Supplementary Information for:

Intracellular Polyphosphate Length Characterization in Polyphosphate Accumulating Microorganisms (PAOs): Implications in PAO Phenotypic Diversity and Enhanced Biological Phosphorus Removal Performance

Dongqi Wanga,b,c,1, Yueyun Lic,d,1, Helen A. Copee, Xiaoxiao Lib, Peisheng Hef, Cong Liub, Guangyu Lic,f, Sheikh M. Rahmanc,g, Nicholas B. Tookerc,h, Charles B. Botti, Annalisa Onnis-Haydenc, Jyoti Singhj,k, Alistair Elficke, Ricardo Marquesl, Henning J. Jessenj, Adrian Oehmenl,m and April Z. Guc,f,\*[[1]](#footnote-1)

a State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi'an University of Technology, Xi'an, Shaanxi, 710048, China

b Department of Municipal and Environmental Engineering, Xi'an University of Technology, Xi'an, Shaanxi, 710048, China

c Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Boston, MA, 02115, United States

d Black and Veatch, 2999 Oak Road #490, Walnut Creek, CA, 94597, United States

e School of Engineering, Institute for Bioengineering, The University of Edinburgh, Edinburgh, United Kingdom

f School of Civil and Environmental Engineering, Cornell University, 220 Hollister Hall, Ithaca, NY, 14853, United States

g Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

h Department of Civil and Environmental Engineering, University of Massachusetts Amherst, Marston Hall, Amherst, MA, 01003, United States

i Hampton Roads Sanitation District, 1434 Air Rail Avenue, Virginia Beach, VA, 23454, United States

j Institute of Organic Chemistry, University of Freiburg, Albertstrasse 21, 79104 Freiburg, Germany

k Department of Chemistry, University College London, 20 Gordon St, Bloomsbury, London WC1H 0AJ, United Kingdom

l UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Campus de Caparica, 2829-516 Caparica, Portugal

m School of Chemical Engineering, The University of Queensland, Brisbane, Queensland 4072, Australia

**Contents**

**1. Methods**

*1.1 31P-NMR spectra-based average polyP chain length determination*

*1.2* *Multivariate analysis of preprocessed Raman spectra*

**2. Tables**

**Table S1** Phosphorus-containing chemicals used in this study.

**Table S2** Composition of the synthetic influent wastewater of the lab-scale EBPR reactor operated at SRT of 3, 5, 10 and 20d.

**Table S3** Summary of the main characteristics of the full-scale EBPR plants for 31P-NMR and PAGE analyses. All results presented are averaged values based on the information provided by the plants for three months of operation at the sampling time.

**Table S4** Summary of average operational data, performance and phylogenetic information of the four S2EBPR facilities for SCRS analysis. Data is retrieved from Onnis-Hayden et al. (2020b).

**3. Figures**

Figure S1 Flow diagrams of the biological treatment process at the investigated S2EBPR facilities: (a) Side-stream RAS fermentation (SSR) configuration at the South Cary Water Reclamation Facility (SC), Apex, North Carolina; (b) Side-stream RAS fermentation with supplemental carbon addition (SSRC) configuration at the Westside Regional Wastewater Treatment Plant (WR), West Kelowna, British Columbia; (c) Side-stream mixed liquor suspended solids (MLSS) fermentation (SSM) configuration at the Cedar Creek Wastewater Treatment Facility (CC), Olathe, Kansas; and (d) Unmixed in-line MLSS fermentation (UMIF) configuration at the Kurt R. Segler Water Reclamation Facility (Hen), Henderson, Nevada. The figure is retrieved from Gu et al. (2019).

Figure S2 Relative frequency histogram and distribution curve of PO2- peak position of polyPs in PAOs from a *Tetrasphaera*-enriched culture fed with sodium casein hydrolysate. (a), (b) and (c) show the results from three biological replicates (50 PAO cells per replicate) in the *Tetrasphaera*-enriched culture. Raman spectra are available upon request. FWHM: full width at half maximum.

Figure S3 Example 31P-NMR spectra of a) EDTA+NaOH extracted and b) EDTA extracted polyP sample. PP1: terminal phosphate groups; PP2 and PP3: the second and third post-terminal phosphate groups; PP4: the core phosphate groups. The 31P-NMR spectra of EDTA+sonication extracted polyPs were similar with that extracted by EDTA only.

Figure S4 Comparison of different extraction methods for polyP analysis with 31P-NMR, on four EBPR-SBRs operated with 3, 5, 10 and 20 days SRT. Primary Y-axis left: Visible polyP content represented by the 31P-NMR integral of the “middle” P signal (PP2+PP3+PP4) of polyPs, normalized by sludge VSS. Secondary Y-axis: PO4-P removal efficiency for each of the SBRs at the extraction sampling time. The asterisk labelled SBRs of 10 and 20 days SRT were experiencing deterioration due to low dissolved oxygen supply for a short period.

Figure S5 Relative frequency histogram and distribution curve of P–O–P peak position of polyPs in SCRS-identified PAOs from (a) a *Tetrasphaera*-enriched culture fed with sodium casein hydrolysate (n=355) and (b) an *Accumulibacter*-like culture fed with acetate (n=60). Raman spectra are available upon request. FWHM: full width at half maximum.

Figure S6 (a) Preprocessed single-cell Raman spectra (n=168) collected from the four lab-scale EBPR systems, clustered by HCA into each sub-OPUs, with the signature peaks for glycogen and PolyP labeled. (b) Dendrogram from HCA of single-cell Raman spectra, with major and sub-OPUs labeled. The spectra and dendrogram are color-coded based on Raman analysis as follows: pink−candidate PAOs; blue−candidate GAOs; black−other cells). The figure is retrieved from Li et al. (2018). Raman spectra are available upon request.

**Supplementary References**

**1. Methods**

*1.1 31P-NMR spectra-based average polyP chain length determination*

PolyP chain length, *n*, was estimated from the 31P-NMR spectra, based on the integrals of the following polyP-associated peaks (Figure S3): A doublet signal at -10 ppm originated from the terminal phosphate group (PP1); The two signals at -25 and -26 ppm corresponding to the second and third post-terminal phosphate groups (PP2 and PP3); A single intense signal at -26.5 ppm corresponding to the core phosphate groups (PP4) beyond these three terminal groups (Pilatus et al. 1989). The average length of heterogeneous polyPs, *n*, can then be estimated by the following, where, the total number of P atoms was estimated as the sum of PP1, 2, 3, 4. Since each molecule of polyP has two terminal entities, 1/2 PP1 represents the number of polyP polymers in the sample.

$Average n=\frac{Total No. of P}{No. of PolyP chain}=\frac{PP1+PP2+PP3+PP4}{\frac{1}{2} PP1}$ (1)

The PP3 peak is usually integrated together with the PP4 peak, and termed “internal P” (Smith et al. 2010). This polyP length determination method was validated with two synthetic polyPs, namely sodium phosphate glass type 45 (SPG45) and sodium hexametaphosphate (SHMP).

*1.2 Multivariate analysis of preprocessed Raman spectra*

The full wavenumber region, 400−1800 cm-1, of each preprocessed Raman spectrum was used for the multivariate analysis. The multivariate analysis procedure consisted of two steps: spectral transformation and classification.

(1) To ensure comparability between the spectra, and to correct for potential minor variation in spectral resolution and quality, each preprocessed spectrum was transformed by binning the Raman intensity over regions of size 5 cm-1, as suggested by Webb-Robertson et al. (2012).

(2) An unsupervised multivariate statistical analysis, hierarchical clustering analysis (HCA), using correlation coefficient as the distance measuring criteria, was performed on all of the single-cell Raman spectra, based on intra-spectral similarities (Meisel et al. 2014). A dendrogram was generated to present the clustering output that depicts different groups of Raman spectra with similar features. And we defined the major Raman-classified groups as the operational phenotypic units (OPUs), which could be further broken down into sub-OPUs (Figure S6). The OPU is defined based on cluster analysis of Raman spectra, using the same rationale as that of the widely accepted operational taxonomic units (OTUs), where OTU is defined based on similarities between nucleotide sequences and OPU is defined based on similarities of single-cell Raman spectra. The cutoff distance for sub-OPUs was set at 0.7.

**2. Tables**

Table S1 Phosphorus-containing chemicals used in this study.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chemical name | Formula | CAS | Abbreviation | Structure |
| Monosodium phosphate | NaH2PO4·2H2O | 13472-35-0 | MSP, P1 |  |
| Tetrasodium pyrophosphate | Na4P2O7·10H2O | 13472-36-1 | TSPP, P2 |  |
| Sodium tripolyphosphate | Na5P3O10 | 7758-29-4 | STPP, P3 |  |
| Sodium tetraphosphate | Na6P4O13 | 14986-84-6 | STP, P4 |  |
| Sodium hexametaphosphate | (NaPO3)6 | 10124-56-8 | SHMP, P6 | âSodium Hexametaphosphateâçå¾çæç´¢ç»æ |
| Ammonium polyPs labeled with alkyne tags | C5H35N7O19P6 | n/a | pentyne-P6 |  |
| C5H39N8O22P7 | n/a | pentyne-P7 |  |
| Sodium phosphate glass type 45 | Nan+2PnO3n +1(n=~45) | n/a | SPG45, P45 |  |
| Polyphosphate, medium chain (p100) | Nan+2PnO3n +1 (n=~100) | n/a | P100 |
| Polyphosphate, long chain (p700) | Nan+2PnO3n +1 (n=~700) | n/a | P700 |

Table S2 Composition of the synthetic influent wastewater of the lab-scale EBPR reactor operated at SRT of 3, 5, 10 and 20d.

|  |  |  |
| --- | --- | --- |
| Parameter | Component | Concentration (mg/L) |
| Carbon | CH3COONa·3H2O | 312 |
| Casamino acids | 30 |
| Yeast extract | 8 |
| Nutrient | NaH2PO4·2H2O | 40.3 |
| NH4Cl | 30.6 |
| Macroelement | KCl | 117 |
| MgCl2·6H2O | 219 |
| MgSO4·7H2O | 14 |
| CaCl2 | 46 |
| Trace element | H3BO3 | 0.061 |
| ZnSO4·7H2O | 0.300 |
| KI | 0.015 |
| CuSO4·5H2O | 0.061 |
| Co(NO3)2·6H2O | 0.075 |
| Na2MoO4·2H2O | 0.031 |
| MnSO4·H2O | 0.340 |
| FeSO4·7H2O | 0.300 |
| Nitrification inhibitor | Allylthiourea (ATU) | 4 |

**Table S3** Summary of the main characteristics of the full-scale EBPR plants for 31P-NMR and PAGE analyses. All results presented are averaged values based on the information provided by the plants for three months of operation around the sampling time.

|  |  |
| --- | --- |
| Main characteristics and parameters | Wastewater treatment plants |
| Clark County | Las Vegas | Atlantic Plant | Nansemond Plant | Virginia Initiative Treatment Plant |
| (CL) | (LV) | (AT) | (Nan) | (VIP) |
| Operational conditions | Configuration | A/O | A2O+MBR | High rate A/O | 5-stage BNR | VIP |
| Average flow (MGD) | 91±3 | 43±1 | 27±4 | 17±1 | 35±5 |
| Temperature (°C) | 20±1 | 20±2 | 17±2 | 16±1 | 16±1 |
| SRT (d) | 7 | n/a | 1.5-3 | n/a | n/a |
| Influenta | cBOD (mg/L) | 135±13 | n/a | n/a | 148±19 | 116±16 |
| TKN (mg N/L) | n/a | n/a | 44±2 | 39±5 | 24±3 |
| NH4–N (mg N/L) | 28±4 | 26±2 | 34±2 | 31±3 | 18±2 |
| TP (mg P/L) | 4.2±0.8 | 3.9±3.1 | 6±2 | 8±1 | 3±0.7 |
| cBOD: N:P | 32:7:1 | n/a | n/a | 19:5:1 | 39:8:1 |
| Efficiencyb | cBOD removal (%) | n/a | 95c | 93c | 97 | 95 |
| TKN removal (%) | n/a | n/a | n/a | 89 | 69 |
| TP removal (%) | 99d | 95d | 44d | 79 | 91 |
| Effluent TP (mg P/L) | n/a | n/a | n/a | 1.4±0.9 | 0.4±0.2 |
| Effluent ortho-P (mg P/L) | 0.03 | 0.09±0.03 | 4±3 | 1.2±0.9 | 0.12±0.10 |
| PAO/GAO populationse | *Accumulibacter* | 11.1% | n/a | n/a | n/a | n/a |
| *Competibacter* | 8.9% | n/a | n/a | n/a | n/a |

SRT: sludge retention time; cBOD: [carbonaceous biochemical oxygen demand](http://en.wikipedia.org/wiki/Carbonaceous_biochemical_oxygen_demand); TKN: total Kjeldahl nitrogen; NH4–N: ammonia; TP: total phosphorus; n/a: not available.

a: influent to biological process;

b: removal efficiency in biological process;

c: BOD removal;

d: ortho-P removal

e: determined by FISH (Onnis-Hayden et al. 2020a).

**Table S4** Summary of average operational data, performance and phylogenetic information of the four S2EBPR facilities for SCRS analysis. Data is retrieved from Onnis-Hayden et al. (2020b).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Facility | Cedar Creek | Henderson | South Cary | Westside Regional |
| (CC) | (Hen) | (SC) | (WR) |
| *System configuration information* |
| S2EBPR configuration | SSM | UMIF | SSR | SSRC |
| Mainstream configuration | Johannesburg | Johannesburg | 4‐stage Bardenpho | Modified MLE |
| Chemical addition | No | Yes | No | Yes |
| VFA addition | No | No | No | Yes, PFO |
| Primary sedimentation | No | No | No | Yes |
| Tertiary filtration | No | Yes | Yes | Yes |
| *System operating parameters* |
| Mainstream sludge age (d) | 13 | 6.4 | 7.3 | 10 |
| Mainstream HRTn (h) | 19 | 16 | 23 | 12 |
| Side-stream sludge age (h) | 47 | NA | 36 | 1.3 |
| Side-stream HRTn (h) | 0.9 | 0.4 | 2.9 | 0.9 |
| Side-stream HRTa (h) | 13.5 | NA | 36 | 1.3 |
| *Influent*a *parameters* |
| Influent flow (MGD)  | 3.0±0.9 | 20.9±1.9 | 5.2±0.8 | 2.6±0.2 |
| BOD (mg/L) | 236±92 | 263±38 | 284.4±69 | 240±53 |
| TKN (mg/L) | 29.4±11 | 43.7±5.4 | 48.2±6.8 | 44.1±4.8 |
| TP (mg/L) | 2.7±5.4 | 5.7±0.8 | 7.1±0.1 | 6.8±1 |
| BOD:P | 102±88 | 46.5±4.2 | 39±6.6 | 38.4±23 |
| *Effluent parameters* |
| BOD (mg/L) | 8.4±3.9 | 5.4±2.2 | 2.8±0.5 | n/a |
| TN (mg/L) | 7.6±3.1 | 16.4±2.3 | 2.1±0.7 | 5.8±4.3 |
| TP (mg/L) | 0.9±0.3 | 0.48±0.3 | 0.4±0.4 | 0.2±0.1 |
| TP removal (%) | 67% | 92% | 94% | 97% |
| *PAO/GAO populations* |
| *Accumulibacter*b | 6.2% | 5.1% | 6.4% | 7.6% |
| *Tetrasphaera*b | 20.2% | 19.2% | 15.3% | 18.1% |
| *Competibacter*+*Defluviicoccus*b | 0.3% | 4.2% | 0.7% | 0.5% |
| *Accumulibacter*c | 0.5% | 0.9% | 0.4% | 0.1% |
| *Tetrasphaera*c | 0.0% | 0.4% | 0.4% | 6.8% |
| *Competibacter*c | 0.0% | 0.0% | 0.0% | 0.0% |
| *Defluviicoccus* c | 0.0% | 0.0% | 0.1% | 0.0% |
| *Propionivibrio*c | 0.2% | 0.1% | 0.1% | 0.0% |

SSR: side-stream return activated sludge (RAS) fermentation; SSRC: side-stream RAS fermentation with supplemental carbon addition; SSM: side-stream mixed liquor suspended solids (MLSS) fermentation; UMIF: unmixed in-line MLSS fermentation; PFO: primary fermenter overflow; HRT: hydraulic retention time; HRTn: nominal HRT; HRTa: actual HRT; BOD: [biochemical oxygen demand](http://en.wikipedia.org/wiki/Carbonaceous_biochemical_oxygen_demand); TKN: total Kjeldahl nitrogen; TP: total phosphorus.

a: influent to the biological process;

b: determined by FISH;

c: determined by 16S rRNA gene amplicon sequencing.

**3. Figures**



Figure S1 Flow diagrams of the biological treatment process at the investigated S2EBPR facilities: (a) Side-stream RAS fermentation (SSR) configuration at the South Cary Water Reclamation Facility (SC), Apex, North Carolina; (b) Side-stream RAS fermentation with supplemental carbon addition (SSRC) configuration at the Westside Regional Wastewater Treatment Plant (WR), West Kelowna, British Columbia; (c) Side-stream mixed liquor suspended solids (MLSS) fermentation (SSM) configuration at the Cedar Creek Wastewater Treatment Facility (CC), Olathe, Kansas; and (d) Unmixed in-line MLSS fermentation (UMIF) configuration at the Kurt R. Segler Water Reclamation Facility (Hen), Henderson, Nevada. The figure is retrieved from Gu et al. (2019).





Figure S2 Relative frequency histogram and distribution curve of PO2- peak position of polyPs in SCRS-identified PAOs from a *Tetrasphaera*-enriched culture fed with sodium casein hydrolysate. (a), (b) and (c) show the results from three biological replicates (50 PAO cells per replicate) in the *Tetrasphaera*-enriched culture. Raman spectra are available upon request. FWHM: full width at half maximum.



Figure S3 Example 31P-NMR spectra of a) EDTA+NaOH extracted and b) EDTA extracted polyP sample. PP1: terminal phosphate groups; PP2 and PP3: the second and third post-terminal phosphate groups; PP4: the core phosphate groups. The 31P-NMR spectra of EDTA+sonication extracted polyPs were similar with that extracted by EDTA only.



Figure S4 Comparison of different extraction methods for polyP analysis with 31P-NMR, on four EBPR-SBRs operated with 3, 5, 10 and 20 days SRT. Primary Y-axis left: Visible polyP content represented by the 31P-NMR integral of the “middle” P signal (PP2+PP3+PP4) of polyPs, normalized by sludge VSS. Secondary Y-axis: PO4-P removal efficiency for each of the SBRs at the extraction sampling time. The asterisk labelled SBRs of 10 and 20 days SRT were experiencing deterioration due to low dissolved oxygen supply for a short period.



Figure S5 Relative frequency histogram and distribution curve of P–O–P peak position of polyPs in SCRS-identified PAOs from (a) a *Tetrasphaera*-enriched culture fed with sodium casein hydrolysate (n=355) and (b) an *Accumulibacter*-like culture fed with acetate (n=60). Raman spectra are available upon request. FWHM: full width at half maximum.



Figure S6 (a) Preprocessed single-cell Raman spectra (n=168) collected from the four lab-scale EBPR systems, clustered by HCA into each sub-OPUs, with the signature peaks for glycogen and PolyP labeled. (b) Dendrogram from HCA of single-cell Raman spectra, with major and sub-OPUs labeled. The spectra and dendrogram are color-coded based on Raman analysis as follows: pink−candidate PAOs; blue−candidate GAOs; black−other cells). The figure is retrieved from Li et al. (2018). Raman spectra are available upon request.

**Supplementary References**

Gu, A.Z., Tooker, N.B., Onnis-Hayden, A., Wang, D., Srinivasan, V., Li, G., Takács, I. and Vargas, E. (2019) Optimization and design of a side-stream EBPR process as a sustainable approach for achieving stable and efficient phosphorus removal, Water Research Foundation.

Li, Y., Cope, H.A., Rahman, S.M., Li, G., Nielsen, P.H., Elfick, A. and Gu, A.Z. (2018) Toward Better Understanding of EBPR Systems via Linking Raman-Based Phenotypic Profiling with Phylogenetic Diversity. Environmental science & technology 52(15), 8596-8606.

Meisel, S., Stöckel, S., Rösch, P. and Popp, J. (2014) Identification of meat-associated pathogens via Raman microspectroscopy. Food Microbiology 38, 36-43.

Onnis-Hayden, A., Majed, N., Li, Y., Rahman, S.M., Drury, D., Risso, L. and Gu, A.Z. (2020a) Impact of solid residence time (SRT) on functionally relevant microbial populations and performance in full-scale enhanced biological phosphorus removal (EBPR) systems. 92(3), 389-402.

Onnis-Hayden, A., Srinivasan, V., Tooker, N.B., Li, G., Wang, D., Barnard, J.L., Bott, C., Dombrowski, P., Schauer, P., Menniti, A., Shaw, A., Stinson, B., Stevens, G., Dunlap, P., Takács, I., McQuarrie, J., Phillips, H., Lambrecht, A., Analla, H., Russell, A. and Gu, A.Z. (2020b) Survey of full-scale sidestream enhanced biological phosphorus removal (S2EBPR) systems and comparison with conventional EBPRs in North America: Process stability, kinetics, and microbial populations. Water Environment Research 92(3), 403-417.

Pilatus, U., Mayer, A. and Hildebrandt, A. (1989) Nuclear polyphosphate as a possible source of energy during the sporulation of Physarum polycephalum. Archives of biochemistry and biophysics 275(1), 215-223.

Smith, S.A., Choi, S.H., Davis-Harrison, R., Huyck, J., Boettcher, J., Rienstra, C.M. and Morrissey, J.H. (2010) Polyphosphate exerts differential effects on blood clotting, depending on polymer size. Blood 116(20), 4353-4359.

Webb-Robertson, B.-J.M., Bailey, V.L., Fansler, S.J., Wilkins, M.J. and Hess, N.J. (2012) Spectral signatures for the classification of microbial species using Raman spectra. Analytical and Bioanalytical Chemistry 404(2), 563-572.

1. \* School of Civil and Environmental Engineering, Cornell University, 220 Hollister Hall, Ithaca, NY, 14853, United States.

E-mail address: aprilgu@cornell.edu (A.Z. Gu)

1 These authors contributed equally to this study. [↑](#footnote-ref-1)