

○ 藤 博幸¹, 山口 敦子²

1. 関西学院大学 生命環境学部 生物科学科, 2. 東京都市大学大学院総合理工学研究科 情報専攻 情報工学領域

目的

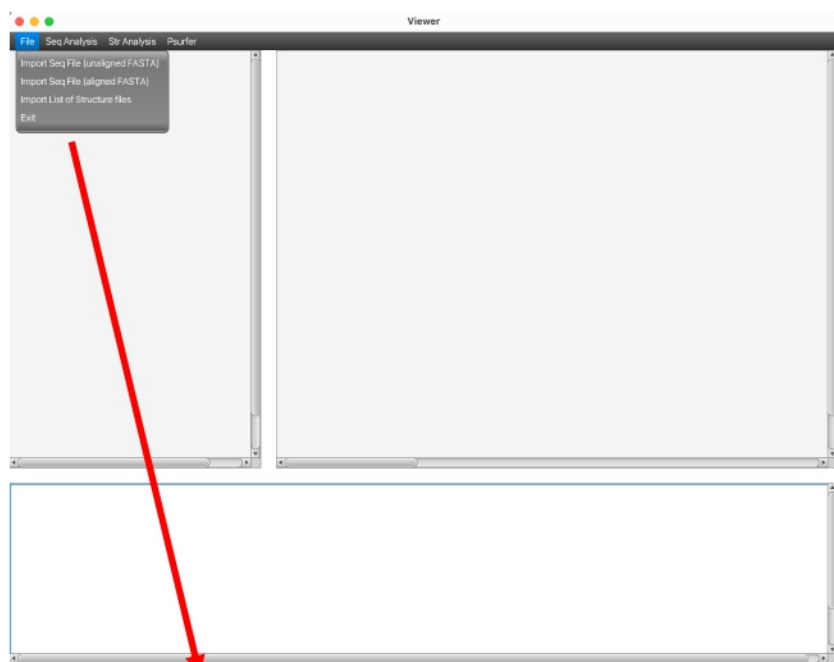
アラインメント配列を何らかの基準 (eg. 基質特異性の違いなど) でグループ化して比較
——→ グループ間の違いを特徴づけるサイト (基質特異性の違いを決定するサイト) の抽出

Linked Open Data (LOD) を利用したアラインメントビューアの開発

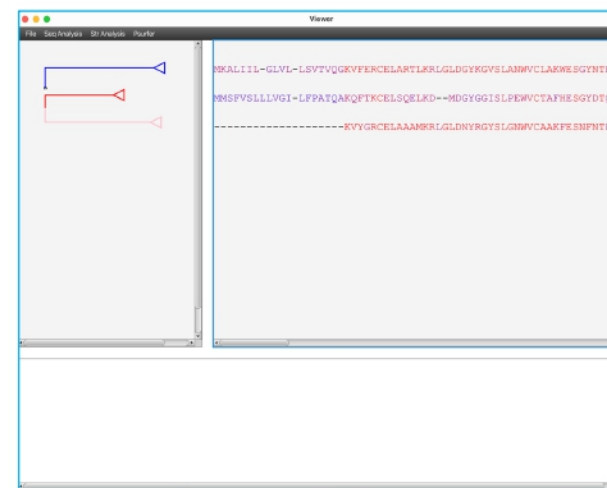
- (1) 相同配列のマルチプルアラインメント作成 (mafft使用)
- (2) LODからタンパク質のGO 情報や機能部位情報を収集 (PSurfer使用)
- (3) NJ法による無根系統樹の囲む同心円の形で収集情報を可視化
- (4) 同心円をクリックすることで、特定のGO情報や機能部位情報に関する相同配列のグループ化を容易にする
- (5) 選択されたグループごとの保存度、またグループ間の差異への寄与の大きさを計算
- (6) (4)のスコアのプロファイルの作成、スコア上位のアミノ酸の立体構造への表示 (PyMol使用)

ビューアはJavaで、ビューアのGUI部分はJava FXで作成

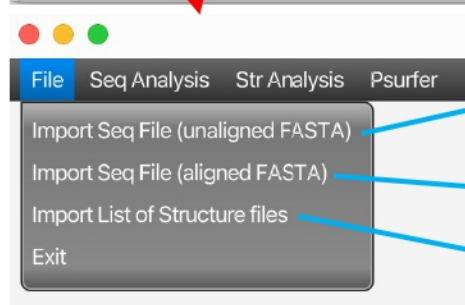
ビューアを起動. File メニューから配列データと構造データの読み込み



アラインメントを自動で
3つのグループの分割



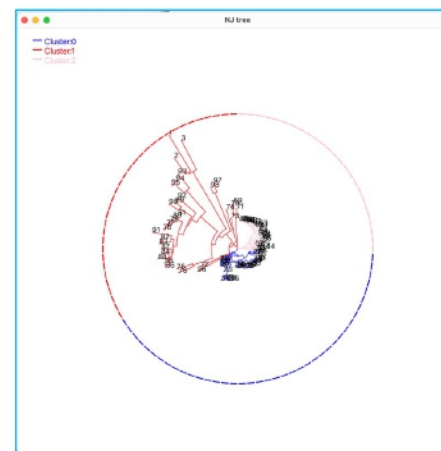
メインウィンドウに粗視化されたアラインメントと系統樹が表示



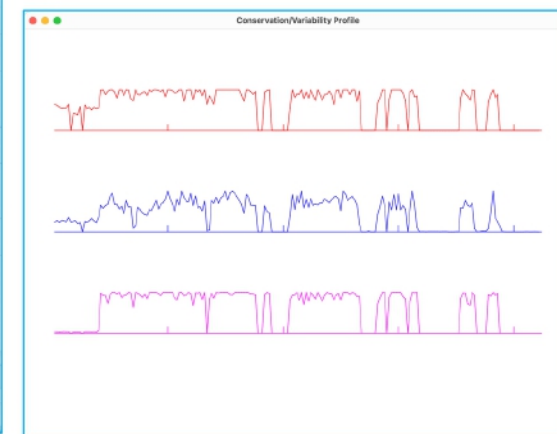
アラインされていない配列データの読み込み
mafftで自動的にアラインメント

マルチプルアラインメントの読み込み

mmCIF ファイルと chain ID の読み込み



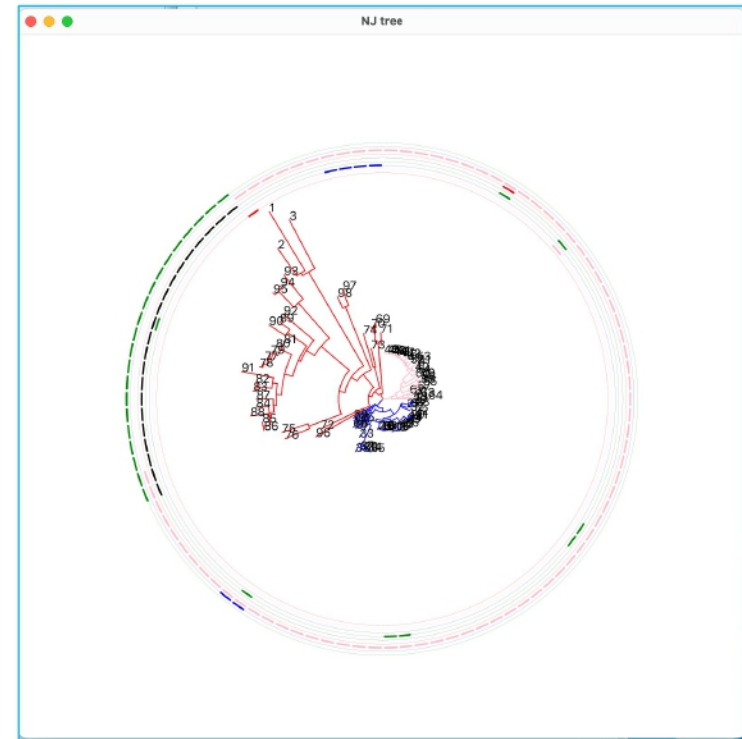
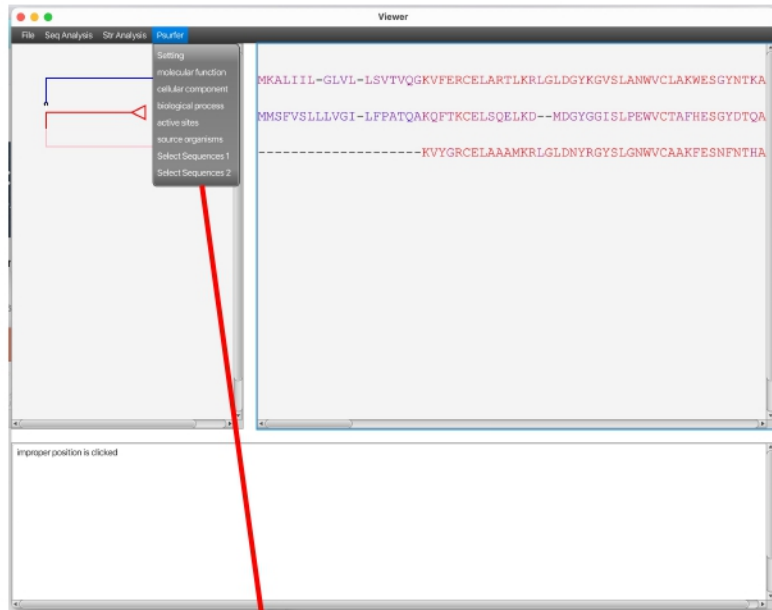
無根系統樹サブウィンドウ



スコアプロファイルサブウィンドウ

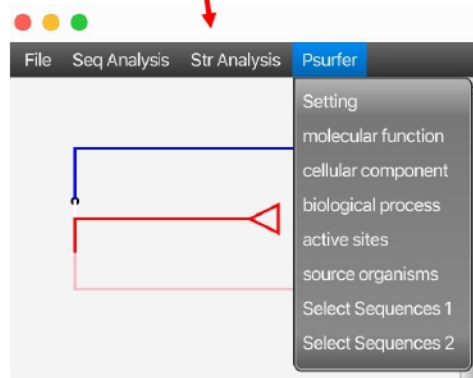
2つのサブウィンドウが生成

PSurferの解析結果の表示



MFを選択した場合の同心円. 各同心円がMFのGO term に対応. 同心円上の太い弧が、そのGO termを持つ配列を示す

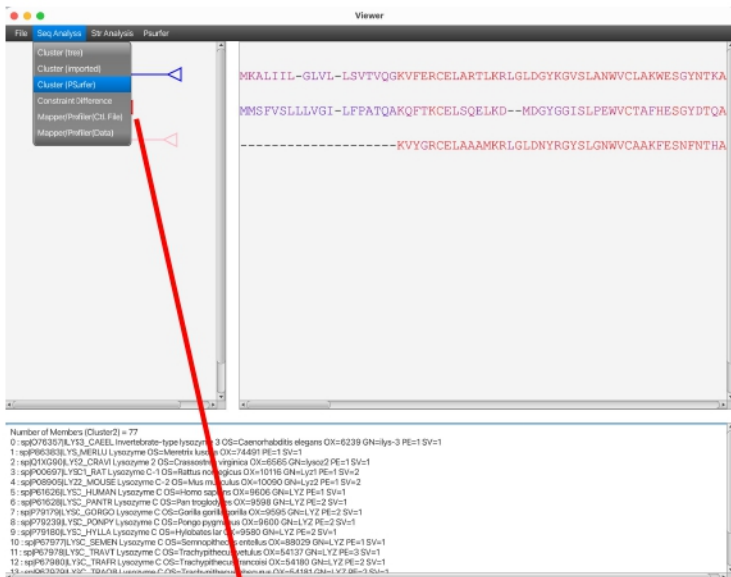
lysozymeファミリーに属す98本の配列のアラインメントの表示. 左上の一群が他のメンバーと異なるGO termを持つことがわかる. これらは触媒活性を失った α ラクトアルブミンよりなる



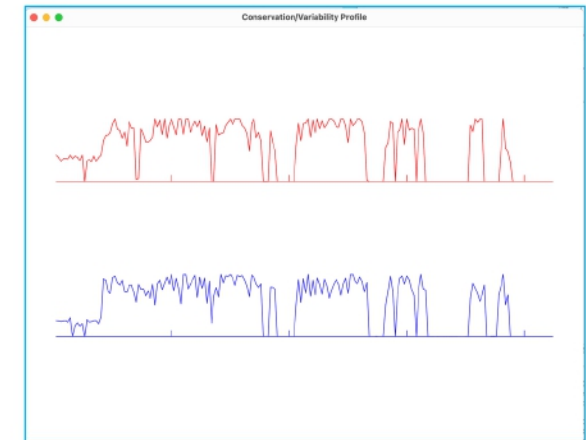
↑ Psurferメニューから、MF、CC、BP、active sitesのいずれかを選択すると、無根系統樹サブウィンドウの系統樹の周辺に同心円の形で選択された情報が表示

↑ Select Sequences 1と2を選択すると同心円がクリック可能になり、同心円上の弧を選択できるようになる. これで比較したい2つのグループを指定できる.

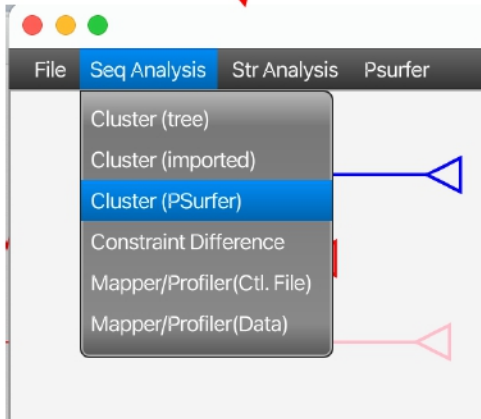
各クラスターの保存度の解析



PSurferメニューで選択した
2つのグループに対して、
選択した計算方法により
保存度プロフィールと
系統樹の同心円が再構築

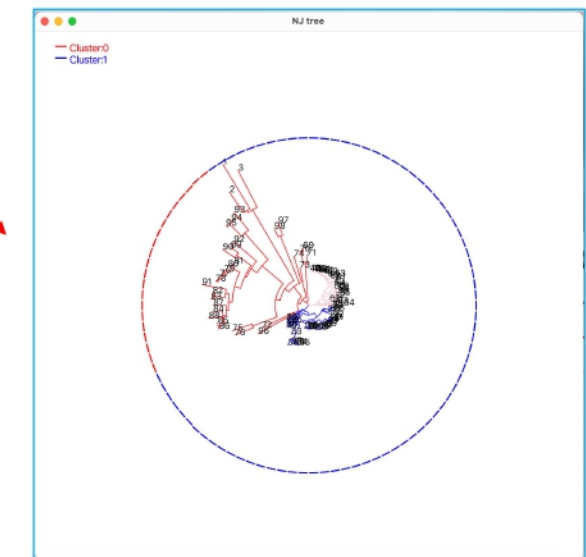
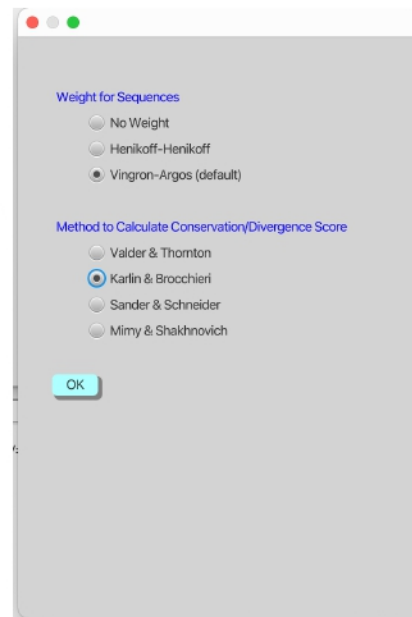


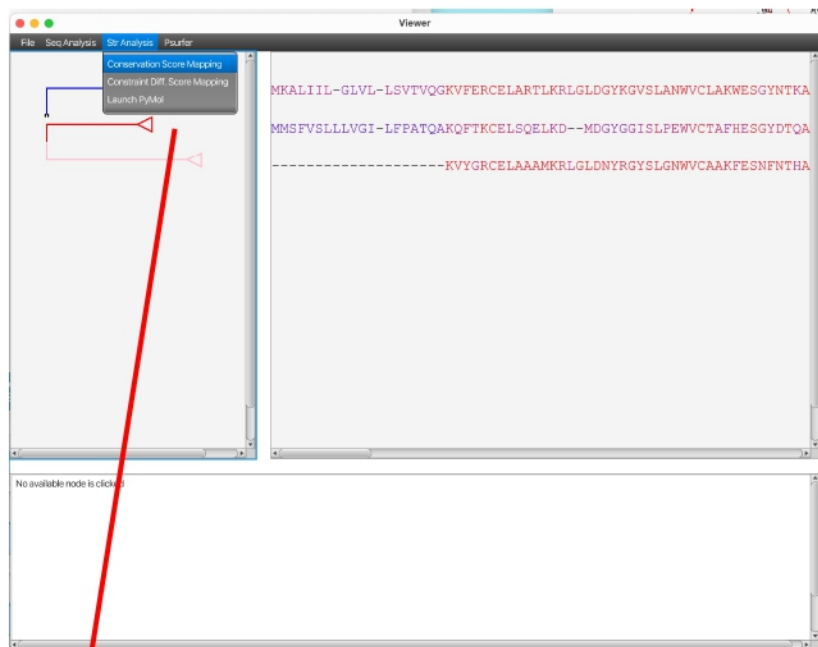
保存度のプロファイルはファイルにも出力される



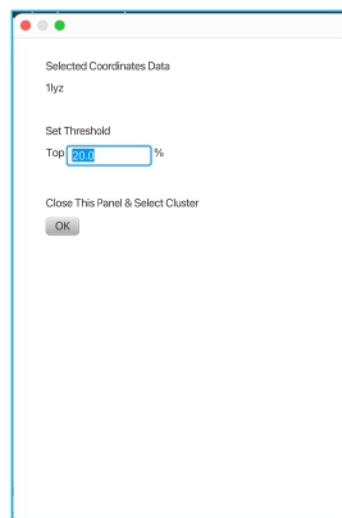
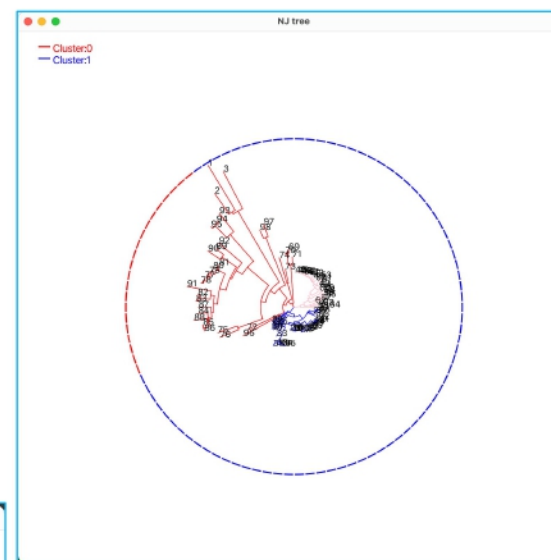
配列選択後に、Seq Analysis
メニューのCluster (Psurfer)
を選択

taxonomic biasの補正方法、
保存度の計算方法を選択する
サブウィンドウから補正
方法、計算方法を選択

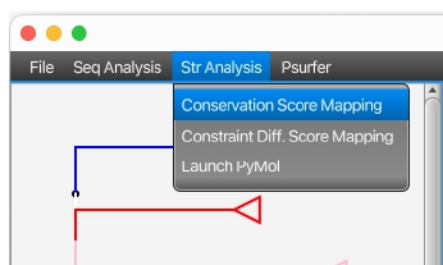




系統樹上部左に表示
されるクラスタの
いずれかをクリック

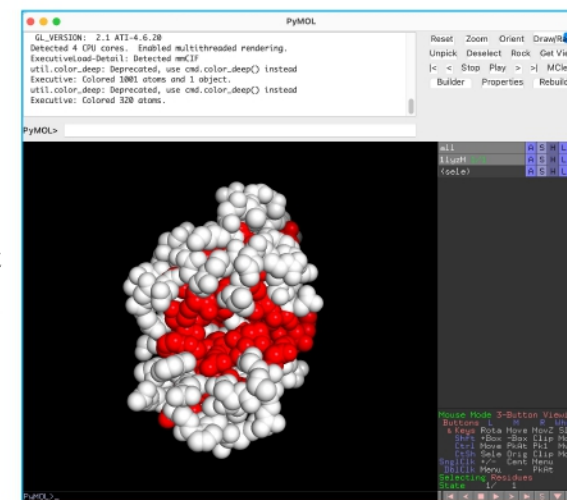


保存スコアの上位何%まで
を表示するかを選択



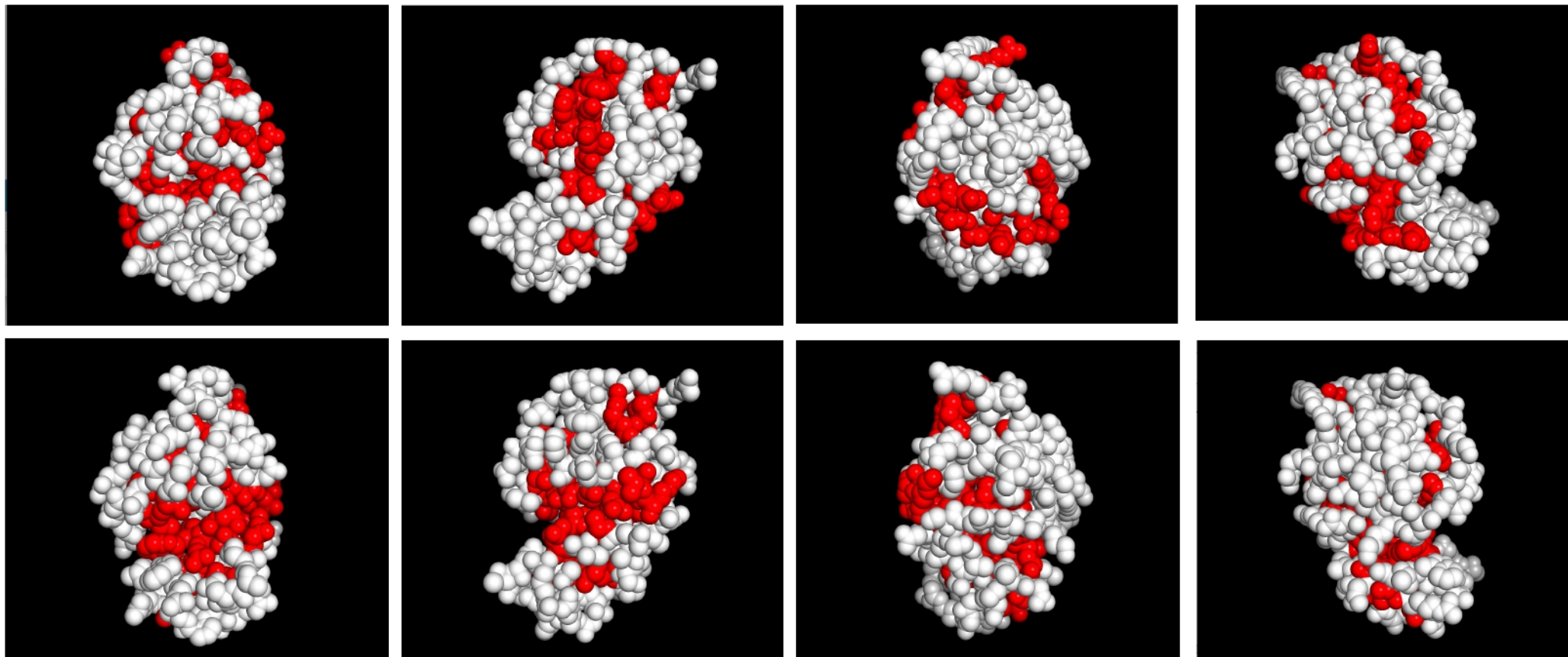
Str Analysis メニューの
Conservation Score Mapping
を選択

PyMolが起動し
指定した立体構造
の上に、スコア
上位のアミノ酸を
赤色で表示



Cluster 1: α lactoalbumin

構造表面に保存残基が分布. β -1,4-galactosyltransferase Iと複合体を形成しlactose synthaseを形成するが、そのインターフェイス部分が検出箇所に含まれる

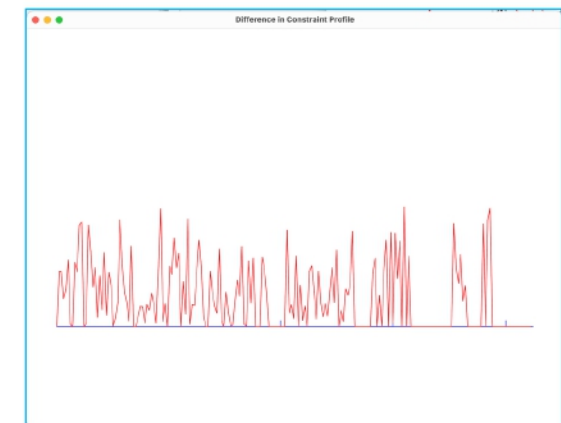
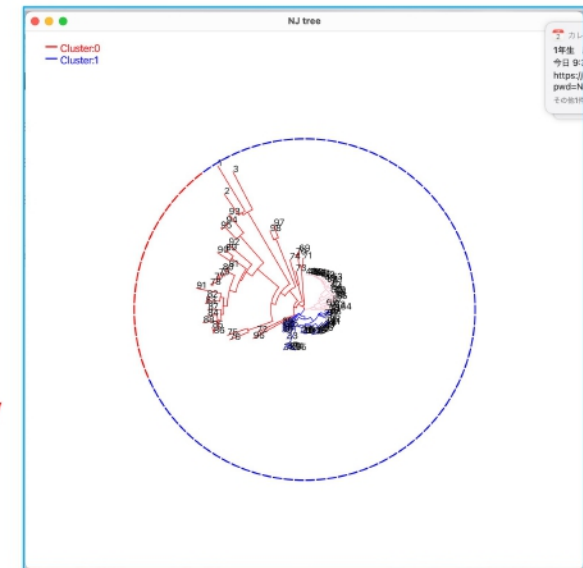
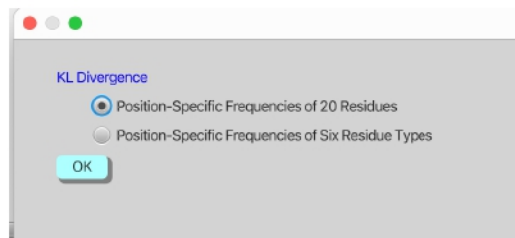
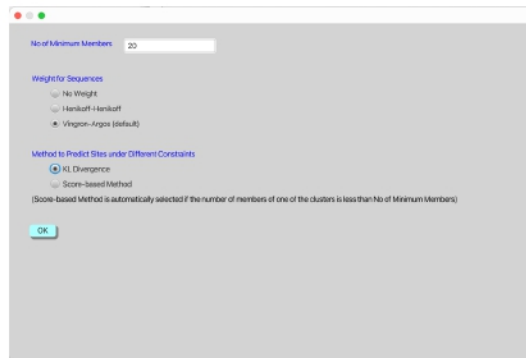
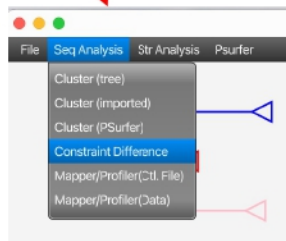
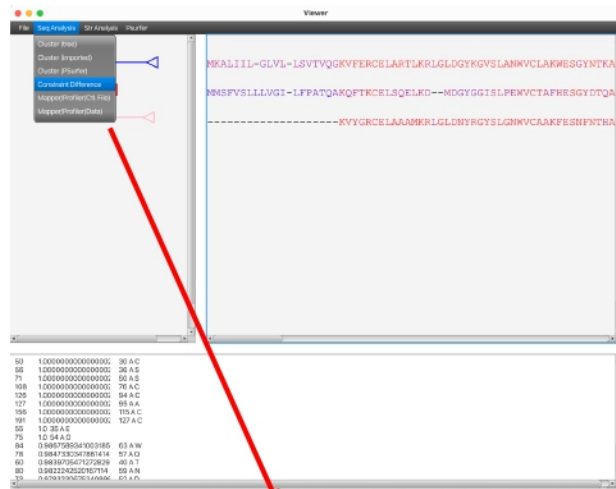


Cluster2: lysozyme

触媒部位であるポケット領域に保存アミノ酸が集中. 左端の図がポケット領域を正面から見たもの

図はY軸周りで90° づつ回転させて作成

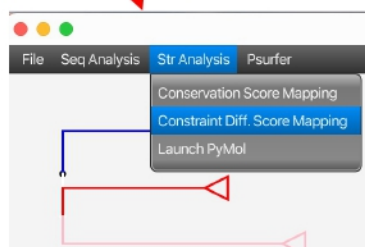
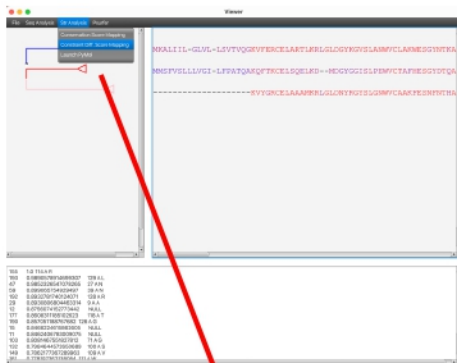
クラスタ間で選択圧の異なるサイトの解析



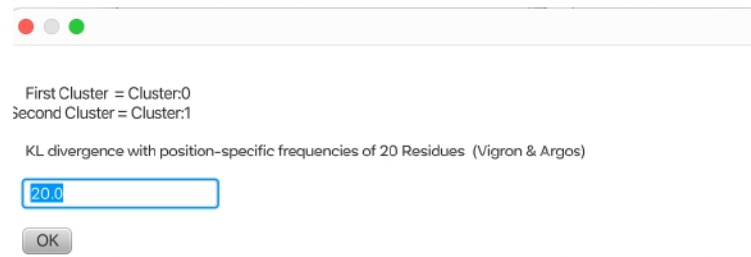
Seq Analysis メニューからConstraint Differenceを選択

計算方法選択のサブウィンドウから方法を選択し、無根系統樹ウィンドウから比較すクラスタ (左上) をクリックして選択. ここではKL情報量を選択.

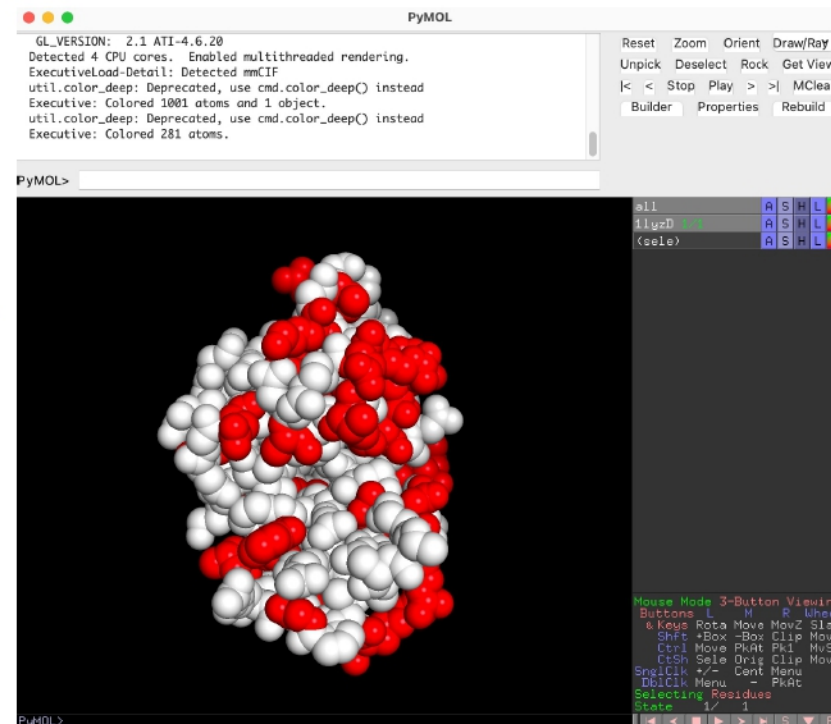
KL情報量のプロファイルが表示される.

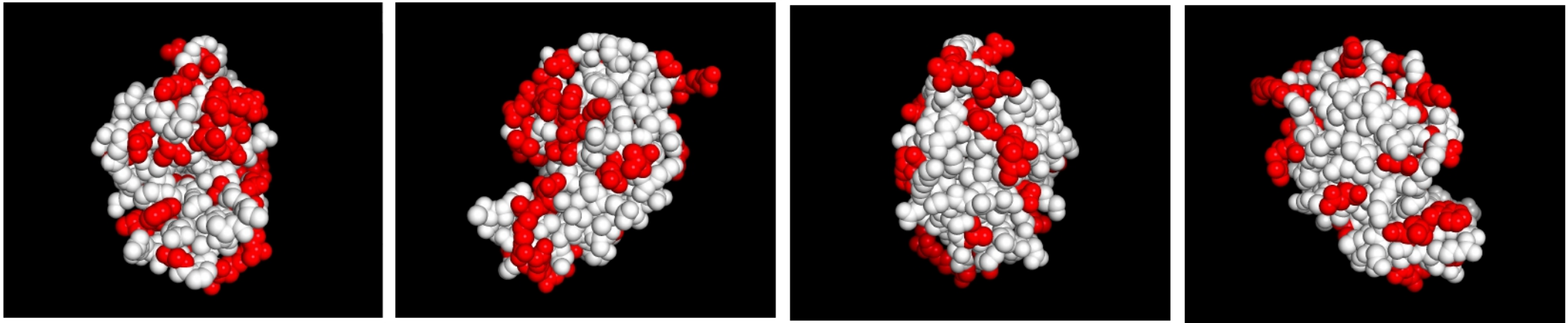


Str Analysisメニューの
Constraint Diff, Score Mapping
を選択



Scoreの上位何%を表示するかを決
PyMolが自動的に立ち上がり、
選択されたアミノ酸をカラーで表示





クラスタ間の比較でトップ20%のKL情報量を持つアミノ酸の立体構造へのマッピング
ポケット領域には少なく、構造表面に多い。インターフェースに対応するものが含まれる

図はY軸周りで90° ずつ回転させて作成

結論：

LOD情報を可視化すると同時に、イベント処理で操作可能にすることで、GO termの違いなどで異なるグループを選択を容易にできるようにした。

これにより、グループ特異的な保存パターンを示すアミノ酸や、グループ間で異なる制約を受けているサイトを検出することが容易になった。