bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Molecular dynamics analysis of Superoxide Dismutase 1 mutations suggests decoupling between mechanisms underlying ALS onset and progression

Munishikha Kalia[1,2], Mattia Miotto[3], Deborah Ness[1,2], Sarah Opie-Martin[1], Thomas P Spargo[1,2], Lorenzo Di Rienzo[3], Tommaso Biagini[4], Francesco Petrizzelli[4], Ahmad Al-Khleifat[2], Renata Kabiljo[1,2], Simon Topp[2,5], Keith Mayl[2], Isabella Fogh[2], Puja R Mehta[2,6], Kelly L Williams[7] , Jennifer Jockel-Balsarotti[8] , Taha Bali[8,9] , Wade Self[8] , Lyndal Henden[7] , Garth A Nicholson[7,10], Nicola Ticozzi[11,12], Diane McKenna-Yasek[13], Lu Tang[14], Pamela Shaw[15], Adriano Chio[16,17], Albert Ludolph[18,19], Jochen H Weishaupt[20,21], John E Landers[13], Jonathan D Glass[22], Jesus S Mora[23], Wim Robberecht[24], Philip Van Damme[24,25], Russell McLaughlin[26], Orla Hardiman[27], Leonard H van den Berg[28], Jan H Veldink[28], Phillippe Corcia[29,30], Zorica Stevic[31], Nailah Siddique[32], Antonia Ratti[11,33], Vincenzo Silani[11,12], Ian P Blair[7], Dong-sheng Fan[14], Florence Esselin[34], Elisa de la Cruz[34], William Camu[34], A Nazli Basak[35], Teepu Siddique[32], Timothy Miller[9], Robert H Brown[13], Peter M Andersen[36], Project MinE ALS Sequencing Consortium, Christopher E Shaw[5,37] Tommaso Mazza[4], Giancarlo Ruocco[38,39], Edoardo Milanetti[38,39], Richard JB Dobson [1,40,41], Ammar Al-Chalabi[2,42] and Alfredo Iacoangeli[1,2,41*]

[1]Department of Biostatistics and Health Informatics, King's College London, London, UK
[2]Department of Basic and Clinical Neuroscience, King's College London, Maurice Wohl Clinical Neuroscience Institute, London, UK
[3]Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy
[4]Bioinformatics Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, S. Giovanni Rotondo, Italy
[5]UK Dementia Research Institute Centre at King's College London, School of Neuroscience, King's College London, Strand, London, 57 WC2R 2LS, United Kingdom[6]UCL Queen Square Motor Neuron Disease Centre, Department of Neuromuscular diseases, UCL Queen Square Institute of Neurology, London, WC1N 3BG, UK
[7]Macquarie University Centre for MND Research, Macquarie Medical School, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, NSW, Australia
[8]Department of Neurology, Washington University School of Medicine, St Louis, Missouri, MO 63110, USA
[9]Department of Neurology, Tufts University School of Medicine, Boston, MA 02111, USA
[10]Concord Clinical School, ANZAC Research Institute, Concord Repatriation Hospital, Sydney, NSW 2139, Australia
[11]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy
[12]Department of Pathophysiology and Transplantation, "Dino Ferrari" Center, Università degli Studi di Milano, Milan, Italy
[13]Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts, MA 02125, USA
[14]Department of Neurology, Peking University Third Hospital, 49 North Garden Road, Haidian District, Beijing, 100191, People's 32 Republic of China
[15]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, United Kingdom, S10 2HQ
[16]Rita Levi Montalcini' Department of Neuroscience, University of Turin, Italy
[17]Neurology 1, AOU Città della Salute e della Scienza of Torino, Turin, Italy
[18]Department of Neurology, Ulm University, Oberer Eselsberg 45, 89081 Ulm, Germany
[19]German Center for Neurodegenerative Diseases, DZNE, Ulm, Germany
[20]Department of Neurology, University of Ulm, Oberer Eselsberg 45, 89081 Ulm, Germany
[21]Division of Neurodegenerative Disorders, Department of Neurology, Mannheim Center for Translational Neuroscience, Medical 40 Faculty Mannheim, Heidelberg University, Germany
[22]Department Neurology, Emory University School of Medicine, Atlanta, GA 30322, US
[23]ALS Unit, Department of Neurology, Hospital San Rafael, 28016, Madrid, Spain
[24]Neurology Department, University Hospitals Leuven, Herestraat 49, 3000, Belgium
[25]Neuroscience Department KU Leuven and Center for Brain & Disease Research VIB Leuven, Belgium
[26]Complex Trait Genomics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin D02 PN40, Ireland

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[27]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin D02 PN40, Ireland
[28]Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, 3584 CX, The 48 Netherlands
[29]Centre de Référence pour la SLA et les autres maladies du motoneurone (FILSLAN), 2 Avenue Martin Luther King, 87042 Limoges 50 Cedex
[30]Centre de Compétences Neuropathies amyloïdes familiales et autres neuropathies périphériques rares (NNERF)
[31]Neurology Clinic, Clinical Center of Serbia, School of Medicine, University of Belgrade, Studentski trg 1, Belgrade, Serbia
[32]Neuromuscular Disorders Program, Northwestern University, Feinberg School of Medicine, IL 60208
[33]Department of Medical Biotechnology and Molecular Medicine, Università degli Studi di Milano, 20122 Milan, Italy
[34]Reference Center for ALS and Other Rare Motoneuron Disorders, University Hospital Gui de Chauliac, Montpellier, France, FRA 34295
[35]Koç University, School of Medicine Translational Medicine Research Center KUTTAM-NDAL, 34450 Sarıyer
[36]Institute of Clinical Sciences, Neuroscience, Umeå University, Umeå, Sweden
[37]Centre for Brain Research, University of Auckland, 85 Park Road, Grafton, Auckland 1023, New Zealand
[38]Center for Life Nano & Neuro Science, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161, Rome, Italy
[39]Department of Physics, Sapienza University, Piazzale Aldo Moro 5, 00185, Rome, Italy
[40]Institute of Health Informatics, University College London, London, UK
[41]National Institute for Health Research Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust King's College London, London, United Kingdom
[42]Clinical Neurosciences, King's College Hospital, Denmark Hill, London, UK

*Correspondence should be addressed to alfredo.iacoangeli@kcl.ac.uk

ABSTRACT

Mutations in the superoxide dismutase 1 (*SOD1*) gene are the second most common known cause of ALS. *SOD1* variants express high phenotypic variability and over 200 have been reported in people with ALS. Investigating how different *SOD1* variants affect the protein dynamics might help in understanding their pathogenic mechanism and explaining their heterogeneous clinical presentation. It was previously proposed that variants can be broadly classified in two groups, 'wild-type like' (WTL) and 'metal binding region' (MBR) variants, based on their structural location and biophysical properties. MBR variants are associated with a loss of SOD1 enzymatic activity. In this study we used molecular dynamics and large clinical datasets to characterise the differences in the structural and dynamic behaviour of WTL and MBR variants with respect to the wild-type SOD1, and how such differences influence the ALS clinical phenotype. Our study identified marked structural differences, some of which are observed in both variant groups, while others are group specific. Moreover, applying graph theory to a network representation of the proteins, we identified differences in the intramolecular contacts of the two classes of variants. Finally, collecting clinical data of approximately 500 *SOD1* ALS patients carrying variants from both classes, we showed that the survival time of patients carrying an MBR variant is generally longer (~6 years median difference, $p < 0.001$) with respect to patients with a WTL variant. In conclusion, our study highlights key differences in the dynamic behaviour of the WTL and MBR SOD1 variants, and wild-type SOD1 at an atomic and molecular level. We identified interesting structural features that could be further investigated to explain the associated phenotypic variability. Our results support the hypothesis of a decoupling between mechanisms of onset and progression of *SOD1* ALS, and an involvement of loss-of-function of SOD1 with the disease progression.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder, characterized by progressive muscle weakness and paralysis, leading to death from neuromuscular respiratory failure typically within 5 years of symptom onset [1, 2]. Mutations in the *superoxide dismutase type 1 (SOD1)* gene are the second most common known cause of ALS and have been found in both familial ALS and sporadic ALS [3]. SOD1 is an antioxidant metalloenzyme, which protects cells from oxidative damage. It catalyses the conversion of the superoxide O_2^- free radical present in the cytoplasm to H_2O_2 and molecular oxygen. SOD1 is a 32 kDa protein which exists as a homodimer and its structure is characterized by the presence of a β-barrel immunoglobin fold, two long loops: metal binding loop (MBL) and electrostatic loop (ESL) [4]. The fully formed SOD1 homodimer is highly stable and has a melting point of 92 °C [5], due to copper and zinc binding, an intrasubunit disulphide bond and dimerization [6].

At present, more than 220 *SOD1* mutations have been reported in people with ALS [7, 8]. Although there is evidence that *SOD1* mutations cause ALS, the exact underlying mechanism of disease onset and progression remains poorly understood [9]. It has been proposed that the pathogenic mutations hinder post-translational maturation, which decreases structural stability, thereby instigating the misfolding of SOD1 protein [10-12]. The ultimate outcome of SOD1 misfolding is the formation of hallmark amyloid aggregates of SOD1 in the affected tissues [13-15]. A noteworthy feature of the misfolded SOD1 is its prion-like behaviour of self-proliferation. It was proposed that a misfolded SOD1 can transverse between cells and cause misfolding of other SOD1 molecules via the protein aggregates released from dying cells [16-18].

Moreover, *SOD1* variants can express high phenotypic variability. For example, A4V (A5V in the HGVS V2.0 nomenclature – traditional nomenclature will be used throughout this manuscript) is a variant which is responsible for 48% of *SOD1* ALS in US familial ALS. It is associated with variable site of onset, rapid disease progression and a mean survival time of 1.1 years from clinical presentation [19-21]. In contrast, H46R is commonly observed in the Japanese, Pakistan and Norwegian populations and is invariably associated with a stereotypic phenotype with slowly ascending paresis beginning in the legs and long survival (approx. 12 years from diagnosis) [22-25]. Investigating the differences between *SOD1* mutations in terms of their structural effect on the protein might help in understanding the mechanism underlying such a heterogeneous clinical presentation.

Disease-associated *SOD1* mutations are found in all domains of the protein and over the years attempts have been made to classify the impact of each variant based on location and subsequent effect on protein structure [26]. Based on their structural location and biophysical characteristics such as metal binding affinity and influence on SOD1 *in vitro* activity, it has been proposed that *SOD1* variants can be classed in two broad groups [26-28]. The first group consists of the 'wild-type like' (WTL) variants such as: A4V, L38V, G37R, G41S, G72S, D76Y, D90A and G93A. These variants bind metal ions tightly and the catalytic activity is not affected *in vitro*, therefore their biophysical behaviour is expected to be similar to that of the wild-type SOD1 (wt-SOD1) [27, 29, 30]. The second group contains 'metal binding region' (MBR) variants, such as: H46R, H48Q, G85R, D124V, D125H, G127X and S134N. These variants are localized in the metal binding sites, metal binding loop and electrostatic loop regions, and cause reduced metal binding and diminished *in vitro* catalytic activity [4, 31-33].

The aim of this study was to characterize the differences in the structural and dynamic behaviour of the WTL and MBR disease-causing SOD1 variants with respect to the wt-SOD1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For this we performed all atomistic molecular dynamics (MD) simulations of the wt-SOD1, 6 MBR and 7 WTL SOD1 variant monomers. We focussed on the metal-free SOD1 monomer, as the conformational changes of the metal binding loop and electrostatic loop are unravelled by the lack of metal stabilisation [33]. In order to quantify and characterise the motion of the whole protein and its loops, we studied root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration, number of hydrogen bonds and performed principal component analysis (PCA). Moreover, two additional analyses were carried out that focus on the study of individual residues. In the first we studied the covariance of the residue motions and in the second, we used a graph-theory based approach to study the complexity of the intermolecular connections that were captured in a simple residue-specific descriptor. Our results suggested that the structural and dynamic behaviour differences between WTL and MBR variants could be linked to the ALS clinical phenotype. To explore this hypothesis, we collected a large clinical dataset of *SOD1* ALS patients [34] and performed survival and age of onset analyses in the patients carrying WTL and MBR variants.

RESULTS

To characterize the impact of one mutation at a time, we focussed only on SOD1 structures with one single amino acid variant. We could identify 13 WTL and MBR mutations in the Protein Data Bank (PDB) for which high quality structures were available. We divided them in the two categories based on whether they were previously reported to belong to one or whether they were in proximity of another mutation of known category. For example, we classified T2D as WTL as it is in proximity of A4V which was reported to exhibited normal SOD activity *in vitro* [27]. The WTL group included T2D, A4V, G37R, L38V, T54R, G93R, I113T and the MBR variants were H46R, C57S, H80R, G85R, D124V and D125H. The structural location of the variants is shown in the Figure 1. The detailed structural information, clinical interpretation, and resolution of the PDB structures used in the study are shown in Table 1.

We have used the monomeric apo SOD1 instead of the fully formed SOD1 homodimer, as it is the least stable form of the protein and an ideal model to study the impact of the pathogenic variants [32, 33]. We performed all atomistic MD simulations of the wt-SOD1, 6 MBR and 7WTL variants in TIP3P water for 100 ns each and each simulation was performed in triplicates.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

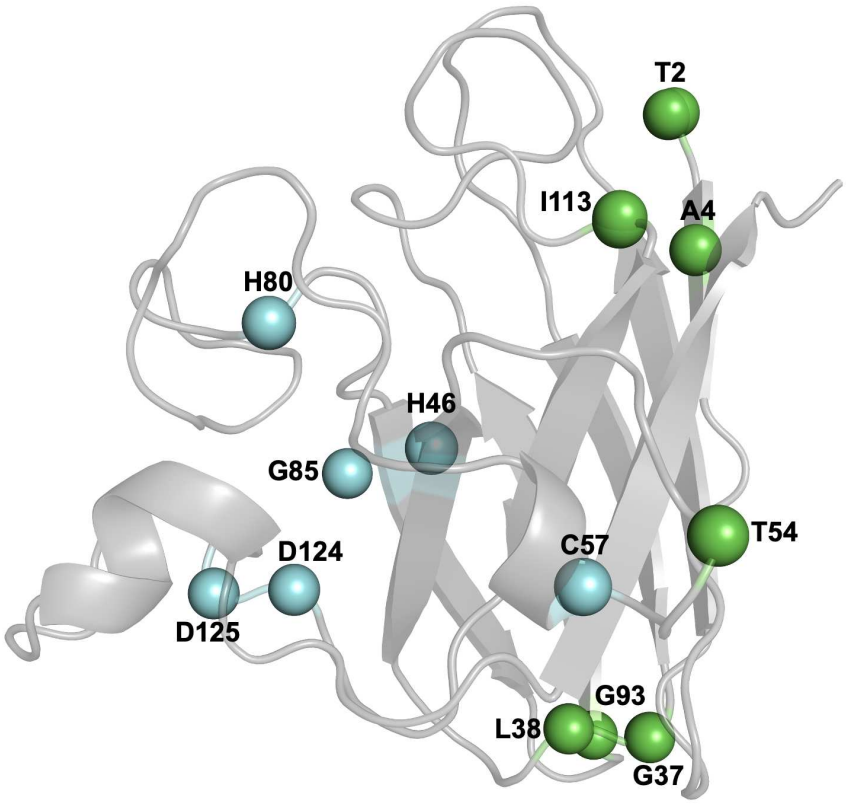

Figure 1. SOD1 monomer structure coloured in grey, the residues which form WTL variants are coloured in blue and the residues which form the MBR variants are coloured in green.

Flexibility of the metal binding and electrostatic loops.

The root mean square fluctuation is a measure of individual protein residue flexibility. A high RMSF indicates higher flexibility and a low RMSF indicates lower flexibility. We analysed the RMSF of Cα atoms of the wt-SOD1 and 13 SOD1 variants. Our analysis highlighted marked differences in the flexibility of the metal binding loop and electrostatic loop regions (Figure 2a, b), while the remaining part of the protein was relatively less flexible (Sup Fig 1). For the metal binding loop region (Figure 2a), the highest flexibility was observed in the H46R variant. Next, we performed the Wilcoxon Rank-Sum Test for statistical comparison between the RMSF of each variant and the wt-SOD1. We noted that the RMSF difference between a variant and the wt-SOD1 was significant for A4V, H46R and D125H. For the remaining variants the RMSF difference between the variants and the wt-SOD1 were non-significant.

Focussing on the electrostatic loop, the MBR variants such as C57S, D125H, H46R H80R (Figure 2b), generally showed higher RMSF with respect to WTL variants, illustrating their increased flexibility. In contrast, WTL variants, such as A4V, G37R, I113T and G93A, showed FMSF values lower or very similar to that of wt-SOD1 (Figure 2b). The only exception was L38V which had very high RMSF values in the electrostatic loop region. The statistical analysis showed that the RMSF difference between A4V, H46R and wt-SOD1 were significant, while for the remaining variants the differences were not significant. A4V causes one of the most aggressive forms of *SOD1* ALS, whilst H46R results in very slow progression. It is interesting to note that their RMSFs corresponded to the lowest and highest values among the investigated variants, suggesting that such dynamics features could be explored to explain the clinical representation of variants.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Finally, we tested the difference between the RMSF distributions of the entire protein for all the variants in each class, i.e. WTL and MBR, and the wt-SOD1 (Sup Fig 2). The RMSF of the WTL variants was not different from the wt-SOD1 (p = 0.54) while the RMSF of the MBR variants was higher than both the wt-SOD1 (p = 5.2x10^-07) and the WTL variants (p = 6.4x10^-08).

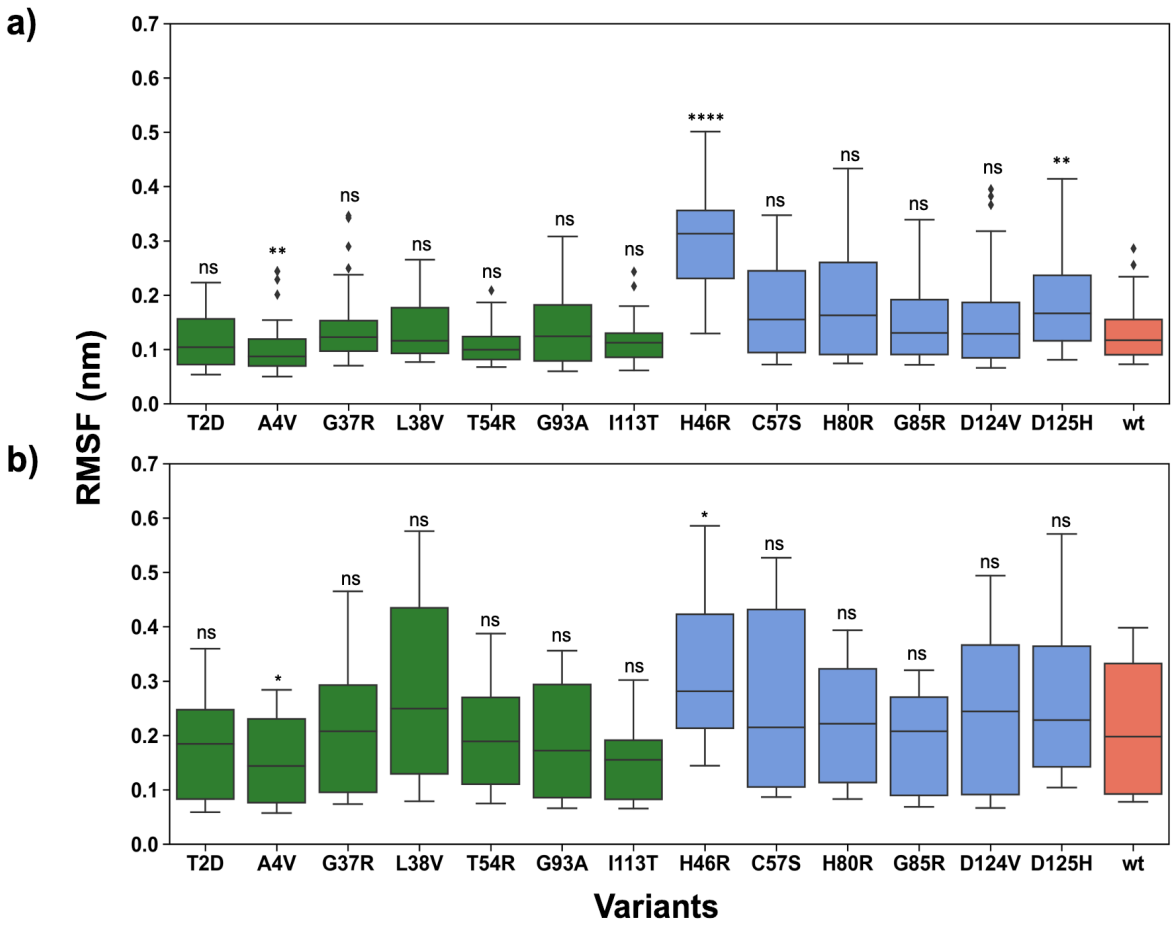

Figure 2. Boxplots depicting the RMSF distribution of the WTL (green), MBR (blue) variants and wt-SOD1 (coral) in the a) MBL b) ESL regions. RMSF is calculated for the Cα atoms. The whisker bars represent the range of minimum and maximum RMSD, the median is represented by a line subdividing the box. Wilcoxon rank-sum test, p-value annotation legend: ns: p <= 1.00e+00, *: 1.00e-02 < p <= 5.00e-02, **: 1.00e-03 < p <= 1.00e-02, ***: 1.00e-04 < p <= 1.00e-03, ****: p <= 1.00e-04.

Comparison of the conformational changes in the WTL and MBR variants

The root mean square deviation is a quantitative measure of the similarity between two superimposed sets of atomic coordinates and is widely used in the MD analysis to highlight the conformational changes observed in the protein structures over the length of simulations. We observed that there are clear differences in the Cα RMSD distribution of the WTL and MBR variants (Figure 3 and Table 2). The RMSD distributions of the WTL variants were similar to that of the wt-SOD1. The mean and median RMSDs observed in all WTL variants were <= than the mean and median RMSD of the wt-SOD1 (both 0.22 nm), while all mean and median RMSDs observed for all MBR variants but G85R, for which the mean was 0.20 nm and median 0.21nm, were higher than the wt-SOD1 and WTL variants. The same trend was observed when only the RMSDs between 20ns and 100ns were considered to account for the equilibration phase (Sup Fig 3).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Subsequently, we investigated the conformational changes adopted by the variants and the wt-SOD1 during the simulations. We observed that the wt-SOD1 adopted one stable conformation during the simulations (Figure 4). Among the WTL variants, L38V and T54R also assumed a single conformation. A4V, G93A and I113T acquired two major conformations, whilst T2D and G37R adopted three major conformations. The MBR variants D124V and D125H presented one major conformation for the entire length of the simulation, H46R and H80R adopted two major conformations and, interestingly, C57S and G85R variants showed four major conformations during the trajectory of the simulations (Figure 4).

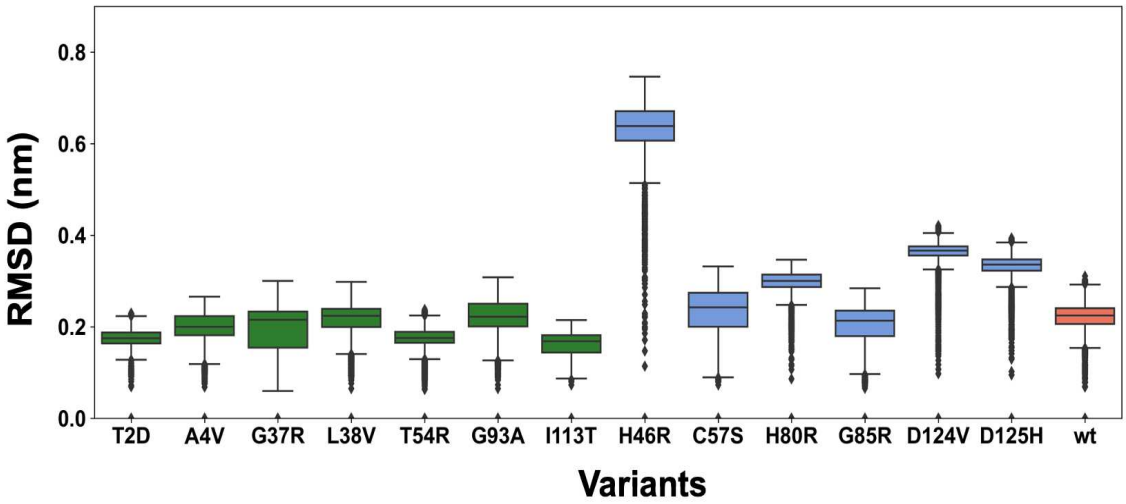

Figure 3. Box plots depicting RMSD analysis of the MD simulations performed on the wt-SOD1 (coral), WTL (green) and MBR (blue) variants. The RMSD is calculated for the Cα atoms of the wt-SOD1 and the variants. The whisker bars represent the range of minimum and maximum RMSD, the median is represented by a line subdividing the box.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

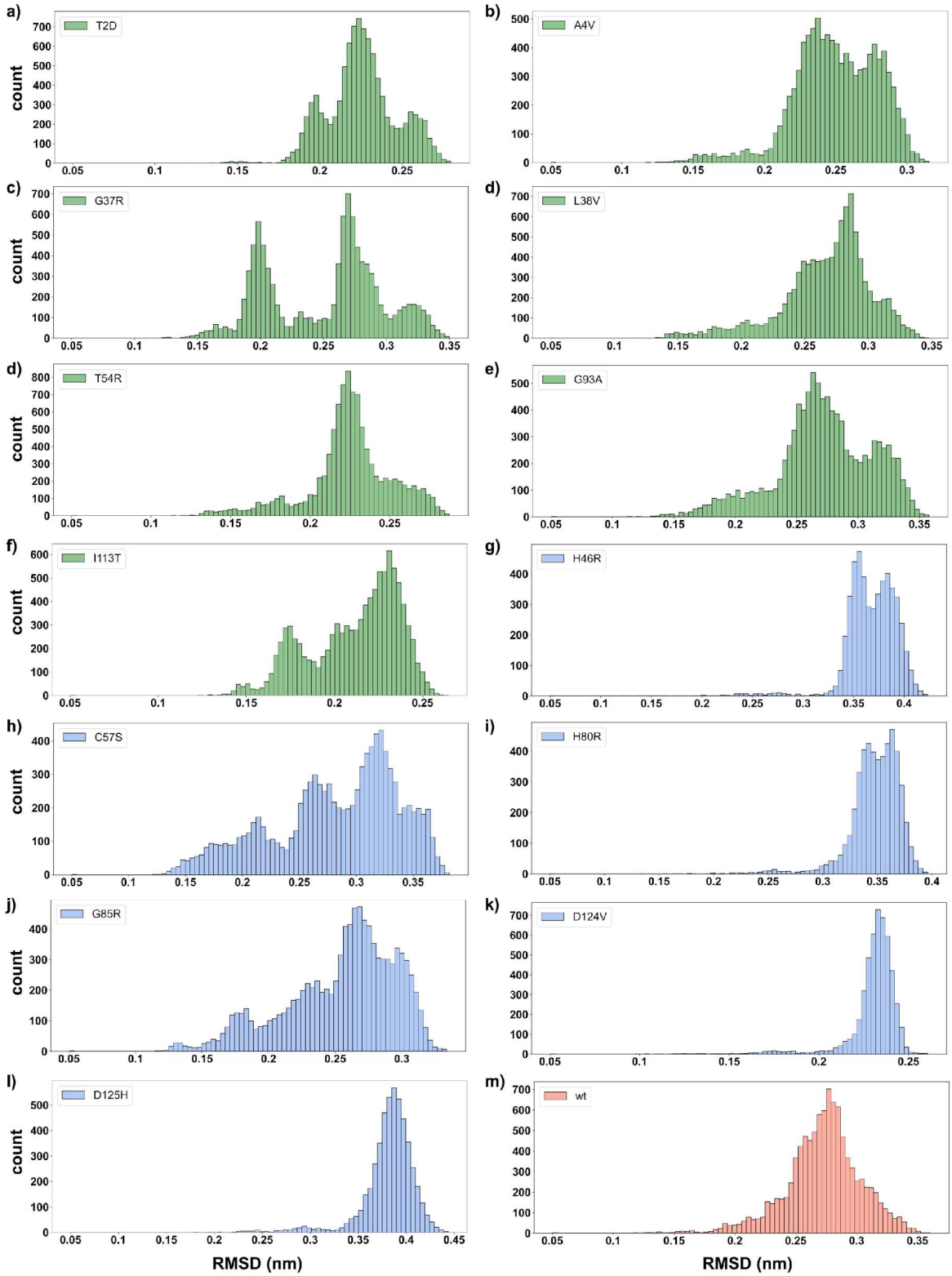

Figure 4. Histograms showing the conformational changes observed in the WTL (green), MBR (blue) variants and wt-SOD1 (coral). a) T2D b) A4V c) G37R d) L38V e) G93A f) I113T g) H46R h) C57S i) H80R j) G85R k) D124V l) D125H m) wt-SOD1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Comparison of the compactness of the WTL and MBR variants

The radius of gyration (Rg) is a measure of the compactness of a protein and is widely utilised to compare the protein's dynamic behaviour. A high radius of gyration is an indication that the protein structure is more extended, whilst a smaller radius of gyration is indicative of a more compact protein structure. We calculated the radius of gyration of the WTL and MBR variants, and wt-SOD1 over the entire length of the simulations. Our analysis showed that the protein structure remains closely packed in the wt-SOD1 and WTL variants (Figure 5), while MBR variants adopt a more extended conformation (Figure 5). The largest difference with respect to the wt-SOD1 was observed in the H46R and H80R variants.

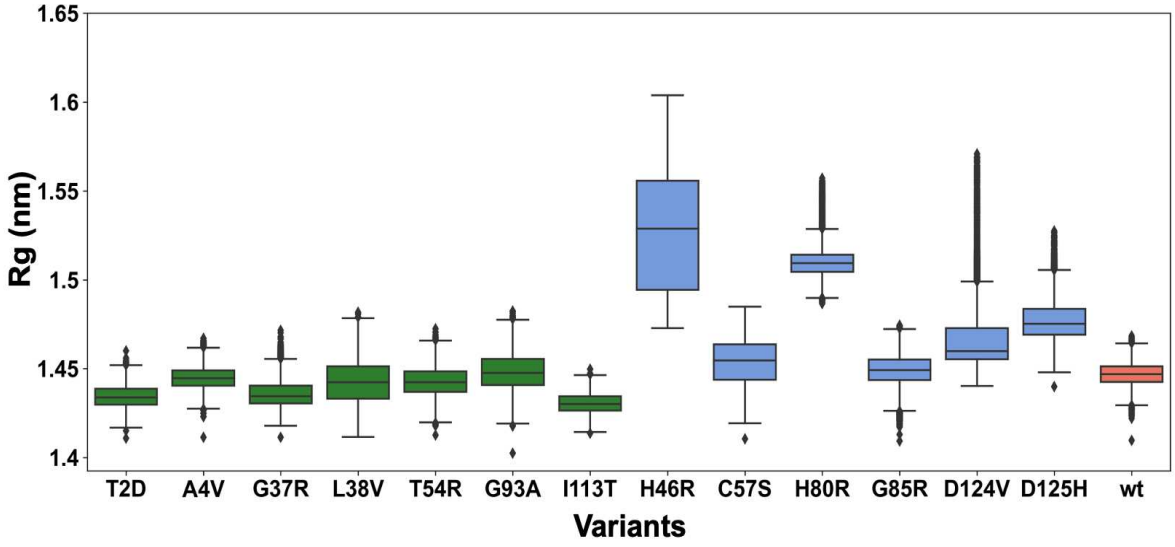

Figure 5. Box plots depicting radius of gyration (Rg) comparison of the wt-SOD1 (coral), WTL (green) and MBR (blue) variants simulations. The whisker bars represent the range of minimum and maximum Rg, the median is represented by a line subdividing the box.

Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique that is frequently used to reduce the number of dimensions that describe the dominant protein domain motion [35-37]. PCA was performed on the WTL, MBR variants and wt-SOD1. The first three eigenvectors explained > 40 % of all the domain motion (Table 3) and such motion was driven by the electrostatic loop for the wt-SOD1 and all variants but T2D and I113T (porcupine plots in Figure 6 a,b). In the MBR SOD1 variants, considerable domain motion was also observed in the metal binding loop in addition to the electrostatic loop. Moreover, the magnitude of the domain motion in the metal binding loop and electrostatic loop regions was substantially larger for the MBR variants than the WTL variants. It was interesting to note that the dynamic motion of metal binding loop and electrostatic loop regions in L38V was more similar to the MBR variants than the WTL variants. The motion variance explained by PC1-4 for each variant and the wt-SOD1 is shown in Table 3.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

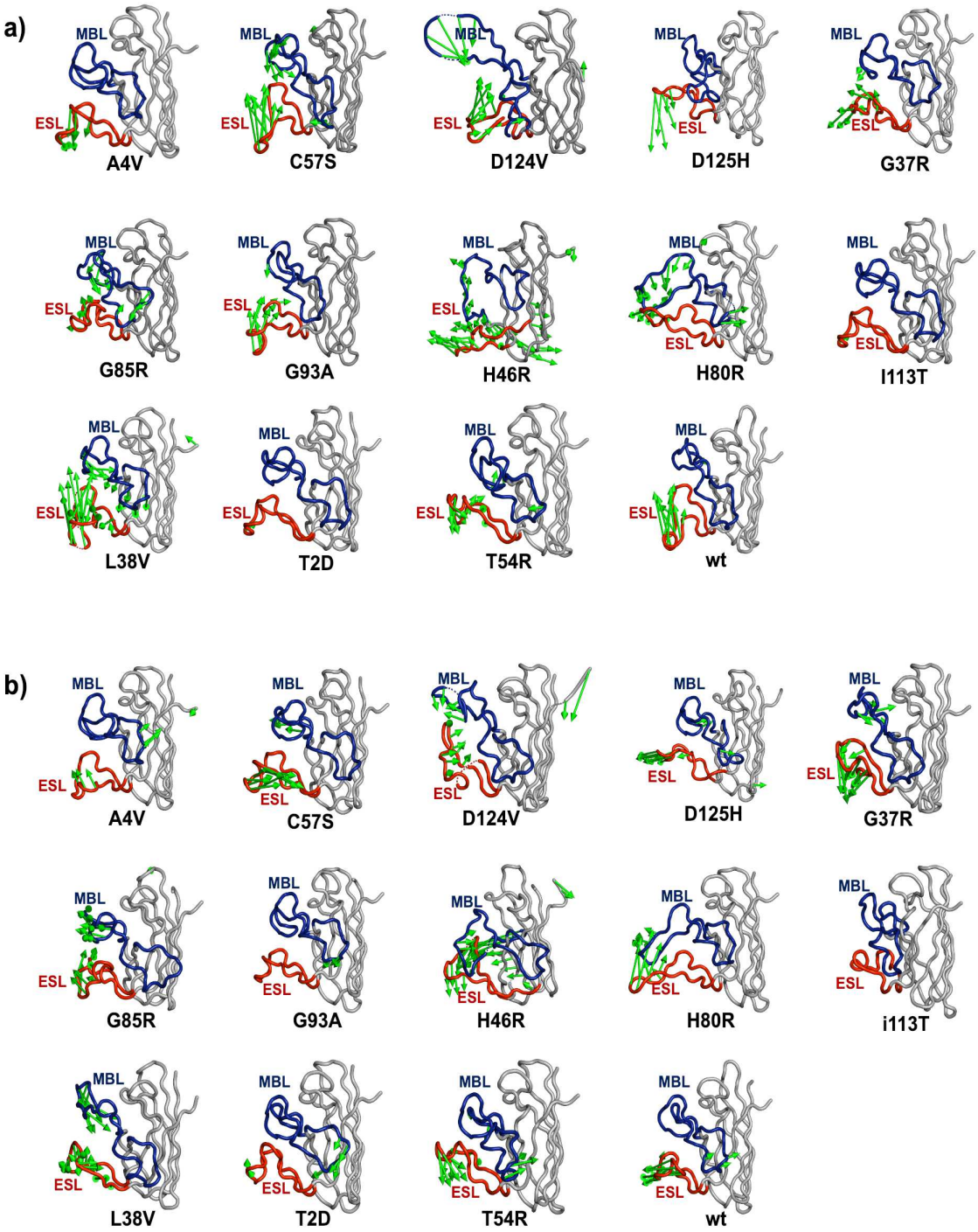

Figure 6. PCA analysis of the concatenated trajectories. Porcupine plots corresponding to the a) PC1 b) PC2. The wt-SOD1 and variant protein backbone is coloured in grey, the MBL in blue and ESL in red. The vectors coloured in green represent the direction and magnitude of the domain motion.

Hydrogen bond analysis

To further investigate the dynamics of the wt-SOD1 and its variants we analysed the total number of hydrogen bonds formed during the simulations. The average number of hydrogen bonds formed in the entire SOD1 protein in the wt-SOD1, WTL and MBR variants are shown in Figure 7 (full distributions in Sup Fig 4). Interestingly, A4V and H46R variants, which

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

commonly represented the two extreme opposites in the other analyses, had a similar number of hydrogen bonds as the wt-SOD1, while all other variants consistently presented a lower mean number with respect to the wt-SOD1.

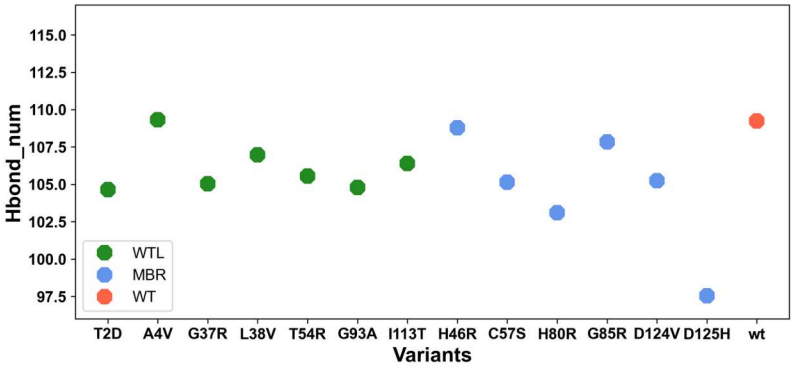

Figure 7. The mean number of hydrogen bonds formed in the wt-SOD1 (coral), WTL (green) and MBR (blue) variants during the entire length simulation.

Graph-Theory based analysis

To better investigate the complex nature of the intramolecular interactions of the simulated systems, we used graph-theory based descriptors to evaluate the role of each residue within their interaction network. To this end, we used a weighted network representation, where each residue of the protein was a node of the network. Two nodes were connected by a link if the distance between their side chain centroids was lower than a threshold. In addition, each link was weighted with the contact frequency as calculated from the MD simulations trajectory. Several descriptors were computed using this network representation of the protein.

Many typical network descriptors (such as degree, strength, clustering coefficient, etc.) can capture local features of the system. Conversely, *shortest path based* descriptors are able to capture the connection between residues, even when they are spatially distant. Indeed, such quantities do not consider direct interactions but a set of consecutive interactions, defined as a path. For this purpose, we selected the closeness centrality descriptor, defined as the reciprocal of the sum of the length of the shortest paths between a given node and all other nodes in the graph. The higher the closeness of a residue, the higher its centrality is in the protein network.

First, we compared the distributions of residue closeness centrality values analysing wt-SOD1 and its 13 variant systems. We studied separately the residues belonging to metal binding loop and electrostatic loop. In Figure 8a, we reported the boxplots for each system, where the boxplots represent the values obtained considering all three simulation replicas.

Overall, the residue closeness centrality values of MBR variants were >= than the wt-SOD1 while WTL variants were <= than the wt-SOD1. H46R and D125H showed a large difference with respect to the wt system considering both metal binding loop and electrostatic loop regions (Figure 8a).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

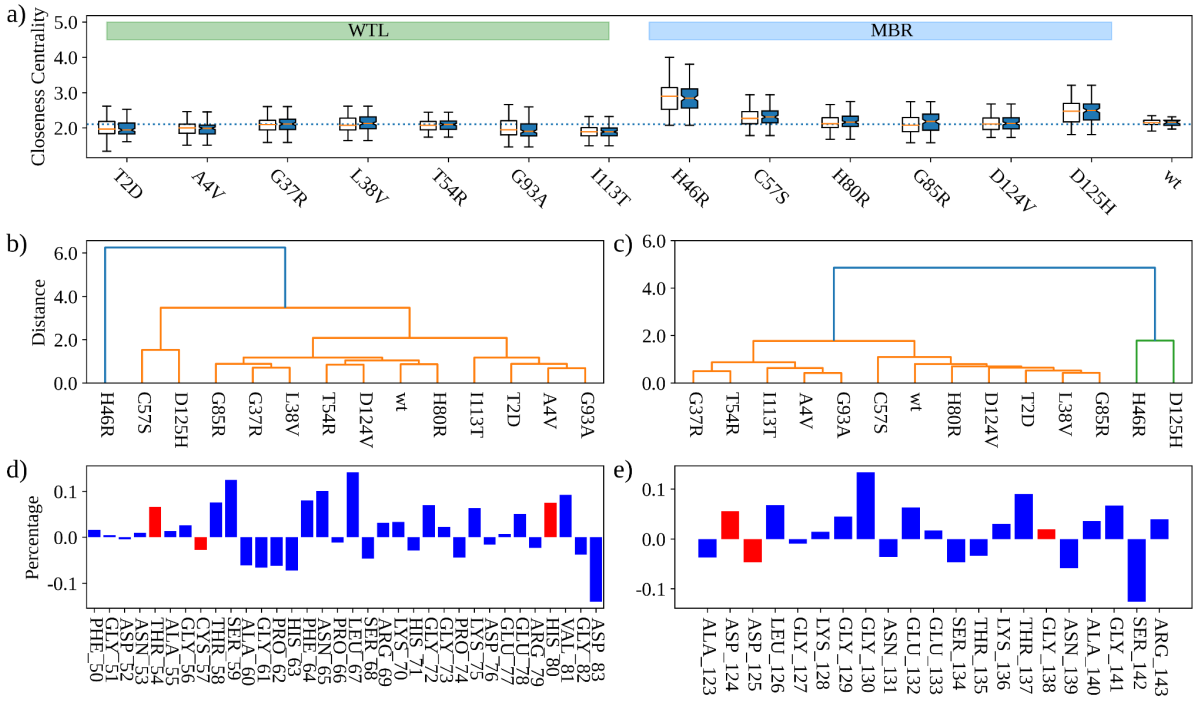

Figure 8. a) Box plots depicting Closeness centrality comparison of the wt-SOD1 and the variants. For each variant the boxplot on the left (respectively on the right) is obtained considering only the residues belonging to the MBL (resp. the ESL). The dotted line represents the mean Closeness centrality of the wt-SOD1 residues. **b)** Hierarchical clustering of the wt-SOD1 and its variants obtained considering the Closeness centrality indexes of the residues belonging to the MBL. **c)** Same as in c) but for the ESL residues. **d)** Mean difference (in percentage) between the closeness centrality indexes of the residues of the variants with respect to the wild-type ones for the MBL. **e)** Same as in d) but for the ESL.

To further compare the wt-SOD1 with the variant systems, we performed a hierarchical clustering analysis (Figure 8b-c), which allowed us to compare each system with all the others. Figure 8b shows the clustering of the systems based on the centrality values of their metal binding loop residues. The metal binding loop region of H46R displayed the most singular behaviour, since it formed a group separated from all the others. When the variants were represented by the centrality values of the electrostatic loop residues, D125H and H46R formed a cluster distinct from all other variants (Figure 8c). In general, the clusters obtained using the metal binding loop and electrostatic loop residues (Figure 8b and Figure 8c) were very similar to each other, maintaining their global structure. In line with the previous results, H46R had a different profile from all the other systems while D125H clustered together with H46R based on metal binding loop residues and with C57S based on electrostatic loop residues.

Lastly, we investigated the contribution of each residue to the observed differences in the centrality values between variants and wt-SOD1. We calculated the mean percentage difference between the closeness centrality indexes of the residues, as calculated from the wt-SOD1 and the variants. The higher this value, the larger is the mismatch between the residues in the two classes of systems. We reported the results of this analysis in Figure 8d-e for metal binding loop and electrostatic loop regions respectively. The red bar represents residues that underwent a mutation in one of the simulated systems. Remarkably, the most marked differences did not regard mutated residues, testifying the complex nature of the interactions occurring in the protein.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Covariance analysis

To study the difference between the dynamics of wt-SOD1 and its variants, we analysed the correlation between the covariance of the residue motion registered in such systems. For each MD simulation we built a covariance matrix calculating the covariance of the motion between each couple of residues. Hence, for each variant we computed the Pearson correlation coefficients between the covariance values regarding the wt and the three replicates of the variant systems. We reported in Figure 9a the mean and the standard error of the means of such correlations, where a high value means that the correlated motion between the wt and the variant systems was similar. T2D, I113R, C57S and G85R were characterised by a high mean correlation value, therefore highlighting an overall similar motion with the wt-SOD1. Conversely, once again H46R exhibits a marked difference with the wt-SOD1 behaviour.

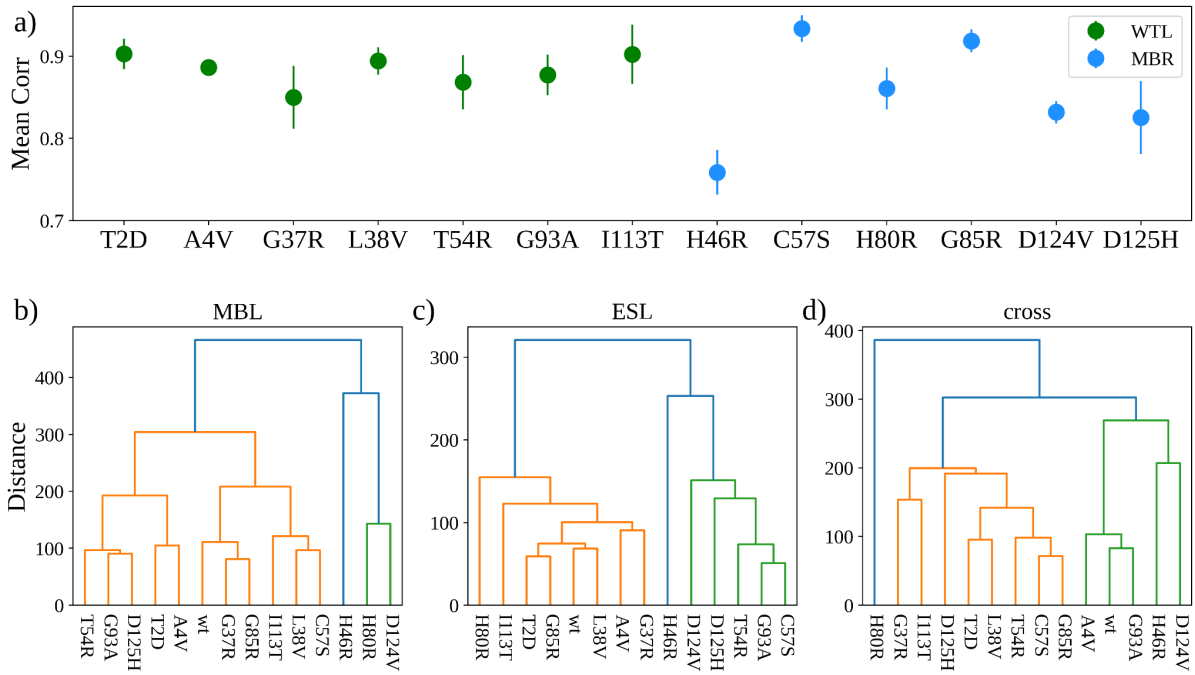

Figure 9. a) Mean correlation between the covariance matrices of the atomic coordinates of the wt-SOD1 and the variants. Bars represent the Standard Error of the Mean. Hierarchical clustering of the wt-SOD1 and 'crystal' variants obtained considering the covariances of the atomic coordinates of all the residues belonging to the **b)** MBL. **c)** ESL. **d)** covariances between the residues of the MBL and ESL.

In Figure 9b-d we showed the results of the hierarchical clustering using the covariance values as descriptors. Looking at panel b, c and d in Figure 9, variants such as G85R and G37R presented correlated motions similar to the wt-SOD1 when considering the two separately, while the cross covariances differed.

Phenotype analysis of WTL and MBR variants

In order to investigate whether WTL and MBR variants associate with different clinical outcomes, we collected clinical data of patients carrying the variants investigated in this study. We were able to retrieve clinical data of 489 ALS patients with A4V, G37R, L38V, G93R, I113T, H46R, H80R, G85R and D125H (see details of individual variants in Table 4) [34]. Cox proportional hazard analysis (Figure 10) showed that patients with MBR variants presented a longer survival time (approximately 6 years median difference, $p < 0.001$) than patients with WTL variants, while no difference ($p = 0.19$) was observed for the age of onset (details in

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 4). Because A4V and I113T represented the great majority of the *SOD1* ALS patients in our sample (312 and 120 respectively), we repeated the survival analysis excluding these two variants from the WTL group. This analysis confirmed a significant difference between WTL variants (without A4V and I113T) and MBR variants (approx. 2.5 years median difference, p = 0.006).

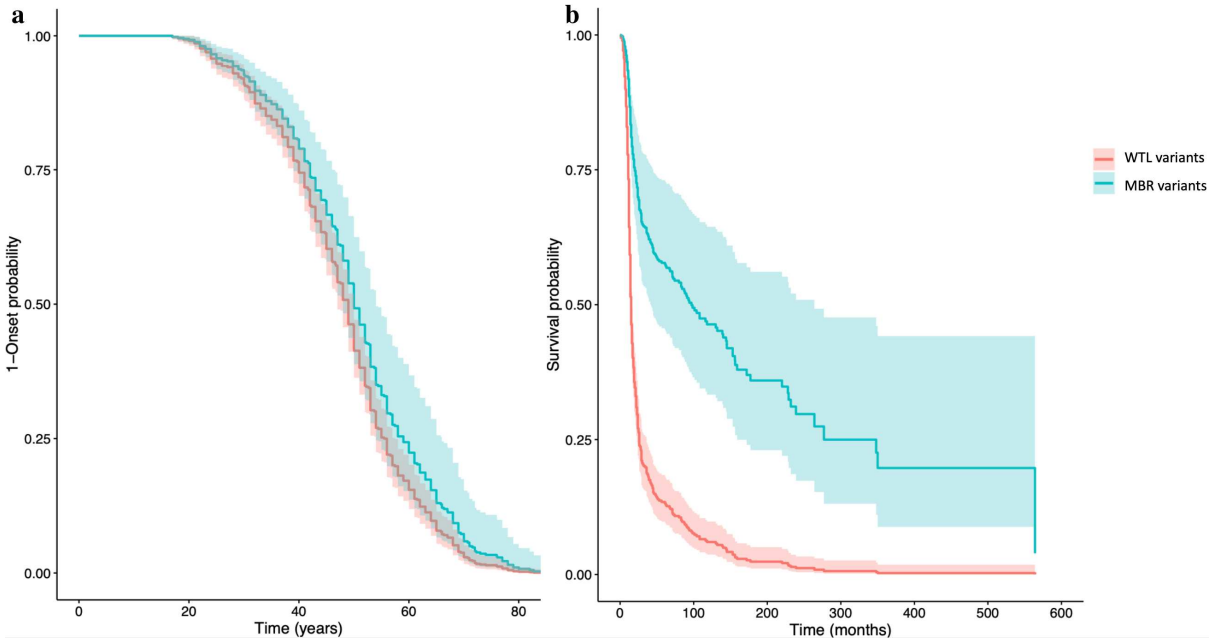

Figure 10. Age at onset (a) and survival (b) curves from the cox proportional hazard analysis (all WTL variants VS all MBR variants).

DISCUSSION

In this study we investigated the differences in the structural and dynamic behaviours of the wt-SOD1 and sets of its WTL and MBR variants associated with ALS. Our results showed the disparate nature of dynamic properties of WTL and MBR variants. We showed that in the two groups there were differences in key features such as flexibility of the metal binding loop and electrostatic loop. It was shown previously that enervating electrostatic loop can lead to the gain of interaction, thereby increasing the formation of SOD1 amyloid fibrils [16, 38, 39]. Destabilization of the electrostatic loop due to disease causing mutations has long been linked to ALS pathogenesis [40, 41]. Akin to these previous studies, we also observed a substantial flexibility in the electrostatic loop in wt-SOD1, WTL and MBR variants. However, the electrostatic loop was substantially more flexible in the MBR variants in comparison to the WTL variants and wt-SOD1, which might be the reason why, in our simulations, MBR variants adopted more extended conformations with respect to wt-SOD1 and WTL variants as suggested by the radius of gyration analysis.

Our study also illustrated that H46R had the largest impact on flexibility, compactness of the protein structure, conformational changes, and domain motions. We also showed that there was a significant difference in the flexibility of the metal binding loop and electrostatic loop in A4V and H46R variants with respect to the wt-SOD1 and these two variants represented the two flexibility extremes across the variants that we studied. Interestingly, these two variants are known to represent extremes of the phenotypic spectrum of *SOD1* ALS. A4V variant causes

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the fast progression of ALS, with a survival time of 2 years [19-21] whilst H46R causes a slow progressing form of ALS, with a mean survival of ~12 years [22-25].

The disease-causing mutations in SOD1 protein can adopt a range of misfolded states [41] which are challenging to sample computationally. Our MD simulations showed that the MBR variants, C57S and G85R, adopted a range of conformational changes, while D124V, D125H and wt-SOD1 adopted one major conformation. In comparison, the WTL variants, T2D, G37R, adopted three conformational changes. The PCA performed on the trajectories of the simulations we performed showed that the range of conformational changes observed in the WTL and MBR variants were led by the domain motions in the electrostatic loop and metal binding loop regions. The motion and flexibility of the electrostatic loop were dynamic features observed ubiquitously in the SOD1 variants we studied.

The hydrogen bond analysis performed in our study, highlighted differences in the behaviour of the wt-SOD1, WTL and MBR at an atomic level. We observed that the largest number of hydrogen bonds were formed in the wt-SOD1 and A4V variant during the simulations. This is a very interesting observation, despite acting very similar to the wt-SOD1 on an atomic level the A4V variant manifests itself as a fast-progressing form of ALS. Previous studies have pointed out that the pathogenic mutations in SOD1 might alter the protein hydrogen bonds and promote the formation of new ones which can increase the probability of misfolding and aggregation in SOD1 [42, 43]. Our results complement these by showing that apart from A4V, SOD1 variants generally form fewer hydrogen bonds which might affect the protein stability and favour misfolding.

We also showed that the dynamic behaviour of the L38V variant was unexpected. Although prior studies have grouped L38V as a WTL variant [27] and L38V and wt-SOD1 have very similar crystal structures (RMSD = 0.074 Å), the difference between their dynamic behaviours was highly noticeable. In our study we observed that the increased flexibility and dominant motion of the electrostatic loop of L38V were similar to the MBR variants. This might be due to the disruption of the β-barrel plug [27]. Interestingly, L38V has been associated with early onset of ALS (mean age of onset = 38 years) [44].

Driven by the consistent differences between WTL and MBR variants highlighted in our analyses, and by the recurrent positioning of A4V and H46R at the far ends of such differences, mirroring the extreme clinical outcomes these two variants are associated with (fast and slow progressing forms of ALS), we attempted to investigate if WTL and MBR variants led to different clinical outcomes. Using a large clinical dataset (almost 500 *SOD1*-ALS patients) we were able to show that the survival of patients with a WTL variant was substantially shorter than the survival of patients with an MBR variant (median difference approximately 6 years, $p < 0.001$).

Experimental data suggests that ALS arises from a toxic gain-of-function mechanism and loss of SOD1 function alone has been linked to a severe phenotype distinct from ALS [45-47]. However, loss of function of SOD1 in ALS has been proposed as a potential modifier. A key difference between WTL and MBR variants is their effect on the SOD1 function. MBR variants, but not WTL variants, cause reduction of SOD1 enzymatic function in *in vitro* experiments of the isolated and engineered SOD1. On such a basis, it follows that our results support the hypothesis of decoupling between the role of gain and loss of SOD1 function in ALS affecting disease onset and duration respectively and potentially independently.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Identifying and understanding which mechanisms contribute to the development of the disease and which to its progression can improve genetic counselling, the development of new therapies and the design of more effective trials. This is particularly timely for *SOD1* given the current clinical trials that are testing the efficacy of treatments based on *SOD1* antisense oligonucleotides [48]. The great difference in the survival time between the patients carrying WTL and MBR variants, and the structural features that we found associated with the two classes, could be used to improve the classification into fast and slow progression of the patients that take part to trials, and to generate better estimates of their expected survival. A correct estimation of the expected survival is essential for the design of trials and for the interpretation of its results, as potential beneficial effects on slow-progressing variants would require lengthier trials to prove them.

A potential limitation of the design and interpretation of our study is the definition of the two classes of variants. The distinction between WTL and MBR variants was based on whether or not the variants were expected to affect the catalytic activity of SOD1. However, this effect was based on *in-vitro* studies of isolated and engineered proteins. An analysis of *SOD1* activity in blood samples from ALS patients showed that SOD1 activity was approximately halved in patients carrying a wide range of variants including some of the variants we classified as WTL [49]. Although these results are apparently in contrast with the *in vitro* studies we have based the classification of WTL and MBR variants on, a possible explanation could be that the loss of SOD1 function observed in the blood of patients is not an intrinsic effect of the variants but the result of a secondary effect linked to other disease pathogenic mechanisms, e.g. SOD1 aggregation. Another possibility is that both WTL and MBR variants affect SOD1 function but with different effect sizes.

In conclusion, our study highlights key differences in the dynamics of the WTL and MBR *SOD1* variants, and wt-SOD1. It sheds light into the behaviour of *SOD1* variants at an atomic and a molecular level, suggesting interesting structural features that could be investigated to explain their associated phenotypic variability, and it supports the role of loss of function of SOD1 as a modifier of disease progression in ALS.

METHODS

Protein structures

The protein structures of wt-SOD1 and 13 variants were obtained from Protein Data Bank (PDB) [50]; the detailed structural information is shown in Table 1. SOD1 variant structures were carefully selected; only those variants which had a single point mutation were included. An in-house python script Mutatpipe.py was utilized to select and the filter the variants (https://github.com/Utilon/MutaPipe_Repo). For the variants which had more than one crystal structure, the protein structure with the highest resolution and without missing residues was selected. SOD1 exists in a homodimer form in nature, however, for the scope of this study we focused only on the SOD1 monomer; the additional chains and the copper/zinc ions present in the PDB structures were manually removed. The missing residues in 1OEZ, 3H2Q, 3H2P and 1P1V monomer structures were modelled in with Modeller 10.2 program [51].

Variants nomenclature

The old nomenclature was used to name *SOD1* variants as this reflects the actual position of the amino acids in the chains of the SOD1 structures we have used in our study. The new Human Genome Variation Society (HGVS) v2 nomenclature numbers the amino acids according to the mRNA reference sequence (GenBank: NM_000454.4) and it includes the start

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

methionine which is cleaved post-translationally. Table 1 reports both nomenclatures for each variant.

MD simulations protocol

We employed MD package GROMACS (version 2020.1) [52-54] to perform all atom MD simulations of wt-SOD1 protein, WTL and MBR variants. All the MD simulations were performed on the GPUs at *Rosalind* (https://rosalind.kcl.ac.uk), the high-performance computing (HPC) facility at King's College London. AMBER99SB-ILDN forcefield was employed for all the MD simulations performed in the study [55]. The structure was solvated in the TIP3P water model in a rhombic dodecahedron box; the protein was placed at least 1.4 nm from the edges in all directions. Na^+ and Cl^- ions were then added to neutralize the system and then the system was energy minimized over 50000 steps using the steepest descent method. The whole system consisted of ~ 30237 atoms; the number slightly varied in different variants.

After energy minimizing the system the solvent and ions around the protein were equilibrated in two steps. The first step was carried out under the NVT ensemble (constant number of particles, volume and temperature) for a length of 100 ps, at 300 K, using Berendsen thermostat. The second step was carried under NPT ensemble (constant number of particles, pressure and temperature), using the Parrinello-Rahman barostat [56] for a length of 100 ps at 1 atm and 300 K.

The final production was performed in triplicates for a length of 100 ns at 300 K for the 13 variants. In case of the wt-SOD1 the simulations were repeated 6 times for a length of 100 ns at 300 K. The covalent bonds were constrained using Linear constraint solver (LINCS) algorithm [57] and Particle Mesh Ewald (PME) algorithm was used to calculate long range interactions [58]. Leap frog integrator was used to calculate the equation of motion with a timestep of 2 fs [58]. The detailed information about the protein structures and total number of MD simulations performed in this study are summarized in Table 1.

MD trajectory analysis

The visual analysis was performed using PyMOL [59] and Visual Molecular Dynamics (VMD) program [60]. We calculated the Cα RMSF, Cα RMSD, radius of gyration and hydrogen bonds using the GROMACS modules; *gmx rmsf, gmx rms, gmx gyrate, gmx hbond.*

Principal Component Analysis

PCA is a highly used technique to reduce the dimensionality of the data obtained from molecular dynamics simulations and extract the dominant motion in the proteins [35, 36]. For PCA, the trajectories from three replicates were concatenated into one trajectory and the protein backbone was selected for analysis. *gmx covar* was then used to calculate and diagonalize the mass weighted covariance matrix. *gmx anaeig* was used to analyse the eigenvectors. The structures corresponding to PC1 and PC2 were extracted. To make the porcupine plots the structures corresponding to PC1 and PC2 were loaded to PyMOL. The vectors representing domain motion and direction were constructed after uploading the python script Modevectors.py on PyMOL (https://pymolwiki.org/index.php/Modevectors).

Statistical Analysis

All the statistical analyses in the study were performed using SciPy.Stats computing library in Python [61]. Wilcoxon Rank-sum *T*est was employed to calculate the differences in the statistical significance of the RMSF between the variants and the wt-SOD1. Age of onset and

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

survival time from disease onset analyses were performed using Cox proportional hazards regression. Models were adjusted for site of onset of symptoms and sex. The coxph() function was utilised with tie resolution at the default setting. When modelling survival time from onset, age of onset was also included as a covariate. For all statistical analyses, $p < 0.05$ was considered significant.

Graph theory analysis

We modelled each protein structure as a network, schematizing each residue as a node of the network and each intramolecular interaction as a network edge [62, 63]. In particular, we defined a weighted graph for each system, weighing each link connecting two residues with the corresponding contact frequency calculated from the molecular dynamics, defining, in this case, contact between two residues if their distance is less than 8.5˚A, similar to previously described procedures [64].

Covariance Analysis

Covariance matrices of the atomic positions of all the atoms of the metal binding loop and electrostatic loop were computed from molecular dynamics trajectories. Covariance between couples of residues were then obtained averaging the covariances of all the couples of atoms belonging to the considered residue couple.

Clinical dataset

Data on people with *SOD1*-ALS were collected from the ALS Online Database (https://alsod.ac.uk) [7, 8], the Project MinE whole-genome sequencing dataset [65], and a number of centres. The corresponding authors of the AlSoD database entries with missing data were contacted to fill the gaps. Anonymised records of people with *SOD1*-ALS were obtained from the following centres: Macquarie University, ANZAC Research Institute, University of Massachusetts, University Hospitals of Montpellier, King's College London, Washington University School of Medicine in St Louis, Peking University Third Hospital, Università degli Studi di Siena, Northwestern Medicine – Feinberg School of Medicine, Istituto Auxologico Italiano IRCCS-University of Milan, University of Belgrade and University of Calabria. Individuals were eligible to be included if their diagnosis of ALS was made by a consultant neurologist, or their data was reported in the literature with an ALS diagnosis. The clinical and demographic features utilised in this study were country of origin, sex at birth, age of onset (in years) of first motor symptoms, site of onset, survival time (in months) defined as time from diagnosis to death or latest visit. In each analysis, individuals with missing data among the subset of clinical features required were discarded. Individuals with the same *SOD1* variant, country of origin, sex, age and site of onset were considered to be duplicates.

FUNDING

The authors are supported by South London and Maudsley NHS Foundation Trust; MND Scotland; Motor Neurone Disease Association; National Institute for Health Research; Darby Rimmer MND Foundation; Spastic Paraplegia Foundation; Rosetrees Trust; Alzheimer's Research UNK; Italian Ministry of Health. M.M. and G.R. acknowledge support from European Research Council Synergy grant ASTRA (n. 855923).

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ACKNOWLEDGEMENTS

We acknowledge use of the research computing facility at King's College London, Rosalind (https://rosalind.kcl.ac.uk), which is delivered in partnership with the National Institute for Health Research (NIHR) Biomedical Research Centres at South London & Maudsley and Guy's & St. Thomas' NHS Foundation Trusts and part-funded by capital equipment grants from the Maudsley Charity (award 980) and Guy's and St Thomas' Charity (TR130505). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, King's College London, or the Department of Health and Social Care. We would like to thank people with ALS and their families for their participation in this project. We thank the Project MinE Sequencing Consortium members for their collaboration and support.

APPENDIX 1

Project MinE ALS Sequencing Consortium group authors

Philip van Damme, Philippe Corcia, Philippe Couratier, Orla Hardiman, Russell McLaughlin, Marc Gotkine, Vivian Drory, Vincenzo Silani, Nicola Ticozzi, Jan H Veldink, Leonard H van den Berg, Mamede Carvalho, Susana Pinto, Jesus Mora Pardina, Monica Povedano, Peter M Andersen, Markus Weber, Nazli Başak, Ammar Al-Chalabi, Chris Shaw, Pamela Shaw, Jonathan Cooper-Knock, Alfredo Iacoangeli, Karen Morrison, John Landers, Jonathan Glass, Patrick Vourc'h.

	SOD1 variant	HGVS V2.0 nomenclature	Pathogenicity	PDB ID	Missing residues	Res (Å)	Protein chain	Run time (ns)	Replicates
1.	T2D	T3D	Likely pathogenic	5K02	-	1.99	A	100	3
2.	A4V	A5V	Pathogenic	6SPH	-	2.25	A	100	3
3.	G37R	G38R	Pathogenic	1AZV	-	1.9	A	100	3
4.	L38V	L39V	Pathogenic	2WYT	-	1.0	A	100	3
5.	H46R	H47R	Pathogenic	1OEZ	66-77, 126-141	2.15	A	100	3
6.	T54R	T55R	Likely pathogenic	3ECW	-	2.15	A	100	3
7.	C57S	C58S	Likely pathogenic	4MCM	-	2.20	A	100	3
8.	H80R	H81R	Pathogenic	3H2Q	67-78, 134-138	1.85	A	100	3
9.	G85R	G86R	Pathogenic	2VR6	-	1.30	A	100	3
10.	G93A	G94A	Pathogenic	3GZO	-	2.10	A	100	3
11.	I113T	I114T	Pathogenic	1UXL	-	1.60	A	100	3
12.	D124V	D125V	Pathogenic	3H2P	67-78, 125-139	1.55	A	100	3
13.	D125H	D126H	Pathogenic	1P1V	129-135	1.40	A	100	3
14.	wt-SOD1	-	-	2C9V	-	1.07	A	100	3

Table 1: List of PDB structures of the wt-SOD1 and its variants used in the study.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	SOD1 variant	Mean (nm)	Median (nm)
1.	T2D	0.18	0.17
2.	A4V	0.20	0.20
3.	G37R	0.20	0.22
4.	L38V	0.22	0.22
5.	T54R	0.17	0.18
6.	G93A	0.22	0.22
7.	I113T	0.16	0.17
8.	H46R	0.63	0.64
9.	C57S	0.23	0.24
10.	H80R	0.30	0.30
11.	G85R	0.20	0.21
12	D124V	0.36	0.37
13.	D125H	0.33	0.34
14.	wt	0.22	0.22

Table 2. The mean and median of the RMSD calculated for the WTL, MBR variants and wt-SOD1.

	SOD1 variant	PC1 (%)	PC2 (%)	PC3 (%)	PC4 (%)
1.	A4V	28.93	19.89	9.31	6.25
2.	C57S	35.31	24.03	5.63	3.44
3.	D124V	30.17	20.63	5.46	4.97
4.	D125H	30.88	15.22	9.25	5.71
5.	G37R	51.56	12.36	5.74	3.24
6.	G85R	37.39	10.61	5.84	4.49
7.	G93A	22.21	13.09	11.35	7.04
8.	H46R	36.99	25.66	6.52	4.24
9.	H80R	29.2	13.27	9.06	5.21
10.	I113T	33.31	20.78	8.04	5.14
11.	L38V	40.32	18.38	7.23	5.29
12.	T2D	32.74	9.58	6.5	5.22
13.	T54R	27.44	13.15	9.17	5.71
14.	wt-SOD1	27.45	21.57	10.59	4.05

Table 3. Percentage of variance represented by PC1, PC2, PC3, PC4.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	SOD1 variant	HGVS V2.0 nomenclature	Total sample size	Patients with disease duration	Number censored	Mean disease duration (months)	Median disease duration (months)	Patients with age of onset	Mean age of onset (years)	Median age of onset (years)
1.	A4V	A5V	312	260	7	15.75	13	298	48.61	49.46
2.	D125H	D126H	3	1	0	8.08	8.08	3	50	52
3.	G37R	G38R	10	6	0	182.33	186	10	35.55	33.98
4.	G85R	G86R	18	17	1	42.24	30	17	60	60
5.	G93R	G94R	1	1	0	55	55	1	34	34
6.	H46R	H47R	19	18	12	318.65	277	19	45.16	43
7.	H80R	H81R	1	1	0	18	18	1	24	24
8.	I113T	I114T	120	86	15	94.36	62	108	53.81	53
9.	L38R	L39R	5	4	0	19.66	23.5	5	44.09	43
10.	All WTL	All WTL	448	357	22	37.32	14	422	49.54	50
11.	All WTL (no A4V and I113T)	All WTL (no A5V and I114T)	16	11	0	111.6	55	16	38.12	38.33
12.	All MBR	All MBR	41	37	13	176.65	84.17	40	51.3	50

Table 4. Characteristics of the clinical dataset.

REFRENCES

1. Cleveland, D.W. and J.D. Rothstein, *From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS.* Nature reviews neuroscience, 2001. **2**(11): p. 806-819.
2. Boillée, S., C.V. Velde, and D.W. Cleveland, *ALS: a disease of motor neurons and their nonneuronal neighbors.* Neuron, 2006. **52**(1): p. 39-59.
3. Perrone, B. and F.L. Conforti, *Common mutations of interest in the diagnosis of amyotrophic lateral sclerosis: how common are common mutations in ALS genes?* Expert Review of Molecular Diagnostics, 2020. **20**(7): p. 703-714.
4. Antonyuk, S., et al., *Structural consequences of the familial amyotrophic lateral sclerosis SOD1 mutant His46Arg.* Protein Science, 2005. **14**(5): p. 1201-1213.
5. Stathopulos, P.B., et al., *Calorimetric analysis of thermodynamic stability and aggregation for apo and holo amyotrophic lateral sclerosis-associated Gly-93 mutants of superoxide dismutase.* Journal of biological chemistry, 2006. **281**(10): p. 6184-6193.
6. Furukawa, Y., et al., *Disulfide cross-linked protein represents a significant fraction of ALS-associated Cu, Zn-superoxide dismutase aggregates in spinal cords of model mice.* Proceedings of the National Academy of Sciences, 2006. **103**(18): p. 7148-7153.
7. Abel, O., et al., *ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics.* Human mutation, 2012. **33**(9): p. 1345-1351.
8. Wroe, R., et al., *ALSOD: the amyotrophic lateral sclerosis online database.* Amyotrophic Lateral Sclerosis, 2008. **9**(4): p. 249-250.
9. Wright, G.S., S.V. Antonyuk, and S.S. Hasnain, *The biophysics of superoxide dismutase-1 and amyotrophic lateral sclerosis.* Quarterly reviews of biophysics, 2019. **52**.
10. Huai, J. and Z. Zhang, *Structural properties and interaction partners of familial ALS-associated SOD1 mutants.* Frontiers in Neurology, 2019. **10**: p. 527.
11. Furukawa, Y., et al., *Conformational disorder of the most immature Cu, Zn-superoxide dismutase leading to amyotrophic lateral sclerosis.* Journal of Biological Chemistry, 2016. **291**(8): p. 4144-4155.
12. Rothstein, J.D., *Current hypotheses for the underlying biology of amyotrophic lateral sclerosis.* Annals of neurology, 2009. **65**(S1): p. S3-S9.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

13. Banci, L., et al., *Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: a possible general mechanism for familial ALS.* Proceedings of the National Academy of Sciences, 2007. **104**(27): p. 11263-11267.
14. Oztug Durer, Z.A., et al., *Loss of metal ions, disulfide reduction and mutations related to familial ALS promote formation of amyloid-like aggregates from superoxide dismutase.* PloS one, 2009. **4**(3): p. e5004.
15. Ray, S.S., et al., *An intersubunit disulfide bond prevents in vitro aggregation of a superoxide dismutase-1 mutant linked to familial amytrophic lateral sclerosis.* Biochemistry, 2004. **43**(17): p. 4899-4905.
16. Healy, E.F., *A prion-like mechanism for the propagated misfolding of SOD1 from in silico modeling of solvated near-native conformers.* PloS one, 2017. **12**(5): p. e0177284.
17. Grad, L.I., et al., *Intermolecular transmission of superoxide dismutase 1 misfolding in living cells.* Proceedings of the National Academy of Sciences, 2011. **108**(39): p. 16398-16403.
18. Grad, L.I., et al., *Intercellular propagated misfolding of wild-type Cu/Zn superoxide dismutase occurs via exosome-dependent and-independent mechanisms.* Proceedings of the National Academy of Sciences, 2014. **111**(9): p. 3620-3625.
19. Saeed, M., et al., *Age and founder effect of SOD1 A4V mutation causing ALS.* Neurology, 2009. **72**(19): p. 1634-1639.
20. Andersen, P.M., *Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene.* Current neurology and neuroscience reports, 2006. **6**(1): p. 37-46.
21. Hays, A.P., et al., *Sporadic amyotrophic lateral sclerosis and breast cancer: hyaline conglomerate inclusions lead to identification of SOD1 mutation.* Journal of the neurological sciences, 2006. **242**(1-2): p. 67-69.
22. Zou, Z.-Y., et al., *H46R SOD1 mutation is consistently associated with a relatively benign form of amyotrophic lateral sclerosis with slow progression.* Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 2016. **17**(7-8): p. 610-613.
23. Holmøy, T., K. Bjørgo, and P.M. Roos, *Slowly progressing amyotrophic lateral sclerosis caused by H46R SOD1 mutation.* European Neurology, 2007. **58**(1): p. 57.
24. Arisato, T., et al., *Clinical and pathological studies of familial amyotrophic lateral sclerosis (FALS) with SOD1 H46R mutation in large Japanese families.* Acta neuropathologica, 2003. **106**(6): p. 561-568.
25. Suzuki, N., et al., *Genetics of amyotrophic lateral sclerosis: seeking therapeutic targets in the era of gene therapy.* Journal of Human Genetics, 2022: p. 1-22.
26. Valentine, J.S., P.A. Doucette, and S. Zittin Potter, *Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis.* Annu. Rev. Biochem., 2005. **74**: p. 563-593.
27. Hayward, L.J., et al., *Decreased Metallation and Activity in Subsets of Mutant Superoxide Dismutases Associated with Familial Amyotrophic Lateral Sclerosis* 210.* Journal of Biological Chemistry, 2002. **277**(18): p. 15923-15931.
28. Valentine, J.S. and P.J. Hart, *Misfolded CuZnSOD and amyotrophic lateral sclerosis.* Proceedings of the National Academy of Sciences, 2003. **100**(7): p. 3617-3622.
29. Pardo, C.A., et al., *Superoxide dismutase is an abundant component in cell bodies, dendrites, and axons of motor neurons and in a subset of other neurons.* Proceedings of the National Academy of Sciences, 1995. **92**(4): p. 954-958.
30. Bakavayev, S., et al., *Cu/Zn-superoxide dismutase and wild-type like fALS SOD1 mutants produce cytotoxic quantities of H2O2 via cysteine-dependent redox short-circuit.* Scientific reports, 2019. **9**(1): p. 1-13.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

31. Hörnberg, A., et al., *The coupling between disulphide status, metallation and dimer interface strength in Cu/Zn superoxide dismutase.* Journal of molecular biology, 2007. **365**(2): p. 333-342.
32. Vassall, K.A., et al., *Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in ALS.* Proceedings of the National Academy of Sciences, 2011. **108**(6): p. 2210-2215.
33. Sekhar, A., et al., *Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy.* Proceedings of the National Academy of Sciences, 2016. **113**(45): p. E6939-E6945.
34. Opie-Martin, S., et al., *The SOD1-mediated ALS phenotype shows a decoupling between age of symptom onset and disease duration.* Nature Communications, 2022. **13**(1): p. 1-9.
35. Amadei, A., et al., *An efficient method for sampling the essential subspace of proteins.* Journal of Biomolecular Structure and Dynamics, 1996. **13**(4): p. 615-625.
36. Amadei, A., A.B. Linssen, and H.J. Berendsen, *Essential dynamics of proteins.* Proteins: Structure, Function, and Bioinformatics, 1993. **17**(4): p. 412-425.
37. Yamaguchi, H., et al., *Essential dynamics of DNA containing a cis. syn cyclobutane thymine dimer lesion.* Nucleic acids research, 1998. **26**(8): p. 1939-1946.
38. Healy, E.F., *A mechanism for propagated SOD1 misfolding from frustration analysis of a G85R mutant protein assembly.* Biochemical and biophysical research communications, 2016. **478**(4): p. 1634-1639.
39. Souza, P.C., et al., *An allosteric pathway in copper, zinc superoxide dismutase unravels the molecular mechanism of the G93A amyotrophic lateral sclerosis-linked mutation.* The journal of physical chemistry letters, 2019. **10**(24): p. 7740-7744.
40. Molnar, K.S., et al., *A common property of amyotrophic lateral sclerosis-associated variants: destabilization of the copper/zinc superoxide dismutase electrostatic loop.* Journal of Biological Chemistry, 2009. **284**(45): p. 30965-30973.
41. Prudencio, M. and D.R. Borchelt, *Superoxide dismutase 1 encoding mutations linked to ALS adopts a spectrum of misfolded states.* Molecular neurodegeneration, 2011. **6**(1): p. 1-19.
42. Alemasov, N.A., et al., *Molecular mechanisms underlying the impact of mutations in SOD1 on its conformational properties associated with amyotrophic lateral sclerosis as revealed with molecular modelling.* BMC structural biology, 2018. **18**(1): p. 1-14.
43. Alemasov, N.A., et al., *Computer analysis of the relation between hydrogen bond stability in SOD1 mutants and the survival time of amyotrophic lateral sclerosis patients.* Journal of Molecular Graphics and Modelling, 2022. **110**: p. 108026.
44. Régal, L., et al., *The G93C mutation in superoxide dismutase 1: clinicopathologic phenotype and prognosis.* Archives of neurology, 2006. **63**(2): p. 262-267.
45. Baskoylu, S.N., et al., *Single copy/knock-in models of ALS SOD1 in C. elegans suggest loss and gain of function have different contributions to cholinergic and glutamatergic neurodegeneration.* PLoS genetics, 2018. **14**(10): p. e1007682.
46. Saccon, R.A., et al., *Is SOD1 loss of function involved in amyotrophic lateral sclerosis?* Brain, 2013. **136**(8): p. 2342-2358.
47. Park, J.H., et al., *SOD1 deficiency: a novel syndrome distinct from amyotrophic lateral sclerosis.* Brain, 2019. **142**(8): p. 2230-2237.
48. Miller, T.M., et al., *Trial of antisense oligonucleotide Tofersen for SOD1 ALS.* New England Journal of Medicine, 2022. **387**(12): p. 1099-1110.
49. Keskin, I., et al., *Comprehensive analysis to explain reduced or increased SOD1 enzymatic activity in ALS patients and their relatives.* Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 2017. **18**(5-6): p. 457-463.

bioRxiv preprint doi: https://doi.org/10.1101/2022.12.05.519128; this version posted December 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

50. Berman, H.M., et al., *The protein data bank.* Nucleic acids research, 2000. **28**(1): p. 235-242.
51. Eswar, N., et al., *Comparative protein structure modeling using Modeller.* Current protocols in bioinformatics, 2006. **15**(1): p. 5.6. 1-5.6. 30.
52. Bekker, H., et al. *Gromacs-a parallel computer for molecular-dynamics simulations*. in *4th International Conference on Computational Physics (PC 92)*. 1993. World Scientific Publishing.
53. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-854.
54. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1**: p. 19-25.
55. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(8): p. 1950-1958.
56. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method.* Journal of Applied physics, 1981. **52**(12): p. 7182-7190.
57. Hess, B., et al., *LINCS: a linear constraint solver for molecular simulations.* Journal of computational chemistry, 1997. **18**(12): p. 1463-1472.
58. Essmann, U., et al., *A smooth particle mesh Ewald method.* The Journal of chemical physics, 1995. **103**(19): p. 8577-8593.
59. Janson, G., et al., *PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL.* Bioinformatics, 2017. **33**(3): p. 444-446.
60. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.
61. Virtanen, P., et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nature methods, 2020. **17**(3): p. 261-272.
62. Miotto, M., et al., *Insights on protein thermal stability: a graph representation of molecular interactions.* Bioinformatics, 2019. **35**(15): p. 2569-2577.
63. Miotto, M., et al., *Simulated epidemics in 3d protein structures to detect functional properties.* Journal of Chemical Information and Modeling, 2020. **60**(3): p. 1884-1891.
64. Miotto, M., et al., *Inferring the stabilization effects of SARS-CoV-2 variants on the binding with ACE2 receptor.* Communications biology, 2022. **5**(1): p. 1-13.
65. *Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis.* European Journal of Human Genetics, 2018. **26**(10): p. 1537-1546.