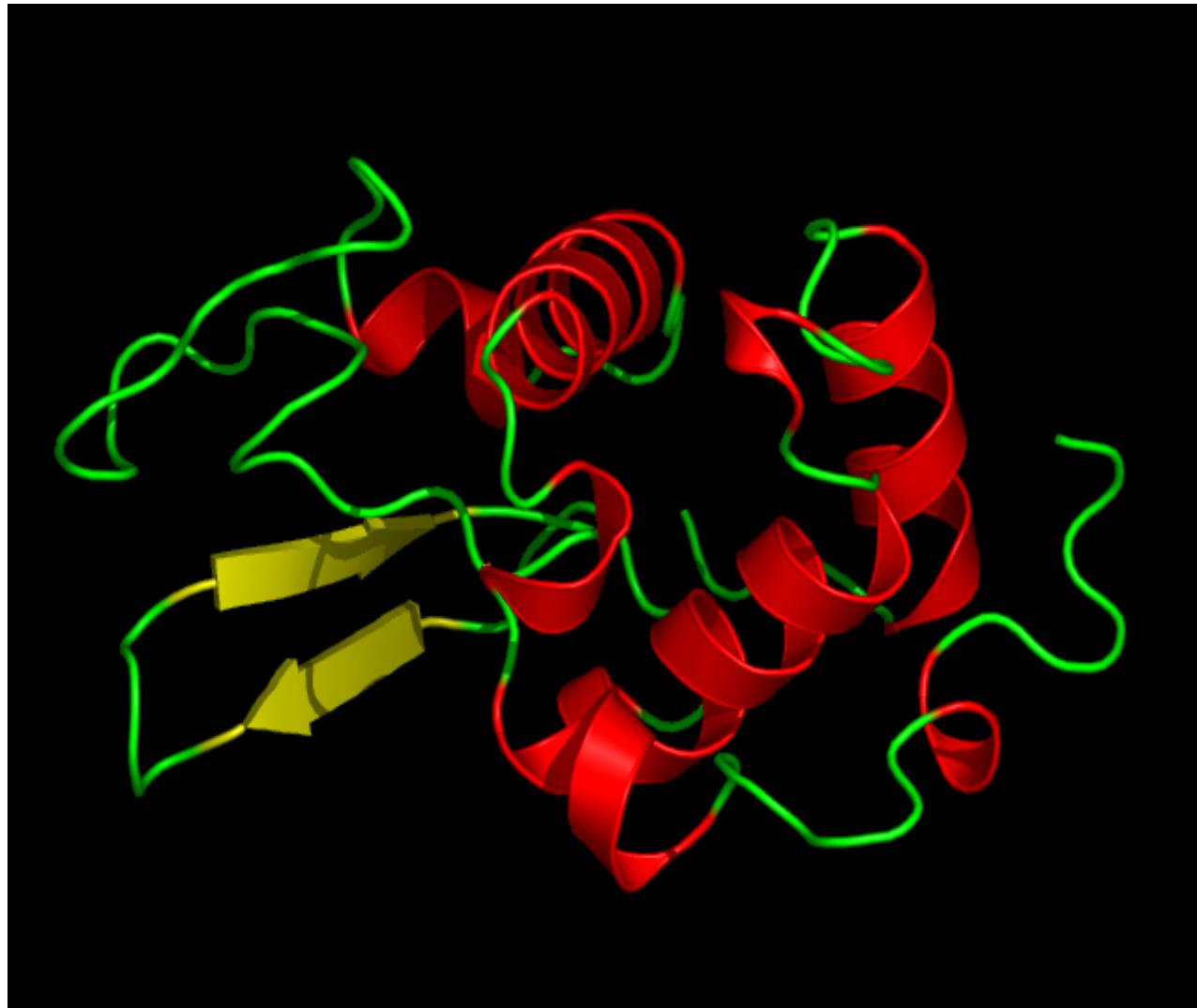
1K7C.A GVEYVDHWSYVDSIYETLGNATVNSYFPIDHT**H**TSPAGAEVVAEAFLKAVVCTGTSL
1WAB. _ RAHFILDADPG---FVHSDG--TISHHDMDYL**H**LSRLGYTPVCRALHSLLLRL---L

Including structure

- Sequence within in a protein superfamily share remote sequence similarity
 - , but they share high structural similarity
 - Structure is known for template
 - Predict structural properties for query
 - Secondary structure
 - Surface exposure
 - Position specific gap penalties derived from secondary structure and surface exposure
-

Predicting local Protein Structure

$\alpha+\beta$: Lysozyme (1jsf)



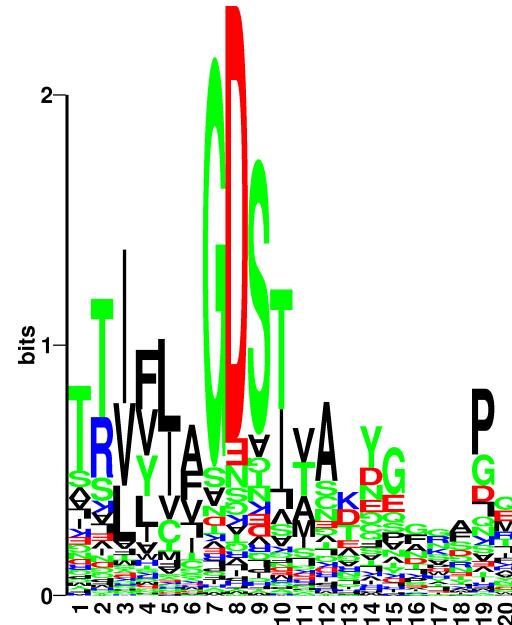
Improvement of accuracy

1974 Chou & Fasman	~50-53%
1978 Garnier	63%
1987 Zvelebil	66%
1988 Quian & Sejnowski	64.3%
1993 Rost & Sander	70.8-72.0%
1997 Frishman & Argos	<75%
1999 Cuff & Barton	72.9%
1999 Jones	76.5%
2000 Petersen et al.	77.9%

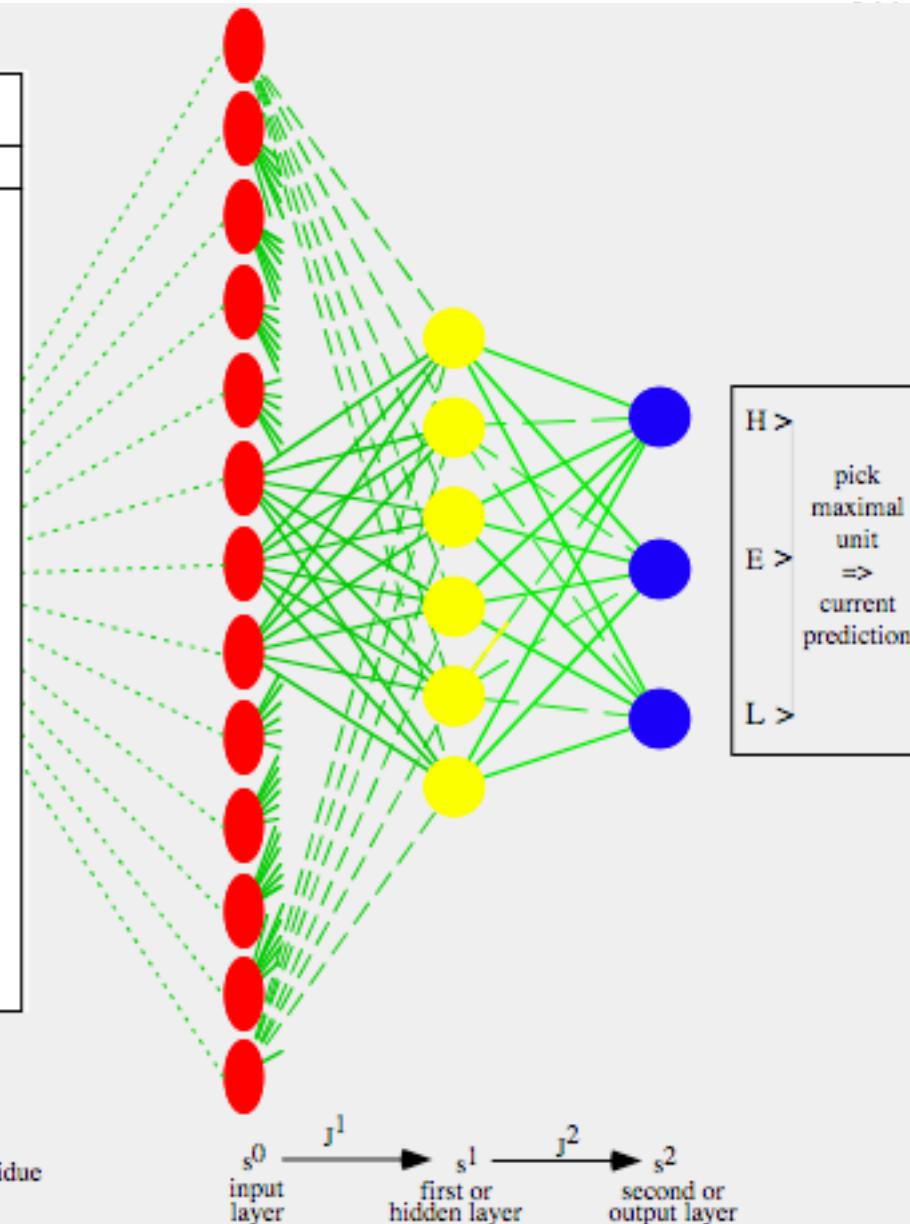
Sequence profiles



	1	50
fyn_human	VTLFVALYDY EARTEDDLSF HKGEKFQILN SSEGDWWEAR SLTTGETQVI	
yrk_chick	VTLFIALYDY EARTEDDLSF QKGEKFHIIN NTEGDWWEAR SLSSGATQVI	
fgr_human	VTLFIALYDY EARTEDDLTF TKGEKFHILN NTEGDWWEAR SLSSGKTQCI	
yes_chick	VTVFVALYDY EARTTDDLSF KKGERFQIIN NTEGDWWEAR SIATGKTQVI	
src_avis2	VTTFVALYDY ESRTETDLSF KKGERIQIVN NTEGDWWLAH SLTTGQTQVI	
src_aviss	VTTFVALYDY ESRTETDLSF KKGERIQIVN NTEGDWWLAH SLTTGQTQVI	
src_avisr	VTTFVALYDY ESRTETDLSF KKGERIQIVN NTEGDWWLAH SLTTGQTQVI	
src_chick	VTTFVALYDY ESRTETDLSF KKGERIQIVN NTEGDWWLAH SLTTGQTQVI	
stk_hydat	VTIFVALYDY EARISEDLSF KKGERIQIIN TADGDWWYAR SLTNSEGQVI	
src_rsvpa ESRIETDLSF KKRERIQIVN NTEGTWWLAH SLTTGQTQVI	
hck_human	..IVVALYDY EAIHHEDLSF QKGDQM/VLE ES.GEWWKAR SLATRKEQVI	
blk_mouse	..FVVALFDY AAVNDRDLQV LKGEKLQVLR .STGDWWLAR SLVTGREGQV	
hck_mouse	.TIVVALYDY EAIHREDLSF QKGDQM/VLE .EAGEWWKAR SLATKKEQVI	
lyn_human	..IVVALYPY DGIHPDDLSF KKGEKMVKLE .EHGEWWKAK SLLTKKECFI	
lck_human	..LVIALHSY EPSHDGDLGF EKGEQIRILE QS.GEWWKAQ SLTTGQEAFI	
ss81_yeast ALYPY DADDDeISF EQNEILQVSD .IEGRWWKAR R.ANGETGII	
abl_mouse	..LFVALYDF VASGDNTLSI TKGEKIRVLG YnnGEWCEAQ ..TKNGQQNV	
abl1_human	..LFVALYDF VASGDNTLSI TKGEKIRVLG YnnGEWCEAQ ..TKNGQQNV	
src1_drome	..VVSLYDY KSRDESDLNF MKGDRMEVID DTESDWWRRVV NLTRRQECLI	
mysd_dicdi ALYDF DAESSMELSF KEGDIITVLD QSSGDWWDAE L..KGRRKV	
yfj4_yeastVALYSF AGEESGDLPF RKGDVITILK ksQNDWWTGKV ..NGREGIF	
abl2_human	..LFVALYDF VASGDNTLSI TKGEKIRVLG YNQNGEWSEV RSKNG.QGV	
tec_human	.EIVVAMYDF QAAEGHDLRL ERGQEYLILE KNDVHWWRAR D.KYGNEGQVI	
abl1_caeel	..LFVALYDF HGVGEEQLSL RKGDQVRILG YNKNNEWCEA R1rLGEIGNW	
txk_humanALYDF LPREPCLNL RRAEELYLILE KYNPHWWKAR D.RLGNEGLI	
yha2_yeast	VRRVVALYDL TTNEPDELSF RKGDVITVLE QVYRDWWKGAL..RGNMGIF	
abp1_sacexAEYDY EAGEDNELTF AENDKIINIE FVDDDWLGE LETTGQKGLF	



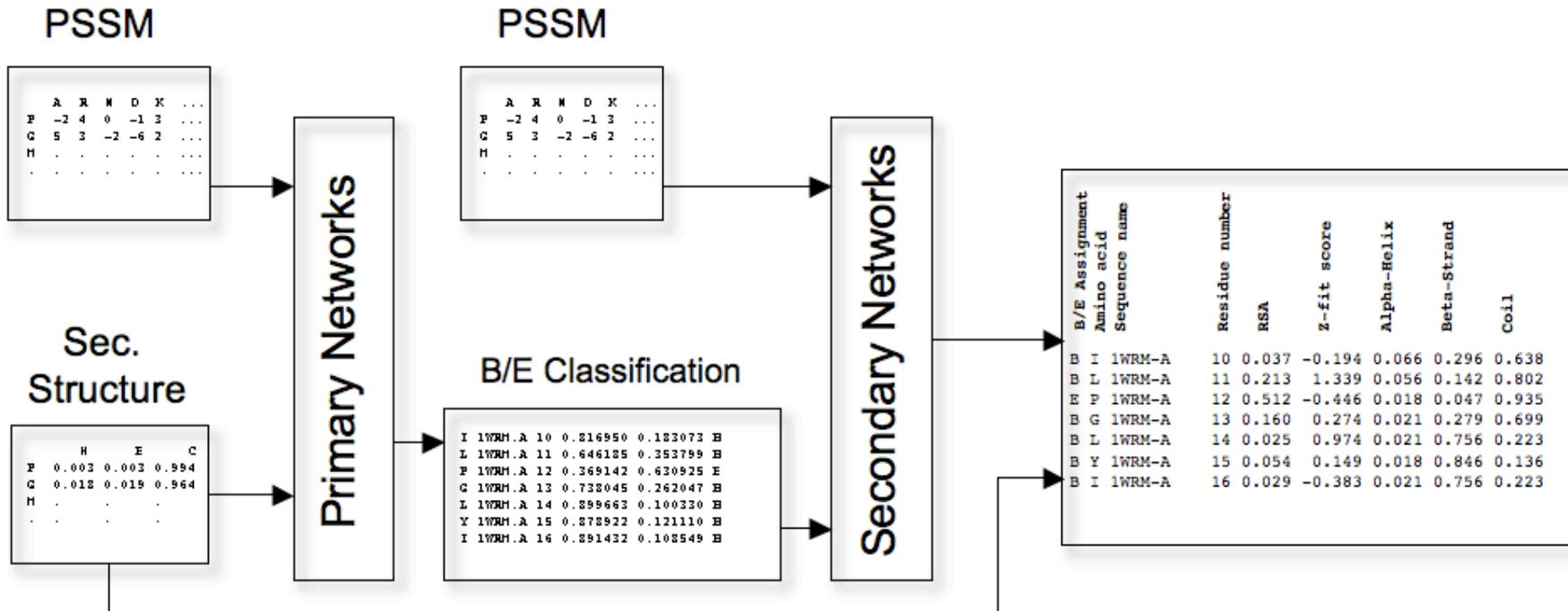
Protein	Alignments	profile table
:	:	GSAPD NTEKQ CVHIR LMYFW
G	GGGG	5.....
Y	YYYY5..
I	IIEE2...3..
Y	YYYY5..
D	DDDD5.....
P	PPPP	...5.....
E	AEAA	..3..2..
D	VVEE1..2..2..
G	GGGG	5.....
D	DDDD5.....
P	PPPP	...5.....
D	DTDD4..1..
D	NQNN1..3...1..
G	GNGG	4....1....
V	VIVV4..1..
N	EPKK1..1..2..
P	PPPP5.....
G	GGGG	5.....
T	TTTT5.....
D	EKSA	..11..1..11..
F	FFFF5..
:	:



corresponds to the the 21×3 bits coding for the profile of one residue

NetSurfP (surface exposure and SS structure)

CENTERFO
RBILOGI
CALSEQU
ENCEANA
LYSIS CBS
CENTERFO
RBILOGI
CALSEQU
ENCEANA
LYSIS CBS



NetSurfP

NetSurfP ver. 1.1 - Protein Surface Accessibility and Secondary Structure Predictions

Version 1.1: Caching system has been implemented.

View the [version history](#) of this server.

NetSurfP server predicts the surface accessibility and secondary structure of amino acids in an amino acid sequence.

The method also simultaneously predicts the reliability for each prediction, in the form of a Z-score. The Z-score is related to the surface prediction, and not the secondary structure.

Instructions

Output format

Dear user, we want your feedback !

Please fill out this two minute questionnaire [Click here to fill in the questionnaire \(opens a new window\)](#)

Speed: Approx 5 min per sequence, but much faster if sequence has been cached previously.

Paste in sequence data (maximum 2000 sequences)

```
>1K7C.A
TTVYLAGDSTMKNCGGSGTNCWGEYLASYLSATVVNDAVAGRSARSYTREGRFENIADV
VTAGDYVIVEFGHNDGGSLSTDNGRTDCSGTGAEVCYSVYDGVNTELTFPAYLENAAKL
FTAKGAKVILSSQTPNNPWETGTFVNSPTRFVEYELAAEVAGVEYVDHWSYVDSIYETL
GNATVNSYFPIDHTHTSPAGAEVVAEFLKAVVCTGTSKSVLTTTSFEGTCL
```

or upload sequence data

Choose File no file selected

Valid format examples: [Fasta](#)

All sequences must be submitted in amino acid format and have a unique sequence id!

Do not cache sequence profile and prediction results.

Submit query

Clear fields

NetSurfP - Protein Surface Accessibility and Secondary Structure Predictions

Technical University of Denmark

```
# For publication of results, please cite:  
# A generic method for assignment of reliability scores applied to solvent accessibility predictions.  
# Bent Petersen, Thomas Nordahl Petersen, Pernille Andersen, Morten Nielsen and Claus Lundegaard  
# BMC Structural Biology 2009, 9:51 doi:10.1186/1472-6807-9-51  
#  
# Column 1: Class assignment - B for buried or E for Exposed - Threshold: 25% exposure, but not based on RSA  
# Column 2: Amino acid  
# Column 3: Sequence name  
# Column 4: Amino acid number  
# Column 5: Relative Surface Accessibility - RSA  
# Column 6: Absolute Surface Accessibility  
# Column 7: Z-fit score for RSA prediction  
# Column 8: Probability for Alpha-Helix  
# Column 9: Probability for Beta-strand  
# Column 10: Probability for Coil
```

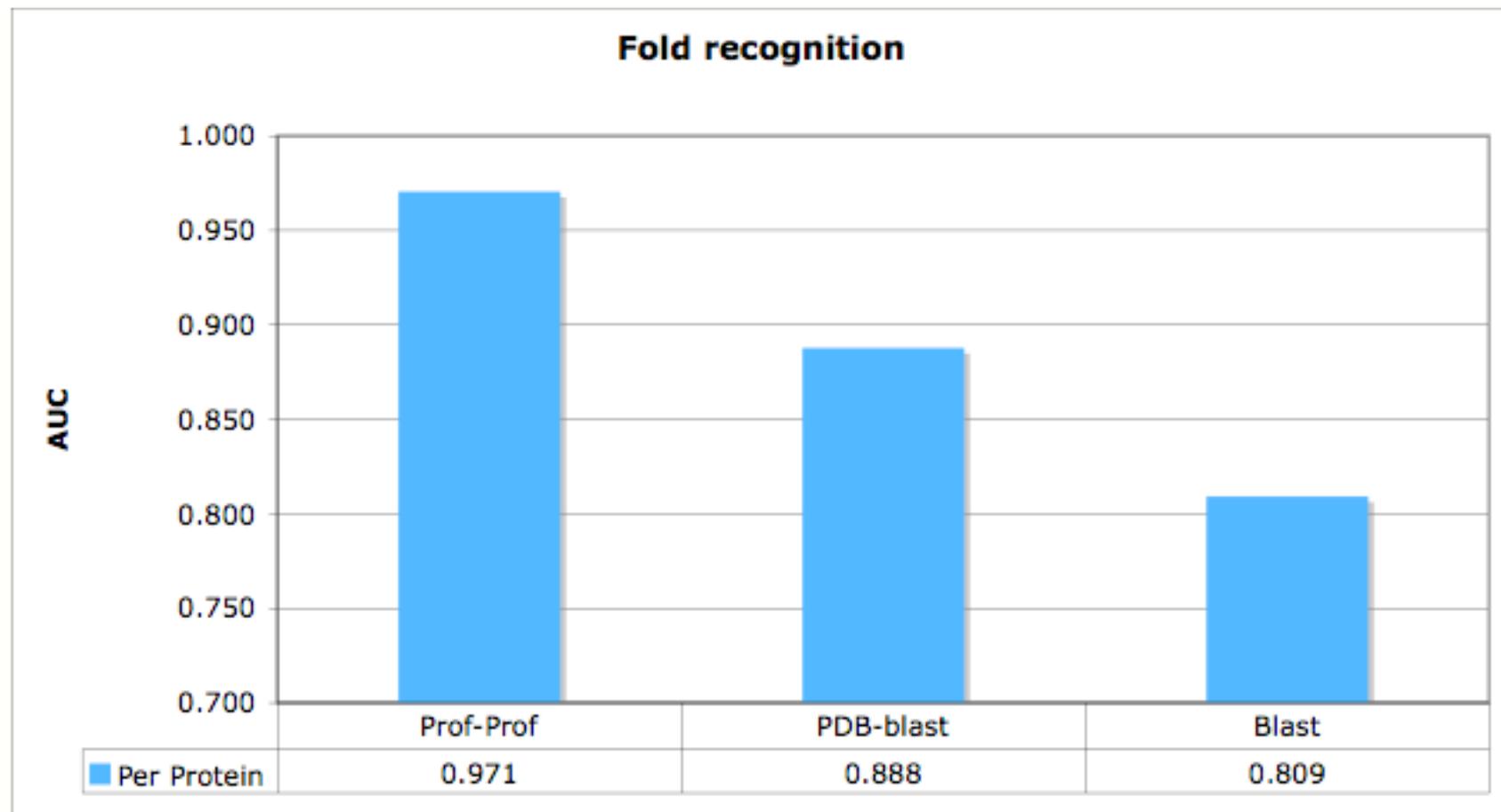
		H/E/C						
E T	1K7C.A	1	0.519	72.041	0.653	0.003	0.003	0.994
B T	1K7C.A	2	0.179	24.883	1.344	0.001	0.900	0.099
B V	1K7C.A	3	0.036	5.595	0.270	0.000	0.983	0.017
B Y	1K7C.A	4	0.037	7.992	0.516	0.000	0.983	0.017
B L	1K7C.A	5	0.021	3.808	0.931	0.000	0.983	0.017
B A	1K7C.A	6	0.024	2.634	1.003	0.001	0.959	0.040
B G	1K7C.A	7	0.028	2.188	-0.279	0.003	0.718	0.279
B D	1K7C.A	8	0.062	8.963	-0.060	0.020	0.205	0.775
B S	1K7C.A	9	0.033	3.879	-0.715	0.021	0.279	0.699
B T	1K7C.A	10	0.049	6.741	0.217	0.074	0.484	0.442
B M	1K7C.A	11	0.041	8.184	-0.316	0.064	0.216	0.721
B A	1K7C.A	12	0.200	22.084	0.164	0.052	0.084	0.864
E K	1K7C.A	13	0.397	81.704	-0.503	0.018	0.088	0.893
E N	1K7C.A	14	0.380	55.676	-0.612	0.018	0.047	0.935
E G	1K7C.A	15	0.549	43.183	-2.134	0.018	0.047	0.935
E G	1K7C.A	16	0.624	49.077	-2.846	0.018	0.019	0.964
E G	1K7C.A	17	0.492	38.681	-1.751	0.018	0.019	0.964

What are the different methods?

- Simple sequence based methods
 - Align (BLAST) sequence against sequence of proteins with known structure (PDB database)
 - Sequence profile based methods
 - Align sequence profile (Psi-BLAST) against sequence of proteins with known structure (PDB, FUGUE)
 - Align sequence profile against profile of proteins with known structure (FFAS)
 - Sequence and structure based methods
 - Align profile and predicted secondary structure against proteins with known structure (3D-PSSM, Phyre)
 - Sequence profiles and structure based methods
 - HHpred, CPHModels
 - Multiple template methods
 - Modeler (via HHpred, 3D jury)
-

CpHModels - Fold recognition performance

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS



CpHModels 3.2

CENTERFO R BIOLOGI CAL SEQU ENCE ANA LYSIS CBS	EVENTS	NEWS	RESEARCH GROUPS	CBS PREDICTION SERVERS	CBS DATA SETS	PUBLICATIONS	EDUCATION
	STAFF	CONTACT	ABOUT CBS	INTERNAL	CBS BIOINFORMATICS TOOLS	CBS COURSES	OTHER BIOINFORMATICS LINKS

[CBS](#) >> [CBS Prediction Servers](#) >> CPHmodels

CPHmodels 3.2 Server

CPHmodels 3.2 is a protein homology modeling server. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions.

New in version 3.2: An improvement in the alignment algorithm in case of remote homology modeling where a structure dependant gap penalty has been introduced. Also, there are changes in the output format. A summary line has been included to make parsing easier.

View the [version history](#) of this server.

[Instructions](#)

[Output format](#)

[Article abstract](#)

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

Submit a file in [FASTA](#) format directly from your local disk:

no file selected

Restrictions:

Only one sequence per submission with not less than 15 and not more than 4,000 amino acids.

Confidentiality:

The sequences are kept confidential and will be deleted after processing.

Take home message

- Identifying the correct fold is only a small step towards successful homology modeling
 - Do not trust % ID or alignment score to identify the fold. Use P-values
 - You **can** do reliable fold recognition **AND** homology modeling even for low sequence homology
 - Use sequence profiles and local protein structure (predictions) to align sequences
-