蛋白質結構模擬
Protein Structure Modeling

呂平江

國立清華大學

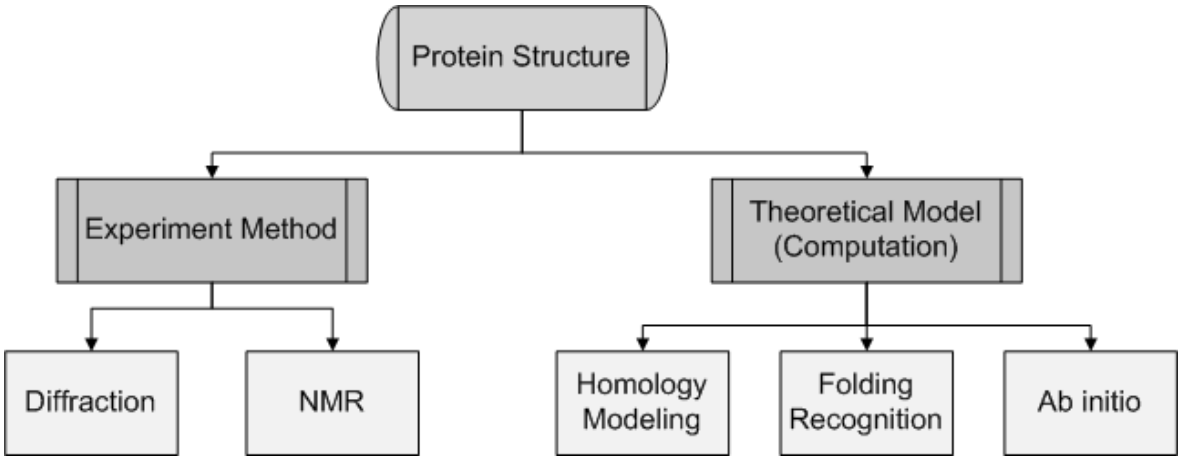
生命科學系/生物資訊與結構生物研究所

2012/06/27

Significance of Protein Structure

What does a structure offer in the way of biological knowledge?

- Location of mutants and conserved residues
- Ligand and functional sites
- Evolutionary Relationships
- Mechanisms



Experimental Method

目前利用實驗方法得到蛋白質的結構有下列兩種方法，但是這些方法有個共同的缺點，就是需要較多的時間，才能得到結構。

方法	原理	缺點
X-光繞射 (X-ray Diffraction)	利用X-光線繞射的特性，對已經結晶的蛋白質進行繞射實驗，然後再將所得到的數據加以分析，就可以得到結構 。	其缺點是要得到蛋白質的結晶是實驗中最重要也是最困難的。
核磁共振方法(NMR)	利用核磁共振現象得到蛋白質的結構。並不是所有原子核都能產生核磁共振的現象。一般常用來偵測的對象包含了^{1}H、^{13}C和^{15}N。	其一方法的缺點是只能對小的蛋白質來作分析。

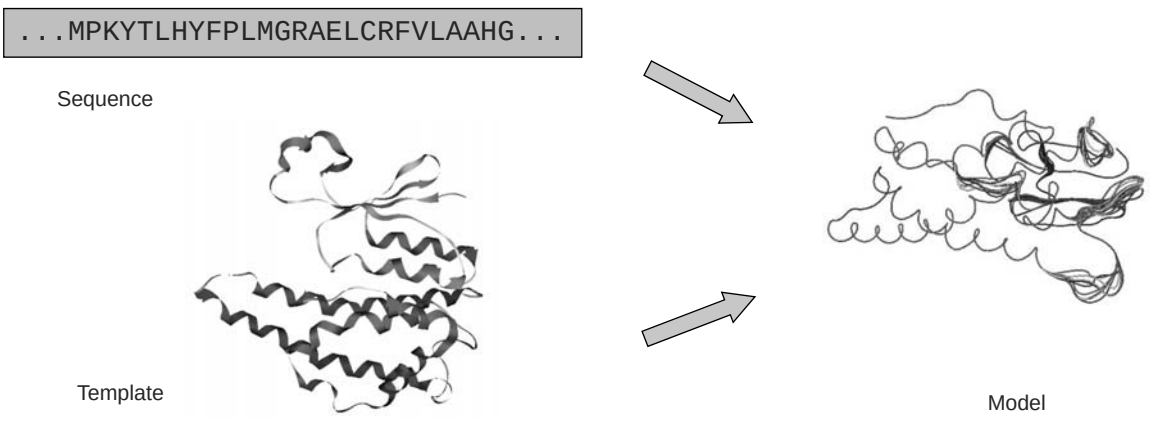
Theoretical Model

- 因此，科學家就想利用電腦計算方式來獲得蛋白質結構，以加快定出蛋白質三度空間結構的速率。由於目前已知資料庫中，完整地蛋白質序列數量已遠遠超過所解出的蛋白質結構數量，而且利用實驗方法解出蛋白質結構不是太耗時間，就是受到一些限制而無法定出結構，因此，利用電腦計算的方式來計算出蛋白質結構已逐漸受到矚目。大致可分為下列三種方式：

- (1) **homology or comparative modeling**; (2) **fold recognition (threading)**; (3) ***ab initio* techniques**.
- Homology or comparative modeling is currently the most accurate method to predict the three-dimensional structure of proteins.

Homology Modeling

- Given: protein sequence
- Aim: model of the 3D structure of the target protein
- Approach: use homologous proteins as templates



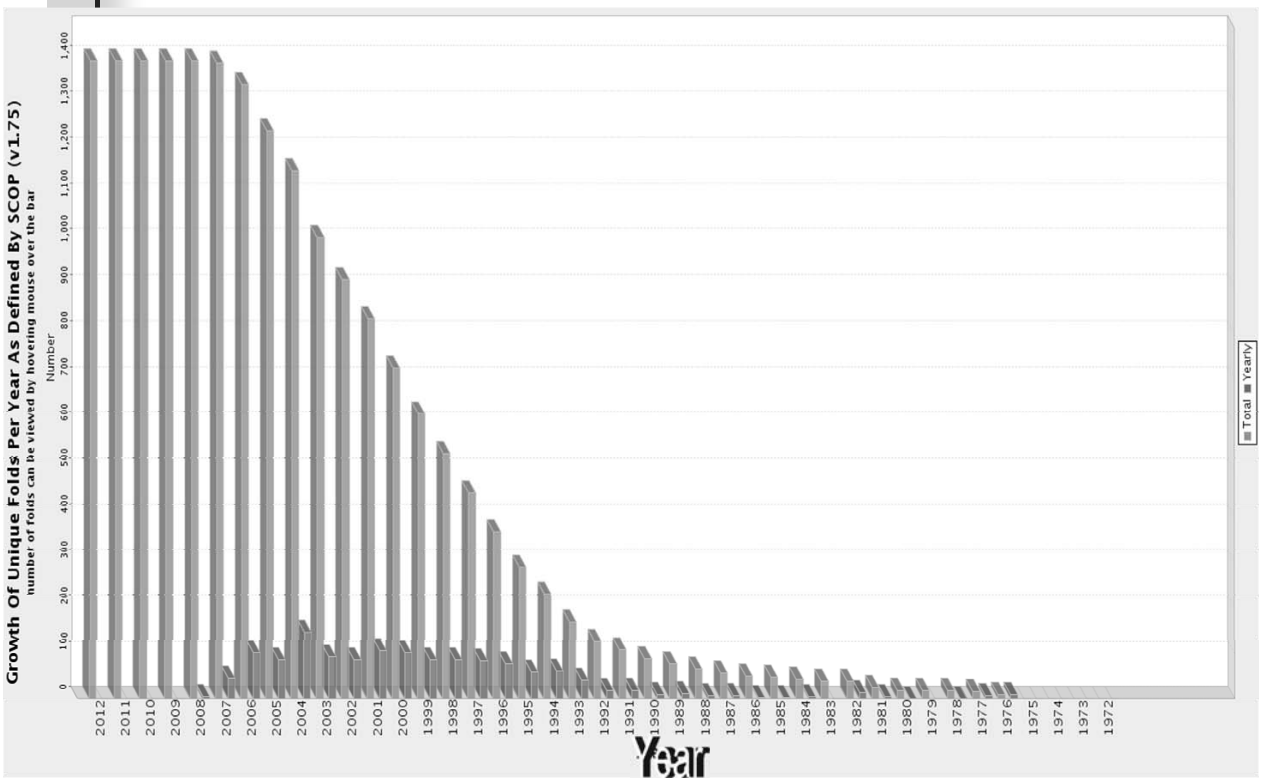
Use of Homology Modeling

- What can a homology model provide …
  - study patterns of conservation
  - spatial proximity of residues to known active sites
  - surface exposure of residues
- … and what not?
  - atomic details of protein geometry
  - exact loop and side chain conformations
  - local shape
  - protein flexibility

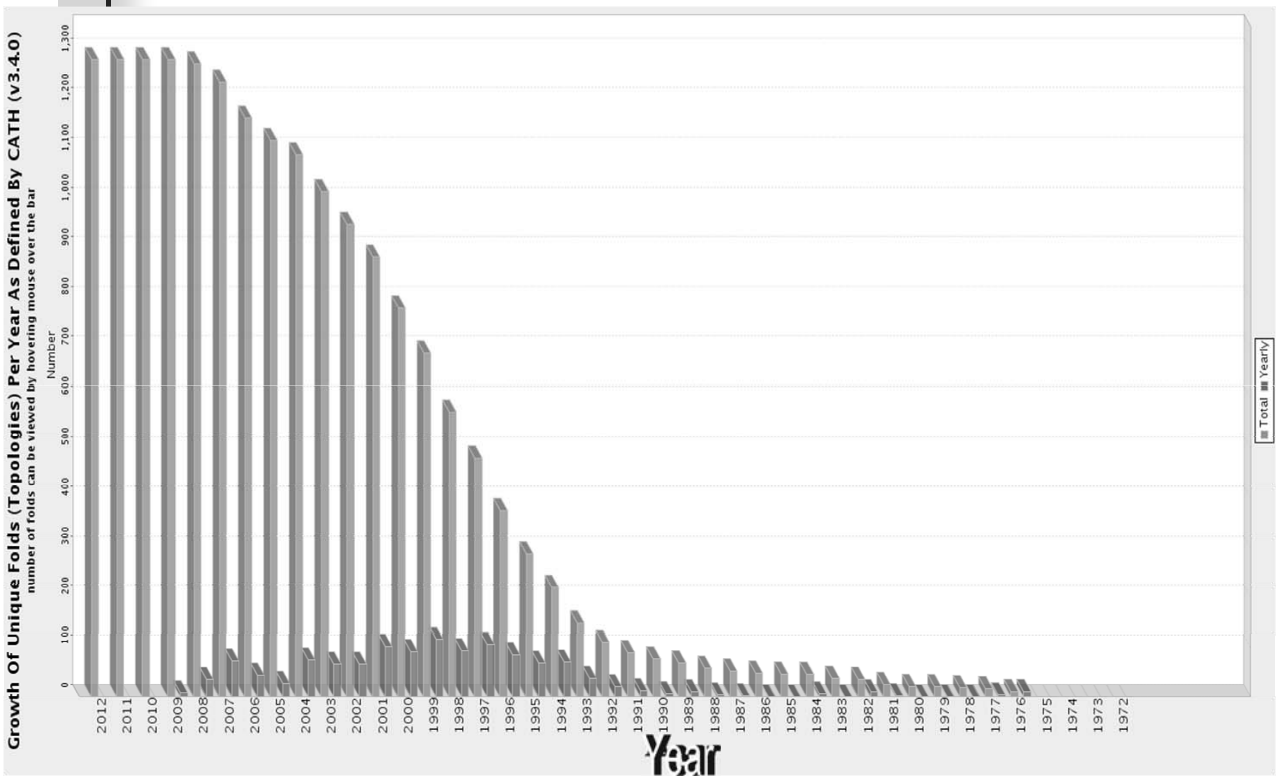
Computational Determination and Analysis

- **Databases**
  - CATH (Class, Architecture, Topology, Homologous superfamily)
  - SCOP (Structural Classification Of Proteins)
  - FSSP (Fold classification based on Structure-Structure alignment of Proteins)
- **Prediction**
  - *ab*-initio, theoretical modeling, and conformation space search
  - Homology modeling and threading
  - Energy minimization, simulation and Monte Carlo

Growth of Unique Folds Per Year (SCOP)

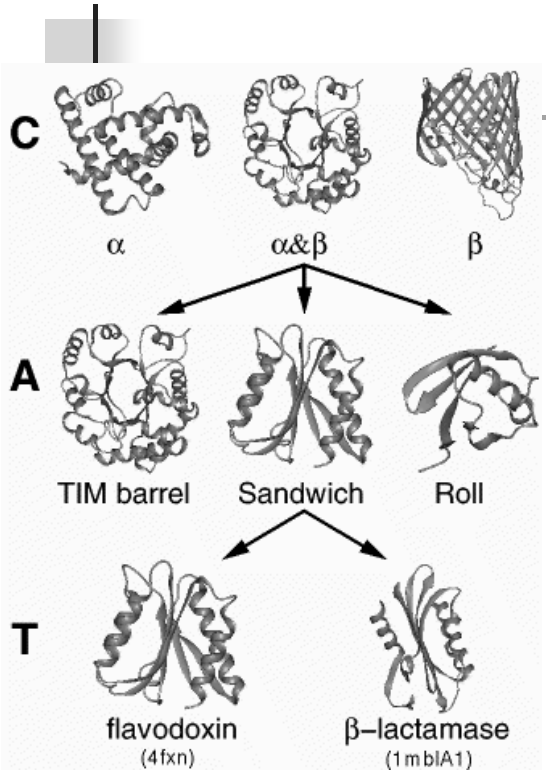


Growth of Unique Folds Per Year (CATH)



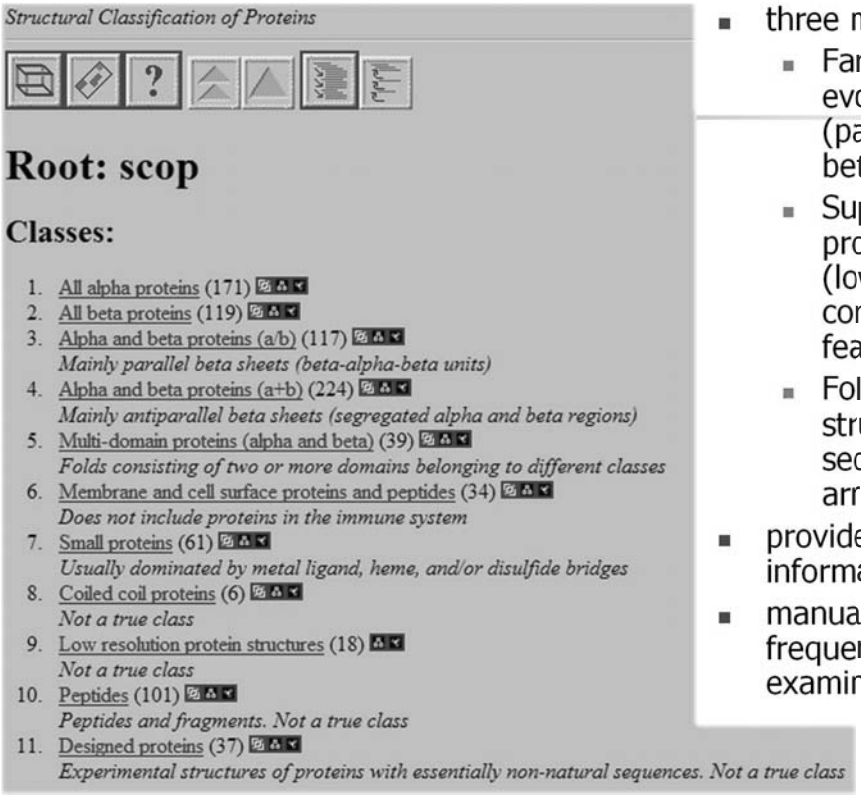
CATH

http://www.cathdb.info/



- a combination of manual and automated hierarchical classification
- four major levels:
  - **Class (C) – based on secondary structure content**
  - **Architecture (A) – based on gross orientation of secondary structures**
  - **Topology (T) – based on connections and numbers of secondary structures**
  - **Homologous superfamily (H) – based on structure/function evolutionary commonalities**
- provides useful geometric information (e.g. architecture)
- partial automation may result in examples near fixed thresholds being assigned inaccurately

http://scop.mrc-lmb.cam.ac.uk/scop/



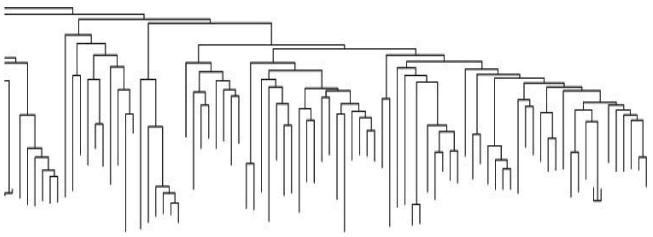
- a purely manual hierarchical classification
- three major levels:
  - Family – based on clear evolutionary relationship (pairwise residue identities between proteins are >30%)
  - Superfamily – based on probable evolutionary origin (low sequence identity but common structure/function features
  - Fold – based on major structural similarity (major secondary structures in same arrangement and topology
- provides detailed evolutionary information
- manual process influences update frequency and equally exhaustive examination

FSSP http://ekhidna.biocenter.helsinki.fi/dali

- a purely automated
- hierarchical classification
- three major levels:
  - representative set – 330 protein chains (less than 30% sequence identity)
  - clustering – based on structural alignment into fold families
  - convergence – cutting at a high statistical significance level increases the number of distinct families, gradually approaching one family per protein chain
- continually updated, presents data and lets user assess
- Without sufficient knowledge, user may not assess data appropriately

Family index	PDB-code	Alignments	compound	
1.1.1.1.1.1	1aak	alignment	"UBIQUITIN CONJUGATING ENZYME (E.C.6.3.2.19)"	
2.1.1.1.1.1	1shaA	alignment	"V-SRC TYROSINE KINASE TRANSFORMING PROTEIN (PHOSPHOTYROSINE"	
3.1.1.1.1.1	1rblM	alignment	"RIBULOSE 1,5 BISPHOSPHATE CARBOXYLASE/OXYGENASE"	
4.1.1.1.1.1	1daaA	alignment	"2 MOLECULE: D-AMINO ACID AMINOTRANSFERASE;	1DAA 6"
5.1.1.1.1.1	3mdsA	alignment	"MANGANESE SUPEROXIDE DISMUTASE (E.C.1.15.1.1)"	
6.1.1.1.1.1	1hmx	alignment	"MIV-1 MATRIX PROTEIN (MIV-1 MA, MIVP17, P17, MA)"	
7.1.1.1.1.1	1gky	alignment	"GUANYLATE KINASE (E.C.2.7.4.8) COMPLEX WITH"	
7.1.1.1.1.2	1ukz	alignment	"URIDYLATE KINASE (E.C.2.7.4.-) COMPLEXED WITH ADP AND AMP"	
7.1.1.1.1.2	1aky	alignment	"ADENYLATE KINASE"	
8.1.1.1.1.1	1ordA	alignment	"2 MOLECULE: ORNITHINE DECARBOXYLASE;	1ORD 4"
8.1.1.1.1.2	1tplA	alignment	"TYROSINE PHENOL-LYASE (E.C.4.1.99.2)"	
8.1.1.1.1.2	2dkb	alignment	"2,2-DIALKYLGLYCINE DECARBOXYLASE (PYRUVATE) (DGD)"	
8.1.1.1.1.2	7aatA	alignment	"ASPARTATE AMINOTRANSFERASE (E.C.2.6.1.1) COMPLEX WITH"	
8.2.1.1.1.1	1hmpB	alignment	"HYPOXANTHINE GUANINE PHOSPHORIBOSYLTRANSFERASE (HGPRTASE)"	
8.2.1.1.2.1	1gph4	alignment	"GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE (PRPP)"	
8.2.1.1.2.2	1asc	alignment	"OROTATE PHOSPHORIBOSYLTRANSFERASE (OPRTASE) (E.C.2.4.2.10)"	
8.3.1.1.1.1	2reb	alignment	"REC A PROTEIN (E.C.3.4.99.37)"	
8.3.1.2.1.1	1poxA	alignment	"PYRUVATE OXIDASE (E.C.1.2.3.3) MUTANT WITH PRO 178"	
8.3.1.2.1.1	1trkA	alignment	"TRANSKETOLASE (E.C.2.2.1.1)"	
8.3.1.2.1.1	1pvdA	alignment	"PYRUVATE DECARBOXYLASE (PDC) (E.C.4.1.1.1)"	
8.4.1.1.1.1	1nbaA	alignment	"N-CARBAMOYLSARCOSINE AMIDOHYDROLASE (E.C.3.5.1.59)"	
8.4.2.1.1.1	1ipd	alignment	"3-ISOPROPYLMALATE DEHYDROGENASE (E.C.1.1.1.85)"	
8.4.2.2.1.1	1phr	alignment	"MOLECULE: LOW MOLECULAR WEIGHT PHOSPHOTYROSINE PROTEIN"	
8.4.2.3.1.1	1esc	alignment	"ESTERASE"	
8.4.2.4.1.1	1dih	alignment	"2 MOLECULE: DIHYDRODIPICOLINATE REDUCTASE;	1DIH 4"
8.4.2.4.2.1	1dhr	alignment	"DIHYDROPTERIDINE REDUCTASE (DHPR) (E.C.1.6.99.10) COMPLEX"	
8.4.2.4.2.1	1hdoA	alignment	"3-ALPHA, 20-BETA-HYDROXYSTEROID DEHYDROGENASE (E.C.1.1.1.53)"	
8.4.2.4.2.1	1udpB	alignment	"URIDINE DIPHOSPHOGALACTOSE 4-EPIMERASE (E.C.5.1.3.2)"	
8.4.2.4.3.1	1scuA	alignment	"SUCCINYL-COA SYNTHETASE (SUCCINATE-COA LIGASE)"	
8.4.2.4.3.2	1scuB	alignment	"SUCCINYL-COA SYNTHETASE (SUCCINATE-COA LIGASE)"	
8.4.2.4.4.1	3chy	alignment	"CHE-Y"	
8.4.2.4.4.2	1bmtA	alignment	"METHIONINE SYNTHASE (B12-BINDING DOMAINS) (E.C.2.1.1.13)"	
8.4.2.4.5.1	2ohxA	alignment	"ALCOHOL DEHYDROGENASE (HOLO FORM) (E.C.1.1.1.1) COMPLEX"	

list of representative set



clustering dendogram

ab-initio, theoretical modeling, and conformation space search

***ab*-initio** = given amino acid primary structure, i.e. sequence, derive structure from first principles (e.g. treat amino acids as beads and derive possible structures by rotating through all possible ϕ, ψ angles using a "reliable" energy function, then optimize globally)

Theoretical modeling = subset of *ab*-initio, given amino acid primary structure and knowledge about characteristic features, derive structure that has that structure and features
(e.g. protein has an iron binding site → possible heme substructure)

Conformation space search = subset of *ab*-initio, but a stochastic search in which the sample space is reduced by initial conditions/assumptions (e.g. reduce sample space to conform to Ramachandran plot)

Homology modeling and threading

Homology modeling = knowledge-based approach, given a sequence database, use multiple sequence alignment on this database to identify structurally conserved regions and construct structure backbone and loops based on these regions, restore side-chains and refine through energy minimization (apply to proteins that have high sequence similarity to those in the database)

Threading = knowledge-based approach, given a structure database of interest (e.g. one that provides a limited set of possible structures per given sequence for fold recognition, one that provides a one structure per given limited set of possible sequences for inverse folding) use scoring functions and correlations from this database to derive structure that is in agreement (apply to proteins with moderate sequence similarity to those in the database)

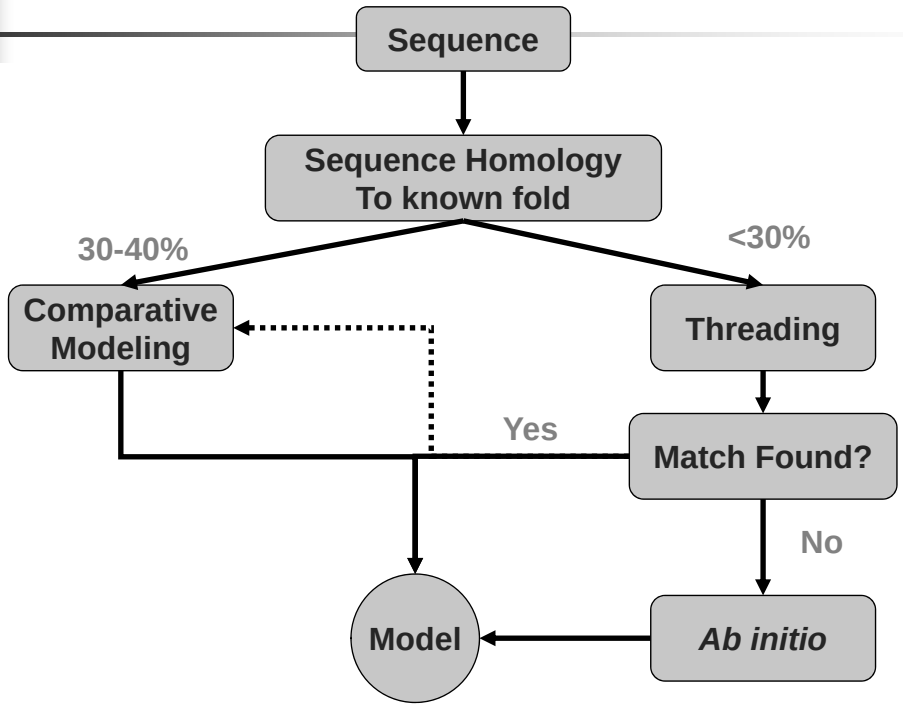
Energy minimization, simulation and Monte Carlo

Energy minimization = select an appropriate energy function and derive conformations that yield minimal energies based on this function

Simulation = select appropriate molecular conditions and derive conformations that are suited to these molecular conditions

Monte Carlo = subset of molecular simulation, but it is an iterated search through a Markov chain of conformations (many iterations → canonical distribution, P(particular conformation)~exp(-E/T)) proposed by N. Metropolis, in which a new conformation is generated from the current one by a small ``move'' and is accepted with a probability P_{acc} = min(1, exp(-ΔE/kT)), which depends on the corresponding change in energy, ΔE, and on an external adjustable parameter, kT

Protein Structure Prediction

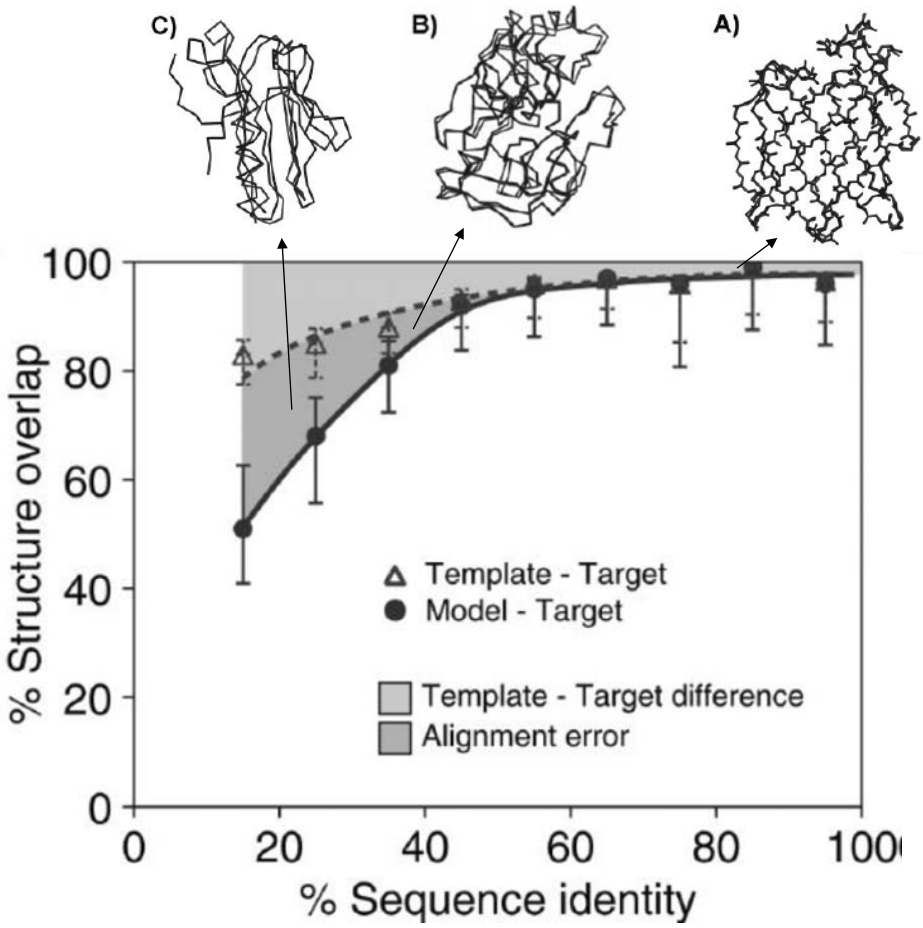


Homology Modeling (同源模擬法)

- **Homology modeling** = knowledge-based approach, given a sequence database, use multiple sequence alignment on this database to identify structurally conserved regions and construct structure backbone and loops based on these regions, restore side-chains and refine through energy minimization (apply to proteins that have high sequence similarity to those in the database)

- 同源性模擬法(Homology modeling)又可稱為比較性模擬法(comparative modeling)或知識基礎性模擬法(knowledge-based modeling)。因為這一個方法可以快速模擬出蛋白質的三度空間結構。因此，目現已有利用這個方法模擬出整個酵母菌(yeast)基因體的所有蛋白質結構。

- 同源性模擬法的基本假設是特定的胺基酸序列會構成特定的蛋白質結構。主要是利用現存已解出的結構為模板(由核磁共振(NMR)或是X光繞射(X-ray diffraction)所解出的結構)，模擬出未知結構蛋白質序列 (protein sequence)的三度空間結構。一般來說，當所要解的蛋白質序列(target protein)和模板(template)之間的序列相似度越高，所模擬出來的結構越正確也越可信。



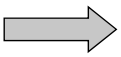
What is Homology Modeling?

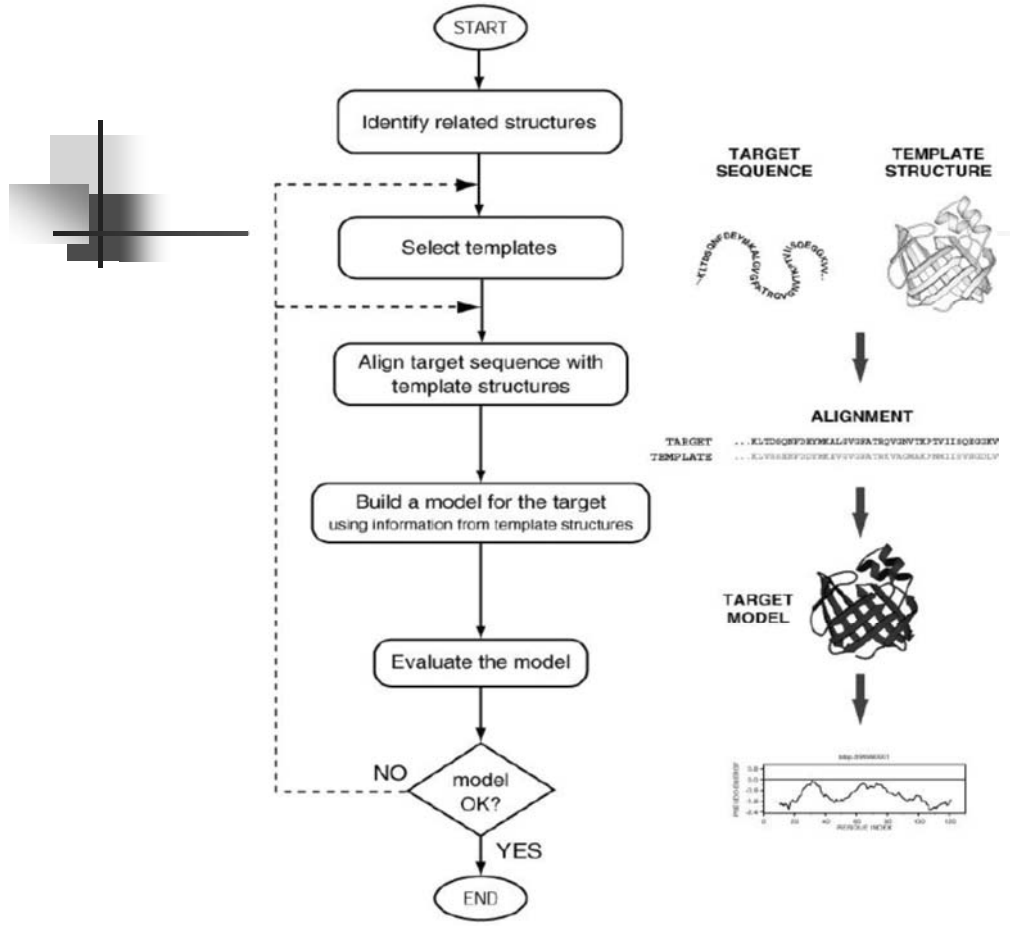
- Predicts the three-dimensional structure of a given protein sequence (TARGET) based on an alignment to one or more known protein structures (TEMPLATES)
- If similarity between the TARGET sequence and the TEMPLATE sequence is detected, structural similarity can be assumed.
- In general, 30% sequence identity is required for generating useful models.

Target

>AAT03210 glyoxalase family protein
MKIEHVALWTTNLEQMKQFYVTYFGATANDLYENKTKGFNSYFLSFEDG
ARLEIMSRTDVTGKTTGENLGWAHIAISTGTKEAVDELTEKLR

Template

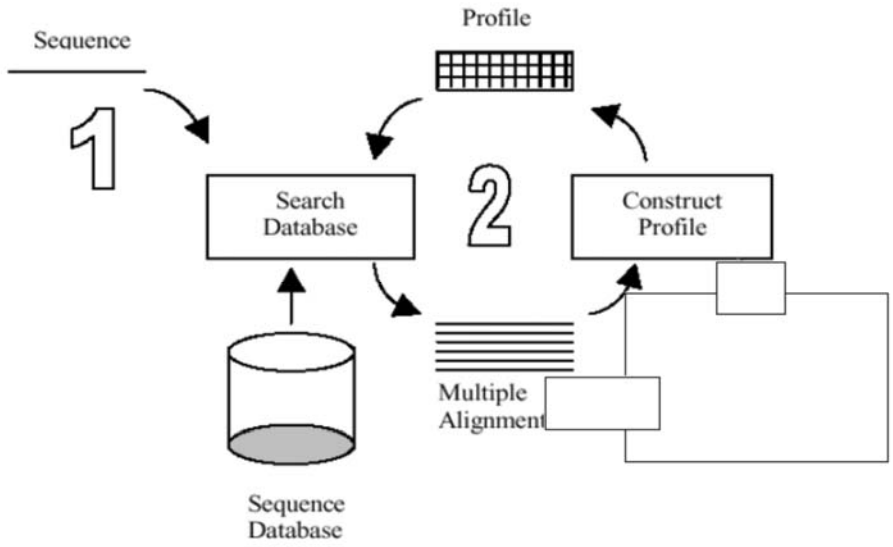




Homology detection and template selection

- Homology detection
  - To detect the fold of a probe sequence from a library of known target fold.
- The three type of methods:
  - Pair-wise sequence-sequence comparison
    - FASTA, BLAST
  - Sequence profile comparison
    - PSI-BLAST, IMPALA, HMMER, SAM
  - Profile-profile comparison
    - prof_sim, COMPASS

PSI-BLAST Overview

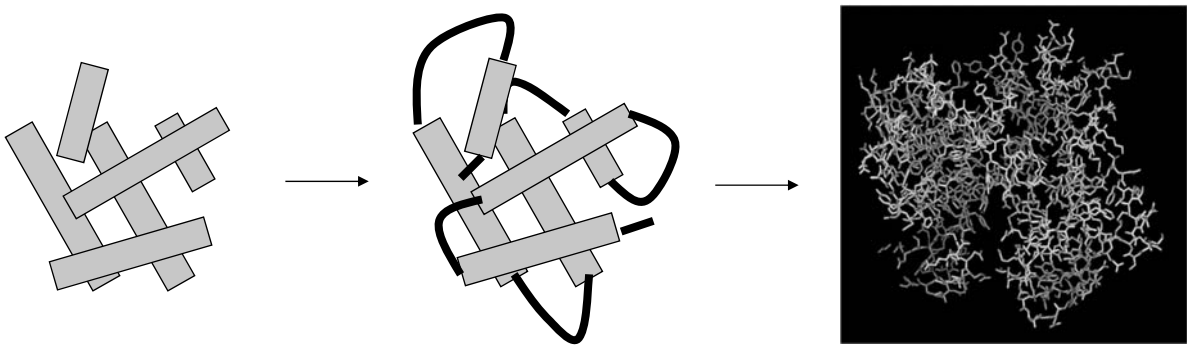


The importance of the sequence alignment

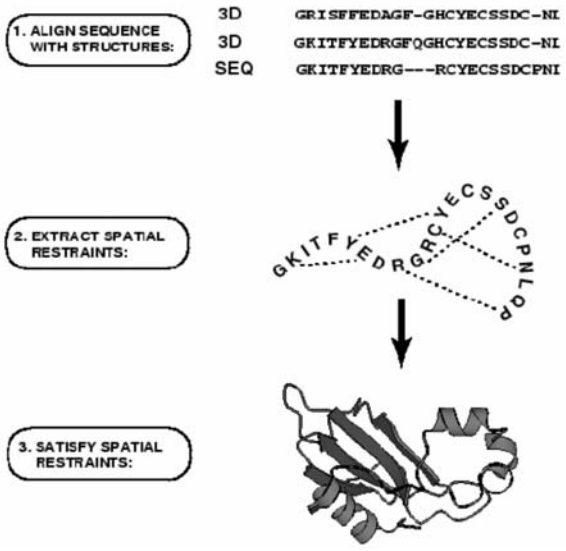
- The quality of the sequence alignment is of crucial importance.
- Misplaced gaps, representing insertions or deletions, will cause residues to be misplaced in space.
- Careful inspection and adjustment on alignment may improve the quality of the modeling.

Model Building

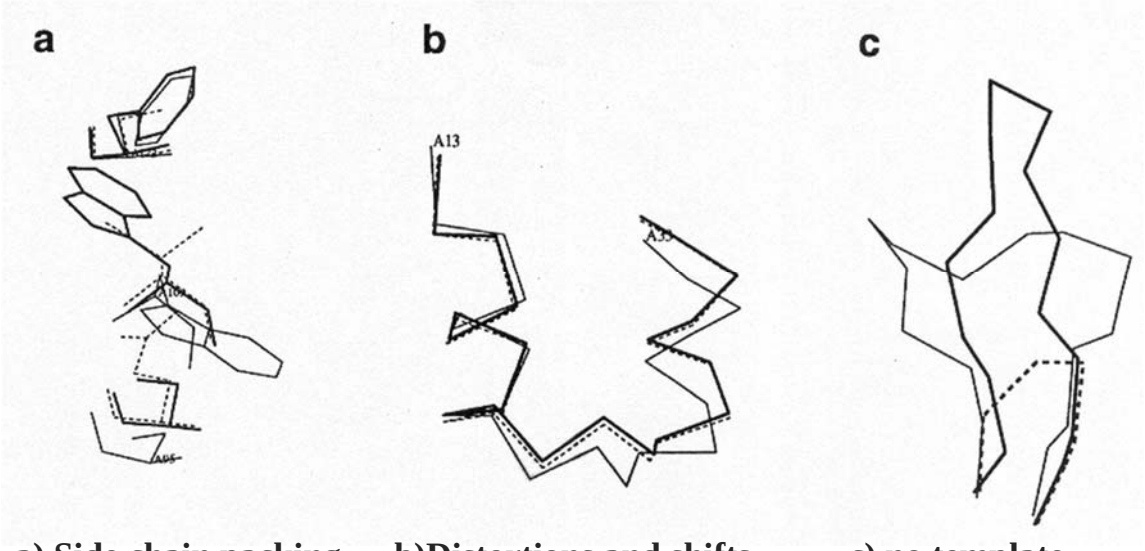
- Steps
  - Partial backbone generation
  - Loop modeling
  - Side-chain modeling



MODELLER

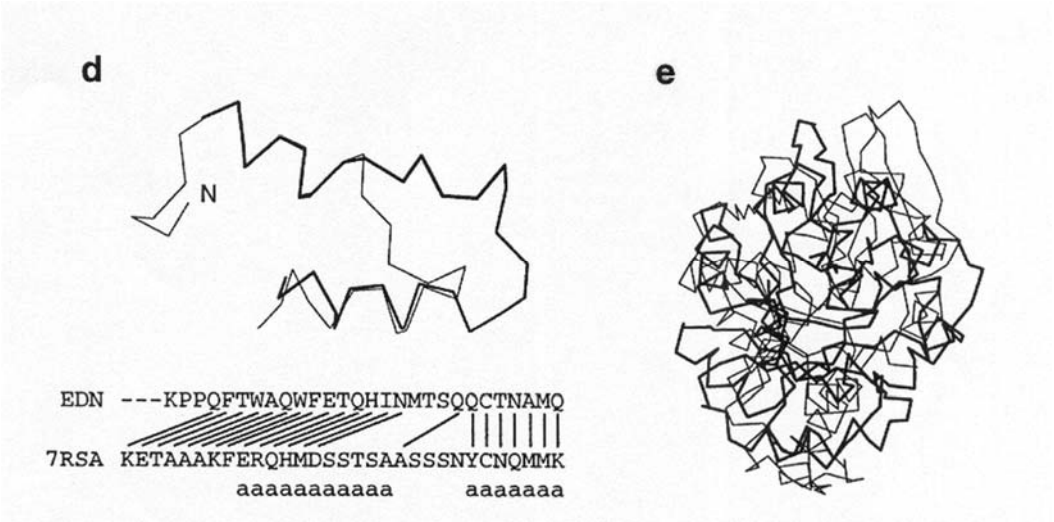


Errors in Homology Modeling



a) Side chain packing **b)Distortions and shifts** **c) no template**

Errors in Homology Modeling



d) Misalignments **e) incorrect template**

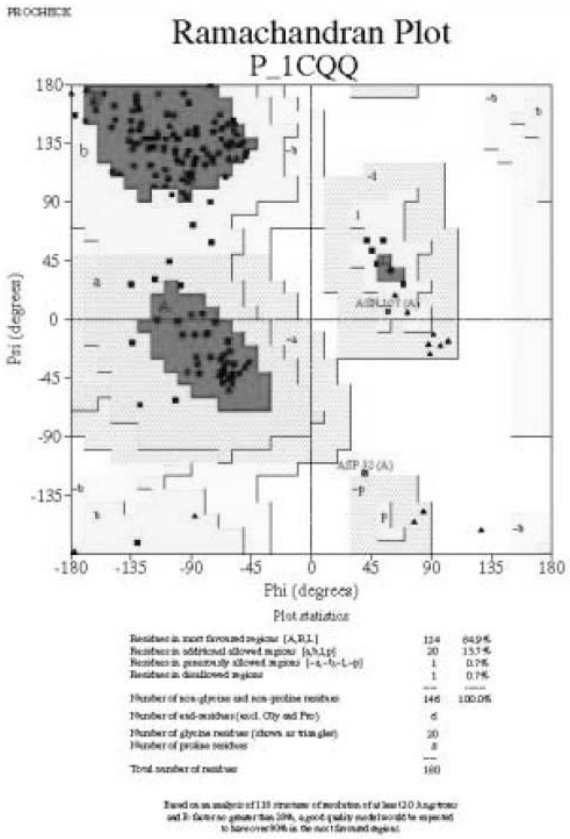
Marti-Renom *et al*., *Ann. Rev. Biophys. Biomol. Struct.*, 2000, 29:291-325.

Detection of Errors

- A stereochemical check on the modeled structure
  - PROCHECK
  - WHATCHECK
  - DISTAN
  - which will show deviations from normal bond lengths, dihedrals, etc.

PROCHECK

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



Modelling Servers:

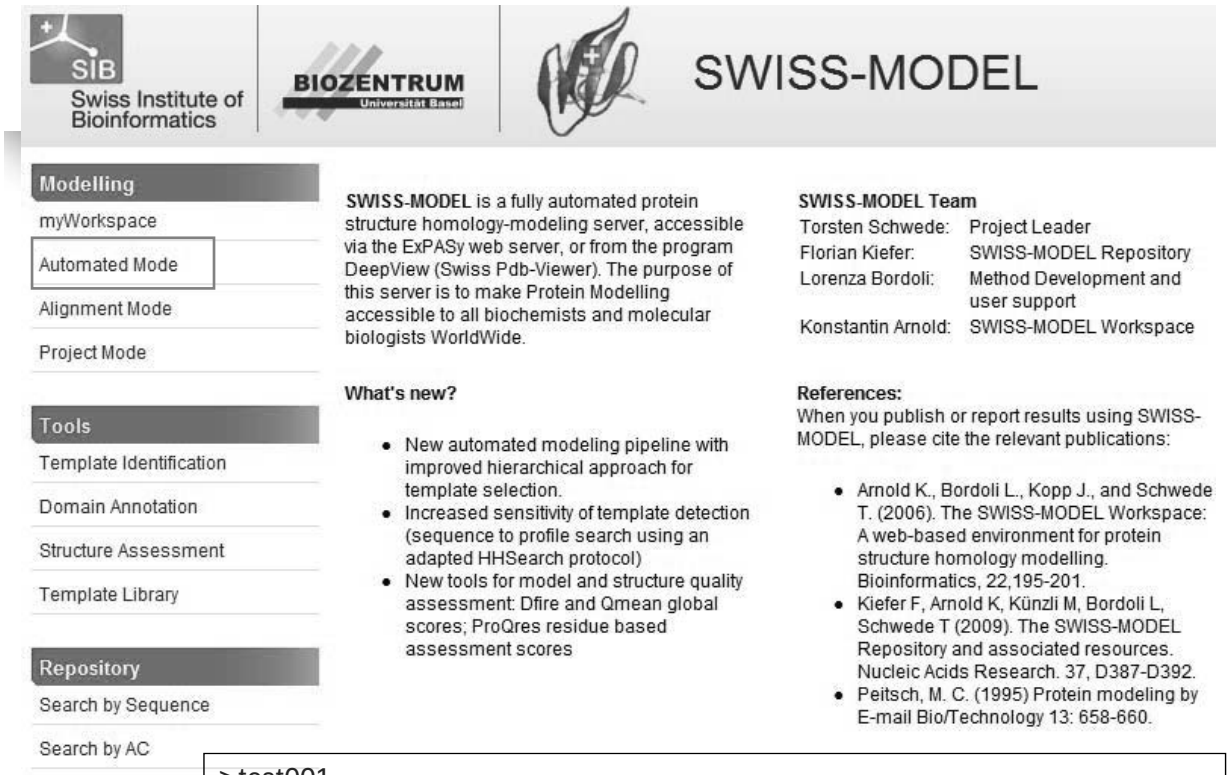
SWISS-MODEL	Server	http://www.expasy.org/swissmod/SWISS-MODEL.html http://www.expasy.ch/swissmod/SWISS-MODEL.html
WHAT IF	Program	http://www.cmbi.kun.nl/whatif/
MODELLER	Program	http://guitar.rockefeller.edu/modeller/modeller.html
InsightII	Program	http://www.accelrys.com/about/msi.html
SDSC1	Server	http://cl.sdsc.edu/hm.html
PredictProtein	Server	http://cubic.bioc.columbia.edu/predictprotein/ http://biobug.life.nthu.edu.tw/predictprotein/

$(PS)^2$: Protein Structure Prediction Server http://ps2.life.nctu.edu.tw/

SWISS-MODEL

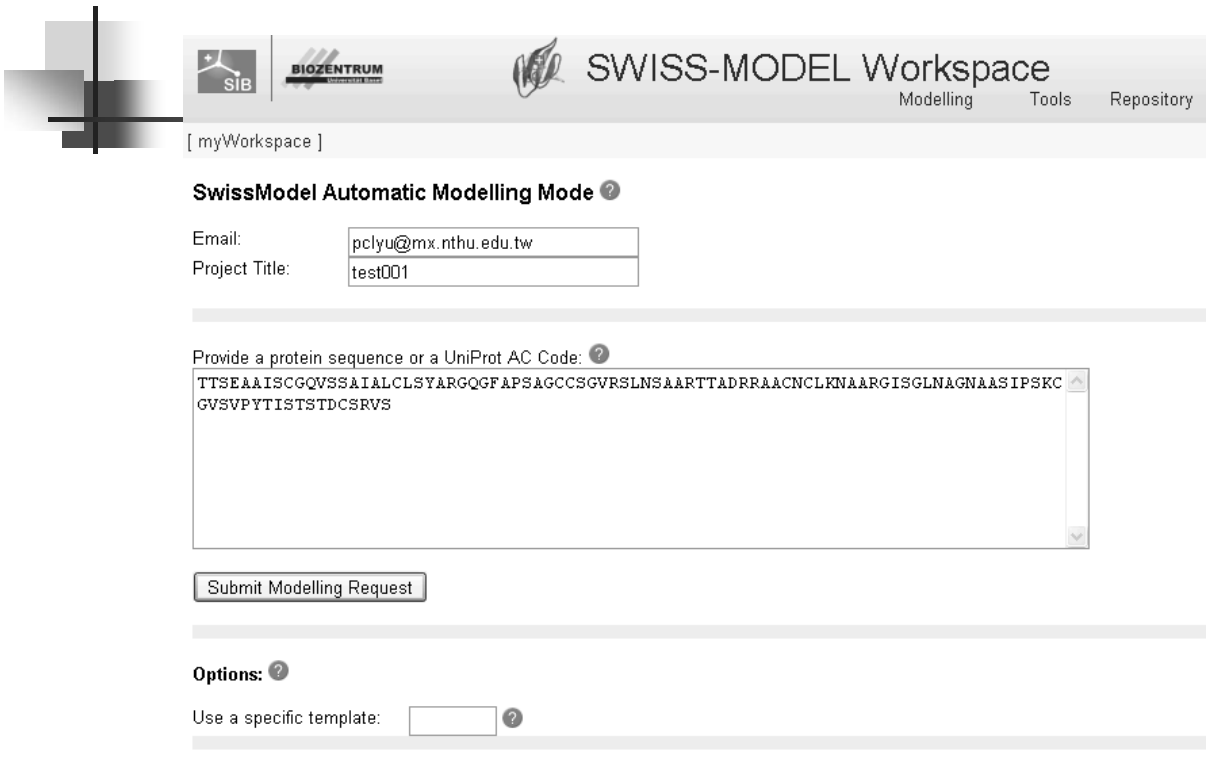
http://www.expasy.org/swissmod/SWISS-MODEL.html

http://www.expasy.org/swissmod/SWISS-MODEL.html



>test001
TTSEAAISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLK
NAARGISGLNAGNAASIPSKCGVSVPYTISTSTDCSRVS

Automatic Modelling Mode

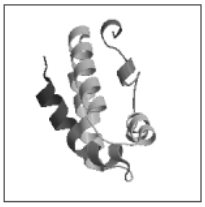


Workunit: P000004 Title:ltp002

1 98

Go to: [Template Selection] [Alignment] [Modelling Log] [Evaluation]

Model Details: Segment 1



Model info:
modelled residue range: 6 to 98
based on template **1fk5A** (1.30 Å)
Sequence Identity [%]: 89.247
Evalue: 1.03e-20

display model: as pdb - as DeepView project
download model: as pdb - as Deepview project - as text

Alignment [top]

```
TARGET    6       AISCGQVS SAIALCLSYA RGQGFAPSAG CCSGVRSLNS AARTTADRRA
1fk5A     1       aiscgqva saiapcisya rgqgsgpsag ccsgvrslnn aarttadrra

TARGET              hhhhh h   hhhhhh h       hh hhhhhhhhhh h   hhhhhh
1fk5A               hhhhh h   hhhhhh h       hh hhhhhhhhhh h   hhhhhh
TARGET    54    ACNCLKNAAR GISGLNAGNA ASIPSKCGVS VPYTISTSTD CSRVS
1fk5A     49    acnclknaaa gvsglnagna asipskcgvs ipytiststd csrvn-

TARGET          hhhhhhhhh         hhhh hhhhhhh
1fk5A           hhhhhhhhh         hhhh    hhhh
```

SWISS-MODEL Workspace

Modelling Tools Repository

[myWorkspace] [Settings] [logout]

Template Identification

Email: pclyu@mx.nthu.edu.tw

Project Title: ltp001

Please provide a protein sequence or a UniPort AC Code:

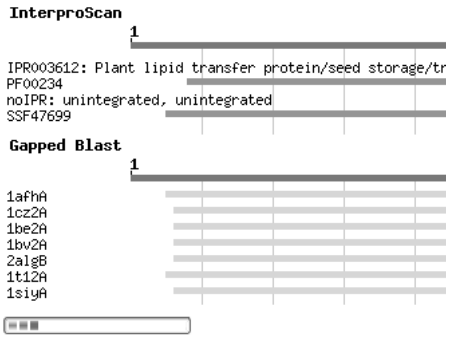
```
TTSEAAISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLKNAARGISGLNAGNAASIPSKC
GVSVPYTISTSTDCSRVS
```

submit

InterPro Domain Scan:

☑ HMMPfam ☑ HMMTigr ☑ ProfileScan ☑ SuperFamily
☑ BlastProDom ☐ FPrintScan ☐ HMMPIR ☐ HMMSmart
☐ ScanRegExp

Workunit: P000003
Title: ltp001
Status: running




```
Query= TARGET
         (98 letters)

Database: SMTL100
          50,266 sequences; 12,621,495 total letters

Searching.................................................done
```

Sequences producing significant alignments:	Score (bits)	E Value
ExPDB|1afhA|99|none	96	4e-21
ExPDB|1cz2A|99|none	92	4e-20
ExPDB|1be2A|99|none	92	5e-20
ExPDB|1bv2A|99|none	87	2e-18
ExPDB|2algB|2.3|Crystal structure of peach Pru p3 the prototypic...	81	1e-16
ExPDB|1t12A|99|none	79	6e-16
ExPDB|1siyA|99|none	73	3e-14

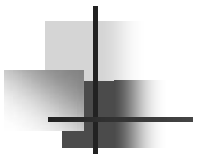
[top]

```
>[Template]|1cz2A|99|none
          Length = 90


 [Display Alignment in DeepView]
 Score = 92.4 bits (228), Expect = 4e-20
 Identities = 53/91 (58%), Positives = 66/91 (72%), Gaps = 1/91 (1%)

Query: 7  ISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLKNAARGIS 66
          I CG V S +  CLSY +G G  PS  CC GV++L++ AR+ +DR++ACNCLK  ARGI
Sbjct: 1  IDCGHVDSLVRPCLSYVQG-GPGPSGQCCDGVKNLHNQARSQSDRQSACNCLKGIARGIH 59

Query: 67 GLNAGNAASIPSKCGVSVPYTISTSTDCSRV 97
           LN  NA SIP KCGV++PYTIS + DCSRV
Sbjct: 60 NLNEDNARSIPPKCGVNLPYTISLNIDCSRV 90
```



SwissModel Alignment Mode

Email: pclyu@mx.nthu.edu.tw

Project Title: ltp003

Alignment Input Format:

FASTA

Cut & paste your Target-Template Alignment:

```
>test001
ISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLKNAARGISGLNAGNAASIPSKCGVSVPY
TISTSTDCSRV

>1cz2A
IDCGHVDSLVRPCLSYVQG-
GPGPSGQCCDGVKNLHNQARSQSDRQSACNCLKGIARGIHNLNEDNARSIPPKCGVNLPYTISLNIDCSRV
```

Or upload an alignment file [瀏覽...]

submit alignment

Workunit: P000006
Title:
Status: submission
back

Please select sequences from your alignment:

Target Sequence: test001
Template Sequence: 1cz2A PDB-Code: 1cz2 Chain-ID: A

submit alignment

Your Request has been Submitted:

The Request has been submitted to the queueing system and will run soon.
Status of your Request is: submitted

The results will be displayed in this page .

Workunit: P000006
Title:
Status: submission
back

Before submission, please ensure that the alignment has been interpreted correctly:

```
CLUSTAL W(1.81) multiple sequence alignment


Target/1-91           ISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLKNAARGIS
Template/1-91         IDCGHVDSLVRPCLSYVQG-GPGPSGQCCDGVKNLHNQARSQSDRQSACNCLKGIARGIH
                      *.**:*.* :  ****.:* * .**. **.**:.*:. **: :**::******. ****


Target/1-91           GLNAGNAASIPSKCGVSVPYTISTSTDCSRV
Template/1-91         NLNEDNARSIPPKCGVNLPYTISLNIDCSRV
                      .** .** ***.****.:***** . *****
```

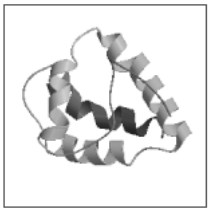
submit alignment

Workunit: P000006 Title:ltp003

1 91

Go to: [Template Selection] [Alignment] [Modelling Log] [Evaluation]

Model Details: **Segment 1**



Model info:
modelled residue range: 1 to 91

display model: as pdb - as DeepView project
download model: as pdb - as Deepview project - as text

Alignment [top]

```
TARGET    1       ISCGQVSS AIALCLSYAR GQGFAPSAGC CSGVRSLNSA ARTTADRRAA
1cz2A     1       idcghvds lvrpclsyvq g-gpgpsgqc cdgvknlhnq arsqsdrqsa

TARGET                hhhh hhhhhhhhhh        hhh hhhhhhhhhh    hhhhhhh
1cz2A                 hhhh hhhhhhhhhh        hhh hhhhhhhhhh    hhhhhhh
TARGET    49    CNCLKNAARG ISGLNAGNAA SIPSKCGVSV PYTISTSTDC SRV
1cz2A     48    cnclkgiarg ihnlnednar sippkcgvnl pytislnidc srv-

TARGET          hhhhhh          hhhhh hh
1cz2A           hhhhhhhhhh      hhhhh hh
```

Anolea / Gromos / Verify3D [top]

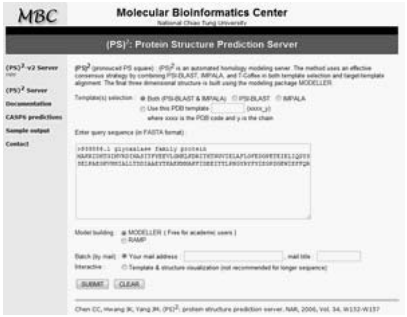
anolea: on off **gromos:** on off **verify3d:** on off show

Check Your Model by PROCHECK

- Go to PDBsum
- http://www.ebi.ac.uk/pdbsum/
- Generate
- Sumbmit your PDB file for PDBsum analysis

- TBI http://www.tbi.org.tw/

$(PS)^2$: Protein Structure Prediction Server



http://ps2.life.nctu.edu.tw/

Overview

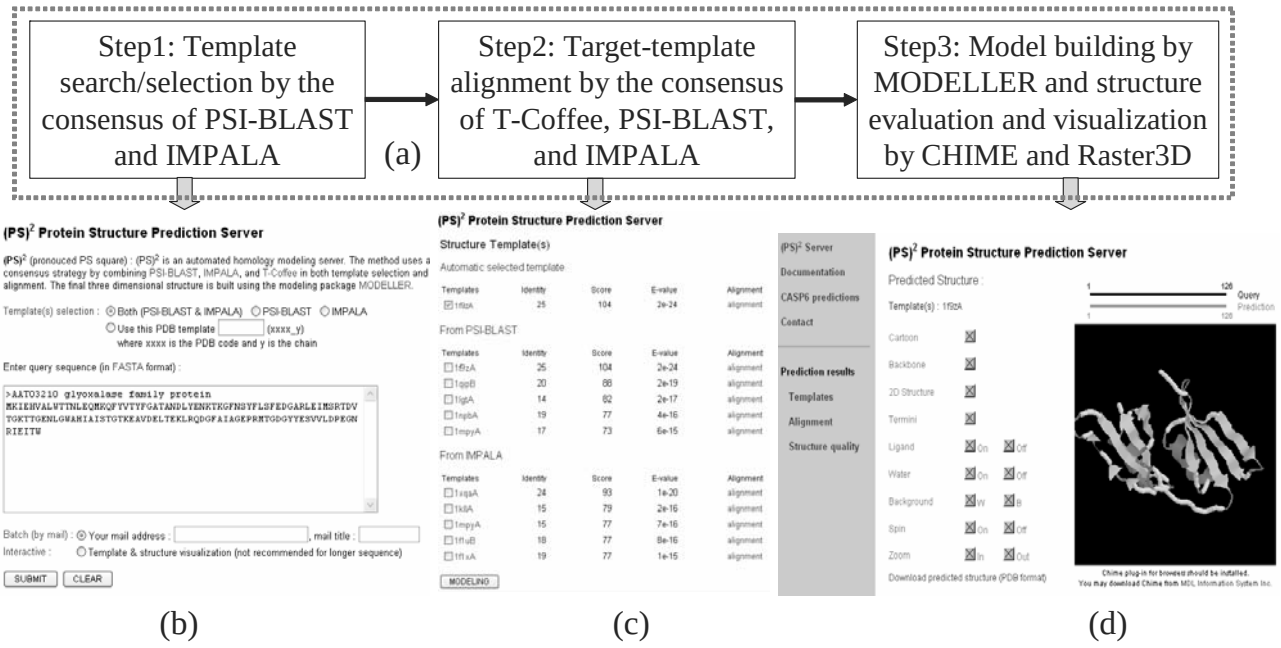
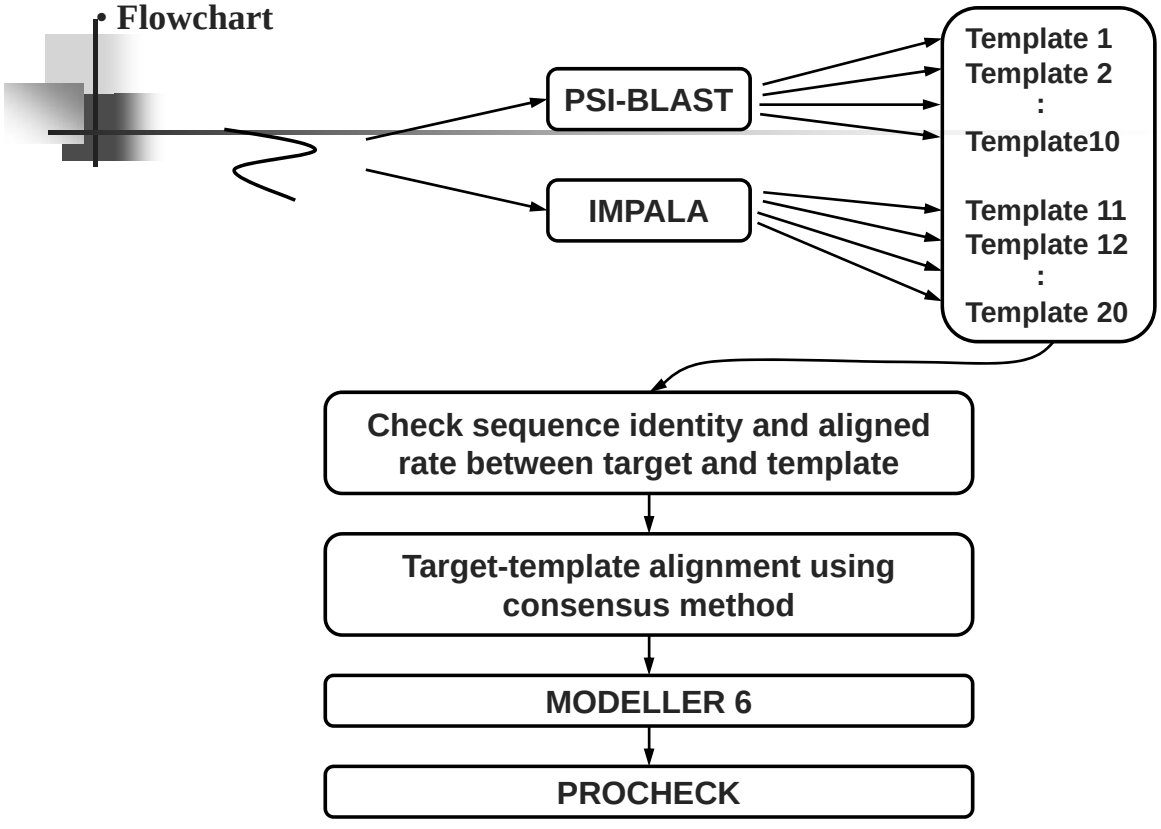


Figure 1. Overview of the protein structure prediction server, $(PS)^2$.

Materials and Methods

- **Flowchart**



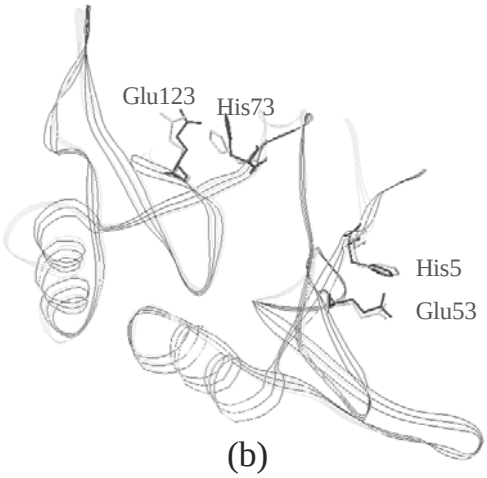
Results

```
AAT03210: 1   MKIEHVALWTTNLEQMKQFYVTYFGATANDLYENKTKGFNSYFLSFE---DGARLEIMSR 57
              M++ H  L    +L++    FY     G        EN     ++  F+ +    + A +E+
1f9zA   : 1   MRLLHTMLRVGDLQRSIDFYTKVLGMKLLRTSENPEYKYSLAFVGYGPETEEAVIELTYN 60

AAT03210: 58  TDVTGKTTGENLGWAHIAISTGTKEAVDELTEKLRQDGFAIAGE-PRMTGDGYYESVVLD 116
                V     G   G  HIA+S         E  EK+RQ+G  +  E   + G     + V D
1f9zA   : 61  WGVDKYELGTAYG--HIALSVDNAA---EACEKIRQNGGNVTREAGPVKGGTTVIAFVED 115

AAT03210: 117 PEGNRIEI 124
              P+G +IE+
1f9zA   : 116 PDGYKIEL 123
```

(a)



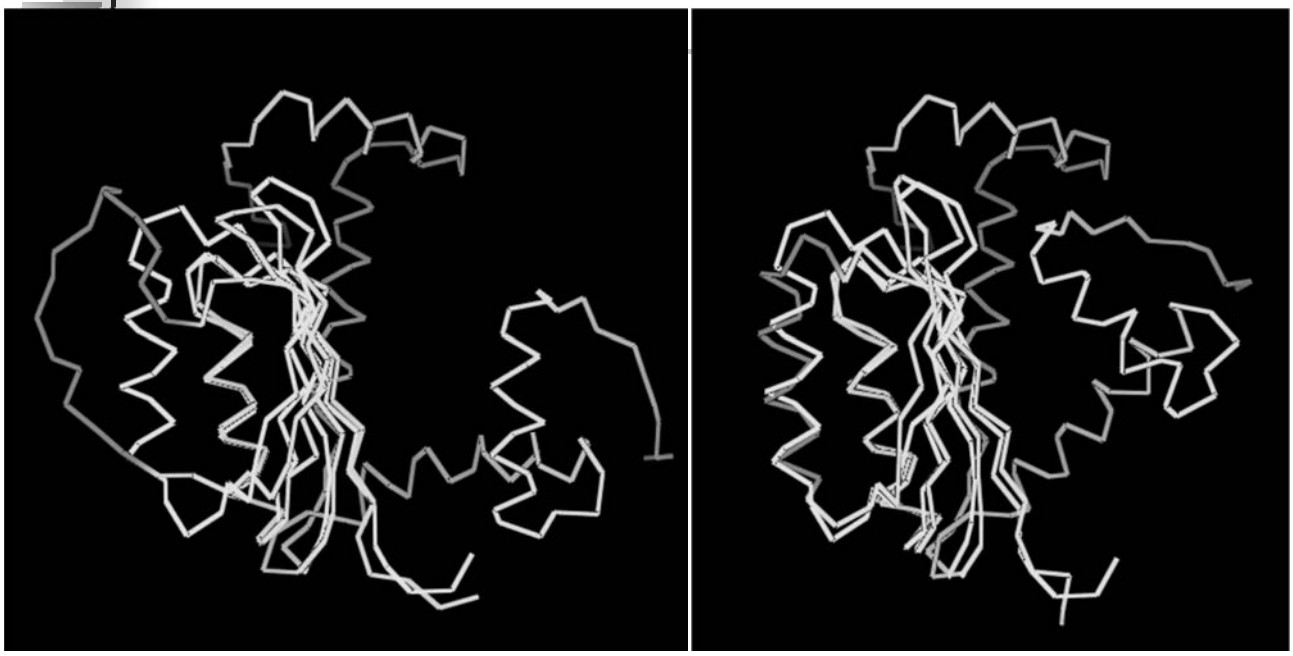
(b)

Figure 7. Predicted structure results of the $(PS)^2$ server using F2365 glyoxalase protein (AAT03210) sequence from *L. monocytogenes* as a query.

Target: **2bj7A**

$(PS)^2$ model

multi-templates



GDT_TS = 45.96

GDT_TS = 60.48

http://ps2.life.nctu.edu.tw

(PS)^2: Protein Structure Prediction Server

(PS)^2 (pronouced PS square) : (PS)^2 is an automated homology modeling server. The method uses an effective consensus strategy by combining PSI-BLAST, IMPALA, and T-Coffee in both template selection and target-template alignment. The final three dimensional structure is built using the modeling package MODELLER.

Template(s) selection : ◉ Both (PSI-BLAST & IMPALA) ○ PSI-BLAST ○ IMPALA
○ Use this PDB template [ ] (xxxx_y)
where xxxx is the PDB code and y is the chain

Enter query sequence (in FASTA format) :

```
>test001
TTSEAAISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLKN
AARGISGLNAGNAASIPSKCGVSVPYTISTSTDCSRVS
```

Model building : ◉ MODELLER (Free for academic users)
○ RAMP

Batch (by mail) : ◉ Your mail address : [ ], mail title : [ ]
Interactive : ○ Template & structure visualization (not recommended for longer sequence)

SUBMIT CLEAR

Chen CC, Hwang JK, Yang JM. (PS)^2: protein structure prediction server. NAR, 2006, Vol. 34, W152-W157

(PS)^2: Protein Structure Prediction Server

Structure Template(s)

Automatic selected template

Templates	Identity	Score	E-value	Alignment	
☑ 1fk5A	89	105	2e-24	alignment	MODELING

From PSI-BLAST

Templates	Identity	Score	E-value	Alignment
☐ 1fk5A	89	105	2e-24	alignment
☐ 1bwoA	58	102	1e-23	alignment
☐ 1midA	57	102	2e-23	alignment
☐ 1rzlA	81	96	1e-21	alignment
☐ 2algA	57	90	1e-19	alignment
☐ 1t12A	58	87	6e-19	alignment
☐ 1siyA	46	81	4e-17	alignment

From IMPALA

Templates	Identity	Score	E-value	Alignment
☐ 2algA	56	111	7e-26	alignment
☐ 1fk5A	89	109	2e-25	alignment
☐ 1t12A	58	107	6e-25	alignment
☐ 1rzlA	80	107	1e-24	alignment
☐ 1bwoA	57	106	1e-24	alignment
☐ 1midA	56	106	2e-24	alignment
☐ 1siyA	45	104	5e-24	alignment

MBC

Molecular Bioinformatics Center

National Chiao Tung University

(PS)^2: Protein Structure Prediction Server

Your job has been successfully submitted to the (PS)^2 server.

The job ID is 2114247571

Submitted at Wed Dec 16 09:30:18 2009
Current time Wed Dec 16 09:30:29 2009

Structure modeling, please wait for 1 minutes. .....

Chen CC, Hwang JK, Yang JM. (PS)^2: protein structure prediction server. NAR, 2006, Vol. 34, W152-W157

(PS)^2: Protein Structure Prediction Server

(PS)^2 Server
Documentation
CASP6 predictions
Sample output
Contact

Prediction results
Templates
Alignment
Structure quality
B-factor

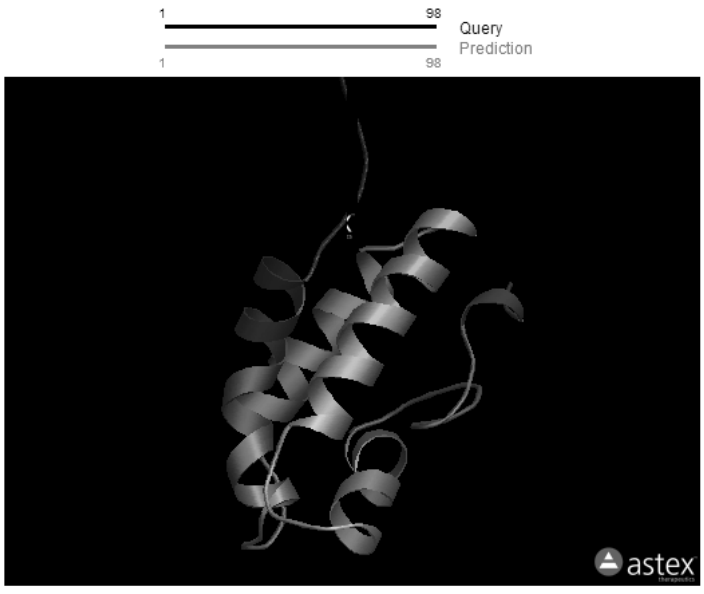
Predicted Structure :

Template(s) : 1fk5A

Cartoon
Cylinders
Spheres
Background W B
Surface On Off
Colour...
Opacity...
Texture...

Show the predicted picture (PNG format)

Download predicted structure (PDB format)



Structure quality

Check quality of the predicted structure by PROCHECK

```
 +----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                             |
 | test   2.0                                                      98 residues |
 |                                                                             |
 | Ramachandran plot:   97.6% core    2.4% allow    0.0% gener    0.0% disall |
 |                                                                             |
 | All Ramachandrans:    0 labelled residues (out of  96)                      |
+| Chi1-chi2 plots:      1 labelled residues (out of  40)                      |
 | Main-chain params:    6 better     0 inside      0 worse                    |
 | Side-chain params:    5 better     0 inside      0 worse                    |
 |                                                                             |
 | Residue properties: Max.deviation:     3.0             Bad contacts:    1 |
+|                     Bond len/angle:    4.9    Morris et al class:  1  1  2 |
 |                                                                             |
 | G-factors           Dihedrals:   0.22  Covalent:  -0.15    Overall:   0.08 |
 |                                                                             |
 | M/c bond lengths: 99.6% within limits   0.4% highlighted                    |
 | M/c bond angles:  93.8% within limits   6.2% highlighted                    |
 | Planar groups:   100.0% within limits   0.0% highlighted                    |
 |                                                                             |
 +-----------------------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

Ideally, the G-factors scores should be above -0.5. Values below -1.0 may need investigation.

(PS)^2: Protein Structure Prediction Server

Structure Template(s)

Automatic selected template

Templates	Identity	Score	E-value	Alignment	
☐ 1fk5A	89	105	2e-24	alignment	MODELING

From PSI-BLAST

Templates	Identity	Score	E-value	Alignment
☐ 1fk5A	89	105	2e-24	alignment
☐ 1bwoA	58	102	1e-23	alignment
☐ 1midA	57	102	2e-23	alignment
☐ 1rzlA	81	96	1e-21	alignment
☐ 2algA	57	90	1e-19	alignment
☐ 1t12A	58	87	6e-19	alignment
☐ 1siyA	46	81	4e-17	alignment

From IMPALA

Templates	Identity	Score	E-value	Alignment
☑ 2algA	56	111	7e-26	alignment
☑ 1fk5A	89	109	2e-25	alignment
☑ 1t12A	58	107	6e-25	alignment
☑ 1rzlA	80	107	1e-24	alignment
☐ 1bwoA	57	106	1e-24	alignment
☐ 1midA	56	106	2e-24	alignment
☐ 1siyA	45	104	5e-24	alignment

MODELING

```
#Query
>P1;test
sequence:test::::::::
TTSEAAISCGQVSSAIALCLSYARGQGFAPSAGCCSGVRSLNSAARTTADRRAACNCLKNAARGISGLNAGNAAS
IPSKCGVSVPYTISTSTDCSRVS*
#Template
>P1;2algA
structureX:2alg:   0 :A:  91 :A:MOL_ID  1;  MOLECULE  NON-SPECIFIC LIPID TRANSFER PROTEIN;
-----MITCGQVSSSLAPCIPYVRGGGAVPPA-CCNGIRNVNNLARTTPDRQAACNCLKQLSASVPGVNPNNAAA
LPGKCGVSIPYKISASTNCATVK*
#Template
>P1;1fk5A
structureX:1fk5:   1 :A:  93 :A:MOL_ID  1;  MOLECULE  NONSPECIFIC LIPID-TRANSFER PROTEIN;
-----AISCGQVASAIAPCISYARGQGSGPSAGCCSGVRSLNNAARTTADRRAACNCLKNAAAGVSGLNAGNAAS
IPSKCGVSIPYTISTSTDCSRVN*
#Template
>P1;1t12A
structureN:1t12:   1 :A:  91 :A:MOL_ID  1;  MOLECULE  NONSPECIFIC LIPID-TRANSFER PROTEIN 1;
-----AITCGQVTSNLAPCLAYLRNT--GPLGRCCGGVKALVNSARTTEDRQIACTCLKSAAGAISGINLGKAAG
LPSTCGVNIPYKISPSTDCSKVQ*
#Template
>P1;1rzlA
structureX:1rzl:   1 :A:  91 :A:MOL_ID  1;  MOLECULE  NONSPECIFIC LIPID TRANSFER PROTEIN;
------ITCGQVNSAVGPCLTYARG-GAGPSAACCSGVRSLKAAASTTADRRTACNCLKNAARGIKGLNAGNAAS
IPSKCGVSVPYTISASIDCSRVS*
```

(PS)²: Protein Structure Prediction Server

(PS)² Server
Documentation
CASP6 predictions
Sample output
Contact

Prediction results
- **Templates**
- **Alignment**
- **Structure quality**
- **B-factor**

Predicted Structure :

Template(s) : 2algA 1fk5A 1t12A 1rzlA

1 98 Query
1 98 Prediction

Cartoon ◉
Cylinders ○
Spheres ○
Background ○ W ◉ B
Surface ○ On ◉ Off

Colour...
Opacity...
Texture...

Show the predicted picture (PNG format)

Download predicted structure (PDB format)