bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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)

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cold plasma treatment [20]. The mechanical properties of zein have been engineered by crosslinking [21, 22]. Recently, excessive aggregation in 4D-printing with zein was prevented by tuning the water content of an ethanol/water supporting bath to 40-75% [16].

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

percentage expressed sequence tags (ESTs) were αz19B1, αz19B3, and αz22z1, see Fig 1. The second-highest expressed clone, αz19B3, corresponds to the cZ19C2 zein examined herein.

Fig 1. Topological tree for α-zein protein sequences. Zein names are from the studies [23, 29, 30]. UniProtKB IDs are given in brackets. Expressed sequence tag (EST) percentages are given for the three most expressed randomly chosen clones in B73 endosperm [30]. The cZ19C2 zein studied here is indicated by a grey background. The tree excludes signal peptides.

Every α-zein protein sequence can be divided into four primary sections: a signal peptide, an N-terminal turn, a variable number of approximately 20-residue homologous repeat units, and a C-terminal turn. This is illustrated in Fig 2 using zein gZ19AB1 as an example [28]. In this article, the numbering of residues does not include the signal peptide.

Fig 2. Overall α-zein sequence organization. As an example, the segmentation of gZ19AB1 [28] is indicated by residue numbers. Numbers in brackets exclude the signal peptide.

There is no general agreement on the definition of the repeat unit and published numbers range from 6 to 9 repeats depending on definition and zein sequence [32]. It has been proposed that Z22 zeins are larger than Z19 zeins due to exactly one extra repeat [32]. However, regardless of repeat definition, they constitute about two-thirds of any α-zein sequence and must largely define the proteins' spatial structure.

Due to the challenges of zein separation [33] most experimental characterizations have been carried out on mixtures of α-zeins in alcoholic solution [34, 35] or in the solid state [36, 37] However, two studies on Z19 zeins offer relevant context for the present work. In the first study, Forato *et al.* [38] studied α-zein extracted from the maize variety BR451 which lacks Z22 expression. Gel experiments

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

showed solely Z19 zeins which were characterized without further purification or sequencing. From solid state FTIR the authors reported α-helix segmentation consistent with ~20 residue repeats and a secondary structure distribution of 46% α-helix, 22% β-sheet, 23% turns, and 13% undetermined structure. An extended structure was inferred from ^{1}H NMR (70% deuterated EtOH) and SAXS (90% EtOH, 3.6 mg zein /mL). The authors proposed a hairpin structure based on αz19D2 (UniProt ID: Q946V7) containing helical segments, sheets, and turns with dimensions of 12 Å × 120 - 130 Å. In the other Z19-focused study, Cabra *et al.* [33] purified a Z19 α-zein from waxy yellow dent corn seeds to give a single MALDI-TOF-MS peak of 24.535 kDa. By sequencing the first 20 amino acids (TIFPQCSQAPIASLLPPYLP) they identified the protein as zein M6 [23] (UniProtKB ID: P04702) using BLAST. The CD secondary structure distribution was 40.0% α-helix, 19.5% β-sheet, 15.4% coils, and 25.1% undetermined secondary structure for the isolated zein (0.08 - 0.25 mg/mL) in 70% (v/v) methanol. CD tertiary structure prediction suggested that the α-helices folded into unspecified compact structures. This differs from the extended structure assignments from SAXS and NMR studies on other α-zein preparations in alcoholic solution [35, 38-40].

Interestingly, a contemporary BLAST search on the 20 first amino acids reported by Cabra *et al.* returned three additional sequence hits from UniProtKB/SwissProt (updated 2022/10/30 https://blast.ncbi.nlm.nih.gov/) including cZ19C2, see S36 Fig. As the calculated molecular weight of cZ19C2 (24.087 kDa) approximates the molecular weight from MALDI-TOF-MS, the spectroscopic findings by Cabra *et al.* may possibly apply to the zein studied here.

Despite decades of research, no experimental atomic-level structures of zeins have been published. Instead, the current understanding of zein molecular structure relies on low-resolution experimental methods such as SAXS and a multitude of structural models. Early idealized "consensus-sequence models" leveraged the α-zein repeat structure: A single ~20 residue consensus sequence represents all repeats and is modelled as a single α-helix, except for terminal residues that

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Methods

AlphaFold2 Model

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

GROMACS All-atom Simulations

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with a single chloride ion. Each protein-solvent box corresponding to a particular ethanol concentration was subjected to the GROMACS simulation protocol outlined below.

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

GROMACS All-atom Simulation Protocol

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Interactions with Glycidyl Methacrylate

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

MD Analysis and Visualization

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

algorithms. Waggawagga [74] and Deepcoil2 [75] were used to scan the sequence for single-α helix (SAH) domains, canonical heptad-repeat coiled-coil domains, and non-canonical coiled coil domains. However, none of these domains were identified.

AlphaFold2 Model

The AlphaFold2 model is shown in Fig 3B. Its loosely arranged structure superficially resembles models from published experimental α-zein characterizations in alcohol/water mixtures, such as Forato *et al*.'s [38] segmented multi-helix model and Tatham *et al*.'s elongated few-helix model [39]. Overall low predicted lDDT [76] scores (S1 Fig) imply high model uncertainty across the sequence, reflecting the absence of α-zein-like structures in the AlphaFold2 training data. The model lacks β-sheets, and has an α-helical content of 68% distributed among seven antiparallel and mostly loosely packed α-helical segments (denoted helix I to VII) of varying length connected by turns or loops. The α-helical segments are initiated by Pro-Ala or, in one case, Arg-Ala. Some conspicuous features of the model include the two central long α-helices (helix II and III), and the bundle of three relatively tightly packed α-helices near the C-terminal (helix IV, V, and VI) which have slightly higher confidence scores, see S1 Fig. The α-helices II and III appear unusually long given their lack of stabilizing interactions with adjacent structure [77],[78] . Interestingly, the *ab initio* method C-QUARK also predicted a long helix for a subset (S56 – A86) of the helix II residues, see Fig B in S2 Appendix.

Fig 3: Consensus sequence alignment and AlphaFold2 model. A: Consensus sequence alignment. Top row (grey background) is the consensus repeat sequence by Geraghty *et al.* [25]. The other rows contain the mature protein sequence of cZ19C2 (UniProtKB ID: P06677) manually aligned with the consensus repeat sequence in segments of up to 20 amino acids. Green background indicates amino

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

acids matching the consensus sequence. Black background indicates gaps. B: AlphaFold2 model. Residues corresponding to the first amino acid of each row in the alignment (A) are indicated by black labels. The α-helices I – VII are annotated in purple. Gln and Asn are shown in green and orange stick representation, respectively. The protein backbone is shown as ribbons coloured by secondary structure. α-carbons are shown as spheres coloured by sequence position (red: N-terminal, blue: C-terminal).

A detailed description of the AlphaFold2 model follows. Starting from the C-terminal, the unstructured tail residues Q210 – F219 are preceded by the short helix VII (P205 – Y209). The preceding residues N144 to D204 span the bundle of the three antiparallel α-helices, helix IV (P145 – A163), helix V (R165 – A182), and helix VI (P205 – Y209) that are laterally associated with each other, forming a mainly hydrophobic core. Interestingly, the sequence of each helix corresponds to segments 9, 10, and 11 of the zein sequence in the alignment to the consensus sequence by Geraghty *et al.* [25], see Fig 3A. Thus Gln-Gln repeats are inside helices and not in turns linking the helices, in contrast to the consensus model by Argos *et al.* [28]. Instead, N144, N164 and N184 reside in the turns and expose their side chains, possibly enabling hydrogen bonding. The C-terminal α-helical bundle is connected to the long helix III (P110-L143) through a hairpin turn containing N144. Comparison of helix III to the consensus sequence alignment in Fig 3A shows that it spans segments 7 and 8. It could have been reasonable to anticipate a break in helix III at the position corresponding to the start of segment 8, in light of the one-to-one segment-to-α-helix mapping for the C-terminal helices IV, V, and VI. However, in contrast to the latter, gaps were used in segments 7 and 8 for alignment with the consensus sequence. The gaps imply that segment 8 lacks residues corresponding to positions 1-3 in the consensus sequence. Since these consensus sequence positions in segments 9 – 11 coincide with the start of helices IV - VI, their lack in segment 8 may explain why AlphaFold2

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

400 ns all-atom MD Simulations

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig 5. Snapshots from 400 ns MD simulations in 2 mol% ethanol. A, D, G: AlphaFold2 structure before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation. C, F, I: Snapshots at 400 ns.

Fig 6. Snapshots from 400 ns MD simulations in 23 mol% ethanol. A, D, G: AlphaFold2 structure before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation. C, F, I: Snapshots at 400 ns.

Fig 7. Snapshots from 400 ns MD simulations in 50 mol% ethanol. A, D, G: AlphaFold2 structure before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each simulation. C, F, I: Snapshots at 400 ns.

Fig 8. Snapshots from 400 ns MD simulations in 100 mol% ethanol. A, D, G: AlphaFold2 structure before simulation. B, E, H: Superimposition of 10 snapshots from the last 100 ns of each 400 ns simulation. C, F, I: Snapshots at 400 ns.

The associated time series for backbone root mean square deviation from the AlphaFold2 structure (RMSD), radius of gyration (R_g) and protein solvent accessible surface area (SASA) are included in S3 Fig. Average values for these time series are shown in Fig 9.

Fig 9. Time series averages for 400 ns MD simulations. A: Backbone RMSD relative to the AlphaFold2 structure. B: Radius of gyration (*R*g). C: Solvent accessible surface area (SASA). D: Total helicity (sum of α- and 3_{10}-helicity). Values for each ethanol concentration were averaged over the last 100 ns of each simulation time series and then over the three simulations (seed 1 - 3) per

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

concentration. Error bars indicate standard deviation across seeds seed 1 - 3. For the associated time series see S3 Fig.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to the final 100 ns of each 400 ns MD trajectory. The resulting plot in Fig 10 depicts RMSD for Cα-atom superposition versus the fraction of the total number of Cα atoms (ϕ).

Fig 10. Least-mobile substructures in the 400 ns MD simulations. The graphs for each simulation show the lowest RMSD of Cα-atoms after superpositions of all subsets of size ϕ (fraction of the total number of Cα atoms). The analysis was carried out on the last 100 ns of each simulation with MDLovoFit [79].

Each fraction corresponds to the lowest attainable RMSD for all possible subsets of size ϕ. The plots show that ~30% of the Cα-atoms (ϕ = 0.3) can be superimposed within a RMSD of 1 Å in the trajectories of all 100 mol% water simulations, all 2 mol% ethanol simulations, and two 23 mol% ethanol simulations (seed 1 and 3). In contrast, only around 10% of the Cα-atoms (ϕ = 0.1) could be superimposed within 1 Å for one 23 mol% ethanol simulation (seed 2), two 50 mol% ethanol simulations (seed 1 and 3), and one 100 mol% ethanol simulation (seed 3). The extensive conformational variation in the two remaining simulations, 100 mol% ethanol (seed 1) and 50 mol% ethanol (seed 2), implied that only a negligible fraction of Cα atoms could be superimposed within ~1 Å RMSD.

Helicity

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Influence of Ethanol Concentration on Helicity

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

β-sheet

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

Comparison with Previous SAXS Experiments

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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-

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

terminal helical assembly, encompassing the lower half of helix III, as well as helices IV, V, VI, and VII, extending approximately halfway around them. Additionally, the first residue of helix II (P57) had approached the first residue of helix V (R165), establishing interactions between these regions enduring throughout the remainder of the simulation. The "stickiness" of this interaction is characteristic of the type of intramolecular interactions observed in the study. From 13 ns to 1000 ns, the overall topology of the structural elements described thus far remained mostly unchanged. However, there was a continued process of condensation yielding denser packing within the lower region of the protein, as evident in Fig 14A. Alongside this condensation, notable rearrangements occurred in other regions of the structure. Particularly, the C-terminal segment (T1 - Q55) and the first residues of helix II (P110, A111) made contact at the top of the C-terminal bundle. This interaction is evident from the proximity of A9 and A111 in Fig 14A at 400 ns, and it persisted throughout the simulation. Further reorganization in the less structured upper part of the structure then took place, as reflected by the RMSD increase at around ~500 ns (S20 Fig A) The rearrangement involved condensation, as evident from comparing the upper part of the structure in the snapshots at 400 ns and 1000 ns in Fig 14 A. This is also seen from reductions in radius of gyration and solvent-accessible surface area (SASA) in S20 Fig B and C. The simulation endpoint at 1000 ns had a globular conformation, characterized by a lower region containing mostly helices and a smaller and less compact upper portion with mostly coil/turn elements. Helicity was maintained in residues Q63-N83 of helix II, except for Q73 which acted as a hinge point.

Fig 14. Snapshots from extended MD simulations. Grey arrows indicate approximate directions of rearrangement. Cα atoms are shown as spheres, coloured by residue position (red: N-terminal, blue: C-terminal). Labelled residues are discussed in the main text. Rotation has been used to highlight features in select snapshots, including the identical initial structures. In the case of 2 mol% ethanol

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

2 mol% Ethanol 2 μs Simulation

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

23 mol% Ethanol 2 µs Simulation

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

Extended Simulations Summary

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Interactions Among Multiple α-zein Monomers

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

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

All-atom Simulations of Multiple Zeins

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

Coarse-grained Simulations of Multiple Zeins

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

simulation snapshots at 375 ns for three simulation replicates in Fig 17 D - I show non-covalent zein aggregates with both branching and linear features, resembling fractal-like aggregates [84]. The aggregates appear consistent with the known capacity of α-zeins to form viscoelastic networks in zein dough systems [85], even though the simulated structures lack β-sheets, which are presumed to have a central role in dough formation [86]. A notable limitation of the coarse-grained simulations is that the initial secondary structure assignment remains fixed throughout the simulations. Thus, further research is needed to explore if there is a connection between these simulated aggregates and actual zein networks.

Fig 17. Coarse grained MD simulations. A: All-atom MD equilibrated structure before coarse graining with heavy atoms and polar hydrogens shown in ball and stick representation. B: Martini coarse-grained structure. C: Initial setup with 73 zein copies in a water box. D-F: Final MD snapshots at 375 ns for simulations started from seed 1, 2, and 3 respectively. G-I: Alternative visualization of the final MD snapshots for seed 1,2, and 3 showing the largest zein cluster across the periodic images of the simulation box. Water and chloride ions are not shown. Each zein monomer has randomly been assigned colors from a blue-red palette.

α-zein Cross-linking

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Cysteine-Cysteine Distances During MD

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

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

Interactions with Glycidyl Methacrylate

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Y128 within 3.3 Å of a GMA epoxide carbon, which may again lead to a pre-reaction complex for functionalization.

Fig 18. Interaction between glycidyl methacrylate (GMA) and α-zein during MD. The inset structures show the initial (t = 0 ns) and final (t = 257 ns) configuration of α-zein (white ribbons) and 48 GMA molecules. Red indicates GMA molecules more than 4 Å from the protein. Green indicates GMA molecules within 4 Å of the protein.

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

Conclusions

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

References

1. Haider TP, Völker C, Kramm J, Landfester K, Wurm FR. Plastics of the Future? The Impact of Biodegradable Polymers on the Environment and on Society. Angewandte Chemie International Edition. 2019;58(1):50-62. doi: https://doi.org/10.1002/anie.201805766.

2. Tortorella S, Maturi M, Vetri Buratti V, Vozzolo G, Locatelli E, Sambri L, et al. Zein as a versatile biopolymer: different shapes for different biomedical applications. RSC Advances. 2021;11(62):39004-26. doi: 10.1039/D1RA07424E.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3. Angelova N, Hunkeler D. Rationalizing the design of polymeric biomaterials. Trends in Biotechnology. 1999;17(10):409-21. doi: https://doi.org/10.1016/S0167-7799(99)01356-6.

4. Mejia CD, Mauer LJ, Hamaker BR. Similarities and differences in secondary structure of viscoelastic polymers of maize α-zein and wheat gluten proteins. Journal of Cereal Science. 2007;45(3):353-9. doi: https://doi.org/10.1016/j.jcs.2006.09.009.

5. Zhang Y, Xu M, Zhang X, Hu Y, Luan G. Application of zein in gluten-free foods: A comprehensive review. Food Research International. 2022;160:111722. doi: https://doi.org/10.1016/j.foodres.2022.111722.

6. Burr B, Burr FA. Zein synthesis in maize endosperm by polyribosomes attached to protein bodies. Proc Natl Acad Sci U S A. 1976;73(2):515-9. Epub 1976/02/01. doi: 10.1073/pnas.73.2.515. PubMed PMID: 1061153; PubMed Central PMCID: PMCPMC335940.

7. Shewry PR, Tatham AS. The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J. 1990;267(1):1-12. Epub 1990/04/01. doi: 10.1042/bj2670001. PubMed PMID: 2183790; PubMed Central PMCID: PMCPMC1131235.

8. Dickey LC, Dallmer MF, Radewonuk ER, Parris N, Kurantz M, Craig JC. Zein Batch Extraction from Dry-Milled Corn: Cereal Disintegration by Dissolving Fluid Shear. Cereal Chemistry. 1998;75(4):443-8. doi: https://doi.org/10.1094/CCHEM.1998.75.4.443.

9. Corradini E, Curti PS, Meniqueti AB, Martins AF, Rubira AF, Muniz EC. Recent Advances in Food-Packing, Pharmaceutical and Biomedical Applications of Zein and Zein-Based Materials. International Journal of Molecular Sciences. 2014;15(12):22438-70. PubMed PMID: doi:10.3390/ijms151222438.

10. de Vries A, Nikiforidis CV, Scholten E. Natural amphiphilic proteins as tri-block Janus particles: Self-sorting into thermo-responsive gels. Europhysics Letters. 2014;107(5):58003. doi: 10.1209/0295-5075/107/58003.

11. Wang Y, Padua GW. Nanoscale Characterization of Zein Self-Assembly. Langmuir. 2012;28(5):2429-35. doi: 10.1021/la204204j.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Weissmueller NT, Lu HD, Hurley A, Prud'homme RK. Nanocarriers from GRAS Zein Proteins to Encapsulate Hydrophobic Actives. Biomacromolecules. 2016;17(11):3828-37. doi: 10.1021/acs.biomac.6b01440.

13. Chaunier L, Leroy E, Della Valle G, Lourdin D. Plasticized protein for 3D printing by fused deposition modeling. AIP Conference Proceedings. 2016;1769(1):190001. doi: 10.1063/1.4963611.

14. Chaunier L, Dalgalarrondo M, Della Valle G, Lourdin D, Marion D, Leroy E. Rheological characterization of plasticized corn proteins for fused deposition modeling. AIP Conference Proceedings. 2017;1896(1):040002. doi: 10.1063/1.5008028.

15. Tavares-Negrete JA, Aceves-Colin AE, Rivera-Flores DC, Díaz-Armas GG, Mertgen A-S, Trinidad-Calderón PA, et al. Three-Dimensional Printing Using a Maize Protein: Zein-Based Inks in Biomedical Applications. ACS Biomaterials Science & Engineering. 2021;7(8):3964-79. doi: 10.1021/acsbiomaterials.1c00544.

16. Zhang Y, Raza A, Xue Y-Q, Yang G, Hayat U, Yu J, et al. Water-responsive 4D printing based on self-assembly of hydrophobic protein "Zein" for the control of degradation rate and drug release. Bioactive Materials. 2023;23:343-52. doi: https://doi.org/10.1016/j.bioactmat.2022.11.009.

17. Luecha J, Hsiao A, Brodsky S, Liu GL, Kokini JL. Green microfluidic devices made of corn proteins. Lab on a Chip. 2011;11(20):3419-25. doi: 10.1039/C1LC20726A.

18. Mattice KD, Marangoni AG. Functionalizing zein through antisolvent precipitation from ethanol or aetic acid. Food Chemistry. 2020;313:126127. doi: https://doi.org/10.1016/j.foodchem.2019.126127.

19. Taylor JRN, Johnson SK, Taylor J, Njila S, Jackaman C. Oxidation of commercial (α-type) zein with hydrogen peroxide improves its hydration and dramatically increases dough extensibility even below its glass transition temperature. Journal of Cereal Science. 2016;70:108-15. doi: https://doi.org/10.1016/j.jcs.2016.05.025.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

20. Chen G, Chen Y, Jin N, Li J, Dong S, Li S, et al. Zein films with porous polylactic acid coatings via cold plasma pre-treatment. Industrial Crops and Products. 2020;150:112382. doi: https://doi.org/10.1016/j.indcrop.2020.112382.

21. Mattice KD, Marangoni AG. Physical properties of zein networks treated with microbial transglutaminase. Food Chemistry. 2021;338:128010. doi: https://doi.org/10.1016/j.foodchem.2020.128010.

22. Kim S, Sessa DJ, Lawton JW. Characterization of zein modified with a mild cross-linking agent. Industrial Crops and Products. 2004;20(3):291-300. doi: https://doi.org/10.1016/j.indcrop.2003.10.013.

23. Viotti A, Cairo G, Vitale A, Sala E. Each zein gene class can produce polypeptides of different sizes. The EMBO Journal. 1985;4(5):1103-10. doi: https://doi.org/10.1002/j.1460-2075.1985.tb03746.x.

24. Geraghty D, Peifer MA, Rubenstein I, Messing J. The primary structure of a plant storage protein: zein. Nucleic Acids Research. 1981;9(19):5163-74. doi: 10.1093/nar/9.19.5163.

25. Geraghty DE, Messing J, Rubenstein I. Sequence analysis and comparison of cDNAs of the zein multigene family. Embo j. 1982;1(11):1329-35. Epub 1982/01/01. doi: 10.1002/j.1460-2075.1982.tb01318.x. PubMed PMID: 6897917; PubMed Central PMCID: PMCPMC553212.

26. Pedersen K, Devereux J, Wilson DR, Sheldon E, Larkins BA. Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell. 1982;29(3):1015-26. doi: https://doi.org/10.1016/0092-8674(82)90465-2.

27. Marks MD, Larkins BA. Analysis of sequence microheterogeneity among zein messenger RNAs. Journal of Biological Chemistry. 1982;257(17):9976-83. doi: https://doi.org/10.1016/S0021-9258(18)33973-5.

28. Argos P, Pedersen K, Marks MD, Larkins BA. A structural model for maize zein proteins. The Journal of biological chemistry. 1982;257 17:9984-90.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

29. Marks MD, Lindell JS, Larkins BA. Nucleotide sequence analysis of zein mRNAs from maize endosperm. Journal of Biological Chemistry. 1985;260(30):16451-9. doi: 10.1016/S0021-9258(17)36258-0.

30. Woo Y-M, Hu DW-N, Larkins BA, Jung R. Genomics Analysis of Genes Expressed in Maize Endosperm Identifies Novel Seed Proteins and Clarifies Patterns of Zein Gene Expression. The Plant Cell. 2001;13(10):2297-317. doi: 10.1105/tpc.010240.

31. Spena A, Viotti A, Pirrotta V. A homologous repetitive block structure underlies the heterogeneity of heavy and light chain zein genes. The EMBO Journal. 1982;1(12):1589-94. doi: https://doi.org/10.1002/j.1460-2075.1982.tb01360.x.

32. Garratt R, Oliva G, Caracelli I, Leite A, Arruda P. Studies of the zein-like α-prolamins based on an analysis of amino acid sequences: Implications for their evolution and three-dimensional structure. Proteins: Structure, Function, and Bioinformatics. 1993;15(1):88-99. doi: https://doi.org/10.1002/prot.340150111.

33. Cabra V, Arreguin R, Galvez A, Quirasco M, Vazquez-duhalt R, Farres A. Characterization of a 19 kDa α-Zein of High Purity. Journal of Agricultural and Food Chemistry. 2005;53(3):725-9. doi: 10.1021/jf048530s.

34. Uzun S, Ilavsky J, Padua GW. Characterization of zein assemblies by ultra-small-angle X-ray scattering. Soft Matter. 2017;13(16):3053-60. doi: 10.1039/C6SM02717B.

35. Bugs MR, Forato LA, Bortoleto-Bugs RK, Fischer H, Mascarenhas YP, Ward RJ, et al. Spectroscopic characterization and structural modeling of prolamin from maize and pearl millet. European Biophysics Journal. 2004;33(4):335-43. doi: 10.1007/s00249-003-0354-3.

36. Forato LA, Bicudo TDC, Colnago LA. Conformation of α zeins in solid state by Fourier transform IR. Biopolymers. 2003;72(6):421-6. doi: https://doi.org/10.1002/bip.10481.

37. Akhmetova A, Lanno GM, Kogermann K, Malmsten M, Rades T, Heinz A. Highly Elastic and Water Stable Zein Microfibers as a Potential Drug Delivery System for Wound Healing. Pharmaceutics. 2020;12(5). Epub 2020/05/24. doi: 10.3390/pharmaceutics12050458. PubMed PMID: 32443445; PubMed Central PMCID: PMCPMC7284525.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

38. Forato LA, Doriguetto AC, Fischer H, Mascarenhas YP, Craievich AF, Colnago LA. Conformation of the Z19 Prolamin by FTIR, NMR, and SAXS. Journal of Agricultural and Food Chemistry. 2004;52(8):2382-5. doi: 10.1021/jf035020+.

39. Tatham AS, Field JM, Morris VJ, I'Anson KJ, Cardle L, Dufton MJ, et al. Solution conformational analysis of the alpha-zein proteins of maize. Journal of Biological Chemistry. 1993;268(35):26253-9. doi: https://doi.org/10.1016/S0021-9258(19)74308-7.

40. Matsushima N, Danno G-i, Takezawa H, Izumi Y. Three-dimensional structure of maize α-zein proteins studied by small-angle X-ray scattering. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology. 1997;1339(1):14-22. doi: https://doi.org/10.1016/S0167-4838(96)00212-9.

41. Momany FA, Sessa DJ, Lawton JW, Selling GW, Hamaker SA, Willett JL. Structural characterization of alpha-zein. J Agric Food Chem. 2006;54(2):543-7. Epub 2006/01/19. doi: 10.1021/jf058135h. PubMed PMID: 16417318.

42. Tan H, Zhou H, Guo T, Zhang Y, Ma L. Integrated multi-spectroscopic and molecular modeling techniques to study the formation mechanism of hidden zearalenone in maize. Food Chemistry. 2021;351:129286. doi: https://doi.org/10.1016/j.foodchem.2021.129286.

43. Liu C, Lv N, Ren G, Wu R, Wang B, Cao Z, et al. Explore the interaction mechanism between zein and EGCG using multi-spectroscopy and molecular dynamics simulation methods. Food Hydrocolloids. 2021;120:106906. doi: https://doi.org/10.1016/j.foodhyd.2021.106906.

44. Arooj M, Shehadi I, Nassab CN, Mohamed AA. Physicochemical stability study of protein-benzoic acid complexes using molecular dynamics simulations. Amino Acids. 2020;52(9):1353-62. Epub 2020/10/03. doi: 10.1007/s00726-020-02897-2. PubMed PMID: 33006112.

45. Hajjari MM, Sharif N. In-silico behavior of dissolved prolamins under electric field effect applied by electrospinning process using molecular dynamics simulation. Journal of Molecular Liquids. 2021;344:117778. doi: https://doi.org/10.1016/j.molliq.2021.117778.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

46. Hajjari MM, Sharif N. In-silico zein/tannic acid colloidal nanoparticles and their activity at oil and water interface of Pickering emulsion using molecular dynamics simulation. Journal of Molecular Liquids. 2022;359:119321. doi: https://doi.org/10.1016/j.molliq.2022.119321.

47. Erickson DP, Ozturk OK, Selling G, Chen F, Campanella OH, Hamaker BR. Corn zein undergoes conformational changes to higher β-sheet content during its self-assembly in an increasingly hydrophilic solvent. International Journal of Biological Macromolecules. 2020;157:232-9. doi: https://doi.org/10.1016/j.ijbiomac.2020.04.169.

48. Erickson DP, Dunbar M, Hamed E, Ozturk OK, Campanella OH, Keten S, et al. Atomistic Modeling of Peptide Aggregation and β-Sheet Structuring in Corn Zein for Viscoelasticity. Biomacromolecules. 2021;22(5):1856-66. doi: 10.1021/acs.biomac.0c01558.

49. Lawton JW. Zein: A History of Processing and Use. Cereal Chemistry. 2002;79(1):1-18. doi: https://doi.org/10.1094/CCHEM.2002.79.1.1.

50. Selling GW, Hamaker SAH, Sessa DJ. Effect of Solvent and Temperature on Secondary and Tertiary Structure of Zein by Circular Dichroism. Cereal Chemistry. 2007;84(3):265-70. doi: https://doi.org/10.1094/CCHEM-84-3-0265.

51. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. doi: 10.1038/s41586-021-03819-2.

52. The UniProt C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. 2021;49(D1):D480-D9. doi: 10.1093/nar/gkaa1100.

53. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nature Methods. 2022;19(6):679-82. doi: 10.1038/s41592-022-01488-1.

54. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996;14(1):33-8. doi: https://doi.org/10.1016/0263-7855(96)00018-5.

55. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal of Chemical Theory and Computation. 2008;4(1):116-22. doi: 10.1021/ct700200b.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

56. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013;29(7):845-54. doi: 10.1093/bioinformatics/btt055.

57. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. The Journal of Physical Chemistry B. 2001;105(28):6474-87. doi: 10.1021/jp003919d.

58. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. The Journal of Physical Chemistry. 1987;91(24):6269-71. doi: 10.1021/j100308a038.

59. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. Journal of Computational Chemistry. 2004;25(9):1157-74. doi: https://doi.org/10.1002/jcc.20035.

60. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983;79(2):926-35. doi: 10.1063/1.445869.

61. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. The Journal of Physical Chemistry B. 2009;113(26):9004-15. doi: 10.1021/jp901540t.

62. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function, and Bioinformatics. 2010;78(8):1950-8. doi: https://doi.org/10.1002/prot.22711.

63. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al., editors. Scalable algorithms for molecular dynamics simulations on commodity clusters. SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing; 2006: IEEE.

64. Schrödinger Release 2018-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2018. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2018.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

65. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. Journal of Chemical Theory and Computation. 2016;12(1):281-96. doi: 10.1021/acs.jctc.5b00864.

66. Schrödinger Release 2018-1: Maestro, Schrödinger, LLC, New York, NY, 2018.

67. de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schäfer LV, et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. Journal of Chemical Theory and Computation. 2013;9(1):687-97. doi: 10.1021/ct300646g.

68. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637. Epub 1983/12/01. doi: 10.1002/bip.360221211. PubMed PMID: 6667333.

69. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Research. 2004;32(suppl_2):W500-W2. doi: 10.1093/nar/gkh429.

70. Postu PA, Ion L, Drochioiu G, Petre BA, Glocker MO. Mass spectrometric characterization of the zein protein composition in maize flour extracts upon protein separation by SDS-PAGE and 2D gel electrophoresis. ELECTROPHORESIS. 2019;40(20):2747-58. doi: https://doi.org/10.1002/elps.201900108.

71. Shen M, Broeckling CD, Chu EY, Ziegler G, Baxter IR, Prenni JE, et al. Leveraging Non-Targeted Metabolite Profiling via Statistical Genomics. PLOS ONE. 2013;8(2):e57667. doi: 10.1371/journal.pone.0057667.

72. Sharma S, Pradhan R, Manickavasagan A, Thimmanagari M, Dutta A. Exploration of corn distillers solubles from selective milling technology as a novel source of plant-based ACE inhibitory protein hydrolysates. Food Chemistry. 2022;388:133036. doi: https://doi.org/10.1016/j.foodchem.2022.133036.

73. Santos-Rosales V, Alvarez-Rivera G, Hillgärtner M, Cifuentes A, Itskov M, García-González CA, et al. Stability Studies of Starch Aerogel Formulations for Biomedical Applications. Biomacromolecules. 2020;21(12):5336-44. doi: 10.1021/acs.biomac.0c01414.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

74. Simm D, Hatje K, Kollmar M. Waggawagga: comparative visualization of coiled-coil predictions and detection of stable single α-helices (SAH domains). Bioinformatics. 2015;31(5):767-9. doi: 10.1093/bioinformatics/btu700.

75. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics. 2019;35(16):2790-5. doi: 10.1093/bioinformatics/bty1062.

76. Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013;29(21):2722-8. Epub 2013/08/30. doi: 10.1093/bioinformatics/btt473. PubMed PMID: 23986568; PubMed Central PMCID: PMCPMC3799472.

77. Barnes CA, Shen Y, Ying J, Takagi Y, Torchia DA, Sellers JR, et al. Remarkable Rigidity of the Single α-Helical Domain of Myosin-VI As Revealed by NMR Spectroscopy. J Am Chem Soc. 2019;141(22):9004-17. Epub 2019/05/24. doi: 10.1021/jacs.9b03116. PubMed PMID: 31117653; PubMed Central PMCID: PMCPMC6556874.

78. Simm D, Hatje K, Kollmar M. Distribution and evolution of stable single α-helices (SAH domains) in myosin motor proteins. PLoS One. 2017;12(4):e0174639. Epub 2017/04/04. doi: 10.1371/journal.pone.0174639. PubMed PMID: 28369123; PubMed Central PMCID: PMCPMC5378345 ONE policies on sharing data and materials.

79. Martínez L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. PLOS ONE. 2015;10(3):e0119264. doi: 10.1371/journal.pone.0119264.

80. Kretschmer CB. Infrared Spectroscopy and Optical Rotatory Dispersion of Zein, Wheat Gluten and Gliadin. The Journal of Physical Chemistry. 1957;61(12):1627-31. doi: 10.1021/j150558a014.

81. Piiadov V, Ares de Araújo E, Oliveira Neto M, Craievich AF, Polikarpov I. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

data measured on a relative scale. Protein Science. 2019;28(2):454-63. doi: https://doi.org/10.1002/pro.3528.

82. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J. 2013;105(4):962-74. Epub 2013/08/27. doi: 10.1016/j.bpj.2013.07.020. PubMed PMID: 23972848; PubMed Central PMCID: PMCPMC3752106.

83. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research. 2007;35(suppl_2):W407-W10. doi: 10.1093/nar/gkm290.

84. Filippov AV, Zurita M, Rosner DE. Fractal-like Aggregates: Relation between Morphology and Physical Properties. Journal of Colloid and Interface Science. 2000;229(1):261-73. doi: https://doi.org/10.1006/jcis.2000.7027.

85. Zhang X, Dong C, Hu Y, Gao M, Luan G. Zein as a structural protein in gluten-free systems: an overview. Food Science and Human Wellness. 2021;10(3):270-7. doi: https://doi.org/10.1016/j.fshw.2021.02.018.

86. Sadat A, Corradini MG, Joye IJ. Vibrational and fluorescence spectroscopy to study gluten and zein interactions in complex dough systems. Curr Res Food Sci. 2022;5:479-90. Epub 2022/03/11. doi: 10.1016/j.crfs.2022.02.009. PubMed PMID: 35265856; PubMed Central PMCID: PMCPMC8899121.

87. Gomes MC, Mano JF. Chemical modification strategies to prepare advanced protein-based biomaterials. Biomaterials and Biosystems. 2021;1:100010. doi: https://doi.org/10.1016/j.bbiosy.2021.100010.

88. Mukherjee S, Kapp EA, Lothian A, Roberts AM, Vasil'ev YV, Boughton BA, et al. Characterization and Identification of Dityrosine Cross-Linked Peptides Using Tandem Mass Spectrometry. Analytical Chemistry. 2017;89(11):6136-45. doi: 10.1021/acs.analchem.7b00941.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

89. Liu C, Hua J, Ng PF, Fei B. Photochemistry of bioinspired dityrosine crosslinking. Journal of Materials Science & Technology. 2021;63:182-91. doi: https://doi.org/10.1016/j.jmst.2020.02.086.

90. Moon SH, Hwang HJ, Jeon HR, Park SJ, Bae IS, Yang YJ. Photocrosslinkable natural polymers in tissue engineering. Front Bioeng Biotechnol. 2023;11:1127757. Epub 2023/03/28. doi: 10.3389/fbioe.2023.1127757. PubMed PMID: 36970625; PubMed Central PMCID: PMCPMC10037533.

91. Allasia M, Sonzogni AS, Vaillard VA, Vaillard SE, Gugliotta LM, Minari RJ. Methacrylated zein as a novel biobased macro-crosslinker for PVCL hydrogels. Polymer. 2021;235:124278. doi: https://doi.org/10.1016/j.polymer.2021.124278.

Supporting information

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S4 Fig. Time series for secondary structure for 100 mol% water 400 ns MD simulations. Values given in the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S5 Fig. Time series for secondary structure for 2 mol% ethanol 400 ns MD simulations. Values given in the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S6 Fig. Time series for secondary structure for 23 mol% ethanol 400 ns MD simulations. Values given in the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S7 Fig. Time series for secondary structure for 50 mol% ethanol 400 ns MD simulations. Values given in the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S8 Fig. Time series for secondary structure for 100 mol% ethanol 400 ns MD simulations. Values given in the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S9 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 100 mol% water 400 ns MD simulations. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2, and 3) 100 mol% water 400 ns MD simulations.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S10 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 2 mol% ethanol 400 ns MD simulations. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2, and 3) 2 mol% ethanol 400 ns MD simulations.

S11 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 23 mol% ethanol 400 ns MD simulations. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2, and 3) 23 mol% ethanol 400 ns MD simulations.

S12 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 50 mol% ethanol 400 ns MD simulations. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2, and 3) 50 mol% ethanol 400 ns MD simulations.

S13 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 100 mol% ethanol 400 ns MD simulations. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the three (Seed 1, 2, and 3) 100 mol% ethanol 400 ns MD simulations.

S14 Fig. Principal Component Analysis of Helicity Profiles from 400 ns MD simulations. Left: Scores for the first two principal components. Right: Loadings for the two first principal components. The PCA was carried out on helicity profiles obtained from the last 100 ns of each 400 ns MD simulations by averaging the helicity per residue across the last 100 snapshots. Simulations are color

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coded based on ethanol concentration. The suffixes #1, #2, and #3 indicate different simulation replicates (seeds).

S15 Fig. Transient β-sheet formation in 400 ns ethanol simulations. A: Seed 1, B: Seed 2. The snapshots are from t = 346 ns in both cases.

S16 Fig. Distance between sulfur atoms in the two cysteines across all 400 ns MD trajectories. Mean and standard deviations in this plot are calculated using the entire 400 ns range.

S17 Fig. SAXS predictions. Predictions for 2 mol% ethanol showing the best fit (t = 2 ns) and the fit to the last MD frame (t = 400 ns).

S18 Fig. SAXS predictions. Predictions for 50 mol% ethanol showing the two best fits at t = 161 ns and t = 160 ns.

S19 Fig. SAXS predictions. Predictions for 100 mol% ethanol showing the two best fits at t = 234 ns and t = 400 ns.

S20 Fig. Time series for extended GROMACS all-atom MD simulations. A, D, G: Backbone RMSD. B, E, H: Radius of gyration. C, F, I: SASA. Values given in the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S21 Fig. Time series for secondary structure for extended GROMACS all-atom MD simulations. Top: 100 mol% water simulation extended to 1000 ns. Middle: 2 mol% ethanol

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the legend for each seed are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S28 Fig. MD time series for an extended GROMACS all-atom MD simulation with the ff99SB*-ILDN force field. A: Backbone RMSD. B: Radius of gyration. C: SASA. The value given in the legend is an average over the last 100 ns, with standard deviation given in brackets.

S29 Fig. Time series for secondary structure for 400 ns GROMACS all-atom MD simulations with the ff99SB*-ILDN force field. Top: 100 mol% water simulation (seed 1). Middle: 100 mol% water simulation (seed 1) extended to 800 ns. Bottom: 100 mol% water simulation (seed 2). Values given in the legend are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S30 Fig. Time series for secondary structure for 400 ns GROMACS all-atom MD simulations with the ff99SB*-ILDN force field. Top: 24 mol% ethanol simulation (seed 1). Bottom: 24 mol% ethanol simulation (seed 2). Values given in the legend are averages over the last 100 ns of each simulation, with standard deviations given in brackets.

S31 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over the last 100 ns of the 100 mol% water 400 ns MD simulation with the ff99SB*-ILDN force field. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity averaged over the last 100 ns of the 100 mol% water 400 ns Seed 1 simulation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S32 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over the last 100 ns of the 100 mol% water 800 ns MD simulation with the ff99SB*-ILDN force field. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity averaged over the last 100 ns of the 100 mol% water Seed 1 MD simulation extended to 800 ns.

S33 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 100 mol% water 400 ns MD simulations with the ff99SB*-ILDN force field. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the two (Seed 1, Seed 2) 100 mol% water 400 ns MD simulations.

S34 Fig. Helicity per residue (sum of 3_{10}- and α-helicity assigned by STRIDE) in the initial AlphaFold2 model and averaged over 24 mol% ethanol 400 ns MD simulations with the ff99SB*-ILDN force field. Orange: Helicity in the initial AlphaFold2 model. Blue: Helicity of the last 100 ns averaged over the two (Seed 1, Seed 2) 24 mol% ethanol 400 ns MD simulations.

S35 Fig. Secondary structure and backbone RMSD for all-atom MD simulations of 6 zein copies in a water box.

S36 Fig. Mature protein sequence alignment and identity matrix for P06677 (19 kDa zein 19C2), P04703 (19 kDa zein A20), P04702 (19 kDa zein M6), P06676 (19 kDa zein 19C1).

S1 Appendix. Overview of major α-zein sequences

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S2 Appendix. Homology and *ab initio* models

S3 Appendix. Additional simulation details

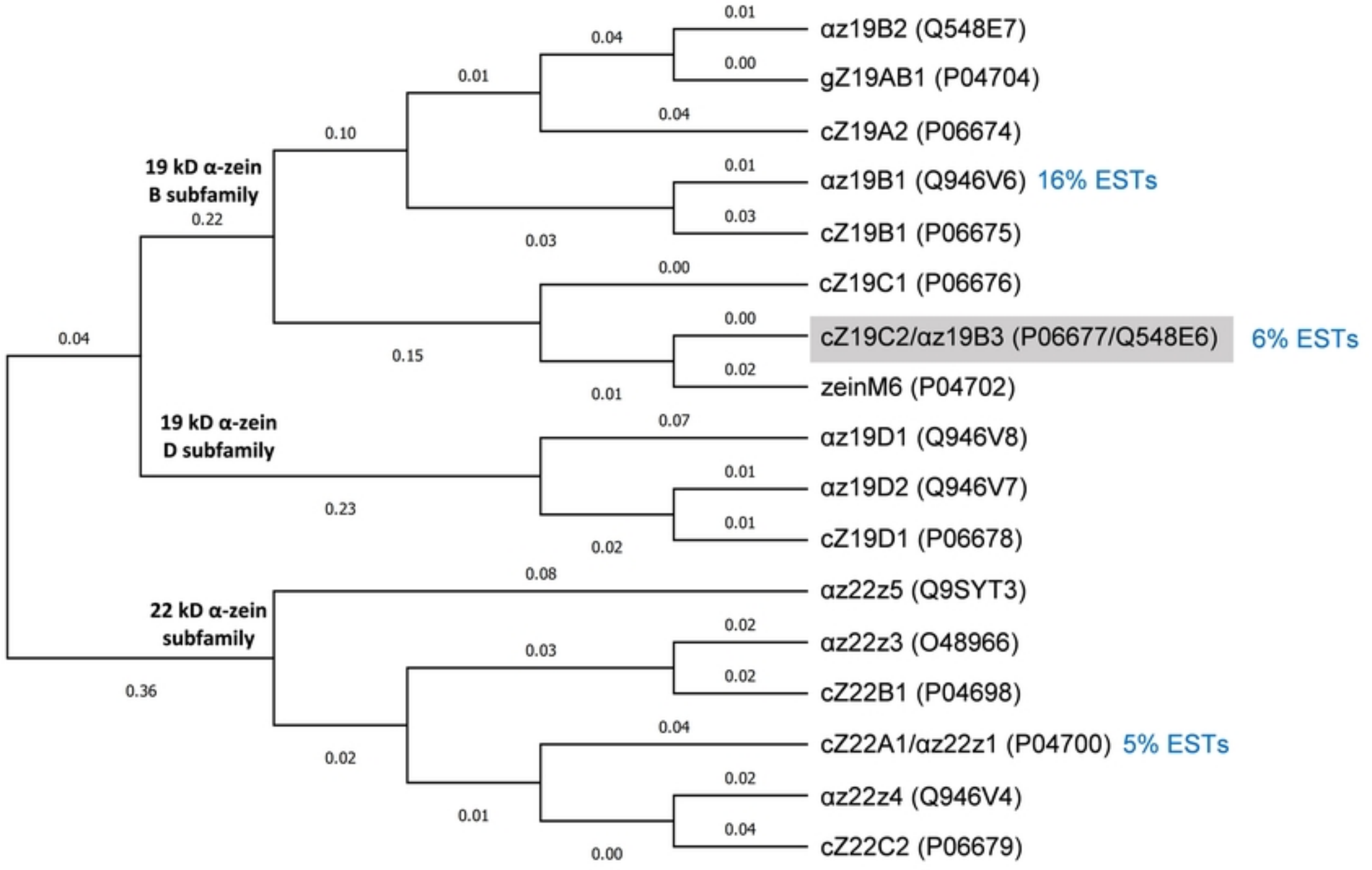

αz19B2 (Q548E7)
gZ19AB1 (P04704)
cZ19A2 (P06674)
αz19B1 (Q946V6) 16% ESTs
cZ19B1 (P06675)
cZ19C1 (P06676)
cZ19C2/αz19B3 (P06677/Q548E6) 6% ESTs
zeinM6 (P04702)
αz19D1 (Q946V8)
αz19D2 (Q946V7)
cZ19D1 (P06678)
αz22z5 (Q9SYT3)
αz22z3 (O48966)
cZ22B1 (P04698)
cZ22A1/αz22z1 (P04700) 5% ESTs
αz22z4 (Q946V4)
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

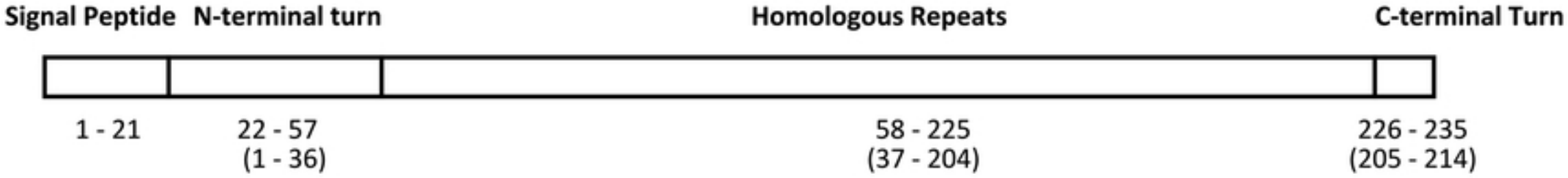

Signal Peptide
N-terminal turn
Homologous Repeats
C-terminal Turn
1 - 21
22 - 57
(1 - 36)
58 - 225
(37 - 204)
226 - 235
(205 - 214)

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

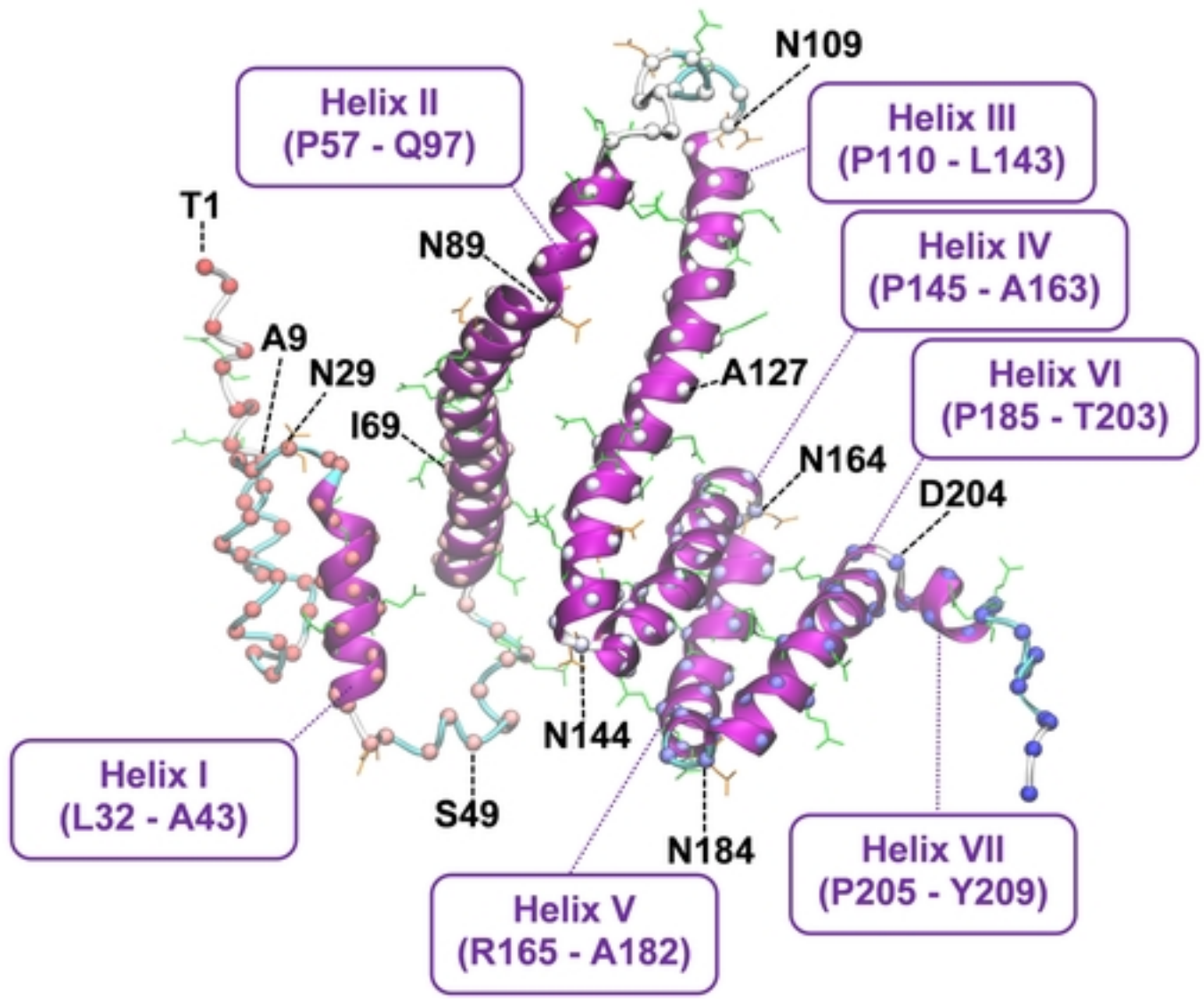


Figure 3

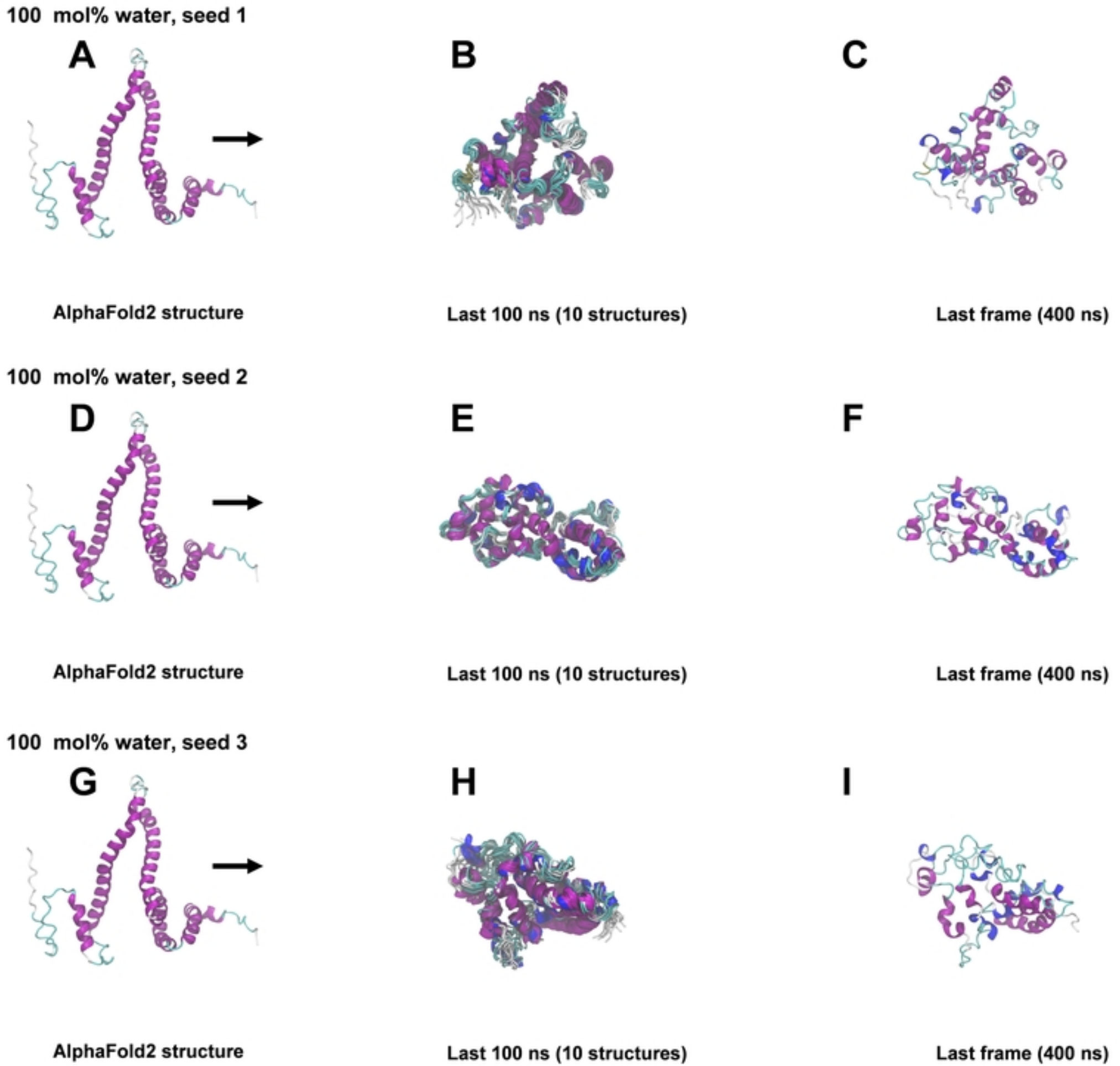

100 mol% water, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% water, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% water, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 4

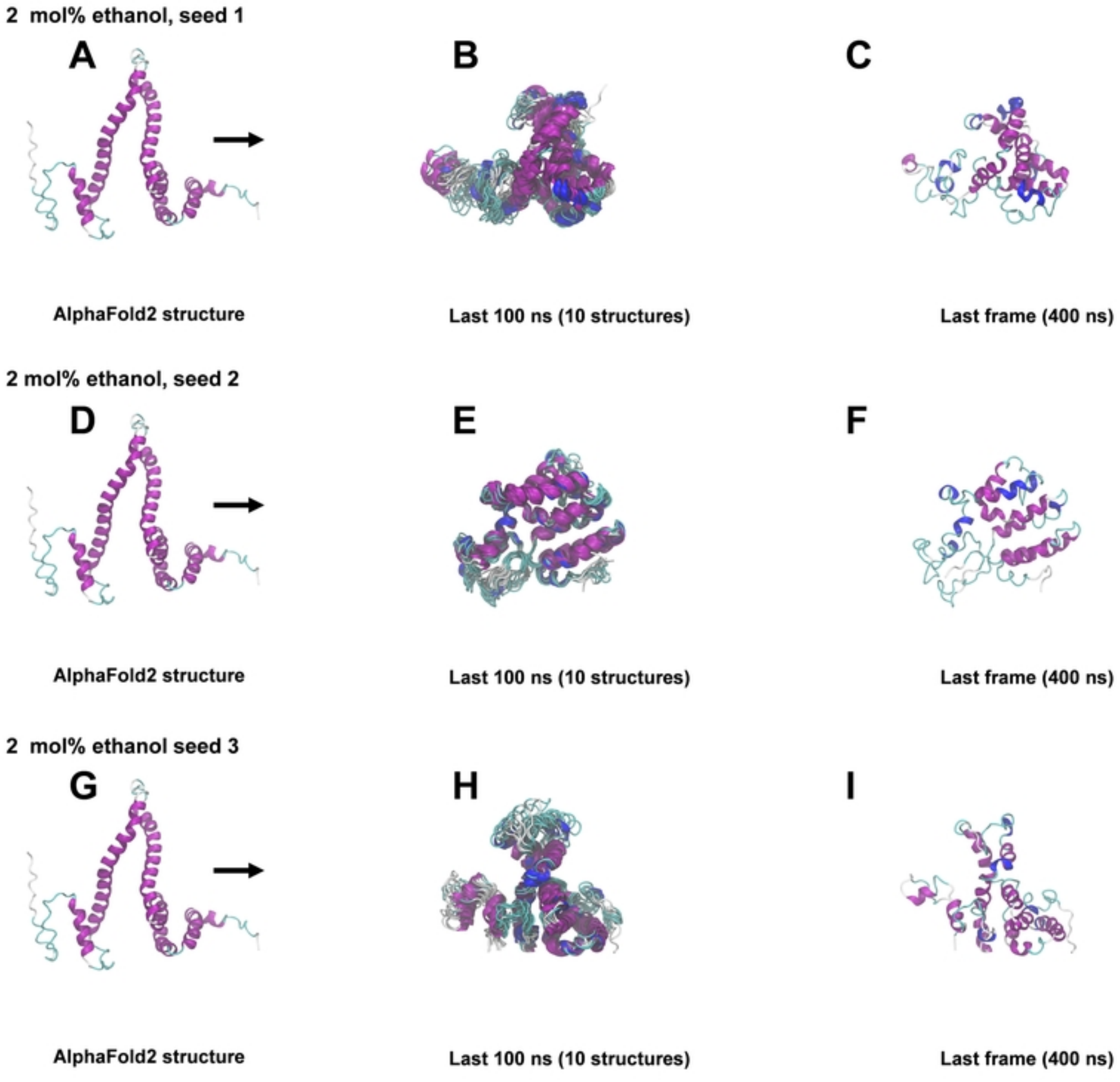

2 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
2 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
2 mol% ethanol seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 5

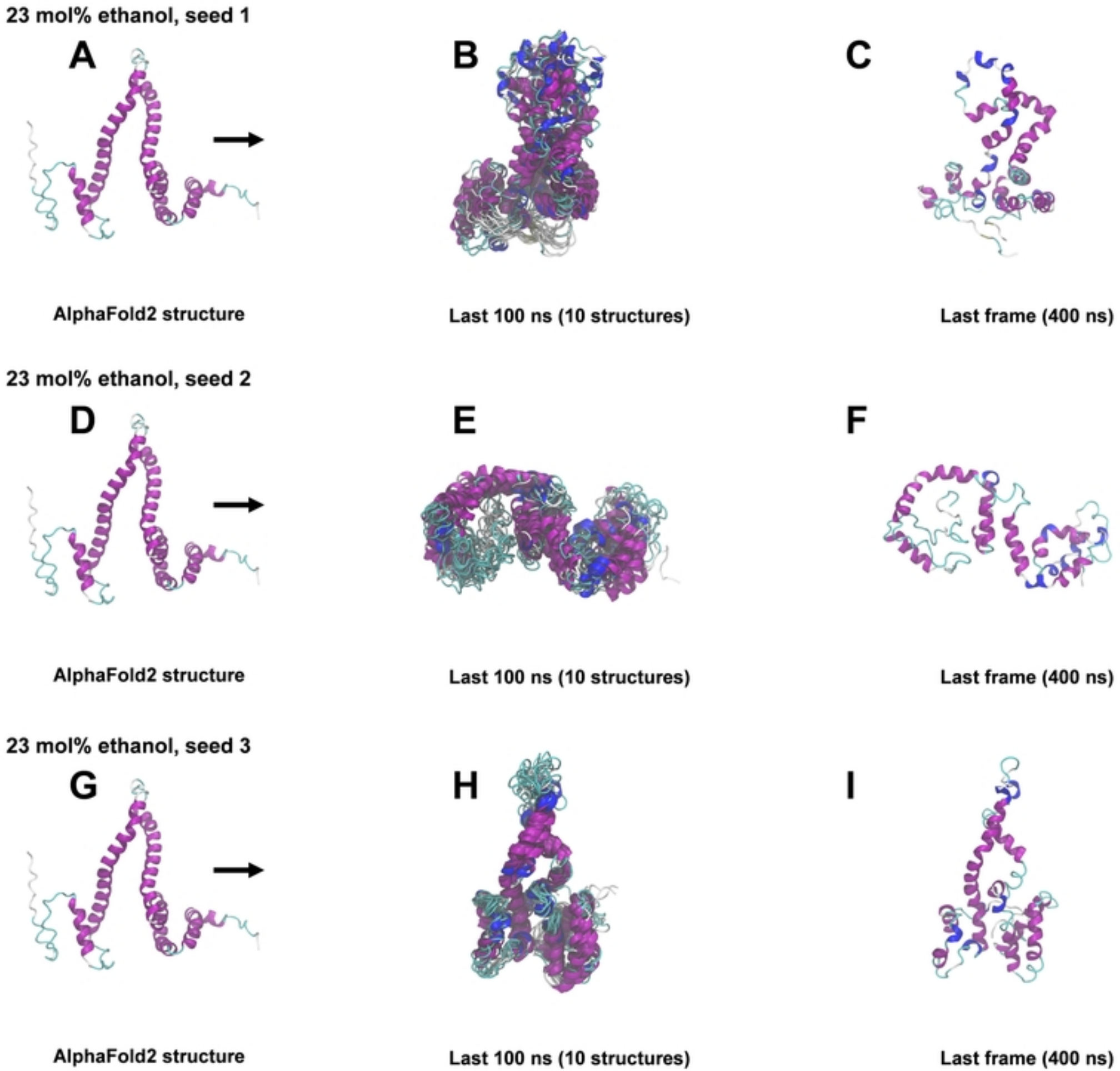

23 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
23 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
23 mol% ethanol, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 6

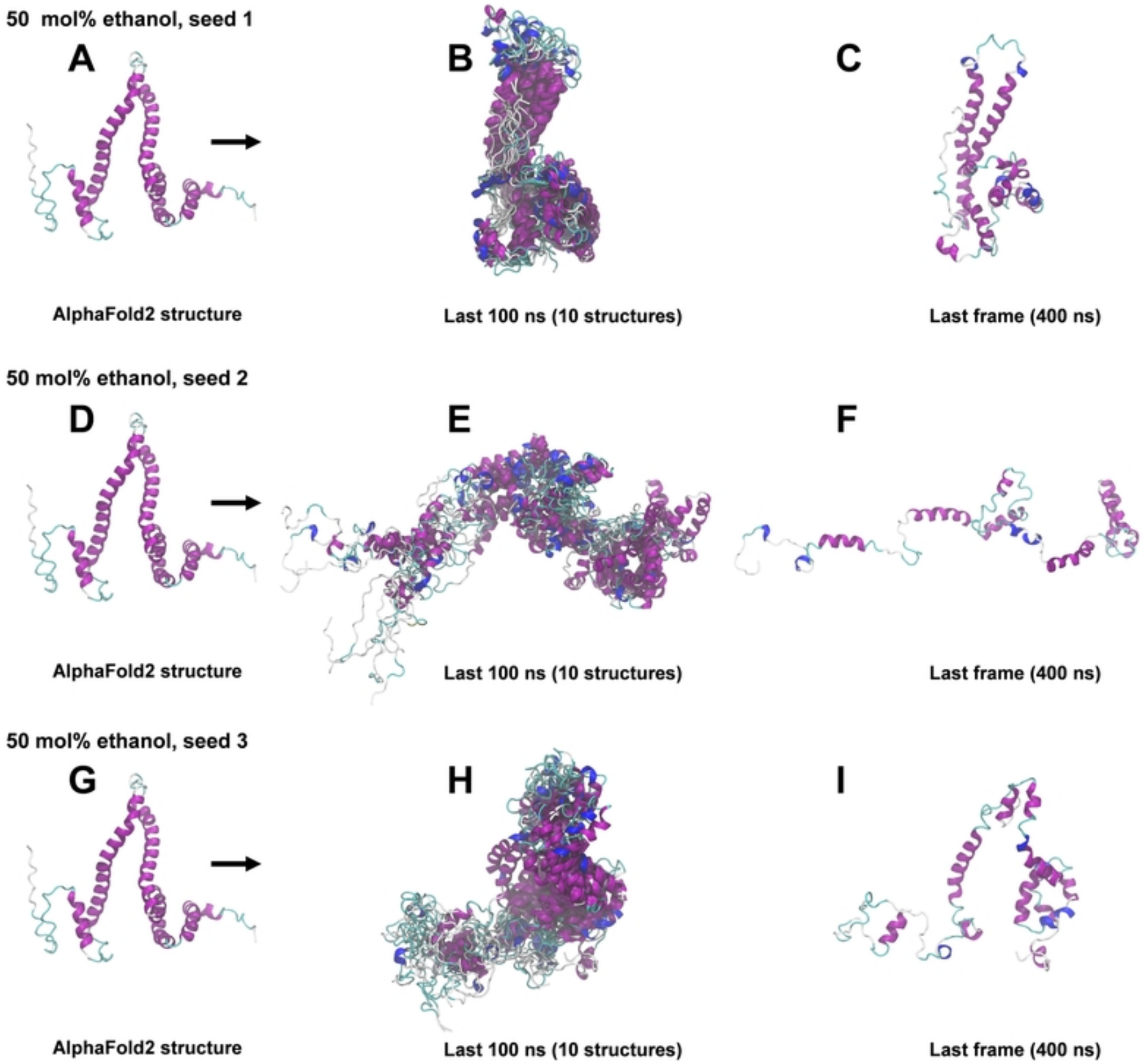

50 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
50 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
50 mol% ethanol, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 7

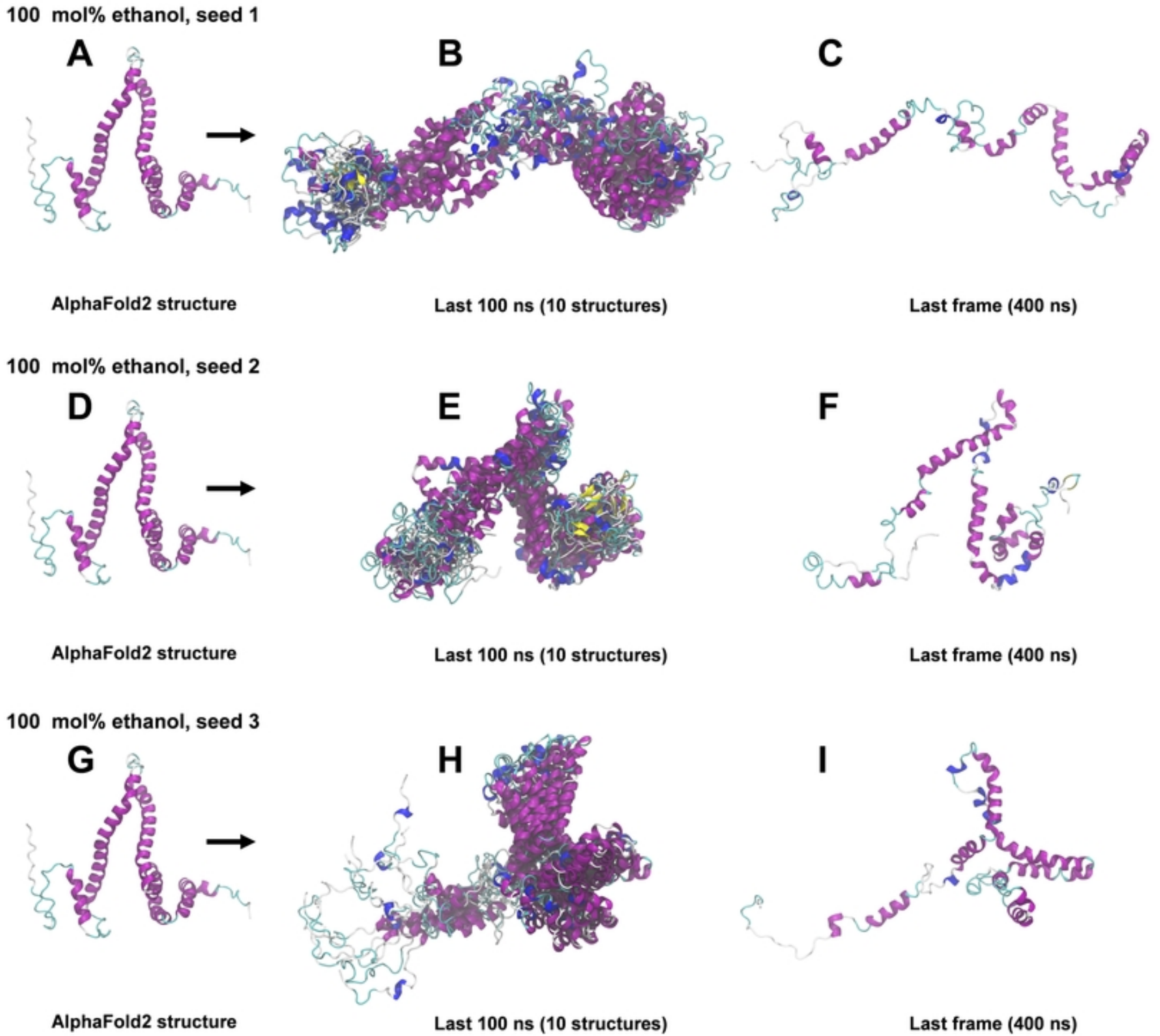

100 mol% ethanol, seed 1
A
B
C
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% ethanol, seed 2
D
E
F
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)
100 mol% ethanol, seed 3
G
H
I
AlphaFold2 structure
Last 100 ns (10 structures)
Last frame (400 ns)

Figure 8

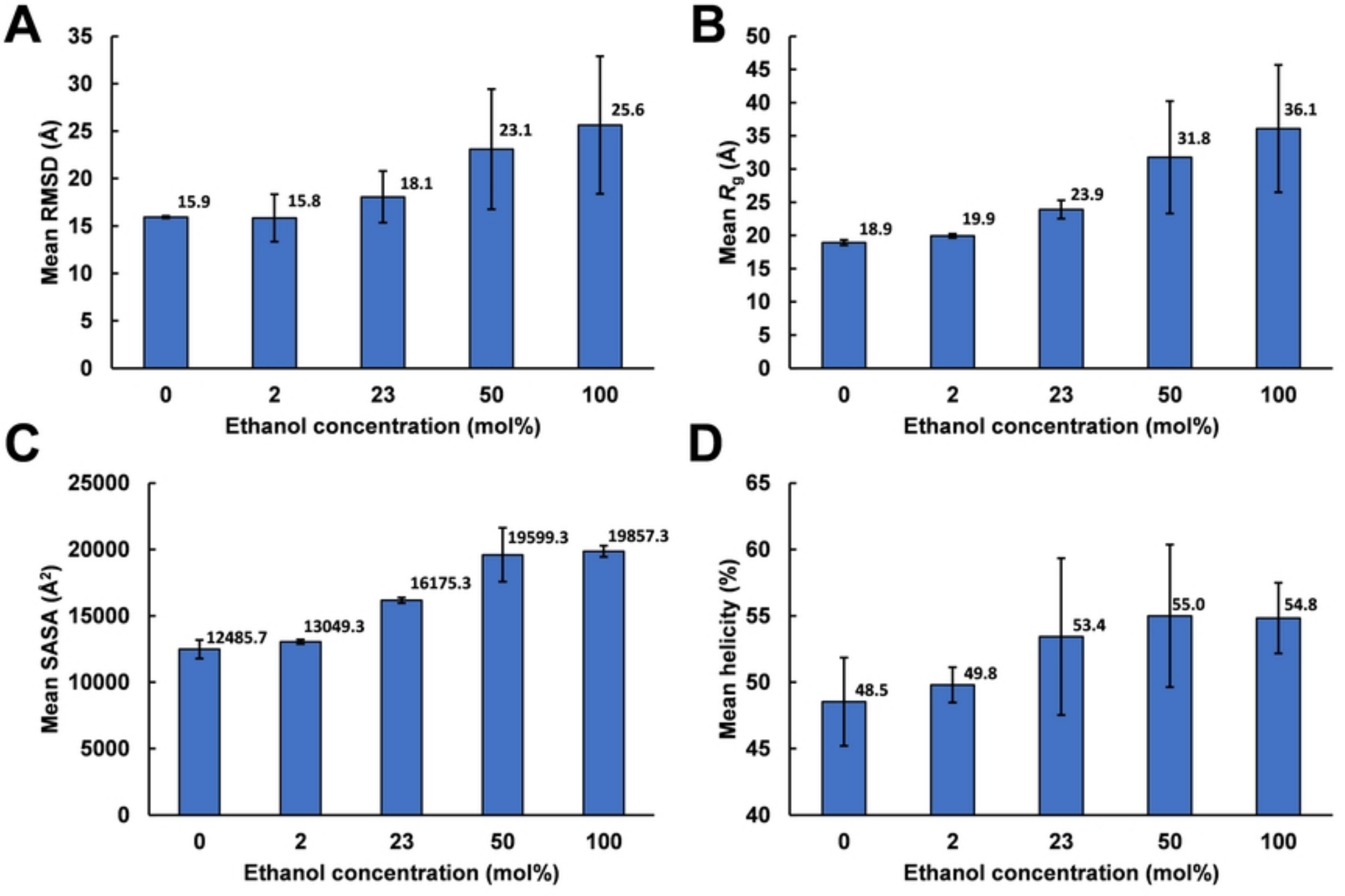

A
Mean RMSD (Å)
35
30
25
20
15
10
5
0
15.9
15.8
18.1
23.1
25.6
0
2
23
50
100
Ethanol concentration (mol%)
B
Mean R_g (Å)
50
45
40
35
30
25
20
15
10
5
0
18.9
19.9
23.9
31.8
36.1
0
2
23
50
100
Ethanol concentration (mol%)
C
Mean SASA (Å²)
25000
20000
15000
10000
5000
0
12485.7
13049.3
16175.3
19599.3
19857.3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

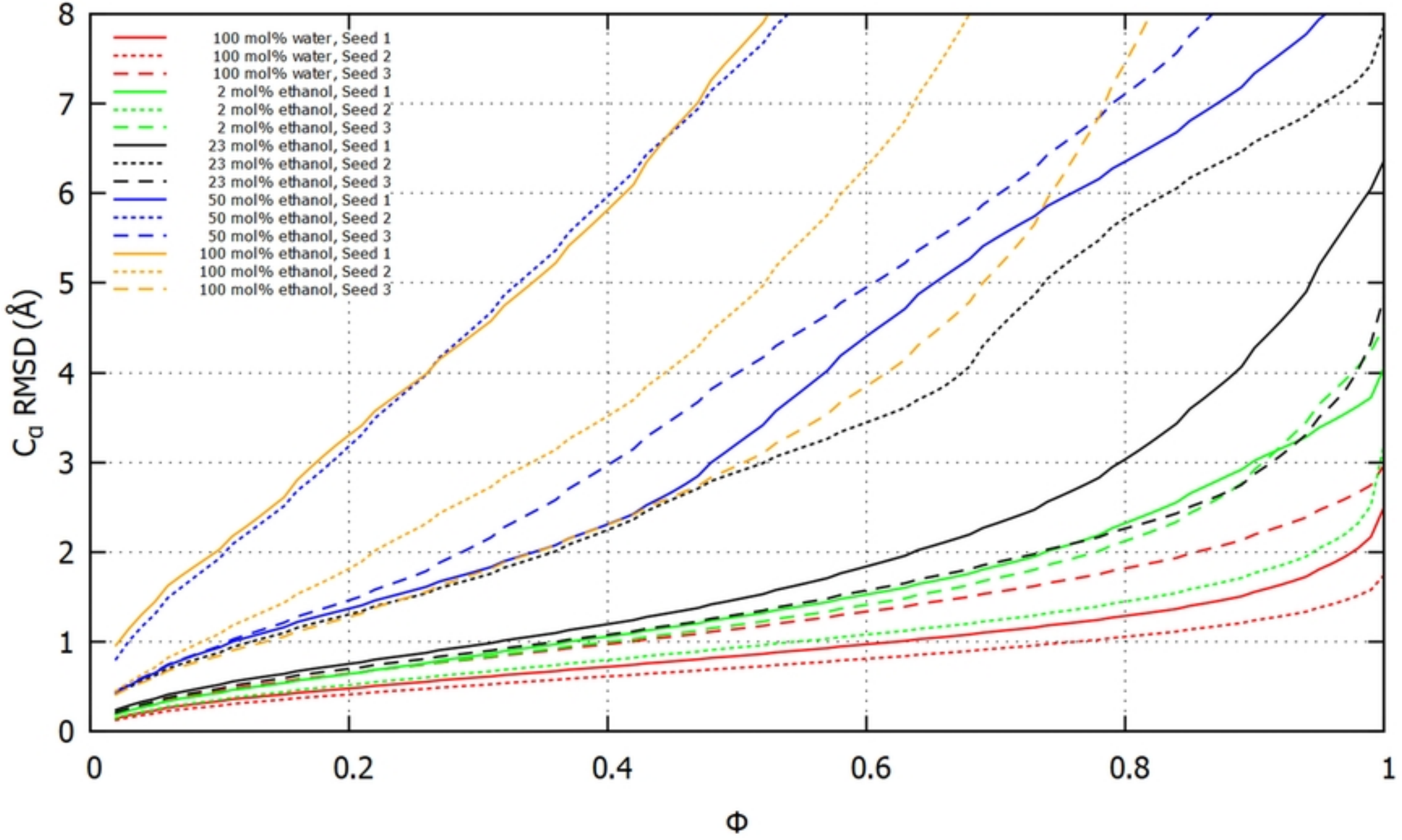

100 mol% water, Seed 1
100 mol% water, Seed 2
100 mol% water, Seed 3
2 mol% ethanol, Seed 1
2 mol% ethanol, Seed 2
2 mol% ethanol, Seed 3
23 mol% ethanol, Seed 1
23 mol% ethanol, Seed 2
23 mol% ethanol, Seed 3
50 mol% ethanol, Seed 1
50 mol% ethanol, Seed 2
50 mol% ethanol, Seed 3
100 mol% ethanol, Seed 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

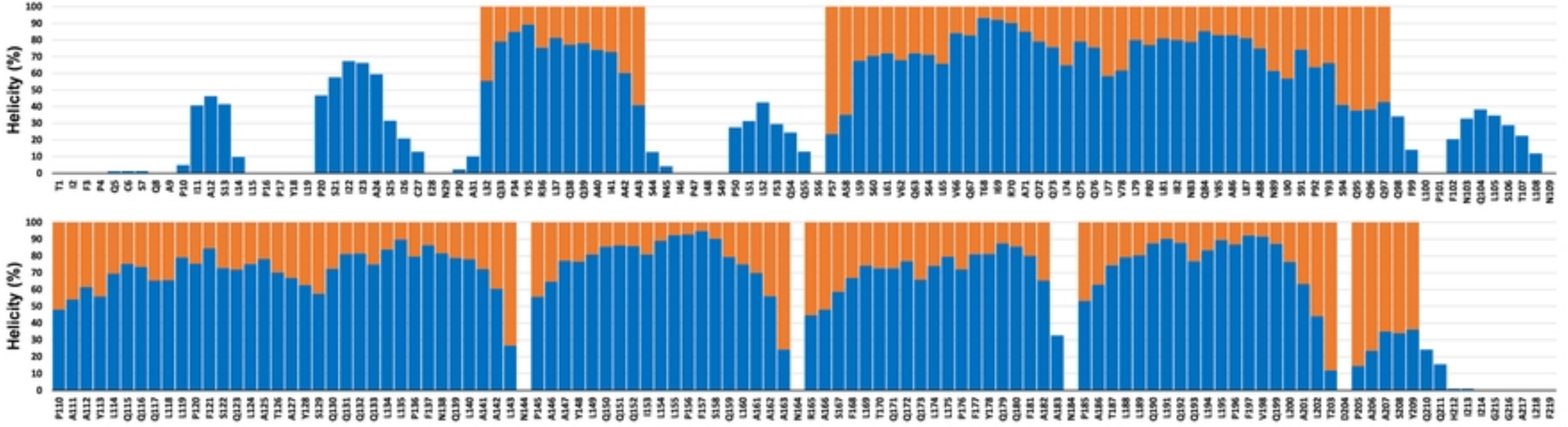


Figure 11

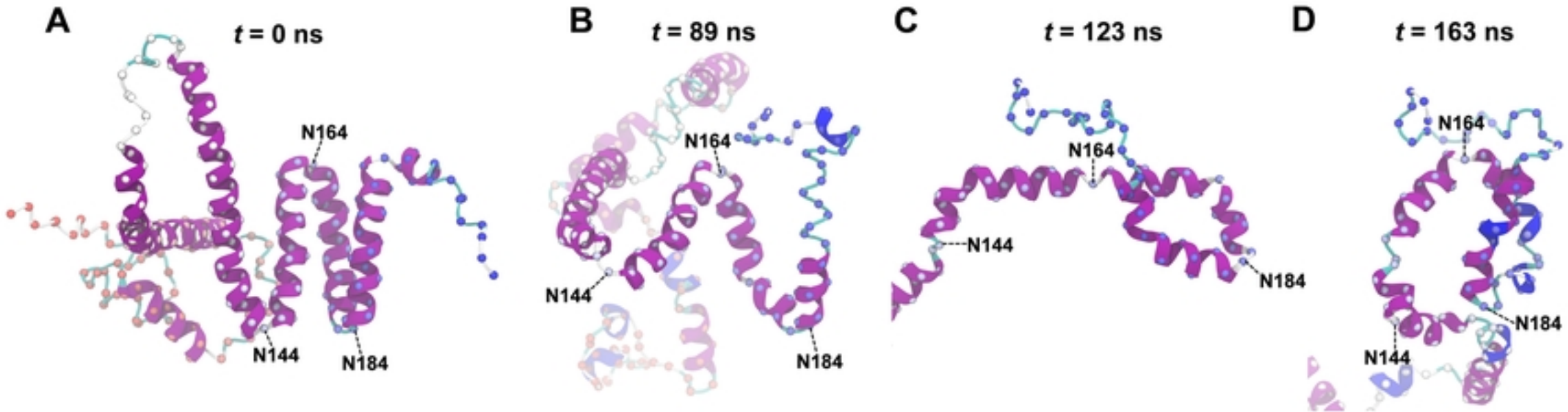
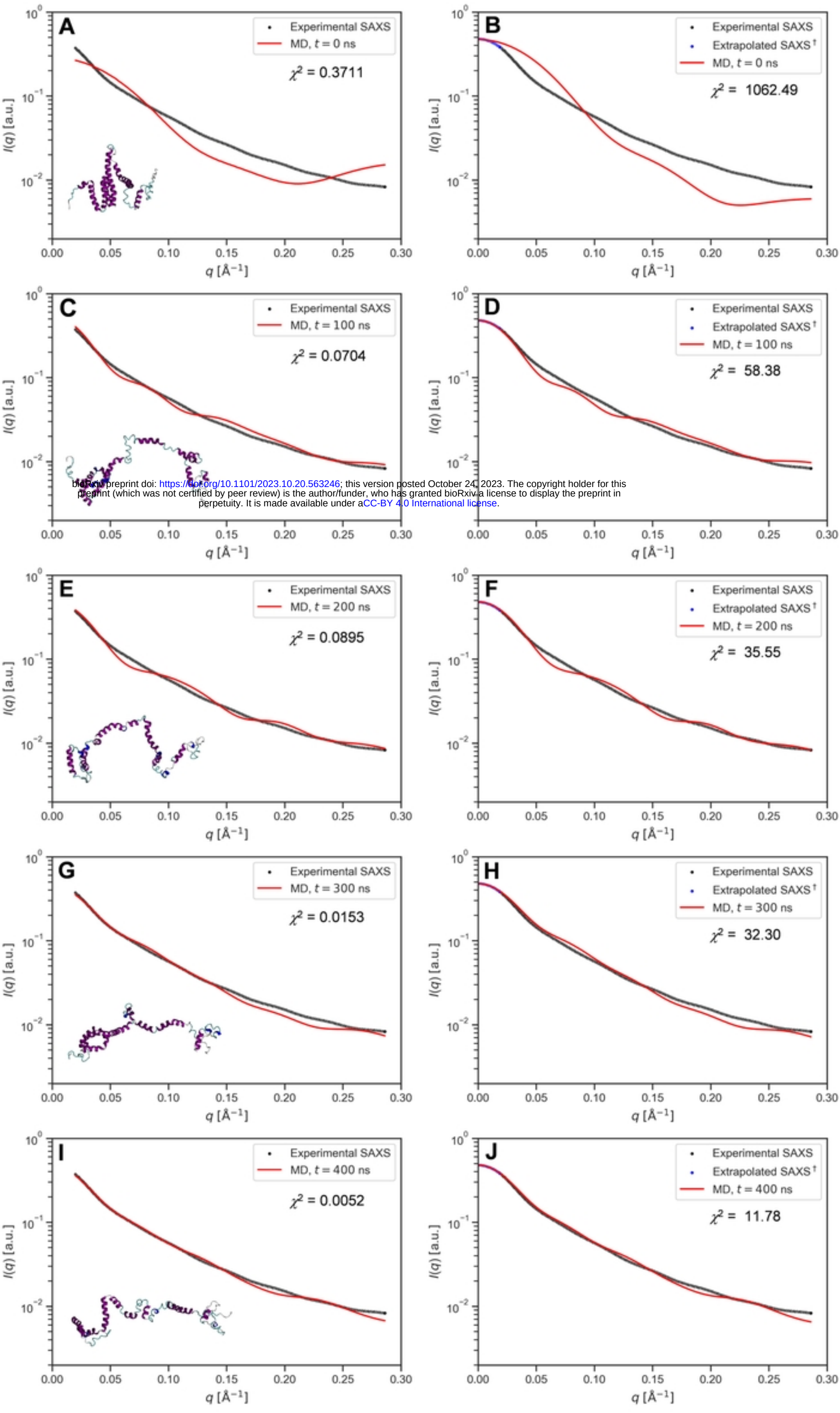

A
t = 0 ns
N164
N144
N184
B
t = 89 ns
N164
N144
N184
C
t = 123 ns
N164
N144
N184
D
t = 163 ns
N164
N184
N144

Figure 12



bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 13

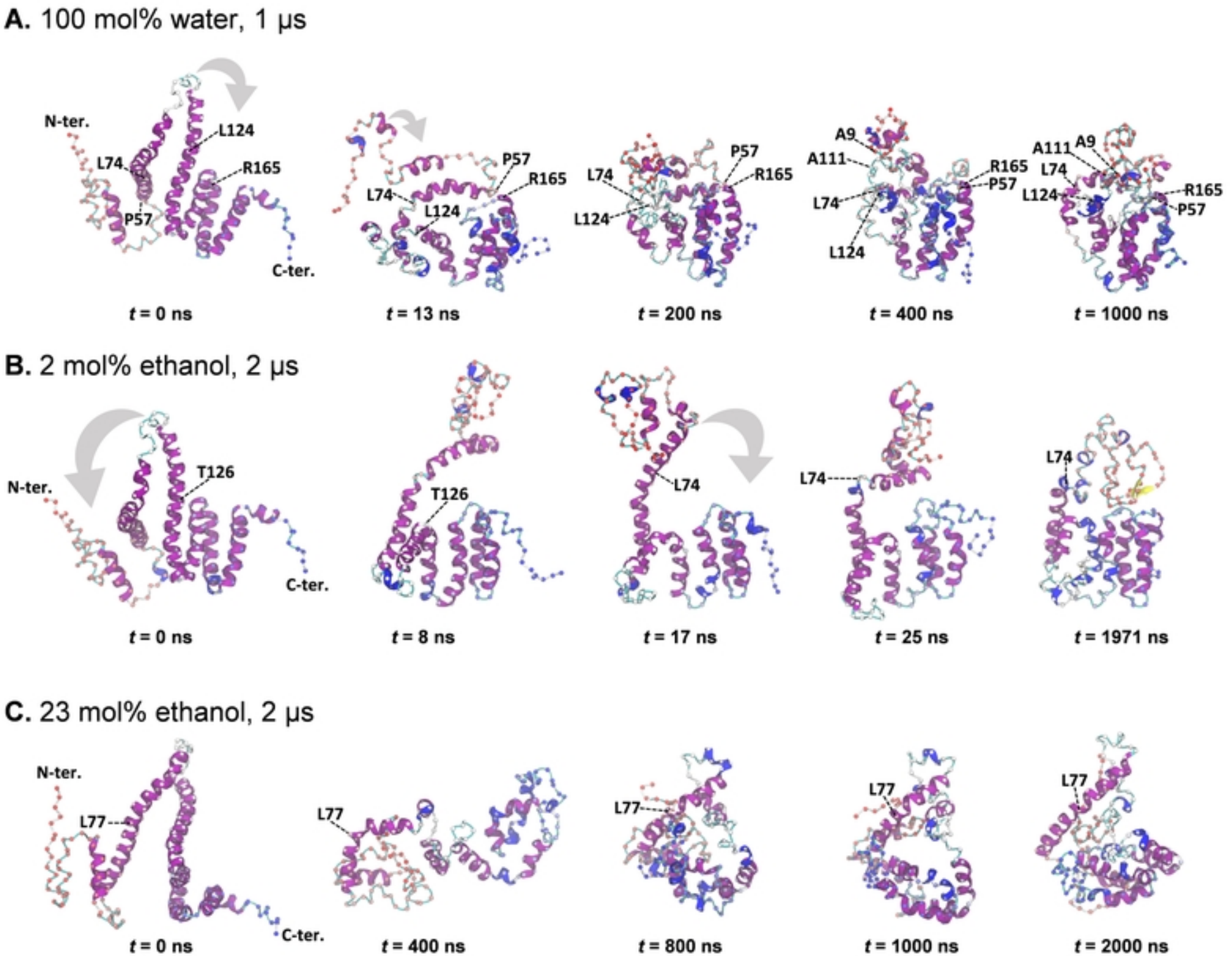

A. 100 mol% water, 1 μs
N-ter.
L124
L74
R165
P57
C-ter.
t = 0 ns
P57
R165
L74
L124
t = 13 ns
P57
R165
L74
L124
t = 200 ns
A9
A111
R165
P57
L74
L124
t = 400 ns
A111
A9
L74
L124
R165
P57
t = 1000 ns
B. 2 mol% ethanol, 2 μs
N-ter.
T126
C-ter.
t = 0 ns
T126
t = 8 ns
L74
t = 17 ns
L74
t = 25 ns
L74
t = 1971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

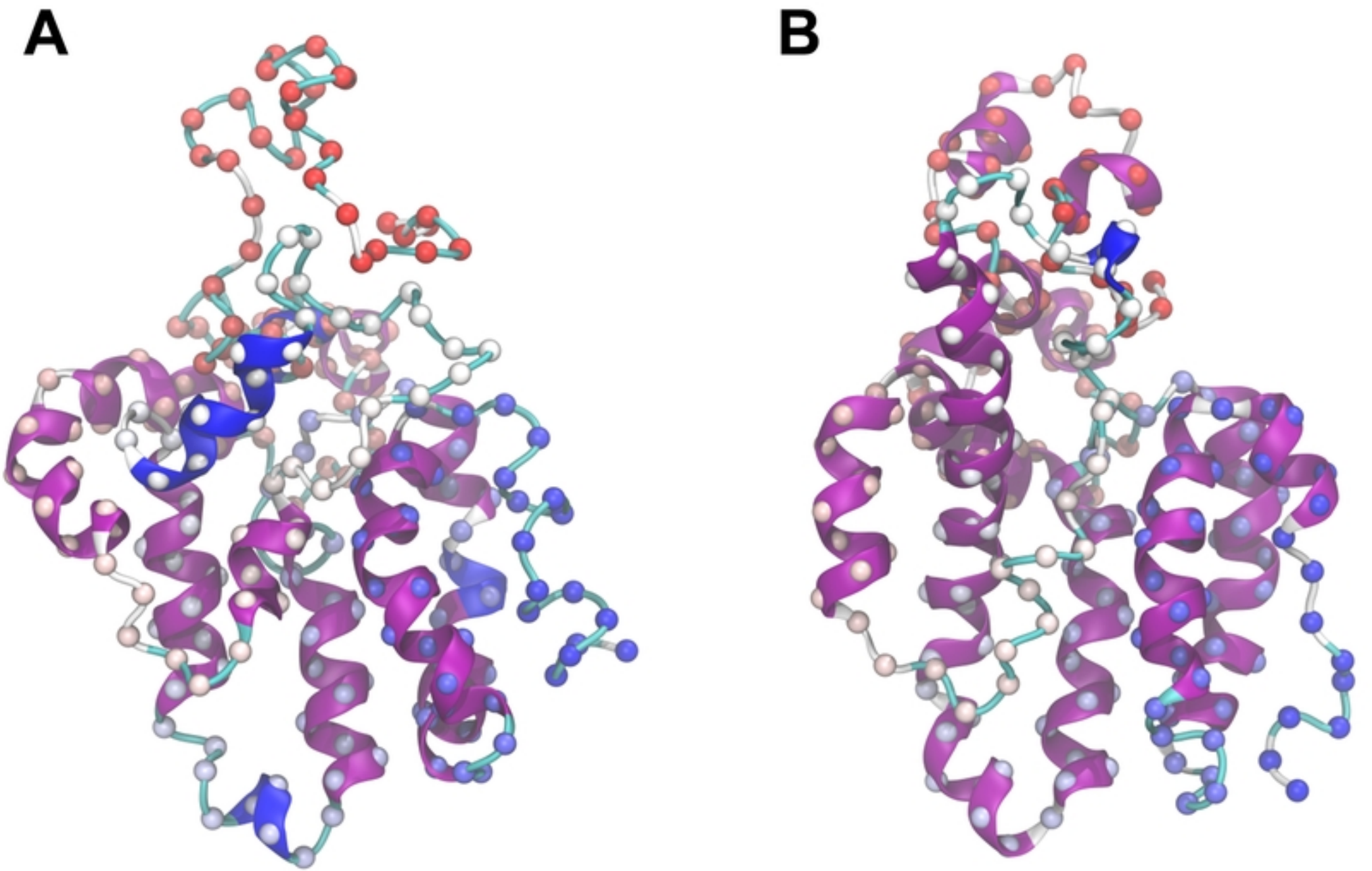


Figure 15

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

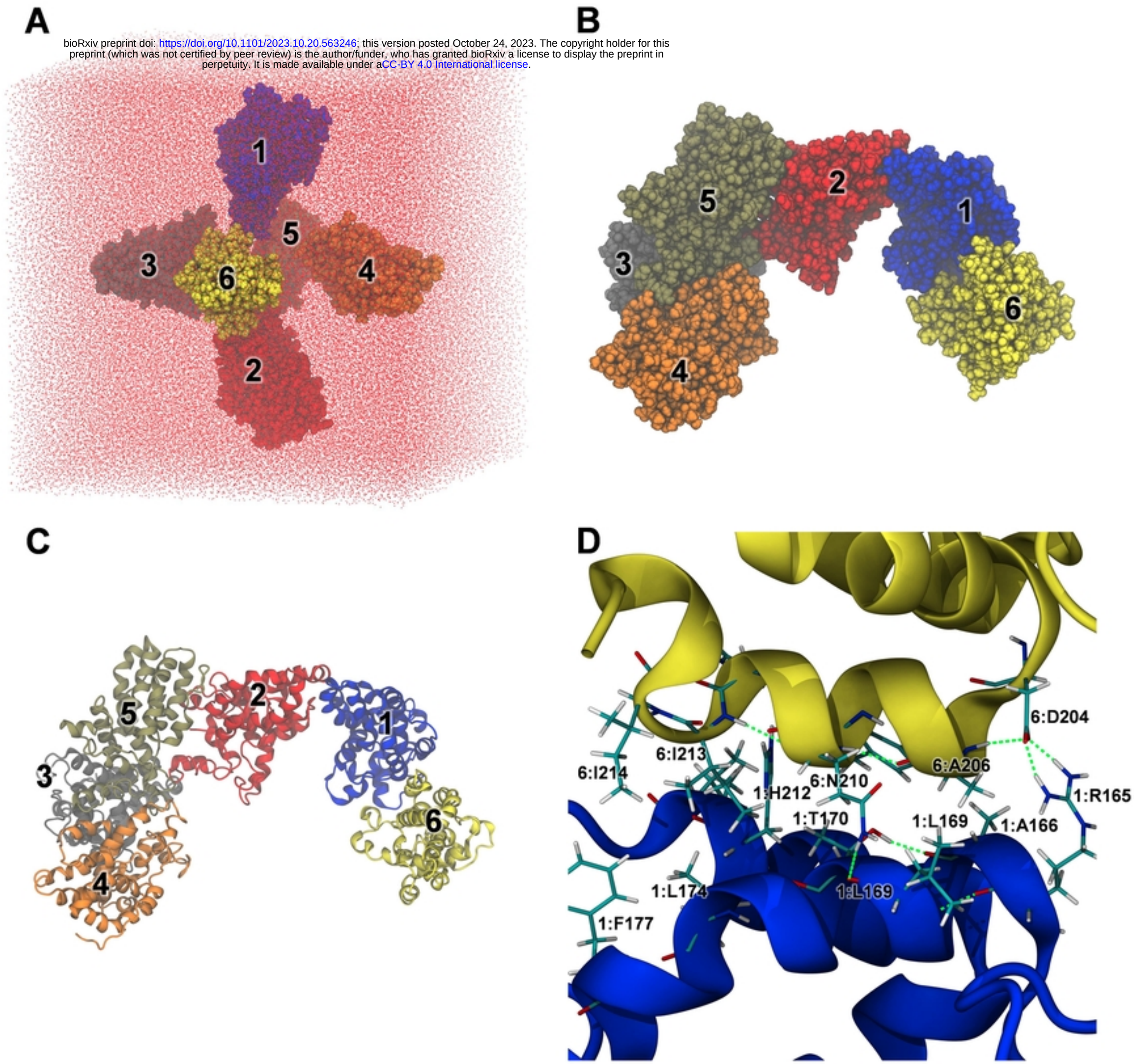


Figure 16

bioRxiv preprint doi: https://doi.org/10.1101/2023.10.20.563246; this version posted October 24, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

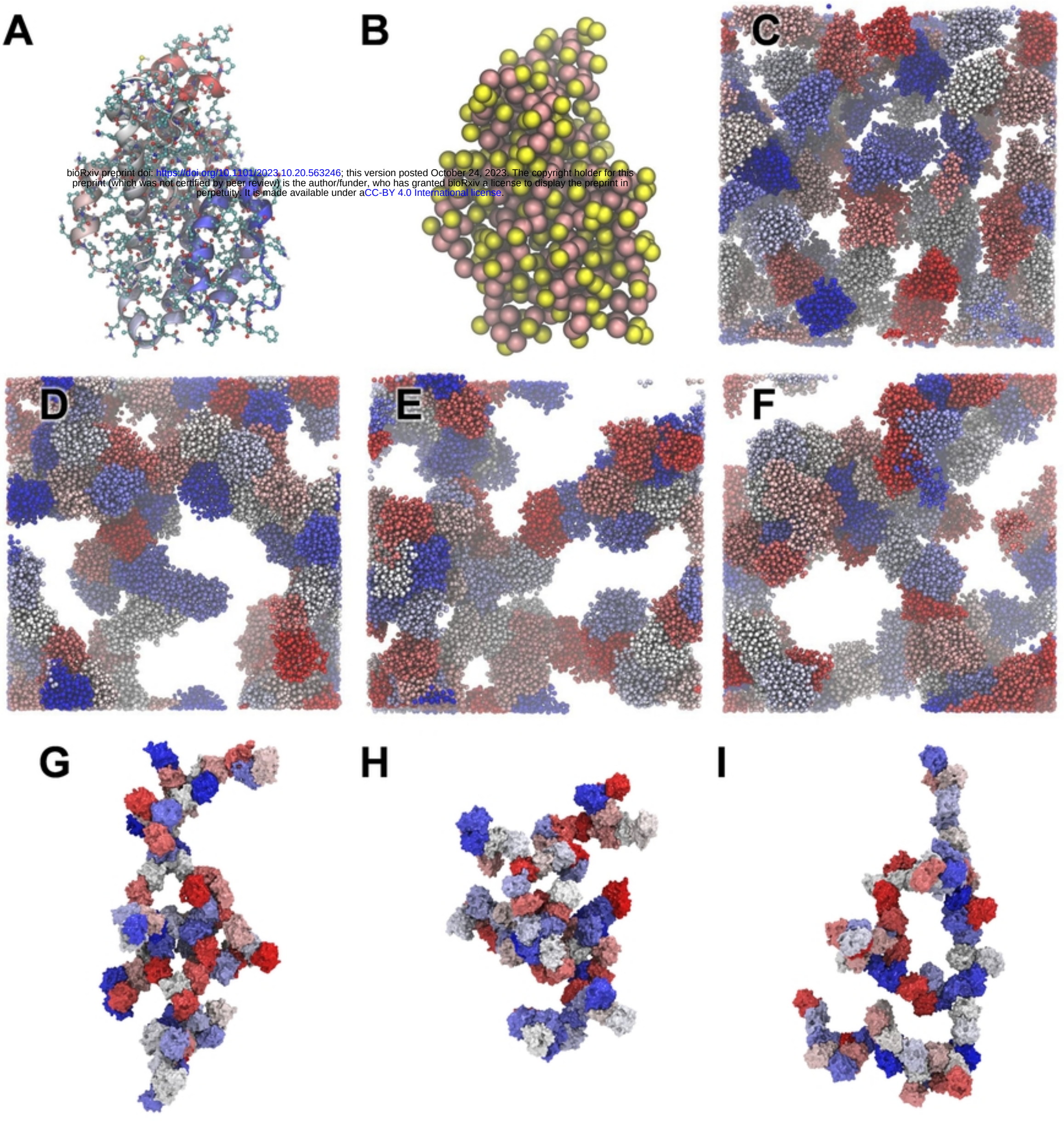


Figure 17

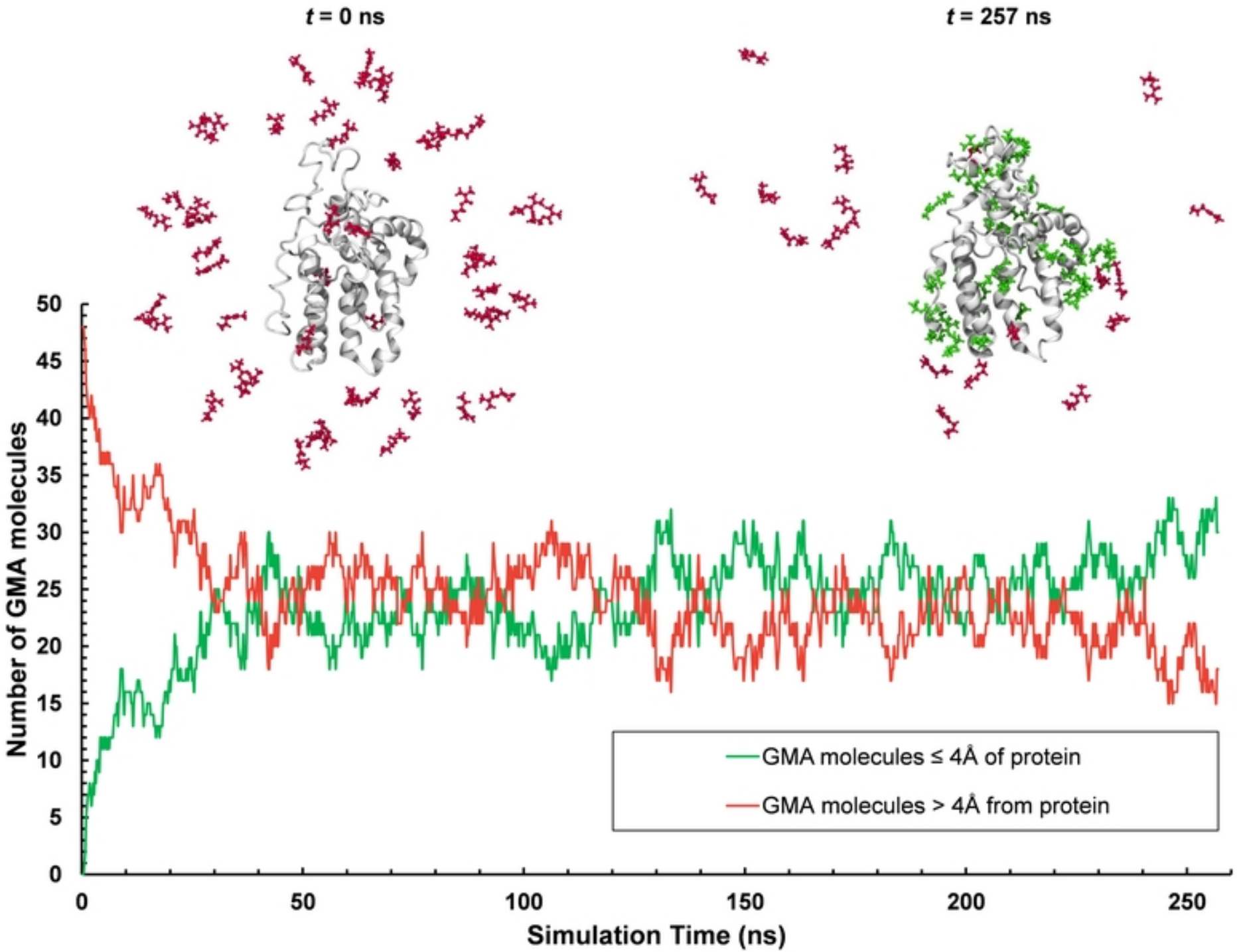

t = 0 ns
t = 257 ns
Number of GMA molecules
Simulation Time (ns)
0
5
10
15
20
25
30
35
40
45
50
0
50
100
150
200
250
GMA molecules ≤ 4Å of protein
GMA molecules > 4Å from protein

Figure 18

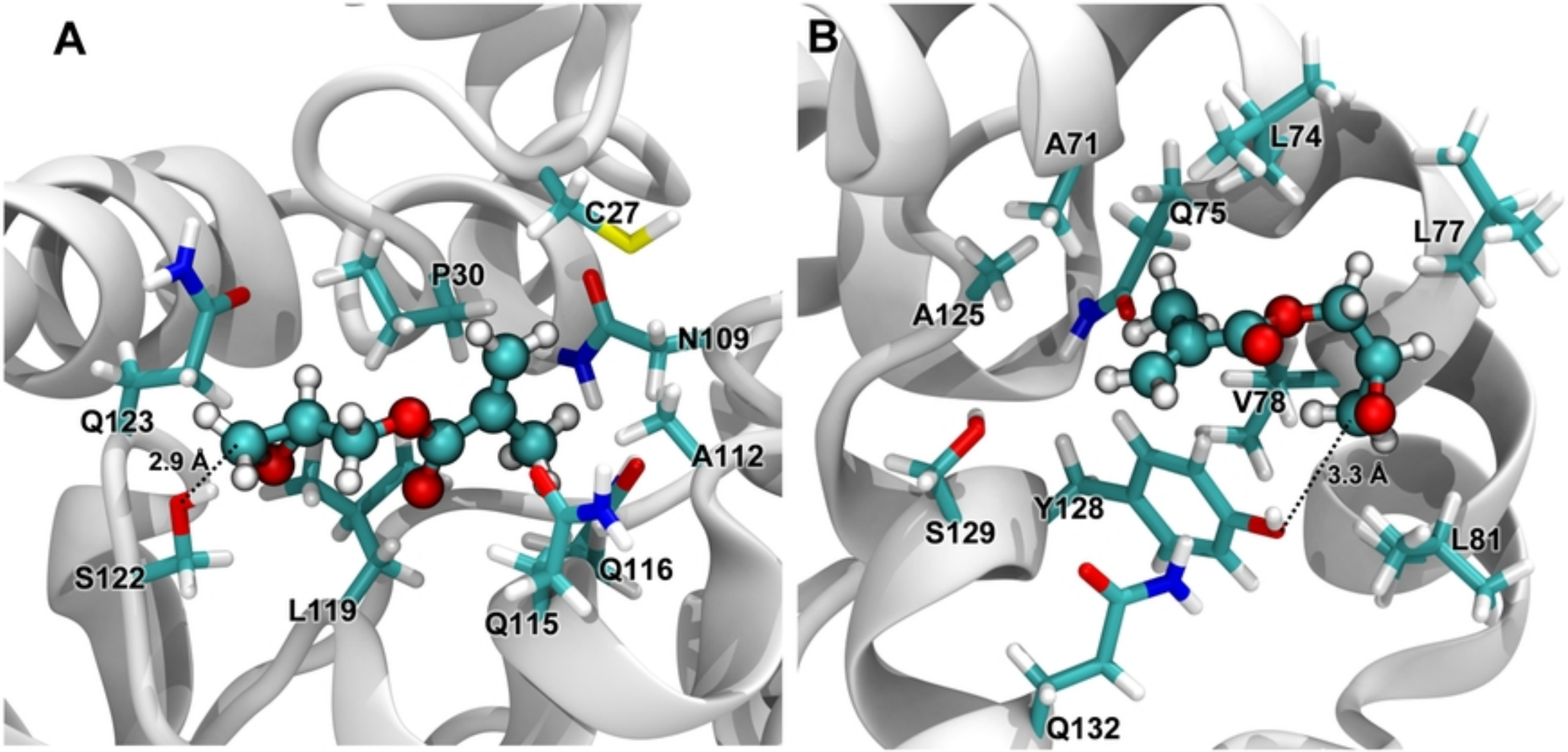

A
C27
P30
N109
Q123
2.9 Å
A112
S122
L119
Q115
Q116
B
A71
L74
Q75
L77
A125
V78
3.3 Å
S129
Y128
L81
Q132

Figure 19