A conformational fingerprint for amyloidogenic light chains.

Cristina Paissoni[1,a], Sarita Puri[1,a], Luca Broggini[2], Manoj K. Sriramoju[3], Martina Maritan[1], Rosaria Russo[1], Valentina Speranzini[1], Federico Ballabio[1], Mario Nuvolone[4], Giampaolo Merlini[4], Giovanni Palladini[4], Shang-Te Danny Hsu[3,5,6], Stefano Ricagno[1,2*], Carlo Camilloni[1*]

[1]Department of Bioscience, University of Milan, Milan, Italy.
[2]Institute of Molecular and Translational Cardiology, IRCCS, Policlinico San Donato, Milan, Italy.
[3]Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan.
[4]Department of Molecular Medicine, University of Pavia, Pavia, Italy; Amyloidosis Research and Treatment Center, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy.
[5]Institute of Biomedical Sciences, National Taiwan University, Taipei, Taiwan.
[6]International Institute for Sustainability with Knotted Chiral Meta Matter (SKCM2), Hiroshima University, Higashi-Hiroshima, Japan.

[a]CP and SP contributed equally to this work and share the first authorship.

*Stefano Ricagno and Carlo Camilloni
Email: stefano.ricagno@unimi.it, carlo.camilloni@unimi.it

Author Contributions: CP performed and analyzed simulations and analyzed SAXS experiments. SP prepared HDX-MS samples and analyzed the data. LB, MM, RR, and VS prepared SAXS samples and performed SAXS experiments. MKS collected HDX-MS data. FB took care of data sharing and deposition and contributed to data analysis. MV, GM, GP, STDH, SR and CC supervised the project. SP, SR and CC wrote the manuscript with contributions from all authors. SR and CC designed the project.

Keywords: Amyloidogenic light chain, Conformational dynamics, Small angle X-ray scattering, Molecular dynamics, Hydrogen deuterium exchange.

Abstract

Immunoglobulin light chain amyloidosis (AL) and multiple myeloma (MM) both share the overproduction of a clonal light chain (LC). However, while LCs in MM remain soluble in circulation, AL LCs misfold into toxic soluble species and amyloid fibrils that accumulate in organs, leading to distinct clinical manifestations. The significant sequence variability of LCs has hindered understanding of the mechanisms driving LC aggregation. Nevertheless, emerging biochemical properties, including dimer stability, conformational dynamics, and proteolysis susceptibility, distinguish AL LCs from those in MM under native conditions. This study aimed to identify a conformational fingerprint distinguishing AL from MM LCs. Using small-angle X-ray scattering (SAXS) under native conditions, we analyzed four AL and two MM LCs. We observed that AL LCs exhibited a slightly larger radius of gyration and greater deviations from X-ray crystallography-determined or predicted structures, reflecting enhanced conformational dynamics. SAXS data, integrated with molecular dynamics (MD) simulations, revealed a conformational ensemble where LCs adopt multiple states, with variable and constant domains either bent or straight. AL LCs displayed a distinct, low-populated, straight conformation (termed H state), which maximized solvent accessibility at the interface between constant and variable domains. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) experimentally validated this H state. These findings reconcile diverse experimental observations and provide a precise structural target for future drug design efforts.

Introduction

Immunoglobulin light-chain (AL) amyloidosis is a systemic disease associated with the overproduction and subsequent amyloid aggregation of patient-specific light chains (LCs) (1-4). Such aggregation may take place in one or several organs, the heart and kidneys being the most affected ones (1). AL originates from an abnormal proliferation of a plasma cell clone that results in LCs overexpression and over-secretion in the bloodstream (1). LCs belonging both to lambda (λ) and kappa (κ) isotypes are associated with AL; however, λ-LCs are greatly overrepresented in the repertoire of AL patients. Specifically, AL-causing LCs (AL-LCs) most often belong to a specific subset of lambda germlines such as *IGLV6* (λ6), *IGLV1* (λ1), and *IGLV3* (λ3) (5-8).

λ-LCs are dimeric in solution with each subunit characterized by two immunoglobulin domains, a constant domain (CL) with a highly conserved sequence and a variable domain (VL) whose extreme sequence variability is the result of genomic recombination and somatic mutations (9-12). VL domains are generally indicated as the key responsible for LC amyloidogenic behavior. The observation that the fibrillar core in most of the structures of *ex-vivo* AL amyloid fibrils consist of VL residues further strengthens this hypothesis (13-17). However, in a recent cryo-electron microscopy (Cryo-EM) structure, a stretch of residues belonging to the CL domain is also part of the fibrillar core, and mass spectrometry (MS) analysis of several *ex vivo* fibrils from different patients indicates that amyloids are composed of several LC proteoforms including full-length LCs (18-21).

Interestingly the overproduction of a light chain is a necessary but not sufficient condition for the onset of AL. Indeed, the uncontrolled production of a clonal LC is often associated with Multiple Myeloma (MM), a blood cancer, but only a subset of MM patients develops AL, thus indicating that specific sequence/biophysical properties determine LC amyloidogenicity and AL onset (12,22-24). To date, the extreme sequence variability of AL-LCs has prevented the identification of sequence patterns predictive of LC amyloidogenicity, however, it has been reproducibly reported that several biophysical properties correlate with LC aggregation propensity. AL-LCs display a lower thermodynamic and kinetic fold stability compared to non-amyloidogenic LCs found overexpressed in MM patients (named hereafter MM-LCs) (12,20-24).

Previous work on LCs has indicated how differences in conformational dynamics can play a role in the aggregation properties of AL-LCs (22-26). Oberti *et al.* have compared multiple λ-LCs

obtained from either AL patients or MM patients identifying the susceptibility to proteolysis as the best biophysical parameter distinguishing the two sets (12). Weber *et al.* have shown, using a murine-derived κ-LC, how a modification in the linker region can lead to a greater conformational dynamic, an increased susceptibility to proteolysis, as well as an increased *in vitro* aggregation propensity (25). Additionally, AL-LC flexibility and conformational freedom have also been correlated to the proteotoxicity observed in patients affected by cardiac AL and experimentally verified in human cardiac cells and a *C. elegans* model (27,28). It is noteworthy that the amyloid LCs analyzed in this study were originally purified from patients with cardiac AL amyloidosis.

Here building on this previous work as well on our previous experience on β2-microglobulin, another natively folded amyloidogenic protein (29-33) we aimed at revealing the structural and dynamic determinants of LC amyloidogenicity. With this goal we investigated the native solution state dynamics of multiple λ-LCs by combining MD simulations, SAXS, and HDX-MS. Interestingly, we found a unique conformational fingerprint of amyloidogenic LCs corresponding to a low-populated state (defined as H-state in the following) characterized by extended linkers, with an accessible VL-CL interface and possible structural rearrangements in the CL-CL interface.

Results

SAXS suggests differences in the conformational dynamics of amyloidogenic and non-amyloidogenic LC. SAXS data were acquired either in bulk or in-line with size-exclusion chromatography (SEC) for a set of six LCs previously described (cf. **Table 1** and **Methods**). Four of these LCs (referred as H3, H7, H18 in ref. (12) and AL55 in ref. (16)) were identified in multiple AL patients (AL-LC), white two (referred as M7, and M10 in ref. (12)), were identified in MM patients (MM-LC). These LCs cover multiple germlines, with H18 and M7 belonging to the same germline (cf. **Table 1**). The sequence identity is the largest for H18 and M7 (91.6%) while is the lowest for AL55 and M7 (75.2%). A table showing the statistics for all pairwise sequence alignments is reported in **Table S1** in the Supporting Information. For H3, H7, and M7 a crystal structure was previously determined (12) while for H18, AL55, and M10 we obtained a model using either homology modeling (H18 and AL55) or AlphaFold2 (M10). Qualitatively, the SAXS curves in **Figure 1** did not reveal any macroscopic deviation of the solution behavior with respect to the crystal or model conformation. For each LC, we compared the experimental and theoretical curves calculated from the LC structures (cf. **Table 1**) analyzing the residuals and the associated χ^2. The analysis indicated a discrepancy between the model conformation and the data in the case of the AL-LCs, which was instead not observed in the case of the MM-LCs. For AL-LC, residuals deviate from normality in the low *q* region, suggesting some variability in the global size of the system. Additionally, a weak trend distinguishing AL-LC from MM-LCs could be identified in the radius of gyration (Rg) as derived from the Guinier approximation (cf. **Table 1**). H3, H7, H18, and AL55 have an Rg that is 0.5 to 0.8 Å larger than M7 and M10, a small but statistically significant difference (p-value < 10^{-5}). Overall, the SAXS measurements point to less compact and more structurally heterogeneous AL-LCs compared to more compact and structurally homogeneous MM-LCs.

MD simulations reveal a conformational fingerprint for amyloidogenic light chains. To investigate the conformational dynamics of the six LCs we performed Metadynamics Metainference (M&M) MD simulations employing the SAXS curves ($q<0.3$ Å) as restraints (cf. **Material and Methods**) (34-37). Metainference is a Bayesian framework that allows the integration of experimental knowledge on-the-fly in MD simulations improving the latter while accounting for the uncertainty in the data and their interpretation. Metadynamics is an enhanced sampling technique able to speed up the sampling of the conformational space of complex systems. The combination of SAXS and MD simulations has been shown to be effective for multi-domain proteins as well as for intrinsically disordered proteins (38-40).

For each LC, we performed two independent M&M simulations coupled by the SAXS restraint, accumulating around 120-180 µs of MD per protein (cf. **Material and Methods** and **Table 2**). The resulting conformational ensembles resulted in a generally improved agreement with the SAXS data employed as restraints (cf. **Table 2** and **Figure 2A**). To investigate differences in the LC local flexibility we first analyzed the root mean square fluctuations (RMSF) for the CL and VL separately, averaging over the chains and the replicates, **Figure 2B**. The RMSF indicates comparable flexibility in most of the regions, with differences localized in the termini and in some loops. The VL of amyloidogenic LCs are generally more flexible than the MM ones, but this may be associated with the lengths of their complementarity-determining regions (CDRs). Indeed, M10 has the longest and most flexible CDR1 and the shortest and least flexible CDR3. Unexpectedly, there are some differences also in the CL domains. Here, in **Figure 2B**, the AL-LCs are always more flexible in at least one region even if these differences are relatively small. Overall, the RMSF does not provide a clear indication to differentiate AL and MM-LCs. To provide a global description of the dynamics of the six LC systems, we then introduced two collective variables, namely the elbow angle, describing the relative orientation of VL and CL dimers, and the distance between the VL and CL dimers center of mass, illustrated in **Figure 2C**.

In **Figure 3,** we report the free energy surfaces (FES) obtained from the processing of the two replicates of each LC as a function of the elbow angle and the CL-VL distance calculated from their center of mass. The visual inspection of the FES indicates converged simulation: in all cases, the replicates explore a comparable free-energy landscape with comparable features. All six LC FES share common features: a relatively continuous low free energy region along the diagonal, spanning configurations where the CL and VL are bent and close to each other (state L_B), and configurations where the CL and VL domains are straight and at relative distance between 3.4 and 4.1 nm (state L_S). A subset of LCs, namely H18, M7, and AL55, display conformations where the domains are straight in line (elbow angle greater than 2.5 rad) and in close vicinity, with a relative distance between the center of mass of less than 3.4 nm (state G). Of note, H18 and M7 belong to the same germline, letting us speculate that this state G may be germline specific. Most importantly, only the AL-LCs display configurations with CL and VL straight in line but well separated at relative distances greater than 4.1 nm, this state H seems to be a fingerprint specific for AL-LCs. A set of configurations exemplifying the four states is reported in **Figure 3**. The estimates of the populations for the four states L_B, L_S, G, and H are reported in **Table 3**. The quantitative analysis indicates that, within the statistical significance of the simulations, states L_B and L_S represent in all cases most of the conformational space. In the case of H18, AL55, and M7, the compact state G is also significantly populated (10-34%). The state H, associated with amyloidogenic LCs, is populated between 5 and 10% in H3, H7, H18, and AL55 and less than 1% in M7 and M10.

To identify additional differences between the conformations observed in state H and the rest of the conformational space, we focused our attention on the VL-VL and CL-CL dimerization interfaces. In **Figure 4,** we showed the free energy as a function of the distance between the CL domains versus the distance between the VL domains for each of the four states for one of the two simulations performed on H3; the same analysis for all other simulations is shown in **Figures S2 to S7** in the Supporting Information. From the comparison of the FESes, it is clear that only in the conformations corresponding to the state H do the CL-CL dimers display an alternative configuration. In the case of H3, the CL-CL domains in the H state are characterized by a shift towards configurations characterized by a larger distance, the same is observed in the case of H18 and AL55, while in the case of H7 the H state is characterized by a smaller distance between the CL domains.

Our conformational ensembles allowed us to hypothesize a conformational fingerprint for AL proteins, namely the presence of a weakly but significantly populated state (H) characterized by a more extended quaternary structure, with VL and CL dimers well separated, and with perturbed CL-CL interfaces.

HDX independently validates the amyloidogenic LC conformational fingerprint. To gain further molecular insight into how the dynamics of the tertiary and quaternary structures can be differentiated in AL- and MM-LCs, HDX-MS was performed on our set of proteins. HDX-MS probes the protein dynamics by monitoring the hydrogen-to-deuterium uptake over time and the obtained data well complement structural, biophysical, and computational data. Four LCs from our set (H3, H7, AL55, and M10) yielded good peptide sequence coverages of 98.6, 92.5, 98.6, and 99.1%, respectively, with redundancy of > 4.0 (**Figures S8A-S11A,** and **Table S2** in the Supporting Information). H18 and M7 were not included in this analysis due to their poor sequence coverage and were not further investigated. Due to sequence heterogeneity among our proteins, the common peptide analyses were not performed. Therefore, the relative deuterium exchange at different HDX time points of 0.5 to 240 min with respect to zero exchange time was used to compare the dynamics of individual proteins. The average uptake at different time points for selected regions including residues 34-50 and 152-180 were included in **Figure S12** in the Supporting Information. The individual peptide mapping at all HDX time points is given in **Figure S13** in the Supporting Information**.** The deuterium uptakes at 30 min HDX time showed the most pronounced differences between different proteins, which were chosen to illustrate the key structural features in the main figure panel (**Figure 5**).

HDX-MS analysis revealed subtle structural dynamics of the individual proteins. The most significant difference between the AL and MM-LCs is observed for residues 34-50, which are part of both the VL-VL dimerization interface and, more importantly in the context of this work, the CL-VL interface. These residues show significantly higher deuterium uptakes in all H-proteins, with H3 being the highest, implying that AL-LCs dimeric interfaces (VL-VL and CL-VL) are more dynamic and hence significantly more destabilized than in M10 (**Figure 5, Figure S12** in the Supporting Information). The highly dynamic VL-VL interface of H3 also correlates well with its open VL-VL interface in a crystal structure (PDB 8P89), which houses two nanobodies interacting with each VL in a dimeric structure (28). On the other hand, residues 54-70, which are not part of either interface, show a higher deuterium uptake and hence more dynamics in M10 protein, which may be a result of redistribution of dynamics in the regions away from the rigid VL-VL interface to stabilize the overall structure as also observed for other proteins previously (41,42) (**Figure 5, Figure S12**). In contrast, the VL-CL hinge regions (residues 110-120) show homogenously high flexibility in all the proteins except H7 indicating their higher accessible surface areas (**Figure 5**). As expected, the CL domains from residues 170-200 also show a similar pattern of average deuterium uptakes and hence flexibility in AL- and MM-LCs, with a minor difference contributed by the rigid VL-CL interface containing residues 165-180 (**Figures S8B-S11B, S12** in the Supporting Information). This region shows significantly less deuterium uptake (rigid) in both AL and M proteins when compared to other peptides in the CL domain (**Figures S8B-S11B** in the Supporting Information). However, comparing the average uptake for this region (residue 152-180) between AL and M proteins shows that H3 and AL55 have higher uptake than M10. In contrast, H7 is an exception with the lowest deuterium uptake in this region (**Figure S12 Panel 152-180** in the Supporting Information). These data are particularly interesting in the light of our simulations. The dimeric conformations identified in the H state, **Figure 3**, are characterized by higher accessibility for the CL-VL interface, which is in agreement with the increased accessibility for the region 34-50 on the VL and 159-180 in the CL observed in the HDX-MS analysis. Notably, the H state of H7 is the only one in which the CL-CL interface is remarkably compact (see **Figure S3** in the Supporting Information), consistent with the lower HDX for residues 152-180 observed in H7. Overall, the HDX-MS data provide an independent validation of the H-state predicted from our conformational ensembles

Discussion

Understanding the molecular determinants of AL amyloidosis has been hampered by its high sequence variability in contrast to its highly conserved three-dimensional structure (8). In this work, building on our previous studies highlighting susceptibility to proteolysis as a property that can discriminate between AL-LC and MM-LC, as well as the role of conformational dynamics in protein aggregation, we characterized LC conformational dynamics under the assumption that

AL-LC proteins, despite their sequence diversity, may share a property that emerges at the level of their dynamics. We combined SAXS measurements with MD simulations under the integrative framework of Metainference to generate conformational ensembles representing the native state conformational dynamics of 4 AL-LC and 2 MM-LC. While SAXS alone already indicated possible differences, its combination with MD allowed us to observe a possible low-populated state, which we refer to as state H, characterized by well-separated VL and CL dimers and a perturbed CL-CL interface, which is significantly populated in AL-LCs while only marginally populated in MM-LCs. Notably, high-energy states associated with amyloidogenic proteins have been previously identified in the case of SH3 (43) and β2m (32). Here, HDX measurements allowed us to independently validate this state by observing increased accessibility in CL-VL interface regions. Importantly, given the limited size of our LC set, we observed in reference (44) that a comparison of lambda-6 LCs by HDX-MS revealed differences between AL and MM LCs in a region overlapping with the one reported here (residues 35–50). Specifically, this region exhibited increased deuterium uptake in AL compared to MM proteins. A T32N substitution was shown to mitigate aggregation propensity and reduce deuterium uptake in the AL protein, while the reverse substitution, N32T, applied to the germline sequence, increased it (44). Finally, the H-state observed here is comparable to that previously observed for an aggregation-prone mutation in a monomeric kappa LC (25).

Having established a conformational fingerprint for AL-LC proteins, it would be tempting to identify possible mutations that could be associated with the presence of the H state. Comparing the sequences and structures of M7 and H18, both of which belong to the *IGLV3-19*01* germline, we can identify a single mutation, A40G, that could easily be associated with the appearance of the H state in H18. This mutation is located in the 37-43 loop, which H/D exchange showed to be more accessible in our three AL-LCs than in our MM-LC (see **Figure 5 and Figure S12**), and it breaks a hydrophobic interaction with the methyl group of T165, as observed in the crystal structure of M7 (PDB 5MVG and **Figure S13**), potentially making T165 more accessible in H18 than in M7 (see **Figure 5, and Figure S12**). Comparing the H18 and M7 sequences with the germline reference sequence, we see that position 40 in *IGLV3-19*01* is a glycine (see **Figure S13**). This would suggest the intriguing interpretation that the G40A mutation in M7 may increase the interdomain stability compared to the germline sequence, making it less susceptible to aggregation. However, it should also be noted that while this framework position is a glycine in H3, H7, and AL55, it is also a glycine in M10. Previous research has often focused on identifying, on a case-by-case basis, the key mutations that may be considered responsible for the emergence of the aggregation propensity, under the assumption that such aggregation propensity should not be present in germline sequences, but this assumption may be misleading given the observation that few germlines are strongly overrepresented in AL, suggesting that these starting germline sequences may be inherently more aggregation-prone than the germline genes that are absent or rarely found in AL patients. More generally, by comparing our AL-LC sequences with their germline references (**Figures S14 to S19** in the Supporting Information), we observe that all mutations fall exclusively in the variable domain, allowing us to exclude for these systems a direct role for residues in the linker region, as observed in ref (25,45), or in the constant domain, as observed in ref (46). Many mutations fall in the CDR regions, as expected, but others are found in the framework regions, both near the dimerization interface and in other regions of the protein. Somatic mutations in the dimerization interfaces have been identified previously as possibly responsible for toxicity of AL-LC (47). Regarding mutations in the CDRs, it has been suggested that AL-LC proteins may exhibit frustrated CDR2 and CDR3 loops, with few key residues populating the left-hand alpha helix or other high-energy conformations (48), resulting in the destabilization of the VL. In **Figures S14 to S19** in the Supporting Information, we have analyzed the Ramachandran plot obtained from our conformational ensembles, focusing only on those residues that most populate the left-hand alpha helix region, which are marked with a red circle symbol and whose Ramachandran is reported. Our data indicate the presence of residues populating left-hand alpha regions in the Ramachandran plot, but these are also found in the case of MM-LCs, so our simulations do not allow to confirm or exclude this mechanism in our set of protein systems.

In conclusion, our study provides a novel, complementary, perspective on the determinants of the misfolding propensity of AL-LCs that we schematize in Figure 6. The identification of a high-energy state, with perturbed CL dimerization interfaces, extended linkers, and accessible regions in both the VL-CL and VL-VL interfaces may be the common feature interplaying with specific properties shown by previous work including the direct or indirect destabilization of both the VL-VL and CL-CL dimerization interfaces (22,23,46-50). Our conformational fingerprint is also consistent with the observation that protein stability does not fully correlate with the tendency to aggregate, whereas susceptibility to proteolysis and conformational dynamics may better to capture the differences between AL-LC and MM-LC. In this context, our data allow us to rationally suggest that targeting the constant domain region at the CL-VL interface, which is more labile in the H state, maybe a novel strategy to search for molecules against LC aggregation in AL amyloidosis.

Materials and Methods

LC production and purification. Recombinant AL- (H3, H7, H18, and AL5) and M- (M7, M10) proteins were produced and purified from the host *E. coli* strain BL21(DE3). Firstly, the competent BL21(DE3) cells were transformed with plasmid pET21(b+), which contains genes encoding H3, H7, H18, AL55, M7, and M10 proteins. The transformed cells were selected for each plasmid by growing them on LB agar plates containing the antibiotic ampicillin at a final concentration of 100 µg/ml. For over-expression of protein, one colony was picked from each plate and grown overnight in 20 ml of LB broth containing ampicillin at a final concentration of 100 µg/ml. The overnight-grown cells were then used to inoculate a secondary culture in one liter of LB broth. The cells were grown until the turbidity (OD_{600nm}) reached between 0.6-0.8 and protein expression was subsequently induced by adding 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 4 h. The bacterial cells containing overexpressed LCs were then harvested using a Backman Coulter centrifuge at 6000 rpm for 20 min at 4 °C. All the proteins were overexpressed as inclusion bodies. For protein purification, the inclusion bodies were isolated by cell lysis induced by sonication. The purification of inclusion bodies was performed by washing them with buffer containing 10 mM Tris (pH 8) and 1% triton X 100. The purified inclusion bodies were unfolded with buffer containing 6.0 M guanidinium hydrochloride (GdnHCl) for 4h at 4 °C. The unfolded LCs were then refolded in a buffer containing reduced and oxidized glutathione to assist in disulfide bond formation. The refolded proteins were subjected to anion exchange and size-exclusion chromatography steps for final purification. The level of protein purity was checked on 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels. The final protein concentration was measured using molecular weight and extinction coefficient of individual proteins. The purified proteins were stored at -20 °C for further use. Size exclusion coupled multi-angle light scattering (SEC-MALS) confirmed the dimeric assembly of purified proteins (**Figure S20**) which were used for SAXS and HDX-MS experiments mentioned below.

Small-angle X-ray scattering (SAXS). For SAXS analysis, H3 was diluted to 3.4 mg/mL, H7 was diluted to 3.4 mg/mL, H18 was diluted to 2.8 mg/mL, AL55 was diluted to 2.6 mg/mL, M7 was diluted to 3.6 mg/mL, in 20 mM TrisHCl, 150 mM NaCl, pH 8. H3, H7 and M7 batch data were collected at the P12 BioSAXS beamline of the EMBL Hamburg Synchrotron (52), while AL55 batch data and H18 and M10 SEC data were collected at the BM29 BioSAXS beamline of the ESRF, Grenoble (53). For SEC-SAXS, H18 and M10 were injected into a superdex 200 increase 10/300 GL column previously equilibrated in 20 mM TrisHCl, 150 mM NaCl, pH 8, at a concentration of 2.8 mg/mL and 6.7 mg/mL, respectively, see also **Figure S21** in the Supporting Information. SAXS data were processed using programs PRIMUS and GNOM within the ATSAS package (54). Data are deposited in the SASBDB (55) and available with accession codes SASDVL4, SASDVM4, SASDVN4, SASDVK4, SASDVP4, and SASDVQ4 (cf. **Dataset S1**).

Molecular dynamics simulations. The available crystallographic structures of H3, H7 and M7 (PDB: 5mtl, 5muh and 5mvg, respectively (12)) were used as starting conformations, using Modeller to add missing residues (56). H18 and AL55 were modelled by homology modelling

using SwissModel (57), while M10 was modelled using AF2 (58). Simulations were performed using GROMACS 2019 (59) and the PLUMED2 software (60-62), using AMBER-DES force field and TIP4P-D water (63,64). During in-vacuum minimization RMSD-restraints were imposed to enhance the symmetry between the two constant and the two variable domains. The systems were solvated in a periodic dodecahedron box, initially 1.2 nm larger than the protein in each direction, neutralized with Na and Cl ions to reach a salt concentration of 10 mM, then minimized and equilibrated at the temperature of 310 K and pressure of 1 atm using the Berendsen thermostat and barostat. Two independent 900 ns long plain MD simulations were run to generate reliable and independent starting conformations for the Metadynamics Metainference simulations (34,35,65). 30 conformations were extracted from each simulation and duplicated by inverting the two chains, to obtain 60 starting conformations symmetrically distributed with respect to chain inversion.

Metadynamics Metainference production simulations were run in duplicate using 60 replicas, each replica evolved for ~1 µs (cf. **Table 2**). Simulations were performed in the NPT ensemble maintaining the temperature at 310 K with the Bussi thermostat (66) and the pressure of 1 atm with the Parrinello-Rahman barostat (67); the electrostatic was treated by using the particle mesh Ewald scheme with a short-range cut-off of 0.9 nm and van der Waals interaction cut-off was set to 0.9 nm. To reduce the computational cost, the hydrogen mass repartitioning scheme was used (68): the mass of heavy atoms was repartitioned into the bonded hydrogen atoms using the heavyh flag in the *pdb2gmx* tool, the LINCS algorithm was used to constraint all bonds, allowing to use a time step of 5 fs. In these simulations Parallel Bias Metadynamics (69) was used to enhance the sampling, combined with well-tempered metadynamics and the multiple-walker scheme, where Gaussians with an initial height of 1.0 kJ/mol were deposited every 0.5 ps, using a bias factor of 10. Five CVs were biased, including combinations of phi/psi dihedral angles of the linker regions (i.e. residues connecting variable and constant domains) in the two chains, combinations of chi dihedral angles of the linker regions in the two chains, combination of inter-domain contacts between the variable and the constant domains. The width of the Gaussians was 0.07, 0.12 and 120 for the combination of phi/psi, of chi dihedral angles and combination of contacts, respectively. Metainference was used to include SAXS restraints, using the hySAXS hybrid approach described in (36,37,70). A set of 13 representative SAXS intensities at different scattering angles, ranging between 0.015 A^{-1} and 0.25 A^{-1} and equally spaced, was used as restraints. These intensities were extracted from experimental data, after performing regularization with the Distance Distribution tool of Primus, based on Gnom (54). Metainference was applied every 5 steps, using a single Gaussian noise per data point and sampling a scaling factor between experimental and calculated SAXS intensities with a flat prior between 0.5 and 1.5. The aggregate sampling from the 60 replicas was reweighted using the final metadynamics bias to obtain a conformational ensemble where each conformation has an associated statistical weight (71). Convergence and error estimates were assessed by the inspection of the two replicated metadynamics metainference run. SAXS data were then recalculated using crysol (54). All relevant data are available on Zenodo (cf**. Dataset S2**).

Hydrogen-deuterium mass exchange spectrometry (HDX-MS). SYNAPT G2-HDMS system (Waters Corporation, USA) equipped with a LEAP robotic liquid handler was used to perform HDX-MS measurements in a fully automated mode as described previously (41,42,72,73). The data collection was carried out by a 20-fold dilution of H3, H7, AL55, and M10 proteins (100 µM) with the labeling buffer 1X phosphate buffer saline (PBS) prepared in D_2O (pD 7.4) to trigger HDX for 0, 0.5, 1,10, 30, 120, and 240 min at 25 °C in technical triplicates. Each reaction was quenched by mixing the labeled protein with quench buffer (50 mM sodium phosphate, 250 mM TCEP, 3.0 M GdnHCl (pH 2)) in a 1:1 ratio at 0 °C. Online digestion was then performed using an immobilized pepsin digestion column (Waters Enzymate BEH Pepsin, 2.1 x 30 mm). The digested peptides were trapped using a C18 trapping column (Acquity BEH VanGuard 1.7 µm, 2.1 x 5.0 mm) and separated by a linear acetonitrile gradient of 5 to 40%. Protein Lynx Global Server (PLGS), and DynamX (Waters Corporation, USA) were used to identify the individual peptides, and subsequently, data processing using parameters: maximum peptide length of 25, the minimum intensity of 1000; minimum ion per amino acid of 0.1; maximum MH+ error of 5 ppm

and a file threshold of three. A reference molecule [(Glu1)-fibrinopeptide B human (CAS No 103213-49-6, Merck, USA)] was used to lock mass with an expected molecular weight of 785.8426 Da. The obtained peptide coverage of H3, H7, AL55, and M10 were 98.6%, 92.5%, 98.6%, and 99.1%, respectively, with a redundancy > 4.0. Backbone amide groups exhibited a relative deuterium uptake (with respect to the zero exchange time data) of 0 to 30% within 4 h of exchange time. The relative deuterium uptake data at different HDX times were then used to generate heat maps for each amino acid. The obtained data was mapped on individual protein structures in a gradient of blue-white-red showing 0 to 30% of uptake, respectively (74). Red denotes the dynamic peptides while peptides colors in blue are rigid. The common peptide analysis between different model LCs was not performed in this analysis as they all generate different peptides both due to sequence heterogeneity in LCs especially in the VL domain and non-specific cleavage of pepsin enzyme used to generate peptides after deuterium exchanged for MS analysis. Therefore, all interpretation was done on individual relative exchange data.

Acknowledgments

We acknowledge Martin A. Schroer and Dmitri Svergun at EMBL Hamburg, and Sonia Longhi at AFMB Marseille, for discussion and support on the SAXS data acquisition and analysis. We acknowledge PRACE for awarding us access to Piz Daint at CSCS, Switzerland. The authors acknowledge CINECA for an award under the ISCRA initiative for the availability of high-performance computing resources and support. This work was supported by the Italian Ministry of Research PRIN 2020 (20207XLJB2), by CARIPLO/TELETHON Foundations (GJC23044) and by Italian Ministry of Health (Ricerca Finalizzata, grant #GR-2018-12368387). SP acknowledges Fondazione Veronesi for a postdoctoral fellowship. STDH is supported by Academia Sinica intramural fund, an Academia Sinica Career Development Award from Academia Sinica (AS-CDA-109- L08), and the National Science and Technology Council (NSTC), Taiwan (NSTC113-2123-M-001-010-, 110-2113-M-001-050-MY3 and 110-2311-B-001-013-MY3). We also acknowledge Dr. Min-Feng Karen Hsu, and Mr. Yong-Sheng Wang for the protein quality controls, and Dr. Shu-Yu Lin, and Mr. Ming-Jie Tsai of the Academia Sinica Common Mass Spectrometry Facilities (AS-CFII-111-209), funded by the Academia Sinica Core Facility and Innovative Instrument Project, for supporting the HDX-MS experiments.

References

1. G. Merlini, *et al.*, Systemic immunoglobulin light chain amyloidosis. *Nat Rev Dis Primers* **4**, 38 (2018).
2. L. M. Blancas-Mejia, *et al.*, Immunoglobulin light chain amyloid aggregation. *Chem. Commun.* **54**, 10664–10674 (2018).
3. P. Cascino, *et al.*, SINGLE-MOLECULE REAL-TIME sequencing of the M protein: Toward personalized medicine in monoclonal gammopathies. *American J Hematol* **97** (2022).
4. T. L. Poshusta, *et al.*, Mutations in specific structural regions of immunoglobulin light chains are associated with free light chain levels in patients with AL amyloidosis. *PLoS ONE* **4**, e5169 (2009).
5. V. Perfetti, *et al.*, The repertoire of λ light chains causing predominant amyloid heart involvement and identification of a preferentially involved germline gene, IGLV1-44. *Blood* **119**, 144–150 (2012).
6. T. V. Kourelis, *et al.*, Clarifying immunoglobulin gene usage in systemic and localized immunoglobulin light-chain amyloidosis by mass spectrometry. *Blood* **129**, 299–306 (2017).
7. R. L. Comenzo, Y. Zhang, C. Martinez, K. Osman, G. A. Herrera, The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig VL germ line gene use and clonal plasma cell burden. *Blood* **98**, 714–720 (2001).
8. R. M. Absmeier, G. J. Rottenaicher, H. L. Svilenov, P. Kazman, J. Buchner, Antibodies gone bad – the molecular mechanism of light chain amyloidosis. *The FEBS Journal* **290**, 1398–1419 (2023).

9. P. C. Bourne, *et al.*, Three-dimensional structure of an immunoglobulin light-chain dimer with amyloidogenic properties. *Acta Crystallogr D Biol Crystallogr* **58**, 815–823 (2002).
10. M. L. Chiu, D. R. Goulet, A. Teplyakov, G. L. Gilliland, Antibody structure and function: The basis for engineering therapeutics. *Antibodies* **8**, 55 (2019).
11. J. M. Di Noia, M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
12. L. Oberti, *et al.*, Concurrent structural and biophysical traits link with immunoglobulin light chains amyloid propensity. *Sci Rep* **7**, 16809 (2017).
13. L. Radamaker, *et al.*, Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. *Nat Commun* **12**, 875 (2021).
14. L. Radamaker, *et al.*, Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM. *Nat Commun* **12**, 6434 (2021).
15. L. Radamaker, *et al.*, Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nat Commun* **10**, 1103 (2019).
16. P. Swuec, *et al.*, Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nat Commun* **10**, 1269 (2019).
17. S. Puri, *et al.*, The cryo-EM structure of renal amyloid fibril suggests structurally homogeneous multiorgan aggregation in AL amyloidosis. *Journal of Molecular Biology* **435**, 168215 (2023).
18. S. Ricagno, *et al.*, Helical superstructures between amyloid and collagen VI in heart-derived fibrils from a patient with Light Chain Amyloidosis. [Preprint] (2023). Available at: https://www.researchsquare.com/article/rs-3625869/v1 [Accessed 16 April 2024].
19. F. Lavatelli, *et al.*, Mass spectrometry characterization of light chain fragmentation sites in cardiac AL amyloidosis: insights into the timing of proteolysis. *Journal of Biological Chemistry* **295**, 16572–16584 (2020).
20. G. Mazzini, *et al.*, Protease-sensitive regions in amyloid light chains: what a common pattern of fragmentation across organs suggests about aggregation. *The FEBS Journal* **289**, 494–506 (2022).
21. S. Dasari, *et al.*, Proteomic detection of immunoglobulin light chain variable region peptides from amyloidosis patient biopsies. *J. Proteome Res.* **14**, 1957–1967 (2015).
22. G. J. Rottenaicher, *et al.*, Molecular mechanism of amyloidogenic mutations in hypervariable regions of antibody light chains. *Journal of Biological Chemistry* **296**, 100334 (2021).
23. E. Rennella, G. J. Morgan, J. W. Kelly, L. E. Kay, Role of domain interactions in the aggregation of full-length immunoglobulin light chains. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 854–863 (2019).
24. E. S. Klimtchuk, *et al.*, Role of complementarity-determining regions 1 and 3 in pathologic amyloid formation by human immunoglobulin κ1 light chains. *Amyloid* **30**, 364–378 (2023).
25. B. Weber, *et al.*, The antibody light-chain linker regulates domain orientation and amyloidogenicity. *Journal of Molecular Biology* **430**, 4925–4940 (2018).
26. X. Sun, H. J. Dyson, P. E. Wright, Role of conformational dynamics in pathogenic protein aggregation. *Current Opinion in Chemical Biology* **73**, 102280 (2023).
27. M. Maritan, *et al.*, Inherent biophysical properties modulate the toxicity of soluble amyloidogenic light chains. *Journal of Molecular Biology* **432**, 845–860 (2020).
28. L. Broggini, *et al.*, Nanobodies counteract the toxicity of an amyloidogenic light chain by stabilizing a partially open dimeric conformation. *Journal of Molecular Biology* **435**, 168320 (2023).
29. L. Visconti, *et al.*, Investigating the molecular basis of the aggregation propensity of the pathological D76N mutant of beta-2 microglobulin: Role of the denatured state. *IJMS* **20**, 396 (2019).
30. B. M. Sala, *et al.*, Conformational stability and dynamics in crystals recapitulate protein behavior in solution. *Biophysical Journal* **119**, 978–988 (2020).
31. C. Camilloni, *et al.*, Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. *Sci Rep* **6**, 25559 (2016).

32. T. Le Marchand, *et al.*, Conformational dynamics in crystals reveal the molecular bases for D76N beta-2 microglobulin aggregation propensity. *Nat Commun* **9**, 1658 (2018).
33. A. Achour, *et al.*, Biochemical and biophysical comparison of human and mouse beta-2 microglobulin reveals the molecular determinants of low amyloid propensity. *The FEBS Journal* **287**, 546–560 (2020).
34. M. Bonomi, C. Camilloni, M. Vendruscolo, Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics. *Sci Rep* **6**, 31232 (2016).
35. M. Bonomi, C. Camilloni, A. Cavalli, M. Vendruscolo, Metainference: A Bayesian inference method for heterogeneous systems. *Sci. Adv.* **2**, e1501177 (2016).
36. C. Paissoni, A. Jussupow, C. Camilloni, Determination of protein structural ensembles by hybrid-resolution SAXS restrained molecular dynamics. *J. Chem. Theory Comput.* **16**, 2825–2834 (2020).
37. C. Paissoni, C. Camilloni, How to determine accurate conformational ensembles by metadynamics metainference: A chignolin study case. *Front. Mol. Biosci.* **8**, 694130 (2021).
38. D. Saad, *et al.*, High conformational flexibility of the E2F1/DP1/DNA complex. *Journal of Molecular Biology* **433**, 167119 (2021).
39. M. C. Ahmed, *et al.*, Refinement of α-synuclein ensembles against SAXS data: Comparison of force fields and methods. *Front. Mol. Biosci.* **8**, 654333 (2021).
40. F. E. Thomasen, K. Lindorff-Larsen, Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins. *Biochemical Society Transactions* **50**, 541–554 (2022).
41. S. Puri, S.-T. D. Hsu, Oxidation of catalytic cysteine of human deubiquitinase BAP1 triggers misfolding and aggregation in addition to functional loss. *Biochemical and Biophysical Research Communications* **599**, 57–62 (2022).
42. K.-T. Ko, I.-C. Hu, K.-F. Huang, P.-C. Lyu, S.-T. D. Hsu, Untying a knotted SPOUT RNA methyltransferase by circular permutation results in a domain-swapped dimer. *Structure* **27**, 1224-1233.e4 (2019).
43. P. Neudecker, *et al.*, Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366 (2012).
44. Peterle, et al., A Conservative Point Mutation in a Dynamic Antigen-binding Loop of Human Immunoglobulin λ6 Light Chain Promotes Pathologic Amyloid Formation. *Journal of Molecular Biology* 3, 167310 (2021)
45. C. N. Nokwe, *et al.*, The antibody light-chain linker is important for domain stability and amyloid formation. *Journal of Molecular Biology* **427**, 3572–3586 (2015).
46. G. J. Rottenaicher, R. M. Absmeier, L. Meier, M. Zacharias, J. Buchner, A constant domain mutation in a patient-derived antibody light chain reveals principles of AL amyloidosis. *Commun Biol* **6**, 209 (2023).
47. M. Garofalo, *et al.,* Machine learning analyses of antibody somatic mutations predict immunoglobulin light chain toxicity. *Nat Commun* **12**, 3532 (2021).
48. T. Pradhan, *et al.*, Mechanistic insights into the aggregation pathway of the patient-derived immunoglobulin light chain variable domain protein FOR005. *Nat Commun* **14**, 3755 (2023).
49. F. C. Peterson, E. M. Baden, B. A. L. Owen, B. F. Volkman, M. Ramirez-Alvarado, A single mutation promotes amyloidogenicity through a highly promiscuous dimer interface. *Structure* **18**, 563–570 (2010).
50. P. Kazman, *et al.*, Fatal amyloid formation in a patient's antibody light chain is caused by a single point mutation. *eLife* **9**, e52300 (2020).
51. B. Al-Lazikani, A. M. Lesk, C. Chothia, Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology* **273**, 927–948 (1997).
52. C. E. Blanchet, *et al.,* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr*. **48**(Pt 2):431-443 (2015).
53. P. Pernot, *et al.,* New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *J. Phys.: Conf. Ser*. **247**, 012009 (2010).

54. K. Manalastas-Cantos, *et al.*, ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Cryst*. **54**, 343-355 (2021).
55. E. Valentini, *et al.,* SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Research* **43**(D1), D357–D363 (2015).
56. B. Webb, A. Sali. Comparative protein structure modeling using modeller. Current protocols in bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.
57. A. Waterhouse, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46, W296-W303. (2018)
58. J. Jumper, *et al.,* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
59. M. J. Abraham, et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Softwarex* **1–2,** 19–25 (2015).
60. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613 (2014).
61. Bonomi, M., Camilloni, C., Integrative structural and dynamical biology with PLUMED-ISDB. *Bioinformatics* **33,** 3999–4000 (2017).
62. Bonomi, M., *et al.,* Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16,** 670–673 (2019).
63. S. Piana, P. Robustelli, D. Tan, S. Chen, D. E. Shaw, Development of a force field for the simulation of single-chain proteins and protein-protein complexes. *J Chem Theory Comput* (2020).
64. S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* **119,** 5113–5123 (2015).
65. Löhr, T., Jussupow, A., Camilloni, C., Metadynamic metainference: Convergence towards force field independent structural ensembles of a disordered peptide. *J. Chem. Phys.* **146,** 165102 (2017).
66. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J Chem Phys* **126,** 014101 (2007).
67. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52,** 7182–7190 (1981).
68. C. W. Hopkins, S. L. Grand, R. C. Walker, A. E. Roitberg, Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput* **11,** 1864–1874 (2015).
69. J. Pfaendtner, M. Bonomi, Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J Chem Theory Comput* **11,** 5062–7 (2015).
70. F. Ballabio, C. Paissoni, M. Bollati, M. de Rosa, R. Capelli, C. Camilloni, Accurate and efficient SAXS/SANS implementation including solvation layer effects suitable for molecular simulations. *J. Chem. Theory Comput.* **19,** 8401–8413 (2023).
71. D. Branduardi, G. Bussi, M. Parrinello, M. Metadynamics with adaptive gaussians. *J Chem Theory Comput* **8,** 2247–54 (2012).
72. G. R. Masson, *et al.*, Recommendations for performing, interpreting and reporting hydrogen-deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat Methods* **16**, 595–602 (2019).
73. S. Puri, *et al.*, Impacts of cancer-associated mutations on the structure–activity relationship of BAP1. *Journal of Molecular Biology* **434**, 167553 (2022).
74. The PyMOL molecular graphics system, Version 1.2, *Schrödinger, LLC.*

Figures and Tables

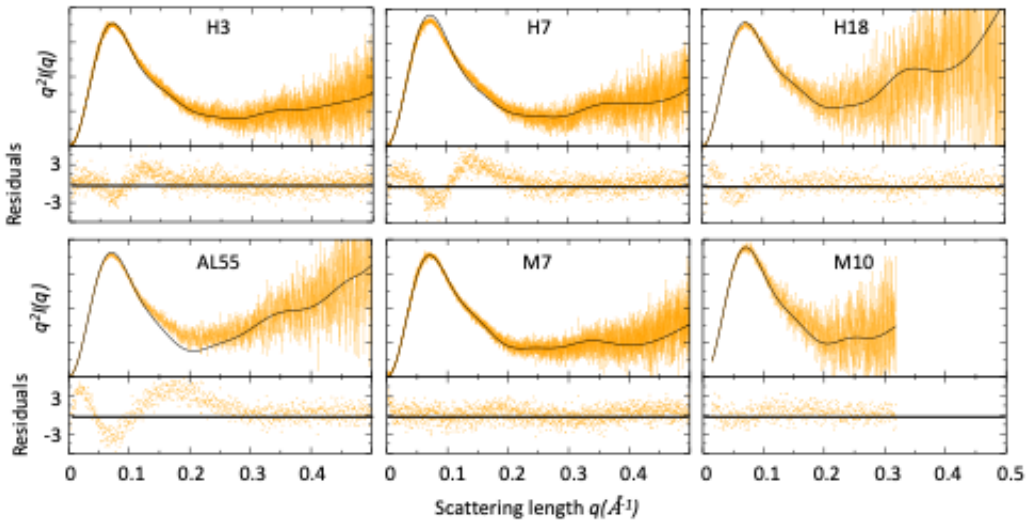

Figure 1. Comparison of SAXS data and single light chain structures. Kratky plots of for AL- and MM light chains. The experimental (orange) and theoretical (black) curves (top panels) and the associated residuals (bottom panels) indicate that AL-LC solution behavior deviates from reference structures more than MM-LC. SAXS was measured as follow: H3 measured in bulk (Hamburg), 3.4 mg/ml; H7 measured in bulk (Hamburg), 3.4 mg/ml; H18 measured by SEC-SAXS (ESRF) with the injection concentration of at 2.8 mg/ml; AL55 measured in bulk (ESRF), 2.6 mg/ml; M7 measured in bulk (Hamburg), 3.6 mg/ml; M10 measured by SEC-SAXS with the injection concentration of 6.7 mg/ml (ESRF). Theoretical SAXS curves were calculated using crysol (54). Log-log plots are shown in Fig. S1 in the Supporting Information.

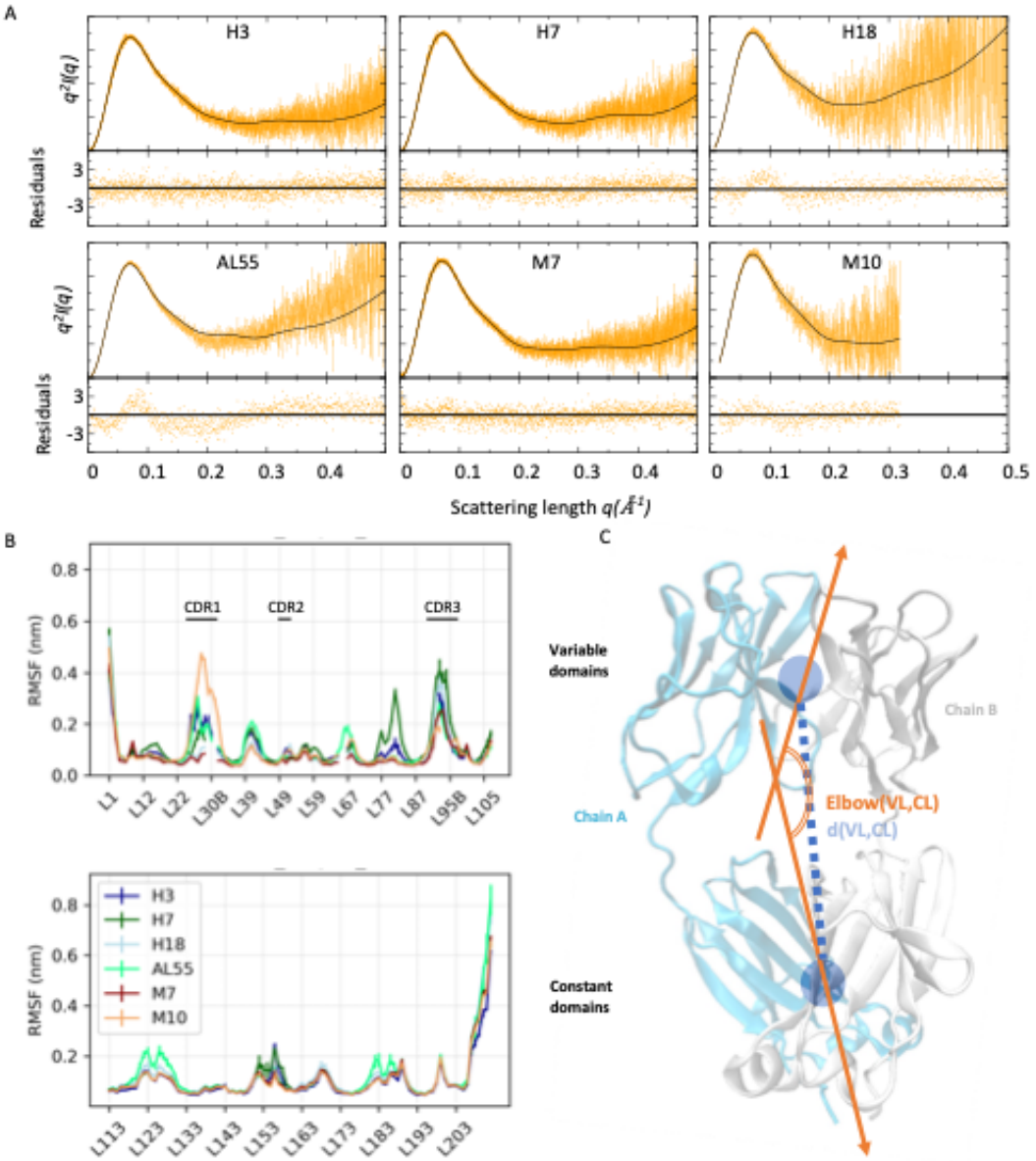

Figure 2. Light chain SAXS driven MD simulations. (A) Kratky plots and associated residuals (bottom panels) comparing experimental (orange), and theoretical (black) curves obtained by averaging over the metainference ensemble for H3, H7, H18, AL55, M7 and M10, respectively. Theoretical SAXS curves were calculated using crysol (54). (B) Residue-wise root mean square fluctuations (RMSF) obtained by averaging the two Metainference replicates and the two equivalent domains for the six systems studied. The top panel shows data for the variable domains, while the bottom panel shows data for the constant domain. Residues are reported using *Chothia and Lesk* numbering (51). (C) Schematic representation of two global collective variables used to compare the conformational dynamics of the different systems, namely the distance between the center of mass of the VL and CL dimers and the angle describing the bending of the two domain dimers.

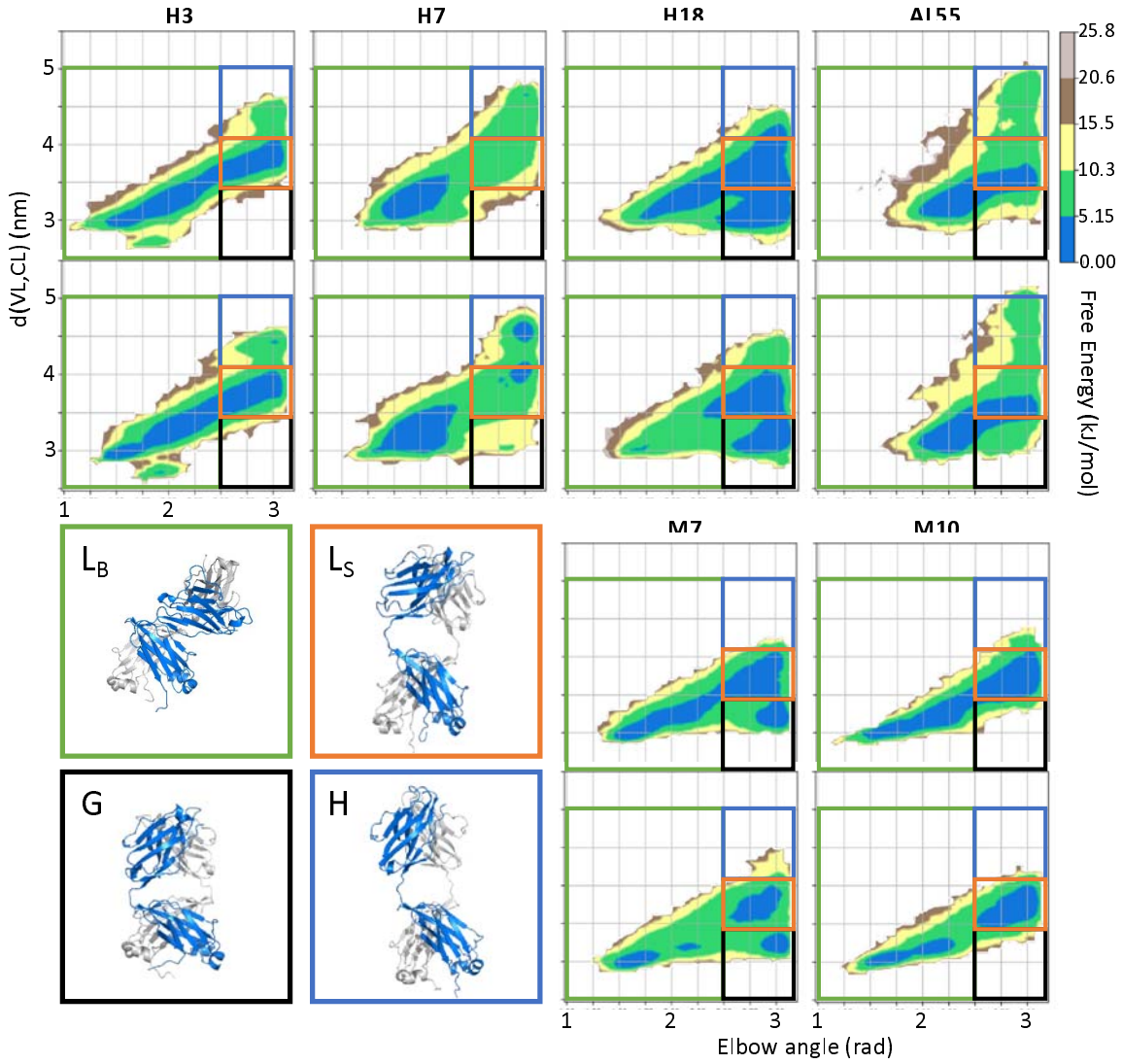

Figure 3. Free Energy Surfaces (FESes) for the six light chain systems under study by Metadynamics Metainference MD simulations. For each system, the simulations are performed in duplicate. The x-axis represents the elbow angle indicating the relative bending of the constant and variable domains (in radians), while the y-axis represents the distance in nm between the center of mass of the CL and VL dimers. The free energy is shown with color and isolines every $2k_BT$ corresponding to 5.16 kJ/mol. On each FES are represented four regions (green L_B-state, red L_S-state, blue H-state, and black G-state) highlighting the main conformational states. For each region, a representative structure is reported in a rectangle of the same color.

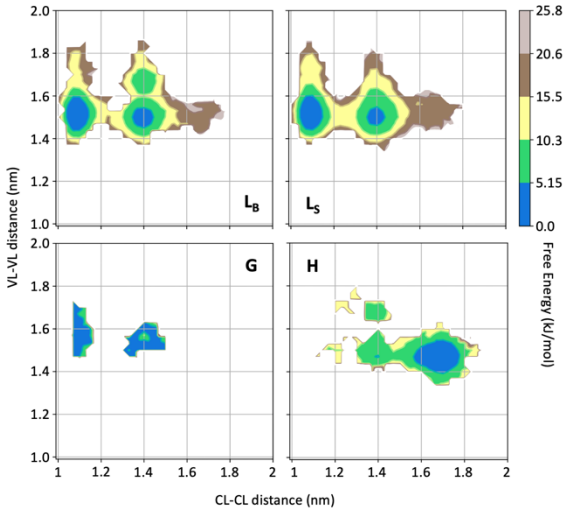

Figure 4. FESes for the four substates identified in Figure 3 in the case of the first H3 Metainference simulation. The x-axis shows the distance between the centers of mass of the constant domains, while the y-axis shows the distance between the centers of mass of the variable domains. The free energy is shown with color and isolines every $2k_BT$ corresponding to 5.16 kJ/mol.

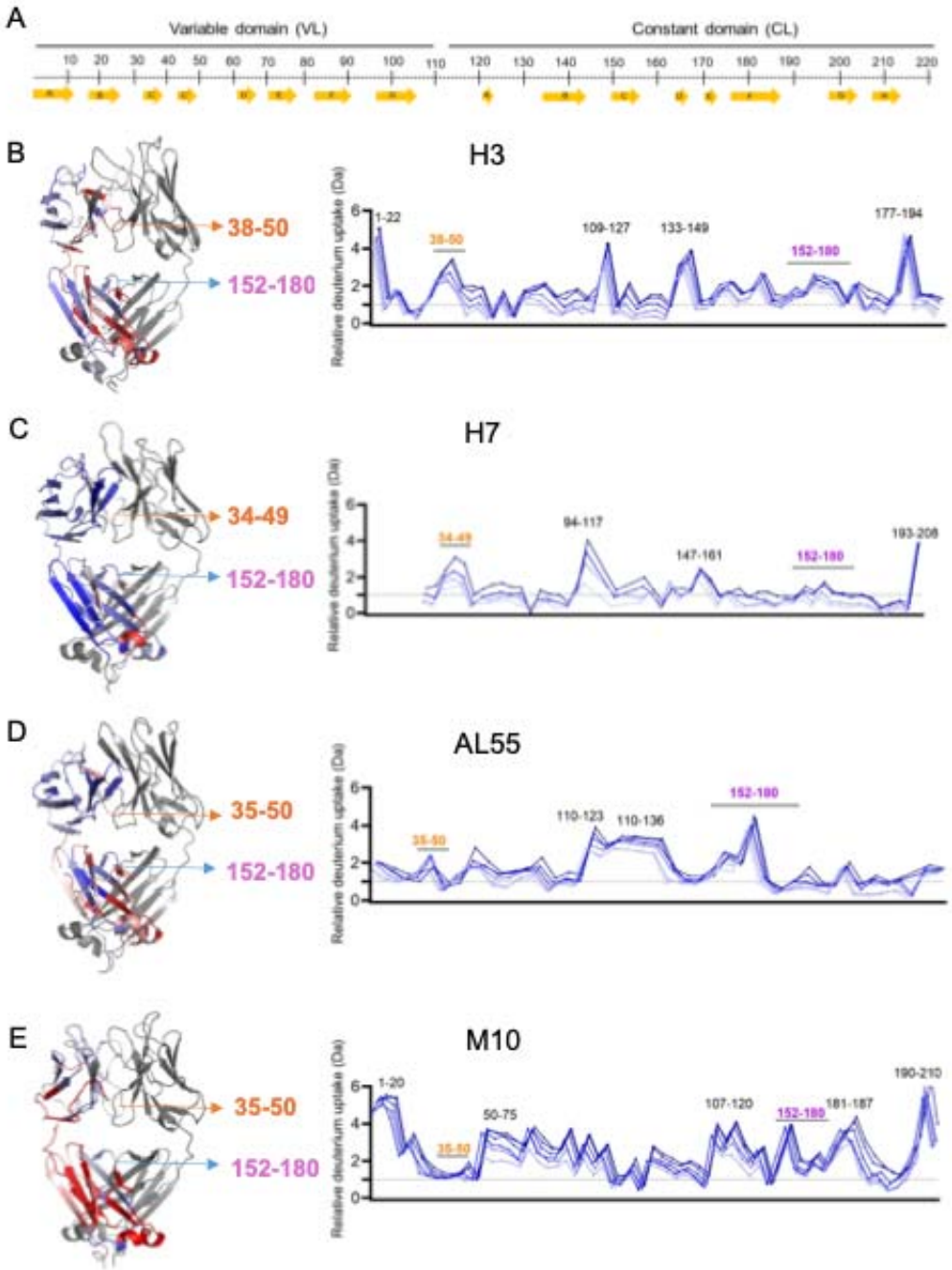

Figure 5. **HDX-MS analysis**. (A) The top panel represents the simplified presentation of the primary structure of an LC including variable domain (VL) and constant domain (CL). The location of β-strands according to *Chothia and Lesk* (51). (B) The structural mapping and butterfly plot of relative deuterium uptake (Da) of H3. The chain A of H3 structure is colored in a gradient of blue-white-red for an uptake of 0 to 30% at an exchange time of 30 min. The chain B is shown in gray (right hand panel). The butterfly plot showing relative deuterium uptake at all time points from 0.5 to 240 min on a gradient of light to dark blue (left-hand panel). (C, D, E) are the figures corresponding to H7, AL55, and M10, respectively with the same color coding as in figure B. The overall sequence coverage for all proteins was > 90% with a redundancy of > 4.0. The VL-VL domains interface peptides covering amino-acid residues 34-50 are labelled in orange while CL-CL interface region containing 152-180 amino acids are labelled in magenta. Collectively, they form VL-CL interface which is important to define H-state.

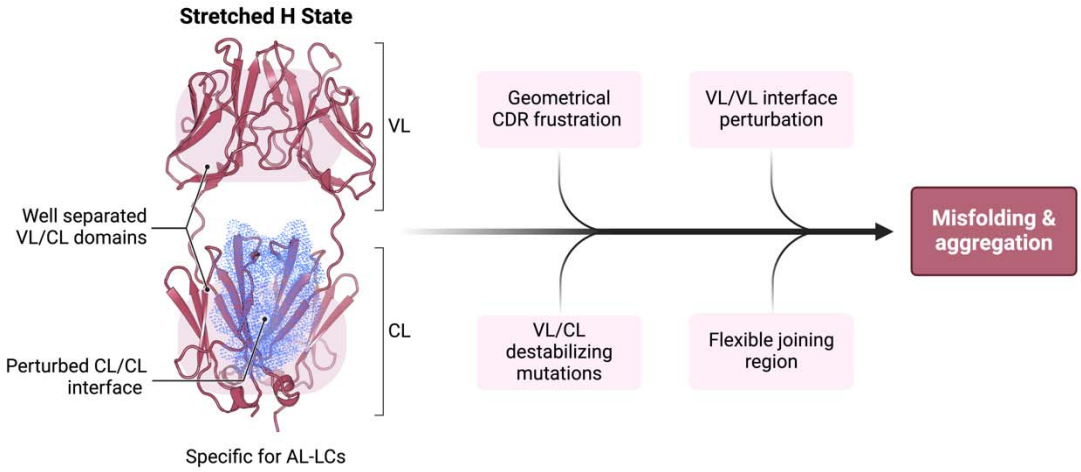

Figure 6. **Schematic representation summarizing our findings in the context of previous work on the biophysical properties of amyloidogenic light chains.** We propose that the H state is the conformational fingerprint distinguishing AL LCs from other LCs, which together with other features contributes to the amyloidogenicity of AL LCs.

Table 1. LC systems studied in this work. The table includes information about the germline, phenotype, method to obtain structure, the SAXS curves, and the radius of gyration derived from the SAXS data for all the model proteins studied in this work.

LC	Germline	Phenotype	Structure	SAXS [2] q<0.5	Rg (SAXS)

				Å (q<0.3)	[nm]
H3	*IGLV**1**-44*01*	AL	5MTL	1.6 (1.9)	2.57 ± 0.02
H7	*IGLV**1**-51*01*	AL	5MUH	2.8 (4.0)	2.56 ± 0.02
H18	*IGLV**3**-19*01*	AL	*Homology*	1.6 (1.9)	2.56 ± 0.01
AL55	*IGLV**6**-57*02*	AL	*Homology*	5.1 (7.8)	2.58 ± 0.04
M7	*IGLV**3**-19*01*	MM	5MVG	1.2 (1.2)	2.50 ± 0.02
M10	*IGLV**2**-14*03*	MM	*AF2*	1.2 (1.2)	2.51 ± 0.01

Table 2. Metainference simulations performed in this work for the 6 systems. For each metainference simulation is reported the simulation time per replica with the number of replicas; the chi-square of the resulting conformational ensemble with the experimental SAXS curve, the range q<0.3 Å is the one used as restraint in the simulation; the average radius of gyration with error estimated by block averaging.

LC code	Simulation	Length per replica [ns] (# replicas)	SAXS χ^2 q<0.5 Å (q<0.3)	Average Rg [nm]
H3	M&M 1	1,530 (60)	1.2 (1.1)	2.56 ± 0.02
	M&M 2	1,520 (60)	1.2 (1.1)	2.56 ± 0.02
H7	M&M 1	1,627 (60)	1.1 (1.2)	2.54 ± 0.04
	M&M 2	1,545 (60)	1.1 (1.2)	2.54 ± 0.05
H18	M&M 1	1,643 (60)	1.4 (1.6)	2.53 ± 0.03
	M&M 2	1,529 (60)	1.4 (1.7)	2.54 ± 0.03
AL55	M&M 1	1,545 (60)	2.7 (3.0)	2.59 ± 0.02
	M&M 2	1,591 (60)	2.5 (2.7)	2.58 ± 0.05
M7	M&M 1	1,623 (60)	1.2 (1.2)	2.52 ± 0.03
	M&M 2	1,530 (60)	1.1 (1.2)	2.52 ± 0.03
M10	M&M 1	987 (60)	1.1 (1.1)	2.53 ± 0.07
	M&M 2	995 (60)	1.1 (1.1)	2.54 ± 0.05

Table 3. Populations of the four states shown in Figure 3 resulting from the two independent Metadynamics Metainference simulations performed for each of the 6 LCs. The population of the H state, which we supposed to be a fingerprint specific for AL-LCs, is in bold.

%	H3	H7	H18	AL55	M7	M10
L_B	62.0±0.4	72.3±2.5	22.9±2.8	48.3±0.1	46.8±0.1	48.4±0.3
L_S	33.0±0.2	15.2±2.7	38.4±2.1	32.5±2.0	35.0±0.1	49.1±1.5
G	0.2±0.1	0.8±0.4	33.8±3.8	10.5±1.4	17.6±0.1	1.8±0.6
H	**4.8±0.5**	**11.7±0.5**	**5.0±1.0**	**8.7±0.5**	0.6±0.2	0.8±0.6