Cellular and Molecular Life Sciences (2019) 76:4145–4154
https://doi.org/10.1007/s00018-019-03108-2

ORIGINAL ARTICLE

Network mapping of the conformational heterogeneity of SOD1 by deploying statistical cluster analysis of FTIR spectra

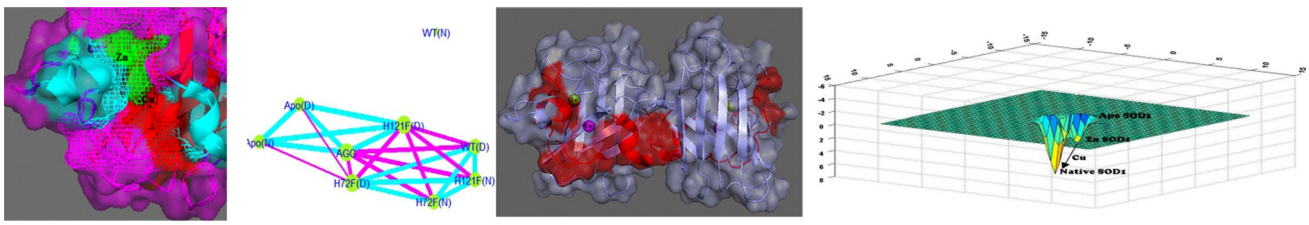
Sourav Chowdhury[1,2] · Sagnik Sen[3] · Amrita Banerjee[1,6] · Vladimir N. Uversky[4,5] · Ujjwal Maulik[2] · Krishnananda Chattopadhyay[1]

Received: 23 October 2018 / Revised: 12 April 2019 / Accepted: 15 April 2019 / Published online: 22 April 2019

Abstract

A crucial contribution to the heterogeneity of the conformational landscape of a protein comes from the way an intermediate relates to another intermediate state in its journey from the unfolded to folded or misfolded form. Unfortunately, it is extremely hard to decode this relatedness in a quantifiable manner. Here, we developed an application of statistical cluster analyses to explore the conformational heterogeneity of a metalloenzyme, human cytosolic copper–zinc superoxide dismutase (SOD1), using the inputs from infrared spectroscopy. This study provides a quantifiable picture of how conformational information at one particular site (for example, the copper-binding pocket) is related to the information at the second site (for example, the zinc-binding pocket), and how this relatedness is transferred to the global conformational information of the protein. The distance outputs were used to quantitatively generate a network capturing the folding sub-stages of SOD1.

Graphical abstract



Keywords ALS · Superoxide dismutase · Infrared spectroscopy · Protein folding · Correlation analysis · Cluster dendrogram · Protein aggregation

Sourav Chowdhury, Sagnik Sen, and Amrita Banerjee are contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00018-019-03108-2) contains supplementary material, which is available to authorized users.

✉ Krishnananda Chattopadhyay
krish@iicb.res.in

[1] Protein Folding and Dynamics Laboratory, Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata 700032, India

[2] Present Address: Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

[3] Department of Computer Science, Jadavpur University, Kolkata 700 032, India

[4] Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, FL, USA

[5] Laboratory of New Methods in Biology, Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino 142290, Moscow Region, Russia

[6] Department of Chemistry, Hiralal Mazumdar Memorial College for Women, Dakshineswar, Kolkata 700035, India

Introduction

Protein-folding studies typically use model systems of relatively low complexity owing to limited experimental resolution and incomplete theoretical understandings. However, the focus is slowly shifting towards more complex protein systems [1, 2]. This transition is needed not only to gather fundamental understanding of the protein-folding problem, but also to obtain greater insights into protein misfolding. Conformational landscapes of large complex proteins can be heterogeneous. Often, there exists considerable overlap between folding and aggregation landscapes [3]. It is non-trivial to determine how different intermediates, which can be accumulated in the folding/aggregation landscape of a protein, relate to each other in a statistical and/or quantifiable manner. It is also not understood how a change in one structural form can be correlated with the variation in other forms of secondary structure in different sub-stages of protein folding and/or aggregation. Here, we studied the relatedness of folding intermediates and aggregated states of superoxide dismutase (SOD1), a key participant in the anti-oxidant defence mechanism. Misfolding and aggregation of SOD1 are implicated in amyotrophic lateral sclerosis (ALS) [4].

SOD1 is a homo-dimeric protein having a mosaic of structural elements at the secondary level [5]. Each of the monomers contains an eight-stranded anti-parallel Greek key β-barrel motif and several loops, with loops IV and VII having an extended stretch (Fig. 1a). An intra-molecular disulphide bond links Cys57 (loop IV) and Cys146 (loop VII). The concerted occupancy of the metal co-ordination sites and the chaperone-mediated entry of Cu responsible for the inner structural corrugation leads to restricted mobility of the two extended loops in SOD1 [6–8]. Among the two metal ion co-factors, copper fluctuating between its Cu^{2+} and Cu^{+} forms makes up the catalytic centre, which is deeply buried [9]. The Cu^{2+} ion remains weakly bound to water molecule and is anchored by four histidine residues, His46, His47, His63, and His120 [10] (Fig. 1b). The zinc ion is co-ordinated by three histidine residues (His63, His71, and His80) and one aspartate residue (Asp83) [11] (Fig. 1b). There are reports suggesting that the absence of metal ion co-factors can lead to SOD1 oligomerization [12–14]. Further reports validate these findings in cellular contexts [6].

In this study, we investigated specifically how conformational information at one particular site of a protein (for example, the copper pocket) is related to the information at the second site (for example, the zinc pocket) and how this relatedness is transferred to the global conformational information of the folded protein. As readout of protein conformation, we used the information on the secondary structure determined using Fourier-transform infrared spectroscopy (FTIR). The amide I region of the IR spectrum, which corresponds to the C=O stretch, contains discrete secondary structural fingerprints. We resorted to discrete metal mutants, such as H121F (Cu-deficient SOD1 with Zn co-ordination intact) and H72F (Zn-deficient SOD1 with Cu co-ordination preserved) along with the de-metallated SOD1 (apo-SOD1, both metals removed). FTIR measurements were also carried out using guanidinium hydrochloride (GdnCl)-unfolded proteins. Finally, FTIR measurements were performed using the protein samples, which were aggregated utilizing mechanical agitation (discussed in "Materials and methods"). Figure 2a, b shows the FTIR spectra of WT and different mutants of SOD1, which were obtained at different solution conditions.

The FTIR spectral outputs were subjected to statistical cluster analysis and distance calculations to have a

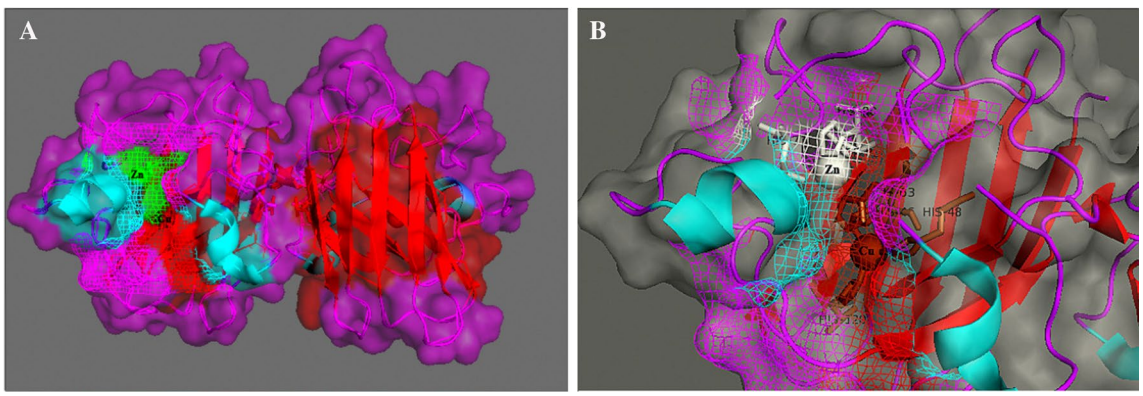


Fig. 1 **a** Transparent surface representation of the crystal structure of dimeric SOD1 (PDB: 4FF9). β-Strands have been shaded with red, α-helical stretches with blue, and loop regions with purple. Cu and Zn metal ion co-factor co-ordination pockets have been shaded with red and green mesh, respectively. **b** Zn (white sphere) and Cu (brick-red sphere) micro-environments have been shown with their respective amino acid co-ordinations

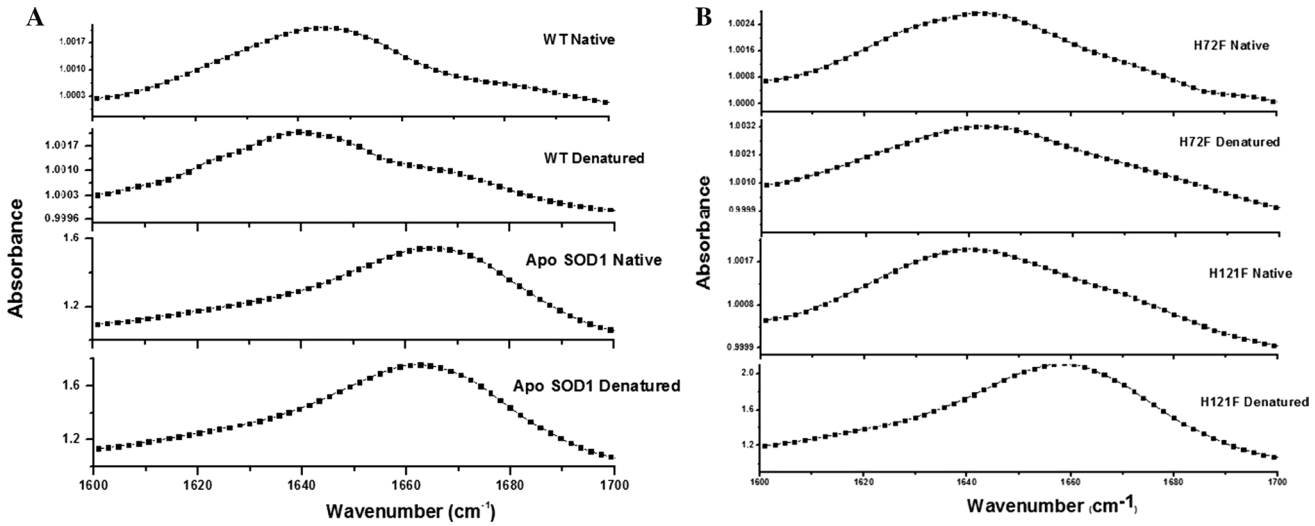


Fig. 2 **a** Amide I (1600–1700 cm^{-1}) FTIR spectra of WT (native and denatured) and apo-SOD1 (native and denatured). **b** Amide I FTIR spectral distribution for native and denatured H72F (Zn-deficient mutant) and H121F (Cu-deficient mutant)

quantitative idea about the relatedness of one protein form to the other. There are reports, where cluster analysis has been used to understand the dynamics associated with non-native trimeric assemblies of SOD1 and its implication on ALS [15, 16]. The raw FTIR data (100 points between 1600 and 1700 cm^{-1}) obtained at every wavenumber were used for the correlation analysis. The linear relationship between *x*- and *y*-axes was used to determine the Pearson's correlation coefficient (*r*), and the colour scale indexes the degree of correlation, where a trending value towards + 1 represents an increasing relationship (increasing correlation), while value trending towards − 1 depicts a negative relationship (anti-correlation).

Materials and methods

Recombinant SOD1 purification

Recombinant SOD1 was over-expressed in *E. coli* (BL21 DE3 strain). The over-expression of SOD1 was induced with 1 M isopropyl-1-thio-*β*-D-galactopyranoside (IPTG). The induction was coupled with metalation, where 1 mM $CuSO_4$ was added directly to the Luria–Bertani culture media so as to ensure proper metal loading over the protein. Followed by induction, the cells were allowed to grow for 3.5 h. The cells were pelleted down by centrifuging at 6000 rpm for 15 min at 4 °C followed by re-suspension in pre-chilled lysis buffer (20 mM Tris–HCl + 500 mM NaCl, pH 8.0). After thorough re-suspension in lysis buffer, the cells were subjected to sonication (20 pulses, each of 30 s pulse time, and an interim time frame of 1 min). Unbroken cells and debris were removed by another act of centrifugation at 10,000 rpm for 10 min. The soluble fraction obtained thereafter was carefully removed and allowed to bind to Ni–NTA agarose resin. The Ni–NTA column was washed using 40 ml wash buffer (20 mM Tris–HCl, 500 mM NaCl and 50 mM imidazole, pH 8.0) followed by elution with 20 mM Tris–HCl, 500 mM NaCl and 500 mM imidazole, pH 8.0. The eluted fractions were pulled according to their tentative protein content as per their absorbance at 280 nm. The post-elution fractions were subjected to dialysis in 20 mM Na-phosphate buffer pH 7.5. The protein concentration of the post-dialyzed fraction was estimated by recording absorbance at 280 nm using the molar extinction coefficient ($\varepsilon_{280} = 10{,}000\ M^{-1}\ cm^{-1}$) [11]. His tagged SOD1 has also been used earlier for structural studies [17]. In our studies, we investigated the impact of His tag on SOD1 by comparing the CD spectrometric and FTIR profiles of over-expressed His tagged SOD1 and commercial SOD1 (untagged, Sigma Aldrich). We found comparable results confirming no impact of His tag on the structural dynamics of SOD1 (Figs. S2, S3).

Site-directed mutagenesis

The recombinant plasmid pET-19b containing the gene for hSOD1 with a polyhistidine tag at the N-terminal end was used as a template for mutagenesis using the Quick-Change XL Site-Directed Mutagenesis Kit (Stratagene, USA). The mutagenic primers containing the mutations (shown in bold type) for the replacement of His72 with a phenylalanine residue (H72F) and His121 with a phenylalanine residue (H121F) are provided in Table S1. Both primers were annealed to the same target sequence on opposite strands of pET-19b. The site-directed mutagenesis was performed as described by the manufacturer. The clones used for the

production of the H72F and H121F mutants were confirmed by DNA sequencing. Proteins were expressed in BL21 (DE3) pLysS *E. coli* cells by induction with 0.5 mM PTG at 37 °C for 4 h. Cells were re-suspended in ice-cold 20 mM Tris–HCl, 500 mM NaCl, pH 8.0, containing protease inhibitor (2 mM PMSF) and lysed by sonication. Unbroken cells and debris were removed by centrifugation at 10000*g* for 10 min. Binding of soluble proteins from the supernatant to Ni–NTA agarose resin (Qiagen, Germany) was done overnight. The Ni–NTA flow through was collected for analysis. The Ni–NTA resin was washed with 40 ml of wash buffer (20 mM Tris–HCl pH 8, 500 mM NaCl, 50 mM imidazole). Elution was done with wash buffer containing 500 mM imidazole.

Preparation of apo-SOD1

Apo-enzyme was prepared from the holo-SOD1 by metal chelation following earlier reported protocol (McCord and Fridovich 1969) with some modifications. Holo-SOD1 was subjected to overnight dialysis in 50 mM Na acetate, 10 mM EDTA, pH 3.8 so as to ensure proper removal of metal ions. EDTA was removed by successive dialysis in 50 mM Na-acetate pH 5.2 and in 20 mM Na-phosphate, pH 7.5. The demetalation was ensured with an activity assay of SOD1 indexing photo-oxidation of pyrogallol at pH 8 (Fig. S4). The activity assay profile of apo-SOD1 matched with that of control (in the absence of protein), indicating complete demetalation (Fig. S4). The metal content was probed with atomic absorption spectroscopy and it showed the presence of 0.460 ppm of Cu and 0.342 ppm of Zn in ppm of WT SOD1 sample. The Zn-starved H72F SOD1 mutant contained 0.462 ppm of Cu in ppm of H72F sample. The Cu starved H121F mutant had 0.330 ppm of Zn in ppm of H121F sample. The apo-form on SOD1 did not give any detectable signal in atomic absorption spectroscopy, showing in turn an effective removal of metal ion co-factors on EDTA-based apo-SOD1 preparation.

All the metal mutants (H72F, H121F, and Apo) were spectrometrically probed with CD for their respective secondary structural fingerprints (Fig. S5). Furthermore, the activity assay is also done for respective metal mutants subsequently revealing absolute loss of enzymatic activity (Fig. S4).

Preparation of SOD1 aggregates

WT SOD1, Apo, H72F, and H121F were subjected to mechanical agitation at 200 rpm for 48 h. The seed concentrations for the aggregate preparation were kept 20 µM in 20 mM sodium phosphate buffer at pH 7.5. The seed-protein species were treated with 1.2 µM tris(2-carboxyethyl) phosphine (TCEP).

Fourier-transform infrared spectroscopy (FTIR)

FTIR spectra of WT SOD1, apo-SOD1, and SOD1 mutants in the absence and the presence of denaturants were acquired using Bruker 600 series FTIR spectrometer. Protein samples with concentration of 20 µM were treated with TCEP in 20 mM sodium phosphate buffer at pH 7.5 and incubated for 16 h (in the absence or presence of denaturants) at room temperature before the measurements. All the FTIR measurements were carried out in H_2O buffer, as H_2O does not impact the protein structure as opposed to D_2O [18]. The experiments were carried out in solution, and the buffer baseline was subtracted before taking each spectrum. The spectral readouts were obtained on absorbance mode with a path length of 0.01 mm following standard methodology [19]. The deconvolution of raw spectra in the amide I region (1700–1600 cm^{-1}) was done using least-squares iterative curve fitting to Gaussian/Lorentzian line shapes. MATLAB and Origin 8.5 software were deployed for the curve fitting and second derivative analysis. The assignment of peaks was done using previously described spectral components associated with different secondary structure elements [18, 20]. FTIR spectra were smoothed using Savistky–Golay method [21, 22].

Statistical cluster analyses

SOD1 upon metal site disruption undergoes significant conformational change at the level of secondary structure, which is evident from the results of our FTIR studies. The secondary structure changes in individual protein types as reflected by the FTIR spectral outputs were further subjected to statistical analyses to get an idea about the extent of relatedness among the protein types. Such analyses were done to figure out as to how metal ion co-factors individually decide the structural fate of SOD1 and mark the transition from demetallated apo-SOD1 to the metallated Cu–Zn–SOD1 state.

Correlation analyses were performed with the deconvoluted FTIR data over amide I band to extract the information about the structural correlation among the individual protein types (Table S2). We performed Pearson's correlation analysis (Eq. 1), where the coefficient 'r_p' represents a statistical measure of the strength of a linear relationship between paired data [23]. Secondary structure elements and their relative abundance (as obtained from the deconvolution of FTIR spectra) comprise the paired data set. Correlation coefficient (r_p) depicts the nature of relationships between two set of variables x and y. In case of Pearson's correlation coefficient, r_p is called linear correlation coefficient, which measures the magnitude and nature of the linear relationship between x and y. Coefficient r_p spans within the interval -1 to $+1$, where values trending towards $+1$ represent an increasing relationship, while value trending towards -1

depicts a negative relationship (anti-correlation). The closer the value is to 1 or – 1, the stronger the linear correlation (see Eq. 1):

$$r_{\mathrm{p}} = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\left\{\sum y^2 - \left(\sum y\right)^2\right\}\right]}}, \quad (1)$$

where x and y are the input data sets, n is the number of pairs of data, and r_{p} is the correlation coefficient.

Next, hierarchical clustering was done to identify the closest neighbours (protein variants) as per distance calculations from distance matrices [24]. Distance matrices were constructed using distance calculation algorithm, i.e., Euclidean distance methods, which measure the point-to-point distance between the two sets [25, 26]. In case of Euclidean method, the point-to-point distances were calculated on Euclidean plane:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}, \quad (2)$$

where x_i and y_i are two points from the plane, and $d(x,y)$ is the Euclidean distances between them. More precisely, we had applied root mean square deviation (RMSD) to calculate the cumulative distance among mutants in terms of FTIR scoring. The resultant distance matrix had been utilized for hierarchical clustering.

Agglomerative hierarchical clustering is a bottom–up clustering technique, where individual clusters are joined together to make a single cluster [27]. In this experiment, average linkage clustering method was being applied.

Network analysis

WT SOD1 and its metal variants (H72F and H121F mutants and apo-SOD1) were statistically analysed. However, relatedness with aggregates of each protein is yet to be established. Here, four generated protein aggregates are analysed in terms of relatedness among them. Following the previous techniques, correlation coefficient among the aggregate types is calculated. As the type of relationship is not known, we use correlation methods for both linear and non-linear relations. Subsequently, distances among all eight previously mentioned protein classes and four aggregates are calculated. Depending on the distances among protein types and all protein aggregates, a cluster dendrogram is designed. There are altogether 12 different cluster modules. Among them, eight modules are from four different protein types at two different stages and four clusters are aggregates of each protein type.

From the similarity score on the basis of correlation among protein samples and its aggregates, network N_{w} is designed considering higher rate of similarity. The similarity matrix is converted to weighted adjacency matrix M, where the size of the matrix is dependent on number of mutants considered for the experiments. The ultimate objective is to establish a relation among protein types and their four aggregates. Therefore, in the network $N_{\mathrm{w}} \in (V, E)$, vertex V represents protein clusters or modules, where each module corresponds to one protein type, and edges E represent relatedness between the two protein types, where the width of edge is changing with similarity score, and the size of nodes is changing with the degree centrality of a node. In such network, stages of protein structural transformation can be shown depending on the similarity scores. On the other hand, we can comment on relatedness just seeing width of the edges. It is easy to infer which mono-metal mutant is closer to aggregates.

Results and discussion

Figure 3a shows the results of the cluster analyses of native and GdnCl-unfolded proteins. The corresponding data for the aggregates are shown in Fig. 3b. Cluster analysis was done on FTIR amide 1 spectral output for all the four protein variants under native and denatured conditions. The readouts were taken from multiple experimental accusations. From the spectral distribution of multiple accusations, the estimated rate of standard deviation is 3.5% as quantified after spectral deconvolutions (Fig. S6). Figure 3a clearly shows that the wild-type protein completely correlates with WT SOD1, which is a necessary condition, as they are the same species. On the contrary, WT SOD1 shows least correlation with the apo-form, while the Cu- and Zn mutants maintain intermediate correlations. Row 1 from the bottom in Fig. 3a compares the native apo-SOD1 with seven other SOD1 variants. It reveals that native apo-SOD1 is not much deviant from the unfolded apo-SOD1. On the contrary, native apo-SOD1 and WT SOD1 are very dissimilar, with a correlation coefficient value close to zero (Fig. 3a). The unfolded WT SOD1, unfolded H72F, native H72F, and native H121F variants have strikingly significant positive correlations among themselves and show almost similar extents of correlation with native apo-SOD1. Unfolded H121F has a very high positive correlation with the native apo-SOD1. Row 3 from the bottom clearly shows that native WT SOD1 has no similarity with the other variants and has significantly lowered positive correlation with its own type. Figure 3a shows low positive correlation between apo-SOD1 (unfolded and native) and four other variants, such as unfolded WT SOD1, unfolded H72F, native H72F, and native H121F. The amide I spectral data of the aggregated species [Fig S1 (ii)] on

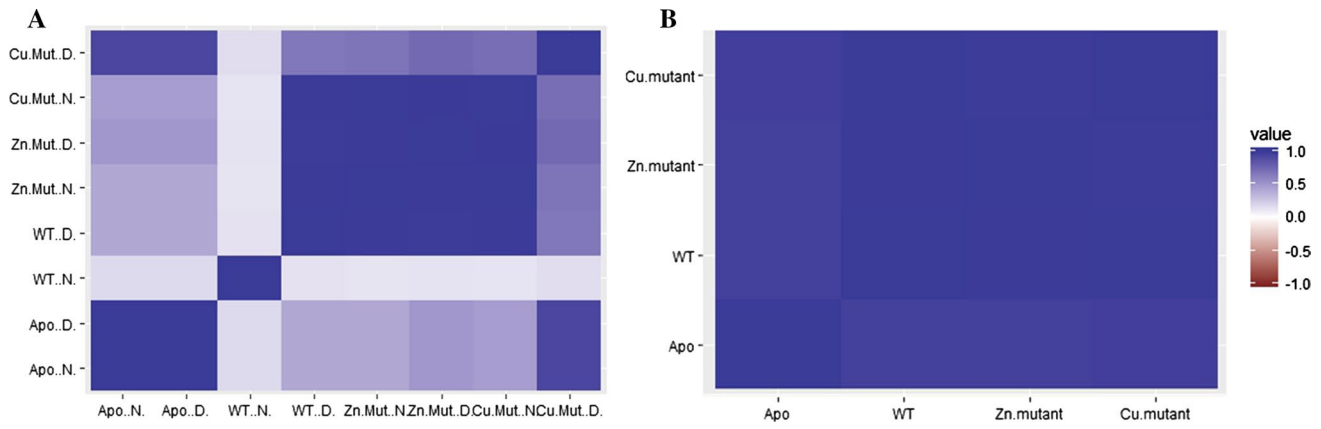


Fig. 3 **a** Combined correlation matrix built on Pearson's correlation algorithm for the raw FTIR spectral data at the amide I range (between 1600 and 1700 cm^{-1}) using native proteins (WT N, Apo-N, Zn Mut N, and Cu Mut N) and their denatured forms (WT D, Apo-D, Zn Mut D, and Cu Mut D). **b** Correlation matrix calculated from the Amide I data for the aggregated species of Apo, WT, and Zn mutant and Cu mutant. The matrix reveals high extent of secondary structural overlap

cluster analyses revealed a highly correlated colour matrix (Fig. 3b). This reflects that upon aggregation, the secondary structure organization of the protein species shares a significant overlap, in turn yielding a comparable amide I spectra.

The cluster dendrograms were subsequently determined using hierarchical clustering models constructed based on Euclidean distance. Euclidean distance method depends on an absolute distance between two points on a Euclidean plane. The two main branches represent conserved regions (comparable secondary structural features), which are further sub-branched (Fig. 4a). Native and unfolded apo-SOD1 are clustered together, suggesting that they have similar FTIR profiles. Unfolded H121F, which is in the vicinity of apo-SOD1 cluster, stands diverse. On the other hand, unfolded WT SOD1, native H121F, native, and unfolded H72F are similar in terms of denaturant-induced conformation. Similar dendrogram was also constructed to understand the hierarchic association of aggregated protein species with the other eight non-aggregated species (Fig. 4b). Aggregated apo-SOD1, Zn-deficient mutant (H72F aggregated), and Cu-deficient mutant (H121F aggregated) were closely branched. Although aggregated WT SOD1 was distanced from the other aggregated species, it got placed in different sub-branches relative to the WT SOD1 in the cluster dendrogram.

A conformational landscape of SOD1 was constructed using the relatedness analysis (Fig. 4c). The interim distance between two discrete clusters is the statistical manifestation of their structural diversity. Since the protein types used in the cluster analysis represent different states of SOD1 ranging from its unfolded apo-state to the complete folded holo-state, we used the distance outputs to quantitatively capture the folding sub-stages (Fig. 4c). Relevant protein types in the clusters were interpreted as the nested knots in the folding landscape, where the smoothness was introduced using spline interpolation [28]. Consisting of polynomial pieces on subintervals joined together with precise continuity conditions, a spline function mathematically bridges each interval between data points [29, 30]. The topology produces nested zones, which clearly reflect the structural transitions. Apo-SOD1 resides at the top of the funnel with H72F (Zn-deficient mutant) occupying a minimum, indicating that it can be considered as a stable intermediate in the pathway leading to the formation of the WT SOD1. The native state, which is also the most stable state, is attained after Cu incorporation. The Zn-deficient mutant is expected to represent a semi-stable mono-metallated state, which upon Cu uptake makes the metalloenzyme. The entry of the metal ions confers an orchestrated reorientation of the secondary structure, at which the loop regions are expected to play an important role.

The network N_w (Fig. 4d) was generated on weighted adjacency matrix of similarity scores. The node corresponding to the native state of WT protein was found separated from the main network. The number of neighbours of a node determined the degree centrality [31]. The presented cluster analyses helped revealing the correlations between the aggregated states and the folding sub-states in a statistically validated manner. Irrespective of the seed-protein species, the aggregates (AGG) showed a high degree of correlation among themselves. This observation is important, suggesting that all these aggregated species shared a high extent of secondary structure similarity, although their origin could be different. The observation also pointed out to a converging mechanism of aggregation, as has also been suggested by the previous reports [32]. Furthermore, there are reports, suggesting that a transition in conformation, which triggers that oligomerization is a feature shared among the

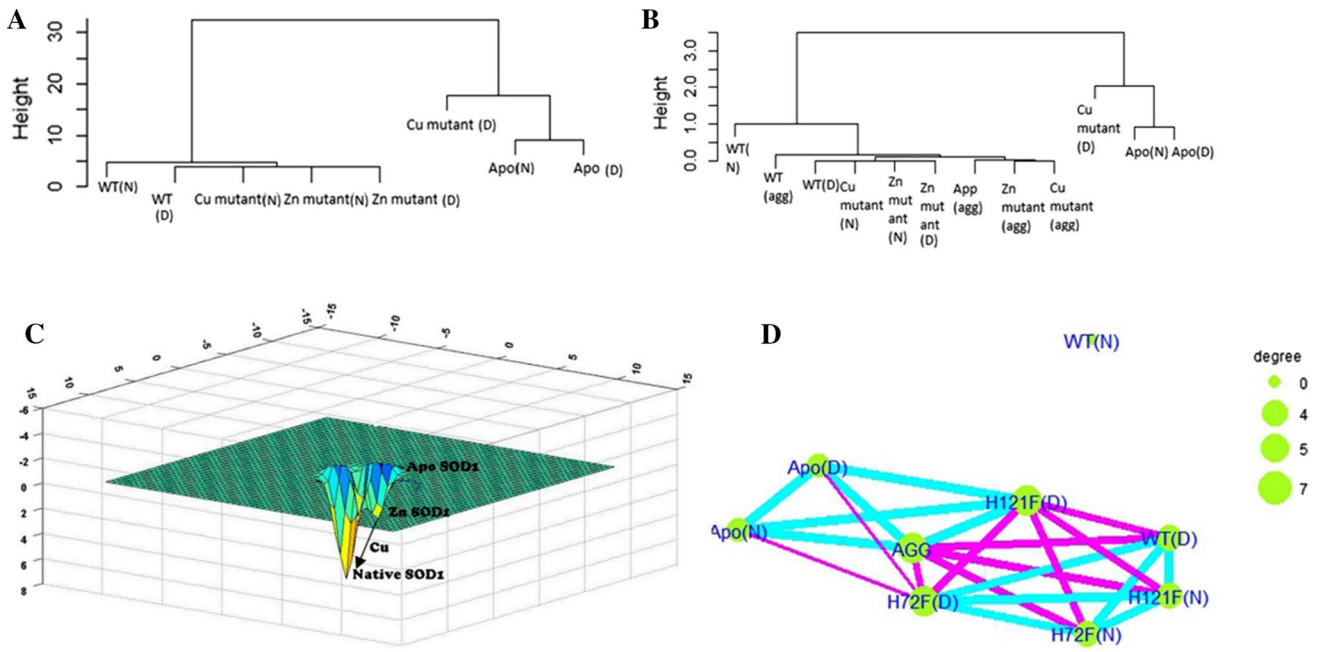


Fig. 4 **a** Cluster dendrograms constructed using Euclidean distance method showing the hierarchic distribution of the native and denatured forms of WT and metal mutants of SOD1. **b** Dendrogram showing the hierarchic distribution of native [WT(N), Cu mutant (N), Zn mutant (N), and Apo (N)], denatured [WT(D), Cu mutant (D), Zn mutant (D), and Apo (D)], and aggregated [WT(agg), Cu mutant (agg), Zn mutant (agg), and Apo (agg)] protein species and their interim distances of separation. **c** Folding landscape of SOD1 as determined from the distance calculation showing the presence of Zn SOD1 as the folding intermediate. **d** Network map, $N_{w,}$ showing the extent of relatedness of WT SOD1 with the other protein species (native, denatured, and aggregate). Cyan lines indicate stronger interconnection relative to purple lines. Bigger spheres indicate higher degree centrality and hence higher nodal connections. Agg node in N_w represents the aggregates

ALS-associated SOD1 mutants [33, 34]. The network map N_w further quantified the interconnections between different conformers and AGG. AGG node on the network map has seven nodal connections (Fig. 4d). Among these nodal junctions, the native and unfolded/denatured forms of SOD1 (Apo-N and Apo-D) have the highest edge width. This is in compliance with the previous reports, which showed that apo-SOD1 has significantly higher propensity to aggregate [35]. H121F (the Cu-deficient mutant) also have a strong nodal connection with the AGG node on the map, further implicating its tendency to aggregate. This observation is in line with the fact that the folding intermediate (Zn SOD1) being mapped in our folding landscape (Fig. 4c). Furthermore, the network, N_w, also revealed a strong connection between WT SOD1 and H121F (*N*). This stands in compliance with the observations from the folding landscape too. The co-ordination site of Zn is distributed only in loop IV, while that of Cu involves both loops IV and VII. Therefore, the occupancy of Cu site imposes greater internal steric restriction relative to the Zn site. However, SOD1 has two extended regions of intrinsic disorder (IDRs, intrinsically disordered regions) spanning from residues 69 to 100 and 121 to 143, and the propensity for intrinsic disorder in these regions is affected by the H72F and H121F mutations (see Fig. 5). Interestingly, three of the four Zn co-ordination sites fall within the first IDR stretch. Therefore, occupancy of the Zn-binding site leads to the transition from partial disorder to partial order in the secondary organization of SOD1. This explains as to why H121F (N) is closely placed to WT in the network map, as well as in cluster dendrogram (Fig. 4a, b). This is further validated by the secondary structure comparisons drawn from the β-structure and non-β-structure quantifications derived from the deconvolution of the Amide I IR data [Fig S1(i): A, B, and C] [36]. Therefore, the two micro-environments within the Zn pocket and Cu pocket have rather different ways of imposing steric restriction viz. Cu leading to a reduced mobility of loops IV and VII and Zn disrupting the disorder of loop IV.

The relevance of the apo-form in the aggregation of SOD1 has been substantiated not only by the cluster analyses presented here, but also by others. An inspection of the three-dimensional structure of SOD1 seems to implicate a possible role of the loop regions of SOD1 (loops IV and VII). To obtain further insights, we resorted to Ab initio calculations using Zhang Lab server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [37], which suggested the presence of helices in both loops IV and VII (Fig S6). In addition, a sequence analysis showed the presence of

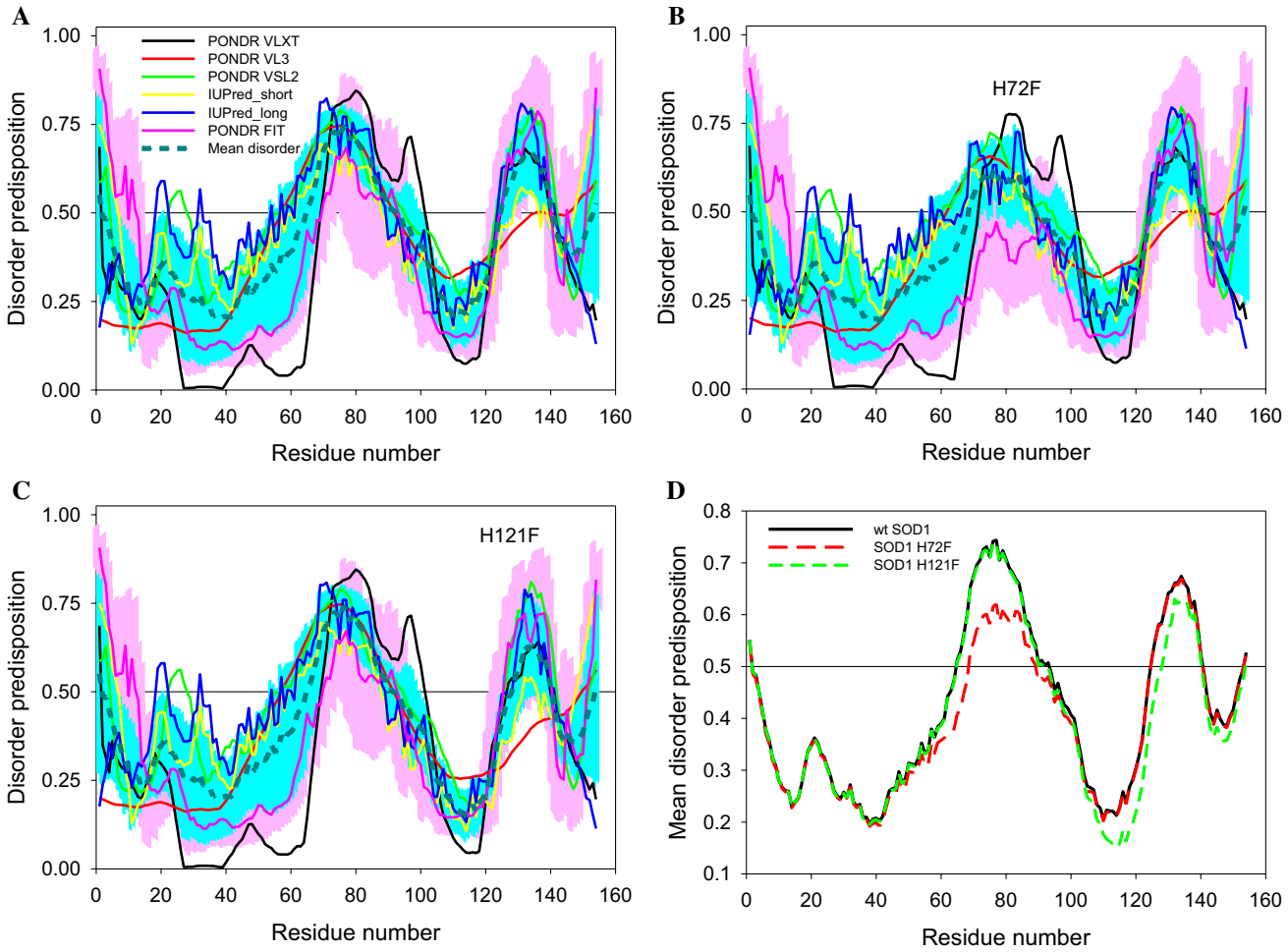


Fig. 5 Multiparametric analysis of the intrinsic disorder predisposition of WT SOD1 (**a**) and its H72F (**b**) and H121F (**C**) mutants evaluated by PONDR® VLXT (black lines), PONDR® VL3 (red lines), PONDR® VSL2 (green lines), PONDR® FIT (pink lines), IUPred_short (yellow lines), and IUPred_long (blue lines). Light pink shadow around PONDR® FIT curves shows error distribution. Bold dashed dark cyan line shows the mean disorder propensity calculated by averaging disorder profiles of individual predictors, whereas light cyan shadow around the corresponding curves represent error distribution. Plot D compares mean disorder profiles calculated for WT SOD1 (solid black curve) and its H72F (dashed red line) and H121F (dashed green line) mutants. In these analyses, the predicted intrinsic disorder scores above 0.5 are considered to correspond to the disordered residues/regions

seven chameleon sequences (Table S3) complementing the finding of ab initio calculations. Chameleon sequences can attain both α-helical and β-sheet organization [38], depending on their neighbourhood sequences and/or solution conditions. The positions of the chameleon sequences were found to have considerable overlap with the loop IV and VII regions, which can adopt transient helical structure (Fig. S7). It can hence be inferred that the absence of metals can prompt secondary structure flips which in turn gets manifested in terms of distance of separation in our cluster dendrograms and network map. It should be noted that the presence of helices in early folding intermediates was found before in different proteins, including β-lactalbumin, the intestinal fatty acid binding protein, and MPT-63 [39–41].

In the absence of metal ion, either in the mono-metallated state or in the apo-state, the extended loops owing to their intrinsic disorder nature support a continuum of conformational states and transitions [42, 43]. Unlike other intrinsically disordered proteins, which undergo disorder-to-order transitions upon ligand binding [44], thereby forming the fuzzy complexes, in proteins such as SOD1 and metallothionein such transitions are dependent on the orchestrated entry of metal ions [45]. There is an entropic cost associated with the disorder-to-order transition that accompanies the binding of an intrinsically unstructured protein to its target. The key thermodynamic driving force for the binding reaction is generally a favourable enthalpy contribution, which gives an example of enthalpy–entropy compensation [46]. These are all internal events, which remain synchronised with the

metal co-ordination in SOD1. This renders a cryptic disorder in proteins such as SOD1, where the metal ion co-factors upon entry conceals the local disorder and locks the loop region in its state of restricted mobility.

The advantage of the present method is that it is model free. This study does not require any detailed analyses of the FTIR spectra. In contrast, the limitation of the present study presumably originates from its sole use of secondary structure as the measurement option ignoring any contribution of the tertiary structure. This is obviously necessitated by the choice of the spectroscopic method. We preferred FTIR over CD, as the former method has higher sensitivity to β-structure and is presumably better suited for deciphering heterogeneities at the secondary levels of organization [36, 47, 48]. Current research effort in our laboratory is devoted to developing applications of cluster analysis to unravel the conformational interplay of several other proteins.

Acknowledgements SC acknowledges University Grants Commission (UGC) for Senior Research Fellowship. SS acknowledges Department of Science and Technology (DST) for doctoral fellowship. AB acknowledges Department of Biotechnology (DBT), Govt. of India for postdoctoral fellowship. KC acknowledges DST SERB EMR/2016/000310 for extramural grant. Authors thank Dr. Dipak Dasgupta of Saha Institute of Nuclear Physics for generously gifting the WT SOD1 construct. The authors acknowledge FTIR facility of CSIR – Indian Institute of Chemical Biology (IICB) and director, IICB.

References

1. Müller DJ, Gaub HE (2017) Membrane proteins scrambling through a folding landscape. Science 355:907–908
2. Nguyen H, Maier J, Huang H, Perrone V, Simmerling C (2014) Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. J Am Chem Soc 136:13959–13962
3. Gershenson A, Gierasch LM, Pastore A, Radford SE (2014) Energy landscapes of functional proteins are inherently risky. Nat Chem Biol 10:884–891
4. Shaw BF, Valentine JS (2007) How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? Trends Biochem Sci 32:78–85
5. Roberts BR, Tainer JA, Getzoff ED, Malencik DA, Anderson SR, Bomben VC, Meyers KR, Karplus PA, Beckman JS (2007) Structural characterization of zinc-deficient human superoxide dismutase and implications for ALS. J Mol Biol 373:877–890
6. Banci L, Barbieri L, Bertini I, Luchinat E, Secci E, Zhao Y, Aricescu AR (2013) Atomic-resolution monitoring of protein maturation in live human cells by NMR. Nat Chem Biol 9:297
7. Banci L, Bertini I, Cantini F, Kozyreva T, Massagni C, Palumaa P, Rubino JT, Zovo K (2012) Human superoxide dismutase 1 (hSOD1) maturation through interaction with human copper chaperone for SOD1 (hCCS). Proc Natl Acad Sci 109:13555–13560
8. Luchinat E, Cantini F, Rubino JT, Barbieri L, Banci L, Kozyreva T (2014) In-cell NMR reveals potential precursor of toxic species from SOD1 fALS mutants. Nat Commun 5:5502
9. Kim B-E, Nevitt T, Thiele DJ (2008) Mechanisms for copper acquisition, distribution and regulation. Nat Chem Biol 4:176–185
10. Bush AI (2002) Is ALS caused by an altered oxidative activity of mutant superoxide dismutase? Nat Neurosci 5:919
11. Potter SZ, Zhu H, Shaw BF, Rodriguez JA, Doucette PA, Sohn SH, Durazo A, Faull KF, Gralla EB, Nersissian AM (2007) Binding of a single zinc ion to one subunit of copper–zinc superoxide dismutase apoprotein substantially influences the structure and stability of the entire homodimeric protein. J Am Chem Soc 129:4575–4583
12. Banci L, Bertini I, Durazo A, Girotto S, Gralla EB, Martinelli M, Valentine JS, Vieru M, Whitelegge JP (2007) Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: a possible general mechanism for familial ALS. Proc Natl Acad Sci 104:11263–11267
13. Banci L, Bertini I, Boca M, Girotto S, Martinelli M, Valentine JS, Vieru M (2008) SOD1 and amyotrophic lateral sclerosis: mutations and oligomerization. PLoS One 3:e1677
14. Khan MAI, Respondek M, Kjellström S, Deep S, Linse S, Akke M (2017) Cu/Zn superoxide dismutase forms amyloid fibrils under near-physiological quiescent conditions: the roles of disulfide bonds and effects of denaturant. ACS Chem Neurosci 8:2019–2026
15. Proctor EA, Fee L, Tao Y, Redler RL, Fay JM, Zhang Y, Lv Z, Mercer IP, Deshmukh M, Lyubchenko YL (2016) Nonnative SOD1 trimer is toxic to motor neurons in a model of amyotrophic lateral sclerosis. Proc Natl Acad Sci 113:614–619
16. Redler RL, Wilcox KC, Proctor EA, Fee L, Caplow M, Dokholyan NV (2011) Glutathionylation at Cys-111 induces dissociation of wild type and FALS mutant SOD1 dimers. Biochemistry 50:7057–7066
17. Bhatia NK, Srivastava A, Katyal N, Jain N, Khan MAI, Kundu B, Deep S (2015) Curcumin binds to the pre-fibrillar aggregates of Cu/Zn superoxide dismutase (SOD1) and alters its amyloidogenic pathway resulting in reduced cytotoxicity. Biochim et Biophys Acta (BBA) Proteins Proteom 1854:426–436
18. Kong J, Yu S (2007) Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim Biophys Sin 39:549–559
19. Combs J, Gonzalez C, Wang C (2016) Surface FTIR techniques to analyze the conformation of proteins/peptides in H 2 O environment. J Phys Chem Biophys 6(2161–0398):1000202
20. Ridgley DM, Ebanks KC, Barone JR (2011) Peptide mixtures can self-assemble into large amyloid fibers of varying size and morphology. Biomacromol 12:3770–3779
21. Dong A, Huang P, Caughey WS (1990) Protein secondary structures in water from second-derivative amide I infrared spectra. Biochemistry 29:3303–3308
22. Susi H, Byler DM (1983) Protein structure by Fourier transform infrared spectroscopy: second derivative spectra. Biochem Biophys Res Commun 115:391–397
23. Wood TC, Pearson WR (1999) Evolution of protein sequences and structures. J Mol Biol 291:977–995
24. Gront D, Kolinski A (2005) HCPM—program for hierarchical clustering of protein models. Bioinformatics 21:3179–3180
25. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
26. Chang D-J, Desoky AH, Ouyang M, Rouchka EC (2009) Compute pairwise manhattan distance and pearson correlation coefficient of data points with gpu. In: Paper presented at the software engineering, artificial intelligences, networking and parallel/distributed computing, 2009 SNPD'09 10th ACIS international conference on
27. Johnson SC (1967) Hierarchical clustering schemes. Psychometrika 32:241–254
28. Maeland E (1988) On the comparison of interpolation methods. IEEE Trans Med Imaging 7:213–217

29. De Boor C, De Boor C, Mathématicien E-U, De Boor C, De Boor C (1978) A practical guide to splines. Springer-Verlag, New York
30. Scott LR, Zhang S (1990) Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math Comput 54:483–493
31. Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc Netw 32:245–251
32. Uversky VN, Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim et Biophys Acta (BBA) Proteins Proteom 1698:131–153
33. Nordlund A, Leinartaitė L, Saraboji K, Aisenbrey C, Gröbner G, Zetterström P, Danielsson J, Logan DT, Oliveberg M (2009) Functional features cause misfolding of the ALS-provoking enzyme SOD1. Proc Natl Acad Sci 106:9667–9672
34. Teilum K, Smith MH, Schulz E, Christensen LC, Solomentsev G, Oliveberg M, Akke M (2009) Transient structural distortion of metal-free Cu/Zn superoxide dismutase triggers aberrant oligomerization. Proc Natl Acad Sci 106:18273–18278
35. Furukawa Y, O'Halloran TV (2005) Amyotrophic lateral sclerosis mutations have the greatest destabilizing effect on the apo-and reduced form of SOD1, leading to unfolding and oxidative aggregation. J Biol Chem 280:17266–17274
36. Yang H, Yang S, Kong J, Dong A, Yu S (2015) Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. Nat Protoc 10:382
37. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738
38. Guo JT, Jaromczyk JW, Xu Y (2007) Analysis of chameleon sequences and their implications in biological processes. Proteins Struct Funct Bioinform 67:548–558
39. Sakurai K, Fujioka S, Konuma T, Yagi M, Goto Y (2011) A circumventing role for the non-native intermediate in the folding of β-lactoglobulin. Biochemistry 50:6498–6507
40. Kundu A, Kundu S, Chattopadhyay K (2016) The presence of non-native helical structure in the unfolding of a beta sheet protein MPT63. Protein Sci 26:536–559
41. Sarkar-Banerjee S, Chowdhury S, Paul SS, Dutta D, Ghosh A, Chattopadhyay K (2016) The non-native helical intermediate state may accumulate at low pH in the folding and aggregation landscape of the intestinal fatty acid binding protein. Biochemistry 55:4457–4468
42. Babu MM, Kriwacki RW, Pappu RV (2012) Versatility from protein disorder. Science 337:1460–1461
43. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends Biochem Sci 33:2–8
44. Uversky VN (2011) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40:1623–1634
45. Faller P, Hureau C, La Penna G (2014) Metal ions and intrinsically disordered proteins and peptides: from Cu/Zn amyloid-β to general principles. Acc Chem Res 47:2252–2259
46. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331
47. Dong A, Matsuura J, Allison SD, Chrisman E, Manning MC, Carpenter JF (1996) Infrared and circular dichroism spectroscopic characterization of structural differences between β-lactoglobulin A and B. Biochemistry 35:1450–1457
48. Barth A (2007) Infrared spectroscopy of proteins. Biochim et Biophys Acta (BBA)-Bioenerg 1767:1073–1101