

遺伝子同時転写単位同定にもとづく新規DNA修復関連タンパク質の予測

●由良 敬

日本原子力研究開発機構 システム計算科学センター シミュレーション技術開発室 量子生命情報解析チーム

＜研究の目的と進め方＞

ゲノムは放射線や化学物質により頻繁に損傷を受けている。これらの損傷の結果、ゲノム配列に変化が生じるが、生物にはこの損傷を修復する機構がある。また、DNA複製の際にも複製ミスが起こっているが、そのミスもほとんどの場合は修正されている。DNAの塩基配列をある程度維持することができなければ、ゲノムにコードされている情報が変化してしまい、その生物は死にいたる。しかし、完全に修復してしまえば、生物は進化しない。DNA損傷の修復は、DNA修復関連タンパク質によって行われていることが解明されている。放射線照射による細胞のガン化は、DNA修復関連タンパク質による修復が正常に機能しない、または間に合わないことが原因の一端となっている。さらに近年はDNAから転写されたRNAも修復の対象となっている場合があることがわかってきている。

本研究では、DNAとRNAの修復機構を解明することを大きな目標とし、*Deinococcus radiodurans*をモデル生物として研究を展開することとする。*D. radiodurans*は放射線抵抗性細菌として知られている。一般に生物が大量の放射線を受けると、DNAが損傷し、その個体の生命活動を維持することができなくなる。ある程度の線量までならば、損傷したDNAはDNA修復関連タンパク質の働きで修復される。*D. radiodurans*高放射線下でも生存できるのは、DNAの修復が効率よくおこなわれていることに起因していることがわかっている。この高効率の原因を知るために*D. radiodurans*のゲノム全塩基配列が決定されたが、効率のよいDNA修復機構は明らかにされていない。本研究では、*D. radiodurans*がもつ高修復機能がどのタンパク質に起因しているのかを明らかにすることにより、生物がもつDNAとRNAの修復機構の解明を目指す。

本研究を進めるにあたり、以下の3項目を明らかにすることで、*D. radiodurans*の高修復能を担う遺伝子の解明を目指す。

1) ゲノム全塩基配列比較にもとづく修復関連タンパク質の比較と分類：生命の基本である遺伝情報の維持を担う修復関連タンパク質は、DNAポリメラーゼやRNAポリメラーゼのように、すべての生物に共通に存在するのだろうか。あるいは、生物ごとに異なる種類のDNA修復関連タンパク質が存在するのであろうか。生物は進化の過程で何種類の修復関連タンパク質を生み出してきたのであろうか。現在、ヒトをはじめとする約200種を越える生物において、そのゲノム塩基配列が判明し、バクテリアにおいてはコードされている全タンパク質のアミノ酸配列も判明している。特にゲノム塩基配列が判明した*D. radiodurans*は修復関連遺伝子の宝庫であると考えられる。本研究は、ゲノム塩基配列情報を用いて、各生物が持つDNA修復関連タンパク質を整理分類し、生物種によってDNA修復メカニズムがどのように異なっているかを俯瞰的に調べることで明らかにすることを目的とする。

2) 進化情報を用いた修復関連タンパク質複合体の相互作用部位予測：ゲノムを修復するDNA修復関連タンパク質は複合体を形成していることが、生化学的な解析および立体構造解析でわかってきた。ところが、各修復システムの全立体構造が判明しているのはまれで、単体の構造もしくはドメインの構造のみが判明している場合が多い。そこでDNA修復関連タンパク質において、立体構造が判明している各サブユニットまたはドメインのどの部分が複合体形成に関与するかを、進化的情報から抽出する方法の開発と適用を目的として本研究を展開する。

3) 遺伝子同時転写単位同定にもとづく新規修復関連タンパク質の予測：従来のホモロジー検索の方法では、*D. radiodurans*のDNA高修復効率に関与する遺伝子を見いだすことができていない。ホモロジー検索の方法は、ゲノム塩基配列の全情報を利用していない。ホモロジー検索で利用している情報はタンパク質をコードしている部分のみであり、ゲノム塩基配列情報を完全に利用しているとはいえない。そこで、ゲノムの構造からDNA修復関連タンパク質と同時に機能するタンパク質をコードする遺伝子の予測を試みる。バクテリアの場合には、遺伝子がオペロン構造を構築し、関連する機能を持つ遺伝子が同時に転写発現される場合があることが知られている。そこで、本研究では、ゲノムの構造からわかる同時転写単位に関する情報を利用して、タンパク質の機能を予測する方法を開発し、*D. radiodurans*の放射線耐性の原因遺伝子を推定する。

＜研究開始時の研究計画＞

1) ゲノム全塩基配列比較にもとづく修復関連タンパク質の比較と分類：修復酵素及び修復に関連するタンパク質の生物学的な情報を収集し、データベースを構築する。データベースをもとにして、各種の修復関連タンパク質の各生物における有無を調べ、修復機構の分子進化を明らかにする。

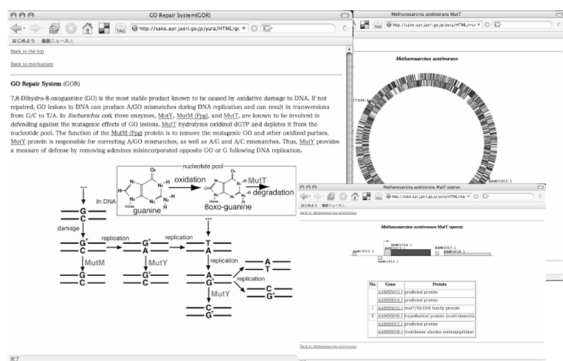
2) 進化情報を用いた修復関連タンパク質複合体の相互作用部位予測：立体構造が判明しており、他のサブユニットと複合体を形成する修復関連タンパク質を選別し、データベース化するとともに、DNA修復関連タンパク質において、マルチプルアラインメントを作成し系統樹を作成する。アミノ酸配列の違いによって有意に分かれるクラスターを検出し、アミノ酸残基の保存部位の違いを同定する。各クラスターに固有の保存部位を、タンパク質の立体構造にマップし、保存部位の空間分布を調べる。さらにDNAやRNAと相互作用する部位を形成するアミノ酸残基をタンパク質立体構造から抽出し、進化的及び物理的な特性を抽出する。

3) 遺伝子同時転写単位同定にもとづく新規修復関連タンパク質の予測：全ゲノム塩基配列を取得し、Open Reading Frame (ORF)を同定する。ORF間のスペー

サ領域の塩基数を測定して、オペロンを形成することがわかっている既知の遺伝子群におけるスペーサ領域の塩基数をもとに、オペロンの推定をおこなう。D. radioduransのゲノムからDNA修復関連タンパク質を見だし、これらのタンパク質と同時に転写される可能性がある遺伝子群をみつける。これらの中に、DNA修復に関連するとすでに判明しているタンパク質がどの程度あるかを調べることで、予測の精度を推定する。DNA修復関連タンパク質と同時に転写される可能性がある機能未知タンパク質を見だし、DNA修復システムとの関連を検討する。新規に予測されたDNA修復関連タンパク質が本当にDNA修復機構と関連があるかは、日本原子力研究開発機構高崎研究所などに実証実験を依頼して確かめる。

〈研究期間の成果〉

1) ゲノム全塩基配列比較にもとづく修復関連タンパク質の比較と分類：DNA修復酵素および修復に関連するタンパク質のアミノ酸配列、ゲノム上の位置、及び相互作用の情報を収集することができ、それらを統合したデータベースを構築することができた。DNA修復酵素および修復に関連するタンパク質のデータベースは、現在世界のどこからも公開されていない。DNA修復のメカニズムは大変に複雑であり、専門の生物学者との密な協力が必要であるために、簡単に構築することはできないためであると想像される。



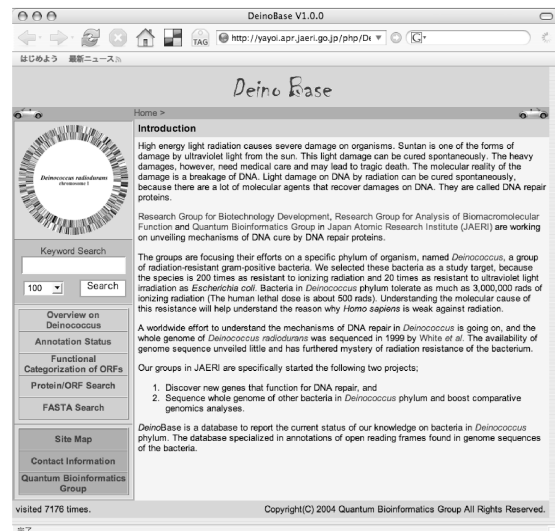
研究の基礎として用いている DNA 修復関連タンパク質のデータベース

DNA修復関連タンパク質のアミノ酸配列、ゲノム上の位置、及び相互作用の情報を収集した結果、現在133種類のDNA修復関連タンパク質が知られていることがわかった。さらに、データベース構築にあたり、各生物種において、各種の修復関連タンパク質の存在の有無を調べたところ、多くの修復関連タンパク質が、保存されていないことが判明した。たとえば、ミスマッチ修復システムであるMutS,MutL,MutMは3つのタンパク質がそろって、初めて機能するはずであるが、生物によっては、3つが完全にそろっていない、または1つが重複している場合があることがわかった。

2) 進化情報を用いた修復関連タンパク質複合体の相互作用部位予測：現在判明している133種類のDNA修復関連タンパク質のうち、82種類については少なくともドメインの立体構造がホモロジーモデリングによって構築可能であることがわかった(4)。これらの修復関連タンパク質はグリコシダーゼ以外のほとんどの場合、タンパク質複合体で機能していることがわかった。アミノ酸配列のマルチプルアラインメントにもとづき、DNA修復関連タンパク質の相互作用

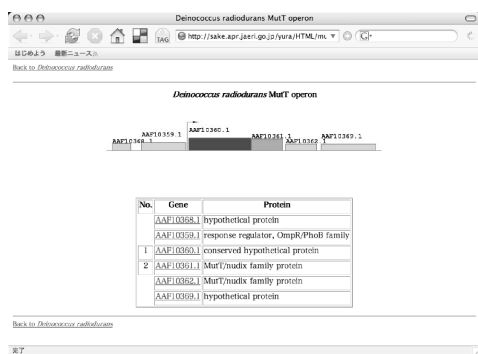
面を調べたところ、RuvAのRuvBとの相互作用面を正しく推定することができた。また塩基除去修復酵素Aagには、他の修復関連タンパク質と相互作用しそうな表面が存在しないことが推定され、Aagが単体で損傷DNAに作用することが推定された。さらに、MutSホモログにおいて、DNA修復に関与することがわかっているホモログにのみ保存されている残基を立体構造にマップしたところ、実験的にMutSのタンパク質複合体相互作用部位と推定されている部位とほぼ一致した。タンパク質間相互作用表面部位を推定する別の手法として、タンパク質立体構造のデータベースを用いて、20種のアミノ酸残基がタンパク質の内部に埋まる傾向を調べた。その結果、従来の疎水性指標とは異なる傾向にあることがわかった。この結果を基にしてアミノ酸残基単位の指標を作成することができた。この指標は修復関連タンパク質に限らず他のタンパク質の複合体形成面推定にも活用できることがわかった(2)。さらに核酸高分子との相互作用面を推定するために、核酸とアミノ酸残基との相互作用をデータベースより見だし、統計処理を行うことで、相互作用ポテンシャルを構築することができた。このポテンシャルを用いると、修復関連タンパク質などが核酸とどの部位で相互作用するかが、タンパク質の立体構造から推定できることがわかった(1)。

3) 遺伝子同時転写単位同定にもとづく新規修復関連タンパク質の予測：D.radioduransのゲノム塩基配列から新規修復関連タンパク質を予測するにあたって、まずD.radioduransのゲノムから遺伝子を推定し、可能な限りのアノテーションを行った。その結果をDeinoBaseとして公開した(6)。



DeinoBase(D.radioduransのゲノムアノテーション)

上記データを用いて、D. radioduransのゲノムから遺伝子の同時転写単位として、Open Reading Frame (ORF)間スペーサー領域の塩基数をもとに、オペロン構造の推定をおこなうことができた。推定オペロンと上記のDNA修復関連タンパク質のデータベースを組み合わせることで、D. radioduransのゲノムからDNA修復関連タンパク質とともにオペロンを形成する機能未知遺伝子を見出すことができた。これら遺伝子はDNA修復関連タンパク質と同時に転写される可能性がある。



ゲノム塩基配列からの DNA 修復関連オペロンの推定例

これら遺伝子のうち、約60%が新規のDNA修復関連タンパク質であることを理論的にみつめることができ、実証実験を日本原子力研究開発機構高崎研究所との共同研究でおこなった。その結果、これらの機能未知タンパク質をコードする遺伝子のうち、少なくとも2つはDNA修復関連タンパク質をコードしていることが実証できた。ひとつはDNA架橋剤による損傷修復に関与する遺伝子であり (Kiuchi, et al., J. Biochem. (Tokyo), 投稿準備中)、金属イオンとの相互作用が推定されている。銅イオンや亜鉛イオンとの相互作用部位には、アミノ酸残基のモチーフが存在することが知られており(4,5)、この遺伝子にコードされているタンパク質には、亜鉛イオンを配位するモチーフが存在した。もう一つは γ 線損傷の修復に関連する遺伝子であった。この遺伝子の立体構造を推定したところ、アセチル基転移酵素と類似の立体構造を持っていることが推定できた (Yura, Narumi, et al., Mol. Microbiol.投稿準備中)。

〈国内外での成果の位置づけ〉

DNA修復関連タンパク質とその関連分野において、ゲノム塩基配列を中心とした網羅的な研究は国内及び国外で発表されている例はすくない。多くの研究者は、あるひとつのDNA修復タンパク質に注目した研究を展開している。ゲノム配列を利用してDNA修復関連タンパク質を俯瞰する試みは、独自の研究である。さらに、ある特定のタンパク質におけるドメイン接触部位や複合体形成部位を予測する研究は行われているが、DNA修復関連タンパク質に予測方法を適用した例は少ない。これらの基礎的な解析をもとにして、D. radioduransにおいてDNA修復に関与するタンパク質を新規に2個同定することに成功した。本研究の成果は、まだ広く公表をしていないためにどのような反響を得るかはわからない。しかし、DNA修復関連タンパク質は生物進化及びDNAの複製という、非常に重要な部分にかかわるタンパク質なので、本研究の公表によりある程度の反響が期待される。

〈達成できなかったこと、予想外の困難、その理由〉

DNA修復関連タンパク質の情報収集が、予想より時間を要したために、データベース公開がまだできていない。これは、DNA修復関連タンパク質が想像以上に多種多様にわたっていたためである。しかしこのことは、DNA修復関連機構を網羅して理解している研究者がいないことをも意味しており、本研究によるデータベースの公開が意義深いことを示唆している。

本研究の大きな目標であったD. radioduransのDNA修復機構理解の前進が達成されたことは大きな意味がある。特にバイオインフォマティクス研究と分子生物学研究が

融合できたことが大きい。本共同研究によるD. radioduransのDNA修復機構解明研究はこれからも継続する。本ゲノム特定領域の期間中に、データベース及び論文の公式発表が実現できなかったのは残念であるが、本研究を進める人材不足に起因するところが大きい。

〈今後の課題〉

D. radioduransがなぜ放射線に強いのかは、まだ解明されていない。本研究は解明の第1歩に過ぎない。これからも遺伝子の探求と、相互作用の推定、及び実証実験による機能解明を続け、その成果をDNA修復関連タンパク質の総合的データベースとして成果を広く公表する。

〈研究期間の全成果公表リスト〉

1) 論文

1. Kim, T.P.O., Yura, K., Go, N., and Harumoto, T., Newly sequenced eRF1s from ciliates: the diversity of stop codon usage and the molecular surfaces that are important for stop codon interactions. *Gene*, 346, 277-286, 2005.
2. Kim, T.P.O., Yura, K., Go, N. and Harumoto, T., Divergent Actins from Karyorelictean, Heterotrich and Litostome Ciliates, *J. Eukaryotic Microbiology* in press (2004).
3. 403290830:Yamaguchi, A., Iida, K., Matsui, N., Tomoda, S., Yura, K., and Go, M.: Het-PDB Navi. : A database for protein-small molecule interactions. *J. Biochem (Tokyo)* 135 (1), 79-84 (2004).
4. 0304282223:Yamaguchi, A., Iwadate, M., Suzuki, E., Yura, K., Kawakita, S., Umeyama, H., and Go, M.: Enlarged FAMSBASE: 3D protein structure models of genome sequences for 41 species. *Nucleic Acids Research*, 31(1), 463-468 (2003).
5. 403290840:Nakamura, K., Kawabata, T., Yura, K. and Go, N., Novel types of two-domain multi-copper oxidases: Possible missing links in the evolution, *FEBS Letters*, 553 (2) 239-244 (2003).

2) データベース

6. DeinoBase: D. radioduransのゲノムアノテーションデータベース <http://yayoi.apr.jaeri.go.jp/php/DeinoBase/index.php>