## **Electronic supplementary information**

Cysteine containing dipeptides show metal specificity that matches the composition of seawater

Luca Belmonte, Daniele Rossetto, Michele Forlin, Simone Scintilla, Claudia Bonfio, and Sheref S. Mansy

CIBIO, University of Trento, via Sommarive 9, Povo, Italy.

# **Materials and Methods**

# **Bioinformatic Calculations**

PDB entries, up to October 2014, that contained coordinated mononuclear metal ions and iron-sulphur clusters were downloaded and analyzed with an in-house script that determined the frequency of residues in positions  $\pm 1$  relative to the ligating cysteine. In total, 13,600 PDB entries were analyzed, including 1,151 iron proteins, 538 cobalt proteins, 384 nickel proteins, 964 copper proteins, 9,529 zinc proteins, 446 [2Fe-2S] proteins, and 588 [4Fe-4S] proteins.

### **DFT Calculations: Geometry Optimization and Bonded Interactions**

Geometry optimization of  $[(CH_3S)_4Fe]^{2^-}$  complexes extracted from the PDB ID 1IRO were carried out with DFT calculations at the B3LYP/TZV+(2d,p) level of theory for the ligands and the LANL2TZ+ plus Effective Core Potentials (ECP)<sup>1-4</sup> for the metals. Screening of DFT calculations with B3LYP, PBE0 and PW91 revealed that B3LYP was faster while giving similar results to PBE0 and PW91. The initial geometry was optimized in vacuum starting with ionic fragments.<sup>7</sup> An additional two optimization steps were then performed with decreasing energy thresholds followed by a third optimization step with the Polarizable Continuum Model (PCM)<sup>8</sup> to account for solvent effects. Optimizations of  $[(CH_3S)_4M]^{2^-}$ , where M was either a copper, zinc, cobalt, or nickel ion, were carried out under the same conditions and at the same level of theory used for the iron complex. Ionic radii for high spin and tetrahedrally coordinated M<sup>2+</sup> were as previously reported.<sup>9</sup> Insight into the quality of the optimized structures was gained by superimposing the calculated structures with PDB deposited structures. Missing parameters for ligand-metal bonded interactions were calculated on the optimized structures as previously described.<sup>10</sup> MacMolPlot<sup>11</sup> and Avogadro<sup>12</sup> were used to manipulate GAMESS-US input files. Basis sets were downloaded from EMSL.<sup>14,15</sup> Calculations were with GAMESS-US.<sup>5,6</sup> All calculated parameters are available as a separately downloadable supplementary files.

# **MP2** Calculations: Non-Bonded Interactions

Interaction energies depicted in Figure 2 were calculated in gas phase as

$$E_{\text{interaction}} = E_{\text{complex}} - (E_{\text{ion}} + E_{\text{ligand}})$$
(1.1)

where  $E_{complex}$  is the point energy of the entire complex, four methanethiolates coordinated to the metal dication, and  $E_{ion}$  and  $E_{ligand}$  are the point energies of the ion alone and of the four methanethiolates alone, respectively. Lennard-Jones potentials were recalculated. To do so, the Simple but Accurate Method proposed by Karplus and co-worker,<sup>18</sup> as later implemented,<sup>19</sup> was used. One of the ligands was fixed ([(CH <sub>3</sub>S)<sub>4</sub>Fe<sup>2+</sup>]<sup>2-</sup>) to the origin of the system and the M<sup>2+</sup> was moved toward the Sulfur atom of the fixed ligand starting from a distance of 10 Å in order to obtain a Potential Energy function. A sampling step of 0.05 Å in the region 1.1 to 2.05 Å was used, which roughly corresponded to the metal ion-ligand distance previously found in optimized geometries. We used a different sampling of 0.1 Å in

the region 2 to 4 Å. The cut-off was fixed at 10 Å. Interaction energies between metal ions and ligands are well described by the usage of the Møller-Plesset perturbation theory (MP2) used in these calculations. The metal ion-ligand energy for the metal ion-methanethiolate system was then calculated as follows<sup>19</sup>:

$$\Delta E = E_{\text{complex}} - (E_{\text{ion}} + E_{\text{ligand}})$$
(1.2)

The obtained functions were then fit to the following potential function:

$$V = 332C \sum_{j} \frac{Q_{i}q_{j}}{r_{ij}^{2}} + \sum_{j} \frac{-A_{1}A_{j}}{r_{ij}^{6}} + \sum_{j} \frac{B_{1}B_{j}}{r_{ij}^{12}} + D \quad [kcal/mol]$$
(1.3)

where  $Q_i$  and  $q_j$  were the calculated Merz-Kollman charges for the metal dication and the coordinated sulfur (table S1) at the distance at which the well of the potential energy function was found (Figure S5). A<sub>1</sub> and B<sub>1</sub> were equal to 793.3 and 25.01, respectively, as previously described.<sup>20</sup> All calculated parameters are available as a separately downloadable supplementary files.

#### **Molecular Dynamics**

The cysteine coordinates of *Clostridium pasteurianum* rubredoxin (PDB ID: 1IRO) along with the coordinated iron centre were extracted. The coordinates were then manipulated with Avogadro to build the other metallocomplexes; i.e. to add different peptide ligands. All of the complexes were solvated in water, neutralized and equilibrated for 20 ps. The complexes were heated to 298 K in a stepwise manner. A constant pressure of 1 atm was used to have a NPT ensemble. The calculation interval for the equations of motion was 2 fs. Long-range electrostatic interactions were calculated using the Ewald approximation and periodic boxes (PBC). The SHAKE<sup>21</sup> procedure was employed to constrain hydrogen atoms. Non-bonded interactions were up to a complete cut-off of 10 Å. The Ewald sum was computed using the Particle-Mesh Ewald (PME)<sup>22</sup>. The Langevin algorithm, as implemented in NAMD,<sup>23</sup> was also used. Molecular Dynamics (MDs) ran for 10 ns. Analysis of the trajectories was performed using VMD.<sup>24</sup> MDs were performed with a customized version of Charmm27 Force Field<sup>25–27</sup> topology and parameter files obtained by the DFT and MP2 calculations described above. The Solvent Accessible Surface Area (SASA) was calculated with VMD, and PDB files were manipulated with UCSF Chimera.<sup>13</sup>

Molecular dynamics data were analysed with an in-house prepared R package *MoDyGliAni* (Molecular Dynamics Global Analysis) available as a separately downloadable supplementary file and at http://smansy.org/modygliani. MoDyGliAni works on a set of tab separated ASCII files that are the result of one or more MD runs. Input files are in the format time step, energy (kcal/mol), RMSD (Å). MoDyGliAni first attempts at the RMSD curve fitting with a double negative exponential function (eq 2.1). If the RMSD distribution is not fit by the double exponential function, MoDyGliAni then attempts fitting with a negative exponential function (eq 2.2). If the RMSD distribution is not fit by the negative exponential function, the user must define the time threshold. The two exponential functions are:

$$RMSD(t) = k \cdot e^{-\lambda_1 \tau_1} \cdot e^{-\lambda_2 \tau_2} \quad (2.1)$$

$$RMSD(t) = k \cdot e^{-\lambda_1 \tau_1}$$
 (2.2)

where k,  $\lambda_1$  and  $\lambda_2$  are constants. This process is iterated for all the input files provided. Then, among the time constants ( $\tau$ ) provided by the fitting, MoDyGliAni searches for the longest time constant ( $\tau_s$ ) to ensure that comparisons are made between systems in the production phase. The potential energy surface (PES) and RMSD are considered to be in the equilibration phase if PES is  $< 5\tau$ , otherwise the PES is considered in the production phase. MoDyGliAni provides as an output a histogram of comparison of the <RMSD> and <E<sub>internal</sub>> for each complex trajectory given as input in the production phase relative to the slowest  $\tau$ ; i.e.  $\geq 5\tau_s$ . Parameters of the histogram are delivered in csv format. MoDyGliAni also provides the complete charts of the RMSD(t), E<sub>internal</sub>(t), histograms of RMSD and of E<sub>internal</sub>> are shown in Figure 3 and Figure S8.

## Materials

All reagents were from Sigma-Aldrich and used without any further purification. Deionized Milli-Q (Millipore) purified water was distilled under nitrogen flow to deoxygenate the solvent. Ligand solutions were obtained under controlled nitrogen atmosphere by using a Schlenk line and Schlenk glassware and transferred to anaerobic sealed Hellma quartz cuvettes with a septum. *In situ* synthesis of peptido-metal complexes was obtained by injecting each metal ion solution through the cuvette septum with Hamilton gastight syringes.

### Solid phase peptide synthesis

The synthesis of C- and N- blocked peptides was performed according to standard Fmoc-based SPPS procedures.<sup>28</sup> N.N-dimethyl formamide (DMF) was used as the solvent and Rink acid-labile (hydroxymethyl)polystyrene resin was used as the starting polymeric support. Fmoc-protected aminoacids were used as building blocks. Peptide elongation was obtained by Fmoc-deprotection of the residue anchored to the resin and Fmoc-AA-OH coupling. Fmoc-deprotection was obtained by washing the mixture with 20% (v/v) solution of piperidine in DMF. For each coupling, an excess (Fmoc-AA-OH: anchored AA, 3:1) of the Fmoc- $\alpha$ -amino acid derivative was added to the resin. Apart from Fmoc-Cys(Trt)-OH, Fmoc- $\alpha$ -amino acid derivatives were activated with a mixture of hydroxyl-benzotriazole (HOBt), N,N,N',N'-tetramethyl-O-(benzotriazol-1-yl)uronium tetrafluoborate (TBTU), and N,Ndiisopropylethyl amine (DIPEA). Fmoc-Cys(Trt)OH was activated with a N.N'diisopropylcarbodiimide (DIC)/HOBt mixture. At the end of the synthesis, the last Fmoc-protecting group was removed, and the acetylation of the N-terminus was performed by shaking with 25% acetic anhydride in DMF in the presence of DIPEA. The blocked-peptides were cleaved from the resin and deprotected by treatment with a solution of trifluoroacetic acid (TFA):H<sub>2</sub>O:triisopropyl silane (TIS):1,2-ethanedithiol (EDT) (volume ratio 37:1:1:1) for 2 h. The volume was reduced under nitrogen atmosphere to avoid cysteinyl-thiol oxidation, and the product was precipitated with a cold solution of diethyl ether followed by washing cycles with diethyl ether or extracted three times with 20% acetic acid/chloroform and finally dried under inert atmosphere. Peptides were confirmed by mass spectrometry.

# Mass spectrometry

Mass spectra were acquired at the Proteomics/MS unit of Cogentech S.c.a.r.l. (Fondazione IFOM-Istituto FIRC di Oncologia Molecolare, Milano). Samples were resuspended in 1 mL of HPLC grade H<sub>2</sub>O and directly infused on a quadrupole Orbitrap Q-Exactive HF mass spectrometer (Thermo Scientific) with H-ESI Ion Max source. Ionization was achieved at a flow rate of 5  $\mu$ L/min in positive ion mode applying +3.5 kV at the entrance of the capillary. Sheat gas was set at 5 psi, capillary temperature 320 °C, S-Lens RF Level 60, S-Lens Voltage 21, FT Resolution 35000.

#### **UV-Visible spectroscopy**

UV-Vis absorption spectra of freshly prepared solutions of peptido-metal complexes were recorded with an Agilent 8453 UV-Vis diode array spectrophotometer with an integration time of 0.5 s and an interval of 1 nm. Baseline subtraction was made at 900 nm.

#### Saturation binding assay

As a general procedure for iron, cobalt, and nickel ions, aliquots of 25 mM metal salt solution (FeCl<sub>2</sub>, CoCl<sub>2</sub>·6H<sub>2</sub>O, NiSO<sub>4</sub>·7H<sub>2</sub>O, respectively) were injected into the cuvette to titrate 1 mL of an aqueous solution containing 2.5 mM ligand (Ac-(AA)<sub>x</sub>-NH<sub>2</sub>, x = 1 or 2) at pH 8.7. UV-Vis spectra were collected upon each addition. Absorbance values at a fixed wavelength (380 nm, 750 nm, and 535 nm for iron, cobalt, and nickel ion titrations, respectively) were monitored until no changes were observed.

#### **Competition binding assay**

To overcome the spectroscopic silence of Zn(II), competitive binding experiments involving a preformed peptido-Co(II) complex, with the characteristic absorbance at 750 nm, was exploited. A decrease in the absorbance at the fixed wavelength was observed upon titration of the spectroscopic probe with 25 mM ZnSO<sub>4</sub>. The concentration of cobalt added to each peptide before the competition assay was coincident with the K<sub>d</sub> value of that complex. UV-Vis spectra were collected upon each addition. Titrations continued until no changes in absorbance at 750 nm were observed.

#### Determination of K<sub>d</sub>

GraphPad Prism v. 6.00 (GraphPad Software, La Jolla California USA) for Windows was used to calculate the  $K_d$  values. For Fe(II), Co(II), and Ni(II) peptido-complexes, the  $K_d$  values were calculated by fitting the absorbance data to the equation:

$$y = \frac{B_{max} * x^h}{K_d^h + x^h} \tag{3.1}$$

where  $B_{max}$  was the absorbance reached at saturation, x the concentration, and h the Hill slope. The K<sub>d</sub> values for zinc(II) complexes were calculated by fitting to a revised Cheng-Prusoff equation, as previously described.<sup>29</sup>

	Fe <sup>2+</sup>			Co <sup>2+</sup>			Ni <sup>2+</sup>			Zn <sup>2+</sup>						
	Estimate	Std.Error	t value	Pr(> t )	Estimate	Std.Error	t value	Pr(> t )	Estimate	Std.Error	t value	Pr(> t )	Estimate	Std.Error	t value	Pr(> t )
Α	-4.38157	0.06781	-64.62	<2e-16	-4.015	0.0475	-84.55	<2e-16	-3.9636	0.03673	-107.91	<2e-16	-4.59856	0.08464	-54.33	<2e-16
В	-78.04319	3.44724	-22.64	<2e-16	-67.08	2.4189	-27.73	<2e-16	-76.92957	2.27465	-33.82	<2e-16	-92.94003	4.32701	-21.48	<2e-16
С	3.26262	0.23849	13.68	<2e-16	2.9341	0.1733	16.93	<2e-16	3.58236	0.14369	24.93	<2e-16	23.54132	1.36813	17.21	<2e-16
D	341.0894	4.94931	68.92	<2e-16	356.76	3.8596	92.43	<2e-16	375.77889	3.09712	121.33	<2e-16	291.38804	6.54115	44.55	<2e-16
ε	-0.00005	-	-	-	-0.00005	-	-	-	-0.00003	-	-	-	-0.00001	-	-	-
$r_{\rm min}/2$	0.9070	-	-	-	0.8973	-	-	-	0.9200	-	-	-	0.9262	-	-	-
Charge	0.5691	-	-	-	0.4934	-	-	-	0.4702	-	-	-	0.9322	-	-	-
Sulfur Charge	-0.3004	-	-	-	-0.3619	-	-	-	-0.3485	-	-	-	-0.042	-	-	-

Table S1. Fitting parameters of the metal dication-methanethiolate potential energy function as given in equation 1.2 and as shown in Fig S7.  $\varepsilon$  and  $r_{min}/2$  are given as mapped into the customized Charmm force field files.

	Zn <sup>2+</sup> (mM)	Co <sup>2+</sup> (mM)	Fe <sup>2+</sup> (mM)	Ni <sup>2+</sup> (mM)
Cys	0.044*	0.304*	0.170	0.186
Cys-Ala	0.046	0.082	0.157	0.215
Cys-Gly	0.029	0.088	0.324	0.380
Cys-Ile	0.008	0.054	0.266	0.212
Cys-Leu	0.031	0.066	0.264	0.139
Cys-Pro	0.026	0.038	0.100	0.142
Cys-Ser	0.010	0.064	0.185	0.351
Cys-Thr	0.016	0.045	0.208	0.136
Cys-Tyr	0.015	0.160	0.148	0.102
Cys-Val	0.045	0.119	0.241	0.183
Phe-Cys	0.060	0.245	0.292	0.212
Pro-Cys*	0.010	0.134	0.108	0.102
Gly-Cys	0.015*	0.093*	0.244	0.141

*Table S2. Experimentally determined K<sub>d</sub> values.* 

\*Measured in the presence of 2.5 mM TCEP (Tris(2-carboxyethyl)phosphine)

*Table S3. Average Solvent Accessible Area Surface (SASA) for the whole trajectory of all the complexes analyzed by MD.* 

	SASA (Å <sup>2</sup> )							
	Fe <sup>2+</sup>	Co <sup>2+</sup>	Ni <sup>2+</sup>	$Zn^{2+}$				
Cys	0.000	0.000	0.333	0.000				
CysAla	0.000	0.001	0.404	0.000				
CysGly	0.000	0.001	0.000	0.000				
CysPro	0.001	0.002	0.000	0.000				
CysThr	0.000	0.002	1.272	0.000				
CysVal	0.000	0.003	3.131	0.000				
GlyCys	0.000	0.000	0.000	0.000				
ProCys	0.000	0.000	0.000	0.000				



Figure S1. Protein ligand preference. Structures from the protein data bank were analyzed to determine the frequency of ligands for each metal centre. (Top) Analysis of iron-sulphur clusters, including 446 PDB entries of [2Fe-2S] proteins, 588 of [4Fe-4S] proteins, and 1151 of mononuclear iron proteins (total = 2185 PDB entries) shows that polynuclear iron-sulphur clusters prefer ligation by cysteine. (Bottom) 1151 PDB entries of iron proteins, 538 structures of cobalt proteins, 384 of nickel proteins, 964 of copper proteins, and 9529 of zinc proteins were analyzed resulting in a data set of 12566 PDB entries. Mononuclear metal coordination of clusters and Fe, Co, Ni, Cu and Zn occurs more frequently through cysteine, aspartate, glutamate, and histidine (not ordered by frequency) ligands, even though other residues can occasionally contribute.



Figure S2. The probability of the 20 amino acids to occupy the -1 and +1 positions in the primary sequence with respect to a  $M^{2+}$  coordinated cysteine. The probability of Val at the -1 position was ~0.34 and ~0.08 for iron and zinc ions, respectively. The probability of a Gly at the +1 position was ~0.33, ~0.28, and ~0.14 for iron, nickel, and zinc ions, respectively.



Figure S3. DFT test on the  $[(CH_3S)_4Fe]^{2-}$  complex. Geometry optimization results in a tetrahedral geometry for all of the performed DFT tests (top panel). The point energy per optimization step for each of the DFT calculation runs is shown (middle panel) and compared (bottom panel). Only a slight difference in energies was observed, i.e. -1.8757e+003, -1.8752e+003, and -1.8760e+003 au for B3LYP, PBE0, and PW91, respectively. In other words, the difference in energy was by 0.016% for B3LYP with respect to PW91 and by 0.042% with respect to PBE0. Additionally, only 38 steps were taken with B3LYP versus 64 steps for PBE0 and 81 steps for PW91. Therefore, the speed of calculation improved more than 2-fold for B3LYP with respect to PW91.



Figure S4. DFT test on the  $[(CH_3S)_4Zn]^{2-}$  complex. Geometry optimization results in a tetrahedral geometry for all of the performed DFT tests (top panel). The point energy per optimization step for each of the DFT calculation runs is shown (middle panel) and compared (bottom panel). The resulting energies were -1.8181e+003 au for B3LYP and -1.8185e+003 au for PW91 (a difference of 0.022%). 58 steps were taken for B3LYP and 60 for PW91.



Figure S5. Potential energy functions at MP2/TZV(2d,p)++ and LANL2TZ+ with ECP for metal centres for Fe<sup>2+</sup> (a) Co<sup>2+</sup> (b) Ni<sup>2+</sup> (c) and Zn<sup>2+</sup> (d). Dashed lines represent the QM calculated function while the non-dashed curve represents the fitted function. Minima are found at 2.8, 2.55, 2.5, and 2.25 Å for Fe<sup>2+</sup>, Co<sup>2+</sup>, Ni<sup>2+</sup>, and Zn<sup>2+</sup>, respectively.



Figure S6. Superposition of calculated structures and crystallographic coordinates of metal substituted rubredoxins.<sup>30</sup> Only crystallographic coordinates are shown on the left: tube representation is for the protein backbone, blue dots are for cysteine sidechain sulfurs. On the right, the superimposed structures are shown: blue tube representation is for the protein backbone, while the ball and stick portion of the figure represents carbon, sulfur, and metal ions from the DFT optimized geometries. Metal centres are always shown as spheres: orange is for iron, green for cobalt, grey for nickel, and magenta for zinc. PDB entries are 11RO for iron rubredoxin (a), 1R0H for cobalt-substituted rubredoxin (b), 1R0J for nickel-substituted rubredoxin (c), and 11RN for zinc-substituted rubredoxin (d). RMSDs are 0.256, 0.252, 0.384, and 0.38 Å, respectively.



Figure S7. Charmm27 customized parameters of (a) average force constants calculated by DFT and (b) charge of the metal ions and sulfur ligands by *ab initio* calculations.



Figure S8. Average RMSD from MD trajectories of the complexes: (a) Cys-Xxx and (b) Xxx-Cys. Histograms were made with MoDyGliAni.



Figure S9. Mass spectra of Cys and the dipeptides.

A) High resolution mass spectrum of Ac-(Cys)-NH<sub>2</sub> (blocked Cysteine) in aqueous solution is shown. The peak found at 163.65 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>5</sub>H<sub>10</sub>N<sub>2</sub>O<sub>2</sub>S, 162.05 Da).



B) High resolution mass spectrum of Ac-(Cys-Ala)-NH<sub>2</sub> (blocked Cysteine-Alanine) in aqueous solution is shown. The peak found at 234.08 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>8</sub>H<sub>15</sub>N<sub>3</sub>O<sub>3</sub>S, 233.08 Da).



C) High resolution mass spectrum for Ac-(Cys-Gly)-NH<sub>2</sub> (blocked Cysteine-Glycine) in aqueous solution is shown. The peak found at 220.07 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula  $C_7H_{13}N_3O_3S$ , 219.07 Da). The peak found at 439.14 m/z is consistent with the oxidized homodimer species.



D) High resolution mass spectrum for Ac-(Cys-Ile)-NH<sub>2</sub> (blocked Cysteine-Isoleucine) in aqueous solution is shown. The peak found at 276.13 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>11</sub>H<sub>21</sub>N<sub>3</sub>O<sub>3</sub>S, 275.13 Da).



E) High resolution mass spectrum for Ac-(Cys-Leu)-NH<sub>2</sub> (blocked Cysteine-Leucine) in aqueous solution is shown. The peak found at 276.13 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula  $C_{11}H_{21}N_3O_3S$ , 275.13 Da). The peak found at 551.26 m/z is consistent with the oxidized homodimer species.



F) High resolution mass spectrum for Ac-(Cys-Pro)-NH<sub>2</sub> (blocked Cysteine-Proline) in aqueous solution is shown. The peak found at 260.1 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>10</sub>H<sub>17</sub>N<sub>3</sub>O<sub>3</sub>S, 259.1 Da).



G) High resolution mass spectrum for Ac-(Cys-Ser)-NH2 (blocked Cysteine-Serine) in aqueous solution is shown. The peak found at 250.08 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>8</sub>H<sub>15</sub>N<sub>3</sub>O<sub>4</sub>S, 249.08 Da).



H) High resolution mass spectrum for Ac-(Cys-Thr)-NH<sub>2</sub> (blocked Cysteine-Threonine) in aqueous solution is shown. The peak found at 264.1 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula  $C_9H_{17}N_3O_4S$ , 263.09 Da). The peak found at 527.19 m/z is consistent with the oxidized homodimer species.

![](_page_22_Figure_0.jpeg)

I) High resolution mass spectrum for Ac-(Cys-Tyr)-NH<sub>2</sub> (blocked Cysteine-Tyrosine) in aqueous solution is shown. The peak found at 326.11 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>14</sub>H<sub>19</sub>N<sub>3</sub>O<sub>4</sub>S, 310.11 Da).

![](_page_23_Figure_0.jpeg)

J) High resolution mass spectrum for Ac-(Cys-Val)-NH<sub>2</sub> (blocked Cysteine-Valine) in aqueous solution is shown. The peak found at 262.11 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula  $C_{10}H_{19}N_3O_3S$ , 261.11 Da).

![](_page_24_Figure_0.jpeg)

K) High resolution mass spectrum for Ac-(Gly-Cys)-NH<sub>2</sub> (blocked Glycine-Cysteine) in aqueous solution is shown. The peak found at 220.07 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>7</sub>H<sub>13</sub>N<sub>3</sub>O<sub>3</sub>S, 219.07 Da).

![](_page_25_Figure_0.jpeg)

L) High resolution mass spectrum for Ac-(Pro-Cys)-NH<sub>2</sub> (blocked Glycine-Cysteine) in aqueous solution is shown. The peak found at 260.1 m/z is consistent with the value of the isotopic mass calculated for the protonated molecule ([M+1], formula C<sub>7</sub>H<sub>13</sub>N<sub>3</sub>O<sub>3</sub>S, 259.1 Da).

![](_page_26_Figure_0.jpeg)

Figure S10. Examples of metal ion titrations with the dipeptide Cys-Leu. (Top)  $\text{Co}^{2+}$ -dependent absorbance changes at 750 nm were monitored in the presence of 2.5 mM Cys-Leu. The data were fit to the Hill equation for saturation binding. (Bottom)  $\text{Zn}^{2+}$ -dependent absorbance changes at 750 nm in the presence of 2.5 mM Cys-Leu and 0.06 mM  $\text{Co}^{2+}$  were monitored. The data were fit to the revised Cheng-Prusoff equation for competitive binding.<sup>29</sup>

![](_page_27_Figure_0.jpeg)

Figure S11. The calculated average internal energy versus the measured  $K_d$  value for each metal ion. The values were averaged for each metal ion over the entire dipeptide set. The whole ensemble (represented by open circles) and the same ensemble from which SASA > 0 was excluded (filled circles) are shown. The correlation was 0.95 for the whole ensemble and 0.98 for the ensemble excluding SASA > 0.

![](_page_28_Figure_0.jpeg)

Figure S12. Scatterplot matrix of the average  $K_d$  value per each metal ion. The Pearson correlation coefficient is given for each pair of metal ions. For example, 0.75 is the coefficient related to  $Zn^{2+}$  and  $Co^{2+}$ , while 0.12 is the correlation coefficient for  $Zn^{2+}$  and  $Fe^{2+}$ . The associated scatter plots of  $K_d$  (mM) are also shown.

### References

- 1. Roy, L. E., Hay, P. J. & Martin, R. L. Revised basis sets for the LANL effective core potentials. *J. Chem. Theory Comput.* **4**, 1029–1031 (2008).
- 2. Wadt, W. R. & Hay, P. J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* **82**, 284–298 (1985).
- 3. Hay, P. J. & Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **82**, 299–310 (1985).
- 4. Hay, P. J. & Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.* **82**, 270–283 (1985).
- 5. Schmidt, M. W. *et al.* General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347–1363 (1993).
- 6. Gordon, M. S. & Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. *Theory Appl. Comput. Chem. first forty years* 1167–1189 (2005).
- 7. Szilagyi, R. K. & Winslow, M. a. On the accuracy of density functional theory for iron-sulfur clusters. *J. Comput. Chem.* **27**, 1385–1397 (2006).
- 8. Mennucci, B. Polarizable continuum model. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 386–404 (2012).
- 9. Shannon, R. D. t. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. Sect. A Cryst. Physics, Diffraction, Theor. Gen. Crystallogr.* **32**, 751–767 (1976).
- 10. Boatz, J. A. & Gordon, M. S. Decomposition of Normal-Coordinate Vibrational Frequencies. J. *Phys. Chem.* **93**, 1819–1826 (1989).
- 11. Bode, B. M. & Gordon, M. S. MacMolPlt: a graphical user interface for GAMESS. J. Mol. Graph. Model. 16, 133–138 (1998).
- 12. Hanwell, M. D. *et al.* Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **4**, 17 (2012).
- 13. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- 14. Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **17**, 1571–1586 (1996).
- 15. Schuchardt, K. L. *et al.* Basis set exchange: a community database for computational sciences. *J. Chem. Inf. Model.* **47**, 1045–1052 (2007).
- 16. Merz, K. M. Analysis of a large data base of electrostatic potential derived atomic charges. *J. Comput. Chem.* **13**, 749–767 (1992).
- 17. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
- 18. Stote, R. H. & Karplus, M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins Struct. Funct. Bioinforma.* **23**, 12–31 (1995).
- Ahlstrand, E., Spångberg, D., Hermansson, K. & Friedman, R. Interaction energies between metal ions (Zn 2+ and Cd 2+) and biologically relevant ligands. *Int. J. Quantum Chem.* 113, 2554–2562 (2013).
- 20. Aqvist, J. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. J. Phys. Chem. 94, 8021–8024 (1990).
- 21. Ryckaert, J.-P., Bellemans, A., Ciccotti, G. & Paolini, G. V. Shear-rate dependence of the viscosity of simple fluids by nonequilibrium molecular dynamics. *Phys. Rev. Lett.* **60**, 128 (1988).
- 22. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

- 23. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
- 24. Humphrey, W., Dalke, A. & Schulten, K. VMD Visual Molecular Dynamics. J. Mol. Graph. 14, 33–38 (1996).
- 25. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
- 26. MacKerel Jr., A. D. *et al.* in (ed. v. R. Schleyer et al., P.) **1**, 271–277 (John Wiley & Sons: Chichester, 1998).
- 27. Brooks, B. R. *et al.* CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
- 28. Jensen, K. J., Shelton, P. T. & Pedersen, S. L. *Peptide synthesis and applications*. (Springer, 2013).
- 29. Cheng, H. C. The power issue: Determination of KB or Ki from IC50 A closer look at the Cheng-Prusoff equation, the Schild plot and related power equations. *J. Pharmacol. Toxicol. Methods* **46**, 61–71 (2001).
- 30. Maher, M., Cross, M., Wilce, M. C. J., Guss, J. M. & Wedd, A. G. Metal-substituted derivatives of the rubredoxin from Clostridium pasteurianum. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 298–303 (2004).