Applied Topology based deep learning for Biomolecular Data

Lin Mu[*], Guowei Wei[†]

I Summary of proposed research

Machine learning is a major driving force behind the current data science, while its full potential in unveiling structure-function relationships of biomolecular systems is yet to be reached. Due to the entanglement of geometric complexity and biological complexity in three-dimensional (3D) biomolecular data sets, there is no competitive deep learning algorithm for predicting protein-ligand binding affinities and protein stability changes upon mutation. Persistent homology, a relatively new branch of algebraic topology, has emerged as a new strategy for data analysis. However, current persistent homology does not have a competitive edge in protein classification over other conventional techniques due to its oversimplification of biological information.[11,59] In this whitepaper, we propose element specific persistent homology (ESPH) to retain crucial biological information in the topological simplification. We further integrate ESPH and machine learning to reveal hidden structure-function relationships of biomolecules.

Preliminary studies indicate that the proposed approach significantly outperforms existing methods in the predictions of protein-ligand binding affinities and protein stability changes upon mutation. In fact, the proposed approach has given rise to the best binding free energy prediction in recent world-wide Grand Challenge 2 competition in drug design. *The remaining challenges that hinder the current predictions of protein-ligand binding affinities and mutation impacts are due to the involvement of multiple ligands, membrane-ligand interaction, ligand binding influenced mutations, and membrane influenced mutations.* These challenges are tackled by carefully designed topological learning strategies. The proposed research will dramatically advance the state-of-the-art predictions of protein-ligand binding affinities and protein stability changes upon mutation established by the PIs.

II Significance and innovation

Significance The driving force behind the current transition from qualitative, phenomenological and descriptive biology to quantitative, analytical and predictive biology is theoretical modeling and computational algorithms, which have their roots in mathematics, statistics, and computer science. Indeed, mathematical theories, statistical methods and computer science algorithms underpin quantitative and predictive biological sciences. However, many biophysical models have been built with too much structural or geometric detail and lead to an excessively large number of degrees of freedom, while most topological methods, including persistent homology, suffer from too much reduction in geometric and/or biological information. Element specific persistent homology proposed in this work has the ability to bridge the gap between traditional geometry and classical topology for data analysis. Built upon our recent work on the topological analysis

[*]Email: mul1@ornl.gov. Computer Science and Mathematics Division Oak Ridge National Laboratory, Oak Ridge, TN, USA.

[†]Email: wei@math.msu.edu. Department of Mathematics, Biochemistry and Molecular Biology, and Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA.

of biomolcules,[11–13,68,74,93,95–101] the PIs propose to develop a host of groundbreaking topological learning strategies by integrating element specific persistent homology and machine learning to reveal the hidden structure-function relationships in biomolecules. The establishment of new topological learning strategies has direct applications in the quantitative predictions of biomarkers and their variations, such as mass, concentration, pKa, pH, solubility, partition coefficient, toxicity, solvation free energies, binding affinities of protein-ligand, protein-protein and protein-DNA/RNA interactions, and kinetics. These predictions are crucial to the understanding of biochemical processes in the cell such as gene transcription, translation and regulation, which are core subject of research in the various omics fields such as transcriptomics, proteomics, metabolomics, and particularly structural and functional genomics.

Innovation The novelties of the proposed research are as follows:

First, synergistic fusion of element specific persistent homology and machine learning is introduced to improve the state-of-the-art predictions of protein-ligand binding affinities and protein mutation impacts. Preliminary studies have demonstrated that topological learning strategies outperform all existing methods in the field on massive and diverse data sets.

Second, topology based convolutional deep learning architecture has been proposed for the first time as a novel approach to 3D biomolecular data. This approach will lead to a new paradigm for revealing structure-function relationships from various very large 3D macromolecular data sets. This new formulation has a rich biophysical origin as well as a rigorous mathematical foundation.

Finally, in addition to the promising preliminary results illustrating the power of the proposed new strategy, the PI proposes the development of unique databases and robust online servers for both protein-ligand binding and protein mutation impact predictions. This development provides the assurance for advancing computational biophysics and disseminating the research findings to the general community of biological sciences. In addition to the promising and extensive preliminary results illustrating the power of this new approach, extensive validation and application have been proposed to ensure that this methodology yields robust and powerful tools for molecular biology, computational biophysics, and structural and functional genomics.

It is believed that the proposed topological learning strategies are **transformative:**

First, the proposed topological learning strategy will bring a surge in similar approaches in 3D biomolecular data predictions in the near future.

Second, to reduce the data complexity, new topology based machine learning approaches can be applied to other fields, such as biochemistry, material science and drug design.

Third, the proposed topological learning strategy will open new directions and research topics in abstract/pure/applied mathematics and computer science to explore optimal mathematical abstractions of complex data sets.

References

[1] H. M. Ashtawy and N. R. Mahapatra. A comparative assessment of ranking accuracies of conventional and machine-learning-based scoring functions for protein-ligand binding affinity prediction. *IEEE/ACM Transactions on computational biology and bioinformatics*, 9(5):1301–1313, 2012.

[2] P. J. Ballester. Machine learning scoring functions based on random forest and support vector regression. *Proceedings of the 7th IAPR international conference on Pattern Recognition in Bioinformatics*, pages 14–25, 2012.

[3] P. J. Ballester and J. B. O. Mitchell. A machine learning approach to predicting protein -ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9):1169–1175, 2010.

[4] U. Bauer, M. Kerber, and J. Reininghaus. Distributed computation of persistent homology. *Proceedings of the Sixteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, 2014.

[5] J. Bennett, F. Vivodtzev, and V. Pascucci, editors. *Topological and statistical methods for complex data: Tackling large-scale, high-dimensional and multivariate data spaces*. Mathematics and Visualization. Springer-Verlag Berlin Heidelberg, 2015.

[6] S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. *ACM Computing Surveys*, 40(4):12, 2008.

[7] P. T. Bremer, V. P. I. Hotz, and R. Peikert, editors. *Topological methods in data analysis and visualization III: Theory, algorithms and applications*. Mathematics and Visualization. Springer International Publishing, 2014.

[8] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. *In Proc. of ICML*, pages 89–96, 2005.

[9] C. J. Burges. From RankNet to LambdaRank to LambdaMART: An overview. *Microsoft Research Technical Report*, 82, 2010.

[10] C. J. C. Burges, R. Ragno, and Q. V. Le. Learning to rank with nonsmooth cost functions. *Advances in Neural Information Processing Systems*, 19:193–200, 2007.

[11] Z. X. Cang, L. Mu, K. Wu, K. Opron, K. Xia, and G.-W. Wei. A topological approach to protein classification. *Molecular based Mathematical Biologys*, 3:140–162, 2015.

[12] Z. X.Cang, L. Mu, and G. W. Wei. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. Preprint.

[13] Z. X. Cang and G. W. Wei. Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction . *International Journal for Numerical Methods in Biomedical Engineering*, Accepted, 2017.

[14] Z. X. Cang and G. W. Wei. Analysis and prediction of protein folding energy changes upon mutation by element specific persistent homology. *Bioinformatics*, in press, 2017.

[15] Z. X. Cang and G. W. Wei. TopologyNet: Topology based deep convolutional neural networks for biomolecular property predictions. *PLOS Computational Biology*, in press, 2017.

[16] G. Carlsson. Topology and data. *Am. Math. Soc*, 46(2):255–308, 2009.

[17] G. Carlsson and V. De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.

[18] G. Carlsson, V. de Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. In *Proc. 25th Annu. ACM Sympos. Comput. Geom.*, pages 247–256, 2009.

[19] G. Carlsson, G. Singh, and A. Zomorodian. Computing multidimensional persistence. In *Algorithms and computation*, pages 730–739. Springer, 2009.

[20] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. *Discrete Computational Geometry*, 42(1):71–93, 2009.

[21] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(2):149–187, 2005.

[22] H. W. Chang, S. Bacallado, V. S. Pande, and G. E. Carlsson. Persistent topology and metastable state in conformational dynamics. *PLos ONE*, 8(4):e58699, 2013.

Physical Chemistry Chemical Physics, 10:471–81, 2008.

[23] T. Cheng, X. Li, Y. Li, Z. Liu, and R. Wang. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.*, 49:1079–1093, 2009.

[24] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

[25] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Persistent homology for kernels, images, and cokernels. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA 09, pages 1011–1020, 2009.

[26] G. E. Dahl, N. Jaitly, and R. Salakhutdinov. Multi-task neural networks for qsar predictions. *arXiv preprint arXiv:1406.1231*, 2014.

[27] I. K. Darcy and M. Vazquez. Determining the topology of stable protein-DNA complexes. *Biochemical Society Transactions*, 41:601–605, 2013.

[28] B. DasGupta and J. Liang. *Models and Algorithms for Biomolecules and Molecular Networks*. John Wiley & Sons, 2016.

[29] O. N. A. Demerdash, M. D. Daily, and J. C. Mitchell. Structure-based predictive models for allosteric hot spots. *PLOS Computational Biology*, 5:e1000531, 2009.

[30] R. L. DesJarlais, R. P. Sheridan, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan. Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.*, 29:2149–2153, 1986.

[31] T. K. Dey, F. Fan, and Y. Wang. Computing topological persistence for simplicial maps. In *Proc. 30th Annu. Sympos. Comput. Geom. (SoCG)*, pages 345–354, 2014.

[32] B. Di Fabio and C. Landi. A mayer-vietoris formula for persistent homology with an application to shape recognition in the presence of occlusions. *Foundations of Computational Mathematics*, 11:499–527, 2011.

[33] C. J. Dickson, V. Hornak, C. Velez-Vega, D. J. McKay, J. Reilly, D. A. Sandham, D. Shaw, R. A. Fairhurst, S. J. Charlton, D. A. Sykes, R. A. Pearlstein, and J. S. Jose S. Duca. Uncoupling the structure-activity relationships of $\beta 2$ adrenergic receptor ligands from membrane binding. *Journal of medicinal chemistry*, 2016.

[34] H. Edelsbrunner and J. Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[35] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.

[36] X. Feng, K. Xia, Y. Tong, and G.-W. Wei. Geometric modeling of subcellular structures, organelles and large multiprotein complexes. *International Journal for Numerical Methods in Biomedical Engineering*, 28:1198–1223, 2012.

[37] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[38] P. Frosini and C. Landi. Persistent betti numbers for a noise tolerant shape-based approach to image retrieval. *Pattern Recognition Letters*, 34:863–872, 2013.

[39] I. Fujishiro, Y. Takeshima, T. Azuma, and S. Takahashi. Volume data mining using 3d field topology analysis. *IEEE Computer Graphics and Applications*, 20(5):46–51, 2000.

[40] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda. Topological measurement of protein compressibility via persistence diagrams. *Japan Journal of Industrial and Applied Mathematics*, 32:1–17, 2014.

[41] R. Ghrist. Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.*, 45:61–75, 2008.

[42] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical journal*, 72(3):1047, 1997.

[43] M. K. Gilson and H. X. Zhou. Calculation of protein-ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structur*, 36:21–42, 2007.

[44] D. S. Goodsell and A. J. Olson. Automated docking of substrates to proteins by simulated annealing. *Protein Struct. Funct. Genet.*, 8:195–202, 1990.

[45] D. Greene, W. M. Botello-Smith, A. Follmer, L. Xiao, E. Lambros, and R. Luo. Modeling membrane protein–ligand binding interactions: The human purinergic platelet receptor. *The Journal of Physical Chemistry B*, 120(48):12293–12304, 2016.

[46] C. Heitsch and S. Poznanovic. Combinatorial insights into rna secondary structure, in N. Jonoska and M. Saito, editors. *Discrete and Topological Models in Molecular Biology*, Chapter 7:145–166, 2014.

[47] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[48] D. Horak, S. Maletic, and M. Rajkovic. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, 2009.

[49] S. Y. Huang and X. Zou. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. derivation of interaction potentials. *J. Comput. Chem.*, 27:1865–1875, 2006.

[50] T. B. Hughes, G. P. Miller, and S. J. Swamidass. Modeling epoxidation of drug-like molecules with a deep machine learning network. 2015.

[51] W. Humphrey, A. Dalke, and K. Schulten. VMD – visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[52] W. L. Jorgensen. Rusting of the lock and key model for protein-ligand binding. *Science*, 254:954–955, 1991.

[53] T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational Homology*, volume 157 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[54] P. M. Kasson, A. Zomorodian, S. Park, N. Singhal, L. J. Guibas, and V. S. Pande. Persistent voids a new structural metric for membrane fusion. *Bioinformatics*, 23:1753–1759, 2007.

[55] S. L. Kinnings, N. Liu, P. J. Tonge, R. M. Jackson, L. Xie, and P. E. Bourne. A machine learning based method to improve docking scoring functions and its application to drug repurposing. *Journal of Chemical Information and Model*, 51(2):408–419, 2011.

[56] B. Krishnamoorthy, S. Provan, and A. Tropsha. A topological characterization of protein structure. In *Data Mining in Biomedicine, Springer Optimization and Its Applications*, pages 431–455, 2007.

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[58] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 161:269–288, 1982.

[59] G. Kusano, C. K. Fukumizu, Y. Hiraoka, and A. TOHOKU. Persistence weighted gaussian kernel for topological data analysis. *Statistics*, 1:1–2, 2016.

[60] L. Larini, L. Lu, and G. A. Voth. The multiscale coarse-graining method. vi. implementation of three-body coarse-grained potentials. *Journal of Chemical Physics*, 132(164107), 2010.

[61] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[62] H. Lee, H. Kang, M. K. Chung, B. Kim, and D. S. Lee. Persistent brain network homology from the perspective of dendrogram. *Medical Imaging, IEEE Transactions on*, 31(12):2267–2277, Dec 2012.

[63] G.-B. Li, L.-L. Yang, W.-J. Wang, L.-L. Li, and S.-Y. Yang. ID-Score: A new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *J. Chem. Inf. Model.*, 53(3):592–600, 2013.

[64] H. Li, K. Leung, P. Ballester, and M. H. Wong. iStar: A web platform for large-scale protein-ligand docking. *Plos One*, 9(1), 2014.

[65] H. Li, K.-S. Leung, M. Wong, and P. J. Ballester. Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: Cyscore as a case study. *BMC Bioinformatics*, 15(291), 2014.

[66] H. Li, K.-S. Leung, M.-H. Wong, and P. J. Ballester. Low-Quality Structural and Interaction Data Improves Binding Affinity Prediction via Random Forest. *Molecules*, 20:10947–10962, 2015.

[67] J. Li, P. Mach, and P. Koehl. Measuring the shapes of macromolecules - and why it matters. *Comput Struct Biotechnol J.*, 8:e201309001, 2013.

[68] B. Liu, B. Wang, R. Zhao, Y. Tong, and G. W. Wei. ESES: software for Eulerian solvent excluded surface. *Journal of Computational Chemistry*, 38:446–466, 2017.

[69] X. Liu, Z. Xie, and D. Yi. A fast algorithm for constructing topological structure in large data. *Homology, Homotopy and Applications*, 14:221–238, 2012.

[70] A. Lusci, G. Pollastri, and P. Baldi. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *Journal of chemical information and modeling*, 53(7):1563–1575, 2013.

[71] G. Mate, A. Hofmann, N. Wenzel, and D. W. Heermann. A topological similarity measure for proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838:1180–1190, 2014.

[72] K. Mischaikow and V. Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete and Computational Geometry*, 50(2):330–353, 2013.

[73] L. Mu, J. Wang, G. W. Wei, X. Ye, and S. Zhao. Weak Galerkin method for second order elliptic interface problems. *Journal of Computational Physics*, 250:106 – 125, 2013.

[74] L. Mu, K. Xia, and G. W. Wei. Geometric and electrostatic modeling using molecular rigidity functions. *Journal of Computational and Applied Mathematics*, 313: 18-37, 2017.

[75] V. Nanda. Perseus: the persistent homology software. Software available at http://www.sas.upenn.edu/~vnanda/perseus.

[76] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

[77] P. Niyogi, S. Smale, and S. Weinberger. A topological view of unsupervised learning from noisy data. *SIAM Journal on Computing*, 40:646–663, 2011.

[78] K. Opron, K. L. Xia, and G. W. Wei. Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis. *Journal of Chemical Physics*, 140:234105, 2014.

[79] A. R. Ortiz, M. T. Pisabarro, F. Gago, and R. C. Wade. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem*, 38:2681–2691, 1995.

[80] S. Y. Oudot and D. R. Sheehy. Zigzag Zoology: Rips Zigzags for Homology Inference. In *Proc. 29th Annual Symposium on Computational Geometry*, pages 387–396, June 2013.

[81] D. Pachauri, C. Hinrichs, M. Chung, S. Johnson, and V. Singh. Topology-based kernels with application to inference problems in alzheimer's disease. *Medical Imaging, IEEE Transactions on*, 30(10):1760–1770, Oct 2011.

[82] J. A. Perea, A. Deckard, S. B. Haase, and J. Harer. Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC Bioinformatics*, 16:257, 2015.

[83] J. A. Perea and J. Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15:799–838, 2015.

[84] M. K. Phillips-Jones. Structural and biophysical characterisation of membrane protein-ligand binding (editorial). *Biochimica et Biophysica Acta–Biomembranes.*, 1838(1):1–2, 2013.

[85] V. Robins. Towards computing homology from finite approximations. In *Topology Proceedings*, volume 24, pages 503–532, 1999.

[86] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[87] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[88] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach. Topological analysis of population activity in visual cortex. *Journal of Vision*, 8(8), 2008.

[89] A. Tausz, M. Vejdemo-Johansson, and H. Adams. Javaplex: A research software package for persistent (co)homology. Software available at http://code.google.com/p/javaplex, 2011.

[90] T. Unterthiner, A. Mayr, G. Klambauer, and S. Hochreiter. Toxicity prediction using deep learning. *arXiv preprint arXiv:1503.01445*, 2015.

[91] H. F. G. Velec, H. Gohlke, and G. Klebe. Knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem*, 48:6296–6303, 2005.

[92] I. Wallach, M. Dzamba, and A. Heifets. Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*, 2015.

[93] B. Wang and G. W. Wei. Object-oriented persistent homology. *Journal of Computational Physics*, 305:276–299, 2016.

[94] B. Wang, Z. Zhao, D. D. Nguyen, and G. W. Wei. Feature functional theory - binding predictor (FFT-BP) for the blind prediction of binding free energy. *Theoretical Chemistry Accounts*, 136:55, 2017.

[95] K. L. Xia, X. Feng, Y. Y. Tong, and G. W. Wei. Persistent homology for the quantitative prediction of fullerene stability. *Journal of Computational Chemsitry*, 36:408–422, 2015.

[96] K. L. Xia, Z. Li, L. Mu. Multiscale persistent functions for biomolecular structure characterization. arXiv:1612.08311.

[97] K. L. Xia and G. W. Wei. Persistent homology analysis of protein structure, flexibility and folding. *International Journal for Numerical Methods in Biomedical Engineerings*, 30:814–844, 2014.

[98] K. L. Xia and G. W. Wei. Multidimensional persistence in biomolecular data. *Journal Computational Chemistry*, 36:1502–1520, 2015.

[99] K. L. Xia and G. W. Wei. Persistent topology for cryo-EM data analysis. *International Journal for Numerical Methods in Biomedical Engineering*, 31:e02719, 2015.

[100] K. L. Xia, Z. X. Zhao, and G. W. Wei. Multiresolution persistent homology for excessively large biomolecular datasets. *Journal of Chemical Physics*, 143:134103, 2015.

[101] K. L. Xia, Z. X. Zhao, and G. W. Wei. Multiresolution topological simplification. *Journal Computational Biology*, 22:1–5, 2015.

[102] Y. Yao, J. Sun, X. H. Huang, G. R. Bowman, G. Singh, M. Lesnick, L. J. Guibas, V. S. Pande, and G. Carlsson. Topological methods for exploring low-density states in biomolecular folding pathways. *The Journal of Chemical Physics*, 130:144115, 2009.

[103] S. Yin, L. Biedermannova, J. Vondrasek, and N. V. Dokholyan. Medusascore: An acurate force field-based scoring function for virtual drug screening. *Journal of Chemical Information and Model*, 48:1656–1662, 2008.

[104] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33:249–274, 2005.