XXXVI Int. Congress of Phys. Sci.
TU-3 July 29, 2009

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

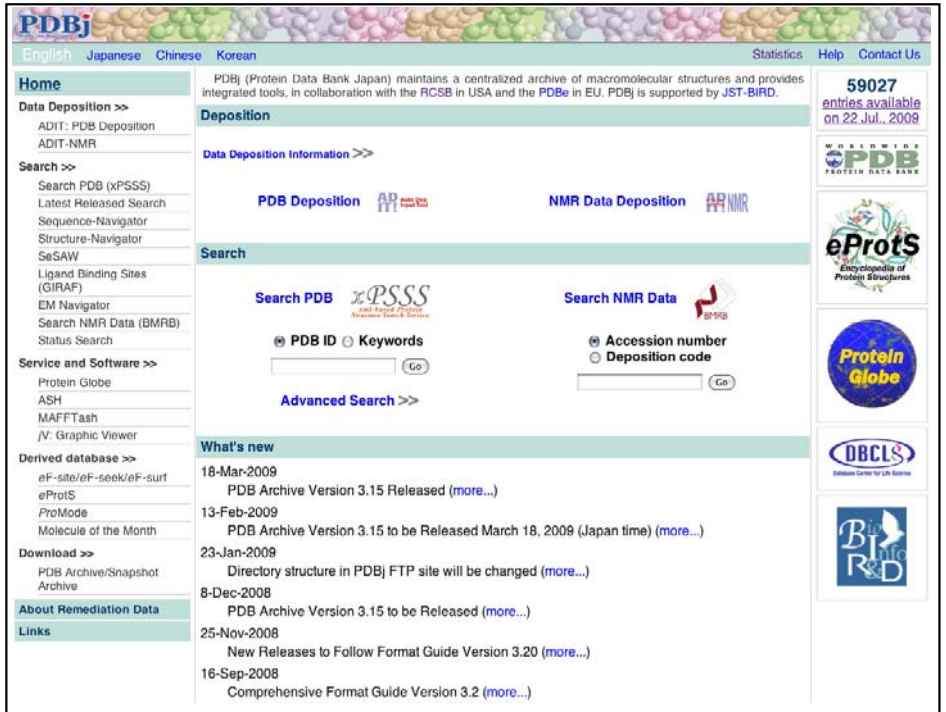
The center for Advanced Medical Engineering and Informatics, Osaka University

http://www.protein.osaka-u.ac.jp/rcsfp/pi/
http://www.pdbj.org/

PDBj
Protein Data Bank Japan

Protein Data Bank Japan

http://www.pdbj.org/



- Structure Data curation and editing
- Structure Data browsing and downloading

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).

WORLDWIDE PDB
PROTEIN DATA BANK

http://www.wwpdb.org

PDBe
PROTEIN DATA BANK EUROPE

PDBj
Protein Data Bank Japan

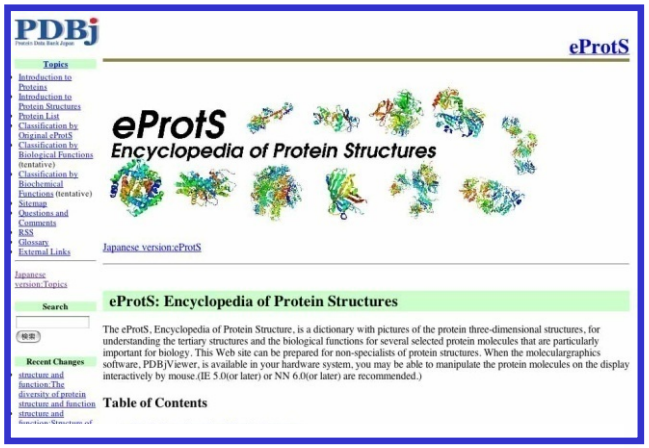
BMRB

RCSB PDB
PROTEIN DATA BANK

wwPDB and wwPDBAC members at EBI, Hinxton on 29 Sept, 2008

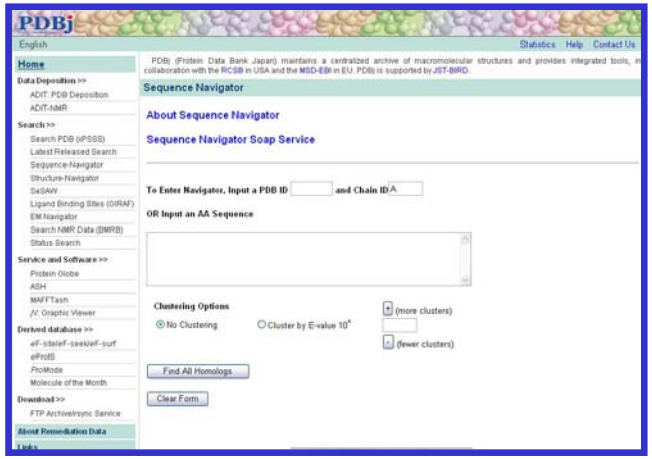
Development of other Databases and Services

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

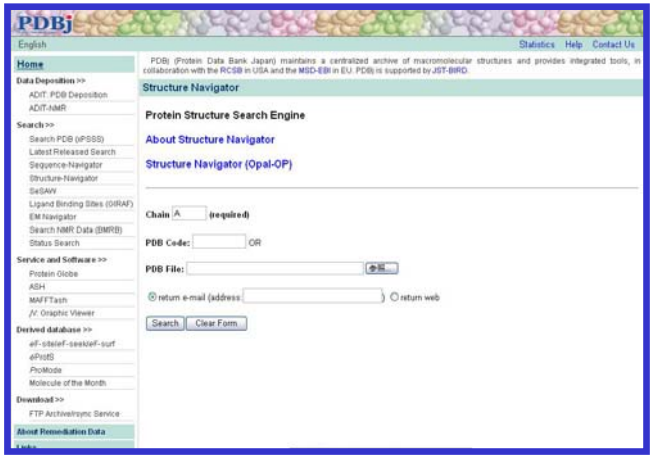


Encyclopedia of Protein Structures, eProtS
(Kinjyo, Kudo, & Ito)

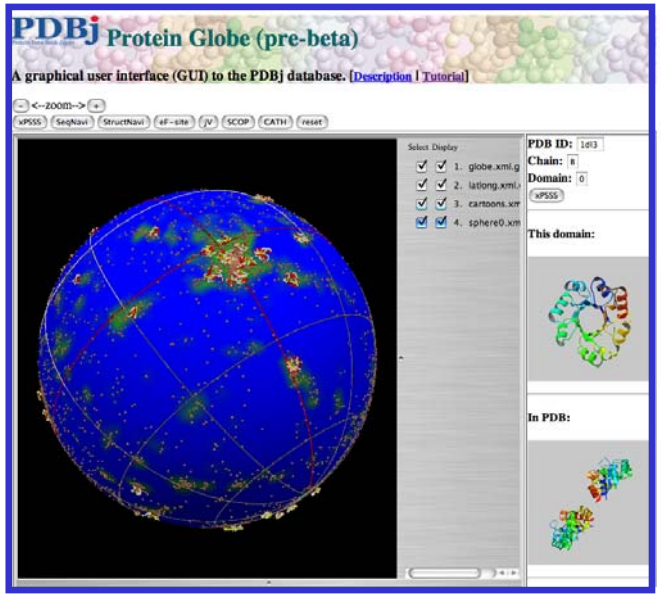
Molecular of the Month, MoM
(Goodsell & Kudo)



Homolog protein search, Sequence Navigator
(Standley)

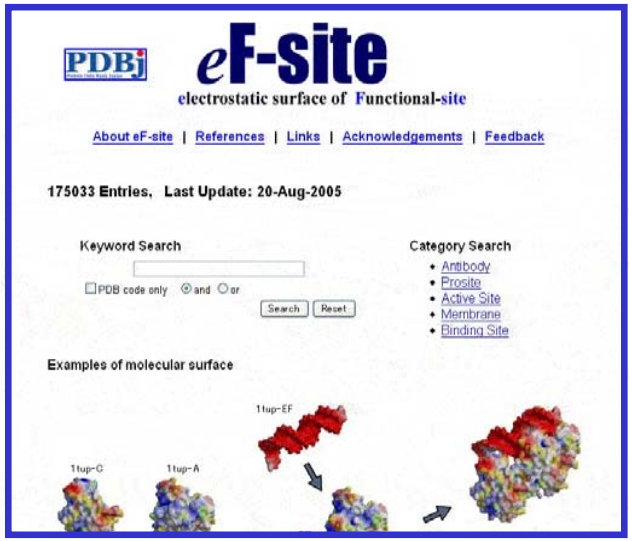


Similar fold search, Structure Navigator
(Standley & Toh)



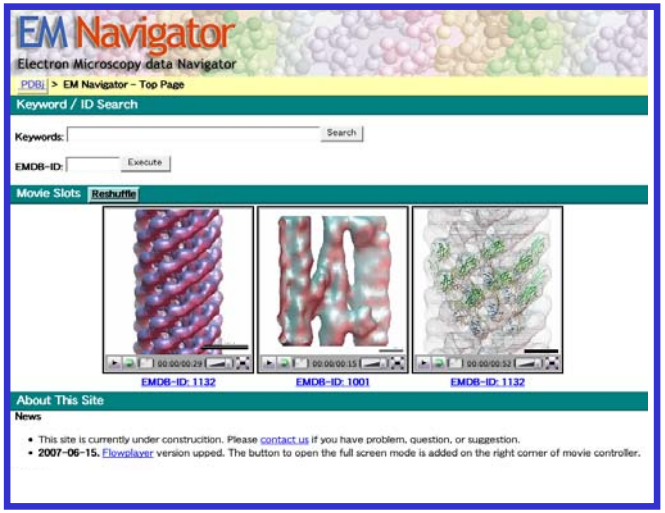
Protein Folds Browser, Protein Globe *(Kinjo & Standley)*

Development of other Databases and Services

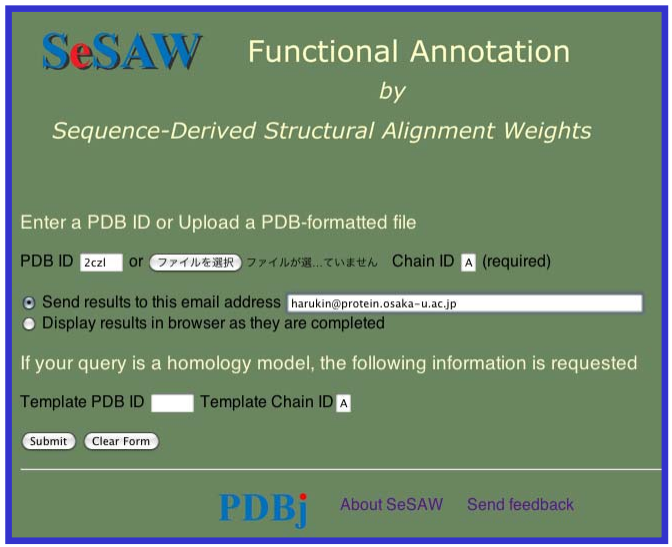


Protein Molecular Surface Database, *e*F-site *(Kinoshita & Nakamura)*

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

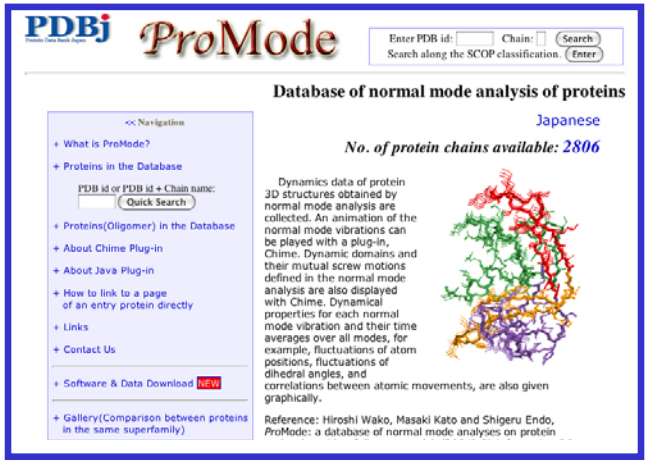


Electron Microscopy Navigator, EM-Navi *(Suzuki)*



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

Ligand Binding Site Search, GIRAF *(Kinjo)*



Protein Dynamics Database, *Pro*Mode *(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)

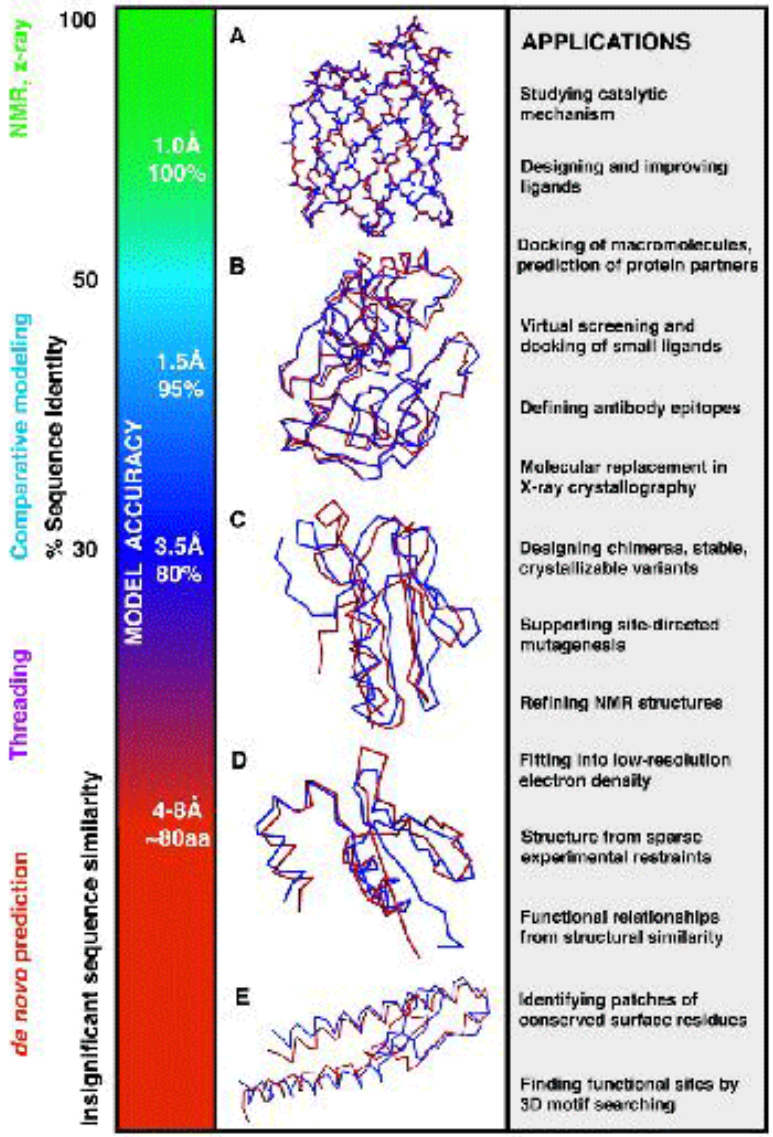
  3D structural model can be made when any structure of the family protein is in DB.

  ・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.



D. Baker & A. Sali (2001) Science 294, 93-96.

Simple Example

Swine Influenza A virus neuraminidase (NA) gene

GenBank: FJ981614.1

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

Comment Features Sequence

```
LOCUS       FJ981614                1410 bp    cRNA     linear   VRL 01-MAY-2009
DEFINITION  Influenza A virus (A/Texas/04/2009(H1N1)) segment 6 neuraminidase
            (NA) gene, complete cds.
ACCESSION   FJ981614
VERSION     FJ981614.1  GI:229299518
PROJECT     GenomeProject:37813
DBLINK      Project:37813
KEYWORDS    .
SOURCE      Influenza A virus (A/Texas/04/2009(H1N1))
  ORGANISM  Influenza A virus (A/Texas/04/2009(H1N1))
            Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
            Influenzavirus A.
REFERENCE   1  (bases 1 to 1410)
  AUTHORS   Shu,B., Balish,A., Garten,R., Smith,C., Emery,S., Barnes,J.,
            Deyde,V., Klimov,A. and Cox,N.
  TITLE     Human infection with novel swine H1N1 influenza
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1410)
  AUTHORS   Shu,B., Balish,A., Garten,R., Smith,C., Emery,S., Barnes,J.,
            Deyde,V., Klimov,A. and Cox,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2009) WHO Collaborating Center for Surveillance
            Epidemiology and Control of Influenza, Influenza Division, Centers
            for Disease Control and Prevention, 1600 Clifton Road, N.E.,
            Atlanta, GA 30333, USA
COMMENT     Swine influenza A (H1N1) virus isolated during human swine flu
            outbreak of 2009.  For more information, see http://www.cdc.gov/.

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

            ##EpifluData-START##
            Isolate                 A/Texas/04/2009
            Subtype                 H1N1
            Segment_name            NA
            Host_gender             M
            Host_age                16
            Passage_history         X/C1
            Adamantane_resistance   resistant
            Zanamivir_resistance    sensitive
            Oseltamivir_resistance  sensitive
            Country                 USA
            State/Province          Texas state
            Collection_day          14
            Collection_month        4
            Collection_year         2009
            EPI_accession           EPI177301
            ##EpifluData-END##
```

Change Region Shown

Customize View

Sequence Analysis Tools

- BLAST Sequence
- Pick Primers

Influenza Viral Resource

Flu-related NCBI resources including sequences, alignments, phylogeny and literature.

Recent Activity

Turn Off Clear

Influenza A virus (A/Texas/04/2009(H1N1))

All links from this record

- Protein
- Taxonomy
- Related Sequences

```
FEATURES             Location/Qualifiers
     source          1..1410
                     /organism="Influenza A virus (A/Texas/04/2009(H1N1))"
                     /mol_type="viral cRNA"
                     /strain="A/Texas/04/2009"
                     /serotype="H1N1"
                     /host="Homo sapiens; gender M; age 16"
                     /db_xref="taxon:641811"
                     /segment="6"
                     /country="USA: Texas state"
                     /collection_date="14-Apr-2009"
     gene            1..1410
                     /gene="NA"
     CDS             1..1410
                     /gene="NA"
                     /codon_start=1
                     /product="neuraminidase"
                     /protein_id="ACQ55360.1"
                     /db_xref="GI:229299519"
                     /translation="MNPNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQN
                     QIETCNQSVITYENNTWVNQTYVNISNTNFAAGQSVVSVKLAGNSSLCPVSGWAIYSK
                     DNSVRIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLMS
                     CPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTIKS
                     WRNNILRTQESECACVNGSCFTVMTDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHY
                     EECSCYPDSSEITCVCRDNWHGSNRPWVSFNQNLEYQIGYICSGIFGDNPRPNDKTGS
                     CGPVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFEMIWDPNGWTGTDNNFSIKQD
                     IVGINEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTIWTSGSSISFCGVNS
                     DTVGWSWPDGAELPFTIDK"
ORIGIN
        1 atgaatccaa accaaaagat aataaccatt ggttcggtct gtatgacaat tggaatggct
       61 aacttaatat tacaaattgg aaacataatc tcaatatgga ttagccactc aattcaactt
      121 gggaatcaaa atcagattga aacatgcaat caaagcgtca ttacttatga aaacaacact
      181 tgggtaaatc agacatatgt taacatcagc aacaccaact ttgctgctgg acagtcagtg
      241 gtttccgtga aattagcggg caattcctct ctctgccctg ttagtggatg ggctatatac
      301 agtaaagaca acagtgtaag aatcggttcc aagggggatg tgtttgtcat aagggaacca
      361 ttcatatcat gctccccctt ggaatgcaga accttcttct tgactcaagg ggccttgcta
      421 aatgacaaac attccaatgg aaccattaaa gacaggagcc catatcgaac cctaatgagc
      481 tgtcctattg gtgaagttcc ctctccatac aactcaagat ttgagtcagt cgcttggtca
      541 gcaagtgctt gtcatgatgg catcaattgg ctaacaattg gaatttctgg cccagacaat
      601 ggggcagtgg ctgtgttaaa gtacaacggc ataataacag acactatcaa gagttggaga
      661 aacaatatat tgagaacaca agagtctgaa tgtgcatgtg taaatggttc ttgctttact
      721 gtaatgaccg atggaccaag taatggacag gcctcataca agatcttcag aatagaaaag
      781 ggaaagatag tcaaatcagt cgaaatgaat gcccctaatt atcactatga ggaatgctcc
      841 tgttatcctg attctagtga aatcacatgt gtgtgcaggg ataactggca tggctcgaat
      901 cgaccgtggg tgtctttcaa ccagaatctg gaatatcaga taggatacat atgcagtggg
      961 attttcggag acaatccacg ccctaatgat aagacaggca gttgtggtcc agtatcgtct
     1021 aatggagcaa atggagtaaa aggattttca ttcaaatacg gcaatggtgt ttggataggg
     1081 agaactaaaa gcattagttc aagaaacggt tttgagatga tttgggatcc gaacggatgg
     1141 actgggacag acaataactt ctcaataaag caagatatcg taggaataaa tgagtggtca
     1201 ggatatagcg ggagttttgt tcagcatcca gaactaacag ggctggattg tataagacct
     1261 tgcttctggg ttgaactaat cagagggcga cccaaagaga acacaatctg gactagcggg
     1321 agcagcatat ccttttgtgg tgtaaacagt gacactgtgg gttggtcttg gccagacggt
     1381 gctgagttgc catttaccat tgacaagtaa
//
```

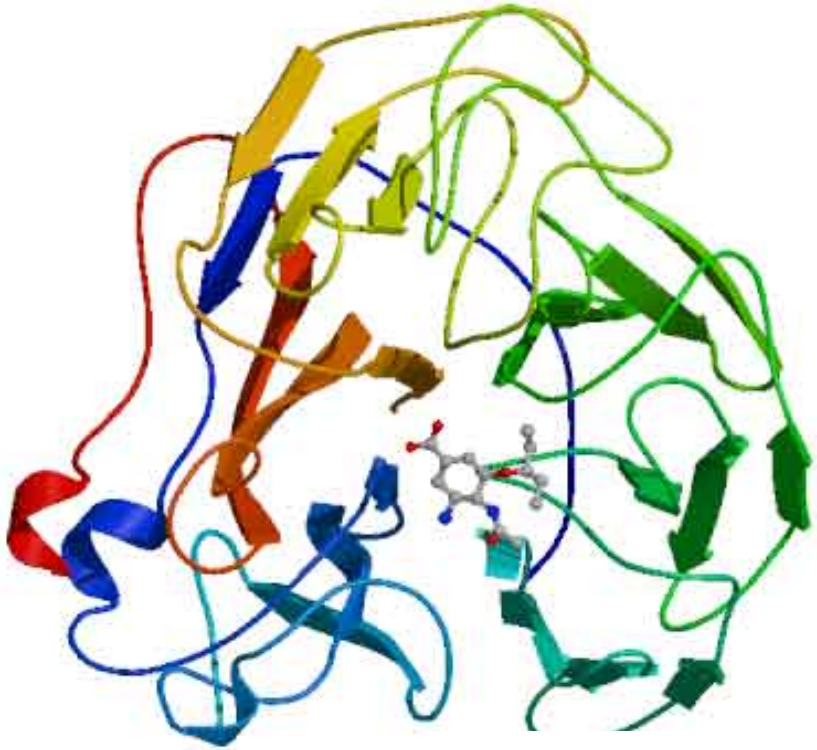
AA sequence

NA sequence

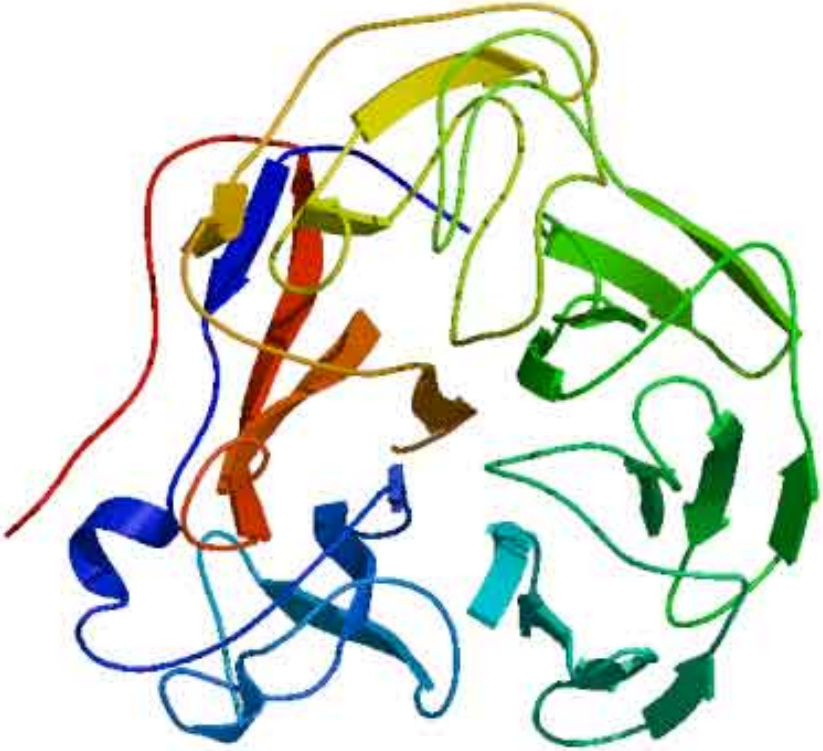
>2qwk chain-A: NEURAMINIDASE
>Influenza A virus neuraminidase(NA) gene, complete cds. FJ981614

```
RDFNNLTKGLCTINSWHIYGKDNAVRIGEDSDVLVTREPYVSCDPDECRFYALSQGTTIRGKHSNGTIHDRSQYRALI
         :: .. : :: ::: ::::   ::.: :::.::: : ::: . :.::  .  :::::::.::: :: :.
---------LCPVSGWAIYSKDNSVRIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLM

SWPLSSPPTVYNSRVECIGWSSTSCHDGKTRMSICISGPNNNASAVIWYNRRPVTEINTWARNILRTQESECVCHNGV
: :.   :. .::::.: ..:: . ::::   ..: :::::.: : ::. ::      : .:  :::::::::: : ::
SCPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGS

CPVVFTDGSATGPAETRIYYFKEGKILKWEPLAGTAKHIEECSCYGERAEITCTCRDNWQGSNRPVIRIDPVANTHTS
:  :::::   : :  .:. .  ::::.:   . .   : :::::: .  :::: ::::: ::::: .  .   .
CFTVMTDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSCYPDSSEITCVCRDNWHGSNRPWVSFNQ-NLEYQI

QYICSPVLTDNPRPNDPTVGKCNDPYPGNNNNGVKGFSYLDGVNTWLGRTISIASRSGYEMLKVPNALTDDKSKPTQG
 :::: .. ::::::: : : :  :  :  :::::::   :   :.::: :: :: :.::.  ::. :      .
GYICSGIFGDNPRPNDKT-GSCG-PVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFEMIWDPNGWTGTDNNFSIK

QTIVLNTDWSGYSGSFMDY—-WAEGECYRACFYVELIRGRPKEDKVWWTSNSIVSMCSSTEFLGQWDWPDGAKIEYFL
: ::   .:::::::::       .: : ::::::::::::. . :: : .:.:        : ::::: . .
QDIVGINEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTI-WTSGSSISFCGVNSDTVGWSWPDGAELPF—
```

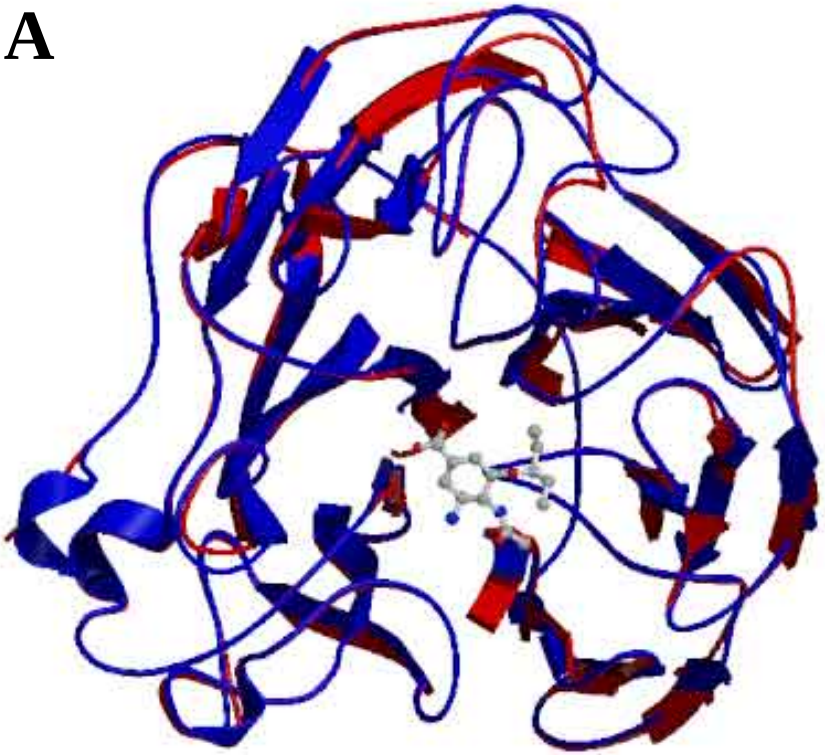
Ins/del



2qwk A



Swine NA model



Blue: 2qwk A

Red:Swine NA model

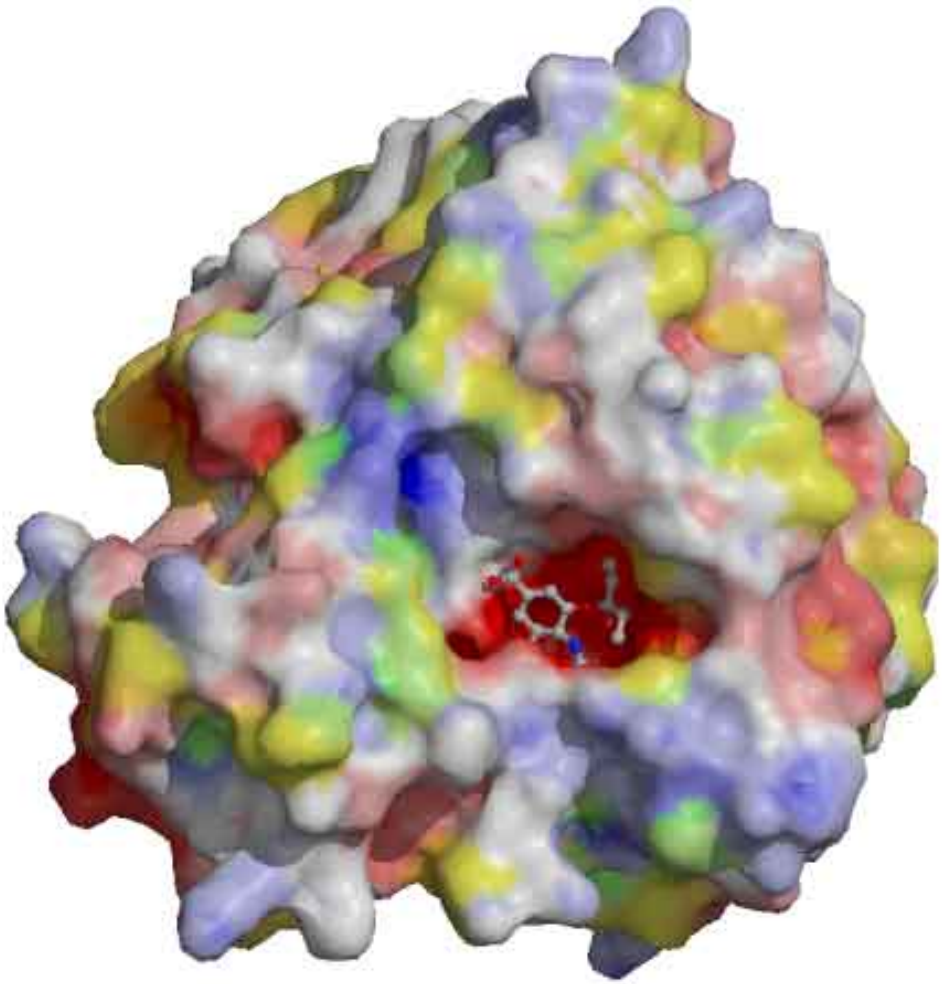
>2qwk chain-A: NEURAMINIDASE
>Influenza A virus neuraminidase(NA) gene, complete cds. FJ981614

```
RDFNNLTKGLCTINSWHIYGKDNAVRIGEDSDVLVTREPYVSCDPDECRFYALSQGTTIRGKHSNGTIHDRSQYRALI
         :: .. : :: ::: ::::   ::.: :::.:: : ::: . :.::  .  :::::::.::: :: :.
---------LCPVSGWAIYSKDNSVRIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLM

SWPLSSPPTVYNSRVECIGWSSTSCHDGKTRMSICISGPNNNASAVIWYNRRPVTEINTWARNILRTQESECVCHNGV
: :.   :. .::::: ..:: . ::::   ..: :::::: : ::. ::      : .:  :::::::::: : ::
SCPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGS

CPVVFTDGSATGPAETRIYYFKEGKILKWEPLAGTAKHIEECSCYGERAEITCTCRDNWQGSNRPVIRIDPVAMTHTS
:  :::::   : :  .:. .  ::::.:   . .   : :::::: .  :::: ::::: ::::: .  .   .
CFTVMTDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSCYPDSSEITCVCRDNWHGSNRPWVSFNQ-NLEYQI

QYICSPVLTDNPRPNDPTVGKCNDPYPGNNNNGVKGFSYLDGVNTWLGRTISIASRSGYEMLKVPNALTDDKSKPTQG
 :::: .. ::::::: : : :  :   :  :::::::   :   :.::: :: :: :.::.  ::. :      .
GYICSGIFGDNPRPNDKT-GSCG-PVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFEMIWDPNGWTGTDNNFSIK

QTIVLNTDWSGYSGSFMDY—-WAEGECYRACFYVELIRGRPKEDKVWWTSNSIVSMCSSTEFLGQWDWPDGAKIEYFL
: ::   .:::::::::.        .: : ::::::::::::. . ::: : .:.:        : ::::: . .
QDIVGINEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTI-WTSGSSISFCGVNSDTVGWSWPDGAELPF—
```

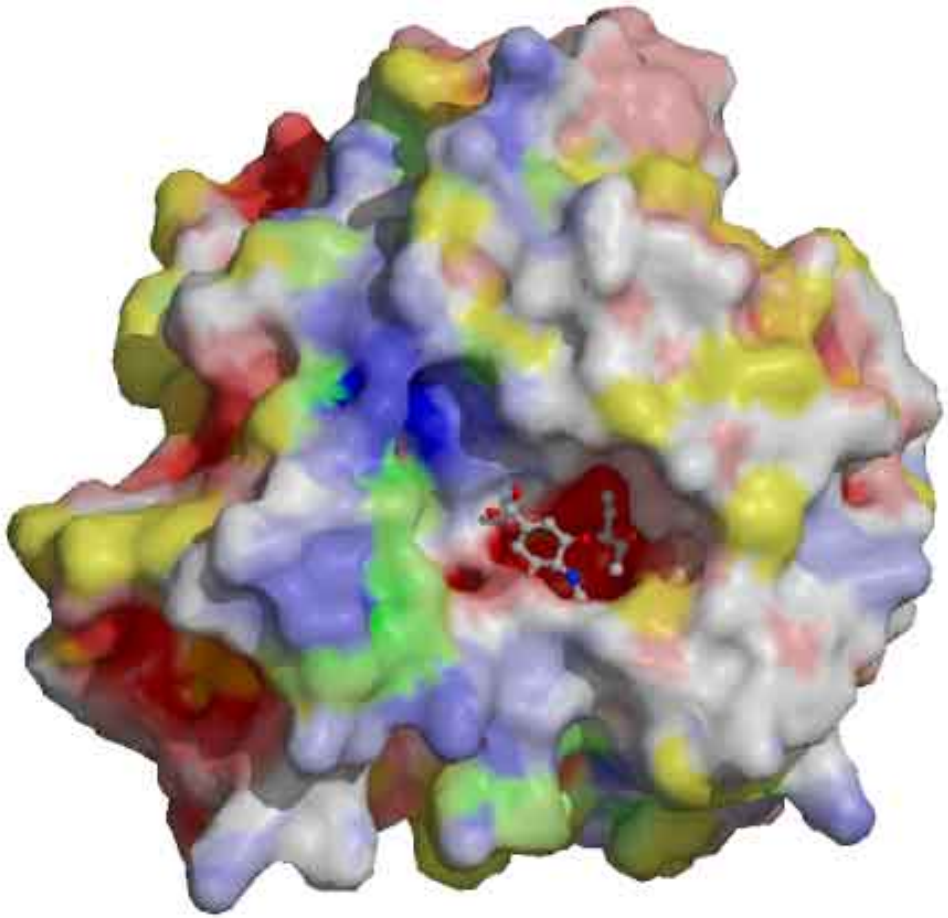
Red: Active site residues surrounding Tamiflu

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

```
RDFNNLTKGLCTINSWHIYGKDNAVRIGEDSDVLVTREPYVSCDPDECRFYALSQGTTIRGKHSNGTIHDRSQYRALI
         :: .. : :: ::: ::::   ::.: :::.::: : ::: . :.::  .  :::::::.::: :: :.
---------LCPVSGWAIYSKDNSVRIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLM

SWPLSSPPTVYNSRVECIGWSSTSCHDGKTRMSICISGPNNNASAVIWYNRRPVTEINTWARNILRTQESECVCHNGV
: :.   :. .::::: ..:: . ::::   ..: :::::: : ::. ::      : .:  :::::::::: : ::
SCPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTIKSWRNNILRTQESECACVNGS

CPVVFTDGSATGPAETRIYYFKEGKILKWEPLAGTAKHIEECSCYGERAEITCTCRDNWQGSNRPVIRIDPVAMTHTS
:  :::::   : :  .:. .  ::::.:   . .   : :::::: .  :::: ::::: ::::: .  .   .
CFTVMTDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSCYPDSSEITCVCRDNWHGSNRPWVSFNQ-NLEYQI
```

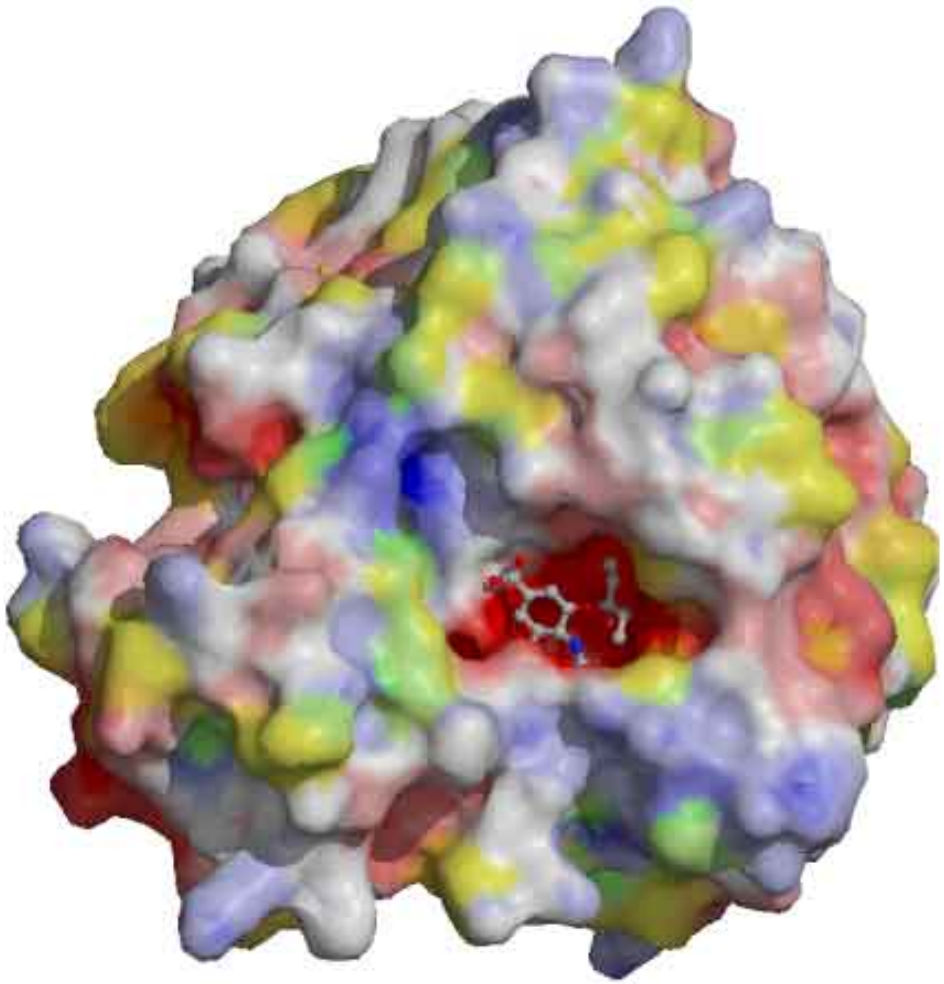
H274Y: Tamiflu resistant

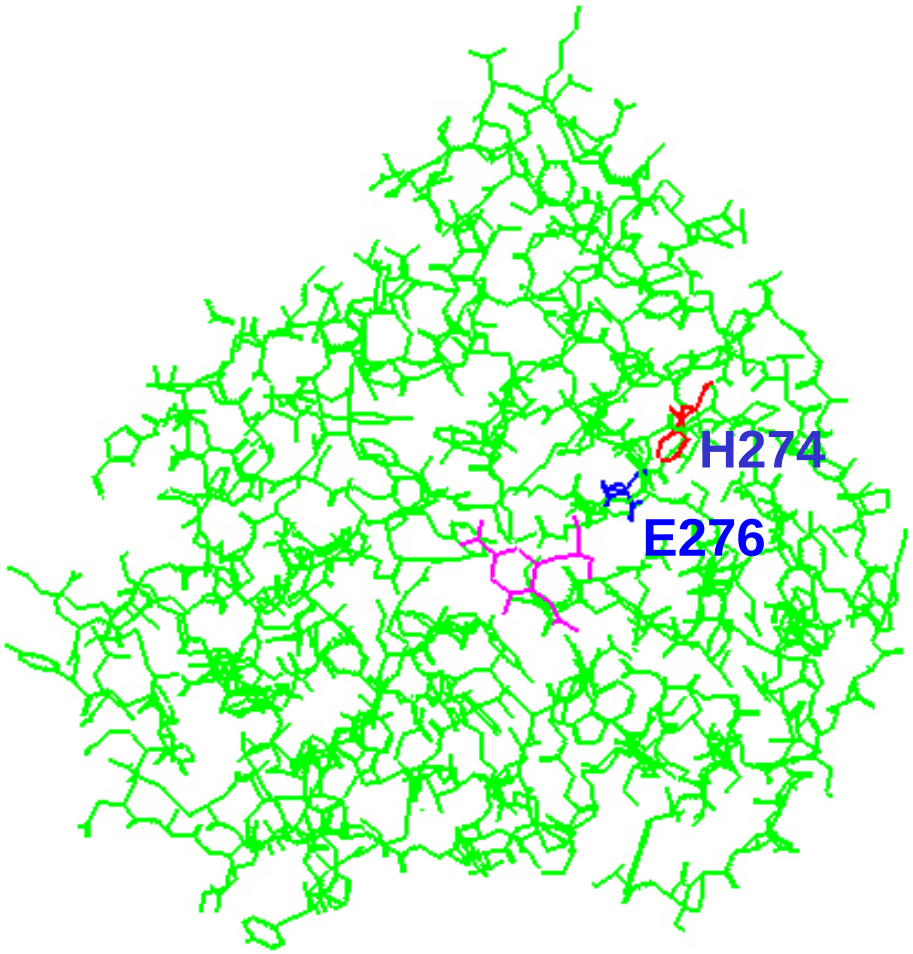
```
QYICSPVLTDNPRPNDPTVGKCNDPYPGNNNNGVKGFSYLDGVNTWLGRTISIASRSGYEMLKVPNALTDDKSKPTQG
 :::: .. ::::::: : : :  :   :  :::::::   :   :.::: :: :: :.::.  ::. :      .
GYICSGIFGDNPRPNDKT-GSCG-PVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFEMIWDPNGWTGTDNNFSIK

QTIVLNTDWSGYSGSFMDY—-WAEGECYRACFYVELIRGRPKEDKVWWTSNSIVSMCSSTEFLGQWDWPDGAKIEYFL
: ::   .:::::::::.        .: : ::::::::::::. . ::: : .:.:        : ::::: . .
QDIVGINEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTI-WTSGSSISFCGVNSDTVGWSWPDGAELPF—
```

Red: Active site residues surrounding Tamiflu

Electrostatic molecular surfaces

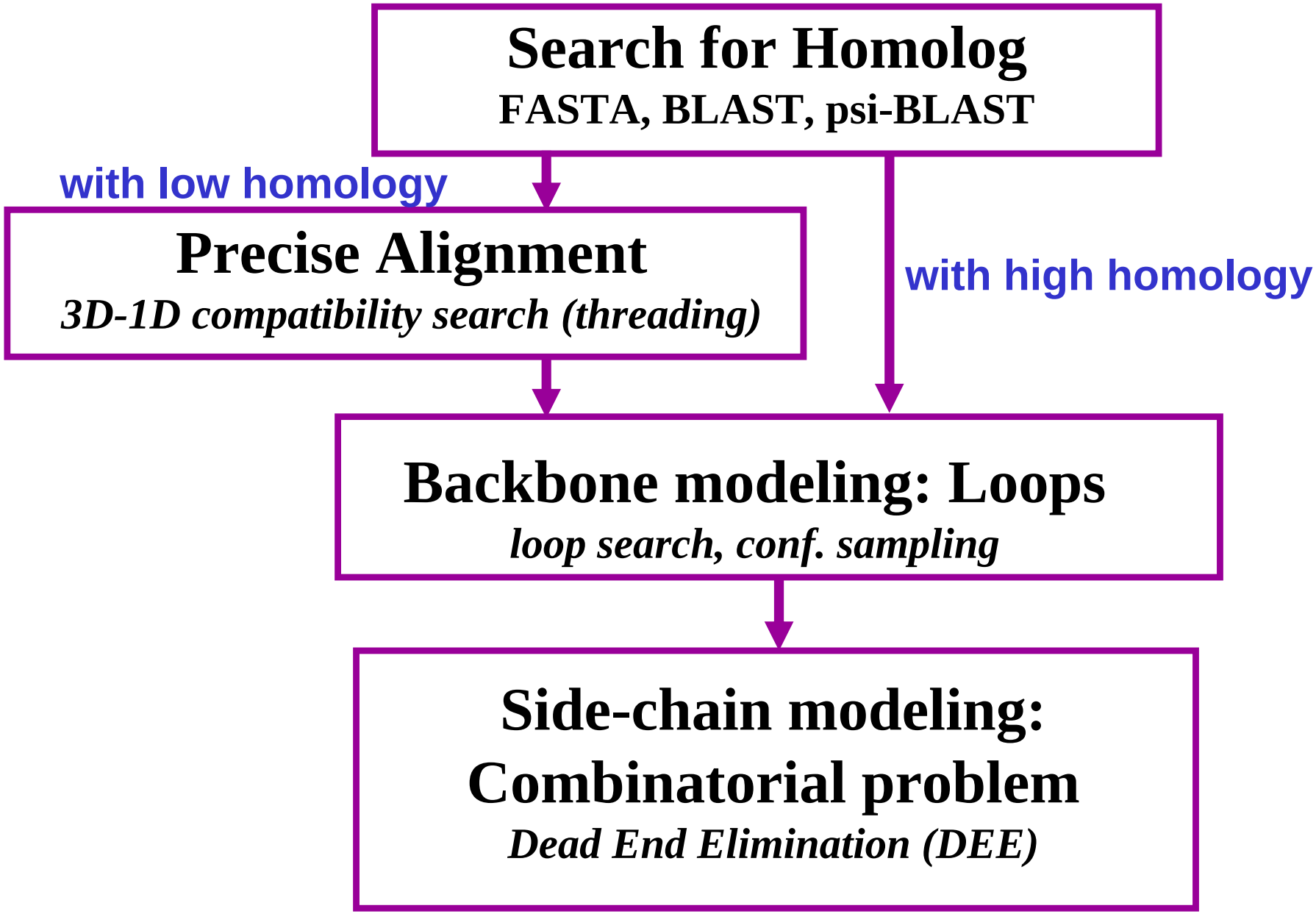
Blue: positive, Red: negative, yellow: hydrophobic





2qwk A

Procedure of Homology Modeling



Homology modeling for a target protein

1) Introduction

2) Search for homolog(s)

3) Threading (3D-1D compatibility)

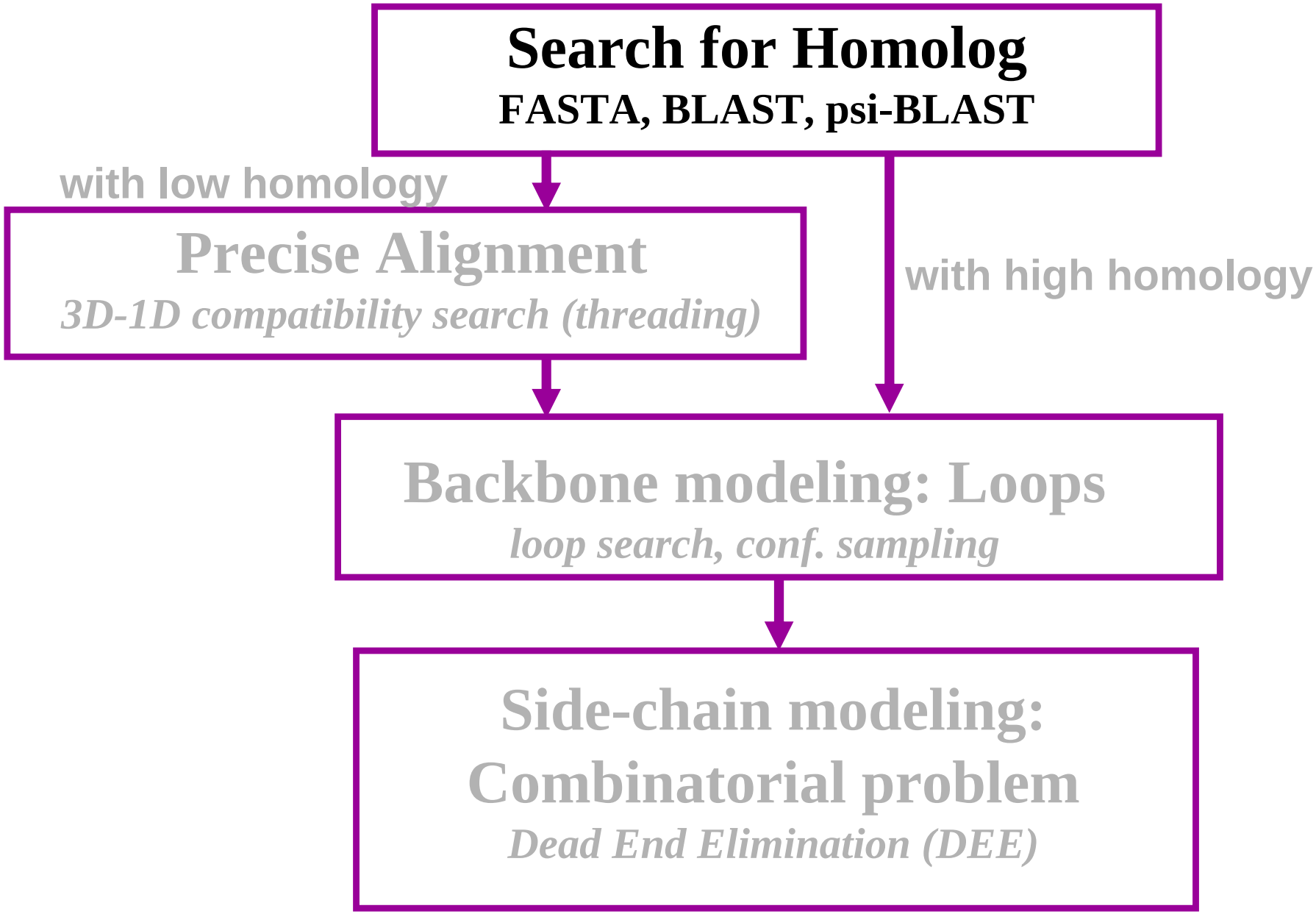
4) Backbone modeling

5) Side-chain modeling

6) Structure optimization

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

Procedure of Homology Modeling



Get amino-acid sequence of hERG channel from NCBI

(*http://www.ncbi.nlm.nih.gov/*)

NCBI Entrez Protein

My NCBI [Sign In] [Register]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search Protein for Go Clear

Limits | Preview/Index | History | Clipboard | Details

Format: GenPept FASTA Graphics More Formats ▼ Download ▼ Save ▼ Links ▼

Try the Graphics report for a more informative view of the biological features.

Swiss-Prot: Q12809.1

RecName: Full=Potassium voltage-gated channel subfamily H member 2; AltName: Full=Voltage-gated potassium channel subunit Kv11.1; AltName: Full=Ether-a-go-go-related gene potassium channel 1; Short=H-ERG; Short=Erg1; Short=Ether-a-go-go-related protein 1; Shor...

Comment Features Sequence

```
LOCUS       Q12809                  1159 aa            linear   PRI 07-JUL-2009
DEFINITION  RecName: Full=Potassium voltage-gated channel subfamily H member 2;
            AltName: Full=Voltage-gated potassium channel subunit Kv11.1;
            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 homolog.
ACCESSION   Q12809
VERSION     Q12809.1  GI:7531135
DBSOURCE    UniProtKB: locus KCNH2_HUMAN, accession Q12809;
            class: standard.
            extra accessions:O75418,O75680,Q9BT72,Q9BUT7,Q9H3P0
            created: May 30, 2000.
            sequence updated: Nov 1, 1996.
            annotation updated: Jul 7, 2009.
            xrefs: U04270.1, AAA62473.1, AB009071.2, BAA37096.1, AF363636.1,
            AAL37559.1, AB044806.1, BAB19682.1, AJ512214.1, CAD54447.1,
            AJ010538.1, CAA09232.1, AJ010539.1, AJ010540.1, AJ010541.1,
            AJ010542.1, AJ010543.1, AJ010544.1, AJ010545.1, AJ010546.1,
            AJ010547.1, AJ010548.1, AJ010549.1, AJ010550.1, AJ010551.1,
            AF052728.1, AAC69709.1, BC001914.1, AAH01914.2, BC004311.2,
            AAH04311.2, I38465, NP_000229.1, NP_742053.1, NP_742054.1, 1BYW_A,
            1UJL_A
            xrefs (non-sequence databases): IPI:IPI00029662, IPI:IPI00172614,
            IPI:IPI00221190, IPI:IPI00221191, UniGene:Hs.647099, PDBsum:1BYW,
            PDBsum:1UJL, IntAct:Q12809, TCDB:1.A.1.20.1, PhosphoSite:Q12809,
            PRIDE:Q12809, Ensembl:ENSG00000055118, GeneID:3757, KEGG:hsa:3757,
            UCSC:uc003wib.1, UCSC:uc003wic.1, UCSC:uc003wie.1,
            GeneCards:GC07M150272, H-InvDB:HIX0007217, HGNC:6251,
            HPA:CAB006838, MIM:152427, MIM:609620, Orphanet:130, Orphanet:768,
            Orphanet:51083, PharmGKB:PA212, HOGENOM:Q12809, HOVERGEN:Q12809,
            OMA:Q12809, DrugBank:DB01118, DrugBank:DB00276, DrugBank:DB00637,
            DrugBank:DB01136, DrugBank:DB00604, DrugBank:DB00204,
            DrugBank:DB01218, DrugBank:DB00308, DrugBank:DB01100,
```

Change Region Shown

Customize View

Sequence Analysis Tools

- BLAST Sequence
- Conserved Domains

Articles about the KCNH2 gene

- Genetic Polymorphism of KCNH2 Confers Predispos [J Cardiovasc Electrophysiol. 2009]
- Breaking the gene barrier in schizophrenia. [Nat Med. 2009]
- Interactions of H562 in the S5 helix with T618 and S621 in the pore helix a [Biophys J. 2009]

» See all...

Identical Proteins for Q12809.1

- Sequence 2 from patent US 7541 [ACS12627]
- Sequence 5 from patent US 7537 [ACS08477]
- Sequence 2 from patent US 7510 [ACQ19114]

» See all...

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 Drosophila ether...

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

Homologs of the KCNH2 gene

The KCNH2 gene is conserved in chimpanzee,

```
     Region          1055
                     /gene="KCNH2"
                     /gene_synonym="ERG"
                     /gene_synonym="ERG1"
                     /gene_synonym="HERG"
                     /gene_synonym="HERG1"
                     /region_name="Variant"
                     /experiment="experimental evidence, no additional details
                     recorded"
                     /note="R -> Q (in dbSNP:rs41307270). /FTId=VAR_036682."
     Site            1137
                     /gene="KCNH2"
                     /gene_synonym="ERG"
                     /gene_synonym="ERG1"
                     /gene_synonym="HERG"
                     /gene_synonym="HERG1"
                     /site_type="mutagenized"
                     /experiment="experimental evidence, no additional details
                     recorded"
                     /note="S->A: Abolishes phosphorylation; when associated
                     with A-283; A-890 and A-895."
ORIGIN
        1 mpvrrghvap qntfldtiir kfegqsrkfi ianarvenca viycndgfce lcgysraevm
       61 qrpctcdflh gprtqrraaa qiaqallgae erkveiafyr kdgscflclv dvvpvknedg
      121 avimfilnfe vvmekdmvgs pahdtnhrgp ptswlapgra ktfrlklpal laltaressv
      181 rsggaggaga pgavvvdvdl tpaapssesl aldevtamdn hvaglgpaee rralvgpgsp
      241 prsapgqlps prahslnpda sgsscslart rsrescasvr rassaddiea mragvlpppp
      301 rhastgamhp lrsgllnsts dsdlvryrti skipqitlnf vdlkgdpfla sptsdreiia
      361 pkikerthnv tekvtqvlsl gadvlpeykl qaprihrwti lhyspfkavw dwlilllviy
      421 tavftpysaa fllketeegp patecgyacq plavvdlivd imfivdilin frttyvnane
      481 evvshpgria vhyfkgwfli dmvaaipfdl lifgsgseel igllktarll rlvrvarkld
      541 ryseygaavl fllmctfali ahwlaciwya ignmeqphmd srigwlhnlg dqigkpynss
      601 glggpsikdk yvtalyftfs sltsvgfgnv spntnsekif sicvmligsl myasifgnvs
      661 aiiqrlysgt aryhtqmlrv refirfhqip nplrqrleey fqhawsytng idmnavlkgf
      721 peclqadicl hlnrsllqhc kpfrgatkgc lralamkfkt thappgdtlv hagdlltaly
      781 fisrgsieil rgdvvvailg kndifgepln lyarpgksng dvraltycdl hkihrddlle
      841 vldmypefsd hfwssleitf nlrdtnmipg spgstelegg fsrqrkrkls frrrtdkdte
      901 qpgevsalgp gragagpssr grpggpwges pssgpsspes sedegpgrss splrlvpfss
      961 prppgeppgg eplmedceks sdtcnplsga fsgvsnifsf wgdsrgrqyq elprcpaptp
     1021 sllniplssp grrprgdves rldalqrqln rletrlsadm atvlqllqrq mtlvppaysa
     1081 vttpgpgpts tspllpvspl ptltldslsq vsqfmaceel ppgapelpqe gptrrlslpg
     1141 qlgaltsqpl hrhgsdpgs
//
```

Get amino-acid sequence of hERG channel from UniProt

(*http://www.uniprot.org/*)

UniProt › UniProtKB Downloads · Contact · Documentation/Help

Search in: Protein Knowledgebase (UniProtKB) Query: [ ] Search Clear Fields »

Search | Blast * | Align * | Retrieve | ID Mapping *

Reviewed, UniProtKB/Swiss-Prot **Q12809** (KCNH2_HUMAN)

Last modified July 7, 2009. Version 109. History...

Contribute
Send feedback
Read comments (1) or add your own

Clusters with 100%, 90%, 50% identity | Documents (6) | Third-party data | Customize display TEXT XML RDF/XML GFF FASTA

Names and origin · Protein attributes · General annotation (Comments) · Ontologies · Alternative products · Sequence annotation (Features) · Sequences · References · Web resources · Cross-references · Entry information · Relevant documents

Names and origin

Hide | Top

Protein names	*Recommended name:* **Potassium voltage-gated channel subfamily H member 2** *Alternative name(s):* Voltage-gated potassium channel subunit Kv11.1 Ether-a-go-go-related gene potassium channel 1 Short name=H-ERG Short name=Erg1 Short name=Ether-a-go-go-related protein 1 Short name=Eag-related protein 1 eag homolog
Gene names	Name: **KCNH2** Synonyms: ERG, ERG1, HERG, HERG1
Organism	**Homo sapiens (Human)** [Complete proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

Protein attributes

Hide | Top

Sequence length	1159 AA.
Sequence status	Complete.
Sequence processing	The displayed sequence is not processed.

Molecule processing

Chain	1 – 1159	1159	Potassium voltage-gated channel subfamily H member 2	PRO_0000053999

Regions

Topological domain	1 – 403	403	Cytoplasmic Potential	
Transmembrane	404 – 424	21	Segment S1 Potential	
Transmembrane	451 – 471	21	Segment S2 Potential	
Topological domain	472 – 495	24	Cytoplasmic Potential	
Transmembrane	496 – 516	21	Segment S3 Potential	
Transmembrane	521 – 541	21	Segment S4 Potential	
Topological domain	542 – 547	6	Cytoplasmic Potential	
Transmembrane	548 – 568	21	Segment S5 Potential	
Transmembrane	639 – 659	21	Segment S6 Potential	
Topological domain	660 – 1159	500	Cytoplasmic Potential	
Domain	41 – 70	30	PAS	
Domain	92 – 144	53	PAC	
Nucleotide binding	742 – 842	101	cNMP	
Region	612 – 632	21	Segment H5 (pore-forming) Potential	
Motif	624 – 629	6	Selectivity filter By similarity	
Compositional bias	297 – 300	4	Poly-Pro	

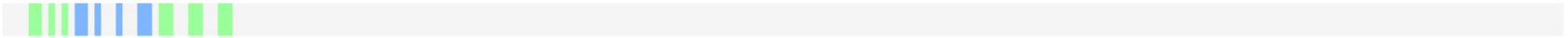
Amino acid modifications

Glycosylation	598	1	N-linked (GlcNAc...) Ref.11	

Natural variations

Alternative sequence	1 – 376	376	MPVRR…EKVTQ → MAAPAGKASRTGALRPRAQK GRVRRAVRISSLVAQE in isoform 2.	VSP_000965
Alternative sequence	139 – 195	57	Missing in isoform 4.	VSP_000966
Alternative sequence	801 – 886	86	KNDIF…SRQRK →	VSP_000967

Secondary structure

1  1159

Helix Strand Turn

Details...

Sequences

Hide | Top

Sequence		Length	Mass (Da)	Tools
Isoform 1 [UniParc].	FASTA	1,159	126,655	Blast go

Last modified November 1, 1996. Version 1.
Checksum: D03BD4F657641FBA

```
        10         20         30         40         50         60
MPVRRGHVAP QNTFLDTIIR KFEGQSRKFI IANARVENCA VIYCNDGFCE LCGYSRAEVM

        70         80         90        100        110        120
QRPCTCDFLH GPRTQRRAAA QIAQALLGAE ERKVEIAFYR KDGSCFLCLV DVVPVKNEDG

       130        140        150        160        170        180
AVIMFILNFE VVMEKDMVGS PAHDTNHRGP PTSWLAPGRA KTFRLKLPAL LALTARESSV

       190        200        210        220        230        240
RSGGAGGAGA PGAVVVDVDL TPAAPSSESL ALDEVTAMDN HVAGLGPAEE RRALVGPGSP

       250        260        270        280        290        300
PRSAPGQLPS PRAHSLNPDA SGSSCSLART RSRESCASVR RASSADDIEA MRAGVLPPPP

       310        320        330        340        350        360
RHASTGAMHP LRSGLLNSTS DSDLVRYRTI SKIPQITLNF VDLKGDPFLA SPTSDREIIA

       370        380        390        400        410        420
PKIKERTHNV TEKVTQVLSL GADVLPEYKL QAPRIHRWTI LHYSPFKAVW DWLILLLVIY

       430        440        450        460        470        480
TAVFTPYSAA FLLKETEEGP PATECGYACQ PLAVVDLIVD IMFIVDILIN FRTTYVNANE

       490        500        510        520        530        540
EVVSHPGRIA VHYFKGWFLI DMVAAIPFDL LIFGSGSEEL IGLLKTARLL RLVRVARKLD

       550        560        570        580        590        600
RYSEYGAAVL FLLMCTFALI AHWLACIWYA IGNMEQPHMD SRIGWLHNLG DQIGKPYNSS

       610        620        630        640        650        660
GLGGPSIKDK YVTALYFTFS SLTSVGFGNV SPNTNSEKIF SICVMLIGSL MYASIFGNVS
```

Homology modeling for a target protein

1) Introduction

2) Search for homolog(s)

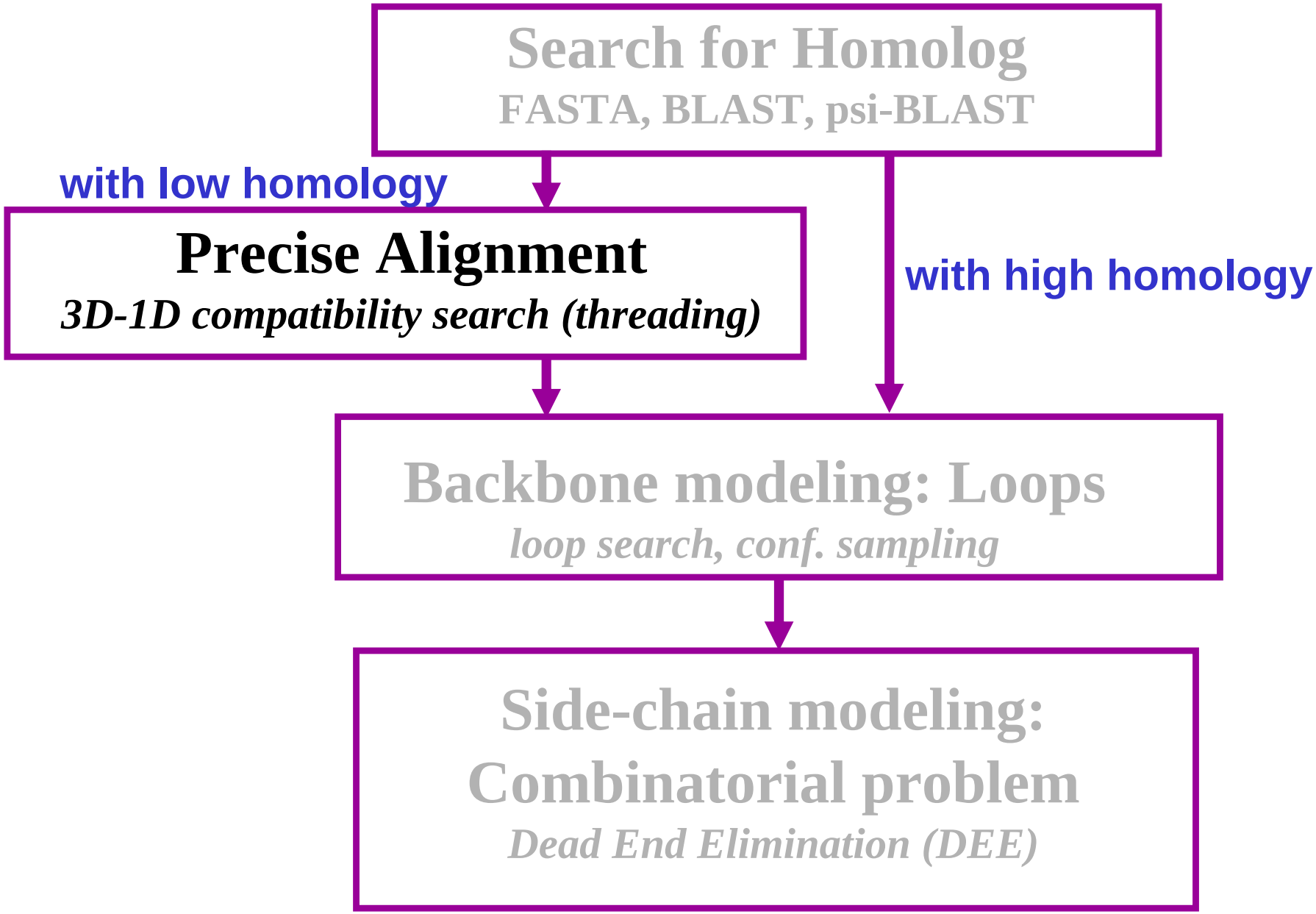
3) Threading (3D-1D compatibility)

4) Backbone modeling

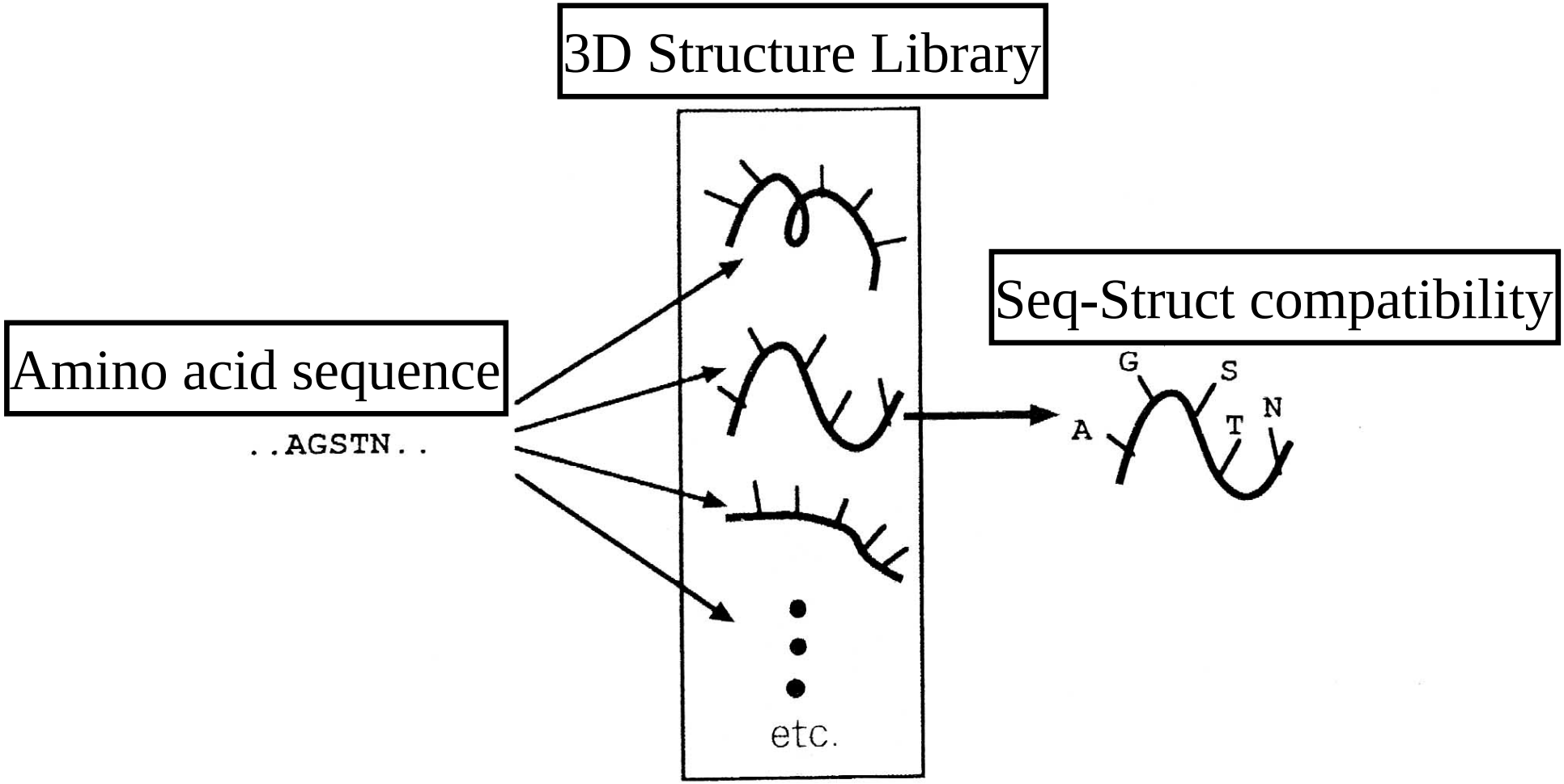
5) Side-chain modeling

6) Structure optimization

Procedure of Homology Modeling



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.

Step2: Get homologs in PDB and have alignments with 3D modes.

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

PDBj

WPI Osaka University iFReC

Sequence to Function Annotation Server

Please enter your query

Name: hERG(550-67

Seqeunce:

```
>hERG | Q12809 | residues 550-671
LFLLMCTFALIAHWLACIWYAIGNMEQPHMDSRIGWLHNLGDQIGKPYNSSGLGGPSIKDKY
VTALYFTFSSLTSVGFGNVSPNTNSEKIFSICVMLIGSLMYASIFGNVSAIIQRLYSGTA
```

Or

Upload a FASTA-formatted sequence file: ファイルを選択 ファイルが選...ていません

Select Alignment methods

☑ Blast	◉ Whole PDB	○ Rep. Domains
☑ PsiBlast	◉ Whole PDB	○ Rep. Domains
☑ HHpred	◉ PDB + SCOP	

Send results to this email address

Input your e-mail address

Submit Clear Form

Send feedback About SFAS

hERG Results

Method	Template		E-value	Coverage	Alignment		Model
HHpred	2r9rB	jV	1.10e-19	75	Text	Jalview	Spanner
HHpred	1orqC	jV	9.70e-20	46	Text	Jalview	Spanner
HHpred	3behA	jV	2.20e-19	75	Text	Jalview	Spanner
HHpred	1xl4A	jV	8.20e-17	47	Text	Jalview	Spanner
HHpred	2a9hA	jV	3.70e-17	49	Text	Jalview	Spanner
HHpred	2ih3C	jV	4.10e-16	48	Text	Jalview	Spanner
HHpred	2q67A	jV	2.40e-16	67	Text	Jalview	Spanner
HHpred	2qksA	jV	5.60e-13	45	Text	Jalview	Spanner
Psiblast	1lnqG	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqH	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqB	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqA	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqF	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqE	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqC	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1lnqD	jV	1.74e-08	40	Text	Jalview	Spanner
Psiblast	1orqC	jV	1.67e-07	42	Text	Jalview	Spanner
Psiblast	2a0lB	jV	4.19e-07	41	Text	Jalview	Spanner
Blast	1ujlA	jV	8.03e-20	33	Text	Jalview	Spanner
Blast	2q6aB	jV	7.93e-04	29	Text	Jalview	Spanner
Blast	2q67B	jV	7.48e-04	29	Text	Jalview	Spanner
Blast	2q67A	jV	7.24e-04	29	Text	Jalview	Spanner
Blast	2ahzB	jV	8.62e-04	29	Text	Jalview	Spanner
Blast	3e89B	jV	1.03e-03	31	Text	Jalview	Spanner
Blast	3e8hB	jV	1.03e-03	31	Text	Jalview	Spanner
Blast	2q6aA	jV	8.13e-04	29	Text	Jalview	Spanner
Blast	2ahyA	jV	8.84e-04	29	Text	Jalview	Spanner
Blast	3e83B	jV	1.03e-03	31	Text	Jalview	Spanner

(*http://sysimm100.protein.osaka-u.ac.jp/tmp/SFAS16483/hERG_top.html*)

Result of SFAS: The best template is 2r9rB

Jalview

Start Jalview

Return to SFAS Results

```
>2r9rB
LGLLIFFLFIGVILFSSAVYFAE---ADER-------------------------
DSQFPSIPDAFWWAVVSMTTVGYGDMVPTTIGGKIVGSLCAIAGVLTIALPVPVI
VSNFNYFYHRET
>hERG
LFLLMCTFALIAHWLACIWYAIGNMEQPHMDSRIGWLHNLGDQIGKPYNSSGLGG
PSIKDKYVTALYFTFSSLTSVGFGNVSPNTNSEKIFSICVMLIGSLMYASIFGNV
SAIIQRLYSGTA
```

/tmp/SFAS12715/2r9rB_hhpred

Java Applet Window

File Edit Select View Format Colour Calculate Help

2r9rB
hERG

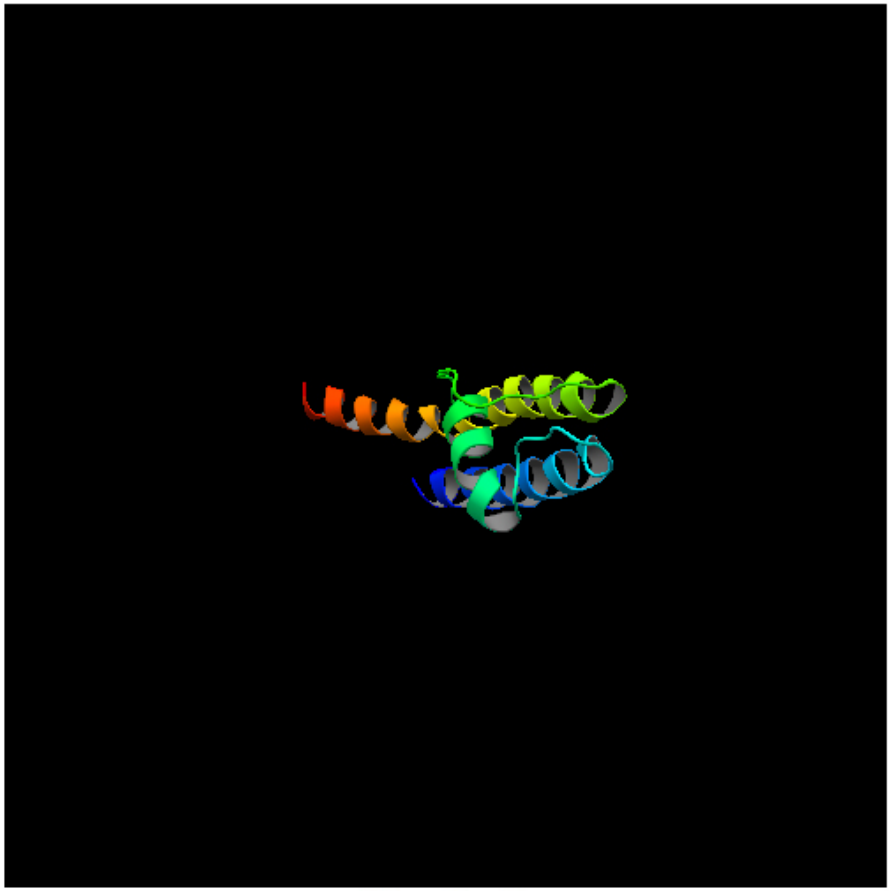
Conservation

Quality

Consensus

Sequence 2 ID: hERG Residue: GLY (99)

2r9rB_hhpred



jV version 3

Homology modeling for a target protein

1) Introduction

2) Search for homolog(s)

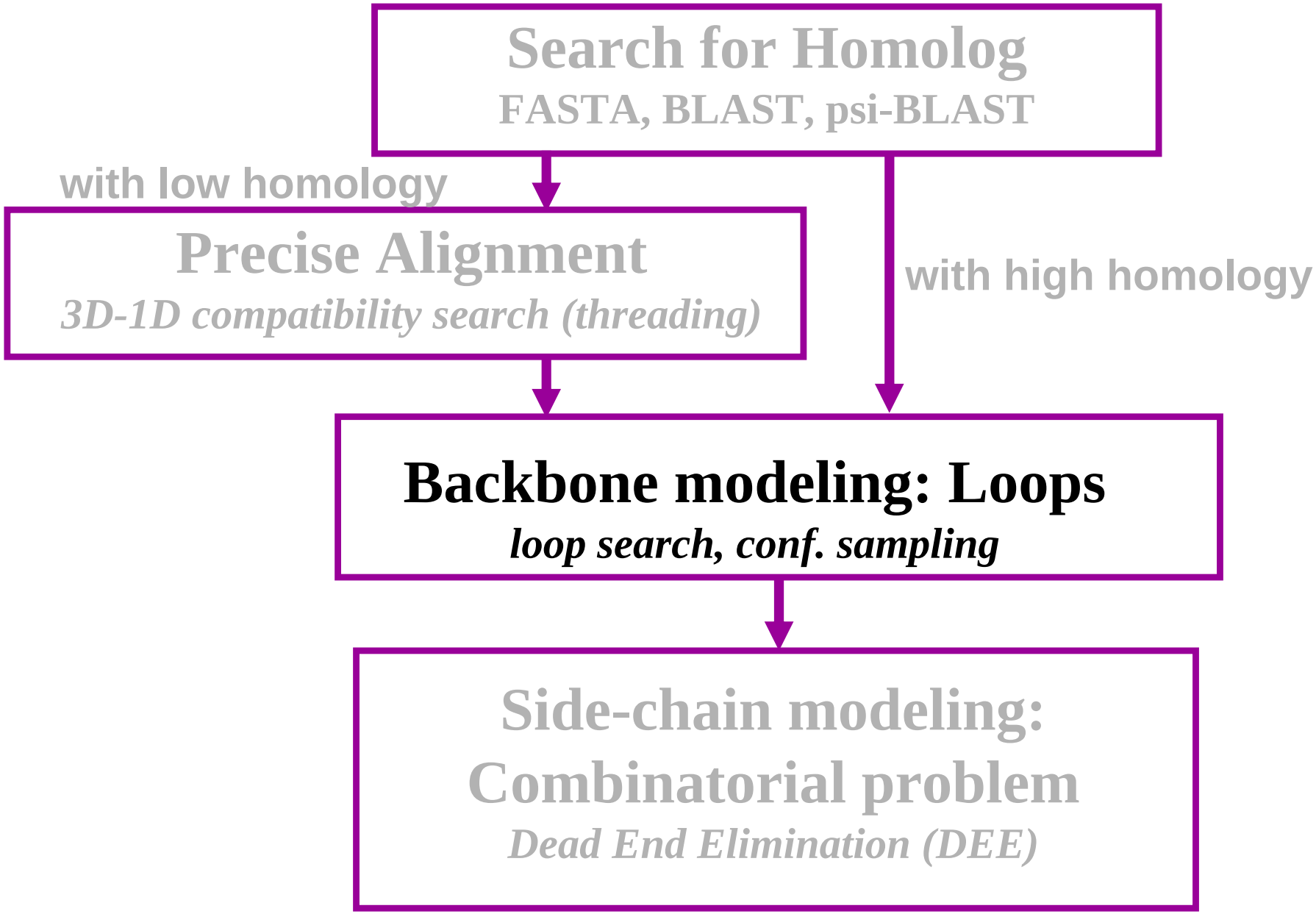
3) Threading (3D-1D compatibility)

4) Backbone modeling

5) Side-chain modeling

6) Structure optimization

Procedure of Homology Modeling

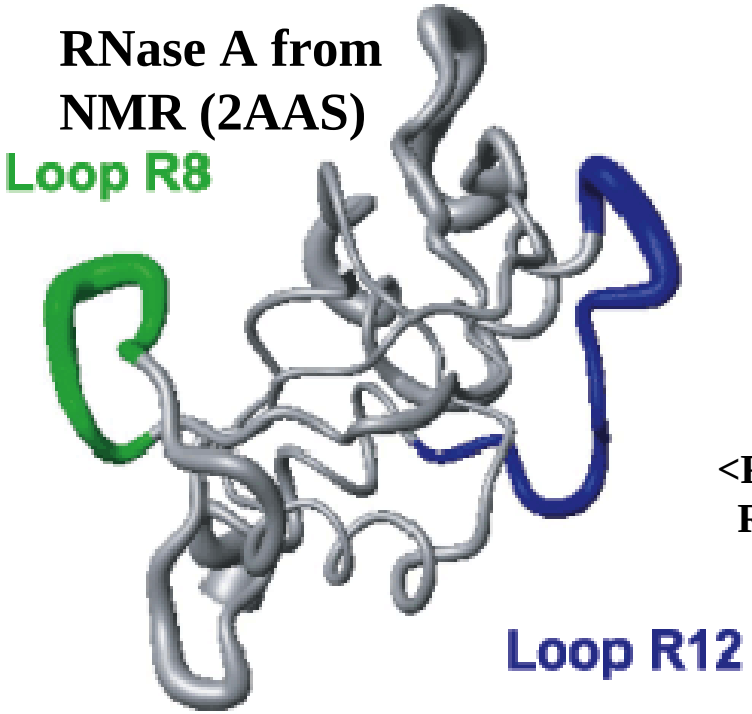


Loop modeling: Modeling for deletion is easy, but for Insertion (in particular, with longer than 7 residues) is difficult.

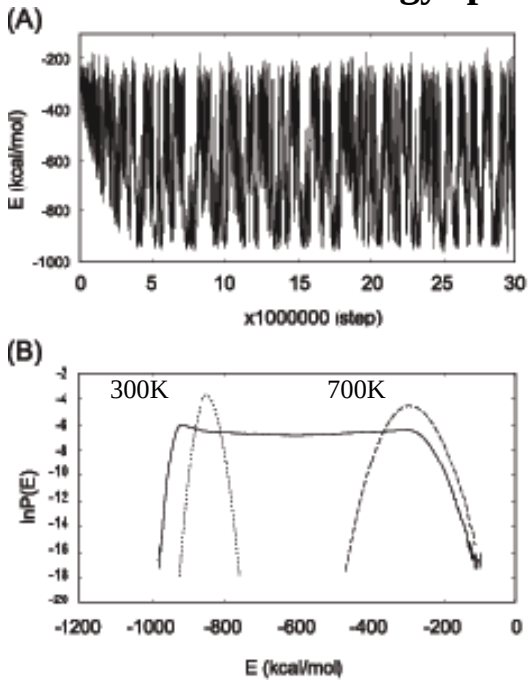
- **Loop Search method: the known loop fragments are used.**
- **Conformational search method: the most stable loop structure is searched from the possible candidates.**

Modeling of a loop structure longer than 10 residues.

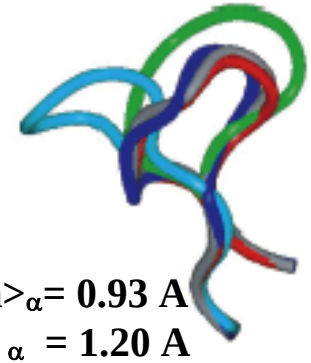
RNase A from NMR (2AAS)



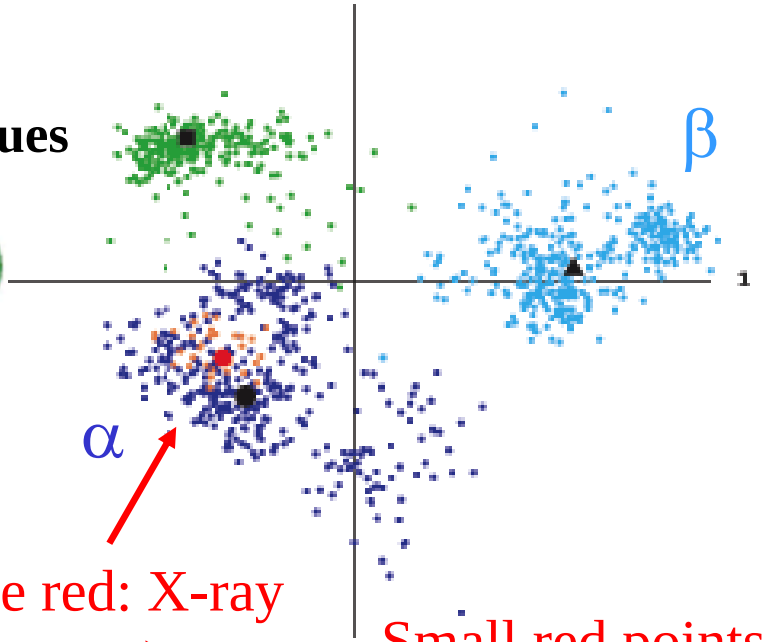
Random-walk in the Energy space



Loop R8 with 8 residues



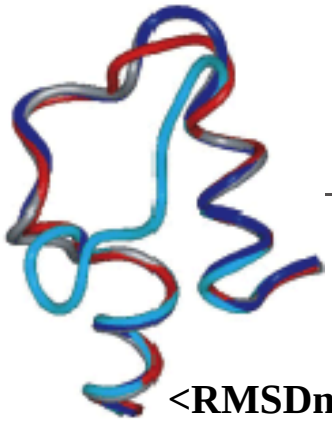
<RMSDn>$_\alpha$= 0.93 A
RMSDx $_\alpha$ = 1.20 A

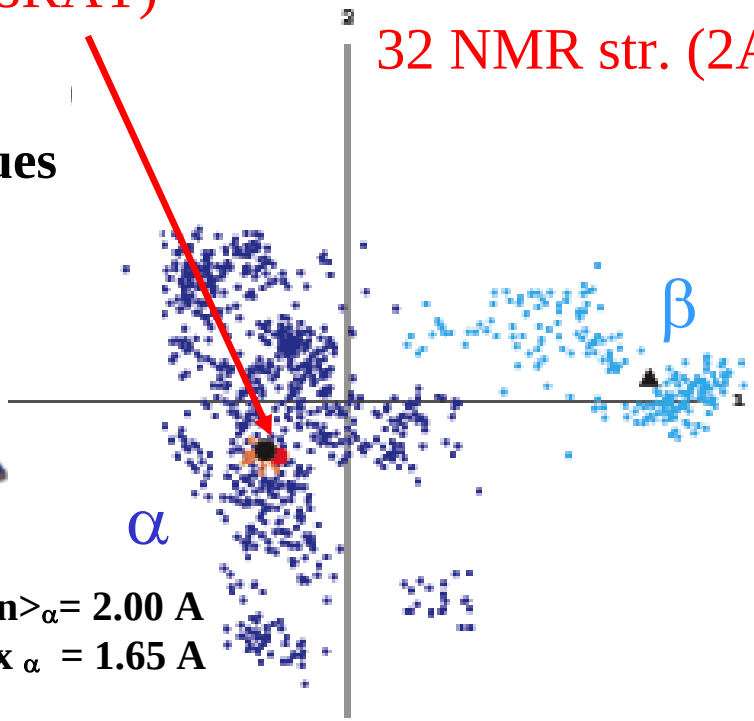


Large red: X-ray str. (8RAT)

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

Loop R12 with 12 residues





<RMSDn>$_\alpha$= 2.00 A
RMSDx $_\alpha$ = 1.65 A

Watanabe, Y. S. et al. (2006) Biophys. 2, 1-12.

Homology modeling for a target protein

1) Introduction

2) Search for homolog(s)

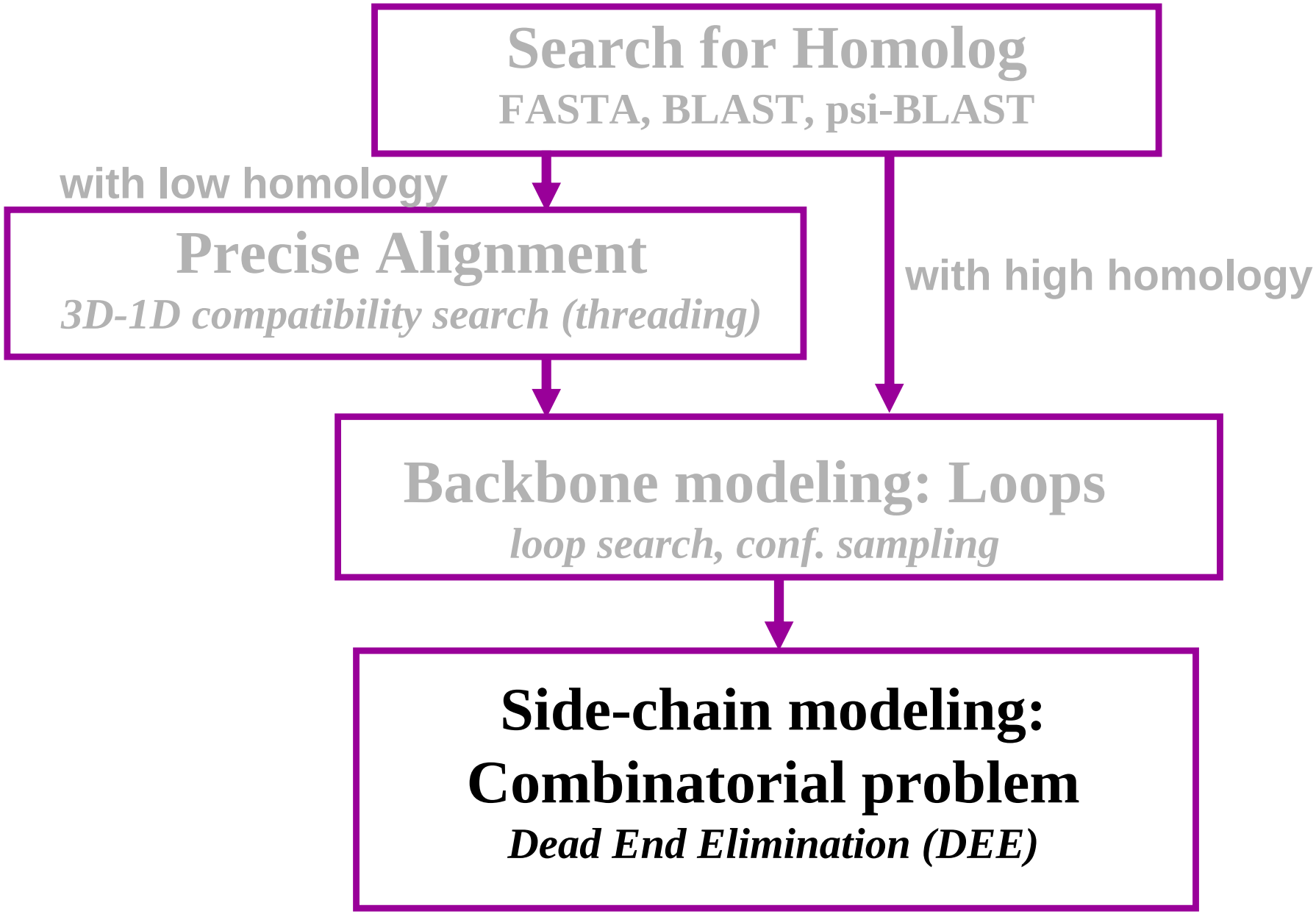
3) Threading (3D-1D compatibility)

4) Backbone modeling

5) Side-chain modeling

6) Structure optimization

Procedure of Homology Modeling



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 3

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

Rotamer	Number	%	Chi 1	Chi 2
Valine				
t	100	67·1	173.5 (9·0)	
−	39	26·2	−63·4 (8·1)	
+	8	5·4	69·3 (9·6)	
Other	2	1·3		
Leucine				
− t	94	63·9	−64·9 (8·2)	176·0 (9·9)
t +	36	24·5	−176·4 (10·2)	63·1 (8·2)
t t	7	4·8	−165·3 (10·0)	168·2 (34·2)
+ +	3	2·0	44·3 (20·0)	60·4 (18·8)
Other	7	4·8		
Isoleucine				
− t	42	45·2	−60·9 (7·5)	168·7 (11·6)
− −	17	18·3	−59·6 (9·6)	−64·1 (14·3)
+ t	15	16·1	61·7 (5·0)	163·8 (16·4)
t t	12	12·9	−166·6 (10·1)	166·0 (8·9)
t +	3	3·2	−174·8 (24·9)	72·1 (10·5)
Other	4	4·3		
Serine				
+	94	48·0	64·7 (16·1)	
−	56	28·6	−69·7 (14·6)	
t	46	23·5	−176·1 (20·2)	
Threonine				
+	81	47·9	62·7 (8·5)	
−	76	45·0	−59·7 (9·4)	
t	8	4·7	−169·5 (6·6)	
Other	4	2·4		
Cysteine				
−	57	60·6	−65·2 (10·1)	
t	23	24·5	−179·6 (9·5)	
+	13	13·8	63·5 (9·6)	
Other	1	1·1		

Ponder & Richards (1987) *J. Mol. Biol.* **193**, 775-791.

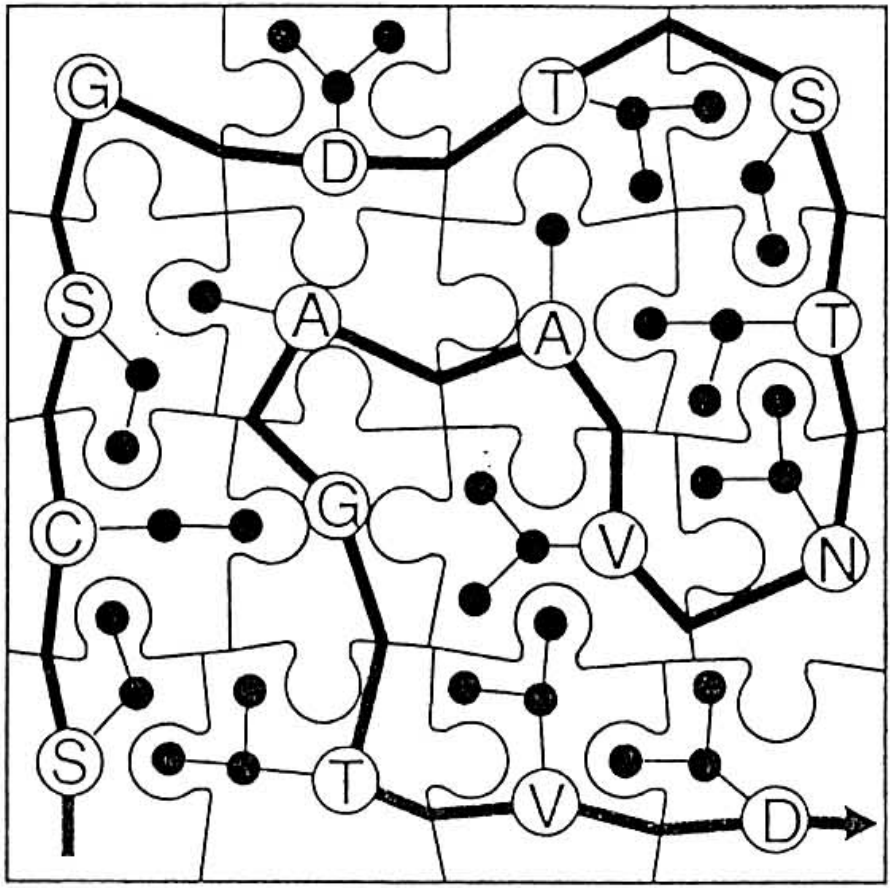
Backbone-dependent rotamer library for proteins

		Φ		ψ		χ-population		
	Number	lower/	upper	lower/	upper	+60°	180°	-60°
V	20	−160	−140	120	140	50	45	5
V	31	−160	−140	140	160	48	10	39
V	12	−160	−140	160	180	17	0	83
V	50	−140	−120	100	120	4	94	2
V	146	−140	−120	120	140	8	86	5
V	99	−140	−120	140	160	12	35	53
V	50	−140	−120	160	180	0	4	96
V	11	−120	−100	−60	−40	0	82	18
V	20	−120	−100	−20	0	5	15	80
V	71	−120	−100	100	120	0	97	3
V	181	−120	−100	120	140	7	88	4
V	49	−120	−100	140	160	14	43	43
V	12	−120	−100	160	180	0	0	100
V	13	−100	−80	−60	−40	8	92	0
V	15	−100	−80	−40	−20	0	53	47
V	13	−100	−80	−20	0	23	15	62
V	43	−100	−80	100	120	7	93	0
V	80	−100	−80	120	140	6	88	6
V	29	−100	−80	140	160	14	41	45
V	207	−80	−60	−60	−40	2	97	0
V	131	−80	−60	−40	−20	11	60	28
V	19	−80	−60	−20	0	32	21	47
V	15	−80	−60	100	120	0	93	7
V	62	−80	−60	120	140	2	94	5
V	27	−80	−60	140	160	0	56	44
V	109	−60	−40	−60	−40	2	92	6
V	39	−60	−40	−40	−20	28	54	18
V	16	−60	−40	120	140	6	75	19

Dunbrack & Karplus (1993) *J. Mol. Biol.* **230**, 543-574.

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.

Taylor, W. (1992) *Nature* **356**, 478-480.

c) Combinatorial approach

Dead-end elimination (DEE) method

Algorithm of dead-end elimination (Desmet et al. Nature, 356, 539-542, 1992)

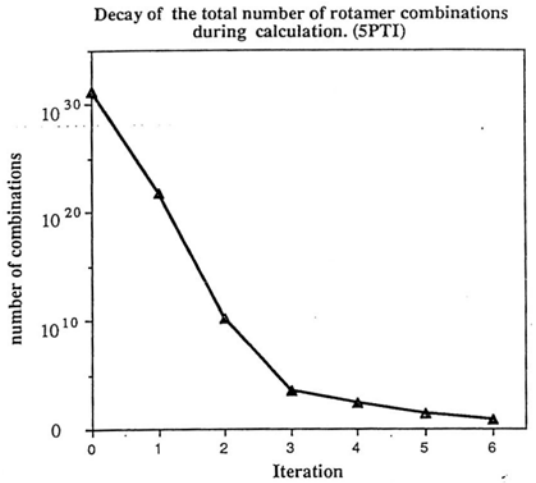
Structural energy for side-chains of N-residues is described by the interaction energy between the backbone and the side-chain, E_1, and the interaction energy between the side-chains, E_2.

$$c\,(r1, r2, \ldots\ldots rN) = \sum_{i} E_1\,(ri) + \sum_{i<j}\sum E_2\,(ri, sj) \qquad \text{—— (1)}$$

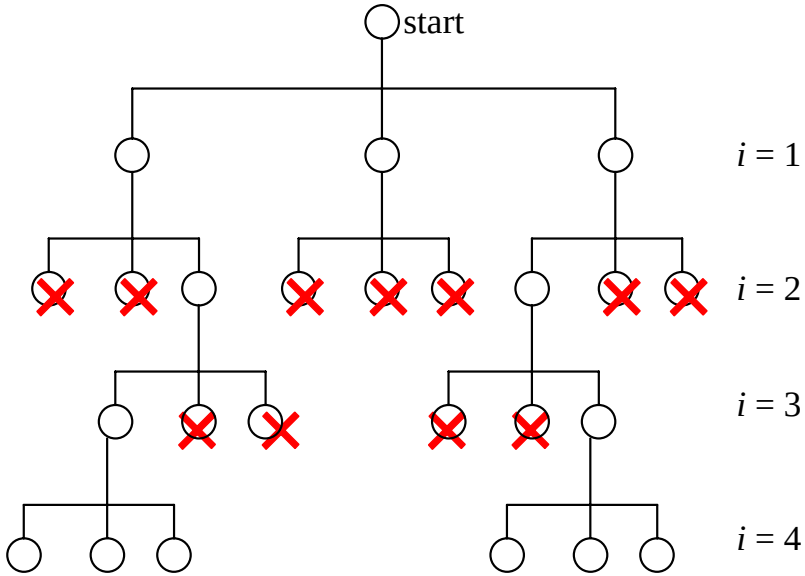
Theorem: When the t'th rotamer (t_i) of the i'th residue is found, satisfying the next equation for the r'th rotamer (r_i) of the i'th residue, then the global energy minimum conformation does not include the r_i.

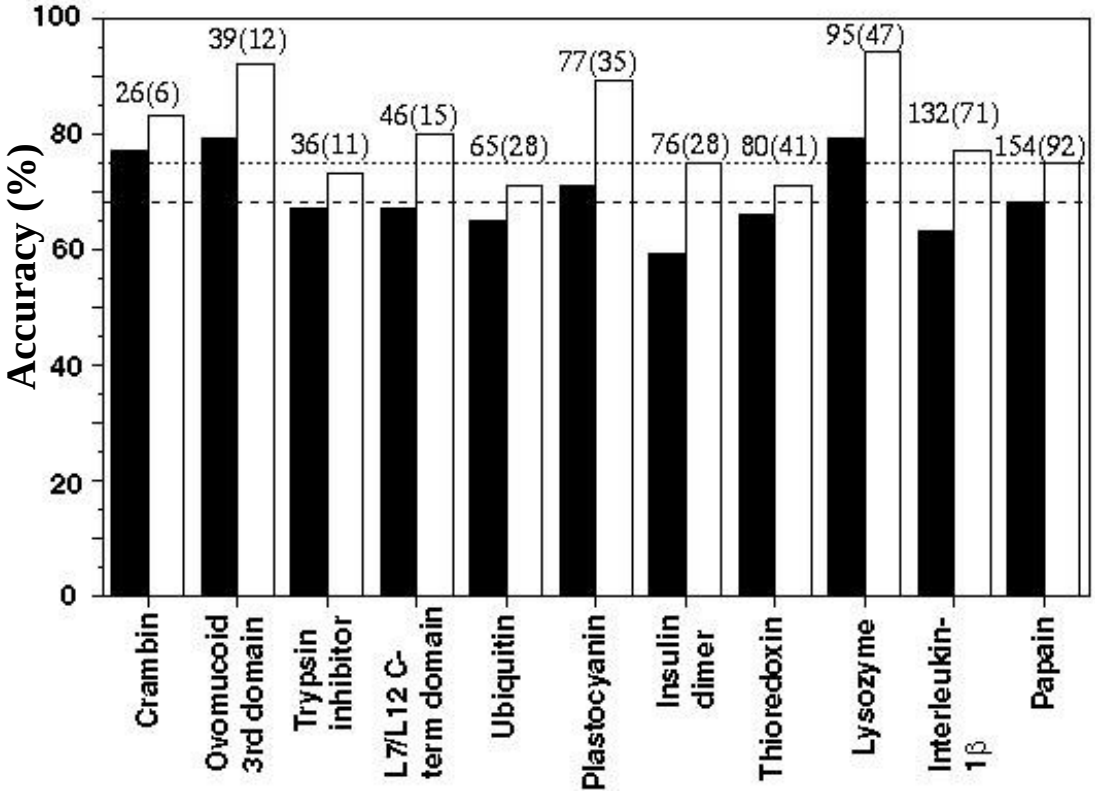
$$E_1\,(ri) + \sum_{i \neq j} \min_{s} \{E_2\,(ri, sj)\} > E_1\,(ti) + \sum_{i \neq j} \max_{s} \{E_2\,(ti, sj)\} \qquad \text{—— (2)}$$

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.

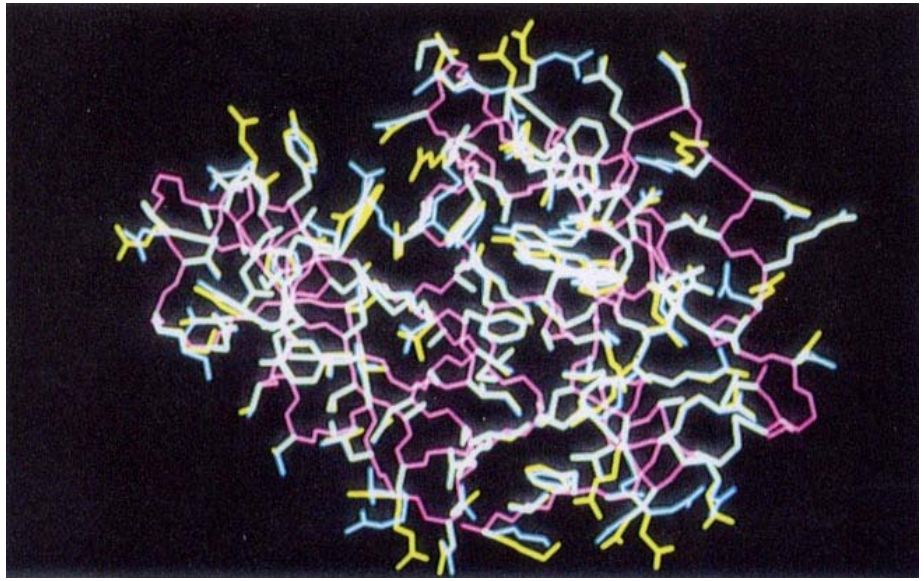


The initial number of rotamer combinations for BPTI (5PTI, 58 a.a.) equals 1.6 x 10^{31}. The number of rotamer combinations was finally reduced to 6 when DEE theorem was applied iteratively.





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.

R. Tanimura et al. Protein Science, 3, 2358-2365, 1994

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

$$U = \sum_{bonds} \frac{1}{2} k_r (r - r_0)^2$$ Bond stretches (1-2)

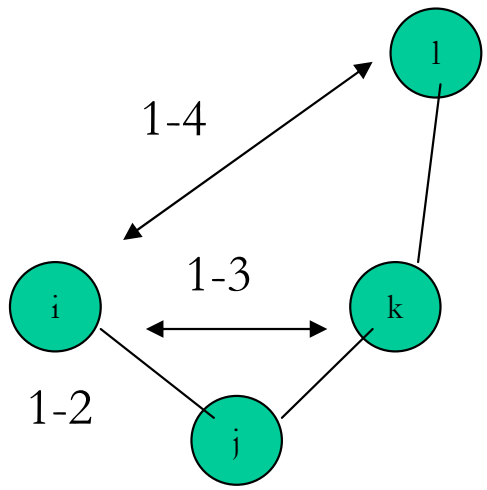
$$+ \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2$$ Angle bending (1-3)

$$+ \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$$ Torsional rotation (1-4)

$$+ \sum_{improper} V(improper \quad torsion)$$ Improper torsion (1-4)

$$+ \sum_{elec} \frac{q_i q_j}{r_{ij}}$$ Electrostatic interaction (1-5)

$$+ \sum_{LJ} [\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}]$$ Lennard-Jones interaction (1-5)



Web site for homology modeling

http://swissmodel.expasy.org/

MENU

Modeling requests:

- First Approach mode
- Alignment Interface
- Project (optimise) mode
- Oligomer modeling
- GPCR mode

Model Database

- SWISS-MODEL Repository, a database for theoretical protein models.

Interactive tools

- SWISS-MODEL Workspace, an interactive working environment for protein structure modelling and

HELP

- Frequently Asked Questions.
- Visualising 3D models.
- Reliability of models.
- How SWISS-MODEL

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-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

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

SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and is now being further developed within the SIB - Swiss Institute of Bioinformatics in collaboration between Torsten Schwede at the Structural Bioinformatics Group, Biozentrum (University of Basel) and Nicolas Guex at GlaxoSmithKline.
The computational resources for the SWISS-MODEL server are provided in collaboration by the Biozentrum (University Basel) and the Advanced Biomedical Computing Center (NCI Frederick, USA).

http://salilab.org/modeller/

To main Sali lab pages

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints

About MODELLER
MODELLER News
Download & Installation
Release Notes
Data file downloads
Registration
Discussion Forum
Subscribe
Browse archives
Search archives
Documentation
FAQ
Tutorial
Online manual
Wiki

About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is available for download for most Unix/Linux systems, Windows, and Mac.

Several graphical interfaces to MODELLER are commercially available from Accelrys. Teaching licenses are also available to those institutions that acquire and maintain a research license.

1. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 200.
2. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
3. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.

Spanner

Spanner is a structural homology modeling program—that is, it threads a specific amino-acid sequence onto a specific PDB structure, patching up the gaps as best it can.

To create a model, you must provide a template structure, as well as an alignment of the sequence you wish to model onto the template sequence. Spanner will replace matching residues, fill any gaps caused by inserted or deleted residues, and thermodynamically optimize the resulting structure.

The resulting PDB, as well as a log file, will be emailed to you when the modeling task finishes. If an error prevented a homology model from being generated (for example, when the alignment you provided does not match the template structure), the log file will explain which part of the modeling sequence failed.

Template PDB structure (PDB format):
ファイルを選択 ファイルが選...ていません

Sequence alignment (FASTA format; first sequence is the template, second sequence is the query):
ファイルを選択 ファイルが選...ていません

Model: (not necessary if PDB file contains only one model)

Chain: (not necessary if PDB file contains only one chain)

Email address for results:

START!

New! **Spanner**

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

Constructed by

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