Supplementary Information: Un-evolved De Novo Proteins Have Innate Tendencies to Bind Transition Metals

Protein Sequence Information:

| PROTEIN | AMINO ACID SEQUENCE |
|---------|---------------------|
| S-824   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| S-824  - HC | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| S-824  - HZ | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 1   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 2   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 3   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 4   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 5   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 6   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 7   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 8   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 9   | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 10  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 11  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 12  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 13  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 14  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 15  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 16  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 17  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 18  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 19  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 20  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
| NMB 21  | MYGKLNLLELDQVBLNKLHKNHGGKDNLHVDNNLGNVMQIEH1HDPMQG |
## Binding data for selected proteins

**Table S2.** Binding screen for 52 well expressed Naïve Metal Binding (NMB) Proteins. The amount of de novo protein that remained bound to the metalated bead after stringent washing was compared to the amount initially loaded. Proteins were grouped into three classes: If 0-33% of the protein remained bound to the bead, the protein was designated a weak binder (+); if 33-66% of the protein remained, it was classified as a moderate metal binder (++); and if 66-100% remained bound, it was classified as a strong binder (+++).

| Protein | Co(II) | Cu(II) | Zn(II) |
|---------|--------|--------|--------|
| S-824   | +      | +++    | ++     |
| S-824-HC| +      | -      | -      |
| S-824-HZ| -      | -      | -      |
| NMB 1   | +++    | ++     | +      |
| NMB 2   | ++     | +      | ++     |
| NMB 3   | ++     | +++    | ++     |
| NMB 4   | +++    | ++     | +      |
| NMB 5   | +      | +      | +++    |
| NMB 6   | ++     | +++    | +      |
| NMB 7   | +      | ++     | +      |
| NMB 8   | +      | ++     | +      |
| NMB 9   | ++     | ++     | +      |
| NMB 10  | +      | ++     | +      |
| NMB 11  | +      | +      | ++     |
| NMB 12  | +      | +++    | +      |
| NMB 13  | +      | ++     | ++     |
| NMB 14  | ++     | +++    | +      |
| NMB 15  | +      | +++    | -      |
| NMB 16  | +      | ++     | +      |
| NMB 17  | +      | +++    | +      |
| NMB 18  | +      | ++     | +      |
| NMB 19  | -      | ++     | +      |
| NMB 20  | +++    | +++    | +      |
| NMB 21  | +      | +++    | -      |
| NMB 22  | +      | -      | +      |
| NMB 23  | +      | ++     | +      |
| NMB 24  | +      | ++     | +      |
| NMB 25  | -      | ++     | -      |
Table S3. Summary of ITC binding information from Table 1. This is compared with the binding strength estimated from binding to the metalated beads. All binding curves are shown in Figure S2.

| Protein | Metal | ∆H<sub>app</sub> (kJ/mol) | -T∆S<sub>app</sub> (kJ/mol) | ∆G<sub>app</sub> (kJ/mol) | N | K<sub>d</sub> (µM) | N<sub>app</sub> | K<sub>d,app</sub> (nM) | Screen |
|---------|-------|--------------------------|-----------------------------|--------------------------|---|----------------|--------------|----------------|--------|
| NMB 39  | Co(II) | -7.36 ± 0.3              | -36.9                       | -44.3                    | 1 | 0.020 ± 0.007 | 1.5          | 700            | +      |
|         |       | -2.25 ± 0.27             | -35.2                       | -37.3                    | 2 | 0.278 ± 0.01  |             |                |        |
|         | Cu(II) | -13.3 ± 2.2              | -18.3                       | -31.6                    | 2 | 1.45 ± 0.82   | 1.5          | 700            | +++    |
|         | Zn(II) | -33.5 ± 3.8              | -1.75                       | -31.8                    | 3 | 2.74 ± 0.94   | 3            | 1000           | ++     |
|         |        |                          |                             |                          |   |                |              |                |        |
| HisZero | Co(II) | -47.4 ± 7.0              | 15.1                        | -31.5                    | 1 | 2.17 ± 0.89   | 2            | 700            | -      |
|         | Cu(II) | -19.6 ± 2.1              | -13.1                       | -32.7                    | 2 | 1.92 ± 0.82   | 1            | 600            | ++     |
|         | Zn(II) | -44.5 ± 3.0              | 11.2                        | -33.3                    | 1.5| 1.50 ± 0.42   | 4            | 1000           | ++     |
|         |        |                          |                             |                          |   |                |              |                |        |

* 20 nM is near the Limit of Detection for ITC. This is the best fit of a two-site model, but should be viewed as an estimate rather than a precise determination.
**Apparent Dissociation Constant:**

The dissociation constant \( K_d \) for one site binding one metal describes the following system at equilibrium:

\[
[M]_{\text{free}} + [\text{Receptor}]_{\text{free}} \xrightleftharpoons[k_{-1}]{k_1} [\text{Metal:Receptor}]_{\text{bound}}
\]  

(1)

With these, the \( K_d \) of a single site is defined as the ratio of dissociation rate to association rate:

\[
K_d = \frac{k_{-1}}{k_1} = \frac{[\text{Metal}]_{\text{free}}[\text{Receptor}]_{\text{free}}}{[\text{Metal:Receptor}]_{\text{bound}}} = \frac{[M]_f[R]_f}{[R]_B}
\]  

(2)

Which can be written in terms of bound receptor ([R]_B) and total receptor ([R]_T):

\[
K_d = \frac{[M]_f ([R]_T - [R]_B)}{[R]_B}
\]  

(3)

Which can be rewritten as:

\[
\frac{[R]_B}{[R]_T} = \frac{[M]_f}{K_d + [M]_f}
\]  

(4)

Equation (5) is what is used to fit the \( K_d \) curve for a single binding site.

For multiple independent binding sites with somewhat overlapping affinity, as we find in our de novo proteins, the observed binding is a linear combination of the contributing events. We denote the apparent affinity across \( n \) similar binding sites \( K_{d,\text{app}} \), and define it as the average of all component dissociation constants \( K_{d,n} \).

\[
K_{d,\text{app}} = \frac{K_{d,1} + K_{d,2} + \cdots + K_{d,n}}{n} = \frac{[M]_f ([R]_T - [R]_{B,1}) + [M]_f ([R]_T - [R]_{B,2}) + \cdots + [M]_f ([R]_T - [R]_{B,n})}{[R]_{B,1} + [R]_{B,2} + \cdots + [R]_{B,n}}
\]  

(5)

Because all receptors are on the same protein of concentration [R]_T and total metal is in equilibrium across the system, the only additional variable is the metal bound to each receptor [R]_{B,n}. This can be rewritten to mirror equation (4):

\[
\frac{1}{n} \left( \frac{[M]_f[R]_T}{[R]_{B,1}} + \frac{[M]_f[R]_T}{[R]_{B,2}} + \cdots + \frac{[M]_f[R]_T}{[R]_{B,n}} \right) = K_{d,\text{app}} + [M]_f
\]  

(6)

\[
\frac{[R]_{B,1} + [R]_{B,2} + \cdots + [R]_{B,n}}{[R]_T} = \frac{[M]_f}{K_{d,\text{app}} + [M]_f}
\]  

(7)

\[
\frac{[R]_{B,\text{app}}}{[R]_T} = \frac{[M]_f}{K_{d,\text{app}} + [M]_f}
\]  

(8)

\([R]_{B,\text{app}} \) is the bound concentration measured in the experiment, and is the total bound across all sites. Thus, the left-hand side of equation (8) is the “Bound Equivalents” y-axis in all equilibrium dialysis plots. Equation (8) is what is used to fit the \( K_{d,\text{app}} \) curve for all de novo proteins.
Binding curves for selected proteins

Figure S1: Binding curves for all proteins characterized by equilibrium dialysis with Co$^{2+}$, Cu$^{2+}$, and Zn$^{2+}$.

HisConserved - Co$^{2+}$

| $K_{d,app}$ (nM) | 900 | Max Equivalents (mol/mol) | 1.5 |
|-----------------|-----|---------------------------|-----|

HisConserved - Cu$^{2+}$

| $K_{d,app}$ (nM) | 300 | Max Equivalents (mol/mol) | 1   |
|-----------------|-----|---------------------------|-----|
HisConserved - Zn$^{2+}$

| $K_{d,app}$ (nM) | 600 | Max Equivalents (mol/mol) | 2 |
|------------------|-----|---------------------------|---|

NMB11 - Co$^{2+}$

| $K_{d,app}$ (nM) | 600 | Max Equivalents (mol/mol) | 3 |
Kd,app (nM) 800 Max Equivalents (mol/mol) 3

NMB11 - Cu²⁺

Kd,app (nM) 300 Max Equivalents (mol/mol) 4

NMB11 - Zn²⁺
K\text{d,app (nM)} & 300 & \text{Max Equivalents (mol/mol)} & 1.5 \\

K\text{d,app (nM)} & 600 & \text{Max Equivalents (mol/mol)} & 2 \\

NMB20 - Co^{2+} \\

NMB20 - Cu^{2+}
**NMB20 - Zn\(^{2+}\)**

| \(K_d,_{\text{app}} \, (\text{nM})\) | 800 | Max Equivalents (mol/mol) | 3 |

**NMB24 - Co\(^{2+}\)**

| \(K_d,_{\text{app}} \, (\text{nM})\) | 2300 | Max Equivalents (mol/mol) | 3 |
NMB25 - Co^{2+}

| $K_{d,\text{app}}$ (nM) | 1200 | Max Equivalents (mol/mol) | 3 |

NMB25 - Cu^{2+}

| $K_{d,\text{app}}$ (nM) | 1200 | Max Equivalents (mol/mol) | 2 |
| K_d,app (nM) | 500 | Max Equivalents (mol/mol) | 4 |
|-------------|-----|--------------------------|---|

NMB25 - Zn^{2+}

| K_d,app (nM) | 1600 | Max Equivalents (mol/mol) | 3 |
|-------------|------|--------------------------|---|

NMB37 - Co^{2+}
NMB37 - Cu$^{2+}$

| $K_{d,\text{app}}$ (nM) | 200 | Max Equivalents (mol/mol) | 2 |

NMB37 - Zn$^{2+}$

| $K_{d,\text{app}}$ (nM) | 1200 | Max Equivalents (mol/mol) | 4 |
NMB39 - Co^{2+}

| $K_{d,app}$ (nM) | 700 | Max Equivalents (mol/mol) | 2 |

NMB39 - Cu^{2+}

| $K_{d,app}$ (nM) | 600 | Max Equivalents (mol/mol) | 1 |
**NMB39 - Zn\textsuperscript{2+}**

| $K_{d,app}$ (nM) | 1000 | Max Equivalents (mol/mol) | 4 |

**S824 - Co\textsuperscript{2+}**

| $K_{d,app}$ (nM) | 700 | Max Equivalents (mol/mol) | 1.5 |
S824 - Cu$^{2+}$

| $K_{d,\text{app}}$ (nM) | 700 | Max Equivalents (mol/mol) | 1.5 |

S824 - Zn$^{2+}$

| $K_{d,\text{app}}$ (nM) | 1000 | Max Equivalents (mol/mol) | 3 |
Binding is proportional to Free metal - nonspecific

**HisZero - Co\(^{2+}\)**

![Graph showing binding is proportional to Free metal - nonspecific for HisZero - Co\(^{2+}\)].

**Graph Details:**
- Equation: $y = 0.083x$
- $R^2 = 0.8426$
- X-axis: [Free] (µM)
- Y-axis: Bound Equivalents (mol / mol)

**HisZero - Cu**

![Graph showing binding is proportional to Free metal - nonspecific for HisZero - Cu].

**Graph Details:**
- Equation: $y = 0.1178x$
- $R^2 = 0.9845$
- X-axis: [Free] (µM)
- Y-axis: Bound Equivalents (mol / mol)
Above 5µM free zinc, the nonspecific binding is apparent:

Accounting for nonspecific binding gives the following curve:
HisZero - Zn\(^{2+}\) - nonspecific adjusted

| K\(_{d,\text{app}}\) (nM) | 1500 | Max Equivalents (mol/mol) | 1 |
Figure S2: Binding curves for all proteins characterized by ITC with Co\textsuperscript{2+}, Cu\textsuperscript{2+}, and Zn\textsuperscript{2+}. Raw data is presented at the top while the fit curve is below. Tabulated below each figure is the protein and metal concentrations used, and the determined affinity and thermodynamic terms, or the raw enthalpy of dilution for nonbinding events. Noise in S-824 binding copper is due to metal-mediated precipitation.

| Protein  | 20µM S-824 | Metal  | 500µM Co(II) |
|----------|------------|--------|--------------|
| $N_1$:   | 0.94 ± 0.03| $K_{D,1}$: | 20 ± 7 nM * |
| $\Delta H_1$: | -7.36 ± 0.3 kJ/mol | $-T\Delta S_1$: | -36.9 kJ/mol | $\Delta G_1$: | -44.3 kJ/mol |
| $N_2$:   | 2.11 ± 0.06| $K_{D,2}$: | 0.278 ± 0.01 µM |
| $\Delta H_2$: | -2.25 ± 0.27 kJ/mol | $-T\Delta S_2$: | -35.2 kJ/mol | $\Delta G_2$: | -37.3 kJ/mol |

* 20 nM is near the Limit of Detection for ITC. This is the best fit of a two-site model, but should be viewed as an estimate rather than a precise determination.
ITC noise at t > 29min is the metal-induced precipitation of the protein.

**Protein:** 10μM S-824  
**Metal:** 250μM Cu(II)  
**N:** 1.81 ± 0.09  
**Kₜ:** 0.853 ± 0.442 μM  
**ΔH:** -12.3 ± 1.2 kJ/mol  
**TΔS:** -22.4 kJ/mol  
**ΔG:** -34.7 kJ/mol
**Protein:** 10µM S-824  
**Metal:** 250µM Zn(II)  

| Parameter | Value |
|-----------|-------|
| N         | 3.02 ± 0.20 |
| K_0       | 2.74 ± 0.94 µM |
| ΔH        | -33.5 ± 3.8 kJ/mol |
| -TΔS      | 1.75 kJ/mol |
| ΔG        | -31.8 kJ/mol |
Protein: 10µM NMB 39  
Metal: 250µM Co(II)  
N: 0.949 ± 0.07  
K_D: 2.17 ± 0.89 µM  
ΔH: -47.4 ± 7.0 kJ/mol  
-TΔS: 15.1 kJ/mol  
ΔG: -31.5 kJ/mol
**Protein**: 10µM NMB 39

**Metal**: 250µM Cu(II)

| Protein | 10µM NMB 39 |
|---------|-------------|
| **N**:   | 2.05 ± 0.10 |
| **K_D**: | 1.92 ± 0.82 µM |
| **ΔH**:  | -19.6 ± 2.1 kJ/mol |
| **-TΔS**:| -13.1 kJ/mol |
| **ΔG**:  | -32.7 kJ/mol |
Protein: 10µM NMB 39
Metal: 250µM Zn(II)

$N: 1.51 \pm 0.05$

$K_D: 1.50 \pm 0.42 \mu M$

$\Delta H: -44.5 \pm 3.0 \text{ kJ/mol}$

$-T\Delta S: 11.2 \text{ kJ/mol}$

$\Delta G: -33.3 \text{ kJ/mol}$
**Protein:** 10µM HisZero  
**Metal:** 250µM Co(II)  
**Raw Heat Average:** $-1.486 \pm 0.573$ kJ/mol
Protein: 10µM HisZero  
Metal: 100µM Cu(II)  
Raw Heat Average: -2.7706 ± 1.2599 kJ/mol
| Protein         | 10µM HisZero |
|-----------------|-------------|
| Metal           | 250µM Zn(II) |
| $N$             | 1.09 ± 0.05 |
| $K_D$           | 0.961 ± 0.350 µM |
| $\Delta H$      | -8.23 ± 0.665 kJ/mol |
| $-T\Delta S$    | -26.1 kJ/mol |
| $\Delta G$      | -34.4 kJ/mol |
Amino Acid Abundance-Affinity Correlation:

Figure S3: Lack of correlation between the abundance of metal binding residues (His, Asp, or Glu) and the observed binding to cobalt immobilized on a bead. The percentage of metal bound (i.e. remaining on the bead after stringent washes) is on the x axis, and the number of potential metal-binding residues is on the y axis.
Figure S4: Lack of correlation between the abundance of metal binding residues (His, Asp, or Glu) and the observed binding to copper immobilized on a bead. The percentage of metal bound (i.e. remaining on the bead after stringent washes) is on the x axis, and the number of potential metal-binding residues is on the y axis.
Figure S5: Lack of correlation between the abundance of metal binding residues (His, Asp, or Glu) and the observed binding to zinc immobilized on a bead. The percentage of metal bound (i.e. remaining on the bead after stringent washes) is on the x axis, and the number of potential metal-binding residues is on the y axis.