

Engineering Yeast for Heavy Metal Waste Remediation

by

George L. Sun

B.S., University of California, Berkeley (2014)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author
Department of Biological Engineering
August 9, 2019

Certified by
Angela M. Belcher
Professor of Biological Engineering and Material Science
Thesis Supervisor

Accepted by
Katharina Ribbeck
Professor of Biological Engineering
Graduate Program Chair

Engineering Yeast for Heavy Metal Waste Remediation

by

George L. Sun

Submitted to the Department of Biological Engineering
on August 9, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

The global rate of waste production has consistently outpaced the world's ability to manage and remediate it. Specifically, global consumption of raw materials, unrenewable energy sources, and disposal of electronic goods have contaminated water sources with heavy metals causing environmental damage and public health concerns. Despite the urgent need to contain and remove metals from the environment, there still does not exist robust and complete remediation technologies. Physicochemical technologies like chemical precipitation, absorption, and ion-exchange lack the specificity for metal capture, produce their own secondary-waste in the form of chemical by-products or sludge, and have a high cost barrier requiring development of dedicated infrastructure and technical expertise.

Instead, this work investigates biologically-derived strategies for managing waste, technologies also known as bioremediation. Principles from chemical precipitation, absorption, and ion-exchange were analogously designed in *S. cerevisiae*—the common baker's yeast. The three analogies were: engineering yeast sulfur metabolic pathways for controlled metal sulfide precipitation; designing new metal trafficking schemes using membrane metal transporters; and engineering supramolecular forming proteins for yeast-protein metal chelation and sequestration. For all methods, metal removal were between 50–90% efficiency for heavy metals such as Cu, Cd, Hg, and Pb. Furthermore, 2–4 rounds of processing eliminated almost 100 μM of metal, 100–1000 fold greater than EPA toxicity thresholds. Strategies to retrieve and recycle captured metals were also investigated, such as precipitating metal sulfide crystals onto the yeast surface, compartmentalizing metals into the yeast vacuole, or sedimenting bound metals into cell-protein complexes.

Relying on yeast takes advantage of their autonomous growth, ease of engineering, and its ubiquitous presence in the household and consumer market. The purpose of this work was to show that the same yeast used for brewing and baking can be harnessed for clean water applications.

Thesis Supervisor: Angela M. Belcher

Title: Professor of Biological Engineering and Material Science

Acknowledgments

To my advisor Angela M. Belcher, and my thesis committee members K. Dane Wittrup and Cathy L. Drennan. They have supported as well as pushed my growth as a scientist and critical thinker. To the members of the Belcher Lab, who have provided me a home to pursue my research ambitions: Peter Jansen, Ngozi Eze, Archana Iyer, Dahyun Oh, John Casey, Matthew Klug, Maraym Moradi, Xiangnan Dang, Hiroshi Atsumi, Briana Dunn, Geran Zhang, Griffin Clausen, Nimrod Heldman, Jifa Qi, Ching-Wei Lin, Swati Kataria, Eric Lehnhardt, William Records, Tyler Toth, Uyanga Tsedev, Jelle Van Der Hilst, and Erin Reynolds. Special thanks to members of the Amon Lab who have patiently taught me the fundamentals of yeast biology, and made me appreciate the rigor and meticulousness that goes into the work of a pure biologist: Christopher Brennan, Summer Morrill, Daniel Corbi, Hilla Weidberg, and Xiaoxue “Snow” Zhou. Also, to the numerous scientists in the MIT scientific community who have helped, taught, and expedited my research with their expertise; they come from: the Center for Material Science & Engineering, the Koch Facilities, the Whitehead Keck Microscopy Facility, the Center for Environmental Health & Sciences Facility, the Department of Chemistry Instrumentation Facility, the Center for Bits and Atoms Media Lab, and the numerous maker spaces around MIT.

To my previous advisors, Mikhail Shapiro and David Schaffer. My academic career at the University of California, Berkeley was arguably unparalleled and helped propel me to where I am today. I only wish I could be at the same intellectual and professional calibur as they.

To the Communication Laboratory, Diana Chen, Prerna Bhargava, and Sean Clarke who have facilitated my growth as a speaker and educator. To David Niño from the Gordon Engineering Leadership Program for his course 6.928J Leading Creative Teams which I helped taught. To Noreen Lyell, Leslie McClain, and Maxine Jonas for teaching an ambitious course, 20.109 Laboratory Fundamentals in Biological Engineering, and allowing me to teach alongside them.

To Anthony Santella, Professor at the University of New England, who enthusi-

astically introduced me to his scientific maker space, and including me in his student out-reach programs. I admire his ambition for promoting STEM outreach, and I was happy to have been involved.

To my adopted academic family, the Biological Engineering PhD class: Ally Huang, Jared Kehe, Aaron and Mary Dy, Brett Geiger, Julia Joung, Manu Kumar, Andee Wallace, Ross and Katy Mitchell-Jones, Allison Tisdale, Erika Handly, Zijay Tang, Linyi Gao, Claire Duvallet, Malvika Verma, Paul Reginato, and everyone else from the program.

To the Boston swing dancing community: Dennis Walsh, Chris Warren, Katy Huang and Ray Zavesky, Kipaya Kapiga, Laura Nakolan, Elisabeth Sepulveda, and Surhbi Dhar. I could not have asked for better moments of laughter, conversation, and fun.

To my old friends from California: Raymong Yang and Lin Lin Phan, James Bui, Nathan Lin, Eric Mao, Felix Huang, Lauren Tang and Brian Paz, Lawrence Huang and Kristin Au, Julie Liang, Juna Song, Helen Sun, Selena Shen, Kunal Mehta, Jess Wang, and Jason Chavez. I hope to be back in the Bay Area, to relive our fun memories and hopefully make more.

What has driven me this far was my family. To my sister, Rosy, who has always been an example of empathy and kindness. To my brother-in-law, Lawrence, whose willingness to help others inspires me. To their daughters, my nieces, Eleanor and Amelia Ong, who I hope I can make a better world for. To my in-laws, Grace and Eddie Ong, who have continuously supported our family. Finally, to my Mom and Dad; it has been my mission to prove to them that the sacrifices they have made, and are still making, will culminate to the success and enjoyment of mine and the future generations after me. I have truly stood on the shoulder of giants, and I hope not to squander the privileges I have been given.

Thank you

This doctoral thesis has been examined by a Committee of the
Department of Biological Engineering as follows:

Professor K. Dane Wittrup	Chairman, Thesis Committee
	Professor of Biological and Chemical Engineering
Professor Angela M. Belcher	Thesis Supervisor
	Professor of Biological Engineering and Material Science
Professor Catherine L. Drennan	Member, Thesis Committee
	Professor of Biology and Chemistry

Contents

List of Figures

List of Tables

Chapter I

Preface

I.I Talking about waste

Waste has inherently been a poor subject for discussion, the term naturally connotes disease, vileness, and ultimately death. Unfortunately these connotations are true, improper handling and negligence of waste do lead to disease and death. As the World Health Organization has estimated in 2016, nearly 12.6 million deaths are associated with unhealthy environments and living conditions. In other words, 1 in 4 of total global deaths are caused by unhealthy living conditions primarily contributed to wasteful and negligent environmental practices [**who2016**].

Ironically, the public perception of waste manifested from centuries of unhealthy and unsustainable waste management practices which gave rise to the associated maladies we describe waste today. Waste on its own is defined as materials eliminated or discarded from a used good or process. Be it excess materials, unwanted by-products, or goods that are consumed and transformed into an unwanted form. The subsequent dangers of waste: effluent, run-off, contamination, etc. are caused by a lack of sufficient handling and containment practices, and in extreme cases, active disregard for safe waste disposal. This meant discarding waste in dumps, landfills, and environmentally vulnerable locations. Until the 1920s, this laissez-faire approach for handling waste manifested into a global spread of contaminated areas. What we see today, waste as an abhorrent and dangerous entity, is the product of centuries of

inaction leading to the ill-definition we describe it today.

The discussion of waste is also actively suppressed, as more salient news eclipse its much needed discussion. Instead, society has been more moved by discussions of emerging technologies, new feats in engineering, and cures for diseases. These topics do deserve their headlines, but underneath every achievement is the hidden mark of waste. For example, putting a cell-phone in every hand has given rise to a new form of waste, electronic waste (e-waste). E-waste is now one of the fastest growing waste streams with a global production of 45 million metric tons equating to 6 kg (13.2 lbs) of discarded e-waste per person per year [**balde2017**].

The environmental and health dangers of mining for raw materials, expanding manufacturing processes, and the short life-cycles of electronic goods [**hong2015**] is often not seen by the consumer and subsequently not a concern for industry. Also, establishing proper waste treatment practices can be harmful to a company's immediate revenue performance, and even with potential long-term gains in building environmentally friendly practices investment in cleaner technologies may not outweigh immediate performance metrics [**king2001**]. Likewise, industries are often defensive, sometimes employing political or marketing tactics to circumvent new environmental policies. In some cases, policies are ignored completely [**nemerow1998, markowitz2013**]. For example, in 2013 Wal-Mart was accused and pleaded guilty of illegally dumping hazardous waste, almost 2 million pounds of pesticides in protected areas from 2006-2008 [**usatoday2013**], despite laws inacted to prevent this. There are many more examples of companies that have attempted to circumvent waste management laws, when in 2009 the New York Times cited that within their 5 year investigation almost 500,000 cases of illegal industrial waste practices were performed [**duhigg2009**]. Even on a city-wide scale, events such as the Flint water crisis and West Virginia chemical leak, both occurred in 2014, are still struggling to remediate their contaminated drinking waters because of slow moving politics and inadequate waste management infrastructure [**haimerl2017, kristabryson2016**].

This work is intended to not only show new waste management technologies, but to also encourage the conversation about our current state of waste. The goal is to

overcome the first barrier preventing productive waste management initiatives, and that is to acknowledge it as a problem.

I.I.I Waste – a historical context

Government waste management practices in the United States did not exist until the early 1800s [elliott2013]. Since then, the main waste management practice tacitly followed two principles. The first was to remove waste out of public sight, and the second was a belief that nature or some other entity would eventually deal with it [elliott2013]. These two mentalities were the hallmark of the Industrial Revolution (circa 1760-1840), in which waste was seen as an annoyance that interferred with productivity [williammcdonough1998]. However, as manufacturing and production grew, so did waste. The quantity of waste, whether it was manufactured by-products, excess chemicals, biological waste etc. were routeinly discarded haphazardly on the streets, buried, or dumped into nearby streams. It was not until the 1880s that waste became visibly obvious. Streets were dirty, waters were soured, and the connection between illnesses and waste became synced in the public's mind [louis2004].

Unfortunately, because municipal waste was disregarded for almost a century, there existed no waste management infrastructure and almost no government investment to build one [elliott2013]. In a reluctant series of finger pointing, the eventual clean up responsibility fell on to the city—but only loosely. Transitioning into the 1900s the municipal government assumed primary responsibility for solid waste streams produced at the local level and established related services such as collection and disposal. Industries on the other hand were held self-accountable of their hazardous waste [elliott2013]. Despite the switch to actively manage waste, there was no capital to build new waste management infrastructure, and sometimes building such infrastructure was insurmountable given city constraints. So solid waste management primarily relied on establishing municipal dumps where waste was collected and stored, with the hope that future technologies would eventually process it [louis2004, elliott2013].

Local municipals took the burden of finding and establishing new dumping

grounds, and industries continued their laissez-faire approach when managing their own. Given the separation of responsibilities there were often clashes when state or federal legislations implemented new waste management regulations.

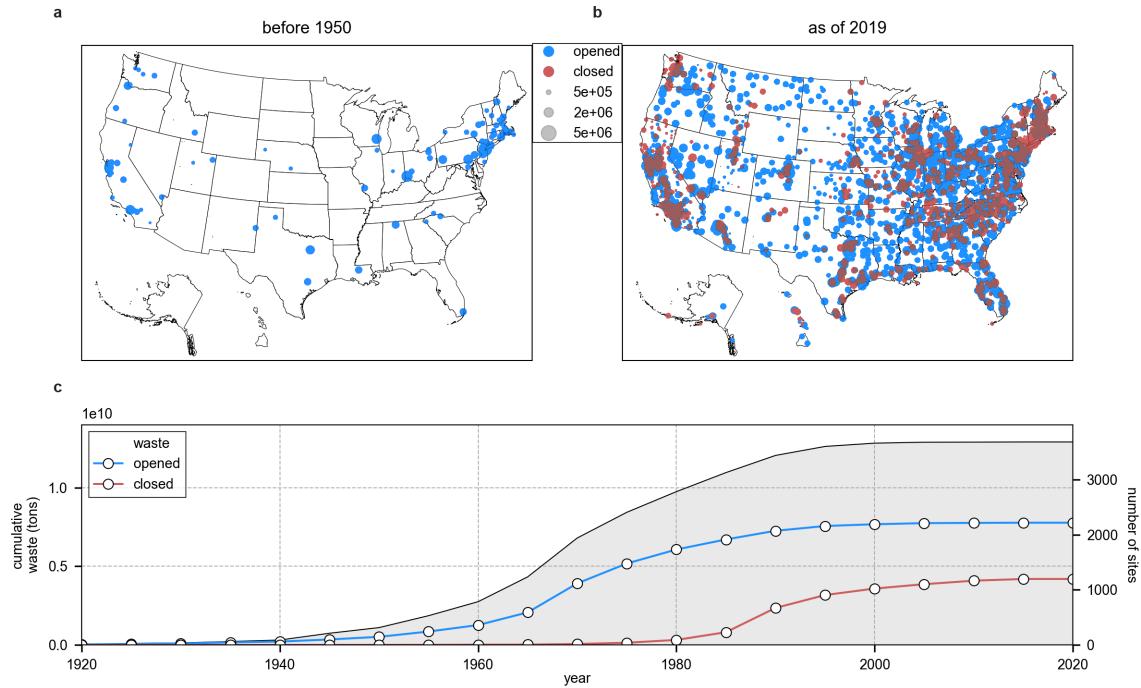


Figure I.I | Mapping and number of landfills in the United States from 1950 to 2019¹. (a) A map of the United States indicating locations of landfills during the first recorded year in 1950. Blue sites indicate opened (during that time), and radius of site corresponds to the amount of waste produced. (b) The same mapping in 2019. Blue indicates opened sites, and red indicates closed sites. (c) More than 10 billion tons of accumulated landfill waste had been discarded on United States' soil since reporting began in 1920. The rate of waste production accelerated between the 1960–80s; however, the production has slowly plateaued to a rate of >250 million tons per year [usepa2017b].

Industries would lobby for looser control and less oversight, hoping to keep their manufacturing practices unperturbed. The outcome was ineffective regulations that focused on specific waste streams rather than broadly tackling the overall inefficient

¹USA landfill and geolocation data available at: https://catalog.data.gov/dataset?metadata_type=geospatial&metadata_type_limit=0&q=Landfill
USA landfill sites by the year available at: <https://www.epa.gov/lmop/landfill-technical-data>

waste process [markowitz2013]. Many loopholes existed allowing many industries to continue their unregulated dumping. Rather than placing new guards around unsustainable waste practices, new legislation instead provoked the rise in organized resistance and industrial lobbying [elliott2013, markowitz2013].

Consequently, regulation of waste disposal did not consistently occur at the municipal, state, or federal level until the 1960s. At that time the quantity of waste became physically impossible to avoid, and the public soon became aware of its daunting health and environmental consequences. In the 1950s, there were 74 municipal landfills across the United States, each on average containing more than 1.2 million tons of waste (Figure ??). Within 10 years, the number of landfills more than doubled to 192 sites.

Public awareness and activism was also spurred by the ongoing social and civil rights movements. The momentum for transforming public opinion into political initiatives also pushed environmental protection and sustainability legislation into the government agenda [cutter1995]. Especially, the book Silent Spring by Rachel Carson (1962) also contributed to a social movement for environmental protection and subsequently helped enact actionable policies, namely the Clean Air Act in 1963 and the eventual formation of the National Environmental Policy Act in 1969. Municipal waste management (MSWM) activities finally received defining legislation in 1976 through the Resource Conservation and Recovery Act (RCRA) which forced closure of open dumps and required regional planning of municipal waste [louis2004].

However, these policies had their loopholes and negative repercussions. With the closing of municipal dumps, waste had to be transported elsewhere such as dedicated processing plants. Unfortunately, handling transportation of waste, the involved labor cost, and the potential tax implications of moving waste between state lines², made waste processing much more logically challenging. Therefore, waste was still largely handled by municipalities, and often municipals would contract private companies to move waste between facilities or statelines much like exporting or importing a good.

²waste was now defined as a good of commerce by the United States' Supreme Court under the Commerce Clause of the Constitution [louis2004]

The net result was exactly the same as before but on the national level, the movement of waste between state lines to government owned landfills [**elliott2013**]. Unfortunately, the growth in waste regulation from 1960-1990 did not correlate to a reduction in waste or improved waste management [**andrews2018**].

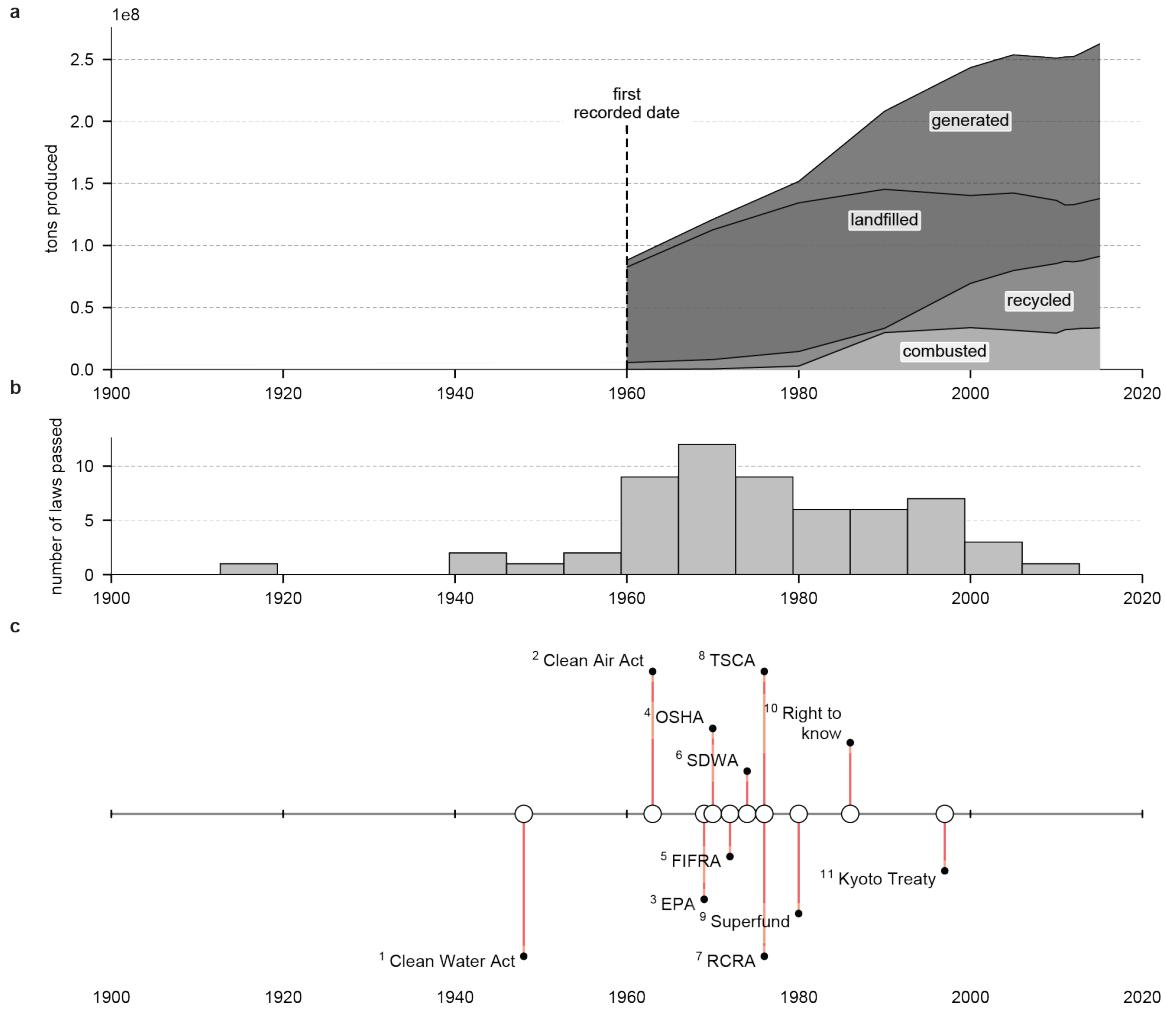


Figure I.II | The rise in waste and government policies³. (a) The growth of recorded municipal waste since 1960 until 2016. The shaded regions breakdown the total generated waste into landfilled, recycled, or combusted (burned). (b) A histogram indicating the breakdown of waste related legislation in the United States starting as early as 1910 to 2019. There were more than 57 major legislative actions concerning waste starting in 1916, with a spike in 1960 during the passing of the Clean Air Act and the EPA. The number of legislation has reduced since the 90s, as legislation at the state and federal level have more relied on global NGO initiatives (e.g. WHO, UNEP, etc.) which has now become the governing body for waste management policies. (c) Highlighted environmental policies enacted that laid the groundwork for waste management in the United States.

³USA wastestream type and quantities: <https://catalog.data.gov/dataset/sustainable-materials-management-smm-materials-and-waste-management-in-the-united-states-key-f>
list of EPA policies: https://en.wikipedia.org/wiki/Environmental_policy_of_the_United_States

In the 1990s, disposal of hazardous waste increased by 25% [**usepa2015**, **usepa2017b**, **andrews2018**]. Loopholes and exemptions encouraged waste generators to dilute their waste with nonhazardous waste streams and incentivized cheating or mis-reporting numbers. Overall, the rise in environmental laws did not curb the amount of waste or disincentivize illegal practices (Figure ??) [**elliott2013**, **colten2010**]. To this day, the result is more than 2000 active and 1000 closed or abandoned landfill sites. Many of the abandoned sites still contain waste and have yet to be properly remediated. Currently, the United States produces more than 250 million tons of waste per year [**usepa2017b**], which is equivalent to approximately 4.5 pounds per person per day (Figure ??).

The villainization of waste was not because waste was inherently thought of as an evil entity, as for almost two centuries waste was mostly viewed as an inconvenience. Instead, the villainization of waste grew through the insufficiency of cities and indifference of industries that allowed waste to grow to an unmanageable quantity we see today. Today's anxiety about waste production and waste management is warranted, but only because waste has now accumulated to a point that is impossible to ignore.

I.I.II Waste – current perspectives

To understand the current and future implications of waste and its growing new forms, a societal picture should be examined to explain how changing moods and public opinion could have a profound impact on waste management decisions. A growing number of literature concerned about a new age of modernity was written as a reaction to the techno-social revolution from the 90s onwards. This techno-social revolution was seen as an increased interconnectivity between countries, industries, and people [**ekberg2007**]. This body of literature suggested that society had entered a new modernity, where nations were no longer bound by geographical lines and risks were inherently made in everyday geopolitical decisions. A framework to view this shift in societal behavior was pioneered by Anthony Giddens, Scott Lash, and Ulrich Beck's theory of the "risk society" [**elliott2013**, **giddens1991**, **beck1992**].

In their definition, a risk society is mainly concerned about future goals, whether policies, technologies, or shifts in public opinion. Society is reactionary rather than preemptive or anticipatory to the risk involved, be it changes in government, the scientific community, or environmental health.

For example, the rise in gene editing technologies have faced morale and ethical questions [**wang2019**], yet the scientific advancement of the technology will continue forward until more tangible consequences emerge. Unfortunately, these “tangible consequences” are ill-defined and are often manufactured in an attempt to provide substance to a future unknown. Contrary to the era of the industrial revolution, risks/concerns were mainly revolved around tangible and quantifiable items, such as products and distribution of goods. In a risk society, risks are conceptual and ultimately manufactured given expectations and opinions. These are called “manufactured risks”. To continue the example, a manufactured risk for gene editing technologies would be the possibility of abusing genetic modifications to unfairly enhance success (beauty, intelligence, immunity, etc.). Other day-to-day manufactured risks may be whether companies will remain solvent, rise and trade of investments and stocks, and the hope of emerging technologies such as new therapeutics or quantum computing to alleviate current stresses (e.g. diseases and aging technologies) [**beck1994, beck2006**]. This concept goes beyond technology to ecological disaster, diseases and epidemics, population growth, food and water limitations, and territorial conflicts [**beck1994, elliot2013**]. Many of these manufactured risks have no historical reference to provide estimation, so they are largely unpredictable.

In Gidden, Lash, and Beck’s discussion, societies reflexive response to risk shapes the new modernity, in other words, “reflexive modernization” [**beck1992, beck1994**]. Reflexive modernization is a main characteristic of risk societies, that is concerns (i.e. risk) are focused on large-scale techno-scientific processes pursued by industrialists, engineers, and policymakers while equally downplaying the negative repurcussions of such actions in order to emphasize economic growth and convey a sense of progress. Given these concepts of reflexive modernity, the rise in risk societies, and the creation of manufacturered risk, there are three main conclusions that can explain today’s

loose and reluctant behavior in addressing the current waste crisis. The first is that globalization has made risk societies extend beyond geographical borders. Risks are now shared between countries, and risks such as waste is increasingly burdened onto transnational forces such as corporations and NGOs like the United Nations and World Health Organization rather than the nations themselves. The second is the redefinition of waste as a manufactured risk. When speaking about waste, it is no longer an object or a thing, but a concept. Rather than handling the current physical mass of waste, policies are aimed at reducing the notion of waste as a harm to society. Statements such as “zero-discharge” and “zero-emission” are intended to convey a proactive stance on waste management, even though producing zero waste is nearly impossible for any process [wang2004]. Lastly is the question of who is responsible. As waste is now interpreted as a manufactured risk, it is easily passed through multiple hands with varying degrees of interest. Ironically, it is the same political, business, and scientific expertise which are called upon to answer the problems of earlier wasteful trends generated from previous political, business, and scientific actions [beck1992]. Examples would be the concern over nuclear waste and plastic waste, which prior to their now known environmental hazard, were used freely without much regard about their disposal. As these same politicians, businessmen, and scientist are more concerned about progress, the nature of this risk is slowly dissipated through cursory arguments leaving the public more entertained by other concerns such as the internet, new technologies, and the economy [andrews2018].

I.I.III Waste – current and future trends

Past actions (or inactions) towards waste and the transition into the reflexive modernity suggest that society as a whole reacts to waste rather than actively managing or preemptively preparing for it. Align with the concept of the risk society, many nations are more concerned with growth, namely the developing world such as China, India, South American, and African countries [cohen2006urbanization]. Trends in a country’s population and gross domestic product (GDP), preliminary metrics to estimate a country’s growth, can be correlated with waste production to help elucidate

the relationship between a country's development and waste output. Data from 2017 on population, GDP, and waste production of 217 countries, broken down to high and low economic groups, were fitted with a linear model. An increase in population and GDP, regardless of economic status, correlates strongly with an increase in waste production (Figure ??). Particularly, high income countries (HIC) have a higher ratio of waste production per capita, on average generating 69 tons of waste per thousand citizens. Whereas for low income countries (LIC), as well as upper and lower-middle income countries (HMC and LMC, respectively) waste production is almost 5 times less per capita (Table ??a).

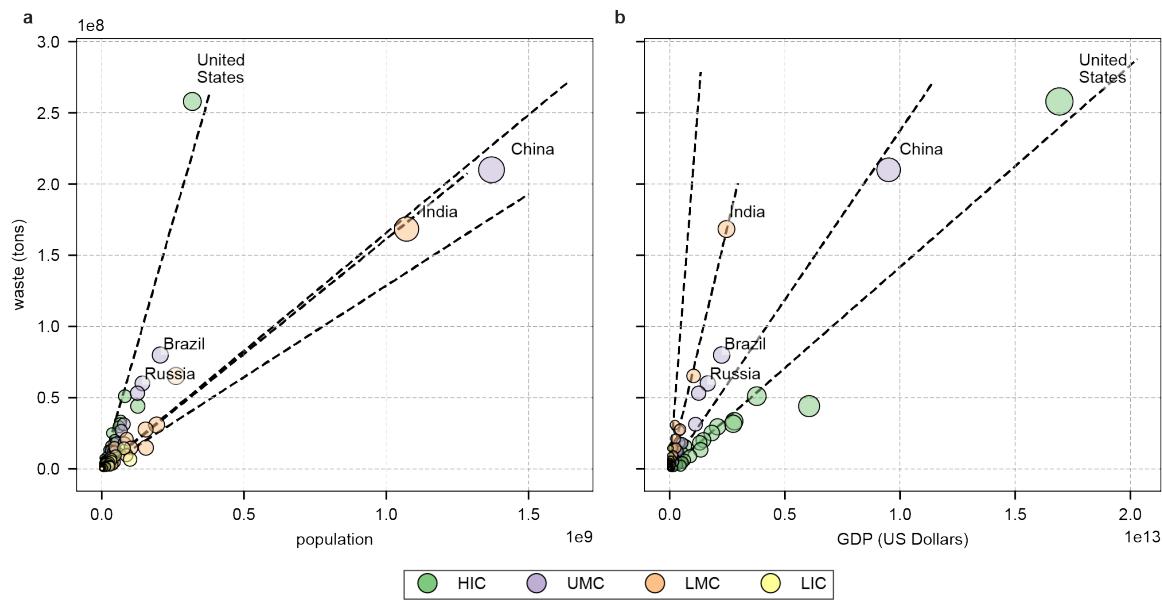


Figure I.III | Correlation between waste production and population & GDP for 217 countries, broken down by economic status⁴. HIC = high-income country LIC = low-income country, LMC = low-middle income country, UMC = upper-middle income country. (a) Trendlines fitting waste production against country population levels as of 2017. (b) Trendlines fitting waste production against country GDP (normalized against the US dollar) as of 2017.

⁴world population data available at: <https://data.worldbank.org/indicator/SP.POP.TOTL>
world GDP data available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?year_high_desc=false

global municipal waste data available at: <https://datacatalog.worldbank.org/dataset/what-waste-global-database>

classification of countries based on economic status: http://faculty.ucr.edu/~jorgea/econ181/wdr_2008.pdf

Looking at the waste relationship with GDP, lower income countries such as India, Brazil, and China produce more waste per GDP than HICs such as the United States (Table ??b). This may suggest that countries with higher economic income can afford, or become more educated, on the impacts of waste and have the social and financial bandwidth to make the appropriate policy or technological changes.

(a)			(b)		
	tons per 1000 people	R ²		tons per million \$	R ²
HIC	69.4	0.94	HIC	14.2	0.96
UMC	16.6	0.91	UMC	23.7	0.96
LMC	16.1	0.97	LMC	67.8	0.99
LIC	12.8	0.68	LIC	206	0.59

Table I.I | Regression model parameters on waste production as a function of population or GDP. (a) Fits between waste production and population level suggest that HIC countries produce more waste per capita. Units are in tons per thousand people. (b) Countries with higher GDPs, such as the United States, correlate to less waste production (HIC < UMC < LMC < LIC). Units are in tons per million US dollars.

A similar trend discovery experiment can be performed for more country-specific metrics such as the level of urbanization, government budget (e.g. EPA), and the advancement of technology (e.g. Moore's law, pre-2010). A multi-variate time-dependent cross correlation on waste production versus these other metrics was performed for the United States. An analysis on the United States' waste trend could help foreshadow the trajectory of waste production for other countries such as China, India, and Brazil; countries that are currently encountering similar growth challenges the United States once did in the past decades.

Waste output for the United States has grown by 3.6 million tons per year, equivalent to 2 kilograms (4.5 pounds) of waste per person per day [usepa2017b]. This trend was cross-correlated with time-series for population growth, GDP, rise in technological power (indirectly through transister counts), and percent urbanization since

the 1970s. The time-dependent Pearson coefficient shows that many of these trends positively correlate with higher waste output (Table ??). One negative correlation was in environmental government spending, specifically the EPA, suggesting that invested capital in regulations may have an impact on waste reduction. However, more policies correlate to an increase in waste production. This reversal may be explained by observing that major political actions in the 1960s–80s coincided with major industrial changes resulting in a positive correlation with waste output (Figure ??) [louis2004, andrews2018].

	Pearson	p-value	Spearman	p-value
Budget (EPA)	-0.705	6.6E-08	-0.791	1.0E-10
policies	0.991	4.9E-39	0.989	1.7E-37
GDP	0.927	5.9E-20	0.990	1.9E-38
moore	0.468	1.2E-03	0.989	1.7E-37
population	0.957	1.1E-24	0.989	1.7E-37
urbanization	0.941	6.7E-22	0.989	1.7E-37

Table I.II | Time-dependent cross correlation of the United States municipal waste output compared to sociotechno-trends using Pearson’s and Spearman’s coefficients⁵. Sociotechno-trends such as population growth, GDP, and urbanization correlate strongly with an increase in waste production. The EPA budget has correlates negatively, primarily due to the decrease in funding since 2010 onwards.

Although it should be emphasized that correlation does not mean causation, these trends do support an intuitive conclusion. More people and more spending lead to higher amounts of waste production. Waste is inevitable, and should be instead accepted as a natural part of growth and development rather than being a mark of shame or reason for accusation. Unfortunately, population and technological growth

⁵EPA budget available at: <https://www.epa.gov/planandbudget/budget> USA GDP available at: <https://fred.stlouisfed.org/series/GDP> Moore’s law based on number of transistors per microprocessor available at: <http://crab.rutgers.edu/~sundares/MIS334Sec40.Sp08/protected/notes/hw/moores.xls> USA population data available at: <https://www.census.gov/programs-surveys/popest/data.html> percent USA population in urbanized cities available at: <https://ourworldindata.org/urbanization>

has risen exponentially, leading to an equivalent rise in waste production that has outpaced society's ability to manage and remediate it. The result, an unrelenting production of waste led by the most powerful and fastest growing countries (Figure ??).

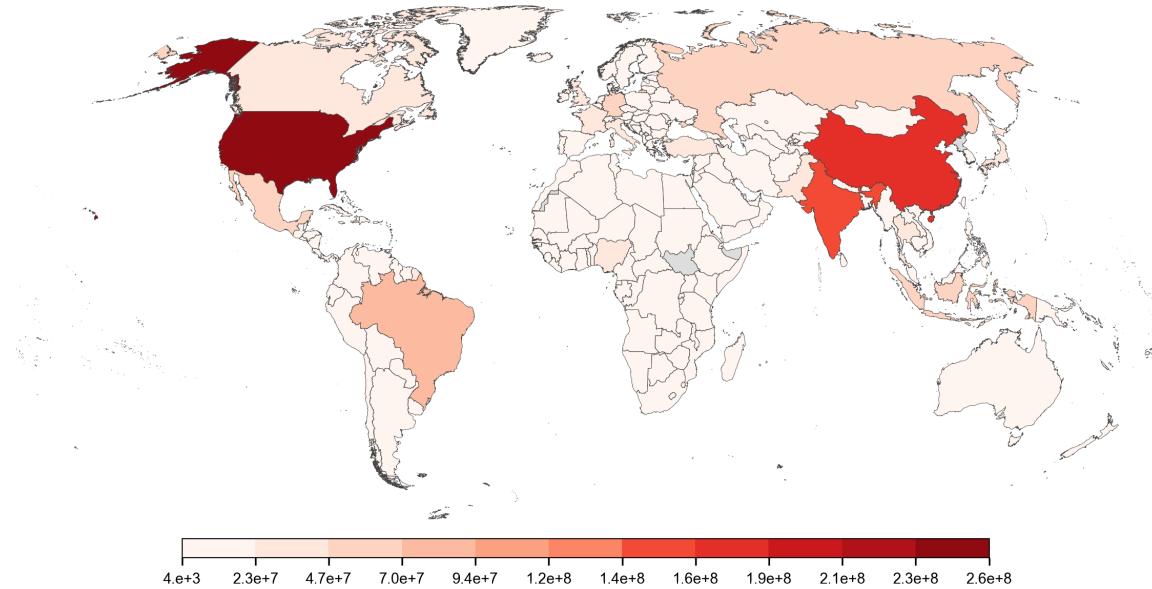


Figure I.IV | Waste generated per country, color-coded by quantity (in billions) as of 2018⁶.

The United States, China, and India produced over 600 million tons of waste in 2019 [**theworldbank2018**], exceeding the production of the next 20 countries combined. This trend will continue to rise as the world population is expected to grow another 40%, totaling 9.8 billion by 2050 [**unitednations2017**]. Fortunately (or unfortunately) the ability to sustain this growing population will be matched by access to the earth's available raw resources. In terms of energy, the earth is projected to have 155 years of processible coal, 51 years for oil, and 65 years for gas given current population and consumption trajectories [**shafiee2008**, **worldcoalassociation2015**].

⁶data on waste production per country: <https://datacatalog.worldbank.org/dataset/what-waste-global-database>

Peak production of oil is expected to begin in 2023, and peak production for natural gas, coal, and uranium will occur by 2050 [**mason2007**]. And despite slowing of Moore’s law, the market for electronic goods will continue to rise by 6% per year until 2024 [**robinson2009, hong2015, zionmarket2018**]. The world’s population and economy is not limited by natural resources or energy requirements. Instead, what may prevent sustainable population and economic growth is the waste produced from all of these processes. Therefore, for sustained growth waste must be managed and mitigated. Unfortunately, policies enacted by individual countries have been ineffective, ultimately transferring responsibilities to NGOs such as the United Nations or the World Health Organization [**elliott2013**]. To tackle the continuing buildup of waste there must be an equivalent advancement and scaling of technologies to counteract its global buildup.

I.II Rise of heavy metal waste

One of the greatest environmental concerns is the rise and pervasiveness of heavy metal waste. Growing populations and an increase in technological demand have driven the activity of mining and manufacturing of raw materials and electronic goods to unprecedented levels. The result is a surge in environmental heavy metal contaminants such as chromium (Cr), copper (Cu), arsenic (As), cadmium (Cd), mercury (Hg), and lead (Pb), to name a few. For example, Mining of fossil fuels have pulled metals from the Earth’s crust to the surface creating tailings that flow into nearby streams and water beds. Similarly, effluent from manufacturing sites are often dumped or left to flow into adjacent soils which contaminate water and agricultural sources [**rattan2005**]. Without any protective barriers, heavy metals eventually return to the public via contaminated water and food. The toxic effects of heavy metals are broad, ranging from gastrointestinal, epidermal, neurological diseases, cancer, and in acute dosages death (Table ??) [**singh2011**].

Heavy metal waste is becoming more of a concern given the exponential rise in electronic waste (e-waste). Production of e-waste is led by the United States,

Europe, and China, generating almost 50 million tons in 2016 [robinson2009, unitednationsuniversity2017].

Metal	Source	Effect	toxic level(ppm)
Chromium (Cr)	mines, chemical industry, metallurgy, dyes and pigments	nervous system damage	0.1
Manganese (Mn)	welding, fuel addition, ferromanganese production	damage to central nervous systems	0.05
Iron (Fe)	naturally occurring in soils, minerals, and the earth's crust. 2nd most abundant metal on the earth's crust	plant growth inhibition, corrosion of the gastrointestinal tract, metabolic interference, accumulation in organs	0.3
Cobalt (Co)	mining, electronic and battery production	inhibits DNA repair, genotoxic, dermatitis	n/a
Nickel (Ni)	metal refineries, coins, electronics, battery, and magnet production	creates free radicals, haemotoxic, neuro and genotoxic	0.1
Copper (Cu)	mining, pesticides, chemical industry, metal piping	anemia, liver, kidney, stomach and intestinal damage	1.3
Zinc (Zn)	refineries, brass manufacture, metal plating, plumbing	skin damage, nervous system	5
Arsenic (As)	pesticides, fungicides, mteal smelters	bronchitis, acute poisoning	0.01
Cadmium (Cd)	electronics, batteries, nuclear power plants, fertilizers, pesticides	Renal damage, lung disease/cancer, bone defects, respiratory damage/cancer, blood, gastrointestinal	0.005
Mercury (Hg)	pesticides, batteries, lighting, (old) electronics, paper industry	tremors, damage to nervous systems, poisoning	0.002
Lead (Pb)	paint, pesticides, smoking, automobiles, mining, burning fossil fuels	mental retardation, development delay, neural diseases, chronic damage to neural, gastro, liver, and kidney	0.015

Table I.III | Sources of heavy metals and their effect on human health⁷.

Globally, this is equivalent to 6.5 kilograms (14 pounds) of generated e-waste per person on average; however some countries such as Australia have higher generation of e-waste per-capita at 17.3 kg (38 pounds) per person. E-waste is expected to grow another 17% by 2021, and will lead as the world's fastest growing waste stream [unitednationsuniversity2017]. Unfortunately, only 20% is collected or recycled, the remaining 80% is landfilled or exported to other countries [schmidt2006, unitednationsuniversity2017]. What this means is that beyond trying to manage already harmful practices such as mining and manufacturing, the advent of electronic waste will only exacerbate the heavy metal waste scenario.

The repercussions of the rise in heavy metal waste is the toxification of both urban and environmental settings. Heavy metal contamination has disproportionately affected developing countries, especially those in Africa, South America, and the Middle East. Access to unsafe and poorly sanitized living areas have contributed to poisoning and death (Figure ??). Deaths rates are more concentrated in LICs such as Niger and Chad, whereas the likelihood of deaths related to heavy metal poisoning are less frequent in higher income countries (Figure ??a). LMCs and LICs are 10–50% more likely of heavy metal related death, and since the 2000s the rate of death has either remained constant or increased (Figure ??b).

⁷table modified from Singh et al., and Jaishankar et al. metal permissible as by the EPA: <https://www.epa.gov/dwstandardsregulations>

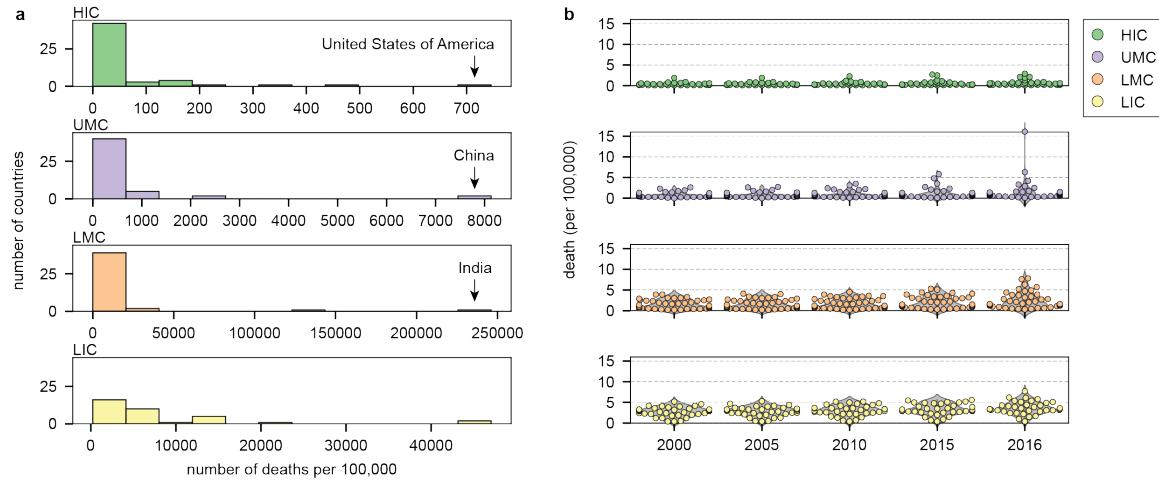


Figure I.V | Deaths related to poor water access linked to heavy metal contamination⁸. Death rates per country broken down by economic status (HIC, LIC, LMC, UMC) **(a)** Number of deaths attributed to poor water sanitation. **(b)** Time serie scatter plots of number of deaths since 2000 attributed to poor water sanitation.

The rise in heavy metal waste has dramatically impacted the quality and safety of drinking waters. Polluted waters due to heavy metal leaching have starkly increased the number of deaths in the African and South American countries (Figure ??). About 2.1 billion people lack access to safe drinking waters and more than 4.5 billion, more than half of the world's population, lack managed sanitation services from their community or government [unitednations2015, unu-inweh2017]. Again, countries such as the United States and European countries face little sanitation or water accessibility challenges, while the burden is disporportional higher in LMC and LICs (Figure ??).

⁸data on mortality rate attributed to unsafe waters available at: <http://apps.who.int/gho/data/node.main.INADEQUATEWSH?lang=en>
data on global causes of death available at: <https://ourworldindata.org/causes-of-death>

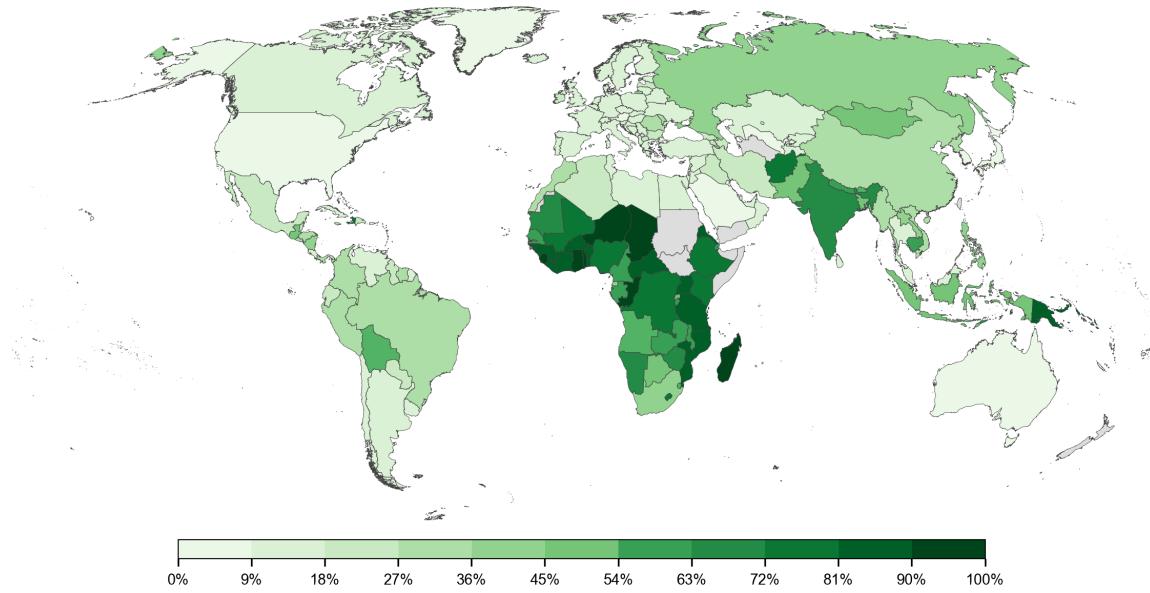


Figure I.VI | Mortality rate per country due to unsafe water, sanitation, and hygiene services⁹. Data from 2016 collected from the World Health Organization. Color map of light-to-dark indicate the rate of death due to exposure or use of unsafe water related sanitation services.

⁹global water sanitation and mortality rate data available at: https://www.who.int/gho/phe/water_sanitation/burden/en/index3.html

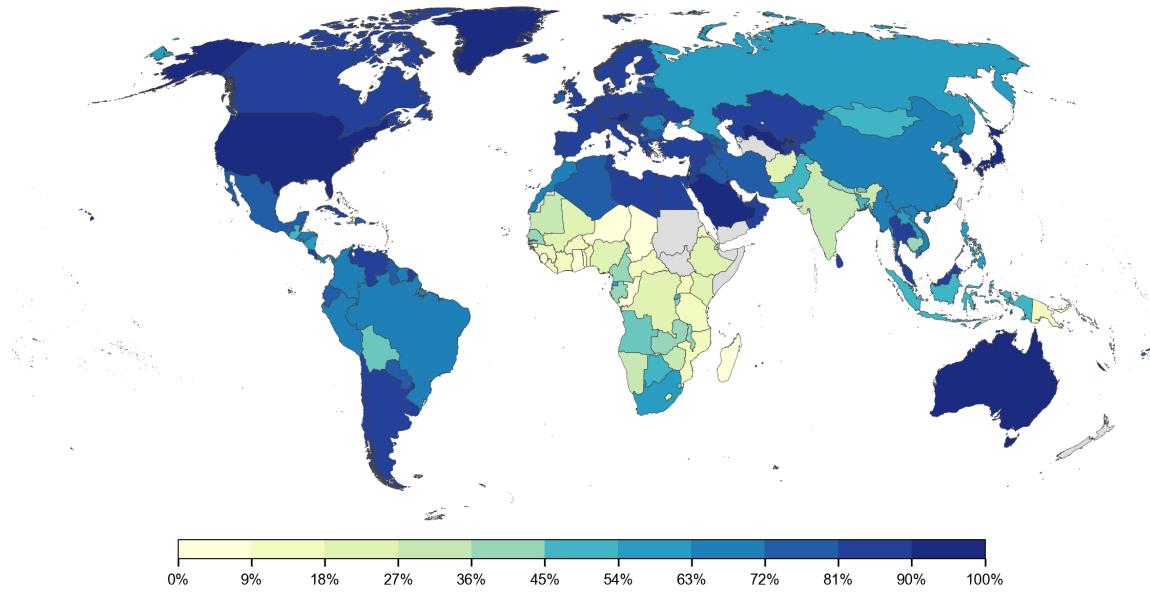


Figure I.VII | Access to clean water per country¹⁰. Almost 850 million people lack access to clean water, many residing in LICs and LMCs.

The abundance of raw materials and energy sources will continue to be available on Earth for the next several decades (Section ??). However, water will become increasingly scarce as the remaining drinkable reservoirs are depleted or actively polluted over these next decades. The water burden, calculated by water usage with respect to the available water per country, show that HICs and UMCs are actively draining their water resources faster than any other countries (Figure ??). If this continues, by 2035 more than 40% of the world will live in seriously water-stressed areas. What makes this situation worse is due to growing population levels and agricultural demand the global need for water will increase by 50% during this time [**unu-inweh2017**]. So not only are countries with low economic power facing extreme water safety challenges, countries with higher economic power are draining the remaining accessible waters at a disproportional rate.

¹⁰global water accessibility data available at: <https://ourworldindata.org/water-use-sanitation>

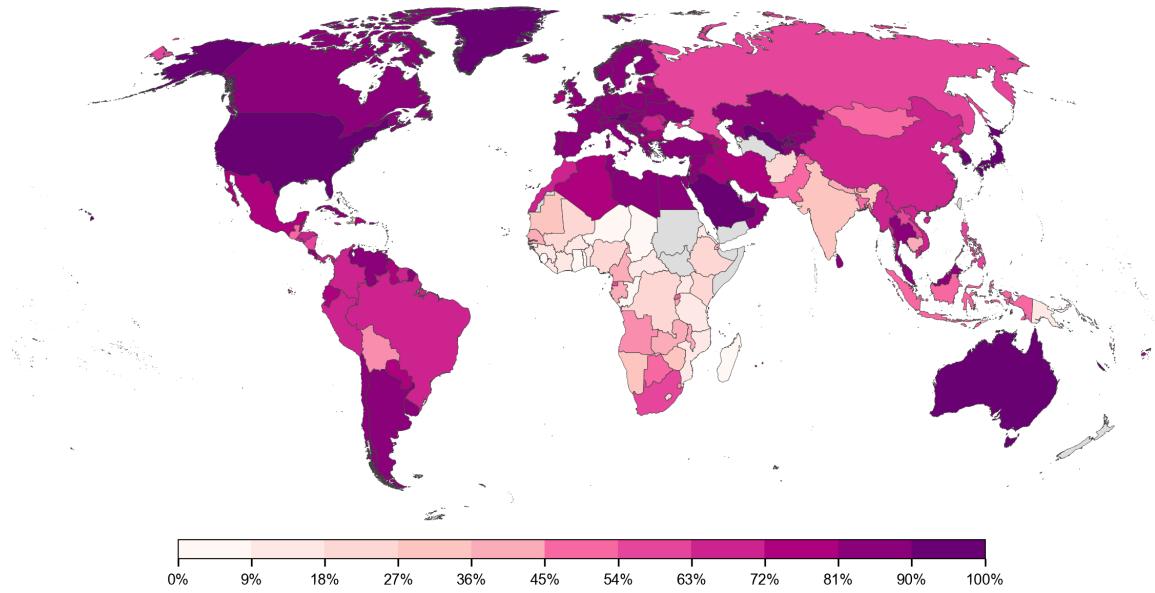


Figure I.VIII | Percent water usage (i.e. water burden) with respects to the total amount of available water per country¹¹.

I.III Concluding remarks

The rise in waste has resulted in a variety of maladies, from contamination of food and water, health concerns, and ecological damage. However, the negative association surrounding waste is anthropogenically rooted and can be traced back to poor decisions made at the national, government, and public level. However, to blame the public or the government for the current waste scenario is not appropriate or even correct. The purpose of the models and trends described in this work was intended to convey that waste is an inevitable—and arguably a natural—by-product rather than an abhorrent phenomenon. No matter how much pressure is made on the national and international level, humans will continue to make waste. It is inevitable. Instead, if a group or entity should be blamed it should be on the scientific and engineering com-

¹¹global water usage data available at: <https://ourworldindata.org/water-use-sanitation>

munity. We are responsible for providing our communities with technological answers to difficult questions, and right now that is waste. So far we have performed well in bringing about the Industrial Revolution and now the Silicon and Telecommunication Revolution, but there has only been a murmur in the waste management field. The work herein attempts to motivate a new technological focus on waste remediation and environmental technology. The same manufacturing and technological revolution that has helped push the world into today's new scientific era deserves a complimentary technology to remove, remediate, and recycle the dangerous by-products that come from it.

Chapter 1

Technologies in waste remediation & management

1.1 Current technological strategies

Heavy metals are typically transition metals ranging from chromium to mercury while also including some group III, IV, and V elements like indium, lead, and arsenic. However, the context at which these elements appear also define whether they are biologically harmful or not. In trace amounts, elements such as zinc, iron, and even chromium have biological function as enzymatic co-factors or essential minerals [althuis2002]. However, in excess these elements can have deleterious effects such as poisoning, cancer, and even death. Metal chemistries have been well-studied scientifically, and knowledge of how they react, bind, and transform into other compounds have been utilized for physical and chemical removal from waste waters—methods also known as physicochemical processes.

1.1.1 Physicochemical strategies

Physicochemical processes include all methods that use chemicals or manufactured materials to react, absorb, or bind onto waste [fu2011]. When dealing with heavy metals, the primary examples are chemical precipitation, ab(b)sorption, and ion-

exchange.

Chemical precipitation

Waste water is subjected to high pH (8.0–11.0) where metal solubilities are low. The water is then treated with hydroxides ($-\text{OH}_2^{1-}$) or sulfides ($-\text{S}^{2-}$) to precipitate metals [charerntanyarak1999heavy, ku2001photocatalytic]. The result is a transformation of dissolved or emulsified waste into a solid precipitated mass. This solid heterogeneous mass is also known as sludge. The conversion from aqueous metals into a solid allows for easier physical separation from the liquid fraction which is typically filtered out afterwards (Figure ??a).

Absorption

Absorption techniques rely on micro- and mesoporous structures to capture metals via sterics and/or binding interactions. The most widely used absorbents are activated carbons (Figure ??b) [jusoh2007simulation, kang2008sorption]. In consumer settings, sorbents such as activated carbon are packed into cartridges or filters where water can be poured from one end and strained at an outlet leaving the contaminants behind. Typically, sorbents can be reused by substituting bound toxins with a competing molecule, or using pressure and/or heat to forcibly expel the captured metals from the sorbent bed. However, sorbents such as activated carbon are often discarded during consumer use, ultimately throwing away the captured toxins back into the environment.

Ion-exchange

Ion-exchange takes advantage of electrostatic interactions to bind charged species onto a column or packed container. The most common synthetic resins are strongly acidic sulfonic groups ($-\text{SO}_3\text{H}^-$) or weakly acidic carboxylic groups ($-\text{COOH}^-$) (Figure ??c) [alyuz2009kinetics]. Ion-exchangers are difficult to customize and manufacture, so scientist have recently looked at naturally available substances (however,

non-biological) exchangers such as zeolites, an aluminosilicate mineral with microporous sieves, for example [motsi2009adsorption].

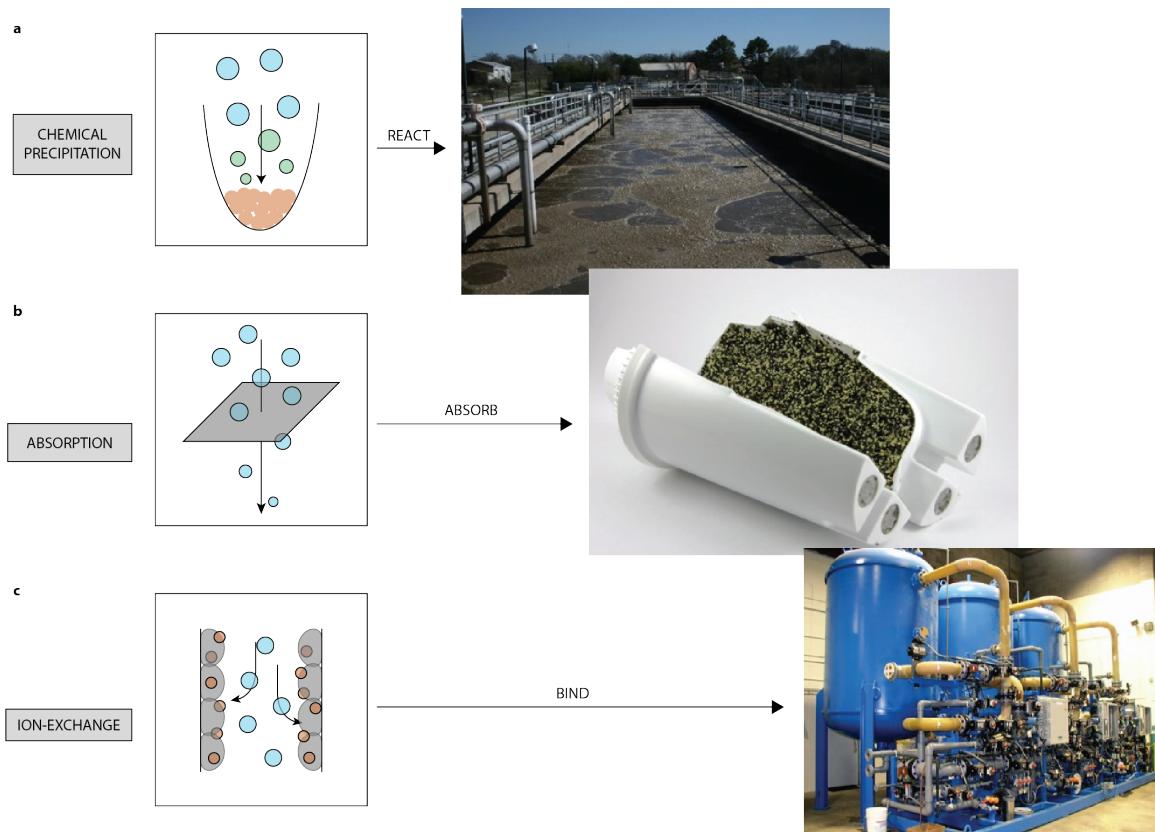


Figure 1.1 | Overview of physicochemical processes used in industry¹. (a) Chemical precipitation relies on high pH and the addition of reactive species to precipitate metals from solutions. The benefit is the conversion of aqueous metals into solids which can be physically handled and removed. (b) Adsorption or absorption is the capture of metals onto or into a membrane or solid matrix, respectively. Well-known examples used in the market are activated carbon or synthesized semi-porous meshes. (c) Ion-exchange relies on the electrostatic interactions of metals and resins. Resins are physically enclosed in columns or chambers in which waste water is poured into and eluted from. Metals captured can be later discharged from the resin and collected.

¹Representative images of chemical precipitation, absorption, and ion-exchange come from online stock photos. All rights belong to their respective creators or owners.

1.1.2 Disadvantages and limitations

Problems current physicochemical methods fail to address are cost, generation of secondary-waste, and technological accessibility. Almost all physicochemical processes require dedicated infrastructure to operate and manage. In addition, these processes require scientific skill and technical expertise to operate with regards to using reactors, synthesizing membranes or resins, or handling and storage of chemicals. Given these responsibilities it is extremely difficult for developing nations to adopt physicochemical technologies. Unfortunately, it is these same countries which have the most desperate need for economical and sustainable remediation technologies (Chapter ??). Without a viable remediation technology, many countries are unable to manage their accumulation of waste. An investigation by the United Nations revealed that almost 80% of global waste water is left untreated and allowed to flow back into drinking waters [unitednations2015, unu-inweh2017]. The problem is not the mechanism of action or removal efficiencies of current physicochemical technologies, but more so on the logistical and accessibility challenges that reduce the ease of adoption and use.

Chemical precipitation

The precipitation of heavy metals is highly reactive and kinetically fast, allowing other particles to non-specifically aggregate within the precipitating mass [fu2011]. The result is the generation of amorphous sludge. Sludge is typically heterogenous and chemically ill-defined, inhibiting further processing or extraction of individual elements. Therefore, typical disposal methods rely on dumping sludge into landfills or pyrolyzed [kurniawan2006, kumargupta2012]. Although chemical precipitation offers the convenience of extracting solid waste from the cleared liquid, the generation of sludge as a secondary-waste continues the cycle of ineffective waste treatment. Currently the handling and disposal of sludge is under investigation as an environmental and human health hazard [EPA_chemprecip_review], as previously buried sludge is beginning to leach back into soil and public waters.

Absorption

Heavy commercial use of current sorbent materials such as activated carbon have increased prices and significantly reduced availability. Research into new sorbents have been limited, as substituting materials such as carbon nanotubes and nanostructure materials are difficult to construct and are costly to synthesize [stafiej2007adsorption]. These barriers: cost, manufacturing difficulty, and material availability have limited widespread adoption of synthetic sorbent materials [fu2011]. Instead, there has been a push to discover natural sorbents such as rocks, minerals, and even composted biomass [razmovski2008biosorption].

Ion-exchange

Resin modifications are difficult to tailor per metal of interest, therefore monolithic chromatography columns are employed to indiscriminately capture free floating metal species. This method easily saturates resin beds with more abundant metals rather than metals that are rarer but acutely toxic (e.g. mercury versus sodium) [ustun2007regeneration]. In addition, resin synthesis, column construction, and the need for dedicated fluid control infrastructure produces an added layer of logistical and technical difficulty for mainstream adoption. In addition, creating resins and modifying its metal-binding functional groups are difficult to engineer and can add cost to research and development [wachinski2016]. Often, the manufacture of resins produce secondary-waste in the form of chemical by-products during resin synthesis or regeneration of resins after use [amin2015].

1.2 New and developing tools in bioremediation

Rather than focusing on man-made techniques for waste management, current research is moving towards cheaper and environmentally friendlier techniques which take advantage of natural waste processing strategies found in nature. Biologically-derived or bio-inspired solutions are attractive methods because nature has already evolved numerous mechanisms for decomposing hazardous substances into less toxic

or usable forms. The concept of waste management is universal in all biological systems, and these processes have been optimized in many organisms allowing them to survive in a variety of environments which are normally toxic to humans. The first example of utilizing biology for targeted remediation efforts was performed in the 1960s by George Robinson. Robinson demonstrated a bio-facilitated waste treatment reaction using bacteria to degrade petroleum and other polluting hydrocarbons [sonawdekar2012bioremediation]. Since then, the field has grown to include other natural and engineered organisms to react, absorb, and compartmentalize toxic metals. Until the 1990s, the methodology for designing bioremediation platforms was to engineer available microorganisms to enhance their tolerance and consumption of waste, specifically for oil and oil spills. However, as results did not improve during the years, from the 1990s onward scientists began to search for natural microorganisms that natively harbored desirable bioremediation characteristics [sonawdekar2012bioremediation, antizar-ladislao2010]. The discovery of new and exotic organisms became possible with the advent of new genetic and sequencing technologies. Examples of adopted microorganisms are bacteria that live in highly corrosive volcanic vents to plants that thrive in polluted soils [bull2000search].

1.2.1 Biological analogies to physicochemical processes

Deconstructing the mechanisms behind physicochemical technologies, namely chemical precipitation, absorption, and ion-exchange, reveal straightforward processes relying on fundamental scientific principles. Simply, these processes are to *react*, *absorb*, and *bind* heavy metals away or out of contaminated waters. These mechanisms are not particularly unique to physicochemical processes. Biological systems routinely perform such functions to remove metals from their environments for survival purposes. In this work, several discoveries were made to find biological analogies to these physicochemical processes, and from there engineered for improved usability, scalability, cost effectiveness, and environmental sustainability (Figure ??).

Physicochemical process		Bioremediation equivalent
chemical precipitation	→	biomediated mineralization
adsorption	→	metal trafficking and cellular uptake
ion-exchange	→	metal binders and cell surface display

Table 1.1 | Analogies between physicochemical and bioremediation strategies

In a biological context, homeostatic balance of metal concentrations is essential—too much or too little will lead to cell death. Therefore, biology has evolved exquisite mechanisms to balance the flow of metals in to and out of a cell. Not surprisingly, the ability to react, absorb, and bind onto metals can be found in most biological systems (Table ??). These biological processes just come in another name; they are biominerization, intracellular metal transport, and cell/protein mediated metal chelation.

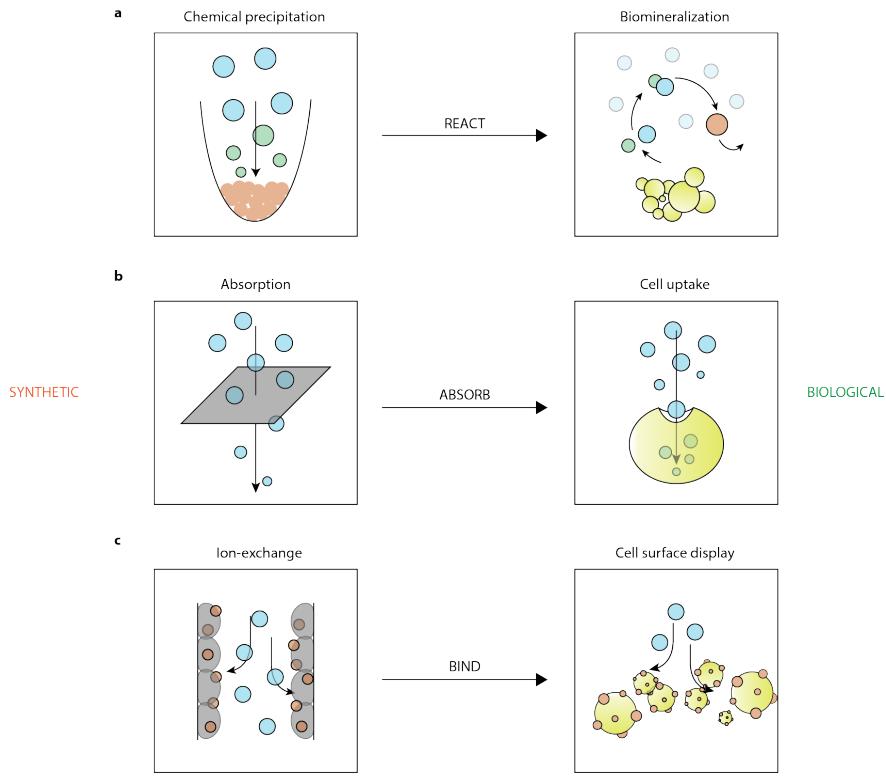


Figure 1.2 | Analogies between physicochemical and biological processes for heavy metal removal. (a) There are a variety of microorganisms that can produce their own chemicals for metal precipitation, a process commonly known as biominerization. Examples are magnetosomes from magnetotactic bacteria, abalone shells, and diatom exoskeletons. (b) The intracellular volume of cells can be used as a compartment to absorb metals. Rather than manufacturing synthetic absorbing matrixes, biology can naturally grow their own. (c) The combination of cells and proteins can be used as a physical substrate to chelate and aggregate metals, much like in ion-exchange.

Biomineralization as an analogy to chemical precipitation

Biology has evolved unique strategies to coexist with inorganic materials, and in some instances, productively utilize these materials for biological purposes—bones, shells, exoskeletons, etc. The process to biologically synthesize and incorporate mineralized materials can be grouped into two modes, biologically induced mineralization (BIM), and biologically controlled mineralization (BCM) [frankel2003biologically].

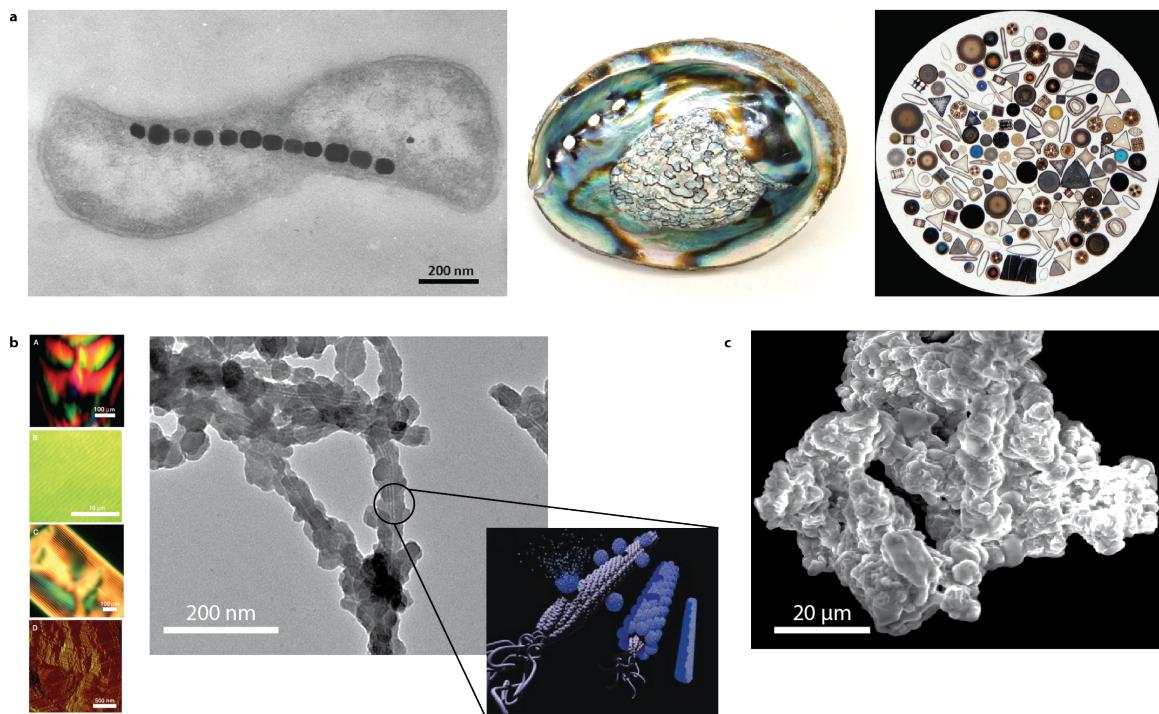


Figure 1.3 | Natural and engineered biominerization processes². (a) (from left to right) Magnetotactic bacteria, abalone shells, and diatoms are examples of natural organisms that mineralize elements such as silicon, iron, and calcium to create biologically functional structures. (b) An example of phage used to template zinc sulfide nanocrystals to create quantum dot materials. (from left to right) Modifying the pVIII coating of the M13 bacteriophage allows controlled biominerlization alignment, as represented by the striations and color of the material. A TEM image of the underlying phage and coat material, with inset showing the process of templating phage with semi-conductor metals. (c) The same engineering approach to coat phage can be performed on other host organisms, such as yeast. SEM image shows yeast aggregated into a matrix of cells and CdS nanoparticles.

BIM typically occurs for organisms that use metals as terminal electron acceptors which subsequently react and participate in mineral formation. BCM controls metal mineralization by facilitating growth in organic matrixes or vesicles within the cell (Figure ??a). An organism discovered in the 1980s that has received incredible interest in the past decades is the magnetotactic bacteria [blakemore1982]. Magnetotactic bacteria are able to create magnetic particles in their bodies with extreme precision and control. Scientist are still uncovering the magnetosome formation mechanism, but research into this phenomenon shows that nature has found a way to naturally manufacture electromagnetic compounds with environmentally available minerals [uebe2016magnetosome].

Using BIM and BCM, biological engineers have attempted to synthesize and create novel biologically derived materials for applications in semiconductor or biomaterial fabrication. A major breakthrough in bridging the gap between organic and inorganic materials was performed by Lee et al. who demonstrated the BIM-like synthesis of ZnS crystals on engineered M13 bacteriophage coat proteins (Figure ??b) [lee2002ordering]. More so, other groups such as Tan et al. and Naik et al. have engineered small peptides to synthesize and control the formation of Au and Ag nanoparticles, respectively, in a BCM-like fashion [tan2010uncovering, naik2002biomimetic].

Alternatively, BIM and BCM-like processes can be exploited to precipitate and remove heavy metals from water. Through the use of phage and bacterial display, researchers have discovered peptides that nucleate minerals such as SiO_2 , TiO_2 , and $\text{Ca}_3(\text{PO}_4)_2$, and a comprehensive review by Chen et al. provides a table listing identified biomineralization peptides [naik2002silica, dickerson2008identification, gungormus2008regulation, chen2010peptide]. Efforts to use biomineralization for the creation of novel materials can be equally applicable for metal precipitation and waste water remediation. These same peptides can be used to help mineralize

^{2a)} magnetotactic bacteria, abalone shell, and diatoms taken from Google images. All rights belong to their respective creators or owners.

^{b)} phage images taken from Lee, Seung-Wuk, et al. "Ordering of quantum dots using genetically engineered viruses." Science 296.5569 (2002): 892-895 [lee2002ordering].

and react metals from environmentally contaminated waters (Figure ??c).

Heavy metal transport as an analogy to adsorption

Plants have evolved one of the more advanced mechanisms for heavy metal tolerance. As plants lack the ability to migrate away from toxic areas, they instead evolved robust defense mechanisms to absorb and sequester metals away from vulnerable tissues [hall2002cellular]. Plants that are able to preferentially absorb and accumulate metals are classified as hyperaccumulators. There are at least 400 species of identified hyperaccumulators. The defintion of hyperaccumulation being the accumulation of 100 mg/kg (0.01% dry wt.) of cadmium or arsenic, 1000 mg/kg of (0.1% dry wt.) of cobalt, copper, chromium, aluminum, nickel and lead, and 10,000 mg/kg (1% dry wt.) of manganese, iron, and zinc [rascio2011heavy, vara2003metal, branquinho2007revisiting, kramer2010metal].

A review by Clemens et al. highlights the significance of metal transporters and chelating agents in contributing to plant's hyperaccumulating capabilities [clemens2002long]. Hypothesized transporters implicated in hyperaccumulation are CDFs (cation diffusion facilitators), special classes of passive transporters such as ZIPs (ZRT and IRT-like proteins), Nramps (natural resistance associated macrophage proteins), heavy metal ATPases, and ABCs (ATP binding cassettes) (Figure ??) [clemens2002long, kramer2010metal].

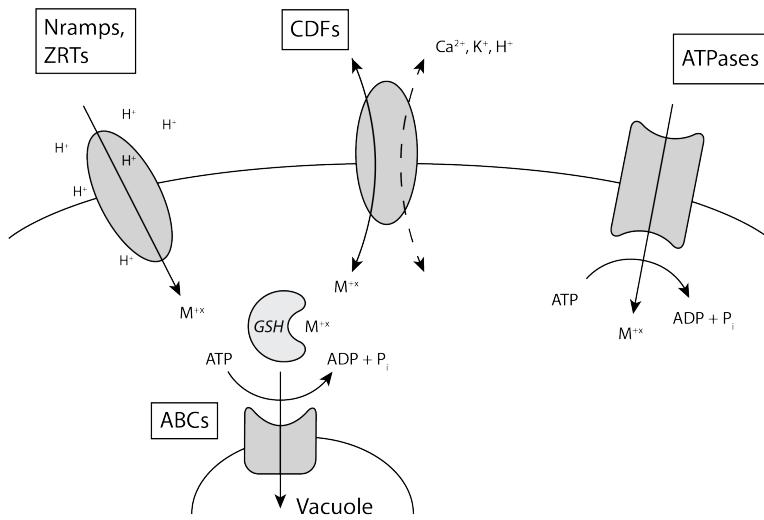


Figure 1.4 | Overview of metal trafficking and sequestration mechanisms in plant cells

Further work by Clemens et al. demonstrated that the ZIP family contributes to Fe^{2+} and Zn^{2+} hyperaccumulation, whereas the Nramp families have a broad spectrum of metal uptake capabilities [clemens2001molecular]. On the other hand, Pittman et al. showed that knockouts of antiporters, namely members of the CDF family led to hypersensitivity and hyperaccumulation of Zn^{2+} , Co^{2+} , and other trace metals [clemens2002transporter, pittman2005managing]. ABC transporters have recently been shown to transport metals into vacuoles after intracellular uptake, and it is hypothesized that ABC transporters recognize metal-phytochelatin complexes (e.g. glutathione or trypanothione) which they then carry into vacuoles as phytochelatin-mediated conjugates [hall2003transition, legare2001leishmania].

Transporter	Metal	Where	Requires
CDF	Mn/Fe/Zn/Co/Cd	varies	antiports counter ions
ZIP	Fe/Zn	membrane	proton gradient
Nramp	Fe/Zn/Mn/Co/Ca/Cu/Ni/Pb	membrane	antiports H ⁺
ATPases	Cu/Ag/Zn/Cd/Pb/Co/Ca/Mg	membrane	ATP
ABCs	phytochelatin + metal complex	vacuole	ATP

Table 1.2 | Transporter types listed by preferred metal, localization, and mechanism of action

The ability to uptake and compartmentalize metals also enables elevated tolerance to toxic levels of environmental metal concentrations. Tolerance is primarily due to compartmentalization inside vacuoles [song2010arsenic] or binding from sequestration agents such as phytochelatins or metallothioneins [cobbett2002phytochelatins]. Phytochelatins are oligomers of glutathione (GSH; γ Glu-Cys-Gly), and metallothioneins (MTs) are a family of cysteine-rich low molecular weight proteins. Both species utilize cysteine's thiol group to sequester heavy metals, and both have promiscuous binding affinities for mono- and divalent metals such as Zn, Hg, Cu, As, and Ag [cobbett2002phytochelatins]. Overall, metal accumulation in plants is possible due to a combination of hyperactive transporters and a defense network of metal chelating molecules.

Protein-metal chelation as an analogy to ion-exchange

Ion-exchange modifies the surface chemistry of resins to electrostatically bind onto metals. The same technique can be done for cell surfaces, in which the exterior cell wall or membrane can be biochemically modified to display chelating moieties for metal capture. This technology called cell surface display has been engineered in several hosts such as bacteria and yeast in which an extracellular anchor (in bacteria LamB/OmcAl; yeast AGA1/2) is fused to a protein of interest (POI) (Figure ??a) [**freudl1986cell, boder1997yeast**]. The POI is then tethered to the cell wall or membrane thereby decorating the surface for extracellular binding. From this point of view, a cell is no different than a resin bead; sizes are approximately 1–10 μm in diameter, and both surfaces can be functionalized with various metal binders (Figure ??b).

The advantage of using cell surface display rather than ion-exchange is that displayed proteins can be easily engineered and modified. Additionally, cell production is autonomous, easily stored and handled, and the resources needed to maintain large quantities of cells may be more cost-effective than handling synthetic compounds. In addition, surface binding proteins can be engineered to become highly sensitive and selective for a specific metal through high-throughput screens and/or directed evolution [**peelle2005**]. A common challenge for ion-exchange is the saturation of resin beds with more abundant metals, such as sodium, over highly toxic yet less abundant heavy metals such as mercury. In a biologically-derived system it is possible to engineer protein-metal interactions to be highly selective, thereby reducing interference of background metals during the remediation process.

Recently, work performed by Ruta et al. functionalized the yeast surface with hexapeptides to capture a range of common divalent metals such as Ni, Cu, Fe, etc [**ruta2017**]. In addition, cells with displaying metal binding proteins tend to be more metal tolerant, as metal captured extracellularly are prevented from entering the cell body [**ruta2017, kuroda2001cell**]. However, current literature data show low binding capacities in the μM range and poor capture to cell weight ratio compared

to ion-exchange resins. Binding capacities on a per weight scale for ion-exchange is 3–6 orders of magnitude greater than using bacteria or yeast display technologies [barakat2011new, stathi2010heavy]. Therefore the number of binding sites per cell is the greatest limiting factor in removing an impactful amount of metals per volume, and so far display technologies for waste removal has not been able to overcome this limitation.

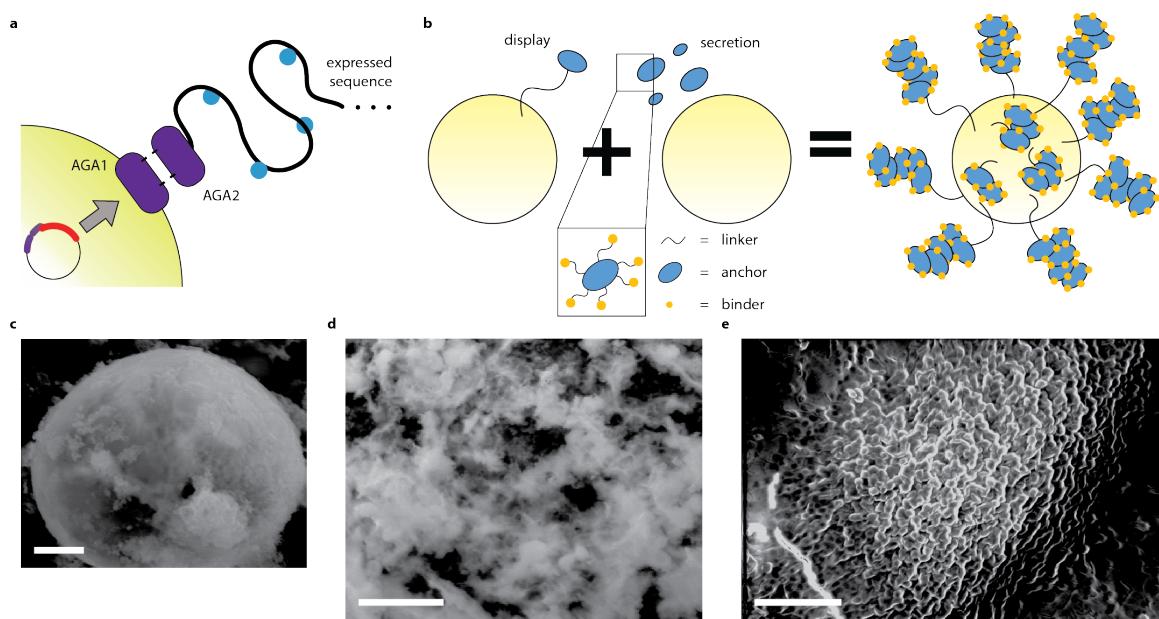


Figure 1.5 | Using cell display and protein-metal binding technologies to remove metals. (a) Example schematic of yeast display using the AGA1/2 system for display of metal binding peptides. Purple ovals represent the AGA1/2 anchoring domains, and blue spheres represent areas of metal binding. (b) To improve metal binding capacities, a system with yeast could both (or simultaneously) secrete and display aggregating proteins (blue) fused with metal binding moieties (orange) to effectively increase the surface number of metal binding domains. Rather than with traditional display technology of a single peptide/protein molecule per display, this system allows for multiple displayed domains to be anchored onto the cell surface. (c) Example of yeast display bound to precipitated copper metal. Scale bar represents 1 μm . (d) Example of yeast display with aggregating proteins to create a yeast-protein matrix for metal binding. Scale bar represents 10 μm . (e) A hybrid approach in which yeast that display and secrete aggregating proteins are embedded in a bacterial biofilm for metal capture³. Nodules represent cluster of yeast colonies. Scale bar represents 50 μm .

1.2.2 A reason for biology, and the focus on yeast

Biology has provided a gambit of organisms and biological pathways to handle and utilize metal compounds in the environment. However, there is no one microorganism that can provide all of the functionality described above (Section ??). Therefore a candidate microorganism needs to be selected for further engineering.

However, rather than taking a bottom-up approach in which requirements and design criterias are first drafted and tested per microorganism of interest, this study focuses on a top-down approach in which high level specifications such as cost, scalability, and engineerability are first considered to then narrow in on a candidate microorganism.

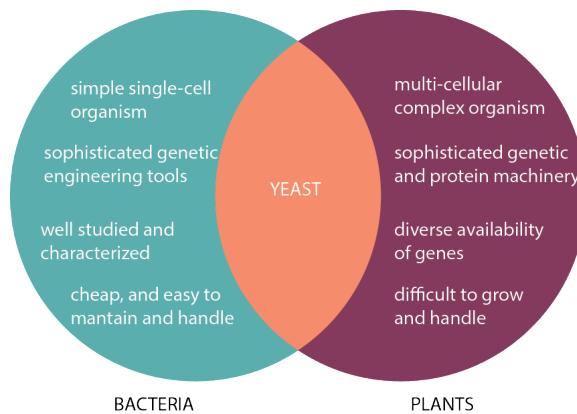


Figure 1.6 | Advantages and disadvantages of using bacteria, yeast, or plants as a model organism for waste remediation research.

So far the high-level organismal categories routinely used in biological studies are bacteria, yeast, and plants. Bacteria, specifically *Escherichia coli* have been the model organism for biological studies given their ease to grow, well understood cellular mechanisms, and sophisticated genetic and protein engineering technologies that were built on centuries of research. However, bacterial systems may be too simple of a starting point to build a complete remediation strategy. The advent of synthetic biology is bringing in a new wave of hybrid bacterial systems to counter this notion, but many of the technologies in genetic circuits and directed evolution are

³sample preparation and image taken with Zijay Tang from the Tim Lu Lab, MIT.

still maturing [blount2012construction]. Plants on the other hand, namely the model organism *Arabidopsis thaliana*, have already proven themselves to be useful players for waste water remediation [rascio2011heavy]. Yet unlike bacteria, plants are difficult to grow and maintain given their diverse growing conditions. Especially, plant genetic and protein engineering technologies have yet to reach the maturity of bacterial based systems primarily because of the complexity of plant cellular biology.

Given the almost two opposite ends of bacteria and plant biology, what is almost a goldilocks compromise between the two model organisms *Saccharomyces cerevisiae*—the common baker’s yeast (Figure ??). It can be argued that yeast were the first microorganism to be deliberately used by man, specifically for food and beer [greig2009, spencer1997]. Like bacteria, yeast have a mature biotechnology platform in which strains are routinely engineered, and not just for research purposes, but also for the consumer and pharmaceutical applications [nielsen2013, chae2001]. But unlike bacteria, yeast are eukaryotic organisms that contain sophisticated metabolic machinery and have a variety of protein homologues that are shared between both bacteria and plants. More so, the past centuries of yeast-based research has unveiled almost an encyclopedia knowledge of yeast genomic, proteomic, and metabolic profiles [guthrie2002guide]. Such a spectrum of information is still unavailable for plants, as plants are much more diverse with more complex biology. Therefore yeast seems to be the best compromise between the two platforms.

Aside from the technology, yeast also come with logistical advantages. Yeast have been used for almost 5000 years [greig2009], and now it has become a common house-hold item that touches all areas of the globe integrating into many national, ethnic, and cultural backgrounds. Yeast-based consumer goods are routinely FDA approved [nutrition2018, usda2014], and the global scaling and reduced cost of yeast has been pioneered by the food and beer industries.

Herein, this work shows that yeast can be a robust and tunable platform for next generation bioremediation technologies. Not only that, the technologies derived in this study are meant to translate into the yeast consumer market, where the same technologies used to make bread and beer can be leveraged to clean today’s rise in

waste waters.

Chapter 2

Using yeast to sustainably remediate and extract heavy metals from wastewaters

Abstract

Our demand for electronic goods and fossil fuels have challenged our ecosystem with contaminating amounts of heavy metals causing numerous water sources to become polluted. To counter heavy metal waste, industry relies on a family of physicochemical processes with hydroxide and sulfide chemical precipitation being one of the most commonly used. However, the disadvantages of chemical precipitation are vast, some of which are the generation of secondary waste, handling of volatile chemicals, and the need for dedicated infrastructures. To circumvent these limitations, biological processes have been sought after to naturally manage waste. Herein, this work shows that yeast can act as a biological alternative to traditional chemical precipitation by controlling naturally occurring production of hydrogen sulfide (H_2S). Sulfur production was harnessed by controlling the yeast sulfate assimilation pathway, where strategic knockouts and controlled culture conditions generate H_2S from 0 to over 1000 ppm. These sulfur producing yeast were able to remove mercury, lead, and copper from

real-world samples taken from the Athabasca Oil Sands. More so, surface display of biominerization peptides allowed for controlled size distribution and crystallinity of metal sulfide nanoparticles which can nucleate on the cell surface. Altogether, this yeast-based platform not only removes heavy metals but also offers a platform for metal re-extraction through precipitated metal sulfide nanoparticles.

2.1 Introduction

Growing consumption of electronic goods and raw materials have pushed mining and manufacturing practices to unprecedented levels that the United Nations Environment Programme (UNEP) declared a global waste challenge in 2015 in order to monitor waste risk and waste crimes[**rucevska2015**]. Because of the growing demand for electronic goods and raw materials such as metals and fuels, 41.8 million metric tonnes (46.1 million tons) of electronic waste (e-waste) was produced globally in 2014, and this amount has risen by 20–25% in 2018 [**rucevska2015, balde2017**]. For the United States, there are more than 13,000 reported active mining sites with an additional 500,000 that are abandoned yet still polluting 16,000 miles of streams [**centerfordiseasecontrolandprevention2018, degraff2007addressing**]. Metal contaminates are typically copper, lead, cadmium, mercury and sometimes zinc [**rucevska2015, rucevska2015**]. Despite these obvious waste sources, industry still continues to unsustainably mine for raw materials, especially given the growing demand and consumption of batteries and electronic devices [**kang2013**]. China alone produces and consumes one of the largest quantities of batteries in the world, and in 2013 generated 570 kilotons of battery waste with less than 2% being collected and recycled [**song2017**]. The repercussions of battery waste, especially with lithium-ion batteries, is leaching of toxic amounts of copper and lead [**kang2013**].

Unfortunately, the progress for sustainable remediation technologies, in particular heavy metal removal, is slow in comparison to the rise of e-waste and the pace of mining [**rucevska2015**]. So far, practical implementation of heavy metal remediation has relied on physicochemical processes, the most ubiquitous industrial method being

chemical precipitation via lime, hydroxide (e.g. NaOH), or sulfide (e.g. FeS or H₂S) chemicals [fu2011removal]. Sulfur is the more desirable precipitation reagent as it is more reactive and has a lower rate of leaching than hydroxide precipitates, but the counter is that sulfur storage and handling is dangerous and costly which makes lime and hydroxides the preferred choice despite being less effective [fu2011removal]. Overall, chemical precipitation is costly, requires dedicated infrastructure, involves handling dangerous compounds and reactive gases, and generates secondary waste in the form of sludge [fu2011removal, kurniawan2006, barakat2011new]. Furthermore, sludge is ineffectively eliminated through pyrolysis or physically transported to landfills for burial [kurniawan2006, kumargupta2012]. Because of the inefficient downstream recycling steps many of the precipitate compounds leach back into ground water and nearby water sources thereby perpetuating this cycle of inefficient cleaning. Thus, physicochemical treatment via chemical precipitation is not an amenable option for developing countries which typically face the biggest challenge for heavy metal removal, and chemical precipitation will most likely need to be replaced with more sustainable and cost-effective processes in the near future [kumargupta2012].

Contrast to physicochemical processes, scientist have discovered the benefits of using biological systems to remediate waste as a natural alternative to current methods. Bioremediation has gained traction for wastewater treatment due to its natural means to process waste in addition to its autonomous growth and reactions allowing humans to distance themselves from constant maintenance and direct waste intervention [gavrilescu2015, singh2015]. In addition, there is hope that with the growing toolkit of molecular biology and bioengineering technologies scientist can further augment biology's capability to manipulate and convert waste. Already, scientist have discovered naturally occurring microorganisms which have been observed to tolerate and accumulate toxic metals, for example metal reducing bacteria [gadd2004microbial, wiatrowski2006novel, silver2005, picard2018, gartman2017]. One particular family of interest are sulfate reducing bacteria (SRB) which use sulfates as their terminal electron acceptor that then create hydrogen sulfide (H₂S) as a by-product leading to precipitation of nearby metals. Connecting

the dots, it is easy to see that biology has already developed a mechanism for biotic chemical precipitation using H₂S producing SRBs. Interesting uses of these organisms have been the design of anaerobic bed or stirred tank reactors for precipitation of metal contaminated effluent. Examples of bioreactors built by Jong et al. and Kieu et al. showed precipitation of 70–99% of introduced metals (dependent on the metal identity) over 2–8 week periods [jong2003, kieu2011]. However, the limiting piece to this technology is the biology itself. SRBs are obligate anaerobes, require precise handling of culture conditions, and grow slowly. In addition, many SRBs are unable to process complex carbon sources and require additional anaerobic microflora to persist [neculita2007] creating a new layer of complexity when managing reactors. To circumvent the stringent culture conditions scientist have begun to extract and transfer their unique behavior into more tractable organisms such as *E. coli* by heterologously expressing enzymes and non-native metal reducing pathways. Examples are expressing mercuric reductases from *Thiobacillus ferrooxidans* into *E. coli* [shiratori1989] or using a combination of protein and metabolic engineering to endow *E. coli* with sulfur generating capabilities much like SRBs [wang2000]. However in either case, attempting to transfer a complex and somewhat unknown mechanism into *E. coli* seems ambitious, and possibly riddled with difficulty.

To avoid the technical hurdles of engineering SRBs or expressing foreign pathways in *E. coli*, a more tractable biological platform was used to develop a biotic method for heavy metal precipitation. More so, an organism that can be easily used by both scientist and non-scientists in addition to having an established presence in industrial and consumer settings was prioritized. Therefore, yeast was chosen. The common baker's yeast, *S. cerevisiae*, is widely used both in the scientific and consumer setting, and by using yeast advantages beyond the biotechnology such as infrastructure to scale, cost, packaging and transport, are already in place [vieira2013, barth-haasgroup.n.d., prnewswire.n.d.]. The goal of this work is to transform yeast into a bioremediation platform for heavy metal removal and tap into the available resources for translating yeast into a usable system for practical waste remediation and recycling for real-world settings. With yeast, rather than assembling complex metabolic circuits or inserting

foreign genes, this work uses its natural metabolic pathways to endogenously generate H₂S much like SRBs. However, unlike SRBs H₂S production can be controlled both in rate and quantity by modifying the sulfate assimilation pathway and tuning culture conditions. These added controls enable yeast to generate uniform metal sulfide precipitates with respect to size distribution and crystallinity, potentially improving downstream filtration and recycling processes. Overall these results show that yeast, an already environmentally friendly and sustainably grown organism conventionally used for food and drink, can also be used as an agent for heavy metal detoxification.

2.2 Results

2.2.1 Metabolic control of yeast H₂S production

The first goal for engineering yeast as an agent for heavy metal remediation was to find a biologically generated product that could be used to precipitate metals from solution. Fortunately, the wine-industry was key in identifying such a compound. Good wine makers know that over-fermenting yeast produces an off-tasting and often off-putting egg smell, and researchers attributed this phenomenon with the build-up of H₂S gas [swiegers2007]. Wine researchers identified that the yeast sulfate assimilation pathway driven under fermentation conditions drives the production of H₂S gas (Figure ??a) [swiegers2007, linderholm2008]. From there yeast wine-strains were engineered to suppress the production of H₂S gas for better quality wine. However, by performing the opposite modifications yeasts' natural sulfur production can be instead harnessed for heavy metal precipitation. During this investigation it was shown that single gene knockouts in the sulfate assimilation pathway promotes H₂S production in a controllable manner. Knockout strains that produced detectable amounts of H₂S were Δ MET2 (accession number #P08465), Δ MET6 (#P05694), Δ MET17 (#P06106), Δ HOM2 (#P13663), Δ HOM3 (#P10869), Δ SER33 (#P40510), and Δ CYS4 (#P32582) (Figure ??a). Knockouts Δ SER33 and Δ CYS4 became auxotrophic to cysteine while Δ HOM3, and Δ MET2 were slow growers on synthetically

defined (SD) media. Δ HOM2, Δ MET17, and Δ CYS4 were chosen as experimental strains for sulfur-induced chemical precipitation due to their consistently high levels of sulfur production and normal growth characteristics in complete synthetically defined media (CSM) compared to the other strains. From Δ HOM2, Δ MET17, and Δ CYS4 double deletions were performed to obtain Δ HOM2 and Δ MET17 (Δ HM217); combination knockouts with Δ CYS4 produced extremely slow growers.

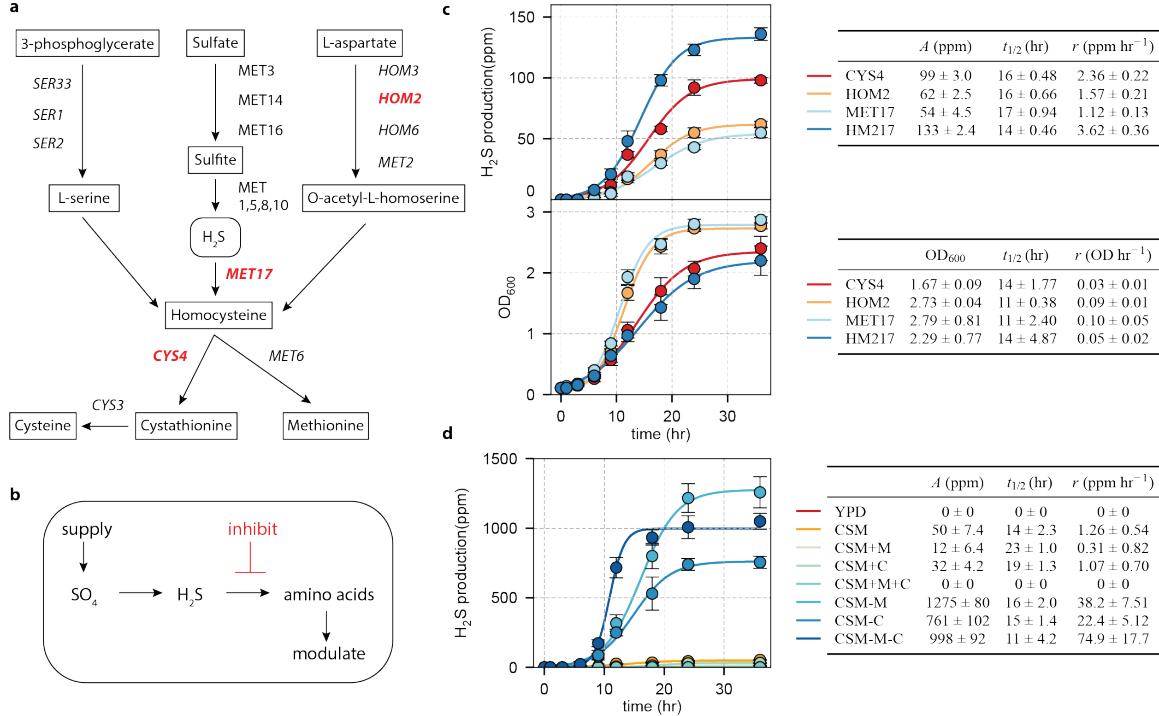


Figure 2.1 | Engineering the yeast sulfate assimilation pathway to generate hydrogen sulfide (H_2S). (a) Schematic adapted from Linderholm et al. [linderholm2008] Genes involved in the conversion of H_2S to amino acids were knocked out. Italicized knockouts were screened for H_2S production, while bolded red knockouts gave noticeable production of H_2S . (b) Deletant $\Delta CYS4$, $\Delta HOM2$, $\Delta MET17$, and $\Delta HM217$ produced sulfur closely following Le Chatelier’s principle. Supplying sulfate (reactants) while limiting nutrients such as cysteine and methionine (products) motivated the production of sulfur. (c) H_2S production (top curves) compared to growth curves in CSM cultures (bottom curves). Fitted parameters A represents the steady-state production of H_2S , $t_{1/2}$ represents the time at which H_2S production reached half-max, and r is the maximum rate of H_2S production. (d) H_2S production as a function of media composition for $\Delta MET17$ with fitted parameters A , $t_{1/2}$, and r . For all data, the mean \pm s.d. of three replicates from different days are shown. Curves were fitted and parametrized against the sigmoid function
$$\frac{A}{1 + \exp -k(t - t_o)}$$
.

Despite the metabolic complexities of the sulfate assimilation pathway, yeast H_2S production was observed to follow Le Chatelier’s Principle. Supplying the necessary nutrients such as nitrogen sources and sulfates, while limiting the amount of products, i.e. cysteine and methionine, stimulates the yeast sulfate assimilation pathway

to produce H₂S (Figure ??b). However, the normal conversion of sulfide to thiol containing biomolecules such as cysteine and methionine is prevented by removing pathway enzymes ΔCYS4, ΔHOM2, ΔMET17, thereby forcing expulsion of the intermediate H₂S. In CSM cultures, ΔCYS4, ΔHOM2, ΔMET17 and ΔHM217 produced 99 ± 3, 62 ± 3, 54 ± 5, and 133 ± 3 ppm of sulfur gas in a 50 mL cultures, respectively (Figure ??b; Supplemental Figure ??a). Sulfur production was optimized by altering the media composition, primarily by removing cysteine and methionine. For ΔMET17, sulfur production was tuned from a negligible amount to over 1000 ppm with maximum production rate of 75 ± 18 ppm hr⁻¹ in 50 mL CSM cultures lacking methionine (Figure ??c; Supplemental Figure ??b,c).

2.2.2 Chemical precipitation of heavy metals using sulfur producing yeast

To test for chemical precipitation, overnight cultures of ΔCYS4, ΔHOM2, ΔMET17 and ΔHM217 were diluted to mid-log in media containing 100 μM copper, zinc, cadmium, lead, or mercury and shaken for 12 hours. The amount of metal precipitated correlated to the mutants capacity to produce H₂S, with ΔHM217 and ΔCYS4 precipitating more than ΔHOM2 and ΔMET17 (Supplemental Figure ??a). Testing ΔMET17 specifically, the amount of metals removed measured using inductively coupled plasma (ICP) was dependent on the media's nutrient content. Cultures grown in YPD were the least effective, while cultures with CSM lacking methionine or cysteine had an overall precipitation efficiency >50% for copper, cadmium, lead, and mercury ($p < .05$), almost doubling the amount of precipitation compared to cultures in CSM supplemented with those amino acids ($p < .05$). The increase in metal precipitation correlates to the increased level of H₂S production in those media (Figure ??a; Supplemental Figure ??b).

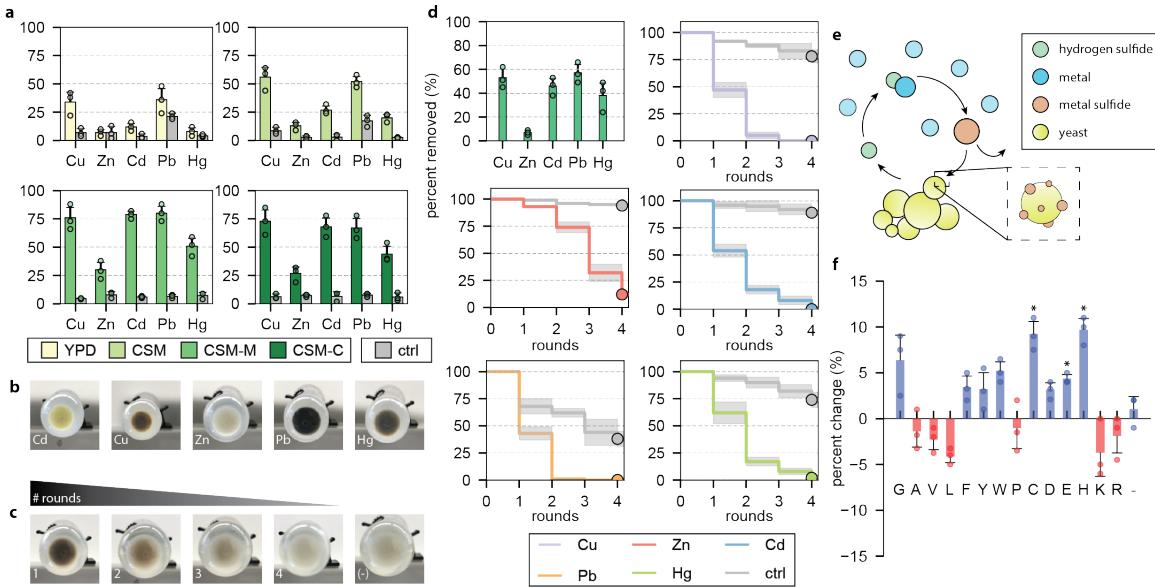


Figure 2.2 | Uptake of Cu, Zn, Cd, Pb and Hg with Δ MET17 sulfur-producing mutant. (a) Bar chart represents percent precipitation of metals under varying culture conditions. -M and -C indicate media without methionine or cysteine, respectively. The same experiment was performed with non-H₂S producing WT strain to test for non-specific metal removal (gray bars). (b) Visual representation of metal sulfide precipitation in cultures incubated with 100 μ M metals. (c) Δ MET17 with Cu, Zn, Cd, Pb and Hg all at 100 μ M were cultured together for multiple rounds of precipitation. Images represent the sequential precipitation of metals out of solution, with the darker precipitated color gradually diminishing with increased number of rounds. (-) represents a control yeast culture without any metals added. (d) Data representing images in (c). Top left plot represents the uptake from the first round. Remaining plots represent the gradual reduction of metal in solution after each round of precipitation. The same experiment was performed for a control WT strain (gray lines). (e) Illustration of the hypothesized reaction of metal sulfides on the surface of yeast. Metals could either precipitate in solution or on the yeast surface. (f) Bar chart represents the percent change in cadmium precipitation given expression of hexa-peptide repeats of the amino acids designated on the x-axis. For all data, the mean \pm s.d. of three replicates are shown. Asterisk above bar charts represent significant increase in metal precipitation compared to WT ($p < .05$).

Yeast culture density (OD₆₀₀) was tested to determine the optimal culture density at which the most amount of metals precipitated. At the extremes, low and high OD₆₀₀ precipitated very little. Cells at low OD₆₀₀ produce low amounts of H₂S per

volume while also struggling to grow against toxic amounts of metals. Whereas at high OD₆₀₀s cells were no longer rapidly growing and driving their sulfate assimilation pathway thereby producing little sulfur. OD₆₀₀s at mid-log had higher amounts of precipitation which correlated to the fast growth and sulfur production rates (Supplemental Figure ??c). This effect was most significant for Cd and Hg for cultures beyond 0.5 OD₆₀₀ ($p < .05$), possibly due to their higher toxicities.

To test for precipitation specificity, cells were incubated in media containing a mixture of copper, zinc, cadmium, lead, and mercury, each at 100 μM . The preference for precipitation was copper, lead, cadmium, mercury, and zinc, in that order, which loosely follows their trends in solubility products, and was observed in several other physicochemical precipitation experiments [**fu2011removal, neculita2007, rickard2006**] (Figure ??c,d). Sequential precipitation experiments were conducted to test the minimum number of iterations required to completely remove metals from solution, a practice normally implemented in industrial water processing via chemical precipitation [**kurniawan2006, kumargupta2012, bolong2009, johnson2005**]. Unprecipitated metals left in solution were mixed with fresh yeast for additional rounds of precipitation. Copper and lead removal below 1% (63 ppb and 207 ppb, respectively) required 2 rounds, 3 rounds for cadmium and mercury (112 ppb and 201 ppb, respectively), and 4 rounds to remove zinc below 20% (1.31 ppm) (Figure ??d). Given these results, yeast-based precipitation of heavy metals closely approached EPA standards for potable waters (i.e. tens to hundreds of ppb) [**usepaa, usepa**].

H_2S producing yeast are also tolerant to high levels of metal contamination, some as high as 100 μM levels of cadmium and lead. Cultures of ΔMET17 and WT diluted to 0.1 OD₆₀₀ were grown with 100 μM of Cu, Zn, Cd, Pb, or Hg. ΔMET17 showed robust growth curves compared to WT (Supplemental Figure ??a). In addition, cells that underwent metal precipitation were regrown without any significant changes in growth rate, and could be used for further precipitation experiments (Supplemental Figure ??a).

2.2.3 Yeast display of amino acids modulates metal sulfide precipitation

To increase the amount of metal precipitation, the yeast cell surface was modified to display amino acid repeats via yeast display. A lab-derived plasmid containing constitutive expression of the AGA1 and AGA2 yeast display constructs under the GAP promoter (vector called pYAGA) was used. AGA2 was fused with a hexa-amino acid repeat consisting of what was hypothesized as either positive (e.g. Asp, Glu, His, Cys, Ser), or negative (e.g. Val, Leu) effectors for metal precipitation. To test for expression the hexa-amino acid repeat was followed by a Myc tag which was stained for flow cytometry analysis (tag was not included in strains used for metal precipitation experiments). Yeast display expression levels were greater than 60% for hexa-repeats of Gly, Ala, Val, Leu, Phe, Tyr, Trp, Pro, Cys, Asp, Glu, His, Lys, Arg (Supplemental Figure ??a). Plasmids were then transformed into Δ MET17 and metal precipitation experiments were repeated in CSM. Amino acids containing Thiol and metal-binding moieties such as cysteine, glutamic acid, and histidine increased the precipitation of cadmium, zinc, and mercury by 5–10% ($p < .05$). Precipitation was negatively affected by more hydrophobic residues such as valine and leucine ($p < .05$) (Figure ??f; Supplemental Figure ??b). Precipitation of copper and lead were not as affected by the presence of these displayed peptides. A hypothesis is that the fast copper/lead sulfide reaction rates favor precipitation in solution rather than the diffusion-limited process of nucleating onto the cell surface.

2.2.4 Chemical precipitation of wastewater taken from the Canadian Oil Sands

To test the efficacy of this system in real-world wastewater, effluent from the Athabasca Oil Sands in Canada was received and subjected to yeast induced metal precipitation. The Athabasca Oil Sands is a well-known deposit of bitumen and crude oil, and for almost a hundred years this area had been a key resource for oils and fossil fuels which still drives the global economy today [sparks2003]. Due to this,

the area has been heavily mined and contaminated with human-driven excavations, drilling, and mining, leading to erosion, pollution, and ecological damage making the Athabasca Oil Sands an area in need of major remediation [kelly2010]. A sample of the effluent was obtained (roughly top one-fourth to one-half meter of the oil sands top layer; Figure ??a) and fractionated with gentle centrifugation to separate the liquid from the solid debris (Figure ??b). A one to one mixture of the liquid phase was mixed with Δ MET17 in 2X CSM-M (1X final) at 1 OD₆₀₀. Cultures were shaken overnight for 12 hours, spun down, and supernatant collected and measured for metal content.

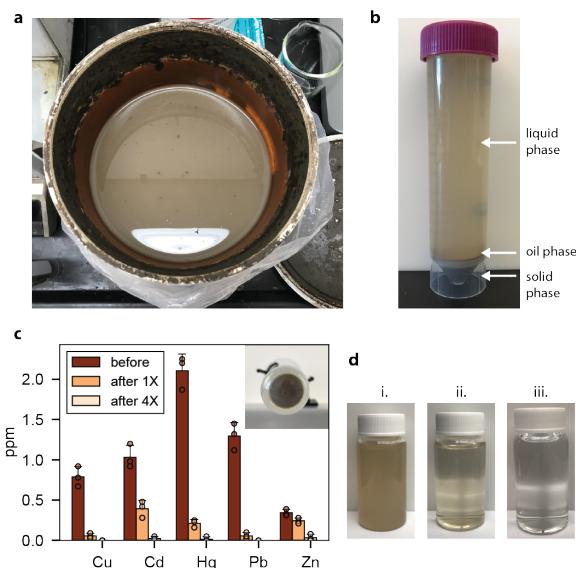


Figure 2.3 | Using yeast to remediate effluent from the Athabasca Oil Sands in Canada. (a) Isolated effluent taken from the Athabasca Oil Sands. (b) Effluent was centrifuged to separate the liquid, oil and solid phase. The liquid phase was taken to test for yeast induced metal precipitation. (c) 1:1 mixture of the effluent liquid phase to CSM-M culture with Δ MET17 were incubated overnight and measured for metal content. Afterwards, the supernatant was transferred to a fresh culture of Δ MET17 and experiment repeated up to 4 times, with each iteration measured for metal content using ICP. Top right inlet image shows pelleted cell culture with precipitated waste after 1 round. (d) Visual inspection of wastewater opacity before (i) and after (ii) one round of yeast induced metal precipitation. (iii) same sample after 4 rounds of yeast induced chemical precipitation. For all data, the mean \pm s.d. of three replicates are shown.

ICP analysis showed that the liquid phase from the Athabasca Oil Sands contained appreciable amounts of copper, cadmium, mercury, lead, and zinc, with the more toxic cadmium, mercury, and lead being more abundant per weight (around 1–2 ppm or mg/L) (Supplemental Figure ??). One round of yeast induced chemical precipitation showed greater than 85% removal of copper, mercury, and lead, and between 30–50% removal of cadmium and zinc (Figure ??c). These results were consistent with the metal removal experiments with metals spiked at 100 μ M in CSM (Figure ??a; 10–20 times more concentrated than what was measured in the oil sand's liquid phase). After 4 rounds, the amount of copper, cadmium, mercury, lead, and zinc levels closely approached undetectable amounts using ICP ($p < .05$) (Figure ??c). Examining the remediated effluent visually, the opacity of effluent dramatically reduced after just one round (Figure ??d; Supplemental Figure ??). Considering the stark difference in coloration, and observing that cadmium, lead, and mercury produce minimal water coloration, it was suspected that other materials in the liquid phase such as silicon-based compounds (rock/sand) contribute to the effluent's darken opacity (Supplemental Figure ??). Therefore, in addition to metal precipitation, other contaminating materials may non-specifically bind to the yeast surface and precipitate as a conglomerate. This shows another use-case for yeast as a platform for bioremediation due to their ability to act as a biosorbent, and several studies have attempted to deploy yeast as a natural biosorbent in contaminated areas [wang2006].

2.2.5 Controlled metal sulfide particle formation for downstream extraction and recycling

In addition to converting metals into non-soluble metal sulfides, the yeast cell wall can also act as a substrate for metal sulfide nucleation and aggregation. Given previous results showing that metal precipitation could be effected by knockouts, culture condition, and displayed peptides, it was hypothesized that these parameters could be tuned to influence precipitation rate and hence precipitate morphology and crystal

quality. After metal precipitation experiments with Δ MET17 and Cd, precipitated CdS were extracted by treating the cell wall with zymolyase and separating cellular debris from metal sulfide particles through liquid-liquid extraction. Using the same experiments yeast were fixed and sectioned to analyze the localization of metal precipitates. Metal sulfides and cell sections were analyzed with TEM and energy-dispersive x-ray microscopy (EDX) to quantitatively measure particle morphology, localization, and chemical composition.

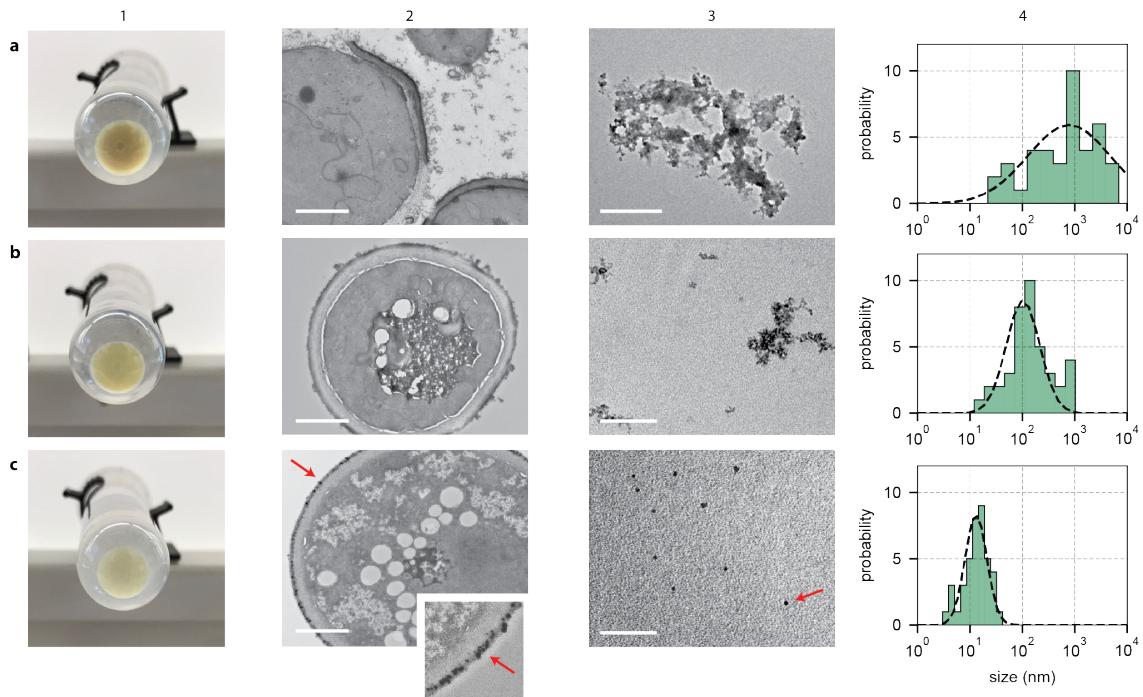


Figure 2.4 | Controlled size distribution of cadmium sulfide particles by engineering yeast strain and culture conditions. Columns are ordered as followed: image of metal precipitate (1), cell sectioning with metal precipitate (2), isolated metal precipitate (3), and counted size distribution of isolated metal precipitate (4). (a) Δ MET17 grown in CSM-M, (b) Δ MET17 grown in CSM-C, and (c) Δ MET17 grown in CSM. Size distribution data were determined by imaging 40 random locations on 3 samples of isolated metal precipitates using TEM. Sizes were measured using ImageJ. Column 2 scale bars represent 1 μ m. Column 3 scale bars represent 1 μ m for row A, and 100 nm for row B and C.

Extracted precipitates were first characterized for size distribution, morphology, and crystal structure without any displayed peptides. Cultures in CSM lacking both

methionine and cysteine with fast H₂S production rates above 50 ppm hr⁻¹ led to precipitates characterized by amorphous structures with average size exceeding 1 μm and size distribution spanning 2–3 orders of magnitude ($p < .05$) (Figure ??a). The precipitates were also shown to damage the cell wall, as TEM analysis of cell sections showed fragmented cell walls surrounded by large metal sulfide aggregates (Figure ??a). As H₂S production rates were slowed by supplementing cultures with methionine and cysteine, the average size of precipitates began to decrease while also uniformly nucleating onto the cell wall without causing visible cell damage as examined under TEM and EDX (Figure ??b; Supplemental Figure ??a). Cultures in fully supplemented CSM with low sulfur production rates below 10 ppm hr⁻¹ produced particles with controlled size distributions between 5–50 nm for CdS ($p < .05$) (Figure ??c). In addition, purified particles had 1:1 metal to sulfur stoichiometry (Supplemental Figure ??b).

These observations suggested that an increase in H₂S production rate correlate to an increased metal sulfide particle size distribution, and vice-versa. Therefore, a desired metal sulfide particle size distribution could be reverse engineered by tuning the rate of H₂S production by selecting the appropriate strain, culture density, and culture growth rate (Figure ??b,c; Figure ??c). A hypothesized mechanism for controlled metal sulfide particle size is that slower H₂S production rates allow time for metals to diffuse and nucleate on to the yeast cell surface. Given that the cell wall consists of negatively charged polysaccharides and proteins, a reasoning is that the electronegative environment allows for somewhat size-controlled nucleation.

Metal nucleation was further explored by displaying nucleating peptides to facilitate metal sulfide growth, a concept that has been successfully employed in other biological organisms such as viruses and bacteria [picard2018, wang2000, sakimoto2016, sweeney2004]. Without any displayed motifs, precipitated cadmium sulfide examined under high resolution TEM (HRTEM) produced large amorphous structures (Figure ??a). Crystalline structures indicated by lattice fringes were first observed with the hexa-cysteine motif, CCCCCC. More prominent lattice fringes were observed with GGCGGC and GCCGCC displayed peptides, glycine-

cysteine motifs generally conserved in metal-binding proteins such as metallothioneins [desilvataram.2002] (Figure ??a,b; higher resolution images can be found in Supplemental Figure ??). Slowing the rate of sulfur production below 10 ppm hr⁻¹ while displaying glycine-cysteine motifs generated cadmium sulfide quantum dot-like nanoparticles in the 50 nm range (Figure ??c,d). With more crystalline features these CdS particles gave a strong excitation peak at 330 nm and an emission peak at 480 nm (Figure ??e).

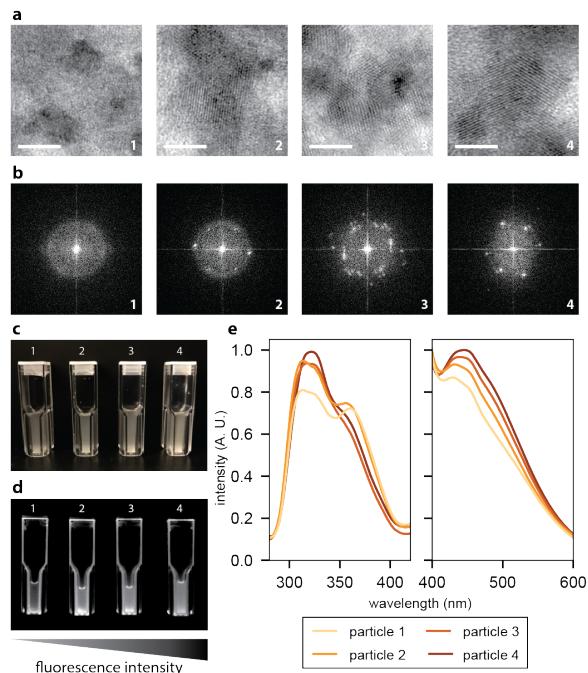


Figure 2.5 | Crystal quality and fluorescence of isolated precipitated CdS nanoparticles as a function of yeast displayed peptides. Numbers representing amino-acid sequence are: 1 = GGGGGG, 2 = CCCCCC, 3 = GGCGGC, 4 = GCGGCC. (a) Rows 1–4 show high resolution TEM images of precipitated CdS particles displaying various degrees of lattice fringes. Scale bars represent 5 nm. (b) Fourier transform of CdS particles showing various degrees of diffraction patterns caused by lattice fringes. (c) Image of isolated CdS particles suspended in water of samples 1 through 4 in ambient light. (d) Same image captured under UV excitation. (e) Excitation and emission spectra of samples 1 through 4. Excitation peak converged towards 330 nm and emission peak towards 450 nm with increasing crystallinity.

The creation of fluorescently active metal sulfide nanoparticle by-products encourages the idea that yeast can be used not only to remediate waste, but also to convert

waste into useful materials. Compared to industrial chemical precipitation, precipitates are generally amorphous, large, and chemically ill-defined making burial, or pyrolysis the only accessible means for removal [**kurniawan2006**, **kumargupta2012**, **bolong2009**]. However, yeast controlled metal sulfide precipitation offers a means to control the precipitate morphology and reduce the likelihood of sludge production. As well, the additional biological handles on strain and culture conditions allow for controlled precipitate size and crystallinity. Metal sulfide nanoparticles that nucleate on the cell wall lend themselves to direct metal re-extraction through cell wall removal, and subsequent metal particle isolation may allow for simpler downstream recycling processes and potential reuse [**kim2003**].

2.2.6 Feasibility in industrial settings

Yeast culture compositions are chemically defined and standard among scientists, with yeast being able to survive on several carbon sources at varying temperatures and at pH's as low as 3–4. In addition, yeast grow in defined culture environments in both aerobic and anaerobic conditions. These factors have made yeast one of the most understood and appreciated organisms not only to scientists, but for bakers, beer makers, and consumers [**barth-haasgroup.n.d.**, **prnewswire.n.d.**]. A typical laboratory only needs 3 dollars to produce 1 L of yeast with respects to the cost of consumables such as glucose, extracts, and buffers [**harrison2015**]. Industrially, the infrastructure to scale and bioreactor optimization done by both the beer and pharmaceutical industries have reduced the cost to 16 cents per liter [**vieira2013**, **harrison2015**, **hoek1998**]. These factors have allowed a global production of more than a million tons of yeast by weight in 2015 [**bccresearch**]. More so, packaging and delivery of yeast through freeze-dried and active-dried packets has allowed the yeast market to touch all areas of the globe, allowing both high-tech industries as well as rural villages the power to brew their own yeast [**prnewswire.n.d.**, **bccresearch**]. If the scale and breadth of the yeast market can be tapped for bioremediation purposes, specifically the precipitation and conversion of heavy metals using sulfur-producing yeast, then the potential impact on heavy metal waste management could be signifi-

cant and profound.

With respects to chemical precipitation, yeast are now vessels for environmentally responsible storage of H₂S where yeast can endogenously produce sulfur on demand rather than storing volatile sulfur compounds in pressurized tanks like in traditional chemical precipitation practices [**fu2011removal, kurniawan2006, kumargupta2012**]. Likewise, the production of sulfur can be turned on or off depending on the media composition and culture conditions, methods that are much simpler than directly handling or storing volatile chemicals. The scale of bio-precipitation of heavy metals is only limited to the accessibility and quantity of yeast that can be grown. Therefore, the millions of tons of yeast produced per year for food and drinks could be leveraged by treatment facilities as a robust and sustainably produced resource to more accessibly and economically treat waste water

2.3 Discussion

Future work will investigate more complex displayed biomineralization peptides in order to improve metal sulfide formation and capture. Further engineering biomineralization peptides could have two major applications: selective precipitation of metals and the creation of unique metal sulfide alloys that mimic doped metal sulfide compounds. Highly toxic elements such as cadmium and mercury in potable water should be removed preferentially to less toxic elements such as sodium or calcium. With engineered biomineralization peptides, it may be possible to selectively precipitate highly toxic metals such as mercury over precipitation of calcium even at disproportionate concentrations by using known heavy-metal binding motifs found in nature [**picard2018, sakimoto2016, sweeney2004, rugh1996**]. Another application is the ability to create useful metal sulfides in a ratiometric manner. Many metal sulfides used industrially are doped with other divalent metals to enhance their physico-chemical properties in semiconductors, solar cells, and magnetic materials [**arai2008, parveen2018, bhattacharya2007**]. Therefore, with further engineering it may be possible to use yeast as a substrate to facilitate ratiometric precipitation of multi-

metal sulfides, a concept that is especially interesting if the dopant metal is already present in the effluent.

Aside from remediation, the natural and autonomous production of sulfur is an attractive solution for curbing reliance for mined extraction of sulfur gas. Currently, sulfur is produced from petroleum, natural gas, and related fossil fuel activities with China, US and Canada being leading producers [**selim2013**, **usdepartmentoftheinterior2018**]. Sulfate however, the metabolic precursor to hydrogen sulfide in the yeast sulfate assimilation pathway [**swiegers2007**, **linderholm2008**], is generally more accessible through natural oxidation of ores, shales, and agricultural runoff [**little2000**], making sulfate more readily accessible than sulfur gas. Therefore, feeding yeast a low value resource such as sulfate, and generating a higher value product such as hydrogen sulfide is a tremendous benefit for industry. These engineered yeast provide a natural, low-cost H₂S source while also simplifying H₂S storage and transportation. Currently H₂S storage is hazardous and costly, but with a yeast-based production system storing H₂S is equivalent to storing yeast themselves.

In conclusion, this work uses yeast to generate H₂S to precipitate heavy metals from contaminated water. Furthermore production of H₂S can be tuned through gene knockouts and adjusting media conditions, thereby allowing control over the quantity of metal precipitation and precipitate size distribution. Crystallinity of metal sulfides can be controlled through displayed biomineralization peptides, and these particles are easily extracted by digesting the yeast cell wall for downstream recycling. This work ultimately shows that yeast could be a viable platform for heavy metal waste remediation and metal re-extraction, and invites the exploration of other yeast-facilitated bioremediation processes.

2.4 Materials and methods

Yeast strain and culture

Yeast strain W303 α was obtained from the Amon Lab at MIT. Synthetically defined dropout media (SD) was made by combining 1.7 g/L yeast nitrogen base without amino acid and ammonium sulfate (YNB) (Fischer), 5 g/L ammonium sulfate (Sigma), 1.85 g/L drop-out mix without methionine and cysteine (US Biological), and 20 g/L glucose (Sigma) and the addition of 10 mL/L 100X adenine hemisulfate stock (1 g/L) (Sigma). Complete synthetically defined media (CSM) was made by adding cysteine and methionine amino acids at a final concentration of 50 mg/L (Sigma). Both SD and CSM were pH'd to 7 with NaOH. Mixtures were stirred and filtered through a .22 μ m filter top (EMD). YPD media was made by adding 10 g/L yeast extract, 20 g/L peptone (Fisher), and 20 g/L glucose (Sigma) and filtered sterilized. Plates were made by adding 20 g/L BactoAgar (Fisher) and autoclaving.

Yeast strain and culture

The pRS303 and pRS305 vectors were used to clone the HIS and LEU markers for gene deletions in W303 α via homologous recombination. Single gene deletions of SER33, SER1, SER2, HOM2, HOM6, MET2, MET6, MET17, CYS3, and CYS4 were deleted by amplifying the LEU marker with 30 bp of the appropriate up and downstream overlaps to their respective gene target using PCR (Supplemental Table ??). Double mutants were created by amplifying the HIS marker with 30 bp of the appropriate overlap to the target gene using PCR and transformed into the single deletant strains. A constitutive yeast display vector constructed in the Belcher lab named pYAGA contains the AGA1 and AGA2 gene, both of which were downstream of a GAP promoter and upstream of a CYC1 terminator. Single stranded sequences coding for hexa-peptide repeats were ordered from IDT and annealed with sticky ends matching the BamHI and PmeI cloning sites of pYAGA (Supplemental Table ??). Hexa-peptide sequences were phosphorylated with T4 PNK prior to ligation using T4 ligase (NEB).

Circularized plasmids were transformed into chemically competent NEB α following the recommended NEB protocol and selected using ampicillin.

Yeast transformations were performed using Frozen-EZ Yeast Transformation Kit II (Zymo). For deletions, transformed cells were plated onto YPD for 1–2 days and replica plated onto drop out media (either HIS, LEU, or both) to select for positive transformants. Otherwise, plasmid transformations were grown directly onto plates with the appropriate drop-out media. Plasmid or genomic DNA was isolated by using silica bead beating and phenol/chloroform (Sigma) extraction. Sequences were confirmed by amplifying the DNA inserted sequence using PCR and sequencing the fragment using QuintaraBio sequencing services.

Screening and quantifying H₂S production

Cultures were initially screened in 5 mL CSM cultures in 14 mL BD culture tubes with taped lead acetate hydrogen sulfide indicator strips (VWR). Lead acetate strips turned brown to black in the presence of gaseous H₂S. Cultures were grown at 30°C over 1–2 days and H₂S was detected by lead acetate strip darkening. Quantitative sulfur detection was monitored using Draeger hydrogen sulfide detection columns (VWR). 50 mL cultures in 250 mL Erlenmeyer flasks were corked with a single-hole rubber stopper in which the hydrogen sulfide detection columns were fitted. Cultures were grown for 1–2 days and were visually inspected at specific time-points to measure sulfur production.

OD₆₀₀ Measurements

Optical density measurements at discrete time points were performed using 2 mL non-frosted cuvettes (VWR) and a table-top DU800 Beckman Coulter spectrophotometer measuring at 600 nm. Continuous growth curve studies were performed on a shaking 96 well BioTek Synergy 2 plate reader held at 30°C with 100 μ L cultures. Cultures were first diluted from overnights to < 0.1 OD₆₀₀ and aliquot into a 96-well round bottom plate (Cellstar) with the appropriate metal and concentration.

Metal precipitation experiments and quantification

Liquid stocks of copper (II) chloride, zinc chloride, cadmium nitrate, lead nitrate, and mercury (II) chloride (Sigma) were made at 100 mM in water. Metal precipitation studies were performed by diluting overnight cultures to varying culture densities in 5 mL of fresh culture containing 100 μ M of metal. Cultures were grown overnight, spun down at 900xg for 3 min in a swinging bucket rotor and the supernatant was collected for measurement of metal content. Metal content was measured on an Agilent ICP-AES 5100 following standard operating procedures. Trace concentrations of metal below 10 μ M were measured on an Agilent ICP-MS 7900. If samples were to be diluted, they were diluted in 3% HNO₃ (Sigma) to fit within the dynamic range of ICP detection. For all experiments, a sample with no cells and metals was measured to account for the background quantities of copper, zinc, cadmium, lead, and mercury in the media. All ICP measurements were subtracted by this blank sample. The ICP measurement of the supernatant was used to subtract the initial amount of metal added (i.e. 100 μ M) to calculate the amount of metal removed. These experiments were done for all strains in addition to a WT control to measure non-specific absorption onto the cell.

Multiple uptake experiments was performed by resuspending 1 OD₆₀₀ of fresh yeast culture with the equivalent volume of supernatant of a previous metal precipitation experiment. This process was repeated at most up to 4 times, with each iteration sampled for ICP measurement and metal removal calculation described previously.

Metal uptake study of effluent taken from Canada's Athabasca Oil Sands

Samples of effluent was taken from the Athabasca Oil Sands in Canada. Liquid was gently centrifuged at 1000xg for 30 minutes to fractionate the liquid, oil, and solid phase. The liquid phase was used as the waste medium to test for yeast-induced metal precipitation. Although not thoroughly investigated in this study, the oil phase contained many organics, aromatics, and oils contributed from mining runoff. The

solid phase contains a heterogeneous mixture of large debris, rocks, and precipitates that are easily spun down during centrifugation or through size-exclusion filtration.

To prepare for precipitation experiments, an overnight of Δ MET17 was grown in CSM-M and spun down. 1 OD₆₀₀ per mL of cells was added to a 1 to 1 mixture of 2X CSM-M (prepared by doubling all ingredients) and liquid phase extracted from the effluent. The mixture was incubated overnight for 12 hours, spun down, and visualized for precipitation. The supernatant was taken for ICP measurement for copper, cadmium, mercury, lead, and zinc following the procedures explained above.

The liquid phase metal profile content was studied using ICP. Commercial ICP multi-element standards was used to multiplex measurements in parallel (VWR or Agilent). Multiple dilutions of the liquid phase in 3% HNO₃ was performed (such as 1 to 1, 1 to 10, etc.) to determine the level of matrix effect, as the liquid phase contained other contaminants not accounted for in the standards and can skew readings. A 1 to 5 dilution gave consistent results and was used to calculate the concentrations of Na, Mg, K, Ca, Sr, Ba, Mn, Fe, Cu, Zn, Si, Cd, Pb, Hg, Cr, As.

Flow cytometry

Displayed peptides were first cloned with a C'-terminus V5 tag followed by a stop codon in a constitutive AGA1 and AGA2 vector which was called pYAGA. Cultures were grown to saturating OD and 0.5 OD₆₀₀ were taken for staining and flow cytometry. Cells were washed and pelleted at 900xg with PBS+1% BSA. Primary antibodies against V5 (Life Technologies) were diluted 1:500 in PBS+1% BSA and incubated at room temperature for 1 hour. Secondary antibodies with AlexaFluor 488 were diluted 1:2000 in PBS+1% BSA and incubated at room temperature for 1 hour. Cells were then washed and diluted to 1e6 cells per mL for FACS. FACS was performed on a BD FACS Canto or LSR II following standard operating procedure provided by the Koch Flow Cytometry Core.

Extraction of metal sulfide precipitates from yeast

Overnight cultures yeast and precipitated metals were pelleted at 900xg for 3 min. Cultures were washed and resuspended in 1 mL sorbitol citrate. 100T Zymolyase (Zymo) was diluted 1 to 100 and added to the suspension and incubated for >1 hour at 30°C while shaking. Digested cells were pelleted at 900xg for 3 min and supernatant was removed, or kept for later analysis of dislodged metal sulfide particles. Cells were resuspended with 1:1 water and oleic acid (organic layer; Sigma). Mixtures were spun down at 900xg for 3 min to pellet the cellular debris while allowing insoluble metal sulfide particles to remain in the organic layer. The organic layer was removed and fresh oleic acid was introduced to further extract the metal sulfide particles. This process was performed between 1–3 times until coloration was completely transferred into the oleic acid. Most organic solvents were observed to work (phenol:chloroform, hexane, octonal, etc), however oleic acid was more cost effective, easier to handle, and safer to use. Samples could be used immediately for analysis or concentrated by spinning down particles at max speed for 15 min and resuspended in a lower volume in either oleic acid or water.

Fluorometry

An Agilent Cary Eclipse Fluorescence Spectrophotometer was used to measure the fluorescence of the isolated metal sulfide particles using disposable PMMA acrylic cuvettes (VWR). Excitation and emission scans were performed following standard operating procedures provided by the Center of Material Science Engineering, MIT.

TEM sample prep

Cells were not digested with zymolayse in order to preserve the cell wall for imaging. Cell fixation, dehydration, embedding, and sectioning followed yeast OTO processing provided by the WhiteHead Institute, MIT [[seligman1966](#)]. The yeast cells were grown to an appropriate optical density and fixed with 2% glutaraldehyde, 3% paraformaldehyde, 5% sucrose in 0.1 M sodium cacodylate buffer (EMS) for 1 hour.

Pelleted cells were washed and stained for 30 minutes in 1% OsO₄, 1% potassium ferrocyanide, and 5 mM CaCl₂ in 0.1 M cacodylate buffer. Osmium staining was followed by washing and staining in 1% thiocarbohydrazide. Pellets were washed and stained again in the reduced osmium solution. The cells were then stained in 2% uranyl acetate (EMS) overnight, serially dehydrated with ethanol, and embedded in EMBED-812 (EMS). Sections were cut on a Leica EM UC7 ultra microtome with a Diatome diamond knife at a thickness setting of 50 nm, stained with 2% uranyl acetate, and lead citrate. The sections were examined using a FEI Tecnai Spirit at 80KV and photographed with an AMT CCD camera.

TEM and EDX analysis

TEM samples of purified metal sulfide particles were prepared on 400 mesh nickel Formvar grids (EMS) by dropping 10 μ L of sample onto the grids for 5 min and wicked dry. TEM images were acquired on a FEI Tecnai. Samples were also monitored by energy-dispersive x-ray (EDX) spectroscopy to qualitatively determine the relative amounts of sulfur and metal. When necessary, for example with copper, the signal background was corrected by subtracting the spectrum with a region without any metal sulfide particles to deconvolve overlapping peaks from the copper grid. High resolution TEM (HRTEM) images were acquired on a JOEL2010 to observe crystal spacing. A JOEL2010F was used for more resolved EDX elemental mapping of metal sulfide particles that nucleated on the cell wall.

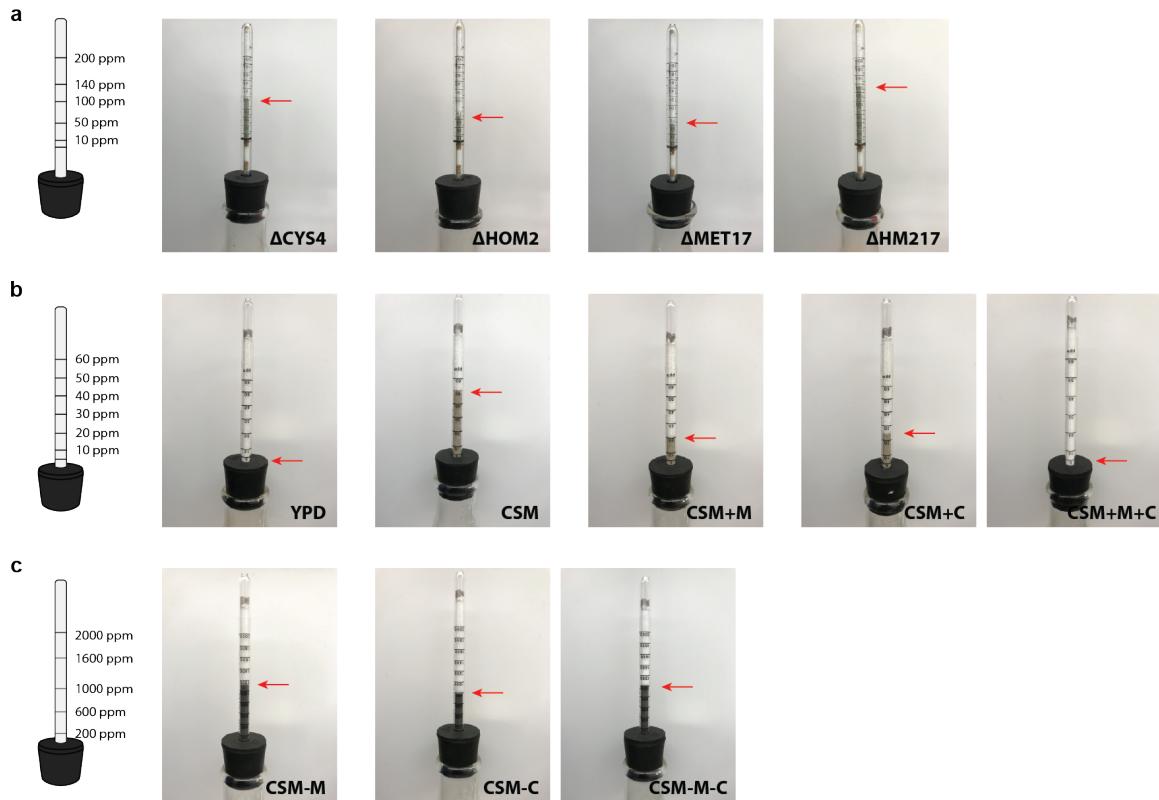
Mathematical analysis and plotting

Raw data was collected and stored as csv or Excel file formats. Data was imported and analyzed with Python using modules such as numpy, pandas, and scipy. Plots were graphed with matplotlib.

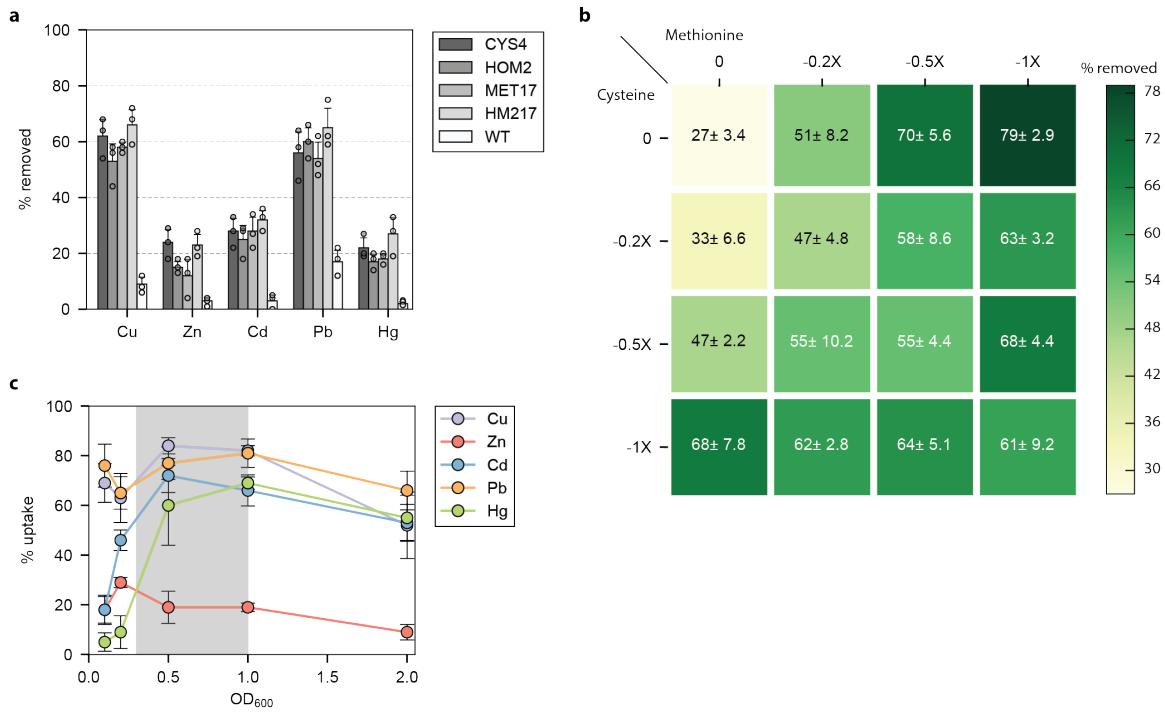
Statistical analysis

Statistical parameters including the the definition and values of n, SDs, and/or SEs are reported in the figures and corresponding figure legends. When reporting significance, a two-tailed unpaired t-test was performed between observations and p-values reported in the text. The significance threshold was set to $p < .05$ for all experiments, or as specified in the text.

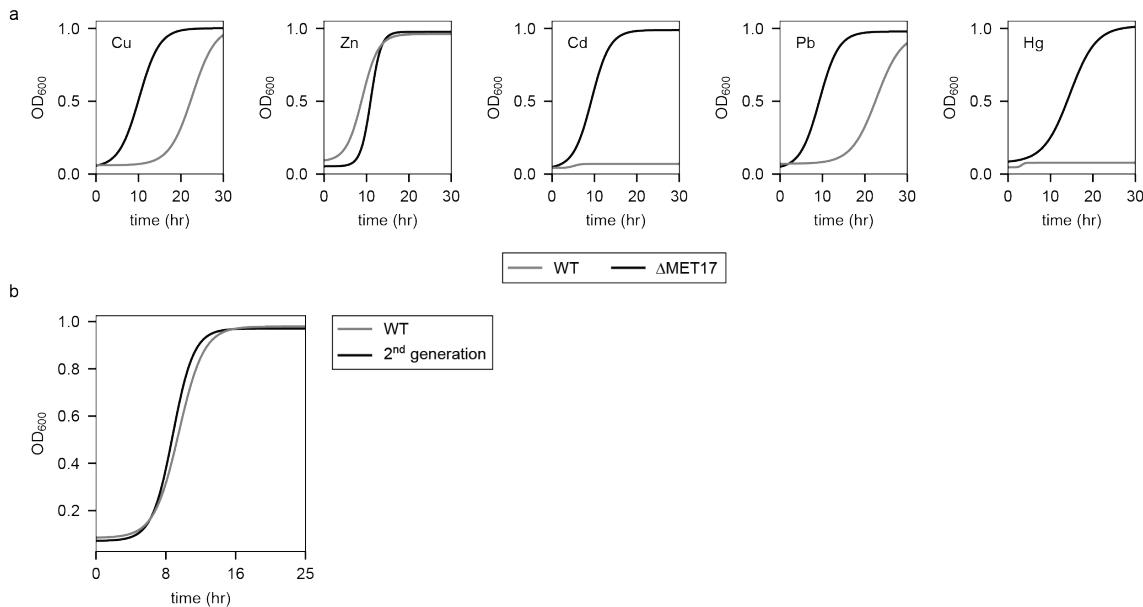
2.5 Supplemental figures



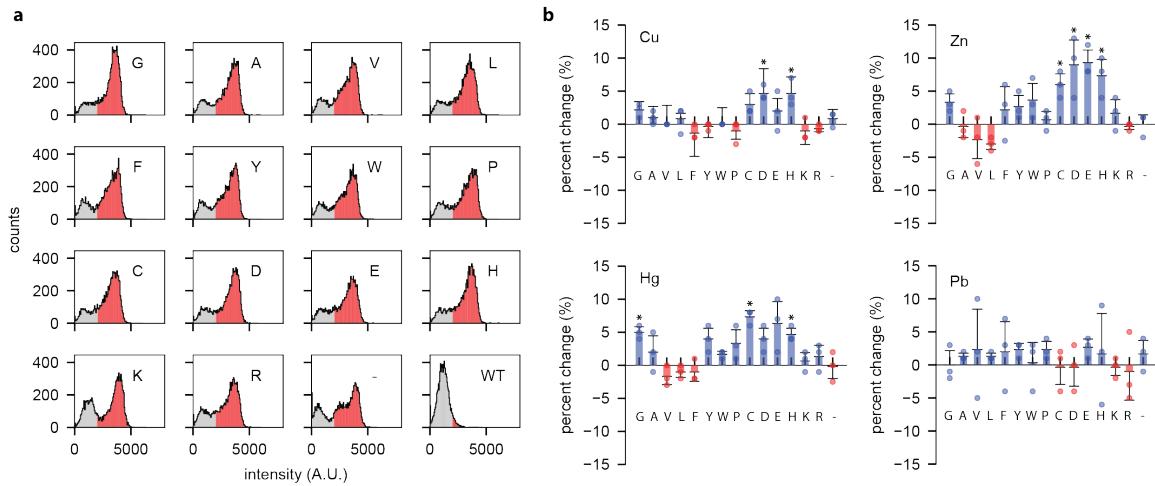
S. Figure 2.1 | Measurement of H₂S production from sulfur producing yeast cultures. Left illustrations represent H₂S detection columns with tick marks indicating the level of sulfur in ppm. (a) Sulfur detection using 200 ppm columns for mutants ΔCYS4 , ΔHOM2 , ΔMET17 , and ΔHM217 . (b) Sulfur detection using 60 ppm columns for ΔMET17 in cultures of YPD, CSM, and CSM with the addition (+) of methionine (M) and cysteine (C). (c) Sulfur detection using 2000 ppm columns for ΔMET17 in CSM cultures lacking (-) methionine or cysteine, or both.



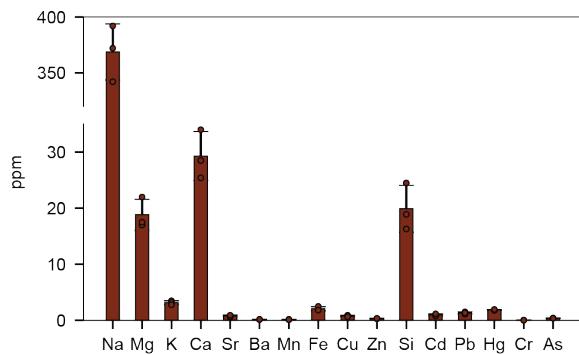
S.Figure 2.2 | Strain, culture density, and media composition effects on metal precipitation. (a) Precipitation of copper, zinc, cadmium, lead, and mercury with mutants Δ CYS4, Δ HOM2, Δ MET17, and Δ HM217, and WT as a control, in CSM. Cultures were grown at 30°C in 100 μ M metal as indicated. (b) Testing the effects of removing methionine (M) and/or cysteine (C) from CMS on precipitation efficacy with Δ MET17 and 100 μ M cadmium. Columns signify removal of M while rows signify removal of C from CSM. 1X stands for 100% removal (i.e. 0.2X = 20% and 0.5X = 50%). Annotated values per grid cell are the percent cadmium removed and standard error. (c) Optimal culture density (marked within grey bounds) was determined by titrating growing cultures of Δ MET17 at different OD₆₀₀ with copper, zinc, cadmium, lead, and mercury. Metal color coding matches those used in the main text. For all data, the mean \pm s.d. of three replicates were taken for each data point.



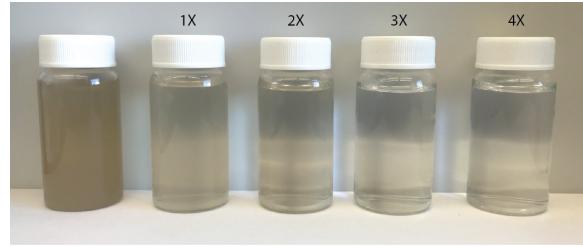
S. Figure 2.3 | Growth curves of ΔMET17 and WT in metal containing cultures. (a) Growth curves of ΔMET17 and WT grown in CSM with 100 μM metals specified. (b) ΔMET17 was used to precipitate 100 μM cadmium overnight. Afterwards, cells were diluted 1 to 100 and grown again (2nd generation) and compared to WT. All data points were measured in a BioTek Synergy 2 plate reader with 100 μL cultures shaken at 30°C. All curves were normalized to 1 within each experiment.



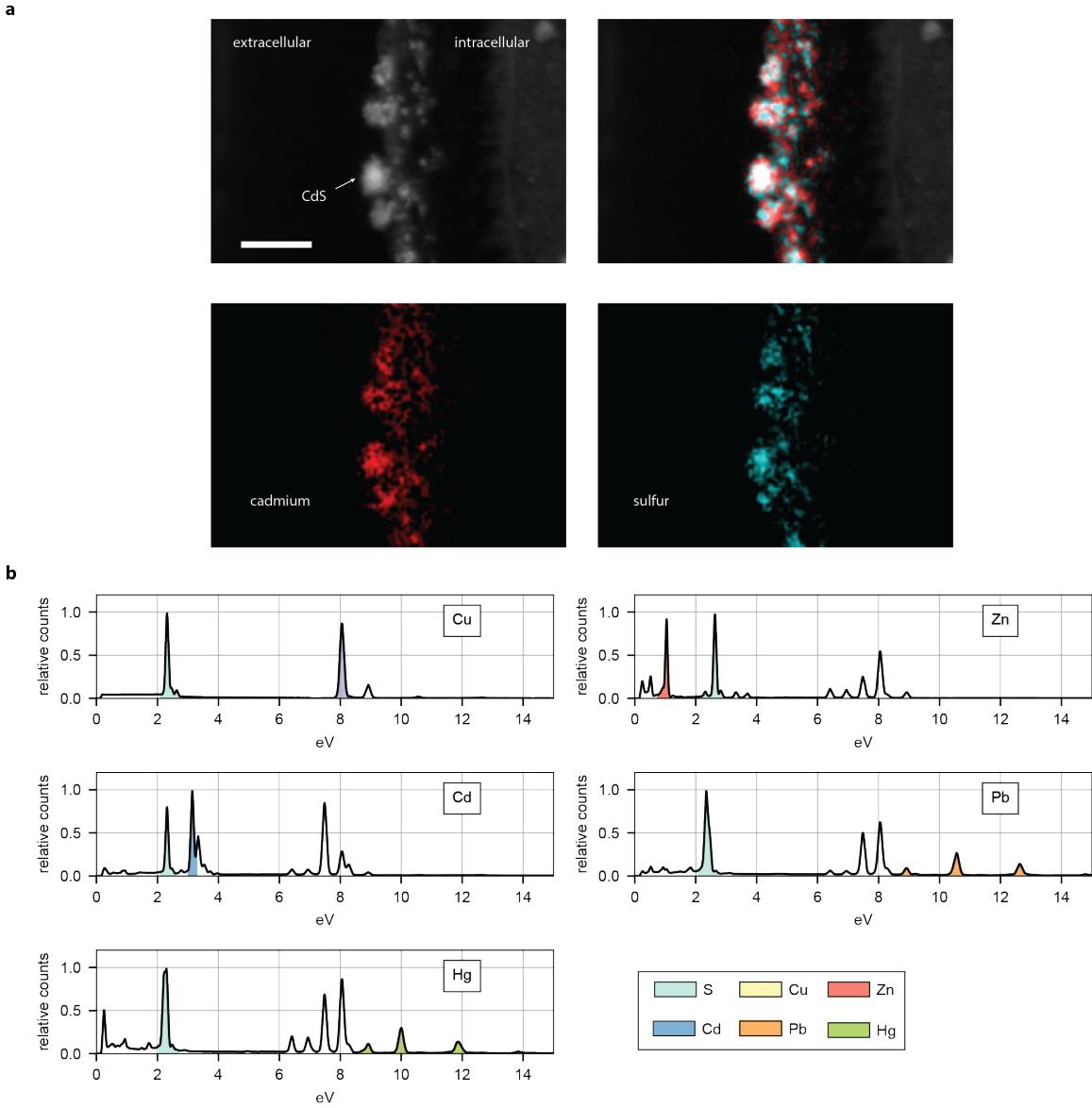
S.Figure 2.4 | Effects of yeast displayed amino acids on metal precipitation. (a) Flow cytometry data showing fluorescence intensity of labelled C'-terminus Myc tag on hexa-amino acid repeats. Expression was compared against an empty displaying pYAGA vector (NA) and non-displaying WT (-) for controls. Positive expression was cutoff at 2000 A.U. for segregating expressing versus non-expressing populations (grey and red, respectively). (b) Δ MET17 transformed with pYAGA with the hexa-amino acid motif (specified on the x-axis) was used to modulate precipitation of copper, zinc, lead, and mercury. Bars represent the percent change in metal precipitation compared to non-displaying Δ MET17 in CSM.



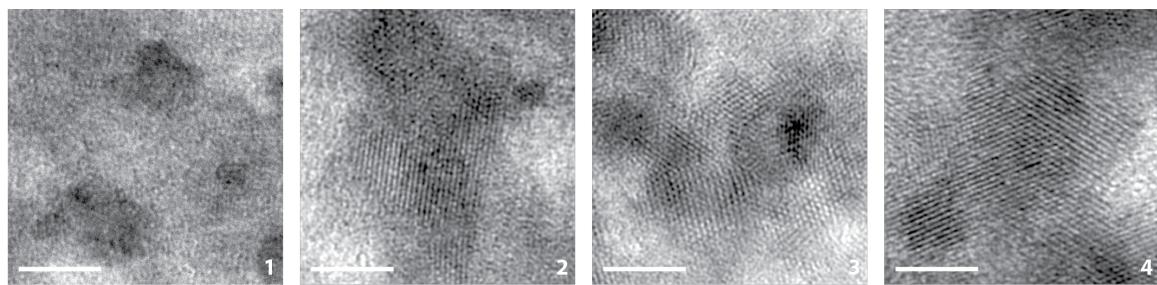
S.Figure 2.5 | Metal content profile of the liquid phase effluent taken from Canada's Athabasca Oil sands. Group I and II elements, such as Na, Mg, K, and Ca, in addition to silicon were strongly present. Heavier and toxic elements such as cadmium, mercury, lead, and arsenic were appreciably present at 1–3 orders of magnitude greater than EPA standards.



S.Figure 2.6 | Visual representation of remediated oil sands after multiple rounds with ΔMET17. Numbers at the top of each sample (1X, 2X, ...) indicate the number of rounds that were performed for removing contaminants from the liquid phase taken from the oil sands.



S. Figure 2.7 | Elemental mapping of precipitate metal sulfide particles. (a) Elemental mapping of HRTEM images of cadmium sulfide nanoparticles deposited on the cell wall of Δ MET17. Cadmium is false colored as red, sulfur as blue. Scale bar represents 50 nm. (b) Elemental dispersive X-ray (EDX) spectroscopy was performed on purified precipitated copper, cadmium, lead, mercury, and zinc sulfide particles under TEM. Elemental K α peaks are colored and highlighted as areas under the curve for qualitative comparisons. Metal color coding of spectral plots match those used in the main text.



S.Figure 2.8 | Enlarged HRTEM images from Figure ??a of lattice fringes of CdS particles precipitated on yeast displayed Δ MET17. 1 = GGGGGG, 2 = CCCCCC, 3 = GGCGGC, 4 = GCCGCC.

2.6 Supplemental tables

Name	direction	primer
pRS303 (HIS)	fwd	TATTACTCTTGGCCTCCTCT
	rev	CCTGATGCGGTATTTCTC
pRS305 (LEU)	fwd	AACTGTGGGAATACTCAG
	rev	GGTCAGGTCATTGAGTG

S.Table 2.1 | Primers used to amplify the HIS and LEU auxotrophic cassettes from the pRS303 and 305 vector. Primers that anneal and amplify the HIS and LEU markers on the pRS303 and pRS305 vectors, respectively. These primers were used as the base primer to then append overhangs for homologous recombination in W303 α strains to delete specific sulfate assimilation pathway genes. Overhangs and fully assembled primers are in Table ??

Name	direction	homology	primer(HIS)	primer(LEU)	primer(seq)
SER33	fwd	AACACTGA	AACACTGA	AACACTGA	TTCACGCT
		TTTCGGGT	TTTCGGGT	TTTCGGGT	GGAACACGA
		ATTCCTC	ATTCCTC	ATTCCTC	G
		CCTAAC	CCTAACTA	CCTAACAA	
			TTACTCTT	CTGTGGGA	
			GGCCTCCT	ATACTCAG	
			CT		
		CTATACAT	CTATACAT	CTATACAT	GAGACAGC
		ATATTTT	ATATTTT	ATATTTT	ACTTTGTG
		ATTATCT	ATTATCT	ATTATCT	GA
	rev	GAGTAA	GAGTAACC	GAGTAAGG	
			TGATGCGG	TCAGGTCA	
			TATTTCT	TTGAGTG	
			C		
		TCATCGAT	TCATCGAT	TCATCGAT	GACATTAA
SER1	fwd	TAACCATT	TAACCATT	TAACCATT	GGAGCCTT
		AGTGATAA	AGTGATAA	AGTGATAA	TGA
		GAAACA	GAAACATA	GAAACAAA	
			TTACTCTT	CTGTGGGA	
			GGCCTCCT	ATACTCAG	
			CT		
		GAACCAAA	GAACCAAA	GAACCAAA	CGTCTCAT
		TTACAGGC	TTACAGGC	TTACAGGC	ATCATATC
		ATATTCCG	ATATTCCG	ATATTCCG	GTATAATT
		CTGATA	CTGATACC	CTGATAGG	CG
	rev		TGATGCGG	TCAGGTCA	
			TATTTCT	TTGAGTG	
			C		

SER2	fwd	CATAGACA TTAAGCAC GACAGCTG TAAAAAA	CATAGACA TTAAGCAC GACAGCTG TAAAAATA	CATAGACA TTAAGCAC GACAGCTG TAAAAAAA	GACAACGT TTCTCGAA TCAG CTGTGGGA
					ATACTCAG
					CT
	rev	CTCTATT CATCTATC TATCATT TTTTCT	CTCTATT CATCTATC TATCATT TTTTCTCC	CTCTATT CATCTATC TATCATT TTTTCTGG	CAGCTGCT ATCAAGCA AG TGATGCGG
					TCAGGTCA
					TATTTTCT
					TTGAGTG
					C
MET1	fwd	TGATAAAAT AAACTAAG AAAATTTC AAAAGA	TGATAAAAT AAACTAAG AAAATTTC AAAAGATA	TGATAAAAT AAACTAAG AAAATTTC AAAAGAAA	GTAGGGCTT CATTAGA ATTGCT TTACTCTT
					CTGTGGGA
					GGCCTCCT
					ATACTCAG
					CT
	rev	TTTGAATG ATATCTTG TCTTTATA TACATA	TTTGAATG ATATCTTG TCTTTATA TACATACC	TTTGAATG ATATCTTG TCTTTATA TACATAGG	GACGACTT GGTTGAAG GA TGATGCGG
					TCAGGTCA
					TATTTTCT
					TTGAGTG
					C
MET2	fwd	AAAGAAAG AAAAAAAC GTAGTATA AAAGGA	AAAGAAAG AAAAAAAC GTAGTATA AAAGGATA	AAAGAAAG AAAAAAAC GTAGTATA AAAGGAAA	CTAGAAC GTCAAGTC TTCG TTACTCTT
					CTGTGGGA
					GGCCTCCT
					ATACTCAG
					CT

		rev	TTATGCCT	TTATGCCT	TTATGCCT	CTACTACT
			GAGGTATG	GAGGTATG	GAGGTATG	ACCGTTAG
			TGTGGTAT	TGTGGTAT	TGTGGTAT	TGTTAC
			CTATCC	CTATCCCC	CTATCCGG	
				TGATGCGG	TCAGGTCA	
				TATTTTCT	TTGAGTG	
				C		
MET5	fwd		GGGAACCA	GGGAACCA	GGGAACCA	CCATCGTC
			GAGAAAAA	GAGAAAAA	GAGAAAAA	TTGATATT
			CAAAAGAT	CAAAAGAT	CAAAAGAT	TATTCAGA
			TGGCGA	TGGCGATA	TGGCGAAA	TC
				TTACTCTT	CTGTGGGA	
				GGCCTCCT	ATACTCAG	
				CT		
		rev	TAGTATGT	TAGTATGT	TAGTATGT	CGAAGGAT
			CCTACTAT	CCTACTAT	CCTACTAT	AACAATAA
			GTCATATG	GTCATATG	GTCATATG	TGAGTCTT
			CTATCA	CTATCAC	CTATCAGG	AAC
				TGATGCGG	TCAGGTCA	
				TATTTTCT	TTGAGTG	
				C		
MET6	fwd		ACCAATAT	ACCAATAT	ACCAATAT	CCATCTTT
			AATTCAA	AATTCAA	AATTCAA	ACTAGCAT
			AGTACATA	AGTACATA	AGTACATA	TAGTTCT
			TCAAAA	TCAAAATA	TCAAAAAA	C
				TTACTCTT	CTGTGGGA	
				GGCCTCCT	ATACTCAG	
				CT		
		rev	ATATCATT	ATATCATT	ATATCATT	GCAAGAGT
			ACTTGCT	ACTTGCT	ACTTGCT	TATGGCTT
			TCCTTTT	TCCTTTT	TCCTTTT	TGT
			AAAACC	AAAACCCC	AAAACCGG	
				TGATGCGG	TCAGGTCA	
				TATTTTCT	TTGAGTG	
				C		

MET8	fwd	AAAATAAG AGAGTGTA TAATAGGA TAAAAAA	AAAATAAG AGAGTGTA TAATAGGA TAAAAATA	AAAATAAG AGAGTGTA TAATAGGA TAAAAAAA	GCTACAAA GTCCGATG AC
				TTACTCTT GGCCTCCT	CTGTGGGA ATACTCAG
			CT		
	rev	CGCGCCCC TTAAAAGA GGAGGCC TGTCGC	CGCGCCCC TTAAAAGA GGAGGCC TGTCGCC	CGCGCCCC TTAAAAGA GGAGGCC TGTCGCG	GTTGATCT GAACAGGC ATTG
				TGATGCGG TATTCT	TCAGGTCA TTGAGTG
			C		
MET10	fwd	TTCCTCGA GGTCACCC AAATATAC AACGAG	TTCCTCGA GGTCACCC AAATATAC AACGAGTA	TTCCTCGA GGTCACCC AAATATAC AACGAGAA	GTTCTCGA GACCACCA TC
				TTACTCTT GGCCTCCT	CTGTGGGA ATACTCAG
			CT		
	rev	TAGATATT TAGTTTT ATTACTAT ATTAAT	TAGATATT TAGTTTT ATTACTAT ATTAATCC	TAGATATT TAGTTTT ATTACTAT ATTAATGG	GCAGCCAA TAGAAAAG CTTG
				TGATGCGG TATTCT	TCAGGTCA TTGAGTG
			C		
MET17	fwd	AGATACAA TTCTATTA CCCCCATC CATACA	AGATACAA TTCTATTA CCCCCATC CATACATA	AGATACAA TTCTATTA CCCCCATC CATACAAA	GGTTAAGT AAAGCGTC TGTTAG
				TTACTCTT GGCCTCCT	CTGTGGGA ATACTCAG
			CT		

		rev	ATACATAA	ATACATAA	ATACATAA	GTTCAAAG
			TTTTACAA	TTTTACAA	TTTTACAA	TACGAGTC
			CTCATTAC	CTCATTAC	CTCATTAC	ACG
			GCACAC	GCACACCC	GCACACGG	
				TGATGCGG	TCAGGTCA	
				TATTTTCT	TTGAGTG	
				C		
HOM2	fwd		AATTAAAT	AATTAAAT	AATTAAAT	GTCAAGCA
			TGTAGAAA	TGTAGAAA	TGTAGAAA	TTGATTGA
			TAAAGCGT	TAAAGCGT	TAAAGCGT	CTCA
			TCTAAA	TCTAAATA	TCTAAAAA	
				TTACTCTT	CTGTGGGA	
				GGCCTCCT	ATACTCAG	
				CT		
		rev	AAGATGAA	AAGATGAA	AAGATGAA	GAAATGAG
			GACATAAC	GACATAAC	GACATAAC	TACCAACA
			TTTGCAAT	TTTGCAAT	TTTGCAAT	GTGCT
			TTTTCC	TTTCCCC	TTTCCGG	
				TGATGCGG	TCAGGTCA	
				TATTTTCT	TTGAGTG	
				C		
HOM3	fwd		ACAGAACG	ACAGAACG	ACAGAACG	CATTGAAG
			TTTCATTT	TTTCATTT	TTTCATTT	GATATTG
			TTTTAAC	TTTTAAC	TTTTAAC	TGTAGC
			TTTTAC	TTTTACTA	TTTTACAA	
				TTACTCTT	CTGTGGGA	
				GGCCTCCT	ATACTCAG	
				CT		
		rev	CTATCATT	CTATCATT	CTATCATT	CAGTATAA
			AAAGTGAA	AAAGTGAA	AAAGTGAA	CCCTGACA
			GAAGAAAG	GAAGAAAG	GAAGAAAG	TTACAT
			GTGGAT	GTGGATCC	GTGGATGG	
				TGATGCGG	TCAGGTCA	
				TATTTTCT	TTGAGTG	
				C		

HOM6	fwd	TAGTATCA	TAGTATCA	TAGTATCA	GGTTAGCG
		TCAATCGA	TCAATCGA	TCAATCGA	ATAGACAA
		ATAATAAA	ATAATAAA	ATAATAAA	TTTGTG
		AAAAAAA	AAAAAATA	AAAAAAA	
			TTACTCTT	CTGTGGGA	
			GGCCTCCT	ATACTCAG	
			CT		
	rev	ACCTATGT	ACCTATGT	ACCTATGT	AGATTGTA
		TTTATAT	TTTATAT	TTTATAT	GAAGATTG
		GTCTGTT	GTCTGTT	GTCTGTT	AGTAGC
CYS3	fwd	ATATACAC	ATATACAC	ATATACAC	CATTACG
		ACAAGACA	ACAAGACA	ACAAGACA	TGATCTCA
		AAACCAAA	AAACCAAA	AAACCAAA	GC
		AAAAAT	AAAATTA	AAAATAA	
			TTACTCTT	CTGTGGGA	
			GGCCTCCT	ATACTCAG	
			CT		
	rev	CGGTCGAA	CGGTCGAA	CGGTCGAA	GAGCGTTA
		GGCAGAGA	GGCAGAGA	GGCAGAGA	CTTCCAAA
		CGTGGCAC	CGTGGCAC	CGTGGCAC	TCG
CYS4	fwd	TGGCGA	TGGCGACC	TGGCGAGG	
			TGATGCGG	TCAGGTCA	
			TATTTCT	TTGAGTG	
			C		

rev	TTTGCTTT	TTTGCTTT	TTTGCTTT	CTGATGTG
	TATTTGAA	TATTTGAA	TATTTGAA	ATGCATGC
	GCGTGGGT	GCGTGGGT	GCGTGGGT	AT
	TCTTAT	TCTTATCC	TCTTATGG	
		TGATGC GG	TCAGGTCA	
		TATTTTCT	TTGAGTG	
	C			

S.Table 2.2 | Primers used to amplify the HIS and LEU markers for specific gene deletions in the sulfate assimilation pathway. Table shows gene name, 30bp homology region for that gene, and the HIS or LEU primers used to homogously recombine and replace (delete) the respective gene. The last column (seq) represents the primer used to confirm deletants by amplifying the deleted region from isolated genomic DNA using PCR and confirming using Sanger sequencing.

Name	AA	primer
mut1	<i>GCCGCC</i>	GGTTGTTGTGGATGCTGT ACAGCATCCACAACAAACC
mut2	<i>GCCGCC</i>	TGTGGCGGCTGTGGCGGC GCCGCCACAGCCGCCACA
mut3	<i>GCCGCC</i>	GGATGCGGAGGCTGCGGA TCCGCAGCCTCCGCATCC
mut4	<i>GCCGCC</i>	GGTGGTTGCGGTGGGTGC GCACCCACCGCAACCACC

S.Table 2.3 | Amino acid and DNA sequences used for cloning glycine-cysteine motifs for yeast display Columns AA, fwd, and rev stand for the hexa-amino acid sequence, forward (5'-3') oligo, and reverse (3'-5') oligo, respectively. Reverse oligos were followed with a TAG stop codon. Forward and reverse oligos were appended with BamHI and PmeI sticky ends, respectively. Oligos were ordered from IDT and annealed to form duplex DNA strands which were then ligated into pYAGA digested with BamHI and PmeI.

Chapter 3

Designing yeast as plant-like hyperaccumulators for heavy metals

Abstract

Hyperaccumulators are plants that absorb and tolerate elevated amounts of heavy metals by using an assortment of metal trafficking and chelating pathways. These metal trafficking mechanisms make plants useful for bioremediation applications; however, compared to bacteria-based bioremediation systems, plant growing conditions are long and difficult to maintain hindering adoption of plant-based bioremediation solutions. In addition, current genetic engineering tools are challenging to employ in plants and lack the sophistication of bacteria or yeast-based cloning. Herein, this work combines the robust growth and engineerability of bacteria with the unique waste management mechanisms of plants by using a more tractable platform—the common baker’s yeast—to create plant-like hyperaccumulators. Through overexpression of metal transporters and engineering metal trafficking pathways, we created yeast strains capable of sequestering metals at concentrations 10–100 times more than established hyperaccumulator thresholds for chromium, arsenic, and cadmium. To achieve these results, a density-based assay was developed to efficiently screen mutants for metal specificity and hyperaccumulating ability based on metal mass accumulation. Using this method the yeast transporter SMF1 was converted to a

cadmium mutant, and separately into a strontium transporter for applications in toxic waste and radioactive Sr⁹⁰ remediation. Overall, this work shows a systematic approach for transforming yeast into metal hyperaccumulators that are as effective as their plant counterparts.

3.1 Introduction

As the world becomes increasingly industrialized, heavy metal contamination is a growing environmental concern. Mining, manufacturing, and electronic goods disposal are the main sources of heavy metal waste; the United States alone adds 289 million tons of waste per year to the growing 850 and more landfills [**powell2016estimates**]. To illustrate the impact of waste generation this work specifically looked at two significant, yet often overlooked, contributors of heavy metal waste which are the textile industry and pollution from nuclear power plants and past fallout. Textile manufacturing employs a variety of heavy metal related processes, in particular dyeing, with many of the 100,00 types of dyes containing metal chelated centers for coloration [**robinson2001**]. Particular regions, such as India and Bangladesh where textile manufacturing is a dominant industrial practice, see high levels of cadmium, chromium, and lead in soils reaching 10–100 times higher than WHO established safety limits [**manzoor2006**]. Other metals such as cobalt, copper, zinc, and nickel are also pervasive and are incorporated at different levels in the textile process [**robinson2001**, **manzoor2006**]. The result, leachate that contains an indiscriminate mixture of metals which is difficult to separate, therefore leaving burial or transport as the only viable waste management option. On the same vein, the problem of nuclear waste and past nuclear fallout, with past catastrophic events of Chernobyl and Fukushima, have refocused attention on radioactive metal contamination, specifically radioactive strontium (Sr⁹⁰), which is of particular interest for its biological implications in bone integration and cancer [**sahoo2016**, **shutov1993**, **mangano2011**]. However, given the molecular similarity of calcium and strontium, and the relative abundance of calcium over strontium, removing just Sr⁹⁰ is chal-

lenging without being overwhelmed by other species. Both waste scenarios expose a unique challenge, how to selectively capture and discriminate metals from one another. Removal of toxic elements such as cadmium and mercury should be prioritized, even if at lower concentrations than more abundant and less harmful elements such as calcium and magnesium. This is particularly true for radioactive elements such as Sr⁹⁰, where strontium is typically masked by large amounts of similar divalent metals like calcium. Therefore, it is extremely difficult to remove low abundance heavy metals without being saturated by background species. Current industrial approaches such as absorption and ion-exchange are not particularly effective for such precise removal of toxic yet low concentration heavy metals as these processes are first saturated by more abundant background metals [fu2011removal, kurniawan2006, kumargupta2012].

Bioremediation strategies have the potential to address the challenges of heavy metal contamination. A promising subset of bioremediation is phytoremediation, the use of plants to sequester pollutants from soils and water [marques2009remediation, singh2015]. Plants have developed mechanisms to uptake heavy metals without suffering major toxic effects, and their abundant and renewable biomass contribute to significant bioaccumulation of toxins from soils and waters [marques2009remediation, singh2015]. Out of all plants, there are more than 400 species that hyperaccumulate heavy metals; the stricter definition being an accumulation of 100 mg/kg (0.01% dry wt.) of cadmium or arsenic, 1,000 mg/kg (0.1% dry wt.) of cobalt, copper, chromium, aluminum, nickel, and lead, and 10,000 mg/kg (1% dry wt.) of manganese, iron, and zinc [prasad2003, kramer2010metal, branquinho2007revisiting]. Not all hyperaccumulators have equal metal preferences. Even in a single family such as Brassicaceae, out of the 87 species 67 are nickel hyperaccumulators, 15 are zinc, and 5 can do both [prasad2003]. Insights to the mechanisms of hyperaccumulation have been attributed to hyperactive metal transporters and a variety of detoxification pathways which include glutathione synthesis and metal compartmentalization in vacuoles and other organelles [clemens2002, hall2002].

What limits wide-spread adoption of plant-based remediation solutions is the complexity to engineer and maintain strains. Plants are complex organisms, with different species requiring strict growing conditions where hyperaccumulators found in one location may not necessarily thrive in others due to surrounding biotic and abiotic factors. More so, current phytoremediation technology takes weeks to years to see signs of remediation, and in this current global waste crisis may be too long of a time scale [**eapen2006**, **yang2004**, **sarma2011**]. There have been attempts to create transgenic plants which incorporate genes from hyperaccumulators which grow faster and are more resistant to environmental factors [**eapen2005**]. However, because plants are multi-cellular with more complex gene clusters, the current state of genetic tools have yet to realize the sophistication and ease of genetic engineering for their single-celled counterparts such as bacteria and yeast [**eapen2005**]. Therefore, design of faster and easier waste management technologies needs to be developed on other platforms that are scalable and cost-effective. Single-cell organisms such as bacteria offer ease and scalability; however, they lack many hyperaccumulating features such as hyperactive metal transporters and useful organelles such as a vacuole. A biological platform at the intersection of these two is the common baker's yeast, *S. cerevisiae*. Current genetic engineering technologies have made it possible to engineer yeast on all levels, from specific proteins to complex metabolic pathways. More so, the infrastructure and ability to scale and distribute yeast are already in place thanks to the beer and pharmaceutical industries [**barth-haasgroup.n.d.**, **bccresearch**, **prnewswire.n.d.**]. The results herein show that taking concepts from plant hyperaccumulators and engineering them into yeast generate similar hyperaccumulating capabilities that are equal or better than their plant-based counterparts. This work demonstrates that yeast are a promising platform for targeted heavy metal remediation.

3.2 Results

3.2.1 Use of membrane transporters for enhanced and selective metal uptake

Several fundamental metal trafficking components are essential for enhanced metal uptake in hyperaccumulating plants, including cell membrane transporters, organelles storage systems, and chelator molecules [clemens2002, hall2002] (Figure ??a). Metal ions enter via highly active membrane transporters, and are compartmentalized into organelles such as the vacuole. To limit cellular toxicity, chelators such as glutathione, metallothionein, and phytochelatins bind and remove metals from sensitive metabolic functions [hall2002]. To mimic these plant hyperaccumulating features, the first step was to identify and express a hyperactive membrane transporter. A set of membrane metal transporters for zinc, copper, iron, and manganese [kramer2010metal, clemens2002, andre1995, luk2003, anthonyvanho2002] were overexpressed in yeast. For this study, native yeast metal transporters ZRT1 (accession number #P32804), ZRT2 (#Q12436), CTR1 (#P49573), CTR3 (#Q06686), FTR1 (#P40088), FET4 (#P40988), SMF1 (#P38925), and SMF2 (#P38778) (ZRT3, CTR2 and SMF3 are vacuole transporters, while FET3 is an oxidoreductase) were cloned and overexpressed using a GAL1 promoter on a 2μ plasmid. When overexpressed, some of the transporters, along with several more described below, did not show uniform expression but instead had punctate patterns when examined under fluorescence microscopy (Supplemental Figure ??a). This may suggest that over-expression led to poor localization, and this factor was considered when selecting a transporter candidate for future engineering.

To measure metal uptake, cells were incubated in 100 μ M metal for 4 hours. Supernatant was collected and measured for remaining metal content using inductive coupled plasma (ICP), and this value was used to calculate the amount of metal removed by the cells. Parallel to each experiment a sample of wild-type (WT) and a sample with no cells were measured to account for non-specific metal binding onto

the cell wall or sample tube. Taking these controls into consideration differential and enhanced uptake of zinc, copper, iron, and manganese was observed across the transporters (Figure ??b). ZRT1,2 and CTR1,3 were highly selective for zinc and copper respectively, increasing metal uptake by 10-fold compared to wild-type (WT) ($p < .05$). FET4 and SMF1 were less metal-specific and increased metal uptake by 3-5 fold across all four metals ($p < .05$; except for FET4 uptake of Zn compared to WT).

A similar study was performed for arsenic and chromium. These metals are typically found in oxy-polyatomic states such as arsenate and chromate. To achieve arsenate and chromate hyperaccumulation a different set of transporters are needed, as arsenate and chromate's -2 charge render them incompatible with the divalent metal transporters used previously. Given the molecular and steric similarity between phosphate (PO_4^{3-}) and arsenate (AsO_4^{2-}), and sulfate (SO_4^{2-}) and chromate (CrO_4^{2-}), a hypothesis was that the overexpression of sulfate and phosphate permeases will allow passage of arsenate and chromate [catarecha2007, pereira2008]. Overexpressing the phosphate permeases Pho84 (#P25297), 87 (#P25360), and 89 (#P38361), and Sul1 (#P38359) and Sul2 (#Q12325) showed increased metal uptake of arsenate and chromate, respectively (Figure ??c; Supplemental Figure ??a). Overall the Pho genes increased arsenate uptake by more than 3-5 fold ($p < .05$), and Sul genes increased chromate uptake by more than 5-fold ($p < .05$). These observations also align with plant hyperaccumulation of arsenate and chromate which show that they trespass into the cell via the phosphate and sulfate assimilation pathways [zhao2009, shanker2005].

Another common group of metal contaminants are trivalent metal ions such as aluminum and rare-earth metals (lanthanides and actinides). The most obvious approach is to use a trivalent metal transport for aluminum and rare-earth metal uptake; however, none exist in yeast or generally, at all. But research in a rice specie, *Oryza sativa*, uncovered a Nramp-like transporter known as Nrat1 (#Q6ZG85) which showed selective uptake of aluminum but not divalent metals [xia2010]. Cloning and heterologously expressing Nrat1 in yeast does indeed promote selective uptake of aluminum

with more than a 5-fold increase in aluminum uptake than compared to WT ($p < .05$) (Figure ??d), and no significant uptake for common divalent metals such as Cu, Zn, Fe, and Mn ($p > .05$) (Supplemental Figure ??). These results align with previous experimental results [xia2010]. The use of Nrat1 for rare-earth metal uptake such as neodymium and ytterbium, precious metals used in magnets and electronics, were tested but gave unreliable results as they precipitated in culture before measurements could be performed. However, the preferential accumulation of aluminum using Nrat1 support the hypothesis that other trivalent metals such as lanthanides and actinides can be potentially hyperaccumulated.

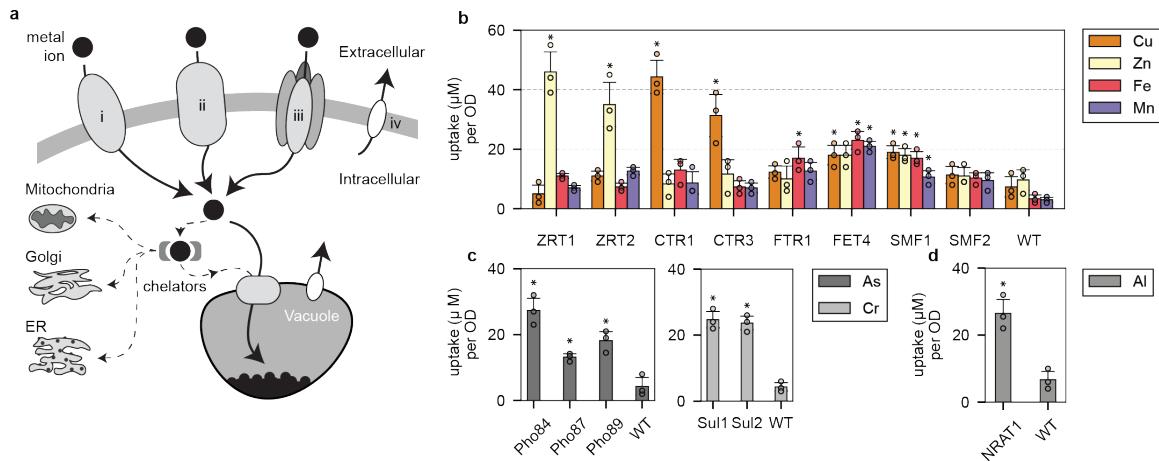


Figure 3.1 | A variety of metal transporters can be used to selectively internalize heavy metals into yeast. (a) A simplified schematic of metal transport in a eukaryotic cell. Membrane transporters can be diivalent metal transporters (i), permeases (ii), metal transporters that are modified or found to have auxillary metal transport function (iii), or antiporters which are used to remove excess metals out of the cell (iv). (b) Bar coloring indicates metal measured, with over-expressed transporter labeled on the x-axis. Values are reported in μM of metal uptake normalized per yeast culture density ($\mu\text{M}/\text{OD}$). Yeast metal transporters for zinc (ZRTs), copper (CTRs), iron (FTRs and FETs), and manganese (SMFs) were overexpressed and studied for metal hyperaccumulation. A WT strain was also tested in parallel for each metal to measure non-specific metal uptake. (c) The same study was performed for phosphate and sulfate permeases (PHOs, and SULs) which showed transport of arsenate and chromate, respectively. (d) The Nrat1 transporter, previously shown to uptake trivalent metals in certain strains of rice, was expressed and showed aluminum(III) transport. Asterisk above bar charts represent significance increase in uptake compared to WT ($p < .05$) for strains mentioned in the text. For all data, the mean \pm s.d. of three replicates are shown.

Simply by overexpressing metal transporters yeast have reached or exceeded hyperaccumulating thresholds established for plants. To compare results with established values, the amount of metal uptake was converted to milligram of metal per gram of culture dry weight (gDW) (Supplemental Figure ??). Given these results, overexpression of CTR1,3 reached 5.43 ± 0.7 and 3.84 ± 0.9 mg/gDW for copper, and overexpression of FTR1 and FET4 reached 1.83 ± 0.4 and 2.48 ± 0.3 mg/gDW for iron, respectively (Supplemental Table ??). All phosphate (Pho84, 87, 89) and sulfate

(Sul1, 2) permeases accumulated beyond the 1 mg/gDW threshold for arsenate and chromate hyperaccumulation. Nrat1 reached 1.33 ± 0.2 mg/gDW of aluminum which is above the 1 mg/gDW threshold [branquinho2007revisiting]. Overall, these results show that hyperaccumulation is not a plant-specific trait but a generalizable feature that can be engineered in yeast by selecting and expressing the appropriate metal transporters.

3.2.2 Increased expression levels of SMF1 enhance metal uptake

SMF1 from the Nramp family was selected for further optimization and engineering because of its broad metal specificity (Figure ??b) and the existing body of research on the Nramp family and its homologous [bozzi2016, bozzi2016a, ehrnstorfer2017, courville2006]. Another selection criteria was SMF1's relatively consistent membrane-localized expression as observed under fluorescent microscopy. When compared to FTR1 for example, expression was punctate and non-uniform, suggesting poor localization due to over-expression (Supplemental Figure ??a). Therefore the inability to over-express and localize transporters eliminated candidates for further engineering. SMF1 was also favored above the other transporters because of its promiscuous activity with several metals such as manganese, iron, nickel, and cobalt [anthonyvanho2002, courville2006, liu1997]. Thus SMF1 was a more appealing candidate to engineer for selective heavy metal uptake rather than converting a highly specific metal transporter which may be less malleable to change. More so, past work by Bozzi et al. and Ehrnstorfer et al. have elucidated crystal structures of multiple Nramps and have shed light on their structure-to-function relationship with respect to metal uptake [bozzi2016a, ehrnstorfer2014]. These insights were leveraged to semi-rationally alter the metal preference of SMF1, which is shown in later results.

Enhancing metal uptake using SMF1 require increasing its expression life-cycle by increasing protein yield and stability. SMF1 (denoted as S), like most nutrient trans-

porters, is tightly regulated to control the flux of metals into the cell, while limiting excess uptake to protect against toxicity. SMF1 expression, for example, is controlled by manganese ions and is post-translationally down-regulated by ubiquitination and endocytosis [nikko2008]. To create a hyperaccumulating yeast these controls need to be removed so that the transporter can be highly expressed without degradation. Therefore, mutations on SMF1's ubiquitination site K33,34 were altered to arginine (mutant denoted as S*) which helped reduce protein degradation [nikko2008]. In addition, BSD2 ubiquitin ligase (#P38356), which post-transcriptionally tags SMF1 for degradation, was deleted to further enhance SMF1 expression levels (deletion strain denoted as B) [liu1997, liu1999]. Finally, SMF1* was integrated (denoted as iS*) under a GAL promoter in BSD2 knockout strains. Expression was measured using both fluorescence microscopy (Supplemental Figure ??), and quantified using flow cytometry by fluorescently labeling a V5 tag fused to the C'-terminus of the SMF1 variants. Populations of fluorescently labelled SMF1 were analyzed to measure the percent of positively expressing cells, and the mean fluorescent intensity was used to qualitatively correlate the expression level between cells to their metal uptake levels (Figure ??a). Changes from S → S* → S*B → iS*B corresponded to increasing uptake of manganese and cadmium which resulted in uptake levels saturating to $85 \pm 6.7 \mu\text{M}$ ($9.0 \pm 0.7 \text{ mg/gDW}$) for manganese and $22 \pm 6.0 \mu\text{M}$ ($6.4 \pm 1.3 \text{ mg/gDW}$) for cadmium given the presence of $100 \mu\text{M}$ manganese or cadmium in culture (Figure ??a).

3.2.3 Use of vacuole transporters increase metal uptake yield

Metal uptake capacity can be further enhanced by expressing vacuole transporters to compartmentalized metals internalized by SMF1. Native yeast vacuole transporters [luk2003, anthonyvanho2002] tested were CCC1 (#P47818), COT1 (#P32798), ZRC1 (#P20107), and SMF3 (#Q12078) which were individually expressed in S*B strains (Supplemental Figure ??b; Supplemental Figure ??). All tested vacuole transporters showed elevated metal uptake for copper, zinc, iron, and manganese, with CCC1 and COT1 being the most significant across all metals ($p <$

.05) (Figure ??b). These results support the role that the vacuole broadly compartmentalizes metals from the cytosol. However, without the expression of SMF1, sole expression of vacuole transporters CCC1, COT1, ZRC1, and SMF3 in WT strains had negligible impact on copper, zinc, iron, and manganese uptake ($p > .05$) (Supplemental Figure ??). This result suggests that the largest barrier to metal uptake is from the membrane transporter, in this case SMF1, which is responsible for initial metal internalization. It is only after metal enters a cell that the vacuole transporters are rendered useful.

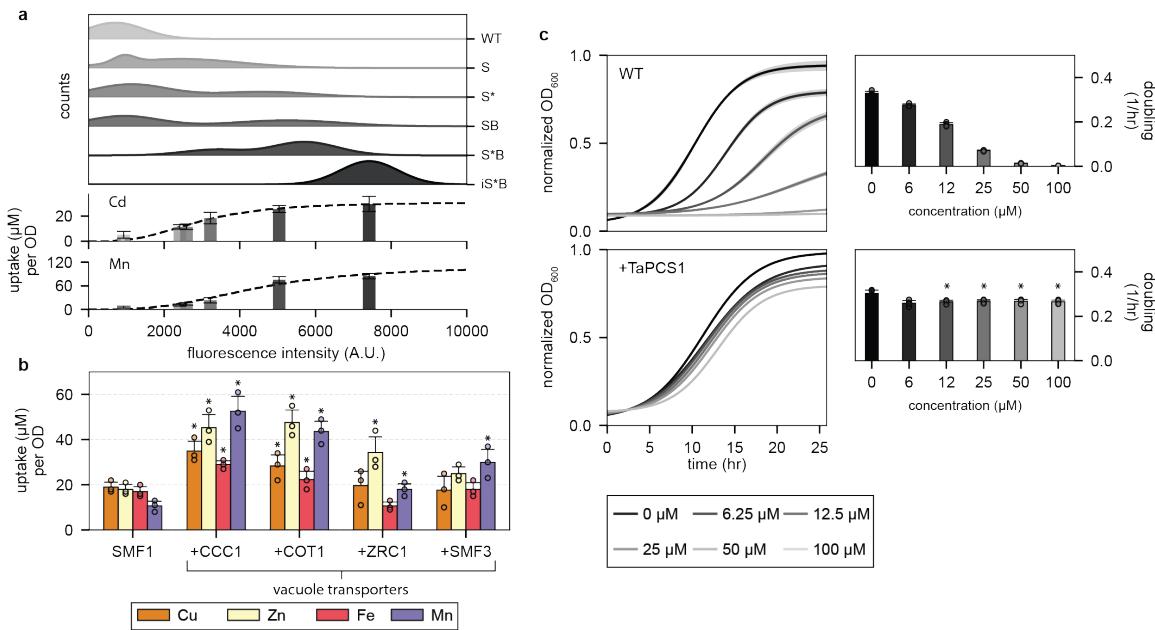


Figure 3.2 | Modifications in the metal trafficking pathway in yeast show enhanced metal uptake and tolerance for cadmium. (a) Top subpanel shows the population distribution of SMF1 variants measured with fluorescently labelled V5-tag using flow cytometry. The weighted average of the fluorescent intensity corresponds to the placement of the lower subpanel bar charts which represent the level of metal uptake for that strain. Increasing expression levels of SMF1 correlate to increased metal uptake of cadmium or manganese; however, up to a certain point indicated by the plateau in uptake. (b) Expressing vacuole transporters CCC1, COT1, ZRC1, and SMF3 in addition to SMF1 enhanced metal uptake. Asterisk above bar charts represent significant increase in uptake compared to SMF1 ($p < .05$). (c) Constitutively expressing wheat phytochelatin synthase, TaPCS1, conferred heavy metal tolerance against cadmium. Asterisk above bar charts represent significant changes in growth rates compared to WT ($p < .01$). For all data, the mean \pm s.d. of three replicates are shown.

3.2.4 Enhanced metal tolerance using phytochelatin synthase, TaPCS1

The purpose for creating a metal hyperaccumulator becomes counterproductive if the cell dies and releases the internalized metals back into the media. Therefore, mechanisms for metal detoxification and tolerance are needed to increase cell viability, and in theory, give cells more time to endure and uptake metals. One of the main mechanisms found in plants for metal detoxification is the production of phytochelatins, oligomers of glutathione (GSH) with cysteine and carboxyl rich moieties that chelate metals such as copper and cadmium [kramer2010metal, clemens2002, eapen2005]. Yeast are able to produce glutathione via the GSH pathway, which naturally protects yeast from accumulation of toxic metals. However, there does not exist a phytochelatin synthase for robust metal detoxification like that in plants. Instead, yeast rely on GSH or cysteine-rich and low molecular weight CUP1 metallothionein to chelate metals. However, past work has shown that metal detoxification is effective only at high copy numbers of CUP1 [jeyaprakash1991]. suggesting that protein production versus chemical synthesis of metal chelating compounds is less effective possibly due to a slower rate of protein synthesis and/or abundance. Therefore to create yeast tolerant to heavy metal environments would require a similar phytochelatin synthase mechanism. Past studies in plant hyperaccumulators have shown that a phytochelatin synthase, TaPCS1 (#Q9SWW5), from wheat improved heavy metal tolerance in both plants and yeast [clemens1999].

Integrating TaPCS1 under constitutive expression using a GAP promoter showed cadmium tolerance beyond 100 μM , whereas WT growth rates are significantly hampered below 10 μM ($p < .01$) (Figure ??c), results which support past observations [clemens1999]. TaPCS1 also improved copper, manganese, zinc, and cobalt tolerance by 2–10 fold than compared to WT (Supplemental Figure ??). The purpose of this work was to demonstrate that TaPCS1, in conjunction with membrane and vacuole transporters, can enhance metal uptake and retention due to enhanced metal tolerance. The subsequent results which combine SMF1, CCC1, and TaPCS1 show

that these modules can act additively to incrementally improve metal hyperaccumulation.

3.2.5 Creating a manganese and cadmium hyperaccumulator

To mimic the characteristics of a plant hyperaccumulator, the final yeast-based system combined expression of the membrane transporter SMF1 (S, or K33,34R mutant S*), vacuole transporter CCC1 (C), metal detoxifying phytochelatin synthase TaPCS1 (T), and deletion of ubiquitin ligase BSD2 (B). All parts were integrated into the genome except for S* which was introduced on a 2μ plasmid under a GAL1 promotor. As each component was added to the system the amount of cadmium uptake increased incrementally. The effect of adding all components together (S*BCT) enhanced cadmium uptake by almost 16-fold than compared to WT ($p < .01$) (Figure ??a). In addition, the rate of uptake increased dramatically with the combination S*BC reaching steady-states within 2–4 hours compared to 10–12 hours for strains lacking an overexpressed vacuole transporter (Figure ??b). The rate of uptake increased by almost 30-fold for S*BCT compared to WT ($p < .01$). Adding T to S*B or S*BC does not significantly enhance metal uptake but instead stabilizes metal internalization (Figure ??a, b). After 12 hours of growth in media containing 100 μ M cadmium, strains without TaPCS1 began to “leak” back out cadmium, possibly due to cell death or activation of divalent antiporters. In terms of viability, during active metal uptake in 100 μ M cadmium, the expression of C slightly improves cell viability, whereas combined expression of C and T fully rescue yeast survival ($p < .01$) (Figure ??c; Supplemental Figure ??).

SMF1 and CCC1 have broad metal specificity primarily for row one transition metals, thereby out-competing the uptake of cadmium if other transition metals such as manganese are present. To analyze the degree of manganese interference against cadmium, S*BCT was titrated at varying concentrations of cadmium with and without a constant background of 100 μ M. Metal uptake values were normalized to percent uptake with respects to the original metal concentration added, and the concentration at which metal uptake was half was termed K_U . The K_U for cadmium with and with-

out the presence of 100 μM went from $127 \pm 12 \mu\text{M}$ to $21 \pm 3.7 \mu\text{M}$ ($p < .01$). The K_U for manganese is almost 8 times higher at $945 \pm 84 \mu\text{M}$ ($p < .01$) (Figure ??d). Therefore, the main mechanism of transport for SMF1 prefers manganese and the uptake of cadmium is inferred to be due to transport “leakiness”.

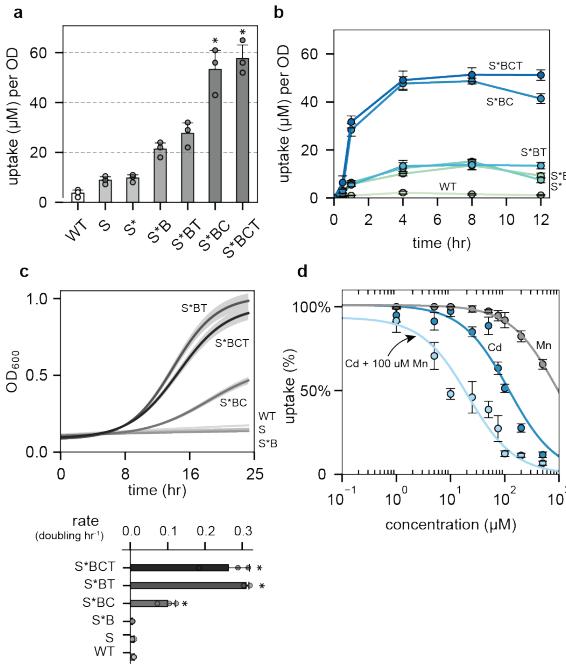


Figure 3.3 | Combining membrane transporter SMF1 and vacuole transporter CCC1 with TaPCS1 improved uptake capacity and metal tolerance.

(a) SMF1 (S) and its modifications (S^* and $\Delta\text{BSD}2$ as B) along with vacuole transporter CCC1 (C) and metal resistance enzyme TaPCS1 (T) incrementally enhanced cadmium uptake. Asterisk above bar charts represent significant increase in cadmium uptake when compared to WT ($p < .01$). (b) Combinations of S^* , B, C, and T showed changes in uptake rate, capacity, and metal retention over 12 hours of metal incubation. (c) In the presence of 100 μM cadmium, the growth rate is rescued with the addition of CCC1 and furthermore with TaPCS1. Subfigure below represents the doubling time of each strain. Asterisk to the side of bar charts represent significant increase in growth rate compared to WT ($p < .01$). (d) S^{BCT} strain was titrated against cadmium, manganese, or cadmium in the constant presence of 100 μM manganese (x-axis). Metal uptake experiments was performed at varying concentrations from 1 to 100 μM , metal content analyzed using ICP, and values reported as percent uptake. S^{BCT} showed a higher preference for manganese than cadmium, with cadmium uptake being dramatically reduced in the background presence of 100 μM manganese (light blue curve). For all data, the mean \pm s.d. of three replicates are shown.

3.2.6 Designing a pipeline to engineer a metal specific transporter

SMF1 was evolved using a newly developed screening pipeline which uses density gradient centrifugation to fractionate mutants based on metal accumulation. Crystal structures and literature on Nramp structure-to-function was used to semi-rationally build libraries to create two variants of SMF1. The first variant was a more specific cadmium transporter, and the other was a strontium transporter for potential application for radioactive Sr⁹⁰ remediation.

The crystal structures of SMF1 homologues *D. radiodurans* (DraNramp) and *S. capitis* (ScaDMT) were used to narrow down transmembrane domains (TM) fundamental for metal recognition and transport [bozzi2016a, bozzi2016, courville2006, ehrnstorfer2014, xu2004]. Specifically, TM regions 1, 4 and 6 in the Nramp family have been identified to confer metal selectivity and movement [bozzi2016a, ehrnstorfer2014]. Without a crystal structure for SMF1, the specific TM regions have to be inferred from known structures or through multi-alignments of conserved regions. Multi-aligning SMF1 protein sequence against a Pfam database of homologous Nramps including DraNramp and ScaDMT revealed region 76-105, 180-200, and 264-287 to represent TM1, 4, and 6, respectively, based on the highest degree of conservation when compared to TM regions in the aligned homologues (Figure ??a; Supplemental Figure ??).

More so, previous work in Nramp mechanistic function showed that mutation M276 in SMF1 (discovered as M230 in DraNramp) confers metal selectivity [bozzi2016]. Outside crystallographic observations, it was empirically shown that mutating TM4 region G189 (discovered as G153 in DraNramp, or G185 in DMT1)33,43 into an arginine exposes a calcium entryway, which is hypothesized to also transport similar group II elements like strontium (Supplemental Figure ??).

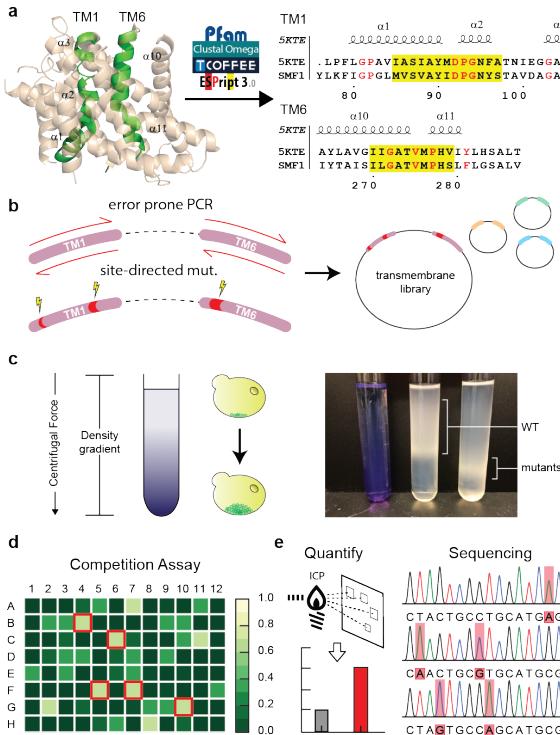


Figure 3.4 | Pipeline to systematically screen and engineer metal transporters for selective metal uptake. (a) Pfam protein database and clustering services such as Clustal ω , TCoffee, and E-Sprint3 were used to align SMF1 with referenced protein crystal structure 5KTE [bozzi2016a]. Through literature searches and multi-alignments, transmembrane 1 and 6 (TM1, 6) were found to be the most significant regions for mutagenesis. The alignment comparing 5KTE with SMF1 shows the TM1 and TM6 region, where yellow highlights indicate conserved regions, and red text indicate highly conserved residues (similarity score > 0.7). (b) Mutations cited to enhance or decrease metal transport were selectively mutated using site-directed mutagenesis. Libraries were then generated on top of these mutations through error-prone PCR. (c) An initial screen was performed through rate-zonal density gradient centrifugation. (d) Fractionated layers were plated, picked, and assayed for metal uptake. A competition assay of the desired metal versus the native metal (e.g. manganese) was performed calorimetrically. Wells with the least amount of native metal uptake (highest signal) were selected and (e) quantitatively measured for metal uptake using ICP. Mutations were sequenced and reintroduced in the pipeline to generate better performing mutants.

Mutating M276C and separately G189R and M276A were performed on SMF1 prior to generating libraries for cadmium and strontium screening, respectively. Given

these base mutations, error-prone PCR was done sequentially on TM1 and TM6 to generate libraries (Figure ??b) which were then transformed into BCT strains. Creating the cadmium and strontium mutant were performed in parallel, where separate libraries were screened for cadmium or strontium uptake. During screening, libraries were subjected to either 100 μ M cadmium or strontium similar to previous metal uptake experiments. Libraries were then screened based on an increase in mass as an indirect measurement for metal uptake. Mutants with higher metal content were fractionated using rate-zonal density gradient centrifugation (Figure ??c; Supplemental Figure ??). Rate-zonal, rather than isopynic density gradient centrifugation was used to fractionate cells based on changes in mass, rather than equilibrium density, as previous studies have shown that yeast maintain a relatively constant density despite external influences⁴⁴. Cells migrating the furthest were isolated, plated, and picked for colonies for a more focused metal assay. Cells were subjected to a competition assay with cadmium or strontium with 100 μ M manganese in a 96 well format. A colorimetric assay specific to manganese was performed on the supernatant, where wells with the highest intensity (low manganese uptake) corresponded with mutants with low manganese preference (Figure ??d). A select number of mutants were then chosen for quantitative metal uptake measurement using ICP, then sequenced, and later re-introduced into the mutagenesis/screening pipeline (Figure ??e). 4 rounds of screening were performed to generate a cadmium and strontium mutant.

3.2.7 Creation of SMF1 transporters specific to cadmium or strontium

The SMF1 mutant with the highest cadmium specificity (denoted as mCd) contained mutations S105C, M276C, and S269T; whereas the SMF1 mutant with the most selectivity for strontium (denoted as mSr) contained mutations G189R, T266S, M276C and G283Q (Figure ??a). To test the contributions of each mutation, SMF1* was systematically mutated at each of the changed residues to reveal their significance and effect on SMF1 expression and function. Many of the mutations on mCd and

mSr were located on TM6 rather than TM1, which supports past observations of the highly sensitive permeation region in the first alpha-helix segment of TM1 (Supplemental Figure ??). In addition, rounds of mutations leading to mCd and mSr did not significantly change expression levels (Supplemental Figure ??).

Supporting Bozzi's et al. work, M276 plays a critical role in metal selectivity³⁴, changing the methionine to cysteine doubles cadmium uptake while halving manganese uptake (Figure ??b) ($p < .05$). Whereas changing the methionine into alanine, and subsequently changing G189 into arginine enhances strontium uptake while dramatically reducing uptake of manganese by almost 8-fold ($p < .01$) (Figure ??b). These modifications, and each subsequent change, reduces Mn uptake while increasing uptake of Cd or Sr for mCd and mSr, respectively (Figure ??b). It should be noted that these mutations could impede Mn uptake allowing increased permissiveness of Cd and Sr transport, rather than strictly increasing sensitivity for Cd or Sr; a subtle yet important distinction. However, in either case, the goal of improving Cd or Sr uptake is shown for mCd and mSr, respectively.

Testing the fully mutated mCd and mSr a titration experiment with cadmium or strontium, respectively, in the background of 100 μM manganese was performed to determine the new KU's. For mCd the K_U for manganese dropped by 40-fold to 26.2 ± 7.6 ($p < .01$), whereas the K_U for cadmium went from 100 ± 3.2 without manganese to 75.8 ± 10.3 in the presence of manganese, a reduction by less than 25 percent ($p < .05$) in comparison to the 5-fold decrease with the non-mutated version ($p < .01$) (Figure ??d; Figure ??c). Similarly for mSr, the K_U for manganese dropped to 17.9 ± 1.6 ($p < .01$) whereas the K_U for strontium was 26.8 ± 5.7 and remained relatively constant at 27.1 ± 11 in the presence of manganese.

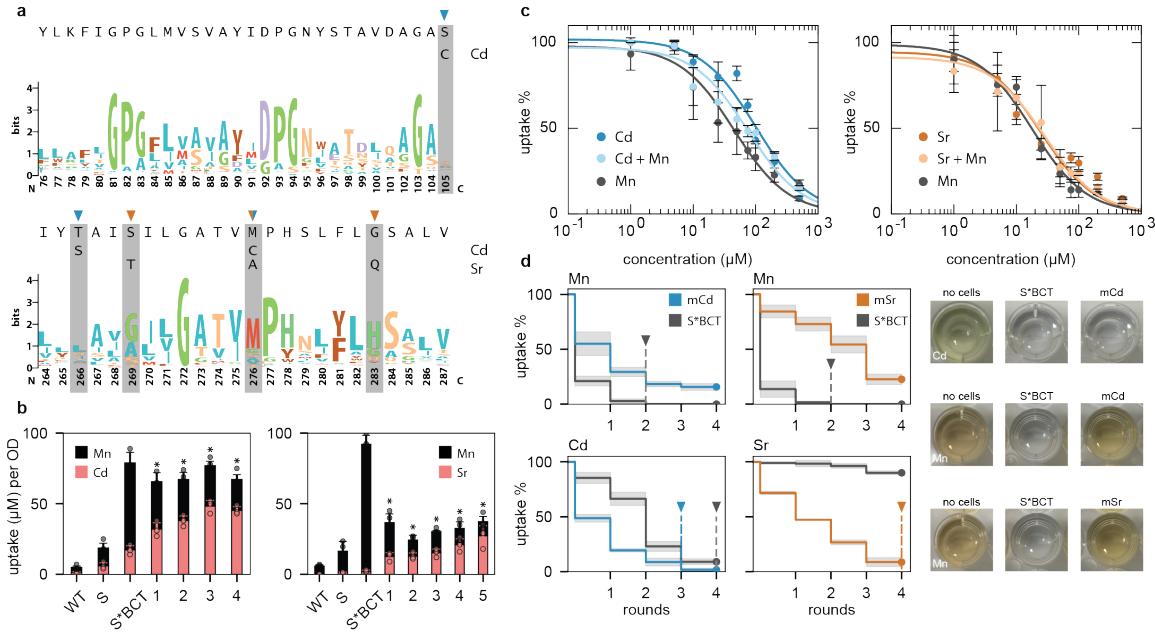


Figure 3.5 | Cadmium and strontium metal transport mutants were generated after 4 rounds of screening. (a) Weblogos of SMF1 TM1,6 from Nramp multi-alignment are displayed, with cadmium and strontium mutations highlighted. Cadmium mutants had S105C, T266S, and M276C. Strontium mutants had G189R, S269T, M276A, and G283Q. (b) Total metal uptake of $100 \mu\text{M}$ cadmium and manganese were measured to assess manganese interference. Cadmium mutant labeling corresponds to 1=M276C, 2=M276C+S105C, 3=M276C+T266S, and 4=M276C+S105C+T266S. Strontium mutant labeling corresponds to 1=M276A, 2=M276A+G189R, 3=M276A+G189R+S269T, 4=M276A+G189R+G283Q, and 5=M276A+G189R+S269T+G283Q. Strain background for all mutants were BCT. Asterisk above bar charts represent significant changes in both Cd and Mn uptake compared to un-mutated S*BCT ($p < .05$). (c) Titration curves of fully mutated cadmium and strontium transporters in strain BCT were performed for Cd or Sr, respectively, with or without $100 \mu\text{M}$ Mn; x-axis represents the concentration of either Cd, Sr, or Cd, Sr with Mn. (d) Sequential uptake experiments, up to 4 rounds, were performed to measure the amount of cycles required for complete elimination of $100 \mu\text{M}$ cadmium or strontium in a mixture of $100 \mu\text{M}$ manganese. Images on the right are colorimetric detection of cadmium and manganese (there are no available colorimetric assays for strontium at this concentration) showing selective preference for cadmium (no coloration) against native metal manganese (darkened well). Samples were performed in a 12-well. For all data, the mean \pm s.d. of three replicates are shown.

The improved preference for cadmium and strontium uptake is more obvious when performing iterative rounds of metal uptake. When comparing uptake of cadmium or strontium in the presence of manganese for mCd and mSr against the un-mutated S*, it took two rounds to fully remove manganese in the un-engineered case, while there still remains >10 percent manganese after 4 rounds for mCd and mSr showing a significant reduction in manganese uptake efficiency ($p < .01$) (Figure ??d). However, when measuring cadmium or strontium uptake, mCd completely removes cadmium after 3 rounds while the un-engineered strain requires 4 because of manganese uptake inhibition. For mSr, strontium incrementally decreases after each round without signs of manganese inhibition. After 4 rounds strontium levels reached below 10 percent, while the un-engineered case had >80 percent strontium remaining ($p < .01$) signifying a significant change in metal preference from manganese to strontium.

3.3 Discussion

This work demonstrates that yeast can be engineered to hyperaccumulate metals by overexpressing and evolving native metal transporters and engineering mechanisms for metal detoxification. The main design requirements for yeast hyperaccumulation are, (1) overexpression and engineered hyperactive membrane transporter activity, (2) overexpression of vacuole (or other organelle) transporters for metal compartmentalization, and (3) enhanced metal tolerance. Co-expression of a cell membrane transporter (SMF1) and a vacuole metal transporter (CCC1), enhanced metal uptake of manganese and cadmium by more than 10-fold, exceeding the plant hyperaccumulating threshold of 10 mg/gDW and 0.1 mg/gDW, respectively. In addition, simultaneous expression of CCC1 and plant phytochelatin synthase TaPCS1 rescue yeast survival in the presence of 100 μ M cadmium. In order to improve metal selectively against the preferred substrate manganese, and more towards cadmium or strontium, information from crystallographic and empirical observations from Nramp point mutations were utilized to strategically engineer relevant SMF1 transmembrane domains. Semi-rational mutagenesis of SMF1 combined with a novel screening pipeline based

on mass changes using rate-zonal centrifugation generated SMF1 variants with either cadmium or strontium preference, and more than 10-fold reduction in manganese selectivity.

Actual application of these yeast strains in real-world settings would require another layer of technological development, such as a container or cartridge to secure yeast in a controllable unit. Fortunately, these technologies exist, such as yeast packaging, freeze-drying, and delivery which are routine technologies found in the consumer market. A potential concept is to grow and store yeast in commercial filter-like cartridges where they can be housed in filtering units with size-exclusion cutoffs to prevent yeast leakage back into the purified waters. An additional layer of safety is to genetically modify these yeast with kill switches, or a metabolic reliance on a controlled nutrient such that removal from these containers will result in cell death [chan2016].

There are yet many more handles that offer better control over metal hyperaccumulation. Expression levels of membrane metal transporters, specifically SMF1, can be enhanced by performing additional ubiquitin associated lysine mutations, such as deleting specific proteases such as PEP4 [nikko2008, liu1999], or integrating multiple copies into the genome with inducible or constitutive expression. Uptake can be further enhanced by trafficking metals into other organelles such as the mitochondria, ER or Golgi which themselves harbor multiple metal transpoters [luk2003]. The same transporter screening pipeline developed in this work can also be used to engineer organelle metal transporters, such as CCC1, to further increase metal uptake and specificity in conjunction with a surface metal transporter like SMF1. A complementary approach, which is currently being investigated, is whether the deletion of metal antiporters can improve metal retention in organelles and enhance overall metal accumulation. It may be a promising strategy to delete antiporters from organelles, such as the Golgi, ER, mitochondria, etc. to gradually build up metal compartmentalization. This is particularly interesting, and also straightforward, if there is no good metal transporter candidate from these organelles or if they are difficult to rationally engineer. Finally, yeast morphology could be altered to allow higher uptake

capacity. Theoretically, the upper limit for metal uptake capacity is restricted to intracellular volume. If needed, organelle size, morphology, copy-number, and size of the entire yeast cell itself, can be controlled with key genes such as VPH1 and VTCs [desfougeres2016, turner2012]. Future work will assess whether increasing the physical volume of yeast or increasing the number of organelles such as the vacuole will lead to higher metal uptake capacity.

The major benefit of using transporters for metal hyperaccumulation and environmental remediation is the control over metal selectivity. Unlike current non-specific physicochemical techniques, biological transporters engineered for metal hyperaccumulation can distinguish less abundant yet more toxic metals over background elements. Biological systems have evolved a repertoire of transporters that can be leveraged for such selective metal uptake. This work demonstrated a focused study of SMF1; however, a similar approach using the same screening pipeline can be employed to other transporters mentioned earlier. Such engineering may be limited by the lack of structure-to-function knowledge and crystal structure availability for some transporters. However, advances in nanobody-aided crystallography, NMR, and cryoTEM may help elucidate transporter crystal structures for better mechanistic understanding [bozzi2016a, ehrnstorfer2014].

There are other areas in which yeast hyperaccumulators can have potential real-world application. Given the customizability of yeast and methods proposed here to engineer metal selectivity, there is a possibility to design yeast strains by demand. Certain geographic areas suffer from specific metal contamination because of specific industries, for example areas in Bangladesh and India suffer from arsenic and chromium poisoning due to the textile industry [robinson2001, manzoor2006]. Therefore, yeast can be tailored to selectively capture and remediate arsenic and chromium from their soils. Rather than providing a monolithic process for metal capture, which is the method employed today, more focused metal remediation processes can be achieved with higher efficiency. The second application is to recycle, or mine out heavy metals back from solutions. Waste typically contains a mix of metals, making it extremely difficult to process and especially difficult to extract and

re-capture precious metals. With this yeast-based approach it may be possible not only to remediate waste, but also to extract, concentrate, and store removed metals in yeast for mining purposes. Therefore if a certain mixture contains X number of metals of interest, it would be possible to design X number of strains to individually target and mine back those metals. Using yeast as a mechanism for metal removal as well as mining and recycling can close the loop between manufacturing, use, and disposal. Therefore, rather than providing a palliative solution for the waste management crisis, yeast can be used as an integral tool for waste treatment processes and recycling.

3.4 Materials and methods

Yeast strain and culture

Yeast strain W303 α was obtained from the Amon Lab at MIT. Synthetically defined dropout media (SD) was made by combining 1.7 g/L yeast nitrogen base without amino acid and ammonium sulfate (YNB) (Fischer), 5 g/L ammonium sulfate (Sigma), 0.6 g CSM-HIS-LEU-TRP-URA powder (MPBio), 20 g/L glucose (Sigma), and 10 mL/L of 100X adenine hemisulfate stock (1 g/L) (Sigma). 100X stocks of His (5 g/L), Leu (10 g/L), Trp (10 g/L), and Ura (2 g/L) (Sigma) were made in ddH₂O and filtered sterilized before supplementing cultures. Alternatively, complete synthetically defined media (CSM) was made with the above ingredients but with 0.79 g/L CSM mix (MPBio). YPD was made with 20 g/L peptone, 20 g/L glucose and 10 g/L yeast extract. CSM/SD-R media was made by replacing 20 g/L glucose with raffinose (VWR). Likewise, CSM/SD-G media was made by replacing 20 g/L glucose with 20 g/L galactose and 20 g/L raffinose. Solutions were stirred and filter sterilized through a .22 μ m filter top (EMD). Agar plates were made by adding 20 g/L BactoAgar (Fisher) and autoclaving before pouring.

Isolating genomic DNA (smash & grab)

Cultures were grown overnight in their appropriate drop-out media. 500 μ L of cells were then transferred and pelleted at 900xg for 3 min. Cells were then resuspended in 250 μ L DNA breakage buffer containing 2% Triton X-100, 1% SDS, 0.1 M NaCl, 10 mM Tris-HCl, 0.5 M EDTA (Sigma) in 100 mL ddH₂O. Approximately 1:1 of acid-washed 420-600 μ m glass beads (Sigma) to cell pellet were added to the tubes. 250 μ L phenol/chloroform/isoamyl alcohol (25:24:1; Sigma) was then layered on top. Tubes were bead beaten for 5 min. and spun down at 14000xg for 5 minutes at 4°C. The aqueous layer was then removed and added to 1 mL of ice-cold 100% EtOH (VWR) and spun down at 14000xg for 5 minutes at 4°C. EtOH was aspirated leaving behind precipitated DNA which was then dried at room temp for 30 min. Cells were then resuspended in TE buffer (Sigma) for downstream cloning.

Cloning metal transporters

Sequences were acquired from the Yeast Genome Database (www.yeastgenome.org) or through NCBI (<https://www.ncbi.nlm.nih.gov/gene>). All cloning steps were first simulated with Snapgene. All enzymes including the commercial non-trademarked Gibson assembly master mix, HiFi, were purchased from NEB. All references to Gibson assembly used the HiFi master mix. Between each PCR step, products were cleaned using the Wizard SV Gel and PCR CleanUp Kit (Promega). The pYES2/CT (Invitrogen) was used as the plasmid backbone for gene expression. The pYES2/CT vector was modified by inserting a stop codon after the V5 tag to eliminate expression of the C'-terminus 6xHis tag. All sequences were confirmed by Sanger sequencing using Quintara Bio.

Metal transporters, CTR1, CTR3, FET4, FTR1, SMF1, SMF2, ZRT1, and ZRT2 were amplified from genomic W303 α DNA using PCR and ligated into pYES2/CT via restriction cloning. Forward and reverse primers (Supplemental Table ??) of the metal transporter genes were flanked with the KpnI and XhoI restriction sites and trailed by *TAAAGCA* junk DNA to enable efficient restriction cleavage. All genes were

followed by the V5 tag native to the pYES2/CT vector.

Permeases Pho84, Pho87, Pho89, Sul1, Sul2 were amplified from genomic W303 α DNA using PCR and Gibson assembled into the pYES2/CT vector. During assembly overhangs contained a HA tag to replace the V5 tag of the pYES2/CT vector (Supplemental Table ??).

Nrat1 protein sequence was retrieved from Uniprot (www.uniprot.org), codon optimized, and synthesized using Genscript. Nrat1 was Gibson assembled into pYES2/CT and immediately followed by the V5 tag (Supplemental Table ??).

CCC1, COT1, ZRC1, and SMF3 were assembled into a modified pYES2/CT vector. The pYES2/CT original URA marker was replaced with a LEU marker taken from the pRS305 vector. The CCC1 gene was first cloned using restriction sites SacI and BamHI. The V5 tag was replaced with a Flag tag by introducing the appropriate primer overhangs. The remaining vacuole transporter genes COT1, ZRC1 and SMF3 were created by replacing the CCC1 via Gibson assembly (Supplemental Table ??).

Engineering S*BCT

A mutated version of SMF1 was performed by mutagenizing the K33,34 region, *AA-GAAA*, into arginines, *AGGAGA*, using QuikChange site-directed mutagenesis (Agilent) (Supplemental Table ??a). The BSD2 ubiquitin ligase gene was deleted by amplifying the HIS cassette from pRS303 with 40 bp overlap with the genomic BSD2 region using PCR. The PCR product was then used to transform competent yeast strains following the transformation protocol described below (Supplemental Table ??b).

TaPCS1 was ordered from Addgene (#49767; deposited by the Julian Schroeder Lab) and inserted into the pD1235 vector (ATUM) via Gibson assembly. The gene along with the TRP marker was amplified with 40 bp overlap over the *trp1-1* region of the W303 strain using PCR and transformed for genomic integration (Supplemental Table ??c).

pYES2/CT with CCC1 was modified to allow proper integration into the yeast genome. The vector was reorganized to have the LEU marker downstream of the

gene, the swap being made via Gibson assembly. CCC1 along with the LEU cassette was amplified with 40 bp overlap over the *leu2-3* region of the W303 strain using PCR and transformed for genomic integration (Supplemental Table ??d).

Identifying SMF1 TM regions and mutagenesis

Pfam (<https://pfam.xfam.org/>) was used to curate the representative proteomes from the Nramp family and were compared using TCoffee’s transmembrane multi-alignment algorithm (<http://tcoffee.crg.cat/>). To check the accuracy of this tool the same dataset was aligned using Clustal Omega (<https://www.ebi.ac.uk/services/teams/clustal-omega>), which showed similar results. The resultant multi-aligned file was visualized using ESPript (<http://escript.ibcp.fr/ESPript/ESPript/>) with reference sequence taken from PDB entry 5KTE to help indicate regions with secondary structure. Red highlighted amino acids indicate highly conserved regions with similarity scores > 0.7 . All other amino acids are colored black. Visualized alignments identified transmembrane regions on SMF1, and mapped residues G153 and M230 found in 5KTE to G189 and M276 on SMF1, respectively. Sequence usage of the Nramp family was also visualized using WebLogo (<https://weblogo.berkeley.edu/logo.cgi>) and mapped onto TM1,4 and 6 of SMF1 to qualitatively understand the significance of mutated regions during screening.

Libraries of SMF1 were generated using primers flanking TM1 and 6 which were then used with Agilent’s GenemorphII EZClone mutagenesis kit (Supplemental Table ??a). Site-directed mutagenesis primers were created using Agilent’s primer design webservice (www.agilent.com/genomics/qcpd) and mutations were introduced using Agilent’s Quikchange lightning or multi-site mutagenesis kits (Supplemental Table ??b).

Transformations

Plasmid constructions were performed in NEB α competent cells (NEB) and transformed following NEB's protocol. Yeast transformations were performed with the Frozen-EZ Yeast Transformation Kit II (Zymo Research). The protocol was modified slightly for integrated constructs. Transformed cells were first plated onto YPAD plates and grown for 1 day. Plates were replica-plated on their respective SD drop-out and grown for an additional 1–2 days. 4–8 colonies were then picked, grown overnight, and smash & grabbed to isolate their genomic DNA. DNA was then amplified with primers flanking the integrated area of interest using PCR, and PCR product ran on a gel to verify proper integration.

Correlating OD₆₀₀ to culture dry weight (gDW)

Wild-type W303 α were grown and diluted to various culture densities ranging from 0.1–2 OD₆₀₀ in 500 mL. Cells were pelleted and washed 3X in ddH₂O. 50 mL conical tubes were pre-weighed on an analytical balance with microgram resolution. Cells were transferred into these tubes, pelleted, and resuspended in 1 mL of ddH₂O. Tubes were then dipped and snap-freezed in liquid nitrogen. Tubes were then capped with a porous cloth and fitted into a lyophilisation chamber (VirTis) and lyophilized for 48 hours. Tubes with cells were weighed with weight of the tube subtracted to calculate cell dry weight. Mass of cells per volume (y-axis) was plotted against measured OD₆₀₀ (x-axis) giving a ratio between OD and culture dry weight per culture volume.

Metal uptake analysis using inductive coupled plasma (ICP)

Liquid stocks of copper (II) chloride, zinc chloride, iron (II) chloride, manganese (II) chloride, cadmium nitrate, and strontium chloride (Sigma) were made at 100 mM in ddH₂O and filtered through a .22 μ m filter. Colonies streaked on SD agar plates were picked and inoculated in SD-R media with the appropriate supplemented amino acids. Overnights were diluted 1:10 in SD-R and grown for 4 hours. Cells were then pelleted and resuspended in SD-G media for induction overnight.

To prepare cells for metal uptake analysis, cells induced with SD-G were diluted to 1 OD₆₀₀ in fresh SD-G and spiked with 100 μM metal and incubated for 4 hours at 30°C. In conjunction, a control sample containing a wild-type W303α (WT) strain was also spiked with 100 μM metal and processed similarly to account for non-specific metal uptake in each experiment. Another sample containing no cells was also spiked with 100 μM metal to test for non-specific metal binding to the test tube and measurement equipment in each experiment. Overall, no significant non-specific binding was observed for the cell-free samples.

After 4 hours of metal incubation OD₆₀₀ was measured again to take into account any changes in culture density. Afterwards, cells were pelleted and supernatant collected for metal analysis. Metal concentrations were measured in an inductive coupled plasma (ICP) Agilent ICP-AES 5100 instrument following standard operating procedures provided by the Center of Material Science facility at MIT. Metal standards were made from ICP-grade stock solutions purchased from Fluka and diluted in a 2–3% HNO₃ matrix. After ICP analysis metal uptake was calculated by subtracting 100 μM (original metal concentration) by the metal concentration measured in the supernatant. The value was then divided by the OD₆₀₀ measurement to give units of $\mu\text{M}/\text{OD}$ in order to equally compare uptake levels between strains and metals. Units were further converted to help compare against literature values which report milligram of metal per gram of culture dry weight (mg/gDW) by multiplying the molarity of metal uptake by the molecular weight of the metal used, and converting the culture OD into gram dry weight using the ratio derived in the OD₆₀₀ to culture dry weight analysis above.

Metal uptake titration experiments were performed following the same method but using different metal concentrations ranging from 1 μM to 100 μM . Metal uptake was normalized to percent uptake with respects to the original metal concentration added. The concentration at which 50% of metal was uptaken was termed K_U. For interference experiments, titrations against the desired metal (cadmium or strontium) was performed in the presence of constant 100 μM manganese.

Iterative metal uptake experiments was performed by taking the supernatant of

a previous metal uptake experiment, and transferring the supernatant directly into a freshly induced culture normalized to 1 OD₆₀₀. Uptake was performed for 4 hours, and supernatant transferred iteratively to a fresh new culture up to 4 times. At each iteration the supernatant was sampled and measured using ICP to calculate the metal uptake per round.

Staining and microscopy

Transporter expression was measured using immunohistochemistry. Cells were induced following the procedure mentioned above and fixed with 3.7% paraformaldehyde (EMS) at 0.5 OD₆₀₀ for 30 minutes at room temperature. Cells were pelleted at 900xg and washed 3X in 1.2 M sorbitol-citrate buffer (Sigma) before resuspending in the same buffer with 1:100 dilution of 100T Zymolyase (Zymo) and incubated at 30°C for 30 minute to 1 hour. Cells were pelleted and washed 3X in PBS+1% BSA before settling on poly-lysine treated 8 well chamber slides (Lab-Tek). Cells were gently permeabilized with 0.1% Tween-20 (Sigma) in PBS+1% BSA on ice for 5 min. Cells were then stained with the appropriate primary antibody against V5, HA, or Flag tag, washed 3X in PBS+1% BSA, and stained with the appropriate secondary antibody conjugated to AlexaFluor488 or 647 (Thermo). DAPI at 5 µg/mL in PBS was used to stain nuclei for 3–5 minutes. Cells were washed and aspirated before removing the wells. A 24x50mm coverslip was placed gently on the slide with 60% glycerol in PBS as the mounting media. Nail polish was used to seal the edges and slides were imaged on an AxioPlan2 within 24 hours.

In all experiments, a non-expressing WT control was stained in parallel to measure non-specific antibody binding and autofluorescence. The same primary and secondary antibodies (V5, HA, Flag tag), and staining conditions were performed similarly with the experimental samples.

Quantifying membrane expression using flow cytometry

SMF1 variants (S, S*, SB, S*B, iS*B) were stained with antibodies following the same steps in the staining and microscopy methods. Cells were diluted to 0.1 OD₆₀₀ in PBS+1% BSA and measured on a BD FACS Canto or LSR II following standard operating procedures provided by the Koch Flow Cytometry Core. Cell counts were plotted against binned fluorescent intensity (x-axis) creating a population distribution histogram of fluorescence (y-axis). The mean fluorescent intensity weighted by cell count was used to quantitatively compare fluorescent intensity (i.e. expression) against metal uptake measured by ICP for those strains.

Cell culture density measurements and viability assays

OD₆₀₀ measurements were performed using 2 mL non-frosted cuvettes and a table-top DU800 Beckman Coulter spectrophotometer measured at 600 nm. OD₆₀₀ values were used to divide metal uptake values measured by ICP to normalize for culture density.

Cell viability was measured at different metal concentrations ranging from 1 μ M to 100 μ M. Cultures were grown overnight and then diluted to < 0.1 OD₆₀₀. Cultures were aliquoted to a total volume of 100 μ L and spiked with varying metal concentrations. Cultures were placed in a 96 well U-bottom plate (Cellstar) and shaken in a BioTek Synergy 2 plate reader held at 30°C for 24-36 hours. Growth rates were calculated by finding the maximum slope in the growth curve.

A live-dead assay was also performed to analyze cell viability by calculating the ratio of live to dead cells after metal uptake experiments. Cells after metal uptake experiments were resuspended in culture media and dyed with a live-dead fluorescent indicator (Thermo). A positive control of freshly grown cells, and a negative control of cells heated to 70°C for 15 minutes, were used to gate the live and dead cell populations, respectively. Counts within those gates were used to calculate ratio of live cells after metal uptake. Cells were analyzed under the FITC and PE channels of an LSR II flow cytometer.

Manganese assay

The manganese colorimetric detection Hach kit was modified to fit a 96-well format. 50 μ L of sample was added to 50 μ L of 2X ascorbic acid provided by the kit. Then 5 μ L of the cyanide and PANI reagent were used to detect manganese given a colorimetric change from yellow to red. Wells were measured at 560 nm. Cyanide was disposed of using guidelines approved by MIT EH&S.

Screening of yeast libraries using rate-zonal density gradient centrifugation and colorimetric assay

Percoll (Sigma) buffered with 1.5 M NaCl was used to make density gradients. A Pharmacia LKB Pump P-1 peristaltic pump joined to a gradient maker (GE) was used to make Percoll gradients. Gradients were formed in Greiner 16 x 100 mm round bottom polystyrene tubes (Sigma) which were first hydrophobically coated with Sigmacote (Sigma). A purple dye was used as a control to visually inspect consistency of gradient formation per batch.

Libraries were transformed into yeast and plated. Single colonies were pooled together using a scraper (Corning) into 10 mL of SD-R with the appropriate amino acids. Cells were grown for 12 hours before being diluted into 50 mL of SD-R for 4 hours. Cells were then pelleted and resuspended in SD-G with the appropriate amino acids for induction overnight. Induced culture was then diluted to 1 OD₆₀₀ into multiple 10 mL SD-G media with spiked 100 μ M of cadmium or strontium. Cultures were grown for 4 hours before washing and resuspending in 150 mM NaCl. Cells, percoll gradient, and an Eppendorf 5804-R swinging bucket centrifuge were chilled to 16°C before spinning. Settings for acceleration and braking were set to 0. Cells were gently layered onto the gradient and spun in increments of 5 min at 100xg. 3–4 spins were sufficient to see segregation of cells which signified a fractionation of heavier cells due to metal uptake. A centimeter below the least visible band was collected and spun down at 1500xg for 3 min before resuspending in SD with the appropriate amino acids. Cells were rescued for 1.5 hours before plating. Collected cells were

plated onto 2-3 plates giving approximately 10-100 colonies each.

After plating roughly 10–50 colonies, depending on the number of colonies formed, were picked in 100 μ L SD-R cultures in a 96-well format and induced following the same protocol as before. Cells were diluted to 1 OD₆₀₀ and spiked with 100 μ M cadmium or strontium with the addition of 100 μ M manganese and shaken for 4 hours. Plates were spun down at 900xg for 3 min and the supernatant was diluted 1:10 in ddH₂O and assayed using the modified manganese Hach detection kit described above. The top 4–6 Wells with the highest readings (most manganese remaining) were selected and plated again. Selected colonies were then subjected to a more thorough metal uptake ICP experiment and sequenced before re-introducing them into the screening pipeline.

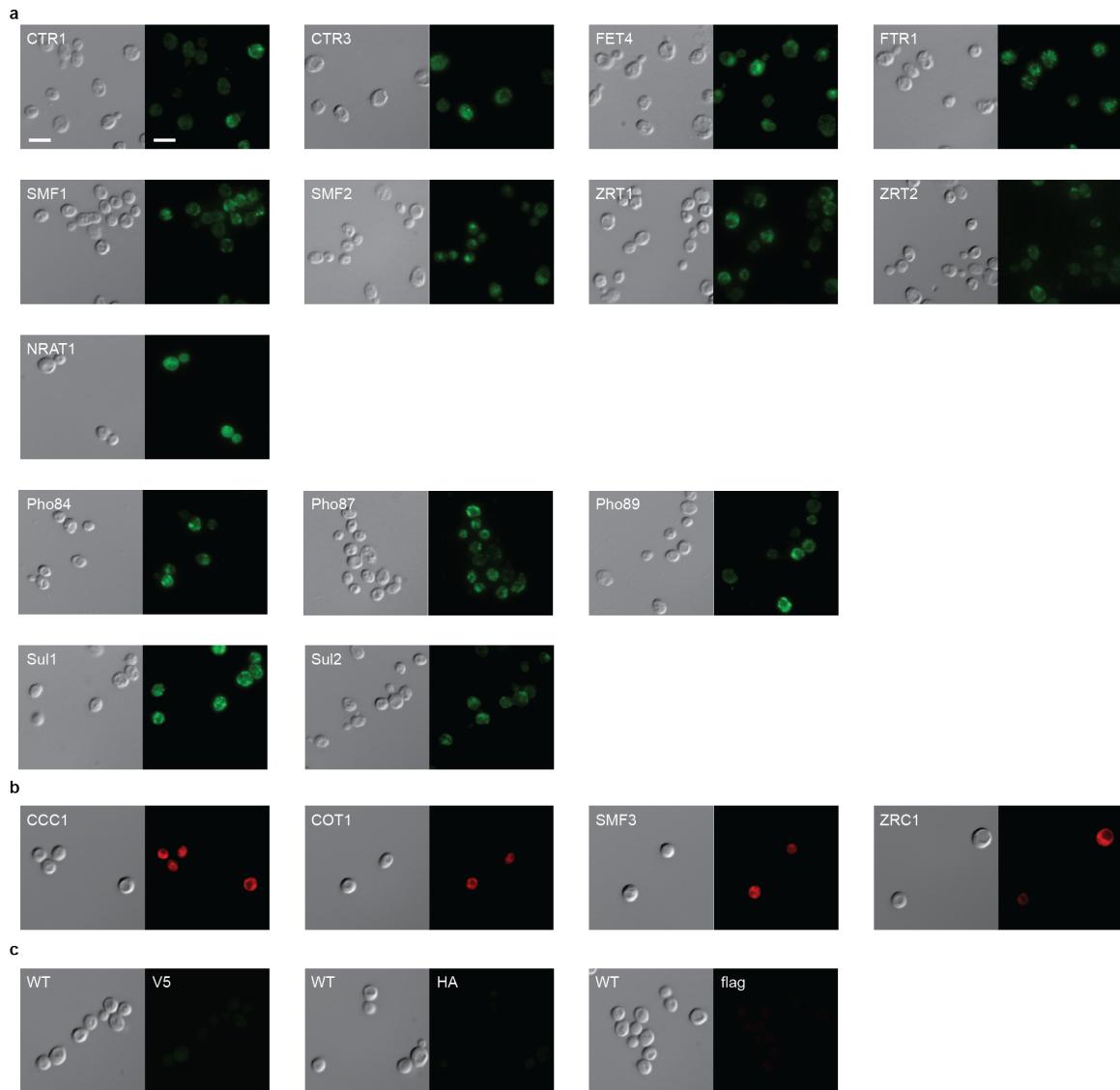
Mathematical analysis and plotting

Raw data was collected and stored as csv or excel file formats. Data was imported and analyzed with python using modules such as numpy, pandas, and scipy. Plots were graphed using matplotlib.

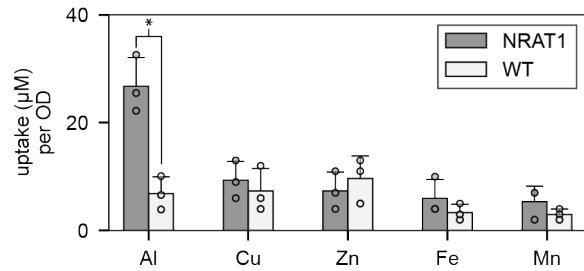
Statistical analysis

Statistical parameters including the the definition and values of n, SDs, and/or SEs are reported in the figures and corresponding figure legends. When reporting significance, a two-tailed unpaired t-test was performed between observations and p-values reported in the text. The significance threshold was set to p < .05 for all experiments, or as specified in the text.

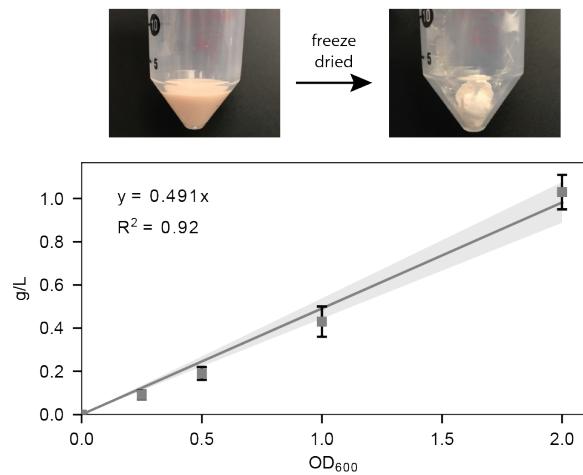
3.5 Supplemental figures



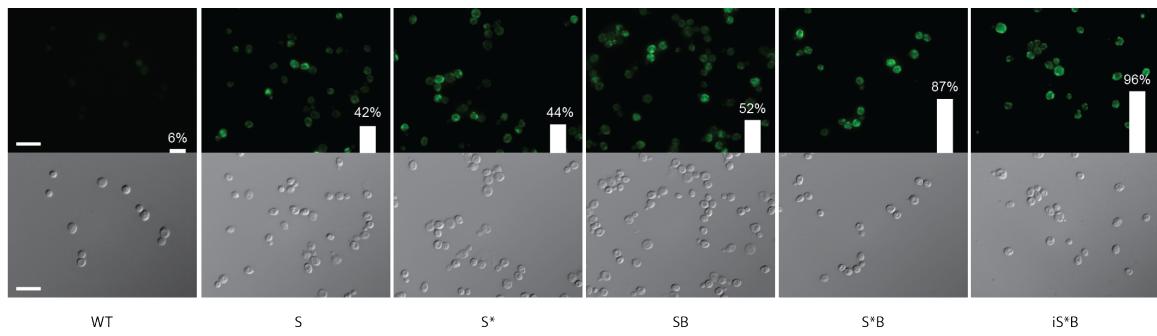
S.Figure 3.1 | Fluorescently stained membrane and vacuole transporters to visualize overexpression. (a) Membrane transporters CTR1, CTR3, FET4, FTR1, SMF1, SMF2, ZRT1, ZRT2 and Nrat1 were fused with a C'-terminus V5 tag and stained. Pho84, Pho87, Pho89, and Sul1 and Sul2 were fused with a C'-terminus HA tag and stained. (b) Vacuole transporters CCC1, COT1, SMF3, and ZRC1 were fused with a C'-terminus flag tag and stained. (c) Negative controls of WT were stained with identical antibodies targeting V5, HA, and flag tag in parallel with the transporters already described. No noticeable background fluorescence was observed. All tags were labelled with the appropriate primary and secondary antibodies conjugated with either an Alex488 (green; membrane) or Alex647(far red; vacuole) dye. Scale bars represent 5 μ m for all images.



S.Figure 3.2 | Nrat1 Al uptake compared to other transition metals used in this study. Metal uptake experiments for Nrat1 was performed with Al, Cu, Zn, Fe, and Mn and compared against non-expressing WT strain. Asterisk above bar chart represent significant uptake when compared to WT ($p < .05$). For all data, the mean \pm s.d. of three replicates are shown.

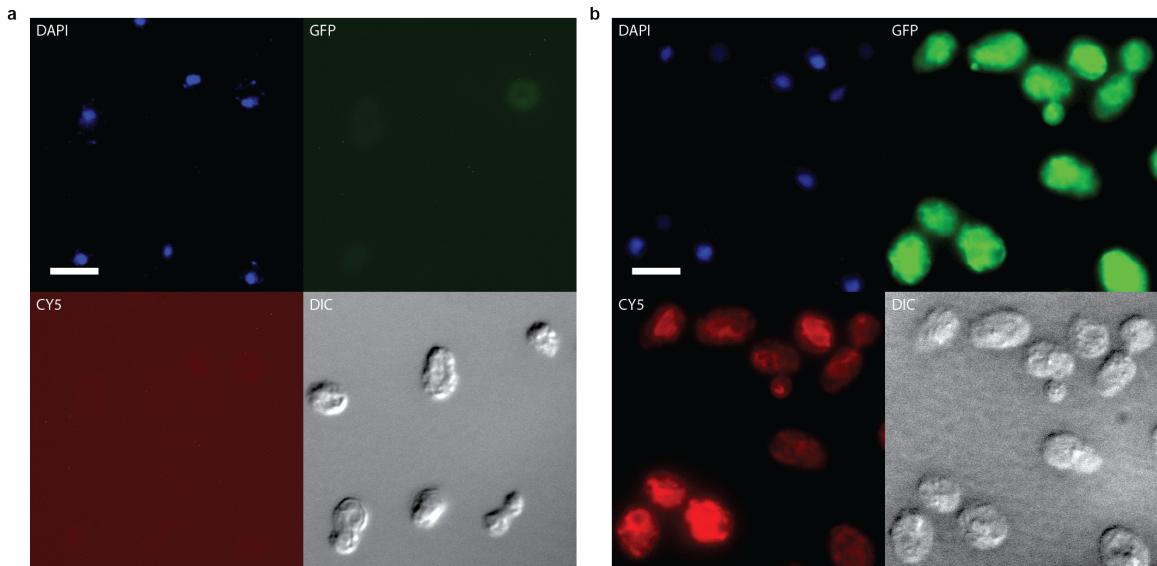


S.Figure 3.3 | Correlating culture optical density (OD₆₀₀) to grams of culture dry weight (gDW). (a) Cells were grown to the appropriate OD₆₀₀, washed, pelleted, and freeze-dried to obtain culture dry weight per culture volume. Masses were weighed on a precision scale with hundredths of milligram resolution. A line of best fit with intercept at 0 was performed to obtain a correlation between OD₆₀₀ and gram of dry weight per culture volume. (b) Relationship for WT was 1 OD₆₀₀ : 0.491 ± 0.05 mg of gDW per liter. (c) Relationship for engineered S*BCT was 1 OD₆₀₀ : 0.506 ± 0.1 gDW per liter. The difference in correlation between WT and S*BCT was negligible. For all data, the mean ± s.d. of three replicates are shown.

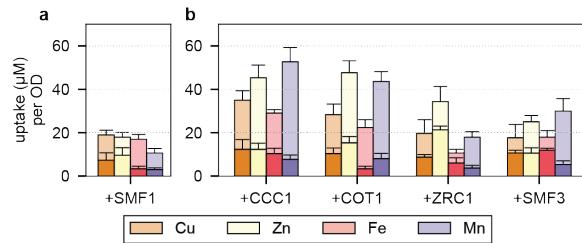


S.Figure 3.4 | Expression of SMF1 increased with increasing modifications.

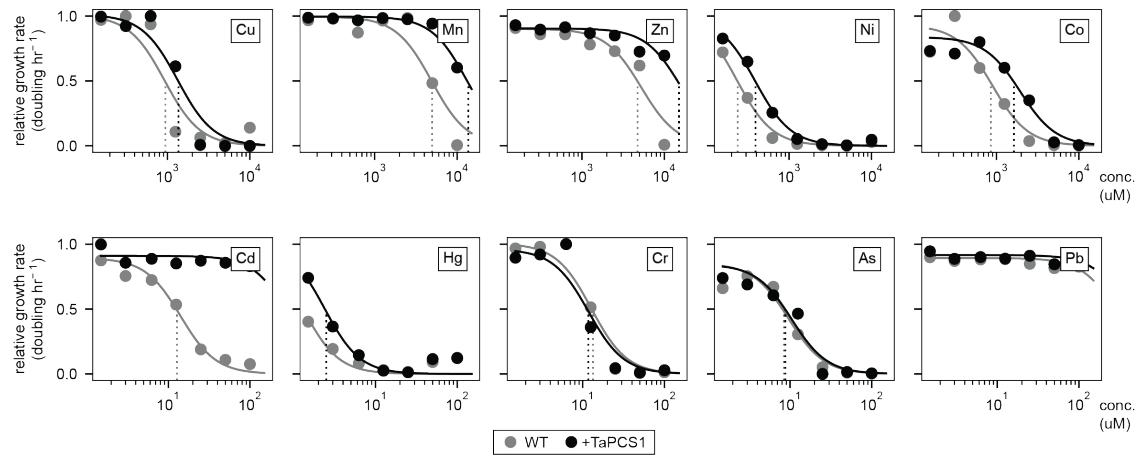
Labels from left to right: WT, overexpression of SMF1 (S), SMF1 with K33,34R mutation (S*), S with BSD2 deletion (SB), S* with BSD2 deletion (S*B), and S integrated with BSD2 deletion (iS*B). Bars and values indicate percent expression after subtracting background signal from WT controls. Images and expression scores were calculated using ImageJ. Scale bars represent $10 \pm m$ for all images.



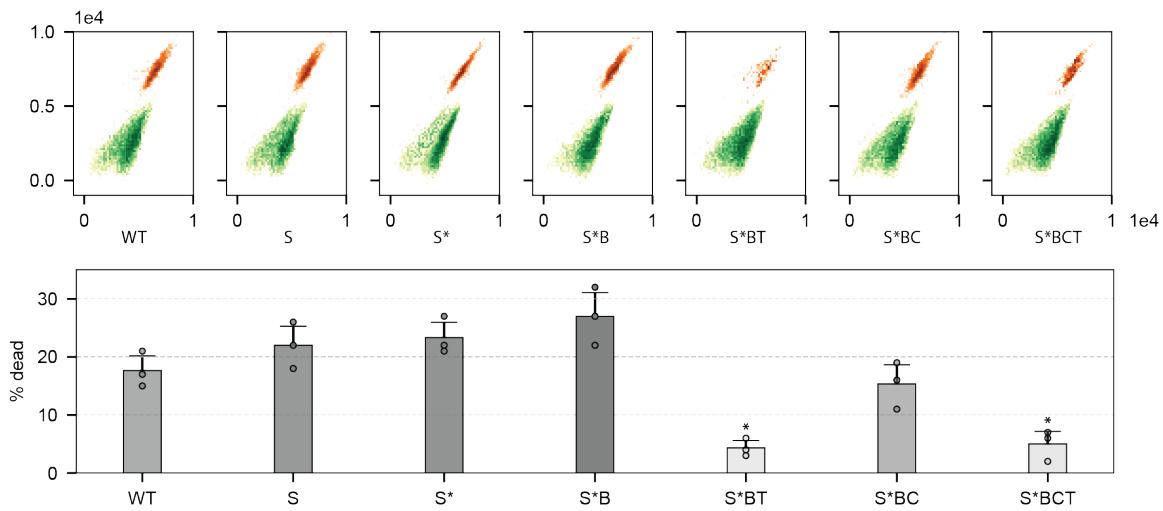
S.Figure 3.5 | Fluorescently labelled SMF1 and CCC1 co-expressing strains. (a) Fluorescent measurements of non-expressing WT strain as a control. (b) SMF1* was fused with a C'-terminus V5 tag, whereas CCC1 was fused with a C'-terminus flag tag. Tags were stained with the appropriate primary and secondary antibodies conjugated with either an Alex488 (green; SMF1*) or Alex647(far red; CCC1) dye, respectively. Scale bars represent 5 μ m for all images.



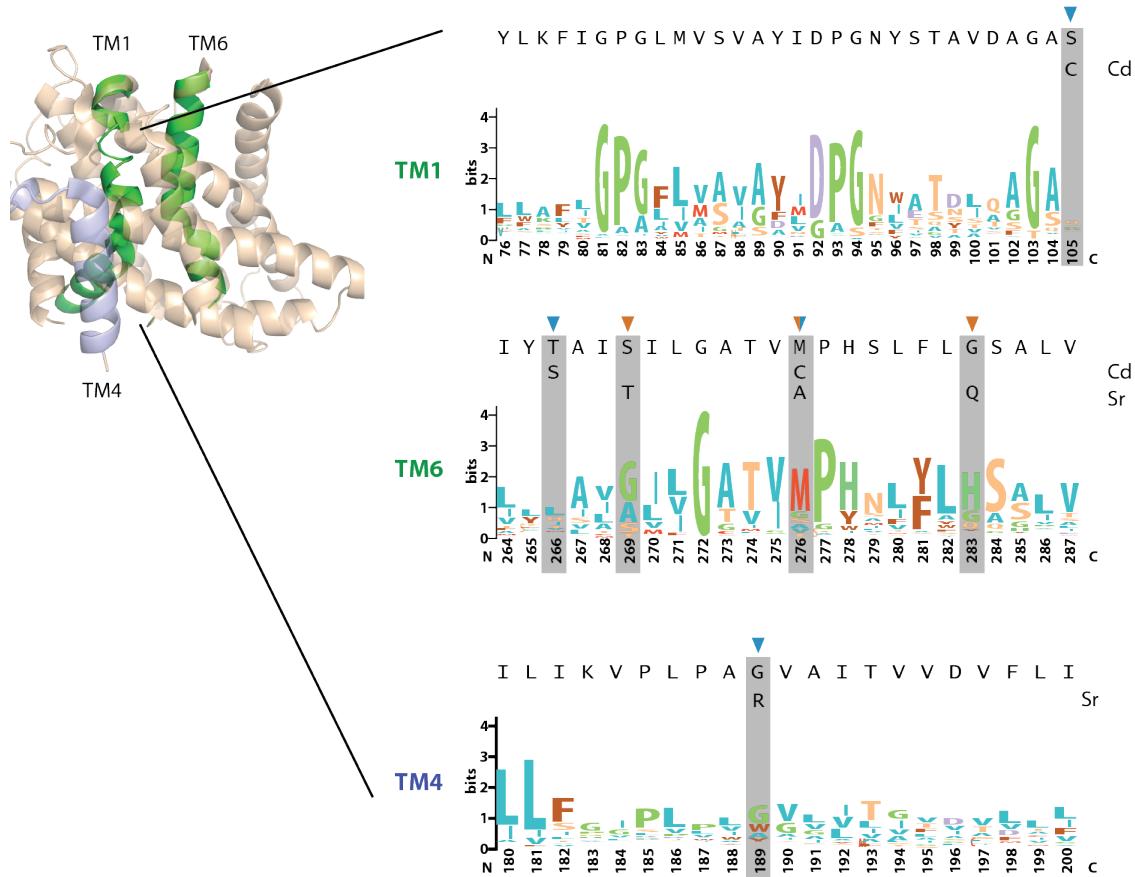
S. Figure 3.6 | Metal uptake increased with the addition of a vacuole transporter. (a) Heavier colored bars represent WT metal uptake. Lighter colored overlaid bars represent metal uptake with SMF1 expression (S; no modifications). (b) Heavier colored bars represent metal uptake with vacuole transporter expression. Lighter colored overlaid bars represent metal uptake with co-expression of SMF1. For all data, the mean \pm s.d. of three replicates are shown.



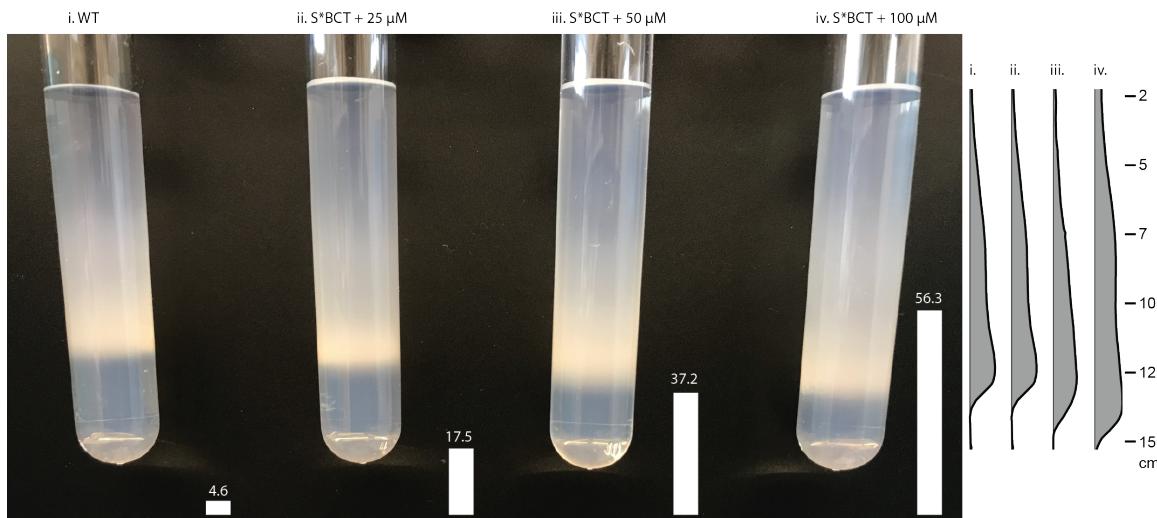
S.Figure 3.7 | Metal effects on WT and TaPCS1 growth rates at varying concentrations. Cu, Mn, Zn, Ni, and Co were assayed starting from 10 mM. Cd, Hg, Cr, As, and Pb were assayed starting from 100 μM . The addition of TaPCS1 slightly enhanced tolerance to Cu, Mn, Zn and Co in the millimolar range. Of the more toxic elements, TaPCS1 specifically conferred tolerance to Cd while changing little against Hg, Cr, and As. Growth rate curves for Pb and Fe were misleading as Fe and Pb precipitated in culture during the 24 hour growth experiments. For all data, the mean \pm s.d. of three replicates are shown.



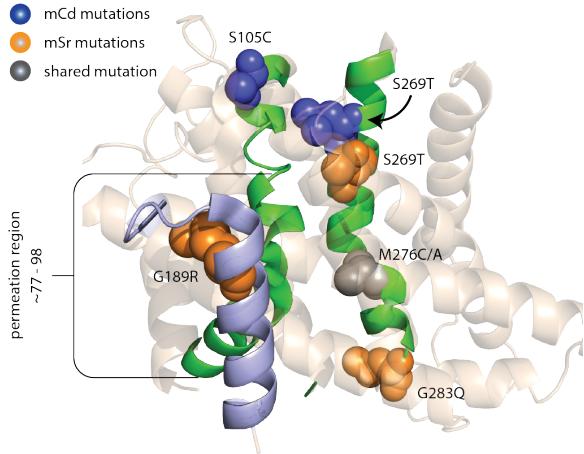
S.Figure 3.8 | Percent survival after metal uptake experiments with 100 μM cadmium using a FACS live-dead assay. Increased transporter expression (Supplemental Figure ??) and cadmium uptake (Figure ??a) loosely correlated to increase cell death during metal uptake experiments (WT \rightarrow S \rightarrow S* \rightarrow S*B). Expression of TaPCS1 (T) and CCC1 (C) enhanced cell viability despite increased cadmium uptake. Asterisk above bar charts represent significance change in survival percentage ($p < .05$) compared to WT. For all data, the mean \pm s.d. of three replicates are shown.



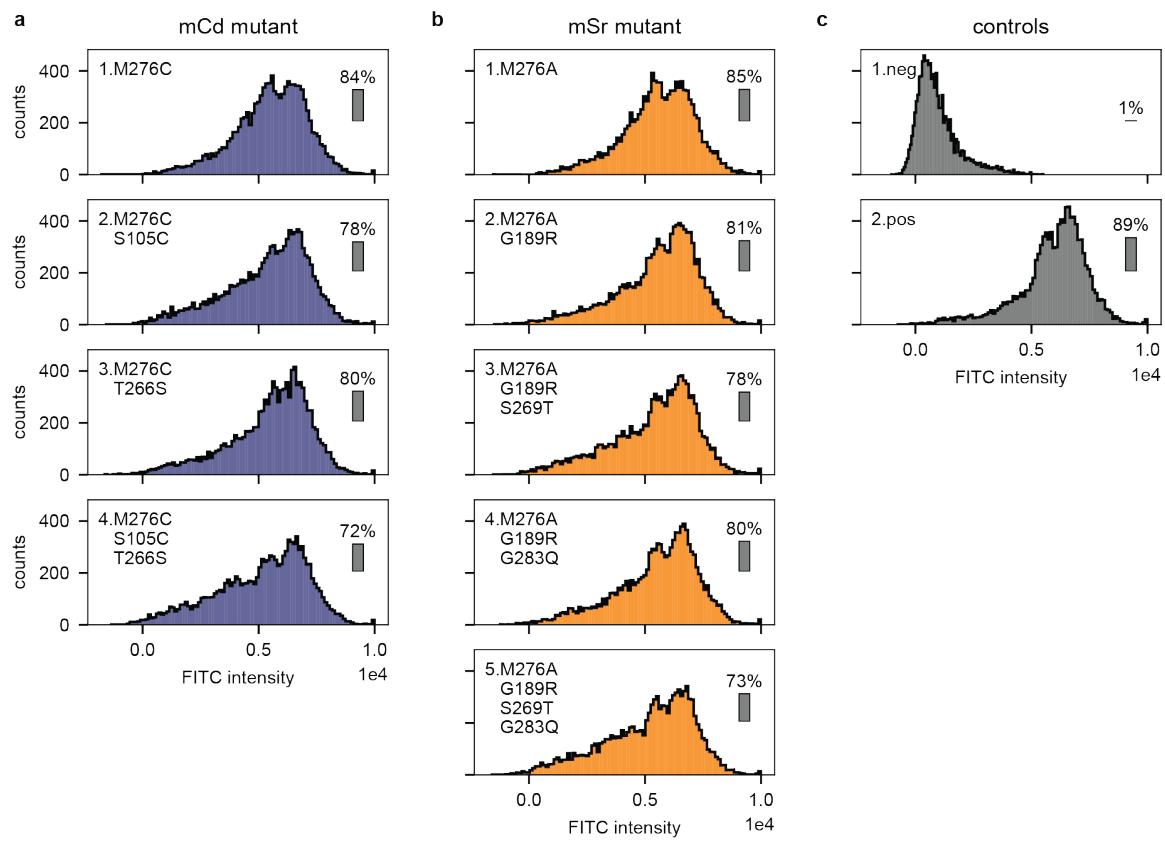
S.Figure 3.9 | Weblogos and mutagenesis annotations of transmembrane regions for TM1 and 6 (green) and TM4 (blue). Mutations introduced based on previous findings [bozzi2016] were M276C for cadmium mutants, and G189R and M276A for strontium mutants. Mutations discovered using the developed transporter assay were S105C, and T266S, for cadmium mutant mCd and S269T and G283Q for mSr.



S.Figure 3.10 | Fractionating cells based on metal uptake using rate-zonal density gradient centrifugation. Conditions from left to right were: (i) WT incubated with 100 μM cadmium; (ii), (iii), and (iv) S*BCT strain incubated with 25, 50, and 100 μM cadmium, respectively. Bottom bar charts indicate the amount of metal uptake per condition. Left chart shows the population distribution of migrated cells by measuring the tube's opacity as a function of height using ImageJ. Ticks represent distance measured from the meniscus to the tube's bottom in centimeters.



S.Figure 3.11 | Approximate mutation locations for mCd and mSr are highlighted on homologue DraNramp (PDB 5KTE). Many of the mutations for both mCd and mSr reside on TM6, or at the entry of TM1. This could suggest that the region 77-98 (41-61 for DraNramp) in the first alpha-helix sequence of TM1 is highly sensitive to mutations, as observed in previous works [bozzi2016a, ehrnstorfer2017, courville2006]. This region, referred to as the permeation region [bozzi2016a], has a highly conserved DPGN sequence which may act as an actuator to transport metals through the inner cavity. Whereas, TM6 and 4 may provide the spacing and environment to select for certain metals.



S. Figure 3.12 | Effect of mCd and mSr mutations on SMF1 expression measured with flow cytometry. (a) Progressive mutations (1-4; corresponding to numberings in Figure 5b) leading to mCd caused little change in transporter expression. (b) Likewise, a slight reduction in mSr transporter expression was observed, but not significant. (c) Negative (WT) and positive controls (S*BCT) were also measured to compare expression changes with non-engineered strains.

3.6 Supplemental tables

name	element	threshold (mg/gDW)	uptake (mg/gDW)	hyper?
ZRT1	Zn ²⁺	10	5.80 ± 0.85	FALSE
ZRT2	Zn ²⁺	10	4.42 ± 0.95	FALSE
CTR1	Cu ²⁺	1	5.43 ± 0.68	TRUE
CTR3	Cu ²⁺	1	3.84 ± 0.86	TRUE
FTR1	Fe ²⁺	1	1.83 ± 0.40	TRUE
FET4	Fe ²⁺	1	2.48 ± 0.32	TRUE
SMF1	Mn ²⁺	10	1.17 ± 0.22	FALSE
SMF2	Mn ²⁺	10	1.01 ± 0.28	FALSE
Pho84	AsO ₄ ²⁻	1	3.91 ± 0.54	TRUE
Pho87	AsO ₄ ²⁻	1	1.91 ± 0.14	TRUE
Pho89	AsO ₄ ²⁻	1	2.63 ± 0.39	TRUE
Sul1	CrO ₄ ²⁻	1	2.48 ± 0.25	TRUE
Sul2	CrO ₄ ²⁻	1	2.38 ± 0.21	TRUE
NRAT1	Al ³⁺	1	1.33 ± 0.21	TRUE
S*BCT	Mn ²⁺	10	57.5 ± 2.62	TRUE
S*BCT	Cd ²⁺	0.1	12.6 ± 1.18	TRUE

S.Table 3.1 | Transporters and strains developed in this work that are within or have exceeded hyperaccumulating thresholds [prasad2003, kramer2010metal] for their respective metal.

gene	direction	primer
ZRT1	fwd	TAAGCAGGTACCATGAGCAACGTTACTACG
	rev	TAAGCACTCGAGAGGCCACTTACCGATC
ZTR3	fwd	TAAGCAGGTACCATGGAAAAAATTCCCAGGTG
	rev	TAAGCACTCGAGAGTGAAAAGGGCACTC
CTR1	fwd	TAAGCAGGTACCATGGAAGGTATGAATATGG
	rev	TAAGCACTCGAGGTTATGAGTGAATTTCG
CTR3	fwd	TAAGCAGGTACCATGAATATGGGAGGCAG
	rev	TAAGCACTCGAGCAAGCAGCATTGC
FET4	fwd	TAAGCAGGTACCATGGTAAAATTGCAGAG
	rev	TAAGCACTCGAGTTTCCAACATCATAACC
FTR1	fwd	TAAGCAGGTACCATGCCTAACAAAGT
	rev	TAAGCACTCGAGAAGAGAGTCGGCTTAAC
SMF1	fwd	TAAGCAGGTACCATGGTGAACGTTGG
	rev	TAAGCACTCGAGACTGATATCACCATGAGAC
SMF2	fwd	TAAGCAGGTACCATGACGTCCAAAGAATAT
	rev	TAAGCACTCGAGGAGGTGTACTTCTTGCG

S.Table 3.2 | Primers used to clone divalent metal transporters into the pYES2/CT vector using restriction sites KpnI and XhoI.

gene	direction	primer
Pho84	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGA GTTCCGTCAATAAAGATAC
	rev	AGCGTAGTCTGGAACGTCGTATGGGTAGGATCCA CCGCCTGCTTCATGTTGAAGTTGAG
Pho87	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGA GATTCTCACACTTCTCA
	rev	AGCGTAGTCTGGAACGTCGTATGGGTAGGATCCA CCGCCAGTGCTACCTTTAAGACCG
Pho89	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGG CTTTACATCAATTGACT
	rev	AGCGTAGTCTGGAACGTCGTATGGGTAGGATCCA CCGCCTGTCATTTGGTATTCCACAC
Sul1	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGTC ACGTAAGAGCTC
	rev	AGCGTAGTCTGGAACGTCGTATGGGTAGGATCCA CCGCCAACGTCCCATTAGAAAAATC
Sul2	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGTC CAGGGAAAGGTAA
	rev	AGCGTAGTCTGGAACGTCGTATGGGTAGGATCCA CCGCCGATATCCCATTAGCAAAATC
pYES2/CT	fwd	GAGCTCGGTACCAAGCTTAATATTC
	rev	GGCGGTGGATCCTACCCATACGACGTTCCAGACTA CGCTTAAGTTAAACCCGCTGATCC

S.Table 3.3 | Primers used to clone permeases into pYES2/CT in addition to substituting the V5 tag with a HA tag given the appropriate primer overhangs.

gene	direction	primer
NRAT1	fwd	CTCACTATAGGAATATTAAGCTTGGTACCATGGA AGGTACTGGTGAAATG
	rev	ACCGAGGAGAGGGTTAGGGATAGGCTTACCCATA CTAGCATCTGCCAAATCTT
pYES2/CT	fwd	GGTAAGCCTATCCCTAACCC
	rev	GGTACCAAGCTTAATATTCCCTATAGTG

S.Table 3.4 | Primers used to clone Nrat1 into pYES2/CT. Nrat1 was first codon-optimized from *O. sativa* and synthesized from Genscript before amplifying and assembling into pYES2/CT.

gene	direction	primer
CCC1	fwd	GTCTTAGAGCTCGTCTTAGAGCTCATGTCCATTGT AGCACTAAAGA
	rev	GTCTTAGGATCCTAACCCAGTAACCTAACAAAGA AC
COT1	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGA AACTCGGAAGCAA
	rev	ATCCTTGTAATCACTCCACCTCCGGATCCATGAT CCTCTAACGAAATCAG
ZRC1	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGA TCACCGGTAAAGAATTG
	rev	ATCCTTGTAATCACTCCACCTCCGGATCCCAGGC AATTGGAAGTATTGC
SMF3	fwd	ATAGGAAATATTAAGCTTGGTACCGAGCTCATGC GATCTTATATGCAGATT
	rev	ATCCTTGTAATCACTCCACCTCCGGATCCAAAAT GGATGTCGGCAC

S.Table 3.5 | Primers used to clone vacuole transporters into a modified pYES2/CT vector with a LEU marker. CCC1 was first cloned via SacI and BamHI restriction enzymes, and later modified to contain a downstream flag tag instead of the V5 tag. COT1, ZRC1, and SMF3 were amplified using PCR and Gibson assembled into the modified pYES2/CT vector replacing CCC1.

(a)

gene	direction	primer
SMF1*	fwd	TCAATTACCACTGTAGAACATCTCCTATCCCTCAG TTCGAATACTTCTTC
	rev	GAAGAAGTATTGAACTGAGGGATAGGAGAGATT CTACAGTGGTAATTGA

(b)

gene	direction	primer
Δ BSD2::HIS	fwd	TGAGAATAACAAGAACACGTAGTCTAGGAAACTA AGCGCTTATTACTCTTGGCCTCCT
	rev	AAAGTTATATATCTCTTTATCATAATGAAGAAG ATGCCCTGATGCGGTATTTCT

(c)

gene	direction	primer
TaPCS1	fwd	CTAAGGGATGGAGGCTCTT
	rev	ATGGAGGTGGCGTCG
pD1235	fwd	CACCGCCGGTACAGCGACGCCACCTCCATTAT
	rev	CCGTCGAAACTAAG GAGGCATGTCAAGAGCCTCCATCCCCTAGCAGGT TAAATCATGTAATTAGTTATG
iTaPCS1	fwd	CACCCGCACGGCAGAGACCAATCAGTAAAAATCAA
	rev	CGGTTTCATTATCAATACTCGCCAT CAAGTGCACAAACAATACTT

(d)

gene	direction	primer
part1-GAL	fwd	CAGTCACGACGTTGTAAAACGACGGCCAGTAG TACGGATTAGAACGCCG
	rev	GGTTTTCTCCTTGACGT
part2-CCC1	fwd	AGTTTCGACGGATTCTAGAACTAGTGGATCCT CATGTCCATTGTAGCACTA
	rev	GCAGCTTGCAAATTAAAGC
part3-LEU	fwd	GACGCTCGAAGGCTTAATTTGCAAGCTGCAA CTGTGGGAATACTCAGG
	rev	CACGTTGAGCCATTAGTATC
part4-pUC18	fwd	TACCTCTACTTTAACGTCAAGGAGAAAAAA CCGTCATAGCTGTTCCGTGT
	rev	ACTGGCCGTCGTTTA
iCCC1	fwd	AACTGTGGGAATACTCAGGTATCGTAAGATGC AAGAGTTCAGTACGGATTAGAACGCCG
	rev	CACGTTGAGCCATTAGTATC

S.Table 3.6 | Primers used to construct the S*BCT strain. (a) Mutagenesis primers altering the K33,34 region (*AAGAAA*) into arginines (*AGGAGA*). (b) Primers used to amplify the HIS3 cassette from pRS303 with appropriate overhangs flanking the genomic BSD2 gene for deletion. (c) Primers used to construct constitutively expressing TaPCS1 in pD1235, which was then amplified using PCR and homologously recombined into the yeast genome by substituting the TRP auxotrophic marker. (d) Primers used to assemble GAL1 inducible CCC1 with a downstream LEU marker into a pUC18 backbone. Construct was then amplified with overhangs flanking the LEU auxotrophic marker and homologously recombined for genomic integration.

(a)

gene	direction	primer
<i>mTM1</i>	fwd	GCAGGTAATGAGAGATATTTGCTAAA
	rev	CAAAATGATAACAAAGTAGGGAAAATTGATT
<i>mTM6</i>	fwd	CCAAATGTTGACCACAATGGT
	rev	GTCATAATCTAAAAGCCTGGCTG

(b)

gene	direction	primer
Cd-S105C	fwd	GTCGATGCAGGTGCCTGTAATCAATTTCCTAC
Cd-T266S	fwd	CCAAATGTTGACCACAATGGTATTATTCTGCTA TTTCCATCTTAGGTGC
Cd-M276C	fwd	CTATTCCATCTTAGGTGCTACTGTTGTCCACAT TCGTTGTTTGCGGTCC
Sr-G189R	fwd	GTGCCCTTCCAGCGAGAGTGGCCATTACTGTT
Sr-S269T	fwd	GACCACAATGGTATTATACCGCTATTACTATCTT AGGTGCTACTGTTA
Sr-M276A	fwd	CTATTCCATCTTAGGTGCTACTGTTGTCCACAT TCGTTGTTTGCGGTCC
Sr-G283Q	fwd	GTTATGCCACATCGTTGTTTGCAATCCGCTTT AGTGCAGCCAA

S.Table 3.7 | Primers used for library generation and site-directed mutagenesis of SMF1*. (a) Primers used to error-prone PCR regions TM1 and 6 of SMF1 to construct libraries using Agilent's GeneMorph(II) protocol. (b) Site-directed mutagenesis primers used to selectively mutate residues identified to have effects on SMF1 metal preference and uptake.

Chapter 4

Engineering supramolecular forming proteins to chelate heavy metals for waste water remediation

Abstract

Physicochemical technologies have dominated water treatment methods; however, impact on reducing contaminated waters has lagged behind the growing accumulation of heavy metal waste. Ion-exchange is one such method which has gained traction in developed countries such as the United States, but remains difficult to deploy in developing nations which are in more need of economical and sustainable remediation technologies. Unfortunately, ion-exchange requires sophisticated design strategies, advanced chemical synthesis techniques, and overall has a high cost and technical barrier for adoption. But with a closer look, the main mechanism of ion-exchange is straightforward and operates through reversible metal chelation, a reaction found almost everywhere in biological systems such as proteins and enzymes. Herein, this work shows that proteins from the CTP synthase and glutamine synthetase family, recently discovered to have polymerization properties, can act as metal chelating agents for water purification applications. Specifically, proteins pyrG and glnA show

metal-inducible polymerization behavior in which they aggregate into solid matrixes when exposed to metals. Their surfaces can be decorated with either a 6xHis tag, metallothionein, or calmodulin fusions to tailor removal of copper, cadmium, mercury, lead, and even calcium. These proteins were further modified to harbor binding tags to allow anchorage onto denser substrates such as agarose, magnetic beads, or yeast to increase metal-protein sedimentation rates. With this system, modified pyrG and glnA showed more than 50% removal of 1 millimolar amounts of metal (more than 10–100 times higher than EPA thresholds for dangerous metal content) and more than 80% recovery of metals removed. The ability for proteins to act as ion-exchange like agents opens the potential for robust heavy metal removal and possible downstream collection and/or recycling.

4.1 Introduction

One of the most common methods for wastewater remediation is through ion-exchange. Ion-exchange is a well-established technology with high removal efficiency and fast removal kinetics [**fu2011removal**, **ustun2007regeneration**, **motsi2009adsorption**, **kumargupta2012**]. When broken down to its fundamental mechanism, ion-exchange uses beads (also known as resins) that are functionalized with multiple copies of commonly used metal binding functional groups (e.g. sulfonic, $-SO_3H$, and carboxylic acids, $-COOH$) [**fu2011removal**]. Unfortunately, what limits the wide-spread adoption of ion-exchange for global water treatment is the cost, handling, and production of ion-exchange resins [**ustun2007regeneration**, **amini2015**]. Resins are designed synthetically, and require rounds of chemical synthesis for testing and production. More so, the design of exchangers is non-trivial, as custom software such as Dow Water's Computer Assisted Design for Ion Exchange (CADIX), Purolite Pure DesignTM, and other programs are sometimes required to aid in the design strategy and synthesis process [**wachinski2016**]. What this design flow suggest is creating, testing, and manufacturing resins is arduous and requires technical skill. More so, during synthesis many of the chemical processes are prone to generating

secondary-waste due to reaction by-products or effluent from regeneration of resins when preparing or after use [**kumargupta2012**, **amin2015**].

Looking again at the fundamental components of ion-exchange there are two main groups to engineer, the resin, and the identity and composition of the metal binding group. The resin provides the physical medium in which exchangers can be packed and stored as a solid unit, while the metal binders perform the metal chelation and water purification. Both are key and mutually dependent. Without the resin, chelated metals would remain free-floating. Without the chelators, resins serve no remediation function. But both components do not have to be synthetic, rather, natural biological systems have already developed a mirror-image approach for handling heavy metal waste. One such example is the removal of metals through chelation by polymerized chains of phytochelatins, or binding onto metallothionein proteins [**clemens2002**, **cobbett2002phytochelatins**]. In addition, the functional groups of strong and weak exchangers (e.g. $-S^{2-}$, $-NH_2$, and $-COOH^-$) are regularly found, and often modifiable, in most peptides/proteins. This approach has been recognized by scientists in the past, as some have attempted to mimic ion-exchange by functionalizing cell surfaces with metal binding peptides and proteins. The cell acts as the physical anchor while the peptides serve as the exchanger. One of the first attempts was in 1996 by Sousa et al. who used bacteria display to anchor poly-his peptides to accumulate metals on the cell surface [**sousa1996**]. Thereon, other works showed similar results by functionalizing bacterial or yeast surfaces with metal-specific proteins such as mercuric reductase or metallothioneins [**bae2003**, **kuroda2003**, **pazirandeh1998**]. Compared to synthetic manufacturing of resins, biology provides an easier platform to produce resins (i.e. cells) and production and engineerability of exchanges (i.e. peptide/proteins). Given that almost a third of proteins harbor some metal binding domain [**waldron2009bacterial**], the diversity and richness of discovering unique protein-metal chelators is enticing, and the advances in genetic and protein engineering make it possible to further augment protein-metal binding interactions for controlled water remediation applications.

However, past results in biologically mimicked ion-exchange is not yet able to com-

pete with synthetic ion-exchange with respects to metal removal capacities. Results for cell surface display metal removal have capture capacities of tens of nanomolar of metal per gram of cell dry weight [**sousa1996, kuroda2003, ruta2017**], whereas ion-exchange have typical capture capacities of millimolar of metal per gram dry weight, almost 3–6 orders of magnitude more [**p.stathi2013, barakat2011new**]. Being that metal removal is purely a numbers game, the limitation in biologically mimicked ion-exchange is the low number of peptide/protein expression per cell per volume. For example, typical values of protein expression for yeast display range from 1 to 100 thousand [**kieke1999, rutherford2006**]. Given modest estimates of yeast culture density (assuming 1 OD₆₀₀ per mL) an upper bound of metal removal assuming each protein has 1 completely occupied binding site is roughly 2–200 nanomolar, an almost insignificant amount when dealing with micro to millimolar levels of metal contaminants.

In order to capitalize on the benefits of using biology, e.g. diversity, engineerability, scale, this work focuses on using proteins recently discovered to have polymerization properties for heavy metal chelation and sedimentation. Using proteins that can polymerize increase protein quantity per volume, and relying on proteins rather than an entire cell for metal capture would increase capture capacities per dry weight. Furthermore, proteins could be modified using facile genetic and protein engineering techniques to express metal binding moieties on their surfaces. Thus when polymerized the supramolecular complex can be decorated with multiple copies of metal chelators that are then pulled down during aggregation. Ultimately the proteins themselves, once aggregated, act as the resin and the fused metal binding moieties act as the exchangers. The two proteins investigated herein are bacterial pyrG from the CTP synthase (CS) family, and glnA from the glutamine synthetase (GS) family. The existence of these proteins have been known for decades, however it was not until recently that CS was observed to form cytoplasmic filaments in a conserved fashion among a variety of organisms [**robertson1995, ingerson-mahar2010, liu2011, carcamo2011, liu2010**]. This peculiar behavior has been associated with subcellular protein organization for enzyme regulation [**ingerson-mahar2010, carcamo2011**,

liu2010]. Given these observations, a multitude of other proteins have been screened and discovered to form similar large cytoplasmic supramolecular complexes in *E. coli*, *S. cerevisiae*, and *D. melanogaster* [**liu2010, narayanaswamy2009, noree2010**]. In these screens another interesting protein family, GS, was determined to also form supramolecular complexes, and some of the first recorded observations were cited as early as the 1970s [**noree2010, eisenberg2000, valentine1968, miller1974**]. In the context of metal removal, the particular enzymatic function, and even the cellular role of these proteins, is irrelevant; the main interest is the fact that both CS and GS have controllable means to initiate polymerization and metal binding, and the focus of this work is aimed to capitalize on this mechanisms for a direct application in heavy metal waste removal.

4.2 Results

4.2.1 Use of CS and GS for toxic metal removal

Representative crystal structures for pyrG (accession number #P0A7E5) and glnA (#P0A9C5) (PDB 1S1M and 1FPY, respectively) reveal that these proteins are monomers of a larger symmetric structure (Figure ??a,b). Past studies have shown that pyrG form monomers, dimers, and tetramers [**liu2011**] whereas glnA often forms 2 hexagonal structures stacked on top of one another to form a 12-mer structure [**valentine1968**]. Genes for pyrG and glnA were isolated from the genome of a DH5 α *E. coli* strain and inserted into IPTG inducible pET28c(+) vector. Confirmed sequences were transformed into BL21(DE3) high expression strain, induced, lysed, and purified with Ni-NTA columns (Supplemental Figure ??) and protein concentrations quantified using calibrated Nanodrop readings (Supplemental Figure ??). Final protein samples and experimental reactions were performed in water. Past studies have shown that glnA aggregates in the presence of Mn(II) [**valentine1968, miller1974**], however a surprising observation is that both pyrG and glnA aggregate in response to a variety of other divalent metals such as Co(II), Ni(II), Cu(II), and

Zn(II) (future reference to metal elements are assumed to be divalent from hereon; Figure ??c; Supplemental Figure ??). At 100 μ M pyrG or glnA added to 1 mM of Mn, Co, Ni, Cu, Zn, Cd, Hg, or Pb visibly turned cloudy and eventually formed aggregates which could be pelleted. To test whether 1 mM of metal was causing protein denaturation or non-specific aggregation, a control of 100 μ M BSA was added to 1 mM of the same metals, and no visible opaqueness formed, except for Pb which naturally formed hydroxides over time (Supplemental Figure ??). Alkaline and alkaline earth metals (Na, Mg, etc.) were not observed to induce aggregation.

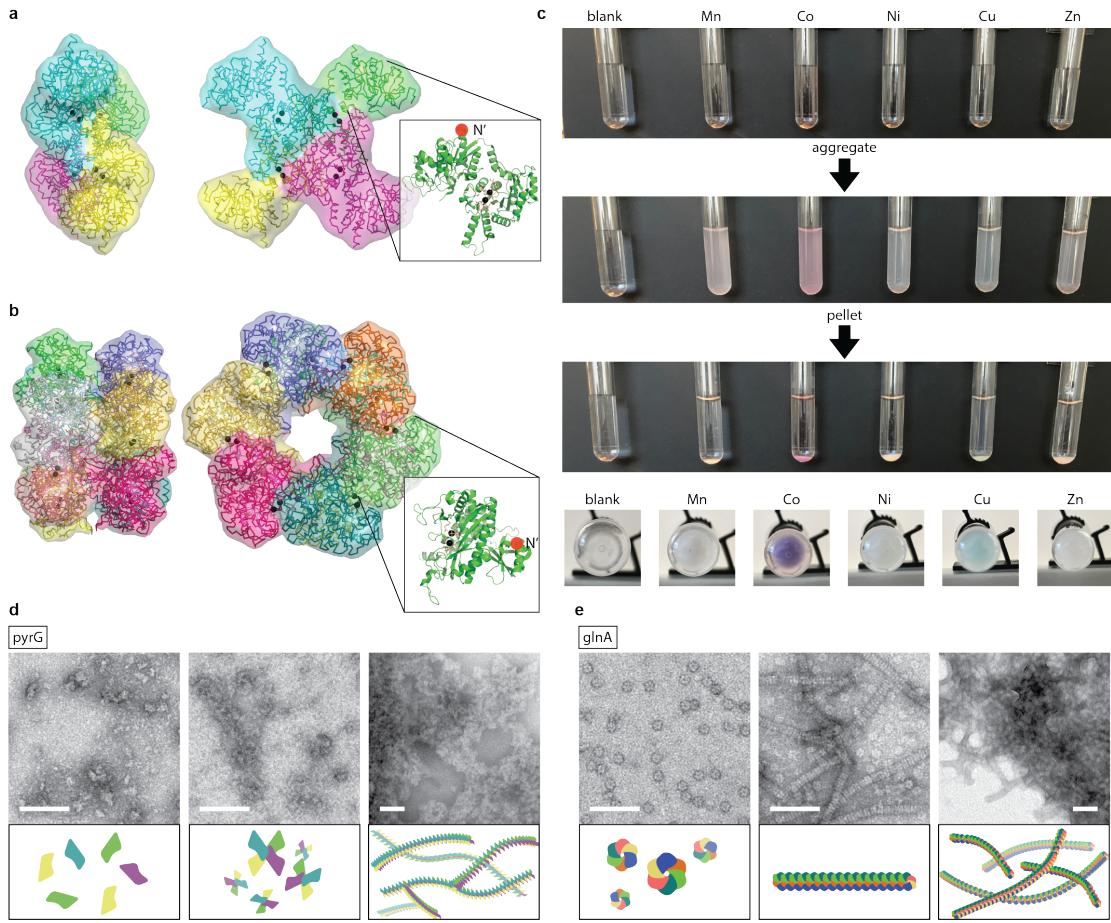


Figure 4.1 | Using pyrG and glnA as aggregating metal binders for heavy metal removal. (a) pyrG representative crystal structure 1S1M is shown, which contains 4 identical monomers shaped into an ‘X’. Each monomer contains 2 divalent metal binding pockets. The N'-terminus is highlighted to show the region in which metal binding appendages can be attached to further enhance metal binding. (b) A similar structure is shown for glnA using representative crystal structure 1FPY modeled as two stacked hexagons. Each monomer has two metal binding pockets, and N'-terminus is highlighted. (c) Visual representation of metal binding and aggregation of glnA. Metal and protein concentration were at 10 mM and 500 μ M respectively. (d) HRTEM images of pyrG and glnA, respectively at different levels of aggregation in the presence of Zn. Illustrations below the hypothesized structure and formation of these aggregating chains. Scale bars represent 50 nm.

The aggregating tendencies of pyrG can be observed visually; pyrG forms spontaneous aggregates at room temperature, and further examination under transmission electron microscopy (TEM) show haphazard connections rather than its canonical

tetrameric structure. Under high resolution TEM (HRTEM), pyrG was observed to not regularly form its tetrameric structures, but instead bundle as monomers. In the presence of metals pyrG forms a combination of tetrameric but more often irregular aggregated structures, and at the extreme aggregates into amorphous supramolecular complexes (Figure ??d; Supplemental Figure ??a). On the other hand, purified glnA examined under HRTEM forms uniform dodecamer structures. Upon addition of metal these structures aggregate into rods, with each hexagon face stacking on top of one another. At the extreme these stacked rods aggregate into bundles to form similar amorphous supramolecular complexes like pyrG (Figure ??e; Supplemental Figure ??b). This study utilizes the metal responsiveness of pyrG and glnA to chelate heavy metals away from waters for bioremediation applications.

4.2.2 Analysis of metal-induced aggregation of pyrG and glnA

pyrG and glnA metal interactions have been shown to modulate their enzyme activity and are shown in this study to induce protein aggregation [**valentine1968, robertson1993, eisenberg2000**]. pyrG and glnA contain 2 metal binding domains per monomer, with binding constants for Mn, Co, and Zn in the low or sub-micromolar range [**robertson1993, denton1969**]. Some researchers have observed that increasing metal concentrations beyond a millimolar retard enzymatic activity [**robertson1993, eisenberg2000**]. In this study, both pyrG and glnA were observed to form aggregates at 1 mM for Mn, Co, Ni, Cu, Zn, Cd, Hg, and Pb.

To quantitatively measure aggregation intensity without having to repeatedly examine structures under TEM, aggregates were measured spectroscopically at 350 nm, a wavelength that produced the highest signal to noise ratio of aggregates to non-aggregated protein and showed a linear dependence on aggregation intensity in solution (Supplemental Figure ??). pyrG and glnA titrated with various concentrations of metals show metal-dependent aggregation following typical binding kinetics. Examined under TEM, the density of the protein network increased with increasing metal concentrations (Figure ??a). The study was performed again by quantifying aggregates at 350 nm to model a binding (in this case, aggregation) curve. Absorbance

readings were normalized and plotted against metal concentrations to form a titration curve fitted with a Hills function. The magnitude of aggregation (A), the metal concentration at which aggregation was half-max $A/2$ (K_D), and Hill coefficient (n) were fitted and tabulated (Table ??)

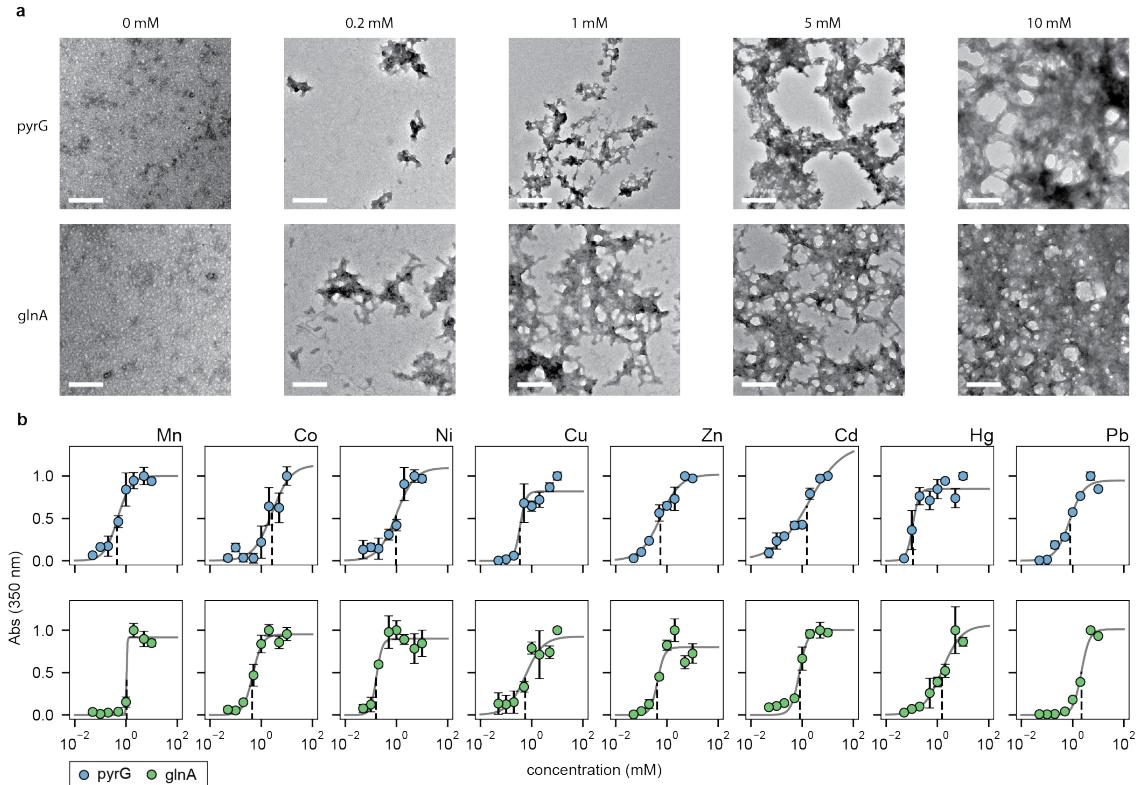


Figure 4.2 | Measuring metal induced aggregation responsiveness and intensity for pyrG and glnA. (a) Top/bottom rows are pyrG and glnA respectively. Columns represent concentrations of Zn which was used to induce aggregation, with progressively higher concentrations leading to higher levels of aggregation. Scale bars represent 200 nm for all images. (b) pyrG and glnA were titrated with metals to measure the level of aggregation as a function of metal concentration. Aggregation was quantitatively measured using 350 absorbance readings (Supplemental Figure ??). For all data, the mean \pm s.d. of three replicates are shown.

The maximum aggregation intensity at 10 mM metal for pyrG followed Pb Cd Cu > Hg > Zn Ni > Co > Mn, whereas for glnA was Pb > Mn Cu Cd > Zn > Co > Ni > Hg. Aggregation sensitivity ranked by K_D (smallest value to highest) for pyrG followed Hg > Cu > Mn Zn > Pb > Cd Ni Co whereas for glnA was Ni >

$\text{Co} > \text{Zn} > \text{Cu} > \text{Cd} > \text{Mn} > \text{Hg} > \text{Pb}$ (Figure ??b; Table ??). The Hill coefficient for pyrG and glnA for most metals exceeded 1, suggesting positive cooperativity for aggregate formation.

Metal aggregation was also found to be reversible. Different dilutions of EDTA was added to samples aggregated with 1 mM of metal. Examination under TEM showed that 1–10 mM of EDTA reversed aggregation and released pyrG and glnA into smaller aggregates or individual monomers (Supplemental Figure ??a). Examining these effects spectroscopically at 350 nm, a steep reduction in absorbance readings was observed after 1 mM, or at an equal molar ratio of metal to EDTA (Supplemental Figure ??b). This finding suggest that pyrG and glnA require some type of metal binding to induce aggregation, and aggregation is reversible upon metal removal or competition.

	Mn	Co	Ni	Cu	Zn	Cd	Hg	Pb
A	0.10 ± 0.01	0.15 ± 0.07	0.20 ± 0.08	0.37 ± 0.04	0.20 ± 0.01	0.39 ± 0.05	0.24 ± 0.01	0.38 ± 0.01
pyrG	K_D	0.5 ± 0.01	4.66 ± 0.43	3.7 ± 0.4	0.42 ± 0.2	0.57 ± 0.03	$2.44 \pm .12$	0.13 ± 0.05
n		1.58 ± 0.21	12.43 ± 1.6	1.67 ± 0.87	11.06 ± 11.4	1.1 ± 0.06	0.66 ± 0.15	4.31 ± 0.20
A	0.22 ± 0.01	0.15 ± 0.01	0.12 ± 0.01	0.21 ± 0.03	0.16 ± 0.02	0.18 ± 0.01	0.08 ± 0.01	1.37 ± 0.01
glnA	K_D	1.05 ± 0.02	0.48 ± 0.06	0.17 ± 0.01	0.61 ± 0.14	0.42 ± 0.05	0.8 ± 0.09	2.11 ± 0.86
n		32.86 ± 3.15	$3.32 \pm .18$	$4.57 \pm .18$	9.93 ± 1.3	3.25 ± 0.5	2.96 ± 0.76	1.33 ± 0.6

Table 4.1 | Values for maximum aggregation intensity (A ; measured at 350 nm), K_D , and fitted cooperativity coefficient (n) for pyrG and glnA for the various metals studied. Coefficients were fitted from data collected in Figure ??b.

4.2.3 Metal removal and sequestration via protein-metal aggregation

Beyond understanding the aggregating behaviors of pyrG and glnA, the main objective of this work was to chelate and sediment metals upon induced aggregation. To increase binding capacities per pyrG and glnA complex, their surfaces were deco-

rated with metal binding domains. The exposed N'-terminus was used to fuse several metal binding domains (Figure ??a,b), namely the 6xHis tag, a well-known metal binding motif [knecht2009]. Further mention of pyrG or glnA have an appended N'-terminus 6xHis tag, unless otherwise noted to be cleaved to measure un-modified pyrG and glnA as controls. All experiments used 100 μ MM of protein and 1 mM metal for aggregation and metal removal studies.

Two mutually dependent factors contribute to the level of metal removal. The first is the degree of metal-induced aggregation which physically assembles the protein complex and allow it to sediment out of solution. The second is the degree of metal binding to the protein and metal binding domain, in this case the 6xHis tag. Without the former, irregardless of the strength of metal binding the protein-metal complex will remain in solution if it does not aggregate and sediment. On the other hand, if there was no metal binding no amount of aggregation will chelate the metal away. Therefore, pyrG and glnA have different metal removal profiles when analyzing removal of Mn, Co, Ni, Cu, Zn, Cd, Hg, and Pb (Figure ??a,b). For example, pyrG does not aggregate strongly with Mn whereas glnA does, therefore metal removal for Mn is almost 4 times higher for glnA ($p < .05$); however, the opposite is true for Hg where aggregation is more pronounced for pyrG, corresponding to 4 times more removal when compared to glnA ($p < .05$). Compared to controls of pyrG and glnA without a 6xHis tag, metal removal is significantly reduced for Ni, Cu, Zn, Cd, Hg and Pb ($p < .01$; except for glnA with Hg, $p < .05$), confirming that the fused metal binding domain increases metal binding capacity and removal. The most amount of metal removed for pyrG were with Cu ($463 \pm 34 \mu\text{M}$), Zn ($413 \pm 13 \mu\text{M}$), and Hg ($670 \pm 33 \mu\text{M}$) in the presence of 1 mM metal. Whereas for glnA were with Zn ($699 \pm 25 \mu\text{M}$), Cd ($373 \pm 28 \mu\text{M}$), and Pb ($433 \pm 51 \mu\text{M}$) in the presence of 1 mM metal. Therefore, despite the same 6xHis tag, metal removal was heavily dependent on the degree of aggregation per metal.

pyrG and glnA were also tested in mixed metal conditions. Mn, Co, Ni, Cu, Zn, Cd, Hg, and Pb were mixed at equal molar to a final metal concentration of 1 mM ($125 \mu\text{M}$ per metal) and mixed with 100 μM pyrG and glnA. The amount of metal

removed for both were approximately >33% (Figure ??c,d). For *pyrG*, the individual metal makeup of what was removed favored Cu, Zn, Cd, Hg, and Pb. A similar analysis for *glnA* showed a similar composition of metal removed from the mixed source (Figure ??c,d; Supplemental Figure ??).

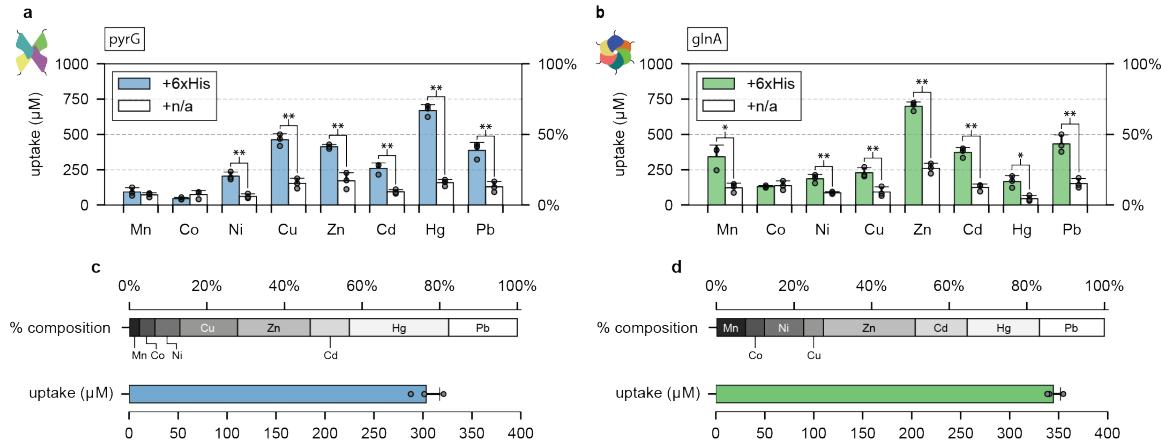


Figure 4.3 | Metal removal of Mn, Co, Ni, Cu, Zn, Cd, Hg, and Pb for *pyrG* and *glnA* modified with a 6xHis tag. (a, b) Removal of added 1 mM metals were individually measured for *pyrG* or *glnA*, respectively. Controls (white bars) represent *pyrG* and *glnA* cleaved of its 6xHis tag fusion and measured for metal removal. (c, d) Metal removal of a metal mixture (totaling 1 mM, or 125 μM each) were measured for *pyrG* or *glnA*, respectively. Top bar represents the percent composition of the metals removed, whereas the bottom bar represents the total amount of metal removed. For all data, the mean \pm s.d. of three replicates are shown. * = $p < .05$, and ** = $p < .01$ represent significant metal removal compared to controls.

Even though *pyrG* and *glnA* have different metal aggregating responses, the presence of all 8 metals strongly encouraged *pyrG* and *glnA* aggregation as the weakest and strongest metal-inducers were present. Therefore aggregation was no longer an issue and the responsibility for metal removal was now determined by the 6xHis tag. Therefore, the metal removal composition profile appear roughly the same for *pyrG* and *glnA* because of the same 6xHis metal-binding domain. To better illustrate this takeaway: *glnA* has a relatively weak aggregating effect with Hg, despite the 6xHis tag having a relatively strong affinity for Hg (Figure ??b). However, if the interest is to remove Hg, then *glnA* can be co-mixed with Zn, a metal that induces a strong

aggregating effect in glnA. The outcome is an increase in Hg removal by encouraging aggregation with another metal, Zn, allowing the 6xHis tag to now pull down both Hg and Zn (Supplemental Figure ??). The results shown in Figure ??c,d is an extension of this example, where the composition of metal removed is more dependent on the 6xHis tag.

Heavy metal removal is ultimately a numbers game, and to understand this metal removal system pyrG and glnA's stoichiometry's were analyzed to theoretically predict upper and lower bounds of metal removal. pyrG and glnA have two native metal binding sites per monomer (Figure ??a,b), and the addition of the 6xHis tag now increases the stoichiometry to at least 3. However, the calculated stoichiometry of pyrG and glnA given the metal removal results (Figure ??) give ratios between 3–8 (depending on the metal), higher than expected. Several explanations can be made: the 6xHis tag could bind to multiple metals, aggregation of proteins may encourage allosteric binding due to the proximity of multiple 6xHis tags leading to positive cooperativity, or the act of aggregation opens other metal binding pockets not observed in the monomer. Data without the 6xHis tag showed removal ratios of almost 1:1, meaning at least one of the binding pockets were occupied, a finding that supports past results for pyrG and glnA binding kinetics [robertson1993, denton1969]. So the increase in metal removal with the addition of the 6xHis tag suggest that the highly electronegative surface of the protein aggregate may positively encourage metal binding to a greater amount than compared to the predicted one 6xHis tag to one metal assumption. Therefore, modifications to the protein surface is an important engineering parameter to design for in order to both enhance metal removal specificity and metal removal capacity.

4.2.4 Tuning metal removal profiles through selection of new metal-binding domains

Aside from the 6xHis tag, there are many other metal-binding peptides/proteins that can serve as metal-binding accomplices with pyrG or glnA. The remainder of this

study focuses on glnA because of its better stability, more defined macrostructure as examined under TEM, and the higher metal-protein binding stoichiometry per macromolecule compared to pyrG (12 compared to 4).

A family of metal-binding proteins used in a variety of cells for metal detoxification are metallothioneins (MTs). MTs are cysteine-rich, low molecular weight proteins with high metal affinity for common toxic metals such as copper, cadmium, and mercury [cobbett2002phytochelatins]. Many of the isoforms of MTs have been studied in plants, and because *E. coli* do not natively have MTs, the MT1A (#P43392) from *A. thaliana* was codon-optimized and fused to the N'-terminus of glnA which also replaced the 6xHis tag. Performing the identical metal removal experiment as before, the profile of metals removed skewed more towards Cu, Zn, Cd, Hg. The removal of Cu, Zn, Cd, and Hg increased by $406 \pm 92 \mu\text{M}$, $110 \pm 13 \mu\text{M}$, $169 \pm 62 \mu\text{M}$, and $147 \pm 53 \mu\text{M}$, respectively, in comparison to using the 6xHis tag (Figure ??a). In a multi-metal removal experiment the enhanced metal removal is more apparent, with total metal removal exceeding 50% and a large composition of the metal removed being Cu, Zn, Cd and Hg (Figure ??b).

A similar experiment can be performed for alkaline-earth metals such as calcium by exchanging the 6xHis/MT1 fusion for an alkaline-earth sensitive binder. Although elements such as magnesium and calcium are not dangerously harmful when compared to cadmium or mercury, they are still harmful for remediation processes as these metals often precipitate which erode infrastructure, calcify around sensitive plumbing, and add more background metals that can decrease remediation efficiency processes like ion-exchange [greenleaf2006]. A commonly known and used protein for calcium binding, calmodulin (CaM), was used to test whether the metal removal profile could be shifted towards the alkaline-earth metals. Yeast CaM (#P06787) was isolated from the genome of *S. Cerevisiae* W303 α strains and cloned into glnA to replace the N'-terminus 6xHis tag. Unfortunately, the alkaline-earth metals do not induce aggregation, therefore 1 mM Zn was co-mixed with either 1 mM Mg, Ca, Sr, or Ba to promote sedimentation and metal removal. Replacing 6xHis for CaM showed increased calcium removal of $226 \pm 51 \mu\text{M}$ compared to $66 \pm 8 \mu\text{M}$ of glnA with 6xHis

($p < .01$), and a slight increase in Sr removal of $116 \pm 30 \mu\text{M}$ compared to $54 \pm 7.8 \mu\text{M}$ of glnA with 6xHis ($p < .05$) (Figure ??c). The metal removal profile loosely corresponded to CaM native preference for Ca, followed by Mg and Sr [haiech1981].

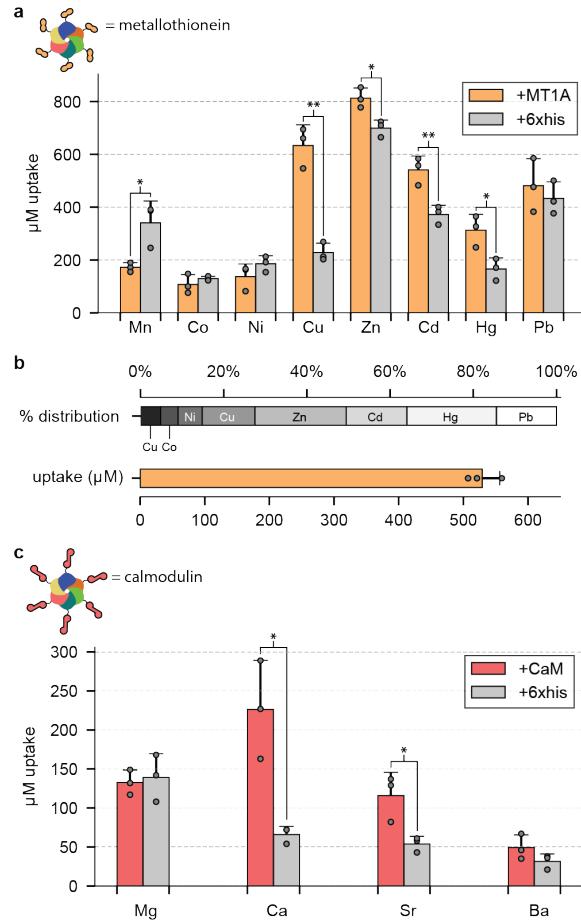


Figure 4.4 | Substituting the 6xHis tag with plant MT1A or yeast calmodulin alters metal binding preference and metal removal capacity. (a) The same metal removal experiment was performed for glnA+MT1 and compared against glnA+6xHis. Individual metal removal preference skewed more towards Cu, Zn, Cd, and Hg ($p < .05$). (b) Likewise, the same metal mixture removal experiment was performed, where the top bar represents the percent composition of removed metals, and the bottom bar represents total metal removed. (c) glnA+CaM were aggregated with Zn and tested for Mg, Ca, Sr, and Ba removal. For all data, the mean \pm s.d. of three replicates are shown. * = $p < .05$, and ** = $p < .01$ represent significant metal removal compared to controls.

4.2.5 Improving sedimentation of metal-protein complexes and metal recovery

The metal-induced precipitation of glnA offers a natural self-filtering mechanism via protein aggregation and sedimentation. However, rather than waiting several hours, or forcibly pelleting aggregates using a centrifuge, sedimentation of glnA can be enhanced by binding aggregated complexes to denser anchors such as beads, or even cells. The effect on sedimentation was tested by binding aggregated glnA with agarose or magnetic beads. In addition, a fully biological route was to display glnA monomers via yeast display and allowing aggregates to grow and anchor onto the yeast cell surface. For bead attachment, glnA was modified to include expression of a 3xFlag tag at the N'-terminus in conjunction with the 6xHis tag. Agarose or magnetic beads were functionalized with anti-flag antibodies to recognize the 3xFlag tag and pull down the aggregates. For the yeast display condition, EBY100 strains were transformed with the pYD1 yeast display vector expressing a glnA monomer (no appended His tag) (Supplemental Figure ??). glnA aggregation was performed together with yeast displaying cells such that aggregates could nucleate onto the yeast cell surface. In this setup the 3xFlag tag was not added to the glnA protein.

Measurements of sedimentation was determined by measuring protein concentration at the liquid's mid-height. When placed in cuvettes at 4 mL (mid-height being at 2 mL), 100 μ M glnA aggregated with 1 mM of Cd took several hours to sediment. At roughly 5 hours almost half of the aggregated content was beneath 50% of the sampled liquid height (Figure ??a). When aggregates were mixed with agarose or magnetic beads, sedimentation was much quicker. Using agarose beads, it took 1 minute for 50% of the aggregates to sediment beneath the sampled liquid height. Whereas, with magnetic beads and using an external magnet for pull down sedimentation was much more dramatic, only requiring < 5 seconds for the same amount of protein removal (Figure ??a,b). Alternatively, glnA and yeast displaying glnA monomers were co-precipitated for 1 hour with 1 mM Cd to allow aggregates to bind onto the yeast surface.

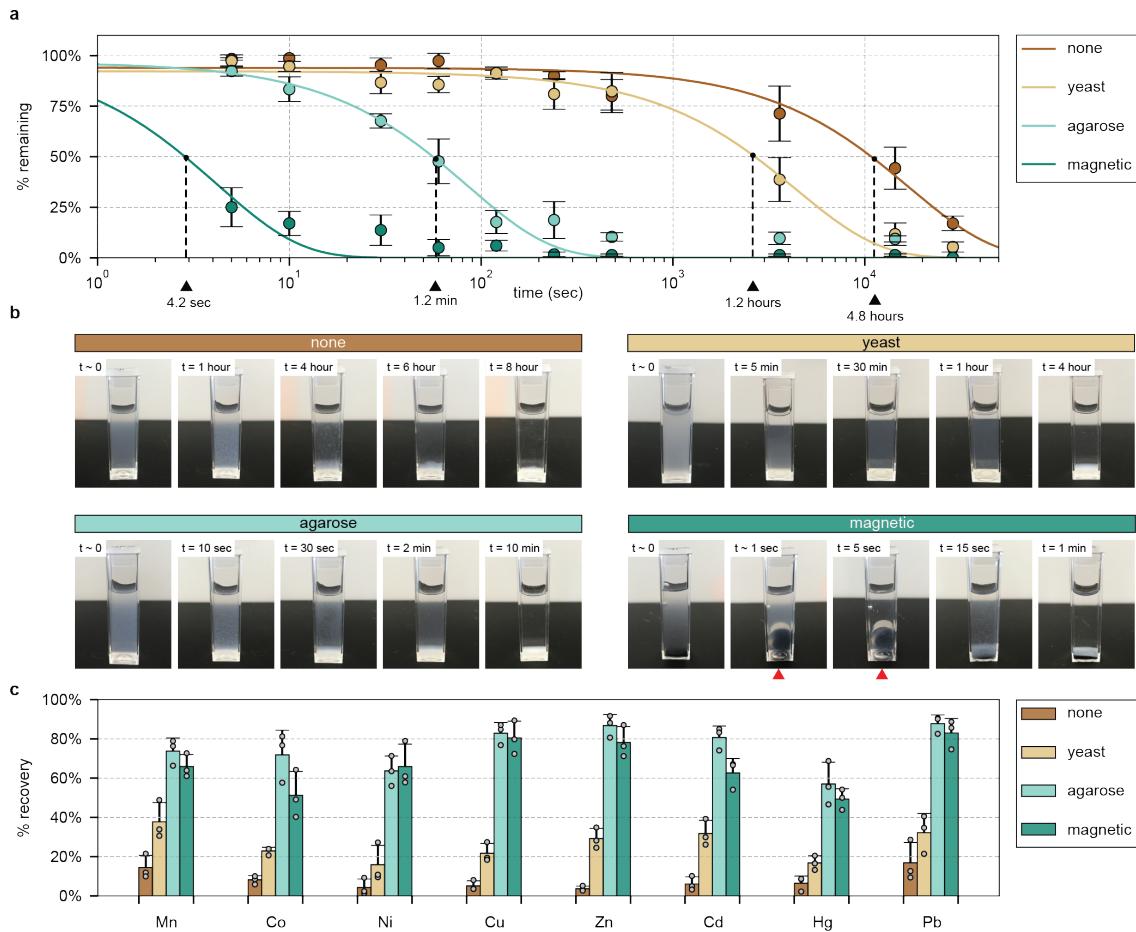


Figure 4.5 | glnA fused with a 3XFlag tag, or bound to yeast displaying glnA monomers, improved sedimentation rates and metal recovery. (a) glnA+3XFlag was aggregated with 1 mM Cd and later mixed with anti-flag agarose or anti-flag magnetic beads. Alternatively, glnA was mixed with yeast displaying glnA monomers. Magnetic beads had the quickest pull down of aggregates, followed by agarose beads, then yeast. (b) Visual representation of protein-metal complex sedimentation with magnetic, agarose beads, and yeast. Red arrows for the magnetic bead condition indicate moments when an external magnet was applied. (c) Mixtures were allowed to sediment for 10 minutes in which the sedimented pellet was isolated, washed, and extracted for metals using EDTA. Bars represent the percent metals recovered relative to the amount of metals removed by the protein aggregates. For all data, the mean \pm s.d. of three replicates are shown.

Metal recovery was also measured using the various sedimentation strategies. To account for the maximum amount of metal recoverable, the amount of metal bound to the glnA complex was first measured. Afterwards, complexes were then resuspended

and mixed with agarose or magnetic beads. For the yeast display condition, glnA and yeast were mixed during metal induced aggregation in order to allow proteins to aggregate onto the yeast surface. After incubation, the protein-metal bead/yeast complex were stirred and allowed to sediment. Samples were allowed to sit for 10 minutes after which the supernatant was fully removed and whatever sedimented was recovered. The sediment was transferred and washed 2X in ddH₂O before resuspending in 10 mM EDTA. The amount of freed metals were measured and compared against the initial measurement of metals bound to the aggregate before the sedimentation experiment. This value gave a metal percent recovery value in the 10 minute span allowed for sedimentation.

Factors that affected recovery were the rate of sedimentation within the 10 minute timeframe, and the binding strength between the aggregates and anchors during the wash steps. Without any anchor, the recovery for just glnA was below 10% for all metals (Figure ??c). For yeast it was observed that flakes of aggregates would dissociate during the wash steps; this was not seen for the agarose or magnetic bead samples. Aggregates with yeast displayed monomers had recoveries between 15–40%. When using beads, the agarose and magnetic anchors were quicker to sediment thereby recovering more aggregates which corresponded to a higher percent recovery of Cu, Zn, Cd and Pb. Metal recoveries ranged between 60–80% for the agarose and magnetic bead conditions. The high recoveries could also be associated to the tight binding of the 3XFlag tag between glnA and the functionalized antibodies on the beads. A takeaway from these experiments is that not only can glnA be modified to increase metal removal and metal removal preferences using appended metal binding domains, but further modifications can lead to other functionality such as attachment to beads or cells for improved sedimentation and recovery.

4.3 Discussion

Biological enzymes previously discovered to have unique structural and enzymatic roles in cells can be re-purposed for bioremediation and water cleaning strategies.

The appeal of using such proteins is their controllable aggregating behavior and the subsequent solid sediment that forms allowing for physical handling and removal. This study shows that pyrG and glnA are responsive to a variety of divalent metals such as Mn, Co, Ni, Cu, Zn, Cd, Hg, and Pb and aggregate within the submillimolar to millimolar range. In addition, pyrG and glnA surfaces are easily modified and appended with metal chelating agents such that metal removal can be physically captured in the aggregated sediment. In this perspective, pyrG and glnA act as vehicles in which their surfaces can be genetically modified to introduce metal chelating agents for heavy metal clean up. For example, adding a 6xHis tag to pyrG and glnA has broad metal specificity for most metals tested, and increased metal removal for Cu, Zn, Cd, Hg and Pb compared to native pyrG and glnA. The 6xHis tag can be substituted for other small peptide/protein metal binders such as a metallothionein or calmodulin which alter metal removal profiles. For example, the metallothionein MT1A from *A. thaliana* increased overall metal removal for Cu, Zn, Cd, and Hg compared to the 6xHis tag. More so, metal preferences were entirely altered from transition metals to alkaline-earth metals using calmodulin from yeast. Appending calmodulin showed preferential calcium removal, whereas the 6xHistag was insensitive. The ability to engineer metal removal preferences from heavy metals to alkaline metals provides another avenue for water cleaning which is for water softening, a frequent problem in municipal drinking waters and damage in water-related infrastructure.

Protein anchor glnA can be further modified to contain binding tags that can adhere to denser anchors such as agarose or magnetic beads for faster sedimentation. Likewise, a fully biological method used yeast display to anchor aggregates onto the yeast surface. With either beads or yeast the time for sedimentation reduced from hours (glnA alone) to minutes (agarose) or even seconds (magnetic beads). Once sedimented, the pull-downed aggregate-metal complex could be dissociated to release the bound metal for recovery purposes. Overall, improved sedimentation due to agarose and magnetic beads correlated to improve metal recovery.

Different metal binding domains, repeat of domains, or combinations of several binders on a single protein can potentially improve protein-based water removal ca-

pacities and tailor metal removal preferences. For example, multiple repeats of the 6xHis tag can be fused to glnA to potentially increase the number of metal binders, and hence improve removal capacities. Alternatively, several proteins with different metal binding preferences can be fused on one chain to customize metal removal profiles. For example, fusing MT1A with CaM onto glnA can potentially remove both transition and alkaline-earth metals. In addition, a variety of tags can be added for additional functionality; in this work, they were denser beads to improve sedimentation rates. However, future possibilities could be to anchor pyrG or glnA on other platforms using well-established protein binding strategies (Flag, Streptavidin, SpyTag, etc.). Several interesting platforms would be biofilms, hydrogels, or other semi-porous matrixes that can harbor aggregating proteins like pyrG or glnA. More possibilities not mentioned (or thought of) here can be easily designed by appropriately modifying the surfaces of pyrG or glnA and combining it with the platform of choice. What may limit the length or complexity of appendages would be a disruption to pyrG or glnA protein expression and folding. However, this concern of overburdening the protein is true for most protein-fusion expression systems and requires fine tuning of expression conditions and purification protocol.

The identity of the protein anchor is not as important as its ability to aggregate and sediment out of solution. There has been much work in the biological community to identify the growing number of enzymes that behave like pyrG and glnA, and so far a list of 33 proteins discovered in yeast may have similar properties [**narayanaswamy2009**], and many more may exist in other organisms [**liu2010, carcamo2011**]. Although not exhaustively tested here, there is a possibility that these identified proteins behave differently in either responsiveness or sensitivity towards aggregation, and these behaviors could be uniquely capitalized for different heavy metal removal conditions. More so, protein stability, stability with respects to surface modifications, and controlled aggregation are desired behaviors of the host protein. Future steps would be to understand mechanistically and biochemically how aggregation can be induced and controlled in this set of discovered proteins. If these underlying mechanisms can be uncovered, more engineering can be done to improve

the aggregation sensitivity and density of aggregates for improved metal removal, sedimentation, and recovery.

The goal of this work was to biologically mimic ion-exchange by recapitulating its underlying mechanism in proteins. The first was to use aggregating proteins to physically mimic the resin bed. The second was to fuse metal binding peptides/proteins onto these aggregates for targeted metal removal, a process similar to functionalizing resins with strong and weak metal exchangers. However, unlike ion-exchange these proteins are autonomously produced in cells and do not require complex chemical synthesis normally used when creating resins. In addition, protein modifications are becoming increasingly easier to perform on the genetic level given the maturing technologies in genetic and protein engineering. This ever growing synthetic biology toolkit opens a range of customizability when designing aggregating proteins and metal binders. Finally, using a fully biological method to create resin like ion-exchangers is appealing because protein production could be more economic and environmentally sustainable than manufacturing its synthetic ion-exchange counterpart. Proteins focused here are naturally available, require little chemical processing, and are robustly produced in *E. coli*. Much like the antibody market has grown to be a mass producer of proteins [kelley2009], the same infrastructure and technology can be leveraged to produce proteins for applications in clean water technology. As biological engineers, the hope is to take advantage of current biological technologies to efficiently and sustainably solve the waste water crisis, and more so, convince the others that many more sustainable solutions may exist with further exploration at the intersection of biology and technology.

4.4 Materials and methods

Gene isolation and plasmid construction

Genomic DNA from *E. coli* DH5 α (NEB) was isolated using DNA Fungal/Bacterial Miniprep Kit (Zymo). Gene sequences of pyrG and glnA were retrieved via Uniprot

and used to create primers (IDT) for Gibson assembly into pET28c(+) IPTG-inducible vector (Supplemental Figure ??). 5' primer contained an overhang adding a TEV protease site (ENLYFQSQS) downstream of the pET28c(+) 6xHis tag before the gene insertion. Primers were used to amplify the appropriate genes using the Q5 polymerase (NEB). All PCR products mentioned were examined under a gel imager (AlphaImager 2200) and cleaned (Promega) before performing subsequent cloning steps. Products were then assembled into linearized pET28c(+) using the HiFi 2X assembly master mix (NEB). Assembled constructs were transformed into NEB α cells (NEB), plated onto 1X kanamycin (50 mg/L) LB plates, picked and miniprepped (Promega), and sequenced to confirm proper gene insertions (Quintara Bio). Confirmed sequences were then transformed into BL21(DE3) (Agilent) for protein expression.

The pET28c(+-)glnA vector was further modified by fusing calmodulin (CaM) or *A. thaliana* metallothionein 1a (MT1A) upstream of the glnA sequence. The DNA sequence for CaM was isolated from the yeast genome and purified using the DNA Fungal/Bacterial Miniprep Kit (Zymo), whereas the MT1A sequence was codon-optimized and synthesized (GenScript). Primers were created for HiFi assembly into the pET28c(+-)glnA vector (Supplemental Table ??) following the protocol already described above.

Alternatively, pET28c(+-)glnA was modified by appending a 3XFlag tag upstream of glnA and downstream of the metal binding fusion (i.e. 6xHis tag). Primers were designed to anneal before the glnA sequence contained overhangs carrying the 3XFlag sequence (DYKDHDGDDYKDHDIDYKDDDDK) (Supplemental Table ??). The linear PCR product was re-circularized by adding T4 polynucleotide kinase with T4 ligase in 1X T4 buffer (NEB) and plasmid transformed in competent *E. coli*.

The pyrG and glnA were also used for yeast display in the pYD1 vector. Primers were designed with a 5' NheI and 3' BamHI restriction cut sites (Supplemental Table ??) and isolated using the protocol already described above. The pYD1 vector was linearized by cutting with NheI-HF and BamHI-HF restriction enzyme (NEB). pyrG or glnA were ligated into linearized pYD1 using T4 ligase and then transformed into competent *E. coli* and yeast W303 α strains.

Protein purification

BL21(DE3) cells were grown in 10 mL starter cultures overnight. The following day cultures were diluted by 1:100 in small scale (100 mL) or large scale (1 L) cultures. Cultures were grown for 5-6 hours at 30°C before induction with 1X IPTG (1 mM; GoldBio). Cultures were induced for 16 hours before harvesting by pelleting cultures at 5000xg for 15 min. Pelleted cultures were stored in -80°C for 1 day before preparing protein samples.

For 100 mL of pelleted culture, 10 mL of 1X Bugbuster (Millipore Sigma) was used for cell lysis. For each mL of Bugbuster, 1 μ L of lysonase bioprocessing reagent (Millipore Sigma), 1 μ L of DNase (NEB), and 10 μ L of 100X Halt Protease Inhibitor Cocktail (Thermo) were added. Resuspended cultures were gently agitated on a vortexer for 15–30 minutes until solutions became clear. Lysed solution were spun down at 16,000xg for 20 minutes at 4°C to remove debris and inclusion bodies. During this process Ni-NTA resin (Qiagen) was prepared by taking 2 mL for every 100 mL of pelleted culture and equilibrating with 25 mL 1X Ni-NTA Buffer (Millipore Sigma). Resins were spun down at 1000xg for 5 min and equilibration buffer removed. After bacterial lysis and centrifugation, supernatant was removed and filtered through a 0.45 μ m syringe filter directly onto equilibrated Ni-NTA resin. Mixtures were gently rocked for at least 2 hours at 4°C.

Mixtures were poured into disposable Econo-Pac chromatography columns (10 or 20 mL, depending on the scale; Biorad). Columns were washed 3 times with 5X column volumes (1 mL NTA resin equates to 5 mL of wash buffer) of 1X Ni-NTA wash buffer (Millipore Sigma). Protein was then eluted with 2X column volume of 1X Ni-NTA elution buffer (Millipore Sigma) separated into 4 fractions (1 mL of NTA resin equates to 0.5 mL of elution buffer, repeated 4 times).

Eluted proteins were desalted using Sephadex G-25 PD-10 desalting columns (GE). Columns were first equilibrated with 1X TBS (100 mM Tris in 150 mM NaCl; Rockland Inc.). 2.5 mL of eluted protein were flowed through, followed by 3.5 mL of 1X TBS which was collected. For protein samples that needed to be cleaved of the

6xHis tag, namely pyrG and glnA controls, and glnA with CaM and MT1A add-ons, samples were instead dialyzed. TEV protease (Sigma) was added to eluted samples at 1:100 (w/w), typically equating to 1:500 (v/v). Samples were fitted in 10K MWCO Snakeskin dialysis membrane (Thermo) and dialyzed against 5 L of 1X TBS for 2 days at 4°C. Dialysis buffer was exchanged every 12 hours. Dialyzed samples were then added back to Ni-NTA columns to remove cleaved 6xHis tag, uncleaved protein, and TEV itself.

Proteins that were below 1 mg/mL were concentrated using Amicon centrifugal filters (Millipore Sigma) with 10K cutoffs. Filters were spun at 4000xg at increments of 5 minutes in a swinging-bucket rotor to reduce sample volume and increase protein concentration.

Protein expression was examined using SDS-PAGE and coomassie stain. Protein samples were diluted to 0.5 mg in 20 μ L sample volume containing 1X LDS buffer and 1X denaturation buffer (Thermo). Samples were boiled at 70°C for 10 min and loaded onto pre-cast Bolt 4-12% Bis Tris gels (Thermo). Gels were ran at 200 V for 30 min in 1X MES buffer (Thermo) and visualized using PageBlue protein stain (Thermo). Band sizes were compared against the PageRuler pre-stained protein ladder (Thermo). pyrG, glnA, glnA+CaM, glnA+MT1, and glnA+3XFlag gave single and clear bands. +CaM showed a strong band at the appropriate size, and another fainter band where glnA is normally found, suggesting that the fusion protein was cleaved during the expression or purification process. However, the alkaline metal removal studies for glnA+CaM should not be effected by the presence of background glnA, as glnA has little sensitivity for the metals tested (Mg, Ca, Sr and Ba).

Protein quantification

Protein concentrations were routinely checked on the Nanodrop. However, to better calculate protein concentrations a relationship between the higher resolution Pierce 660 nm protein assay (Thermo) and Nanodrop readings was performed. Serial dilutions of pyrG, glnA, glnA+CaM, glnA+MT1A, glnA+3XFlag starting at 2 mg/mL were read on both Nanodrop and 660 nm assay. A line of best fit ($y = m \cdot x + b$)

was constructed to give a relationship between the 660 nm reading and Nanodrop, i.e. $[660] = m_{\text{nano}}[\text{Nano}] + b_{\text{nano}}$. The Pierce protein assay was then used to create a calibration curve against BSA standards (Thermo), which gave a calibration curve of protein concentration (mg/mL) versus 660 nm, i.e. $[\text{mg/mL}] = m_{600}[660] + b_{660}$. Substituting the relationship between the Nanodrop and the Pierce 660 nm protein readings, the final equation relating the Nanodrop to the BSA calibrated protein concentration reading of mg/mL gave $[\text{mg/mL}] = (m_{660} \cdot m_{\text{nano}})[\text{Nano}] + m_{660} \cdot b_{\text{nano}}$. This new line of best fit was used to correlate readings from the Nanodrop to mg/mL as determined by the more reliable Pierce 660 nm protein assay (Supplemental Figure ??).

Protein concentrations of mg/mL were converted to μM by using the monomer molecular weight. pyrG = 62.5 kDa, and glnA = 54 kDa.

Aggregation quantification using absorbance measurement

pyrG and glnA without a 6xHis tag were used in absorbance reading experiments. Purified proteins were aggregated using 10 mM of metals and serially diluted in 96-well plates. Absorbance scans at increments of 5 nm were measured for each sample on a plate reader (Tecan M200 Pro). One wavelength at each dilution was measured and fitted with absorbance (y-axis) versus dilution (x-axis). The wavelength 350 nm gave the best fit between absorbance intensity and protein aggregation versus non-aggregated protein, hence highest signal-to-noise ratio (Supplemental Figure ??).

Proteins at 100 μM were aggregated in 100 μL in a 96-well plate at varying metal concentrations starting at 10 mM. Aggregation was allowed to occur for 1 hour before measuring at 350 nm. Absorbance measurements at 350 nm (y-axis) were plotted against metal concentrations (x-axis) to fit a Hill equation in order to parametrize the K_D for metal induced aggregation, intensity of aggregation (A), and Hill coefficient (n) (Table ??).

Testing reversibility of protein aggregation was performed by aggregating proteins at 1 mM metal for 1 hour. Resuspended aggregates were aliquoted in 100 μL in a 96-well plate at a final concentration of 100 μM . In each well a different concentration

of EDTA was added, with the highest at 10 mM and serially diluted by factors of 10. EDTA and protein aggregate were mixed for 10 minutes. The plate was then measured at 350 nm to measure intensity of protein aggregation after metal removal due to EDTA.

Metal removal experiments

Liquid stocks of manganese (II) chloride, cobalt (II) chloride, nickel (II) chloride, copper (II) chloride, zinc (II) chloride, cadmium (II) nitrate, mercury (II) chloride, and lead (II) nitrate (Sigma) were made at 100 mM in water. The same for the alkaline earth metals magnesium (II) chloride, calcium (II) chloride, strontium (II) chloride, and barium (II) chloride (Sigma) were made at 100 mM in water.

Protein samples were diluted to 100 μ M in water for metal removal experiments.

Protein aggregates were first formed by mixing 100 μ M of protein with 1 mM of metal. The reaction was allowed to occur for 1 hour before spinning down samples at 10,000xg for 5 min at 4°C. The supernatant was collected and diluted 1:10 in 3% HNO₃ solution for inductively coupled plasma (ICP) measurement. An ICP-OES (Agilent 5100) was used to measure metal concentrations of the supernatant. Standards were made from ICP-quality metal stocks (Fluka). In addition, samples with metal added but no protein were measured to test for natural metal precipitation or non-specific metal binding onto the sample tubes. The only metal with appreciable precipitation was Pb, which naturally formed hydroxides after several minutes. This value was subtracted from the ICP measured in the supernatant in order to adjust for the actual metal removed due to aggregate formation. Finally, the amount of protein captured by the pelleted aggregates was calculated by subtracting the original metal concentration (i.e. 100 μ M) with the adjusted ICP measurement from the supernatant.

For multi-element removal experiments, the eight studied metals, Mn, Co, Ni, Cu, Zn, Cd, Hg, and Pb were mixed in equimolar ratios with a combined metal concentration of 1 mM (concentration of each metal being 125 μ M). The same metal removal experiment was performed as described above. Wavelengths for ICP analysis with the minimal amount of cross-over were Mn (257.610 nm), Co (230.786 nm), Ni (216.555

nm), Cu (327.395 nm), Zn (213.857 nm), Cd (226.502 nm), Hg (194.164 nm), and Pb (220.353 nm). The amount of metals removed was calculated as described above for each metal; the total amount of metal removed being the sum of all calculated values.

For metal removal experiments of control pyrG and glnA without a 6xHis tag were used. 6xHis tag of purified pyrG and glnA were cleaved using TEV protease, dialyzed, and re-purified using Ni-NTA to isolate non-tagged 6xHis proteins. Non-tagged pyrG and glnA controls were used in parallel with tagged proteins during the metal removal experiments.

Similarly, glnA+CaM and glnA+MT1 were first cleaved of its 6xHis tag used for protein purification. For glnA+CaM alkaline-earth metal removal experiments, samples were co-mixed with 100 μ L of either Mg, Ca, Sr or Ba and 100 μ M Zn. Metal removal experiments with glnA+CaM and glnA+MT1 followed the same experimental outline as described above. ICP wavelengths for alkaline-earth measurements were Mg (279.553 nm), Ca (396.847 nm), Sr (407.771 nm), and Ba (233.527 nm).

Transmission electron microscopy sample preparation and imaging

All TEM samples were prepared with pyrG and glnA without a 6xHis tag to observe native aggregation effects. 7 μ L of sample was removed for transmission electron microscopy (TEM). Samples were dropped on a 400 mesh copper grid coated on carbon film (EMS). Grids were left for 60 seconds before removing excess solution by touching the grid on a kimwipe. 10 μ L of negative staining solution phosphotungstic acid (Sigma) at 1% was dropped on the grid and immediately removed with a kimwipe. Another 10 μ L of negative stain was immediately dropped after and left for 40 seconds before removing excess liquid. Grids were left to air dry at room temperature for more than 30 minutes.

A FEI Tecnai was used at 120 kV to image protein aggregates with roughly 100 nm resolution. High resolution TEM (HRTEM) was performed on a JOEL 2100 FEG microscope at 200 kV with assistance from Koch's Nanotechnology Core.

Yeast display of CS and GS monomers

Yeast EBY100 strain were made competent using Frozen-EZ Yeast Transformation II Kit (Zymo) and either stored at -80°C, or used immediately for transformation. Competent EBY100 was transformed with pYD1 vectors containing pyrG or glnA without a 6xHis tag. Transformed cells were plated on SDCAA media (Teknova) and grown for 1.5-2 days. Transformants were confirmed by extracting DNA via bead-beating with 420-600 μ m glass beads (Sigma) and phenol:chloroform (Sigma) extraction and ethanol precipitation. A region of the pYD1 vector was amplified with primers flanking the AGA2 and T7 promoter using PCR and checked via gel electrophoresis for correct insertion.

Transformed strains were grown overnight in SDCAA media in 30°C, and diluted 1 to 10 the next day. Cultures continued to grow for another 4-6 hours to mid-log phase before spinning down cultures and resuspending in SGCAA (Teknova). Cultures were induced overnight at room temperature before harvesting. To check for positive expression, induced strains were washed in PBS + 1% BSA and tagged with primary antibodies against the N'-HA tag and the C'-flag tag for 1 hour at 4°C. Secondary antibodies conjugated with 488 or 647 dye were used against the primary antibodies and stained for 1 hour at 4°C. Fluorescently tagged cells were then analyzed using flow cytometry on a LSR II (BD Bioscience) and plots of FITC versus PE-Cy5 were analyzed to measure populations of expressing cells with respects to the WT non-expressing samples.

Sedimentation analysis

For agarose and magnetic bead sedimentation studies, a 3xFlag tag was added after the 6xHis tag (e.g. N'-6xHis-3XFlag-glnA-C'). glnA at 100 μ M was aggregated with 1 mM of Cd in 5 mL for 1 hour. Samples were then mixed with 100 μ L of pure resin pre-equilibrated in ddH₂O of either Anti-Flag M2 Affinity Gel or Anti-Flag M2 Magnetic Beads (Sigma). Aggregated samples with beads were incubated for 1 hour before transferring 4 mL to a fluorimeter cuvette with 4 flush clear sides (Sigma).

Samples were left to settle, whereas for the magnetic beads a magnet was dragged to the bottom of the cuvette and left for approximately 5 seconds before removing to magnetically pull down beads with bound proteins.

For yeast display conditions, 1 OD₆₀₀ of expressing EBY100 was mixed with 100 μM glnA and 1 mM of Cd in 5 mL during aggregation. Aggregation was allowed to continue for 1 hour before transferring to a 4 mL fluorometer cuvette.

For sedimentation studies, samples were collected at the mid-height (e.g. 2 mL for a 4 mL cuvette). 20 μL aliquots were taken at specific time points and quickly washed with 0.2 M glycine HCl pH 3.5 and shaken for 2–5 minutes to dissociate any bound proteins. Tubes were quickly centrifuged and the top 10 μL was measured for protein content using the Pierce 660 nm protein assay. The percent ratio of protein concentration measured per time point versus original concentration (i.e. 100 μM) was plotted to analyze the sedimentation rate of protein aggregates when bound to agarose, magnetic beads, or yeast displaying glnA. Experimental controls removed the connector between glnA and the denser anchor. So for agarose and magnetic beads, glnA without 3xFlag was used. For yeast display conditions an EBY100 strain displaying an empty pYD1 vector was used. The same sedimentation experiment was performed for these controls.

Metal recovery

The same glnA constructs in the sedimentation experiments were used; glnA with an added 3xFlag tag for agarose and magnetic pull-down, and EBY100 displaying a glnA monomer for yeast pull-down. For the agarose and magnetic samples, 100 μM glnA+3xFlag and 1 mM metal were allowed to aggregate for 1 hour before spinning the complex down. The supernatant was sampled to calculate the amount of metal captured in the protein-metal complex. This value was set as the initial amount of metal captured (i.e. maximum amount of metal recoverable, 100%). As a control, samples with no protein were measured to account for the natural precipitation of metal, namely Pb. This value was subtracted from the initial metal measurement to isolate the actual amount of metal removed from the protein-metal complex. After-

wards, aggregates were resuspended and then mixed with 100 μL of pure agarose or magnetic resin pre-equilibrated in ddH₂O. Mixtures were gently shaken for 1 hour before aliquoting 1 mL into Eppendorf tubes. Samples were allowed to settle for 10 minutes, the same time window as in the sedimentation experiments. For samples with magnetic beads, a magnet was dragged to the bottom of the cuvette and left for approximately 5 seconds. Afterwards, the top 900 μL of sample was removed. The remaining volume was diluted with 1 mL of ddH₂O, spun down, and washed once more to remove any metals not bound to protein. 1 mL of 10 mM EDTA was then added to the pellet and mixed for 10 minutes. The tube was spun once more to remove any beads/yeast or protein debri. The supernatant was sampled and measured for metal using ICP. This value was set as the amount of metal recovered. The amount of metal recovered was divided by the initial amount of metal captured to give a recovery percent.

For yeast anchored samples, 100 μM glnA was mixed with 1 mM metal and 1 OD₆₀₀ of induced EBY100 displaying glnA monomers. The mixture was allowed to aggregate for 1 hour before spinning it down. The supernatant was sampled to calculate the amount of metal captured in the protein-yeast-metal complex. This value was set as the initial amount of metal captured. Controls with just induced EBY100 displaying glnA monomers and 100 μM metal were used to measure non-specific binding onto the yeast surface and natural precipitation of metal. This value was subtracted from the initial amount of metal captured to isolate the actual amount of metal removed from just the protein-metal complex. 1 mL of yeast-protein-metal mixture was aliquoted into Eppendorf tubes. The same protocol mentioned with the agarose and magnetic bead was performed to calculate the amount of metal recoverable. Percent recovery values for all conditions: agarose, magnetic, and yeast along with all metals used in this work were plotted and compared with one another.

Mathematical analysis and plotting

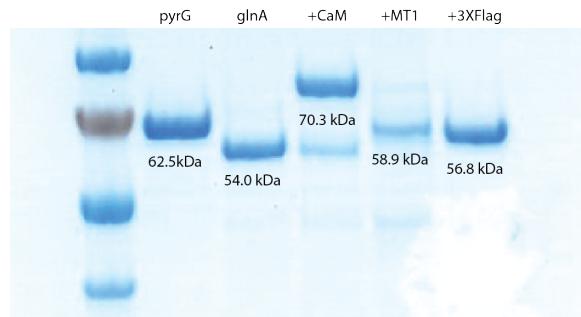
Raw data was collected and stored as csv or Excel file formats. Data was imported and analyzed with Python using modules such as numpy, pandas, and scipy. Plots

were graphed with matplotlib.

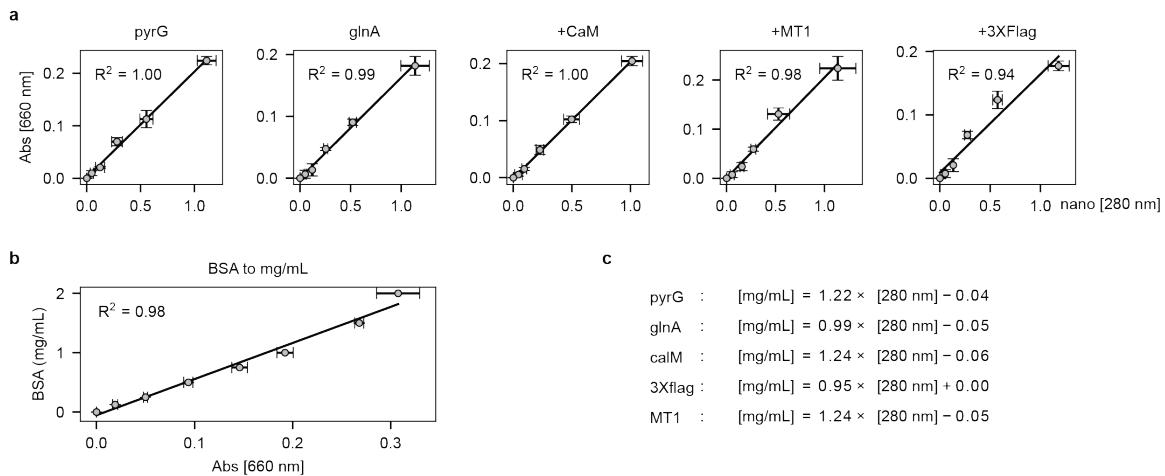
Statistical analysis

Statistical parameters including the the definition and values of n, SDs, and/or SEs are reported in the figures and corresponding figure legends. When reporting significance, a two-tailed unpaired t-test was performed between observations and p-values reported in the text. The significance threshold was set to $p < .05$ for all experiments, or as specified in the text. In the figures, * = $p < .05$ and ** = $p < .01$.

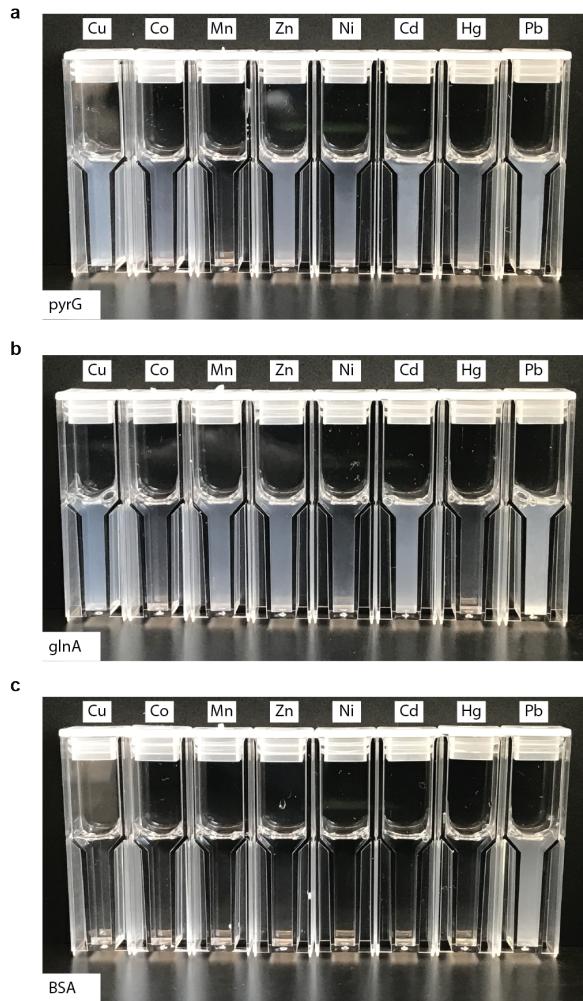
4.5 Supplemental figures



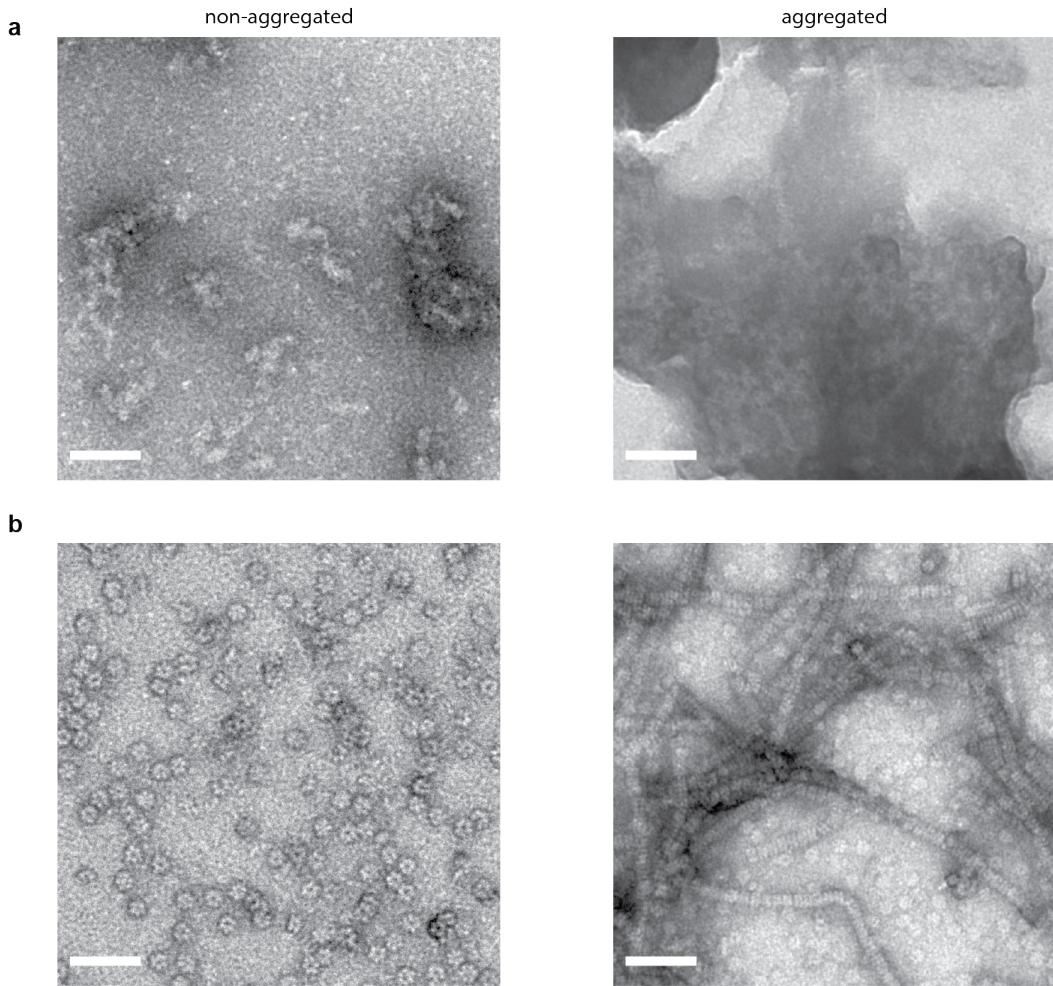
S.Figure 4.1 | Examining protein expression and purity using SDS-PAGE and coomassie staining. Purified pyrG, glnA, glnA+CaM, glnA+MT1, and glnA+3XFlag showed the correct band size. Only glnA+CaM showed a fainter second band which corresponds to un-modified glnA, suggesting that the glnA+CaM fusion was cleaved either during expression or during the purification process.



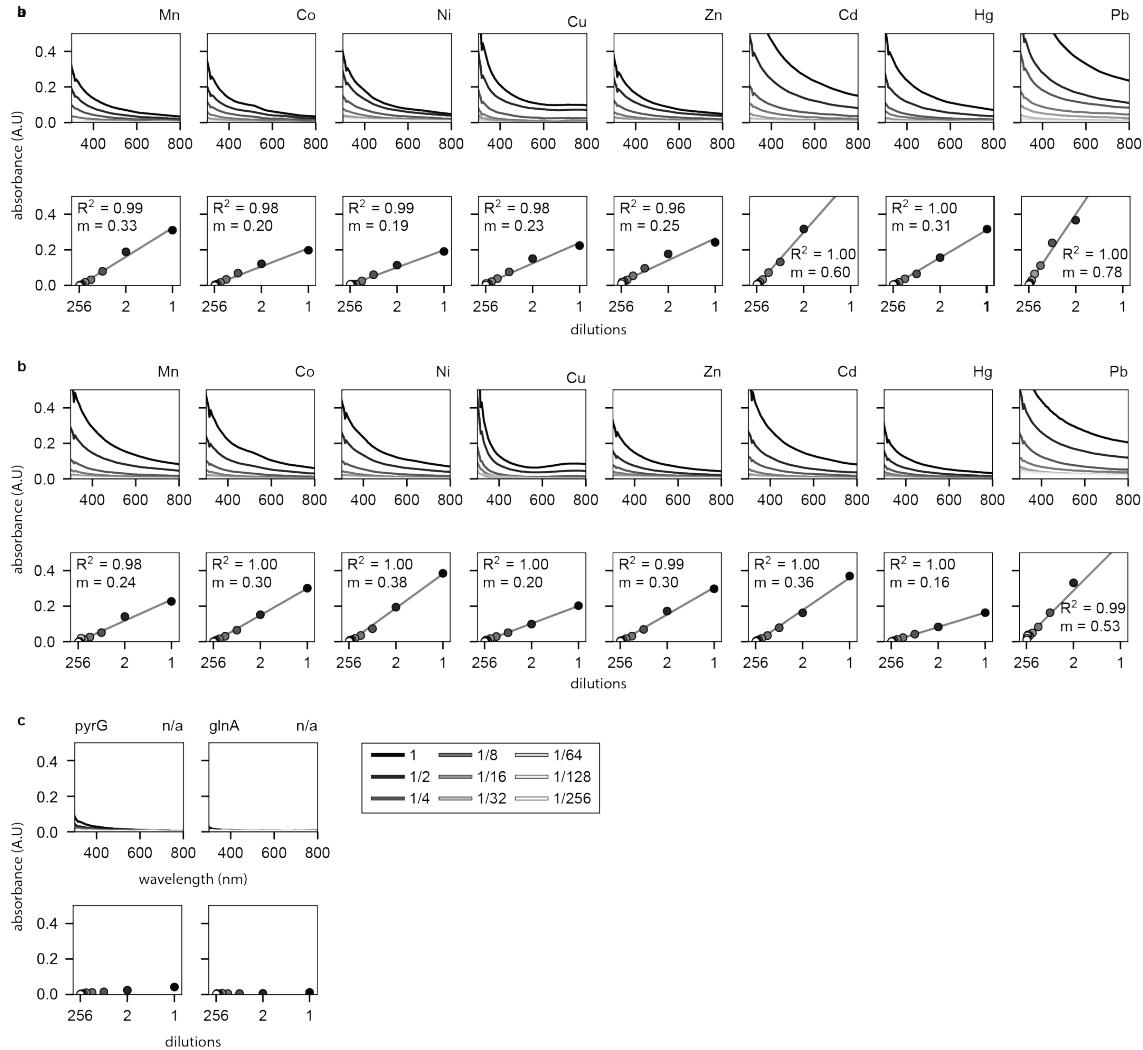
S.Figure 4.2 | Protein concentrations measured on Nanodrop were correlated with Pierce 660 Protein Assay. (a) Serial dilutions of proteins pyrG, glnA, +CaM, +MT1, and +3XFlag were measured on Nanodrop (x-axis) and Pierce 660 nm protein assay (y-axis). (b) Pierce assay was then used to create a calibration curve against BSA standards. The calibration curve, as well as the relationship between the Nanodrop and Pierce 660 nm protein assay was used to create a linear relationship (c) to better quantify concentrations for each protein. For all data, the mean \pm s.d. of three replicates are shown.



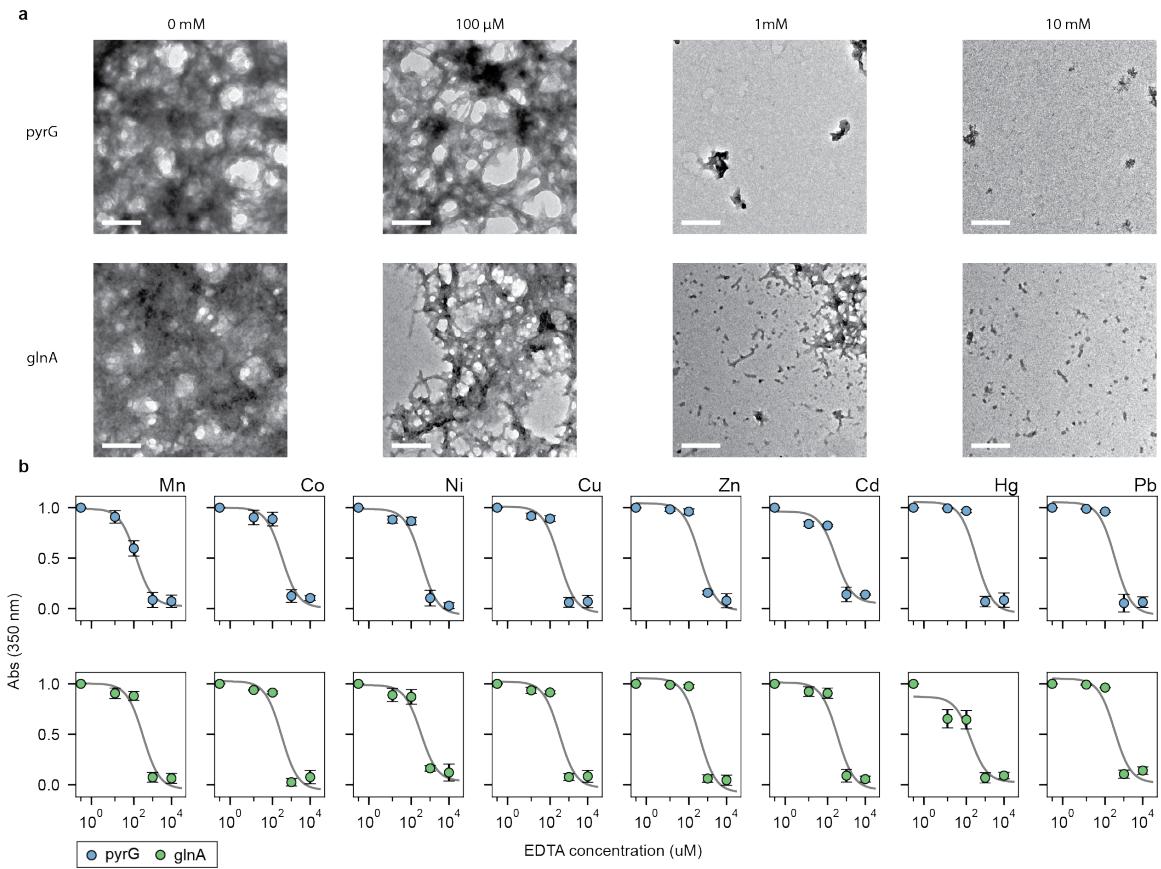
S.Figure 4.3 | Images of 100 μM pyrG, glnA and BSA mixed with 1 mM of metals. (a, b) Aggregation of pyrG and glnA. (c) As a control, BSA was also mixed with 1 mM of metals to measure aggregation due to protein denaturation, or metal precipitation. The only visible change in opacity was with Pb, as it spontaneously formed hydroxides in neutral pH over time.



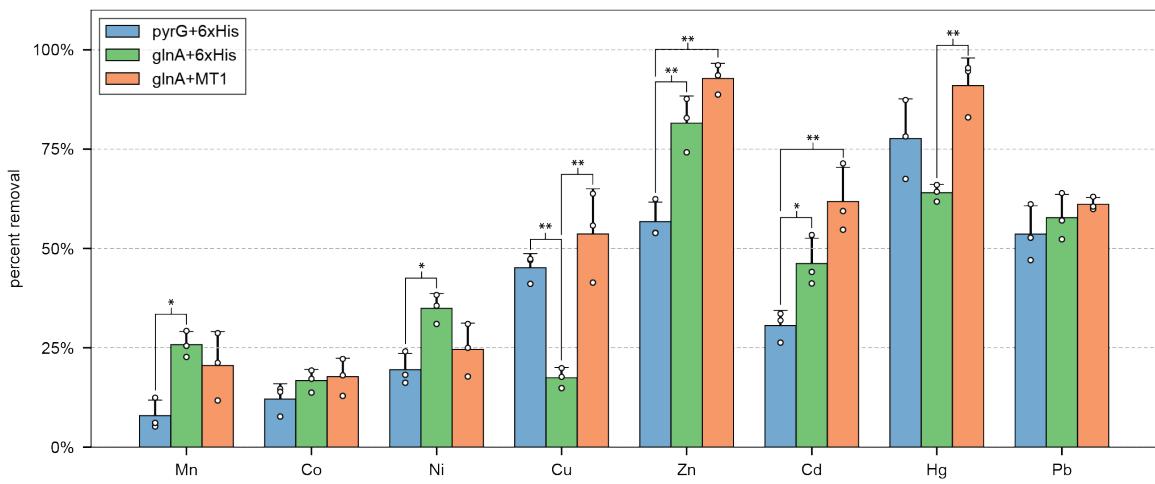
S.Figure 4.4 | HRTEM images of pyrG and glnA. (a) High resolution of pyrG show no definitive structure, but rather monomers that haphazardly form small aggregates. When aggregation is induced with metal, pyrG produces densely packed structures as seen by the dark contrast. (b) glnA forms the predicted structure of two hexamers stacked on top of one another. When aggregated, glnA forms a cob-like structures with glnA stacked on top of one another in a vertical fashion.



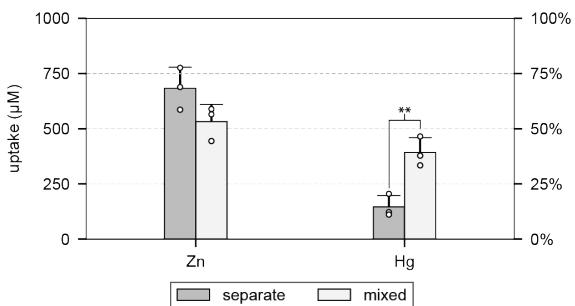
S. Figure 4.5 | Absorbance scans of pyrG and glnA at varying degrees of metal-induced aggregation. (a, b) pyrG and glnA absorbance scans for all metals tested in this study at increasing dilutions as indicated by the lower panel x-axis and increasing line opacity. Scans at 350 nm gave a linear relationship with good signal to noise for the various dilutions of protein aggregation. (c) Absorbance scans of non-aggregated pyrG and glnA absent of any metals.



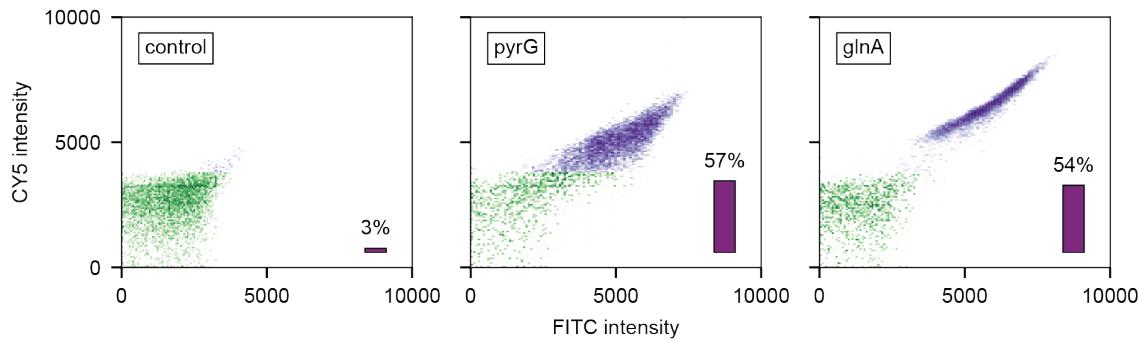
S. Figure 4.6 | Reversibility of aggregated pyrG and glnA after metal removal using EDTA. (a) Top/bottom row are pyrG and glnA, respectively. Columns represent concentrations of EDTA added after 100 μ M of protein was aggregated with 1 mM of Zn for 1 hour. Scale bars represent 200 nm for all images. (b) After protein aggregation, concentrations of EDTA (x-axis) were added to compete for metal chelation, and eventual dissociation of protein subunits. The intensity of aggregation was measured at 350 nm after adding EDTA for 10 min. For all data, the mean \pm s.d. of three replicates are shown.



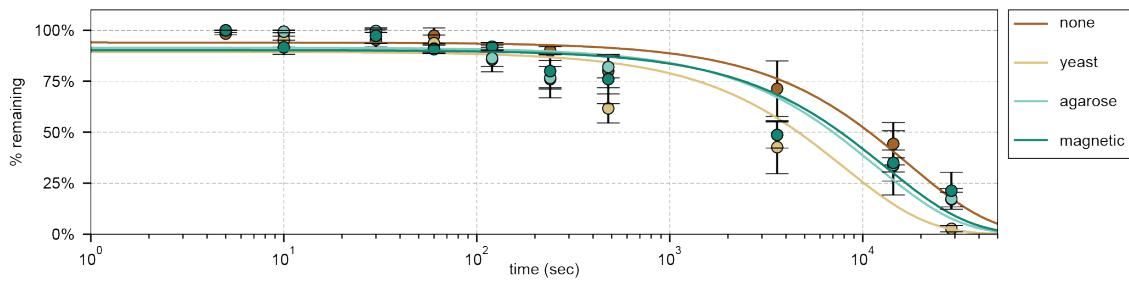
S.Figure 4.7 | Percent metal removal for multi-metal removal experiments using *pyrG*, *glnA*, and *glnA* with MT1. Multi-metal removal data reported in Figure ??c,d and Figure ??e were converted to percent fraction of metal removed. For all data, the mean \pm s.d. of three replicates are shown. * = $p < .05$, and ** = $p < .01$ represent significant metal removal compared to controls.



S.Figure 4.8 | Mixing Zn and Hg alters metal removal for glnA. Individually, Zn is removed significantly more than Hg when mixed with 100 μM glnA with a 6xHis tag. However, when Zn and Hg are mixed together, Zn removal decreases while Hg increases ($p < .01$). An explanation for this trend is that glnA does not aggregate as strongly with Hg it does with Zn. However, when mixed together, glnA aggregates in the presence of Zn but the 6xHis tag preferentially pulls down Hg over Zn. Therefore Hg removal piggy-backs Zn aggregation of glnA. For all data, the mean \pm s.d. of three replicates are shown. ** = $p < .01$ represent significant metal removal compared to controls.



S.Figure 4.9 | Yeast display of *pyrG* and *glnA* monomers using EBY100 and pYD1 vector. Expression was analyzed using flow cytometry by staining the N'-terminus HA tag and C'-terminus Flag tag with antibodies conjugated with 488 (FITC) and 647 (Cy5) dyes. A WT control was measured in parallel to properly bin the population of non-expressing cells to expressing cells. The percent expression is shown as the lower right inlet bar chart.



S.Figure 4.10 | Control sedimentation study of glnA using agarose or magnetic beads, or yeast. glnA without a 3XFlag tag was aggregated and mixed with anti-flag agarose or magnetic beads. The yeast control had EBY100 strains transformed with an empty pYD1 vector and mixed with aggregated glnA. The sedimentation rates for each control was comparable to the sample containing no beads or yeast. For all data, the mean \pm s.d. of three replicates are shown.

4.6 Supplemental tables

name	direction	sequence
pyrG	fwd	CAGCAGCGCGAAAACCTGTATTTCAGAGCACAA CGAACTATATTTTGACCC
	rev	CCACCAGTCATGCTAGCCATATGTTACTTCGCCTG ACGTTT
glnA	fwd	CAGCAGCGCGAAAACCTGTATTTCAGAGCTCCG CTGAACACGTAC
	rev	CCACCAGTCATGCTAGCCATATGTTAGACGCTGTA GTACAGC
pET28c(+)	fwd	ACAGGTTTCGCCGCTGCTGTGATGATG
	rev	CATATGGCTAGCATGACTGGTG

S.Table 4.1 | Primers to isolate pyrG and glnA from *E. coli* genomic DNA
 5' fwd primer contain a 6xHis tag and TEV protease site that is added upstream to the gene. Purified PCR products were then Gibson assembled into pET28c(+) linearized by the primers shown.

name	direction	sequence
+CaM	fwd	CGGCGAAAACCTGTATTTCAGAGCTCCTCCAATC TTACCGAAG
	rev	CGGAGCCGCTACCGCCTTAGATAACAAAGCAGCG A
+MT1a	fwd	CGGCGAAAACCTGTATTTCAGAGCATGGCTGATT CTAATTGTGG
	rev	CGGAGCCGCTACCGCCACAATTACAGTTAACCA CAA
glnA (bB)	fwd	GGCGGTAGCGGCTCCGCTGAACACGTACT
	rev	GCTCTGAAAATACAGGTTTCG

S.Table 4.2 | Calmodulin (CaM) and plant metallothionein (MT1A) were added to constructed pET28c(+) glnA vector using primers shown. CaM was isolated from yeast genomic DNA, whereas MT1A was codon-optimized and isolated from the synthesized plasmid. Primers contain the appropriate overhangs to Gibson assemble into the constructed pET28c(+) glnA vector which was linearized with the primers shown.

name	direction	sequence
+3xFlag	fwd	CATGACATCGATTACAAGGATGACGATGACAAGTC CGCTGAACACGTACT
	rev	ATCTTTATAATCACCGTCATGGTCTTGTAGTCGC TCTGAAAATACAGGTTTCG s

S.Table 4.3 | A 3XFlag tag was added to the constructed pET28c(+) glnA vector. Primers were used to linearize the pET28c(+) vector with each primer overhang encoding half of the 3XFlag sequence. Linearized product was then blunt-end ligated to re-circularize the plasmid.

name	direction	sequence
glnA	fwd	GCTAGCTCAGCCGAACACGTATTAAC
	rev	GGATCCA ACTGAGTAATACAATTCAAATTCAAC
pyrG	fwd	GCTAGCACAA CGAACTATATTTTGACCc
	rev	GGATCCCTTCGCCTGACGTTCTG

S.Table 4.4 | Primers used to insert bacterial glnA and pyrG genes into pYD1 yeast display vector. Primers were used to amplify glnA and pyrG with NheI and BamHI overhangs. The pYD1 vector was cut with NheI and BamHI and ligated with the digested pyrG or glnA product.

Chapter 5

Considerations and future work

5.1 Use of other biological processes for heavy metal removal

Electrochemical treatment

Another actively performed physiochemical strategy is electrochemical treatment. Similar to chemical precipitation, the intention is to convert dissolved metals into another form such as a solid. The mechanism of action is through oxidation or reduction of metals by applying current/voltage in solution to plate-out metal ions onto a cathode (reducing) or anode (oxidizing) surface. Once deposited, the plated metal can be later stripped, collected, and cathode/anode renewed for future rounds of use [wang2007mechanism].

Difficulties occur when reducing high redox potential species such as mercury and arsenic. Moreover, conditions need to be optimized to select the correct solvent and cathode/anode composition for specific metal deposition [fu2011removal]. Lastly, electrochemical treatment is a unique process among the physicochemical processes in which external energy needs to be injected into the system.

Enzymatic reduction or oxidation of metals as an analogy to electrochemical treatment

The biological analogy to electrochemical treatment would be to use enzymatic reductases or oxidases to alter a metal's valent state, or convert metals to their ground state M^0 as a solid. Reductases and oxidases are enzymes that catalyze reducing or oxidizing reactions by facilitating the transfer of electrons from substrate to product¹. To facilitate the transfer of electrons many reductases and oxidases contain a porphyrin, a molecular co-factor which contains a flat ring of four heterocyclic groups with a central metal atom; examples being heme (Fe^{2+}) and chlorophyll (Mg^{2+}). Common electron transfer enzymes are cytochromes which contain a heme co-factor that participates in the electron transport chain during respiration. During respiration the terminal electron acceptor is oxygen, which provides the energy to then allow the catalytic formation of ATP.

There are other microorganisms that use metals as their terminal electron acceptor rather than oxygen. In a way, these microorganisms breathe metal. This phenomenon is typically found in a family of bacteria and archaea known as dissimilatory metal-reducing bacteria (DMRBs). DMRBs utilize electrochemical gradients generated from metal reduction to create chemical energy for cellular growth [lovley1993dissimilatory]. What drives dissimilatory metal reduction are the numerous copies of c-type cytochromes on the periplasmic and cell membranes that facilitate electron transfer to surrounding metals [lovley1993dissimilatory, weber2006microorganisms]. Cytochrome families *mtr*, *omc*, *ppcA* have been implicated in $Fe^{3+} \rightarrow Fe^{2+}$, and even $Hg^{2+} \rightarrow Hg^0$ $Cr^{6+} \rightarrow Cr^{3+}$ $As^{5+} \rightarrow Fe^{3+}$ [liu2002reduction, wiatrowski2006novel, stolz1999bacterial] The electron donor is usually an organic molecule such as NAD(P)H, formate, lactate, or pyruvate which is replenished by dehydrogenases or other metabolic processes [du2007state] (Figure ??a).

¹For the majority of metals which are in the cationic state, reductases (transfer of electrons to the metal) are generally more desirable (except for dealing with polyatomic metal anions such as CrO_4^{2-} and AsO_4^{3-}).

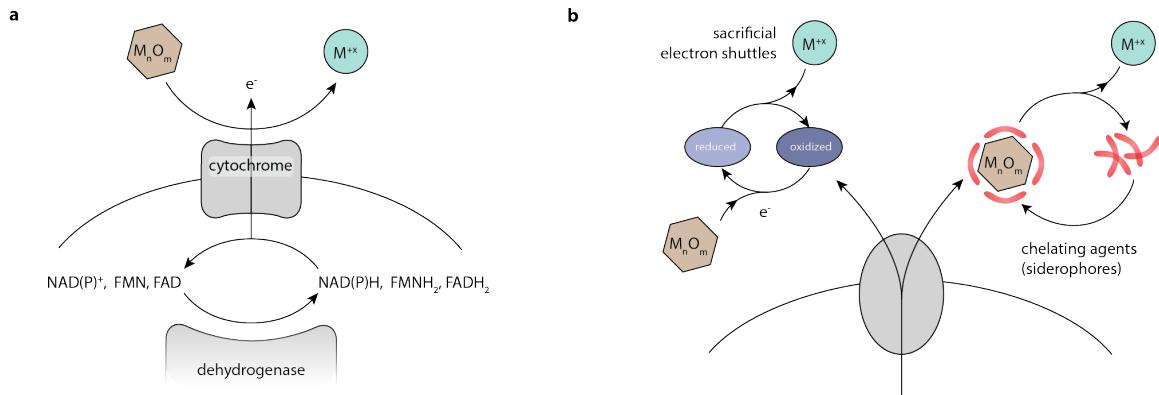


Figure 5.1 | Overview of metal reducing mechanisms found in nature. (a) Cytochromes facilitate the transfer of electrons derived from biological processes that convert extracellular metals and metal oxides to those of lower oxidation states. (b) Other mechanisms of metal reduction use exported sacrificial electron shuttles and chelators to reduce environmental metals.

Other mechanisms such as assimilatory metal-reduction use sacrificial electron shuttles like sulfide containing molecules or chelating agents like siderophores to react or entrap metal species [rajkumar2010potential, hao1996sulfate] (Figure ??b). These reacted compounds can either be stored in the cell for later metabolic processing or removed by precipitating captured metals out of solution [gadd2004microbial]. Overall, both methods of dis- and assimilatory metal reduction provide energy-efficient and natural pathways to convert metals from highly reactive valencies to a less reactive or usable form without the use of an external power source commonly needed in electrochemical treatment.

However, a major difficulty with using DMRBs is their exotism. Many DMRBs live in environments that cannot be easily recapitulated in a laboratory setting, and some culture conditions are still unknown to scientists. In addition, many DMRBs are obligate anaerobes, making ambient oxygen concentrations lethal to cell growth. However, one microorganism that has been a popular model organism for dissimilatory metal reduction is *Shewanella oneidensis*. *S. oneidensis* is able to grow in both aerobic and anaerobic conditions, and can be cultured in LB broth commonly used for *E. coli* cultures. In anaerobic environments, *S. oneidensis* have been shown

to reduce metals such as iron, lead, chromates, arsenates, and even radioactive elements such as uranium [lies2005, holmes2002enrichment, lloyd2001]. More so, several studies have used *S. oneidensis* for microbial fuel cells, a method that leverages the microbes' electron-shuttling mechanism to naturally provide electrical power between an anode and cathode material much like a battery [allen1993microbial, schaetzle2008, du2007state]. However, like using *E. coli*, using *S. oneidensis* has its limitations when taking into account environmental waste conditions, scalability, and engineerability (Figure ??). Given the positive experimental results using yeast with the other physicochemically-inspired strategies (Chapters ??–??), a follow up study would be to ask whether electrochemical treatment of waste can be transferred to yeast using enzymes and pathways found in DMRBs.

5.2 Modular and synergistic combination of yeast-based strategies

Three separate physicochemical strategies were discussed in Chapter 1, and their mechanism of action were compared to processes found in nature. Mechanistically examining these physicochemical strategies, the goal was to endow yeast with biologically equivalent capabilities for heavy metal waste removal by creating mirrored biological analogies. They were: chemical precipitation → biomimetic mineralization, absorption → membrane transporters, ion-exchange → and metal-protein aggregation (Table ??; Figure ??). Chapters ??–?? individually discuss experimental work and results to design yeast with biomimetic mineralization, internalized metal trafficking, and metal-protein binding capabilities.

In conventional physicochemical systems, these strategies are truly independent, requiring separate chemical synthesis processes, reactors, and infrastructure to house them. However, for a biologically-based approach these strategies are not mutually exclusion. Each strategy could be swapped, mixed, and combined with one another (in addition to future strategies still to be developed) by simply providing the correct

genetic instructions. Rather than having individual strains perform biomineratization or metal transport, a single strain can be modified to perform them in tandem. With the advent of synthetic biology, the power of logic—the logic typically referred to in microprocessors and computing—can be incorporated into the genetic instructions to broaden the sophistication of such combined processes where one process operates in certain scenarios, while the other is modulated, and vice-versa. Abstractly, yeast are the factories in which they compile and build from the instructions given to them by the engineer. As *this* engineer, we have the freedom to modularly and synergistically design strategies that improve upon one another, rather than in conventional waste treatment strategies which is forced to silo each process to an individual bucket.

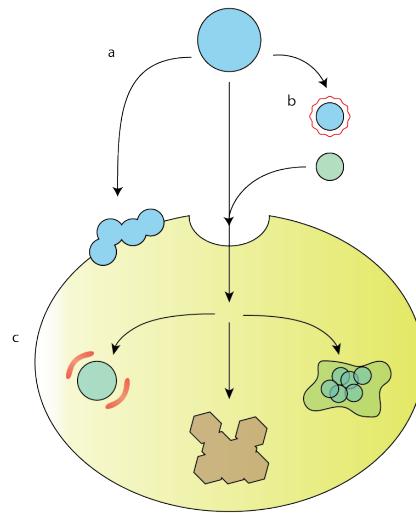


Figure 5.2 | Example of synergistic combinations of metal remediation strategies in yeast. (a) Metals could be precipitated onto the cell surface, or transported into the cell. (b) Alternatively, metals could be extracellularly modified to enable sequestration in solution, facilitated precipitation onto the cell surface, or improve recognition for internalization. (c) In the cell, several of the pathways that were performed extracellularly could be similarly performed intracellularly. Some examples would be to chelate internalized metals, chelate and transport metals into compartment organelles such as the vacuole, or chemically convert metals into useful structures or materials.

5.2.1 Genetic circuits and logic

So far the strategies proposed in this work have a single mode of operation which is perpetually on. However, there are several advantages if yeast can be controlled in an analog fashion that is responsive to the given environment or from a controlled external stimuli. Such “input”–“output” behavior is possible through the construction of genetic circuits. To achieve sensing and actuating behavior, i.e. logic, a combination of promoters, repressors, and other genetic components could be identified and appropriately engineered to control the flow of protein expression or metabolic pathways.

An example of using a genetic circuit is to turn on or off the bioremediation activity. This may be important if yeast are to be stored or grown to scale in the absence of toxic material. In this case the cellular machinery should instead focus on growing efficiently. However, the desired genes should be activated in the presence of metal. Metal-induced activation could be possible by taking advantage of the metal-sensitive CUP1 promoter [**jeyaprakash1991**], or metal-responsive transcription factors ZAP1 (#P47043), AFT1 (#P22149), and Mac1 (#P47043) which are sensitive to copper, zinc, and iron concentrations [**waldron2009**]. The degree of protein expression could also be determined by the level of metal contaminants. If the level of contamination is low, then expression should be equally regulated such that energy is appropriately distributed between metal removal and cell growth. More complex sensing and actuation would involve transitioning between different states of action for multi-functional yeast. For example, having yeast express membrane and vacuole transporters and then transition to H₂S production to precipitate the internalized metals (Figure ??c).

Turning yeast off is just as important. When yeast have reached their metal removal capacities (e.g. yeast that are fully mineralized, metal internalized, or metal bound) they should be capable of removing themselves autonomously from the media in order to allow fresher more recently grown yeast to replace them. A possible removal mechanism is to encourage flocculation. The expression of the FLO1 gene

(#P32768) promotes yeast-specific adhesion that causes yeast to aggregate and sediment [verstrepen2006, ellis2009diversity]. When yeast have sensed that they have removed the maximum amount of metal, a genetic toggle could turn on the expression of the FLO1 gene to aggregate older cells, sedimenting them from solution and allowing recently grown yeast to reoccupy the media. A simple genetic circuit would be to place a timer, in which irregardless of metal capture, cells flocculate after a set amount of time when exposed to metal. A more sophisticated design would be to sense the amount of internalized or reacted metals in or on the cell, at which above a certain threshold would activate flocculation through the use of metal responsive promoters or transcription factors.

5.2.2 Yeast containment strategies

A follow up question is how to retrieve yeast that have been distributed onto a waste site. One method would be to size-exclude yeast from the remediated water, as the average diameter of yeast ($1\text{--}10 \mu\text{m}$) lends itself to simple filtration. This technique is often employed in the beer and consumer industry where yeast, along with other particulates, are strained from the liquid fraction [harrison2015]. However, containment can be more precise and controlled. Strategies include physical containment into cartridges, or biological containment in solid biological matrixes such as biofilms or hydrogels. A follow-up discussion would be how to control yeast that have escaped their containment and reduce the likelihood of contamination of “rogue” yeast into the environment.

Physical containment

One of the most common methods to segregate biological contaminants is to exclude cells based on size using filters with defined pore-sizes. This may serve as a first pass to remove yeast from solution; however, this method may become time-consuming and less effective over time as the flow rate would gradually decrease as more yeast accumulate on the filter. In addition, many filters are often thrown out after use,

as attempting to clean filters could potentially damage them. Another possibility is to promote sedimentation. As mentioned in Chapter ?? and Section ??, yeast that have reached their maximum removal capacities should sediment or flocculate out of solution [verstrepen2006]. From here, it is somewhat straightforward to segregate the top-layer of the remediated liquid from the bottom sediments. As an additional layer of safety, the top-layer could be flowed through a size-exclusion filter to remove any lingering yeast, in this case the quantity of yeast should be minimal and not negatively impact the filter integrity or flow rate.

Another strategy would be to collect yeast by having them chemically or electrostatically bind onto an externally introduced material such as a blanket or mesh. Yeast can be surface functionalized with ligands, antibodies, or magnetic particles that when in contact with the appropriate binding partner are then attached or attached to the blanket/mesh material and bound. This approach is very similar to anchoring proteins with agarose or magnetic beads to increase sedimentation rates for ion-exchange like removal (Chapter ??).

Conversely, rather than adding yeast to a waste site, waste can be instead added to yeast. Yeast can be stored in cartridges or columns which have an inlet and outlet. In these cartridges yeast could be packed and secured using size-excluding membranes at the in/outlets, or have yeast solidly cross-linked using cross-linking peptides via yeast display or using the FLO1 gene [verstrepen2006]. Waste could be flowed through the inlet, and remediated water collected from the outlet. Much like current industrial practices, waste water is taken from their original location and transferred to a treatment plant in which water is handled at different processing stations. In this example, yeast housed in a container could be one of these processing stations with a modular inlet and outlet connecting adjacent stations. Alternatively, for small-scale applications, cartridges of yeast can be used to filter water from local sources such as a faucet or stream. Water could be incubated in these cartridges for a period of time, and then gravity flowed to elute cleaned water from the outlet.

Biofilms and hydrogels

There are also several fully biological methods to contain yeast into solid and containable formats. One such method is to embed yeast into biofilms through biochemical or microbiological methods. Biofilms contain consortium of microorganisms that stick together on a surface through adhesion of secreted extracellular polymeric substances. The physical properties of biofilms have led many scientist to use biofilms to filter water much like synthetic membrane filtration systems [**simpson2008, shannon2010science**]. Alone, yeast do not have biofilm-forming capabilities. However, a symbiotic culture of bacteria and yeast (SCOBY), the physical biofilm referred to as a pellicle, can form a thick membrane often used in creating kombucha drink [**jayabalan2014review**]. Recently, some scientist have attempted to use pellicles to act as natural filters for clean water applications (Figure ??e)². Formation of pellicles offer a physical handle on the bacteria/yeast biofilm; incorporating the engineered yeast would then add the heavy metal removal machinery of precipitation, absorption, metal chelation, or combinations thereof.

In addition to naturally created biofilms, another approach is to artificially construct a biological environment using hydrogels to contain yeast. Hydrogels are a network of polymeric chains that are highly absorbent and are frequently used in tissue engineering applications [**lee2001hydrogels**]. Mammalian cells are often used, primarily for their bio-compatibility when creating tissue scaffolds or for creating 2–3 dimensional organoids; however, yeast can also be easily impregnated into hydrogels using similar biochemical processes. Much like the previous biofilm example, hydrogels can also be used to flow and filter waste water. In addition, because hydrogels have porous membranes and relatively high water content, hydrogels can also be used as sponges in which a semi-dry or lyophilized sample can be re-hydrated with waste water. Contaminants are absorbed into the hydrogel matrix while cleaned water can be squeezed or flowed back out.

²work currently investigated in collaboration with Zijay Tang, Tim Lu Lab at MIT.

Kill switches

A common concern for many biologically engineered systems is the possibility of creating uncontrollable or “rogue” strains that enter the environment. The containment strategies proposed previously are meant to secure yeast and prevent escape; however, the smallest possibility of yeast escaping into the environment should be acknowledged and the appropriate engineering controls should be designed for preemptive measures. One common strategy is to produce deficient yeast, in other words, yeast that are unable to grow in natural environments because they lack essential genes to function. This is routinely performed in laboratory settings in which lab strains lack auxotrophic markers associated with amino acid synthesis such as tryptophan (*trp1*), leucine (*leu2*), histidine (*his3*), and other biological compounds such as adenine (*ade1*) and uracile (*ura3*)³. Removal of these nutrients halt growth, and in prolong cases cause cell death due to nutrient starvation. During laboratory production, yeast are grown in cultures containing these nutrients, but when used in environmental settings these nutrients should be omitted.

A second method is to employ kill switches, a genetic circuit when turned on induces cellular apoptosis [chan2016]. Yeast naturally contain apoptotic mechanisms [skulachev2002programmed], and this mechanism can be engineered to turn on or off under specific circumstances. An example is culturing yeast in a medium which contains an inexpensive nutrient that halts the apoptotic mechanism (e.g. an inhibitor or repressor). When yeast escape into the environment where this nutrient is no longer available, the apoptotic mechanism would now be activated and prompt cell death. This strategy could be added in addition to the other strategies discussed previously to create a secondary or tertiary layer of security. These layers would be physical containment, chemical and nutrient control over yeast growth, to finally kill-switches which irreversibly terminate the cell upon escape.

³list of common auxotrophic markers: https://wiki.yeastgenome.org/index.php/Commonly_used_auxotrophic_markers

5.3 Brief economics and scalability of yeast

5.3.1 Consumer use and practical applications

For thousands of years yeast have been a staple consumer good for its use in food, beverages, and drugs. It was not until the 1850s that yeast were “domesticated” and scientifically studied for controlled beer, food, and chemical production [greig2009]. Thereon, yeast have been sold as a commodity from freeze-dried packets to cultures shared between consumers, hobbyist, and industries. The scientific work here leverages this mature market so that findings from the bench can be easily transferred and adopted in industry. An example is growing, packaging, and distributing yeast for clean water applications. Yeast can be grown in dedicated facilities, freeze-dried, and either stored or distributed with shelf-lives of up to months to years until when needed [lodato1999viability] (Figure ??).



Figure 5.3 | Overview of the yeast consumer good industry. (a) Yeast can be grown in cultures or fermentors routinely used in the consumer and pharmaceutical industry. (b) Once grown, yeast can be freeze-dried into packets and reconstituted when needed. (c) Alternatively, yeast can be compressed into blocks for larger scale distribution. (d) When not in need, or for disaster scenarios, large quantities of yeast can be stored for emergency situations.

There are two main methods of distribution and use for bioremediating yeast. The first is direct to consumers, where individuals are responsible for purchasing, growing, and using these yeast for their own water cleaning purposes. In these scenarios a variety of engineering controls should be set up to allow ease of use and safety. One of the first requirements is the ability to grow and contain yeast. For this, a separate unit could be provided such as a small user-friendly culture flask (or tub, vat, container, etc.) and a packet of nutrients to enable proper yeast growth (see Section ?? for containment strategies). Instructions normally used for baking or brewing recipes could also be provided to help guide users to optimally grow their yeast. When grown, several other add-ons can be provided such as a cartridge in which yeast can be packed into and used as a filtration unit for routine water needs (e.g. from a faucet, well, or nearby water stream). Alternatively, yeast could be packed into a separate package and stored for later use, or stored as a starter culture for subsequent yeast cultures. As for disposal, local removal could mean destroying the yeast by heating, bleaching, or adding a chemical to induce apoptosis (i.e. kill switch; Section ??). Possibly a safer removal strategy would be to collect used yeast and have it processed by a local site which contains the infrastructure to separate yeast from the collected metal. Much like we have biological, trash, and plastic processing sites today. This site can then recycle both the metal and yeast, and potentially repackage and recycle the yeast for later consumer use.

The second distribution method would be direct to businesses such as factories or waste treatment centers. Instead, the full yeast-mediated waste treatment process would be handled on site in parallel to all other cleaning processes in the facility. Yeast can be grown on site or provided by a partnering supplier such as a brewing company where they often do not reuse spent yeast. Therefore, the yeast can be stored in a similar brewing container in which waste water enters from an inlet, and the outlet elutes the treated water. Operations will include how to remove yeast after reaching maximum metal removing capacities, cycling used yeast with fresher stock, and optimizing flow rate and incubation times. In addition, used yeast should then be treated to separate collected metal for downstream recycling. Non-destructive

methods would be to fractionate precipitated or bound metals from the supernatant (Chapter ??), leaving yeast intact and reusable. A destructive method, typically for absorbed metals, would be the lyse yeast and separate the biological debri from the metals. Even with destroyed yeast, the biological residual can be processed and used as yeast extract which can feed later cultures; a common process used in both industrial and laboratory settings [greig2009].

5.3.2 Techno-economic analysis and scaling

The yeast market is extremely mature and has grown with the growing consumer market. The global yeast market was valued at 4.2 billion dollars in 2017 with a compound annual growth rate (CAGR) of 9.2%, making it a 9–10 billion dollar market by 2026 [prnewswire.n.d.]. From 2012–2017 the worldwide beer production has consistently produced almost 200 billion liters of yeast, roughly equivalent to 1 million tons of yeast per year [thierrygodard2018, barth-haasgroup.2019]. These values do not include the amount of yeast produced by the craft beer market, which was roughly another 5 billion liters [demetergroup.n.d.2014], or the amount of yeast used in the consumer or pharmaceutical market.

The yeast industry has scaled its infrastructure to produce, package, and distribute yeast to meet global demands, making the price to culture and maintain yeast extremely cost-effective. Considering the chemical and physical cost to contain, grow, and feed yeast, it takes \$4 to grow 1 kg of yeast, and equivalently \$3 to supply the appropriate chemical nutrients (e.g. glucose, extracts, buffers, water, etc.) [harrison2015]. On a per liter basis, this is equivalent to 16 cents per liter of yeast (Appendix ??) [harrison2015]. This cost increases to approximately \$7.94–\$12.62 per liter of yeast for home brewers or for laboratory settings that do not have the economics of scale to reduce cost (Table ??).

Overall, producing yeast has been an economically viable market, that if tapped into, can dramatically impact the global waste water crisis. For today's physicochemical processes it roughly takes an investment of \$20–\$500 for chemical precipitation, \$50–\$150 for adsorption, and \$50–\$200 for ion-exchange to treat 1 million liters of wa-

ter [kumargupta2012]. Performing the same analysis for yeast, estimating a metal removal capacity of 1% of metal mass per mass of yeast (typical removal capacities achieved by hyperaccumulating strains were in Chapter ??), it would take an equivalent investment of \$160 for cadmium, \$64 for mercury, and \$480 for lead to remove 1 ppm, or 100–500 times more than EPA levels for toxic levels (Table ??).

These cost were estimated by assuming 1% of metal absorbed per mass of yeast. However, the other methods such as chemical precipitation and ion-exchange have higher per weight ratios of metal capture of up to 2–5% (Chapter ??, ??; Appendix ??, ??) which could further reduce cost. A note, however, less toxic metals which are allowable at higher concentrations by the EPA such as iron, copper and zinc will require more investment, meaning more yeast, which may be prohibitively expensive or difficult to scale. Therefore, major applications of yeast should target more toxic yet less abundant metals such as cadmium, mercury, and lead. In addition, these targets are more favorable as yeast can be finely tuned to target specific and more toxic metals than compared to physicochemical methods. Physicochemical methods lack such fine tuning and should instead be used for monolithic metal removal, or as a concluding treatment after yeast processing to remove any lingering contaminants.

(a)						
	method	cost (US\$)				
	chemical precipitation	\$ 20–500				
	ab(d)sorption	\$ 50–150				
	ion-exchange	\$ 50–200				

metal	EPA limit (ppm)	metal mass (g)	yeast mass (g)	culture volume (L)	cost (US\$)
Cr	0.1	100	10000	20000	\$ 3,200
Mn	0.05	50	5000	10000	\$ 1,600
Fe	0.3	300	30000	60000	\$ 9,600
Co ⁴	0.1	100	10000	20000	\$ 3,200
Ni	0.1	100	10000	20000	\$ 3,200
Cu	1.3	1300	130000	260000	\$ 41,600
Zn	5	5000	500000	1000000	\$ 160,000
As	0.01	10	1000	2000	\$ 320
Cd	0.005	5	500	1000	\$ 160
Hg	0.002	2	200	400	\$ 64
Pb	0.015	15	1500	3000	\$ 480

Table 5.1 | Estimate cost to process 1 million liters of waste using yeast, compared to current physicochemical platforms (a) Cost of physicochemically processing 1 million liters of waste waters in the United States [kumargupta2012]. (b) Theoretical cost of scaling and using engineered yeast to process 1 million liters of waste waters.

Rather than scaling yeast production from scratch, what if yeast could be syphoned from the already robust yeast-producing consumer market. In a completely hypothetical scenario, if the entire yeast produced from the beer industry were to be

⁴Cobalt toxicity limit was not defined by the EPA, so threshold was inferred from its toxicity relative to the other metals.

used for metal removal, then approximately 1 million tons of metal could be removed (Appendix ??). Assuming this 1 million tons came from water contaminated with 1 ppm of metal (roughly the metal content in the Athabasca Oil Sands; Chapter ??), then approximately 9.75 trillion liters of water could be remediated. To put this in a relative context, this is equivalent to processing 3.9 million olympic size swimming pools⁵. In the United States, there are only 309,000 swimming pools, meaning that this quantity of water could be processed ten-times over in a given year⁶. On a micro-analysis, this means that given 1 liter of yeast almost 50X more volume of waste can be processed at 1 ppm metal contamination (Appendix ??).

5.4 Conclusion

The overarching theme of this work was to take a well known biological platform—yeast—and transform it to a bioremediation agent for heavy metal removal. Principles from physicochemical technologies (Chapter 1) as well as processes found in plants and bacteria were engineered in yeast to endow novel waste removal functionality. Three principal methods: chemical precipitation, absorption, and ion-exchange were recapitulated in yeast using current genetic and protein engineering techniques. Direct analogies to these physicochemical processes were chemical precipitation using H₂S generated from the yeast sulfate assimilation pathway (Chapter ??), absorption via native and non-native engineered membrane transporters (Chapter ??), and ion-exchange using supramolecular metal chelating proteins (Chapter ??) (Table ??; Figure ??). The discoveries made in this work are not limited to just these three examples. A variety of other methods, combinations of methods, and higher level programming through genetic circuits and sensing mechanisms can further augment yeast’s capabilities to remove heavy metals. Simply put, this work showed that the power of biological engineering with the power to bake and brew yeast can together

⁵olympic size swimming pools typically contain 2.5 million liters: <https://www.livestrong.com/article/350103-measurements-for-an-olympic-size-swimming-pool/>

⁶census on pool and spa numbers: <https://www.thespruce.com/facts-about-pools-spas-swimming-safety-2737127>

be harnessed for waste water remediation.

Technology aside, yeast was chosen because of its global presence as a staple consumer good, and the industrial market has proven itself as a sustainable product that can be scaled, packaged, and distributed to all parts of the world. These practical considerations were ultimately the deciding factor as to why yeast was chosen as the model organism for waste water cleanup; as the discoveries made in this lab are intended to be translated to real-world applications by leveraging the yeast market. However, the future waste trends and upcoming technological advancement will decide whether or not this work can be integrated into current waste treatment practices. Whether or not society succeeds or fails to bring a stop to the uncontrolled production of waste, the underlying intention of this work was to provoke a new idea and to excite readers that many more possibilities and solutions exist in waste treatment technologies if only we, the public and scientific community, open it up for discussion.

Appendix A

Relevant yeast values and calculations

A.1 Upper limit of yeast display capture

The amount of molecules (or moles) a single yeast displaying strain can bind to can be calculated based on the number of displayed binding moieties (N_E), the number of binding sites per moiety (N_B), and the bound perfect occupancy. Assuming perfect binding (all binding sites are occupied regardless of ligand concentration), and using OD_{600} to indirectly calculate the number of cells per culture (N_{cells}), the equation for the upper limit of yeast display capture can be calculated by,

$$N_{\text{binders}} = N_E \times N_B \quad [\#/L] \quad (\text{A.1})$$

$$\bar{N}_{\text{cells}} = \text{OD}_{600} \times \lambda_{\text{OD}} \quad [\text{cells}] \quad (\text{A.2})$$

$$\bar{N}_{\text{bound}} = \bar{N}_{\text{cells}} \times N_{\text{binders}} \quad [\#/L] \quad (\text{A.3})$$

$$\bar{M}_{\text{bound}} = \bar{N}_{\text{bound}} / N_A \quad [\text{M}] \quad (\text{A.4})$$

where variables annotated with \bar{X} represent bulk values for the total yeast culture, whereas all other variables are relative to a single yeast cell. Relevant variables are:

N_E = 1e3–1e6, number of displayed moieties per yeast

N_B = 1–..., number of binding sites per domain

N_{binders} = number of metals bound per cell

$$\begin{aligned}
\text{OD}_{600} &= 0 - \dots, \text{optical density measured at } 600 \text{ nm} \\
\bar{N}_{\text{cells}} &= \text{number of cells per culture density} \\
\lambda_{\text{OD}} &\approx 1e7 \text{ cells per mL, ratio of optical density to} \\
&\quad \text{OD}_{600} \\
N_A &= 6.022e23, \text{ Avagadro's number, molecules per} \\
&\quad \text{mole} \\
\bar{M}_{\text{bound}} &= \text{total metal capture per OD}_{600}
\end{aligned}$$

Using frequently achievable values in experimental settings, such as an expression level of 100,000 metal binding domains per cell [**kieke1999**], 1 binding site per domain, and a typical OD_{600} of 1, yields an astonishingly low ≈ 2 nanomolar (10^{-9}) of bound metals. If these parameters were pushed to an extreme, using an expression level of 1 million, 10 binding sites per domain, and an OD_{600} of 10 (either by growing to saturation or packing yeast) would yield ≈ 2 micromolar (10^{-6}) of bound metals, or 3 orders of magnitude more (Table ??).

OD_{600}	expression (#)	binders (#)	capture (#)	capture (Molarity)
1	1E+5	1	1E+12	2E-9
1	1E+6	1	1E+13	2E-8
1	1E+5	10	1E+13	2E-8
10	1E+5	1	1E+13	2E-8
1	1E+6	10	1E+14	2E-7
10	1E+6	10	1E+15	2E-6

Table A.1 | Number of metals bound given yeast display parameters. Molarity of metal removed using yeast display ranges from nanomolar (10^{-9}) to micromolar (10^{-6}). These values are 3–6 orders of magnitude smaller than typical ion-exchange capacities if comparing yeast display as a biological analogy [**barakat2011new**].

Therefore, in order to obtain environmentally relevant values, the number of metals bound would have to increase by another 3–6 orders of magnitude (in the high μM to mM range).

metal	method	reported (nmol/mg)	expression (#)	ref
Cd	YT	27.10	1.01E+9	[kuroda2003bioadsorption]
Cd	YT	16.60	6.20E+8	[kuroda2003bioadsorption]
Cd	YT	10.00	3.73E+8	[kuroda2006]
Cu	YT	1.70	6.35E+7	[kuroda2001cell]
Cu	YT	25.80	9.63E+8	[ruta2017]
Zn	YT	48.80	1.82E+9	[ruta2017]
Cd	BT	1.10	1.41E+6	[pazirandeh1998]
Cd	BT	15.00	4.40E+6	[sousa1996enhanced]
Cd	BT	7.00	2.06E+6	[sousa1996enhanced]
Cd	BT	1.00	2.94E+5	[sousa1996enhanced]
Cd	BT	93.75	2.75E+7	[yoshida2002]
Cd	BT	14.90	4.37E+6	[kuroda2003bioadsorption]
Cd	BT	6.40	1.88E+6	[kuroda2003]
Cu	BT	0.55	7.04E+5	[pazirandeh1998]
Cu	BT	19.20	5.64E+6	[yoshida2002]
Hg	BT	1.30	1.66E+6	[pazirandeh1998]
Hg	BT	17.30	5.08E+6	[bae2003]
Hg	BT	3.10	9.10E+5	[bae2003]
Hg	BT	12.98	3.81E+6	[bae2003]
Pb	BT	0.95	1.22E+6	[pazirandeh1998]
Zn	BT	51.54	1.51E+7	[yoshida2002]

Table A.2 | Back-calculating cell surface display removal capacities citing previously published metal removal results. YT = yeast display. BT = bacterial display. Using metal removal values reported in previous literature, the amount of displayed groups are beyond what is typically seen in cellular display technology (tens to hundred thousands) by 1–3 orders of magnitude (calculations ranging from millions to billions). Therefore, past reports of cellular display mediated metal removal could have been overestimated possibly due to background binding or cellular uptake.

To do so would require massive optimization in protein expression, designing proteins with multiple binding sites, and improving yeast culture densities. Given these circumstances, yeast display is unfortunately not a viable method for significant metal capture, and publications that have shown promising results may be observing other binding phenomenon such as non-specific cell surface adsorption or absorption into the cell (Table ??).

A.2 Upper limit of yeast metal absorption

Rather than using cell display technologies, cell volume is a much greater container for substances than its surface area given the surface-to-volume ratio. This exercise is to estimate the bulk uptake capacity of yeast as a whole, not considering the biological impact of cell death, cytosolic metal binding, or metal trafficking into other cytoplasmic organelles. To estimate the theoretical bulk capacity maximum limit of yeast uptake is to understand the geometry of yeast (diameter, d , hence volume, V), number of cells (i.e. cell culture density, OD_{600}), and nominal uptake values (Chapter ??). To determine the upper limit requires a top-down approach. The strategy is to fix a metal uptake concentration to then calculate the internal metal concentration per yeast and assessing the feasibility given typical metal concentrations found in a cell. The metal uptake concentration is then changed, and this process is iterated as necessary until a physically plausible metal uptake concentration range is established.

Assuming a metal uptake concentration of \bar{M}_{uptake} , the amount of metal atoms per yeast can be calculated by

$$\bar{N}_{\text{atoms}} = \bar{M}_{\text{uptake}} \times V \times N_A \quad [\#/L] \quad (\text{A.5})$$

$$\bar{N}_{\text{cells}} = \text{OD}_{600} \times \lambda_{\text{OD}} \quad [\text{cells}] \quad (\text{A.6})$$

$$N_U = \frac{\bar{N}_{\text{atoms}}}{\bar{N}_{\text{cells}}} \quad [\#/cell] \quad (\text{A.7})$$

where:

\bar{M}_{uptake} = 0–10 mM, range of uptaken metal concentrations

\bar{N}_{atoms} = number of metal atoms uptaken in culture

\bar{N}_{cells} = number of yeast cells

N_U = number of atoms uptaken per cell

The number of atoms uptaken can be converted to moles (mol_U) or molarity (M_U) as follows:

$$\text{mol}_U = N_U / N_A \quad (\text{A.8})$$

$$M_U = \text{mol}_U / V \quad (\text{A.9})$$

where:

$$d \approx 1\text{--}10 \mu\text{m}, \text{diameter of yeast}$$

$$V \approx 0.52\text{--}524 f\text{L} (1\text{E-}15 \text{ L})$$

For a given amount of metal uptake of the bulk culture (\bar{M}_{uptake}), an equivalent concentration of metal uptake per cell can be calculated (M_U ; Equation ??). Work by Bryan et al. provides accurate yeast morphology parameters such as average cell volume, density, and dry and wet mass weight that can be used to calculate uptake values per cell [bryan2010measurement]. Table ?? provides a list intracellular metal concentrations considering the amount of metal uptaken, culture density, and average cell volume.

Given these calculations an uptake of 1 μM would amount to 3 mM of intracellular metal content. For comparison, nominal metal concentrations in yeast for K^+ is approximately 300 mM, for Na^+ is 30 mM, and for Mg^{2+} is 50 mM [philips2015]. Therefore, hyperaccumulating yeast consume almost an order of magnitude less of

metals as it has essential salts such as potassium and sodium. However, these levels of intracellular metal content is presumably toxic. Past studies have shown that media containing metals such as cadmium and mercury at dosages in the low micromolar range are lethal [wang2012] (Chapter ??). Therefore, there must be other physiological changes in the cell that occur during hyperaccumulation, and these calculations do not consider other cellular changes such as changes in volume, mass, or density.

uptake (μM)	OD ₆₀₀	diameter (μm)	capture (moles)	internalized concentration (mM)
10	1	4	1E-15	30
50	1	4	5E-15	149
100	1	4	1E-14	298
50	1	10	5E-15	9.55
50	10	4	5E-16	14.9
50	10	10	5E-16	0.95

Table A.3 | Calculated intracellular metal concentrations after metal uptake experiments. Depending on the culture density, cell volume, and metal added to the media, intracellular metal concentrations can range from 1–300 mM given typical uptake values reported in Chapter ??.

A.3 Uptake induced density changes

The purpose of this exercise is to determine the feasibility of using density gradient centrifugation performed in Chapter ?? for screening new hyperaccumulator mutants. A secondary goal is to further elaborate on the calculations presented in Section ?? by considering other factors such as weight and volume changes during metal uptake.

To calculate the degree of density change a top-down approach, much like what was performed in Section ??, could be done to iteratively narrow in on a range of physiological plausible values for cell density changes due to metal uptake. Calculations performed consider both a fixed cell volume, and cell volume changes as a function of maintaining isotonicity for the increase in dissolved solutes (i.e. internalized metal).

Constant volume

For a fixed volume (V), cell density would increase by the additional mass accumulated from the uptaken metals. The change in mass (Δm) is determined by the amount of internalized metal (molarity, M , or atoms, N_{atoms}), and its molecular weight (MW).

If cell volume (V) remained constant,

$$\Delta m = \text{mol}_U \times \text{MW} \quad [\text{g}] \quad (\text{A.10})$$

$$\rho' = \frac{m_o + \Delta m}{V} \quad [\text{g/mL}] \quad (\text{A.11})$$

$$\rho' = \rho_o + \frac{\Delta m}{V} \quad [\text{g/mL}] \quad (\text{A.12})$$

where variable suffix's $_o$ and $'$ denote original and new values, respectively, and:

Δm = mass accumulated from uptaken metals

mol_U = moles of uptaken metal (Equation ??)

MW = metal molecular weight

m_o = original cell mass

V = yeast volume

ρ_o = original yeast density

ρ' = new yeast density

To calculate density changes, a hypothetical experimental condition of 1 OD₆₀₀, with average yeast diameter of 4 μm and density of 1.102 g/L was used [bryan2010measurement]. Overall, the contributions to density change is minimal, as experimental uptake measurements of tens of μM barely amount to a percent change, even for the heavier elements such as cadmium, mercury, and lead. At 100 μM uptake, density changes are roughly between 1–6%. These small changes

may be possible to distinguish with isopynic density gradient centrifugation, with past studies showing fractionation of cell populations with just 5% density differences [baldwin1984, kjeldsen1999subcellular]. However, experiments performed for screening yeast hyperaccumulator mutants using isopynic density gradient centrifugation did not yield consistent results, as the bands were difficult to distinguish after metal uptake experiments.

	\bar{M}_U (μM)	Δm (g)	ρ' (g/mL)	$\left(\frac{\rho' - \rho_o}{\rho_o} \right)$
Mn	10	5.49E-14	1.104	0.18%
Co	10	5.89E-14	1.104	0.18%
Ni	10	5.87E-14	1.104	0.18%
Cu	10	6.35E-14	1.104	0.18%
Zn	10	6.54E-14	1.104	0.18%
Cd	10	1.12E-13	1.105	0.27%
Hg	10	2.01E-13	1.108	0.54%
Pb	10	2.07E-13	1.108	0.54%
<hr/>				
Mn	100	5.49E-13	1.118	1.45%
Co	100	5.89E-13	1.12	1.63%
Ni	100	5.87E-13	1.12	1.63%
Cu	100	6.35E-13	1.121	1.72%
Zn	100	6.54E-13	1.122	1.81%
Cd	100	1.12E-12	1.136	3.09%
Hg	100	2.01E-12	1.162	5.44%
Pb	100	2.07E-12	1.164	5.63%

Table A.4 | Density change as a function of metal uptake given constant cell volume. It is possible to induce density changes of 1–6% given metal uptake above 10 μM with heavier elements such as cadmium, mercury, and lead.

Volume as a function of isotonicity

Holding volume constant in Section ?? made calculations more straightforward, but may not be an appropriate assumption. Rather, volume changes should also be considered, especially if the amount of intracellular dissolved solutes increases due to metal uptake. To maintain osmotic equilibrium, the increase in dissolved content would encourage diffusion of water into the cell in order to maintain isotonicity (Equation ??). Osmotic equilibrium is achieved by maintaining the osmotic pressure inside and outside of the cell by passively or actively transporting ions or water into or out of the cell.

$$\Pi_i = iC_iRT \quad [\text{atm}] \quad (\text{A.13})$$

where:

i = van't Hoff factor (assumed to be ≈ 1)

C_i = concentration of dissolved species in a cell

R = 0.08206 liter atm/mol/K, universal gas constant

T = 303 K, temperature at 30°C

A change in volume can be calculated by equating the original osmotic pressure to the osmotic pressure of cells after metal uptake and solving for the new volume V' (Equation ??).

$$\Pi_o = \Pi' \quad (\text{A.14})$$

$$iC_oRT = iC'RT \quad (\text{A.15})$$

$$\frac{\text{mol}_o}{V_o} = \frac{\text{mol}_o + \text{mol}_U}{V'} \quad (\text{A.16})$$

rearranging,

$$V' = V_o \times \left(\frac{\text{mol}_U}{\text{mol}_o} \right) \quad (\text{A.17})$$

where,

- Π = cellular osmotic pressure
- C = cellular concentration of dissolved species
- V = cell volume
- mol_o = moles of dissolved species in the cell
- mol_U = moles of metal uptake (Equation ??)

The volume change is proportional to the amount of metal taken up (mol_U) relative to the original solute concentration of the cell (mol_o). Therefore, the increase in density due to a mass increase of taken-up metals is counteracted by the volume increase in order to maintain cellular isotonicity. The new equation to calculate density change taking into account these two effects is derived in Equation ??.

$$\Delta V = V' - V_o \quad [\text{L}] \quad (\text{A.18})$$

$$\Delta V = \left(1 + \frac{\text{mol}_U}{\text{mol}_o} \right) \quad [\text{L}] \quad (\text{A.19})$$

$$\Delta m_w = \rho_{\text{water}} \times \Delta V \quad [\text{g}] \quad (\text{A.20})$$

Therefore,

$$\rho' = \frac{m_o + \Delta m_w + \Delta m_U}{V'} \quad [\text{g/L}] \quad (\text{A.21})$$

Even though the effect of metal uptake on density change is much less significant, the overall increase in mass accumulation, primarily due to the contribution of water intake, is much more significant.

The increase in mass and volume due to water is a better indicator of metal uptake and can be distinguished using rate-zonal density gradient centrifugation, rather than isopynic density gradient centrifugation. As seen in the hyperaccumulator work in Chapter ??, using rate-zonal density gradient centrifugation yeilded better results in fractionating cells with higher metal uptake content than using isopynic density gradient centrifugation.

metal	uptake (μM)	$\Delta V/V_o$	$\Delta m_{\text{U}}/m_o$	$\Delta m_w/m_o$	$\left(\frac{\rho' - \rho_o}{\rho_o} \right)$
Mn	10	11%	0.15%	11%	-0.91%
Sr	10	37%	0.24%	11%	-0.91%
Cd	10	54%	0.30%	11%	-0.91%
Mn	50	11%	0.75%	54%	-2.73%
Sr	50	37%	1.19%	54%	-2.73%
Cd	50	54%	1.52%	54%	-2.73%
Mn	100	11%	1.49%	108%	-3.64%
Sr	100	37%	2.38%	108%	-3.64%
Cd	100	54%	3.05%	108%	-3.64%

Table A.5 | Mass, volume, and density changes as a function of cellular metal uptake. Metals used for these calculations were Mn, Sr, and Cd, metals which were engineered for in Chapter ???. The contribution of mass change due to metal uptake is less significant than the mass and volume gained from the intake of water.

Appendix B

Back of the envelope calculations

B.1 Percent metal removed per yeast dry weight

The units reported in this work were μM of metal removed with respects to the quantity of cells used, i.e. OD_{600} . The units were $\mu\text{M}/\text{OD}_{600}$. However, the phytoremediation community routinely reports mass of metal removed per plant dry weight. In other words g/gDW.

In Chapter ??, the relationship between yeast OD_{600} and dry weight was derived (Supplemental Figure ??). The ratio was approximately $\sim 0.5 \text{ gDW L}^{-1} \text{ OD}_{600}^{-1}$. The reported units of $\mu\text{M}/\text{OD}_{600}$ were converted to g/gDW by multiplying the numerator's molarity by the metal's molecular weight, and multiplying the denominator OD_{600} by the conversion ratio of $0.5 \text{ gDW L}^{-1} \text{ OD}_{600}^{-1}$. The units of volume shared between the numerator and denominator cancel out.

$$= \frac{\mu\text{M} [1\text{E}-6 \text{ mol L}^{-1}]}{\text{OD}_{600}} \times \frac{\text{MW} [\text{g mol}^{-1}]}{\lambda} \quad (\text{B.1})$$

where $\lambda = \text{gDW L}^{-1} \text{ OD}_{600}^{-1}$

$$= 1\text{E}-6 \times \frac{\text{g} \cdot \text{mol} \cdot \text{mol}^{-1} \cdot \text{L}^{-1}}{\text{gDW} \cdot \text{OD}_{600} \cdot \text{OD}_{600}^{-1} \cdot \text{L}^{-1}} \quad (\text{B.2})$$

canceling units

$$= \frac{g}{gDW} \quad (B.3)$$

A one-to-one comparison of hyperaccumulating thresholds and metal removal quantities per yeast dry weight are tabulated in Table ???. These values were derived from data in Chapters ??–???. Specifically Figure ??, Figure ?? for yeast induced H₂S metal precipitation, and Figure ??, ??, ??, ?? for hyperaccumulating uptake.

	Cr	Mn	Fe	Cu	Zn	As	Cd	Hg	Pb
MW	51.9	54.9	55.8	58.9	63.38	74.9	112.41	200.5	207.2
low	0	10	10	50	10	0	60	40	60
high	25	500	30	100	50	25	100	60	80
hyper threshold	0.10%	1.00%	0.10%	0.10%	1.00%	0.01%	0.01%	0.10%	1.00%
percent capture (low)	0.00%	0.11%	0.11%	0.59%	0.13%	0.00%	1.35%	1.60%	2.49%
percent capture (high)	0.26%	5.49%	0.56%	1.18%	0.63%	0.32%	2.25%	2.41%	3.32%
mutants	Sul1, Sul2	S*BC, S*BCT	FTR1, FTR4, S+C1	CTR1, CTR3, S+C1	n/a	Pho84, Pho87, Pho89	S*BCT, mCd, ΔM17	ΔM17, ΔH2, ΔC4, ΔHM	ΔM17, ΔH2, ΔC4, ΔHM

Table B.1 | Percent weight of metal captured compared to hyperaccumulating thresholds.

The amount of metal removed represented by units g/gDW may be deceiving as the molecular weight of the metal plays a significant role in determining hyperaccumulation. For example, an equal molar uptake of Zn and Hg would result in an almost 3-fold increase in g/gDW reported uptake for Hg, as Hg is three times heavier than Zn. More so, hyperaccumulation thresholds are based on the metal's toxicity

levels. So hyperaccumulation for Mn and Zn are much more difficult to achieve than for Hg and Pb, even if on a molecular standpoint the same quantity of atoms were to be removed.

The same removal per weight analysis was performed for the heavy metal protein chelation strategy used in Chapter ???. The units reported were μM of metal captured per μM protein (typically $100 \mu\text{M}$). To convert to g/gDW with respect to protein mass the numerator's molarity was multiplied by the metal's molecular weight, and the denominator's molarity was multiplied by the protein's molecular weight. The units of volume shared between the numerator and denominator cancel out. Specifically Figure ??, and ?? for proteins pyrG+6xHis, glnA+6xHis, and glnA+CaM and glnA+MT1A were converted from μM of metals removed to $100 \mu\text{M}$ protein used to g/gDW.

$$= \frac{\mu\text{M} [1\text{E}-6 \text{ mol L}^{-1}]}{\text{mM} [1\text{E}-3 \text{ mol L}^{-1}]} \times \frac{\text{MW}_{\text{metal}} [\text{g}_{\text{metal}} \text{ mol}^{-1}]}{\text{MW}_{\text{protein}} [\text{g}_{\text{protein}} \text{ mol}^{-1}]} \quad (\text{B.4})$$

$$= 1\text{E}-3 \times \frac{\text{g}_{\text{metal}} \cdot \text{mol} \cdot \text{mol}^{-1} \cdot \text{L}^{-1}}{\text{g}_{\text{protein}} \cdot \text{mol} \cdot \text{mol}^{-1} \cdot \text{L}^{-1}} \quad (\text{B.5})$$

canceling units

$$= \frac{\text{g}_{\text{metal}}}{\text{g}_{\text{protein}}} \quad (\text{B.6})$$

On average, the removal quantities for proteins were roughly less than 2% of total protein dry weight, almost a percent less than yeast g/gDW values (Table ??). Although it may seem that proteins should have a higher metal capture per mass, on a per weight standpoint proteins such as pyrG and glnA are almost 3-orders of magnitude heavier than the metal's captured (metals being tens of daltons; proteins like pyrG in the kilo-daltons).

	Mn	Co	Ni	Cu	Zn	Cd	Hg	Pb	Ca
MW	54.94	58.93	58.69	63.546	65.38	112.411	200.592	207.2	40.078
pyrG (62.5 kDa)	92.7 0.08%	45 0.04%	204.7 0.19%	462.7 0.47%	413.1 0.43%	259.7 0.47%	669.7 2.15%	387 1.28%	- -
glnA (54 kDa)	341.3 0.35%	130 0.14%	186.3 0.20%	228.3 0.27%	699.7 0.85%	372.7 0.78%	166 0.62%	433.3 1.66%	- -
+CaM (70.3 kDa)	- -	- -	- -	- -	- -	- -	- -	- -	226.3 0.13%
+MT1A (58.9 kDa)	172.3 0.16%	108 0.11%	137.3 0.14%	634.3 0.68%	813.7 0.90%	541.7 1.03%	313.3 1.07%	482 1.70%	- -

Table B.2 | Percent weight of metal captured for engineered proteins pyrG, glnA and their derivatives.

Yeast may have a higher per weight capture ratio because the mechanisms of H₂S metal precipitation and metal trafficking are not limited to ligand-binding stoichiometry. In other words, H₂S precipitation solely relies on the production of H₂S and these quantities could exceed 1000 ppm, or equivalently more than 30 mM in the culture headspace (Figure ??). Metal internalization and trafficking is also not a stoichiometric process. Transporters are effectively catalytic, moving metals from the extracellular to intracellular space as a function of transporter expression (Figure ??). Therefore, metal removal is only limited by the available space and changes in cell volume (Appendix ??). However, because cellular display strategies rely on stoichiometric binding of metals to displayed metal binders, the per weight metal capture can be extremely low in comparison (Appendix ??).

B.2 Cost of yeast production

An analytical breakdown for cost and scaling biological production was studied by Harrison et al. [harrison2015]. The main components comprising yeast culture are glucose, yeast extract, buffers, nutrients, and water. The estimated cost of each component to grow 1 kg of yeast is shown in Table ???. Cumulatively, the cost is a little more than \$3 per kg of yeast dry weight.

To make a cost comparison at the laboratory scale, nutrients such as amino acids (CSM)¹, and yeast buffers (YNB)² were purchased at 100 g quantities, and common stocks such as glucose³ and water were purchased as the maximum bulk orders, then the cost to grow 1 kg of yeast would take approximately \$12.62 (Table ??).

item	laboratory cost	industrial cost
glucose	\$ 0.20	\$ 0.50
yeast extract	\$ 2.12	\$ 0.50
buffers/nutrients	\$ 9.30	\$ 2.00
water	\$ 1.00	\$ 0.01
total	\$ > 12.62	\$ > 3.01

Table B.3 | Cost to grow 1 kilogram of yeast. Comparison of the cost to scale yeast in a laboratory setting, where resources are sourced from different vendors at bulk scale and low price points.

In Harrison et al.'s bioseparation analysis, typical yeast cultures at saturation can be as dense as 100 OD₆₀₀. Assuming a yeast culture to dry weight ratio of 0.5 gDW L⁻¹ OD₆₀₀⁻¹ (an assumption also used in Section ??), it would take approximately 20 liters to achieve 1 kg of yeast dry weight (Equation ??).

¹purchased from Sunrise Scientific: <https://sunrisescience.com/shop/growth-media/amino-acid-supplement-mixtures/csm-formulations/csm-powder-100-grams/>

²purchased from Fisher Scientific: <https://www.fishersci.com/shop/products/bd-difco-dehydrated-culture-media-yeast-nitrogen-base-without-amino-acids-6/p-4901538>

³purchased from Millipore-Sigma: <https://www.sigmaldrich.com/catalog/product/sigma/g8270?lang=en®ion=US>

$$= \frac{1 \text{ [kg]} (1000 \text{ [g]})}{0.5 \text{ [gDW L}^{-1} \text{ OD}_{600}^{-1}] \cdot 100 \text{ [OD}_{600}\text{]}} \quad (\text{B.7})$$

eliminating units

$$= \frac{1000 \text{ [g]}}{50 \text{ [g L}^{-1}\text{]}} \quad [\text{L}] \quad (\text{B.8})$$

$$= 20$$

If it costs approximately a little more than \$3 to generate 1 kg of yeast, and 1 kg of yeast requires approximately 20 liters, then the cost per liter would approximtely take 16 cents (if rounding up).

B.3 Potential scale and impact of yeast towards global wastewater remediation

To begin, a few assumptions based on the data represented in this work, and nominal values typically found in industry, are made to perform the calculations below. These assumptions are that yeast cultures are grown to saturating densities of approximately 10 OD₆₀₀, and that each cell has a 1% metal removal capacity per dry weight (Section ??). At 1 ppm of metal contamination (typical metal contamination levels found in the Athabasca Oil Sands; Figure ??), a single liter of yeast would be able to fully remediate 50 L of contaminated waters. A 1 to 50 ratio on a per volume basis (Equation ??).

$$V_{\text{yeast}} \text{ [L]} \cdot 0.5 \text{ [gDW L}^{-1} \text{ OD}_{600}^{-1}\text{]} \cdot 10 \text{ [OD}_{600}\text{]} \cdot 1\% \left[\frac{\text{g}}{\text{gDW}} \right] = 1 \text{ [ppm]} \cdot V_{\text{water}} \quad (\text{B.9})$$

Rearranging to solve for volume

$$\frac{V_{\text{water}}}{V_{\text{yeast}}} = \frac{0.5 \cdot 10 \cdot .01}{1 \times 10^{-3}} \quad (\text{B.10})$$

$$= 50$$

The global production of yeast from the beer industry was 1.95 billion hectoliters (195 billion liters) in 2017 [barth-haasgroup.2019]. If yeast cultures were approximately 10 OD₆₀₀ during industrial fermentation, this translates to 975 billion grams, or 1.07 million tons of yeast produced per year. Assuming a nominal per weight metal removal of 1% (Table ??, ??), this translates to approximately 107 thousand tons of metals removed. If these metals come from a contaminated water source of 1 ppm contaminants then this would equate to 9.75 trillion liters of remediable water (Equation ??).

$$975 \times 10^9 [\text{gDW}] \cdot 1\% \left[\frac{\text{g}}{\text{gDW}} \right] = 1 [\text{ppm}] \cdot \text{volume} [\text{L}] \quad (\text{B.11})$$

solving for volume

$$\text{volume} = \frac{975 \times 10^9 [\text{g}] \cdot 1\%}{1 \text{ ppm}} [\text{L}] \quad (\text{B.12})$$

$$= 9.75 \times 10^{12}$$

More than 9 trillion liters of water could be processed given the quantities of yeast produced by the global beer industry. This equates to processing almost 3.9 million Olympic size swimming pools⁴. In the United States, there are roughly 309,000 swimming pools, with another 10 million smaller residential pools⁵. With the mass

⁴average Olympic pool volume is 2.5 million liters: <https://www.livestrong.com/article/350103-measurements-for-an-olympic-size-swimming-pool/>

⁵census data on US pool sizes and number: <https://www.thespruce.com/facts-about-pools->

of yeast produced per year, it would be possible to clean almost all swimming pools in the United States. These calculations provide a theoretical underestimation. The above calculations do not include the yeast production from the craft industry which produced roughly 5.09 billion liters in 2018 [**demetergroup.n.d.2014**], the consumer market such as yeast packets and bread making, and the pharmaceutical industry.

B.4 Considerations and limitations

The calculations above are mainly to prove a point, that using the current yeast market may be a viable way to scale yeast-based technologies for bioremediation purposes. Sections ?? and ?? only give crude estimates to the economical and scaling implications of processing waste water using yeast production quantities typical found in industry. One major assumption was the cost of yeast per liter. These calculations only consider the chemical cost; however labor cost, infrastructure maintenance, and pre- and post-processing such as quality control and testing should also be factored in. Other assumptions needed, but go beyond the scope of this work, is the logistical cost of transferring the technology in the brewing and baking industry to an operable waste-processing facility. New ideas on how to convert or modify already in-place infrastructure to accommodate wastewater flow should be considered.

There are also scientific assumptions that need to be further validated. The volume to dry weight ratio of yeast cultures should be thoroughly validated, especially for specific yeast strains or at different stages of yeast growth (lag, mid-log, saturation, diauxic shift). The next is the nominal culture density that yeast could grow at scale. In these calculations, a culture density of 10 OD₆₀₀ was used. Typical saturating OD₆₀₀s in the lab can range between 2–10. The upper OD₆₀₀ level of 10 was used in these calculations because it was assumed that brewing and baking fermentors routinely push cultures to their saturating limits. On the same vein, the remediation performance, whether H₂S production, metal transport, or protein production need to be tested at these extreme culture densities and conditions. Performance

may decline as cells become more stressed at higher culture densities given the lower nutrient and higher biological waste-product content. If so, further metabolic engineering may need to be done to counter the loss of function, or a different culturing strategy may be required. A final consideration is the background chemical content that yeast are normally grown in. As in, the high salt and mineral content. These chemicals, although necessary for cell growth, also need to be removed after waste water processing. This is where a merge of biological and physicochemical processes may be synergistic. Physicochemical processes may be useful for removing residual content in the yeast culture after the yeast have removed the bulk of the heavier elements. This may be economically advantageous since yeast can be more specific and cost-effective at remediating less abundant but more toxic elements such as Cd, Hg, and Pb (Chapter ??).

Appendix C

De-novo design of metal binding moieties using machine learning

C.1 Inspiration from bioinformatics and computational modeling

Advances in software engineering normally found in data science, robotics, and telecommunication have been increasingly adopted in the biological sciences for -omic level analysis. The field of biology, traditionally thought of as a ‘wet’ science where discoveries are empirically derived, is now benefiting from software engineering to automate and process the now numerous datasets coming from genomics, proteomics, metabolomics, and so on. This new field, known as bioinformatics, uses data science techniques to analyze datasets from the life sciences to make conclusions which were typically reserved for empirically conducted experiments. In particular, statistical and machine learning methods are routinely employed in biological datasets to either uncover insights or train algorithms. The intention is to make conclusions, create trends, derive robust machine learned models to analyze future or unknown datasets. Many bioinformatic applications are directed towards diagnostics, or developing new drugs with predicted antibody-ligand binding interactions [larranaga2006machine].

There are two factors that contribute to the success of bioinformatic analysis.

The first is the quality and size of the dataset. The onus is primarily on the experimental work, such as preparing, processing, and measuring samples. For genomics, this is typically whole-genome sequencing. The rise in new sequencing technologies and knowledge of new culturing and cell isolation methods have contributed to the widespread success of genomic-based bioinformatics. The second factor is the efficiency and accuracy of the algorithm being used to analyze such datasets. New software tools and pipelines are constantly being made and improved upon¹, but the real crux of all bioinformatic tools is the number of biological assumptions that are made and whether or not these assumptions or models accurately describe the dataset. With poor assumptions, such as fitting a linear curve to a trend that is non-polynomial, bioinformatic methods would add little value to data interpretation. In the worst case, the results would be coherent but entirely inaccurate, causing a cascade of poor decisions inspired by errors.

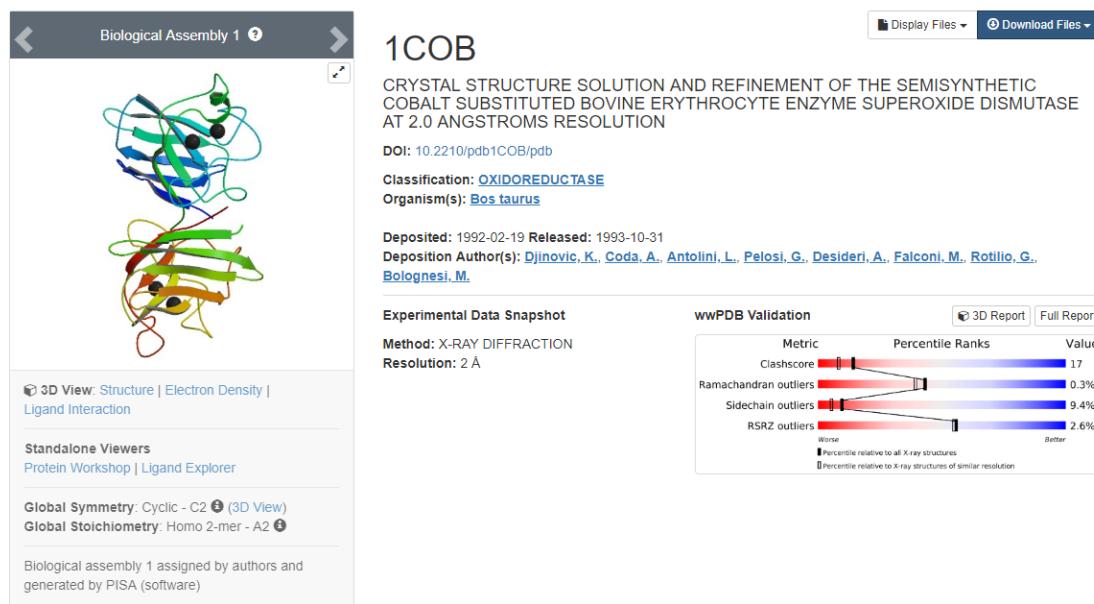
These types of errors are a result of either underfitting, or overfitting. Underfitting a dataset fails occurs when a model fails to sufficiently predict or align with the trends or patterns of the dataset. This may be caused by poor model selection (e.g. linear versus non-linear models), or too few or too general variables. What may be worse than failing to construct a model is to construct one that is wrong. This type of error is overfitting, and may occur when there are too many parameters (i.e. assumptions) that are tweaked to force a fit that is not inherent in the dataset. What happens is a propagation of inaccurate conclusions for subsequent datasets, as an overfitted model will most likely not fit other datasets with its own nuances.

For the purpose of this work, the question is whether or not current datasets and bioinformatic tools are capable of modeling protein-metal binding interactions. The goal is to determine which configurations of amino acids confer the greatest metal-binding affinity. In other words, which sequence of amino acids is capable of capturing heavy metals from the environment with high specificity for bioremediation purposes. Rather than using well-known metal binding moieties for metal capture (examples in

¹one out of many curated list of bioinformatic software tools: <https://github.com/danielecook/Awesome-Bioinformatics>

Chapter ??), is it instead possible to create de-novo peptide sequences tailored for a specific metal of interest? Already, many scientists in the protein engineering space know that the 6xHis tag or cysteine rich domains are good metal binders. Given that humans have learned some basic intuition on metal-peptide binding patterns, can a machine also learn these patterns?

To begin, a dataset of known metal-binding proteins should be derived or curated. Such dataset exists crystallographically thanks to the efforts of the Protein Data Bank (PDB) where solved protein crystal structures are publically deposited in an online server² [berman2008] (Figure ??). In the PDB, more than 30% of all proteins contain some metal-binding domain [waldron2009], and these 3-dimensional interactions could be used to generate datasets for statistical analysis and machine learning on protein-metal binding interactions. A filtered PDB only containing structures with metal ligands is called the metal PDB (*m*PDB)³ [andreini2013]. The *m*PDB provides additional metadata on the metal-protein structure such as metal location, possible amino acid binding partners, and parameters relevant to the quality of the 3D crystal structure (Figure ??).



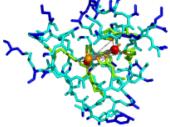
²PDB url: <https://www.rcsb.org/>

³metalPDB url: <http://metalweb.cerm.unifi.it/>

Metal Site

Click on the Image to run Jmol

PDB Chain	Molecule Name	Organism Name	UniProt Id	EC Number
1cob_A	Superoxide dismutase [Cu-Zn]	Bos taurus	P00442	1.15.1.1

Site Name	Nuclearity	Location	Physiological Relevance	Site Image
1cob_1	Dinuclear	Within a Chain	Modified Physiological Site (Substituted)	

Metal	Metal Id in PDB	Coordination Number	Coordination Geometry	Endogenous Ligands	Exogenous Ligands
Copper (Cu)	CU 152(A) CU	5	irregular (n/a)	HIS_44(A), HIS_46(A), HIS_61(A), HIS_118(A)	HOH_237(A)
Cobalt (Co)	CO 153(A) CO	5	irregular (n/a)	HIS_61(A), HIS_69(A), HIS_78(A), ASP_81(A)	-

CATH Id	SCOP Id	Pfam Domain
2.60.40.200	b.1.8.1	Sod_Cu

Figure C.1 | Overview of protein database resources. The Protein Data Bank (PDB) contains all published protein crystal structures and downloadable files. File formats are text readable .mmCIF or .pdb formats which contain experimental metadata and atomic locations. The metal PDB (*m*PDB) only contain entries with metal ligands. Entries also provide additional data derived from the PDB files such as metal location, metal site environment, and metadata of the protein origin using other database servers such as Uniprot and Pfam.

The goal is to somehow reorganize the 3-dimensional information of each protein structure into a processable dataset for statistical analysis. Once an adequate dataset is derived, the next step would be to apply statistical or machine learning framework to discover patterns between metals and their preferred (or most occurring) protein-binding environment. It may not be necessary to deconstruct the entire protein structure, but to only analyze the vicinity around the metal binding site. Within these binding sites, assumptions around the relevance of each parameter (i.e. features) can be made, such as the importance of the distances between amino acids, identity of the amino acid, types of bonding, bonding geometry, and so on. With the appropriate machine learning tools it may be possible to train a model to understand these features and how it may impact metal binding affinity. The potential results can help answer two specific questions: 1) given an unknown binding site, how likely is it to be a metal binder, and if so, which metal? and 2) given a desired metal, what is the optimal sequence and arrangement of amino acids that confer the greatest binding affinity?

C.2 Methods, algorithms, & data processing

The interface between biological data and designing an analytical pipeline to consume, transform, and interpret such data is possibly one of the most difficult steps in bioinformatics. The question arises as to how biological data, which is typically unordered, noisy, and sometimes uninterpretable by both machine and human, can be extracted of its fundamental features and then analyzed as a sequence of numbers or categories. Throughout these transformations, a caution is to maintain as much biological relevance without introducing contrived data that is not present in the original dataset, or eliminating/ignoring relevant data for the sake of efficiency or simplicity.

C.2.1 Extracting data from the metal PDB

List of protein structures with metal ligands was extracted from the *mPDB* by automating a webscraper that fetched all protein files that contained either Ag, Al, Au, Ba, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, Hg, In, K, Li, Mg, Mn, Na, Ni, Pb, Pd, Pt, Rb, Sr, W, and Zn (there were no files that contained As, Si)⁴.

The data for each file contained a PDB ID (a unique 4 digit alphanumeric code representing the protein structure), a number next to the PDB ID representing the metal instance (one protein can contain multiple metal binding sites, hence sufficing the ID code with incrementing numbers); the metal or metals found in that binding pocket; a string denoting the molecule which the metal belongs to (often it is just the metal, but for example a heme group would be included because of its Fe center) the numeric location of the metal (all atoms and molecules in a PDB file format are numbered), the chain letter, and the metal identity; and finally the ligands determined by the *mPDB* to be significant binding partners. These metal binding partners were delimited by a 3 amino acid or molecular code, followed by the numeric location of the binding partner, followed by a chain letter in which the binding partner was part

⁴Rare earth metals for Ce, Er, Eu, Gd, Ho, La, Lu, Os, Pa, Pr, Re, Sm, Ta, Tb, U, Yb were also extracted but not used in this analysis.

of in the overall protein structure. The data structure for the raw data taken from the *mPDB* can be found in Table ??.

	SiteName	Metal(s)	Metal(s) in pdb file	Ligand(s)
<i>format</i>	PDBID_no.	metal _x	molecule_loc(chain)_metal	residue_loc(chain)
<i>example</i>	1apq_2	Cu	CU_125(A)_CU	TYR_76(A), HIS_105(A)
<i>example</i>	1arm_1	Hg, Cu	CU_315(A)_CU HG_310(A)_HG	GLU_270(A), HOH_320(A) HIS_69(A), GLU_72(A), HIS_196(A), TRS_319(A), HOH_320(A)

Table C.1 | Data format of protein files from the *mPDB*. Data from the *mPDB* of each metal instance contained the PDB ID, the identity of the metal for that instance, the molecular location of that metal, and the molecular location of the ligands that bind to that metal.

C.2.2 Filtering and cleaning data

Although the data from the *mPDB* contained valuable information, it does contain missing entries, redundancies, and is of a format which is difficult to parse for basic analytical pipelines that expect numerical values or consistent categorical strings. Therefore, a filtering step was performed to remove any erroneous entries, and the data was transformed to fit a particular format more amenable for downstream analysis (Figure ??). *mPDB* entries were further validated by ping’ing the original PDB database to double check the existance of the protein structure.

The first filter was to remove any entries with multiple metals per metal binding site, as this would confuse the analysis as to which metal was more significant in the metal binding pocket. The second filter was whether or not the PDB file exists. Unfortunately some PDB files were either removed, archived, or did not exist when checked on <https://www.rcsb.org/>. The third filter was whether or not the

metal PDB data contained valid annotations of where the metal resides, both in the chain lettering and metal numbering. In some cases, some annotations in the *mPDB* were incorrect, and some metals did not exist in a small fraction of provided protein structures.

During the filtering and cleaning step, more features were extracted from the *mPDB* and PDB database. Useful features such as metal valency, metal binding geometry, the geometry quality (distorted or regular) were extracted. Additional metadata was extracted to paint a better picture of the protein (although not necessary for the analytical framework) such as UNIPROT ID, organism, and enzyme commission annotations (example of the new dataset structure is tabulated in Table ??).

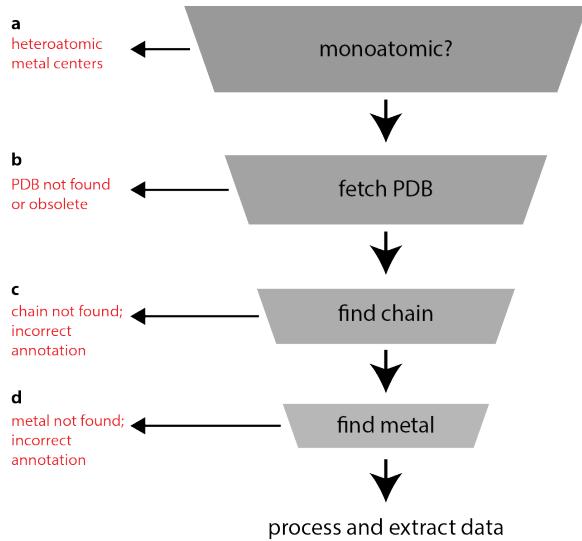


Figure C.2 | Pipeline for parsing and extracting PDB information from the metal PDB. (a) Entries that contained multiple metal atoms per metal site were removed to avoid confusion during analysis. (b) PDB files that were unfetchable (e.g. obsolete, archived, or non-existent) were ignored. (c, d) Annotations provided by the metal PDB were sometimes incorrect. Often this would occur with wrong chain numberings or non-existent metal locations. This filtering step was performed after retrieving the protein structure from the PDB (<https://www.rcsb.org/>).

(a)					
	ID	instance	chain	location id	metal id
<i>format</i>	PDBID	metal number	A, B, ...	loc(chain)	metal atom location
<i>example</i>	1q06	1	B	300(B)	24985
(b)					
	protein	organism	Uniprot	EC	
<i>format</i>	UNIPROT protein name	UNIPROT protein name	Uniprot accession number	enzyme commission number	
<i>example</i>	HTH-type transcriptional regulator CueR	Escherichia coli	P0A9G4	-	
(c)					
	valency	geometry	idealized	ligands	
<i>format</i>	0–9	linear, trigonal, ...	-, distored, regular	residue_loc(chain)	
<i>example</i>	2	linear	regular	CYS_112(B), ...	

Table C.2 | Tranformed dataset structure of protein entires filtered from the *mPDB*. (a) The original metal PDB entries were cleaned to individually represent the PDB ID, metal instance for that PDB file, chain location, molecular location of the metal, and atomic location of the metal. (b) Additional metadata was extracted such as the UNIPROT ID, organism name, accession number, etc. (c) Categorical data such as metal valency, metal binding geometry, and ligands were also taken from the metal PDB.

C.2.3 Creating datasets processable by machine learning algorithms

The filtered and cleaned data from the *mPDB* had yet to undergo another transformation in order to be inputted into a statistical or machine learning framework. To create such an input–output pipeline the data structure was converted to numerical or categorical features. More so, the previously processed features were not feature rich

(Table ??), as in, they did not quantitatively explain the metal-protein structure relationship in detail. An example would be the sparse 1–5 amino acid binding partner description per metal coordination sphere. Therefore, a new algorithm was developed to extract 3-dimensional data from the protein crystal structure which bared more fine-grained atomic data. The algorithm developed was a variation of the nearest-neighbor algorithm [arya1998optimal] in which an imaginary radii stretched from the metal center is modelled to extract the closest molecule/amino acid from that metal center (Figure ??). The discovered neighbors were further processed to determine which atom from that molecule were closest, and the distance between the atom and the molecule from the metal center were also calculated⁵.

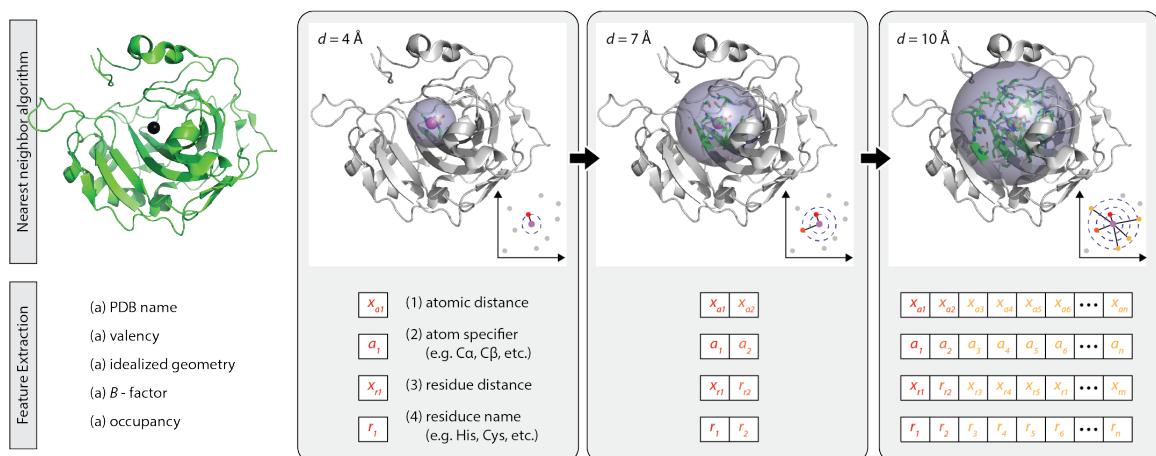


Figure C.3 | Variation of the nearest-neighbor algorithm to extract molecular and atomic neighbors from the metal center of a protein structure. Molecular and atomic neighbors at incremental radii away from the metal center were identified using the nearest-neighbor algorithm and stored in arrays. These arrays were later concatenated to create either a 2D dataset or 1D row of features. Additional parameters were extracted during the algorithm. These parameters were the valency and geometry of the metal-binding environment, the metal's β -factor and its percent occupancy.

The features collected were the atoms closest to the metal center, and their distances away from the metal center. The same process was performed for nearest neighbor molecules and their distances. These features were ordered from closest to

⁵Bio.PDB package in python was used to handle PDB structural data: https://biopython.org/wiki/The_Biopython_Structural_Bioinformatics_FAQ

furthest, the closest being 1 Å and furthest being 50 Å away. All features were concatenated to create a 1D array with each analyzed protein structure representing a row of a larger 2-dimensional data structure. These datasets were segregated by metal’s analyzed. For example, there exists separate datasets of acquired nearest-neighbor information for Li, Na...Pb, where each row of each dataset represents the features extracted from a single protein structure.

In addition, several metadata parameters were included in the feature list. They were the metal’s β -factor and percent occupancy. These values helped score the confidence of the metal location in the protein structure, and future work would use these values to under-weigh or ignore potential outliers or bad instances.

(a)								
	metal	ID	valency	geometry	idealized	β factor	occupancy	anisotropy
<i>example</i>	Ag	1q06	2	linear	regular	26.6	1	-1
(b)								
	atom name	atom distance (Å)		molecule name	molecule distance (Å)			
<i>example</i>	SG, CB, CB, CB, CA, O	2.35, 3.2, 3.37, 3.5, 3.51, 3.6	CYS, CYS, CYS, SER, CYS, SER		4.17, 3.51, 4.17, 3.88, 3.51, 3.88			

Table C.3 | Training set data structure as input to machine learning frameworks. (a) Several metal-protein specific parameters such as the metal’s valency and geometry, as well as its β -factor and percent occupancy were collected. (b) Arrays of atom names, atom distances, residue names, and residue distances sorted from closest to furthest from the metal center were concatenated into a single array.

C.3 Results

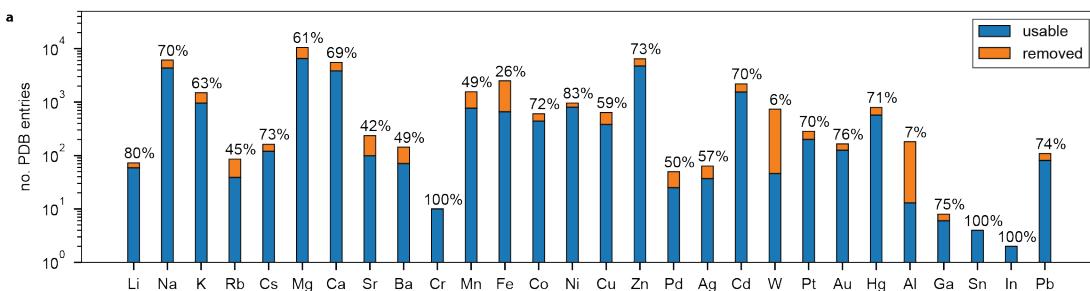
C.3.1 Analysis of protein-metal data curated by the *mPDB*

A significant portion of the curated protein structures from the *mPDB* were not usable, primarily because there were multiple metals per metal binding site, and some of the chain and metal location annotations were incorrect. For metals that

had more than 10 PDB entires, the filtering step removed almost 25–40% of entires (Figure ??a).

The metals with the most PDB entries were Na, Mg, Ca, and Zn. However, alkaline and alkaline-earth metals like Na and Mg may be over-represented because these metals are frequently present in buffers during crystallization. More so, during PDB submission many of the crystal structure solute environment remains annotated, scoring hundreds to thousands of spectator ions. Unfortunately, the *mPDB* does not filter these entries, and in this work these entries propagated through the analysis. To remove these false-positive metal-bound proteins would require differentiating metals in the buffer from metals bound to the protein. To do so would require accessing the PDB structure directly and querying every metal. Alternatively, a threshold could be set that if a protein contains more than X number of metals, specifically Na, Mg, etc., then it should be eliminated. However, this assumption is crude, and may falsely eliminate good protein structures.

The fact that many metals in the protein structure entries were solutes rather than bound metals help explains the high β -factors and low percent occupancy for most of the alkaline and alkaline-earth metals. Also, many metals which are rarely found in proteins such as the metalloid and noble metals like Pt and Hg had poor β -factors and percent occupancies (Figure ??b). When looking at the metal-bound protein structures holistically, on average each structure contains 2 or less binding sites (Figure ??c). In other words, it is likely that a protein structure containing a particular metal will have one or two binding pockets for that metal.



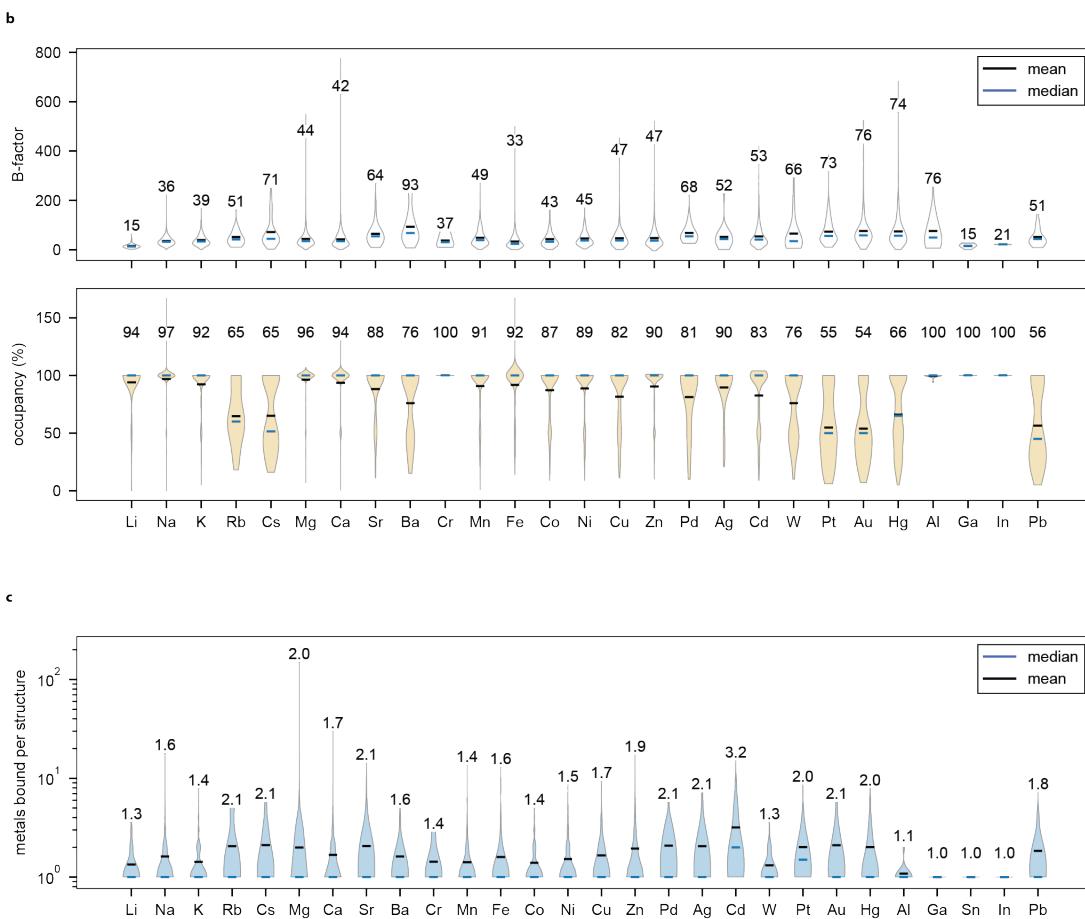


Figure C.4 | Number of usable metal PDB entries and their statistics.

(a) Several protein structures provided by the *mPDB* were not usable and had to be filtered (see Figure ?? on filtering criteria). Note the y-axis is in log scale, and that bar heights are not linearly proportional. Some metals such as Ga, Sn, and In, particularly from the semi-metals contained very few entries (less than 10). Alkaline and alkaline-earth metals were highly represented, most likely because these metals are often present as dissolved solutes during protein crystallization. (b) Plot showing β -factors and percent occupancy of each metal. Low β -factors suggest more accurate metal position, and higher percent occupancy suggest greater confidence in metal localization. Metals that are not often found in proteins such as Rb, Cs, Hg, etc. have high β -factors and low percent occupancies, which may suggest that these are not natural or favorable binding environments. (c) Plot showing the average number of metal binding sites per protein for a given metal. Overall, the average metal-binding proteins usually contain less than 2 metal binding sites.

In most metal-binding environments many metals, especially the alkaline and alkaline-earth metals, have valencies of 1 or 2. For transition metals they may be

found in several valent states ranging from 1–7 because of the electron donating d-orbitals; however, still many transition metal co-factors are found in the divalent state such as Mn^{2+} , Fe^{2+} , Zn^{2+} etc. So it was surprising to see that the valencies calculated by the *mPDB* were not between 1–2, but rather varying dramatically.

no.	val.	abbrv.	full geometry name	instances	percent
1	0	-	-	4086	15%
2	0	IRR	irregular	7530	28%
3	2	TRV	trigonal plane with a vacancy	2209	8%
4	2	LIN	linear	230	1%
5	3	TRI	trigonal plane	214	1%
6	3	SPV	square plane with a vacancy	634	2%
7	3	TEV	tetrahedron with a vacancy	1047	4%
8	4	BVP	trigonal bipyramidal with a vacancy (equatorial)	234	1%
9	4	BVA	trigonal bipyramidal with a vacancy (axial)	505	2%
10	4	SPL	square plane	537	2%
11	4	PYV	square pyramid with a vacancy (equatorial)	1124	4%
12	4	TET	tetrahedron	1405	5%
13	5	SPY	square pyramid	1247	5%
14	5	TBP	trigonal bipyramidal	202	1%
15	5	TPV	trigonal prism with a vacancy	87	0%
16	6	OCT	octahedron	3219	12%
17	6	TPR	trigonal prism	49	0%
18	6	PVP	pentagonal bipyramidal with a vacancy (equatorial)	491	2%
19	6	CTF	trigonal prism, square-face monocapped with a vacancy (capped face)	40	0%
20	6	CTN	trigonal prism, square-face monocapped with a vacancy (non-capped edge)	95	0%
21	6	PVA	pentagonal bipyramidal with a vacancy (axial)	129	0%
22	6	CON	octahedron face monocapped with a vacancy (non-capped face)	77	0%
23	6	COF	octahedron face monocapped with a vacancy (capped face)	71	0%
24	7	HVP	hexagonal bipyramidal with a vacancy (equatorial)	49	0%
25	7	CUV	cube with a vacancy	4	0%
26	7	CTP	trigonal prism square-face monocapped	99	0%
27	7	PBP	pentagonal bipyramidal	544	2%
28	7	HVA	hexagonal bipyramidal with a vacancy (axial)	2	0%
29	7	COC	octahedron face monocapped	148	1%
30	7	SAV	square antiprism with a vacancy	83	0%
31	8	BTT	trigonal prism triangular-face bicapped	0	0%
32	8	BOC	octahedron trans-bicapped	0	0%
33	8	BTS	trigonal prism square-face bicapped	56	0%
34	8	SQA	square antiprism	79	0%
35	8	CUB	cube	3	0%
36	8	HBP	hexagonal bipyramidal	5	0%
37	9	CSA	square antiprism square-face monocapped	0	0%
38	9	TPP	trigonal prism square-face tricapped	0	0%

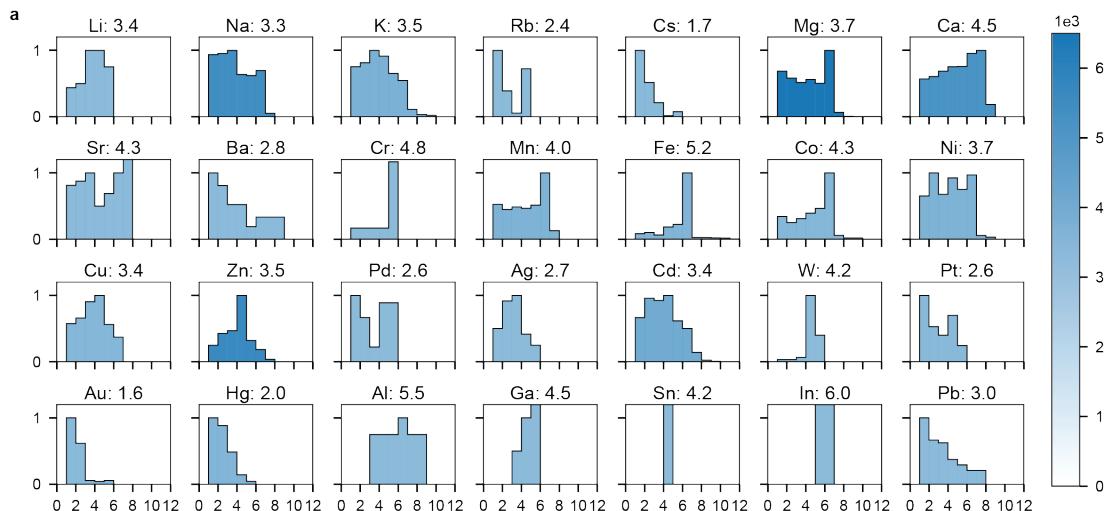
Table C.4 | Table of metal binding geometries for all metals examined from the metal PDB. Many metal geometries could not be identified, or were irregular (>40%). The most prominent geometries were trigonal planar with a vacancy (valency of 2), and octahedral (valency of 8).

Elements such as Li, Na, and K had valencies above 3, and many of the metalloids had valencies above 4 (Figure ??a). A hypothesis is that these valencies were calculated indirectly by the number of binding partners found in the protein structure. The number of metal-binders predicted by the *mPDB* may have simply counted the number of ligands and summed them to generate a valency value (Table ??; “Ligand(s)” column).

The overall representation of metal geometries were either irregular (28%), not identifiable (15%), octahedral (12%), or trigonal planar with a vacancy (8%) (Table ??). What this data suggest is that valency could be a poor identifier to distinguish different types of metal binding environments.

C.3.2 Differentiating protein-metal interactions by clustering steric and ligand data

Data strictly derived from the *mPDB* were used as input datasets for statistical and clustering analysis. The intention was to use features that were filtered and cleaned (Section ??) from the *mPDB* to help elucidate patterns that could help differentiate metals and their metal-protein binding interactions from one another. The most basic objective was whether or not the curated data from the *mPDB* could discern between alkaline, transition, metalloids, and noble metals from one another.



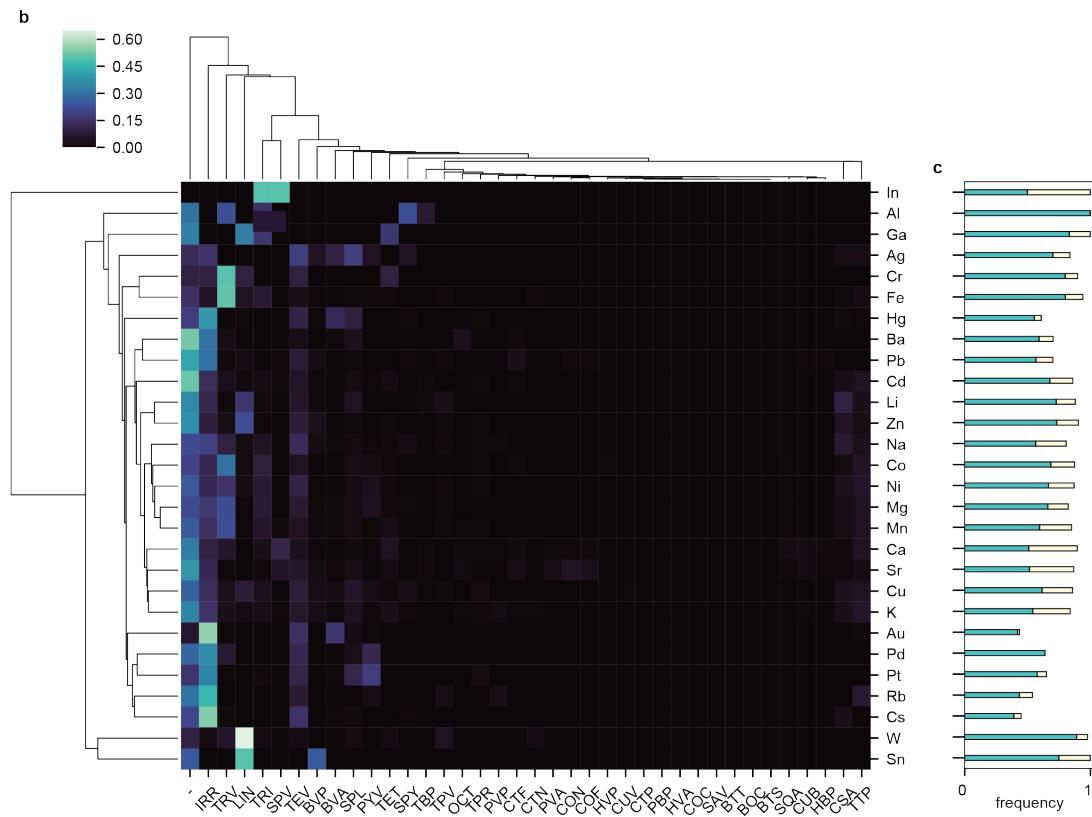


Figure C.5 | Analyzing metal valency and geometry data to determine differences between metal groups (a) Histogram of valencies for each metal. Typically, most metals found biologically have valencies between 1–2; however, data from the *mPDB* suggest valencies above 2–, on average 3. (b) Binned geometry occurrences for each metal were clustered to identify any discrimination between metal groups. There were no distinct groupings observed. (c) Each geometry was either deemed regular, distorted, or unknown. For each metal, the annotation of each geometry was summed and plotted. The horizontal bar chart represent the frequency of regular (turquoise bar), and frequency of distored (beige bar) for each metal presented in the heatmap b). The remainder up to 1 represents the frequency of unknown.

Data on the metal's sterics, such as the metal's valency and geometry, were used for clustering analysis. Clusters were generated by using a single linkage method and euclidian distance as the distance metric. Unfortunately, there were no discernable groupings of alkaline metals from transition metals from metalloids, etc. (Figure ??b). Overall, there was no consistent clustering pattern that could differentiate a metal-protein binding interaction using observations based on the steric environment alone.

Instead of analyzing the metal's sterics, the metal-binding ligands (i.e. amino acid residues) represented in each *mPDB* entry (Table ?? were used instead for clustering analysis. Although the ligand data lacked statistical power (many residue entries were zero), the clustering did show discrimination between certain metal groups (Figure ??). In particular, the majority of alkaline and alkaline-earth metals were clustered together, and some of the transition, metalloid, and noble metals were segregated with statistical significance.

These preliminary findings suggest that it may be possible to systematically differentiate metals based on their amino acid-binding environment. Simply, a straightforward counting and binning of nearby binding residues was enough to superficially differentiate metals based on their periodic grouping (Figure ??). To further this investigation, the data provided by the *mPDB* could be more feature rich if counts of neighboring residues in the metal binding environment were accounted for at varying distances away from the metal, and the same goes for nearby atoms. From here it would be possible to either construct a supervised or unsupervised machine learning model to predict the most common amino acid configuration away from the metal. With these models it may be possible to eventually develop autoencoders or generative models [goodfellow2014generative] to create de-novo metal-protein binding sites given robust predictions of the metal's nearest-neighbor data.

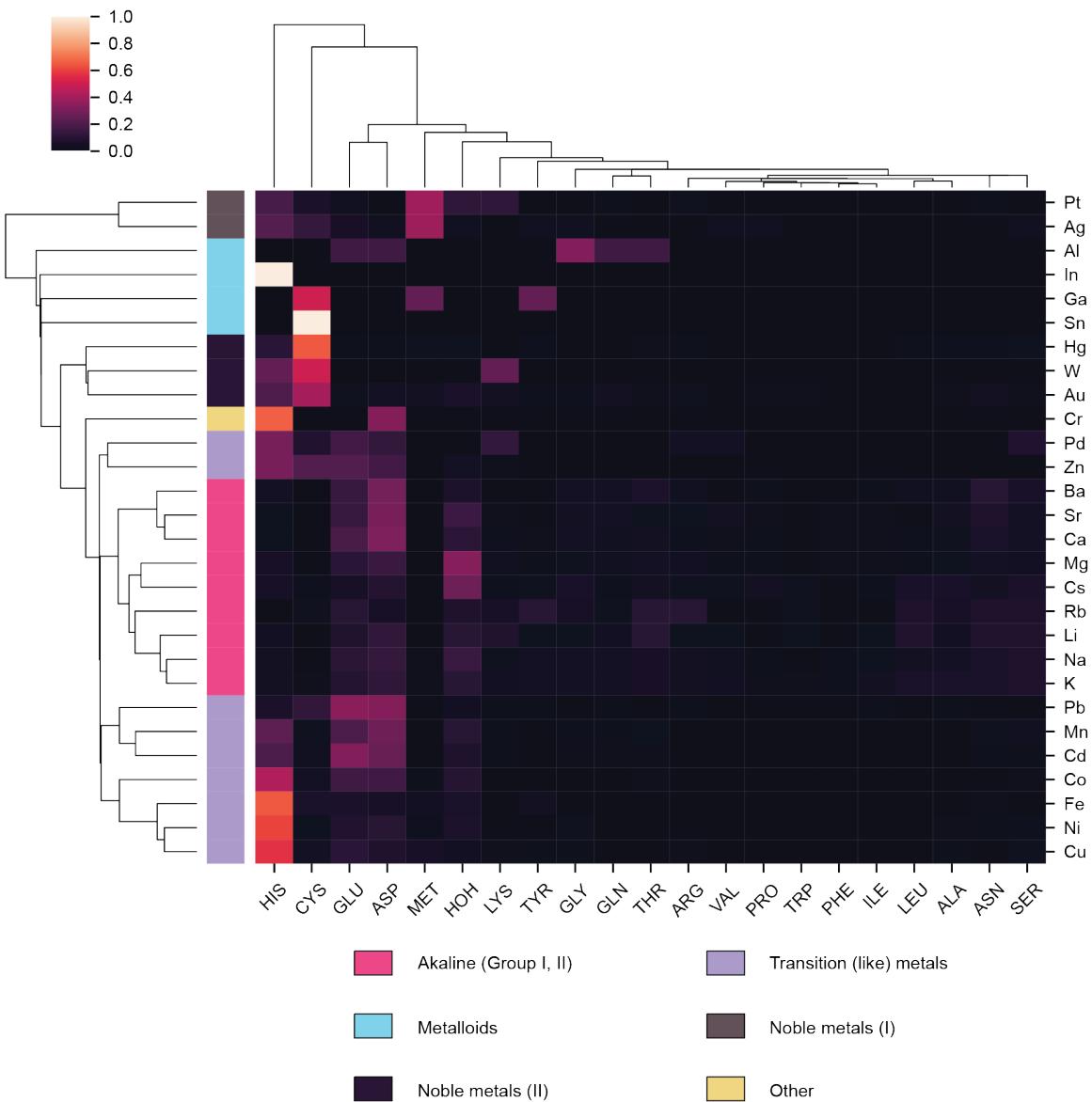


Figure C.6 | Clustering metals based on their metal-binding ligands given data from the *mPDB*. Alkaline and alkaline-earth metals clustered together (pink rows), whereas some of the transition metals (light purple rows) were separate from the metalloid (turquoise rows) and noble metals (dark pink and brown rows).

C.3.3 Extracting and engineering features for better protein-metal discrimination

Preliminary results taken only from the *mPDB* annotations provided somewhat of a glimpse as to what features were important to differentiate metals from their protein structure. Information on sterics such as valency and binding geometry could not adequately cluster metals (Figure ??); however, straightforward binning the frequency of occurrence of residues nearest the metal did product discernible clusters. The next step was to make the binned nearest-neighbor residue data more feature rich. This meant to re-process the filtered data from the *mPDB* by fetching the entire protein structure from the PDB and using the modified nearest-neighbor algorithm (Figure ??) to extract more granular data on the metal-binding environment.

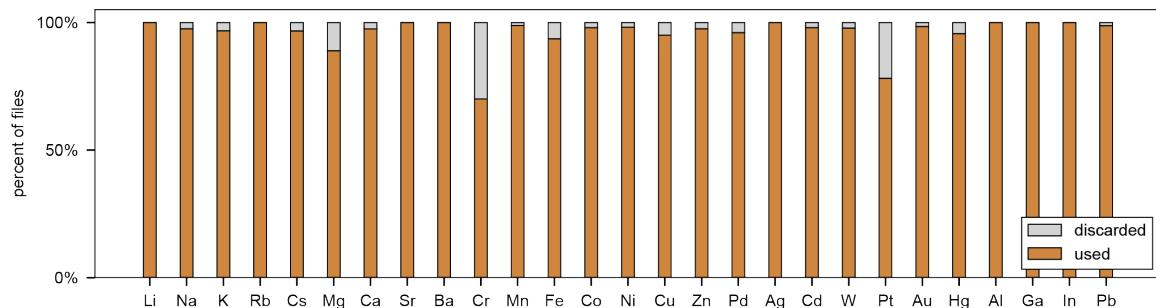


Figure C.7 | Filtering metal PDB entries for incorrectly annotated or removed PDB entries. The chain letter and metal location were used to locate the metal center from the PDB entry. However, some annotations provided by the *mPDB* were incorrect and led to wrongly assigned metal IDs or non-existent locations.

Unfortunately, this additional step did eliminate more protein entries, as some of the annotations provided by the *mPDB* were incorrect, or some protein structures were out of date (additional filtering step was discussed in Figure ??c,d). On average, approximately 2–10% of files were discarded when attempting to retrieve protein structures from the PDB using file information taken from the *mPDB* (Figure ??).

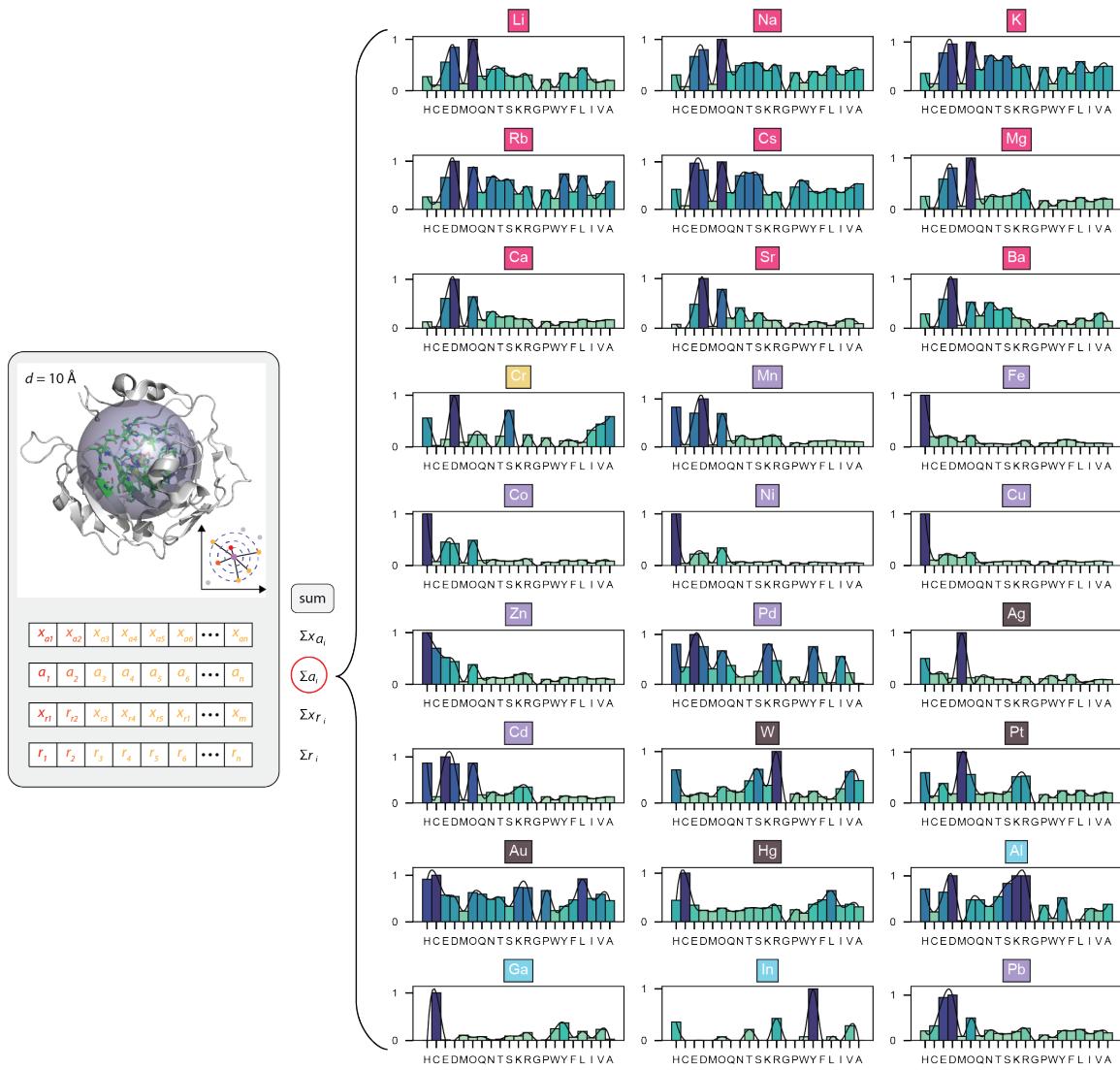
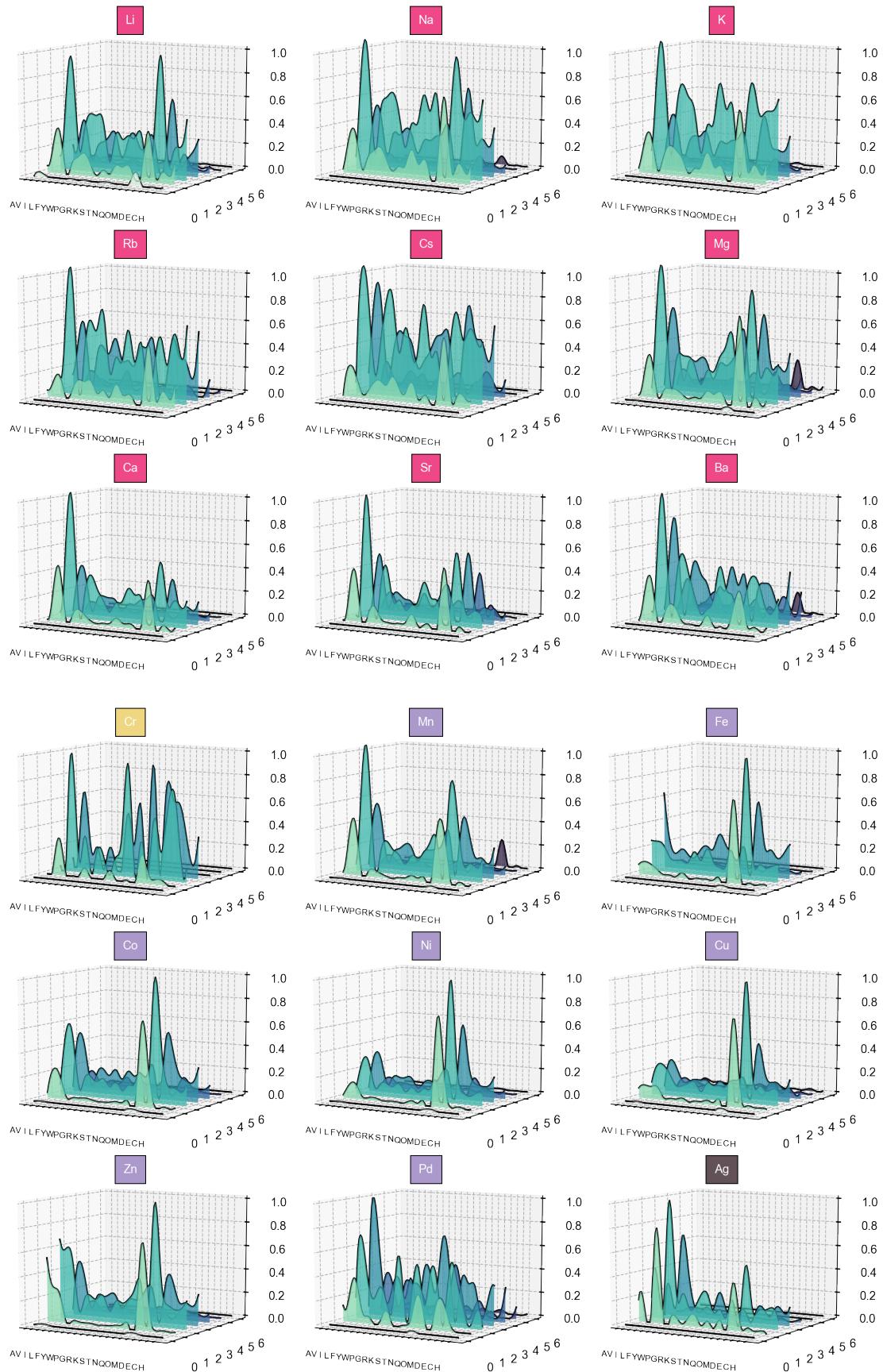


Figure C.8 | Histogram of binned residue counts at 10 Å away from the metal center. The frequency of occurrence of residues encountered 10 Å away from the metal center. The frequency plot represents a “residue profile” of the most common amino acids present in the vicinity of the metal-binding environment. Metal coloring for each plot title corresponds to the cluster they belong to, analyzed in Figure ??

From the usable files, the metal center of each PDB entry was located, and the nearest-neighbors were tabulated for a given radii away from the metal. At each radii, an array was constructed by tabulating the nearest atom, distance, and the same for residues and their distances away from the metal (the data structure format was shown in Table ??).



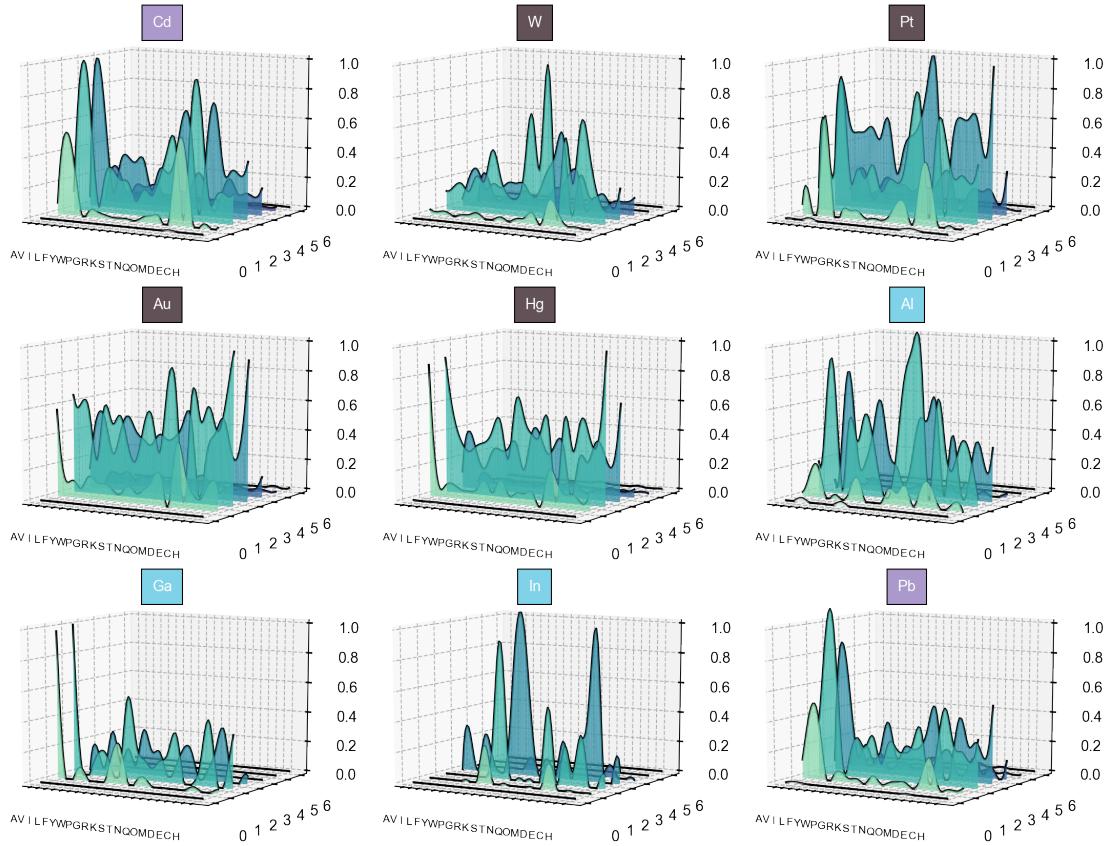


Figure C.9 | 3-dimensional histogram of residue counts as a function of distance away from the metal center. x-axis (left most axis) are single code-letterings of amino acids. y-axis (right most axis) represent the distance in Å away from the metal center. z-axis (vertical axis) represent the percent occurrence of the amino acids at the given distance away, noramlized to 1.

Each row represents a particular metal-protein interaction, and features were concatenations of atoms, residues, and their distances (in ascending order) away from the metal. Similar to counting the representation of ligands annotated by the *mPDB*, the occurrences of residues were counted and binned at each radii to create stacks of 2-dimensional histograms (Figure ??). These histograms represent a ‘residue profile’, in other words the statistical representation of residues encountered as one moves away from the metal center.

Histograms of each radii slice can be stacked together to generate a 3-dimensional plot of residues encountered (x-axis) versus distance (y-axis) versus frequency of encountered residue (z-axis) (Figure ??).

Similarly, these same plots can be flattened to 2-dimensions by color encoding the z-axis as a heatmap (Figure ??).

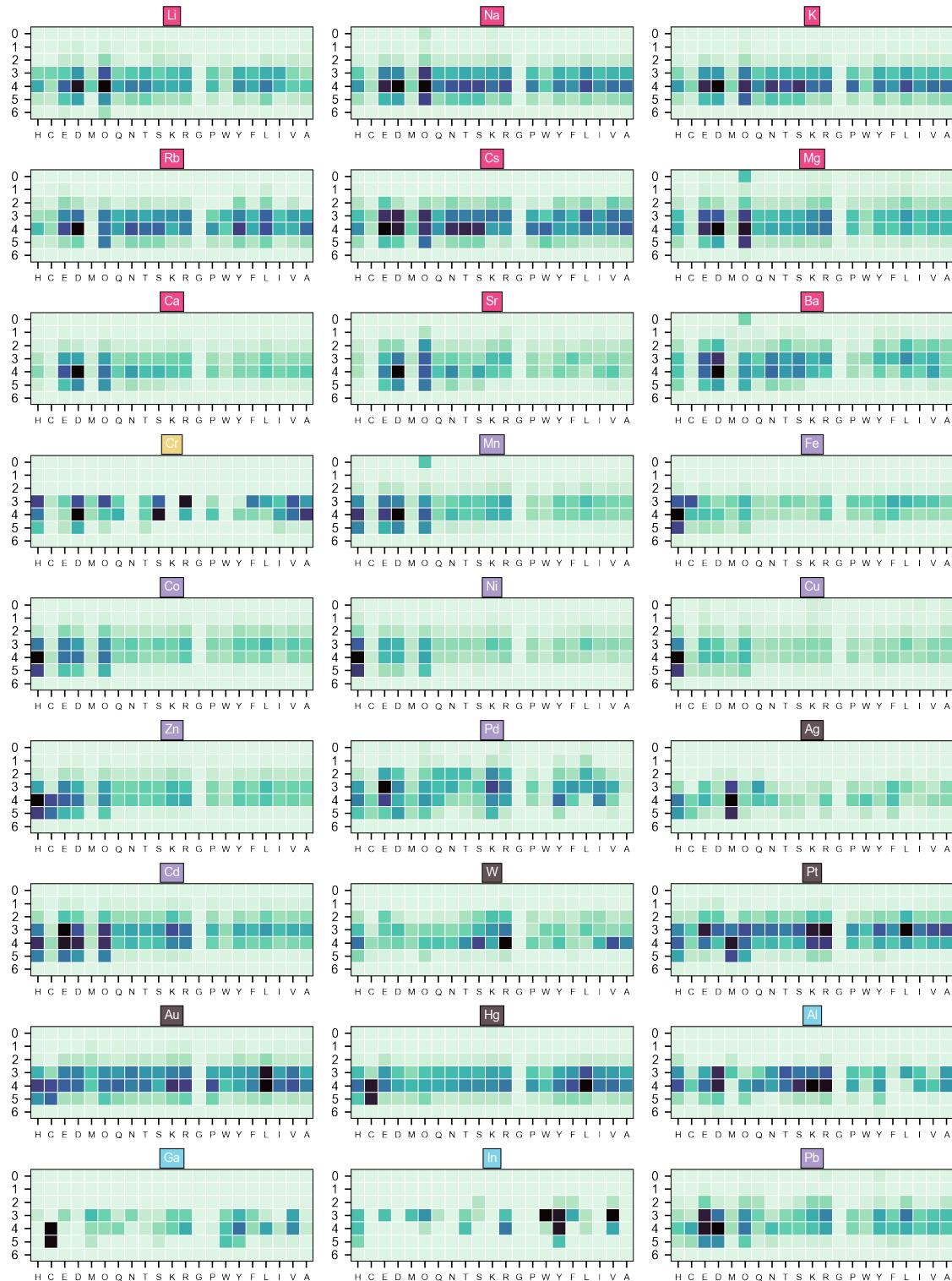


Figure C.10 | 2-dimensional heatmap of residue counts as a function of distance away from the metal center

The 2-dimensional dataset for each metal can be further flattened to a 1-dimensional array with each row concatenated into a single vector.

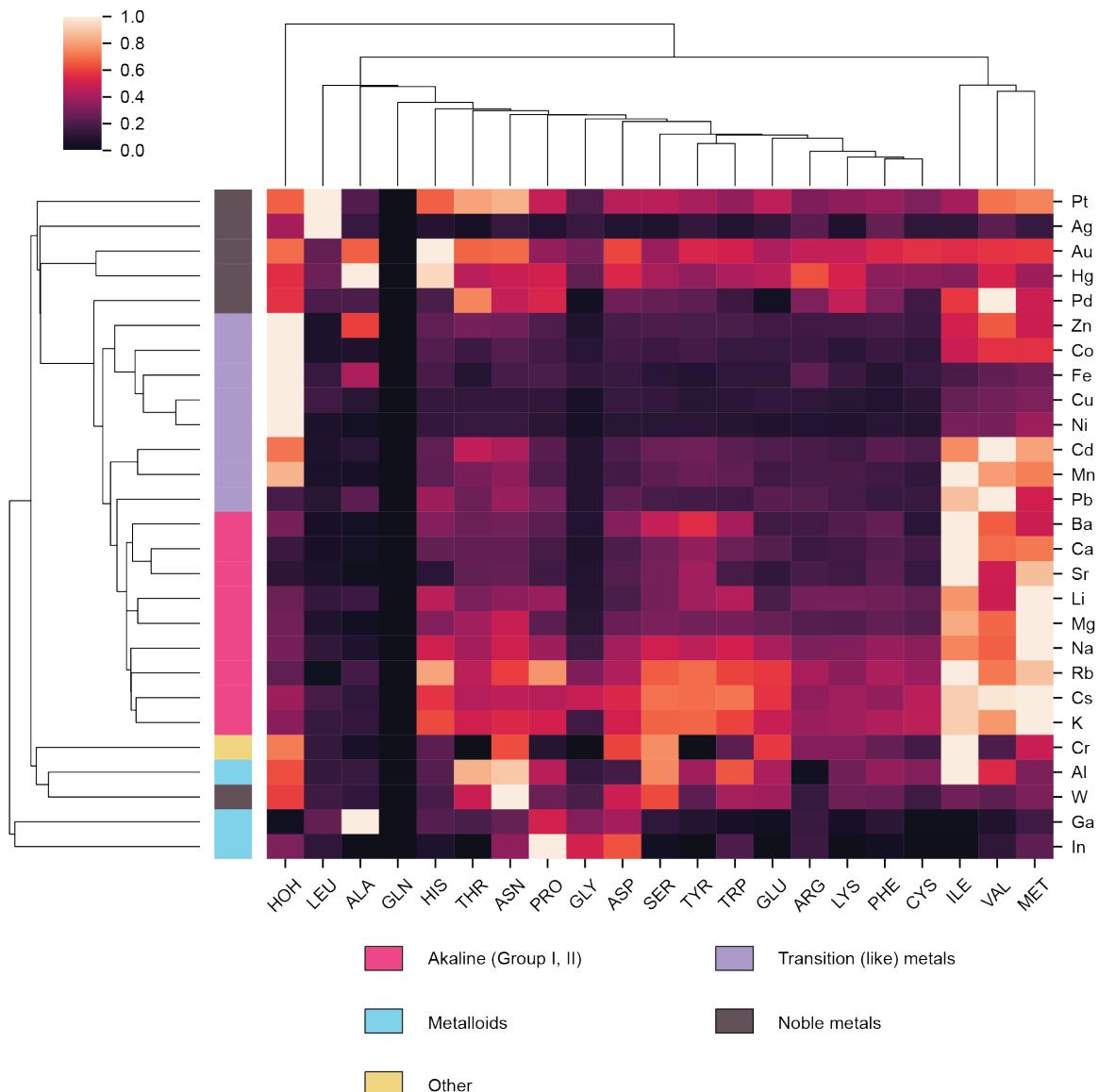


Figure C.11 | Improved metal clustering using new datasets derived from the modified nearest-neighbor algorithm. Data on amino acid occurrences as a function of distance away from the metal center provided a more data rich analysis compared to data provided from the *mPDB* (Figure ??). In addition, the segregation of metal groups was more distinct, as each periodic grouping: alkaline/alkaline-earth metals (pink rows), transition metals (light purple rows), noble metals (brown rows), and metalloids (turquoise rows), were segregated from one another with statistical significance. The only outliers were W (tungsten), which grouped into the metalloid cluster, and Cr (chromium), which was distinguished as its own group.

With this, the same clustering analysis performed for Figure ?? could be done. The result is an improved segregation of metal groups given the more feature rich dataset (Figure ??).

C.4 Statistical analysis and machine learning; future work

Future work needs to be done to further investigate different feature extraction strategies and appropriate usage of statistical and machine learning models. This work is under current investigation and is under active research. The goal of this future work is to take the insights from the previous sections and machine learning models for predictive or generative purposes. Examples include predicting the metal identity in unknown metal-binding pockets, or generating de-novo protein-metal binding environments.

Future steps to take include:

- **Further feature extraction and engineering**

Only residue counts as a function of distance were used as features for the clustering analysis (Figure ??). Yet there are many features not used, such as the distance between metal and residues, and the same for the nearest atoms. Other features can be calculated such as the euclidean angle between metal and residues, or trigonometric parameters between residues in the same binding sphere. More sophisticated 3-dimensional analytics can be performed to solve for the sphere of hydration between the metal and residues, or the accessible surface topology of the metal binding pocket. In addition, metadata such as β -factor, percent occupancy, and other intrinsic data to the PDB structure can be used to weigh specific entries such that outliers are suppressed during analysis.

- **Dimensionality reduction**

Too many features add noise or unnecessarily increase computational time. In

most cases too many features may contribute to overfitting. Therefore, dimensional reduction techniques such as principal component analysis should be used to determine the most fundamental linear combinations of features that still fully represent the dataset. Dimensionality reduction can also project the dataset onto another plane that better discriminates instances; for example, projecting a cone from the top to form can separate the heights of the cone on a 2-dimensional plane. However, a poor projection from the side would overlap the surfaces of the cone into a triangle. So testing a variety of dimensional reduction techniques should be checked before proceeding to training machine models .

- **Experiments with a variety of classifier models**

There exist numerous machine learning models, many with their own benefits and drawbacks. Each model should be tested empirically to determine which performs the best. For this purpose classifiers would be appropriate, as the model should be able to classify which arrangement of residues have the highest probability of containing a certain metal. Examples of classifiers are decision trees or random forests.

- **Hyperparameter tuning using gridsearch**

Mostly all models do not work “out of the box” and need fine tuning to adjust for the nuances of a particular dataset. These parameters, such as rate of learning, variable weights, and so on, are called hyperparameters. Often, these hyperparameters are brute-forced optimized through grid searching, where several hyperparameters are permuted for a single training run, and performed repeatedly over the entire combinatorial space of hyperparameters until a value is selected that optimizes the training output, or reduces the error.

- **Creating pipelines for ensemble learning**

Several algorithms can work in conjunction to provide better modeling and decision power. Examples include large or convolutional neural networks, algorithms sub-classified as deep learning algorithms. With these new algorithms,

the pipeline of feature engineering, dimensionality reduction, and hyperparameter tuning may need to be further optimized.

One overarching concern is whether the quantity and quality of PDB entries is enough to robustly train a model to accurately predict metal-binding interactions. So far there are 151,754 protein structures in the PDB⁶. However, some are redundant, and the quality varies dramatically between entires. It actually may be too early, or ambitious, to study every protein structure as the data is not yet sufficient to train a suitable machine model. Another setback may be the difficulty to extract useful features from a complex 3-dimensional crystal structure, and that algorithms have not yet been developed. However, the work so far suggest that it may be possible to cluster groups of similar metals together, and that may be enough to help create custom peptide/proteins with affinities for alkaline and alkaline-earth metal from the transition, noble, and metalloids.

⁶statistics on the PDB can be found here: <https://www.rcsb.org/stats>