Supplementary Information (SI) for Environmental Science: Nano.
This journal is © The Royal Society of Chemistry 2024

Supporting Information for *Environmental Science: Nano* article:

Formation and Transformation of Iron oxy-hydroxide Precursor Clusters to Ferrihydrite

Alireza Namayandeh[a,d], Olaf J. Borkiewicz[b], Michel Sassi[c], , Kevin Rosso[c], and F. Marc Michel[a,e]

[a]Department of Geosciences, Virginia Tech, Blacksburg, VA 24061, USA

[b]Advanced Photon Source, Argonne National Laboratory, IL60439, USA

[c]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[d]Department of Earth System Science, Stanford University, Stanford, CA 94305, USA

[e]Division of Nanoscience, Academy of Integrated Science, Virginia Tech, Blacksburg, VA 24061, USA

Authors e-mail addresses: arnama@stanford.edu, borkiewicz@anl.gov, michel.sassi@pnnl.gov, kevin.rosso@pnnl.gov, marc.michel@vt.edu

Corresponding author:

Alireza Namayandeh- Current address: Stanford University; 367 Panama St, Stanford, CA 94305 arnama@stanford.edu

10 pages, including 1 text, 1 table, 6 figures, and supporting references.

List of Contents.

Text S1. Linear Combination Fitting: Linear combination fitting (LCF) is used to estimate the relative proportions of distinct phases in a sample containing a mixture of various phases. Mathematically, LCF can be represented as:

Measured Data = (Comp 1 * Coeff 1) +(Comp 2 * Coeff 2) + ... + (Comp n * Coeff n) (1)

In Equation 1, the data for each component (Comp) is multiplied by a coefficient (Coeff). The best fit is achieved by varying the coefficients for each component freely or within defined limits.

In this study, LCF is used to test if the experimental PDFs are made up of one or more of the structures proposed by Das[1], including the Fe_{13} delta-Keggin structure. Fitting involves representing the experimental PDF as a linear combination of up to 18 component PDFs corresponding to different structures obtained from Das[1]. The experimental PDFs for the samples synthesized at pH 1.0, 1.5, 2.5, and 4.5 were evaluated using LCF analysis. While LCF typically uses ordinary least squares regression to minimize the sum of squared residuals, incorporating many component spectra can introduce challenges such as overfitting and collinearity. To mitigate these effects, we used two different approaches that are detailed in the following.

L1-Regularization LCF: In the first approach, L1 regularization, also known as Lasso (Least Absolute Shrinkage and Selection Operator) regression, was incorporated into LCF fitting to assist in automatically selecting a subset of the 18 components that best describe the experimental data. L1 regularization adds a penalty to the objective function proportional to the absolute value of the coefficients. This has the property of driving some coefficients to zero, effectively selecting a subset of candidates and thereby minimizing the impacts of the potential issues noted above. The objective and penalty can be represented as:

$$\text{Objective} = ||Y - X\beta||\frac{2}{2} + \lambda||\beta||_1 \quad (2)$$

In Equation 2, Y is the experimental PDF, X is the matrix of candidate spectra, β is a vector of coefficients for each candidate spectrum, and λ is the regularization parameter controlling the strength of the penalty. A λ of 0.005 was chosen for the fit results reported here. **Combinatorial LCF:** In a second approach, we combined LCF with combinatorial analysis and ordinary least squares regression to evaluate all possible combinations of the components. Given 18 components, the total number of possible combinations amounts to 262,143. Using Python (v3.11), the code calculated and compared evaluation metrics, including Mean Squared Error, Root Mean Squared Error, and R-squared for each fit. After evaluating all possible combinations, the code identified the combination and corresponding coefficients that provided the best fit based on the highest R-squared value.

Table S1. Selected details for the DFT-optimized structures reported by Das[1].

Number	Structure ID	Description	Composition
1	Fe1	Monomer	$Fe(H_2O)_6^{3+}$
2	Fe2a	Dimer (ES, O)	$Fe_2(H_2O)_6(OH)_4^{2+}$
3	Fe2b	Dimer (ES, O)	$Fe_2(H_2O)_6(OH)_4^{2+}$
4	Fe2c	Dimer (ES, O)	$Fe_2(H_2O)_8(OH)_2^{4+}$
5	Fe2d	Dimer (ES, O)	$Fe_2(H_2O)_8(OH)_2^{4+}$
6	Fe2e	Dimer (ES, O)	$Fe_2(H_2O)_6(OH)_4^{2+}$
7	Fe2f	Dimer (CS, O)	$Fe_2O(H_2O)_{10}^{4+}$
8	Fe2g	Dimer (CS, O)	$Fe_2OH(H_2O)_{10}^{5+}$
9	Fe3a	Trimer (ES, O)	$Fe_3O(OH)_3(H_2O)_9^{4+}$
10	Fe3b	Trimer (ES, O)	$Fe_3O(OH)_3(H_2O)_9^{4+}$
11	Fe3c	Trimer (ES, O)	$Fe_3(OH)_4(H_2O)_{10}^{5+}$
12	Fe3d	Trimer (ES, O)	$Fe_3(OH)_4(H_2O)_{10}^{5+}$
13	Fe4a	Tetramer (ES, O)	$Fe_4O_2(OH)_4(H_2O)_{10}^{4+}$
14	Fe4b	Tetramer (ES, O)	$Fe_4(OH)_8(H_2O)_8^{4+}$
15	Fe5	Pentamer (ES, O)	$Fe_5O_3(OH)_5(H_2O)_{11}^{4+}$
16	Fe5	Pentamer (ES, O)	$Fe_5O_3(OH)_5(H_2O)_{11}^{4+}$
17	Fe7	Heptamer (ES, CS, OT)	$Fe_7O(OH)_{12}(H_2O)_{12}^{7+}$
18	Fe13a	Keggin ion	$[FeO_4(Fe(OH)_2(H_2O))_{12}]^{7+}$

19	Fe13b	Keggin ion	$[FeO_4(Fe(OH)_2(H_2O))_{12}]^{7+}$

ES = Edge-sharing, CS = Corner-sharing, O = Octahedral, OT = Mix of Octahedral & Tetrahedral

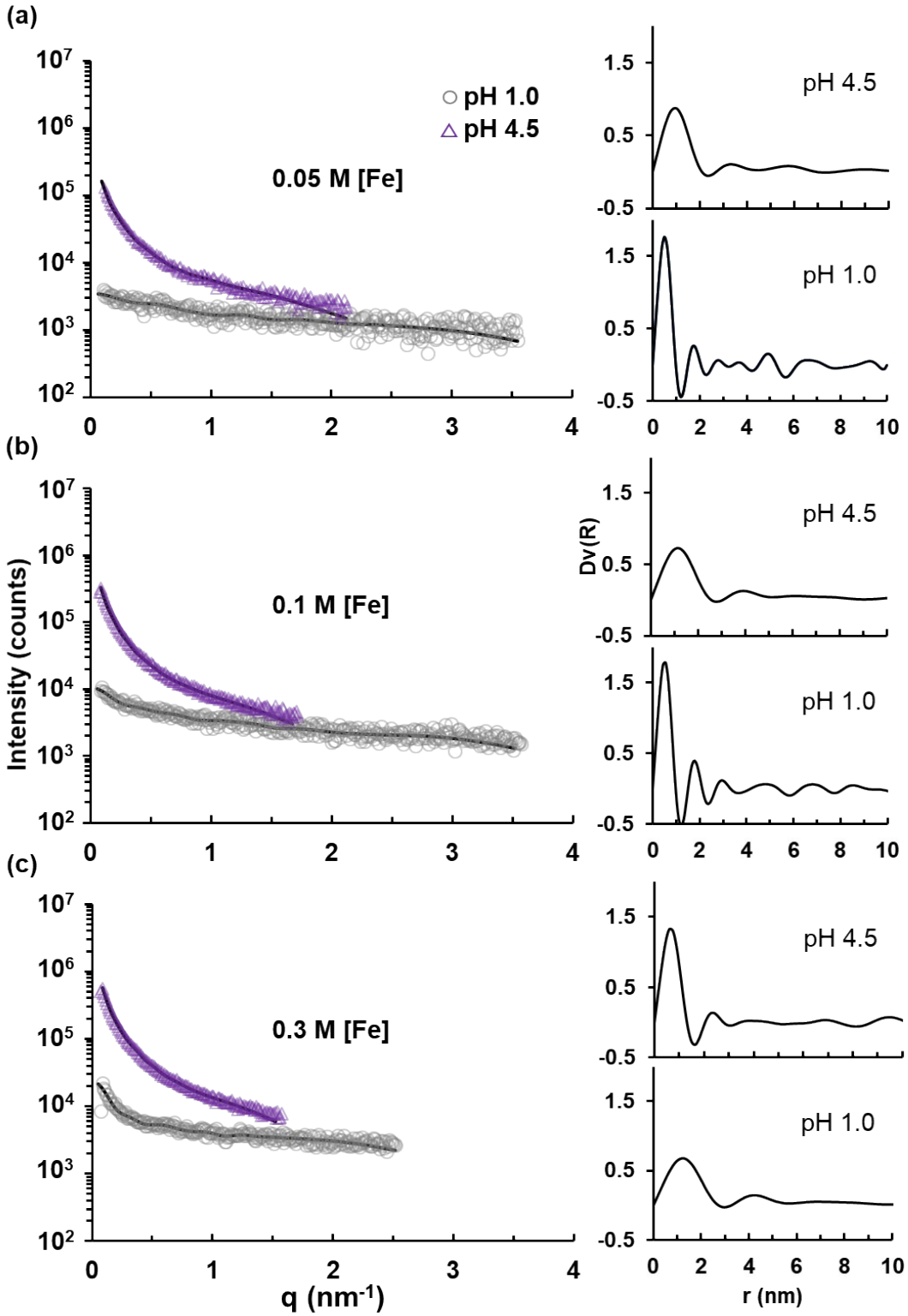

Figure S1. Selected SAXS profiles of background subtracted samples at pH 1.0 (open circle) and 4.5 (triangles) and their Dv(R) fits (solid lines) for (a) 0.05 M (b) 0.1 M, and (c) 0.3 M Fe concentrations.

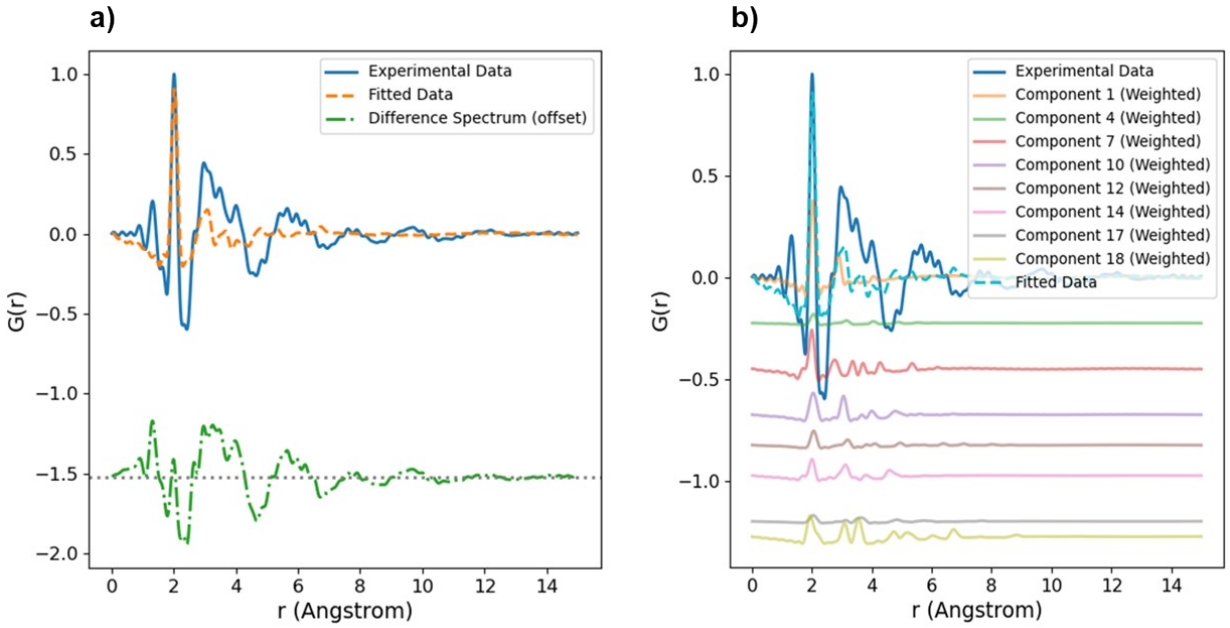

Figure S2. a) LCF fit and different spectrum and b) Weighted component spectra and calculated some for pH 1.0 sample

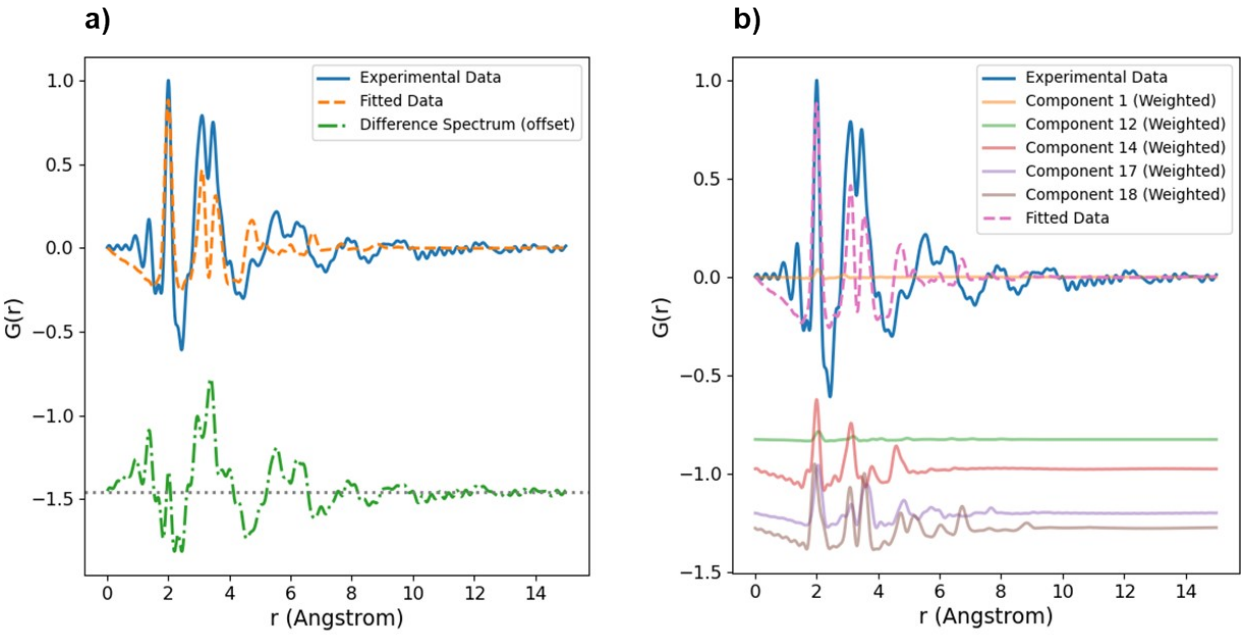

Figure S3. a) LCF fit and different spectrum and b) Weighted component spectra and calculated some for pH 1.5 sample

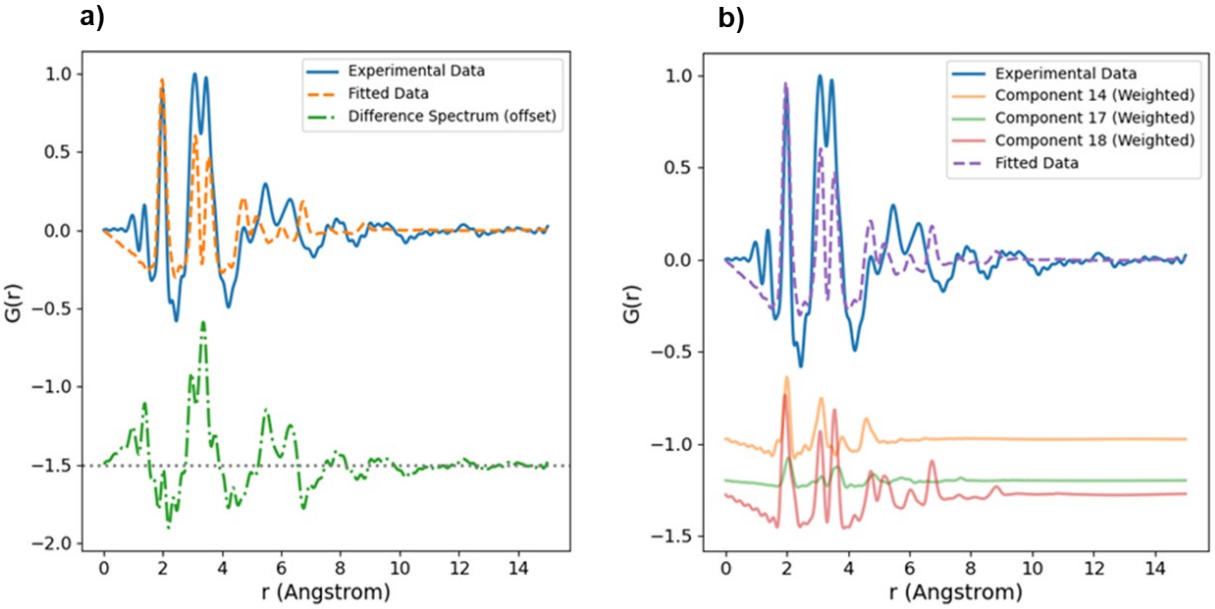

Figure S4. a) LCF fit and different spectrum and b) Weighted component spectra and calculated some for pH 2.5 sample

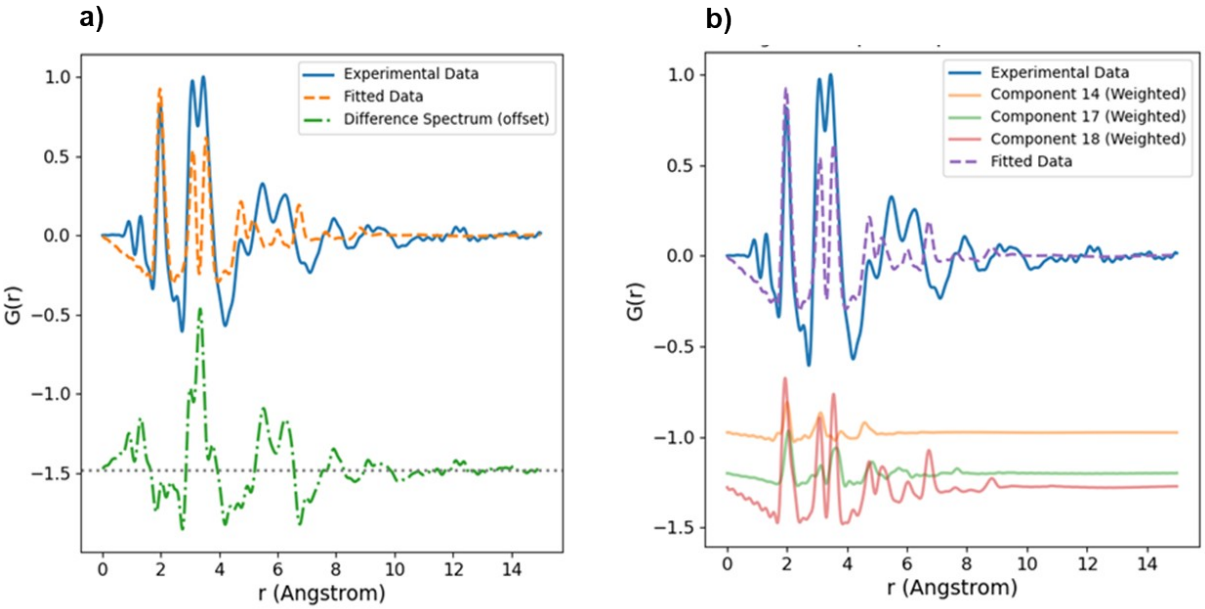

Figure S5. a) LCF fit and different spectrum and b) Weighted component spectra and calculated some for pH 4.5 sample

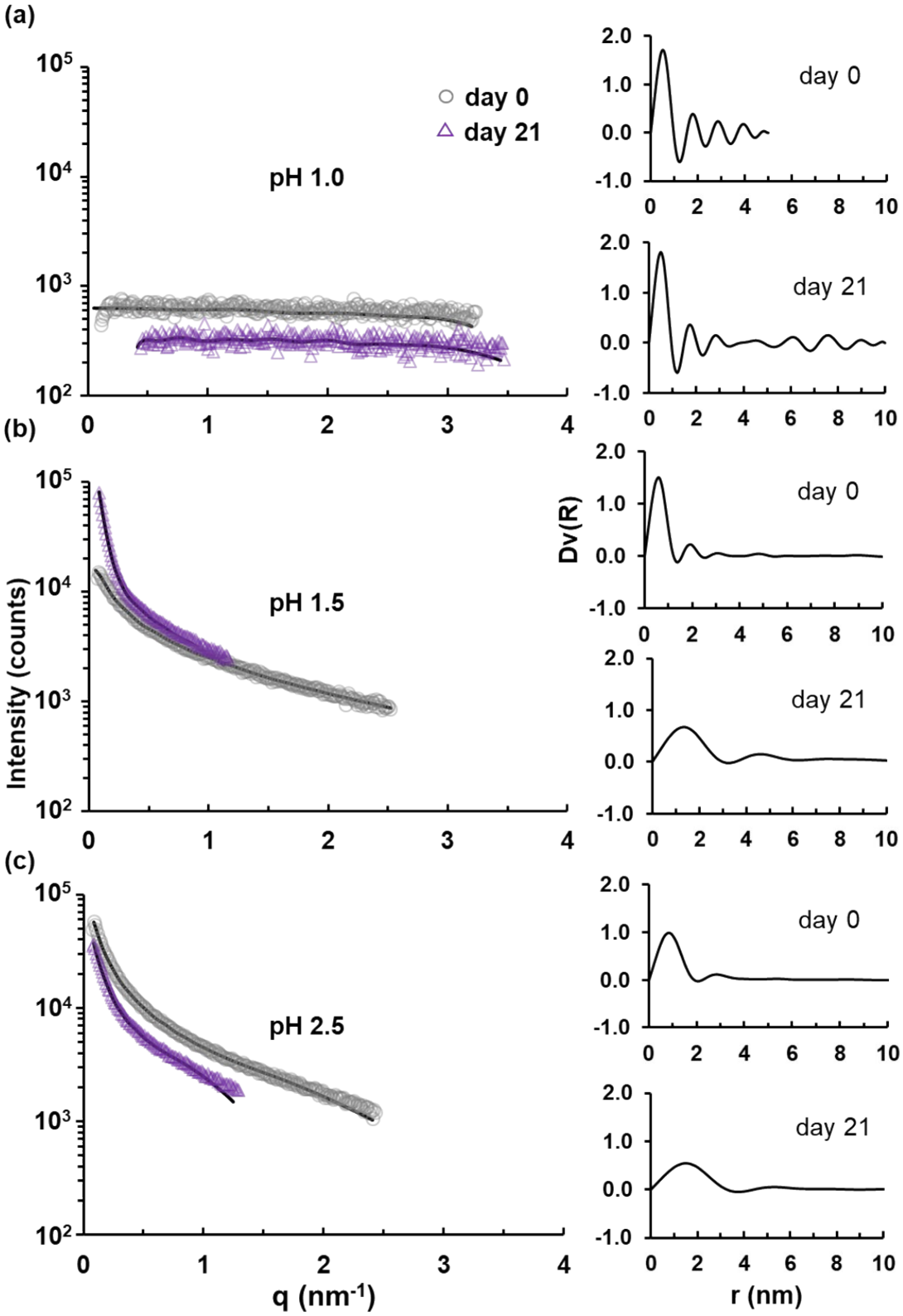
(a)
day 0
day 21
pH 1.0
day 0
day 21
(b)
pH 1.5
Intensity (counts)
Dv(R)
day 0
day 21
(c)
pH 2.5
day 0
day 21
q (nm-1)
r (nm)

Figure S6. Selected SAXS profiles of background subtracted samples on days 0 (open circle) and 30 (triangles) and their Dv(R) fits (solid lines) for pH (a) 1.0, (b) 1.5, and (c) 2.5

References

1. Das, B., Theoretical Study of Small Iron–Oxyhydroxide Clusters and Formation of Ferrihydrite. *The Journal of Physical Chemistry A* **2018,** *122* (2), 652-661.