Robustness Measure for an Adeno-Associated Viral Shell Self-Assembly is accurately predicted by Configuration Space Atlasing using EASAL *

Ruijin Wu
Department of Computer and Information Science and Engineering
University of Florida
ruijin@cise.ufl.edu

Aysegul Ozkan
Department of Computer and Information Science and Engineering
University of Florida
aozkan@cise.ufl.edu

Antonette Bennett
Department of Biochemistry and Molecular Biology
University of Florida
dendena@ufl.edu

Mavis Agbandje-Mckenna
Department of Biochemistry and Molecular Biology
University of Florida
mckenna@ufl.edu

Meera Sitharam
Department of Computer and Information Science and Engineering
University of Florida
sitharam@cise.ufl.edu

ABSTRACT

Viral shell assembly occurs spontaneously in solution, caused by weak inter-atomic interactions between the identical coat-protein monomers that constitute the shell. A sound measure of robustness of the assembly process is therefore based on determining the weak interactions whose disruption significantly disrupts successful assembly.

We used data isolated from the X-ray structure of a T=1, Adeno-Associated Virus as input to a new software suite of algorithms, EASAL, (Efficient Atlasing and Search of Assembly Landscapes, developed by 3 of the authors). Since assembly is entropically driven, we predicted and ranked the crucial interactions for assembly, purely by focusing on key changes in the geometry of the assembly configuration space when specific interactions are dropped. Using the same data for mutagenesis experiments towards assembly disruption, the 2 other authors experimentally verified these predictions successfully.

This paper briefly describes the problem background for supramolecular assembly, the key features and novelty of EASAL, the geometric features of the configuration space chosen to measure robustness of assembly, the process by which the input data for EASAL is extracted from X-ray crystallography viral structure data, and a validation of EASAL's robustness predictions with mutagenesis results for the specific case of the AAV-2 virus. Extensive validation using other viruses is underway.

Categories and Subject Descriptors

F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and problems—*Computations on polynomials*; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Geometrical problems and computations*; J.2 [**Physical Sciences and Engineering**]: Physics,Chemistry; J.3 [**Life and Medical Sciences**]: Biology

General Terms

Theory, Algorithms, Design, Performance, Verification

Keywords

Virus Assembly, Supramolecular Assembly, Robustness, Molecular, Configuration Space, Free Energy Landscape, Configurational Entropy, Algorithmic Performance Guarantees, Atlas, Geometric Constraints, Optimal Parameterization, Cayley Parameters, Convex Programming, Distance Geometry, Structural Rigidity, Stratification, Semi-Algebraic sets

1. INTRODUCTION

1.1 Motivation

Supramolecular self-assemblies occur spontaneously and widely in nature and are increasingly important in healthcare and engineering. Yet due to their rapidity, they are poorly understood. Viral shell self-assembly is a necessary part of the viral life-cycle. Understanding it illuminates the pathophysiology of infectious disease, and could help in engineering viral capsids - used, for example, for gene therapy and other medical applications.

1.2 Factors influencing a Supramolecular Assembly process

*This research was supported in part by NSF Mathematical Biology Grant DMS0714912, a University of Florida computational biology seed grant and NSF CCF-1117695.

ACM-BCB'12 October 7-10, 2012, Orlando, FL, USA

We focus here on icosahedral T=1 viruses (60 identical coat protein monomers) that autonomously assemble into empty shells (not necessarily enclosing genomic material, and without the aid of chaperones or scaffolding proteins). The self-assembly process starts from the monomers and culminates in a specific *successful molecular assembly configuration* with remarkable efficacy and robustness.

The assembly process is caused and governed by (a) weak, inter-atomic forces between and within monomers - a monomer is viewed as a small collection of *rigid motifs* (rigid configurations of points representing atom centers) that are tethered together; and (b) implicit solvent effects, lumping together all other relevant weak forces.

The efficacy of assembly is largely a consequence of the geometric structure of the equilibrium free-energy landscape of the assembly configuration space, and is determined by the factors below.

(1) Potential Energy. The first factor is the depth of the potential energy *well* containing the successful molecular assembly configuration (the potential energy contributions come from relevant inter-atomic interactions mentioned above).

(2) Entropy. The second factor is the volume of the potential energy *basin* consisting of all possible assembly configurations that lead to the successful assembly at equilibrium. This, in turn, depends on 2 factors:
(2a) Interface Configurational Entropy of various intermolecular interface configuration spaces occuring in various intermediate subassemblies that are present in a successful assembly configuration.
(2b) Combinatorial Entropy, i.e, the number of possible ways in which a successful assembly configuration can be recursively decomposed (reverse of assembled) into subassembly intermediates down to the rigid motifs in the monomers [28, 29, 5, 6]; this is largely determined by the combinatorics and symmetry of the overall successful assembly configuration.

(3) Kinetics. The third factor influencing the efficacy of assembly is the rates of formation and concentrations of the various assembly intermediates from their constituent subassemblies.

1.3 A Measure of Robustness

The robustness of the assembly process can be viewed as the (in)sensitivity - of the free energy landscape - to changes in the governing inter-atomic interactions. A wet lab experimental method of measuring and ranking sensitivity of the assembly process to a specific interaction is by selectively mutating those specific residues that participate in that interaction, and observing the effect on assembly efficacy. This measures the combined influences of all 3 factors (1-3) above. However, selective mutagenesis is a highly time-consuming process, even after candidate interactions have been isolated (see Section 2.3 for a brief description of this isolation procedure).

1.4 Contribution and Organization

In this paper, we describe a novel computational approach to quantifying, determining and ranking sensitivity of molecular self-assembly processes to specific inter-atomic interactions. We use our new suite of algorithms EASAL (Efficient Atlasing and Search of Assembly Landscapes, described in Section 2.1) to obtain an *atlas* of the configuration space for each of the 3 relevant interfaces in a T=1 icosahedral viral shell: dimeric, trimeric and pentameric, using which we obtain a *sensitivity ranking* of each candidate interaction. To do this, we only take into account the change in interface configurational entropy, Factor (2a) above, when an inter-atomic interaction is removed. Moreover, we use a coarse approximation of the full-blown entropy computation which is a notoriously difficult problem in general with a vast literature [15, 34, 2, 13, 14, 12, 16, 11, 27, 17]. Yet our ranking tallied remarkably well with the mutagenesis results obtained at the Agbandje-Mckenna lab as described in the previous section. Previous coarse-grained computational modeling of the dynamics of viral shell assembly include [4, 36, 35, 25, 26, 21] However, to the best of our knowledge, ours is the first attempt at computational predictions of sensitivity of virus assembly processes, to the removal of specific inter-atomic interactions.

Organization

Section 2.1 gives a brief description of the key features of EASAL, its novelty, and the use of EASAL to obtain a sensitivity ranking as described above, by atlasing interface configuration spaces. Section 2.3 explains the validation methodology, specifically how we obtained and isolated the set of candidate interactions from the X-ray structure of AAV2. These are used both as input to EASAL and for the mutagenesis validation. Section 3 gives EASAL's sensitivity rankings and their mutagenesis validation. Section 4 concludes with future work.

2. DETAILS OF APPROACH

2.1 Description of EASAL and its Configuration Space Atlas

A *interface assembly system* consists of (i) a small number m (at most 10) rigid molecular motifs, each specified as the positions of the atom centers in a 3 dimensional local coordinate system; (ii) a potential energy function whose terms include sterics as Hard-Sphere potentials, weak inter-atomic interactions as Lennard Jones potentials, and implicit solvent terms [20]. A *interface assembly configuration* lives in a $6(m-1)$ dimensional cartesian space representing the rotations and translations of the local coordinate systems of $m-1$ of the rigid molecular motifs with respect to one fixed rigid motif. EASAL (Efficient Atlasing and Search of Assembly Landscapes) is our new suite of algorithms that completely maps configuration spaces with a high dimensionality and geometric complexity while providing provable guarantees. It leverages the following features.

Active geometric constraint regions. For the type of potential energy function described above, the geometry and topology of potential energy basins - that are of interest a given entropy computation - can be completely partitioned into *active geometric constraint regions*. I.e, regions of the configuration space that satisfy (bounds on) sets of inter-residue distances or angles (a restricted class of semi-algebraic sets). EASAL's operation is based entirely on these constraints. The choice of geometric constraints depends on the potential energy function and partition into the approximately constant potential energy regions of the configura-

tion space. Then each region is uniquely labeled by a set of geometric constraints that are explicitly active in that region.

Thom-Whitney Stratification, Convexification and Atlas A classical way to describe a partition into active geometric constraint regions is as a topological complex using the so-called Thom-Whitney stratification of semi-algebraic sets [18]. See Figure 1. Intuitively, the configuration space is partitioned into strata. Each stratum consists of active geometric constraint regions of the same effective dimension. The "children" of a given region are 1 lower dimensional regions obtained when one additional geometric constraint is satisfied or active. The zero dimensional regions thus consists of rigid configurations. See Figure 1. In [22], we have shown that assembly configuration spaces can be *atlased*, i.e, they have *convexifiable* active constraint regions using so-called *Cayley*, or distance parameters, following a new theory developed by some of the authors [30, 7], who also showed that cartesian configurations that correspond to a Cayley configuration can be computed efficiently using an optimally parametrized algebraic system [24].

Novelty There is extensive literature for using MC and MD to uniformly sample and thereby compute configurational entropy and free nergy for all types of molecular configuration spaces [15, 34, 2, 13, 14, 12, 16, 11, 27, 17]. There has also been some research on infering the topology of the configuration space [10, 31, 19] starting from MC and MD samples, and using the topology to guide dimensionality reduction, [33]. However, to the best of our knowledge, none of these methods have the above features of EASAL, e.g., active geometric constraint regions, stratifications, convexifiable regions, etc., which are required for provably efficient and complete atlasing and obtaining our type of sensitivity ranking.

A possible explanation: most of these methods do not specifically target assembly, although we now know that even simple folding configuration spaces, (e.g., the classic cycloheptane or cyclooctane) *cannot be atlased*, a stark difference from assembly configuration spaces.

2.2 Sensitivity Ranking from Interface Configuration Space Atlas

The potential energy basins of an interface assembly system are centered around the configurations in the zero-dimensional active constraint regions of the configuration space atlas. These regions cannot be found by EASAL without finding the higher dimensional regions of the atlas. Furthermore, each configuration in such a region is rigid and stable. It satisfies at least $6(m-1)$ of the input constraints (for m rigid motifs), i.e, the corresponding inter-atomic distances fall within their respective Lennard-Jones wells. The number of (similar) copies of configurations could be viewed as an approximate measure of the size or volume of a potential energy basin (configurational entropy associated with that basin).

However many of these configurations (and corresponding energy basins) are very different from the successful interface assembly configuration. We first compute the ratio of the number of copies of successful configurations to the number of copies of all configurations for the entire atlas, i.e, with all the input inter-atomic interactions being present. Our measure of sensitivity of a given input inter-atomic interaction is obtained by computing the same ratio restricted to a portion of the atlas, i.e., those regions where the given inter-atomic interaction is not part of the active constraint set. A single atlas computation is sufficient to compute sensitivity to each given inter-atomic inteaction. The interactions are then ranked based on the sensitivity.

2.3 Data and Validation Methodology

We started from simplified potential energy landscapes designed from known X-ray structure of AAV2 coat protein monomers and interfaces [1, 8, 23] (data provided by Mavis Agbandje-Mckenna's lab, see Figure 1). For each of the 3 interfaces (2-fold, 3-fold and 5-fold), we determined the pairs of interacting residues that are conserved in related viruses (10-20 pairs for each interface). These were used as the candidate interactions for the sensitivity rankings.

For the mutagenesis experiment, these candidate interactions were disabled one by one, by mutating one of the residues in the pair. The effect of the mutation on assembly efficacy was determined by measuring concentration of successfully assembled viral shells via cryo-electron microscopy. *This experiment took at least 2 years.*

For EASAL's predictions, we treated monomers as single rigid motifs in the interface assembly systems. We used Lennard-Jones potentials for the above pairs of interacting residues and hard spheres for the sterics of the remaining residues. No solvent effects were considered. For each interface, for of these interactions, the sensitivity of interface assembly was computed as described in the previous section. In fact, for each of the interfaces we generated a new atlas and sensitivity results for more than one assembly system obtained from different pairs of participating multimers - see Section 3 for a detailed description. The rationale was that the same interface drives assembly of different types of multimer-pairs during the formation of larger intermediate subassemblies. We obtained a cumulative sensitivity ranking for each interaction, over all of the relevant interface assembly systems for that interaction. *This computation took 1 week.*

3. RESULTS

We tabulate EASAL's sensitivity rankings below, for the AAV2 dimer, trimer and pentamer interface assembly systems for monomers. Figure 2 shows the change of the number and ratio of successful interface assembly configurations after removing each interaction. See the tabulated results in Tables 1 and 2. The highest ranked interactions output by EASAL (i.e, assembly is most sensitive to these interactions), were validated by mutagenesis resulting in assembly disruption (the "Confirmed" column). Note that blank entries in the "Confirmed" column indicate that mutagenesis was not performed for disabling those interactions, i.e, it is as yet unknown whether EASAL's predictions are correct.

Note concerning the Trimer interface: As can be seen from Figure 2 we could not obtain useful sensitivity rankings for the trimer interface due to heavy influence of sterics caused by interdigitation). This tallied with the fact that mutagenesis of the any of the trimer interface interactions could not disrupt assembly. We do not believe that assembly of the AAV2 shell is sensitive to any of the trimer interactions. We conjecture that the assembly proceeds primarily by dimeric and pentameric interface interactions. Trimers interdigitate and contribute to stability of the capsid after the assembly is complete.

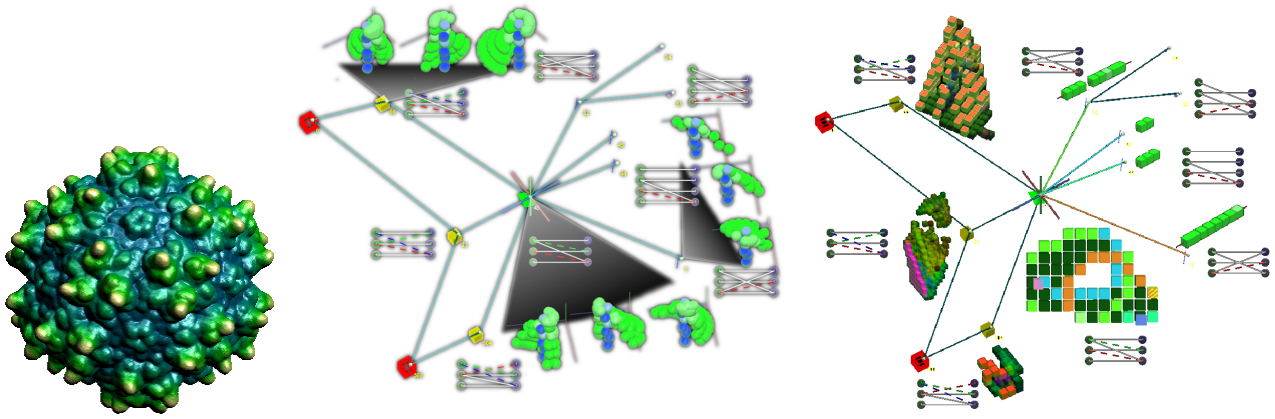

Figure 1: Left: AAV2 X-ray structure - Courtesy VIPER DB[9]; Mid: Portion of an atlas with green, 2dim active constraint region in the center; regions in 4,3,2,1,0 dimensional strata from left to right are colored red, yellow, green etc. Regions are labeled by their active constraint graphs (dark edges), which increase with dimension; the regions are shown as sweeps around a stationary reference molecule. Right: Cayley parameters (dotted colored edges) yield convexifiable charts for active constraint regions; note intersection with the complement of a convex subregion in the center.

Table 1: Sensitivity ranking: Dimer Interface

Residue1	Residue2	Confirmed
P293	W694, P696	Yes[3]
R294	E689, E697	Yes[3, 32]
E689	R298	Yes[3]
W694	P293, Y397	Yes[3]
P696	P293	Yes[3]
Y720	W694	Yes[3]

Table 2: Sensitivity ranking: Pentamer Interface

Residue1	Residue2	Confirmed
N227	Q401	Yes[32]
R389	Y704	
K706	N382	
M402	Q677	Yes[3]
K706	N382	
N334	T337,Q319	
S292	F397	Yes[32]

Pentamer interface with participating Multimers

In Figure 2, we only considered interface assembly systems with two participating monomers. But during the formation of larger assembly intermediates two multimers could assemble across the same interface. We obtained a new pentamer interface atlas for a monomer and a dimer. While the weak-force interactions remain the same, the number of sterics increases and affects the atlas regions and topology significantly.

Factoring this into the rankings, we found two other crucial interactions for the pentamer interface: S292-F397 and N227-Q401. Both of them were confirmed by assembly disruption through mutagenesis, and have been included in the above tables.

4. CONCLUSIONS AND FUTURE WORK

Our results indicate that our method of sensitivity ranking using EASAL's interface configuration space atlasing could be remarkably effective in studying the robustness of viral shell assembly. This is a surprising result since our approach ignored the influence of Factors (1) (2b) and (3) described earlier, and only used a coarse approximation of Factor (2a). Further computational robustness predictions and validation are underway for other T=1 and T=3 viruses, and initial results for the MVM (minute virus of mice, another T=1 virus) are promising.

So far we have (i) considered an entire monomers as a single rigid motif (ii) only considered pairs of participating multimers while atlasing interface configuration spaces. Hence it was sufficient to atlas the configuration space of only 2 rigid motifs at a time. One natural extension of the algorithm is taking more than 3 rigid motifs into account simultaneously. Another is to consider more than 2 participating multimers, i.e, atlas the assembly configuration space of 2 or more interfaces simultaneously.

5. REFERENCES

[1] M. Agbandje-McKenna, A. Llamas-Saiz, F. Wang, P. Tattersall, and M. Rossmann. Functional implications of the structure of the murine parvovirus, minute virus of mice. *Structure*, 6:1369–1381, 1998.

[2] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics*, 115(14):6289, 2001.

[3] A. Bennett. N/a. Unpublished manuscript, N/A 2012.

[4] B. Berger, P. Shor, J. King, D. Muir, R. Schwartz, and L. Tucker-Kellogg. Local rule-based theory of virus shell assembly. *Proc. Natl. Acad. Sci. USA*, 91:7732–7736, 1994.

[5] M. Bona and M. Sitharam. Influence of symmetry on probabilities of icosahedral viral assembly pathways. *Computational and Mathematical Methods in Medicine: Special issue on Mathematical Virology, Stockley and Twarock Eds*, 2008.

[6] M. Bona, M. Sitharam, and A. Vince. Tree orbits under the action of the icosahedral group and

Figure 2: Right: EASAL found the succesful interface assembly configurations (within its zero-dimensional regions) starting from 2 randomly positioned, separated monomers, one blue and one green (in the trimer and pentamer cases, the remaining green monomers are shown merely for identifying the type of interface). Left: each row in the chart shows the situation after removing the interaction in the row label. The three horizontal bars in each row represent the total number of zero-dimensional or rigid configurations, the number of configurations close to the successful interface assembly configuration, and their ratio.

enumeration of macromolecular assembly pathways. *Bull. Math. Bio, Special Issue on Algebraic Biology, to appear*, 2011.

[7] U. Chittamuru. Sampling configuration space of partial 2-trees in 3d. 2011.

[8] P. E., V. Bowman, N. Kaludov, L. Govindasamy, H. Levy, P. Nick, R. McKenna, N. Muzyczka, J. A. Chiorini, T. S. Baker, and M. Agbandje-McKenna. Structure of adeno-associated virus type 4. *Journal of Virology*, 79:5047–58, 2005.

[9] J. et al. http://viperdb.scripps.edu/info_page.php?vdb=1lp3.

[10] D. Gfeller, D. M. De Lachapelle, P. De Los Rios, G. Caldarelli, and F. Rao. Uncovering the topology of configuration space networks. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, 76(2 Pt 2):026113, 2007.

[11] M. S. Head, J. A. Given, and M. K. Gilson. Mining minimaÂ£: Direct computation of conformational free energy. *The Journal of Physical Chemistry A*, 101(8):1609–1618, 1997.

[12] U. Hensen, O. F. Lange, and H. GrubmÃijller. Estimating absolute configurational entropies of macromolecules: The minimally coupled subspace approach. *PLoS ONE*, 5(2):8, 2010.

[13] V. Hnizdo, E. Darian, A. Fedorowicz, E. Demchuk, S. Li, and H. Singh. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *Journal of Computational Chemistry*, 28(3):655–668, 2007.

[14] V. Hnizdo, J. Tan, B. J. Killian, and M. K. Gilson. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *Journal of Computational Chemistry*, 29(10):1605–1614, 2008.

[15] M. Karplus and J. Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14(2):325–332, 1981.

[16] B. J. Killian, J. Yundenfreund Kravitz, and M. K. Gilson. Extraction of configurational entropy from molecular simulations via an expansion approximation. *The Journal of chemical physics*, 127(2):024107, 2007.

[17] B. M. King, N. W. Silver, and B. Tidor. Efficient calculation of molecular configurational entropies using an information theoretic approximation. *The Journal of Physical Chemistry B*, 0(ja):null, 0.

[18] T.-C. Kuo. On thom-whitney stratification theory. *Mathematische Annalen*, 234:97–107, 1978. 10.1007/BF01420960.

[19] Z. Lai, J. Su, W. Chen, and C. Wang. Uncovering the properties of energy-weighted conformation space networks with a hydrophobic-hydrophilic model. *International Journal of Molecular Sciences*, 10(4):1808–1823, 2009.

[20] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *Proteins*, 35(2):133–152, 1999.

[21] C. J. Marzec and L. A. Day. Pattern formation in icosahedral virus capsids: the papova viruses and nudaurelia capensis β virus. *Biophys*, 65:2559–2577, 1993.

[22] A. Ozkan and M.Sitharam. Easal: Efficient atlasing and search of assembly landscapes. In *Proceedings of BiCoB*, 2011.

[23] E. Padron, R. McKenna, N. Muzyczka, N. Kaludov, J. A. Chiorini, and M. Agbandje-McKenna. Structurally mapping the diverse phenotype of adeno associatedvirus serotype 4. *Journal of Virology*, 80:11556–570, 2006.

[24] J. Peters, J. Fan, M. Sitharam, and Y. Zhou. Elimination in generically rigid 3d geometric constraint systems. In *Proceedings of Algebraic Geometry and Geometric Modeling*, pages 27–29, Nice, September 2004. Springer Verlag,1-16,2005.

[25] D. Rapaport, J. Johnson, and J. Skolnick. Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation. *Comp Physics Comm*, 1998.

[26] V. S. Reddy, H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks, and J. E. Johnson. Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses. *Biophys*, 74:546–558, 1998.

[27] G. S. and Chirikjian. Chapter four - modeling loop entropy. In M. L. Johnson and L. Brand, editors, *Computer Methods, Part C*, volume 487 of *Methods in Enzymology*, pages 99 – 132. Academic Press, 2011.

[28] M. Sitharam and M. Agbandje-Mckenna. Sampling virus assembly pathway: Avoiding dynamics. *Journal of Computational Biology*, 13(6), 2006.

[29] M. Sitharam and M. Bona. Combinatorial enumeration of macromolecular assembly pathways. In *Proceedings of the International Conferecnce on bioinformatics and applications*. World Scientific, 2004.

[30] M. Sitharam and H.Gao. Characterizing graphs with convex cayley configuration spaces. *Discrete and Computational Geometry*, 2010.

[31] G. Varadhan, Y. J. Kim, S. Krishnan, and D. Manocha. Topology preserving approximation of free configuration space. *Robotics*, (May):3041–3048, 2006.

[32] P. Wu, W. Xiao, T. Conlon, J. Hughes, M. Agbandje-McKenna, T. Ferkol, T. Flotte, and N. Muzyczka. Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *Journal of virology*, 74(18):8635–47, Sept. 2000.

[33] Y. Yao, J. Sun, X. 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(14):144115, 2009.

[34] H.-X. Zhou and M. K. Gilson. Theory of free energy and entropy in noncovalent binding. *Chemical Reviews*, 109(9):4092–107, 2009.

[35] A. Zlotnick. To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.*, 241:59–67, 1994.

[36] A. Zlotnick, J. Johnson, P. Wingfield, S. Stahl, and D. Endres. A theoretical model successfully identifies features of hepatitis b virus capsid assembly. *Biochemistry*, 38:14644–14652, 1999.