

# Анализа протеинских секвенци

# Циљеви часа

- Предвиђање основних биохемијских особина протеина
- Претрага мотива у протеинским секвенцама
- Идентификација протеинских домена, и њихова веза са функцијом протеина

# План часа

1. In-silico биохемија
2. Предвиђање трансмембранских домена
3. Предвиђање пост-транслационих модификација PROSITE-ом
4. Коришћење колекција домена

# In-silico биохемија

- Online сервери омогућавају да се одреде многе биохемијске особине протеина
  - Молекуларна тежина
  - Време полуживота
- Све ове програме за анализу можете наћи на
  - [www.expasy.ch](http://www.expasy.ch)

# Пример

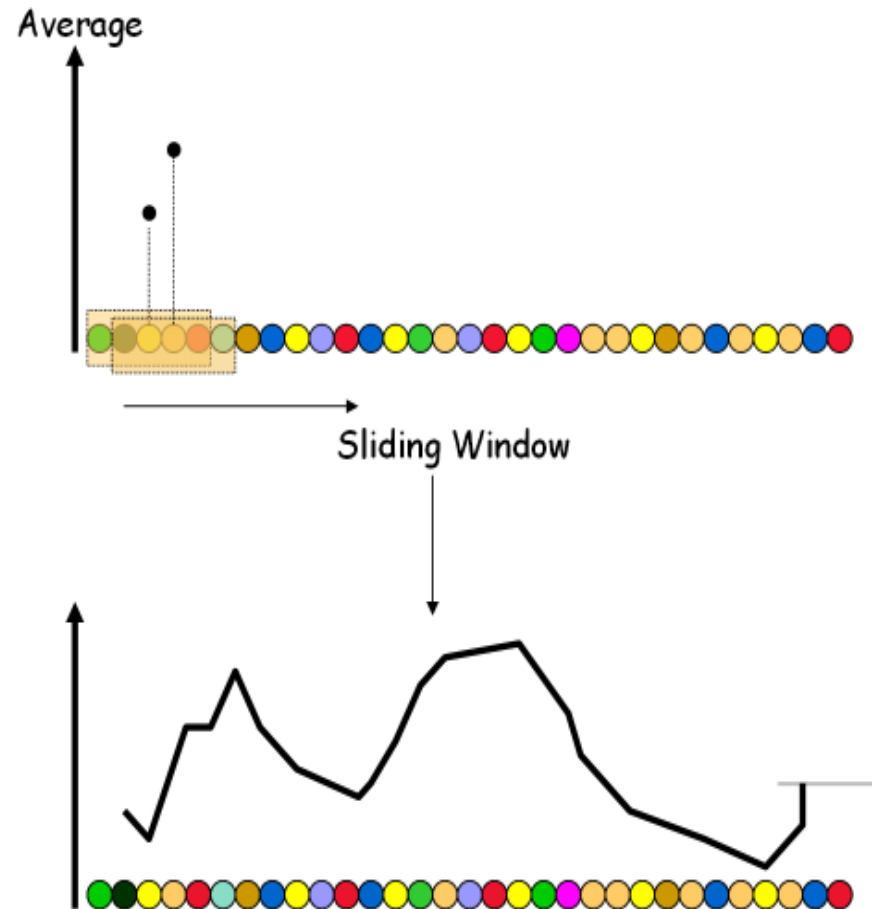
- Одредите основне биохемијске особине за dUTPase E coli (Swiss prot accession number је P06968).
- Користите web сајт <http://www.expasy.ch/tools/#primary>
- Идите на ProtParam
- Примедба: Instability index – ако је вредност индекса испод 40, протеин је обично стабилан
- Примедба: Молекулска маса не узима у обзир могуће пост-транслационе модификације
- Примедба: Време полуживота може бити знатно краће ако је деградација протеина део регулаторног процеса

# Анализа локалних особина протеина

- Многе локалне особине су битне за функцију вашег протеина
  - Хидрофобни региони су потенцијални трансмембрански домени
  - Коил-коил региони су потенцијална места за интеракцију између протеина
  - Хидрофилни региони су потенцијална места за протеинске петље (loops).
- Можете открити те регионе
  - Користећи технику покретног прозора (једноставан начин)
  - Постоје и сложенији методи који показују већу тачност

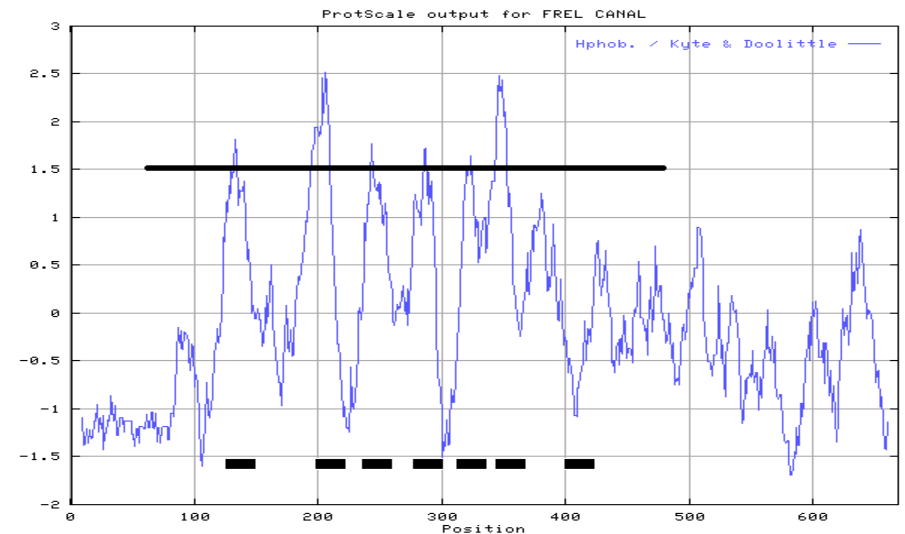
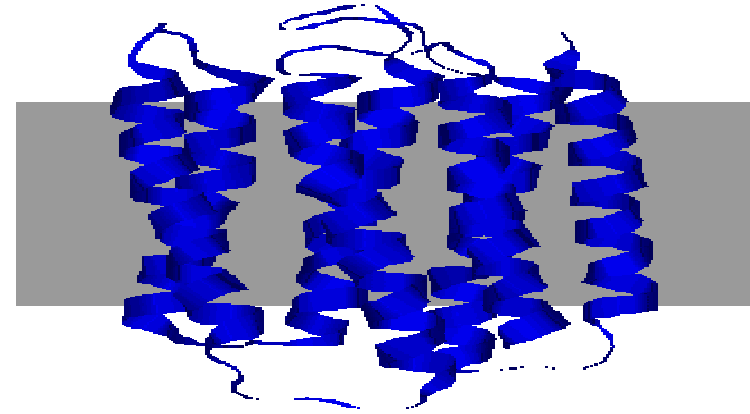
# Техника покретног прозора

- Идеална за идентификацију јаких сигнала
- Врло једноставан метод
- Можете да га испробате на [www.expasy.org](http://www.expasy.org) (ProtScale)
- Направите прозор исте величине као и особина коју испитујете (нпр. за трансмембранске домене ~ 20 ).



# Трансмембрански домени

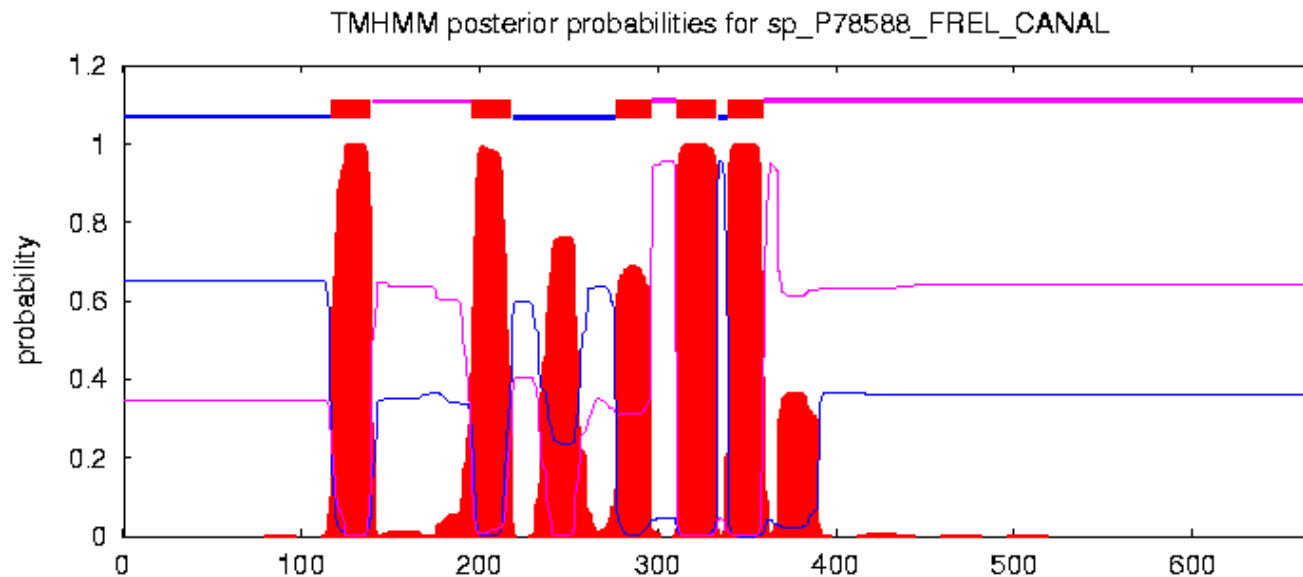
- Откривање трансмембранског сегмента вам каже доста о протеину
- Нпр. неки битни рецептори имају 7 трансмембранских сегмената
- Трансмембрански сегменти могу бити откривени користећи технику покретног прозора
- Тачније предикције се могу добити помоћу ТМНММ



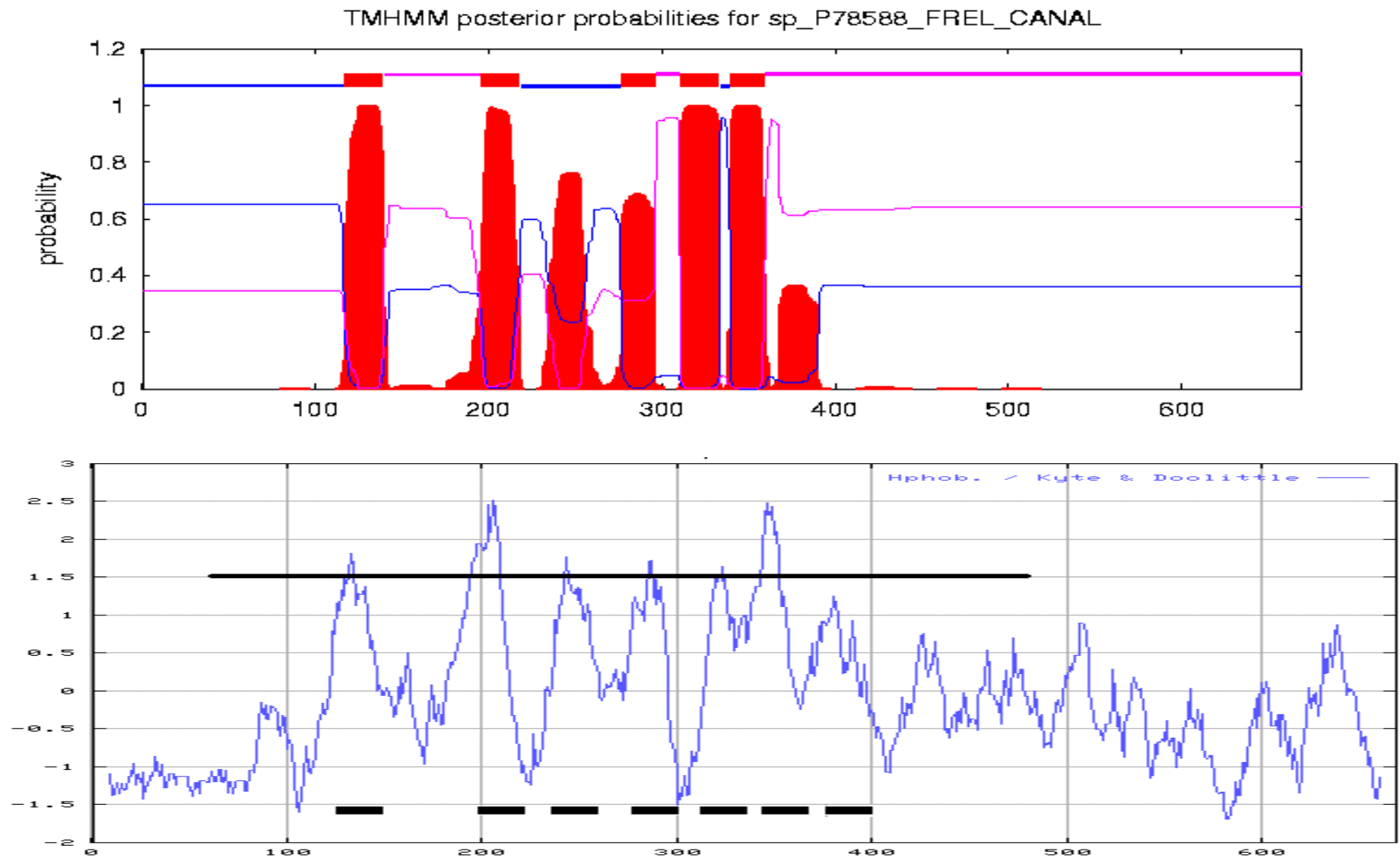


# Коришћење ТМНММ

- ТМНММ је један од најбољих метода за предвиђање трансмембранских домена
- Принцип на коме ради је различит од метода покретног прозора
- ТМНММ даје предвиђање



# TMHMM vs. метод покретног прозора



# Пример ТМНММ

- Предвидите трансмембранске регионе за протеин са Swiss Prot accession бројем P78588.
- Прво искористите Swiss Prot да би запамтили секвенцу овог протеина у FASTA формату
- Идите на [www.cbs.dtu.dk/services/TMHMM-2.0](http://www.cbs.dtu.dk/services/TMHMM-2.0) и генеришите предвиђања за трансмембранске домене.

# Предвиђање посттранслационих модификација

- Посттранслационе модификације су често повезане са сличним мотивима који се јављају у различитим протеинима
- PROSITE је база познатих мотива, при чему је сваки повезан са одговарајућом функцијом или посттранслационом модификацијом
- Можете да претражујете PROSITE тако што проверите да ли се сваки мотив који се налази у овој бази налази у вашем протеину (то сервер аутоматски ради за вас)
- PROSITE уноси су повезани са екстензивном документацијом

# Претраживање PROSITE мотива

- Претражите ваше протеине помоћу PROSITE-а на ExPASy
  - [www.expasy.org/tools/scanprosite](http://www.expasy.org/tools/scanprosite)
- PROSITE предвиђања треба да посматрате у биолошком контексту
  - Кратки мотиви сами по себи нису много информативни
  - Треба их посматрати само као могућности
  - Комбинујте их са другим информацијама да би извели крајњи закључак
- Такође запамтите: Није сваки мотив у PROSITE-у

# Интерпретирање PROSITE предвиђања

- Проверите функцију мотива: Да ли је компатибилна са протеином?
  - Некада мотиви сугеришу непостојеће функције протеина
  - На пример: Ако нађете мотив за миристилацију (myristylation) у бактерији, игноришите га; код бактерија нема миристилације!
- Кратки мотиви су поузданији ако можете да их нађете у више хомологних секвенци
  - У том случају можете да направите multiple-sequence alignment – поравнање више секвенци
  - Доња слика показује пример

```
chite  ---ADKPKRPLSAFMLWLNLSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNKPKRAPSAFFVFMGEFREFKQKNPKNKSVAAVGKAAGERWKS LSE
trybr  KKDSNA PKRAMTSMFFSSDFRS----KHSDLS-IVEMSKAAGAAWKE L GP
mouse  -----KPKRPRSAYNIVSESFQ----EAKDDS-AQGK LKLVNEAWKNLSP
          ***.  :::  .:  ..  .  :  .  .  *  .  *:  *
```

# Пример: претрага PROSITE базе

- Идентификујте који мотиви из PROSITE базе се налазе у протеину са Swiss Prot accession number P00579
- Идите на [www.expasy.org/tools/scanprosite](http://www.expasy.org/tools/scanprosite)
- Шта можете да кажете о функцији овог протеина на основу идентификованих мотива.
- Да ли сви кратки мотиви имају биолошки значај.

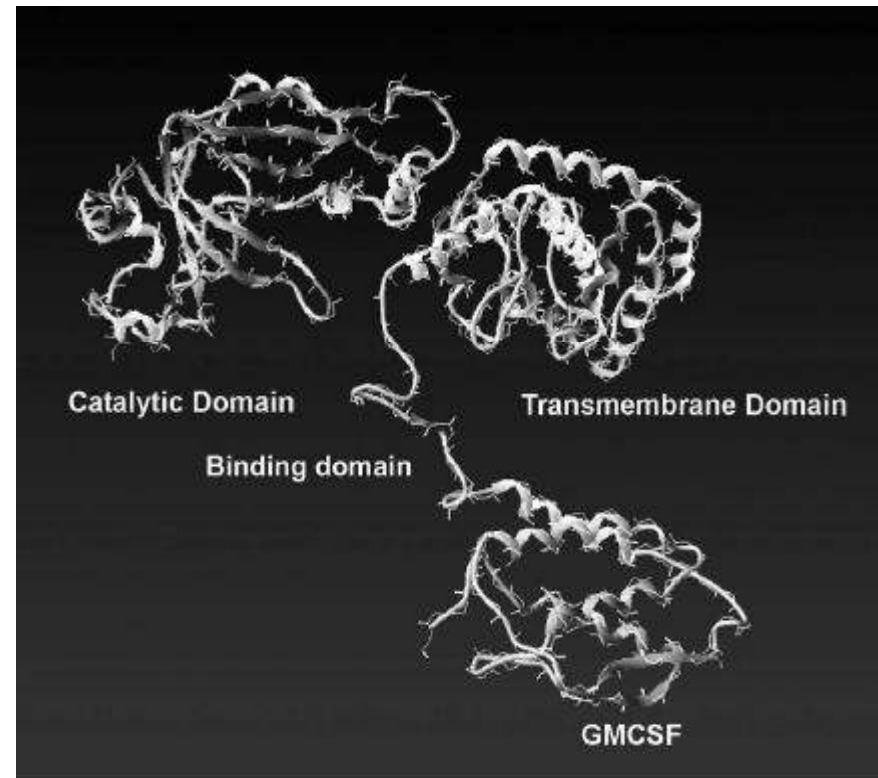
# Мотиви и домени

- Мотиви су генерално карактеристични делови ширих структура који се називају домени
- Домени су мање конзервирани него мотиви али су обично дужи
- Код протеина, анализа домена постепено замењује анализу мотива



# Домени протеина

- Протеини се обично састоје од домена
- Домен је јединица која се аутономно склапа у 3D структуру
- Домени су дужи од 50 аминокиселина
- Често се следећи домени налазе заједно:
  - Регулаторни домен
  - Домен за везивање
  - Каталитички домен



Crystal structure of DTGM.

# Проналажење домена

- Домени се откривају тако што:
  - се упоређују протеини са сличним функцијама
  - секвенце тих протеина се поравнавају
  - идентификују се очувани сегменти
- Биоинформатички, домен је поравнање више секвенци које се формулише као мотив
- За сваку колону, тај мотив указује која аминокиселина је вероватнија да се појави у оквиру домена

# Колекције домена

- Научници откривају и одређују функције домена већ више од 20 година
- Тренутно су установљене 8 различитих колекција домена
  - Ручно састављене колекције су веома прецизне али мале
  - Аутоматске колекције су врло екстензивне, али мање информативне
- Ове колекције
  - Се преклапају
  - Састављене су од стране различитих људи
  - Имају различите предности и мане
- Најбоље да искористите што више њих!

# Величанствених 8

<i>Name</i>	<i>Web Address</i>	<i>Size</i>	<i>Generation</i>
PROSITE-Profile (IP)	<a href="http://www.expasy.org/prosite">www.expasy.org/prosite</a>	616	Manual
PfamA (IP)	<a href="http://www.sanger.ac.uk/Software/Pfam">www.sanger.ac.uk/Software/Pfam</a>	7973	Manual
PRINTs (IP)	<a href="http://www.bioinf.man.ac.uk/dbbrowsers/PRINTS">www.bioinf.man.ac.uk/dbbrowsers/PRINTS</a>	1900	Manual
PRODOM (IP)	<a href="http://protein.toulouse.inra.fr/prodom/current/html/home.php">protein.toulouse.inra.fr/prodom/current/html/home.php</a>	736000	Automatic
SMART (IP)	<a href="http://smart.embl-heidelberg.de">smart.embl-heidelberg.de</a>	685	Manual
COGs	<a href="http://www.ncbi.nlm.nih.gov/COG/new/">www.ncbi.nlm.nih.gov/COG/new/</a>	4852	Manual
TIGRFAM (IP)	<a href="http://www.tigr.org/TIGRFAMs">www.tigr.org/TIGRFAMs</a>	2453	Manual
BLOCKS	<a href="http://blocks.fhcrc.org/">blocks.fhcrc.org/</a>	12542	Automatic








- Pfam је најобимнија ручно састављена колекција
- Pfam се често користи као референца

# Претраживање колекције домена

- Домени у Pfam-у често садрже познате функције
- Поклапање између вашег протеина и домена је врло пожељно
  - Поклапање је потенцијална индикација функције
  - То је **ВРЛО** информативно за даље истраживање!
- Постоје три сервера који омогућавају да се упореде протеини и колекције домена:
  - InterProScan [www.ebi.ac.uk/interproscan](http://www.ebi.ac.uk/interproscan)
  - CD-Search [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
  - Motif Scan [www.ch.embnet.org](http://www.ch.embnet.org)

# Коришћење InterProScan

- InterProScan је најобимнији метод за претраживање база са доменима
- Не наводи статистичку поузданост

SEQUENCE: <a href="#">FOSB_HUMAN</a> CRC64: DDF827C5047850F LENGTH: 338 aa	
InterPro <a href="#">IPR000209</a> Domain	Peptidase S8 and S53, subtilisin, kexin, sedolisin <a href="#">PS00138</a>  SUBTILASE_SER
InterPro <a href="#">IPR000837</a> Family	Fos transforming protein <a href="#">PR00042</a>  LEUZIPRFOS
InterPro <a href="#">IPR004827</a> Domain	Basic-leucine zipper (bZIP) transcription factor <a href="#">SM00338</a>  BRLZ <a href="#">PS00038</a>  BZIP_BASIC <a href="#">PS50217</a>  BZIP
InterPro <a href="#">IPR008817</a> Domain	Eukaryotic transcription factor, DNA-binding <a href="#">SSF47454</a>  Euk_transcr_DNA
InterPro <a href="#">IPR011700</a> Domain	Basic leucine zipper <a href="#">PF07716</a>  bZIP_2

# Пример - претрага са InterProScan

- [www.ebi.ac.uk/InterProScan/](http://www.ebi.ac.uk/InterProScan/)
- Помоћу InterProScan-а предвидите домене за  $\sigma 70$  код E coli (Swiss Prot accession број P00579)
- Ради веће брзине изоставите ProDom и базу из претраге
- Да запамтите резултат, притисните Tool Output дугме и запамтите фајл као web страну.

# Резултат CD претраге

- CD претрага је мање екстензивна него InterProScan
- CD претрага омогућава приступ COGs
- Резултати се наводе са статистичким значајем (E вредност)
  - $10^{e-15}$  **Ниска E-вредност**    Дobar погодак
  - 2.1    **Висока E-вредност**    Лош погодак

The screenshot displays a search result for the domain BRLZ. At the top, there are radio buttons for 'Concise Result' (selected), 'Full Result', and a checkbox for 'Show Search Information'. Below this is a protein domain alignment bar with positions 1, 75, 150, 225, 300, and 338 marked. A red box labeled 'BRLZ' is positioned over the alignment. Below the alignment is a yellow tooltip that reads 'BRLZ -- Click for more options' and 'smart00338, BRLZ, basic region leucin zipper; .'. Underneath the tooltip is a table with the following data:

Name	Title	PssmId	Accession	E-value
<a href="#">(+BRLZ</a>	smart00338, BRLZ, basic region leucin zipper; .	47665	smart00338	2e-15

At the bottom of the interface is a button labeled 'Search for similar domain architectures'.



# Пример - претрага са CD search

- <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>
- Помоћу CD-а предвидите домене за  $\sigma 70$  код E coli (Swiss Prot accession број P00579)
- Ради веће брзине изоставите ProDom и Super Family из претраге
- Да запамтите резултат, притисните Tool Output дугме и запамтите фајл као web страну.

# Задаци за вежбу

- 1) Испитајте биохемијске особине за dUTPase human (Swiss Prot Accession number P33316). Запамтите резултат као web страну.
- 2) Предвидите трансмембранске сегменте за протеин са Swiss Prot accession бројем P78588. Запамтите резултат као web страну. Колико има трансмембранских сегмената, и који део протеина је унутар, а који изван ћелије?
- 3) Идентификујте Prosite мотиве унутар протеина са Swiss Prot accession P12259. Запамтите резултат као web страну.

- 4) Помоћу InterProScan предвидите домене за протеин са SwissProt accession бројем P53539. Напомена: морате прво преко Swiss-Prot-а да запамтите секвенцу протеина као FASTA фајл. Запамтите екран и сачувајте га у PowerPoint-у.
- 5) Поновите горњи задатак користећи NCBI CD сервер. Запамтите резултат користећи Print Screen са тастатуре, а онда Cut и Paste у Power Point презентацију.