PNAS RESEARCH ARTICLE | SYSTEMS BIOLOGY
BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Compatibility of intracellular binding: Evolutionary design principles for metal sensors

Nicolas Lenner[a,1], Logan Chariker[a,b,1], and Stanislas Leibler[a,c,1]

Contributed by Stanislas Leibler; received December 28, 2024; accepted March 31, 2025; reviewed by Yamuna Krishnan, Rob Phillips, and Nigel Robinson

In the common cellular space, hundreds of binding reactions occur reliably and simultaneously without disruptive mutual interference. The design principles that enable this remarkable compatibility have not yet been adequately elucidated. In order to delineate these principles, we consider the intracellular sensing of transition metals in bacteria—an integral part of cellular metal homeostasis. Protein cytosolic sensors typically interact with metals through three types of lateral chain residues, containing oxygen, nitrogen, or sulfur. The very existence of complete sets of mutually compatible sensors is a nontrivial problem solved by evolution, since each metal sensor has to bind to its cognate metal without being "mismetallated" by noncognate competitors. Here, based solely on theoretical considerations and limited information about binding constants for metal-amino acid interactions, we are able to predict possible "sensor compositions," i.e., the residues forming the binding sites. We find that complete transition-metal sensor sets are severely limited in their number by compatibility requirements, leaving only a handful of possible sensor compositions for each transition metal. Our theoretical results turn out to be broadly consistent with experimental data on known bacterial sensors. If applicable to other cytosolic binding interactions, the results generated by our approach imply that compatibility requirements may play a crucial role in the organization and functioning of intracellular processes.

cellular metal sensors | protein sensors | metallo-proteins | Irving–Williams series | mismetallation

Significance

Metal ions participate in the catalysis of many intracellular chemical reactions. Cells regulate metal concentrations using a set of protein sensors, each detecting a particular metal ion. These sensors not only need to be sensitive but also mutually compatible, so as not to bind any metal other than their own. Here, we show how small differences in the relative binding preferences of the amino acid residues, which form metal binding sites, are enough to distinguish among different metal ions, despite their similar sizes and charges. Within a simple theoretical model, we outline design principles underlying mutual compatibility of known bacterial metal sensors. Our approach can possibly be generalized to investigate mutual compatibility of other interactions between cellular components.

About 20 chemical elements are used as building blocks for all known living organisms. Much research has been dedicated to understanding the cycling and recycling of carbon, nitrogen, and phosphorus, which are the components of DNA, RNA, proteins, and lipids. Less is known about the regulation of the equally essential transition metals. This is surprising, since many of all known proteins are metallo-proteins (1, 2), and the presence of metal ions is required to catalyze numerous reactions in the cell (3, 4).

Transition metals bind to biomolecules with vastly different binding constants K_B (units: $1/\mathrm{M} \equiv \mathrm{L/mol}$). Seventy years ago, Irving and Williams postulated that most biomolecules follow a fixed order in the binding preferences for transition metals (5). This series is now called the Irving–Williams series and is, following the order in the periodic table, schematically represented as

$$\mathrm{Mn^{II}} < \mathrm{Fe^{II}} < \mathrm{Co^{II}} < \mathrm{Ni^{II}} < \mathrm{Cu^{I}} > \mathrm{Zn^{II}}, \text{ with } \mathrm{Ni^{II}} \approx \mathrm{Zn^{II}},$$

where the inequality signs refer to the relative values of their binding constants. Here, the superscripts denote the typical oxidation state in the reducing environment of the cytosol. For example, $\mathrm{Zn^{II}}$ binds about 10^4 times more strongly to the amino acid cysteine than does $\mathrm{Mn^{II}}$ (6). Thus, to ensure equal binding probability to a protein through cysteine, the $\mathrm{Mn^{II}}$ concentration would need to be at least a factor 10^4 times higher than the $\mathrm{Zn^{II}}$ concentration. In reality, this effect is even stronger since the binding sites for metal ions are formed by several amino acids. For instance, in *Salmonella enterica*, a factor as large as 10^{12} has been reported between the affinities of $\mathrm{Mn^{II}}$ to the MntR sensor and of $\mathrm{Cu^{I}}$ to the CueR sensor (7). This disparity in binding affinities of transition metals to biomolecules creates "a metal binding problem" (8): How can a cell ensure the correct binding of different metal ions if their binding affinities are so vastly different? Or, to give a concrete example, why are not all $\mathrm{Mn^{II}}$ binding sites permanently clogged with much more strongly binding $\mathrm{Cu^{I}}$ or $\mathrm{Zn^{II}}$ ions?

Metallo-proteins acquire their metals directly from the cytosol, in some cases helped by specialized proteinaceous metallo-chaperones (9–11). Therefore, the crucial adjustable factor that determines metal ions' binding events in a cell is their cytosolic concentration. The regulation of intracellular transition metal concentrations is a complex and rather

Author affiliations: [a]Simons Center for Systems Biology, School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540; [b]Department of Mathematics, Duke University, Durham, NC 27708; and [c]Laboratory of Living Matter, The Rockefeller University, New York, NY 10065

Author contributions: N.L. and S.L. designed research; N.L., L.C., and S.L. performed research; and N.L., L.C., and S.L. wrote the paper.

Reviewers: Y.K., The University of Chicago; R.P., California Institute of Technology; and N.R., Durham University.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: lenner@ias.edu, clc144@duke.edu, or livingmatter@rockefeller.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2427151122/-/DCSupplemental.

Published April 30, 2025.

poorly understood phenomenon (11, 12). It involves different mechanisms, such as regulation of ion efflux and influx (11, 13, 14), shuffling between different cellular compartments (13), or local buffering (15), all acting at different time scales. It is important at this point to recall that the use of concentrations might be problematic since the "free cytosolic concentrations" of Zn^{II}, Ni^{II}, and Cu^{I} can be extremely low, even below one ion per cell volume. In such situations, the metal ions that bind to sensors do not come from exchange reactions with the water dissolved pool, but rather from what is often called the "labile pool" of weakly bound ions (16). These ions are readily accessible: The exchange reactions are fast and the size of the labile pool is relatively large. It is reasonable to assume that the labile pool is equilibrated on the time scales we are considering here. As a consequence, we can use an effective chemical potential, or effective cytosolic concentrations, for each of the metal species bound to the labile pool. This is what we will assume in our modeling, although for simplicity we will use the term "cytosolic concentrations" to describe such effective concentrations.

The long-term adjustment of intracellular concentrations has been studied more extensively, on a metal-by-metal basis, in bacteria (7, 17–19). This typically involves transcriptional regulation (Fig. 1*A*) (22, 23), where protein transcription factors regulate the expression of uptake (influx) and efflux systems for transporting metals into and out of the cell. Uptake systems (22) are typically regulated by repressors: Under low buffered cytosolic metal concentrations, no metal is bound to the transcription factor, and the transcription is in its "inactive state," which, in general, implies that the transcription factor is not bound to the DNA (Fig. 1 *A*, *Top*). Under such conditions, the expression of the uptake system is enhanced, allowing an increase of the cytosolic metal concentration. Above a certain "set-point" concentration, the transcription factor binds its cognate metal (Fig. 1 *A*, *Bottom*). This triggers an allosteric response of the transcription factor, typically in the form of a conformational change. The altered conformation then allows efficient binding of the transcription factor to the DNA and, consequently, repression of the genes that control the expression of the uptake system. This state is the "active state" of the transcription factor (24, 25).

A good illustrative example is provided here by ZuR, a transcription factor that regulates Zn^{II} uptake in most-studied gram-negative bacteria (26). In its inactive state, ZuR occurs as a homodimer (27, 28). Upon increase of the cytosolic concentration, two Zn^{II} ions bind to the homodimer (one per monomer) at structurally identical regulatory binding sites (20). These two binding sites are placed at hinge-like regions of ZuR, where metal binding stabilizes its closed conformation (29). With two metals bound, the transcription factor is in its active state and binds strongly to the DNA (20). Fig. 1*B* shows one homodimer and one of the two regulatory sites (as an *Inset*) for ZuR found in *Escherichia coli*. As typical for transition metal binding sites, Zn^{II} is coordinated by nitrogen (N), oxygen (O), and sulfur (S) centers belonging to the amino acid side-chains located in the sensing binding site (Fig. 1 *B* and *C*) (30, 31).

Transcription factors that regulate metal efflux are typically less specific to the metal they bind to, and the reported regulatory principles are more diverse than those of influx systems (22). In general, the regulatory principles of transition metal homeostasis are similar to those observed for regulating small molecules. This comes as no surprise, as the transcription factors which regulate transition metals are evolutionarily related orthologs to the transcription factors that regulate the cellular abundances of small molecules (22).

It has long been assumed that cytosolic transition metal concentrations have to follow the inverse Irving–Williams series: The stronger a metal binds according to this series, the lower its cytosolic concentration has to be. Such adjustment of the cytosolic concentrations of metals effectively equalizes their binding probabilities. Deviations from these cytosolic "reference concentrations" can then be sensed in a distinctive way that depends on detailed characteristics of the sensors, such as their preferred binding geometry, the particular amino acid residues that form the metal-binding sites (which we shall call here the "sensor composition"), etc. In a beautiful and seminal work, Deenah Osman and her collaborators demonstrated that in the cytosol of *S. enterica* a set of transcription factors indeed binds their cognate metals following the inverse Irving–Williams series (5, 7). They further showed that all sensors are specific with respect to their cognate metals, assuming a cognate metal occupation of roughly 50% for each sensor (7), i.e., the cytosolic metal concentrations are kept in the range of maximal sensor sensitivity.

Osman et al. have implicitly laid bare what is required for joint transition metals sensing: Sensors must be both sensitive and specific with respect to their cognate metal to be compatible. Here, we will show that these two simple requirements impose strong constraints on the binding free energies of all sensors involved. A simple description of metal-amino acid interactions will further reveal how the requirement for compatibility can be used to make ab initio predictions about the sensor composition of each particular metal-binding site. We will test our predictions against a large set of experimentally uncovered transition-metal sensors compiled from the literature.

1. Theory

For a cell to distinguish between different transition metals, each sensor must bind to the cognate metal with a significantly higher probability than to any mismatched noncognate metal. Since sensor occupancy increases with metal concentration, each sensor can in principle be overwhelmed by high concentrations of noncognate metals. Sensors must therefore jointly regulate transition metal concentrations to protect each other from mismetallation. From an evolutionary perspective, this suggests that the cognate and noncognate binding constants of all metal sensors are jointly adapted to produce a set of mutually compatible sensors.

To formulate the constraints underlying the evolutionary search for a set of compatible sensors, we start with the two fundamental requirements for a sensor. A sensor must be i) sensitive to variations in the cytosolic metal concentrations it controls and ii) specific, so as to preferentially bind its cognate metal.

1.1. Sensitivity. In the following, we will consider a collection of m sensors indexed by subscripts $i = 1, \ldots, m$ as well as m metals indexed by superscripts $j = 1, \ldots, m$, where the cognate metal of sensor i has the index $j = i$. The sensor–metal binding probability in a reaction kinetics model for noncooperative binding for all of the metals $j = 1, \ldots, m$ to the binding site i is given by (32, 33)

$$\Theta_i = \frac{\sum_{j=1}^{m} K_i^j [M^j]}{1 + \sum_{j=1}^{m} K_i^j [M^j]}, \qquad [1]$$

where K_i^j denotes the equilibrium binding constant of the metal j to the sensor i (in units 1/Molar), and $[M^j]$ its cytosolic

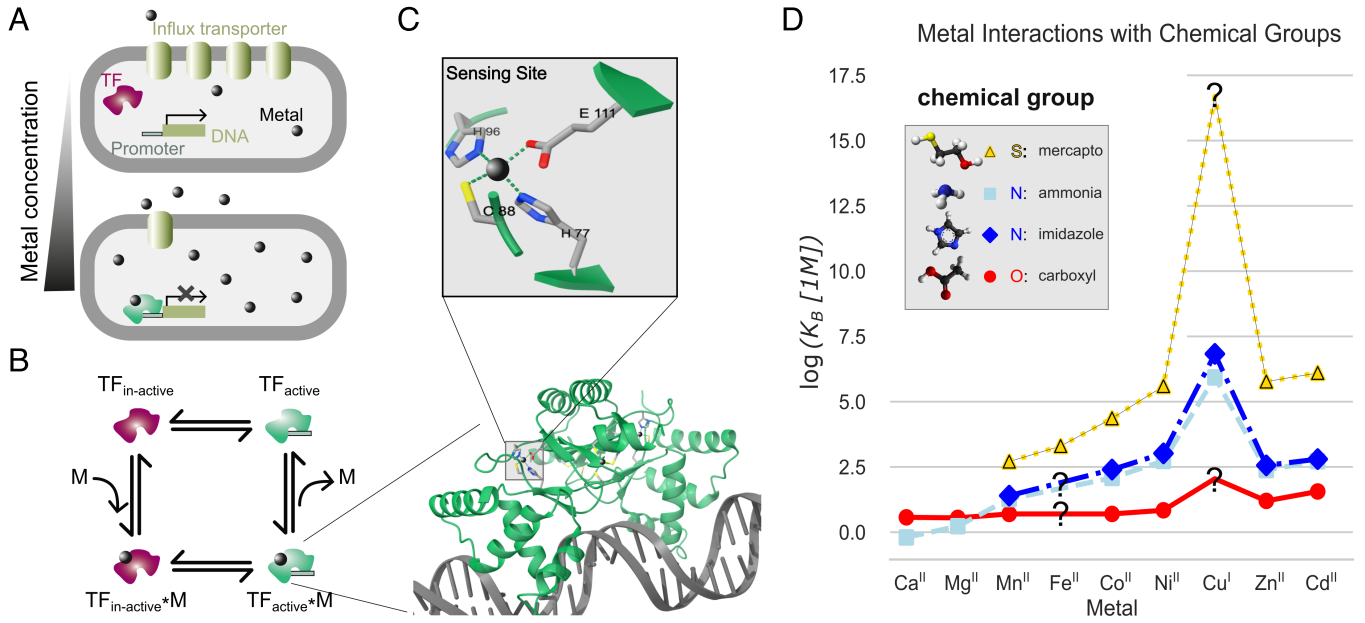

Fig. 1. Transition-metal sensing by proteinaceous bacterial transcription factors. (*A*) Schematic view of the regulation of cytosolic metal concentrations. In this example, an increase in cytosolic metal concentration leads to the repression of influx mediating genes. The repression is mediated by metal-sensing transcription factors (TFs). (*B*) Thermodynamic cycle of allosteric transition metal sensing. The binding and unbinding in the inactive state (bordeaux colored TFs) is discussed in the main text, while the whole cycle is treated in *SI Appendix*, section 14. (*C*) Metal sensing occurs at binding sites formed by the amino acid residues of glutamic and aspartic acid, histidine, terminal amine groups, and cysteine. Shown here is the main sensing site of the Zn^{II} uptake regulator ZuR of *E. coli* (20). Visualization with ChimeraX (21). (*D*) The binding strengths of biologically relevant transition metals (including Ca^{II} and MgII), to a given chemical group, follow the Irving–Williams series. Shown here are the previously measured stability constants of chemical groups that approximate the amino acid residues involved in metal–protein binding. Question marks represent model-dependent estimates (details in *SI Appendix*, section 8). For a given metal (except for Ca^{II} and Mg^{II}) we observe that the binding strengths to different chemical groups follow another order, which we term the Williams-Heuristic (WH): These metals prefer the S-group (yellow) over the N-groups (blue) over the O-group (red).

concentration. We depict schematically the one sensor–one metal version of the model in the *Left*-hand side of Fig. 1*B*. The theoretical occupation probability for a sensor i as function of its cognate metal concentration $[M^i]$ is shown in Fig. 2 *A*, *Top*. Throughout this text, for simplicity of the presentation, we ignore the additional complication of allostery. Indeed, we assume that the cell is unable to distinguish whether a sensor is bound to its cognate metal or, instead, to a noncognate metal, so that it responds only to the fraction of sensors bound to any metal. That is, we consider the case where metals compete only for the binding site in the inactive state, and that different metals have similar effects on the allosteric switching to the active state. In *SI Appendix*, section 14, we include more realistic allosteric effects and explain how "allosteric filtering" can in fact help to distinguish between bound cognate and noncognate metals. For the sake of simplicity we also ignore here any effects of cooperative binding of metal ions.

The function that quantifies the sensitivity S_i^j of a sensor i upon relative concentration changes of a metal M^j is defined as

$$S_i^j([M^j]) := [M^j]\frac{\partial \Theta_i}{\partial [M^j]}. \qquad [2]$$

Using Eq. **1**, we can express the sensitivity function as

$$S_i^j([M^j]) = \frac{K_i^j[M^j]}{\left(1+\sum_k K_i^k[M^k]\right)^2}. \qquad [3]$$

We depict the theoretical sensitivity function of cognate metal binding in Fig. 2 *A*, *Bottom*. The maximum of $S_i^i([M^i])$, the sensitivity of a sensor to its cognate metal, is found at the concentration

$$[M^i]^* = \frac{1+\eta_i}{K_i^i}, \quad \text{with} \quad \eta_i = \sum_{k\neq i} K_i^k[M^k]. \qquad [4]$$

The quantity η_i can be thought of as the ratio between the number of mismetallated and unmetallated sensors (*SI Appendix*, section 1). $[M^i]^*$ coincides with the point where the number of unmetallated sensors is equal to the number of sensors occupied by the cognate metal (Fig. 2*A*). The classical assumption of maximum sensitivity at $[M^i] \approx 1/K_i^i$ for independent sensors still holds approximately, but, more precisely, it is increased by the multiplicative factor $(1+\eta_i)$.

Clearly, the buffered metal concentration will not stay exactly at the point of maximum sensitivity, $[M^i]^*$, but most of the time it will deviate from this concentration. How big can these deviations be until a sensor loses its sensitivity? We may tentatively define the "sensitivity range" as the $[M^i]$ concentrations for which the sensitivity function stays above half of its maximum value. Deviations well outside of this concentration range would lead to loss of sensitivity and consequently to loss of regulatory control with potential system failures. From Eq. **3**, we find that this range corresponds to (*SI Appendix*, section 2):

$$\frac{1}{3+2\sqrt{2}} < \lambda_i := \frac{[M^i]}{[M^i]^*} < (3+2\sqrt{2}). \qquad [5]$$

Thus, the sensitivity range of concentrations, expressed in terms of $[M^i]^*$, does not depend on any sensor binding-constants. In other words, six-fold concentration changes roughly mark the sensitivity range for all sensors (see Fig. 2*A*, gray region).

1.2. Specificity. With the sensitivity range of a sensor thus characterized, we next define the specificity of a sensor i. The

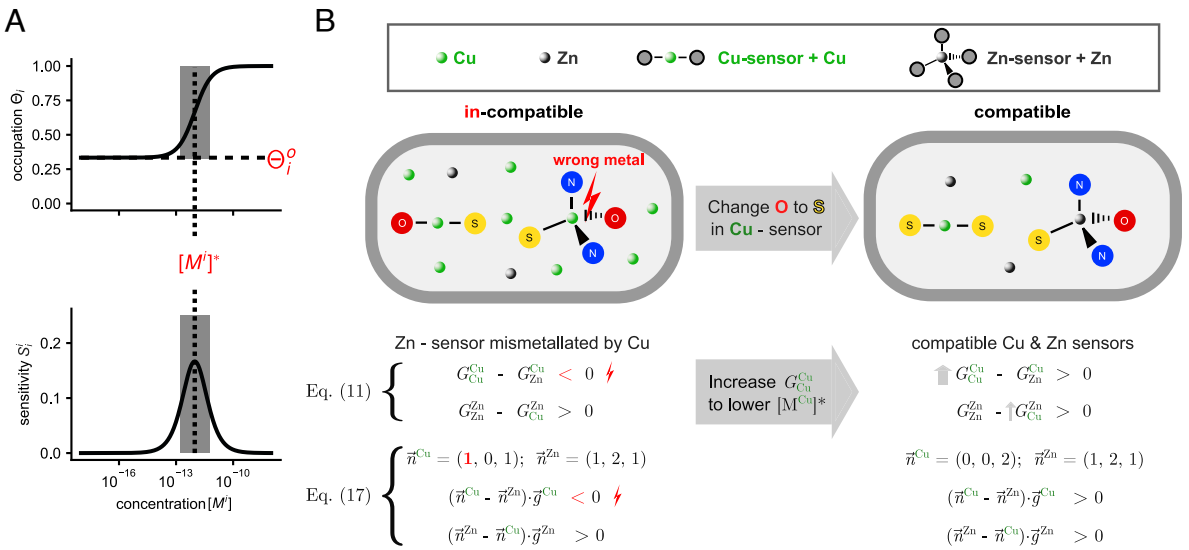

Fig. 2. Compatible transition metal sensing requires specific and sensitive sensors. (*A*) *Top* panel: The occupation probability of a sensor in response to its cognate metal concentration, with all other metal concentrations fixed. In the absence of cognate metal ions, the sensor is entirely occupied by noncognate metals at a level Θ_i^o, the baseline occupation (horizontal dashed line). Sensors are defined to be "specific" when their baseline occupation stays below 1/2. *Bottom* panel: The sensitivity function of a sensor to its cognate metal. The maximum sensitivity occurs at the concentration $[M^i]^*$, where the occupation probability is midway between Θ_i^o and 1. In both panels $[M^i]^*$ is marked by dotted lines, and a surrounding "sensitivity range" lying between two half-max-sensitivity concentrations is shaded in gray. (*B*) The amino acid configurations of all sensors (amino acids are schematically depicted as large circles) jointly define which sensors are selective for their cognate metal. Here, two hypothetical protein sensors (*Left* panel)–one for Cu^I and another for Zn^{II}–are respectively specific and nonspecific, and thus mutually incompatible. A change in the amino acid configuration of the Cu–sensor binding site (marked by an arrow) can lead, however, to a compatible set (pair) of specific Zn^{II} and Cu^I sensors. The compatibility Eqs. **11** and **17** are presented here explicitly to show how the change of the amino acid leads to the increase in the binding free energy G^{Cu}_{Cu} for the Cu^I-sensor. A consequence of this increase is lowering of the Cu^I concentration, which in turn stops the mismetallation of the Zn^{II}-sensor. Note that in this example, G^{Zn}_{Cu} also increases, but not enough to affect the second inequality in Eq. **11**.

more a sensor is bound to noncognate metals, the less it responds specifically to its own cognate concentration. The fraction of a sensor bound to noncognate metals is the largest when the cognate metal concentration is low. Therefore, as a measure of how specific a sensor can in principle bind to its cognate metal i, we consider the fraction of the sensor occupied in the absence of this metal:

$$\Theta_i^0 = \Theta_i\Big|_{M^i=0} = \frac{\sum_{j\neq i} K_i^j [M^j]}{1+\sum_{j\neq i} K_i^j [M^j]} = \frac{\eta_i}{1+\eta_i}. \qquad [6]$$

We call sensor i specific when $\Theta_i^0 < 1/2$, i.e., no more than half of the sensors are occupied by noncognate metals even when the cognate concentration $[M^i]$ is low. Θ_i^0 can also be interpreted as the limit of the response curve Θ_i as $[M^i] \to 0$, as seen in Fig. 2 *A*, *Top*.

1.3. Compatibility. To learn more about the relationship between the involved binding constants, on the one hand, and the specificity and sensitivity of a sensor, on the other hand, we derived the following lower bound on the ratio between K_j^j and K_i^j (details in *SI Appendix*, section 5):

$$\frac{K_j^j}{K_i^j} > \lambda_j\left(\frac{1}{\Theta_i^0}-1\right). \qquad [7]$$

This inequality is a direct consequence of the above definitions of η_i and Θ_i^0, without any further assumptions applied (other than $[M^j] > 0$ and $K_i^j > 0$). Taking the base-10 logarithm of both sides yields:

$$G_j^j - G_i^j > \log(\lambda_j) + \log\left(\frac{1}{\Theta_i^0}-1\right). \qquad [8]$$

Note that we replaced the logarithmic equilibrium binding constants by

$$G_i^j := \log(K_i^j[1M]), \qquad [9]$$

which we call here, somewhat loosely, "free energies" (Gibbs free energies can be obtained from G_i^j by multiplying by -2.303 RT, where R is the universal gas constant and T is the temperature).

We will now require that the sensors in the compatible set are both sensitive and specific. The particular, biologically plausible choice for these two conditions will allow us then to establish lower bounds for the difference between cognate and noncognate binding free energies G_j^j and G_i^j.

(A) For a sensor to be sensitive to its cognate metal, we require that the concentration $[M^i]$ varies around the point of highest sensitivity. That is, $[M^i]^*/\Lambda < [M^i] < [M^i]^*\Lambda$, for some $\Lambda > 1$, or, equivalently, $1/\Lambda < \lambda_i < \Lambda$ (Fig. 2 *A*, *Bottom* panel).

(B) For a sensor i to be specific against a noncognate metal j, we require at least that its baseline occupation by noncognate metals is no more than 50% when concentration λ_j is being varied in the sensitivity range of the sensor j. That is, $\Theta_i^0 \leq 1/2$, when $1/\Lambda < \lambda_j < \Lambda$.

Applying the specificity condition to Eq. **8**, we arrive at a necessary condition for (at least marginal) specificity of sensor i under perturbation of any metal j,

$$G_j^j - G_i^j > 0. \qquad [10]$$

If this condition is violated, then sensor i does not stay specific under variations in the concentration of metal j. Somewhat surprisingly, this condition for sensor i is independent of its cognate metal binding constant. Instead, it requires ensuring that other metals bind better to their cognate sensors than to sensor i. Note that this is only a necessary condition. Realistically, $G_j^j - G_i^j$ will have to be larger than a certain value of $\epsilon > 0$. For example, Eq. **8** tells us that in order to ensure that the baseline occupation remains below $\Theta_i^0 < 1/4$ while the noncognate metal concentration varies less than two-fold, the value of ϵ would need to be larger than $\log(5) \approx 0.7$. This has an interesting synthetic-biological implication: Sensors cannot be made from arbitrary chemical compounds but only from those that have sufficient binding strength to fulfill the above conditions.

For metal sensing to be operational for a whole sensor set, Eq. **10** must hold for every transition metal sensor. In other terms only the joint fulfillment of Eq. **10** creates a complete set of compatible sensors. Compatibility thus requires a delicate balance between all cognate and noncognate binding constants. With our definition of (positive) binding free energies, the compatibility condition for the sensor set then reads as

$$G_j^j - G_i^j > 0\,, \quad \text{for all } i,j \text{ with } i \neq j\,. \qquad [11]$$

Eq. **11** is our first main result. In Fig. 2*B*, we illustrate schematically an example of coexisting Zn^{II} and Cu^{I} sensors and explore their compatibility conditions. We show how changing the binding site of the Cu^{I} sensor (and as a consequence its free energies of binding to Cu^{I} and to Zn^{II}), can make the Zn^{II} sensor selective to its cognate metal. As a result of this change, one obtains a set of two compatible metal sensors. Once again, it is important to note that the conditions described in Eq. **11** are only necessary and not sufficient, i.e. they do not guarantee that there exist metal concentrations such that all of the sensors simultaneously satisfy the conditions (A) and (B) listed above. In fact, to guarantee sufficiency, the difference between G_j^j and G_i^j may have to be substantially larger than 0. For instance, we found that $G_j^j - G_i^j > \log(2(m-1))$ is sufficient to guarantee that (A) and (B) are satisfied (details in *SI Appendix*, section 6). For an explicit example, consider six sensors ($m = 6$) and $\Lambda \gtrsim 1$. If $G_j^j - G_i^j > 1$ for all i and j, then concentrations can be found so that conditions (A) and (B) above are simultaneously satisfied for six sensors.

It might seem difficult to judge the relevance of the compatibility condition for biological systems. One can ask whether Eq. **11** is a strong constraint that confines sensor evolution to a small set of solutions or it only prevents some special cases from being functional in the cell. To explore this question, we study a model of metal–binding site interactions that captures the essential contributions to the binding free energies G_i^j.

1.4. Thermodynamic Additivity. Within sensing binding sites, three chemical elements act as binding partners to transition metals. These elements are oxygen (O), typically from aspartic or glutamic acid (rarely from carbonyl groups), nitrogen (N) from the imidazole group of histidine or the amide group from the terminal ends of a protein, and sulfur (S) from cysteine (see, e.g. ref. 34 and *SI Appendix*, Table S2). Thus, the binding site of a transition metal can be characterized by the numbers of its O- (n_O), N- (n_N), and S- (n_S) residues involved in sensing. We define the binding site of a transition metal i accordingly, via a counting vector:

$$\vec{n}^i := \left(n_O^i, n_N^i, n_S^i\right)\,. \qquad [12]$$

We synonymously use the vector $\vec{n}$ as a label of the sensor, e.g. $\vec{n}^{\mathrm{Zn}}$ is the "Zn-sensor." Fig. 1*B* shows an example of a Zn^{II} sensor binding site with n-vector $\vec{n} = (1, 2, 1)$, as it is composed of one O residue (red), as well as two N (blue) and one S (yellow). The sum of the n-vector components, i.e. the number of atoms directly interacting with the metal, is called the coordination number, $N := \sum_{\alpha\in\{O,N,S\}} n_\alpha$. For transition-metal binding, the observed and theoretically expected coordination numbers are typically between 2 and 6 (35).

The characterization of binding sites in terms of their elemental composition n_α^i suggests a decomposition of the corresponding binding free energies G_i^j. Following (36), we will assume here that the total binding free energies G_i^j can be decomposed into three additive terms using the hydrated form of the metal ions as a reference state:

$$\begin{aligned} G_i^j &= \sum_\alpha n_\alpha^i g_{\kappa(i),\alpha}^j - \sum_\alpha f_\alpha(n_\alpha^i) + C \\ &\equiv \vec{n}^i \cdot \vec{g}_{\kappa(i)}^j - F(\vec{n}^i) + C \end{aligned} \qquad [13]$$

(see *SI Appendix*, section 7 for a detailed discussion of this approximation. Note that using the labile pool as a reference state would only complicate matters as i) not every metal binds to the same pool and ii) measurements for the different components of the following decomposition of G_i^j have so far only been performed with water as reference state). Based on the assumption of thermodynamic additivity, we have written the first (leading-order) term as a sum over the pairwise free energy contributions $g_{\kappa(i),\alpha}^j$ of each of the coordinated elements α and the bound metal j. The index $\kappa(i)$ accounts for the possibility of different binding strengths in different steric arrangements of the binding site. We express this possibility as

$$\vec{g}_{\kappa(i)}^j \equiv \gamma_{\kappa(i)}^j \vec{g}^j, \qquad [14]$$

where $\gamma_{\kappa(i)}^j$ is a constant that depends on the metal j and the sensor i it is bound to. We chose $\gamma_{\kappa(i)}^j = 1$ if the binding site i is arranged in a chemically preferred geometry of metal j, and $0 \leq \gamma_{\kappa(i)}^j < 1$ otherwise. The residual pairwise binding free energies

$$\vec{g}^j := \left(g_O^j, g_N^j, g_S^j\right)\,. \qquad [15]$$

depend then exclusively on the metal and the chemical group (O, N, S) it is bound to, but nothing else.

The second term is the shielding function $F(\vec{n}^i)$ that accounts for decreasing contributions of successive (O,N,S)-residues to the total binding free energy. By construction, it is only a correction to the leading-order term and thus is always (much) smaller in magnitude. For each chemical group α, the term $f_\alpha(n_\alpha^i) \geq 0$, and it increases with n_α^i (36). This modeling assumption agrees well with experimental data on metal ligand interactions in solution, and, interestingly, it is largely independent of the metal identity

(36, 37). We show explicit examples of this function in the (*SI Appendix*, Fig. S4). The existence of the shielding term implies that the contributions to the binding free energy of individual chemical residues α decrease with increasing site coordination number.

Finally, the third term C in Eq. **13** accounts for the entropic contribution of freeing up the water molecules previously bound in the hydration shell around the dissolved metal upon its binding to the sensor. Since most transition metals have similar hydration shells (38), and the hydration shell is supposed to be completely lost upon binding, we assume that this contribution to the total free energy is a constant.

In the following, we will first consider the "plain model," in which only the leading order term is included

$$G_i^j := \sum_{\alpha\in\{O,N,S\}} n_\alpha^i g_\alpha^j \equiv \vec{n}^i \cdot \vec{g}^j, \qquad [16]$$

and where we have assumed $\gamma^j_{\kappa(i)} = 1$ for all i, j and $F(\vec{n}^i) \approx 0$. We have also ignored the constant C as it cancels out in all versions of the compatibility conditions. To exclude repulsive binding interactions, we also assume $g_\alpha^j \geq 0$. The contributions due to geometric preferences and shielding, present in Eq. **13**, will be considered later as perturbations to the plain model.

The assumption of thermodynamic additivity, as stated in Eq. **16**, simplifies the general condition of sensor compatibility, Eq. **11**. The resulting $m(m-1)$ linear inequalities

$$\left(\vec{n}^j - \vec{n}^i\right) \cdot \vec{g}^j > 0 \quad \text{for all } i, j \text{ with } i \neq j \qquad [17]$$

that jointly need to be fulfilled for all m metals and sensors, thus provide the means to determine possible sets of compatible sensors. This is our second main result.

We depict schematically the compatibility improvement process in Fig. 2*B*. We show how changes to the entries of the n-vectors, which define the binding sites' compositions, can make previously incompatible sensors compatible. If one is only interested in the specificity of sensor i, it is sufficient to keep i fixed and test whether the conditions defined by Eq. **17** are fulfilled for all $j \neq i$. It is important to note that the inequalities in Eq. **17** do not rely on the lengths of the g-vectors but only on their orientations. We use this property later when we try to determine the sets of mutually compatible sensors. We will see that in the absence of precise experimental values for the g-vectors, the ratios $g_O : g_N : g_S$ can often provide sufficient information.

1.5. Geometric Interpretation of Compatibility. Describing sensors as n-vectors allows for a geometric interpretation of compatible sensors in a coordinate space spanned by the unit vectors $\vec{e}_O = (1, 0, 0)$, $\vec{e}_N = (0, 1, 0)$ and $\vec{e}_S = (0, 0, 1)$. In this "sensor space" any conceivable sensor is represented as a point on a grid of natural numbers (integers ≥ 0). All conceivable sensors lie on or below the surface defined by $\sum_\alpha n_\alpha = N_{\max}$. Consider, for example, all conceivable sensors with maximal coordination number $N_{\max} = 6$ (dashed line in Fig. 3*A*) and binding sites that rely only on two amino acid residues (e.g. S and N). In such a 2-dimensional (2D) sensor space one finds exactly 27 (generally, $\frac{1}{2}(N_{\max}+2)(N_{\max}+1) - 1$) conceivable sensors (gray dots in Fig. 3*A*). For a 3-dimensional (3D) sensor-space, i.e. for sensors based on 3 distinguishable amino acid residues (O,N,S) with

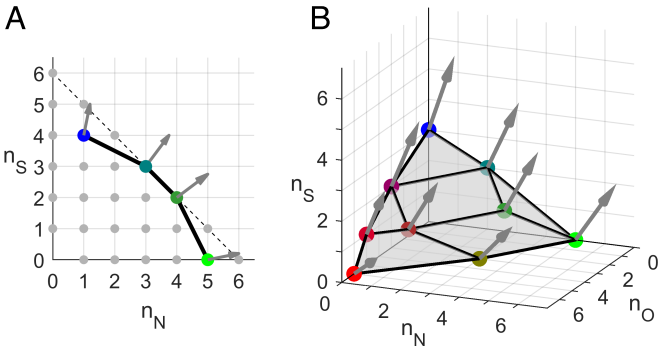

Fig. 3. Convex hulls of mutually compatible sensors. (*A*) Sensors for which only N and S residues interact with metals can be represented in two-dimensional (2D) n-spaces. The 27 possible n-vectors with coordination number $N_{\max}$ between 1 and 6 are shown as gray discs. The dashed line runs through the possible n-vectors with $N_{\max} = 6$. The compatibility conditions allow a set of at most 4 mutually compatible sensors. An example set of compatible n-vectors $\vec{n}^i$ (colored discs) is shown. For each sensor, its cognate metal g-vector $\vec{g}^i$, anchored at $\vec{n}^i$, is also shown. A geometric consequence of the compatibility conditions is that all sensors must lie at the vertices of their convex hull, the upper part of which is shown here in black. (*B*) Sensors for which all three residues, O, N, and S, interact with metals are represented in three-dimensional (3D) n-space. Nine sensor n-vectors $\vec{n}^i$ (colored discs) are shown. Again, for each sensor, its cognate metal g-vector $\vec{g}^i$ is anchored at $\vec{n}^i$, and the upper part of the convex hull of sensors is shown.

$N_{\max} = 6$, the number of conceivable sensors increases to 83 (generally, $\frac{1}{6}(N_{\max}+3)(N_{\max}+2)(N_{\max}+1) - 1$).

The requirement that sensors are mutually compatible and thus satisfy the inequalities of Eq. **17**, reduces the number of possible coexisting sensors. A direct consequence of these equations is that all compatible sensors must lie at the vertices of their convex hull (black line in Fig. 3*A*; gray surface in Fig. 3*B*). We prove this statement in *SI Appendix*, section 3. In the 2D case discussed above, we calculated that, out of the 27 conceivable sensors, only 4 at most can be mutually compatible (example shown in Fig. 3*A*). Applying constraints to the orientation of the g-vectors can reduce this number further (*SI Appendix*, Fig. S2*A*). This very low maximum number of compatible sensors in the 2D case might be the reason why real sensors rely on 3 chemically distinguishable groups. In the 3D case, numerical evaluation of plausible scenarios of metal binding preferences shows that maximally 10 sensors can be simultaneously compatible (*SI Appendix*, Fig. S2*B*). Depending on the binding preferences of the metals to O, N, and S, the number of jointly compatible sensors will be even lower. For example, if all metals prefer to bind S over N over O ($g_S > g_N > g_O$), the maximum number of compatible sensors decreases to 7.

1.6. Sensor Arrangement. Up until this point we characterized the global arrangement of compatible sensors within sensor space as lying at the vertices of a convex hull. We now discuss how the arrangement of the sensors is coupled to the arrangement of their g-vectors. In Fig. 3, we show the n-vectors $\vec{n}^i$ of compatible sensors as well as their corresponding g-vectors $\vec{g}^i$ anchored to them. In the 2D example (Fig. 3*A*), the angular orientations of the g-vectors "mimics" the order of the sensors on their convex hull (shown in black in Fig. 3*A*): The more a g-vector points toward the $\vec{e}_S$ direction, the closer the sensor is to the $\vec{e}_S$ axis. This relationship between vectors $\vec{n}$ and $\vec{g}$ is generally valid for the 2D case. In the 3D case, the notion of an ordering becomes more involved (see *SI Appendix*, section 4 for a rigorous treatment) but, for all practical purposes, it can be said that the angular

orientations of the g-vectors mimics the arrangement of the n-vectors representing the sensors on the convex hull (Fig. 3*B*).

2. Data

To test our theory, we require experimentally determined values for both g-vectors and n-vectors.

2.1. g-vectors. g-vectors, as such, have not been tabulated until now, yet measurements which can be used as approximations of their entries, g_O, g_N, and g_S have been made. We followed Martell and Hancock, who suggested approximating amino acid–metal interactions by the mono-dentate interactions of the chemical residues involved (36). In practice, this approach amounts to using acetic acid as a proxy for glutamic and aspartic acids, ammonia, and imidazole as a proxy for terminal amine groups and histidine, and mercapto-ethanol as a proxy for cysteine (36). In detail, for a given metal i, we use $g_O^i = \log K^i_{\text{acetic acid}}[1M]$, $g_N^i = \log K^i_{\text{imidazole}}[1M]$, and $g_S^i = \log K^i_{\text{mercapto-ethanol}}[1M]$. The values of these binding constants or, as they are often called in the literature stability constants, are displayed for each metal in Fig. 1*D*. Ammonia and imidazole are chemically different, yet the almost perfect agreement of the reported binding constants for ammonia and imidazole in Fig. 1*D* suggests that this simplification is a reasonable first approximation. Using this simplification, discussed in more detail in *SI Appendix*, section 8, we derived a curated set of g-vectors based on the published stability constants tables (6, 39). We state the entries of our set of g-vectors in *SI Appendix*, Table S1 and Fig. 1*D*, while the g-vectors themselves, for clarity rescaled to terminate on the 111 plane, are plotted in Fig. 4*B*.

Can we build trustworthy g-vectors based on experimentally obtained stability constants from the literature? Most likely not. Although the reported measurement errors are typically below 10%, the large variability across different experimental setups, reported in ref. 39, suggests much larger error values. Increasing confidence in the values of the stability constants may require renewed effort to carefully measure their values under identical and biologically plausible conditions, using protocols as laid out in refs. 36, 40, and 41.

Although some particular values of the reported stability constants may not be accurate, two phenomenological properties of the reconstructed g-vectors are likely to be quite solidly established. First, transition metals prefer binding S over N over O. This requires the g-vector components to obey the inequality

$$\text{WH1:}\quad g_S > g_N > g_O, \qquad \textbf{[18]}$$

which we refer to as an example of Williams Heuristics (WH) (see Fig. 1*D* where this ordering is clearly visible). We coin this term after R. J. P. Williams, who was the first to point out the apparent universality of such an ordering. This ordering seems to hold across different compounds that contain the same reactive group (5, 42). Only for Mn^{II}, a preference for O over N cannot be ruled out (5). We call this hypothetical ordering the second Williams Heuristic

$$\text{WH2:}\quad g_S > g_O > g_N, \qquad \textbf{[19]}$$

The alkaline earth metals Ca^{II} and Mg^{II} seem to exclusively bind to O, corresponding to the third Williams Heuristic with

$$\text{WH3:}\quad g_O > g_N \approx g_S \approx 0. \qquad \textbf{[20]}$$

The second, more subtle phenomenological property of g-vectors that is likely to be well established is that transition metals have distinct relative preferences for O, N, and S (g_O:g_N:g_S; see Fig. 1*D*). Graphically, this translates to g-vectors pointing in different directions (Fig. 4*B*).

2.2. Sensors. Extensive experimental effort is required to resolve the coordination number, geometry, and amino acid composition of a binding site. The structures of several different bacterial transition-metal sensors with high resolution of their binding site have been characterized experimentally. Each of these sensors

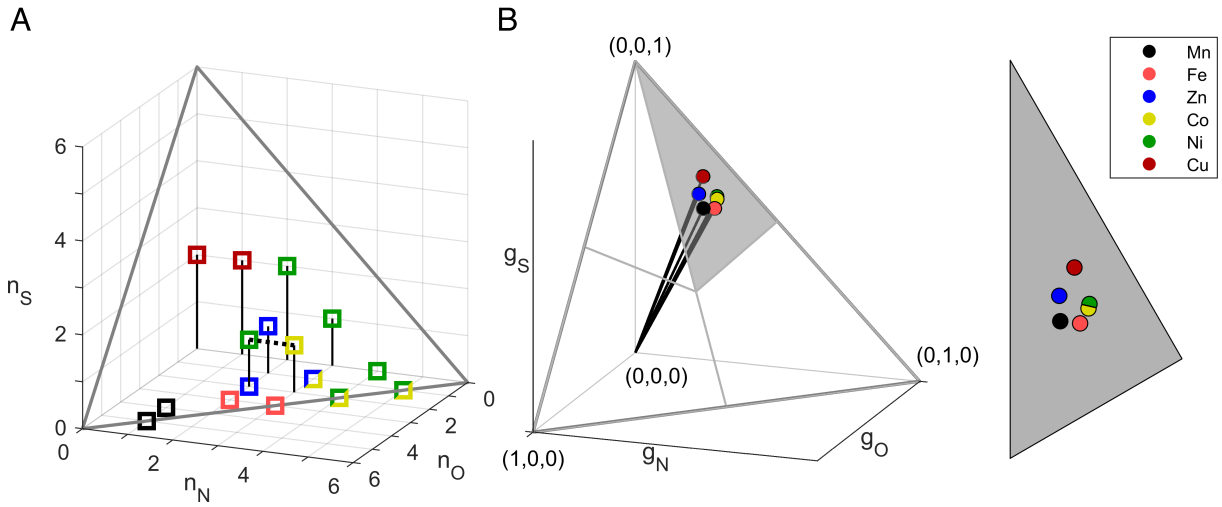

Fig. 4. Experimentally determined bacterial metal-sensors characterized by the residue composition of their primary binding sites (n-vectors) and the corresponding binding strengths (g-vectors). (*A*) The primary binding site of each sensor as described in Table 1 is represented by a small square in n-space. A triangle has been drawn between the points (6, 0, 0), (0, 6, 0), and (0, 0, 6) to more easily gauge the relative amounts of O, N, and S residues in each sensor ("sensor composition"). Viewed from atop, the *Lower Left* corner of the triangle marks O-rich sensors, the *Lower Right* corner N-rich sensors and the *Top* corner S-rich sensors. Different colors represent different metals (shown in the legend). Sensors that bind to multiple metals are indicated by bicolor squares. The dashed line between (2, 2, 1) and (2, 3, 1) represents a single joint Co^{II}/Ni^{II} sensor where different residues are recruited into the binding site depending on which metal is present. (*B*) G-vectors of all metals approximated by stability constants from the literature (*SI Appendix*, Table S1) and g-vectors as projected onto the (1, 1, 1) plane (the gray triangle on the *Right*).

Table 1. Bacterial metal sensors

Metal	Primary binding site n_O	n_N	n_S	Species
Mn^{II}	4	1	0	MT
Mn^{II}	5	1	0	BS, EC, SE
Fe^{II}	3	2	0	MT
Fe^{II}	3	3	0	BS, EC, SE
Co^{II} (Ni^{II})	2	3	1	EC, SE
Zn^{II}	1	2	1	MT, BS, SE, EC, SC
Zn^{II}	2	2	0	SP6, SP7
Zn^{II}, Co^{II}	1	3	0	SP, LL, SA
Ni^{II}	0	2	2	SP6
Ni^{II}	0	3	1	EC, SE
Ni^{II}	0	4	0	SC
Ni^{II} (Co^{II})	2	2	1	EC, SE
Co^{II}, Ni^{II}	1	5	0	MT
Co^{II}, Ni^{II}	2	4	0	MT
Cu^{I}	0	0	2	EC, SE
Cu^{I}	0	1	2	GT, MT, BS

The primary binding site is the tightest allosterically active binding site. If several metals are named in a row this implies that a sensor (or binding site in general) is responsive to more than one metal. Metals named in parenthesis also bind to this sensor but recruit (slightly) different amino acid residues. The species code is as follows: SA: *S. aureus*, SP7: *Synechocystis PCC 7942*, EC: *E. coli*, SC: *S. coelicolor*, BS: *B. subtilis*, SP6: *Synechocystis PCC 6803*, MT: *M. tuberculosis*, SE: *S. enterica*, SP: *S. pneumoniae*, LL: *L. lactis*, and GT: *G. thermodenitrificans*. A more detailed version of the table can be found in *SI Appendix*, Table S2.

has typically up to three functionally distinguishable types of binding sites. These binding sites can occur as multiple identical duplicates in the same sensor. These binding sites are typically sorted into three separate categories. The binding sites of category 1 are the tightest allosterically active binding sites for their cognate metal and are called here primary metal binding sites (Table 1 and *SI Appendix*, Table S2). Binding sites of category 2 are also allosterically active yet bind metals more weakly and their function is considered as regulatory to the primary binding event. We call these sites ancillary metal binding sites. These binding sites do not exist in every sensor. Finally, binding sites of category 3, which we call structural binding sites, are exclusively relevant for the structural integrity of a sensor and exist only in some of the sensors. Throughout this text, we only use the primary metal binding sites in comparison to our theory. Only a handful of bacteria have been thoroughly studied with respect to their metal sensors. There is significant overlap in the binding sites used by these species, which we also report in Table 1.

Depicting experimentally determined sensors from Table 1 in $(\vec{e}_O, \vec{e}_N, \vec{e}_S)$-sensor space reveals a clear clustering of sensors by their metal preferences (Fig. 4). The relative arrangement of these clusters seems to be preserved for all studied bacterial species. However, it should be noted that despite all the experimental efforts, there still remains some uncertainty about the amino acid composition of some of these sensors (see discussion in *SI Appendix*, section 9).

Strikingly, the arrangement of the sensors seen in Fig. 4*A* mimics the g-vector arrangement shown in Fig. 4*B*. For example, the n-vectors of the sensors Mn^{II}, Fe^{II}, Ni^{II}, and Cu^{I} follow a circular arrangement, similar to the g-vectors for these metals. The vectors associated with Co^{II} and Ni^{II} overlap both in sensor space and in g-vector space. The observation that the g-vectors and n-vectors follow similar arrangements is in agreement with our theoretical prediction (see Fig. 3 for a graphic- and *SI Appendix*, section 4 for a formal- treatment).

3. Comparison of Theory and Data: The Case of *E. coli*

A core prediction of our theory is that compatible sets of metal sensors should be found in $(\vec{e}_O, \vec{e}_N, \vec{e}_S)$-sensor space on a convex surface. In Sections 3.1–3.3 below, we test this prediction against existing microbial data. Our theory also predicts where in sensor space particular sensors could be found if additional information about the g-vectors for metals is taken into account. Although it is difficult to infer the precise locations of g-vectors from the published experimental data, in Section 3.4 we nevertheless examine how well theoretical predictions of sensor locations compare with known sensors, given this imprecise and scarce data.

3.1. Testing the Convex Hull Condition. To test the prediction about location of the compatible sensors on a convex hull in sensor space, one should consider metal sensors found within the same bacterial species. According to our data summary displayed in Table 1 and *SI Appendix*, Table S2, the most extensive experimental data were obtained for sensors in *S. enterica*, *Bacillus subtilis*, *E. coli*, and *Mycobacterium tuberculosis*. We focus here on *E. coli*, and consider the other three bacteria (with almost identical metal binding sites) in *SI Appendix*, Fig. S3. In Fig. 5, we compare our theoretical results with the known sensors, each intervening in the regulation of only a single metal among Mn^{II}, Fe^{II}, Zn^{II}, Ni^{II}, and Cu^{I}. We find that in $(\vec{e}_O, \vec{e}_N, \vec{e}_S)$-sensor space all of the 5 displayed sensors lie on the same plane. However, the convexity requirement implies that no n-vector can be equal to a weighted average of the other n-vectors. As the sensors in Fig. 5 are coplanar and the Zn^{II} sensor is surrounded by the other four sensors, this condition is clearly not fulfilled. Therefore, the reported sensors of *E. coli* fail the compatibility conditions, although interestingly, as we will see, they do it only by a small margin.

3.2. Compatibility Enhancing Mechanisms. Within the present theory, metal sensors in *E. coli* can easily be made compatible by relaxing some of the assumptions made earlier thus extending the plain model. In particular, we assumed that (I) the prefactor in Eq. **14** representing the effects of binding site geometry is 1, $\gamma^{j}_{\kappa(i)} = 1$; that (II) the shielding effect Eq. **13** can be neglected, i.e., $F(\vec{n}^i) = 0$; that (III) every metal interacts with the full set of binding partners present in the binding site, i.e., that the n-vector of a sensor is the same independent of the metal bound; and that (IV) the allosteric response of a sensor binding to different metals is the same. We discuss below how these four aspects of metal binding modify the simple version of the compatibility condition stated in Eq. **17**, and enhance the compatibility of a set of metal sensors. We will show that, mathematically, all of these compatibility-enhancing mechanisms amount alternatively to reducing the entries of n-vectors, to rescaling n-vectors when considering noncognate metals, or to changing the threshold value on the RHS of the compatibility conditions away from zero.

(I) Geometric selection: The binding site of a sensor is typically arranged according to the preferred geometry (35, 43, 44) of its cognate metal. In consequence, noncognate metals that do not share this geometric preference may bind less well to the sensor. Sensors can thus "geometrically select" against some of the noncognate metals. Although noncognate metals may distort the binding site to adapt it to their preferred geometries, in some cases, it might not

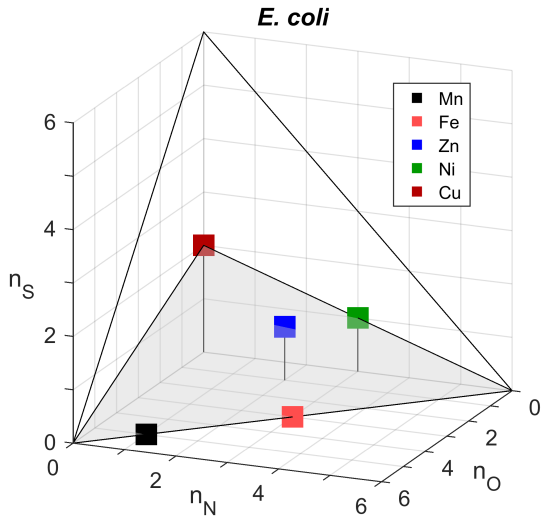

Fig. 5. Sensors of *E. coli* marginally fail the compatibility conditions. Shown are known primary binding sites of *E. coli* sensors uniquely responding to a single metal. They all lie on the same plane (indicated in gray) in n-space, with the Zn^{II} sensor surrounded by the four others. The convexity requirement is violated and compatibility enhancing mechanisms have to be considered.

be possible for them to fully assume this geometry, due to residual strains in the protein. To formalize this concept, we keep the cognate geometry factor $\gamma^j_{\kappa(j)} = 1$, and set $\gamma^j_{\kappa(i)} < 1$ to represent the geometry of sensor i disfavoring metal j. The corresponding compatibility condition then becomes

$$\vec{n}^j \cdot \vec{g}^j_{\kappa(j)} - \vec{n}^i \cdot \vec{g}^j_{\kappa(i)} = (\vec{n}^j - \gamma^j_{\kappa(i)} \vec{n}^i) \cdot \vec{g}^j > 0\,, \quad [21]$$

so that only the n-vector $\vec{n}^i$ is rescaled.

In *SI Appendix*, Fig. S5*C* shows an example of geometric selection for $i = Ni^{II}$ and $j = Zn^{II}$, where the geometry of the Ni^{II} sensor disfavors Zn^{II} binding ($\gamma^{Zn}_{\kappa(Ni)} < 1$). In consequence, the Ni^{II} sensor becomes compatible with the Zn^{II} sensor; with the Ni^{II} sensor's n-vector rescaled, all of the sensors in *SI Appendix*, Fig. S5*C* now lie on a convex hull.

(II) Shielding: Increasing the coordination number N of a metal leads to a decreasing contribution of each additional chemical group to the total binding free energy (36, 37). For a particular sensor j, this shielding effect is represented by the term $-F(\vec{n}^j)$ in Eq. **13**. If we include the shielding effect, the pairwise compatibility conditions become

$$(\vec{n}^j - \vec{n}^i) \cdot \vec{g}^j > \epsilon^{ji} \quad [22]$$

$$(\vec{n}^i - \vec{n}^j) \cdot \vec{g}^i > -\epsilon^{ji} \quad [23]$$

with $\epsilon^{ji} = F(\vec{n}^j) - F(\vec{n}^i)$. Shielding is thus equivalent to shifting the threshold of the pairwise compatibility conditions, from zero to $\pm\epsilon^{ji}$. In consequence, shielding is a sort of trade-off mechanism that makes one metal sensor more compatible at the expense of another. In general, $\epsilon^{ji} > 0$ will hold for $N^j > N^i$ as $F(\vec{n}^j)$ increases with the coordination number N. Including shielding will thus affect mostly the compatibility between sensors of different coordination numbers.

(III) Incomplete binding site assembly: A metal binding to a noncognate sensor may not be able to assemble the entire binding site (45–47). Zn^{II} mismetallation of the Ni^{II}/Co^{II} sensor NmtR is a prime example (47). In such a case, the metal interacts with a subset of the possible binding partners in the sensor, and we model this scenario by replacing the compatibility condition for i and j with

$$(\vec{n}^j - \vec{n}^{i\,\text{binding}\,j}) \cdot \vec{g}^j > 0 \quad [24]$$

where the entries of $\vec{n}^{i\,\text{binding}\,j}$ are less than or equal to the entries of $\vec{n}^i$.

(IV) Allosteric effects: If we include in the model the differences in allosteric responses triggered by different metals, then the zero threshold on the RHS of the compatibility conditions can be relaxed (*SI Appendix*, section 14). For the Zn^{II}/Ni^{II} binding example, discussed above for geometric selection, this implies that even though Zn^{II} can bind to the Ni^{II} sensor, it does not assume the planar binding site geometry. This presumably prevents Zn^{II} from triggering the same allosteric response as Ni^{II}. If we include the effect of allostery, the pairwise compatibility conditions can be written

$$(\vec{n}^j - \vec{n}^i) \cdot \vec{g}^j > A^j_i, \quad [25]$$

where A^j_i can be negative when metal j triggers a weaker allosteric effect in sensor j compared to i. In *SI Appendix*, section 14, we describe this mechanism in detail. For sufficiently negative A^{Zn}_{Ni}, the Ni^{II} sensor becomes compatible with Zn^{II}.

3.2.1. Other mechanisms. Finally, we should note that recent studies of bacterial sensors suggest additional compatibility enhancing mechanisms that lie outside the scope of the present theory, such as the use of cross-regulation (48, 49), or multistate sensors depending on the cellular state, e.g., the intracellular concentration of reactive oxygen species (ROS) (50).

3.3. Constraints Dictated by Binding Constants: Williams Heuristics. Until now, in analyzing the compatibility of metal sensors in *E. coli*, we have not considered constraints coming from the binding (free) energy values, i.e., g-vectors $\vec{g}^i$. Ideally, the g-vectors would be constrained by precise measurements of the binding constants. In Section 3.4 below, we consider the best g-vector estimates derived from published data. Before that, however, let us consider the less constraining, but better established, phenomenological relations between the binding constants, which we named Williams Heuristics (WH1, WH2, and WH3; see Eqs. **18**–**20**). For example, for the case study of *E. coli*, the relevant metals Fe^{II}, Ni^{II}, Cu^{I}, Zn^{II}, and Mn^{II} are considered to obey the WH1 heuristic.

For a given set of sensors, we can explore whether the compatibility conditions can now be fulfilled. By evaluating the compatibility conditions under these extra constraints, we find that some sensors will always be mismetallated by Zn^{II} and Mn^{II}, no matter the exact values of the g-vectors. We analyze this in more detail in *SI Appendix*, section 10 from a geometric perspective. For *E. coli*, no conflict arises from the Fe^{II}, Ni^{II}, and Cu^{I} sensors. This implies that without compatibility enhancing mechanisms, some sensors will be mismetallated by Zn^{II} and Mn^{II}.

To explore which sensors are in particular prone to such mismetallation by Zn^{II} and Mn^{II}, we analyzed the way in which the compatibility conditions are violated. We found that the conditions $\vec{g}^j \cdot (\vec{n}^j - \vec{n}^i) > 0$ for $(j, i) = (\mathrm{Mn}, \mathrm{Fe})$ and $(j, i) = (\mathrm{Zn}, \mathrm{Ni})$ cannot be satisfied if both Zn^{II} and Mn^{II} satisfy the WH1 heuristics. Without compatibility enhancing mechanisms acting in *E. coli*, the Ni^{II} sensor would be unspecific against Zn^{II}, and the Fe^{II} sensor would be unspecific against Mn^{II}.

The current experimental evidence suggests that, indeed, for the Ni^{II} sensor, a specificity enhancing mechanism against Zn^{II} mismetallation is operational (51). In low spin state, Ni^{II} preferentially binds in a planar configuration, while Zn^{II}, in turn, typically prefers a tetrahedral configuration. The Ni^{II} sensor [(0, 3, 1) in Table 1] seems to be functional if the detected metal is bound in a planar configuration. The binding signal is only then allosterically transmitted further, leading to a change in the DNA-binding affinity (29). This elegant mechanism seems to assure the mutual compatibility of the Zn^{II} and Ni^{II} sensors. This is in line with the geometric compatibility enhancing mechanism and allosteric filtering described above in Section 3.2 (I) and (IV).

The situation with the Mn^{II} and Fe^{II} sensors seems more subtle. Based on the currently available experimental measurements of the stability constants, it is not clear whether the incompatibility of Mn^{II} and Fe^{II} sensors could not be just an artifact of the assumption that the Mn^{II} g-vector satisfies the WH1 relations. Based on current data, there remains the possibility that Mn^{II} satisfies the WH2 relations (5). Assigning the WH2-heuristic to Mn^{II} would resolve the mutual incompatibility of Mn^{II} and Fe^{II}. We demonstrate this graphically in *SI Appendix*, Fig. S5. If this possibility is not confirmed, then one will have to invoke alternative compatibility enhancing mechanisms beyond the scope of the present model (see the discussion in *SI Appendix*, section 11).

3.4. Constraints Dictated by Binding Constants: Measured g-vectors. The best way to compare the predictions of our theory with experiments would be to use precise measurements of g-vectors, i.e., free energy binding constants measured for all the metals. The problem is that experimental measurements of g-vectors come with large uncertainties. It turns out that if we use the published values of the g-vectors for Mn^{II}, Fe^{II}, Zn^{II}, Ni^{II}, and Cu^{I} (listed in *SI Appendix*, Table S1), then there are no sets of compatible sensors that could be deduced from our theory based solely on Eq. **17** without evoking possible compatibility enhancing mechanisms. This is true for all four bacterial species, *S. enterica*, *B. subtilis*, *M. tuberculosis*, and *E. coli*, for which many of the sensors were found and characterized.

In view of the lack of precise experimental measurements, before conjecturing additional compatibility mechanisms for metal sensors in these bacteria, it might be worth checking whether—by slightly perturbing the orientation of the experimentally determined g-vectors—one could solve this apparent puzzle. Therefore, we perturbed the g-vector orientations multiple times randomly and uniformly by not more than 0.03 radians (*SI Appendix*, section 13), as shown in the *Left* part of Fig. 6. We found in fact that many of the combinations of such small perturbations led to sets of fully compatible sensors. We depict the sensors generated in this way in Fig. 6 (color-coded circles, details in the caption) and compare them with experimentally

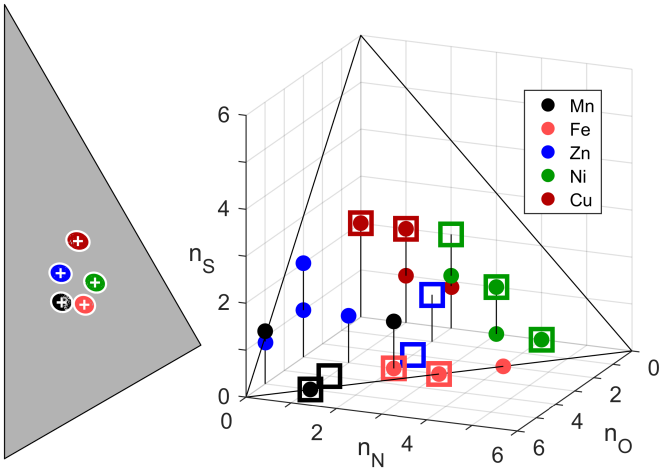

Fig. 6. Prediction of primary binding site composition of different sensors based on the measured g-vectors. Predicting Mn^{II}, Fe^{II}, Zn^{II}, Ni^{II}, and Cu^{I} sensors. In the *Left* triangle, the g-vectors corresponding to experimentally found sensors (white crosses; see Fig. 4*B*) are projected onto the 111 plane. Around them, outlined in white are the perturbation regions (see main text). Those of the g-vector perturbations that successfully lead to compatible sensor sets are indicated as small colored-dots; taken together they form quasi-continuous colored patches inside the perturbation regions. In the 3D n-space on the *Right*, all n-vectors describing the compatible sensors, obtained through the perturbations, are shown as filled circles, and are color-coded by the metal they sense. Sensors occurring at the same location are shown vertically stacked. Sensors which occurred in less than 5% of the compatible sets are not shown. For comparison, the corresponding experimentally determined sensors are marked as open squares, each responding to one metal only.

studied sensors of Mn^{II}, Fe^{II}, Zn^{II}, Ni^{II}, and Cu^{I} (colored squares). Interestingly, many of the real bacterial sensors align well with the predictions of our theory (see colored circles inside the squares). The only case for which none of the theoretically generated sensors corresponds to the one observed in bacteria is the case of the Zn^{II} sensor (blue symbols in Fig. 6). This confirms the analysis presented above (in Section 3.2), which showed that a compatibility enhancing mechanism is needed to obtain the Zn^{II} sensor as part of a compatible sensor set.

Despite a large uncertainty associated with the published measurements, our analysis based on small perturbations of the experimental orientations of g-vectors suggests that some aspects of the relative binding preferences of each metal might still be captured. This encouraged us to further explore the predictive power of our theory by including the g-vector for Ca^{II} and Mg^{II} and repeating the perturbation analysis (*SI Appendix*, Fig. S7 and section 12).

Another interesting case is that of Co^{II}. To our knowledge, there are no bacterial sensors that sense Co^{II} only; the known examples characterize some sensors as responding to Co^{II} and Ni^{II}, or to Co^{II} and Zn^{II} (52, 53). Such double sensors can be easily included in our theory—we show this explicitly in *SI Appendix*, section 11 (see, in particular, *SI Appendix*, Fig. S6).

4. Conclusions, Limitations, and Discussion

4.1. Conclusions. In this work, we have tried to uncover the evolutionary design principles that lead to the compatibility of intracellular binding reactions. We chose as the case study bacterial transition-metal sensors, which participate in regulation of cytosolic metal concentrations. For these sensors it has been shown that even small metal concentration variations can lead to

mismetallation of noncognate binding sites (54). To predict the actual chemical "composition" of transition-metal binding sites, we used a simple equilibrium-thermodynamic model and treated chemical interactions as thermodynamically additive. We described each sensor as a 3D vector, where each vector component corresponded to the number of three different types of participating chemical residues (O: carboxyl, N: imidazole/ammonia, and S: mercapto). Each sensor could then be represented as a point in the positive quadrant of a 3D vector space. Our theory predicts that compatible sensors lie on a convex hull in such a "sensor-space." It further predicts that their arrangement on the convex hull is determined by the relative preference of each metal to bind to O, N, and S residues (Williams Heuristics). Interestingly, we were also able to show that a 3D sensor space is just big enough to ensure compatible sensing of all essential transition metals.

Using actual metal preferences, based on experimentally determined stability constants for metallo-organic group interactions, we found good agreement between the general arrangement in the sensor space of the predicted and experimentally determined sensors. In particular, sensor compositions are quite well predicted by the proposed simple design principles. In accordance with experimental observations, our theory predicts Cu^{I} sensors to be rich in S residues with coordination numbers between two and three. Similarly, Ni^{II} sensors are predicted to be rich in N residues. Mn^{II} and Fe^{II} sensors are neighbors in sensor space, having coordination number five or six, and are typically built exclusively from O and N residues. An intriguing finding of our theory is that the known Zn^{II} and Ni^{II} sensors of *E. coli* (and also *B. subtilis* and *S. enterica*), as well as their Fe^{II} and Mn^{II} sensors, should not be mutually compatible according to our ("plain") theoretical model. Additional "compatibility enhancing mechanisms" are thus necessarily required to distinguish these metal pairs. Interestingly, such mechanisms, described in this paper, are strongly suggested by experimental observations made for both metal pairs.

Another conclusion of our study is a relatively small number of compatible-yet-distinguishable sensor compositions that the sensor binding site for a specific metal can assume. We find this assertion qualitatively in agreement with the compiled data on experimentally determined bacterial metal sensors. If confirmed, this conclusion may imply that the variability of cytosolic transition-metal concentrations is strongly restricted across different bacterial species (and potentially also across other organisms). Changes in the make-up of binding sites are possible, yet the requirement for compatibility induces correlations between realizable binding sites. This lack of evolutionary biochemical adaptability even under extreme environmental variations might be the reason why phagocytes, which are a part of the human immune system, can successfully attack bacterial cells with the use of a mixed strategy of iron deprivation and zinc and copper poisoning (55, 56). This mechanism may be an efficient addition to the usually evoked mechanism of the creation of reactive species that denature proteins and damage DNA (57).

It is also important to stress that our theory, similarly to most cellular models, oversimplifies the nature of the cytosolic medium. The phenomena we consider take place in crowded spaces and involve a small number of molecules, where the usual notion of concentration may not be applicable. The properties of intracellular media, such as cytosolic pH, vary among species and should thus be included in future theories. It will be particularly interesting to generalize our theory to metal sensing in acidophiles or alkaliphiles, and by extension also thermophiles and barophiles.

There are also some shortcomings of our work, connected with assumptions made in order to simplify the model. For example, we assumed metal concentrations to be close to where cognate sensors are the most sensitive. We also set the threshold specificity Θ_i^0 equal to 1/2. Obviously, these assumptions should be confronted with experimental data, once the latter become more extended and precise.

4.2. Perspectives. We hope that our theory of compatibility, based on the geometry of the sensor space, will serve as a heuristic to reason about metallo-regulatory processes. As a first step, one could measure the cognate and, importantly, the noncognate binding constants of all sensors that jointly operate in one species. The paper by Osman et al. (7) provides these measurements for the cognate metals; likewise, an extension to noncognate metals would be informative and interesting. However, to make our theory truly quantitatively predictive, a renewed effort to measure stability constants for metallo-organic group interactions with much better precision will be required. Additionally, to meaningfully include steric selectivity in our theory, one would need to experimentally quantify the energetics of the steric preferences and of allosteric interactions for each metal. Finally, in our opinion, the effect of shielding, which can affect the binding contribution of each amino acid residue as the coordination number increases, has not been adequately explored experimentally.

The theory presented here makes a strong case for the convergent evolution of sensor binding sites. The coarse positioning of sensors in the sensor space should to a large extent be determined by the rules of mutual compatibility. This seems in agreement with the experimental finding that evolutionarily unrelated metal sensors, which sense the same metal in different bacterial species, often present identical binding sites. Similarly, one finds that evolutionarily related metal sensors have diversified into sensors for different metals.

By its very design, the approach presented here does not apply exclusively to metal–protein binding in bacteria. In principle, it should be possible to adapt it to other organisms and to more general cytosolic binding interactions. If this hypothesis turns out to be true, we might want to study cellular interactions not under the usual lens of "what binds best," but rather "what binds well enough while being fully compatible"—a conceptual shift with potential implications for our basic understanding of cellular processes, as well as for possible practical applications.

Data, Materials, and Software Availability. Previously published data were used for this work (Multiple articles, cited in the article and *SI Appendix*).

ACKNOWLEDGMENTS. This work was stimulated by experimental work of Deenah Osman (to whose memory we wish to dedicate this article), Nigel Robinson, and their collaborators. We would like to thank them, as well as Sirio Belga Fedeli, Tankut Can, and Daniel R. Weilandt for helpful discussions. We truly appreciate thoughtful comments of all three reviewers, who helped us to clarify many points in the original manuscript. The work of N.L. was supported by a stipend funded by the Simons Foundation, the Schmidt Fellow Fund, and the Eric and Wendy Schmidt Fund. The work of L.C. was supported by a grant from the Simons Foundation (691552, as well as Simons Math+X Investigators Award 400837).

1. A. J. Thomson, H. B. Gray, Bio-inorganic chemistry. *Curr. Opin. Chem. Biol.* **2**, 155–158 (1998).
2. S. E. Janisse, R. L. Fernandez, M. C. Heffern, Characterizing metal-biomolecule interactions by mass spectrometry. *Trends Biochem. Sci.* **48**, 815–825 (2023).
3. C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, Metal ions in biological catalysis: From enzyme databases to general principles. *JBIC J. Biol. Inorg. Chem.* **13**, 1205–1218 (2008).
4. C. Andreini, I. Bertini, A. Rosato, Metalloproteomes: A bioinformatic approach. *Acc. Chem. Res.* **42**, 1471–1479 (2009).
5. H. Irving, R. Williams, The stability of transition-metal complexes. *J. Chem. Soc.* **1953**, 3192–3210 (1953).
6. A. E. Martell, R. M. Smith, *Critical Stability Constants* (Springer, 1974), vol. 1.
7. D. Osman *et al*., Bacterial sensors define intracellular free energies for correct enzyme metalation. *Nat. Chem. Biol.* **15**, 241–249 (2019).
8. A. W. Foster, D. Osman, N. J. Robinson, Metal preferences and metallation. *J. Biol. Chem.* **289**, 28095–28103 (2014).
9. D. A. Capdevila, K. A. Edmonds, D. P. Giedroc, Metallochaperones and metalloregulation in bacteria. *Essays Biochem.* **61**, 177–200 (2017).
10. D. E. Kunkle, E. P. Skaar, Moving metals: How microbes deliver metal cofactors to metalloproteins. *Mol. Microbiol.* **120**, 547–554 (2023).
11. U. Krämer, Metal homeostasis in land plants: A perpetual balancing act beyond the fulfilment of metalloproteome cofactor demands. *Annu. Rev. Plant Biol.* **75**, 27–65 (2024).
12. M. R. Jordan, J. Wang, D. A. Capdevila, D. P. Giedroc, Multi-metal nutrient restriction and crosstalk in metallostasis systems in microbial pathogens. *Curr. Opin. Microbiol.* **55**, 17–25 (2020).
13. G. L. Sun, E. E. Reynolds, A. M. Belcher, Designing yeast as plant-like hyperaccumulators for heavy metals. *Nat. Commun.* **10**, 5080 (2019).
14. N. Kwiatos, K. J. Waldron, In a state of flux: New insight into the transport processes that maintain bacterial metal homeostasis. *J. Bacteriol.* **206**, e00146–24 (2024).
15. A. W. Foster *et al*., A tight tunable range for Ni (II) sensing and buffering in cells. *Nat. Chem. Biol.* **13**, 409–414 (2017).
16. J. D. Helmann, Metals in motion: Understanding labile metal pools in bacteria. *Biochemistry* **64**, 329–345 (2025).
17. D. H. Nies, "Bacterial transition metal homeostasis" in *Molecular Microbiology of Heavy Metals*, D. H. Nies, S. Silver, Eds. (Springer, 2007), pp. 117–142.
18. D. H. Nies, G. Schleuder, D. Galea, M. Herzberg, A flow equilibrium of zinc in cells of *Cupriavidus metallidurans*. *J. Bacteriol.* **206**, e00080–24 (2024).
19. P. Chandrangsu, C. Rensing, J. D. Helmann, Metal homeostasis and resistance in bacteria. *Nat. Rev. Microbiol.* **15**, 338–350 (2017).
20. B. A. Gilston *et al*., Structural and mechanistic basis of zinc regulation across the *E. coli* Zur regulon. *PLoS Biol.* **12**, e1001987 (2014).
21. E. F. Pettersen *et al*., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
22. Z. Ma, F. E. Jacobsen, D. P. Giedroc, Coordination chemistry of bacterial metal transport and sensing. *Chem. Rev.* **109**, 4644–4681 (2009).
23. K. A. Baksh, D. B. Zamble, Allosteric control of metal-responsive transcriptional regulators in bacteria. *J. Biol. Chem.* **295**, 1673–1684 (2020).
24. S. Marzen, H. G. Garcia, R. Phillips, Statistical mechanics of monod-wyman-changeux (mwc) models. *J. Mol. Biol.* **425**, 1433–1460 (2013).
25. M. Razo-Mejia *et al*., Tuning transcriptional regulation through signaling: A predictive theory of allosteric induction. *Cell Syst.* **6**, 456–469 (2018).
26. A. Mikhaylina, A. Z. Ksibe, D. J. Scanlan, C. A. Blindauer, Bacterial zinc uptake regulator proteins and their regulons. *Biochem. Soc. Trans.* **46**, 983–1001 (2018).
27. D. Lucarelli *et al*., Crystal structure and function of the zinc uptake regulator FurB from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **282**, 9914–9922 (2007).
28. J. H. Shin *et al*., Graded expression of zinc-responsive genes through two regulatory zinc-binding sites in Zur. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5045–5050 (2011).
29. D. P. Giedroc, A. I. Arunkumar, Metal sensor proteins: Nature's metalloregulated allosteric switches. *Dalton Trans.*, 3107–3120 (2007).
30. S. Barber-Zucker, B. Shaanan, R. Zarivach, Transition metal binding selectivity in proteins and its correlation with the phylogenomic classification of the cation diffusion facilitator protein family. *Sci. Rep.* **7**, 1–12 (2017).
31. X. Cao *et al*., Identification of metal ion binding sites based on amino acid sequences. *PLoS One* **12**, e0183756 (2017).
32. R. Phillips, J. Kondev, J. Theriot, H. Garcia, *Physical Biology of the Cell* (Garland Science, 2012).
33. D. L. Nelson, A. L. Lehninger, M. M. Cox, *Lehninger Principles of Biochemistry* (Macmillan, 2008).
34. M. Babor, S. Gerzon, B. Raveh, V. Sobolev, M. Edelman, Prediction of transition metal-binding sites from apo protein structures. *Proteins Struct. Funct. Bioinf.* **70**, 208–217 (2008).
35. C. E. Housecroft, A. G. Sharpe, *Inorganic Chemistry* (Pearson Education, 2018), vol. 1.
36. A. E. Martell, R. D. Hancock, *Metal Complexes in Aqueous Solutions* (Springer Science & Business Media, 2013).
37. R. D. Hancock, F. Marsicano, The chelate effect: A simple quantitative approach. *J. Chem. Soc. Dalton Trans.* **1976**, 1096–1098 (1976).
38. I. Persson, Structure and size of complete hydration shells of metal ions and inorganic anions in aqueous solution. *Dalton Trans.* **53**, 15517–15538 (2024).
39. E. Sl, A. E. Martell, *Stability Constants of Metal-Ion Complexes* (The Chemical Society, London, 1964), vol. Vol. 1.
40. Z. Xiao, A. G. Wedd, The challenges of determining metal-protein affinities. *Nat. Prod. Rep.* **27**, 768–789 (2010).
41. T. R. Young, Z. Xiao, Principles and practice of determining metal-protein affinities. *Biochem. J.* **478**, 1085–1116 (2021).
42. J. F. Da Silva, R. J. P. Williams, *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life* (Oxford University Press, 2001).
43. J. Stanley-Gray, Z. Zhang, D. Venkataraman, Updated coordination geometry table of the d-block elements and their ions. *J. Chem. Educ.* **98**, 2476–2481 (2021).
44. M. J. Winter, Chemdex: Quantification and distributions of valence numbers, oxidation numbers, coordination numbers, electron numbers, and covalent bond classes for the elements. *Dalton Trans.* **53**, 493–511 (2024).
45. H. T. Huang *et al*., Co (II) and Ni (II) binding of the *Escherichia coli* transcriptional repressor RcnR orders its N terminus, alters helix dynamics, and reduces DNA affinity. *J. Biol. Chem.* **293**, 324–332 (2018).
46. F. M. J. Chang *et al*., Cu (I)-mediated allosteric switching in a copper-sensing operon repressor (CsoR). *J. Biol. Chem.* **289**, 19204–19217 (2014).
47. M. A. Pennella, J. E. Shokes, N. J. Cosper, R. A. Scott, D. P. Giedroc, Structural elements of metal selectivity in metal sensor proteins. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3713–3718 (2003).
48. E. F. Bosma, M. H. Rau, L. A. van Gijtenbeek, S. Siedler, Regulation and distinct physiological roles of manganese in bacteria. *FEMS Microbiol. Rev.* **45**, fuab028 (2021).
49. C. H. Steingard, A. Pinochet-Barros, B. M. Wendel, J. D. Helmann, Iron homeostasis in *Bacillus subtilis* relies on three differentially expressed efflux systems. *Microbiology* **169**, 001289 (2023).
50. J. D. Helmann, Specificity of metal sensing: Iron and manganese homeostasis in *Bacillus subtilis*. *J. Biol. Chem.* **289**, 28112–28120 (2014).
51. Z. Ma, F. E. Jacobsen, D. P. Giedroc, Metal transporters and metal sensors: How coordination chemistry controls bacterial metal homeostasis. *Chem. Rev.* **109**, 4644 (2009).
52. D. R. Campbell *et al*., Mycobacterial cells have dual nickel-cobalt sensors: Sequence relationships and metal sites of metal-responsive repressors are not congruent. *J. Biol. Chem.* **282**, 32298–32310 (2007).
53. A. Gaballa, J. D. Helmann, *Bacillus subtilis* CPx-type ATPases: Characterization of Cd, Zn, Co and Cu efflux systems. *Biometals* **16**, 497–505 (2003).
54. D. Osman *et al*., Fine control of metal concentrations is necessary for cells to discern zinc from cobalt. *Nat. Commun.* **8**, 1884 (2017).
55. G. T. Antelo, A. J. Vila, D. P. Giedroc, D. A. Capdevila, Molecular evolution of transition metal bioavailability at the host-pathogen interface. *Trends Microbiol.* **29**, 441–457 (2021).
56. C. C. Murdoch, E. P. Skaar, Nutritional immunity: The battle for nutrient metals at the host-pathogen interface. *Nat. Rev. Microbiol.* **20**, 657–670 (2022).
57. C. M. C. Andrés, J. M. Pérez de la Lastra, C. A. Juan, F. J. Plou, E. Pérez-Lebeña, The role of reactive species on innate immunity. *Vaccines* **10**, 1735 (2022).