# Investigating the Bioactive Conformation of Angiotensin II Using Markov State Modeling Revisited with Web-Scale Clustering

Emmanouil Christoforou, Hari Leontiadou, Frank Noé, Jannis Samios, Ioannis Z. Emiris, Zoe Cournia

▶ **To cite this version:**

Emmanouil Christoforou, Hari Leontiadou, Frank Noé, Jannis Samios, Ioannis Z. Emiris, et al.. Investigating the Bioactive Conformation of Angiotensin II Using Markov State Modeling Revisited with Web-Scale Clustering. Journal of Chemical Theory and Computation, 2022, 18 (9), pp.5636-5648. 10.1021/acs.jctc.1c00881 . hal-03895590

**HAL Id: hal-03895590**
**https://hal.science/hal-03895590**

Submitted on 9 Aug 2023

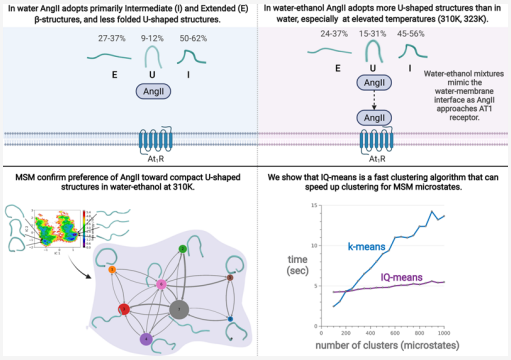
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Investigating the Bioactive Conformation of Angiotensin II Using Markov State Modeling Revisited with Web-Scale Clustering

Emmanouil Christoforou, Hari Leontiadou, Frank Noé, Jannis Samios, Ioannis Z. Emiris,* and Zoe Cournia*

**ABSTRACT:** Molecular dynamics simulation is a powerful technique for studying the structure and dynamics of biomolecules in atomic-level detail by sampling their various conformations in real time. Because of the long timescales that need to be sampled to study biomolecular processes and the big and complex nature of the corresponding data, relevant analyses of important biophysical phenomena are challenging. Clustering and Markov state models (MSMs) are efficient computational techniques that can be used to extract dominant conformational states and to connect those with kinetic information. In this work, we perform Molecular Dynamics simulations to investigate the free energy landscape of Angiotensin II (AngII) in order to unravel its bioactive conformations using different clustering techniques and Markov state modeling. AngII is an octapeptide hormone, which binds to the AT1 transmembrane receptor, and plays a vital role in the regulation of blood pressure, conservation of total blood volume, and salt homeostasis. To mimic the water−membrane interface as AngII approaches the AT1 receptor and to compare our findings with available experimental results, the simulations were performed in water as well as in water−ethanol mixtures. Our results show that in the water−ethanol environment, AngII adopts more compact U-shaped (folded) conformations than in water, which resembles its structure when bound to the AT1 receptor. For clustering of the conformations, we validate the efficiency of an inverted-quantized *k*-means algorithm, as a fast approximate clustering technique for web-scale data (millions of points into thousands or millions of clusters) compared to *k*-means, on data from trajectories of molecular dynamics simulations with reasonable trade-offs between time and accuracy. Finally, we extract MSMs using various clustering techniques for the generation of microstates and macrostates and for the selection of the macrostate representatives.




## INTRODUCTION

The renin−angiotensin (RAS) system is a hormone system implicated in hypertension. The octapeptide Angiotensin II (Asp-Arg-Val-Tyr-Ile-His-Pro-Phe, AngII) is a physiologically active component of the RAS system, which regulates vasoconstriction, electrolyte and water absorption, as well as blood pressure and total blood volume.[1−3] AngII results from the conversion of its precursor angiotensin-I (Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu, AngI) into AngII by the angiotensin-I converting enzyme located on the surface of vascular endothelial cells, predominantly those of the lungs. AngII then binds to the AngII transmembrane receptor, AT1, which promotes various intracellular signaling pathways, resulting in hypertension, endothelial dysfunction, vascular remodeling, and end organ damage.[4]

The bound structure of AngII on its receptor AT1 remains so far elusive although many efforts have aimed to establish the bioactive bound conformation of AngII. A recent crystallographic structure of AT1 in complex with the nonpeptide antagonist ZD7155 shows that AT1-ZD7155 is well buried inside the protein, and thus it could be assumed that, similarly, AngII most probably interacts primarily with the protein environment than with the solvent. It has been suggested that as AngII enters the receptor, the loss of aqueous solvation is compensated by the intermolecular interactions of polar residues resulting in the folding of the peptide into a compact conformation.[6] Indeed, an X-ray structure of the Angiotensin II (AngII)-Fab complex shows that the hormone peptide adopts a compact U-shaped structure (Figure 1a) when it is bound to the receptor[7] with the membrane when bound to the AT1 receptor.[5] However, photolabeling studies of [Bpa[3]]AngII with the AT1 receptor indicate a model where the peptide in its bound form adopts a rather extended $\beta$-strand structure

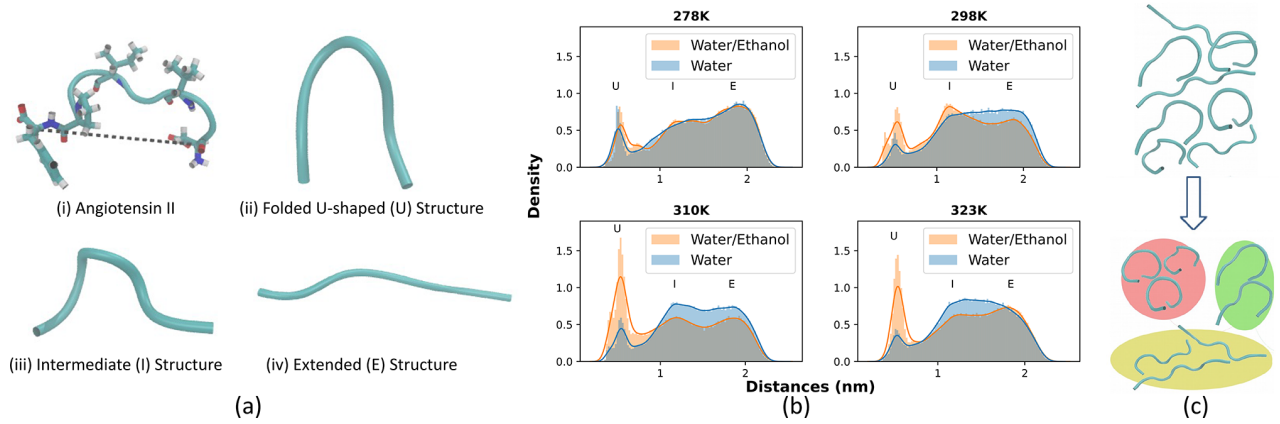



**Figure 1.** (a) AngII representative structures. The dotted line indicates the end-to-end distance of terminal residues. The conformations are visualized as cartoon. (b) Histogram and probability densities, $P(r)$, of end-to-end distances of the terminal residues in water (blue) and water/ethanol (warm pink) simulations. Conformations concentrate mainly in three peaks U (u-shaped) = 0.5 nm, I (intermediate) = 1.1−1.3 nm and E (extended) = 1.8−2.0 nm. (c) Pool of AngII conformations clustered by GROMOS.

(Figure 1a).[8] Furthermore, several nuclear magnetic resonance (NMR) studies of AngII in aqueous and organic solvents suggest the existence of a mixture of unfolded and folded conformations ranging from extended $\beta$-structures to the more compact U-shaped ones.[3,6,9−11]

Organic solvents such as 2,2,2-trifluoroethanol or simple temperature annealing have been reported to favor the fold of the angiotensin peptide in a rather compact structure determined through circular dichroism measurements.[12] However, recent NMR and molecular dynamics (MD) studies of the [Val5] AngII analogue in a water/ethanol (35% v/v) binary solvent at low temperatures suggest that the peptide is primarily in an extended $\beta$-structure and that most regions of the peptide are preferentially solvated by ethanol molecules.[13,14] Besides the inherent flexibility of the octapeptide, another reason for the divergent structural models proposed over the years could be the use of different solvents, such as water/ethanol and dimethyl sulfoxide as well as different experimental conditions. Key structures of AngII, such as folded (U-shape) and extended ($\beta$-strand), are illustrated in Figure 1a.

Mixtures of ethanol and water can influence the stability and conformational properties of biological molecules in diverse ways and, importantly, act as a mimic for the peptide−membrane interface. Using such solvents in the presence of biomolecules that usually function in environments that are only partly aqueous may produce results that help in understanding the role(s) of water in maintaining the macromolecule structure and activity and can act as mimics of the water−membrane interface. In enzyme catalysis, mixtures of water and ethanol provide reaction media that enable enzymes to act on substrates that are not soluble in water, although substrate specificity, reaction rates, and protein stability may be altered by the solvent mixture. The observed effects may be the result of direct interactions of solvent alcohol molecules with a protein. In general, preferential solvation of proteins by organic cosolvents and clustering of organic molecules in the binary solution are considered possible mechanisms that could either allow or block proteins from jumping between different conformational states.[15−17]

The aim of this paper is to describe the conformations of AngII using clustering techniques and Markov state modeling in water−ethanol mixtures as a water−membrane model interface mimic to mimic the approach of the AngII peptide to the water−membrane interface. Also, we use and validate the efficiency of an inverted-quantized $k$-means algorithm (IQ-means) as a fast approximate clustering technique for web-scale data compared to $k$-means. Moreover, we aim to provide an atomic-level picture of the intermolecular interactions governing AngII−water and AngII−water−ethanol mixtures using MD simulations. The structural and dynamic properties of these systems are outlined in the context of identifying the solvent microenvironment around the AngII peptide and in describing the representative conformations of AngII in these two media. Our results are in excellent agreement with relevant experimental data and provide insights into the bioactive conformation of this hormone. Moreover, since identifying the transition states by simply sampling MD simulations is not sufficient, we apply Markov state models (MSMs),[18−23] a powerful framework to analyze MD simulations and extract the long-time statistical conformational dynamics. A step of this framework is to discretize the trajectories from the simulations using clustering techniques, such as $k$-means. For long simulations resulting in big data, clustering may consist of a time-consuming and computationally intensive process. To reduce the computational effort, we employ IQ-means, a fast approximate clustering method designed for web-scale clustering (e.g., hundreds of millions of points, such as images and web documents into thousands or millions of clusters). Our results validate the efficiency of IQ-means in clustering MD simulations and discretizing trajectories for the MSM and provide a new scheme for efficient big data handling.

## METHODS

**System Preparation.** Human AngII was simulated in water and ethanol/water-35% (v/v) solvents at different temperatures, $T$ = 278, 298, 310, and 323 K (Table S1). The initial structure of the octapeptide was retrieved from the Protein Data Bank, 1N9V.pdb.[3] The apparent pH was chosen to be 4 in order to match the experimental conditions of refs 13, 14. Therefore, in our model, the peptide termini were ionized. Asp was not protonated and charged (−1), Arg was protonated and charged (+1) and His was protonated and charged (+1). The system was solvated in pure water or 35% ethanol-1,1-d2-water (v/v) and neutralized with 1 $Cl^-$ ion. In

both systems one AngII peptide was solvated in a 15 Å solvent box.

**MD Simulations.** MD simulations were performed using GROMACS 4.6.1.[24] The AMBER99SB-ILDN force field[25] was chosen for the peptide, and TIP4P[26] was used for the water model. The model for the ethanol molecule was the one previously studied in ref 13. A 2 fs time step was used together with PME for the calculation of the long-range electrostatics and a cutoff of 1.4 nm for both electrostatics and van der Waals interactions. The nonbonded neighbor list is updated every five steps. LINCS[27] is used as a constraint algorithm for the C−H bond, and Berendsen coupling was used for the pressure and temperature control in the equilibration phase, while Nosé−Hoover was employed for the production runs. The simulated systems are summarized in Table S1. The analysis of the trajectories dynamical and structural properties such as distances, number of hydrogen bonds, mean squared displacements, and radial distribution functions (RDFs) was performed with GROMACS 4.6.1 tools using default options. In the first 50 ns the systems equilibrated and thus this time was discarded. The reported self diffusion coefficients, $D$, were calculated by fitting the linear regime of the mean squared displacement curves of each system and using the Einstein relationship:

$$6D(t) = \lim_{t\to\infty} \langle \Delta \vec{r}_i(t) \rangle$$

Initially, MD simulations for AngII using single starting conformation were performed up to 600 ns (Figures S1−S3). In order to increase the sampling for the construction of MSMs, 10 conformations chosen from each system were selected based on root mean squared deviation (RMSD) >0.15 nm and were used to initiate new, independent simulations (80 × 100 ns = 8 $\mu$s total simulation time, Table S1, Figure S5−S8). Conformational snapshots were saved every 10 ps so that finally 100,000 conformations for each system were selected for MSM construction and subsequent analysis.

**MSMs.** To unravel the bioactive conformation of AngII, we processed the MD simulations using Markov state modeling; MSMs were constructed using the PyEMMA software.[28] MSMs are discrete-time models based on the kinetic exchange between states that describe a decomposition of the conformational space into small metastable regions.[29−31] MSMs provide the means to understand and gain an insight from simulation data with complex nature by predicting long timescale dynamics from long or multiple short trajectories. The construction of an MSM is far from trivial, since it involves a lot of decisions. The key steps to build an MSM (Figure 2) are as follows:

(1) The selection of features from the MD trajectories and the application of time-lagged independent component analysis (TICA) transformation, to prepare the state space.
(2) The "geometric" clustering step, to discretize the trajectories into finite states, the microstates.
(3) The estimation of a transition probability matrix for the microstates with proper lag time for the Markov model.
(4) The "kinetic" clustering step, to group the microstates by the transition probability matrix into sets of kinetically related states, the macrostates.
(5) The coarse-graining of the kinetic model based on the produced macrostates.

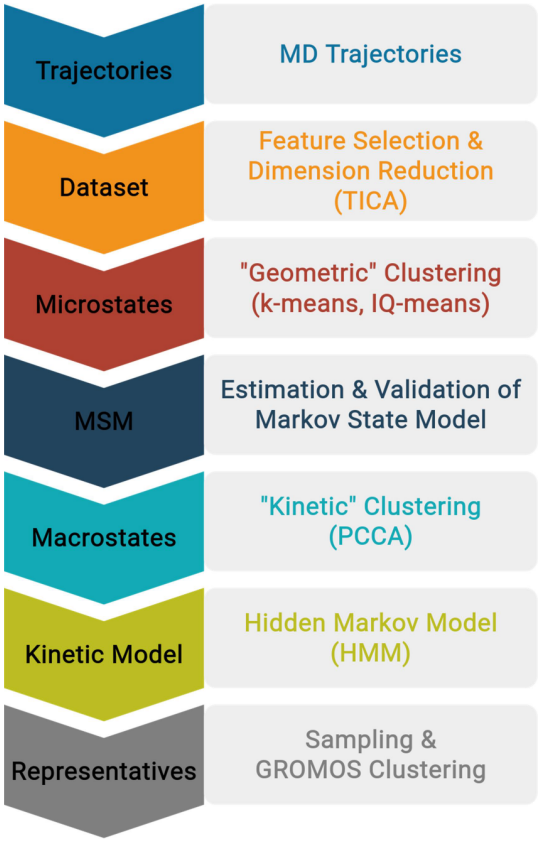



**Figure 2.** MSM construction steps.

**Feature Selection and Dimension Reduction.** Instead of using raw MD trajectory conformations, we select a set of features in order to characterize best the rare event transitions. Distance metrics, such as RMSD between heavy atom or C$\alpha$ coordinates[21,31,52] or backbone dihedral angles,[31,38] are often candidates for the construction of Markov models.[23,28] Here, we experimented with different featurizations and TICA lag times by checking representations against end-to-end distances and by estimating and validating the resulting MSMs, and we keep all the dihedral angles (all backbone phi/psi and chi1 angles) and the minimum distances between peptide heavy atoms (Table S2), which, compared to AngII end-to-end distances (Figure 1a), appear to finely discretize key structures. It should be noted that we do not directly use dihedral angles but their cosine and sine values (doubling the number of selected angle coordinates) because, after the cos/sin transformation, standard operations such as calculating and subtracting the mean can be performed.[28,53] The resulting dataset consists of 59 features.

For the efficient generation of the microstates, it is recommended to reduce the dimensions of the selected features[28] in order to improve the quality of the discretizations[32] and the CPU time.

**TICA.** TICA is a linear transformation method, which finds coordinates of maximal autocorrelation at a given lag time and has been shown to be optimal in approximating the relevant slow reaction coordinates from MD simulations.[33,34] Thus, TICA is considered to be ideal to construct MSMs. TICA parameters, such as lag time, features, and remaining model's parameters, are historically chosen heuristically and can be evaluated using scoring functions such as the generalized

matrix Rayleigh quotient[54] to pick the best model.[35] In this work, we selected a TICA lag time of 50 steps (0.5 ns), reducing the dimension of our data to 22 and 26 for AngII simulations at 310 K for water and water/ethanol systems, respectively, while preserving 95% of the kinetic variance. This was achieved by testing and evaluating the resulting state space on the resulting free-energy landscape free energy computed over the first two TICA coordinates, its discretization ability against the examined key structures of AngII, and by evaluating the quality of the estimated MSMs.

**Microstates.** The discretization of the conformational space can be accomplished by clustering techniques such as *k*-means, Ward's method,[55] *k*-medoids, GROMOS, etc. The number of microstates typically ranges from 100 to 100,000.[23] Here, a *k*-means clustering with 100 centers was carried out. In Figures S22 and S23, we demonstrate that using more microstates (250, 500) appeared to provide worse descritizations, increasing the errors of the constructed MSMs. The results were compared with those of a fast approximate clustering algorithm, IQ-means.[36]

***k*-means.** For the discretization of the trajectories, it is suggested to use *k*-means,[28] which appears to produce the best MSMs for protein folding along with Ward's method. *k*-means produces a balanced clustering and, combined with *k*-means++[37] initialization, consists a fast solution that converges in only a few iterations with reproducible results.

**IQ-means.** IQ-means is a fast approximation clustering method for web-scale clustering. It uses multiple ingredients from advanced approximate *k*-means variants, designed for large-scale datasets (big data). Some key ingredients of IQ-means are: (i) the fine representation of data in a 2D grid, (ii) the multiindex-based inverted search from centroids to cells, and (iii) the dynamic version of the algorithm that comes as a natural extension from Expanding Gaussian Mixtures (EGM).[38] It has been reported to achieve the clustering of 100 million images on a single machine in less than an hour in calculations using deep learned representations for the images.[36]

IQ-means starts by learning a codebook for data representation, as in the inverted multi-index,[56] using a small sample of the data. Pretrained codebooks can be also used, if available. All points are then quantized on the codebook's grid, like DRVQ,[57] and a discrete two-dimensional distribution of points over cells is constructed. The algorithm alternates between an assignment and an update step, similarly to *k*-means. During the assignment step, searches for a set of individual points in the nearest cells of each centroid are made following a reverse approach, which makes it extremely fast. Instead of looking for the nearest centroid of every cell, it looks for the nearest cells (points) for each centroid using a window (ranged search) in the 2D grid.

**Estimation of MSMs.** MSMs are the models that describe the kinetics of molecules by a reversible transition matrix[54] of conditional transition probabilities among the microstates. An MSM is composed of the conditional probabilities for a state space, that consists of $s(t)$ discrete trajectories, jumping between $n$ microstates at lag time $\tau$ ($p_{ij} = \Pr(s(t+\tau) = j \vee s(t) = i)$). The probability of jumping between states is computed by the maximum likelihood estimator.

**Implied Timescales.** To ensure the accuracy of the MSM, we select a lag time, so that the implied relaxation timescales are approximate constant within the statistical error. The behavior of the implied timescales consists of a way to check if the model is Markovian.[19] The implied timescales refer to the relaxation timescales of a molecule implied by the transition matrix of a Markov model at a lag time $\tau$ and is given by $t_i(t) = \frac{-\tau}{\ln|\lambda_i(\tau)|}$, where $\lambda_i(\tau)$ is the $i$th eigenvalue of the transition matrix $P(\tau)$. For a Markovian system, $\lambda_i(\tau)$ should be constant and independent of the lag time $\tau$.

**Model Validation.** The model, at the selected lag time, is validated using the Chapman−Kolmogorov test[23] a formulation for the Chapman−Kolmogorov equation ($P(k\tau) = P^k(\tau)$). Because of this test, a Markov model estimated at a lag time $\tau$ should be able to predict estimates performed at longer timescales $k\tau$ within the statistical error.

**Macrostates.** The model is coarse-grained using the transition probabilities among the microstates to a simpler, "humanly" readable, model. There are a variety of coarse-graining techniques (Perron cluster cluster analysis (PCCA)[18] and BACE[39]) to simplify the model, exploiting the kinetic relevance of the states. Here, we performed PCCA++.[40]

**PCCA.** PCCA coarse-grains MSMs exploiting the eigenspectrum of the transition probability matrix. Starting with all the microstates merged into one big macrostate, it iteratively splits the most kinetically diverse macrostates, until the requested number of states is reached. The most common approaches for MSMs are PCCA+[41] and PCCA++.[41]

**Kinetic Modeling.** The final coarse-grained kinetic model and the estimation of the transition rates between the metastable states are generated using hidden Markov models (HMMs). HMMs consist of an efficient approximation of the kinetics on discretized molecular state spaces.[42]

**Macrostate Representatives.** In this work, the macrostate representatives are selected using GROMOS clustering.[43] Instead of randomly sampling from the macrostates, GROMOS is applied on large groups of sampled conformations from each macrostate and the medoid of the bigger cluster is assumed as a "dominant" representative. We use large samples from the metastable states to minimize the required computational power and time, ensuring quality of the results (larger samples have smaller approximation error).

**GROMOS.** GROMOS[43] is a method for clustering using a given distance cutoff and comparing the RMSD of their atomic positions. Figure 1c shows an example of clustering a pool of conformations using their RMSD distances.

**Metadata Representatives.** The metadata representations for protein conformations in a trajectory, as suggested by Zhang et al.,[44] are simple 3D points representatives produced by multidimensional scaling (MDS). To generate these representations a distance matrix of the backbone atoms (for an example of 3 out of the 276 used AngII backbone atom distances for MDS, see Figure S10) is computed for each conformation and then classical MDS is applied to reduce the dimension of the matrix to 3. The eigenvalues from the MDS seem to preserve the amount of variations in data, consisting of a fine metadata representation with a single 3D point for each conformation. Such representations allow us to efficiently visualize our simulations into noncomplex plots.

## RESULTS AND DISCUSSION

**Conformational Sampling of AngII in Different Solvents and Temperatures.** The RMSD compared to the initial structure has been calculated for all different systems (Figure S1). The average RMSD for both systems in water and water/ethanol solvent remains the same. However, it is

observed that for elevated temperatures (310, 323 K) the RMSD fluctuations increase, indicating a folding/unfolding pattern of the peptide. The folding/unfolding behavior of the peptide is noticeable also in the plots of the radius of gyration of AngII in all different systems (Figure S2). In our simulations, we observe that the folding/unfolding events of AngII at elevated temperatures (310, 323 K) seem to be significantly more frequent in the binary solvent (water/ethanol) than in water (Figures 1b and S5−S7). Also, we notice the impact of the starting conformations on the folding behavior of the peptide. In Figure 3, the probability densities of

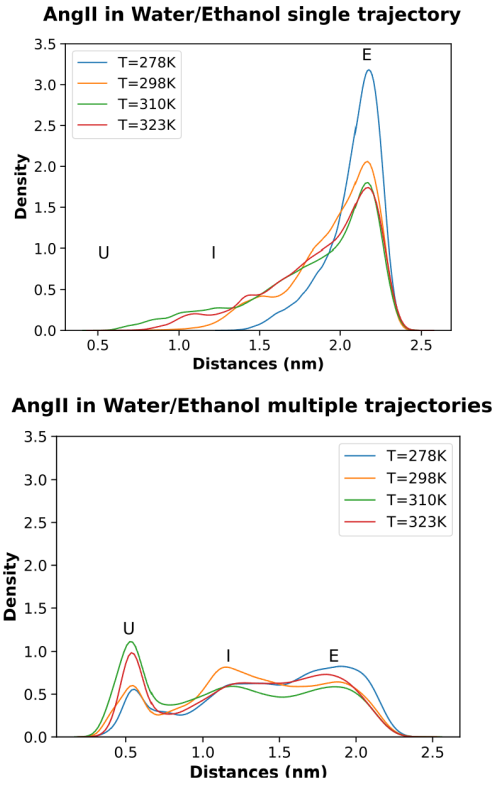



**Figure 3.** Probability densities of end-to-end distances of the terminal residues for MD simulations with single trajectory and multiple trajectories (10 different starting conformations).

end-to-end distances, namely, distances of C$\alpha$ terminal atoms (Figure 1a), show that the conformational samples were significantly increased using multiple trajectories, providing both folded and unfolded structures (Figures S5−S7). On the contrary, during simulations initiated using a single conformation, the molecule adopts mainly an extended coil structure (Figure S3). To differentiate between the three major conformations that AngII adopts, we denote a "U" structure for the U-shaped folded conformations, an "E" structure for an extended conformation and an "I" structure for intermediate conformations.

**Representative Conformations of AngII Observed in Water and in Water/Ethanol.** A cluster analysis was performed for the peptide in order to find out the representative conformations of AngII in pure water and in water/ethanol during the simulation at different temperatures ($T$ = 278, 298, 310, and 323). The conformations were clustered by comparing the RMSD of the C$\alpha$ atoms of the peptide according to the GROMOS method implemented in Gromacs, using a cutoff of 0.15 nm. The medoids from the most populated clusters of each system are selected as the representative conformations and are shown in Figure 4. We

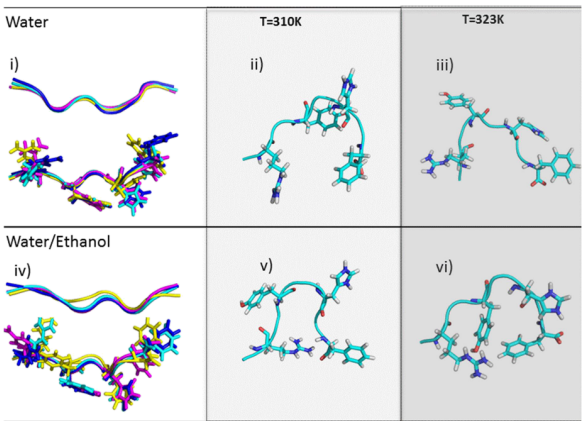



**Figure 4.** First cluster representative conformations of AngII (i) in water and in (iv) water/ethanol are shown at various temperatures. U-shaped conformations are less populated in water at (ii) 310 K and (iii) 323 K and in water/ethanol at (v) 310 K and (vi) 323 K. Peptides are represented in cartoon while some key residues are shown in sticks. Structures at 278, 298, 310, and 323 K are colored in yellow, magenta, cyan, and blue, respectively.

observe that the dominant conformation of AngII in water and water/ethanol is mainly an extended, coil structure (Figure 4i,iv). Note that at the lowest simulated temperature, 278 K, the AngII peptide is mainly in its open extended conformation, which remains stable throughout the simulation. This is in agreement with previous NMR and MD simulation studies.[8,9,13] However, at elevated temperatures, there is a small population of folded conformations that resemble the U-shaped molecule found in other studies.[7,45,46] More specifically, the percentage of these compact structures in water is 7% at 310 K (Figure 4ii) and 6% at 323 K (Figure 4iii). In water/ethanol, the same percentage is found to be 6% at 310 K (Figure 4v) and 17% at 323 K (Figure 4vi). In Figure 1a, all the key structures of AngII (U-shaped, intermediate, and extended) are shown.

**MSM for AngII in Pure Water and in Water/Ethanol.** In order to build a kinetic model for the AngII conformational states in water and in water/ethanol and to provide the different metastable states of these systems and compare with the above-mentioned representative conformations, we performed an MSM. For this purpose, we combined (a) the trajectories of AngII in pure water and (b) in water/ethanol in 310 K in two different datasets, respectively. Because of the short length of AngII and its intrinsic flexibility, the peptide is mainly unstructured. However, probability densities of end-to-end distances indicate a preference for specific states. As seen in Figure 1b, the probability densities, $P(r)$, calculated for all different systems can be roughly divided into three categories that correspond to three major peaks. The peak corresponding to U at 0.5 nm represents U-shaped folded conformations, the peak corresponding to I at 1.1−1.3 nm corresponds to intermediate structures, and the peak corresponding to E at 1.8−2.0 nm to extended ones. This categorization gives us a qualitative and coarse-grained description of the conformational preferences of AngII under different conditions. It is, however, obvious from the plots of Figure 1b (also Figures

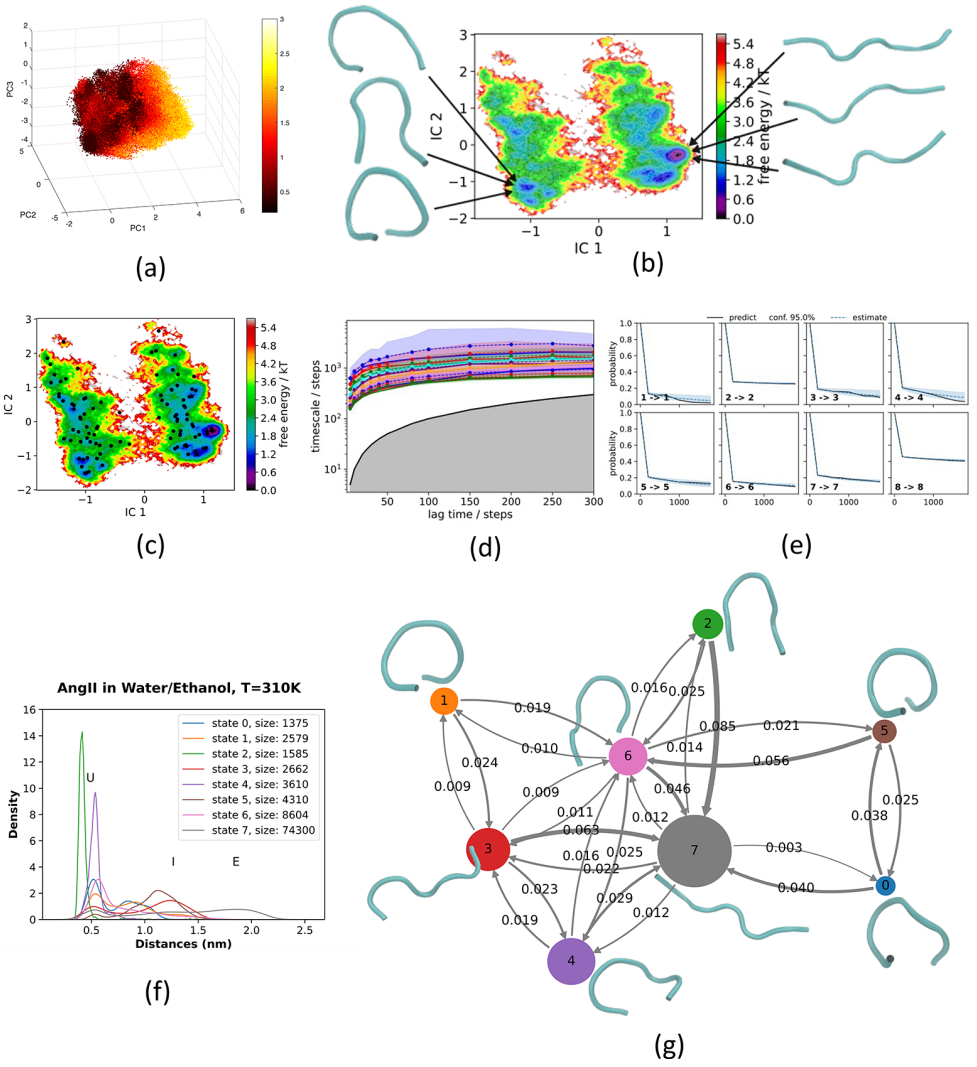



**Figure 5.** MSM construction for AngII in water/ethanol at 310 K. (a) Feature selection. Features colored by end-to-end distances (darker color, lower distance); dimension reduced to three using PCA for visualization. (b) Dimension reduction using TICA. (c) Mircostate clustering. (d) Estimation of implied timescales; different colors correspond to implied timescales. Confidence intervals are depicted with shaded areas. Implied timescales at 2 ns converge and are constant within error. (e) Validation of estimated MSM at 2 ns (200 steps) lag time with the Chapman–Kolmogorov test. (f) Probability Density plots for each macrostate. (g) Kinetic model of MSM with representatives for each state.

S5–S8 and Table S3) that in the presence of ethanol molecules in the solvent the peptide adopts more U-shaped structures than in the absence of ethanol. This is especially obvious in the magnitude of the $P(r)$ peaks at higher temperatures (310 and 323 K).

To further examine the dynamic behavior of AngII folding and unfolding, we built MSMs for AngII in water and in water/ethanol at 310 K. We chose this temperature as it is more relevant to the conditions encountered by AngII in cells in the human body. The features (cos and sin of dihedral angles and minimum distances between heavy atoms) used to construct the MSM seem to preserve the key structures (U, I, E) of the conformations. Figure 5a illustrates the used features colored by their end-to-end distances, reduced by PCA in three dimensions for visualization. Similar end-to-end distances are concentrated together, which verifies that the selected features are suitable to represent the key structures of AngII. Notice that in the front plane of the visualized dimensions (Figure 5a) are the U-shaped conformations (darker colors, lower end-to-end distances), while as we move backward are the intermediate and then the extended conformations (lighter colors, higher end-to-end distances). TICA is applied on the selected features (initial feature space) with a lag time of 50 steps (0.5 ns) to prepare the statespace for the "geometric" clustering. The adapted description of the conformational states using selected features and TICA transformation is verified on the free-energy surface of the first two TICA coordinates (Figure 5b), where the discretization of state space seems to properly describe the folding process of AngII in the water–ethanol solvent. In Figure 5b, sampled conformations of states with high probability are shown. On the left state lie folded U-shaped conformations, and on the right are extended states. Then, a $k$-means clustering with 100 centers was carried out (Figure 5c). Using more microstates (such as 250, 500) appeared to provide worse discretizations, increasing the errors of the constructed MSMs (Figures S22 and S23).

Implied timescales (shown in Figure 5d) appear to converge at 200 steps, indicating a lag time $\tau$ = 2 ns. The estimated Markov state model at 2 ns (200 steps) lag time is validated using the Chapman–Kolmogorov test (Figure 5e). We use the

implied timescales plots (Figures 5d and S27) as a heuristic to estimate a number of metastable states. From the timescale separation (Figure S27b), we observe some comparably large timescale gaps, within the time resolution of the model, between the 6th and 7th, and the 8th and 9th relaxation timescales, suggesting seven to nine metastable states. We evaluated coarse-grained models with seven to nine metastable states (retaining six to eight relaxation timescales accordingly) using the Chapman−Kolmogorov test. The model with the assumed eight metastable states yields a passing Chapman−Kolmogorov test (Figure 5e), having better predictions within the statistical error, in comparison to the other models (Figures S28 and S29). Thus, the microstates are coarse-grained to eight macrostates with PCCA++ (Figure 5f,g), which appear to produce a fine coarse-grained MSM for the assumed eight states of AngII. Finally, the transition probabilities for each macrostate are shown over the arrows in the kinetic model estimated by an HMM using the eight metastable states found by PCCA++ (Figure 5g).

The representatives shown in Figure 5g are selected using GROMOS clustering, with a cutoff of 0.15 nm, at samples of 500 conformations from each macrostate. For each macrostate, the dominant representative is shown, based on the cluster's population (the top two representatives are shown in Figure S34 with the number of neighbors in each cluster, for clarity). Notice that each metastable state in Figure 5g corresponds to a probability density graph in Figure 5f, which shows the structural preference of the peptide in that state. The metastable states may exhibit conformational heterogeneity, since they are produced by coarse-graining, simplifying, MD simulations, combining both structural and kinetic similarity of conformations, and thus dominant cluster representatives should be taken into account. More plots and details for the selection of parameters, the construction of MSMs for AngII in water and water−ethanol at 310 K, and the resulting models (including plots such as metastable states against time, probability density plots for the macrostate samples used for GROMOS, and the top two macrostate representatives with their number of neighbors in cluster) can be found in Figures S18−S46.

Based on our kinetic model, the presence of ethanol does not seem to affect the kinetics of the folding/unfolding events. The peptide appears to present a folding/unfolding behavior in both environments, jumping between folded (U-shaped) and unfolded (Intermediate, Extended) states. Only the preference of AngII toward more compact structures appears to be affected in the presence of ethanol. Figure 6 shows the resulting MSM for AngII in water at 310 K (the detailed steps are shown in Figures S35−S46). The comparison of IQ-means with *k*-means on TICA coordinates and the MSM using IQ-means discretization is shown in Figures S47−S63. Also, further calculations on the generation of the macrostates using BACE before PCCA++ are shown in Figures S64−S70.

**MSM of AngII in Pure Water and in Water/Ethanol Using IQ-Means.** Before generating the MSMs with IQ-means discretization, we compare IQ-means to *k*-means clustering on the reduced dimensions produced by TICA. We apply both clustering methods for various number of clusters (microstates) on the TICA coordinates of AngII in water/ethanol at 310 K. The results are compared in terms of time and quality. The quality is assessed using the average distortion, calculated by the cost function of *k*-means, which is the average squared distance of each point from its cluster's center. In Figure 7 (also Figure S47), IQ-means appears to outperform *k*-means in time for many clusters, with fair loss in distortion (1.5 times worse than *k*-means), making it a reasonable trade-off. Every experiment was performed three times for reliability and the mean values are reported. In Figure 7a there are three plots for IQ-means, because some steps can be considered as preprocessing. Only the clustering step is mandatory for every run. Experiments were run on a computer with an Intel(R) Xeon(R) CPU E5-2620 v4 at 2.10 GHz using 64 GB of RAM. More calculations to demonstrate the

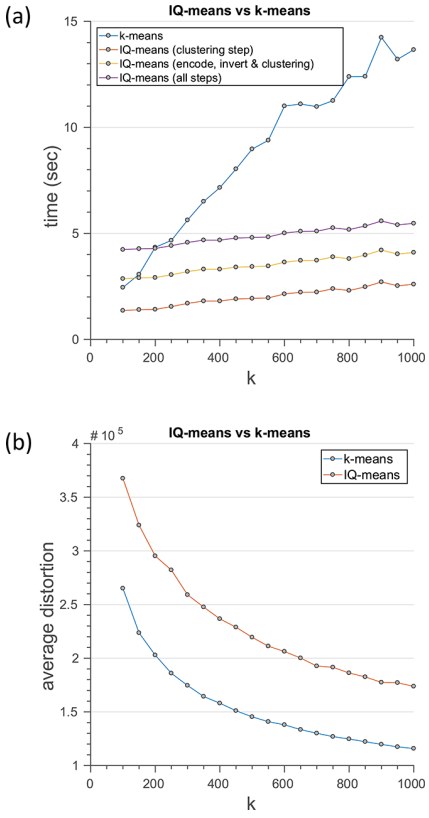



**Figure 7.** Comparison of IQ-means with *k*-means on TICA coordinates. (a) *k* (number of clusters) versus time (s). Max time for both methods is at 900 clusters, where *k*-means runs for 14.2 s, while IQ-means needs only 5.6 s for all steps (with preprocessing) from which only 2.7 s are used for clustering. Min time for both is at 100 clusters, where *k*-means runs for 2.4 s, while IQ-means for 4.2 s for all steps , but only 1.4 s for the clustering step. (b) Average distortion versus *k* (number of clusters).

efficiency of IQ-means against *k*-means on higher dimensions, such as backbone atoms distances are shown in Figure S9.

To validate the microstates produced by IQ-means, we compare the resulting metastable states (Figures 8 and 9), since similar discretization should result to analogous MSMs. The IQ-means microstates are generated using same features and TICA components with previous experiments (Figure 5a,b). The probability densities of the MSMs with *k*-means and IQ-means appear to be similar in both solvents at 310 K (Figures 8a,c and 9a,c), especially for the U-shaped metastable states. The MSM with IQ-means microstates appear to result to macrostates that correspond to the ones that were generated by *k*-means microstates. This can be also verified by the distribution of the macrostates through time. In Figure S52, conformations' end-to-end distances are plotted against time and colored according to their macrostate membership. The distribution of metastable states through time is similar for both MSMs (same colors are used for similar macrostates generated by both models). Thus, the structural preference of the peptide in the metastable states that were identified using *k*-means and IQ-means microstates, appear to be comparable. More details for the construction of MSM with IQ-means, and further comparisons to MSM with *k*-means (for AngII in water and water−ethanol at 310 K) are provided in Figures S48−S63.

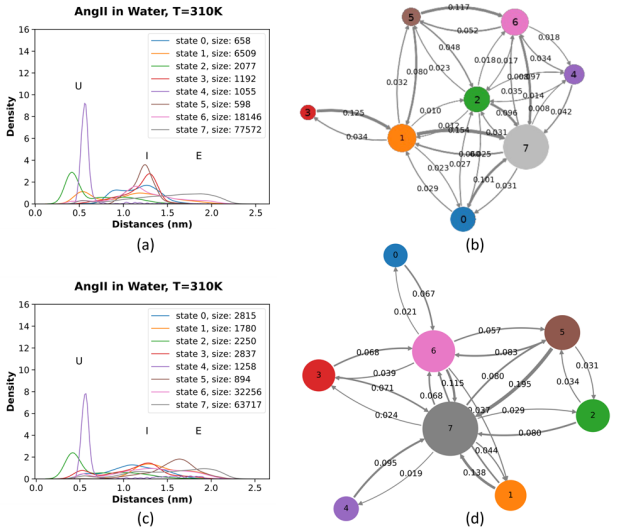



**Figure 9.** MSM for AngII in water, $T$ = 310 K. (a,b) Microstates generated with *k*-means. (a) Probability density plots for each macrostate. (b) Kinetic model of the MSM. (c,d) Microstates generated with IQ-means. (c) Probability density plots for each macrostate. (d) Kinetic model of the MSM. Probability densities of the MSMs appear to be similar especially for the U-shaped metastable states.

**Metadata Representation.** Metadata representations, as proposed by Zhang et al.,[44] are generated by applying MDS on a matrix with 276 distances between the, totally, 24 backbone atom distances (Figure 10a shows a sample of the distances) for each conformation of AngII in water/ethanol at 310 K. Coloring each 3D point by the corresponding end-to-end distance in Figure 10b, we observe that the representations appear to distinguish the key structures of AngII (U-shaped,

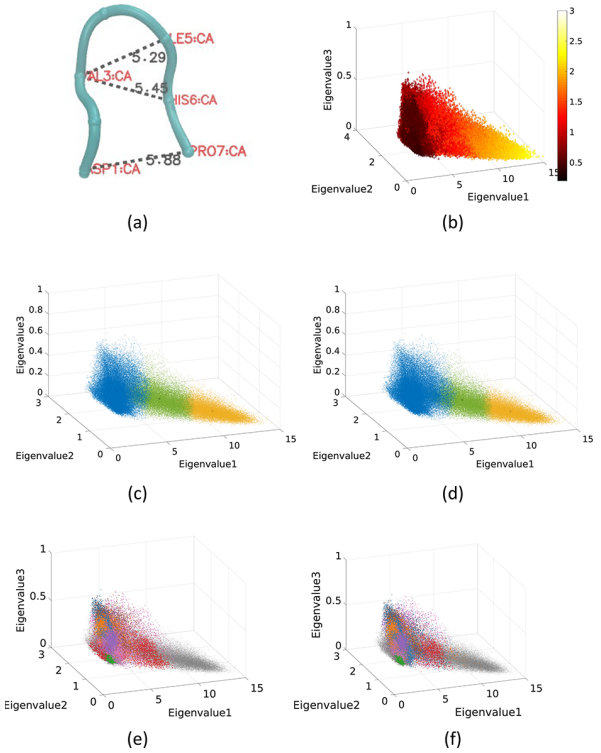



**Figure 10.** Metadata representatives. (a) Illustrates 3 from the 276 backbone atom distances used for MDS. (b) 3D point metadata representations colored by end-to-end distances; red/black points (smaller end-to-end distances) are U-shaped conformations; yellow/white points (bigger end-to-end distances) are Intermediate (I) and Extented (E). (c,d) Comparison of (c) *k*-means and (d) IQ-means on these representations with three cluster centers. (e,f) Distribution of metastable states from MSMs (AngII in water/ethanol, 310 K), shown on the metadata representatives of the conformations. (e) MSM with *k*-means. (f) MSM with IQ-means. Metastable states appear to be similarly distributed on metadata feature space.

Intermediate, and Extended structures). Conformations with similar end-to-end distances are concentrated together, which verifies the representatives Also, we further exploit these representations to visualize the resemblance of resulting clusterings using *k*-means and IQ-means. Figure 10c,d presents the results of clustering with three centers, which appear to be similar. Also the metastable states from the constructed MSMs with *k*-means and IQ-means are validated on these representatives. In Figure 10e,f, we notice how the metastable states are distributed to structurally similar conformations. More clustering calculations, using these representations with *k*-means, IQ-means, and a hierarchical fuzzy *c*-means algorithm[44] are shown in Figures S11−S17.

**Diffusion of AngII Peptide in Water and in Water/Ethanol.** In Tables 1 and 2, the calculated translational diffusion coefficients for the AngII octapeptide and for ethanol are presented. Diffusion coefficients were calculated from mean squared displacements (see Methods and Figure S71). Comparison with previous available experimental and simulation data shows a very good agreement, which allows us to explore solvent structural properties in a more detail.

**Preferential Binding of Solvent Molecules at the Peptide Surface.** It is well known that preferential interactions between the peptide residues and the different solvent components are important for peptide/protein

**Table 1. Diffusion of AngII $D_{peptide} \times 10^{-5}$ cm$^2$ s$^{-1}$**

| temperature | in water | in water/ethanol | simulation of ref 13[a] |
|---|---|---|---|
| 278 K | 0.1846 (±0.0458) | 0.0579 (±0.0025) | 0.053 (±0.007) |
| 298 K | 0.6045 (±0.0245) | 0.3497 (±0.0899) | 0.11 (±0.02) |
| 310 K | 0.9412 (±0.3155) | 0.1563 (±0.0141) | |
| 323 K | 2.7192 (±0.6804) | 0.4776 (±0.1569) | |

[a]Simulations and experiments for AngII in ethanol/water-35% (v/v), reported in ref 13.

**Table 2. Diffusion of Ethanol Molecules $D_{ethanol} \times 10^{-5}$ cm$^2$ s$^{-1}$**

| temperature | in water/ethanol | simulation of ref 13[a] | exp.[a] |
|---|---|---|---|
| 278 K | 0,7344 (±0,0034) | 0.342 (±0.013) | 0.22 |
| 298 K | 12809 (±0,0105) | 0.839 (±0.022) | 0.68 |
| 310 K | 17711 (±0,0645) | | |
| 323 K | 23184 (±0,0908) | | |

[a]Simulations and experiments for AngII in ethanol/water-35% (v/v), reported in ref 13.

structure and dynamics.[14,16,47] Therefore, the RDFs of important residues/atoms of the system have been calculated for all systems at all temperatures. The RDF between the center of mass (COM) of each residue in the peptide and the atoms of the solvent and cosolvent has been calculated. The temperature dependence of the RDF is shown in Figure S71 for the two terminal residues. According to the results, the temperature increase does not affect the positions of the peaks and the valleys of the RDFs but only slightly their heights and depths. Furthermore, comparison of the RDFs of the residues COM with the oxygen atom of water and ethanol molecules at the same temperature clearly indicates differences at the heights but not the positions of the first peaks (Figure S72). A slight preference for water molecules is observed for Phe8, where the RDF for the oxygen of water molecules starts at a distance of 0.3 nm, indicating probably the formation of a hydrogen bond (Figure S72, last figure). In addition, comparison of the RDFs of peptide residues COM with water molecules in pure water and water molecules in the binary water/ethanol solvent does not indicate any change in the positions of the peaks and valleys that could be due to the presence of the ethanol (data not shown). In simulations of [val5]AngII in water/ethanol (35% v/v), Gerig[13] has reported preferential solvation of peptide hydrogens by ethanol molecules. More specifically, he reports that hydrogen atoms of valine and tyrosine side chains preferentially interact with ethanol molecules, while hydrogens of Arg1, Arg2, Phe8, and His6 side chains are in a solvent environment rich in water molecules. In our system, which is similar to the one of Gerig[13] except for the fact that we have used the human [Ile5]AngII analogue, we have assessed the preferential solvation of the peptide by calculating the RDF of the COM of each peptide residue with the $O_{water}$ and $O_{ethanol}$ of the solvent. Integration of the RDF plots up to the first valleys gives an indication of the coordination numbers for the first solvation shell around each residue (Figure S72). As seen in Table 3, most of the peptide residues interact mainly with water molecules except for the apolar residues of Val3 and Ile5. The strong interaction of the peptide with water is seen also in the number of hydrogen bonds formed between the peptide and the solvent molecules (Figure S73). There are on average ∼15 hydrogen bonds

**Table 3. Coordination Numbers of each Peptide Residue with $O_{ethanol}$ and $O_{water}$ in the First Solvation Shell for the System of AngII in Water/Ethanol**

| | Res. 1 | Res. 2 | Res. 3 | Res. 4 | Res. 5 | Res. 6 | Res. 7 | Res. 8 |
|---|---|---|---|---|---|---|---|---|
| $O_{water}$ | 16 | 11 | 2 | 14 | 2 | 14 | 15 | 12 |
| $O_{ethanol}$ | 2 | 2 | 1 | 8 | 1 | 2 | 3 | 2 |

formed between the peptide and water and only ∼2 hydrogen bonds between the peptide and ethanol molecules.

**Solvent Structural Properties.** As mentioned, one of the aims of the present study is to provide an atomic-level picture of the intermolecular interactions governing AngII−water and AngII−water−ethanol mixtures using MD simulations. To describe the solvent microenvironment around the AngII peptide in the two solvents, we performed polar analysis interactions of the solvent around AngII.

The hydrophobic clustering of ethanol molecules in bulk aqueous solutions has been experimentally investigated[48−51] showing that ethanol is primarily monomeric for mole fractions of $\chi_{eth} < 0.07$, while it self-associates when $0.07 < \chi_{eth} < 0.74$. Mass spectrometry analysis indicates that the ethanol rich clusters consist of less than eight ethanol molecules.[50] In our system, where $\chi_{eth}$ is around 0.14, we do observe minor ethanol aggregation. The percentage of ethanol molecules that self-associate is around 14% and the average size of the clusters formed is two ethanol molecules independent of the temperature. Larger aggregates of ethanol are also observed comprising three to four molecules but with a low percentage (1−3%) of occurrence. A detailed analysis of the polar interactions between water and ethanol molecules has been performed for an area with $r$ = 0.4 nm around the peptide. Figure S74 presents the number of hydrogen bonds formed between water−water, ethanol−ethanol, and water−ethanol molecules in the proximity of the peptide. There are on average two hydrogen bonds formed between ethanol molecules around the peptide, while the ethanol−water hydrogen bonds are >15 indicating a clear preference of ethanol to associate with water molecules in the vicinity of AngII.

**Intrapeptide Hydrogen Bonds.** To further characterize the microenvironment of AngII, we calculated the percentages of intrapeptide hydrogen bond formation in pure water and in water/ethanol that are presented in Table S4. As mentioned before, during all different simulations, the peptide is mainly in the open, extended conformation where the formation of intrapeptide hydrogen bonds does not seem to be favored. However, it is important to mention the occurrence of some intrapeptide hydrogen bonds that are characteristic of the AngII octapeptide and compare these with known experimental models. Therefore, common features between our model and the X-ray structure of AngII bound to mAb Fab13126 as well as the NMR structure of AngII in aqueous solution are (a) the formation of a hydrogen bond between Asp1-side-OD1 and Arg2-main-NH. In fact, in our simulations we additionally observed Asp1-Arg2 side chain-side chain hydrogen bonds resulting in a total percentage of occurrence of the Asp1-Arg2 hydrogen bond of about 1−4% in pure water and 3−12% in water/ethanol (see Table S4), (b) the formation of a His6-side-ND1 and Pro7-main-O hydrogen bond with a percentage of occurrence of 6−13% in water and 2−8% in water/ethanol. Interestingly, the Asp1-Arg2 and His6-Pro7 hydrogen bonds discussed above are common for all simulated systems, sampling both extended and U-shaped conformations, although the former is marginally present in pure water. A hydrogen bond that appears only in the water/ethanol system at elevated temperatures (310, 323 K) is the one formed between Arg2 and Phe8 residues with an occurrence of 3−8%. This hydrogen bond could bring the two peptide termini in a close distance and stabilize a compact U-shape structure (Figure 4v).

## CONCLUSIONS

In this paper, we provide an atomic-level picture of the intermolecular interactions governing AngII−water and AngII−water−ethanol mixtures using MD simulations. Both mixtures are solvents that can influence the stability and conformational properties of biological molecules in diverse ways. The structural and dynamical properties of these systems are outlined in the context of identifying the solvent microenvironment around the AngII peptide and in describing the representative conformations of AngII in these two media. While the phospholipid bilayer is more polarized and its 2D structure may influence the AngII conformation and dynamics, here we have chosen to perform our simulations in water and water−ethanol mixtures to compare our findings with available experimental results in this solvent mixture. Our results are in excellent agreement with relevant experimental data and provide insights into the bioactive conformation of this hormone.

In order to increase sampling for the construction of MSMs, 10 different conformations were chosen for each system to initiate new, independent simulations. Results validate that the samples for the various structures of AngII were significantly increased using multiple trajectories, while using a single trajectory resulted to mainly an extended coil structure. Here, we are interested in whether we have enough structural data to characterize the free-energy landscape of the AngII folding and kinetically connect the different states with a Markov state model. By using multiple trajectories with different starting configurations, we increase the conformational variability of the sample and, as evident by the good discretization of states shown in Figure 5b, this is achieved with the MSM model produced by this sampling.

Our cluster analysis, in agreement with previous studies, validates that the dominant conformation of AngII in water and water/ethanol is mainly an extended structure, but at elevated temperatures there is a small population of folded (U-shaped) conformations. Also, the RMSD fluctuations and the radius of gyration of AngII indicate that the folding/unfolding events of AngII at elevated temperatures (310, 323 K) seem to be significantly more frequent in the binary solvent (water/ethanol) than in water. The kinetic models produced by MSMs also confirm that the presence of ethanol does not seem to affect the kinetics of the folding/unfolding events but only the preference of AngII toward more compact structures. This points to the fact that because the water−ethanol interface mimics the water−lipid interface, the water−membrane model interface induces the peptide to assume its bioactive conformation, as the peptide has to assume the bioactive conformation before it docks to the receptor.

While the kinetic models of AngII in water and in water/ethanol may differ, based on our kinetic model, the presence of ethanol does not seem to affect the kinetics of the folding/unfolding events. This means that the peptide appears to present a folding/unfolding behavior in both environments.

The presence or absence of ethanol does not seem to affect the folding events, since the peptide appears to have structurally similar folded (U-shaped) and unfolded (intermediate, extended) states, jumping from one to another. Only the frequency of the folding and the preference of the peptide toward more compact (folded) structures are affected having more conformations in a U-shaped structure in the water/ethanol mixtures.

Experiments with IQ-means suggest that it could consist an efficient approximation for *k*-means in discretizing MD trajectories for MSMs, with a fair trade-off between time and accuracy. IQ-means could be used for long simulations resulting in big data to reduce the computational effort, either by replacing *k*-means as a proper approximation, or by performing fast approximations to experiment with the microstates, before using *k*-means. Also, GROMOS seems to provide intuitive representatives for the MSMs and the metadata representations are able to produce simple, straightforward 3D point visualizations for the trajectories.


### Corresponding Authors

**Ioannis Z. Emiris** − *ITMB, Department of Informatics & Telecommunications, National and Kapodistrian University of Athens, Athens 15772, Greece; Athena Research Center, Marousi 15125, Greece*; Email: emiris@di.uoa.gr

**Zoe Cournia** − *ITMB, Department of Informatics & Telecommunications, National and Kapodistrian University of Athens, Athens 15772, Greece; Biomedical Research Foundation, Academy of Athens, Athens 11527, Greece*; orcid.org/0000-0001-9287-364X; Email: zcournia@bioacademy.gr

### Authors

**Emmanouil Christoforou** − *ITMB, Department of Informatics & Telecommunications, National and Kapodistrian University of Athens, Athens 15772, Greece; Biomedical Research Foundation, Academy of Athens, Athens 11527, Greece*

**Hari Leontiadou** − *Biomedical Research Foundation, Academy of Athens, Athens 11527, Greece*

**Frank Noé** − *Fachbereich Mathematik und Informatik, Freie Universität Berlin, Berlin 14195, Germany*

**Jannis Samios** − *Department of Chemistry, Laboratory of Physical Chemistry, National & Kapodistrian University of Athens. Athens 15772. Greece*


## ■ ACKNOWLEDGMENTS


This research work was also supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 1780) awarded to Z.C. Computational time was granted from the Greek Research & Technology Network (GRNET) in the National HPC facility ARIS, under project ID pr007007/a-syn. E.C. and I.Z.E. would like to acknowledge support of this work by the project "INSPIRED-The National Research Infrastructures on Integrated Structural Biology, Drug Screening Efforts and Drug target functional characterization" (MIS 5002550) which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure," funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund). Data for this manuscript can be donwloaded at 10.5281/zenodo.5887613.


## ■ REFERENCES


(1) Lavoie, J. L.; Sigmund, C. D. Minireview: Overview of the Renin-Angiotensin System—An Endocrine and Paracrine System. *Endocrinology* **2003**, *144*, 2179−2183.

(2) Unal, H.; Karnik, S. S. Constitutive activity in the angiotensin II type 1 receptor: discovery and applications. *Adv. Pharmacol.* **2014**, *70*, 155−174.

(3) Spyroulias, G. A.; Nikolakopoulou, P.; Tzakos, A.; Gerothanassis, I. P.; Magafa, V.; Manessi-Zoupa, E.; Cordopatis, P. Comparison of the solution structures of angiotensin I & II. Implication for structure-function relationship. *Eur. J. Biochem.* **2003**, *270*, 2163−2173.

(4) Kawai, T.; Forrester, S. J.; O'Brien, S.; Baggett, A.; Rizzo, V.; Eguchi, S. AT1 receptor signaling pathways in the cardiovascular system. *Pharmacol. Res.* **2017**, *125*, 4−13.

(5) Zhang, H.; Unal, H.; Gati, C.; Han, G. W.; Liu, W.; Zatsepin, N. A.; James, D.; Wang, D.; Nelson, G.; Weierstall, U.; Sawaya, M. R.; Xu, Q.; Messerschmidt, M.; Williams, G. J.; Boutet, S.; Yefanov, O. M.; White, T. A.; Wang, C.; Ishchenko, A.; Tirupula, K. C.; Desnoyer, R.; Coe, J.; Conrad, C. E.; Fromme, P.; Stevens, R. C.; Katritch, V.; Karnik, S. S.; Cherezov, V. Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell* **2015**, *161*, 833−844.

(6) Matsoukas, J. M.; Hondrelis, J.; Keramida, M.; Mavromoustakos, T.; Makriyannis, A.; Yamdagni, R.; Wu, Q.; Moore, G. J. Role of the NH2-terminal domain of angiotensin II (ANG II) and [Sar1]-angiotensin II on conformation and activity. NMR evidence for aromatic ring clustering and peptide backbone folding compared with [des-1,2,3]angiotensin II. *J. Biol. Chem.* **1994**, *269*, 5303−5312.

(7) Garcia, K. C.; Ronco, P. M.; Verroust, P. J.; Brünger, A. T.; Amzel, L. M. Three-dimensional structure of an angiotensin II-Fab complex at 3 A: hormone recognition by an anti-idiotypic antibody. *Science* **1992**, *257*, 502−507.

(8) Boucard, A. A.; Wilkes, B. C.; Laporte, S. A.; Escher, E.; Guillemette, G.; Leduc, R. Photolabeling identifies position 172 of the human AT(1) receptor as a ligand contact point: receptor-bound angiotensin II adopts an extended structure. *Biochemistry* **2000**, *39*, 9662−9670.

(9) Chatterjee, C.; Martinez, D.; Gerig, J. T. Interactions of trifluoroethanol with [val5]angiotensin II. *J. Phys. Chem. B* **2007**, *111*, 9355−9362.

(10) Matsoukas, M. T.; Potamitis, C.; Plotas, P.; Androutsou, M. E.; Agelis, G.; Matsoukas, J.; Zoumpoulakis, P. Insights into AT1 receptor activation through AngII binding studies. *J. Chem. Inf. Model.* **2013**, *53*, 2798−2811.

(11) Tzakos, A. G.; Bonvin, A. M.; Troganis, A.; Cordopatis, P.; Amzel, M. L.; Gerothanassis, I. P.; van Nuland, N. A. On the molecular basis of the recognition of angiotensin II (AII). NMR structure of AII in solution compared with the X-ray structure of AII bound to the mAb Fab131. *Eur. J. Biochem.* **2003**, *270*, 849−860.
(12) Fermandjian, S.; Morgat, J. L.; Fromageot, P. Studies of angiotensin-II conformations by circular dichroism. *Eur. J. Biochem.* **1971**, *24*, 252−258.
(13) Gerig, J. T. Investigation of Ethanol−Peptide and Water−Peptide Interactions through Intermolecular Nuclear Overhauser Effects and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117*, 4880−4892.
(14) Gerig, J. T. Solvent Interactions with [Val5]angiotensin II in Ethanol−Water. *J. Phys. Chem. B* **2008**, *112*, 7967−7976.
(15) Ghosh, R.; Roy, S.; Bagchi, B. Solvent Sensitivity of Protein Unfolding: Dynamical Study of Chicken Villin Headpiece Subdomain in Water−Ethanol Binary Mixture. *J. Phys. Chem. B* **2013**, *117*, 15625−15638.
(16) Ortore, M. G.; Mariani, P.; Carsughi, F.; Cinelli, S.; Onori, G.; Teixeira, J.; Spinozzi, F. Preferential solvation of lysozyme in water/ethanol mixtures. *J. Chem. Phys.* **2011**, *135*, 245103.
(17) Roccatano, D.; Colombo, G.; Fioroni, M.; Mark, A. E. Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: A molecular dynamics study. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12179.
(18) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146−168.
(19) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a $\beta$-Hairpin Peptide. *J. Phys. Chem. B* **2004**, *108*, 6582−6594.
(20) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.* **2007**, *126*, 155102.
(21) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.
(22) Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.
(23) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.
(24) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(25) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.
(26) Jorgensen, W. L.; Blake, J. F.; Buckner, J. K. Free energy of TIP4P water and the free energies of hydration of CH4 and Cl- from statistical perturbation theory. *Chem. Phys.* **1989**, *129*, 193−200.
(27) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(28) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(29) Prinz, J.-H.; Keller, B.; Noé, F. Probing molecular kinetics with Markov models: metastable states, transition pathways and spectroscopic observables. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912−16927.
(30) Bowman, G. R.; Huang, X.; Pande, V. S. Network models for molecular kinetics and their initial applications to human health. *Cell Res.* **2010**, *20*, 622−630.
(31) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* **2009**, *131*, 124101.
(32) Noé, F.; Clementi, C. Kinetic distance and kinetic maps from molecular dynamics simulation. *J. Chem. Theory Comput.* **2015**, *11*, 5002−5011.
(33) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, No. 015102.
(34) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(35) Harrigan, M. P.; Sultan, M. M.; Hernández, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; McGibbon, R. T.; Pande, V. S. MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys. J.* **2017**, *112*, 10−15.
(36) Emiris, I. Z.; Avrithis, Y.; Kalantidis, Y.; Anagnostopoulos, E. In *Web-scale image clustering revisited*, ICCV 2015 - International Conference on Computer Vision, Santiago, Chile, December 11, 2015; Santiago: Chile, 2015.
(37) Arthur, D.; Vassilvitskii, S., k-means++: the advantages of careful seeding. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2007; pp 1027−1035.
(38) Avrithis, Y.; Kalantidis, Y. In *Approximate Gaussian Mixtures for Large Scale Vocabularies*, Computer Vision − ECCV, Fitzgibbon, A.; Lazebnik, S.; Sato, Y.; Schmid, C., Eds.; Springer, 2012.
(39) Bowman, G. R. Improved coarse-graining of Markov state models via explicit consideration of statistical uncertainty. *J. Chem. Phys.* **2012**, *137*, 134111.
(40) Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data Anal. Classif.* **2013**, *7*, 147−179.
(41) Deuflhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* **2003**, *398*, 161−184.
(42) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *J. Chem. Phys.* **2013**, *139*, 184114.
(43) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(44) Zhang, B.; Estrada, T.; Cicotti, P.; Taufer, M. In *Enabling In-Situ Data Analysis for Large Protein-Folding Trajectory Datasets*, 2014 IEEE 28th International Parallel and Distributed Processing Symposium, 19−23 May 2014; pp 221−230.
(45) Carpenter, K. A.; Wilkes, B. C.; Schiller, P. W. The octapeptide angiotensin II adopts a well-defined structure in a phospholipid environment. *Eur. J. Biochem.* **1998**, *251*, 448−453.
(46) Jalili, P.; Dass, C. Determination of the structure of lipid vesicle-bound angiotensin II and angiotensin I. *Anal. Biochem.* **2008**, *374*, 346−357.
(47) Fioroni, M.; Diaz, M. D.; Burger, K.; Berger, S. Solvation phenomena of a tetrapeptide in water/trifluoroethanol and water/ethanol mixtures: a diffusion NMR, intermolecular NOE, and molecular dynamics study. *J. Am. Chem. Soc.* **2002**, *124*, 7737−7744.
(48) Matsumoto, M.; Nishi, N.; Furusawa, T.; Saita, M.; Takamuku, T.; Yamagami, M.; Yamaguchi, T. Structure of Clusters in Ethanol-Water Binary Solutions Studied by Mass Spectrometry and X-Ray Diffraction. *Bull. Chem. Soc. Jpn.* **1995**, *68*, 1775−1783.
(49) Nishi, N.; Koga, K.; Ohshima, C.; Yamamoto, K.; Nagashima, U.; Nagami, K. Molecular association in ethanol-water mixtures studied by mass spectrometric analysis of clusters generated through adiabatic expansion of liquid jets. *J. Am. Chem. Soc.* **1988**, *110*, 5246−5255.