Structure and Dynamics of Tetrhis, a Novel Metal-dependent Polyhistidine Tetramerization Motif

Robert Healey, Louise Couillaud, François Hoh, Assia Mouhand, Sebastien Granier, Cedric Leyrat

Structure and Dynamics of Tetrhis, a Novel Metal-dependent Polyhistidine Tetramerization Motif

Robert Healey
Sosei-Heptares
Louise Couillaud
INSERM
francois hoh
CBS
Assia Mouhand
University Montpellier https://orcid.org/0000-0002-8800-4633
Sebastien Granier
IGF, University of Montpellier, CNRS, INSERM, Montpellier https://orcid.org/0000-0003-1550-3658
Cedric Leyrat (cedric.leyrat@igf.cnrs.fr)
INSERM, Institut de Genomique fonctionnelle https://orcid.org/0000-0003-0189-0562

Article

Keywords:

Posted Date: February 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2548431/v1

Abstract

The polyhistidine (6XHis) tag is one of the most ubiquitous fusion tags used for the purification of recombinant proteins. The 6XHis motif enables the tagged protein to bind metal ions such as Co^{2+}, Ni^{2+}, Zn^{2+} and Cu^{2+}, which can be advantageously used for protein purification with immobilized metal affinity chromatography (IMAC). Despite its widespread use, high-resolution protein structures encompassing a metal-bound 6XHis motif rarely occur in the worldwide Protein Data Bank (PDB). Here, we obtained a 2.5 Å resolution crystal structure of a single chain variable fragment (scFv) antibody bearing a C-terminal tag composed of a sortase motif (LPETG), followed by a 6XHis and a TwinStrep tag (WSHPQFEK$[G_3S]_3$WSHPQFEK). The protein crystallized in the presence of cobalt (II) and its structure reveals a novel tetramerization motif (TetrHis) stabilized by 8 Co^{2+} ions clustered within a small region of space. The TetrHis motif contains two 6 residues-long β-strands related by a twofold symmetry axis, and the 8 metal centers coordinate 3 to 5 protein residues each, including all of the 6XHis histidines, a glutamate residue from the sortase motif and an additional histidine residue from the TwinStrep tag. We reviewed metal-bound homopolymeric histidine tracts in deposited PDB structures and found no similar motifs. Using a combination of small angle x-ray scattering (SAXS), molecular dynamics (MD) simulations and ensemble analysis, we studied the structural dynamics of the tagged scFv in the absence and presence of Co^{2+} ions, demonstrating an open/close equilibrium of the monomer and the formation of a cobalt-stabilized tetramer in solution. The structure of the novel metal-dependent tetramerization motif reported here might provide a useful starting point for designing metal-loaded biomaterials, with potential applications in the fields of biosensors, bioanalytical devices, or biocatalysts.

Introduction

Polyhistidine 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 [1]. These polyhistidine sequences bind various metal ions such as Co^{2+}, Ni^{2+}, Zn^{2+} and 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 [2, 3], such as various zinc transporters [4], bacterial nickel metabolism proteins [5, 6], human transcription factors [7] or snake venoms [8]. A number of isolated peptide sequences have been studied for their metal-binding properties, ranging from a simple 6XHis [9] to more complex sequences such as the poly-His/poly-Gly peptide pHpG-1 ($EDDH_9GVG_{10}$) isolated from the venom of the African viper *Atheris squamigera* [8, 10, 11], or the peptide $H_2ASHGH_2NSH_2PQH_{11}$ corresponding to residues 33–57 of human Forkhead box protein G1 [7]. These peptides form complexes with metal ions that usually display polymorphic binding states [7–9, 12], sometimes coupled with the stabilization of α-helical structure [8].

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.

In this study, we report the serendipitous discovery of TetrHis, a metal-dependent tetramerization motif with minimal sequence ETGHHHHHHWSHPQ observed in the cobalt-bound crystallographic structure of a single chain variable fragment (scFv) antibody. The sequence of the 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). 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 [13] and have a tendency to form dimers and higher order oligomers depending on the length of the linker peptide [14–16]. 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 [17] onto a murine scFv scaffold with a $[GGGGS]_4$ linker (taken from PDB entry 5LX9 chain B [18]), 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 structure and dynamics of the tagged scFv molecule in the absence or presence of Co^{2+} ions using a combination of small angle x-ray scattering (SAXS), molecular dynamics simulations (MDS) and ensemble analysis, revealing its conformational equilibria in atomistic details.

Results

The X-ray structure of scFv 2A2 reveals a cobalt-stabilized tetrameric assembly. ScFv 2A2 crystallized in space group *I*121 in mother liquor containing 10 mM $CoCl_2$ 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 and Figure S1). Each scFv monomer adopts the classical arrangement of the V_H and V_L domains, and further dimerizes through a ~ 650 $Å^2$ V_L- V_L interface that is stabilized by 15 hydrogen bonds (Fig. 1D). This V_L- V_L interface has previously been observed in the crystal packing of a number of scFv structures [18, 19] and isolated V_L domains [20], as can be seen from the structural alignment shown in Figure S1. 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 [21]).

The most striking and unexpected feature of the structure is the presence of electron density for the C-terminal sortase recognition motif (^{266}LPETG^{270}), the 6XHis and the first Strep tag (^{277}WSHPQFEK^{284} or

only ^{277}WSHPQ^{281} in one of the two chains) of the Twin-strep tag (WSHPQFEK($G_3S)_3$WSHPQFEK). These 16 to 19 residues form an extension that adopts a tetrameric arrangement stabilized by 8 Co^{2+} ions (Fig. 1C and Fig. 2). The motif, named TetrHis, contains two 6 residue-long β-strands related by the two-fold symmetry axis, and the 8 Co^{2+} ions coordinate 3 to 5 protein residues each, resulting in 4 unique metal binding sites (Fig. 2A). In site 1 and 2, Co^{2+} ions are coordinated by residues belonging exclusively to the two 1st protein chains, thereby stabilizing the dimeric assembly. In contrast, the Co^{2+} coordination at site 3 and 4 connects protein chains across the two-fold crystallographic axis and thus stabilizes the tetramer (Fig. 2B). The Co^{2+} ion of site 1 is coordinated by His279 (Strep tag) of the 1st chain and His273 and His275 (6XHis) of the 2nd 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 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 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 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 β-strand from the 1st chain covers the ^{275}HHWSHP^{280} sequence and shares backbone-backbone hydrogen bonds with the 2nd chain β-strand which has the sequence ^{272}HHHHHW^{277} (Fig. 2); (2) while there is no electron density for residues after Gln281 of the 1st chain, in the 2nd chain residues ^{278}SHPQFEK^{284} interact with both the tetramerization motif and residues from the scFv framework, forming an α-helical turn (Fig. 1C,E and Fig. 2A). From both chains, Trp277 play an important stabilizing role in the complex, creating a hydrophobic core between the tetramerization motif and the V_L 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).

Table 1
Data collection and refinement statistics (molecular replacement)

	scFv 2A2
Data collection	
Space group	I 1 2 1
Cell dimensions	
a, *b*, *c* (Å)	97.50, 81.82, 111.04
α, β, γ (°)	90.00, 111.14, 90.00
Resolution (Å)	45.47–2.50 (2.60–2.50)*
R_{merge}	0.213 (1.293**)
$CC_{1/2}$	0.989 (0.491)
$I / \sigma I$	5.6 (1.1)
Completeness (%)	99.2 (96.8)
Redundancy	6.9 (6.2)
Refinement	
Resolution (Å)	45.47–2.50
No. reflections	28159
R_{work} / R_{free}	19.02/24.83
No. atoms	
Protein	3847
Cobalt	4
Water	196
B-factors	
Protein	47.8
Cobalt	60.4
Water	46.0
R.m.s. deviations	
Bond lengths (Å)	0.008

	scFv 2A2
Bond angles (°)	1.318

*Values in parentheses are for highest-resolution shell.

** R_{merge} value over 1 is statistically meaningless

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 [22] 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 (Table S1). 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 motifs 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 unknown whether the crystallographically observed assembly can also form in solution in the presence of Ni^{2+} ions, however entries 3CGM and 4ODP are crystal structures of SlyD from Thermus thermophiles, a protein well known for its metallochaperone activity [23–25]. Taken together, these analyses indicate that no linear metal-dependent polyhistidine tetramerization motif exists in the PDB, making TetrHis a novel sequence motif.

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 different protein concentrations in gel filtration buffer or in the presence of additives: 5 mM EDTA, 5 mM $NiSO_4$, or 5mM $CoCl_2$ (Fig. 4A and Table 2). We also attempted to measure data in the presence of 5mM $ZnCl_2$, but these showed severe signs of aggregation. The measured radius of gyration (Rg) in the presence of EDTA increased from 24 Å at 2 mg/ml of protein to 29 Å at 8 mg/ml, while the Rg values in regular buffer were comprised between 28 and 32 Å, at the respective concentrations. This concentration dependence of the Rg indicated significant interparticle attraction, however the measured values were roughly consistent with the Rg calculated from monomers extracted from the crystal structure, particularly at low protein concentration (theoretical Rg ≈ 20–28 Å dependent on whether the missing

residues have been added or not, and their respective conformations). In the presence of Co^{2+} or Ni^{2+} ions, the shape of the measured SAXS profiles changed slightly, with the appearance of a small bump at intermediate Q, particularly visible at high protein concentration (Fig. 4A), 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). Similar to the data measured in the absence of added metal ions, the measured Rg showed important protein concentration-dependent variations with values of 38–51 Å in the presence of Ni^{2+}, and 36–60 Å in the presence of Co^{2+} (Table 2). The 36 Å Rg value obtained at low protein concentration was also roughly consistent with the Rg calculated from the crystallographic tetramer (up to 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 the lowest protein concentrations in the presence of EDTA versus Co^{2+} ions suggests that the metal ions induce a transition from a mostly globular to a multidomain protein (Fig. 4B).

Table 2
SAXS-derived parameters.

	Buffer conditions	c (mg/ml)	MW (kDa)	R_g (Å)
SAXS				
	Buffer A*	8.5	47.6	32.3 ± 0.2
	Buffer A	4.2	40.6	30.0 ± 0.2
	Buffer A	2.1	35.8	27.6 ± 0.3
	Buffer A + 5mM EDTA	8.5	39.7	29.0 ± 0.3
	Buffer A + 5mM EDTA	4.2	29.4	25.6 ± 0.2
	Buffer A + 5mM EDTA	2.1	26.1	24.1 ± 0.3
	Buffer A + 5mM $CoCl_2$	8.5	184.6	50.6 ± 0.4
	Buffer A + 5mM $CoCl_2$	4.2	124.7	44.7 ± 0.4
	Buffer A + 5mM $CoCl_2$	2.1	81.3	38.0 ± 0.2
	Buffer A + 5mM $NiSO_4$	8.5	248.5	60.2 ± 0.5
	Buffer A + 5mM $NiSO_4$	4.2	126.4	45.2 ± 0.7
	Buffer A + 5mM $NiSO_4$	2.1	75.4	36.3 ± 0.8
SEC-SAXS				
Frame 464–524	Buffer A	8.5**	34.7	26.5 ± 0.1
Frame 390–455	Buffer A + 5mM $CoCl_2$	8.5**	55.4	31.7 ± 0.1
*Buffer A: 50 mM HEPES pH 7.5 150 mM NaCl				

** Concentration of the sample that was injected on the SEC.

Because the strong concentration dependence of the measured Rg values precluded more detailed analysis, we next measured SEC-SAXS data with or without addition of Co^{2+} ions in the SEC buffer. We hypothesized that the size-based separation of the SEC would result in higher data quality by decreasing interparticle interactions and removing potential signal contributions from small amounts of higher order oligomers or aggregates. ScFv 2A2 eluted from the SEC as a single, symmetric peak as can be seen from the average intensity and UV_{280} profiles (Fig. 4C and Figure S2). The measured Rg was mostly constant across the peak with a slight downward trend in the second half of the peak and an average value of 26.5 Å, in agreement with the Rg value expected for a monomeric species (Table 2 and Fig. 4C). The estimated molecular weight (MW) was also mostly constant ranging from 28 to 33 kDa (Figure S2B) in agreement

with the theoretical MW of 30.5 kDa. In contrast, the profile measured in the presence of 5 mM $CoCl_2$ displayed a wider, assymmetric peak that was slightly shifted towards lower elution volumes (Fig. 4C). The measured Rg showed important variations across the peak with values ranging from ~ 25 to 34 Å (Fig. 4C) and the estimated MW was comprised between 30 and 60 kDa (Figure S2B). 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 2 independent trajectories of the monomer (Fig. 5A and Figure S3A). 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. 5A). This spontaneous transition 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 [26, 27]. We also ran multiple independent MD trajectories of the tetrameric scFv complex (Fig. 5B and Figure S3B) in which the cobalt ions were replaced by zinc ions (as parameters for 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. 5B). 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 the SEC-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 to 20) ensemble of models that fits the experimental SAXS data from a larger pool of structural models (> 1000) [28]. This method was used to fit SAXS curves extracted from the SEC-SAXS experiments using the pool of MD models of scFv 2A2 (Fig. 6). The data measured in the absence of cobalt was well fitted using the monomeric ensemble of models with a χ_{exp} value of 1.08 (Fig. 6A and C). The distribution of Rg for models of the selected ensemble shows a single peak centered on the experimental Rg value of 26.5 Å (Fig. 6A). The ensemble is composed of 85% of closed conformers and 15% of open conformers, and the C-terminal extension behaves as an intrinsically disordered tail adopting mostly extended conformations (Fig. 6E).

For the data measured in the presence of cobalt ions, a SAXS curve was extracted from the first part of the asymmetric elution peak (frame 390–455, see Fig. 4C). The data could not be adequately fitted using the monomeric scFv ensemble (χ_{exp} = 4.64) and required a pool ensemble containing both monomeric and tetrameric forms (Fig. 6B) to reach a satisfactory χ_{exp} value of 1.03 (Fig. 6C). The Rg distribution

reveals a complex landscape induced by the cobalt ions (Fig. 6B). 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 - Fig. 6D). Analysis of the selected ensemble indicates that 20% of scFv molecules are assembled as tetramers, which are in equilibrium with close state (50%) and open state monomers (30%) (Fig. 6F). Interestingly, the selected close state monomers are much more compact than those selected in the absence of cobalt ions (Fig. 6A vs Fig. 6B, and Fig. 6E vs Fig. 6F), 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.

Discussion

In this study we characterized the structure and dynamics of scFv 2A2, uncovering a novel metal-dependent tetramerization motif containing a 6XHis sequence, which we called TetrHis. The motif possesses a unique β-stranded architecture and harbours 8 metal binding sites clustered within a small region of space, with inter-site distances of 10 to 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 β-sheet dimers at the V_L-V_L interfaces, a common type of weak homotypic antibody interface [21].

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 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 [26, 27], 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, and the formation of tetramers stabilized by the metal-bound TetrHis motif. In the conditions we used (5 mM $CoCl_2$), the proportion of tetrameric states extracted from the SEC-

SAXS data was ~ 20%, however the solution SAXS measurements suggested that a larger proportion of tetramers might be present in solution, but partially dissociate on the SEC column due to sample dilution.

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. This would enable the use of the TetrHis motif as a chemically switchable fiducial and size enhancer for cryo-EM structural determination of scFv complexes, providing a tetravalent platform with a fixed orientation of bound particles.

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 [29]. 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 [17]) into a mouse scFv scaffold with a (Gly4Ser)4 linker and by adding a sortase motif LPETG followed by a 6x-Histag and a Twin-Strep-tag [30] at the C-terminus. The corresponding synthetic gene was synthesized (Eurofins Genomics) and cloned into Drosophila melanogaster S2 expression vector for scFv [31]. This resulted in a mature secreted construct with the following primary structure: V_H - $(Gly_4Ser)_4$ - V_L – sortase – His_6 – TwinStrep. Drosophila S2 cells were transfected as reported previously[32], amplified, and scFv expression was induced with 4 µM $CdCl_2$ at a density of ~ 10x 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 $mg.ml^{-1}$ and frozen at -80°C.

X-ray crystallography

Crystallization was carried out by vapor diffusion using a Cartesian Technologies pipetting system [33]. 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 0.01M cobalt chloride, 0.1M MES, pH6.5, 1.8 M ammonium sulfate after about 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 [34]. Structural determination was initiated by

molecular replacement using an homology model of scFv 2A2 obtained via SWISS-MODEL [35] as a search model in PHASER [36]. The solution was subjected to repetitive rounds of restrained refinement in PHENIX [37] and Autobuster [38] and manual building in COOT [39]. 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.

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 [40]. 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 5mM $CoCl_2$. 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 [41] and PRIMUS [42].

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 $(Gly_4Ser)_4$ were manually added in Coot. Both systems were simulated in GROMACS [43] using the amber99SBws forcefield [44], 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 ion and the coordinating residues. 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 CRYSOL [45] and ensemble optimization fitting was performed with GAJOE [28].

Declarations

Acknowledgements

We acknowledge the 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, and the staff of SWING beamline at SOLEIL for assistance with SAXS data acquisition.

Competing interests

The authors declare that no competing interests exist.

References

1. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. **60**(5): p. 523-33.
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, 2009. **5**(3): p. e1000397.
3. Faux, N.G., et al., *Functional insights from the distribution and role of homopeptide repeat-containing proteins.* Genome Res, 2005. **15**(4): p. 537-51.
4. Taylor, K.M. and R.I. Nicholson, *The LZT proteins; the LIV-1 subfamily of zinc transporters.* Biochim Biophys Acta, 2003. **1611**(1-2): p. 16-30.
5. Seshadri, S., S.L. Benoit, and R.J. Maier, *Roles of His-rich hpn and hpn-like proteins in Helicobacter pylori nickel physiology.* J Bacteriol, 2007. **189**(11): p. 4120-6.
6. Fu, C., J.W. Olson, and R.J. Maier, *HypB protein of Bradyrhizobium japonicum is a metal-binding GTPase capable of binding 18 divalent nickel ions per dimer.* Proc Natl Acad Sci U S A, 1995. **92**(6): p. 2333-7.
7. Hecel, A., et al., *Histidine tracts in human transcription factors: insight into metal ion coordination ability.* J Biol Inorg Chem, 2018. **23**(1): p. 81-90.
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, 2015. **54**(16): p. 7692-702.
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, 2014. **53**(13): p. 6675-83.
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, 2014. **43**(44): p. 16680-9.
11. Watly, J., et al., *Uncapping the N-terminus of a ubiquitous His-tag peptide enhances its Cu(2+) binding affinity.* Dalton Trans, 2019. **48**(36): p. 13567-13579.
12. Brasili, D., et al., *The unusual metal ion binding ability of histidyl tags and their mutated derivatives.* Dalton Trans, 2016. **45**(13): p. 5629-39.

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, 2017. **22**(10).
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, 2016. **6**: p. 34515.
15. Arndt, K.M., K.M. Muller, and A. Pluckthun, *Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment.* Biochemistry, 1998. **37**(37): p. 12918-26.
16. Ludel, F., et al., *Distinguishing Between Monomeric scFv and Diabody in Solution Using Light and Small Angle X-ray Scattering.* Antibodies (Basel), 2019. **8**(4).
17. Rotolo, J., et al., *Anti-ceramide antibody prevents the radiation gastrointestinal syndrome in mice.* J Clin Invest, 2012. **122**(5): p. 1786-90.
18. Vasiliauskaite-Brooks, I., et al., *Structural insights into adiponectin receptors suggest ceramidase activity.* Nature, 2017. **544**(7648): p. 120-123.
19. 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, 2019. **93**(8).
20. Nymalm, Y., et al., *Antiferritin VL homodimer binds human spleen ferritin with high specificity.* J Struct Biol, 2002. **138**(3): p. 171-86.
21. Yin, Y., et al., *Antibody interfaces revealed through structural mining.* Comput Struct Biotechnol J, 2022. **20**: p. 4952-4968.
22. Fan, S., et al., *Solution structure of synbindin atypical PDZ domain and interaction with syndecan-2.* Protein Pept Lett, 2009. **16**(2): p. 189-95.
23. 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, 2009. **276**(16): p. 4529-44.
24. Low, C., et al., *Crystal structure determination and functional characterization of the metallochaperone SlyD from Thermus thermophilus.* J Mol Biol, 2010. **398**(3): p. 375-90.
25. Quistgaard, E.M., et al., *Molecular insights into substrate recognition and catalytic mechanism of the chaperone and FKBP peptidyl-prolyl isomerase SlyD.* BMC Biol, 2016. **14**(1): p. 82.
26. Worn, A. and A. Pluckthun, *Different equilibrium stability behavior of ScFv fragments: identification, classification, and improvement by protein engineering.* Biochemistry, 1999. **38**(27): p. 8739-50.
27. Honegger, A., *Engineering antibodies for stability and efficient folding.* Handb Exp Pharmacol, 2008(181): p. 47-68.
28. Bernado, P., et al., *Structural characterization of flexible proteins using small-angle X-ray scattering.* J Am Chem Soc, 2007. **129**(17): p. 5656-64.
29. Saif, B. and P. Yang, *Metal-Protein Hybrid Materials with Desired Functions and Potential Applications.* ACS Appl Bio Mater, 2021. **4**(2): p. 1156-1177.
30. 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, 2013. **92**(1): p. 54-61.

31. 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, 2012. **25**(2): p. 59-66.
32. Johansson, D.X., T. Krey, and O. Andersson, *Production of recombinant antibodies in Drosophila melanogaster S2 cells.* Methods Mol Biol, 2012. **907**: p. 359-70.
33. 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, 2005. **61**(Pt 6): p. 651-7.
34. Winter, G., C.M. Lobley, and S.M. Prince, *Decision making in xia2.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1260-73.
35. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.
36. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
37. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
38. Blanc, E., et al., *Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2210-21.
39. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
40. Franke, D., et al., *ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.* J Appl Crystallogr, 2017. **50**(Pt 4): p. 1212-1225.
41. Panjkovich, A. and D.I. Svergun, *CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data.* Bioinformatics, 2018. **34**(11): p. 1944-1946.
42. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of Applied Crystallography, 2003. **36**(5): p. 1277-1282.
43. Pall, S., et al., *Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS.* J Chem Phys, 2020. **153**(13): p. 134110.
44. Best, R.B., W. Zheng, and J. Mittal, *Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association.* J Chem Theory Comput, 2014. **10**(11): p. 5113-5124.
45. Svergun, D., C. Barberato, and M.H.J. Koch, *CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates.* Journal of Applied Crystallography, 1995. **28**(6): p. 768-773.

Figures

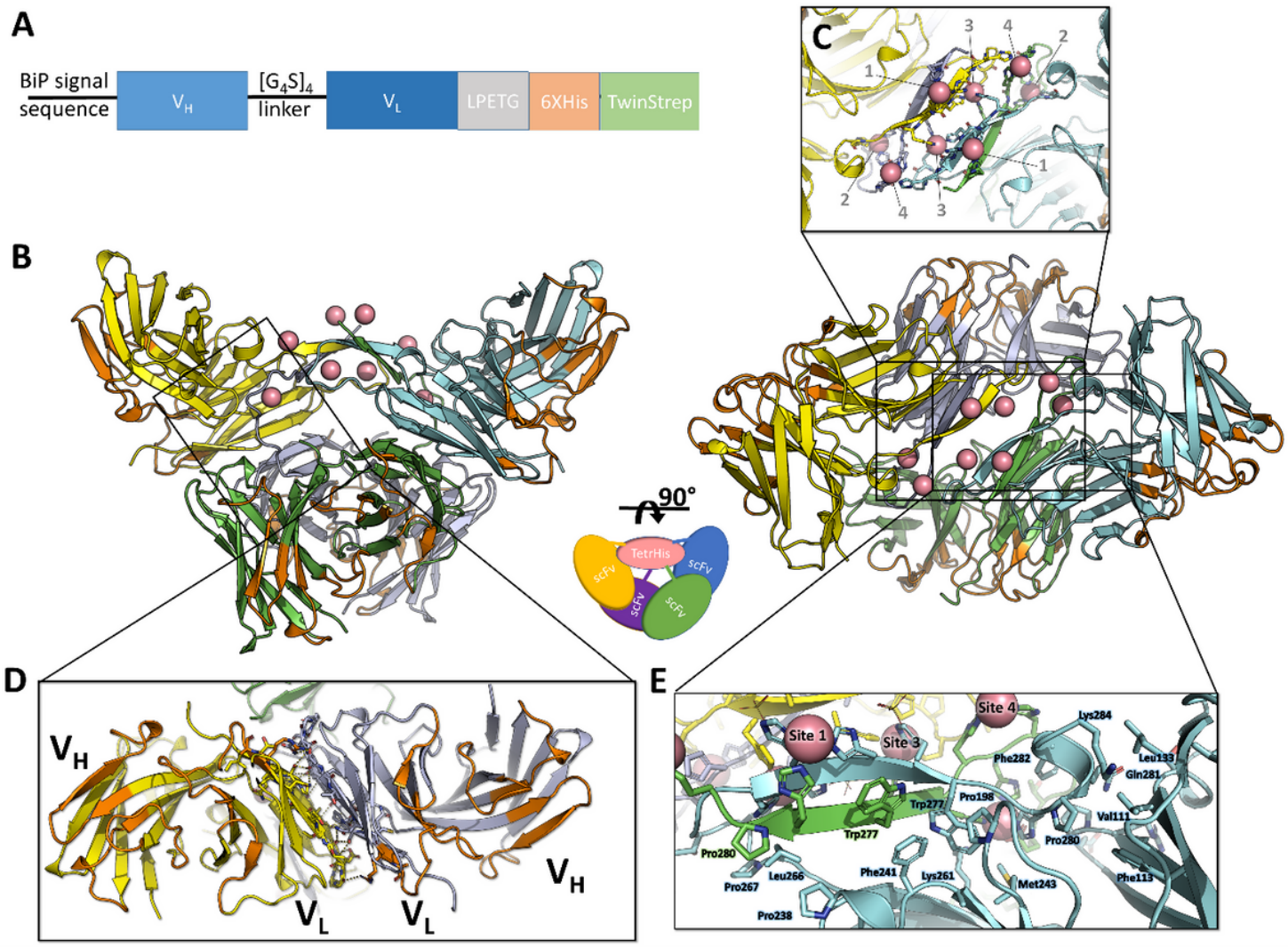

Figure 1

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$[G_3S]_3$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 metal cluster showing the 4 types of cobalt binding sites related by the twofold symmetry axis. D. Close up view of the V_L - V_L interface. E. Hydrophobic interactions stabilizing the interface between the TetrHis motif and the scFv V_H and V_L domains.

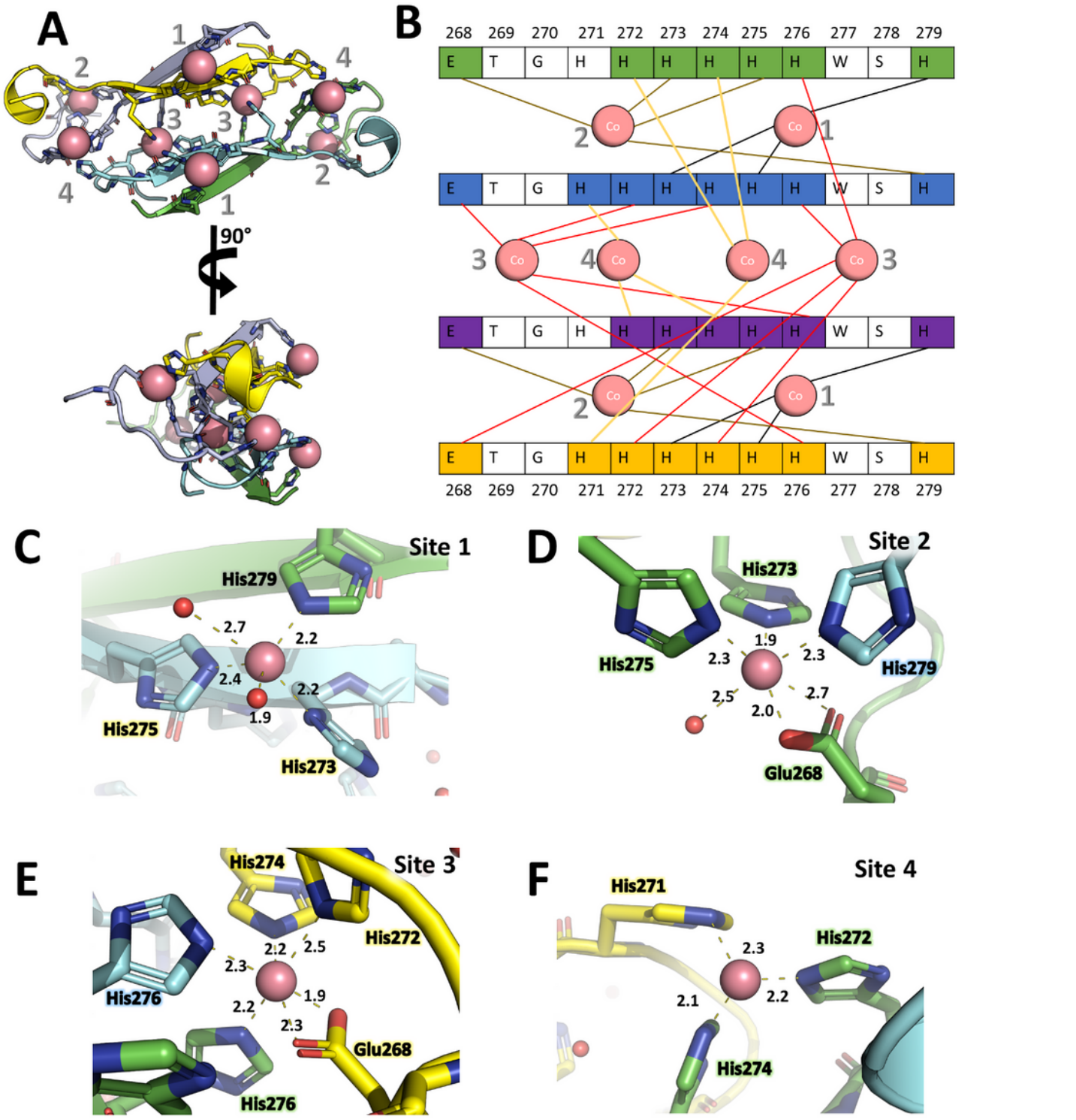

Figure 2

Architecture of the cobalt-bound tetrameric polyhistidine (TetrHis) motif. 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 figure 1. B. Schematic representation of the metal-protein contacts within the tetrameric motif. C, D, E and F. Close ups of the 4 cobalt binding sites showing the coordination geometries and distances.

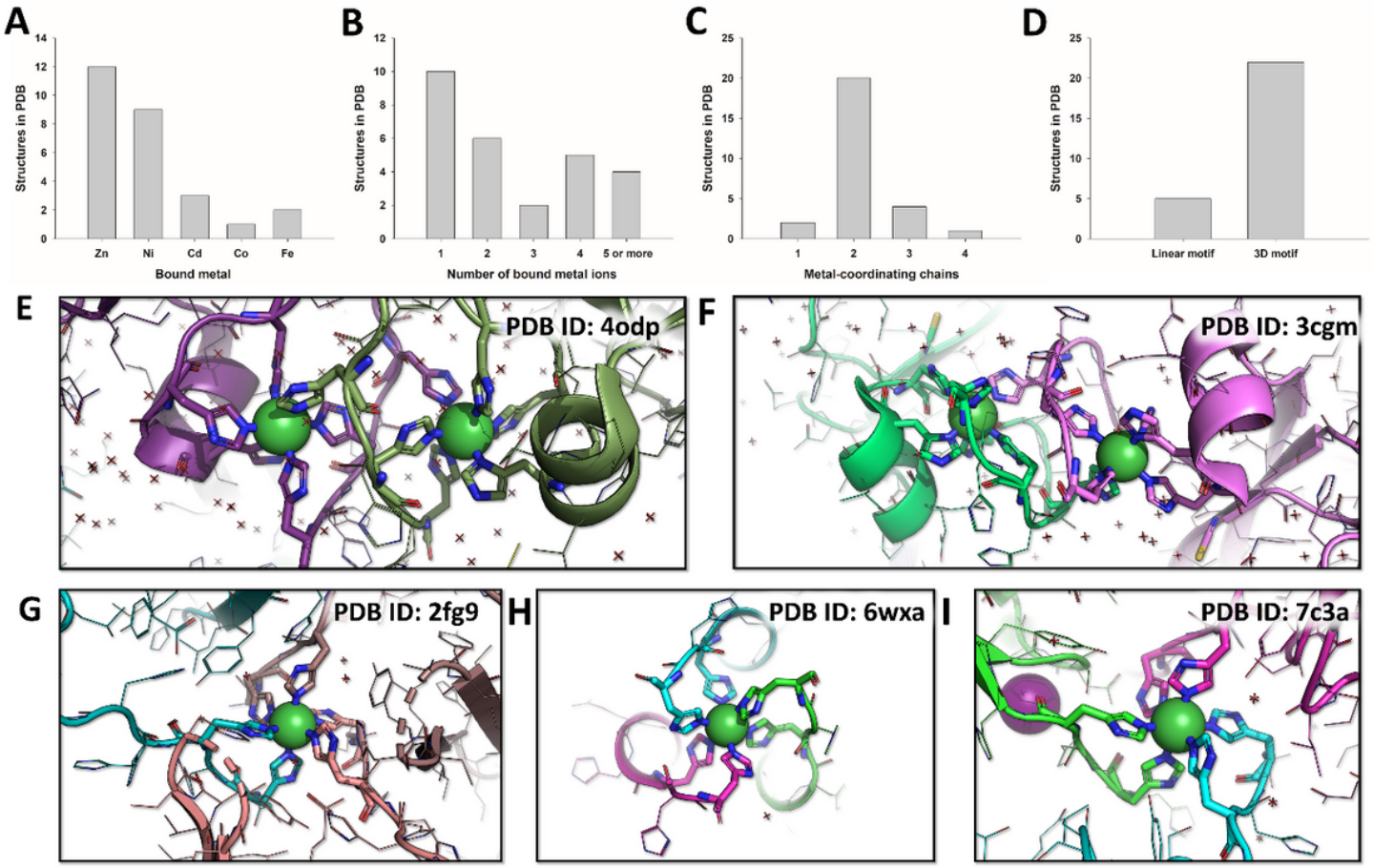

Figure 3

Survey of metal-bound polyhistidine tags in the PDB. A, B, C and 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 3-dimensional motif (with residues located far away within the protein sequence) (D). E and F. Nickel-dependent dimeric assemblies of SlyD from Thermus thermophiles, 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, H and I. Nickel-dependent trimeric assemblies observed in PDB entries 2fg9, 6wxa and 7c3a. In each case, 3 x 2 His residues from 6XHis tags assemble to form a octahedral coordination sphere around a single nickel ion.

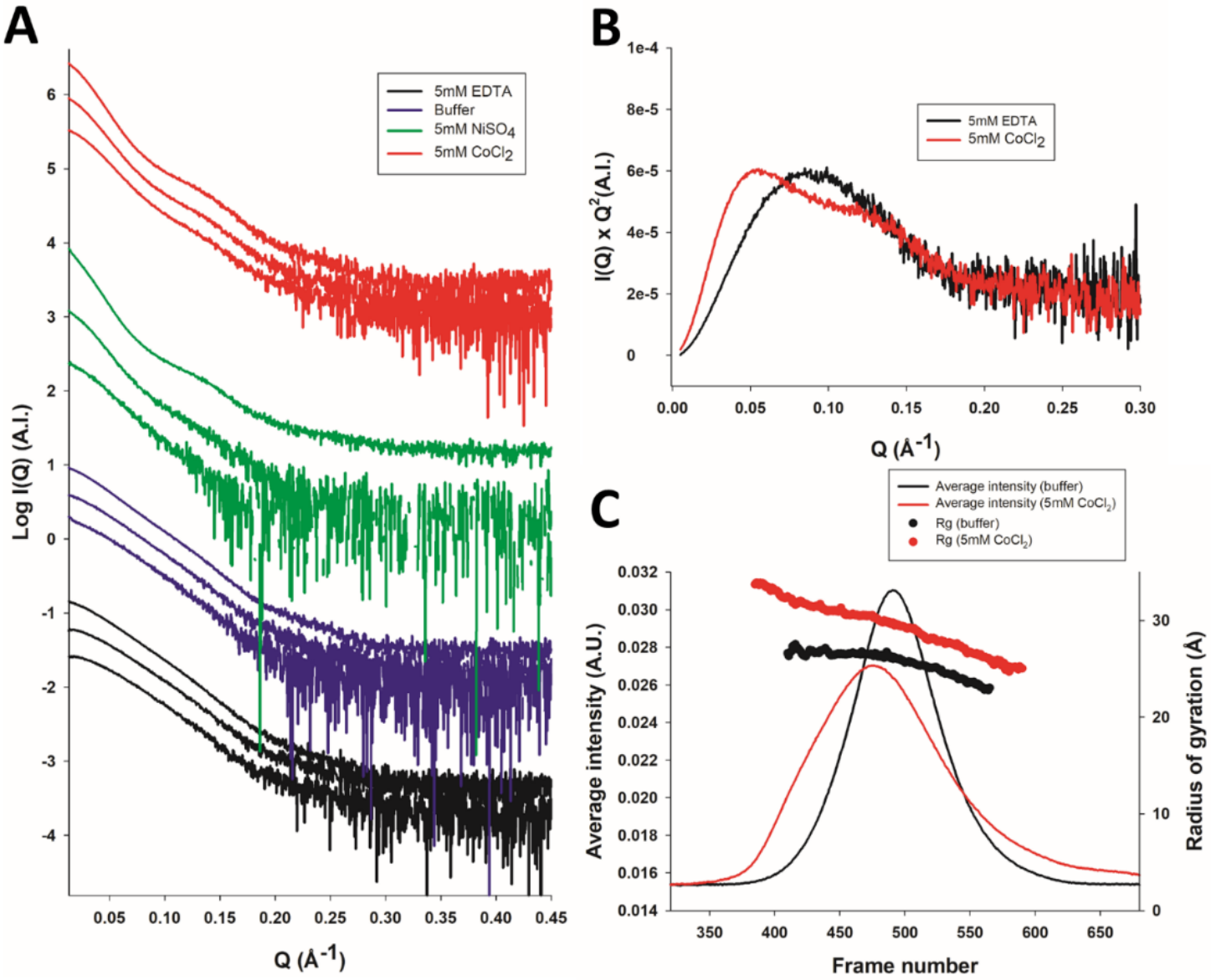

Figure 4

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 $NiSO_4$ or 5 mM $CoCl_2$ (green and red curves, respectively). The data were measured at increasing protein concentrations of 2.1, 4.2 and 8.5 mg/ml. B. Kratky plots the EDTA and $CoCl_2$ data using the 2.1 mg/ml SAXS profiles. C. SEC-SAXS profiles of scFv 2A2 in the presence (red) or absence of 5 mM $CoCl_2$ (black curve), and analysis of Radius of gyration versus frame number.

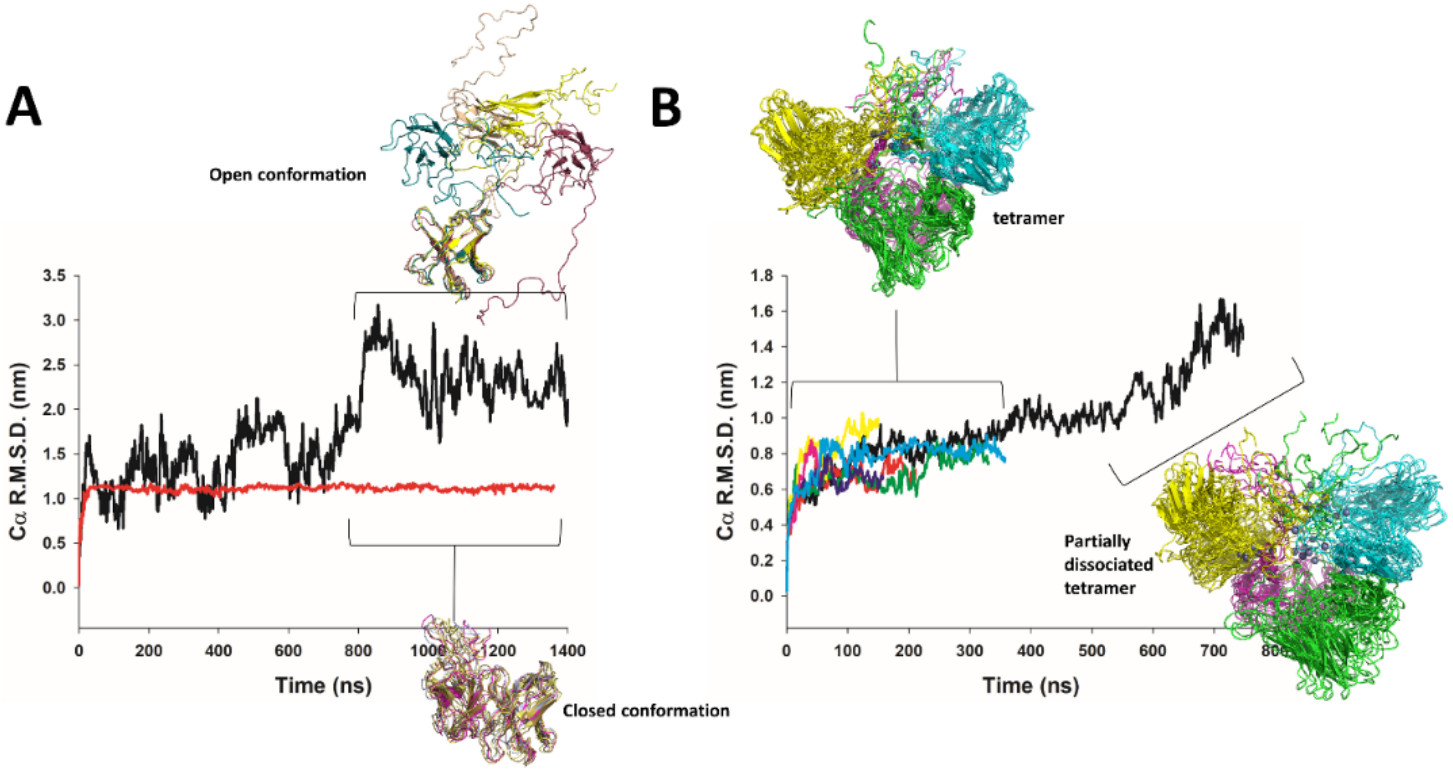

Figure 5

Molecular dynamics simulations analysis of scFv 2A2 in its monomeric and tetrameric forms. A. Cα root mean square deviation (R.M.S.D) as a function of simulation time for two independent MD trajectories of the scFv monomer. 5 representative conformers of the open and closed states sampled during the simulations are superimposed and shown as cartoon. B. A. Cα 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.

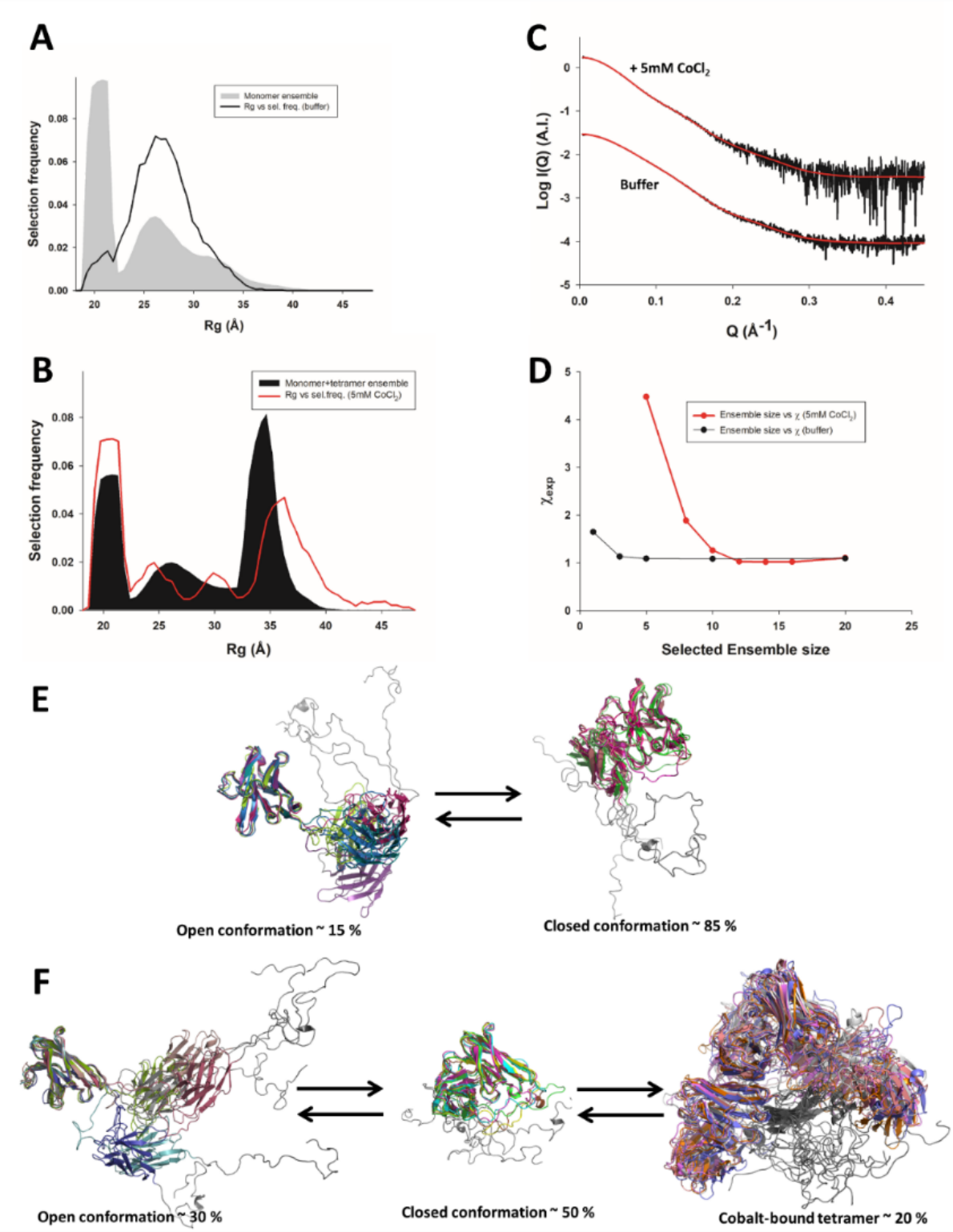

Figure 6

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. Radius of gyration distributions for the MD-generated ensemble of monomeric and tetrameric models (black area) and for the OE that fits the SEC-SAXS data (red curve) in the presence of 5 mM $CoCl_2$. C.

Fitted SAXS profiles of scFv 2A2 measured in the presence (top) or absence (bottom) of 5 mM $CoCl_2$. Experimental scattering curves are shown as black lines with OEs fits shown as red lines. D. EOM Goodness-of-fit χ_{exp} as a function of the number of models in the OEs, in the presence (red) or absence (black) of 5 mM $CoCl_2$. E. 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 grey. The percentage of open and closed models observed in the OEs is indicated. F. Conformational equilibrium of scFv 2A2 in the presence of 5 mM $CoCl_2$. 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 grey.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiguresandtables.docx