Tutorial-3  #2

Structural modeling of proteins: Principle and application to an ion channel

Haruki Nakamura
PDBj, Institute for Protein Research, Osaka University

Daron M. Standley
Immunology Frontier Research Center, Osaka University

Narutoshi Kamiya
The center for Advanced Medical Engineering and Informatics, Osaka University

http://www.protein.osaka-u.ac.jp/rcsfp/pi/
http://www.pdbj.org/
Protein Data Bank Japan
http://www.pdbj.org/

At Institute for Protein Research, Osaka Univ. since 2001 supported from the Institute for Bioinformatics Research and Development, Japan Science and Technology Agency (BIRD-JST).

● Structure Data curation and editing
● Structure Data browsing and downloading
E-MSD is supported by grants from the Wellcome Trust, the EU (TEMBLOR, NMRQUAL and IIMS), CCP4, the BBSRC, the MRC and EMBL.

The BMRRB is supported by NIH grant LM05799 from the National Library of Medicine.

PDBj is supported by grant-in-aid from the Institute for Bioinformatics Research and Development, Japan Science and Technology Agency (BIRD-JST), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

The RCSB PDB is supported by grants from the National Science Foundation, National Institute of General Medical Sciences, the Office of Science-Department of Energy, the National Library of Medicine, the National Cancer Institute, the National Center for Research Resources, the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and the National Institute of Diabetes & Digestive & Kidney Diseases.

Markley, JL  Henrick, K  Berman, HM  Nakamura, H

wwPDB and wwPDBAC members at EBI, Hinxton on 29 Sept, 2008
Development of other Databases and Services

Alignment of Sequence and Structures. MAFFTash (Kato, Toh & Standley)

Homolog protein search, Sequence Navigator (Standley)

Similar fold search, Structure Navigator (Standley & Toh)

Encyclopedia of Protein Structures, eProtS (Kinjo, Kudo, & Ito)

Molecular of the Month, MoM (Goodsell & Kudo)

Protein Folds Browser, Protein Globe (Kinjo & Standley)
Development of other Databases and Services

**Protein Molecular Surface Database, eF-site** (Kinoshita & Nakamura)

**Search for Similar Surface, eF-seek** (Kinoshita & Nakamura)

**Electron Microscopy Navigator, EM-Navi** (Suzuki)

**Function Annotation from Folds and Sequences, SeSAW** (Standley)

**Ligand Binding Site Search, GIRAF** (Kinjo)

**Protein Dynamics Database, ProMode** (Wako & Endo)
Homology modeling for a target protein

1) Introduction
2) Search for homolog(s)
3) Threading (3D-1D compatibility)
4) Backbone modeling
5) Side-chain modeling
6) Structure optimization
Homology modeling
/Comparative modeling

A structural model of a target protein is constructed based on the homolog protein structure as a template, using the similarity of amino acid sequences.

- **Requirement for the homology modeling**
  - Sequence information: 9.2 M (UniProt)
  - 3D Structure: 59 K (wwPDB)
  - 3D structural model can be made when any structure of the family protein is in DB.
    - Total family number: about 30,000 (30% identity)
    - Total folds: about 2,000 (loose definition)

- **Principle**
  - When sequence is similar, structure is similar.

- **Modeling procedure**
  - Search homolog proteins.
  - Construct (multiple) sequence alignment
  - Differences in the backbone and the side-chains are modeled.

Swine Influenza A virus neuraminidase (NA) gene

**Simple Example**

**Influenza A virus (A/Texas/04/2009[H1N1])** segment 6 neuraminidase (NA) gene, complete cds

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>FJ891614.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>LENGTH</td>
<td>1410 bp</td>
</tr>
<tr>
<td>cDNAlinear</td>
<td>VRL 01-MAY-2009</td>
</tr>
</tbody>
</table>

**REFERENCE**

| TITLE | Human infection with novel swine H1N1 influenza |
| JOURNAL | Unpublished |

**COMMENT**

Swine influenza A (H1N1) virus isolated during human swine flu outbreak of 2009. For more information, see [https://www.cdc.gov/](https://www.cdc.gov/).

Some of the information does not have homolog feature identifiers and is being provided in the comment section.

**AA sequence**

<table>
<thead>
<tr>
<th>AA sequence</th>
<th>NA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 atgatccac ccgacagat aatccattc cggctgtgt gattgatc atgcatgc atgatgc</td>
<td></td>
</tr>
<tr>
<td>61 atccattc tctacatgg acatactg tctagctga acataatc</td>
<td></td>
</tr>
<tr>
<td>121 ggtgctagc acatggact gccagccggt cggagtggg cggagtggg</td>
<td></td>
</tr>
<tr>
<td>181 tcggtatcct acataaggc tgcgcgcgcgcc gcgcgccgcc cgccgccgcc</td>
<td></td>
</tr>
<tr>
<td>241 ggtgctagc acatggact gccagccggt cggagtggg cggagtggg</td>
<td></td>
</tr>
<tr>
<td>301 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>361 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>421 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>481 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>541 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>601 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>661 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>721 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>781 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>841 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>901 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>961 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1021 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1081 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1141 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1201 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1261 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1321 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1381 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
<tr>
<td>1441 aagccagc aagccagc aagccagc aagccagc aagccagc</td>
<td></td>
</tr>
</tbody>
</table>
Red: Active site residues surrounding Tamiflu
Electrostatic molecular surfaces
Blue: positive, Red: negative, yellow: hydrophobic

2qwk A
Swine NA model
>2qwk chain-A: NEURAMINIDASE
>Influenza A virus neuraminidase(NA) gene, complete cds. FJ981614

RDFNNLTKGLCTINGWHYKDNARVGEDSDLVLVREREYVSCDPDECRFYALSQGTIRGKHSNGTIHDRSQRALI
---------LCPVSGWAIYSKDINSVRIGSKGDVFVIREPFISCPLECTCFFLTQGALLNDKHSNGTIKDRSPYRTLM

SWPLSSSPTVYNSRVECIGWSSTSTCHDGKTRMSICISGPNNNASAVIWYNNRPVTEINTvarNIRLRTQELSECVCHNGV
SCPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPNDGAVAVLKYNGIITDTIKSWRNNIRLRTQELSEACVNGS

CPVVFTDGSAETPAETRIYYFKEGKILKWEPLAGTAKHIECSXYGERAEITCTCRDWNQGSNRPVIRIDPVAMTHTS
CFTVMTDGSPNQQASYKIFRIEKGKIVKSVEMNAPNYHYEESCYPDSSEITCVCRDNWHGSNRPWVSFNQ-NLEYQI

↑ H274Y: Tamiflu resistant

QYICSPVLTDNPRPDPTVGCNDPYPGNNNNGVKGFSYLDGVNTLGRSTISARSGYEMLKVPNALTDDKSKPTQG
GYICSGIFGDNPRPDNTGSC-PVSSNGANNGVKGFSFKYNGNVGRTRKSISSRNGFEMIWDPNGWGTDNDNSSIK

QTIVLNTDWSGYSGSFMDY---WAEGECYRCFYVELIRGRPKEDKVVWTSNSIVSMCSSTEFLGQWDWPDAQIEYFL
QDIVGINWSGYSGSFVQHELITGLDCIRPCFWVELIRGRPKENTI-WTSGSSISFCGVNSDTVGWSWPDGAELPF---

Red: Active site residues surrounding Tamiflu
Electrostatic molecular surfaces
Blue: positive, Red: negative, yellow: hydrophobic

2qwk A
Procedure of Homology Modeling

Search for Homolog
FASTA, BLAST, psi-BLAST

Precise Alignment
3D-1D compatibility search (threading)

Backbone modeling: Loops
loop search, conf. sampling

Side-chain modeling:
Combinatorial problem
Dead End Elimination (DEE)
Homology modeling for a target protein

1) Introduction
2) Search for homolog(s)
3) Threading (3D-1D compatibility)
4) Backbone modeling
5) Side-chain modeling
6) Structure optimization

Goal of this Tutorial:
To construct the homology model of hERG channel
Procedure of Homology Modeling

Search for Homolog
FASTA, BLAST, psi-BLAST

Precise Alignment
3D-1D compatibility search (threading)

with low homology

Backbone modeling: Loops
loop search, conf. sampling

with high homology

Side-chain modeling:
Combinatorial problem
Dead End Elimination (DEE)
Get amino-acid sequence of hERG channel from NCBI
(http://www.ncbi.nlm.nih.gov/)

RecName: Full=Potassium voltage-gated channel subfamily H member 2; AltName: Full=Voltage-gated potassium channel subunit Kv1.1; AltName: Full=Ether-a-go-go-related gene potassium channel 1; Short=H-ERG; Short=ErG1; Short=Ether-a-go-go-related protein 1; Short=Eag-related protein 1; AltName: Full=eag homology.

Sequence Analysis Tools
- BLAST Sequence
- Conserved Domains

Articles about the KCNH2 gene
- Genetic Polymorphism of KCNH2 Confers Pydrosopia [Cardiovasc Electrophysiol. 2009]
- Breaking the gene barrier in schizophrenia. [Nat Med. 2009]
- Interactions of H562 in the S5 helix with T518 and S621 in the pore helix a[Biophys J. 2009]

Identical Proteins for Q12809.1
- Sequence 2 from patent US 75411 [ACS12627]
- Sequence 5 from patent US 7537 [AC308477]
- Sequence 2 from patent US 7510 [AC019114]

RefSeq Protein Isoforms
See 3 reference sequence protein isoforms for the KCNH2 gene.

More about the KCNH2 gene
This gene encodes a voltage-activated potassium channel belonging to the eag family. It shares sequence similarity with the Dro sophila ether...

Also Known As: ERG1, HERG, HERG1, Kv11...

Homologs of the KCNH2 gene
The KCNH2 gene is conserved in chimpanzee,
Get amino-acid sequence of hERG channel from UniProt (http://www.uniprot.org/).

<table>
<thead>
<tr>
<th>Names and origin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein names</strong></td>
</tr>
<tr>
<td><em>Recommended name:</em></td>
</tr>
<tr>
<td><em>Alternative name(s):</em></td>
</tr>
<tr>
<td>Short name(s):</td>
</tr>
<tr>
<td><strong>Gene names</strong></td>
</tr>
<tr>
<td>Name:</td>
</tr>
<tr>
<td>Synonyms:</td>
</tr>
<tr>
<td><strong>Organism</strong></td>
</tr>
<tr>
<td><strong>Taxonomic identifier</strong></td>
</tr>
<tr>
<td><strong>Taxonomic lineage</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence length</strong></td>
</tr>
<tr>
<td><strong>Sequence status</strong></td>
</tr>
<tr>
<td><strong>Sequence processing</strong></td>
</tr>
<tr>
<td>Regions</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Topological domain</td>
</tr>
<tr>
<td>Transmembrane</td>
</tr>
<tr>
<td>Transmembrane</td>
</tr>
<tr>
<td>Topological domain</td>
</tr>
<tr>
<td>Transmembrane</td>
</tr>
<tr>
<td>Transmembrane</td>
</tr>
<tr>
<td>Topological domain</td>
</tr>
<tr>
<td>Transmembrane</td>
</tr>
<tr>
<td>Transmembrane</td>
</tr>
<tr>
<td>Topological domain</td>
</tr>
<tr>
<td>Domain</td>
</tr>
<tr>
<td>Domain</td>
</tr>
<tr>
<td>Nucleotide binding</td>
</tr>
<tr>
<td>Region</td>
</tr>
<tr>
<td>Motif</td>
</tr>
<tr>
<td>Compositional bias</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino acid modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycosylation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Natural variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative sequence</td>
</tr>
<tr>
<td>Alternative sequence</td>
</tr>
</tbody>
</table>
Secondary structure

Helix [ ] Strand [ ] Turn [ ]

Details...

Sequences

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Mass (Da)</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoform 1 [UniParc]</td>
<td>1,159</td>
<td>126,655</td>
<td>Blast</td>
</tr>
</tbody>
</table>

Checksum: D03BD4F657841FBA

10
MPVRGRHVP QNFIFDITIR KFEGQRKRFI

20
IANARNVENC VIYCNDFQCE LCGYSAEUVN

30
QRFCTCDFLH GPRQQRRAAE QIAQALGGAE

40
ERKVEIAFGR KDGSCFLCLV DVVEFKNEDG

50
AVIMFILNE VVMEEDMVGS FAHDTNHRGF

60
PTSWLAPGRA KTIRKLPLLAL LALTARESSV

70
RSGCAGAGAAA PGAVVVDVDL TPAAPSESEL

80
ALDEVTAMDN HVAGLPHead RRALVDPGSP

90
PRAHSLNPDA SGSSCSLART RSRESCAVSR RASSADDIEA MRAGVLPPPP

100
SKIPQITLNF VDLKGDPLFA SPTSDREIIA

110
PKIKERTHV IYKTOVSLG CADVLPFEYL QAPRIRFRWI LHYPFKAIV DWLILLVYI

120
TAVFTPYSAA PLLKETEEGF RATEGACQ OPLAVDLIVD IMFIVDILIN PRRTVNVN

130
EVRVHPGRIA UHYFKEGWFIL DMWAAIFPFLL LFMTGGSSEKL IGLKARTLL RLVRVARKLD

140
RYSEYGAAVL FILMCTFALI AHWLACIWYA IGNMEQPHMD SRIGWLNHLG DQIGKPYNSS

150
CILGSIKD KYYTALYFTPS SLTSVFGCNV SFMTSEKFIP SICVNLIGEL MYASIFGNV
Homology modeling for a target protein

1) Introduction
2) Search for homolog(s)
3) Threading (3D-1D compatibility)
4) Backbone modeling
5) Side-chain modeling
6) Structure optimization
Procedure of Homology Modeling

Search for Homolog
FASTA, BLAST, psi-BLAST

Precise Alignment
3D-1D compatibility search (threading)

Backbone modeling: Loops
loop search, conf. sampling

with low homology

with high homology

Side-chain modeling: Combinatorial problem
Dead End Elimination (DEE)
3D-1D compatibility search (*Threading method*)

The amino acid sequence of a target protein is threaded on many known 3D structures, and the most compatible 3D structure is searched.
Step 2: Get homologs in PDB and have alignments with 3D modes.

(http://sysimm100.protein.osaka-u.ac.jp/sfas/)

Sequence to Function Annotation Server

Please enter your query

Name: hERG(550-671)
Sequence:

Or

Upload a FASTA-formatted sequence file:

Select Alignment methods

- Blast
- PDBj
- HHpred

Send results to this email address

Input your e-mail address

(hhttp://sysimm100.protein.osaka-u.ac.jp/tmp/SFAS16483/hERG_top.html)
Result of SFAS: The best template is 2r9rB

sequence:  2r9rB  
LGLLIFFLFIFGILFSSAVYFAE---ADER---  
DSQFPSIPDFAFWAVVSMTTGYGDVPTTTIGKIVGSLCAIAAVLTIALPVPVI  
VSNFNYFHYRET  
>hERG  
LFLLMCETFALIAHWLACIYNAIGMNEQPHMDSRIGWLHNLGDQIGKPYNSGSLGG  
PSIKDKYVTALYTFESSLTSVGFGNVSPNTNSEKIFSICVMLIGSLMYASIGFQNV  
SAIQRFLYSGTA

Java Applet Window

Sequence 2 ID: hERG  Residue: GLY (99)

jV version 3
Homology modeling for a target protein

1) Introduction
2) Search for homolog(s)
3) Threading (3D-1D compatibility)
4) Backbone modeling
5) Side-chain modeling
6) Structure optimization
Procedure of Homology Modeling

Search for Homolog
FASTA, BLAST, psi-BLAST

Precise Alignment
3D-1D compatibility search (threading)

Backbone modeling: Loops
loop search, conf. sampling

Side-chain modeling:
Combinatorial problem
Dead End Elimination (DEE)
Loop modeling: Modeling for **deletion** is easy, but for **Insertion** (in particular, with longer than 7 residues) is difficult.

- **Loop Search method**: the known loop fragments are used.

- **Conformational search method**: the most stable loop structure is searched from the possible candidates.
Modeling of a loop structure longer than 10 residues.

RNase A from NMR (2AAS)

Loop R8
with 8 residues

<\text{RMSD}_\alpha> = 0.93 \text{ A}
RMSD_{\text{x}\alpha} = 1.20 \text{ A}

Large red: X-ray str. (8RAT)

Loop R12
with 12 residues

<\text{RMSD}_\alpha> = 2.00 \text{ A}
RMSD_{\text{x}\alpha} = 1.65 \text{ A}

Small red points:
32 NMR str. (2AAS)

Random-walk in the Energy space

\text{(A)}

\text{(B)}

\text{Watanabe, Y. S. et al. (2006) Biophys. 2, 1-12.}
Homology modeling for a target protein

1) Introduction
2) Search for homolog(s)
3) Threading (3D-1D compatibility)
4) Backbone modeling
5) Side-chain modeling
6) Structure optimization
Procedure of Homology Modeling

Search for Homolog
FASTA, BLAST, psi-BLAST

Precise Alignment
3D-1D compatibility search (threading)

Backbone modeling: Loops
loop search, conf. sampling

Side-chain modeling:
Combinatorial problem
Dead End Elimination (DEE)
Side-chain modeling

a) Local stable conformations for individual residue at the energy minima

b) Local stable conformations for individual residue from Statistics in PDB

<table>
<thead>
<tr>
<th>Rotamer</th>
<th>Number</th>
<th>%</th>
<th>Chi 1</th>
<th>Chi 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>39</td>
<td>28.2</td>
<td>$-65.4$ (8.1)</td>
<td>89.3 (9.9)</td>
</tr>
<tr>
<td>-</td>
<td>8</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+$</td>
<td>36</td>
<td>24.5</td>
<td>$-176.4$ (10.2)</td>
<td>$83.1$ (9.2)</td>
</tr>
<tr>
<td>$-$</td>
<td>7</td>
<td>4.8</td>
<td>$183.3$ (10.0)</td>
<td>$188.2$ (34.2)</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>2.0</td>
<td>$44.3$ (20.0)</td>
<td>$60.4$ (18.8)</td>
</tr>
<tr>
<td>Isoleucine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+$</td>
<td>3</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
<td>4.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+$</td>
<td>94</td>
<td>63.9</td>
<td>$-64.4$ (8.2)</td>
<td>176.0 (9.9)</td>
</tr>
<tr>
<td>$-$</td>
<td>36</td>
<td>24.5</td>
<td>$-176.4$ (10.2)</td>
<td>83.1 (9.2)</td>
</tr>
<tr>
<td>$+$</td>
<td>7</td>
<td>4.8</td>
<td>$183.3$ (10.0)</td>
<td>$188.2$ (34.2)</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>2.0</td>
<td>$44.3$ (20.0)</td>
<td>$60.4$ (18.8)</td>
</tr>
<tr>
<td>Threonine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+$</td>
<td>15</td>
<td>16.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-$</td>
<td>12</td>
<td>12.9</td>
<td>$-186.8$ (10.1)</td>
<td>186.0 (9.9)</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>3.2</td>
<td>$-174.9$ (24.9)</td>
<td>72.1 (10.5)</td>
</tr>
<tr>
<td>Cysteine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+$</td>
<td>35</td>
<td>60.6</td>
<td>$-65.2$ (10.1)</td>
<td>82.5 (9.6)</td>
</tr>
<tr>
<td>$-$</td>
<td>23</td>
<td>42.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$+$</td>
<td>13</td>
<td>24.3</td>
<td>$-179.6$ (9.5)</td>
<td>82.5 (9.6)</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Side-chain angles from the rotamer library chi values and standard deviations

Backbone-dependent rotamer library for proteins

<table>
<thead>
<tr>
<th>Rotamer</th>
<th>Number</th>
<th>$\phi$ lower/upper</th>
<th>$\psi$ lower/upper</th>
<th>$\chi$-population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>+60° 180° -60°</td>
<td>+60° 180° -60°</td>
<td>+60° 180° -60°</td>
</tr>
<tr>
<td>Valine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>20</td>
<td>$-180$ 120</td>
<td>$140$ 120</td>
<td>50 45 5</td>
</tr>
<tr>
<td>-</td>
<td>31</td>
<td>$-180$ 140</td>
<td>$180$ 140</td>
<td>48 10 39</td>
</tr>
<tr>
<td>$+$</td>
<td>12</td>
<td>$-180$ 120</td>
<td>$180$ 120</td>
<td>17 0 83</td>
</tr>
<tr>
<td>$+$</td>
<td>50</td>
<td>$-140$ 120</td>
<td>$120$ 120</td>
<td>4 94 2</td>
</tr>
<tr>
<td>$+$</td>
<td>148</td>
<td>$-140$ 120</td>
<td>$120$ 120</td>
<td>8 86 5</td>
</tr>
<tr>
<td>$+$</td>
<td>99</td>
<td>$-140$ 120</td>
<td>$120$ 120</td>
<td>12 35 53</td>
</tr>
<tr>
<td>$+$</td>
<td>50</td>
<td>$-140$ 120</td>
<td>$180$ 180</td>
<td>0 4 96</td>
</tr>
<tr>
<td>$+$</td>
<td>11</td>
<td>$-120$ 100</td>
<td>$-60$ 40</td>
<td>0 82 18</td>
</tr>
<tr>
<td>$+$</td>
<td>20</td>
<td>$-120$ 100</td>
<td>$-20$ 0</td>
<td>5 15 80</td>
</tr>
<tr>
<td>$+$</td>
<td>71</td>
<td>$-120$ 100</td>
<td>$100$ 120</td>
<td>0 97 3</td>
</tr>
<tr>
<td>$+$</td>
<td>181</td>
<td>$-120$ 100</td>
<td>$120$ 140</td>
<td>7 88 4</td>
</tr>
<tr>
<td>$+$</td>
<td>49</td>
<td>$-120$ 100</td>
<td>$120$ 140</td>
<td>4 43 43</td>
</tr>
<tr>
<td>$+$</td>
<td>12</td>
<td>$-120$ 100</td>
<td>$180$ 180</td>
<td>0 4 100</td>
</tr>
<tr>
<td>$+$</td>
<td>13</td>
<td>$-100$ 80</td>
<td>$-60$ 40</td>
<td>8 92 0</td>
</tr>
<tr>
<td>$+$</td>
<td>15</td>
<td>$-100$ 80</td>
<td>$-40$ 20</td>
<td>0 53 47</td>
</tr>
<tr>
<td>$+$</td>
<td>13</td>
<td>$-100$ 80</td>
<td>$20$ 0</td>
<td>23 15 62</td>
</tr>
<tr>
<td>$+$</td>
<td>43</td>
<td>$-100$ 80</td>
<td>$130$ 7</td>
<td>93 0</td>
</tr>
<tr>
<td>$+$</td>
<td>80</td>
<td>$-100$ 80</td>
<td>$120$ 140</td>
<td>6 88 6</td>
</tr>
<tr>
<td>$+$</td>
<td>29</td>
<td>$-100$ 80</td>
<td>$140$ 160</td>
<td>14 41 45</td>
</tr>
<tr>
<td>$+$</td>
<td>207</td>
<td>$-80$ 60</td>
<td>$-60$ 40</td>
<td>2 97 0</td>
</tr>
<tr>
<td>$+$</td>
<td>131</td>
<td>$-80$ 60</td>
<td>$-40$ 20</td>
<td>11 60 28</td>
</tr>
<tr>
<td>$+$</td>
<td>19</td>
<td>$-80$ 60</td>
<td>$-20$ 0</td>
<td>32 21 47</td>
</tr>
<tr>
<td>$+$</td>
<td>15</td>
<td>$-80$ 60</td>
<td>$100$ 120</td>
<td>0 93 7</td>
</tr>
<tr>
<td>$+$</td>
<td>62</td>
<td>$-80$ 60</td>
<td>$120$ 140</td>
<td>2 94 5</td>
</tr>
<tr>
<td>$+$</td>
<td>27</td>
<td>$-80$ 60</td>
<td>$140$ 160</td>
<td>0 56 44</td>
</tr>
<tr>
<td>$+$</td>
<td>109</td>
<td>$-60$ 40</td>
<td>$-40$ 20</td>
<td>2 92 6</td>
</tr>
<tr>
<td>$+$</td>
<td>39</td>
<td>$-60$ 40</td>
<td>$-20$ 0</td>
<td>28 54 18</td>
</tr>
<tr>
<td>$+$</td>
<td>16</td>
<td>$-60$ 40</td>
<td>$120$ 140</td>
<td>6 75 19</td>
</tr>
</tbody>
</table>

Side-chain modeling

c) Combinatorial approach (Monte Carlo method, GA, DEE, etc.)

The protein jigsaw puzzle. At first sight the solution is easy because there is a known backbone structure (green) to copy. But packing the side-chains (small red and black circles) is difficult, because for each piece there are a number of alternatives (rotamers) only one of which will appear in the completed picture at any position. The approach of Desmet et al. can be explained, in simplified terms, by considering the options for the residue (C) at the second position. If there are three rotamers for C and two rotamers for S, then each C is tried with each S at the first and third positions. If there is a rotamer of C that will not fit with any S at either adjacent position (or with G at the thirteenth position), then that piece cannot be part of the final picture and can be thrown away. This test is applied to all positions, so reducing the number of pieces that need to be considered when it comes to the final (combinatorial) assembly stage.

c) Combinatorial approach

**Dead-end elimination (DEE) method**


Structural energy for side-chains of N-residues is described by the interaction energy between the backbone and the side-chain, $E_1$, and the interaction energy between the side-chains, $E_2$.

$$ c (r_1, r_2, ..., r_N) = \sum_{i} E_1 (r_i) + \sum_{i<j} E_2 (r_i, r_j) \quad (1) $$

**Theorem:** When the $t$'th rotamer ($t_i$) of the $i$'th residue is found, satisfying the next equation for the $r$'th rotamer ($r_i$) of the $i$'th residue, then the global energy minimum conformation does not include the $r_i$.

$$ E_1 (r_i) + \min_{i\neq j} |E_2 (r_i, s_j)| > E_1 (t_i) + \max_{i\neq j} |E_2 (t_i, s_j)| \quad (2) $$

Using the above theorem, it is possible to find the global energy minimum conformation, gradually rejecting the non-probable side-chain structures that cannot be included in the global energy minimum conformation.
a) White: Core regions. Black: all side-chains

b) Result of DEE for lysozyme. Blue: X-ray crystal structure, Yellow: DEE model structure, white: the coincident side-chain structures in between the crystal and the DEE model.

Homology modeling for a target protein

1) Introduction
2) Search for homolog(s)
3) Threading (3D-1D compatibility)
4) Backbone modeling
5) Side-chain modeling
6) Structure optimization
Stress in the 3D structural model is removed by Minimization/MD with the potential energy $U$. 

**Force Fields**

\[
U = \sum_{\text{bonds}} \frac{1}{2} k_r (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{\text{improper}} V(\text{improper torsion}) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}} + \sum_{\text{LJ}} \left[ \frac{A_{ij}}{r_{12}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right]
\]

- Bond stretches (1-2)
- Angle bending (1-3)
- Torsional rotation (1-4)
- Improper torsion (1-4)
- Electrostatic interaction (1-5)
- Lennard-Jones interaction (1-5)
Web site for homology modeling

http://swissmodel.expasy.org/

http://salilab.org/modeller/

**SWISS-MODEL**

An Automated Comparative Protein Modelling Server

**SIB** - Biozentrum Basel site provided by:

**BIOZENTRUM SIB**

SWISS-MODEL is a fully automated protein-structure homology-modelling server, accessible via the EMBL web server, or from the program *Dredd* (Swiss-Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists worldwide.

The present version of the server is 3.5 and is under constant improvement and debugging. In order to help us define the sequence analysis and modelling algorithms, please report of possible bugs and problems with the modelling procedure.

SWISS-MODEL was initiated in 1990 by Manuel Hutter, and is now being further developed within the SIB-EMBL Institute of Bioinformatics in collaboration between Thomas Schwede at the Structural Bioinformatics Laboratory at the University of Basel and Mieszko Lis at SWISS-MODEL.

The computational resources for the SWISS-MODEL server are provided in collaboration by the Biozentrum (University Basel) and the Advanced Bioimaging Computing Center (NCI-Frederick, USA).

**Modeller**

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints

New! **Spanner**

http://www.pdbj.org/spanner/

**Constructed by**

Daron M. Standley (iFREC, Osaka U),
Mieszko Lis (MIT),
Haruki Nakamura (IPR, Osaka U)