作成:2024 年 06 月 10 日
改定:2026 年 03 月 18 日

# YASARA チュートリアル　タンパクモデリング

株式会社アフィニティサイエンス

**概要**:YASARA Structure(Version 25.9.17)による、アミノ酸配列データ(一次構造)からタンパク質立体構造(三次構造)を予測するタンパクモデリングの実行方法をチュートリアル形式で説明します。

※オフライン時に実行する際の注意事項

オフラインで操作を行う場合、予め、BLAST 配列データベース(UniRef90)やテンプレートとして使用する PDB ファイルをダウンロードする必要があります(オンラインの場合、初回実行時に UniRef90 は自動的にダウンロードされます)。また、オフラインで実行する場合、PDB_REDO データベースのテンプレート自動ダウンロードや EBI の AlphaFold データベースの利用ができない等、一部の機能が制限されます。

## 1. はじめに

YASARA でタンパクモデリングを実行するには、マクロを用いて実行する方法と、GUI から直接条件設定を行う方法(Experiment メニューから実行)の 2 つの手段があります。タンパクモデリング計算に関しては、どちらで実行しても設定パラメータや出力ファイルに大きな違いはないので、お好みの方法を選択してください。

YASARA ver.24.4.10 から、従来のホモロジーモデリング手法に加え、AI ベースのフォールディング手法(ESMFold、OmegaFold、AlphaFold)を利用したタンパクモデリングが可能となりました。ESMFold と OmegaFold は構造予測プログラム本体(Python パッケージ)が YASARA に組み込まれており、YASARA を実行中の計算機単独で立体構造の予測が可能です(オフライン解析可)。一方、AlphaFold では、初回使用時に約 2 億の配列情報が含まれる EBI AlphaFold 配列データベース(75 GB)がダウンロードされます。構造予測時には、BLAST 検索後、ターゲット配列にもっとも近い配列の立体構造がダウンロードされ、テンプレートとして採用されます(ターゲット配列情報は外部に送信されませんが、構造データのダウンロードにインターネット環境が必要)。

さらに、YASARA Version 25.9.17 からは AI ベースのフォールディング手法として新たに Boltz-2 の構造予測プログラムが追加され、ESMFold や OmegaFold と同様に、予測構造をタンパクモデリングのテンプレート構造に利用することができるようになりました。

なお、YASARA でタンパクモデリングを実行する場合、構造最適化が自動で行われるため、生成されたモデル構造のリファインなどは特に必要ありません。

### 1.1 本チュートリアルの流れ

本チュートリアルの内容は以下になります。

**全体の流れ**



まず、前半の【2.タンパクモデリングの実行】【3.モデリング結果の確認・可視化方法】では、AI ベースのフォールディング手法（AlphaFold、 ESMFold、OmegaFold、Boltz-2）を利用したタンパクモデリングの基本的な実行方法を紹介します。続いて、後半の【4. 応用：独自のテンプレートやアライメントを使って実行する】では、独自に用意したテンプレートやアライメントを使用したい場合の実行方法を紹介します。こちらはタンパクモデリングに必須の操作ではないので、必要に応じてご参考ください。【6.参考情報】では、ユーザーマニュアルの案内や各パラメータの説明など、補足となる参考情報を記載しています。また、本チュートリアルやタンパクモデリングにおいて用いられる基本用語を【6.5 基本用語】にまとめています。適宜ご参照ください。

## 2. タンパクモデリングの実行

YASARA の GUI から直接条件設定を行う方法（Experiment メニューから実行）【手順 2.1】と、マクロを用いて実行する方法【手順 2.2】をそれぞれ紹介します。

任意になりますが、モデリング精度を向上させたい場合は、UniRef90 シーケンスデータベースの最新バージョン（大容量）を利用することもできます。詳細については、補足の【6.4 HugeUniRef90 データベースのダウンロード】をご参照ください。

### 2.1 Experiment メニューから実行する方法

この項目では、Experiment メニューから実行する方法を紹介します。YASARA の操作画面上で各種パラメータ設定を行ってタンパクモデリングを実行することができます。

必要なファイルは以下になります。（このあとの【手順 2.1.2】でサンプルファイルを作成します。サンプルファイルは本チュートリアルの関連ファイル「YourModel.fasta」をダウンロードして入手することもできます。）

| | |
|---|---|
| 必要なファイル： | ターゲット配列「（任意のファイル名）.fasta」（FASTA 形式） |

#### 2.1.1 作業ディレクトリの作成と設定

入出力データを格納する任意の作業ディレクトリを作成します（Windows ではエクスプローラ、Linux では mkdir コマンド等を使用）。作成場所やファイル名は任意ですが、YASARA 上で表示が乱れるため、日本語を含まない方が好ましいです。

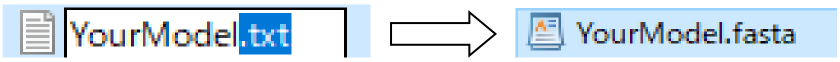
例
(Windows) C:\Users\username\Desktop\YASARA_DATA\tuto_protein-modeling
(Linux)　/home/username/YASARA_DATA/tuto_protein-modeling

操作画面のメニューから「Options」→「Working directory」を選択し、作成したディレクトリを指定し作業ディレクトリを設定します。
(ここで設定したディレクトリはカレントディレクトリとなります。)

### 2.1.2 ターゲット配列ファイルの準備

　モデルを作成したい構造のアミノ酸配列(ターゲット配列)を FASTA 形式で準備し、作業ディレクトリに保存します。
今回は以下の配列情報（PDB:2J24 と配列が一致している）をサンプルとして用います。
以下のテキストをメモ帳などにコピーしてプレーンテキストとして保存し、拡張子を.fasta に変更して使用します。
（RTF、PDF、DOC 形式のファイル等は利用できません。）
ファイル名は任意です。本チュートリアルでは「YourModel.fasta」としました。

```
>YourModel FASTA example
SKPQPIAAANWKCNGSQQSLSELIDLFNSTSINHDVQCVVASTFVHLAMTKERLSHPKFV
IAAQNAIAKSGAFTGEVSLPILKDFGVNWIVLGHSERRAYYGETNEIVADKVAAAVASGF
MVIACIGET
```




※FASTA 形式：　タンパク質のアミノ酸配列や DNA の塩基配列を文字および記号によって記述するファイルの形式。
">" で始まる 1 行のヘッダー行と、2 行目以降の実際のシーケンス文字列で構成される。

### 2.1.3 Experiment 実行

次に、Experiment メニューからタンパクモデリングを実行します。

1. メニューから「Options」→「Choose experiment」→「Protein modeling」を選択します。

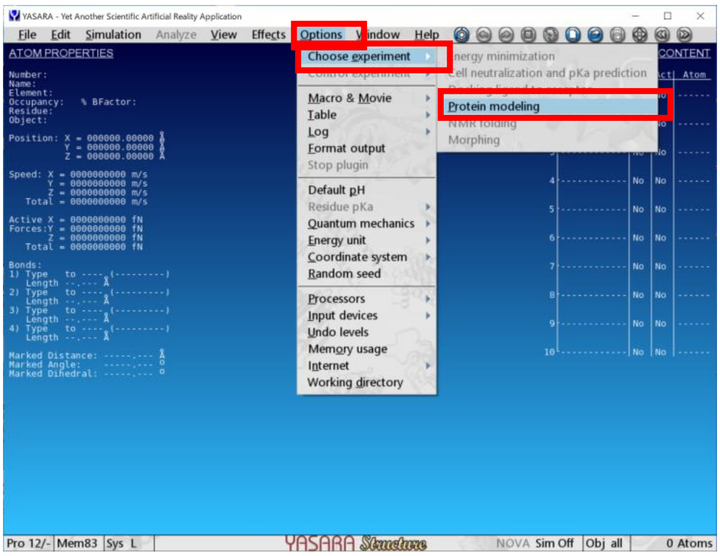



Figure 2.1 Experiment の実行メニュー

2. ダイアログが表示されるので、【手順 2.1.2】で準備した FASTA 形式ファイルを選択して「OK」をクリックします。

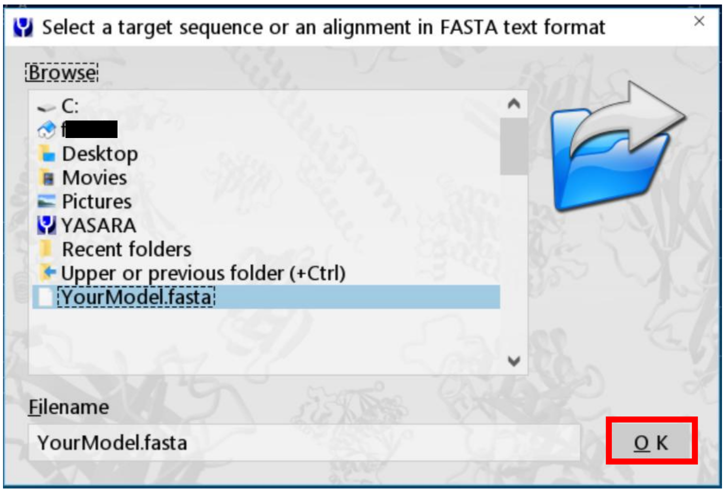



Figure 2.2 ターゲットファイルの選択

3. 次の画面では独自のテンプレートタンパクを使用する場合に指定しますが、今回は使用しないので何も選択せずに「OK」をクリックします。（独自のテンプレートを利用する方法は、【手順 4.2】をご覧ください。）

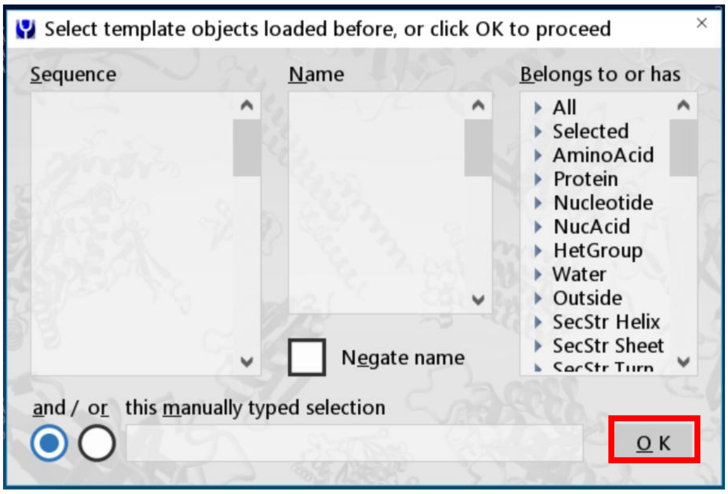



Figure 2.3 テンプレートオブジェクト選択

4. 続いて出力ファイルの保存先と名前を指定します。本チュートリアルでは変更せずに「OK」をクリックして次へ進みます。【手順 2.1.1】で設定した作業ディレクトリが出力ファイル場所に設定され、「YourModel」が出力ファイル名として使われます。（任意で変更することも可能です。出力ファイルの名前（Filename）は fasta ファイル名と異なっていてもよく、出力ファイルの保存先も fasta ファイルの保存先と違う場所を指定できます。）

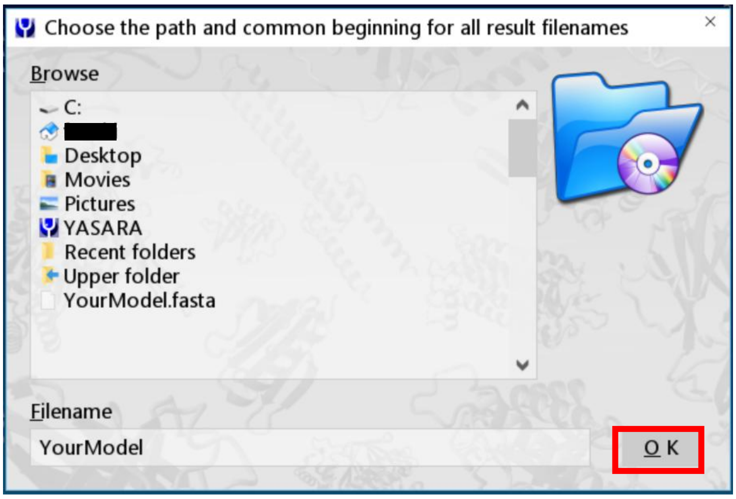



Figure 2.4 出力ファイルの保存先とファイル名の指定

5. 次にパラメータ設定画面が表示されるので各種設定を行います。ここでは、デフォルト設定のまま「OK」ボタンを押して次の設定画面に進みます。各種パラメータの詳細については、補足の【6.2 各種パラメータの概要と…】に記載しています。

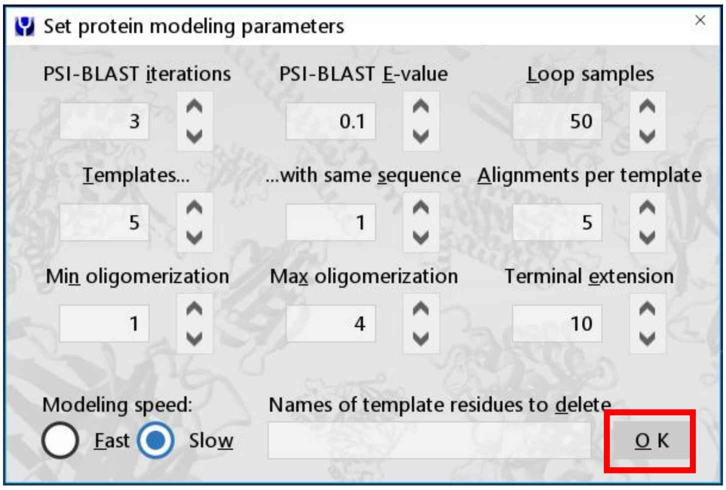



Figure 2.5 パラメータ設定

6. 次に、使用する AI フォールディング手法を選択します。モデリングに使用するテンプレート構造として追加することができます。ここでは、以下のデフォルト設定のまま「OK」を押してタンパクモデリングを実行します。

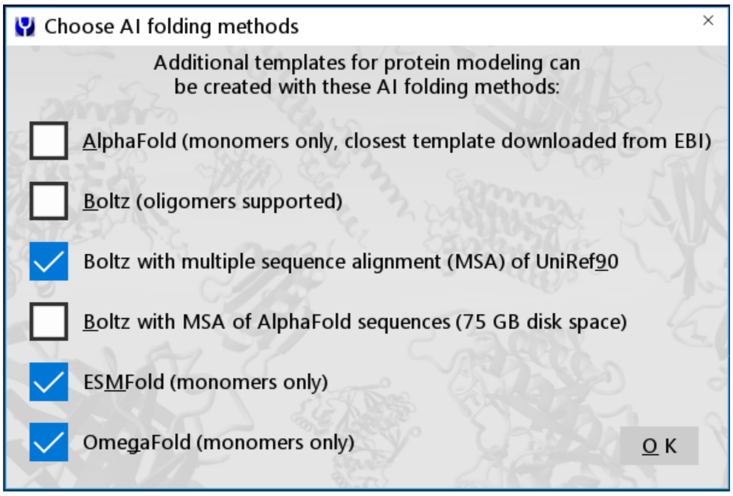



Figure 2.6 利用する AI フォールディング手法の設定

- ❑ AlphaFold
  モノマーのみに対応。入力構造と最も近い構造 EBM データベースから検索・ダウンロードされてテンプレート構造に使用。
- ❑ Boltz
  オリゴマーもサポート。入力としてターゲット配列のみを使用して Boltz-2（YASARA に内包されている）による構造予測を実行し、テンプレート構造に使用。
- ❑ Boltz with multiple sequence alignment (MSA) of UniRef90
  「Boltz」に加え、構造予測時に入力として UniRef90 データベースを用いた MSA を使用。
- ❑ Boltz with MSA of AlphaFold sequences
  「Boltz」に加え、構造予測時に入力として AlphaFold データベースを用いた MSA を使用。
- ❑ ESMFold
  モノマーのみに対応。YASARA に含まれている ESMFold による構造予測を実行し、テンプレート構造に使用。
- ❑ OmegaFold
  モノマーのみに対応。YASARA に含まれている OmegaFold による構造予測を実行し、テンプレート構造に使用。

【注意】

- 「AlphaFold 」または「Boltz with MSA of AlphaFold sequences」を有効にすると、EBI の AlphaFold 配列データベース（約 75 GB）がダウンロードされるため、事前に空き容量を確認してください。
- huge UniRef90 database 及び UniRef90 database のダウンロードを行っていない場合は、初回実行時、自動的に UniRef90 database（約 4.7GB）のダウンロードが行われます。

7. 計算が終了すると、【手順 2.1.2】で指定したディレクトリに各種解析データや HTML 形式の結果レポート（例：YourModel.html）が作成されます。結果レポートの内容は、【手順 3.1】をご覧ください。

### 2.2 マクロを用いて実行する方法

この項目では、YASARA 付属のマクロファイルを用いて実行する方法を紹介します。Experiment から実行する場合と異なり、パラメータ設定は操作画面から行うことはできず、マクロファイルを編集する必要があります。同じ条件で繰り返し実行する場合や、コマンドラインから実行する場合などはこちらの方法が便利です。
必要なファイルは以下になります。

| | |
|---|---|
| 対応マクロファイル： | 「pm_build.mcr」、「hm_build.mcr」または「hm_buildfast.mcr」（下記参照） |
| 必要なファイル： | ターゲット配列「（任意のファイル名）.fasta」（FASTA 形式） |

※（任意のファイル名）の文字列がマクロターゲットとなります。

タンパクモデリングのための付属マクロファイルは3種類あります。

- pm_build.mcr　：標準マクロ。PDB テンプレート、ESMFold による予測構造、OmegaFold による予測構造、Boltz による予測構造、AlphaFold の AI ベースの ab initio 予測（オプション）を利用したモデリング
- hm_build.mcr　：テンプレートベースのホモロジーモデリング
- hm_buildfast.mcr　：簡易的なテンプレートベースのホモロジーモデリング

本チュートリアルでは、標準マクロである「pm_build.mcr」を使用します。他のマクロファイルを使用する場合も、基本的には同じ流れで実行できます。
各マクロファイルの違いの詳細については、補足の【6.2 各種パラメータの概要と…】に記載しています。

#### 2.2.1 作業ディレクトリの作成と設定

手順「2.1.1」と同様の操作を行ってください。（任意）

#### 2.2.2 ターゲット配列ファイルの準備

手順「2.1.2」と同様の操作を行ってください。

#### 2.2.3 マクロファイルの準備

本チュートリアルでは、付属の標準マクロ「pm_build.mcr」を複製、編集し、以下のパラメータを調整したマクロファイルを作成します。なお、マクロ中の各種パラメータの詳細については、補足の【6.2 各種パラメータの概要と…】を参照してください。他のパラメータを変更する場合も同様の操作で準備可能です。
本チュートリアルでは、マクロファイルの編集方法を説明するためにパラメータを変更していますが、通常は付属のマクロファイルをそのまま実行して問題ありません。パラメータ変更方法についての説明が不要な場合は、この項目をスキップして【手順 2.2.4】から再開してください。

・変更内容：「ESMFold+OmegaFold+BoltzMSA90」 → 「ESMFold+OmegaFold+Boltz」
（利用する AI ベース手法を、ESMFold、OmegaFold、BoltzMSA90 から、「BoltzMSA90」→「Boltz （MSA を使用しない）」に変更してみます。）

【マクロファイルのパラメータ変更方法】

1) マクロファイルの複製
インストールフォルダ内に保存されているマクロファイル pm_build.mcr（インストールフォルダ> yasara > mcr > pm_build.mcr）をコピーし、適当なフォルダにペーストします。本チュートリアルでは、【手順 2.2.1】で作成した作業ディレクトリにコピーします。

2) マクロファイルの表示
複製したマクロファイルをテキストエディタで開きます。（Windows であればメモ帳など）
マクロファイルの冒頭部分には概要等が記載されています。その下の「=」行で挟まれている部分がパラメータセクションとなっ

ているので、計算条件などを変更する場合はここの内容を書き換えます。「＃」で始まる行はコメントとなっており、処理は実行されません。

3) 条件設定の変更
本チュートリアルでは例として、「foldmethod」(利用する AI フォールディング手法)のパラメータを変更してみます。マクロファイルの該当箇所を探し、以下のように書き換えます。

(変更前) `foldmethod='ESMFold+OmegaFold+BoltzMSA90'`

↓

(変更後) `foldmethod='ESMFold+OmegaFold+Boltz'`

【注意】「foldmethod」で「AlphaFold 」または「BoltzMSAAF」を有効にする場合、EBI の AlphaFold 配列データベース(約 75 GB)がダウンロードされるため、事前に空き容量を確認してください。
※各種パラメータの詳細については、補足の【6.2 各種パラメータの概要と…】を参照してください。

4) マクロファイルの保存
編集したファイルを上書き保存し、作業を終了します。
(テキストファイル「.txt」として保存してしまった場合は、ファイル名を編集して拡張子を「.mcr」に変更してください。)
任意ですが、編集済マクロファイルは混同しないようにファイル名を変更しておくとよいです。本チュートリアルでは「pm_build_tuto.mcr」としておきます。

### 2.2.4 マクロターゲットの指定

マクロファイルの準備ができたら、マクロを適用するターゲットファイルを指定します。メニューから「Options」→「Macro&Movie」→「Set target」を選択し、【手順 2.2.2】で作成した FASTA ファイル(本チュートリアルでは YourModel.fasta)をクリック後、右側の Remove…「file extension」(拡張子を除く)にチェックを入れて拡張子を除外して指定します。

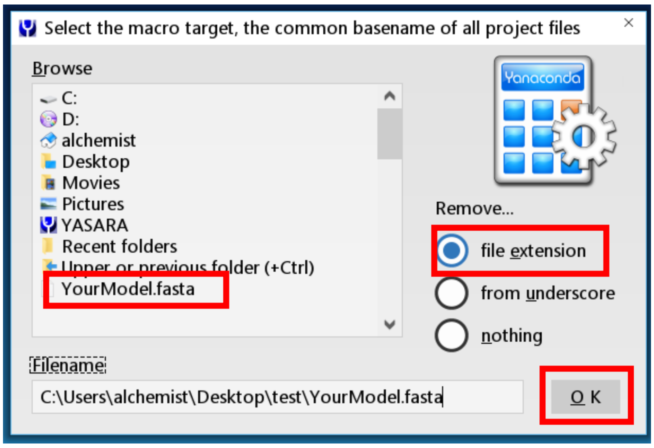



Figure 2.6 マクロターゲット設定

### 2.2.5 マクロ実行

メニューから「Options」→「Macro&Movie」→「Play macro」を選択し、【手順 2.2.3】で準備したマクロファイル(例：作業ディレクトリに保存した pm_build_tuto.mcr)を選択して「OK」をクリックし、タンパクモデリングを実行します。
(デフォルト設定で実行する場合は標準の「pm_build.mcr」、AI ベース手法を利用しない場合は「hm_build.mcr」または「hm_buildfast.mcr」を選択してください。)
(huge UniRef90 database 及び UniRef90 database のダウンロードを行っていない場合は、初回マクロ実行時、自動的に UniRef90 database(約 4.7GB)のダウンロードが行われます。また、AlphaFold または BoltzMSAAF を有効にする場合は、EBI の

AlphaFold 配列データベース(75 GB)がダウンロードされます。)
タンパクモデリングの処理の詳細やリファレンスについては、YASARA ユーザーマニュアル(Help > Show user manual)の、Recipes - Perform complex tasks > Build a protein model > How YASARA builds protein models を参照してください。大まかな流れについては参考情報の【6.3 タンパクモデリングの計算プロセス】にも記載しています。

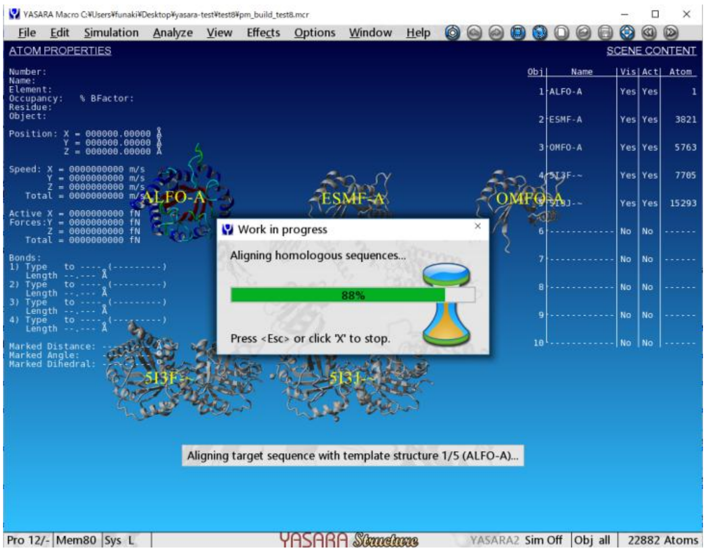



Figure 1.7 タンパクモデリングデリング計算中画面

【Tips】UniRef データベース（UniProt Reference Clusters）とは？

UniProt ナレッジベース（www.uniprot.org/uniprotkb）と一部の UniParc レコード（www.uniprot.org/uniparc）からクラスタリングされたタンパク質の配列セットを提供するデータベースです。これにより、冗長な配列を除外しつつ、複数の解像度（100%、90%、50%の同一性）で配列空間を解析することができます。

UniRef90 は、UniProt データベースの一部で、タンパク質配列をクラスタリングして冗長性を減らし、効率的な類似性検索を可能にするためのものです。具体的には、UniRef90 は、すべてのタンパク質配列をそのまま保持する UniRef100 の配列をクラスタリングし、各クラスタが少なくとも 90%の配列同一性と、クラスタ内の最長配列（シード配列）と 80%の重複を持つように構成されています。

UniRef50 は、UniRef90 のシード配列をクラスタリングし、各クラスタが少なくとも 50%の配列同一性をもち、80%以上の重複を持つように構成されています。

2024 年 9 月現在、UniRef の各 DB のクラスタ数は、100%(412,245,268)、90%(192,983,315)、50%(66,075,574)となっており、全 671,304,157 レコードとなっています（www.uniprot.org/uniref）。

Huge UniRef90 は、UniRef90 の拡張版で、より多くの配列を含むように設計されています。大規模なデータ解析やメタゲノム解析に対応するよう設計されており、環境サンプルやメタゲノムデータセットからの配列を多く取り込み、より広範な生物多様性をカバーしています。標準的な UniRef90 と同様に、少なくとも 90%の配列同一性をもつ配列をクラスタリングしており、より多くのデータを処理するための最適化が施されています。

各データベースの詳細は、次のヘルプページをご確認ください。

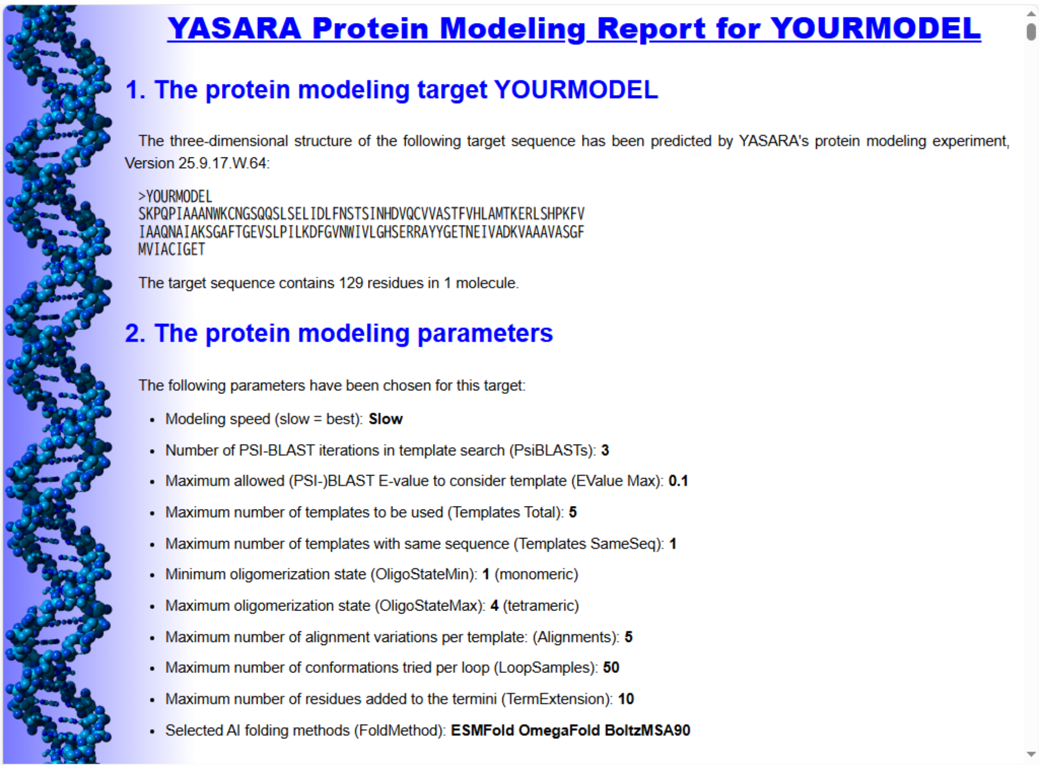
UniProtKB
https://www.uniprot.org/help/uniprotkb
UniParc
https://www.uniprot.org/help/uniparc
UniRef
https://www.uniprot.org/help/uniref

## 3. モデリング結果の確認・可視化方法

計算が終了すると、作成されたモデルが YASARA 上に表示され、作業ディレクトリ内に各モデルの構造ファイル（YOB 形式）やスクリーンショット（PNG 形式）、およびモデリングの結果を記したレポート（HTML 形式）が作成されます。

### 3.1 解析結果（HTML 形式レポート）の見方

作業ディレクトリ内に HTML 形式の結果レポート（YourModel.html）が作成されるので、開いて内容を確認します。
（マクロを用いて実行した場合は、計算終了時に自動的にレポートファイルが開きます。）


YASARA Protein Modeling Report for YOURMODEL

1. The protein modeling target YOURMODEL

The three-dimensional structure of the following target sequence has been predicted by YASARA's protein modeling experiment, Version 25.9.17.W.64:

```
>YOURMODEL
SKPQPIAAANWKCNGSQQSLSELIDLFNSTSINHDVQCVVASTFVHLAMTKERLSHPKFV
IAAQNAIAKSGAFTGEVSLPILKDFGVNWIVLGHSERRAYYGETNEIVADKVAAAVASGF
MVIACIGET
```

The target sequence contains 129 residues in 1 molecule.

2. The protein modeling parameters

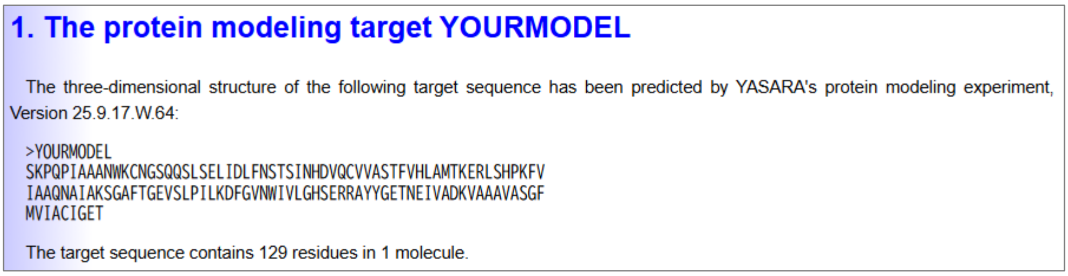
The following parameters have been chosen for this target:

- Modeling speed (slow = best): **Slow**
- Number of PSI-BLAST iterations in template search (PsiBLASTs): **3**
- Maximum allowed (PSI-)BLAST E-value to consider template (EValue Max): **0.1**
- Maximum number of templates to be used (Templates Total): **5**
- Maximum number of templates with same sequence (Templates SameSeq): **1**
- Minimum oligomerization state (OligoStateMin): **1** (monomeric)
- Maximum oligomerization state (OligoStateMax): **4** (tetrameric)
- Maximum number of alignment variations per template: (Alignments): **5**
- Maximum number of conformations tried per loop (LoopSamples): **50**
- Maximum number of residues added to the termini (TermExtension): **10**
- Selected AI folding methods (FoldMethod): **ESMFold OmegaFold BoltzMSA90**

Figure 3.1　モデリングレポート

【手順 2.2 Experiment メニューから実行する方法】で実行した際に作成されたレポートを例に、各項目内容を順に紹介します。
（マクロを用いて実行した場合も、パラメータの一部が異なるだけでレポート内容は同じです。）

1. The protein modeling target（マクロターゲット）


1. The protein modeling target YOURMODEL

The three-dimensional structure of the following target sequence has been predicted by YASARA's protein modeling experiment, Version 25.9.17.W.64:

```
>YOURMODEL
SKPQPIAAANWKCNGSQQSLSELIDLFNSTSINHDVQCVVASTFVHLAMTKERLSHPKFV
IAAQNAIAKSGAFTGEVSLPILKDFGVNWIVLGHSERRAYYGETNEIVADKVAAAVASGF
MVIACIGET
```

The target sequence contains 129 residues in 1 molecule.

Figure 3.2 ターゲット配列情報

計算を行った YASARA のバージョンと、ターゲット配列、その残基数、分子数が記載されています。

2. The protein modeling parameters

## 2. The protein modeling parameters

The following parameters have been chosen for this target:

- Modeling speed (slow = best): **Slow**
- Number of PSI-BLAST iterations in template search (PsiBLASTs): **3**
- Maximum allowed (PSI-)BLAST E-value to consider template (EValue Max): **0.1**
- Maximum number of templates to be used (Templates Total): **5**
- Maximum number of templates with same sequence (Templates SameSeq): **1**
- Minimum oligomerization state (OligoStateMin): **1** (monomeric)
- Maximum oligomerization state (OligoStateMax): **4** (tetrameric)
- Maximum number of alignment variations per template: (Alignments): **5**
- Maximum number of conformations tried per loop (LoopSamples): **50**
- Maximum number of residues added to the termini (TermExtension): **10**
- Selected AI folding methods (FoldMethod): **ESMFold OmegaFold BoltzMSA90**

Figure 3.3 パラメータ情報

計算実行時の各種設定が記載されています。

3. The protein modeling templates

## 3. The protein modeling templates

The following template structures have been generated using AI folding methods:

| Template | Object | Name | Method | Amino acids |
|---|---|---|---|---|
| 1 | 1 | ESMF-A | ESMFold | 129 |
| 2 | 2 | OMFO-A | OmegaFold | 129 |
| 3 | 3 | BM90-A | BoltzMSA90 | 129 |

AI ベース手法を利用して準備されたテンプレート

Since only 3 template structures were provided or AI generated, but 5 were requested, additional possible templates were identified by running 3 PSI-BLAST iterations to extract a position specific scoring matrix (PSSM) from UniRef90, and then searching the PDB for a match (i.e. hits with an E-value below the homology modeling cutoff 0.1).

The following 173 hits were found:

PDB の検索結果一覧（スコア順）

| Template | Total score | BLAST E-value | Align score | Cover | ID | Resolution | Header |
|---|---|---|---|---|---|---|---|
| 4 | 596.39 | 7e-033 | 659.0 | 100% | 5I3F-D | 1.72 A | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (249 residues with quality score 0.905), released 2016-05-18 |
| 5 | 595.08 | 8e-033 | 659.0 | 100% | 5I3J-B | 1.80 A | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (249 residues with quality score 0.903), released 2016-05-18 |
| - | 594.42 | 8e-033 | 659.0 | 100% | 5I3G-B | - | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (249 residues with quality score 0.902), released 2016-05-18. **NOTE: This template was** |

Figure 3.4 テンプレート情報

タンパクモデリング計算のテンプレートとし得るテンプレートの情報が記載されています。
AI ベース手法を用いて用意されたテンプレートに加えて、UniRef90 に対して PSI-BLAST を行って作成された配列プロファイルを用いた PDB の検索結果の一覧が、「Total score」の高い順に表示されます。「Total score」は、BLAST のアライメントスコア「Align score」、PDBFinder2 データベースにおける WHAT_CHECK のクオリティスコア（0.000～1.000）、ターゲットの網羅範囲「Cover」の単純な積で表されます。テンプレートの選択においては基本的に total score を基準とし、その際、最高スコアの 30％を切るテンプレートは除外されます。

本チュートリアルでは、テンプレート数を「5」に指定していたため、3 種の AI ベース手法により生成された 3 つのテンプレートに加え、残りの 2 つ分が PDB 構造から採用されています。

4. The secondary structure prediction

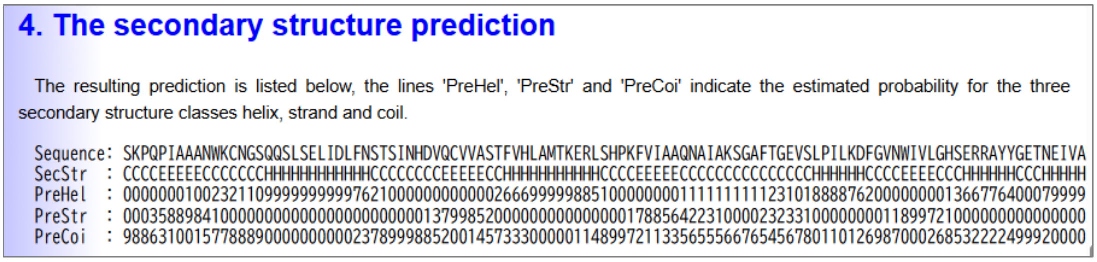



Figure 3.5 二次構造予測

アライメントの補正とループモデリングを支援するために行われた、ターゲット配列の二次構造予測の結果が記載されています。一段目のターゲット配列（Sequence）の各残基に対して、二段目（SecStr）にヘリックス構造（H）、ストランド構造（S）、コイル構造（C）いずれかの構造が示されます。三～五段目には、ヘリックス構造（PreHel）、ストランド構造（PreStr）、コイル構造（PreCoi）となる可能性が 0～9 のスコアで示され、各残基ごとのスコア合計が9になるよう評価されます。

5. The target sequence profile

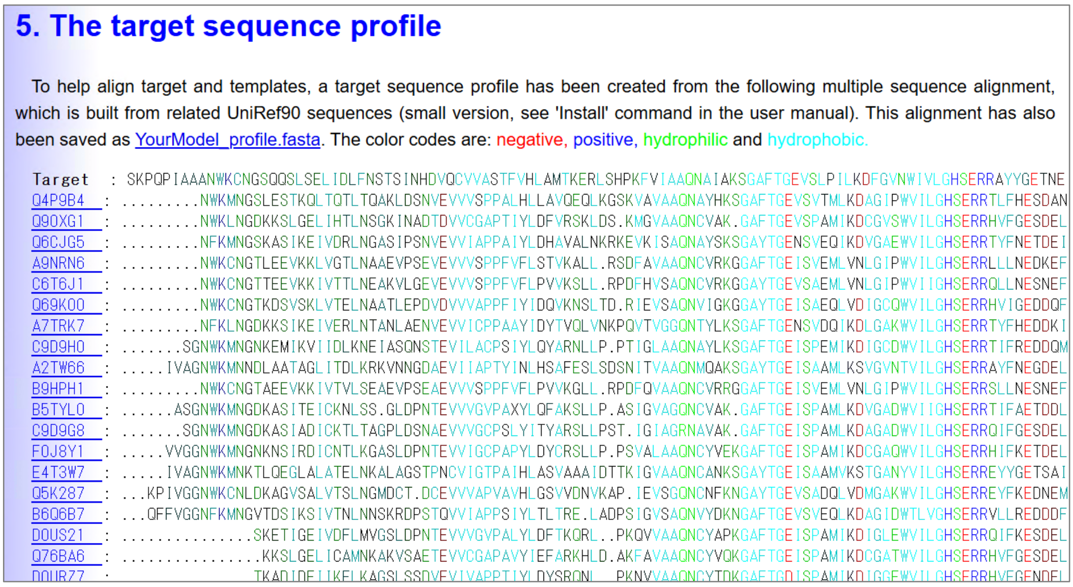



Figure 3.6 マルチプルアライメントの結果

ターゲット配列とマルチプルアライメントの結果の一部が一覧化されます。配列データは UniRef90 から得られており、青字をクリックしてリンクページに移動することもできます。UniRef90 の検索ヒット順に表示（間隔をあけて順にピックアップ）され、負電荷を帯びたアミノ酸は赤、正電荷を帯びたアミノ酸は青、親水性のアミノ酸は緑、疎水性のアミノ酸は水色の文字で表されています。この.html 形式の表示では、スペースの都合上、表示される配列の数は省略されていますが、全ての情報は作業ディレクトリに「. . . _profile.fasta」として保存されています。このマルチプルアライメントから、プロファイルが作成されています。

6. The initial protein models

各テンプレートから予測されるモデルの詳細がそれぞれ小項目に分けられて示されます。各テンプレートから、アライメントが確実な場合は単一のモデルが、あいまいな場合は複数のモデルが作成されます。

- プロファイル-プロファイルアライメントの結果

  ターゲット配列とテンプレートの 2 つのプロファイルをアライメントすることで、ターゲット配列とテンプレートの二次構造の相同性と、テンプレートに含まれる構造に関する情報が考慮されます。

  真ん中の「Match」という行にターゲット-テンプレートアライメントの結果が記載されています。

  なおリストの下方には、テンプレート構造を基にモデルを作成した際の処理について記載されています。

Figure 3.7 プロファイル-プロファイルアライメントの結果

リストの内容は上から順に以下のようになっています。

| | |
|---|---|
| UniPlotID（10 個） | ターゲット配列を基に得られたマルチプルアライメントの結果の一部が表示される(配列は UniRef90 より検索される)。「(MacroTarget)_profile.fasta」ファイルの上から 10 個分が、上下逆の並びになって記載されている。 |
| SeqStr | 推測されたターゲット配列の二次構造 |
| Target | ターゲット配列 |
| Match | ターゲット配列—テンプレート配列間のアライメント結果 |
| Template | テンプレート構造の配列 |
| SeqStr | テンプレート配列の二次構造 |
| PDB ID（20 個） | テンプレートプロファイルの一部が表示される。全ての結果は「(MacroTarget)_(PDB ID)-~_profile.ali.」ファイルに保存されている。 |

- エネルギー最小化と Z-score

  続いて、エネルギー最小化を行った際の Z-score の情報が記載されています。

  エネルギー最小化の前半は残基のバックボーンを固定したまま行われ（構造は～_refined50.yob として保存)、後半は残基を固定せずに行われます（構造は～_refined100.yob として保存)。レポートには、それぞれ Z-score が記載された表が出力され、総合（Overall）の Z-score が良い方がモデルに採用されます。

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.435 | Optimal |
| Packing 1D | -0.463 | Good |
| Packing 3D | -0.162 | Good |
| Overall | -0.048 | Good |

| Z-score | Description |
|---|---|
| ＜-5 | disgusting |
| ＜-4 | terrible |
| ＜-3 | bad |
| ＜-2 | poor |
| ＜-1 | satisfactory |
| ＜0 | good |
| ＞0 | optional |

Figure 3.8 Z-score の出力例(左)と評価基準(右)

- 最終的なモデルの構造と Z-score プロット

  作成されたモデルの構造は、作業ディレクトリに（マクロターゲット）_（モデル ID）.yob として保存されます。

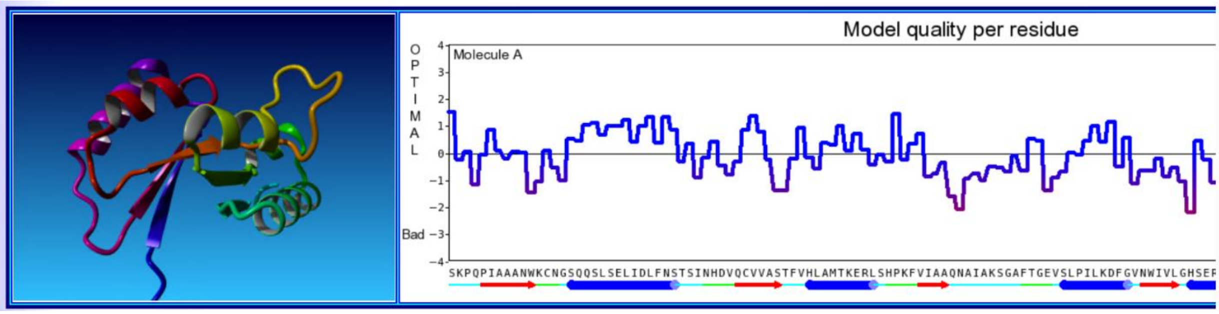



Figure 3.9　作成されたモデル構造（左）と残基ごとの Z-score（右）

7. The model ranking

　作成されたモデルが Z-score によってランク付けされた一覧表が記載されている。各構造は、作業ディレクトリに（共通のファイル名）.sce としても保存されています。

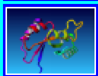
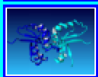
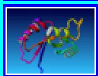
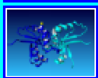
| Rank | Z-score | Structure | State | Model ID | Filename | Original number | Residues | Comment |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.033 |  | monomer | OMFO-A | YourModel_omfo-a.yob | 2 | 1-129 | Optimal |
| 2 | 0.026 |  | homodimer | 5I3F-~ | YourModel_5i3f-~.yob | 4 | 1-129 | Optimal |
| 3 | 0.023 |  | monomer | ESMF-A | YourModel_esmf-a.yob | 1 | 1-129 | Optimal |
| 4 | -0.036 |  | homodimer | 5I3J-~ | YourModel_5i3j-~.yob | 5 | 1-129 | Good |
| 5 | -0.049 |  | monomer | BM90-A | YourModel_bm90-a.yob | 3 | 1-129 | Good |

Figure 3.10　作成されたモデルのランキング（Z-score 順）

8. The hybrid model

　作成されたモデルの最適な部分を組み合わせて形成されたハイブリッドモデルの結果が表示されます。

1つ目のリストでは、ハイブリッドモデルの構築に使用されたフラグメントの一覧が記載されています。

| Transfer | First residue | Last residue | Length | From model | Score |
|---|---|---|---|---|---|
| 1 | 1 | 129 | 129 | OMFO-A | **-1.499** |
| 2 | same | same | same | same | **-1.231** oligomerized |
| 3 | 74 | 88 | 15 | 5I3J-~ | -1.230 rejected |
| 4 | 71 | 83 | 13 | 5I3F-~ | **-1.226** accepted |
| 5 | 69 | 86 | 18 | 5I3J-~ | **-1.213** accepted |
| 6 | 71 | 76 | 6 | 5I3J-~ | **-1.197** accepted |
| 7 | 72 | 75 | 4 | 5I3J-~ | -1.214 rejected |
| 8 | 93 | 96 | 4 | 5I3F-~ | -1.206 rejected |

Figure 3.11　ハイブリッドモデルのフラグメント一覧

続いて、ハイブリッドモデルの Z-score と 3D 構造、Z-score プロットが出力されます。

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.250 | Optimal |
| Packing 1D | -0.028 | Good |
| Packing 3D | -1.007 | Satisfactory |
| Overall | -0.298 | Good |

Figure 3.12　ハイブリッドモデルの Z-score の出力例

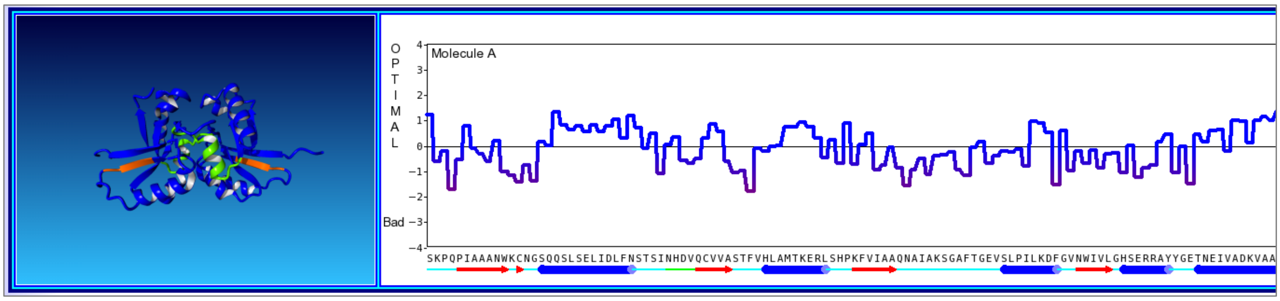



Figure 3.13 ハイブリッドモデルの構造（左）と残基ごとの Z-score（右）

作成されたハイブリッドモデルの構造は、（共通のファイル名.yob）として作業ディレクトリに保存されます。それぞれの断片は異なる色になるようにカラーリングされています。

### 3.2 モデル構造の確認（残基毎 Z スコアの可視化）

計算が終了すると、以下のモデル構造ファイルが作業ディレクトリに保存されます。

- （マクロターゲット）.sce　：　各テンプレート構造から作成されたモデル構造
- （マクロターゲット）.yob　：　ハイブリッドモデルの構造

試しに、（マクロターゲット）.sce　ファイル（本チュートリアルでは YourModel.sce）を開いて見てみます。

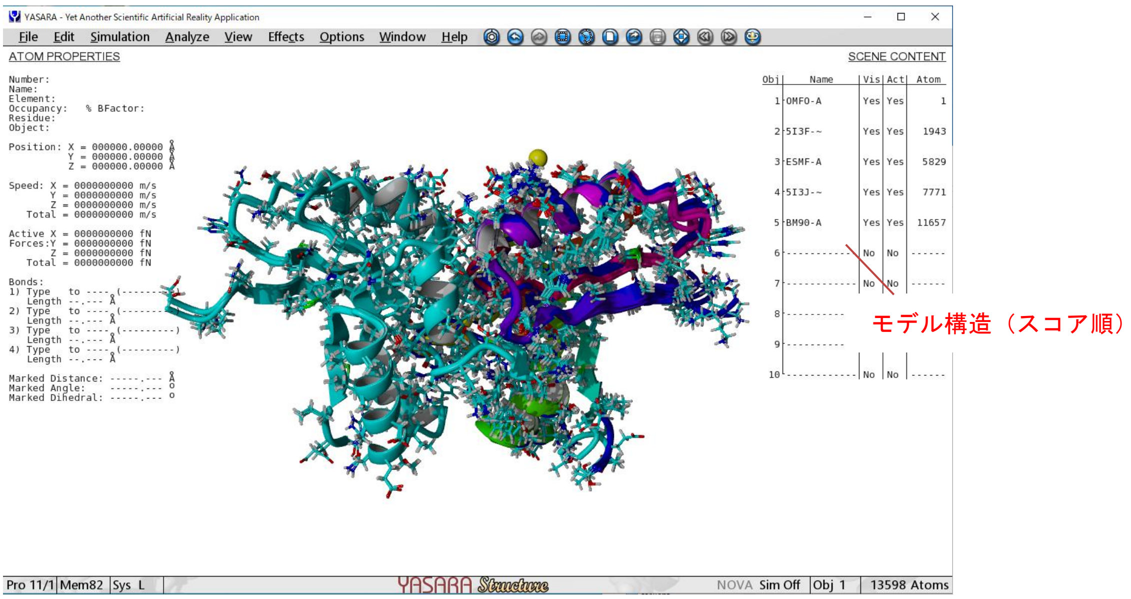



Figure 3.14 各モデル構造一覧（～.sce ファイル）

このファイルには、採用された各テンプレート構造を基に作成されたモデル構造が保存されています。
右側の「SCENE CONTENT」では、スコアが高い順にリスト化された各モデル構造を確認することができます。モデル構造の残基ごとの Z スコアは-20 倍されて B Factor 値に格納されているので、以下の方法でカラーマッピングして可視化することができます。
【注意】B Factor（温度因子）は、本来原子の熱振動の大きさを表すパラメータですが、タンパクモデリングの実行時には、Z スコアをカラーリングにより可視化できるようにするために、一時的に B Factor 値のパラメータに Z スコアを反映した値が格納されます。

【B Factor（Z スコア）の可視化】

1. メニュー画面から、View > Color > All を選択します。（Object や Molecule などで個別に選択してもよいです）

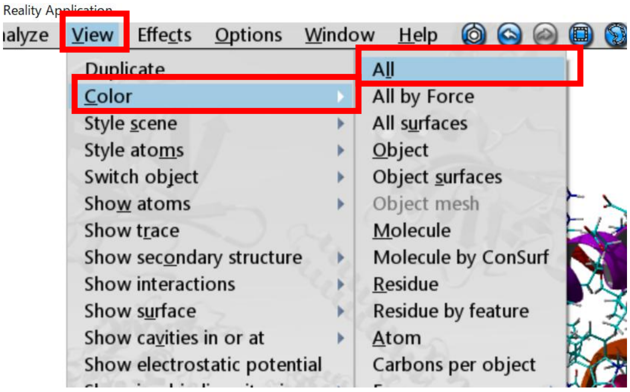



Figure 3.15 カラー設定メニュー

2. 続いて、色の指定画面が表示されるので、右側の Name リストから、「B Factor」を選択し、「Apply unique color」をクリックします。

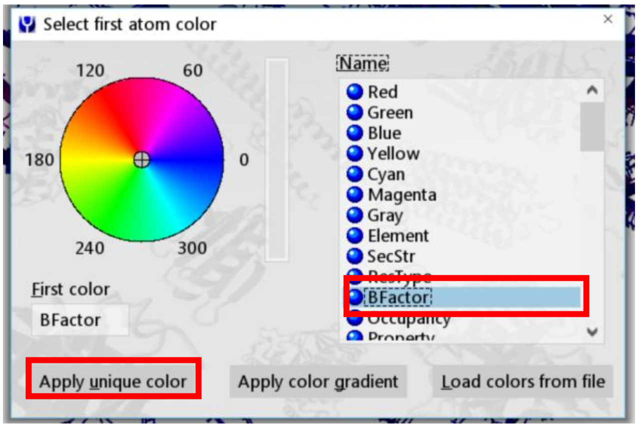



Figure 3.16 色の選択画面

3. すると、B Factor 値を可視化することができます。

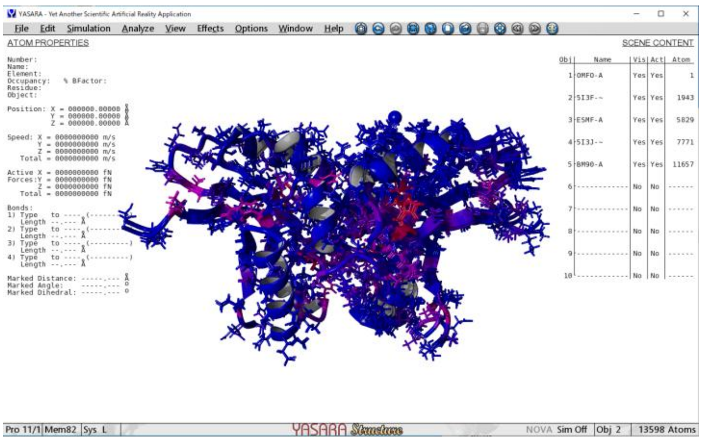



Figure 3.17 B Factor の可視化

B Factor の配色は、青(0)から黄色(75)のグラデーションになっています。ここでは Z スコアを-20 倍した値が B Factor 値に格納されており、青色(0)に近い方が良い結果ということになります。

| B Factor | Z-score | Description |
|---|---|---|
| ＞100 | ＜-5 | disgusting |
| ＞80 | ＜-4 | terrible |
| ＞60 | ＜-3 | bad |
| ＞40 | ＜-2 | poor |
| ＞20 | ＜-1 | satisfactory |
| ＞0 | ＜0 | good |
| ＜0 | ＞0 | optional |

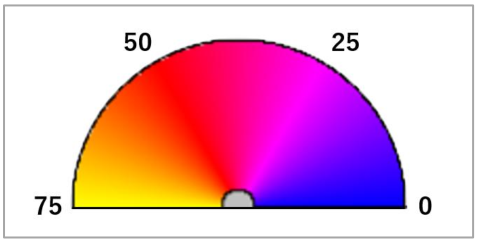



Figure 3.18 B Factor 値と Z スコアの対応(左)と B Factor の配色(右)

また、B Factor に格納された値は、画面左の「ATOMPROPERTIES」から確認することができます。確認したいアミノ酸に含まれるいずれかの原子をクリックしてマーク状態にすると、画面に出力されます。

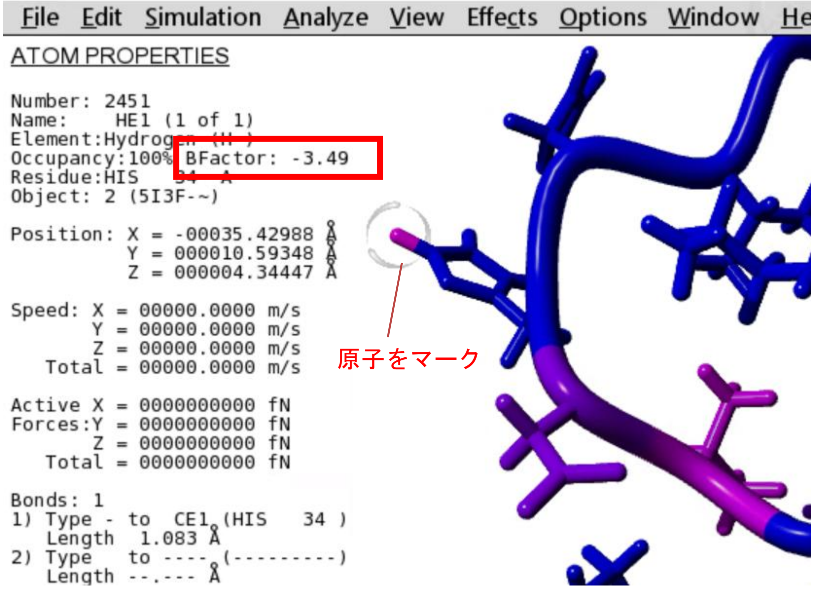



Figure 3.19 B Factor 値の確認

## 4. 応用：独自のテンプレートやアライメントを使って実行する

これまでに紹介した内容で基本的なタンパクモデリングは実行可能ですが、独自のテンプレートやアライメントを使用してタンパクモデリングを行うこともできます。この項目では、以下の内容を順にご紹介します。

1. テンプレートに使用する PDB 構造を指定する方法
2. 独自のテンプレート構造を使用する方法
3. 独自のアライメントを使用する方法
4. テンプレートとアライメントのどちらも独自に用意したものを使用する方法

これらは基本操作+αの内容となりますので、必要な際に参考にしてください。

### 4.1 テンプレートに使用する PDB 構造を指定する方法

| | |
|---|---|
| 対応マクロファイル： | 「pm_build.mcr」または「hm_build.mcr」 |
| 必要なファイル： | ターゲット配列「（任意のファイル名）.fasta」（FASTA 形式） |

※（任意のファイル名）の文字列がマクロターゲットとなります。

特定の PDB 構造をテンプレートに指定したい場合は、マクロファイル（「pm_build.mcr」または「hm_build.mcr」）を編集し、「TemplateList」パラメータ（テンプレートに使用する PDB 構造のリスト）を有効にして実行します。必要に応じて「Templates」（使用するテンプレートの総数）パラメータも変更します。

ここではマクロファイルを使った実行方法を紹介しますが、Experiment メニューからも【手順 4.2】の方法で、独自のテンプレートとして PDB 構造を読み込めば実行できます。（この場合は独自に準備したテンプレートとして扱われ、テンプレートとして採用される順番は AI ベースの構造よりも優先されます。）

#### 4.1.1 マクロファイルの編集（TemplateList）

「TemplateList」パラメータは、マクロファイルのはじめの方にあるパラメータセクション（「=」行で挟まれている部分）にはなく、後半にあるので、「TemplateList」で検索をかけて探してみてください。続いて、「TemplateList」をアンコメント（文頭の「#」を削除）し、テンプレートに使用したい PDB 構造を入力します。本チュートリアルでは「pm_build.mcr」を編集して以下の変更を行い、「5I3F」と「2V2C」を指定してみます。（マクロファイルの編集方法については【手順 2.2.3】を参照してください）

（変更前）　　　　　　　　　　（変更後）

`# TemplateList 2ZPY,2HE7,2I1K`　→　`TemplateList 5I3F,2V2C`

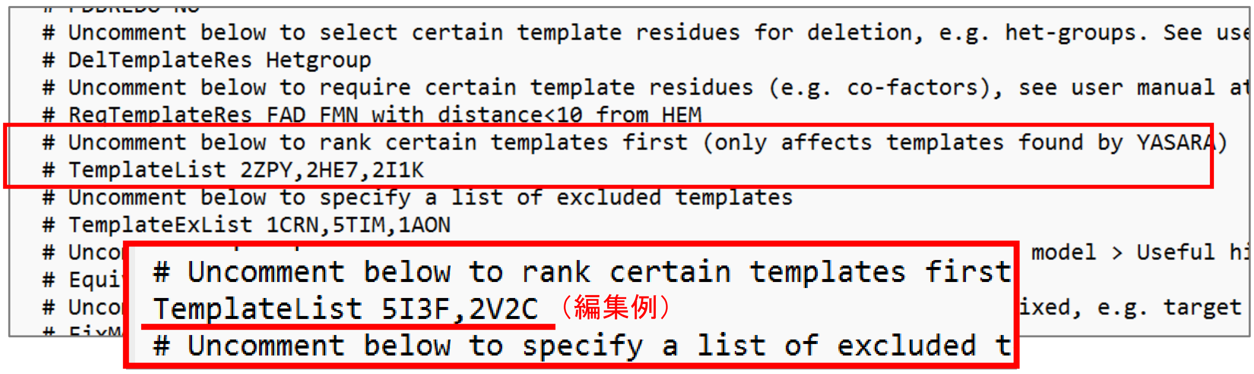



Figure 4.1 マクロファイル編集例：TemplateList の変更

#### 4.1.2 マクロファイルの編集（Templates）（任意）

※この操作は必要に応じて行ってください。本チュートリアルではスキップします。

指定した PDB 構造のみをテンプレート構造として利用したい場合は、マクロ内の「Templates」(使用するテンプレートの総数)パラメータを、「TemplateList」にて指定した PDB 構造の数と一致するように変更します。デフォルトでは「Templates=5」となっているため、指定した構造の数がこれより少ない場合は、自動的に残りの数のテンプレートが PDB から検索されてしまいます。
また、AI ベース手法を利用したテンプレート構造が不要な場合は、「hm_build.mcr」(テンプレート構造の準備に AI ベース手法を利用しない)を利用するとよいです。

マクロファイルを開き、前半部分にある「# Maximum number of templates to use…」と記載された項目を探してください。その下の行、「template=(数値)」でテンプレート数を指定しているので、(数値)部分を用意したテンプレート数に書き換え、保存します。(マクロファイルの編集方法については、前述の【手順 2.2.3】を参照してください。)

```
# Maximum PSI-BLAST Evalue allowed for templates
evalue=0.5

# Maximum number of templates to use (if you provide fewer templates, the others will be picked by YASARA)
templates=5

# Maximum number of ambiguous alignments to consider per template
alignments=5

# Maximum oligomerization state, build at most decameric models
```

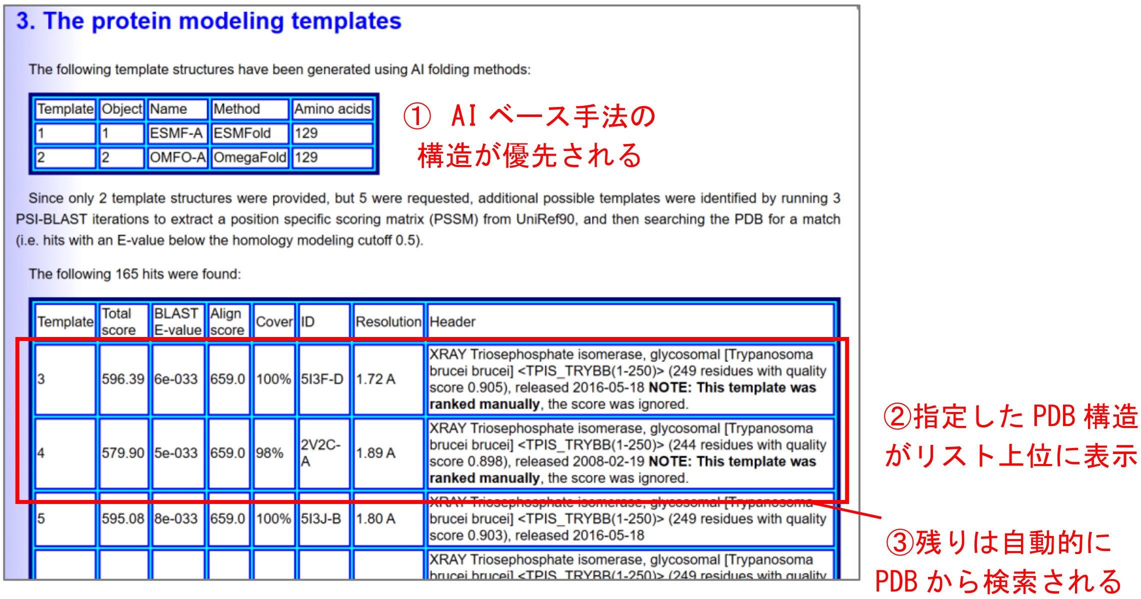
Figure 4.2 マクロファイル編集例:templates の変更

### 4.1.3 マクロ実行

マクロファイルが準備できたら、【手順 2.2】と同様に進めていき、マクロファイルの選択画面で、上で準備した編集済みマクロファイルを選択して実行します。

【Note】
「pm_build.mcr」を使用し、AI ベース手法を利用した場合は、テンプレートとしてまず AI ベース手法の構造が優先して採用されます。続いて、TemplateList にて指定した PDB 構造が採用され、指定したテンプレートの総数に余りがあれば、残りの分は PDB から検索されます。(下図は foldmethod パラメータで ESMFold と OmegaFold を指定した時の例)

Figure 4.3 テンプレート採用時の優先順位


**3. The protein modeling templates**

The following template structures have been generated using AI folding methods:

| Template | Object | Name | Method | Amino acids |
|---|---|---|---|---|
| 1 | 1 | ESMF-A | ESMFold | 129 |
| 2 | 2 | OMFO-A | OmegaFold | 129 |

Since only 2 template structures were provided, but 5 were requested, additional possible templates were identified by running 3 PSI-BLAST iterations to extract a position specific scoring matrix (PSSM) from UniRef90, and then searching the PDB for a match (i.e. hits with an E-value below the homology modeling cutoff 0.5).

The following 165 hits were found:

| Template | Total score | BLAST E-value | Align score | Cover | ID | Resolution | Header |
|---|---|---|---|---|---|---|---|
| 3 | 596.39 | 6e-033 | 659.0 | 100% | 5I3F-D | 1.72 A | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (249 residues with quality score 0.905), released 2016-05-18 **NOTE: This template was ranked manually**, the score was ignored. |
| 4 | 579.90 | 5e-033 | 659.0 | 98% | 2V2C-A | 1.89 A | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (244 residues with quality score 0.898), released 2008-02-19 **NOTE: This template was ranked manually**, the score was ignored. |
| 5 | 595.08 | 8e-033 | 659.0 | 100% | 5I3J-B | 1.80 A | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (249 residues with quality score 0.903), released 2016-05-18 |
| | | | | | | | XRAY Triosephosphate isomerase, glycosomal [Trypanosoma brucei brucei] <TPIS_TRYBB(1-250)> (249 residues with quality |

### 4.2 独自の構造をテンプレートに使用する方法

独自の構造（自社作成の構造や、手動で修正を加えた構造など）をテンプレートとして利用したい場合には、事前に作業ディレクトリにテンプレート構造を保存しておくだけで、マクロ実行時に自動的にファイルが読み込まれてモデリングに使用されます。
また、事前にテンプレート構造を YASARA 操作画面上にロードしておけば、Experiment メニューからも実行することができます。
なお、いずれの場合も必要に応じてマクロ内の「Templates」（使用するテンプレートの総数）パラメータを変更してください。（【手順 4.1.2】を参照）なお、独自に準備したテンプレート構造は、AI ベース手法の構造よりも優先して採用されます。

#### 4.2.1 マクロを用いて実行する方法

| | |
|---|---|
| 対応マクロファイル： | 「pm_build.mcr」または「hm_build.mcr」 |
| 必要なファイル： | ①ターゲット配列「（任意のファイル名）.fasta」（FASTA 形式） |
| | ②テンプレートファイル「（任意のファイル名）_t001.pdb」、「（任意のファイル名）_t002.pdb」…　（.yob 形式でも可） |
| マクロターゲット： | 作業ディレクトリ/（任意のファイル名） |

※（任意のファイル名）部分はすべて統一してください。この文字列はマクロターゲットとして使用します。

1. まず、テンプレートファイルを準備します。作業ディレクトリ（ターゲット配列が保存してあるフォルダ）に、用意したテンプレート構造を.pdb または.yob 形式で保存します。
2. ファイル名は、「(任意のファイル名）_ t001.pdb (or .yob)」としてください。（さらに増やす場合には末尾の 3 桁の番号を 1 つずつ増やして保存してください。）（任意のファイル名）部分には、ターゲット配列の FASTA ファイル名と共通の任意の文字列を指定します。この文字列は、後ほどマクロ実行時にマクロターゲットとして指定します。

（例：配列ファイル…YourModel.fasta　テンプレートファイル…YourModel_t001.pdb（1 つ目）、YourModel_t002.pdb（2 つ目）…）

3. 以上の準備ができたら、【手順 2.2】と同様に進めていき、マクロファイルを実行すると、作業ディレクトリに保存したテンプレート構造が自動的に読み込まれてモデリングに使用されます。

#### 4.2.2 Experiment メニューから実行する方法

| | |
|---|---|
| 必要なファイル： | ①ターゲット配列「（任意のファイル名）.fasta」（FASTA 形式） |
| | ②テンプレート構造（オブジェクト） |

まず、使用したいテンプレート構造を YASARA の操作画面にロードします。複数ある場合は全て読み込んでください。テンプレート構造となるファイルやオブジェクト名などに指定はありません。
続いて、手順「」と同様の操作で進めていくと、テンプレートに使用するオブジェクトの選択画面が出てきます。この画面から、使用するテンプレート構造のオブジェクトを選択（Shift キーや Ctrl キーを使って複数選択可）し、「OK」ボタンを押してください。残りの操作は【手順 2.1.3】と同様にすすめ、タンパクモデリングを実行します。なお、「Templates」（使用するテンプレートの総数）パラメータについては前述の通り、必要に応じて変更してください。（【手順 4.1.2】を参照）

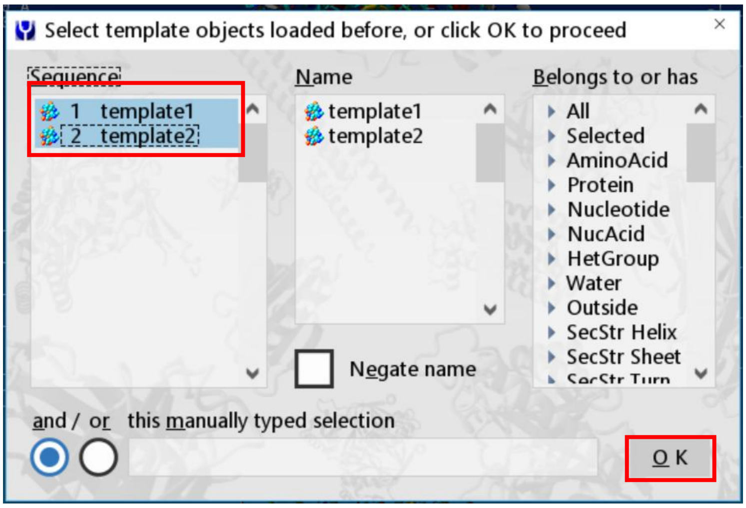


Figure 4.4 テンプレートオブジェクトの選択

### 4.3 独自のアライメントを使用する方法

| | |
|---|---|
| 対応マクロファイル: | 「pm_build.mcr」または「hm_build.mcr」 |
| 必要なファイル: | アライメントファイル「(任意のファイル名)_align.fasta」(FASTA 形式) |
| マクロターゲット: | 作業ディレクトリ/(任意のファイル名)　(「_align.fasta」部分は不要) |

フォールド予測が困難なモデルの場合など、時には手動でアライメントの調整をした方が便利な場合もあります。そのような場合は、ターゲットファイルの代わりにアライメントファイルを用意することで、独自に用意したアライメントを利用できます。
独自のアライメントを使用する場合は、マクロファイルを使って実行します。

#### 4.3.1 アライメントファイルの準備

まず、アライメントファイル(FASTA ファイル)を用意します。アライメントファイルの作成において、以下の点に注意してください。作成したアライメントファイルはファイル名に注意し、作業ディレクトリ(【手順 2.1.1】を参照)に保存します。

- ファイル名の末尾は「_align.fasta」とする
- アライメントされた状態のターゲット配列とテンプレート配列の組み合わせが 1 つ以上含まれている必要がある。
- 先にターゲット配列、次にテンプレート配列を記述する。
- 先頭や、シーケンスを中断する場所に空行を入れない。(他の部分なら空行を入れてもよい)
- 複数アライメントの場合、ターゲットシーケンスのヘッダーは統一する。(以下の例、「>YourModel」の部分、名前は任意)
- テンプレートシーケンスのヘッダーは PDB ID で開始する。ハイフンの後にチェーン(分子名)を指定できる。(例:2OUC-A や 2OUC-B など)
- 一行あたりのシーケンス文字数は関係なく、行を折り返す必要はない。
- ヘテロオリゴマーをテンプレートに使用するには、シーケンスのチェーンの区切り位置に'|' を挿入し、ヘッダーの PDB ID にチェーン情報を追記する。(例: PDB ID:1XYZ のチェーン(分子名)A と B を使用したい場合は、PDB ID を「1XYZ-AB」、オブジェクト内の全てのチェーン(分子)を使用する場合は「1XYZ-~」とする。)

アライメントファイルの例 (ファイル名:(任意のファイル名_align.fasta)

```
>YourModel ―― ターゲット配列のヘッダーは統一する
MIGTQIVTERLVALLESGTEKVLLIDSRPFVEYNTSHILEAININC-SKLMKRRLQQDKVLITELIQH---
SAKHKVDIDCSQKVVVYDQSSQDVASLSSDCFLTVLLGKLEKSFNSVHLLAGGFAEFSRCFPGLCEGKSTLVPTCISQPAH
HHHHH
>2OUC-A: First alignment variant with PDB file 2OUC-A―― テンプレート配列のヘッダーは PDB ID で開始する
---KIIYPNDLAKKMTKP----VIIDCRPFMEYNKSHIQGAVHINCADKISRRRLQQGKITVLDLISCREG--
KDSFKRIFSKEIIVYDENTNEPSRVMPSQPLHIVLESLKREGKEPLVLKGGLSSFKQNHENLCDNSKE----------------

>YourModel ―― ターゲット配列のヘッダーは統一する
MIGTQIVTERLVALLESGTEKVLLIDSRPFVEYNTSHILEAININC-SKLMKRRLQQDKVLITELIQHSA-
KHKVDIDCSQKVVVYDQSSQDVASLSSDCFLTVLLGKLEKSFNSVHLLAGGFAEFSRCFPGLCEGKSTLVPTCISQPAHHH
HHH
>2OUC-A: Second alignment variant with PDB file 2OUC-A
----KIIYPNDLAKKMTKP---
VIIDCRPFMEYNKSHIQGAVHINCADKISRRRLQQGKITVLDLISCREGKDSFKRIFSKEIIVYDENTNEPSRVMPSQPLHIV
LESLKREGKEPLVLKGGLSSFKQNHENLCDNSKE----------------
```

Figure 4.5 アライメントファイル構成例

#### 4.3.2 マクロ実行

【手順 2.2】と同様の操作でマクロファイルを実行します。なお、マクロターゲットには、アライメントファイルの「(任意のファイル名)_align.fasta」の(任意のファイル名)部分だけを指定します。アライメントファイルを選択後、Remove オプションにて「from underscore」

（アンダースコア以降を除く）にチェックを入れればよいです。

### 4.4 独自のテンプレートとアライメントを使用する方法

| | |
|---|---|
| 対応マクロファイル： | 「pm_build.mcr」または「hm_build.mcr」 |
| 必要なファイル： | アライメントファイル「（任意のファイル名）_align.fasta」（FASTA 形式）<br>テンプレートファイル「（任意のファイル名）_t001.pdb」、「（任意のファイル名）_t002.pdb」… （.yob 形式でも可） |
| マクロターゲット： | 作業ディレクトリ/（任意のファイル名）　（「_align.fasta」部分は不要） |

※（任意のファイル名）部分はすべて統一してください。この文字列はマクロターゲットとして使用します。

独自のテンプレート構造と、事前に用意したアライメントをどちらも使用するには、【手順 4.2】と【手順 4.3】を組み合わせて実行します。

#### 4.4.1 テンプレートファイルの準備

テンプレートファイルの準備方法については、手順「」に記載した通りです。こちらも同様に、ファイル名は「（任意のファイル名）_t001.pdb」、「（任意のファイル名）_t002.pdb」…のように、複数ある場合は末尾の数字を増やしてください。

#### 4.4.2 アライメントファイルの準備

アライメントファイルの準備方法についても、基本的には手順「」に記載した通りですが、独自のテンプレート構造を使用する場合は、追加で以下の点に注意してください。

- テンプレートシーケンスのヘッダーには、PDB ID の代わりに、【手順 4.4.1】で作成したテンプレートファイルの末尾の4文字を使用する。（例：T001、T002 など）また、ハイフンの後にチェーン（分子名）を指定できる 。（チェーン A を指定したい場合は「T001-A」、全体を指定したい場合は「T001-~」など）

アライメントファイルの例 （ファイル名：（任意のファイル名）_align.fasta）

```
>YourModel target sequence
---------------------MIGTQIVTERLVALLESGTEKVLLIDSRPFVEYNTSHILEAININCSKLMKRRLQQDKVLITEL---
IQHSAKHKVDIDC---SQKVVVVYDQSSQD--------
VASLSSDCFLTVLLGKLEKSFNSVHLLAGGFAEFSRCFPGLCEGKSTLVPTCISQPAHHHHHH
>T001-A: Alignment with custom template t001 ―― テンプレート配列のヘッダーはファイルの末尾の 4 文字で開始する
MIDTLRPVPFASEMAISKTVAW----------
LNEQLELGNERLLLMDCRPQELYESSHIESAINVAIPGIMLRRLQKGNLPVRALFTR--GEDRDRFTRRCGTD--
TVVLYDESSSDWNENTGGE-------SLLGLLLKKLKDEGCRAFYLEGGFSKFQAEFSLHCETNLDGS--------------
```

Figure 4.6 アライメントファイル構成例（独自テンプレートを使用）

#### 4.4.3 マクロ実行

【手順 2.2】と同様の操作でマクロファイルを実行します。なお、マクロターゲットには、アライメントファイルの「（任意のファイル名）_align.fasta」の（任意のファイル名）部分だけを指定します。アライメントファイルを選択後、Remove オプションにて「from underscore」（アンダースコア以降を除く）にチェックを入れればよいです。

## 5. おわりに

ターゲット配列の fasta ファイルさえあれば、簡単な操作でタンパクモデリングを行うことができます。また、HTML 形式の詳細なレポートが自動的に作成されるため、後から計算条件やパラメータを参照するのも容易です。
また、タンパクモデリングによって得られた構造を用いて低分子ドッキングや分子動力学計算にスムーズに移行でき、YASARA Structure 単独で構造ベース手法の多くをご利用いただけます。
弊社 HP では、低分子ドッキングや分子動力学計算の各種チュートリアルも公開していますので、よろしければそちらもご参照ください。

YASARA 日本語チュートリアル：https://www.affinity-science.com/yasara-tutorial

ぜひ研究にお役立てください。

## 6. 参考情報

### 6.1 ユーザーマニュアルについて

タンパクモデリング実行についての詳細は、ユーザーマニュアル（Help > Show user manual）の、Recipes - Perform complex tasks > Build a protein model に記載されています。

- ・Protein modeling the easy way　:基本的なタンパクモデリングの実行手順
- ・Providing your own templates　:独自のテンプレートを利用する場合の実行方法やポイントなど
- ・Providing your own alignments　:独自のアライメントを利用する場合の実行方法やポイントなど
- ・How YASARA builds protein models　:タンパクモデリングの処理内容の詳細とリファレンス
- ・Useful protein modeling hints　:タンパクモデリングを実行するうえでの注意事項やヒントなど

特に、「Useful protein modeling hints」には、タンパクモデリングに関するヒントが記載されているので、疑問点がある場合やうまくいかない場合などは、参考にしてみてください。

### 6.2 各種パラメータの概要とマクロ（pm_build.mcr、hm_build.mcr、hm_buildfast.mcr）の違いについて

タンパクモデリングに対応する YASARA 付属マクロは、pm_build.mcr、hm_build.mcr、hm_buildfast.mcr、の 3 種類があります。

- pm_build.mcr

  YASARA ver.24.04.10 から追加されました。AI ベース手法を利用した予測構造をテンプレートとして用いてホモロジーモデリングを行うことができます。AI ベース手法は ESMFold、OmegaFold、Boltz、AlphaFold の 4 つを利用できます。ESMFold、OmegaFold、Boltz は YASARA に含まれており、生成された予測構造をテンプレートとして利用します。Boltz は、①MSA なし、②Uniref90 データベースを用いた MSA を加える、③AlphaFold データベースを用いた MSA を加える、の 3 つのオプションを利用できます。一般的には MSA を利用した方がより高精度の構造予測が可能です。デフォルトでは、ESMFold、OmegaFold、Boltz（Uniref90 データベースを用いた MSA を利用）の 3 つの AI ベース手法が有効になっています。オプションで AlphaFold も選択できますが、こちらは AlphaFold を利用して予測構造を生成するのではなく、AlphaFold により予測された約 2 億の配列データベース（EBI）をダウンロードし、ターゲット配列と最も近い配列をもつ立体構造をホモロジーモデリングのテンプレートとして採用します。テンプレート数に余りがあれば、残りのテンプレートは従来の hm_build.mcr と同じように、PDB からスコアの良いものが検索・採用されます。AI ベース手法はテンプレート構造の準備に利用され、その後の流れは hm_build.mcr と同じようになっています。

  詳細は、以下の YASARA 公式サイト、タンパクモデリングの機能紹介ページをご覧ください。

  https://www.yasara.org/proteinfolding.htm

- hm_build.mcr

  ターゲット配列情報を元に、PDB を検索してスコアが上位の構造をテンプレートとしてホモロジーモデリングを実行します。テンプレート構造の準備に AI ベース手法を利用しなくてよい場合は、こちらのマクロをお使いください。

  詳細については、以下の YASARA 公式サイト、ホモロジーモデリングの機能紹介ページをご覧ください。

  https://yasara.org/homologymodeling.htm

- hm_buildfast.mcr

  上記の hm_build.mcr の簡易版です。以下のショートカットが行われます。この他にもいくつか異なるパラメータがあるので、比較表をご覧ください。

  1. テンプレート数（templates パラメータ）とテンプレートごとに許容される異なるアライメントの数（alignments パラメータ）が減る。
  2. アライメントは配列プロファイル（計算コストが高い PSI-BLAST を必要とする）を使用せずに作成される。
  3. ループサンプル数（Loop samples パラメータ）が減る。
  4. 側鎖の配座予測には SCWRL アルゴリズムと最急降下法による最適化だけを使用し、計算コストの高いランダムサンプリングのステップを省略する。
  5. モデルのウォーターシェル内での構造最適化を行わない。
  6. 複数のモデルの断片を組み合わせるハイブリッドモデルは作成されない。

変更可能な各種パラメータと、3 種類の付属マクロのデフォルト値の一覧を下に示します。「パラメータ」欄には、Experiment から実行する際に表示されるパラメータ設定画面での表記を記載しており、()内には対応するマクロ内変数を記載しています。

| パラメータ<br>(マクロ内変数) | 説明 | pm_build | hm_build | hm_buildfast |
|---|---|---|---|---|
| AlphaFold/ESMFold/OmegaFold<br>(foldmethod) | テンプレート作成のために利用する AI ベース手法を指定。 | ESMFold+<br>OmegaFold+<br>BoltzMSA90 | 設定なし | 設定なし |
| PSI-BLAST iterations<br>(psiblasts) | ターゲット配列との相同な配列を探索する反復回数を指定。「1」にすると、単純な BLAST 検索のみが行われる。 | 3 | 3 | 3 |
| PSI-BLAST E-value<br>(evalue) | ターゲット配列との相似範囲を指定。値が小さいほど、相似性が高くなる。 | 0.5 | 0.5 | 0.1 |
| Templates...<br>(templates) | ホモロジーモデリングを行うテンプレートの最大数を指定。 | 5 | 5 | 3 |
| Alignments per template<br>(alignments) | 1 つのテンプレートにつき、異なるアライメントの許容数を指定。 | 5 | 5 | 1 |
| Min oligomerization<br>(oligostatemin) | ターゲット配列が種々のオリゴマー(重合体)を形成する場合を考慮し、生成するモデルのオリゴマーの最小値を指定。 | 1 | 1 | 設定なし |
| Max oligomerization<br>(oligostatemax) | オリゴマーを形成し得る最大の分子数を指定。 | 4 | 4 | 設定なし |
| Oligomerization state<br>(oligostate) | オリゴマーを形成し得る最大の分子数を指定。 | 設定なし | 設定なし | 4 |
| Terminal extension<br>(termextension) | テンプレート配列がターゲット配列より短い場合に、末端に付加する残基の最大数を指定。 | 10 | 10 | 10 |
| ...with same sequence<br>(SameSeq) | 同じ配列を持つテンプレートの最大数を指定。 | 1 | 1 | 1 |
| Loop samples<br>(LoopSamples) | ループの構造最適化の際に計算するループのコンフォメーション数を指定。 | 50 | 50 | 25 |
| Use PSSP<br>(structprofile) | 有効にすると、より詳細な PSSP (Profiles from Sequence- and Structurally related Proteins)データベースを利用することができる。 | try | try | 設定なし |
| Modeling speed | Experiment の Speed 設定行を「fast」に設定すると、いくつかのステップが省略され、スピードが優先される。 | 設定なし<br>(slow) | 設定なし<br>(slow) | fast |

Table 6.1 各マクロとパラメータのデフォルト値

### 6.3 タンパクモデリングの計算プロセス

| 手順 | 計算内容 | 補足事項(関連パラメーター等) |
|---|---|---|
| 1 | ＜AI ベース手法を有効にした場合のみ＞<br>AI ベース手法を利用し、テンプレート構造を生成する。 | foldmethod:テンプレート作成のために利用する AI ベース手法を指定する。ESMFold・OmegaFold・Boltz は YASARA に含まれており、構造予測を行う。Boltz は、MSA なし「Boltz」、UniRef90 に対する MSA 使用「BoltzMSA90」、AlphaFold データベースに対する MSA 使用「BoltzMSAAF」の 3 つのオプションを指定可。AlphaFold を有効にすると、EBI の配列データベース(初回にダウンロードされる)に対して BLAST 検索が行われ、ターゲット配列ともっとも近い立体構造がテンプレート構造として採用される。 |
| 2 | 相同配列の探索。<br>ターゲット配列の PSSM(position-specific scoring matrix)に基づき、データベース「UniRef90」を用いて PSI-BLAST を行う。 | Psiblast:PSI-BLAST の回数を指定する。<br>E-Value:合致精度を指定する。 |
| 3 | テンプレートとなる PDB の探索と順位付け。<br>手順 2 で探索された相同配列を元に、モデリングのテンプレートとして利用できる PDB を探索し、アライメントスコアに基づいて順位を付ける。 | 計算結果の HTML ファイルには、探索結果として出された全ての PDB が相同配列との合致スコア順に表示される。上から「Templates」で指定された数までの PDB がテンプレートとなり、以降の計算が行われる。<br>Templates:テンプレートの数を指定する。<br>TemplateList:独自の順位の付け方を用いる。 |

| | | |
|---|---|---|
| 4 | ターゲット配列の二次構造予測。<br>ターゲット配列の各残基に対して、ヘリックス(H)、ストランド(S)、コイル(C)構造となる可能性を 0〜9 のスコアで示し、高スコアの構造が採用される。 | 計算結果の HTML ファイルには、3 種の構造のスコアと二次構造が示される。<br>SecStrFile：手動で二次構造を入力する。 |
| 5 | ターゲット配列とテンプレートの配列プロファイルの作成。<br>データベース「UniRef90」を用いてターゲット配列に対して BLAST を行い、得られたヒットとアライニングすることで、配列プロファイルを作成する。 | テンプレートに関する配列プロファイルの一部は計算結果の HTML ファイルに、全結果は*profile.fasta ファイルに表示される。 |
| 6 | ターゲット配列とテンプレートのアライメント。<br>2 つのプロファイルをアライメントすることで、ターゲット配列の二次構造とテンプレートの二次構造の相同性とテンプレートに含まれる構造に関する情報を考慮する。 | 配列アライメント結果の一部は計算結果の HTML ファイルに、全結果は*profile.ali ファイルに表示される。<br>Equivalence：手動でのアライメント、アライメントの修正を行う。<br>RequireRes：テンプレートを削除する独自の基準を指定する。<br>Alignments：アライメントが上手くいかず、代わりのモデルを作成する場合の基準を指定する。<br>oligostatemin：ターゲット配列が種々の重合体を形成する場合を考慮し、形成し得るオリゴマーの最小数を指定する。<br>oligostatemax：形成し得るオリゴマーの最大数を指定する。<br>oligostate：形成し得るオリゴマーの最大数を指定する。 |
| | ＜Fast 計算のみ＞＜残基の挿入や欠陥がある場合＞<br>最適なループ末端を考慮してループ構造を作成する。 | LoopLenMax：ループの長さにテンプレートを除外する基準を指定する。<br>Slow 計算では、ハイブリッドモデル形成時に同様の計算を行うため、この段階でループは作成されない。 |
| | ＜ターゲット配列がテンプレート配列より長い場合＞<br>テンプレート配列の N 末端及び C 末端に残基を加える。 | TermExtension：末端に加える残基の最大数を指定する。 |
| | ＜テンプレートがリガンドを含む場合＞<br>リガンドとなる分子をパラメータ化し、ホモロジーモデリングで考慮する。 | DelTemplateRes：ホモロジーモデリングで考慮する内容をコメントする。 |
| 7 | ロータマー最適化。<br>dead-end elimination を用いて側鎖のロータマー最適化を行う。 | FixModelRes：手動で修正を加える。 |
| 8 | ループ構造最適化。<br>異なるコンフォメーションを計算することで、ループ構造の最適化を行う。 | LoopSamples：計算するコンフォメーションの数を指定する。 |
| 9 | 静電気的な相互作用や溶媒効果などの相互作用を考慮した側鎖ロータマー最適化の再計算を行う。 | |
| 10 | エネルギー最適化。<br>YASARA の力場を用いて、溶媒シェル内でのエネルギー最適化計算を行う。 | 前半の最適化計算(残基を固定した計算)結果は*_refined50.yob に、後半の計算(残基の可変性を考慮した計算)結果*_refined100.yob に保存される。最終モデルとアライメント結果は*˜.yob に保存され、計算結果の HTML ファイルにも表示される。 |
| 11 | 各テンプレートに対し 5〜9 が行われた後、モデルの順位付けを行う。 | 計算結果は*.sce に保存され、一部は HTML ファイルに表示される。 |
| | ＜Slow 計算のみ＞<br>ハイブリッドモデルの形成。<br>pH 依存性やリガンドを考慮したハイブリッドモデルを形成する。 | 計算結果は HTML ファイルに表示される。<br>ハイブリッドモデルは*.yob として保存される。 |

Table 6.2 モデリングの流れ

### 6.4 HugeUniRef90 データベースのダウンロード

使用マシンが solid-state disk を利用しており、かつ約 20GB の空き容量が存在する場合で、処理時間が長くなってもモデリング精度を約 2%向上させたい場合は、「Help」→「Install」→「huge UniRef90 database」をクリックして、UniRef90 シーケンスデータベースの最新バージョン（大容量）をダウンロードすることができます。

※huge UniRef90 database のダウンロードを行わなかった場合は、初回マクロ実行時に通常の UniRef90 database（合計約 6 GB、ダウンロードファイル.gz:約 1.8 GB＋展開後のファイル:約 4.3 GB）がダウンロードされます。

### 6.5 基本用語

タンパクモデリングにおいて用いられる基本用語を以下に記載します。

| 用語 | 説明 |
|---|---|
| ターゲット（Target） | 立体構造（三次構造）を予測したいタンパク質 |
| テンプレート（Template） | 立体構造が既知で、構造予測の鋳型として使用するタンパク質 |
| モデル（Model） | テンプレートを基に作成されたターゲットタンパク質の予測構造 |
| アライメント（Alignment） | 複数の配列を相互に比較するために、同じ配列を示す部分をそろえたり、ギャップを挿入したりして並べること。 |
| プロファイル（profile） | PSSM を参照 |
| ホモロジーモデリング | 立体構造未知のタンパク質（ターゲット）の立体構造を、類似の配列（相同性）を持つ立体構造既知のタンパク質（テンプレート）を参照して予測する手法 |
| モノマー（サブユニット） | タンパク質が複合体を形成する場合の各ポリペプチド鎖。タンパク質の最小構造単位 |
| オリゴマー | 複数のモノマーが集合してできた複合体 |
| ホモオリゴマー | 同じ種類のモノマーが複数集まってできたオリゴマー |
| ヘテロオリゴマー | 異なる種類のモノマーが集まってできたオリゴマー |
| ESMFold | Meta 社によって開発された AI ベースのフォールディング手法。YASARA に内包されている。 |
| OmegaFold | 中国の Helixon 社によって開発された AI ベースのフォールディング手法。YASARA に内包されている。 |
| AlphaFold | Google/Alphabet により開発された AI ベースのフォールディング手法。YASARA では EBI に収集されている配列と予測構造のデータベースを利用している。 |
| Boltz-2 | AlphaFold3（AF3）のオープンソース代替モデルである Boltz-1 を基盤とし、分子複合体の構造予測と結合親和性予測を同時に実行するために開発された、生体分子基盤モデル（Biomolecular Foundation Model）と呼ばれる新しい AI プログラム。YASARA に内包されている。YASARA のタンパクモデリングでは、AI 手法の中で唯一オリゴマーの予測が可能。<br>※現時点（2025/12/01）では、YASARA に結合親和性予測機能は実装されておりません。<br>※本チュートリアルおよび YASARA 内で「Boltz」と記載されている場合は、Boltz-2 と同義です。 |
| UniRef90 | UniProt の配列データベースの一つ。UniRef100 の配列をクラスタリングすることにより構築されており、各クラスタは、最長配列（シード配列）と少なくとも 90%の配列同一性を持ち、80%の重複を持つ配列で構成されている。 |
| BLAST | 配列のシーケンスアライメントを行うためのアルゴリズムの一つ。BLAST を使ってデータベースに対して検索を行い、類似する配列を見つけることができる。 |
| PSI-BLAST<br>(Position-Specific<br>Iterated BLAST) | BLAST で見つかった配列から PSSM(プロファイル)を作り、それをもとに検索するという作業を繰り返す。配列自体の類似度が低い場合でも機能的に関連している配列を検出できる場合がある。 |
| PSSM<br>位置特異的スコア行列 | モチーフ（機能に直結する特定の配列パターン）の位置ごとのアミノ酸の出現率から算出されるスコア行列（プロファイル）。 |
| Z-score | モデルの品質が平均的な高分解能 X 線構造から何標準偏差離れているかを表す。モデルの全体的なスコアは、Overall = 0.145*Dihedrals + 0.390*Packing1D + 0.465*Packing3D という式を使用して、個々の Z スコアの加重平均として計算される。 |
| PDBFinder2 | タンパク質構造解析プログラム「WHAT IF」に含まれている、PDB ファイルを検索するためのデータベースで、クオリティスコアなどの情報が含まれる。 |
| WHAT_CHECK | タンパク質構造解析プログラム「WHAT IF」に含まれている、タンパク質の構造妥当性評価ツール |

Table 6.3 基本用語一覧

以上

## バージョン履歴

| 改定日 | 主な改定内容 |
|---|---|
| 2024 年 06 月 10 日 | 初版公開 |
| 2024 年 09 月 27 日 | AI ベースのフォールディング手法（ESMFold、OmegaFold、AlphaFold）に関する内容を追加 |
| 2025 年 12 月 02 日 | AI ベースのフォールディング手法（Boltz-2）に関する内容を追加 |
| 2026 年 03 月 18 日 | 日本語チュートリアルの URL を変更 |