Quarterly Reviews of Biophysics

www.cambridge.org/qrb

Single-particle cryoEM: Theory and practice

Ardan Patwardhan[1], Peter B. Rosenthal[2], Christopher J. Russo[3] and Richard Henderson[3]

[1]Cellular Structure & 3D Bioimaging, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), UK; [2]Structural Biology of Cells and Viruses Laboratory, The Francis Crick Institute, London, UK and [3]MRC Laboratory of Molecular Biology, Cambridge, UK

Review

Cite this article: Patwardhan A, Rosenthal PB, Russo CJ and Henderson R (2026). Single-particle cryoEM: Theory and practice. *Quarterly Reviews of Biophysics*, **59**, e16, 1–13
https://doi.org/10.1017/S0033583526100249

Received: 20 July 2026 UTC
Revised: 20 July 2026 UTC
Accepted: 05 September 2026 UTC

Keywords:
Electron Microscopy Data Bank (EMDB); Protein Data Bank (PDB); cryogenic electron microscopy; single particle analysis; tomography; radiation damage; B factor

Corresponding author:
Richard Henderson;
Email: rh15@mrc-lmb.cam.ac.uk

Abstract

Radiation damage is the fundamental limitation to the useful exposure in single-particle electron cryomicroscopy (cryoEM), so the goal of the field is to record near-perfect images, to extract as much information as possible from them, and then use this information to determine the three-dimensional structure of the desired biological molecule or molecular complex. This review presents a theoretical estimate of the number of single-particle images required to reach any desired resolution assuming perfect performance of all aspects of the imaging process. This target is then compared with the current state of cryoEM in practice, as represented by published structures in the Electron Microscopy Data Bank (EMDB). Considering recent structures, it becomes clear that the best structures in the 2.5–3.0 Å resolution range come within a factor of 3 of theoretical expectations. More typical datasets still require 100 times more images in practice than necessary in theory. This also applies to the higher resolution structures where larger datasets are essential. The reasons for this discrepancy are discussed alongside possible strategies to bring experimental structures closer to theoretical perfection. Finally, even the best cryoEM structures have significant radiation damage, which can only be ameliorated by recovering information from the best possible images at the beginning of the exposure at a fraction of the total dose currently used.

Table of contents

Introduction **2**
Theory **3**
Calculation of the number of particles required for different resolution ranges 3
Different ways to estimate resolution loss; different types of B-factors 3
Angular error estimation and contribution to B-factor 4

The Electron Microscopy Data Bank (EMDB): Number of particles currently required in practice **6**
Combined influence of particle size and shape **7**
Narrowing the gap between theory and practice **7**
Some future challenges **8**
Tomography and crowded environments **8**
Conclusion **8**
Appendix: Analysis of how the optimum particle number varies with structure size and shape in single-particle cryoEM **10**
Introduction **10**
Methods **10**
Initial data selection 10
Procedure for selection of entries to download 10
Manual data curation 10
Manual checks performed 10
Size and shape measurement 10
Cube root width 11
Asphericity 11
Data analysis and model fitting 11

Results **12**
Width 12
Asphericity 12
Can the width and asphericity models be combined? 12

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

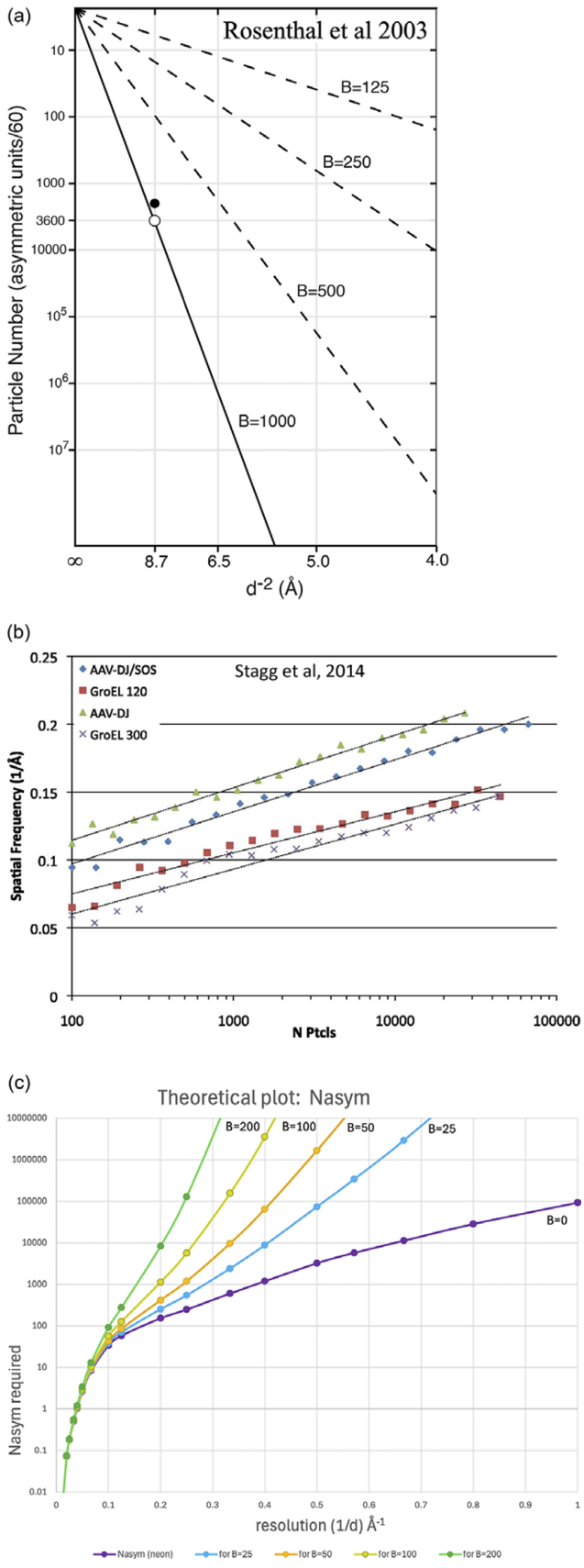


Figure 1. Two examples of earlier plots of the number of particles or asymmetric units needed to attain specific resolutions. (*a*) Rosenthal plot showing log (*Nasym*) versus spatial frequency squared ($1/d^2$) for theoretical datasets with several imposed *B*-factors.

Introduction

There have been occasional attempts to make quantitative estimates of the efficiency of electron microscope (EM) imaging and the number of images needed to obtain atomic resolution structures of organic or biological structures, since the beginning of the field. An early publication (Frank, 1975) discussed the theoretical possibility of detecting the position of non-crystalline biological molecules that all had the same orientation, and a subsequent paper (Saxton and Frank, 1977) estimated the sensitivity needed in terms of particle size. The contrast in early images, recorded on photographic film, fell well below that predicted from the intensities in electron diffraction patterns of two-dimensional crystalline specimens whether radiation-sensitive or radiation-insensitive (Henderson and Glaeser, 1985). That was partly due to the poor detective quantum efficiency (DQE) of the photographic film, but also due to other problems with the imaging process including motion of the specimen (Brilot *et al.*, 2012). Modern direct electron detectors (McMullan *et al.*, 2014) have DQEs much nearer to 100%, can record movies instead of static images, and new specimen supports eliminate the movement of the specimen upon irradiation (Naydenova *et al.*, 2020).

An early review (Henderson, 1995) gave an estimate of particle size and number of required images to reach near-atomic resolution, based on a calculation from elastic and inelastic cross-sections, alongside known radiation damage dose limitations. This suggested 10,000 single-particle images would be required to reach 3 Å resolution for particles large enough to be detectable and with enough signal to allow orientation determination (37 kDa). Later calculations, using more realistic, less stringent criteria indicated lower numbers for the required number of asymmetric units (*Nasym*) of 1,400 (Glaeser, 1999) or 600 (Rosenthal and Henderson, 2003; Vinothkumar and Henderson, 2016). A recent review (Patwardhan *et al.*, 2025) outlined how one might make a more sophisticated calculation that included resolution-dependent parameters that affect the number of particles required. The purpose of the present article is to describe more completely this calculation of the number of particles required to determine a structure to a particular resolution in cryoEM, and to compare this to the number used in typical and best-case recent structure determinations.

Two earlier publications introduced scaling laws in the form of semi-logarithmic plots relating the number of particles included in a single-particle cryoEM structure to the resolution achieved. First Rosenthal and Henderson (2003) plotted the natural logarithm of asymmetric units ln(*Nasym*) versus spatial frequency squared ($1/d^2$). For a structure blurred by a pure Gaussian function in real space, this translates into a straight line in reciprocal space, as shown in Figure 1*a*, which is copied (with permission) from that publication. Second, Stagg *et al.* (2014) introduced a second empirical scaling law plotting log(*Nptcls*) versus spatial frequency ($1/d$). This is also reproduced here in Figure 1*b*. Note that *Nasym* and

Figure 1. (Continued)
The *y* axis scaling assumes a molecule with icosahedral (60-fold) point group symmetry for particle numbers but the general principle applies to any point group. (*b*) Experimental Stagg ResLog plot (Stagg *et al.*, 2014) showing spatial frequency ($1/d$) versus log (*Nptcls*). Note that *Nptcls* and *Nasym* are related by the particle point group symmetry. The Rosenthal plot becomes linear above 10 Å when the decrease in signal-to-noise ratio with increased spatial frequency can be described by a single Gaussian blurring function in real space (Rosenthal and Henderson, 2003). The Stagg ResLog plot also looks linear over the restricted resolution range of 5–15 Å. (*c*) Plot of the number of asymmetric units (Nasym) on a log scale required to reach a given resolution ($1/d$) according to the theory described in Section 'Theory' (Eq. (1)) assuming perfect images and computations ($B = 0$) and for several values of *B*-factor.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

Nptcls in those two publications are related by the particle point group symmetry, and they differ in using resolution or resolution squared as the abscissa or ordinate. The Rosenthal plot for experimental data becomes linear above 10 Å when the decrease in signal-to-noise ratio with increased spatial frequency can be described by a single Gaussian blurring function in real space (Rosenthal and Henderson, 2003). The ResLog plot also looks linear over the restricted resolution range of 5–15 Å. In this article, we consider a much wider range of resolution from 100 to 1 Å, and a much larger variation in particle number from 1 up to 10^8. Because there is now more experimental data on single-particle cryoEM, it is also timely to revisit and extend the comparison between theory and practice, to examine how well we are doing and what we can do to make further improvements.

Theory

Given a desired resolution, and knowing the allowed exposure limited by radiation damage, it is possible to estimate the number of particles required to reach that resolution assuming that the positions and orientations of the particles in the images can be accurately determined. Depending on the definition of resolution, this results in slightly different numbers. Two earlier calculations (Henderson, 1995; Rosenthal and Henderson, 2003) determined that 600 particles are needed to reach 3 Å resolution if the critical dose *Ne* to reduce the diffracted intensity at 3 Å to $1/e$ is 5 el/Å^2. This also assumes the 0.143 Fourier Shell Correlation (FSC) resolution threshold and corresponds to an average phase error of 60°, equivalent to a figure of merit (mean cosine of phase error) of 0.5 (Blow and Crick, 1959; Rosenthal and Henderson, 2003). We now use 680 because the experimental value of *Ne* at 3 Å obtained by Naydenova *et al.* (2022) for specimens at 80 K is 4.4 rather than 5 el/Å^2. The value of 4.4 el/Å^2 was extrapolated from the value of 7.0 el/Å^2 at 4 Å for the purple membrane in Naydenova *et al.* (2022).

Calculation of the number of particles required for different resolution ranges

The factors that need to be considered are

(a) The theoretical atomic scattering amplitudes for electrons at different resolutions (*International Tables Volume C*, Prince, 2006, p. 263); because of the finite size of the electron distribution inside atoms, these decrease at higher resolution. We used the atomic scattering amplitudes σ_{neon} for elastic scattering of electrons by neon neutral atoms, rather than carbon, because neon has closed shells, which should be similar to carbon atoms that form bonds in real protein structures rather than having their outer electrons unbonded as in the independent atom model used for the calculations in the International Tables.

(b) The experimental values of *Ne* at different resolutions for specimens at 80 K, the temperature of liquid nitrogen (Naydenova *et al.*, 2022), extrapolated to higher resolutions. *Ne* is the electron dose measured in el/Å^2 at which the intensity falls to $1/e$ of its initial value because of radiation damage. (See Table 1).

(c) The minimum number of views needed at different resolutions, $\pi D/d$ where D is the diameter of the particle and d the resolution (Crowther *et al.*, 1970).

(d) The envelope function for scattering amplitudes at lower resolutions; this is the resolution range below 8–12 Å where the contrast between solvent and protein gives rise to structure factors that exceed those determined by Wilson statistics (Wilson, 1949), which assumes random atom positions.

Equation (1) shows how the theoretical number of particles at 3 Å resolution (680) can be extended to other resolution values.

$$N_{asym}(d) = 680\left(\frac{1.405^2}{\sigma_{neon}^2}\right)\left(\frac{4.4}{N_e}\right)\left(\frac{3}{d}\right)\frac{1}{G_A{}^2} \quad (1)$$

The values of the critical dose *Ne* (el/Å^2) and the elastic scattering factor for neon σ_{neon} (Å) for different resolutions d are shown in Table 1. The product of factors in Eq. (1) includes numerators for σ_{neon} (1.405 Å), *Ne* (4.4 el/Å^2), and d at 3 Å resolution and denominators indicating these parameters at defined resolution as well as the Guinier plot amplitude G_A relative to Wilson statistics, which does not affect the data beyond 8 Å resolution. The resulting theoretical numbers of particles N_{asym} are plotted as a function of resolution in Figures 1*c* and 2 (purple for $B = 0$).

More particles are needed to get to higher resolution and fewer for a lower resolution target. In the best case, around 100,000 well-preserved identical particles (asymmetric units) are needed to reach 1 Å resolution, but 20 Å resolution can be reached with only a single particle (Rosenthal and Henderson, 2003). The image of this single particle could be acquired as a dose-fractionated tilt series as is done in electron cryotomography (Oikonomou and Jensen, 2017) to give a low-resolution 3D map or as a single image, which can then be compared at high-resolution against simulated images of known structures as is done in 2D template matching (Rickgauer et al., 2017).

Multiple conformational states of a particle, if they can be identified, need an increase in the number of images to obtain good enough statistics per conformer, so may be considered within the same framework.

Different ways to estimate resolution loss; different types of B-factors

Resolution loss in practice has many causes, including image blurring, molecular disorder, partial denaturation, poor determination of individual positional or angular parameters, or instrumental limitations due to electron optics or detector performance, and these provide the experimental backdrop to the theory in this review. Resolution loss can often be characterized by a *B*-factor, which we discuss in this section.

The original definition of *B*-factor or Debye–Waller factor was introduced by Debye and Waller in two early publications (Debye, 1913; Waller, 1923) to describe the attenuation of X-ray scattering due to thermal motion of atoms. The scattering intensity is described by $I = I_0 \exp(-B/2d^2)$, where $B = 8\pi^2\langle u^2\rangle$ and u^2 is the mean-square atomic displacement. Thermal motion of the atoms in crystals such as sodium chloride at room temperature (300 K) results in atomic motion of a fraction of an Ångstrom, described by a *B*-factor of around 1.5 Å^2. Lipson and Cochran in their textbook (Lipson and Cochran, 1966) cite Bradley and Lu (1937) and Fowweather and Hargreaves (1950) with *B*-factors of 1.1 and 0.4 Å^2 for highly ordered crystals of metatolidine and $AlCr_2$ respectively. A *B*-factor of 4.9 Å^2 was used to describe disorder in methyl groups by Roth and Harker (1948).

Wilson statistics (Wilson, 1949), together with atomic scattering form factors derived from the electron or potential distribution in atoms can be used to predict the theoretically expected value of the

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

Table 1. Values of the critical radiation dose *Ne* (el/Å^2) for fading of purple membrane electron diffraction spots to an intensity of 1/*e* (Naydenova *et al.*, 2022), the elastic scattering amplitude σ_{neon} at different resolutions (Prince, 2006), and the Guinier plot amplitude G_A at different resolutions, as used in this article

Resolution *d* (Å)	Critical dose *Ne* for 1/*e* fading (el/Å^2)	Elastic scattering amplitude for neon σ_{neon} (Å)	sin θ/λ = 1/2*d*	Guinier amplitude, G_A relative to Wilson statistics (E2CD)	Guinier amplitude for smaller protein (catalase)
100.0	1,000.0	1.652	0.005	5	40
50.0	400.0	1.651	0.010	2	14
40.0	250.0	1.650	0.0125	1.9	5
30.0	150.0	1.649	0.0167	1.6	3.6
25.0	100.0	1.648	0.020	1.5	2.4
20.0	65.0	1.645	0.025	1.3	1.4
15.0	39.0	1.634	0.033	1.1	1.2
10.0	17.5	1.626	0.050	1	1
9.0	16.0	1.619	0.055	1	1
8.0	13.0	1.612	0.0625	1	1
5.0	8.5	1.552	0.100	1	1
4.0	7.0	1.500	0.125	1	1
3.0	**(4.4)**	**1.405**	0.167	1	1
2.5	(3.05)	1.313	0.200	1	1
2.0	(1.75)	1.176	0.250	1	1
1.75	(1.33)	1.083	0.286	1	1
1.5	(0.98)	0.971	0.333	1	1
1.25	(0.68)	0.805	0.400	1	1
1.0	(0.44)	0.619	0.500	1	1

The values at 3 Å resolution, which are highlighted in bold, are those used in the numerator of Eq. (1), whereas the full tabulation is used to produce the theoretical curves shown in Figs 1*c* and 2. Values of *Ne* in brackets at 3 Å resolution and beyond are extrapolated.

average scattering intensity versus resolution. The distribution in real space results in an effective *B*-factor in reciprocal space. Again, these are small values in the range of 2–3 Å^2.

Statistical disorder in typical protein structures or other macromolecular assemblies due to soft fluctuations can be characterized by *B*-factors, which are often in the range 10–30 Å^2.

Rosenthal and Henderson (2003) characterized the effect of inaccurate determination of the particle orientations of single particles as $B_{\text{computation}}$, and any intrinsic specimen or image disorder as B_{image}. They emphasized disorder in real space that can be represented by a single Gaussian distribution can be characterized by a single *B*-factor.

There is also the effect of radiation damage, which causes bond breakage with different bonds being broken in different molecules, so that on average there is an increase in disorder proportional to dose. This can be characterized by an increased *B*-factor of about 5 Å^2 per electron/ Å^2 for specimens at 80 K, or 3.5 Å^2 per electron/ Å^2 at 13 K (Dickerson *et al.*, 2025). The increased disorder is accompanied by an increase in molecular volume, which is due to the breakage of covalent bonds and replacement with van der Waals contacts (Garman, 2010).

For ice-embedded specimens, the water molecules in the embedding medium also move and are modified during irradiation (McMullan *et al.*, 2014), though at liquid nitrogen temperature (80 K) or indeed at temperatures below 120 K there is no water motion at all in the absence of irradiation and therefore no dose-rate effect. The water molecules only move during irradiation in what can be described as pseudo-Brownian motion. This pseudo-Brownian motion of the irradiated water displaces the embedded macromolecules and causes a blurring of their image, which again can be characterized by a *B*-factor, which gets bigger for smaller proteins. The motion of the average water molecule in bulk irradiated amorphous ice is about 1 Å for each 1 el/Å^2 of irradiation, or 5 Å during a typical exposure of 25 el/Å^2. For a 100,000 Da protein molecule, which is 6,000 times more massive than a water molecule, the resulting *B*-factor increase for an image averaged over a typical cryoEM exposure has been estimated to be about 25 Å^2 (McMullan *et al.*, 2014), though this will be much less in a dose-weighted calculation.

Angular error estimation and contribution to B-factor

In practice, it is never possible to determine the exact orientation of images of single particles from low-dose images. There is always an uncertainty in the orientation determination, which leads to an angular inaccuracy and blurring of the 3D map that is reconstructed from the images. This blurring can be characterized by a *B*-factor such as listed in Table 2 or measured experimentally (Henderson *et al.*, 2011) using tilt pairs, which relates angular error to additional *B*-factor. Rosenthal and Henderson (2003) drew a distinction between the angular error that comes from physical motion during the imaging exposure and the computational error that arises later from the stochastic noise in the image. The physical beam-induced motion causes translational and rotational motions, both of which cause image blurring.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

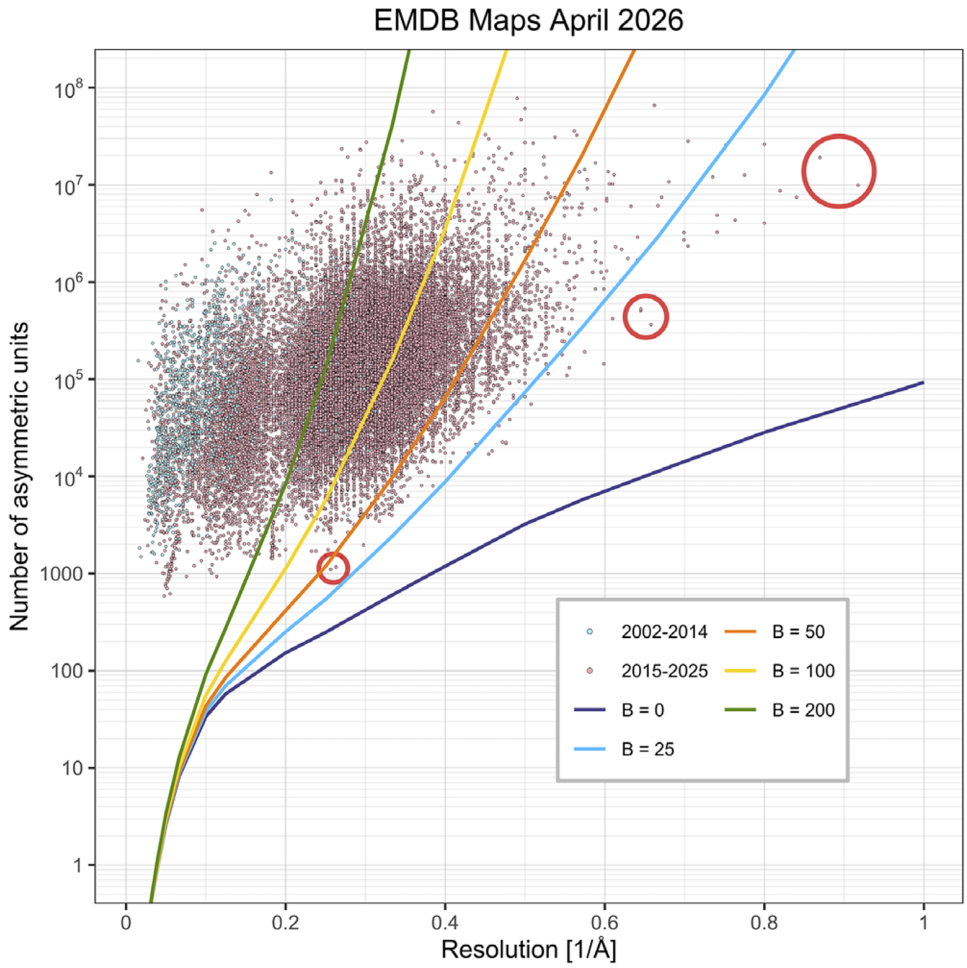


Figure 2. Logarithm of the number of asymmetric units (*Nasym*) versus resolution (1/*d*) plot for all the experimentally determined 3D maps in the EMDB in April 2026. These are assigned to time periods before (2002–2014) and after (2015–2025) the resolution revolution. Three apoferritin maps, one NiFe-hydrogenase map, and three ribosome maps are circled in red. These maps are among those with the highest resolution for the number of asymmetric units in the dataset, so they come closest to the theoretical curve. The theoretical curves for perfect images and computations ($B = 0$) as well as those with *B*-factors of 25, 50, 100, and 200 Å^2 are superimposed. The deposition numbers for the red-circled structures follow. The two highest-resolution structures at 0.9–1.0 Å^{-1} are both apoferritin: EMD-19436 and EMD-11668. The middle group entries are: EMD-27661 0.66 Å^{-1} [NiFe]-hydrogenase, EMD-11121 0.65 Å^{-1} apoferritin, and EMD-15793 0.65 Å^{-1} ribosome. The low-resolution group entries are: EMD-25405 0.26 Å^{-1} ribosome and EMD-25415 0.26 Å^{-1} ribosome. The manual curation described in the Appendix was also applied to the data used in this figure. Additionally, entries with fewer than 1,000 asymmetric units and with an implied *B* factor lower than 25 were also checked.

Table 2. Relation between orientation angular error and *B*-factor

Source publication or dataset	MW	Diameter *D* (Å)	Angular error $\Delta\theta$ (°)		*B*-factor (Å^2)	Resoln (Å)
			Tilt pairs or estimated	RELION		
Rosenthal and Henderson, 2003	E2CD 1.5 MDa	245	5.0		700, 1,000	8.7
Henderson *et al.*, 2011	70S ribo 2.6 MDa	250	4.0		450	11
Henderson *et al.*, 2011	β-gal 450 kDa	130	10.0		700	6.9
Henderson *et al.*, 2011	CAV 2.7 MDa	255	2.5		180	11
Takanori apoF 5,000 ptcls	apoF 450 kDa	137	2.0 (realistic)	0.2 (RELION)	45	1.9, 1.2
catT Krios dataset	catT 240 kDa	118	2.5 (realistic)	0.3 (RELION)	50	1.73
catL Krios dataset	catL 240 kDa	118	3.0 (realistic)	0.4 (RELION)	45	1.89
Fromm et al., 70S ribosome	70S ribo 3 MDa	250	0.9 (estimated)		22	1.55
Naydenova et al., DPS	DPS 200 kDa	100	3.5 (realistic)	0.7 (RELION)	69	1.94

The formula used in Henderson *et al.* (2011) to relate angular error to the *B*-factor contribution was $B = (\Delta\theta \cdot D)^2/2{,}200$, where *D* is particle diameter in Ångstroms and $\Delta\theta$ is measured in degrees. There is a full derivation of the equation, assuming a spherical protein, at the end of the Materials and Methods section of that paper.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

It is possible to calculate the effect of the angular error $\Delta\theta$ in particle orientation on the effective *B*-factor in the resulting map by using the formula from Henderson *et al.* (2011), namely.

$$B = (\Delta\theta \cdot D)^2/2{,}200 \quad (2)$$

where D is the particle diameter in Ångstroms and $\Delta\theta$ is measured in degrees. In practice it is not simple to obtain an accurate value for the angular error, but there are several ways to try to estimate it.

Single-particle cryoEM programs such as RELION (Kimanius *et al.*, 2021), cisTEM (Grant *et al.*, 2018), or cryoSPARC (Punjani *et al.*, 2017) estimate the accuracy of the angular parameter fit by comparing internally the shape of the function being minimized with the fluctuating error during sampling. Tilt pair analysis (Rosenthal and Henderson, 2003; Henderson *et al.*, 2011; Russo and Passmore, 2014; Wasilewski and Rosenthal, 2014) gives experimental errors from successive images of the same particles recorded with known exact angular differences. These experimental estimates calibrated using the known tilt angles are always bigger than those estimated internally during processing. Tilt pair analysis of larger structures where the angular accuracy is higher revealed the major reason for this discrepancy. The particles and the thin film of ice move during the exposures by angles of up to 4 degrees due to beam-induced motion (Henderson *et al.*, 2011; Campbell *et al.*, 2012). The translational component of this beam-induced motion is removed by motion corrections (Zheng *et al.*, 2017) but the angular change is not normally treated, though it has been done occasionally (Campbell *et al.*, 2012; Dimitry Tegunov apoF, unpublished)). Consequently, the tilt pair analysis error estimate is a more realistic value for the angular uncertainty that gives rise to blurring in real single-particle images. These two measures of angular accuracy are listed in separate columns in Table 2.

An estimate of the overall *B*-factor that describes the real experimental data can be calculated from the fall-off of map structure factors with spatial frequency squared ($1/d^2$) or from a Rosenthal plot obtained through a series of map calculations using dataset subsets including, for example, 1/2, 1/4, 1/8, 1/16th of the full particle dataset (e.g. McMullan *et al.*, 2023) showing how $\ln(N_{asym})$ varies with $1/d^2$. Finally, *B*-factors can also be estimated by refinement of an atomic model against the experimental cryoEM potential maps (e.g. Nicholls *et al.*, 2018; Terwilliger *et al.*, 2018; Joseph et al., 2020).

Thus, the theoretical plots in Figure 2 should be viewed keeping in mind the difficulty in determining an absolute value for the *B*-factor of the particle images that contribute to the final map. There is also the well-known observation that cryoEM structures of large complexes like the ribosome (Maji *et al.*, 2017) or spliceosome (Nguyen *et al.*, 2016) often have one or more domains that are more highly disordered, so there is often a need for application of a local model *B*-factor (Beton *et al.*, 2024).

It is also clear that the angular error will be smaller for larger particles such as ribosomes or viruses, and much bigger for smaller particles that are near the limit of detectability. To take this into account, Figures 1*c* and 2 include plots of the number of particles versus resolution as a function of the *B*-factor due to all sources of contrast loss. These plots show that smaller particles need more images to reach the same resolution or conversely that maps of smaller structures typically have lower resolution, as observed (Patwardhan *et al.*, 2025).

For example, good 3D maps of apoferritin, with a MW of 500 kDa, can have *B*-factors using resolution-dependent dose-weighted movie frames as low as 30 Å^2, whereas DPS and catalase, with MWs around 240 kDa, have *B*-factors around 60 Å^2. Structures of molecules with MW below 100 kDa have higher *B*-factors. The smallest structure in the EMDB determined by single-particle cryoEM is Ara2, to which the authors chose to apply a sharpening *B*-factor of 90 Å^2, but we expect the Rosenthal *B*-factor to be higher. Its resolution of 2.6 Å from over 600,000 asymmetric units places it in the most populated region in Figure 2.

The Electron Microscopy Data Bank (EMDB): Number of particles currently required in practice

In a recent paper, Patwardhan *et al.* (2025) examined all the three-dimensional maps produced by single-particle cryoEM that have been deposited in EMDB up to 2024 comparing resolution with particle size. They noted that there is an optimum size of structure that yields the best resolution. Structures of smaller proteins have lower resolution due to higher alignment *B*-factors. Structures of larger proteins also tend to have lower resolution. This is likely due to more flexibility so that the focused alignment needed to align the domains of larger complexes suffers from the same limitation as for smaller structures. Basically, the amount of information in the image of each domain is too small for accurate alignment, again resulting in higher *B*-factors and lower resolution. The optimal size for single-particle cryoEM at present falls in the range 0.2 to 3.5 MDa with the peak at 0.5 MDa. In this range, the particle orientations are accurately determined without too much handicap from flexibility. There is also the possibility of loss of intrinsic symmetry due to perturbations during specimen preparation, such as squashing, bending, or partial denaturation at the air-water interface. This possibility has not been systematically investigated.

In this review, we extend this examination of deposited structures to find the relationship between the number of asymmetric units and the resolution achieved in current practice (Figure 2). In this plot, the particle size affects the location of individual structures indirectly through its effect on orientation accuracy, which feeds through into the *B*-factor, which describes the blurring, and the need for more images.

The distribution of structures shown in Figure 2 shows how the resolutions have improved since the 'Resolution Revolution' (Kühlbrandt, 2014) from typically 5–10 to 3 Å or better, due to improved instrumentation and image analysis software. A few of the best structures indicated with red circles are either apoferritin, rigid enzymes, or ribosomes, which have optimum size or additional advantages such as higher contrast from RNA and slower radiation damage to nucleic acids. These are within a factor of 3 of the best that can ever be expected according to theory. It is likely that about half of this gap between theory and practice arises because all the structures use bright-field defocus phase contrast to create the signal. This means that the contrast transfer function (CTF) has periodic zeros with only some Fourier components being recorded with near-maximum contrast. The availability of a perfect Zernike quarter-wave phase plate would roughly double the signal-to-noise ratio on average with an even larger increase at low resolution (Schwartz et al., 2019; Remis *et al.*, 2024).

On the other hand, the centroid of the distribution in Figure 2 falls around 3 Å resolution for 100,000 asymmetric units, which is about 100 times more images than would be necessary if they were of higher quality. A significant proportion of structures still required over 1,000 times more data than strictly necessary. There are likely to be many reasons for this, and these problems will undoubtedly be the target of future technical improvements.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

Finally, there are many structures that are apparently well-behaved in solution but never make it onto grids because of specimen preparation problems among which interaction with the air-water interface may be the most important. These will all be targets for future improvements.

Combined influence of particle size and shape

In studying the number of particles required to reach a particular resolution, a related question is whether this is influenced by the size and shape of the macromolecule being studied. The earlier review (Patwardhan *et al.*, 2025) explored the relationship between particle size and resolution, but the number of particles and the shape were not investigated explicitly. The Appendix presents an analysis used to make Figure 3. Briefly, a new parameter called PSR is defined as the product of the number of particles, the symmetry, and the resolution achieved, and this is plotted against the particle width defined as the cube root width of a three-dimensional ellipsoid that was fitted to structures in EMDB. A small value of PSR indicates more efficient data collection. The plot makes use of 6% of the EMDB deposited maps with a curve generated to define the lower envelope of the distribution, chosen to reflect the best structures at each particle size in cube root width bins of 20 Å. Figure 3 shows that PSR decreases until it reaches a minimum for a particle size of ~190 Å.

The fitted ellipsoid is also used to define the asphericity (Θ), which describes the deviation of the particle's shape from a sphere. The Appendix shows that more particles are needed to achieve a comparable resolution when asphericity is large. Many of these are oblate or ring-like structures with high cyclic point group symmetries. In practice these could have more uneven Euler angular distributions due to issues such as preferential orientation from interaction with the air-solvent interface. The analysis confirms that roughly spherical particles in the size range 0.2–3.5 MDa produce the highest resolution maps with the smallest number of particles.

Narrowing the gap between theory and practice

What are the innovations needed to bring us nearer to perfect single-particle cryoEM structure determinations? On the technical side, which is targeted at improving the quality of all the images, there are several developments that should become available in the next few years. Some of these will be expensive but others will not.

The three developments likely to be more expensive are colder specimens such as by using liquid helium temperature microscope stages (Naydenova *et al.*, 2020; Dickerson *et al.*, 2025), laser phase plates (Schwartz et al., 2019; Remis *et al.*, 2024; Petrov *et al.*, 2026) and chromatic aberration correctors (Kabius *et al.*, 2009; Hosokawa *et al.*, 2013; Dickerson *et al.*, 2025; Wu *et al.*, 2025). Each of these may allow improvements in the signal-to-noise ratio of images of up to a factor of two, depending on specimen thickness and target resolution. Less expensive innovations involve using lower-energy electrons where the signal is greater for the same amount of radiation damage (Peet *et al.*, 2019; McMullan *et al.*, 2023) or installing an objective lens with lower chromatic aberration (Alves *et al.*, 2026). Together, all the above may provide a factor of 3–4 improvement. About half of this potential improvement (due to phase plates and Cc correctors) should help to bring theory and practice closer, but the other half (lower temperature and lower

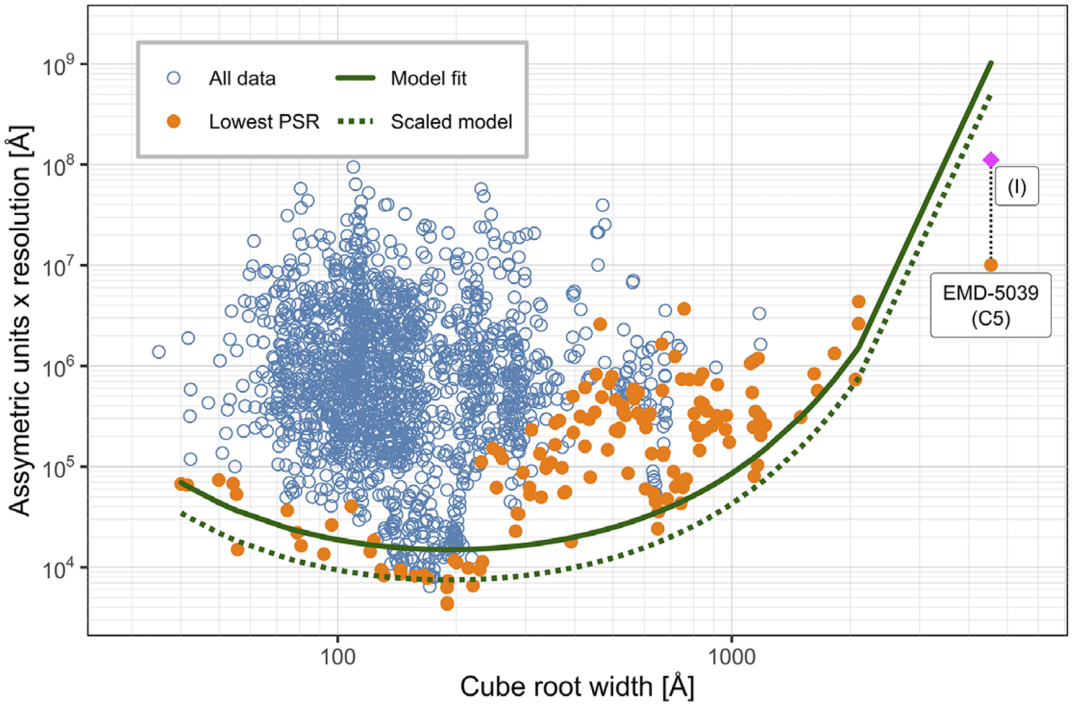


Figure 3. PSR (the product of the number of particles, symmetry, and resolution) varies with particle size using the cube root width as the size measure. PSR decreases toward a minimum around 190 Å, which is dominated by ribosomes and then gradually increases linearly (on the vertical logarithmic scale) as the width increases from there on. For modeling, entries with the three lowest PSR values (shown in orange) in cube root width bins of 20 Å are used to get an even sampling along the curve. There is substantial variation in the PSR within and between subsequent bins. A dotted curve defines the lower envelope. EMD-5039, with a cube root width of 4,550 Å is a reconstruction of a giant mimivirus; it has been plotted using both C5 and icosahedral symmetry, which differ by a factor of 12.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

energy electrons) should extend the theoretical range by improving the useful information in an exposure.

Aside from these entirely technical or electron-optical improvements, the most important goal for future work will be specimen preparation to minimize the problems with denaturation at the air-water interface, or damage at the surface of specimens during thinning using other approaches such as FIB milling (Mahamid *et al.*, 2016; Schaffer et al, 2017), plasma FIB milling (Berger et al, 2023) or CEMOVIS (Al-Amoudi *et al.*, 2004; Eltsov *et al.*, 2018). The gap between the number of particles needed for an average structure and the best obtained so far is likely due to damage during specimen preparation, which should be a tractable problem and not a fundamental limitation (Gusach *et al.*, 2025).

Some future challenges

As illustrated by a recent publication (Dickerson *et al.*, 2025), beam-induced motion in cryoEM samples at temperatures near that of liquid helium is a challenging problem but preparing the specimen in very small holes in a supporting gold matrix can reduce motion to the desired minimum. The study of many biological specimens will not in all cases be able to make use of these small-hole (100 nm) grids. There is also still a strong demand to obtain single-particle cryoEM structures for smaller proteins where improvements in the methodology either in physical specimen preparation or in computation will have the most impact. The pseudo-Brownian motion of water molecules in amorphous ice, which appears to be similar at liquid nitrogen (McMullan *et al.*, 2014) and liquid helium temperature (Dickerson *et al.*, 2025), has a bigger impact on the motion of smaller proteins, so cryoEM structures of smaller proteins will always have an intrinsic *B*-factor and resolution limitation.

Over many years, there have been attempts to show that diffraction-based or scanning transmission (STEM) imaging methods may have some advantages over bright-field phase contrast imaging with or without a perfect Zernike phase plate (Zernike, 1942) in terms of dose efficiency (Dwyer, 2023; Dwyer and Paganin, 2024), but the present consensus is that bright-field phase contrast still offers the highest information content for a given exposure that is limited by radiation damage (Bennemann *et al.*, 2026). It has also been shown experimentally (Peet *et al.*, 2019) that there is a moderate increase in information content obtained for a given amount of radiation damage, by using lower electron energy, such as 100 keV, and possibly even more at even lower electron energy.

Almost all the cryoEM structures in EMDB have significant radiation damage, since even the first frame of typical cryoEM images involves a radiation dose of ~3 MGy. In cases where radiation damage is a particular concern, a useful goal might be to record 10 times more images so that 3D maps at only 0.3 MGy could be calculated from the first frames, ensuring that any motion in the first few frames is minimized by using specimen supports that restrict movement of the water during imaging.

Tomography and crowded environments

Although most of the above discussion has concerned the theoretical foundations of single-particle cryoEM and how to extend its range, a substantial portion of the cryoEM community has already moved on to tackle the broader challenges of tomography or cellular imaging in a crowded environment (Rickgauer *et al.*, 2017; Russo *et al.*, 2022; Lucas *et al.*, 2023; Dickerson *et al.*, 2025). Such specimens usually do not involve highly purified molecules or molecular complexes. The molecules of interest are sometimes present at low abundance yet surrounded by a high concentration of other proteins or complexes. The challenge is to identify the location of the molecules of interest so that the power of averaging or classification can be used to extract structural information. This identification, whether based on a two-dimensional template matching (2DTM) search or on a sub-tomogram feature search, is intrinsically limited by the signal-to-noise ratio in the image or tomogram and is subject to false positives and false negatives as the structures become less abundant and as the environment becomes more crowded. In the last analysis, tomography is limited by the same fundamental signal-to-noise problems that restrict the range of single-particle cryoEM. The effect of a crowded environment where cellular structures are closely packed, such as in the structure of chromatin in the nucleus, is to reduce the strength of the low-resolution Fourier components and remove the contribution of stronger feature in the resolution range below 8–10 Å. In the final two columns of Table 1 for catalase and pyruvate dehydrogenase E2CD, this represents a two- to fivefold signal reduction. The same four technical improvements discussed above (liquid helium temperature, phase-plates, chromatic aberration correction, and lower energy electrons) are likely to help here too.

Conclusion

Single-particle cryoEM has had enormous impact on our ability to determine structures that were beyond the reach of earlier methods. Current developments, including lower electron energy cryoEM, phase plates, lower specimen temperature, and the use of chromatic aberration correction will help to extend the reach of cryoEM, but after that, future improvements will depend on new methods of specimen preparation, which is an important topic of discussion that is beyond the scope of this article. In the meantime, practitioners may find it interesting to find where each new structure falls in comparison to current practice as well as theoretical criteria on the plot shown in Figure 2.

Acknowledgments. We thank Dr Paul Midgley for valuable discussions.

Financial support. A.P. is supported by the Wellcome Trust (212977/Z/18/Z) and EMBL with funding from its member states. P.B.R. is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (CC2106), the UK Medical Research Council (CC2106), and the Wellcome Trust (CC2106). This work is supported by Medical Research Council Grant No. MC_UP_120117 (C.J.R.) and Wellcome Trust Grant No. 220526/B/20/Z (R.H. and C.J.R.).

Competing interests. There are no conflicts of interest.

References

Al-Amoudi A *et al.* (2004) Cryo-electron microscopy of vitreous sections. *The EMBO Journal* **23**(18), 3583–3588. http://doi.org/10.1038/sj.emboj.7600366.

Alves GM *et al.* (2026) The Dublin lens: a c = 1.0 mm objective lens intended for CryoEM at 100 keV. *Methods in Microscopy* **3**(1), 19–25. http://doi.org/10.1515/mim-2025-0019.

Barron JT (2019) A general and adaptive robust loss function. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4326–4334.

Bennemann F *et al.* (2026) Detective quantum efficiency-based comparison of HRTEM and ptychography phase imaging. *Microscopy and Microanalysis* **32**(2), ozag018. http://doi.org/10.1093/mam/ozag018.

Berger C *et al.* (2023) Plasma FIB milling for the determination of structures in situ. *Nature Communications* **14**(1), 629. http://doi.org/10.1038/s41467-023-36372-9.

Beton JG *et al.* (2024) Cryo-EM structure and B-factor refinement with ensemble representation. *Nature Communications* **15**(1), 444. http://doi.org/10.1038/s41467-023-44593-1.

Blow DM and Crick FHC (1959) The treatment of errors in the isomorphous replacement method. *Acta Crystallographica* **12**(10), 794–802. http://doi.org/10.1107/S0365110X59002274.

Bradley AJ and Lu SS (1937) The crystal structures of Cr_2Al and Cr_5Al_8. *Zeitschrift für Kristallographie* **96**(1-6), 20–37. http://doi.org/10.1524/zkri.1937.96.1.20.

Brilot AF *et al.* (2012) Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* **177**(3), 630–637. http://doi.org/10.1016/j.jsb.2012.02.003.

Campbell MG *et al.* (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**(11), 1823–1828. http://doi.org/10.1016/j.str.2012.08.026.

Carlson BC (1995) Numerical computation of real or complex elliptic integrals. *Numerical Algorithms* **10**(1), 13–26. http://doi.org/10.1007/BF02198293.

Chau J (2025) gslnls: GSL multi-start nonlinear least-squares fitting. R package version 1.4.1. http://doi.org/10.32614/CRAN.package.gslnls; https://CRAN.R-project.org/package=gslnls.

Crowther RA *et al.* (1970) The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society A* **317**(1530), 319–340. http://doi.org/10.1098/rspa.1970.0119.

Debye P (1913) Interferenz von Röntgenstrahlen und Wärmebewegung (Interference of X-rays and thermal motion). *Annalen der Physik.* **348**(1), 49–92. http://doi.org/10.1002/andp.19133480105.

Dickerson JL *et al.* (2025) Reducing the effects of radiation damage in cryo-EM using liquid helium temperatures. *Proceedings of the National Academy of Sciences of the United States of America* **122**(17), e2421538122. https://doi.org/10.1073/pnas.2421538122.

Dwyer C (2023) Quantum limits of transmission electron microscopy. *Physical Review Letters* **130**(5), 056101. http://doi.org/10.1103/PhysRevLett.130.056101.

Dwyer C and Paganin DM (2024) Quantum and classical Fisher information in four-dimensional scanning transmission electron microscopy. *Physical Review B* **110**(2), 024110. http://doi.org/10.1103/PhysRevB.110.024110.

Eltsov M *et al.* (2018) Nucleosome conformational variability in solution and in interphase nuclei evidenced by cryo-electron microscopy of vitreous sections. *Nucleic Acids Research* **46**(17), 9189–9200. http://doi.org/10.1093/nar/gky670.

Fowweather F and Hargreaves A (1950) The crystal structure of meta-tolidine dihydrochloride. *Acta Crystallographica* **3**(2), 81–87. http://doi.org/10.1107/S0365110X50000239.

Frank J (1975) Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* **1**(2), 159–162. http://doi.org/10.1016/S0304-3991(75)80020-9.

Garman EF (2010) Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica* **D66**, 339–351.

Geman S and McClure DE (1987) Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistics Institute* **LII-4**, 5–21.

Glaeser RM (1999) Review: Electron crystallography: present excitement, a nod to the past, anticipating the future. *Journal of Structural Biology* **128**(1), 3–14. http://doi.org/10.1006/jsbi.1999.4172.

Grant T *et al.* (2018) cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383. http://doi.org/10.7554/eLife.35383.

Gusach A *et al.* (2025) Outrunning protein diffusion to the air-water interface in cryoEM. *Proceedings of the National Academy of Sciences of the United States of America* **122**(43), e2516900122. http://doi.org/10.1073/pnas.2516900122.

Henderson R (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics* **28**(2), 171–193. http://doi.org/10.1017/S003358350000305X.

Henderson R and Glaeser RM (1985) Quantitative analysis of image contrast in electron micrographs of beam-sensitive crystals. *Ultramicroscopy* **16**(2), 139–150. http://doi.org/10.1016/0304-3991(85)90069-5.

Henderson R *et al.* (2011) Tilt-pair analysis of images from a range of different specimens in single particle electron cryomicroscopy. *Journal of Molecular Biology* **413**(5), 1028–1046. http://doi.org/10.1016/j.jmb.2011.09.008.

Hosokawa K *et al.* (2013) Development of Cs and Cc correctors for transmission electron microscopy. *Microscopy* **62**(1), 23–41. http://doi.org/10.1093/jmicro/dfs134.

Joseph AP *et al.* (2020) Comparing Cryo-EM reconstructions and validating atomic model fit using difference maps. *Journal of Chemical Information and Modeling* **60**(5), 2552–2560. http://doi.org/10.1021/acs.jcim.9b01103.

Kabius B *et al.* (2009) First application of Cc-corrected imaging for high-resolution and energy-filtered TEM. *Journal of Electron Microscopy* **58**(3), 147–155. http://doi.org/10.1093/jmicro/dfp021.

Kimanius D *et al.* (2021) New tools for automated cryo-EM single-particle analysis in RELION-4.0. *The Biochemical Journal* **478**(24), 4169–4185. http://doi.org/10.1042/BCJ20210708.

Kühlbrandt W (2014) The resolution revolution. *Science* **343**(6178), 1443–1444. http://doi.org/10.1126/science.1251652.

Lipson H and Cochran W (1966) *The Determination of Crystal Structures*, Vol. **1966**, 3rd Edn. Pub. Bell & Sons, pp. 72–74.

Lucas BA *et al.* (2023) Baited reconstruction with 2D template matching for high-resolution structure determination in vitro and in vivo without template bias. *eLife* **12**, RP90486. http://doi.org/10.7554/eLife.90486.3.

Mahamid J *et al.* (2016) Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**(6276), 969–972. http://doi.org/10.1126/science.aad8857.

Maji S *et al.* (2017) Quantitative characterization of domain motions in molecular machines. *The Journal of Physical Chemistry. B* **121**(15), 3747–3756. http://doi.org/10.1021/acs.jpcb.6b10732.

McMullan G *et al.* (2014) Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156–163. http://doi.org/10.1016/j.ultramic.2014.08.002.

McMullan G *et al.* (2023) Structure determination by cryoEM at 100 keV. *Proceedings of the National Academy of Sciences of the United States of America* **120**(49), e2312905120. http://doi.org/10.1073/pnas.2312905120.

Naydenova K *et al.* (2020) Cryo-EM with sub–1 Å specimen movement. *Science* **370**(6513), 223–226. http://doi.org/10.1126/science.abb7927.

Naydenova K *et al.* (2022) On the reduction in the effects of radiation damage to two-dimensional crystals of organic and biological molecules at liquid-helium temperature. *Ultramicroscopy* **237**, 113512. http://doi.org/10.1016/j.ultramic.2022.113512.

Nguyen THD *et al.* (2016) Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**(7590), 298–302. http://doi.org/10.1038/nature16940.

Nicholls RA *et al.* (2018) Current approaches for the fitting and refinement of atomic models into cryo-EM maps using CCP-EM. *Acta Crystallographica* **74**(Pt 6), 492–505. http://doi.org/10.1107/S2059798318007313.

Oikonomou CM and Jensen GJ (2017) Cellular electron cryotomography: toward structural biology in situ. *Annual Review of Biochemistry* **86**(1), 873–896. http://doi.org/10.1146/annurev-biochem-061516-044741.

Patwardhan A *et al.* (2025) Extending the reach of single-particle cryoEM. *Current Opinion in Structural Biology* **92**, 103005. http://doi.org/10.1016/j.sbi.2025.103005.

Peet M *et al.* (2019) The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules. *Ultramicroscopy* **203**, 125–131. http://doi.org/10.1016/j.ultramic.2019.02.007.

Petrov PN *et al.* (2026) Laser phase plate improves structure determination of small proteins by cryo-EM. *Science* **393**(6807), 195–199. http://doi.org/10.1126/science.aeh0665.

Prince, E. (2006) *International Tables for Crystallography Volume C, Mathematical, Physical and Chemical Tables*, section 4.3, p. 263. Electron diffraction. ISBN: 978-1-4020-1900-5. Third edition publisher: John Wiley & Sons on behalf of the International Union of Crystallography (IUCr).

Punjani A *et al.* (2017) cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**(3), 290–296. http://doi.org/10.1038/nmeth.4169.

Remis J *et al.* (2024) Cryo-EM phase-plate images reveal unexpected levels of apparent specimen damage. *Journal of Structural Biology* **216**(4), 108150. http://doi.org/10.1016/j.jsb.2024.108150.

Rickgauer JP *et al.* (2017) Single-protein detection in crowded molecular environments in cryo-EM images. *eLife* **6**, e25648. http://doi.org/10.7554/eLife.25648.

Rosenthal PB and Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single particle electron

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

cryomicroscopy. *Journal of Molecular Biology* **333**(4), 721–745. http://doi.org/10.1016/j.jmb.2003.07.013.

Roth WL and Harker D (1948) The crystal structure of Octamethylspiro [5″5] Pentasiloxane' rotation about the ionic silicon-oxygen bond. *Acta Crystallographica* **1**(1), 34–42. http://doi.org/10.1107/S0365110X48000077.

Russo CJ and Passmore LA (2014) Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**(6215), 1377–1380. http://doi.org/10.1126/science.1259530.

Russo CJ *et al.* (2022) Cryomicroscopy in situ: what is the smallest molecule that can be directly identified without labels in a cell? *Faraday Discussions* **240**, 277–302. http://doi.org/10.1039/D2FD00076H.

Saxton WO and Frank J (1977) Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy* **2**(2-3), 219–227. http://doi.org/10.1016/S0304-3991(76)91385-1.

Schaffer M *et al.* (2017) Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *Journal of Structural Biology* **197**(2), 73–82. http://doi.org/10.1016/j.jsb.2016.07.010.

Schwartz O *et al.* (2019) Laser phase plate for transmission electron microscopy. *Nature Methods* **16**, 1016–1020. http://doi.org/10.1038/s41592-019-0552-2.

Stagg SM *et al.* (2014) ResLog plots as an empirical metric of the quality of Cryo-EM reconstructions. *Journal of Structural Biology* **185**(3), 418–426. http://doi.org/10.1016/j.jsb.2013.12.010.

Terwilliger TC *et al.* (2018) Map segmentation, automated model-building and their application to the Cryo-EM model challenge. *Journal of Structural Biology* **204**(2), 338–343. http://doi.org/10.1016/j.jsb.2018.07.016.

Vinothkumar, K.R. and Henderson, R. (2016) Single particle electron cryomicroscopy: trends, issues and future perspective. *Quarterly Reviews of Biophysics* **49**, e13, 1–25.

Waller I (1923) Zur, Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen (On the question of the influence of thermal motion on the interference of X-rays). *Zeitschrift für Physik A* **17**, 398–408.

Wasilewski S and Rosenthal PB (2014) Web server for tilt-pair validation of single particle maps from electron cryomicroscopy. *Journal of Structural Biology* **186**(1), 122–131. http://doi.org/10.1016/j.jsb.2014.02.012.

Wilson AJC (1949) The probability distribution of X-ray intensities. *Acta Cryst* **2**(5), 318–321. http://doi.org/10.1107/S0365110X49000813.

Wu J *et al.* (2025) Chromatic aberration (Cc) corrected cryo-EM: the structure of pseudorabies virus (PRV) using both zero-loss and energy loss electrons. *Ultramicroscopy* **276**, 114182. http://doi.org/10.1016/j.ultramic.2025.114182.

Zernike F (1942) Phase contrast, a new method for the microscopic observation of transparent objects. *Physica* **9**(7), 686–698. http://doi.org/10.1016/S0031-8914(42)80035-X.

Zheng SQ *et al.* (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**(4), 331–332. http://doi.org/10.1038/nmeth.4193.

Appendix: Analysis of how the optimum particle number varies with structure size and shape in single-particle cryoEM

Introduction

Although the size and shape of a structure are not metrics available in EMDB, they can be computationally estimated from the main maps of EMDB entries. For any given size or shape the number of asymmetric units among EMDB entries may vary by several orders of magnitude. The spread can be attributed to various factors such as suboptimal data collection and data processing, molecular heterogeneity, and flexibility. Hence, the analysis focuses on entries with the lowest numbers of asymmetric units for any given size or shape, that is a small selection of the *circa* 47 K single-particle entries (as of April 2026) in EMDB.

As shown in the main text, see for example Eqs. (1) and (2), and Tables 1 and 2, the number of asymmetric units depends on several parameters including the *B* factor, the critical dose, neon elastic scattering factor, and the Guinier amplitude. There is also a direct dependence on resolution d via the Crowther term ($\pi D/d$) and an indirect dependence via the other parameters. Every single-particle entry in EMDB has a resolution estimate, which was used as the abscissa in Figure 2 and even though the resolution may have been measured using different methods, over time the field has converged to using the 0.143 criteria (Rosenthal and Henderson, 2003). For the size and shape analysis we use the product of the number of particles, symmetry order, and resolution (PSR) as the dependent variable; all three parameters can be obtained from EMDB metadata, and the PSR relates to Eq. (1) if the $1/d$ term on the right-hand side is moved to the left and combined with Nasym, that is PSR = Nasym × *d*.

Methods

Initial data selection

Scripts written in R (https://www.r-project.org/) were used to download EMDB entry metadata and to analyse and filter entries. 2,872 entries comprising ~6% of the ~47 K released single-particle entries were selected. Map data for the selected entries were downloaded using the download_emdb Python (https://www.python.org/) program in the em_map_utils package (https://github.com/ardanpat/em-map-utils; https://pypi.org/project/em-map-utils/).

Procedure for selection of entries to download

(a) Use the EMDB API to download the metadata for all released single-particle EMDB entries (100 metadata entries are downloaded at a time until there are no more entries): https://www.ebi.ac.uk/emdb/api/search/structure_determination_method%3A%22singleParticle%22%20AND%20current_status%3AREL?rows=100

(b) Check the metadata, keeping only entries that have an aggregation state of 'particle' and valid values for the number of particles used as well as the resolution, and removing entries that are marked as composite maps, that are negatively or positively stained and entries missing a PubMed ID for the associated publication.

(c) To get a relatively even sampling of different symmetries and resolutions, for every point group symmetry, data were grouped by resolutions 'bins' with a width of 1/10 of the base 10 logarithm of the reciprocal resolution. Data in each resolution bin were sorted by PSR in ascending order, and a number (10–15) of the entries with the lowest PSR values in each bin were selected for further analysis.

Manual data curation

Several manual checks were performed on the selected entries resulting in a further reduction to 1,857 entries or ~ 4% of the released single-particle entries.

Manual checks performed

(a) Omit entries which have associated publications that cannot be accessed, e.g., due to paid subscriptions not supported by EMBL-EBI and the unavailability of open-source copies.

(b) Cross-check the number of particles, point group symmetry, and resolution with the associated publication. If there are discrepancies, update values to match paper or omit entry from analysis if no corroboration can be found.

(c) If the entry is a part of a structure, e.g., part of a phage, omit it from the analysis.

(d) If the structure is bound to a nano-disc, liposome, or micelle with pronounced contrast of these peripheral structures in the reconstruction, omit them from the analysis.

(e) Omit any misannotated entries that managed to slip through the automated selection process, e.g. composite, negatively stained, tomography, and helical structures.

Size and shape measurement

The size and shape of structures were determined using the structure_size_and_shape program in the em_map_utils package using the following procedure:

(a) Threshold the map so that any map values below the threshold are set to zero, and the threshold value is subtracted from any map values above the threshold.

(b) Calculate the three principal axes of the map value distribution, then rotate the map so that the principal axes align with the box axes.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

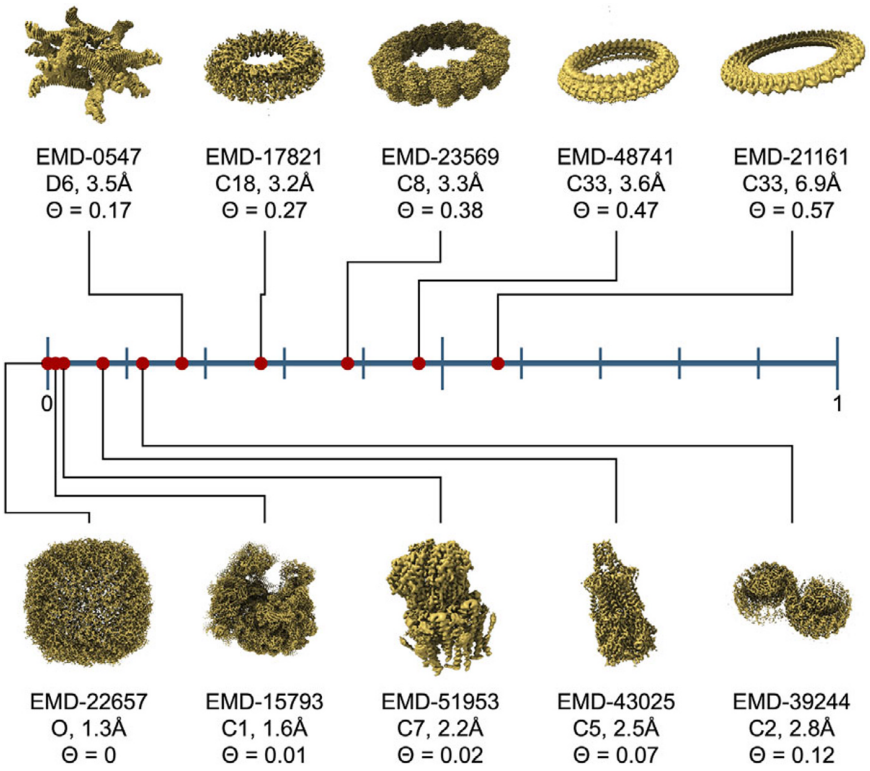


Figure A1. Example structures from EMDB showing how the asphericity metric (Θ) varies with the shape of macromolecules. The range of Θ varies from 0 to 1.

(c) Calculate three 1D line projections of the map value distribution along the three box axes by first finding 2D projections and then reducing the 2D projection to a 1D one, e.g., $p(x,y,z) \rightarrow p(x,z) \rightarrow p(z)$.
(d) Estimate the widths of the line profiles and use them as measures of the extent of the structure along the three axes.

Cube root width

The volume of an ellipsoid spanned by the three width measurements is given by:

$$V = \frac{1}{6}\pi w_1 w_2 w_3 \tag{A1}$$

The volume of a sphere with the equivalent volume is given by:

$$V = \frac{1}{6}\pi w_r^3 \tag{A2}$$

Combining Eqs. (A1) and (A2) one can express the cube root width in terms of the three principal axes widths:

$$w_r = \sqrt[3]{w_1 w_2 w_3} \tag{A3}$$

The cube root width provides a measure of the overall size of the structure.

Asphericity

The surface area to volume ratio of an ellipsoid is at a minimum when the three axes are equal in length, that is for the special case of a sphere. As the axes become more dissimilar, e.g. the ellipsoid becomes more oblate or prolate, the ratio increases. One can therefore define a simple measure for the shape of the molecule solely by using the widths along the principal axes and the volume-to-surface area ratio. Here we define 'asphericity' as:

$$\theta = 1 - \left(\frac{V/S}{V_0/S_0}\right) \tag{A4}$$

where V and S are the volume and surface area for an ellipsoid, and V_0 and S_0 are the same for a sphere with the cube root width calculated from the ellipsoid axes. In terms of widths along the principal axes and the cube root width, the ratios are defined as:

$$\frac{V_0}{S_0} = \frac{w_r}{6} \tag{A5}$$

and

$$\frac{V}{S} = \frac{1}{6R_G\left(\frac{1}{w_1^2}, \frac{1}{w_2^2}, \frac{1}{w_3^2}\right)} \tag{A6}$$

where R_G is the completely symmetric elliptic integral of the second kind (Carlson, 1995) as implemented by the 'ellprg' function in the SciPy package (https://scipy.org/). The asphericity can then be expressed as a dimensionless value between 0 and 1:

$$\theta = 1 - \frac{1}{w_r R_G\left(\frac{1}{w_1^2}, \frac{1}{w_2^2}, \frac{1}{w_3^2}\right)} \tag{A7}$$

Figure A1 shows how the asphericity varies with the shape of macromolecules in EMDB.

Data analysis and model fitting

Data analysis and model fitting were mainly done using R script. The gslnls package (Chau, 2025) was used for non-linear least squares fitting. The Barron loss function (Barron, 2019) was used for fitting with $\alpha = -2$ (which is essentially a Geman–McClure loss function (Geman and McClure, 1987), which helps reduce the effect of outliers. The scaling parameter c of the loss function was adjusted empirically by visually observing the overlap of the fit to the data as well as checking modeling statistics.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

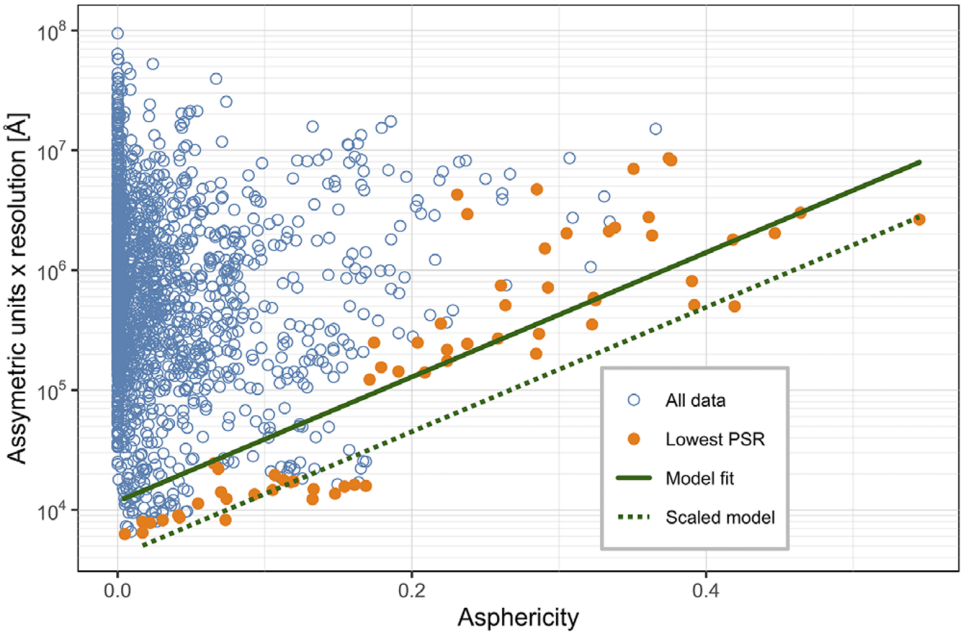


Figure A2. Variation of PSR (Product of the number of particles, symmetry and resolution) with asphericity. Points represent EMDB structures: the orange ones were used for model fitting. The solid green line is the model fit and the dashed line is identical except for a manually tweaked scaling factor to make the curve skim the lower envelope of the data points.

Results

Width

Figure 3 shows how the PSR varies with the cube root width. PSR decreases toward a minimum around 190 Å, which is dominated by ribosomes and then gradually increases linearly (on the logarithmic PSR scale) as the width increases from there on. For modeling, entries with the three lowest PSR values (shown in orange) in cube root width bins of 20 Å are used to get an even sampling along the curve. There is substantial variation in the PSR within and between subsequent bins, but still a low-order curve may be traced to define the lower envelope of the variation:

$$PSR = C_1 \, exp\left[(w + C_2) \cdot \left(\frac{1}{w} + C_3\right)\right] \tag{A8}$$

The estimates and standard errors for the fitting parameters are C_1 = 1,526 (60), C_2 = 98 (10), and $C_3 = 2.666 \times 10^{-3}$ (4×10^{-6}). The Barron loss function was used for fitting with the parameters $\alpha = -2$ and $c = 0.06$. The model fitting process by its nature tries to place the fitted curve so that points to which it has been fitted are distributed equally on both sides of it. The scaled model is equivalent to the fitted model with the exception that the C_1 constant has been manually adjusted to 763 (50% of the original value) so that the curve skims the bottom envelope of the data.

The minima of the fitted and scaled models occur at:

$$w_{min} = \sqrt{\frac{C_2}{C_3}} \tag{A9}$$

which for the given constants is 192 Å. The points defining the minima in the chart consist chiefly of ribosomes, and this width is consistent with the size of a ribosome. If one assumes a protein density of ~0.81 Da $Å^{-3}$ and a spherical protein, a width of 192 Å corresponds to a molecular weight of ~3 MDa, which is also consistent with that of ribosomes.

There is one data point (EMD-5039) with a cube root width of 4,550 Å that lies well below the curve by two orders of magnitude. EMD-5039 is a reconstruction of the giant mimivirus using C5 symmetry at 65 Å resolution. It was in fact deposited as having icosahedral symmetry despite the title stating that it has C5 symmetry and the map clearly showing a non-icosahedral fivefold feature, and hence manually modified in the analysis as having C5 symmetry. Had the original icosahedral symmetry been kept for this entry, the PSR would be 12 times higher than its current value and closer to the curve (although still almost an order of magnitude better). In common with many virus entries in EMDB, EMD-5039 was first processed using a higher point group symmetry (icosahedral in this case) and subsequently relaxed to a lower symmetry, and the results indicate that in such cases remnants of the originally applied symmetry still are manifested and visible in the final map.

Given the dependence of the *B* factor exponential and other factors on the right-hand side of Eq. (1) on resolution, it is reasonable to inquire whether Eq. (A8) should have some inverse proportional dependence on resolution, d^{-a}. Indeed, this was tested for values of '*a*' ranging from 0.25 to 2 but did not result in any visible improvement of the fit.

Asphericity

Figure A2 shows the variation of PSR with asphericity. For modeling, entries with the three lowest PSR values (shown in orange) in asphericity bins of 0.02 are used to get an even sampling along the curve. The lower envelope of the points shows an approximate linear relationship between the logarithm of the PSR and asphericity:

$$PSR = C_1 \, exp(C_2 \cdot \theta) \tag{A10}$$

The estimates and standard errors for the fitting parameters are C_1 = 11,700 (387) and C_2 = 11.95 (0.07). The Barron loss function was used for fitting with the parameters $\alpha = -2$ and $c = 0.5$. In Fig. A2, the scaled model is equivalent to the fitted model with the exception that the C_1 constant has been manually adjusted to 4,120 (35% of the original value) so that the curve skims the bottom envelope of the data. As with the cube root width fit, attempts to add a resolution term d^{-a} where *a* was varied between 0.25 and 2 did not result in any visible improvement of the fit.

Can the width and asphericity models be combined?

Given that both models are exponentials it is natural to inquire whether they can be combined into one model:

$$PSR(w_r, \theta) = C_1 \, exp\left[(w + C_2) \cdot \left(\frac{1}{w} + C_3\right) + C_4 \cdot \theta\right] \tag{A11}$$

This requires that the assumption of separability is valid:

$$PSR(w_r, \theta) = f(w_r) g(\theta) \tag{A12}$$

If one considers the PSR measurements as a sparse matrix of width and asphericity values then one way to test whether the data is consistent with this assumption is to

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press

check that the matrix has rank 1, that is all rows are scalar multiples of one another, and the same applies to all the columns. A matrix is rank 1 if a singular value decomposition yields a solitary non-zero singular value. To reduce the sparseness of the matrix, the width and asphericity values can be binned, and for each unique bin combination the lowest PSR can be used to populate the matrix. This analysis was performed using an asphericity bin size of 0.05 and a width bin size of 100 Å resulting in a sparse matrix with 48 non-zero elements out of 506 (90% sparseness). In contrast to expectations for a separable function, there were around 6 significant singular values out of 11 instead of one. However similar results were seen in simulations performed using a defined separable function as the level of noise and matrix sparseness were increased. The conclusion is that more data (i.e. more structures) are needed before Eq. (A11) can be confirmed.

https://doi.org/10.1017/S0033583526100249 Published online by Cambridge University Press