



# New Techniques for Ancient Proteins: Direct Coupling Analysis Applied on Proteins Involved in Iron Sulfur Cluster Biogenesis

Marco Fantini<sup>1</sup>, Duccio Malinverni<sup>2</sup>, Paolo De Los Rios<sup>2\*</sup> and Annalisa Pastore<sup>3,4\*</sup>

<sup>1</sup> BioSNS, Faculty of Mathematical and Natural Sciences, Scuola Normale Superiore, Pisa, Italy, <sup>2</sup> Institute of Physics, School of Basic Sciences, and Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, <sup>3</sup> Maurice Wohl Institute, King's College, London, United Kingdom, <sup>4</sup> Molecular Medicine Department, University of Pavia, Pavia, Italy

## OPEN ACCESS

### Edited by:

Elena Papaleo,  
University of Copenhagen, Denmark

### Reviewed by:

Julian Echave,  
National University of General San  
Martín, Argentina  
Elodie Laine,  
Université Pierre et Marie Curie,  
France

### \*Correspondence:

Paolo De Los Rios  
paolo.delosrios@epfl.ch  
Annalisa Pastore  
annalisa.pastore@crick.ac.uk

### Specialty section:

This article was submitted to  
Mathematics of Biomolecules,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 19 March 2017

**Accepted:** 24 May 2017

**Published:** 15 June 2017

### Citation:

Fantini M, Malinverni D, De Los Rios P  
and Pastore A (2017) New Techniques  
for Ancient Proteins: Direct Coupling  
Analysis Applied on Proteins Involved  
in Iron Sulfur Cluster Biogenesis.  
*Front. Mol. Biosci.* 4:40.  
doi: 10.3389/fmolb.2017.00040

Direct coupling analysis (DCA) is a powerful statistical inference tool used to study protein evolution. It was introduced to predict protein folds and protein-protein interactions, and has also been applied to the prediction of entire interactomes. Here, we have used it to analyze three proteins of the iron-sulfur biogenesis machine, an essential metabolic pathway conserved in all organisms. We show that DCA can correctly reproduce structural features of the CyaY/frataxin family (a protein involved in the human disease Friedreich's ataxia) despite being based on the relatively small number of sequences allowed by its genomic distribution. This result gives us confidence in the method. Its application to the iron-sulfur cluster scaffold protein IscU, which has been suggested to function both as an ordered and a disordered form, allows us to distinguish evolutionary traces of the structured species, suggesting that, if present in the cell, the disordered form has not left evolutionary imprinting. We observe instead, for the first time, direct indications of how the protein can dimerize head-to-head and bind 4Fe4S clusters. Analysis of the alternative scaffold protein IscA provides strong support to a coordination of the cluster by a dimeric form rather than a tetramer, as previously suggested. Our analysis also suggests the presence in solution of a mixture of monomeric and dimeric species, and guides us to the prevalent one. Finally, we used DCA to analyze interactions between some of these proteins, and discuss the potentials and limitations of the method.

**Keywords:** co-evolution, computational methods, direct coupling analysis, iron-sulfur cluster biogenesis, molecular machines, protein folding

## INTRODUCTION

Protein sequences determine the folds of proteins and what interactions they may form with their partners. The logic connecting residue-residue contacts to evolutionary correlation is very simple: residues in contact cannot evolve independently. If one residue gets larger, the other needs to be smaller in a concerted and not necessarily pairwise way. Charges must be compensated in the same way. Stabilizing/destabilizing amino acid substitutions need to be compensated by substitution of other interacting positions to retain function. In principle, one can thus use a comparative analysis of the primary sequences of proteins as a powerful way to predict their structures and interactions.

This idea has been an “elusive Holy Grail” for decades (Altschuh et al., 1987; Göbel et al., 1994; Pazos et al., 1997). More recently, an effective method, called direct coupling analysis (DCA) (Weigt et al., 2009; Morcos et al., 2011), has been proposed as a powerful approach to determine from an evolutionary perspective which residues interact, exploiting the large, and growing, number of available protein sequences. The method has been used successfully to acquire constraints for structural, dynamical and functional analysis (Dago et al., 2012; Hopf et al., 2012, 2015; Marks et al., 2012; Espada et al., 2015; Malinverni et al., 2015; Sutto et al., 2015), multimerization (Hopf et al., 2014; Ovchinnikov et al., 2014), and to shed light on interaction specificity (Bitbol et al., 2016) and inter-pathway cross-talk in bacterial signal transduction (Procaccini et al., 2011; Kensche et al., 2012).

Here, we have applied DCA to explore the nature of the interactions between proteins involved in the biosynthesis of iron-sulfur (FeS) clusters, which are essential prosthetic groups that provide electrons in reduction/oxidation reactions and/or stabilize protein folds. This biosynthesis is a complex process involving specialized machinery that mediates the recruitment of sulfur and free iron from the cellular environment, catalyzes the synthesis, and delivers the newly formed clusters to acceptor proteins. In bacteria, the systems able to perform these tasks belong to the *nif* (nitrogen fixation, *Nif* *iscA-nifSU*), *isc* (iron-sulfur complex, *iscRSUA-hscBA-fdx*), and *suf* (mobilization of sulfur, *sufABCDSE*) operons. Amongst these, the most universal is the *isc* operon, whose gene products have direct orthologs in eukaryotes. Malfunction in FeS cluster assembly has direct effects on health (Beilschmidt and Puccio, 2014; Rouault, 2015). Elucidation of the structures and the interactions between the various proteins involved in this process can thus provide valuable insights in the origin of several diseases.

The central players in the *isc* machine are IscS (or Nfs1 in eukaryotes) and IscU (Isu) (Figure 1). IscS is a desulfurase that converts cysteine to alanine and forms the persulfide that is incorporated into the cluster. IscU is the scaffold protein on which the cluster is assembled. Together, IscS and IscU form a complex in which two IscU monomers are bound to the IscS obligate dimer. It has been suggested that IscU exists in two conformational states in the cell, one folded and ordered (S state), the second being partially unfolded, or disordered (D state) (Bothe et al., 2015). However, all crystal structures of IscU in isolation and in complexes with zinc or IscS capture the protein in its ordered state. IscU was described as a dimer when bound to a 4Fe4S cluster (Agar et al., 2000) but it is observed as a monomer when isolated in solution and when bound to IscS (Prischi et al., 2010; Shi et al., 2010; Marinoni et al., 2012). Two regulatory proteins are CyaY (frataxin), which is the protein involved in Friedreich's ataxia in humans, and IscA, which is thought to be an alternative scaffold protein. CyaY/frataxin is a globular monomeric protein formed from a conserved domain preceded in eukaryotes by an intrinsically unfolded mitochondrial import sequence. It is highly conserved from bacteria to primates (Gibson et al., 1996), acts as a regulator of the enzymatic activity of IscS and binds it in a site close to the active site (Adinolfi et al., 2009; Shi et al., 2010). Puzzlingly, its

presence seems to inhibit the activity of IscS in prokaryotes but to activate it in eukaryotes (Gakh et al., 2010; Prischi et al., 2010; Tsai and Barondeau, 2010; Iannuzzi et al., 2011). IscA is an ancient protein thought to be an alternative scaffold for cluster formation. The IscA family is characterized by a conserved CX<sub>n</sub>CGCG motif thought to be involved in iron and/or 2Fe-2S binding (Kaut et al., 2000). In all available structures, IscA is either dimeric or tetrameric, but different symmetries and cluster coordination have been suggested (Bilder et al., 2004; Cupp-Vickery et al., 2004; Morimoto et al., 2006).

We found that DCA is able to describe the proteins considered in great detail. We used frataxin, which is monomeric and globular, to benchmark the method. We questioned whether evolution could tell us if IscU is folded or a mixture of ordered and disordered species; we also addressed the question of how the IscU dimer, which undoubtedly must exist to allow coordination of the cluster, assembles; finally, we wondered if DCA could give us indications of which of the reported structures of IscA is more representative of the protein in solution.

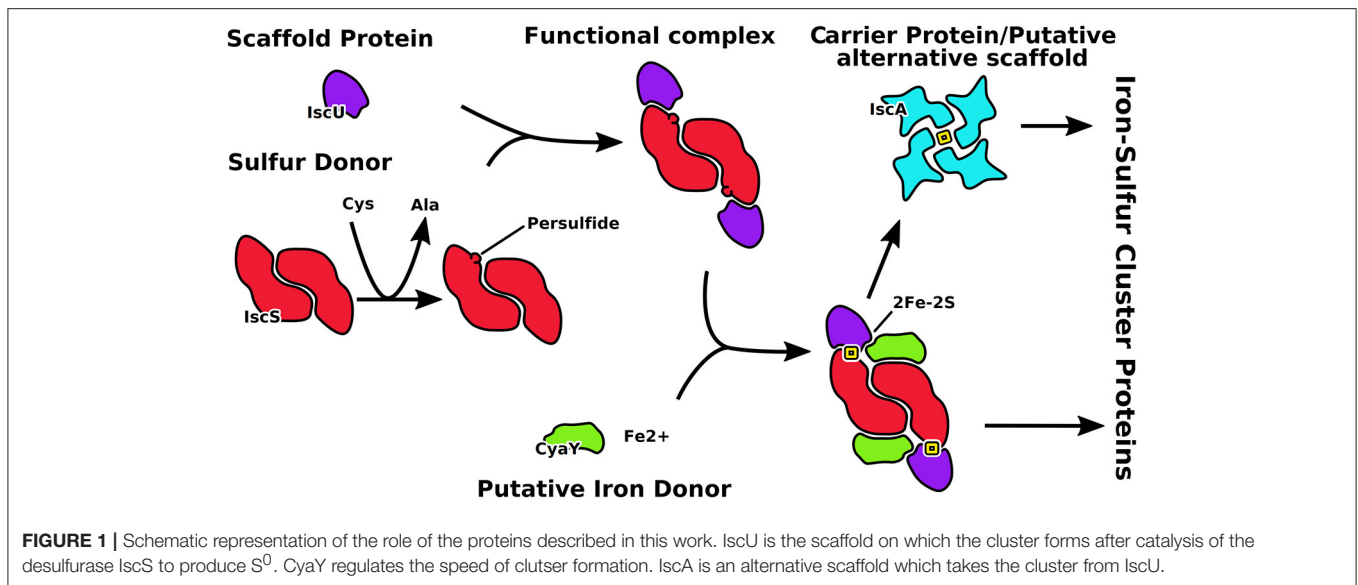
We observed no trace of the D state of IscU whereas the S state is clearly represented in the contact co-evolution. Instead we observed structural evidence which hints at a head-to-head dimerization of IscU. This is in agreement with what is required by cluster coordination. We also found that not all the IscA structures in the PDB database match the conserved contacts, which suggests that the coordination of the FeS cluster was likely misattributed. Finally, we were able to predict successfully interactions between IscU and the functional partner IscS, whereas contacts predicted for CyaY do not match our current knowledge. These observations are likely to reflect the possibilities but also the limitations of DCA.

## RESULTS

### Benchmarking the Method on the Frataxin Family

The major sequence divergence within the CyaY/frataxin family is in the non-conserved and mainly unstructured N-terminus (Prischi et al., 2009; Popovic et al., 2015). The evolutionary conserved C-terminal domain forms a compact globular structure in which two  $\alpha$ -helices pack against a  $\beta$ -sheet composed of 5–7 strands arranged in a  $\alpha\beta\beta\beta\beta(\beta\beta)\alpha$  motif. The available structures of this region are all similar (average RMSD  $\sim 2.3$  Å) with minor differences in details (Supplementary Table S1). Different orthologs differ in the length of the C-terminus, which is longer in human frataxin and shorter in yeast. This difference contributes to the thermodynamic stability of the protein (Adinolfi et al., 2004). Experimental evidence suggests that the region interacting with iron and with the desulfurase IscS/Nfs1 is located in  $\alpha 1$  and  $\beta 1$  (Figure 2A, Nair et al., 2004; Pastore et al., 2007; Prischi et al., 2010).

We retrieved all the sequences matching a Hidden Markov Model (HMM) from the Uniprot database, constructed from a seed made of the 196 CyaY entries of the Swiss-Prot database. We then built a multiple sequence alignment (MSA) containing 3,459 sequences, defining 109 consensus residue



positions where 1,102 sequences were from eukaryotes and 2,326 from bacteria. The number of retrieved sequences is relatively small for a successful application of DCA but reflects the absence of frataxin in several species (Huynen et al., 2001). We then performed DCA on this MSA using the pseudo-likelihood approximation described in Balakrishnan et al. (2011), a method that estimates the joint probability distribution of a collection of random variables. The predicted contacts are displayed in contact maps that have the protein sequence numbering on both axes (**Figure 2B**). Contacts are displayed as spots that indicate interactions between residues. Traces perpendicular to the diagonal indicate that this region forms an antiparallel secondary structure. Parallel traces reflect interactions between parallel strands. Contacts which do not line up in parallel or perpendicular fashion but cluster in various regions of the plot correspond to contacts between distal elements.

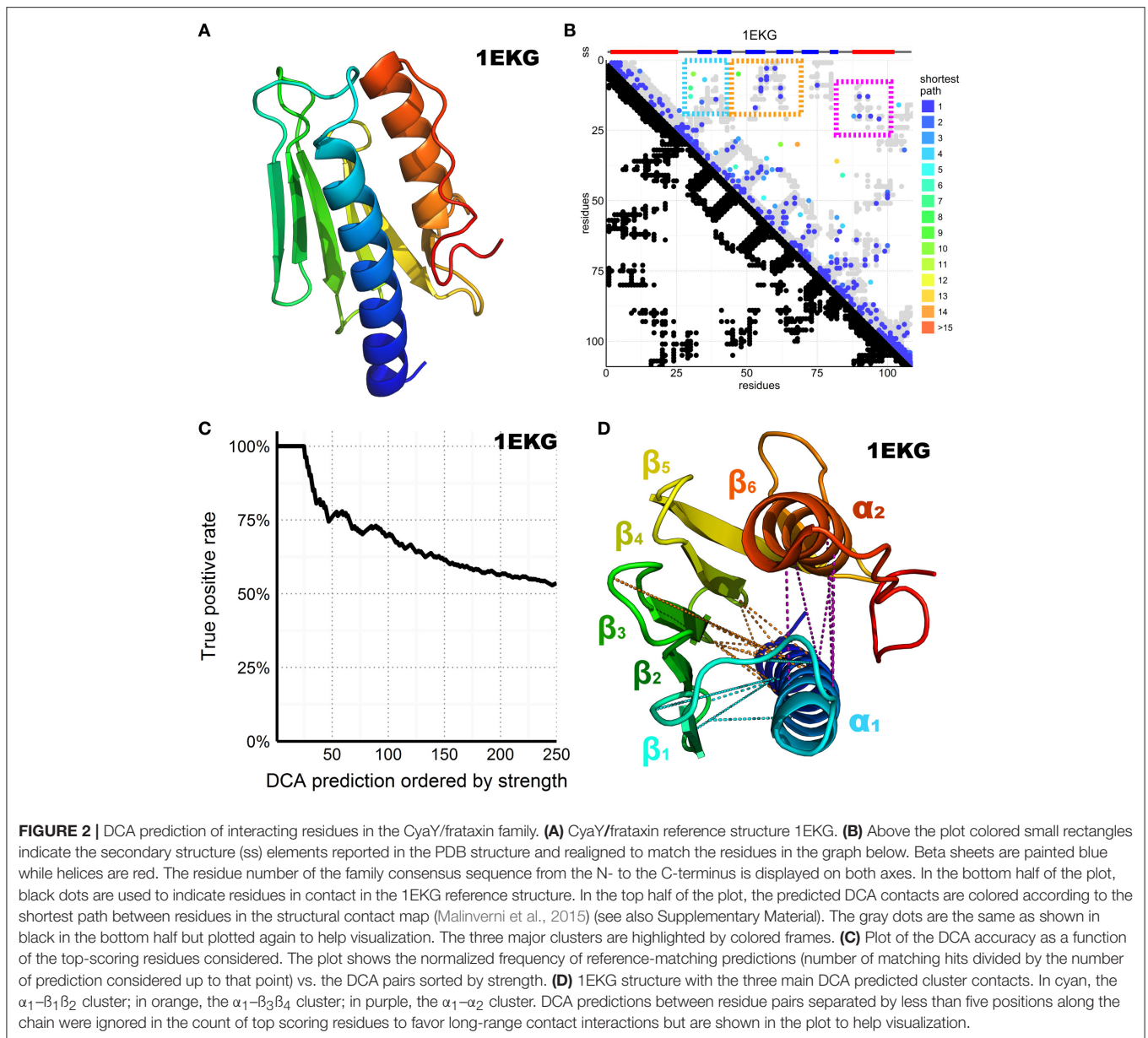
Despite the relatively small dataset of sequences, the DCA pairs show a satisfactory true positive rate overall (**Figure 2C**). We retained the top 109 DCA contacts with the highest scores, which correspond to *ca.* 2% of the total 5460 possible contacts. The retained contacts correlate well with the secondary structure of the protein. Additionally, three clusters were observed, all involving the amino terminus of the domain (**Figure 2D**). Two clusters reflect packing of  $\alpha_1$  against  $\beta_1$ – $\beta_2$  and  $\beta_3$ – $\beta_4$ . The third reflects contacts between the two helices. This tells us how important  $\alpha_1$  is for this protein fold. The only other tertiary interactions between distant secondary structure elements involve  $\beta_4$ – $\beta_5$  and the C-terminal  $\alpha_2$ . This interaction is reflected in the DCA analysis by a small cluster visible at the very bottom of the DCA plot.

These results support the confident use of DCA for the analysis of FeS proteins: even though the number of retrieved sequences is suboptimal, we were able to capture most of the important features of the CyaY/frataxin fold.

## Structure of IscU Proteins and N-Terminal Localization

IscU is a more complex case. Twelve structures are available from 8 different species (Supplementary Table S1). They can be divided into three groups. All the X-ray structures, which are available for isolated cluster-loaded (holo, 2Z7E), zinc-loaded IscU (1SU0 and 2QQ4) and for complexes with IscS/Nfs1 (3LVL, 4EB5, and 4EB7), have a compact ordered structure with a  $\beta$ -sheet packing against two  $\alpha$ -helices (**Figure 3A**). The N-terminus (residues 1–21) does not contain regular secondary structure elements apart from a two-turn helix ( $\alpha_1$ ) between residues 5–12 which packs against the other helix anchoring the N-terminus to the rest of the structure. In one of the structures (2Z7E), the N-terminus adopts different orientations in the different protomers of a homo-trimer. In the solution structures, (1R9P, 1Q48, 2L4X, 2KQK, and 1WFZ), the fold is similar, but the N-terminus is disordered and completely solvent-exposed (**Figure 3B**). Some of these structures are thought to contain a zinc atom in the same position as where the cluster is coordinated (i.e., on the tip of the approximate ellipsoid where three conserved cysteines are). However, zinc is NMR-silent and cannot be observed directly. Only two crystallographic structures (1SU0 and 2QQ4) contain zinc explicitly. Finally, one zinc-free NMR structure (2L4X) is supposed to be representative of a partially unfolded state. However, it is probably more correct to describe this entry as a nascent chain or a molten globule rather than a structure as we normally intend. Its presence in the PDB is misleading.

DCA on 13,148 IscU sequences resulted in a clear co-evolutionary prediction of contacts (**Figures 3C,D**). The analysis was characterized by an excellent true positive rate of the prediction vs. the reference structure 3LVL (88%) and due to the necessity to visualize some weaker interaction, the number of DCA contact kept threshold was set to double the usual amount (74% true positive rate). We observe contacts between  $\beta_1$ – $\beta_2$ ,  $\beta_2$ – $\beta_3$ ,  $\beta_3$ – $\alpha_2$ ,  $\alpha_2$ – $\alpha_3$ , and  $\alpha_3$ – $\alpha_6$  (according to the nomenclature used in Liu et al., 2005) as traces perpendicular to the diagonal,

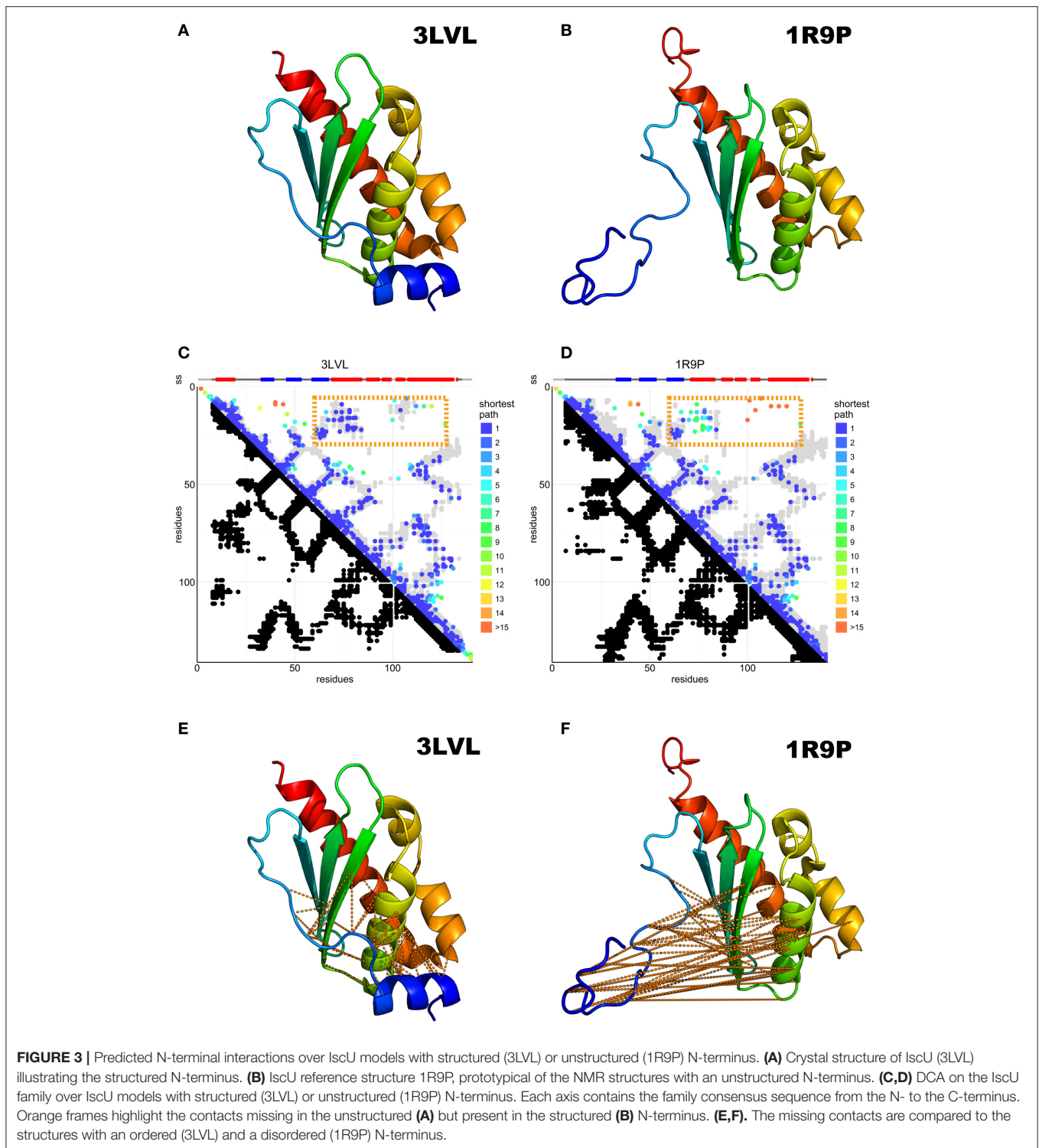


while the  $\beta_2$ - $\alpha_2$ ,  $\beta_3$ - $\alpha_6$ ,  $\alpha_2$ - $\alpha_6$  interactions are reflected by three traces parallel to the diagonal. All secondary structure elements from  $\beta_1$  to  $\alpha_6$  form contacts with the previous and the subsequent secondary elements, forming hairpins. The parallel traces reflect interactions between parallel strands. Helix  $\alpha_1$  does not conform to this pattern and forms interactions with several strands, suggesting a transversal orientation across the sheet.

Most experimental structures agree with these predicted contacts (Figures 3C,D) with the exception of the N-terminal region (up to *ca.* residue 16), which is also where the structures differ most. Contacts between the N-terminus and the  $\beta_2$ - $\beta_3$ - $\alpha_2$  region are conserved, in support of a structured state in the  $\alpha_1$  region (Figures 3E,F). This does not, however, preclude the existence or the functional relevance of a disordered

conformation of the N-terminus: disordered regions have a weaker co-evolutionary signal and are thus difficult to probe in current DCA predictions (Toth-petroczy et al., 2016).

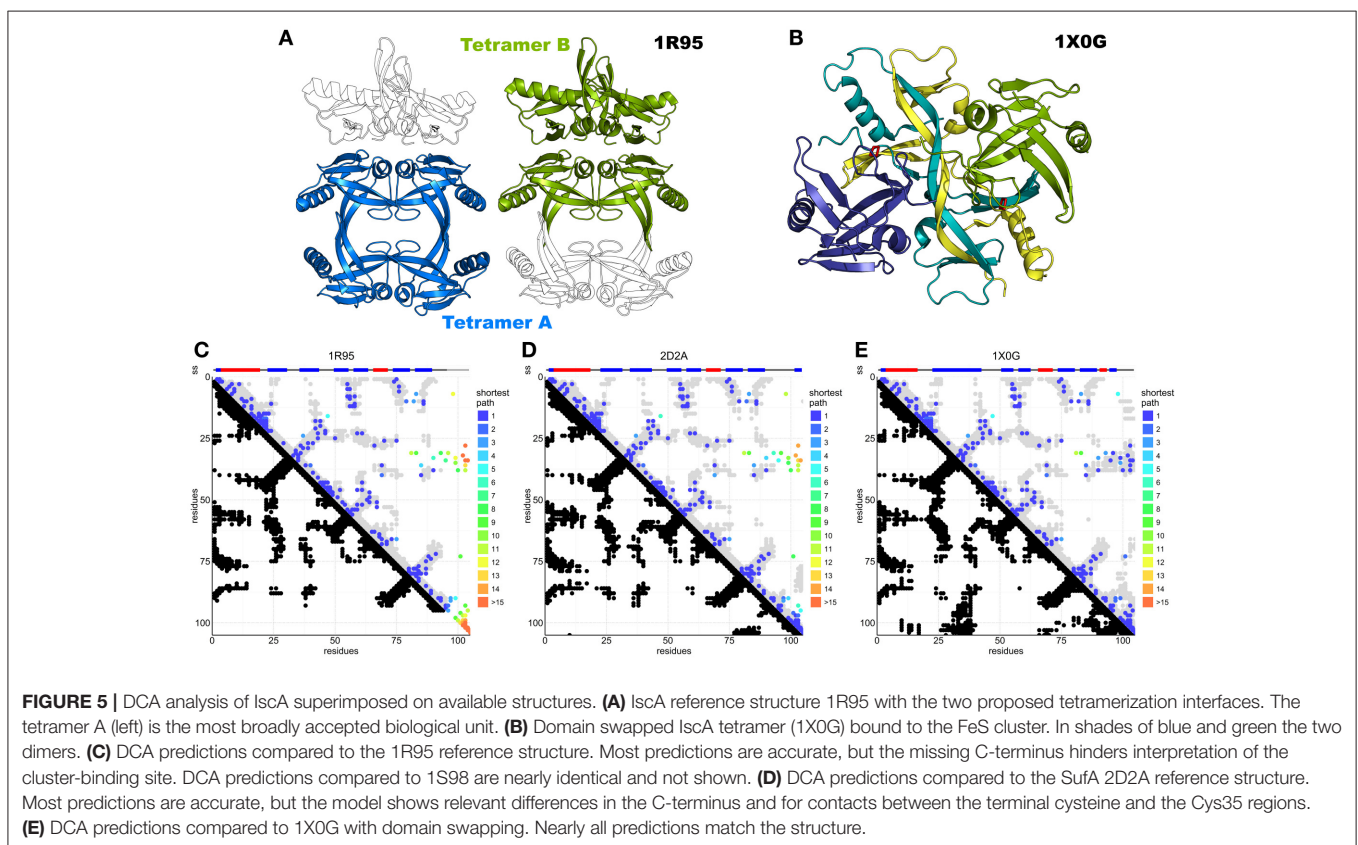
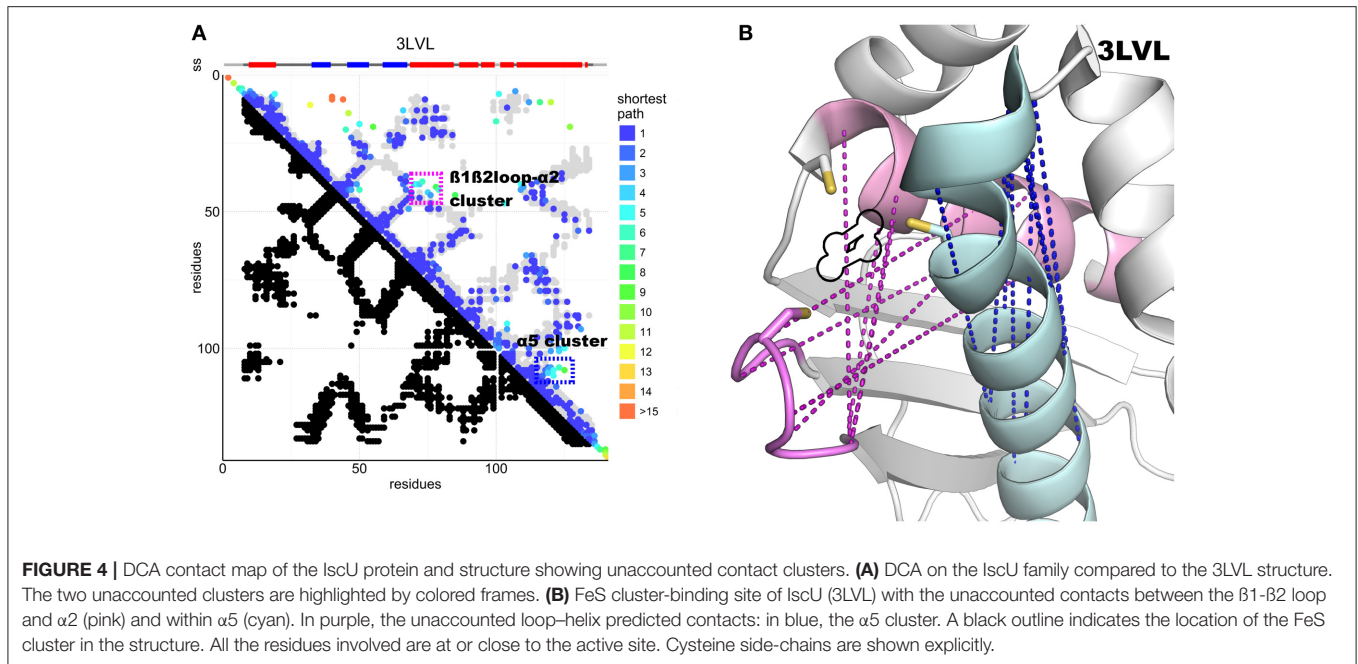
The N-terminus also forms contacts with the  $\beta$ -sheets and the  $\alpha_1$ - $\beta_1$  loop. Superposition of the predicted contacts onto the deposited structures leaves two unaccounted predicted contact clusters, one between  $\alpha_2$  and the  $\beta_1$ - $\beta_2$  loop, the other within the  $\alpha_5$  region (Figure 4A). These contacts are incompatible with the inter-molecular interactions observed in the crystal structures of the cluster-loaded trimer (2Z7E) or of a decamer (2QQ4) (Figure S1) and include areas involved in or surrounding the FeS cluster-binding site (Figure 4B). A different explanation could be that these contacts reflect formation of a head-to-head dimer with an interface located



around the conserved cysteines. This hypothesis would be fully consistent with the necessity of at least a dimer to coordinate a 4Fe4S cluster (Adrover et al., 2015) according to an oxidative mechanism previously proposed (Chandramouli et al., 2007).

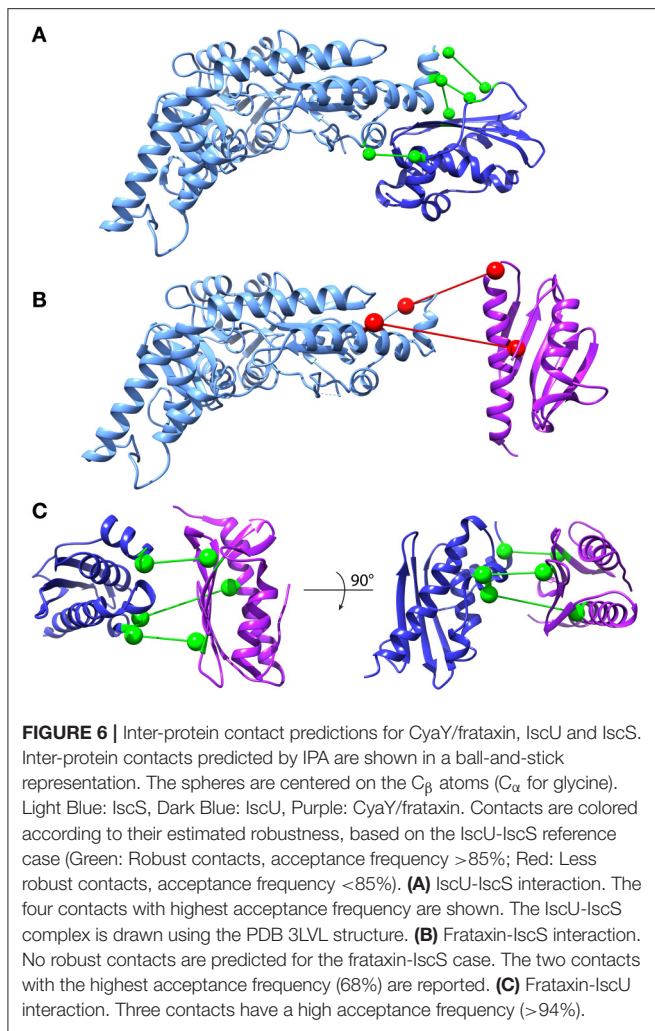
### Multimerization and FeS Cluster Coordination of IscA

Seven structures of IscA-like proteins are available (Supplementary Table S1). The first published structure (1R95) (Bilder et al., 2004) has an internal 2-fold symmetry with tandem



pseudo-symmetric motifs ( $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\beta 3/\beta 5$ - $\alpha 2$ - $\beta 6$ - $\beta 7$ ) separated by a quasi-palindromic hinge ( $E_{43}FVDEPTPEDIVFE_{56}$  in the  $\beta 3$ - $\beta 4$  region). The fold of each protomer consists of a  $\beta$ -sandwich of a mixed twisted four-stranded  $\beta$ -sheet,  $\beta 4$ - $\beta 5$ - $\beta 2$ - $\beta 3$ ,

packed against a three-stranded  $\beta 1$ - $\beta 6$ - $\beta 7$  sheet. The protomers could form a dimer or two possible tetramers or dimer of dimers (tetramers A and B, **Figure 5A**). The electron density around the C-terminus (where two of the three cysteine residues are)



is fuzzy, indicating disorder or conformational exchange. An alternative apo IscA crystal structure (Cupp-Vickery et al., 2004) has individual protomers nearly identical to those observed in 1R95, but the dimer interface, described as an  $\alpha_1\alpha_2$  dimer with minor differences between protomers, is different. The overall tetrameric ( $\alpha_1\alpha_2$ )<sub>2</sub> structure is similar to the 1R95 A tetramer. Also, this structure lacks a defined C-terminus but the authors modeled it based on stereochemical parameters. The authors concluded that the cysteines of the dimer would be unable to coordinate the FeS cluster and that tetramer formation is necessary to stabilize coordination (Cupp-Vickery et al., 2004). They also suggested that of the three cysteines of the CX<sub>n</sub>CGCG motif, only the last two (Cys99 and Cys101 in *E. coli*) are involved in cluster coordination, whereas Cys35 would remain idle. The only fully resolved holo IscA is from *T. elongatus* (1X0G) (Morimoto et al., 2006). This structure has a structured C-terminus that allows coordination of the FeS cluster. It is a dimer of asymmetric dimers ( $\alpha\beta$ )<sub>2</sub> and has domain swapping between two of the protomers ( $\beta$  and  $\beta'$ ) which exchange their central domain forming a long intertwined  $\beta$ -sheet (Figure 5B). The unusual asymmetry imposes asymmetric interfaces, one of

which (the one between  $\alpha$  and the domain-swapped  $\beta'$ ) forms the pocket that accommodates the FeS cluster. The pocket itself is asymmetric with the cysteine motif (Cys37, Cys101, Cys103) contributed both by the  $\alpha$  protomer and the swapped domain of the  $\beta$  protomer [Cys103( $\beta_{sw}$ )] (Figure 5B and Supplementary Figure S1).

In our analysis (84% true positive rate versus 1X0G reference structure), most of the sequences belong either to the IscA or to the ErpA subfamilies but comprise also SufA and the eukaryotic paralogs IscA1/IscA2 (ca. 11,000 sequences). These proteins are all part of the A-type carrier (ATC) family and should have overlapping functions. Structurally, SufA (2D2A) and IscA (1R95, 1S98) have similar contact maps except for two regions, which account for contacts within the C-terminus and between the C-terminus and residues 30–40 (Figures 5C,D). These regions contain the three conserved cysteines. Cluster coordination is thought to occur inter-molecularly because none of the structures allow intra-molecular coordination (Krebs et al., 2001), so we hypothesize that these contacts reflect inter-molecular interactions. None of the inter-chain contact maps is able to match convincingly the contacts observed in the analysis (Supplementary Figure S2), suggesting that in solution there might be different species in mutual equilibrium or that none of the available structure represents the functional species. The first hypothesis is also in agreement with the diversity of packing observed in the crystal structures.

The contacts within C-terminal residues show the characteristic pattern of  $\beta$ -sheets or loop conformations. These patterns could be in agreement with the swapped dimer of 1X0G, where the loop harboring the first cysteine of the CX<sub>n</sub>CGCG motif (Cys37) is bent toward the C-terminus and stabilized by steric hindrance from the swapped central twisted  $\beta$ -sheets. In this structure, cluster coordination is asymmetric and achieved by Cys37 and Cys101 of the  $\alpha$  protomer and Cys103 of the  $\beta$  protomer. The evolutionary trace of contacts between the C-terminus (residues 98–112) and the loop between residues 33–41 (Figure 5E) suggests the existence of a conformation which allows the proximity of the first cysteine (Cys37) to the terminal cysteine pair (Cys101 and Cys103) (Supplementary Figure S3), supporting a contribution of Cys37 in cluster coordination. This conclusion is strongly at variance with the previous belief that only the C-terminal cysteines participate in coordination and implies that cluster coordination can occur at the level of the dimer without invoking the formation of a tetramer. The 1X0G structure is currently the only available structure able to describe cluster coordination, although domain swapping may not be required to explain the interactions: domain swapping could easily be replaced by a non-swapped protomer in a symmetric dimer (Morimoto et al., 2006).

We can thus conclude that DCA of IscA suggests important new hypotheses that can change drastically our views on the coordination properties of this protein cluster.

## Protein–Protein Interactions

DCA can in principle be extended to predict conserved contacts between interacting proteins on the basis of MSAs of protein pairs that are known to interact. In the absence of such a curated

set, several matching strategies have been developed (Hopf et al., 2014; Ovchinnikov et al., 2014; Bitbol et al., 2016; Gueudré et al., 2016). Among these, two independent implementations have recently been suggested in back-to-back publications (Bitbol et al., 2016; Gueudré et al., 2016). We adopted the Iterative Paralog Matching (IPA) method (Bitbol et al., 2016) to investigate the interactions between frataxin, IscU, and IscS. IPA is an iterative process that allows finding matchings between paralogs of two protein families in an organism by maximizing the inter-protein co-evolutionary signal. Briefly, for each organism the retrieved sequences of the first protein are randomly matched with the sequences of the other protein and the first MSA is built. Mean-Field DCA is used to infer the model and the resulting couplings are used to score all possible matchings of paralogs. The pairs who show the highest inter-protein co-evolution score are added to the MSA, which is then fed as input to MF-DCA for the next iteration. This procedure is repeated until all paralog pairs are matched. The resulting matched MSAs can be used to perform standard DCA and record the strongest inter-protein contacts (see the “Iterative Paralog Matching and inter-protein predictions” section of Materials and Methods). We performed multiple IPA runs and scored protein-protein contacts according to the number of times they are accepted among all the runs (acceptance frequency). The most frequently predicted contacts were then selected for further analysis. We first analyzed the interactions between IscU and IscS, because a high-resolution crystal structure of this complex is available (3LVL). We observed that the four most often accepted contacts do indeed lie in the interface of the IscU-IscS dimer. These contacts have acceptance frequencies between 100 and 85% (**Figure 6A**, Supplementary Figure S4). Contacts with lower acceptance frequencies are mainly incompatible with the structural model of the IscU-IscS dimer (i.e., false-positives). We also observed at least one contact (V17-L383, accepted in 17% of IPA runs) that lies in the IscU-IscS interface. In the absence of an absolute scale quantifying the reliability of predicted contacts, and of known structures for the IscU-frataxin and IscS-frataxin complexes, we used the IscU-IscS case as a reference. We assumed that contacts being accepted in more than 85% of IPA simulations would be in excellent agreement with an experimental model, while contacts with lower acceptance frequency display high variability and false positive rates.

We observed absence of contacts with high acceptance frequency for the IscS-frataxin pair (compared to the IscU-IscS case) (**Figure 6B**, Supplementary Figure S5A). The 68% acceptance frequency of the two most frequent contacts falls in the range where, in the case of IscU-IscS, most contacts are false positives. Therefore, even though the two contacts have geometrical compatibility, i.e., they could in principle be satisfied by a docked pose, their high statistical uncertainty prevents drawing conclusions about their biological relevance.

In the case of interactions between frataxin and IscU, IPA identified three contacts with very high acceptance frequencies (>94%) (**Figure 6C**, Supplementary Figure S5B) and potential geometric compatibility with a docked complex. However, there is no overlap between these three co-evolutionary predicted contacts and the interaction interface between frataxin and

IscU in an available model of the IscU-IscS-frataxin trimer (di Maio et al., 2016). It must, however, be noted that the number of sequences in the IscU and IscS families are significantly higher than for the frataxin family. This should contribute to a higher statistical robustness of the predictions for the IscU-IscS complex.

## DISCUSSION

DCA is a powerful method, by now shown to be robust and reliable as long as a sufficiently high number of independent protein sequences are available (Morcos et al., 2011; Marks et al., 2012; Ekeberg et al., 2013). In this work, we have interrogated evolution through DCA to gain new insights into the molecular machinery involved in FeS cluster biosynthesis. We selected three essential components: the scaffold protein IscU, the alternative scaffold IscA and the regulator of cluster formation, CyaY/frataxin. Apart from the medical and biological interest of the latter, the choice of CyaY/frataxin turned out to validate the method for our purposes since this protein has a well compact and stable fold with high structural conservation. The smaller number of sequences available for CyaY/frataxin reflects the origin of this protein, which goes back only to the alpha-beta-gamma proteobacteria (Huynen et al., 2001). In contrast, IscU is at least 200 million years older (Hwang et al., 1996). Nonetheless, we observed that, despite the relatively small number of sequences, we can reproduce most features of the CyaY/frataxin fold. This gives us confidence with the other two much better represented proteins. We then applied DCA to resolve questions that could allow us to understand cluster coordination and assembly of the other two proteins.

Much has been said about the presence of partially unstructured structures of IscU which could be in equilibrium with the fully folded form in solution (Markley et al., 2013). There is no doubt that IscU is a marginally stable protein: in the absence of partners like zinc, the FeS cluster or IscS, it is prone to unfold not only at high but also at low temperatures (Iannuzzi et al., 2014). The N-terminus is either flexible or in a conformational exchange in solution even in the presence of zinc. We do not find traces of the unstructured conformation in our analysis, while the signal from the structured form is clear and unmistakable. Even more interestingly, we found for the first time indications that directly support the reported existence of a head-to-head IscU dimer whose interface would involve the conserved cysteines (Chandramouli et al., 2007). This dimer was suggested to be the result of an oxidative event occurring in the later stages of FeS cluster formation, after the cluster-loaded IscU has detached from IscS (Agar et al., 2000; Chandramouli et al., 2007). IscU dimerization agrees with the consideration that the only way to reach sufficient coordination groups and enable formation of the 4Fe4S cubane, which would instead be too unstable to be coordinated by the IscU monomer, the the formation of a dimer (Iannuzzi et al., 2014). This event, so far only inferred indirectly, leads to the formation of a 4Fe4S cluster of which we now observe a direct indication.

DCA of IscA suggests new hypotheses about the structure of this otherwise still obscure protein. Because IscA binds both iron



and FeS clusters, the protein has alternatively been suggested to be a scaffold protein or the carrier protein that delivers iron to the desulfurase (Krebs et al., 2001; Ollagnier-De-Choudens et al., 2001; Ding et al., 2007). What is certain is that IscA contains three conserved cysteines, which are excellent candidates for both ion and cluster coordination. The crystal structures of IscA have been relatively uninformative about the type of molecular assembly and cluster/metal coordination. Our DCA data rely on a large number of sequences, just a little bit inferior to those retrieved for IscU. We observed a signal that is compatible with formation of the  $\alpha\beta$  fold observed in all available structures. However, we also observed contacts that cannot easily be explained by only one structure, suggesting the presence of several different species, at least in the absence of a cluster or cations. This is consistent with experimental evidence (Popovic and Pastore, 2016), which clearly supports the presence of an equilibrium between at least two species in a range of concentrations compatible with those expected in the cell. After analyzing different structures we conclude that the co-presence of structures such as 1X0G and 1R95 would match what we observe in the DCA analysis. These conclusions strongly suggest that, while not necessarily giving domain swapping, cluster coordination can be mediated by the dimeric form of IscA rather than the tetramer.

Finally, we applied a recent DCA application (Bitbol et al., 2016) to investigate the binary complexes between frataxin, IscU, and IscS. Prediction of the IscS-IscU interface is in excellent agreement with the crystal structures (Shi et al., 2010). In contrast, predictions of frataxin-IscS interactions did not display results with sufficient robustness to allow strong conclusions. They led to three strong signals which are incompatible with the experimental interaction-interface (di Maio et al., 2016). This co-evolutionary analysis raises intriguing questions about the uniqueness of the frataxin-IscU interaction, which calls for more extended experimental and computational investigations.

In conclusion, we found that DCA is a methodology which can enhance our knowledge of specific protein families and provide new information that can address unresolved questions. We can thus confidently add DCA to the tools that can allow us to study the FeS cluster machinery.

## MATERIALS AND METHODS

### Multiple Sequence Alignments

Multiple sequence alignments (MSAs) for each of the studied protein families were constructed using the following protocol: We first gathered all sequences from Uniprot with gene names corresponding to the canonical members of the families (CYAY or FXN for frataxin, ISCA for IscA, ISCS for IscS, ISCU for IscU). We then aligned the sequences of each seed using MAFFT (<http://mafft.cbrc.jp/alignment/software/>) (Katoh, 2002). The resulting MSA was then used to generate a Hidden Markov Model using the HMMER package (<http://hmmer.org/>) (Finn et al., 2011). The Uniprot database was then searched using the HMMs to extract homologous sequences. The resulting MSAs were further filtered, removing all sequences containing more than 10% of gapped positions.

### Direct Coupling Analysis

DCA (Weigt et al., 2009; Morcos et al., 2011) was performed using an in-house code of the asymmetric version of the Pseudo-likelihood method to infer the parameters of the Potts model (Balakrishnan et al., 2011; Ekeberg et al., 2013). Sequences were reweighed using a maximum 90% identity threshold. L2 regularization parameters were used (Ekeberg et al., 2013). This is achieved by fitting the parameters  $h_i$  and  $J_{ij}$  of the generalized Potts model to the sequences in a MSA according to Equation 1:

$$P(X) = \frac{1}{Z} \exp\left[\sum_i^N h_i(X_i) + \sum_{i,j}^{N,N} J_{ij}(X_i, X_j)\right] \quad (1)$$

where  $X$  is a sequence of the MSA and  $Z$  is a normalizing constant, known as the partition function in statistical physics. The sequences were reweighed using a maximum 90% identity threshold to partially remove phylogenetic and sampling biases in the MSA. A standard L2 regularization was added to the learning procedure of the parameters, with the original regularization weights of Ekeberg et al. (2013) ( $\lambda = 0.01$ ). We used the scoring scheme for DCA contacts introduced in Markley et al. (2013). Specifically, the DCA scores  $S_{ij}$  were computed as the Frobenius norm of the local coupling matrices  $J_{ij}$  of the Potts model. The modifications introduced in Feinauer et al. (2014) were adopted, which consist in ignoring couplings with gaps in the local  $J_{ij}$  matrices. This modification of the original Frobenius norm scoring scheme improves the prediction quality, by removing non-functional predictions raising from strong correlations in MSAs introduced by the presence of long gap stretches (Feinauer et al., 2014). The average product correction (APC) term was subtracted (Dunn et al., 2008). The  $N$  top scoring predictions ( $N$  being fixed as the MSA sequence length) were compared with the contact map of reference structures in which two residues were considered to be in contact if they have at least one heavy-atom less than 8.5 Å apart. Different cut-offs do not modify the results but only the interactions we consider. Decreasing the cut-off results in more stringent interactions but also in loss of information about the neighborhood (Supplementary Figure S6). We ignored DCA predictions between residue pairs separated by less than five positions along the chain to favor visualization of long-range contact interactions.

### Iterative Paralog Matching and Inter-Protein Predictions

To perform DCA analysis of pairs of (putatively) interacting protein families, a concatenated MSAs consisting of interacting sequences in families A and B must first be built. The main challenge rises when multiple paralogs of proteins A and B are present in an organism. In this case, it is not straightforward to match correctly interacting paralogs for all organisms. To build matched MSAs of two interacting protein families (denoted A and B), we used the Iterative Paralog Matching (IPA) strategy (Bitbol et al., 2016). The rationale of this procedure is to find matchings between paralogs of two protein families in an organism, such that the inter-protein co-evolutionary signal is self-consistently maximized. The steps of the procedure can be

summarized as follows (see Bitbol et al., 2016 for a detailed analysis and benchmark of the method):

1. An initial random seed is built, such that for each organism, the sequences of protein A are randomly matched with sequences of protein B, yielding the matched MSA of the 0th iteration.
2. Mean-Field DCA (Morcos et al., 2011) is performed to infer the statistical model (Equation 1).
3. The inferred inter-protein coupling  $J_{ij}^{\text{Inter}}$  couplings are used to score all possible matchings of paralogs for all organisms, yielding an inter-protein co-evolution score for each pair of paralogs of families A and B for each organism.
4. All pairs of paralogs of families A and B are then ranked based on their inter-protein co-evolution score.
5. A number  $N_{\text{Select}}$  of the top ranking sequence pairs is added to the MSA matched MSA of the next iteration.
6.  $N_{\text{Select}}$  is increased by 6 at each iteration.
7. Steps 2–6 are repeated until all possible  $N_{\text{Select}} = N_{\text{max}}$ , which ends the iterative procedure and yields the final matched MSA.

Note that the random seed is discarded after the first iteration, therefore at iteration 1, the MSA will contain  $N_{\text{inc}}$  paired sequences, and will be grow by additional  $N_{\text{inc}}$  paired sequences at each iteration. The procedure stops when there are  $N_{\text{max}}$  matched sequence pairs in the MSA, where  $N_{\text{max}}$  is defined by

$$N_{\text{max}} = \sum_{\text{Organisms}} \min(N_A, N_B) \quad (2)$$

where  $N_A$  (resp.  $N_B$ ) denote the number of paralogs of class A (resp. B) in a given organism. This means that for each organism, all sequences belonging to the family with less paralogs in the current organism will be matched to a single paralog of the opposite class.

The procedure is repeated  $N_{\text{IPA}}$  times with different initial random sequence pairings. For each of these  $N_{\text{IPA}}$  MSA, we performed DCA with the Pseudo-Likelihood method described above to perform contact prediction. The strongest inter-protein contacts were recorded for all iterations. To select the strongest co-evolving inter-protein contacts for a given iteration, we used a selection criterion introduced in Hopf et al. (2014), which renormalizes DCA predictions to allow comparison between different protein families:

$$\tilde{S}_{ij} = \frac{S_{ij}}{\left| \min(S_{i,j}^{\text{Inter}}) \right| \left( 1 + \sqrt{\frac{N}{N_{\text{eff}}}} \right)} \quad (3)$$

where  $S_{ij}$  is the APC corrected score described above,  $N$  the length of the MSA and  $N_{\text{eff}}$  the effective number of sequences in the MSA (which takes into account the weights of the sequences).  $\text{Min}$  stands for the minimum over all inter-protein residue pairs. Note that the minimum is only taken over the inter-protein scores  $S_{ij}^{\text{Inter}}$ . This normalized score (Equation 3) partially removes the dependency of the scores on the length of the protein and on the depth of the alignments. As discussed in Ovchinnikov et al. (2014), to account for possible variations in

evolutionary rates between the two interacting protein families, the average product correction is taken asymmetrically, i.e., the two averages over positions  $(i,j)$  in the APC are taken over the two protein families separately (see Ovchinnikov et al., 2014 for details). For each DCA calculation, we retained all inter-protein contacts which had a normalized score  $S_{ij}$  above 0.8, a criterion introduced in Hopf et al. (2014). Note that the use of this renormalized scores does not change the ranking of the contacts, i.e., it is an alternative way of selecting the strongest inter-protein co-evolving contacts. It is in fact equivalent to setting a threshold on the number of contacts to select, or a threshold on the DCA score.

To obtain an estimate of the robustness of inter-protein DCA predictions, we ranked all possible inter-protein contacts by the normalized number of times they were accepted in the  $N_{\text{IPA}}$  iterations (acceptance frequency). Contacts being accepted more often across several IPA runs should reflect more robustness and higher statistical significance.

We used  $N_{\text{IPA}} = 200$  iterations for the IscU-IscS system, and  $N_{\text{IPA}} = 300$  for the frataxin-IscU and frataxin-IscS systems. The inter-protein structural contact maps were built with the same contact threshold as intra-protein maps, i.e., inter-protein pairs of residues were considered to be in contact if at least one pair of heavy-atoms between the two were less than 8.5 Å apart.

## Shortest-Path Analysis

We used a shortest-path (SP) analysis introduced in (Malinverni et al., 2015) to quantify the agreement between DCA predictions and structural contacts. The shortest path for a DCA-predicted contact between residue  $i$  and  $j$  is defined as the minimal number of contacts in the structural contact map needed to join these two residues. DCA-predicted contacts which are native contacts in the structural map therefore have an SP of one. This analysis helps to highlight the mediation of contacts, and is a natural measure of the topological propagation of contact information (see Malinverni et al., 2015 for further details).

## AUTHOR CONTRIBUTIONS

MF did the experiments, analyzed the data and wrote a first draft. DM did the experiments and analyzed the data. PD analyzed the data. AP analyzed the data and wrote the final version.

## FUNDING

The research described from AP lab was supported by MRC (U117584256). DM thanks the Swiss National Science Foundation (<http://www.snf.ch/>) for grants 2012\_149278 & 20020\_163042/1.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2017.00040/full#supplementary-material>

## REFERENCES

- Adinolfi, S., Iannuzzi, C., Prisch, F., Pastore, C., Iametti, S., Martin, S. R., et al. (2009). Bacterial frataxin CyaY is the gatekeeper of iron-sulfur cluster formation catalyzed by IscS. *Nat. Struct. Mol. Biol.* 16, 390–396. doi: 10.1038/nsmb.1579
- Adinolfi, S., Nair, M., Politou, A., Bayer, E., Martin, S., Temussi, P., et al. (2004). The factors governing the thermal stability of frataxin orthologues: how to increase a protein's stability. *Biochemistry* 43, 6511–6518. doi: 10.1021/bi036049+
- Adrover, M., Howes, B. D., Iannuzzi, C., Smulevich, G., and Pastore, A. (2015). Anatomy of an iron-sulfur cluster scaffold protein: understanding the determinants of [2Fe-2S] cluster stability on IscU. *Biochim. Biophys. Acta Mol. Cell Res.* 1853, 1448–1456. doi: 10.1016/j.bbamcr.2014.10.023
- Agar, J. N., Krebs, C., Frazzon, J., Huynh, B. H., Dean, D. R., and Johnson, M. K. (2000). IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry* 39, 7856–7862. doi: 10.1021/bi000931n
- Altschuld, D., Leskx, A. M., Bloomer, A. C., and Klug, A. (1987). Correlation of coordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. 193, 693–707.
- Balakrishnan, S., Kamisetty, H., Carbonell, J. G., Lee, S. I., and Langmead, C. J. (2011). Learning generative models for protein fold families. *Proteins* 79, 1061–1078. doi: 10.1002/prot.22934
- Beilschmidt, L. K., and Puccio, H. M. (2014). Mammalian Fe-S cluster biogenesis and its implication in disease. *Biochimie* 100, 48–50. doi: 10.1016/j.biochi.2014.01.009
- Bilder, P. W., Ding, H., and Newcomer, M. E. (2004). Crystal structure of the ancient, Fe-S scaffold IscA reveals a novel protein fold. *Biochemistry* 43, 133–139. doi: 10.1021/bi035440s
- Bitbol, A. F., Dwyer, R. S., Colwell, L. J., and Wingreen, N. S. (2016). Inferring interaction partners from protein sequences. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12180–12185. doi: 10.1073/pnas.1606762113
- Bothe, J. R., Tonelli, M., Ali, I. K., Dai, Z., Frederick, R. O., Westler, W. M., et al. (2015). The complex energy landscape of the protein IscU. *Biophys. J.* 109, 1019–1025. doi: 10.1016/j.bpj.2015.07.045
- Chandramouli, K., Uncileac, M. C., Naik, S., Dean, D. R., Boi, H. H., and Johnson, M. K. (2007). Formation and properties of [4Fe-4S] clusters on the IscU scaffold protein. *Biochemistry* 46, 6804–6811. doi: 10.1021/bi6026659
- Cupp-Vickery, J. R., Silberg, J. J., Ta, D. T., and Vickery, L. E. (2004). Crystal structure of IscA, an iron-sulfur cluster assembly protein from *Escherichia coli*. *J. Mol. Biol.* 338, 127–137. doi: 10.1016/j.jmb.2004.02.027
- Dago, A. E., Schug, A., Procaccini, A., Hoch, J. A., Weigt, M., and Szurmant, H. (2012). Structural basis of histidine kinase autophosphorylation deduced by integrating genomics molecular dynamics and mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1733–E1742. doi: 10.1073/pnas.1201301109
- Ding, H., Yang, J., Coleman, L. C., and Yeung, S. (2007). Distinct iron binding property of two putative iron donors for the iron-sulfur cluster assembly: IscA and the bacterial frataxin ortholog CyaY under physiological and oxidative stress conditions. *J. Biol. Chem.* 282, 7997–8004. doi: 10.1074/jbc.M609665200
- di Maio, D., Chandramouli, B., Yan, R., Brancato, G., and Pastore, A. (2016). Understanding the role of dynamics in the iron sulfur cluster molecular machine. *Biochim. Biophys. Acta* 1861, 3154–3163. doi: 10.1016/j.bbagen.2016.07.020
- Dunn, S. D., Wahl, L. M., and Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* 24, 333–340. doi: 10.1093/bioinformatics/btm604
- Ekeberg, M., Lökvist, C., Lan, Y., Weigt, M., and Aurell, E. (2013). Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Phys. Rev. E* 87:012707. doi: 10.1103/PhysRevE.87.012707
- Espada, R., Parra, R. G., Mora, T., Walczak, A. M., and Ferreiro, D. U. (2015). Capturing coevolutionary signals in repeat proteins. *BMC Bioinformatics* 16:6483. doi: 10.1186/s12859-015-0648-3
- Feinauer, C., Skwark, M. J., Pagnani, A., and Aurell, E. (2014). Improving contact prediction along three dimensions. *PLoS Comput. Biol.* 10:3847. doi: 10.1371/journal.pcbi.1003847
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Gakh, O., Bedekovics, T., Duncan, S. F., Smith, I. V., Berkholz, D. S., and Isaya, G. (2010). Normal and *Friedreich ataxia* cells express different isoforms of frataxin with complementary roles in iron-sulfur cluster assembly. *J. Biol. Chem.* 285, 38486–38501. doi: 10.1074/jbc.M110.145144
- Gibson, T. J., Koonin, E. V., Musco, G., Pastore, A., and Bork, P. (1996). Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction. *Trends Neurosci.* 19, 465–468. doi: 10.1016/S0166-2236(96)20054-2
- Göbel, U., Sander, C., Schneider, R., and Valencia, A. (1994). Correlated mutations and residue contacts in proteins. *Proteins Struct. Funct. Genet.* 18, 309–317. doi: 10.1002/prot.340180402
- Guedré, T., Baldassi, C., Zamparo, M., Weigt, M., and Pagnani, A. (2016). Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12186–12191. doi: 10.1073/pnas.1607570113
- Hopf, T. A., Colwell, L. J., Sheridan, R., Rost, B., Sander, C., and Marks, D. S. (2012). Theory three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 149, 1607–1621. doi: 10.1016/j.cell.2012.04.012
- Hopf, T. A., Morinaga, S., Ihara, S., Touhara, K., Marks, D. S., and Benton, R. (2015). Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. *Nat. Commun.* 6:6077. doi: 10.1038/ncomms7077
- Hopf, T. A., Schärfe, C. P., Rodrigues, J. P., Green, A. G., Kohlbacher, O., Sander, C., et al. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* 3:e03430. doi: 10.7554/eLife.03430
- Huynen, M. A., Snel, B., Bork, P., and Gibson, T. J. (2001). The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly. *Hum. Mol. Genet.* 10, 2463–2468. doi: 10.1093/hmg/10.21.2463
- Hwang, D. M., Dempsey, A., Tan, K. T., Liew, C. C. (1996). A modular domain of NifU, a nitrogen fixation cluster protein, is highly conserved in evolution. *J. Mol. Evol.* 43, 536–540. doi: 10.1007/BF02337525
- Iannuzzi, C., Adinolfi, S., Howes, B. D., Garcia-Serres, R., Clémancey, M., Latour, J. M., et al. (2011). The role of cyay in iron sulfur cluster assembly on the *e. coli* iscu scaffold protein. *PLoS ONE* 6:e21992. doi: 10.1371/journal.pone.0021992
- Iannuzzi, C., Adrover, M., Puglisi, R., Yan, R., Temussi, P. A., and Pastore, A. (2014). The role of zinc in the stability of the marginally stable IscU scaffold protein. *Protein Sci.* 23, 1208–1219. doi: 10.1002/pro.2501
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kaut, A., Lange, H., Diekert, K., Kispal, G., and Lill, R. (2000). Isa1p is a component of the mitochondrial machinery for maturation of cellular iron-sulfur proteins and requires conserved cysteine residues for function. *J. Biol. Chem.* 275, 15955–15961. doi: 10.1074/jbc.M909502199
- Kensche, P. R., Duarte, I., and Huynen, M. A. (2012). A three-dimensional topology of complex I inferred from evolutionary correlations. *BMC Struct. Biol.* 12:19. doi: 10.1186/1472-6807-12-19
- Krebs, C., Agar, J. N., Smith, A. D., Frazzon, J., Dean, D. R., Huynh, B. H., et al. (2001). IscA, an alternate scaffold for Fe-S cluster biosynthesis. *Biochemistry* 40, 14069–14080. doi: 10.1021/bi015656z
- Liu, J., Oganessian, N., Shin, D. H., Jancarik, J., Yokota, H., Kim, R., et al. (2005). Structural characterization of an iron-sulfur cluster assembly protein IscU in a zinc-bound form. *Proteins Struct. Funct. Genet.* 59, 875–881. doi: 10.1002/prot.20421
- Malinverni, D., Marsili, S., Barducci, A., and de Los Rios, P. (2015). Large-scale conformational transitions and dimerization are encoded in the amino-acid sequences of Hsp70 chaperones. *PLoS Comput. Biol.* 11:e262. doi: 10.1371/journal.pcbi.1004262
- Marinoni, E. N., Oliveira, J. S., De Nicolet, Y., Raulfs, E. C., Amara, P., Dean, D. R., et al. (2012). (IscS-IscU) 2 complex structures provide insights into Fe 2S 2 biogenesis and transfer. *Angew. Chem. Int. Ed. Engl.* 51, 5439–5442. doi: 10.1002/anie.201201708
- Markley, J. L., Kim, J., Dai, Z., Bothe, J. R., Cai, K., Frederick, R. O., et al. (2013). Metamorphic protein IscU alternates conformations in the course of its role as the scaffold protein for iron-sulfur cluster biosynthesis and delivery.

- FEBS Lett. Federation of European Biochemical Societies; 587, 1172–1179. doi: 10.1016/j.febslet.2013.01.003
- Marks, D. S., Hopf, T. A., and Sander, C. (2012). Protein structure prediction from sequence variation. *Nat. Biotechnol.* 30, 1072–1080. doi: 10.1038/nbt.2419
- Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., et al. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1293–E1301. doi: 10.1073/pnas.1111471108
- Morimoto, K., Yamashita, E., Kondou, Y., Lee, S. J., Arisaka, F., Tsukihara, T., et al. (2006). The asymmetric IscA homodimer with an exposed [2Fe-2S] cluster suggests the structural basis of the Fe-S cluster biosynthetic scaffold. *J. Mol. Biol.* 360, 117–132. doi: 10.1016/j.jmb.2006.04.067
- Nair, M., Adinolfi, S., Pastore, C., Kelly, G., Temussi, P., and Pastore, A. (2004). Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites. *Structure* 12, 2037–2048. doi: 10.1016/j.str.2004.08.012
- Ollagnier-De-Choudens, S., Mattioli, T., Takahashi, Y., and Fontecave, M. (2001). Iron-sulfur cluster assembly. Characterization of IscA and evidence for a specific and functional complex with ferredoxin. *J. Biol. Chem.* 276, 22604–22607. doi: 10.1074/jbc.M102902200
- Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* 2014, 1–21. doi: 10.7554/eLife.02030
- Pastore, C., Franzese, M., Sica, F., Temussi, P., and Pastore, A. (2007). Understanding the binding properties of an unusual metal-binding protein—A study of bacterial frataxin. *FEBS J.* 274, 4199–4210. doi: 10.1111/j.1742-4658.2007.05946.x
- Pazos, F., Helmer-Citterich, M., Ausiello, G., and Valencia, A. (1997). Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.* 271, 511–523. doi: 10.1006/jmbi.1997.1198
- Popovic, M., and Pastore, A. (2016). Chemical shift assignment of the alternative scaffold protein IscA. *Biomol. NMR Assign.* 10, 227–231. doi: 10.1007/s12104-016-9672-0
- Popovic, M., Sanfelice, D., Pastore, C., Prischi, F., Temussi, P. A., and Pastore, A. (2015). Selective observation of the disordered import signal of a globular protein by in-cell NMR protein by in-cell NMR: the example of frataxins. *Protein Sci.* 24, 996–1003. doi: 10.1002/pro.2679
- Prischi, F., Giannini, C., Adinolfi, S., and Pastore, A. (2009). The N-terminus of mature human frataxin is intrinsically unfolded. *FEBS J.* 276, 6669–6676. doi: 10.1111/j.1742-4658.2009.07381.x
- Prischi, F., Konarev, P. V., Iannuzzi, C., Pastore, C., Adinolfi, S., Martin, S. R., et al. (2010). Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nat. Commun.* 1:95. doi: 10.1038/ncomms1097
- Procaccini, A., Lunt, B., Szurmant, H., Hwa, T., and Weigt, M. (2011). Dissecting the specificity of protein-protein interaction in bacterial two-component signaling: orphans and crosstalks. *PLoS ONE* 6:19729. doi: 10.1371/journal.pone.0019729
- Rouault, T. A. (2015). Mammalian iron-sulphur proteins: novel insights into biogenesis and function. *Nat. Rev. Mol. Cell. Biol.* 16, 45–55. doi: 10.1038/nrm3909
- Shi, R., Proteau, A., Villarroya, M., Moukadiri, I., Zhang, L., Trempe, J. F., et al. (2010). Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-protein interactions. *PLoS Biol.* 8:354. doi: 10.1371/journal.pbio.1000354
- Sutto, L., Marsili, S., Valencia, A., and Gervasio, F. L. (2015). From residue coevolution to protein conformational ensembles and functional dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13567–13572. doi: 10.1073/pnas.1508584112
- Toth-petroczy, A., Palmedo, P., Ingraham, J., Hopf, T. A., Berger, B., Sander, C., et al. (2016). Structured states of disordered proteins from genomic sequences. *Cell* 167, 158–170.e12. doi: 10.1016/j.cell.2016.09.010
- Tsai, C. L., and Barondeau, D. P. (2010). Human frataxin is an allosteric switch that activates the Fe-S cluster biosynthetic complex. *Biochemistry* 49, 9132–9139. doi: 10.1021/bi1013062
- Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A.* 106, 67–72. doi: 10.1073/pnas.0805923106

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Fantini, Malinverni, De Los Rios and Pastore. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.