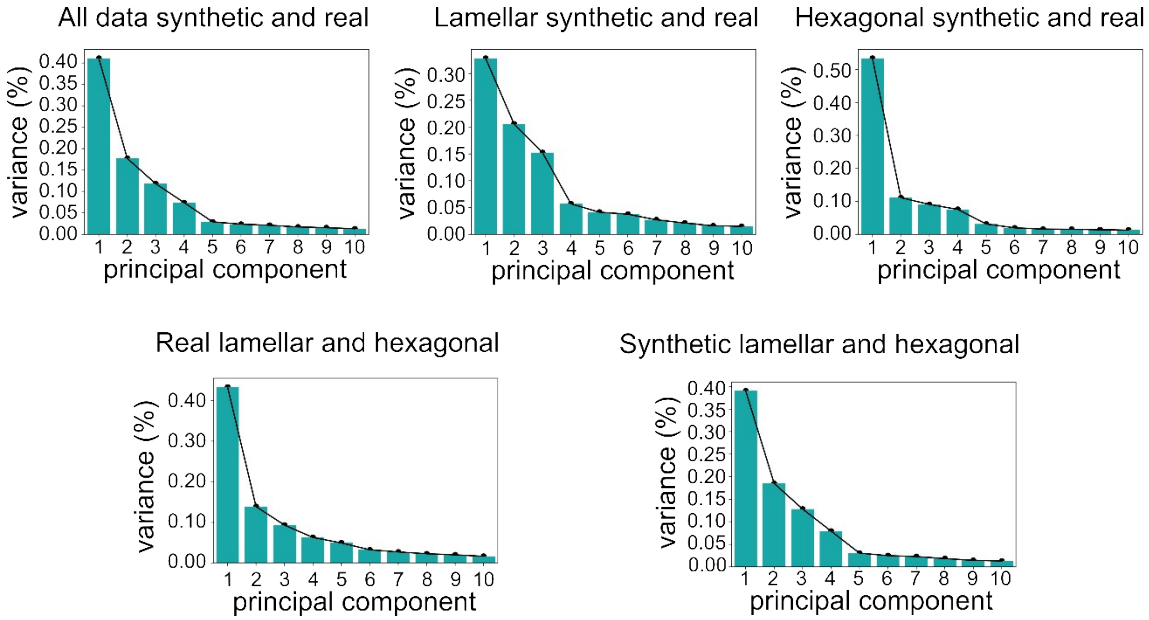
Electronic Supplementary Material (ESI) for Digital Discovery.

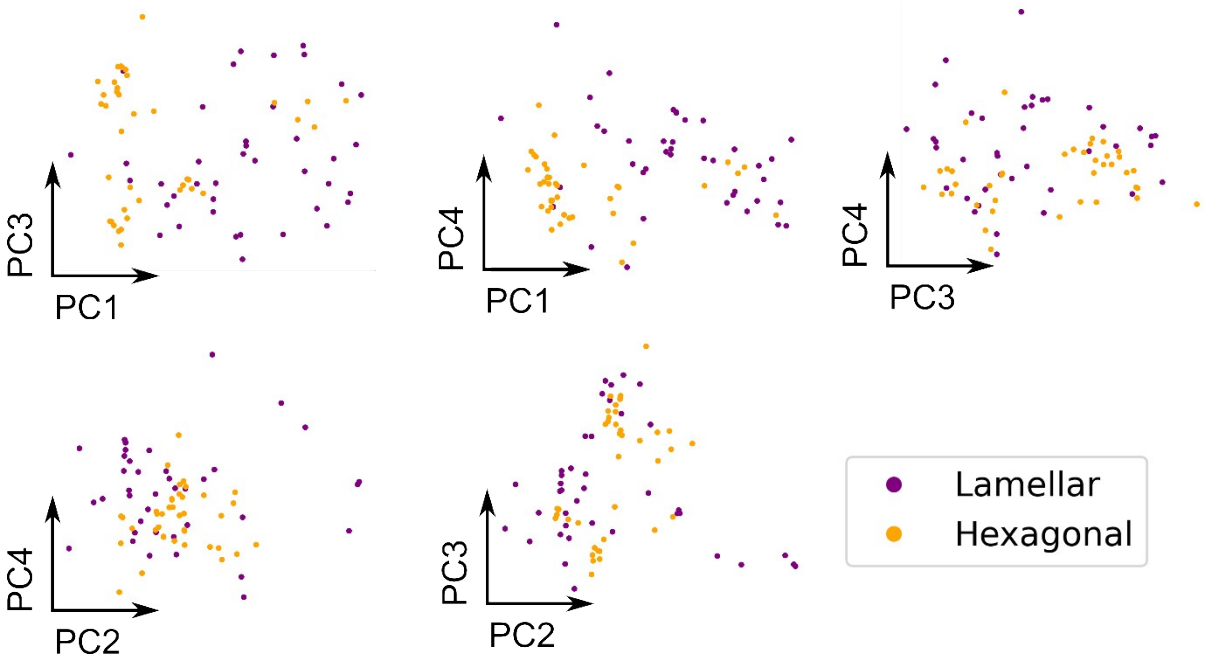
Supplementary

S.Table 1 – Variance captured for PCA components

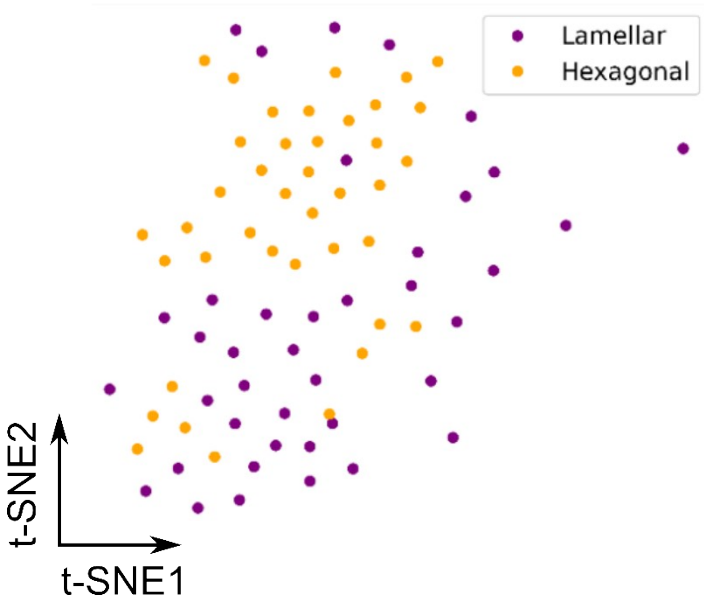
Figure	Description	Variance % (PCA 1)	Variance % (PCA 2)
Figure 2 A	All data synthetic and real	41.13	17.74
Figure 2 B	Lamellar synthetic and real	32.86	20.60
Figure 2 C	Hexagonal synthetic and real	53.44	11.11
Figure 3 A	Real lamellar and hexagonal	43.30	13.92
Figure 3 B	Synthetic lamellar and hexagonal	39.04	18.51



S.Figure 1: **Scree plots for each PCA distribution.** Elbow method reveals that the first 4 components capture significant variance of the dataset with the first 2 components deemed reasonable for a first assessment. The first 3 components in the lamellar synthetic and real plot appear significant.



S.Figure 2: **Further PCA distributions for real lamellar and hexagonal patterns.** No significant further clustering present in the first 4 components. PC1 vs PC4 shows minor separation by label although no significant clustering.



S.Figure 3:**Real data t-SNE distribution following a PCA dimensionality reduction.** A similar clustering pattern was observed as in figure 3 A. The PCA captured 89.94 % variance. Clustering was reduced upon changing perplexity from default (30).

Synthetic spectra modelling

Meshed electron density models were initially built in real and Fourier space to represent the lamellar phase (in 1D), the hexagonal phase (in 2D) and the cubic phase (in 3D).

Lamellar phase modelling

The lamellar phase Bravais lattice can be modelled in 1 dimension using the below function (1) where δ is the delta function, *a* is the lattice parameter and n is the number of points within the model.

$$s(x) = \sum_{n=-\infty}^{\infty} \delta(x - na) \quad (1)$$

The Fourier transform of equation 1 was then calculated to produce equation 2 which represents the reciprocal lamellar lattice.

$$S(q_x) = \frac{2\pi}{a} \sum_{m=-\infty}^{\infty} \delta\left(q_x - m\frac{2\pi}{a}\right) \quad (2)$$

Along with the lattice model the electron density function of the lamellar unit cell was modelled in 1 dimension (3). The electron density function is comprised of 3 different Gaussian functions that represent the bilayer headgroup positions at the positions $\pm\frac{L}{2}$, and at in the centre of the unit cell. Within this equation L represents the bilayer diameter, σ_1 represents the head group electron density signal and σ_2 represents the electron density signal of the terminal methyl group.

$$\rho(x) = \exp\left(-\frac{1}{\sigma_1}\left(x - \frac{L}{2}\right)^2\right) + \exp\left(-\frac{1}{\sigma_1}\left(x + \frac{L}{2}\right)^2\right) - \exp\left(-\frac{1}{\sigma_2}(x)^2\right) \quad (3)$$

Once more the Fourier transform of equation 3 was calculated resulting in equation 4, the reciprocal lamellar unit cell electron density function.

$$F(q_x) = \sqrt{\pi\sigma_1}\exp\left(-\pi^2\sigma_1(q_x)^2\right)\left(\exp\left(-\frac{iq_xL}{2}\right) + \exp\left(+\frac{iq_xL}{2}\right)\right) + \sqrt{\pi\sigma_2}\exp\left(-\pi^2\sigma_2(q_x)^2\right) \quad (4)$$

Inverse Hexagonal phase modelling

The inverse hexagonal phase is composed on micellar lipid cylinders that are packed hexagonally along the z-coordinate. This symmetry enables the inverse hexagonal phase to be modelled in 2 dimensions. However, we can collapse the 2 dimensions into a single scattering radius by taking the Euclidian distance of the reciprocal lattice points (equation 5). By taking a set of reciprocal lattice points along the radius from the origin a 1-dimensional set of scattering points in reciprocal space is obtained.

$$S(q_r) = \sqrt{(s(q_x) + s(q_y))^2} \quad (5)$$

Meanwhile the electron density function for the inverse hexagonal unit cell was modelled in polar coordinates by two gaussian functions modelling the lipid head and tail (equation 6). Within this equation r_h represents the position of the headgroups, r_l represents the position of the tails while σ_h and σ_l represent the thickness of the headgroups and tails respectively. Finally, I_1 and I_2 describe the relative intensity values of the two Gaussian functions, in this model $I_1 = 1$ and $I_2 = 0.2$

$$\rho(r,\theta) = I_1 \exp\left(-\frac{1}{2}\left(\frac{r - r_h}{\sigma_h}\right)^2\right) - I_2 \exp\left(-\frac{1}{2}\left(\frac{r - r_l}{\sigma_l}\right)^2\right) (6)$$

The reciprocal electron density function was once again found by taking the Fourier transform of the above equation. This results in equation 7 where the final equation is a function of the frequency radius k and angle ψ [2].

$$F(k) = \int_0^{\infty}\int_{-\pi}^{\pi} \rho(r,\theta) \exp\left[- irk\cos\left(\psi - \theta\right)\right] r dr d\theta \ (7)$$

As equation 7 is radially symmetric ($\rho(r,\theta) = \rho(r)\rho(\theta)$), the equation can be reduced to the form seen in equation 8.

$$F(k) = \int_0^{\infty} r\rho(r) \int_{-\pi}^{\pi} \exp\left[- irk\cos\left(\psi - \theta\right)\right] d\theta s \ (8)$$

Equation 8 can be simplified further through applying the zeroth-order Bessel function (equation 9) reducing the equation to the form seen within equation 10.

$$J_0(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left[- ix\cos\left(\psi - \theta\right)\right] d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left[- ix\cos\left(\alpha\right)\right] d\alpha \ (9)$$

$$F(k) = 2\pi \int_0^{\infty} \rho(r) J_0(kr) r dr \ (10)$$

This final equation is the Henkel transform and can be solved numerically to give an output value for the reciprocal electron density.

Inverse bi-continuous cubic phase modelling

As given previously by Garstecki et al[3,4], the reciprocal electron density in 3-dimensions is defined by the below equation:

$$F(q) = \int \rho(x) \exp\left[iq \cdot x\right] dx \ (11)$$

This electron density model was then modified by decorating the surface with two gaussians or a Gaussian convolved with 2 Kronecker deltas instead of using a rectangular bilayer function. The Kronecker delta functions were of the form detailed in equation 12 Where the distance between a and $-a$ is L. The Fourier transform of this function is then given by equation 13.

$$f_{kd}(x) = \delta(x - a) + \delta(x - a) \ (12)$$

$$F_{kd}(q \cdot n_j) = 2\cos\left(q \cdot n_j \frac{L}{2}\right) (13)$$

Meanwhile the Fourier transform of the new Gaussian function (equation 14) where the standard deviation is σ can be described by equation 15.

$$f_{gauss}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) (14)$$

$$F_{gauss}(q \cdot n_j) = \exp\left(- 2\pi^2\sigma^2 (q \cdot n_j)^2\right) (15)$$

By incorporating equations 13 and 15 into the original reciprocal electron density function and through neglecting fluctuations and modifying the shape of the bilayer, the final reciprocal electron density for the inverse bicontinuous cubic phases can be modelled as indicated below (equation 16). Where the he surface area ^{S_j} is the surface area of the surface triangle along the cubic surface, the positions of the midplane of the triangle is ^{x_j}, and the normal at that point is ^{n_j}. The wave vector q is defined here with the miller indices $q = \frac{2\pi}{a}[h,k,l]$.

$$F(q) = 2\sum_{j=1}^{N} s_j \exp\left[iq \cdot r_j\right] \cos\left(q \cdot n_j \frac{L}{2}\right) \exp\left(-2\pi^2\sigma^2(q \cdot n_j)^2\right) \quad (16)$$

This equation was them combined with the appropriate reciprocal inverse bicontinuous lattice model to produce the final intensity function.

Synthetic data generation

After producing equations to represent both the reciprocal electron density and reciprocal lattice for the phases modelled, model SAXS spectra were produced using equation 17. This equation gave the SAXS intensity $I(q)$ at a given wave number q and required the input of the reciprocal electron density function (the form factor) $F(q)$ of a single unit cell combined with the reciprocal lattice parameter $S(q)$ for any given phase.

$$I(q) = |F(q)|^2 S(q) \quad (17)$$

For the lamellar phase model SAXS spectra were produced by varying the lattice parameter a between 40 to 104 Å, the bilayer diameter L from 25 to 35 Å, the head group signal σ_1 from 2 to 4, and the terminal methyl signal σ_2 from 1 to 3. Each of these parameters were stepped by 10 increments creating 10,000 total samples of model lamellar SAXS spectra.

For the inverse hexagonal phase model SAXS spectra were produced by varying the lattice parameter a from 20 to 78 Å, head position from 5 to 30 Å,the headgroup thickness from 0.5 to 3 and the tail thickness from 0.5 to 5. Each of these parameters were stepped by 10, generating 10,000 samples of model hexagonal phase SAXS spectra.

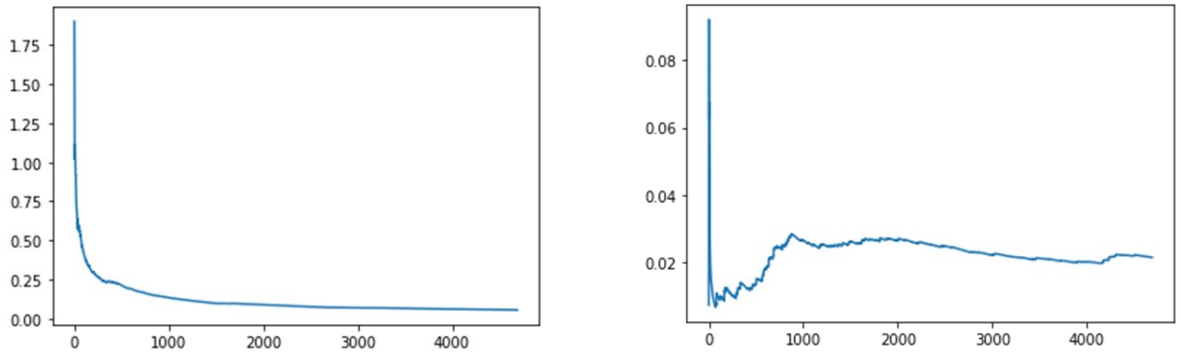
For the cubic phases SAXS spectra were produced by varying the lattice parameter a from 5 to 30 Å, length of lipid from 1 to 5 Å,and lipid head from 0.0001 to 0.01. Each of these parameters were stepped by 10.

To represent real SAXS spectra, a Voigt signal (equation 18) with the parameters varied randomly along a normal distribution with values: [0.0001,0.005) for the gaussian alpha and [0.0001,0.005) for the Lorenz gamma for both hexagonal and cubic phases. For the lamellar these parameter ranges were: [0.0001,0.001) for the gaussian alpha and [0.0001,0.005) for the Lorenz gamma respectively was added to the intensity peaks to model imperfect crystal scattering. Furthermore the Debye-Waller dampening factor (equation 19) with parameters randomly ranging from [0.005, 0.02) for hexagonal and cubic, and [0.02,0.1) for lamellar was applied to introduce dampening of peaks present at high q values[1]. Finally, a peak representing scattering from blank polycarbonate tubes was added from a randomly selected real blank sample. These additions resulted in the final synthetic SAXS spectra.

$$V(q) = \frac{Re[w(z)]}{\sigma\sqrt{2\pi}} \quad \text{where} \quad z = \frac{q + i\gamma}{\sigma\sqrt{2}} \quad (18)$$

$$I(q) = I_0 e^{-\beta q^2} \quad (19)$$

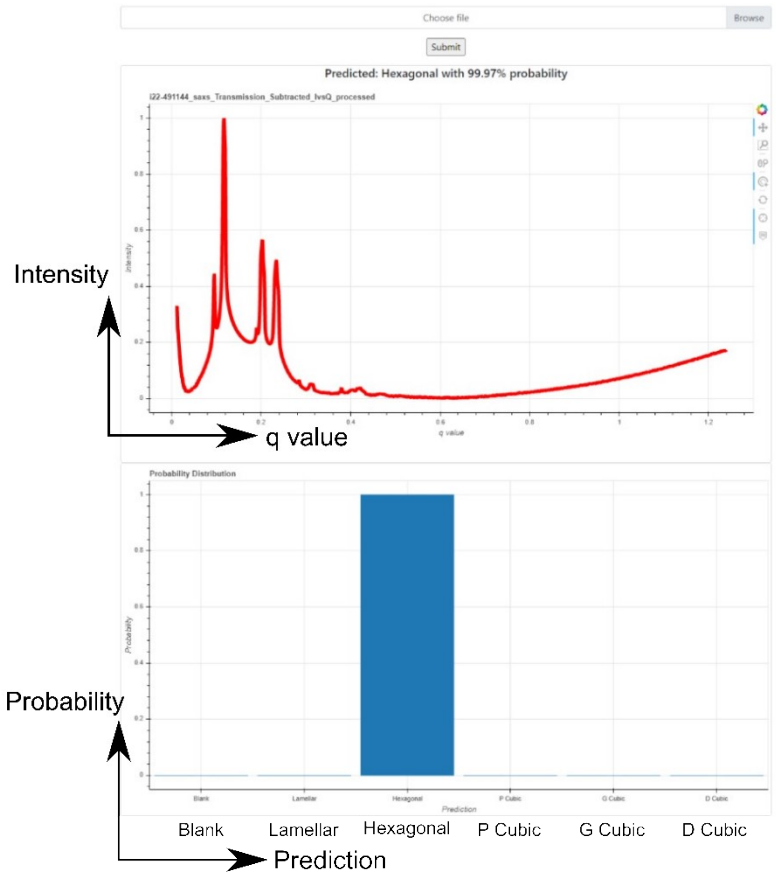
Pre-train Validation Loss



S.Figure 4: Validation loss vs. iteration number for Epochs 1 (left) and 2 (right) of pre-training on synthetic data

SAXS prediction examples

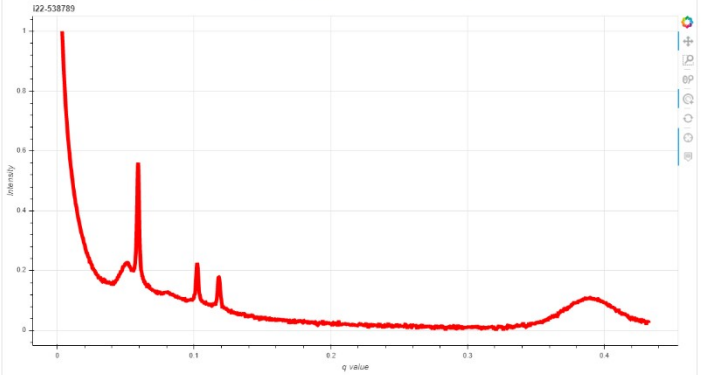
The following figures are samples of our model predicting the lipid phase of real, experimentally obtained SAXS patterns – these are all screenshots of our in-house webserver in action. To aid visualisation, the first example has enlarged labels and tags.

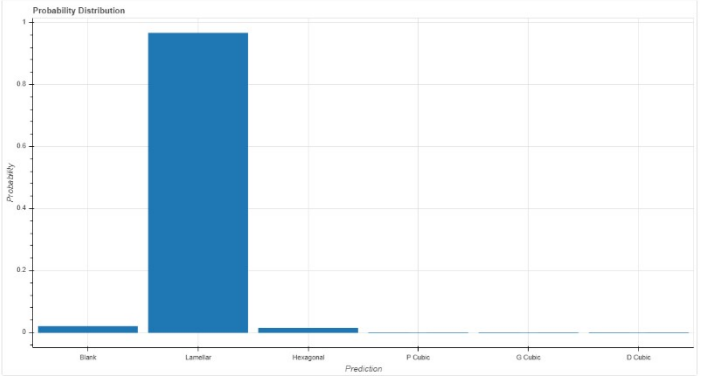


Choose file Browse

Submit

Predicted: Lamellar with 96.60% probability

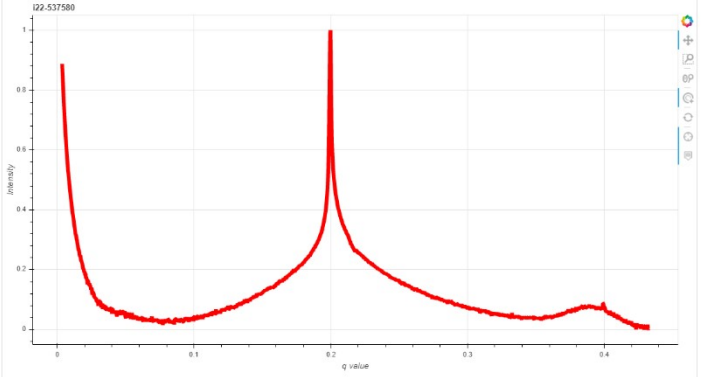


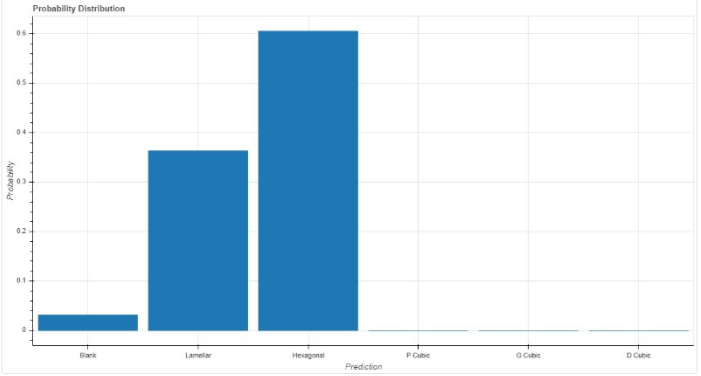


Choose file Browse

Submit

Predicted: Hexagonal with 60.53% probability

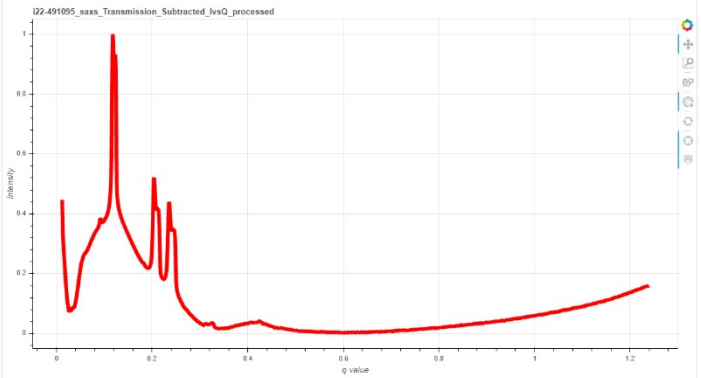


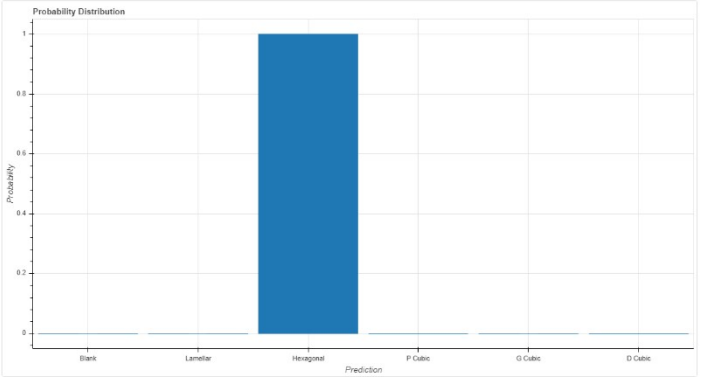


Choose file Browse

Submit

Predicted: Hexagonal with 100.00% probability





Choose file Browse

Submit

Predicted: Lamellar with 76.73% probability