bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Distribution of disease-causing germline mutations in coiled-coils suggests essential role of their N-terminal region

Zsofia E. Kalman[1,2], Bálint Mészáros[3], Zoltán Gáspári[1,*], Laszlo Dobson[1,4*]
[1]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/A, 1083 Budapest, Hungary
[2]3in-PPCU Research Group, 2500 Esztergom, Hungary
[3]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany
[4]Research Centre for Natural Sciences, Magyar Tudósok Körútja 2, 1117 Budapest, Hungary
*To whom correspondence should be addressed: gaspari.zoltan@itk.ppke.hu, dobson.laszlo@ttk.mta.hu

Abstract

Next-generation sequencing resulted in the identification of a huge number of naturally occurring variations in human proteins. The correct interpretation of the functional effects of these variations necessitates the understanding of how they modulate protein structure. Coiled-coils are α-helical structures responsible for a diverse range of functions, but most importantly, they facilitate the structural organization of macromolecular scaffolds *via* oligomerization. In this study, we analyzed a comprehensive set of disease-associated germline mutations in coiled-coil structures. Our results highlight the essential role of residues near the N-terminal part of coiled-coil regions, possibly critical for superhelix assembly and folding in some cases. We also show that coiled-coils of different oligomerization states exhibit characteristically distinct patterns of disease-causing mutations. Our study provides structural and functional explanations on how disease emerges through the mutation of these structural motifs.

Introduction

Advances in sequencing resulted in the identification of a huge number of different Single Nucleotide Variations (SNVs) of various genomic positions among individuals in healthy and disease states. Variations can impact proteins on several levels, ranging from polymorphisms (PMs) with negligible effect on fitness to lethal mutations through increasingly strong phenotypes. According to their origin, disease-causing genetic alterations can be broadly categorized into either somatic or germline mutations. Somatic mutations, most notably responsible for tumorigenesis, are confined to the cell they originated in and its daughter cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In accord, their phenotypic change can be extreme with abolishing cell-cycle control, escaping apoptosis and achieving cellular immortality. In contrast, disease-associated germline mutations (DMs) persist in the whole organism and are transmitted from generation to generation. Thus, DMs cause relatively weak changes in phenotypes, yet they still have a noticeable negative impact on the quality of life.

Coiled-coils are oligomeric helical structural units in proteins connected to a wide range of functions. Several coiled coil proteins have been shown to have catalytic activity[1] or undergo oligomerization, yet their most common functions arise directly from their structure: they are molecular spacers separating or connecting domains[2]. They can bridge large distances and connect proteins of different sides of large structural organizations like the postsynaptic density where the Homer coiled coil serves as a direct connection between the plasma membrane and intracellular proteins through EVH1 domains binding to various scaffolding partners[3].

Coiled-coils are α-helical domains consisting of two or more helices packed together in a specific knobs-into-holes manner[4], with interhelical interactions playing a dominant role in folding[5]. Coiled-coils can have parallel or antiparallel arrangement, and they can be formed by intrachain interaction of the same subunit, or by interchain bonds between distinct polypeptide chains[6]. Regardless of their oligomerization state, the main forces driving their interactions are the formation of hydrophobic contacts at their inside, often supported by electrostatic interactions aiding the stability from the outside[7]. Coiled-coil residues can be classified into register positions according to their role in complex formation with the opposing helices: the most common repeat pattern is heptad ('abcdefg'), with 'a' and 'd' positions being responsible for hydrophobic interactions, while 'e' and 'g' positions contain (oppositely) charged residues[8]. Notably, other variants (e.g., hendecads are 11 residue repeats) were also discovered[9]. Folding studies of selected coiled-coils indicated the importance of a specific segment, the trigger sequence, that is required for initiating the proper interaction between the helices[10]. However, it is not yet entirely clear whether specific sequence patterns are required for assembly, or the accumulation of interaction promoting residues at critical positions generally aids coiled-coil formation.

In recent decades many studies addressed how DMs perturb protein structure. The majority of significantly occurring structural elements (e.g., transmembrane [11] and intrinsically disordered protein regions[12]), as well as various structurally distinct functional regions (e.g., protein-protein interfaces, buried domains[13]) were analyzed in detail. However, coiled-coils are a largely understudied class in this respect with only individual cases discussed. To our knowledge, only one large-scale study has been published, highlighting the critical role of register positions and pointing out mutations frequently associated with pleiotropy[14]. In this study we integrate multiple prediction algorithms and structural information for an in-depth analysis to assess how non-synonymous disease-associated germline mutations affect coiled-coil structures and thereby their functions. Our work revealed that disrupting hydrophobic and electrostatic interactions impairs coiled-coil structure and disease-associated mutations accumulate near the N-terminal of coiled-coil regions. We also showed that even if their destabilizing effect is small, DMs are enriched in antiparallel homodimer coiled-coils. On one hand, understanding how these variations modulate the structure and function of proteins may improve prediction algorithms. On the other hand, the rational coiled-coil design can be achieved through a detailed

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

understanding of the sequence-structure relationship[15]. However, a missing piece of the puzzle is how DMs perturb coiled-coil structures.

Results

DMs primarily target coiled-coil segment containing proteins outside the coiled-coil region

To obtain an overall picture of how disease-associated mutations (DM) and coiled-coils are related, we determined the number of proteins in the human proteome that contain coiled-coil regions and at least one DM or polymorphism (PM). To define coiled-coil regions, we used four different predictors (Deepcoil, Marcoil, Ncoils and Paircoil). We calculated all statistics separately for each method (detailed in the supplementary material) and also for the consensus of the predictions.

In general, PMs have a higher overall abundance in the human proteome than DMs, as observed in many previous studies. At the protein level, there is a slight increase in the frequency of DMs in coiled-coil containing proteins compared to other proteins in the proteome, yielding an odds ratio of 1.03 against PMs (Supplementary Figure 1). Notably, prediction methods exhibit some inconsistency regarding these statistics: DeepCoil predicts less coiled-coils and produces a negative odds ratio, while the other three methods show the inverse trend. Surprisingly in the light of the above observation, coiled-coil regions are depleted in DMs compared to non-coiled-coil residues, agreed by all prediction methods. A possible explanation for this phenomenon might be that coiled-coils are generally less vulnerable to harmful variations. However, these proteins carry essential functions as they are more frequently occupied by DMs.

Coiled-coils are often perturbed by DMs affecting charged residues

The main driving force of protein domain folding and stability is achieved through hydrophobic interactions. Coiled-coils are special structural units where the balanced contribution of hydrophobic interactions and electrostatic interactions aid the stability together. This is reflected by the different amino acid preferences of the different positions in the heptad repeat unit, corresponding to the distinct spatial position and role of these within the superhelical structure. To assess how residues are affected by variations, we grouped amino acids based on their basic physico-chemical features (positive: HKR, negative: DE, hydrophobic: AILMV, other: CFGNPQSTWY), then we calculated the log ratio of the substitutions observed in DMs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

From \ To	Hydrophobic	Positive	Negative	Other		Σ
Hydrophobic	-0.014	0.019	-0.062	0.042		-0.016
Positive	0.108	0.086	0.219	0.154		0.567
Negative	0.131	0.235	0.264	0.101		0.731
Other	-0.270	-0.167	-0.233	-0.147		-0.818

Coiled-coil ← → Human proteome

Figure 1: Amino acid changes in coiled-coils. *Left) Residue change preferences by DMs in the proteome (negative values, also marked with the shades of blue) and in coiled-coil regions (positive values, also marked with the shades of red). Values show the logarithm of ratio of DMs changing given residues types Right) Targeted residue type preferences by DMs in the proteome (negative values) and in coiled-coil regions (positive values).*

Figure 1 shows preferred residue type changes in coiled-coils relative to other proteins of the proteome. We calculated the relative frequency of amino acid substitutions in the coiled-coil regions, and in the proteome, then calculated the log ratio of substitution frequencies. According to our results hydrophobic residues are targeted in similar proportions. However, in the case of coiled-coils, charged residues aiding electrostatic interactions are much more frequently affected by DMs (Figure 1/right).

In contrast, several residue types indispensable for stable domain structure (e.g., cysteines forming disulfide bridges) do not influence coiled-coil formation, thus their replacement does not cause stability problems (Figure 1/left, for more details see Supplementary Figure 2). The most prevalent changes in coiled-coil regions by DMs are replacements by oppositely charged residues. Interestingly, the negatively charged Glu and Asp are generally not interchangeable residues in coiled-coils, in contrast to the positively charged residues Lys and Arg. In coiled-coils DMs most likely target A,E,I,K,L,M,N and Q residues, as opposed to C, G and P residues being more often targeted in other proteins in the proteome (Supplementary Figure 2).

DMs accumulate at the N-terminal region of coiled-coils

We investigated the distribution of variations in coiled-coils, considering their coverage, abundance in the N-terminal region, and coiled-coil length. We divided coiled-coil regions into five equal parts, and calculated the proportion of variations in these parts. Although the first half of the sequences contain slightly more DMs, the difference is not significant compared to PMs (Supplementary Figure 3).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

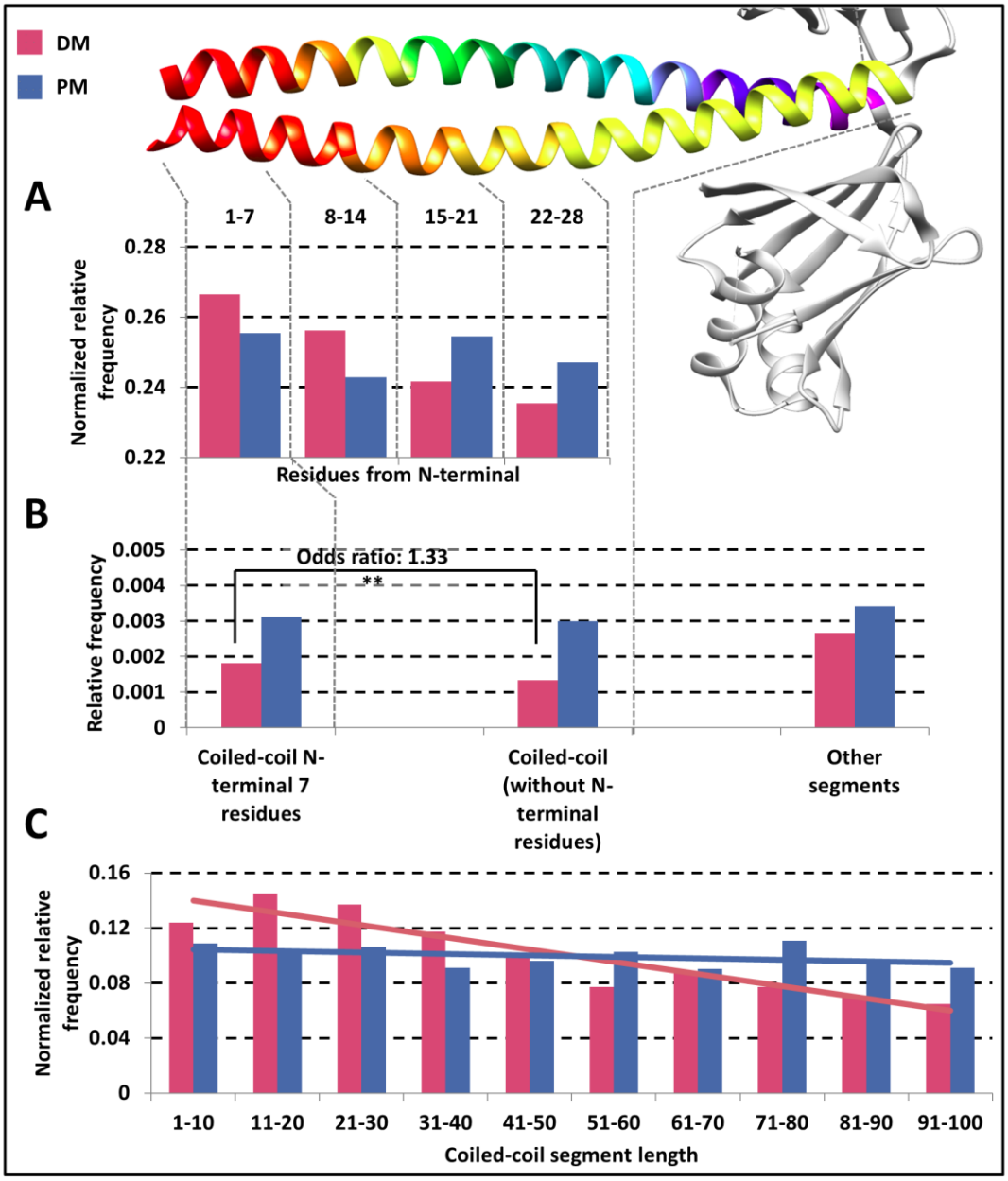


Figure 2: Variations along coiled-coil segments. *Distribution of variations in the sequence. A) X-axis shows the coverage of the N-terminal of coiled-coils. B) Relative frequency of coiled-coil residues targeted at N-terminal of the coiled-coil, other coiled-coil residues and other segments of proteins, respectively. C) Distributions of variations in coiled-coils with different lengths (linear trend lines were aligned to the data). Red: DMs; blue: PMs.*

To reduce bias originating from the varied length of coiled-coils, we performed the enrichment calculation considering only the first 28 residues of all coiled-coils, this time by dividing sequences into four equal parts, i.e., using seven residue bins - keeping in mind that predictors were optimized for heptad repeats (Figure 2/A). Using this approach, the accumulation of DMs

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

at the N-terminal became visible, showing a monotonous decline of DMs towards the C-terminal.

To demonstrate that the first seven residues of coiled-coils contain significantly more DMs compared to the rest of the coiled-coil regions, we counted the number of DMs and PMs in the first seven residues of coiled-coils and in their succeeding part. The result is significant (χ^2 test, $p<0.01$), and the odds ratio between DMs and PMs is 1.33 (Figure 2/B).

Notably, this effect is strong enough to influence the distribution of variations considering coiled-coils with different lengths. The relative frequency of DMs is significantly higher in shorter coiled-coils (Figure 2/C), as they utilize most of their residues as “N-terminal” segment that may contribute to stability, while in longer coiled-coils, other residues have a lesser role in sustaining the complex form. In contrast, PMs show uniform distribution in coiled-coil regions with different lengths. Notably, it is arguable whether very short predicted coiled-coil segments (below 10 residues) are biologically relevant, however we did not want to tailor prediction outputs. Moreover, omitting the first bin only further strengthens our result.

We performed the same analysis around C-terminal residues, however according to our results the accumulation of DMs is only detectable at the N-terminal region (data not shown).

Oligomerization state affects which register positions are vulnerable

The periodic property of coiled-coils enables a position type classification of residues, grouping amino acid positions based on their location in the helix, uncovering preferred physico-chemical features and interaction types. Regardless of oligomerization state, residues at “a” and “d” positions are often hydrophobic and face each other, forming the core of the complex, while “e” and “g” residues may be charged and promote stability via electrostatic interactions on the outer face of the superhelical structure.

We analyzed the distribution of variations on the different positions, considering different oligomerization states. As expected from early results, PMs are more abundant in every heptad position. However, considering DMs only, residues falling into “stabilizing” positions are more vulnerable to variations (Figure 3, left). Interestingly, residue type changes affect the heptad positions differently: replacement of amino acids in “a” and “d” positions likely perturb the structure, even when the substitution seems conserving on the basis of physicochemical properties (i.e., variation replacing a hydrophobic residue to another one is also often harmful). In contrast, “e” and “g” positions seem to be slightly more resilient, and residue type change (i.e., charge change) is more often required to disrupt the structure. PMs change the residue type to a lesser extent.

The oligomerization state of the coiled-coil also affects its vulnerability. In general, antiparallel formations (both dimers and tetramers) are slightly more preferred targets of DMs. Oligomerization also influences which positions are modulated (Figure 3, right): “e” and “g” (charged) positions are more often affected by DMs in parallel dimers, while “a” and “d” (hydrophobic) positions are primarily targeted in antiparallel dimers. Hydrophobic interaction promoting positions are less likely to be targeted by DMs in parallel dimers. When these positions are mutated, the mutation changes the type of the residue in almost every case, showing an opposite trend compared to other oligomerization modes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Variations in trimeric and tetrameric coiled-coils are similar: in these cases, structures are most often perturbed via amino acids in “a”, “g” and “e” positions and also often replace residue type. DMs on “d” positions are rare.

In general, there seems to be an opposing trend, that in positions where the DM frequency is lower, any change can carry disease, while in positions where the DM frequency is higher, the mutations more likely change the physico-chemical property of the residue.

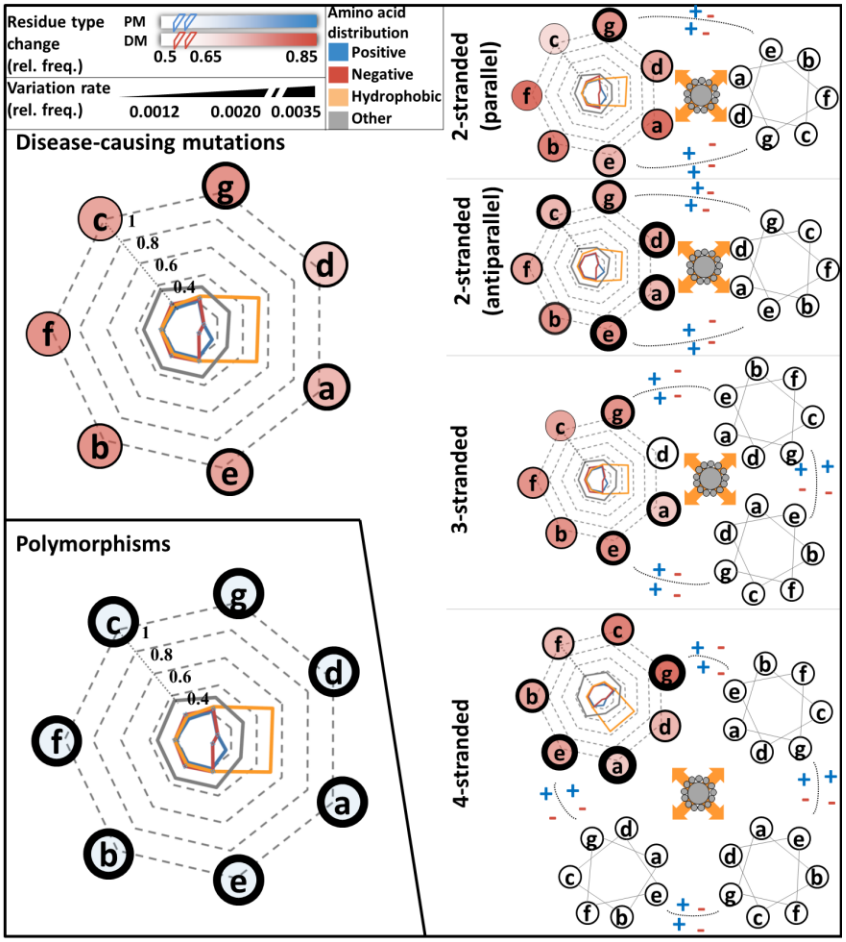


Figure 3: Distribution of variations in coiled-coils. *Amino acids were grouped according to their physico-chemical properties (positive: HKR, negative: DE, small hydrophobic: AILMV, other: CFGNPQSTWY). Radars represent the amino acid distributions in different* positions*. Line thickness around* positions *is proportional to variation frequency. The opacity of positions is proportional to the rate of variations changing the physico-chemical features of the targeted residue. Left) DMs and PMs in coiled-coils. Middle) DMs in coiled-coils with different oligomerization states.*

Structure analysis reveals most DMs occur in homooligomeric coiled-coils with a subtle destabilizing effect

To gain detailed insights on how DMs perturb the formation of coiled-coils, we searched for structures in the PDB and identified coiled-coil segments using SOCKET. Although the number

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of variations falling into characterized coiled-coil structures is rather low, and sometimes insufficient for performing reliable statistical tests to draw convincing conclusions, such analysis can open prospects to recognize interesting trends.

First, we analyzed the distribution of DMs in different heptad positions. As the number of cases was low, we classified the positions into three categories: responsible for hydrophobic stabilization (a,d), electrostatic stabilization (e,g) and outward facing/solvent exposed (b,c,f). Disease-associated mutations are enriched on residues responsible for forming the hydrophobic core of coiled-coils, have nearly the same occurrence as PMs in positions reserved for charged residues, and show decline on outward facing residues (Figure 4/A). Although at first glance this does not seem to confirm prediction data where DMs have a higher frequency on 'e' and 'g' positions. However, this discrepancy is due to the very different composition of the two datasets with regard to oligomerization state: the most prevalent class of structures are two-stranded antiparallel coiled-coils, the only class where mutations on hydrophobic positions dominate in prediction data too (Figure 3).

Next, we investigated whether N-terminal segments of the coiled-coils gather more variations. Although both types of variations (PMs and DMs) seem to accumulate in the first seven residues of coiled-coils, the two kinds of variations exhibit an opposing trend, with a higher frequency of DMs around the N-terminal and PMs in other residues (Figure 4/B). Moreover, while PM data is not significant, the distribution of DMs is slightly significant according to the $\chi 2$ test ($p<0.1$).

Sequence data alone can be rather difficult to utilize for defining the monomeric or oligomeric state of coiled-coil assemblies, and predictions are also limited in detecting how many strands the coiled-coils are composed of. However, from structural data we can readily classify coiled-coils as monomers (both strands are part of the same protein, typically antiparallel coiled coils with a short linker between the two helices), homooligomers (interaction of identical proteins) or heterooligomers (interaction between different proteins). While PMs show a uniform distribution among these classes, DMs mainly occur in homooligomers (Figure 4/C). The rationale behind this can be that a single mutation might (but in a heterozygous case, not necessarily) affect multiple constituent helices simultaneously, so their effect is instantly multiplied, in contrast to heterooligomers and monomers where the interacting partner/segment does not amplify the impact of the mutation.

The energy change calculated by the introduced mutation can be used as an approximation of the contribution of a mutation to the overall stability of the coiled coil. Figure 4/D shows the calculated energy changes upon mutation in the proteome and in coiled-coil structures. Generally, the mean energetic contribution of PMs can outline the range of changes a protein can tolerate without damage. In both cases (proteome, and coiled-coil proteins), DMs have an average higher $\Delta\Delta G$. However, in the case of coiled-coils, despite keeping the same trend, both variation types seem to have a lower effect compared to those of other proteins of the proteome.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

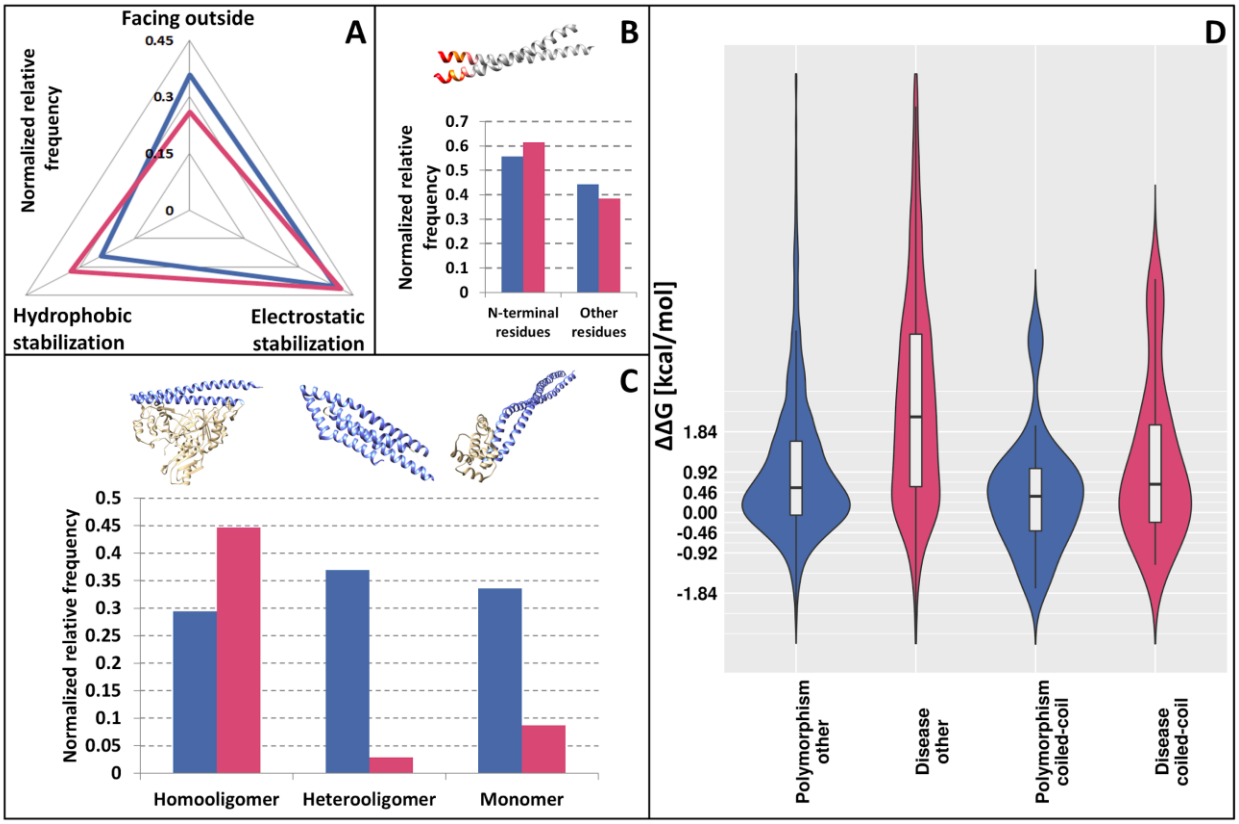


Figure 4: Distribution of variation on PDB structures. *A) Distribution of variations based on the register position types. B) Distribution of variations in the N-terminal seven residues and in other segments of coiled-coils. C) Distribution of variations according to the oligomerization state of structures. D) Energy change distributions in coiled-coils (right) and in other proteins from the human proteome (left). Red: DMs; blue: PMs.*

Further context can be added by the joint analysis of structural data. Figure 5 shows how DMs are distributed according to their features. Heptad positions with the highest standard deviation corresponding to energetic changes (positions ‘b’, ‘d’, ‘f’) exhibit the most heterogeneous distribution of different types of coiled-coils. Mutations in positions contributing to the hydrophobic core of coiled-coils (‘a’ and ‘d’), as well as the most outward facing residue type (‘f’) available for interactions operate with the most destabilizing energy changes: mutations here likely have more critical effect compared to other positions where there is more (spatial) room for substitutions. Mutations on ‘a’ position more likely affect two-stranded antiparallel coiled-coils (77%; 10 mutations affect two-stranded antiparallel coiled-coils out of total 13 mutations on position ‘a’), which is interestingly not true for the other hydrophobic residue promoting position ‘d’ (22%), also confirmed by the more comprehensive prediction data. Negative (stabilizing) energetic change is somewhat more likely occurring in homooligomeric coiled-coils (80%), with most cases on position ‘e’. We mapped the variations to only one chain of PDB structures, thus the real energetic contribution of a mutation may be even more stabilizing in homooligomers, abolishing transient interactions. In contrast, most mutations affecting monomeric coiled-coils are definitely highly destabilizing (94%), suggesting greater energetic effect is required to disrupt the overall structure that also includes intrachain interactions outside the coiled-coil, in

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

contrast to mutations of complexes where coiled-coil interchain interactions are the only forces keeping the complex together. Last, but not least, destabilizing mutations more frequently occur in the possible N-terminal regions (57%), compared to stabilizing mutation (33%).

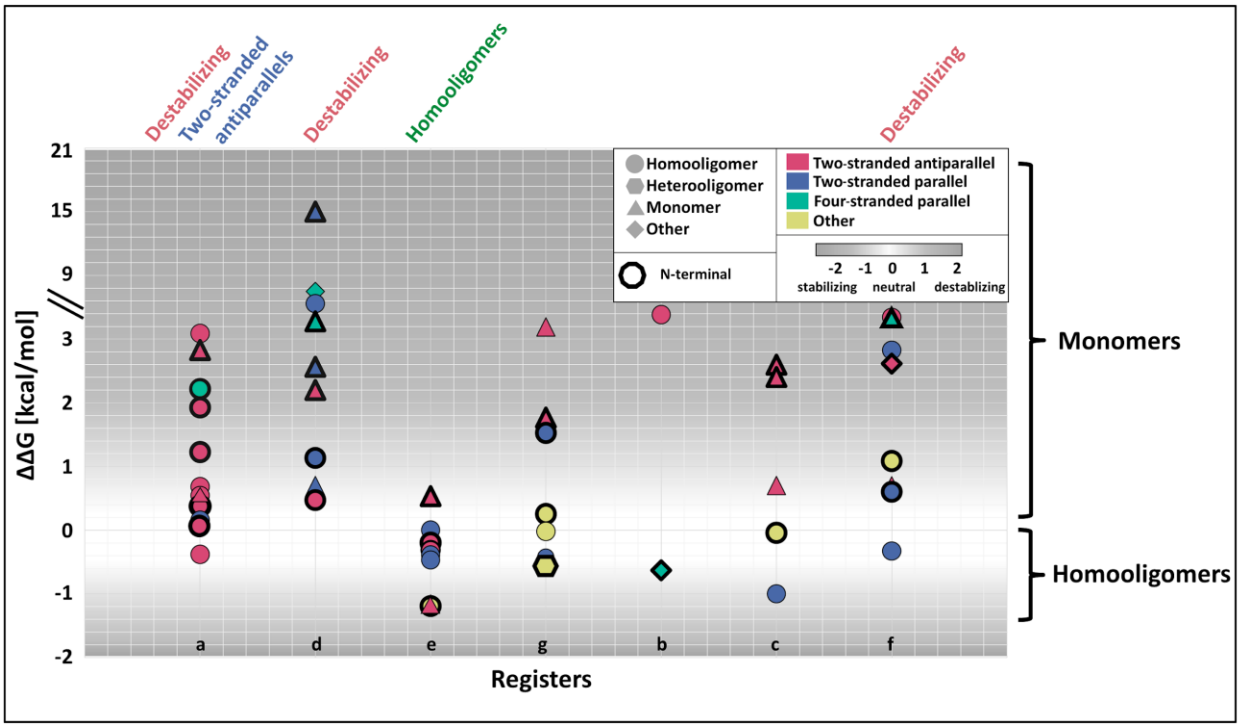


Figure 5: Distribution of DMs with respect to various structural features. *Y-axis shows the calculated energy change caused by mutations. X-axis shows the register position DMs fall into. Colors and shapes highlight further properties (see legend).*

Impaired coiled-coils are mostly associated with central nervous system diseases

To get a better picture of what types of diseases emerge from the analyzed variations, we used DiseaseOntology to group diseases into categories based on their MIM identifier. Such ontologies often have some level of annotation bias, as the included proteins are not necessarily annotated to the same extent. To avoid drawing false conclusions we did not calculate p-values, and only counted the number of occurrences (and also highlighted the expected number of occurrences for comparison). Table 1 shows the most commonly occurring disease classes, where a DM directly disrupts the structure of a coiled-coil. According to our analysis, the most enriched disease terms are skin diseases, muscular diseases, carbohydrate metabolic diseases, and central nervous system diseases. Coiled-coil DMs are also enriched in more general terms, such as integumentary system diseases, musculoskeletal system diseases, and nervous system diseases. There is also a subtle abundance of conditions related to metabolic diseases that can be acquired according to the annotation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: DiseaseOntology term analysis. *Only terms responsible for at least 5% of all annotated diseases are shown in the top 3 levels of the ontology. N is the number of mutations disrupting coiled-coil structures (in parenthesis expected values are shown).*

Level 1	N	Level 2	N	Level 3	N
disease of anatomical entity	734 (778)	integumentary system disease	72 (21)	skin disease	71 (21)
		musculoskeletal system disease	165 (120)	muscular disease	161 (55)
		nervous system disease	397 (377)	central nervous system disease	294 (156)
				sensory system disease	74 (115)
		cardiovascular system disease	60 (85)	heart disease	60 (83)
syndrome	52 (43)				
genetic disease	56 (160)				
disease of metabolism	103 (468)	acquired metabolic disease	63 (41)	carbohydrate metabolism disease	63 (51)

Discussion

Several studies investigated the structural consequences of inherited disease-causing mutations[16], one of them also including coiled-coils[14]. In this manuscript, we used four different predictors to assess the structural consequences of variations in coiled-coils, then extended our analysis by incorporating structural data and features responsible for the proper assembly of coiled-coil complexes. We found that DMs accumulate in heptad positions critical for the assembly of coiled-coils (in line with the findings published by Mohanasundaram *et al.*[14]), N-terminal parts of coiled-coils are more abundant in DMs, and mutations mostly strongy affect homooligomeric coiled-coils.

Sequence properties are often used to characterize substitutions, as they often can be connected to structural changes. In this case, grouping amino acids based on their possible role in coiled-coil formation highlights the critical role of charged residues. While mutations on hydrophobic residues impair coiled-coil structures to the same extent as in the case of globular proteins, charge changes often perturb coiled-coil formation. Notably, not only the change of net charge influences coiled-coils, but residues bearing negative charges also do not seem to be interchangeable. This effect is attributable to the helix formation tendency of glutamic acid[17] that was also proposed in the case of single-α helices[18]. The most characteristic feature of coiled-coils is their repeated register position pattern. Steric clashes and loss of hydrophobic interactions dominate in ‘a’ and ‘d’ positions, while the loss of electrostatic interactions mostly occurs in ‘e’ and ‘g’ positions. Outward-facing residues can also carry essential roles

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sometimes: they can serve as outside staples that stabilize the alpha-helix by electrostatic interactions, or they can provide a binding site for other molecules (Figure 6, top). For example, the ubiquitin-binding domain (UBAN), conserved in optineurin (OPTN) is part of a coiled-coil, specifically recognizing ubiquitin chains binding to the accessible surface of the coiled-coil[19]. The nuclear factor-κB (NF-κB) pathway plays an important role in regulating inflammation, adaptive and innate immune responses, and cell death via transcriptional targets, such as IL-1β[20]. In the canonical pathway, NF-κB factors are retained in an inactive state via binding to OPTN[19]. The E478G mutation in the UBAN of OPTN abolishes its NF-κB suppressive activity[21], as residues involved in linear ubiquitin-binding correspond to the residues crucial for keeping NF-κB inactive[22]. The mutations result in significant up-regulation of IL-1β, causing neuroinflammation and neuronal cell death of motor neurons, leading to Amyotrophic Lateral Sclerosis[23].

From a functional point of view, mutations falling into distinct structural categories may have different effects. DMs harboring residues contributing to the hydrophobic core usually have an indirect consequence (Figure 6, bottom). In the first scenario, the effect of the mutation manifests outside the coiled-coil region. Desmins are large scaffolding proteins connecting the Sarcolemma, Z-discs, and the nucleus[24]. They consist of elongated coiled-coil regions, with a head and tail unit at their termini. Mutations in the coiled-coil regions disrupt the coiled-coil structure (e.g, DESM: L345P), eventually leading to the disorganization of Z discs and affecting the integrity of the cellular IF network[25]. Mutations often impair coiled-coils directly, so they lose (some of) their binding affinity to molecules interacting with them. The H486R mutation in OPTN perturbs the structure of the UBAN domain and causes low-grade inflammation that leads to glaucoma[26]. However, in contrast to other mutants that were shown to have a direct role in interacting with ubiquitin, this residue points inside the coiled-coil, and only reduces the binding affinity[27]. In the third scenario, mutations are occurring in the coiled-coil, however on a residue facing outward. An example of disruption of direct binding is the mutation affecting interaction of PIK3CA-PIK3R2-glycerol complex. PIK3R2 possesses a two-stranded coiled-coil and forms a heterodimer regulatory unit with PIK3CA via H-bond between N345 of PIK3CA and D557 of PIK3R2. Their complex structure also preserves a groove, providing a room for binding glycerol[28], which is perturbed by the D557H mutation. The lost direct contact of Asp sidechain with the glycerol, as well as the lack of negative charge positioning the molecule (which is abolished with the positively charged and larger histidine) were proposed to impact binding negatively[29]. PIK3R2 was associated with Megalencephaly-Polymicrogyria-Polydactyly-Hydrocephalus[30], although the molecular details of the disease were not revealed yet. Besides perturbing structural stability and folding leading to toxic conformations, mutations may also modulate degradation or lead to improper trafficking [31]. For example, assembly of the Non-POU domain-containing octamer-binding protein is mediated via antiparallel coiled-coil domains and single-α helices[32,33]. The R293H mutation in the coiled-coil domain was shown to lead to subnuclear mislocalization and resulting in endocrine-related tumors[34].

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

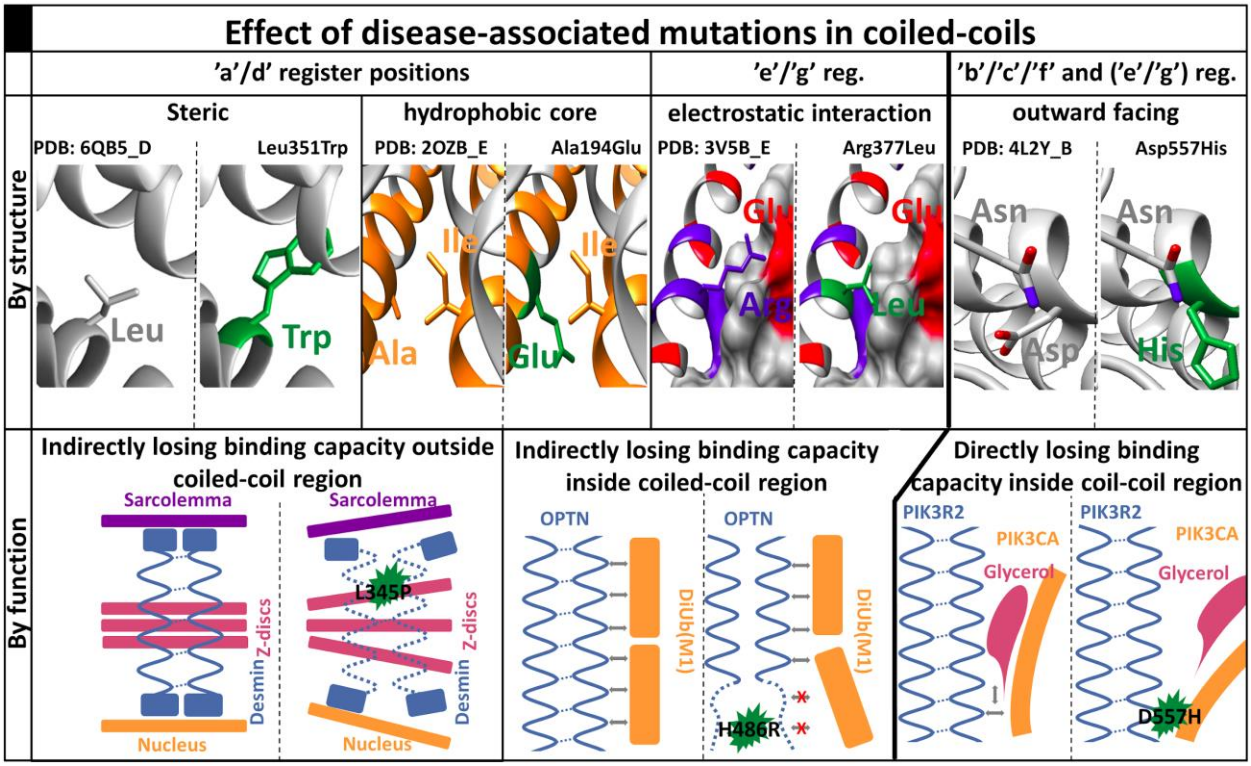


Figure 6: Flavors of disease-causing mutations in coiled-coils. *Top row: the structural effect of variations (the 'Rotamers' function of Chimera were used to visualize the mutation). Bottom row: functional consequences of mutations. Left side: structure/function in healthy people; Right side: perturbed structures/function having pathological condition. Green color indicates disease-causing mutation; On the structure, orange: hydrophobic; blue: positive; red: negative.*

The different types of coiled-coils utilizing different strategies to achieve a folded state. Many studies already suggested that highly conserved sequence patterns (so-called trigger sequences) are responsible for initiating coiled-coil assembly: for example a seven residue highly conserved motif is required for the folding of the Human Macrophage Scavenger Receptor oligomerization domain[35], and germline mutation in this region is associated with prostate cancer risk[36]. Another way to initialize assembly occurs during co-translation, as in the case of Peripherin including two-stranded parallel coiled-coils[37], which also accomodate a disease-causing mutation at the N-terminal region in one of it's coiled-coil regions[38]. A possible interpretation for the N-terminal accumulation of DMs might lie in the co-translational initiation of the folding and stabilization of α-helices as they emerge from the ribosome[39]. Although abolishing the process likely affects superhelix assembly, this phenomenon only serves as an explanation for mutations in parallel coiled-coils. The critical role of terminal regions are also well-marked in antiparallel coiled-coils: SMC1 forms a complex with SMC3 via their globular N- and C-terminal domains. In both proteins the head and tail regions are connected by antiparallel coiled-coils, and most of the identified DMs gather at their beginning/end of the coiled-coil domains[40]. The proposed antiparallel intramolecular coiled-coil of KIF21A gathers several DMs, predominantly occupying the termini of the coiled-coil[41], responsible for congenital fibrosis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Thus, although the exact molecular background was not revealed yet, there is a number of evidence supporting the critical role of certain segments in coiled-coils (trigger sites or terminal regions), with an underlied role of N-terminal residues.

Oligomer coiled coils are formed via mutually synergistic folding (MSF), which means interacting partners are disordered in monomer form and the folding and the binding process cannot be separated[42]. The energetic spectrum of such complexes is heterogeneous, including strongly bound structures and weak associations too. Coiled-coil proteins form a particular subclass having relatively weak stabilizing energies, where the interaction plays an essential role in forming a stable complex[43]. It is expected that these proteins are often bound by transient interactions. This is somewhat supported by the fact that both neutral and harmful mutations cause a lower average energetic change compared to other proteins of the proteome. Notably, some of these changes may have a double effect not represented in our calculations, as a significant proportion of mutations affect homooligomers, where both subunits may be exposed to the same mutation.

Our results suggest that mutations in coiled-coils are responsible for several muscular, skin, and neurodegenerative diseases. The role of postsynaptic density (PSD) proteins in neural degeneration is often revisited in various papers[44,45]. In our recent article, we showed that postsynaptic density proteins are exceptionally modular, and they are highly enriched in coiled-coil regions[46]. These results further underline the critical role of coiled-coil containing proteins in the PSD, and advocates the importance of modular protein structures in disease development.

A handful of popular methods are available to predict the effect of variations[47,48] or to highlight vulnerable regions in proteins[49,50], yet most of these are based on purely statistical approaches. Methods incorporating structural information are largely limited to general features of PDB structures, or prediction of transmembrane domains or disordered segments, although no currently available methods incorporate features of coiled-coils. We showed that basic properties of coiled-coils, such as register position, oligomerization state and position along the region significantly influence the formation of coiled-coils. Since coiled-coil region prediction typically has short run times, we suggest that including such data into state-of-the-art predictors to increase their accuracy would be feasible.

Methods

Datasets

The human proteome was downloaded from UniProt[51], germline variations were obtained from humsavar[51] (Supplementary Table 1). For redundancy filtering CD-HIT[52] was applied on the human proteome in an incremental manner, filtering identical proteins to 90, 70, 50 and finally to 40% identity using 5,4,3 and 2 word lengths, respectively (Supplementary Table 2).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Coiled-coil predictions

Coiled-coil regions were determined using DeepCoil[53], MarCoil[54], Ncoils[55] and Paircoil[56] (Supplementary Table 3), applying default cutoff values suggested in their descriptive articles. In the case of DeepCoil we utilized the 'PSSM' flavor: we generated PSSM for each sequence, using PSI-BLAST with three iterations and 10^{-5} e-value cutoff on the SwissProt database. Coiled-coil positions were predicted using MarCoil, Ncoils and Paircoil (Supplementary Table 4). Oligomerization states were defined using LogiCoil (Supplementary Table 4). Single-α Helix regions[57] were used as a filter, to reduce false positive hits (Supplementary Table 5).

Statistical tests

χ^2 tests were performed in contingency tables (Table 2). Odds ratios were defined as:

OR= $(x_1/x_2)/(x_4/x_5)$

Enrichments on Supplementary figure 1 were defined as:

Enrichment (DMs) = $(x_4/(x_4+x_1))/(x_6/(x_6+x_3))$
Enrichment (PMs) = $(x_5/(x_5+x_2))/(x_6/(x_6+x_3))$

χ^2 was applied to find the significance of the relation between DMs and coiled-coils (Supplementary Table 6) and the significant importance of the first seven residues of coiled-coil sequences (Supplementary Table 7).

Kolmogorov-Smirnov tests were used to estimate the significance of the distribution of mutations at the first 28 residues of coiled-coils (Supplementary Table 8) and along the coiled-coil sequences (Supplementary Table 9).
χ^2 test was performed to find the significance of distribution of DMs into coiled-coils with different lengths (Supplementary Table 10)

To estimate the significance of residue changes, we eliminated the sporadic error of the data by performing bootstrap analysis. We randomly selected 80% of the data 100 times and the significance was determined by calculating the average and standard deviations of the data according to the 68-95-99.7 rule (Supplementary Table 11-12).

χ^2 was used to find the significance of the distribution of DMs into different coiled-coils positions (Supplementary Table 13)

All tests and analysis were performed to the different predictors separately. To produce figures, in each case, we calculated the mean of different predictors (Supplementary Table 6-13). Although we plotted mean values, if predictors showed different trends, we noted it in the text.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DiseaseOntology term analysis

Disease ontology terms were mapped using MIM identifiers from humsavar and DiseaseOntology[58]. Only identifiers linked to DMs, where all methods predicted coiled-coil were used. For the analysis the top three level of the ontology was applied, and the number of mutations were counted in each disease category - only terms occurring in coiled-coil containing proteins are shown in Supplementary Table 14, and only terms responsible for at least 5% of all annotated diseases shown in Table 1. Next we mapped all mutations in a similar manner. Expected values were calculated by normalizing these numbers on each term with the proportion of all coiled-coil mutations.

Assigning structures to amino acid sequences

We used BLAST on sequences from the non-redundant human proteome against the PDB with 10^{-5} e-value. Chimeric proteins were discarded. We used the greedy algorithm to select structures with 100% identity, with the most variations mapped on them (Supplementary Table 15). On all PDB structures we considered biomatrix transformations as defined in the PDB files to detect all possible coiled-coils.

Calculating structural and energetic properties

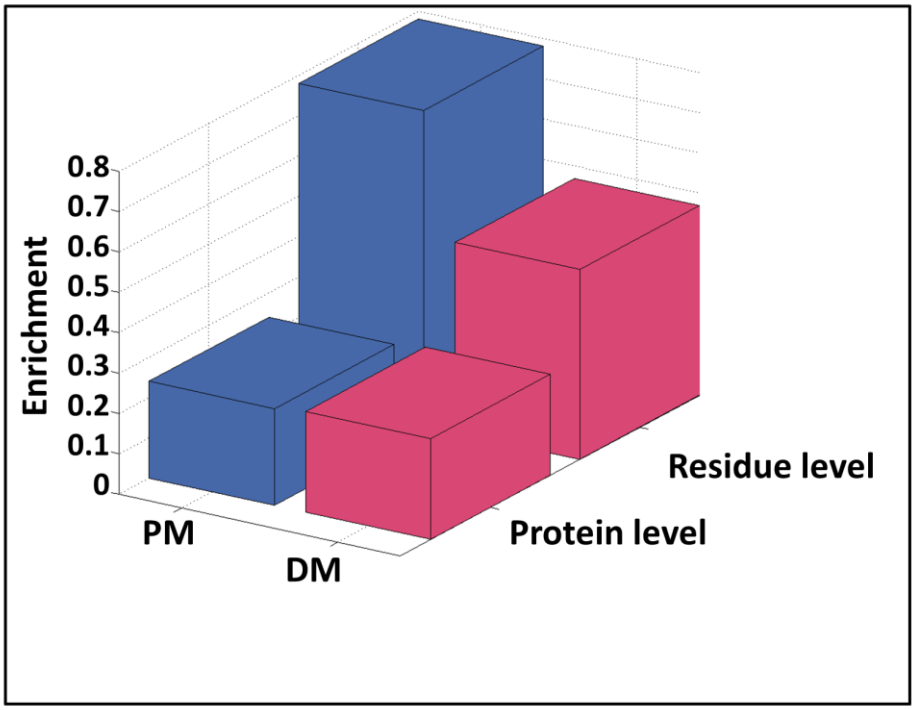
We detected coiled-coils using SOCKET[59] with default settings. Coiled-coil features (heptad positions, the number of strands, angle of strands (Supplementary Table 15-16)) were determined based on SOCKET output. For monomer/homooligomer/heterooligomer assignment, we checked which BLAST query corresponds to the detected coiled-coil regions (Supplementary Table 18). Energy calculations were performed using FoldX[60]. $\Delta\Delta G$ calculations were executed on previously optimized structures and were performed five times. All reported $\Delta\Delta G$ values represent the average of these independent runs. In 76 cases (less than 1%) we experienced problems with FoldX, these cases were omitted (Supplementary Table 19). Calculated structural features shown on Figure 4 are based on values from Supplementary Table 20. Energetic changes on figure are categorized as highly stabilizing (<-1.84 kcal/mol), stabilizing (-1.84 to -0.92 kcal/mol), slightly stabilizing (-0.92 to -0.46 kcal/mol), neutral (-0.46 to $+0.46$ kcal/mol), slightly destabilizing ($+0.46$ to $+0.92$ kcal/mol), destabilizing ($+0.92$ to $+1.84$ kcal/mol) and highly destabilizing ($> +1.84$ kcal/mol).

Visualization

Images were prepared using UCSF Chimera[61].

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figures

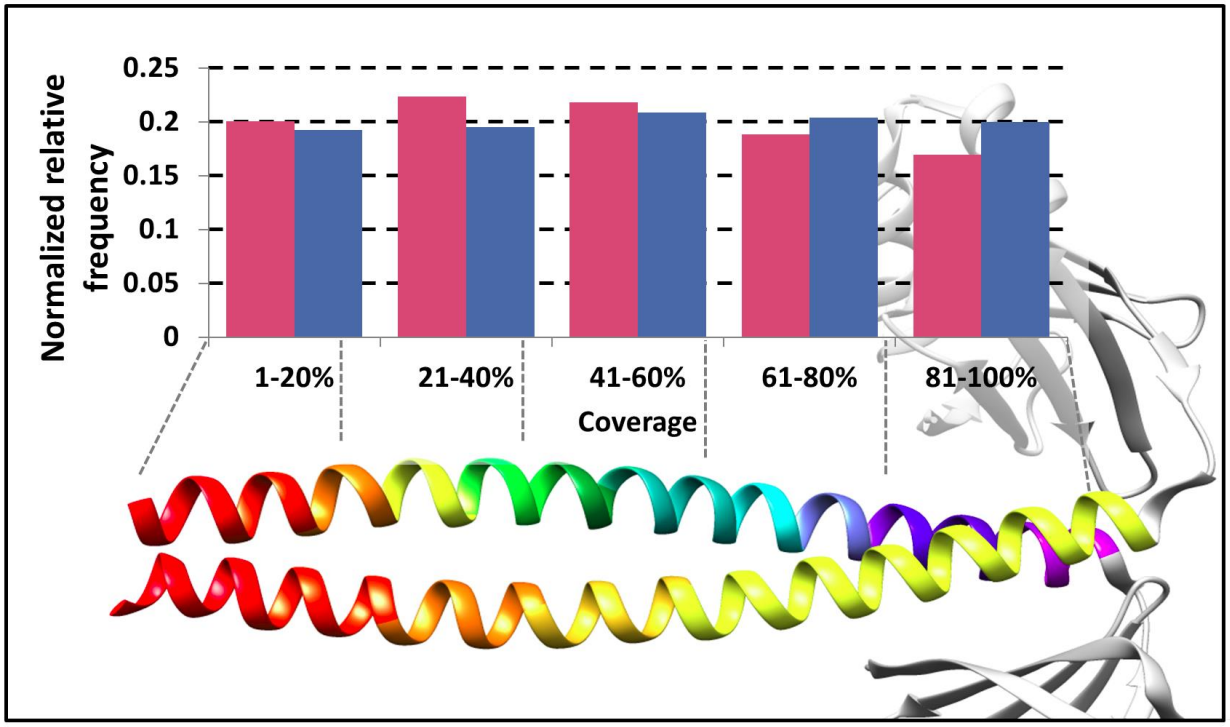


SFigure 1: Relation between DMs and coiled-coils. *A) Enrichment of PMs (blue) and DMs (red) in proteins and in residues.*

from\to	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
A	0	0.729	0.053	-0.16	0.428	0.197	0	0	0.428	0	0	0	0.093	0	0	0.012	0.129	0.135	0	0	2.05
C	0	0	0	0	-0.7	-0.44	0	0	0	0	0	0	0	0	-0.18	-0.69	0	0	-0.38	-0.48	-2.87
D	0.127	0	0	0.187	0	0.114	-0.05	0	0	0	0	-0.08	0	0	0	0	0	-0.1	0	0.032	0.23
E	0.461	0	0.335	0	0	0.342	0	0	0.278	0	0.906	0	0	0.305	0	0	0	0.148	0	0	2.77
F	0	-0.06	0	0	0	0	0	0.03	0	-0.23	0	0	0	0	0	-0.14	0	-0.25	0	0	-0.66
G	-0.39	-0.74	-0.4	-0.48	0	0	0	0	0.428	0	0	0	0	0	-0.35	-0.34	0	-0.45	-0.69	0	-3.42
H	0	0	-0.01	0	0	0	0	0	0	-0.02	0	-0.21	-0.15	-0.09	-0.13	0	0	0	0	-0.22	-0.84
I	0	0	0	0	-0.04	0	0	0	-0.3	0.108	-0.14	-0.04	0.428	0	-0.01	0.076	0.067	0.252	0	0	0.41
K	0	0	0	0.266	0	0	0	-0.42	0	0	0.349	0.17	0	0.335	0.252	0	0.393	0	0	0	1.35
L	0	0	0	0	0.08	0	0.316	0.185	0	0	0.276	0	0.166	0.301	0.118	0.242	0	0.161	0.09	0	1.94
M	0	0	0	0	0	0	0	-0.06	0.01	0	0	0.729	0	0	0.162	0	0.104	0.058	0	0	1
N	0	0	0.112	0	0	0	0.165	0.012	0.105	0	0	0	0	0	0	0.032	0.163	0	0	0.207	0.8
P	-1.29	0	0	0	0	0	0	0	0	-1.39	0	0	0	0	-1.05	-0.88	0	0	0	0	-4.61
Q	0	0	0	0.326	0	0.729	0.192	0	0.268	0.428	0	0	0.327	0	0.106	0	0	0	0	0	2.38
R	0	0.117	0	0	0	0.11	0.09	0.538	0.252	0.09	-0.24	0	0.223	0.159	0	0.235	0.09	0	0.186	0	1.85
S	0.048	-0.42	0	0	-0.1	-0.31	0	-0.12	0	-0.07	0	-0.38	-0.03	0	-0.02	0	0.176	0	-0.29	0.021	-1.49
T	0.002	0	0	0	0	0	0	-0.15	0.173	0	-0.17	-0.15	-0.14	0	-0.12	0.298	0	0	0	0	-0.24
V	-0.07	0	-0.1	-0.17	0.038	-0.06	0	-0.15	0	0.06	-0.18	0	0	0	0	0	0	0	0	0	-0.64
W	0	-0.38	0	0	0	-0.14	0	0	0	-0.01	0	0	0	0	-0.45	-0.27	0	0	0	0	-1.25
Y	0	-0.16	-0.16	0	-0.05	0	-0.35	0	0	0	0	-0.27	0	0	0	-0.08	0	0	0	0	-1.07

SFigure 2: Amino acid changes in coiled-coils. *Left) Residue change preferences by DMs in the proteome (negative values, also marked with the shades of blue) and in coiled-coil regions (positive values, also marked with the shades of red). Values show the logarithm of ratio of DMs changing given residues types Right) Targeted residue type preferences by DMs in the proteome (negative values) and in coiled-coil regions (positive values).*

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



SFigure 3: Variations along coiled-coil segments. *Distribution of variations in the sequence. X-axis shows the coverage of the sequences.*

Funding

Zs. K. has been supported by the European Union and cofinanced by the European Social Fund (EFOP- 3.6.2- 16- 2017- 00013, Thematic Fundamental Research Collaborations Grounding Innovation in Informatics and Infocommunications). Z.G. acknowledges support from the National Research, Development and Innovation Office through grant OTKA 124363.

Conflicts of interest

None declared.

References

1. Dong, G., Medkova, M., Novick, P. & Reinisch, K. M. A catalytic coiled coil: structural insights into the activation of the Rab GTPase Sec4p by Sec2p. *Mol. Cell* **25**, 455–462 (2007).
2. Truebestein, L. & Leonard, T. A. Coiled-coils: The long and short of it. *BioEssays* vol. 38 903–916 (2016).
3. Hayashi, M. *et al.* The postsynaptic density proteins homer and shank form a polymeric network structure. *Neuroscience Research* vol. 68 e339 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4. Crick, F. H. C. The packing of α-helices: simple coiled-coils. *Acta Crystallographica* vol. 6 689–697 (1953).
5. Burkhard, P., Stetefeld, J. & Strelkov, S. V. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88 (2001).
6. Moutevelis, E. & Woolfson, D. N. A periodic table of coiled-coil protein structures. *J. Mol. Biol.* **385**, 726–732 (2009).
7. Lupas, A. N., Bassler, J. & Dunin-Horkawicz, S. The Structure and Topology of α-Helical Coiled Coils. *Subcell. Biochem.* **82**, 95–129 (2017).
8. Mason, J. M. & Arndt, K. M. Coiled coil domains: stability, specificity, and biological implications. *Chembiochem* **5**, 170–176 (2004).
9. Hicks, M. R., Holberton, D. V., Kowalczyk, C. & Woolfson, D. N. Coiled-coil assembly by peptides with non-heptad sequence motifs. *Fold. Des.* **2**, 149–158 (1997).
10. Kammerer, R. A. *et al.* An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13419–13424 (1998).
11. Ng, D. P., Poulsen, B. E. & Deber, C. M. Membrane protein misassembly in disease. *Biochim. Biophys. Acta* **1818**, 1115–1122 (2012).
12. Pajkos, M., Mészáros, B., Simon, I. & Dosztányi, Z. Is there a biological cost of protein disorder? Analysis of cancer-associated mutations. *Mol. BioSyst.* vol. 8 296–307 (2012).
13. Gao, M., Zhou, H. & Skolnick, J. Insights into Disease-Associated Mutations in the Human Proteome through Protein Structural Analysis. *Structure* **23**, 1362–1369 (2015).
14. Mohanasundaram, K. A., Grover, M. P., Crowley, T. M., Goscinski, A. & Wouters, M. A. Mapping genotype-phenotype associations of nsSNPs in coiled-coil oligomerization domains of the human proteome. *Hum. Mutat.* **38**, 1378–1393 (2017).
15. Woolfson, D. N. Coiled-Coil Design: Updated and Upgraded. *Subcell. Biochem.* **82**, 35–61 (2017).
16. Dobson, L., Mészáros, B. & Tusnády, G. E. Structural Principles Governing Disease-Causing Germline Mutations. *J. Mol. Biol.* **430**, 4955–4970 (2018).
17. Chou, P. Y. & Fasman, G. D. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**, 45–148 (1978).
18. Süveges, D., Gáspári, Z., Tóth, G. & Nyitray, L. Charged single alpha-helix: a versatile protein structural motif. *Proteins* **74**, 905–916 (2009).
19. Rahighi, S. *et al.* Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. *Cell* **136**, 1098–1109 (2009).
20. Cogswell, J. P. *et al.* NF-kappa B regulates IL-1 beta transcription through a consensus NF-kappa B binding site and a nonconsensus CRE-like site. *J. Immunol.* **153**, 712–723 (1994).
21. Li, F. *et al.* Structural insights into the interaction and disease mechanism of neurodegenerative disease-associated optineurin and TBK1 proteins. *Nat. Commun.* **7**, 12708 (2016).
22. Nakazawa, S. *et al.* Linear ubiquitination is involved in the pathogenesis of optineurin-associated amyotrophic lateral sclerosis. *Nat. Commun.* **7**, 12547 (2016).
23. Liu, Z. *et al.* ALS-Associated E478G Mutation in Human OPTN (Optineurin) Promotes Inflammation and Induces Neuronal Cell Death. *Front. Immunol.* **9**, 2647 (2018).
24. Hnia, K., Ramspacher, C., Vermot, J. & Laporte, J. Desmin in muscle and associated diseases: beyond the structural function. *Cell Tissue Res.* **360**, 591–608 (2015).
25. Goldfarb, L. G. & Dalakas, M. C. Tragedy in a heartbeat: malfunctioning desmin causes skeletal and cardiac muscle disease. *J. Clin. Invest.* **119**, 1806–1813 (2009).
26. Toth, R. P. & Atkin, J. D. Dysfunction of Optineurin in Amyotrophic Lateral Sclerosis and Glaucoma. *Front. Immunol.* **9**, 1017 (2018).
27. Li, F. *et al.* Structural insights into the ubiquitin recognition by OPTN (optineurin) and its regulation by TBK1-mediated phosphorylation. *Autophagy* **14**, 66–79 (2018).
28. Zhao, Y. *et al.* Crystal Structures of PI3Kα Complexed with PI103 and Its Derivatives: New

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Directions for Inhibitors Design. *ACS Med. Chem. Lett.* **5**, 138–142 (2014).
29. Terrone, G. *et al.* De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia. *Eur. J. Hum. Genet.* **24**, 1359–1362 (2016).
30. Mirzaa, G. M. *et al.* Characterisation of mutations of the phosphoinositide-3-kinase regulatory subunit, PIK3R2, in perisylvian polymicrogyria: a next-generation sequencing study. *Lancet Neurol.* **14**, 1182–1195 (2015).
31. Thomas, P. J., Qu, B.-H. & Pedersen, P. L. Defective protein folding as a basis of human disease. *Trends in Biochemical Sciences* vol. 20 456–459 (1995).
32. Passon, D. M. *et al.* Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4846–4850 (2012).
33. Dobson, L., Nyitray, L. & Gáspári, Z. A conserved charged single α-helix with a putative steric role in paraspeckle formation. *RNA* **21**, 2023–2029 (2015).
34. Kharade, S. S., Parekh, V. I. & Agarwal, S. K. Functional Defects From Endocrine Disease–Associated Mutations in HLXB9 and Its Interacting Partner, NONO. *Endocrinology* vol. 159 1199–1212 (2018).
35. Frank, S., Lustig, A., Schulthess, T., Engel, J. & Kammerer, R. A. A distinct seven-residue trigger sequence is indispensable for proper coiled-coil formation of the human macrophage scavenger receptor oligomerization domain. *J. Biol. Chem.* **275**, 11672–11677 (2000).
36. Xu, J. *et al.* Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. *Nat. Genet.* **32**, 321–325 (2002).
37. Chang, L., Shav-Tal, Y., Trcek, T., Singer, R. H. & Goldman, R. D. Assembling an intermediate filament network by dynamic cotranslation. *J. Cell Biol.* **172**, 747–758 (2006).
38. Kohl, S. *et al.* RDS/peripherin gene mutations are frequent causes of central retinal dystrophies. *J. Med. Genet.* **34**, 620–626 (1997).
39. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16**, 589–597 (2009).
40. Deardorff, M. A. *et al.* Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am. J. Hum. Genet.* **80**, 485–494 (2007).
41. Bianchi, S. *et al.* Structural basis for misregulation of kinesin KIF21A autoinhibition by CFEOM1 disease mutations. *Sci. Rep.* **6**, 30668 (2016).
42. Mészáros, B. *et al.* Sequential, Structural and Functional Properties of Protein Complexes Are Defined by How Folding and Binding Intertwine. *J. Mol. Biol.* **431**, 4408–4428 (2019).
43. Mészáros, B., Dobson, L., Fichó, E. & Simon, I. Sequence and Structure Properties Uncover the Natural Classification of Protein Complexes Formed by Intrinsically Disordered Proteins via Mutual Synergistic Folding. *Int. J. Mol. Sci.* **20**, (2019).
44. Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21 (2011).
45. Roy, M. *et al.* Proteomic analysis of postsynaptic proteins in regions of the human neocortex. *Nat. Neurosci.* **21**, 130–138 (2018).
46. Kiss-Tóth, A. *et al.* Occurrence of Ordered and Disordered Structural Elements in Postsynaptic Proteins Supports Optimization for Interaction Diversity. *Entropy* vol. 21 761 (2019).
47. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit7.20 (2013).
48. Kulandaisamy, A., Zaucha, J., Sakthivel, R., Frishman, D. & Michael Gromiha, M. Pred-MutHTP: Prediction of disease- causing and neutral mutations in human transmembrane

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.07.029165; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proteins. *Human Mutation* vol. 41 581–590 (2020).
49. Mészáros, B., Zeke, A., Reményi, A., Simon, I. & Dosztányi, Z. Systematic analysis of somatic mutations driving cancer: uncovering functional protein regions in disease development. *Biol. Direct* **11**, 23 (2016).
50. Niu, B. *et al.* Protein-structure-guided discovery of functional mutations across 19 cancer types. *Nature Genetics* vol. 48 827–837 (2016).
51. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
52. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* vol. 26 680–682 (2010).
53. Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V. & Dunin-Horkawicz, S. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **35**, 2790–2795 (2019).
54. Delorenzi, M. & Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625 (2002).
55. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
56. McDonnell, A. V., Jiang, T., Keating, A. E. & Berger, B. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* **22**, 356–358 (2006).
57. Dudola, D., Tóth, G., Nyitray, L. & Gáspári, Z. Consensus Prediction of Charged Single Alpha-Helices with CSAHserver. *Methods Mol. Biol.* **1484**, 25–34 (2017).
58. Schriml, L. M. *et al.* Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* **40**, D940–6 (2012).
59. Hillar, A. *et al.* Detection of α-Helical Coiled-Coil Dimer Formation by Spin-Labeled Synthetic Peptides: A Model Parallel Coiled-Coil Peptide and the Antiparallel Coiled Coil Formed by a Replica of the ProP C-Terminus†. *Biochemistry* vol. 42 15170–15178 (2003).
60. Delgado, J., Radusky, L. G., Cianferoni, D. & Serrano, L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**, 4168–4169 (2019).
61. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).