Characterizing discrete and continuous 3-D particle heterogeneity from cryo EM images

Peter C. Doerschuk

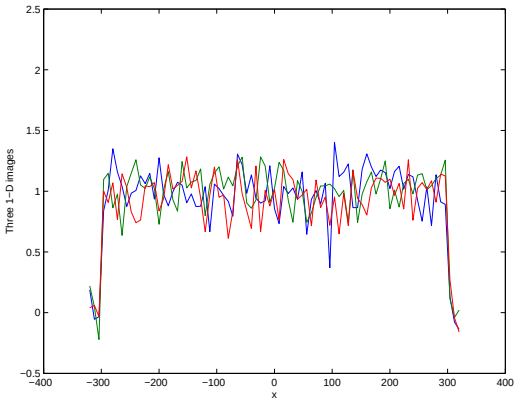
Meinig School of Biomedical Engineering
School of Electrical and Computer Engineering
Cornell University

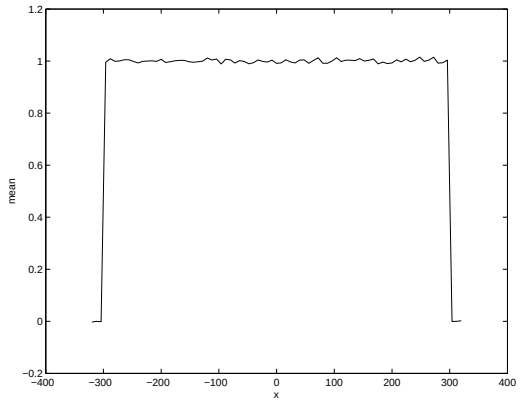
March 31, 2016

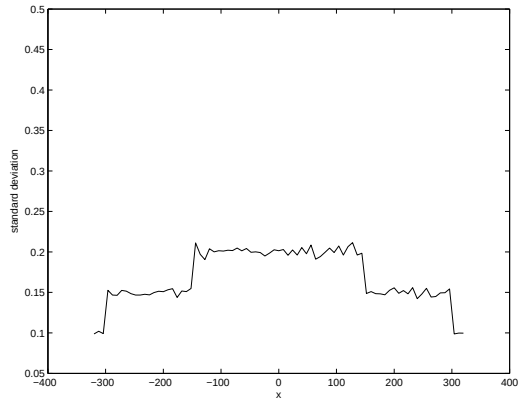
Thanks to

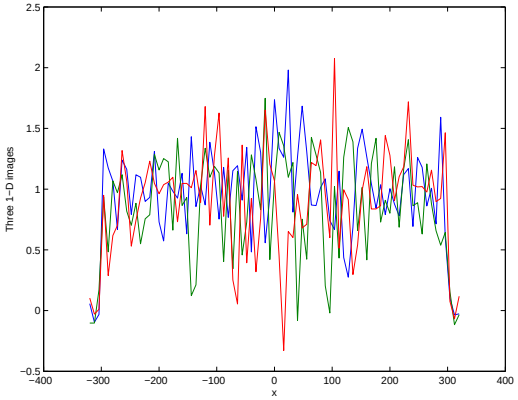
- John E. Johnson, Tatiana Domitrovic, Tsutomu Matsui, David Veesler
- Yunye Gong, Qiu Wang, Nan Xu, Yili Zheng
- NSF

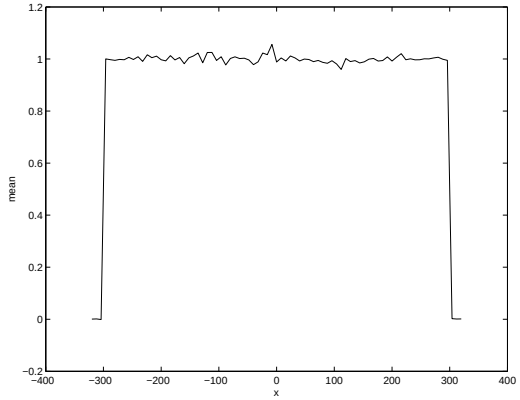
Imaging particle heterogeneity

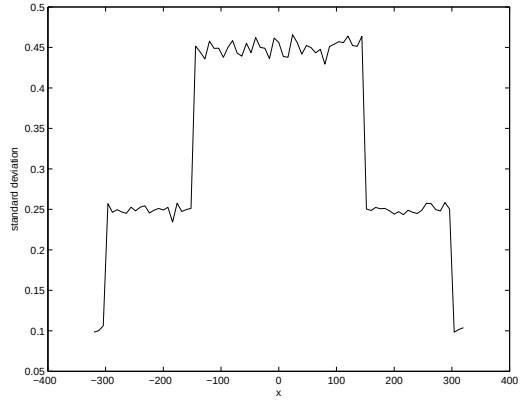












Topics

1. Computational methods (Zheng, Wang, Doerschuk, *J. Optical Soc. Am. Series A* 29(6):959–970, 2012).
2. Maturation of different quasi-equivalent subunits of Nudaurelia Capensis Omega Virus (NωV) has different kinetics (Wang, Matsui, Domitrovic, Zheng, Doerschuk, Johnson, *J. Structural Biol.* 181(3):195–206, 2013).
3. The viral maturation protease of Hong Kong 97 (HK97) stabilizes the δ domain before cleavage (Gong, Veesler, Doerschuk, Johnson, *J. Structural Biol.* 193(3):188–195, 2016).

Description of the particle

$$\rho(\mathbf{x}) = \sum_{\alpha} c_\alpha \psi_\alpha(\mathbf{x}) \quad (1)$$

$$\rho(\mathbf{x}) = \text{3-D electron scattering intensity in real space} \quad (2)$$

$$c_\alpha = \alpha^{\text{th}} \text{ coefficient} \quad (3)$$

$$\psi_\alpha(\mathbf{x}) = \alpha^{\text{th}} \text{ basis function} \quad (4)$$

Goals, prior and current

1. Homogeneous particles: Coefficients are unknown numbers and the goal is to compute numerical values from the data.
2. Heterogeneous particles: Coefficients are random variables and the goal is to compute properties of their probability density function (pdf) from the data.

For Gaussian random variable coefficients, their pdf is characterized by computing

1. the mean vector (i.e., average value).
2. the covariance matrix (i.e., average of the product of the deviations of two random variables away from their mean values)

Methods

1. Maximum likelihood (ML) estimation of mean vector and covariance matrix
2. An expectation-maximization algorithm to compute the ML estimates
    1. Expectation: By numerical computation
    2. Maximization: By solving a linear system of equations (mean vector) or by numerical computation (covariance matrix)
3. Matlab software
    1. Multi core.
    2. Not gp-GPU coprocessor or distributed memory capable

NωV background

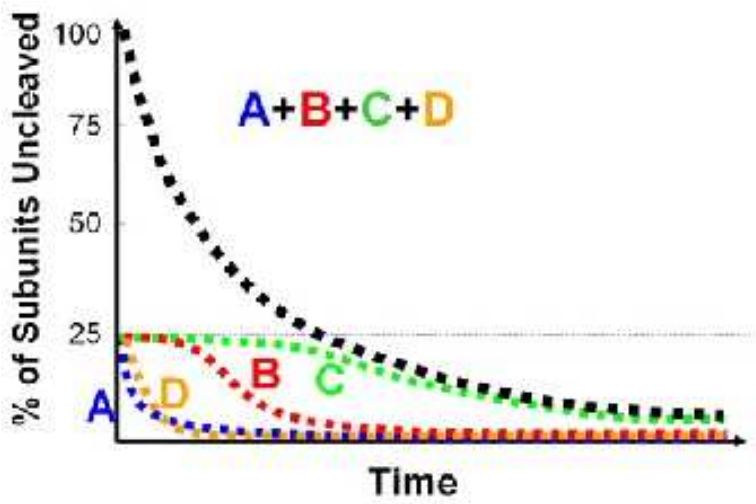
1. RNA virus with icosahedral symmetry and $T = 4$ quasi symmetry.
2. A capsid with 240 copies of the alpha peptide assembles.
3. Maturation involves shrinking of the capsid followed by autocatalytic cleavage of alpha to yield the major capsid protein beta and the small gamma peptide which remains associated with the particle.

NωV maturation

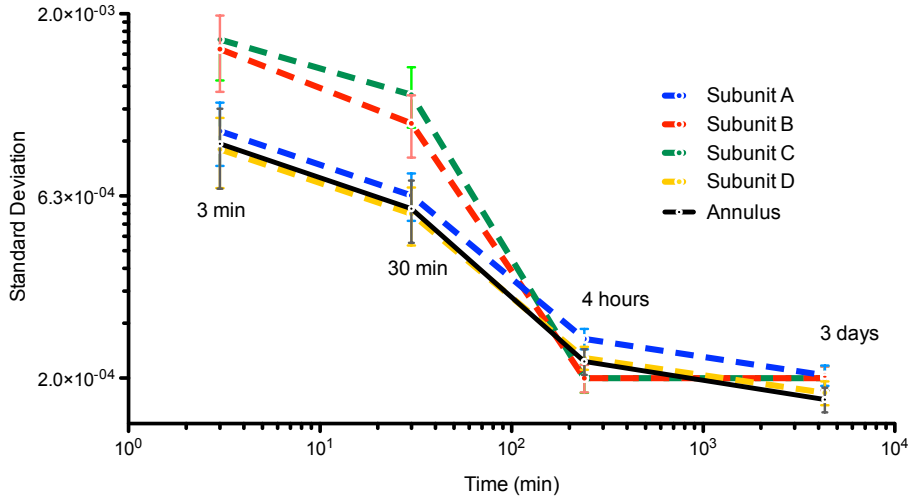
1. Maturation can be triggered by a pH drop to 5.0.
2. Cleavage is slow: Can perform time-resolved cryo EM at 3 min, 30 min, 4 hour, and 3 days (Matsui, Lander, Khayat, Johnson, *PNAS* 107(32):14111–14115, 2010).
3. Overall geometry during cleavage is unchanging but, by difference maps (10Å), can visualize change at the autocatalytic sites corresponding to different kinetics at the B and C subunits (fast cleavage) and the A and D subunits (slow cleavage)

Can the kinetics be imaged without using difference maps?

NωV heterogeneity

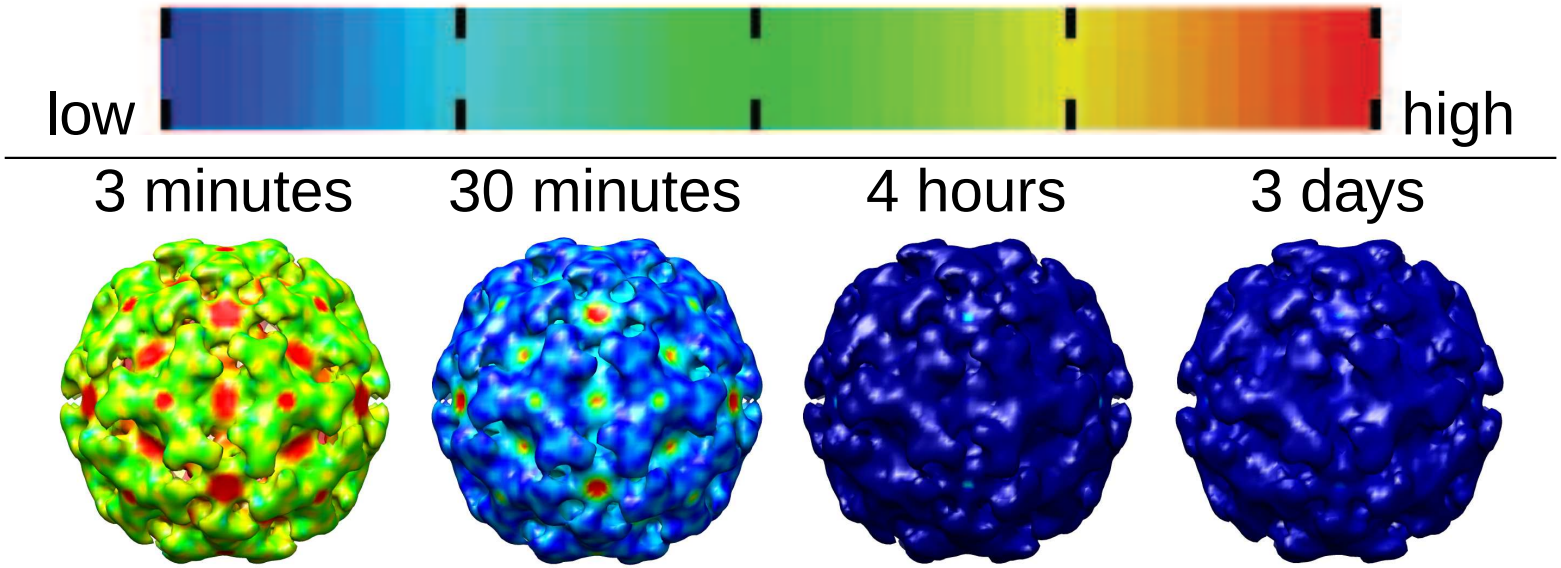


Matsui, Lander, Khayat, Johnson (10Å)



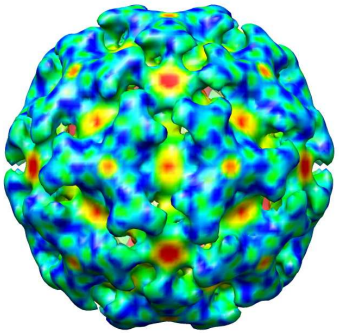
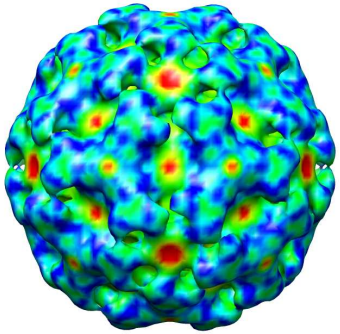
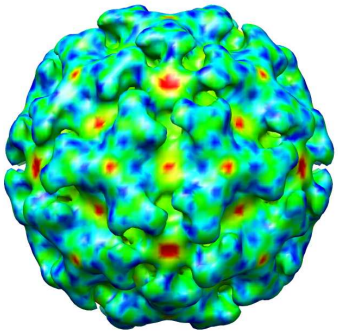
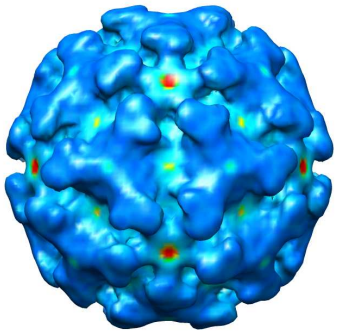
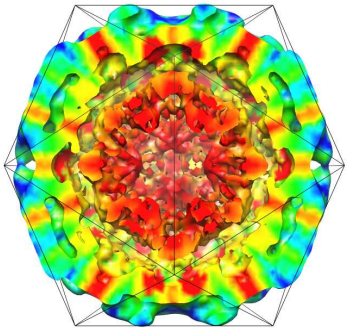
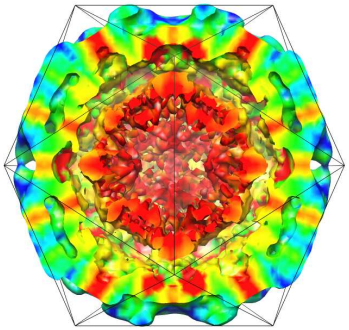
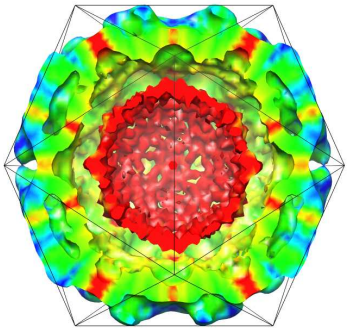
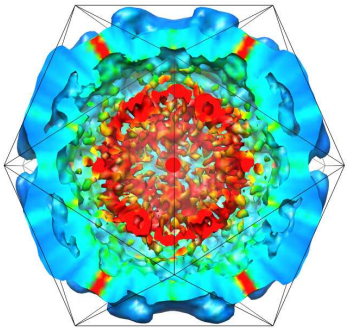
Wang, Matsui, Domitrovic, Zheng, Doerschuk, Johnson (21Å)

NωV heterogeneity at 21Å

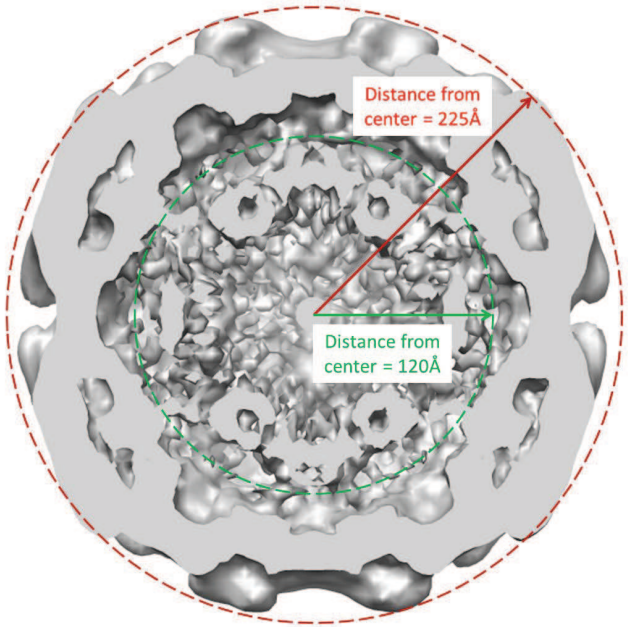


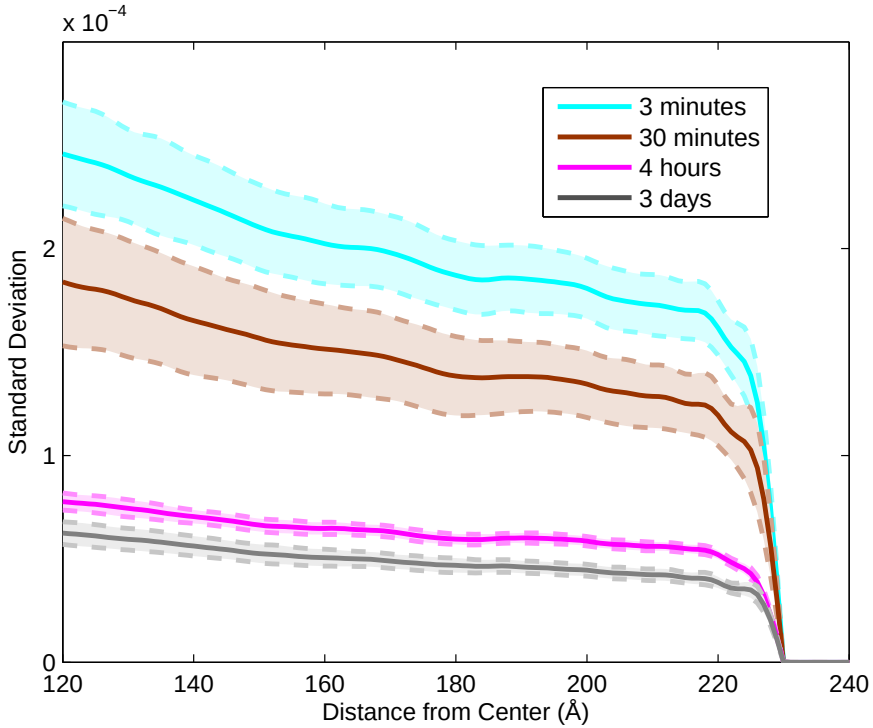
All four panels: [blue–red]=[1–5.5] $\times$ 10^{-7}.

NωV heterogeneity at 21Å

3 minutes	30 minutes	4 hours	3 days
			
			
$[3–10] \times 10^{-7}$	$[1.5–4] \times 10^{-7}$	$[0.38–1] \times 10^{-7}$	$[0.04–0.9] \times 10^{-7}$

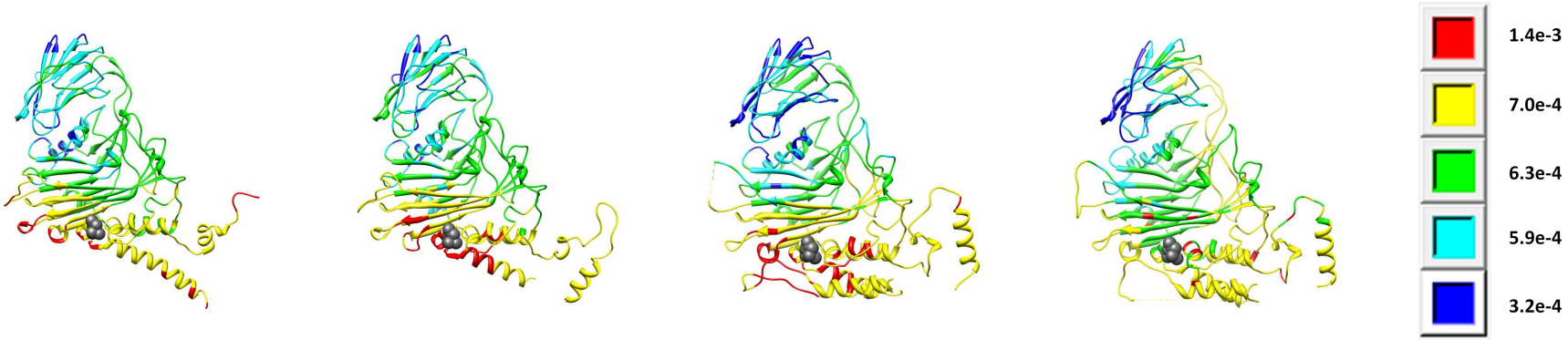
NωV heterogeneity at 21Å: Spherical average in capsid region



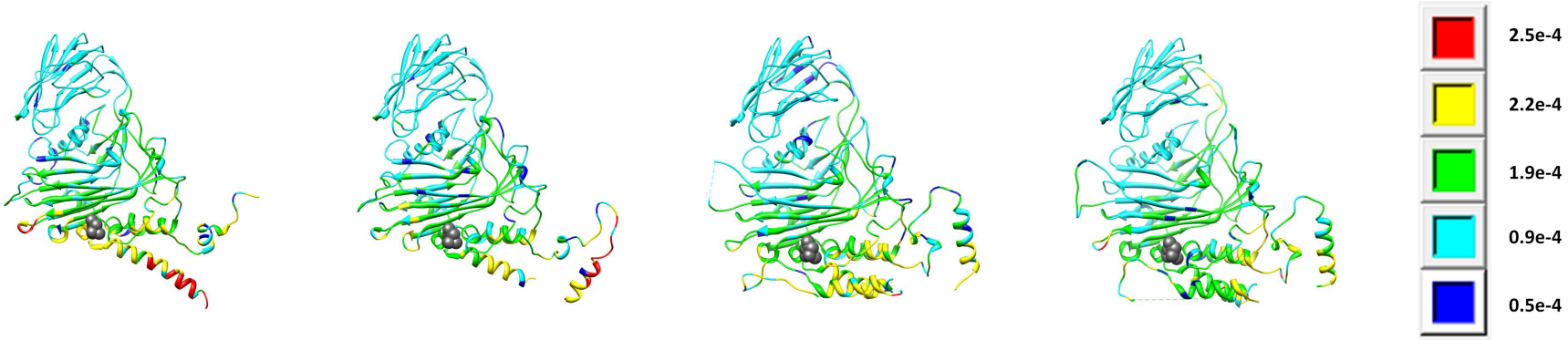


NωV heterogeneity at 21Å: Paint with heterogeneity

Subunit A | Subunit B | Subunit C | Subunit D



3 min



Capsid

HK97 background

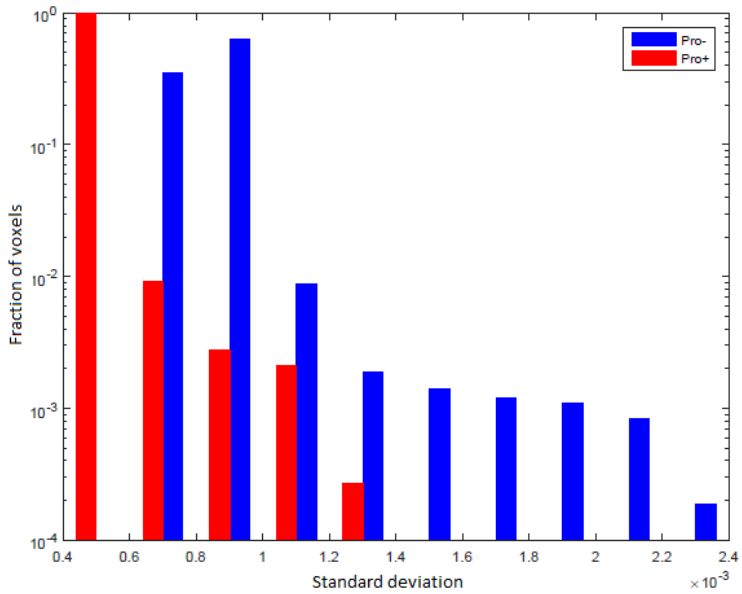
1. Virus-like particles (VLPs) of dsDNA bacteriophage HK97.
2. Co-expression of capsid protein and viral protease leads to assembly of 420 capsid protein copies and about 100 protease copies in Prohead I.
3. VLP: icosahedral symmetry and $T = 7l$ quasi symmetry.
4. Protease is activated by completion of assembly and digests (1) 103 AA (δ domain) from N-termini of capsid peptides and (2) itself.
5. Isolate Prohead I by packaging a defective protease (PhI$^{\mathrm{Pro}+}$) or by not packaging protease (PhI$^{\mathrm{Pro}-}$).

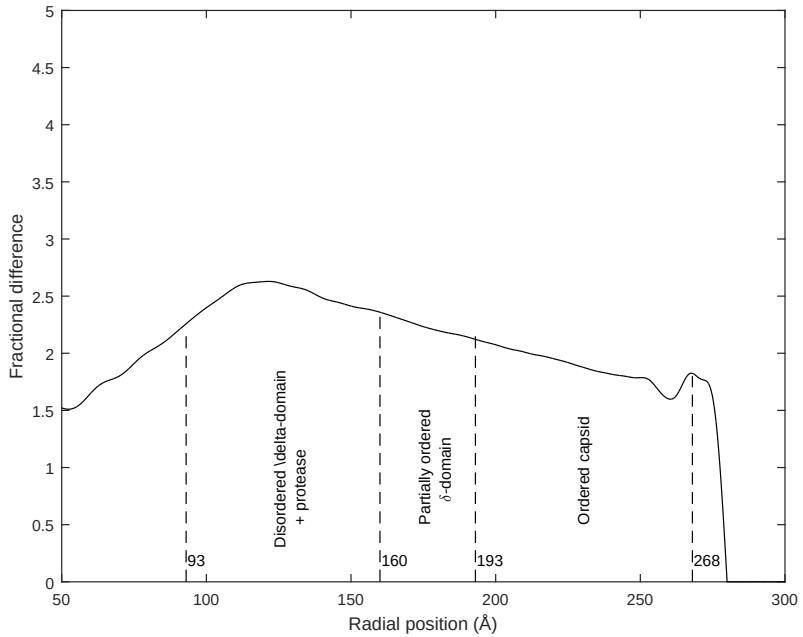
Comparison of $\text{PhI}^{\text{Pro}+}$ and $\text{PhI}^{\text{Pro}-}$

1. Structures of $\text{PhI}^{\text{Pro}+}$ (8.3Å) and $\text{PhI}^{\text{Pro}-}$ (7.8Å) are similar (Veesler, Khayat, Krishnamurthy, Snijder, Huang, Heck, Anand, Johnson, *Structure* 22(2):230–237, 2014).
2. Biochemistry of $\text{PhI}^{\text{Pro}+}$ and $\text{PhI}^{\text{Pro}-}$ are different:
   1. Hydrogen-deuterium exchange: δ domain residues protected in $\text{PhI}^{\text{Pro}+}$ compared to $\text{PhI}^{\text{Pro}-}$.
   2. Urea and other disassembly conditions: $\text{PhI}^{\text{Pro}+}$ more resistant than $\text{PhI}^{\text{Pro}-}$.

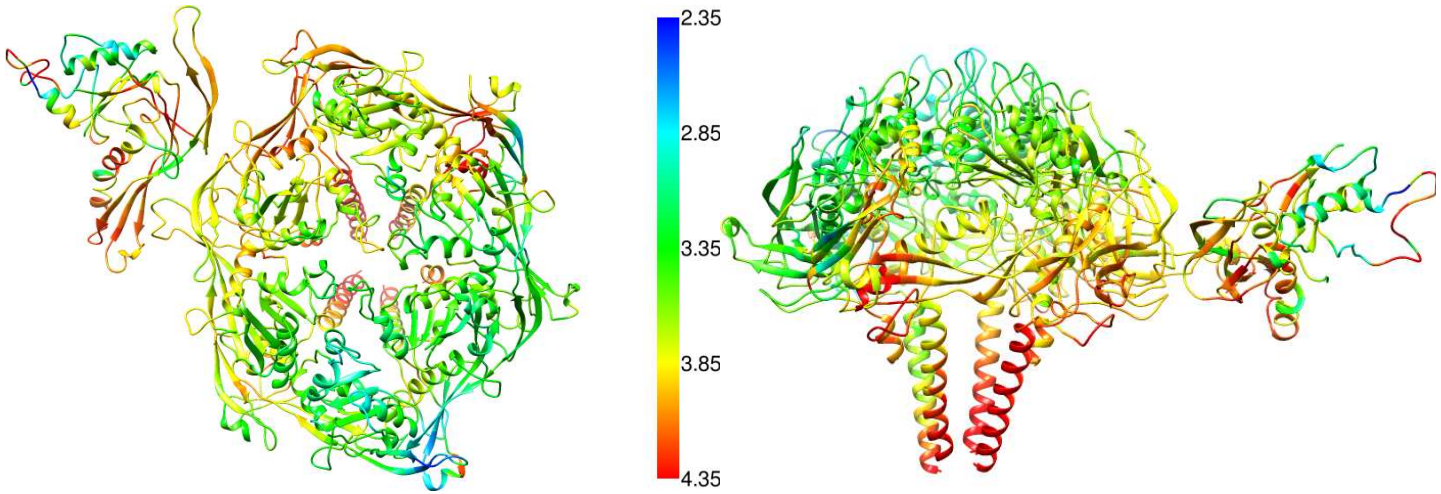
Can these differences be imaged?

HK97 heterogeneity at 17Å: Histogram and spherical average

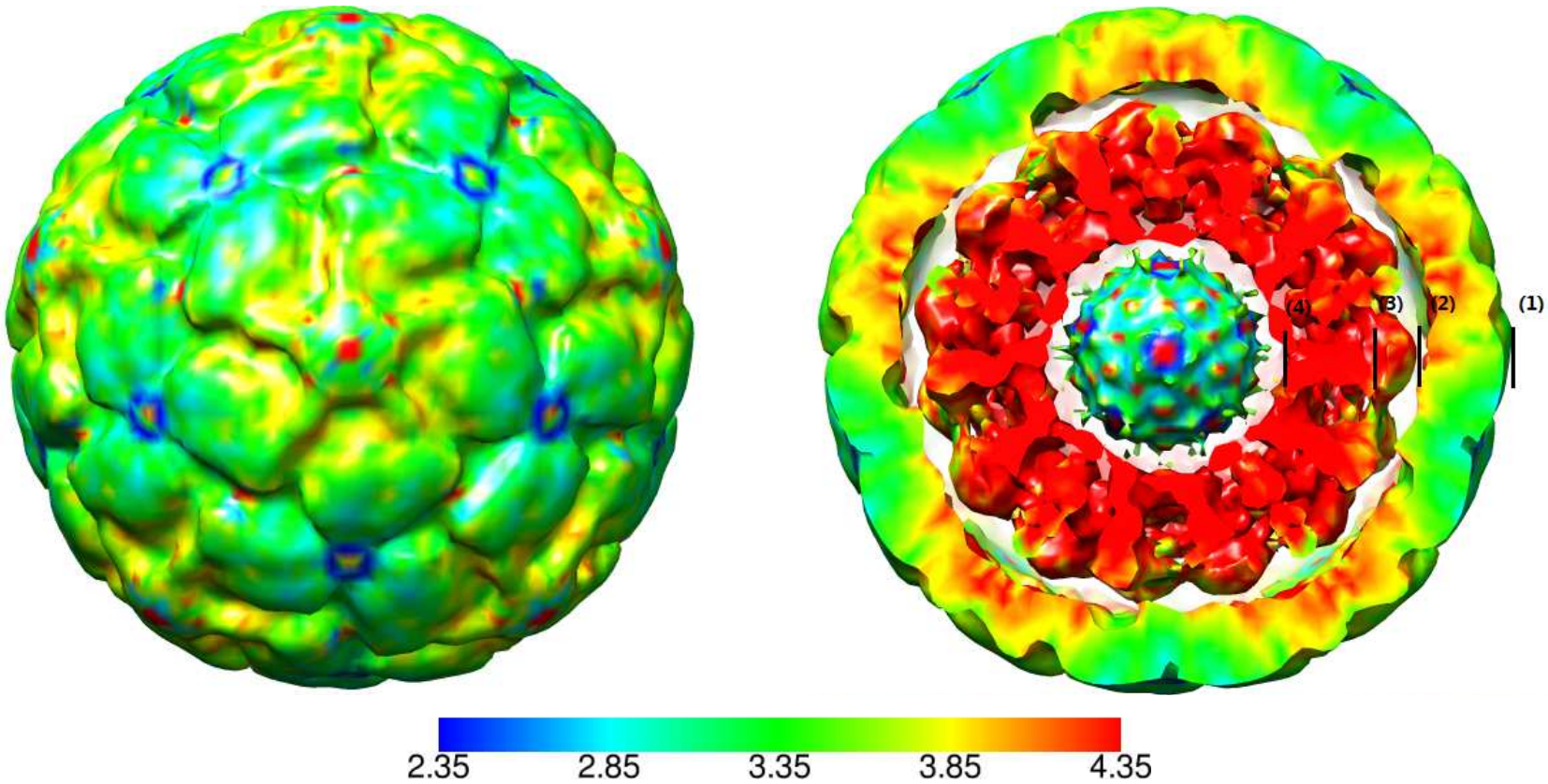




HK97 heterogeneity at 17Å: Paint with the heterogeneity difference



HK97 heterogeneity at 17Å: Surface and cross section of $\mathrm{PhI}^{\mathrm{Pro}-}$ painted with the heterogeneity difference



Thank you!
Questions?