

## 霊長類ゲノムの多様性と反復配列進化

●植田 信太郎

東京大学大学院理学系研究科

### <研究の目的と進め方>

ヒトの進化をもたらした遺伝的背景を探る研究は1960年代より進められてきたが、ヒトとヒトに最も近縁なチンパンジーの間の塩基配列での違いは1%余りに過ぎず、逆にその謎は深まるばかりである。そこで本研究では、上述の1%という値を求めた「ゲノム中の単一コピー配列」ではなく、ヒトへの進化の背景となったヒトと他の霊長類の間でのゲノム情報の違いを、反復配列に着目して比較解析をおこなう。このために、(1)全ゲノム配列を対象としたドットプロット法を用いてゲノム中で直列型の構造をとる反復配列を検出する。これによりゲノム中に存在する「長い配列を単位とする反復配列」と「多くの繰り返しからなる反復配列」のすべてを抽出する。これらの領域についてヒトとチンパンジーの間等での比較により「ヒトに特異的な反復配列領域」を抽出する。(2)アミノ酸をコードする遺伝子において、ヒトで単一のアミノ酸リピード構造をもつ遺伝子に着目して、原猿から新世界ザル・旧世界ザル・類人猿に至る様々な霊長類におけるそのorthologous 遺伝子での様態を明らかにする。同様の分析を単一のアミノ酸リピード構造をもつ様々な遺伝子についておこない、これにより、ヒトへの進化の過程における単一アミノ酸リピード構造の多様性とその進化動態を、最終的には単一のアミノ酸リピード構造をもつ遺伝子全体の動態とヒトへの進化の係わりを探る。(3)種に特異的に存在する単一アミノ酸リピード構造が蛋白ネットワークの中で果たす機能的役割ならびに表現型レベルでの役割を、*in vitro*での再構成系実験および遺伝子ノックイン・モデル動物により明らかにする。

### <2008年度の研究の当初計画>

前年度までに“霊長類進化の過程でヒトへ向かって複雑化していった変異”の解析に関し、アミノ酸をコードする遺伝子においてヒトでは単一アミノ酸反復配列をもつのに対しマウスでは相同な位置に単一アミノ酸反復配列をもたない遺伝子を抽出し、それらの反復配列領域に着目し、原猿から新世界ザル・旧世界ザル・類人猿に至る様々な霊長類のゲノム情報を明らかにしてきた。2008年度は、この生物学的意味を(1)哺乳動物と他の脊椎動物との比較、(2)霊長類と他の哺乳動物との比較、(3)高等霊長類と他の霊長類との比較、そして、(4)ヒトと類人猿との比較、すなわち生物分類段階を追った比較をおこない、進化の系譜との関連を探る。また、単一アミノ酸リピードが蛋白立体構造に及ぼす影響、とくに、“進化的影響の変化”の意味を中心に解析をおこなう。これにより、ヒトへの進化の過程で単一アミノ酸反復配列が果たした意味を考察する。一方、これら“霊長類進化の過程でヒトへ向かって複雑化していった変異”の機能的意味の理解は不可欠である。そこで、種に特異的に存在する単一アミノ酸リピード構造の個体レベルにおける機能的意味を理解するために、昨年度までにHoxD13転写因子を対象として、“哺乳類においては非常によく保存されているにもかかわらず、哺乳類以外の脊椎動物では完全に欠失している転写因子の転写活性化ドメイン上に存在する単一アミノ酸リピード構造を人為的に欠失させたノックインマウス”を作成し、たった一つの単一アミノ酸リピード構造の欠

失だけでも表現型に影響を及ぼすことを世界で初めて示した。しかし、これが普遍的な意味をもつ(他の転写因子でも同様あるいは同等以上の機能的影響を示す)のかどうかを明らかにする必要がある。本年度では、さらに別の転写因子、すなわち、「ヒトにおいて蛋白コード内に単一アミノ酸反復配列を有し、かつ、脳に特異的に発現しているBrain-2転写因子の単一アミノ酸反復配列を人為的に完全欠失させたノックインマウス」を作成し、この進化モデル改変実験動物に、生理的あるいは表現型(行動)レベルで何らかの変異が生じるのかに関しての解析をおこなう。以上の成果から、“ヒトへの進化の過程で複雑化していったと言える反復配列”が“ヒトへの進化”で果たした生物学的な役割を考察する。

### <2008年度の成果>

ゲノムワイドの網羅的解析から各生物分類群における単一アミノ酸反復配列の特徴を抽出するために、ヒトにおいて連続して7残基以上、同一のアミノ酸が並ぶ“単一アミノ酸リピード構造”を網羅的に探索した結果、678遺伝子由来の861“単一アミノ酸リピード構造”が抽出され、それらの構造的特徴として、(1)グルタミン酸、グルタミン、ロイシン、プロリン、アラニン、グリシン、セリン、それぞれのアミノ酸残基からなる単一アミノ酸リピードが多数を占めていること、(2)その多くは転写ならびに翻訳に係わっている遺伝子であること、(3)進化的視点から見ると、約85%は哺乳類の系統で単一アミノ酸リピードの長さが保存され、霊長類の系統だけで保存されていたのは約10%であること、が示された。これら“単一アミノ酸リピード構造”における“霊長類進化の過程でヒトへ向かって複雑化していった変異”の特徴を抽出するために、データベースから霊長類を含む哺乳動物におけるオルソログ遺伝子の塩基配列をダウンロードした。しかしこれだけでは原猿から新世界ザル・旧世界ザル・類人猿に至る段階的な霊長類ゲノム情報が著しく不足するため、実験的にオルソログ遺伝子の塩基配列を決定した。これにより上記“単一アミノ酸リピード構造”を、(a)哺乳動物全般において保存されている“単一アミノ酸リピード構造”、(b)霊長類の中では保存されている“単一アミノ酸リピード構造”、(c)霊長類の中で変異が見られる“単一アミノ酸リピード構造”に分類し、各生物分類段階を追った比較をおこない、“哺乳類から霊長類そしてヒトへの進化の過程で変異していった単一アミノ酸リピード”の構造的特徴の抽出を試みた。その結果、上記(a)~(c)のいずれにおいても、ヒトで見られた“単一アミノ酸リピード構造”の特徴である、(1)グルタミン酸、グルタミン、ロイシン、プロリン、アラニン、グリシン、セリン、それぞれのアミノ酸残基からなる単一アミノ酸リピードが多数を占めていること、(2)その多くは転写ならびに翻訳に係わっている遺伝子であること、は共通した特徴であることが示された。しかし、それぞれのアミノ酸残基が占める割合は各群で大きな差異が存在していた。

従来、タンパク質は立体構造を形成し、その立体構造の特徴により他の分子と相互作用することによって機能を発揮すると考えられてきた。しかし最近の研究により、ある特定の立体構造を形

成しない領域 (disorder 領域: natively unfolded と呼ばれる) が存在することが知られるようになった。また、これら disorder 領域の中には、機能発現に関与する領域があることが実験的に明らかになると共に、このような領域が高等生物で特に多く見られること、それらは転写調節に関与するタンパク質やDNA結合タンパク質中に多く存在することが示されている。そこで、disorder 領域を取りやすい“単一アミノ酸リピート構造”の有無を検索した結果、おおきく3つのアミノ酸残基グループに分けられることが明らかとなった。そこで、上記(a)~(c)群についてタンパク質中の disorder 領域の割合分布を調べたところ、各群間に違いが、とくに、「霊長類の中で変異が見られる“単一アミノ酸リピート構造”」である(c)群において disorder 領域の割合分布が他の2群と大きく異なっていることが明らかとなった。

次に、昨年度までと同様に霊長類における単一アミノ酸リピート構造の多様性とその進化動態をより詳細なデータから分析することを目的として、ヒトにおいてCAGトリプレットリピートによってコードされるQリピートをもち、その異常なリピート回数の拡張により ataxia を引き起こすSCA遺伝子について、類人猿4属5種、旧世界ザル7属13種、新世界ザル5属7種、原猿2属2種について、それらのリピート構造を明らかにした。その結果、SCA6遺伝子のグルタミンリピートは霊長類の共通祖先で既に獲得されていたこと、SCA12遺伝子のグルタミンリピートの出現は原猿と真猿類の分岐後、真猿類の共通祖先で生じていたことなどが明らかになった。

第三に、昨年度までに明らかにした「Hoxd-13 転写因子における単一アミノ酸リピート構造の人為的欠失が表現型に及ぼす影響」の研究成果を踏まえ、単一アミノ酸リピート構造の欠失が表現型に影響を及ぼすことが他の転写因子においても観察できるのかどうかを検証するために、ヒトにおいて蛋白コード内に単一アミノ酸反復配列を有し、かつ、脳に特異的に発現しているBrain-2 転写因子の単一アミノ酸反復配列のうちで哺乳動物に特徴的な複数の単一アミノ酸反復配列のみを人為的に完全欠失させたノックインマウスの作出をおこない、不致死であるノックイン・ホモ個体においてどのような表現型レベルにおける変異が現れるかを調べた。ヒトのBrain-2 転写因子の転写活性化ドメイン上に存在する単一アミノ酸リピート構造のうち、アラニン・グルタミン・プロリンそれぞれ1つずつ、計3つの単一アミノ酸リピート構造はマウスを含む哺乳類全般において非常によく保存されている。一方、これらの単一アミノ酸リピート構造は哺乳類以外の脊椎動物ではほぼ完全に欠けている。そこで、哺乳動物群に特異的に存在するBrain-2 転写因子の単一アミノ酸リピート構造の機能的意味を探るために、哺乳動物群に特異的な上記3つの単一アミノ酸リピート構造すべてを人為的に欠失させたノックインマウス”を作出した。基本的行動指標としての自発活動量を測定したところ、単一アミノ酸リピート欠失の雌ホモ個体で明期の活動量が統計学的に有意に低かった。そこで、エタノール誘発性睡眠時間の測定をおこなった結果、脳内のドーパミン量が低くなっていることが示唆された。生物群特異的な単一アミノ酸リピートの脳の発達への関与を明らかにするべく、現在、更に詳細な解析を進めている。

#### <国内外での成果の位置づけ>

ヒトに続いてヒトに最も近縁な現生霊長類であるチンパンジーの全ゲノム配列のドラフト解読も発表されるなど、世界のゲノム研究は日々邁進しているが、反復配列に着目した比較研究は、非反復配列に関する研究の場合と比較すると相対的に著しく少ない。過去にはゲノム中に散在する反復配列についての個別的研究が進んだ時期があったが、近年はすっかり隅に追いやられた感があった。しかし、再び反復配列が見直され始めている。特に、そ

の進化上の機能的意味に注目が集まっている。この機能進化の本質の解明には連続した生物種(最も近縁な生物から遠い系統関係の生物種まで一連の生物種)を用いた網羅的比較解析が不可欠であり、本研究はその意味で貴重なデータを提供するものである。

#### <達成できなかったこと、予想外の困難、その理由>

特になし

#### <今後の課題>

“霊長類進化の過程でヒトへ向かって複雑化していった変異”を明らかにするべく、アミノ酸をコードする遺伝子においてヒトで単一アミノ酸反復配列をもつ遺伝子(の単一アミノ酸リピート構造)に着目し、原猿から新世界ザル・旧世界ザル・類人猿に至る様々な霊長類のオルソログ領域のゲノム情報を明らかにしてきた。それと共に、単一アミノ酸リピート構造の多様性(リピート構造の有無、リピートの長短など)の生物学的意味を、(1)哺乳動物内、(2)霊長類と他の哺乳動物との間、(3)霊長類内、における比較、すなわち生物分類段階を追った比較をおこない、進化の系譜との関連を探ってきた。これにより、ヒトへの進化の過程で単一アミノ酸反復配列が構造的に大きく変動してきた動態ならびに単一アミノ酸リピートが蛋白立体構造に及ぼす影響を明らかにすることができたが、今後は、単一アミノ酸リピート構造がそれを包含するタンパク質全体へ及ぼす影響を明らかにし、包括的な解釈へと展開させる必要がある。

また、種に特異的に存在する単一アミノ酸リピート構造の個体レベルにおける機能的意味を理解するために、昨年度までにHoxd-13 転写因子を対象として、“哺乳類においては非常によく保存されているにもかかわらず、哺乳類以外の脊椎動物では完全に欠失している転写因子の転写活性化ドメイン上に存在する単一アミノ酸リピート構造を人為的に欠失させたノックインマウス”を作出し、たった一つの単一アミノ酸リピート構造の欠失だけでも形態的表現型に影響を及ぼすことを世界で初めて示した。今年度は新たに作出した「ヒトにおいて蛋白コード内に単一アミノ酸反復配列を有し、かつ、脳に特異的に発現しているBrain-2 転写因子の単一アミノ酸反復配列を人為的に完全欠失させたノックインマウス」の解析により、行動的表現型へも影響を及ぼすことを示した。今後は、これらの影響が生物群特異的と従来定義されてきた表現型へも影響を伴うものであるのかを明らかにしていく必要がある。以上の課題に加え、来年度は最終年度であること、既に膨大な量の解析結果が蓄積しているので、論文として成果をまとめて公表する作業にも力点を置いていく。

#### <成果公表リスト>

- 1) 論文
  1. 0807231524  
Kurosaki T., Matsuura T., Ohno K., and Ueda S.: Long-range PCR for the diagnosis of spinocerebellar ataxia type 10. Neurogenetics 9 (2), 151-152 (2008)
  - 2) データベース/ソフトウェア  
なし
  - 3) 特許など  
なし
  - 4) その他顕著なもの  
なし