1 Supplemental Material

1.1 Peak adjustment at cluster boundaries

When clusters combine SrMise performs a recursive search for peaks, but there are two separate recursive cases. In the first case, which is performed when at least one peak exists, the peaks in the two clusters are jointly fit and extraction is then performed on the residuals. In the second case, which occurs when each of the adjacent clusters contains at least one peak, no fit is performed. However, if the total value of the peaks at the the boundary between clusters exceeds the value of the data at that point, then the peak parameters in the adjacent clusters are adjusted so that they equal the experimental value. Recursion is then performed over the residuals. Details of adjusting the peaks for the second case follow.

Suppose we have adjacent clusters A, B with n_A, n_B peaks and where the ith peak in each is given by the peak function A_i, B_i. The value of the PDF at the boundary $r = r_b$ is approximated by linear interpolation

$$r_b = \frac{1}{2}(r_{\text{edgeA}} + r_{\text{edgeB}})$$
$$G(r_b) = \frac{1}{2}(G(r_{\text{edgeA}}) + G(r_{\text{edgeB}}))$$

where r_{edgeA} and r_{edgeB} are the r-coordinates of the data point in each cluster nearest to the other cluster.

If $\sum_i A_i(r_b) + \sum_j B_j(r_b) > G(r_b)$ the peak parameters are adjusted to meet the following conditions, assuming a solution exists.

1. The sum of the adjusted peaks at r_b is $G(r_b)$.
2. The relative contribution from each peak at r_b is unchanged.
3. The position and height of the local maximum of each peak (ignoring any effect from the baseline) is unchanged.

Define the constant $s > 1$, which determines how the contribution of the peaks must change at r_b.

$$s = \frac{\sum_i A_i(r_b) + \sum_j B_j(r_b)}{G(r_b)} \tag{1}$$

Given an initial Gaussian over r peak function

$$f(r; \mu_1, \sigma_1, a_1) = \frac{a_1}{r\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(r-\mu_1)^2}{2\sigma_1^2}\right) \tag{2}$$

the new parameters (μ_2, σ_2, a_2) are given by the solution to following system of equations.

$$\begin{aligned} f(r_b; \mu_1, \sigma_1, a_1) &= s f(r_b; \mu_2, \sigma_2, a_2) \\ r_{\max} \text{ s.t. } \left.\frac{df(r; \mu_1, \sigma_1, a_1)}{dr}\right|_{r=r_{\max}} &= \left.\frac{df(r; \mu_2, \sigma_2, a_2)}{dr}\right|_{r=r_{\max}} = 0 \\ f(r_{\max}; \mu_1, \sigma_1, a_1) &= f(r_{\max}; \mu_2, \sigma_2, a_2) \end{aligned}$$

A solution exists if $\mu_1^2 \geq 4\sigma_1^2$, equivalent to the physically trivial condition that the peak function has a local maximum. Defining the intermediate quantities

$$\begin{aligned} r_{\max} &= \frac{1}{2}(\mu_1 + \sqrt{\mu_1^2 - 4\sigma_1^2}) \\ k &= \log(s) + \frac{(r_b - r_{\max})(r_b - 2\mu_1 + r_{\max})}{2\sigma_1^2} \end{aligned}$$

the solution is

$$\begin{aligned} \sigma_2 &= \sqrt{\frac{r_{\max}(r_b - r_{\max})^2}{(r_b - r_{\max} + k r_{\max})}} \\ \mu_2 &= \frac{\sigma_2^2 + r_{\max}^2}{r_{\max}} \\ a_2 &= a_1 \frac{\sigma_2}{\sigma_1} \exp\left(-\frac{(r_{\max} - \mu_1)^2}{\sigma_1^2}\right) / \exp\left(-\frac{(r_{\max} - \mu_2)^2}{2\sigma_2^2}\right). \end{aligned}$$

The conversion to SrMise's internal parameterization is straightforward. In fact, any peak initially expressible in that width-limited parameterization remains expressible after this transformation because the peak's full-width at half-maximum decreases for $s > 1$.

1.2 PDF Baseline of PbTe nanoparticle

For a well normalized PDF the baseline is $4\pi\rho_0\gamma_0(r)$, as in the main text. We approximated the baseline for the PbTe nanoparticle by treating it as a simple sphere. Given a sphere of radius R, $\gamma_0(r) = 1 - \frac{3r}{4R} + \frac{r^3}{16R^3}$ on the interval $[0, R]$, and 0 elsewhere (Guinier *et al.*, 1955). Since the number density may not be known, and a PDF is not necessarily well normalized, SrMise subsumes the leading constants into a fittable scale factor.

The particular baseline used during the SrMise trials were $R = 11.36$ Å with scale factor of $\sim$.281. These ad hoc values were determined visually by varying the baseline over a plot of the experimental PDF, and we do not justify them beyond the barest physical considerations. This baseline is adequate only for rough exploration, study of gross features, and the benchmarking performed. Next we consider a PDF where the baseline is estimated much more carefully.

1.3 Interparticle Correlation for C_{60}

The PDF of the C_{60} sample has a substantial contribution from interparticle correlations. It is well known that in the solid phase at room temperature these form an FCC lattice with lattice constant ≈14.11 Å, and at each lattice site the molecules spin in a random orientation (Heiney *et al.*, 1991). The interparticle correlation therefore averages to a PDF much like that of atomic FCC, while the precise atomic correlations with a single molecule are preserved. The diameter of C_{60} is roughly 7 Å, but because the nearest neighbor distance is about 10 Å, interparticle correlations become evident past approximately 3 Å, well within the region of interest for a single molecule.

For the present study we estimated the interparticle contribution to the PDF by modeling C_{60} as hollow spheres of fixed density and calculated the radial distribution function (RDF) analytically in terms of the inner and outer radius, as well as the lattice constant. The assumption is that the random orientation and spinning of the molecules makes this a good approximation to the effect on the experimental PDF. This form is clearly presaged on some prior knowledge of the molecule's structure, but as discussed in the main text such knowledge (or a reasonable guess) is essentially required in all but the simplest cases. The strength of the structural assumptions in this case permits a detailed model with very few free parameters.

The calculation proceeds in three steps. First, the RDF between two non-overlapping sphere is found, followed by that for two hollow spheres, and finally an FCC lattice of these spheres are fit to the data.

1.3.1 RDF between two spheres

Given two spheres with radii r_1 and r_2, respectively, and whose centers are separated by the distance $d \geq r_1 + r_2$, calculate the RDF $R(r)$. Assume the spheres consist of the same material of homogenous density, fixed to 1 for simplicity. Due to the spherical symmetry the RDF depends only on the distance between them, and so no angle-averaging or other special consideration of orientation is required. The strategy is selecting a point $\vec{p} = \vec{p}(p, \theta, \phi)$ within the sphere 1 and calculating the partial contribution $R(r, p, \theta)$ to the full RDF between $\vec{p}$ and points within sphere 2 a distance r away. Integrating over all points within sphere 1 will then obtain RDF. The coordinate system for $\vec{p}$ is the standard spherical-polar coordinates with origin at the center of sphere 1 and the positive z-axis along the line which connects it to the center of sphere two.

By the definition of the radial distribution function

$$\begin{aligned} R(r)dr &= \int_0^{r_1} \int_0^{2\pi} \int_0^{\pi} R(r, p, \theta) p^2 \sin(\theta)\, d\theta\, d\phi\, dp\, dr \\ &= 2\pi \int_0^{r_1} \int_0^{\pi} R(r, p, \theta) p^2\, \sin(\theta)\, d\theta\, dp\, dr. \end{aligned} \tag{3}$$

Denote the distance between $\vec{p}$ and the center of sphere 2 as d_p. The line between these points defines the primed coordinate system (r, θ', ϕ'). Therefore, $R(r, p, \theta)$ is the contribution from thin spherical shells of thickness dr which intersect sphere 2. For any given shell this is dr times the surface area of a sphere of radius r where it passes through sphere 2.

Now, the intersection of any two spheres A and B of radius a and b and separated by a distance d_{ab} defines a circle of intersection (at the surface of both spheres) of radius

$$c = \frac{1}{2d_{ab}}\sqrt{4d_{ab}^2a^2 - (d_{ab}^2 - b^2 + a^2)^2}. \tag{4}$$

In this case $d_{ab} = d_p$, $a = r$, and $b = r_2$ since sphere A is the one centered at point $\vec{p}$ of radius r and sphere 2. Therefore

$$c = \frac{1}{d_p}\sqrt{4d^2r^2 - (d_p^2 - r_2^2 + r^2)^2}. \tag{5}$$

Then $\theta' = \arcsin(\frac{c}{r})$ defines the angle subtended by the intersection. Therefore

$$\begin{aligned} R(r, p, \theta)dr &= \int_0^{2\pi}\int_0^{arcsin(\frac{c}{r})} r^2 \sin(\theta')\, d\theta'\, d\phi' \\ &= 2\pi r^2 \int_0^{arcsin(\frac{c}{r})} \sin(\theta')\, d\theta' \\ &= 2\pi r^2 \left(1 - \sqrt{1 - \frac{c^2}{r^2}}\right) \\ &= \pi r^2 \left(2 - \sqrt{\frac{(r^2 + d_p^2 - r_2^2)^2}{r^2 d_p^2}}\right). \end{aligned}$$

Furthermore, from the law of cosines $d_p = \sqrt{d^2 + p^2 - 2dp\cos(\theta)}$, so

$$R(r, p, \theta) = \pi r^2 \left(2 - \sqrt{\frac{(d^2 + p^2 + r^2 - 2dp\cos(\theta) - r_2^2)^2}{r^2(d^2 + p^2 - 2dp\cos(\theta))}}\right). \tag{6}$$

This partial RDF is only physical for

$$\sqrt{d^2 + p^2 - 2dp\cos(\theta)} - r_2 \leq r \leq \sqrt{d^2 + p^2 - 2dp\cos(\theta)} + r_2 \tag{7}$$

(the zeros occur at the equalities), otherwise the spheres overlap. Since point $\vec{p}$ is within sphere 1, it is also the case that $d_p > r_2$, $d > r_2$, and $d > p\cos(\theta) + r_2$. We define the partial PDF as a piecewise function that is identically 0 outside the bounds just specified.

To integrate equation 6 over θ the proper boundaries of integration must be determined. This is dependent on d, p, and r due to the inequalities on the partial RDF just described.

Solving 7 for θ at the equalities yields

$$\theta_1 = \arccos\left(\frac{d^2+p^2-(r-r_2)^2}{2dp}\right) \tag{8}$$

$$\theta_2 = \arccos\left(\frac{d^2+p^2-(r+r_2)^2}{2dp}\right) \tag{9}$$

This is not quite complete, however, because the argument to arccos can fall outside the range $[-1,1]$ for physical values of d, p, r, and r_2. This corresponds to situations where these parameters are not necessary to constrain θ to the usual range, and leads the above expression to return a complex result. This can be removed by keeping only the real portion of 8 with the helpful identity

$$\Re(\arccos(x+\mathrm{i}y)) = \arccos\left(\frac{1}{2}\sqrt{(x+1)^2+y^2}-\frac{1}{2}\sqrt{(x-1)^2+y^2}\right) \tag{10}$$

for real x and y. The argument in 8 is always real, therefore

$$\theta_1 = \arccos\left(-\frac{1}{2}\sqrt{\left(-1+\frac{d^2+p^2-(r-r_2)^2}{2dp}\right)^2}+\frac{1}{2}\sqrt{\left(1+\frac{d^2+p^2-(r-r_2)^2}{2dp}\right)^2}\right)$$

$$\theta_2 = \arccos\left(-\frac{1}{2}\sqrt{\left(-1+\frac{d^2+p^2-(r+r_2)^2}{2dp}\right)^2}+\frac{1}{2}\sqrt{\left(1+\frac{d^2+p^2-(r+r_2)^2}{2dp}\right)^2}\right).$$

Evaluating the integral yields the unwieldy expression

$$R(r,p,d,r_2) = \frac{\pi r p}{6\sqrt{2}d}\Bigg(\sqrt{2(d^2+p^2)+|\alpha_1|-|\alpha_2|}\left(-\beta+3\sqrt{2}r\sqrt{2(d^2+p^2)+|\alpha_1|-|\alpha_2|}+|\alpha_2|-|\alpha_1|\right)+$$
$$\sqrt{2(d^2+p^2)+|\alpha_3|-|\alpha_4|}\left(\beta-3\sqrt{2}r\sqrt{2(d^2+p^2)+|\alpha_3|-|\alpha_4|}+|\alpha_3|-|\alpha_4|\right)\Bigg) \tag{11}$$

where

$$\begin{aligned}\alpha_1 &= d^2-2dp+p^2-(r+r_2)^2\\ \alpha_2 &= d^2+2dp+p^2-(r+r_2)^2\\ \alpha_3 &= d^2-2dp+p^2-(r-r_2)^2\\ \alpha_4 &= d^2+2dp+p^2-(r-r_2)^2\\ \beta &= 2d^2+2p^2+6r^2-6r_2^2.\end{aligned}$$

To integrate equation 11 over p it is helpful to first remove the absolute values by splitting the function into cases corresponding to whether the αs are positive or negative. We define a 4-symbol signature for each of the cases. For example, the function $R_{++--}(r,p,d,r_2)$

corresponds to positive α_1 and α_2, and negative α_3 and α_4. The values of r where each α is positive is given below.

$$\begin{aligned}
\alpha_1 &: 0 \le r \le d - p - r_2 \\
\alpha_2 &: 0 \le r \le d + p - r_2 \\
\alpha_3 &: 0 \le r \le d - p + r_2 \\
\alpha_4 &: 0 \le r \le d + p + r_2
\end{aligned}$$

Therefore, with respect to increasing r, α_1 is always the first to become negative, and α_4 the last. When $p < r_2$ then α_2 becomes negative second and α_3 third. When $p > r_2$, then α_3 becomes negative second and α_2 third. Given that, enumerate the 6 possible values of the integrand.

$$\begin{aligned}
R_{++++} &: 0 \\
R_{-+++} &: -\frac{\pi p r(-d+p+r-2r_2)(-d+p+r+r_2)^2}{3d} \\
R_{--++} &: -\frac{2\pi p^2 r(p^2+3(d-r)^2-3r_2^2)}{3d} \\
R_{-+-+} &: \frac{4\pi p r r_2^3}{3d} \\
R_{---+} &: -\frac{\pi p r(d+p-r-2r_2)(d+p-r+r_2)^2}{3d} \\
R_{----} &: 0
\end{aligned}$$

These integrands were determined by conditions on r, but to integrate over p the conditions on p where the limits of integration hold must also be calculated from the inequalities.

$$R(r,p,d,r_2) = \begin{cases}
R_{--++} & : r \le d \text{ and } 0 \le p \le -d + r_2 + r \\
R_{-+++} & : r \le d \text{ and } \sqrt{(d-r_2-r)^2} \le p \le d + r_2 + r \\
R_{-+-+} & : r \le d \text{ and } p \ge d + r_2 - r \\
R_{--++} & : r \ge d \text{ and } 0 \le p \le d + r_2 - r \\
R_{---+} & : r \ge d \text{ and } \sqrt{(d+r_2-r)^2} \le p \le -d + r_2 + r \\
R_{-+-+} & : r \ge d \text{ and } p \ge -d + r_2 + r \\
0 & : \text{elsewhere.}
\end{cases} \tag{12}$$

The limits given for R_{-+++}, R_{-+-+}, and R_{---+}, need no further alteration since they are physical for all values of the parameters. Unfortunately, the limits given for R_{--++} are not yet enough to perform the integration because for some values of d, r, and r_2 the expression $-d + r_2 + r$ may be a negative value which has no physical meaning.

Since R_{--++} is even in p it is possible to rewrite the integral in a way which avoids this

difficulty. For example, for the case where $r \leq d$

$$\begin{aligned}\int_0^{-d+r_2+r(\geq 0)} R_{--++}\,dp &= \frac{1}{2}\left(\int_0^{-d+r2+r} R_{--++}\,dp + \int_0^{\sqrt{(-d+r_2+r)^2}} R_{--++}\,dp\right)\\ &= \frac{1}{2}\int_{-\sqrt{(-d+r_2+r)^2}}^{-d+r_2+r} R_{--++}\,dp.\end{aligned}$$

Since the new lower bound is guaranteed to be negative but has the same magnitude as the upper bound, the integral is identically 0 for unphysical situations, and correctly adjusted by the additional factor of half in physical situations where the upper bound is positive. The case $r \leq d$ is similar.

Finally, since the integral of interest is $\int_0^{r_1} R(r,p)\,dp$ but $0 \leq r_1 \leq d - r_2$, any of the piecewise portions might be truncated before reaching its "natural" upper integration limit. Therefore, the integrals are calculated analytically with respect to an arbitrary upper bound u. For $r \leq d$ these are

$$\begin{aligned}\frac{1}{2}\int_{-|u|}^{u} R_{--++}(r,p)\,dp &= -\frac{\pi r}{15d}(u^5 + 5u^3((d-r)^2 - r_2^2) + 5((d-r)^2 - r_2^2)|u|^3 + |u|^5)\\ \int_{\sqrt{(d-r_2-r)^2}}^{u} R_{-+++}(r,p)\,dp &= \frac{\pi r}{60d}(-4u^5 + 15u^4(d-r) + 10u^2(-d+r+r_2)^2(d-r+2r_2)\\ &\quad - 5(-d+r+r_2)^4(5d-5r+4r^2) - 20u^3(d^2-2dr+r^2-r_2^2)\\ &\quad + 20(d^2-2dr+r^2-r_2^2)|-d+r+r_2|^3 + 4|-d+r+r_2|^5)\\ \int_{\sqrt{(d-r_2-r)^2}}^{u} R_{-+-+}(r,p)\,dp &= \frac{2\pi r}{3d} r_2^3(u^2 - (d-r+r_2)^2)\end{aligned}$$

and for $r \geq d$ they are

$$\begin{aligned}\frac{1}{2}\int_{-|u|}^{u} R_{--++}(r,p)\,dp &= -\frac{\pi r}{15d}(u^5 + 5u^3((d-r)^2 - r_2^2) + 5((d-r)^2 - r_2^2)|u|^3 + |u|^5)\\ \int_{\sqrt{(d+r_2-r)^2}}^{u} R_{---+}(r,p)\,dp &= \frac{\pi r}{60d}(-4u^5 - 15u^4(d-r) - 10u^2(d-r-2r_2)(d-r+r_2)^2\\ &\quad + 5(5d-5r-4r_2)(d-r+r^2)^4 - 20u^3(d^2-2dr+r^2-r_2^2)\\ &\quad + 20(d^2-2dr+r^2-r_2^2)|d-r+r_2|^3 + 4|d-r+r_2|^5)\\ \int_{\sqrt{(-d+r_2+r)^2}}^{u} R_{-+-+}(r,p)\,dp &= \frac{2\pi r}{3d} r_2^3(u^2 - (-d-r+r_2)^2).\end{aligned}$$

By examining the various intervals where these expressions are valid, one can write down the following surprisingly compact form for the (unnormalized) RDF.

$$R(r) = \begin{cases} \frac{\pi}{60d} r(-d+r+r_1+r_2)^3(d^2+r^2-3r(r_1+r_2)-4(r_1^2-3r_1r_2+r_2^2)+d(-2r+3(r_1+r_2)))\\ -\frac{2\pi}{15d} r r_1^3(5(d-r)^2+r_1^2-5r_2^2)\\ -\frac{2\pi}{15d} r r_2^3(5(d-r)^2+r_2^2-5r_1^2)\\ \frac{\pi}{60d} r(d-r+r_1+r_2)^3(d^2+r^2+3r(r_1+r_2)-4(r_1^2-3r_1r_2+r_2^2)-d(2r+3(r_1+r_2)))\\ 0 \end{cases} \tag{13}$$

with corresponding bounds

$$\begin{cases} d - r_1 - r_2 \leq r \leq d - |r_1 - r_2| \\ d - |r_1 - r_2| \leq d + |r_1 - r_2| \text{ and } r_1 \leq r_2 \\ d - |r_1 - r_2| \leq d + |r_1 - r_2| \text{ and } r_2 \leq r_1 \\ d + |r_1 - r_2| \leq r \leq d + r_1 + r_2 \\ \text{elsewhere.} \end{cases} \tag{14}$$

The correctness of this expression can be easily verified by numerical integration of equation 3 using equation 6.

The normalization constant is

$$\int_0^\infty R(r)\, dr = \int_{d-r_1-r_2}^{d+r_1+r_2} R(r)\, dr = \frac{8\pi}{9} r_1^3 r_2^3, \tag{15}$$

which is, as expected, dependent only on the size of the spheres.

1.3.2 RDF between two hollow spheres

The RDF between two hollow spheres with inner radii $r_{1,i}$ and $r_{2,i}$, outer radii $r_{1,o}$ and $r_{2,o}$, and separated by a distance $d \geq r_{1,o} + r_{2,o}$ can be calculated from the RDF between solid spheres. This is given by the RDF from the non-hollow region of sphere 1 to the entirety of sphere 2, less any contributions which are missing due to the hollow regions. Given $R(r) = R(r; r_1, r_2)$, for hollow spheres

$$R_{\text{hollow}}(r) = (R(r; r_{1,o}, r_{2,o}) - R(r; r_{1,i}, r_{2,o})) - (R(r; r_{1,o}, r_{2,i}) - R(r; r_{1,i}, r_{2,i})). \tag{16}$$

The normalization constant follows directly.

$$\int_0^\infty R_{\text{hollow}}(r)\, dr = \frac{8\pi}{9}(r_{1,o}^3 - r_{1,i}^3)(r_{2,o}^3 - r_{2,i}^3) \tag{17}$$

1.3.3 Fitting to experimental C_{60} PDF

The distance to the lattice point in the nth FCC shell is $a\frac{\sqrt{2n}}{2}$ for lattice parameter a. We treat the approximating hollow sphere to have a maximum radius $\frac{7.13+1.44}{2}$ Å, where the second term in the numerator approximates the "thickness" of the C_{60}'s walls as the carbon-carbon nearest neighbor distance. The first non-zero contribution from the nth cell occurs therefore occurs at approximately $(14.11\frac{\sqrt{2n}}{2} - (7.13 + 1.44))$ Å. For the 8th and 9th shells these are roughly 19.6 Å and 21.4 Å, so for a PDF out to 20 Å it suffices to fit through the 8th shell.

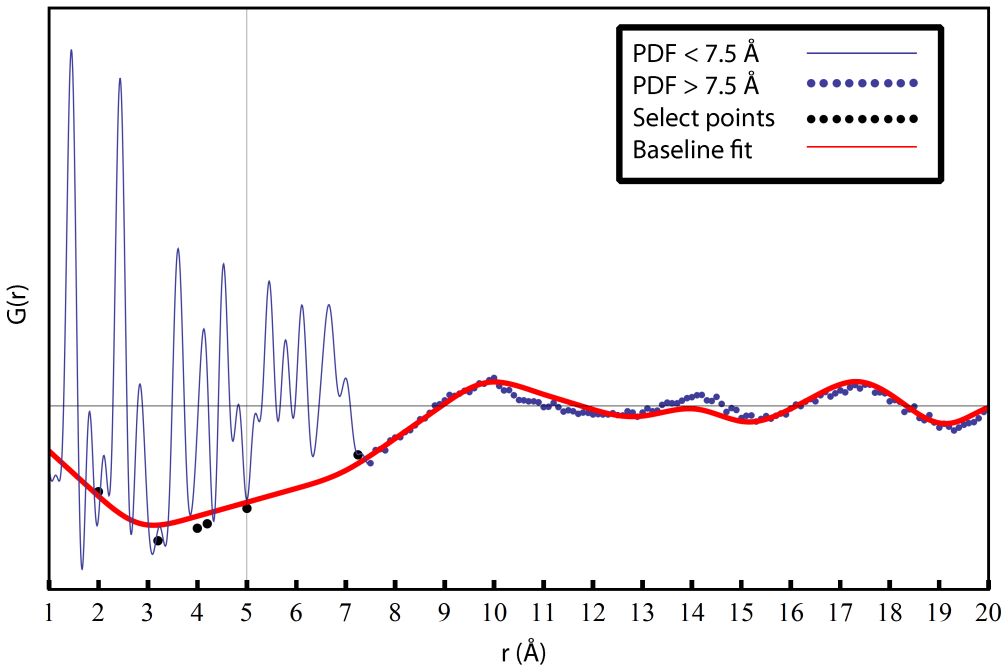

Figure 1: The fit baseline (in red) for C_{60} as determined by approximating its interparticle correlation with an FCC lattice of hollow spheres, up to a scale factor. Above 7.5 Å the single molecule PDF no longer contributes, so all those points are fit. Only the 6 points shown in black are included in the fit below 7.5 Å, but they are sufficient to constrain the fit because improvements above that range tend to induce very large costs below it.

The RDF of the interparticle correlation is then easily calculated, up to a scale factor, as the sum of the contribution from each shell, weighted by the number of contributing spheres. Starting from the nearest neighbor these weights are 12, 6, 24, 12, 24, 8, 48, and 6.

This model was fit using unweighted least squares to the experimental PDF from 7.5 Å to 20 Å (i.e. beyond the single molecule contribution) along with a few selected points below 7.5 Å (figure 1). These last points constrain the fit in the region where the interparticle correlation approaches 0, and they were chosen because points in the troughs of well separated peaks should be near the baseline. The approximate results of the fit are an inner radius of 3.07 Å, an outer radius of 4.06 Å, and lattice parameter 14.11 Å.

References

Guinier, A., Fournet, G., Walker, C. & Yudowitch, K. (1955). *Small-angle scattering of x-rays*. New York: John Wiley & Sons, Inc.

Heiney, P., Fischer, J., McGhie, A., Romanow, W., Denenstein, A., McCauley Jr., J., Smith, A. & Cox, D. (1991). *Phys. Rev. Lett.* **66**(22), 2911–2914.