Conformations of a highly expressed Z19 α-zein studied with AlphaFold2 and MD simulations

Niels Johan Christensen[1*]

[1]Department of Chemistry, University of Copenhagen, Frederiksberg, Denmark

[*]Corresponding author

E-mail: njc@chem.ku.dk (NJC)

Abstract

α-zeins are amphiphilic maize proteins with interesting material properties suitable for numerous applications, e.g., in renewable plastics, foods, therapeutics and additive manufacturing (3D-printing). To exploit their full potential, molecular-level insights are essential. Since α-zeins have resisted experimental atomic-resolution characterization, molecular models have been central to elucidating their structure, but deep-learning α-zein models are largely unexplored. Therefore, this work studies an AlphaFold2 model of a highly expressed zein with molecular dynamics (MD) simulations. The mature protein sequence of the α-zein cZ19C2 gave a loosely packed model with 7 α-helical segments connected by turns/loops and 68% total α-helicity. Compact tertiary structure was limited to a C-terminal bundle of three α-helices, each aligning well with the published repeat sequence NPAAYLQQQQLLPFNQLA(V/A)(L/A). The model was subjected to MD simulations in water containing 0, 3, 23, 50, and 100 mol% ethanol for 400 ns, with extension of selected simulations to the μs timescale. Although the simulations gave structurally diverse endpoints, several patterns were observed: In water and ≤ 2 mol% ethanol, the model rapidly formed compact globular structures, largely preserving the C-terminal bundle. At ≥ 50 mol% ethanol, extended conformations prevailed, consistent with previous SAXS studies. Tertiary structure was partially stabilized in water and in 2 mol% and 23 mol% ethanol, but was disrupted in ≥ 50 mol% ethanol. Averaged results indicated slightly increased helicity with ethanol concentration. Multiple copies of a globular model conformation rapidly formed branched aggregates in aqueous all-atom and coarse-grained MD simulations. β-sheet content was typically <1% across all simulations. In aqueous simulations with glycidyl methacrylate (GMA), 55% of GMA was found within 4 Å of the model. Pre-reaction complexes for methacrylation were indicated by GMA epoxide carbons within 2.9 – 3.3 Å of side chain hydroxyl oxygens, suggesting accessibility of reactive sites in compact α-zein conformations.

Introduction

The growing awareness of the harmful effects of fossil-derived plastics on the Earth's ecosystem has sparked a quest for new biomaterials as sustainable alternatives [1]. Agriculture is one source of such materials, as biopolymers from crops favourably balance accessibility, biocompatibility, and sustainability [2]. Moreover, the physicochemical properties of biopolymers can often be tailored to extend their range of applications [3].

Zeins are examples of biopolymers with desirable material characteristics such as hydrophobicity, thermoplasticity and the ability to form non-covalent viscoelastic networks [4]. They are maize grain storage proteins (prolamins) comprising more than half of the ~10% protein content of maize seeds [5] where they are deposited inside endosperm cells in membrane-enveloped organelles known as protein bodies [6]. There are four major zein groups (α, β, γ, δ) differing by hydrophobicity and apparent molecular weights in gel experiments. α-zein is the most hydrophobic and abundant of the groups, comprising 75 - 85% of the total zein [7].

The properties of zeins have been harnessed in a wide range of applications, such as edible coatings [8], medicinal biomaterials [9], thermoresponsive gels [10], nanoparticles with tuneable morphologies [11], and nanocarriers for hydrophobic drugs [12]. In the field of additive manufacturing (3D-printing) the thermoplasticity of zein has been exploited in fused deposition modelling (FDM) where zein is melted and extruded through a nozzle [13, 14]. Additionally, ethanolic zein solutions have been used for 3D bioprinting [15] and for 4D printing in drug release applications [16]. The tailorability of zein, as evidenced in various studies, has facilitated its broader application. For example, brittleness in microfluidic systems made from zein was reduced by oleic acid and monoglyceride as plasticizers [17]. Reduced zein brittleness and enhanced yield strength have also been achieved by anti-solvent precipitation [18]. Hydration, adhesion, and extensibility have been improved by oxidation with hydrogen peroxide and horseradish peroxidase [19] and by

68 cold plasma treatment [20]. The mechanical properties of zein have been engineered by crosslinking
69 [21, 22]. Recently, excessive aggregation in 4D-printing with zein was prevented by tuning the water
70 content of an ethanol/water supporting bath to 40-75% [16].
71 The α-zein fraction has been extensively studied. Early gel experiments found that it comprises two
72 major α-zein size classes with apparent molecular weights of 19 kDa and 22 kDa, leading to their
73 respective names Z19 and Z22 [23]. It was later discovered that their actual molecular weights were
74 higher, but the names persisted [7]. The first complete zein sequence determined was a Z19 zein
75 obtained in 1981 by Geraghty *et al.* by reverse transcription of a protein body polysome zein mRNA
76 [24]. This protein sequence named A30 (UniProtKB ID: P02859) displayed characteristics later found
77 in all α-zeins. Notably, there was an abundance of dipeptide repeats such as Gln-Gln and the last two-
78 thirds of the protein consisted of 7-8 repeats of a degenerate ~20 amino acid repeat unit. Similar
79 repeats were later found in the Z19 zein A20 (UniProtKB ID: P04703) and the Z22 zein B49
80 (UniProtKB ID: P05815). These repeats were represented by the consensus sequence
81 NPAAYLQQQQLLPFNQLA(V/A)(L/A) [25]. Parallel efforts by Pedersen *et al.* gave the sequence
82 of the Z19-zein ZG99, alternatively named gZ19AB1, with UniProtKB ID: P04704 [26]. This
83 sequence, in conjunction with the sequence of Z22-zein cZ22B1 (UniProtKB ID: P04698) [27], was
84 used to propose the consensus sequence LQQ(F/L)LP(A/F)NQL(A/L)(A/V)ANSPAYLQQ, which
85 formed the basis of a widely referenced structural model of zein [28] described below. Collectively,
86 the studies by Marks *et al.* [29] and Woo *et al.* [30] summarizes major α-zein protein sequences, as
87 outlined in S1 Appendix. The topological tree in Fig 1 shows how these sequences fall into the three
88 subfamilies named “19 kD B”, “19 kD D”, and “22 kD” [30]. Importantly, these studies concur that
89 while there are on the order of 100 zein genes [31] only a subset is detectably expressed in the
90 common inbred maize varieties W64A [29] and B73 [30]. In the latter study, only 8 α-zein genes
91 were substantially expressed. The top-three expressed random selected clones measured by

92 percentage expressed sequence tags (ESTs) were αz19B1, αz19B3, and αz22z1, see Fig 1. The
93 second-highest expressed clone, αz19B3, corresponds to the cZ19C2 zein examined herein.
94
95 **Fig 1. Topological tree for α-zein protein sequences.** Zein names are from the studies [23, 29, 30].
96 UniProtKB IDs are given in brackets. Expressed sequence tag (EST) percentages are given for the
97 three most expressed randomly chosen clones in B73 endosperm [30]. The cZ19C2 zein studied here
98 is indicated by a grey background. The tree excludes signal peptides.
99
100 Every α-zein protein sequence can be divided into four primary sections: a signal peptide, an N-
101 terminal turn, a variable number of approximately 20-residue homologous repeat units, and a C-
102 terminal turn. This is illustrated in Fig 2 using zein gZ19AB1 as an example [28]. In this article, the
103 numbering of residues does not include the signal peptide.
104
105 **Fig 2. Overall α-zein sequence organization.** As an example, the segmentation of gZ19AB1 [28] is
106 indicated by residue numbers. Numbers in brackets exclude the signal peptide.
107
108 There is no general agreement on the definition of the repeat unit and published numbers range from
109 6 to 9 repeats depending on definition and zein sequence [32]. It has been proposed that Z22 zeins
110 are larger than Z19 zeins due to exactly one extra repeat [32]. However, regardless of repeat
111 definition, they constitute about two-thirds of any α-zein sequence and must largely define the
112 proteins' spatial structure.
113 Due to the challenges of zein separation [33] most experimental characterizations have been carried
114 out on mixtures of α-zeins in alcoholic solution [34, 35] or in the solid state [36, 37] However, two
115 studies on Z19 zeins offer relevant context for the present work. In the first study, Forato *et al.* [38]
116 studied α-zein extracted from the maize variety BR451 which lacks Z22 expression. Gel experiments

117 showed solely Z19 zeins which were characterized without further purification or sequencing. From
118 solid state FTIR the authors reported α-helix segmentation consistent with ~20 residue repeats and a
119 secondary structure distribution of 46% α-helix, 22% β-sheet, 23% turns, and 13% undetermined
120 structure. An extended structure was inferred from ^{1}H NMR (70% deuterated EtOH) and SAXS (90%
121 EtOH, 3.6 mg zein /mL). The authors proposed a hairpin structure based on αz19D2 (UniProt ID:
122 Q946V7) containing helical segments, sheets, and turns with dimensions of 12 Å × 120 - 130 Å. In
123 the other Z19-focused study, Cabra *et al.* [33] purified a Z19 α-zein from waxy yellow dent corn
124 seeds to give a single MALDI-TOF-MS peak of 24.535 kDa. By sequencing the first 20 amino acids
125 (TIFPQCSQAPIASLLPPYLP) they identified the protein as zein M6 [23] (UniProtKB ID: P04702)
126 using BLAST. The CD secondary structure distribution was 40.0% α-helix, 19.5% β-sheet, 15.4%
127 coils, and 25.1% undetermined secondary structure for the isolated zein (0.08 - 0.25 mg/mL) in 70%
128 (v/v) methanol. CD tertiary structure prediction suggested that the α-helices folded into unspecified
129 compact structures. This differs from the extended structure assignments from SAXS and NMR
130 studies on other α-zein preparations in alcoholic solution [35, 38-40].
131 Interestingly, a contemporary BLAST search on the 20 first amino acids reported by Cabra *et al.*
132 returned three additional sequence hits from UniProtKB/SwissProt (updated 2022/10/30
133 https://blast.ncbi.nlm.nih.gov/) including cZ19C2, see S36 Fig. As the calculated molecular weight
134 of cZ19C2 (24.087 kDa) approximates the molecular weight from MALDI-TOF-MS, the
135 spectroscopic findings by Cabra *et al.* may possibly apply to the zein studied here.
136 Despite decades of research, no experimental atomic-level structures of zeins have been
137 published. Instead, the current understanding of zein molecular structure relies on low-resolution
138 experimental methods such as SAXS and a multitude of structural models. Early idealized
139 "consensus-sequence models" leveraged the α-zein repeat structure: A single ~20 residue consensus
140 sequence represents all repeats and is modelled as a single α-helix, except for terminal residues that

141 reside in turns connecting to neighbor helices. Different spatial arrangements of multiple identical
142 consensus helices give distinct models. The consensus sequence by Argos *et al.* mentioned above has
143 been used in several published models, including their own 1983 model [28]. The model is a hollow
144 cylinder formed by lateral association of nine anti-parallel consensus-helices linked end-to-end by
145 glutamine-repeat loops. The glutamine side chains enable hydrogen bonds in the axial direction
146 between zein monomers, while lateral interactions between helices stabilize the monomer and enable
147 lateral self-assembly of zein monomers. The 1993 model by Garratt *et al.* [32] for Z22 zeins also
148 assumes lateral association of antiparallel consensus α-helices, but on a two-dimensional hexagonal
149 grid. This grid model can grow by association of additional monomers while minimizing voids, to
150 emulate α-zein packing in protein bodies. Finally, the 1997 consensus model by Matsushima *et al.*
151 [40] consists of antiparallel repeat helices laterally stacked perpendicular to the helical axis, giving a
152 linear ribbon-like model with dimensions 130 Å × 12 Å × 30 Å. The model was used to explain both
153 limited aggregation in SAXS experiments (see below) and extensive aggregation assumed in protein
154 bodies.

155 After the consensus models, all-atom models incorporating complete sequence information
156 have become prevalent. A widely cited example is the 2005 model by Momany *et al.* [41] for the
157 cZ19C2 α-zein. The authors combined human intuition and molecular modelling to build the α-zein
158 monomer as a superhelix consisting of three axially stacked heterotrimeric coiled coils, respectively
159 containing the residue ranges (37-54, 55-77, 78-92), (93-112, 113-127, 128-146), and (147-166, 167-
160 185, 186-204). Each coiled coil was stabilized by nonpolar interactions with a molecule of the
161 pigment lutein. In MD simulations, the model was structurally stable in a mixture of 90% methanol
162 and 10% water, but degraded in pure water.

More recently, there has been an increase in automatically generated all-atom α-zein models from homology modelling (e.g., [42],[43],[44]) and *ab initio* modelling (e.g., [45],[46]). These studies moved beyond static single structures by subjecting the models to MD simulations.

Recent applications of MD simulations directly targeting α-zein structural understanding are exemplified by the research by Erickson *et al.* [47, 48]. They undertook [47] comparative MD studies of zein M6 (UniProtKB: P04702) constructed both as the superhelical model by Momany *et al.* [41] and as the extended hairpin molecular model by Forato *et al.* [38] in both 75% and 45% (v/v) ethanol/water. Only the hairpin model agreed with their CD-studies on commercial zein mixtures by giving increased β-sheet in more hydrophilic solvent. The study [48] focused on β-sheet propensities of peptide segments from the three most expressed α-zein clones αz19b1, αz19b3 (=cZ19C2) and αz22z1 found by Woo *et al*. The cZ19C2 peptides were from positions 9-19, 21-28, 75-90 and 148-157. The 21-28 peptide had the largest β-sheet propensity (~30 – 45%) whereas the other peptides had a maximum β-sheet content of ~10%.

In summary, the large body of modelling studies outlined above delineates three main groups of α-zein monomer models: consensus-sequence models, manually built atomistic models, and automatically generated atomistic models (including homology and ab initio models). While all models have importantly shaped the thinking about α-zein structure, it is useful to consider some of their possible limitations. A potential caveat for models in the first two groups is their reliance on human judgment for establishing a consensus sequence and/or assigning secondary and tertiary structure. Furthermore, in case of consensus models, their inherent neglect of variations in repeat sequences renders them unable to capture any conformational diversity that may arise from such heterogeneity. The automatically generated atomistic models also have limitations. The lack of templates for homology modelling of α-zeins can manifest as generation of widely different models from identical sequences. This is illustrated in Fig A in S2 Appendix for the zein examined herein.

187 Ab initio models do not have this issue, but they are susceptible to inadequate sampling or theoretical
188 treatment.

189 The present approach does not eliminate all of these concerns and could potentially introduce
190 new issues. Nonetheless, as the modelling route differs from previous research this study contributes
191 by offering a new perspective. The modelling herein draws inspiration from the large conformational
192 variability of α-zein under different environmental conditions indicated by numerous spectroscopic
193 studies. This implies that a single native structure, if such exists, is unlikely to represent the diverse
194 array of α-zein conformations in practical applications. Therefore, the main goal of the present study
195 is not to suggest a single native-like state of zein, but rather to contribute to charting the
196 conformational landscape of α-zein in various solution environments. The environments addressed
197 here are ethanol, water and their binary mixtures, as the latter are common solution systems [49, 50].
198 The findings may serve as a groundwork for a comprehensive mapping of α-zein conformations
199 across a broad range of conditions.

200 The deep-learning method AlphaFold2 [51] was used to build the α-zein structure.
201 AlphaFold2 is not a homology- or an *ab initio* method, but may be seen as an extension of both
202 approaches, expanding their capabilities through deep-learning integration of evolutionary, physical,
203 and geometric constraints of known proteins. It was hypothesized that this level of data integration
204 might partially compensate for the lack of homologous experimental structure templates for α-zein.
205 All-atom MD simulations in varied ethanol/water solvents were used to sample the conformations of
206 the AlphaFold2 model. In this context, a particular benefit of the model was its loose arrangement of
207 secondary structure elements. This allowed large-scale rearrangements during MD without bias
208 towards a densely packed tertiary structure as often found in homology- or *ab initio* α-zein models.
209 Illustrating the latter, *ab initio* models generated from the same sequence (Fig B in S2 Appendix)

210 show a dense packing that may hinder conformational sampling without obviously ensuring a more
211 realistic initial state.

212 Methods

213 AlphaFold2 Model

214 The protein sequence for the α-zein cZ19C2 [29] excluding signal peptide (residues 1-21) was
215 obtained via the accession number P06677 from UniProtKB [52]. The sequence was submitted via
216 the ColabFold [53] web-interface to generate AlphaFold2 [51] models. The best ranked model was
217 used for all further studies. The protonation state of the model was assigned with the GROMACS
218 pdb2gmx program, giving a net charge of +1. The assigned protonation states of ionizable groups are
219 given in Table A in S3 Appendix. To ensure adequate solvation, the principal molecular axes of the
220 model were aligned with the Cartesian axes in VMD [54] prior to solvation.

221 GROMACS All-atom Simulations

222 Solvent boxes containing of 0, 2, 23, 50, and 100 mol% ethanol in water were made by substituting
223 water for the required number of ethanol molecules in a pre-equilibrated pure ethanol box
224 (ethanol_T298.15.pdb.gz from https://virtualchemistry.org/ff/GAFF-ESP-2012.zip) using
225 GROMACS tools. The resulting simulation box compositions are shown in Table 1. Each box was
226 subjected to energy minimization followed by MD equilibration (100 ps NVT and 40.1 ns NPT at T
227 = 300 K) to give stable densities. The equilibrated solvent boxes were used to solvate the AlphaFold2
228 model to give cubic protein-solvent boxes with at least 15 Å of solvent on each side of the protein.
229 The +1 charge of the protein was neutralized by substitution of a randomly chosen solvent molecule

230 with a single chloride ion. Each protein-solvent box corresponding to a particular ethanol
231 concentration was subjected to the GROMACS simulation protocol outlined below.
232
233 **Table 1. System composition for GROMACS all-atom simulations.**

System	Ethanol molecules	Water molecules	Cl^-
100 mol% water	0	57033	1
2 mol% ethanol	985	54796	1
23 mol% ethanol	8788	30174	1
50 mol% ethanol	16050	16301	1
100 mol% ethanol	17728	0	1
100 mol% water (ff99SB*-ILDN)	0	57033	1
24 mol% water (ff99SB*-ILDN)	8919	28902	1

234

235 **GROMACS All-atom Simulation Protocol**

236 The GROMACS all-atom simulation protocol consisted of (1) energy minimization to relax steric
237 clashes using a target maximum force of 1000 kJ mol^{-1} nm^{-1}, (2) 100 ps MD equilibration with protein

238 heavy atom position restraints in the NVT ensemble at T = 300 K using the velocity rescaling (V-
239 rescale) thermostat, (3) 100 ps MD equilibration in the NPT ensemble (P = 1 bar, T = 300 K) with
240 protein heavy atom position restraints using the Parrinello-Rahman barostat pressure coupling, (4)
241 production MD run in the NPT ensemble (P = 1 bar, T = 300 K). The leap-frog algorithm was used
242 to integrate the equations of motion using a timestep of 2 fs and holonomic constraints were imposed
243 with LINCS [55]. Periodic boundary conditions and Particle Mesh Ewald long-range electrostatics
244 were used with cubic interpolation and a Fourier spacing of 0.16 nm. System densities after 100 ps
245 NPT equilibration (step 3) are listed in Table B in S3 Appendix.

246 **All-atom Simulations with the OPLS-AA/L Force Field**

247 For the 400 ns simulations, three simulation replicates labelled Seed 1, Seed 2, and Seed 3, were
248 carried out for each of the 5 ethanol concentrations giving a total of 15 × 400 ns MD simulations.
249 Each simulation replicate differed by initial atomic velocity randomization to improve sampling, The
250 simulations were carried out with GROMACS 2022 [56] using the OPLS-AA/L force field [57] for
251 protein and the SPC/E water model [58]. GAFF [59] parameters for ethanol were obtained from
252 virtualchemistry.org as ethanol.itp from the package https://virtualchemistry.org/ff/GAFF-ESP-
253 2012.zip.

254 **All-atom Simulations with the ff99SB*-ILDN Force Field**

255 To probe the behaviour of the α-zein model with a newer force field, four independently prepared
256 400 ns control simulations (2 seeds for 100 mol% water, and 2 seeds for 24 mol% ethanol) were
257 carried out in GROMACS with the TIP3P water model [60] and the ff99SB*-ILDN force field for
258 protein [61],[62]. The parameters were imported from
259 (https://github.com/bestlab/force_fields/blob/master/gromacs_format/README). The 100 mol%

water (seed 1) simulation was extended to 800 ns. Time series for these supplementary simulations are given in S27 Fig, S28 Fig, S29 Fig, S30 Fig, S31 Fig, S32 Fig, S33 Fig, S34 Fig.

Additional Equilibration of a Globular α-zein Model

Based on its globularity and structural stability in a long MD trajectory, the 1 μs all-atom GROMACS simulation endpoint in water (seed 3) was chosen as a starting point for multiple zein simulations. The structure was further equilibrated in a water box in Desmond [63, 64] for 500 ns in the NPT ensemble at 300 K and 1 bar with default simulation protocol settings, using the OPLS3 [65] force field for protein and the TIP3P water model.

All-atom Simulations of Multiple α-zeins

The Maestro [66] graphical user interface and Desmond was used for preparing and executing all-atom simulation of the multi-zein system. In Maestro, a non-interacting configuration of six copies of the globular α-zein was constructed by placing the monomer structures at positions corresponding to midpoints of the faces of a sufficiently sized virtual cube, see Fig 16A. This structure was solvated in a cubic box, giving a total of 489201 atoms (156161 water molecules) followed by a 500 ns MD simulation in Desmond with the standard pre-simulation relaxation protocol and NPT simulation settings. The OPLS3 force field was used for the protein and the TIP3P [60] model for water.

Coarse-grained Simulations of Multiple α-zeins

Coarse-grained simulations on multiple copies of the globular α-zein conformation were carried out in GROMACS with the Martini 2.2 force field [67]. The all-atom α-zein monomer was converted to coarse-grained coordinates and topology using the Martini script Martinize.py. The -dssp option was used with the script, providing secondary structure assignment using DSSP-based [68] analysis of backbone dihedrals and hydrogen bonds. The structure before and after coarse-graining is shown in

282 Fig 17A and B, respectively. To build a crowded model system mimicking a pre-aggregation state,
283 73 copies of the coarse-grained α-zein monomer were randomly inserted into a 250 × 250 × 250 $Å^3$
284 box, with a minimum van der Waals distance of 4 Å between proteins. The box was solvated with a
285 pre-equilibrated 160 × 160 × 160 $Å^3$ box of 35937 coarse-grained Martini water molecules, using a
286 minimum van der Waals distance of 2.1 Å between waters. The net positive charge of +73 was
287 neutralized by automatically replacing 73 randomly selected waters with chloride ions. After energy
288 minimization, the system in Fig 17C was subjected to three NPT (P = 1 bar, T = 303 K) simulations
289 differing by initial velocity randomization, that were each run for 375 ns (15000000 steps of 25 fs).
290 The simulation endpoints are shown in Fig 17D – I.

291 **Interactions with Glycidyl Methacrylate**

292 The 1 μs all-atom GROMACS simulation endpoint in water (seed 3) was used to probe interactions
293 with glycidyl methacrylate (GMA). GMA was built and minimized in Maestro, and a spherical
294 distribution of 48 GMA molecules was set up around the α-zein with 47 Å between the centroids of
295 each GMA molecule and the α-zein centroid. This ensured a minimum initial distance of 7 Å between
296 any GMA molecule and the α-zein surface. This configuration was solvated in a cubic box of 61370
297 TIP3P molecules giving a buffer of 10 Å on all sides of the protein-GMA configuration. The system
298 was subjected to the default multistage minimization and NPT simulation protocol in Desmond for
299 257 ns.

300 **MD Analysis and Visualization**

301 Prior to analysis, the GROMACS MD trajectories were downsampled to give a time interval of 1 ns
302 between each frame. Molecular figures were made with VMD [54] and Maestro. VMD was used for
303 MD trajectory analysis, including calculation of backbone root mean square deviation (RMSD),

solve-accessible surface area (SASA), radius of gyration (R_g), and secondary structure based on STRIDE [69].

Results and discussion

Sequence

Although all sequences in Table A in S1 Appendix are candidates for investigation, here the practical choice was made to study the Z19 α-zein cZ19C2 also known as αz19b3 (UniProtKB ID: P06677). This decision was motivated by several factors, including the expected prevalence of this zein in actual preparations. As noted above, it ranked among the top three zeins in terms of expression level in common maize lines. It is also one of only two zeins currently registered in UniProtKB with protein-level evidence, see Table C in S1 Appendix. Furthermore, the relatedness with other zeins, as depicted in Fig 1, suggests that insights from cZ19C2 could potentially apply to other zeins as well. Moreover, the sequence has been cited in several studies, e.g., [70],[71],[72],[73], and was used for the atomistic model by Momany *et al.* [41] mentioned above. Finally, cZ19C2 is homologous or possibly identical with the purified α-zein in Cabra *et al.*'s study [33] which makes it a relevant reference point and enables comparison with previous research. In Fig 3A the sequence bas been manually aligned with the consensus sequence from Geraghty *et al.* [25], yielding 12 segments each containing up to 20 residues.

Initial Scan for Self-assembly Domains

Initially, an attempt was made to predict regions with self-assembly potential within the mature cZ19C2 protein sequence. This effort was motivated by a desire to increase the transparency of the modelling process. Furthermore, it was of interest to see if coiled-coil interactions of the type central to the cZ19C2 model by Momany *et al.* [41] could be detected with contemporary prediction

327 algorithms. Waggawagga [74] and Deepcoil2 [75] were used to scan the sequence for single-α helix
328 (SAH) domains, canonical heptad-repeat coiled-coil domains, and non-canonical coiled coil domains.
329 However, none of these domains were identified.
330
331 **AlphaFold2 Model**
332 The AlphaFold2 model is shown in Fig 3B. Its loosely arranged structure superficially resembles
333 models from published experimental α-zein characterizations in alcohol/water mixtures, such as
334 Forato *et al*.'s [38] segmented multi-helix model and Tatham *et al*.'s elongated few-helix model [39].
335 Overall low predicted lDDT [76] scores (S1 Fig) imply high model uncertainty across the sequence,
336 reflecting the absence of α-zein-like structures in the AlphaFold2 training data. The model lacks β-
337 sheets, and has an α-helical content of 68% distributed among seven antiparallel and mostly loosely
338 packed α-helical segments (denoted helix I to VII) of varying length connected by turns or loops. The
339 α-helical segments are initiated by Pro-Ala or, in one case, Arg-Ala. Some conspicuous features of
340 the model include the two central long α-helices (helix II and III), and the bundle of three relatively
341 tightly packed α-helices near the C-terminal (helix IV, V, and VI) which have slightly higher
342 confidence scores, see S1 Fig. The α-helices II and III appear unusually long given their lack of
343 stabilizing interactions with adjacent structure [77],[78] . Interestingly, the *ab initio* method C-
344 QUARK also predicted a long helix for a subset (S56 – A86) of the helix II residues, see Fig B in S2
345 Appendix.
346
347 **Fig 3: Consensus sequence alignment and AlphaFold2 model.** A: Consensus sequence alignment.
348 Top row (grey background) is the consensus repeat sequence by Geraghty *et al.* [25]. The other rows
349 contain the mature protein sequence of cZ19C2 (UniProtKB ID: P06677) manually aligned with the
350 consensus repeat sequence in segments of up to 20 amino acids. Green background indicates amino

351 acids matching the consensus sequence. Black background indicates gaps. B: AlphaFold2 model.
352 Residues corresponding to the first amino acid of each row in the alignment (A) are indicated by
353 black labels. The α-helices I – VII are annotated in purple. Gln and Asn are shown in green and orange
354 stick representation, respectively. The protein backbone is shown as ribbons coloured by secondary
355 structure. α-carbons are shown as spheres coloured by sequence position (red: N-terminal, blue: C-
356 terminal).
357
358 A detailed description of the AlphaFold2 model follows. Starting from the C-terminal, the
359 unstructured tail residues Q210 – F219 are preceded by the short helix VII (P205 – Y209). The
360 preceding residues N144 to D204 span the bundle of the three antiparallel α-helices, helix IV (P145
361 – A163), helix V (R165 – A182), and helix VI (P205 – Y209) that are laterally associated with each
362 other, forming a mainly hydrophobic core. Interestingly, the sequence of each helix corresponds to
363 segments 9, 10, and 11 of the zein sequence in the alignment to the consensus sequence by Geraghty
364 *et al.* [25], see Fig 3A. Thus Gln-Gln repeats are inside helices and not in turns linking the helices, in
365 contrast to the consensus model by Argos *et al.* [28]. Instead, N144, N164 and N184 reside in the
366 turns and expose their side chains, possibly enabling hydrogen bonding. The C-terminal α-helical
367 bundle is connected to the long helix III (P110-L143) through a hairpin turn containing N144.
368 Comparison of helix III to the consensus sequence alignment in Fig 3A shows that it spans segments
369 7 and 8. It could have been reasonable to anticipate a break in helix III at the position corresponding
370 to the start of segment 8, in light of the one-to-one segment-to-α-helix mapping for the C-terminal
371 helices IV, V, and VI. However, in contrast to the latter, gaps were used in segments 7 and 8 for
372 alignment with the consensus sequence. The gaps imply that segment 8 lacks residues corresponding
373 to positions 1-3 in the consensus sequence. Since these consensus sequence positions in segments 9
374 – 11 coincide with the start of helices IV - VI, their lack in segment 8 may explain why AlphaFold2

375 instead merges segments 7 and 8 into the single long α-helix III. Indeed, all α-helices in the
376 AlphaFold2 model are either initiated by Pro-Ala or Arg-Ala, whereas the first two amino acids in
377 segment 8 are Ala-Tyr. Helix III connects via the disordered residues Q98-N109 to the other long
378 helix, helix II (P57-Q97). Unlike helix III, the terminal residues of helix II do not correspond to
379 segment start and stops in the consensus sequence alignment (Fig 3A). Instead, helix II begins in the
380 middle of segment 4, spans segment 5 and terminates in the middle of segment 6. Except for segment
381 6, these segments share little homology with the consensus sequence. This reflects a gradual loss of
382 recognizable repeat structure towards the α-zein N-terminus. Helix II is linked via the unstructured
383 residues S44-S56 to the short helix I (L32-A43) which again is linked to the N-terminal tail residues
384 T1 - A31 which exhibit less ordered structure consisting of coils and turns. This region contains the
385 only two cysteines, C6 and C27, which are in close proximity with a S-S distance of 3.4 Å.

386 **400 ns all-atom MD Simulations**

387 The structural snapshots in Fig 4 - 8 represent the behavior of the AlphaFold2 model during the fifteen
388 400 ns MD simulations. In each figure, each row corresponds to a simulation replicate differing by
389 the seed (seed 1, seed 2, seed 3) used for randomizing initial atom velocities. The first column (subfigs
390 A, D, E) shows the initial AlphaFold2 structure as reference. The middle column (subfigs B, E, H)
391 shows 10 superimposed snapshots from the last 100 ns of each simulation, to illustrate conformational
392 variation. The last column (subfigs C, F, I) displays the final structural snapshots at 400 ns.

393

394 **Fig 4. Snapshots from 400 ns MD simulations in 100 mol% water.** A, D, G: AlphaFold2 structure
395 before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation.
396 C, F, I: Snapshots at 400 ns.

397

398 **Fig 5. Snapshots from 400 ns MD simulations in 2 mol% ethanol.** A, D, G: AlphaFold2 structure
399 before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation.
400 C, F, I: Snapshots at 400 ns.
401
402 **Fig 6. Snapshots from 400 ns MD simulations in 23 mol% ethanol.** A, D, G: AlphaFold2 structure
403 before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation.
404 C, F, I: Snapshots at 400 ns.
405
406 **Fig 7. Snapshots from 400 ns MD simulations in 50 mol% ethanol.** A, D, G: AlphaFold2 structure
407 before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation.
408 C, F, I: Snapshots at 400 ns.
409
410 **Fig 8. Snapshots from 400 ns MD simulations in 100 mol% ethanol.** A, D, G: AlphaFold2
411 structure before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each
412 400 ns simulation. C, F, I: Snapshots at 400 ns.
413
414 The associated time series for backbone root mean square deviation from the AlphaFold2 structure
415 (RMSD), radius of gyration (R_g) and protein solvent accessible surface area (SASA) are included in
416 S3 Fig. Average values for these time series are shown in Fig 9.
417
418 **Fig 9. Time series averages for 400 ns MD simulations.** A: Backbone RMSD relative to the
419 AlphaFold2 structure. B: Radius of gyration (*R*g). C: Solvent accessible surface area (SASA). D:
420 Total helicity (sum of α- and 3_{10}-helicity). Values for each ethanol concentration were averaged over
421 the last 100 ns of each simulation time series and then over the three simulations (seed 1 - 3) per

422 concentration. Error bars indicate standard deviation across seeds seed 1 - 3. For the associated time
423 series see S3 Fig.
424
425 Simulations differing only by the randomization seed used for initial velocities gave distinct
426 structures at 400 ns, as evident from the last columns of Fig 4 - 8. This agrees with the expected
427 sensitivity to perturbations in the loosely packed AlphaFold2 structure. The two long α-helices (helix
428 II and III) did not maintain their initial forms in any of the simulations, but rearranged into various
429 mixtures of secondary structure elements. In both water and in 2 mol% ethanol, the model rapidly
430 collapsed into compact globular forms that were largely retained for the remainder of the simulations.
431 This is reflected in the fast convergence to relatively low average values in the RMSD, R_g and SASA
432 time series (S3 Fig) and by the similarity of superimposed MD snapshots within each simulation (Fig
433 4 and S3 Fig B, E, and H). The simulation endpoints (Fig 4 and Fig 5 C, F, and I) show the distinct
434 compact structures. With increasing ethanol concentration, a wider range of α-zein conformations
435 were sampled. The increased sampling was subtle at 2 mol% ethanol, but became prominent at 23
436 mol% ethanol and above. This is seen from the larger fluctuations and higher average values in the
437 RMSD, Rg, and SASA time series (S3 Fig) and by less congruent superimposed MD snapshots at
438 higher ethanol concentrations, see Fig 4 – Fig 8, subfigs B, E, and H. Relative to water simulations,
439 simulations with added ethanol showed structural expansion. This is evident from the time series
440 averages of radius of gyration (Rg) and SASA, calculated across the three simulation seeds for each
441 ethanol concentration, as depicted in Fig 9. For instance, the average R_g almost doubled from water
442 (18.9 Å) to ethanol (36.1 Å). Also, the average SASA in water (12486 Å^2) and ethanol (19857 Å^2)
443 implies a ~59 % SASA increase in ethanol. At high ethanol concentration, these changes were
444 accompanied by disruption of interactions between secondary structure elements (i.e., tertiary
445 structure). This disruption was particularly evident for the elongated structures observed at 50 mol%

446 ethanol (seed 2) and 100 mol% ethanol (seed 1), see Fig 7F and Fig 8C. At 50 mol% ethanol, the
447 three 400 ns simulation replicates sampled expanded α-zein conformations, as indicated by the time
448 series for R_g and SASA (S3 Fig), with one replicate (seed 2) showing pronounced elongation, see Fig
449 7F. The onset and extent of expansion varied across the replicates, occurring at the simulation start
450 for seed 2 but only after ~140 ns for seed 1, underlining the stochastic nature of MD simulations. This
451 highlights the significance of using multiple seeds and conducting long simulations to explore
452 conformational changes. At 100 mol% ethanol, the time series (S3 Fig) and structural snapshots
453 resemble the results at 50 mol%. Thus, there was once again an increased exploration of expanded
454 conformations, as reflected by disordered superposition of snapshots. Also, one simulation (seed 1)
455 produced very elongated conformations. The SASA averaged over seeds (19857 $Å^2$) was similar to
456 the 50 mol% ethanol simulation (19599 $Å^2$), while the average radius of gyration (36.1 Å) was
457 significantly larger compared to the 50 mol% simulation (31.8 Å). Nevertheless, the general
458 resemblances between the simulations at 100 mol% ethanol and 50 mol% ethanol may indicate that
459 the impact of ethanol on conformational sampling has reached a plateau at around 50 mol%. A
460 possible concern arising from the pronounced structural expansion at high ethanol concentration (≥
461 50 mol%) is that the very elongated zein conformations in some simulations could lead to non-
462 physical interactions between α-zein and its own image in neighboring simulation cells. Thus, these
463 limiting cases should be interpreted with some care. Nevertheless, the elongated conformations are
464 consistent with previous SAXS experiments, as shown in a later section.

465 Identification of least-mobile protein substructures using MDLovoFit [79] offers an
466 alternative view of how solvent composition influences the α-zein model. MDLovoFit maps rigid and
467 mobile regions in a MD trajectory by aligning progressively larger fractions of Cα-atoms (denoted
468 by ϕ) across the trajectory and computing the RMSD of the aligned atoms. The method was applied

469 to the final 100 ns of each 400 ns MD trajectory. The resulting plot in Fig 10 depicts RMSD for Cα-
470 atom superposition versus the fraction of the total number of Cα atoms (ϕ).
471
472 **Fig 10. Least-mobile substructures in the 400 ns MD simulations.** The graphs for each simulation
473 show the lowest RMSD of Cα-atoms after superpositions of all subsets of size ϕ (fraction of the total
474 number of Cα atoms). The analysis was carried out on the last 100 ns of each simulation with
475 MDLovoFit [79].
476
477 Each fraction corresponds to the lowest attainable RMSD for all possible subsets of size ϕ. The plots
478 show that ~30% of the Cα-atoms (ϕ = 0.3) can be superimposed within a RMSD of 1 Å in the
479 trajectories of all 100 mol% water simulations, all 2 mol% ethanol simulations, and two 23 mol%
480 ethanol simulations (seed 1 and 3). In contrast, only around 10% of the Cα-atoms (ϕ = 0.1) could be
481 superimposed within 1 Å for one 23 mol% ethanol simulation (seed 2), two 50 mol% ethanol
482 simulations (seed 1 and 3), and one 100 mol% ethanol simulation (seed 3). The extensive
483 conformational variation in the two remaining simulations, 100 mol% ethanol (seed 1) and 50 mol%
484 ethanol (seed 2), implied that only a negligible fraction of Cα atoms could be superimposed within
485 ~1 Å RMSD.

486 **Helicity**

487 An averaged representation of helicity in all 15 × 400 ns simulations is shown in Fig 11. For each
488 residue, the sum of α-helicity and 3_{10}-helicity averaged over the last 100 ns of all 15 × 400 ns
489 simulations is plotted with blue bars. The initial AlphaFold2 α-helicity is shown as orange bars.
490 Individual plots for each ethanol concentration are provided in S9 Fig, S10 Fig, S11 Fig, S12 Fig,
491 S13 Fig.

492

493 **Fig 11. Helicity per residue.** Orange: α-helicity in the initial AlphaFold2 model. Blue: Helicity (sum
494 of 3_{10} and α-helicity) of the last 100 ns of each 400 ns simulation averaged over all 15 × 400 ns MD
495 simulations.

496

497 The plot in Fig 11 shows a substantial retention of helical structure from the AlphaFold2 model.
498 Moreover, some simulations display helicity in regions that were initially non-helical in AlphaFold2
499 model (e.g., P20 - C27). Furthermore, specific residues never became helical. Notable examples are
500 N144, N164, and N184, which are located in the turns connecting helix III and IV, helix IV and V,
501 and helix V and VI, respectively. Non-helicity for these residues is readily explained for simulations
502 with high water content. In those cases, stabilization of helix-helix interactions in the C-terminal α-
503 helical-bundle restrained N144, N164, and N184 to reside near their initial positions in the turns
504 connecting the helices. Interestingly, these residues remained non-helical even at high ethanol
505 concentration where helix-helix interactions were destabilized, permitting large-scale rearrangement
506 of the constituting helical-hairpins. This is illustrated in Fig 12 for the 100 mol% ethanol 400 ns
507 simulation (seed 1). The initial compact packing of the helices in Fig 12A is clearly disrupted at 89
508 ns (Fig 12B) and at 123 ns the turn between helix IV and V is fully extended (Fig 12C). Nevertheless,
509 N164 did not become helical, but functioned as a pivotal point connecting the two helices. Thus at
510 163 ns, there was a partial reconstitution of the turn, see Fig 12D.

511

512 **Fig 12. Large-scale motion of helices in C-terminal bundle during a 400 ns ethanol simulation**
513 **(seed 1).** A: Initial AlphaFold2 structure. B: Disruption of the α-helical bundle. C: Full extension of
514 the helix-N164-helix hairpin. D: Partial reconstitution of the hairpin.

515 **Influence of Ethanol Concentration on Helicity**

516 Principal components analysis (PCA) applied to the helicity profiles from the simulations did not
517 produce any clustering of similar ethanol concentrations, as demonstrated in S14 Fig. Thus, ethanol
518 concentration did not seem to systematically influence the distribution of helicity along the α-zein
519 sequence. Nevertheless, the average time series plot of helicity in Fig 9D shows a slight increase (up
520 to ~7%) in average helicity at higher ethanol concentrations, albeit with large standard deviations.
521 Taken together, these findings suggest a minor influence of ethanol concentration on the secondary
522 structure in the simulations. This aligns with previous reports of constant α-helicity in the Z19 zein
523 from maize BR451 across ethanol concentrations ranging from 28% to 70% [35]. Also, a recent
524 circular dichroism (CD) study found no notable variations in the secondary structure of FZ3121 zein
525 with ethanol concentrations from 60% to 90% [50].

526 **β-sheet**

527 Most experimental studies on zeins have reported some β-sheet content, even for highly purified Z19
528 zein [33]. In contrast, most published models of α-zein monomers lack β-structure. Mechanisms for
529 α-helix to β-sheet transitions have been proposed for increases in solvent hydrophilicity [11] and
530 mechanical treatment [80]. However, the extent to which β-sheet structure is inherently present in
531 isolated α-zein monomers remains uncertain. MD simulations on isolated zeins, such as in this study,
532 may potentially offer insights. Hence an examination of β-structure in the simulations follows below,
533 which should be viewed in light of the potential limitations of the approach such as limited sampling
534 and force field bias.

535 In all 400 ns GROMACS simulations β-sheet structure in any of the forms assigned by
536 STRIDE (i.e., "Extended", or "Isolated bridge") was sparse and transient. This can be appreciated by
537 considering the average content during the last 100 ns of each simulation (S4 Fig, S5 Fig, S6 Fig, S7

538 Fig, S8 Fig) For the water simulations, all seeds lacked extended structure while seed 1 had the highest
539 isolated bridge content of ~0.9%. For all seeds in the 2 mol% ethanol simulation series, extended
540 structure lacked, whereas a maximum of 0.2% isolated bridge structure occurred for seed 1. Similar
541 trends were found for the 23 mol% ethanol simulations, where extended structure was absent, but one
542 simulation (seed 1) had 0.9% isolated bridge. The 50 mol% series lacked extended structure and had
543 only transient populations of isolated bridge structure. Finally, for the 100 mol% ethanol series, seed
544 1 and 2 had 0.2% and 0.8% extended structure and 1.3% and 0.4 % isolated bridge, respectively,
545 whereas seed 3 lacked β-structure. Snapshots of the β-structure in seed 1 and 2 are shown for both
546 simulations in S15 Fig. For seed 1 an interaction involves β-strands formed by residues C6 – Q8 and
547 I46 – L48. Similarly, for seed 2 there is an interaction between β-strands formed by residues G215 -
548 G216 and Q211 - H212. These β-structures occur near the N- and C-termini, respectively, within
549 sequence regions that were assigned coil or loop secondary structure in the AlphaFold2 model. While
550 the 400 ns time frame apparently allows for the transition to β-structure from these relatively
551 unstructured regions, transitioning from an α-helical structure might require longer simulations. Thus,
552 given the potentially long timescale for transition to β-sheet structure from the initial α-helix dominant
553 state of the AlphaFold2 model it was of particular interest to examine the last 100 ns of the three
554 extended MD trajectories. The extension of the 100 mol% water simulation (seed 3) to 1 μs did not
555 introduce any additional β-structure, see S21 Fig, top panel. In case of the 2 mol% ethanol simulation
556 (seed 2) the extension to 2 μs caused the extended β-sheet content to increase from 0.0% to 0.1%
557 while the isolated bridge content increased from 0.0 to 1.3% (S21 Fig, middle panel). This β-structure
558 is described later in the section on extended simulations. For the 23 mol% ethanol simulation (seed
559 2) the lack of extended structure persisted until 2 μs, while there was a marginal increase in isolated
560 bridge content relative to the first 400 ns (from 0.0 to 0.3%), see S21 Fig, bottom panel. The small

561 set of additional simulations carried out with the newer ff99SB*-ILDN force field [62] also showed
562 little β-structure content, see S29 Fig.

563 A brief examination of the possible induction of β-structure due to zein-zein interactions was
564 provided by the all-atom MD simulation of 6 interacting copies of identical α-zein monomers.
565 Although these monomers associated early in the simulation, no significant β-structure developed
566 during the 500 ns trajectory, as evident from the time series plots in S35 Fig. None of the monomers
567 contained extended structure, while small amounts (up to 0.7%) of isolated bridge structures were
568 observed for some monomers.

569 Thus, the MD simulations herein generally contain less than 1 percent β-structure. Several
570 possible factors could account for this scarcity. Methodological factors might include force field
571 biases or insufficient sampling, as the transition to β-structure may occur on longer timescales.
572 Furthermore, specific perturbations in e.g., temperature or structure, which have not been taken into
573 account, might also be important. On a more speculative note, one could consider that if α-zein could
574 be isolated to the same extent as in the simulations, experimental findings might indicate a similarly
575 low presence of β-structure.

576 **Comparison with Previous SAXS Experiments**

577 Small-angle X-ray scattering (SAXS) studies have provided several well-known models of Z19
578 and/or Z22 zein in solution. Tatham *et al.* [39] studied mixed Z19/Z22 zein from commercial Sigma
579 zein in 70% (v/v) methanol/water at a concentration of 8 mg/mL and modelled the solution structure
580 as a 150 Å × 14 Å rigid rod comprised by a single α-helix without consensus-repeat segmentation.
581 Matsushima *et al.* [40] studied a 2:1 mixture of Z19 and Z22 zeins extracted from Royal Dent 120
582 maize at 2 to 40 mg/mL zein concentrations in 70% (v/v) ethanol/water with added 2-mercaptoethanol
583 to disrupt disulfide-stabilized dimers. They found that SAXS agreed with a tetramer of their ribbon

584 like-model with dimensions 130 Å × 42 Å × 30 Å. Forato *et al.* [38] studied Z19 zein from the Z22-
585 supressed BR451 maize strain in 90% ethanol and 10% water at a zein concentration of 3.6 mg/mL.
586 They modelled the SAXS data as an extended 130 Å × 12 Å structure resembling a folded hairpin
587 with mixed helix, turn and sheet content. Despite significant variation among these SAXS structures,
588 the studies concur that α-zein assumes extended conformations in water/alcohol mixtures with a high
589 alcoholic content.

590 Since only Forato *et al.* studied an isolated Z19 fraction, that work was compared with the
591 simulations herein. The SAXS data was provided by the authors and used both as received and after
592 extrapolation of the low q range to $q = 0$ using SAXSMoW 2.0 [81] for comparison with SAXS
593 predictions from MD simulations using FoXS [82]. Given that the authors had proposed an extended
594 model, here an MD trajectory with notably extended conformations (100 mol% ethanol, 400 ns, seed
595 1) was first used for comparison. As shown in Fig 13A, the initial AlphaFold2 structure gave a poor
596 fit to the experimental SAXS data. However, as the MD simulation progressed, leading to an
597 extension of the protein, the fit improved as shown in Fig 13 A, C, E, G, I. The same behaviour was
598 observed in the fits against experimental SAXS extrapolated to $q = 0$, shown in Figs 13 B, D, F, H, J.
599 Moreover, extended snapshots from the 50 mol% ethanol simulation (seed 2) gave fits of similar good
600 quality, see S18 Fig. This is relevant, as the 50 mol% ethanol concentration corresponds to 78 % (v/v)
601 and thus approaches the experimental 90% concentration. In contrast, SAXS predicted for compact
602 structures from water-rich simulations deviated markedly from the experimental curves and gave poor
603 fits, see S17 Fig.

604

605 **Fig 13. SAXS predictions from MD simulations and fitting to published experimental data.**
606 Predictions from MD frames at $t = 0$ ns, 100 ns, 200 ns, 300 ns, 400 ns of the (seed 1) 400 ns 100%
607 ethanol simulation (red) and fit to experimental SAXS (black) from [38]. (A, C, E, G, I): MD

predictions against experimental SAXS. (B, D, F, H, J): MD predictions against experimental SAXS extrapolated to $q = 0$ (blue) from the linear low q region of the Guinier plot using SAXSMoW2 [81]

Extended Simulations

Extended α-zein MD trajectories were analyzed in their totality, i.e., including the first 400 ns already discussed above, to identify and present key structural changes along the full simulation trajectory. To this end, MD frames were superimposed on less dynamic substructures within each trajectory. They were the α-helical bundle (N144 - D204) in case of the water 1 μs (seed 3) simulation and the 2 mol% ethanol 2 μs (seed 2) simulation. For the 23 mol% ethanol 2 μs (seed 2) simulation, helix II and III were used for superimposition.

Water 1 μs Simulation

Both long helices II and III underwent almost simultaneous destabilization early in the simulation. They bent near the middle positions L74 and L124, which coincided with a decrease in α-helicity for residues L74 - L124. These residues encompass approximately the upper portion of the AlphaFold2 helix II-loop-helix III motif, with the model's orientation at 0 ns in Fig 14A. The grey arrow at 0 ns indicates the initial movement of these residues, which was part of a progression of structural condensation and association with the C-terminal α-helical bundle (N144 - D204). At 13 ns there was increased association among the C-terminal helices IV, V, VI, and VII. Furthermore, reduced distance between helix IV of this bundle and the lower helical segment of the adjacent helix III (Q133 - L143) resulted in their partial association. Also, the residues of helix II (P57 - Q97) had moved closer to the growing C-terminal helix assembly. This segment maintained considerable α-helicity and now exhibited a notable curvature, partly due to a kink involving residues L74, Q75, Q76. The curved nature of this segment gave the visual impression that it loosely encircled the upper portion of the C-

631 terminal helical assembly, encompassing the lower half of helix III, as well as helices IV, V, VI, and
632 VII, extending approximately halfway around them. Additionally, the first residue of helix II (P57)
633 had approached the first residue of helix V (R165), establishing interactions between these regions
634 enduring throughout the remainder of the simulation. The "stickiness" of this interaction is
635 characteristic of the type of intramolecular interactions observed in the study. From 13 ns to 1000 ns,
636 the overall topology of the structural elements described thus far remained mostly unchanged.
637 However, there was a continued process of condensation yielding denser packing within the lower
638 region of the protein, as evident in Fig 14A. Alongside this condensation, notable rearrangements
639 occurred in other regions of the structure. Particularly, the C-terminal segment (T1 - Q55) and the
640 first residues of helix II (P110, A111) made contact at the top of the C-terminal bundle. This
641 interaction is evident from the proximity of A9 and A111 in Fig 14A at 400 ns, and it persisted
642 throughout the simulation. Further reorganization in the less structured upper part of the structure
643 then took place, as reflected by the RMSD increase at around ~500 ns (S20 Fig A) The rearrangement
644 involved condensation, as evident from comparing the upper part of the structure in the snapshots at
645 400 ns and 1000 ns in Fig 14 A. This is also seen from reductions in radius of gyration and solvent-
646 accessible surface area (SASA) in S20 Fig B and C. The simulation endpoint at 1000 ns had a globular
647 conformation, characterized by a lower region containing mostly helices and a smaller and less
648 compact upper portion with mostly coil/turn elements. Helicity was maintained in residues Q63-N83
649 of helix II, except for Q73 which acted as a hinge point.
650
651 **Fig 14. Snapshots from extended MD simulations.** Grey arrows indicate approximate directions of
652 rearrangement. Cα atoms are shown as spheres, coloured by residue position (red: N-terminal, blue:
653 C-terminal). Labelled residues are discussed in the main text. Rotation has been used to highlight
654 features in select snapshots, including the identical initial structures. In the case of 2 mol% ethanol

655 (B), instead of the 2 μs endpoint the structure at 1971 ns is presented to provide a clearer view of the
656 β-structure.

657 **2 mol% Ethanol 2 μs Simulation**

658 The structure rapidly converged to a compact conformation which was largely conserved for the
659 remainder of the simulation. Thus, at 25 ns, the overall structure of the simulation endpoint at 2 μs
660 was already mostly formed, as illustrated by the snapshots in Fig 14B. The state at 25 ns was reached
661 by successive bending of each of the long helices III and II at hinge points roughly at the middle of
662 each helix. First, helix III bent near residue T126, causing this helix to fold back on itself, forming a
663 helix-turn-helix motif. This movement is indicated by a grey arrow at 0 ns in Fig 14B. The movement
664 brought the upper half of the helix II-coil-helix III helix hairpin into lateral contact with the C-terminal
665 helical bundle (helix IV - VI) yielding the state shown at 8 ns. This was followed by bending of helix
666 II around L74, as indicated by a grey arrow at 17 ns. This motion placed the substructure containing
667 residues T1 - L74 on top of the previously compacted substructure containing the remaining residues
668 Q75 - F219. The resulting configuration at 25 ns thus appears to divide the protein into an upper part
669 (T1 - L74) and a lower part (Q75 - F219). The most notable change in the remainder of the simulation
670 was reorganization of the upper part, as noticed by comparing the snapshots at 25 ns and 1971 ns.
671 The reorganization can loosely be described as a continued condensation within T1 - L74, along with
672 their packing against residues in the lower domain. This was accompanied by a notable loss of α-
673 helicity (from ~70% to ~30%) over the first 1000 ns of the simulation, followed by more stable α-
674 helicity until the endpoint at 2000 ns, see S21 Fig, middle panel. The loss of α-helicity occurred
675 mainly in the upper part of the protein. During the upper part reorganization, a small amount of β-
676 sheet was formed by the backbone hydrogen bond pairs A9/I26 and I11/A24, as shown in the snapshot
677 at 1971 ns and in more detail in S25 Fig.

678 **23 mol% Ethanol 2 μs Simulation**

679 Consistent with the general trends in simulations with significant ethanol concentration, the α-zein
680 rapidly explored of a wide range of conformations, especially in the early stages. This is evident from
681 the pronounced RMSD fluctuations during the first ~500 ns, see S20 Fig G. As a result, any two
682 sparsely sampled MD snapshots were connected by a complex sequence of rearrangements that
683 escape concise summary. However, the net rearrangement from 0 ns to 400 ns involved conversion
684 of the middle residues (L119 - Q130) of helix III to turn, while helix II largely maintained helicity.
685 An ensuing buckling collapse of the helix II-coil-helix III hairpin caused a downward motion of its
686 upper part. Consequently, at 400 ns the apex residues (P110 – P120) of the hairpin appeared lodged
687 between the N-terminal residues (T1 - N89) to the left and the C-terminal residues (S122 - F219) to
688 the right, see Fig 14C. This unstable structure reconfigured substantially into the form at 800 ns,
689 which maintained its overall shape until the simulation endpoint at 2 μs. The endpoint had contiguous
690 α-helicity for residues L61-S91, thus preserving most α-helicity of the long helix II of the AlphaFold2
691 starting point. In contrast, helix III lost most of its α-helicity when Y113 - Q130 changed to coil. This
692 enabled a denser packing of the involved residues, as reflected in the overall compact structure at 2
693 μs shown in Fig 14C. An interesting aspect of this simulation is its apparent convergence to a
694 relatively compact structure despite significant ethanol concentration (corresponding to ~50 vol%
695 ethanol).

696 **Extended Simulations Summary**

697 The extended simulations corroborated certain trends observed in the 400 ns simulations, notably the
698 models' tendency to condense in water and expand with increasing ethanol concentration, along with
699 the persistence of specific intramolecular interactions once they were formed. This was evident in
700 both the water and the 2 mol% ethanol simulations from the similarity between their 400 ns simulation

701 endpoints and the structures at the end of the extended simulations, as seen in S2 Fig. Although these
702 persistent interactions may reflect the aggregation propensity of actual α-zeins, they pose a barrier to
703 conformational sampling. Consequently, enhanced MD sampling studies would be an obvious
704 extension of the present work.

705 Notwithstanding major structural differences, both the water and 2 mol% extended
706 simulations gave structures with a more condensed lower region and a less dense upper region. The
707 more compact lower region included the C-terminal α-helical bundle onto which other parts of the
708 protein associated during the simulations. This helical bundle was relatively conserved from the
709 AlphaFold2 model. Additionally, as discussed above, this part of the structure mapped well onto the
710 sequence segmentation resulting from alignment of cZ19C2 with the consensus sequence by
711 Geraghty *et al*. [25].

712 The unusually long helices II and III in the AlphaFold2 model rearranged into shorter helical
713 segments connected by turns or loops during the extended simulations. This was most pronounced
714 for helix III, where no significant common stretch of residues remained helical across the three
715 simulations. In helix II, about half of the residues maintained helicity during MD. Thus, at the
716 simulation endpoints, predominant α-helicity was observed for residues Q63 – N83 in the water
717 simulation, Q72 – S91 in the 2 mol% ethanol simulation, and L61 – S91 in the 23 mol% ethanol
718 simulation.

719 Protein quality assessment by ProSA-web [83] gave z-scores for the extended simulation
720 endpoints that were comparable with the experimental reference structures of ProSA, with the water
721 simulation giving the best agreement, see S26 Fig.

722 **Interactions Among Multiple α-zein Monomers**

723 Given that the properties of zein materials depend on interactions between zein monomers, it is of
724 interest to characterize these interactions. While a comprehensive analysis surpasses the scope of this
725 study, the following represents a first step towards examining interactions between the cZ19C2 model
726 monomers. The interactions were probed by both atomistic and coarse-grained simulations of
727 multiple copies of a globular α-zein monomer structure representative of simulations in water. As
728 described in the Methods section, this monomer structure was obtained by subjecting the 1 μs water
729 simulation endpoint to further 500 ns NPT equilibration in Desmond. The additional equilibration
730 caused increased α-helical content while maintaining a globular structure overall similar to the
731 GROMACS structure, see Fig 15A and B.

732

733 **Fig 15. Monomer preparation for multiple α-zein simulations.** A: Endpoint for the extended 1 μs
734 GROMACS MD simulation (seed 3) in water. B: Structure after additional 500 ns Desmond MD
735 simulation in water. This structure was used for both all-atom and coarse-grained simulations of
736 multiple monomers. α-carbons are shown as spheres coloured by sequence position (red: N-terminal,
737 blue: C-terminal).

738

739 **All-atom Simulations of Multiple Zeins**

740 During MD the initial configuration (Fig 16A) rapidly formed an aggregate (Fig 16 B, C) which
741 remained relatively static for the remainder of the simulation. The fast aggregation was a product of
742 the close initial proximity of the α-zein monomers and the presence of hydrophobic and polar groups
743 on their surfaces, leading to amphiphilic association among neighboring monomers. Aggregation
744 began as two sub-aggregates formed through the condensation of adjacent monomers. The first sub-
745 aggregate consisted of monomers 2, 3, 4, and 5, while the second sub-aggregate contained monomers

746 1 and 6. The latter sub-aggregate crossed the periodic boundaries of the simulation box, leading to its
747 interaction and association with monomers 2, 3, 4, and 5. These interactions between α-zein
748 monomers demonstrated a persistence or "stickiness" reminiscent of the intrachain interactions noted
749 above in monomer simulations. Thus, once again, as the interaction geometry practically becomes
750 locked upon the first monomer-monomer contact, future studies charting the landscape of zein-zein
751 interactions will benefit from enhanced sampling techniques.

752 The detailed view of the simulation endpoint in Fig 16D illustrates typical hydrophobic and
753 polar interactions between monomers in the aggregate. The interactions between monomers 1 and 6
754 include a salt-bridge between the side chains of 6:D204 and 1:R165, hydrophobic interactions
755 between 6:A206, 1:L169, and 1:A166, a hydrogen bond between the sidechain of 6:N210 and the
756 backbone oxygen of 1:L169, and additional hydrophobic interactions between residues from both
757 monomers. The backbone RMSD timelines (S35 Fig, second column) indicate some rearrangement
758 within the monomers of the aggregate. Nonetheless, this was not associated with β-sheet formation,
759 as evident from the lack of extended structure in the secondary structure timelines (S35 Fig, first
760 column). On the other hand, the simulated aggregate showed conserved or even slightly increased
761 helicity in the monomers.

762

763 **Fig 16. All-atom MD simulations of 6 copies of the globular α-zein monomer.** A: Initial
764 configuration. B: Structure after 500 ns MD simulation in water (space filling representation). C:
765 Ribbon representation of (B). D: Representative interactions between zein monomers 1 and 6.

766

767 **Coarse-grained Simulations of Multiple Zeins**

768 Oligomerization of the α-zein monomer on a larger scale was probed with coarse-grained MD
769 simulations started from the random arrangement of 73 monomer copies shown in Fig 17C. The final

770 simulation snapshots at 375 ns for three simulation replicates in Fig 17 D - I show non-covalent zein
771 aggregates with both branching and linear features, resembling fractal-like aggregates [84]. The
772 aggregates appear consistent with the known capacity of α-zeins to form viscoelastic networks in zein
773 dough systems [85], even though the simulated structures lack β-sheets, which are presumed to have
774 a central role in dough formation [86]. A notable limitation of the coarse-grained simulations is that
775 the initial secondary structure assignment remains fixed throughout the simulations. Thus, further
776 research is needed to explore if there is a connection between these simulated aggregates and actual
777 zein networks.
778
779 **Fig 17. Coarse grained MD simulations**. A: All-atom MD equilibrated structure before coarse
780 graining with heavy atoms and polar hydrogens shown in ball and stick representation. B: Martini
781 coarse-grained structure. C: Initial setup with 73 zein copies in a water box. D-F: Final MD snapshots
782 at 375 ns for simulations started from seed 1, 2, and 3 respectively. G-I: Alternative visualization of
783 the final MD snapshots for seed 1,2, and 3 showing the largest zein cluster across the periodic images
784 of the simulation box. Water and chloride ions are not shown. Each zein monomer has randomly been
785 assigned colors from a blue-red palette.

786 **α-zein Cross-linking**

787 Crosslinking is widely used in protein biomaterials engineering [87] but often targets primary amines,
788 sulfhydryl, or carboxyl groups which are scarce in α-zeins. However, the overview of α-zein amino
789 acid composition in Table A in S1 Appendix suggests other groups that may be modified. For
790 instance, α-zeins contain between 7 and 10 Tyr residues that could undergo oxidative dityrosine
791 formation [88],[89]. Additionally, 7 to 17 Ser residues could be functionalized to enable crosslinking.

792 Moreover, although most Z19 zeins lack methionine, Z22 zeins contain 3 to 5 Met residues which
793 might undergo oxidative crosslinking.

794 **Cysteine-Cysteine Distances During MD**

795 The 1 to 3 cysteines in α-zeins can facilitate limited crosslinking. This is experimentally supported
796 by Matsushima *et al.* who reported species with twice the molecular weight of Z19 and Z22
797 monomers for Royal Dent 120 α-zein [40]. Similarly, Cabra *et al.* reported 2 - 3 monomer aggregates
798 of their purified Z19 zein in 70% (v/v) aqueous methanol at unspecified α-zein concentration in DLS
799 and gel studies [33].

800 Herein, the only two cysteines in the mature cZ19C2 sequence, C6 and C27, were in close
801 proximity in the AlphaFold2 model, with a S - S distance of 3.4 Å, indicating the possibility of an
802 intrachain disulfide bridge. However, such short S-S distances were not maintained in the 400 ns
803 simulations, see S16 Fig. Instead, the average S - S distance increased with ethanol concentration,
804 consistent with the overall protein expansion at higher ethanol concentrations noted above. Increasing
805 structural fluctuations at higher ethanol concentrations are also noted from the S-S distance plots,
806 where the most pronounced oscillations occurred at ethanol concentrations ≥ 50 mol%. For instance,
807 in the 50 mol% ethanol (seed 3) simulation the S-S distance changed from 38.8 Å at 252 ns to 3.4 Å
808 at 284 ns, thus transiently approaching the initial S-S distance. Given these large variations in S-S
809 distances during MD an interesting future research direction is to explore the impact of an intrachain
810 S-S bond between C6 and C27 on the overall conformational dynamics of the α-zein.

811 **Interactions with Glycidyl Methacrylate**

812 Introduction of methacryloyl groups is a popular route to expand the cross-linking repertoire of
813 biopolymers while retaining biocompatibility. Additionally, this modification facilitates
814 photopolymerization, which allows light-based 3D printing [90]. Methacrylation has been achieved

815 for α-zein by reaction with glycidyl methacrylate (GMA) which introduced an average of 3.5
816 methacryloyl groups per zein monomer in mixtures of 70% ethanol and 30% water [91]. The authors
817 noted lower methacrylation at higher ethanol concentrations. They attributed this reduction to a
818 decrease in solvent polarity, rather than a reduced exposure of reactive groups. This interpretation
819 finds support in both the current and prior studies, as they show that higher ethanol concentrations
820 cause an increase rather than decrease in overall structural exposure. Conversely, since simulations
821 in solvent with high water content gave compact α-zein conformations, it became intriguing to
822 investigate the compatibility of these conformations with the experimentally demonstrated GMA
823 functionalization in hydrophilic environments.

824 MD simulations on the compact α-zein model in water containing GMA demonstrated both
825 exposure of α-zein polar groups and their interactions with GMA. This is illustrated in Fig 18 by the
826 time development of the number of GMA molecules within 4 Å of the protein (green) or beyond
827 (red). The top left inset figure shows the initial spherical placement of 48 GMA molecules around the
828 α-zein and the top right inset figure shows the simulation endpoint where multiple GMA molecules
829 are associated with the protein. The last 100 ns of the time series showed an average partitioning of
830 26 GMA molecules (55%) within 4 Å of the protein and 22 (45%) beyond. Representative snapshots
831 of interactions of long duration between GMA and α-zein are shown in Fig 19**.** In the first case (Fig
832 19A) several hydrophobic/amphiphilic interactions with P30, A112, L119, and other residues tether
833 the GMA molecule to the zein surface site for ~50% of the simulation time. The hydroxyl oxygen of
834 S122 was within 2.9 Å of an GMA epoxide carbon. This configuration may potentially lead to epoxide
835 ring opening and subsequent functionalization. Another notable interaction between α-zein and GMA
836 is shown in Fig 19B. Here, GMA was associated for ~15% of the simulation time with the
837 hydrophobic patch formed by A71, L74, L77, V78, L81, A125. This placed the hydroxyl oxygen of

838 Y128 within 3.3 Å of a GMA epoxide carbon, which may again lead to a pre-reaction complex for
839 functionalization.
840
841 **Fig 18. Interaction between glycidyl methacrylate (GMA) and α-zein during MD.** The inset
842 structures show the initial (t = 0 ns) and final (t = 257 ns) configuration of α-zein (white ribbons) and
843 48 GMA molecules. Red indicates GMA molecules more than 4 Å from the protein. Green indicates
844 GMA molecules within 4 Å of the protein.

845 **Fig 19. α-zein sites with high residence times for glycidyl methacrylate (GMA) during MD.**
846 GMA is shown as ball and stick models, zein residues within 4 Å of GMA are shown in stick
847 representation, zein backbone shown in ribbon representation. **A:** GMA binding persisting for ~50%
848 of the simulation time**. B**: GMA binding persisting for ~15% of the simulation time.

849 Conclusions

850 The α-zein cZ19C2 which is one of the three most highly expressed clones previously identified [30]
851 was modelled with AlphaFold2. This gave a structure with 7 loosely packed α-helical segments
852 connected by turns or loops, with a total of 68% α-helicity and no β-structure. Compact tertiary
853 structure was limited to a C-terminal bundle of three α-helices. Notably, there was a strong
854 correspondence between the individual α-helices in this bundle and the consensus repeat sequence
855 NPAAYLQQQQLLPFNQLA(V/A)(L/A) by Geraghty *et al.* [25]. This marks a deviation from
856 previously published models [28],[32],[40] based on the consensus sequence proposed by Argos *et*
857 *al.* [28].

858 To systematically examine the impact of solution environment, the AlphaFold2 model was
859 used as a starting point for multiple 400 ns MD simulations. Triplicate simulations (seed 1, 2, and 3)
860 differing by random initialization of atomic velocities were conducted in a mixed water/ethanol

861 solvent, with varying concentrations of ethanol (0 mol%, 3 mol%, 23 mol%, 50%, and 100 mol%).
862 Selected simulations (seed 3 for 100 mol% water, seed 2 for 2 mol% ethanol, and seed 2 for 23 mol%
863 ethanol) were extended to 1 μs, 2 μs, and 2 μs respectively, to enhance sampling. All simulations
864 gave markedly different structural endpoints, reflecting the loosely packed initial structure's
865 sensitivity to perturbations. In water and low (≤ 2 mol%) ethanol concentration, the α-zein model
866 rapidly collapsed to compact structures that reorganized little and slowly during the remainder of the
867 simulation. In contrast, higher ethanol concentration was associated with structural expansion and
868 increased conformational sampling. Tertiary structure, in particular the C-terminal helical bundle,
869 was partially preserved in water and at low to intermediate ethanol concentrations. At ethanol
870 concentrations ≥ 50 mol%, tertiary structure was mostly or entirely disrupted.

871 Elongated structures from MD simulations with high ethanol concentrations gave the best
872 agreement with previously published SAXS experiments. In contrast, compact conformations from
873 simulations with high water content did not align with the experimental data. These findings
874 corroborate the previous hypothesis [38] that α-zein predominantly assumes elongated and flexible
875 conformations in a binary ethanol/water solvent with a substantial ethanol concentration.

876 Principal components analysis did not reveal a clear coupling between ethanol concentration
877 and the percentage distribution of secondary structural elements (helix, sheet-, etc.). However,
878 averaged time series data indicated a subtle overall helicity increase with ethanol concentration.

879 Interactions among identical zein monomers were briefly examined with both all-atom and
880 coarse-grained simulations using the globular structural endpoint of the 1 μs water simulation (seed
881 2) after further 500 ns equilibration with the alternative force fields OPLS3 and TIP3P. In all-atom
882 MD simulations in water, 6 copies of the monomer rapidly formed an aggregate. Multiple amphiphilic
883 interactions between monomers stabilized the aggregate, hindering reorganization during the

884 remainder of the simulation. Larger scale aggregation was probed by coarse-grained MD simulations
885 on 73 α-zein monomers in water, yielding branched non-covalent networks of monomers.

886 In contrast to most published experiments, very little (generally < 1%) β-sheet or isolated
887 bridge structure was found in any of the all-atom simulations (fifteen 400 ns OPLS-AA/SPCE
888 simulations of which three simulations were extended to the microsecond timescale, four additional
889 ff99SB*-ILDN/TIP3P simulations, and one OPLS3/TIP3P simulation of 6 interacting zein monomers
890 in water.) Although this lack agrees with most published α-zein models, more exhaustive
891 conformational sampling is needed to ascertain if this reflects an inherent low β-sheet propensity for
892 the studied α-zein.

893 Interactions between the common crosslinking agent glycidyl methacrylate (GMA) and a
894 compact, globular α-zein conformation representative water simulations were probed by all-atom
895 simulations in water. After equilibration, the total 48 GMA molecules initially located ≥ 4 Å from the
896 protein were distributed with 26 GMA (55%) within 4 Å of the protein and 22 (45%) beyond,
897 suggesting high affinity between GMA and compact α-zein conformations in water. Inspecting two
898 specific GMA-zein interactions persisting for 15% and 50% of the simulation time, showed
899 arrangements where a GMA epoxide carbon was within 3.3 Å and 2.9 Å of the hydroxyl group
900 oxygens of Y128 and S122, respectively. These are potential pre-reaction complexes for
901 methacrylation of α-zein.

902 Overall, the loosely packed AlphaFold2 model allowed relatively unhindered reorganization
903 on the timescale of the simulations, thus providing a platform for α-zein conformational exploration.
904 Attractive interactions established in water-rich simulations both within the zein monomer and
905 between zein monomers tended to be highly persistent or "sticky". While possibly reflecting the
906 known aggregation propensity of α-zeins, future studies may probe the robustness of these results and
907 examine their connection to experiments. The simulations suggest directions for further

908 computational studies. For instance, the use of long simulations and multiple seeds/replicates for the
909 studied α-zein model likely extends as a guideline for α-zein models in general. In addition, "sticky"
910 interactions may potentially be overcome by enhanced sampling methods. Future simulations may
911 also explore alternative protonation states, and the impact of a disulfide bond between C6 and C27.
912 Finally, as only a single highly expressed α-zein (cZ19C2) was studied, a next step towards
913 computational elucidation of practical zein mixtures could entail simulations on the other two highly
914 expressed zeins clones [30].

915 In summary, the modelling approach for the cZ19C2 α-zein monomer herein offers a new
916 perspective relative to previous α-zein models, by subjecting an initial AlphaFold2 structure to
917 thorough MD sampling in binary ethanol/water solvents. Instead of approximating a single "native-
918 like" structure, this study presents ensembles of α-zein conformations characteristic for each ethanol
919 concentration. Such ensembles, together with the body of previous α-zein models, contribute to an
920 increasingly detailed map of α-zein conformational variation across a multitude of environments. This
921 knowledge may e.g., inform strategies for introducing chemical modifications to the α-zein monomer,
922 and in turn pave the way for advances in α-zein engineering, including the development of new
923 biomaterials.

924 **References**

925 1. Haider TP, Völker C, Kramm J, Landfester K, Wurm FR. Plastics of the Future? The
926 Impact of Biodegradable Polymers on the Environment and on Society. Angewandte Chemie
927 International Edition. 2019;58(1):50-62. doi: https://doi.org/10.1002/anie.201805766.
928 2. Tortorella S, Maturi M, Vetri Buratti V, Vozzolo G, Locatelli E, Sambri L, et al. Zein as
929 a versatile biopolymer: different shapes for different biomedical applications. RSC Advances.
930 2021;11(62):39004-26. doi: 10.1039/D1RA07424E.

931 3. Angelova N, Hunkeler D. Rationalizing the design of polymeric biomaterials. Trends in
932 Biotechnology. 1999;17(10):409-21. doi: https://doi.org/10.1016/S0167-7799(99)01356-6.
933 4. Mejia CD, Mauer LJ, Hamaker BR. Similarities and differences in secondary structure
934 of viscoelastic polymers of maize α-zein and wheat gluten proteins. Journal of Cereal Science.
935 2007;45(3):353-9. doi: https://doi.org/10.1016/j.jcs.2006.09.009.
936 5. Zhang Y, Xu M, Zhang X, Hu Y, Luan G. Application of zein in gluten-free foods: A
937 comprehensive review. Food Research International. 2022;160:111722. doi:
938 https://doi.org/10.1016/j.foodres.2022.111722.
939 6. Burr B, Burr FA. Zein synthesis in maize endosperm by polyribosomes attached to
940 protein bodies. Proc Natl Acad Sci U S A. 1976;73(2):515-9. Epub 1976/02/01. doi:
941 10.1073/pnas.73.2.515. PubMed PMID: 1061153; PubMed Central PMCID: PMCPMC335940.
942 7. Shewry PR, Tatham AS. The prolamin storage proteins of cereal seeds: structure and
943 evolution. Biochem J. 1990;267(1):1-12. Epub 1990/04/01. doi: 10.1042/bj2670001. PubMed PMID:
944 2183790; PubMed Central PMCID: PMCPMC1131235.
945 8. Dickey LC, Dallmer MF, Radewonuk ER, Parris N, Kurantz M, Craig JC. Zein Batch
946 Extraction from Dry-Milled Corn: Cereal Disintegration by Dissolving Fluid Shear. Cereal Chemistry.
947 1998;75(4):443-8. doi: https://doi.org/10.1094/CCHEM.1998.75.4.443.
948 9. Corradini E, Curti PS, Meniqueti AB, Martins AF, Rubira AF, Muniz EC. Recent
949 Advances in Food-Packing, Pharmaceutical and Biomedical Applications of Zein and Zein-Based
950 Materials. International Journal of Molecular Sciences. 2014;15(12):22438-70. PubMed PMID:
951 doi:10.3390/ijms151222438.
952 10. de Vries A, Nikiforidis CV, Scholten E. Natural amphiphilic proteins as tri-block Janus
953 particles: Self-sorting into thermo-responsive gels. Europhysics Letters. 2014;107(5):58003. doi:
954 10.1209/0295-5075/107/58003.
955 11. Wang Y, Padua GW. Nanoscale Characterization of Zein Self-Assembly. Langmuir.
956 2012;28(5):2429-35. doi: 10.1021/la204204j.

957 12. Weissmueller NT, Lu HD, Hurley A, Prud'homme RK. Nanocarriers from GRAS Zein
958 Proteins to Encapsulate Hydrophobic Actives. Biomacromolecules. 2016;17(11):3828-37. doi:
959 10.1021/acs.biomac.6b01440.
960 13. Chaunier L, Leroy E, Della Valle G, Lourdin D. Plasticized protein for 3D printing by
961 fused deposition modeling. AIP Conference Proceedings. 2016;1769(1):190001. doi:
962 10.1063/1.4963611.
963 14. Chaunier L, Dalgalarrondo M, Della Valle G, Lourdin D, Marion D, Leroy E. Rheological
964 characterization of plasticized corn proteins for fused deposition modeling. AIP Conference
965 Proceedings. 2017;1896(1):040002. doi: 10.1063/1.5008028.
966 15. Tavares-Negrete JA, Aceves-Colin AE, Rivera-Flores DC, Díaz-Armas GG, Mertgen
967 A-S, Trinidad-Calderón PA, et al. Three-Dimensional Printing Using a Maize Protein: Zein-Based
968 Inks in Biomedical Applications. ACS Biomaterials Science & Engineering. 2021;7(8):3964-79. doi:
969 10.1021/acsbiomaterials.1c00544.
970 16. Zhang Y, Raza A, Xue Y-Q, Yang G, Hayat U, Yu J, et al. Water-responsive 4D printing
971 based on self-assembly of hydrophobic protein "Zein" for the control of degradation rate and drug
972 release. Bioactive Materials. 2023;23:343-52. doi: https://doi.org/10.1016/j.bioactmat.2022.11.009.
973 17. Luecha J, Hsiao A, Brodsky S, Liu GL, Kokini JL. Green microfluidic devices made of
974 corn proteins. Lab on a Chip. 2011;11(20):3419-25. doi: 10.1039/C1LC20726A.
975 18. Mattice KD, Marangoni AG. Functionalizing zein through antisolvent precipitation from
976 ethanol or aetic acid. Food Chemistry. 2020;313:126127. doi:
977 https://doi.org/10.1016/j.foodchem.2019.126127.
978 19. Taylor JRN, Johnson SK, Taylor J, Njila S, Jackaman C. Oxidation of commercial (α-
979 type) zein with hydrogen peroxide improves its hydration and dramatically increases dough
980 extensibility even below its glass transition temperature. Journal of Cereal Science. 2016;70:108-15.
981 doi: https://doi.org/10.1016/j.jcs.2016.05.025.

982 20. Chen G, Chen Y, Jin N, Li J, Dong S, Li S, et al. Zein films with porous polylactic acid
983 coatings via cold plasma pre-treatment. Industrial Crops and Products. 2020;150:112382. doi:
984 https://doi.org/10.1016/j.indcrop.2020.112382.
985 21. Mattice KD, Marangoni AG. Physical properties of zein networks treated with microbial
986 transglutaminase. Food Chemistry. 2021;338:128010. doi:
987 https://doi.org/10.1016/j.foodchem.2020.128010.
988 22. Kim S, Sessa DJ, Lawton JW. Characterization of zein modified with a mild cross-
989 linking agent. Industrial Crops and Products. 2004;20(3):291-300. doi:
990 https://doi.org/10.1016/j.indcrop.2003.10.013.
991 23. Viotti A, Cairo G, Vitale A, Sala E. Each zein gene class can produce polypeptides of
992 different sizes. The EMBO Journal. 1985;4(5):1103-10. doi: https://doi.org/10.1002/j.1460-
993 2075.1985.tb03746.x.
994 24. Geraghty D, Peifer MA, Rubenstein I, Messing J. The primary structure of a plant
995 storage protein: zein. Nucleic Acids Research. 1981;9(19):5163-74. doi: 10.1093/nar/9.19.5163.
996 25. Geraghty DE, Messing J, Rubenstein I. Sequence analysis and comparison of cDNAs
997 of the zein multigene family. Embo j. 1982;1(11):1329-35. Epub 1982/01/01. doi: 10.1002/j.1460-
998 2075.1982.tb01318.x. PubMed PMID: 6897917; PubMed Central PMCID: PMCPMC553212.
999 26. Pedersen K, Devereux J, Wilson DR, Sheldon E, Larkins BA. Cloning and sequence
1000 analysis reveal structural variation among related zein genes in maize. Cell. 1982;29(3):1015-26.
1001 doi: https://doi.org/10.1016/0092-8674(82)90465-2.
1002 27. Marks MD, Larkins BA. Analysis of sequence microheterogeneity among zein
1003 messenger RNAs. Journal of Biological Chemistry. 1982;257(17):9976-83. doi:
1004 https://doi.org/10.1016/S0021-9258(18)33973-5.
1005 28. Argos P, Pedersen K, Marks MD, Larkins BA. A structural model for maize zein
1006 proteins. The Journal of biological chemistry. 1982;257 17:9984-90.

1007 29. Marks MD, Lindell JS, Larkins BA. Nucleotide sequence analysis of zein mRNAs from
1008 maize endosperm. Journal of Biological Chemistry. 1985;260(30):16451-9. doi: 10.1016/S0021-
1009 9258(17)36258-0.
1010 30. Woo Y-M, Hu DW-N, Larkins BA, Jung R. Genomics Analysis of Genes Expressed in
1011 Maize Endosperm Identifies Novel Seed Proteins and Clarifies Patterns of Zein Gene Expression.
1012 The Plant Cell. 2001;13(10):2297-317. doi: 10.1105/tpc.010240.
1013 31. Spena A, Viotti A, Pirrotta V. A homologous repetitive block structure underlies the
1014 heterogeneity of heavy and light chain zein genes. The EMBO Journal. 1982;1(12):1589-94. doi:
1015 https://doi.org/10.1002/j.1460-2075.1982.tb01360.x.
1016 32. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P. Studies of the zein-like α-prolamins
1017 based on an analysis of amino acid sequences: Implications for their evolution and three-dimensional
1018 structure. Proteins: Structure, Function, and Bioinformatics. 1993;15(1):88-99. doi:
1019 https://doi.org/10.1002/prot.340150111.
1020 33. Cabra V, Arreguin R, Galvez A, Quirasco M, Vazquez-duhalt R, Farres A.
1021 Characterization of a 19 kDa α-Zein of High Purity. Journal of Agricultural and Food Chemistry.
1022 2005;53(3):725-9. doi: 10.1021/jf048530s.
1023 34. Uzun S, Ilavsky J, Padua GW. Characterization of zein assemblies by ultra-small-angle
1024 X-ray scattering. Soft Matter. 2017;13(16):3053-60. doi: 10.1039/C6SM02717B.
1025 35. Bugs MR, Forato LA, Bortoleto-Bugs RK, Fischer H, Mascarenhas YP, Ward RJ, et al.
1026 Spectroscopic characterization and structural modeling of prolamin from maize and pearl millet.
1027 European Biophysics Journal. 2004;33(4):335-43. doi: 10.1007/s00249-003-0354-3.
1028 36. Forato LA, Bicudo TDC, Colnago LA. Conformation of α zeins in solid state by Fourier
1029 transform IR. Biopolymers. 2003;72(6):421-6. doi: https://doi.org/10.1002/bip.10481.
1030 37. Akhmetova A, Lanno GM, Kogermann K, Malmsten M, Rades T, Heinz A. Highly
1031 Elastic and Water Stable Zein Microfibers as a Potential Drug Delivery System for Wound Healing.
1032 Pharmaceutics. 2020;12(5). Epub 2020/05/24. doi: 10.3390/pharmaceutics12050458. PubMed
1033 PMID: 32443445; PubMed Central PMCID: PMCPMC7284525.

1034 38. Forato LA, Doriguetto AC, Fischer H, Mascarenhas YP, Craievich AF, Colnago LA.
1035 Conformation of the Z19 Prolamin by FTIR, NMR, and SAXS. Journal of Agricultural and Food
1036 Chemistry. 2004;52(8):2382-5. doi: 10.1021/jf035020+.
1037 39. Tatham AS, Field JM, Morris VJ, I'Anson KJ, Cardle L, Dufton MJ, et al. Solution
1038 conformational analysis of the alpha-zein proteins of maize. Journal of Biological Chemistry.
1039 1993;268(35):26253-9. doi: https://doi.org/10.1016/S0021-9258(19)74308-7.
1040 40. Matsushima N, Danno G-i, Takezawa H, Izumi Y. Three-dimensional structure of maize
1041 α-zein proteins studied by small-angle X-ray scattering. Biochimica et Biophysica Acta (BBA) -
1042 Protein Structure and Molecular Enzymology. 1997;1339(1):14-22. doi:
1043 https://doi.org/10.1016/S0167-4838(96)00212-9.
1044 41. Momany FA, Sessa DJ, Lawton JW, Selling GW, Hamaker SA, Willett JL. Structural
1045 characterization of alpha-zein. J Agric Food Chem. 2006;54(2):543-7. Epub 2006/01/19. doi:
1046 10.1021/jf058135h. PubMed PMID: 16417318.
1047 42. Tan H, Zhou H, Guo T, Zhang Y, Ma L. Integrated multi-spectroscopic and molecular
1048 modeling techniques to study the formation mechanism of hidden zearalenone in maize. Food
1049 Chemistry. 2021;351:129286. doi: https://doi.org/10.1016/j.foodchem.2021.129286.
1050 43. Liu C, Lv N, Ren G, Wu R, Wang B, Cao Z, et al. Explore the interaction mechanism
1051 between zein and EGCG using multi-spectroscopy and molecular dynamics simulation methods.
1052 Food Hydrocolloids. 2021;120:106906. doi: https://doi.org/10.1016/j.foodhyd.2021.106906.
1053 44. Arooj M, Shehadi I, Nassab CN, Mohamed AA. Physicochemical stability study of
1054 protein-benzoic acid complexes using molecular dynamics simulations. Amino Acids.
1055 2020;52(9):1353-62. Epub 2020/10/03. doi: 10.1007/s00726-020-02897-2. PubMed PMID:
1056 33006112.
1057 45. Hajjari MM, Sharif N. In-silico behavior of dissolved prolamins under electric field effect
1058 applied by electrospinning process using molecular dynamics simulation. Journal of Molecular
1059 Liquids. 2021;344:117778. doi: https://doi.org/10.1016/j.molliq.2021.117778.

1060 46. Hajjari MM, Sharif N. In-silico zein/tannic acid colloidal nanoparticles and their activity
1061 at oil and water interface of Pickering emulsion using molecular dynamics simulation. Journal of
1062 Molecular Liquids. 2022;359:119321. doi: https://doi.org/10.1016/j.molliq.2022.119321.
1063 47. Erickson DP, Ozturk OK, Selling G, Chen F, Campanella OH, Hamaker BR. Corn zein
1064 undergoes conformational changes to higher β-sheet content during its self-assembly in an
1065 increasingly hydrophilic solvent. International Journal of Biological Macromolecules. 2020;157:232-
1066 9. doi: https://doi.org/10.1016/j.ijbiomac.2020.04.169.
1067 48. Erickson DP, Dunbar M, Hamed E, Ozturk OK, Campanella OH, Keten S, et al.
1068 Atomistic Modeling of Peptide Aggregation and β-Sheet Structuring in Corn Zein for Viscoelasticity.
1069 Biomacromolecules. 2021;22(5):1856-66. doi: 10.1021/acs.biomac.0c01558.
1070 49. Lawton JW. Zein: A History of Processing and Use. Cereal Chemistry. 2002;79(1):1-
1071 18. doi: https://doi.org/10.1094/CCHEM.2002.79.1.1.
1072 50. Selling GW, Hamaker SAH, Sessa DJ. Effect of Solvent and Temperature on
1073 Secondary and Tertiary Structure of Zein by Circular Dichroism. Cereal Chemistry. 2007;84(3):265-
1074 70. doi: https://doi.org/10.1094/CCHEM-84-3-0265.
1075 51. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly
1076 accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. doi:
1077 10.1038/s41586-021-03819-2.
1078 52. The UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids
1079 Research. 2021;49(D1):D480-D9. doi: 10.1093/nar/gkaa1100.
1080 53. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold:
1081 making protein folding accessible to all. Nature Methods. 2022;19(6):679-82. doi: 10.1038/s41592-
1082 022-01488-1.
1083 54. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of
1084 Molecular Graphics. 1996;14(1):33-8. doi: https://doi.org/10.1016/0263-7855(96)00018-5.
1085 55. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal
1086 of Chemical Theory and Computation. 2008;4(1):116-22. doi: 10.1021/ct700200b.

1087 56. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5:
1088 a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics.
1089 2013;29(7):845-54. doi: 10.1093/bioinformatics/btt055.
1090 57. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and
1091 Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum
1092 Chemical Calculations on Peptides. The Journal of Physical Chemistry B. 2001;105(28):6474-87.
1093 doi: 10.1021/jp003919d.
1094 58. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair
1095 potentials. The Journal of Physical Chemistry. 1987;91(24):6269-71. doi: 10.1021/j100308a038.
1096 59. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a
1097 general amber force field. Journal of Computational Chemistry. 2004;25(9):1157-74. doi:
1098 https://doi.org/10.1002/jcc.20035.
1099 60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of
1100 simple potential functions for simulating liquid water. The Journal of Chemical Physics.
1101 1983;79(2):926-35. doi: 10.1063/1.445869.
1102 61. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the
1103 Helix−Coil Transition of Polypeptides. The Journal of Physical Chemistry B. 2009;113(26):9004-15.
1104 doi: 10.1021/jp901540t.
1105 62. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved
1106 side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function,
1107 and Bioinformatics. 2010;78(8):1950-8. doi: https://doi.org/10.1002/prot.22711.
1108 63. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al., editors.
1109 Scalable algorithms for molecular dynamics simulations on commodity clusters. SC'06: Proceedings
1110 of the 2006 ACM/IEEE Conference on Supercomputing; 2006: IEEE.
1111 64. Schrödinger Release 2018-1: Desmond Molecular Dynamics System, D. E. Shaw
1112 Research, New York, NY, 2018. Maestro-Desmond Interoperability Tools, Schrödinger, New York,
1113 NY, 2018.

1114 65. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field
1115 Providing Broad Coverage of Drug-like Small Molecules and Proteins. Journal of Chemical Theory
1116 and Computation. 2016;12(1):281-96. doi: 10.1021/acs.jctc.5b00864.
1117 66. Schrödinger Release 2018-1: Maestro, Schrödinger, LLC, New York, NY, 2018.
1118 67. de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schäfer LV, et al.
1119 Improved Parameters for the Martini Coarse-Grained Protein Force Field. Journal of Chemical
1120 Theory and Computation. 2013;9(1):687-97. doi: 10.1021/ct300646g.
1121 68. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of
1122 hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637. Epub 1983/12/01.
1123 doi: 10.1002/bip.360221211. PubMed PMID: 6667333.
1124 69. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from
1125 known atomic coordinates of proteins. Nucleic Acids Research. 2004;32(suppl_2):W500-W2. doi:
1126 10.1093/nar/gkh429.
1127 70. Postu PA, Ion L, Drochioiu G, Petre BA, Glocker MO. Mass spectrometric
1128 characterization of the zein protein composition in maize flour extracts upon protein separation by
1129 SDS-PAGE and 2D gel electrophoresis. ELECTROPHORESIS. 2019;40(20):2747-58. doi:
1130 https://doi.org/10.1002/elps.201900108.
1131 71. Shen M, Broeckling CD, Chu EY, Ziegler G, Baxter IR, Prenni JE, et al. Leveraging
1132 Non-Targeted Metabolite Profiling via Statistical Genomics. PLOS ONE. 2013;8(2):e57667. doi:
1133 10.1371/journal.pone.0057667.
1134 72. Sharma S, Pradhan R, Manickavasagan A, Thimmanagari M, Dutta A. Exploration of
1135 corn distillers solubles from selective milling technology as a novel source of plant-based ACE
1136 inhibitory protein hydrolysates. Food Chemistry. 2022;388:133036. doi:
1137 https://doi.org/10.1016/j.foodchem.2022.133036.
1138 73. Santos-Rosales V, Alvarez-Rivera G, Hillgärtner M, Cifuentes A, Itskov M, García-
1139 González CA, et al. Stability Studies of Starch Aerogel Formulations for Biomedical Applications.
1140 Biomacromolecules. 2020;21(12):5336-44. doi: 10.1021/acs.biomac.0c01414.

1141 74. Simm D, Hatje K, Kollmar M. Waggawagga: comparative visualization of coiled-coil
1142 predictions and detection of stable single α-helices (SAH domains). Bioinformatics. 2015;31(5):767-
1143 9. doi: 10.1093/bioinformatics/btu700.
1144 75. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. DeepCoil—a fast
1145 and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics.
1146 2019;35(16):2790-5. doi: 10.1093/bioinformatics/bty1062.
1147 76. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for
1148 comparing protein structures and models using distance difference tests. Bioinformatics.
1149 2013;29(21):2722-8. Epub 2013/08/30. doi: 10.1093/bioinformatics/btt473. PubMed PMID:
1150 23986568; PubMed Central PMCID: PMCPMC3799472.
1151 77. Barnes CA, Shen Y, Ying J, Takagi Y, Torchia DA, Sellers JR, et al. Remarkable
1152 Rigidity of the Single α-Helical Domain of Myosin-VI As Revealed by NMR Spectroscopy. J Am Chem
1153 Soc. 2019;141(22):9004-17. Epub 2019/05/24. doi: 10.1021/jacs.9b03116. PubMed PMID:
1154 31117653; PubMed Central PMCID: PMCPMC6556874.
1155 78. Simm D, Hatje K, Kollmar M. Distribution and evolution of stable single α-helices (SAH
1156 domains) in myosin motor proteins. PLoS One. 2017;12(4):e0174639. Epub 2017/04/04. doi:
1157 10.1371/journal.pone.0174639. PubMed PMID: 28369123; PubMed Central PMCID:
1158 PMCPMC5378345 ONE policies on sharing data and materials.
1159 79. Martínez L. Automatic Identification of Mobile and Rigid Substructures in Molecular
1160 Dynamics Simulations and Fractional Structural Fluctuation Analysis. PLOS ONE.
1161 2015;10(3):e0119264. doi: 10.1371/journal.pone.0119264.
1162 80. Kretschmer CB. Infrared Spectroscopy and Optical Rotatory Dispersion of Zein, Wheat
1163 Gluten and Gliadin. The Journal of Physical Chemistry. 1957;61(12):1627-31. doi:
1164 10.1021/j150558a014.
1165 81. Piiadov V, Ares de Araújo E, Oliveira Neto M, Craievich AF, Polikarpov I. SAXSMoW
1166 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS

1167 data measured on a relative scale. Protein Science. 2019;28(2):454-63. doi:
1168 https://doi.org/10.1002/pro.3528.
1169 82. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile
1170 computation and its assessment by contrast variation experiments. Biophys J. 2013;105(4):962-74.
1171 Epub 2013/08/27. doi: 10.1016/j.bpj.2013.07.020. PubMed PMID: 23972848; PubMed Central
1172 PMCID: PMCPMC3752106.
1173 83. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of
1174 errors in three-dimensional structures of proteins. Nucleic Acids Research. 2007;35(suppl_2):W407-
1175 W10. doi: 10.1093/nar/gkm290.
1176 84. Filippov AV, Zurita M, Rosner DE. Fractal-like Aggregates: Relation between
1177 Morphology and Physical Properties. Journal of Colloid and Interface Science. 2000;229(1):261-73.
1178 doi: https://doi.org/10.1006/jcis.2000.7027.
1179 85. Zhang X, Dong C, Hu Y, Gao M, Luan G. Zein as a structural protein in gluten-free
1180 systems: an overview. Food Science and Human Wellness. 2021;10(3):270-7. doi:
1181 https://doi.org/10.1016/j.fshw.2021.02.018.
1182 86. Sadat A, Corradini MG, Joye IJ. Vibrational and fluorescence spectroscopy to study
1183 gluten and zein interactions in complex dough systems. Curr Res Food Sci. 2022;5:479-90. Epub
1184 2022/03/11. doi: 10.1016/j.crfs.2022.02.009. PubMed PMID: 35265856; PubMed Central PMCID:
1185 PMCPMC8899121.
1186 87. Gomes MC, Mano JF. Chemical modification strategies to prepare advanced protein-
1187 based biomaterials. Biomaterials and Biosystems. 2021;1:100010. doi:
1188 https://doi.org/10.1016/j.bbiosy.2021.100010.
1189 88. Mukherjee S, Kapp EA, Lothian A, Roberts AM, Vasil'ev YV, Boughton BA, et al.
1190 Characterization and Identification of Dityrosine Cross-Linked Peptides Using Tandem Mass
1191 Spectrometry. Analytical Chemistry. 2017;89(11):6136-45. doi: 10.1021/acs.analchem.7b00941.

1192 89. Liu C, Hua J, Ng PF, Fei B. Photochemistry of bioinspired dityrosine crosslinking.
1193 Journal of Materials Science & Technology. 2021;63:182-91. doi:
1194 https://doi.org/10.1016/j.jmst.2020.02.086.
1195 90. Moon SH, Hwang HJ, Jeon HR, Park SJ, Bae IS, Yang YJ. Photocrosslinkable natural
1196 polymers in tissue engineering. Front Bioeng Biotechnol. 2023;11:1127757. Epub 2023/03/28. doi:
1197 10.3389/fbioe.2023.1127757. PubMed PMID: 36970625; PubMed Central PMCID:
1198 PMCPMC10037533.
1199 91. Allasia M, Sonzogni AS, Vaillard VA, Vaillard SE, Gugliotta LM, Minari RJ.
1200 Methacrylated zein as a novel biobased macro-crosslinker for PVCL hydrogels. Polymer.
1201 2021;235:124278. doi: https://doi.org/10.1016/j.polymer.2021.124278.
1202

1203 ## Supporting information

1204

1205 **S1 Fig. AlphaFold2 per-residue estimate of confidence (plDDT) for the rank 1 model of α-zein**
1206 **cZ19C2 (UniProtKB ID: P06677).** The scale runs from 0 (minimum confidence) to 100 (maximum
1207 confidence).

1208

1209 **S2 Fig. Overview of the all-atom GROMACS simulations with the OPLS-AA/L and SPCE force**
1210 **fields.** 15 × 400 ns simulations of which 3 simulations were extended to at least 1 μs.

1211

1212 **S3 Fig. Time series for 400 ns GROMACS all-atom MD simulations.** A, D, G, J, M: Backbone
1213 RMSD. B, E, H, K N: Radius of gyration. C, F, I, L O: SASA. Values given in the legend for each
1214 seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

1215

1216 **S4 Fig. Time series for secondary structure for 100 mol% water 400 ns MD simulations.** Values
1217 given in the legend for each seed are averages over the last 100 ns of each simulation, with standard
1218 deviations given in brackets.
1219
1220 **S5 Fig. Time series for secondary structure for 2 mol% ethanol 400 ns MD simulations.** Values
1221 given in the legend for each seed are averages over the last 100 ns of each simulation, with standard
1222 deviations given in brackets.
1223
1224 **S6 Fig. Time series for secondary structure for 23 mol% ethanol 400 ns MD simulations.** Values
1225 given in the legend for each seed are averages over the last 100 ns of each simulation, with standard
1226 deviations given in brackets.
1227
1228 **S7 Fig. Time series for secondary structure for 50 mol% ethanol 400 ns MD simulations.** Values
1229 given in the legend for each seed are averages over the last 100 ns of each simulation, with standard
1230 deviations given in brackets.
1231
1232 **S8 Fig. Time series for secondary structure for 100 mol% ethanol 400 ns MD simulations.**
1233 Values given in the legend for each seed are averages over the last 100 ns of each simulation, with
1234 standard deviations given in brackets.
1235
1236 **S9 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1237 **AlphaFold2 model and averaged over 100 mol% water 400 ns MD simulations.** Orange: Helicity
1238 in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2,
1239 and 3) 100 mol% water 400 ns MD simulations.
1240

1241 **S10 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1242 **AlphaFold2 model and averaged over 2 mol% ethanol 400 ns MD simulations.** Orange: Helicity
1243 in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2,
1244 and 3) 2 mol% ethanol 400 ns MD simulations.
1245
1246 **S11 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1247 **AlphaFold2 model and averaged over 23 mol% ethanol 400 ns MD simulations.** Orange: Helicity
1248 in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2,
1249 and 3) 23 mol% ethanol 400 ns MD simulations.
1250
1251 **S12 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1252 **AlphaFold2 model and averaged over 50 mol% ethanol 400 ns MD simulations.** Orange: Helicity
1253 in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2,
1254 and 3) 50 mol% ethanol 400 ns MD simulations.
1255
1256 **S13 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1257 **AlphaFold2 model and averaged over 100 mol% ethanol 400 ns MD simulations.** Orange:
1258 Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three
1259 (Seed 1, 2, and 3) 100 mol% ethanol 400 ns MD simulations.
1260
1261 **S14 Fig. Principal Component Analysis of Helicity Profiles from 400 ns MD simulations.** Left:
1262 Scores for the first two principal components. Right: Loadings for the two first principal components.
1263 The PCA was carried out on helicity profiles obtained from the last 100 ns of each 400 ns MD
1264 simulations by averaging the helicity per residue across the last 100 snapshots. Simulations are color

1265 coded based on ethanol concentration. The suffixes #1, #2, and #3 indicate different simulation
1266 replicates (seeds).
1267
1268 **S15 Fig. Transient β-sheet formation in 400 ns ethanol simulations.** A: Seed 1, B: Seed 2. The
1269 snapshots are from t = 346 ns in both cases.
1270
1271 **S16 Fig. Distance between sulfur atoms in the two cysteines across all 400 ns MD trajectories.**
1272 Mean and standard deviations in this plot are calculated using the entire 400 ns range.
1273
1274 **S17 Fig. SAXS predictions.** Predictions for 2 mol% ethanol showing the best fit (t = 2 ns) and the
1275 fit to the last MD frame (t = 400 ns).
1276
1277 **S18 Fig. SAXS predictions.** Predictions for 50 mol% ethanol showing the two best fits at t = 161
1278 ns and t = 160 ns.
1279
1280 **S19 Fig. SAXS predictions.** Predictions for 100 mol% ethanol showing the two best fits at t = 234
1281 ns and t = 400 ns.
1282
1283 **S20 Fig. Time series for extended GROMACS all-atom MD simulations.** A, D, G: Backbone
1284 RMSD. B, E, H: Radius of gyration. C, F, I: SASA. Values given in the legend for each seed are
1285 averages over the last 100 ns of each simulation, with standard deviations given in brackets.
1286
1287 **S21 Fig. Time series for secondary structure for extended GROMACS all-atom MD**
1288 **simulations.** Top: 100 mol% water simulation extended to 1000 ns. Middle: 2 mol% ethanol

simulation extended to 2000 ns. Bottom: 23 mol% ethanol simulation extended to 2000 ns. Values given in the legend are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S22 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) for the 100 mol% water 1 µs simulation. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity during the last 100 ns.

S23 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) for the 2 mol% ethanol 2 µs simulation. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity during the last 100 ns.

S24 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) for the 23 mol% ethanol 2 µs simulation. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity during the last 100 ns.

S25 Fig. Example of β-sheet structure in the extended 2 mol% (seed 2) ethanol simulation. Left: MD snapshot at 1971 ns. Right: Enlarged view of the backbone hydrogen bond pairs A9/I26 and I11/A24. The protein backbone is shown in ribbon representation coloured by secondary structure. α-carbons are shown as spheres coloured by sequence position (red: N-terminal, blue: C-terminal).

S26 Fig. ProSA-web quality checks for extended MD simulation endpoints

S27 Fig. MD time series for 400 ns GROMACS all-atom MD simulations with the ff99SB*-ILDN force field. A, D: Backbone RMSD. B, E: Radius of gyration. C, F: SASA. Values given in

1314 the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations
1315 given in brackets.
1316
1317 **S28 Fig. MD time series for an extended GROMACS all-atom MD simulation with the ff99SB*-**
1318 **ILDN force field.** A: Backbone RMSD. B: Radius of gyration. C: SASA. The value given in the
1319 legend is an average over the last 100 ns, with standard deviation given in brackets.
1320
1321 **S29 Fig. Time series for secondary structure for 400 ns GROMACS all-atom MD simulations**
1322 **with the ff99SB*-ILDN force field.** Top: 100 mol% water simulation (seed 1). Middle: 100 mol%
1323 water simulation (seed 1) extended to 800 ns. Bottom: 100 mol% water simulation (seed 2). Values
1324 given in the legend are averages over the last 100 ns of each simulation, with standard deviations
1325 given in brackets.
1326
1327 **S30 Fig. Time series for secondary structure for 400 ns GROMACS all-atom MD simulations**
1328 **with the ff99SB*-ILDN force field.** Top: 24 mol% ethanol simulation (seed 1). Bottom: 24 mol%
1329 ethanol simulation (seed 2). Values given in the legend are averages over the last 100 ns of each
1330 simulation, with standard deviations given in brackets.
1331
1332 **S31 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1333 **AlphaFold2 model and averaged over the last 100 ns of the 100 mol% water 400 ns MD**
1334 **simulation with the ff99SB*-ILDN force field.** Orange: Helicity in the initial AlphaFold2 model.
1335 Blue: Helicity averaged over the last 100 ns of the 100 mol% water 400 ns Seed 1 simulation.
1336

1337 **S32 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1338 **AlphaFold2 model and averaged over the last 100 ns of the 100 mol% water 800 ns MD**
1339 **simulation with the ff99SB*-ILDN force field.** Orange: Helicity in the initial AlphaFold2 model.
1340 Blue: Helicity averaged over the last 100 ns of the 100 mol% water Seed 1 MD simulation extended
1341 to 800 ns.
1342
1343 **S33 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1344 **AlphaFold2 model and averaged over 100 mol% water 400 ns MD simulations with the**
1345 **ff99SB*-ILDN force field.** Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the
1346 last 100 ns averaged over the two (Seed 1, Seed 2) 100 mol% water 400 ns MD simulations.
1347
1348 **S34 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial**
1349 **AlphaFold2 model and averaged over 24 mol% ethanol 400 ns MD simulations with the**
1350 **ff99SB*-ILDN force field.** Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the
1351 last 100 ns averaged over the two (Seed 1, Seed 2) 24 mol% ethanol 400 ns MD simulations.
1352
1353 **S35 Fig. Secondary structure and backbone RMSD for all-atom MD simulations of 6 zein copies**
1354 **in a water box.**
1355
1356 **S36 Fig. Mature protein sequence alignment and identity matrix for P06677 (19 kDa zein 19C2),**
1357 **P04703 (19 kDa zein A20), P04702 (19 kDa zein M6), P06676 (19 kDa zein 19C1).**
1358
1359 **S1 Appendix. Overview of major α-zein sequences**
1360

1361 **S2 Appendix. Homology and *ab initio* models**

1362

1363 **S3 Appendix. Additional simulation details**

1364

1365

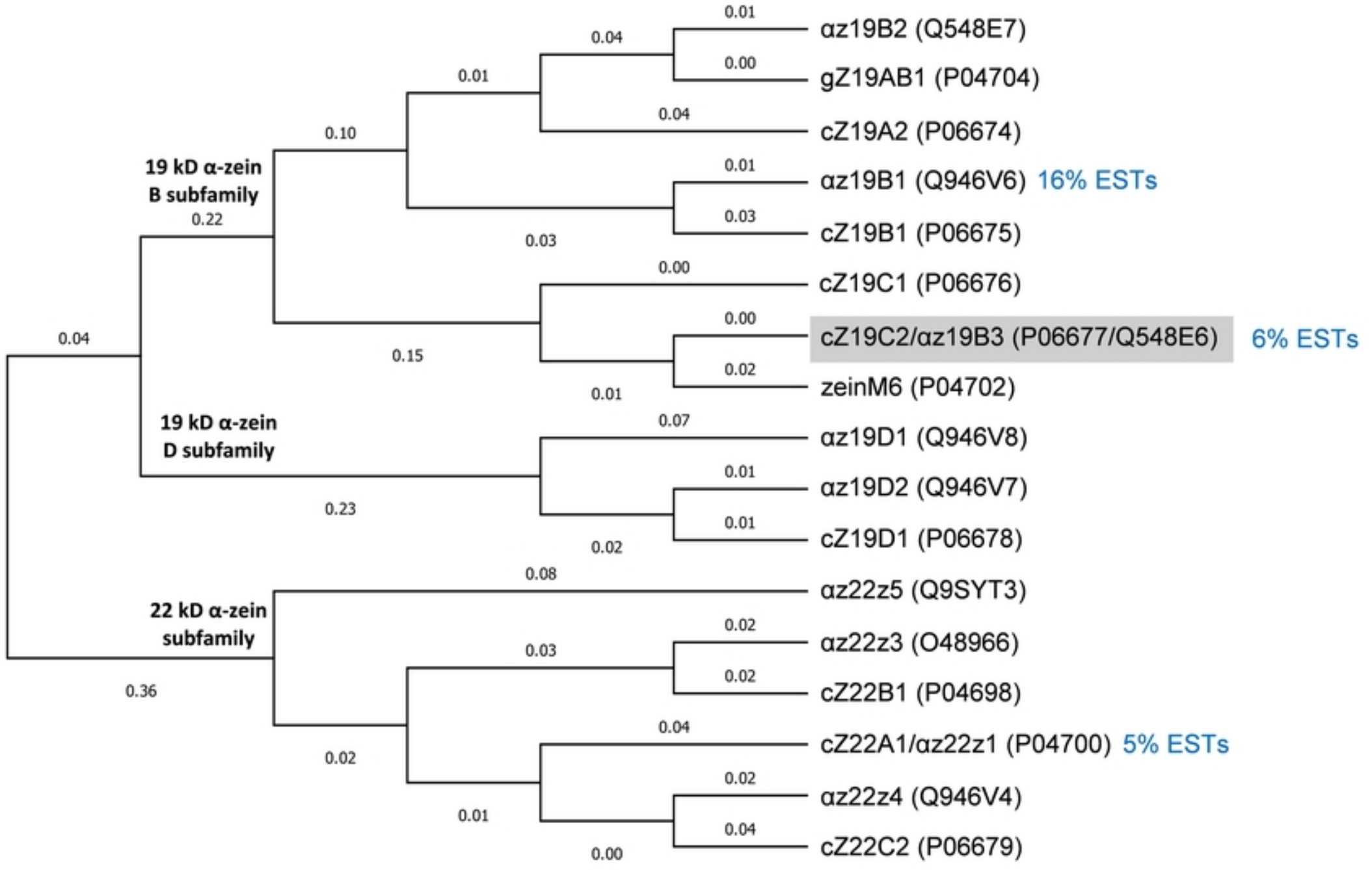
αz19B2 (Q548E7)
gZ19AB1 (P04704)
cZ19A2 (P06674)
αz19B1 (Q946V6) 16% ESTs
cZ19B1 (P06675)
cZ19C1 (P06676)
cZ19C2/αz19B3 (P06677/Q548E6) 6% ESTs
zeinM6 (P04702)
αz19D1 (Q946V8)
αz19D2 (Q946V7)
cZ19D1 (P06678)
αz22z5 (Q9SYT3)
αz22z3 (O48966)
cZ22B1 (P04698)
cZ22A1/αz22z1 (P04700) 5% ESTs
αz22z4 (Q946V4)
cZ22C2 (P06679)
19 kD α-zein B subfamily
19 kD α-zein D subfamily
22 kD α-zein subfamily
0.01
0.04
0.00
0.01
0.04
0.10
0.01
0.03
0.22
0.03
0.00
0.00
0.04
0.15
0.02
0.01
0.07
0.01
0.23
0.01
0.02
0.08
0.02
0.03
0.02
0.36
0.04
0.02
0.02
0.01
0.04
0.00

Figure 1

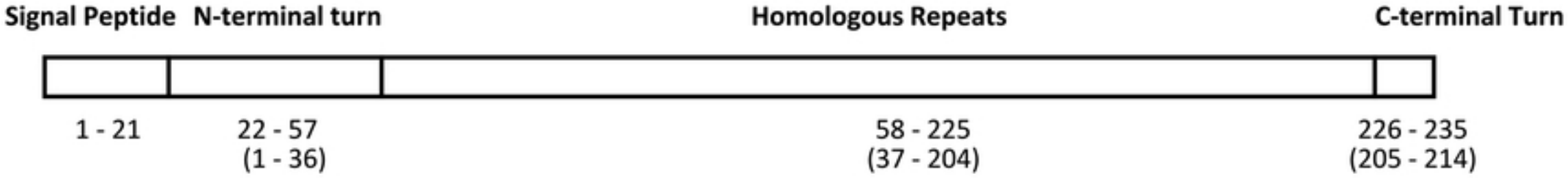
Signal Peptide
N-terminal turn
Homologous Repeats
C-terminal Turn
1 - 21
22 - 57
(1 - 36)
58 - 225
(37 - 204)
226 - 235
(205 - 214)

Figure 2

A

Segment	N	P	A	A	Y	L	Q	Q	Q	Q	L	L	P	F	N	Q	L	A	A/V	A/L
1													T1	I2	F3	P4	Q5	C6	S7	Q8
2	A9	P10	I11	A12	S13	L14	L15	P16	P17	Y18	L19	P20	S21	I22	I23	A24	S25	I26	C27	E28
3	N29	P30	A31	L32	Q33	P34	Y35	R36	L37	Q38	Q39	A40	I41	A42	A43	S44	N45	I46	P47	L48
4	S49	P50	L51	L52	F53	Q54	Q55	S56	P57	A58	L59	S60	L61	V62	Q63	S64	L65	V66	Q67	T68
5	I69	R70	A71	Q72	Q73	L74	Q75	Q76	L77	V78	L79	P80	L81	I82	N83	Q84	V85	A86	L87	A88
6	N89	L90	S91	P92	Y93	S94	Q95	Q96	Q97	Q98	F99	L100	P101	F102	N103	Q104	L105	S106	T107	L108
7	N109	P110	A111	A112	Y113	L114	Q115	Q116	Q117		L118	L119	P120	F121	S122	Q123	L124	A125	T126	
8				A127	Y128	S129	Q130	Q131	Q132	Q133	L134	L135	P136	F137	N138	Q139	L140	A141	A142	L143
9	N144	P145	A146	A147	Y148	L149	Q150	Q151	Q152	I153	L154	L155	P156	F157	S158	Q159	L160	A161	A162	A163
10	N164	R165	A166	S167	F168	L169	T170	Q171	Q172	Q173	L174	L175	P176	F177	Y178	Q179	Q180	F181	A182	A183
11	N184	P185	A186	T187	L188	L189	Q190	L191	Q192	Q193	L194	L195	P196	F197	V198	Q199	L200	A201	L202	T203
12	D204	P205	A206	A207	S208	Y209	Q210	Q211	H212	I213	I214	G215	G216	A217	L218	F219				

B

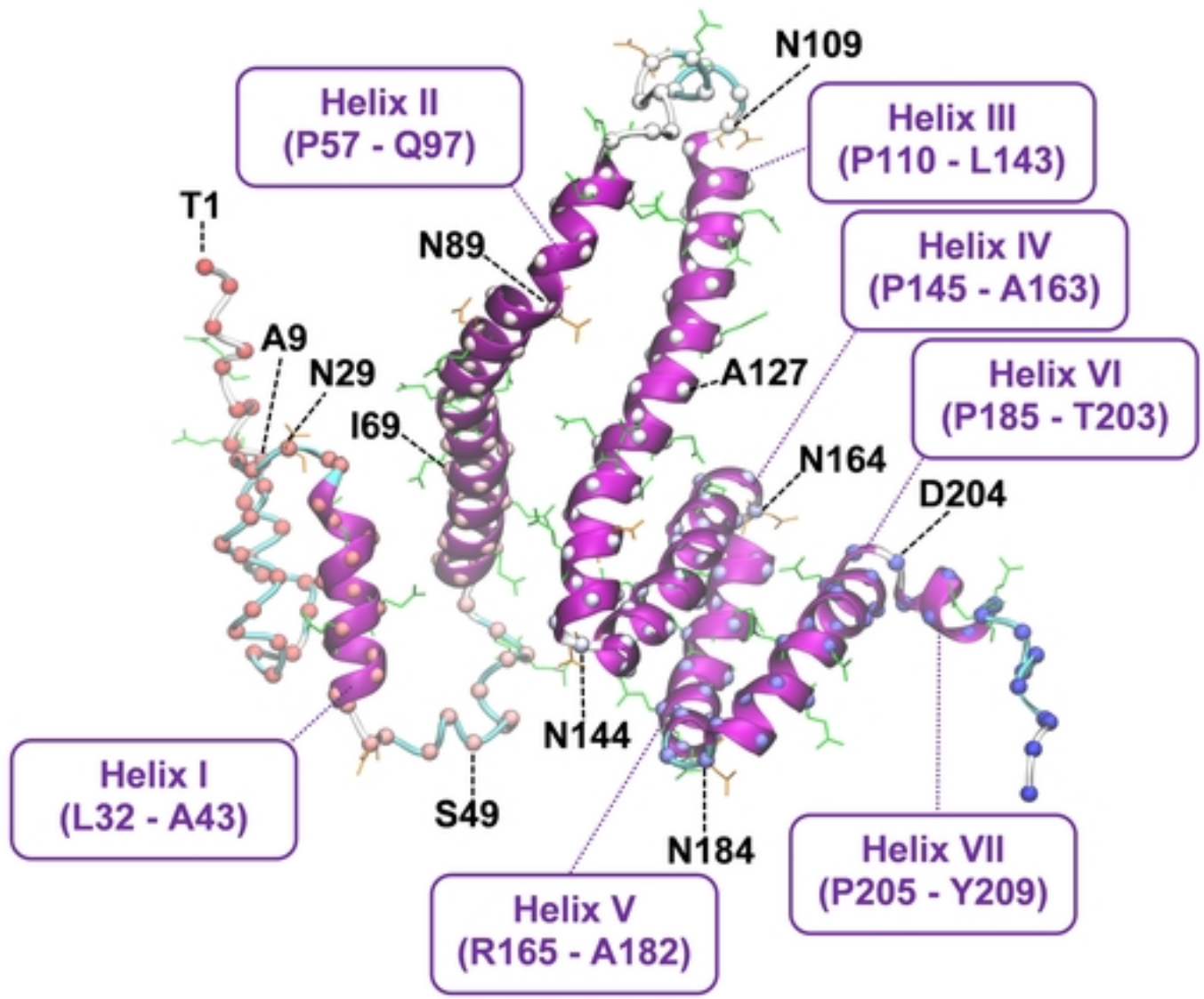

Figure 3

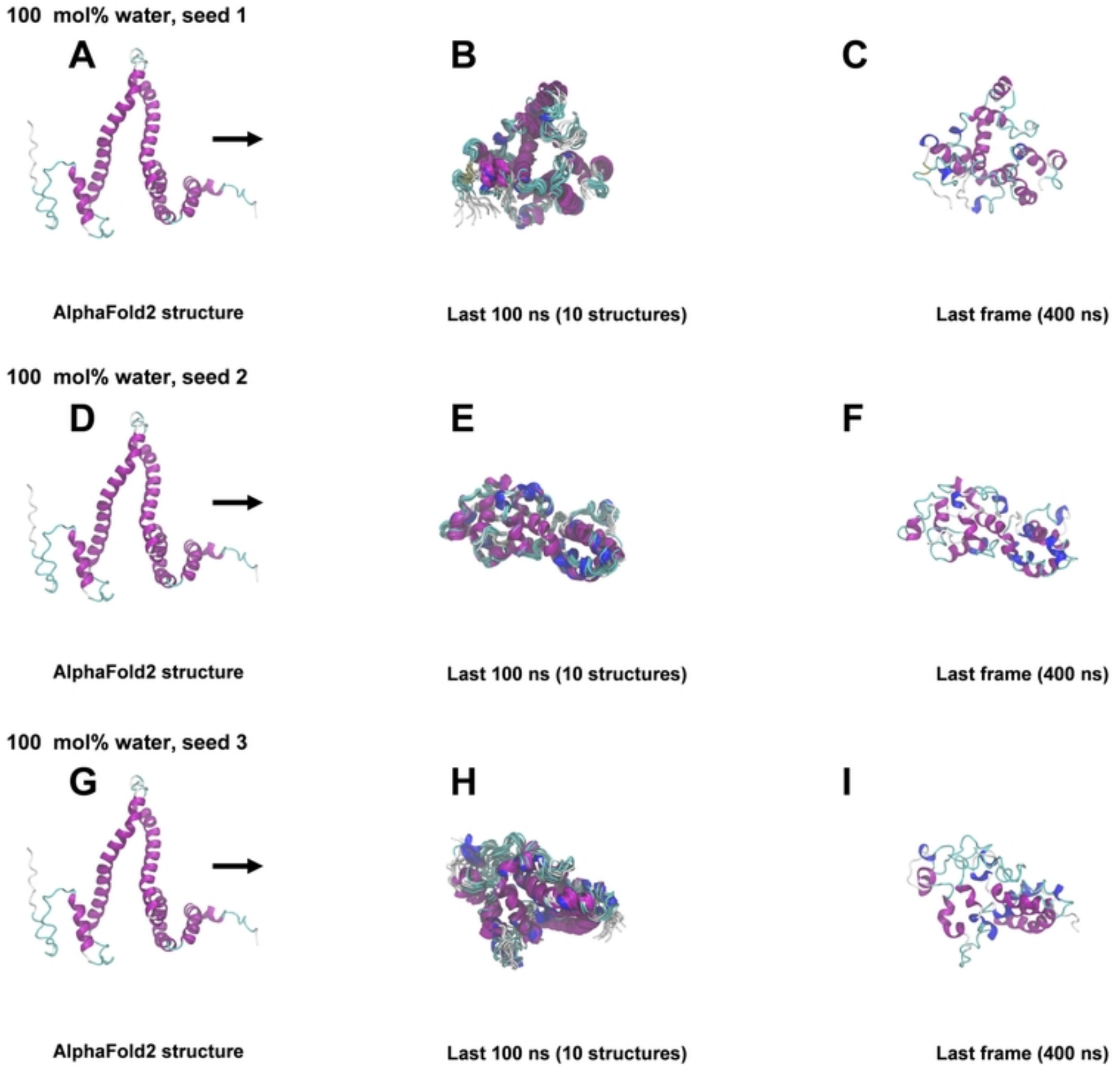
100 mol% water, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% water, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% water, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 4

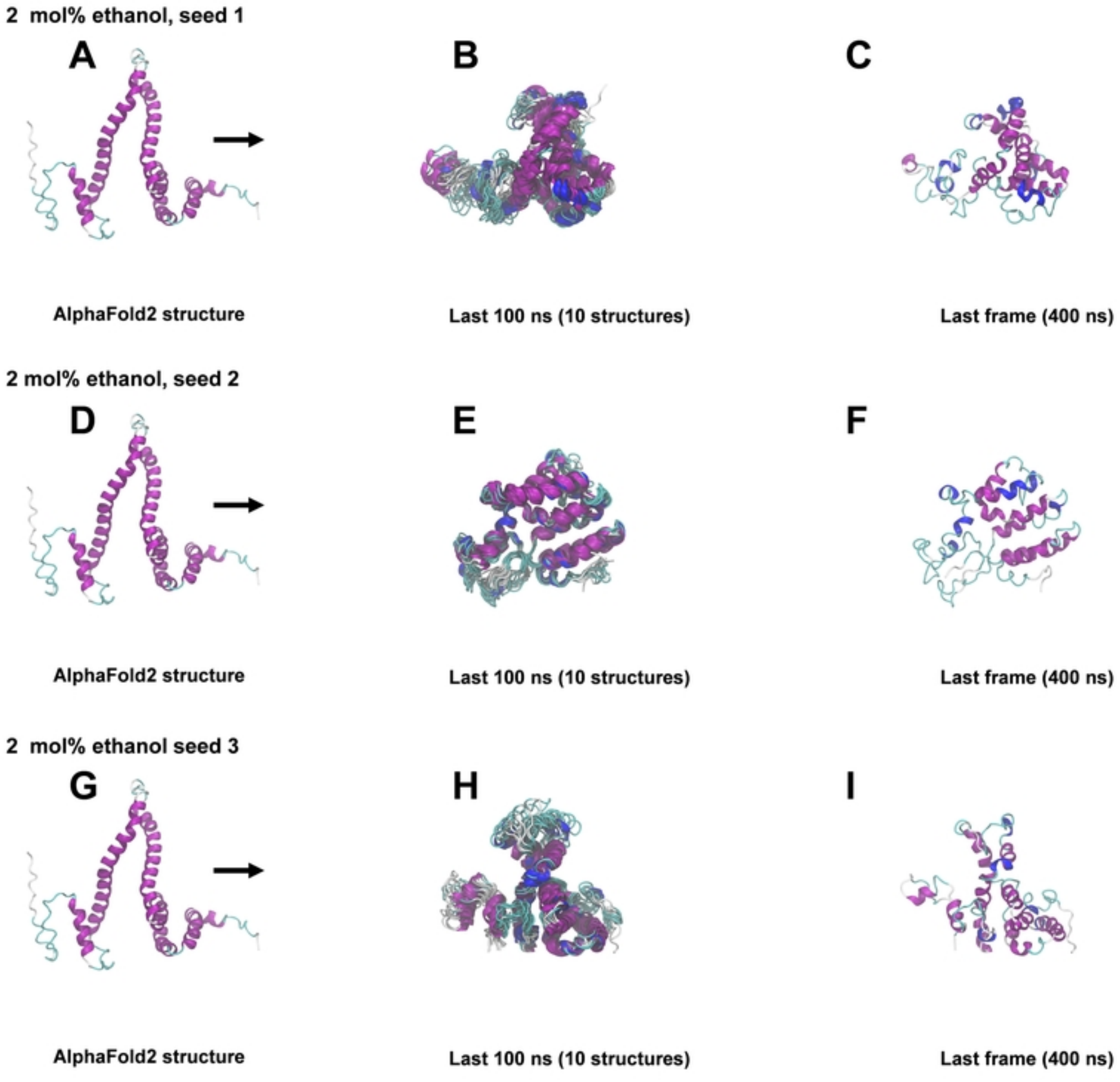
2 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
2 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
2 mol% ethanol seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 5

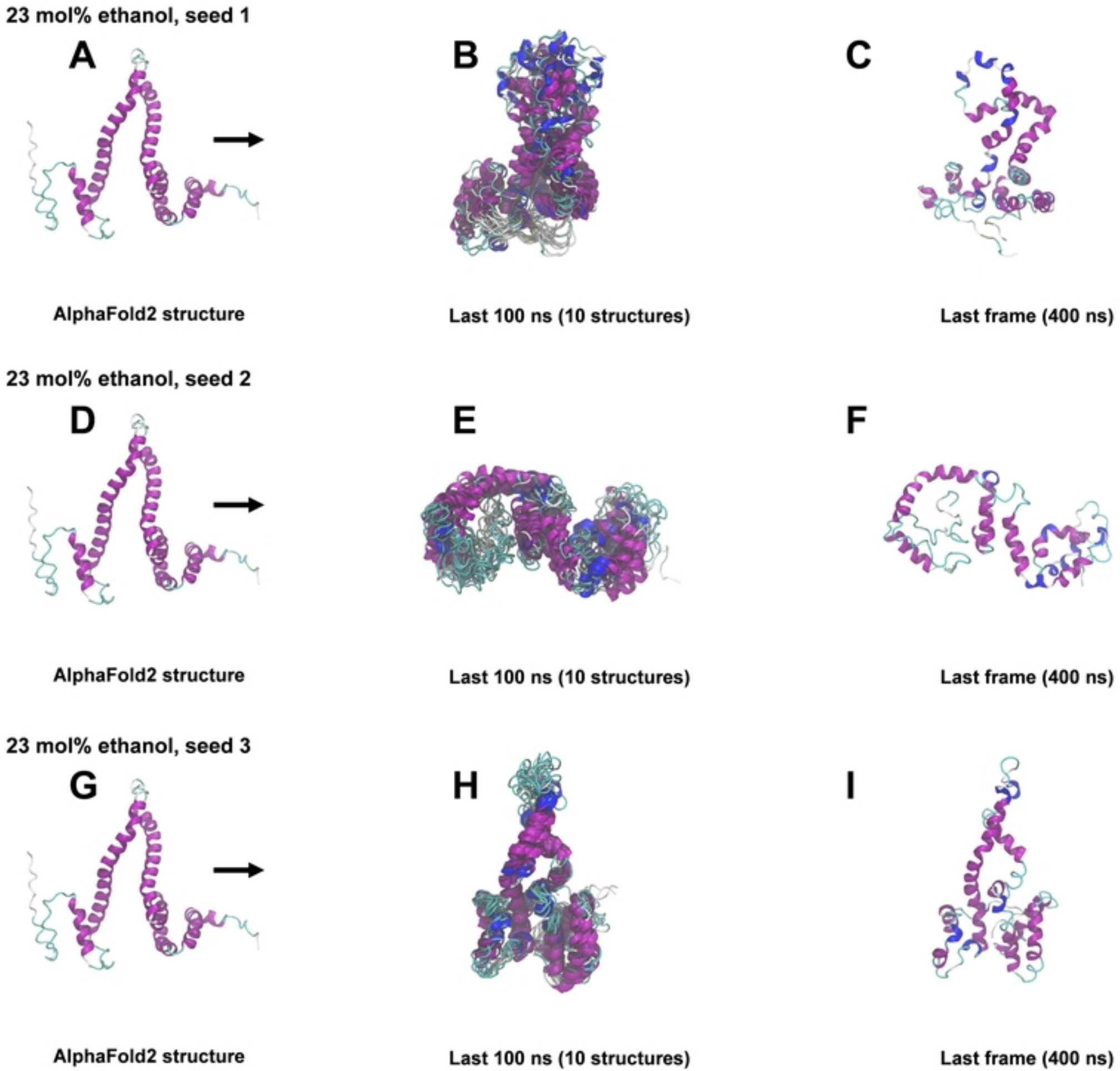
23 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
23 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
23 mol% ethanol, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 6

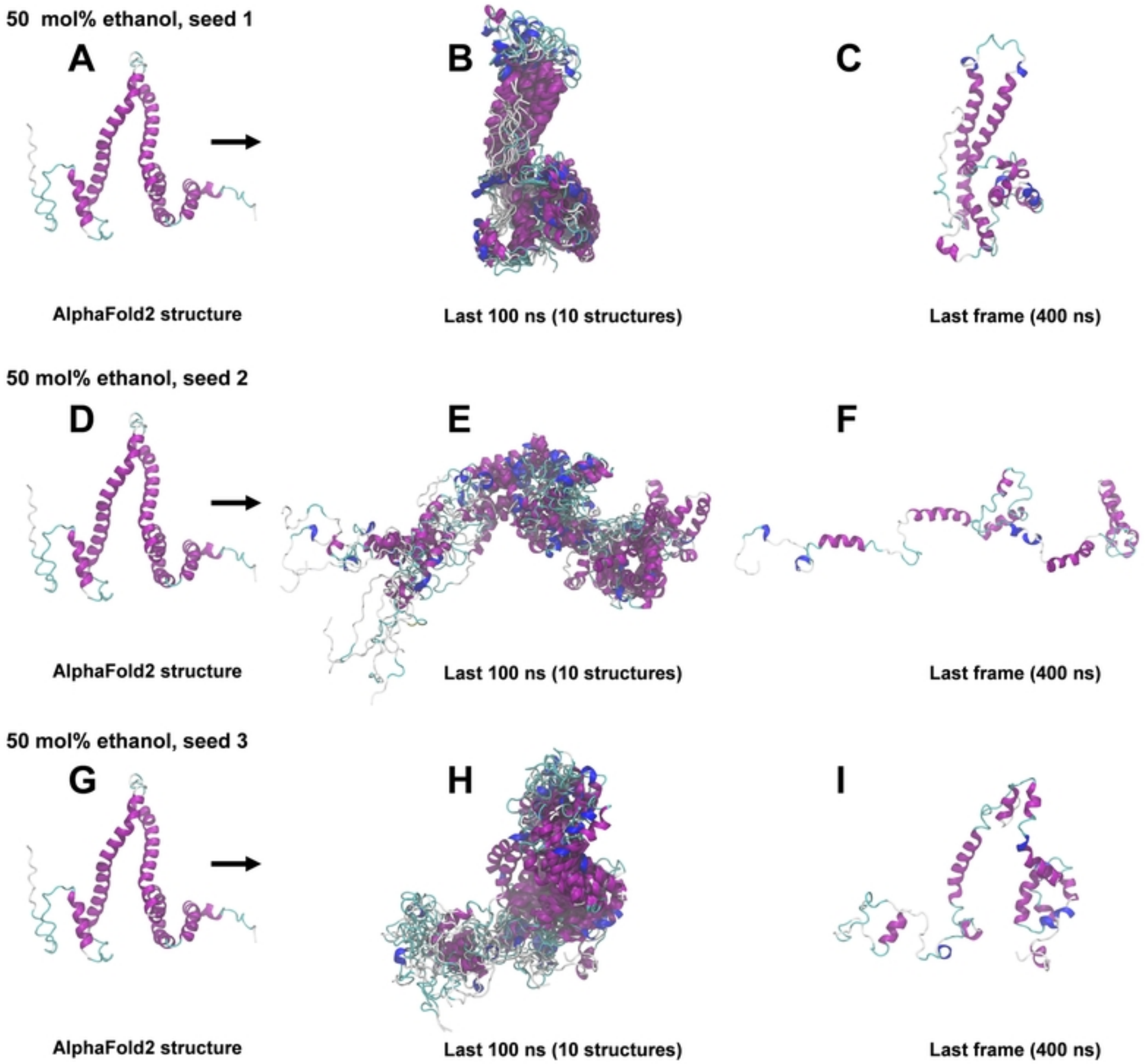
50 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
50 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
50 mol% ethanol, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 7

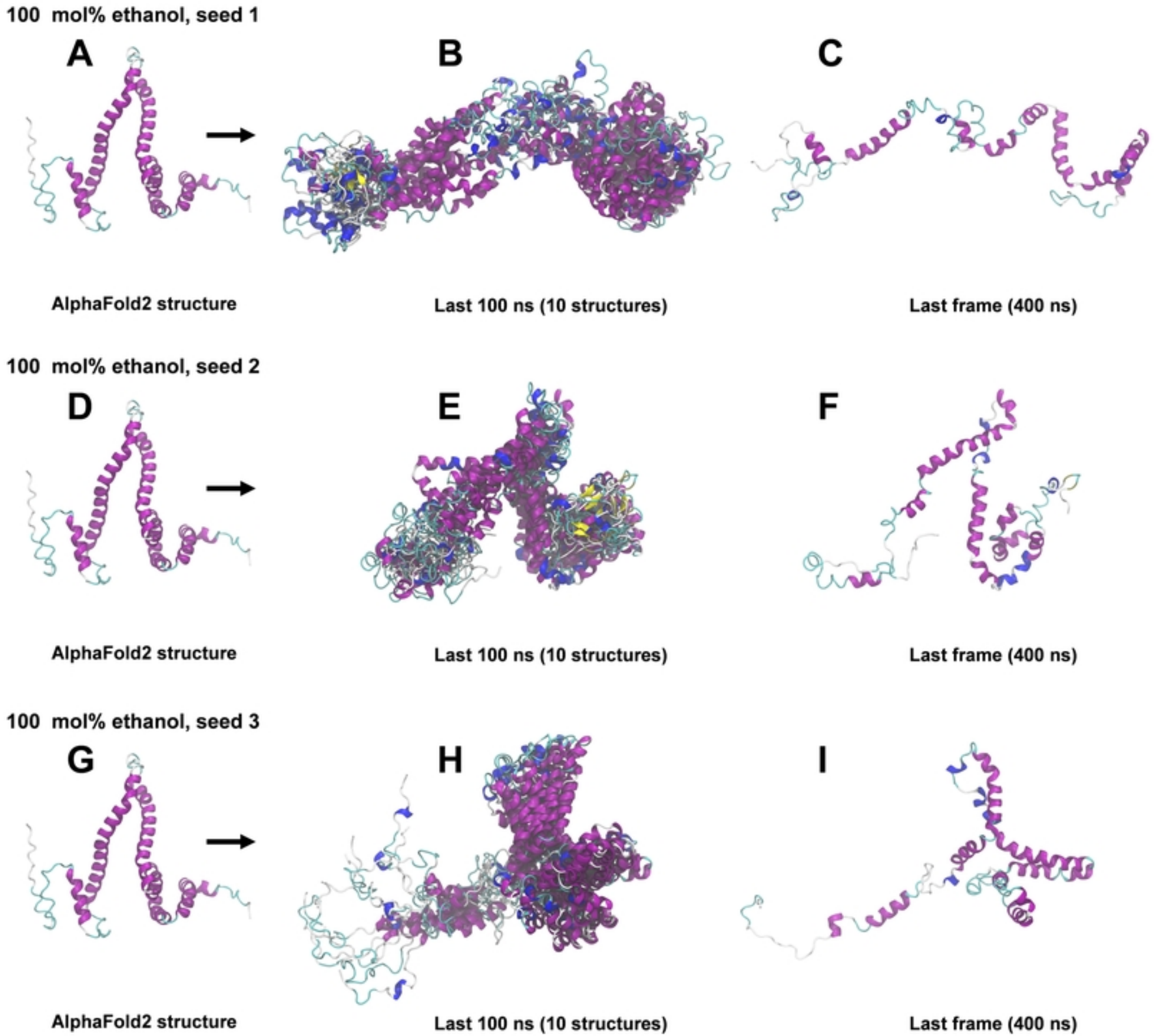
100 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% ethanol, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 8

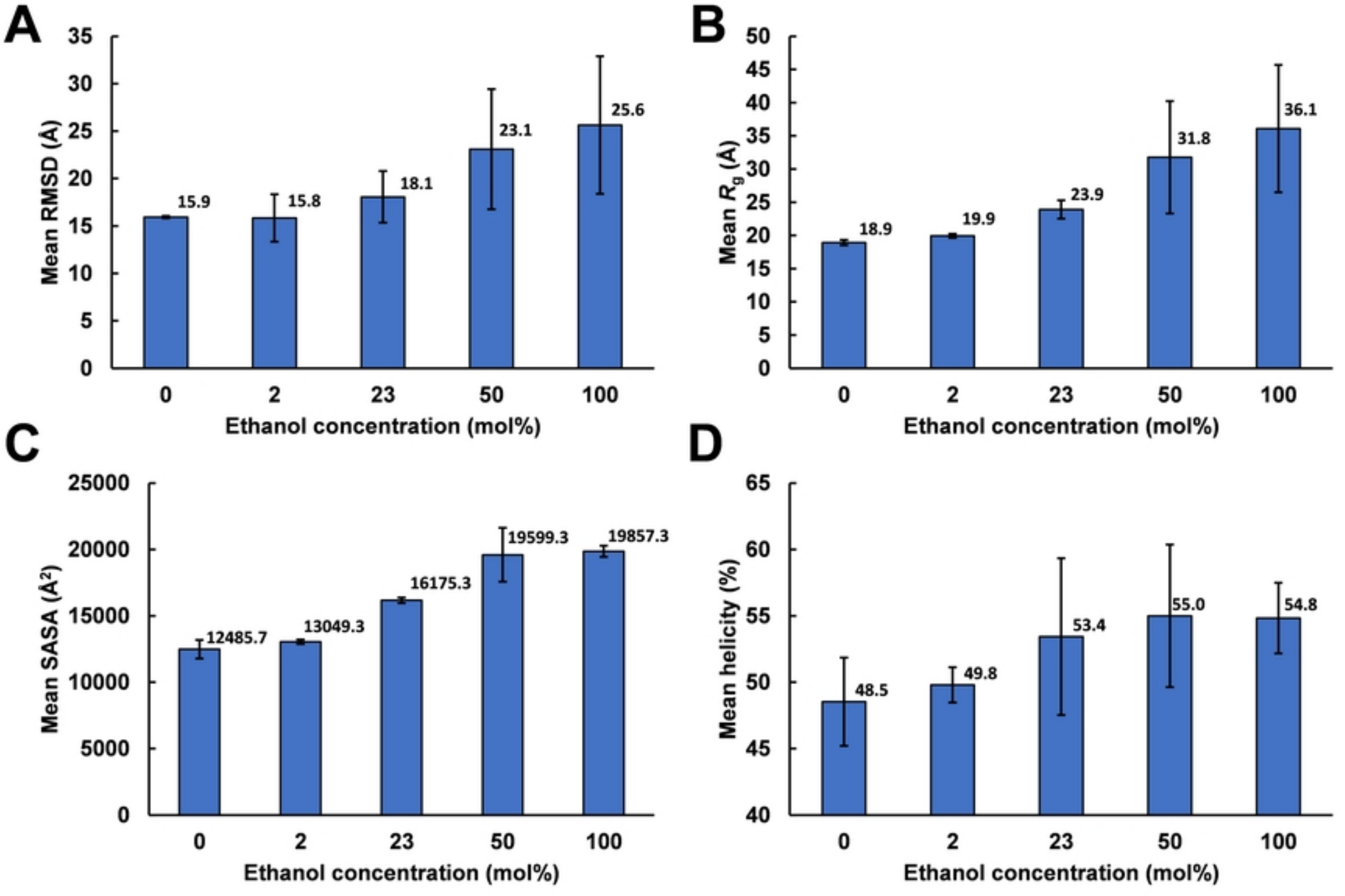
A
Mean RMSD (Å)
35
30
25
20
15
10
5
0
15.9
15.8
18.1
23.1
25.6
0
2
23
50
100
Ethanol concentration (mol%)
B
Mean R_g (Å)
50
45
40
35
30
25
20
15
10
5
0
18.9
19.9
23.9
31.8
36.1
0
2
23
50
100
Ethanol concentration (mol%)
C
Mean SASA (Å²)
25000
20000
15000
10000
5000
0
12485.7
13049.3
16175.3
19599.3
19857.3
0
2
23
50
100
Ethanol concentration (mol%)
D
Mean helicity (%)
65
60
55
50
45
40
48.5
49.8
53.4
55.0
54.8
0
2
23
50
100
Ethanol concentration (mol%)

Figure 9

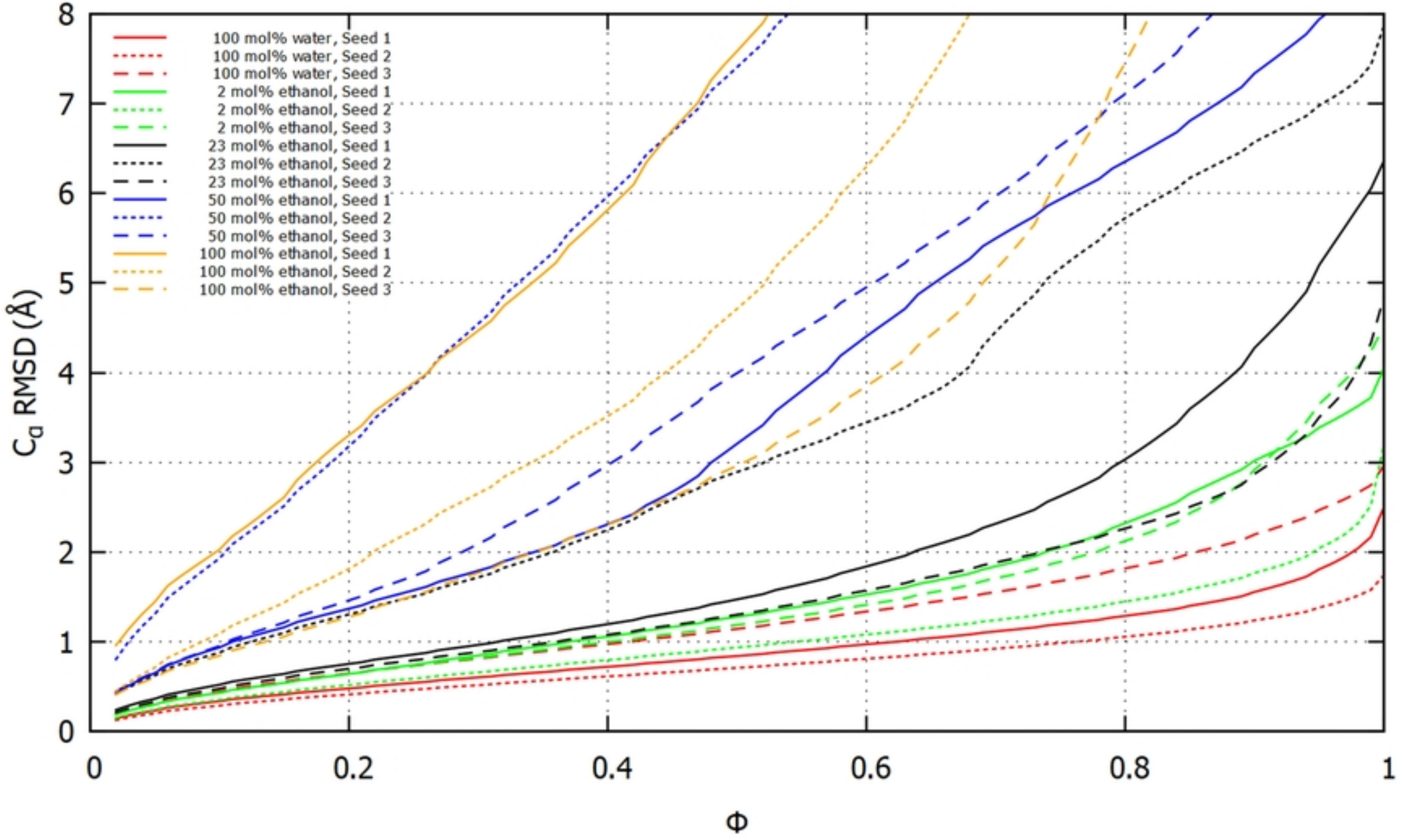
100 mol% water, Seed 1
100 mol% water, Seed 2
100 mol% water, Seed 3
2 mol% ethanol, Seed 1
2 mol% ethanol, Seed 2
2 mol% ethanol, Seed 3
23 mol% ethanol, Seed 1
23 mol% ethanol, Seed 2
23 mol% ethanol, Seed 3
50 mol% ethanol, Seed 1
50 mol% ethanol, Seed 2
50 mol% ethanol, Seed 3
100 mol% ethanol, Seed 1
100 mol% ethanol, Seed 2
100 mol% ethanol, Seed 3
C_α RMSD (Å)
Φ
0
1
2
3
4
5
6
7
8
0
0.2
0.4
0.6
0.8
1

Figure 10

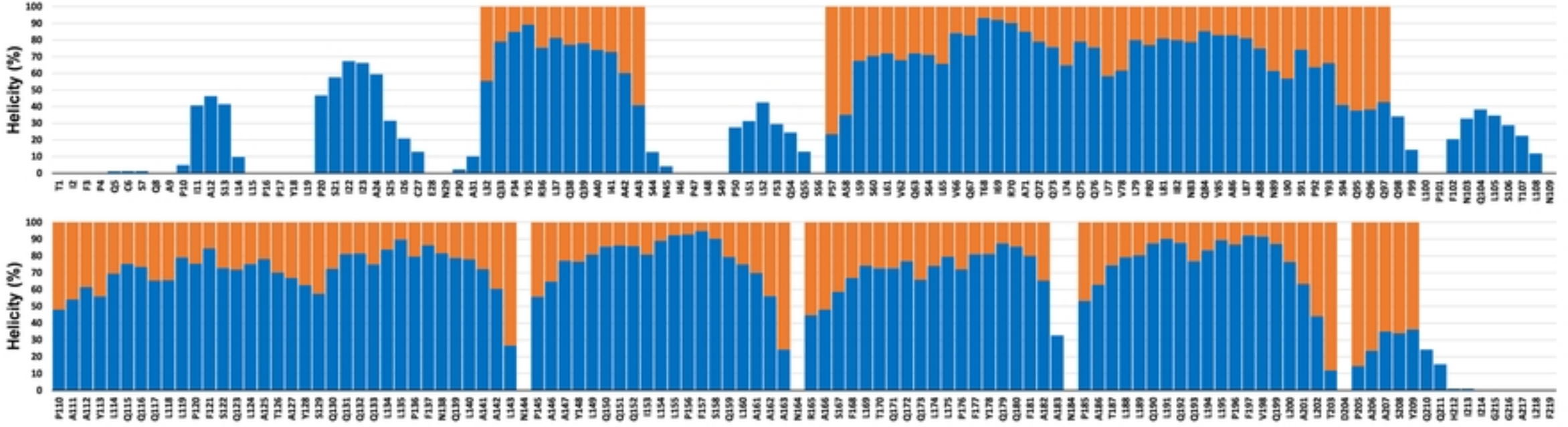

Figure 11

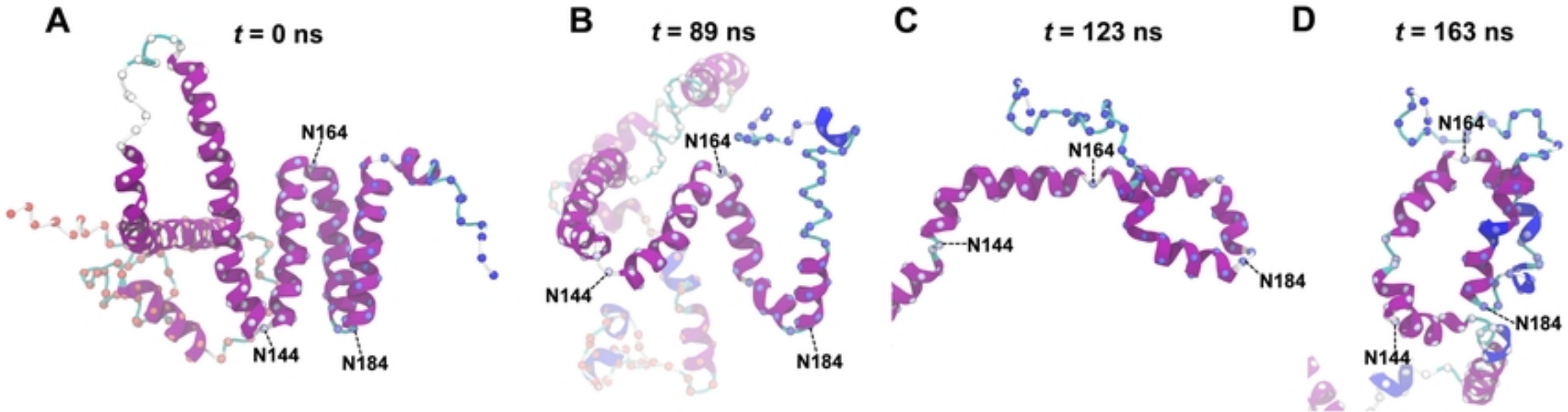
A
t = 0 ns
N164
N144
N184
B
t = 89 ns
N164
N144
N184
C
t = 123 ns
N164
N144
N184
D
t = 163 ns
N164
N184
N144

Figure 12

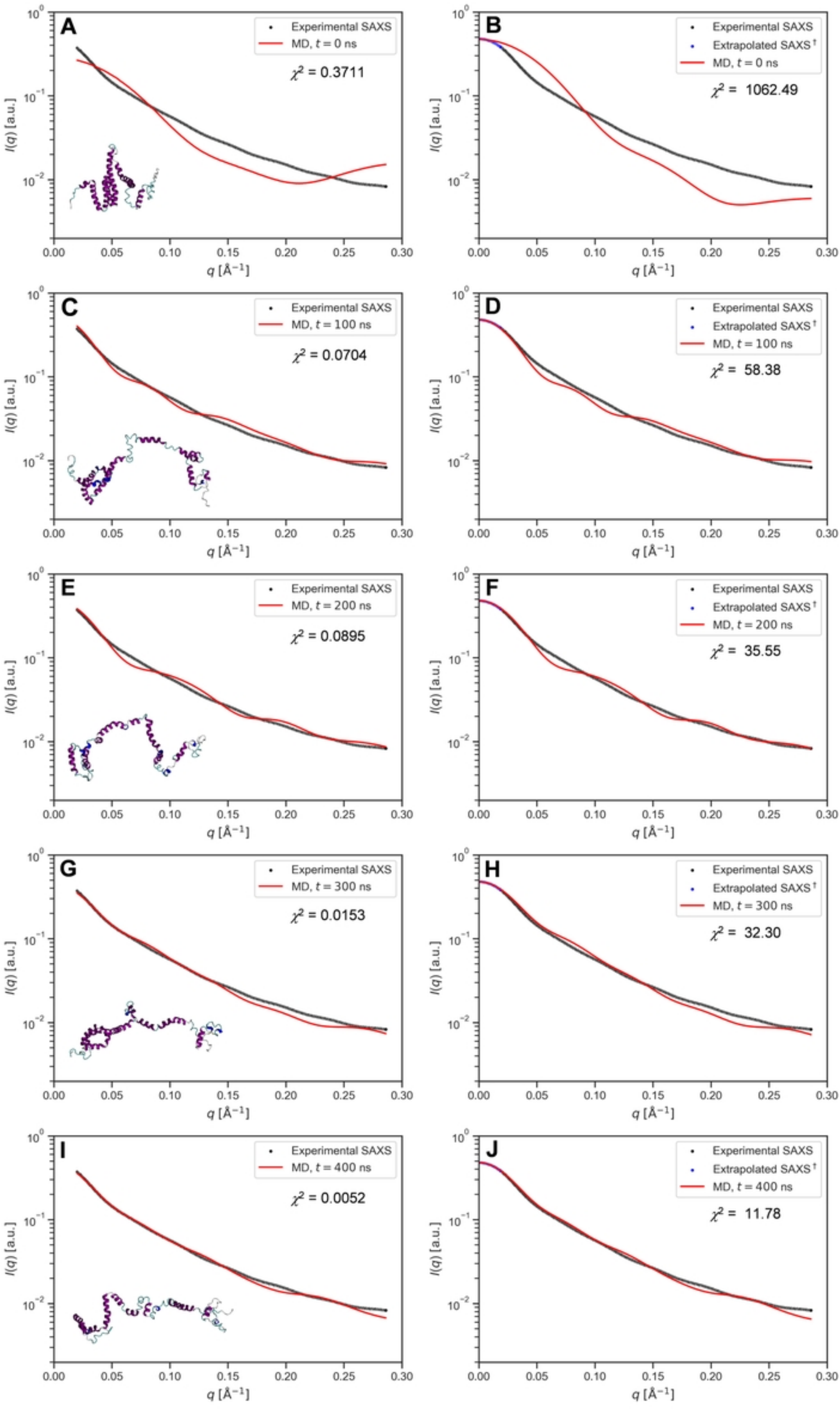
A
Experimental SAXS
MD, t = 0 ns
χ² = 0.3711
B
Experimental SAXS
Extrapolated SAXS†
MD, t = 0 ns
χ² = 1062.49
C
Experimental SAXS
MD, t = 100 ns
χ² = 0.0704
D
Experimental SAXS
Extrapolated SAXS†
MD, t = 100 ns
χ² = 58.38
E
Experimental SAXS
MD, t = 200 ns
χ² = 0.0895
F
Experimental SAXS
Extrapolated SAXS†
MD, t = 200 ns
χ² = 35.55
G
Experimental SAXS
MD, t = 300 ns
χ² = 0.0153
H
Experimental SAXS
Extrapolated SAXS†
MD, t = 300 ns
χ² = 32.30
I
Experimental SAXS
MD, t = 400 ns
χ² = 0.0052
J
Experimental SAXS
Extrapolated SAXS†
MD, t = 400 ns
χ² = 11.78
I(q) [a.u.]
q [Å⁻¹]

Figure 13

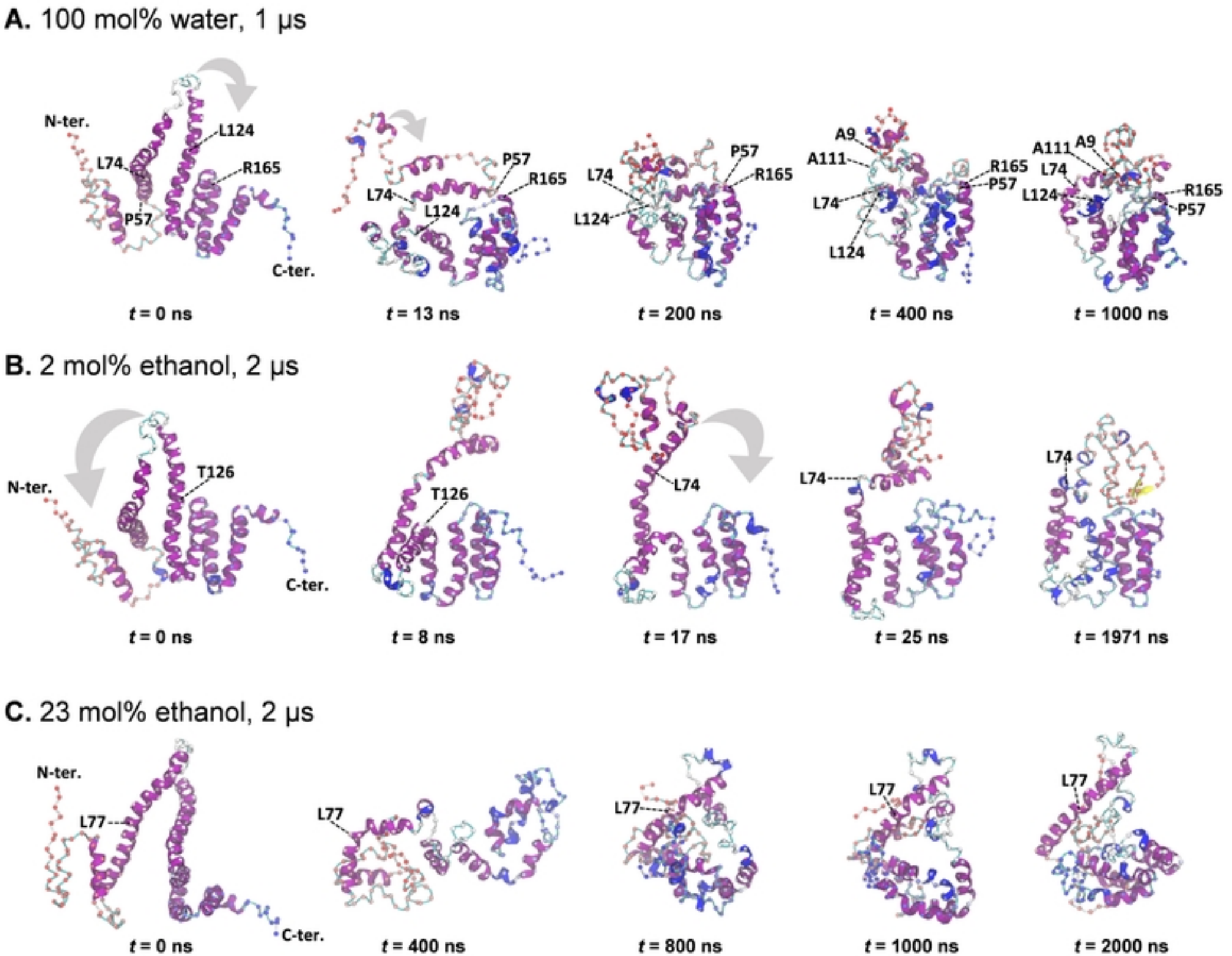
A. 100 mol% water, 1 μs
N-ter.
L124
L74
R165
P57
C-ter.
t = 0 ns
P57
R165
L74
L124
t = 13 ns
P57
R165
L74
L124
t = 200 ns
A9
A111
R165
P57
L74
L124
t = 400 ns
A111
A9
L74
L124
R165
P57
t = 1000 ns
B. 2 mol% ethanol, 2 μs
N-ter.
T126
C-ter.
t = 0 ns
T126
t = 8 ns
L74
t = 17 ns
L74
t = 25 ns
L74
t = 1971 ns
C. 23 mol% ethanol, 2 μs
N-ter.
L77
C-ter.
t = 0 ns
L77
t = 400 ns
L77
t = 800 ns
L77
t = 1000 ns
L77
t = 2000 ns

Figure 14

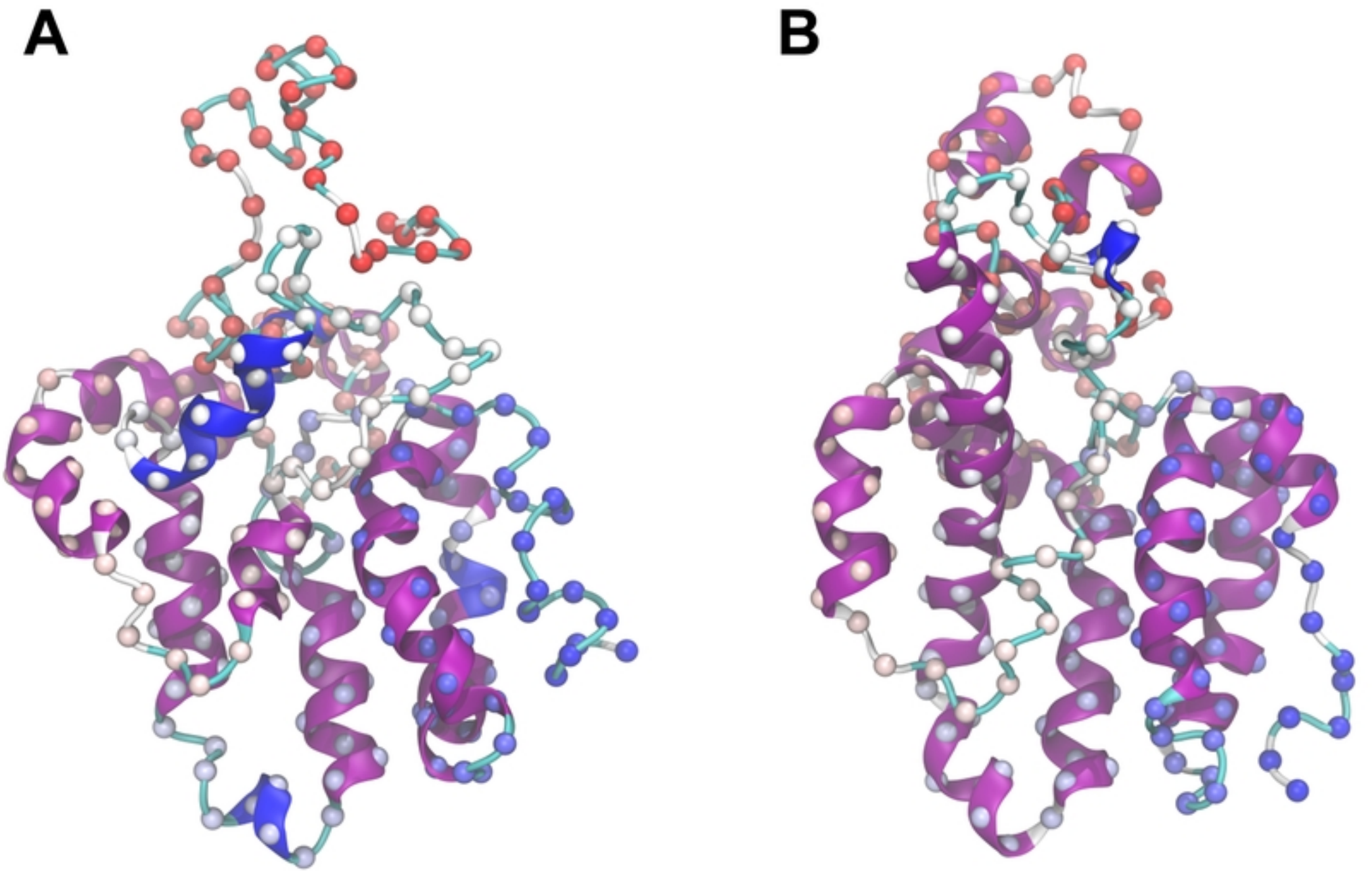

Figure 15

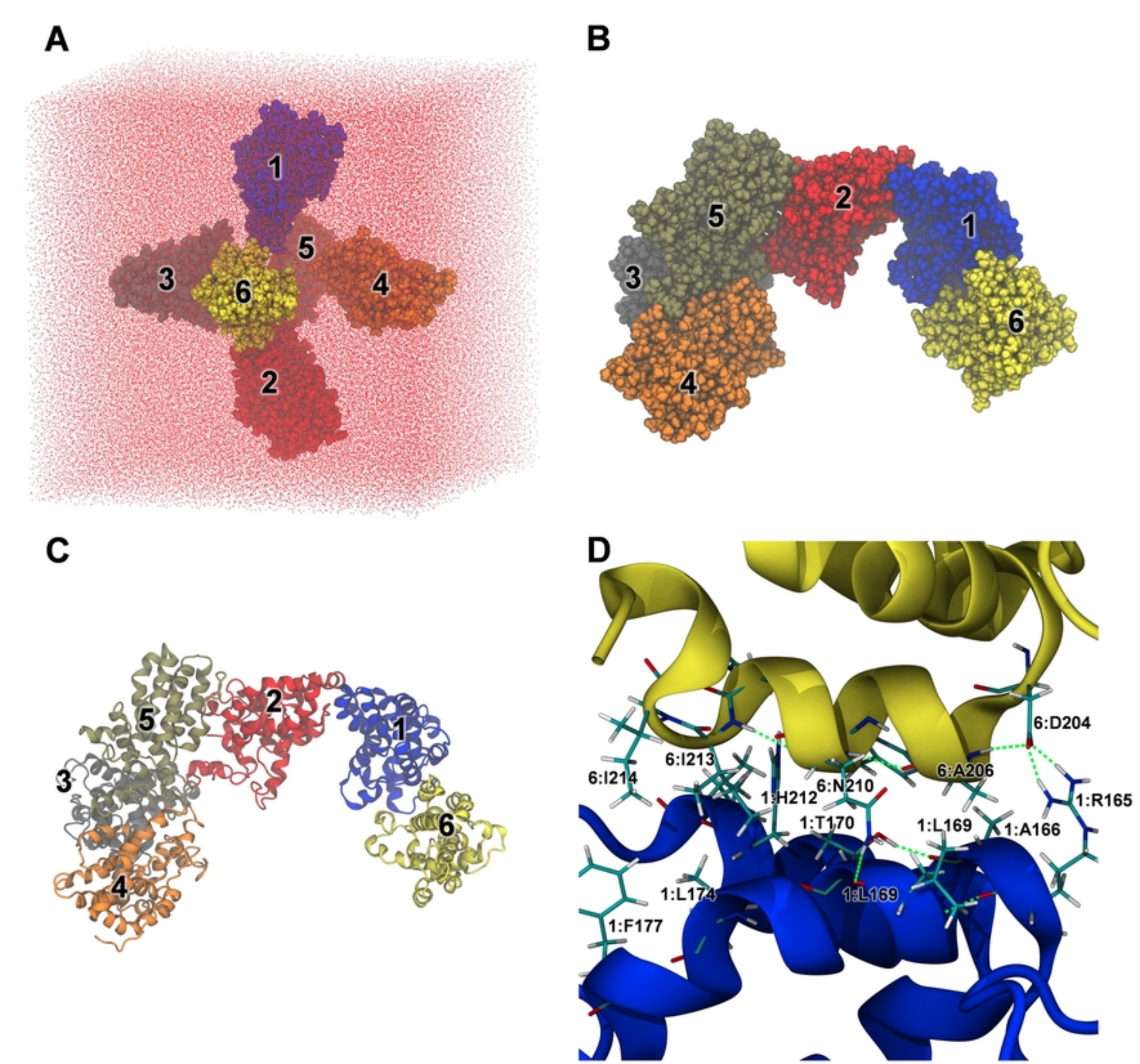
A
1
5
3
6
4
2
B
5
2
1
3
6
4
C
5
2
1
3
6
4
D
6:D204
6:I213
6:I214
6:A206
6:N210
1:H212
1:R165
1:T170
1:L169
1:A166
1:L174
1:L169
1:F177

Figure 16

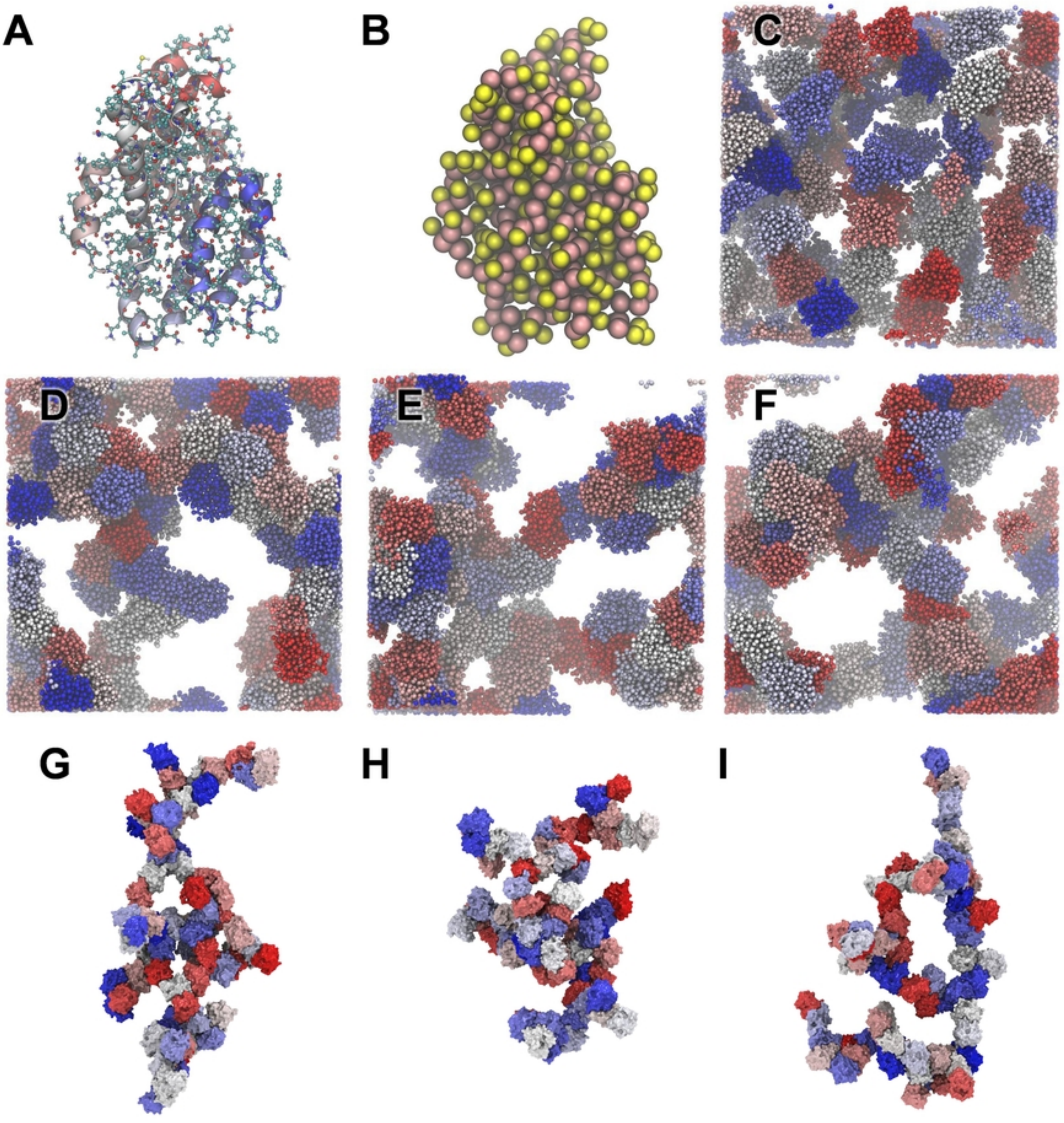

Figure 17

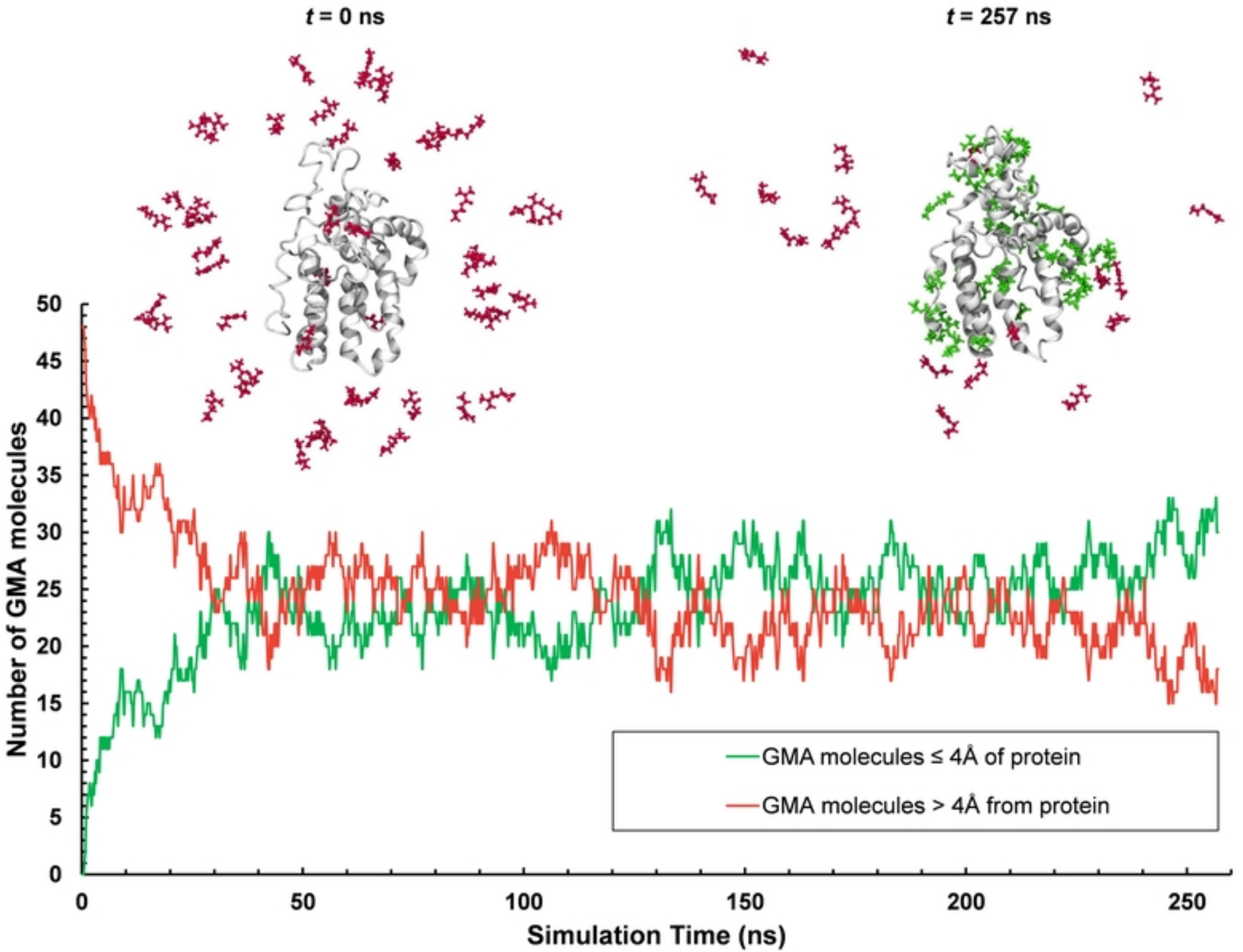
t = 0 ns
t = 257 ns
Number of GMA molecules
Simulation Time (ns)
0
5
10
15
20
25
30
35
40
45
50
0
50
100
150
200
250
GMA molecules ≤ 4Å of protein
GMA molecules > 4Å from protein

Figure 18

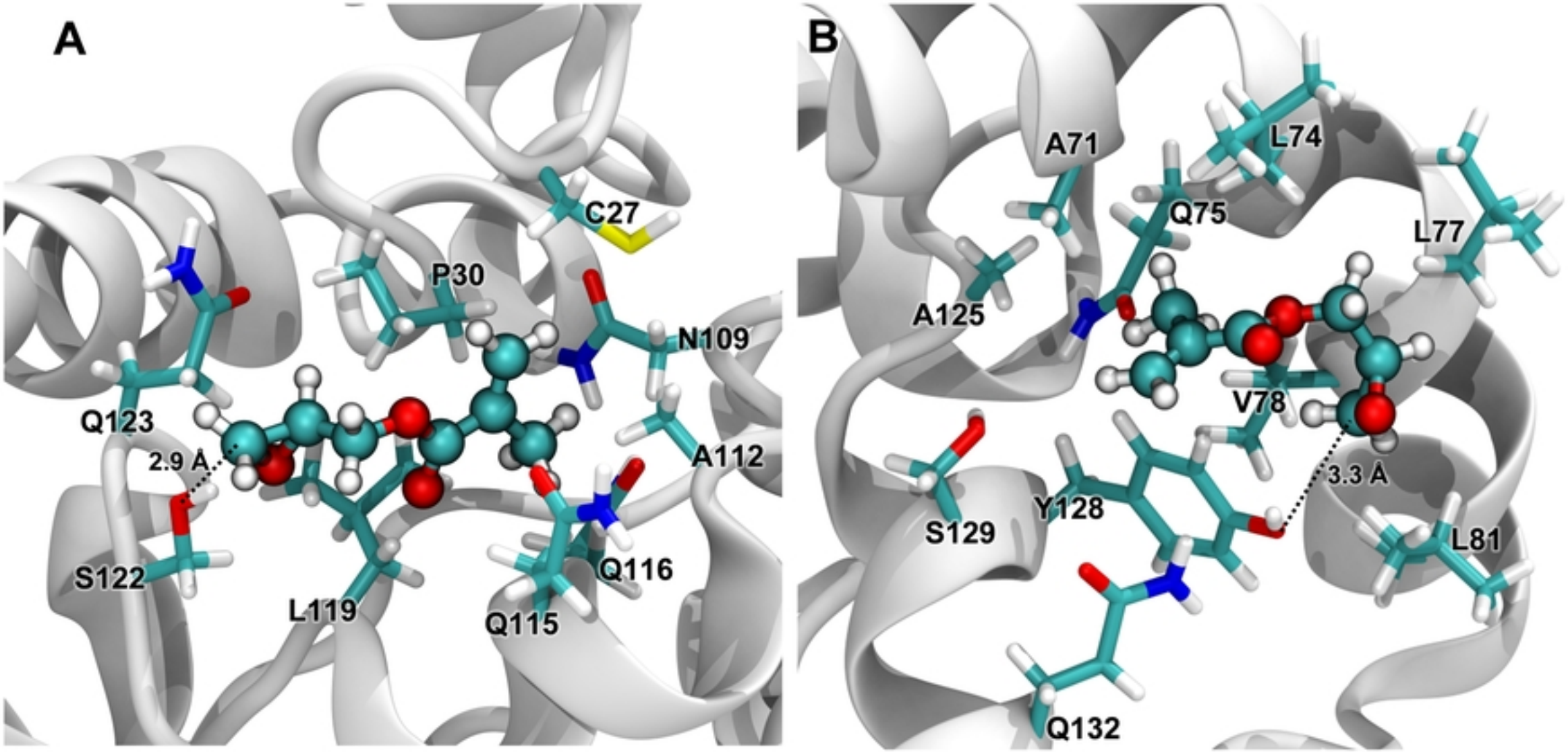
A
C27
P30
N109
Q123
2.9 Å
A112
S122
L119
Q115
Q116
B
A71
L74
Q75
L77
A125
V78
3.3 Å
S129
Y128
L81
Q132

Figure 19