

令和6年度
修士学位論文

博士前期課程 理学専攻
生物物理分野

指導教員 倭 剛久教授

令和4年入学

学籍番号 262301298
氏名 齊藤 詩恋

目次

第 1 章

序論

1.1 はじめに

タンパク質はアミノ酸が多数繋がって構成されている高分子化合物であり、タンパク質全体が分子機械として働く。しかしこのアミノ酸単位やアミノ酸間の相互作用という”部分”としての局所的挙動とドメイン単位やタンパク質という”全体”としての大域的挙動には時空間の大きな隔たりがある。それが顕著に表れている具体的な話でいうと、タンパク質のアロステリー現象が挙げられる。タンパク質のアロステリー現象は、リガンド結合や外部刺激によって生じる構造変化が刺激受容部位から遠隔の活性部位に影響を及ぼす現象であり、そのメカニズム解明は生命現象の理解と創薬研究の中心的課題の一つである。アロステリーはタンパク質の機能を制御する重要な特性であり [1]、リガンド結合など外部刺激のシグナルが残基間相互作用を介して活性部位に伝達する仕組みを提供する。この過程の特徴を 2 点挙げる。

1. 活性部位がサブ Å ~ 数十 Å 離れた場所でのリガンド結合や微小環境の摂動を総じた情報を受け取る点。
2. ピコ秒オーダーの残基間エネルギー移動過程 [2] がミリ秒以上のアロステリック遷移 [3] を引き起こす点。つまりこの過程は、サブピコ秒からミリ

秒の時間スケールにわたるダイナミクスと、サブÅから数十Åの空間スケールの相互作用が連動して行われることが興味深い点であり大きな謎 [4] となっている。この離れた場所間のコミュニケーションを解明する一つの方法として、グラフ理論を用いたアプローチが注目されている。グラフ理論は、分子内の残基間の相互作用をネットワークとして表現し、複雑な動的挙動を解析するための強力なツールとして広く利用されてきた [5][6][7]。タンパク質を構造ネットワークとしてモデル化すると、シグナル伝達機構を解釈しやすくなる。そのようなモデルは、ネットワーク内の最短経路の存在の重要性を強調 [8] しており、それらが離れた部位間での効率的な情報伝達に寄与していることを示している。また、その前提のもとで、高い中心性を持つ残基 [9] やネットワーク上における最短経路上によく現れる残基 [10] を解析することで、機能的残基を同定してきた。実際にこれらの残基はタンパク質の折りたたみにおける重要なアミノ酸や酵素ファミリーの活性部位残基であると関連付けられている。

しかし、残基間エネルギー移動過程とアロステリック遷移の時間スケールの違いを考慮すると、単純な「最短経路モデル」だけではアロステリーの情報伝達を十分に説明できない可能性がある。実際に、残基集団の協調的な運動 [11] や、複数経路の存在 [12] を示す文献もあり、この視点はアロステリーのより現実的で包括的な理解を提供する可能性がある。

本研究では、アロステリーの解析対象として β_2 アドレナリン受容体 (β_2 AR) を選定した。まず分子動力学 (MD) シミュレーションを用いて得られた inactive 状態と active 状態におけるトラジェクトリー解析から、残基間距離を反映したネットワークを構築した。さらに、Louvain 法 [13] によるコミュニティ検出を適用し、コミュニティによるシグナル伝達機構を定量的に解析した。

その結果、active 状態において新たなコミュニティの生成が有意に認められ、これがシグナル伝達において重要な役割を果たすことが示唆された。また、膜タンパク質内のエネルギー的に保存された水分子 [14] が果たす役割を

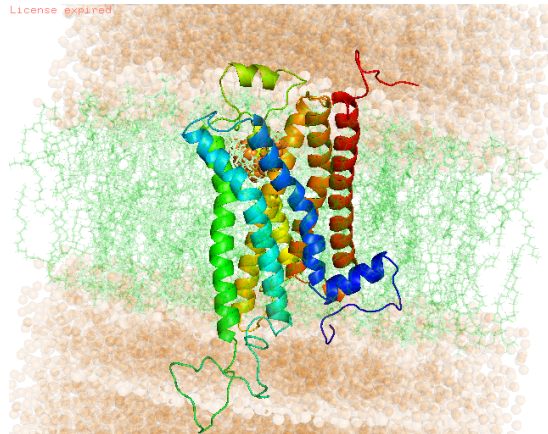
確認した。

本論文の構成は以下の通りである。はじめに、本研究の対象タンパク質である β_2 AR について??節で記述する。次に、分子動力学シミュレーションの方法を??節で、ネットワーク構築とコミュニティ検出の方法を??節で説明する。??章以降は、シミュレーションとネットワーク構築の方法とその結果、考察をまとめる。

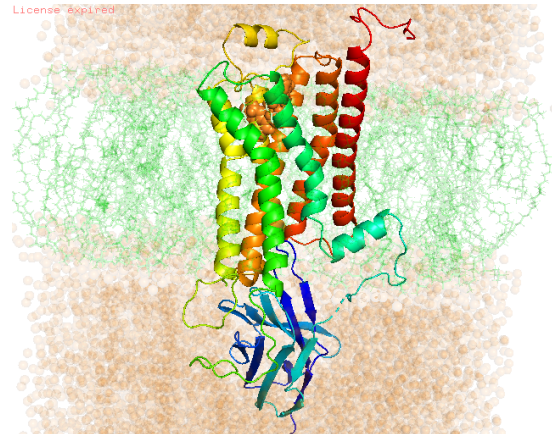
1.2 β_2 アドレナリン受容体 (β_2 AR)

基本情報

β_2 アドレナリン受容体 (β_2 AR) は、G タンパク質共役受容体 (GPCR) の一種である。GPCR は細胞膜に存在する膜タンパク質であり、ホルモンや神経伝達物質などの細胞外刺激を認識し、それを細胞内のシグナルに変換する役割を担っている。また、視覚、嗅覚、味覚といった感覚にも関与し、生体内の多様なシグナル伝達経路において中心的な役割を果たしている。GPCR は 7 回膜貫通構造を持つことが特徴であり、各膜貫通ヘリックス (TM1 TM7) は、細胞外ループ (ECL) と細胞内ループ (ICL) を介して他のヘリックスと連結されている。この構造により、細胞外でのリガンド認識と細胞内でのシグナル伝達を効率的に行うことが可能となる。



(a) β_2 AR の inactive 状態。



(b) β_2 AR の active 状態。

図 1.1: β_2 AR の inactive 状態と active 状態の初期構造。

膜タンパク質の構造データは、Protein Data Bank(PDB) から取得した。 β_2 AR の不活性状態として 2RH1[cherezov2007], 活性状態として 3P0G[rasmussen2011] を用いた。

活性化に伴う構造的変化

β_2 AR の inactive 状態から active 状態への活性化は、リガンド結合部位に結合したアゴニストによって引き起こされる。すると受容体が活性化され、特に G_s といった G タンパク質の結合が促進される。 G_s はヘテロ三量体型タンパク質であり、その活性化により細胞内でアデニル酸シクラーゼが刺激され、cAMP(サイクリック AMP) の産生が増加する。この過程 [philip2007] を通じて、細胞内のシグナル伝達経路が活性化される。

β_2 AR のリガンド結合部位は膜貫通ヘリックス (TM) の間に位置し、細胞外の刺激を感知する。一方で、G タンパク質との結合は細胞内ループおよび細胞質側ドメインを介して行われる。

アゴニストが結合した β_2 AR は、顕著な構造変化を引き起こす [rasmussen2011crystal] ことがわかっている。主な変化として以下が挙げられる：

1. **TM6 の移動:** TM6 の細胞質側末端が、約 14 Å 外側に移動する。この移動により、G タンパク質が結合するための空間が確保される。
2. **TM5 の変化:** TM5 の細胞質側末端部分において、 α ヘリックスの延長が観察される。これにより、シグナル伝達に関与する構造が再編成される。

活性化に関わる重要な残基

β_2 AR の活性化において、特定のモチーフ [nygaard2009ligand][lee2013mapping] が重要な役割を果たしていることが知られている。

モチーフとは、タンパク質中の特定の機能に関連する保存されたアミノ酸配列のことであり、GPCR ではアロステリックシグナル伝達やコンフォメーション変化を介して受容体の機能を調節する。 β_2 AR を含むクラス A GPCR では、以下の 4 つの主要な保存モチーフ (DRY、CWxP、NPxxY、

PIF) と「イオンロック」が重要な役割を果たしている。

表 1.1: 主要な保存モチーフ

モデル名
役割
DRY モチーフ
inactive 状態では TM6 の Glu とイオンロックを形成し構造を安定化させていたのが、
CWxP モチーフ
inactive 状態では Trp がリガンド結合ポケットを開いた状態に維持させていたのが、a
NPxxY モチーフ
Tyr が不活性状態と活性状態で異なる立体配座間を回転し、構造の変化を媒介させる
PIF モチーフ
Phe がスイッチとして機能し、活性化時に異なる立体配座間で方向を変える
イオンロック

また水分子も、生物学的システムにおいて重要な役割を果たすことが知られており、特に GPCR の活性化メカニズムにも深く関与している。ロドプシンをはじめとする GPCR において、保存された水は活性化過程においてアロステリックを仲介する機能を果たすことが示されている [angel2009conserved]。β₂AR においても、膜貫通ドメイン内にはいくつかの保存された水分子が確認され、これらの水分子はアロステリックシグナル伝達に関与していると考えられる。らに、inactive 状態と active 状態それぞれの β₂AR 構造間で水分子の位置や配置がどのように変化するかを比較することで、水の動態が受容体の機能に与える影響を明らかにすることができる可能性がある。

そこで本研究でも、モチーフと保存された結晶水がコミュニティ構造に与える影響を解析し、シグナル伝達機構における役割を明らかにすることを目指す。

第 2 章

材料と方法

2.1 分子動力学シミュレーション

本研究では、大きく分けてモデリング、構造最適化、熱平衡化、サンプリング、解析の 5 つのステップを経た。

2.1.1 モデリング

モデリングでは、欠損部位の補完、結晶水の同定、膜タンパク質の構造準備の 3 つのステップを経た。

欠損部位の補完

PDB に登録されている膜タンパク質の構造データは、全ての残基位置が定まっているわけではない。なぜなら X 線結晶構造解析において、揺らぎが大きい部位は見えないからである。 β_2 AR では、MODELER ソフトウェアを用いて細胞内ループ 3(ICL3)、N 末端、C 末端の欠損部位を補完した。MODELER ではホモロジーモデリングを利用して、既存の PDB データから欠損部位の座標を予測した。

結晶水の同定

膜タンパク質内の結晶水が重要な役割を果たしていることから、DOWSER を用いてエネルギー的に安定な水分子を同定し、構造に含める作業を行った。DOWSER はエネルギー的に安定な水分子の位置を計算し、これを構造データに追加することが可能である。この手順により、機能的に重要な膜タンパク質の構造を構築した。

膜タンパク質の構造準備

初めに、膜タンパク質と脂質膜の複合体構造は、CHARMM-GUI を用いて作成した。複合体構築には置換法 (Replacement Method) を採用し、タンパク質を囲むように脂質二重膜を配置した。以下に CHARMM-GUI の設定手順を示す。

1. 初期設定: CHARMM-GUI サイトで *Protein/Membrane System* を選択し、準備した PDB ファイルをアップロード。
2. 分子選択: リガンドと結晶水を含めるように選択。
3. プロトン化とジスルフィド結合の設定: プロトン化残基やジスルフィド結合を指定。
4. 脂質二重膜の構築: Heterogeneous Lipid を選択し、脂質の XY 軸長を $80 \text{ \AA} \times 80 \text{ \AA}$ に設定。
5. イオン追加: NaCl 濃度を 0.15M に設定し、Na⁺ と Cl⁻ 追加。

分子動力学計算は Amber 力場で行うため、CHARMM-GUI で出力された pdb ファイルを Amber 形式に変換する。

続いて、リガンドの力場構築には、AmberTools の `antechamber` と `parmchk2` ツールを使用した。

続いて、実際にシミュレーションを行うシステムに対して、プロトン化状態の修正、ジスルフィド結合の形成を行う。プロトン化状態の修正に関して

は、タンパク質構造が与えられた環境下で適切なプロトン化状態を取るように、残基の名前を変更する必要がある。本研究では、脂質環境下にある Glu122 は GLH に変更し、膜表面側にある残りの GLu, Asp, Arg, Lys は中性 (pH=7) での状態を採用した。

最後に、タンパク質-リガンド複合体の分子動力学シミュレーションで用いるためのシミュレーション用のファイルを準備した。シミュレーションに使用する力場関数に関して、タンパク質には ff12SB、脂質には lipid21、リガンドには gaff2、水分子には tip3p モデルを用いた。

系のユニットセルサイズは CHARMM-GUI で得られた情報に基づいて、以下のように指定する。

表 2.1: シミュレーションに用いた系の大きさ

モデル名	原子数	ボックスの大きさ (x, y, z)
inactive(3P0G)	84225	(80.2377, 80.2377, 113.064)
active(2RH1)	70608	(80.0239, 80.0239, 136.043)

2.1.2 構造最適化

構造最適化は 3 つのステップに分けて行った。全てのステップにおいて、最急降下法で最小化を開始し、200 ステップ後に共役勾配法 (CG 法) へ切り替えている。初めに、水素以外の全ての原子に位置制約を課し、タンパク質周辺の脂質や水分子の構造最適化を行った。続いて、膜タンパク質の主鎖原子、リガンドの重原子、脂質の head 部分に位置制約を課し、膜タンパク質の側鎖原子や水分子の構造最適化を行った。最後に、上記の束縛力を弱めて、全体の構造を最適化した。

2.1.3 熱平衡化

熱平衡化は4つのステップに分けて行った。初めに、NVT アンサンブルで系を徐々に加熱させた。初期速度は温度 $T = 0[\text{K}]$ のマクスウェル分布に従って与え、100 ps で温度を $T = 310[\text{K}]$ まで上昇させた。タンパク質とリガンドに対して $2.0 \text{ kcal/mol}/\text{\AA}^2$ の拘束を加えた。続いて、NVT 条件下で 200 ps のシミュレーション $= 310 [\text{K}]$ で設定した。タンパク質とリガンドの位置拘束の重みを $0.1 \text{ kcal/mol}/\text{\AA}^2$ に設定 $= 310 [\text{K}]$ で設定した。位置拘束を外し、膜系のシミュレーション用に表面張力 $\gamma = 17 \text{ dyne/cm}$ で設定した。

2.1.4 サンプルング

ここでは $\text{NP}\gamma\text{T}$ シミュレーションによるサンプルングを 56 ns 行った。設定は、熱平衡化で行った $\text{NP}\gamma\text{T}$ シミュレーションと同様である。

2.1.5 保存された結晶水の同定

保存された結晶水を同定するために、サンプルングで得られたシミュレーションデータから水分子の位置とその密度を分析した。具体的には、シミュレーションの各タイムステップで得られる水分子の座標を格子状に集約し、サンプルングで得られたトラジェクトリの半分以上で保存されている水分子を同定した。

2.1.6 NVE

最後に NVE シミュレーションによるサンプルングを行った。初期状態として、 $\text{NP}\gamma\text{T}$ シミュレーションによるサンプルングで得られた座標と速度を用いた。保存した座標と速度のうち、2 つを使って $\Delta t = 0.5 \text{ fs}$ 、1000 ps

のシミュレーションを実行した。最終的に、inactive 状態 active 状態双方で 1000ps トラジェクトリを計 10 本取得した。

2.2 グラフ理論

2.2.1 ネットワーク構築

グラフ理論では、タンパク質のアミノ酸残基の相互作用を、ノードとエッジを用いたアミノ酸ネットワークとして表すことができる。

- ノード：アミノ酸残基（原子やタンパク質全体も可能）
- エッジ：ノード同士の相互作用

そしてエッジに重みづけをすることで、エッジを同じ距離 1 とするのではなく、各結合ごとに強度を振り分けることが可能である。

本研究では、ノードをアミノ酸残基とし、以下の条件に基づいてエッジを形成した。1. 残基ペア間の最短距離が 3 Å 未満である場合、エッジを形成する。2. アミノ酸配列上で隣接する残基間のエッジは削除する。ここで、残基ペア間の最短距離は、NVE トラジェクトリの全フレームの座標を時間平均することで得られた平均構造を用いて計算したものである。隣接する残基間のエッジを削除した理由は、本研究では inactive 状態から active 状態へのアロステリック転移に寄与する、より長距離で特異的な native contact に焦点を当てるためである。

エッジの重みづけとして、残基間最短距離の 2 乗逆数の平均 $\langle \frac{1}{d^2} \rangle$ を用いた。これは、エネルギー移動速度は残基間距離の 2 乗平均に反比例し、距離が短いほど速い。以下がネットワークの詳細である。

表 2.2: ネットワーク上のノード数と、残基ペア間の最短距離が 3 Å 未満のエッジ数

モデル名	ノード数	エッジ数
inactive(3P0G)	351	16336
active(2RH1)	349	15433

2.2.2 Louvain 法によるコミュニティ検出

Louvain 法は、ネットワークにおけるコミュニティ構造を検出するための効率的なアルゴリズムである。本研究では、Louvain 法 [13] を実装してネットワークデータを解析し、最適なコミュニティ分割とモジュラリティの評価を行った。

モジュラリティの定義

モジュラリティ Q は、ネットワーク内で検出されたコミュニティ構造の質を評価する尺度である。具体的には、同一コミュニティ内に存在するエッジの密度が、ランダムに生成されたネットワークと比較してどれだけ顕著に密集しているかを示す。モジュラリティは次式で定義される：

$$Q = \sum_{c \in \text{コミュニティ}} \left(\frac{m_c}{m} - \left(\frac{K_c}{2m} \right)^2 \right), \quad (2.1)$$

ここで、 m_c はコミュニティ c 内のエッジの重み合計、 m はネットワーク全体のエッジの重み合計、 K_c はコミュニティ c 内のノードの次数の合計である。この定義により、同一コミュニティ内でのエッジ密度が高い場合にモジュラリティの値が大きくなり、ネットワークのコミュニティ構造がより明確であることを示す。

アルゴリズムの流れ

Louvain 法は以下の 2 つのステップを繰り返してコミュニティを検出する：

1. **局所移動ステップ：** 各ノードを隣接するコミュニティに移動させ、モジュラリティの増加が最大となる配置を探索する。このプロセスにおいて、モジュラリティの変化量 ΔQ は以下で計算される：

$$\Delta Q = \Delta Q_{\text{削除}} + \Delta Q_{\text{挿入}}, \quad (2.2)$$

ここで、 $\Delta Q_{\text{削除}}$ はノードを元のコミュニティから削除した際のモジュラリティの変化量、 $\Delta Q_{\text{挿入}}$ はノードを新たなコミュニティに追加した際のモジュラリティの変化量である。

2. **集約ステップ：** 検出されたコミュニティを 1 つのノードとして扱い、新しいネットワークを構築する。この操作を通じて、階層的なコミュニティ構造を得ることが可能である。

本研究では、Python を用いて Louvain 法を実装した。モジュラリティの増加が小さい場合に収束するよう、収束条件として閾値 $\epsilon = 10^{-7}$ を設定した。また、ノードの移動においてランダム性を導入し、複数回の試行によりモジュラリティの最適解を探索した。そして、100 回の試行の中で最大のモジュラリティおよびその標準偏差と標準誤差を算出した。

第 3 章

結果と考察

3.1 分子動力学 (MD) シミュレーション

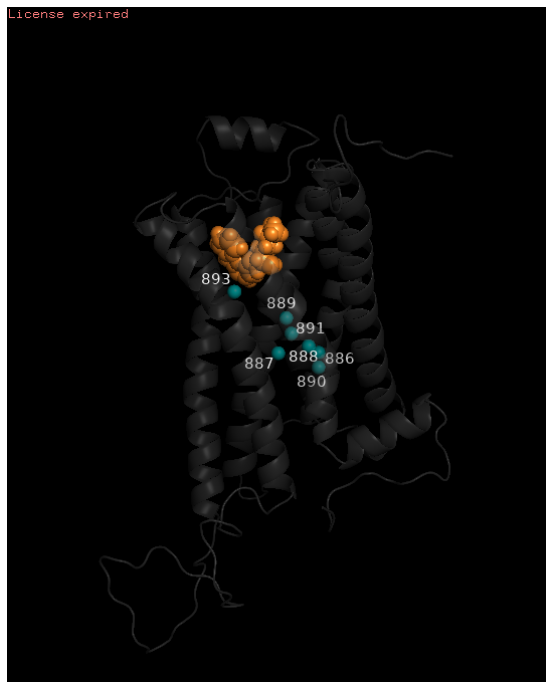
β_2 AR の inactive 状態 active 状態双方で 1000ps トラジェクトリを計 10 本取得した。また保存された結晶水は inactive 状態で 7 個、active 状態で 5 個同定され、その位置は図??に示す。

3.2 β_2 AR の inactive および active 状態のコミュニティ検出

Louvain 法の信頼性

Louvain 法の結果の信頼性は、モジュラリティ Q の値を用いて評価される。モジュラリティ Q の値は通常-1 から 1 の範囲を取り、以下のように解釈される：

- Q が負：分割がネットワーク構造と一致しておらず、不適切な分割である。
- Q が 0 に近い：ネットワークがランダム構造に近い。



(a) β_2 AR の inactive 状態における保存された結晶水。



(b) β_2 AR の active 状態における保存された結晶水。

図 3.1: 保存された結晶水

- Q が正：ネットワーク内にコミュニティ構造が存在する。

本研究の解析対象である β_2 AR の inactive および active 状態のコミュニティ検出におけるモジュラリティ Q の最良値および標準偏差を計算した。

■ 試行回数 100 回におけるモジュラリティ値の平均および標準偏差

- inactive 状態： $Q = 0.5172 \pm 0.0059$
- active 状態： $Q = 0.5239 \pm 0.0044$

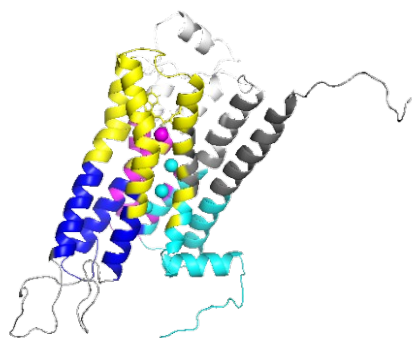
さらに、モジュラリティ値の標準誤差は以下の通りです：

- inactive 状態： ± 0.0006
- active 状態： ± 0.0004

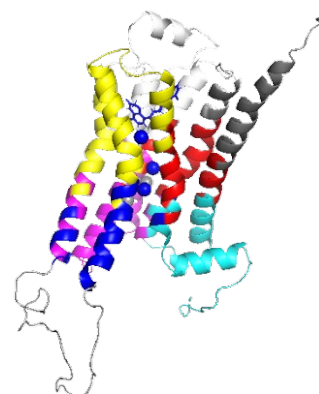
inactive 状態と active 状態のモジュラリティ値はどちらも正の値を示しており、ネットワーク内に明確に分割されたコミュニティ構造があることを示唆している。

3.2.1 Louvain 法で検出されたコミュニティ

図??には β_2 AR の inactive 構造で検出されたコミュニティ、図??には active 構造で検出されたコミュニティを示す。



(a) β_2 AR の inactive 状態において検出されたコミュニティ構造を色分けして示した図。各色と名付けたコミュニティは以下に表す：黒 (start)、シアン (end)、白 (helix)、ピンク (back)、グレー (loop)、黄色 (ligand)、青 (gprotein)。



(b) β_2 AR の active 状態において検出されたコミュニティ構造を色分けして示した図。各色と名付けたコミュニティは以下に表す：黒 (start)、シアン (end)、白 (helix)、ピンク (back)、グレー (loop)、黄色 (ligand)、青 (gprotein)、赤 (new)。

図 3.2: コミュニティ構造

まず双方のコミュニティに共通することとして、リガンド結合部位と活性部位である G タンパク質結合部位に対応する独立したコミュニティが、黄色 (ligand)、青 (gprotein) として検出された。また図??から図??への変化として、青 (gprotein) コミュニティの再編成が起きていることと、赤 (new)

コミュニティが新規に形成されたことが挙げられる。

青 (gprotein) コミュニティと赤 (new) コミュニティに属する残基の比較を行うと、以下のようなになる。

active 状態になると、青 (gprotein) コミュニティにはリガンドと保存された結晶水 3 つ、赤 (new) コミュニティにはモチーフが追加された。

ここで、図??と図??で示されているネットワークの全体密度を計算する。全体エッジ密度 D_{global} の計算式は以下のように表される。

$$W_{\text{actual}} = \sum_{(u,v) \in E} w(u,v)$$

W_{actual} は実際のエッジ重みの合計である。E は、グラフ内のエッジ集合を示す。 $w(u, v)$ は、エッジ (u, v) の重みを示す。

$$W_{\text{max}} = \sum_{i=1}^N \sum_{j=i+1}^N w(i, j)$$

W_{max} は理論上可能な最大エッジ重みの合計である。N は、グラフ内のノード数を示す。 $w(i, j)$ は、ノード i と j の間のエッジ重みを示す。

$$D_{\text{global}} = \frac{W_{\text{actual}}}{W_{\text{max}}}$$

最終的に、ネットワークの全体エッジ密度 D_{global} は以上のように示される。

この式に基づいて、inactive 構造と active 構造のネットワークの全体エッジ密度を計算すると、それぞれ以下のようなになる。

- inactive 状態 : $Q = 0.3083$
- active 状態 : $Q = 0.8579$

inactive 状態は低い密度を示しており、最短距離が 3 Å 未満の残基ペアが相対的に少なかった。active 状態は、inactive 状態に比べて高い全体エッジ

密度を持っており、より多くの相互作用が存在することを示している。この結果は、活性状態におけるタンパク質の構造や機能がより密接に結びついていることを示唆している。

3.2.2 コミュニティ内およびコミュニティ間のエッジ密度

active 状態で再編成されたコミュニティや新しく検出されたコミュニティの役割を定量的に分析するために、ネットワークの全体密度 D_{global} の計算で用いた密度の概念を用いたさらなる計算を行った。続いて、inactive 状態と active 状態双方においてそれぞれコミュニティ内のエッジ密度、コミュニティ間のエッジ密度を計算した。

コミュニティ内エッジ密度

コミュニティ内エッジ密度 D_{internal} の計算式は以下のように表される。

まずコミュニティごとにサブグラフを作成する。ただしコミュニティ間のエッジは削除し、独立したコミュニティを表現している。続いてコミュニティ内エッジ密度を計算する。

$$D_{\text{internal}} = \frac{\sum_{(u,v) \in E_C} w_{uv}}{\sum_{(u,v) \in E_C} w_{uv}^{\text{max}}}$$

ここで E_C はコミュニティ C 内の全エッジの集合、 w_{uv} はノード u とノード v の間の実際のエッジ重み、 w_{uv}^{max} はノード u とノード v の間のエッジの最大可能重みである。分子の $\sum_{(u,v) \in E_C} w_{uv}$ はコミュニティに属するノード間に存在する実際のエッジの重みの合計を示している。分母の $\sum_{(u,v) \in E_C} w_{uv}^{\text{max}}$ は、そのコミュニティ内の全てのノードが完全に接続している場合のエッジの最大可能重みの合計を示している。ただし、自己ループは除く。

コミュニティ内エッジ密度 D_{internal} は、ネットワーク全体のエッジ密度

を基準としてコミュニティ内のエッジ密度がネットワーク全体と比べて相対的にどれだけ高い (または低い) かを示し、コミュニティ内部でどれだけ密接に関連しているかを測定することを目的としている。上記の式に基づいて、inactive 構造と active 構造のコミュニティ内エッジ密度を計算すると、それぞれ以下ようになる。

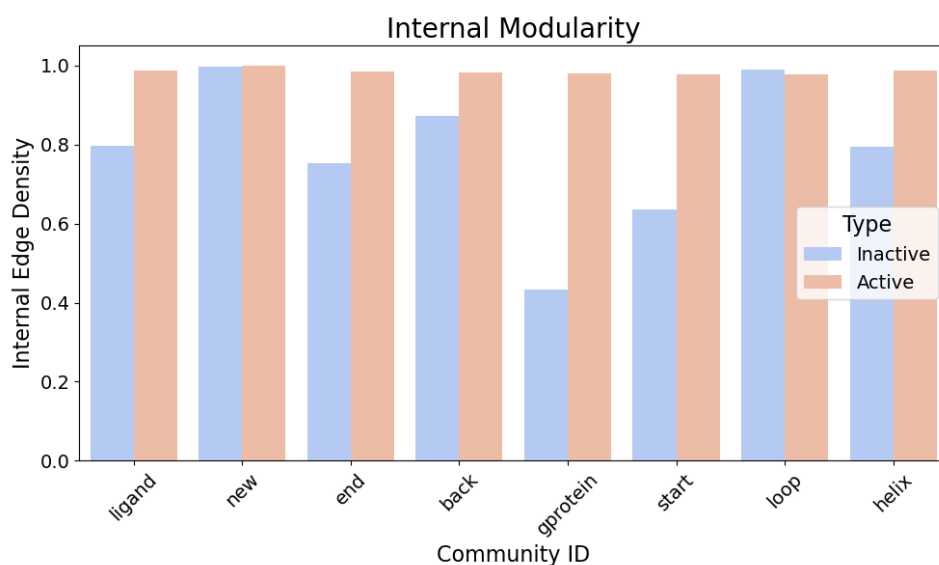


図 3.3: β_2 AR の inactive 状態と active 状態において検出されたコミュニティのコミュニティ内エッジ密度を示した図。コミュニティの名称は、図??と図??で名付けたものである。

コミュニティ内エッジ密度 D_{internal} の値は以下のように解釈される：

- D_{internal} が 0 に近い：コミュニティ内のノード間の相互作用は希薄であり、結びつきが弱い。
- D_{internal} が 1 に近い：コミュニティ内のノード間には強い結びつきがあり、密に相互作用している。

inactive 構造では、loop, new コミュニティのコミュニティ内エッジ密度が特に大きい値を示した。一方で active 構造においては、全てのコミュニ

ティのコミュニティ内エッジ密度が 1.0 付近という値を示し、特に強い結びつきを形成している構造であることがわかる。inactive 構造から active 構造への変化に着目すると、どのコミュニティも値が大幅に大きくなり、特に gprotein コミュニティが顕著であった。

コミュニティ間エッジ密度

コミュニティ間エッジ密度 D_{inter} の計算式は以下のように表される。

$$D_{\text{inter}} = \frac{\sum_{\substack{u \in C_i \\ v \in C_j}} w_{uv}}{\sum_{\substack{u \in C_i \\ v \in C_j}} w_{uv}^{\text{max}}}$$

ここで C_i と C_j はそれぞれ異なるコミュニティ i と j 、 w_{uv} はノード u とノード v の間の実際のエッジ重み、 w_{uv}^{max} はノード u とノード v の間のエッジの最大可能重みである。分子の $\sum_{\substack{u \in C_i \\ v \in C_j}} w_{uv}$ は異なるコミュニティに属するノード間の実際の

コミュニティ間エッジ密度は、異なるコミュニティ間のエッジの結びつきの強さを示しており、お互いにどれだけ相互作用が強いかを測定することを目的としている。上記の式に基づいて、inactive 構造と active 構造のコミュニティ間エッジ密度を計算すると、それぞれ以下ようになる。

ただし、inactive 状態と active 状態の双方でコミュニティ間エッジ密度が 0 だったコミュニティペアは、図から除外している。該当するコミュニティペアは start-loop、helix-loop、new-loop である。

コミュニティ間エッジ密度 Q の値は以下のように解釈される：

- Q が 0 に近い：コミュニティ間でエッジが少なく、各コミュニティがほぼ独立している。
- Q が 1 に近い：コミュニティ間で多数のエッジが存在し、コミュニティ間の結びつきが強い。

inactive 状態では、コミュニティペア back-gprotein、helix-start、end-

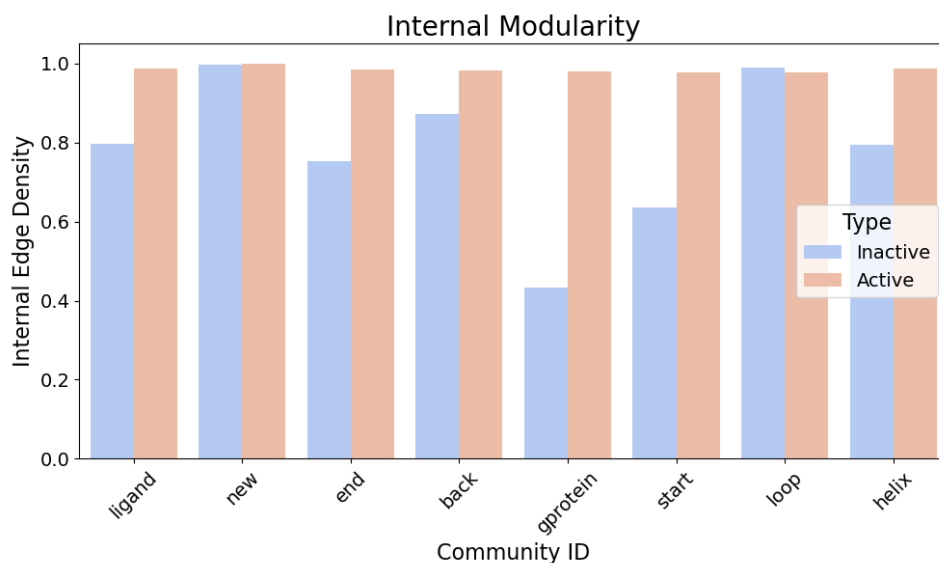


図 3.4: β_2 AR の inactive 状態と active 状態において検出されたコミュニティのコミュニティ間エッジ密度を示した図。コミュニティの名称は、図??と図??で名付けたものである。

back、start-ligand が高い値を示した。これらのコミュニティペアは隣り合っており、妥当な結果を示している。active 状態では、リガンド結合部位と活性部位の間に存在するコミュニティ back, gprotein, ligand, new を含むコミュニティペアの値が特に顕著に高い値を示した。特に new コミュニティとペアを形成しているコミュニティペアに関しては、全て 0.6 以上と高い値を示した。inactive 構造から active 構造への変化に着目すると、リガンド結合部位と活性部位を結ぶ経路から離れているコミュニティペア (end,start) 以外のどのコミュニティも、値が大幅に大きくなった。

inactive 状態と active 状態の全体密度、コミュニティ密度の考察

■密度に関わる 3 つの変数における、inactive 状態から active 状態への変化
 ここまで計算したコミュニティ内エッジ密度、コミュニティ間エッジ密度、全体エッジ密度の inactive 状態から active 状態への変化を示したのが以下

の表である。

- コミュニティ内エッジ密度：増加
- コミュニティ間エッジ密度：増加
- 全体エッジ密度：増加

inactive 状態では、loop コミュニティや new コミュニティにおいて相互作用がコミュニティ内部に局在化しており、gprotein コミュニティは内部の結びつきが弱かった。またコミュニティ間のエッジ密度が低いため、各コミュニティの協調的な動きが抑制されている状態が観察された。一方で active 状態になると、ほぼすべてのコミュニティ内の相互作用が増加し、コミュニティ間の相互作用も大幅に増加し、その結果全体密度も増加する状態が観察された。特に new コミュニティに関わるコミュニティペア間の相互作用の強さが特徴的であった。まとめると、active 状態になるとコミュニティ内部と間の相互作用が増加するようにコミュニティの再編成が起き、結果的に全体の協調的な動きが活性化している様子が見てとれる。これらの結果から、分子全体の協調的な相互作用が増加し情報伝達の効率が向上するアロステリック遷移のメカニズムが示唆される。またこの遷移において、active で新たに形成された new コミュニティが、リガンド結合部位や活性部位である G タンパク質結合部位間の情報伝達を促進する導管として働いている可能性がある。

3.3 ノード削除による active ネットワーク接続性への影響

あるノードが active ネットワーク全体または所属するコミュニティの接続性に与える影響を定量化に評価するために、impact score という変数を導入した。これは、特定のノードを active ネットワークから削除した際に生

じる全体エッジ密度とコミュニティ内エッジ密度の変化量を基に計算した。全体エッジ密度とコミュニティ内エッジ密度はそれぞれ前述の式 (式 ??) と式 (式 ??) に従っている。以降、全体エッジ密度の変化量を global impact score、コミュニティ内エッジ密度の変化量を community impact score と表現する。

また、global impact score と community impact score について、Z スコアを算出することで影響の大小を統計的に評価する。Z スコアは、あるデータ点で平均からどれだけ標準偏差の単位で離れているかを示す指標である。Z スコアの計算式は以下のとおりである。

$$z = \frac{x - \mu}{\sigma}$$

ここで x はデータ点の値、 μ はデータセットの平均、 σ はデータセットの標準偏差を示している。Z スコアの解釈は以下のとおりである。

- $z < 0$: データ点は平均よりも小さい値である。
- $z = 0$: データ点は平均と一致している。
- $z > 0$: データ点は平均よりも大きい値である。
- $z > 2$: データ点は平均から 2 標準偏差以上離れており、全体の約 2.5% にあたるくらい高い値である。

global impact score

global impact score の変化量が大きかったノードのうち、Z スコアが 2 以上であったノードを昇順に並び替えたのが以下の表である。

ここで Community の名称は、図??と図??で名付けたものである。Node type は以下の

上位 10 このうち 7 このノードが new コミュニティに属しており、active 構造で検出された新しいコミュニティが重要な役割を果たしていることが示

Node	Global Impact Score	Community	Node type
303	0.000775	new	Ligand-Site
62	0.000773	new	Other
66	0.000766	new	Motif
273	0.000753	ligand	Motif
111	0.000747	new	Other
96	0.000740	back	Ligand-Site
69	0.000739	new	Motif
107	0.000724	new	Motif
104	0.000683	new	Ligand-Site
108	0.000678	ligand	Other
58	0.000669	helix	Other
100	0.000647	new	Ligand-Site
105	0.000640	back	Ligand-Site
110	0.000635	helix	Other
339	0.000635	end	Other
114	0.000626	helix	Other
186	0.000622	back	Ligand-Site
342	0.000622	end	Other

表 3.1: Top 18 Nodes by Impact Score

唆された。また、リガンド結合部位や new コミュニティに属しているモチーフに関連するノードが多く、これらの領域が全体のネットワーク密度に強い影響を与えていることが示唆される。

community impact score

community impact score が大きかったノードのうち、Z スコアが 2 以上であったノードを昇順に並び替えたのが以下の表である。

22 このうち 18 このノードが gprotein コミュニティもしくは loop コミュニティに属しており、active 状態において再編成を起こした 2 つのコミュニティ内の重要な結束性に寄与している可能性がある。また、上位 4 つを保存された結晶水が占めていることから、これらが g タンパク質結合領域内で強い影響力を持つことを示している。

global impact score、community impact score の考察

global impact score が目立ったりリガンド結合部位やモチーフに関わる残基はネットワーク全体の構造を保つ役割を果たしている。つまり、リガンド結合部位やモチーフはネットワーク全体の「骨格」として、リガンド結合によるネットワーク全体の構造的再編成を引き起こし、アロステリックな影響を広げる重要な起点となっていることが示唆される。community impact score が目立ったりリガンドや g タンパク質結合部位、保存された結晶水に関わる残基は局所的なコミュニティ構造を保つ役割を果たしている。つまり、リガンドや g タンパク質結合部位、保存された結晶水は局所的な「柔軟性」を提供し、G タンパク質の結合やシグナル伝達の効率化を高めていることが示唆される。

3.4 シグナル伝達機構の考察

ここまで inactive 状態と active 状態におけるコミュニティ内密度とコミュニティ間密度、active 状態のネットワークにおけるノードが active ネットワーク全体または所属するコミュニティの接続性に与える影響について解析をしてきた。これらの解析結果を踏まえると、以下のような段階的なシグ

ナル伝達機構が考えられる。

まずリガンドがリガンド結合部位に結合すると、まず結合近傍の局所的なコミュニティ再編成が起きる。この再編成はリガンドが高い community impact score を示したことに反映されている。続いて再編成されたりガンド結合部位が、局所的な変化を他のコミュニティへと伝播させることで、ネットワーク全体の構造変化が促進される。この過程は、ligand コミュニティと他のコミュニティにおけるコミュニティ間密度が大幅に増加したことや、リガンド結合部位が高い global impact score を示したことに反映されている。最後に、再編成されたネットワーク全体が g タンパク質結合部位や保存された結晶水などの局所的な結束性によって安定化され、シグナル伝達が効率的に進む。この過程は、gprotein コミュニティのコミュニティ内密度が大幅に増加したことや、gprotein コミュニティに含まれる保存された結晶水が高い community impact score を示したことに反映されている。

Node	Community Impact Score	Community	Node type
346	0.002723	gprotein	X-Water
347	0.002642	gprotein	X-Water
349	0.002634	gprotein	X-Water
348	0.002609	gprotein	X-Water
343	0.002576	gprotein	Ligand
241	0.002372	loop	Other
1	0.002248	start	Other
258	0.002208	gprotein	Gprotein-Site
231	0.002121	loop	Other
259	0.002074	gprotein	Other
243	0.002047	loop	Other
236	0.001998	loop	Other
17	0.001990	start	Other
242	0.001974	loop	Other
345	0.001969	gprotein	X-Water
239	0.001962	loop	Other
2	0.001876	start	Other
222	0.001826	loop	Other
255	0.001815	gprotein	Motif
339	0.001767	end	Other
217	0.001761	loop	Gprotein-Site
261	0.001746	gprotein	Gprotein-Site

表 3.2: Top 22 Nodes by Impact Score

第 4 章

まとめ

本研究では、 $\beta 2AR$ の inactive 状態および active 状態におけるネットワーク構造を解析し、Louvain 法を用いて検出されたコミュニティの特性を比較した。まず検出されたコミュニティを比較すると、g タンパク質結合部位が再編成され、新規のコミュニティが形成されていることが明らかになった。続いてそれぞれのネットワークの全体エッジ密度、コミュニティ内エッジ密度、コミュニティ間エッジ密度を比較した。全体エッジ密度に関しては、inactive 状態では 0.3083、active 状態では 0.8579 であった。active 状態のネットワークの方がより高い相互作用密度を示し、構造的および機能的に密接な結びつきが形成されていることが確認された。全てのコミュニティでコミュニティ内エッジ密度が向上し、特に gprotein コミュニティが顕著に増加した。コミュニティ間エッジ密度の解析では、リガンド結合部位や活性部位に関連するコミュニティ間の相互作用が強化されており、活性状態における重要な相互作用ネットワークが形成されていることが分かった。最後にノード削除が全体エッジ密度とコミュニティ内エッジ密度に与える影響の解析では、リガンド結合部位や new コミュニティに属しているモチーフに関するノードが全体エッジ密度に、リガンドや g タンパク質結合部位、保存された結晶水がコミュニティ内エッジ密度に与える影響が大きかった。前者

はネットワーク全体の「骨格」としてアロステリックな影響を広げる重要な起点となっていることが、後者は局所的な「柔軟性」を提供し G タンパク質の結合やシグナル伝達の効率化を高めていることが示唆された。

4.1 今後の展望

本研究ではノード間の距離を重みとした構造ネットワークを用いた。しかし活性化によって相互作用のみが変化したノードに関しては、構造ネットワークではその変化を捉えることが困難である。そのため、構造のみならず、ダイナミクスや相互作用も反映した変数である熱伝導度を重みとした物理的な熱ネットワークの構築により、 $\beta 2AR$ のアロステリー機構の解明をより詳細に理解することが期待できる。また、Louvain 法によるコミュニティ検出では、特定の時間スケールでの 1 つのコミュニティ分割しか検出しておらず、異なる時間スケールでの過渡現象を観察できない。そのため、マルコフ安定性のようなネットワーク内のさまざまなスケールに存在する多層コミュニティ構造を識別できる動力学ベースのマルチスケール方の導入が必要である。

謝辞

学部3年次の生物物理セミナーからはじめ、3年間の研究の遂行にあたり、指導教官として終始多大なご指導を賜った、倭剛久先生先生に深謝致します。研究室選びに迷っている中で学部3年次に出席した生物物理分野に興味を持ち、急遽研究室見学を受け入れてくださった時から今まで、生物物理に関する深い知識と経験に基づく多くの有益な助言をいただきました。木村明洋先生にも、講義やセミナー、研究室での議論を通じて、多くのご指導をいただきました。またB研の皆様には、本研究の遂行にあたり多大なご助言、ご協力頂きました。特に王婷婷氏、杉浦航氏、在田陽一氏には、右も左もわからなかった自分に、生物物理やシミュレーションの基礎をはじめ、私の疑問や悩みをたくさん解決していただきました。大久保雄大氏、吉村風汰氏、大櫃丈氏には、研究室内発表時の議論を通じて多くの刺激をもらいました。

最後に、研究ができる環境を提供していただき、精神的な支えにもなっていただいた両親や祖父、姉に深く感謝いたします。

参考文献

- ¹Q. Cui and M. Karplus, “Allostery and cooperativity revisited”, *Protein Sci.* **17**, 1295–1307 (2009).
- ²M. Lim, T. A. Jackson, and P. A. Anfinrud, “Femtosecond near-ir absorbance study of photoexcited myoglobin: dynamics of electronic and thermal relaxation”, *J. Phys. Chem.* **100**, 12043–12051 (1996).
- ³J.-P. Changeux and S. J. Edelstein, “Allosteric mechanisms of signal transduction”, *Science* **308**, 1424–1428 (2005).
- ⁴A. W. Fenton, “Allostery: an illustrated definition for the ‘second secret of life’ ”, *Trends Biochem. Sci.* **33**, 420–425 (2008).
- ⁵N. T. Doncheva, K. Klein, F. S. Domingues, and M. Albrecht, “Analyzing and visualizing residue networks of protein structures”, *Trends Biochem. Sci.* **36**, 179–182 (2011).
- ⁶A. J. M. Martin, M. Vidotto, F. Boscariol, T. D. Domenico, I. Walsh, and S. C. E. Tosatto, “Ring: networking interacting residues, evolutionary information and energetics in protein structures”, *Bioinformatics* **27**, 2003–2005 (2011).
- ⁷N. T. Doncheva, Y. Assenov, F. S. Domingues, and M. Albrecht, “Topological analysis and interactive visualization of biological networks and protein structures”, *Nat. Protoc.* **7**, 670–685 (2012).

- ⁸A. Ghosh and S. Vishveshwara, “A study of communication pathways in methionyl-trna synthetase by molecular dynamics simulations and structure network analysis”, *Proc. Natl. Acad. Sci.* **104**, 15711–15716 (2007).
- ⁹G. Amitai, A. Shemesh, E. Sitbon, M. Shklar, D. Netanel, I. Venger, and S. Pietrokovski, “Network analysis of protein structures identifies functional residues”, *J. Mol. Biol.* **344**, 1135–1146 (2004).
- ¹⁰A. Del Sol, H. Fujihashi, D. Amoros, and R. Nussinov, “Residues crucial for maintaining short paths in network communication mediate signaling in proteins”, *Mol. Syst. Biol.* **2** (2006).
- ¹¹A. P. Kornev and S. S. Taylor, “Dynamics-driven allostery in protein kinases”, *Trends Biochem. Sci.* **40**, 628–647 (2015).
- ¹²A. del Sol, C.-J. Tsai, B. Ma, and R. Nussinov, “The origin of allosteric functional modulation: multiple pre-existing pathways”, *Sci. Direct* **17**, 1042–1050 (2009).
- ¹³V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks”, *J. Stat. Mech.* **2008**, P10008 (2008).
- ¹⁴T. E. Angel, M. R. Chance, and K. Palczewski, “Conserved waters mediate structural and functional activation of family a (rhodopsin-like) g protein-coupled receptors”, *Proc. Natl. Acad. Sci.* **106**, 8555–8560 (2009).
- ¹⁵S. Buchenberg, D. M. Leitner, and G. Stock, “Scaling rules for vibrational energy transport in globular proteins”, *J. Phys. Chem. Lett.* **7**, 25–30 (2015).