# Introduction to Protein Analysis Tools and Prediction

呂平江博士/ 鄭兆勝博士
清大生資所


# ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server

***Switzerland:*** **http://www.expasy.org/** at Swiss Institute of Bioinformatics, Geneva
***Australia:*** **http://au.expasy.org/** at Australian Proteome Analysis Facility, Sydney
***Brazil:*** **http://br.expasy.org/** at Laboratório Nacional de Computação Científica, Petrópolis
***Canada:*** **http://ca.expasy.org/** at Canadian Bioinformatics Resource, Halifax
***Taiwan:*** **http://tw.expasy.org/** at National Health Research Institute
***China:*** **http://cn.expasy.org/** at Peking University
***Korea:*** **http://kr.expasy.org/** at Yonsei Proteome Research Center, Seoul

# Introduction

- ExPASy是由瑞士生物資訊機構 (Swiss Institute of Bioinformatics， SIB)所架設的伺服器。此伺服器包含了資料庫、工具&軟體、教育服務及一些相關連結。因為此伺服器所收集的資料及所提供的服務相當的龐大，無法一一概述，因此，我們在本課程中只針對資料庫及一部分的工具和軟體加以說明。其餘部分將由有同學自行參考。

**Databases**

- **UniProt Knowledgebase (Swiss-Prot and TrEMBL)** - Protein knowledgebase
- **PROSITE** - Protein families and domains
- **SWISS-2DPAGE** - Two-dimensional polyacrylamide gel electrophoresis
- **ENZYME** - Enzyme nomenclature
- **SWISS-MODEL Repository** - Automatically generated protein models
- Links to many other molecular biology databases

**Tools and software packages**

- **Proteomics and sequence analysis tools**
  - Identification and characterization (Aldente, FindMod, Popitam, Phenyx, pI/Mw, ProtParam...)
  - DNA -> Protein
  - Similarity searches (BLAST...)
  - Pattern and profile searches (ScanProsite...)
  - Post-translational modification and topology prediction
  - Primary structure analysis
  - Secondary and tertiary structure tools (Swiss-PdbViewer...)
  - Alignment and Phylogenetic analysis
- **ImageMaster / Melanie** - Software for 2-D PAGE analysis
- **MSight** - Mass Spectrometry Imager
- **Roche Applied Science's Biochemical Pathways**

**Education and services**

- **The ExPASy FTP server**
- **Swiss-Shop** - automatically obtain (by email) new sequence entries relevant to your field(s) of interest
- **Vital-IT** - The HPC Center for Life Sciences
- **e-Proxemis** - Bioinformatics Learning Portal for Proteomics
- **Master's degree in Proteomics and Bioinformatics**
- **Proteomics Core Facility** (previously SWISS-2DSERVICE) - get your 2-D Gels performed according to Swiss standards

**Documentation**

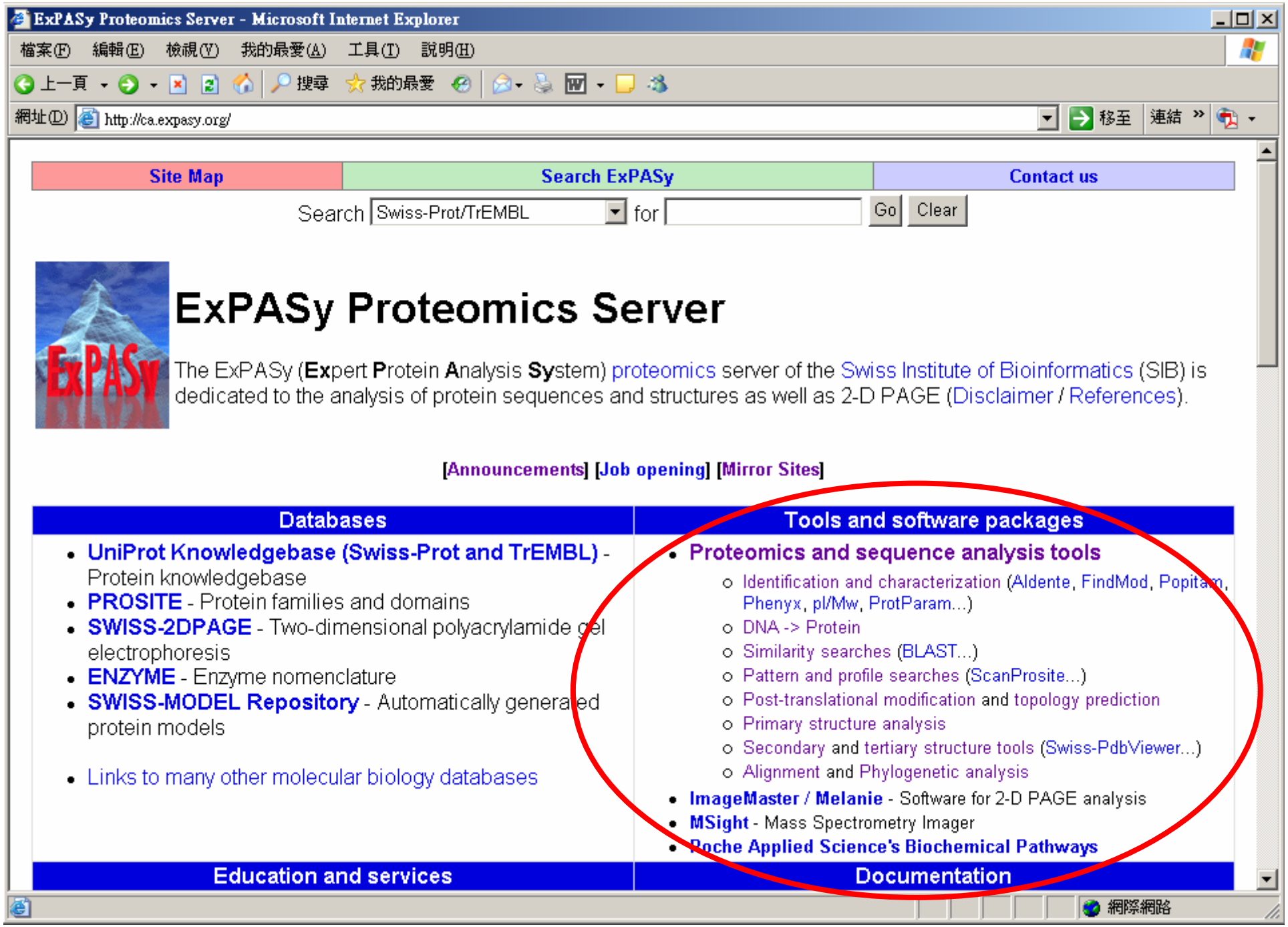
- **What's New on ExPASy**
- **SWISS-FLASH** electronic bulletins
- **Swiss-Prot documents**
- **How to create HTML links to ExPASy**
- **Complete table of available documents**

**Links to lists of molecular biology resources**

- **ExPASy Life Science Directory** - The ExPASy list of biomolecular servers
- **BioHunt** - Search the internet for molecular biology information
- **WORLD-2DPAGE list** - Links to 2-D PAGE database servers and 2-D PAGE related servers and services
- **World-2DPAGE portal** new - A dynamic portal to query simultaneously World-Wide proteomics databases
- **CMS-SDSC** - The CMS-SDSC Molecular Biology Resource
- **Biology links** - from Harvard University
- **Yahoo - Science:Biology**

**Links to some major molecular biology servers**

- **European Bioinformatics Institute (EBI)**
- **National Center for Biotechnology Information (NCBI)**
- **Japanese GenomeNet**
- **Australian National Genomic Information Service (ANGIS)**
- **BIOSCI/bionet Electronic Newsgroup Network for Biology**


ExPASy Proteomics Server - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛
網址(D) http://ca.expasy.org/
移至 連結
Site Map
Search ExPASy
Contact us
Search Swiss-Prot/TrEMBL for Go Clear
ExPASy Proteomics Server
The ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References).
[Announcements] [Job opening] [Mirror Sites]
Databases
UniProt Knowledgebase (Swiss-Prot and TrEMBL) - Protein knowledgebase
PROSITE - Protein families and domains
SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
ENZYME - Enzyme nomenclature
SWISS-MODEL Repository - Automatically generated protein models
Links to many other molecular biology databases
Tools and software packages
Proteomics and sequence analysis tools
Identification and characterization (Aldente, FindMod, Popitam, Phenyx, pI/Mw, ProtParam...)
DNA -> Protein
Similarity searches (BLAST...)
Pattern and profile searches (ScanProsite...)
Post-translational modification and topology prediction
Primary structure analysis
Secondary and tertiary structure tools (Swiss-PdbViewer...)
Alignment and Phylogenetic analysis
ImageMaster / Melanie - Software for 2-D PAGE analysis
MSight - Mass Spectrometry Imager
Roche Applied Science's Biochemical Pathways
Education and services
Documentation
網際網路

# 十二大分類項目

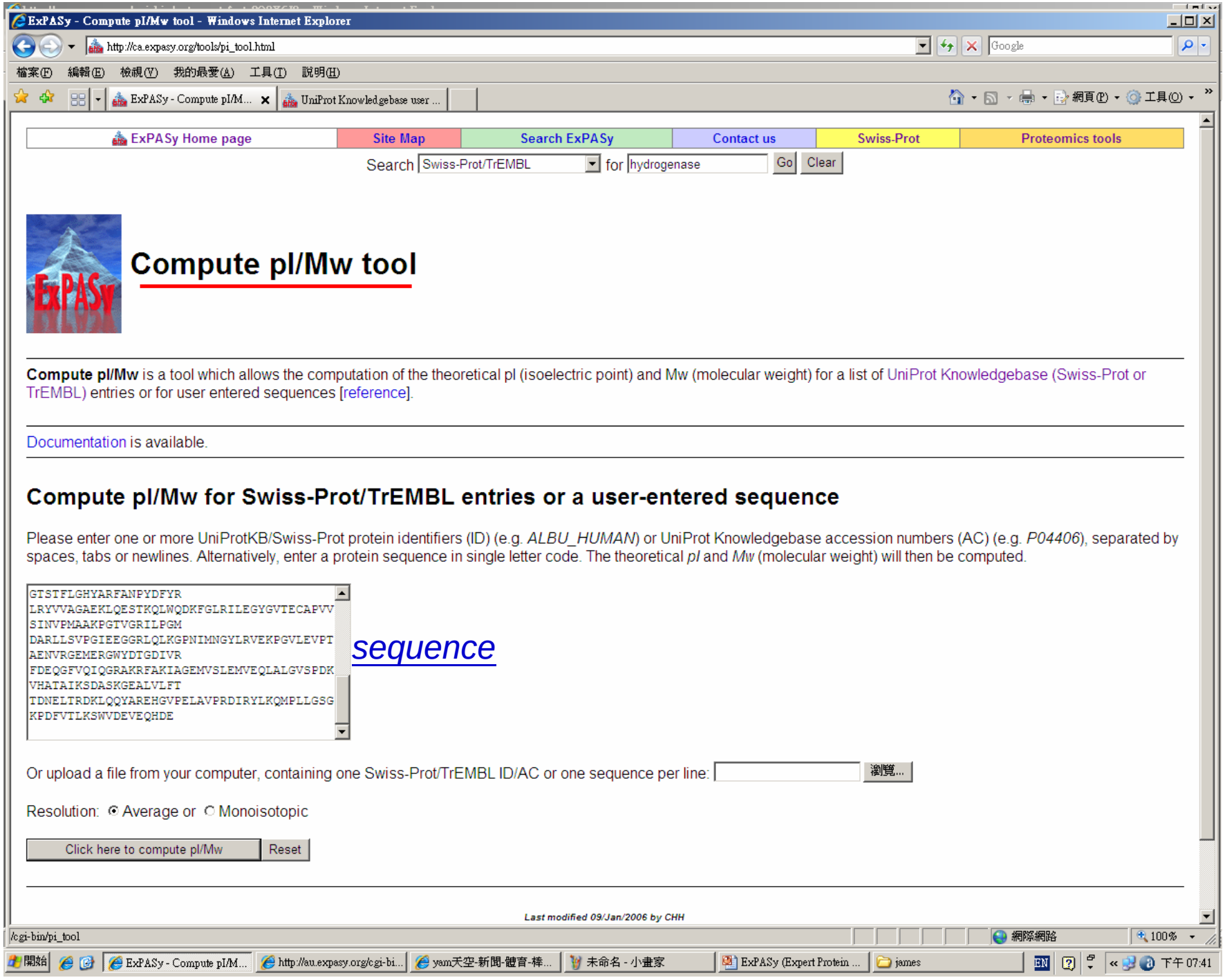
- 蛋白質身份辨識與理化特性分析 (Identification and characterization)
- 人工轉譯分析 (DNA -> Protein)
- 相似序列搜尋 (Similarity searches)
- 樣板序列搜尋與分析 (Pattern and profile searches)
- 轉譯後修飾預測 (Post-translational modification prediction)
- 拓樸特性預測分析(Topology prediction)
- 一級結構分析 (Primary structure analysis)
- 二級結構分析 (Secondary structure prediction)
- 三級結構分析 (Tertiary structure)
- 序列比對 (Sequence alignment)
- 演化樹分析 (Phylogenetic analysis)
- 生物學關鍵字分析 (Biological text analysis)

# 蛋白質身份辨識與理化特性分析 (Identification and characterization)

- 以氨基酸序列預測質譜指紋 (mass fingerprinting) 或以質譜指紋來確認蛋白質成份。

- 分子量(MW)、等電點(pI)計算與氨基酸成分比例分析
  - Compute pI/Mw, AACompSim

- 蛋白分解酵素切位與片段分析
  - PeptideCutter, PeptideMass


ExPASy - Compute pI/Mw tool - Windows Internet Explorer
http://ca.expasy.org/tools/pi_tool.html
ExPASy Home page
Site Map
Search ExPASy
Contact us
Swiss-Prot
Proteomics tools
Search Swiss-Prot/TrEMBL for hydrogenase Go Clear
Compute pI/Mw tool
Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for a list of UniProt Knowledgebase (Swiss-Prot or TrEMBL) entries or for user entered sequences [reference].
Documentation is available.
Compute pI/Mw for Swiss-Prot/TrEMBL entries or a user-entered sequence
Please enter one or more UniProtKB/Swiss-Prot protein identifiers (ID) (e.g. ALBU_HUMAN) or UniProt Knowledgebase accession numbers (AC) (e.g. P04406), separated by spaces, tabs or newlines. Alternatively, enter a protein sequence in single letter code. The theoretical pI and Mw (molecular weight) will then be computed.
GTSTFLGHYARFANPYDFYR
LRYVVAGAEKLQESTKQLWQDKFGLRILEGYGVTECAPVV
SINVPMAAKPGTVGRILPGM
DARLLSVPGIEEGGRLQLKGPNIMNGYLRVEKPGVLEVPT
AENVRGEMERGWYDTGDIVR
FDEQGFVQIQGRAKRFAKIAGEMVSLEMVEQLALGVSPDK
VHATAIKSDASKGEALVLFT
TDNELTRDKLQQYAREHGVPELAVPRDIRYLKQMPLLGSG
KPDFVTLKSWVDEVEQHDE
sequence
Or upload a file from your computer, containing one Swiss-Prot/TrEMBL ID/AC or one sequence per line:
瀏覽...
Resolution: Average or Monoisotopic
Click here to compute pI/Mw
Reset
Last modified 09/Jan/2006 by CHH
/cgi-bin/pi_tool
網際網路
100%
開始
ExPASy - Compute pI/M...
http://au.expasy.org/cgi-bi...
yam天空-新聞-體育-棒...
未命名 - 小畫家
ExPASy (Expert Protein ...
james
下午 07:41

# Results

## Compute pI/Mw

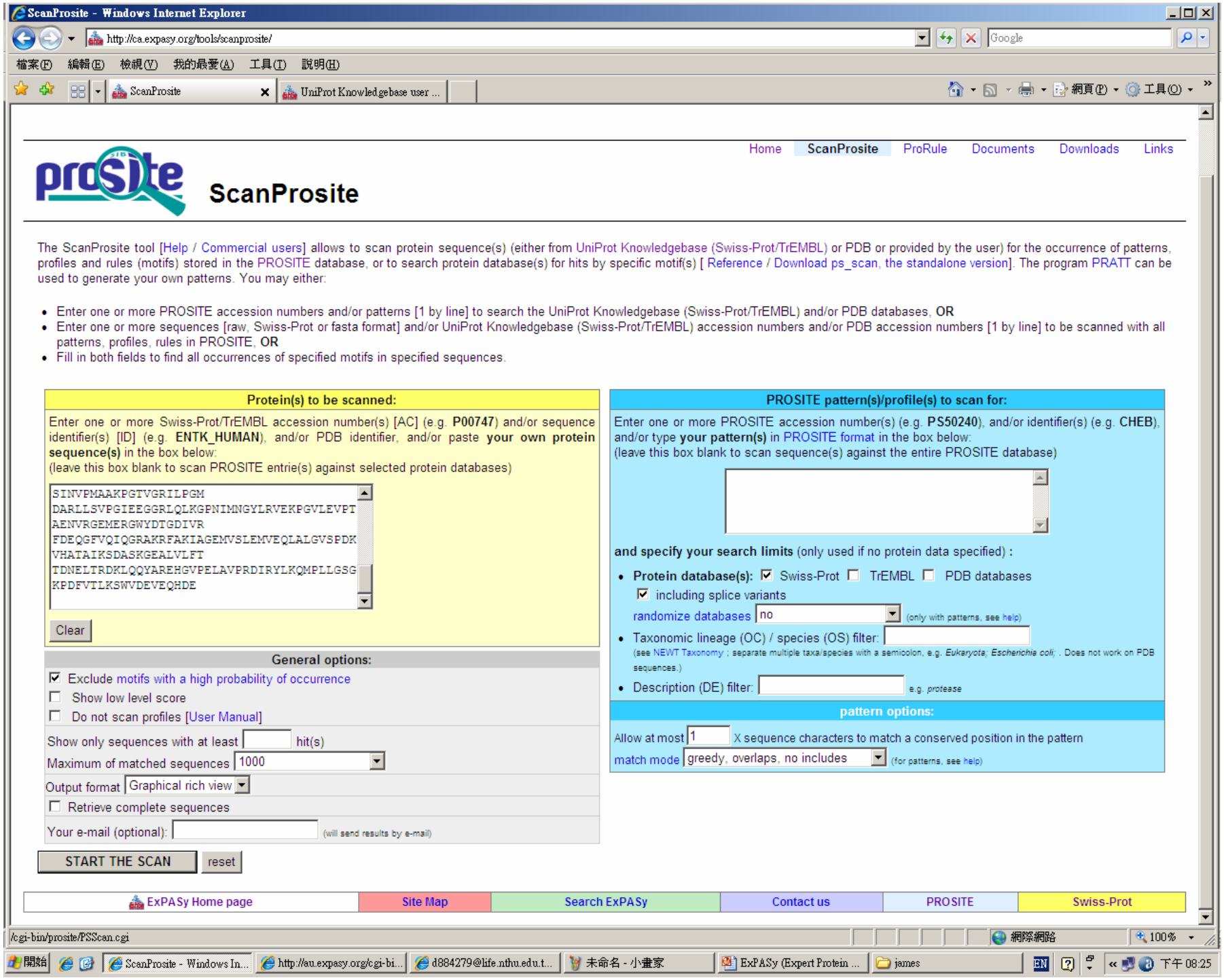
### Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MLFSFFRNLC RVLYRVRVTG DPQALKGERV LITPNHVSFI DGILLGLFLP VRPVFAVYTS
        70         80         90        100        110        120
ISQQWYMRWL KSFIDFVPLD PTQPMAIKHL VRLVEQGRPV VIFPEGRITT TGSLMKIYDG
       130        140        150        160        170        180
AGFVAAKSGA TVIPVRIEGA ELTHFSRLKG LVKRRLFPQI TLHILPPTQV EMPDAPRARD
       190        200        210        220        230        240
RRKIAGEMLH QIMMEARMAV RPRETLYESL LSAMYRFGAG KKCVEDVNFT PDSYRKLLTK
       250        260        270        280        290        300
TLFVGRILEK YSVEGERIGL MLPNAGISAA VIFGAIARRR IPAMMNYTAG VKGLTSAITA
       310        320        330        340        350        360
AEIKTIFTSR QFLDKGKLWH LPEQLTQVRW VYLEDLKADV TTADKVWIFA HLLMPRLAQV
       370        380        390        400        410        420
KQQPEEEALI LFTSGSEGHP KGVVHSHKSI LANVEQIKTI ADFTTNDRFM SALPLFHSFG
       430        440        450        460        470        480
LTVGLFTPLL TGAEVFLYPS PLHYRIVPEL VYDRSCTVLF GTSTFLGHYA RFANPYDFYR
       490        500        510        520        530        540
LRYVVAGAEK LQESTKQLWQ DKFGLRILEG YGVTECAPVV SINVPMAAKP GTVGRILPGM
       550        560        570        580        590        600
DARLLSVPGI EEGGRLQLKG PNIMNGYLRV EKPGVLEVPT AENVRGEMER GWYDTGDIVR
       610        620        630        640        650        660
FDEQGFVQIQ GRAKRFAKIA GEMVSLEMVE QLALGVSPDK VHATAIKSDA SKGEALVLFT
       670        680        690        700        710
TDNELTRDKL QQYAREHGVP ELAVPRDIRY LKQMPLLGSG KPDFVTLKSW VDEVEQHDE
```

Theoretical pI/Mw: 9.27 / 80764.09

# 樣板序列搜尋與分析
# (Pattern and profile searches)

- 搜尋經常被重複性使用的功能或結構性樣版序列。例如 Calcium binding domain, ATPase active site 等。
- ScanProsite, PPSEARCH, MotifScan


ScanProsite - Windows Internet Explorer
http://ca.expasy.org/tools/scanprosite/
Google
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
ScanProsite
UniProt Knowledgebase user ...
網頁(P) 工具(O)
Home ScanProsite ProRule Documents Downloads Links
ScanProsite
The ScanProsite tool [Help / Commercial users] allows to scan protein sequence(s) (either from UniProt Knowledgebase (Swiss-Prot/TrEMBL) or PDB or provided by the user) for the occurrence of patterns, profiles and rules (motifs) stored in the PROSITE database, or to search protein database(s) for hits by specific motif(s) [ Reference / Download ps_scan, the standalone version]. The program PRATT can be used to generate your own patterns. You may either:
Enter one or more PROSITE accession numbers and/or patterns [1 by line] to search the UniProt Knowledgebase (Swiss-Prot/TrEMBL) and/or PDB databases, OR
Enter one or more sequences [raw, Swiss-Prot or fasta format] and/or UniProt Knowledgebase (Swiss-Prot/TrEMBL) accession numbers and/or PDB accession numbers [1 by line] to be scanned with all patterns, profiles, rules in PROSITE, OR
Fill in both fields to find all occurrences of specified motifs in specified sequences.
Protein(s) to be scanned:
Enter one or more Swiss-Prot/TrEMBL accession number(s) [AC] (e.g. P00747) and/or sequence identifier(s) [ID] (e.g. ENTK_HUMAN), and/or PDB identifier, and/or paste your own protein sequence(s) in the box below:
(leave this box blank to scan PROSITE entrie(s) against selected protein databases)
SINVPMAAKPGTVGRILPGM
DARLLSVPGIEEGGRLQLKGPNIMNGYLRVEKPGVLEVPT
AENVRGEMERGWYDTGDIVR
FDEQGFVQIQGRAKRFAKIAGEMVSLEMVEQLALGVSPDK
VHATAIKSDASKGEALVLFT
TDNELTRDKLQQYAREHGVPELAVPRDIRYLKQMPLLGSG
KPDFVTLKSWVDEVEQHDE
Clear
PROSITE pattern(s)/profile(s) to scan for:
Enter one or more PROSITE accession number(s) (e.g. PS50240), and/or identifier(s) (e.g. CHEB), and/or type your pattern(s) in PROSITE format in the box below:
(leave this box blank to scan sequence(s) against the entire PROSITE database)
and specify your search limits (only used if no protein data specified) :
Protein database(s): Swiss-Prot TrEMBL PDB databases
including splice variants
randomize databases no (only with patterns, see help)
Taxonomic lineage (OC) / species (OS) filter:
(see NEWT Taxonomy ; separate multiple taxa/species with a semicolon, e.g. Eukaryota; Escherichia coli; . Does not work on PDB sequences.)
Description (DE) filter: e.g. protease
General options:
Exclude motifs with a high probability of occurrence
Show low level score
Do not scan profiles [User Manual]
Show only sequences with at least hit(s)
Maximum of matched sequences 1000
Output format Graphical rich view
Retrieve complete sequences
Your e-mail (optional): (will send results by e-mail)
pattern options:
Allow at most 1 X sequence characters to match a conserved position in the pattern
match mode greedy, overlaps, no includes (for patterns, see help)
START THE SCAN
reset
ExPASy Home page
Site Map
Search ExPASy
Contact us
PROSITE
Swiss-Prot
/cgi-bin/prosite/PSScan.cgi
網際網路
100%
開始
ScanProsite - Windows In...
http://au.expasy.org/cgi-bi...
d884279@life.nthu.edu.t...
未命名 - 小畫家
ExPASy (Expert Protein ...
james
EN
下午 08:25

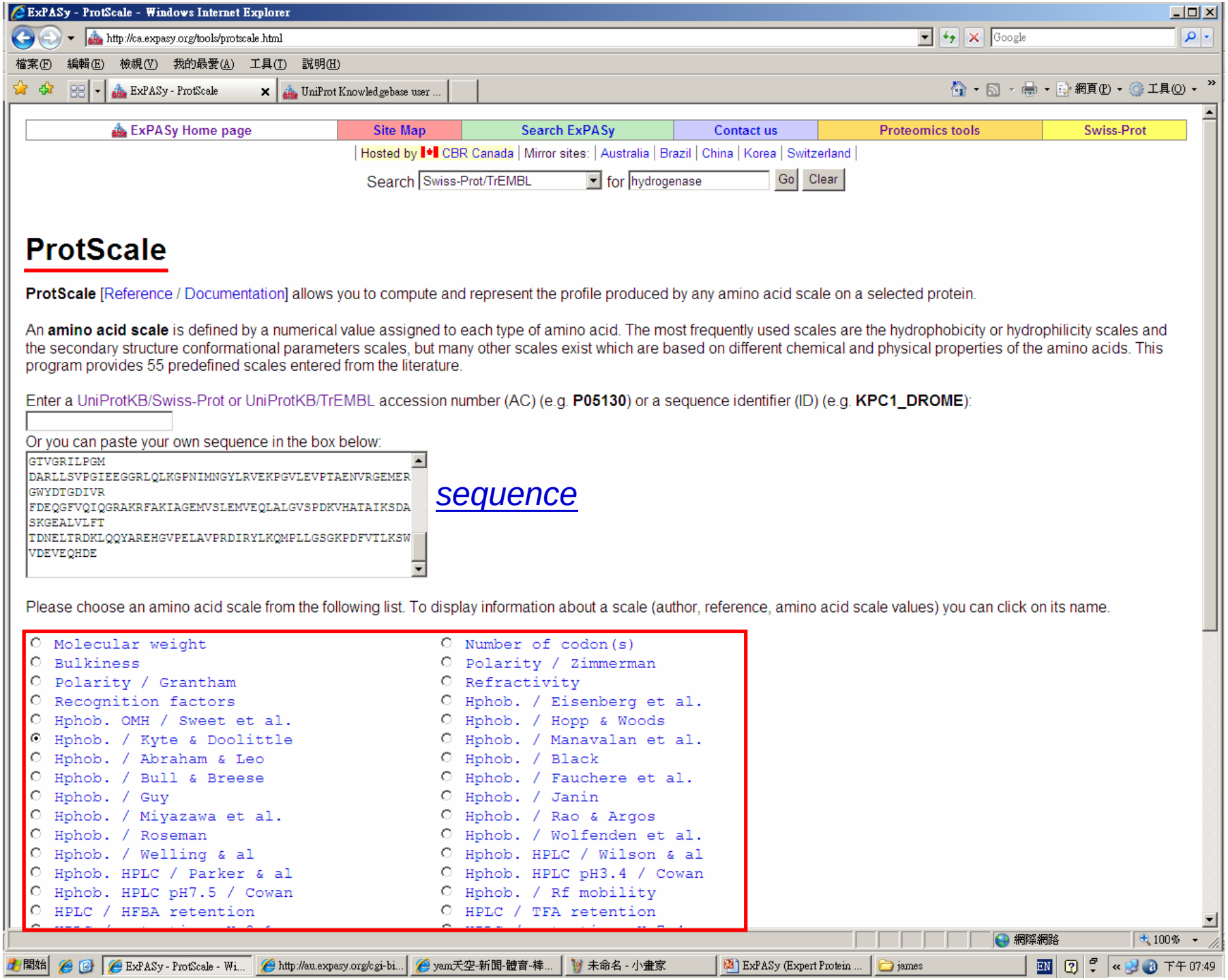
# 拓樸特性預測分析
# (Topology prediction)

- 利用氨基酸序列預測蛋白質的拓樸學特性。例如：
- 胞內分佈位置(subcellular localization)預測
  - PSORT, TargetP
- 穿膜區域預測
  - PredictProtein, SOSUI, TMHMM, TMpred
- 蛋白質骨架走勢分析
  - TopPred


SOSUI/submit a protein sequence - Windows Internet Explorer
http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
SOSUI/submit a protein se...
UniProt Knowledgebase user ...
網頁(P) 工具(O)
SOSUI: Submit a protein sequence
[Sample Sequences] [References]
Enter a title or comment for the sequence :
None
Enter your sequence with one-letter symbol (by copy & paste) :
(Minimum: 20 a.a., Maximum: 5000 a.a.)
GFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGC
NLEGFFATLG
GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAP
PLVGWSRYIP
EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTV
KEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTI
PAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA
To execute the query, press this button : Exec
To clear the form, press this button : Clear
sosui@proteome.bio.tuat.ac.jp
[Sample Sequences]
Rhodopsin (bovine)
Cytochrome c oxidase I subunit (bovine)
Myoglobin (human)
[References]
Hirokawa T., Boon-Chieng S., and Mitaku S., Bioinformatics, 14 378-9 (1998)
SOSUI: classification and secondary structure prediction system for membrane proteins.
[Abstruct] [Full Text(PDF)]
Mitaku S., Hirokawa T. Protein Eng. 11 (1999) Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length
[Abstruct] [Full Text(PDF)]
./cgi-bin/adv_sosui.cgi
網際網路
100%
開始
SOSUI/submit a prot...
http://au.expasy.org/...
d884279@life.nthu.e...
Sequence Manipulati...
未命名 - 小畫家
ExPASy (Expert Prot...
james
EN
下午 08:37

# 一級結構分析
# (Primary structure analysis)

- 基本理化特性計算分析
  - ProtParam, Compute pI/Mw
- 特殊序列搜尋分析
  - 重複性: REP
  - coiled coil: Coils
- 親水性、親油性、SSE預測分析
  - ProtScale, Drawhca


ExPASy - ProtScale - Windows Internet Explorer
http://ca.expasy.org/tools/protscale.html
ExPASy Home page
Site Map
Search ExPASy
Contact us
Proteomics tools
Swiss-Prot
Hosted by CBR Canada | Mirror sites: | Australia | Brazil | China | Korea | Switzerland |
Search Swiss-Prot/TrEMBL for hydrogenase Go Clear
ProtScale
ProtScale [Reference / Documentation] allows you to compute and represent the profile produced by any amino acid scale on a selected protein.
An amino acid scale is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity scales and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 55 predefined scales entered from the literature.
Enter a UniProtKB/Swiss-Prot or UniProtKB/TrEMBL accession number (AC) (e.g. P05130) or a sequence identifier (ID) (e.g. KPC1_DROME):
Or you can paste your own sequence in the box below:
GTVGRILPGM
DARLLSVPGIEEGGRLQLKGPNIMNGYLRVEKPGVLEVPTAENVRGEMER
GWYDTGDIVR
FDEQGFVQIQGRAKRFAKIAGEMVSLEMVEQLALGVSPDKVHATAIKSDA
SKGEALVLFT
TDNELTRDKLQQYAREHGVPELAVPRDIRYLKQMPLLGSGKPDFVTLKSW
VDEVEQHDE
sequence
Please choose an amino acid scale from the following list. To display information about a scale (author, reference, amino acid scale values) you can click on its name.
Molecular weight
Bulkiness
Polarity / Grantham
Recognition factors
Hphob. OMH / Sweet et al.
Hphob. / Kyte & Doolittle
Hphob. / Abraham & Leo
Hphob. / Bull & Breese
Hphob. / Guy
Hphob. / Miyazawa et al.
Hphob. / Roseman
Hphob. / Welling & al
Hphob. HPLC / Parker & al
Hphob. HPLC pH7.5 / Cowan
HPLC / HFBA retention
Number of codon(s)
Polarity / Zimmerman
Refractivity
Hphob. / Eisenberg et al.
Hphob. / Hopp & Woods
Hphob. / Manavalan et al.
Hphob. / Black
Hphob. / Fauchere et al.
Hphob. / Janin
Hphob. / Rao & Argos
Hphob. / Wolfenden et al.
Hphob. HPLC / Wilson & al
Hphob. HPLC pH3.4 / Cowan
Hphob. / Rf mobility
HPLC / TFA retention

Please choose an amino acid scale from the following list. To display information about a scale (author, reference, amino acid scale values) can click on its name.

- Molecular weight
- Bulkiness
- Polarity / Grantham
- Recognition factors
- Hphob. OMH / Sweet et al.
- Hphob. / Kyte & Doolittle
- Hphob. / Abraham & Leo
- Hphob. / Bull & Breese
- Hphob. / Guy
- Hphob. / Miyazawa et al.
- Hphob. / Roseman
- Hphob. / Welling & al
- Hphob. HPLC / Parker & al
- Hphob. HPLC pH7.5 / Cowan
- HPLC / HFBA retention
- HPLC / retention pH 2.1
- % buried residues
- Hphob. / Chothia
- Ratio hetero end/side
- Average flexibility
- beta-sheet / Chou & Fasman
- alpha-helix / Deleage & Roux
- beta-turn / Deleage & Roux
- alpha-helix / Levitt
- beta-turn / Levitt
- Antiparallel beta-strand
- A.A. composition
- Relative mutability
- Number of codon(s)
- Polarity / Zimmerman
- Refractivity
- Hphob. / Eisenberg et al.
- Hphob. / Hopp & Woods
- Hphob. / Manavalan et al.
- Hphob. / Black
- Hphob. / Fauchere et al.
- Hphob. / Janin
- Hphob. / Rao & Argos
- Hphob. / Wolfenden et al.
- Hphob. HPLC / Wilson & al
- Hphob. HPLC pH3.4 / Cowan
- Hphob. / Rf mobility
- HPLC / TFA retention
- HPLC / retention pH 7.4
- % accessible residues
- Hphob. / Rose & al
- Average area buried
- alpha-helix / Chou & Fasman
- beta-turn / Chou & Fasman
- beta-sheet / Deleage & Roux
- Coil / Deleage & Roux
- beta-sheet / Levitt
- Total beta-strand
- Parallel beta-strand
- A.A. comp. in Swiss-Prot

# Results

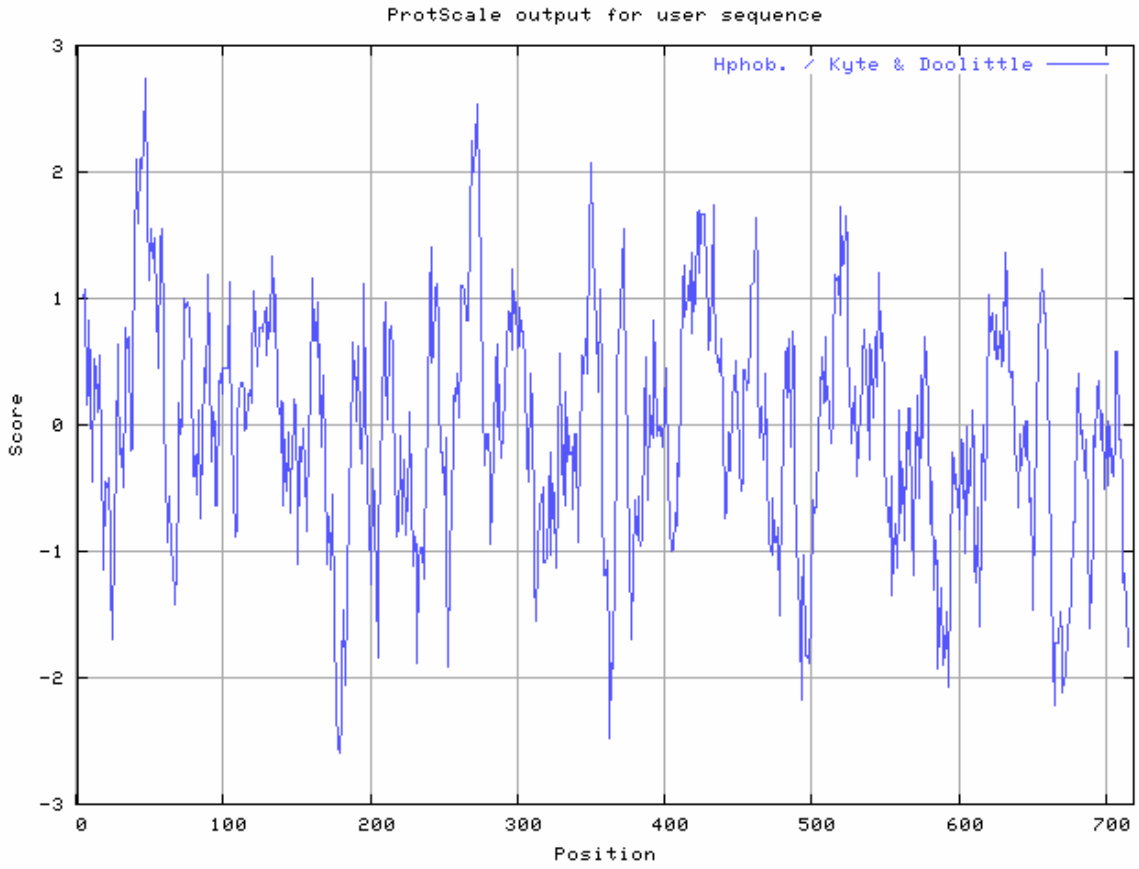
Using the scale **Hphob. / Kyte & Doolittle**, the individual values for the 20 amino acids are:

```
Ala:  1.800  Arg: -4.500  Asn: -3.500  Asp: -3.500  Cys:  2.500  Gln: -3.500
Glu: -3.500  Gly: -0.400  His: -3.200  Ile:  4.500  Leu:  3.800  Lys: -3.900
Met:  1.900  Phe:  2.800  Pro: -1.600  Ser: -0.800  Thr: -0.700  Trp: -0.900
Tyr: -1.300  Val:  4.200  Asx: -3.500  Glx: -3.500  Xaa: -0.490
```

Weights for window positions 1,..,9, using **linear weight variation model**:

```
   1     2     3     4     5     6     7     8     9
 1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 edge                    center                edge
```

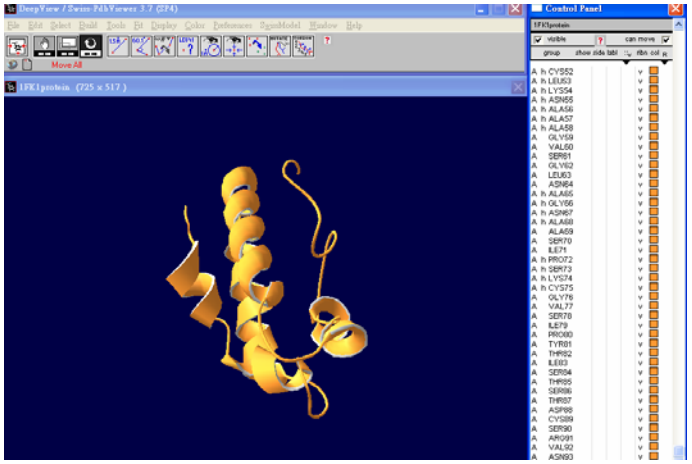
```
MIN: -2.600
MAX: 2.744
```

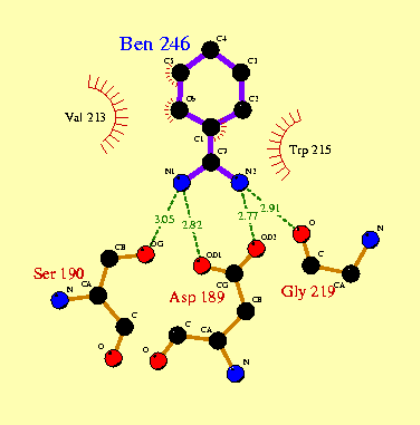


# SWISS PDB Viewer

# SWISS PDB Viewer

1. Distance/angle
2. Ribbon
3. Ramachandran plot
4. Mutation/rotamer
5. Superimpose
6. Compute H-bond/energy
7. Energy minimization
8. Cavity






# LIGPLOT v.4.4.2

## Program for automatically plotting protein-ligand interactions

Written by Andrew Wallace & Roman Laskowski

http://www.biochem.ucl.ac.uk/bsm/ligplot/ligplot.html

# Ligplot

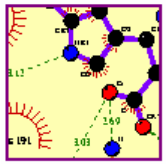
- Automatically generates **schematic diagrams** of **protein-ligand interactions** for a given PDB file. The interactions shown are those mediated by **hydrogen bonds** and by **hydrophobic contacts**. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact.

- **Availability**
- Available free to academic institutions by anonymous ftp from: *ftp.biochem.ucl.ac.uk*.
- We also recommend you pick up the following:-
- HBPLUS - program for calculating hydrogen bonds and hydrophobic contacts for plotting by **LIGPLOT**.
- Het Group Dictionary - dictionary of Het Groups
- **NACCESS** - program for computing solvent accessible areas
- **Windows version**

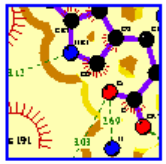
# Sample output

## Example 1: Chymotrypsin (8gch)

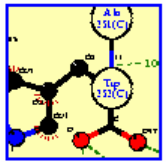
The **LIGPLOT** diagrams below illustrate the catalytic triad (**His 57**, **Asp 102** and **Ser 195**) in the active site of the serine protease **chymotrypsin**. The ligand bound is a 3-residue inhibitor **Gly-Ala-Trp**. The plots show the ligand's **Trp 252** residue nestling in the highly hydrophobic specificity pocket of the enzyme's active site.



Standard colour **LIGPLOT**. (Black-and-white version).



With atomic accessibilities calculated by **NACCESS**.



Schematic peptide representation.

# Introduction

- The program automatically generates **schematic diagrams** of protein-ligand interactions from the 3D coordinates in a **PDB** file.

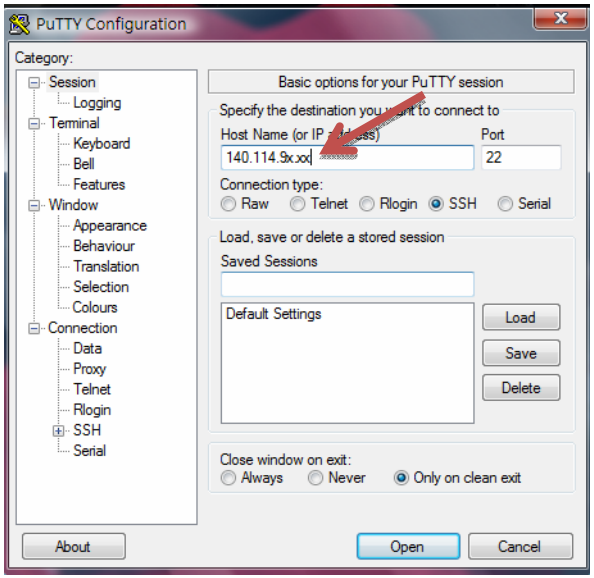
- **LIGPLOT algorithm:**
  - In principle, it reads in the 3D structure of the ligand from the **PDB** file, together with the protein residues it interacts with, and `unrolls' each object about its rotatable bonds, flattening them out onto the 2D page.

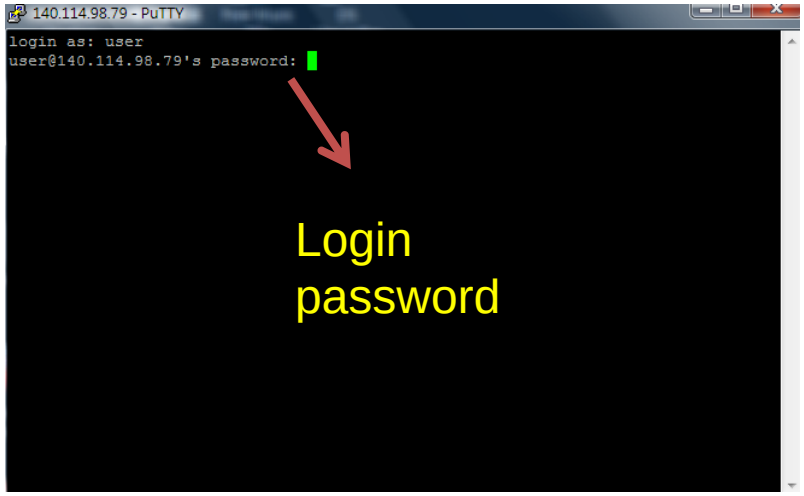
- By default, **LIGPLOT** expects the hydrogen bonds and nonbonded contacts to be calculated by the **HBPLUS** program and can read the files output by that program.

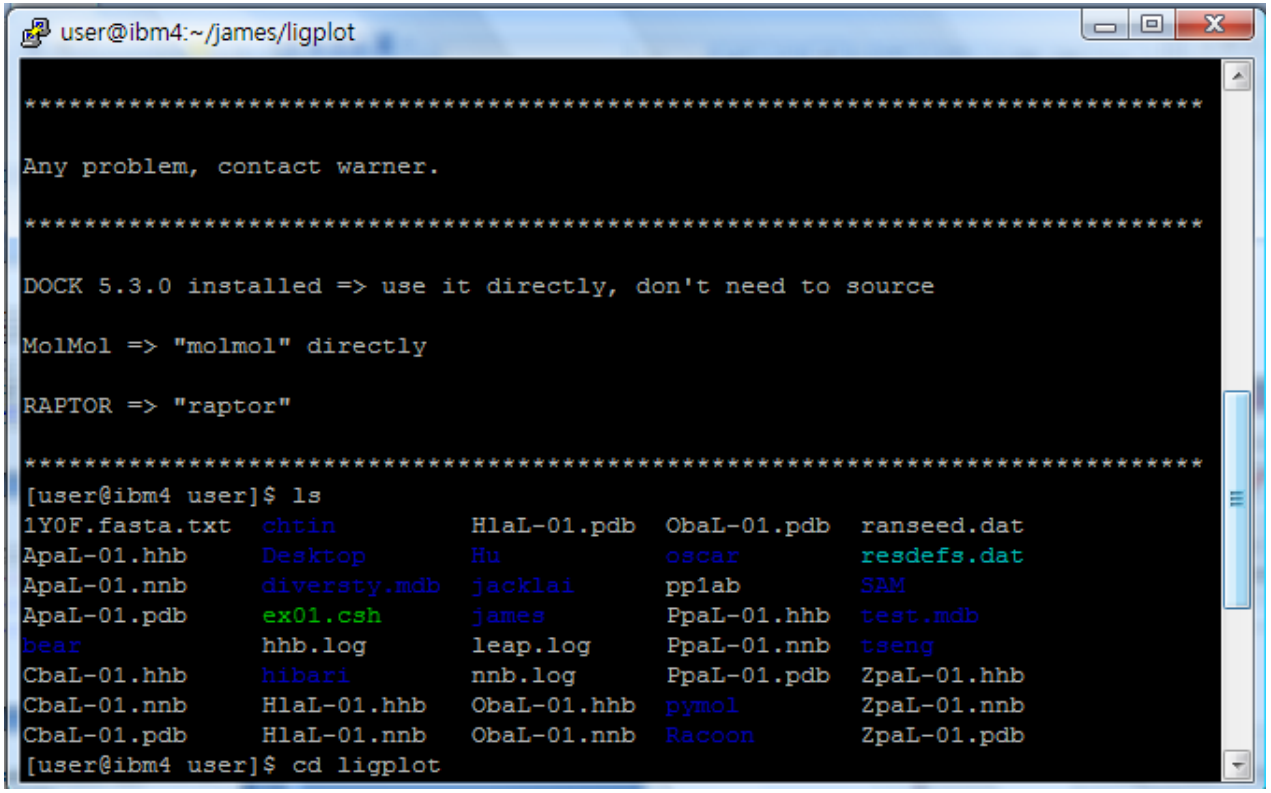
- The major drawback of **HBPLUS** is unable to recognize the majority of ligands in the **PDB**. As a result, it may miss certain hydrogen bonds between protein and ligand, and then **LIGPLOT** will not plot these absent interactions.

- **Het Group Dictionary (HBADD)** aims to cut out the manual effort of creating the input file for **HBPLUS**. The **HBADD** program identifies all the HETATM groups in your **PDB** file and searches for them in the **Het Group Dictionary**, available from the **PDB** in either PDB format:


putty.exe


PuTTY Configuration
Category:
Session
Logging
Terminal
Keyboard
Bell
Features
Window
Appearance
Behaviour
Translation
Selection
Colours
Connection
Data
Proxy
Telnet
Rlogin
SSH
Serial
Basic options for your PuTTY session
Specify the destination you want to connect to
Host Name (or IP address)
Port
140.114.9x.xx
22
Connection type:
Raw
Telnet
Rlogin
SSH
Serial
Load, save or delete a stored session
Saved Sessions
Default Settings
Load
Save
Delete
Close window on exit:
Always
Never
Only on clean exit
About
Open
Cancel

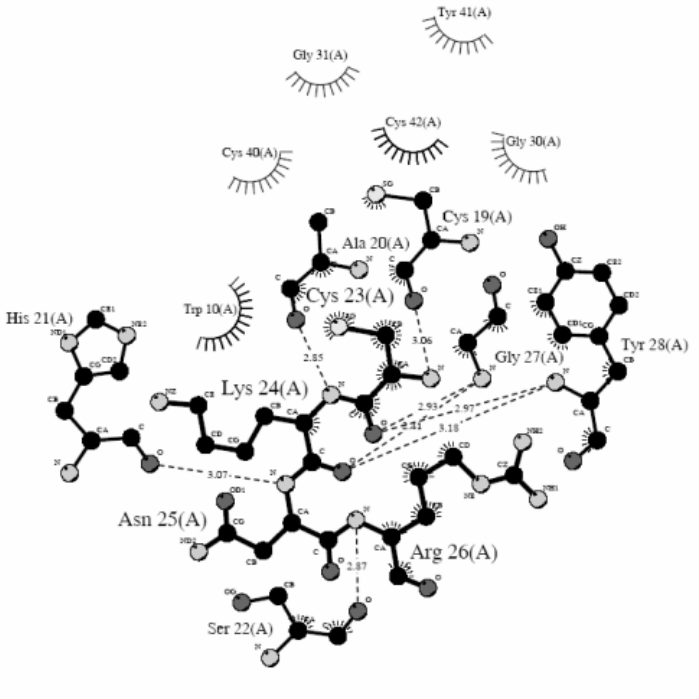

140.114.98.79 - PuTTY
login as: user
user@140.114.98.79's password:
Login
password


user@ibm4:~/james/ligplot
**********
Any problem, contact warner.
**********
DOCK 5.3.0 installed => use it directly, don't need to source
MolMol => "molmol" directly
RAPTOR => "raptor"
**********
[user@ibm4 user]$ ls
1YOF.fasta.txt chtin HlaL-01.pdb ObaL-01.pdb ranseed.dat
ApaL-01.hhb Desktop Hu oscar resdefs.dat
ApaL-01.nnb diversty.mdb jacklai pp1ab SAM
ApaL-01.pdb ex01.csh james PpaL-01.hhb test.mdb
bear hhb.log leap.log PpaL-01.nnb tseng
CbaL-01.hhb hibari nnb.log PpaL-01.pdb ZpaL-01.hhb
CbaL-01.nnb HlaL-01.hhb ObaL-01.hhb pymol ZpaL-01.nnb
CbaL-01.pdb HlaL-01.nnb ObaL-01.nnb Racoon ZpaL-01.pdb
[user@ibm4 user]$ cd ligplot

# 2. How to run LIGPLOT

The script file that runs **LIGPLOT** assumes the following:

1. That you have installed the **HBPLUS** program in accordance with the Installation Instructions.
2. That you have installed the **Het Group Dictionary**, as **het_dictionary.txt** or as **components.cif**, in the **LIGPLOT** directory
3. LIGPLOT program, **ligplot.scr** can automatically run **HBPLUS** prgram to generate two files
   *filename***.hhb** and *filename***.nnb**



# 2. How to run LIGPLOT

a. Running LIGPLOT (ligplot + hbplus)

To run LIGPLOT, type the following:

**ligplot filename [residue1] [residue2] [chain_id] [-w] [-m]**

**[user@ibm4 ligplot]$ligplot 1TI5_removeSG.pdb 23 26 A**

filename: ← your protein structure

[residue1] and [residue2]: ← its first and last residues

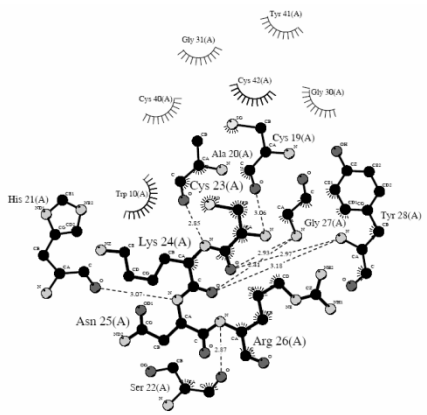
ex: 18 20 ← residue number

NAG 18 MAN 20 ← residue name and residue number

[chain_id]: ← must be entered unless ligand's chain is blank.

[-w]: ← ligand is a water molecule

[-m]: ← ligand is a single metal ion



# Inputs to LIGPLOT

- The **input** files:
  - *filename*.**pdb**
    Input **PDB** file holding the coordinates of the protein and ligand.
  - *filename*.**hhb**
    List **hydrogen-bonds** in structure. (generated by **HBPLUS** program)
  - *filename*.**nnb**
    List **nonbonded contacts** in structure. (generated by **HBPLUS** program)
  - **ligplot.prm**
    Parameter file to govern the final appearance of the plot.

- Note: should **not** use the *filename* "**ligplot.pdb**" as the input file

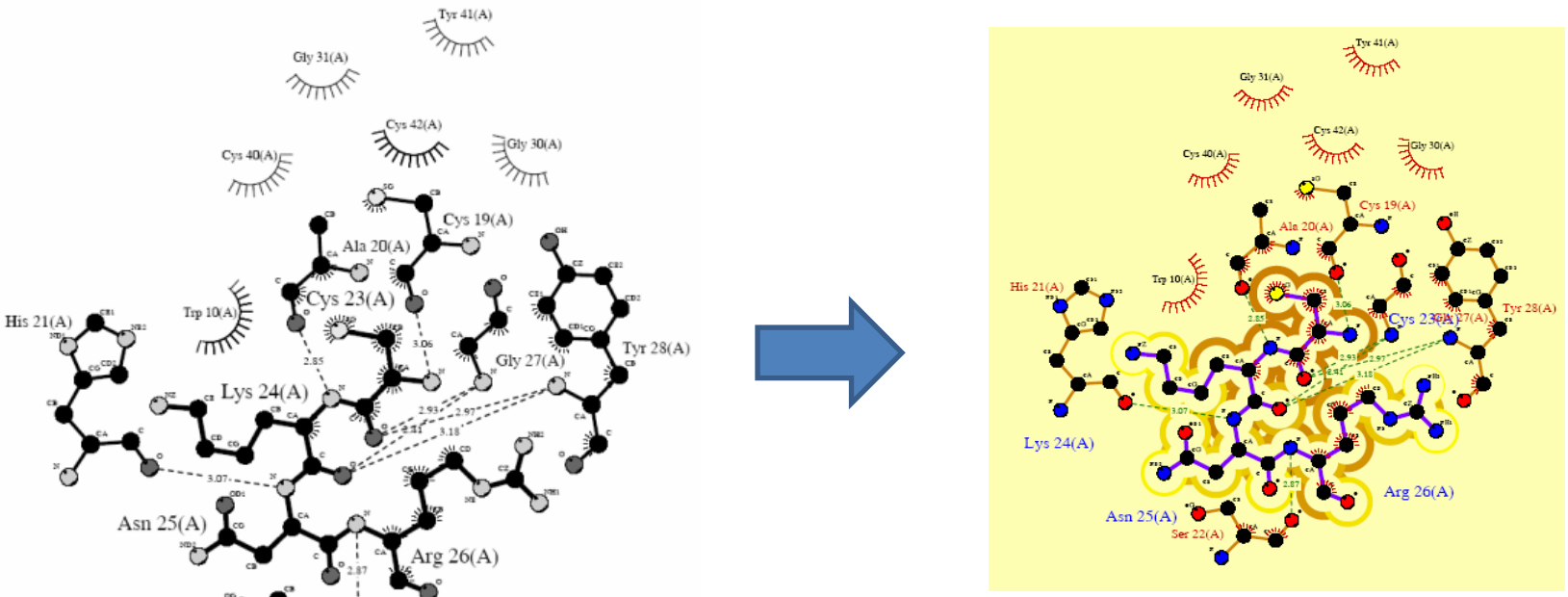
# Edit parameters

[user@ibm4 ligplot]$vi ligplot.prm

```
LIGPLOT v.4.0  -  Parameter file (ligplot.prm)
--------------------------------

PRINT OPTIONS
-------------
N     <- Produce a colour PostScript file (Y/N)?
P     <- Orientation of plot: (P)ortrait or (L)andscape?
0.0      <- Rotation angle (clockwise) for final plot

PLOT PARAMETERS
---------------
Y     <- Include: Hydrophobic interactions - (Y/N)?
Y     <- Include: Water molecules - (Y/N)?
Y     <- Include: Non-ligand mainchain atoms - (Y/N)?
Y     <- Include: Linked residues listed below - (Y/N)?
Y     <- Plot: Hydrogen bonds - (Y/N)?
N     <- Plot: Internal H-bonds in ligand - (Y/N)?
N     <- Plot: External groups covalently bonded to ligand - (Y/N)?
N     <- Plot: Bonds showing hydrophobic interactions - (Y/N)?
N     <- Plot: Schematic ligand representation [see Note 1] - (Y/N)?
N     <- Plot: Schematic non-ligand residues [see Note 1] - (Y/N)?
N     <- Plot: Accessibility shading [see Note 2] - (Y/N)?
Y     <- Plot: Ligand atoms (as spheres) - (Y/N)?
Y     <- Plot: Nonligand atoms (as spheres) - (Y/N)?
Y     <- Plot: Double- and triple bonds (for ligplot.pdb only) - (Y/N)?
Y     <- Print: Key to symbols in PostScript output - (Y/N)?
Y     <- Print: Residue names/numbers - (Y/N)?
Y     <- Print: Atom names - (Y/N)?
Y     <- Print: H-bond lengths on hydrogen bonds - (Y/N)?
Y     <- Print: Filename as title if title not explicitly defined - (Y/N)?
Y     <- Plot: Solid lines for covalent bonds to external groups - (Y/N)?
0     <- Non-bonded contacts option [see Note 3]
Y     <- Plot: Water atoms (as spheres) - (Y/N)?
Y     <- Plot: Accessibility shading for the ligand only - (Y/N)?
```

# 2. How to run LIGPLOT

```
[user@ibm4 ligplot]$ligplot 1TI5_removeSG.pdb 23 26 A
```



# Outputs produced by LIGPLOT

- The **output** files:
- **ligplot.ps** - Colour or black-and-white **PostScript**
- **ligplot.pdb** - Output file, in **PDB** format, of the final flattened molecules (ligand and interacting protein residues) as shown in the plot.
- **ligplot.hhb** - Output file of just those hydrogen bonds in the original *filename***.hhb** file that were used by **LIGPLOT** in producing the final picture
- **ligplot.nnb** - Output file of just those hydrophobic contacts in the original *filename***.nnb** file that were used by **LIGPLOT** in producing the final picture
- **ligplot.bonds** - Output file listing of bonds and bond-types in the final **LIGPLOT** picture.
- **ligplot.frm** - Output file, in **PDB** format, of the molecules. You can view this file using **Rasmol** to see only those residues that interact with the ligand.
- **ligplot.rcm** - Output file listing all the residues on the plot.
- **ligplot.drw** - File for input to the Java-based LIGPLOT editor, **LigEd**.

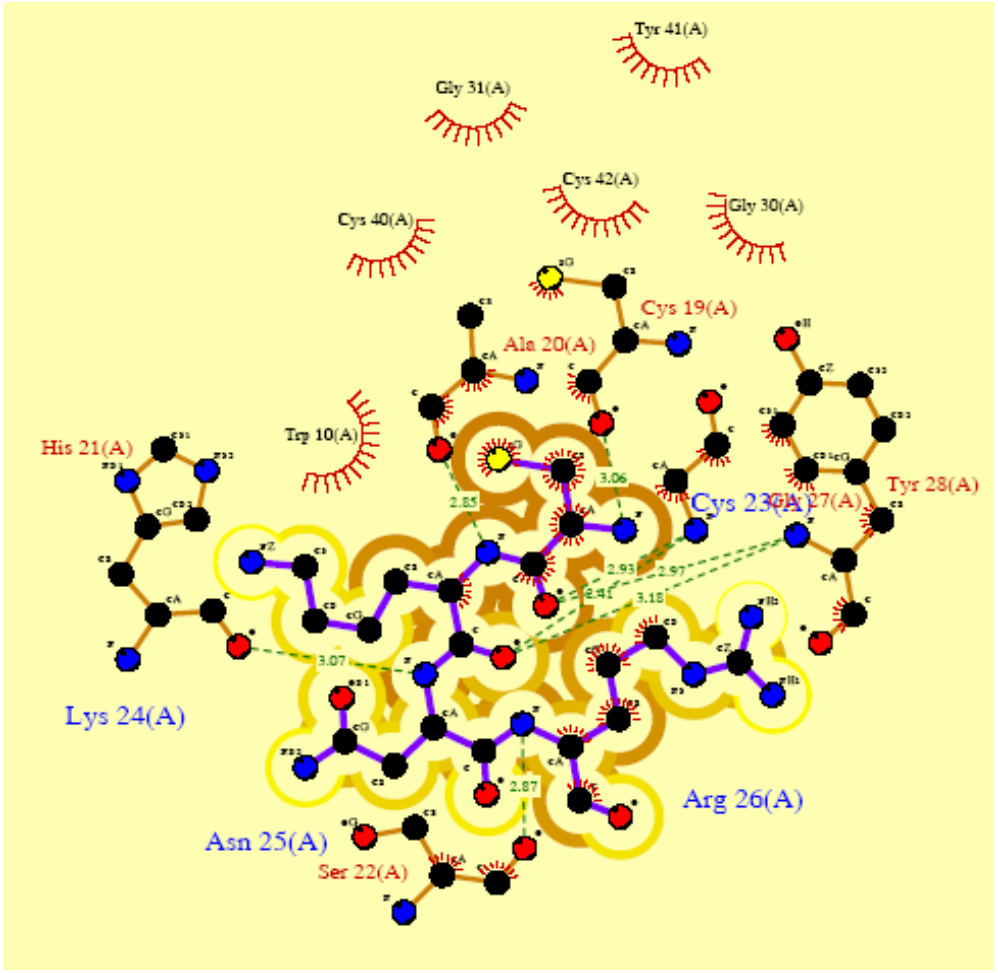
# PostScript （**.ps**）

## Ghostscript, Ghostview and GSview

- http://pages.cs.wisc.edu/~ghost/index.htm

| PostScript | (.ps) | Ghostscript, Ghostview and GSview |
|---|---|---|
| 1 .點選上方連結，可連至 Ghostscript, Ghostview and GSview 的網頁。 | | |
| 2 .點選 Obtaining AFPL Ghostscript 7.04 連結，前往下載 Ghostscript 7.04 的網頁。 | | |
| (1)將網頁往下拉至 Windows 95, 98, ME, NT, 2000 or XP | | |
| (2)下載 gs704w32.exe<br>[下載至您的電腦裡，請先進行安裝後，再執行下列步驟] | | |
| 3 .請再點選上方連結，連至 Ghostscript, Ghostview and GSview 的網頁。 | | |
| 4 .點選 GSview 4.3 連結，前往下載 GSview 4.3 的網頁。 | | |
| (1)下載 gsv43w32.exe<br>[即可下載至您的電腦裡，進行安裝後即可觀看PostScript檔] | | |
| | | |

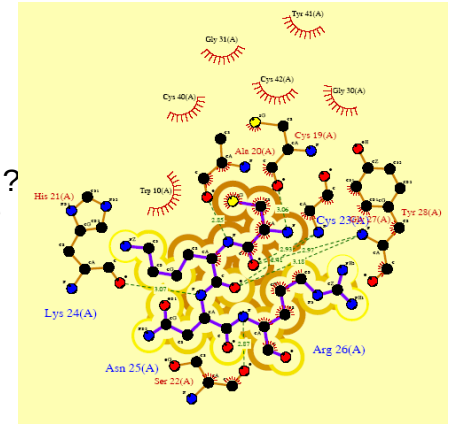
# Ligplot.ps


Tyr 41(A)
Gly 31(A)
Cys 42(A)
Cys 40(A)
Gly 30(A)
Cys 19(A)
Ala 20(A)
Trp 10(A)
His 21(A)
Tyr 28(A)
Cys 23(A)
Lys 24(A)
Arg 26(A)
Asn 25(A)
Ser 22(A)
3.06
2.85
2.93
2.97
2.41
3.18
3.07
2.87

# b. Showing atom solvent accessibilities computed by NACCESS
# < Naccess (xxxx.asa) + Ligplot >

Edit parameters

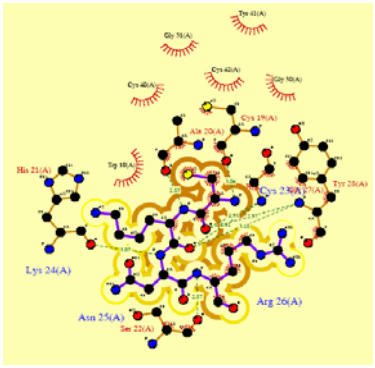
```
PLOT PARAMETERS
---------------
Y     <- Include: Hydrophobic interactions - (Y/N)?
Y     <- Include: Water molecules - (Y/N)?
Y     <- Include: Non-ligand mainchain atoms - (Y/N)?
Y     <- Include: Linked residues listed below - (Y/N)?
Y     <- Plot: Hydrogen bonds - (Y/N)?
N     <- Plot: Internal H-bonds in ligand - (Y/N)?
N     <- Plot: External groups covalently bonded to ligand - (Y/N)?
N     <- Plot: Bonds showing hydrophobic interactions - (Y/N)?
N     <- Plot: Schematic ligand representation [see Note 1] - (Y/N)?
N     <- Plot: Schematic non-ligand residues [see Note 1] - (Y/N)?
Y     <- Plot: Accessibility shading [see Note 2] - (Y/N)?
Y     <- Plot: Ligand atoms (as spheres) - (Y/N)?
Y     <- Plot: Nonligand atoms (as spheres) - (Y/N)?
Y     <- Plot: Double- and triple bonds (for ligplot.pdb only) - (Y/N)?
Y     <- Print: Key to symbols in PostScript output - (Y/N)?
Y     <- Print: Residue names/numbers - (Y/N)?
Y     <- Print: Atom names - (Y/N)?
Y     <- Print: H-bond lengths on hydrogen bonds - (Y/N)?
Y     <- Print: Filename as title if title not explicitly defined - (Y/N)?
Y     <- Plot: Solid lines for covalent bonds to external groups - (Y/N)?
0     <- Non-bonded contacts option [see Note 3]
Y     <- Plot: Water atoms (as spheres) - (Y/N)?
Y     <- Plot: Accessibility shading for the ligand only - (Y/N)?
```




Orange: Buried atoms
YELLOW: Accessible atoms

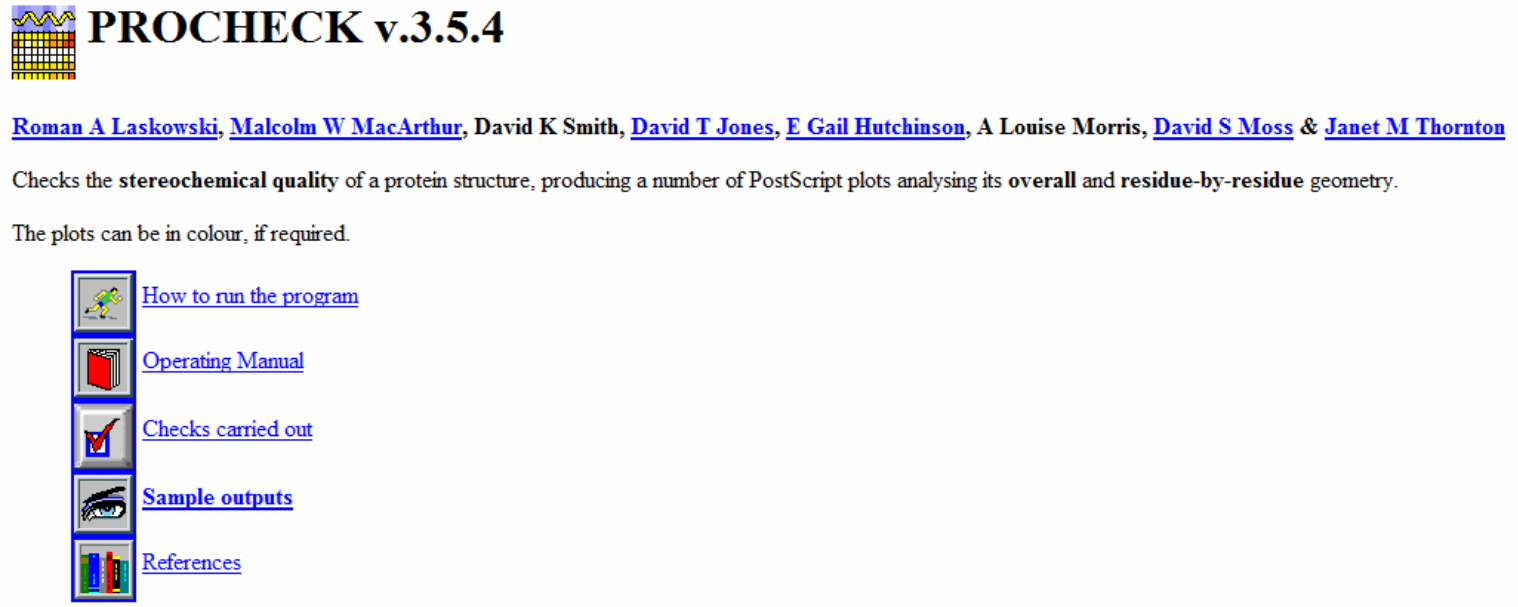
# < Naccess (xxxx.asa) + Ligplot >

```
[user@ibm4 ligplot]$./naccess 1TI5_removeSG.pdb -h –f
-f: "full" output format
-h: HETATMs are to be included in the accessibility
calculations

[user@ibm4 naccess2.1.1]$ ls
1TI5_removeSG.asa 1TI5_removeSG.pdb 1TI5_removeSG.log
1TI5_removeSG.rsa (for HADOCK)

[user@ibm4 ligplot]$ligplot 1TI5_removeSG.asa 23 26 A
```



# Procheck

# Protein Structure Analysis




http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html

# Introduction

- Programs to check the Stereochemical Quality of Protein Structures
- The aim of PROCHECK is to assess both the overall stereochemical quality of a given protein structure and to give an indication of its local, residue-by-residue reliability.

- The checks also make use of “ideal” bond lengths and bond angles, as derived from CSD Database- now numbering over 100,000 structures.

- The PROCHECK programs produce a number of plots, together with a detailed residue-by-residue listing.

- The input to PROCHECK is a single PDB file

# Availability

- Available by anonymous ftp on: *ftp.biochem.ucl.ac.uk* Source code can be picked up from:
  - **pub/procheck/tar3_5**
    **pub/procheck/source3_5**
  - **pub/procheck/tar3_5/manual.tar.Z**
  - Users must sign a Confidentiality Agreement and post or fax it to:-
  - *Roman Laskowski*
    *European Bioinformatics Institute,*
    *Wellcome Trust Genome Campus,*
    *Hinxton,*
    *Cambridge, CB10 1SD,*
    *United Kingdom*
    **Fax:-** +44 (0)1223 494 468

- **Note:** A version of the **PROCHECK** programs running under **Windows NT** has been prepared by **Bernhard Rupp** of the Lawrence Livermore National Laboratory and is available by anonymous ftp from *http://ruppweb.dyndns.org/ftp_warning.html*.

# Procheck programs

- **CLEAN - cleaning PDB file**
  - corrects any mislabelled atoms and creates a new coordinates file (XXX.new)
- **SECSTR - assigning secondary structure**
- **NB - identifying non-bonded interactions**
- **ANGLEN - calculating bond lengths and bond angles**
- **TPLOT, PPLOT, BPLOT - graphical output**

# Running Procheck

To run the program on a PDB file, type

**procheck** *filename [chain] resolution*

where

*filename* = the coordinates file in Brookhaven format
*[chain]* = an optional one-letter chain-ID
*resolution* = a real number giving the resolution of the structure

For example:-

**procheck /data/pdb/p1amt.pdb A 1.5**

Prerequisites:-

The following environment variables and aliases must be included in your .cshrc file:-

```
# PROCHECK
# -------
set    prodir = /procheck_directory
setenv prodir  '/procheck_directory'
alias  procheck  $prodir'/procheck.scr'
alias  proplot   $prodir'/proplot.scr'
alias  gfac2pdb  $prodir'/gfac2pdb.scr'
```

where **/procheck_directory** is the name of the directory holding the **PROCHECK** scripts, executables and data files.

# Input requirements

- The only input required for **PROCHECK** is the **PDB file** holding the coordinates of the structure of interest. For **NMR** structures, each model in the ensemble should be separated by the correct **MODEL** and **ENDMDL** records. Only the first model will be analysed. A separate program, **PROCHECK-NMR** deals specifically with the analysis of **NMR** structures.

# PROCHECK outputs

- **PostScript files (XXX.ps)**
- The plots show each of the different **PROCHECK** analyses generated by the run.
- **Residue-by-residue listing (xxx.out)**
- The residue-by-residue lists all the computed stereochemical properties, by residue, in a printable ASCII text file.
- **Other output files**
- xxx.lan: Main-chain bond lengths and bond angles used by the plotting programs
- .nb List of atom-pairs making near-neighbour contacts
- .new "Cleaned-up" version of the original coordinates file
- .pln Coordinates of atoms in planar groups
- .rin Residue information used by the plotting programs
- .sco Main-chain and side-chain properties
- .sdh Residue-by-residue G-factors
  - **.new file** The .new file holds the `cleaned-up' version of the original **PDB** file, with any wrong atom-labels corrected in accordance with the **IUPAC** naming conventions
  - **.sum file** The .sum file gives a short summary of the overall **PROCHECK** RESULTS.
- **Log files**
- Each program in the suite also produces its own log file. Should the **PROCHECK** suite crash, or give strange-looking results, these log files should be the first place you look for a reason for the problem. The **7** files are:
- anglen.log clean.log pplot.log tplot.log bplot.log nb.log secstr.log

# Procheck output

- a. Ramachandran plot quality – percentage of the protein's residues that are in the core regions of the Ramachandran plot.
- b. Peptide bond planarity – standard deviation of the protein structure's omega torsion angles.
- c. Bad non-bonded interactions – number of bad contacts per 100 residues.
- d. C**a tetrahedral distortion – standard** deviation of the **z torsion angle (Ca, N, C,** and C**b).**
- e. Main-chain hydrogen bond energy - standard deviation of the hydrogen bond energies for main-chain hydrogen bonds.
- f. Overall G-factor - average of different Gfactors for each residue in the structure.

# Procheck output