

Electronic Supplementary Information:
***Ab-initio* Electronic Structure Calculations of *Entire* Blue
Copper Azurins**

C. Romero-Muñiz,^{1,*} M. Ortega,¹ J. G. Vilhena,^{1,2} I. Díez-Pérez,^{3,4}

Juan Carlos Cuevas,^{1,5} Rubén Pérez,^{1,5,†} and Linda A. Zotti^{1,5,‡}

¹*Departamento de Física Teórica de la Materia Condensada,
Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

²*Department of Physics, University of Basel,
Klingelbergstrasse 82, CH-4056 Basel, Switzerland*

³*Department of Chemistry, Faculty of Natural & Mathematical Sciences, King's College
London, Britannia House, 7 Trinity Street, London SE1 1DB, United Kingdom.*

⁴*Institute of Theoretical and Computational
Chemistry (IQTC), University of Barcelona, Spai*

⁵*Condensed Matter Physics Center (IFIMAC),
Universidad Autónoma de Madrid, E-28049 Madrid, Spain*

* carlos.romero@uam.es

† ruben.perez@uam.es

‡ linda.zotti@uam.es

CONTENTS

S1. Analysis of MD simulations	S3
S2. MD frame selection criteria	S9
S3. Protonation state of the protein	S11
S4. Additional information on DFT results	S16
A. Comparison between single- ζ and double- ζ basis sets	S16
B. PDOS	S17
C. New version of Fig. 3	S17
D. Orbital localization	S17
E. Total DOS	S17
References	S25

S1. ANALYSIS OF MD SIMULATIONS

In Fig S1 we show the initial configurations for all the Azurin structures (a-e). Here is clearly observed that the three mutated residues (K41, L120, S89) are not at the same distance from the Cu coordination sphere, but they are all in the vicinity of the copper atom. We have analyzed the dynamics of the holo, apo and K41C structures during 500 ns of NVT simulations, and the dynamics of the L120C and S89C structures during 300 ns. Looking at the evolution of the total energy, see Fig S2a, we realize that it is stable during the last 50 ns of simulation for all the structures here considered, which is consistent with thermally equilibrated structures. This is further confirmed by the evolution of the radius of gyration and the root-mean-square deviation (R_g and RMSD respectively), see Fig S2b-c. In the five MD simulations, the values of the RMSD and R_g barely change during the last 100 ns of simulation. This indicates that the azurin is a relatively stiff molecule in contrast to the most abundant plasma proteins such as IgG and BSA. [1-3]

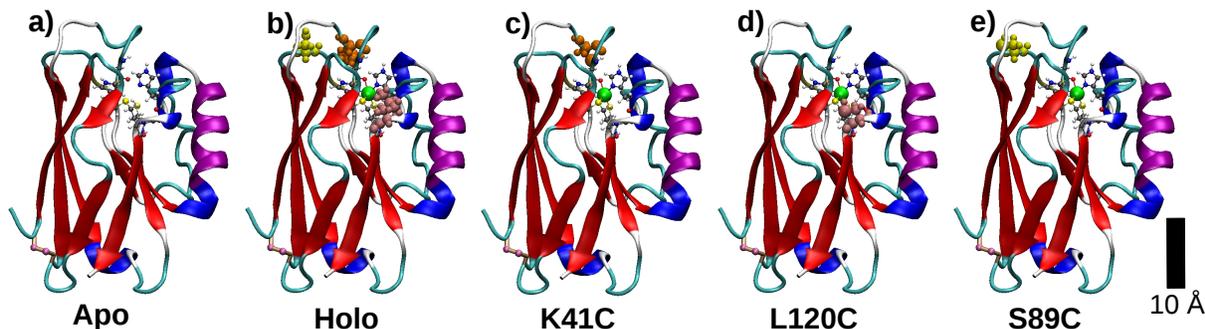


FIG. S1. Initial configurations of the (a) Apo, (b) Holo, (c) K41C mutant, (d) L120C mutant and (e) S89C mutant structures. The Azurin is represented with its secondary structure: beta-sheets (red), alpha-helix (purple), 310-helix (dark-blue), turns (cyan), and random-coils (white). The copper atom is shown using its van-der-Waals representation in an opaque green color, and its coordination residues are represented with a ball-and-stick model. The disulfide bridge is shown in a bond representation colored in light orange, and the main chain of the two cysteines which formed it is colored in pink. The positions of the residues in which the mutations have been performed (K41, S89 and L120) have been marked with a van-der-Waals representation in an opaque color (orange, yellow and pink, respectively)

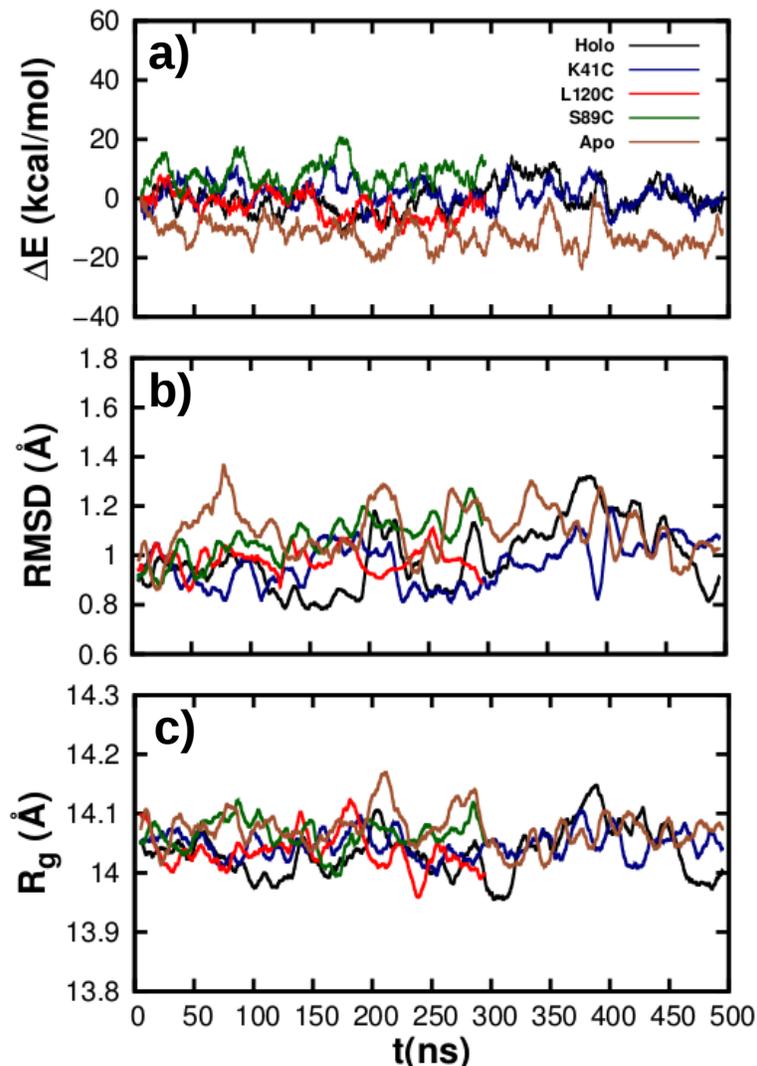


FIG. S2. Evolution of the total energy change (top), the root-mean-square deviation, RMSD, (middle) and the radius of gyration, R_g , (bottom) for the holo (black), K41C mutant (blue), L120C mutant (red), S89C mutant (green) and apo (brown) structures during the simulation of its free dynamic. The average of the total energy in the first 10 ns has been taken as our energy reference. RMSD has been calculated for all the backbone atoms of the protein. The reference configuration used for computing the RMSD evolution is the crystallographic azurin structure. The L120C and S89C results go only until 300 ns as these simulations are shorter than the other three.

Once the convergence of the structure is confirmed, we have used the last 50 ns of each simulation (25000 frames) to compute an averaged configuration for each Azurin structure. In Fig S3a-c we compare the averaged configuration of the holo structure with the averaged configurations of the three mutated structures. Looking at this figure, it can be observed that

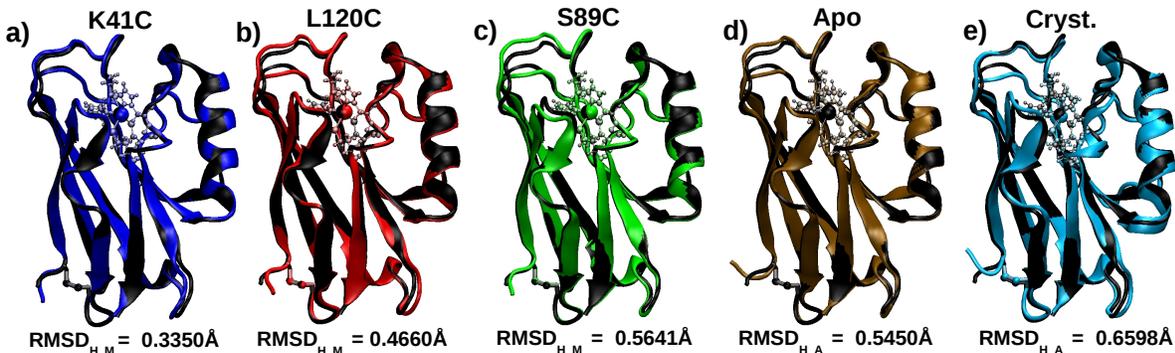


FIG. S3. Comparison between the averaged configuration of the holo structure (black) and the averaged configuration of: (a) the K41C structure (blue), (b) the L120C structure (red), (c) the S89C structure (green) and (d) the apo structure (brown). The comparison of the averaged configuration of the holo structure with the crystallographic structure (cyan) is also shown in (e). In all the cases, the two structures compared have been aligned with the program Visual Molecular Dynamics, and the value of the root-mean-square deviation (RMSD) between both structures is shown. The averaged configurations have been obtained by averaging the position of each atom in the last 50 ns of simulation. The azurin is represented with its secondary structure. The copper atom is shown in its Van-der-Waals representation and its coordination residues with a ball-stick model. The disulfide bridge is also shown in a bond representation.

the three mutant structures are practically superimposed to the holo structure. Moreover, the root-mean-square deviation between the holo averaged configuration and each mutant averaged configuration (RMSD_{H-M}) is smaller than 0.6 Å. Therefore, from Fig S3 we conclude that the mutations do not affect the overall protein structure. This agrees with the final values of the RMSD and the R_g observed in Fig S2, which are practically identical for the five Azurin structures. Concerning the apo, see Fig S3d, the same arguments exposed before for the mutant structures apply, i.e. one barely notices any structural change with respect to the holo structure.

To assure that the five azurin structures obtained (holo, apo, K41C mutant, L120C mutant and S89C mutant) are compatible with the crystal structure of the protein, we have also compared the azurin crystallographic structure with the holo averaged structure (see Fig S3e). In this figure, we observe that the holo averaged structure is nearly superim-

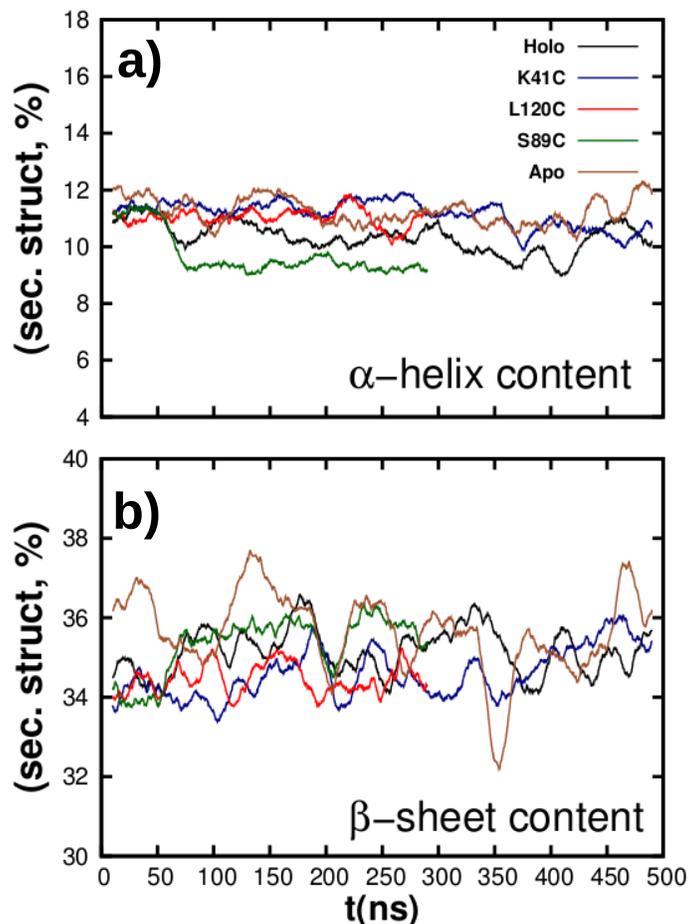


FIG. S4. Time evolution of the percentage of α -helix and β -sheet content for the holo (black), K41C mutant (blue), L120C mutant (red), S89C mutant (green) and apo (brown) structures. As in Fig S2 the L120C and S89C results go only until 300 ns as these simulations are shorter than the other three.

posed to the crystallographic structure, which indicates that both are practically identical. Moreover, that conclusion can be extended to the azurin mutant structures obtained, as the structural differences between averaged mutant structures and the holo one are negligible ($\text{RMSD}_{\text{H-M}} < 0.6 \text{ \AA}$). This also agrees with the RMSD evolution from the crystal structure for the five azurin structures, see Fig. S2, as all of them are smaller than 1.4 \AA during the whole MD simulation.

We have also analyzed the evolution of the azurin secondary structure in the five MD simulations, see Fig S4. The final values obtained for the holo structure are in agreement with

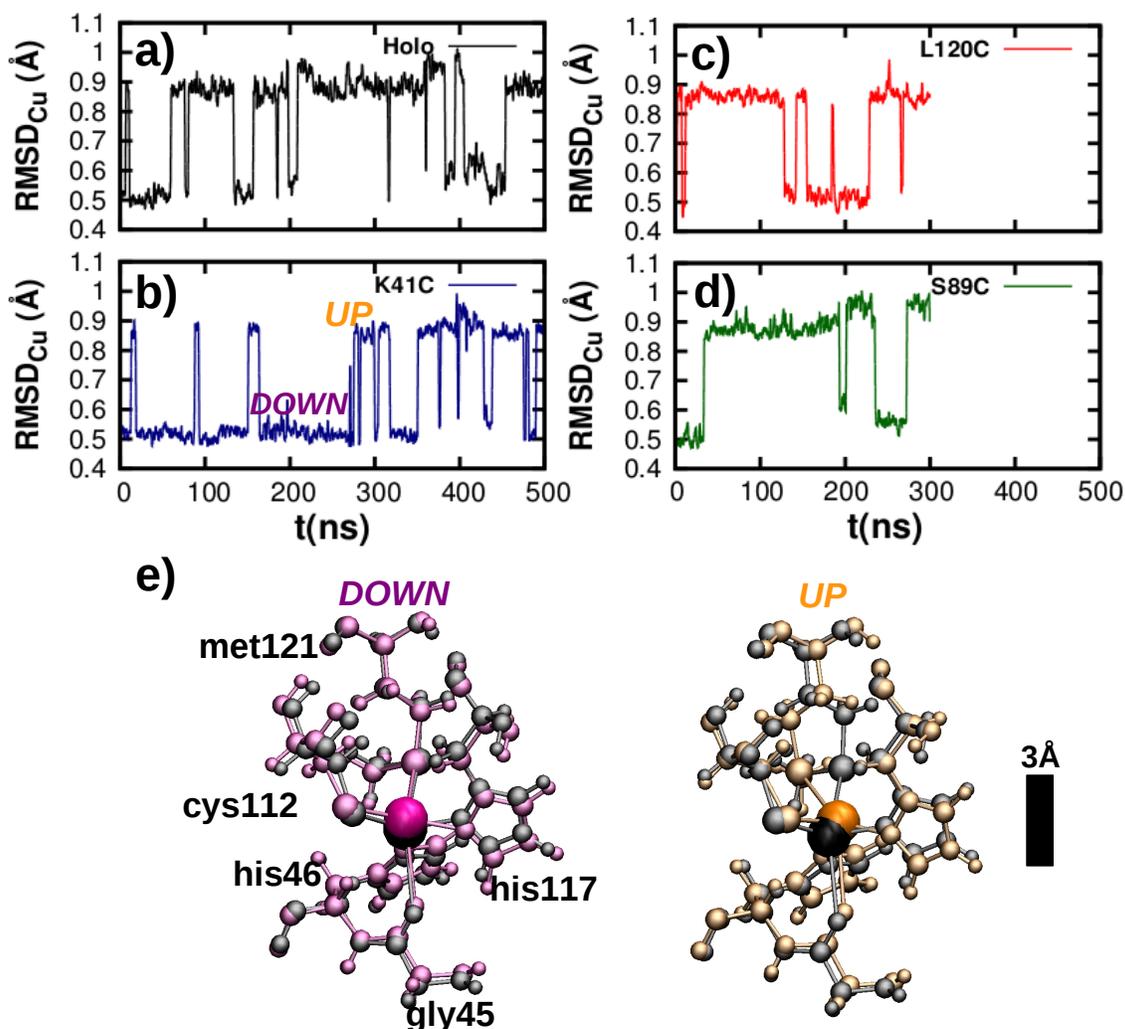


FIG. S5. RMSD evolution of the position of the copper coordination residues ($RMSD_{Cu}$) during the MD simulation for the (a) Holo (black), (b) K41C mutant (blue), (c) L120C mutant (red) and (d) S89C mutant (green) structures. The L120C and S89C evolution go only until 300 ns as these simulations are shorter than the other two. For the four azurin structures, we observe two distinct RMSD values during the whole simulation: UP ($RMSD_{Cu} \sim 0.9 \text{ \AA}$) and DOWN ($RMSD_{Cu} \sim 0.5 \text{ \AA}$). The two corresponding atom configuration of that two RMSD values are shown in (e). The UP configuration is represented in orange, the DOWN configuration in purple and the reference configuration (crystallographic azurin structure) in black color.

previously reported results [4, 5], i. e. α -helices $\sim 10\%$ and β -sheets $\sim 35\%$. Concerning the mutated structures, the difference between the final values obtained for these structures and the holo one is smaller than the changes induced by the thermal fluctuations during the

holo dynamics ($\sim 2\%$). This is also observed when comparing the apo and holo secondary structure content evolution. Therefore, we can conclude that the azurin secondary structure barely changes by introducing single-aminoacid mutations on its structure or without the copper atom. This agrees with the subtle differences observed in Fig. S3 between the averaged configurations of the holo and mutated structures

Finally, we have analyzed the evolution during the five MD simulations of the RMSD of the copper coordination sphere, i.e the residues met121, cys112, his46, gly45, his117 and Cu129. In Fig. S5a-d, we show the results obtained for the holo, K41C mutant, L120C mutant and S89C mutant structures. From that figure we can observe that for all the azurin structures, the RMSD of the copper coordination sphere (RMSD_{Cu}) essentially takes only two values: $\text{RMSD}_{\text{UP}} \sim 0.9 \text{ \AA}$ and $\text{RMSD}_{\text{DOWN}} \sim 0.4 \text{ \AA}$. To understand the origin of these two values, we have analyzed the configuration of the copper coordination sphere for both RMSD values in the Holo structure, see Fig. S5e. When $\text{RMSD}_{\text{Cu}} = \text{RMSD}_{\text{DOWN}}$, the position of the atoms is practically identical to the position of the atoms in the reference structure, i.e. the crystallographic structure (in black). However, when $\text{RMSD}_{\text{Cu}} = \text{RMSD}_{\text{UP}}$, we observe a substantial difference with the reference structure, which is that the met121 residue is nearer to the cys112. This approach changes the value of the RMSD_{Cu} but does not significantly change the distance between the copper atom and the S atom of the methionine (see Fig. S5e).

S2. MD FRAME SELECTION CRITERIA

In Fig. S6 we show both the Azurin root-mean-square deviation of all the backbone atoms (RMSD) and of the copper coordination residues ($RMSD_{Cu}$) for the holo and K41C Mutant structures. In this figure, it is shown again that the amplitude of the RMSD oscillation during the MD simulations is very small (less than 0.6 Å), indicating that the azurin is a relatively stiff molecule and therefore subjected to little structural changes during the

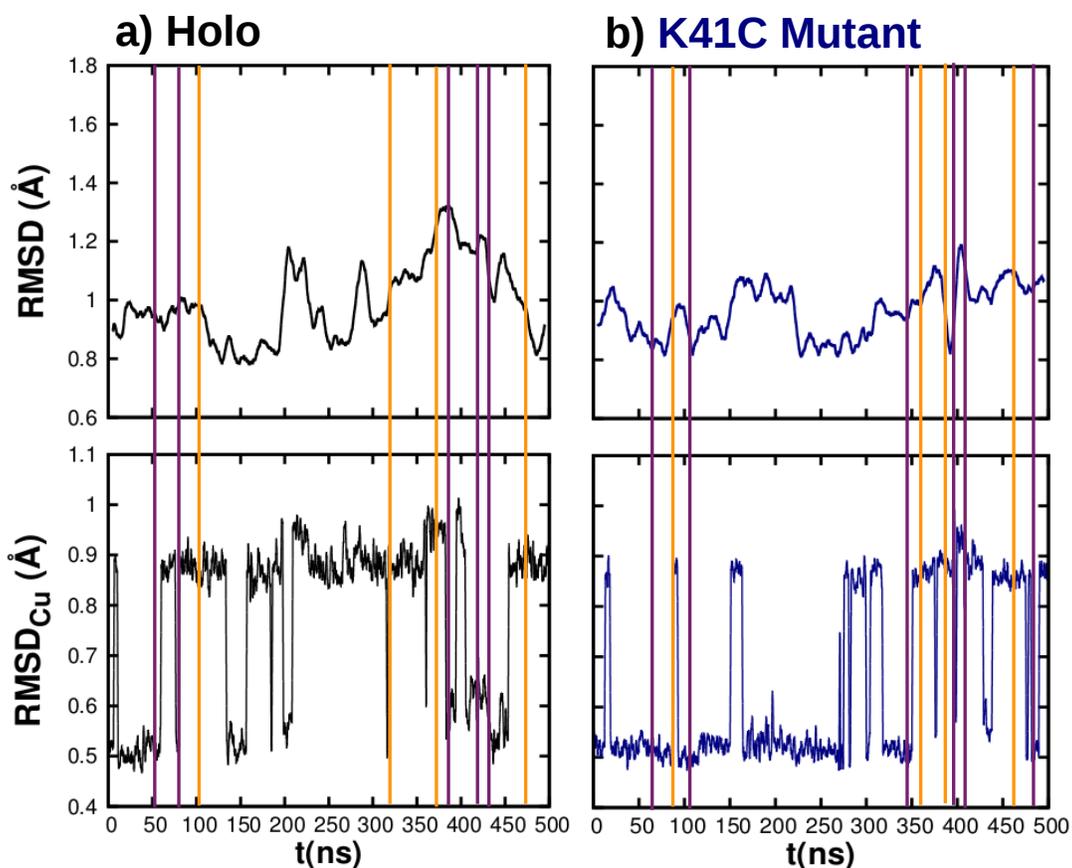


FIG. S6. Structural characteristics of the frames selected for computing the electronic properties of the (a) holo (black) and (b) K41C mutant (blue) Cu-Azurin structures. Both the evolution of the root-mean-square deviation of the all the backbone atoms of the protein (RMSD, top row) and of the position of the copper coordination residues ($RMSD_{Cu}$, bottom row) are represented. The reference configuration used for computing both RMSD evolutions is the crystallographic azurin structure. The frames selected are marked with purple/orange straight lines as they correspond to a DOWN/UP configuration (see Fig. S5)

whole dynamics. Therefore, we think that using ~ 10 frames for computing the electronic structure of both the holo and mutated azurin configurations, we are considering all the relevant azurin structural variations due to thermal fluctuations.

The chosen frames are shown in Fig. S6 with straight lines. As can be observed, we have sampled configurations that correspond to the extreme situations of the RMSD (holo:386.2 ns ; K41C:408.2 ns) to see possible differences in the electronic structure of the protein. We have selected also configurations corresponding to the average value of the RMSD (holo:323.8 ns ; K41C:363.4 ns). Moreover, we both have chosen frames with $\text{RMSD}_{\text{Cu}} = \text{RMSD}_{\text{UP}}$ (colored in orange) and with $\text{RMSD}_{\text{Cu}} = \text{RMSD}_{\text{DOWN}}$ (colored in purple), as we wanted to analyze if these two different configurations of the copper coordination sphere lead to different azurin electronic properties.

S3. PROTONATION STATE OF THE PROTEIN

The estimation of the protonation state of the different amino acids has been made by means of the H++ program [6] which estimates the pKa values of the different residues of the protein. The following residues, listed in Table S1 are susceptible of changing its charge state via protonation/deprotonation. The spatial localization of some of them is shown in Fig. S7. The probability of those residues of being protonated is not the same as if they were isolated and is given by their pKa values (collected also in Table S1) , which are defined as the pH for which that residue is 50 % occupied by a hydrogen. This means that if the pKa of a given residue is smaller than the pH, the most probable configuration of that group is the deprotonated state. According to these results we can assign the appropriate protonation state to all amino acids constituting the azurin (128) (see Table S2). Note that the tritrable amino acids are highlighted in red while the five residues forming the Cu complex are displayed in blue.

The usual charged amino acids (i.e. negative GLU/ASP and positive LYS/ARG) have the normal protonation state except ASP 23 which possesses one more H atom because it has pKa=4.7. There are two CYS residues (3 and 26) which are deprotonated, but they do not contribute to the net charge because they are bonded via a disulfide bond. There are other two HIS residues (35 and 83) which are not involved in the Cu complex bonding. His35 is protonated (positively charged) while His83 is not, in agreement with M. van den Bosch et al. [7]. In any case, the protonation state of His83 is not going to affect the structure of the protein as it is in the outer sphere of the protein (exposed to the solvent).

Finally, the five residues involved in the Cu complex deserve a particular comment. Since the Cu center is not taken into account during the calculation of the H++ program, we must ignore the results for this five amino acids given in Table S1. Both HIS residues (46 and 117) are deprotonated since they act as ligands (N-Cu). GLY45 and MET121 are both uncharged and CYS112 has negative charge since it is deprotonated, forming a thiolate ligand.

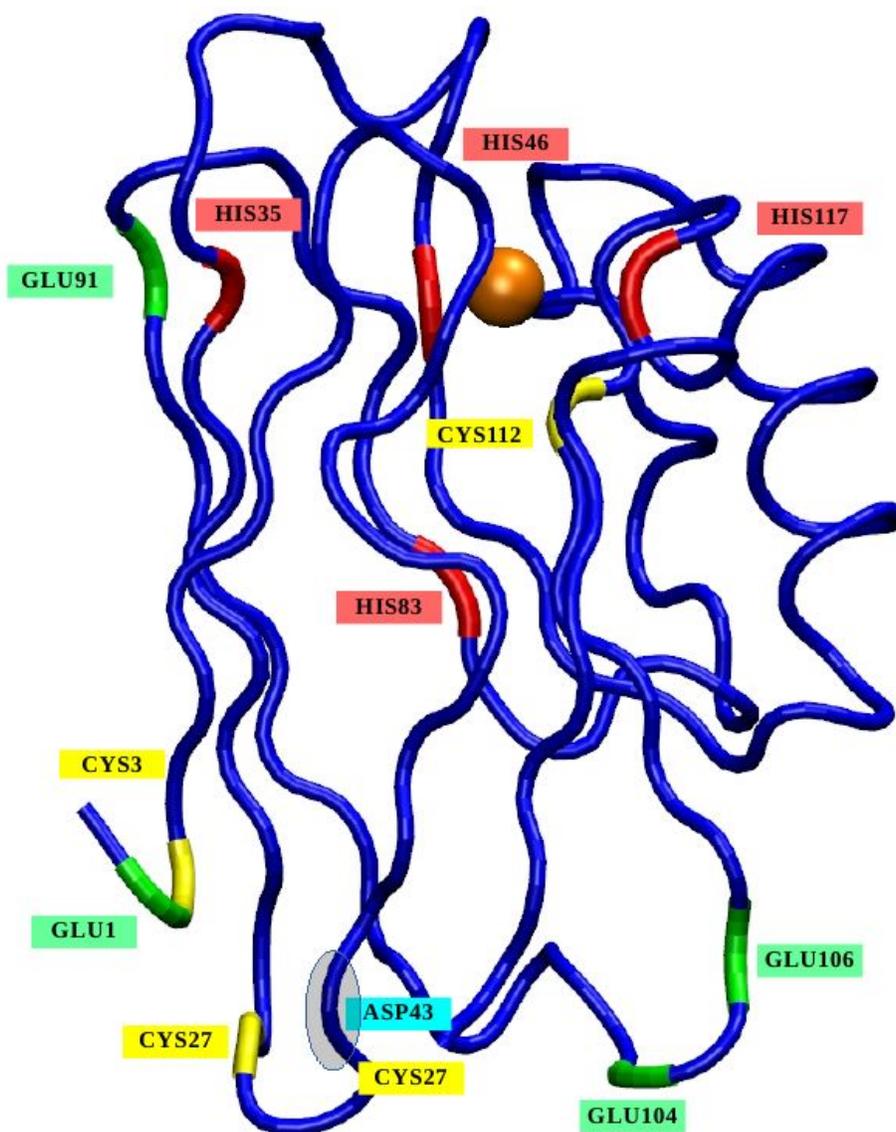


FIG. S7. Azurin tritable sites: the positions of all the histidines (HIS, in red), glutamic acids (GLU, in green), and cysteines (CYS, in yellow) are represented

Residue	pKa	Result
NTALA-1	7.713	Protonated form (positively charged)
GLU-2	4.291	Deprotonated form (negatively charged)
ASP-6	3.238	Deprotonated form (negatively charged)
ASP-11	1.793	Deprotonated form (negatively charged)
ASP-23	4.745	Protonated form (neutral)
LYS-24	11.688	Protonated form (positively charged)
LYS-27	10.721	Protonated form (positively charged)
HIS-35	9.593	Protonated form (positively charged)
LYS-41	11.459	Protonated form (positively charged)
HIS-46	< 0.000000	Deprotonated form (neutral)
ASP-55	3.099	Deprotonated form (negatively charged)
ASP-62	2.593	Deprotonated form (negatively charged)
ASP-69	3.378	Deprotonated form (negatively charged)
LYS-70	11.106	Protonated form (positively charged)
ASP-71	2.455	Deprotonated form (negatively charged)
TYR-72	11.599	Protonated form (neutral)
LYS-74	11.125	Protonated form (positively charged)
ASP-76	3.795	Deprotonated form (negatively charged)
ASP-77	0.493	Deprotonated form (negatively charged)
ARG-79	> 12.000000	Protonated form (positively charged)
HIS-83	7.783	Protonated form (positively charged)
LYS-85	11.713	Protonated form (positively charged)
GLU-91	3.709	Deprotonated form (negatively charged)
LYS-92	10.985	Protonated form (positively charged)
ASP-93	2.536	Deprotonated form (negatively charged)
ASP-98	2.746	Deprotonated form (negatively charged)
LYS-101	11.02	Protonated form (positively charged)
LYS-103	11.319	Protonated form (positively charged)
GLU-104	3.482	Deprotonated form (negatively charged)
GLU-106	4.068	Deprotonated form (negatively charged)
TYR-108	> 12.000000	Protonated form (neutral)
CYS-112	> 12.000000	Protonated form (neutral)
HIS-117	5.654	Protonated form (positively charged)
LYS-122	10.618	Protonated form (positively charged)
LYS-128	10.024	Protonated form (positively charged)
CTLYS-128	1.338	Deprotonated form (negatively charged)

TABLE S1. Protonation state of all the titratable residues of the holo-azurin structure at pH=4.5 according to its pKa value. The pKa values here shown have been computed with the H++ server [6]. Each residue is colored in agreement with its amino acid type (ALA = blue; GLU = green; ASP = cyan; LYS = orange; HIS = red; TYR = magenta; ARG = olive; CYS = black)

Number	Residue	H No.	Charge	Number	Residue	H No.	Charge
1	ALA	7	1	65	ALA	5	0
2	GLU	6	-1	66	SER	5	0
3	CYS	4	0	67	GLY	3	0
4	SER	5	0	68	LEU	11	0
5	VAL	9	0	69	ASP	4	-1
6	ASP	4	-1	70	LYS	13	1
7	ILE	11	0	71	ASP	4	-1
8	GLN	8	0	72	TYR	9	0
9	GLY	3	0	73	LEU	11	0
10	ASN	6	0	74	LYS	13	1
11	ASP	4	-1	75	PRO	7	0
12	GLN	8	0	76	ASP	4	-1
13	MET	9	0	77	ASP	4	-1
14	GLN	8	0	78	SER	5	0
15	PHE	9	0	79	ARG	13	1
16	ASN	6	0	80	VAL	9	0
17	THR	7	0	81	ILE	11	0
18	ASN	6	0	82	ALA	5	0
19	ALA	5	0	83	HIS	7	0
20	ILE	11	0	84	THR	7	0
21	THR	7	0	85	LYS	13	1
22	VAL	9	0	86	LEU	11	0
23	ASP	5	0	87	ILE	11	0
24	LYS	13	1	88	GLY	3	0
25	SER	5	0	89	SER	5	0
26	CYS	4	0	90	GLY	3	0
28	GLN	8	0	92	LYS	13	1
29	PHE	9	0	93	ASP	4	-1
30	THR	7	0	94	SER	5	0
31	VAL	9	0	95	VAL	9	0
32	ASN	6	0	96	THR	7	0
33	LEU	11	0	97	PHE	9	0
34	SER	5	0	98	ASP	4	-1
35	HIS	8	1	99	VAL	9	0
36	PRO	7	0	100	SER	5	0
37	GLY	3	0	101	LYS	13	1
38	ASN	6	0	102	LEU	11	0
39	LEU	11	0	103	LYS	13	1
40	PRO	7	0	104	GLU	6	-1
41	LYS	13	1	105	GLY	3	0
42	ASN	6	0	106	GLU	6	-1
43	VAL	9	0	107	GLN	8	0
44	MET	9	0	108	TYR	9	0

Table S2 (Continued on next page)

Continued from previous page

Number	Residue	H No.	Charge	Number	Residue	H No.	Charge
45	GLY	3	0	109	MET	9	0
46	HIS	7	0	110	PHE	9	0
47	ASN	6	0	111	PHE	9	0
48	TRP	10	0	112	CYS	4	-1
49	VAL	9	0	113	THR	7	0
50	LEU	11	0	114	PHE	9	0
51	SER	5	0	115	PRO	7	0
52	THR	7	0	116	GLY	3	0
53	ALA	5	0	117	HIS	7	0
54	ALA	5	0	118	SER	5	0
56	MET	9	0	120	LEU	11	0
57	GLN	8	0	121	MET	9	0
58	GLY	3	0	122	LYS	13	1
59	VAL	9	0	123	GLY	3	0
60	VAL	9	0	124	THR	7	0
61	THR	7	0	125	LEU	11	0
62	ASP	4	-1	126	THR	7	0
63	GLY	3	0	127	LEU	11	0
64	MET	9	0	128	LYS	13	0

TABLE S2: Protonation state of all holo-azurin residues in our MD simulations. The tritable residues are marked in red

S4. ADDITIONAL INFORMATION ON DFT RESULTS

A. Comparison between single- ζ and double- ζ basis sets

In order to ensure the reliability of the basis sets used in this work (see methods section in the main text for details) we have performed the same calculation using a double- ζ basis set for all atoms in the protein. Fig. S8 shows the PDOS calculated with the two basis sets for different relevant atoms: Cu-center, some of the surrounding atoms of the first coordination sphere and the oxygen atoms belonging to the carboxylate group of the residue ASP83 responsible for some electronic levels close to the HOMO. The comparison confirms that there are no significant differences in the electronic level structure of these atoms.

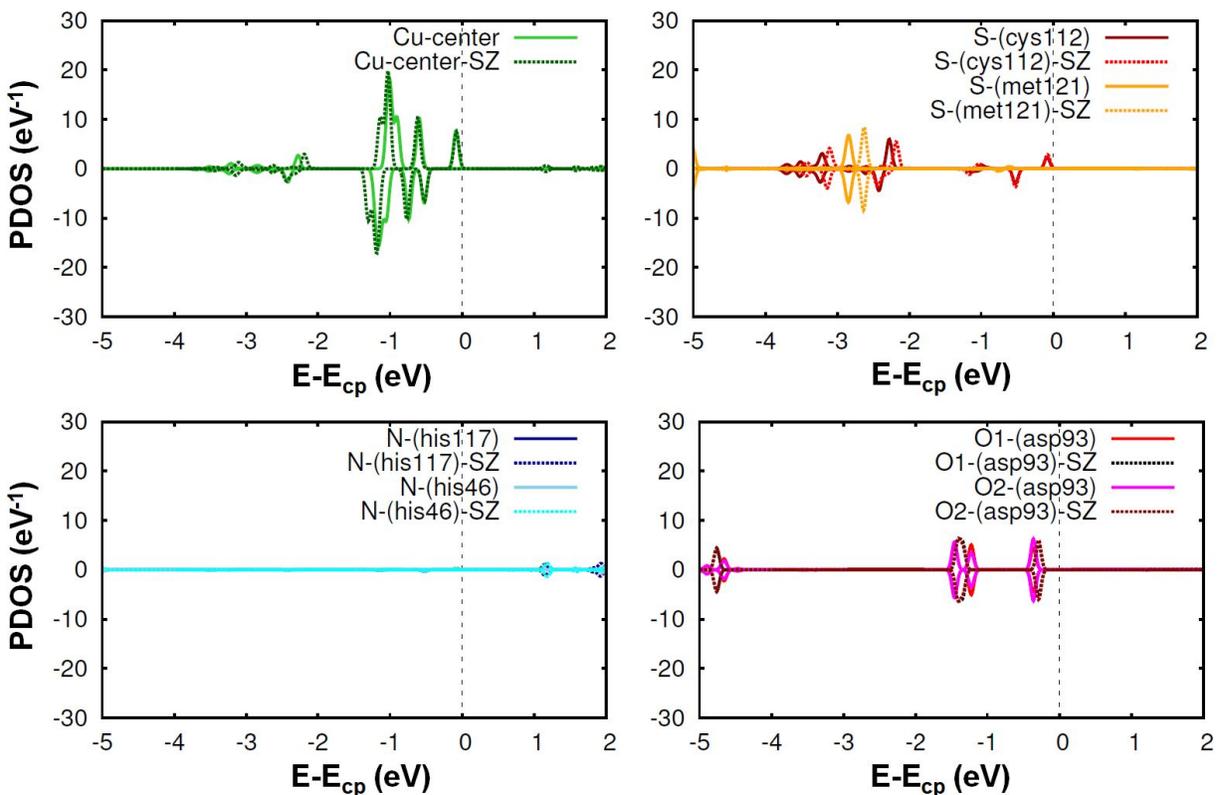


FIG. S8. Comparison of the PDOS for selected atoms (see insets) using mainly a single- ζ basis set for all atoms except the Cu-center and some oxygen atoms as specified in the Methods section (labeled as SZ) and a fully double- ζ calculation.

B. PDOS

In this section, we report the PDOS of the same atoms as those chosen for Fig. 2 of the main text, but this time for many more time frames extracted from the molecular dynamics trajectories. This section includes Figs. [S9](#), [S10](#), [S11](#) and [S12](#).

C. New version of Fig. 3

Fig. [S13](#) is an alternative version of Fig. 3 using ball-and-stick models for the whole protein in order to distinguish the exact locations of the different molecular orbitals.

D. Orbital localization

In this section, we show the spatial distribution of 14 different electronic levels below the HOMO orbital for the holo azurin. This section includes Fig. [S14](#).

E. Total DOS

In this section, the total DOS of different mutants of the azurin and the apo protein are compared. For this purpose, various graphs are presented where the DOS curves belonging to different pairs of proteins are compared. This section includes Fig. [S15](#).

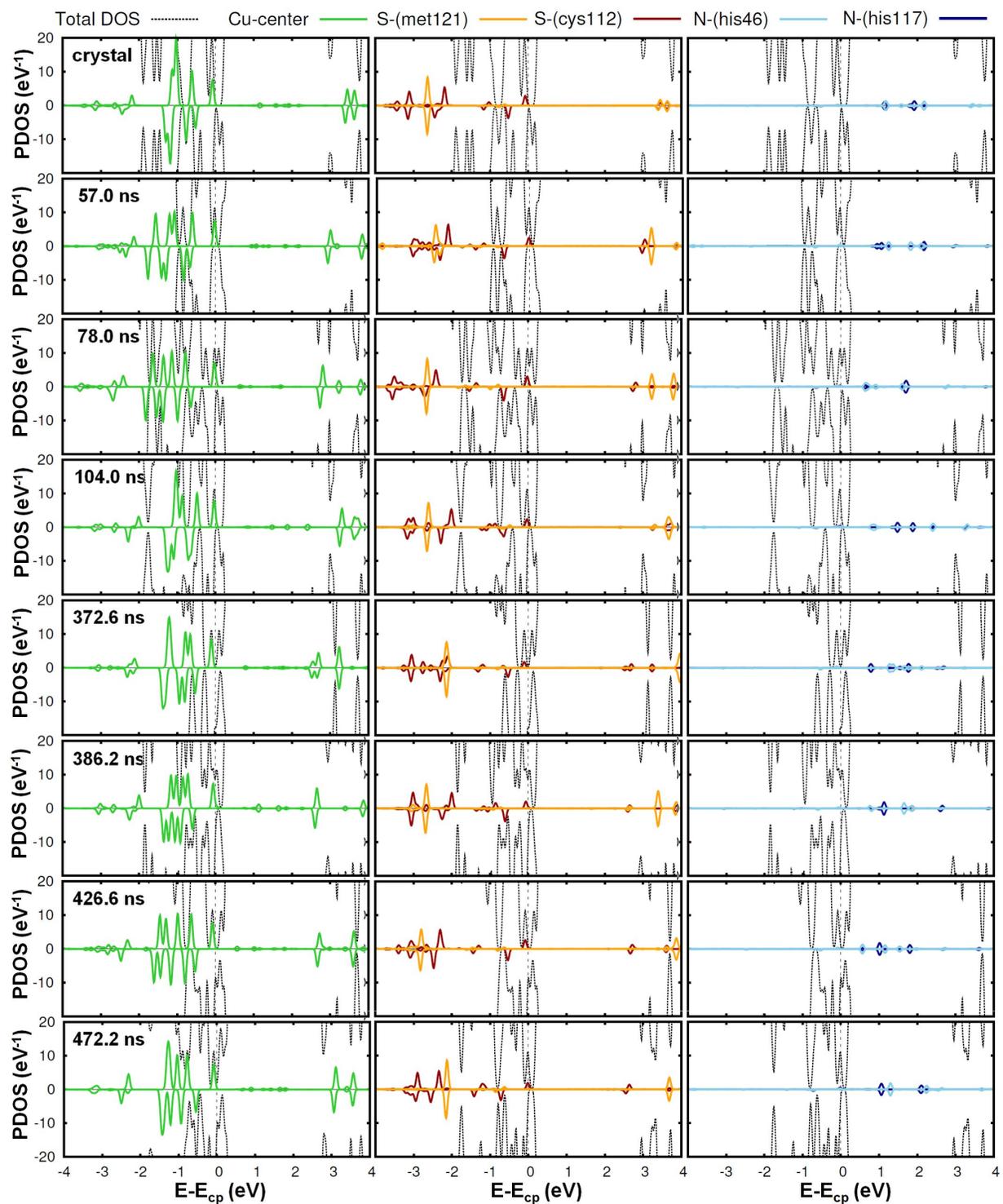


FIG. S9. Spin resolved projected density of states of the Cu center (left panel) and other atoms involved in the coordination bonds: S-met121, S-cys112, (middle panel) N-his46 and N-his117 (right panel) for different frames of the holo protein.

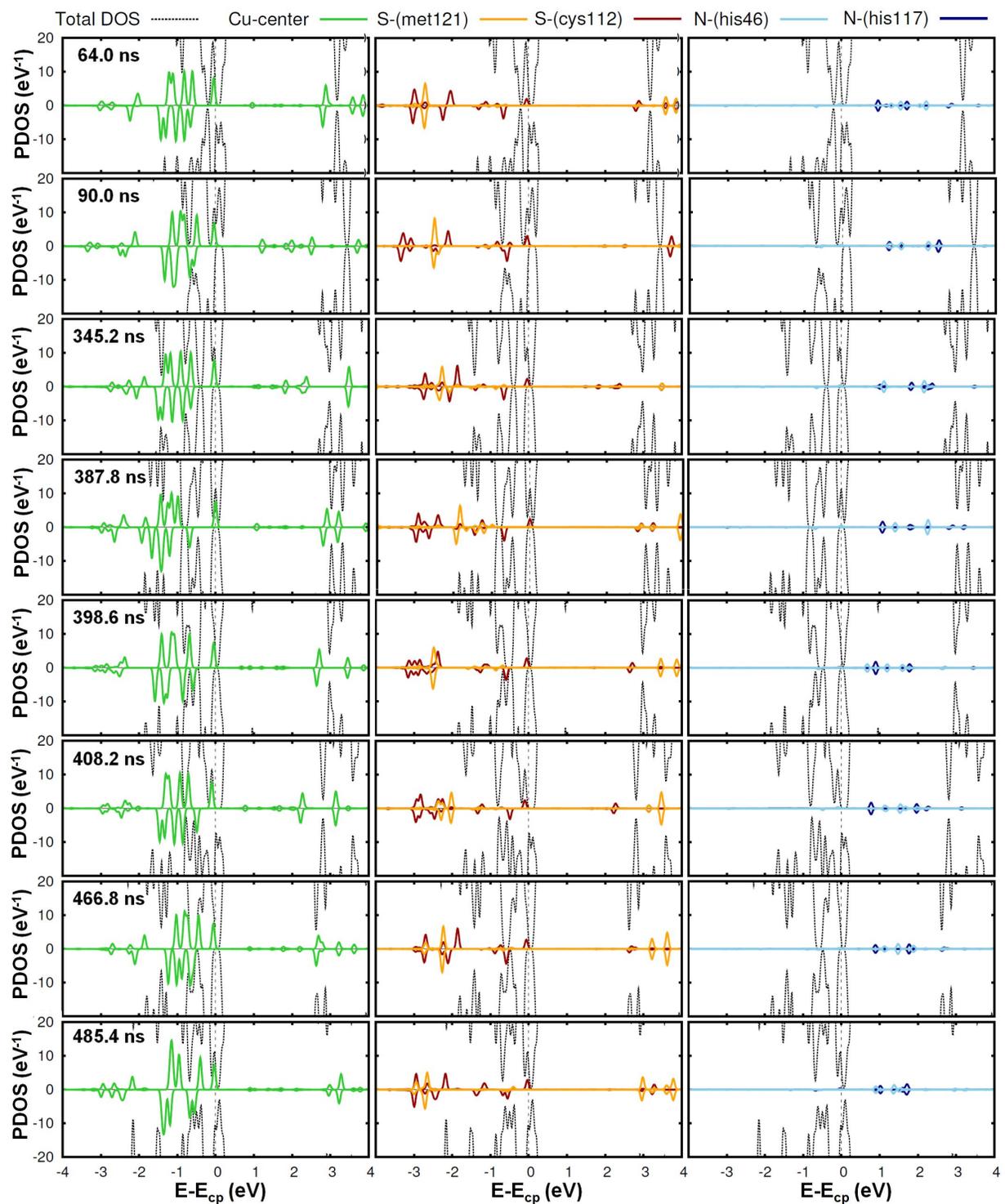


FIG. S10. Spin resolved projected density of states of the Cu center (left panel) and other atoms involved in the coordination bonds: S-met121, S-cys112, (middle panel) N-his46 and N-his117 (right panel) for different frames of the K41C azurin.

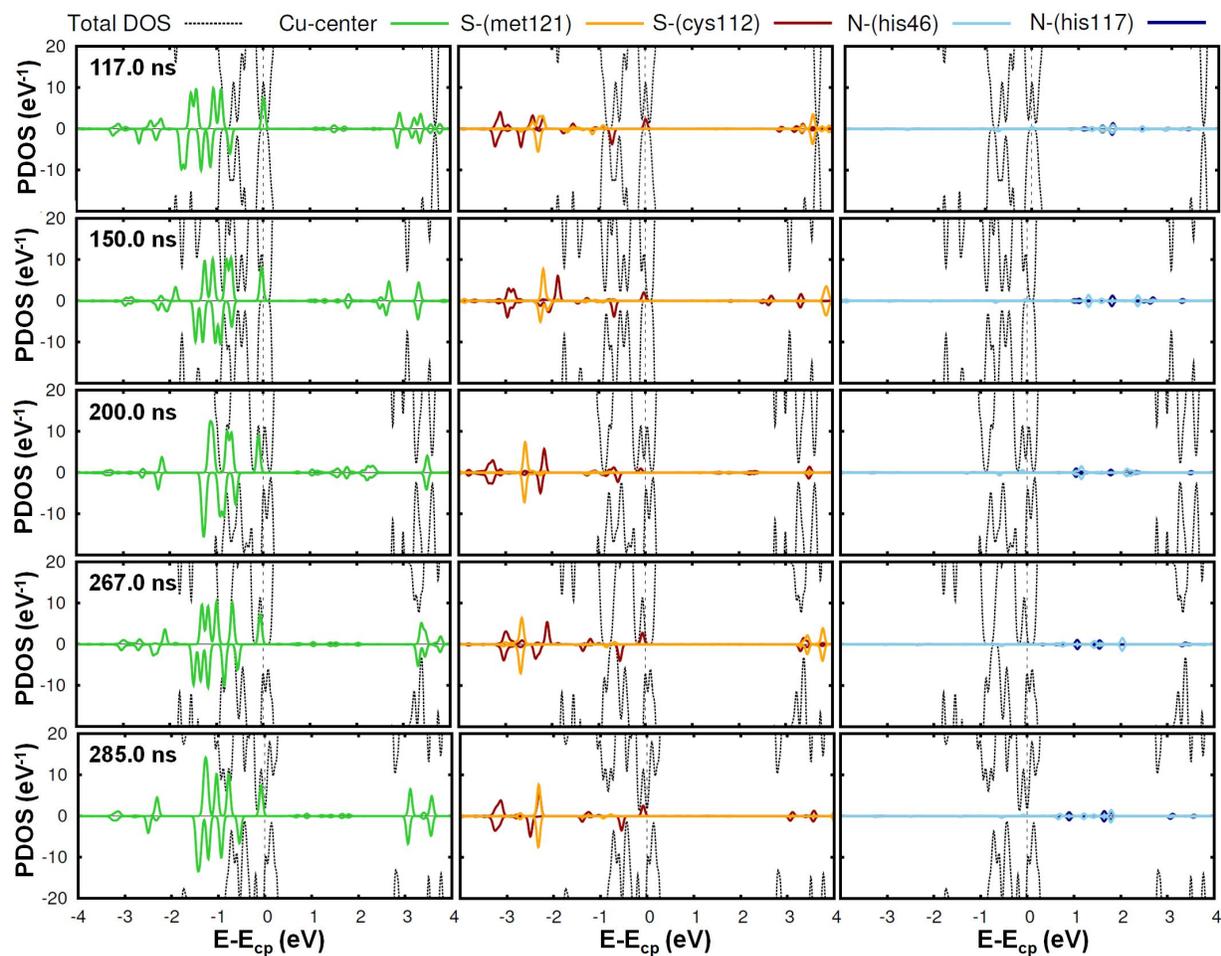


FIG. S11. Spin resolved projected density of states of the Cu center (left panel) and other atoms involved in the coordination bonds: S-met121, S-cys112, (middle panel) N-his46 and N-his117 (right panel) for different frames of the S89C azurin.

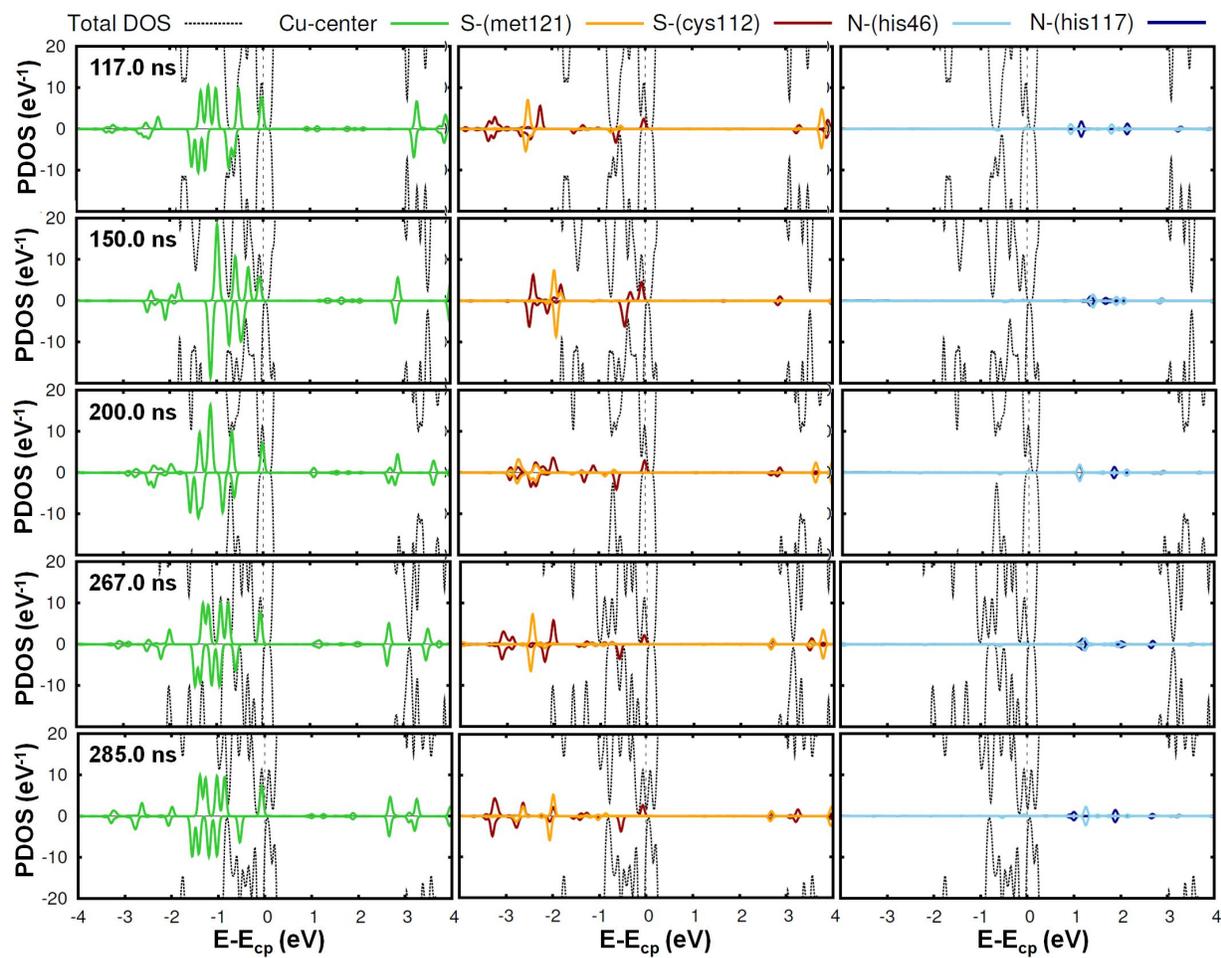


FIG. S12. Spin resolved projected density of states of the Cu center (left panel) and other atoms involved in the coordination bonds: S-met121, S-cys112, (middle panel) N-his46 and N-his117 (right panel) for different frames of the L120C azurin.

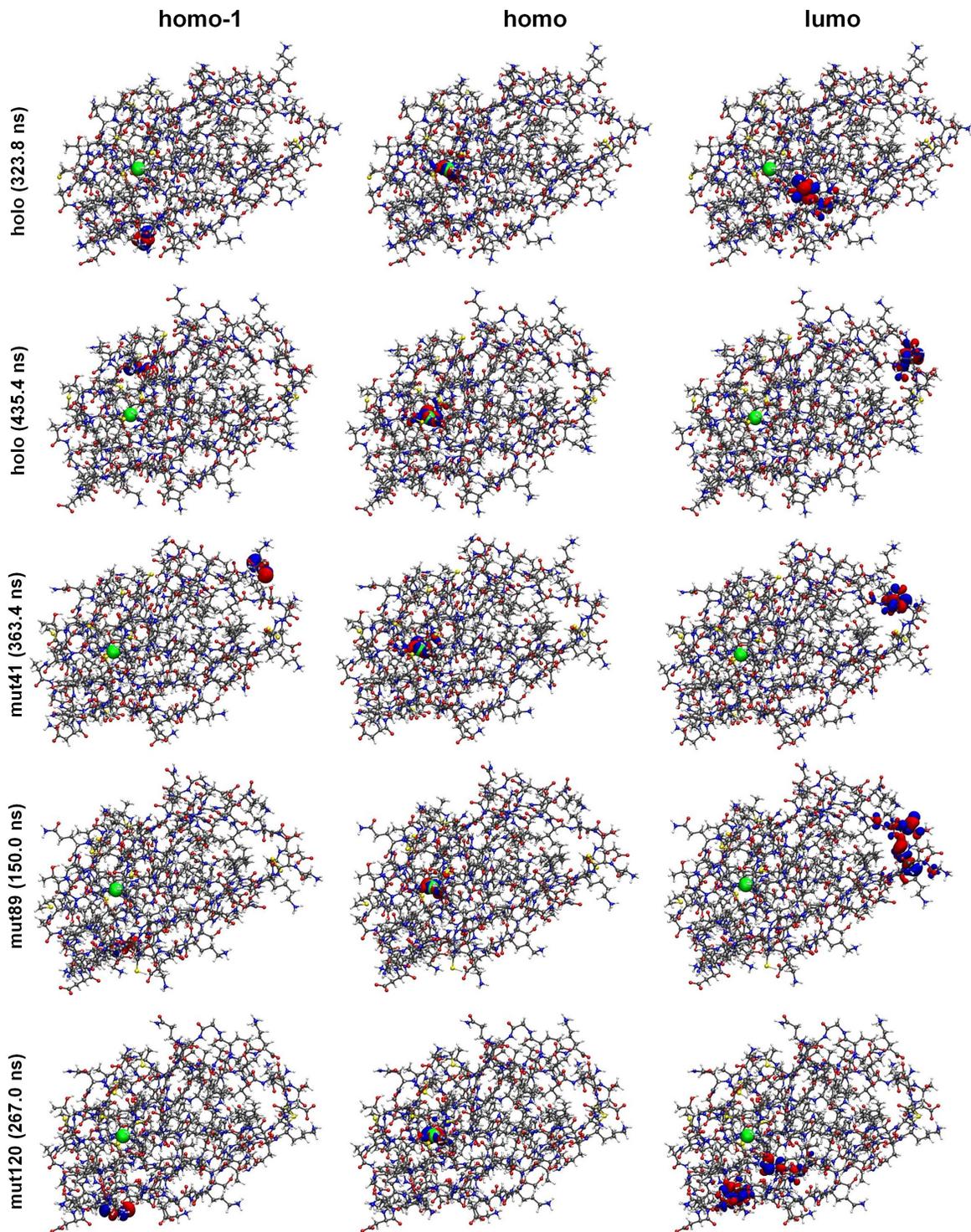


FIG. S13. Spatial distribution of HOMO-1, (left panel), HOMO (middle panel) and LUMO (right panel) orbitals for the same protein structures as in Fig. 2 of the main text. An isosurface value of ± 0.02 was adopted. This is a different version of the Fig. 3 of the main text using ball-and-stick models instead of the usual ribbon representation. This all-atom representation allows a better visualization of the individual atoms involved in each molecular orbital.

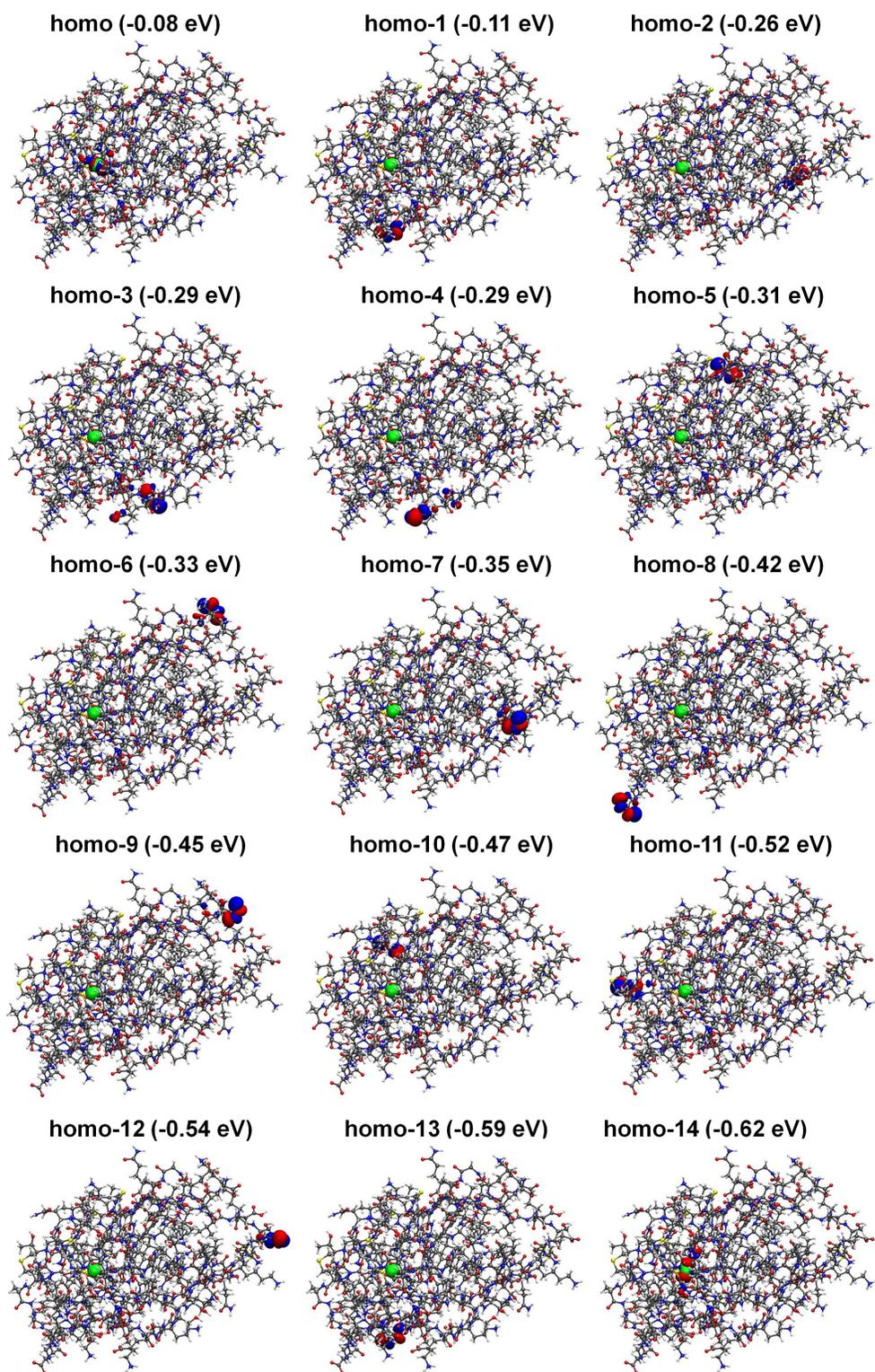


FIG. S14. Spatial distribution of selected occupied molecular orbitals in the holo azurin protein. All orbitals comprised between HOMO and HOMO-14 are shown, covering all electronic states between two consecutive peaks of Cu. Notice that these orbitals are all localized on oxygen atoms of carboxyl groups.

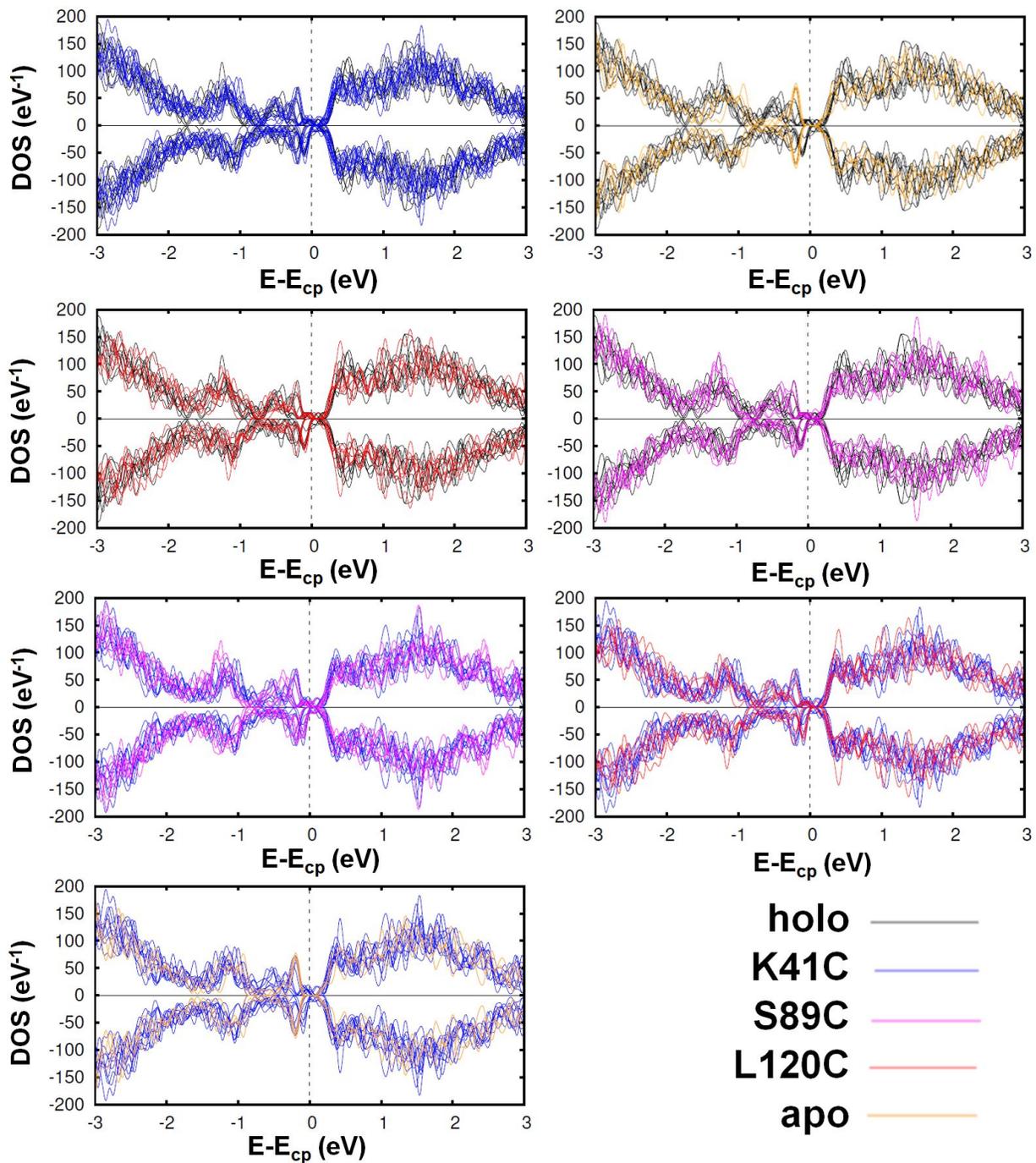


FIG. S15. Comparison between total DOS of different azurin proteins. In each panel, the DOS of two different proteins are reported. It can be observed that all curves have some main features in common and that the apo azurin is the one which differs the most (especially regarding the HOMO-LUMO gap). However, some minor discrepancies are also found between the holo and the other mutants, especially around $E - E_F \sim -2$ eV.

-
- [1] M. P. Ruiz, A. C. Aragonès, N. Camarero, J. G. Vilhena, M. Ortega, L. A. Zotti, R. Pérez, J. C. Cuevas, P. Gorostiza, and I. Díez-Pérez, *J. Am. Chem. Soc.* **139**, 15337 (2017).
- [2] J. G. Vilhena, A. C. Dumitru, E. T. Herruzo, J. I. Mendieta-Moreno, R. Garcia, P. A. Serena, and R. Pérez, *Nanoscale* **8**, 13463 (2016).
- [3] J. G. Vilhena, P. Rubio-Pereda, P. Velloso, P. A. Serena, and R. Pérez, *Langmuir* **32**, 1742 (2016).
- [4] J. W. Curtis, *Proteins: Struct. Funct. Bioinf.* **35**, 307 (1999).
- [5] A. R. Bizzarri, C. Baldacchini, and S. Cannistraro, *Langmuir* **33**, 9190 (2017).
- [6] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev, *Nucleic Acids Res.* **33**, W368 (2005).
- [7] M. van den Bosch, M. Swart, J. G. Snijders, H. J. C. Berendsen, A. E. Mark, C. Oostenbrink, W. F. van Gunsteren, and G. W. Canters, *ChemBioChem* **6**, 738 (2005).