Structural bioinformatics

A computational methodology to diagnose sequence-variant dynamic perturbations by comparing atomic protein structures

Lorenza Pacini [1,2] and Claire Lesieur [1,2,*]

[1]AMPERE, CNRS, Université de Lyon, 69622, Lyon, France
[2]Institut Rhônalpin des systèmes complexes (IXXI), École Normale Supérieure de Lyon, Lyon 69007, France

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The objective is to diagnose dynamics perturbations caused by amino acid mutations as prerequisite to assess protein functional health or drug failure, simply using network models of protein X-ray structures.
Results: We find that the differences in the allocation of the atomic interactions of each amino acid to 1D, 2D, 3D, 4D structural levels between variants structurally robust, recover experimental dynamic perturbations. The allocation measure validated on two B-pentamers variants of AB_5 toxins having 17 mutations, also distinguishes dynamic perturbations of pathogenic and non-pathogenic Transthyretin single-mutants. Finally, the main proteases of the coronaviruses SARS-CoV and SARS-CoV-2 exhibit changes in the allocation measure, raising the possibility of drug failure despite the main proteases structural similarity.
Availability: The Python code used for the production of the results is available at `github.com/lorpac/protein_partitioning_atomic_contacts`. The authors will run the analysis on any PDB structures of protein variants upon request.
Contact: `claire.lesieur@ens-lyon.fr`
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Proteins are constituted by amino acids chemically interacting via their atoms in a three-dimensional structure. The biological function of a protein relies on the protein dynamics, i.e. the motions of the atoms that constitute the protein structure (Henzler-Wildman and Kern, 2007). Thus, amino acid mutations that impact the protein structure and/or dynamics can impact the protein function (Thusberg and Vihinen, 2009). Computational and experimental approaches are available to investigate the impact of mutations on structure and dynamics (Achoch *et al.*, 2016; McLaughlin Jr *et al.*, 2012; Naganathan, 2019; Vuillon and Lesieur, 2015; Haliloglu and Bahar, 2015; Degiacomi *et al.*, 2013; Ramaswamy *et al.*, 2021; Leitner and Yamato, 2018; Zhang *et al.*, 2021; Kamada *et al.*, 2011). Nevertheless, investigating the impact of mutations on the slow dynamics underlying function and folding is still a challenge because it involves multiple time and length scales. There are large collective motions (e.g. interface, chain, large domain) with fluctuations from millisecond to second and small collective motions (e.g. 3D/2D structures) with fluctuations from microsecond to milliseconds (Henzler-Wildman and Kern, 2007; Munoz and Cerminara, 2016). Such timescales are difficult to simulate at high resolution using molecular dynamics (Lesieur and Schulten, 2015; Perilla *et al.*, 2015). New ultra-fast spectroscopies combined with computational approaches are efficient but remain complicated techniques (Herbst *et al.*, 2002; Kolano *et al.*, 2006; Mizutani and Kitagawa, 2001; Fang *et al.*, 2009; Schneider *et al.*, 2019; Findsen *et al.*, 1985). Another difficulty is to monitor how the dynamics propagate from the local scale (amino acid level) to larger scales in the protein in order to assess how mutations impact the dynamic propagation paths (Buchenberg *et al.*, 2017).

Atomic motions and dynamics can be tracked by measuring changes in atomic interactions (link weights) inferred from network models of

protein structures combined with MD simulations (Gheeraert *et al.*, 2019; Leitner and Yamato, 2018). More generally, dynamics of networked system is susceptible to link perturbations (Unicomb *et al.*, 2018; Battiston *et al.*, 2012). Experimentally, broadband dielectric spectroscopy (BDS) monitors slow protein dynamics with a dielectric signal sensitive from the local dipole environments to the global organization of the protein structure such that sequence mutations local perturbations to large-scale thermal unfolding perturbations can be studied, showing the relation between interaction changes and dynamics changes (Bourgeat *et al.*, 2019, 2021).

The atomic interactions made by an amino acid with its first shell neighbors depict how the atoms occupy the space locally and in turn what is the remaining empty space where local side chain motions can take place. We have shown that in a protein structure every amino acid has a similar moderate average number of atomic interactions (link weights) with its first-shell neighbors, regardless its position and type (Dorantes-Gilardi *et al.*, 2018). This means there is always empty space available for atomic motions around amino acids. Moreover, the local atomic average is conserved for variants maintaining structural integrity. Keeping the average constant while modifying the number and position of the side-chain atoms at the mutation site, necessarily modifies the way atomic interactions are allocated among the neighbors. It follows that variants with structural integrity carve the local empty space differently by using different link solutions that lead to the same average number of atomic interactions and structural integrity but may impact the protein dynamics.

We therefore made the hypothesis that allocating atomic interactions to structural levels (1D, 2D, 3D and 4D) and measuring the difference in allocation between variants will reveal areas with structural-level dynamic perturbations associated with the mutations. This hypothesis is consistent with the scale of dynamics associated with structural levels and with experimental evidences showing the impact of mutations on individual structural levels or a combination of them (Munoz and Cerminara, 2016; Degiacomi *et al.*, 2013; Gersting *et al.*, 2008). To verify the hypothesis, protein variants that maintain structural integrity but have different folding dynamics are analyzed. The structural integrity guarantees that differences in the atomic-contacts allocation are limited to dynamics perturbation and do not involve structural perturbations. If this condition was not fulfilled, there would be ambiguity on the consequences of the allocation changes. Such dynamic-centered perturbations are common in diseases related to misfolding and aggregation (Bemporad and Chiti, 2012) and are also certainly at play in drug failure since local protein dynamics impact the thermodynamics and kinetics of drugs binding (Amaral *et al.*, 2017).

We have selected case studies that are involved in human diseases and exhibit slow folding dynamics perturbations. Firstly, the B-pentamers of two AB_5 toxins, cholera toxin ($CtxB_5$) and heat labile (LTB_5) distinguished by 18 mutated amino-acids over 103 resulting in different folding/unfolding mechanisms (Dorantes-Gilardi *et al.*, 2018; Zrimi *et al.*, 2010). Secondly, single amino acid variants of Transthyretin (TTR), that have different propensity to misfolding and amyloidosis (Saelices *et al.*, 2015; Saraiva, 2001). Lastly, the allocation measure is applied to the main proteases (Mpro's) of the coronaviruses SARS-CoV and SARS-CoV-2 to diagnose dynamic perturbations that might be responsible for higher virulence or drug resistance.

2 Materials and Methods

2.1 Protein structures

The Protein Data Bank (PDB, `http://www.rcsb.org/pdb/`) codes of the protein structures used in this study are the following. $CtxB_5$: 1eei, LTB_5: 1ltr, WT TTR: 1f41, T119Y TTR: 4tne, L55P TTR: 3djz, V30M TTR: 3kgs, T119M TTR: 1bze, SARS-CoV Mpro: 2h2z, SARS-CoV-2 Mpro: 6y2e. Chains D to H have been studied for $CtxB_5$ and LTB_5, from position 1 to position 102. Chains A and B have been studied for the TTR variants, from position 10 to position 124. Chain A has been studied for the Mpro of SARS-CoV and SARS-CoV-2, from position 1 to position 305.

2.2 Sequence and structure alignment

TopMatch(Sippl and Wiederstein, 2012) has been employed to calculate the percentage of sequence identity and the Root Mean Square Deviation (RMSD) between two protein variants.

2.3 Protein structure visualization

Figures of protein structures have been produced with VMD (Humphrey *et al.*, 1996). The secondary structure information for the visualization of protein structures in the cartoon representation has been read from the secondary structure records in the corresponding PDB files though Andrew Dalke's VMD script that can be found at `www.ks.uiuc.edu/Research/vmd/script_library/scripts/vmd_use_pdb_ss`.

2.4 Amino acid network (AAN)

The Amino Acid Networks (AANs) of protein structures are created as described in (Dorantes-Gilardi *et al.*, 2018). In a nutshell, the AAN of a protein structure is an undirected network where nodes represent amino acids and weighted links exist between amino acids that are at distance lower than 5 Å, where the link weight is given by the number of atomic interactions between the two amino acids (number of atomic couples at distance lower than 5 Å).

2.5 Links classification

A (i, j) link between amino acid i and amino acid j in the AAN of a protein is classified as: 1D if i and j belong to the same chain and are first neighbors in the protein amino-acids sequence; 2D if i and j belong to the same secondary structure element (α-helix, β-sheet or coil as retrieved from the PDB file of the protein structure) and i and j are distant by less than five positions in the protein amino-acids sequence; 3D if i and j belong to different secondary structure elements or are distant by at least five positions in the amino-acids sequence; 4D if i and j belong to two different chains.

2.6 Local structural-level allocation of the atomic interactions

For each amino acid i and for each link category x, with x = 1D, 2D, 3D or 4D, we define the fraction of atomic interactions allocated by i to the category x as $f_{w_{x,i}} = w_{x,i}/w_i$, where $w_{x,i}$ is the number of atomic interactions of category x made by i (i.e. the sum of the weights $w(i, j)$ of links of category x in the AAN, for all neighbors j of node i) and w_i is the total number of atomic interactions made by i (i.e., the sum of the weights $w(i, j)$ in the AAN, for all neighbors j of node i). Using fractions of atomic interactions allocated to a structural level instead of the number of atomic interactions allocated to a structural level allows us to measure the contribution of each level compared to one another.

2.7 Protein variants comparison

The local structural-level allocation of atomic interactions in two protein variants $var1$ and $var2$ is compared through the differences $\Delta f_{w_{x,i}} = f^{var2}_{w_{x,i}} - f^{var1}_{w_{x,i}} = w^{var2}_{x,i}/w^{var2}_i - w^{var1}_{x,i}/w^{var1}_i$ with x = 1D, 2D, 3D, 4D for each amino acid i of the protein structure. Because all amino acids

in globular proteins are expected to allocate some interactions to the 1D, 2D and 3D structural levels, $f_{w_{x,i}}$ is expected to be significantly smaller than 1 for each x = 1D, 2D, 3D, 4D. Thus, $\Delta f_{w_{x,i}}$, which is also for each structural level is likewise expected to be significantly lower than 1 and $\Delta f_{w_{x,i}} = \pm 0.4$ is considered a large change in allocation of atomic contacts to the x structural level.

3 Results and Discussion

Our goal is to generate a tool capable of identifying multi-scale slow dynamics perturbations resulting from mutations using only the X-ray structures of two protein variants. To do so, we compared the allocation of atomic interactions to structural levels, in protein variants that have similar structure (RMSD $<$ 1Å) and different experimental dynamics. Precisely, the fraction of atomic interactions involved in the 1D, 2D, 3D and 4D structural levels $\left(f_{w_{1D,i}}, f_{w_{2D,i}}, f_{w_{3D,i}}, f_{w_{4D,i}}\right)$, is computed for each amino acid i of a protein variant and compared between the two variants through the measure of $\left(\Delta f_{w_{1D,i}}, \Delta f_{w_{2D,i}}, \Delta f_{w_{3D,i}}, \Delta f_{w_{4D,i}}\right)$. The atomic interactions and the atomic interaction allocations are inferred from modeling the PDB X-ray structure by a spatial network (Materials and Methods). We use three cases of study to show the match between non-zero $\Delta f_{w_{x,i}}$ regions and experimental dynamic differences.

The first case study is the comparison of the B-pentamers of two AB_5 toxins, Cholera toxin ($CtxB_5$) and human heat labile (LTB_5), which have 82% of sequence identity, structural and functional similarity but different folding/unfolding mechanisms: the formation of assembly intermediates is the rate-limiting step for $CtxB_5$ (fly-casting folding mechanism) while the folding of the monomers is the rate-limiting step for LTB_5 (induced-fit folding mechanism) (Ruddock *et al.*, 1996; Lesieur *et al.*, 2002). Since the differences in folding dynamics of $CtxB_5$ and LTB_5 are known, they represent a good case-study to test the hypothesis that differences in protein dynamics are encoded by different local structural-level allocation in the protein structure.

The second case study is Transthyretin (TTR) with non-pathological and pathological variants that are involved in neurological and cardiac genetic diseases due to amyloid fiber formation (Cendron *et al.*, 2009; Ruberg and Berk, 2012; Saraiva, 2001). The exact mechanism of amyloid formation of TTR is unknown and probably depends on the particular TTR variant and on the experimental conditions (Dasari *et al.*, 2020, 2019; Hamilton and Benson, 2001; Schmidt *et al.*, 2019; Yee *et al.*, 2019). The study of TTR variants challenges the sensitivity of the allocation measure to detect the impact of single amino acid mutations.

The third case study is the comparison of the Mpro of two coronaviruses: SARS-CoV and SARS-CoV-2. SARS-CoV is responsible of the Severe Acute Respiratory Syndrome (SARS) pandemic that caused 774 deaths in 2003 (`www.who.int/csr/sars/country/table2004_04_21/en/`, last visited on 8th June 2020) and SARS-CoV-2 is responsible of the actual COVID-19 global pandemic that has caused more than three million deaths since December 2019 (`covid19.who.int/`, last visited on 20th May 2021). The inhibition of Mpro is one of the targets for the anti-viral treatment of COVID-19 thanks to its functional importance in virus replication (Dai *et al.*, 2020; Yang *et al.*, 2003). Moreover, the structural similarity between the SARS-CoV Mpro and the SARS-CoV-2 Mpro has motivated the research for common drugs that could inhibit the Mpro activity of a broad spectrum of coronaviruses (Burley, 2020; Jin *et al.*, 2020; Yang *et al.*, 2005). We apply our methodology, validated through the first two case studies, to determine if a different dynamical behavior can be expected for the two Mpro, questioning the common drug strategy.

3.1 B-pentamers of AB_5 toxins

$CtxB_5$ and LTB_5 have 18 mutated residues over 103 (82% of sequence identity and very similar structure (RMSD = 0.59 Å, Figure 1A), but different folding/unfolding mechanisms. We have compared the local structural-level allocation of the atomic interactions in LTB_5 with respect to $CtxB_5$, for all amino acid positions, through the measure of $\Delta f_{w_{1D,i}} = f^{LTB_5}_{w_{1D,i}} - f^{CtxB_5}_{w_{1D,i}}$ (and similarly for $\Delta f_{w_{2D,i}}$, $\Delta f_{w_{3D,i}}$ and $\Delta f_{w_{4D,i}}$). The pentameric unit has been employed for the calculations, whereas the results (Figure 1) are plotted only for chain E, for simplicity. The plots for all five chains are reported in Figure S1. The results that were not consistent in at least four chains over five have been considered as potential X-ray artifacts and not taken into consideration.

Differences in $f_{w_{2D}}$,$f_{w_{3D}}$ and $f_{w_{4D}}$ are found at the N-terminal, the C-terminal and at the central α-helix of the pentamers (Figure 1C, boxes), all involved in the interfaces of the pentamers (I_1, I_2 and I_3, respectively, Figure 1A and 1B). LTB_5 allocates more atomic interactions to the 4D structural level at the I_2 and I_3 interfaces compared to $CtxB_5$ ($\Delta f_{w_{4D}} > 0$), while it allocates less atomic interactions to the 2D structural level and more to the 3D structural level at the I_1 and I_3 interfaces compared to $CtxB_5$ ($\Delta f_{w_{2D}} < 0$ and $\Delta f_{w_{3D}} > 0$). The fact that more atomic interactions are allocated to the maintenance of the interface compared to the secondary-structure fold means that the secondary structure will be more susceptible to unfolding compared to the tertiary and quaternary structure: when the interfaces are disrupted, the monomers are already unfolded. This is consistent with the experimental evidence that assembly intermediates are not detected by BDS during the thermal unfolding of LTB_5, contrary to $CtxB_5$ (Bourgeat *et al.*, 2019, 2021). The BDS results also corroborate the allocation measures because it detects dynamics perturbations in the N-terminal, C-terminal and central α-helix (Bourgeat *et al.*, 2021). More precisely, the allocation measure detects perturbations in the 2D, 3D and 4D structural levels known to cover fast microsecond to millisecond collective motions of small length scale (2D-3D structures) to slow millisecond to second collective motions of larger length scale (large domain, interfaces/chain). The same is observed by BDS, fast and slow collective motions for the N- and C-terminal domains and slow collective motions for the central α-helix which presents an allocation perturbation over a larger sequence domain (residue 58 to 81, Figure 1C) than the C- and N-terminal domains.

In addition, the central α-helix, where most of the changes in the allocation measure are observed, has only one mutation (ALA75THR), showing that the allocation measure tracks perturbations from the local scale, i.e. the site of mutations, to larger scales. The lack of mutations in the central α-helix makes it difficult to predict the helix dynamic perturbation from sequence and/or structural analysis. It emphasizes the specificity of the allocation measure in detecting multiscale dynamics perturbations.

The comparison of LTB_5 and $CtxB_5$ validates our hypothesis that differences in protein dynamics can be retrieved by differences in the local structural-level allocation of atomic interactions into the 1D, 2D, 3D and 4D classes. The next case studies challenge the sensitivity of the measure to detect the impact of milder sequence changes on protein folding dynamics.

3.2 Transthyretin variants

The allocation measure is applied on T119Y and T119M non-pathogenic TTR variants (Saelices *et al.*, 2015; Saraiva, 2001) and on L55P and V30M pathogenic TTR variants that have higher propensity to misfolding and amyloidosis than WT TTR (Saraiva, 2001). Again, the four TTR variants verify the condition of structural integrity with RMSD values: 0.58 Å, 0.58 Å, 0.54 Åand 0.34 Åfor L55P TTR, V30M TTR, T119Y TTR and T119M TTR, respectively, Figure 2A). The dimeric unit has

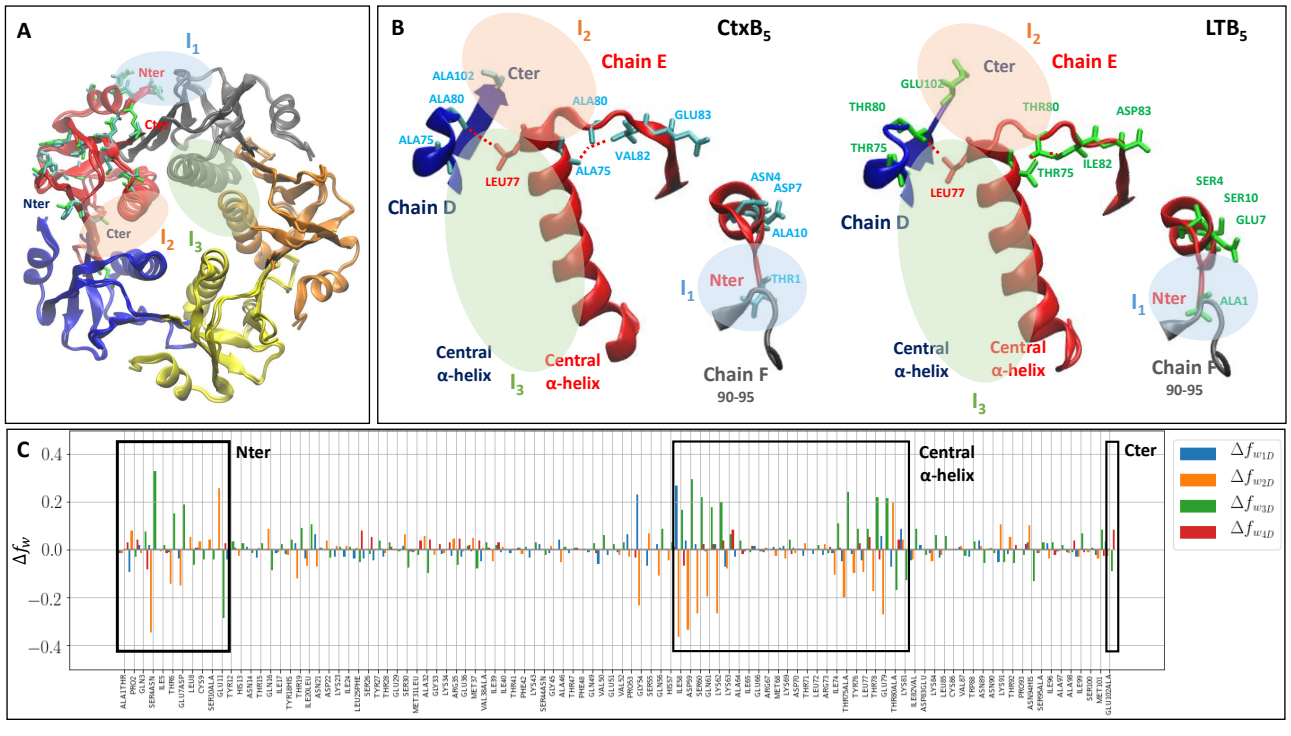

Fig. 1. B subunits of AB_5 toxins. A: Superposition of the X-ray structures of the $CtxB_5$ pentamer (PDB: 1eei) and the LTB_5 pentamer (PDB: 1ltr). The 18 mutated positions are highlighted in one chain (cyan amino acids for $CtxB_5$ and green amino acids for LTB_5). B: Details of the X-ray structures of $CtxB_5$ and the LTB_5. C: Changes in the local structural-level allocation of atomic interactions in LTB_5 compared to $CtxB_5$ (chain E), where $\Delta f_{w_{1D},i} = f^{LTB_5}_{w_{1D},i} - f^{CtxB_5}_{w_{1D},i}$ (and similarly for $\Delta f_{w_{2D},i}$, $\Delta f_{w_{3D},i}$ and $\Delta f_{w_{4D},i}$). Positive (negative) values of $\Delta f_{w_{x},i}$ mean that residue i allocates a higher (lower) fraction of contacts to the x structural level in LTB_5 compared to $CtxB_5$.

been employed for the calculations, whereas the results (Figure 2C) are plotted only for chain A, for simplicity. The plots for both chains are reported in Figure S2 and S3. The results that were not consistent for both chains have been considered as potential X-ray artifacts and not taken into further consideration.

The T119Y TTR variant shows very little changes in the allocation measure compared to the WT TTR, apart from the 98-103 segment (Figure 2B and 2C). Changes in this segment, which is part of a long coil, are observed for the four variants, with changes in chain A different from the changes in chain B (Figure S2 and S3), suggesting a perturbation related to an X-ray bias rather than a dynamic fault. The absence of changes in the allocation measure of T119Y TTR is consistent with the fact that T119Y does not have dynamic perturbations underlying a propensity to amyloid formation (Saelices *et al.*, 2015).

In contrast, the pathogenic L55P TTR variant shows large differences compared to the WT due to the reallocation of the atomic interactions of many residues within the segment 50-58 (Figure 2C). The residues in the mutant devote more resources to 2D interactions and less to 3D interactions. This may increase the 3D-mobility of the segment 50-58 by having less atomic space occupancy in the mutant and hence more empty space for atomic motions, and in turn make the segment 50-58 available for intermolecular interactions with another monomer, pre-requisite for amyloid fiber formation. Intra-molecular perturbations, suggesting a folding perturbation, agree with the L55P fiber diameter similar to the TTR tetramer size (Lashuel *et al.*, 1999), meaning that dissociation can happen but is not necessary for fiber formation and is consistent with L55P TTR amyloid fiber models (Pacini *et al.*, 2020; Sebastião *et al.*, 1998).

The changes in allocation of the other pathogenic variant, V30M TTR, are distinct from the ones of L55P TTR: lower absolute value and 2D and 3D perturbations in a segment (residues 86-92) at the dimer interface (Figure 2). This is consistent with the fact that ex-vivo studies have shown that V30M TTR unfolds prior to fiber formation, contrary to WT TTR (Dasari *et al.*, 2020; Schmidt *et al.*, 2019). Moreover, the V30M interface perturbations is also compatible with the mutation leading to dissociation prior fiber formation, in agreement with the diameter of the V30M fiber higher than the size of the TTR tetramer (Serpell *et al.*, 1995).

We find little perturbation for the non-pathogenic T119Y variant and substantial perturbation for the pathogenic L55P and V30M variants pointing out a quantitative relation between allocation perturbations and pathology as previously shown by the localized frustration measure (Ferreiro *et al.*, 2007; Kumar *et al.*, 2013). Moreover, the L55P and V30M variants exhibit different allocation perturbations at different areas, adding a qualitative relation between the perturbations and pathology. The consistency between the different allocation measures for T119Y, L55P and V30M and their different experimental dynamics supports the reliability of the allocation measure in detecting dynamics perturbations even for single-mutation variants.

T119M exhibits the most perturbed allocation measure with three segments having changes in the allocations of 2D and 3D interactions. This result reveals that perturbations do not necessarily imply pathology, meaning that a prognostic of pathology cannot rely only on a diagnostic of perturbation quantification. Here prognostic is defined as 'knowledge beforehand of how a situation is likely to turn out' (`www.merriam-webster.com/dictionary/prognosis#note-1`, last visited on 17th May 2021). For a prognostic, the perturbations diagnosed with the allocation measure must be somehow

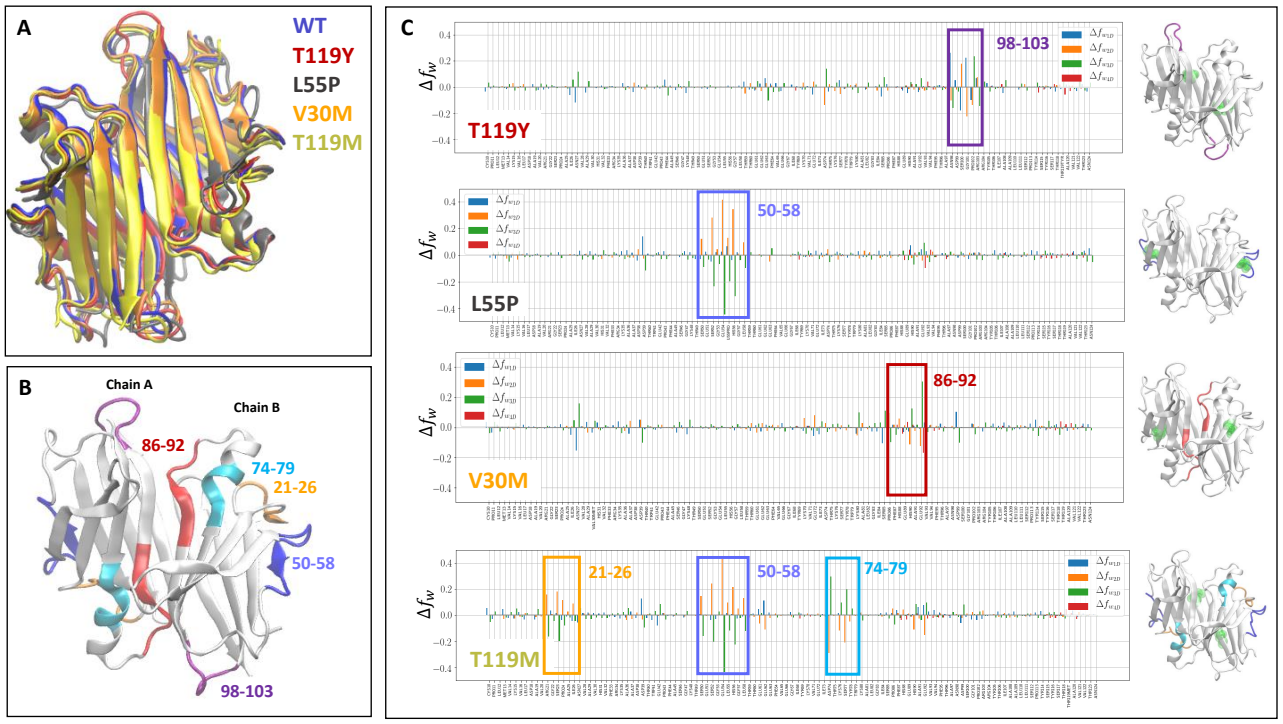

Fig. 2. Transthyretin (TTR) variants. A: Superposition of the X-ray structures of the TTR dimer variants: WT (PDB: 1f41), T119Y (PDB: 4tne), L55P (PDB: 3djz), V30M (PDB: 3kgs) and T119M (PDB: 1bze). B: X-ray structure of the WT TTR dimer. The segments involved in changes in local structural-level allocation in the variants with respect to the WT are highlighted. C: Changes in the local structural-level allocation of atomic interactions in the TTR variants compared to WT TTR (chain A), where $\Delta f_{w_{1D,i}} = f^{variant}_{w_{1D,i}} - f^{WT}_{w_{1D,i}}$ (and similarly for $\Delta f_{w_{2D,i}}$, $\Delta f_{w_{3D,i}}$ and $\Delta f_{w_{4D,i}}$). Positive (negative) values of $\Delta f_{w_{x,i}}$ mean that residue i allocates a higher (lower) fraction of contacts to the x structural level in the variant compared to the WT. Next to each plot, the segment involved in changes in local structural-level allocation of atomic interactions in the variants compared to the WT TTR are highlighted in the variant's X-ray structure, with the same color code as the boxes in the plots and panel B. The mutated position is highlighted in green.

classified as functionally (folding) tolerated or functionally (folding) harmful. Three perturbed segments introduce the possibility that changes on one segment compensate changes in others, resulting in functionally (folding) tolerated perturbations. The fact that L55P and T119M present perturbations in the 50-58 segment, but only L55P promotes fiber formation, supports the presence of a compensatory mechanism in T119M (Dorantes-Gilardi *et al.*, 2018). We calculated that T119M has a higher number of atomic interactions between these two segments than the WT TTR, hence the mobility of the segment 50-58 in L55P might be prevented in T119M by tightening it to the 21-26 segment. Yet, the higher atomic interactions could also provide more backup links contributing to the higher stability of T119M reported experimentally, and have thus no relation with the 50-58 segment allocation perturbations in L55P (McCutchen *et al.*, 1995; Quintas *et al.*, 2001; Dorantes-Gilardi *et al.*, 2018). T119M is also known to prevent amyloidogenesis when present together with the V30M mutation in heterozygous individuals (Alves *et al.*, 1996) and to protect the V30M TTR tetramer from dissociation in vitro when some V30M chains are replaced by T119M chains (Hammarstrom, 2001). This is another type of compensatory mechanism where the dynamics perturbation caused by the V30M mutation on one chain is corrected by a dynamics perturbation caused by the T119M mutation on another chain. Again having three perturbed segments in T119M is consistent with a T119M-V30M compensatory mechanism but the allocation measure is not enough to determine the exact mechanism. In fact, the TTR cases show the complexity of establishing a prognostic that will link a perturbation diagnostic to pathology.

The analysis of the B-pentamers of AB_5 toxins and of TTR variants has proven that the comparison of local structural-level allocation of atomic interactions is a good measure of differences in protein slow folding dynamics and of the propagation of perturbations from local to larger scales. To explore the issue of common drug design for protein variants sharing structural integrity, we compared the allocation measure of the main proteases (Mpro's) of two coronaviruses, SARS-CoV and SARS-CoV-2.

3.3 Main proteases of SARS coronaviruses

The Mpro's of SARS-CoV and SARS-CoV-2 have very similar sequence (12 mutations over 305 residues, corresponding to 96% identity, Figure 3B) and similar crystalline structure (RMSD = 0.71 Å for one Mpro chain, Figure 3B). The percentage of sequence identity between the two Mpro's is intermediate with respect to the previous case studies and the RMSD is as low (¡ 1 Å) so we assume that differences in the allocation measure will measure differences in protein dynamics. The main differences in atomic interactions allocation are located at the C-terminal domain (Figure 3A), where a higher fraction of atomic interactions is devolved to 2D and a lower fraction to 3D ($\Delta f_{w_{2D}} > 0$, $\Delta f_{w_{3D}} < 0$ and longer α-helix, Figure 3C) in SARS-CoV-2 Mpro compared to SARS-CoV Mpro. Moreover, the C-terminal coil points towards the active domain in SARS-CoV-2 Mpro and towards the rest of the extra domain in SARS-CoV Mpro (Figure 3C). This is reflected by $\Delta f_{w_{3D}} < 0$ at residues residues 213 to 215 and GLN256 of the extra domain and $\Delta f_{w_{3D}} > 0$ at residue TYR154 of the active domain (Figure 3A and 3C).

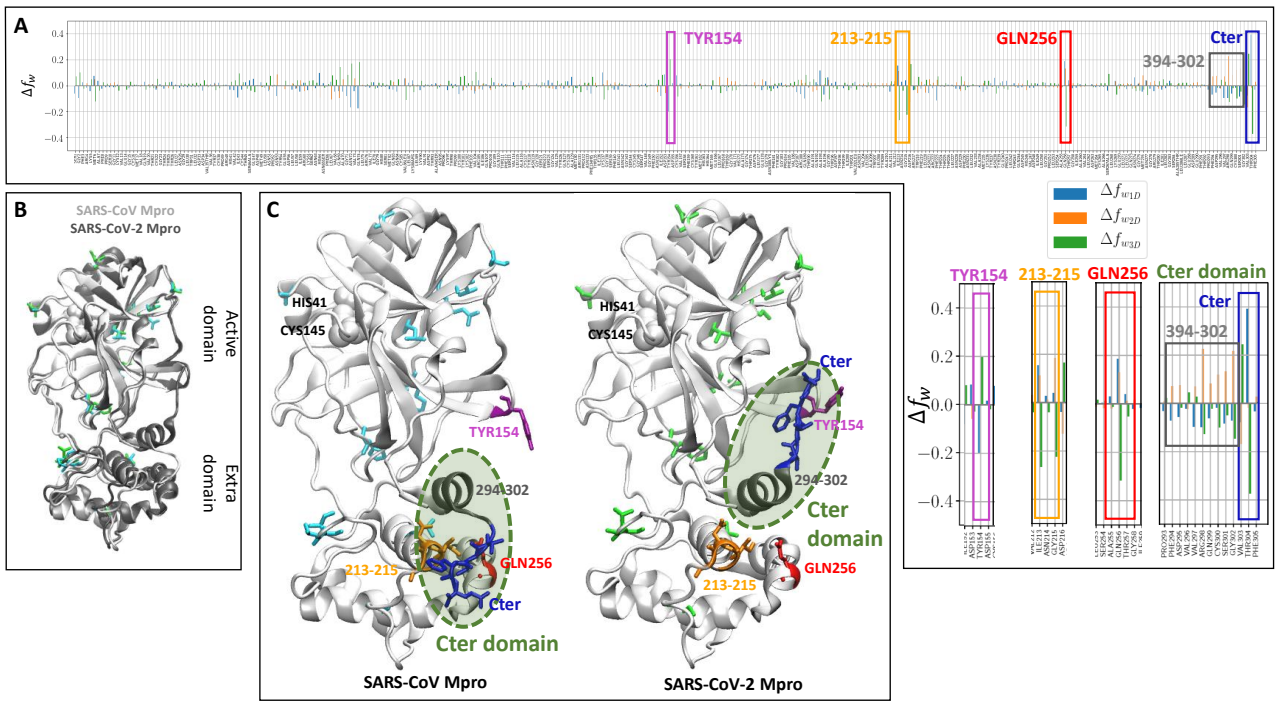

Fig. 3. Main proteases (Mpro) of coronaviruses SARS-CoV and SARS-CoV-2. A: changes of local structural-level allocation of atomic interactions along the Mpro monomer, where $\Delta f_{w1D,i} = f_{w1D,i}^{SARS-CoV-2} - f_{w1D,i}^{SARS-CoV}$ (and similarly for $\Delta f_{w2D,i}$ and $\Delta f_{w3D,i}$). Positive (negative) values of $\Delta f_{w_{x},i}$ mean that residue i allocates a higher (lower) fraction of contacts to the x structural level in SARS-CoV-2 compared to SARS-CoV. B: Superposition of the X-ray structures of the Mpro monomer of SARS-CoV (PDB: 2h2z) and SARS-CoV-2 (PDB: 6y2e). The 12 mutated positions are highlighted in one chain (cyan amino acids for SARS-CoV and green amino acids for SARS-CoV-2). C: Details of the X-ray structures of SARS-CoV Mpro and SARS-CoV-2 Mpro. The segments involved in changes in local structural-level allocation of atomic interactions are highlighted with different colors and the active sites (HIS 41 and CYS145) are represented in space-fill.

Much effort is devolved to the design of drugs to inhibit the Mpro's of a broad spectrum of coronaviruses (Burley, 2020; Jin *et al.*, 2020; Yang *et al.*, 2005; Carli *et al.*, 2020), but our results suggest differences in the enzyme dynamics, particularly of the C-terminal dynamics, a region whose role in the enzymatic activity has been shown experimentally for SARS-CoV (Shi and Song, 2006). This result supported by the other cases of study advises against a broad-spectrum drug strategy due to the large-scale dynamics differences that are likely to prevent same-drug recognition. It therefore counsels on more strain-customized drugs to inhibit the function of the Mpro of coronaviruses.

4 Conclusion

Our hypothesis that the difference in the allocation of atomic interactions to structural levels (1D, 2D, 3D and 4D) between variants will reveal areas with structural-level dynamic perturbations associated with the mutations is verified for the four cases of study. The methodology is fast and computationally-light and diagnoses robust and functionally-sensitive mutations in terms of their impact on the multi-scale protein dynamics. As an example, the calculation for the 305 amino acids of the two SARS-CoV Mpro structures takes less than 2 seconds on a laptop with an Intel Core i7 processor. From the results, we can deduce that the allocation of the atomic interactions into structural levels contains information on the protein dynamics. Further application to a large database of protein variants will establish the ranges of structural allocation changes that associate with different dynamic perturbations. This could open-up novel directions towards the decoding of a protein dynamics from the static information contained in its crystalline structure. The allocation measure provides information which will allow digging into the molecular mechanisms underlying pathogenic mutations and compensatory mechanisms involving error-correcting mutations, and thus assist in assigning drugs to variants for personalized therapy.

Acknowledgements

We thank the CNRS MITI for funding the work.

Funding

This work has been supported by the CNRS MITI (Grant OASIC project Go-Pro2018).

References

Achoch, M., Dorantes-Gilardi, R., Wymant, C., Feverati, G., Salamatian, K., Vuillon, L., and Lesieur, C. (2016). Protein structural robustness to mutations: an in silico investigation. *Physical Chemistry Chemical Physics*, **18**(20), 13770–13780.

Alves, I., Jacobson, D., Torres, M., Coelho, T., Holmgren, G., and Saraiva, M. (1996). Compound heterozygoty in patients with ttr-related amyloidosis. *Neuromuscular Disorders*, **6**, S19.

Amaral, M., Kokh, D., Bomke, J., Wegener, A., Buchstaller, H., Eggenweiler, H., Matias, P., Sirrenberg, C., Wade, R., and Frech, M. (2017). Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nature communications*, **8**(1), 1–14.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. (2012). Debtrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific reports*, **2**(1), 1–6.

Bemporad, F. and Chiti, F. (2012). Protein misfolded oligomers: Experimental approaches, mechanism of formation, and structure-toxicity relationships. *Chemistry & Biology*, **19**(3), 315–327.

Bourgeat, L., Serghei, A., and Lesieur, C. (2019). Experimental protein molecular dynamics: Broadband dielectric spectroscopy coupled with nanoconfinement. *Scientific reports*, **9**(1), 1–12.

Bourgeat, L., Pacini, L., Serghei, A., and Lesieur, C. (2021). Experimental diagnostic of sequence-variant dynamic perturbations revealed by broadband dielectric spectroscopy. *Structure (in press)*.

Buchenberg, S., Sittel, F., and Stock, G. (2017). Time-resolved observation of protein allosteric communication. *Proceedings of the National Academy of Sciences*, **114**(33), E6804–E6811.

Burley, S. K. (2020). How to help the free market fight coronavirus. *Nature*, **580**(7802), 167–167.

Carli, M., Sormani, G., Rodriguez, A., and Laio, A. (2020). Candidate binding sites for allosteric inhibition of the sars-cov-2 main protease from the analysis of large-scale molecular dynamics simulations. *The Journal of Physical Chemistry Letters*, **12**, 65–72.

Cendron, L., Trovato, A., Seno, F., Folli, C., Alfieri, B., Zanotti, G., and Berni, R. (2009). Amyloidogenic potential of transthyretin variants: INSIGHTS FROM STRUCTURAL AND COMPUTATIONAL ANALYSES. *Journal of Biological Chemistry*, **284**(38), 25832–25841.

Dai, W., Zhang, B., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L.-K., Xu, Y., Yang, H., and Liu, H. (2020). Structure-based design of antiviral drug candiyears targeting the SARS-CoV-2 main protease. *Science*, page eabb4489.

Dasari, A. K., Hung, I., Gan, Z., and Lim, K. H. (2019). Two distinct aggregation pathways in transthyretin misfolding and amyloid formation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1867**(3), 344–349.

Dasari, A. K. R., Hung, I., Michael, B., Gan, Z., Kelly, J. W., Connors, L. H., Griffin, R. G., and Lim, K. H. (2020). Structural characterization of cardiac ex vivo transthyretin amyloid: Insight into the transthyretin misfolding pathway in vivo. *Biochemistry*, **59**(19), 1800–1803.

Degiacomi, M. T., Iacovache, I., Pernot, L., Chami, M., Kudryashev, M., Stahlberg, H., Van Der Goot, F. G., and Dal Peraro, M. (2013). Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nature chemical biology*, **9**(10), 623–629.

Dorantes-Gilardi, R., Bourgeat, L., Pacini, L., Vuillon, L., and Lesieur, C. (2018). In proteins, the structural responses of a position to mutation rely on the goldilocks principle: not too many links, not too few. *Physical Chemistry Chemical Physics*, **20**(39), 25399–25410.

Fang, C., Frontiera, R. R., Tran, R., and Mathies, R. A. (2009). Mapping gfp structure evolution during proton transfer with femtosecond raman spectroscopy. *Nature*, **462**(7270), 200–204.

Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007). Localizing frustration in native proteins and protein assemblies. *Proceedings of the National Academy of Sciences*, **104**(50), 19819–19824.

Findsen, E. W., Scott, T. W., Chance, M. R., Friedman, J. M., and Ondrias, M. R. (1985). Picosecond time-resolved raman studies of photodissociated carboxymyoglobin. *Journal of the American Chemical Society*, **107**(11), 3355–3357.

Gersting, S. W., Kemter, K. F., Staudigl, M., Messing, D. D., Danecka, M. K., Lagler, F. B., Sommerhoff, C. P., Roscher, A. A., and Muntau, A. C. (2008). Loss of function in phenylketonuria is caused by impaired molecular motions and conformational instability. *The American Journal of Human Genetics*, **83**(1), 5–17.

Gheeraert, A., Pacini, L., Batista, V. S., Vuillon, L., Lesieur, C., and Rivalta, I. (2019). Exploring allosteric pathways of a v-type enzyme with dynamical perturbation networks. *The Journal of Physical Chemistry B*, **123**(16), 3452–3461.

Haliloglu, T. and Bahar, I. (2015). Adaptability of protein structures to enable functional interactions and evolutionary implications. *Current Opinion in Structural Biology*, **35**, 17–23.

Hamilton, J. A. and Benson, M. D. (2001). Transthyretin: a review from a structural perspective. *Cellular and Molecular Life Sciences CMLS*, **58**(10), 1491–1521.

Hammarstrom, P. (2001). Trans-suppression of misfolding in an amyloid disease. *Science*, **293**(5539), 2459–2462.

Henzler-Wildman, K. and Kern, D. (2007). Dynamic personalities of proteins. *Nature*, **450**(7172), 964–972.

Herbst, J., Heyne, K., and Diller, R. (2002). Femtosecond infrared spectroscopy of bacteriorhodopsin chromophore isomerization. *Science*, **297**(5582), 822–825.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, **14**(1), 33–38.

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., and Yang, H. (2020). Structure of mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*.

Kamada, R., Nomura, T., Anderson, C. W., and Sakaguchi, K. (2011). Cancer-associated p53 tetramerization domain mutants: quantitative analysis reveals a low threshold for tumor suppressor inactivation. *Journal of Biological Chemistry*, **286**(1), 252–258.

Kolano, C., Helbing, J., Kozinski, M., Sander, W., and Hamm, P. (2006). Watching hydrogen-bond dynamics in a β-turn by transient two-dimensional infrared spectroscopy. *Nature*, **444**(7118), 469–472.

Kumar, S., Clarke, D., and Gerstein, M. (2013). Localized structural frustration for evaluating the impact of sequence variants. *Nucleic acids research*, **44**(21), gkw927.

Lashuel, H. A., Wurth, C., Woo, L., and Kelly, J. W. (1999). The most pathogenic transthyretin variant, l55p, forms amyloid fibrils under acidic conditions and protofilaments under physiological conditions. *Biochemistry*, **38**(41), 13560–13573.

Leitner, D. M. and Yamato, T. (2018). Mapping energy transport networks in proteins. *Reviews in Computational Chemistry, Volume 31*, pages 63–113.

Lesieur, C. and Schulten, K. (2015). Editorial overview: Theory and simulation. *Current opinion in structural biology*, **31**, v–vi.

Lesieur, C., Cliff, M. J., Carter, R., James, R. F., Clarke, A. R., and Hirst, T. R. (2002). A kinetic model of intermediate formation during assembly of cholera toxin b-subunit pentamers. *Journal of Biological Chemistry*, **277**(19), 16697–16704.

McCutchen, S. L., Lai, Z., Miroy, G. J., Kelly, J. W., and Colon, W. (1995). Comparison of lethal and nonlethal transthyretin variants and their relationship to amyloid disease. *Biochemistry*, **34**(41), 13527–13536.

McLaughlin Jr, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature*, **491**(7422), 138–142.

Mizutani, Y. and Kitagawa, T. (2001). Ultrafast dynamics of myoglobin probed by time-resolved resonance raman spectroscopy. *The chemical record*, **1**(3), 258–275.

Munoz, V. and Cerminara, M. (2016). When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochemical Journal*, **473**(17), 2545–2559.

Naganathan, A. N. (2019). Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Current Opinion in Structural Biology*, **54**, 1–9.

Pacini, L., Vuillon, L., and Lesieur, C. (2020). Induced perturbation network and tiling for modeling the l55p transthyretin amyloid fiber. *Procedia Computer Science*, **178C**, 8–17.

Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., Yu, H., Wu, Z., and Schulten, K. (2015). Molecular dynamics simulations of large macromolecular complexes. *Current opinion in structural biology*, **31**, 64–74.

Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J. M., and Brito, R. M. (2001). Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *Journal of Biological Chemistry*, **276**(29), 27207–27213.

Ramaswamy, V. K., Musson, S. C., Willcocks, C. G., and Degiacomi, M. T. (2021). Deep learning protein conformational space with convolutions and latent interpolations. *Physical Review X*, **11**(1), 011052.

Ruberg, F. L. and Berk, J. L. (2012). Transthyretin (TTR) cardiac amyloidosis. *Circulation*, **126**(10), 1286–1300.

Ruddock, L. W., Coen, J. J., Cheesman, C., Freedman, R. B., and Hirst, T. R. (1996). Assembly of the b subunit pentamer of escherichia coli heat-labile enterotoxin: Kinetics and molecular basis of rate-limiting steps in vitro. *Journal of Biological Chemistry*, **271**(32), 19118–19123.

Saelices, L., Johnson, L. M., Liang, W. Y., Sawaya, M. R., Cascio, D., Ruchala, P., Whitelegge, J., Jiang, L., Riek, R., and Eisenberg, D. S. (2015). Uncovering the mechanism of aggregation of human transthyretin. *Journal of Biological Chemistry*, **290**(48), 28932–28943.

Saraiva, M. J. M. (2001). Transthyretin mutations in hyperthyroxinemia and amyloid diseases. *Human mutation*, **17**(6), 493–503.

Schmidt, M., Wiese, S., Adak, V., Engler, J., Agarwal, S., Fritz, G., Westermark, P., Zacharias, M., and Fändrich, M. (2019). Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nature Communications*, **10**(1), 5008.

Schneider, R., Blackledge, M., and Jensen, M. R. (2019). Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using nmr spectroscopy. *Current opinion in structural biology*, **54**, 10–18.

Sebastião, M. P., Saraiva, M. J., and Damas, A. M. (1998). The crystal structure of amyloidogenic leu55→ pro transthyretin variant reveals a possible pathway for transthyretin polymerization into amyloid fibrils. *Journal of Biological Chemistry*, **273**(38), 24715–24722.

Serpell, L., Sunde, M., Fraser, P., Luther, P., Morris, E., Sangren, O., Lundgren, E., and Blake, C. (1995). Examination of the structure of the transthyretin amyloid fibril by image reconstruction from electron micrographs.

Shi, J. and Song, J. (2006). The catalysis of the SARS 3c-like protease is under extensive regulation by its extra domain. *FEBS Journal*, **273**(5), 1035–1045.

Sippl, M. J. and Wiederstein, M. (2012). Detection of spatial correlations in protein structures and molecular complexes. *Structure*, **20**(4), 718–728.

Thusberg, J. and Vihinen, M. (2009). Pathogenic or not? and if so, then how? studying the effects of missense mutations using bioinformatics methods. *Human Mutation*, **30**(5), 703–714.

Unicomb, S., Iñiguez, G., and Karsai, M. (2018). Threshold driven contagion on weighted networks. *Scientific reports*, **8**(1), 1–10.

Vuillon, L. and Lesieur, C. (2015). From local to global changes in proteins: a network view. *Current Opinion in Structural Biology*, **31**, 1–8.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., and Rao, Z. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences*, **100**(23), 13190–13195.

Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M., and Rao, Z. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biology*, **3**(10), e324.

Yee, A. W., Aldeghi, M., Blakeley, M. P., Ostermann, A., Mas, P. J., Moulin, M., de Sanctis, D., Bowler, M. W., Mueller-Dieckmann, C., Mitchell, E. P., Haertlein, M., de Groot, B. L., Boeri Erba, E., and Forsyth, V. T. (2019). A molecular mechanism for transthyretin amyloidogenesis. *Nature Communications*, **10**(1), 925.

Zhang, S., Krieger, J. M., Zhang, Y., Kaya, C., Kaynak, B., Mikulska-Ruminska, K., Doruker, P., Li, H., and Bahar, I. (2021). Prody 2.0: Increased scale and scope after 10 years of protein dynamics modelling with python. *Bioinformatics*.

Zrimi, J., Ng Ling, A., Giri-Rachman Arifin, E., Feverati, G., and Lesieur, C. (2010). Cholera toxin b subunits assemble into pentamers - proposition of a fly-casting mechanism. *PLoS ONE*, **5**(12), e15347.