SUPPORTING INFORMATION

Electrostatic Regulation of Blue Copper Sites

*Daniel Bím,[a] Anastassia N. Alexandrova[a,b,]**

[a]Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095-1569, USA; [b]California NanoSystems Institute, University of California, Los Angeles, 570 Westwood Plaza, Los Angeles, California 90095-1569, USA

Corresponding author:

*E-mail: ana@chem.ucla.edu

Supporting Information Contents

Computational Details

Local Electric field analysis

For the statistical analysis of local electric fields in the blue copper proteins, we used a snapshot of the RCSB protein data bank[1] acquired on October 21, 2020. We used Erebus PDB protein substructure search server[2] to identify the Cu sites with 2His, Cys, Met coordination. We then excluded proteins with 90% sequence identity using EMBOSS package,[3] and we removed multinuclear proteins having T1 site in contact with other metal (such as multi-copper oxidases). In the resulting 36 proteins (see **Table S1**), the missing hydrogen atoms were added using PyMOL Molecular Graphics System[4] and the Cu-ligating residues (2His, Cys, Met) were replaced with Ala for the scaffold field calculations.

For each of the proteins, we have assigned atomic charges to each atom at the 'EX-NPA_6-31Gd_PCM level' as implemented in the web-based Atomic Charge Calculator,[5] and we calculated the electric field at the position of Cu atom (with the Cu charge zeroed) using the Coulomb's law:

$$\vec{E} = \frac{1}{4\pi\varepsilon_0}\sum_i \frac{q_i \vec{r}_i}{r_i^2 r_i} \qquad (1)$$

Density functional theory calculations

For the DFT calculations, we have prepared a cluster model of the BC active site from the crystal structure of the oxidized poplar plastocyanin – PDB code: 4dp9 (**Figure S10**). All of the calculation were carried out using Turbomole 6.6 program.[6] The structures were optimized using TPSS functional[7], def2-TZVP basis set[8], and zero-damping dispersion correction (D3)[9]. The effect of solvation on geometry optimizations was included by employing the implicit conductor-like screening model (COSMO)[10] with a dielectric constant of $\varepsilon = 4$, and the default radii for all atoms (for Cu, we used radius of 2.2 Å). The calculations were accelerated by resolution-of-identity approximation (RI).[11] The geometry optimizations were followed by the single-point calculations at the RI-B3LYP/def2-TZVP(cosmo; $\varepsilon = 4$) level of theory. The effect of the oriented external electric field was evaluated as implemented in the Turbomole 6.6.

The absorption spectra were calculated using time-dependent (TD-)DFT using B3LYP/def2-TZVP level of theory at the equilibrium geometry from the two-dimensional potential energy surface scan from **Figure 3** in the main text.

Molecular dynamics simulations

Discrete molecular dynamics (DMD)[12] simulation method was utilized to rapidly sample conformations of various azurin variants (see **Note S3**). The models for DMD simulations were prepared from the wild-type azurin (PDB: 4azu) by: (i) removing chains B-D; keeping only

monomeric unit, (ii) removing the solvent molecules (note that DMD is using implicit solvation model), (iii) introducing target mutations, (iv) protonating all of the residues. During the simulations, the ligating atoms ($2N_{His}$,S_{Cys},S_{Met}) and Cu were kept fixed at the azurin crystal structure positions. Otherwise, the sampling of the protein structure would require much more accurate treatment of the Cu active site (e.g., QM/MM MD approach). For each protein variant, we have performed 3 replicate DMD simulations consisting of 400 000 time steps (roughly corresponding to 20 ns). Along each trajectory, we saved every 10 000th structure that was subjected to the local electric field analysis (*vide supra*). The $\boldsymbol{F}$(Cu-S_{Cys}) projection of the local electric field at the Cu site, plotted in **Figure S7**, is the average of the 3 replicate DMD trajectories.

Table S1. The local electric fields (LEFs) and the $\boldsymbol{F}$(Cu-S_{Met}), $\boldsymbol{F}$(Cu-S_{Cys}), and $\boldsymbol{F}$(Cu-N_{His}) projections of the LEFs discovered in the series of BC proteins.

PDB code	Protein	Total LEF	$\boldsymbol{F}$(Cu-S_{Met})	$\boldsymbol{F}$(Cu-S_{Cys})	$\boldsymbol{F}$(Cu-N_{His})[a]	d(Cu-S_{Met})	d(Cu-S_{Cys})	τ
1a3z	rusticyanin	14.0	9.7	–10.0	3.2	2.75	2.25	0.81
1aac	amicyanin	48.5	–22.2	–30.1	24.0	2.91	2.11	0.79
1ag6	plastocyanin	81.9	–40.9	–57.1	47.1	2.88	2.15	0.78
1azb	azurin	121.4	–78.0	–82.6	42.5	3.40	2.18	0.74
1b3i	plastocyanin	63.9	–33.9	–39.8	30.8	2.95	2.35	0.92
1bq5	nitrite reductase	43.5	–8.6	–31.5	24.4	2.69	2.17	0.78
1bqk	pseudoazurin	69.5	–26.2	–52.2	41.9	2.71	2.13	0.78
1bxu	plastocyanin	95.8	–46.4	–65.6	54.7	2.94	2.14	0.76
1byo	plastocyanin	94.2	–42.1	–69.1	52.9	2.74	2.19	0.76
1dyz	azurin	136.2	–84.4	–79.9	47.2	3.26	2.14	0.75
1ezl	azurin	164.5	–130.2	–72.6	21.0	3.38	2.30	0.82
1f56	plantacyanin	50.1	–15.9	–36.6	30.8	2.72	2.17	0.75
1id2	amicyanin	43.8	–8.0	–27.4	20.9	2.88	2.04	0.77
1iuz	plastocyanin	77.6	–30.8	–53.6	50.3	2.69	2.18	0.79
1joi	azurin	105.8	–59.4	–63.0	44.0	3.23	2.10	0.71
1jxg	plastocyanin	84.3	–34.2	–60.4	52.4	2.76	2.17	0.78
1mg2	amicyanin	73.8	–5.3	–53.2	38.6	2.78	2.21	0.77
1n70	nitrite reductase	38.8	–18.6	–21.3	23.2	2.46	2.24	0.78
1nwo	azurin	84.2	–48.4	–45.0	36.7	3.01	2.13	0.78
1ov8	auracyanin	54.1	–32.1	–31.4	27.9	2.90	2.07	0.74
1plb	plastocyanin	46.8	–12.7	–36.2	30.4	3.10	2.02	0.73
1pmy	pseudoazurin	52.6	–14.6	–40.7	31.8	2.66	2.15	0.74
1rkr	azurin	159.9	–111.5	–92.2	34.5	3.14	2.14	0.80
2aan	auracyanin	63.4	–16.3	–53.4	39.0	2.90	2.22	0.75
2avf	nitrite reductase	34.3	–4.3	–25.1	20.2	2.35	2.34	0.82
2cbp	cucumber basic protein	70.8	–23.1	–50.3	41.0	2.61	2.16	0.78
2gim	plastocyanin	79.8	–38.7	–50.1	47.6	2.77	2.13	0.76
2plt	plastocyanin	76.7	–29.7	–58.9	47.4	2.89	2.11	0.76
2yqb	nitrite reductase	60.7	–36.3	–29.2	33.2	2.55	2.21	0.77
2zoo	nitrite reductase	62.3	–35.8	–31.1	33.7	2.54	2.22	0.77
3erx	pseudoazurin	81.4	–28.8	–61.6	49.2	2.63	2.22	0.82
3tu6	pseudoazurin	48.0	–11.6	–35.9	27.8	2.50	2.24	0.77
3wia	nitrite reductase	53.1	–33.3	–23.1	28.5	2.63	2.09	0.76
4dp0	plastocyanin	112.3	–50.4	–71.2	66.9	2.59	2.26	0.82
4kns	nitrite reductase	51.1	–31.9	–25.8	27.8	2.36	2.33	0.77
4paz	pseudoazurin	69.9	–43.4	–37.8	37.9	2.76	2.14	0.81
Average		74.1	–35.5	–47.4	36.4	2.80	2.17	0.78
Std.		33.3	28.6	19.5	12.6	0.26	0.08	0.04

[a] Since the coordinating His residues are in principle indistinguishable, we provide the average projection of the LEF values in both Cu-N_{His} directions.

Table S2. Calculated Cu(II) α HOMO and Cu(II) β LUMO energies and the Mulliken spin densities of the selected atoms (at the B3LYP/def2-TZVP level) for the equilibrium structure from **Figure 3** in the main text without the applied field, with the average field experienced in the BC sites, and with the individual projections in the $\mathrm{Cu - \vec{S}Met}$ and $\mathrm{Cu - \vec{S}Cys}$ orientation.

Orbital energy [eV]	w/out OEEF	average OEEF (74.1 MV cm^{-1})	$Cu\text{-}S_{Met}$ projection (–35.5 MV cm-1)	$Cu\text{-}S_{Cys}$ projection (–47.4 MV cm^{-1})
Cu(II) α HOMO	-7.21	-6.50	-7.28	-6.84
Cu(II) β LUMO	-4.75	-5.25	-4.78	-5.08
Mulliken spin density [e]	w/out OEEF	average OEEF (74.1 MV cm^{-1})	$Cu\text{-}S_{Met}$ projection (–35.5 MV cm-1)	$Cu\text{-}S_{Cys}$ projection (–47.4 MV cm^{-1})
Cu	0.34	0.49	0.31	0.46
S_{Cys}	0.51	0.27	0.57	0.30
S_{Met}	0.08	0.05	0.04	0.10
N_{His}[a]	0.06	0.17	0.06	0.12

[a] Combined spin density on both coordinating His residues.

Figure S1. A correlation between Cu-S_{Met} bond dinstances (d(Cu-S_{Met})) vs. LEFs and the individual $\boldsymbol{F}$(Cu-S_{Cys}), $\boldsymbol{F}$(Cu-S_{Met}), and $\boldsymbol{F}$(Cu-N_{His}) projections for the BC proteins from **Table S1**. The empty triangles represent proteins that were crystallized in the reduced form.

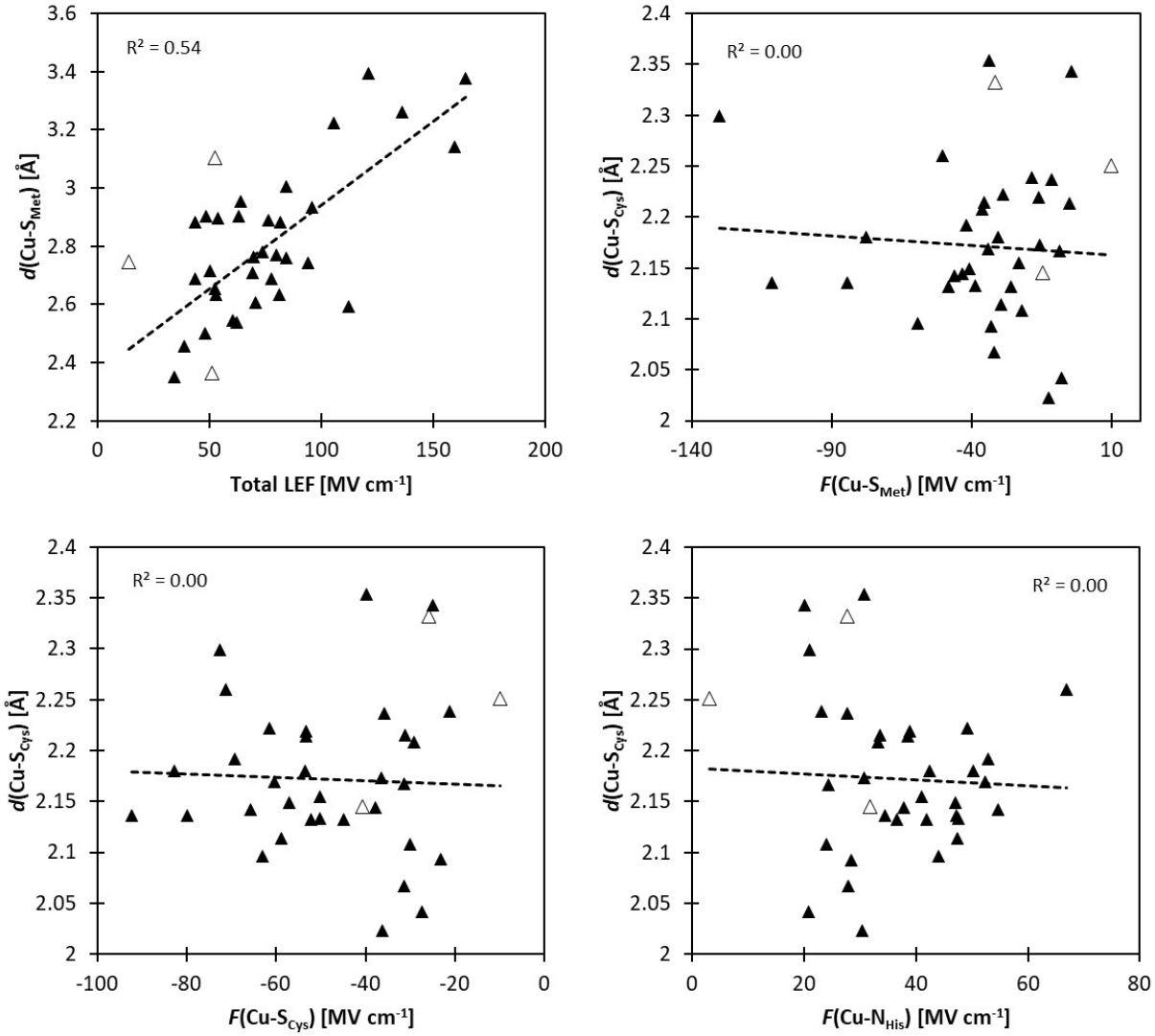


Figure S2. A correlation between Cu-S_{Cys} bond dinstances (d(Cu-S_{Cys})) vs. LEFs and the individual $\boldsymbol{F}$(Cu-S_{Cys}), $\boldsymbol{F}$(Cu-S_{Met}), and $\boldsymbol{F}$(Cu-N_{His}) projections for the BC proteins from **Table S1**. The empty triangles represent proteins that were crystallized in the reduced form.

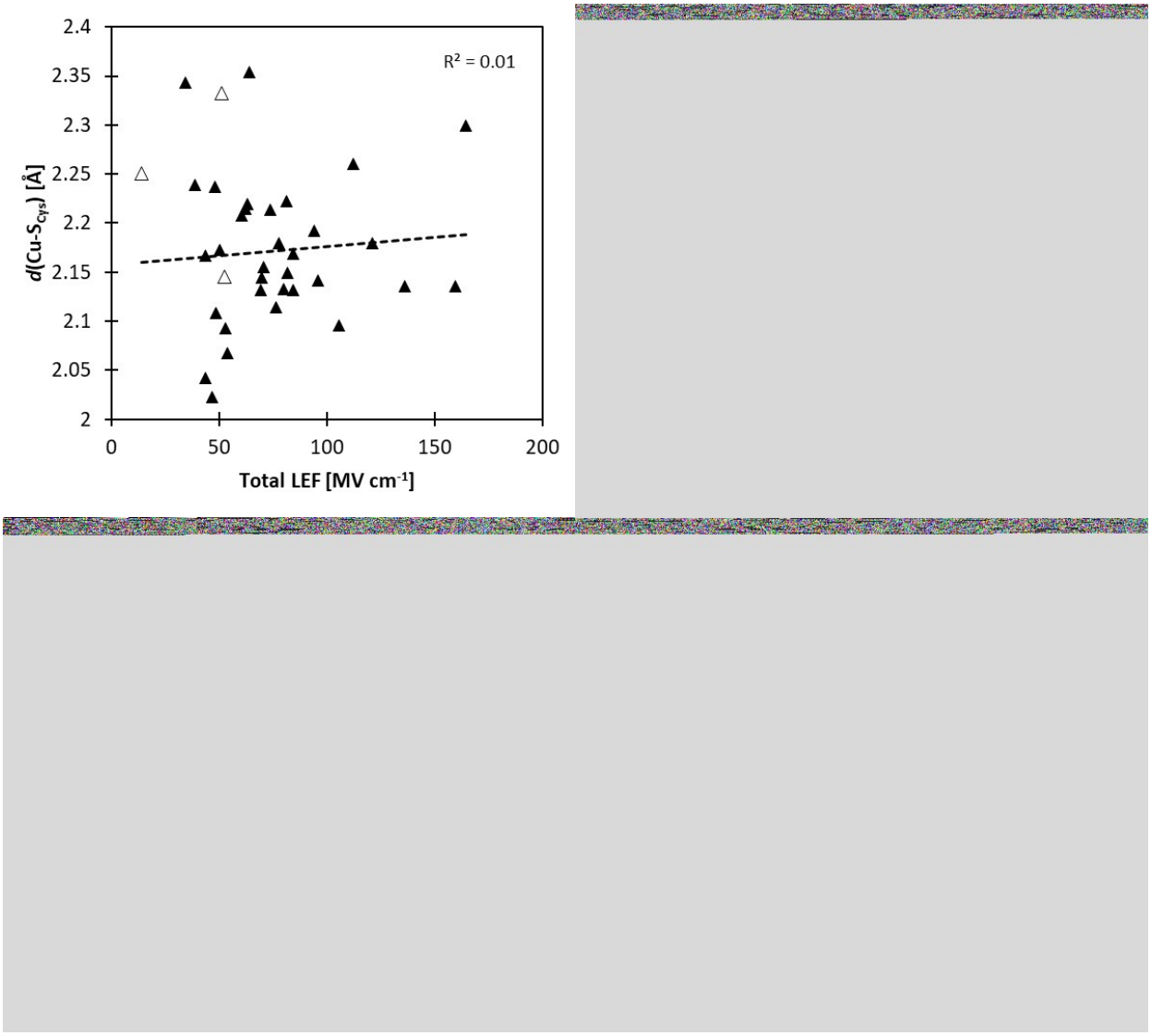


Figure S3. A correlation between Cu-S_{Cys}/Cu-S_{Met} bond dinstances ratio (d(Cu-S_{Cys})/d(Cu-S_{Met})) vs. LEFs and the individual $\boldsymbol{F}$(Cu-S_{Cys}), $\boldsymbol{F}$(Cu-S_{Met}), and $\boldsymbol{F}$(Cu-N_{His}) projections for the BC proteins from **Table S1**. The empty triangles represent proteins that were crystallized in the reduced form.

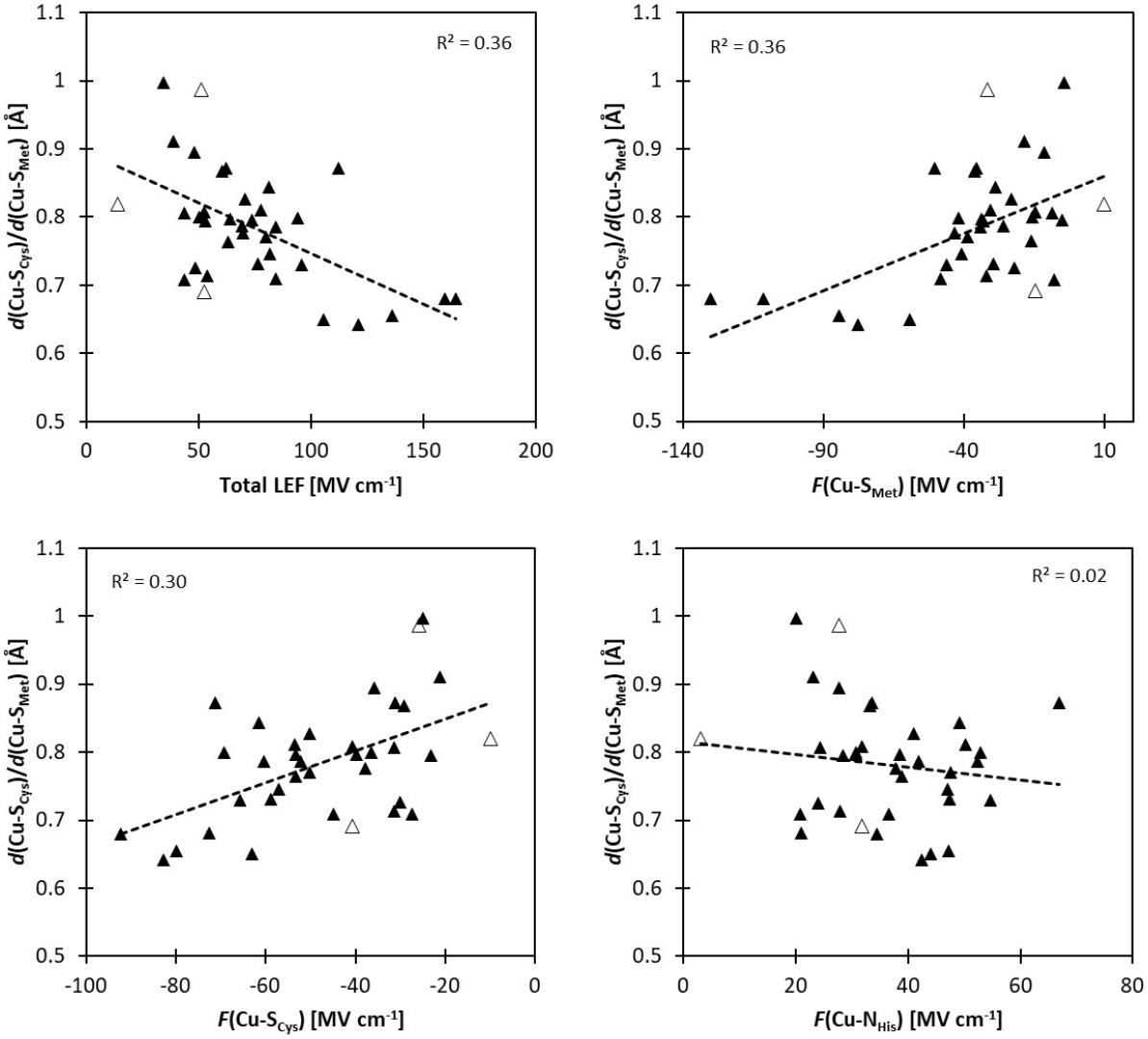
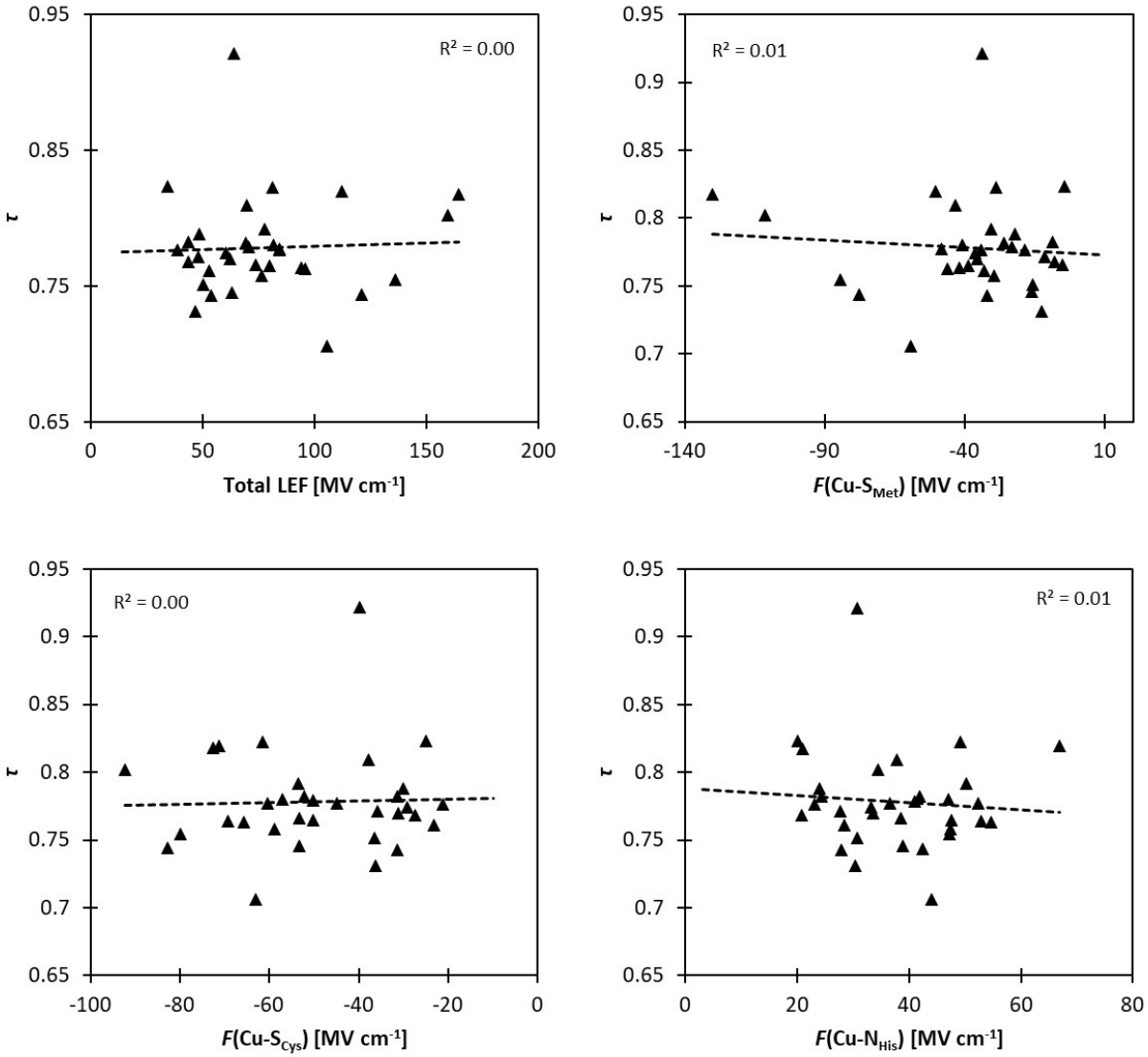


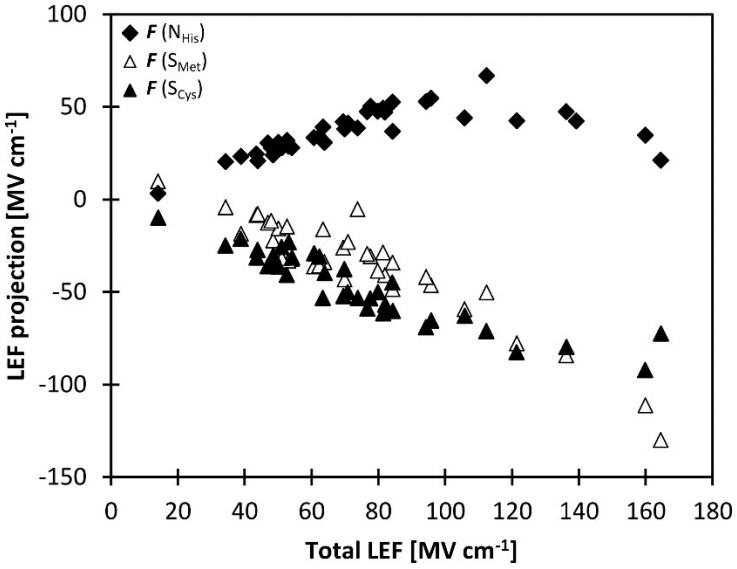
Figure S4. A correlation between geometry index τ vs. LEFs and the individual $\boldsymbol{F}$(Cu-S_{Cys}), $\boldsymbol{F}$(Cu-S_{Met}), and $\boldsymbol{F}$(Cu-N_{His}) projections for the BC proteins from **Table S1**. Note that $\tau = (360° - (\alpha + \beta))/(141°)$, where α and β are the two largest angles between the Cu ligating atoms. The parameter τ was introduced by Yang et al.[13] as a four-coordinate geometry index ranging from perfect tetrahedral (τ = 1) to perfect square planar (τ = 0).



Note S1: Correlation between individual LEF projections

The LEFs in the series of examined BC proteins are oriented in a definite way, as demonstrated in **Figure 1** in the main text. The non-orthogonality of the $\mathrm{Cu}-\vec{\mathrm{S}}\mathrm{Cys}$, $\mathrm{Cu}-\vec{\mathrm{S}}\mathrm{Met}$, and $\mathrm{Cu}-\vec{\mathrm{N}}\mathrm{His}$ vectors thus lead to a pronounced correlation between the LEF and its individual $\boldsymbol{F}(\mathrm{Cu\text{-}S_{Cys}})$, $\boldsymbol{F}(\mathrm{Cu\text{-}S_{Met}})$, and $\boldsymbol{F}(\mathrm{Cu\text{-}N_{His}})$ projections (**Figure S5**). Not surprisingly, the geometric features in **Figures S1-S4** thus correlate with all LEF projections exhibiting a comparable regression coefficient (R^2). However, the origin of such correlations can be traced back to individual bond strengths affected by the particular LEF projections, whereas the projections in the other orientations contribute only negligibly.

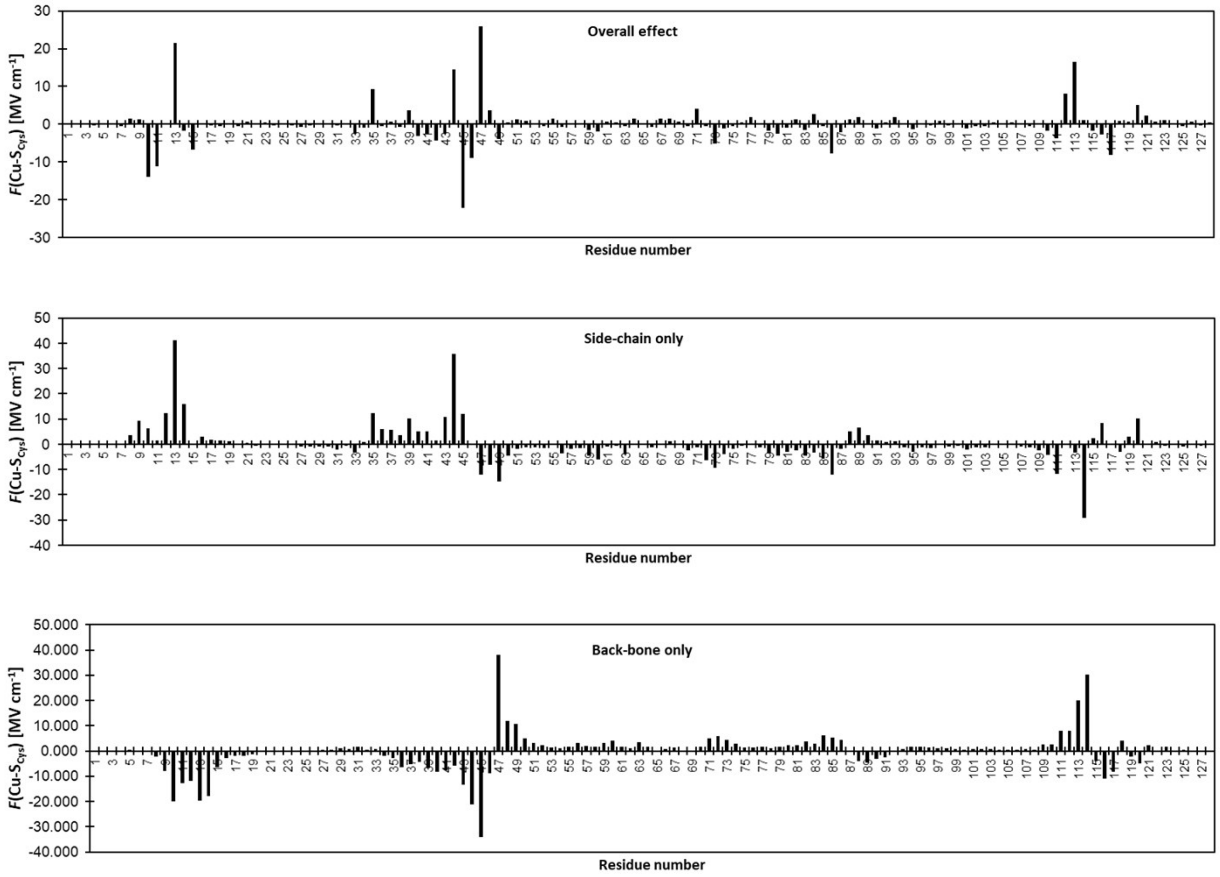
Figure S5. Individual $\boldsymbol{F}(\mathrm{Cu\text{-}S_{Cys}})$, $\boldsymbol{F}(\mathrm{Cu\text{-}S_{Met}})$, and $\boldsymbol{F}(\mathrm{Cu\text{-}N_{His}})$ projections of the LEF vs. the total LEF for the BC proteins from **Table S1**.



Note S2: Individual residues' contributions to the F(Cu-S_{Cys}) in wild-type azurin

The individual contributions of each residue in the wild-type azurin crystal structure (PDB code: 4azu) to F(Cu-S_{Cys}) projection of the LEF was analyzed by calculating F(Cu-S_{Cys}) using Coulomb's law in Eq. (1) with iteratively turning off the charges at each residue. In **Figure 5B** in the main text, the effect of zeroing only the charges on the residues' side chains is presented, which is essential for mutagenetic experiments. Most importantly, the side chains of the residues M13, M14, and F114 contribute to the overall F(Cu-S_{Cys}). By their variation, it should be possible to alter F(Cu-S_{Cys}) and thus electronic absorption intensities or Cu(II/I) reduction potentials of azurin protein. In **Figure S6**, we compare the per-residue contributions to F(Cu-S_{Cys}) originating from the residues' side chains to those from the back-bone atoms and the full residues. This demonstrates how the overall F(Cu-S_{Cys}) is build-up from the individual residual contributions. While the modification of the F(Cu-S_{Cys}) by mutagenesis would be more difficult for residues contributing via back-bone atoms, we note that N47 residue was successfully utilized in the past to modify the Cu(II/I) reduction potential of azurin.[14] Noticeably, the N47 residue has the highest individual contribution to F(Cu-S_{Cys}) (~26 MV cm^{-1}) originating from the dipole moment near Cu-S_{Cys} generated by the back-bone carbonyl.

Figure S6. Contributions to F(Cu-S_{Cys}) from the individual azurin residues.



Note S3: Evaluating of the local electric field for different azurin variants

An agreement between the experimental Cu(II/I) reduction potential and $\boldsymbol{F}(\text{Cu-S}_{\text{Cys}})$ component of the local electric field in various azurin variants in **Figure 5A** suggests that, with the knowledge of the protein crystal structure, we can predict the change of the reduction potential upon introduction of mutations in the wild-type azurin. In Figure 5B in the main text, we have proposed that M13/M44 mutations of the wild-type azurin might lead to significant change in the $\boldsymbol{F}(\text{Cu-S}_{\text{Cys}})$ component of the local electric field at the Cu active site, which could alter the Cu(II/I) reduction potential. However, there are no accessible crystal structures for M13/M44 azurin mutations. Instead, we thus performed molecular dynamics simulations (using discrete molecular dynamics method; DMD – see Computational Details section), and evaluate $\boldsymbol{F}(\text{Cu-S}_{\text{Cys}})$ along the DMD trajectories.

First, we note that for experimentally determined azurin variants from ref. [14] (wild-type azurin, F114P, F114N, N47S, N114N/N47S), the correlation between the Cu(II/I) reduction potentials and $\boldsymbol{F}(\text{Cu-S}_{\text{Cys}})$ from DMD simulations holds relatively well (**Figure S8**). This suggests we can quite well predict the effects of mutations, even without knowing the crystal structures. The absolute results are a bit different from the fields analyzed directly from the crystal structures in **Figure 5A**. This can be due to using simplified protein models in MD simulations; e.g., we included only chain A (monomer) for the simulations (instead of tetramer). Also, we have kept the coordinating atoms (2N_{His},S_{Cys},S_{Met}) and Cu fixed at the original crystal positions of wild-type azurin (*vide supra*).

Next, we have evaluated the possible mutations of M13 and M44 residues, as proposed in the main text. We show that they are capable of shifting the local electric field at Cu (i.e., the $\text{Cu} - \vec{\text{S}}\text{Cys}$ projection) negative *wrt* to wild-type azurin. Namely, the single-point mutations, such as M44F, M13E, M44S, and M44D are predicted to shift the $\boldsymbol{F}(\text{Cu-S}_{\text{Cys}})$ by ~12.5 – 17.5 MV cm^{-1} to more negative values (**Figure S7**), corresponding to an increase in the Cu(II/I) reduction potential by as much as 150 mV. Note that such change is comparable to the experimentally tested single-point mutations N47S and F114N. Additionally, the M13/M44 double mutants, such as M13F/M44F can lower the $\boldsymbol{F}(\text{Cu-S}_{\text{Cys}})$ by ~26 MV cm^{-1}, and the reduction potential of wild-type azurin by ~250 mV; again comparable to the best experimental (second-sphere) double mutant F114N/N47S.

Figure S7. The relative change of the $F(\text{Cu-S}_{\text{Cys}})$ projection of the local electric field at the Cu site of the wild-type azurin protein (PDB code: 4azu), upon introduced mutations.

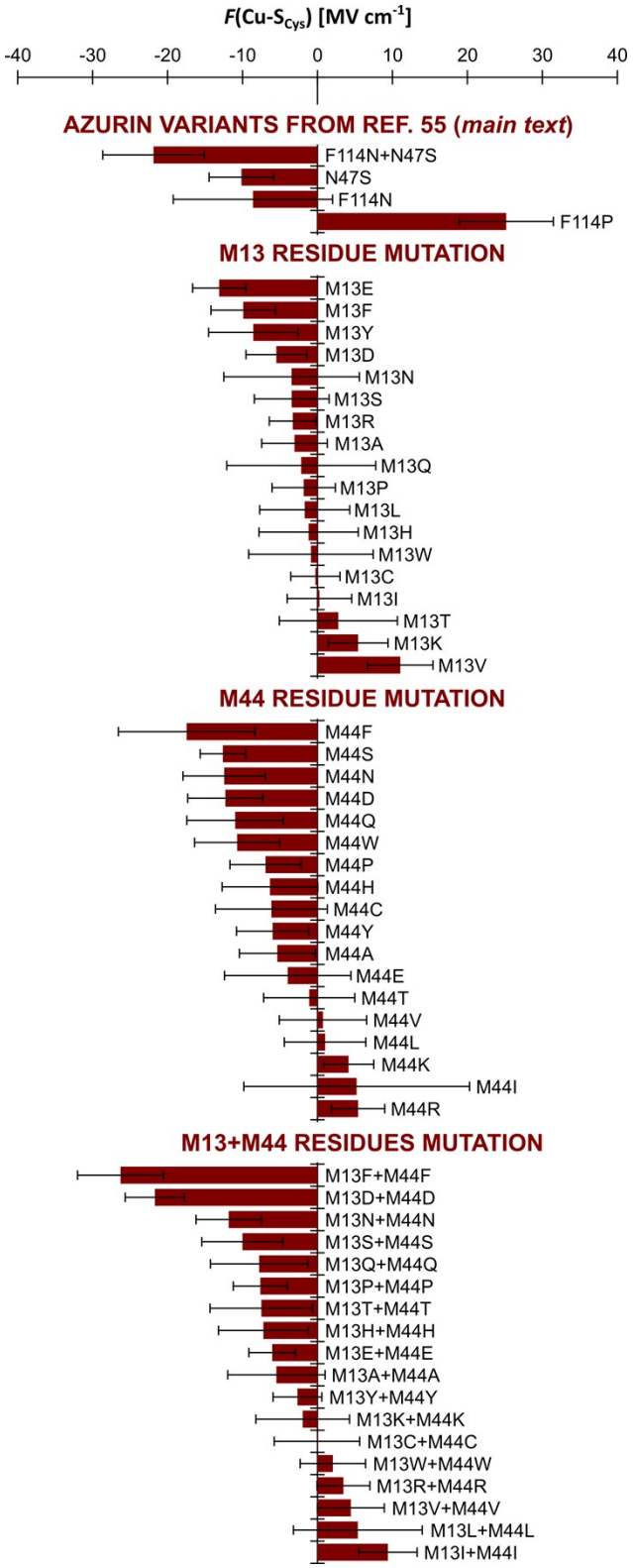


Figure S8. Correlation between experimental one-electron Cu(II/I) reduction potentials (referenced to normal hydrogen electrode) vs. the change of $\boldsymbol{F}$(Cu-S_{Cys}) projection of the LEF in a series of azurin variants from ref. 55 (*main text*), that were obtained from the molecular dynamics simulations.

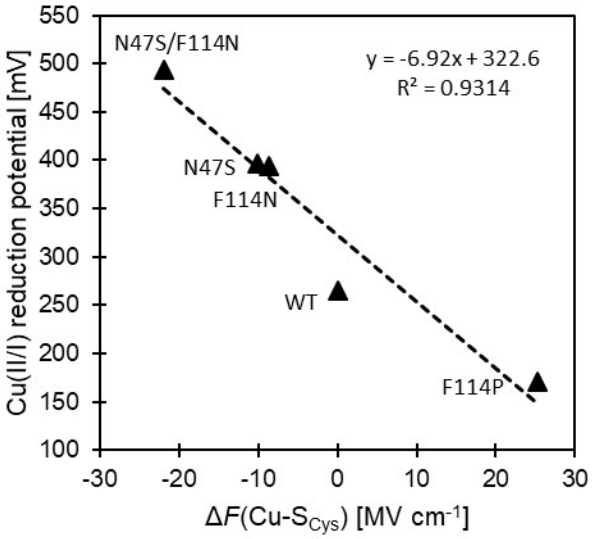
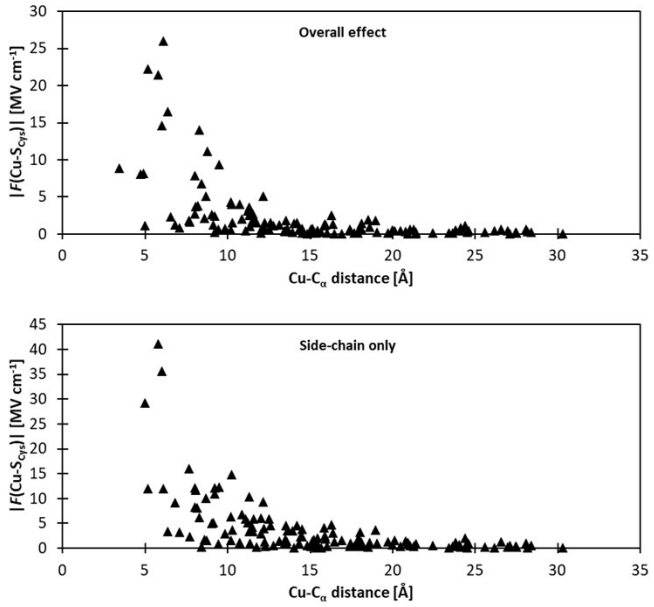


Figure S9. The absolute contributions to $\boldsymbol{F}$(Cu-S_{Cys}) are significantly reduced with the individual residues' distance from the Cu active site. Irrespective of the $\boldsymbol{F}$(Cu-S_{Cys}) origin (side chain or backbone), the contributions become negligible beyond ca. 15 Å distance between Cu and C_α carbon of each residue.



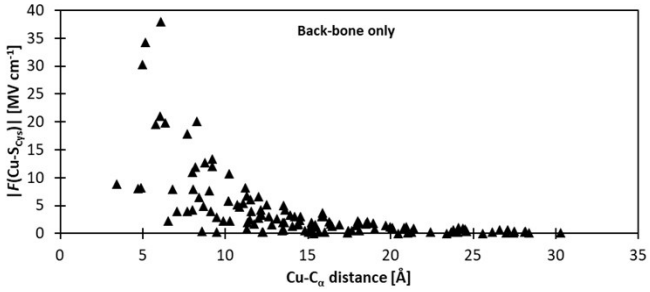
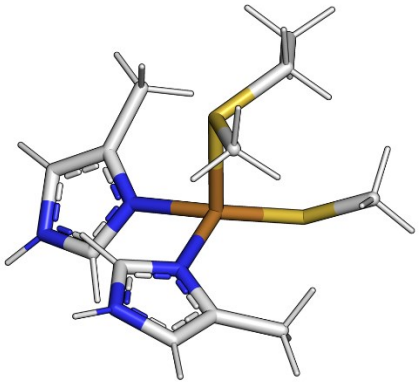


Figure S10. Cluster model of the BC active site utilized in the DFT calculations. The cluster model was generated from the oxidized poplar plastocyanin crystal structure (PDB code: 4dp9).



References

1. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
2. Shirvanyants, D.; Alexandrova, A. N.; Dokholyan, N. V. *Bioinformatics* **2011**, *27*, 1327–1329.
3. Rice, P.; Longden, I.; Bleasby, A. *EMBOSS: The European Molecular Biology Open Software Suite (2000). Trends in Genetics 16, (6) p.276-277*.
4. Schrödinger, L. *The {PyMOL} Molecular Graphics System, Version~2.3.4*; 2019.
5. Ionescu, C.-M.; Sehnal, D.; Falginella, F. L.; Pant, P.; Pravda, L.; Bouchal, T.; Svobodová Vařeková, R.; Geidl, S.; Koča, J. *J. Cheminform.* **2015**, *7*, 50.
6. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
7. Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
8. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
9. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*.
10. Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, No. 5, 799–805.
11. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
12. Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.
13. Yang, L.; Powell, D. R.; Houser, R. P. *Dalt. Trans.* **2007**, No. 9, 955–964.
14. Marshall, N. M.; Garner, D. K.; Wilson, T. D.; Gao, Y. G.; Robinson, H.; Nilges, M. J.; Lu, Y. *Nature* **2009**, *462*, 113–116.