# Cover Sheet

| Project Title | Advanced Biorecovery of Critical Minerals through AI/ML-Guided Design of Microbial Chassis and Bioadsorbent |
| --- | --- |
| Prime Lab | Lawrence Berkeley National Laboratory |
| Partner University | University of California, Berkeley |
| Principal Investigator’s Name, Email address, Phone number | Ning Sun, [nsun@lbl.gov](mailto:nsun@lbl.gov)  510-495-8024 |
| Co-Principal Investigators’ Name, Email | [Marcin Joachimiak](mailto:mjoachimiak@lbl.gov), [joachimiak@lbl.gov](mailto:joachimiak@lbl.gov)  Rebecca Abergel, [rjabergel@lbl.gov](mailto:rjabergel@lbl.gov)  [Romy Chakraborty](mailto:rchakraborty@lbl.gov),[rchakraborty@lbl.gov](mailto:rchakraborty@lbl.gov)  [Yasuo Yoshikuni](mailto:yyoshikuni@lbl.gov), [yyoshikuni@lbl.gov](mailto:yyoshikuni@lbl.gov)  N. Cecilia Martinez-Gomez, cecimartinez@berkeley.edu |
| Requested DOE Funding Amount | $850,000 |
| Period of Performance (months) | 12 |

# **Abstract**: The goal of this project is to understand the fundamental mechanisms of biorecovery and develop new microbial chassis and bioadsorbent harnessing their natural capacity to efficiently recover rare earth elements (REEs) through a modular, AI/ML-guided technology that integrates bioleaching, biomineralization, and bioadsorption. Building on LBNL’s established research expertise and leadership in synthetic biology, artificial intelligence, f-element coordination, and lab automation, the technology will provide an efficient and flexible alternative to conventional CM mining and solvent extraction. AI/ML will be embedded across the workflow—from identifying and engineering metal-sequestering microorganisms, novel REE chelators and proteins, to optimizing bioprocess parameters–and continuously synchronized through the unified CM/REE Knowledge Graph (KG‑CMREE). All data: from genes to high-throughput (HTP) strain screening, flows into KG‑CMREE that drives AI‑powered Design–Build–Test–Learn (DBTL) loops across every task. Ultimately, this biology-based, AI-guided technology aims to strengthen the domestic supply of CMs by enabling clean and adaptive metal recovery. Our primary focus in Year 1 will be on REE chelators and trafficking peptides/proteins. Once established for REEs, the technology can be adapted to other CMs in subsequent years. The AI/ML-guided biorecovery technology offers transformative benefits for U.S. supply chain resilience and scientific innovation.

# **Work Plan:** We will validate the design rules for microbial chassis and establish a modular, AI/ML-guided bio-recovery technology that integrates bioleaching, microbial biomineralization, and bioadsorption to reclaim REEs through four interrelated yet independent major tasks described below. **Task 1.0: Technical tool development to enable high-throughput (HTP) strain screening and validation** This task focuses on the development of technical tools, including computational pipelines to guide microbial/microbial consortium discovery, spectroscopic characterization assays for fast quantification of REEs, and HTP strain screening and validation through lab automation. ***Task 1.1: Construct KG and Establish AI/ML framework*** The AI/ML framework for microbial biomineralization centers on constructing a comprehensive AI-ready data substrate in the form of a Knowledge Graph for Critical Minerals and Rare Earth Elements (KG-CMREE). This resource integrates multi-scale biological information spanning genomic sequences, metabolic pathways, spectroscopic measurements, and bioprocess parameters through semantic data modeling, KG construction tools, and established ontologies [3,4](https://sciwheel.com/work/citation?ids=15134432,13181338&pre=&pre=&suf=&suf=&sa=0,0&dbf=0&dbf=0). Working with this core data resource we will apply both symbolic rule-mining to extract interpretable IF-THEN rules and advanced graph learning techniques including Graph Transformers[5](https://sciwheel.com/work/citation?ids=18048398&pre=&suf=&sa=0&dbf=0) to rank candidate taxa, pathways, and proteins for REE recovery. The hybrid computational-experimental framework will be orchestrated by an autonomous multi-agent system comprising specialized agents such as a Literature and Data Mining Agent that continuously updates KG-CMREE with LLM-assisted extraction from lathanome publications; an Experiment Design Agent that uses historical outcomes and predictive models; a Failure Analysis Agent that learns from unsuccessful experiment; and so on. The cooperating agents will enable semi-autonomous Design-Build-Test-Learn cycles for rapid strain and process optimization, automatically identifying promising genetic targets such as phosphate starvation responses (e.g., *pstS* deletion) that have demonstrated improvements in REE extraction efficiency. ***Task 1.2 Spectroscopic assay development.*** To enable HTP detection of REE-binding and accumulation, we will develop a time-resolved luminescence (TRL) assay that leverages the sharp, long-lived emissions of lanthanide ions (e.g., Eu³⁺ and Tb3+) stemming from f-f transitions [10,11](https://sciwheel.com/work/citation?ids=5424366,4750843&pre=&pre=&suf=&suf=&sa=0,0&dbf=0&dbf=0). Such assay will be implemented in microplates, with TRL collected from microorganisms incubated with low micromolar concentrations of metal salts. For improved sensitivity and specificity, sensitizing antenna ligands such as β-diketonates or aromatic chelators (e.g., hydroxypyridinone derivatives) will be included [12,13](https://sciwheel.com/work/citation?ids=18048433,18048419&pre=&pre=&suf=&suf=&sa=0,0&dbf=0&dbf=0). Confocal and electron microscopies will enable the localization of REE within the cells and REE uptake mechanism. Fluorescence-activated cell sorting (FACS) will be used to identify and isolate library members that show high REE binding. To engineered strains that can show high-selectivity binding against competing monovalent and divalent cations and to release REEs upon specific environmental triggers, we will apply HTP TRL and ICP-OES REE-binding analysis methods. ***Task 1.3 HTP strain screening pipeline*** We have developed modular high-throughput automation pipelines for strain engineering, capable of processing up to 6,000 samples per day [14](https://sciwheel.com/work/citation?ids=18048579&pre=&suf=&sa=0&dbf=0). Our pipeline supports a range of tasks, including automated *E.* *coli* transformation, plasmid validation, host colony picking, and sample processing for both genotyping (e.g., variant sequencing) and phenotyping (e.g., RapidFire assays, proteomics, fluorescent detection). By expanding the capacity of these automated workflows to support REE recovering strains, we aim to reduce processing time by over 90% compared to manual methods. This technology will accelerate the identification of high-performing strains for bioleaching or biomineralization. We will implement the TRL assay developed in *Task 1.2* for rapid quantification of REEs. Identified microbial strains as well as engineered mutants from Task 2 will be screened and validated using this workflow.

**Task 2.0: AI/ML-guided design of microbes for REE recovery through bioengineering** Biotechnology offers a promising path to reduce costs and environmental impacts through bioleaching, biosorption, and biomineralization. To fully unlock this potential, however, we need to engineer more robust, efficient, and versatile microbial chassis capable of operating in harsh environments typical of REE recovery processes.

***Task 2.1 Bacterial strain engineering to improve REE selectivity.*** We will engineer bacterial strains to hyper-accumulate with high selectivity, specific REEs from complex feedstocks. Microbes that require REE for growth at neutral pH must produce REE-chelators, also known as lanthanophores. We have characterized methylolanthanin (MLL) from the strain *Methylobacterium extorquens* AM1 [17](https://sciwheel.com/work/citation?ids=16733616&pre=&suf=&sa=0&dbf=0), yet the presence of genes encoding biosynthetic pathways for alternative lanthanophores is widespread and unexplored. Strains from the LBNL and the Martinez-Gomez collection (~20) that contain at least two mechanisms for REE use and transport and that grow in neutral pH or alkaline conditions will be grown comparing the use of REE in a soluble (chloride) and an insoluble (oxide) form. Transcriptomic profiling along with genomic analysis will validate the identification of novel biosynthetic clusters producing novel lanthanophores as previously achieved). We will replace native promoters with strong promoters upstream of the predicted biosynthetic clusters. We have demonstrated a 4-fold increase in REE bioleaching and bioaccumulation when upregulating the promoter of *mll* [8](https://sciwheel.com/work/citation?ids=18048428&pre=&suf=&sa=0&dbf=0). In addition, the strains with novel biosynthetic clusters will be screened for growth rate and growth yield using different REEs. We predict that different lanthanophores will have different affinities for each REE and can be monitored with growth, as demonstrated between methylolanthanin and rodopetrobactin. ***Task 2.2 Engineering next-generation bacterial and fungal chassis for REE recovery using AI-guided design.*** Building on strains from Task 2.1, we will apply Chassis-independent Recombinase-Assisted Genome Engineering (CRAGE)[19–21](https://sciwheel.com/work/citation?ids=9909268,11671138,7640687&pre=&pre=&pre=&suf=&suf=&suf=&sa=0,0,0&dbf=0&dbf=0&dbf=0) to domesticate and optimize both well-characterized extremophiles (e.g., *Acidithiobacillus ferrooxidans*, *A. thiooxidans*, *Leptospirillum ferrooxidans*, *Sulfobacillus thermosulfidooxidans*) and newly discovered bacterial isolates for high-efficiency bioleaching, biosorption, and biomineralization of REEs. CRAGE will enable precise integration of synthetic pathways, regulatory circuits, and barcoded libraries to enhance acid production, metal tolerance, biofilm formation, leaching/sorption kinetics, and biomineralization capacity. Graph learning techniques including Graph Transformers operating on KG-CMREE—combining sequence/context embeddings, genome architecture, and pH-resolved phenotypes—will predict CRAGE integration efficiency and engineered pathway stability under acidic conditions within the neurosymbolic loop. Predictions are continually validated through the publish/subscribe architecture linking all agents, with results feeding back into KG-CMREE to refine the neurosymbolic loop and guide design choices.. In parallel, we will engineer acid-, salt-, and temperature-tolerant fungal chassis such as *Pichia kudriavzevii*, leveraging prior success in biomineralization[22](https://sciwheel.com/work/citation?ids=17948483&pre=&suf=&sa=0&dbf=0).

**Expected Outcomes:** By the end of Year 1, we will validate key biological design rules and identify at least one novel microbial chassis, along with previously uncharacterized components of the lanthanome, to enable efficient REE recovery from e-waste. Major deliverables include:

* Establish spectroscopic assays for REE quantification in a high-throughput manner
* Identify and validate at least one microbial strain with high REE bioaccumulation through AI/ML
* Establish HTP pipeline to screen and validate high performance microbial strain/consortia for REE bioaccumulation
* Establish an AI/ML-guided framework to support core task activities and serve as a unified source of data integration and model predictions

**Appendix 2 – Bibliography**

[1.    Kaboli, E. *Critical Minerals and Materials for Selected Energy Technologies*. (2025).](https://sciwheel.com/work/bibliography/18048546)

[2.    Phogat, P., Kumar, S. & Wan, M. A scientometrics study of advancing sustainable metal recovery from e-waste: processes, challenges, and future directions. *RSC. Sustain.* **3**, 2434–2454 (2025).](https://sciwheel.com/work/bibliography/18048585)

[3.    Caufield, J. H. *et al.* KG-Hub-building and exchanging biological knowledge graphs. *Bioinformatics* **39**, (2023).](https://sciwheel.com/work/bibliography/15134432)

[4.    Unni, D. R. *et al.* Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. *Clin. Transl. Sci.* **15**, 1848–1855 (2022).](https://sciwheel.com/work/bibliography/13181338)

[5.    Dwivedi, V. P. *et al.* Relational Graph Transformer. *arXiv preprint* (2025).](https://sciwheel.com/work/bibliography/18048398)

[6.    Caufield, H. *et al.* CurateGPT: A flexible language-model assisted biocuration tool. *arXiv* (2024) doi:10.48550/arxiv.2411.00046.](https://sciwheel.com/work/bibliography/18048536)

[7.    Joachimiak, M. P., Caufield, J. H., Harris, N. L., Kim, H. & Mungall, C. J. Gene set summarization using large language models. *arXiv* (2024) doi:10.48550/arxiv.2305.13338.](https://sciwheel.com/work/bibliography/14876808)

[8.    Good, N. M. *et al.* Scalable and Consolidated Microbial Platform for Rare Earth Element Leaching and Recovery from Waste Sources. *Environ. Sci. Technol.* **58**, 570–579 (2024).](https://sciwheel.com/work/bibliography/18048428)

[9.    Schmitz, A. M. *et al.* High efficiency rare earth element bioleaching with systems biology guided engineering of Gluconobacter oxydans. *Commun. Biol.* **8**, 815 (2025).](https://sciwheel.com/work/bibliography/18048418)

[10.   Bünzli, J.-C. G. & Eliseeva, S. V. Intriguing aspects of lanthanide luminescence. *Chem. Sci.* **4**, 1939 (2013).](https://sciwheel.com/work/bibliography/5424366)

[11.   Bünzli, J.-C. G. Lanthanide luminescence for biomedical analyses and imaging. *Chem. Rev.* **110**, 2729–2755 (2010).](https://sciwheel.com/work/bibliography/4750843)

[12.   Tian, L., Dai, Z., Ye, Z., Song, B. & Yuan, J. Preparation and functionalization of a visible-light-excited europium complex-modified luminescent protein for cell imaging applications. *Analyst* **139**, 1162–1167 (2014).](https://sciwheel.com/work/bibliography/18048433)

[13.   Arnedo-Sanchez, L. *et al.* Combining the Best of Two Chelating Titans: A Hydroxypyridinone-Decorated Macrocyclic Ligand for Efficient and Concomitant Complexation and Sensitized Luminescence of f-Elements. *ChemPlusChem* **86**, 483–491 (2021).](https://sciwheel.com/work/bibliography/18048419)

[14.   Kuo, R. *et al.* A Modular High-Throughput Pipeline for Automated Microbial Strain Engineering: From Colony Picking to Phenotypic and Genotypic Analysis v2. (2025) doi:10.17504/protocols.io.261gerqkdl47/v2.](https://sciwheel.com/work/bibliography/18048579)

[15.   Jing, J., Garbeva, P., Raaijmakers, J. M. & Medema, M. H. Strategies for tailoring functional microbial synthetic communities. *ISME J.* **18**, (2024).](https://sciwheel.com/work/bibliography/16599038)

[16.   Chen, M., Acharya, S. M., Yee, M. O., Cabugao, K. G. M. & Chakraborty, R. Developing stable, simplified, functional consortia from Brachypodium rhizosphere for microbial application in sustainable agriculture. *Front. Microbiol.* **15**, 1401794 (2024).](https://sciwheel.com/work/bibliography/16487663)

[17.   Zytnick, A. M. *et al.* Identification and characterization of a small-molecule metallophore involved in lanthanide metabolism. *Proc Natl Acad Sci USA* **121**, e2322096121 (2024).](https://sciwheel.com/work/bibliography/16733616)

[18.   Good, N. M. *et al.* Hyperaccumulation of Gadolinium by Methylorubrum extorquens AM1 Reveals Impacts of Lanthanides on Cellular Processes Beyond Methylotrophy. *Front. Microbiol.* **13**, 820327 (2022).](https://sciwheel.com/work/bibliography/12749804)

[19.   Wang, B. *et al.* CRAGE-Duet Facilitates Modular Assembly of Biological Systems for Studying Plant-Microbe Interactions. *ACS Synth. Biol.* **9**, 2610–2615 (2020).](https://sciwheel.com/work/bibliography/9909268)

[20.   Ke, J. *et al.* CRAGE-CRISPR facilitates rapid activation of secondary metabolite biosynthetic gene clusters in bacteria. *Cell Chem. Biol.* **29**, 696-710.e4 (2022).](https://sciwheel.com/work/bibliography/11671138)

[21.   Wang, G. *et al.* CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol.* **4**, 2498–2510 (2019).](https://sciwheel.com/work/bibliography/7640687)

[22.   Müller, I. E. *et al.* Cost-effective urine recycling enabled by a synthetic osteoyeast platform for production of hydroxyapatite. *Nat. Commun.* **16**, 4216 (2025).](https://sciwheel.com/work/bibliography/17948483)