

Topological Measurement of Protein Compressibility via Persistence Diagrams

Gameiro, Marcio

Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo

Hiraoka, Yasuaki

Institute of Mathematics for Industry, Kyushu University

Izumi, Shunsuke

Department of Mathematical and Life Sciences, Hiroshima University

Kramar, Miroslav

Department of Mathematics, Hill Center-Busch Campus, Rutgers, The State University of New Jersey

他

<https://hdl.handle.net/2324/22032>

出版情報 : MI Preprint Series. 2012-6, 2012-06-04. 九州大学大学院数理学研究院
バージョン :
権利関係 :

MI Preprint Series
Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

**Topological Measurement of
Protein Compressibility via
Persistence Diagrams**

**Marcio Gameiro,
Yasuaki Hiraoka,
Shunsuke Izumi,
Miroslav Kramar,
Konstantin Mischaikow and
Vidit Nanda**

MI 2012-6

(Received June 4, 2012)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Topological Measurement of Protein Compressibility via Persistence Diagrams

Marcio Gameiro* Yasuaki Hiraoka[†] Shunsuke Izumi[‡] Miroslav Kramar[§]
Konstantin Mischaikow[¶] Vidit Nanda^{||}

Abstract

We exploit recent advances in computational topology to study the compressibility of various proteins found in the Protein Data Bank (PDB). Our fundamental tool is the *persistence diagram*, a topological invariant which captures the sizes and robustness of geometric features such as tunnels and cavities in protein molecules. Based on certain physical and chemical properties conjectured to impact protein compressibility, we propose a topological measurement C_P for each protein molecule P .

C_P can be efficiently computed from the PDB data of P . Our main result establishes a clear linear correlation between C_P and the experimentally measured compressibility of most proteins for which both PDB information and experimental compressibility data are available.

Keywords. Protein Compressibility, Persistent Homology, Persistence Diagram, Weighted Alpha Complex

1 Introduction

A protein is a soft material which acts as an enzyme and/or functions as a receptor. The softness of a protein is related to cavities present in its molecular structure. However, various definitions of “cavity” are used depending on which particular biochemical phenomena are being investigated (e.g., active sites). Our approach uses computational topology to deal with cavities in a rigorous manner.

Topology is a branch of mathematics which studies holes of various dimensions as well as how these holes fit together in a given geometric object. Recent progress in computational topology (e.g., [7, 14, 15]) provides pure mathematical tools which may be applied to data-driven problems in a wide variety of scientific contexts (e.g., [1, 2]). In this paper, we apply computational topology to study a physical and chemical property of proteins.

The stiffness of a protein is related to its tertiary structure and has an effect on structural stability, spatial fluctuations and functionality. One of the quantities which characterizes the stiffness of a protein is its *compressibility* [4, 6, 9]. The compressibility of some globular proteins has been studied experimentally

*Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Caixa Postal 668, 13560-970, São Carlos, SP, Brazil (e-mail: gameiro@icmc.usp.br).

[†]Institute of Mathematics for Industry, Kyushu University, 744, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan (e-mail: hiraoka@imi.kyushu-u.ac.jp).

[‡]Department of Mathematical and Life Sciences, Hiroshima University, 1-3-1, Kagamiyama, Higashi-Hiroshima, 739-8526, Japan (e-mail: sizumi@sci.hiroshima-u.ac.jp).

[§]Department of Mathematics, Hill Center-Busch Campus, Rutgers, The State University of New Jersey, 110 Frelinghusen Rd, Piscataway, NJ 08854-8019, USA (e-mail: miroslav@math.rutgers.edu).

[¶]Department of Mathematics and BioMaPS, Hill Center-Busch Campus, Rutgers, The State University of New Jersey, 110 Frelinghusen Rd, Piscataway, NJ 08854-8019, USA (e-mail: mischaik@math.rutgers.edu).

^{||}Department of Mathematics, Hill Center-Busch Campus, Rutgers, The State University of New Jersey, 110 Frelinghusen Rd, Piscataway, NJ 08854-8019, USA (e-mail: vidit@math.rutgers.edu).

[4]. From these experimental studies, it is conjectured that cavities in protein molecules play an important role in spatial fluctuations even though proteins have a relatively dense packed structure.

The main objective of this paper is to study compressibility by *persistence diagrams* [1, 12]. In particular, we discuss geometric and topological sub-structures of protein molecules whose presence impacts compressibility. Persistence diagrams indicate the presence of holes as well as the robustness of each hole to the locations and sizes of atoms in a given protein molecule. Moreover, these diagrams can be efficiently computed [11, 12] from X-ray crystallography data found in the Protein Data Bank [17] via the *alpha complex* representation [3]. Based on certain physical and chemical properties which are conjectured to impact compressibility, we derive a topological measurement C_P from the persistence diagrams.

Our main result shows that C_P exhibits a remarkable linear correlation with most experimental compressibility data present in [4]. We also show that the feature size of cavities which impact compressibility may be estimated from such topological computations.

2 Topological Methods

In this section we describe the abstract geometric objects called *alpha complexes* which will be used to model protein molecules. We also provide a heuristic introduction to *persistent homology* which is our primary topological tool for understanding the structure of each cavity present in a protein molecule.

2.1 Alpha Complex Models of Protein Molecules

A standard geometric representation of an atom is a ball with the van der Waals radius. Each protein is built from a one dimensional sequence of amino acids and can be modeled as a union of the van der Waals balls corresponding to the atoms in these amino acids. This model is used in [10] to estimate surface areas and the volumes of proteins, and the key tool in that analysis is the alpha complex [3].

Given a protein molecule P , let \mathcal{X}_P be a finite set of points in \mathbb{R}^3 denoting the *centers* of atoms present in the amino acid sequence which comprises P . Associate to each $\xi \in \mathcal{X}_P$ a number $w(\xi) > 0$ which equals the van der Waals radius of the atom centered at ξ .

For each $x \in \mathbb{R}^3$ and $r > 0$, let $B_r(x)$ denote the three dimensional solid ball centered at x with radius r . Then, $\mathcal{B}(P) = \{B_{w(\xi)}(\xi) \mid \xi \in \mathcal{X}_P\}$ is a collection of solid balls which provides a simple geometric model for the molecule P . However, this static model is very sensitive to errors in the measurement of atom centers \mathcal{X}_P as well as the radii $w(\xi)$. In order to provide some robustness to these types of errors, we consider a one-parameter family of such models.

For each $\xi \in \mathcal{X}_P$, define the radius function $r_\xi : [0, \infty) \rightarrow \mathbb{R}$ by

$$r_\xi(\alpha) = \sqrt{\alpha + w(\xi)^2}, \quad (2.1)$$

and define $\mathcal{B}_\alpha(P) = \{B_{r_\xi(\alpha)}(\xi) \mid \xi \in \mathcal{X}_P\}$. For any given value of α , $\mathcal{B}_\alpha(P)$ is a collection of balls centered at the points in \mathcal{X}_P ; at $\alpha = 0$ we recover the static model $\mathcal{B}(P)$ defined above and as α increases, so does the radius of each ball in $\mathcal{B}_\alpha(P)$.

For each $\xi \in \mathcal{X}_P$, we define the *weighted distance function* $\mathbf{d}_\xi : \mathbb{R}^3 \rightarrow \mathbb{R}$ by

$$\mathbf{d}_\xi(x) = \|x - \xi\|^2 - w(\xi)^2,$$

where $\|\cdot\|$ denotes the standard Euclidean distance. These functions may be used to partition \mathbb{R}^3 into finitely many pieces $\{\mathbf{V}_\xi \mid \xi \in \mathcal{X}_P\}$ called *Voronoi cells* defined as follows:

$$\mathbf{V}_\xi = \{x \in \mathbb{R}^3 \mid \mathbf{d}_\xi(x) \leq \mathbf{d}_{\xi'}(x) \text{ for each } \xi' \neq \xi \text{ in } \mathcal{X}_P\}.$$

That is, \mathbf{V}_ξ is the set of all those points in \mathbb{R}^3 whose weighted distance to ξ is less than or equal to their weighted distances to all other points in \mathcal{X}_P . It is easy to check that each point x in \mathbb{R}^3 lies in at least one such \mathbf{V}_ξ whenever \mathcal{X}_P is non-empty.

The Voronoi cells also partition the union \mathcal{U}_α of balls in $\mathcal{B}_\alpha(P)$ for each α . In order to keep track of these partitions systematically, we define new sets $\mathbf{W}_\xi(\alpha) = B_{r_\xi(\alpha)}(\xi) \cap \mathbf{V}_\xi$. Then, for each $\alpha \geq 0$ we have the decomposition

$$\mathcal{U}_\alpha = \bigcup_{B \in \mathcal{B}_\alpha(P)} B = \bigcup_{\xi \in \mathcal{X}_P} \mathbf{W}_\xi(\alpha),$$

and moreover, $\mathbf{W}_\xi(\alpha) \subset \mathbf{W}_\xi(\alpha')$ whenever $\alpha \leq \alpha'$ for any $\xi \in \mathcal{X}_P$.

Definition 2.1 Given a topological space \mathcal{U} and a finite collection of subsets $\mathcal{W} = \{\mathbf{W}_k \mid 1 \leq k \leq n\}$ whose union $\bigcup_k \mathbf{W}_k$ equals \mathcal{U} , the *nerve* of \mathcal{W} is the abstract simplicial complex $\mathcal{N}(\mathcal{W})$ defined as follows: to each collection of $p+1$ distinct elements of \mathcal{W} whose intersection is non-empty, we associate a p -simplex.

In particular, for each subset of \mathcal{U} in \mathcal{W} , there is a *vertex* of $\mathcal{N}(\mathcal{W})$. There is an *edge* between two such vertices if the corresponding subsets intersect. Each *triangle* of $\mathcal{N}(\mathcal{W})$ corresponds to three such subsets with at least a point of common intersection, and so on for higher dimensional simplices.

It is a consequence of the well-known *Nerve lemma* that if each subset of \mathcal{U} in the collection \mathcal{W} is *convex*, then the simplicial complex $\mathcal{N}(\mathcal{W})$ has the same *homology groups* as those of \mathcal{U} . In particular, the finite combinatorial object $\mathcal{N}(\mathcal{W})$ accurately captures all the connected pieces, tunnels and cavities of the topological object \mathcal{U} . With this in mind, we define the alpha complex.

Definition 2.2 (Alpha Complex) For each $\alpha \geq 0$, the *alpha complex* X_α associated to $\mathcal{B}_\alpha(P)$ is the nerve of the collection of subsets $\{\mathbf{W}_\xi(\alpha) \mid \xi \in \mathcal{X}_P\}$ of \mathcal{U}_α .

We show an example of the alpha complex in Figure 1.

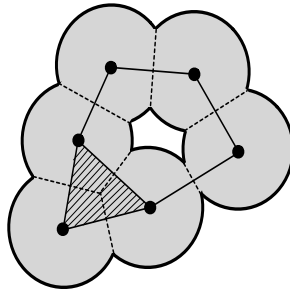


Figure 1: A union of balls in the plane. The dashed lines indicate partitions by Voronoi cells and the associated alpha complex is overlaid.

It should be noted that the alpha complexes X_α and $X_{\alpha'}$ with $\alpha \leq \alpha'$ satisfy an inclusion relation $X_\alpha \subset X_{\alpha'}$. This follows from the fact that each $\mathbf{W}_\xi(\alpha)$ is a subset of $\mathbf{W}_\xi(\alpha')$ and therefore each p -simplex of X_α uniquely determines a p -simplex of $X_{\alpha'}$. It is helpful to think of α as a *scale*. Simplices introduced at larger values of α correspond to larger feature sizes than those introduced at smaller values of α .

In the next section, we see how the collection of simplicial complexes $\{X_\alpha \mid \alpha \geq 0\}$ can be used to study topological structures (e.g., tunnels and cavities) in the unions of balls $\{\mathcal{U}_\alpha \mid \alpha \geq 0\}$. More importantly, the inclusion induced by increasing α provides information regarding the size and robustness of these topological structures.

2.2 Persistent Homology

The *homology groups* $H_n(X)$ of a simplicial complex X provide information about the topological features of X . The n -dimensional holes may be heuristically thought of as connected components, tunnels and

cavities for $n = 0, 1$, and 2 , respectively (see [5, Ch. 2] for a thorough introduction). Let X_α be the alpha complex defined in Section 2.1. Then, the homology groups $H_n(X_\alpha)$ with \mathbb{R} -coefficients may be represented as

$$H_n(X_\alpha) \cong \begin{cases} \mathbb{R}^{\beta_n(\alpha)}, & n = 0, 1, 2, \\ 0, & n > 2. \end{cases}$$

Here, the dimension $\beta_n(\alpha)$ expresses a count of n -dimensional holes in \mathcal{U}_α and is called the n -th *Betti number* of \mathcal{U}_α .

A *filtration* is an increasing sequence of simplicial complexes. Recall from the previous section that the collection \mathbb{X} of alpha complexes defined by $\{X_\alpha \mid \alpha \geq 0\}$ is naturally a filtration because $X_\alpha \subset X_{\alpha'}$ whenever $\alpha < \alpha'$. The inclusion of simplices across the sequence \mathbb{X} allows one to systematically track values of α at which n -dimensional holes appear and disappear.

More precisely, pick any n -dimensional hole \mathbf{z} in X_{α_0} . It is possible to unambiguously define positive real numbers $\mathbf{b}_z \leq \alpha_0 \leq \mathbf{d}_z$ called the *birth* and *death* scales of \mathbf{z} such that \mathbf{b}_z is the smallest value of α for which \mathbf{z} is an n -dimensional hole in X_α and \mathbf{d}_z is the largest such value. The quantity $\mathbf{d}_z - \mathbf{b}_z$ provides an effective measure of the *robustness* of \mathbf{z} to fluctuations in the scale α .

The *persistence diagram* $\text{PD}_n(\mathbb{X})$ of the filtration \mathbb{X} is defined to be the set of points $(\mathbf{b}_z, \mathbf{d}_z) \in \mathbb{R}^2$ where \mathbf{z} ranges over all the n -dimensional holes in \mathbb{X} . An important feature of persistence diagrams is their *stability* [8]. If one perturbs each point in the set \mathcal{X}_P of atom centers by an amount $\delta > 0$, the points in the associated persistence diagrams $\text{PD}_n(\mathbb{X})$ are perturbed by an amount smaller than δ . In this precise sense, persistence diagrams are stable to errors in measurement.

Figure 2 shows a sample filtration \mathbb{X} and Figure 3 shows its corresponding persistence diagram $\text{PD}_1(\mathbb{X})$. Note that the two cycles which appear in the filtration are represented by the two points in the persistence diagram.

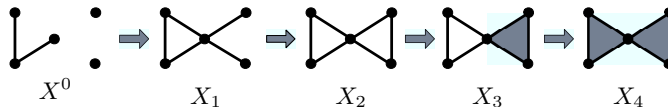


Figure 2: Filtration \mathbb{X}

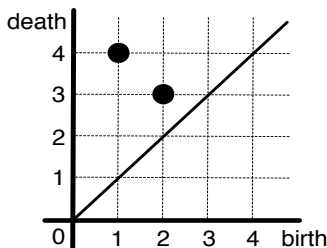


Figure 3: Persistence diagram $\text{PD}_1(\mathbb{X})$

Finally, we remark that persistence diagrams of a given filtration \mathbb{X} are efficiently computable [11, 12].

3 Main Results

In this section we outline the construction of persistence diagrams for a given protein P , define a topological quantity C_P and show that this quantity is closely related to experimentally available compressibility data of P .

3.1 Persistence Diagrams of Proteins

For each protein P in the Protein Data Bank (PDB), we extract the set \mathcal{X}_P of atom centers. The van der Waal radius associated to each atom type in \mathcal{X}_P is provided in Table 1 below. The alpha complexes are constructed using the CGAL software library [13] and the parameter α ranges over $[0, 12]$. Note that $\alpha = 0$ corresponds to the van der Waal ball model of P by (2.1). The persistence diagrams are computed using the Perseus software available at [16].

atom	radius(\AA)
C	1.70
N	1.55
O	1.52
P	1.80
S	1.80

Table 1: van der Waals radii

Figure 4 shows the persistent diagrams $\text{PD}_n, n = 1, 2$, of ovalbumin (PDB ID: 1OVA). As explained in Section 2, each point in PD_1 corresponds to a tunnel and each point in PD_2 corresponds to a cavity in the filtration \mathbb{X} . The x and y coordinates of each point provide the birth and death scale respectively of the tunnel or cavity being represented by that point.

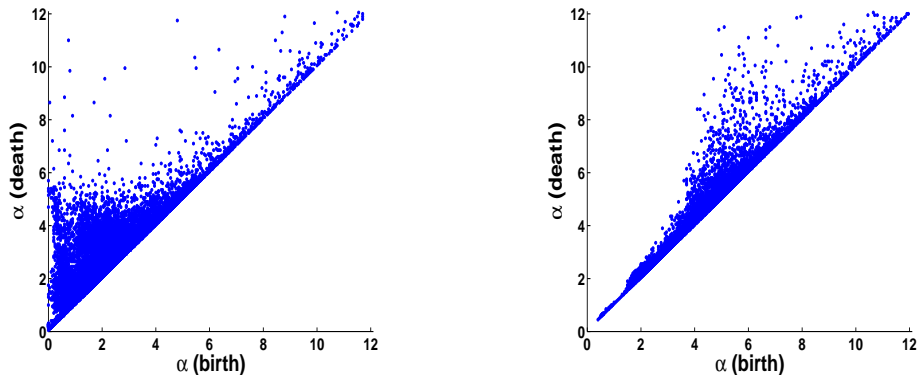


Figure 4: PD_1 (left) and PD_2 (right) of 1OVA.

Such persistence diagrams provide topological insight into protein structures and are of independent interest. For example, the points near the diagonal in PD_1 and PD_2 correspond to non-robust tunnels and cavities, respectively. More importantly, the tunnels and cavities which are robust to fluctuations of scale and locations of atom centers are located far from the diagonal.

We are constructing a database of persistence diagrams of PDB proteins. We encourage the readers to freely download these computational results from our website [14].

3.2 Compressibility from Persistence Diagrams

We define a topological quantity which is well-correlated to the experimental compressibility of proteins as follows. Fix a protein P and let PD_n denote the n dimensional persistence diagram associated to the alpha complex filtration \mathbb{X} of P . Given real numbers $\ell \leq u$, let us use $|\text{PD}_n(\ell, u)|$ to denote the number of points (x, y) in PD_n such that $\ell \leq x \leq u$ and $y - x \geq 1$. Then, define

$$C_P := \frac{|\text{PD}_2(5.9, 8.8)|}{|\text{PD}_1(5.9, 8.6)|}. \quad (3.1)$$

For comparison, we use the experimental compressibility data of proteins from [4]. We restrict our analysis to only those proteins whose PDB data under similar experimental conditions is also available. The numerical result is summarized in Figure 5. The experimental compressibility data from [4] is laid along the vertical axis of the figure and the topological quantity from equation (3.1) is on the horizontal axis. Each plotted point in the figure represents a protein and is labeled with its corresponding PDB ID.

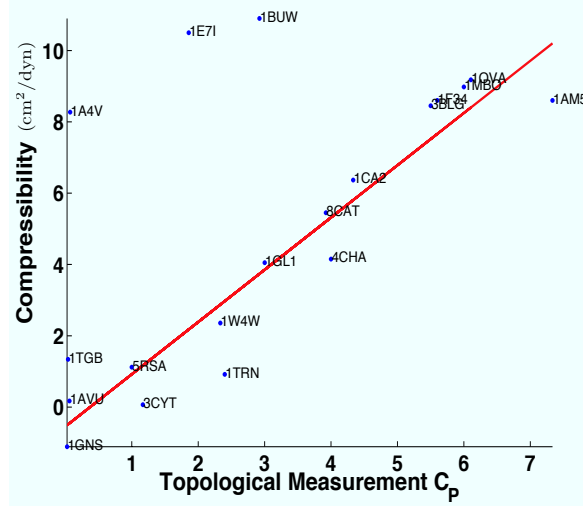


Figure 5: Topological measurement C_P and the compressibility of the proteins

The figure clearly shows a linear correlation between the topological quantity C_P and the compressibility for most such proteins, with the only exceptions being 1A4V, 1E7I, and 1BUW.

The points used in computing C_P for 1OVA are colored red in Figure 6.

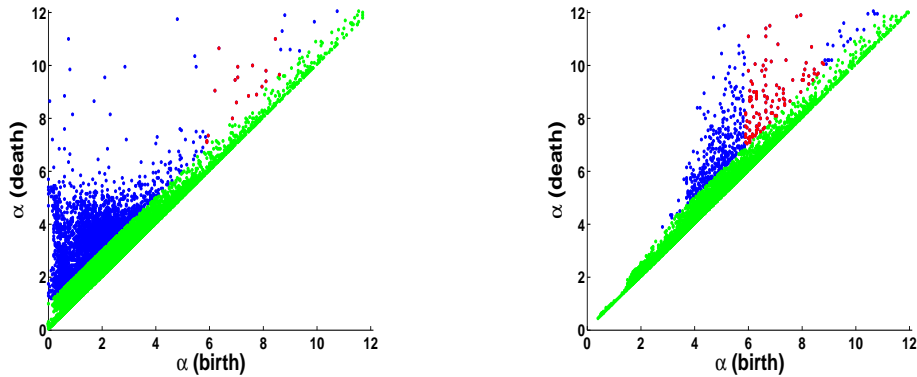


Figure 6: PD_1 (left) and PD_2 (right) of 1OVA. The points used in the computations of C_P are colored red. The green points correspond to non-robust features and are excluded.

3.3 The Derivation of C_P

The derivation of C_P has two fundamental aspects. First, we investigate some structural features of protein molecules which are reasonably expected to impact compressibility. Once these features have been determined, we conduct parameter searches to isolate those pieces of persistence diagrams which correspond to the relevant features.

Before providing details, we remark that even though it is reasonable to expect the presence of holes in protein molecules to impact compressibility, measurements derived from Betti numbers alone did not yield any quantities which correlated nicely with experimental compressibility. Only after we incorporated the stable framework of persistent homology did it become possible to identify robust noise-independent features in a given molecule. As it turns out, isolating those robust features was an extremely important step towards obtaining the linear correlation shown in Figure 5.

One begins the analysis with the reasonable assumption that non-robust topological features do not have a large effect on compressibility. Hence, we introduce a *noise parameter* $\delta > 0$ such that for $i \in \{1, 2\}$, the points

$$N_i(\delta) := \{(b, d) \in \text{PD}_i \mid d - b < \delta\}$$

will be treated as a topological *noise* and be removed from consideration.

Persistent homology enables us to systematically treat the *sparseness* of holes as a by-product of knowing the birth scales. Let us consider the cross-sections of two such holes in Figure 7. Clearly, the sparse hole on the right is deformable to a much larger extent than the dense hole on the left. Thus, we assume that a larger number of sparse holes leads to greater compressibility. The distinction between dense and sparse holes is also readily captured by persistence diagrams: the denser the hole, the smaller its birth scale. In light of this assumption, we introduce four more *sparseness* parameters $(\ell^{(i)}, u^{(i)})$ for $i \in \{1, 2\}$ so that the point sets

$$S_i = \{(x, y) \in \text{PD}_i \mid \ell^{(i)} \leq x \leq u^{(i)}\}$$

comprise the *effective sparse holes* which are expected to increase compressibility.

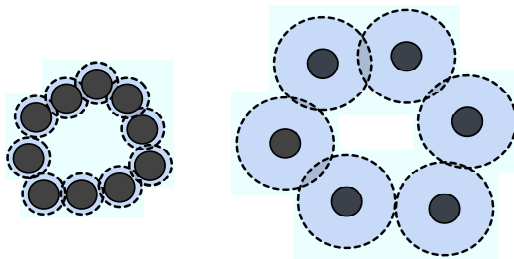


Figure 7: Dense hole (left) and sparse hole (right). Solid balls correspond to the van der Waals radii and dashed balls have radii slightly larger than the birth scale.

We do not introduce similar parameters to control the death scales of holes for the following reasons. The death scale of a hole is closely related to the *size* of that hole (see Figure 8). Introducing an

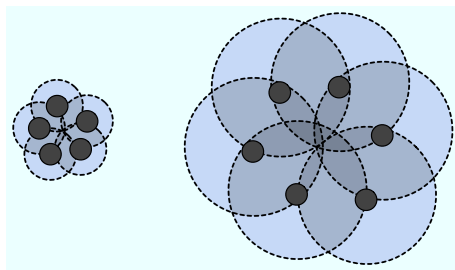


Figure 8: Death radii of holes and their size.

upper bound on the size would remove large holes from consideration and compromise the analysis of

compressibility. On the other hand, a lower bound on death scales is readily available using the existing parameters. Note that the death scale of any hole corresponding to a point in S_i must exceed the value $\ell_i + \delta$.

The final structural feature under consideration is the *length of tunnels*. By a tunnel we mean a cylindrical structure inside a protein molecule. A longer tunnel generically surrounds more space than a shorter one, and so we expect the presence of longer tunnels to contribute towards higher compressibility.

Even though tunnels themselves appear as points in PD_1 , their lengths are not encoded directly into PD_1 . In our analysis, we focus on increasing the radii of balls comprising a given tunnel. This results in the formation of cavities as the expanded walls of the tunnel get pinched together. In general, longer tunnels correspond to more undulant surface regions, and hence to a larger number of cavities generated. This phenomenon is illustrated in Figures 9 and 10.

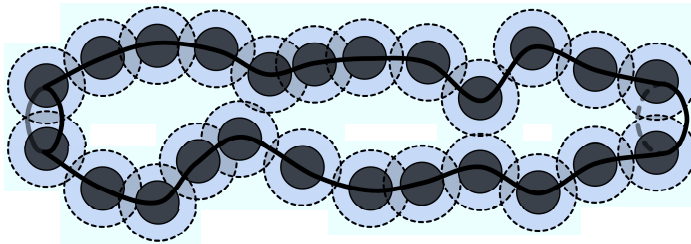


Figure 9: Cross section of a long cylinder

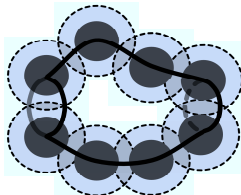


Figure 10: Cross section of a short cylinder

Recall that cavities are points in PD_2 . Thus, the value of interest is the *ratio of cavities to tunnels* in relevant regions of the associated persistence diagrams.

Based on the above explanation, we executed a parameter search in $|\text{PD}_2(\ell^{(2)}, u^{(2)})|/|\text{PD}_1(\ell^{(1)}, u^{(1)})|$ by changing the parameters $\ell^{(i)}, u^{(i)}$, $i = 1, 2$, and δ in order to obtain the best-fitting line in Figure 5. The optimal set of the parameters found by the parameter search is provided in the definition of C_P given by (3.1). In Figure 11, we show the plot the least square errors against the deviation from optimal parameters.

4 Conclusions

Figure 5 clearly indicates that the topological measurement C_P successfully extracts at least some of the essential structural features which determine protein compressibility. We believe that further modifications of C_P by studying physical, chemical and geometric properties of proteins might yield even better fits to experimental compressibility.

Moreover, it is possible to estimate the sizes of holes which have a significant impact on the compressibility of a given protein in the following manner. Pick a point in $\text{PD}_2(5.9, 8.8)$ with the death scale α_d and recall that this point represents a cavity. For simplicity, let us assume that this cavity is constructed

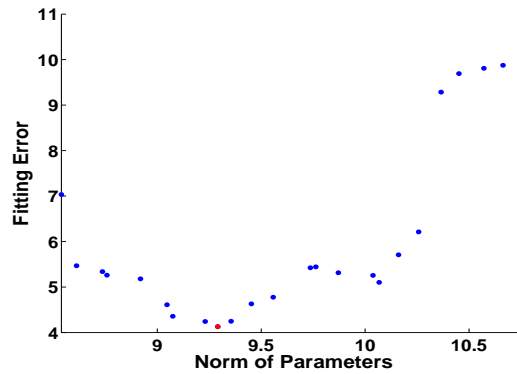


Figure 11: Least squares error against parameter deviation, both in terms of the L^2 norm. The red point indicates the minimal error at the parameters used to define C_P .

by piecewise linear convex regions whose vertices are given by the atoms which have an average van der Waals radius w . At the scale α_d , the union of these van der Waals balls has trivial topology and therefore these balls must have a point of common intersection. Then, we may use equation (2.1) to estimate the diameter of the cavity by $2\sqrt{\alpha_d + w^2}$.

We hope that this estimate will be useful for further understanding the connection between the geometry of a protein molecule and its compressibility. We leave such analysis to future work.

Acknowledgments

The authors would like to thank Fumihide Nouno for valuable discussions. M. G. was partially supported by Fapesp Processo 2010/00875-9 and CNPq Processo 306453/2009-6. Y. H. and S. I. were partially supported by JSPS Grant-in-Aid for Challenging Exploratory Research. M. K., K. M., and V. N. were partially supported by NSF grants DMS-0915019, DMS-1125174, and CBI-0835621 and by contracts from DARPA and AFOSR.

References

- [1] G. Carlsson, Topology and Data, *Bulletin of the Amer. Math. Soc.*, 46(2), 255–308 (2009).
- [2] V. de Silva and R. Ghrist, Homological sensor networks, *Notices Amer. Math. Soc.*, 54(1), 10–17 (2007).
- [3] H. Edelsbrunner, The Union of Balls and Its Dual Shape, *Discrete Comput. Geom.* Vol. 13, 415–440 (1995).
- [4] K. Gekko and Y. Hasegawa, Compressibility-Structure Relationship of Globular Proteins, *Biochemistry*, Vol. 25, 6563-6571 (1986).
- [5] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.
- [6] M. Michael Gromiha and P. K. Ponnuswamy, Relationship Between Amino Acid Properties and Protein Compressibility, *J. Theor. Biol.*, Vol. 165, 87–100 (1993).
- [7] T. Kaczynski, K. Mischaikow, and M. Mrozek, *Computational Homology*, Applied Mathematical Sciences 15, Springer, 2004.

- [8] D. Cohen-Steiner, H. Edelsbrunner and J. Harer, Stability of Persistence Diagrams, *Discrete Comput. Geom.* Vol. 37, 103–120 (2007).
- [9] D. Kharakoz, Protein Compressibility, Dynamics, and Pressure, *Biophysical Journal*, Vol. 79, 511–525 (2000).
- [10] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudharkar and S. Subramaniam, Analytic Shape Computation of Macromolecules I: Molecular Area and Volume Through Alpha Shape, *Proteins: Structure, Function, and Genetics* Vol. 33 , 1–17 (1998).
- [11] K. Mischaikow and V. Nanda, Morse theory for filtrations and efficient computation of persistent homology, in preparation.
- [12] A. Zomorodian and G. Carlsson, Computing Persistent Homology, *Discrete Comput. Geom.*, 33, 249–274 (2005).
- [13] CGAL webpage: <http://www.cgal.org/>
- [14] CHomP webpage: <http://chomp.rutgers.edu/>
- [15] CompTop webpage: <http://comptop.stanford.edu/>
- [16] Perseus webpage: <http://www.math.rutgers.edu/~vidit/perseus.html>
- [17] <http://www.rcsb.org/>

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

- MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata
- MI2008-2 Eiji ONODERA
The initial value problem for a third-order dispersive flow into compact almost Hermitian manifolds
- MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space
- MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme
- MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in infinite extensions over a p -adic field
- MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields
- MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited
- MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition
- MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

- MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE
- MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds
- MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator
- MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials
- MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality
- MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions
- MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings
- MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors
- MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization
- MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions
- MI2009-5 Toshiro HIRANOUCI & Yuichiro TAGUCHI
Flat modules and Groebner bases over truncated discrete valuation rings

- MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations
- MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow
- MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization
- MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity
- MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption
- MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve
- MI2009-12 Tetsu MASUDA
Hypergeometric q -functions of the q -Painlevé system of type $E_8^{(1)}$
- MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination
- MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications
- MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

- MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring
- MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions
- MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force
- MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application
- MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions
- MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations
- MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces
- MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions
- MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map
- MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2 -projection

- MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)
- MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space
- MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme
- MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance
- MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemo-taxis
- MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants
- MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type
- MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect
- MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type $(A_2 + A_1)^{(1)}$
- MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

- MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes
- MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model
- MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes
- MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations
- MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models
- MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons
- MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine
- MI2010-6 Shuichi KAWANO, Toshihiro MISUMI & Sadanori KONISHI
Semi-supervised logistic discrimination via graph-based regularization
- MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation
- MI2010-8 Takahiro ITO
Abstract collision systems on groups
- MI2010-9 Hiroshi YOSHIDA, Kinji KIMURA, Naoki YOSHIDA, Junko TANAKA & Yoshihiro MIWA
An algebraic approach to underdetermined experiments

- MI2010-10 Kei HIROSE & Sadanori KONISHI
Variable selection via the grouped weighted lasso for factor analysis models
- MI2010-11 Katsusuke NABESHIMA & Hiroshi YOSHIDA
Derivation of specific conditions with Comprehensive Groebner Systems
- MI2010-12 Yoshiyuki KAGEI, Yu NAGAFUCHI & Takeshi SUDOU
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Poiseuille type flow
- MI2010-13 Reiichiro KAWAI & Hiroki MASUDA
On simulation of tempered stable random variates
- MI2010-14 Yoshiyasu OZEKI
Non-existence of certain Galois representations with a uniform tame inertia weight
- MI2010-15 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of a Rotating Flow Driven by Precession of Arbitrary Frequency
- MI2010-16 Yu KAWAKAMI & Daisuke NAKAJO
The value distribution of the Gauss map of improper affine spheres
- MI2010-17 Kazunori YASUTAKE
On the classification of rank 2 almost Fano bundles on projective space
- MI2010-18 Toshimitsu TAKAESU
Scaling limits for the system of semi-relativistic particles coupled to a scalar bose field
- MI2010-19 Reiichiro KAWAI & Hiroki MASUDA
Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling
- MI2010-20 Yasuhide FUKUMOTO, Makoto HIROTA & Youichi MIE
Lagrangian approach to weakly nonlinear stability of an elliptical flow

- MI2010-21 Hiroki MASUDA
Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test
- MI2010-22 Toshimitsu TAKAESU
A Generalized Scaling Limit and its Application to the Semi-Relativistic Particles System Coupled to a Bose Field with Removing Ultraviolet Cutoffs
- MI2010-23 Takahiro ITO, Mitsuhiko FUJIO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Composition, union and division of cellular automata on groups
- MI2010-24 Toshimitsu TAKAESU
A Hardy's Uncertainty Principle Lemma in Weak Commutation Relations of Heisenberg-Lie Algebra
- MI2010-25 Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators
- MI2010-26 Reiichiro KAWAI & Hiroki MASUDA
On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling
- MI2010-27 Chikashi ARITA & Daichi YANAGISAWA
Exclusive Queueing Process with Discrete Time
- MI2010-28 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Motion and Bäcklund transformations of discrete plane curves
- MI2010-29 Takanori YASUDA, Masaya YASUDA, Takeshi SHIMOYAMA & Jun KOGURE
On the Number of the Pairing-friendly Curves
- MI2010-30 Chikashi ARITA & Kohei MOTEGI
Spin-spin correlation functions of the q -VBS state of an integer spin model
- MI2010-31 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and spike detection via Gaussian basis expansions

- MI2010-32 Nobutaka NAKAZONO
Hypergeometric τ functions of the q -Painlevé systems of type $(A_2 + A_1)^{(1)}$
- MI2010-33 Yoshiyuki KAGEI
Global existence of solutions to the compressible Navier-Stokes equation around parallel flows
- MI2010-34 Nobushige KUROKAWA, Masato WAKAYAMA & Yoshinori YAMASAKI
Milnor-Selberg zeta functions and zeta regularizations
- MI2010-35 Kissani PERERA & Yoshihiro MIZOGUCHI
Laplacian energy of directed graphs and minimizing maximum outdegree algorithms
- MI2010-36 Takanori YASUDA
CAP representations of inner forms of $Sp(4)$ with respect to Klingen parabolic subgroup
- MI2010-37 Chikashi ARITA & Andreas SCHADSCHNEIDER
Dynamical analysis of the exclusive queueing process
- MI2011-1 Yasuhide FUKUMOTO & Alexander B. SAMOKHIN
Singular electromagnetic modes in an anisotropic medium
- MI2011-2 Hiroki KONDO, Shingo SAITO & Setsuo TANIGUCHI
Asymptotic tail dependence of the normal copula
- MI2011-3 Takehiro HIROTSU, Hiroki KONDO, Shingo SAITO, Takuya SATO, Tatsushi TANAKA & Setsuo TANIGUCHI
Anderson-Darling test and the Malliavin calculus
- MI2011-4 Hiroshi INOUE, Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling via Compressed Sensing
- MI2011-5 Hiroshi INOUE
Implications in Compressed Sensing and the Restricted Isometry Property
- MI2011-6 Daeju KIM & Sadanori KONISHI
Predictive information criterion for nonlinear regression model based on basis expansion methods
- MI2011-7 Shohei TATEISHI, Chiaki KINJYO & Sadanori KONISHI
Group variable selection via relevance vector machine

- MI2011-8 Jan BREZINA & Yoshiyuki KAGEI
Decay properties of solutions to the linearized compressible Navier-Stokes equation around time-periodic parallel flow
Group variable selection via relevance vector machine
- MI2011-9 Chikashi ARITA, Arvind AYYER, Kirone MALLICK & Sylvain PROLHAC
Recursive structures in the multispecies TASEP
- MI2011-10 Kazunori YASUTAKE
On projective space bundle with nef normalized tautological line bundle
- MI2011-11 Hisashi ANDO, Mike HAY, Kenji KAJIWARA & Tetsu MASUDA
An explicit formula for the discrete power function associated with circle patterns of Schramm type
- MI2011-12 Yoshiyuki KAGEI
Asymptotic behavior of solutions to the compressible Navier-Stokes equation around a parallel flow
- MI2011-13 Vladimír CHALUPECKÝ & Adrian MUNTEAN
Semi-discrete finite difference multiscale scheme for a concrete corrosion model: approximation estimates and convergence
- MI2011-14 Jun-ichi INOBUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Explicit solutions to the semi-discrete modified KdV equation and motion of discrete plane curves
- MI2011-15 Hiroshi INOUE
A generalization of restricted isometry property and applications to compressed sensing
- MI2011-16 Yu KAWAKAMI
A ramification theorem for the ratio of canonical forms of flat surfaces in hyperbolic three-space
- MI2011-17 Naoyuki KAMIYAMA
Matroid intersection with priority constraints
- MI2012-1 Kazufumi KIMOTO & Masato WAKAYAMA
Spectrum of non-commutative harmonic oscillators and residual modular forms
- MI2012-2 Hiroki MASUDA
Mighty convergence of the Gaussian quasi-likelihood random fields for ergodic Levy driven SDE observed at high frequency
- MI2012-3 Hiroshi INOUE
A Weak RIP of theory of compressed sensing and LASSO

- MI2012-4 Yasuhide FUKUMOTO & Youich MIE
Hamiltonian bifurcation theory for a rotating flow subject to elliptic straining field
- MI2012-5 Yu KAWAKAMI
On the maximal number of exceptional values of Gauss maps for various classes of surfaces
- MI2012-6 Marcio GAMEIRO, Yasuaki HIRAOKA, Shunsuke IZUMI, Miroslav KRAMAR, Konstantin MISCHAIKOW & Vidit NANDA
Topological Measurement of Protein Compressibility via Persistence Diagrams