

論文の内容の要旨

論文題目 Solute Carrier (SLC) 輸送体の遺伝子変異および
立体構造/機能変異の計算科学による接続

氏 名 樋口 明子

膜輸送体は細胞膜に発現し、細胞や組織に必要な栄養素を輸送するために重要な役割を果たす。Solute Carrier (SLC) 輸送体は、ヒトの膜輸送体の中でも遺伝子数が最も多いスーパーファミリーであり、現在までに、52ファミリーから成り立つ400以上のヒトSLC遺伝子が同定されている。SLC輸送体は幅広い基質を輸送することができる。各ファミリーは、アミノ酸、ペプチド、糖、イオン、脂質、ホルモンなど、異なる基質選択性をもつ。また、SLC輸送体は血液脳関門に多く発現しており、神経活動に欠かせない栄養素を脳に供給している。多くの先行研究により、SLC遺伝子における変異と疾患の関係が指摘されてきた。例えば、ファミリー2 (SLC2) は重要なエネルギー源であるグルコースを脳に輸送している。SLC2の遺伝子変異によるアミノ酸置換は、乳幼児期に痙攣やてんかんなどを引き起こすことが報告されている。また、ファミリー22 (SLC22) のカルニチン輸送体における遺伝子変異は、全身性カルニチン欠乏症や乳幼児突然死と関連することが報告されている。しかし多くの場合、アミノ酸置換が病気を引き起こす物理化学的な機序は明らかになっていない。SLC輸送体を含む膜タンパク質の立体構造や動態の解析実験は、技術的に容易ではない。また、配列多様性および立体構造に関する情報が十分に関連付けて解析されていないことも原因の一つであると考えられる。そこで、ゲノム配列・変異・立体構造・機能に関する情報を接続し、計算科学の視点から機能変異の理解を試みることは有効である。

タンパク質の立体構造は配列情報よりも保存されることが知られている。つまり、アミノ酸配列の変化が機能に与える影響を明らかにするためには、まず立体構造やその動的変化に、アミノ酸配列の変化がどのように関連するかを理解する必要がある。このような統合解析には、各SLC輸送体のアミノ酸配列、変異、構造、機能および表現型に関する情報を関連付けたデータベースを構築しなければならない。

本研究は、ヒトSLC輸送体に関するデータベース構築を通して、アミノ酸置換と立体構造や機能障害の情報とを接続し、疾患に繋がるタンパク質の動的構造変化を計算科学により明らかにすることを目的とする。そのため、下記に記す三つの課題について研究を行った。

- 1) ヒトSLC輸送体に関するリレーショナルデータベースの構築と統合解析
- 2) アミノ酸置換がタンパク質動態に与える影響の分子動力学シミュレーションによる解析
- 3) アミノ酸保存度に基づく進化的重要性の検証

1) ヒトSLC輸送体に関するリレーショナルデータベースの構築と統合解析

収集したヒトSLCのアミノ酸配列、遺伝子名、ファミリー分類、立体構造情報、トポロジー予測結果、変異および疾患情報を用いて、リレーショナルデータベース iMusta4SLC (Integrated Mutational and STRuctural Analysis FOR SLC) を構築し、公開した (<http://cib.cf.ocha.ac.jp/slc/>)。

データベースに格納した情報をもとに、立体構造上の変異の位置と機能障害/疾患との間に相関があるのかを検証するため、トポロジーに着目して疾患関連変異と非関連変異の頻度を比較した。膜タンパク質のトポロジーとは本来、細胞膜に対するタンパク質のN末端・C末端・膜貫通領域の位置情報を指す。本研究では、TopCons2(Tsirigos, et al. 2015)によるトポロジー予測に基づき、新たに独自のトポロジー領域R1-5を定義した(図1A)。また、疾患および非疾患の区別に関しては、ClinVarの臨床的意義の記述を参考に、‘risk factor’もしくは‘pathogenic’の記述を含む変異を「疾患変異」、‘benign’と定義された変異を「非疾患変異」と見なした。疾患変異(p) および非疾患変異(n)のそれぞれについて、各トポロジー領域($R_i, i=1,2,\dots,5$)における変異頻度を調べた。頻度は、ログオッズ比により表した。

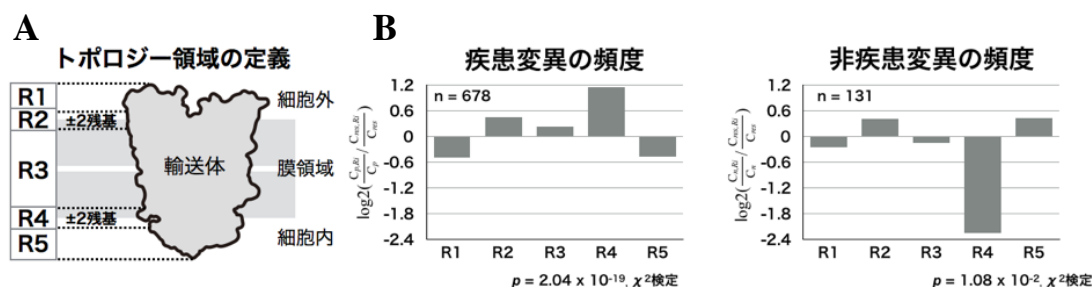


図1. トポロジー領域の定義と変異頻度

各トポロジー領域における変異の頻度を解析した結果、R4領域では疾患変異が存在する頻度が有意に高いことが分かった(図1B)(疾患, $p=2.04 \times 10^{-19}$; 非疾患, $p=1.08 \times 10^{-2}$)。アミノ酸置換パターンの頻度を確認したところ、R4領域ではArgからTrpへの変異が最頻出であることも明らかになった(11.36%)。

2) アミノ酸置換がタンパク質動態に与える影響の解析

R4領域におけるArgからTrpへの変異が立体構造および動態にもたらす影響を詳しく解析するため、分子動力学シミュレーションにより、野生型SLCと変異型SLCの動態を比較した。本計算では異なる二つのグルコース輸送体の立体構造、つまり細胞質内側に開いた構造をもつSLC2A1 (PDB ID: 5EQI)、および細胞外側に開いた構造をもつSLC2A3 (PDB ID: 4ZW9) を使用した。これらを野生型モデルとみなし、SLC2A1には変異 (G91D, R92W, R92Q, R93W, R93Q, R333W, R333Q) を、SLC2A3には変異 (G89V, R90W, R91C, R91H, R331K, R331S, R331W) をPyMOLにより導入し、それぞれの変異モデルを作成した (いずれもR4領域相当に位置するSNP)。また、3種類のシミュレーションモデル、全原子モデル・粗視化モデル・保存性弾性ネットワークモデルを作成し、全原子モデルに対しては10ns(分子力場: CHARMM36, n=3)、粗視化および保存性弾性ネットワークモデルに対しては1 μ s(分子力場: MARTINI, n=5)の計算を、GROMACS(v.5.0.7)を用いて実行した。

得られたタンパク質の軌道から、まず平均二乗偏差(RMSD)を算出し、構造が安定していることを確認した。野生型モデルのシミュレーション結果から得られたタンパク質の継時的な構造変化を主成分分析したところ、動態に最も影響が大きい第一主成分は、内開き・外開きの相互運動を示すことが分かった。これらの第一主成分および第二主成分を軸とする二次元上に、各シミュレーションの軌跡をプロットした(図2)。射影像を比較した結果、保存性弾性ネットワークモデルは、全原子モデルや粗視化モデルよりも短い計算時間で、内開き状態から外開き状態まで観察できることが明らかになった。

次に、保存性弾性ネットワークモデルの結果から、野生型と変異型の動態比較を行った。開閉の度合いを表す軸において、野生型と変異型の間には有意な差が見られ、変異型は野生型と比較して内開き構造の出現頻度が減少していることが分かった (図3)。

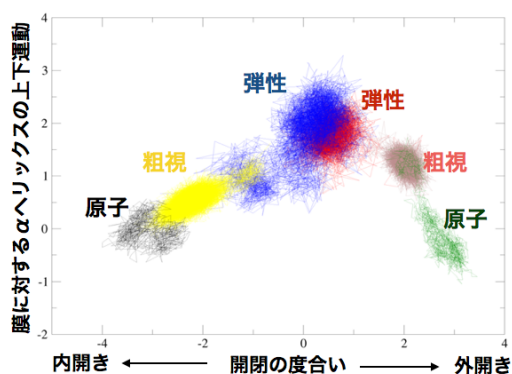


図2. シミュレーションモデルの比較

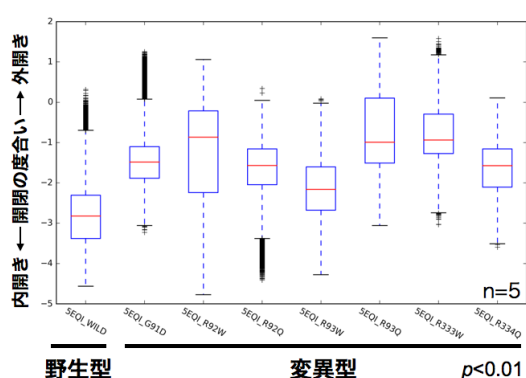


図3. 保存性弾性ネットワークモデル

3) アミノ酸保存度に基づく進化的重要性の検証

膜領域と細胞質の境界におけるアルギニン残基は他の生物種でも同様に重要なのかを検証するため、SLC2ホモログにおけるアミノ酸残基保存度を調べた。

遠縁なホモログの検索は、隠れマルコフモデル (HMM) の原理を用いたホモロジー検索プログラムHMMer(v3.0) (Eddy, 2011) により行った。検索クエリーとしてSLC2およびSLC22 (SLC2と相同性の高いファミリー) の配列情報を用いた。この際、これら輸送体のアミノ酸配列をN末端側およびC末端側に分割し、N末端由来のクエリー (N-side profile)、C末端由来のクエリー (C-side profile)、N末端とC末端の情報を両方用いたクエリー (N+C profile) を生成して、ホモロジー検索を行った。このような工夫により、SLC2の片方みのドメインを奇数個保有するホモログも検索の対象とした。

得られたホモログの生物種は、全ての生物種ドメイン (細菌・古細菌・真核生物) に由来していた。また、20%以上のホモログのアミノ酸配列において保存された残基をSLC2A1の立体構造上にマッピングしたところ (図4, A, N+C profile; B, N-side profile; C, C-side profile)、膜領域と細胞質の境界面において残基が保存されていることが明らかになった。特に、SLC2A1において92残基目のアルギニンにあたる残基は、全ての生物種ドメインにおいて高い保存性を示した。

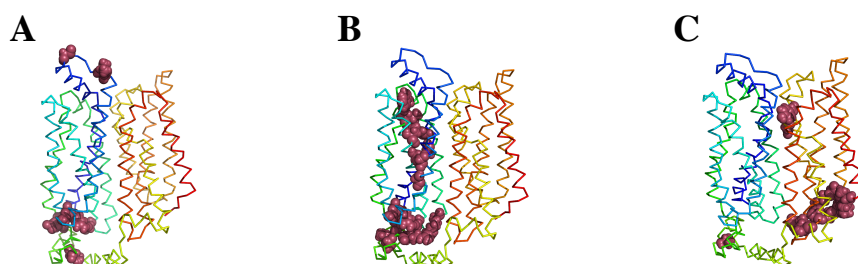


図4. SLC2ホモログの20%以上において保存された残基のマッピング

まとめとして、iMusta4SLCの統合解析により、細胞質と膜領域の境界面におけるアルギニン変異は、SLC輸送体の機能に高頻度で影響を与える疾患変異であることが分かった。このような細胞質と膜貫通領域の境界面におけるアルギニンからのアミノ酸置換は、分子動力学シミュレーションの主成分分析から、内開き構造の出現頻度を減少させ、基質輸送効率を低下させている可能性が示唆された。さらに、SLC2ホモログの配列解析の結果から、細胞質と膜領域の境界における残基はヒトだけでなく、他の生物においても同様に保存されていることが明らかになった。以上の結果から、SLC輸送体の変異情報と構造/機能情報を計算科学的に接続することで、疾患変異が引き起こす動的構造変化への影響を明らかにすることができた。