

博士論文

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

目次

略語	1
序章	3
生体における輸送体の重要性	4
SLC輸送体の分類と基質	8
SLC輸送体の局在と機能	9
SLC輸送体の立体構造	12
SLC輸送体の遺伝子変異と疾患との関連	16
本研究の概要	19
第1章 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	22
1.1 序	23
1.2 方法	24
1.2.1 データ収集	24
1.2.2 リレーショナルデータベースの構築	25
1.2.3 インターフェースの構築	27
1.2.4 アミノ酸配列類似度の算出と可視化	30
1.2.5 アミノ酸保存度の算出および保存残基の立体構造マッピング	30
1.2.6 トポロジー領域の定義	31
1.2.7 構造および変異情報の統合解析	31

1.2.8	アミノ酸置換パターンと頻度の可視化	34
		35
1.3	結果および考察	36
1.3.1	iMusta4SLCの概要	36
1.3.2	SLC輸送体の配列類似度の算出と可視化	39
1.3.3	SLC輸送体における保存残基の探索およびマッピング	41
1.3.4	トポロジー領域ごとの変異頻度	43
1.3.5	トポロジー領域ごとのアミノ酸置換パターン	45
1.3.6	疾患変異と保存残基の位置関係	47
第2章	分子動力学シミュレーションによるグルコース輸送体の変異の影響解析	50
2.1	序	50
2.2	方法	53
2.2.1	シミュレーションモデルの構築	53
2.2.2	分子動力学シミュレーション	62
2.2.3	RMSDの算出	63
2.2.4	主成分分析	64
2.2.5	RMSFの算出	64
2.3	結果および考察	66
2.3.1	RMSDによる構造安定性の観察	68

2.3.2	シミュレーションモデルの検証	65
2.3.3	保存型弾性ネットワークモデルを用いた野生型と変異型の比較	72
2.3.4	RMSFによる各残基のゆらぎ解析	76
第3章 グルコース輸送体のホモログにおける保存残基の探索		
		86
3.1	序	86
3.2	方法	87
3.2.1	SLC2のアミノ酸配列および立体構造の重複に関する検証	87
3.2.2	隠れマルコフモデルを用いたホモロジー検索	89
3.2.3	ホモログの保存部位解析	91
3.3	結果および考察	92
3.3.1	立体構造の対称性の検証	92
3.3.2	隠れマルコフモデルを用いたホモロジー検索	99
3.3.3	ヒットタンパク質の生物種	104
3.3.4	SLC2ホモログの保存部位解析	106
3.3.5	保存残基の立体構造マッピング	113
終章 総合討論		
		115
謝辞		
		123
参考文献		
		125

略語

ABC	ATP-binding cassette
SLC	Solute carrier
MATE1	Multidrug and toxin extrusion 1
MATE2-K	Multidrug and toxin extrusion 2 - splice variant
HUGO	Human Genome Organization
HGNC	Human Genome Organization Gene Nomenclature Committee
ADME	Absorption, distribution, metabolism and elimination
CNS	Central nervous system
BBB	Blood-brain barrier
GLUT1	Glucose transporter 1
GLUT3	Glucose transporter 3
GLUT9	Glucose transporter 9
SNP(s)	Single nucleotide polymorphism(s)
NGS	Next generation sequencer
NHGRI	National Human Genome Research Institute
URAT1	Urate transporter 1
OCTN2	Organic cation/carnitine transporter
MFS	Major facilitator superfamily
APC	Amino acid polyamine organocation
ATP	Adenosine triphosphate
NMR	Nuclear magnetic resonance

SAIL	Stereo-array isotope labeling
iMusta4SLC	Integrated mutational and structural analysis for SLC proteins
RDBMS	Relational database management system
ERD	Entity-relationship diagram
MD	Molecular dynamics
AT	Atomistic model
CG	Coarse-grained model
CON	Conserved elastic network model
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
PCA	Principal component analysis
ED	Essential dynamics
HMM	Hidden Markov Model

- 序章 -

序章

生体内における輸送体の重要性

細胞の内部環境は、細胞膜により外部環境から隔てられることで、安定して生体反応を行えるように恒常性が保たれている。細胞膜は、脂質分子が疎水性部位を内側に向け、親水性部位を細胞外もしくは細胞内側に向けた状態で二層に重なった構造、いわゆる脂質二重層と呼ばれる構造をしている。この細胞膜の脂質二重層は、極性分子および荷電分子の通過を防ぎ、半透過性の拡散バリアとして機能している。水を含む小さな疎水性および極性分子（一酸化窒素や酸素分子など）は、脂質二重層を透過することができる。しかし、分子サイズが大きくなるほど、小分子に対する透過性は減少する。また、細胞膜は流動的な自己再生機能をもち、互いに融合したり、小さな小胞を分裂させたり、細胞分裂の際に二つに分裂したりすることができる。これらの細胞膜の特性は、外部環境の変化や刺激を緩衝し、細胞や組織内の生体反応を安定的に維持するうえで、重要な役割を果たしている。

このようにして細胞は、外部環境から独立した状況のもと、種々の生体反応を進めることができる。細胞内部は水分および酵素やアミノ酸などの栄養素を含む細胞質基質に満たされており、核や小胞体、ミトコンドリア、リボソーム、リソソームなどの細胞小器官が存在する。ここでは、DNAの複製や転写、翻訳や、タンパク質のフォールディング、エネルギー生成、不要物の分解などが行われる。しかし、継続的にこれらの機能を維持してゆくためには、当然、栄養素の取り込みや不要物の排出が常に必要不可欠となる。そうでなければ、すぐに栄養素は枯渇し、不要物が山積して細胞内部の環境は急激に悪化してしまうだろう。では、細胞膜により外部から独

立している細胞内部は、一体どのようにして栄養素を取り込み、不要物を排出しているのだろうか。

細胞膜には様々な膜タンパク質が埋め込まれており、その一部は細胞膜を介して小分子を輸送することができる（図0-1）。驚くべきことに、全ゲノム解析の結果から、生物のゲノムにコードされた遺伝子のうち、平均的に三分の一が膜タンパク質に関連するものであることが示されている[Lukas, *et al.* 2016]。ヒトの場合は、遺伝子の30%弱が膜タンパク質に関するものであると言われている。つまり、ヒトゲノムの中にコードされた遺伝子20,687 (ENCODE Consortium, 2012) のうち、約6,000種が膜において機能するタンパク質であることを意味する[Lukas, *et al.* 2016]。これらの報告から、生物は、細胞膜において機能するタンパク質に多くのリソースを割いていることがわかる。細胞膜に局在する膜タンパク質は、大きく分けて四つの機能カテゴリー、1) 輸送体、2) 受容体、3) 酵素、4) 構造タンパク質に分類することができる[Lukas, *et al.* 2016]。輸送体とは、イオンや小分子の輸送を担うタンパク質であり、外部環境と内部環境の仲介役、もしくはゲートキーパー（門番）などに例えられることもある。受容体とは、外界から情報を受け取る（つまり、特定の基質と結合する）と、構造変化などを通して内部にその情報を伝達することができる膜タンパク質を指す。酵素には、加水分解酵素や、種々の合成酵素や酸化還元酵素などが含まれる。構造タンパク質とは、カドヘリンやインテグリン、タイトジャンクションを構成するタンパク質などに代表される、他の細胞と結合する際に必要となる膜タンパク質を指す。

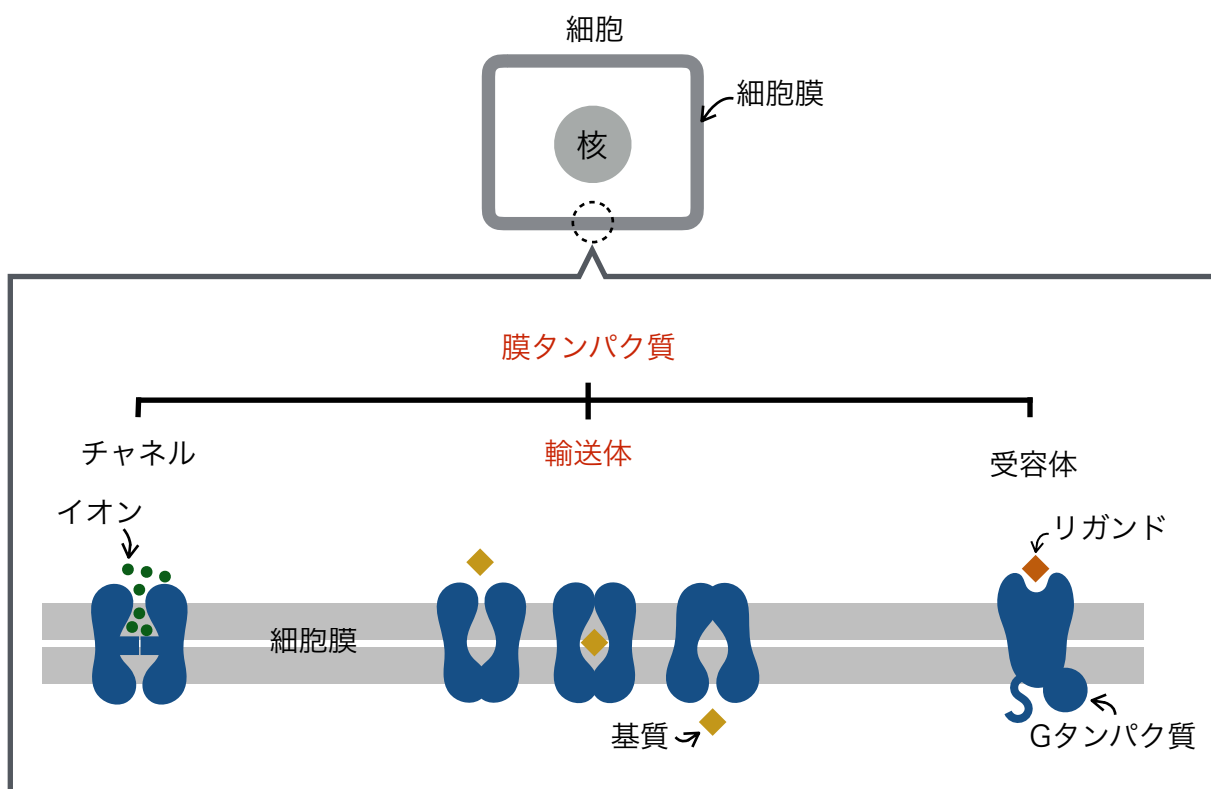


図0-1 代表的な膜タンパク質の模式図

代表的な膜タンパク質の一つであるチャネル（左）は、イオンや水分子を細胞膜を介して受動的に輸送する。輸送体（中央）は、受動的もしくは能動的に、構造変化を伴いながら基質を輸送する。受容体（右）は、リガンドとの結合により活性化し、細胞内シグナル伝達を開始する。

細胞内へ栄養素を取り込んだり、細胞外へ不要物を排出する輸送体は、チャネルやキャリア、ポンプといった種類がある。チャネルは、受動的な拡散によりイオンや水分子を輸送する[Gouaux, *et al.* 2005]。一方で、キャリアやポンプといった輸送体は、チャネルより幅広い基質選択性をもち、受動的もしくは能動的に基質を輸送する。輸送体には、二つの大きなスーパーファミリーであるATP binding cassette (ABC) 輸送体もしくはsolute carrier (SLC) 輸送体に分類できる[Keogh, *et al.* 2016]。これらのスーパーファミリーに属する輸送体は、合計で500以上あると言われている[Keogh, *et al.* 2016]。哺乳類におけるABCスーパーファミリーは、現在7つのファミリーに分類され、合計で48種類の輸送体が同定されている(<http://www.genenames.org/genefamilies/ABC>)。遺伝子名はファミリーに応じて、ABCAからABCGの表記で記される。哺乳類のABC輸送体は、全て排出に関する輸送体であり、ATP加水分解により得たエネルギーを利用して能動的に基質を輸送する。一方、ヒトのSLCスーパーファミリーは、52ファミリーに分類された400以上の遺伝子から成り立ち、輸送体の中では最も多いスーパーファミリーを形成している。SLC輸送体は、ATPを利用せずに構造変化を伴いながら基質を輸送する[Tanaka, *et al.* 2017]が、ABC輸送体と異なり、受動輸送するタイプと、能動輸送するタイプが両方含まれている[Keogh, *et al.* 2016]。能動輸送の場合は、電気化学的な勾配に反して基質を細胞内に取り込み、ナトリウム勾配を利用することが多い。

さらに輸送体は、輸送の方向に応じて、単輸送体 (Uniporter; Facilitated transporter)、共輸送体 (Symporter; Cotransporter)、対向輸送体 (Antiporter; Exchanger) などに分類できる。単輸送体は、イオンや小分子をカップリングせず、電気化学的な濃度勾配に従って、単一の基質を細胞内に取り込む。例えば、糖類、

アミノ酸、ヌクレオチドなどの輸送はこの方式で行われる。共輸送体は、濃度勾配に逆らって一つのイオンもしくは小分子を輸送する際に、同時にもう一つのイオンもしくは小分子を同方向に輸送する。このタイプの輸送体には、Na⁺共役型アミノ酸共輸送体やNa⁺共役型グルコース輸送体などがある。最後に、対向輸送体は、二つのイオンもしくは小分子を、膜を介してそれぞれ異なる方向に輸送する。Na⁺/Ca²⁺対向輸送体や、Na⁺/H⁺対向輸送体などがこの輸送方式をとる。通常の生理学的条件下では、SLC輸送体は基質を細胞へ取り込む方向で輸送するが、例外として、MATE1 (Multidrug and toxin extrusion 1) (SLC47A1) とMATE2-K (Multidrug and toxin extrusion 2 - splice variant) (SLC47A2) は排出の方向に輸送する[Lukas, *et al.* 2016]。このように、多様な輸送体のはたらきにより、細胞内への栄養素の取り込みや、不要物の排出が円滑に行われ、細胞内環境を一定に維持することができる。

SLC輸送体の分類と基質

SLC輸送体は、前述の通り、輸送体の中でも最も多くの遺伝子を含む重要なスーパーファミリーである。現在、52ファミリー (SLC1-SLC52) から成り立つ400以上のヒトSLC遺伝子が同定されている。その分類法は、塩基配列の一致度、機能、基質の類似性、膜貫通領域の数（多くは10から14）などに基づいている[Fredriksson, *et al.* 2008; Schlessinger, *et al.* 2014; Schlessinger, *et al.* 2013]。Human Genome Organization (HUGO) およびHuman Genome Organization Gene Nomenclature Committee (HGNC) により、現在800を超える輸送体遺伝子が遺伝子ファミリーのサブクラス分類が決められ、SLCはその半数を占めている[Marcus & Mak, 2016]。Hedigerらにより設立されたBioparadigms (<http://www.bioparadigms.org>) により制作されたSLC Tables (<http://slc.bioparadigms.org>) [Hediger, *et al.* 2013]というデータベー

スにおいて、ファミリー分類、遺伝子名、タンパク質名、基質、機能などの情報を参照することができる。なお、SLC21に関しては、現在ではSLCOと名称が変更されている。

SLCスーパーファミリーによって輸送される基質は、多岐に渡る。各ファミリーが異なる基質選択性をもち、アミノ酸、ペプチド、糖、イオン、脂質、ホルモンなどが含まれる。また、生体内に自然に存在する物質、いわゆる内因性物質の輸送だけでなく、薬剤や環境ホルモンなど、生体内には自然に存在しない物質、すなわち外因性物質も、SLC輸送体を通して各細胞に取り込まれることが明らかになっている [Keogh, *et al.* 2016]。薬剤が体内に吸収・分解され、標的組織や標的細胞に取り込まれる過程や、薬物の血中濃度の推移などについて、薬理学や薬物動態学の分野では、薬剤のADME (Absorption (吸収)、Distribution (分布)、Metabolism (代謝)、Excretion (排泄)) と呼ぶ。SLC輸送体はこのADMEに深く関わっており、薬物動態と特に関連の深いSLC輸送体の一部のファミリー (e.g SLCO, SLC22, SLC47) は、製薬業界から強い関心を持たれ、精力的に研究されてきた [Keogh, *et al.* 2016]。

SLC輸送体の局在と機能

細胞膜を介した物質輸送が特に重要とされる組織 (腸、肝臓、腎臓、血液脳関門など) においては、SLC輸送体の局在とはたらきが、組織の機能に大きな影響を及ぼす [Nałęcz, 2017; Morris, *et al.* 2017; Suhy, *et al.* 2017]。SLC輸送体は、生体におけるほとんど全ての細胞に発現している。一部のSLC輸送体 (SLC25など) は、ミトコンドリアの膜など細胞外膜以外の膜にも発現する。しかし、ここでは主に細胞外膜において局在するSLC輸送体に限定して述べる。

脳における神経組織は、全身の組織の中でも最重要かつ厳密に恒常性を制御する必要のある組織だ。中枢神経（Central nervous system: CNS）は、物理的な分離体により、全身のその他の部位から断絶されることで内部の環境を保つことができる [Dickens, *et al.* 2016]。この物理的な分離体は、イオン、神経伝達物質、高分子、神経毒、および栄養素の濃度や組成、ならびに異物の出入りを厳密に制御している [Dickens, *et al.* 2016]。CNSの物理的バリアとして、血液と脳間質液とを分離する血液脳関門（Blood-brain barrier: BBB）や血液と心室脳脊髄液との間にある脈絡叢、血液とくも膜下における心室脳脊髄液を分離するくも膜上皮（髄膜障壁）が存在する。

これら三種類のバリアの中で、BBBは特に重要であると考えられている。BBBの構造を図0-2に示す。BBBが重要であるとされる理由の一つとして、BBBは脳内の毛細血管に張り巡らされていることから表面積が非常に大きいという性質がある。その表面積は、実に15~25m²に及ぶとも言われており [Wong, *et al.* 2013]、脳が消費する膨大なエネルギー（全身の15-20%）を継続的に生成するための栄養素の取込みに役立っている。同時に、BBBは、不要物や毒物など異物の神経組織への流入を防ぐ役割も担っている。これまで行われた臨床研究から、98%以上の小分子薬剤とほぼ全ての大きな分子を用いた生物学的治療法は、神経組織への輸送がBBBにより遮断され、ドラッグデリバリーに失敗していると報告されている [Pardridge, 2005]。また、BBBが重要とされるもう一つの理由として、BBBは、CNSの細胞外液と血液を直接隔てていることも挙げられる。

脳の毛細血管断面図

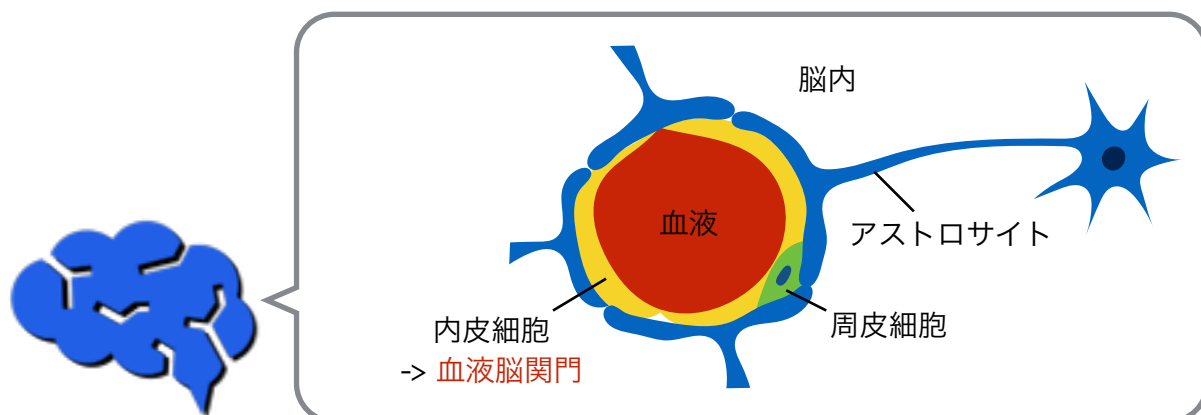


図0-2 血液脳関門の構造

脳の毛細血管の断面において、血管とアストロサイトの間に内皮細胞がぎっしりと張り付き、バリアを形成している。この組織を血液脳関門と呼び、神経組織に異物や有害物質が侵入するのを防ぐ一方、必要な栄養素のみ選択的に取り込んだり、不要な物質を排出する役割がある。

このBBBにおいて、SLC輸送体が神経組織への小分子の輸送を担っていることがわかっている。各SLC輸送体がどの程度BBBにおいて発現・機能しているかについては、まだ正確には明らかになっていない。しかし、少なくとも、一部のファミリー（SLC2, SLC7, SLC16, SLC22, SLCO）はBBBに発現し、重要な機能を果たしていることが先行研究から報告されている [Uchida, *et al.* 2011; Geier, *et al.* 2013; Mastroberardino, *et al.* 1998; Kalliokoski & Niemi, 2009; Boado, *et al.* 1999; Okura, *et al.* 2008; Geldenhuys, *et al.* 2012]。特に、SLC2（グルコース輸送体）のメンバーであるGLUT1（SLC2A1）は、BBBにおいて多く発現しており、多くの生物にとって主要なエネルギー源となるグルコースをCNSに輸送していると考えられている。

SLC輸送体の立体構造

タンパク質は、生体内の生理学的条件下において、エネルギー的に安定な立体構造を形成して熱振動しており、その揺らぎはタンパク質の機能と密接に関与している。SLC輸送体は、複数の α ヘリックスとループにより構成される膜貫通型タンパク質であり、膜を完全に貫いた構造をしている。脂質膜を介して基質を輸送するためには、溶液中の基質を捉え、疎水性の膜領域を貫通させて基質を運び、再び溶液中に基質を放つ過程が必要であると考えられる。この機能が物理化学的にどのような仕組みにより実行されるのか理解するためには、輸送体の立体構造や可動性を詳しく分析することが重要である。

ところが、現在のところ、SLC輸送体の立体構造はほとんど決定されていない。その理由は膜タンパク質の立体構造解明は技術的に難しいとされており、SLC輸送体も例外ではないためだ。タンパク質の立体構造の決定方法として、最も一般的に用いられるのがX線結晶構造解析法がある。この手法では測定したいタンパク質を一度

可溶化した後、結晶化する必要がある。膜タンパク質を可溶化する際に、周囲の脂質膜を分解するため界面活性剤を用いるが、膜タンパク質は脂質二重膜に埋め込まれた生理学的な条件下でなければ構造が不安定になるため、結晶化の大きな障壁となってしまうのだ。仮に結晶化に成功し、立体構造の情報が得られたとしても、細胞膜を貫通した状態（生理的な条件下での環境）においても同様の立体構造をとっているかは保証されない。また、核磁気共鳴法（Nuclear magnetic resonance: NMR）もタンパク質の立体構造や動的構造の解明によく用いられる手法である [Tate. 2013; McDermott. 2009]。NMRを用いる利点は、タンパク質の動的な構造解析に優れていることだ。X線結晶構造解析はタンパク質を結晶化するため、静的な立体構造を解き明かす一方、NMRは、溶液中におけるタンパク質の立体構造や揺らぎ、リガンドとの相互作用について解析することができる。ただし、NMRの欠点は、解析対象とできるタンパク質は分子量が比較的小さいもの（一般的に、40kDa以下）に限られる。膜タンパク質の分子サイズは40-200kDaのものが多い [Keogh, *et al.* 2016] ため、NMRの応用は難しい。これらの課題を克服するため、膜タンパク質を少数の脂質二重層で囲った状態で可溶化する技術のNanodisc法 [Liang. 2016; Hagn. 2013; Bayburt, 2002] や、タンパク質の構成原子を高度選択的に安定同位体標識することでNMR情報を効率よく取得するSAIL法（Stereo-array isotope labeling: 立体整列同位体標識） [Kainosho, *et al.* 2006; Miyanoiri, *et al.* 2016] などが開発されており、今後の発展および応用が期待されている。

現在、タンパク質の立体構造データベースProtein Data Bank (PDB) [Bernstein, *et al.* 1977] には、X線結晶構造解析により決定された立体構造が121,394件（2017年12月現在）登録されている。このうち、膜タンパク質の登録数は、4,417件で、割合

にすると全体のわずか3.6%に過ぎない。 α ヘリックスをもつ膜タンパク質に関しては、登録数は2,683件で、全体における割合は約2.2%である。さらに、ヒトの膜タンパク質に限定すると、126件のみであった。前述の通り立体構造決定手法の技術革新が進んでいるものの、ヒトにコードされた遺伝子の約30%弱、6,000ほどが膜タンパク質関連遺伝子であることを考慮すると、実際には膜タンパク質の立体構造のほとんどは、未だに未知のままである。

ヒトSLC輸送体の立体構造に関して、現時点でX線結晶構造解析により立体構造が決定されているヒトSLC輸送体を表1にまとめた。この表には一部のドメイン構造しか明らかになっていないSLCも含まれているが、わずか4つのファミリー（SLC1, SLC2, SLC4, SLC25）の一部メンバーの構造が明らかになっているのみである。ヒト以外の生物由来であれば、大腸菌 (*Escherichia coli*) 由来のラクトース透過酵素（SLC37ホモログ）の結晶構造（PDB ID: 4OAA）[Kumar, *et al.* 2014]および超好熱性真正細菌 (*Aquifex aeolicus*) 由来のロイシン輸送体（SLC6ホモログ）をはじめとした立体構造（PDB ID: 4FXZ）[Wang, *et al.* 2012]などが、動的構造や基質輸送メカニズムの優れた研究対象とされている[Kazmier, *et al.* 2017]。

これまで、限られた膜タンパク質の立体構造のスナップショットをもとにして、基質輸送メカニズムの仮説が提案されてきた。Colasらの報告によれば、SLC輸送体の基質輸送メカニズムとして、以下のうような3種類の交互アクセス機構（もしくは、相互アクセス機構とも呼ばれる）

- 1) ロッカースイッチ (Rocker switch)
- 2) ゲートポア (Gated-pore)
- 3) エレベーター (Elevator)

が提唱されている[Colas, *et al.* 2016]。ロッカースイッチ機構は、前述のラクトース透過酵素やグルコース輸送体に代表される基質輸送メカニズムである。ロッカースイッチ輸送体では、対称的な構造をもつN末端側ドメインがC末端側ドメインが、外開き状態から内開き状態への構造変化を繰り返すと考えられている。ゲートポア輸送体では、輸送体の結合部位は外向きゲートおよび内向きゲートによって囲まれている。基質は、これら2つのゲートが交互に開閉することによって細胞内へと放出される。エレベータシステムは、基質を捕捉して細胞質側に移動する輸送ドメインと、ほとんど動かない静的ドメインとの2つのドメインで構成される。まさにエレベーターのように、輸送ドメインが細胞外側から細胞質側へと降りてくる仕組みである。SLC1のホモログとされる *Pyrococcus horikoshii* 由来のグルタミン酸・アスパラギン酸輸送体 GltPh (PDB ID: 2NWW) [Boudker, *et al.* 2006]は、このエレベーター方式により機能する輸送体である。

PDBに登録された立体構造情報だけでなく、膜タンパク質のトポロジー情報も、タンパク質の構造特性を理解するうえで非常に重要である。トポロジーとは、タンパク質の立体構造を単純化し、一次構造や二次構造の相対的な位置を二次元において模式的に示したものである。特に、膜タンパク質におけるトポロジーは、膜貫通領域の数やアミノ酸配列における膜貫通領域の位置、およびN末端とC末端が細胞内外のどちらに位置しているかなどの情報を指す[Melén, *et al.* 2003]。先行研究では、実験から明らかになった膜タンパク質のトポロジー情報を蓄積するデータベースPDBTM (Protein Data Bank of Transmembrane Proteins) (<http://pdbtm.enzim.hu>) や、トポロジーを配列情報などから予測するための様々なアルゴリズムが開発されてきた。しかし、PDBTMは、PDBの情報を基盤としているため、PDBと同様にSLC輸送体のほとんどがまだ未登録であり、今後の発展が期待される。また、既存のトポロ

ジー予測手法は必ずしも正確に膜貫通領域の数と位置を当てられるわけではない。実験により決定された膜領域の数と位置を、KyteとDoolittleの疎水性スケール、DAS (Dense alignment surface)、TMHMM (Transmembrane hidden markov model)、HMMTOP (Hidden markov model algorithm for topology) で予測した結果と比較した研究がある[Ott and Lingappa. 2002; Lukas, *et al.* 2016]。報告によれば、どの予測手法も部分的に正確ではあるものの、膜貫通領域の数および位置の両方を完璧には当てられたものはなかった。したがって、膜タンパク質のトポロジー解析を行う際は、様々な予測手法の結果（存在すれば、ホモログの実験結果も含めて）を比較しながら慎重に検討する必要があると考えられる。

SLC輸送体の遺伝子変異と疾患との関連

ヒトゲノム解析の結果から、個人の遺伝子は一塩基多型 (Single nucleotide polymorphism: SNP) や挿入・欠失・繰り返し配列などをもつ多様性に富んだ配列であることが明らかになった[Sachidanandam, *et al.* 2001; Green, *et al.* 2003]。タンパク質をコードする領域 (エクソン) はヒトゲノム全体の約2%に過ぎないようだが、その中には20,000種を超える遺伝子が含まれ、実際には選択的スプライシングにより、遺伝子数より遥かに多くのスプライスバリエントが転写されていると考えられる。ヒトゲノム解析が終了して以降、次世代シーケンサー (Next generation sequencer: NGS) の登場により、ゲノム解析にかかる時間や費用などのコストが大幅に減少した。米国の国立ヒトゲノム研究所 (National Human Genome Research Institute: NHGRI) の算定によると、2001年1月の時点でヒト一人分の全ゲノム解読に\$95,263,072もかかっていた費用は、2017年7月の時点で\$1,121まで低下し、1メガバイトあたりのシーケンスコストはわずか\$0.012となった (Source: <https://>

www.genome.gov/sequencingsdata/)。このまま技術革新が続けば、\$1,000を下回るのも時間の問題である。それに伴い、様々な民族グループに属する被験者の遺伝的な多様性を検証するため、1,000人ゲノムプロジェクトをはじめとした大規模な国際研究が精力的に続けられた。2013年頃からゲノムに関するデータ量は急激に増え始め、現在はデータ総量が百万テラバイトを超える‘Million-Genome Era’に突入しつつある[Morgan, 2015]。これら大規模プロジェクトの結果から、次第に、個人の遺伝的多様性が幅広い疾患の発症と関連することも明らかになってきた。

遺伝的多型と疾患に関する研究成果は、国際的な公共データベース、ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) [Landrum, *et al.* 2014; Landrum, *et al.* 2016]、dbSNP (<https://www.ncbi.nlm.nih.gov/projects/SNP/>) [Sherry, *et al.* 2001]、COSMIC (<http://cancer.sanger.ac.uk/cosmic>) [Forbes, *et al.* 2016; Forbes, *et al.* 2017]などに蓄積されてきた。これらのデータベースから、アミノ酸置換の中でも疾患に深く関与すると疑われる変異もあれば、構造や機能にほとんど影響を与えず良性と見なされる変異もあることがわかってきた。この事実は、もちろん、SLC輸送体においても例外ではない。

SLC輸送体の遺伝子変による機能障害は、主に神経疾患や代謝疾患を引き起こすことが明らかになっている[Benarroch, 2010; Palmieri, 2013; Bröer & Palacín, 2011; Pramod, *et al.* 2013; Wright, *et al.* 2007]。現在、SLC輸送体において発見された疾患関連アミノ酸置換は、ClinVarにおいて679 (2017年12月現在) 登録されている (臨床的意義が‘Risk factor’, ‘Pathogenic’もしくは‘Likely pathogenic’と定義された単一遺伝子上に起きたミスセンス変異のみを集計)。例えば、GLUT1 (SLC2A1) におけるアミノ酸置換は、乳幼児期におけるグルコース輸送体欠損症候群、ジストニア、

てんかんを引き起こす[Wang, *et al.* 2000; Jiang, *et al.*, 2010; Wong, *et al.*, 2007]。脳が急激に発達する乳幼児期において、GLUT1の機能障害により脳へのグルコース供給が不足することで、このような疾患が引き起こされると考えられている。また、URAT1 (SLC22A12) およびGLUT9 (SLC2A9) の変異は遺伝性の腎性低尿酸血症を引き起こし、急性腎不全などの重篤な合併症を誘引する[Enomoto, *et al.* 2002; Matsuo, *et al.* 2008]。OCTN2 (SLC22A5) はカルニチン欠損症や時に乳幼児突然死と関連する[Nezu, *et al.* 1999; Sharma & Black. 2009]。SLC22はカルニチン輸送体であり、活性化脂肪酸をミトコンドリア内膜を介して輸送し、エネルギー生産や毒性のある酸化脂肪酸を除去する機能を果たしている。カルニチン輸送体の機能障害を持つ患者は、代謝性心筋症、中枢神経系異常、反復性腹痛および下痢などの胃腸症状を患うことが明らかになっている[Nezu, *et al.* 1999; Sharma & Black. 2009]。

SLC遺伝子のアミノ酸置換により引き起こされる疾患にはもう一つ特徴がある。それは、これらの疾患の多くが先天性あるいは乳幼児期や小児期において発症するという点だ。つまり、これまで報告されたSLC輸送体におけるアミノ酸置換は、生体へ及ぼす影響が大きく、環境要因が強く関与していないことが示唆される。がんや生活習慣病のように発症までに平均的に長い年月を要し、複数の遺伝的要因と環境要因の複雑な組み合わせにより疾患発症に至る場合、各遺伝子変異や環境要因が疾患発症にどれほど寄与したのか分析することは容易ではない。一方、SLC輸送体の単一のアミノ酸置換により引き起こされるの多く疾患は、発症時期が非常に早い。例えば、SLC26A4の変異により発症するペンドレッド症候群は先天性の難聴が認められる。SLC52の変異も小児期に進行性球麻痺を発症し、聴力に障害を生じる。前述の通り、SLC6やSLC22における変異は乳幼児突然死と関連し、SLC2におけるア

ミノ酸変異は乳幼児期に痙攣やてんかんを引き起こす。前述のURAT1 (SLC22A12)の変異が引き起こす腎性低尿酸血症も家族性であることが分かっている。

SLC輸送体におけるアミノ酸置換と疾患について、以上のように明確な因果関係が多数報告されている。それにも関わらず、アミノ酸置換から疾患発症までの間の物理化学的な機序は、ほとんど明らかにされていない。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、タンパク質の立体構造は配列情報より保存される[Olivella, *et al.* 2013]ことを踏まえ、まず立体構造やその動的変化に着目する必要がある。しかし、前述の通り、NMRなどを用いた実験的な膜タンパク質の動態解析は、技術的に困難でコストが高いという障壁がある。

本研究の概要

上述のように、SLC輸送体は、あらゆる生体反応で必要とされる栄養を各組織や細胞に輸送するための重要なスーパーファミリーであるが、遺伝子変異が入った場合、先天性もしくは早期に深刻な神経疾患や代謝疾患を生じることがある。しかし、膜タンパク質の立体構造解明や動態解析が実験的に難しいため、アミノ酸置換がSLC輸送体の動的構造に与える影響はほとんど明らかになっていない。

本研究では、計算科学的アプローチにより、SLC輸送体のアミノ酸置換が動的構造へ与える影響の分析を試みた。SLC輸送体のアミノ酸配列の変化が動的構造および機能の変化にどのように影響するか計算科学的手法による解析を試みた研究はこれまで例がない。そこで、SLC輸送体のアミノ酸配列、アミノ酸置換情報、立体構造、および表現型に関する情報を計算機を用いて接続し、統合解析することで、これらの一見異なる情報間における関係を明らかにしようとした。また、アミノ酸置

換が疾患につながる具体的な例に注目し、動的構造にどのような影響を与えるのか明らかにする。

以上の方針を踏まえ、本論文では、以下の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) アミノ酸置換がグルコース輸送体の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検証
- 3) アミノ酸保存度に基づく進化的重要度の解析

第1章では、本研究で新たに構築したヒトSLC輸送体に関するリレーショナルデータベースiMusta4SLC (Integrated MUtational and STRuctural Analysis FOR SLC proteins) に関する紹介と、統合解析の結果について述べる。第2章では、第1章において明らかになったアミノ酸置換と疾患の関係に着目し、疾患変異がSLC輸送体の動的変化に与える影響を分子動力学シミュレーションを用いて解析する。第3章では、ヒト以外のSLC輸送体のホモログに着目し、進化の過程において蓄積されたアミノ酸変異と保存度の関係から輸送機能に重要な部位を明らかにした。このように、タンパク質の構造やアミノ酸保存度の視点を加えた統合解析を行うことで、遺伝子変異とその表現型との関係を考察するうえで、より説得力のある考察を与えることができると考えられる。近年は、目覚ましい技術革新により蓄積された膨大なゲノムの多様性と疾患情報の医療応用について、技術だけでなく、個人情報保護法を中心とした法整備の側面からも激しい議論が続いている。今後も個人ゲノムの医療応用には様々な課題を解決していく必要があると考えられる。技術の精度、信頼性、教育、患者とのコミュニケーションや精神的ケア、医療制度の整備など、医療スタッフにとどまらず、基礎研究者と臨床スタッフ、公的機関などによる高度で緻密な連携が求められる。しかし、これら課題が解決されれば、最終的に国民は多大な恩恵を享受

することができると考えられ、個人のゲノム情報に基づいた疾患リスク管理や最適な予防法・治療法の選択、大規模な追跡調査、一遺伝子多型と疾患リスクの相関解析などの実現が期待される。本論文の成果は、このような新しい医療の実現に向けて、基礎研究と臨床研究の相互発展に大きく貢献できるものであると確信している。

- 第1章 -

第1章 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

1.1 序

膜輸送体タンパク質は、脂質二重膜を介して様々な基質を輸送することができる。SLCスーパーファミリーは、膜輸送体の中で最も多いスーパーファミリーを形成しており、その配列多様性や基質多様性は序章において紹介した通りである。SLC輸送体のはたらきは、栄養の取り込み、不用物の排出、シグナル伝達、細胞や組織の恒常性維持などに深く関与し、生体内の物質輸送に関する中心的な役割を果たしている。一方で、小分子の輸送に多大な影響を有するがゆえ、SLC輸送体の遺伝子変異は代謝疾患や神経疾患、時には突然死に繋がるような重篤な疾患を引き起こすことも明らかになっている[Nezu, *et al.* 1999; Wang, *et al.* 2000; Wright, *et al.* 2007; Wong, *et al.* 2007; Sharma & Black. 2009; Jiang, *et al.* 2010; Benarroch, 2010; Bröer & Palacín, 2011; Palmieri, 2013; Pramod, *et al.* 2013]。このようなアミノ酸置換はSLC輸送体の動態に何らかの影響を与えている可能性が考えられるが、具体的な動的構造への影響は明らかになっていない。その理由の一つとして、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ことが挙げられる。全52ファミリーのうち、ヒトの立体構造が明らかになっているのはわずか4ファミリーしかない。また、SLCの配列多様性や動態解析は、これまで個々に研究されることが多く、SLCスーパーファミリー全体を通して、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・表現型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と考えられる。タンパク質の立体構造は、配列情報よりも高度に保存されることが知られている[Olivella, *et*

al. 2013]。つまり、配列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のため、各SLC輸送体に対する配列・変異・構造・機能・表現型の情報を紐付けた新しいデータベースを確立した。本章では、SLC輸送体の構造的特徴と変異を統合解析するための研究補助ツールとして構築した新しいリレーショナルデータベース、iMusta4SLC (Integrated MUtational and SStructural Analysis FOR SLC proteins) の利用方法、活用例、統合解析結果について紹介する。

1.2 方法

1.2.1 データ収集

iMusta4SLCには、既存のデータベースから抽出したSLC輸送体の遺伝子名、アクセッション番号、遺伝子座、ファミリー分類、配列類似度、立体構造、変異情報、疾患情報が登録されている。SLCファミリーの分類および遺伝子名は、BioParadigms (<http://www.bioparadigms.org>)により構築された既存のデータベースであるSLC Tables (<http://slc.bioparadigms.org>) [Hediger, *et al.* 2013] および、Human Gene Nomenclature Committee (HGNC) (<https://www.genenames.org>)の命名法に従った。SLC遺伝子は一般に、SLCnXmの形で命名される。nはファミリー番号、Xはサブファミリーを表すアルファベット、mは個々の遺伝子に割り当てられた数字を示す。これはHGNCによって定められた標準的な命名法である。SLCファミリーの塩基配列およびアミノ酸配列は、NCBIから取得した[Brown, *et al.* 2015]。SLC輸送体の変異情報およびそれらに関連する疾患情報は、公共データベースであるClinVar[Landrum, *et al.* 2014; Landrum, *et al.* 2016] から、キーワード ‘Solute Carrier’

を用いて抽出した。立体構造情報は、Protein Data Bank (PDB) [Bernstein, *et al.* 1977]に登録されたタンパク質に対して、各SLC輸送体と類似の配列をもつタンパク質をBLAST [Johnson, *et al.* 2008]により検索・取得した。

1.2.2 リレーショナルデータベースの構築

取得した情報は、オープンソースのリレーショナルデータベース・マネジメントシステム (Relational Database Management System: RDBMS) として最も多く用いられているMySQL (<https://www.mysql.com>) により、データベースに格納・接続した。本研究で構築したデータベースの実体関連図 (Entity-relationship diagram: ERD) を図1-1に示す。ここでは、属性・配列・機能・構造・変異などに関する15種類のテーブルを作成した。各テーブルには、主キー (プライマリーキー) と呼ばれるカラムがある。主キーは、一行のデータ (レコード) につき一つの値が割り振られることで、データを重複せずに管理できる。また、主キーを指定することで単一のデータを特定でき、確実に抽出できる仕組みをもつ。さらに、他のテーブルと情報を接続するためには、外部キーと呼ばれるカラムを用意した。外部キーは、関連づけされた先のテーブルでは主キーになる。このように、主キーと外部キーの設定により、複数のテーブルから結合したデータを同時に抽出することができる。本データベースでは、図1-1のERDにおけるテーブルSLCのカラムslc_idが、各SLC輸送体に割り振られたIDであり、主キーにあたる。そして、その他のテーブルの多くが、外部キーとしてslc_idというカラムを含んでいる。このIDにより、各SLC輸送体に対して、異なるテーブルに登録された情報を関連づけ、同時に情報を引き出して解析することができる。

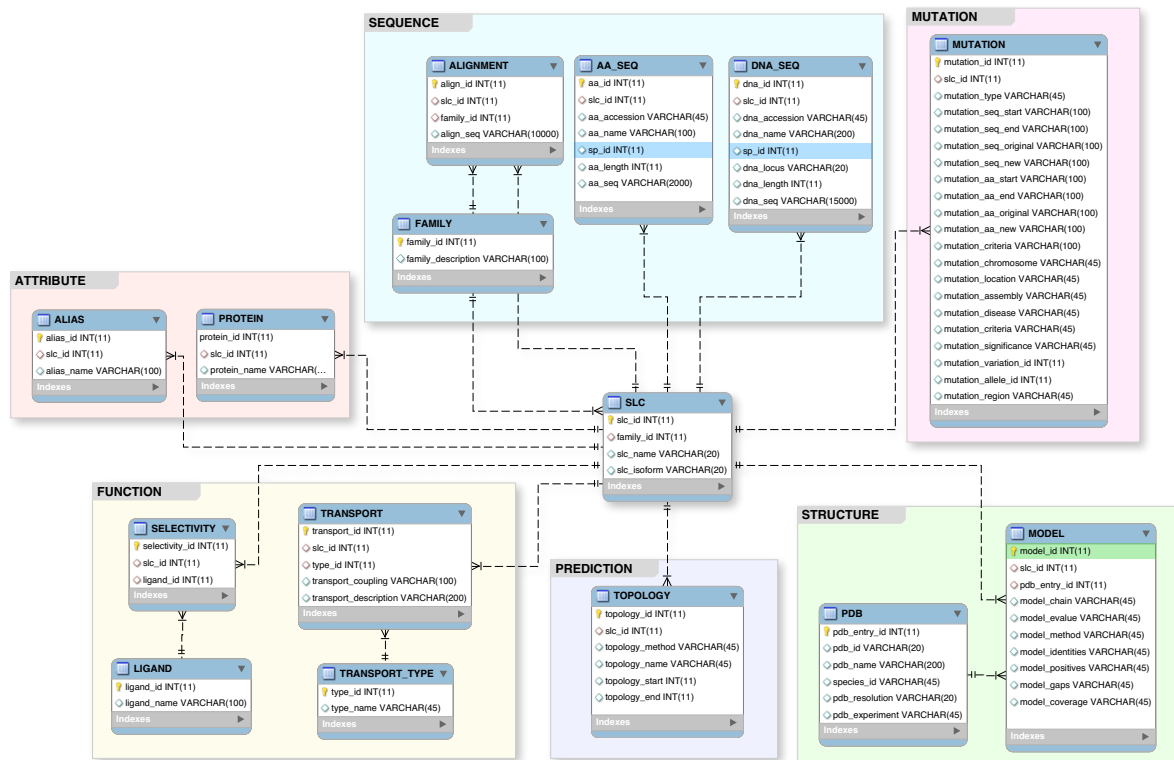


図 1-1 iMusta4SLCの実体関連図 (ERD)

iMusta4SLCのデータベースは、15種類のテーブルから構成されている。ここではテーブルの内容により6つのグループ（属性、オレンジ；配列、水色；変異、ピンク；構造、緑；機能、黄色；トポロジー予測、青）に色分けして示した。中央のマスターテーブル‘SLC’は、各SLC輸送体に対するIDが主キーとして登録されている。このSLC輸送体に割り振られたIDは、他のテーブルにおいては外部キーとして登録されており、異なるテーブル同士の情報を紐付ける際の結合要素として用いられる。各テーブルの枠内には、カラム名とデータのフォーマット（INT(n), n桁以内の整数; VARCHAR(n), n字以内の文字列）が記されている。

1.2.3 インターフェースの構築

データベースのインターフェースはFlask (<http://flask.pocoo.org>) により制作し、ウェブブラウザを通して様々なクエリーを設定し、ユーザーが必要とする情報を検索・ダウンロードできるような設計にした。検索結果は、その内容に応じて、ウェブブラウザの各ページに表示される。このウェブサイトの各ページでは、検索オプションがサイドバーに設定されている。検索オプションおよび抽出できるデータの種類に関する詳細は、‘How to Use’のページに記述した。

‘Family list’のページでは、ファミリーの番号、SLC輸送体の名前、アイソフォーム（スプライスバリエント）、タンパク質名、別名、ファミリーの代表的な機能を記したアノテーションを参照することができる。これらのデータは、上述のSLC Tables [Hediger, *et al.* 2013] およびHGNCに基づき収集した。

‘Function list’のページは、SLC Tablesの情報に基づき、各SLC輸送体に対する基質、輸送の際のカップリングイオン、輸送タイプに関する情報を掲載した。SLC輸送体の輸送タイプは、Cotransporter（共輸送体）、Exchanger（対向輸送体）、Facilitated transporter（単輸送体）、Orphan transporter、Electrogenicの5種類に分類されている。それぞれ構造変化パターンや基質結合部位の位置や数などが異なる可能性があるため、輸送タイプを把握することは重要であると考えられる。

‘Sequence similarity’のページにおいては、全SLC配列間における類似度をFASTA program (v36.3.6) [Pearson, *et al.* 1988; Pearson, *et al.* 2016]により計算した結果を、ユーザーインタラクティブな図により表現している。この図は、D3.js (<https://d3js.org>) [Bostock, *et al.* 2011]による ‘the hierarchical Edge Bundling protocol’を参考に制作している。類似度の閾値はサイドバーにおいて、 1.0×10^{-20} から 1.0×10^{-1} の範囲でユーザーが設定することができる。閾値よりE-valueの低いSLC輸送体

のペアは、青い線で結ばれるような仕様になっている。SLC輸送体はファミリーごとにクラスターを組んで配置されているため、基本的には同クラスター内で青い線が接続されていることが多い。しかし、青い線が外部のクラスター、つまり、他のファミリーと接続していることが多い場合、それらのファミリーは配列類似度が高く、進化的に近いホモログであると言える。進化的に近いホモログを把握するということは、立体構造や変異情報を解析する上で解析対象が広がるばかりでなく、足りない情報を補ったり推測できる可能性があるという点でも極めて重要である。

‘Sequence list’のページでは、配列のアクセッションID、遺伝子座、配列長、その配列の機能などを記した情報を公開している。アクセッションIDには、NCBIへのハイパーリンクが貼られており、簡単に情報源にアクセス・ダウンロードできるよう配慮した。

‘Alignment’のページでは、各ファミリーごとのマルチプルアライメントと保存度を表示している。保存度は、100%保存されている場合はアスタリスク (*)、80%以上保存されている場合はコロン (:)、50%以上保存されている場合はドット (.) で表示した。また、Missense変異が報告されているアミノ酸残基については背景色をつけた。ピンクは疾患変異、シアンは非疾患変異、グレーは臨床的意義が未知である変異を表している。

‘Template list’のページは、PDBデータベースに対して、各SLC輸送体のBLASTを行った結果を表示している。この結果に基づき、各SLC輸送体に対するテンプレート構造の検索を行うことができる。

‘Template coverage’のページでは、‘Template list’の詳細、つまり、PDBに対して各SLC輸送体のBLASTを行った結果が図式的に示されている。図の上部に位置する赤いボックスはクエリー配列を示し、青いストライプのボックスは、クエリー配

列に類似したPDBに登録された配列を表している。青いストライプは、配列の一致度に応じた四段階の青色から構成されている。最も濃い青色は、クエリー配列とヒット配列のアミノ酸残基が一致していることを示す。一方、最も薄い青色は、ギャップを示している。中間の二色は、BLOSUM62（相同性スコア行列）において正のスコアに相当すればやや濃い青色、そうでなければやや薄い青色に分類した。

‘Mutation list’のページでは、iMusta4SLCに登録されたSLC輸送体に関する全ての変異情報を参照することができる。変異の種類（ミスセンス、フレームシフト、欠失、挿入、終了）や、臨床的意義（risk factor, pathogenic, benignなど）、関連する疾患名や症状を参照することができる。各変異には、variant IDが割り振られており、variant IDに貼られたリンクから情報源にアクセスすることもできる。

‘Mutation and Topology’のページでは、変異位置とタンパク質の立体構造の統合解析をすることができる。SLCのそれぞれのファミリーにおいて、特定のトポロジー領域ごとのアミノ酸残基の置換パターンがヒートマップに表示される。ヒートマップの色は置換頻度と相関しており、特定のトポロジー領域において、頻度の高いアミノ酸置換パターンを視覚的に把握することができる。また、検索オプションから、アミノ酸置換の臨床的意義を指定することができるため、例えば疾患に関連するアミノ酸置換パターンを表示することも容易にできる。個々のアミノ酸置換情報は、ヒートマップ下のテーブルで参照することができる。

‘Mapping’のページでは、立体構造のテンプレートが存在し、かつ変異が報告されているファミリー（SLC1, 2, 5, 6, 7, 14, 22, 25, 37）に限定して、Missense変異のテンプレート構造へのマッピングを閲覧することができる。疾患変異（ピンク）・非疾患変異（シアン）・臨床的意義未知（白）のSphere形式で表示される。

テーブル形式で表示しているデータに関しては、'Export' ボタンから、PDF, CSV, Excelのフォーマットで結果をダウンロードすることができる。

1.2.4 アミノ酸配列類似度の算出と可視化

ヒトSLC輸送体は400種類以上存在し、52ファミリーに分類されているが、それぞれのSLC輸送体は互いにどれほど類似しているのだろうか。ファミリー同士の配列類似度や、同一ファミリー内のSLC輸送体の配列類似度を確認しておくことは、解析対象の把握や不足データを補うために重要である。そこで、iMusta4SLCの 'Sequence similarity' のページにおいて、全SLC輸送体間の配列類似性を一目で把握できるように可視化した。配列類似性はFASTA36 [Pearson, *et al.* 1988; Pearson, *et al.* 2016] により計算した。ユーザーが類似性の閾値としてE-valueを設定し、その設定値よりもE-valueが小さいSLC輸送体同士をホモログと見なした。D3.js (<https://d3js.org>) [Bostock, *et al.* 2011]の 'Hierarchical Edge Bundling' のプロトコルを利用し、各SLC輸送体をファミリーごとにグループ分けして円状に並べ、ホモログのSLC輸送体のペアが青い線で結ばれるような図を作成した。図はユーザーインタラクティブに拡大・縮小・ドラッグできるようにした。

1.2.5 アミノ酸保存度の算出および保存残基の立体構造マッピング

SLC輸送体の立体構造において進化的に保存された部位と、疾患変異の位置が関連するか確かめるため、SLCファミリーのアミノ酸保存度の算出と、保存残基の構造マッピングを行った。400種類のSLC輸送体は配列多様性があり、全て一緒にマルチプルアライメントすることは困難であった。したがって、立体構造が既知で、かつファミリーメンバーが多いSLC2とSLC22に着目し、解析ターゲットとした。これら

二つのファミリーは、前述の配列類似度の可視化結果からも分かるように、非常に類似度の高いホモログである。したがって、SLC2とSLC22のメンバーをClustalW (<http://www.genome.jp/tools-bin/clustalw>) [Thompson, *et al.* 1994; Larkin, *et al.* 2007] によりマルチプルアライメントし、入力ファイルとした。次に、ConSurf Server (<http://consurf.tau.ac.il/2016/>) [Glaser, *et al.* 2003; Landau, *et al.* 2005; Ashkenazy, *et al.* 2010; Celniker, *et al.* 2013; Ashkenazy, *et al.* 2016]を活用して、各残基の保存度をスコアリングした。保存度が最も高いと評価されたアミノ酸残基グループを、SLC2A1の立体構造 (PDB ID: 5EQI) およびSLC2A3の立体構造 (PDB ID: 4ZW9) にPyMOL (<https://pymol.org>) を用いてマッピングした。

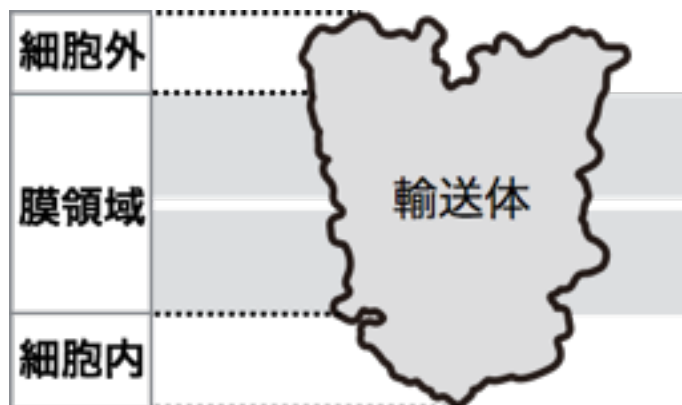
1.2.6 トポロジー領域の定義

膜タンパク質のトポロジーは、本来、膜タンパク質の膜に対する位置情報を説明する用語である[Melén, *et al.* 2003]。例えば、タンパク質のN末端やC末端が膜に対して内側・外側のどちらに位置しているか、また、膜貫通領域の数や開始・終了位置はどこか、という説明がトポロジー情報には含まれる。膜タンパク質のトポロジー情報はデータベースPDBTM (Protein Data Bank of Transmembrane Proteins) [Tusnady, *et al.* 2004; Kozma, *et al.* 2005; Kozma, *et al.* 2013] や、TOPDB (Topology Data Bank of Transmembrane Proteins Topology, Structure and Prediction) [Tusnady, *et al.* 2008; Dobson, *et al.* 2015]に蓄積されている。しかし、これらのデータベースはPDBのデータを基にしているため、SLC輸送体の登録数はまだ30件程度しかない。一方で、膜タンパク質の配列情報からトポロジーを予測する手法も開発されている。本研究では、残基ごとにこれら既存手法、Philius [Reynolds, *et al.* 2008],

PolyPhobius [Kall, *et al.* 2005], SPOCTOPUS [Viklund, *et al.* 2008], OCTOPUS [Viklund, *et al.* 2008] and SCAMPI [Bernsel, *et al.* 2008]のコンセンサスをとることができるトポロジー予測プログラムTopCons2 [Tsirigos, *et al.* 2015]を採用し、SLC輸送体の配列を「細胞外」、「膜領域」、「細胞内」の三つの領域に分類した。

しかし、トポロジー情報を用いてSLC輸送体の構造特性と変異の関係を探るためには、従来のような細胞外・膜領域・細胞内の三分類では不十分であると考えた。なぜなら、上述の実験（アミノ酸保存度の算出および保存残基の立体構造マッピング）の結果から、アミノ酸の保存度は膜領域と細胞内の境界において特に高いことが分かったためだ。つまり、従来のトポロジー分類法では、特徴的な部位が細胞内と膜領域に分断されてしまい、保存部位の性質を上手く捉えられない恐れがあった。したがって、TopCons2の結果から、新たに五つのトポロジー領域R1-R5を再定義した（図1-2）。

A



B

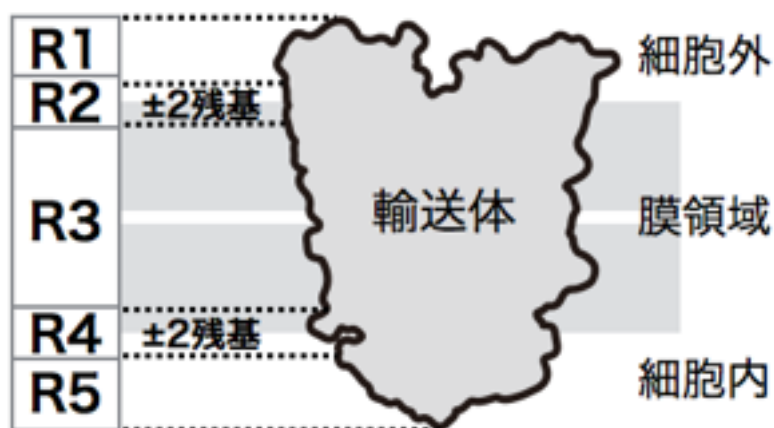


図1-2 トポロジー領域の定義

A, 従来のトポロジー分類法 本研究で使用したトポロジー予測プログラムTopCons2をはじめとして、従来の予測法は細胞外・膜領域・細胞内の三分類が主流である。

B, 本研究で採用したトポロジー分類法 従来の三分類法の結果を踏まえ、さらに細胞外と膜領域の境界面および膜領域と細胞内の境界面に位置する残基の特徴を反映させるべく、トポロジー予測をR1-R5の五分類に改変した。R2は三分類法で得られた細胞外と膜領域の境界面の ± 2 残基、R4は同様に膜領域と細胞内の境界面の ± 2 残基と定めた。

1.2.7 構造および変異情報の統合解析

ここでは、疾患に深く関連する変異が、立体構造上のどのような場所に多く位置するのか明らかにするため、先に定義したトポロジー領域R1-R5ごとの変異頻度を算出した。疾患に関連する変異の定義に関しては、ClinVar [Landrum, *et al.* 2014; Landrum, *et al.* 2016] に記された変異の臨床的意義 (ClinVarにおいては‘Clinical significance’と表現される) の記述を参考にして判断した。ClinVarの臨床的意義の項目において、‘pathogenic’ もしくは ‘risk factor’ という用語を含む変異を「疾患変異」、‘benign’ という用語を含む変異を「非疾患変異」と見なした。

疾患変異とトポロジー領域の関係を探るべく、データベースiMusta4SLCに登録した全ての変異情報を用いて、各トポロジー領域 R_i ($i = 1, 2, \dots, 5$)における疾患変異の頻度を次のように計算した。

$$f_{p,R_i} = \frac{C_{p,R_i}}{C_p}$$

ここで、 C_{p,R_i} はトポロジー領域 R_i における疾患変異のカウント数を表し、 C_p は疾患変異の総数を表す。同様に、トポロジー領域 R_i における非疾患変異の頻度は、

$$f_{n,R_i} = \frac{C_{n,R_i}}{C_n}$$

のように計算した。また、トポロジー領域 R_i における変異の期待頻度は、

$$f_{\text{exp},R_i} = \frac{C_{\text{res},R_i}}{C_{\text{res}}}$$

のように計算され、ここで、 C_{res,R_i} はトポロジー領域 R_i におけるアミノ酸残基の総数を、 C_{res} は全SLC輸送体配列のアミノ酸残基数を示す。最後に、トポロジー領域 R_i に

における疾患変異頻度の対数オッズ比は $\log_2 (f_{p,Ri} / f_{exp,Ri})$ と表し、非疾患変異の場合は同様に $\log_2 (f_{n,Ri} / f_{exp,Ri})$ とした。

1.2.8 アミノ酸置換パターンと頻度の可視化

トポロジーと変異の関係を統合解析するにあたり、具体的にどのようなアミノ酸置換パターンの頻度が多いのか確認するために、アミノ酸置換パターンの頻度を示すヒートマップを出力した。SLCファミリー、特定のトポロジー領域および疾患の臨床的意義を指定した上で、データベースから該当する変異情報を検索し、アミノ酸置換パターンを集計した。検索でヒットした変異の総数に対する割合を算出し、頻度の高いアミノ酸置換パターンほど濃い青で示されるようなヒートマップを生成した。縦軸に記されたアミノ酸はオリジナル（変異ではない）のアミノ酸、横軸に記されたアミノ酸は置換後のアミノ酸を表している。

1.3 結果および考察

1.3.1 iMusta4SLCの概要

iMusta4SLC (Integrated MUtational and STuctural Analysis FOR SLC) のウェブサイトは (URL: <http://cib.cf.ocha.ac.jp/slc/>) よりアクセスすることができる。参考として、Topology & Mutationのページを図1-3に示した。本データベースに登録されたSLC輸送体は、先行研究[Hediger, *et al.* 2013]およびHGNCによって定められた通り、52ファミリーに渡る。アイソフォーム (スプライスバリエント) を含むSLC転写物の総数は573であった。

ClinVarに登録されたSLC輸送体のミスセンス変異は1,216であった。これらの変異の臨床的意義の内訳は以下のようになっていた; Risk factor, 10 (0.8%); Pathogenic, 524 (44.4%); Likely pathogenic, 122 (10.3%); Benign, 64 (5.4%); Likely benign, 55 (4.7%); Uncertain significance, 309 (26.2%); not provided, 78 (6.6%); and Conflicting interpretations, 17 (1.4%)。Pathogenicの数が最も多い点については、現在の遺伝子配列収集制度・仕組みでは、疾患を呈する患者のDNA配列ほど多く収集されるようなバイアスがかかっている可能性が考えられる。疾患に関しては、SLC遺伝子における疾患変異は少なくとも100種類以上の疾患を引き起こすことが明らかになっており、その中には、代謝疾患や神経疾患、先天性異常、難聴、発達障害、免疫不全、腎不全、がんなどが含まれていた。ヒトSLC52ファミリーのうち、立体構造がすでにPDBに登録されているものはわずか4ファミリー (SLC1, SLC2, SLC4, SLC25) のみであった (表1-1)。しかし、その他26種類のファミリーは、少なくとも一つの配列が、PDB中のヒト以外のタンパク質構造に対して、20-30%の配列一致度を示した。

Topology & Mutation

Specify Topological Region → Find Mutations → Get Substitution Heatmap

Mutation Data

SLD name	Type	Significance	Criteria	Disease	Seq.start	Seq.end	Seq.original
SLD040	Measles	Conflicting interpretations of pathogenicity	criteria provided, conflicting interpretations	not specified provided	888	888	C
SLD141	Measles	risk factor, last reviewed: Apr 1, 2006	no assertion criteria provided	Burkitt's lymphoma	1827	1827	S
SLD142	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	197	197	S
SLD143	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	197	197	S
SLD144	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	197	197	S
SLD145	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	436	436	C
SLD146	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	436	436	C
SLD147	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	436	436	C
SLD148	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	514	514	A
SLD149	Measles	Uncertain significance, last reviewed: Jun 16,	criteria provided, single submitter	Hypochromic microcytic anemia with iron overload	514	514	A

Showing 1 to 10 of 1,750 entries

source: Higuchi, A. *et al.* (2018)

図1-3 iMusta4SLCのウェブサイト (Topology & Mutationのページ)

ブラウザ上部には各ページにつながるメニュー、左側には検索条件の設定領域、右側には結果の表示領域が設置されている。検索の結果表示されるテーブルにおいて、さらにキーワード検索や並び替えにより情報を絞り込むことも可能である。テーブル内のデータをExportボタンからCSVもしくはExcel形式でダウンロードすることにより、そのままユーザー独自の解析用データとして用いることもできる。

表 1 - 1 立体構造が決定されたヒトSLC輸送体

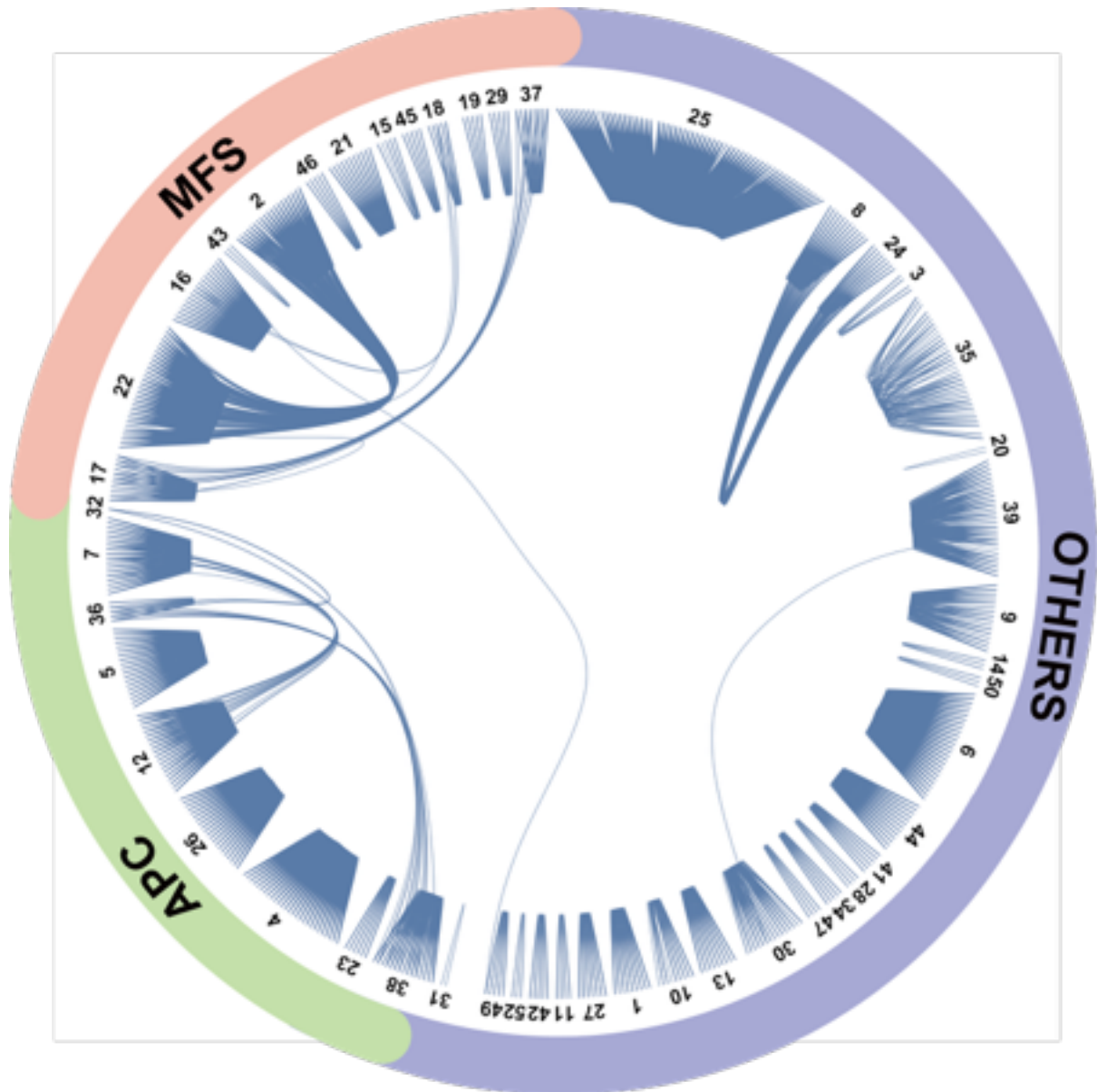
SLC遺伝子	タンパク質名	PDB ID	解像度	状態	参考文献
<i>SLC1A3</i>	Excitatory amino acid transporter 1 (EAAT1)	5LLM	3.25	The thermostabilized EAAT1 cyst mutant in complex with L-ASP and the allosteric inhibitor UCPII01	Canul-Tec, J.C., et al. Nature (2017).
		5LLU	3.32	The thermostabilized EAAT1 cyst-II mutant in complex with L-ASP and the allosteric inhibitor UCPII01	
		5LM4	3.1	The thermostabilized EAAT1 cyst mutant in complex with the competitive inhibitor TFB-TBOA and the allosteric inhibitor UCPII01	
		5MJU	3.71	MFS transporter in complex with inhibitor (Inward-open)	
		5EQG	2.90	MFS transporter in complex with inhibitor (Inward-open)	
<i>SLC2A1</i>	Glucose transporter 1 (GLUT1)	5EQH	2.99	MFS transporter in complex with inhibitor (Inward-open)	Kapoor, K., et al. Proc. Natl. Acad. Sci. USA (2016).
		5EQI	3.00	MFS transporter in complex with inhibitor (Inward-open)	
		4PYP	3.17	Human glucose transporter GLUT1 (Inward-open)	
<i>SLC2A3</i>	Glucose transporter 3 (GLUT3)	4ZW9	1.50	Human GLUT3 bound to D-glucose (outward-occluded)	Deng, D., et al. Nature. (2014).
		4ZWB	2.40	Maltose-bound human GLUT3 (outward-occluded)	
		4ZWC	2.60	Maltose-bound human GLUT3 (outward-open)	
		5C65	2.65	Human glucose transporter GLUT3 / SLC2A3 (outward-open)	
<i>SLC4A1</i>	Anion Exchanger 1	4KY9	2.23	The Cytosolic Domain of Human Anion Exchanger 1	Pike, A. C. W. et al. (To Be Published)
		4H87	1.55	FHA domain of kanadapain (SLC4A1/AP)	
<i>SLC4A8</i>	Sodium-driven chloride/bicarbonate exchanger (NDCBE)	5JHO	2.8	Regulatory domain of the sodium driven chloride bicarbonate exchanger.	Alvadia, C.M., et al. Sci. Rep. (2017).
		4P5W	2.4	The N- and C-terminal domain fusion of the human mitochondrial aspartate/glutamate carrier Citrin in the calcium-bound state	
<i>SLC25A13</i>	Mitochondrial aspartate/glutamate carrier	4P5X	2.26	The N-terminal domain of the human mitochondrial aspartate/glutamate carrier Aralar in the calcium-bound state	Thangaratnarajah, C., et al. Nat Commun. (2014).
		4P60	2.4	The N-terminal domain of the human mitochondrial aspartate/glutamate carrier Aralar in the apo state	
		4ZCU	2.1	Calcium-bound regulatory domain of the human ATP-Mg/Pi carrier in the P2 form	
		4ZCV	2.8	Calcium-bound regulatory domain of the human ATP-Mg/Pi carrier in the P2 2 21 form	
<i>SLC25A24</i>	Mitochondrial ATP-Mg/Pi carrier	4N5X	2.1	N-terminal calmodulin-like Calcium sensor of human mitochondrial ATP-Mg/Pi carrier SCaMC1	Yang, Q., et al. Structure. (2014).

1.3.2 SLC輸送体の配列類似度の算出と可視化

各SLC輸送体のアミノ酸配列同士を比較し、類似度を可視化したものを図1-4に示した。アミノ酸配列の一致度が高い輸送体同士は、立体構造や動的構造変化について共通点が多く、遺伝子変異による影響も類似する可能性がある。配列類似度は、前述の通り、FASTA36 [Pearson, *et al.* 1988; Pearson, *et al.* 2016] により算出し、全てのSLC輸送体のアミノ酸配列同士の配列類似度を調べた。ここでは、E-valueの閾値を $1.0E-5$ とし、それより低いE-valueをもつSLC輸送体ペアをホモログと見なした。

各SLC輸送体をファミリーごとにグループ分けして円状に並べ、ホモログのSLC輸送体のペアが青い線で結ばれるような図を作成した。この際、ドメインMFS (Major facilitator superfamily)、もしくはドメインAPC (Amino acid polyamine organocation) を保有しているファミリーと、その他のファミリーでグループ分けし、類似度が高いと予想されるファミリーが近くに配置されるようにした。

MFSを保有するファミリーの中では、まずSLC2 (グルコース輸送体) とSLC22 (カルニチン輸送体) の間において、E-valueが $1.0E-5$ 以下であり、配列類似度が非常に高いことが分かった。また、SLC37 (糖-リン酸交換体) とSLC17 (小胞グルタミン酸輸送体) の輸送体同士も全てではないが、いくつかはE-valueが $1.0E-5$ 以下のホモログであることが分かった。APCを保有するファミリーにおいては、SLC32 (アミノ酸輸送体)、SLC36 (H^+ カップリング型アミノ酸輸送体)、SLC38 (Na^+ カップリング型アミノ酸輸送体) の三つのファミリー間で類似度が高いことが分かった。また、SLC7 (カチオン性アミノ酸輸送体) とSLC12 (Cl^- 共輸送体) の間でも一部の輸送体同士のE-valueが $1.0E-5$ を下回っていた。その他のファミリーに関しては、SLC8 (Na^+/Ca^{2+} 交換体) とSLC24 ($Na^+/Ca^{2+}, K^+$ 交換体) に属する輸送体は、互いに類似度が高いことが分かった。



source: Higuchi, A. *et al.* (2018)

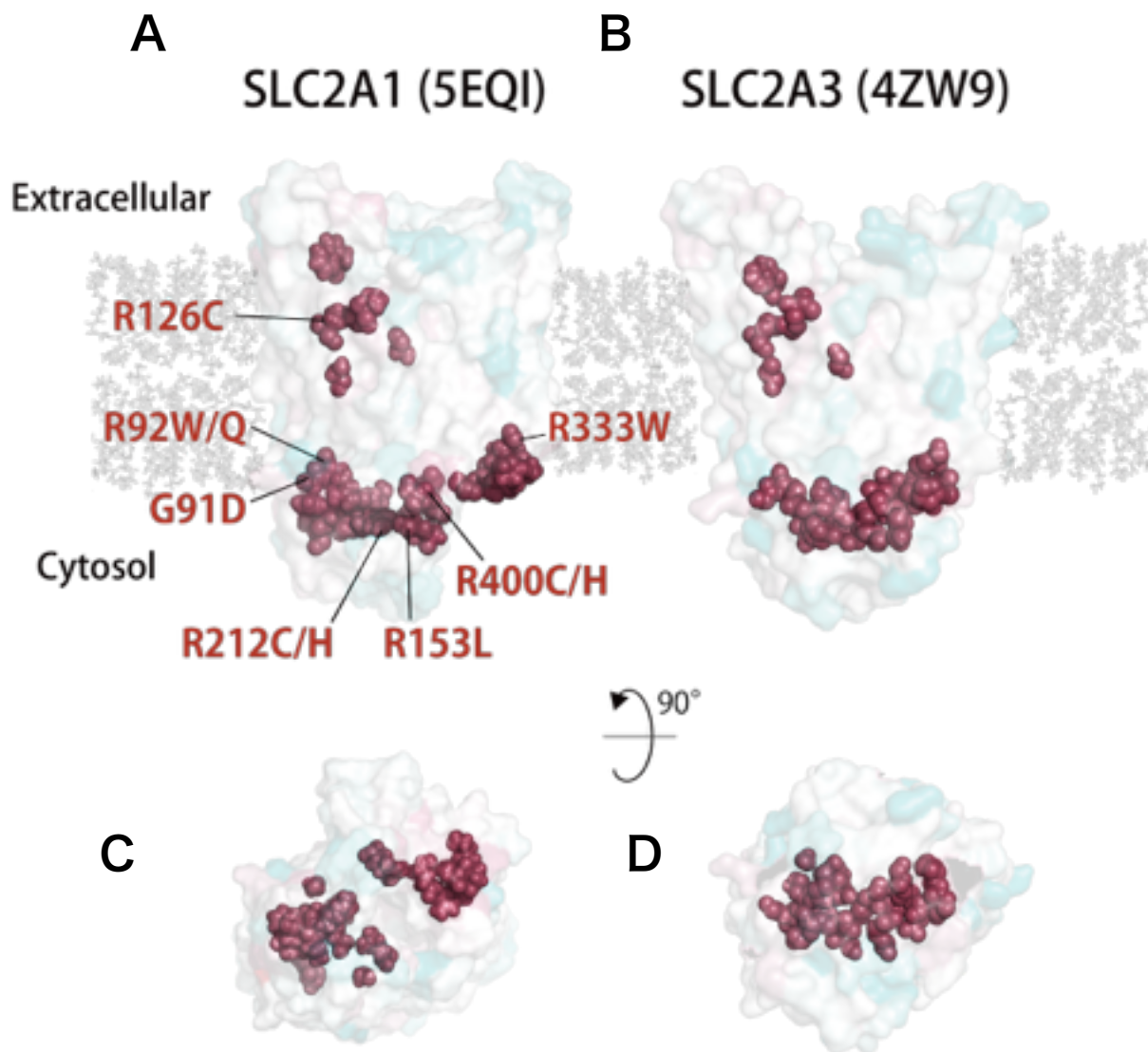
図1-4 SLC輸送体のアミノ酸配列類似度

SLC輸送体の配列類似度を円型マップにより示した。まず、SLC輸送体は三種類のグループ、1) Major Facilitator Superfamily (MFS) ドメインをもつSLC輸送体、2) Amino Acid-Polyamide-Organocation (APC) ドメインをもつSLC輸送体、3) その他の輸送体、に分類された。各SLC輸送体は円型に並べられたノードとして表現し、上記の三種類のグループに分類後、さらにファミリーごとに分類して配置した。類似度が高いSLC輸送体同士は青い線で結ばれている。この図で示す類似度の閾値は、FASTA36による計算で得られたE-valueが 1.0×10^{-5} 以下（データサイズ：572配列、全319,831残基）と設定している。

1.3.3 SLC輸送体における保存残基の探索およびマッピング

次に、400種類以上同定されたSLC輸送体のうち、具体的な例を取り上げながら、さらに詳しく変異情報と構造特性の関係を見ていきたい。立体構造および疾患変異が既に報告されている代表的なファミリーとして、SLC2（グルコース輸送体）が挙げられる。X線結晶構造解析の結果から、SLC2に属する二種類のタンパク質SLC2A1およびSLC2A3の立体構造は、それぞれ異なる開閉状態で報告されている。SLC2A1は、「内開き構造」（細胞質側に空洞を有する構造状態）（PDB ID：5EQI）[Kapoor, *et al.* 2016]、およびSLC2A3は「外開き構造」（細胞外側に空洞を有する構造状態）（PDB ID：4ZW9）[Deng, *et al.* 2015]を有する。

ここではまず、疾患変異と相関をもつ可能性のある配列保存度を考慮に入れ、構造特性との関係を検証し、よく保存された残基がSLCタンパク質の表面上に集まっていることを見出した。SLC2およびSLC22において高度に保存された残基を上述の二つの立体構造上にマッピングしたところ、高度に保存された残基の大部分が、膜領域と細胞質の境界面であるR4領域周辺で見出された（図1-5AB）。さらに、R4の保存された残基は、特に外開き構造上で一つのクラスターを形成した（図1-5B）。クラスターは内開き構造では4つのクラスターに分割され、中央にスペースを形成する（図1-5C）。この空間は、おそらく基質の経路であると考えられるが、外向き構造においては、この部分は完全にふさがれていた（図1-5D）。



source: Higuchi, A. *et al.* (2018)

図1-5 保存残基のマッピング

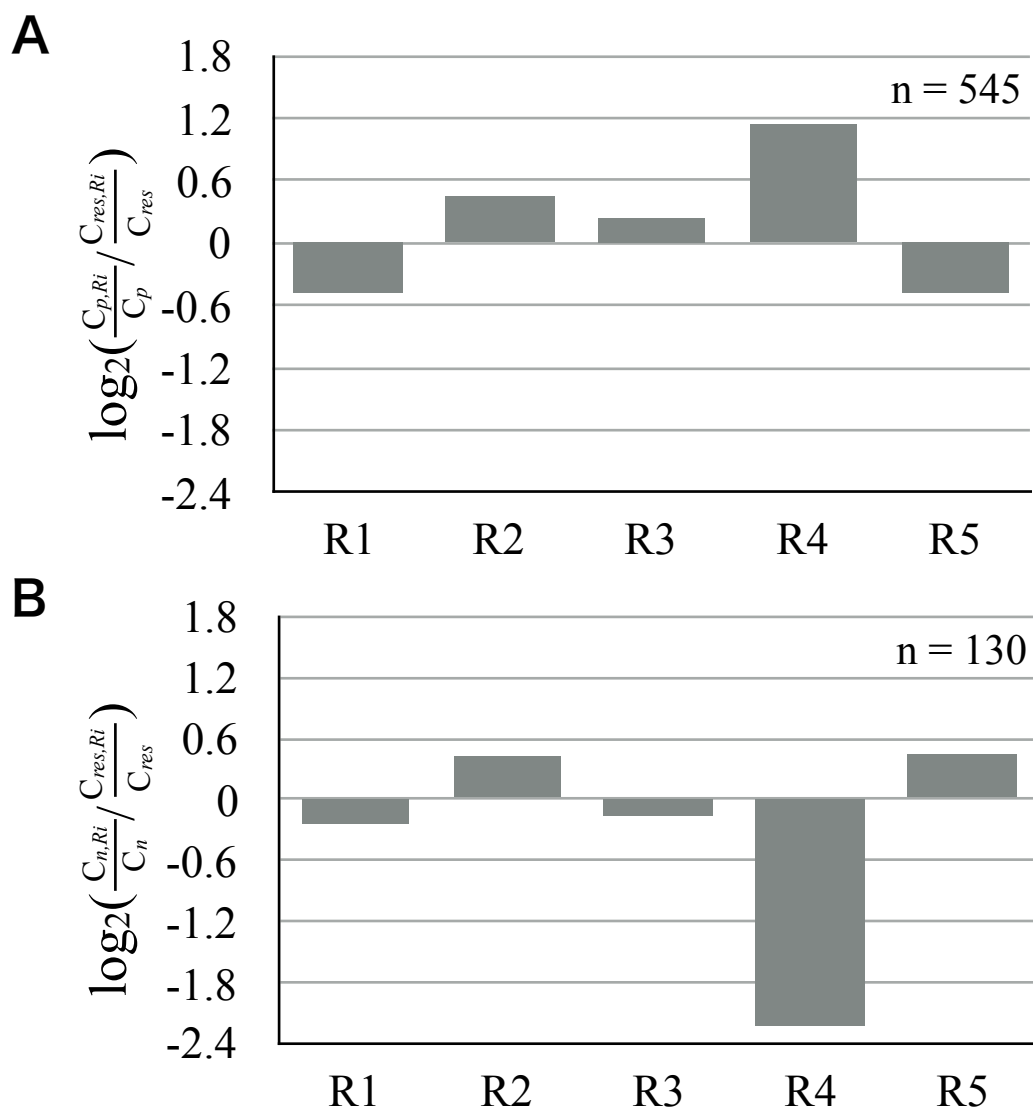
SLC2およびSLC22において高度に保存された残基を、SLC2A1の立体構造（PDB ID: 5EQI）およびSLC2A3の立体構造（PDB ID: 4ZW9）にマッピングし、赤い球体で示した。保存度の計算はConSurf [Glaser, *et al.* 2003; Landau, *et al.* 2005; Ashkenazy, *et al.* 2010; Celniker, *et al.* 2013; Ashkenazy, *et al.* 2016] により算出し、最も保存度が高いと評価されたグループに含まれる残基を保存残基とみなした。また、ClinVarにおいて報告されているSLC2A1の疾患変異を赤字で示した。

1.3.4 トポロジー領域ごとの変異頻度

まず、各トポロジー領域、つまりR1-R5における疾患変異と非疾患変異の発生頻度を比較した。iMusta4SLCには、545の疾患変異および130の非疾患変異が登録されていた。疾患変異に関する情報が、非疾患変異の4倍以上登録されている点については、健常者の変異よりも、疾患変異が積極的に研究・収集・公開されているためであると考えられる。

各トポロジー領域における全SLC輸送体の総アミノ酸数および割合は、R1, 91,115 (28.4%); R2, 22,400 (7.0%); R3, 87,652 (27.3%); R4, 23,424 (7.3%); R5, 96,527 (30.1%)であった。また、疾患変異の場合は、R1, 111 (20.4%); R2, 52 (9.5%); R3, 176 (32.3%); R4, 88 (16.1%); R5, 118 (21.7%)であった。同様に、非疾患変異の場合は、R1, 31 (23.8%); R2, 12 (9.2%); R3, 32 (24.6%); R4, 2 (1.5%); R5, 53 (40.8%)という結果が得られた。

異なるタイプの突然変異の分布を、これらの値の対数オッズ比を用いて比較した(図1-6)。その結果、明らかにR4領域が特異な性質を有することが示された。疾患変異は、R4領域に高頻度で局在しており、逆に非疾患変異はR4領域ではまれであった。つまりこの結果は、R4領域の変異はほとんど病原性であることを意味している。R4領域におけるこの偏った疾患変異の分布は、統計的に有意であり、カイ二乗分析(疾患、非疾患でそれぞれ $p < 2.04 \times 10^{-19}$ および 1.08×10^{-2}) によって統計検定を行った。



source: Higuchi, A. *et al.* (2018)

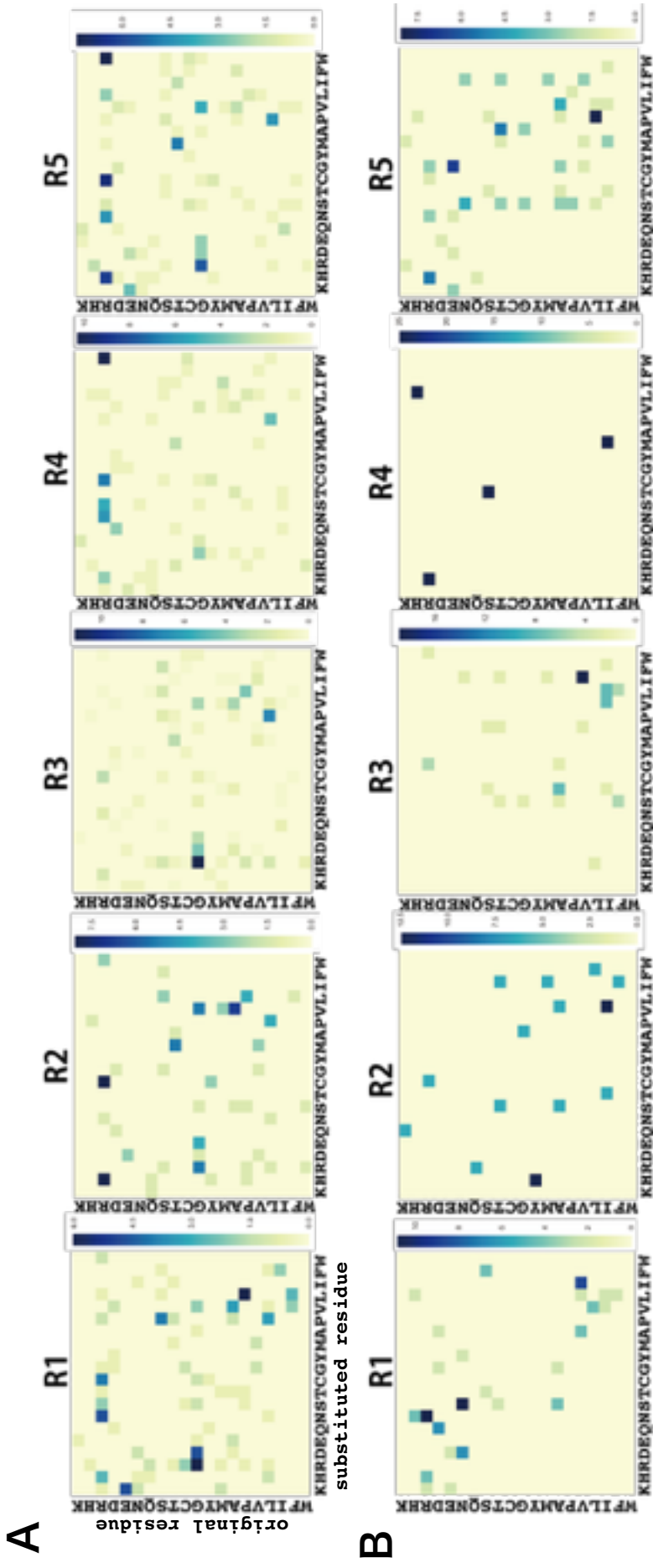
図1-6 トポロジー領域R1-R5ごとの変異頻度

A, 各トポロジー領域における疾患変異の頻度 疾患変異の観測数は545であった。
 B, 各トポロジー領域における非疾患変異の頻度 非疾患変異の観測数は130であった。
 変異頻度は、ログオッズ比により示した。期待度数は、全アミノ酸残基数に対する各トポロジー領域における残基数の比率を使用した。また、観測度数は全疾患変異数に対する各トポロジー領域における疾患変異数の比率を使用した。

1.3.5 トポロジー領域ごとのアミノ酸置換パターン

さらに、iMusta4SLCの「トポロジーおよび突然変異」のページにおいて、各トポロジー領域のアミノ酸置換パターンを分析することができる。それぞれのトポロジー領域における疾患変異および非疾患変異のヒートマップを比較した（疾患変異，1-7A；非疾患変異，1-7B）。また、ヒートマップの横のカラーバーは、割合を示している。R4領域の疾患変異におけるヒートマップ上で、アルギニンからトリプトファンへのアミノ酸置換が高頻度で起こっていた（カウント：11，割合：11.36%）（図1-7A, R4）。

膜領域と細胞質との間の境界面周辺のアルギニンからトリプトファンへの置換は、膜タンパク質のフォールディングおよびタンパク質安定性に重大な影響を与え得る。膜タンパク質のフォールディングに関して、細胞質領域の正に荷電した残基が細胞外領域の残基よりも多いことを示す‘positive inside rule’が提案されている[Heijne, 1986; Heijne, 1992; Baeza-Delgado, *et al.* 2013]。正に荷電した残基数の不均衡は、トポロジーのバランスを崩す。つまり、荷電残基を中性のものに置き換えることは、タンパク質トポロジーの形成に影響を及ぼす可能性があるのだ[Bogdanov, *et al.* 2009; Bogdanov, *et al.* 2014]。したがって、R4領域における疾患変異は、トポロジー形成に影響を及ぼす可能性が高いと考えられる。R4領域に位置するアルギニンはまた、極性分子である脂質および水分子のリン基と相互作用する。アルギニンからトリプトファンへの置換は、アルギニンの側鎖とリン基または水分子中の酸素原子との間の安定な極性相互作用を奪い、SLC輸送体が膜中で不安定化する可能性がある。タンパク質構造におけるこれらの摂動は、最終的にSLC輸送体の機能が妨害され疾患に至る過程における最初の段階であると考えられる。



source: Higuchi, A. *et al.* (2018)

図1-7 各トポロジ領域ごとのアミノ酸置換パターン頻度

各トポロジ領域における疾患変異もしくは非疾患変異のアミノ酸置換パターンの頻度を、ヒートマップにより示した。ヒートマップの縦軸は、野生型の残基、横軸は置換された残基を示す。濃い青は観測頻度が高いことを示す。

1.3.6 疾患変異と保存残基の位置関係

トポロジー領域R4およびその周辺では、疾患変異が高い割合で見出され、高度に保存された残基のクラスターが存在することが明らかになった。この相関関係の具体的な例は、iMusta4SLCでも参照することができる。表1-2は、SLC2A1の全疾患変異をリストアップしたものである。そして、これらの疾患変異の3分の1、すなわちGly91Asp、Arg92Trp/Gln、Arg126Cys、Arg153Leu、Arg212Cys/His、Arg333TrpおよびArg400Cys/Hisは、図1-5に示された保存された残基と重複している点に興味深い。SLC2と配列類似度が高いSLC22（カルニチン輸送体）においても同様に、Arg169やArg399が高度に保存されており、これらの残基はSLC2A1におけるArg92およびArg333に相当する。表1-2に列挙したSLC2A1の変異と同様に、SLC22でもArg169Trp/GlnおよびArg399Trp/Glnなどのアミノ酸置換が全身性カルニチン欠乏症という疾患を引き起こすことが報告されている。

これらの保存残基の機能はいったいどのようなものなのか。基質結合部位との関係はあるのか、という疑問に対しては、バクテリア由来のSLC2ホモロジー(XyleE)のD-グルコース結合部位から、ヒト由来のグルコース輸送体1（GLUT1）におけるD-グルコース結合部位残基が、Phe26、Gln161、Ile164、Gln282、Gln283、Asn288、Tyr292、Phe379、Gly384、Trp388、Asn411、Trp412であると報告されている[Hediger, *et al.* 2013]。同様に、GLUT1の阻害剤が結合するアミノ酸残基（Thr137、Gln161、Gln282、Asn288、Phe379、Gly384、Trp388、Asn411）がグルコースの結合部位とオーバーラップしていることも報告されている[Kapoor, *et al.* 2016; Deng, *et al.* 2015]。しかし、いずれも保存されたクラスターに含まれておらず、これらの保存残基の機能はいまだ不明である。

表 1-2 Clin Varに登録されたSLC2A1の疾患変異

Amino acids substitution		significance	Clinical phenotypes diseases	Position		allele ID		
position	original substituted			chr.	location		variation ID	
34	Asn	Ile	Pathogenic	GLUT1 deficiency syndrome 2	1	42943239	16115	31154
91	Gly	Asp	Pathogenic	GLUT1 deficiency syndrome 1	1	42931049	16110	31149
92	Arg	Trp	Pathogenic	GLUT1 deficiency syndrome 2	1	42931047	16119	31158
93	Arg	Trp	Pathogenic	GLUT1 deficiency syndrome 2	1	42930865	16117	31156
95	Ser	Ile	Pathogenic	GLUT1 deficiency syndrome 2	1	42930858	16116	31155
126	Arg	Cys	Pathogenic	Dystonia 9, GLUT1 deficiency syndrome 1, GLUT1 deficiency syndrome 2, Epilepsy, idiopathic generalized	1	42930766	16118	31157
126	Arg	His	Pathogenic	GLUT1 deficiency syndrome 1, Epilepsy, idiopathic generalized	1	42930765	16111	31150
126	Arg	Leu	Pathogenic	GLUT1 deficiency syndrome 1	1	42930765	16109	31148
130	Gly	Arg	Pathogenic	not provided	1	42930754	207190	201162
140	Val	Met	Pathogenic	not provided	1	42930724	372507	359361
153	Arg	Leu	Pathogenic	not provided	1	42930684	207227	201160
165	Val	Ile	Pathogenic	not provided	1	42930649	379258	365285
212	Arg	Cys	Pathogenic	Dystonia 9	1	42929918	37300	45874
212	Arg	His	Pathogenic	not provided	1	42929917	265386	259669
223	Arg	Trp	Pathogenic	not provided	1	42929885	207193	201157
223	Arg	Pro	risk factor	Epilepsy, idiopathic generalized	1	42929884	39702	48301
232	Arg	Cys	Pathogenic, risk factor	Epilepsy, idiopathic generalized	1	42929766	37301	45875
275	Ala	Thr	Pathogenic	GLUT1 deficiency syndrome 1, GLUT1 deficiency syndrome 2	1	42929637	16114	31153
283	Gln	Arg	Pathogenic	not provided	1	42929612	207197	201152
286	Gly	Asp	Pathogenic	Stomatin-deficient cryohydrocytosis with neurologic defects	1	42929603	218333	215041
295	Thr	Met	Pathogenic	GLUT1 deficiency syndrome 1	1	42929298	207229	201150
313	Ser	Pro	Pathogenic	not provided	1	42929245	280423	264006
314	Gly	Ser	Pathogenic	GLUT1 deficiency syndrome 1, GLUT1 deficiency syndrome 2, Epilepsy, idiopathic generalized	1	42929242	16113	31152
324	Ser	Leu	Pathogenic	not provided	1	42929211	207201	201147
330	Arg	Ter	Pathogenic	not provided	1	42929018	207196	201142
333	Arg	Trp	Pathogenic	GLUT1 deficiency syndrome 1, Epilepsy, idiopathic generalized	1	42929009	198842	196002
400	Arg	His	Pathogenic	not provided	1	42927684	280046	264004
400	Arg	Cys	Pathogenic	not provided	1	42927685	207212	201132
411	Asn	Ser	risk factor	Epilepsy, idiopathic generalized	1	42927651	96709	102598
458	Arg	Trp	risk factor	Epilepsy, idiopathic generalized	1	42927148	96708	102597
468	Arg	Trp	Pathogenic	GLUT1 deficiency syndrome 1	1	42927118	16120	31159

source: Higuchi, A. *et al.* (2018)

- 第2章 -

第2章 アミノ酸置換がグルコース輸送体の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検証

2.1 序

前章において、SLC輸送体の膜領域と細胞質の境界領域では、疾患変異が有意に高い頻度で存在することを明らかにした（図1-6）。その領域において、最も高い頻度で観測される疾患変異パターンは、アルギニンからトリプトファンへの変異であった（図1-7）。さらに、互いに相同性の高いファミリーであるSLC2およびSLC22の配列を比較した結果、膜領域と細胞質の境界面においてアミノ酸残基が高度に保存されており、その多くがアルギニンであることも明らかにした（図1-5）。これらの結果から、膜領域と細胞質の境界面において保存された残基は、SLC輸送体の立体構造の形成・維持もしくは機能に関する重要な役割を担っていると考えられる。

このようにSLC輸送体の膜領域と細胞質の境界に多く観測されている疾患変異は、輸送体の立体構造にどのような影響を及ぼし、疾患を引き起こしているのだろうか。多くの変異情報は、疾患名や症状などの臨床的意義が報告されていたとしても、タンパク質の動的構造に与える影響は明らかになっていない。膜タンパク質は脂質二重膜に埋め込まれた環境下でなければ構造を正常に保つことが難しいことから、実験的にその動的構造を観察することも容易ではない。

そこで本章では、膜領域と細胞質の境界において保存された残基、特にアルギニンにアミノ酸置換が起きた場合、SLC輸送体の構造や機能にどのような影響を与えるのかを計算科学的な手法を用いて検証することを目的とする。一残基置換の構造および機能への影響を解析する手法として、分子動力学シミュレーションを用いる。

分子動力学シミュレーションとは、ニュートンの運動方程式に基づいて、原子や分子の動きをコンピューターにより再現する手法である。まず、原子の座標を定めた立体構造情報をもとにシミュレーションシステムを構築し、原子にはたらく様々な力の合計を計算する。ここでは原子にたらく力とは、大きく分けて結合性相互作用と非結合性相互作用の二種類に分かれる。結合性相互作用とは、1) 結合の伸縮振動、2) 結合の変角振動、3) 面外変角、4) ねじれ角の四種類が挙げられ、非結合性相互作用は、1) ファンデルワールス力、2) 静電気力（クーロン力）の二種類が考慮される。そして、算出された力から、次のタイムステップにおいて原子がどれほど動くかをニュートンの運動方程式に基づき計算することで、分子の動きやエネルギーを明らかにする計算科学的手法である。通常、時間間隔（タイムステップ）は1もしくは2フェムト秒で設定されることが多い。分子動力学シミュレーションを行うためのソフトウェアはCHARMM（Chemistry at HARvard Macromolecular Mechanics）（www.charmm.org）、NAMD（www.ks.uiuc.edu/Research/namd/）、myPresto（Medicinally Yielding PRotein Engineering SimulaTOR）（<http://presto.protein.osaka-u.ac.jp/myPresto4/>）、GROMACS（www.gromacs.org）[Berendsen, *et al.* 1995]などが代表的である。

膜タンパク質の動態解析に分子動力学シミュレーションが用いられ始めたのは、1990年代中頃である。水溶性タンパク質の分子動力学シミュレーションと比較すると、その歴史は20年ほど浅い[Iwata, 2013; Yao, *et al.* 2010]。しかし、X線結晶構造解析では明らかにできない膜タンパク質の動的な性質、つまりダイナミクスを分析できるとして、これまで多くの先行研究で使用され、成果が上げられてきた。実際にグルコース輸送体の動態を分子動力学シミュレーションにより調べた先行研究 [Dwyer, *et al.* 2001] には、Mohanらが行ったGLUT4（GLUT1の構造データから作成したホモ

ロジモデルを使用)の研究[Mohan, *et al.* 2010] や、Fuらが行ったGLUT1 (構造は4PYPおよび4GBZを使用し、6種類の異なる構造状態を再現)の基質輸送メカニズム解析 [Fu, *et al.* 2016] がある。

しかし、膜タンパク質のフォールディングや小胞融合などの膜関連プロセスを含む複雑な生体分子現象のタイムスケールは、全原子分子動力学シミュレーションでは実行することが難しい[Ash, *et al.* 2004]。膜タンパク質は分子量が大きいことに加え、脂質二重膜・水分子・イオンを含めたシミュレーションシステムは原子数が多く、計算負荷が大きい。さらに先行研究から、SLC輸送体は立体構造変化をして開閉する速度が1ターンあたりおよそ1マイクロ秒かかることが報告されている。1マイクロ秒という時間は、先ほど述べたタイムステップから計算すると、数百億回分の計算を必要とし、現在の計算機の性能では、数ヶ月から数年を要するほどの長時間シミュレーションになるであろう。システムの大きさに加え、シミュレーション時間の観点からも、やはり計算負荷が非常に大きいことが課題となる。グルコース輸送体の分子動力学シミュレーションに関する先行研究において、シミュレーション時間は、長くても60ns程度であった。グルコース輸送体のターンオーバーに要する時間と比較すると、まだ十分ではない。

先行研究では、このような計算負荷の大きいシステムや長時間のシミュレーションを扱うために、システムや計算を簡素化する「粗視化」という技術が開発されてきた。いくつかの原子群を単一の粒子として処理する粗視 (coarse-grained, CG) モデルは、生体分子の長時間シミュレーションを実現させるための有望なアプローチとされている。粗視化技術や力場は、既に様々なものが開発されている。代表的な粗視化シミュレーションの分子力場として、Marrinkらが開発したMARTINI [Monticelli, *et al.* 2008; Periole, *et al.* 2009; Yesylevskyy, *et al.* 2010; de Jong, *et al.* 2012;

Arnarez, *et al.* 2015; Hsu, *et al.* 2016]がある。通常的全原子モデルの場合、各原子を一つの点とみなし、それぞれの点に座標が与えられる。一方、MARTINIを用いた粗視化モデルの場合、脂質および膜 に広範に適用されているCGモデル[Marrink *et al.*, 2004]では、各CG粒子は平均して4つの重原子（すなわち水素原子は入らない）を表す。多くの場合、アミノ酸骨格を1つの点とし、さらに側鎖をアミノ酸の種類により0から4つの点の組み合わせで表現する。

以上の先行研究を踏まえ、本章では、膜領域と細胞質の境界における疾患変異がSLC輸送体の動的構造に与える影響について、分子動力学シミュレーションにより明らかにする。GROMACS (v5.0.7) [Berendsen, *et al.* 1995]を用いて、タンパク質構造変化の可視化および比較や、構造の安定性の指標となるRMSD (Root mean square deviation)、各残基の振動幅を表すRMSF (Root mean square fluctuation)などを算出することができる。

2.2 方法

2.2.1 シミュレーションモデルの構築

本研究では、シミュレーション時間を短縮すべく、SLC2（グルコース輸送体）の二種類の立体構造、すなわち「内開き構造」と「外開き構造」を用意し、それぞれ分子動力学シミュレーションを行った。それにより、内開き構造をスタート地点とし、細胞質側の空洞が次第に閉じていく構造変化と、外開き構造をスタート地点とし、細胞外側の空洞が次第に閉じていく構造変化をそれぞれ観察することができると考えた。内開き構造にはヒトSLC2A1（グルコース輸送体1; GLUT1）のX線結晶構造解析により明らかにされた立体構造 (PDB ID: 5EQI) を用いた[Kapoor, *et al.* 2016]。また、SLC2A1の膜領域と細胞質の境界における七種類の点変異

(Gly91Asp、Arg92Trp、Arg92Gln、Arg93Trp、Arg93Gln、Arg333Trp、Arg334Gln) (表2-1) をPyMOLを用いてそれぞれ導入し、変異型立体構造モデルを作成した(図2-1A)。外開き構造には、ヒトSLC2A3(グルコース輸送体3; GLUT3)の同じくX線結晶構造解析により明らかにされた立体構造(PDB ID: 4ZW9)を用いた[Deng, *et al.* 2015]。4ZW9に導入した七種類の点変異(Gly89Val、Arg90Trp、Arg91Cys、Arg91His、Arg331Trp、Arg331Lys、R331S)(表2-1)は5EQIに導入した点変異と対応する位置に存在する(図2-1B)。

ヒトSLC2A1(GLUT1)とヒトSLC2A3(GLUT3)は配列一致度が67%のホモログである。両者ともグルコースを基質とするSLC2のメンバーであるが、局在場所やグルコースとの親和性がやや異なる。GLUT1は、体組織において広く発現され、細胞膜を介したグルコース取り込みに関与する[Macheda, *et al.* 2005]。一方、GLUT3は、特に神経において主要な役割を果たし、「ニューロン・グルコース・トランスポーター」と呼ばれることもある[Simpson, *et al.* 2008]。通常、グルコースが血液循環から神経細胞まで輸送されるには、まずGLUT1により血液脳関門(BBB)の内皮細胞を通過し、つぎにGLUT3により神経の原形質膜を通過する必要がある[Yan, *et al.* 2014]。他のGLUTと比較して、GLUT3は基質への高い親和性、高い活性、および長い半減期をもち、高いエネルギー需要を有する神経細胞にグルコースを供給するのに最適な特徴を有している[Yan, *et al.* 2014]。

さらに、本研究では、以下のような三種類のシミュレーションモデル

- 1) 全原子モデル (Atomistic model)
- 2) 粗視化モデル (Coarse-grained model)
- 3) 保存性弾性ネットワークモデル (Conserved elastic network model)

を構築し、観察に最適なシステムを比較検証した（図2-2）。これらのシステムでは、精度と計算時間がトレードオフの関係にあり、全原子は最も精度が高いものの計算コストが高く、保存性弾性ネットワークは最も計算負荷がかからない一方で精度が低くなる。全原子シミュレーションシステムは前述の通り、全ての原子に座標を与え、各タイムステップごとの原子にかかる力や動き、エネルギーなどを計算するものである。SLC2A1およびSLC2A3の立体構造に脂質二重膜、水分子、イオン（Na⁺、Cl⁻）を含めた全原子モデルは、CHARMM-GUIのMembrane Builder (<http://www.charmm-gui.org/?doc=input/membrane>) [Jo, *et al.* 2007; Jo, *et al.* 2009; Wu, *et al.* 2014; Lee, *et al.* 2016]を用いて生成した（図2-2A）。また、粗視化モデルも前述の通り、CHARMM-GUIのMARTINI Bilayer Maker (<http://www.charmm-gui.org/?doc=input/mbilayer>) [Qi, *et al.* 2015; Hsu, *et al.* 2017]を利用して生成した（図2-2B）。最後に、保存性弾性ネットワークは、粗視化モデルを基盤として、本研究で独自に開発したシミュレーションモデルである。本来の単純な弾性ネットワークモデルでは、N残基からなるタンパク質を、共有結合を無視したN個のC α 原子のみからなる点の集合と考える。この際、相互作用は、C α 原子の近傍の原子との間に結晶構造の位置では0、ゆらぎによる位置の変化 Δr に従い調和バネ（一様なバネ定数 γ ）で元に戻すような力 $-\gamma(\Delta r_1 - \Delta r_2)$ がかかるものとする。本研究においてモデルとして用いるSLC2A1 (PDB ID: 5EQI) およびSLC2A3 (PDB ID: 4ZW9) は、同じ基質・機能を有するホモログであるものの、それぞれのアミノ酸配列・二次構造・三次構造が一部異なる他、開閉状態も内開き構造と外開き構造で大きく異なっている。それらの違いを逆に利用し、開閉状態に関わらず両者で保存された構造や結合部位のみを抽出することで、構造変化の制約を取り払った新たな簡素化モデルを作成した（図2-2C）。内開き構造と外開き構造において、構造や結合が保存さ

れた部位を図2-2Dに示した。このように、構造変化によらず保存される部位のみを抽出することで、おおまかな構造変化の様子を捉えられるのではないかと考えた。以上の三種類のシミュレーションモデルを用いて、動態観察に最適なモデルをGROMACS (v.5.0.7) [Berendsen, *et al.* 1995]により検討した。

表 2-1 モデルに導入した変異

Variation name	Gene name	Data Source	Seq Start	Seq End	Seq From	Seq To	AA Start	AA End	AA From	AA To	Var ID	Allele ID	refSNP	Chr	Loc	Assembl	Significance	Disease
G91D	SLC2A1	ClinVar	272	272	G	A	91	91	Gly	Asp	16110	31149	rs80359814	1	42931049	GRCh38	Pathogenic	GLUT1 deficiency syndrome 1
R92W	SLC2A1	ClinVar	274	274	C	T	92	92	Arg	Trp	16119	31158	rs202060209	1	42931047	GRCh38	Pathogenic/Likely pathogenic	GLUT1 deficiency syndrome 2
R92Q	SLC2A1	dbSNP	275	275	G	A	92	92	Arg	Gln	NA	NA	rs779073410	1	42931046	GRCh38	NA	NA
R93W	SLC2A1	ClinVar	277	277	C	T	93	93	Arg	Trp	16117	31156	rs267607061	1	42930865	GRCh38	Pathogenic	GLUT1 deficiency syndrome 2
R93Q	SLC2A1	dbSNP	278	278	G	A	93	93	Arg	Gln	NA	NA	rs80359815	1	42930864	GRCh38	NA	NA
R333W	SLC2A1	ClinVar	997	997	C	T	333	333	Arg	Trp	198842	196002	rs80359825	1	42929009	GRCh38	Pathogenic/Likely pathogenic	GLUT1 deficiency syndrome 1, Epilepsy, idiopathic generalized
R334Q	SLC2A1	dbSNP	1001	1001	G	A	334	334	Arg	Gln	NA	NA	rs892715050	1	42929005	GRCh38	NA	NA
G89V	SLC2A3	dbSNP	266	266	G	T	89	89	Gly	Val	NA	NA	rs758117298	1	7932990	GRCh38	NA	NA
R91C	SLC2A3	dbSNP	271	271	C	T	91	91	Arg	Cys	NA	NA	rs756172777	1	7931484	GRCh38	NA	NA
R91H	SLC2A3	dbSNP	272	272	G	A	91	91	Arg	His	NA	NA	rs145936296	1	7931483	GRCh38	NA	NA
R331K	SLC2A3	dbSNP	992	992	G	A	331	331	Arg	Lys	NA	NA	rs770855736	1	7924486	GRCh38	NA	NA
R331S	SLC2A3	dbSNP	993	993	A	C	331	331	Arg	Ser	NA	NA	rs749200071	1	7924485	GRCh38	NA	NA

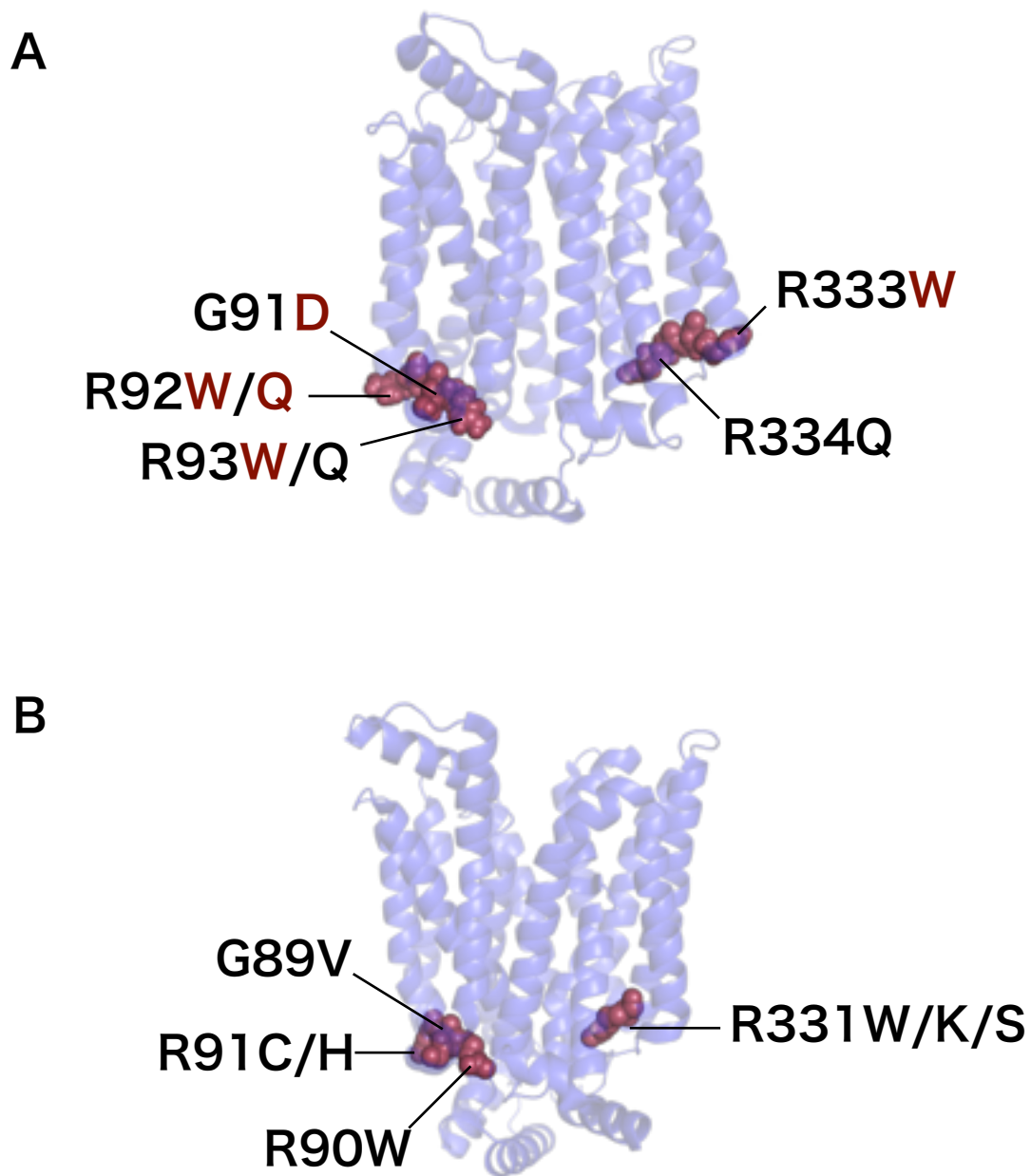


図2-1 シミュレーションモデルに導入した変異の立体構造上の位置

A, SLC2A1 (GLUT1) の内開き構造 (PDB ID: 5EQI) に導入した変異 ClinVarにおいて疾患変であると報告されているものに関しては、赤字で示している。

B, SLC2A3 (GLUT3) の外開き構造 (PDB ID: 4ZW9) に導入した変異 SLC2A3の変異に関しては、ClinVarにおける臨床的意義の報告がないため、疾患との関連は未知である。ただし、SLC2A1の変異導入部位と相同な残基部位に、実際にClinVarもしくはdbSNPにおいて報告された変異を中心として導入した。

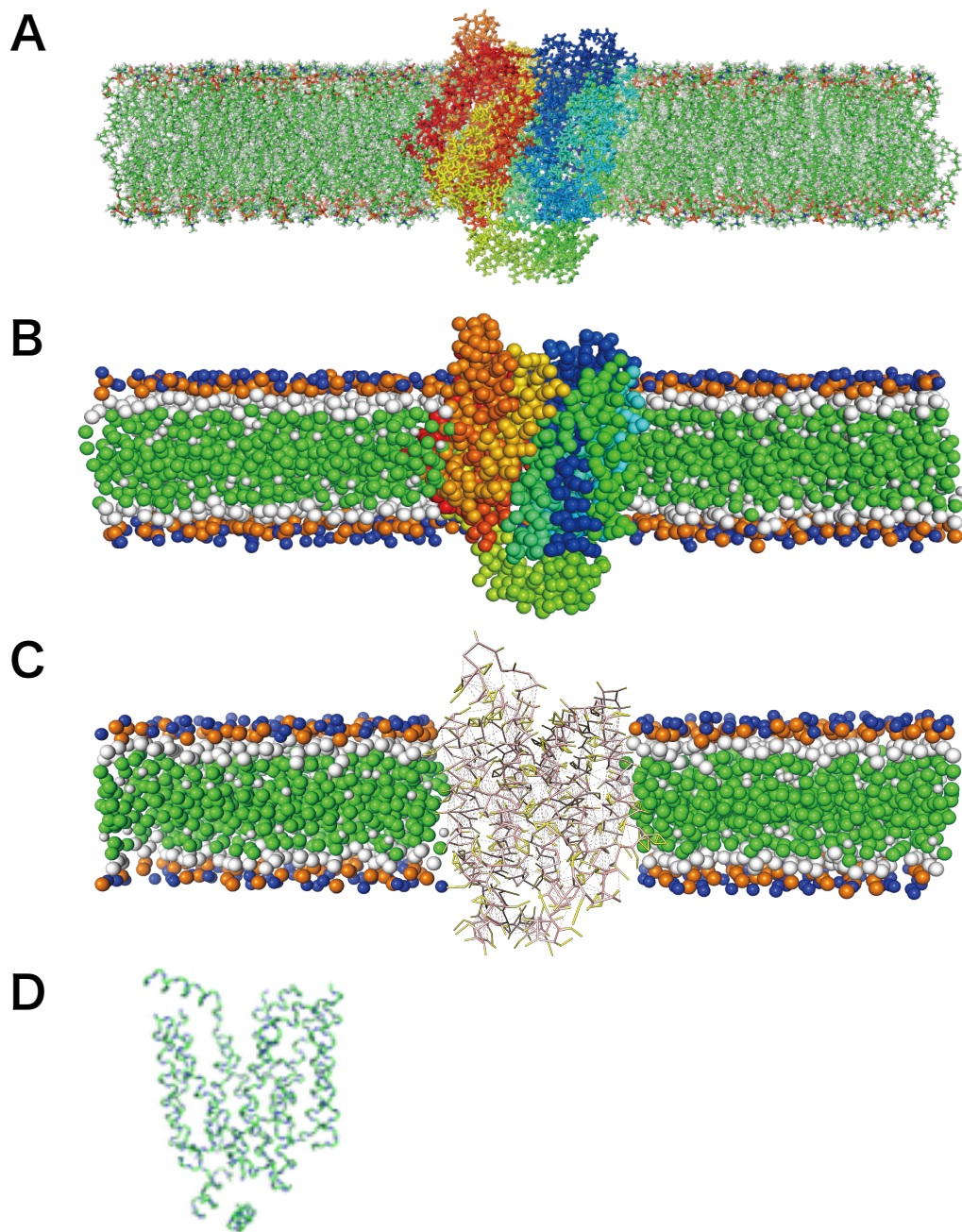


図2-2 本研究で用いたシミュレーションモデル

A, 全原子モデル 全ての原子を点とみなし、継時的に座標を計算する。

B, 粗視化モデル アミノ酸の骨格を一つの点とみなし、側鎖に関してはアミノ酸の種類に応じて0~4つの点の組み合わせで表現する。全原子モデルと比較すると、計算精度は低下するが、計算速度は50-100倍（本研究での観測）ほど高速化する。

C, 保存性弾性ネットワークモデル 粗視化モデルをさらに簡素化させることにより、計算速度および構造変化のしやすさを追求したモデル。内開き構造と外開き構造の両方で保存された構造および結合のみを抽出し、構造変化に対する制約を除いた。

D, 保存性弾性ネットワークモデルにおける保存部位 内開き構造と外開き構造の両方で保存された残基のCa原子を抽出した。

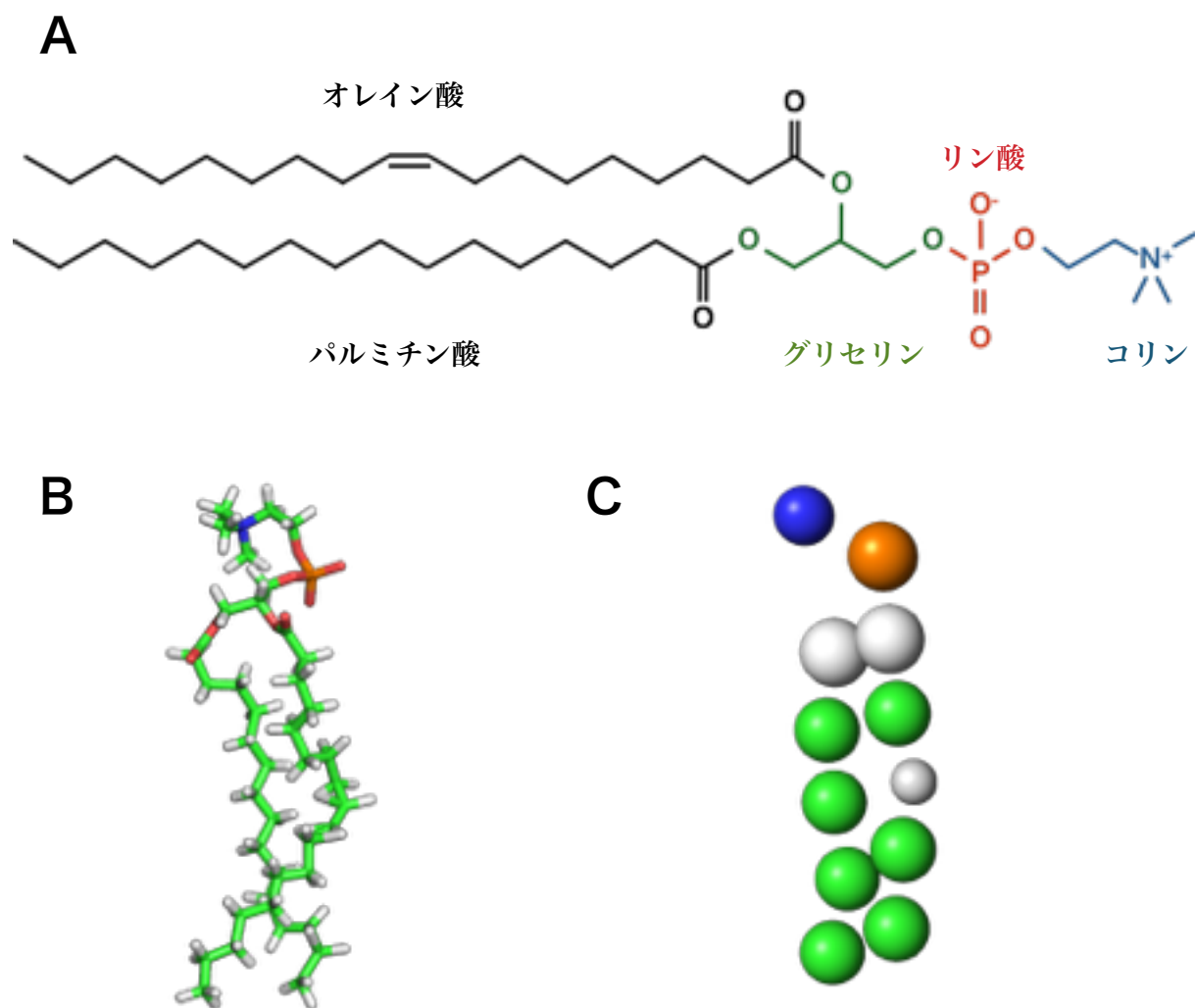


図2-3 本研究で用いた脂質分子POPC（ホスファチジルコリン）の構造

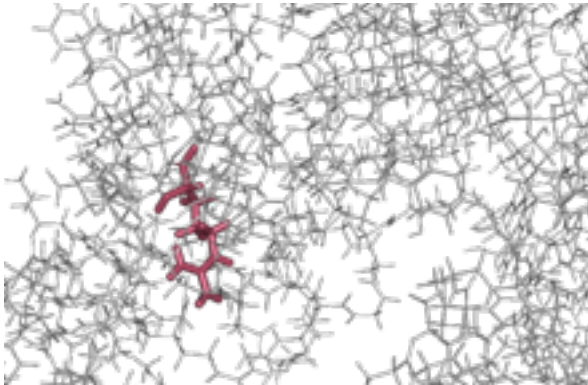
A, POPCの分子構造 疎水性のオレイン酸およびパルミチン酸により構成された尾部と、親水性のリン酸およびコリンにより構成された頭部がグリセリンにより結合された構造をもつ。

B, POPCの全原子モデル 各原子に座標が与えられている。上部が親水性の頭部、下部が疎水性の尾部となっている。

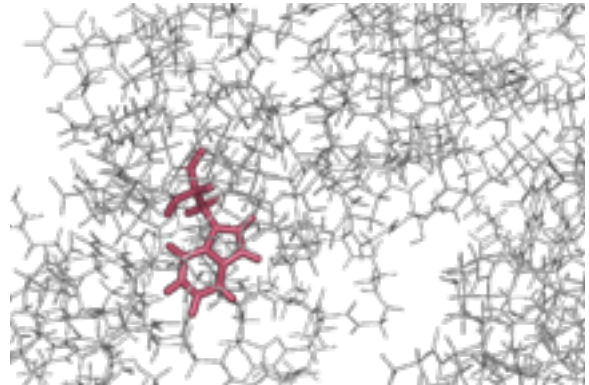
C, POPCの粗視化モデル 全原子モデルを簡素化したモデル。保存性弾性ネットワークモデルにおいても同様の脂質分子モデルを使用している。

A

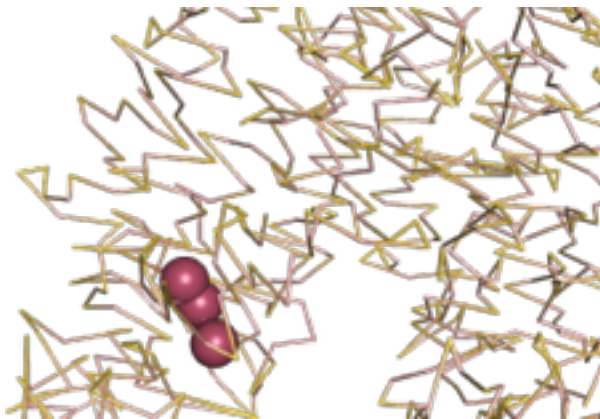
野生型



R92W

**B**

野生型



R92W

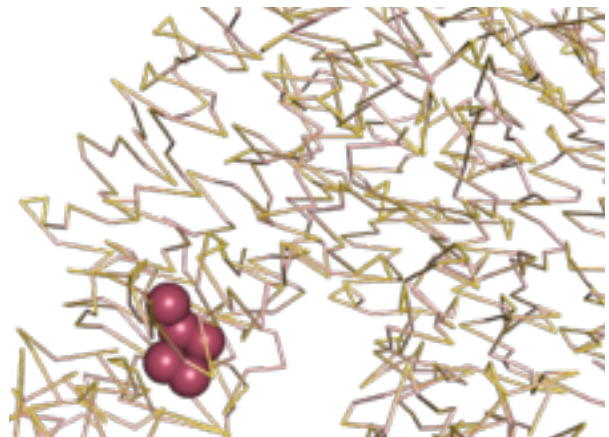


図2-4 変異を導入した残基の初期構造

A, 全原子モデルの変異導入部位の例 野生型における92残基目はアルギニンであるのに対し、右の変異モデルR92Wにおいては、トリプトファンに置換されている。

B, 粗視化モデルおよび保存性弾性ネットワークモデルにおける変異導入部位の例 アルギニン残基は、アミノ酸骨格を一つの球、さらに側鎖は二つの球を組み合わせることにより表現される。一方、トリプトファンの側鎖は、四つの球により構成される。

2.2.2 分子動力学シミュレーション

解析に用いる軌道計算に先立ち、CHARMM-GUI [Jo, *et al.* 2008]を利用し、まず以下のように入力ファイルの生成を行った。RCSB (PDBのウェブサイト) より、SLC2A1の立体構造座標 (PDB ID: 5EQI) およびSLC2A3の立体構造座標 (PDB ID: 4ZW9) を取得し、それぞれの座標の読み取る。つぎに、膜に関してタンパク質を適切に配向させることが必要になる。このステップでは、Zに沿ってタンパク質の主軸を整列させる。また、Z軸に沿ったタンパク質の断面積を計算する。さらに、システムサイズを決定する。システム形状は長方形を採用し、使用する脂質分子POPCの比は全体に対して10%、XYシステムの初期推定サイズを20nmと設定した。脂質分子の設定に関しては、POPCを脂質二重膜の要素として膜タンパク質の分子動力学シミュレーションを行っている先行研究に従った[Mohan, *et al.* 2010; Fu, *et al.* 2016; Kamaraj, *et al.* 2016] (図2-3)。Z軸方向のシステムサイズは、タンパク質の上部および底部からの水分量によって決定される。最終的なシミュレーションボックスのサイズは、 $19.49931 \times 19.49931 \times 10.44057$ (nm)となった。前のステップで決定されたシステムサイズに基づいて、このステップでは、

- 1) タンパク質周囲の脂質二重層
- 2) タンパク質を完全に溶媒和するための追加の水分子
- 3) イオン (モンテカルロサンプリングまたは距離に基づくアルゴリズム)

をそれぞれ構築する。そして、すべての要素 (タンパク質、脂質二重層、水、イオン) の座標が一つのシミュレーションボックス内に配置される。最後に、GROMACSの入力ファイルとして、エネルギー最小化 (ステップ6.0)、平衡化 (ステップ6.1-6.6) および本計算 (ステップ7.0) のためのファイルが提供される[Brooks, *et al.* 2009]。

このような段階を経て生成された初期構造において、変異導入部位を図2-4に示した。また、変異残基や、その他のタンパク質内部におけるアミノ酸側鎖の二面角 (χ_n) が30度以上変化していないことを確認した。

本計算 (ステップ7.0) では、一回のタイムステップを2フェムト秒とし、全原子の場合は10ナノ秒 (5,000ステップ)、粗視化および保存性弾性ネットワークの場合は1マイクロ秒 (500,000ステップ) 分の軌道を計算した。全原子シミュレーションに対する分子力場はCHARMM-36[Klauda, *et al.* 2010; Venable, *et al.* 2014]、粗視化および保存性弾性ネットワークシミュレーションに対する分子力場はMARTINIを使用した[Monticelli, *et al.* 2008; Periolo, *et al.* 2009; Yesylevskyy, *et al.* 2010; de Jong, *et al.* 2012; Arnarez, *et al.* 2015; Hsu, *et al.* 2016]。温度設定は303.15K (berendsen, v-rescale) とし、座標は200ステップごとに記録した。

2.2.3 RMSDの算出

分子動力学シミュレーションにより得られた軌道結果から、まず構造安定性の指標となる立体構造の平均二乗偏差 (Root mean square deviation: RMSD)を算出した。これは、シミュレーションの過程において、タンパク質立体構造のフォールディングが解けるなどのエラーによりシミュレーションモデルが破壊していないか確認するために重要である。

$$RMSD_t = \sqrt{\langle (R_{i,t} - R_o)^2 \rangle}$$

ここで、 $R_{i,t}$ は時間 t における残基 i の座標を表し、 R_o は初期構造における座標を表す。

2.2.4 主成分分析

つぎに、主成分分析に基づき、タンパク質の動きを様々な成分に分解した後、各成分における野生型と変異型の差を比較した。ED分析は、先行研究[Amadei, *et al.* 1993; Stepanova, *et al.* 2007; Kamaraj & Purohit. 2016]などで提唱・実践されている。タンパク質の動態は、原子の位相空間挙動によって最もよく特徴づけられ、共分散行列の固有ベクトルを主成分と呼ぶ。この射影により、各固有ベクトルに沿った特定の軌道の変化が得られる。この分析のその他の基準は、先行研究[Purohit, 2014; Rajendran and Sethumadhavan, 2014]に基づいている。Srikumarらが行った先行研究では、野生型ヒトLaforinモデルと変異型ヒトLaforinモデルの軌道を分子動力学シミュレーションで計算後、その動態の差を主成分分析により検出することに成功している[Srikumar, *et al.* 2014]。また、Buslaevらは脂質分子の動態を分子動力学シミュレーションで再現し、やはり主成分分析により分子の動きを複数の成分に分解して観察を行っている[Buslaev, *et al.* 2016]。このように、ED分析は、タンパク質以外の分子にも応用可能な、汎用性の高い分子軌道の分析手法である。

2.2.5 RMSFの算出

ED分析の他に、平衡状態に達してからのタンパク質構造の揺らぎを確認するため、各アミノ酸残基ごとにRMSF (Root Mean Square Fluctuation) を次のように算出した。

$$RMSF_i = \sqrt{\langle (R_{i,t} - \langle R_i \rangle_t)^2 \rangle_t}$$

$R_{i,t}$ は残基 i の時間 t における座標、 $\langle R_i \rangle$ は残基 i の平均座標を表す。ここで、残基 i の座標 $R_{i,t}$ およびは $\langle R_i \rangle$ の座標は、その残基に含まれる全原子の平均を算出している。

2.3 結果および考察

2.3.1 RMSDによる構造安定性の観察

前述の通り、分子動力学シミュレーションにより得られたタンパク質の軌道をもとに、まずは時間軸に沿ったRMSDの推移を算出し、タンパク質の構造安定性を確認する必要がある。全原子モデルのシミュレーション (n=3) および保存性弾性ネットワークモデルのシミュレーション (n=5) に対するRMSD結果を算出した (図2-5)。全原子モデルは三種類のシミュレーションモデルの中で、構造的・時間的に最も動きの制約が強く、RMSDは小さいと考えられる。一方、保存性弾性ネットワークモデルは、最も動きの制約が少なく、長時間のシミュレーションを行っているため、RMSDの値は大きくなるはずである。

全原子モデルの結果 (図2-5 A) によれば、観測された全てのモデルおよびタイムステップにおいて、RMSDが0.25nm以下であることがわかった。時間ごとのRMSDの推移に関しては、シミュレーション開始から5ナノ秒間はRMSDの値が増加していき、その後は微増していた。動きの制約が最も少ない保存性弾性ネットワークモデル (図2-5 C) においても、ほぼすべてのRMSD測定値が0.5nm以下であった (0.5nmを超えるステップもわずかに存在した)。シミュレーション開始から数十ナノ秒間は、RMSDの値が急激に上昇し、その後は大きな増減はなく、ほぼ一定の値を保つか、もしくは微増していた。

これらの結果から、分子動力学シミュレーションで得られたタンパク質立体構造の動態は安定しており、タンパク質の立体構造が崩壊するなどのエラーは起きていないことがわかった。

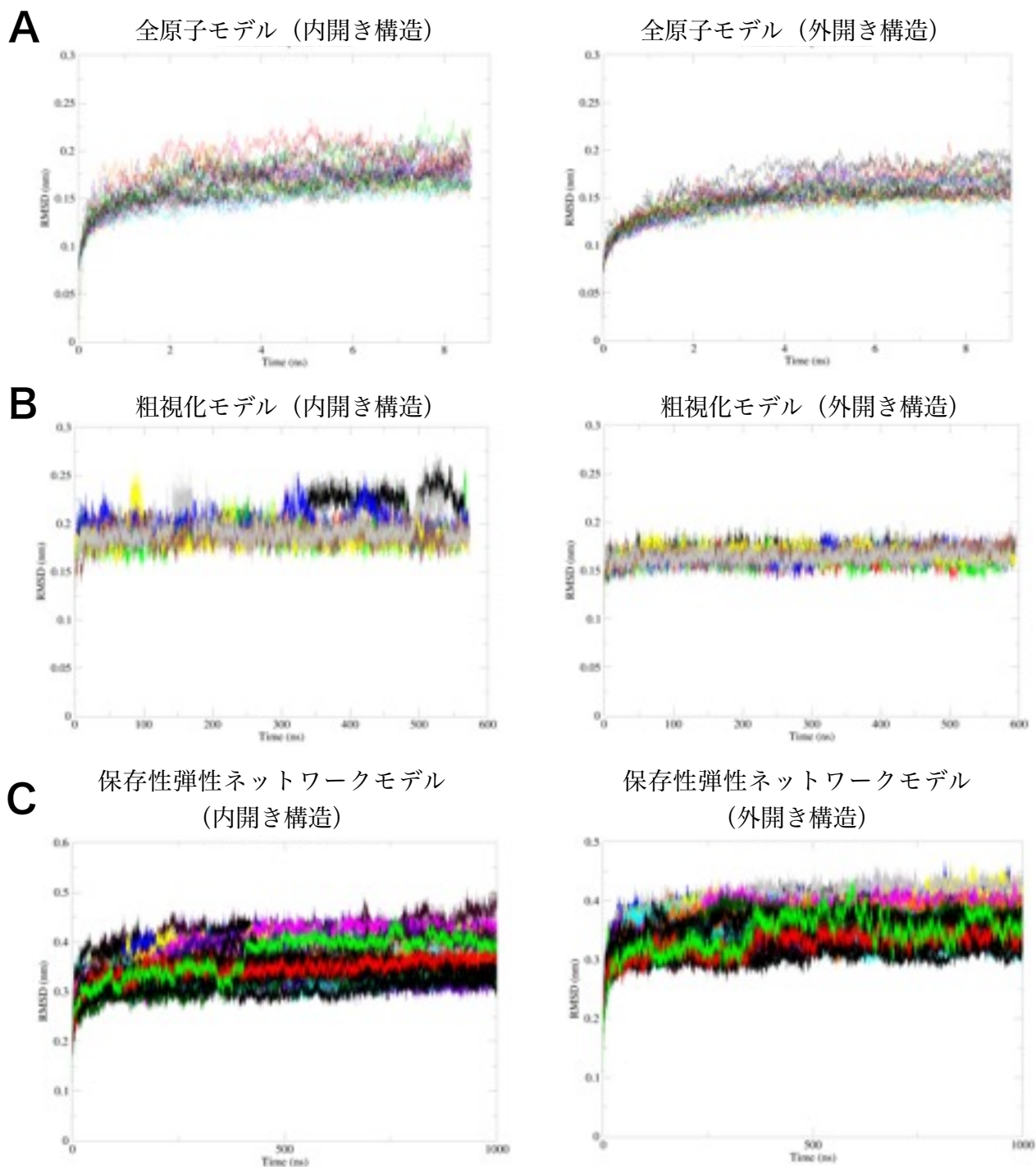


図2-5 各シミュレーションモデルにおけるRMSD

A, 全原子モデルのシミュレーションにおいて観測されたRMSD 内開き構造および外開き構造のそれぞれにおいて、全シミュレーションについての継時的RMSDを算出し、プロットした。

B, 粗視化モデルのシミュレーションにおいて観測されたRMSD Aと同様に、粗視化モデルを用いた全てのシミュレーションの継時的なRMSDを算出し、プロットした。

C, 保存性弾性ネットワークモデルのシミュレーションにおいて観測されたRMSD Aと同様に、保存性弾性ネットワークモデルによる全てのシミュレーションについて、継時的なRMSDを算出し、プロットした。

2.3.2 シミュレーションモデルの検証

本研究では、分子動力学シミュレーションによるタンパク質の動的構造を解析するにあたり、下記のような三種類のシミュレーションモデルを用意した。それぞれの結果や計算速度などを比較しながら、動態観察に最適なモデルを検証した。

- 1) 全原子モデル (Atomistic model)
- 2) 粗視化モデル (Coarse-grained model)
- 3) 保存性弾性ネットワークモデル (Conserved elastic network model)

ここでは、内開き構造のSLC2A1 (GLUT1, 5EQI) の野生型モデルおよび外開き構造のSLC2A3 (GLUT3, 4ZW9) の野生型モデルを上記の三種類のモデルそれぞれにおいて用意し、分子動力学シミュレーションのパフォーマンスを比較した。

各モデルにおいて、分子動力学シミュレーションの結果により得られた軌道を、前述の主成分分析により解析した。全原子モデルの座標から共分散行列を生成して、固有値をプロットしたものを図2-6に示した。成分1の固有値は8.273であり、全固有値の中で最も高いことが分かった。一方、その他の固有値は全て0.25を下回る値だった。したがって、全原子モデルの動的構造変化は、成分1の動きによる影響が大きいということが分かった。最も動的構造に対する影響が大きい成分1を横軸、成分2を縦軸にとり、各モデルの軌道を射影し、その像を比較した (図2-7)。

全原子モデルに関しては、内開き構造をもつSLC2A1は図2-7Aの左側 (黒)、横軸が-4から-2の間にプロットされた。外開き構造をもつSLC2A3は図2-7Aの右側 (緑)、横軸が2から4の間にプロットされた。粗視化モデルに関しては、内開き構造をもつSLC2A1は図2-7Bの左側 (黄)、外開き構造をもつSLC2A3は図2-7Bの右側 (ピンク) にプロットされた。保存性弾性ネットワークモデルは、内開き構

造をもつSLC2A1は図2-7Cの左側（青）、外開き構造をもつSLC2A3は図2-7Cの右側（赤）にプロットされた。

これらの結果を全て重ねた場合（図2-7D）、内開き状態から外開き状態までの動的構造変化を網羅していることが分かった。このような上に突き出た「く」の字の射影像は、全原子シミュレーションによる先行研究においても、非常によく似たものが報告されている。Kamarajらの研究では、SLC45A2遺伝子によってコードされる膜関連トランスポータータンパク質（Membrane-Associated Transporter Protein: MATP）の分子動力学シミュレーションおよび主成分分析を行っている[Kamaraj, *et al.* 2016]。本研究と同様、横軸を第一成分、縦軸を第二成分とした二次元にタンパク質軌道を射影した結果[Kamaraj, *et al.* 2016. Figure4a]では、やはり上に突き出た「く」の字のような射影像が得られている。つまり、この第一主成分に対して分散が最も広く、第一主成分の中間値付近において、第二主成分の値が増加する特徴があることを意味する。これを動的構造変化に置き換えて述べると、内向き構造から外向き構造（あるいはその逆）の構造変化が最も大きな輸送体の構造変化要素であるが、その中間状態においては、第二主成分の構造変化、すなわち細胞内側最大ループのひねり運動が活発になっていることを表す。構造変化が本計算で使用した粗視化モデルや保存性弾性ネットワークモデルの精度は、全原子モデルと比較して低いものの、第一成分および第二成分の射影像に限れば、先行研究と矛盾していないことが分かった。したがって、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を用いた計算では再現が困難な動的構造変化を観察するために有用であることが示唆された。

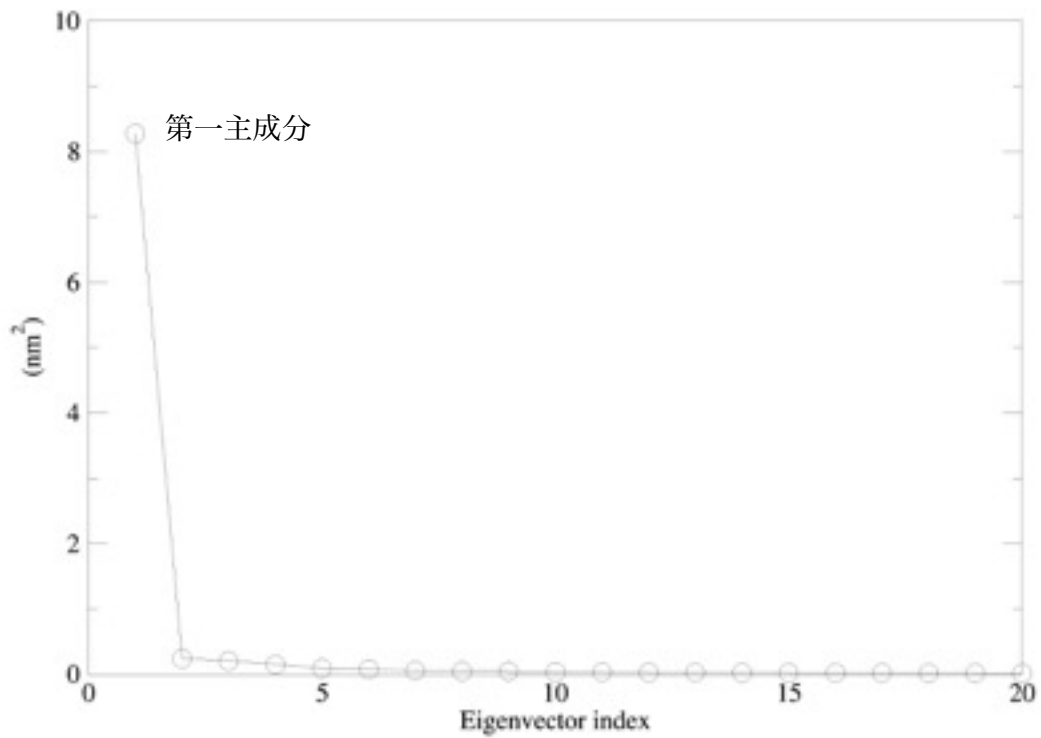


図2-6 野生型全原子モデルの各主成分に対する固有値

野生型の全原子モデルシミュレーション結果を主成分分析することで得られた第一主成分から第二十主成分までの固有値を示す。横軸が主成分のインデックスを表し、縦軸が固有値を表す。

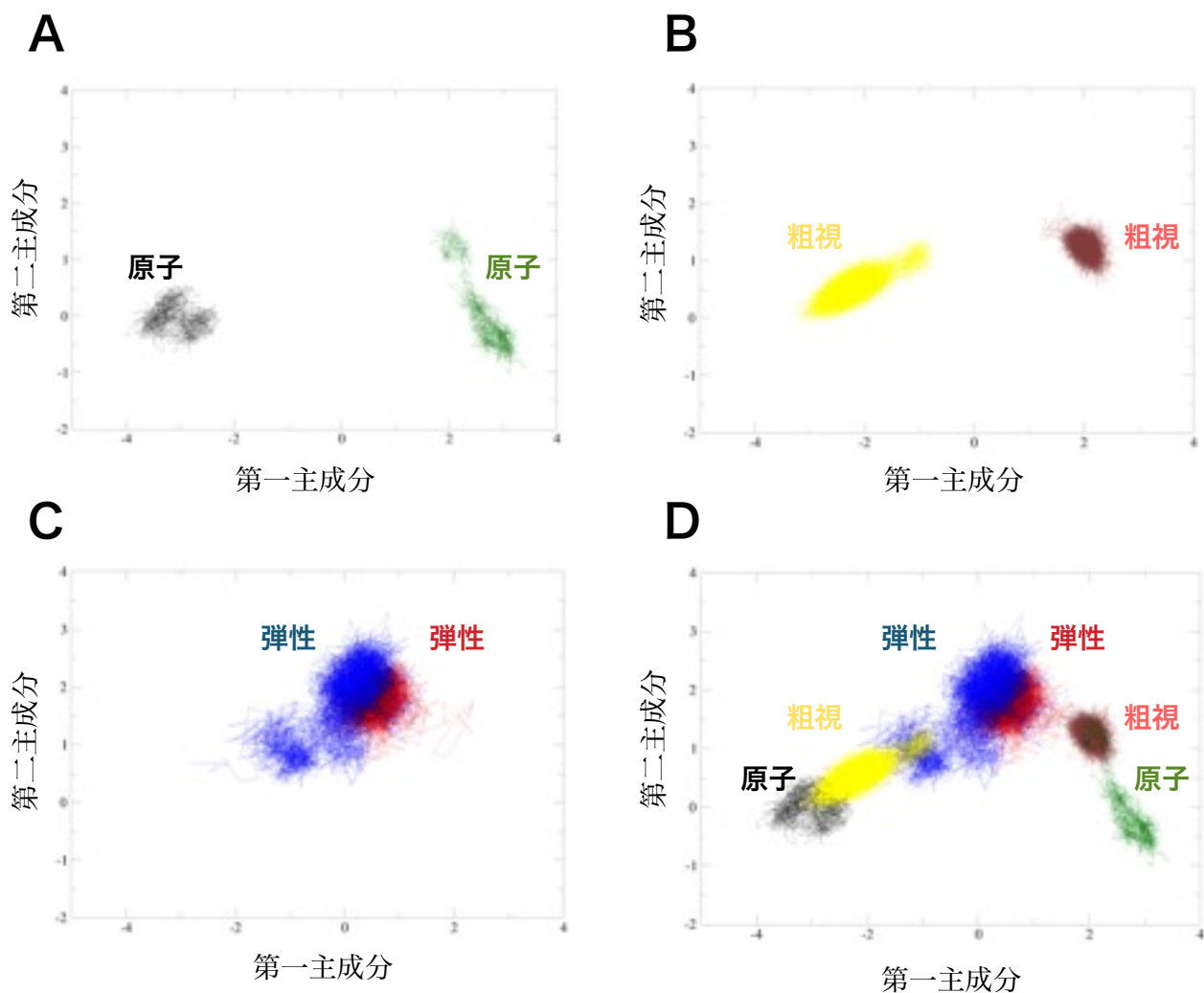


図2-7 第一主成分および第二主成分に対する軌道の射影像

A, 全原子シミュレーションの軌道の射影像 内開き構造と外開き構造の全原子シミュレーションを射影した（黒, 内開き構造; 緑, 外開き構造）。第一主成分は内開き構造と外開き構造の違いを示し、軌道はそれぞれ中央に向かって進む。

B, 粗視化シミュレーションの軌道の射影像 内開き構造と外開き構造の粗視化シミュレーションを射影した（黄, 内開き構造; ピンク, 外開き構造）。

C, 保存性弾性ネットワークシミュレーションの軌道の射影像 内開き構造と外開き構造の保存性弾性ネットワークシミュレーションの射影像（青, 内開き構造; 赤, 外開き構造）。

D, 全ての射影像を重ねたプロット 全てのモデルの射影像を重ねて比較すると、各モデルにおいて観察できる構造状態がそれぞれ異なっている。

2.3.3 保存型弾性ネットワークモデルを用いた野生型と変異型の比較

前計算では、保存性弾性ネットワークモデルを用いることにより、少ない計算コストで長時間分のシミュレーションモデルの構造変化を観測できることを示した。

本計算では、この保存性弾性ネットワークモデルのシミュレーション結果により、野生型と変異型の動的構造の比較を行った。まずは、保存性弾性ネットワークモデルのシミュレーション全体の座標変化から、共分散行列を生成し、固有値をプロットしたものを図2-8に示した。第一主成分の固有値は5.29であり、第二主成分の固有値は1.73、第三主成分は1.00、第四主成分は0.90、第五主成分は0.68、第六主成分は0.64、第七主成分は0.56、残りの固有値は全て0.5を下回る値だった。したがって、全原子モデルの動的構造変化は、第一主成分による動きの影響が大きいことが分かった。

ここで、第一主成分および第二主成分は実際にどのような動的構造変化を意味しているのかに関しては、それぞれの成分軸に沿って10個の平均構造を生成し、これらを繋げて映像にすることで確かめた(図2-9)。この図から、内開き構造から外開き構造へ構造が変化していく様子が確認できる。つまり、第一主成分は、N末端側構造とC末端側構造の方向が互いに変化する動的構造変化を表していることが分かった。

主成分1-8をそれぞれ縦軸にとり、全てのモデルの軌道をプロットし、箱ひげ図として表した結果を図2-10に示す。この中で、第一主成分における内開き構造の野生型と比較して、全ての変異型の固有値が上昇しており、有意な差であることが確認された ($p < 0.01$, Tukey-Kramer test)。したがって、内開き構造の変異型は、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明らかになった。

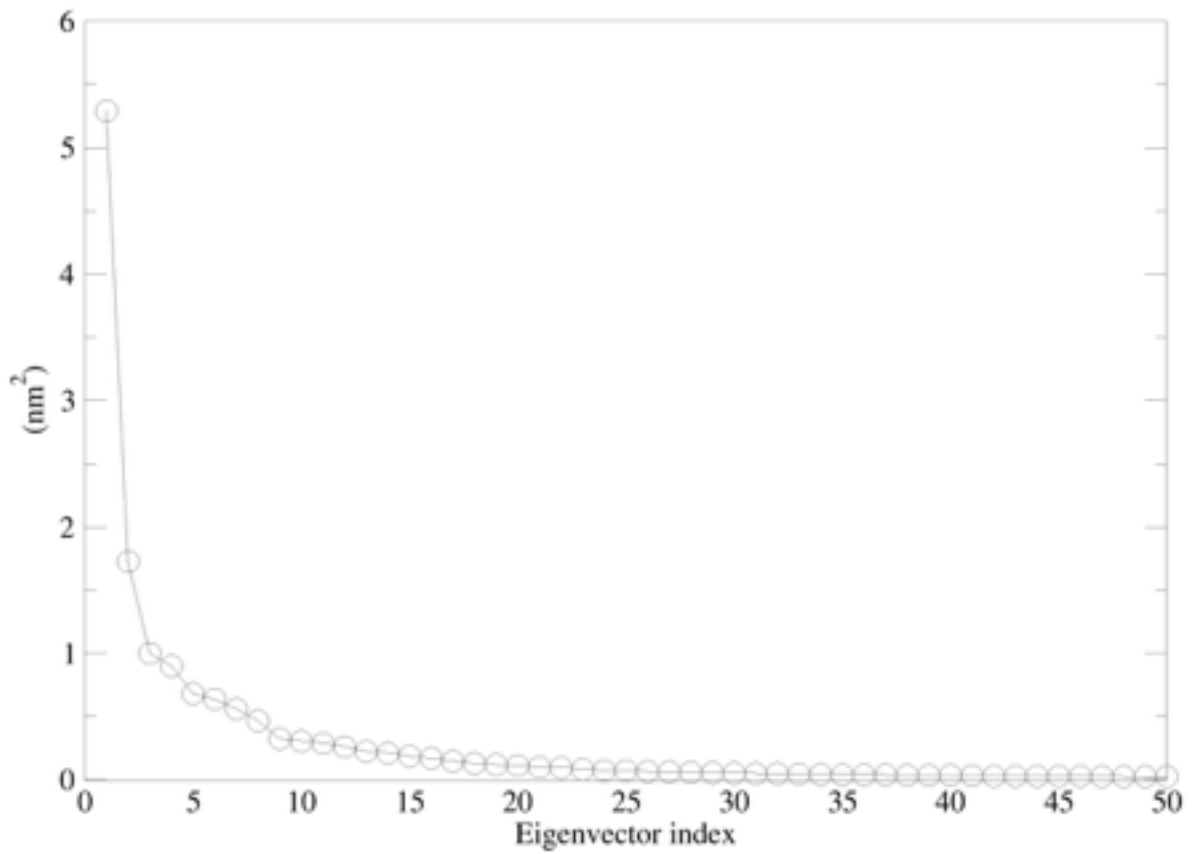


図2-8 保存性弾性ネットワークモデルの各主成分に対する固有値

野生型および変異型の保存性弾性ネットワークシミュレーション結果を主成分分析することで得られた第一主成分から第五十主成分までの固有値を示す。横軸が主成分のインデックスを表し、縦軸が固有値を表す。

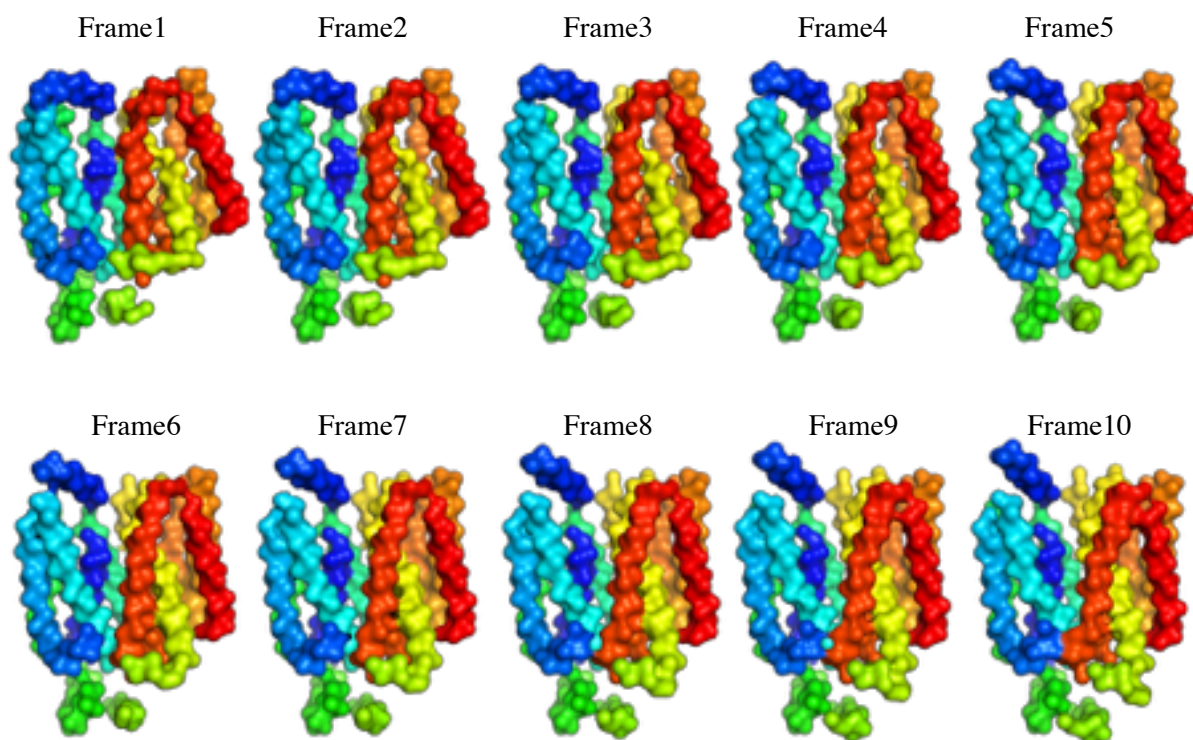


図2-9 保存性弾性ネットワークモデルの第一主成分が示す動的構造変化

保存性弾性ネットワークシミュレーションの主成分分析より得られた第一主成分軸に沿って、最小値から最大値に至るまでを十のフレームに分割し、平均的な構造を生成した。Frame1-10の構造は、全てN末端を青、C末端を赤として虹色で示した。Frame1は最も内開きの構造、Frame10は最も外開きの構造をもつ。

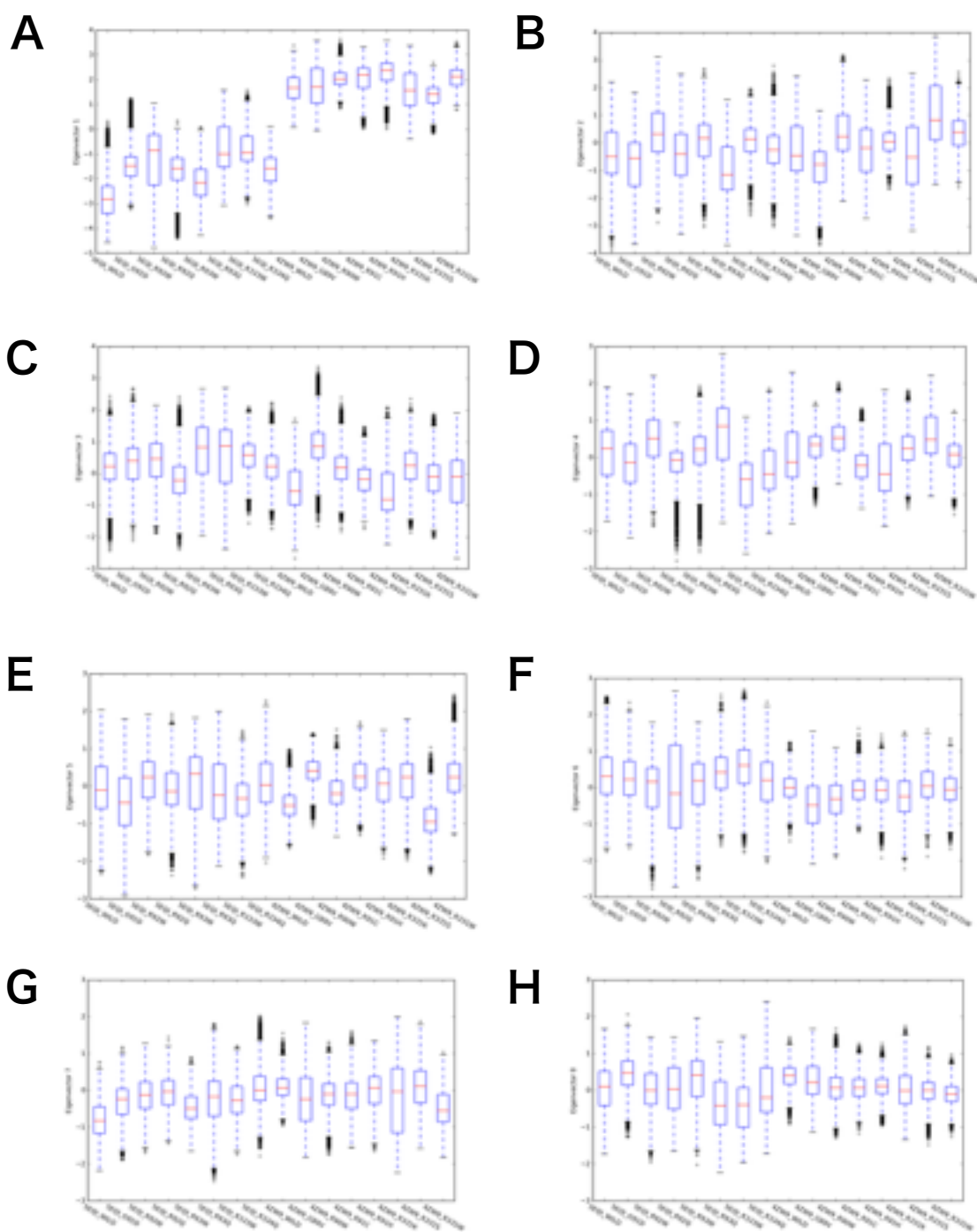


図2-10 主成分ごとのモデル比較

A, 第一主成分 B, 第二主成分 C, 第三主成分 D, 第四主成分 E, 第五主成分 F, 第六主成分 G, 第七主成分 H, 第八主成分

第一主成分から第八主成分について、全ての保存性弾性ネットワークモデルの軌道をプロットした (5EQI, 内向き構造モデル; 4ZW9, 外向き構造モデル)。

2.3.4 RMSFによる各残基のゆらぎ解析

タンパク質構造の揺らぎの大きさを野生型と変型の間で比較するため、RMSFを算出した。ここでは、各残基に含まれる全原子のRMSFの平均を、その残基におけるRMSFとして求めている。それぞれのモデルについて、シミュレーションを5回繰り返しているため、それらの平均および標準偏差を図2-11に示した。野生型のRMSFを黒い線、変異型のRMSFを赤い線で表示した。

まず、全体の傾向として、細胞内領域および細胞外領域の残基はRMSFの値が大きいことが確認できた。例えば、5EQIにおける残基番号0-10, 140-150, 200-270, 440-の周辺残基の多くは、細胞内領域に位置すると考えられる残基である。同様に、残基番号40-50, 165-175, 290-300, 350-360, 415-425あたりの残基は細胞外領域に位置するとトポロジー予測がなされている。これらの残基番号におけるRMSF値はその他の残基と比較して大きくなる傾向にあることが認められた。特に、残基番号200-270あたりはSLC2のN末端側とC末端側をつなぐ大きな細胞内ループを形成しており、RMSF値が大きい残基が多いことが分かった。前述の主成分分析に基づく構造変化の解析や、これらのRMSFの結果から、この大きな細胞内ループの部位が開閉構造の蝶番としての役割を果たしている可能性があると考えられる。

また、野生型と変異型のRMSFの差をより明瞭に示すため、変異型RMSFから野生型RMSFを引いた差のプロットを図2-12に示した。全ての残基について、野生型と変異型を比較して、有意にRMSFに差が認められる残基を調べた。統計検定はT検定を用いて、有意水準は0.01とした。変異が導入された残基番号には実線が引かれ、有意差が認められた残基番号には破線が引かれている。変異導入位置および有意にRMSFが増大もしくは縮小している残基を、トポロジー図および立体構造図にマッピングした(図2-12)。

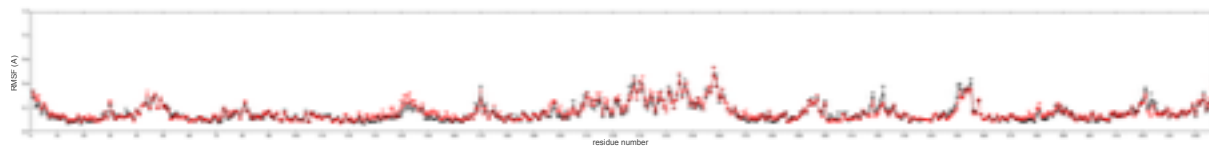
内開き構造のG91DやR92W, R92Q, R93W, R93Qは、変異導入位置が互いに近いものの、ゆらぎの変化の特徴はそれぞれ異なることが分かった。野生型との差が最も顕著なのはR92Wで、主にN末端側の構造において、野生型と比較して有意にRMSFが増大している残基が15残基も検出された。R92Wは公共データベースClinVarによると、グルコース欠乏症2の発症に関与する疾患変異であることが分かっている。このN末端側の残基の揺らぎ増大が、グルコース輸送体の立体構造の不安定化を招き、動的構造変化に影響を与えている可能性が考えられる。G91DおよびR93Wも、ClinVarにおいて疾患変異であると報告されている。両者とも、196PのRMSFの有意な増大が認められる。このプロリンは、6回目膜貫通領域の中ほどに位置し、輸送体表面に面しているため、脂質膜と相互作用する可能性が高い残基である。

内開き構造に導入したR333Wと、外開き構造に導入したR91Hの変異では、RMSFが野生型と比較して縮小している残基が多数観測された。その他の変異に関しては、野生型と比較して、RMSFが縮小する残基は1,2残基またはゼロであり、多くは増大していた。R333WもClinVarにおいて疾患変異であると報告されている変異の一つであるが、RMSFの増減に関しては、その他の疾患変異とは異なった特徴を見せた。

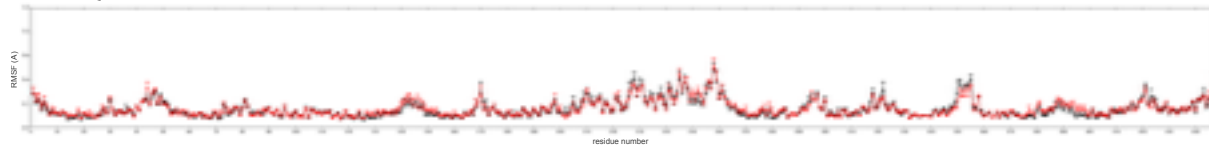
外開き構造に導入したR90W, R91C, R331K, R331S, R331Wの変異では、C末端側の構造の揺らぎを増大させていることが分かった。特に、R90Wは先ほど述べた蝶番としての役割を担っていると考えられる残基200-270の細胞内ループにおける5つの残基 (257V, 258L, 260L, 262R, 264S) のRMSFを有意に増大させていた。R90Wの第一主成分の分布は、野生型と比較するとやや外開き傾向が見られることから、細胞内ループのゆらぎの増大が外開き構造の出現頻度増加につながった可能性が考えられる。

A

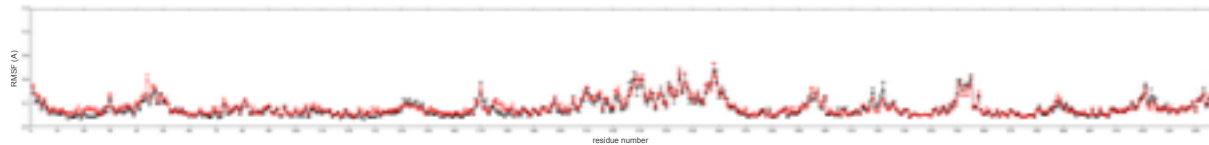
G91D



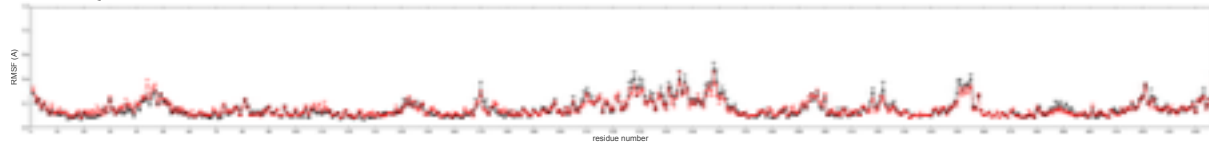
R92Q



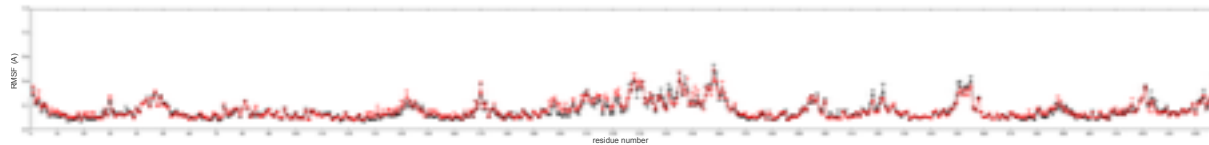
R92W



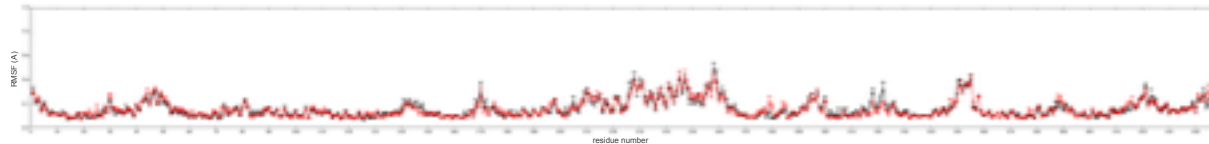
R93Q



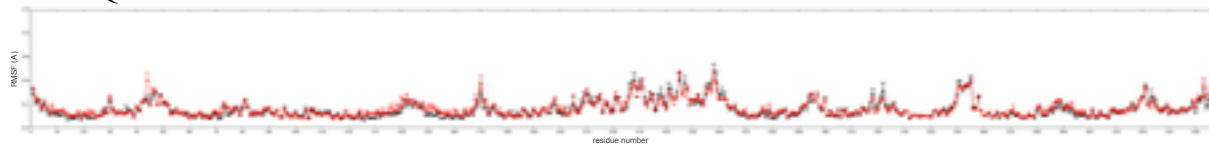
R93W



R333W

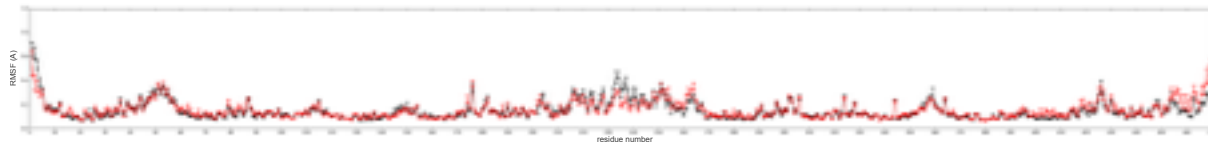


R334Q

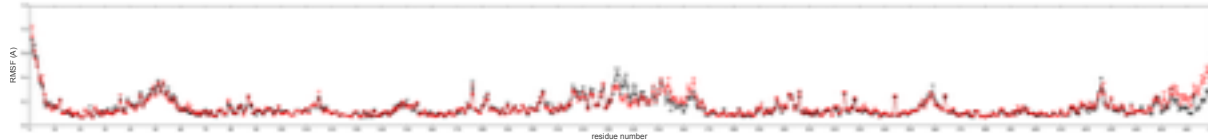


B

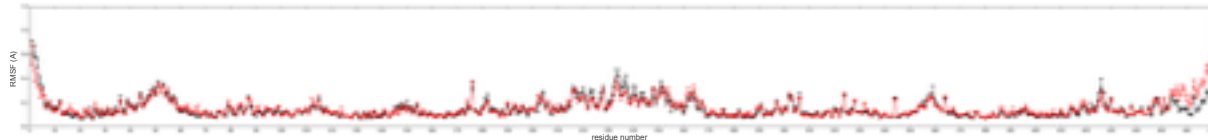
G89V



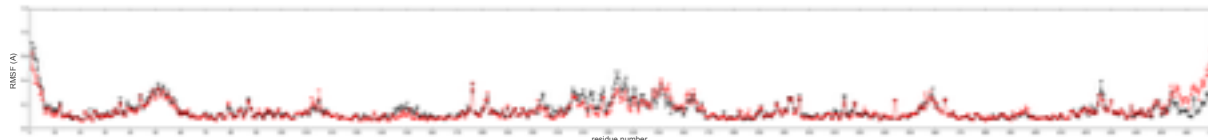
R90W



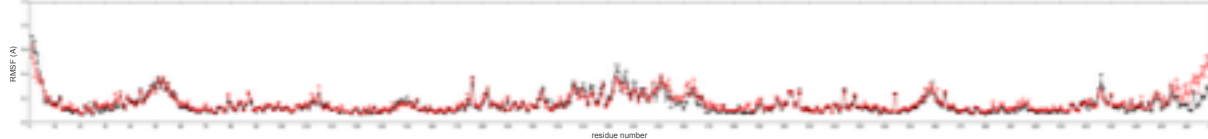
R91C



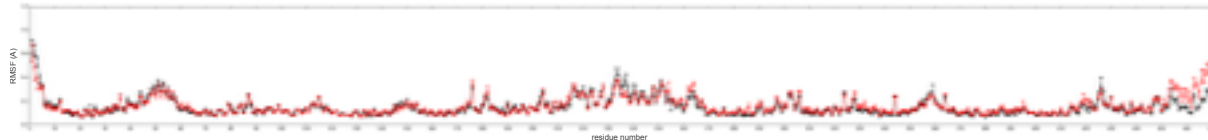
R91H



R331K



R331S



R331W

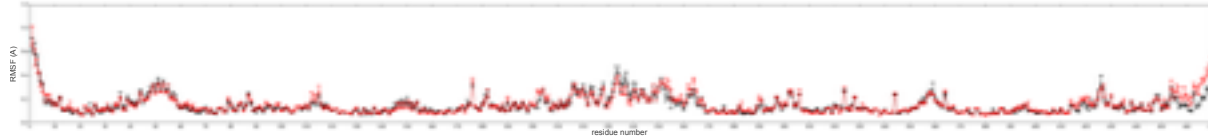


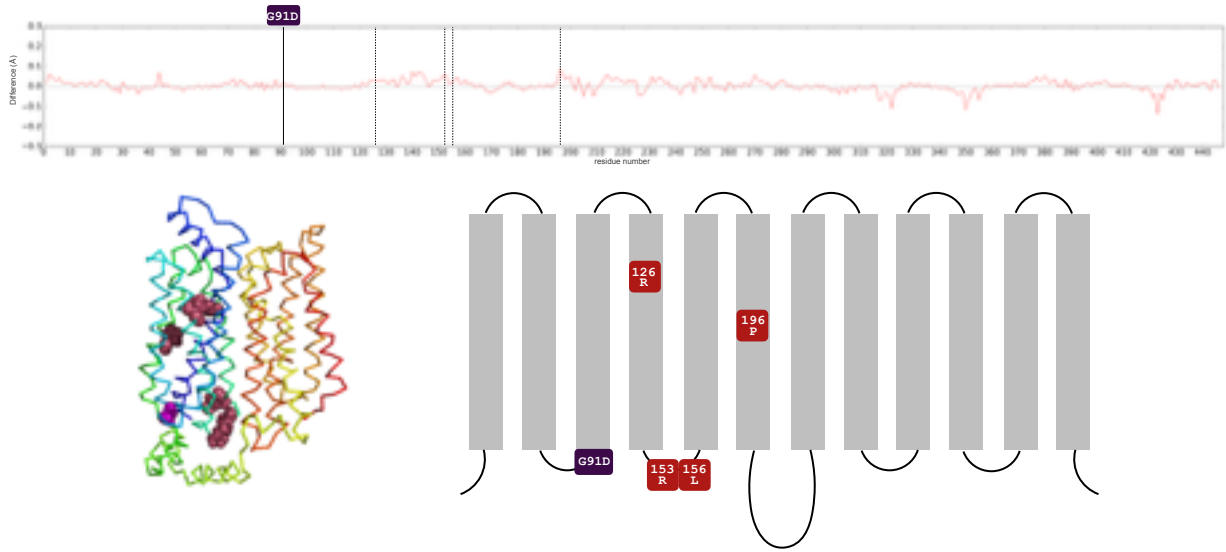
図2-11 各モデルにおけるRMSF

A, 内開き構造の変異型モデルのシミュレーションから算出されたRMSF

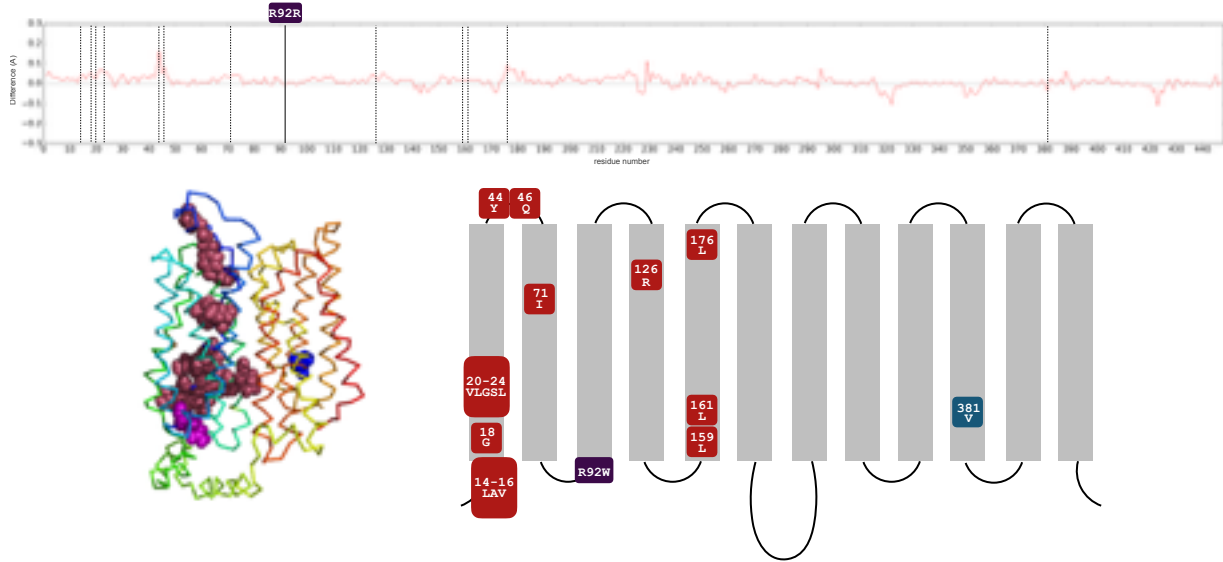
B, 外開き構造の変異型モデルのシミュレーションから算出されたRMSF

横軸は残基番号、縦軸はRMSFの値 (Å) を示し、野生型と変異型を比較した (赤線, 変異型; 黒線, 野生型)。

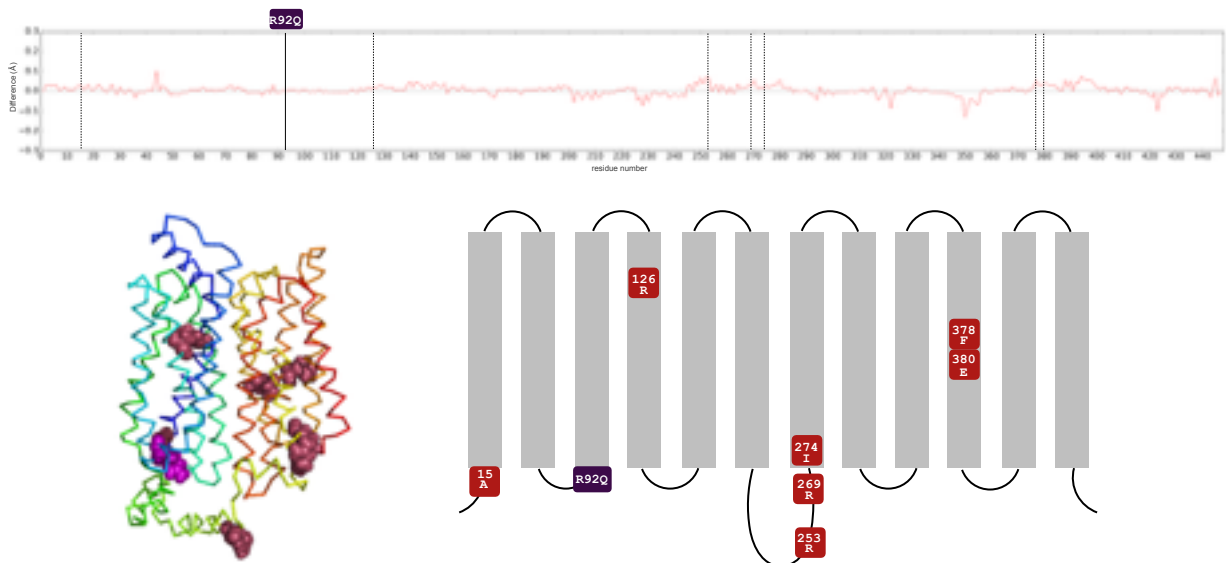
G91D (内開き構造)



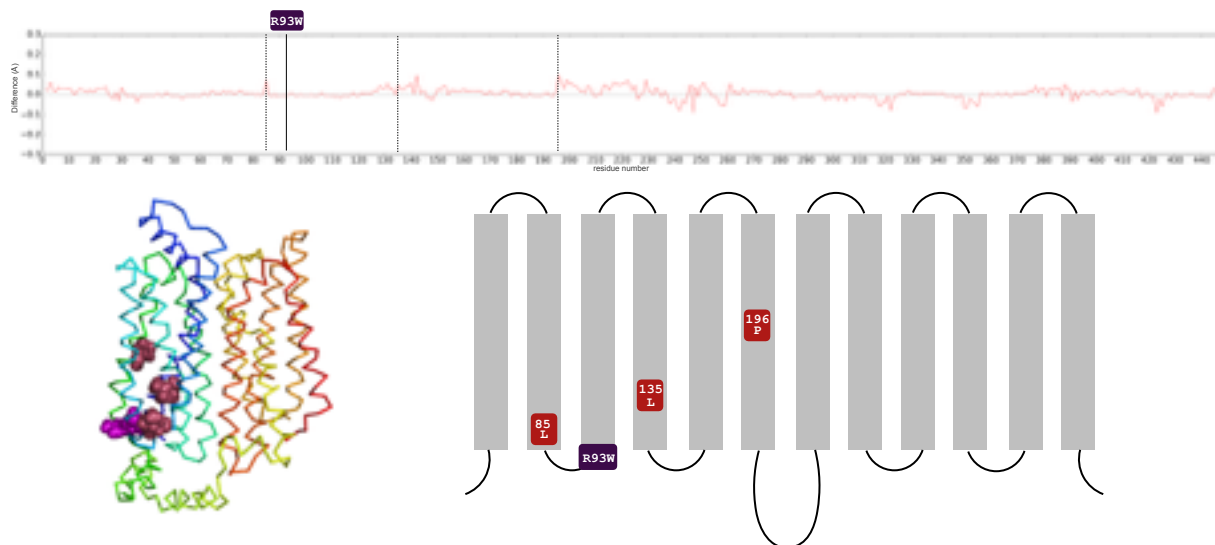
R92W (内開き構造)



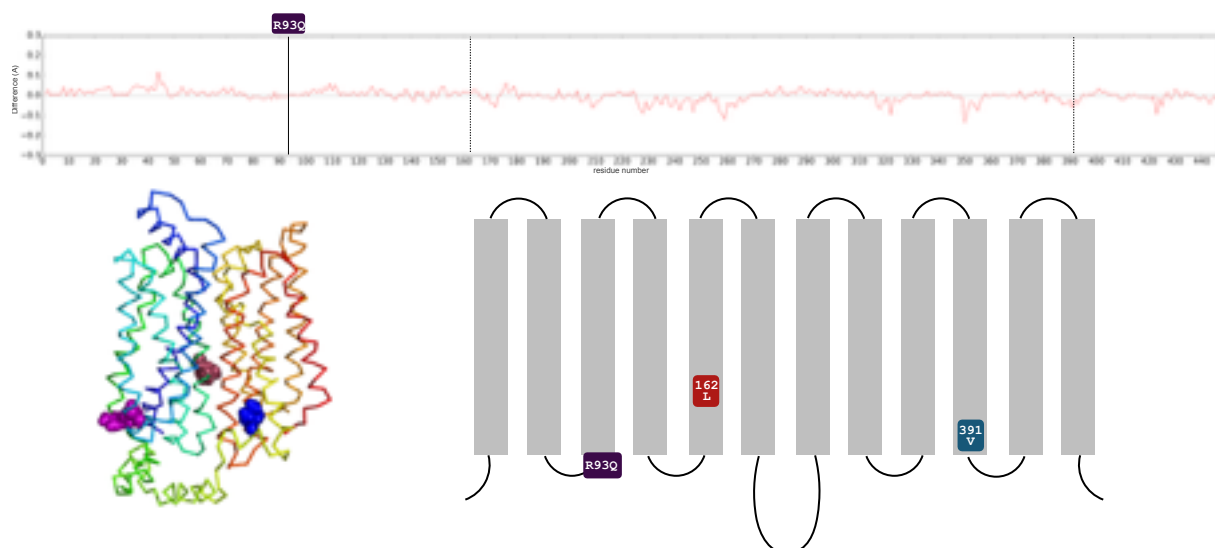
R92Q (内開き構造)



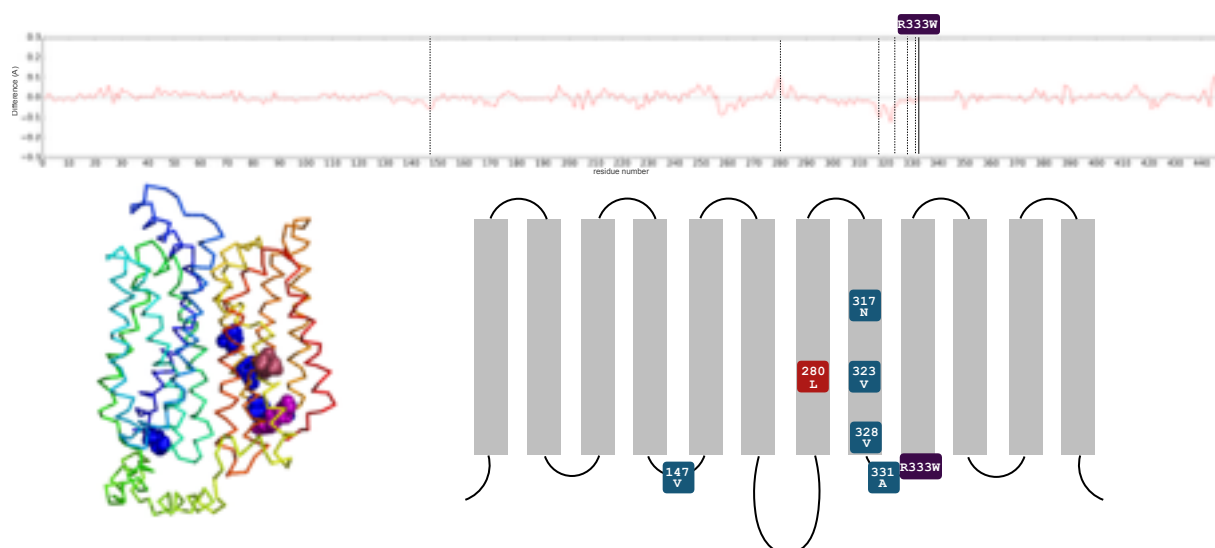
R93W (内開き構造)



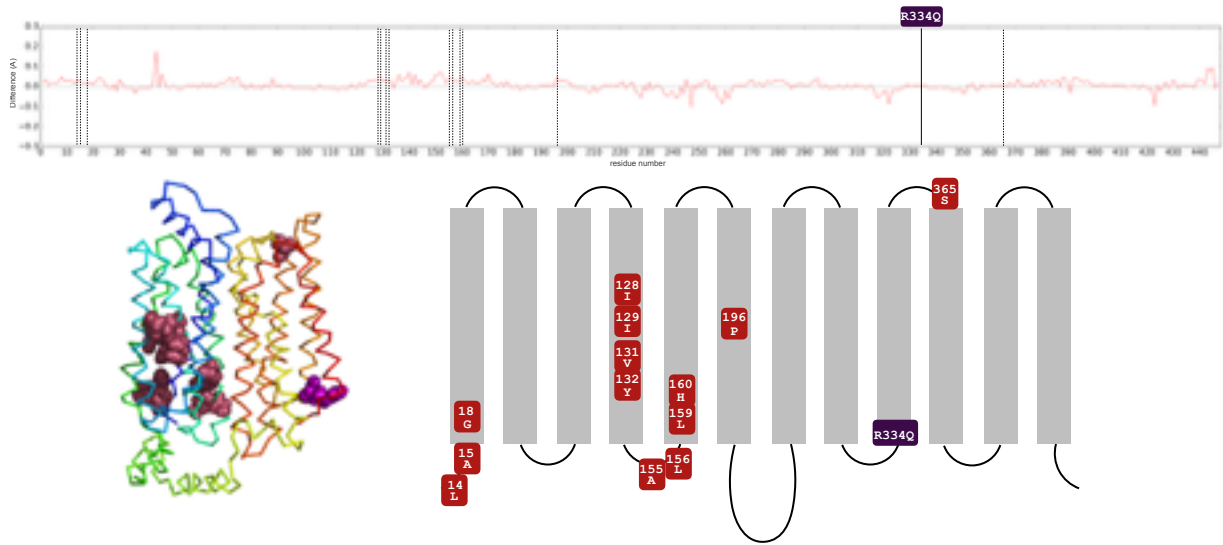
R93Q (内開き構造)



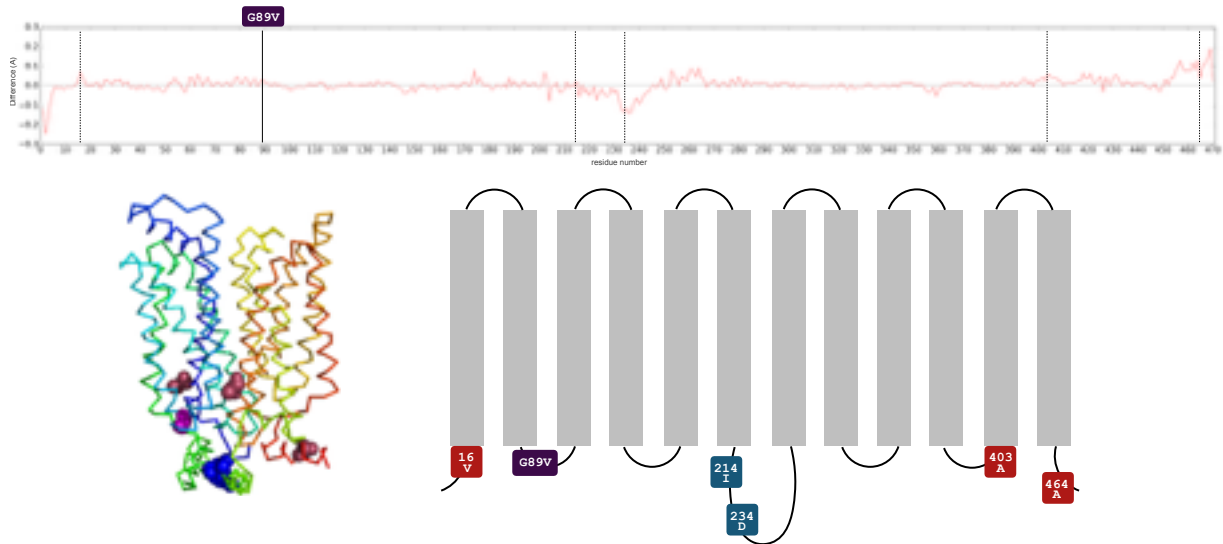
R333W (内開き構造)



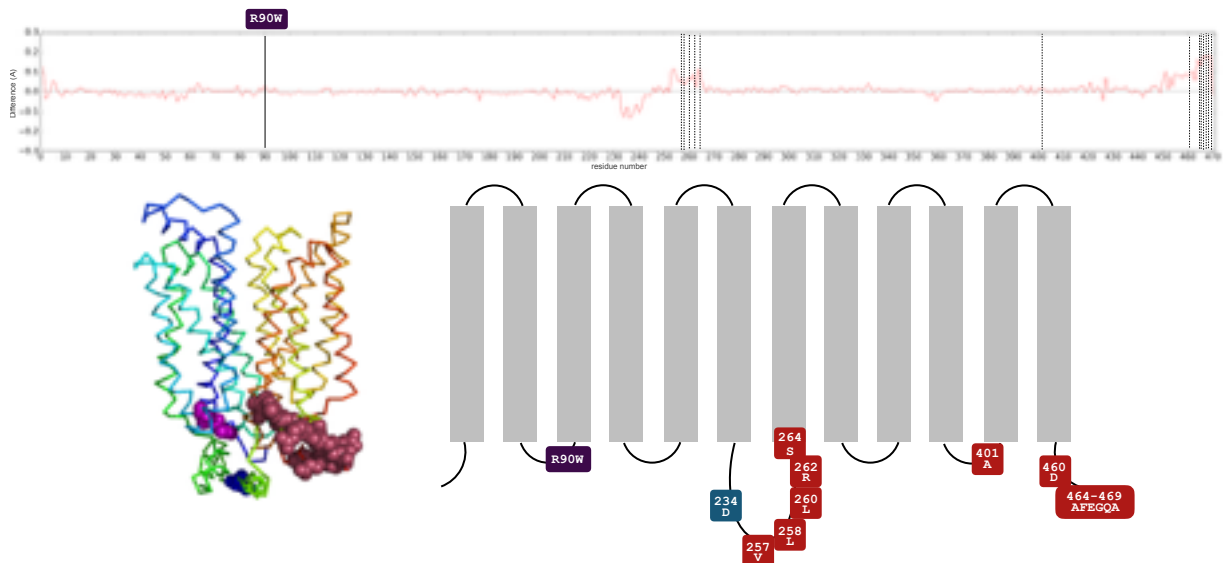
R334Q (内開き構造)



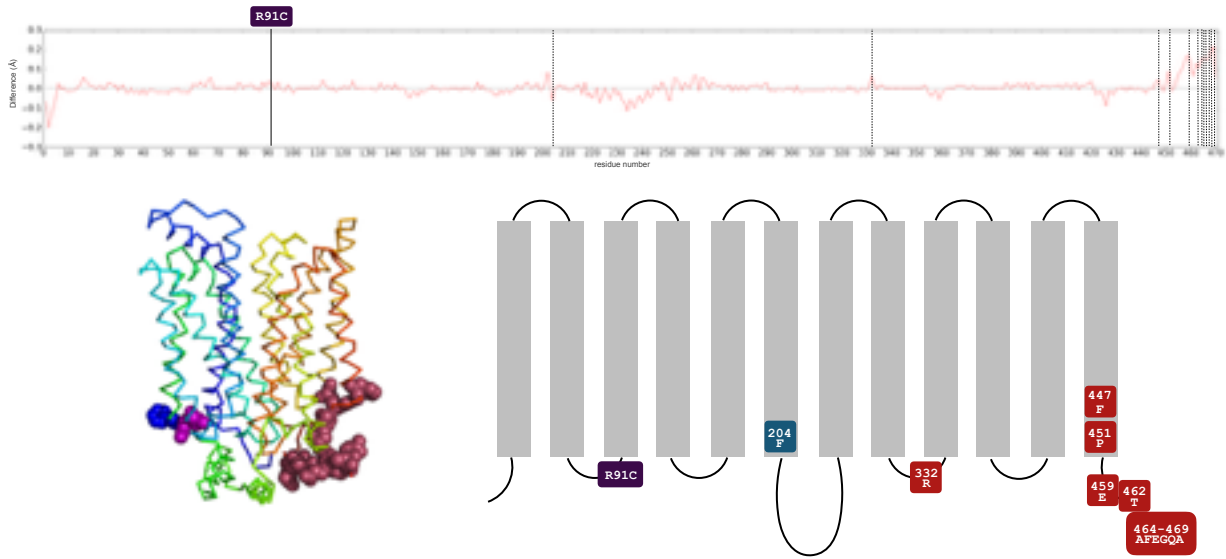
G89V (外開き構造)



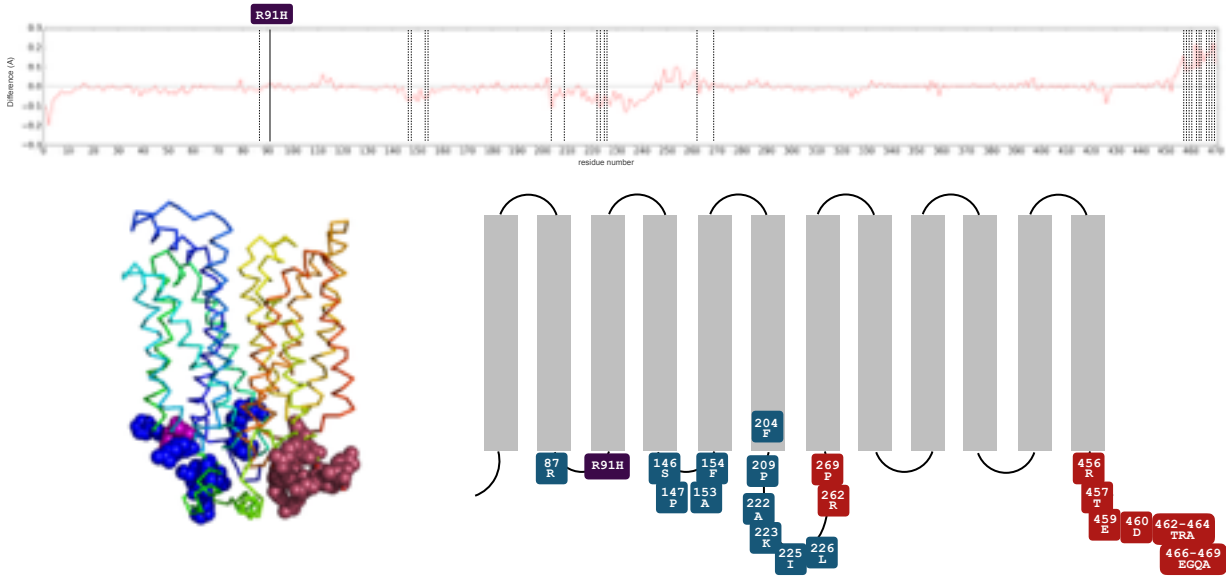
R90W (外開き構造)



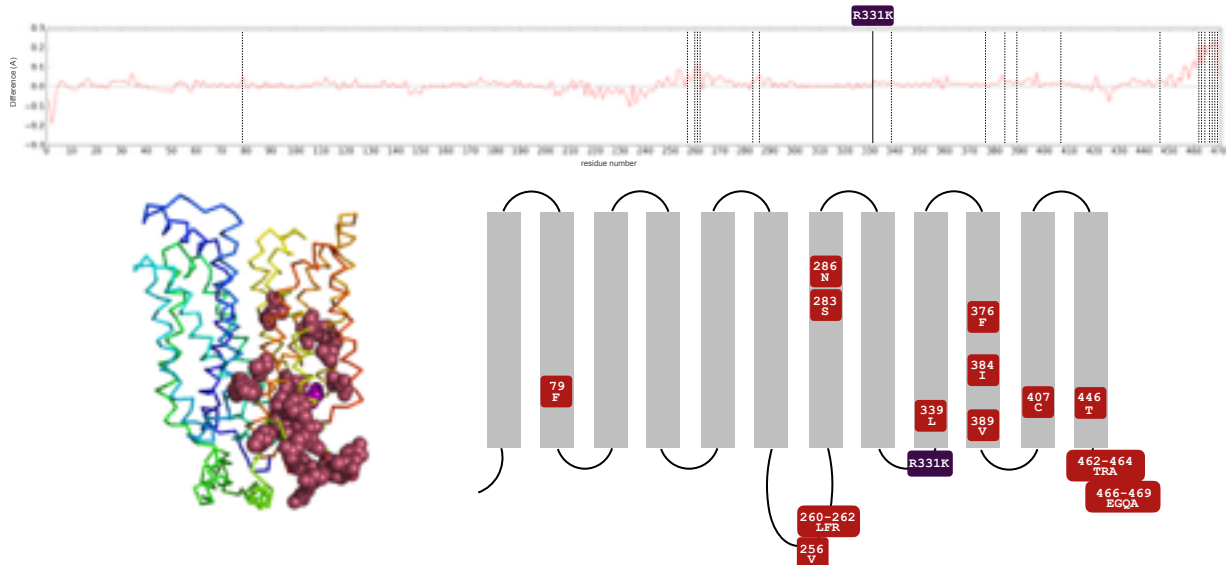
R91C (外開き構造)



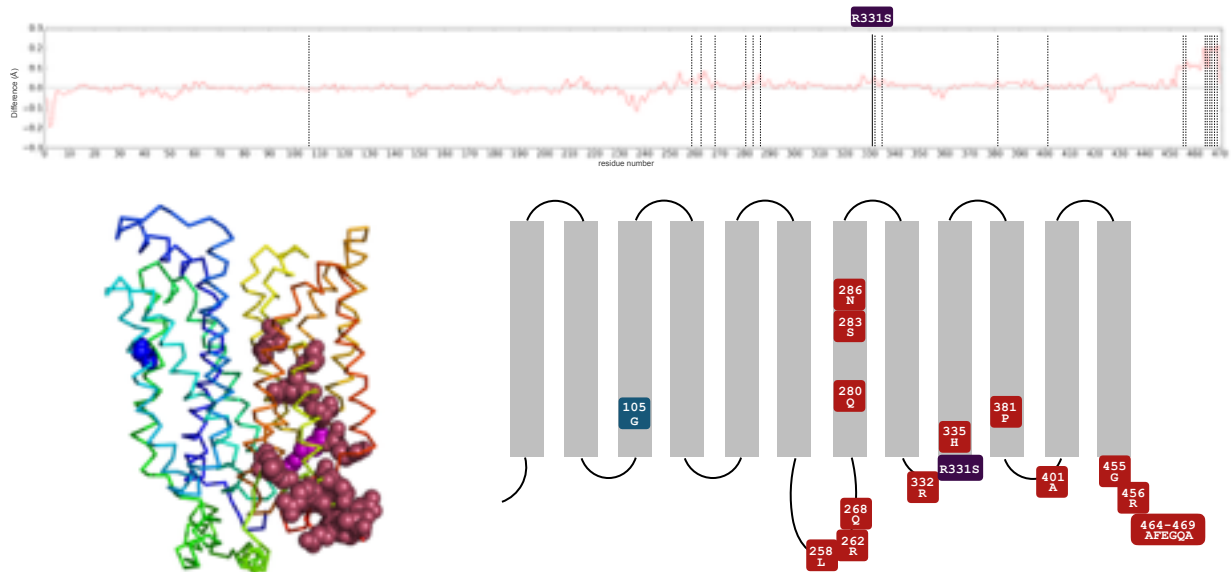
R91H (外開き構造)



R331K (外開き構造)



R331S (外開き構造)



R331W (外開き構造)

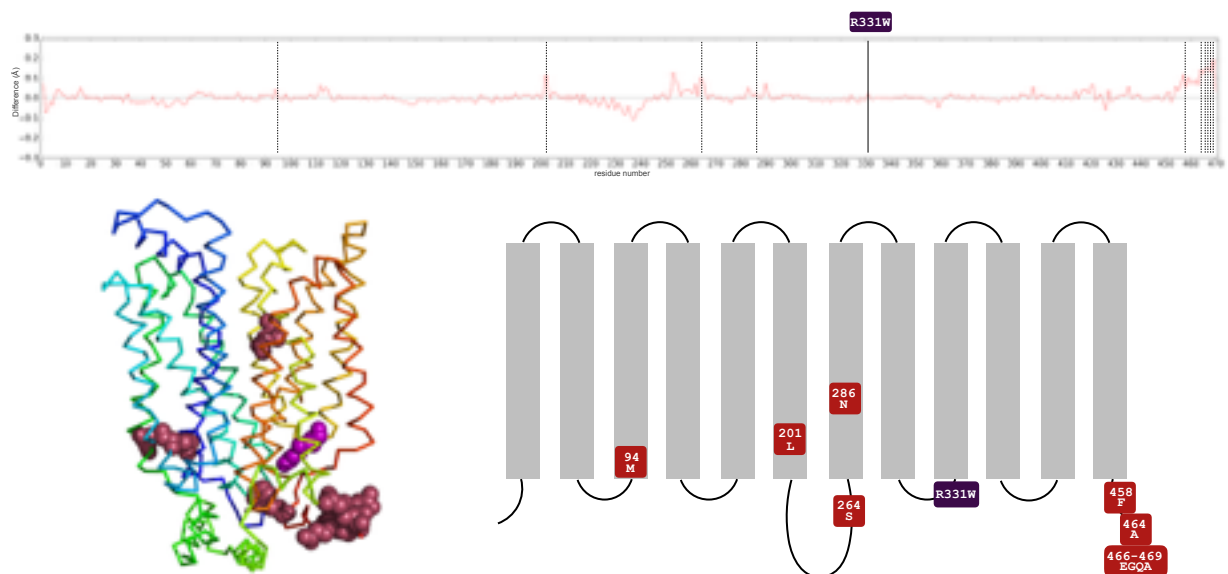


図2-12 野生型と変異型のRMSF比較

図2-11において示した野生型と変異型のRMSFの差をプロットした。また、有意な差が認められた残基を、立体構造およびトポロジー上にマッピングした(紫, 変異残基; 赤, RMSFが野生型と比較して有意に増大した残基; 青, RMSFが野生型と比較して有意に減少した残基)。

- 第3章 -

第3章 グルコース輸送体のホモログにおける保存残基の探索

3.1 序

前章までは、SLC輸送体の変異と表現型の関係を理解する上で、変異が動的構造変化に及ぼす影響を計算科学的アプローチにより検証することの有効性を示してきた。本章では、第1章とは異なる観点から疾患と関連の高い部位を明らかにすべく、グルコース輸送体のホモログにおいて保存されたアミノ酸残基を探索した。

タンパク質のアミノ酸配列や、構造/機能には、非常に長い時間スケール、数億年単位の時間をかけて積み上げられてきたもう一つの重要な変化「分子進化」による影響が反映される[Yamato & Yura. 2013]。ゲノムにコードされた遺伝子配列は、常にランダムな突然変異を受けて少しずつ変化を蓄積していくが、この際、タンパク質の構造や機能を決定的に壊してしまうアミノ酸変異は集団から排除される傾向にある。しかし、タンパク質の構造や機能を損なうことなく、関連タンパク質との相互作用も大きく損なわないような変異は、集団の中に固定される可能性が高い。このような分子進化の仕組みにより、タンパク質の配列・構造・機能は少しずつ変化を遂げ、多様性を獲得する。このような分子進化の歴史を紐解くには、データベースに蓄積された共通祖先由来のアミノ酸配列・構造・機能を比較することが有効である。タンパク質の各残基ポジションにおける進化的なアミノ酸多様性を調べることで、分子進化の過程でどのような影響を受けてきたのか明らかにすることができる。アミノ酸多様性が認められるポジションでは、それらのアミノ酸間における置換はタンパク質の構造や機能を損なうことなく、何らかの形で生命活動に貢献できることを意味し、構造や機能に多様性をもたらしているかもしれない。しかし、アミノ酸多様

性が認められないポジションにおいては、その他のアミノ酸では代替できないような何らかの構造的・機能的役割を特定のアミノ酸が担っている可能性が高い。

本章では、第2章で研究対象としていたSLC2のアミノ酸配列および立体構造情報を活用し、ヒト以外の幅広い生物種を対象として、進化的保存部位を検証した。その結果から、膜領域と細胞質の境界に位置するアルギニン残基は、ヒト以外においても同様に重要なのかという問いに対して考察した。

3.2 方法

3.2.1 SLC2のアミノ酸配列および立体構造の重複に関する検証

SLC2の進化解析を行うにあたり、まずSLC2のN末端側とC末端側が重複により生じたりリピート構造であるかどうかを確かめた。Forrestらの報告によれば、SLC2は左右対象に近い立体構造をもつ [Forrest and Rudnick, 2009; Forrest, *et al*, 2011]。SLC2は12回膜貫通型タンパク質であるが、N末端側とC末端側のそれぞれ6回膜貫通領域の立体構造が保存されているならば、これらは同一祖先由来である可能性が高い。その場合、SLC2の配列や構造を、重複構造の最小単位（つまり、N末端側もしくはC末端側の6回膜貫通領域）に分割してホモロジー検索を行うことで、重複回数によらず、より遠縁なホモログや祖先配列を探索できる可能性があると考えた。

まず、N末端側/C末端側の配列類似性および立体構造類似性を検証するため、2種類の相同なSLC2（SLC2A1とSLC2A3）のアミノ酸配列および立体構造を中央で分割し、それぞれ比較した。過去に重複が起きていれば、N末端側とC末端側の立体構造はある程度保存されている可能性が高い。一方、アミノ酸配列は、重複を起こした時期が古ければ古いほど、保存度は低いと予想される。ここで、使用した配列は、SLC2A1（アクセッション：NP_006507.2）のアミノ酸配列1-239をN末端側、アミ

ノ酸配列240-492をC末端側とした。同様に、SLC2A3（アクセッション：NP_008862.1）のアミノ酸配列1-234をN末端側、アミノ酸配列235-496をC末端側とした。立体構造を比較する場合は、SLC2A1の立体構造（PDB ID: 5EQI）の残基番号9-239をN末端側、残基番号240-455をC末端側とした。同様に、SLC2A3の立体構造（PDB ID: 4ZW9）の残基番号1-234をN末端側、残基番号235-470をC末端側とした。いずれも6回膜貫通領域と7回膜貫通領域の間にあるループ中央部をN末端側とC末側の境界として設定した。

アミノ酸配列のマルチプルアライメントには、MAFFT (v7.305b) [Kato, 2002; Kato, 2013]を用いた。構造アライメントには、MATRAS (Markovian TRAnsition of Structure evolution) [Kawabata & Nishikawa, 2000; Kawabata, 2003]を使用した。

次に、中心軸に沿ってSLC2A1の立体構造を少しずつ回転させ、N末端側とC末端側の構造が最も良く重なる角度、すなわちRMSDが最小になる角度を調べた。タンパク質の立体構造を回転させるためには、まず中心軸を算出する必要がある。そのため、前述のN末端側とC末端側の構造アライメントを用いて、N末端側とC末端側で対応するアミノ酸のペアを抽出した。これらペア残基同士を線で結び、その二等分点の位置をプロットした（図3-3A、青い球体が二等分点を表す）。青い球体は、SLC2A1の立体構造中央部において線状に並ぶことがわかり、これらの近似曲線をタンパク質立体構造の中心軸と見なした。中心軸の式は、

$$\text{中心座標} + t \times \text{方向ベクトル}$$

により表される（ t は定数）。全ての二等分点の平均座標は $[-39.42, 3.59, 9.38]$ 、近似曲線の方法ベクトルは $[-9.46, 35.90, 14.90]$ であったことから、SLC2A1の立体構造の中心軸は、

$$[-39.42, 3.59, 9.38] + t \times [-9.46, 35.90, 14.90]$$

と表すことができる (図 3-3B)。この中心軸に沿って、SLC2A1の立体構造座標を5度ずつ回転させ、初期構造と比較した。構造類似度の指標は、立体構造のRMSDを用いた。RMSDの定義および算出方法については、第2章で述べた通りである。ここでは、各回転角度につきN末端側およびC末端側の構造類似度を、各残基のC α およびC β についてそれぞれ算出した。

3.2.2 隠れマルコフモデルを用いたホモロジー検索

タンパク質のホモロジー検索の感度は、隠れマルコフモデル (Hidden Markov model: HMM) を応用することで高めることができる [Eddy, 1995; Mistry, *et al.* 2013]。HMMは確率モデルの一種であり、状態は文字通り「隠れ」ているため外部から観察することはできず、結果のみが出力される。しかし、状態が隠れていようとも、一定の遷移確率と出力確率に従い、各状態の分布は時系列に沿ってに変化していく。HMMは、1990年代から遺伝子発見やタンパク質配列のモデル化など生物配列の解析にも応用されるようになった。本研究では、HMMを応用した代表的なホモロジー検索アプリケーション、HMMer [Eddy, 1995; Mistry, *et al.* 2013]を用いて、幅広いSLC2のホモログを検索することにした。

HMMerによりホモロジー検索をするためには、HMMを基盤とした「プロファイルHMM」を用いてタンパク質のモチーフ解析を行う [Eddy, 1995; Mistry, *et al.* 2013]。プロファイルHMMとは、マルチプルアライメントの情報を記述するために考案されたHMMである。ギャップの取り扱いに優れており、高い感度でホモロジー検索を実行することができる。プロファイルHMMは、一致状態 (M $_i$) と、挿入状態 (I $_i$) と欠失状態 (D $_i$) からなる構成単位を繰り返し線状に連結させた構造をもつ。繰り返

この回数は、プロファイルHMMの長さと呼ばれる。プロファイルHMMの長さは、マルチプルアライメントの保存カラムの数 N が適用される。 M_i と I_i からは20種類のアミノ酸や4種類の塩基が、 D_i からはギャップのみが出力される。マルチプルアライメントのうち、保存されたカラムの残基は M_i および D_i を、その他の非保存カラムは I_i を用いて、それぞれ遷移確率と共に記述される。

本研究では、SLC2のN末端側とC末端側の進化的特徴に差があるかもしれないことを考慮して、以下の3種類のプロファイルHMMを作成した。

- 1) N末端側の配列のみ (N-side profile)
- 2) C末端側の配列のみ (C-side profile)
- 3) N末端側の配列とC末端側の配列を両方含む場合 (N+C profile)

HMMerにおいてSLC2のHMMプロファイルを生成すべく、マルチプルアライメントをstockholm形式にて保持したファイルを入力値として、コマンドhmmbuiltを実行した。この際、精度を向上させるため、SLC2の配列だけでなく、進化的に近い相同なファミリーSLC22の配列も含めてマルチプルアライメントおよびプロファイルHMMを生成した。得られたプロファイルHMMを入力値として、コマンドhmmsearchを実行することによりホモロジー検索を行った。

HMMerのホモロジー検索でヒットしたタンパク質のうち、本研究ではE-valueが 1.0^{-4} 以下のものをSLC2のホモログタンパク質とした。ヒットタンパク質について、プロファイルHMMと相同なドメインの保持数や、アミノ酸配列長、E-valueの分布などを集計した。プロファイルの種類ごとに、どのようなホモログが得られたのか、その特徴をまとめた。

3.2.3 ホモログの保存部位解析

hmmsearchの結果ファイルには、クエリーとヒットタンパク質との間で相同と判断された部位について、アミノ酸配列のペアワイズアライメントの結果が書き込まれている。このアライメント情報から、クエリーとヒットタンパク質の一致残基と残基番号を全て抽出した。集計はプロファイルの種類ごとに行い、保存度が高い残基とそのポジションを明らかにした。

また、保存度が高いアミノ酸残基が、SLC2の立体構造上でどのような位置にあるのか調べるべく、PyMOLを用いて立体構造上に保存度が高いアミノ酸をマッピングした。プロファイルHMMと立体構造上の位置の対応関係を確認するため、クエリーとして用いたプロファイルHMMの代表配列と、SLC2A1の立体構造 (PDB ID: 5EQI) およびSLC2A3の立体構造 (PDB ID: 4ZW9) のアミノ酸配列をMAFFT (v7.305b) [Katoh, 2002; Katoh, 2013]によりマルチプルアライメントした。このアライメント結果をもとに、マッピングすべき残基を判断した。

3.3 結果および考察

3.3.1 立体構造の対称性の検証

SLC2のN末端側とC末端側が重複により生じたものであるか調べるため、アミノ酸配列のマルチプルアライメントおよび構造アライメントを行った。SLC2A1のN末端側、C末端側、SLC2A3のN末端側、C末端側のそれぞれの配列一致度を図3-1Aに示す。組合わせ1はSLC2A1のN末端側とSLC2A1のC末端側の比較、組合わせ2はSLC2A3のN末端側とSLC2A3のC末端側の比較、組合わせ3はSLC2A1のN末端側とSLC2A3のN末端側の比較、組合わせ4はSLC2A1のC末端側とSLC2A3のC末端側の比較を表す。その結果、両SLC2において、N末端側とC末端側のアミノ酸配列一致度はともに25.0%であった。この一致率の値は、一般に α ヘリックスで構成される膜タンパク質と配列一致度の関係において‘twilight zone’と呼ばれる領域である。この領域では、配列一致度が低くとも、RMSDが低く構造類似度が高い場合があると報告されている[Olivella, *et al.* 2013]。一方、両SLC2間のN末端同士もしくはC末端同士はアミノ酸配列の一致度が63.6%および65.0%と非常に高く、RMSDは0.5-1.0 Å程度であると予想される[Olivella, *et al.* 2013]。

次に、構造アライメントにより算出したRMSDの結果を表3-1 (A, $C\alpha$ を比較した場合のRMSD値; B, $C\beta$ を比較した場合のRMSD値) および図3-1Bに示す。組合せ1では、 $C\alpha$ のRMSDが2.5 Å、 $C\beta$ のRMSDが2.1 Åであった。組合せ2では、 $C\alpha$ のRMSDが3.4 Å、 $C\beta$ のRMSDが2.6 Åであった。一般的な膜タンパク質のアミノ酸配列一致度とRMSDの関係と比較すると、これらRMSDの値は平均的か、もしくはやや小さいことが分かった。配列一致度が14-5%と低いことを考慮すれば、RMSDが2.0-3.0 Åというまた、組合せ3では、 $C\alpha$ のRMSDが0.9 Å、 $C\beta$ のRMSDが0.8 Åとなり、組合せ4では、 $C\alpha$ のRMSDが1.3 Å、 $C\beta$ のRMSDが1.0 Åであった。

さらに、構造アライメントによりそれぞれの立体構造ペアを重ね合わせたものを図3-2に示す。この構造は、各アミノ酸残基のC α 原子の座標を繋ぐことにより、立体構造の骨格をトレースしている。組合せ3 (図3-2C) および組合せ4 (図3-2D) は、前述の通り低いRMSDが観測されただけに、構造が非常によく一致していることが目視でも確認できる。一方、組合せ1 (図3-2A) および組合せ2 (図3-2B) に関しては、 α ヘリックスのおおよその方向が保存されていることが分かった。個々のC α 原子の位置にずれがあり、特にループ部分に関しては細胞内側および細胞外側のいずれにおいても差が大きいことが確認できる。しかし、膜貫通領域に相当すると予想される6本の α ヘリックスの領域において、 α ヘリックスの方向や長さが酷似していることが分かった。

したがって、SLC2A1のN末端側とC末端側、もしくはSLC2A3のN末端側とC末端側の間では、アミノ酸配列の一致度が14-15%と非常に低いにも関わらず、膜貫通領域における立体構造は保存されていることが分かった。

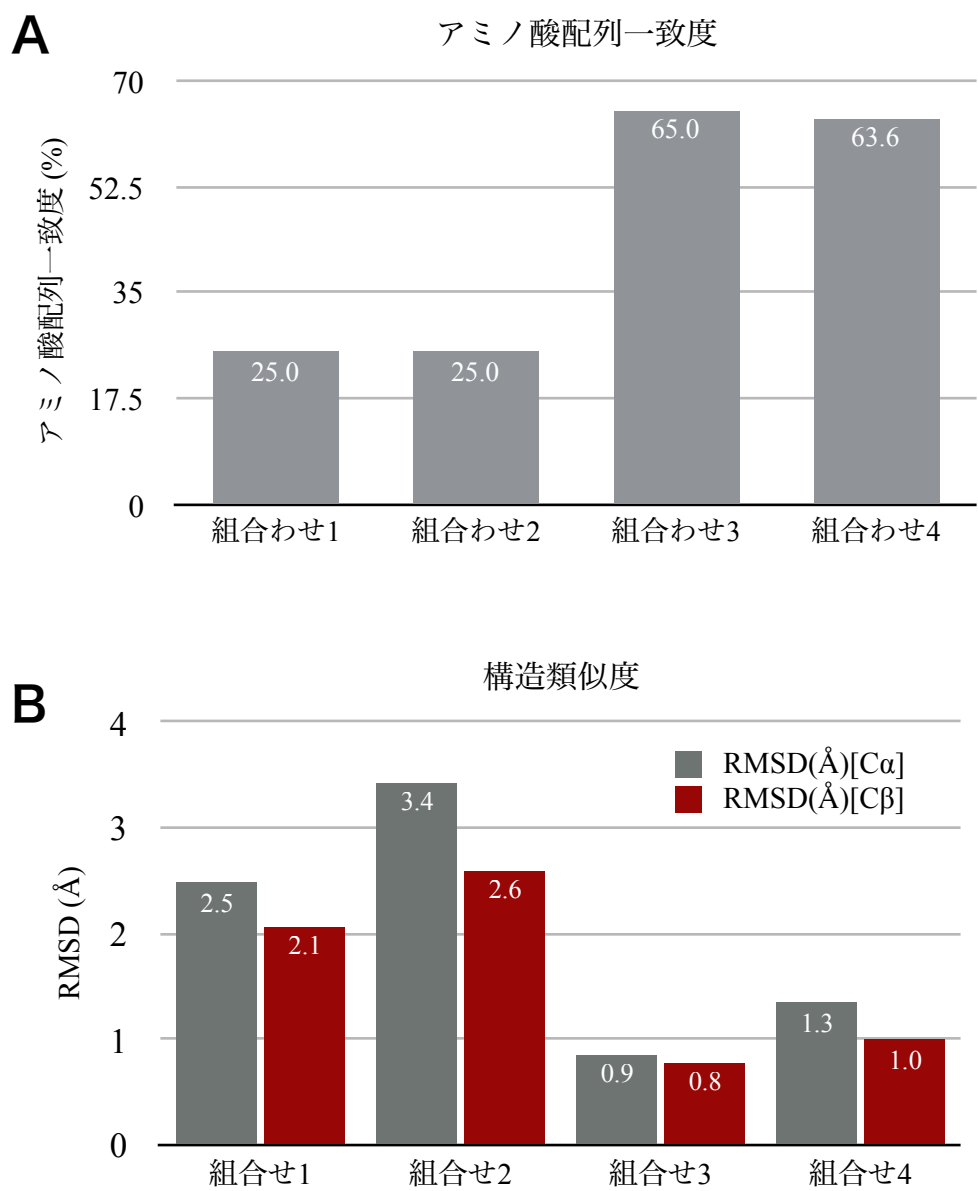


図3-1 N末端側およびC末端側における配列類似度と構造類似度の比較

A, アミノ酸配列の一致度 B, RMSDを指標とした立体構造類似度

組合わせ1, SLC2A1のN末端側とSLC2A1のC末端側の比較; 組合わせ2, SLC2A3のN末端側とSLC2A3のC末端側の比較; 組合わせ3, SLC2A1のN末端側とSLC2A3のN末端側の比較; 組合わせ4, SLC2A1のC末端側とSLC2A3のC末端側

表 3-1 SLC2A1およびSLC2A3のN末端/C末端構造を比較した場合のRMSD値 (Å)

A

使用した配列	比較対象			
	SLC2A1 N末端側	SLC2A1 C末端側	SLC2A3 N末端側	SLC2A3 C末端側
SLC2A1 N末端側	0	2.477	0.855	3.208
SLC2A1 C末端側	2.477	0	3.367	1.345
SLC2A3 N末端側	0.855	3.367	0	3.421
SLC2A3 C末端側	3.208	1.345	3.421	0

B

使用した配列	比較対象			
	SLC2A1 N末端側	SLC2A1 C末端側	SLC2A3 N末端側	SLC2A3 C末端側
SLC2A1 N末端側	0	2.064	0.774	2.183
SLC2A1 C末端側	2.064	0	2.574	0.984
SLC2A3 N末端側	0.744	2.574	0	2.583
SLC2A3 C末端側	2.183	0.984	2.583	0

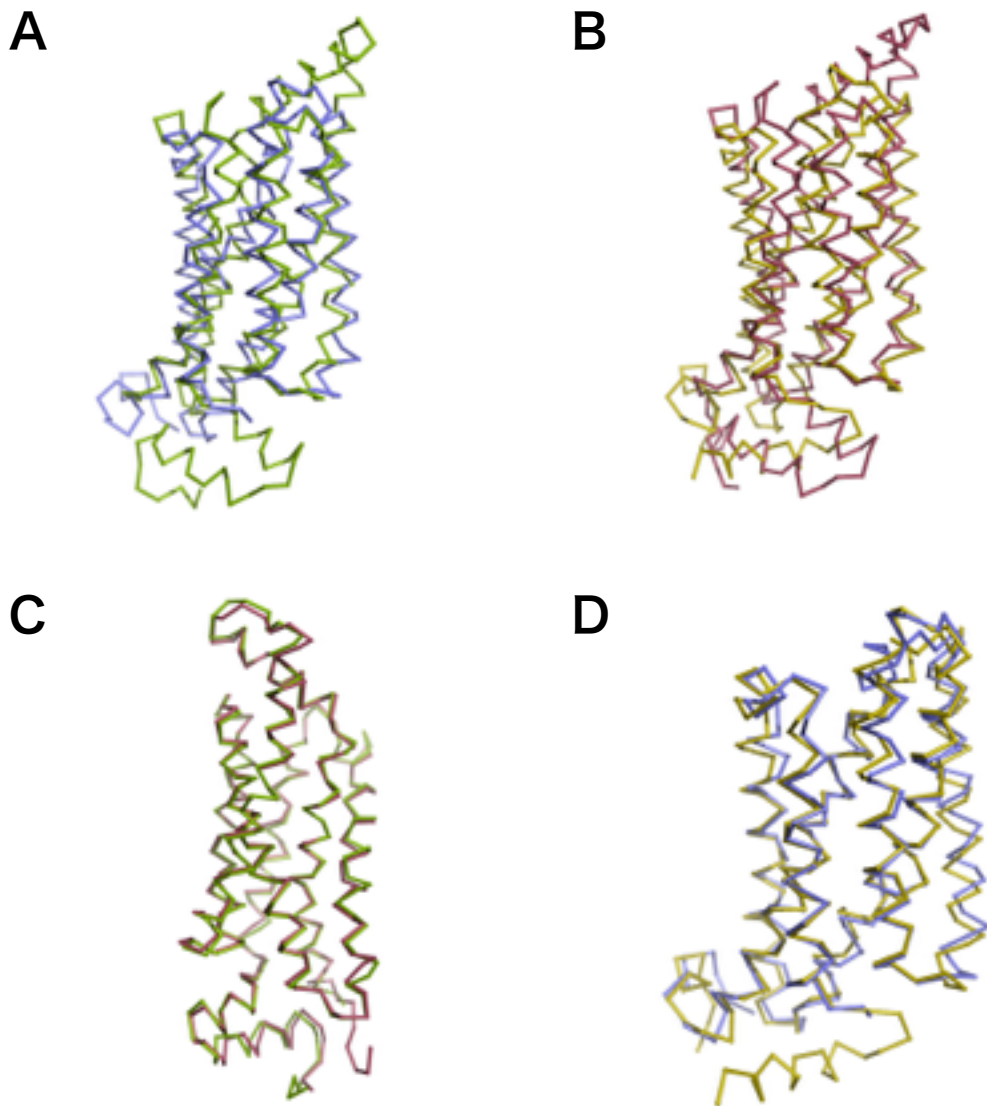


図 3-2 構造の重ね合わせ

A, 組合わせ1 (SLC2A1のN末端側とSLC2A1のC末端側の比較) の構造アライメント
 B, 組合わせ2 (SLC2A3のN末端側とSLC2A3のC末端側の比較) の構造アライメント
 C, 組合わせ3 (SLC2A1のN末端側とSLC2A3のN末端側の比較) の構造アライメント
 D, 組合わせ4 (SLC2A1のC末端側とSLC2A3のC末端側の比較) の構造アライメント

次に、SLC2A1の立体構造を中心軸に沿って回転させ、どの角度において構造の一致度が上がるのか検証した。回転角度によるRMSDの推移を図3-3Cに示す。このRMSDは、初期構造と回転構造の各C α 同士に関して、残基番号に関わらず互いに5Å以内に位置する全ての組合せを抽出し、その距離の根平均二乗偏差を算出することにより求めた。したがって、回転角度0度の場合（初期構造と全く同じ場合）でも、RMSDはゼロにはならない。隣り合う残基のC α が5Å以内に位置する場合、RMSDの計算に組込まれてしまうためである。しかし、こうすることで二つの立体構造間における残基番号や対応するアミノ酸を考慮することなく、全体の構造類似度の推移を観測できるため、このような計算手法を用いた。

その結果、回転角度が20-170度および190-340度においては、RMSDは常に3.8Å以上であったが、回転角度が175-185度の時に、RMSDが3.75Å前後まで低下することが分かった（図3-3C）。回転角度が175-185度の時のRMSD値は、回転角度が15度および345度の時よりも小さくなる。つまり、約180度回転させたところで構造の一致度が上がるということが分かった。

初期構造および180度回転構造の重ね合わせを図3-3Dに示した（初期構造、青；180度回転構造、緑）。細胞内および細胞外のループ部分に構造の重なりが見られない一方で、膜貫通領域に相当する部分において、 α ヘリックス構造の重なりが確認できた。構造アライメントの結果と同様に、 α ヘリックスの方向や長さは保存されていた。

つまりSLC2のN末端側とC末端側は、アミノ酸配列の一致度が低いにも関わらず、 α ヘリックス構造が類似していた。したがって、これらは、過去に共通祖先配列から遺伝子重複によって生じたことが示唆された。

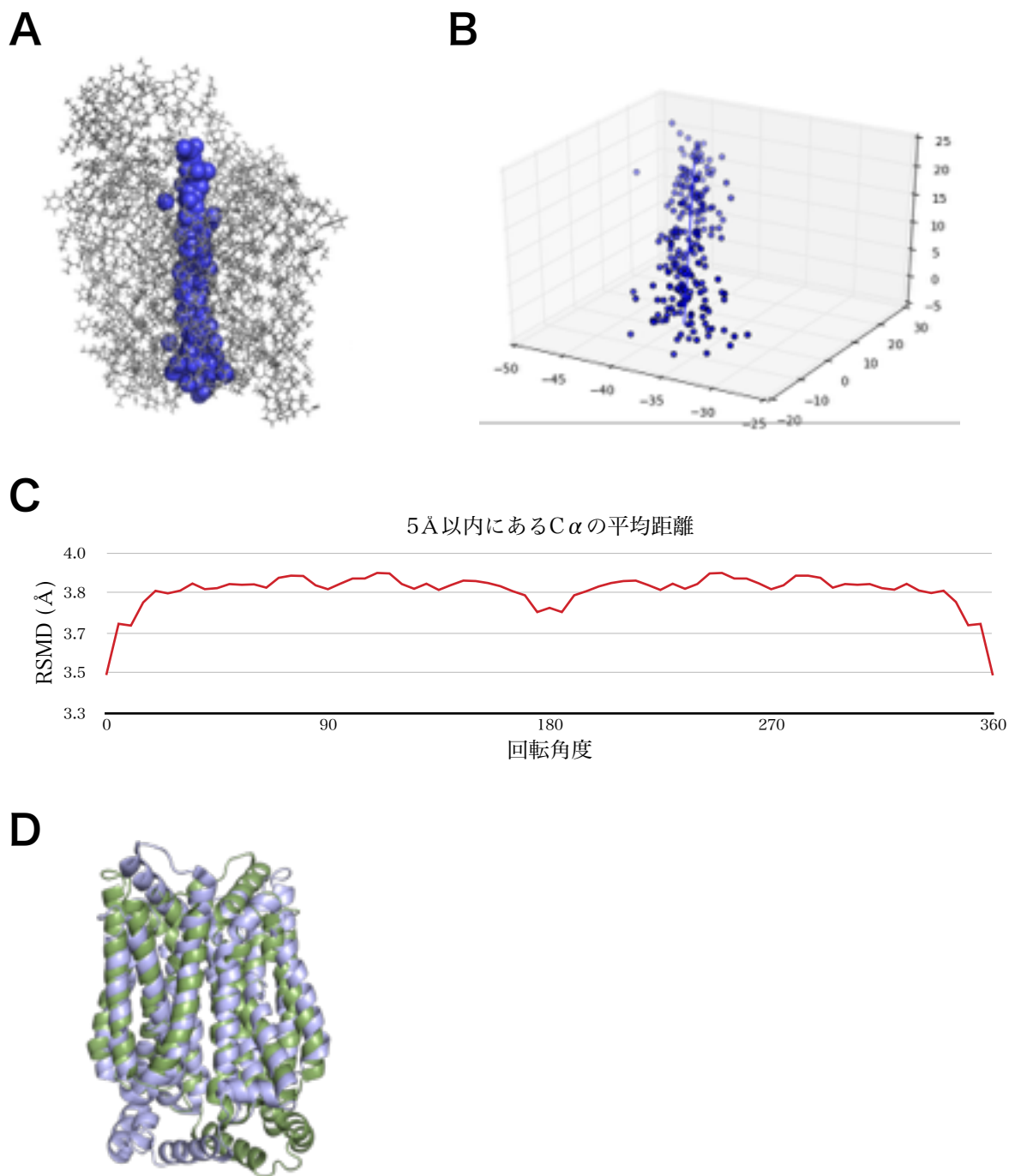


図3-3 回転角度ごとのRMSDの算出と構造の比較

- A, N末端側配列とC末端側配列のアライメントにおける残基ペアの二等分点 (青球)
- B, 二等分点 (青点) とそこから導出された近似線 (青線)
- C, 初期構造と中心軸に沿って5°ずつ回転させた構造を比較したRMSDの推移
- D, 初期構造と180°回転させた構造の構造を重ね合わせ (青, 初期構造; 緑, 180°回転構造)

3.3.2 隠れマルコフモデルを用いたホモロジー検索

前述の検証から、SLC2のN末端側とC末端側の配列は、共通祖先由来である可能性が示唆された。したがって、ホモロジー検索を行うにあたり、SLC2の配列情報をN末端側とC末端側の間で分割したハーフサイズの配列情報を検索クエリーとして用いることとした。そうすることで、重複を起こさずに現在まで存続しているハーフサイズのままの配列や、逆に重複を何度も起こしたホモログ、異なるドメインと融合したホモログなども含めて、幅広いホモログを検索できる可能性があると考えた。

ここでは、N末端側の配列情報から作成したプロファイルHMM (N-side profile)、C末端側の配列情報から作成したプロファイルHMM (C-side profile)、N末端側とC末端側の配列情報を両方用いて作成したプロファイルHMM (N+C profile) の三種類のプロファイルHMMをクエリーとして、HMMerに装備されたホモロジー検索プログラムhmmsearchによりホモログタンパク質を検索した。また、E-valueが 10^{-4} 以下のものをヒットタンパク質と見なした。

ホモロジー検索の結果ヒットしたタンパク質の総数と、ドメイン数ごとの集計を表3-2および図3-4Aに示した。N+C profileをクエリーとして用いた場合、ヒットタンパク質の総数は267,208であった。同様に、N-side profileをクエリーとして用いた場合は226,403、C-side profileをクエリーとして用いた場合は94,903であった。N+C profileを用いた場合において、最も多くのホモログタンパク質を検索することができた。C-side profileをクエリーとして用いた場合の方が、N-side profileをクエリーとした場合よりも、ホモログのヒット数が少なかった。このことから、N末端側よりもC末端側は進化的に後から発展し、進化の歴史が浅い可能性が示唆される。また、ホモログが保持するプロファイルHMMと相同なドメインの数は、1つか2つが多かった。しかし、少数ではあるが、3つ以上の重複をもつものも確認された。

次に、ヒットタンパク質のアミノ酸配列長の分布を全体・ドメイン数1・ドメイン数2の場合に分けてヒストグラムにより表した（図3-4B）。保持するドメインの数に限らず、ヒットタンパク質のアミノ酸配列長は400-450残基前後にピークがくることが分かった。クエリーとして用いたプロファイルHMMの配列長は約230-250残基であったため、多くのヒットタンパク質はドメインを二つ保持しているか、一つしか保持していない場合でも他のドメインと融合しているなどして全体の配列長が400残基以上となっていると考えられる。逆に、配列長が380残基以下のタンパク質は一見非常に少ないように見えるが、ドメイン数1の場合の配列長分布（図3-4B, one domain）を注意深く見てみると、配列長200残基前後に数千単位のヒットタンパク質が分布していることが確認できた。これらは配列長の都合から考えても、プロファイルHMMと相同なドメインのみをもつホモログであると考えられる。このようなホモログは、C-side profileよりも、N-side profileとN+C profileをクエリーに用いた場合に多くヒットしていた。

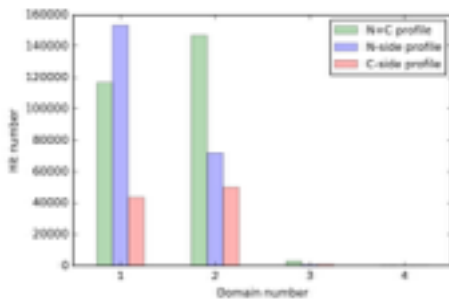
ヒットタンパク質のE-valueの分布を全体・ドメイン数1・ドメイン数2の場合に分けてヒストグラムにて示した（図3-4C）。全体としては、E-valueが下がるほどヒット数は下がっていく傾向にあるが、C-side profileは特定のE-valueで小さなピークを数回あることが確認できた。また、N+C profileのE-valueはドメイン数2の場合に、その他のプロファイルよりも低いE-valueでタンパク質がヒットしていることが分かった。

各プロファイルによるヒットタンパク質の重複をベン図により示した（図3-5）。N+C profileのみヒットしたホモログの数は87,336であった。N末端側とC末端側の配列を混合したプロファイルを使用することで、どちらか片方だけのプロファイルを使用するよりも広い範囲のホモログ検索ができることを確認した。

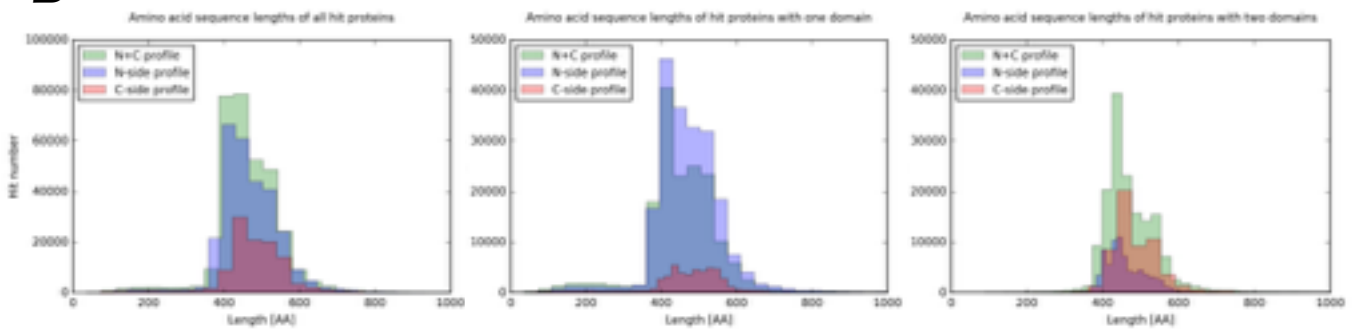
表 3-2 HMMerを用いたホモロジー検索の結果

使用したプロファイル	N+C profile	N-side profile	C-side profile
ヒット数合計	267,208	226,403	94,903
1	116,784	153,423	43,651
2	146,779	72,015	50,182
3	3,229	841	912
4	305	101	104
5	53	10	40
6	38	6	11
7	9	3	1
8	8	3	2
9	3	1	0

A



B



C

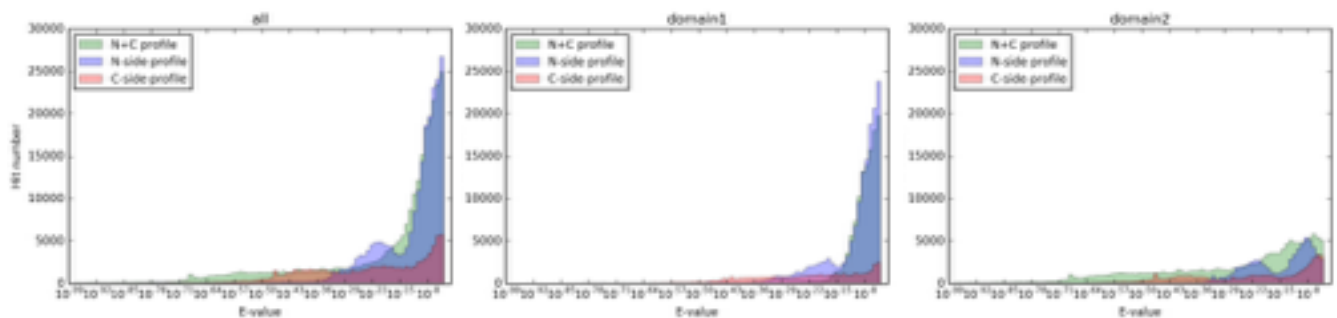


図3-4 HMMerによるホモロジー検索の結果比較

A, 各プロファイルにおけるドメイン数ごとのヒットタンパク質数 ドメイン数とは、ヒットタンパク質のアミノ酸配列における、クエリーと相同な配列の内包数を示す。

B, ヒットタンパク質におけるアミノ酸配列長の分布 全てのヒットタンパク質 (左)、ドメインを1つのみ内包するヒットタンパク質 (中央)、ドメインを2つ内包するヒットタンパク質 (右) のアミノ酸配列長とヒット数の関係をヒストグラムにて示した。

C, ヒットタンパク質におけるE-valueの分布 全てのヒットタンパク質 (左)、ドメインを1つのみ内包するヒットタンパク質 (中央)、ドメインを2つ内包するヒットタンパク質 (右) のHMMerにより算出されたE-valueとヒット数の関係をヒストグラムにて示した。

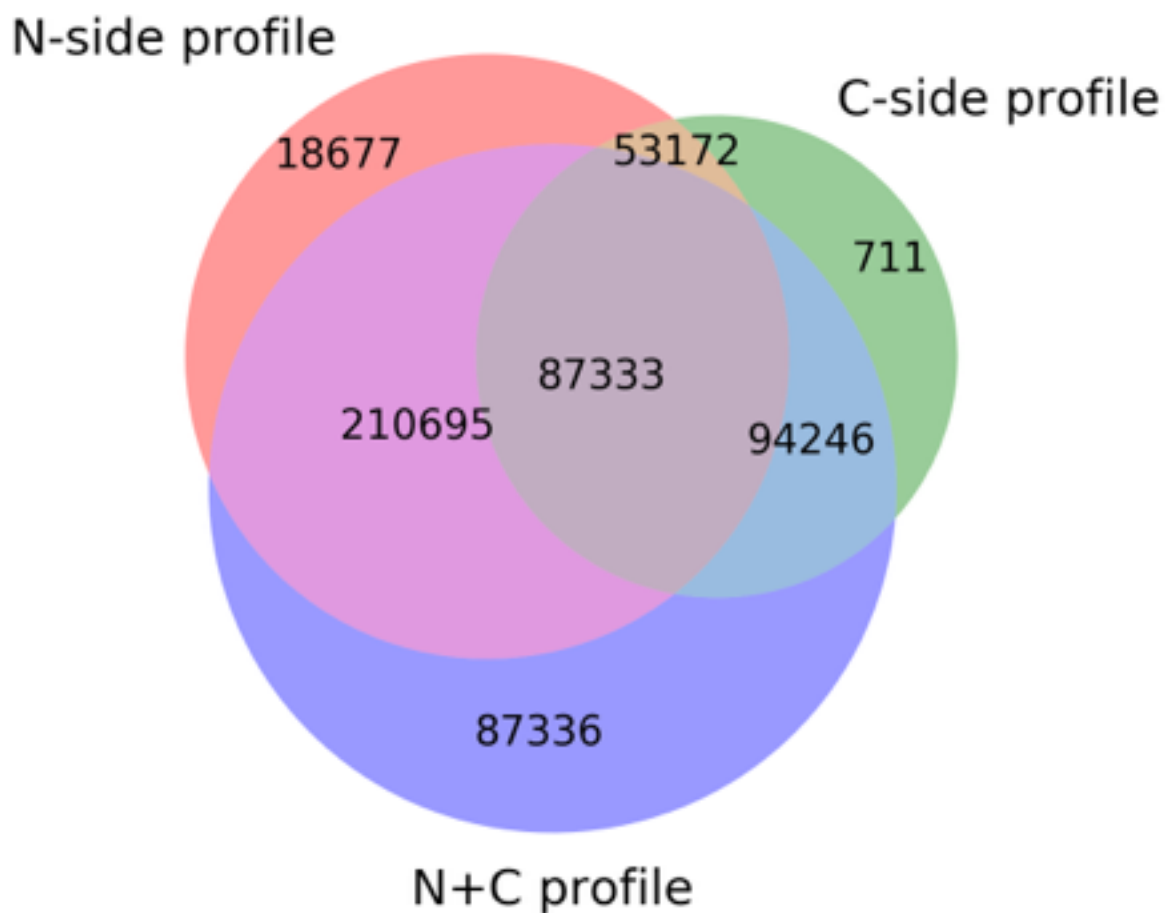


図3-5 各プロファイルのヒットタンパク質の重複

三種類のHMMプロファイル（N+C profile, N-side profile, C-side profile）をクエリーとしてホモロジー検索を行った場合の、ヒットタンパク質の重複および数をベン図により示した。

3.3.3 ヒットタンパク質の生物種

ヒットタンパク質がどのような生物由来であるか確認するため、N+C profileを用いた場合のヒットタンパク質の生物種の分類を図3-6に示した。この生物種は、hmmsearchの結果に含まれるUniProt IDをもとに、UniProtに登録されたTaxonomy情報を参考にした。抽出したTaxonomyのうち、ドメイン (Domain) > 門 (Phylum) > 綱 (Class) の各階層における情報のみを使用した。

ドメインは、Archaea (古細菌)、Bacteria (真正細菌)、Eukaryota (真核生物)、Viruses (ウイルス) が含まれていた。全てのドメインにおける生物が、SLC2のホモログを有していることが明らかになった。ただし、ウイルスはわずか4種類のホモログしか見つからなかったことから、広く普及しているわけではなかった。

これらの生物種に含まれる代表的な例としては、例えばヒトなどの哺乳類が属する脊索動物 (Eukaryota > Metazoa > Chordata)、大腸菌が属するγプロテオバクテリア (Bacteria > Proteobacteria > Gammaproteobacteria)、メタン菌 (嫌気条件下でメタンを合成する古細菌) (Archaea > Euryarchaeota > Methanobacteria) なども含まれていた。各ドメインにおけるヒットタンパク質の数は、古細菌で3,593種、真正細菌で217,717種、真核生物で105,280種、ウイルスで4種であった。

SLC輸送体の進化起源や詳細な進化過程は、まだほとんど明らかになっていない。先行研究では、主にヒトやその他の真核生物を対象に、SLC輸送体ファミリーの進化起源が古く、初期真核生物から分岐が進んでいたであろうと述べられている [Höglund, *et al.* 2011]。しかし、本研究では、古細菌や真正細菌に多くホモログが見つまっていることから、SLC2の進化起源は真核生物初期よりもさらに古いと考えられる。

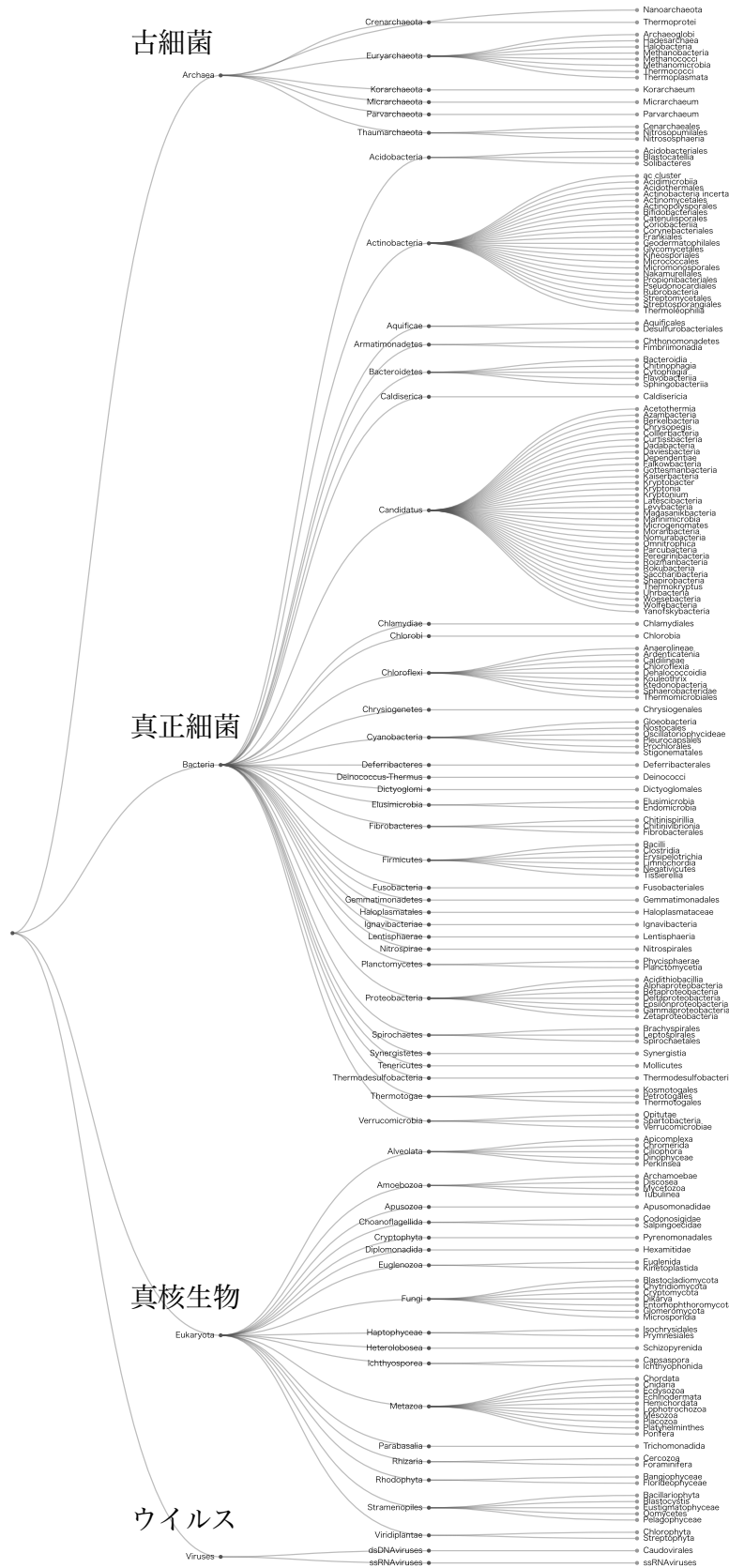


図 3-6 ヒットタンパク質の生物種の分類

N+C profileをクエリーとして用いた際のヒットタンパク質について、由来生物の情報を抽出・分類した。分類は、ドメイン (Domain) > 門 (Phylum) > 綱 (Class) の順に示した。

3.3.4 SLC2ホモログの保存部位解析

最後に、SLC2ホモログの中で進化的に保存された残基およびそのポジションを解析した。hmmsearchの出力ファイルに記載されるアライメント結果から、クエリーとヒットタンパク質の間で一致しているアミノ酸残基を集計した。

まず、N+C profileのヒットタンパク質の保存されたアミノ酸残基を集計を表3-3Aおよび図3-7Aに示す。さらに、ポジションごとの保存アミノ酸の観測結果を図3-7Aに示す。この中で、最も保存箇所が多い残基はアルギニンで、全体の29.7%だった。次いでグルタミン酸23.8%、グリシン22.8%、プロリン10.0%、チロシン9.0%、ロイシン4.7%の順で保存されていた。アルギニン、グルタミン酸、グリシンの3つの残基は、保存度の高いポジションが主に二箇所に分かれていた。アルギニンで最も保存度が高かったポジションは、クエリー配列の116残基目にあたり、観測数は402,907、全ヒットタンパク質の97.24%で一致していた。次いで、177残基目のアルギニンも、402,597回観測され、一致率は97.16%であった。グルタミン酸の保存度が最も高かったポジションは、170残基目で、観測数は403,316（一致率97.34%）であった。同様に、グリシンでは、115残基目が保存されており、観測数は402,503（一致率97.14%）であった。231残基目のグルタミン酸は、観測数が240,863（一致率58.13%）で、半分以上のヒットタンパク質で保存されていることが分かった。同様に、230残基目のプロリン、74残基目のチロシンは過半数のヒットタンパク質で保存されていた（一致率60%前後）。

次に、N-side profileのヒットタンパク質における保存されたアミノ酸残基を集計を表3-3Bおよび図3-7Bに示す。さらに、ポジションごとの保存アミノ酸の観測結果を図3-8Bに示す。観測された全ての保存箇所の39.8%はグリシンで、1,410,199箇所が検出された。次に観測数が多かったのは、アルギニンで、981,643

箇所であった。そして、ロイシンが356,241箇所、グルタミン酸が295,498箇所、トリプトファンが285,881箇所と続いた。これらの中で、最も保存度の高いポジションの一つは177残基目のグリシンで、295,556箇所が検出され、99.1%で保存されていた。また、129残基目、141残基目および214残基目のグリシンも90%以上のヒットタンパク質の中で保存されていた。アルギニンに関しては、保存度の高い部位が3箇所検出された。まず、173残基目は、カウント数が295,583箇所（一致率99.11%）、142残基目は293,287（一致率98.34%）、200残基目で293,476箇所（一致率98.41%）であった。ロイシンは231残基目の保存度が高く、272,211箇所（一致率91.28%）が検出された。193残基目のグルタミン酸や228番目のトリプトファンも90%以上の保存度を示した。

同様に、C-side profileのヒットタンパク質における保存されたアミノ酸残基の集計を表3-3Cおよび図3-7Cに示す。また、ポジションごとの保存アミノ酸の観測結果を図3-7Cに示す。アルギニンの検出数が最も多く、328,169箇所（24.5%）、二番目のロイシンは283,693箇所（21.2%）、三番目にグリシンが212,583箇所（15.9%）、四番目にグルタミン酸199,218箇所（14.9%）が続いた。アルギニンに関してはその他のN-side profileと同様、3箇所が高度に保存されていた。まず、残基番号114は観測数102,932（一致率91.52%）、残基番号115は観測数102,948（一致率91.53%）、残基番号181の観測数は102,688（一致率91.30%）でいずれも90%を超えるヒットタンパク質でこれら残基が保存されていることが分かった。ロイシンは4つのポジションで保存されており、それぞれの保存率は40-80%であった。グリシンは、113残基目で102,946箇所が検出され、91.53%で保存されていた。また、95残基目のグリシンも97,017箇所（一致率86.26%）が観測された。174番目のグルタミン酸も、102,834箇所（一致率91.43%）で保存度が高かった。

各プロファイルと比較すると、N+C profileにおける保存箇所は9箇所のみで、その他2種類のプロファイルの保存箇所数（N-side profile, 14箇所; C-side profile, 15箇所）より少なかった（図3-8）。また、N-side profileおよびC-side profileにおいて、ロイシンが保存残基の上位に位置しているのに対し（N-side profile, 3位（10.1%）； C-side profile, 2位（21.2%））、N+C profileではその割合は6位（4.7%）であった（図3-7）。このような違いが生じる理由として、N+C profileは、N末端側とC末端側の配列情報を両方保持しているため、N末端側とC末端側で共通して保存された部位が検出される傾向にあることが考えられる。たとえば、N-side profileの保存残基77G, 124RとC-side profileの保存残基330G, 331Rは左右対称な位置にある相同な残基であり、N+C profileの保存残基115G, 116Rに相当する。

表3-3 ヒットタンパク質において保存された残基の内訳

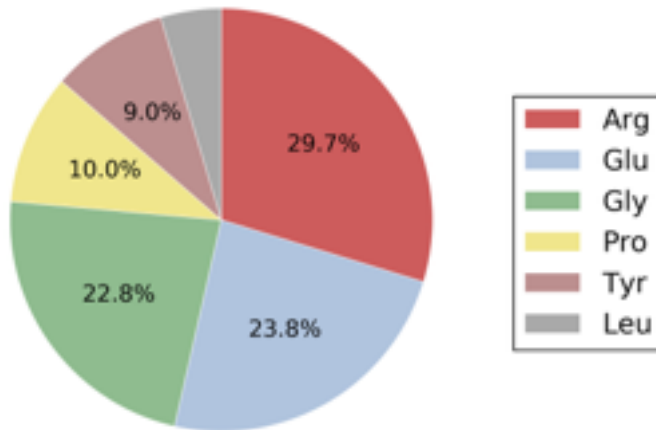
A	保存されたアミノ酸	一致箇所合計	残基番号			残基番号		
			一致数	一致率 (%)	一致数	一致率 (%)		
	Arg (R)	816,771	116	402,907	97.24	177	402,597	97.16
	Glu (E)	654,932	170	403,316	97.34	231	240,863	58.13
	Gly (G)	627,693	115	402,503	97.14	67	218,747	52.79
	Pro (P)	274,103	230	268,917	64.90			
	Tyr (Y)	247,207	74	246,197	59.42			
	Leu (L)	127,429	237	122,638	29.60			

B	保存されたアミノ酸	一致箇所合計	残基番号			残基番号		
			一致数	一致率 (%)	一致数	一致率 (%)		
	Gly (G)	1,410,199	177	295,556	99.10	214	290,505	97.41
141			293,131	98.29	129	272,790	91.47	
125			250,767	84.09				
	Arg (R)	981,643	142	293,287	98.34	173	295,583	99.11
200			293,476	98.41	254	96,079	32.22	
	Leu (L)	356,241	231	272,211	91.28	256	61,488	20.62
	Glu (E)	295,498	193	295,556	99.10			
	Trp (W)	285,881	228	276,233	92.63			

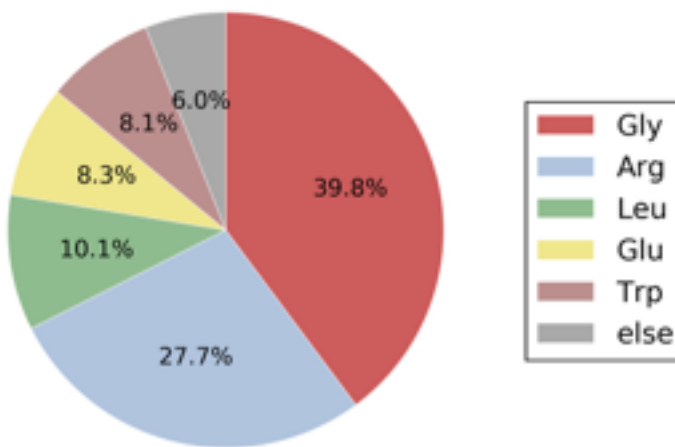
C	保存されたアミノ酸	一致箇所合計	残基番号			残基番号		
			一致数	一致率 (%)	一致数	一致率 (%)		
	Arg (R)	328,169	114	102,932	91.52	115	102,948	91.53
181			102,688	91.30				
	Leu (L)	283,693	13	45,579	40.52	47	69,568	61.85
232			88,452	78.64	240	68,348	60.77	
	Gly (G)	212,583	95	97,017	86.26	113	102,946	91.53
	Glu (E)	199,218	174	102,834	91.43	234	84,005	74.69
	Pro (P)	92,710	233	86,735	77.12			
	Thr (T)	88,624	235	82,630	73.47			
	Tyr (Y)	87,522	72	87,379	77.69			
	Ala (A)	44,503	9	44,511	39.57			

A

Conserved residue found in hit proteins [N+C profile]

**B**

Conserved residue found in hit proteins [N-side profile]

**C**

Conserved residue found in hit proteins [C-side profile]

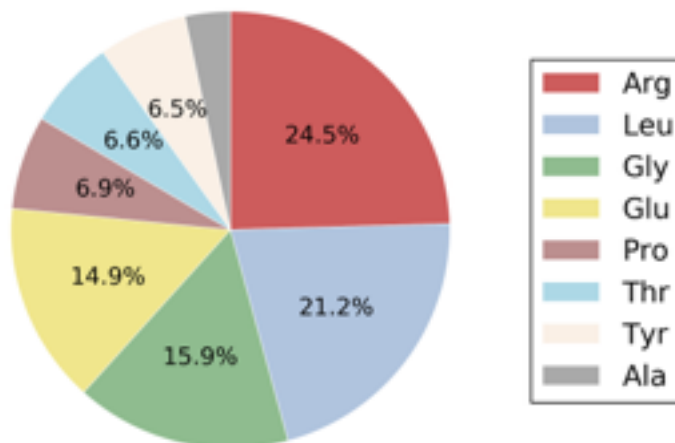
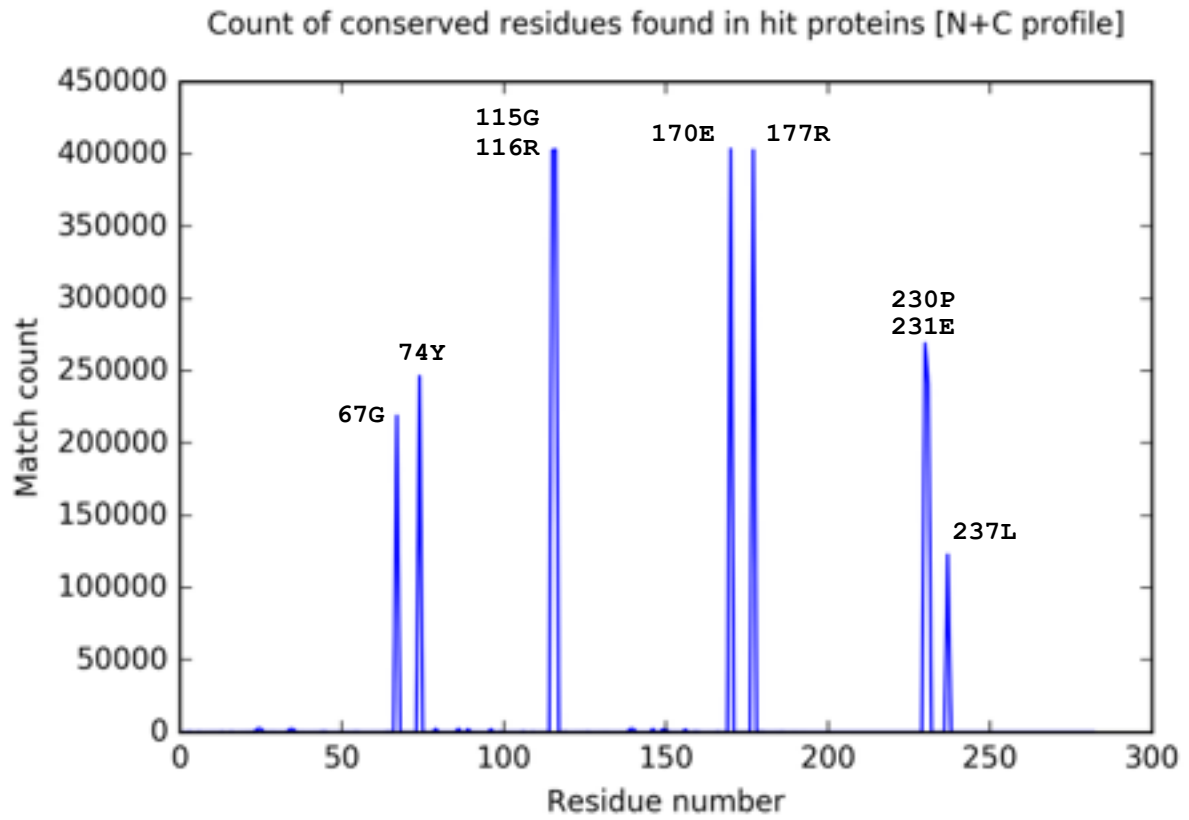
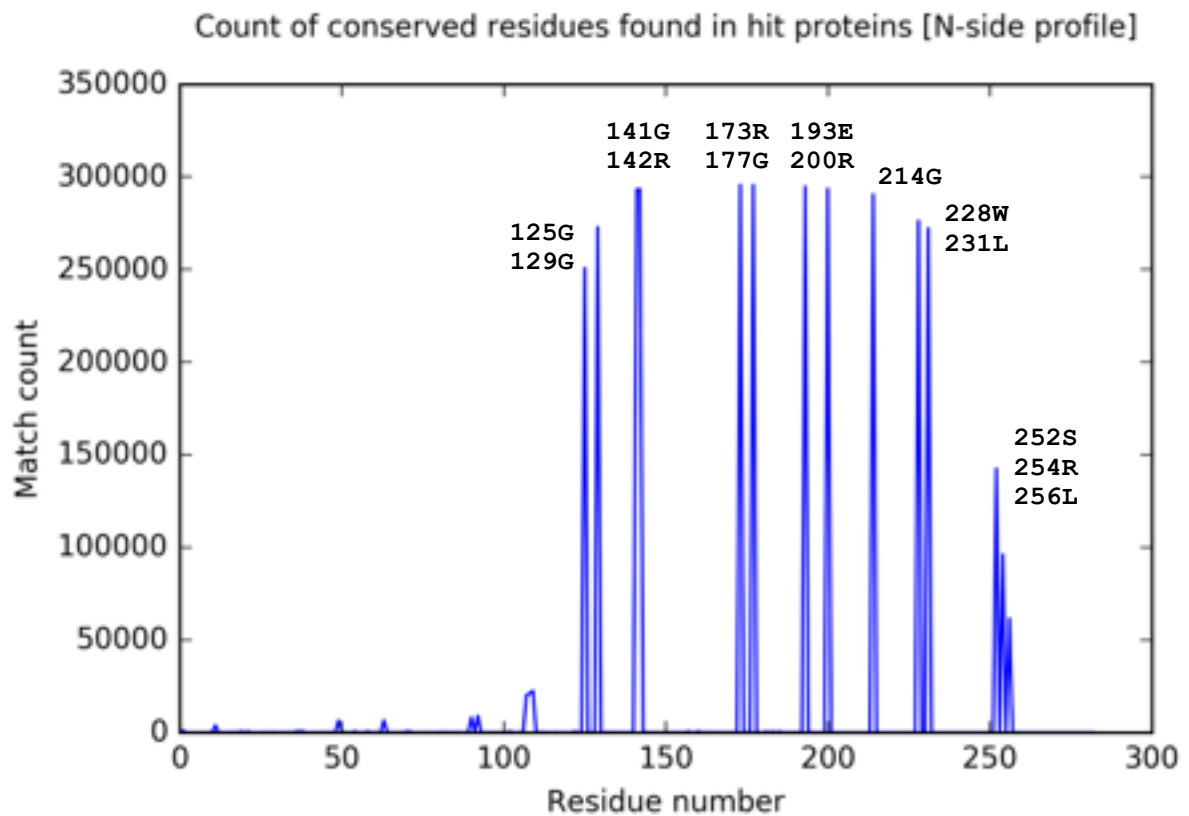


図3-7 ヒットタンパク質において保存された残基の内訳

A, N+C profileをクエリーとした場合のヒットタンパク質において保存された残基と割合

B, N-side profileをクエリーとした場合のヒットタンパク質において保存された残基と割合

C, C-side profileをクエリーとした場合のヒットタンパク質において保存された残基と割合

A**B**

C

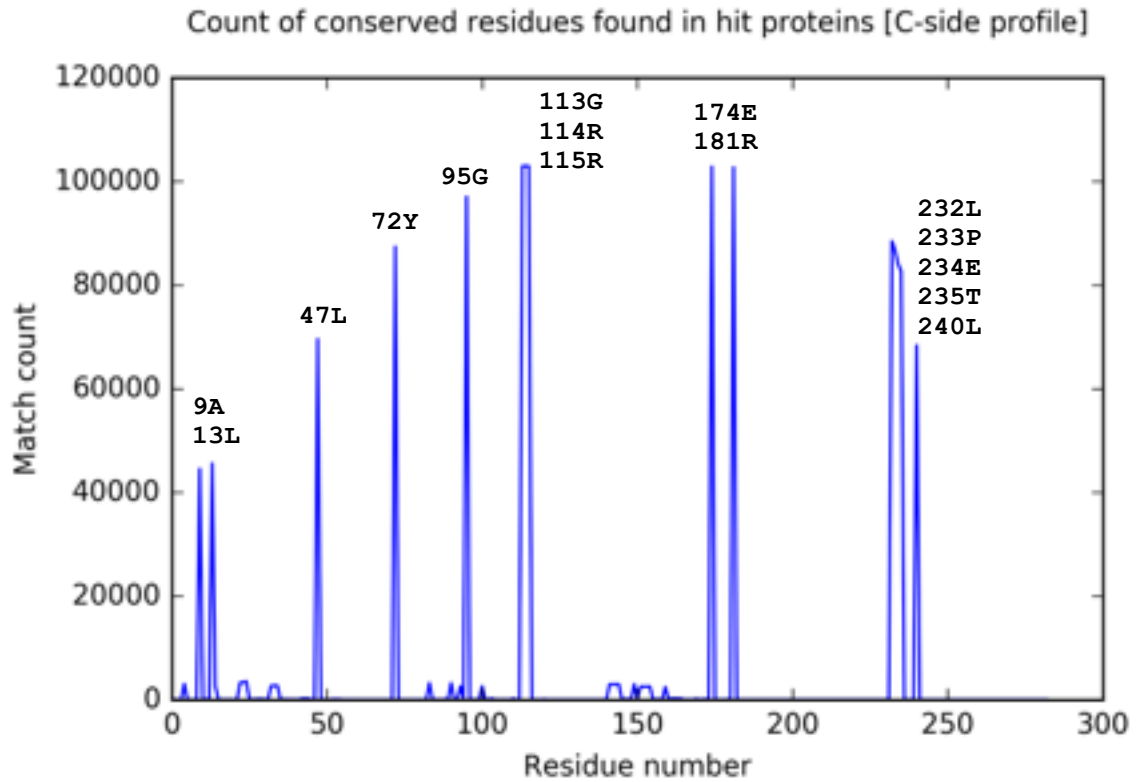


図3-8 ヒットタンパク質の中で保存された残基

- A, N+C profileをクエリーとした場合のヒットタンパク質における保存残基および一致数
- B, N-side profileをクエリーとした場合のヒットタンパク質における保存残基および一致数
- C, C-side profileをクエリーとした場合のヒットタンパク質における保存残基および一致数

3.3.5 保存残基の立体構造マッピング

ヒットタンパク質の中で高度に保存されている残基とその残基番号について、前述の通りまとめた結果を踏まえ、ここではこれら残基の立体構造上の位置を確かめた。マッピングに用いる立体構造はSLC2A1 (PDB ID: 5EQI) およびSLC2A3 (PDB ID: 4ZW9) を使用し、保存されたアミノ酸に対応する残基をそれぞれ赤い球で表した (図3-9)。

N+C profileを用いた場合のヒットタンパク質の保存残基 (図3-9A) は、細胞外ループに2箇所 (67G, 74Y)、細胞質と膜領域の境界面において6箇所 (115G, 116R, 170E, 177R, 230P, 231E)、細胞内ループに1箇所 (237L) 位置していた。特に、細胞質と膜領域の境界面において保存された残基は、互いに距離が近く、一箇所に集中しているように見えた。

N-side profileを用いた場合のヒットタンパク質の保存残基 (図3-9B) は、細胞質と膜領域の境界面において7箇所 (89G, 90R, 144E, 151R, 208S, 210R, 212L)、膜領域から細胞外への境界にかけて7箇所 (73G, 77G, 124R, 128G, 165G, 184W, 187L) が位置していた。151残基目のアルギニンは外開き構造の際はC末端側のドメインに食い込むような形を取っていた。

C-side profileを用いた場合のヒットタンパク質の保存残基 (図3-9C) は、細胞内ループに1箇所 (224A)、細胞質と膜領域の境界面において9箇所 (260L, 330G, 331R, 391E, 398R, 450V, 451P, 452E, 453T)、膜領域から細胞外への境界にかけて2箇所 (291Y, 312G) が位置していた。

これらの結果から、いずれのプロファイルHMMを用いた場合でも、膜領域と細胞質の境界面の残基 (特に一部のグリシン、アルギニン、グルタミン酸) は高度に保存されていることが明らかになった。

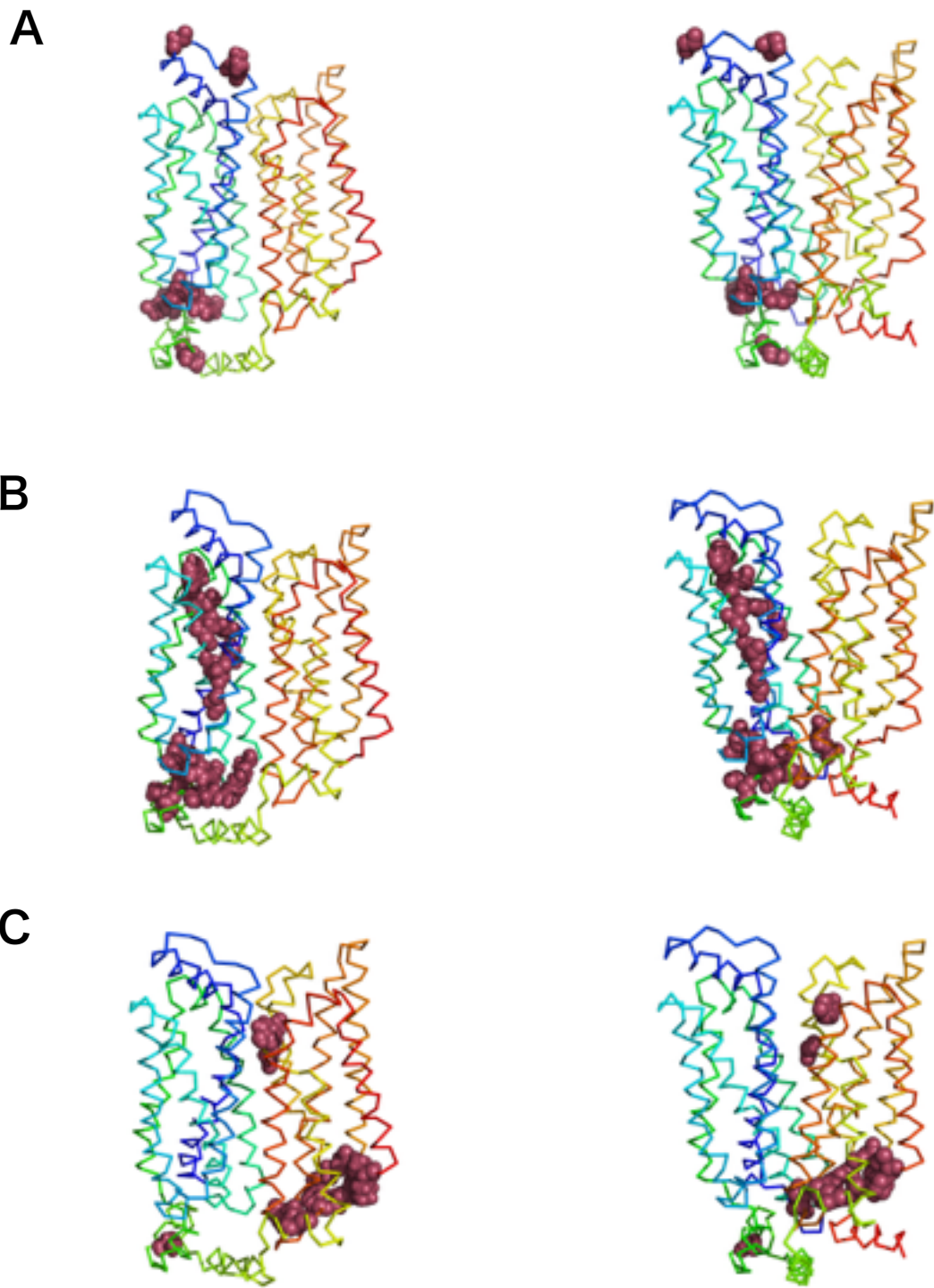


図3-9 ヒットタンパク質で保存された残基の立体構造上へのマッピング

A, N+C profileから得たホモログ保存部位の5EQI (左) および4ZW9 (右) へのマッピング
 B, N-side profileから得たホモログ保存部位の5EQI (左) および4ZW9 (右) へのマッピング
 C, C-side profileから得たホモログ保存部位の5EQI (左) および4ZW9 (右) へのマッピング

- 終章 総合討論 -

総合討論

SLC輸送体は生体内における小分子輸送、特に脳への栄養供給に重要な役割を果たすが、アミノ酸置換により、先天性もしくは乳幼児期など発達の早い段階で神経疾患や代謝疾患を引き起こすことがある。そのような症例は、アミノ酸配列の変化と臨床的意義の関係を紐付けた公共データベースClinVar[Landrum, *et al.* 2014; Landrum, *et al.* 2016]などに蓄積されてきた。しかし、アミノ酸配列変化が立体構造や動的变化に与える影響はほとんど明らかになっていない。そこで、これらを明らかにすることにより、アミノ酸配列変化と臨床的意義の関係に関して、深い妥当性や信頼性をもたせることができると考えた。このような背景を踏まえ、本研究全体の目的は、SLC輸送体におけるアミノ酸置換が、動的構造変化に与える影響を明らかにすることとした。SLC輸送体に関する先行研究では、本研究のような統合データベースの構築から疾患変異の特徴の絞り込みを行い、さらにその影響を分子動力学シミュレーションで検証したという例はない。

本研究の構成として、まず第1章では、ヒトのSLC輸送体に関する情報を紐付けし、計算機により統合解析を実装するためのプラットフォームを構築した。制作したデータベースおよびアプリケーションを使用して、SLC輸送体の変異に関する統計を算出し、立体構造上のどのような部位に疾患変異が集中しているのか明らかにした。第2章では、第1章において特定したSLC輸送体に特有の疾患変異が、動的变化にどのような影響を与えるのか、分子動力学シミュレーションにより分析した。このように、第1章および第2章ではヒトのSLC輸送体、特にヒトSLC2に着目した分析を行った。

第3章では、ヒト以外の生物に分析対象を広げ、進化的に保存されたアミノ酸残基の探索を行った。

第1章の研究において、ヒトSLC輸送体のリレーショナルデータベースiMusta4SLCを完成させた。iMusta4SLCは、性質の異なる複数のデータベースからSLC輸送体に特化したデータを抽出し、計算機により統合解析ができるように設計した。先行研究におけるSLC輸送体特化型データベースは、ユーザがクエリーをかけたり、ダウンロードできる機能がなかったため、手動で大量のデータを整理する必要があった。また、立体構造の情報と配列・変異情報を紐付けたSLC輸送体のデータベースはこれまで存在しなかった。iMusta4SLCの構築により、これらの問題が解消され、変異が立体構造のどの位置に存在し、どのような頻度で観測されているのか、容易に調べられるようになった。

SLC輸送体の多くは未だに立体構造が解明されておらず、本来ならば、変異が立体構造上のどの位置に存在するのか知ることはできない。そこで、本研究ではアミノ酸配列情報から膜タンパク質の大まかな構成、つまり、トポロジーを予測する手法を用いることにより、変異の位置を評価した。また、従来の膜タンパク質のトポロジー予測は、「細胞外」、「膜領域」、「細胞内」の三分類法であったが、本研究では「細胞外と膜領域の境界」および「膜領域と細胞内の境界」を加えた五分類法に改変した。その結果、疾患変異の有意な偏在を捉えることに成功した。

iMusta4SLCの制作には、リレーショナル・データベース・マネジメントシステム(RDBMS)を構築するためのオープンソースシステムであるMySQLを採用した。RDBMSは、異なる属性の情報をテーブルごとに分割して保持し、主キーおよび外部キーの設定により、情報を紐付けできるようにするシステムである。本研究で構築したデータベースは、複数の異なる属性のデータを保持している。具体的には、SLC

輸送体のDNA配列・アミノ酸配列・変異・疾患情報・立体構造情報・トポロジー情報などが挙げられる。RDBMSにより、これら様々な情報を自在に紐付けることにより、統計計算やマッピング、分子動力学シミュレーションのモデル探索を行うことが可能になった。変異とトポロジーの統合解析からは、有意にR4領域（膜領域と細胞内領域の境界）に疾患変異が偏在していることが明らかになった($p < 0.01$)。一方で、R4領域における非疾患変異は、他の領域と比較して頻度が少ないことも分かった($p < 0.05$)。これらの結果から、SLC輸送体のR4領域における変異は、疾患と関連が深いことが示唆された。

ただし、本研究で行ったトポロジー予測は必ずしも正確とは限らないという点には十分注意する必要がある。本研究で使用したトポロジー予測プログラムTopCons2や、その結果をもとに定義したR1-R5領域は、あくまでもトポロジーの予測結果である。したがって、変異の立体構造上における位置のおおまかな指標として用いたに過ぎない。実験的に測定された、より正確なトポロジー情報については、立体構造の解明と同時に、今後次第に情報がPDBTMやOPMに蓄積されていくと考えられる。現時点では、PDBTMやOPMにおけるSLC輸送体の登録数が少なかったため、解析に用いることができなかった。しかし、立体構造解析の発展とともに、将来的には精密な予測値を用いることが可能になると考えられる。同時に、変異の登録数も今後増え続けると考えられる。現在のSLC輸送体の変異データ量については、非疾患変異の登録数（130）が疾患変異（545）と比較して不足している。これは、健常者よりも、疾患に罹患している患者を対象としたゲノム解析が多く実施されているなどの理由で、被験者にバイアスが生じているためであると考えられる。

第2章の研究においては、SLC2（グルコース輸送体）を例として、野生型と変異導入型の動的構造を分子動力学シミュレーションにより比較した。先行研究では、主に野生型SLC2の分子動力学シミュレーションを中心に解析が進められてきたが、疾患変異を導入したモデルについてはまだ十分に検証されていない。本研究では、第1章で明らかにしたSLC輸送体で高頻度に観測される疾患変異を中心とした変異導入モデルを作成し、野生型との比較を行った。

一般的に、分子動力学シミュレーションのモデルには、全原子モデルもしくは粗視化モデルが使用される。しかし、本研究では、外開き構造と内開き構造の両方で保存された構造および結合のみを抽出し、保存性弾性ネットワークモデルを作成した。さらに、計算時間短縮のため、内開き構造を初期構造としたシミュレーションと、外開き構造を初期構造としたシミュレーションを並列に実施した。このような工夫により、短時間で動的構造を検証することが可能となり、限られた計算リソースで野生型と変異型の比較を行うことができた。

保存性弾性ネットワークモデルにおける観測した野生型SLC2および変異型SLC2の軌道は、主成分分析により比較した。第一主成分は、内開き構造と外開き構造を繰り返す開閉運動を表していた。この開閉運動は、構造全体が下から上に突き出るような上下運動を示していた。第二主成分は、N末端側ドメインとC末端側ドメインを結ぶ大きな細胞内ループがねじれ、それにより各ドメインが前後に倒れるような動きを示した。これは、雑巾を絞るときのような、横ひねりの動きのようにも捉えられた。第一主成分における比較の結果、変異型は野生型と比較して、内開き構造の出現頻度が有意に減少していることが明らかになった。内開き構造の出現頻度が減ることにより、取り込んだ基質を細胞質側に放出できる機会が減少し、結果的に

基質輸送効率が低下する可能性が示唆された。第二主成分以降に関しては、野生型と変異型の間で差は見られなかった。

考察において留意すべき点として、保存性弾性ネットワークモデルは全原子モデルと比較してシミュレーション精度が低いことが挙げられる。そのため、本来は、バリデーションのために、全原子モデルでも同様の結果が得られるか確認する必要がある。しかしながら、全原子モデルで本研究と同様の分析を行うには、現在の計算リソースが十分ではない。各野生型および変異型の全原子モデルを1マイクロ秒単位でシミュレーションし、統計計算のために少なくとも3回以上再現実験をこなす必要があるが、計算リソースを考慮すればあまり現実的ではない。本研究におけるSLC2の野生型および変異型の比較は、あくまで動的構造に本質的に重要な部位だけを抽出したモデルとして見なすべきである。

また、本計算における脂質モデルは、先行研究に従ってPOPC（フォスファチジルコリン）を採用した。しかし、実際の生体内における脂質二重層は、様々な種類の脂質分子により構成されている。膜タンパク質の動態は、周囲の脂質の種類や存在率に影響を受ける。そのため、生体内では、本計算で得られた動的変化よりも、さらに複雑な動きをしているはずである。最終的に、変異型において基質輸送効率が低下していることを裏付けるためには、生体内に近い脂質環境の再現と、基質を加えた状態での分子動力学シミュレーション、もしくは実験的に精密な検証が必要となるだろう。

第3章では、ヒトSLC2のホモログタンパク質において、立体構造上のどの部位が高度に保存されているか検証した。ヒトを含め、全生物の輸送体のアミノ酸配列は、分子進化の過程において、祖先配列からアミノ酸置換が蓄積されてきた結果とみな

せる。高度に保存された残基は、立体構造・動的構造変化・機能のいずれかに重要な役割を果たし、代替の利かない部位であると考えられる。また、そのような残基にアミノ酸置換が起これば、生存に不利な表現型を発現する可能性が高い。そこで、幅広い生物におけるSLCホモログのアミノ酸配列を検索し、保存部位を調べた。

ホモロジー検索には、隠れマルコフモデルによる配列検索プログラムHMMer [Eddy, 1995; Mistry, *et al.* 2013]を使用した。入力値としては、SLC2のN末端側の情報から作成したプロファイルHMM (N-side profile)、C末端側の情報から作成したプロファイルHMM (C-side profile)、およびN末端側とC末端側の情報を混合して作成したプロファイルHMM (N+C profile) を用いた。いずれの結果においても、膜領域と細胞質の境界面における残基が保存されていた。HMMerによる検索結果から、ヒットタンパク質の中にはプロファイルと相同なドメインを多数もしくは奇数個もつものがあり、重複を何度も起こしている痕跡が見られた。進化解析の観点からは、今後、本研究で収集したSLC2ホモログの情報から、SLC2の祖先配列に近い形をしたタンパク質の探索などもできるかもしれない。

まとめとして、リレーショナルデータベースを構築し、変異と立体構造情報を含めた統合解析を行った結果、疾患と深く関連していると考えられる変異の特徴をつかむことができた。このような結果は、個々のSLC輸送体を分析しただけでは明らかにできない。したがって、配列・構造・機能・変異の情報を統合したデータベースを構築することは、タンパク質ファミリー全体におけるアミノ酸置換と機能変異の傾向を掴むために非常に有効であることが分かった。さらに、本研究で開発した保存性弾性ネットワークモデルにより、従来よりも短時間の計算で輸送体の動的構造変化を捉えることに成功した。ここでは、SLC2を分析対象にして、動的構造変化に与

える影響について計算および考察を行った。SLC2を対象に選んだ理由は、ヒトSLC2の立体構造がすでに解明されており、疾患変異であるとの報告が豊富に登録されていたからである。SLC1もヒトの立体構造が解明されているが、変異の報告数が不十分であると判断し、本研究では扱わなかった。ヒト以外の立体構造を用いる場合、例えばSLC7はホタル由来 (*lampyris turkestanicus*) の立体構造が明らかになっており (PDB ID: 4M48)、疾患変異および非疾患変異が報告されている。SLC7の疾患変異はシスチン尿症やリジン尿性蛋白不耐症を引き起こすことが知られており、いずれも先天性もしくは乳幼児期に発症・診断されることが多い。ただし、ヒトにおける疾患変異が、必ずしも他の生物で同様の疾患を引き起こすとは限らない。立体構造の由来となった生物種を用いて、予め変異の影響を実験的に確認する必要があるため、変異と立体構造の由来生物が異なる場合には、結果の評価に十分に注意すべきである。今後は立体構造の解明および計算資源の向上に伴い、その他のSLCファミリーについても、本研究と同様の解析が可能になると期待される。

今後の研究構想として、タンパク質におけるアミノ酸変異が、動的構造変化に及ぼす影響について予測した結果を、統合データベースiMusta4SLCに搭載を検討している。各変異についての臨床的知見と計算科学的予測を照合・学習することで、立体構造および動的構造変化の視点から、アミノ酸変異と表現型の関係について、さらなる統合解析を通して原理を理解したい。また、未知の変異や報告の少ない変異についても、従来より高い精度でその影響を予測できるようにしたい。

以上により、SLC輸送体におけるアミノ酸置換と構造/機能変異の関係を検証するためには、計算機によりこれら情報を接続することが有用であることが示された。

謝辞

本研究は、私を指導委託学生として快く受け入れてくださったお茶の水女子大学由良敬教授の大変温かく親身なご指導ご鞭撻により実現したものです。未熟で至らぬ点ばかりの私に、研究者としての基本的な心構え及び生命情報学研究の基礎を丁寧且つ辛抱強くご教授下さいました。心より厚く御礼申し上げます。

主査の菅野純夫先生は、どんなに困難な状況でも寛大に見守って下さり、常に心強いお力添えをして下さいました。これまで、何不自由なく最後まで研究生活を続けることができたのも、一重に菅野先生のご配慮によるものです。この場を借りて、深く御礼申し上げます。

加納信吾先生には、学位論文の副審査員として、専攻の副アドバイザーとして、さらにリーディング大学院での副指導教員として、この五年間多くのご指導ご鞭撻を賜りました。前向きな激励に何度も励まされ、研究生活を続けて参ることができました。深く感謝申し上げます。佐藤均先生には、専攻の学位論文の副審査員及び副アドバイザーとして、大変丁寧で細やかなご指導と、温かい励ましをいただきました。特に、本論文の仕上げに際しましては、佐藤先生の貴重なご指摘ご意見を多く参考にさせていただきました。深く感謝申し上げます。富田耕造先生及び渋谷哲朗先生には、学位論文の副審査員をしていただき、短い時間ながらも的確かつ貴重なご助言を多くいただきました。本論文の仕上げに際しましては、先生方のご助言を大変参考にさせていただきました。

共同研究者のアムステルダム自由大学Anton K. Feenstra先生には、大変貴重な技術面からのご協力ご支援と、多くの独創的なご助言をいただきました。さらに、最初は不安も多かった留學生活でも、常に温かく前向きな励ましを何度もいただいた

お陰で、明るく実りの多い思い出となりました。厚く御礼申し上げます。また、共同研究者の野中尚輝博士には、惜しみない技術支援と研究討論、生活面及び精神面における多大な激励を最後まで賜りました。心より御礼申し上げます。

お茶の水女子大学由良研究室の皆様には、熱心で自由闊達な議論を通して多くの励ましとご助言をいただきました。いつも真剣な研究討論を交え、応援して下さった現名古屋工業大学助教授山田大智先生、坂本美佳氏、Ha Duong氏、大島実莉氏、川上夢香氏、高野怜美氏、牧野美希氏、その他後輩の皆様には、心より感謝申し上げます。

最後に、これまで、あらゆる面で惜しみなく支援し応援し続けてくれた両親、妹、祖母に、心より感謝の意を表します。

- 参考文献 -

Lukas, K. Buehler. CELL MEMBRANES. Garland Science (2016).

Colas, C, Unga, P. M-U, and Schlessinger, A. SLC transporters: structure, function, and drug discovery. *Med. Chem. Commun.* **7**, 1069–1081 (2016).

Keogh, J., Hagenbuch, B., Rynn, C., Stieger, B. and Nicholls, G. Membrane Transporters: Fundamentals, Function and Their Role in ADME. RSC Drug Discovery Series No. 54. Drug Transporters: Volume 1: Role and Importance in ADME and Drug Development. chapter 1. The Royal Society of Chemistry 2016.

Dickens, D., Radisch, S. and Primohamed, M. Drug Transporters at the Blood-Brain Barrier. RSC Drug Discovery Series No. 54. Drug Transporters: Volume 1: Role and Importance in ADME and Drug Development. chapter 5. The Royal Society of Chemistry 2016.

Pardridge, W. M. The Blood-Brain Barrier: Bottleneck in Brain Drug Development. *NeuroRx.* **2**, 3-14 (2005).

Uchida, Y., Ohtsuki, S., Katsukura, Y., Ikeda, C., Suzuki, T., Kamiie, J. and Terasaki, T. Quantitative targeted absolute proteomics of human blood-brain barrier transporters and receptors. *J. Neurochem.* **117**, 333-345 (2011).

E. G. Geier, E. C. Chen, A. Webb, A. C. Papp, S. W. Yee, W. Sadee and K. M. Giacomini. Profiling Solute Carrier Transporters in the Human Blood-Brain Barrier. *Clin. Pharmacol. Ther.* **94**, 636-639 (2013).

L. Mastroberardino, B. Spindler, R. Pfeiffer, P. J. Skelly, J. Loffing, C. B. Shoemaker and F. Verrey. Amino-acid transport by heterodimers of 4F2hc/CD98 and members of a permease family. *Nature.* **395**, 288-291 (1998).

R. J. Boado, J. Y. Li, M. Nagaya, C. Zhang and W. M. Pardridge. Selective expression of the large neutral amino acid transporter at the blood–brain barrier. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12079-12084 (1999).

Kalliokoski, A. and Niemi, M. Br. Impact of OATP transporters on pharmacokinetics. *J. Pharmacol.* **158**, 693-705 (2009).

Gao, B., Hagenbuch, B., Kullak-Ublick, G. A. , Benke, D., Aguzzi A. and Meier, P. J. J. *Pharmacol. Exp. Ther.* **294**, 73-79 (2000).

T. Okura, A. Hattori, Y. Takano, T. Sato, M. Hammarlund-Udenaes, T. Terasaki and Y. Deguchi. Involvement of the Pylramine Transporter, a Putative Organic Cation Transporter, in Blood-Brain Barrier Transport of Oxycodone. *Drug Metab. Dispos.* **36**, 2005-2013 (2008).

W. J. Geldenhuys and D. D. Allen. Molecular determinants of blood–brain barrier permeation. *Cent. Nerv. Syst. Agents Med. Chem.* **12**, 95-99 (2012).

Marcus, D. and Mak, L. Membrane Transporters: Fundamentals, Function and Their Role in ADME. RSC Drug Discovery Series No. 54. Drug Transporters: Volume 2, chapter7: Methods and Resources for Transport Proteins in Bioinformatics and Cheminformatics, The Royal Society of Chemistry 2016.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B.J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., Altshuler, D.; International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. **409**, 928-33 (2001).

Green E. D., Watson J. D. and Collins F. S. Human Genome Project: Twenty-five years of big biology. *Nature*. **526**, 29-31 (2015).

Hediger, M. A., Clémenton, B., Burrier, R. E. and Bruford, E. A. The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Mol. Aspects Med.* **34**, 95–107 (2013).

Gouaux, E. and Mackinnon, R. Principles of selective ion transport in channels and pumps. *Science*. **310**, 1461-5 (2005).

Tanaka, K. J., Song, S., Mason, K. and Pinkett, H. W. Selective substrate uptake: The role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochim. Biophys. Acta*. **S0005-2736**, 30261-4 (2017).

Diallinas, G. Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. *Front. Pharmacol.* **5**, 207 (2014).

Fredriksson, R., Nordström, K. J., Stephansson, O., Hägglund, M. G. and Schiöth, H. B. The solute carrier (SLC) complement of the human genome: Phylogenetic classification reveals four major families. *FEBS Lett.* **582**, 3811–3816 (2008).

Schlessinger, A., Yee, S. W., Sali, A. and Giacomini, K. M. SLC Classification: An Update. *Clin. Pharmacol Ther.* **94**, 19-23 (2013).

Nałęcz, K. A. Solute Carriers in the Blood-Brain Barrier: Safety in Abundance. *Neurochem. Res.* **42**, 795-809 (2017).

Morris, M. E., Rodriguez-Cruz, V. and Felmler, M. A. SLC and ABC Transporters: Expression, Localization, and Species Differences at the Blood-Brain and the Blood-Cerebrospinal Fluid Barriers. *The AAPS J.* **19**, 1317-1331 (2017).

- Suhy, A. M., Webb, A., Papp, A. C., Geier, E. G. and Sadee, W. Expression and splicing of ABC and SLC transporters in the human blood-brain barrier measured with RNAseq. *Eur. J. Pharm. Sci.* **103**, 47-51 (2017).
- Benarroch, E. E. Glutamate transporters: diversity, function, and involvement in neurologic disease. *Neurology.* **74**, 259-64 (2010).
- Palmieri, F. The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol. Aspects Med.* **34**, 465-84 (2013).
- Bröer, S. and Palacín, M. The role of amino acid transporters in inherited and acquired diseases. *Biochem. J.* **436**, 193-211 (2011).
- Pramod, A. B., Foster, J., Carvelli, L. and Henry, L. K. SLC6 transporters: structure, function, regulation, disease association and therapeutics. *Mol. Aspects Med.* **34**, 197-219 (2013).
- Wright, E. M., Hirayama, B. A. and Loo, D. F. Active sugar transport in health and disease. *J. Intern. Med.* **261**, 32-43 (2007).
- Wang, D., Kranz-Eble, P. and De Vivo, D. C. Mutational analysis of GLUT1 (SLC2A1) in Glut-1 deficiency syndrome. *Hum. Mutat.* **16**, 224-231 (2000).
- Jiang, X., McDermott, J. R., Ajees, A. A., Rosen, B. P. and Liu, Z. Trivalent arsenicals and glucose use different translocation pathways in mammalian GLUT1. *Metallomics.* **2**, 211-219 (2010).
- Wong, H. Y., Law, P. Y. and Ho, Y. Y. Disease-associated Glut1 single amino acid substitute mutations S66F, R126C, and T295M constitute Glut1-deficiency states in vitro. *Mol. Genet. Metab.* **90**, 193-198 (2007).
- Nezu, J., Tamai, I., Oku, A., Ohashi, R., Yabuuchi, H., Hashimoto, N., Nikaïdo, H., Sai, Y., Koizumi, A., Shoji, Y., Takada, G., Matsuishi, T., Yoshino, M., Kato, H., Ohura, T., Tsujimoto, G., Hayakawa, J., Shimane, M. and Tsuji, A. Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter. *Nat. Genet.* **21**, 91-94 (1999).
- Sharma, S. and Black, S. M. CARNITINE HOMEOSTASIS, MITOCHONDRIAL FUNCTION, AND CARDIOVASCULAR DISEASE. *Drug Discov. Today Dis. Mech.* **6**, e31–e39. (2009).
- Wang, H., Elferich, J. and Gouaux, E. Structures of LeuT in bicelles define conformation and substrate binding in a membrane-like context. *Nat. Struct. Mol. Biol.* **19**, 212-219 (2012).
- Kazmier, K., Claxton, D. P. and Mchaourab, H. S. Alternating access mechanisms of LeuT-fold transporters: trailblazing towards the promised energy landscapes. *Curr. Opin. Struct. Biol.* **45**, 100–108 (2017).

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W. and Maglott, D. R. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862-D868 (2016).

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M. and Maglott, D. R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980-D985 (2014).

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C. G., Ward, S., Dawson, E., Ponting, L., Stefancsik, R., Harsha, B., Kok, C. Y., Jia, M., Jubb, H., Sondka, Z., Thompson, S., De, T. and Campbell, P. J. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777-D783 (2017).

Forbes, S. A., Beare, D., Bindal, N., Bamford, S., Ward, S., Cole, C.G., Jia, M., Kok, C., Boutselakis, H., De, T., Sondka, Z., Ponting, L., Stefancsik, R., Harsha, B., Tate, J., Dawson, E., Thompson, S., Jubb, H. and Campbell, P. J. COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer. *Curr. Protoc. Hum. Genet.* **91**, 10.11.1-10.11.37 (2016).

Brown, G. R., Hem, V., Katz, K. S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K. D., Maglott, D. R. and Murphy, T. D. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.* **43**, D36-42 (2015).

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542 (1977).

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S. and Madden, T. L. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).

Tusnády G. E., Dosztányi Z. and Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics.* **20**, 2964-2972 (2004).

Kozma D., Simon I., and Tusnády G. E. *Nucleic Acids Res.* **33**, D275-D278 (2005).

Kozma D., Simon I., and Tusnády G. E. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524-D529 (2013).

Tsirigos, K. D., Peters, C., Shu, N., Käll, L. and Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401-W407 (2015).

- Tusnády G. E., Kalmár L. and Simon I. TOPDB: Topology Data Bank of Transmembrane Proteins. *Nucleic Acids Research*. **36**, D234-D239 (2008).
- Dobson L., Langó T., Reményi I. and Tusnády G. E. Expediting topology data gathering for the TOPDB database. *Nucleic Acids Research*. **43**, D283-D289 (2015).
- Reynolds, S. M., Kall, L., Riffle, M. E., Bilmes, J. A. and Noble, W. S. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS. Comput. Biol.* **4**, e1000213 (2008).
- Kall, L., Krogh, A. and Sonnhammer, E. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics*. **21**, i251–i257 (2005).
- Viklund, H., Bernsel, A., Skwark, M. and Elofsson, A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics*. **24**, 2928–2929 (2008).
- Viklund, H. and Elofsson, A. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*. **24**, 1662–1668 (2008).
- Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. and Elofsson, A. Prediction of membrane-protein topology from first principles. *Proc. Natl. Acad. Sci. U S A*. **105**, 7177–7181 (2008).
- Pearson, W. R. Finding Protein and Nucleotide Similarities with FASTA. *Curr. Protoc. Bioinformatics*. **53**, 3.9.1–25 (2016).
- Pearson, W. R. and Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U S A*. **85**, 2444-2448 (1988).
- Melén, K., Krogh, A. and von Heijne, G. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.* **327**, 735-44 (2003).
- Bostock, M., Ogievetsky, V. and Heer, J. D³ Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*. **17**, 2301-2309 (2011).
- Baeza-Delgado, C., Marti-Renom, M. A. and Mingarro, I. Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.* **42**, 199–207 (2013).
- Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–3027 (1986).
- Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487-94 (1992).

Bogdanov, M., Xie, J. and Dowhan, W. Lipid-Protein Interactions Drive Membrane Protein Topogenesis in Accordance with the Positive Inside Rule. *J. Biol. Chem.* **284**, 9637–9641 (2009).

Bogdanov, M., Dowhan, W. and Vitrac, H. Lipids and topological rules governing membrane protein assembly. *Biochim. Biophys. Acta.* **1843**, 1475–1488 (2014).

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350 (2016).

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T. and Ben-Tal, N. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **53**, 199-206 (2013).

Ashkenazy, H., Erez, E., Martz, E., Pupko, T. and Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529-533 (2010).

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299-W302 (2005).

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor, D., Martz, E. and Ben-Tal, N. ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics.* **19**, 163-164 (2003).

Thompson, J. D., Higgins, D. G. and Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics.* **23**, 2947-2948 (2007).

Kapoor, K., Finer-Moore, J. S., Pedersen, B. P., Caboni, L., Waight, A., Hillig, R. C., Bringmann, P., Heisler, I., Müller, T., Siebeneicher, H. and Stroud, R. M. Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides. *Proc. Natl. Acad. Sci. U S A.* **113**, 4711-4716 (2016).

Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S. and Yan, N. Molecular basis of ligand recognition and transport by glucose transporters. *Nature.* **526**, 391-396 (2015).

Jo, S., Kim, T., Iyer, V. G., and Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **29**, 1859-1865 (2008).

Brooks, B. R., Brooks, C. L. III, MacKerell, A. D. Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoseck, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M.. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545-1614 (2009).

Jo, S., Kim, T., and Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE*. **2**, e880 (2007).

Jo, S., Lim, J. B., Klauda, J. B., and Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **97**, 50-58 (2009).

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., Klauda, J. B., and Im, W. CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **35**, 1997-2004 (2014).

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., We, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks C. L. III, MacKerell A. D. Jr, Klauda, J. B., and Im, W.. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **12**, 405-413 (2016).

Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D. Jr., and Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B.* **114**, 7830–7843 (2010).

Venable, R. M., Sodt, A. J., Rogaski, B., Rui, H., Hatcher, E., MacKerell, A. D. Jr., Pastor, R. W., and Klauda, J. B. CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature. *Biophys. J.* **107**, 134-45 (2014).

Hsu, P. C., Jefferies, D., and Khalid, S. Molecular Dynamics Simulations Predict the Pathways via Which Pristine Fullerenes Penetrate Bacterial Membranes. *J. Phys. Chem. B.* **120**, 11170-11179 (2016).

Arnarez, C., Uusitalo, J. J., Masman, M. F., Ingólfsson, H. I. de Jong, D. H., Melo, M. N., Periole, X., De Vries, A. H. and Marrink, S. J. Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *J. Chem. Theory Comput.* **11**, 260-275 (2015).

de Jong, D. H., Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., Periole, X., Tieleman, D. P. and Marrink S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **9**, 687–697 (2012).

Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D., and Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comp. Biol.* **6**, e1000810 (2010).

Periole, X., Cavalli, M., Marrink, S. J. and Ceruso, M. A. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **5**, 2531-2543 (2009).

Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S. J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory and Comput.* **4**, 819–834 (2008).

Qi, Y., Ingólfsson, H. I., Cheng, X., Lee, J., Marrink, S. J., and Im, W. CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* **11**, 4486-4494 (2015).

Hsu, P-C., Bruininks, B. M. H., Jefferies, D., Souza, P. C. T., Lee, J., Patel, D. S., Marrink, S. J., Qi, Y., Khalid, S., and Im, W. CHARMM-GUI Martini Maker for Modeling and Simulation of Complex Bacterial Membranes with Lipopolysaccharides. *J. Comput. Chem.* **38**, 2354-2363 (2017).

Macheda, M. L., Rogers, S. and Best, J. D. Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. *J. Cell Physiol.* **202**, 654–662 (2005).

Yan, Y. E., Zhang, J., Wang, K., Xu, Y., Ren, K., Zhang, B. Y., Shi, M., Chen, C., Shi, Q., Tian, C., Zhao, G. and Dong, X. P. Significant reduction of the GLUT3 level, but not GLUT1 level, was observed in the brain tissues of several scrapie experimental animals and scrapie-infected cell lines. *Mol. Neurobiol.* **49**, 991-1004 (2014).

Simpson, I. A., Dwyer, D., Malide, D., Moley, K. H. , Travis, A., Vannucci, S. J. The facilitative glucose transporter GLUT3: 20 years of distinction. *Am. J. Physiol Endocrinol Metab.* **295**, E242–E253 (2008).

Simpson, I. A., Dwyer, D., Malide, D., Moley, K. H., Travis, A., Vannucci, S. J. The facilitative glucose transporter GLUT3: 20years of distinction. *Am. J. Physiol. Endocrinol. Metab.* **295**, E242–E253 (2008).

Buslaev, P., Gordeliy, V., Grudinin, S. and Gushchin, I. Principal Component Analysis of Lipid Molecule Conformational Changes in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **12**, 1019–1028 (2016).

Srikumar, P. S., Rohini, K., Rajesh, P. K. Molecular Dynamics Simulations and Principal Component Analysis on Human Laforin Mutation W32G and W32G/K87A. *Protein J.* **33**, 289–295 (2014).

Dwyer, D. S. Model of the 3-D structure of the GLUT3 glucose transporter and molecular dynamics simulation of glucose transport. *Proteins*. **42**, 531–541 (2001).

Mohan, S., Sheena, A., Poulouse, N., Anilkumar, G. Molecular Dynamics Simulation Studies of GLUT4: Substrate-Free and Substrate-Induced Dynamics and ATP-Mediated Glucose Transport Inhibition. *PLoS ONE*. **5**, e14217 (2010).

Fu, X., Zhang, G., Liu, R., Wei, J., Zhang-Negrerie, D., Jian, X., Gao, Q. Mechanistic Study of Human Glucose Transport Mediated by GLUT1. *J. Chem. Inf. Model.* **56**, 517-26 (2016).

Bond, P. J., Holyoake, J., Ivetac, A., Khalid, S., Sansom, M. S. P. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J. Struct. Biol.* **157**, 593–605 (2007).

Amadei, A., Linssen, A. B., Berendsen, H. J. Essential dynamics of proteins. *Proteins*. **17**, 412-25 (1993).

Stepanova, M. Dynamics of essential collective motions in proteins: theory. *Phys. Rev. E. Stat. Nonlin. Soft. Matter: Phys.* **76**, 051918 (2007).

Kamaraj, B. and Purohit, R. Mutational Analysis on Membrane Associated Transporter Protein (MATP) and Their Structural Consequences in Oculocutaneous Albinism Type 4 (OCA4)—A Molecular Dynamics Approach. *J. Cell Biochem.* **117**, 2608-19 (2016).

倭剛久, 由良敬. 分子動力学とデータベースによる統合解析. 生化学. **85**, 646-655 (2013).

Japanese Society for Bioinformatics. Encyclopedia of Bioinformatics. (2006).

Iwata, S. Structural Biology of Membrane Proteins. (2013).

Yao, X. Q., Kenzaki H., Murakami, S., Takada, S. Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations. *Nat. Commun.* **1**, 117 (2010).

Fredriksson, R., Nordström, K. J. V, Stephansson, O., Hägglund, M. G. A., & Schiöth, H. B. The solute carrier (SLC) complement of the human genome: Phylogenetic classification reveals four major families. *FEBS Letters*. **582**, 3811–3816 (2008).

Sundberg, B. E., Wååg, E., Jacobsson, J. A., Stephansson, O., Rumaks, J., Svirskis, S., Fredriksson, R. The evolutionary history and tissue mapping of amino acid transporters belonging to solute carrier families SLC32, SLC36, and SLC38. *J. Mol. Neurosci.* **35**, 179–193 (2008).

Höglund, P. J., Nordström, K. J. V, Schiöth, H. B., and Fredriksson, R. The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species. *Mol. Biol. Evol.* **28**, 1531–1541 (2011).

- Sreedharan, S., Stephansson, O., Schiöth, H. B., and Fredriksson, R. Long evolutionary conservation and considerable tissue specificity of several atypical solute carrier transporters. *Gene*. **478**, 11–18 (2011).
- Schiöth, H. B., Roshanbin, S., Hägglund, M. G. A., and Fredriksson, R. Evolutionary origin of amino acid transporter families SLC32, SLC36 and SLC38 and physiological, pathological and therapeutic aspects. *Molecular Aspects of Medicine*. **34**, 571–585 (2013).
- Korycinski, M., Albrecht, R., Ursinus, A., Hartmann, M. D., Coles, M., Martin, J. and Lupas, A. N. STAC - A New Domain Associated with Transmembrane Solute Transport and Two-Component Signal Transduction Systems. *Journal of Molecular Biology*. **427**, 3327–3339 (2015).
- Deamer, D., Dworkin, J. P., Sandford, S. A., Bernstein, M. P. and Allamandola, L. J. The first cell membranes. *Astrobiology*. **2**, 371-81 (2002).
- Pohorilleab, A. and Deamer, D. Self-assembly and function of primitive cell membranes. *Res. Microbiol.* **160**, 449-456 (2009).
- Forrest, L. R. and Rudnick, G. The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology (Bethesda)*. **24**, 377–386 (2009).
- Forrest, L. R., Krämer, R. and Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim Biophys Acta*. **1807**, 167–188 (2011).
- Katoh, K., Misawa, K., Kuma, K. and Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. **30**, 3059–3066 (2002).
- Katoh, K., Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-80 (2013).
- Kawabata, T. MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Res.* **31**, 3367-9 (2003).
- Kawabata, T. and Nishikawa, K. Protein tertiary structure comparison using the Markov transition model of evolution. *Proteins*. **41**, 108-122 (2000).
- Eddy, S. R. Multiple Alignment Using Hidden Markov Models. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **3**, 114-120 (1995).
- Mistry, J., Finn, R. D. and Eddy, S. R. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions., A. Bateman, M. Punta. *Nucleic Acids Research*. **41**, e121 (2013).

楯真一. NMR によるタンパク質の動的構造解析. 生化学 第85巻 第8号, 638—645 (2013).

McDermott, A. Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR. *Annu. Rev. Biophys.* **38**, 385–403 (2009).

Liang, B. and Tamm, L. K. NMR as a tool to investigate the structure, dynamics and function of membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 468-74 (2016).

Hagn, F., Etzkorn, M., Raschle, T. and Wagner, G. Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *J. Amer. Chem. Soci.* **135**, 1919–1925 (2013).

Bayburt, T. H., Grinkova, Y. V. and Sligar, S. G. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters.* **2**, 853–856 (2002).

Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. M. and Güntert, P. *Nature.* **440**, 52–57 (2006).

Miyanoiri, Y., Takeda, M. and Kainosho, M. Recent advances in NMR methods for determining the structures and dynamics of larger proteins. *J. Japan. Biochem. Soci.* **88**, 452-464 (2016).

Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. and Gouaux, E. Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature.* **445**, 387-393 (2007).

Kumar, H., Kasho, V., Smirnova, I., Finer-Moore, J. S., Kaback, H. R., Stroud, R. M. Structure of sugar-bound LacY. *Proc. Natl. Acad. Sci. U S A.* **111**, 1784-1788 (2014).

Ott, C. M. and Lingappa, V. R. Integral membrane protein biosynthesis: why topology is hard to predict. *J. Cell Sci.* **115**, 2003-2009 (2002).

Enomoto, A., Kimura, H., Chairoungdua, A., Shigeta, Y., Jutabha, P., Cha, S. H., Hosoyamada, M., Takeda, M., Sekine, T., Igarashi, T., Matsuo, H., Kikuchi, Y., Oda, T., Ichida, K., Hosoya, T., Shimokata, K., Niwa, T., Kanai, Y., Endou, H. Molecular identification of a renal urate–anion exchanger that regulates blood urate levels. *Nature.* **417**, 447–452 (2002).

Kato, Y., Sugiura, M., Sugiura, T., Wakayama, T., Kubo, Y., Kobayashi, D., Sai, Y., Tamai, I., Iseki, S., Tsuji, A. Organic cation/carnitine transporter OCTN2 (Slc22a5) is responsible for carnitine transport across apical membranes of small intestinal epithelial cells in mouse. *Mol Pharmacol.* **70**, 829-37 (2006).

Alvadia, C. M., Sommer, T., Bjerregaard-Andersen, K., Damkier, H. H., Montrasio, M., Aalkjaer, C., Morth, J. P. The crystal structure of the regulatory domain of the human sodium-driven chloride/bicarbonate exchanger. *Sci. Rep.* **7**, 12131-12131 (2017).

Thangaratnarajah, C., Ruprecht, J. J., Kunji, E. R. Calcium-induced conformational changes of the regulatory domain of human mitochondrial aspartate/glutamate carriers. *Nat. Commun.* **5**, 5491-5491 (2014).

Harborne, S. P., Ruprecht, J. J., Kunji, E. R. Calcium-induced conformational changes in the regulatory domain of the human mitochondrial ATP-Mg/Pi carrier. *Biochim. Biophys. Acta.* **1847**, 1245-1253 (2015).

Yang, Q., Bruschiweiler, S., Chou, J. J. A Self-Sequestered Calmodulin-like Ca(2+) Sensor of Mitochondrial SCaMC Carrier and Its Implication to Ca(2+)-Dependent ATP-Mg/Pi Transport. *Structure.* **22**, 209-217 (2014).

Canul-Tec, J. C., Assal, R., Cirri, E., Legrand, P., Brier, S., Chamot-Rooke, J., Reyes, N. Structure and allosteric inhibition of excitatory amino acid transporter 1. *Nature.* **544**, 446-451 (2017).

Shnitsar, V., Li, J., Li, X., Calmettes, C., Basu, A., Casey, J. R., Moraes, T. F., Reithmeier, R. A. A substrate access tunnel in the cytosolic domain is not an essential feature of the solute carrier 4 (SLC4) family of bicarbonate transporters. *J. Biol. Chem.* **288**, 33848-33860 (2013).

Higuchi, A., Nonaka, N., Yura, K. iMusta4SLC: Database for the structural property and variations of solute carrier transporters. *Biophysics and Physicobiology.* **15**, (2018). (to be published)