UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie

“Lazzaro Spallanzani”

Laurea Magistralis in Molecular Biology and Genetics

NMR Structural Characterization of the LB6 Peptidomimetic Inhibitor Targeting G2-D187N Aggregation in AGel Amyloidosis

Supervisor:

Prof. Vittorio Bellotti

Co-supervisor:

Dr. Laura Ragona

Experimental thesis by

Fatemeh Amini

Academic Year 2025/2026

UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie

“Lazzaro Spallanzani”

Laurea Magistralis in Molecular Biology and Genetics

Caratterizzazione strutturale mediante NMR del peptidomimetico LB6, inibitore dell’aggregazione del dominio G2-D187N della gelsolina, alla base dell’amiloidosi AGel

Supervisor:

Prof. Vittorio Bellotti

Co-supervisor:

Dr. Laura Ragona

Experimental thesis by

Fatemeh Amini

Academic Year 2025/2026

Work very hard on big problems, success doesn’t happen all at once, be open to unexpected discoveries.

Kurt Wüthrich (Nobel Laureate)

Abstract

Gelsolin amyloidosis (AGel) is a rare multisystemic disorder caused by mutations in the gelsolin gene. Clinical presentation is heterogeneous and often includes peripheral neuropathy, muscle weakness, and cardiac or renal involvement. The first identified and the most common mutation responsible for gelsolin amyloidosis is the single-point substitution D187N of plasma gelsolin, which make the second gelsolin domain (G2) susceptible to aberrant proteolysis. The G2 fragments are amyloid-prone as they include a highly amyloidogenic sequence, spanning residues 182-192, known as the gelsolin amyloidogenic core ($GAC_{182–192}$), which in the folded domain pairs in an antiparallel fashion with the flanking sequence 194-204[4]. In the search for potential inhibitors of the amyloid aggregation underlying AGel pathology, selective peptide-mimetics were designed to interfere with the formation of amyloidogenic aggregates in the tissues of AGel patients, starting from the localization and interactions of the amyloidogenic GAC sequence. The binding of the peptidomimetics to the GAC sequence, could indeed avoid the aberrant protein-protein interactions underlying the aggregation process. The best performing peptidomimetic inhibitor, LB6, consisted of two arms, N185–L191 and H198–N204, joined by a non-natural piperidine–pyrrolidine β-turn inducer. LB6 suppresses the aggregation of $GAC_{182–192}$ and of the isolated D187N G2 domain at sub-stoichiometric ratios.

It had not yet been established, however, whether LB6 inhibitory activity depends on the preorganization of the molecule into the β-hairpin structure for which it was designed. The aim of this thesis was to addresses this question by solution-state NMR spectroscopy.

For LB-6 peptide resonance assignment a full set of NMR spectra, including scalar and dipolar correlation experiments (TOCSY, ROESY, NOESY 1H 13C HSQC) were recorded on natural abundance LB-6 sample at 600 MHz spectrometer equipped with a cryoprobe. Both oxidizing and reducing conditions were investigated to assess the possibility of the formation of an intramolecular Cys188–Cys201 disulfide bridge formation. The absence of short-range inter-strand NOEs in NOESY/ROESY spectra excluded the presence of an intramolecular Cys188-Cys201 disulfide bridge.

The full assignment of LB6 peptidic NMR resonances revealed a doubling of the resonances of six residues, namely I190, L191, H198, W200, C201, S203, flanking the piperidine–pyrrolidine linker. This observation was consistent with the existence in solution of two conformational states interconverting in slow exchange on the NMR timescale, arising from the non-peptidic β-turn inducer conformations.

Hα and Cα secondary chemical shifts were evaluated against four independent random-coil references, highlighting a high degree of flexibility and disorder characterizing both peptide arms and showing that the designed β-hairpin is not populated in solution. NMR data revealed predominantly random-coil behavior, with only a transiently structured segment at the level of residues C201–G202–S203, as derived based on a sign inversion of ΔδHα between C201 and G202 and by large positive ΔδCα values of approximately +7 ppm. The same feature was independently recovered by chemical shift analysis through CheSPI (Chemical shift Secondary structure Population Inference) program, which assigns a non-folded population of about 75% throughout the molecule and a 3_{10}-helical element only at this segment. LB6 thus behaves as a highly flexible, dynamic peptidomimetic with only local, transient structural elements. Its inhibitory activity is unlikely to rest on rigid structural mimicry rather an adaptable ensemble of conformations appears to be the more plausible basis contributing to its ability to interact with misfolded G2 D187N intermediates.

Italian Abstract

L'amiloidosi da gelsolina (AGel) è una rara patologia multi-sistemica causata da mutazioni nel gene della gelsolina, caratterizzata da manifestazioni cliniche eterogenee che possono includere neuropatia periferica, debolezza muscolare e coinvolgimento cardiaco o renale. La mutazione D187N nella gelsolina plasmatica, rappresenta la prima mutazione identificata e la più comune associata alla malattia. Tale mutazione favorisce l'esposizione del frammento amiloidogenico, noto come gelsolin amyloidogenic core ($GAC_{182–192}$), localizzato nel secondo dominio (G2) dei sei che costituisco la proteina gelsolina. La maggior esposizione di questo frammento favorisce il taglio ad opera di proteasi specifiche e l'aggregazione di frammenti amiloidogenici a formare fibre che si depositano nei tessuti alterandone la funzionalità. Nella ricerca di potenziali inibitori dell'aggregazione amiloide alla base della patologia AGel, sono stati progettati dei peptidomimetici selettivi in grado di interferire con la formazione di aggregati amiloidogenici nei tessuti dei pazienti affetti da AGel, a partire dalla localizzazione e dalle interazioni della sequenza amiloidogenica GAC. Il legame dei peptidomimetici alla sequenza GAC potrebbe infatti impedire le interazioni proteina-proteina alla base del processo di aggregazione.Tra questi, LB6, costituito da due bracci peptidici, N185–L191 e H198–N204 collegati mediante un induttore di β-turn sintetico si è dimostrato il più efficace. LB6 è in grado di inibire l'aggregazione di $GAC_{182–192}$ e del dominio G2 D187N isolato anche a rapporti sub-stechiometrici.

Non è tuttavia noto il meccanismo di azione e in particolare se LB6 adotti effettivamente in soluzione la conformazione a β-hairpin per la quale è stato progettato, né se tale struttura sia necessaria per la sua attività antiaggregate.L'obiettivo di questa tesi è stato pertanto quello di caratterizzare la struttura di LB

6 in soluzione mediante spettroscopia NMR.

A tale scopo è stata effettuata l'assegnazione completa delle risonanze NMR del peptide attraverso l'analisi di una combinazione di esperimenti di correlazione scalare e dipolare, tra cui TOCSY, ROESY, NOESY e $^{1}H–^{13}C$ HSQC, acquisiti su un campione in abbondanza

naturale ad uno spettrometro operante a 600 MHz e dotato di sonda cryo-raffreddata. Il peptidomimetico è stato studiato in condizioni ossidanti e riducenti per verificare la possibile formazione di un ponte disolfuro intramolecolare tra Cys188 e Cys201 che potesse stabilizzare il β-hairpin. L'assenza di effetti dipolari negli spettri NOESY e ROESY ha escluso la formazione di tale legame disolfuro tra i due bracci peptidici che potesse essere indicativo della presenza in soluzione di una frazione di peptidomimetico in conformazione β-hairpin.

L'assegnazione delle risonanze ha inoltre evidenziato la presenza di due sistemi di spin per ciascuno dei residui I190, L191, H198, W200, C201 e S203, localizzati in prossimità del linker piperidina–pirrolidina. Questo fenomeno ha dimostrato la presenza di due stati conformazionali che interconvertono lentamente sulla scala dei tempi del NMR, probabilmente associati alle diverse conformazioni assunte dall'induttore di β-turn non peptidico.

L'analisi degli chemical shift secondari di Hα e Cα, calcolati rispetto a quattro differenti liste di chemical shift tipici delle conformazioni random coil, ha fornito importanti informazioni sulla tendenza di LB6 a formare strutture secondarie stabili in soluzione. I risultati evidenziano un marcato carattere flessibile e disordinato di entrambi i bracci peptidici e indicano che il β-hairpin non rappresenta una conformazione significativamente popolata in soluzione.

Nel complesso, gli aminoacidi di LB6 sono sostanzialmente in random coil, con la presenza di un solo elemento di struttura locale in corrispondenza dei residui C201–G202–S203. La presenza di questo elemento è supportata sia dall'inversione di segno dei ΔδHα di C201 e G202 e dagli elevati valori positivi di ΔδCα, pari a circa +7 ppm. In analogia, la previsione di popolazione di struttura secondaria fatta con il programma CheSPI (Chemical shift Secondary structure Population Inference). ha stimato una popolazione random coil di circa il 75% lungo la sequenza e ha evidenziato la presenza di un elemento di struttura elicoidale 3_{10} esclusivamente nel segmento C201–G202–S203.

In sintesi i risultati ottenuti dimostrano che LB6 non adotta in soluzione la conformazione β-hairpin prevista ma si comporta piuttosto come una molecola con un alto grado di flessibilità, caratterizzata da un insieme di conformazioni che inter-convertono e da pochi elementi strutturali locali. Questi risultati suggeriscono che l'attività inibitoria di LB6 non dipenda specificamente dalla formazione della struttura β-hairpin presente nel dominio G2. Al contrario, l'elevata plasticità conformazionale di LB6 potrebbe rappresentare un elemento funzionale, consentendo alla molecola di adattarsi strutturalmente e di interagire efficacemente con gli intermedi misfolded del dominio G2. La flessibilità conformazionale emerge pertanto come una possibile caratteristica chiave alla base dell'attività di LB6 e fornisce una nuova prospettiva per l'interpretazione del rapporto tra struttura, dinamica e funzione nei peptidomimetici diretti contro proteine amiloidogeniche.

Table of Contents

Introduction

1. Amyloidosis

Amyloidoses constitute a heterogeneous family of diseases that share a common molecular root: a protein that would normally remain soluble instead misfolds and self-assembles into insoluble fibrillar aggregates that accumulate within tissues and organs, progressively impairing their function.

Approximately one third of these diseases are always hereditary and generally present at an early age; the causative mutations are typically located within the polypeptide chain that undergoes aggregation and are autosomal dominant, so that a mutation in a single copy of the gene is sufficient to cause disease (Chiti and Dobson, 2017).

Forty-two peptides and proteins are currently recognized as forming amyloid deposits in human disease (Buxbaum *et al.*, 2024). These precursor proteins display remarkable structural diversity in their soluble, non-pathological states, sharing no obvious sequence homology or functional overlap. They encompass intrinsically disordered proteins (such as α-synuclein and tau), tightly folded globular proteins (such as transthyretin, lysozyme, and β-microglobulin) and fragments generated via aberrant proteolysis from larger precursors, as seen in amyloid-β and gelsolin amyloidosis (Chiti and Dobson, 2017).

The apparent paradox is resolved by recognizing that the word amyloid does not denote a substance but a structural state. It describes a specific, highly ordered arrangement that a polypeptide chain can adopt when it leaves its soluble form, and which shares characteristic structural features whichever protein produced it. The capacity to adopt this state is not encoded in a particular sequence in the way that a native fold is; it is an inherent property of the polypeptide backbone (Dobson, 2003). Under suitable conditions in vitro, proteins with no connection to disease, myoglobin among them, can be converted into fibrils with all the characteristics of those extracted from patients (Dobson, 2003; Chiti and Dobson, 2017), The same is true of polypeptides containing a single type of residue, such as polylysine and polythreonine (Fändrich and Dobson, 2002), and even of single amino acids such as L-phenylalanine (Adler-Abramovich, 2012).

Amyloid aggregation is therefore best understood not as a chemical peculiarity of certain proteins but as an alternative structural destination available, in principle, to all of them Figure 1. The capacity to aggregate is not what sets these forty-two proteins apart. What sets them apart is that in each tissue, under the right conditions, nothing prevents aggregation from taking place.

Three criteria must be satisfied simultaneously for an aggregate to be classified as amyloid: a fibrillar morphology when imaged by electron or atomic force microscopy; a cross-β architecture, detectable by X-ray fiber diffraction, solid-state NMR, X-ray crystallography or infrared spectroscopy; and characteristic binding of the dyes thioflavin T and Congo red, which are thought to form ordered arrays along the length of the fibril and give rise to diagnostic spectral changes (Iadanza *et al.*, 2018); (Chiti and Dobson, 2017). Aggregates that fail any of these tests, such as the amorphous deposits of light-chain deposition disease or the native-like assemblies of Berger disease, are associated with human pathology but are not amyloid.

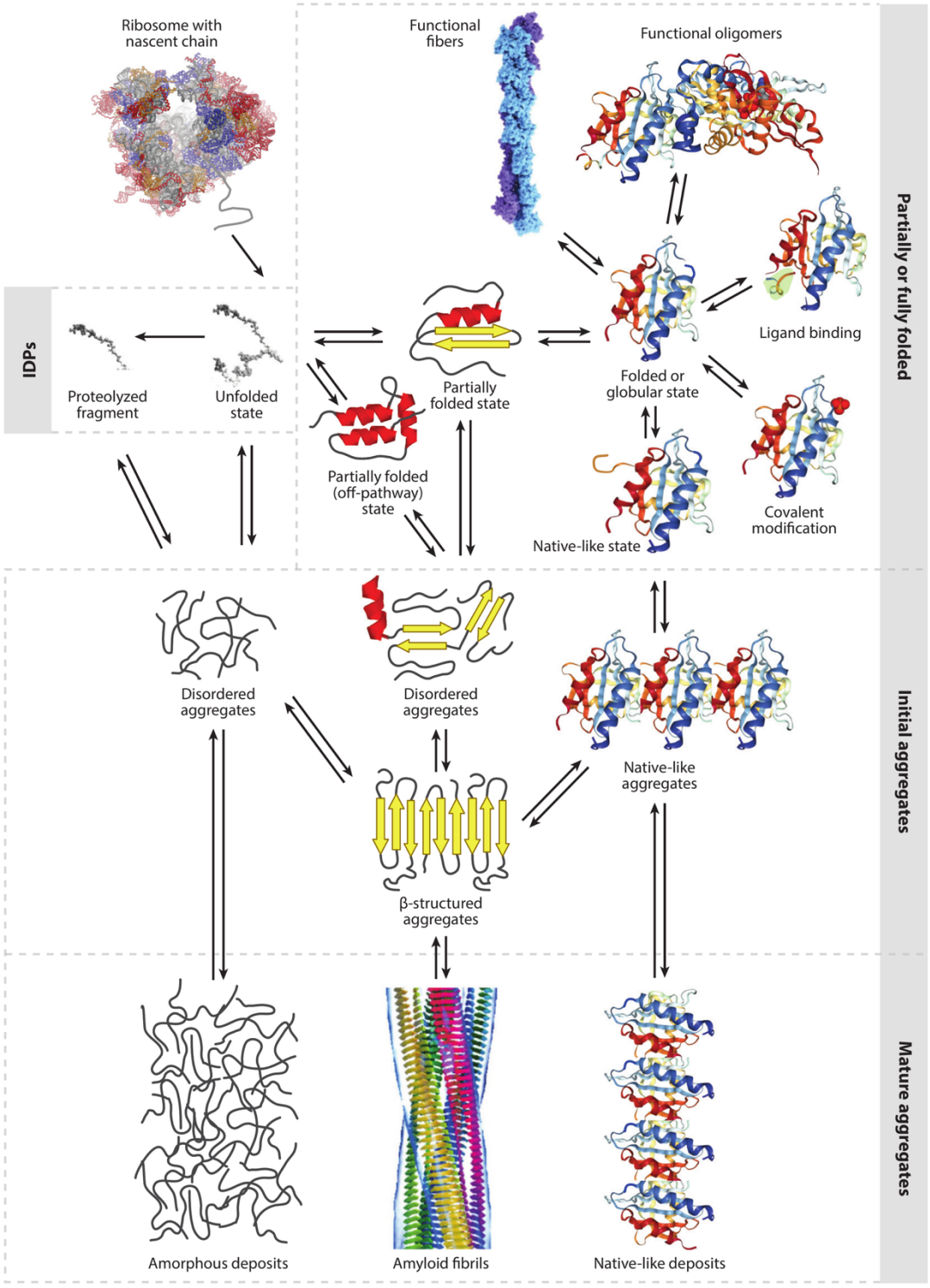


Figure 1 The conformational states accessible to a polypeptide chain following biosynthesis, and the transitions between them. The soluble native state, whether folded or intrinsically disordered, is one of several accessible states; aggregation competes with folding and is normally suppressed by the cellular proteostasis network. Adapted from (Chiti and Dobson, 2017).

From the point of view of structural biology, the defining feature of the amyloid state is the cross-β fold **Error! Reference source not found.**A. In this arrangement the β-strands lie perpendicular to the long axis of the fibril, and the hydrogen bonds that link them run parallel to that axis. Each molecule therefore contributes a single rung to a continuous molecular ladder that propagates along the fibril. The rung-to-rung distance is 4.7– 4.8 Å, fixed by the geometry of the backbone hydrogen bonds that hold neighboring strands together (Iadanza *et al.*, 2018).

X-ray fiber diffraction established this repeat in 1968–1969, and the resulting diffraction pattern is what gives the motif its name. Fibrils aligned in the beam produce two reflections at right angles to one another. The first is a sharp meridional reflection, parallel to the fibril axis, which arises from the regular spacing between hydrogen-bonded β-strands. The second is a more diffuse equatorial reflection, perpendicular to the first, which arises from the separation between stacked β-sheets. The cross formed by these two reflections is the origin of the term cross-β. For fibrils of the model peptide transthyretin (105–115), used to illustrate this section, these reflections occur at 4.67 Å and 8.86 Å respectively (Fitzpatrick *et al.*, 2013) Figure 2.

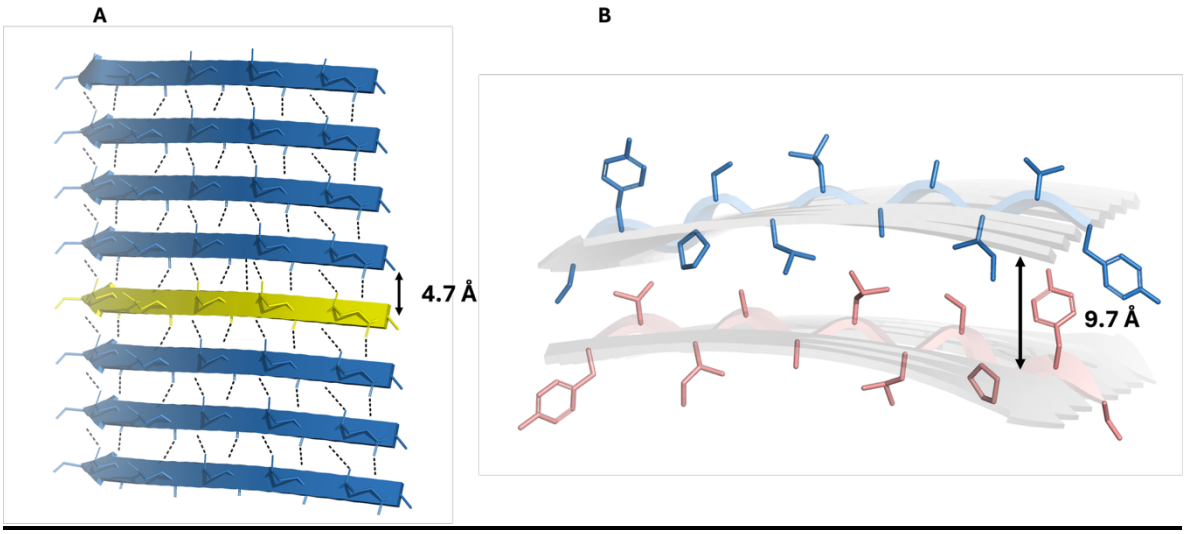


Figure 2 The cross-β architecture of the amyloid protofilament, illustrated using the solid-state NMR structure of a protofilament formed by residues 105–115 of transthyretin (PDB 2M5N;(Fitzpatrick et al., 2013). (A) View perpendicular to the protofilament axis, showing a single β-sheet. The β-strands lie perpendicular to the axis and are linked to their neighbors by backbone hydrogen bonds (dashed lines) that run parallel to it, so that each peptide molecule, shown in yellow, contributes a single rung to a continuous molecular ladder. Successive strands are separated by 4.7 Å, the characteristic

spacing of the cross-β fold. (B) The two-layered organization of the protofilament: two β-sheets (grey ribbons) are packed face to face and separated by 9.7 Å, with the side chains of the first sheet shown in blue and those of the second in salmon. Figure prepared in PyMOL.

The cross-β conformation has since been found in every amyloid fibril examined, whether disease-associated or functional, by X-ray crystallography, cryo-electron microscopy and Fourier-transform infrared spectroscopy (Iadanza *et al.*, 2018).

Infrared spectroscopy is particularly diagnostic. The strong, regular hydrogen bonds of the cross-β fold shift the amide I band to lower wavenumbers than are observed in native β-sheet: amyloid fibrils absorb between 1611 and 1630 cm^{-1}, whereas globular β-sheet proteins cluster between 1630 and 1643 cm^{-1}. Full-length transthyretin demonstrates the effect within a single sequence, the native tetramer absorbing at 1630 cm^{-1} and its fibrils at 1615 cm^{-1} in D_2O. The shift cannot be attributed to a change in the amount of β-structure, since the two states have almost identical β-sheet content (53% and 54% respectively); it reflects the cross-β architecture itself (Zandomeneghi *et al.*, 2004).

Above the level of the single sheet the organization is hierarchical Figure 3. Two β-sheets pack face to face, their side chains tightly interdigitated in an arrangement described as a steric zipper ((Nelson *et al.*, 2005); (Eisenberg and Sawaya, 2017)), to form a protofilament of some 2–7 nm in diameter. Protofilaments then associate with one another, either twisting together or packing laterally as flat ribbons, to give filaments and ultimately the mature fibril, which is typically 7–13 nm across and may extend for several micrometers; between two and eight protofilaments are found in the fibrils characterized to date (Chiti and Dobson, 2017).

The model peptide transthyretin (105–115) illustrates this progression particularly clearly, since the same sequence forms three distinct fibril types containing pairs of two, three or four interconnected protofilaments, with a filament height of approximately 3.9 nm. The interface between protofilaments is of a different character from the dry steric zipper within them. cryo-electron microscopy reveals a nearly constant 16 Å backbone-to-backbone cavity between adjacent protofilaments, wide enough to accommodate their exposed side chains together with an intervening layer of ordered water (Fitzpatrick *et al.*, 2013).

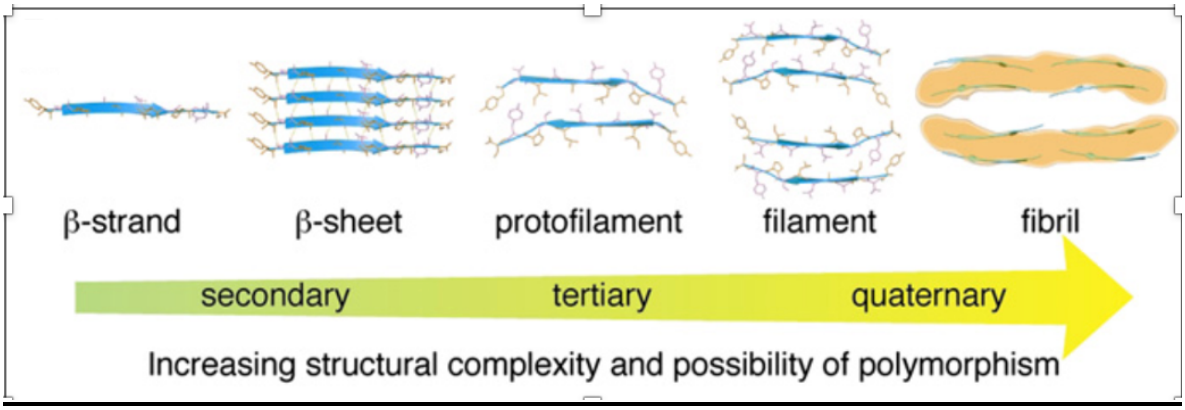


Figure 3 Hierarchical assembly of an amyloid fibril, shown for fibrils formed by residues 105–115 of transthyretin. The individual β-strand assembles into a β-sheet (secondary structure); two sheets pack face to face to form a protofilament and protofilaments associate into filaments (tertiary structure); filaments then twist around one another to produce the mature fibril (quaternary structure). Structural complexity, and with it the possibility of polymorphism, increases at each level. Reproduced from (Fitzpatrick et al., 2013).

In most structures determined to date the strands within a sheet are parallel and in-register, meaning that all strands share the same N to C terminal orientation and that residue i of one strand lies directly over residue i of the next. This arrangement maximizes both the number of backbone hydrogen bonds and the burial of hydrophobic side chains along the fibril axis, and additionally ensures that chemically similar residues are stacked upon one another rather than alternating between polar and apolar contacts; out-of-register arrangements necessarily generate alternating hydrophobic and hydrophilic interactions along the fibril direction and are correspondingly less stable (Chiti and Dobson, 2017).

Two consequences of this architecture deserve emphasis. The first is mechanical: amyloid fibrils have persistence lengths on the order of micrometers and a tensile strength approaching that of steel. A single hydrogen bond is weak, but in the cross-β fold thousands are aligned in register along a single axis and reinforced by multiple stacked sheets, so that cooperativity converts many weak interactions into an exceptionally robust material. The second is thermodynamic: the amyloid form of a polypeptide can be more stable than its native state, whether that state is folded or disordered, even under physiological conditions. This is particularly true of short sequences, which can commit most of their length to the ordered core, whereas longer chains incorporate only a fraction of their residues and are increasingly unable to order the remainder. Consistent with this, the polypeptides that form

amyloid in disease are short, half have fewer than 100 residues, and none exceeds 700, while those forming amorphous or native-like deposits are on average considerably longer (Chiti and Dobson, 2017).

Taken together, these properties define the therapeutic problem. Once the cross-β architecture has formed it is, for practical purposes, irreversible: the structure is stronger than the native protein and more stable than it. Any intervention must therefore act upstream, at a stage when the process remains reversible. For many years no disease-modifying treatment existed for any amyloid disease. That position has recently changed with the approval of the anti-amyloid antibodies lecanemab (2023) and donanemab (2024) for early Alzheimer's disease, which provide the first clinical demonstration that removing amyloid from the brain can slow cognitive decline(Sims, 2023; van Dyck *et al.*, 2023). These agents act on deposits that have already formed, and their benefit is modest, but they establish that intervention against the aggregation process is a viable therapeutic strategy rather than a theoretical one.

Until relatively recently the atomic structures available were those of short peptide fragments, together with models of full-length fibrils built from experimental restraints. Between 2016 and 2018, developments in cryo-electron microscopy and solid-state NMR made it possible to determine near-atomic structures of fibrils formed from intact disease proteins, including amyloid-β, tau and α-synuclein, and in several cases from material purified directly from patient tissue; the number of such structures has grown considerably in the years since.

The structures proved considerably more intricate than had been anticipated and revealed a diversity of architectures all conforming to the cross-β fold. The comparison drawn by Iadanza and colleagues is instructive: some 1,375 distinct folds are recognized among globular proteins, and the amyloid state may prove comparably varied. It is therefore more accurate to speak of amyloid folds in the plural than of a single amyloid structure (Iadanza *et al.*, 2018).

A direct corollary is polymorphism: the same sequence can assemble into structurally distinct fibrils, differing in the fold of the subunit, the number and arrangement of protofilaments, the fibril width and the crossover distance (Iadanza *et al.*, 2018).

The transthyretin (105–115) peptide illustrates this at the simplest possible level, forming doublet, triplet and quadruplet fibrils that differ only in the number of protofilament pairs they contain (Fitzpatrick *et al.*, 2013).

This behavior is in marked contrast to that of globular proteins, for which a given sequence adopts the same fold every time. The explanation lies in the absence of evolutionary selection. A native fold is the outcome of an optimization carried out over evolutionary time and has a dominant native energy basin; the amyloid architecture has never been selected for, so many arrangements of comparable energy are accessible. The energy landscape of fibril formation is correspondingly more rugged, with parallel assembly pathways and multiple end products. Once a given polymorph has been initiated it nevertheless tends to propagate, because ordered repetitive structures are stable and the kinetic barriers between polymorphs are high, Chiti and Dobson draw the analogy with crystalline systems ((Chiti and Dobson, 2017); (Iadanza *et al.*, 2018)).

The link between fibril structure and disease has since been placed on a considerably firmer footing. Cryo-electron microscopy of tau filaments extracted directly from post-mortem brain tissue has shown that many tauopathies are associated with characteristic and reproducible filament folds, and on this basis Shi and colleagues have proposed a hierarchical, structure-based classification in which related diseases are grouped by the folds they share (Shi *et al.*, 2021) ; whether the fold itself causes the clinical phenotype has not yet been established (Vaquer-Alicea, Diamond and Joachimiak, 2021).

Comparable findings have been reported for α-synuclein. Filaments extracted from the brains of patients with multiple system atrophy differ from those of dementia with Lewy bodies (Schweighauser *et al.*, 2020), and assemblies amplified from patients with Parkinson's disease, multiple system atrophy and dementia with Lewy bodies each carry a distinct signature: inoculated into rat brain, the multiple system atrophy strains proved considerably more potent than those from dementia with Lewy bodies in inducing neurodegeneration and spreading pathology (Van Der Perren *et al.*, 2020).

A comparable picture is emerging for transthyretin. Cardiac fibrils from patients with polyneuropathic ATTRv amyloidosis show marked structural variability, whereas those from patients presenting with cardiomyopathy are structurally homogeneous even across different genotypes(Nguyen, Singh, Afrin, Yakubovska, *et al.*, 2024a); a common fold was subsequently confirmed in the hearts of four patients with the wild-type disease (Nguyen, Singh, Afrin, Singh, *et al.*, 2024b). Fibril structure in this system therefore appears to track clinical phenotype more closely than genotype.

In each case the structures determined from patient material differ from those obtained by aggregation in vitro, confirming that the cellular environment determines which polymorph is formed and reinforcing the caution that a compound optimized against a laboratory-grown fibril may not encounter the same structure in a patient.

Polymorphism has consequences beyond structural biology. Different polymorphs of amyloid-β and of α-synuclein display different cytotoxicity in vitro and different rates of deposition in vivo (Van Der Perren *et al.*, 2020).

Amyloid-β fibrils seeded from the brains of patients with distinct clinical subtypes of Alzheimer's disease adopt different structures, and more than one polymorph may be present within a single patient (Qiang *et al.*, 2017). Nor is the structure necessarily fixed. Time-resolved cryo-electron microscopy has shown that tau filaments pass through polymorphic intermediates before reaching their final fold (Lövestam *et al.*, 2024), and since fibril growth can take years to approach equilibrium, the structure that predominates may change over the course of a disease (Iadanza *et al.*, 2018).

Together these observations offer a possible explanation for the long-standing difficulty that amyloid plaque load correlates poorly with clinical severity, and they carry a warning for drug development: a compound optimized against a polymorph grown in vitro may not encounter the same structure in a patient.

A protein does not pass directly from its soluble state to a fibril. The species that form first are small clusters that retain a structural memory of the monomer they came from, so that

disordered monomers give highly disordered oligomers, partially folded monomers give partially structured ones, and folded monomers give native-like assemblies Figure 1.

Only weak intermolecular contacts hold these early aggregates together. They are correspondingly unstable, and frequently dissociate again to regenerate soluble protein (Chiti and Dobson, 2017).

If aggregation proceeds, two outcomes are possible. The oligomers may reorganize internally into more compact, β-sheet-rich species which subsequently grow, often with further substantial rearrangement, into ordered cross-β fibrils; or they may grow without any major conformational conversion, retaining the structure they already possessed and giving rise to amorphous or native-like deposits (Chiti and Dobson, 2017).

Amyloid is thus not the automatic destination of aggregation but requires a specific conformational conversion step.

The kinetics are characteristically sigmoidal: a lag phase in which nuclei form, a growth phase dominated by elongation and secondary nucleation, and a plateau at steady state Figure 4. Progress is usually followed by thioflavin T fluorescence or by turbidity(Bollati, Natale, *et al.*, 2026).

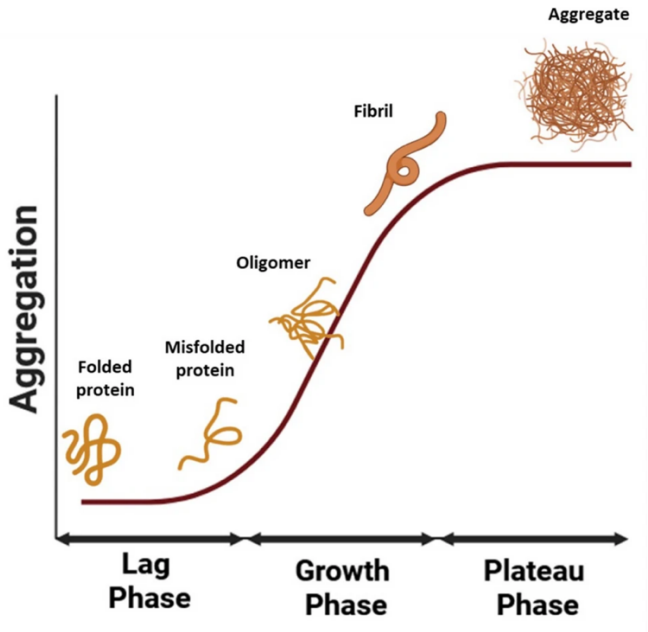


Figure 4 Molecular events underlying amyloid aggregation. A misfolding event exposes aggregation-prone regions, which drive the formation of nuclei and of soluble oligomers already containing β-structure; these grow into fibrils and are eventually deposited as mature aggregates. The kinetics comprise a lag phase in which nuclei form, a growth phase in which elongation and secondary nucleation occur, and a plateau at steady state. Reproduced from Bollati et al. (2026), Cellular and Molecular Life Sciences, 83:177, distributed under CC BY 4.0.

Quantitative analysis of aggregation kinetics has shown that the two phases do not correspond straightforwardly to nucleation and elongation; each result from a combination of primary nucleation, elongation, fragmentation and secondary nucleation, the last of these being the catalysis of new nuclei on the surface of existing fibrils. Fragmentation and secondary nucleation are autocatalytic, since they increase the number of growth-competent species as the reaction proceeds, and for the amyloid-β peptide secondary nucleation has been shown to be the principal contributor to the rapid growth phase (Chiti and Dobson, 2017).

A further practical consequence is that the earliest oligomers bind neither thioflavin T nor Congo red, so that the species most likely to be pathogenic are precisely those to which the standard assay is least sensitive.

This distinction matters because the pathogenic species in the neurodegenerative amyloidosis are now generally held to be the prefibrillar oligomers rather than the mature deposits. Their toxicity correlates with small size, which confers a high diffusion coefficient and hence ready

access to cellular components, and with the extent of hydrophobic surface exposed to solvent (Bollati, Natale, *et al.*, 2026; Chiti and Dobson, 2017).

Mature fibrils are not harmless either. They sequester components of the proteostasis network, they can release oligomers, and they catalyze the formation of further oligomers by secondary nucleation. But the acute damage is attributed to the smaller species.

Solution NMR spectroscopy offers a complementary view of these early stages. Thioflavin T fluorescence reports on the appearance of cross-β structure; NMR instead follows the concentration of free monomer, which starts to fall as soon as aggregation begins. Work on the amyloid-β(1–40) peptide has shown that such measurements can be made at micromolar concentrations under near-physiological conditions, and without the fluorescent or chromophoric probes that may themselves perturb the process (Bellomo *et al.*, 2018; Pagano *et al.*, 2024).

Diffusion-ordered spectroscopy has since been used to follow how the population of oligomeric species is distributed along the same pathway, and to assess the effect of an aggregation inhibitor upon it (Pagano *et al.*, 2024).

Gelsolin amyloidosis is classified on a different basis from the systems described above. Its variants are grouped by mutation site and molecular mechanism rather than by fibril fold, and its aggregation process has not been thoroughly characterized (Bollati et al., 2026).

What determines the phenotype here appears to be settled before a fibril exists, during the processing and aggregation of the precursor. That is also where a therapeutic molecule has to act, and the target is not a straightforward one: proteolytic processing at different domains can release more than one amyloidogenic peptide, and a compound intended to interfere should avoid simply blocking fibril elongation, since that risks allowing the more toxic soluble oligomers to accumulate (Bollati, Natale, *et al.*, 2026)

1.2 Gelsolin Amyloidosis

Gelsolin amyloidosis (AGel), also known as Familial amyloidosis of the Finnish type (FAF), is a rare inherited systemic disease that follows an autosomal dominant pattern.

It is caused by single point mutations affecting different domains of the gelsolin (GSN) protein, which promote its misfolding and the aggregation of the protein or its proteolytic fragments into amyloid deposits that accumulate in various tissues of the body, depending on the mutation, through either a proteolysis-dependent or a proteolysis-independent mechanism (Schmidt, 2020).

Clinically, AGel, first described in Finland in 1969 by the ophthalmologist Meretoja presents a characteristic combination of eye, nerve, and skin symptoms, most commonly corneal lattice dystrophy, cutis laxa, and progressive cranial or peripheral neuropathies. Other organs can also be affected, leading to cardiac conduction problems, kidney involvement, and facial nerve paralysis.

Gelsolin (GSN) is a Ca^{2+} regulated actin-severing and capping protein that functions both intracellularly, where it regulates cytoskeletal remodeling, and extracellularly, where it scavenges actin released from damaged cells.

GSN is composed of six homologous domains (G1–G6), each built around a β-sheet core flanked by α-helices, forming the characteristic "gelsolin fold."

The N-terminal (G1–G3) and C-terminal (G4–G6) halves interact via Ca^{2+}-sensitive β-sheet latches between G2–G1 and G6–G4 (Feldt, 2018). In the absence of Ca^{2+}, GSN adopts a compact, self-inhibited conformation that buries its actin-binding surfaces; Ca^{2+} binding relaxes this packing and exposes the actin-binding interfaces on G1, G2, and G4, enabling actin severing and capping Figure 5 ((Feldt *et al.*, 2018);(Bollati, Peqini, *et al.*, 2026)).

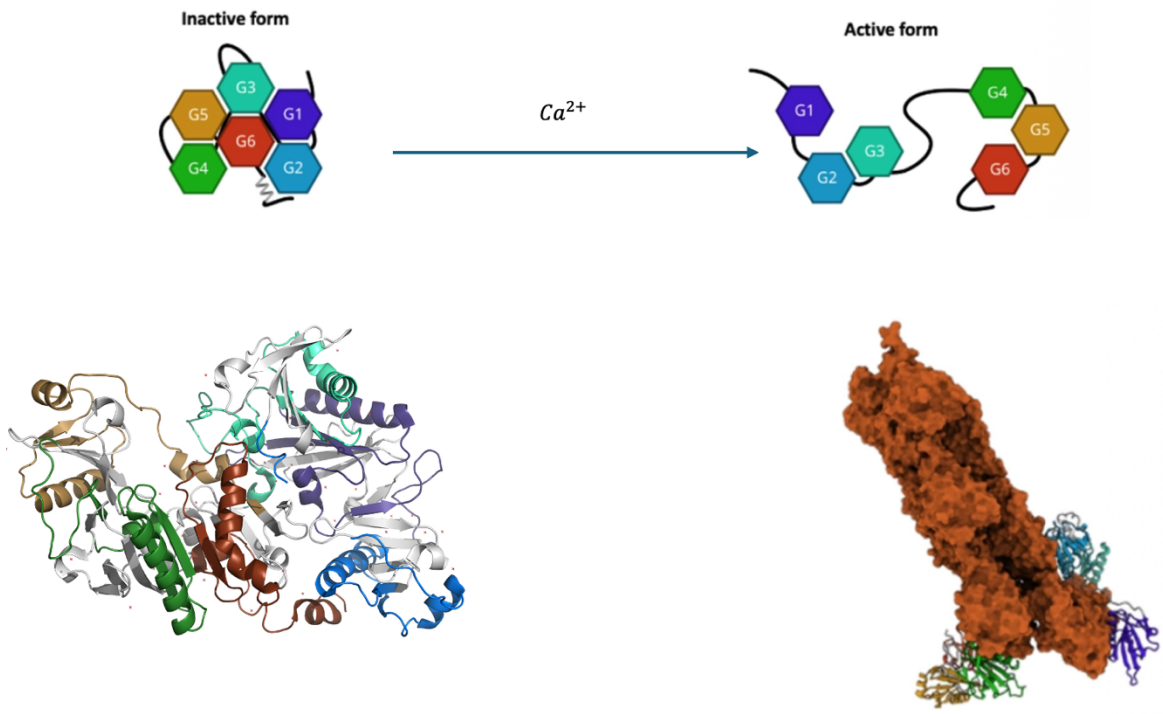


Figure 5 Calcium-dependent conformational transition of gelsolin. (Top) Schematic representation of the compact, Ca^{2+}-free inactive conformation and the open, Ca^{2+}-bound active conformation, in which the domains reorganize into two triplets connected by a flexible linker; reproduced from (Bollati et al., 2026), distributed under CC BY 4.0. (Bottom left) Crystal structure of the closed, Ca^{2+}-free conformation of full-length gelsolin (PDB 3FFN), colored by domain (G1–G6), rendered independently in PyMOL for this thesis. (Bottom right) Cryo-EM structure of the open, Ca^{2+}-bound conformation in complex with the actin filament (shown as surface, PDB 8VIZ), illustrating the capping activity of gelsolin; reproduced from (Bollati et al., 2026).

Amyloidogenic GSN mutations have been identified at several different sites of the protein. The G2 domain hosts the classical hot spot: D187N and D187Y cause the systemic Finnish and Danish forms of AGel, while N184K and G167R are instead associated with a kidney-localized disease. A second cluster of variants (A551P, E553K, M517R, W466R, Y447H/N) lies at the G4:G5 interface and appears to aggregate independently of proteolytic cleavage. More recently, frameshift mutations have been linked to an Alzheimer's disease-like phenotype lacking the classical AGel features, and a single sporadic case has been reported in which wild-type GSN deposits formed around a pituitary tumor (Bollati, Peqini, *et al.*, 2026). Among all these, the G2 domain, and the D187N substitution in particular, is by far the best characterized, both clinically and molecularly, and represents the focus of this thesis.

Within this group, hereditary gelsolin amyloidosis (AGel) is a particularly informative case study: rather than arising from a single well-defined mutation, it results from point

substitutions scattered across different domains of the GSN protein, each triggering aggregation and deposition of the full-length protein or its fragments through apparently distinct molecular routes, and producing a correspondingly heterogeneous clinical picture that can include peripheral neuropathy, muscle weakness, and cardiac or renal involvement.(Bollati, Peqini, *et al.*, 2026).

To date, no specific pharmacological therapy is available for AGel amyloidosis, and clinical management remains largely symptomatic. Surgical and invasive interventions can temporarily alleviate symptoms, but they do not address the underlying molecular cause, the misfolding and aggregation of mutant gelsolin fragments.

Since the pathological process involves both the aggregation-prone peptides derived from furin and metalloprotease cleavage, or the full-length protein, inhibition of these aggregation events represents a promising therapeutic direction.

Loss of calcium binding by the G2 domain has consequences that extend beyond the domain itself. D187 contributes one carboxyl oxygen to the hexacoordinated calcium ion of G2, together with G186, E209, D259 and two water molecules, while its second carboxyl oxygen forms a hydrogen bond with the side chain of K166. Substitution by asparagine or tyrosine abolishes calcium binding and removes these interactions.

The domain is locally destabilized and its association with G3 is disrupted. Because G3 sterically shields a furin cleavage site at 169-RVVR↓A-173, its displacement leaves that site accessible to proteolysis during transit through the trans-Golgi network.

Cleavage releases a 68 kDa C-terminal fragment, C68, which upon secretion is further processed by matrix metalloproteases into two shorter peptides of 8 and 5 kDa, corresponding to residues A173–M243 and A173–R225. Both retain the gelsolin amyloidogenic core (GAC), a segment with a high intrinsic propensity to adopt β-sheet structure. It is these fragments, rather than the intact protein, that assemble into oligomers and subsequently into fibrils. Post-mortem analysis of deposits from D187N and D187Y patients supports this: only the 5 and 8 kDa peptides are recovered, and full-length gelsolin is not detected (Feldt *et al.*, 2018)(Bollati, Peqini, *et al.*, 2026).

The origin of the cascade is a local defect rather than a general loss of stability. Full-length D187N and D187Y gelsolin are structurally almost indistinguishable from the wild-type protein. Their thermal stability is comparable, they adopt the same calcium-free fold, they show no propensity to aggregate when assayed with thioflavin T, and they retain normal actin-severing activity. The single property that changes is their susceptibility to furin, which is markedly increased. The isolated G2 domain behaves differently: once separated from the remainder of the protein it aggregates readily, with no further processing required (Bonì *et al.*, 2016) (Bollati *et al.*, 2021).

Taken together, these observations indicate a two-step mechanism in which local destabilization of G2 permits aberrant proteolysis, and aggregation follows, driven by the released fragments rather than by the parent protein. The pathology therefore does not originate in the misfolding of gelsolin, but in the excision of a small and intrinsically unstable region from an otherwise functional molecule.

1.3 Peptidomimetics as inhibitors of G2 D187N aggregation

Several strategies have been explored to counteract gelsolin misfolding and aggregation, each acting at a different point along the pathogenic pathway. Nanobodies raised against the G2 domain stabilize the native fold of gelsolin and prevent aberrant furin proteolysis, while a second class of nanobodies targets the C68 fragment and protects it from subsequent cleavage. Antigen-binding antibody fragments (Fabs), generated by phage display, intervene later still, recognizing the aggregation-prone 5 and 8 kDa peptides after their release and blocking their self-assembly. A few small molecules have also been reported to modulate gelsolin aggregation in vitro, among them curcumin, emetine, plant-derived cytokinin and methylene blue and the polyphenol EGCG, although their behavior is inconsistent. Curcumin, for instance, accelerates fibril formation rather than preventing it, and most compounds of this class act through broad, non-selective mechanisms with limited specificity for gelsolin(Bollati, Peqini, *et al.*, 2026).

Each approach involves a trade-off. Antibodies and nanobodies are highly specific and effective at stabilizing gelsolin or sequestering its fragments, but their size limits penetration into the tissues where amyloid deposits accumulate; smaller drug-like molecules or peptides retaining chaperone-like function have therefore been proposed as a more efficient alternative (Bollati, 2026).

Small molecules are readily delivered but poorly selective, which raises concerns about off-target effects and translational potential. Peptidomimetics occupy an intermediate position. They are short synthetic constructs designed to reproduce functional peptide motifs or elements of protein secondary structure, and typically show improved chemical stability, protease resistance and pharmacokinetic behavior relative to natural peptides (Bollati *et al.*, 2021). Because they can be built around the sequence and conformation of the amyloidogenic region itself, they combine the selectivity of a designed binder with the size and delivery properties of a conventional drug-like molecule. For this reason, the present work focuses on the peptidomimetic strategy.

The design of these compounds is based on the conformation adopted by the amyloidogenic region within the folded protein. The gelsolin amyloidogenic core, residues 182-SFNNGDCFILD-192 ($GAC_{182\ 192}$), adopts a β- strand conformation and pairs in an antiparallel fashion with a flanking sequence, residues 194- GNNIHQWCGSN-204 (FS_{194} N_{204}), which presents an extended interaction surface. The region encompassing the L191–H198 and F189–W200 pairs approach ideal β-sheet geometry, with the two strands tethered by four hydrogen bonds. The remaining half of the hairpin is more loosely associated, the strands being distorted, although a disulfide bond between Cys188 and Cys201 may contribute to its stabilization (Bollati *et al.*, 2022). The rationale underlying the design is that a molecule reproducing these interactions can compete with the aberrant protein–protein contacts that drive aggregation and thereby preserve the monomeric state. Three peptidomimetics, LB-5, LB-6 and LB-7, were designed on this basis, following an approach previously applied to the aggregation of amyloid-β and of human islet amyloid polypeptide(Bollati *et al.*, 2022). Each incorporates fragments of variable length taken from $GAC_{182\ 192}$ – and $FS_{194\ 204}$, joined by a synthetic piperidine–pyrrolidine scaffold (X) that

imposes a β-turn geometry and – stabilizes the hairpin conformation. LB-6 spans the longest portion of the core (H_2N-NGDCFIL-X-HQWCGSNCONH₂), LB-5 covers its C-terminal region (H_2N-DCFIL-X-HQWCG-CONH₂), and LB-7 only the central segment (H_2N-NGDCF-X-WCGSN-CONH₂). All three were prepared by microwave-assisted solid-phase peptide synthesis on Rink-amide resin using DIC/Oxyma coupling Figure 6.

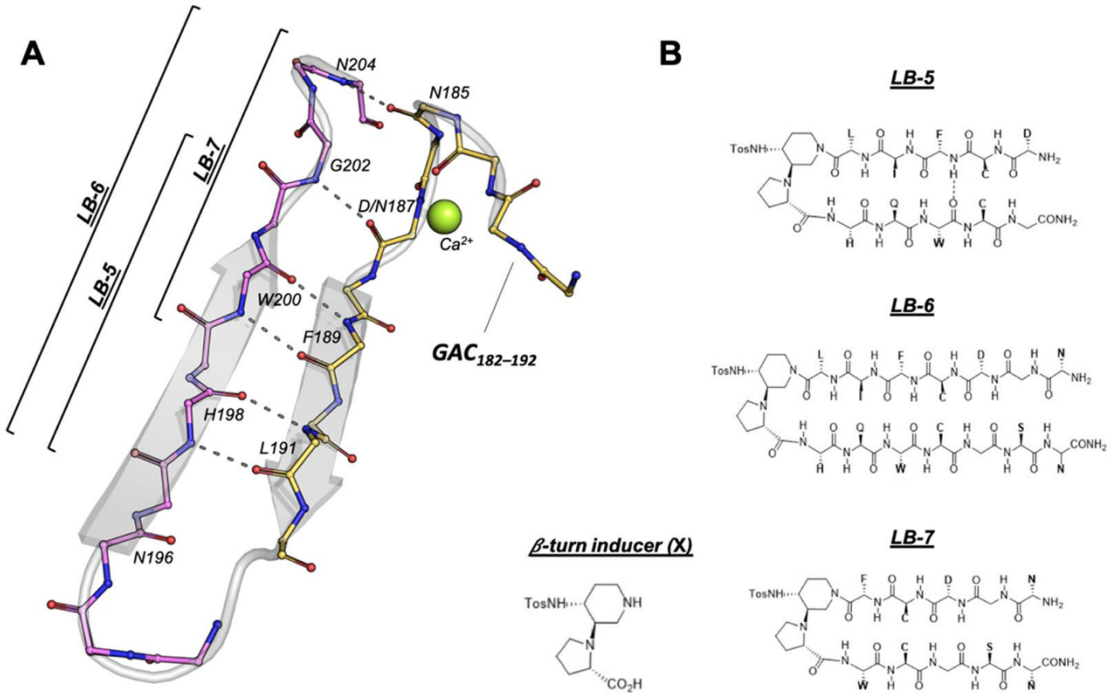


Figure 6 Design and structure of the LB peptidomimetics. (A) Backbone conformation and localization of the gelsolin amyloidogenic core ($GAC_{182–192}$) within the native 3D structure of GSN domain G2 (PDB ID: 6QW3), shown alongside the flanking sequence (residues 194–204) used as a template for peptidomimetic design. (B) Chemical structures of the peptidomimetics LB-5, LB-6, and LB-7, and of the piperidine–pyrrolidine β-turn inducer scaffold (X). Adapted from (Bollati et al., 2022).

Thioflavin T screening distinguished the three compounds clearly. LB-5 and LB-6 inhibited aggregation of $GAC_{182\ 192}$ at sub-stoichiometric concentrations, reducing the end-point fluorescence by approximately 60% at molar ratios of 0.05:1 and 0.1:1 respectively, with complete suppression at 0.5:1 for LB-5 and 1:1 for LB-6. LB-7 was substantially less potent and required a tenfold molar excess for complete inhibition. The activity of LB-6 was not restricted to the isolated core peptide: against the full isolated G2 domain carrying the D187N substitution, a 0.1:1 molar ratio reduced the end-point signal by more than 70%, and equimolar LB-6 abolished it Figure 7.

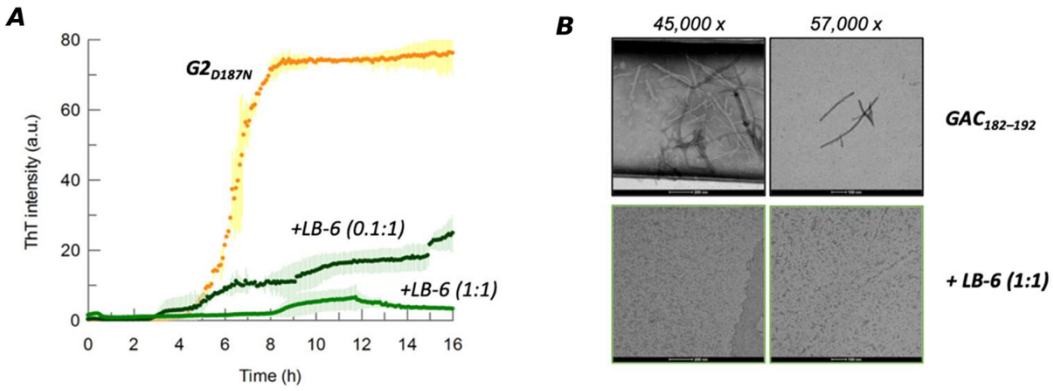


Figure 7 Inhibition of gelsolin amyloid aggregation by LB-6. (A) ThT fluorescence of 75 µM isolated G2 domain carrying the D187N substitution (G2D187N), monitored alone and in the presence of LB-6 at 7.5 µM (0.1:1) and 75 µM (1:1). (B) Representative transmission electron micrographs of 100 µM $GAC_{182–192}$ incubated for 96 h alone (top) and in the presence of 100 µM LB-6 (1:1, bottom), at 45,000× and 57,000× magnification. Adapted from Bollati et al. (2022).

Transmission electron microscopy confirmed the effect at the morphological level, the mature fibrils formed by $GAC_{182\ 192}$ alone being replaced by amorphous, non-fibrillar material in the presence of equimolar LB-6 (Bollati *et al.*, 2022). The same ranking was observed in vivo in a Caenorhabditis elegans proteotoxicity model. LB-5 and LB-6 were well tolerated when administered alone at 2 and 20 µM and produced no measurable pharyngeal toxicity, whereas LB-7 was intrinsically toxic even at the lowest concentration tested. When co-incubated with aggregating $GAC_{182\ 192}$ and then administered, LB-5 and LB-6 both restored pharyngeal pumping to control levels, while LB-7 afforded no protection (Bollati *et al.*, 2022). LB-5 and LB-6 therefore emerged as the two viable candidates, LB-7 being excluded on account of its intrinsic toxicity and weak inhibitory activity.

LB-6 was selected for further structural characterization. Its longer sequence reproduces the native β-strand pairing between the amyloidogenic core and its flanking region more completely than that of LB-5, it showed the strongest inhibitory activity overall, and its efficacy extended to the disease relevant isolated G2 domain carrying the D187N substitution rather than to the core peptide alone.

Despite this biochemical and functional characterization, the molecular basis of the inhibitory activity of LB-6 had not been established. It was not known whether the compound adopts

in solution the β-hairpin conformation for which it was designed, or whether it remains conformationally flexible. The distinction is not merely descriptive. A preorganized hairpin would act as a structural mimic competing directly with the native strand pairing of the amyloidogenic core, whereas a largely disordered molecule would depend on conformational adaptability, engaging an ensemble of aggregation-prone states rather than a single defined structure. The two possibilities imply different strategies for any subsequent optimization of the scaffold, and the work described in this thesis was undertaken to establish which of them applies.

Solution-state nuclear magnetic resonance spectroscopy was employed to address this question. LB-6 is a small, soluble molecule; crystallography requires a well-ordered lattice and cryo-electron microscopy a large assembly, so neither is applicable, whereas solution NMR provides structural and dynamic information at atomic resolution and is particularly well suited to molecules that populate an ensemble of conformations rather than a single structure. Secondary chemical shift analysis was used to detect residual secondary structure, and nuclear Overhauser effect measurements to identify through-space contacts. Together these experiments were intended to determine whether LB-6 adopts a β-hairpin-like conformation in solution or behaves as an intrinsically disordered peptide.

2. Solution Nuclear Magnetic Resonance Spectroscopy for structural analysis of peptides

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical technique that exploits the magnetic properties of atomic nuclei to provide structural and dynamic information on molecules in solution. Among the NMR-active nuclei relevant to biopolymers, four isotopes with spin I = 1/2 are of particular importance: ^{1}H, ^{13}C, ^{15}N, and ^{31}P. Solution-state NMR of biological macromolecules relies predominantly on the observation of ^{1}H resonances, owing to the outstanding properties of this isotope. First, ^{1}H exhibits the highest relative and absolute sensitivity among these nuclei, making it the most readily observable experimentally. Second, because hydrogen atoms occupy peripheral positions within molecular structures, ^{1}H signals are markedly sensitive to non-bonding interactions, such as those between distinct molecular segments in globular proteins or

between individual strands in nucleic acid duplexes. For ^{13}C and ^{15}N, the low natural abundance lowers their intrinsic sensitivity; this same property, however, can be turned to advantage, enabling site-specific observation through isotope labelling.

NMR spectroscopy offers distinctive advantages as a complementary approach to X-ray crystallography. Whereas X-ray diffraction requires crystalline samples, NMR studies are carried out on non-crystalline samples in solution in aqueous or non-aqueous solvents or on detergent-solubilized biopolymers in mixed micelles. When NMR assignments and spatial structures are determined independently, without reference to a corresponding crystal structure, a meaningful comparison of molecular conformation in the crystalline and non-crystalline states becomes possible. Moreover, NMR can be applied to molecules for which no crystals are available. Solution conditions such as pH, temperature, ionic strength, and buffer composition can be varied over a wide range, opening the way to comparative studies under native and denaturing conditions and to the investigation of intermolecular interactions with other species in solution. For the characterization of internal dynamics, NMR provides direct measurements of motional processes, yielding at least semi-quantitative information on the high-frequency dynamics that X-ray crystallography cannot readily access.

2.1 NMR Experiments for peptide resonance assignment: 1D, TOCSY, ROESY, NOESY, and HSQC

A preliminary one-dimensional (1D) ^{1}H spectrum is generally acquired on peptide sample to evaluate sample quality, assess resonance dispersion across the chemical shift range, and detect the possible presence of conformational heterogeneity. The 1D spectrum alone, however, is insufficient for the complete resonance assignment of all but the simplest peptides. The simultaneous projection of backbone amide (HN), Hα, and aliphatic side-chain resonances onto a single chemical shift axis spanning 0 –12 ppm produces extensive spectral overlap that renders the identification of individual signals practically impossible. This limitation does not arise from field strength but reflects the intrinsic chemical similarity of backbone proton environments across residues of different sequence.

Two-dimensional NMR spectroscopy overcomes this problem by distributing resonances across two orthogonal frequency axes, such that each signal is characterized by a unique pair of chemical shift coordinates. Moreover, cross-peaks in 2D spectra arise exclusively between nuclei sharing a defined physical relationship, either scalar coupling through covalent bonds or through-space dipolar coupling occurring for nuclei at short distance in space transforming the spectrum into a rich source of structural connectivity information.

Total Correlation Spectroscopy (TOCSY) establishes correlations between all protons within a common scalar coupling network, thereby identifying the complete spin system of each amino acid residue(Hinds and Norton, 1997). Cross-peaks connect the backbone amide proton HN with all aliphatic side-chain protons (Hα, Hβ, Hγ, Hδ) of the same residue through continuous J-coupling pathways, irrespective of the number of intervening bonds. The characteristic spin-system topologies on which this identification rests are described in Section 2.3; in TOCSY-based analysis, the observed cross-peak multiplicity patterns are matched to the expected topology of each residue type.

Nuclear Overhauser Effect Spectroscopy (NOESY) reports through-space dipolar interactions between proton pairs located within approximately 5.0 Å in the three-dimensional structure, independent of covalent connectivity ((Wüthrich, 1986);(Hinds and Norton, 1997)). Sequential cross-peaks between protons of adjacent residues (d_αN, d_NN connectivity) provide the basis for sequential assignment, while medium- and long-range NOEs between sequentially distant residues report directly on three-dimensional folding.

Rotating-Frame Overhauser Effect Spectroscopy (ROESY) provides equivalent through-space distance information measured in the rotating frame, rendering it particularly advantageous for molecules of intermediate molecular weight such as peptides, where the standard NOE approaches zero owing to the molecular tumbling regime(Hinds and Norton, 1997)(Bax and Davis, 1985). Because ROESY cross-peaks are invariably of opposite sign relative to the diagonal, the experiment permits their unambiguous discrimination from TOCSY relay artefacts that may otherwise contaminate the same spectral regions.

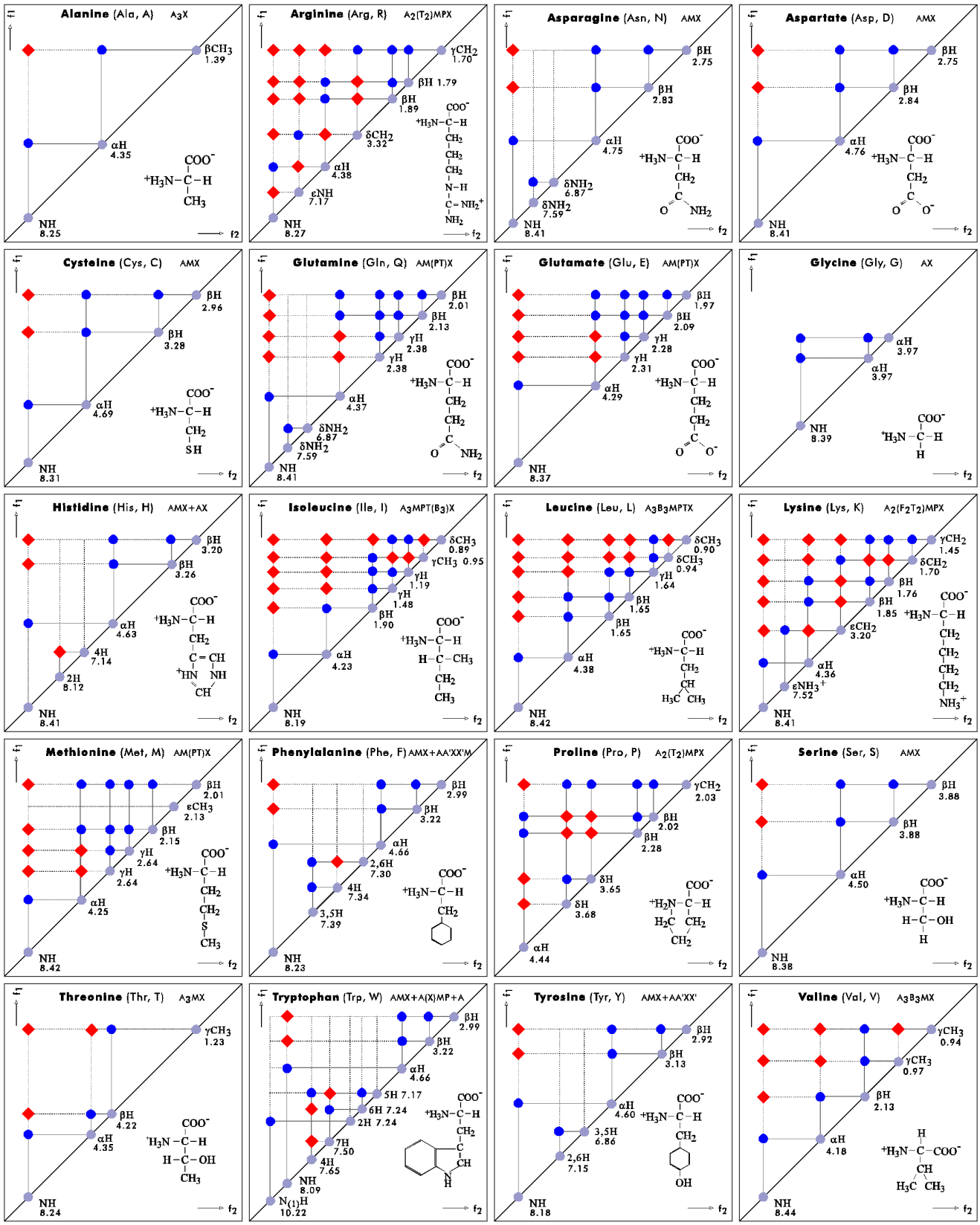
^{1}H–^{13}C Heteronuclear Single Quantum Coherence spectroscopy (HSQC) correlates each ^{1}H nucleus with its directly bonded ^{13}C through one-bond scalar coupling. The resulting fingerprint spectrum resolves H–C pairs across both proton and carbon chemical shift dimensions, overcoming overlaps that persist in the homonuclear spectra. The Hα–Cα cross-peaks are of structural significance, as Cα resonances are shifted downfield in α-helical regions and up-field in β-strand regions relative to random coil reference values. In case the peptide contains 13C at natural isotopic abundance (1.1%), acquisition of the 1H–13C HSQC spectrum requires the use of a cryogenic probe to maximize sensitivity and sample concentrations of at least ~300 μM to obtain spectra with an adequate signal-to-noise ratio(Chen, Freedberg and Keire, 2015).

2.2 Spin Systems in Amino Acid Residues

A central concept in the NMR analysis of peptides and proteins is that of the spin system: a group of spins connected by scalar (through-bond) spin–spin J coupling. In biological macromolecules, such couplings are observable in practice only between hydrogen atoms separated by three or fewer covalent bonds (Wüthrich, 1986) and can be visualized through COSY experiments. TOCSY (Total Correlation Spectroscopy) experiment uses an isotropic mixing period during which magnetization is transferred through a network of scalar-coupled spins. As a result, magnetization is relayed stepwise from one nucleus to the next, allowing correlations to be observed between nuclei that are not directly coupled but belong to the same spin system. Consequently, TOCSY can reveal connectivity extending well beyond three covalent bonds, provided that a continuous pathway of scalar couplings links the nuclei.

The spin systems of all common amino acid residues are well characterized. Glycine, alanine, valine, leucine, isoleucine, threonine, and lysine each possess a unique spin system that permits their identification in a protein spectrum. related protons, while histidine and tryptophan show different symmetry, reflecting their distinct ring structures. the aromatic residues present particularly distinctive patterns: phenylalanine and tyrosine contain two pairs of symmetry-related protons, while histidine and tryptophan show different symmetry, reflecting their distinct ring structures.

Figure 8 Schematic spin-system diagrams reproduced from the NMR course material of the Biophysics group, University of Bayreuth (bp.uni-bayreuth.de/NMR).



2.3 Strategies for peptide resonance assignment

Two nuclei are said to belong to the same spin system when they are connected by scalar (J) coupling. Because such coupling does not extend across the peptide bond, each amino acid residue in a peptide chain constitutes an independent spin system, comprising the backbone amide proton, Hα, Hβ, and the remaining side-chain protons specific to that residue. This absence of scalar coupling between adjacent residues is what allows individual amino acids to be recognized and assigned separately within a crowded NMR spectrum.

Resonance assignment accordingly proceeds in two stages: the identification of the spin system belonging to each residue, followed by the sequential placement of these spin systems along the peptide backbone.

Spin systems are identified by tracing the network of correlations present in the TOCSY spectrum. Beginning from the amide NH resonance or, for residues lacking an amide proton, from a distinctive side chain signal all scalar-coupled protons of that residue are identified in turn, from Hα through Hβ and, where applicable, further along the side chain. The resulting connectivity pattern is frequently diagnostic of residue type in itself: glycine, possessing no side chain, displays only an Hα resonance; valine, leucine and isoleucine each give rise to a characteristic cluster of up field methyl signals; and residues bearing an aromatic ring are readily distinguished by resonances occurring further downfield, in the aromatic region of the spectrum. COSY cross peaks are used to confirm connectivity within each spin system, such as the Hα–Hβ coupling and cross-referencing the TOCSY and COSY data helps to identify signal overlaps that might otherwise be misinterpreted. As an initial check on residue assignment, the observed NH, Hα and Hβ chemical shifts are compared with the random-coil values reported by (Wüthrich, 1986), Table 2.3); the reference spin-system diagrams of the twenty standard amino acids **Error! Reference source not found.**, provide additional support, particularly in assigning the remaining protons of longer side chains.

Identification of the spin systems alone does not establish their order within the sequence; this requires information obtained from NOESY or, in the case of small peptides ROESY since NOE intensities can approach zero near the correlation-time crossover characteristic of

low-molecular-weight compounds. Both experiments detect through-space proximities between protons of neighboring residues. Two categories of cross peak are of relevance: the sequential NH(i)–Hα(i−1) correlation, which links a given residue to its predecessor in the chain, and the intra residue Hβ(i)-HN(i+1) correlation, which corroborates the identity of the spin system itself.

By comparing the resulting pattern of sequential contacts with the known amino acid sequence, each spin system can be assigned to a specific position within the peptide.

Where spectral overlap prevents unambiguous distinction between two spin systems, spectra recorded under altered conditions of temperature or pH can improve peak dispersion sufficiently to resolve the ambiguity. The assignment is regarded as complete only once every spin system has been mapped to a single residue and all observed NOE or ROE contacts are accounted for within the resulting sequence.

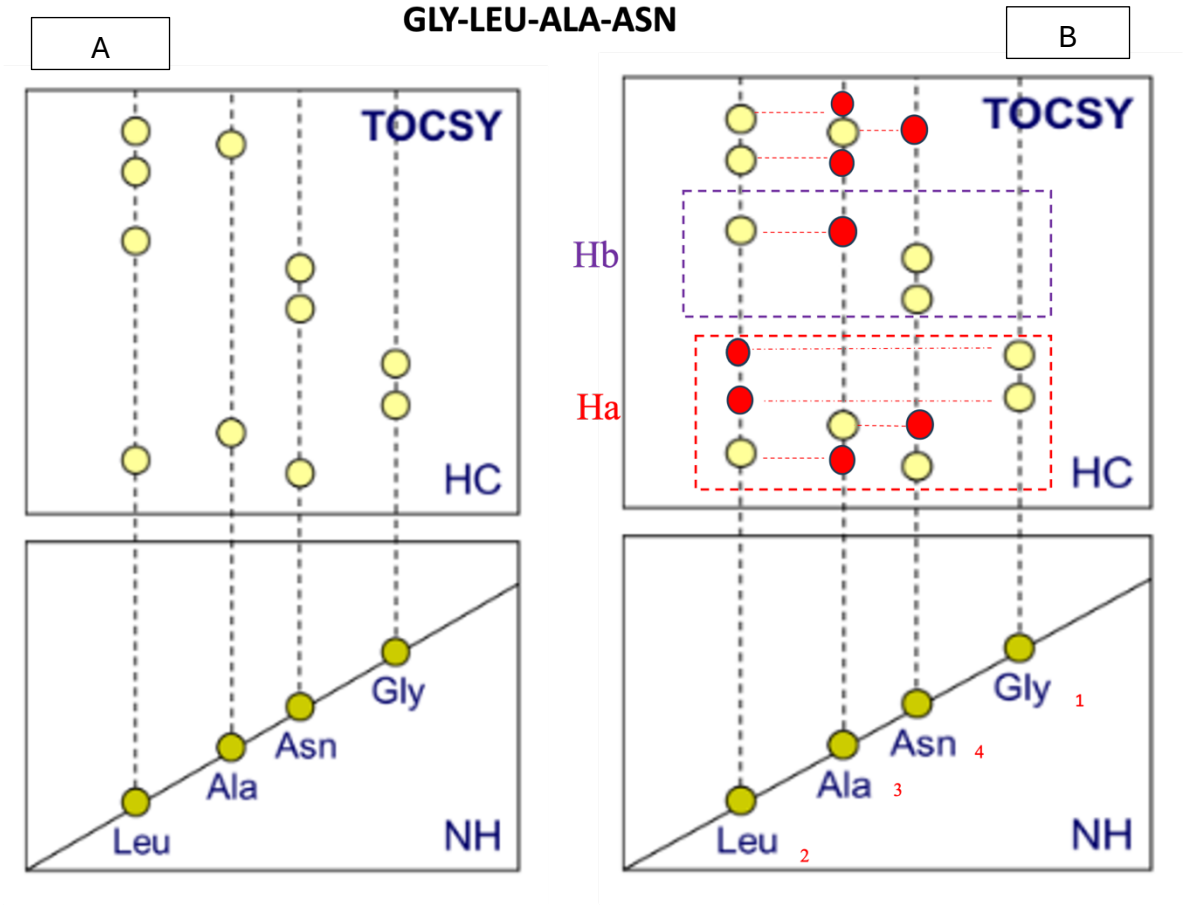


Figure 10 Sequential assignment strategy for the model peptide GLY-LEU-ALA-ASN, illustrated on homonuclear 2D NMR spectra. In both panels the upper box corresponds to the fingerprint (HN-aliphatic) region and the lower box to the HN diagonal region; vertical dashed lines mark the amide proton frequency of each residue. (A) TOCSY spectrum. Cross-peaks (yellow) connect the amide proton of each residue with all the protons of the same spin system, so that the residue type can be identified from its typical pattern of Hα and side-chain signals. (B) Sequential analysis. NOE cross-peaks (red, joined by red dashed lines) give the inter-residue connectivities Hα(i)-HN(i+1) and Hβ(i)-HN(i+1); the corresponding Hα and Hβ regions are enclosed in the red and purple dashed boxes. Following these connectivities from one residue to the next (numbering 1-4) allows the chain to be walked along and the sequence-specific assignment Gly1-Leu2-Ala3-Asn4 to be obtained.

2.4 Random Coil Chemical Shifts and Secondary Chemical Shift Analysis

The secondary chemical shift, defined as $\Delta\delta = \delta_{obs} - \delta_{rc}$, quantifies the deviation of an observed nucleus from the random coil baseline expected for a fully disordered polypeptide. These deviations are amongst the most sensitive and widely used reporters of local conformational propensity in solution NMR spectroscopy. Negative $\Delta\delta$ values for $^{1}H\alpha$ nuclei

are diagnostic of α-helical character, whilst positive deviations indicate extended or β-strand conformations (Bax and Davis, 1985). For $^{13}C\alpha$ and $^{13}C'$ nuclei, the relationship is reversed: downfield shifts relative to random coil report on helical propensity, whilst up field shifts indicate extended conformation (Wishart, 2011).

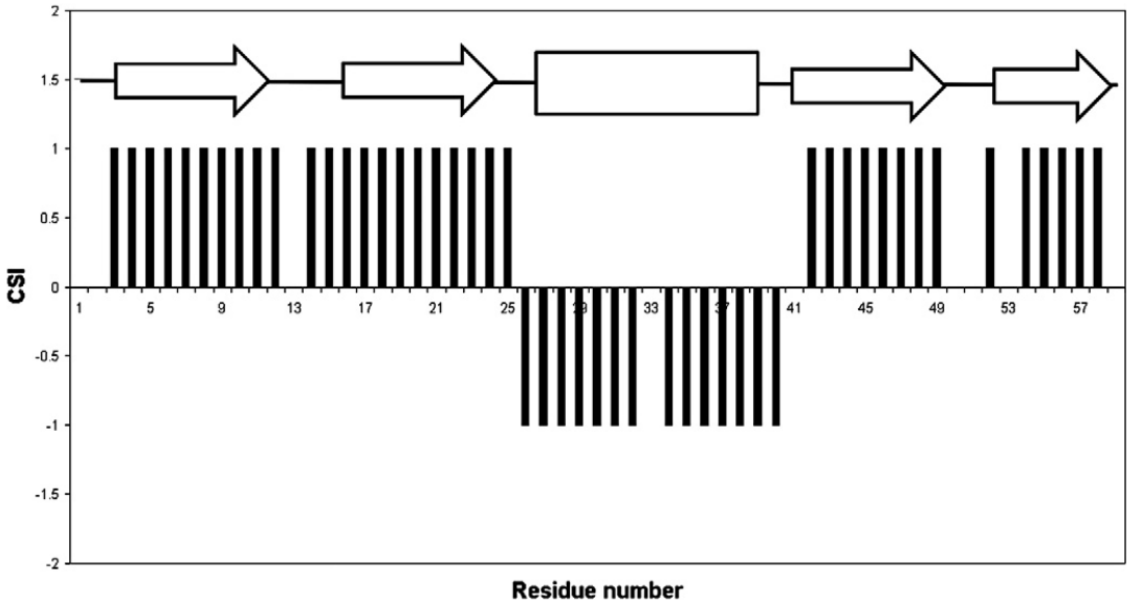


Figure 11 A consensus chemical shift index (CSI) plot generated for protein G (PDB ID 1GB1). The actual secondary structure of the protein is indicated by arrows (b-strands) and rectangles (helices). Note that in a CSI plot the helices are identified by "1" bars and b-strands are identified by "+1" bars. (D.S. Wishart / Progress in Nuclear Magnetic Resonance Spectroscopy 58 (2011) 62–87)

The diagnostic power of secondary chemical shifts derives from the exquisite sensitivity of backbone chemical shifts to the local electronic environment, which is modulated by backbone dihedral angles φ and ψ. Even in intrinsically disordered systems, where no single conformation predominates, secondary chemical shift analysis can reveal transient structural propensities and population-weighted conformational preferences, information inaccessible by other solution-state techniques at comparable resolution.

The reliability of secondary chemical shift analysis depends critically on the quality and appropriateness of the random coil reference dataset against which observed shifts are compared. An ideal reference must satisfy several conditions simultaneously: it must be measured under solution conditions (pH, temperature, ionic strength) comparable to those of

the system under study; it must be derived from genuinely unstructured model systems that faithfully represent the intrinsic chemical shift of each amino acid in the absence of secondary or tertiary structure; and it must account for nearest-neighbors sequence effects, since the chemical shift of a given residue is modulated by the identity of its flanking neighbors even in a disordered context. These requirements have motivated successive generations of random coil reference datasets, each addressing specific shortcomings of its predecessors. Four principal reference datasets were used in this thesis as reference and are discussed **below.**

1) Kjaergaard et al. volume 49 (2011), Experimental dataset at 25°C, pH6.5

This dataset represents a significant methodological advance over earlier references. Random coil chemical shifts were measured directly on a systematic series of Ac-GGXGG-NH_2 pentapeptides, where X is each of the 20 canonical amino acids, in aqueous solution containing 1 M urea at pH 6.5 and 25°C. The use of near-neutral pH is particularly important: earlier peptide-based references, most notably those of Schwarzinger et al. (2001) and Wishart et al. (1995), were acquired at acidic pH values (pH 2–5), under which conditions ionizable residues, Asp, Glu, and His in particular, are fully or partially protonated. Protonation alters the electron density around the backbone nuclei of these residues and their immediate neighbors, introducing systematic errors of up to 0.5 ppm in secondary chemical shift profiles for sequences containing these residues. The Kjaergaard dataset substantially reduces this source of bias by working at pH 6.5, where Asp and Glu (sidechain pKa ≈ 4) are fully deprotonated. Histidine, whose side-chain pKa was determined by titration to be 6.9, remains partially protonated at pH 6.5; rather than assuming a single protonation state, (Kjaergaard, Brander and Poulsen, 2011)performed a dedicated pH titration of the histidine-containing peptide and fitted the data to extract the random coil shifts of the fully protonated and fully deprotonated species, allowing the histidine reference value to be calculated accurately at any pH rather than relying on an uncontrolled, partially protonated average. In addition, the dataset provides explicit per-residue nearest-neighbour sequence correction factors for all backbone nuclei ($^{1}H\alpha$, ^{1}HN, $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$), enabling correction for the sequence-context dependence of random coil shifts that is otherwise a source of systematic

noise in secondary chemical shift profiles.

Table 1 Random coil chemical shifts and temperature coefficients at pH 6.5, 1 M urea and 25°C.*n.d., not determined; –, not applicable (Pro has no amide NH; Gly has no Cβ).*

	δRC (ppm)						**Temperature coefficients (ppb/K)**					
	Cα	**Cβ**	**C'**	**N**	**H^N**	**Hα**	**Cα**	**Cβ**	**C'**	**N**	**H^N**	**Hα**
A	52.77	19.18	178.64	124.14	8.35	4.37	-2.2	4.7	-7.1	-5.3	-9.0	0.7
C	58.64	28.12	175.49	118.74	8.42	4.59	-0.9	1.3	-2.6	-8.2	-7.0	0.0
D	54.42	41.29	177.25	120.43	8.39	4.65	2.8	6.5	-4.8	-3.9	-6.2	-0.1
E	57.02	30.09	177.52	121.03	8.55	4.32	0.9	4.6	-4.9	-3.7	-6.5	0.3
F	58.06	39.55	176.76	120.35	8.26	4.65	-4.7	2.4	-6.9	-11.2	-7.5	0.4
G	45.37	–	175.08	109.04	8.42	4.02	3.3	–	-3.2	-6.2	-9.1	0.0
H	55.88	29.83	175.75	118.78	8.41	4.73	7.8	15.5	3.1	3.3	-7.8	0.4
I	61.57	38.71	177.29	120.08	8.13	4.21	-2.0	4.6	-8.7	-12.7	-7.8	0.4
K	56.54	32.98	177.48	120.99	8.33	4.37	-0.8	2.4	-7.1	-7.6	-7.5	0.4
L	55.48	42.32	178.39	121.83	8.27	4.38	1.7	4.9	-8.2	-2.9	-7.5	0.1
M	55.71	32.75	177.18	119.95	8.40	4.54	4.1	9.4	-8.2	-6.2	-7.1	-0.5
N	53.63	38.97	176.14	118.88	8.49	4.78	2.8	5.1	-6.1	-3.3	-7.0	-2.9
P	63.71	32.09	178.08	n.d.	–	4.46	1.1	-0.2	-4.0	n.d.	–	0.0
Q	56.14	29.38	176.88	120.04	8.43	4.39	2.3	3.6	-5.7	-6.5	-7.2	0.3
R	56.38	30.71	177.23	120.72	8.35	4.38	-1.4	3.5	-6.9	-5.3	-7.1	0.4
S	58.60	63.90	175.43	116.04	8.41	4.51	-1.7	4.4	-4.7	-3.8	-7.6	0.1
T	62.06	69.84	175.64	113.05	8.24	4.41	0.0	2.2	-5.2	-6.7	-7.3	0.0
V	62.32	32.64	177.21	119.30	8.13	4.17	-1.9	1.6	-8.1	-14.2	-7.6	0.5
W	57.59	29.51	177.18	121.26	8.16	4.70	-2.7	3.1	-7.9	-10.1	-7.8	0.4
Y	58.24	38.73	176.82	120.48	8.21	4.58	-5.0	2.9	-7.7	-12.0	-7.7	0.5

Table 2 Sequence correction factors for 13C at pH 6.5, 1 M urea and 25°C.

	Cα				**C'**			
	A	**B**	**C**	**D**	**A**	**B**	**C**	**D**
A	-0.01	-0.15	0.08	0.02	-0.11	-0.75	-0.04	-0.02
C	0.00	-0.05	0.12	0.01	-0.08	-0.46	-0.24	-0.06
D	-0.09	-0.04	0.33	0.07	-0.11	-0.75	0.16	0.05
E	0.00	-0.14	0.14	0.05	-0.11	-0.44	-0.02	0.01
F	-0.05	-0.24	0.07	0.01	-0.30	-0.83	-0.23	-0.10
G	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H	-0.05	-0.11	0.07	-0.01	-0.14	-0.71	-0.04	-0.09
I	-0.05	-0.19	0.03	0.02	-0.23	-0.56	-0.15	-0.03
K	0.02	-0.01	-0.06	0.01	-0.08	-0.49	-0.19	-0.03
L	-0.06	-0.19	0.05	-0.05	-0.15	-0.49	-0.10	-0.01
M	0.02	-0.03	0.11	0.01	-0.09	-0.42	-0.16	-0.02
N	-0.05	-0.02	0.24	0.02	-0.09	-0.71	-0.08	-0.03
P	-0.22	-0.79	0.03	0.05	-0.56	-2.80	-0.06	-0.02
Q	0.00	0.00	0.03	0.02	-0.04	-0.48	-0.15	-0.03
R	0.02	-0.03	0.00	0.01	-0.04	-0.44	-0.19	-0.02
S	-0.01	-0.05	0.13	0.00	-0.07	-0.40	-0.13	-0.05
T	0.00	-0.04	0.13	0.00	-0.08	-0.20	-0.13	-0.05
V	-0.04	-0.17	0.02	0.02	-0.23	-0.53	-0.15	-0.02
W	-0.08	-0.19	0.11	-0.08	-0.30	-0.56	-0.31	-0.19
Y	-0.03	-0.24	0.07	0.00	-0.30	-0.86	-0.23	-0.16

2) Kjaergaard et al. volume 49 (2011), temperature-corrected dataset at 15°C

Backbone ^{13}C chemical shifts show a measurable, approximately linear dependence on temperature. Applying a 25°C random coil reference directly to data recorded at 15°C, a 10°C mismatch, therefore introduces a systematic offset that varies by nucleus and by residue. To correct for this, (Kjaergaard, Brander and Poulsen, 2011)measured per-residue temperature coefficients (dδ/dT) for all backbone nuclei (Table 1 of that study), which allow the reference values to be extrapolated to any desired temperature.

The size of this effect differs considerably between nuclei. $^{13}C\alpha$ coefficients are generally small, on the order of a few ppb/K, and can be of either sign, so a temperature mismatch is unlikely to bias the analysis in a consistent direction. $^{13}C'$ and $^{13}C\beta$ coefficients, in contrast, are larger and consistently signed (Fig. 2), meaning that a 10°C mismatch can shift these reference values by up to approximately 0.1 ppm, large enough to obscure genuine secondary-structure propensities, particularly in disordered systems where the secondary shifts of interest are inherently small.

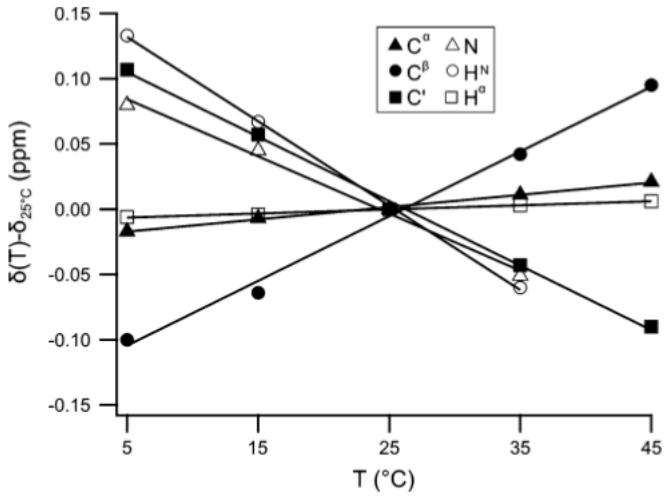


Figure 12 Temperature dependence of the random coil chemical shifts of glutamate at pH 6.5, illustrating the larger, consistently signed drift of C' and Cβ relative to Cα (adapted from Kjaergaard, Brander & Poulsen, 2011).

For this reason, temperature correction was applied to all data recorded away from 25°C, using the following equation:

(1) $\delta rc\ (T) = \delta rc\ (25°C) + a \times (T - 25)$

3) Wishart (2011), consensus experimental values (Table 4)

This dataset is a consensus of experimentally measured random coil chemical shifts for all 20 amino acids, built by pooling a series of independent peptide-based studies (Gly-Gly-Xaa-Gly-Gly and related pentapeptides, measured under a range of solvent conditions) that were uniformly referenced to the International Union of Pure and Applied Chemistry / International Union of Biochemistry and Molecular Biology (IUPAC/IUBMB) standards (DSS for $^{1}H/^{13}C$, liquid ammonia for ^{15}N) and consolidated into a single internally consistent table at 25°C, pH 5.0. Because it draws on multiple independent experimental measurements rather than a single study, and because of its long-standing adoption in the field, it remains one of the most widely used benchmarks for comparison with the broader protein NMR literature.

The choice of pH 5.0 does, however, introduce an important interpretive consideration for the pH-sensitive residues. Using their sidechain pKa values (Asp ≈ 3.9, Glu ≈ 4.1), pH 5.0 leaves both residues roughly 90% deprotonated but still around 10% protonated, a small residual population rather than complete deprotonation. His (pKa ≈ 6.0) shows the mirror-image effect: at pH 5.0 it is roughly 90% protonated, predominantly in its positively charged imidazolium form, rather than simply "intermediate." Kjaergaard et al. (2011) quantified this explicitly when comparing their own pH 6.5 dataset to Wishart's pH 5.0 values, confirming that both Asp/Glu and His carry a measurable, pH-dependent bias relative to data collected at neutral pH. Comparison with the Wishart (2011) dataset therefore allows pH-sensitive contributions to the secondary chemical shift profile to be identified, while still benchmarking against the wider NMR literature, for which pH 5.0 reference data remain the most widely available.(Wishart, 2011)

Table 3

"Random coil" chemical shifts (in ppm) for the 20 common amino (and phosphor amino) acids measured at 25 °C, pH 5.

Amino acid	**^{1}HN**	**^{15}N**	**^{1}Hα**	**^{13}Cα**	**^{13}Cβ**	**^{13}C'**
Ala	8.24	123.8	4.32	52.5	19.1	177.8
Cys (red)	8.32	118.8	4.55	58.2	28.0	174.6
Cys (ox)	8.43	118.6	4.71	55.4	41.1	174.6
Asp	8.34	120.4	4.64	54.2	41.1	176.3
Glu	8.42	120.2	4.35	56.6	29.9	176.6
Phe	8.30	120.3	4.62	57.7	39.6	175.8
Gly	8.33	108.8	3.96	45.1	N/A	174.9
His	8.42	118.2	4.73	55.0	29.0	174.1
Ile	8.00	119.9	4.17	61.1	38.8	176.4
Lys	8.29	120.4	4.32	56.2	33.1	176.6
Leu	8.16	121.8	4.34	55.1	42.4	177.6
Met	8.28	119.6	4.48	55.4	32.9	176.3
Asn	8.40	120.4	4.73	53.1	38.9	175.2
Pro (trans)	N/A	N/A	4.42	63.3	32.1	177.3
Pro (cis)	N/A	N/A	4.42	62.8	34.5	177.3
Gln	8.32	119.8	4.34	55.7	29.4	176.0
Arg	8.23	120.5	4.34	56.0	30.9	176.3
Ser	8.31	115.7	4.47	58.3	63.8	174.6
pSer	*8.65*	*115.5*	*4.60*	*57.4*	*66.8*	*174.9*
Thr	8.15	113.6	4.35	61.8	69.8	174.7
pThr	*8.43*	*113.4*	*4.43*	*61.8*	*74.0*	*175.1*
Val	8.03	119.2	4.12	62.2	32.9	176.3

Amino acid	^{1}HN	^{15}N	^{1}Hα	^{13}Cα	^{13}Cβ	^{13}C'
Trp	8.25	121.3	4.66	57.5	29.6	176.1
Tyr	8.12	120.3	4.55	57.9	37.8	175.9
pTyr	*8.21*	*120.1*	*4.61*	*57.0*	*38.7*	*176.7*

4) <u>Zhang, Neal & Wishart (2003), statistically derived database values, as tabulated in Wishart (2011, Table 5)</u>

Rather than measuring random coil shifts directly on model peptides, this approach derives reference values statistically from a large, curated database of protein chemical shift assignments. The underlying calculation was originally performed by Zhang, Neal and Wishart (2003), who extracted secondary structure-stratified chemical shift averages separately for α-helix, β-sheet, and random coil from RefDB (Re-referenced Protein Chemical shift Database), a repository built from BioMagResBank entries whose chemical shift referencing had been systematically corrected using predicted shifts from X-ray and NMR coordinate data. (Wishart, 2011)reproduces this dataset in Table 5, alongside a review of its main findings.

A total of 229,518 diamagnetic backbone shift assignments were used to compute the reference averages, with the corresponding secondary structure for each residue calculated directly from PDB coordinate files using the program Volume Area Dihedral Angle Reporter(VADAR)(Willard, 2003). The random coil column of this table therefore reflects the average chemical shift of each residue type specifically in segments of real proteins that were computationally assigned no regular secondary structure, rather than in a short, isolated peptide. Because helix and β-sheet averages are calculated on the same basis, the table also provides structure-class-specific benchmarks directly comparable to canonical secondary structure signatures: for ^{13}Cα, helical residues are shifted approximately 2.6 ppm downfield of random coil, whilst β-strand residues are shifted approximately 1.1 ppm upfield.

A key distinction from the Kjaergaard dataset is therefore the source material itself: these values are derived from folded proteins deposited under diverse, uncontrolled experimental

conditions, rather than from a single controlled series of model peptides measured at fixed pH and temperature. This has two consequences. On one hand, the database values may capture genuine environmental effects present in real protein contexts, packing, hydrogen bonding, local electrostatics, that a short, isolated pentapeptide cannot reproduce. On the other, the dataset lacks the explicit pH control and the residue-specific nearest-neighbor correction framework built into the Kjaergaard dataset, so systematic contributions from pH or sequence context are folded into the averages rather than separated out.

Table 4 *Statistically derived reference chemical shifts (in ppm) for backbone shifts in helix, β-sheet and random coil secondary structures.*

Residue type	**$^{13}C_{\alpha}$**			**$^{13}C_{\beta}$**			**$^{13}C'$**			**^{15}N**			**$^{1}H_{N}$**			**$^{1}H_{\alpha}$**		
	Helix	**Sheet**	**Coil**	**Helix**	**Sheet**	**Coil**	**Helix**	**Sheet**	**Coil**	**Helix**	**Sheet**	**Coil**	**Helix**	**Sheet**	**Coil**	**Helix**	**Sheet**	**Coil**
Ala	54.8	51.5	52.8	18.3	21.1	19.1	179.4	176.1	177.7	121.4	124.5	123.6	8.08	8.44	8.15	4.03	4.77	4.26
Cys (ox)	58.0	55.0	55.6	39.4	43.9	41.0	176.2	173.6	174.9	117.7	121.0	118.0	8.20	8.80	8.25	4.15	5.15	3.96
Cys (red)	61.3	56.9	57.5	27.8	30.2	29.4	176.2	173.6	174.9	117.7	121.0	118.0	8.20	8.80	8.25	4.15	5.15	4.65
Asp	56.7	53.9	54.2	40.5	42.3	40.9	178.0	175.5	176.3	119.2	122.2	120.0	8.18	8.51	8.36	4.43	4.94	4.60
Glu	59.1	55.5	56.9	29.4	32.0	30.2	178.6	175.4	176.4	119.0	122.1	120.4	8.22	8.53	8.37	4.07	4.78	4.28
Phe	60.8	56.7	58.0	38.8	41.5	39.5	177.1	174.3	175.6	119.2	121.1	119.7	8.18	8.75	8.17	4.16	5.09	4.54
Gly	46.9	45.2	45.5	N/A	N/A	N/A	175.5	172.6	173.9	107.5	109.3	109.1	8.29	8.34	8.33	3.81	4.20	3.96
His	59.0	55.1	55.9	29.5	31.9	30.0	177.0	174.2	174.8	118.0	120.5	118.7	8.10	8.62	8.21	4.33	5.06	4.53
Ile	64.6	60.1	61.0	37.6	39.9	38.7	177.7	174.9	175.6	119.7	122.8	120.9	8.02	8.68	7.98	3.67	4.68	4.15
Lys	58.9	55.4	56.6	32.3	34.6	32.8	178.4	175.3	176.3	119.2	122.0	120.4	7.99	8.48	8.23	4.09	4.69	4.26
Leu	57.5	54.1	54.9	41.6	43.8	42.4	178.5	175.7	176.9	119.6	124.1	121.5	8.05	8.60	8.08	4.00	4.82	4.36
Met	58.1	54.6	55.7	32.3	35.1	33.4	178.0	174.8	175.4	118.2	121.7	119.7	8.09	8.60	8.18	4.07	4.96	4.38

Residue type	$^{13}C_{\alpha}$			$^{13}C_{\beta}$			$^{13}C'$			^{15}N			$^{1}H_{N}$			$^{1}H_{\alpha}$		
	Helix	Sheet	Coil	Helix	Sheet	Coil	Helix	Sheet	Coil	Helix	Sheet	Coil	Helix	Sheet	Coil	Helix	Sheet	Coil
Asn	55.5	52.7	53.2	38.6	40.1	39.1	176.9	174.6	175.1	117.3	121.6	118.2	8.22	8.60	8.40	4.48	5.06	4.66
Pro	65.5	62.6	63.5	31.5	32.3	31.9	178.3	176.2	176.9	N/A	N/A	N/A	N/A	N/A	N/A	4.22	4.60	4.37
Gln	58.5	54.8	56.1	28.5	31.3	29.1	178.0	174.9	175.9	118.4	121.1	119.5	8.04	8.48	8.23	3.99	4.80	4.26
Arg	58.9	55.1	56.4	30.1	32.2	29.1	178.3	175.1	176.0	118.9	122.3	120.2	8.07	8.56	8.25	3.99	4.74	4.24
Ser	60.9	57.5	58.4	63.1	65.2	64.0	175.9	173.6	174.5	114.9	116.9	115.6	8.14	8.50	8.23	4.25	4.91	4.47
Thr	65.6	61.1	61.6	68.9	70.8	70.1	175.9	173.7	174.7	114.6	116.5	113.4	8.04	8.51	8.16	4.00	4.86	4.45
Val	66.2	60.8	62.1	31.5	33.9	32.7	177.7	174.8	175.7	119.2	121.9	119.8	8.02	8.62	8.04	3.58	4.60	4.12
Trp	60.0	56.4	57.8	29.3	31.5	29.7	178.1	175.4	176.2	119.8	122.1	120.2	8.12	8.59	7.92	4.38	5.19	4.55
Tyr	61.0	56.8	58.0	38.3	41.0	39.0	177.4	175.4	175.4	119.2	121.4	119.5	8.07	8.59	8.06	4.09	5.10	4.52

2.5 CheSPI (Chemical shift Secondary structure Population Inference) program

The four reference datasets described in the preceding section all address the same problem, namely the determination of an appropriate value of δ_rc against which an observed chemical shift may be compared. They do not, however, address the separate and equally important question of how the resulting secondary chemical shifts should be interpreted once they have been obtained.

The analysis presented in Section 2.5 treats each nucleus independently: a profile of Δδ Hα is inspected, a profile of Δδ Cα is inspected alongside it, and conformational conclusions are drawn from the sign, the magnitude and, critically, the persistence of the deviations across consecutive residues. This approach is transparent and long established, but it discards information, and the nature of what is lost becomes limiting precisely in the regime relevant to a small, flexible peptidomimetic (Nielsen and Mulder, 2021).

Three limitations may be identified. The first is that each nucleus reports on the same underlying backbone geometry, so the individual Δδ profiles are strongly correlated rather than independent; inspecting them separately therefore neither exploits the redundancy between them nor provides a principled way of weighting one against another. The second is that the conventional requirement for at least three consecutive residues deviating consistently in the same direction, whilst a necessary safeguard against over-interpretation of isolated excursions, is a binary criterion applied to a continuous quantity. It classifies a segment as structured or not structured and offers no estimate of the fraction of the conformational ensemble that populates the element in question. For a molecule that is expected to interconvert rapidly between conformations, this is the quantity of principal interest.

The third limitation is more subtle and was set out explicitly by (Nielsen and Mulder, 2021) Methods that reduce the set of secondary chemical shifts to a single propensity value on a scale running from −1 for β-sheet to +1 for α-helix, of which the secondary structure propensity (SSP) approach of (Marsh *et al.*, 2006) and its later reparameterization by (Tamiola and Mulder, 2012) are the best known examples, implicitly resolve the problem of correlated shifts, but at a cost.

A rigid loop situated between a sheet and a helix will yield contributions of opposite sign that cancel, producing a propensity close to zero and inviting the erroneous conclusion that the segment is disordered. Neither the index-based nor the propensity-based approaches can discriminate between an ordered turn and genuine disorder, although the information required to do so is in principle contained in the combination of secondary chemical shifts observed.

CheSPI (Chemical shift Secondary structure Population Inference), introduced by(Nielsen and Mulder, 2021), was developed to address these points. It was therefore applied in this thesis as a complement to, and not a replacement for, the reference-based secondary chemical shift analysis described above.

POTENCI: a sequence, temperature and pH-specific random coil reference

CheSPI does not employ a tabulated random coil dataset of the kind discussed above. Instead, it calls POTENCI (Prediction Of Temperature, Neighbor and pH-Corrected shifts), described by (Nielsen and Mulder, 2021), which computes a random coil chemical shift for every nucleus of every residue of the specific sequence under study, at the specific temperature, pH and ionic strength of the experiment. Nearest-neighbor effects are treated at the pentapeptide level, that is, the predicted shift of a residue depends on the identity of the two residues on either side of it; temperature dependence is handled through fitted coefficients rather than by a single global correction; and the protonation state of ionizable side chains is obtained by an explicit pKa calculation at the requested pH and ionic strength, rather than by assuming full protonation or full deprotonation.

CheSPI components

The core of the method is a supervised dimensionality reduction. Two chemical shift databases were used for its parametrization: the first, assembled for the development of POTENCI, contains predominantly disordered residues, whose entries were labelled as disordered or ordered according to whether their CheZOD, Z-score fell below or above a value of 5.0; the second, derived from RefDB (Zhang, Neal and Wishart, 2003) and comprising 809 proteins, contains predominantly structured residues, whose entries were labelled with the eight-class secondary structure assignment returned by DSSP (Kabsch and Sander, 1983). Secondary chemical shifts were computed for the backbone nuclei together with Cβ and Hβ by subtracting the corresponding POTENCI predictions.

Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA), (Bylesjö *et al.*, 2006) was then applied to these labelled data. The procedure resembles principal component analysis but differs from it in an important respect: the first component is constructed to capture the variance between the structural classes, whilst the variance within each class is relegated to the subsequent components. The resulting linear combinations of secondary chemical shifts are termed the CheSPI components Figure 13.

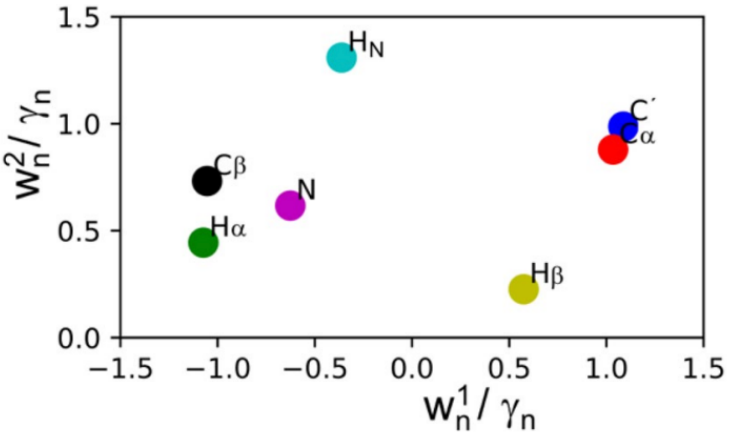


Figure 13 CheSPI component loading plot, showing the optimized weights of each nucleus scaled by its gyromagnetic ratio. Note the dominance of Hα, C′, Cα and Cβ in the first component and of HN and C′ in the second. Reproduced from Nielsen and Mulder (2021).

The optimized weights, scaled by the gyromagnetic ratio of each nucleus, reproduce the sensitivities already familiar from the classical literature: for the first component the ordering is Hα / C′ / Cα / Cβ ≫ N / Hβ / HN, with Hα the most structurally informative proton and the three carbon-13 nuclei of comparable weight. The nuclei N, HN, Hα and Cβ carry weights of opposite sign to C′, Cα and Hβ, which is the formal expression of the well-known inversion of the sign convention between proton and carbon secondary shifts. It is worth noting that the two nuclei on which the analysis in this thesis is based, Hα and Cα, are among those carrying the greatest weight, so that the CheSPI treatment and the reference-based secondary chemical shift analysis described above rest on the same physical information, differently combined.

For the second component all weights are positive, with the ordering HN > C′ / Cα / Cβ > N > Hα > Hβ. The two dominant contributors, HN and C′, are the nuclei previously shown to undergo the largest chemical shift changes upon formation of a backbone hydrogen bond (Nielsen and Mulder, 2021a), and the second component may accordingly be read as a reporter of local hydrogen bonding and backbone rigidity. A third component was found to add nothing of significance to the classification and is not used.

The behavior of the two components is straightforward to summarize. The first component separates helix from extended structure almost perfectly, whereas non-folded conformations sit near the middle of the plot and overlap both classes. The second component is negative for turns and positive for helices and strands, which is what allows an ordered turn to be distinguished from an unstructured segment: the two are resolved along the second component even where the first alone would not separate them. Residues that are genuinely disordered take values close to zero in both components, so that disorder occupies a defined position in the plane rather than being inferred from the absence of a signal Figure 14.

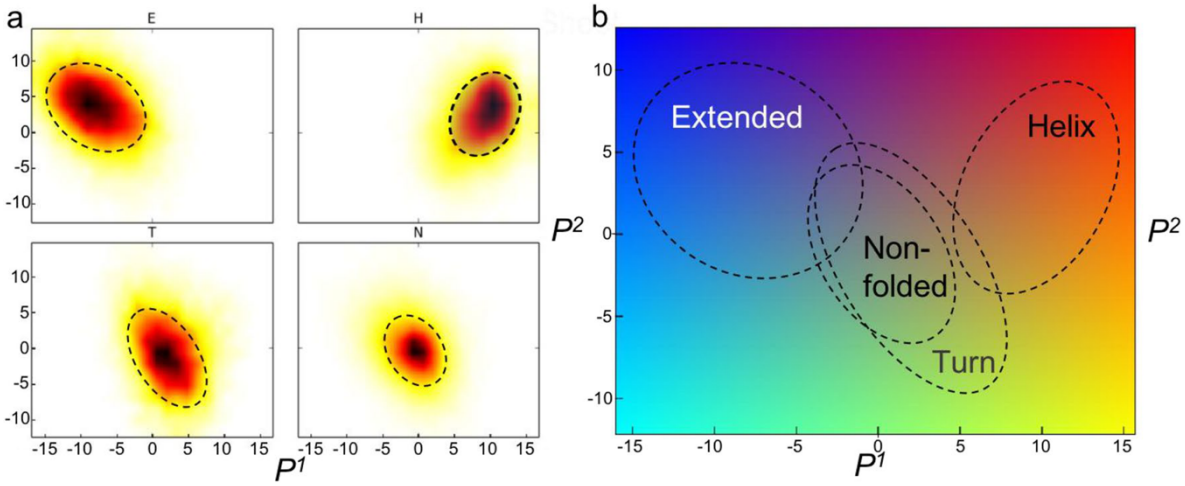


Figure 14 Basis of the CheSPI components. ***(a)*** *Experimental distributions of the first two CheSPI components (P^1 and P^2) for the structural classes extended (E), helix (H), turn (T) and non-folded backbone conformations (N); dashed ellipses delimit each class.* ***(b)*** *Overlay of the four class ellipses on the CheSPI color scale, obtained as a linear combination of the same two components. The first component separates helix from extended structure almost completely, whereas turn and non-folded conformations both center on $P^1 \approx 0$ and are resolved from one another only along the second component, which is negative for turns and close to zero for disorder. On this basis helices are rendered in red, strands in blue, turns in green and disordered residues in gray. Reproduced from Nielsen and Mulder (2021),* Journal of Biomolecular NMR ***75****:273–291, with permission from Springer Nature.*

Across the full eight-class DSSP classification the first component orders the classes as E < B < C < S < T < G < I < H, the more extended conformations taking the more negative values (Nielsen and Mulder, 2021).

The CheZOD Z-score

Alongside the components, CheSPI reports the CheZOD Z-score, a per-residue measure of local order derived from the statistical analysis of sums of squared secondary chemical shifts

(Nielsen and Mulder, 2016, 2020). The score increases with the magnitude of the deviation from random coil, so that low values indicate residues whose shifts are close to random-coil expectations. Nielsen and Mulder applied a cut-off of 5.0 in labeling the training set used to derive the CheSPI components, and a cut-off of 8.0 in classifying disordered regions in the proteins they analyzed. The value of the score lies less in any particular threshold than in its continuity: it returns a quantitative estimate rather than a binary classification, so that a residue may be described as substantially, partially, or negligibly ordered, and adjacent residues may be compared with one another on a common scale.

Because the score is computed from deviations between observed shifts and a predicted random-coil reference, it is sensitive to systematic error in chemical shift referencing, a constant offset in one nucleus being indistinguishable from a genuine, uniform structural deviation. CheSPI limits the influence of such artifacts by truncating the secondary chemical shifts with nucleus-specific cut-offs before they enter the calculation, which prevents extreme values, whether from referencing error or from misassignment, from dominating the result (Nielsen and Mulder, 2021).

CheSPI populations and eight-class secondary structure prediction

The discriminating power of the components is exploited to infer, by statistical inference, the populations of four structural categories for each residue: extended structure, corresponding to the DSSP classes of sheet and bridge (E and B); helical structure, comprising all three helix types (G, H and I); turn (T); and non-folded structure, comprising bend and coil (S and C). These CheSPI populations were validated by comparison with simulated conformational ensembles for four distinct intrinsically disordered proteins. They constitute the quantity that a nucleus-by-nucleus $\Delta\delta$ analysis cannot supply, in that they express the local conformational preference as a fraction of the ensemble rather than as a yes-or-no assignment.

Most methods for inferring secondary structure from chemical shifts return a three-class prediction, and the principal exception, CSI 3.0, returns four classes. CheSPI extends the inference to the full set of eight structural classes defined by DSSP (Kabsch and Sander

1983): α-helix (H), 3_{10}-helix (G), π-helix (I), extended β-strand (E), isolated β-bridge (B), turn (T), bend (S) and coil (C).

This is of direct relevance to a peptidomimetic in which the anticipated structural elements are short and local, since the distinction between a 3_{10}-helical turn, an isolated bridge and an unstructured bend is precisely the distinction that a three-class scheme cannot express.

Finally, the first two components are converted into a color, by means of a linear combination that places well-formed helices in red, well-formed strands in blue, turns in green and disordered residues in gray, with intermediate hues reflecting intermediate conformations. The color scale may be applied to bar plots along the sequence and mapped directly onto a three-dimensional structure, providing an immediate visual summary of local order and its variation.

CheSPI is freely available, either through the web interface at www.protein-nmr.org or as a Python program obtainable from https://github.com/protein-nmr and run locally from the command line. In this thesis the local implementation was used, since it permits the treatment of individual sequence fragments and gives access to the numerical output files. Chemical shifts are supplied in NMR-STAR format together with the sequence and the experimental conditions of temperature, pH and ionic strength, from which the POTENCI reference is generated internally.

The output comprises a set of plain text files reporting, for each residue, the CheZOD Z-score and the first CheSPI components, the four structural populations, the probabilities and the most probable assignment for the three-class and eight-class schemes, and the CheSPI colors; a script for coloring a three-dimensional structure; and a composite figure in which the CheSPI colors and Z-scores, the population distributions and the eight-class prediction are plotted along the sequence (Figure 15 a-c). The two lower panels of that figure derive from the corresponding structural ensemble rather than from the chemical shifts and serve to validate the prediction.

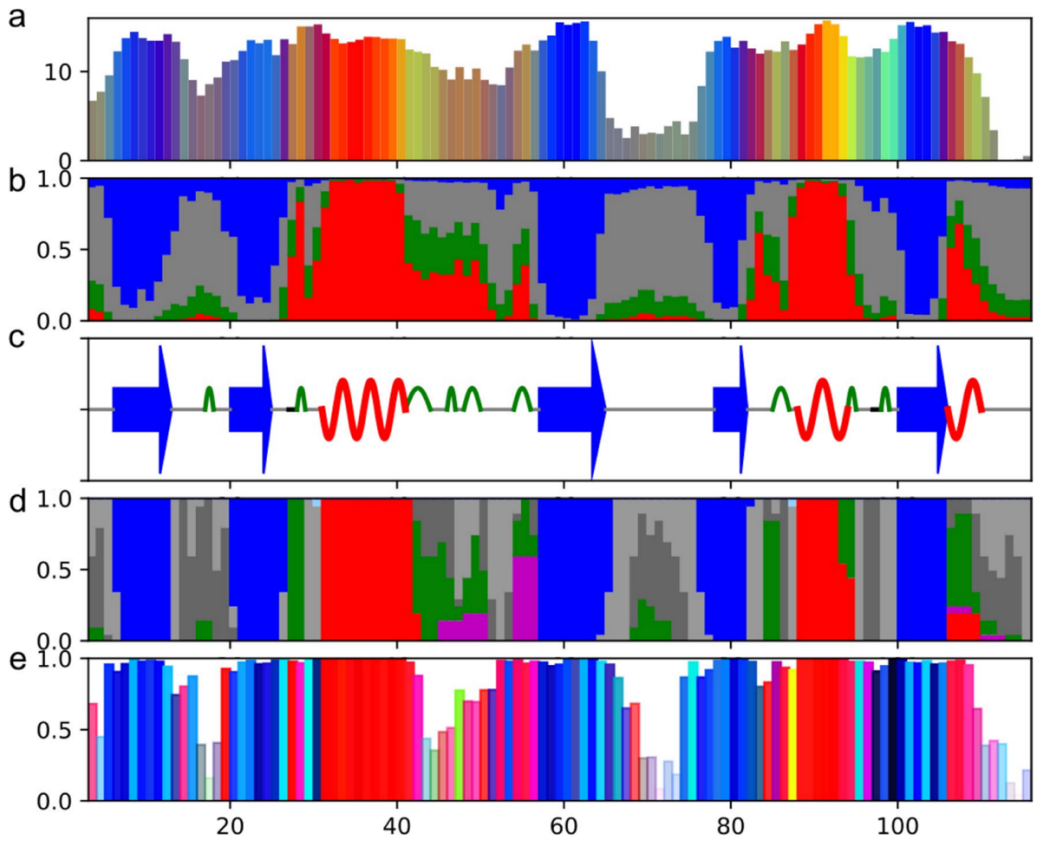


Figure 15 Representative multi-panel output produced by CheSPI, shown for the RA2 domain of phospholipase Cε using the chemical shifts of BMRB entry 6635, together with the corresponding experimental structural ensemble (PDB 2BYF) for validation. ***(a)*** *Bar plot in which the height of each bar is the CheZOD Z-score and its color encodes the first two CheSPI components.* ***(b)*** *Stacked bar plot of the CheSPI populations of extended (blue), helical (red), turn (green) and non-folded (grey) local structure.* ***(c)*** *Cartoon of the most confident eight-class DSSP prediction, using red coils for α-helix (H), magenta for 3_{10}-helix (G), blue arrows for strands and bridges (E and B), green arcs for turns (T), and grey and black lines for coil (C) and bend (S).* ***(d)*** *Stacked bar plot of the conformations observed in the 2BYF ensemble, classified by DSSP into the same eight classes and drawn with the same colors; comparison of (c) with (d) provides a direct assessment of the accuracy of the prediction.* ***(e)*** *Bar plot of backbone angular order, in which the bar height is the geometric mean of the squared angular order parameters for the φ and ψ torsion angles, values close to unity indicating a well-defined local conformation and values close to zero indicating large angular variation; the colors are taken from a two-dimensional scale reflecting the position of each residue in the Ramachandran plot. Reproduced from Nielsen and Mulder (2021)*

Applicability to short sequences

Two constraints follow from the construction of the method and set a lower limit on the length of sequence to which it may usefully be applied.

The first concerns the inference itself. The structural assignment made for a given residue does not rest on its own chemical shifts alone but draws on those of its neighbors within a window of approximately four residues on either side. Residues lying close to a chain terminus are therefore supported by correspondingly less information than residues in the interior of a chain, and the terminal residues themselves are excluded from the analysis

altogether. For a fragment of n residues, results are consequently returned for n–2 positions, and for short fragments a substantial proportion of those positions lies within the region of reduced support.

The second concerns chemical shift referencing. The CheZOD Z-score and the CheSPI components are computed from deviations between observed shifts and a POTENCI random-coil reference, so a systematic referencing error in one nucleus is indistinguishable from a genuine, uniform structural deviation. In a sequence of sufficient length such an offset can be recognized from the behavior of the deviations along the chain and corrected. In a short fragment this redundancy is absent, no correction can be evaluated, and the referencing of the input shifts must therefore be established independently before the output is interpreted.

The consequences differ between the several quantities that the method reports. The CheZOD Z-scores and the four-class populations are computed residue by residue and remain informative for short fragments, subject to the referencing caveat above.

The eight-class secondary structure assignment is more sensitive, since it is obtained by a stochastic optimization over the fragment as a whole, in which the grammar of permitted secondary structure elements acts as a constraint; with few residues available that optimization is correspondingly less well determined, and repeated runs may return different assignments of comparable likelihood.

For sequences of the length of a short peptide, the populations therefore constitute the more robust output, and the eight-class assignment is best regarded as indicative rather than definitive.

AIMS of The THESIS WORK

1) Analyze solution NMR spectra of LB6 molecule, formed by two eptapeptides connected by a synthetic β-inducer, to assign all the 1H and Cα peptidic resonances in both reducing and oxidizing conditions
2) Assess the formation of intramolecular Cys188-Cys201 disulfide bridge from dipolar correlation spectra analysis
3) Calculate Hα and Cα secondary chemical shifts against independent reference random-coil chemical shifts experimentally or statistically derived to infer the existence of secondary structural elements
4) Estimate secondary structure population by Chemical shift Secondary structure Population Inference program

Materials and Methods

4.1 Sample Preparation

LB-6 was synthesized by Prof. Sara Pellegrino at University of Milan, by microwave-assisted Fmoc/tBu-based solid phase peptide synthesis, piperidine–pyrrolidine scaffold was prepared and coupled to peptidic fragments in accordance with previously published procedures reported in Bollati 2022.

LB-6 was dissolved in 200 mM sodium phosphate buffer at pH 5.0 to a final concentration of 1 mM. This buffer composition was selected to match the conditions used in prior G2-gelsolin aggregation assays, ensuring experimental consistency across the biophysical study. To each sample, 25 µL of D2O was added for deuterium lock and ~2 µL of dioxane as an internal ^{1}H NMR reference standard (δ 3.76 ppm) to a final total volume of 600 µL. To prevent oxidative aggregation and maintain peptide in the monomeric state, all NMR experiments were conducted in the presence of 5 mM DTT (dithiothreitol) [full rationale explained in Section 4.1.1].

4.2 NMR Data Acquisition

NMR experiments were performed on a Bruker NEO equipped with a Prodigy cryoprobe operating at 600 MHz. All spectra were recorded at 288 K (15°C) on the reducing LB-6 sample described in Section 3.1. Two-dimensional homonuclear ^{1}H-^{1}H TOCSY, ROESY, and NOESY spectra, together with a natural-abundance ^{1}H-^{13}C HSQC, were collected to enable sequential resonance assignment and secondary chemical shift analysis. Acquisition parameters for each experiment are summarized in Table 3.1.

Table 3.1. NMR acquisition parameters for LB-6.

Experiments	TOCSY	ROESY	NOESY	^{13}C-HSQC
NS	48	96	64	128
NE	400	400	400	256
TD	2048	2048	2048	1024
$SW_{F2}\,(ppm)$	14	14	14	14
$SW_{F1}(ppm)$	14	14	14	200
$R_d^{F1}\,(Hz/pt)$	8.004611	8.004611	8.004611	16.009222
T(K)	288	288	288	288

$t_{sl}\ (ms)$	80			
$t_{mix}(ms)$		250	100	

NS: number of scans; NE: number of experiments (increments); TD: number of time-domain data points; SW: spectral width; R_d: digital resolution; t_sl: spin-lock time (TOCSY); t_mix: mixing time (ROESY, NOESY).

Spectra were processed using Bruker Topspin (version 4.3.0). Sequential resonance assignment was performed using NMRFAM-SPARKY (version 1.470, powered by Sparky 3.190).

Results and Discussion

5.1 Preliminary stability assessment and optimization of redox conditions

The peptide is formed by two arms, each including 7 amino-acids. The peptide arms are derived from the amyloidogenic core (NGDCFIL, residues 185–191) and from the flanking sequence (HQWCGSN, residues 198–204) are joined through a non-natural piperidine–pyrrolidine amino acid (X) that was design to act as a β-turn inducer, to reproduce the reversal of chain direction observed in the folded domain Figure 6. The piperidine nitrogen is acylated by the first arm, and the pyrrolidine carbonyl carries the second, while a tosyl group (Tos) is borne on the piperidine ring. The N-terminus is free, and the C-terminus is amidated. The schematic representation of the peptide sequence is reported in Figure 16, together with the chemical structure of the non-natural piperidine–pyrrolidine β-turn inducer:

H_2N–N_{185}–G_{186}–D_{187}–C_{188}–F_{189}–I_{190}–L_{191}–X–H_{198}–Q_{199}–W_{200}–C_{201}–G_{202}–S_{203}–N_{204}–$CONH_2$

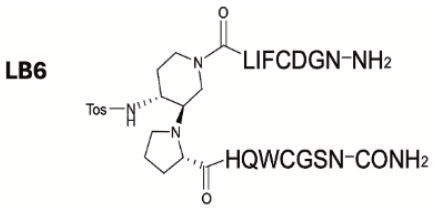


Figure 16 Chemical structure of the peptidomimetic LB6. The two peptide arms are joined by a non-natural piperidine–pyrrolidine amino acid (X), designed to act as a β-turn inducer and to reproduce the reversal of chain direction observed in the folded G2 domain. The first arm, corresponding to residues N185–L191 of the gelsolin amyloidogenic core, is attached through its C-terminal carbonyl to the piperidine nitrogen and is therefore drawn in the C→N direction (LIFCDGN), terminating in the free α-amino group. The second arm, corresponding to residues H198–N204 of the flanking sequence, is carried by the pyrrolidine carbonyl and is drawn in the conventional N→C direction (HQWCGSN), terminating in an amidated C-terminus. A tosyl group (Tos) is borne on the piperidine ring. Adapted from Bollati et al. (2026).

NMR spectra recorded on LB-6 revealed a very complex pattern of resonances. A higher number of spin systems were observed with respect to the 13 expected, considering the two arms, which together comprise 14 amino acids, and the expected absence of the amide resonance of the first residue (N185) owing to fast exchange with water.

The high number of signals strongly complicated sequence-specific resonance assignment. The source of the multiple signals could be the presence of two Cysteines. Based on the designed peptidomimetic architecture, LB-6 incorporates two cysteines (Cys188 and Cys201) that could form intramolecular and /or intermolecular disulfide bonds.

To investigate in detail this possibility and identify optimal experimental conditions, LB6 was analyzed in two different conditions:

1. **<u>Oxidizing condition (no reducing agent)</u>**

One sample was prepared without a reducing agent to permit spontaneous cysteine oxidation and disulfide bond formation.

If formed, such a Cys188–Cys201 disulfide bridge would enforce a rigid β-hairpin geometry, predicted to exhibit characteristic NMR signatures: downfield-shifted ^{1}Hα resonances (diagnostic of β-sheet structure) and short- to medium-range NOEs (diagnostic of defined secondary structure). However, the observed NMR spectra were inconsistent with this

expectation. The $^{1}H\alpha$ resonances remained at chemical shifts characteristic of disordered peptide ($\Delta\delta \approx 0$), and the expected β-sheet-characteristic NOEs were absent. After 24 h in solution, the peptide resonance intensity progressively decreased over days, accompanied by visible precipitation. This behavior is inconsistent with intramolecular disulfide bond formation; rather, it indicates intermolecular disulfide bond formation between separate LB-6 molecules, leading to peptide oligomerization and polymerization that progressively removes monomeric species from solution.

2. **Reducing condition (with DTT)**

To prevent unwanted intermolecular disulfide cross-linking while allowing investigation of the intrinsic structure of monomeric LB-6, a second sample was prepared with 5 mM DTT added. DTT is a reducing agent that maintains both Cys188 and Cys201 in the fully reduced state (−SH form) and effectively prevents both inter- and intramolecular disulfide bond formation. Under these reducing conditions, LB-6 remained monomeric and did not aggregate. NMR spectra were stable over the entire acquisition period (24–48 hours per temperature), yielded reproducible chemical shifts across multiple acquisitions, and were of sufficient signal intensity to permit complete spectral analysis. All subsequent NMR experiments described in this chapter were therefore conducted under reducing conditions (5 mM DTT). Interestingly the spectra recorded in oxidizing conditions within the first 24 hours from LB6 dissolution were perfectly superimposable to those recorded in the presence of DTT, allowing to fully investigate the peptide on a sample stable over time.

Interpretation. The comparative analysis of oxidizing versus reducing conditions reveals that LB-6 does not spontaneously form an intramolecular disulfide bond under the experimental conditions (pH 5.0, 288 K). This is a significant structural finding: the β-hairpin geometry stabilized by the non-natural piperidine–proline β-turn inducer is sufficient to organize the peptide structure without disulfide bond reinforcement. The rapid aggregation observed under oxidizing conditions further demonstrates that the two cysteine residues are readily available for intermolecular cross-linking, confirming their surface accessibility in the monomeric structure. This finding is consistent with the design goal of maintaining LB-6 as

a flexible, accessible inhibitor capable of adapting structure upon interaction with gelsolin targets.

5.2 Sequential assignment of LB6 peptide

Sequence-specific resonance assignments for monomeric LB-6 were performed using a systematic three-stage SPARKY-based workflow combining information from TOCSY, ROESY, NOESY, and HSQC spectra. The strategy exploited the characteristic spectral connectivity patterns of each amino acid type to build a residue-by-residue map of the peptide sequence.

Stage 1: Intra-residue spin system identification from TOCSY:

TOCSY spectra acquired at 60 ms and 80 ms spinlock mixing times establish scalar coupling relationships between all protons within a single residue. For each residue position in the known LB-6 sequence, the expected TOCSY cross-peaks form a distinctive "spin system" a characteristic pattern of peaks whose multiplicities and spacing directly reflect the residue's side-chain structure. Each amino acid type (Asn, Gly, Asp, Cys, Phe, Ile, Leu, His, Gln, Trp, Ser) produces a unique spin system topology. The assignment process began by systematically matching observed TOCSY multiple patterns against the expected topologies of the 14 residues of LB-6, of which 13 give an observable amide resonance. Peaks arising from the backbone amide proton (HN) and the α proton (Hα) of each residue were identified first (as these define the entry point into the spin system), then expanded to include the complete set of side-chain protons (Hβ, Hγ, aromatic protons, etc.) The assignment was initiated from the aliphatic fingerprint region, where Ile190 and Leu191 give rise to the most distinctive spin systems of the sequence. Both residues carry branched side chains with two methyl groups resonating between 0.5 and 0.9 ppm.

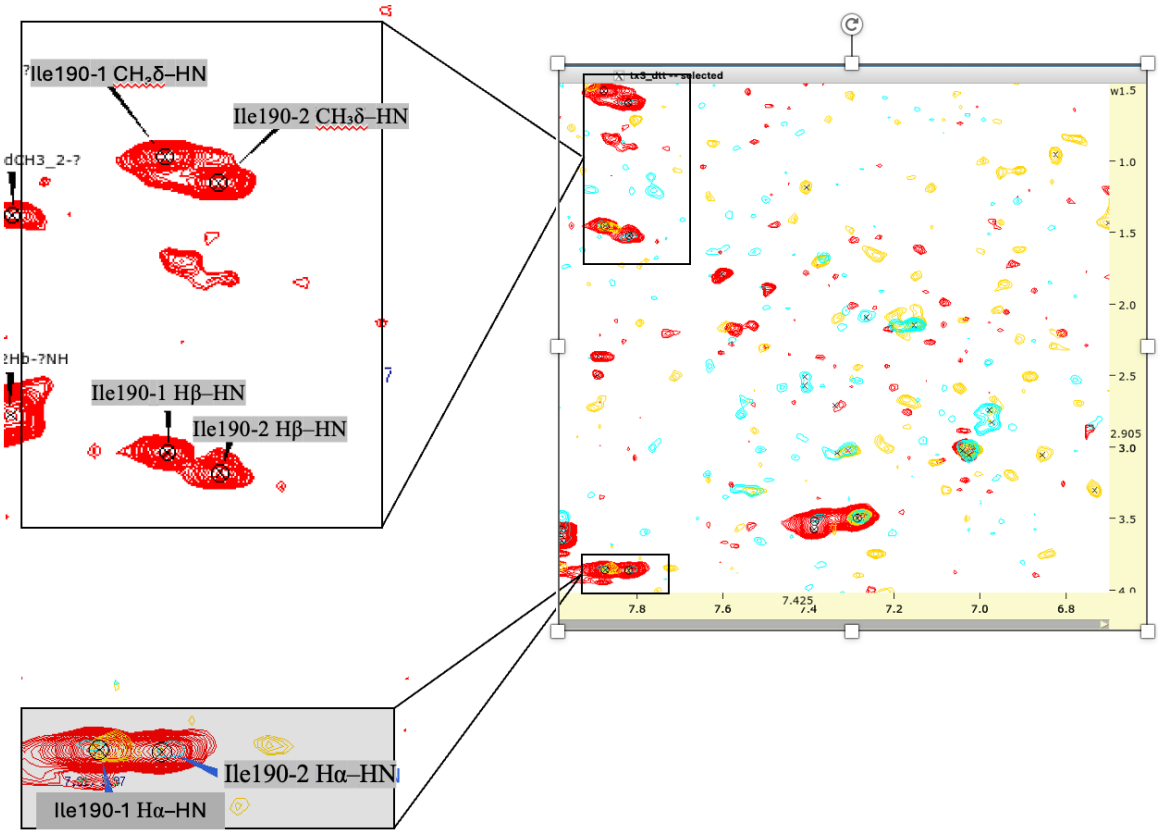


Figure 17 Ile190 spin system in the ^{1}H–^{1}H TOCSY spectrum of LB6 (600 MHz, 288 K, pH 5.0). The boxed regions of the overview (right) are expanded on the left: the amide-to-side-chain region (top) and the Hα–HN region (bottom). All resonances are doubled, corresponding to the conformational states LB6-1 (HN 7.874, Hα 3.859, Hβ 1.456 ppm) and LB6-2 (HN 7.818, Hα 3.861, Hβ 1.522 ppm), which arise from slow interconversion of the linker.

Side-chain identification in the upfield aliphatic region. While the amide-to-aliphatic region establishes that a spin system belongs to a given residue, it does not reveal the shape of the side chain, since all cross-peaks are referred to a single amide proton. Residue-type assignment was therefore carried out in the upfield aliphatic region of the TOCSY spectrum (0–2 ppm in both dimensions), where the side-chain protons correlate with one another and the branching pattern of the spin system becomes apparent Figure 17.

Identification was based on comparison with reference spin-system diagrams, which display the characteristic COSY and TOCSY cross-peak patterns of the twenty standard amino acids together with their random-coil chemical shifts. Isoleucine is distinguished by a topology not reproduced by any other residue present in LB6: two methyl groups carried on separate

branches of the side chain, γCH₃ bonded directly to Cβ, and δCH₃ at the end of the ethyl arm, both resonating below 1.0 ppm, in the least crowded part of the spectrum. Leucine, the only other residue in the sequence with two methyl groups, carries both on the same carbon and therefore gives a different connectivity pattern, which allowed Ile190 and Leu191 to be distinguished despite the partial overlap of their resonances.

The complete spin system of Ile190 was traced in this region, from δCH₃ and γCH₃ through γCH₂ and Hβ to Hα and the amide proton. Every one of these resonances was observed in duplicate Figure 18.

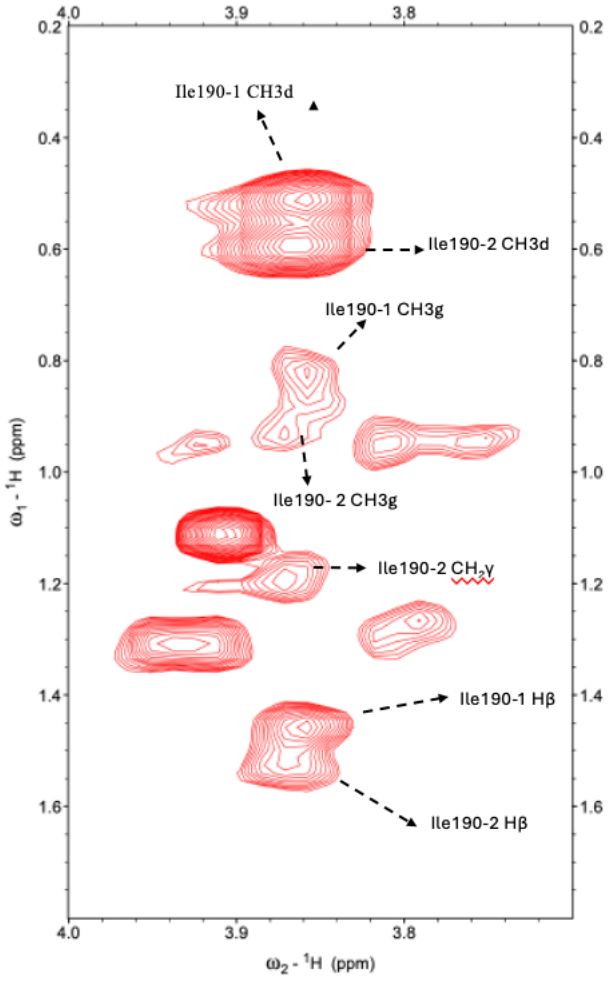


Figure 18 Identification of the Ile190 side chain in the ¹H–¹H TOCSY spectrum of LB6 (600 MHz, 288 K, 200 mM sodium phosphate pH 5.0, 5 mM DTT). Expansion taken at the Hα frequency (ω₂ 4.0–3.7 ppm, ω₁ 0.2–1.8 ppm), in which the correlations from Hα to Hβ, γCH₂, γCH₃ and δCH₃ place the entire side chain of Ile190 in a single vertical column. The branched topology characteristic of isoleucine, two methyl groups carried on separate arms of the side chain, γCH₃ on Cβ and δCH₃ at the end of the ethyl branch, resonating at clearly distinct chemical shifts, is directly apparent, and distinguishes

Ile190 from Leu191, whose two methyl groups are borne on the same carbon and are correspondingly close in shift. Every resonance of the spin system is observed in duplicate, corresponding to the two conformational states LB6-1 and LB6-2: δCH₃ 0.512/0.596, γCH₃ 0.819/0.898, γCH₂ 1.152/1.196, Hβ 1.456/1.522 and Hα 3.859/3.861 ppm for LB6-1 and LB6-2 respectively. The duplication is therefore not confined to the backbone amide resonances but extends through the whole side chain as far as the terminal methyl groups. For LB6-2 the γCH₃ and one of the two γCH₂ protons resonate at 0.898 and 0.904 ppm and are not resolved from one another. Cross-peaks lying outside the Ile190 column belong to the spin systems of other residues whose Hα resonances fall within this spectral region.

Stage 2: Sequential connectivity from NOE cross-peaks:

To order the individually assigned spin systems along the known 14-residue sequence, inter-residue NOE connectivities were searched in ROESY and NOESY spectra. The identification of sequential connectivities mainly relied on two characteristic NOE patterns: (i) dαN connectivities, in which the Hα proton of residue i exhibits an NOE cross-peak to the backbone amide HN of residue i+1, indicative of their spatial proximity across a peptide bond; and (ii) dβN connectivities, linking Hβ of residue i to HN of i+1. By systematically identifying these inter-residue NOE peaks in the spectra and tracing them sequentially from the known N-terminus through the C-terminus, the complete spine of backbone connectivities was established. This sequential chain was verified against the two arms of the peptide, and the connection was interrupted by the presence of the linker region.

ROESY spectra were particularly valuable for this step due to their superior sensitivity for peptides of LB-6's molecular weight (compared to standard NOESY), allowing reliable identification of weak inter-residue NOE peaks. Each matched spin system was individually labelled in SPARKY using a four-character code (e.g., "Phe189" for phenylalanine at position 189).

The first entry point was provided by Phe189, which was located through the sequential NOE connecting its Hα to the amide proton of Ile190, a residue already assigned unambiguously from its branched side-chain spin system Figure 19. Phe189 itself gives a single set of resonances, and the two sequential cross-peaks observed from its Hα, at the amide frequencies of Ile190-1 (7.874 ppm) and Ile190-2 (7.818 ppm), place the N-terminal limit of the doubled region at the amide of Ile190.

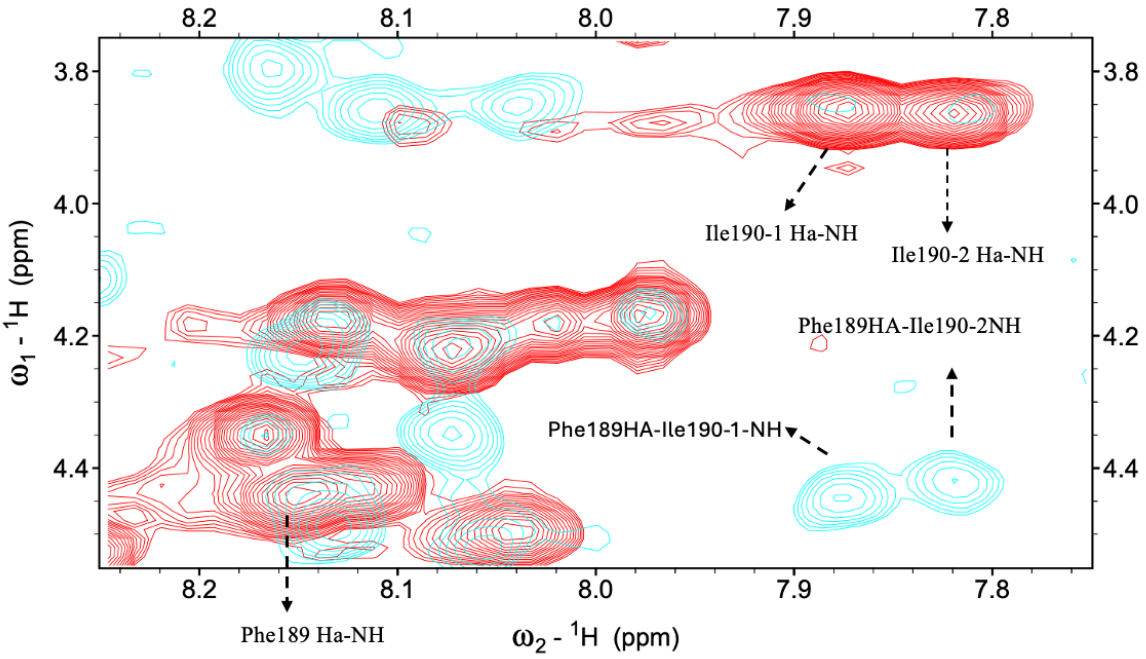


Figure 19 Sequential connectivity from Phe189 to Ile190 in the fingerprint region of LB6 (600 MHz, 288 K, 200 mM sodium phosphate pH 5.0, 5 mM DTT). TOCSY in red, ROESY in cyan. The intra-residue Hα–HN cross-peaks of Phe189 and of the two Ile190 conformers appear in the TOCSY, whereas the two sequential d_αN cross-peaks from the Hα of Phe189 to the amide protons of Ile190 are observed only in the ROESY, confirming that they arise from through-space contacts between adjacent residues rather than from scalar coupling within a single spin system. Phe189 gives a single set of resonances, but its Hα shows two sequential cross-peaks, at the amide frequencies of Ile190-1 (7.874 ppm) and Ile190-2 (7.818 ppm) and at the same ω1. This connectivity identified Phe189 from the previously assigned Ile190 spin system and locates the N-terminal boundary of the doubled region between Phe189 and Ile190.

Aromatic side chains were attributed by a strategy different from that used for the aliphatic residues and were approached from three independent points of entry lying outside the crowded aromatic region.

The difficulty arises from the nature of the coupling network. Magnetization transfer in TOCSY proceeds through ^{3}J couplings, and in Phe, His and Trp this network is interrupted at Cγ, which bears no proton. The transfer therefore stops before reaching the ring, and each aromatic side chain constitutes a spin system isolated from its own backbone: the ring protons can be observed and their residue type recognized, but TOCSY alone cannot establish to which residue of the sequence they belong. The connection had to be made through NOE contacts instead. In LB6 this general difficulty was compounded by the composition of the

aromatic region itself, which is unusually crowded for a peptide of this length. Besides the phenyl ring of Phe189 and the benzene ring of the indole of Trp200, it contains the tosyl ring of the β-turn inducer, which belongs to no backbone spin system at all, and the side-chain amide protons of Asn204. The latter are not aromatic, but the primary amide group in which they reside possesses partial double-bond character through conjugation with the carbonyl, which both deshields the two protons and renders them inequivalent, placing them near 7.6 and 6.9 ppm Figure 20.

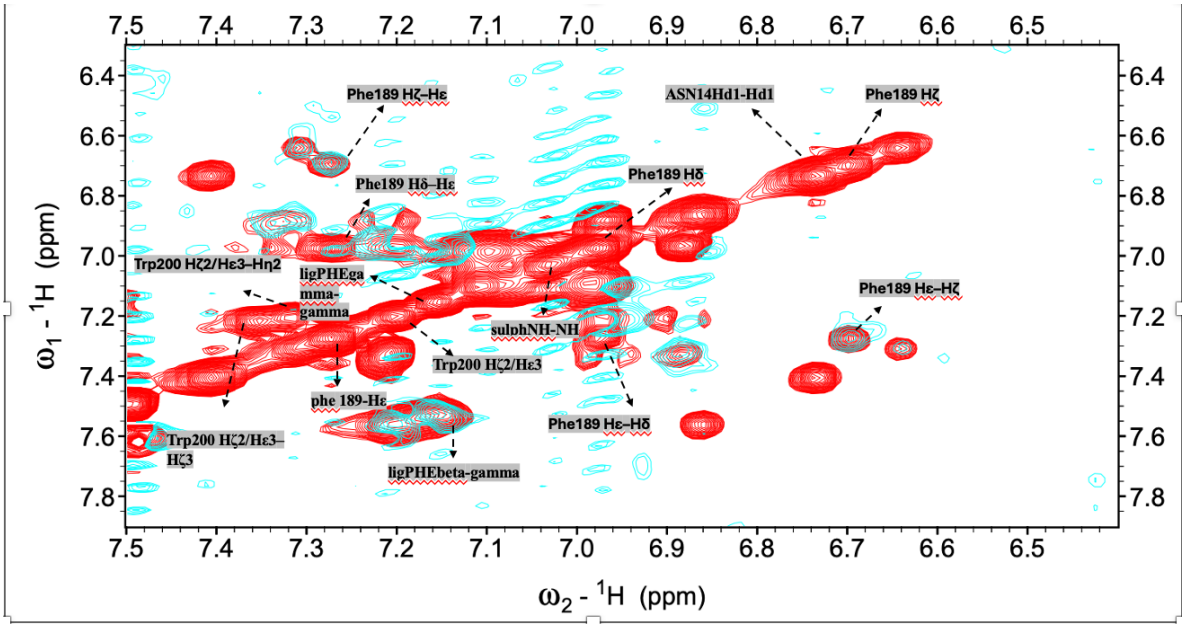


Figure 20 Aromatic region of LB6 (600 MHz, 288 K, 200 mM sodium phosphate pH 5.0, 5 mM DTT). TOCSY in red, ROESY in cyan. The region contains the phenyl ring of Phe189, with correlations among its ortho (Hδ), meta (Hε) and para (Hζ) protons; the benzene ring of the Trp200 indole; the tosyl ring and sulfonamide NH of the β-turn inducer; and the side-chain amide protons of Asn204. Rapid rotation of the phenyl ring about the Cβ–Cγ bond exchanges the two ortho protons with one another and the two meta protons with one another, so that the ring gives three signals rather than five, in the intensity ratio 2:2:1.

The second entry point was the indole NH of Trp200. This proton is not an ordinary amide: its nitrogen lone pair forms part of the aromatic system of the indole, so that the N–H bond is strongly polarized and the proton correspondingly deshielded. It was observed at 9.979 ppm, more than one ppm downfield of the most deshielded backbone amide of the peptide (His198, 8.618 ppm), in a part of the spectrum that is otherwise entirely empty. Since no

other proton of LB6 resonates in this region and the sequence contains a single tryptophan, the assignment was immediate, and this signal served as the anchor from which the remaining protons of the indole system, Hδ1, Hζ2, Hζ3 and Hη2, were traced.

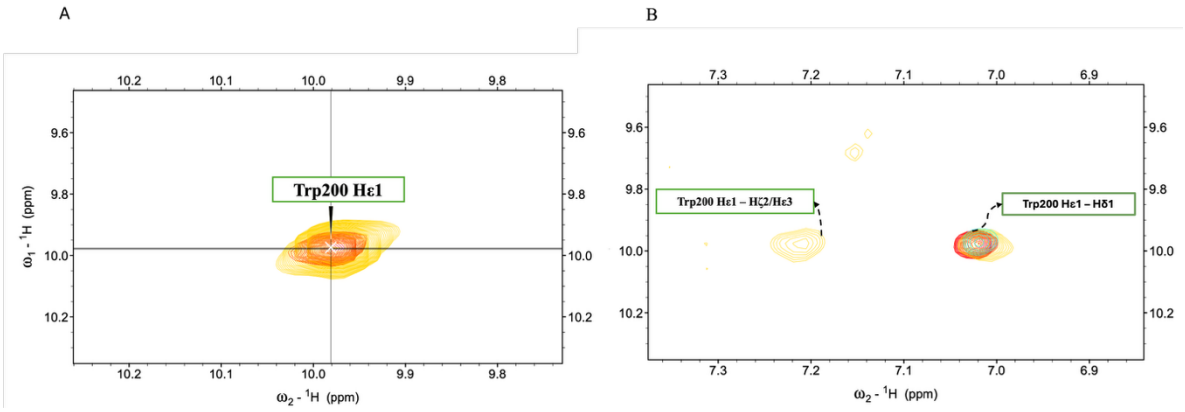


Figure 21 Entry into the Trp200 side chain from the indole NH (600 MHz, 288 K, 200 mM sodium phosphate pH 5.0, 5 mM DTT). TOCSY in red, NOESY in yellow. (A) The Hε1 diagonal peak at 9.979 ppm, isolated from all other resonances of LB6, which made the indole NH an unambiguous entry point into the spin system. (B) Strip taken at ω_1 = 9.979 ppm, showing the NOE cross-peaks from Hε1 to Hδ1 (7.02 ppm) and to Hζ2/Hε3 (7.21 ppm), the two ring protons closest in space to the indole NH. These contacts provided the starting point for the assignment of the indole spin system.

The third entry point was His198. The two protons of the imidazole ring do not resonate together with the other aromatic signals: Hε1 was observed at 8.351 ppm, within the amide region, and Hδ2 at 7.039 ppm. Their mutual cross-peak consequently falls at (8.35, 7.04), in a sparsely populated part of the spectrum lying between the amide–amide region and the aromatic diagonal and could be identified without recourse to the sequential procedure. The position of Hε1 is consistent with an imidazole ring that is largely protonated at the pH of the sample, the pKa of the histidine side chain being approximately 6.9 against a sample pH of 5.0; protonation displaces Hε1 downfield and thereby increases the separation of this cross-peak from the aromatic resonances Figure 22.

Working from these three anchors, the internal connectivity of the aromatic region was traced and the ring systems recognized by comparison with reference spin-system diagrams, which show the expected cross-peak pattern of each aromatic residue. In the case of Phe189 the two ortho protons give a single resonance, as do the two meta protons, because rapid rotation of the ring about the Cβ–Cγ bond exchanges them with one another on a timescale short compared with their chemical shift difference. The phenyl ring therefore produces three

signals rather than five, in the intensity ratio 2:2:1, and this reduced pattern is itself diagnostic of the residue. Each ring was finally connected to its own backbone through the Hβ–Hδ NOE. This contact is the standard means of linking an aromatic side chain to its own spin system, since the Hδ protons lie only two bonds from Cγ and therefore remain close in space to the Hβ protons in any side-chain conformation, so that the cross-peak is reliably observed.

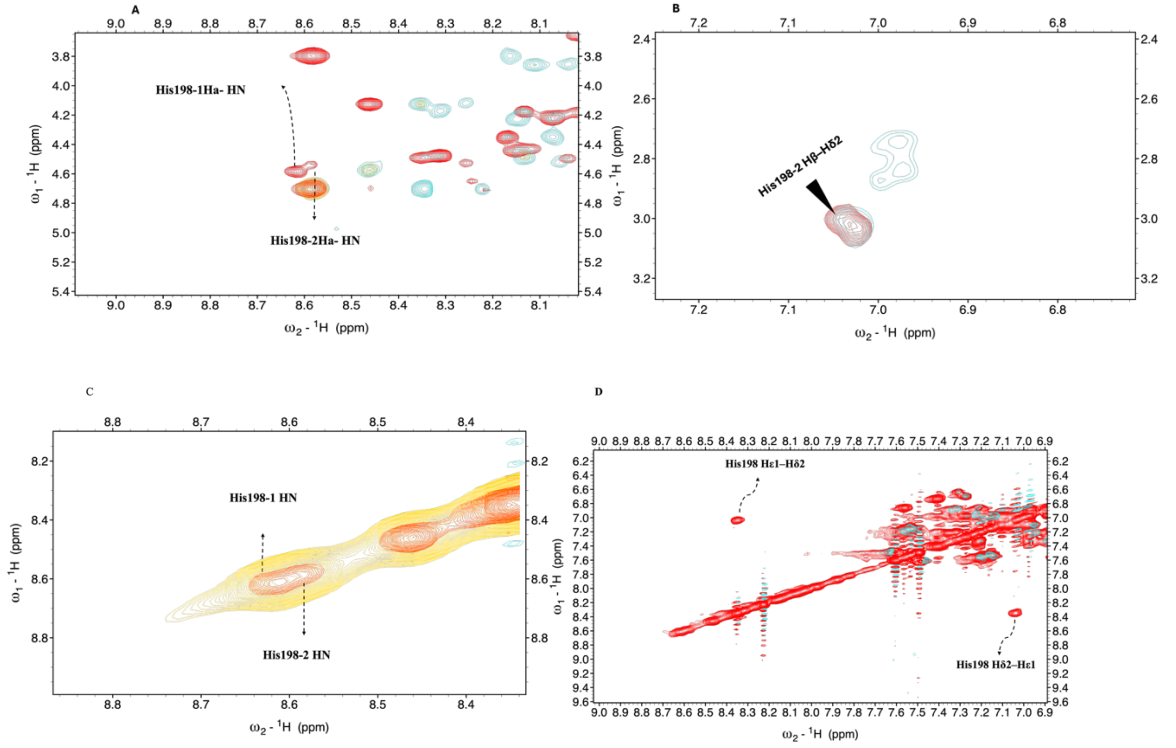


Figure 22 Identification of His198 and of its conformational doubling (600 MHz, 288 K, 200 mM sodium phosphate pH 5.0, 5 mM DTT). TOCSY in red, ROESY in cyan, NOESY in yellow. (A) Fingerprint region. His198 gives two distinct Hα–HN cross-peaks, at (4.584, 8.618) and (4.538, 8.585) ppm, showing that the conformational doubling extends to this residue. (B) Region correlating sidechain Hβ protons with aromatic resonances. The NOE between Hβ at 3.027 ppm and Hδ2 at 7.041 ppm links the imidazole ring to its own backbone: since no scalar coupling pathway crosses the Cβ–Cγ bond, this through-space contact is what establishes that the ring belongs to His198. (C) Amide–amide region. The two backbone amide resonances of His198 lie only 0.027 ppm apart, at 8.612 and 8.585 ppm, and are therefore only partially resolved on the diagonal. (D) Overview of the amide and aromatic regions. The pair of symmetry-related cross-peaks at (8.35, 7.04) and (7.04, 8.35) ppm arises from the imidazole ring, connecting Hε1 at 8.351 ppm, which resonates within the amide region because the ring is largely protonated at pH 5.0 to Hδ2 at 7.039 ppm. Their position, isolated from both the amide–amide region and the aromatic diagonal, allowed His198 to be identified directly, without recourse to the sequential assignment procedure. The vertical ridges at $\omega_2 \approx 8.2$ and 7.5 ppm are t_1 noise.

A critical feature of the spectral analysis was the selective peak doubling of the piperidine–proline β-turn inducer and its immediate flanking residues: Ile190–Leu191–[TosNH-pip-pro]–His198–Trp200–Cys201 -Ser203). These residues, exhibited two independent sets of peaks arising from two conformational states, designated as LB6-1 and LB6-2. These

duplicated signals indicate the presence of two conformational states in slow exchange on the millisecond timescale, most likely corresponding to alternate orientations of the piperidine linker. The peptide resonances of both conformers were assigned de novo. Peak assignment was performed simultaneously for both sets: after identifying and labelling a TOCSY spin system for LB6-1 (residue Cys201, conformation 1), the corresponding TOCSY peaks for the same residue in conformation 2 (LB6-2) were immediately located nearby in the spectrum, marked with an identical residue label but with a conformer suffix ("Cys201_1" and "Cys201_2"). The two conformational states are not equally populated as observed from the difference in signal intensity. An example of the observed resonances for selected residues is shown in the Figure 23:

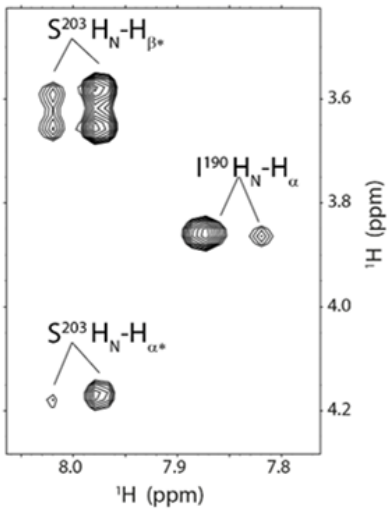


Figure 23 Selected region of the ¹H–¹H TOCSY spectrum of LB6, showing the two distinct patterns of resonances observed for Ile190 and Ser203. Each amide proton gives a separate set of HN–Hα and HN–Hβ cross-peaks for the two conformational states; signals belonging to the second state are marked with an asterisk. Reproduced from Bollati et al. (2026)

Other residues (Asn185, Gly186, Asp187, Cys188 and Phe189 in the first arm, and Gln199, Gly202 and Asn204 in the second) appeared as single resonances in all spectra, reflecting their distance from the piperidine–pyrrolidine linker. A graphical summary of the residues showing double forms I offered in the Figure 24:

H_2N–N_{185}–G_{186}–D_{187}–C_{188}–F_{189}–I_{190}–L_{191}–X–H_{198}–Q_{199}–W_{200}–C_{201}–G_{202}–S_{203}–N_{204}–$CONH_2$

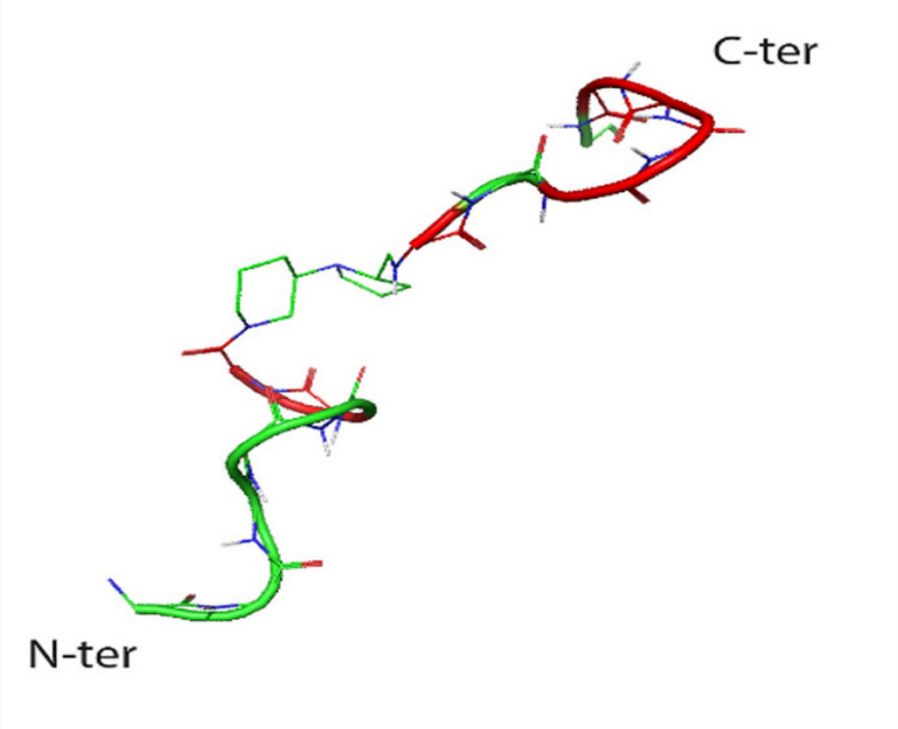


Figure 24 Conformational doubling in LB6. Top: sequence of LB6, The six residues that give two distinct sets of resonances in all spectra, Ile190, Leu191, His198, Trp200, Cys201 and Ser203, are shown in red; the non-natural piperidine–pyrrolidine β-turn inducer is indicated as X. Bottom: representative conformer of LB6 obtained from all-atom molecular dynamics simulation, with the backbone shown as a ribbon and the side chains as sticks, in which residues showing two distinct conformations are again colored in red. The doubled residues are confined to the two segments immediately flanking the linker, consistent with an origin in alternate orientations of the piperidine ring rather than in any intrinsic property of the two arms. The structure is reproduced from Bollati et al. (2026).

The non-natural TosNH-piperidine-proline linker (position 8) remained unassigned due to severe spectral overlap and low signal intensity in the crowded aliphatic region (δ ≈ 3–4 ppm). Despite the linker's unresolved proton resonances, sequential connectivity of the flanking residues (Leu190 and His198) was unambiguous, and the linker's conformational effects are evident in the chemical shifts and NOE patterns of these neighboring residues.

5.3 LB6 Cα resonance assignment

^{1}H–^{13}C correlation spectrum ^{1}H–^{13}C HSQC have been recorded on the natural abundance peptide exploiting the high sensitivity ensured using a cryoprobe. The interest in assigning

Cα resonances lies in their ability to provide further information about the peptide's secondary structure.

All ^{1}H assignments obtained from the homonuclear (TOCSY, ROESY, NOESY) spectra were transferred to the natural-abundance $^{1}H–^{13}C$ HSQC spectrum recorded in the spectral region corresponding to Hα–Cα, correlations. In $^{1}H–^{13}C$ HSQC spectrum each cross-peak directly correlates a ^{1}H nucleus (F2 dimension) to the ^{13}C nucleus it is bonded to (F1 dimension). For each identified pair, the $^{1}H\alpha$ chemical shift was confirmed to match the value from the homonuclear spectra, and the orthogonal ^{13}C coordinate was extracted, providing the ^{13}C chemical shift. For backbone nuclei, this yielded unambiguous Hα–Cα all 12 residues.

The derived resonance assignment is summarized in the Table 4.1 where the direct readout of SPARKY is reported after calibrating spectra with respect to internal dioxane resonance. Based on spectra Hz/pt resolution, chemical shift values are measured with a precision of 0.01 ppm (^{1}H) and 0.1 ppm (^{13}C).

Table 4.1 H and $^{13}C\alpha$ chemical shifts of LB6 in aqueous solution. Values are given in ppm. For residues showing two sets of resonances, the two conformational states are reported separately and denoted by the suffixes 1 and 2 (LB6-1 and LB6-2). The direct readout of SPARKY is reported, however the estimated precision, determined by the digital resolution of the spectra, is 0.01 ppm for ^{1}H and 0.1 ppm for ^{13}C. Residues for which no resonance was observed are indicated as n.o.

	C(Alpha)	**NH**	**Hα**	**Hβ**	**Others**
Asn (N)-185	n.o	n.o	n.o	n.o	n.o
Gly(G)-186	42.571	8.582	3.789		
Asp(D)-187	51.496	8.168	4.350	2.379	
Cys(C)-188	55.562	8.073	4.215	2.598	
Phe(F)-189	54.948	8.147	4.438	b1:2.826 b2: 2.767	Hδ: 6.97, H£: 7.268, H**ζ**:6.695
iLe(I)-190-1	57.746	7.874	3.859	1.456	**δ**: 0. 512, CH3γ: 0.819, CH2$\boldsymbol{\gamma}$: 1.152,
					δ:0.579

iLe(I)-190-2		7.818	3.861	1.522	CH3**δ**: 0. 596, CH3***γ***:0.898, CH2*γ*:1.196, CH2*γ*:0.904,
					CH2*γ*:1.145
Leu(L)-191_1	54.533	8.113	4.427	1.346	**δ**:0.652, H*γ*: 1.063, **δ**:0.703
Leu(L)-191_2		8.043	4.497	1.336	δCH3:0.703, δCH3:0.671 H*γ*:1.075
Linker	_	_	_	_	_
His(H)-198-1	51.330	8.618	4.584	3.009	Hδ: 7.039, H*γ*:8.351
His(H)-198-2	54.465	8.585	4.538	3.005	
Gln(Q)-199	53.332	8.462	4.126	1.759, 1.817	H**γ**: 2.091
Trp(W)-200-1	54.486	8.354	4.492	3.031	H*δ*:7.02, Nε:9.979, H**Z**2 :7.212, H***Z***3: 7.351, Hη:7.317
Trp(W)-200-2		8.257	4.527	3.059	
Cys(C)-201-1	55.471	8.133	4.174	2.533	
Cys(C)-201-2		8.057	4.198	2.584	
Gly(G)-202-1	49.784	7.967	4.169		
Ser(S)-203-1	63.139	8.021	4.182	3.594, 3.592	
Ser(S)-203-2		7.974	4.169		
Asn (N)-204	50.238	8.314	4.477	2.593, 2.587	Hδ:6.739, Hδ:7.406

5.4 NMR chemical shift analysis

Secondary chemical shift analysis was performed on the Hα and Cα resonances of LB6, calculated as Δδ = δobs − δrc, the difference between the experimentally observed chemical shift and the expected random-coil value for each residue type. To verify that any local deviations reflect a genuine property of LB6 rather than an artefact of the reference dataset chosen, this analysis was repeated independently against two experimentally derived random-coil reference sets reported by Kjargaard (ref) and Wuthrich (ref) and a statistically derived reference (Zhang, Neal & Wishart, 2003), that have been extensively described in Chapter 2. In addition, a correction factor related to experiment temperature of 15°C was also evaluated.

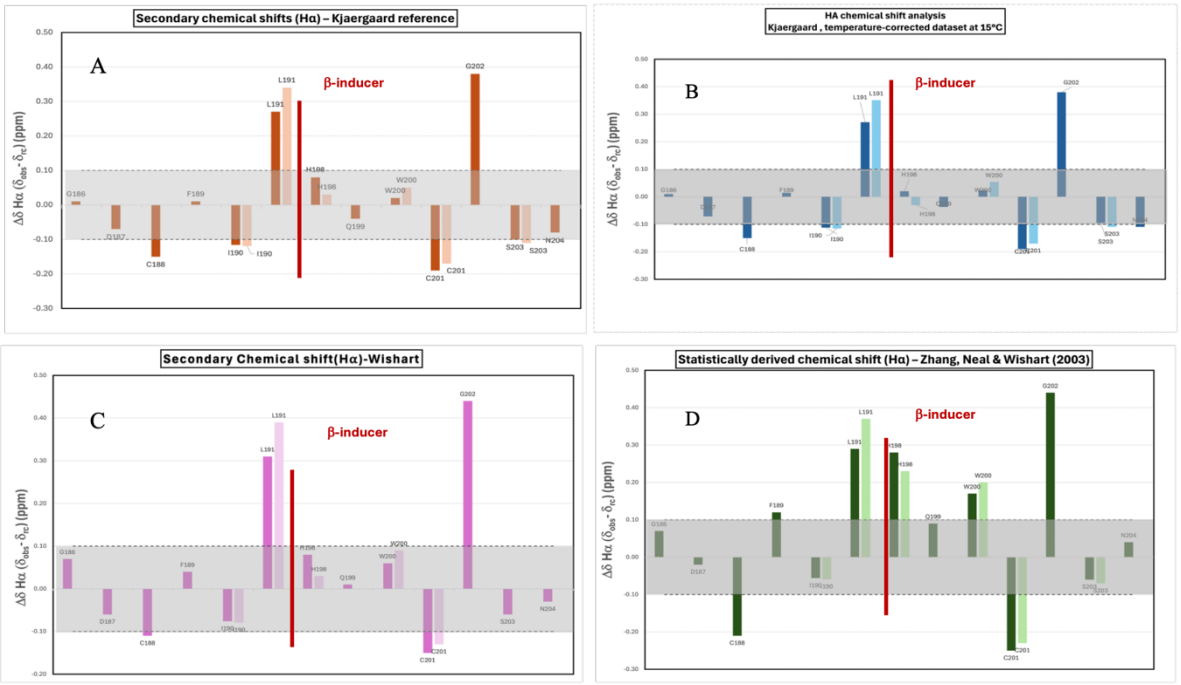


Figure 25 Secondary Hα chemical shifts (Δδ Hα = δ_obs − δ_rc) for LB6-1 and LB6-2 across residues G186–N204, calculated using four different random-coil reference datasets. (A) Kjaergaard random-coil reference; LB6-1 in orange, LB6-2 in light orange. (B) Kjaergaard temperature-corrected reference (15 °C); LB6-1 in blue, LB6-2 in light blue. (C) Wishart random-coil reference; LB6-1 in pink, LB6-2 in light pink. (D) Statistically derived random-coil reference of Zhang, Neal & Wishart (2003); LB6-1 in green, LB6-2 in light green. The shaded grey band (±0.1 ppm) marks the threshold below which deviations are considered insignificant; dashed lines indicate its limits. The vertical red line marks the position of the β-inducer. Positive deviations are indicative of extended/β-strand conformation, whereas negative deviations indicate helical propensity. Across all four referencing schemes, L191 and G202 show consistent positive deviations, while C188 and C201 show consistent negative deviations, indicating that these conformational features are robust to the choice of random-coil reference.

A clear common feature emerges across all four analyses: the pattern of deviations is highly reproducible regardless of which reference dataset is used. Even the temperature correction (panel B) does not change the global observed behavior.

The same residues deviate in the same direction and by comparable magnitude in every case, validating that the observed secondary shifts are not biased by the choice of random-coil values.

For most residues, the Hα deviation from random coil is small, generally falling within a ±0.1 ppm band, consistent with a predominantly disordered conformation. Two residues stand out as exceptions across all four reference sets: L191 and G202. However, neither of these deviations is structurally significant on its own. Assignment of a genuine secondary structure element (such as an α-helix) by chemical shift requires a sustained run of several consecutive residues deviating consistently in the same direction, by convention, at least three residues in a row, rather than an isolated single-residue excursion; L191 and G202 both appear as isolated deviations, not part of such a run, and are therefore not indicative of a stable folded element.

A closer look at the C-terminal region nonetheless reveals a real, localized feature. The Hα secondary shift inverts sign between C201 (−0.19 ppm) and G202 (+0.38 ppm), a residue-to-residue sign flip that is inconsistent with a smoothly propagating helix or strand, where a

consistent sign is expected to be sustained over the residue run. This kind of local sign inversion instead reflects a change in backbone direction, characteristic of a turn.

Secondary chemical shifts were also calculated for Cα carbons and are summarized in Figure 25:

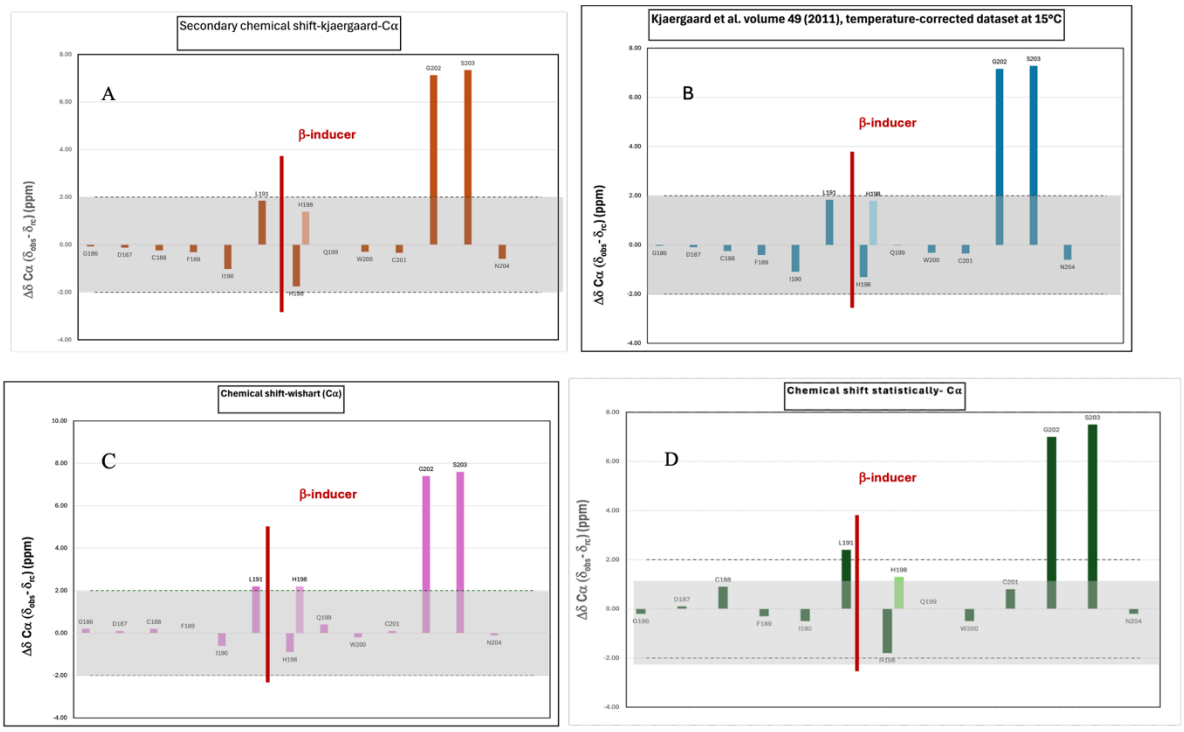


Figure 26 Secondary Cα chemical shifts (Δδ Cα = δ_obs − δ_rc) for LB6-1 and LB6-2 across residues G186–N204, calculated using four different random-coil reference datasets. (E) Kjaergaard random-coil reference; LB6-1 in orange, LB6-2 in light orange. (F) Kjaergaard et al., volume 49 (2011) temperature-corrected reference (15 °C); LB6-1 in blue, LB6-2 in light blue. (G) Wishart random-coil reference; LB6-1 in pink, LB6-2 in light pink. (H) Statistically derived random-coil reference; LB6-1 in green, LB6-2 in light green. The shaded grey band (±2.0 ppm) marks the threshold below which deviations are considered insignificant; dashed lines indicate its limits. The vertical red line marks the position of the β-inducer. For Cα, positive deviations indicate helical propensity, whereas negative deviations indicate extended/β-strand conformation. G202 and S203 show large positive deviations in all four referencing schemes, and L191 lies at or above the significance threshold throughout, indicating that these features are robust to the choice of random-coil reference.

Cα secondary shifts observed for G202 and S203 are both large and consistently positive (approximately +7.1 and +7.3 ppm, respectively) across all four reference datasets, a signal robust enough to be considered significant. Because a genuine α-helix requires the paired signature of a negative Hα together with a positive Cα sustained over consecutive residues,

and this pairing is not observed at G202 (whose Hα deviation is positive, not negative), this feature is better described as a localized turn than a helical element.

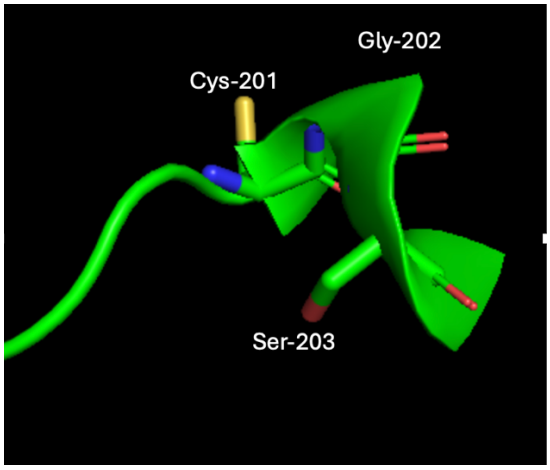


Figure 27 Representative MD snapshot of the C-terminal C201–G202–S203 segment of LB6, shown in cartoon representation with the side chains of Cys-201, Gly-202 and Ser-203 displayed as sticks (carbon, green; oxygen, red; nitrogen, blue; sulfur, yellow). The backbone reverses direction over these three residues, forming a localized turn rather than a continuous helical or extended element. This is consistent with the NMR secondary chemical shifts, which show a residue-to-residue sign inversion in Δδ Hα between C201 (−0.19 ppm) and G202 (+0.38 ppm) together with large positive Δδ Cα at G202 and S203 (≈ +7.1 and +7.3 ppm), a combination incompatible with a propagating α-helix, for which a sustained negative Hα paired with a positive Cα would be expected. MD simulations were performed in collaboration with Dr. Matteo De Rosa and Prof. Stefano Pieraccini.

This interpretation is consistent with the turn identified in the parent structural characterization of LB6 performed by MD calculation by Prof. Stefano Pieraccini, University of Milan,

Taken together, the chemical shift data support a picture of LB6 as a largely disordered peptidomimetic, rather than a peptide populating any sustained helical or extended conformation.C201–S203 turn can be considered a robust, local structural feature, cross-validated by experimental solution NMR data and MD simulations.

5.5 Inference of secondary structure populations from chemical shifts

The secondary chemical shift analysis of the preceding section reports, residue by residue, the deviation of an observed nucleus from its random coil value, and the sign and magnitude of that deviation were interpreted according to the conventions set out in Chapter 2. That

treatment establishes where along the sequence the local conformation departs from a fully disordered chain, but it does not express the result as a population.

A deviation of a given size may arise either from a small fraction of the ensemble adopting a well-formed element or from a larger fraction adopting a partially formed one, and the two cannot be distinguished from a single Δδ profile.

For a molecule such as LB6, which the NMR data already show to interconvert between two conformational states and which is expected based on its size and composition to sample a broad ensemble within each of them, the population is the quantity of principal interest.

A second limitation of the reference-based treatment is that it examines each nucleus separately. The Hα and Cα profiles report on the same backbone geometry and are therefore not independent but inspecting them side by side provides no principled way of weighting one against the other, nor of recognizing the combination of deviations that distinguishes an ordered turn from a genuinely disordered segment. Since the structural features anticipated for a peptidomimetic of this design are precisely short local turns rather than extended regular elements, this limitation bears directly on the interpretation.

The chemical shift data were therefore analyzed a second time with CheSPI (Nielsen and Mulder, 2021), whose principles are described in Section 2.5. In place of a tabulated reference, CheSPI generates its own random coil values with POTENCI for the exact sequence and experimental conditions; it combines the secondary chemical shifts of all available nuclei into two multivariate components; and it returns, for each residue, a CheZOD Z-score quantifying local order, the populations of four structural classes, and a prediction on the eight-class DSSP scheme.

CheSPI is parametrized on natural polypeptides and treats its input as a continuous chain of standard amino acids. The non-natural piperidine–pyrrolidine linker of LB6 cannot be represented in such a scheme, and submitting the concatenated sequence NGDCFILHQWCGSN would require the program to treat Leu191 and His198 as covalently adjacent residues. Because the random coil prediction for any residue depends explicitly on the identity of the two residues on either side of it, this would introduce a systematic error

precisely in the region of the peptide that is of greatest interest, namely the segments flanking the linker in which the conformational doubling is observed. The two arms were therefore submitted as separate sequences, NGDCFIL for the first and HQWCGSN for the second, and POTENCI generated an independent random coil reference for each.

Four input files were prepared in NMR-STAR format, corresponding to the two arms in each of the two conformational states. The experimental conditions were specified as 288 K, pH 5.0 and an ionic strength of 0.3 M, matching the sample composition described in Section 3.1, and the referencing corrections established in Section 5.4 were applied to the input shifts before submission. No data were entered for Asn185, for which no amide resonance was observed owing to fast exchange with the solvent at the free N-terminus.

One consequence of the way the data were acquired should be made explicit. Several residues give two sets of ^{1}H resonances but only a single Cα value, since the natural-abundance ^{1}H–^{13}C HSQC does not resolve the two states at every position; in these cases, the same Cα was entered in both files. In the first arm all six Cα values are therefore common to the two states, and in the second arm only that of His198 differs between them. The two Hβ resonances of Ser203 were likewise observed only for the first state, so that the dataset for the second contains two fewer values at this position. The differences reported below between LB6-1 and LB6-2 consequently rest principally on the ^{1}H data and are smaller than they would be if a complete set of carbon shifts were available for both states.

Three distinct limitations affect the results presented in this section, and it is useful to separate them, since they act in different ways and have different remedies.

The first arises from the set of nuclei available. CheSPI computes its components from the secondary chemical shifts of seven nuclei, and the loading plot published with the method shows that the weights of the first component follow the order Hα / C′ / Cα / Cβ ≫ N / Hβ / HN, whilst the second component is dominated by HN and C′. The dataset assembled here contains HN, Hα, Hβ and Cα. The three nuclei that are absent, N, Cβ and C′, are not a random selection: C′ and Cβ are the second and fourth most heavily weighted contributors to the first component, and C′ is one of the two dominant contributors to the second. Their absence is

intrinsic to the sample rather than accidental, since the carbonyl carbon bears no proton and is therefore invisible in a natural-abundance ^{1}H–^{13}C HSQC, Cβ lies outside the Hα–Cα region that was recorded, and the direct observation of ^{15}N at natural abundance is not practical at the concentrations available.

The practical consequence is that the first component is determined here essentially by Hα and Cα, and the second by HN and Hβ, that is, by its two weakest contributors. Since it is the second component that separates an ordered turn from an unstructured segment, the discrimination between these two possibilities is less well determined in the present analysis than it would be for a uniformly labelled sample.

The second limitation follows from the length of the fragments. CheSPI excludes the terminal residues of a sequence from the analysis, so that results are returned for five of the seven residues of each arm: Gly186 to Ile190 in the first and Gln199 to Ser203 in the second. Asn185 and Leu191 in the first arm, and His198 and Asn204 in the second, therefore fall outside the analysis. It should be emphasized that the chemical shifts of these residues were nevertheless included in the input files, since the prediction made for any position draws on the shifts of its neighbors; their omission from the tables is a property of the method and not an indication that they were left out.

A further consequence of the same constraint is that the automatic re-referencing correction implemented in CheSPI could not be evaluated. That procedure estimates a candidate offset for each nucleus from a running average of the observed deviations along the sequence, using windows of nine residues, and neither arm is long enough for such a window to be defined; the program reported explicitly that no running offset could be estimated. The internal consistency check that CheSPI would normally perform on the referencing of its input was therefore unavailable, and the correctness of the referencing rests entirely on the independent determination described in Section 5.3.

The third limitation is a consequence of the decision to treat the arms separately. In the intact molecule neither Leu191 nor His198 is a chain terminus: the carboxyl group of Leu191 is amidated to the nitrogen of the piperidine ring, and the amino group of His198 receives an

amide bond from the pyrrolidine carbonyl. Asn204 is likewise amidated rather than a free carboxylate. When the arms are submitted as isolated peptides, POTENCI applies the corrections appropriate to free termini at these three positions, and the reference values generated for them, and to a lesser extent for their immediate neighbors Ile190 and Gln199, therefore carry a systematic error whose sign and magnitude are not known. Of the four termini of the two fragments only that of Asn185, which is genuinely a free amine, is correctly modelled. This limitation is intrinsic to the approach and cannot be removed without a random coil predictor able to accept non-natural residues; it is noted here because the affected positions lie adjacent to the region in which structure is detected.

Before the results are presented, one point of nomenclature should be clarified. LB6-1 and LB6-2 do not denote two states of an individual residue but two states of the whole molecule, and every residue is present in both. What distinguishes the doubled residues is that their resonances are resolved between the two states, so that a different chemical shift can be entered for each; for the remaining residues a single set of shifts was observed, and the same values were used in both datasets.

Since CheSPI analyses each dataset in its entirety, a result is returned for every analyzed residue in both, and the two may differ even for a residue that is not itself doubled, because the calculation for any position draws on the shifts of its neighbors. The tables that follow therefore state separately, for each residue, whether doubling was observed experimentally and what each of the two analyses returned.

The CheZOD Z-scores are collected in Table 4.2. On this scale a value below 5.0 identifies a residue as disordered, values between about 5 and 8 indicate a partially ordered residue, and values above 8 indicate a well-ordered one.

Residue	Doubled	Z (LB6-1)	Z (LB6-2)
Asn185	—	not observed	not observed
Gly186	—	2.92	2.92
Asp187	—	4.91	4.91

Cys188	—	6.88	6.88
Phe189	—	7.35	7.57
Ile190	yes	6.13	6.40
Leu191	yes	terminal position	terminal position
X (linker)	—	not represented	not represented
His198	yes	terminal position	terminal position
Gln199	—	5.34	6.23
Trp200	yes	6.18	7.32
Cys201	yes	6.90	7.98
Gly202	—	8.41	7.55
Ser203	yes	7.85	6.51
Asn204	—	terminal position	terminal position

Table 4.2 CheZOD Z-scores obtained for the two arms of LB6. The complete sequence is listed, so that the residues for which no value is returned remain visible. The column headed “Doubled” indicates whether the residue itself gives two sets of resonances in the NMR spectra; LB6-1 and LB6-2 designate the two datasets, each of which describes the peptide in one of its conformational states. Values below 5.0 indicate a disordered residue. Residues marked “terminal position” lie at the end of a fragment and are excluded by the method, although their chemical shifts were included in the input.

Two features of the table deserve comment. The first is that the values obtained for Gly186, Asp187 and Cys188 are identical in the two conformational states, as they must be, since none of these residues is doubled and the corresponding input values are the same in both files. The agreement therefore serves as an internal check that the four input files were correctly constructed. Differences between the two states appear only from Phe189 onwards in the first arm and from Gln199 onwards in the second, which is where the doubled resonances begin to contribute.

The second is that the profile describes a peptide that is disordered but not uniformly so. Gly186 is fully disordered by the CheZOD criterion and Asp187 lies below the threshold, while the remaining residues occupy the intermediate range, indicating a measurable but

modest degree of local ordering. Only Gly202 in the first conformational state exceeds a value of 8. The gradient along the first arm, from 2.92 at Gly186 to values above 7 at Phe189, is consistent with the greater conformational freedom expected close to the free N-terminus and provides a useful internal scale: the same molecule contains, within five residues, positions that the method classifies as fully disordered and positions that it classifies as partially ordered.

The four-class populations are collected in **Table 4.3**. The classes correspond to groupings of the eight DSSP designations: helical structure comprises the α-, 3₁₀- and π-helices, extended structure comprises β-strand and isolated β-bridge, turn is taken alone, and non-folded structure comprises bend and coil.

Residue	Doubled	Dataset	Helical	Turn	Non-folded	Extended
Asn185	—	not observed				
Gly186	—	both	0.031	0.112	0.792	0.065
Asp187	—	both	0.031	0.120	0.777	0.071
Cys188	—	both	0.028	0.125	0.749	0.098
Phe189	—	LB6-1	0.028	0.124	0.740	0.108
		LB6-2	0.035	0.140	0.737	0.088
Ile190	yes	LB6-1	0.026	0.119	0.749	0.107
		LB6-2	0.033	0.135	0.747	0.085
Leu191	yes	terminal position				
X (linker)	—	not represented				
His198	yes	terminal position				
Gln199	—	LB6-1	0.023	0.112	0.776	0.090

		LB6-2	0.025	0.109	0.785	0.081
Trp200	yes	LB6-1	0.023	0.100	0.783	0.094
		LB6-2	0.027	0.102	0.789	0.082
Cys201	yes	LB6-1	0.178	0.292	0.489	0.040
		LB6-2	0.216	0.287	0.467	0.030
Gly202	—	LB6-1	0.901	0.051	0.043	0.005
		LB6-2	0.386	0.249	0.349	0.016
Ser203	yes	LB6-1	0.955	0.030	0.015	0.000
		LB6-2	0.689	0.126	0.174	0.011
Asn204	—	terminal position				

Table 4.3 CheSPI populations of the four structural classes. The column headed “Doubled” indicates whether the residue itself gives two sets of resonances. LB6-1 and LB6-2 designate the two datasets; “both” is entered where they give identical values. Populations are normalized to unity for each residue. Residues marked “terminal position” are excluded by the method, and the linker cannot be represented in a sequence of standard amino acid.

The first arm establishes what a disordered residue looks like in this analysis, and it is worth setting out that baseline explicitly, because it is the reference against which the second arm must be read. At every one of its five analyzed positions, and in both conformational states, the non-folded class accounts for approximately 74 to 79 per cent of the population, the turn class for 11 to 14 per cent, the extended class for 7 to 11 per cent and the helical class for less than 4 per cent. A fully disordered residue is therefore not one for which the non-folded population reaches 100 per cent: even in the absence of any persistent structure, a residue transiently samples turn-like and extended backbone angles, and the populations reported here quantify that sampling. There is no position in the first arm at which any ordered class becomes dominant, and the two conformational states differ by no more than a few per cent at any residue. On this evidence the first arm of LB6 does not populate a defined secondary structure element in solution.

The second arm behaves in the same way over its first two analyzed residues. Gln199 and Trp200 are indistinguishable from the residues of the first arm, with the non-folded class close to 78 per cent in both states. From Cys201 onwards the profile changes, and the departure is best expressed relative to the baseline just defined. At Cys201 the turn population reaches 29.2 and 28.7 per cent in the two states, between two and three times the baseline value, and the helical population reaches 17.8 and 21.6 per cent, six to seven times the baseline; the non-folded population falls correspondingly below 50 per cent. At Gly202 and Ser203 the helical class becomes dominant, reaching 90.1 and 95.5 per cent in the first conformational state, some thirty times the baseline value.

This region corresponds to the C201–S203 turn reported previously for LB6 based on the positive Δδ Cα values at Gly202 and Ser203 (Bollati et al., 2026), and the classification is consistent with that assignment. A helical population in the CheSPI scheme includes the 3_{10}-helix, and a 3_{10}-helical element extending over three residues is, in geometric terms, a tight turn rather than a propagating helix; the two descriptions refer to the same local structure and differ only in the classification scheme applied to it. The difference between the two conformational states at these positions, with populations of 90.1 and 95.5 per cent in LB6-1 against 38.6 and 68.9 per cent in LB6-2, indicates that the element is more fully formed in the first state, although the caveat of Section 5.4 should be recalled: the Cα values entering the two datasets at Trp200, Cys201 and Ser203 are identical, and the difference therefore rests on the ^{1}H data alone, together with the two Hβ values of Ser203 that were observed only for the first state.

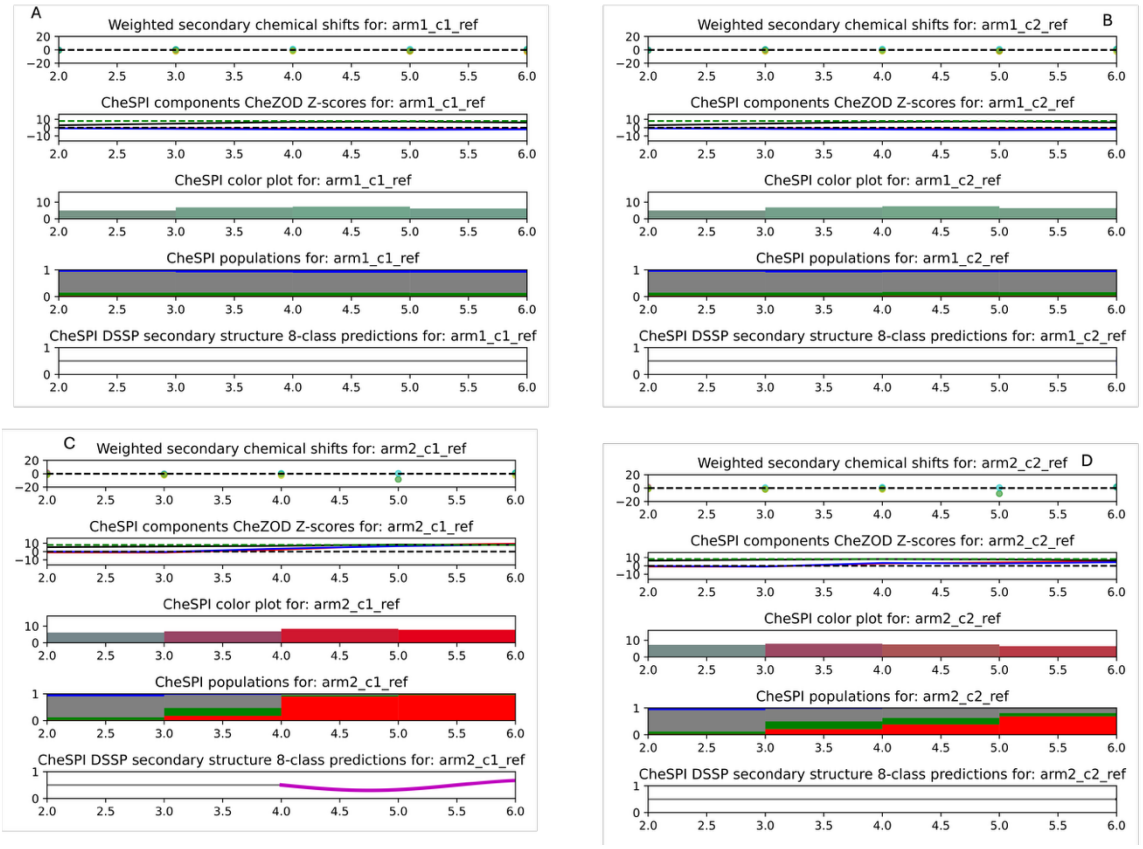


Figure 28 CheSPI output for the two arms of LB6 in the two conformational states: the first arm above and the second below, LB6-1 on the left and LB6-2 on the right. In each panel the upper trace is the CheSPI color plot, in which the color encodes the first two CheSPI components and the height the CheZOD Z-score; the middle trace is the stacked distribution of the four structural populations, with helical structure in red, extended structure in blue, turn in green and non-folded structure in grey; and the lower trace is the most probable eight-class DSSP prediction, drawn with magenta for the 3_{10}-helix. The uniform grey of the first arm, in both conformational states, contrasts with the progression from grey to red along the second arm and identifies the C-terminal segment of that arm as the only region of the peptide showing appreciable local order.

The most probable eight-class assignments returned for the four datasets were as follows. For the first arm, the first conformational state gave no secondary structure element at any position and the second gave an isolated β-bridge at the C-terminal residue only. For the second arm, the first conformational state gave a 3_{10}-helix spanning Gly202 to Asn204 together with an isolated bridge at the N-terminal residue, and the second gave only a turn and a bend. These assignments are consistent with the populations of Table 4.3 and with the conclusion that the sole region of LB6 showing appreciable local order lies at the C-terminal end of the second arm.

Because this element rests on the secondary chemical shifts of a small number of nuclei, its robustness was tested by repeating the analysis for the second arm with the Cα values of Gly202 and Ser203 excluded from the input, all other data being unchanged. The result is informative. The turn population at Cys201 falls from 0.292 to 0.084, the helical populations at Gly202 and Ser203 fall from 0.901 and 0.955 to 0.018 and 0.019, and the non-folded population becomes uniform along the whole arm at between 0.71 and 0.78, matching the profile of the first arm. The eight-class prediction loses the 3₁₀-helix entirely.

Two conclusions follow. The first is that the structural element identified at the C-terminal end of the second arm is carried entirely by the Cα shifts of these two residues, the Hα, HN and Hβ data alone being insufficient to support it. This is not in itself surprising, given that Cα is the only heavily weighted carbon available in the present dataset and that the two nuclei of comparable weight, C′ and Cβ, are absent; but it does mean that the finding cannot be regarded as independently corroborated within the CheSPI analysis itself. The second is a caution about the eight-class assignment for fragments of this length. With five residues reported per arm, and with the assignment obtained by a stochastic optimization over the whole fragment subject to the grammar of permitted elements, the removal of two values is sufficient to change it completely, and repeated runs on the reduced dataset returned different assignments of comparable likelihood. The populations, which are computed residue by residue, are the more robust of the two outputs, and the eight-class assignment should be regarded as indicative rather than definitive.

The CheSPI analysis of the two arms of LB6, carried out independently of the reference-based treatment of Section 5.3 and using a random coil prediction generated for the exact experimental conditions, describes a peptidomimetic that is predominantly disordered in solution. In the first arm the non-folded class accounts for roughly three quarters of the population at every analyzed residue in both conformational states, with no ordered element at any position, and the same is true of Gln199 and Trp200 in the second arm. The segment from Cys201 to Ser203 carries a local element which CheSPI classifies as 3₁₀-helical and which corresponds to the turn reported previously for this region; it is more fully populated

in the first conformational state than in the second, and it is the only feature of the peptide that departs from the disordered baseline.

The analysis therefore reproduces, by an independent route and in quantitative form, the conclusion drawn from the secondary chemical shifts in Section 5.3 and adds to it a per-residue estimate of the populations involved.

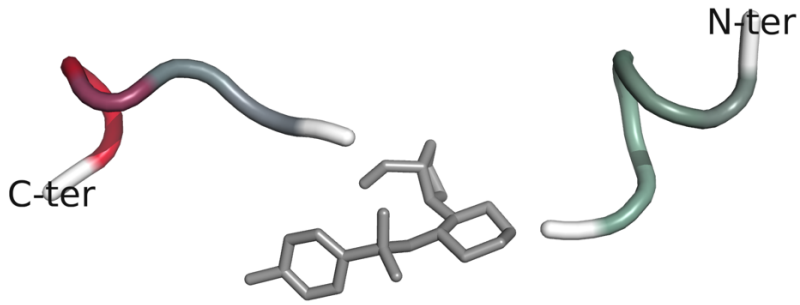


Figure 29 CheSPI colors mapped onto a representative conformer of LB6 taken from the all-atom molecular dynamics' simulation. The color of each residue encodes the first two CheSPI components, on the scale defined in Section 2.6, in which well-formed helices appear red, well-formed strands blue, turns green and disordered residues grey. Residues for which no CheSPI value is returned, namely Asn185, Leu191, His198 and Asn204, together with the piperidine–pyrrolidine linker, are shown in neutral grey. The uniform pale coloring of the first arm and of the N-terminal half of the second reflects the disordered populations of Table 4.3, whilst the segment from Cys201 to Ser203 carries the only coloring indicative of local order. The molecular dynamics coordinates were provided by Dr Matteo de Rosa and Prof. Stefano Pieraccini.

Conclusion

Solution NMR spectroscopy studies were performed on LB-6, a peptidomimetic inhibitor which is effective against the amyloid aggregation of both the gelsolin amyloidogenic core ($GAC_{182–192}$) and the full G2 - D187N domain of gelsolin, to derive structural information that helps to unravel the peptide's mechanism of action and understand the molecular basis of its inhibitory effect on aggregation.

Solution NMR was chosen as it is the elective technique giving atomic-resolution structural information on small soluble peptides that may populate several conformations.

Residue level NMR investigation highlighted the presence of two distinct patterns of resonances for many residues located in both peptide arms. These duplicated signals indicate the presence of two conformational states in slow exchange on the millisecond timescale, most likely corresponding to alternate orientations of the piperidine linker.

Complete chemical-shift assignments of peptidic resonances and their analysis in terms of Hα and Cα secondary chemical shifts, to derive information on stable secondary structure elements, revealed predominantly random-coil behavior of LB-6 peptide. A transiently structured segment was observed in the second arm at the level of residues C201–S203, possibly forming a rather stable turn. This turn is further confirmed by secondary structure population predictions obtained by CheSPI program, which returns a non-folded population of about 75% throughout the molecule and a 3_{10}-helical element at the level of this segment.

The absence of short-range inter-strand NOEs in NOESY/ROESY spectra excluded the presence of an intramolecular Cys188-Cys201 disulfide bridge.

Altogether NMR data demonstrate that the β-hairpin structure, typical of the peptidic segment within the G2 domain, is scarcely populated in solution and LB6 behaves as a highly flexible, dynamic molecule with only local, transient structural elements in the C-terminal arm. The structural experimental data collected on LB6 by NMR suggest that its conformational flexibility is an important determinant of its inhibitory potency.

It can be speculated that the populated dynamic ensemble of LB6 conformations is crucial for interacting with transient and polymorphic species, such as early G2-D187N oligomers, thus counteracting their further fibrillation and pathological deposit in tissues.

REFERENCES

Adler-Abramovich, L. (2012) “Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria,” *Nature Chemical Biology*, 8(8), pp. 701–706. Available at: https://doi.org/10.1038/nchembio.1002.

Bax, A. and Davis, D.G. (1985) “Practical aspects of two-dimensional transverse NOE spectroscopy,” *Journal of Magnetic Resonance (1969)*, 63(1), pp. 207–213. Available at: https://doi.org/10.1016/0022-2364(85)90171-4.

Bellomo, G. *et al.* (2018) “Aggregation kinetics of the Aβ1–40 peptide monitored by NMR,” *Chemical Communications*, 54(55), pp. 7601–7604. Available at: https://doi.org/10.1039/C8CC01710G.

Bollati, M. *et al.* (2021) “A novel hotspot of gelsolin instability triggers an alternative mechanism of amyloid aggregation,” *Computational and Structural Biotechnology Journal*, 19, pp. 6355–6365. Available at: https://doi.org/10.1016/j.csbj.2021.11.025.

Bollati, M. *et al.* (2022) “Rational Design of a Peptidomimetic Inhibitor of Gelsolin Amyloid Aggregation,” *International Journal of Molecular Sciences*, 23(22), p. 13973. Available at: https://doi.org/10.3390/ijms232213973.

Bollati, M., Natale, C., *et al.* (2026) “A molecular perspective of gelsolin amyloidosis: An old foe with new faces,” *Cellular and Molecular Life Sciences*, 83(1), p. 177. Available at: https://doi.org/10.1007/s00018-026-06172-7.

Bollati, M., Peqini, K., *et al.* (2026) “Modulation of early aggregation events by flexible peptidomimetics inhibits gelsolin amyloid formation,” *International Journal of Biological Macromolecules*, 375, p. 153423. Available at: https://doi.org/10.1016/j.ijbiomac.2026.153423.

Bollati, M. (2026) “Modulation of early aggregation events by flexible peptidomimetics inhibits gelsolin amyloid formation.”

Bonì, F. *et al.* (2016) “Molecular basis of a novel renal amyloidosis due to N184K gelsolin variant,” *Scientific Reports*, 6(1), p. 33463. Available at: https://doi.org/10.1038/srep33463.

Buxbaum, J.N. *et al.* (2024) “Amyloid nomenclature 2024: update, novel proteins, and recommendations by the International Society of Amyloidosis (ISA) Nomenclature Committee,” *Amyloid*, 31(4), pp. 249–256. Available at: https://doi.org/10.1080/13506129.2024.2405948.

Bylesjö, M. *et al.* (2006) “OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification,” *Journal of Chemometrics*, 20(8–10), pp. 341–351. Available at: https://doi.org/10.1002/cem.1006.

Chen, K., Freedberg, D.I. and Keire, D.A. (2015) “NMR profiling of biomolecules at natural abundance using 2D 1H–15N and 1H–13C multiplicity-separated (MS) HSQC spectra,” *Journal of Magnetic Resonance*, 251, pp. 65–70. Available at: https://doi.org/10.1016/j.jmr.2014.11.011.

Chiti, F. and Dobson, C.M. (no date) “Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade.”

Dobson, C.M. (2003) “Protein folding and misfolding.”

Eisenberg, D.S. and Sawaya, M.R. (2017) “Structural Studies of Amyloid Proteins at the Molecular Level,” *Annual Review of Biochemistry*, 86(1), pp. 69–95. Available at: https://doi.org/10.1146/annurev-biochem-061516-045104.

Fändrich, M. and Dobson, C.M. (2002) “The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation,” *The EMBO Journal*, 21(21), pp. 5682–5690. Available at: https://doi.org/10.1093/emboj/cdf573.

Feldt, J. *et al.* (2018) “Structure, regulation and related diseases of the actin-binding protein gelsolin,” *Expert Reviews in Molecular Medicine*, 20, p. e7. Available at: https://doi.org/10.1017/erm.2018.7.

Fitzpatrick, A.W.P. *et al.* (2013) “Atomic structure and hierarchical assembly of a cross-β amyloid fibril,” *Proceedings of the National Academy of Sciences*, 110(14), pp. 5468–5473. Available at: https://doi.org/10.1073/pnas.1219476110.

Hinds, M.G. and Norton, R.S. (no date) “NMR spectroscopy of peptides and proteins.”

Iadanza, M.G. *et al.* (2025) “A new era for understanding amyloid structures and disease.”

Kabsch, W. and Sander, C. (1983) “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features,” *Biopolymers*, 22(12), pp. 2577–2637. Available at: https://doi.org/10.1002/bip.360221211.

Kjaergaard, M., Brander, S. and Poulsen, F.M. (2011) “Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH,” *Journal of Biomolecular NMR*, 49(2), pp. 139–149. Available at: https://doi.org/10.1007/s10858-011-9472-x.

Lövestam, S. *et al.* (2024) “Disease-specific tau filaments assemble via polymorphic intermediates,” *Nature*, 625(7993), pp. 119–125. Available at: https://doi.org/10.1038/s41586-023-06788-w.

Marsh, J.A. *et al.* (2006) “Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation,” *Protein Science*, 15(12), pp. 2795–2804. Available at: https://doi.org/10.1110/ps.062465306.

Nelson, R. *et al.* (2005) “Structure of the cross-β spine of amyloid-like fibrils,” *Nature*, 435(7043), pp. 773–778. Available at: https://doi.org/10.1038/nature03680.

Nguyen, B.A., Singh, V., Afrin, S., Singh, P., *et al.* (2024) “Cryo-EM confirms a common fibril fold in the heart of four patients with ATTRwt amyloidosis,” *Communications Biology*, 7(1), p. 905. Available at: https://doi.org/10.1038/s42003-024-06588-6.

Nguyen, B.A., Singh, V., Afrin, S., Yakubovska, A., *et al.* (2024) “Structural polymorphism of amyloid fibrils in ATTR amyloidosis revealed by cryo-electron microscopy,” *Nature Communications*, 15(1), p. 581. Available at: https://doi.org/10.1038/s41467-024-44820-3.

Nielsen, J.T. and Mulder, F.A.A. (2016) “There is Diversity in Disorder—‘In all Chaos there is a Cosmos, in all Disorder a Secret Order,’” *Frontiers in Molecular Biosciences*, 3. Available at: https://doi.org/10.3389/fmolb.2016.00004.

Nielsen, J.T. and Mulder, F.A.A. (2020) “Quantitative Protein Disorder Assessment Using NMR Chemical Shifts,” in B.B. Kragelund and K. Skriver (eds.) *Intrinsically Disordered Proteins*. New York, NY: Springer US (Methods in Molecular Biology), pp. 303–317. Available at: https://doi.org/10.1007/978-1-0716-0524-0_15.

Nielsen, J.T. and Mulder, F.A.A. (2021a) “CheSPI: chemical shift secondary structure population inference,” *Journal of Biomolecular NMR*, 75(6–7), pp. 273–291. Available at: https://doi.org/10.1007/s10858-021-00374-w.

Nielsen, J.T. and Mulder, F.A.A. (2021b) “CheSPI: chemical shift secondary structure population inference,” *Journal of Biomolecular NMR*, 75(6–7), pp. 273–291. Available at: https://doi.org/10.1007/s10858-021-00374-w.

Pagano, K. *et al.* (2024) “Characterizing the Oligomers Distribution along the Aggregation Pathway of Amyloid Aβ1-40 by NMR,” *Chemistry – A European Journal*, 30(38), p. e202400594. Available at: https://doi.org/10.1002/chem.202400594.

Qiang, W. *et al.* (2017) “Structural variation in amyloid-β fibrils from Alzheimer’s disease clinical subtypes,” *Nature*, 541(7636), pp. 217–221. Available at: https://doi.org/10.1038/nature20814.

Schweighauser, M. *et al.* (2020) “Structures of α-synuclein filaments from multiple system atrophy,” *Nature*, 585(7825), pp. 464–469. Available at: https://doi.org/10.1038/s41586-020-2317-6.

Shi, Y. *et al.* (2021) “Structure-based classification of tauopathies,” *Nature*, 598(7880), pp. 359–363. Available at: https://doi.org/10.1038/s41586-021-03911-7.

Sims, J.R. (2023) “Donanemab in Early Symptomatic Alzheimer Disease,” *JAMA*, 330(6), pp. 512–527. Available at: https://doi.org/10.1001/jama.2023.13239.

Tamiola, K. and Mulder, F.A.A. (2012) “Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins,” *Biochemical Society Transactions*, 40(5), pp. 1014–1020. Available at: https://doi.org/10.1042/BST20120171.

Van Der Perren, A. *et al.* (2020) “The structural differences between patient-derived α-synuclein strains dictate characteristics of Parkinson’s disease, multiple system atrophy and dementia with Lewy bodies,” *Acta Neuropathologica*, 139(6), pp. 977–1000. Available at: https://doi.org/10.1007/s00401-020-02157-3.

Van Dyck, C.H. *et al.* (2023) “Lecanemab in Early Alzheimer’s Disease,” *New England Journal of Medicine*, 388(1), pp. 9–21. Available at: https://doi.org/10.1056/NEJMoa2212948.

Vaquer-Alicea, J., Diamond, M.I. and Joachimiak, L.A. (2021) “Tau strains shape disease,” *Acta Neuropathologica*, 142(1), pp. 57–71. Available at: https://doi.org/10.1007/s00401-021-02301-7.

Willard, L. (2003) “VADAR: a web server for quantitative evaluation of protein structure quality,” *Nucleic Acids Research*, 31(13), pp. 3316–3319. Available at: https://doi.org/10.1093/nar/gkg565.

Wishart, D.S. (2011) “Interpreting protein chemical shift data,” *Progress in Nuclear Magnetic Resonance Spectroscopy*, 58(1–2), pp. 62–87. Available at: https://doi.org/10.1016/j.pnmrs.2010.07.004.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley-Interscience (Baker Lecture Series). Available at: https://www.wiley.com/en-us/NMR+of+Proteins+and+Nucleic+Acids-p-9780471828938.

Zandomeneghi, G. *et al.* (2004) “FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils,” *Protein Science*, 13(12), pp. 3314–3321. Available at: https://doi.org/10.1110/ps.041024904.

Zhang, H., Neal, S. and Wishart, D.S. (2003) “RefDB: A database of uniformly referenced protein chemical shifts,” *Journal of Biomolecular NMR*, 25(3), pp. 173–195. Available at: https://doi.org/10.1023/A:1022836027055.

Sims, J. R., Zimmer, J. A., Evans, C. D., et al. (2023). Donanemab in Early Symptomatic Alzheimer Disease. *JAMA*, *330*(6), 512–527.

Adler-Abramovich, L. (2012) “Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria,” *Nature Chemical Biology*, 8(8), pp. 701–706. Available at: https://doi.org/10.1038/nchembio.1002.

Bax, A. and Davis, D.G. (1985) “Practical aspects of two-dimensional transverse NOE spectroscopy,” *Journal of Magnetic Resonance (1969)*, 63(1), pp. 207–213. Available at: https://doi.org/10.1016/0022-2364(85)90171-4.

Bellomo, G. *et al.* (2018) “Aggregation kinetics of the Aβ1–40 peptide monitored by NMR,” *Chemical Communications*, 54(55), pp. 7601–7604. Available at: https://doi.org/10.1039/C8CC01710G.

Bollati, M. *et al.* (2021) “A novel hotspot of gelsolin instability triggers an alternative mechanism of amyloid aggregation,” *Computational and Structural Biotechnology Journal*, 19, pp. 6355–6365. Available at: https://doi.org/10.1016/j.csbj.2021.11.025.

Bollati, M. *et al.* (2022) “Rational Design of a Peptidomimetic Inhibitor of Gelsolin Amyloid Aggregation,” *International Journal of Molecular Sciences*, 23(22), p. 13973. Available at: https://doi.org/10.3390/ijms232213973.

Bollati, M. *et al.* (2026) “Modulation of early aggregation events by flexible peptidomimetics inhibits gelsolin amyloid formation,” *International Journal of Biological Macromolecules*, 375, p. 153423. Available at: https://doi.org/10.1016/j.ijbiomac.2026.153423.

Bollati, M. (2026) “Modulation of early aggregation events by flexible peptidomimetics inhibits gelsolin amyloid formation.”

Bonì, F. *et al.* (2016) “Molecular basis of a novel renal amyloidosis due to N184K gelsolin variant,” *Scientific Reports*, 6(1), p. 33463. Available at: https://doi.org/10.1038/srep33463.

Buxbaum, J.N. *et al.* (2024) “Amyloid nomenclature 2024: update, novel proteins, and recommendations by the International Society of Amyloidosis (ISA) Nomenclature Committee,” *Amyloid*, 31(4), pp. 249–256. Available at: https://doi.org/10.1080/13506129.2024.2405948.

Bylesjö, M. *et al.* (2006) “OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification,” *Journal of Chemometrics*, 20(8–10), pp. 341–351. Available at: https://doi.org/10.1002/cem.1006.

Chen, K., Freedberg, D.I. and Keire, D.A. (2015) “NMR profiling of biomolecules at natural abundance using 2D 1H–15N and 1H–13C multiplicity-separated (MS) HSQC spectra,” *Journal of Magnetic Resonance*, 251, pp. 65–70. Available at: https://doi.org/10.1016/j.jmr.2014.11.011.

Chiti, F. and Dobson, C.M. (no date) “Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade.”

Dobson, C.M. (2003) “Protein folding and misfolding.”

Eisenberg, D.S. and Sawaya, M.R. (2017) "Structural Studies of Amyloid Proteins at the Molecular Level," *Annual Review of Biochemistry*, 86(1), pp. 69–95. Available at: https://doi.org/10.1146/annurev-biochem-061516-045104.

Fändrich, M. and Dobson, C.M. (2002) "The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation," *The EMBO Journal*, 21(21), pp. 5682–5690. Available at: https://doi.org/10.1093/emboj/cdf573.

Feldt, J. *et al.* (2018) "Structure, regulation and related diseases of the actin-binding protein gelsolin," *Expert Reviews in Molecular Medicine*, 20, p. e7. Available at: https://doi.org/10.1017/erm.2018.7.

Fitzpatrick, A.W.P. *et al.* (2013) "Atomic structure and hierarchical assembly of a cross-β amyloid fibril," *Proceedings of the National Academy of Sciences*, 110(14), pp. 5468–5473. Available at: https://doi.org/10.1073/pnas.1219476110.

Hinds, M.G. and Norton, R.S. (no date) "NMR spectroscopy of peptides and proteins."

Iadanza, M.G. *et al.* (2025) "A new era for understanding amyloid structures and disease."

Kabsch, W. and Sander, C. (1983) "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, 22(12), pp. 2577–2637. Available at: https://doi.org/10.1002/bip.360221211.

Kjaergaard, M., Brander, S. and Poulsen, F.M. (2011) "Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH," *Journal of Biomolecular NMR*, 49(2), pp. 139–149. Available at: https://doi.org/10.1007/s10858-011-9472-x.

Marsh, J.A. *et al.* (2006) "Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation," *Protein Science*, 15(12), pp. 2795–2804. Available at: https://doi.org/10.1110/ps.062465306.

Nelson, R. *et al.* (2005) "Structure of the cross-β spine of amyloid-like fibrils," *Nature*, 435(7043), pp. 773–778. Available at: https://doi.org/10.1038/nature03680.

Nielsen, J.T. and Mulder, F.A.A. (2020) "Quantitative Protein Disorder Assessment Using NMR Chemical Shifts," in B.B. Kragelund and K. Skriver (eds.) *Intrinsically Disordered Proteins*. New York, NY: Springer US (Methods in Molecular Biology), pp. 303–317. Available at: https://doi.org/10.1007/978-1-0716-0524-0_15.

Nielsen, J.T. and Mulder, F.A.A. (2021) "CheSPI: chemical shift secondary structure population inference," *Journal of Biomolecular NMR*, 75(6–7), pp. 273–291. Available at: https://doi.org/10.1007/s10858-021-00374-w.

Pagano, K. *et al.* (2024) "Characterizing the Oligomers Distribution along the Aggregation Pathway of Amyloid Aβ1-40 by NMR," *Chemistry – A European Journal*, 30(38), p. e202400594. Available at: https://doi.org/10.1002/chem.202400594.

Schweighauser, M. *et al.* (2020) "Structures of α-synuclein filaments from multiple system atrophy," *Nature*, 585(7825), pp. 464–469. Available at: https://doi.org/10.1038/s41586-020-2317-6.

Shi, Y. *et al.* (2021) "Structure-based classification of tauopathies," *Nature*, 598(7880), pp. 359–363. Available at: https://doi.org/10.1038/s41586-021-03911-7.

Sims, J.R. (2023) "Donanemab in Early Symptomatic Alzheimer Disease," *JAMA*, 330(6), pp. 512–527. Available at: https://doi.org/10.1001/jama.2023.13239.

Tamiola, K. and Mulder, F.A.A. (2012) "Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins," *Biochemical Society Transactions*, 40(5), pp. 1014–1020. Available at: https://doi.org/10.1042/BST20120171.

Van Der Perren, A. *et al.* (2020) "The structural differences between patient-derived α-synuclein strains dictate characteristics of Parkinson's disease, multiple system atrophy and dementia with Lewy bodies," *Acta Neuropathologica*, 139(6), pp. 977–1000. Available at: https://doi.org/10.1007/s00401-020-02157-3.

Van Dyck, C.H. *et al.* (2023) "Lecanemab in Early Alzheimer's Disease," *New England Journal of Medicine*, 388(1), pp. 9–21. Available at: https://doi.org/10.1056/NEJMoa2212948.

Vaquer-Alicea, J., Diamond, M.I. and Joachimiak, L.A. (2021) "Tau strains shape disease," *Acta Neuropathologica*, 142(1), pp. 57–71. Available at: https://doi.org/10.1007/s00401-021-02301-7.

Willard, L. (2003) "VADAR: a web server for quantitative evaluation of protein structure quality," *Nucleic Acids Research*, 31(13), pp. 3316–3319. Available at: https://doi.org/10.1093/nar/gkg565.

Wishart, D.S. (2011) "Interpreting protein chemical shift data," *Progress in Nuclear Magnetic Resonance Spectroscopy*, 58(1–2), pp. 62–87. Available at: https://doi.org/10.1016/j.pnmrs.2010.07.004.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. New York: Wiley-Interscience (Baker Lecture Series). Available at: https://www.wiley.com/en-us/NMR+of+Proteins+and+Nucleic+Acids-p-9780471828938.

Zandomeneghi, G. *et al.* (2004) "FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils," *Protein Science*, 13(12), pp. 3314–3321. Available at: https://doi.org/10.1110/ps.041024904.

Zhang, H., Neal, S. and Wishart, D.S. (2003) "RefDB: A database of uniformly referenced protein chemical shifts," *Journal of Biomolecular NMR*, 25(3), pp. 173–195. Available at: https://doi.org/10.1023/A:1022836027055.

van Dyck, C. H., Swanson, C. J., Aisen, P., et al. (2023). Lecanemab in Early Alzheimer's Disease. *New England journal of medicine*, *388*(1), 9–21. https://doi.org/10.1056/nejmoa2212948

Claude (Anthropic, Claude Opus 5) was used only for grammatical correction of the author's own text; no content or data were generated by the tool.

Acknowledgements

In this soil, in this soil, in this pure field,
let us sow no seed but affection, but love.
Rumi, Divan-e Shams, ghazal 1475

Rumi wrote of a field in which nothing, but love is planted, and the two years behind this thesis have felt like that kind of ground. What grew here was not mine alone.

I would like to thank my family and Mr. Fayazi for their constant support, and my friends Somayeh, Alireza, Mahshid, Dario, Najmeh and Gianluca.

I am grateful to Dear Dr. Laura Ragona and to all the members of the CNR-SCITEC, and to Dear Dr. Katiuscia Pagano, for their guidance and their welcome. Dr. Matteo De Rosa and Dr. Michela Bollati (IBF –CNR), Prof. Sara Pellegrino (University of Milan) and Prof. Stefano Pieraccini (University of Milan) are acknowledged for collaborating in LB6 characterization.

Finally, thank you to everyone who made these two years lighter, more meaningful, or simply more memorable.