生命現象の時空間動態情報の可視化*

京田 耕司**, ホー ケネス**, 糸賀 裕弥**, 遠里 由佳子***, 大浪 修一**

Visualization of Spatiotemporal Biosystems Dynamics Information

Koji KYODA, Kenneth H.L. Ho, Hiroya ITOGA, Yukako TOHSATO and Shuichi ONAMI

1. 緒 論

生細胞イメージング技術の発達により、分子、オルガネラ、細胞などの時空間動態を顕微鏡により観察し撮影することが可能になった。これらの動態を観察することにより、生命現象の背後にある細胞・発生メカニズムの理解が飛躍的に促進している。3次元空間における生命現象の動態は、主には2次元画像がスタックされたタイムシリーズの画像群として撮影される(Fig. 1)。このような画像群のみから時空間動態を理解することは難しいため、3次元空間上に情報を再構築して可視化することが必要となる。

3次元空間上で再構築する可視化手法のひとつとして、スタックされた画像を透過させて表示する手法がある。例えば、UCSD で開発されている ChimeraX¹⁾では、読み込んだスタック画像を3次元空間上に再構築することができる(Fig. 2 a)。また、Allen Institute が公開している Cell Feature Explorer (<https://cfe.allencell.org>)では、約4万のヒト iPS 細胞のスタック画像をブラウザ上で3次元空間上に再構築して表示させることができる(Fig. 2 b)。しかしながら、細胞内の分子やオルガネラの密度が高くなると、これらの境界や関係性を理解するのが困難になる場合がある。この問題を解決するために、各分子、オルガネラや細胞などの物体の境界の決定がしばしば行われる。

生命現象の時空間動態における分子、オルガネラや細胞などの物体の位置や形状を座標情報として計測するために、バイオイメージ・インフォマティクスと呼ばれる分野が2000年代後半より急速に発展してきた。本稿では、バイオイメージ・インフォマティクス技術を用いて計測される生命現象の時空間動態に対する座標情報を記述する統一的なオープンデータフォーマットとその可視化の現状と課題について報告する。

Fig. 1 Example of 3D time-lapse microscopy images. The entire sample can be recorded by changing the focal plane.

Fig. 2 3D visualization tools for microscopy images. a) ChimeraX, b) Cell Feature Explorer.

* 原稿受付 2019 年 10 月 28 日

** 理化学研究所生命機能科学研究センター 発生動態研究チーム

*** 立命館大学 情報理工学部 情報理工学科

2. 生命現象における時空間動態の定量化

バイオイメージ・インフォマティクス技術のひとつに、

画像解析技術を用いて顕微鏡画像から生物学的な新たな知見を得るための技術がある²⁾。顕微鏡画像データから、分子やオルガネラ、細胞などの位置や形状などを座標情報として抽出する手法の開発が盛んに行われている³⁾。生物学の分野において、画像データからこのような定量データを抽出するための汎用的なソフトウェアツールとして、ImageJ⁴⁾、Fiji⁵⁾、CellProfiler⁶⁾、Segmenter⁷⁾などが開発されている。これらのソフトウェアツールの他にもさまざまなツールが開発されており、生命現象の時空間動態を計測した定量データが世界中で産出されている。

3. 定量データを記述するフォーマット

これまで生命現象における時空間動態を定量的に計測したデータは、各研究グループが独自に開発したフォーマットにより記述されてきた。これらのフォーマットで記述された定量データを利用して、新たな解析を行ったり比較を行ったりする場合、データの値が独自の方法で格納されているために値の読み出しに特殊な変換が必要であったり、データの並びに意味があるがその説明が書かれていないことがあったりした。このため、データは公開されているにも関わらず、再利用がなかなか進まないという問題があった。

この問題を解決する方法のひとつが、生命現象の時空間動態に対する定量データを統一的に記述することが可能なデータフォーマットを開発して普及させることである。定量データの可視化や解析のためのツールを開発する側は、そのフォーマットに対応するだけで全てのデータを利用できるため、ツールの開発と検証の効率化につながるという利点がある。また、データを産出する側にとっては、すべてのツールを利用することが可能となり、他の研究グループが産出したデータとの比較を行うことも容易となる利点がある。

そこで我々は、生命現象の時空間動態の定量データを記述するためのオープンフォーマットである Biological Dynamics Markup Language (BDML) を開発した⁸⁾。BDML は Extensible Markup Language (XML) を基盤としている。XML はテキストベースであるので、作成や閲覧に特定のツールを必要とせず、またデータの持つ意味がタグや属性から認識可能となっている。このため、実験生物学者からのデータへのアクセスが容易である。また、シリアルライズやデシリアルライズを含めた XML 対応の各種ライブラリ、ツール群を利用することが可能であることから、計算生物学者をはじめとするツールの開発を効率的に進めることができる。さらに、XML が持つ拡張性を継承しているため、定量データに対するメタ情報の追加、変更、さらには記述要素の追加などにも柔軟に対応できる。

我々は、BDML を開発するにあたり、今までに世界中で計測された生命動態の時空間情報において標的とし

ている分子、オルガネラ、細胞などの物体の記述方法に対する網羅的な調査を行なった。この調査の結果、生命現象における物体の記述には、主に 5 種類の要素、点、線、面、円、球が使われていることがわかった (Fig. 3 a)。例えば、一分子動態は点の動きとして記述されており、また細胞は閉曲線の集合として記述されていることが多い。そこで BDML では、〈point〉、〈line〉、〈circle〉、〈sphere〉、〈face〉のタグを準備し、タグの組み合わせも可能にすることで、既存および将来得られる生命動態の定量データの記述に対応している (Fig. 3 b)。また、これらの物体は時間経過とともにその形状や位置が変化するため、時点間で同一の物体であることを表現する時点間追跡情報の記述も必要となる。BDML では、各時点における個々の物体に対してユニークな ID を付加することが必須となっており、追跡情報は、ある時点のある ID を持つ物体が前の時点でのどの ID を持つ物体と同一であるかを記述することで表現している。このような方法により、BDML ではさまざまな生命現象の時空間動態を記述することを可能にしている (詳細は、<http://ssbd.qbic.riken.jp/bdml/>を参照のこと)。

BDML は XML を基盤としているため、高い人間・機械可読性や柔軟な拡張性を備えているが、大規模なデータの記述を行う際に問題が生じる場合がある。XML を扱う一般的なライブラリでは、巨大なファイルを扱うことを想定しておらず、解析しながらの読み込みに長い時間がかかり、処理のために大量のメモリを必要とする場合がある。BDML も例外ではない。この問題を解決するために、〈series〉というタグを準備し、ファイルを分割することによりひとつひとつのファイルの大きさを調整する仕組みを導入した (BDML v0.18 より採用)。これにより、大規模な定量データの読み込

Fig. 3 BDML data format. a) Five entities available in BDML v3.0, b) Example BDML file.

みにも対応することが可能となった。

上記の問題を解決するもうひとつの方法は、データのうち必要な部分だけを高速に読み出すことができる仕組みを提供することである。我々は、その仕組みを実現するために、Hierarchical Data Format (HDF5; <http://www.hdfgroup.org/hdf5/>)を基盤としたフォーマットである BD5 の開発を進めている。HDF5 は階層構造を持つ科学技術データを格納するのに広く利用されている汎用バイナリフォーマットである。BD5 では、BDML と同じ表現能力を確保しつつデータを階層的に HDF5 のフォーマットで格納することにより、計算コストを低減させてデータの読み出しを行うことを可能にしている(詳細は、<http://ssbd.qbic.riken.jp/bdml/> を参照のこと)。

4. 定量データの可視化

BDML で記述された定量データを可視化することにより、分子、オルガネラや細胞の位置や形状の時間変化を 3 次元空間上で観察することが可能になる。これらの情報を可視化することで、細胞生物学・発生生物学における新たな知見の取得につながることを期待される。定量データの可視化により表現型解析、すなわち生命現象における分子、オルガネラ、細胞などの形態、構造や移動などの解析を行うための必要最低限の要件は、3 次元空間において対象となる物体の位置や形状をさまざまな角度から観察できることと、その物体の位置や形状の時間変化を観察できることが挙げられる。

この目的を達成するために、我々は、BDML ファイルで記述された生命動態の時空間情報を可視化するツール BDML4DViewer を開発した(Fig. 4)。このツールは、主に実験生物学者を対象としているため、現在広く普及している ImageJ のプラグインとして開発した。ImageJ は、オープンソースの画像処理ソフトウェアで、プラグインやマクロによる高い拡張性を持っている。BDML4DViewer により、ImageJ で顕微鏡画像データとそこから画像解析により抽出された定量データを切り替えながら観察することができる。

生命動態の時空間定量データは 3 次元空間上の分子、オルガネラや細胞などの物体の位置や形状の経時的変化

を捉えたものであるため、これらの物体の 3 次元空間上への表示とそれらの変化が観察できることが必要である。BDML4DViewer では、3 次元空間における物体の可視化については Java OpenGL (JOGL)を用いて実現し、時点(時間)の移動についてはキーボードの矢印での操作により実現している。マウス操作による視点の移動を行うことにより、あらゆる角度から生命現象の時空間動態を確認することができるため、スタックされた画像データのみからは困難であったさまざまな角度からその動態を観察することが可能になる。今まで観察することが困難であったことを見るのが可能となるため、新しい細胞・発生生物学機構の発見につながることを期待できる。

また、ウェブブラウザ上で定量データを表示させる試みも行われている。SSBD:database (<http://ssbd.qbic.riken.jp/>)⁹⁾では、定量データをウェブブラウザで表示することができる。前途のとおり巨大な BDML ファイルは扱いにくく、小さいサイズで扱うためにはデータの分割が必要である。そこで、現在の SSBD:database では、時点ごとにデータを分割することにした。その際、JavaScript から扱いやすかった JSON フォーマットで分割したデータを記述することにした。ユーザーの利便性を考慮し、JSON ファイルの読み込みは AJAX を用いて非同期に行い、3 次元空間上への表示には WebGL のライブラリである Three.js を用いて実装している。

一方、Worm Developmental Dynamics Database (WDDD; <http://so.qbic.riken.jp/wddd/>)¹⁰⁾では、線虫 *C. elegans* の初期胚発生における顕微鏡画像データとそこから画像解析を用いて抽出された定量データを同期させて観察することができる。顕微鏡画像はスタックされた 2 次元画像で胚の上から下までをキーボード操作により観察でき、定量データはマウス操作により好きな視点から観察することが可能となっている。また時点の変更もキーボード操作で行うことができる。画像データの可視化においては、ファイルサイズの関係上、ネットワークを介した表示には時間がかかるため、オフラインキャッシュを導入することにより表示におけるユーザー

Fig. 4 BDML4DViewer. Visualization of embryogenesis in *C. elegans* (left) and *D. melanogaster* (right).

Fig. 5 Visualization of microscopy image and quantitative data of nuclear division dynamics of *C. elegans* embryo in the WDDD.

のストレスを低減させている。

5. 議 論

本稿では、生命現象の時空間動態に対する定量データを記述するデータフォーマットとその可視化の現状について報告した。定量データを記述するための統一データフォーマットは、データを生産する側、データ可視化・解析のツールを作成する側の両方に利点があることを議論した。また、可視化に必要な要件とそれを実現したいいくつかのソフトウェアツールを紹介した。

生物学分野の解析の現場において、ある表現型の特徴に対して興味深い現象を見つけた場合に、元の画像データに戻ってその現象を確認するという作業が発生することが多い。このため、生命現象の時空間動態に対する定量データと画像データを同期して観察することができる、解析の効率が飛躍的に向上する。一部の WDDD などのデータベースの可視化ではこのような機能が実装されているが、その他のソフトウェアツールでは実装されていないのが現状である。また、既存の定量データの可視化ツールにおいては、可視化法もまだまだ貧弱である。2次元画像から計測した物体の輪郭を表示するモードのみを備えている場合が多く、例えば、サーフェスレンダリングなどの選択を行うことができない。さまざまな可視化法を備えたソフトウェアツールの開発が期待される。

また、定量データと画像データを同時に、ネットワークを介してウェブブラウザ上で可視化できるシステムの開発も期待されている。これらのデータのファイルサイズは肥大化する傾向にあり、ネットワークを介した際には表示が追いつかずユーザーに対するストレスが大きくなることが予想される。これらの問題を解決するために、ファイルサイズを圧縮したりキャッシュ機構をうまく組み込んだりするなどの工夫が必要であろう。

バイオイメージ・インフォマティクス技術の発展により、画像解析のパイプライン化が進展している⁷⁾。このようなパイプライン化により、大規模な顕微鏡画像データが産出され、そこから大量の時空間動態の定量データが計測されており、生命現象の時空間動態を理解するためのビッグデータが取得されつつある。これらのビッグデータから生物学における新しい知見や発見を得るには、

データを解析する技術だけでなく、データが持つ情報の中で何に注目すべきかをインタラクティブに探索することが可能な可視化技術の開発が今まで以上に期待されている。

参 考 文 献

- 1) Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E.: UCSF ChimeraX: Meeting modern challenges in visualization and analysis, *Protein Sci.*, Vol.27, No.1 (2018) pp.14-25.
- 2) Peng, H.: Bioimage informatics: a new area of engineering biology, *Bioinformatics*, Vol.24, No.17 (2008) pp.1827-1836.
- 3) Keller, P.J.: Imaging morphogenesis: technological advances and biological insights, *Science*, Vol.340, No.6137 (2013) pp.1234168.
- 4) Schneider, C.A., Rasband, W.S., Eliceiri, K.W.: NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods*, Vol.9, No.7 (2012) pp.671-675.
- 5) Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A.: Fiji: an open-source platform for biological-image analysis, *Nat. Methods*, Vol.9, No.7 (2012) pp.676-682.
- 6) Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., Sabatini, D.M.: CellProfiler: image analysis software for identifying and quantifying cell phenotypes, *Genome Biol.*, Vol.7, No.10 (2006) pp.R100.
- 7) Chen, J., Ding, L., Viana, M.P., Hendershott, M.C., Yang, R., Mueller, I.A., Rafelski, S.M.: The Allen Cell Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images, *bioRxiv*, (2018) doi: <https://doi.org/10.1101/491035>
- 8) Kyoda, K., Tohsato, Y., Ho, K.H.L., Onami, S.: Biological Dynamics Markup Language (BDML): an open format for representing quantitative biological dynamics data, *Bioinformatics*, Vol.31, No.7 (2015) pp.1044-1052.
- 9) Tohsato, Y., Ho, K.H.L., Kyoda, K., Onami, S.: SSBD: a database of quantitative data of spatiotemporal dynamics of biological dynamics of biological phenomena, *Bioinformatics*, Vol.32, No.22 (2016) pp.3471-3479.
- 10) Kyoda, K., Adachi, E., Masuda, E., Nagai, Y., Suzuki, Y., Oguro, T., Urai, M., Arai, R., Furukawa, M., Shimada, K., Kuramochi, J., Nagai, E., Onami, S.: WDDD: Worm Developmental Dynamics Database, *Nucleic Acids Res.*, Vol.41 (Database issue), (2013) pp.D732-D737.