

タンパク質の構造予測 (おまけ)

清水謙多郎
shimizu@bi.a.u-tokyo.ac.jp

実習の資料について

- 「バイオインフォマティクス 実習」で検索
– <https://lecture.ecc.u-tokyo.ac.jp/~ashimizu/>

SARS-CoV2メインプロテアーゼの構造予測

- ターゲット: SARS-CoV2メインプロテアーゼ(Mpro) → 講義のページ
 - アミノ酸配列は「cov2.fasta」
 - 正解構造: PDB ID: 6lu7 → 構造が分からないと仮定
- テンプレート: SARS-CoVのMpro → PDBから取得
 - PDB ID: 2a5kを使用
- テンプレート(鉄型)の選択、アラインメント、モデル構造の構築は、すべてChimeraを介して、Modellerで実行

ターゲットの配列(cov2.fasta)

```
SGFRKMAFESGKVEGMVQVTCGETTNGLWLDVVYCP
RATCTSEDMMNVNEVLLPQLVSLVWQVNLRLV1
GSHMQNCVLKLKVDTLCKPKYKKLQPFQTSVLAQ
YNGSPSGVYOCAMRPNTI1KGSPFLNGSCSGVGENIDYC
VSPFCYMHMELPTGVHAGTDLEGKFYGPFFVDRQTAAG
TDTTTITLNLAWLVAAYAVINGDRWFLNRTTLLNDNLVA
MKVNEYEPLTQDHDVILGLSQAQGIAVLDMCASLKELLQ
NGMNGNRGTTLGS1LEDETFDVFVQQCSGVTFQ
```

テンプレート(2a5k Aチェイン)の配列

```
SGFRKMAFESGKVEGMVQVTCGETTNGLWLDVVYCP
RATCTSEDMMNVNEVLLPQLVSLVWQVNLRLV1
GSHMQNCVLKLKVDTLCKPKYKKLQPFQTSVLAQ
YNGSPSGVYOCAMRPNTI1KGSPFLNGSCSGVGENIDYC
VSPFCYMHMELPTGVHAGTDLEGKFYGPFFVDRQTAAG
TDTTTITLNLAWLVAAYAVINGDRWFLNRTTLLNDNLVA
MKVNEYEPLTQDHDVILGLSQAQGIAVLDMCASLKELLQ
NGMNGNRGTTLGS1LEDETFDVFVQQCSGVTFQ
```

ターゲットとテンプレートの配列一致度は96.1%

SARS-CoV2 Mproの構造予測

3

テンプレートの選択

- タンパク質SARS CoV2 Mpro(以下、CoV2 Mpro)と配列が似た構造既知のタンパク質をBLASTで検索
- ChimeraでBLASTを利用
- CoV2 Mproのアミノ酸配列「cov2.fasta」を読み込む
 - 「File」→「Open」で、cov2.fastaをオープン
 - 「cov2.fasta」の配列が表示される
 - p2.fastaのウィンドウの「info」→「Blast Protein」で、BLASTを実行
 - BLASTパラメータはデフォルトのまま
 - 検索対象データベースDatabaseは「pdb」
 - BLASTの検索結果が表示される

SARS-CoV2 Mproの構造予測

4

ターゲットとテンプレートとのアラインメント

BLASTの結果のアラインメントをそのまま用いる

2A5K_A(2A5K Aチェイン)を選択

SARS-CoV2 Mproの構造予測

5

モデル構造の構築(1)

- MultAlignViewerのウィンドウの「Structure」→「Modeller (homology)」を実行
- Modellerの起動画面が表示される

SARS-CoV2 Mproの構造予測

6

モデル構造の構築(4)

最初、テンプレート(2a5k)の構造が表示される

SARS-CoV2 Mproの構造予測

7

モデル構造の構築(5)

SARS-CoV2 Mproの構造予測

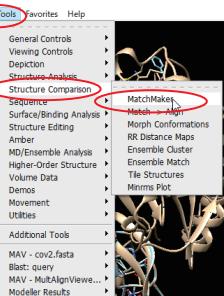
8

結晶構造との重ね合わせ(1)

- 正解の結晶構造(PDB ID:6lu7)との比較
- タンパク質の重ね合わせ計算はChimeraを利用
- メニューの「File」→「Open」でモデル構造のファイル(例えば「cov2-1.pdb」)を開く(モーデリングを行ったままの状態で重ね合わせを実行する場合は不要)
- 「File」→「Fetch by ID」を指定し、その「PDB」に正解構造の「6lu7」を指定
- 「Tools」→「Structure Comparison」→「MatchMaker」を選択

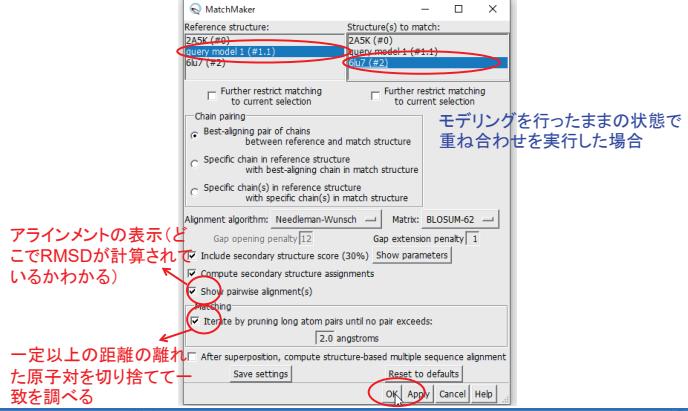
モーデリングを行ったままの状態で重ね合わせを実行した場合

SARS-CoV2 Mproの構造予測



結晶構造との重ね合わせ(2)

- 「Reference structure」として「query model1.pdb (#1.1)」、
「Structure(s) to match」として「6lu7 (#2)」を選択

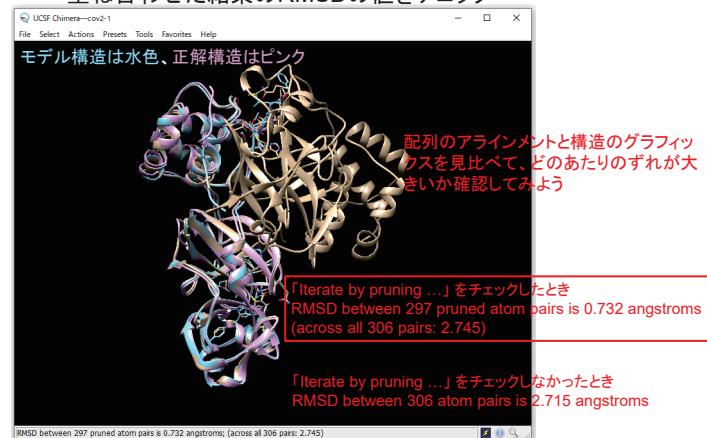


SARS-CoV2 Mproの構造予測

モーデリングを行ったままの状態で重ね合わせを実行した場合

結晶構造との重ね合わせ(3)

- 重ね合わせた結果のRMSDの値をチェック



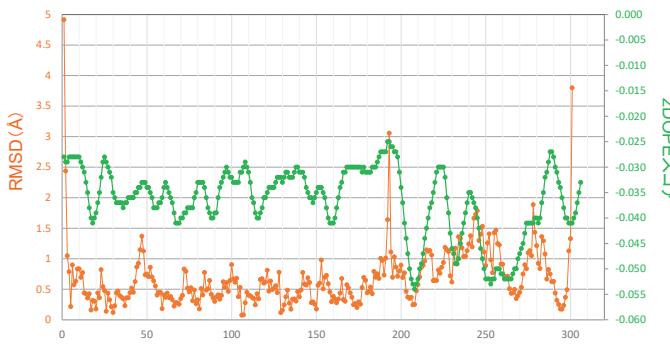
SARS-CoV2 Mproの構造予測

「Iterate by pruning ...」をチェックしなかったとき
RMSD between 306 atom pairs is 2.715 angstroms

参考

結晶構造との重ね合わせ(9)

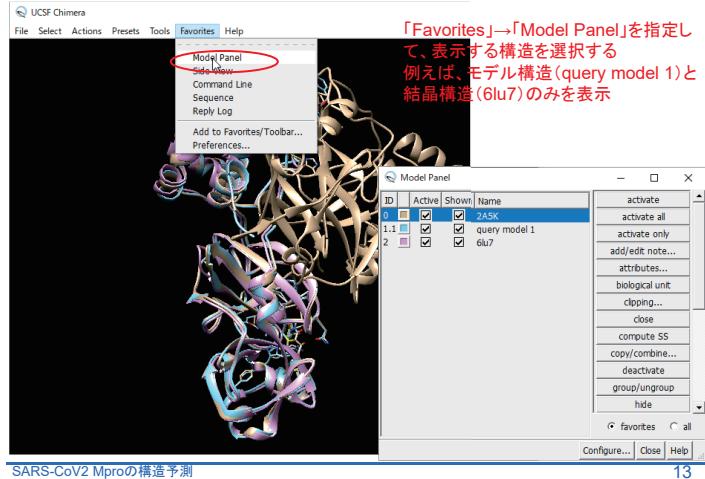
結晶構造と予測構造の差とzDOPEスコアの関係



SARS-CoV2 Mproの構造予測

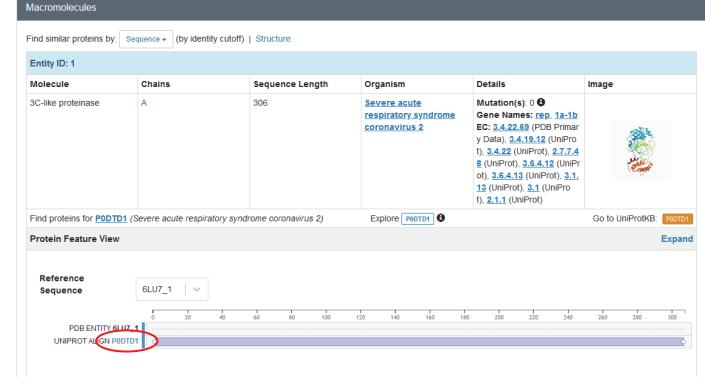
構造バイオ第4回(おまけ)修正0727

結晶構造との重ね合わせ(5)



結晶構造との重ね合わせ(4)

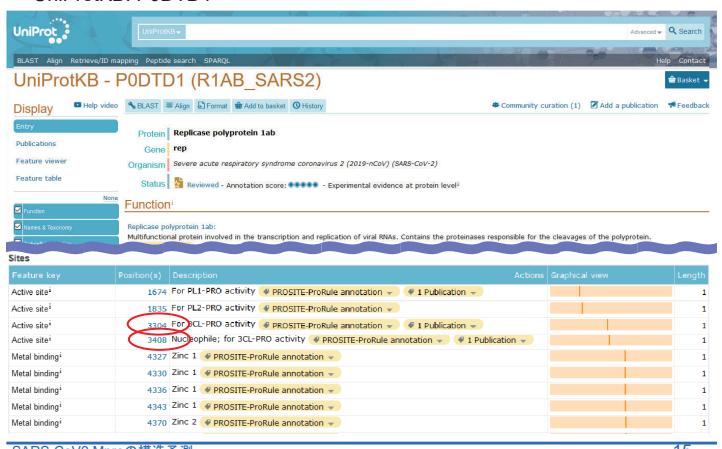
PDB: 6lu7



SARS-CoV2 Mproの構造予測

結晶構造との重ね合わせ(5)

UniProtKB: P0DTD1



SARS-CoV2 Mproの構造予測

結晶構造との重ね合わせ(5)

PROSITEの検索

16

SARS-CoV2 Mproの構造予測

©清水謙多郎

参考 SWISS-MODELによる構造予測(1)

•SARS-CoV2 Mproの構造予測

25

SWISS-MODELによる構造予測(2)

SARS-CoV2 Mproの構造予測

26

SWISS-MODELによる構造予測(3)

•SARS-CoV2 Mproの構造予測

27

SWISS-MODELによる構造予測(4)

•SARS-CoV2 Mproの構造予測

28

SWISS-MODELによる構造予測(5)

•SARS-CoV2 Mproの構造予測

29

SWISS-MODELによる構造予測(6)

•SARS-CoV2 Mproの構造予測

30

SWISS-MODELによる構造予測(7)

•SARS-CoV2 Mproの構造予測

31

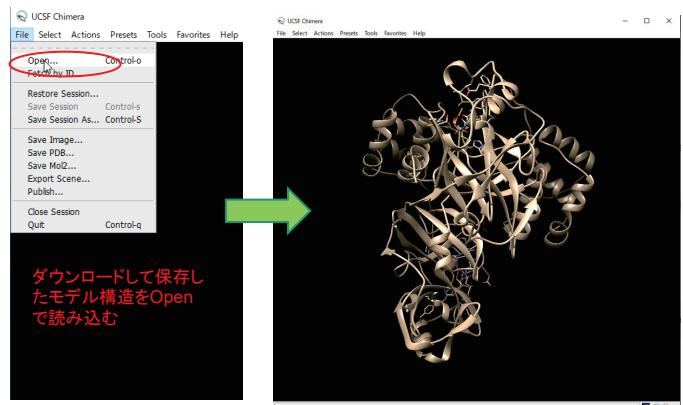
SWISS-MODELによる構造予測(8)

•SARS-CoV2 Mproの構造予測

32

SWISS-MODELによる構造予測(9)

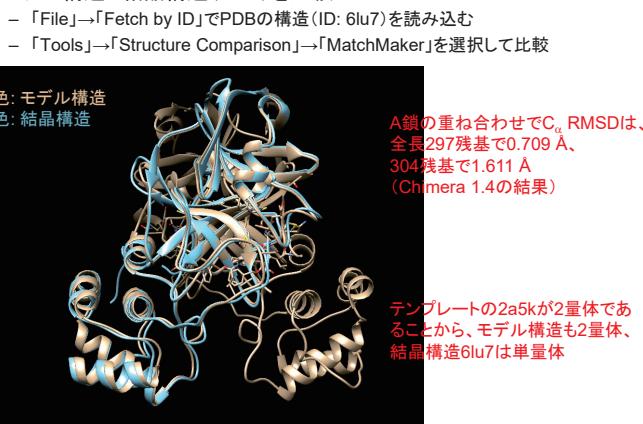
- Chimeraにモデル構造を読み込む



33

SWISS-MODELによる構造予測(10)

- モデル構造と結晶構造(6lu7)を比較



34

SARS CoV2とSARS CoVの比較

```

CoV2_Mpro  SGFRKMAFPSGKVEGCMVQVTCGTTTLLNLWLDDVVYCPVRHVICTEDMLNPNYEDLLIRKSNNHFLVQAGNVQLRVIGH
2a5k_1_A   35 46 65
#SGFRKMAFPSGKVEGCMVQVTCGTTTLLNLWLDDVVYCPVRHVICTEDMLNPNYEDLLIRKSNNHFLVQAGNVQLRVIGH
CoV2_Mpro  SMQNCVLKLKVDTANPKTPKVKFVR1QPGQTFSVLAQYNGSPSGVYQCAMRPNTIKGSFLNGSCSGVGFDYDCVSFC
2a5k_1_A   86 88 94 134
#SMQNCVLKLKVDTANPKTPKVKFVR1QPGQTFSVLAQYNGSPSGVYQCAMRPNTIKGSFLNGSCSGVGFDYDCVSFC
CoV2_Mpro  YMHMELPTGVHAGTDLEGNFYQGPFFVDRQTAAGTDTITITVNLAWLYAAVINGDRNLFNRTFTTLNDNFNLVAMKVNYE
2a5k_1_A   202
#YMHMELPTGVHAGTDLEGKFYQGPFFVDRQTAAGTDTITITVNLAWLYAAVINGDRNLFNRTFTTLNDNFNLVAMKVNYE
CoV2_Mpro  PLTQDHV1LGLGPLSAQPTGIAVLDMCALKEELLQNQMGNGRTILSALLEDEFPPDPDVVRQCSGVTFQ
2a5k_1_A   267 285 286
#PLTQDHV1LGLGPLSAQPTGIAVLDMCALKEELLQNQMGNGRTILSALLEDEFPPDPDVVRQCSGVTFQ

```

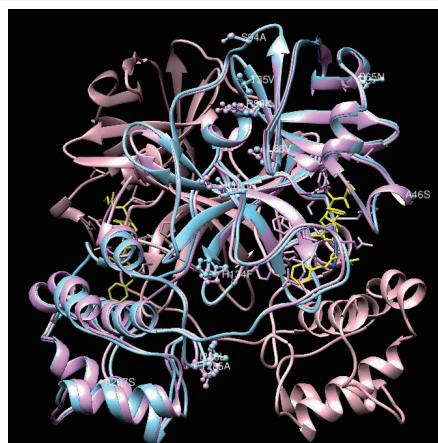
Chimeraの保存イメージ
cov2-cov1.py

水色: CoV2 Mpro (6lu7)
ピンク: CoV Mpro (2a5k)

SARS-CoV2 Mproの構造予測

35

SARS CoV2とSARS CoVの比較



36

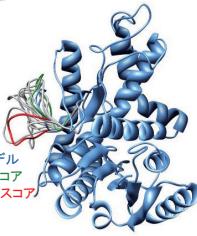
参考

ModLoop

- Saliのグループが開発したループモデリングツール
- 残基位置で指定された「ループ部分」をモデリング
- Modellerのスクリプトで利用可能
- <http://modbase.compbio.ucsf.edu/modloop/>

Sali Lab Home ModWeb ModBase ModEval PCSS FoxS IMP ModPipe
Login Modeler Home Current ModLoop queue Help Contact
ModLoop: Modeling of Loops in Protein Structures
ModLoop is a web server for automated modeling of loops in protein structures. The server relies on the loop modeling routine in MODELLER that predicts the loop conformations by satisfaction of spatial restraints, without relying on a database of known protein structures.
Developer: Andreas Fiser
Acknowledgements: Ben Webb Ursula Pfeifer Andrij Sali Version r169
General information Email address (optional):
Modeler license key:
Upload coordinate file:
Enter loop segments:
Name your model: loop
Process Reset
A. Fiser, N.K.G. Do and A. Sali, Prot Sci (2009) 9, 1753-1773
A. Fiser and A. Sali, Bioinformatics (2003) 19, 2598-2601

Modellerのループモデリング



37

構造予測に関連したツール

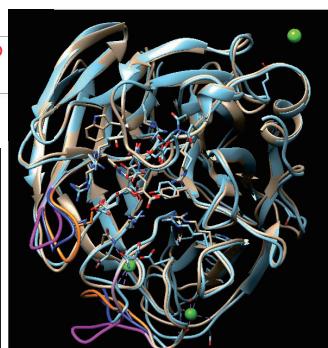
ModLoopの適用例(1)

ModLoop: Modeling of Loops in Protein Structures

ModLoop is a web server for automated modeling of loops in protein structures. The server relies on the loop modeling routine in MODELLER that predicts the loop conformations by satisfaction of spatial restraints, without relying on a database of known protein structures.

ターゲットのモデル構造
shimizu@bioinfo.tohoku.ac.jp
p1-1.pdb
ループモデリングの適用範囲を指定
p1-1
Process Reset

水色: 結晶構造(3b7e Aチェイン)
茶色: もとのモデル構造(p1-1.pdb)
ピンク色: ループモデリング適用後
RMSD between 364 pruned atom pairs is 0.674 angstroms (across all 385 pairs: 0.956)
RMSD between 359 pruned atom pairs is 0.670 angstroms; (across all 385 pairs: 1.105)



38

構造予測に関連したツール

ModLoopの適用例(2)

ループモデリングのためのスクリプト loop_model.py

```

# Run this script with something like
#   python loop.py N > N.log
# where N is an integer from 1 to the number of models.
#
# ModLoop does this for N from 1 to 300 (it runs the tasks in parallel on a
# compute cluster), then returns the single model with the best (lowest)
# value of the Modeller objective function.

from modeller import *
from modeller.automodel import *
import sys

# to get different starting models for each task
taskid = int(sys.argv[1])
env = environ(rand_seed=1000+taskid)

class MyLoop(LoopMode):
    def select_loops(self):
        rns = (
            self.residue_range('163: ', '169: '),
            self.residue_range('251: ', '257: ')
        )
        for rns in rns:
            if len(rns) > 30:
                raise ModellerError("loop too long")
        s = selection(rns)
        if len(s.only_no_topology()) > 0:
            raise ModellerError("some selected residues have no topology")
        return s

    m = MyLoop(env, inmodel='input.pdb',
               sequences='loop')
    m.loop.md_level = refine.slow
    m.loop.starting_model = m.loop.ending_model = taskid
    m.make()

```

構造予測に関連したツール

側鎖モデリングツールSCWRL

- Side Chains With Rotamer Library
- Fox Chase Cancer CenterのDunbrackらによって開発された側鎖モデリングツール
- オープンソースのフリーウェア
- <http://dunbrack.fccc.edu/scwrl4/index.php>

Dunbrack Lab

>>SCWRL4

Home
People
Projects
Publications
Software
Libraries:
Rotamer
Roster Dataset
Ramach
Peptide u/o/p
Backbone
Software:
Scwrl4
BioAssemMod
Pisces
BioDownloader

A baby squirrel from the Fox Chase section of Philadelphia]

Registered Scwrl users: 8771

License for SCWRL4 for non-profit users: Click here

Description
SCWRL Availability
Installation
Usage
Contact

MolIDE: You may be interested in our graphical user interface for homology modeling with SCWRL, that includes PSL-BLAST searches of the PDB, PSIPRED secondary structure predictions, structurally assisted alignment editing, and loop and side-chain modeling. You need to obtain the two programs with separate licenses.

39

構造予測に関連したツール

SARS-CoV2 Mproの構造予測

構造バイオ第4回(おまけ)修正0727

40

©清水謙多郎

参考

HHpredの利用(1)

H1N1ノイラミニーラーp1の配列を入力として、フォールドライブラリを検索。

Input Parameters
Align two sequences/MSAs
Select structural domain databases
PDB mmcif70_29 Now 検索対象のデータベースが選択可能

Submit Job

「Submit Job」をクリック

フォールド認識

参考

HHpredの利用(5)

5n4z_Bをテンプレートとしてモデリング
アライメントはHHpredの結果を利用

Model using selection
「Model using selection」を指定

Visualization
5n4z_Bをテンプレートとして選択
(チェックを入れる)

Submit Section

61

参考

HHpredの利用(2)

「Forward to MODELLER」をクリック

フォールド認識

参考

HHpredの利用(6)

ターゲットとテンプレートの配列が表示される

Forward to MODELLER
「Forward to MODELLER」をクリック

62

フォールド認識

参考

HHpredの利用(3)

Modelerの入力形式のアライメントが生成され、表示される

フォールド認識

参考

HHpredの利用(7)

最初は、Modelerのキーが要求される
(今回は入力済み)

Submit job
「Submit job」をクリック

62

フォールド認識

参考

HHpredの利用(4)

Modelerの入力形式のアライメントが生成され、表示される

フォールド認識

参考

HHpredの利用(8)

「Download PDB file」
PDB形式(座標データ)のファイルのダウンロード

生成されたモデル構造のグラフィックス表示

モデル構造と正解構造3b7e Aチェインとの比較
RMSD between 385 pruned atom pairs is 0.272 angstroms

63

フォールド認識

構造バイオ第4回(おまけ)修正0727

©清水謙多郎

60

64

Phyreの利用(1)

H1N1ノイラミニダーゼp1の配列を入力として、フォールドライブラリを検索

The screenshot shows the Phyre2 search results for the query sequence "c2htuA.p1.fasta". It includes the search parameters, sequence alignment, predicted secondary structures, and a 3D model of the protein.

Phyreの利用(2)

The screenshot shows the Phyre2 search results for the query sequence "c2htuA.p1.fasta". It includes the search parameters, sequence alignment, predicted secondary structures, and a 3D model of the protein.

Phyreの利用(3)

The screenshot shows the Phyre2 search results for the query sequence "c2htuA.p1.fasta". It highlights the 3D model and alignment sections, showing various homologous structures and their alignments.

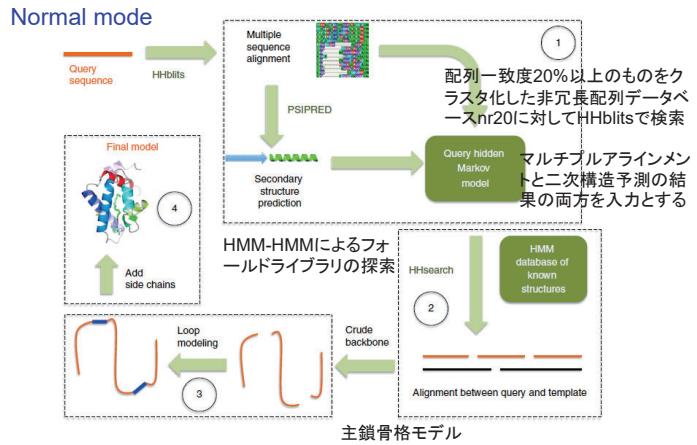
Phyreの利用(4)

The screenshot shows the Phyre2 search results for the query sequence "c2htuA.p1.fasta". It highlights the alignment and conservation sections, showing the alignment of the query sequence with the template sequences.

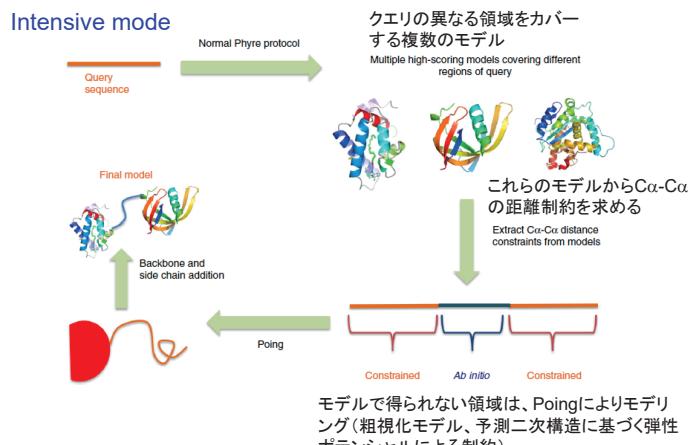
Phyreの利用(5)

The screenshot shows the Phyre2 search results for the query sequence "c2htuA.p1.fasta". It compares the predicted model with the crystal structure and shows the RMSD values between them.

Phyre2の手法(1)



Phyre2の手法(2)



Robettaの利用(1)

The screenshot shows the Robetta search results for the query sequence "c2htuA.p1.fasta". It includes the search parameters, sequence alignment, predicted secondary structures, and a 3D model of the protein.

