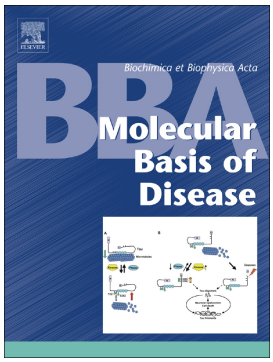

Structure and energetic basis of overrepresented λ light chain in systemic light chain amyloidosis patients

Jun Zhao, Baohong Zhang, Jianwei Zhu, Ruth Nussinov, Buyong Ma

PII: S0925-4439(17)30458-1
DOI: doi:10.1016/j.bbadis.2017.12.009
Reference: BBADIS 64985

To appear in:

Received date: 30 September 2017
Revised date: 29 November 2017
Accepted date: 4 December 2017

Please cite this article as: Jun Zhao, Baohong Zhang, Jianwei Zhu, Ruth Nussinov, Buyong Ma , Structure and energetic basis of overrepresented λ light chain in systemic light chain amyloidosis patients. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Bbadis(2017), doi:10.1016/j.bbadis.2017.12.009

Structure and energetic basis of overrepresented λ light chain in systemic light chain amyloidosis patients

Jun Zhao[1], Baohong Zhang[2], Jianwei Zhu[2,3], Ruth Nussinov[4,5], Buyong Ma[4*]

[1]Cancer and Inflammation Program, National Cancer Institute, Frederick, Maryland, 21702, USA

[2]School of Pharmacy, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China

[3]Jecho Laboratories, Inc. 7320A Executive Way, Frederick, Maryland, 21704, USA

[4]Basic Science Program, Leidos Biomedical Research, Inc., Cancer and Inflammation Program, National Cancer Institute, Frederick, Maryland, 21702, USA

[5]Sackler Inst. of Molecular Medicine Department of Human Genetics and Molecular Medicine Sackler School of Medicine, Tel Aviv University, Tel Aviv, 69978, Israel

* Corresponding author: E-mails: mabuyong@mail.nih.gov

Running title: Sequence and interface preference in the AL amyloidosis

Abstracts

Amyloid formation and deposition of immunoglobulin light-chain proteins in systemic amyloidosis (AL) cause major organ failures. While the κ light-chain is dominant (λ/κ=1:2) in healthy individuals, λ is highly overrepresented (λ/κ=3:1) in AL patients. The structural basis of the amyloid formation and the sequence preference are unknown. We examined the correlation between sequence and structural stability of dimeric variable domains of immunoglobulin light chains using molecular dynamics simulations of 24 representative dimer interfaces, followed by energy evaluation of conformational ensembles for 23 AL patients' light chain sequences. We identified a stable interface with displaced N-terminal residues, provides the structural basis for AL protein fibrils formation. Proline isomerization may cause the N-terminus to adopt amyloid-prone conformations. We found that λ light-chains prefer misfolded dimer conformation, while κ chain structures are stabilized by a natively folded dimer. Our study may facilitate structure-based small molecule and antibody design to inhibit AL.

Highlights:

1. λ is highly overrepresented in AL amyloid formation.

2. Simulation identified a stable interface with displaced N-terminal.

3. Energy and conformations of 23 AL patients' light chain structure were simulated.

4. κ AL protomers have higher unfolding energies and are stabilized by native dimer interface.

Keywords: systemic light chain amyloidosis; amyloid; molecular dynamics; energy landscape; protein misfolding; antibody mis-folding and aggregation

1. Introduction

As a protein misfolding disease, systemic light chain amyloidosis (AL) is characterized by extracellular deposition of immunoglobulin light chain (LC) aggregates of amyloid fibrils[1]. Without effective treatment, amyloid fibrils resulting from over-production [1] and abnormal somatic mutations in immunoglobulin light chains[2-4] cause heart, kidneys, liver, spleen, and peripheral nerves failures[5]. Both types of LC, κ and λ, are in the pathogenic aggregates[6-10]. While the κ chain is dominant (λ/κ=1:2) in healthy individuals, λ is highly overrepresented (λ/κ=3:1) in AL patients[11]. Full length LC contains two sub-domains, constant (C_L) and variable (V_L). V_L is found in amyloid deposits in over 80% of AL patients[12]. Here we ask (1) why are λ chains highly prone to amyloid formation; (2) what are the structural characteristic of the AL amyloid fibril; and (3) how can we prevent or reverse the light chains amyloid formation. Perturbation of stable folded protein appears the first step in misfolding and amyloid formation.

Most amyloid fibrils are made up of N-terminal fragments corresponding to the variable domain, indicating that proteolysis cuts off the constant domain and subsequent destabilization may trigger amyloid formation[13]. Like most amyloidogenic proteins, immunoglobulin light chains are sensitive to mutations, a major factor in the loss of stability and increased amyloidogenicity[14, 15]. Unlike other amyloidogenic proteins, every patient may have different LC sequence due to somatic recombination of immunoglobulins. *In vitro* aggregates suggest that proteins derived from patients with different light chain may follow different aggregation pathways[16]. Thus, finding a common mechanism for light chain amyloidosis within the framework of hundreds of sequences of both κ and λ chains is challenging[17]. Interestingly, somatic hypermutations on the complementarity determining regions (CDRs) of V_L do not significantly affect the monomeric structure of light chains, and the locations of these mutations

are more important than the number of the mutations[18, 19]. Numerous studies focused on identification of key residues which initiate amyloid formation, from either the sequence or structural standpoint. However, the responsible AL "hot spots" have not been identified.

Most crystal structures of V_L in amyloidogenic AL are dimeric, providing important information about LC polymerization[14, 20-29]. The dimeric crystal interfaces fall into three categories: canonical dimers, non-canonical dimers and dimers with partially misfolded protomers. Based on these, protomers can be classified into misfolded or folded. In canonical dimer interfaces, the two monomers are folded and their packing interface is native-like as between Fab's VL and VH domains. In non-canonical dimer interfaces, the two monomers are folded, but with non-native packing interfaces. In interfaces of partially misfolded protomers, such as domain-swapped interfaces[23, 30], the protomers are partially unfolded. The C-terminal of one monomer is swapped, forming intermolecular β-sheet with the other. Recently solid-state NMR spectroscopy revealed that the most of the sequence of AL-09 V_L is immobilized in the fibrils and that the N- and C-terminal portions are well-structured[31]. Fragments from other light chain regions can also form tight hydrophobic interactions [32].

The involvement of N-terminal residues in AL amyloid fibrils has been delineated by AL amyloid fibril-antibody interaction[33]. The antibody 11-1F4 can target N-terminal 1-18 residues of the AL protein; however, the epitope is linear and discrete. This antibody can recognize AL deposits; but not the free antibody light chains. Interestingly, 11-1F4 can also target Aβ 1-40 or Aβ 1-42 although Aβ shared limited sequence similarity with AL proteins[34]. This suggests it recognizes structural rather than sequence motif. As the β-strand-turn-β-strand U-shape motif is common among β sheet arrangement in Aβ amyloids, it is likely that AL fibril also share this motif.

Currently, treatments of systemic light chain amyloidosis aim to reduce or eliminate amyloidogenic immunoglobulin light chain cell clones[35]. While there is no approved drug that directly inhibits or reverses the amyloid, small molecules and antibodies that target them are actively studied; and the 11-1F4 antibody is already in phase 1 clinical trials [36]. However, lack of light chain fibril structural information hampers understanding of AL amyloid formation and drug design.

Here, we systematically examined the correlation of sequence variation and structural stabilities of dimeric variable domains of immunoglobulin light chains. We simulated six types of dimer interfaces, including three crystal dimers with folded protomers and three dimers with misfolded protomers. With four different sequences for each of the six structures, a total of 24 structure/sequence combinations were examined using molecular dynamics simulations. Based on the extensive samplings of the conformations generated by the simulations, we mapped 27 patients' sequences (15 λ and 12 κ) onto these six interfaces to investigate the energy profiles considering sequence and structural preferences. The energy profiles indicated a sequence-dependent dimer preference. For κ AL, all twelve sequences preferred a non-canonical interface with the folded protomers. For λ AL, the 15 sequences showed preference in five dimer interfaces except the canonical interface, while dimers with misfolded protomers seem more favored (9 out of 15). Further energy evaluation of the native and two misfolded protomers suggested that the native protomer has the lowest energy while the partially misfolded protomer presented a lower energy than the domain-swapped protomer. Interestingly, the energy gap between a partially misfolded and native protomer in κ ALs is higher than that in λ ALs. Our study suggested that λ protomers are more prone to unfold and aggregate than κ protomers, explaining why λ ALs are more common than κ ALs in AL patients.

2. Results

2.1. Proline 8 isomerization may control AL dimer association.

Dimer interfaces from available crystal structures (PDB code: 1bre, 1bww ,1pq1, 1rei, 2kqn, 2q20, 5c9k, 4aix, and 4unt) were extracted and superimposed by Chimera[37]. Most of these crystal structures have only one dimer in the unit cell while some unit cells contain multiple dimers, i.e. 1bre, 4aix, 4unt. We extracted these multiple interfaces into separate independent ones. Thus, there are 17 dimer interfaces, and 10 of these have very similar packing to the canonical dimer interface (Fig. 1a) of 1rei[26]. It has been argued that canonical VL dimers mimic the non-pathological conformation of physiologically native Fab antigen-binding domains, thus may not form amyloid fibers prior to a structural rearrangement[23]. Other 5 interfaces maintain the well folded monomer structure, with different overall packing compared with the canonical dimer interfaces. We name these non-canonical dimer interfaces (Fig. 1a).

Domain-swapping was proposed as one possible mechanism in light chain amyloid formation[23, 38]. The amyloidogenic mcg mutant protein (PDB: 4unt) forms domain-swapped dimer (Fig. 1a), in which the C-terminal tail of one protomer swaps into another protomer and forms integrated intermolecular β-strands[23]. Another similar domain-swapped dimer of immunoglobulin light-chain-like domain, CTLA-4 was suggested as the molecular basis for immunoglobulin domain aggregation[38]. Interestingly, CTLA-4 domain-swapped dimer displays cis–trans proline isomerization. Proline isomerization may be the rate controlling step in protein folding[39, 40], particularly in antibodies and other immunoglobulin-like β-sandwich proteins[41-43]. Human β2-microglobulin (β2m) amyloids also involve proline isomerization[44, 45].

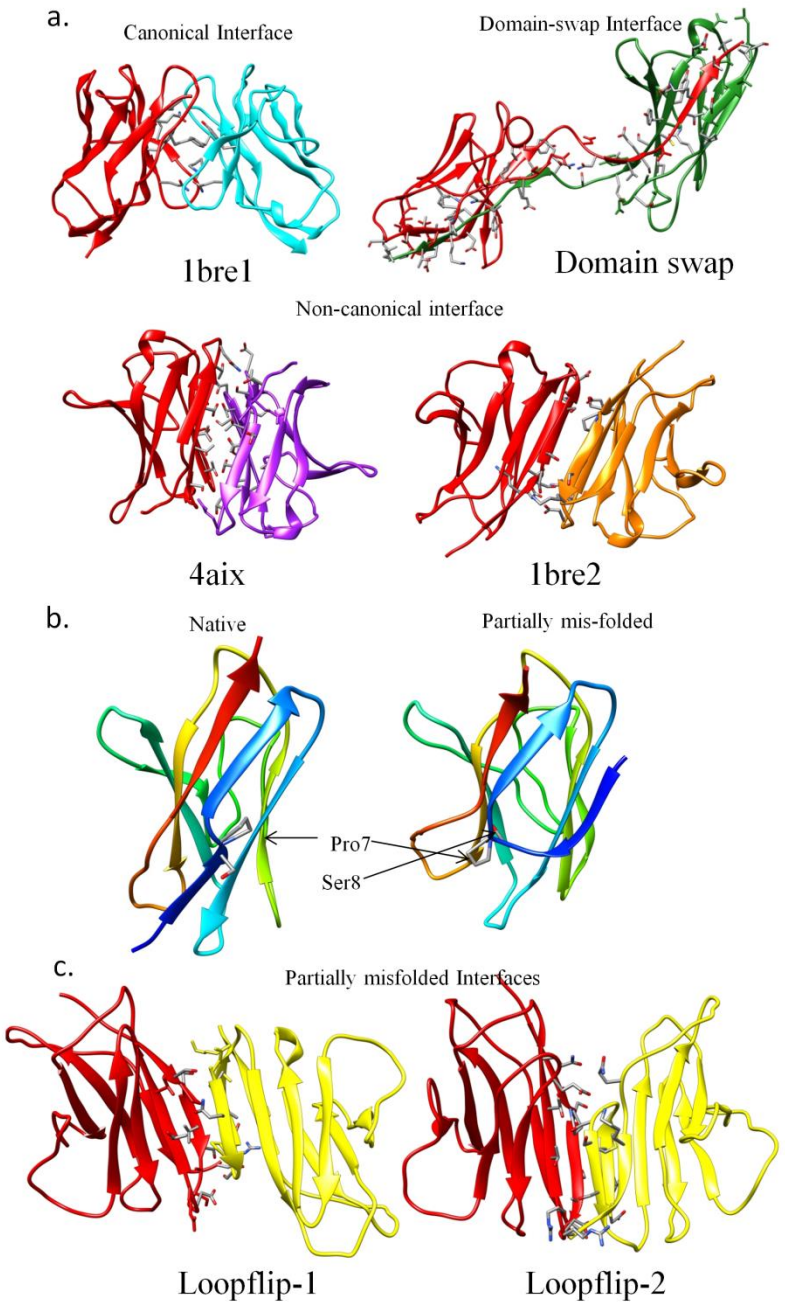

Fig. 1 Dimeric interfaces between AL protomers suggest diverse V_L interfaces. a. The four representative dimer interfaces from crystal structures. b. the proline isomerization is potentially leading to VL protomer partially misfolding. c. Two stable dimers formed by one normally folded protomer and one partially misfolded protomer. The interfacial residues are represented by ball and sticks in grey.

Proline isomerization may also play an important role in AL amyloid formation. Early on, it was noticed that two cis-prolines (Pro8 and Pro95) in the Bence-Jones protein (the protein involved in systemic light chain amyloidosis) the cis-pro-bend characterizes the stable light chain fold[26]. With cis-trans switch, Pro95 at a hinge position can trigger a domain-swapped dimer. Pro8 is conserved (100%) among κ AL sequences and many λ sequences as well. While κ Pro8 is mostly cis-proline, the corresponding λ proline can undergo transition.

Loop flip at Pro8 was previously proposed to lead to a β-sheet conformation by inserting residues 1-7 into the β-strands between residue 9-15 and 16-26[33]. However, we found that there is not enough space between these residues to allow the insertion. Considering that the 11-1F4 antibody can recognize LC amyloid around residues 1-18 and the antibody 11-1F1 have cross reactivity with Aβ1-40 or Aβ1-42 fibrils[34], a classic Aβ-like β-strand-turn-β-strand U-turn structure can be formed with Pro8 isomerization and a subsequent loop flip (Fig. 1b). In this loop flip monomer, residues 1 to 6 form intra-molecular parallel β-sheet with residues 16 to 24, with Ser7 and Pro8 on the turn region (Fig. 1b). Indeed, flipping the U-turn structure into the 1bre dimer interface revealed that the interface is compatible with the U-turn monomer structure. We investigated 8 possible dimer interfaces containing at least one loop-flipped protomer (Fig. S1 and Table S1) with the constraints that AL residues 1-18 form a compact epitope that can be recognized by antibody (like11-1F4).

The simulations identified two stable conformers (Fig. 1c). In "loopflip-1" model, Pro8 is still in cis conformation, residues 1 to 6 of the misfolded protomer form an inter-molecular parallel β-sheet with residue 9 to 14 from the neighboring folded protomer. In the "loopflip-2" model, pro8 flips to trans, and residues 1-6 of the misfolded protomer form an inter-molecular anti-parallel β-sheet with residue 9-14 from the neighboring folded protomer.

2.2. AL dimeric association is sequence-dependent

Among these 17 interfaces, we focused on structures with multiple dimer interfaces as these may represent the AL fibril structure. Among them, we selected 1bre, 4aix, and 4unt as representative crystal structures for canonical (1bre), non-canonical (4aix), and domain swapped-dimers (4unt). 1bre structure contains one canonical dimer interface and additional intermolecular β-strands between monomers. The intermolecular β-strand is characteristic of amyloid fibrils[46], and this dimer interface is possible for folded LC monomers. We selected the 1bre dimer (1bre-2) for a β-sheet extension mechanism of folded monomers in AL. Two loop-flip U-turn dimer structures are also included to test sequence dependence. While these tested interfaces may not exhaustively represent the vast conformational spaces potentially involved in the AL fibril formation. In addition, crystallization conditions such as solvent, pH, temperature, and ionization strength will significantly affect the folding and polymerization states of crystal structures. Nevertheless, our screen of 17 dimer interfaces in crystal structures (PDB code: 1bre, 1bww, 1pq1, 1rei, 2kqn, 2q20, 5c9k, 4aix, and 4unt) found the protein-protein interactions in the LC dimer are restricted by characteristic shapes and sequences in the immunoglobulin folds. For each of the selected six dimer structures, we systematically modeled four LC sequences (1bre, 4aix, 4unt, and 4unu), leading to 24 unique dimer systems. For each of the twenty-four, to sufficiently sample the local potential conformational space, we run 100 ns MD simulations in explicit water to examine their energies and structural stabilities. Therefore, our simulations focus on the sampling relevant dimerization states that could lead to LC polymerization.

Table 1. Average total solvent-accessible surface area ($SASA_{TOTAL}$), interface area (ΔSASA), number of hydrogen bond (H-bond), hydrophobic contacts (Hphob), and salt bridges (Charged) between the two protomers of the dimers in as calculated using last 50 ns of the trajectories.

Interface	Sequence	$SASA_{TOTAL}$ ($Å^2$)	ΔSASA($Å^2$)	H-bond	Hphob	Charged
1bre-1	1BRE	10880.7±189.8	1759.8±97.1	4.2±1.3	16.4±1.9	0±0.1
	4AIX	10325±150.3	1717.8±107.8	5.6±1.5	14.2±2.5	0±0
	4UNU	11700±202.8	786.9±109.2	3±1.1	5.1±1.5	0.4±0.5
1bre-2	1BRE	10749.9±161.4	1539.8±56.9	20±1.9	7.8±0.4	2±0.1
	4AIX	11237.2±195.2	1409.9±124.4	7.9±1.7	3.4±1.4	0±0
	4UNU	11303±247.7	1675.2±126.7	5.5±2.2	7±1.8	0±0.2
4unu	1BRE	11392.5±183.7	1278.8±72.8	15.1±2.1	7.4±0.7	0.6±0.5
	4AIX	11164.1±138.3	938.8±77.2	5.1±1.7	0.6±0.5	0.2±0.5
	4UNU	10850.5±180.7	1614.7±91.7	4.7±2	8.2±1.3	0±0
Domain-swap	1BRE	12474.8±183.5	4732.9±113.7	38.9±2.2	39.5±2	0±0.2
	4AIX	11867.3±162.1	4349.1±92.5	40.7±2.2	54.9±1.9	2.7±0.5
	4UNT	11916.8±276.3	4174±72.2	49±2.8	41.5±1.7	3.1±0.9
Loopflip-1	1BRE	11792.9±206.4	1206.3±116.5	10.7±2.6	7.6±0.6	0.8±0.7
	4AIX	11482.6±181.4	912.5±122.2	8.9±1.7	2.3±1.1	2.6±1.3
	4UNU	12184.3±379.5	732.4±209.2	4.5±3.5	1±1.2	1.1±1.6
Loopflip-2	1BRE	11368.7±175.7	1405.1±75.6	13.9±2.6	6.5±0.6	2.6±1
	4AIX	11790.7±189	1100.3±65.8	7.6±1.3	5.3±0.7	0±0
	4UNU	11716.9±196.1	942.2±166.5	2.8±2.2	6.1±2	0±0

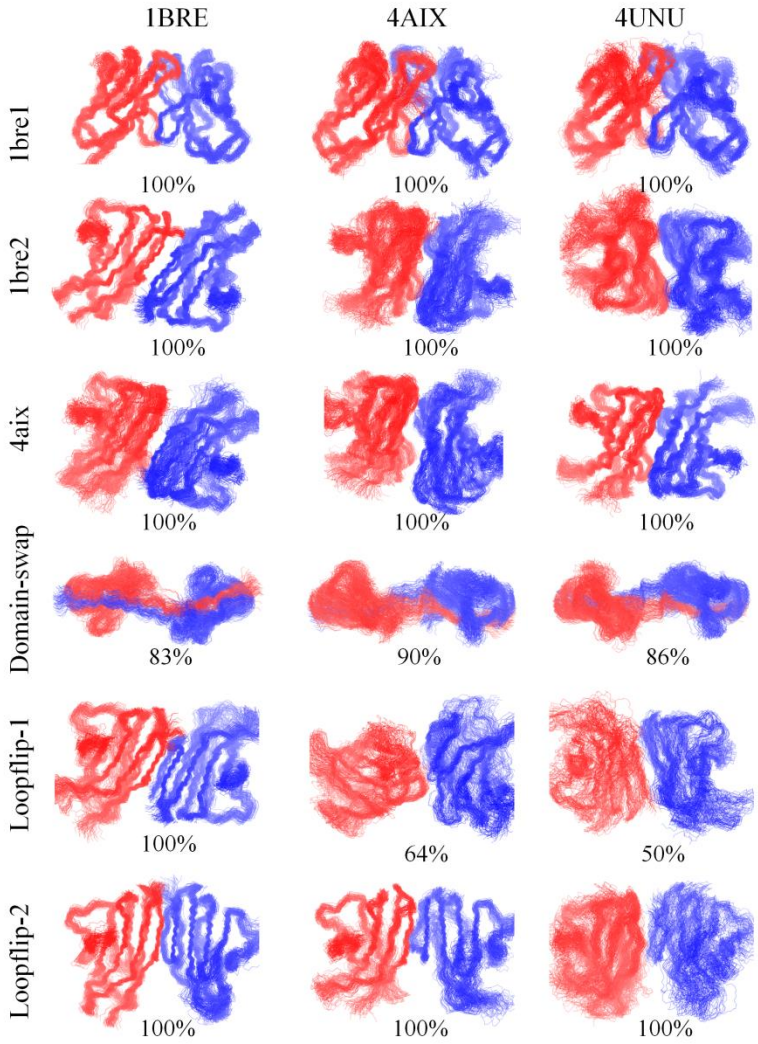

Fig. 2 Cluster analysis of the dimer interfaces reveals stable interfaces between V_L protomers. Structures with backbone RMSD <5 Å are considered as one cluster. The two protomers are colored by red and blue respectively. The population percentage of the most populated clusters listed. Each row indicated the different dimer interface and each column are from different light chain sequences.

RMSDs analysis (Fig. S2) showed that most of the 24 systems reach equilibrium after ~50 ns MD simulations. In most of the systems, the two protomers are well associated, though there is structure reorientation in some, e.g. loopflip-1 interface. However, the domain-swapping mcg mutant protein sequence is only compatible with its native domain-swapped structure (PDB:

4unt); all other three structures modelled using the mcg mutant protein sequence disassociated during our simulations. However, other LC sequences are compatible with the domain-swapped dimer structure. Generally, domain-swapped dimers have larger total contact surface area (Table 1). The interface areas ΔSASA of the domain-swapped dimers are approximately three times larger than other dimers. The numbers of intermolecular hydrogen bonds, hydrophobic contacts, and salt bridges in domain-swapped dimers are the largest among the four dimer interfaces. Domain-swapped dimers are more flexible due to non-rigid strands connecting two domains. We performed the cluster analysis on all the 24 dimer systems (Fig. 2). For most of the systems, the dimers have only one dominant conformation cluster (backbone RMSD < 5Å), while for the domain-swapped interfaces, more than 80% of the structures are in one dominant cluster and for the loopflip-1 interfaces, more than 50% of the structures are in one dominant cluster.

Table 2 Essential residues different between κ and λ chains

κ		λ	
[1]Res#	sequence	[2]Res#	sequence
7-8	ProSer	7-8	ProPro
15	Val	14	Pro/Leu/Ser
17	Asp	16	Gln
25	Ala	24	Gly
55	Glu/Gln/Lys	57	Pro/Ala
66	Gly/Gln	68	Lys/Asn
71	Phe/Tyr	73	Ala
80	Pro/Leu	82	Ala/Thr
83	Phe/Ile	85	Glu
85	Thr	87	Asp
95	Pro/Leu	97	Asp/Asn/Gly/Leu/Thr/Ser

1. residue number from 1BRE; 2. residue number from 4UNT

Dimeric structural flexibilities are sequence-dependent. The RMSFs (Fig. S3) showed six peaks corresponding to the six loops in LC. The CDR3 loop (residue 95-102) near the C-terminal

showed larger fluctuation in the interface of 1bre1 and 1bre2 than 4unu and 4unt_dw. The 24 systems indicated that both sequence and dimer packing influence the secondary structure (Fig. S4). In most dimeric structures, the 1BRE sequence adopts more β structure than 4AIX and 4UNT sequences.

We studied the differential conservation between λ and κ sequences with respect to interface preference (Table 2, and Fig. S5). On the 1bre2 interface, Ser9-Ile21 is highly conserved among the κ sequences while in the λ sequences, Ser9 and Val15 are mutated to proline, which undermined the intermolecular β-strand. Sequence variation between λ and κ may explain the 1bre2 interface preference of κ.

Thr106-Leu111 are more conserved in the κ than in the λ sequences. On the CDR3 loop, both λ and κ sequences showed high diversity, however, there is a conserved proline on site 95 of κ sequences (mostly in cis-conformation). Pro80 is highly conserved in κ but not in λ sequences. The two prolines are on the loop and maintained the secondary structure of the β-sheets in κ sequences while their mutations undermine the stability of the VL protomer and thus increase the possibility of its unfolding.

Overall, our simulations (with cumulative 2.4 μs) of the twenty-four systems indicated that while the crystal dimer structures prefer their native sequences in which the protein is crystalized, the newly found dimer structure with loop-flipped monomer structure could be a candidate for LC polymerization. Amyloid growth along the fibril axis may consist of alternation of loop-flipped and native canonical interface. Among the VL protein structures with more than one interfaces, the canonical interface is always present[20, 23, 24] although the energy indicates that this interface is not the most favorable compared with the domain-swapped and 1bre-2

interfaces. However, the domain-swapped and 1bre-2 interfaces always have higher energies in both Vλ and Vκ sequences. Superimposing the 1bre and 4unt complex structures indicated that both complexes shared the canonical dimer interface (Fig. 3a). Thus, an organized structure requires the canonical interface for fibril extension. Interface residues and multiple alignment analysis suggested that canonical dimer contact interface residues 38-55 are conserved, suggesting that AL fibrils contain at least two dimer interfaces: misfolded and the canonical dimer interfaces of both λ and κ chains. In this growth mechanism, a misfolded loop-flipped monomer can recruit a native monomer for amyloid formation (Fig. 3b), indicating the sensitivity of amyloid formation. The diameter of the modeled fibril is ~2.5nm (Fig. 3b), which agrees with the experimental data[47].

2.3.κ sequences have higher unfolding energies and are stabilized by native dimer interface

Amyloid structures of AL patients are still a mystery. We selected 23 LC sequences (11 κ and 12 λ sequences) from more than 100 AL patients LC sequences to examine their dimer structure preference (Fig. 4a). To examine the structural preferences of the LC sequences found in AL patients, we mapped the 23 sequences onto the conformation ensemble library generated from our 2.4 μs MD simulations of the six representative interfaces. The criteria for the 23 sequences selection were similar sequence length as in the conformation ensemble library, with a maximum of two residues gap (deletion) to minimize structural perturbation during homology modeling.

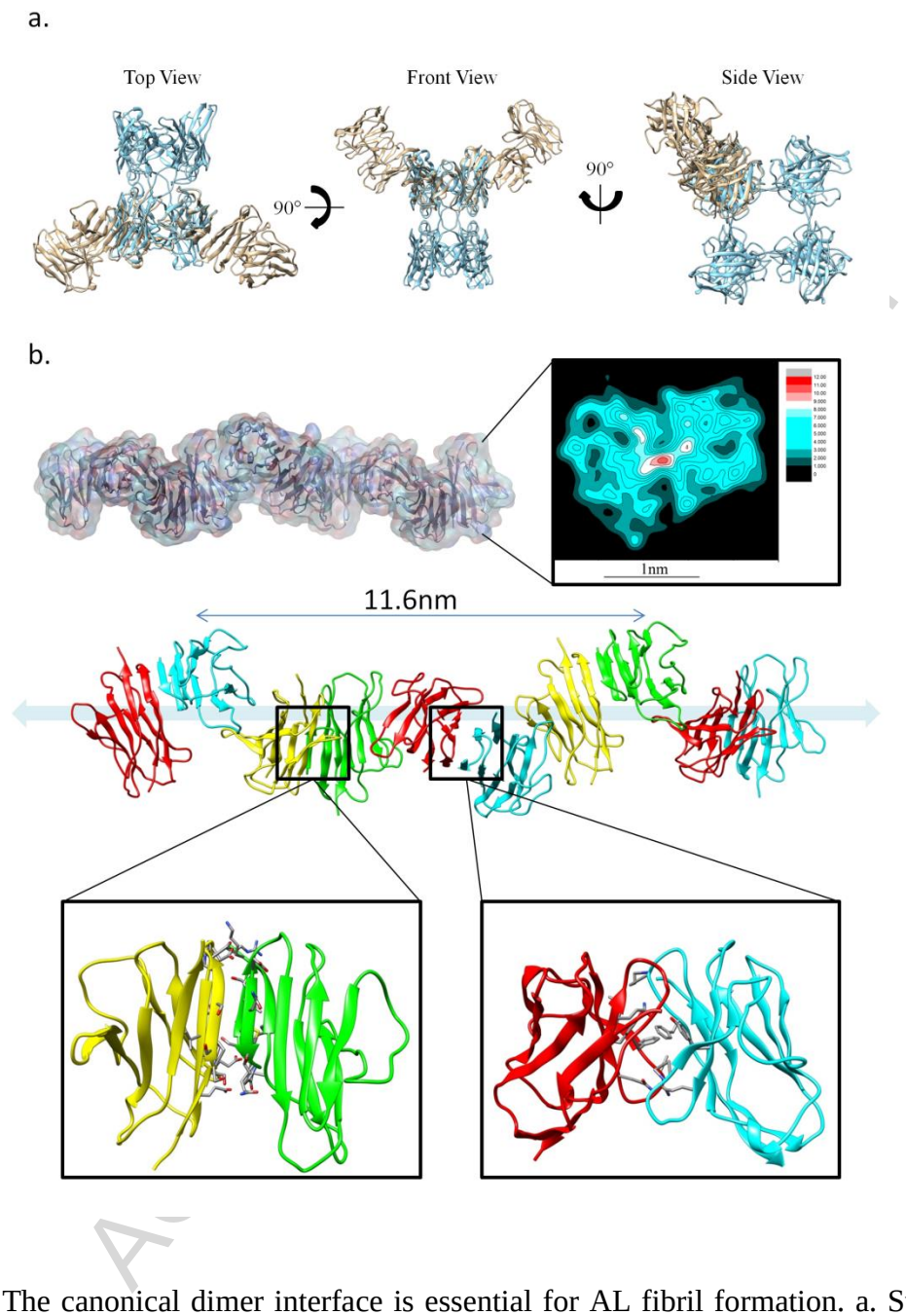

Fig. 3 The canonical dimer interface is essential for AL fibril formation. a. Superimposition of the two fibril structures (1bre and 4unt) with multiple dimer interface. The 1bre complex is colored in brown and 4unt complex is colored in cyan. b. The fibril structure of AL with the partially misfolded protomer and native protomer.

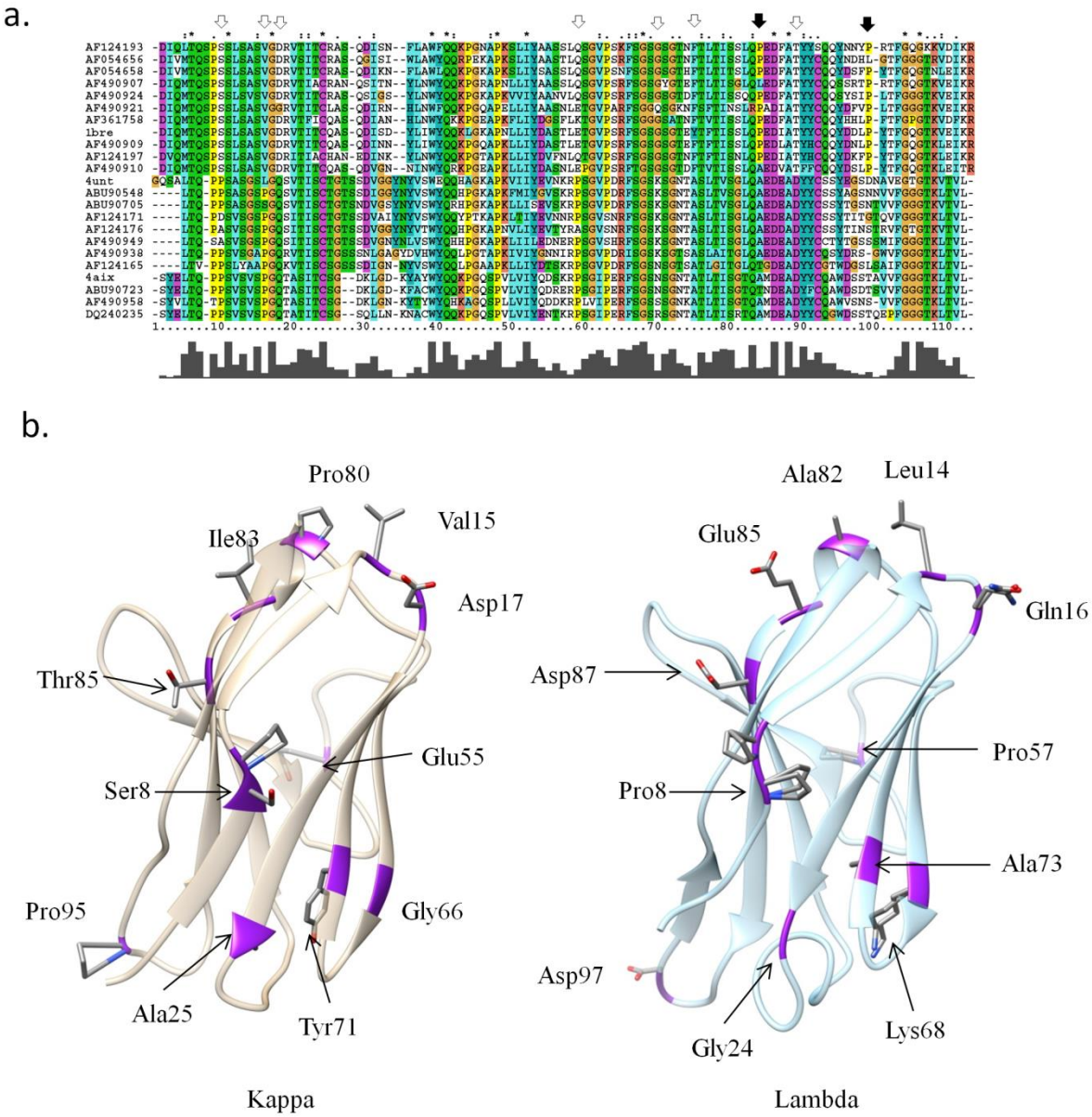

Fig. 4 Sequence alignment of (a) κ and λ sequences. a. Sequence alignment of the 23 AL patients' sequences simulated in this work. The arrows indicated the major different residues between κ and λ. The black arrows are the two prolines from κ sequences that maintain the β-structure. b. The major different residues between κ and λ sequences are mapped onto the κ (1BRE) and λ (4UNU) monomer structures.

Overall, the AL patients' sequences have different relative stabilities (Fig. 5), consistent with earlier finding that proteins derived from AL patients might follow different aggregation pathways[16]. However, we see a trend where λ sequences prefer misfolded interfaces (loop-flipped and domain-swapped interfaces) while κ sequences prefer one of the native dimer interface 1bre-2. The canonical and non-canonical 4unu dimer interface are not preferred with most sequences. For comparison, in all λ structures, Pro8 were switched to trans-conformation by homology modeling, while κ sequences have both cis- and trans- Pro8. The results indicated that the trans-conformations are overall more stable than cis-proline conformers.

The first step in formation of an amyloidogenic interface is perturbation of the folded monomer to a misfolded conformation. To evaluate the relative stabilities of the folded and unfolded monomers, we simulate and evaluate the conformational energies of the folded monomer, partially misfolded monomer and the C-terminus flipped (pre-domain swapping) monomer. As expected, RMSDs and cluster analysis suggest that the C-terminus flipped monomer (from the domain-swapped dimer) has the largest structural variation (Fig. S6-S8).

We then map all the 23 AL sequences onto the three monomeric conformational ensembles. Except the designed domain-swapped sequence (4UNT), all sequences encounter energy barriers to form misfolded monomers (Fig. 6). On average, the loop-flipped monomer (42.7 kcal/mol in κ sequences and 38.2 kcal/mol in λ sequences) has higher energy than the native monomer. The domain-swapped protomer showed higher energy than the native protomer, with 70.6 kcal/mol in κ sequences and 75.6 kcal/mol in λ sequences. Clearly, the λ sequences form the loop-flipped monomer, a precursor for further amyloidogenic oligomerization, more easily.

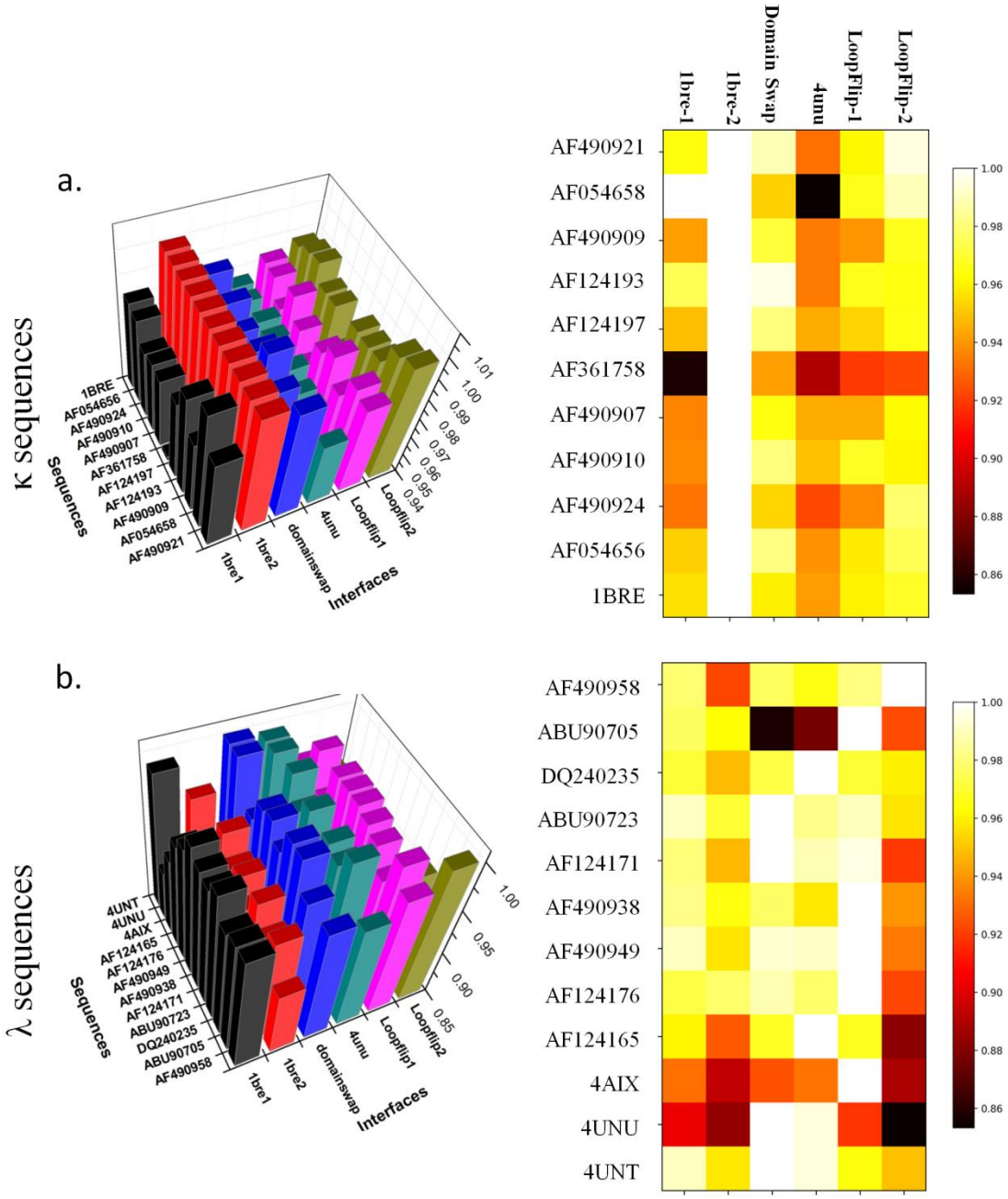

Fig. 5 Relative energies of the six dimer interfaces with diverse sequences. (a) κ and (b) λ show preference of a non-amyloidogenic interface from κ sequences. The relative energy is calculated by the equation of $E_R = E/E_{min}$ where E_R is the relative energy, E is the energy of specific sequence on specific interfaces, and E_{min} is the minimal energy of certain sequence on six different dimer interfaces. The left panel is the 3D bar figure and the right panel is the heatmap figure.

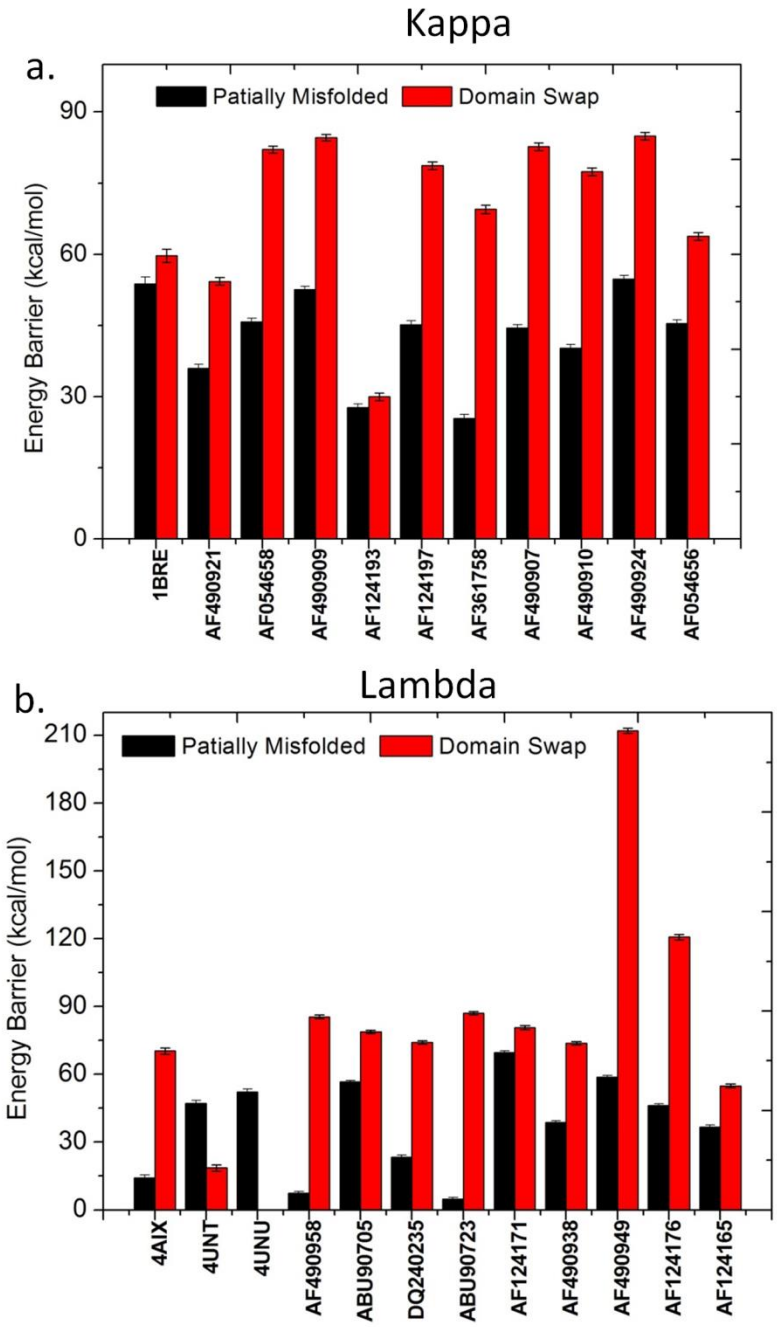

Fig. 6 Energy gap between misfolded and native protomers for (a) κ and (b) λ sequences suggest that the partially misfolded protomer can easily become a loop-flipped dimer. The Energy gap is calculated by the equation of $E_G = E - E_{native}$ where E_G is the Energy Gap, E is the energy of specific sequence on specific protomer and E_{native} is the energy of certain sequence on the native protomer.

3. Discussion:

Fibers in systemic light chain amyloidosis mainly consist of immunoglobulin light chains. Besides the fibrillar aggregates, the light chain can be also deposited in amorphous, dense granular bodies (LCDD)[13, 48-50]. This suggested that, the interactions between light chain protomers is potentially polymorphic. Numerous interfaces between AL proteins have been identified [20-27], confirming the polymorphic nature of AL deposits. Most dimer interfaces are derived from intact VL monomer structures, but misfolded monomers contribute to domain-swapped dimers[23].

Epitope mapping indicates that the mAb 11-1F4 binding site is contained within the first (N-terminal) 18 amino acids of AL proteins and residues of 1-4 and 13-18 are mainly in the epitope recognized by this antibody[33, 51]. Pro8 is the crucial and conserved residue whose mutation causes dramatic binding affinity loss. These results suggested that N-terminal residues of 1-4 ,8, and 13-18 are in proximity and lead to the loop-flipped protomer, in which N-terminal residues of 1-4 and 13-18 form β-strands. The important role of the N-terminal residues in light chain amyloidosis resembles another amyloidogenic immunoglobulin: β2-microglobulin which is involved in type 2 diabetes. A small perturbation of a folded β_2-microglobulin can generate certain conformations that can act as templates. Domain opening of β_2-microglobulin was observed in early molecular dynamics simulations, even when the disulfide bond was left intact[52]. Deletion of the first six β_2-microglobulin residues can shift the equilibrium toward amyloidogenic conformations, which can serve as templates to select (catalyze) amyloid formation by wild-type β_2-microglobulin[53]. Proline cis-trans isomerization may also accelerate conformational changes of β_2-microglobulin[54].

Sequence based analysis of key amyloidogenic residues showed that in specific secondary structure elements, there are significant differences in the number of non-conserved mutations between non-amyloidogenic sequences and AL sequences[19]. In AL, λ is overrepresented (3:1) as compared to healthy individuals (λ:κ=1:2)[11]. Poshusta et al. reported that for the total 50κ and 91λ AL sequences from the patients, AL Vλ sequences have more non-conserved mutations than AL Vκ sequences[19]. Moreover, the locations of the non-conserved mutation hot spots distributed differently between Vλ and Vκ sequences. For example, Vλ sequences are 1.4 and 3.3 times more likely to have non-conserved mutations than AL Vκ sequences on CDR3, strand A and strand G. This mutation preference suggested that Vλ and Vκ monomer stabilities differ while having similar interfaces for polymerization. Consistently, our energy analysis of AL protomers showed that Vλ sequences showed a smaller energy gap between the native state and the misfolded state. The larger number of non-conserved mutations in Vλ sequences could be the result of the smaller energy gap.

The combination of monomer destabilization and misfolded dimer stabilization of the light-chain have also been identified in other systems [24, 55]. It was reported that a single mutation can cause normal VL protein to form AL protein aggregates. For example, P7S at the sheet switch region induced conformational changes which lead to λ6 light-chain fibrillogenesis. A salt bridge between R61 and D82 of Bence-Jones protein REI (κ chain) is critical to prevent the REI protein from aggregating. Several mutations on these two sites cause 1REI VL to form either amorphous aggregates or fibrils. Moreover, *in silico* study suggested that these mutations do not only destabilize the native dimer interface by increasing exposure of certain hydrophobic residues, but also shift the monomers from native structures to amyloid-like structures[56, 57].

Our energy evaluation suggested that different dimer preferences for Vλ and Vκ sequences. Firstly, this result suggests that interfaces with misfolded protomers are preferred by Vλ sequences. Secondly, Vκ sequences preferred a non-canonical interface with an intermolecular β-strands on the N-terminal. This interface was proposed by Schormann et al as an amyloidogenic interface. As the interface might only exist in dimers and cannot form fibrils, it protects VL from fibril formation. This also explain why κ chains are low in AL patients. However, the 11-1F4 antibody cannot recognize native folded soluble antibody light chain. If this interface is in the fibril and 11-1F4 recognizes the exposed motif, the 11-1F4 should also recognize the native VL, as the protomers in this interface are shared. 11-1F4 can also target Aβ. Combined, this evidence suggests that 11-1F4 targets an amyloid misfolded motif which is also in the AL fibrils.

4. Conclusion

In conclusion, we conducted extensive molecular dynamics simulations and conformation energy analysis of more than 30 sequences related to systemic light chain amyloidosis. We found that N-terminus loop flipping may reshape the immunoglobulin light chains dimer interface and lead to amyloid formation. Our study revealed that amyloid formation is easier for the λ light chain. Firstly, the energy gap between misfolded protomer and native protomer is lower in λ than in κ light chain. Secondly, κ light chains are dominantly stabilized by a non-amyloidogenic interface and this interface inhibits the κ light chain from forming other amyloidogenic interactions. Thus, the energy gap between amyloidogenic and non-amyloidogenic interfaces is larger in κ light chains while becoming small or disappearing in Vλ sequences. The stable loop-

flipped interface can provide the structural basis for AL fibril formation and drug design to inhibit its formation.

5. Materials and Methods

5.1. Structural and sequence analysis of interfaces

Four representative dimer interfaces from crystal structures, the canonical dimer (pdb:1bre), the domain-swapped dimer (pdb:4unt), two non-canonical dimer (pdb:1bre and 4aix), and two new dimer interfaces with partially misfolded protomer(s), loop-flipp1 and loop-flipp2, were selected for modeling (definition of the three interface types are given in the Introduction). Missing residues in the corresponding pdb files were modeled by template-based homology modeling using SWISS-MODEL Server[58]. The amino acid sequence of 1bre, 4aix, and 4unt was mapped onto other five interfaces and similarly modeled. The wild type protomer, partially misfolded protomer, and the domain-swapped protomer were extracted from the corresponding dimer interfaces. We selected 11 sequences of κ V_L and 12 sequences of λ V_L from the sequence list from Poshusta et al[19]. The selected sequences were in the same length or with two deletions compared with the sequences of the six interfaces. The sequence alignments were performed using blastx and blastp from NCBI[59].

5.2. Molecular simulation protocol

In the simulations, the N- and C-termini of the VL protomers were charged as NH_3^+ and COO^- groups, respectively. The conserved intra-domain disulfide bonds were constructed. The systems were then solvated by TIP3P water molecules, and sodium and chlorides were added to neutralize the system and to achieve a total concentration of ~150 mM. The resulting solvated

systems were energy minimized for 5000 conjugate gradient steps, with the protein fixed and water molecules and counterions were allowed to move, followed by additional 5000 conjugate gradient steps, where all atoms were allowed to move. In the equilibration stage, each system was gradually relaxed by performing a series of dynamic cycles, in which the harmonic restraints on the proteins were gradually removed to optimize the protein-water interactions. In the production stage, all simulations were performed using the NPT ensemble at 310 K. All MD simulations were performed using the NAMD software [60] with CHARMM36 force field [61]. MD trajectories were saved by every 2 ps for analysis. A summary of all simulation systems is listed in Table S4.

5.3. Energy evaluation of the systems

To evaluate the energy of the dimers and protomers, the trajectory of each system was extracted from the last 80 ns of explicit solvent MD to remove water molecules and ions. For the dimers/protomers with the 23 selected sequences, side chains of the corresponding mutated residues were changed directly based on the sequence alignments. The solvation energies of all systems were calculated using the generalized Born method with molecular volume (GBMV) [62] after 500 steps of energy minimization to relax the local geometries caused by the thermal fluctuations which occurred in the MD simulations. In the GBMV calculation, the dielectric constant of water is set to 80 and no distance cutoff is used. The GBMV energy of all the systems simulated in the work is summarized in Table S3-S4.

Acknowledgements

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. This

research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Jun Zhao was supported in part by the Intramural Research Program of the NIH, NIDCD. This research was supported in part by National Science Foundation of China with grant number 81773621.

Competing interests

There is no competing interest.

References

[1] F.J. Stevens, F.A. Westholm, A. Solomon, M. Schiffer, Self-association of human immunoglobulin kappa I light chains: role of the third hypervariable region, Proceedings of the National Academy of Sciences of the United States of America, 77 (1980) 1144-1148.
[2] J. Wall, M. Schell, C. Murphy, R. Hrncic, F.J. Stevens, A. Solomon, Thermodynamic instability of human lambda 6 light chains: correlation with fibrillogenicity, Biochemistry, 38 (1999) 14101-14108.
[3] P.W. Stevens, R. Raffen, D.K. Hanson, Y.L. Deng, M. Berrios-Hammond, F.A. Westholm, C. Murphy, M. Eulitz, R. Wetzel, A. Solomon, et al., Recombinant immunoglobulin variable domains generated from synthetic genes provide a system for in vitro characterization of light-chain amyloid proteins, Protein science : a publication of the Protein Society, 4 (1995) 421-432.
[4] Y. Kim, J.S. Wall, J. Meyer, C. Murphy, T.W. Randolph, M.C. Manning, A. Solomon, J.F. Carpenter, Thermodynamic modulation of light chain amyloid fibril formation, The Journal of biological chemistry, 275 (2000) 1570-1574.
[5] M.B. Pepys, Pathogenesis, diagnosis and treatment of systemic amyloidosis, Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 356 (2001) 203-210; discussion 210-201.
[6] R.L. Comenzo, Y. Zhang, C. Martinez, K. Osman, G.A. Herrera, The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig V(L) germ line gene use and clonal plasma cell burden, Blood, 98 (2001) 714-720.
[7] V. Perfetti, S. Casarini, G. Palladini, M.C. Vignarelli, C. Klersy, M. Diegoli, E. Ascari, G. Merlini, Analysis of V(lambda)-J(lambda) expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r (lambdaIII) as a new amyloid-associated germline gene segment, Blood, 100 (2002) 948-953.
[8] R.S. Abraham, S.M. Geyer, T.L. Price-Troska, C. Allmer, R.A. Kyle, M.A. Gertz, R. Fonseca, Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain-associated amyloidosis (AL), Blood, 101 (2003) 3801-3808.
[9] A. Solomon, B. Frangione, E.C. Franklin, Bence Jones proteins and light chains of immunoglobulins. Preferential association of the V lambda VI subgroup of human light chains with amyloidosis AL (lambda), The Journal of clinical investigation, 70 (1982) 453-460.

[10] S. Ozaki, M. Abe, D. Wolfenbarger, D.T. Weiss, A. Solomon, Preferential expression of human lambda-light-chain variable-region subgroups in multiple myeloma, AL amyloidosis, and Waldenstrom's macroglobulinemia, Clinical immunology and immunopathology, 71 (1994) 183-189.
[11] R.A. Kyle, M.A. Gertz, Primary systemic amyloidosis: clinical and laboratory features in 474 cases, Seminars in hematology, 32 (1995) 45-59.
[12] K.E. Olsen, K. Sletten, P. Westermark, Extended analysis of AL-amyloid protein from abdominal wall subcutaneous fat biopsy: kappa IV immunoglobulin light chain, Biochemical and biophysical research communications, 245 (1998) 713-716.
[13] J. Buxbaum, Mechanisms of disease: monoclonal immunoglobulin deposition. Amyloidosis, light chain deposition disease, and light and heavy chain deposition disease, Hematology/oncology clinics of North America, 6 (1992) 323-346.
[14] E.M. Baden, B.A. Owen, F.C. Peterson, B.F. Volkman, M. Ramirez-Alvarado, J.R. Thompson, Altered dimer interface decreases stability in an amyloidogenic protein, The Journal of biological chemistry, 283 (2008) 15853-15860.
[15] L. del Pozo Yauner, E. Ortiz, R. Sanchez, R. Sanchez-Lopez, L. Guereca, C.L. Murphy, A. Allen, J.S. Wall, D.A. Fernandez-Velasco, A. Solomon, B. Becerril, Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains, Proteins, 72 (2008) 684-692.
[16] L.A. Sikkink, M. Ramirez-Alvarado, Biochemical and aggregation analysis of Bence Jones proteins from different light chain diseases, Amyloid, 15 (2008) 29-39.
[17] K. Bodi, T. Prokaeva, B. Spencer, M. Eberhard, L.H. Connors, D.C. Seldin, AL-Base: a visual platform analysis tool for the study of amyloidogenic immunoglobulin light chain sequences, Amyloid, 16 (2009) 1-8.
[18] F.J. Stevens, Four structural risk factors identify most fibril-forming kappa light chains, Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis, 7 (2000) 200-211.
[19] T.L. Poshusta, L.A. Sikkink, N. Leung, R.J. Clark, A. Dispenzieri, M. Ramirez-Alvarado, Mutations in specific structural regions of immunoglobulin light chains are associated with free light chain levels in patients with AL amyloidosis, PLoS One, 4 (2009) e5169.
[20] N. Schormann, J.R. Murrell, J.J. Liepnieks, M.D. Benson, Tertiary structure of an amyloid immunoglobulin light chain protein: a proposed model for amyloid fibril formation, Proceedings of the National Academy of Sciences of the United States of America, 92 (1995) 9490-9494.
[21] L.K. Steinrauf, M.Y. Chiang, D. Shiuan, Molecular structure of the amyloid-forming protein kappa I Bre, Journal of biochemistry, 125 (1999) 422-429.
[22] M.I. Villalba, J.C. Canul-Tec, O.D. Luna-Martinez, R. Sanchez-Alcala, T. Olamendi-Portugal, E. Rudino-Pinera, S. Rojas, R. Sanchez-Lopez, D.A. Fernandez-Velasco, B. Becerril, Site-directed mutagenesis reveals regions implicated in the stability and fiber formation of human lambda3r light chains, The Journal of biological chemistry, 290 (2015) 2577-2592.
[23] B. Brumshtein, S.R. Esswein, M. Landau, C.M. Ryan, J.P. Whitelegge, M.L. Phillips, D. Cascio, M.R. Sawaya, D.S. Eisenberg, Formation of amyloid fibers by monomeric light chain variable domains, The Journal of biological chemistry, 289 (2014) 27513-27525.
[24] A. Hernandez-Santoyo, L. del Pozo Yauner, D. Fuentes-Silva, E. Ortiz, E. Rudino-Pinera, R. Sanchez-Lopez, E. Horjales, B. Becerril, A. Rodriguez-Romero, A single mutation at the sheet switch region results in conformational changes favoring lambda6 light-chain fibrillogenesis, Journal of molecular biology, 396 (2010) 280-292.
[25] I. Uson, E. Pohl, T.R. Schneider, Z. Dauter, A. Schmidt, H.J. Fritz, G.M. Sheldrick, 1.7 A structure of the stabilized REIv mutant T39K. Application of local NCS restraints, Acta crystallographica. Section D, Biological crystallography, 55 (1999) 1158-1167.

[26] O. Epp, E.E. Lattman, M. Schiffer, R. Huber, W. Palm, The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0-A resolution, Biochemistry, 14 (1975) 4943-4952.
[27] P.R. Pokkuluri, A. Solomon, D.T. Weiss, F.J. Stevens, M. Schiffer, Tertiary structure of human lambda 6 light chains, Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis, 6 (1999) 165-171.
[28] F.C. Peterson, E.M. Baden, B.A. Owen, B.F. Volkman, M. Ramirez-Alvarado, A single mutation promotes amyloidogenicity through a highly promiscuous dimer interface, Structure, 18 (2010) 563-570.
[29] K. Annamalai, F. Liberta, M.T. Vielberg, W. Close, H. Lilie, K.H. Guhrs, A. Schierhorn, R. Koehler, A. Schmidt, C. Haupt, U. Hegenbart, S. Schonland, M. Schmidt, M. Groll, M. Fandrich, Common Fibril Structures Imply Systemically Conserved Protein Misfolding Pathways In Vivo, Angew Chem Int Ed Engl, 56 (2017) 7510-7514.
[30] N. Sinha, C.J. Tsai, R. Nussinov, A proposed structural model for amyloid fibril elongation: domain swapping forms an interdigitating beta-structure polymer, Protein engineering, 14 (2001) 93-103.
[31] D.W. Piehl, L.M. Blancas-Mejia, J.S. Wall, S.J. Kennel, M. Ramirez-Alvarado, C.M. Rienstra, Immunoglobulin Light Chains Form an Extensive and Highly Ordered Fibril Involving the N- and C-Termini, ACS Omega, 2 (2017) 712-720.
[32] A. Schmidt, K. Annamalai, M. Schmidt, N. Grigorieff, M. Fandrich, Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril, Proc Natl Acad Sci U S A, 113 (2016) 6200-6205.
[33] B. O'Nuallain, A. Allen, S.J. Kennel, D.T. Weiss, A. Solomon, J.S. Wall, Localization of a conformational epitope common to non-native and fibrillar immunoglobulin light chains, Biochemistry, 46 (2007) 1240-1247.
[34] R. Hrncic, J. Wall, D.A. Wolfenbarger, C.L. Murphy, M. Schell, D.T. Weiss, A. Solomon, Antibody-mediated resolution of light chain-associated amyloid deposits, Am J Pathol, 157 (2000) 1239-1246.
[35] A. Dispenzieri, G. Merlini, Immunoglobulin Light Chain Systemic Amyloidosis, Cancer Treat Res, 169 (2016) 273-318.
[36] C.V. Edwards, J. Gould, A.L. Langer, M. Mapara, J. Radhakrishnan, M.S. Maurer, S. Raza, J.G. Mears, J. Wall, A. Solomon, S. Lentzsch, Interim analysis of the phase 1a/b study of chimeric fibril-reactive monoclonal antibody 11-1F4 in patients with AL amyloidosis, Amyloid, 24 (2017) 58-59.
[37] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis, Journal of computational chemistry, 25 (2004) 1605-1612.
[38] A.F. Sonnen, C. Yu, E.J. Evans, D.I. Stuart, S.J. Davis, R.J. Gilbert, Domain metastability: a molecular basis for immunoglobulin deposition?, J Mol Biol, 399 (2010) 207-213.
[39] W.J. Wedemeyer, E. Welker, H.A. Scheraga, Proline cis-trans isomerization and protein folding, Biochemistry, 41 (2002) 14637-14644.
[40] U. Weininger, R.P. Jakob, B. Eckert, K. Schweimer, F.X. Schmid, J. Balbach, A remote prolyl isomerization controls domain assembly via a hydrogen bonding network, Proc Natl Acad Sci U S A, 106 (2009) 12335-12340.
[41] M.J. Feige, L.M. Hendershot, J. Buchner, How antibodies fold, Trends Biochem Sci, 35 (2010) 189-198.
[42] H. Lilie, R. Rudolph, J. Buchner, Association of antibody chains at different stages of folding: prolyl isomerization occurs after formation of quaternary structure, J Mol Biol, 248 (1995) 190-201.
[43] L. Rognoni, T. Most, G. Zoldak, M. Rief, Force-dependent isomerization kinetics of a highly conserved proline switch modulates the mechanosensing region of filamin, Proc Natl Acad Sci U S A, 111 (2014) 5568-5573.

[44] V.Y. Torbeev, D. Hilvert, Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate beta2-microglobulin amyloid assembly, Proc Natl Acad Sci U S A, 110 (2013) 20051-20056.
[45] M.F. Calabrese, C.M. Eakin, J.M. Wang, A.D. Miranker, A regulatable switch mediates self-association in an immunoglobulin fold, Nat Struct Mol Biol, 15 (2008) 965-971.
[46] B. Ma, R. Nussinov, Selective molecular recognition in amyloid growth and transmission and cross-species barriers, J Mol Biol, 421 (2012) 172-184.
[47] C. Ionescu-Zanetti, R. Khurana, J.R. Gillespie, J.S. Petrick, L.C. Trabachino, L.J. Minert, S.A. Carter, A.L. Fink, Monitoring the assembly of Ig light-chain amyloid fibrils by atomic force microscopy, Proceedings of the National Academy of Sciences of the United States of America, 96 (1999) 13175-13179.
[48] D. Ganeval, L.H. Noel, J.L. Preud'homme, D. Droz, J.P. Grunfeld, Light-chain deposition disease: its relation with AL-type amyloidosis, Kidney international, 26 (1984) 1-9.
[49] J.N. Buxbaum, J.V. Chuba, G.C. Hellman, A. Solomon, G.R. Gallo, Monoclonal immunoglobulin deposition disease: light chain and light and heavy chain deposition diseases and their relation to light chain amyloidosis. Clinical features, immunopathology, and molecular analysis, Annals of internal medicine, 112 (1990) 455-464.
[50] M. Cogne, C. Silvain, A.A. Khamlichi, J.L. Preud'homme, Structurally abnormal immunoglobulins in human immunoproliferative disorders, Blood, 79 (1992) 2181-2195.
[51] B. O'Nuallain, A. Allen, D. Ataman, D.T. Weiss, A. Solomon, J.S. Wall, Phage display and peptide mapping of an immunoglobulin light chain fibril-related conformational epitope, Biochemistry, 46 (2007) 13049-13058.
[52] B. Ma, R. Nussinov, Molecular dynamics simulations of the unfolding of β_2-microglobulin and its variants, Protein Eng, 16 (2003) 561-575.
[53] T. Eichner, A.P. Kalverda, G.S. Thompson, S.W. Homans, S.E. Radford, Conformational conversion during amyloid formation at atomic resolution, Mol Cell, 41 (2011) 161-172.
[54] F. Fogolari, A. Corazza, N. Varini, M. Rotter, D. Gumral, L. Codutti, E. Rennella, P. Viglino, V. Bellotti, G. Esposito, Molecular dynamics simulation of beta-microglobulin in denaturing and stabilizing conditions, Proteins, 79 (2011) 986-1001.
[55] L.R. Helms, R. Wetzel, Specificity of abnormal assembly in immunoglobulin light chain deposition disease and amyloidosis, Journal of molecular biology, 257 (1996) 77-86.
[56] M. Nowak, Immunoglobulin kappa light chain and its amyloidogenic mutants: a molecular dynamics study, Proteins, 55 (2004) 11-21.
[57] M. Bhavaraju, U.H. Hansmann, Effect of single point mutations in a form of systemic amyloidosis, Protein science : a publication of the Protein Society, 24 (2015) 1451-1462.
[58] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: An automated protein homology-modeling server, Nucleic Acids Res, 31 (2003) 3381-3385.
[59] Database resources of the National Center for Biotechnology Information, Nucleic Acids Res, 44 (2016) D7-19.
[60] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, NAMD2: greater scalability for parallel molecular dynamics, J. Comput. Phys., 151 (1999) 283-312.
[61] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B, 102 (1998) 3586-3616.

[62] M.S. Lee, M. Feig, F.R. Salsbury, Jr., C.L. Brooks, 3rd, New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations, Journal of computational chemistry, 24 (2003) 1348-1356.

Graphical abstract