

「特集 生体高分子の揺らぎとダイナミクス —シミュレーションと実験の統計解析—」について

伊庭 幸人¹・藤崎 弘士²・松永 康佑³ (オーガナイザー)

1. 特集の内容とその背景

この特集はタンパク質や核酸などの生体分子について、シミュレーションと実験のデータ解析の最近の進展をまとめたものである。これらと関連する話題として、通常の分子動力学では数値的に観測しにくいような準安定状態の間の遷移(レアイベント)を調べるためのシミュレーション技法についても触れている。全部で9編の論文(いずれも「研究詳解」)が収録されている。

生物の基本的構成部品としてのタンパク質・ペプチドや核酸の重要性は論を待たないが、そのイメージは大きく変化してきている。タンパク質を例にとれば、従来のイメージは、高温あるいは変性状態ではランダムに広がっているアミノ酸の鎖が、ほぼ一定の形に3次的に折りたたまれることで、2次構造・3次構造が形成され、タンパク質としての機能が発揮されるというものである。今日でも、情報関係・統計関係の研究者の多くはそのような「フォールディング=最適化」的なイメージを持っているかもしれない。しかしながら、最近の研究によれば、タンパク質の分子は生理的な温度でも大きくダイナミックに揺らいでおり、その状態で生物に必要な機能が発揮されていると考えられている。また、通常はほどけた状態で存在していて、他の分子と相互作用をして機能を発揮する場合にのみ、明確な立体構造を示す分子も数多く発見されており、天然変性タンパク質(IDP)と呼ばれている。

生体高分子の生物学的なイメージが変化する一方で、スーパーコンピュータの発達や専用計算機の登場により、シミュレーションによって、揺らぎ・準安定状態・ダイナミクスを詳しく調べることが可能になりつつある。また、実験でも、1分子計測のような手段が使えるようになって、生体高分子のダイナミクスを時系列として観察できるようになった。そこで必要とされるのが、シミュレーションや実験のデータを解析するための統計科学・統計的機械学習の手法である。また、現在の計算機によっても容易には再現できないレアイベントのダイナミクスを調べるための計算手法も重要である。

2012年の3月に統計数理研究所で開かれた小さな共同利用研究集会「高次元のサンプリングとデータ解析からみたシミュレーション科学」では、本特集のオーガナイザーを含む数名が関連した話題で講演した。その後も研究交流が続き、その上に立って本特集が企画された。結果として、依頼した大部分の方から寄稿していただくことができ、わが国における関連分野の研究者の多くが紙上で一堂に会することになったのは、望外のことであった。寄稿して下さった方々に深くお礼をすると同時に、極めて貴重な記録として、関心を持つ読者にお勧めしたい。

2001年の統計数理特集「地図を描く・風景を眺める」では、タンパク質のシミュレーションの解析について、北尾(2001)と肥後 他(2001)の2本の研究詳解を掲載したが、当時から考えると、

¹ 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

² 日本医科大学 医学部：〒180-0023 東京都武蔵野市境南町 1-7-1

³ 理化学研究所 計算科学研究機構：〒650-0047 兵庫県神戸市中央区港島南町 7-1-26

計算機の能力や生物物理的な内容も、用いられている手法の複雑さも、隔世の感がある。新しい特集が生体複雑系のダイナミクスと統計科学・統計的機械学習の手法が交錯するこの分野への誘いとなれば幸いである。

2. 掲載論文の紹介

掲載論文について、オーガナイザーの立場から簡単な紹介をしておく。いずれの論文も一義的に分類されるわけではなく、あくまで仮に整理しただけなので、その点は誤解をされないようお願いしたい。また文中の手法については、論文の著者の開発したものと既存の手法の応用や解説があるが、この説明ではその区別には重点を置かなかったため、その点については各論文を参照されたい。

2.1 PCA を超えて

高次元空間のデータを扱う手法として、いまの文脈で最も基本的なのは主成分分析(PCA)であろう。PCAでは特異値分解を用いてデータを低次元空間に射影し、次元の圧縮を行う。タンパク質のシミュレーションの解析への応用は1980年代にはじまり、1990年代はじめから本格的に導入された(北尾, 2001, および本特集の桜庭の論文参照)。現在、この分野で提案されている多くの手法は、「通常のPCAでは飽き足りない点がある」ということから出発して、解釈することができる。

タンパク質の揺らぎとダイナミクスを表現することを考えたとき、通常のPCAの主な問題点は次のようになる。

- (1) 非ガウス性や準安定状態の扱いが不十分
- (2) 時系列的な要素を正面から扱っていない
- (3) 生物学的な機能を考慮していない

これらを解決するために、PCAとは異なる枠組みの解析を導入したり、通常とは違う対象についてPCAを適用したりするわけである。

2.2 射影と次元縮約

桜庭の論文では、主に(1)の問題に関して、独立成分分析(ICA)、それを発展させた独立部分空間分析(ISA)の手法が解説され、簡単な例で比較されている。また、(3)に関しても、機能モード分析(FMA)という回帰分析や正準相関分析に似た手法が紹介されている。小山の論文では、位置座標でなく、相互作用エネルギーの各成分について主成分分析を行う手法(PEPCA)、および原子間力と平均溶媒原子数の積に対して関数主成分分析を実行する手法(DIPA)が導入され、シニョリンのフォールディングのシミュレーションが解析されている。小山による枠組みでは、これらの手法は相互作用についての摂動解析として理解されている。

鎌田・戸田の論文は、動的な側面(2)の理解がテーマであるが、そのためにウェーブレット変換を利用している点に特徴がある。ウェーブレット変換と特異値分解を利用した射影を組み合わせることで、複数の時空間階層を持つ複雑なダイナミクスを解明しようと試みている。

2.3 緩和モード解析と時間構造に基づいたICA

高野の論文も、動的な側面(2)の理解が主題であるが、より伝統的な観点に沿って「振幅の大きい運動」の代わりに「遅い運動」の検出をめざしている。高野の論文は、マルコフ連鎖の遷移作用素の固有関数を変分的に求めることで、ダイナミクスを考慮した射影を構成する手法を導

入し、それを緩和モード解析(RMA)と呼んでいる。RMAは1990年代に高野と共同研究者によって開発され、ランダムスピン系やホモポリマーに応用されたものであるが、この論文では、タンパク質のシミュレーションへの応用が解説されている。測上の論文で解説されている手法は、数式的には高野の提案しているものに近いが、時間構造に基づいた独立成分分析(tICA)に基づいており、解釈が異なっている。シミュレーションで得られたタンパク質主鎖の運動に適用した結果が示されている。

2.4 隠れマルコフモデルなど

ここまでで紹介した論文で使われている手法は、対象とする量や部分空間を選ぶ規準は様々であるが、基本的には射影による次元圧縮がベースになっている。これに対して、岡本の論文では、FRETによって得られた1分子時系列データの背後に隠れマルコフモデルを想定して、変分ベイズ法などによってあてはめることで、データを理解しようとしている。なお、岡本のこの論文はシミュレーションデータではなく、実際の実験データ(ここではタンパク質でなくDNA)を解析しているという点でも他と違っている。類似の方法としては、本特集には所収されていないが、確率的な有限状態機械(Epsilon-Machine)による小松崎らの解析手法がある(Li and Komatsuzaki, 2008)。

2.5 レアイベントサンプリング

生体分子シミュレーションにおいて、準安定な形態の間の遷移や分子間の相互作用などの遅い時間スケールの過程をシミュレーションで直接再現することは、困難な課題である(レアイベント問題)。ひとつの解決策として、単純に物理的な過程を模倣することから脱して、計算手法の工夫によって積極的にレアイベントの再現をはかるという考え方がある。残り3つの論文では、こうしたレアイベント・シミュレーションの方法論をめぐる話題が論じられている。

西原・原田・北尾の論文では、並列シミュレーションを使った手法である **Parallel Cascade Selection Molecular Dynamics(PaCS-MD)** がレアイベントのサンプリングに使われている。これは、広義の逐次モンテカルロ法(粒子モンテカルロ法)の一種と考えられる。また、この方法によって得られた結果をクラスター分析し、クラスターによって定義された状態間の遷移をマルコフ連鎖であらわす **Markov State Model(MSM)** の方法が論じられている。

松永の論文では、レアイベントシミュレーションの一方法であるストリング法が、著者の開発した部分を含めて解説され、アデニル酸キナーゼへの応用が示されている。藤崎の論文は、もうひとつの大きな流れであるパスサンプリング法を扱っているが、単なるサンプリングだけではなく、拡散定数や反応座標をシミュレーション結果からどのように推定するかということが主題になっている。

近年、この分野のハードウェアの進歩は著しく、専用計算機 **Anton** の登場によって数万原子系で数ミリ秒に相当する計算が実現した。**Anton 2** や **MDGRAPE-4** のような次世代の専用計算機ではさらに長時間・多原子の計算が可能となることが期待されるが、依然として手法的な工夫も重要と思われる。また、ハードウェア技術とレアイベントサンプリング法は、原理的には併用可能であり、相補的なものと考えられる。

3. 手法に関する補足とコメント

オーガナイザーの1人(伊庭)は、研究ノートの投稿を予定していたが、体調不良などで断念した。代わりに、この場を借りて、いくつか簡単なコメントを述べておきたい。

3.1 独立成分と緩和モード

多変量の場合に「独立」か否かをデータから判定するのは簡単でない。従って、独立成分分析(ICA)といっても、実際には特定の種類の非独立性をターゲットにしていることが多い。時間構造に基づいた独立成分分析(tICA)では、時間遅れを考慮することでPCAでうまく分解できない成分が分離できる可能性があるが、2次相関(線形相関)を利用していることによる限界がある。たとえば、以下のような時系列信号 $\{x_t\}, \{y_t\}$ は互いに独立ではないが、2次相関に基づくtICAでは分離できない。

例 1. $\{u_t\}$ は $[0, 2\pi)$ に一様に分布する独立な列。 $x_t = \cos u_t, y_t = \sin u_t$.

例 2. $\{u_t\}$ は角度 $u_1 = 0$ からスタートした円周上のランダムウォーク。 $x_t = \cos u_t, y_t = \sin u_t$.

いずれの場合も、加法定理より

$$\cos u_t \sin u_{t+\tau} = \frac{1}{2} \{ \sin(u_{t+\tau} + u_t) + \sin(u_{t+\tau} - u_t) \}$$

となることから、すべての τ, t について

$$\mathbb{E}(x_t y_{t+\tau}) = 0$$

となる。また、

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N x_t y_{t+\tau} = 0$$

も成り立つ。すべての2次相関が、サンプル数が多い極限、観測時間の長い極限でゼロになるので、tICAによる分離は明らかに不可能である。しかし、 $x_t^2 + y_t^2 = 1$ なので、 $\{x_t\}, \{y_t\}$ は互いに独立ではない。

本特集の高野論文でいう「緩和モード」(マルコフ連鎖の遷移作用素の固有関数)ではあるが、そこへの射影が独立成分ではない例もあげておく。tICAによっても緩和モードと同様のものが求まると思われるが、それらは独立成分としては解釈できない。

物理的にわかりやすい例としては次がある。

例 3. 動的イジング模型の磁化 $\{x_t\}$ とエネルギー $\{y_t\}$ 。状態変化を定義するマルコフ連鎖としては、たとえば「ランダムに選んだ1スピンを更新する熱浴法(Gibbs Sampler)」を使用する。

外場なしのモデルでは、磁化とエネルギーは対称性が異なるため、異なる緩和モードに属する。しかし、結合定数が正のモデル(強磁性モデル)の場合、エネルギーの値が小さいと秩序が発達して磁化の絶対値(あるいは2乗)が大きくなる傾向があるので、両者は独立ではない。

より単純化すると、以下の例となる。

例 4. 3状態 $i = 1, 2, 3$ のマルコフ連鎖で、遷移行列 $P_{ij} = P(j \rightarrow i)$ が

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

で定義されるものを考える。この場合、右固有ベクトル(縦ベクトル)は ${}^t(1, 1, 1)$, ${}^t(1, 0, -1)$, ${}^t(1, -2, 1)$ となり、それぞれに対応する固有値は $1, \frac{2}{3}, 0$ である(t は転置の意味)。定常分布は固有値1に対応する固有ベクトル ${}^t(1, 1, 1)$ に対応し、3状態の上の一様分布となる。

このとき、固有ベクトル ${}^t(1, 0, -1)$ との内積の値を X , ${}^t(1, -2, 1)$ との内積の値を Y とすると、

表 1. X^2 と Y .

状態	X の値	X^2 の値	Y の値	確率
1	1	1	1	1/3
2	0	0	-2	1/3
3	-1	1	1	1/3

表 1 のようになる。 X と Y は明らかに独立ではない。また、 $\mathbb{E}(X^2Y) \neq 0$ かつ $\mathbb{E}(X^2)\mathbb{E}(Y) = 0$ となることから、 X と Y に非線形相関があり、独立ではないことがわかる。

なお、マルコフ連鎖を議論する場合の「射影」は左右の固有ベクトルを区別する必要があるが、この例では遷移行列が対称なのでどちらでも同じである。

自己相関や自己共分散を利用した成分分解・次元圧縮の手法は他にも提案されている。たとえば、**Maximum Autocorrelation Factor (MAF)** と呼ばれる手法がある (Cunningham and Ghahramani, 2014)。これはもともと、空間データについて提案されたものだが (Switzer and Green, 1984)、時系列にも応用されている (Shapiro and Switzer, 1989)。空間版について、ICA との関係も指摘されている (Larsen, 2002)。さらに古いものとしては、多変量 AR モデルを背景として考えられた類似の手法がある (Box and Tiao, 1977)。

tICA が機械学習で生まれた考え方であるのに対し、高野らの方法での「緩和モード」(マルコフ連鎖の遷移作用素の固有関数) という解釈は、機械学習ではあまり一般的でないようである。そこで、シミュレーションの解析以外にもその解釈を一般化することが考えられるが、その場合に問題になるのは「定常性」の要求である。データが背後にある巨大次元のマルコフ連鎖から生じた定常時系列からのサンプルであると仮定しないと、高野らの解釈は成立しない。応用分野によってはこれが制約になるだろう。

高野の方法での「緩和モード」は本来は高次元の状態空間(分子の形態の空間)で定義されており、低次元の基底で展開することで、はじめて主成分分析と比較される。しかし、逆の見方をすれば、主成分分析についてカーネル主成分分析や関数主成分分析を考えることで、同じレベルで比較することもできる。オーガナイザーの 1 人(伊庭)は、学会発表のみであるが、高野の方法を再生カーネル法と融合する試みを行ったことがある (伊庭, 2008, 2009)。これは tICA とカーネル法を逐次処理の形で結合した手法 (Harmeling et al., 2003) に近いが、概念上、両者がより一体化された形になっている。

3.2 主成分分析のための数値的方法

本特集の小山の論文の付録 A には、主成分分析のためのアルゴリズムとして、共分散行列の対角化、中心化データ行列の特異値分解、中心化グラム行列の対角化があげられており、次元数(特徴量の数)とデータ数に応じて使いわけることが示唆されている。

これに対して、機械学習の文献 (Shawe-Taylor, 2004; 福水, 2010 の 3.6 節) では、残差 2 乗和をモニターしながら中心化グラム行列の不完全コレスキー分解を行い、残差 2 乗和が十分小さくなったところでやめる方法が解説されている。この方法は、ランダム(あるいはあらかじめ決めた順)に新しいデータ点を取り入れて、これまでの点の全体と新しい点が定める基底が張る空間を逐次的に考えていくことに相当する。この方法を生体高分子のシミュレーションの解析に取り入れることもできるが、注意すべきなのは、シミュレーションではデータ点が時系列で与えられ、独立サンプルではないことである。時間間隔をおくほどデータ点同士の独立性が高く、データの広がりについての独立な情報を多く持っているとすれば、「適当な間隔でデータをサブサンプルしたもので作った中心化グラム行列を対角化し、その間隔を小さくして収束を見

る」といった既存手法のほうが、シミュレーションからのデータに対しては良い可能性がある。

これとは別に、次元数およびデータ点の数がきわめて多い場合の特異値分解やグラム行列の扱いについてはランダムネスを含む算法が機械学習などの分野で提案されており、それらを使う方法は今後検討の余地がある(福水, 2010 の 3.6 節; Halko et al., 2011)。

逆に、機械学習や統計科学の側からみると、データ点が時系列で与えられ時間相関のある場合に効率の良い算法を開発するのは、興味ある課題であろう。

3.3 逐次モンテカルロ法(粒子モンテカルロ法)

西原・原田・北尾の論文で提案されているレアイベントサンプリングの手法、および、そこで関連手法として引用されている **Foward Flux Sampling**, **Weighted Ensemble** などの手法(Allen et al., 2009 参照)は、統計的重み付けの厳密性には違いがあるが、広い意味で逐次モンテカルロ法(**Sequential Monte Carlo**)、あるいは「粒子モンテカルロ法」「ポピュレーション型のモンテカルロ法」などと呼ばれる手法群(Iba, 2001; 伊庭, 2005)の一例とみなすことができる。

このタイプの手法では状態空間の点をあらわす「粒子」に重みを与え、重みに応じて分裂させあるいは消去することで、重点サンプリングを実現する。「粒子」は「レプリカ」と呼ばれることもあるが、レプリカ交換モンテカルロ法の「レプリカ」とはかなり性質が異なっている。こうした考え方の起源は、Kalos による量子系のモンテカルロ手法(拡散モンテカルロ法, グリーン関数モンテカルロ法)(Kalos, 1962)、および、自己回避乱歩のシミュレーション手法(Wall, 1959)などにあるが、近年では多くの分野で使われるようになってきている。以下では、読者が興味があると思われる分野を中心に文献を紹介する。

- 逐次モンテカルロ法の応用分野で、近年もっとも注目を集めているのは状態空間モデルによる時系列データの解析である(Doucet et al., 2001; 北川, 2005; 伊庭, 2005; 樋口, 2011a)。シミュレーションそのものを状態空間での遷移と考えた場合、データに合致する初期値、経路、パラメータを選んでゆくことになるが、そのような手法はデータ同化と呼ばれる(樋口, 2011b)。
- 物理系のレアイベントサンプリング手法としての逐次モンテカルロ法については既に上で述べたが、少し違った観点から非線形力学系や非平衡の確率モデルに応用した一連の研究については Kurchan et al., 2011 を参照。
- 福島と伊庭(Hukushima and Iba, 2003)は、温度を1方向的に変化させながら、システム全体の状態を表現する粒子(レプリカ)の分裂と消滅を行うことで、アニール速度が有限でも正しい熱平均が効率よく計算できることを示し、スピニンググラスに応用した(**Population Annealing**)。一般化と分子動力学への応用は Rousset and Stoltz(2006)参照。同様の手法で温度の代わりにエネルギーの閾値を用いたものは **Nested Sampling** と呼ばれる(Skilling, 2004, 2006)。
- 自己回避乱歩のシミュレーションの手法の延長線上で、成長するシステムのレアイベントサンプリングを考えたものとして、Grassberger の **PERM (Prune Enriched Rosenbluth Method)**がある(Hsu and Grassberger, 2011)。PERM は格子タンパクの折りたたみの計算にも使われている(同上)。

大規模な生体分子シミュレーションの世界では、分子動力学法がもっぱら主流で、それに、マルコフ連鎖モンテカルロ法(動的モンテカルロ法)の世界で開発されたレプリカ交換モンテカルロ法、マルチカノニカル法などが組み合わされている、というのが現状であるが、逐次モンテカルロ法(粒子モンテカルロ法)の柔軟さと並列性の高さは魅力的であり、今後重要となる可能性がある。

謝 辞

本特集に寄稿して頂いた皆様への謝意は既に本文に記したところですが、改めてお礼を申し上げます。この序文の執筆にあたって、小山洋平氏からは、3.2 節に関して、ご自身のデータと経験に基づいた有益なご示唆を頂きました。また、3.1 節の文献 Cunningham and Ghahramani, 2014; Shapiro and Switzer, 1989; Switzer and Green, 1984; Larsen, 2002, および、3.2 節の文献 Halko et al., 2011 をご教示頂きました。また、高野宏氏には 3.1 節について有用なご指摘を頂きました。

参 考 文 献

- Allen, R. J., Valeriani, C. and ten Wolde, P. R. (2009). Forward flux sampling for rare event simulations, *Journal of Physics: Condensed Matter*, **21**, 463102.
- Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series, *Biometrika*, **64**(2), 355–365.
- Cunningham, J. P. and Ghahramani, Z. (2014). Unifying linear dimensionality reduction, arXiv: 1406.0873 [stat.ML].
- Doucet, A., Freitas, N. and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*, Springer, Berlin, Heidelberg.
- 福水 健次 (2010). 『カーネル法入門—正定値カーネルによるデータ解析』, 朝倉書店, 東京.
- Giardina, C., Kurchan, J., Lecomte, V. and Tailleur, J. (2011). Simulating rare events in dynamical processes, *Journal of Statistical Physics*, **145**(4), 787–811.
- Halko, N., Martinsson, P. G. and Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.*, **53**(2), 217–288.
- Harmeling, S., Ziehe, A., Kawanabe, M. and Müller, K.-R. (2003). Kernel-based nonlinear blind source separation, *Neural Computation*, **15**, 1089–1124.
- 樋口知之 (2011a). 『予測にいかす統計モデリングの基本—ベイズ統計入門から応用まで』, 講談社サイエンスフィク, 東京.
- 樋口知之 (編著) (2011b). 『データ同化入門—次世代のシミュレーション技術—』, 朝倉書店, 東京.
- 肥後順一, 小野 聡, 中島伸介, 中村春木 (2001). 多次元空間解析で迫る生体高分子の 3 次元立体構造と Energetics, *統計数理*, **49**(1), 57–75.
- Hsu, H.-P. and Grassberger, P. (2011). A review of Monte Carlo simulations of polymers with PERM, *Journal of Statistical Physics*, **144**(3), 597–637.
- Hukushima, K. and Iba, Y. (2003). Population annealing and its application to a spin glass, *AIP Conference Proceedings*, **690**, 200–206.
- Iba, Y. (2001). Population Monte Carlo algorithms, *人工知能学会誌*, **16**, 279–286.
- 伊庭幸人 (2005). 逐次モンテカルロ法入門, 『計算統計 II』(統計科学のフロンティア 12), 293–326, 岩波書店, 東京.
- 伊庭幸人 (2008). カーネル法による緩和モード解析, 日本物理学会第 63 回年次大会(東大阪).
- 伊庭幸人 (2009). 緩和モード解析の改良, 日本物理学会 64 回年次大会(東京).
- Kalos, M. H. (1962). Monte Carlo calculation of the ground state of three- and four-body nuclei, *Physical Review*, **128**, 1791–1795.
- 北川源四郎 (2005). 『時系列解析入門』, 岩波書店, 東京.
- 北尾彰朗 (2001). 主成分分析を使って眺めた蛋白質のエネルギー地形, *統計数理*, **49**(1), 43–56.
- Larsen, R. (2002). Decomposition using maximum autocorrelation factors, *Journal of Chemometrics*, **16**(8-10), 427–435.
- Li, C.-B., Yang, H. and Komatsuzaki, T. (2008). Multiscale complex network of protein conformational

- fluctuations in single-molecule time series, *Proceedings of the National Academy of Sciences of the United States of America*, **105**(2), 536–541.
- Rousset, M. and Stoltz, G. (2006). Equilibrium sampling from nonequilibrium dynamics, *Journal of Statistical Physics*, **123**(6), 1251–1272.
- Shapiro, D. E. and Switzer, P. (1989). Extracting time trends from multiple monitoring sites, Technical Report, Department of Statistics, Stanford University, Stanford.
- Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge.
- Skilling, J. (2004). Nested sampling, *AIP Conference Proceedings*, **735**, 395–405.
- Skilling, J. (2006). Nested sampling for general Bayesian computation, *Bayesian Analysis*, **1**(4), 833–860.
- Switzer, P. and Green, A. A. (1984). Min/max autocorrelation factors for multivariate spatial imagery, Technical Report, Department of Statistics, Stanford University, Stanford.
- Wall, F. T. and Erpenbeck, J. J. (1959). New method for the statistical computation of polymer dimensions, *The Journal of Chemical Physics*, **30**(3), 634–637.