



**HAL**  
open science

# Structure, dynamics and transferability of the metal-dependent polyhistidine tetramerization motif TetrHis for single-chain Fv antibodies

Robert Healey, Louise Couillaud, François Hoh, Assia Mouhand, Aurelien Fouillen, Pierre Couvineau, Sébastien Granier, Cédric Leyrat

► **To cite this version:**

Robert Healey, Louise Couillaud, François Hoh, Assia Mouhand, Aurelien Fouillen, et al.. Structure, dynamics and transferability of the metal-dependent polyhistidine tetramerization motif TetrHis for single-chain Fv antibodies. *Communications Chemistry*, 2023, 6 (1), pp.160. 10.1038/s42004-023-00962-x . hal-04181870

**HAL Id: hal-04181870**

**<https://hal.science/hal-04181870>**

Submitted on 17 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Structure, dynamics and transferability of the metal-dependent polyhistidine tetramerization motif TetrHis for single-chain Fv antibodies

Robert D. Healey<sup>1,3</sup>, Louise Couillaud<sup>1</sup>, François Hoh<sup>2</sup>, Assia Mouhand<sup>2</sup>, Aurelien Fouillen<sup>1</sup>, Pierre Couvineau<sup>1</sup>, Sébastien Granier<sup>1</sup>✉ & Cédric Leyrat<sup>1</sup>✉

The polyhistidine (6XHis) motif is one of the most ubiquitous protein purification tags. The 6XHis motif enables the binding of tagged proteins to various metals, which can be advantageously used for purification with immobilized metal affinity chromatography. Despite its popularity, protein structures encompassing metal-bound 6XHis are rare. Here, we obtained a 2.5 Å resolution crystal structure of a single chain Fv antibody (scFv) bearing a C-terminal sortase motif, 6XHis and TwinStrep tags (LPETGHHHHHHWSHPQFEK[G<sub>3</sub>S]<sub>3</sub>WSHPQFEK). The structure, obtained in the presence of cobalt, reveals a unique tetramerization motif (TetrHis) stabilized by 8 Co<sup>2+</sup> ions. The TetrHis motif contains four 6 residues-long β-strands, and each metal center coordinates 3 to 5 residues, including all 6XHis histidines. By combining dynamic light scattering, small angle x-ray scattering and molecular dynamics simulations, We investigated the influence of Co<sup>2+</sup> on the conformational dynamics of scFv 2A2, observing an open/close equilibrium of the monomer and the formation of cobalt-stabilized tetramers. By using a similar scFv design, we demonstrate the transferability of the tetramerization property. This novel metal-dependent tetramerization motif might be used as a fiducial marker for cryoelectron microscopy of scFv complexes, or even provide a starting point for designing metal-loaded biomaterials.

<sup>1</sup>IGF, University of Montpellier, CNRS, INSERM, Montpellier, France. <sup>2</sup>Centre de Biologie Structurale, University of Montpellier, CNRS, INSERM, Montpellier, France. <sup>3</sup>Present address: Sosei-Heptares, Steinmetz Building, Granta Park, Cambridge CB21 6DG, UK. ✉email: [sebastien.granier@igf.cnrs.fr](mailto:sebastien.granier@igf.cnrs.fr); [cedric.leyrat@igf.cnrs.fr](mailto:cedric.leyrat@igf.cnrs.fr)

**P**olyhistidine tags are peptides consisting of 6 or more consecutive histidine residues (6XHis) that are commonly used for protein purification by immobilized metal affinity chromatography (IMAC) due to their metal binding properties<sup>1</sup>. These polyhistidine sequences bind various metal ions such as  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Zn}^{2+}$ , and  $\text{Cu}^{2+}$  and are usually genetically fused to the N- or C-terminus of recombinant proteins. Histidine-rich motifs (including homopolymeric histidine tracts) also occur in diverse natural proteins and peptides<sup>2,3</sup>, such as various zinc transporters<sup>4</sup>, bacterial nickel metabolism proteins<sup>5,6</sup>, human transcription factors<sup>7</sup> or snake venoms<sup>8</sup>. A number of isolated peptide sequences have been studied for their metal-binding properties, ranging from a simple 6XHis<sup>9</sup> to more complex sequences such as the poly-His/poly-Gly peptide pHpG-1 (EDDH<sub>9</sub>GVG<sub>10</sub>) isolated from the venom of the African viper *Atheris squamigera*<sup>8,10,11</sup>, or the peptide H<sub>2</sub>ASHGH<sub>2</sub>NSH<sub>2</sub>PQH<sub>11</sub> corresponding to residues 33–57 of human Forkhead box protein G1<sup>7</sup>. These peptides form complexes with metal ions that usually display polymorphic binding states<sup>7–9,12</sup>, sometimes coupled with the stabilization of  $\alpha$ -helical structure<sup>8</sup>.

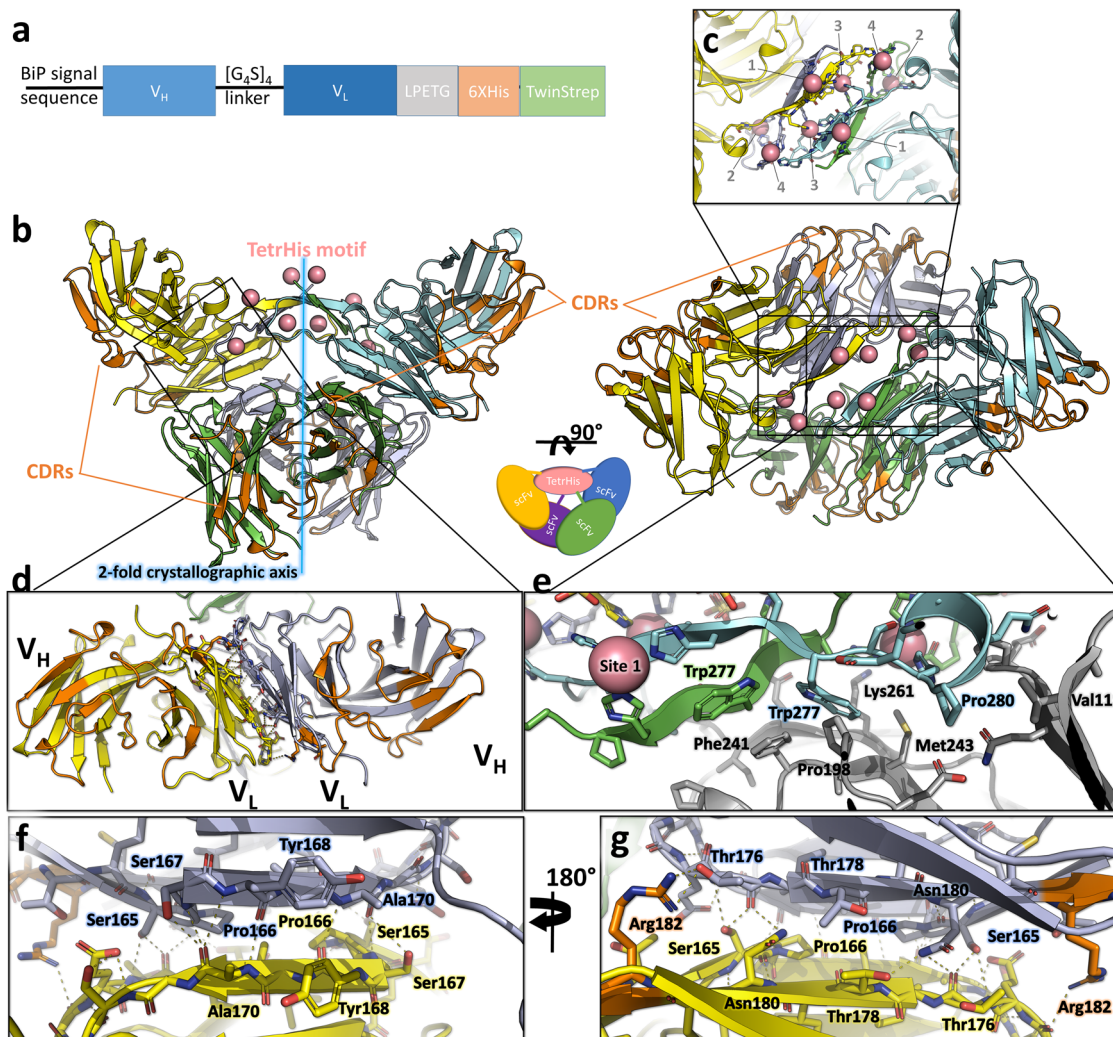
A sequence motif search of protein structures containing a 6XHis in the PDB yields ~42k results, indicating the sequence is highly prevalent in constructs used for structure determination. However, the 6XHis motif usually behaves as an intrinsically disordered region, adopting various conformations, and is thus extremely rare in electron density maps. For this reason, a structure motif search for 4 consecutive histidines yields only 250 structures. Restricting the search to structures containing any zinc, copper, nickel, cobalt, iron, or cadmium returns 81 entries, i.e. 0.002% of entries where the 6XHis was present.

In this study, we report the serendipitous discovery of TetrHis (acronym for Tetrameric His-Tag), a metal-dependent tetramerization motif with minimal sequence ETGHHHHHHWSHPQ observed in the cobalt-bound crystallographic structure of a single-chain variable fragment (scFv) antibody. ScFvs are composed of the variable regions of the antibody heavy and light chains ( $V_H$  and  $V_L$  domains) connected by a flexible peptide linker, and they represent the smallest engineered antibody fragment containing the parental specificity. ScFvs can display significant inter-domain flexibility between  $V_H$  and  $V_L$  domains<sup>13</sup> and have a tendency to form dimers and higher order oligomers depending on the length of the linker peptide<sup>14–16</sup>. However, the identification and characterization of a metal-dependent tetramerization motif utilizing a polyhistidine sequence in scFvs represents a novel and unexplored area of research. The sequence of the TetrHis motif corresponds to a 6XHis tag with a preceding ETG sequence from a sortase motif (LPETG) and a trailing sequence WSHPQ from a TwinStrep tag that were fused at the C-terminus of the expressed scFv construct (Fig. 1a). The scFv used in this work, 2A2, was designed by grafting the complementarity determining regions (CDRs) of the 2A2 antibody described to bind the lipid ceramide<sup>17</sup> onto a murine scFv scaffold with a [GGGG]<sub>4</sub> linker (taken from PDB entry 5LX9 chain B<sup>18</sup>), with the initial aim of studying its binding to ceramide. Below we describe the crystal structure of scFv 2A2 in its cobalt-stabilized tetrameric state that revealed the TetrHis motif. We next characterize the dimensions, structure, and dynamics of the tagged scFv molecule in the absence or presence of  $\text{Co}^{2+}$  ions using a combination of dynamic light scattering (DLS), small angle x-ray scattering (SAXS), molecular dynamics simulations (MDS), and ensemble analysis, revealing its conformational equilibria in atomistic details. Finally, we use SAXS to show that the tetramerization property can be transferred to other scFvs by characterizing a new construct harboring the same framework sequence and C-terminal tags, but in which the CDRs were substituted by those from a previously described anti-FLAG M2 scFv<sup>19</sup>.

## Results

**The X-ray structure of scFv 2A2 reveals a cobalt-stabilized tetrameric assembly.** ScFv 2A2 crystallized in space group I121 in mother liquor containing 10 mM  $\text{CoCl}_2$  and the structure was solved by molecular replacement at 2.5 Å resolution using a homology model of the scFv  $V_H$  and  $V_L$  domains as the search model (Table 1). The asymmetric unit is composed of 2 scFv molecules that form a tetramer with a twofold crystallographic axis, i.e. a dimer of dimers (Fig. 1b, Supplementary Data 1 and Supplementary Fig. 1). Each scFv monomer adopts the classical arrangement of the  $V_H$  and  $V_L$  domains, and further dimerizes through a  $\sim 650 \text{ \AA}^2$   $V_L$ – $V_L$  interface that is stabilized by 15 hydrogen bonds (Fig. 1d, f, g). This  $V_L$ – $V_L$  interface has previously been observed in the crystal packing of a number of scFv structures<sup>18,20</sup> and isolated  $V_L$  domains<sup>21</sup>, as can be seen from the structural alignment shown in Supplementary Fig. 1. A recent analysis of antibody fragment antigen-binding (Fab) X-ray structures in the PDB found that this  $V_L$ – $V_L$  interface was the 5th most common type of Fab–Fab interface with a prevalence of 3% (defined as VL-11 in ref. 22). The most striking and unexpected feature of the structure is the presence of electron density for the C-terminal sortase recognition motif (<sup>266</sup>LPETG<sup>270</sup>), the 6XHis and the first Strep tag (<sup>277</sup>WSHPQFEK<sup>284</sup> or only <sup>277</sup>WSHPQ<sup>281</sup> in one of the two chains) of the Twin-strep tag (WSHPQFEK(G<sub>3</sub>S)<sub>3</sub>WSHPQFEK). These 16–19 residues form an extension that adopts a tetrameric arrangement stabilized by 8  $\text{Co}^{2+}$  ions (Figs. 1c and 2). The motif, named TetrHis (for Tetrameric His-Tag), packs against the scFv surface making hydrophobic contacts mainly involving framework residues Pro238, Phe241, Met243, Lys261, Glu263, and Leu266 and Trp277 from the motif (Fig. 1e). The tetrameric assembly contains two 6 residue-long  $\beta$ -strands related by the twofold symmetry axis, and the 8  $\text{Co}^{2+}$  ions coordinate 3–5 protein residues each, resulting in 4 unique metal binding sites (Fig. 2a). In site 1 and 2,  $\text{Co}^{2+}$  ions are coordinated by residues belonging exclusively to the two 1st protein chains, thereby stabilizing the dimeric assembly. In contrast, the  $\text{Co}^{2+}$  coordination at site 3 and 4 connects protein chains across the twofold crystallographic axis and thus stabilizes the tetramer (Fig. 2b). The  $\text{Co}^{2+}$  ion of site 1 is coordinated by His279 (Strep tag) of the 1st chain and His273 and His275 (6XHis) of the 2<sup>nd</sup> chain (Fig. 2c). The coordination sphere is completed by 2 water molecules resulting in a distorted trigonal bipyramidal geometry. Site 2 adopts a distorted octahedral geometry with  $\text{Co}^{2+}$  coordination by Glu268 (sortase), His273 and His275 (6XHis) of the 1st chain, His279 (Strep tag) of the 2nd chain, and a water molecule (Fig. 2d). Site 3  $\text{Co}^{2+}$  ion also exhibits an octahedral geometry and is coordinated by 5 protein residues (Fig. 2e): His276 from both the 1st and 2nd chains (6XHis), Glu268 (sortase), His272 and His274 (6XHis) from the 4th chain (symmetry equivalent of the 2nd chain). Finally, site 4 involves His272 and His274 from the 1st chain and His271 from the 4th chain (Fig. 2f). No additional water density is visible, resulting in a trigonal pyramidal geometry.

Overall, the sidechains of Glu268 from the sortase recognition motif, His279 from the strep tag, and all the histidines from the 6XHis (His271–276) are directly involved in binding to  $\text{Co}^{2+}$  ions, with the exception of His271 from the 1st chain (Fig. 2b). Several important differences exist between the two non crystallographically related chains at the level of the tetramerization motif: (1) the  $\beta$ -strand from the 1st chain covers the <sup>275</sup>HHWSHP<sup>280</sup> sequence and shares backbone-backbone hydrogen bonds with the 2nd chain  $\beta$ -strand which has the sequence <sup>272</sup>HHHHHW<sup>277</sup> (Fig. 2); (2) while there is no electron density for residues after Gln281 of the 1st chain, in the 2nd chain residues <sup>278</sup>SHPQFEK<sup>284</sup> interact with both the tetramerization



**Fig. 1** Crystal structure of the cobalt-bound scFv 2A2 tetramer. **a** Protein construct used for crystallization of scFv 2A2. The expressed sequence contains a (cleaved) N-terminal BiP signal sequence and a C-terminal tag composed of a sortase motif (LPETG), a 6XHIS and a twinStrep tag (WSHPQFEK[G<sub>3</sub>S]<sub>3</sub>WSHPQFEK). **b** Overall architecture of the tetramer in 2 views rotated by 90°. The protein chains are shown in cartoon representation and colored in green, cyan, yellow, and purple. The complementarity determining regions (CDRs) are colored in orange. The 8 cobalt ions are shown as salmon-colored spheres. **c** Close up view of the cobalt-bound TetrHis motif showing the 4 types of metal binding sites related by the twofold symmetry axis. **d** Close up view of the dimeric V<sub>L</sub>–V<sub>L</sub> interface (defined as VL-11 in ref. 22). **e** Hydrophobic interactions stabilizing the interface between the TetrHis motif and the scFv V<sub>H</sub> and V<sub>L</sub> domains. Framework residues belonging to the V<sub>L</sub> domain are shown in gray. **f, g** Details of the V<sub>L</sub>–V<sub>L</sub> interface interactions, in 2 views rotated by 180°.

motif and residues from the scFv framework, forming an  $\alpha$ -helical turn (Figs. 1c, e and 2a). From both chains, Trp277 play an important stabilizing role in the complex, creating a hydrophobic core between the tetramerization motif and the V<sub>L</sub> domain of the 2nd chain, however Trp277 from the 1st chain remains largely solvent-exposed while Trp277 from the 2nd chain is buried between the scFv and the TetrHis motif (Fig. 1e).

#### Analysis of metal-bound homopolymeric histidine tracts in the PDB shows that the TetrHis motif has no known homologs.

In order to determine whether similar tetramerization motifs involving metal-interacting polyhistidine tags existed in the PDB, we performed a survey of available experimental structures using the PDB structure motif search tool. We used four consecutive histidines from the 6XHIS of entry 2JSN<sup>23</sup> as a search motif and an arbitrarily high RMSD cut-off in order to avoid excluding structures in which histidines adopt a different conformation. Out of the 81 entries that contained at least four consecutive visible

histidines and any bound zinc, copper, nickel, cobalt, cadmium, or iron, we identified 63 structures in which at least one metal ion interacts directly with the 6XHIS motif. We curated this initial dataset by removing redundancy, grouping highly similar structures together and keeping only a single representative, which yielded 27 unique X-ray crystallographic structures (Supplementary Table 1). We further omitted structures in which less than two histidines from the 6XHIS were involved in metal coordination. Basic statistics regarding the nature, number of bound metal and their interactions with the protein are summarized in Fig. 3. Most of these structures contain bound zinc (12) or nickel (9), and the number of metal ions is generally four or less (23 out of 27 structures). These metals usually stabilize the crystal packing by connecting two protein chains (20 out of 27 structures). We found only five occurrences of linear metal-binding motifs (i.e. short amino acid sequences typically ranging from 3 to 10 residues in length) out of the 27 structures, all of which involve one or two bound nickel ions, stabilizing either a dimeric or trimeric assembly (Fig. 3e–i). In all five cases, it is

**Table 1 Data collection and refinement statistics (molecular replacement).**

	scFv 2A2
Data collection	
Space group	I 1 2 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	97.50, 81.82, 111.04
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 111.14, 90.00
Resolution (Å)	45.47–2.50 (2.60–2.50)
<i>R</i> <sub>merge</sub>	0.213 (1.293) <sup>a</sup>
CC <sub>1/2</sub>	0.989 (0.491)
<i>I</i> / $\sigma$ <i>I</i>	5.6 (1.1)
Completeness (%)	99.2 (96.8)
Redundancy	6.9 (6.2)
Refinement	
Resolution (Å)	45.47–2.50
No. reflections	28,159 (1387) <sup>b</sup>
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	19.02/24.83
No. atoms	
Protein	3847
Cobalt	4
Water	196
<i>B</i> -factors	
Protein	47.8
Cobalt	60.4
Water	46.0
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.318
Ramachandran statistics	
Favored	97.31 %
Outliers	0.41 %

<sup>a</sup>Values in parentheses are for highest-resolution shell.  
<sup>b</sup>Values in parentheses are for the test set.

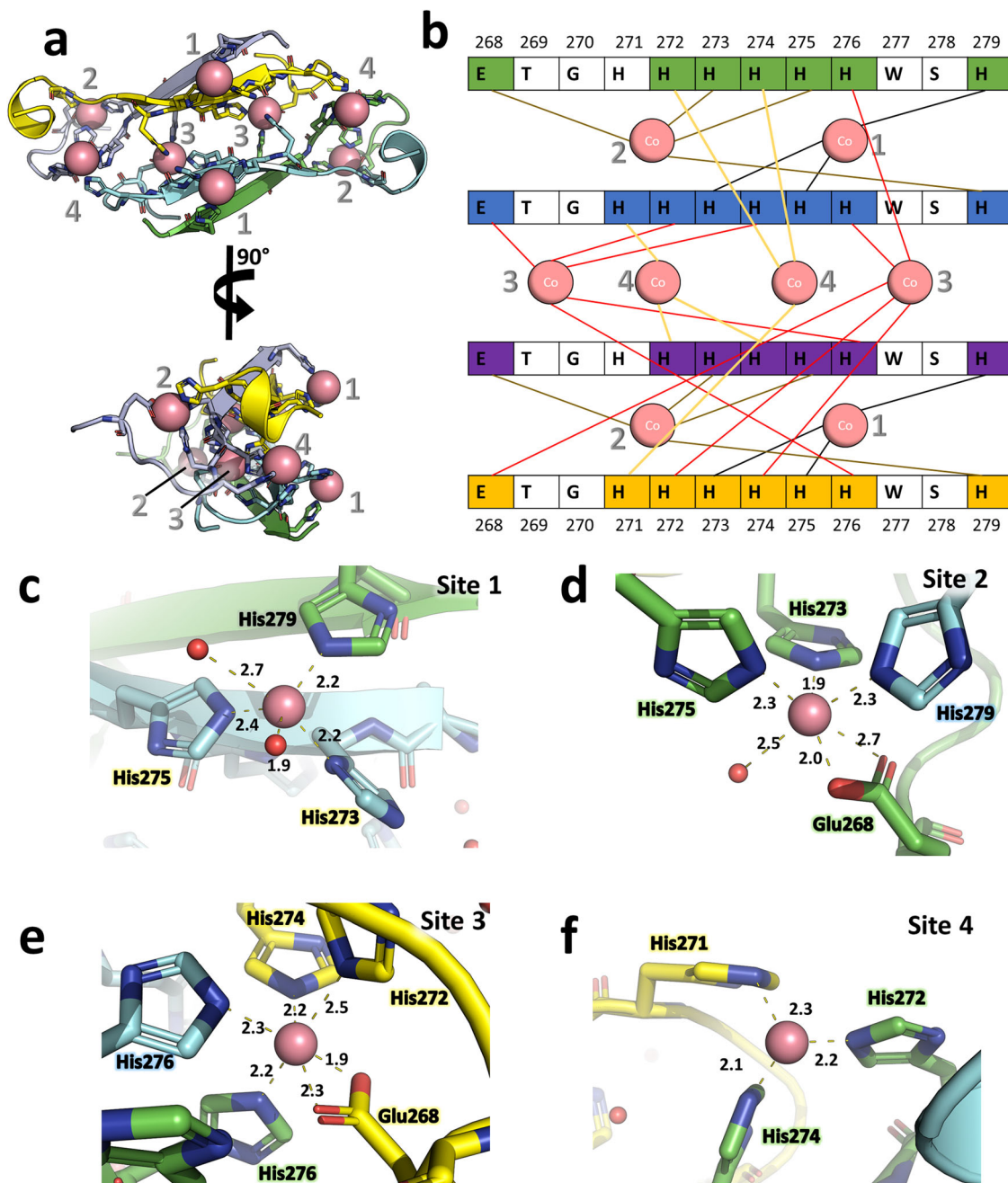
unknown whether the crystallographically observed assembly can also form in solution in the presence of Ni<sup>2+</sup> ions, however entries 3CGM and 4ODP are crystal structures of SlyD from *Thermus thermophilus*, a protein well known for its metallo-chaperone activity<sup>24–26</sup>. Taken together, these analyses indicate that no linear metal-dependent polyhistidine tetramerization motif exists in the PDB, making TetrHis a novel sequence motif.

**DLS analyses show that the size of scFv 2A2 increases as a function of cobalt concentration and that the changes are mediated by the TetrHis motif.** In order to assess the ability of cobalt to induce changes in the conformation or oligomeric state of scFv 2A2 in solution, we performed a cobalt titration by DLS (Fig. 4). In addition, we used an anti-ADIPOR scFv bearing a C-terminal twin strep tag<sup>18</sup> as a negative control to determine whether the observed changes are indeed related to the presence of the TetrHis motif. We found that addition of 500  $\mu$ M to 5 mM of cobalt ions led to a significant shift in the size distribution and calculated hydrodynamic radius (*R*<sub>h</sub>) of scFv 2A2 (Fig. 4 and Supplementary Fig. 2). The estimated *R*<sub>h</sub> increased from ~3.3 nm in the absence of cobalt to ~4.8 nm at 10 mM Co<sup>2+</sup>, indicating the formation of larger species. These values compare well with the ones calculated from structural models of monomers (3.2 nm) and tetramers (4.4 nm) extracted from the x-ray structure, after addition of missing residues (interdomain linker and second strep tag). On the contrary, cobalt seemed to induce a compaction of the control scFv from a *R*<sub>h</sub> value of 5.6 nm down to 3.4 nm, with a concomitant decrease in the width of the size distribution (Supplementary Fig. 2). Although the initial compaction from 5.5 to

4.0 nm in 200  $\mu$ M CoCl<sub>2</sub> was unexpected, all the data measured in the presence of cobalt for this scFv are consistent with a monomeric state. These data demonstrate that the TetrHis motif can mediate cobalt-dependent changes in the conformation and/or oligomeric state of scFv 2A2 in solution, which are compatible with tetramer formation.

**SAXS analysis indicates metal-induced changes in the structure and oligomeric state of scFv 2A2.** Intrigued by the peculiar tetrameric architecture of scFv 2A2 observed in the crystal structure, we turned to SAXS in order to assess the structure of the scFv directly in solution. SAXS profiles were measured at three or five different protein concentrations in gel filtration buffer or in the presence of additives: 5 mM EDTA, 5 mM NiSO<sub>4</sub>, or 5 mM CoCl<sub>2</sub> (Fig. 5a and Table 2). We also attempted to measure data in the presence of 5 mM ZnCl<sub>2</sub>, but these showed severe signs of aggregation. The measured radius of gyration (*R*<sub>g</sub>) in the presence of EDTA increased from 24 Å at 2 mg/ml of protein to 29 Å at 8 mg/ml, while the *R*<sub>g</sub> values in regular buffer were comprised between 25 and 32 Å, at 0.5 and 8 mg/ml of protein, respectively. This concentration dependence of the *R*<sub>g</sub> suggested possible interparticle attraction, however the measured values were roughly consistent with the *R*<sub>g</sub> calculated from monomers extracted from the crystal structure, particularly at low protein concentration (theoretical *R*<sub>g</sub>  $\approx$  20–28 Å dependent on whether the missing residues have been added or not, and their respective conformations). In the presence of Co<sup>2+</sup> or Ni<sup>2+</sup> ions, the measured SAXS profiles showed a noticeable change of slope between  $Q \approx 0.1$ – $0.2 \text{ \AA}^{-1}$ , particularly visible at high protein concentration (Fig. 5a), suggesting significant hollowness of the scattering object. This feature was consistent with the presence of a large void within the tetrameric scFv structure (Fig. 1b), with the presence of a channel of 10–20 Å width in between the TetrHis motif and scFv domains. Similar to the data measured in the absence of added metal ions, the measured *R*<sub>g</sub> showed important protein concentration-dependent variations with values of 36–60 Å in the presence of Ni<sup>2+</sup>, and 34–50 Å in the presence of Co<sup>2+</sup> (Table 2). The 34 to 36 Å *R*<sub>g</sub> values obtained at low protein concentration were also roughly consistent with the *R*<sub>g</sub> calculated from the crystallographic tetramer (32–36 Å after addition of the scFv linker and missing residues at the C-terminus). Comparison of the Kratky plots from SAXS data measured at protein concentrations of 2 mg/ml in the presence of EDTA versus Co<sup>2+</sup> ions suggests that the metal ions induce a transition from a mostly globular to a multidomain (or oligomeric) protein (Fig. 5b). The pair-distance distribution functions *p*(*r*) calculated from the SAXS profiles measured in the presence or absence of 5 mM Co<sup>2+</sup> at different protein concentrations are shown in Fig. 5c. The addition of cobalt leads to a pronounced shift of the *p*(*r*) towards larger interatomic distances, which further increases with protein concentration. At 1–2 mg/ml of protein, the presence of 2 overlapping peaks in the distribution is clearly visible, while higher protein concentrations seem to induce the formation of even larger species.

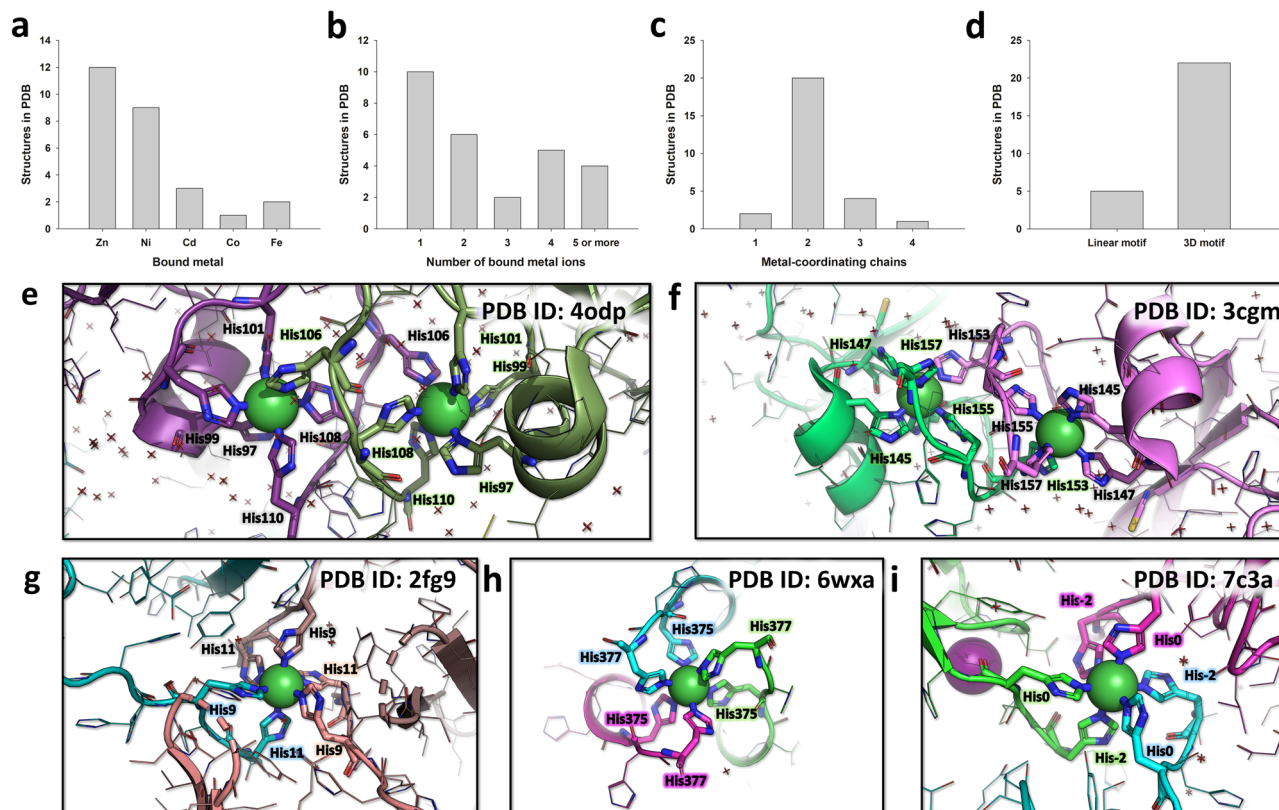
Because the strong concentration dependence of the measured *R*<sub>g</sub> values made analysis of solution data less straightforward, we next measured SEC-SAXS data with or without addition of Co<sup>2+</sup> ions in the SEC buffer. We hypothesized that the size separation of the SEC would remove potential interparticle interference or signal contributions from small amounts of higher order oligomers or aggregates, and thus provide a useful comparison with the solution-based data. ScFv 2A2 eluted from the SEC as a single, symmetric peak as can be seen from the average intensity and UV<sub>280</sub> profiles (Fig. 5c and Supplementary Fig. 3). The measured *R*<sub>g</sub> was mostly constant across the peak with a slight



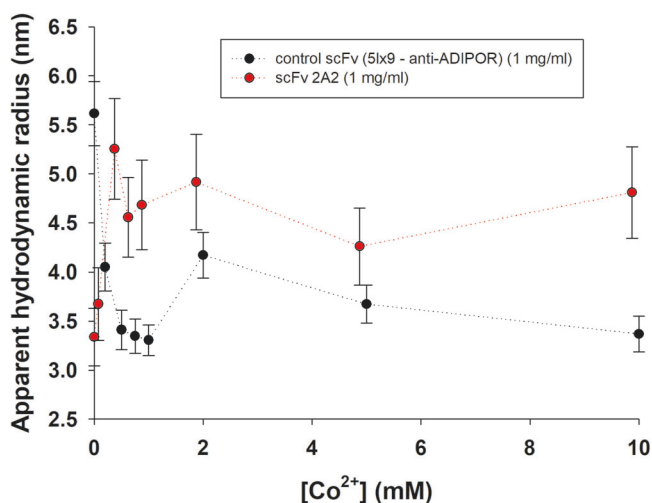
**Fig. 2 Architecture of the cobalt-bound tetrameric polyhistidine (TetrHis) motif.** **a** Close up of the TetrHis motif, in two views rotated by 90°. The protein chains are shown in cartoon representation and the color code is the same as in Fig. 1. **b** Schematic representation of the metal-protein contacts within the tetrameric motif. **c–f** Close ups of the 4 cobalt binding sites showing the coordination geometries and distances.

downward trend in the second half of the peak and an average value of 26.5 Å, in agreement with the  $R_g$  value expected for a monomeric species (Table 2 and Fig. 5d). The estimated molecular weight (MW) was also mostly constant ranging from 28 to 33 kDa (Supplementary Fig. 3b) in agreement with the theoretical MW of 30.5 kDa. In contrast, the profile measured in the presence of 5 mM  $\text{CoCl}_2$  displayed a wider, asymmetric peak that was slightly shifted towards lower elution volumes (Fig. 5d). The measured  $R_g$  showed important variations across the peak with values ranging from ~ 25 to 34 Å (Fig. 5d) and the estimated MW was comprised between 30 and 60 kDa (Supplementary Fig. 3b). All these observations indicate cobalt-induced heterogeneity in the scFv oligomeric state.

**Molecular dynamics simulations suggest significant scFv inter-domain flexibility.** In order to analyze the conformational landscape of scFv 2A2 using the measured SAXS data, we generated the ensembles of conformers required for SAXS-based ensemble optimization using classical explicit solvent MD simulations. We ran two independent trajectories of the monomer (Fig. 6a and Supplementary Fig. 4a). In the first simulation, the 39 residue-long C-terminal extension composed of the LPETG, 6XHis, and TwinStrep tag quickly packed against the surface of the scFv and remained stable for the remainder of the simulation. In the second simulation, the C-terminal extension remained highly flexible and we observed complete dissociation of the  $V_H$  and  $V_L$  domains in the 2nd half of the trajectory (Fig. 6a). This spontaneous transition



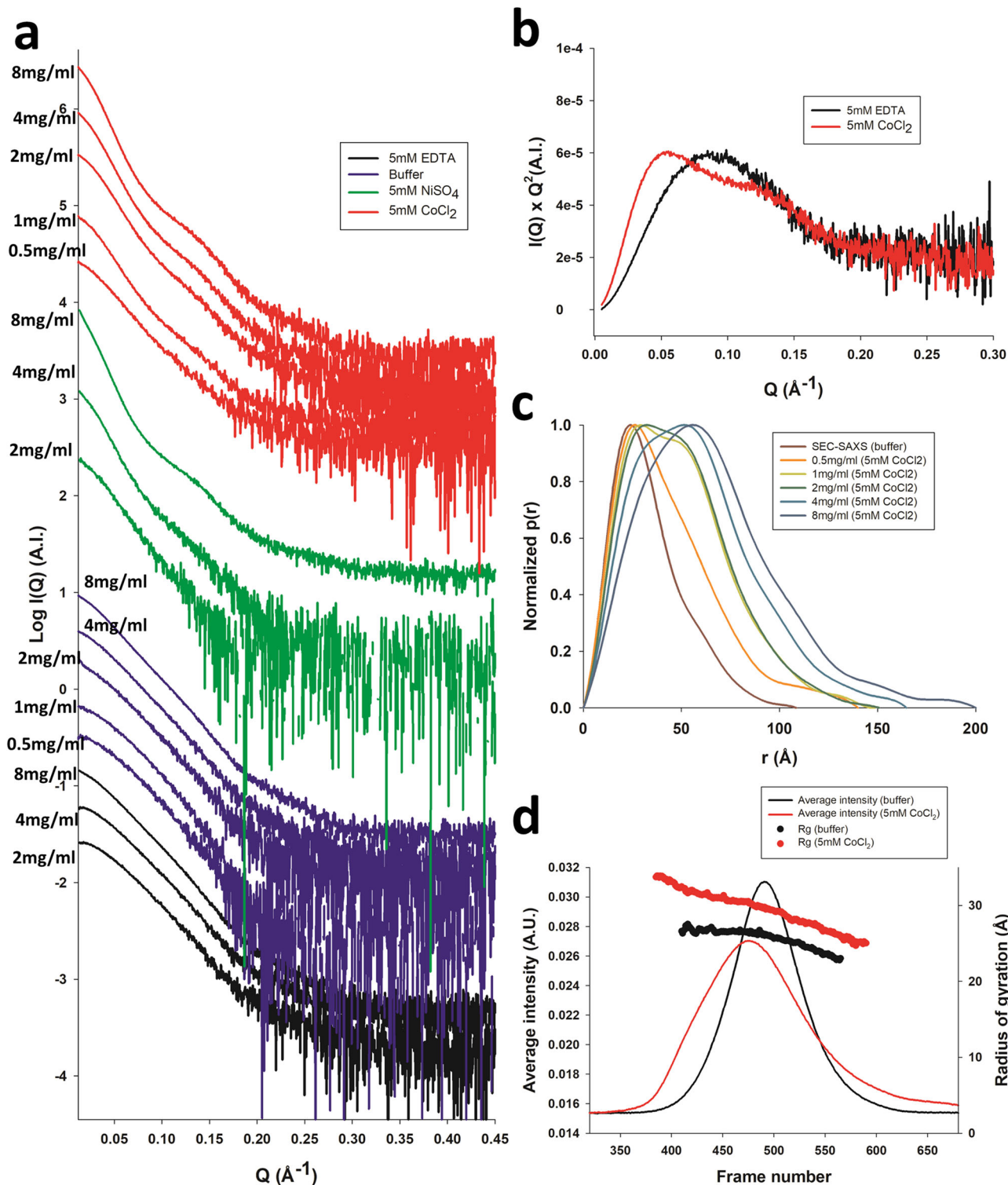
**Fig. 3** Survey of metal-bound polyhistidine tags in the PDB. **a–d** Basic statistics extracted from the 27 unique PDB structures containing at least 4 consecutive histidines and a bound metal. The structures are classified according to the nature of the bound metal (**a**), the number of bound metal ions in the assembly (**b**), the number of protein chains that contribute coordinating residue sidechains (**c**), and whether the assembly involves a short linear sequence motif or a three-dimensional motif (with residues located far away within the protein sequence) (**d**). **e, f** Nickel-dependent dimeric assemblies of SlyD from *Thermus thermophilus*, observed in PDB entries 4odp and 3cgm. 3 His residues located at the C-terminus of the construct are involved in metal coordination, along with 2 His from a 6XHis tag located 3 residues downstream, and 1 His from the 6XHis tag of a symmetric-related molecule. **g–i** Nickel-dependent trimeric assemblies observed in PDB entries 2fg9, 6wxa, and 7c3a. In each case,  $3 \times 2$  His residues from 6XHis tags assemble to form an octahedral coordination sphere around a single nickel ion.



**Fig. 4** Cobalt titrations of scFv 2A2 and anti-ADIPOR scFv by dynamic light scattering (DLS) at 20 °C in 50 mM HEPES pH 7.5, 150 mM NaCl using a protein concentration of 1 mg/ml. Each point corresponds to the average of 3 independent measurements, with standard error shown as a bar. The apparent hydrodynamic radii were calculated from the intensity autocorrelations using the Stokes–Einstein equation.

to an open conformation suggested limited stability of the scFv  $V_H-V_L$  interface, which is a relatively frequent property in scFvs that have not been engineered for stability<sup>27,28</sup>. We also ran multiple independent MD trajectories of the tetrameric scFv complex (Fig. 6b and Supplementary Fig. 4b) in which the cobalt ions were replaced by zinc ions (as parameters for  $\text{Co}^{2+}$  are not available in standard MD force fields). In all trajectories but one, we applied distance restraints to the metal coordination centers in order to maintain the integrity of the crystallographically observed tetramerization motif, resulting in stable tetramers. In the absence of restraints, we observed that zinc ions tended to dissociate over time, leading to partially dissociated tetrameric states at longer timescales (Fig. 6b). We then extracted all monomeric and tetrameric scFv conformers observed in the MD trajectories to create a large ensemble of models for SAXS-based ensemble optimization.

**Optimization of MDS-derived ensembles against SAXS data reveals the conformational landscape of scFv 2A2 in the absence and presence of cobalt(II) ions.** The ensemble optimization method (EOM) uses a genetic algorithm to optimize a small (usually 5–20) ensemble of models that fits the experimental SAXS data from a larger pool of structural models (>1000)<sup>29</sup>. This method was used to fit SAXS curves extracted from both the solution-based and SEC-SAXS experiments using the pool of MD models of scFv 2A2 (Fig. 7). The data measured in the absence of cobalt was generally well fitted using the monomeric ensemble of models with

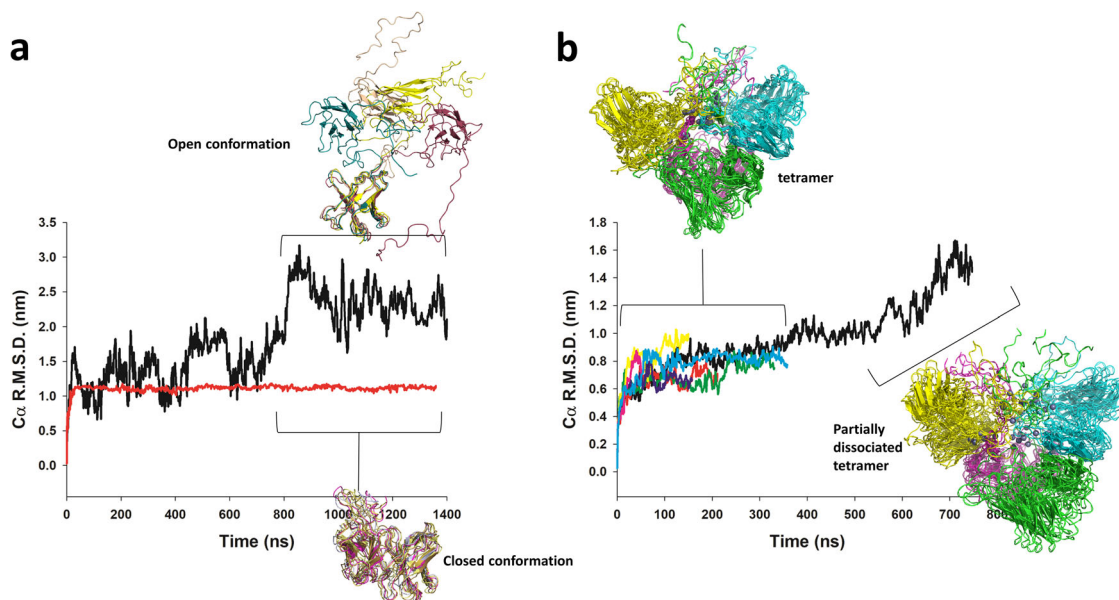


**Fig. 5 SAXS characterization of scFv 2A2.** **a** SAXS profiles of scFv 2A2 in the presence of 5 mM EDTA (black curves), in regular buffer (50 mM HEPES pH 7.5 and 150 mM NaCl, blue curves), or in the presence of 5 mM NiSO<sub>4</sub> or 5 mM CoCl<sub>2</sub> (green and red curves, respectively). The data were measured at increasing protein concentrations of 2.1, 4.2, and 8.5 mg/ml for EDTA and NiSO<sub>4</sub>, while additional curves were measured at 0.5 and 1 mg/ml in regular buffer and CoCl<sub>2</sub> conditions. **b** Kratky plots the EDTA and CoCl<sub>2</sub> data using the 2.1 mg/ml SAXS profiles. **c** Pair-distance distribution functions  $p(r)$  calculated from the SAXS profiles measured in the presence or absence of 5 mM Co<sup>2+</sup> at different protein concentrations. The SAXS curves extracted from the SEC-SAXS profile is used as the cobalt-free reference. **d** SEC-SAXS profiles of scFv 2A2 in the presence (red) or absence of 5 mM CoCl<sub>2</sub> (black curve), and analysis of Radius of gyration versus frame number.



**Table 2 SAXS-derived parameters.**

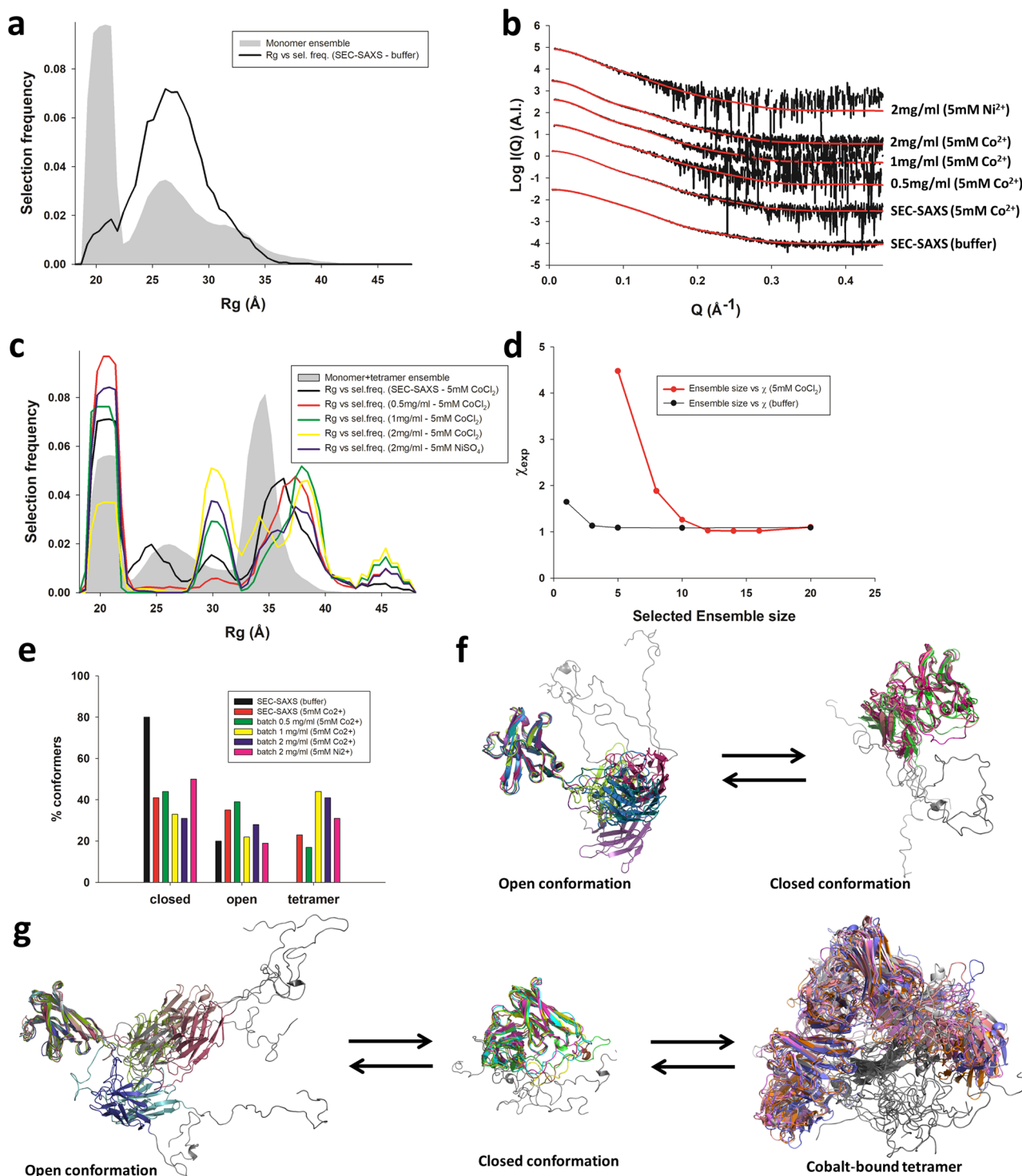
	Buffer conditions	c (mg/ml)	MW (kDa)	R <sub>g</sub> (Å)	χ <sub>EOM</sub>
ScFv 2A2 SAXS	Buffer A <sup>a</sup>	8.5	47.6	32.3 ± 0.2	1.604
	Buffer A	4.2	40.6	30.0 ± 0.2	1.391
	Buffer A	2.1	35.8	27.6 ± 0.3	1.079
	Buffer A	1.0	35.6	25.5 ± 0.2	0.943
	Buffer A	0.5	29.9	24.7 ± 0.3	0.901
	Buffer A + 5 mM EDTA	8.5	39.7	29.0 ± 0.3	2.185
	Buffer A + 5 mM EDTA	4.2	29.4	25.6 ± 0.2	1.110
	Buffer A + 5 mM EDTA	2.1	26.1	24.1 ± 0.3	1.039
	Buffer A + 5 mM CoCl <sub>2</sub>	8.5	172.1	49.9 ± 0.2	10.279
	Buffer A + 5 mM CoCl <sub>2</sub>	4.2	122.8	44.2 ± 0.2	4.770
	Buffer A + 5 mM CoCl <sub>2</sub>	2.1	72.3	38.1 ± 0.2	1.695
	Buffer A + 5 mM CoCl <sub>2</sub>	1.0	69.9	38.4 ± 0.2	1.274
	Buffer A + 5 mM CoCl <sub>2</sub>	0.5	58.1	33.9 ± 0.2	0.927
	Buffer A + 5 mM NiSO <sub>4</sub>	8.5	248.5	60.2 ± 0.5	N.D.
	Buffer A + 5 mM NiSO <sub>4</sub>	4.2	126.4	45.2 ± 0.7	2.233
Buffer A + 5 mM NiSO <sub>4</sub>	2.1	75.4	36.3 ± 0.8	1.034	
SEC-SAXS					
	Frame 464-524	Buffer A	0.85 <sup>b</sup>	34.7	26.5 ± 0.1
Frame 390-455	Buffer A + 5 mM CoCl <sub>2</sub>	0.85 <sup>b</sup>	55.4	31.7 ± 0.1	1.031
ScFv748 SAXS					
	Buffer A	0.5	32.6	25.1 ± 0.3	0.916
	Buffer A + 5 mM CoCl <sub>2</sub>	0.5	47.1	34.9 ± 0.4	0.946

<sup>a</sup>Buffer A: 50 mM HEPES pH 7.5 150 mM NaCl.<sup>b</sup>Concentration of the sample that was injected on the SEC assuming a 10-fold dilution factor.

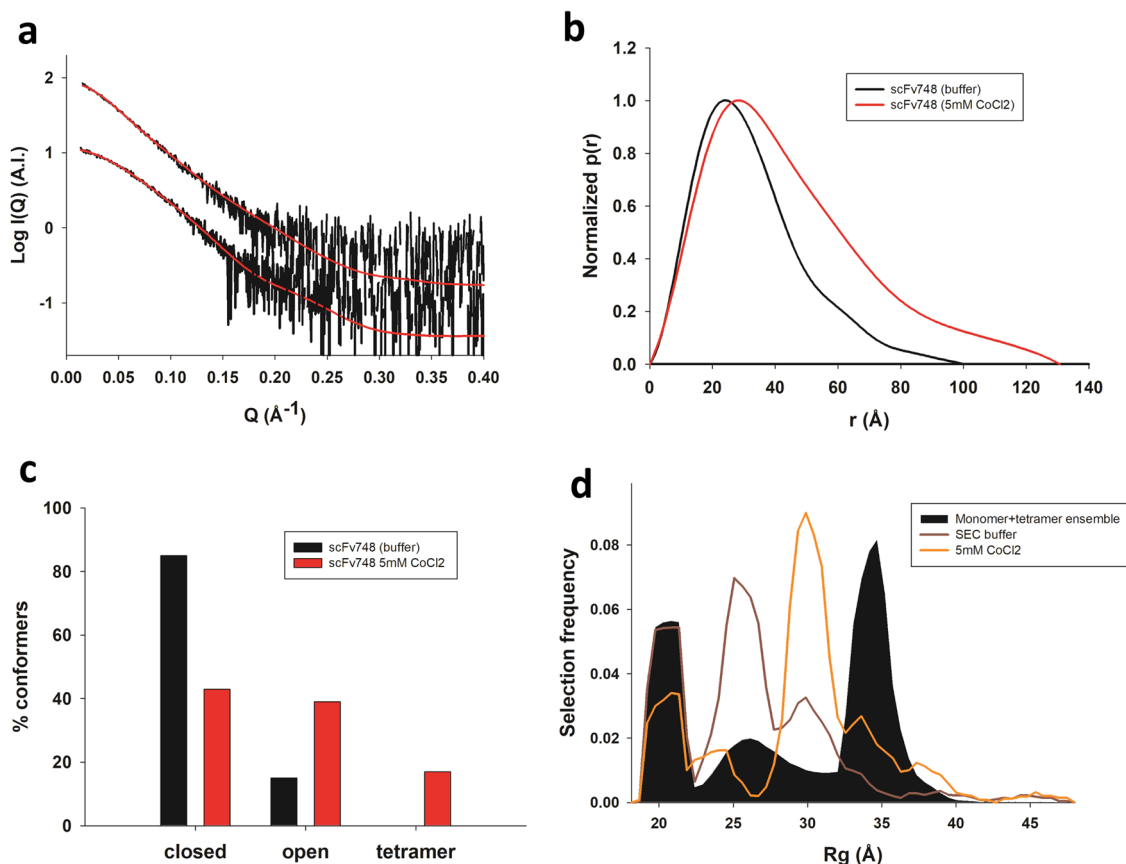
**Fig. 6 Molecular dynamics simulations analysis of scFv 2A2 in its monomeric and tetrameric forms.** **a** C $\alpha$  root mean square deviation (R.M.S.D.) as a function of simulation time for two independent MD trajectories of the scFv monomer. Five representative conformers of the open and closed states sampled during the simulations are superimposed and shown as cartoon. **b** C $\alpha$  R.M.S.D. as a function of simulation time for the MD trajectories of the scFv tetramer. Representative conformers along the trajectories are shown in cartoon representation and colored by chain. The longer trajectory (black curve) was run without distance restraints on metal binding sites, which resulted in the sampling of partially dissociated tetrameric states.

χ<sub>EOM</sub> values of 1.0–1.6, except for a slight worsening of the quality of fit at high protein concentration in the presence of EDTA (χ<sub>EOM</sub> = 2.185; Table 2). The R<sub>g</sub> distribution obtained from analysis of the SEC-SAXS data is shown in Fig. 7a. The distribution shows a single peak centered on the experimental R<sub>g</sub> value of 26.5 Å for models of the selected ensemble. The ensemble is composed of 80–85% of closed conformers and 15–20% of open conformers (Fig. 7e), and the C-terminal extension behaves as an intrinsically disordered tail adopting mostly extended conformations (Fig. 7f).

For the SEC-SAXS data measured in the presence of cobalt ions, a SAXS curve was extracted from the first part of the asymmetric elution peak (frames 390–455, see Fig. 5d). The data could not be adequately fitted using the monomeric scFv ensemble (χ<sub>EOM</sub> = 4.64) and required a pool ensemble containing both monomeric and tetrameric forms (Fig. 6b) to reach a satisfactory χ<sub>EOM</sub> value of 1.03 (Fig. 7b). We also systematically fitted all solution-based SAXS profiles against this monomer-tetramer ensemble and found that good agreement with the



**Fig. 7 Ensemble optimization (EOM) analysis of the SAXS curves extracted from the SEC-SAXS profiles of scFv 2A2.** **a** Radius of gyration distributions for the MD-generated ensemble of monomeric models (gray area) and for the optimized ensemble (OE) that fits the SEC-SAXS data in the absence of cobalt (black curve). **b** Fitted SAXS profiles of scFv 2A2 measured in the presence or absence of 5 mM NiSO<sub>4</sub> or CoCl<sub>2</sub> in solution or in SEC-SAXS. The protein concentration for each curve is indicated on the figure panel. Experimental scattering curves are shown as black lines with OEs fits shown as red lines. **c** Radius of gyration distributions for the MD-generated ensemble of monomeric and tetrameric models (black area) and for the OEs that fits the solution-based or SEC-SAXS data (curves) in the presence of 5 mM CoCl<sub>2</sub> or 5 mM NiSO<sub>4</sub>. **d** EOM Goodness-of-fit  $\chi_{exp}$  as a function of the number of models in the OEs, in the presence (red) or absence (black) of 5 mM CoCl<sub>2</sub>. **e** Histograms showing the percentage of tetramers, open and closed state monomers observed in the best fitting OEs for different experimental conditions. **f** Conformational equilibrium of monomeric scFv 2A2 in 50 mM HEPES pH 7.5 and 150 mM NaCl. For each state, 5 representative models extracted from the OEs are superimposed and shown as cartoon. The disordered C-terminal tail corresponding to the sortase-His6-Twinstrep sequence is colored in gray. **g** Conformational equilibrium of scFv 2A2 in the presence of 5 mM CoCl<sub>2</sub>. For each state, 5 representative models extracted from the OEs are superimposed and shown as cartoon. The sortase-His6-Twinstrep tail encompassing the TetrHis motif is colored in gray.



**Fig. 8 SAXS characterization and ensemble optimization of scFv 748.** **a** Fitted SAXS profiles of scFv 748 measured in the presence or absence of 5 mM  $\text{CoCl}_2$  in solution at 0.5 mg/ml protein concentration. **b** Pair-distance distribution functions  $p(r)$  calculated from the SAXS profiles measured in the presence or absence of 5 mM  $\text{CoCl}_2$ . **c** Proportion of open and closed states monomers and tetramers in the OEs in the presence or absence of 5 mM  $\text{CoCl}_2$ . **d** Radius of gyration distributions for the MD-generated ensemble of monomeric and tetrameric models (black area) and for the OEs that fit the solution SAXS data (brown and orange curves) in the presence or absence of 5 mM  $\text{CoCl}_2$ .

experimental data could be obtained for protein concentrations of up to 2 mg/ml (Fig. 7b and Table 2). The  $R_g$  distributions reveal a complex landscape induced by the cobalt or nickel ions (Fig. 7c). The increased complexity of the scFv conformational landscape is also apparent in the larger number of models required to fit the SAXS data (~ 15 versus 5 - as illustrated for the SEC-SAXS measurements in Fig. 7d). Analysis of the selected ensembles indicate a dynamic equilibrium between tetramers, closed state and open state monomers (Fig. 7e, g). In 5 mM  $\text{Co}^{2+}$ , the population of tetramers increases from roughly 20–40% when protein concentration goes from 0.5 to 1 or 2 mg/ml (Fig. 7e), whereas about 25% of tetramers are present in the SEC-SAXS data, which is consistent with the ~ 10-fold dilution that occurs on the SEC. At 2 mg/ml of protein in the presence of 5 mM  $\text{Ni}^{2+}$ , we also observed a comparable proportion of tetrameric states of ~ 30%, suggesting that both metals are able to induce tetramer formation in a similar way. Interestingly, the selected closed state monomers are much more compact than those selected in the absence of cobalt ions (Fig. 7a vs Fig. 7c, and Fig. 7f vs Fig. 7g), suggesting cobalt-induced structural preorganization of the C-terminal extension in the monomeric state. In addition, the overall ratio of open to close monomeric states is higher in the presence of cobalt, which is consistent with a depletion of the close state conformers through tetramerization. Taken together, these results indicate that the presence of cobalt ions remodels the conformational landscape of scFv 2A2 by inducing tetramerization and affecting its intrinsic open-close state equilibrium in the monomeric form.

**SAXS characterization of scFv 748 demonstrates transferability of the TetrHis motif to other scFvs.** Based on the crystal structure of the cobalt-bound tetramer, all of the structural elements required for tetramerization (i.e. TetrHis motif,  $V_L$ - $V_L$  interface and the scFv region onto which the TetrHis motif packs) are located outside of the CDR loops. An exception to this is the 1<sup>st</sup> residue of CDR1 of the  $V_L$  domain (Arg182), which contributes 1 hydrogen bond at the periphery of the  $V_L$ - $V_L$  interface. In order to test whether other scFvs can also use the TetrHis motif for tetramerization, we designed a new scFv sequence by replacing the CDR loops in scFv 2A2 by those from an anti-FLAG M2 Fab<sup>19</sup>, for which a crystal structure is available (PDB ID:2g60). The protein (named scFv 748), which shares 84.5% sequence identity with the original scFv, was produced with a pelB leader sequence and purified from *e.coli* periplasm (see methods). Although the purification yield obtained by this approach was limited, most probably due to lack of optimization of the framework sequence for bacterial expression, we could purify enough scFv 748 to test the effect of 5 mM of cobalt ions onto the measured SAXS profiles at a protein concentration of 0.5 mg/ml (Fig. 8a). The changes in the  $p(r)$  functions recapitulated the observations made in the same buffer and at the same protein concentration for scFv 2A2 (Figs. 8b and 5c). The curves could be well fitted by EOM using the monomer-tetramer ensemble, and yielded similar proportions of tetramers and monomers, as well as similar  $R_g$  distributions (Fig. 8c, d). Taken together, these results indicate that TetrHis - cobalt mediated tetramerization can be applied to various scFvs, provided that the same framework sequence and C-terminal tags are used.

## Discussion

In this study we characterized the structure and dynamics of scFv 2A2, uncovering a metal-dependent tetramerization motif containing a 6XHis sequence, which we called TetrHis. The motif possesses a unique  $\beta$ -stranded architecture and harbors 8 metal binding sites clustered within a small region of space, with inter-site distances of 10–30 Å. This discovery was made possible thanks to the serendipitous combination of (1) the right C-terminal tags; (2) an antibody framework sequence capable of stabilizing the motif; (3) successful crystallization in the presence of a high concentration of cobalt ions. The tetrameric assembly was further stabilized by the formation of VL-11  $\beta$ -sheet dimers at the  $V_L$ – $V_L$  interfaces, a common type of weak homotypic antibody interface<sup>22</sup>.

Using the PDB structure motif search tool, we found metal-coordinated polyhistidine sequences to be relatively rare in experimental structures. In most cases, a 6X His sequence was involved in crystal packing, stabilizing a dimeric state through metal coordination in conjunction with other residues located far away in the protein sequence. We were unable to identify any motif structurally homologous to TetrHis in the PDB, nor did we find any examples of polyhistidine motifs stabilizing tetrameric states.

Our analysis of DLS, solution, and SEC-SAXS data unambiguously showed that the existence of the TetrHis motif is not limited to the crystal, and that the tetrameric complex can also form in solution. In regular buffer conditions, we found that scFv 2A2 mostly adopted the classical scFv fold, but was in equilibrium with a minor population of open states resulting from the dissociation of  $V_H$  and  $V_L$  domains. This appears to be a frequent property of scFvs that have not been engineered for stability<sup>27,28</sup>, although to the best of our knowledge, such equilibrium has not previously been described in atomistic details by SAXS-based ensemble analysis. In addition, we observed that the C-terminal extension behaved as a typical intrinsically disordered region.

In contrast, when  $Co^{2+}$  ions were added to the buffer, we observed structural changes in the intrinsically disordered C-terminal extension with the appearance of a compact close state population in which the tags pack onto the scFv surface. The formation of tetramers, stabilized by the metal-bound TetrHis motif, was found to increase with protein concentration in the 0.5 to 1 mg/ml range, and plateau at 40 % between 1 and 2 mg/ml. For data measured at higher protein concentrations, we found that the quality of fit deteriorated, probably due to the formation of a small population of higher order oligomers and/or inter-particle interference. However, the analysis of  $p(r)$  distributions, and even the composition of the (poorly fitting) optimized ensembles, suggests that the population of tetramers keeps increasing at these concentrations. Furthermore, the fact that the protein crystallized as a tetramer at 4–8 mg/ml in 5–10 mM  $CoCl_2$  seems to indicate that this should be the dominant scFv conformation in these conditions. Taken together, the SAXS analysis and DLS measurements suggest that tetramer formation is favored at high protein concentrations with cobalt concentrations in the mM range (e.g., 2–10 mM).

Because the atomic contacts between the TetrHis motif and the scFv tetramer are located in the scFv framework regions rather than the CDRs, it is reasonable to assume that the metal-dependent tetramerization property could be transferred to other scFvs provided the same antibody framework and C-terminal tags are used. We tested this hypothesis by producing scFv 748, in which the CDRs present in scFv 2A2 were replaced by those from a previously crystallized anti-FLAG M2 Fab<sup>19</sup>. Our SAXS data analysis showed that the new scFv behaved very similarly to the original one in the presence of cobalt ions, indicating that the tetramerization property conferred by the TetrHis motif should

be transferable to a wide range of scFvs with various binding specificities. This suggests that the TetrHis motif could be used as a chemically switchable fiducial and size enhancer for cryo-EM structural determination of scFv complexes, providing a tetra-valent platform with a fixed orientation of bound particles.

Although the TetrHis motif requires additional scFv framework residues to function, the crystal structure of scFv 2A2 provides information about the parts of the scFv framework that are required to stabilize the motif, and parts that are dispensable and can be engineered to improve, for example, the stability of the monomer to increase bacterial expression levels. In addition, the structural information can be used to design mutations that further improve the stability range of the cobalt-stabilized tetramer by creating additional interactions.

Metal–protein nanohybrid materials are a class of fast emerging functional nanomaterials with a broad range of potential applications such as biomineralization, catalysis, drug delivery, tumor imaging and therapy, and others<sup>30</sup>. In the era of protein design, the TetrHis motif could also be used as a template to design dense metal clusters embedded within larger protein structures, or metal-dependent self-assembling repeating units, thereby generating new types of hydrogels, or even protein-based metal-organic frameworks (MOF).

## Materials and methods

**Construct design, expression, and purification of scFv 2A2.** ScFv 2A2 was designed by inserting the CDRs sequences described for anti-ceramide antibody (<https://patents.google.com/patent/US20190389970A1/en> and<sup>17</sup>) into a mouse scFv scaffold with a (Gly<sub>4</sub>Ser)<sub>4</sub> linker and by adding a sortase motif LPETG followed by a 6x-Histag and a Twin-Strep-tag<sup>31</sup> at the C-terminus. The corresponding synthetic gene was synthesized (Eurofins Genomics) and cloned into *Drosophila melanogaster* S2 expression vector for scFv<sup>32</sup>. This resulted in a mature secreted construct with the following primary structure:  $V_H - (Gly_4Ser)_4 - V_L - \text{sortase} - His_6 - \text{TwinStrep}$ . *Drosophila* S2 cells were transfected as reported previously<sup>33</sup>, amplified, and scFv expression was induced with 4  $\mu\text{M}$   $CdCl_2$  at a density of  $\sim 10 \times 10^6$  cells per ml for 6–8 days for large-scale production. The protein was purified from the supernatant by affinity chromatography using a Strep-Tactin resin (IBA) according to manufacturer's instructions followed by SEC on a Superdex200 column (GE Healthcare). Pure monomeric scFv was concentrated to 8.5 mg.ml<sup>-1</sup> and frozen at  $-80^\circ\text{C}$ .

**Construct design, expression, and purification of scFv 748.** ScFv 748 was designed by performing a structural alignment of the scFv 2A2 monomer onto the crystal structure of anti-FLAG M2 Fab<sup>19</sup> (PDB ID 2G60), and replacing the CDR loops from scFv 2A2. The synthetic gene corresponding to the designed sequence was ordered from Genecust, custom cloned into a pET26b vector with a pelB leader sequence for periplasmic expression in *E. coli*. The plasmid was transformed into BL21 DE3 *E. coli* cells and grown on kanamycin-supplemented plates. A single colony was used to inoculate liquid cultures overnight. The scFv was then expressed by overnight incubation under shaking at 18 °C following 0.25 mM IPTG induction of 4 L of LB after the OD<sub>600nm</sub> reached 0.6. Cells were harvested by centrifugation and the resulting cell pellets were resuspended in 20 mM Tris, pH 7.5, 150 mM NaCl. Cells were lysed by sonication, and the lysate was centrifuged for 25 min at 4 °C and 50,000 $\times$ g to remove cell debris. The supernatant was loaded onto a column containing 2 ml of pre-equilibrated Ni-NTA superflow (QIAGEN). After extensive washes, the protein was eluted in 20 mM Tris, pH 7.5, 150 mM NaCl, 400 mM imidazole. The eluate was then diluted to <200 mM imidazole concentration and loaded onto a column containing 2 ml of pre-equilibrated Strep-Tactin resin (IBA). After washing, the protein was eluted using 2.5 mM desthiobiotin. Concentrated eluate was then subjected to size exclusion chromatography on a S200 column equilibrated in 50 mM HEPES pH 7.5 and 150 mM NaCl.

**X-ray crystallography.** Crystallization was carried out by vapor diffusion using a Cartesian Technologies pipetting system<sup>34</sup>. scFv 2A2 was concentrated to 8.5 mg/ml in 50 mM HEPES pH 7.5 and 150 mM NaCl. The protein crystallized at 20 °C in mother liquor containing 0.01 M cobalt chloride, 0.1 M MES, pH6.5, 1.8 M ammonium sulfate after  $\sim 5$ –10 days. Crystals were frozen in liquid nitrogen after being cryoprotected with mineral oil. Diffraction data were collected at X06SA beamline of the Swiss Light Source (SLS), Villigen, Switzerland, with a wavelength of 0.999 Å. A single dataset at 2.5 Å resolution was obtained. All data were automatically processed by xia2<sup>35</sup>. Structural determination was initiated by molecular replacement using an homology model of scFv 2A2 obtained via SWISS-MODEL<sup>36</sup> as a search model in PHASER<sup>37</sup>. The solution was subjected to repetitive rounds of restrained refinement in PHENIX<sup>38</sup> and Autobuster<sup>39</sup> and manual building in

COOT<sup>40</sup>. TLS parameters were included in the final round of refinement. Data collection and refinement statistics are provided in Table 1, and the final refined coordinates and structure factors have been deposited in the PDB with accession code 8CGE.

**Dynamic light scattering.** DLS measurements were performed at 20 °C using the Malvern Zetasizer Nano S instrument (Malvern, Worcestershire, England) equipped with a Peltier temperature controller. Data analysis was performed using the Zetasizer Nano S DTS software package. The hydrodynamic radii calculated from the intensity size distributions were compared to theoretical values obtained from structural models of scFv 2A2 using the HullRad web server<sup>41</sup>.

**Small-angle x-ray scattering.** Small-angle X-ray scattering measurements of scFv 2A2 at 2.1, 4.2, and 8.5 mg/ml were performed at the SWING beamline of the French national synchrotron facility (SOLEIL). Data was collected at 15 °C, a wavelength of 1.0332 Å and a sample-to-detector distance of 1.99 m. For batch measurements, the scattering from the buffer alone was measured before and after each sample measurement and was used for background subtraction with PRIMUS from the ATSAS package<sup>42</sup>. For SEC-SAXS measurements, scFv 2A2 samples at 8.5 mg/ml were loaded onto a Superdex200 5/150 column (GE Healthcare) previously equilibrated in 50 mM HEPES pH 7.5 and 150 mM NaCl with or without addition of 5 mM CoCl<sub>2</sub>. The measured SAXS images were normalized to the transmitted intensity and azimuthally averaged by using the in-house software Foxtrot (<https://www.synchrotron-soleil.fr/en/beamlines/swing>). The resulting SAXS curves were analyzed using CHROMIXS<sup>43</sup> and PRIMUS<sup>44</sup>.

Additional SAXS measurements were performed on beamline BM29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Samples were kept at 20 °C and data were collected at a wavelength of 0.0995 nm and a sample-to-detector distance of 1 m. 1D scattering profiles were generated and buffer subtraction was carried out by the automated data processing pipeline available at BM29.

**Molecular dynamics and ensemble optimization of scFv 2A2.** Classical explicit solvent molecular dynamics simulations were used to generate conformational ensembles of scFv 2A2 in its monomeric and cobalt-stabilized tetrameric states, in order to fit the SAXS data. The starting models were extracted from the scFv 2A2 crystal structure and missing (disordered) residues from the (Gly<sub>4</sub>Ser)<sub>4</sub> were manually added in Coot. Both systems were simulated in GROMACS<sup>45</sup> using the amber99SBws forcefield<sup>46</sup>, which was designed to reproduce the properties of intrinsically disordered proteins. In order to maintain the geometry of the 8 cobalt binding sites in the tetrameric form, the metal centers were replaced by zinc and harmonic distance restraints (cutoff 2.5 Å) were applied between each Zn<sup>2+</sup> ion and the coordinating residues. The zinc (II) parameters are available as part of the standard amber99SBws forcefield distribution. At the beginning of each simulation, the protein was immersed in a box of TIP4P2005 water, with a minimum distance of 1.0 nm between protein atoms and the edges of the box. The genion tool was used to add 150 mM NaCl. Long-range electrostatics were treated with the particle-mesh Ewald summation. Bond lengths were constrained using the P-LINCS algorithm. The integration time step was 5 fs. The v-rescale thermostat and the Parrinello–Rahman barostat were used to maintain a temperature of 300 K and a pressure of 1 atm. Each system was energy minimized using 1,000 steps of steepest descent and equilibrated for 500 ps with restrained protein heavy atoms prior to production simulations. Multiple independent MD trajectories were calculated for each system, for a total aggregated simulation time of ≈2.8 μs for the monomer and ≈2.0 μs for the tetramer. Snapshots were extracted every 1 ns from each trajectory, leading to the generation of 2760 models of the monomer and 2042 models of the tetramer. For each model from both ensembles, theoretical SAXS patterns were calculated with CRY SOL<sup>47</sup> and ensemble optimization fitting was performed with GAJOE<sup>29</sup>.

**Reporting summary.** Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The final refined coordinates and structure factors of scFv 2A2 crystal structure (Supplementary Data 1) have been deposited in the PDB with accession code 8CGE.

Received: 2 March 2023; Accepted: 21 July 2023;

Published online: 28 July 2023

## References

1. Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533 (2003).
2. Salichs, E. et al. Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genet.* **5**, e1000397 (2009).
3. Faux, N. G. et al. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* **15**, 537–51 (2005).
4. Taylor, K. M. & Nicholson, R. I. The LZT proteins; the LIV-1 subfamily of zinc transporters. *Biochim. Biophys. Acta* **1611**, 16–30 (2003).
5. Seshadri, S., Benoît, S. L. & Maier, R. J. Roles of His-rich hpn and hpn-like proteins in *Helicobacter pylori* nickel physiology. *J. Bacteriol.* **189**, 4120–6 (2007).
6. Fu, C., Olson, J. W. & Maier, R. J. HypB protein of *Bradyrhizobium japonicum* is a metal-binding GTPase capable of binding 18 divalent nickel ions per dimer. *Proc. Natl Acad. Sci. USA* **92**, 2333–7 (1995).
7. Hecel, A. et al. Histidine tracts in human transcription factors: insight into metal ion coordination ability. *J. Biol. Inorg. Chem.* **23**, 81–90 (2018).
8. Watly, J. et al. African Viper Poly-His Tag Peptide Fragment Efficiently Binds Metal Ions and Is Folded into an alpha-Helical Structure. *Inorg. Chem.* **54**, 7692–702 (2015).
9. Watly, J. et al. Insight into the coordination and the binding sites of Cu(2+) by the histidyl-6-tag using experimental and computational tools. *Inorg. Chem.* **53**, 6675–83 (2014).
10. Pontecchiani, F. et al. The unusual binding mechanism of Cu(II) ions to the poly-histidyl domain of a peptide found in the venom of an African viper. *Dalton Trans.* **43**, 16680–9 (2014).
11. Watly, J. et al. Uncapping the N-terminus of a ubiquitous His-tag peptide enhances its Cu(2+) binding affinity. *Dalton Trans.* **48**, 13567–13579 (2019).
12. Brasili, D. et al. The unusual metal ion binding ability of histidyl tags and their mutated derivatives. *Dalton Trans.* **45**, 5629–39 (2016).
13. Fukuda, N. et al. Production of single-chain Fv antibodies specific for GA-pyridine, an advanced glycation end-product (AGE), with reduced inter-domain motion. *Molecules* **22**, 1695 (2017).
14. Kim, J. H. et al. Crystal structures of mono- and bi-specific diabodies and reduction of their structural flexibility by introduction of disulfide bridges at the Fv interface. *Sci. Rep.* **6**, 34515 (2016).
15. Arndt, K. M., Muller, K. M. & Pluckthun, A. Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment. *Biochemistry* **37**, 12918–26 (1998).
16. Ludel, F. et al. Distinguishing between monomeric scFv and diabody in solution using light and small angle X-ray scattering. *Antibodies (Basel)* **8**, 48 (2019).
17. Rotolo, J. et al. Anti-ceramide antibody prevents the radiation gastrointestinal syndrome in mice. *J. Clin. Invest.* **122**, 1786–90 (2012).
18. Vasiliauskaitė-Brooks, I. et al. Structural insights into adiponectin receptors suggest ceramidase activity. *Nature* **544**, 120–123 (2017).
19. Roosild, T. P., Castronovo, S. & Choe, S. Structure of anti-FLAG M2 Fab domain and its use in the stabilization of engineered membrane proteins. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62**, 835–9 (2006).
20. Yang, X. et al. Molecular basis of a protective/neutralizing monoclonal antibody targeting envelope proteins of both tick-borne encephalitis virus and louping ill virus. *J. Virol.* **93**, e02132–18 (2019).
21. Nymalm, Y. et al. Antiferritin VL homodimer binds human spleen ferritin with high specificity. *J. Struct. Biol.* **138**, 171–86 (2002).
22. Yin, Y. et al. Antibody interfaces revealed through structural mining. *Comput. Struct. Biotechnol. J.* **20**, 4952–4968 (2022).
23. Fan, S. et al. Solution structure of synbindin atypical PDZ domain and interaction with syndecan-2. *Protein Pept. Lett.* **16**, 189–95 (2009).
24. Martino, L. et al. The interaction of the *Escherichia coli* protein SlyD with nickel ions illuminates the mechanism of regulation of its peptidyl-prolyl isomerase activity. *FEBS J.* **276**, 4529–44 (2009).
25. Low, C. et al. Crystal structure determination and functional characterization of the metallochaperone SlyD from *Thermus thermophilus*. *J. Mol. Biol.* **398**, 375–90 (2010).
26. Quistgaard, E. M. et al. Molecular insights into substrate recognition and catalytic mechanism of the chaperone and FKBP peptidyl-prolyl isomerase SlyD. *BMC Biol.* **14**, 82 (2016).
27. Worn, A. & Pluckthun, A. Different equilibrium stability behavior of ScFv fragments: identification, classification, and improvement by protein engineering. *Biochemistry* **38**, 8739–50 (1999).
28. Honegger, A. Engineering antibodies for stability and efficient folding. *Handb. Exp. Pharmacol.* **181**, 47–68 (2008).
29. Bernado, P. et al. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–64 (2007).
30. Saif, B. & Yang, P. Metal-protein hybrid materials with desired functions and potential applications. *ACS Appl. Biol. Mater.* **4**, 1156–1177 (2021).
31. Schmidt, T. G. et al. Development of the Twin-Strep-tag(R) and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expr. Purif.* **92**, 54–61 (2013).

32. Gilmartin, A. A. et al. High-level secretion of recombinant monomeric murine and human single-chain Fv antibodies from *Drosophila* S2 cells. *Protein Eng. Des. Sel.* **25**, 59–66 (2012).
33. Johansson, D. X., Krey, T. & Andersson, O. Production of recombinant antibodies in *Drosophila melanogaster* S2 cells. *Methods Mol. Biol.* **907**, 359–70 (2012).
34. Walter, T. S. et al. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 651–7 (2005).
35. Winter, G., Lobley, C. M. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–73 (2013).
36. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
37. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
38. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–21 (2010).
39. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–21 (2004).
40. Emsley, P. et al. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
41. Fleming, P. J. & Fleming, K. G. HullRad: fast calculations of folded and disordered protein and nucleic acid hydrodynamic properties. *Biophys. J.* **114**, 856–869 (2018).
42. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
43. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
44. Konarev, P. V. et al. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
45. Pall, S. et al. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **153**, 134110 (2020).
46. Best, R. B., Zheng, W. & Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **10**, 5113–5124 (2014).
47. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLO-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

## Acknowledgements

We acknowledge the ESRF, PSI, and SOLEIL for provision of synchrotron radiation facilities and we would like to thank the staff of beamline X06SA at the PSI for assistance

with crystal testing and data collection, the staff of BM29 beamline at ESRF and the staff of SWING beamline at SOLEIL for assistance with SAXS data acquisition.

## Author contributions

R.D.H. designed, expressed, purified and crystallized scFv 2A2 preparations with the help of L.C. and F.H.; F.H. collected x-ray crystallography data. C.L. solved and refined the x-ray structure. C.L. and A.M. measured SAXS data. C.L. analyzed SAXS data and performed the computational studies. C.L. and P.C. prepared anti-Flag M2 scFv with the help of A.F. C.L. and A.F. measured and analyzed DLS data. C.L. and S.G. wrote the manuscript and jointly supervised the overall project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42004-023-00962-x>.

**Correspondence** and requests for materials should be addressed to Sébastien Granier or Cédric Leyrat.

**Peer review information** *Communications Chemistry* thanks Hsiao-Ching Yang, Antonio Rosato, Peng Yang, and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023