

まえがき

本書はタンパク質を題材にしたホモロジー群とパーシステントホモロジー群の入門書である。

与えられた幾何学的対象を解析する際、その目的に応じたいくつかの特徴付けをもとに、細部を調べたり他の対象との比較を行う。図形の面積や体積といったものや、曲率などはその例である。本書で考察するホモロジー群やパーシステントホモロジー群は、「穴」に着目した特徴付けを行う道具である。

タンパク質科学の発達により、タンパク質の機能と立体構造は密接に関係していることが明らかになってきており、様々な特徴付けを用いて立体構造が調べられている。立体構造の詳細を調べる際の基礎となる情報はタンパク質を構成している各原子の空間座標データであるが、現在多くのタンパク質についてこれらは調べられておりデータベースとして一般公開されている。Protein Data Bank[29, 30]はこのようなデータベースの1つであり、ここからタンパク質の各原子の空間座標データを手に入れることができる。よってタンパク質を構成している原子をファンデルワールス半径をもつ球として表すと、Protein Data Bankのデータを用いて図1のようにタンパク質（ここではヘモグロビン）を球の和集合としてモデル化することができる。本書で紹介するホモロジー群やパーシステントホモロジー群は、このようなタンパク質の幾何構造の中に存在する「穴」について調べることを可能にする。

本書の構成としては、各章の最初に話題を総括した後、最初の数節は数学的準備、後半の節でタンパク質へ応用、という形でまとめている。1章は幾何学

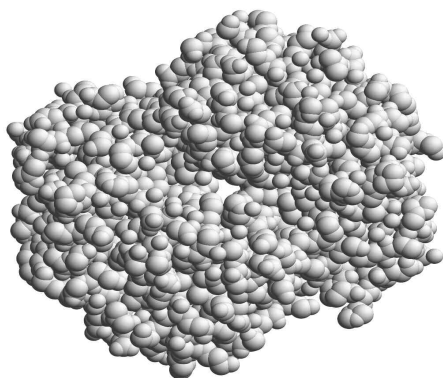


図1 ヘモグロビン (1BUW)

的な準備であり，2, 3章がホモロジー群とパーシステントホモロジー群の解説にそれぞれあてられている．

まず1章では本書を通じて用いられる幾何学的対象である単体複体について解説を行う．単体複体とは単体とよばれる三角形を抽象化したものを適当に張り合わせてできる幾何学的対象である．単体複体はトポロジーの研究に登場する代表的な幾何学的対象であり，また近年では計算機上で表現しやすいという性質から，計算機を用いたトポロジーの解析でもよく用いられている．ここでは単体複体やホモトピーといった基本的な概念の導入を行った後，図1に示したようなタンパク質の球体モデルを単体複体へ変換するいくつかの手法を紹介する．またタンパク質の立体構造データベースである Protein Data Bank の紹介を行い，ここに蓄えられているデータからタンパク質の単体複体モデルが構成できることを示す．

2章ではホモロジー群を導入しその基本的な性質を解説する．実際の応用で取り扱う単体複体は有限個の単体で構成されている場合が多く，その場合ホモロジー群は有限個の情報で記述できる．これは有限生成 \mathbb{Z} 加群の構造定理から導かれるが，本書ではここに力点を置いてホモロジー群の導入を行う．特にスミス標準形を用いて構成的に構造定理を導出することで，ホモロジー群の計算アルゴリズムの基礎となる部分も理解できるように配慮した．さらにタンパク質球体モデルのホモロジー群についての議論から，具体的な問題にホモロジー

群を応用する際の改善すべき性質を述べ、なぜパーシステントホモロジー群が必要になるかという点について説明をする。

3章ではパーシステントホモロジー群を扱う。2章と同じく、最初に代数的に必要な事柄をまとめている。特にパーシステント区間やパーシステント図といった概念は2章と同様の構造定理がもとになるが、ここでもスミス標準形を用いた構成的な導入を行い、計算アルゴリズムの基礎を理解できるようにしている。また2章と並行させて議論を展開しているのも、ホモロジー群と比較することで両者の理解が深まると思われる。後半ではタンパク質の球体モデルの半径の増大列が定めるフィルトレーションに対してパーシステントホモロジー群を応用する。具体的な応用例としては、タンパク質の圧縮率の問題とタンパク質の幾何学的な分類問題に対してパーシステントホモロジー群を応用している。

本書は、生物学や工学などの非数学系研究者や企業研究者に、ホモロジー群やパーシステントホモロジー群を解説してきた経験をもとに作成している。本書を執筆する動機の1つに、非数学者が1冊でホモロジー群やパーシステントホモロジー群を勉強できるようなテキストが欲しい、という要望に応えなかったことがある。結果的に本書がそのような読者のための一冊になれているかどうかは多いに不安が残るところではあるが、非数学者向けのユーザーガイドとしての側面にも配慮を行ったつもりである。特に2, 3章ともに代数の予備知識はほぼ仮定せずに、初歩的な議論からすべて展開している。また2, 3章の構造定理は数学的には統一的に「単項イデアル整域上の単因子論」として片付けることが可能である。しかしこのような事情から過度の一般化は避けることにした。

本書のテーマであるパーシステントホモロジー群は、数学的には比較的新しい概念である。このパーシステントホモロジー群を1つの軸とし、画像処理、離散データ解析、センサーネットワークなどの分野への代数的トポロジーの応用を総称して、近年「応用トポロジー」という分野ができつつある。本書がこの新しい応用数学の分野を知るきっかけになれば幸いである。

本書を執筆するにあたって多くの方々のお世話になった。パーシステントホモロジー群のタンパク質構造解析への応用に関する共同研究者である泉俊輔氏

iv ま え が き

からはタンパク質科学について，Marcio Gameiro 氏からはホモロジー群の計算理論について多くのご教示をいただいた．玉木大氏，荒井迅氏，一宮尚志氏には原稿に目を通してもらい有益な助言をいただいた．國府寛司先生と京都大学の学生である川口澄恵氏，坂中大志氏，中西和音氏，浜田達也氏，三宅隼斗氏からも本書を通読して少なからぬ不備を指摘していただいた．以上の方々に，心から感謝の意を表したい．