bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pyCapsid: Identifying dominant dynamics and quasi-rigid mechanical units in protein shells

Colin Brown[1,2], Anuradha Agarwal[1,3], and Antoni Luque[1,3,4,*]

[1]Viral Information Institute, San Diego State University, 5500 Campanile Drive, San Diego, CA, 92116, USA.

[2]Department of Physics, San Diego State University, 5500 Campanile Drive, San Diego, CA, 92116, USA.

[3]Computational Science Research Center, San Diego State University, 5500 Campanile Drive, San Diego, CA, 92116, USA.

[4]Deparment of Mathematics and Statistics, San Diego State University, 5500 Campanile Drive, San Diego, CA, 92116, USA.

[*]Corresponding author: aluque@sdsu.edu.

Abstract

pyCapsid is a python package developed to facilitate the characterization of the dynamics and mechanical units of protein shells and other protein complexes. The package was developed in response to the rapid increase of high-resolution structures, particularly capsids of viruses, requiring multiscale biophysical analyses. Given a protein shell, pyCapsid generates the collective vibrations of its amino-acid residues, identifies quasi-rigid mechanical regions, and maps the results back to the input proteins for interpretation. pyCapsid's source code is available under MIT License on GitHub (https://github.com/luquelab/pycapsid). It has also been deployed in the two leading python package-management systems, PIP (https://pypi.org/project/pyCapsid/) and Conda (https://anaconda.org/luque_lab/pycapsid). Installation instructions and tutorials are available in the GitHub Page-style online documentation (https://luquelab.github.io/pyCapsid). In addition, users can post issues regarding pyCapsid in the GitHub repository (https://github.com/luquelab/pyCapsid/issues).

Contact: Antoni Luque (aluque@sdsu.edu).

Supplementary information (SI): Further details and figures on the performance results reported in this article are available in the GitHub repository https://github.com/luquelab/pyCapsid/tree/main/results/performance). A gallery displaying the application of pyCapsid to various protein shells is available on the online documentation (https://luquelab.github.io/pyCapsid/gallery/).

1 Introduction

Viruses protect their infective genomes in protein shells called capsids (Twarock and Luque 2019). The number of capsids resolved structurally has increased exponentially in the last two decades, partly thanks to cryo-electron microscopy advances (Callaway 2020; Johnson and Olson 2021; Montiel-Garcia et al. 2021). These three-dimensional reconstructions combined with computational algorithms and complementary experimental techniques are leading to a mechanistic characterization of the assembly,

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

dynamics, and stability of viral capsids, opening the doors to new antiviral strategies (Johnson et al. 2021; Twarock and Stockley 2019; Organtini et al. 2017; Kizziah, Rodenburg, and Dokland 2020; Yeager et al. 1990; Qazi et al. 2018; Li et al. 2008; Mata et al. 2020; de Pablo and San Martín 2022; X. Zhang et al. 2013; Bayfield, Steven, and Antson 2020; Montiel-Garcia et al. 2021; Podgorski et al. 2020; Hua et al. 2017; Mohajerani et al. 2022; Wilson and Roof 2021; Bruinsma, Wuite, and Roos 2021; Luque and Reguera 2013; Lee et al. 2022; Grime et al. 2016; Plavec et al. 2021). Among the computational methods, molecular dynamics algorithms have improved dramatically in the last decades and can infer the dynamics of large protein complexes. However, they resolve relatively short timescales ($\lesssim 1$ μs) and require specialized computational resources (Jana and May 2021; Hadden et al. 2018; Perilla and Schulten 2017; Bryer et al. 2022). The fact that capsids are assembled from 60 to more than 60,000 proteins further limits the application of molecular dynamics (Twarock and Luque 2019; Luque et al. 2020; Berg and Roux 2021). Alternatively, the combination of normal mode analysis (NMA), molecular coarse-graining, and elastic network models (ENM) offers a more scalable solution (Bahar et al. 2010; Romo and Grossfield 2011a). This approach has successfully estimated the collective motion of proteins in complexes (Bahar et al. 2010) and identified structural conformational changes in capsids (Tama and Brooks 2005). Nonetheless, no easy-to-use computational packages are currently available to characterize the dynamics and mechanical properties of protein shells. The bioinformatics software presented here, pyCapsid, aims to address this issue.

pyCapsid is inspired by prior publications that applied ENM, NMA, and clustering methods to three-dimensional reconstructions to extract the quasi-rigid regions associated with mechanically relevant units in protein shells (Ponzoni et al. 2015; Polles et al. 2013). These methods are very insightful, but their implementation in packages that generate the dynamics of protein complexes, such as NRGTEN (Mailhot and Najmanovich 2021), ClustENMD (Kaynak et al. 2021), WebPSN (Seeber et al. 2015), or the popular ProDy ProDy (S. Zhang et al. 2021), is not trivial. To address this issue, we introduce pyCapsid, an accessible Python package that identifies the dominant dynamics and quasi-rigid regions of protein shells. Moreover, since the methods are generic, pyCapsid can also be applied to other protein complexes. Yet, in this first release of pyCapsid, we have focused on the characterization of protein shells, such as viral capsids, cellular protein compartments like encapsulins, and gene-transfer agents (Giessen et al. 2019; Bárdy et al. 2020; Johnson and Olson 2021; Montiel-Garcia et al. 2021).

2 Methods and Features

The python package pyCapsid is divided into five independent modules: The PDB (protein data bank) module, the CG (coarse-graining module), the NMA (normal mode analysis) module, the QRC (quasi-

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rigid clustering) module, and the VIS (visualization) module (Figure 1). The role and technical aspects of each module are briefly described below.

PDB module. This module retrieves and loads structural data from the Protein Data Bank (using the PDB ID) or a local file in PDB or PDBx/mmCIF formats (Westbrook et al. 2022; Berman et al. 2000). The PDB module builds on functions from the Python packages Biotite (Kunzmann and Hamacher 2018).

CG module. This module coarse-grains the proteins at the amino-acid level and establishes an elastic force field between amino acids, offering four standard elastic models' options: The anisotropic network model (ANM), the gaussian network model (GNM), the unified anisotropic and gaussian network model (U-ENM), which is the default, and the backbone-enhanced elastic network model (bb-ENM). Each amino acid is coarse-grained as a point mass on the alpha-carbon. The links in the network connect amino acids that are closer than a certain threshold. The default value is 15Å for ANM and 7.5Å for the other models. These values are based on prior studies of elastic models reproducing empirical molecular thermal fluctuations (B-factors) (Eyal, Yang, and Bahar 2006; Zheng 2008; Micheletti, Carloni, and Maritan 2004; Romo and Grossfield 2011b). The small threshold distance leads to a sparse network. This network, combined with the elastic strength values of the elastic model, defines the Hessian matrix. Given that the network is sparse, the calculations to build the matrix are accelerated using Numba (Lam, Pitrou, and Seibert 2015).

NMA module. This module obtains the motions of the macromolecular complex by decomposing the dynamics into independent sinusoidal motions called normal modes (Goldstein, Poole, and Safko 2002). The normal modes and associated frequencies are obtained from the Hessian matrix derived in the CG module. It is well established that only low-frequency modes are relevant to the global dynamics of macromolecules (Bahar et al. 2010). The default number of modes calculated in pyCapsid is 200. This number was selected by comparing the results with simulations using a larger number of modes (as many modes as one-hundredth of the number of residues, that is, 1000 modes for a structure containing 100,000 residues). pyCapsid also provides an optional dependency to accelerate the calculations further in GPUs using CUDA via solvers in the cupy package (Okuta et al. 2017).

QRC module. This module estimates the amino acids that tend to fluctuate as a single mechanical unit (quasi-rigid cluster) using the Spectrus algorithm (Ponzoni et al. 2015). A cluster contains groups of residues that minimize the distance fluctuations between residues. The default clustering method in pyCapsid is the default discretize method from scikit-learn and offers k-means clustering as an alternative option (Pedregosa et al. 2012; Yu and Shi 2003). pyCapsid explores from four to n_cluster_max number of clusters, which the user sets.

VIS module. The results obtained from pyCapsid are stored as data files, figures, and movies in the same running folder. pyCapsids' online tutorial (https://luquelab.github.io/pyCapsid/tutorial/)

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

provides instructions and scripts to visualize the results in two popular molecular visualization tools NGLview and ChimeraX (Nguyen, Case, and Rose 2018; Pettersen et al. 2021).

3 Applications

pyCapsid's performance and accuracy were obtained for 25 protein shells on an HPC cluster core with Intel Xeon CPU E5-2650 v4 (2.20 GHz) and 128 GB of RAM. The size of the protein shells ranged from 16,000 to over 400,000 amino acid residues. The peak memory usage ranged from 800 MB to 90 GB and increased with the number of residues following a power law (exponent = 1.46±0.06 and R^2 = 0.97). The runtime ranged from 2 minutes to 36 hours and increased with the number of residues following a power law (exponent = 2.20±0.10 and R^2 = 0.95). The accuracy was estimated from the correlation coefficient between the simulated and empirical thermal motions (B-factors) of the amino acids. pyCapsid's accuracy ranged from 0.10 to 0.88 (out of 1) and decreased linearly for structures with lower experimental resolution (slope = –0.20 ± 0.05 1/Å and R^2 = 0.40). The regression projected a perfect accuracy for structures with an ideal experimental resolution of 0 Å (intercept = 1.23 ± 0.18). The accuracy was independent of the number of residues (Spearman's coefficient = –0.09 and p-value = 0.66). Four additional small capsids (PDB IDs 2ms2, 1za7, 1a34, and 3nap) were analyzed, obtaining quasi-rigid domain decompositions consistent with similar analyses of capsids published previously (Polles et al. 2013; Ponzoni et al. 2015). A subset of 10 of the 25 protein shells in the original performance analysis was used to compare the speed of pyCapsid with ProDy, from loading the PDB to generating the normal modes analysis (NMA). pyCapsid displayed an average speed increase of 3.0±1.5 that was independent of capsid size (Spearman's coefficient = 0.11 and p-value = 0.76). This increase in speed was due to the use of Numba and the invert shift mode in SciPy. However, this led to a similar increase in memory usage. The PDB entries, statistical methods, and figures are available in the Supplementary Information.

4. Concluding Remarks

pyCapsid can generate the collective motion and extract the quasi-rigid functional regions of protein shells and other protein complexes. The underlying algorithm of pyCapsid generates the dynamical modes faster than popular protein dynamics packages, like ProDy, which facilitates the quasi-rigid domain decomposition of large complexes like protein shells in the range of minutes to over a day for standard computers. The computational efficiency of pyCapsid, combined with its accessibility via Python distribution packages and online tutorials, aims to facilitate its adoption among researchers interested in physical virology, structural bioinformatics, and related fields.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments. The authors thank the insights from the SDSU Biomath Group, particularly professors Arlette Baljon and Parag Katira. The authors also thank the SDSU Computational Research Science Center for the HPC used to test and benchmark pyCapsid.

Funding information. The authors' research was supported by the National Science Foundation (Award #1951678) and the Gordon and Betty Moore Foundation (Award #GBMF9871, https://doi.org/10.37807/GBMF9871). The HPC facilities were supported by the NSF Office of Advanced Cyberinfrastructure grant 1659169.

References

Bahar, Ivet, Timothy R. Lezon, Lee-Wei Yang, and Eran Eyal. 2010. 'Global Dynamics of Proteins: Bridging Between Structure and Function'. *Annual Review of Biophysics* 39 (1): 23–42. https://doi.org/10.1146/annurev.biophys.093008.131258.

Bárdy, Pavol, Tibor Füzik, Dominik Hrebík, Roman Pantůček, J. Thomas Beatty, and Pavel Plevka. 2020. 'Structure and Mechanism of DNA Delivery of a Gene Transfer Agent'. *Nature Communications* 11 (1): 3034. https://doi.org/10.1038/s41467-020-16669-9.

Bayfield, Oliver W, Alasdair C Steven, and Alfred A Antson. 2020. 'Cryo-EM Structure in Situ Reveals a Molecular Switch That Safeguards Virus against Genome Loss'. *ELife* 9 (April). https://doi.org/10.7554/eLife.55517.

Berg, Maureen, and Simon Roux. 2021. 'Extreme Dimensions — How Big (or Small) Can Tailed Phages Be?' *Nature Reviews Microbiology* 19 (7): 407–407. https://doi.org/10.1038/s41579-021-00574-z.

Berman, Helen M., John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. 2000. 'The Protein Data Bank'. *Nucleic Acids Res.* 28 (1): 235–42. https://doi.org/10.1093/nar/28.1.235.

Bruinsma, Robijn F., Gijs J. L. Wuite, and Wouter H. Roos. 2021. 'Physics of Viral Dynamics'. *Nature Reviews Physics* 3 (2): 76–91. https://doi.org/10.1038/s42254-020-00267-1.

Bryer, Alexander J., Tyler Reddy, Edward Lyman, and Juan R. Perilla. 2022. 'Full Scale Structural, Mechanical and Dynamical Properties of HIV-1 Liposomes'. *PLOS Computational Biology* 18 (1): e1009781. https://doi.org/10.1371/journal.pcbi.1009781.

Callaway, Ewen. 2020. 'Revolutionary Cryo-EM Is Taking over Structural Biology'. *Nature* 578 (7794): 201–201. https://doi.org/10.1038/d41586-020-00341-9.

Eyal, E, L.-W. Yang, and I Bahar. 2006. 'Anisotropic Network Model: Systematic Evaluation and a New Web Interface'. *Bioinformatics* 22 (21): 2619–27. https://doi.org/10.1093/bioinformatics/btl448.

Giessen, Tobias W, Benjamin J Orlando, Andrew A Verdegaal, Melissa G Chambers, Jules Gardener, David C Bell, Gabriel Birrane, Maofu Liao, and Pamela A Silver. 2019. 'Large Protein Organelles Form a New Iron Sequestration System with High Storage Capacity'. *ELife* 8 (July). https://doi.org/10.7554/eLife.46070.

Goldstein, Herbert, Charles Poole, and John Safko. 2002. 'Classical Mechanics'. American Association of Physics Teachers.

Grime, John M. A., James F. Dama, Barbie K. Ganser-Pornillos, Cora L. Woodward, Grant J. Jensen, Mark Yeager, and Gregory A. Voth. 2016. 'Coarse-Grained Simulation Reveals Key Features of HIV-1 Capsid Self-Assembly'. *Nature Communications* 7 (1): 11568. https://doi.org/10.1038/ncomms11568.

Hadden, Jodi A, Juan R Perilla, Christopher John Schlicksup, Balasubramanian Venkatakrishnan, Adam Zlotnick, and Klaus Schulten. 2018. 'All-Atom Molecular Dynamics of the HBV Capsid Reveals

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Insights into Biological Function and Cryo-EM Resolution Limits'. *ELife* 7 (April). https://doi.org/10.7554/eLife.32478.

Hua, Jianfei, Alexis Huet, Carlos A. Lopez, Katerina Toropova, Welkin H. Pope, Robert L. Duda, Roger W. Hendrix, and James F. Conway. 2017. 'Capsids and Genomes of Jumbo-Sized Bacteriophages Reveal the Evolutionary Reach of the HK97 Fold'. *MBio* 8 (5). https://doi.org/10.1128/mBio.01579-17.

Jana, Asis K., and Eric R. May. 2021. 'Atomistic Dynamics of a Viral Infection Process: Release of Membrane Lytic Peptides from a Non-Enveloped Virus'. *Science Advances* 7 (16). https://doi.org/10.1126/sciadv.abe1761.

Johnson, John E., Tatiana Domitrovic, Tsutomu Matsui, Roger Castells-Graells, and George Lomonossoff. 2021. 'Dynamics and Stability in the Maturation of a Eukaryotic Virus: A Paradigm for Chemically Programmed Large-Scale Macromolecular Reorganization'. *Archives of Virology* 166 (6): 1547–63. https://doi.org/10.1007/s00705-021-05007-z.

Johnson, John E., and Arthur J. Olson. 2021. 'Icosahedral Virus Structures and the Protein Data Bank'. *Journal of Biological Chemistry* 296 (January): 100554. https://doi.org/10.1016/j.jbc.2021.100554.

Kaynak, Burak T, She Zhang, Ivet Bahar, and Pemra Doruker. 2021. 'ClustENMD: Efficient Sampling of Biomolecular Conformational Space at Atomic Resolution'. *Bioinformatics* 37 (21): 3956–58. https://doi.org/10.1093/bioinformatics/btab496.

Kizziah, James L., Cynthia M. Rodenburg, and Terje Dokland. 2020. 'Structure of the Capsid Size-Determining Scaffold of "Satellite" Bacteriophage P4'. *Viruses* 12 (9): 953. https://doi.org/10.3390/v12090953.

Kunzmann, Patrick, and Kay Hamacher. 2018. 'Biotite: A Unifying Open Source Computational Biology Framework in Python'. *BMC Bioinformatics* 19 (1): 346. https://doi.org/10.1186/s12859-018-2367-z.

Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert. 2015. 'Numba: A LLVM-Based Python JIT Compiler'. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2833157.2833162.

Lee, Diana Y., Caitlin Bartels, Katelyn McNair, Robert A. Edwards, Manal A. Swairjo, and Antoni Luque. 2022. 'Predicting the Capsid Architecture of Phages from Metagenomic Data'. *Computational and Structural Biotechnology Journal* 20: 721–32. https://doi.org/10.1016/j.csbj.2021.12.032.

Li, Long, Shee-Mei Lok, I-Mei Yu, Ying Zhang, Richard J. Kuhn, Jue Chen, and Michael G. Rossmann. 2008. 'The Flavivirus Precursor Membrane-Envelope Protein Complex: Structure and Maturation'. *Science* 319 (5871): 1830–34. https://doi.org/10.1126/science.1153263.

Luque, Antoni, Sean Benler, Diana Y Lee, Colin Brown, and Simon White. 2020. 'The Missing Tailed Phages: Prediction of Small Capsid Candidates'. *Microorganisms* 8 (12): 1944.

Luque, Antoni, and David Reguera. 2013. 'Theoretical Studies on Assembly, Physical Stability and Dynamics of Viruses'. In *Mateu, M. (Eds) Structure and Physics of Viruses*, 68:553–95. Subcellular Biochemistry. https://doi.org/10.1007/978-94-007-6552-8_19.

Mailhot, Olivier, and Rafael Najmanovich. 2021. 'The NRGTEN Python Package: An Extensible Toolkit for Coarse-Grained Normal Mode Analysis of Proteins, Nucleic Acids, Small Molecules and Their Complexes'. *Bioinformatics* 37 (19): 3369–71. https://doi.org/10.1093/bioinformatics/btab189.

Mata, Carlos P., Javier M. Rodríguez, Nobuhiro Suzuki, and José R. Castón. 2020. 'Structure and Assembly of Double-Stranded RNA Mycoviruses'. In , 213–47. https://doi.org/10.1016/bs.aivir.2020.08.001.

Micheletti, Cristian, Paolo Carloni, and Amos Maritan. 2004. 'Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models'. *Proteins: Structure, Function, and Bioinformatics* 55 (3): 635–45. https://doi.org/10.1002/prot.20049.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mohajerani, Farzaneh, Botond Tyukodi, Christopher J. Schlicksup, Jodi A. Hadden-Perilla, Adam Zlotnick, and Michael F. Hagan. 2022. 'Multiscale Modeling of Hepatitis B Virus Capsid Assembly and Its Dimorphism'. *ACS Nano* 16 (9): 13845–59. https://doi.org/10.1021/acsnano.2c02119.

Montiel-Garcia, Daniel, Nelly Santoyo-Rivera, Phuong Ho, Mauricio Carrillo-Tripp, Charles L Brooks III, John E Johnson, and Vijay S Reddy. 2021. 'VIPERdb v3.0: A Structure-Based Data Analytics Platform for Viral Capsids'. *Nucleic Acids Research* 49 (D1): D809–16. https://doi.org/10.1093/nar/gkaa1096.

Nguyen, Hai, David A Case, and Alexander S Rose. 2018. 'NGLview–Interactive Molecular Graphics for Jupyter Notebooks'. *Bioinformatics* 34 (7): 1241–42. https://doi.org/10.1093/bioinformatics/btx789.

Okuta, Ryosuke, Yuya Unno, Daisuke Nishino, Shohei Hido, and Crissman Loomis. 2017. 'CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations'. *Workshop on Machine Learning Systems (LearningSys) at Neural Information Processing Systems (NIPS)*, no. Nips.

Organtini, Lindsey J., Kristin L. Shingler, Robert E. Ashley, Elizabeth A. Capaldi, Kulsoom Durrani, Kelly A. Dryden, Alexander M. Makhov, James F. Conway, Marie C. Pizzorno, and Susan Hafenstein. 2017. 'Honey Bee Deformed Wing Virus Structures Reveal That Conformational Changes Accompany Genome Release'. *Journal of Virology* 91 (2). https://doi.org/10.1128/JVI.01795-16.

Pablo, Pedro J de, and Carmen San Martín. 2022. 'Seeing and Touching Adenovirus: Complementary Approaches for Understanding Assembly and Disassembly of a Complex Virion'. *Current Opinion in Virology* 52 (February): 112–22. https://doi.org/10.1016/j.coviro.2021.11.006.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2012. 'Scikit-Learn: Machine Learning in Python', January. http://arxiv.org/abs/1201.0490.

Perilla, Juan R., and Klaus Schulten. 2017. 'Physical Properties of the HIV-1 Capsid from All-Atom Molecular Dynamics Simulations'. *Nature Communications* 8 (1): 15959. https://doi.org/10.1038/ncomms15959.

Pettersen, Eric F, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, Tristan I Croll, John H Morris, and Thomas E Ferrin. 2021. 'UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers'. *Protein Science* 30 (1): 70–82.

Plavec, Zlatka, Ina Pöhner, Antti Poso, and Sarah J Butcher. 2021. 'Virus Structure and Structure-Based Antivirals'. *Current Opinion in Virology* 51 (December): 16–24. https://doi.org/10.1016/j.coviro.2021.09.005.

Podgorski, Jennifer, Joshua Calabrese, Lauren Alexandrescu, Deborah Jacobs-Sera, Welkin Pope, Graham Hatfull, and Simon White. 2020. 'Structures of Three Actinobacteriophage Capsids: Roles of Symmetry and Accessory Proteins'. *Viruses* 12 (3): 294. https://doi.org/10.3390/v12030294.

Polles, Guido, Giuliana Indelicato, Raffaello Potestio, Paolo Cermelli, Reidun Twarock, and Cristian Micheletti. 2013. 'Mechanical and Assembly Units of Viral Capsids Identified via Quasi-Rigid Domain Decomposition'. *PLoS Computational Biology* 9 (11). https://doi.org/10.1371/journal.pcbi.1003331.

Ponzoni, Luca, Guido Polles, Vincenzo Carnevale, and Cristian Micheletti. 2015. 'SPECTRUS: A Dimensionality Reduction Approach for Identifying Dynamical Domains in Protein Complexes from Limited Structural Datasets'. *Structure* 23 (8): 1516–25. https://doi.org/10.1016/j.str.2015.05.022.

Qazi, Shefah, Christopher J. Schlicksup, Jonathan Rittichier, Michael S. VanNieuwenhze, and Adam Zlotnick. 2018. 'An Assembly-Activating Site in the Hepatitis B Virus Capsid Protein Can Also Trigger Disassembly'. *ACS Chemical Biology* 13 (8): 2114–20. https://doi.org/10.1021/acschembio.8b00283.

Romo, Tod D., and Alan Grossfield. 2011a. 'Validating and Improving Elastic Network Models with Molecular Dynamics Simulations'. *Proteins: Structure, Function, and Bioinformatics* 79 (1): 23–34. https://doi.org/10.1002/prot.22855.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

———. 2011b. 'Validating and Improving Elastic Network Models with Molecular Dynamics Simulations'. *Proteins: Structure, Function, and Bioinformatics* 79 (1): 23–34. https://doi.org/10.1002/prot.22855.

Seeber, Michele, Angelo Felline, Francesco Raimondi, Simona Mariani, and Francesca Fanelli. 2015. 'WebPSN: A Web Server for High-Throughput Investigation of Structural Communication in Biomacromolecules'. *Bioinformatics* 31 (5): 779–81. https://doi.org/10.1093/bioinformatics/btu718.

Tama, Florence, and Charles L. Brooks. 2005. 'Diversity and Identity of Mechanical Properties of Icosahedral Viral Capsids Studied with Elastic Network Normal Mode Analysis'. *Journal of Molecular Biology* 345 (2): 299–314. https://doi.org/10.1016/j.jmb.2004.10.054.

Twarock, Reidun, and Antoni Luque. 2019. 'Structural Puzzles in Virology Solved with an Overarching Icosahedral Design Principle'. *Nature Communications* 10 (1): 1–9. https://doi.org/10.1038/s41467-019-12367-3.

Twarock, Reidun, and Peter G. Stockley. 2019. 'RNA-Mediated Virus Assembly: Mechanisms and Consequences for Viral Evolution and Therapy'. *Annual Review of Biophysics* 48 (1): 495–514. https://doi.org/10.1146/annurev-biophys-052118-115611.

Westbrook, John D., Jasmine Y. Young, Chenghua Shao, Zukang Feng, Vladimir Guranovic, Catherine L. Lawson, Brinda Vallat, et al. 2022. 'PDBx/MmCIF Ecosystem: Foundational Semantic Tools for Structural Biology'. *Journal of Molecular Biology* 434 (11): 167599. https://doi.org/10.1016/J.JMB.2022.167599.

Wilson, David P., and Danielle A. Roof. 2021. 'Viral Phrenology'. *Viruses* 13 (11): 2191. https://doi.org/10.3390/v13112191.

Yeager, M, K A Dryden, N H Olson, H B Greenberg, and T S Baker. 1990. 'Three-Dimensional Structure of Rhesus Rotavirus by Cryoelectron Microscopy and Image Reconstruction.' *Journal of Cell Biology* 110 (6): 2133–44. https://doi.org/10.1083/jcb.110.6.2133.

Yu, Stella X, and Jianbo Shi. 2003. 'Multiclass Spectral Clustering'. *Proceedings Ninth IEEE International Conference on Computer Vision*, 313–19 vol.1.

Zhang, She, James M Krieger, Yan Zhang, Cihan Kaya, Burak Kaynak, Karolina Mikulska-Ruminska, Pemra Doruker, Hongchun Li, and Ivet Bahar. 2021. 'ProDy 2.0: Increased Scale and Scope after 10 Years of Protein Dynamics Modelling with Python'. *Bioinformatics* 37 (20): 3657–59. https://doi.org/10.1093/bioinformatics/btab187.

Zhang, Xing, Huatao Guo, Lei Jin, Elizabeth Czornyj, Asher Hodes, Wong H Hui, Angela W Nieh, Jeff F Miller, and Z Hong Zhou. 2013. 'A New Topology of the HK97-like Fold Revealed in Bordetella Bacteriophage by CryoEM at 3.5 Å Resolution'. *ELife* 2 (December). https://doi.org/10.7554/eLife.01299.

Zheng, Wenjun. 2008. 'A Unification of the Elastic Network Model and the Gaussian Network Model for Optimal Description of Protein Conformational Motions and Fluctuations'. *Biophysical Journal* 94 (10): 3853–57. https://doi.org/10.1529/biophysj.107.125831.

bioRxiv preprint doi: https://doi.org/10.1101/2023.02.27.529640; this version posted March 1, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures

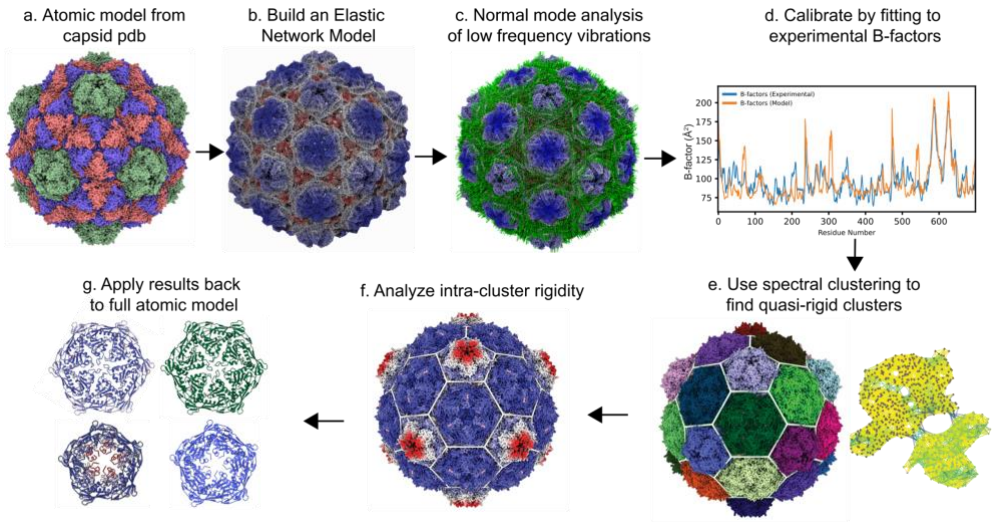
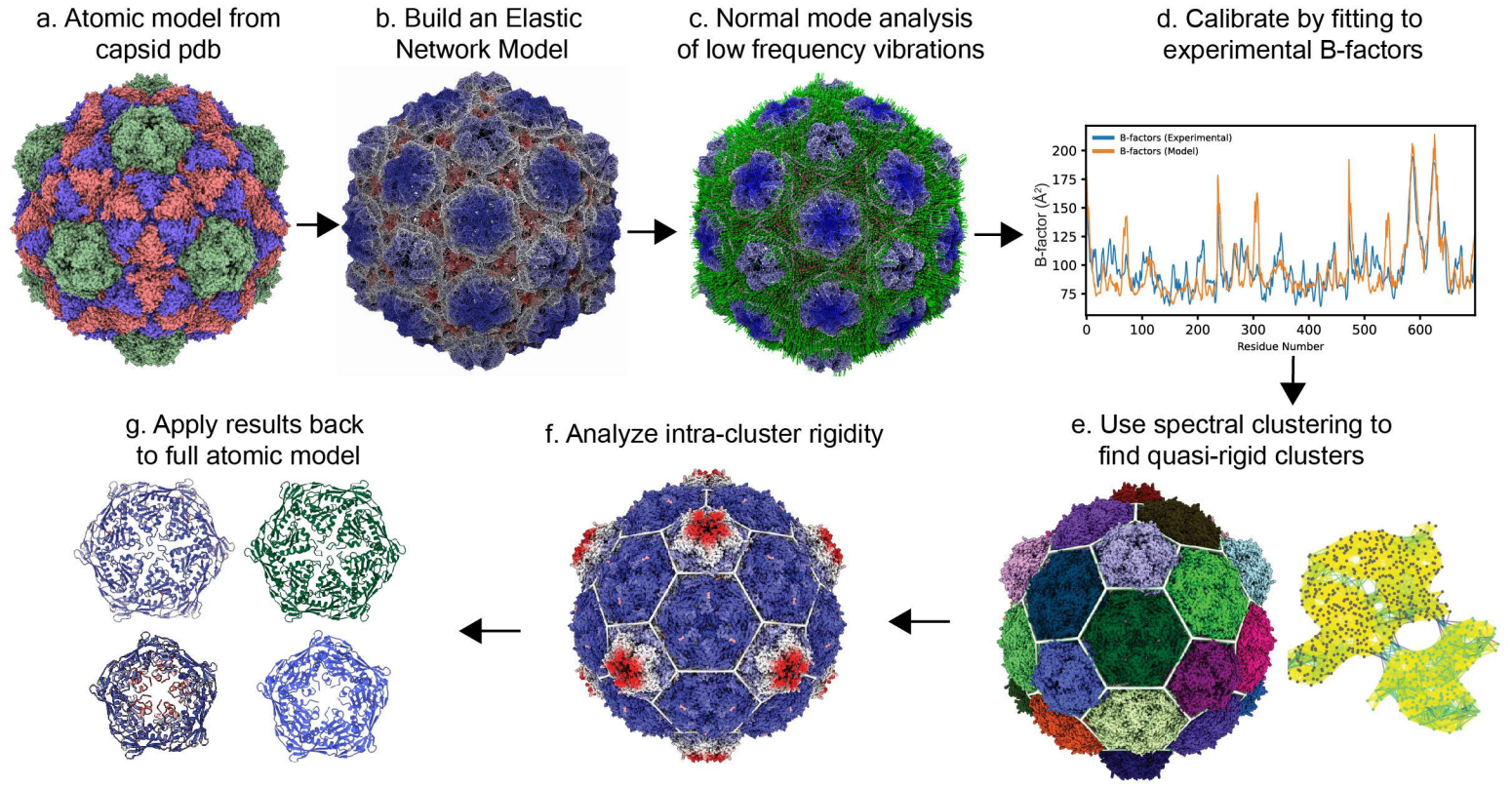


Figure 1. pyCapsid's key steps. a. The protein shell (atom coordinates, atom types, and B-factors) is inputted using a PDB file (or a PDB ID). b. The elastic network model (ENM) is calibrated (parametrized). c. Normal mode analysis (NMA) determines the distance fluctuations between residues. d. The correlation coefficient of predicted and experimental B-factors is generated for quality control. e. Spectral clustering is applied to obtain the quasi-rigid molecular domains in the capsid. f. The fluctuations between residues within the rigid clusters are calculated. g. The results are mapped back to the capsid protein atomic model for structural interpretation.


a. Atomic model from capsid pdb
b. Build an Elastic Network Model
c. Normal mode analysis of low frequency vibrations
d. Calibrate by fitting to experimental B-factors
B-factors (Experimental)
B-factors (Model)
B-factor (Å²)
200
175
150
125
100
75
0
100
200
300
400
500
600
Residue Number
e. Use spectral clustering to find quasi-rigid clusters
f. Analyze intra-cluster rigidity
g. Apply results back to full atomic model