



北海道公立大学法人
札幌医科大学
Sapporo Medical University

SAPPORO MEDICAL UNIVERSITY INFORMATION AND KNOWLEDGE REPOSITORY

Title 論文題目	Novel Motifs and Molecular Evolution of Leucine Rich Repeats in Proteins ① Novel leucine rich repeat domains in proteins from unicellular eukaryotes and bacteria ② Horizontal gene transfer of plant-specific leucine-rich repeats between plants and bacteria (蛋白質ロイシンリッチリピートの新規なモチーフと分子進化) 1 真核単細胞生物とバクテリアの新規なロイシンリッチリピート蛋白質 2 Plant-specific ロイシンリッチリピート遺伝子の水平進化)
Author(s) 著者	宮下, 博樹
Degree number 学位記番号	甲第 2761 号
Degree name 学位の種別	博士 (医学)
Issue Date 学位取得年月日	2014-03-31
Original Article 原著論文	1 Protein & Peptide Letters. 2014 Mar; 21(3): 292-305. 2 Natural Science. 2013; 5(5): 580-598
Doc URL	
DOI	
Resource Version	Author Edition

博士論文の要約

報告番号 甲第 2761 号 氏名 宮下博樹

Novel Motifs and Molecular Evolution of Leucine Rich Repeats in Proteins

(蛋白質ロイシンリッチリピートの新規なモチーフと分子進化)

研究目的

ロイシンリッチリピート(LRR)は、同じまたは似た配列を繰り返すタンデムリピートの一つであり、ウイルスからヒトまで2万個以上の蛋白質に存在する。多くのLRRは蛋白質やリガンドと相互作用し、動物や植物の自然免疫や神経システムの発生などの機能に関与している。また、グラム陰性菌のタイプIIIの分泌システム、アポトーシス、オートファジー、mRNAの核輸送、ユビキチンプロセスにも関与している。

LRRの繰り返し単位は20から30アミノ酸残基である。この繰り返し単位は、前半部分の保存性がよいHighly conserved segment (HCS)と、後半部分のVariable segment (VS)に分けられる。HCSはLxxLxLxx(C/N)x(x/-)Lの11または12残基からなる。VSのコンセンサス配列の違いにより、8つのクラス "Typical", "SDS22-like", "IRREKO", "Bacterial", "plant-specific (PS)", "TpLRR", "RI-like", "Cysteine-containing" が存在する(これらのモチーフを canonical LRR と名付ける)。

LRRドメインは、馬蹄形またはらせん構造をとる。HCSの3から5番目のアミノ酸配列が短いβ-ストランドを取り、隣り合う繰り返し配列とスタッキングして、馬蹄形/らせんの内側を裏打ちしている。VS部分は、α-ヘリックス、 3_{10} -ヘリックス("SDS22-like"において)、ポリプロリンIIヘリックス、β-ターン、β-ストランド、伸びた構造("IRREKO"において)の多様な二次構造をとる。また、多くのLRRドメインの両末端は、疎水性のコアを覆うキャップ構造(LRRNTおよびLRRCT)をとる。

これまで、蛋白質アミノ酸配列からLRRを検出するプログラム(PFAM, SMART, ProSite, LRR_{finder})が知られている。しかしながら、これらのプログラムでみいだせない他の多くのLRRモチーフも存在することが認識され始めている。一方、LRR蛋白質の分子進化についての多くの研究がなされ、遺伝子重複、融合、遺伝的組み換え、不等交差、遺伝子変換などの存在が報告されている。しかしながら、LRRが1つの共通祖先から発散進化してきたか、また、LRR遺伝子の水平進化が起こったかは明らかにされていない。

本研究は、これらの問題を明らかにするために、バイオインフォマティクスの手法を駆使して、アミノ酸配列データベースを用いた徹底的な配列解析を実行した。次の2つの結果を得た。1つは、単細胞真核生物と細菌由来蛋白質から、これまでLRRモチーフと同定されていない新規なLRR(Non-canonical LRRと名付けた)を見出した。Non-canonical LRRのHCSは、VxGx(L/F)x(L/C)xxNxLにより特徴づけられる。2つ目は、植物にのみ存在するPS-LRRが細菌にも存在することを見出し、このLRR遺伝子が植物から細菌に水平伝播(HGT)したことを示した。

[I] Novel leucine rich repeat domains in proteins from unicellular eukaryotes

and bacteria (真核単細胞生物とバクテリアの新規なロイシンリッチリピート蛋白質)

[I-1] 方法

新しい non-canonical LRR を含む蛋白質の同定: *Croceibacter atlanticus* の CA2559_05580 は LRR 蛋白質である。我々が開発した LRR_{pred} による解析は、LRR が 25 回繰り返していることを示した。その 10 回はこれまで LRR と同定されないモチーフ (Non-canonical LRR) であった。この HCS は LxxxLxLxxNxL と表され、canonical な HCS とは異なり 1 残基長い。我々は、この新規な LRR モチーフのみからなる LRR ドメインをもつ蛋白質が存在すると考えた。そこで non-canonical LRR 蛋白質を以下の手順で検出した。① CA2559_05580 の 10 個の non-canonical LRR をクエリーとして、FASTA による similarity 検索を実行した。② LRR_{pred} を用いて、検出した LRR 蛋白質の LRR を同定した。③ 新たに同定した LRR 蛋白質の non-canonical LRR を用いて①と②の手順を繰り返した。また、同様な方法を、*Myxococcus Xanthus* の蛋白質 [Q8VQZ6]、*Listeria monocytogenes* C5850J7 蛋白質に対して適用した。さらに、データベースのキーワード検索も実行した。

配列解析: 蛋白質の二次構造予測をプログラム SABLE と Porter を用いて行った。マルチプルシーケンスアラインメントは ClustalW を用いて行った。シグナルペプチド予測には SignalP を用いた。膜貫通領域予測には TMHMM を用いた。他のドメインは PFAM を用いて同定した。

[I-2] 結果

non-canonical LRR を含む蛋白質 : 新規な non-canonical LRR を持つ蛋白質を、30 種の真核単細胞生物と 73 種のバクテリアから 324 個同定した。Non-canonical LRR を 4 つのクラスに分類した。① Non-canonical “SDS22-like” LRR, ② Non-canonical “IRREKO” LRR, ③ Non-canonical “Vibrio-predominant” LRR, ④ Non-canonical “bacteria-specific” LRR。これらの HCS は VxGx(L/F)x(L/C)xxNx(x/-)L の 12 または 13 残基からなる。3 番目の位置に Gly が挿入されている。①と②の VS は、canonical な “SDS22-like” と “IRREKO” と同じであった。

配列解析 : 二次構造予測は、4 クラス全てで HCS の 4 から 6 番目のアミノ酸配列はβ-ストランドを示した。また、①と②は VS の C 末端側と HCS の N 末端側で、他のもう一つのβ-ストランドを予測した。さらに、多くの蛋白質に LRRNT と LRRCT が存在した。

[I-3] 考察

立体構造: 4 クラス全てにおいて、HCS の 4 から 6 番目がβ-ストランドをとる。このことは、canonical LRR において、HCS の 3 から 5 番目がβ-ストランドをとることと一致する。上皮成長因子受容体の LRR ドメインは、non-canonical モチーフを持っていると推測される。その結晶構造解析の結果は、2 つのβ-ストランドをとっている。このことは、二次構造予測の結果と一致している。①の VS は、β-ストランドと 3₁₀-ヘリックス、②の VS はβ-ストランドと伸びた構造をとるものと思われる。また、4 つのクラスで保存されている疎水性残基は、疎水性コアを形成していると考えられる。

機能: 同定した蛋白質は真菌 ECM33 蛋白質や *Monosiga brevicollis* LRR 受容体キナーゼを含むが、他の多くの蛋白質の機能は不明である。しかしながら、non-canonical LRR も、canonical LRR と同様に、さまざまなリガンドと相互作用し、免疫機能に関与しているかも知れない。

分子進化: canonical な “SDS22-like” と “IRREKO” は共通祖先から進化したことが提案されている。本研究は、39 個の蛋白質の LRR ドメインにおいて、non-canonical な “SDS22-like”

と“IRREKO”が共存することを示した。したがって、これら 4 つのクラスは 1 つの共通祖先から進化してきたと思われる。

[I-4] 結論

本研究において、我々は 324 個の蛋白質に新規な non-canonical LRR を発見した。non-canonical な“SDS22-like”と“IRREKO”は、canonical なものと同一祖先から進化したことが示唆された。Non-canonical な“Vibrio-predominant” LRR や“bacteria-specific” LRR を含む蛋白質は、コレラ等のバクテリアに特徴的にみられた。これら LRR 蛋白質は、バクテリアの生存に重要と思われる。本研究の結果はこれらの蛋白質の構造、機能に関する実験的研究を促進することが期待される。

[II] Horizontal gene transfer of plant-specific leucine-rich repeats between plants and bacteria (Plant-specific ロイシンリッチリピート遺伝子の水平進化)

[II-1] 方法

バクテリア PS-LRR のデータベース検索 : 以下の手順で検出した。① 植物蛋白質 cf-2 の PS-LRR ドメインのアミノ酸配列をクエリーに用いて、プログラム FASTA による similarity 検索を行い、バクテリア PS-LRR を検索した。② LRR_{pred} により同定したバクテリア LRR の塩基配列をクエリーに用いて FASTA による similarity 検索を行った。

配列解析 : LRR のコンセンサス配列を WebLogo を用いて決定した。他のドメインは PFAM と SMART を用いて同定した。また、PS-LRR を持つバクテリアゲノムの塩基配列を用いたプログラム EGID 解析により、Genomic Islands を検出した。

[II-2] 結果

Synechococcus sp. や *Beggiatoa sp. PS.* など 11 種のバクテリアから 20 個の PS-LRR 蛋白質を同定した。また、これらと significant similarity (E-values < 10⁻¹⁰) を持つ 83 個の真核生物の蛋白質(植物 82 個、珪藻 1 個) を同定した。バクテリア PS-LRR のコンセンサス配列は、他のクラスのどれよりも植物 PS-LRR と似ていた。BACCOP_00862 を含むバクテリア蛋白質は、LRR 以外に他のドメイン(AhpC-TSA, IgA peptidase M64, Ig ドメインなど)を持っていた。プログラム EGID は、*Synechococcus sp. CYA_1022* が他の生物種から水平伝播したことを予測した。以下に結果をまとめる。

- ① PS-LRR 蛋白質は植物には広く分布しているが、バクテリアでは 11 種からしか見つからなかった。古細菌には PS-LRR 蛋白質は存在しなかった。
- ② バクテリア PS-LRR のコンセンサス配列は、明らかに PS-LRR であることを示す。
- ③ バクテリア PS-LRR が 83 個の真核生物由来蛋白質と非常に強い significant similarity を示す。また、いくつかのバクテリア PS-LRR は、他のバクテリアのそれよりもむしろ真核生物の蛋白質に似ている。
- ④ CYA_1022 は、他の生物種から水平伝播したことが予測される。
- ⑤ ゲノム配列が解読された 31 種のシアノバクテリアのうち、2 種のみが PS-LRR を持つ。
- ⑥ バクテリア PS-LRR 蛋白質に存在する AhpC-TSA, IgA peptidase M64, Ig ドメインは、バクテリア PS-LRR とは異なり、真核生物蛋白質と significant similarity を示さない。
- ⑦ PS-LRR 蛋白質を持つバクテリアのほとんどは、水環境に生息している。また、*Beggiatoa*

は、コメの根圏に生息している。

[II-3] 考察

植物と細菌の間での HGT : 結果①から③は、細菌と植物の間で HGT が起こったことを示す。また、結果④と⑤から、HGT に関与した細菌は *Synechococcus sp.* である可能性が高い。さらに、結果⑥と⑦も HGT を支持する。ここで対立仮説を考える。植物と細菌の PS-LRR は、1つの共通祖先から発散進化した。また、限られた細菌に PS-LRR が存在することから、遺伝子の喪失が起こったとする。この仮説は、この共通祖先遺伝子が、少なくとも真核生物と原核生物の分岐の前、三十数億年前に存在したことを意味する。一方、ロイシンは6つのコドンを持っている。もし、三十数億年前に祖先遺伝子が存在したのであれば、その間、非同義置換が多数回起こったはずである。しかしながら、これは結果③と矛盾する。収斂進化の対立仮説も考えることができる。この仮説も結果③を説明しない。

植物から細菌の HGT : HGT の起こった向きは二つの可能性が考えられる。1つ目は、細菌から植物への水平伝播である。知られている全ての植物は PS-LRR をもつ。従って、この場合、PS-LRR 遺伝子は全ての祖先植物に水平伝播したと考えられる。しかし、この HGT は、少なくとも一億五千万年前に起こったことになり、結果③を説明しない。2つ目は、植物からいくつかの細菌に、最近 HGT が起こったという仮説である。この仮説は結果①から⑦の結果と矛盾しない。

細菌間での HGT : 生物種の属をまたいだ細菌からの PS-LRR 蛋白質の間で、塩基配列が非常に近い組み合わせが存在した。これは11種の細菌間でも HGT が起こったことを示唆する。

[II-4] 結論

細菌に存在する PS-LRR 蛋白質は、植物からの HGT に由来することを示した。最も parsimonious なシナリオは、植物と細菌の間の PS-LRR 遺伝子の HGT と、それに引き続き細菌間の HGT の少なくとも2回以上の HGT である。

まとめ

本研究は、次の2つのことを明らかにした。1つ目は、これまで LRR モチーフと同定されていなかった新規な Non-canonical LRR を、単細胞真核生物と細菌由来の324個の蛋白質に見出した。Non-canonical LRR は4つのクラスに分類され、non-canonical な“SDS22-like”と“IRREKO”は、canonical なものと同一祖先から進化してきたことが示唆された。2つ目は、植物にのみ存在する PS-LRR が細菌にも存在することを見出し、この PS-LRR 遺伝子が植物から細菌に水平伝播したことを示した。

これらの研究成果は、多様な LRR がどのように生まれ、また、さまざまな生物にどのように拡散していったかを示すメカニズムを明らかにしたと思われる。