

ライフサイエンスデータベース統合推進事業
統合化推進プログラム(統合データ解析トライアル)
研究開発課題
「高精度全電子計算に基づくレクチン-糖鎖間相互
作用解析」

研究開発終了報告書

研究開発期間：平成27年5月15日～平成28年3月31日

研究代表者：中野 祥吾
(静岡県立大学 食品栄養科学部 助教)



§ 1 研究開発の概要

本課題ではタンパク質立体構造データベース、及び糖鎖構造データベースの利用拡大を目指し、タンパク質分子ビューアーである PyMOL plug-in, PyPaics の開発を行った。PyPaics はフラグメント分子軌道法 (FMO) のソフトウェアである PAICS の実行ファイルを、PDB 及び MOL2 形式のファイルから自動的に作成する plug-in, PyPaics_input と、PAICS から算出される相互作用エネルギー (inter fragment interaction energy, IFIE) を PyMOL 上で立体構造に反映し、可視化する plug-in, PyPaics_output (ruby script 含む) で構成される。開発した PyPaics を用いてレクチン、及びレクチン-糖鎖結合構造の FMO 計算、及び PyMOL 上での IFIE の可視化を行うことに成功した。加えてレクチン以外の生体高分子についても開発したツールを適用可能であることを確かめた。開発したツールにより、データベースに登録されている生体高分子の立体構造情報と、FMO 法で第一原理計算により算出した相互作用エネルギーを、PyMOL を介して相互リンクさせることが可能となった。

§ 2 研究開発のねらい

タンパク質構造データベースには 10 万個を超える立体構造データが登録されている。我々は Web を介してデータベースにアクセスし、これらの構造を自由に利用できる。タンパク質立体構造を適切な分子ビューアーで表示し解析することで、タンパク質とリガンド間の相互作用を調べることができる。この相互作用情報を基に、リガンドやタンパク質側の化学構造を変化させることで、薬剤設計や分子プローブの合理的開発が可能となる。一方でこれらの相互作用の評価、例えばタンパク質-リガンド間で形成されている水素結合の評価は、各研究者の主観により判断されることが多い。相互作用を定量的に評価する指標となる相互作用エネルギーを、タンパク質立体構造データ (PDB) と相互にリンクさせることができれば将来、より高精度に分子設計を行うことが可能になると期待される。

相互作用エネルギーについては、フラグメント分子軌道法 (FMO) に代表される第一原理計算により、高い精度で算出可能である。FMO 法を用いて、タンパク質内のアミノ酸残基間の相互作用や、タンパク質内の各アミノ酸とリガンド間の相互作用を、inter fragment interaction energy (IFIE) というパラメータで定量的に示すことができる。これは FMO 法を用いて算出できる特徴的なパラメータで、その他の第一原理計算では算出が難しい。IFIE の値からタンパク質の熱力学的な安定性や、タンパク質-リガンド間の相互作用を、各アミノ酸残基レベルで定量的に見積もることが可能となる。

このように FMO 法は、タンパク質立体構造データベースを扱う、様々な研究分野での応用が期待されている。しかし、FMO 法が実際に応用されている例は少ない。この原因の一つとして、算出される IFIE (2 次元のデータ) のデータとタンパク質立体構造データ (3 次元のデータ) をリンクさせるツールが開発されていないことが挙げられる。IFIE データをタンパク質立体構造データに反映し、可視化するツールを開発できれば、FMO 法を用いたタンパク質立体構造データベースの更なる利用拡大が期待できる。タンパク質分子の 3D イメージ表示には、様々な分子グラフィックスソフトが利用されている。特に PyMOL は世界的に広く利用されている。PyMOL の plug-in として、FMO 計算により算出される相互作用エネルギーの大きさを、タンパク質分子上に可視化するツールを開発できれば、タンパク質立体構造データベース (PDBj) や 糖鎖構造データベース (LfDB) の利用拡大につながることを期待される。本課題では、これらデータベースの利用拡大を目指したツール開発、PyMOL plug-in の開発を第一目標に、研究開発を進めてきた。

本研究開発では、主に糖鎖構造データベースに登録されているレクチン構造を、開発したツールの適用対象として用いた。具体的には PDBj や LfDB に登録されているレクチン、及びレクチン-糖鎖構造を用いた。レクチン中のアミノ酸残基間、及びレクチン-糖鎖間の相互作用エネルギーを FMO 法により計算、その結果を PyMOL 上でレクチン分子上に反映し、エネルギー値を可視化するツールの開発を目指した。

開発したツールを用いた、データベースの応用利用についても検討を行った。例えば、共同研究者の平林、舘野らが開発を進めているレクチンデータベースに登録されている実験結果と、開発したツールから算出される計算結果を統合することで、変異導入などによる糖鎖の結合能の变

化を予測するアルゴリズムが開発可能か検討した。

§ 3 研究開発計画

(1) 当初の研究開発計画

糖鎖およびレクチンデータベースに登録されている糖鎖、レクチンおよび糖鎖-レクチン複合体構造の第一原理計算を行う。糖鎖の動的の構造ゆらぎや、レクチンの点変異を考慮した糖鎖への結合能を理論的に予測する新規アルゴリズムを開発し、実際のレクチン-糖鎖系への適用を行う。異分野の研究者による糖鎖およびレクチンデータベースの利用拡大を目指して、可視化ソフトウェア開発を行い、将来的に相互作用解析結果を疾患・医薬品・環境物質データベースなどと相互リンクさせる足がかりとする。

研究開始時には以下の項目について、開発を行うことを、研究開発計画書の中で記述した。以下にその詳細を記述する。

[研究代表者(中野)]

主に以下に示した PyMOL プラグインの開発を進める。

1. PyPaics_input の開発：PDB や MOL 形式のファイルを PyMOL 上で PAICS の input 形式に変換するプラグインを開発する。PyMOL ライブラリである cmd に関する情報を集め、コーディングを進める。
2. PyPaics_output の開発：PAICS の出力結果を PyMOL 上でレクチン及び糖鎖構造に反映するプラグインを開発する。こちらも 1. と同様に情報収集ののち、コーディングを進める。

[共同研究者 1(常盤)]

3. 相互作用解析用 Ruby スクリプトの開発：PDB と PAICS の出力結果を入力データとし、レクチン-糖鎖間の相互作用エネルギーを抽出する Ruby script。

[共同研究者 2(館野)]

4. 作成したツールに適用するレクチン-糖鎖の構造情報を提供する。及びツールで出力した結果を評価し、ツール開発グループに情報をフィードバックする。

[研究代表者及び共同研究者 1, 2]

5. 開発したソフトウェア（プラグイン等）の公開：開発したソフトウェアの公開は、論文発表後に行う。

(2) 新たに追加・修正など変更した研究開発計画

第一原理計算を実行するにあたり、部分構造の欠損がなく、かつ高い分解能で決定されたレクチンの X 線結晶構造が必要になることが、研究を進める中で判明した。計算に適したレクチンの構造情報を得るため、レクチンデータベースに加えて蛋白質立体構造データベース (PDBj) の利用についても検討を行った。

研究を進める中で、開発したソフトウェアの有効性を確かめるため、開発したツールの適用範囲をレクチンだけでなく、その他のタンパク質構造にも広げた方が良いと考えた。特に酵素については、構造機能相関研究が広く行われており、部位特異的変異体による機能変化（例えば変異体の活性パラメータ変化など）について、大量のデータが報告されている。これらの実験結果を計算結果から説明できるか確認するためにも、酵素を計算対象に取り入れる必要があった。よって研究代表者のグループで、以下の研究を実施することにした。

- PDB に登録されている、産業応用が期待される酵素群の基質認識機構の開発ツールを用いた計算化学的解明

§ 4 研究開発成果

PDB や LfDB に登録されている構造情報から IFIE を算出し、PyMOL 上でその情報を反映するには、Fig. 1 に示したプロセスを行う必要がある。当プロジェクトでは、ユーザー自身が PDB や LfDB にアクセスし、その構造データを処理して IFIE を得ることサポートするツール群を、主に開発した。以下にその概要を記述する。

(1) 研究代表者グループ

① 研究開発題目

生体分子の FMO 計算用 PyMOL プラグイン開発

② 研究開発成果

当初の研究計画に沿って PyMOL plug-in, PyPaics_input と PyPaics_output の開発を進めてきた。この 2 つの plug-in を利用することで、タンパク質立体構造データベースや糖鎖構造データベースに登録されている構造について、IFIE を PyMOL 上で立体構造上に反映し、結果を可視化することが可能となる。以下に開発した個々の plug-in の概要と成果を記載する。なお、いずれの plug-in のインストールについて、PyMOL の機能を用いて GUI を通して実行可能である (Fig. 2)。

•PyPaics_input: 本ツールは蛋白質構造データベース (PDBj) 及び LfDB から取得した PDB, MOL2 形式の 3 次元構造データを、PyMOL 上で国産の FMO ソフトウェア、PAICS の実行ファイルへと全自動で変換するツールである (Fig. 3A)。本ツールはタンパク質分子を変換する PyPaics_input_protein, 糖鎖や FAD などのサイズの大きいリガンド分子を変換する PyPaics_input_ligand, 溶媒である水分子を変換する PyPaics_input_solvent の 3 つから構成されている。

本ツールは PyMOL で分子を表示可能なファイル形式であれば、分子種を問わず、PAICS 実行ファイルへと変換することができる (Fig. 3B)。よって本ツールの汎用性は高いと考えている。本ツールを介して、構造情報が登録されているデータベースにアクセスし、FMO 法による第一原理計算が可能となる。

•PyPaics_output (Fig. 5 ~ 7): 本ツールは PAICS で算出される IFIE を、PyMOL 上でタンパク質構造に反映し、可視化するツールである。PyPaics_output には次の 2 つの機能がある: 1. All interaction mode (Fig. 6), 2. Ligand & protein mode (Fig. 7)。All interaction mode では、タンパク質分子内の各アミノ酸残基間の IFIE 値をタンパク質構造上で可視化することができる。こちらの機能を利用することで、レクチンの各ドメイン間に存在する安定化に寄与するアミノ酸残基間の相互作用を明らかにすることができる (Fig.6)。Ligand & protein mode では、リガンドとタンパク質分子の各アミノ酸残基間の IFIE 値を可視化することができる。この機能により、レクチンの各アミノ酸残基と糖鎖間の相互作用を明らかにすることができる (Fig. 7)。

本ツールは PAICS で出力されたデータであれば、オリゴマー構造やジスルフィド結合を持つタンパク質についても、PyMOL 上で IFIE を可視化することが可能である。よって本ツールの汎用性は高い。また IFIE をタンパク質構造上に反映するツールは報告が無く、新規性も高い。

•ソフトウェアについては、以下の URL の管理下で公開している。

<http://dfns.u-shizuoka-ken.ac.jp/labs/proeng/>

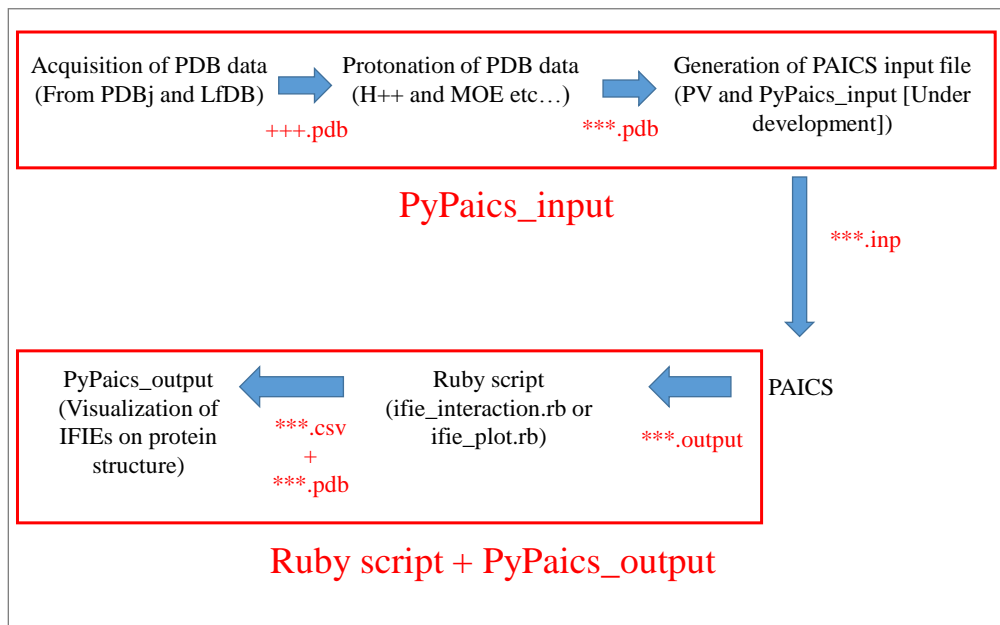


Fig. 1, Flow of how to access database and convert the structural data.

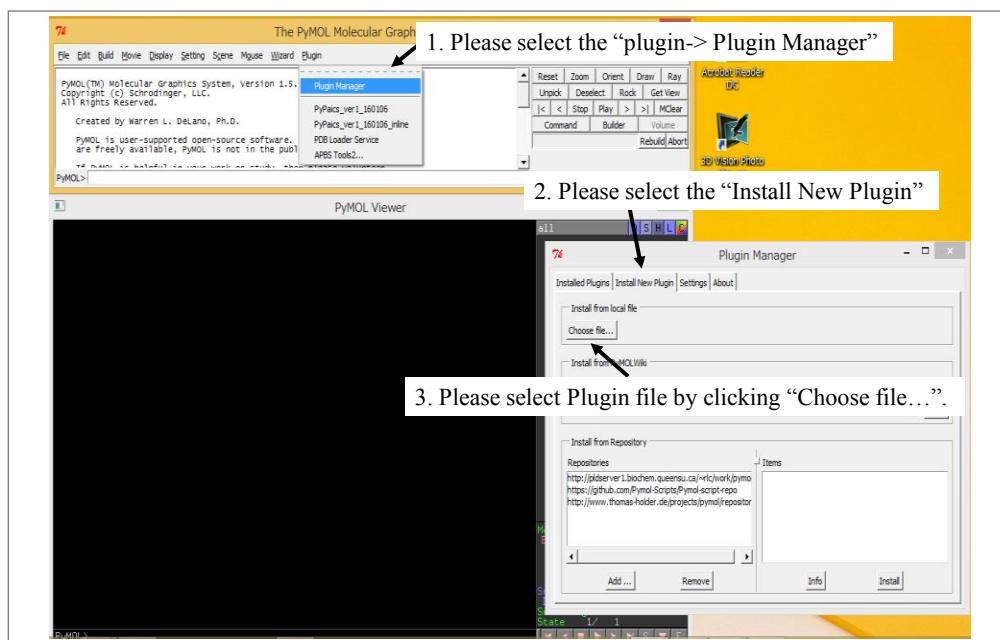


Fig. 2, Installation of PyPaics_input and PyPaics_output
 → Installation of the plugin could be completed by clicking PyMOL menu bar.

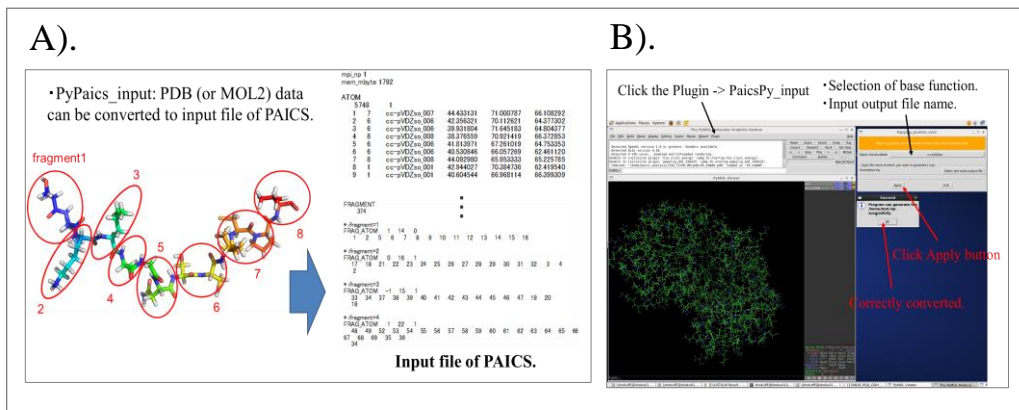


Fig. 3, A). Outline of PyPaics_input. B). GUI of PyPaics_input.

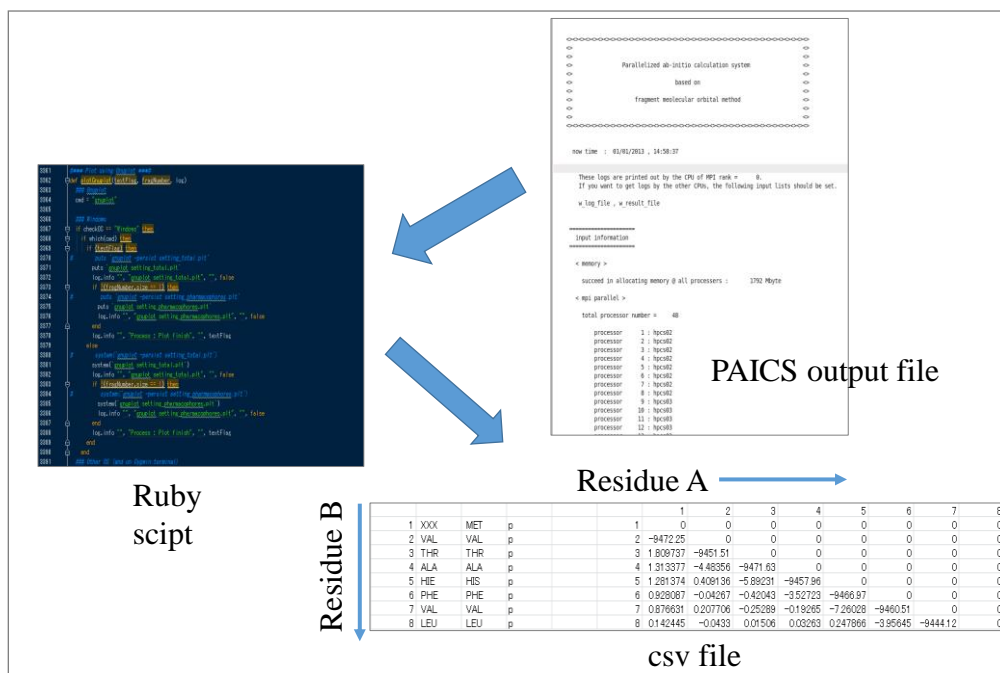


Fig. 4, Data curation from PAICS output to csv file by ruby script.

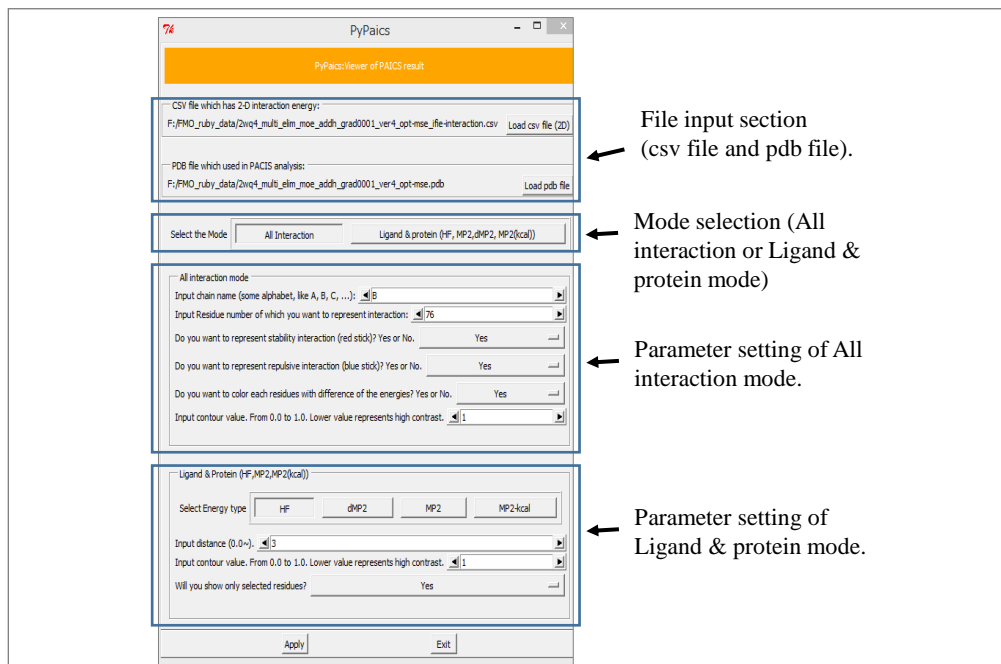


Fig. 5, Control window of PyPaics_output

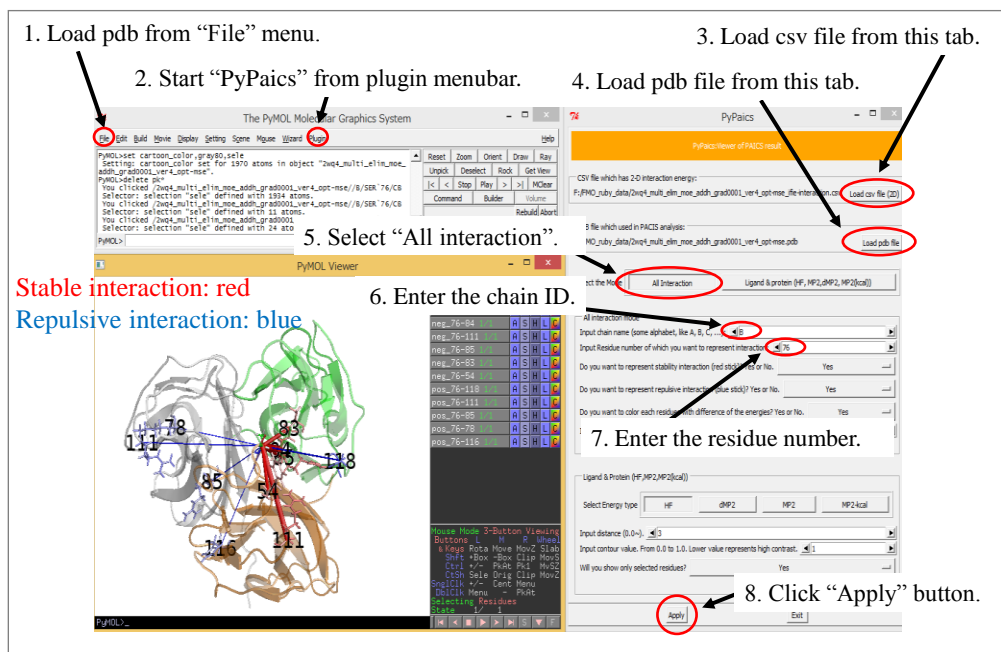


Fig. 6, How to use All interaction mode of PyPaics_output

1. Load pdb from "File" menu. 2. Start "PyPaics" from plugin menubar. 3. Load csv file from this tab. 4. Load pdb file from this tab.

5. Select "Ligand & protein".

6. Select energy type.

7. Enter distance.

8. Click "Apply" button.

Stable interaction: red
Repulsive interaction: blue

Fig. 7, How to use Ligand & protein mode of PyPaics_output

(2) 共同研究者グループ1 (常盤グループ)

① 研究開発題目

生体分子の FMO 計算用 Ruby スクリプトの開発

② 研究開発成果

当初の研究計画・目標に従い、生体分子の FMO 計算用 Ruby スクリプト(ife_plot.rb と ifie_interaction.rb)の開発を進めてきた。この 2 つの Ruby スクリプトは、FMO 法のソフトウェアである PAICS によって計算された IFIE を解析し、データベース化を見据えた Excel 形式の CSV ファイルに結果を出力する。さらに、タンパク質立体構造データベースである PDB ファイルを使用することによりアミノ酸毎に IFIE との関連付けを行い、ターゲットと相互作用が大きなアミノ酸に関する IFIE を棒グラフに自動的に出力することが可能である。以下に本研究で開発した 2 つの Ruby スクリプトの概要を記載する。

•ife_plot.rb:本スクリプトは、純国産 FMO のソフトウェアである PAICS の計算結果ファイルとタンパク質構造データベース (PDBj)から取得した PDB ファイルから IFIE をタンパク質内のアミノ酸毎に解析し、1 次元表データの CSV・TXT のファイル及び Gnuplot を用いて相互作用エネルギーを棒グラフに出力するスクリプトである。このスクリプトの 1 次元表の CSV 形式の出力ファイルは研究代表グループのツールである PyPaics_output の「Ligand & protein mode」に対応し、リガンドタンパク質分子間の IFIE 値をアミノ酸残基毎に 1 次元表データとして出力される。このことから、ターゲットのリガンドがタンパク質分子内のアミノ酸残基のどの部分と相互作用するかを解明することが可能となる。

•ife_interaction.rb:本スクリプトは、純国産 FMO のソフトウェアである PAICS の計算結果ファイルとタンパク質構造データベース (PDBj)から取得した PDB ファイルから IFIE をタンパク質内のアミノ酸毎に ife_plot.rb と同様に解析し、2 次元表データの CSV・TXT のファイルを出力するスクリプトである。このスクリプトの 2 次元表の CSV 形式の出力ファイルは研究代表グループのツールである PyPaics_output の「All interaction mode」に対応し、タンパク質分子内の各アミノ酸残基間同士の IFIE 値をアミノ酸残基毎に 2 次元表データとして出力される。このことから、タンパク質を構成するアミノ酸残基間の相互作用を明らかにできるため、タンパク質分子の安定化に寄与するアミノ酸残基を直接解明することが可能となる。

また、2 つの本スクリプトは、計算に用いた PDB ファイルと FMO の計算結果のファイルがあれば、ターゲットの分子種やタンパク質の構造を問わず、解析することができることから汎用性が高い。IFIE をタンパク質内のアミノ酸毎に解析するスクリプトは報告が無く、新規性も高い。

本スクリプトは、研究代表グループと同様に以下の URL の管理下で公開している。

<http://dfns.u-shizuoka-ken.ac.jp/labs/proeng/>

(3) 共同研究者グループ 2 (鎗野グループ)

① 研究開発題目

計算用のレクチン-糖鎖構造の構造および生化学情報の提供、及び出力結果の評価

② 研究開発成果

当初の計画に従い、計算対象とするレクチンとして、*Burkholderia cenocepacia* 由来の TNF-like lectin domain (PDB ID: 2WQ4) の構造情報を提供した。本課題で作成した PyPaics と ruby script を用いてエネルギー解析を行い、ドメイン境界にて形成されている相互作用と、リガンドである糖とレクチンの相互作用について評価した。解析により明らかになった結果を以下に記述する。

[レクチンドメイン境界領域に存在する相互作用について]

2WQ4 は 3 つの chain から構成されるレクチンである。この境界部位の相互作用は、本レクチンの安定性に重要であると考えられているが、その詳細は不明である。今回、開発した PyPaics を用いて、X 線結晶構造解析の結果から安定化に寄与していると考えられている Ser76 について、相互作用解析を行った。結果、結晶構造だけでは見出すことができなかった隣接 chain に属するアミノ酸残基、Thr83 及び Arg85 が安定化に大きく寄与することが予測された。また相互作用についても、今回の解析結果から、“Gly84 > Arg111 > Arg85 > Thr83 > Phe54”の順に大きくなることを明らかにした。

[レクチン-糖鎖相互作用の定量化について]

2WQ4 には糖鎖として SFU が結合していることが、X 線結晶構造解析の結果から示されている。一方で、どの残基が糖鎖と強く相互作用を形成しているかは不明である。そこで PyPaics を用いて、SFU と相互作用を形成しているアミノ酸残基を探索した。結果、Arg111 と Arg85 が持つ、正の電荷を持ったグアニジノ基が形成する水素結合（イオン結合）が SFU との結合に大きく寄与することが判明した。また、Thr83 の側鎖のヒドロキシル基は、SFU との相互作用について、若干不利に働く、つまり不安定化に寄与することが判明した。

以上のように開発した PyPaics を用いることで、X 線結晶構造解析だけでは明らかにできない、分子間相互作用について、定量的に示すことができた。これらの情報は PDB や LfDB 上の構造データに新たな価値を付加できると考えられる。

§ 5 研究開発計画に対する達成状況と将来展望

(1) 達成状況

ツール開発については、当初の目標を達成できたと考えている。開発計画で 3 つのツールを開発することを記述したが、その内 2 つは開発を完了し、もう一つは 3 月末までに完了できる予定である。前者に関する、“第一原理計算を用いて算出した相互作用エネルギー値を PyMOL 上で可視化するツール、PyPaics_output 及び Ruby script”の開発は、当初の計画通り開発を完了した。現在はその成果を論文にまとめている。後者に関する、“レクチンを含むタンパク質構造データを FMO 解析用ソフトウェア、PAICS で実行可能なファイル形式に PyMOL 上で変換するツール、PyPaics_input”についても、オリゴマー構造への適用ができない問題点を残して、ほぼ開発を完了した。

ツール開発は計画通りに実施できたが、レクチンへ開発ツールを適用し、FMO 計算を行うことには時間がかかった。この点については中間審査の時にも指摘を受けた。計算に時間がかかった理由として、登録されているレクチン構造の PDB データ中に、計算実行上の問題点が存在したためである。例えば構造の欠損や非天然アミノ酸の導入、異方性温度因子の記述やオリゴマー構造を持つことが挙げられる。このような問題点を解決するための機能をツールに追加する、いわゆるデバッグ作業に時間を取られたため、開発に時間がかかり、計算が十分に行えなかった。しかし本研究の最終段階でこれらの問題点を解決でき、*Burkholderia cenocepacia* 由来の TNF-like lectin domain (PDB ID: 2WQ4) について、開発ツールを用いて FMO 解析と IFIE の立体構造への可視化を達成することができた。また産業利用が期待されている酵素についても開発ツールを適用し、新たな知見を得ることができた。計算例を増やし、計算結果とタンパク質立体構造データベース及びレクチンデータベースの情報を統合するために、現在も計算解析を進めている。

(2) ツール等の将来展望

本研究で開発した PyPaics は、PyMOL で表示可能な、レクチンを含むタンパク質の構造

に広く適用可能であり、PyMOL 上で立体構造に相互作用エネルギー (IFIE) という情報を付加することができる。将来、計算機の演算速度が上昇することにより、タンパク質立体構造データベース及び糖鎖構造データベースに登録されている個々のデータに関して、FMO 計算を行い、IFIE を算出することが可能になるかもしれない。これにより計算された IFIE を一つのデータとして、既存の立体構造情報を扱うデータベースに追加することが可能となる。薬剤の合理的設計やタンパク質設計を行う上で重要だが、現在は定性的に評価されているタンパク質-リガンド間およびタンパク質内の相互作用について、定量的に評価する時代が到来するはずである。IFIE を算出しデータ化するための、かつ IFIE と立体構造情報データベースをリンクするためのツールとして、PyPaics は基礎から応用分野の研究まで、広く貢献できると考えている。

§ 6 研究参加者

①研究代表者グループ

氏名	所属	役職	研究開発項目	参加時期
中野 祥吾	静岡県立大学	助教	生体分子の FMO 計算用 PyMOL プラグイン開発	H27.5-H28.3

②共同研究者グループ1(常盤グループ)

氏名	所属	役職	研究開発項目	参加時期
常盤 恭樹	東北大学	博士課程後期 1 年	生体分子の FMO 計算用 Ruby スクリプトの開発	H27.5-H28.3

③共同研究者グループ2(舘野グループ)

氏名	所属	役職	研究開発項目	参加時期
舘野 浩章	産業技術総合研究所	主任研究員	計算用のレクチン-糖鎖構造の構造および生化学情報の提供、及び出力結果の評価	H27.5-H28.3
平林 淳	産業技術総合研究所	首席研究員	計算用のレクチン-糖鎖構造の構造および生化学情報の提供、及び出力結果の評価	同上

§ 7 成果発表等

(1)原著論文発表 (国内(和文)誌 1 件、国際(欧文)誌 1 件)

- Shogo Nakano *et al.*, Structural and computational analysis of peptide recognition mechanism of class-C type penicillin binding protein, alkaline D-peptidase from *Bacillus cereus* DF4-B., *Scientific reports*, **5**, 13836, DOI: 10.1038/srep13836

(2)その他の著作物(総説、書籍など)

なし

(3)国際学会発表及び主要な国内学会発表

- 招待講演 (国内会議 0 件、国際会議 0 件)

- ② 口頭発表 (国内会議 1 件、国際会議 0 件)
1. 中野 祥吾 (静岡県立大学)、*R*-選択的な FAD 依存型アミン酸化酵素の基質認識機構解明、BMB2015、神戸、2015 年 12 月 4 日
- ③ ポスター発表 (国内会議 2 件、国際会議 0 件)
1. Shogo Nakano (Univ. of Shizuoka), Development of PyMOL plug-in, PyPaics, to evaluate protein-ligand interaction by FMO method, 第 20 回静岡健康・長寿学術フォーラム, 静岡, 2015 年 10 月 30 日
 2. 中野 祥吾 (静岡県立大学)、*R*-選択的な FAD 依存型アミン酸化酵素の基質認識機構解明、BMB2015、神戸、2015 年 12 月 4 日

(4)知財出願
該当なし

(5)受賞・報道等
該当なし

§ 8 自己評価

本課題で開発したツールにより、タンパク質内の相互作用エネルギーという定量的指標を、立体構造データに反映することが可能となった。これにより生体分子の立体構造を扱うデータベースに、新たな情報を追加する環境を整えることができたと考えている。一方で、変異導入による糖鎖結合能の変化を予測するアルゴリズム開発については、難航した。これの原因は、開発開始時に、レクチン構造の PDB に、計算実行上の問題点があることを、十分に把握していなかったためと考えている。この点について調査不足であった点は素直に反省したい。プロジェクト終了後もアルゴリズム開発を進めていく計画である。

以上