

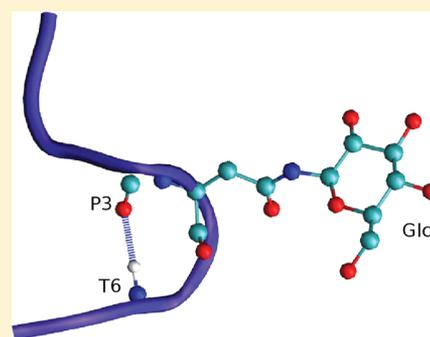
Conformational Landscape of N-Glycosylated Peptides Detecting Autoantibodies in Multiple Sclerosis, Revealed by Hamiltonian Replica Exchange

Carlo Guardiani,[†] Giorgio F. Signorini,[†] Roberto Livi,[‡] Anna Maria Papini,[§] and Piero Procacci^{*,†}

[†]Department of Chemistry, [‡]Department of Physics and Center for the Study of Complex Dynamics (CSDC), and [§]Laboratory of Peptide & Protein Chemistry & Biology, University of Florence, Sesto Fiorentino, Italy 50019

S Supporting Information

ABSTRACT: Synthetic N-glycosylated CSF114(Glc) and related peptides were proved to be able to recognize specific and high-affinity autoantibodies circulating in blood of relapsing-remitting multiple sclerosis (MS) patients and correlating with disease activity. The effect of these peptides has been linked to the β -turn structure around the minimal epitope Asn(Glc). In this work we performed Hamiltonian replica exchange molecular dynamics simulations on the central heptapeptide fragment of a CSF114(Glc)-derived peptide in water and in a water/hexafluoroacetone mixture, confirming a significant incidence of β -turn structures in both solvents. The structural similarity of the glycosylated and unglycosylated forms in all environments proves that the conformation of the heptapeptide is only marginally affected by the presence of the sugar. Moreover, the presence of a significant amount of bioactive hairpin-like conformations in the water environment suggests a possible use not only in the diagnosis but also in the treatment of MS.



INTRODUCTION

Multiple sclerosis (MS) is a chronic, highly disabling disease characterized by an autoimmune reaction destroying the myelin sheath that electrically insulates neurons.¹ Subsequent problems in the transmission of nerve signals result in a situation of progressive paralysis that in some cases determines a premature death. Multiple sclerosis currently affects over 400 000 patients in Europe,² typically in the most productive age of life (20–40). The progressive impairment of work ability and constant efforts required from caregivers both contribute to the high social burden of this disease.

While a specific therapy is not yet available, an early treatment with interferon β -1b was proved to delay progression of the disease,³ highlighting the importance of early diagnosis. Diagnosis of MS, however, is far from trivial since the initial symptoms are heterogeneous and often mild. Magnetic resonance imaging (MRI), a technique allowing physicians to identify lesions in the patients' brain, proved to be a powerful diagnostic tool.⁴ Unfortunately the high cost of the equipment for MRI prevents the use of this technique for a large-scale screening of the population. This makes immunologic assays a viable and an attractive alternative. Immunoassays are based on the evidence that autoantibodies appear in the blood long before the onset of clinical symptoms,⁵ so that monitoring of these biomarkers not only enables an early diagnosis but also anticipates relapses of the disease, allowing the physician to take action to delay or alleviate the effects of the attack.

A significant headway along this line of investigation was achieved by Papini and co-workers, who developed a synthetic glycopeptide, CSF114(Glc), capable of identifying autoanti-

bodies in sera of patients of a MS disease form possibly corresponding to the relapsing-remitting form.⁶ The seminal observation leading to the design of CSF114(Glc) was that while the immuno-dominant epitope of MOG [Asn³¹(Glc)] hMOG(30–50) was able to detect autoantibodies in MS patients, its unglycosylated analogue was inactive.⁷ This suggested that the glucose moiety could be possibly an aberrant post-translational modification reminiscent of a bacterial and/or viral infection, leading to autoantibodies production. The hypothesis of a direct interaction of the Glc unit with the autoantibodies was confirmed by the fact that the glycosylated and unglycosylated variants of hMOG(30–50) adopted similar solution conformations,⁸ thus ruling out the possibility that the presence of the sugar could stabilize a different native state.

Recognition of the importance of the sugar led Papini to engineer the sequence of hMOG(30–50) in order to stabilize the β -turn structure around the minimal, fundamental epitope Asn(Glc).

The resulting antigenic probe, CSF114(Glc), proved to be able to recognize, by ELISA, the presence of specific IgM autoantibodies in the sera of a MS patient population but not in blood donors and other autoimmune conditions.

The affinity of CSF114(Glc) for MS autoantibodies was further substantially increased by changing the amino-acid sequence around the glycosylated residue to ones that are

Received: February 13, 2012

Revised: April 17, 2012

Published: April 20, 2012

known to stabilize β -turn type I, II, or II'.⁹ Moreover, it was proved that the N-glycosylated heptapeptide [*Ac*-ERP(*N*H₂)-HTV-NH₂], apical fragment of an engineered version of CSF114(*Glc*), retains the immunological properties of the parental peptide and the active β -hairpin conformation, as revealed by solid-phase ELISA.

The structure of CSF114(*Glc*) and the derived peptides was determined through NMR spectroscopy in two different solvents:^{6,9} water and a 50% vol/vol mixture of water and hexafluoroacetone (HFA). While in pure water the peptides under investigation have been found mainly unstructured, in the water/HFA mixture they are believed to adopt the active hairpin conformation. The use of fluoro-alcohols as stabilizing agents for peptides is a common practice in biochemistry^{10,11} that relies upon the assumption that the fluoro-alcohol mimics some specific biological or experimental environment. In the case of the family of peptides derived from CSF114(*Glc*) it was suggested that the water/HFA solution might reproduce the environment of the solid substrate of the ELISA test.¹² However, interpretation of NMR spectra of highly flexible molecules such as the CSF114(*Glc*)-derived peptides in solution could be a rather challenging task since torsions about a chemical bond (that occur on a time scale of nanoseconds) cause most NMR parameters (e.g., NOE, coupling constant, chemical shifts) to be averaged out rather than giving a superposition of values as, e.g., in optical spectroscopy.¹³ As a consequence, it might well be that CSF114(*Glc*) adopts the β -hairpin conformation also in water, where, however, the structure is not detectable through NMR because the conformations interconvert too rapidly.

This issue can effectively be addressed through an accurate and unbiased exploration of the equilibrium ensemble of the glycosylated heptapeptide in the two solvents by means of accurate molecular dynamics simulations at the atomistic level. Also, the effect of the glucose unit on the structure of the heptapeptide can be assessed by comparing the simulations of glycosylated and unglycosylated forms. As stated above, such engineered peptide is as effective as its parent CSF114(*Glc*) compound in immunological assays and, due to its small size, lends itself well to a computational study. A realistic and detailed description of the structural and kinetic properties of *Ac*-ERP(*N*H₂)-HTV-NH₂, such as that provided in the present study, may be of great help in rationalizing the bioactivity of this class of MS-related peptides.

Efficient sampling of the conformational space of proteins and peptides using molecular dynamics (MD) simulations is a challenging task because these molecules are characterized by a high-dimensional, rugged energy landscape featuring a huge number of local minima that may act as kinetic traps. This problem can be overcome using generalized ensemble methods such as replica exchange molecular dynamics (REMD) that allows a random walk in energy and temperature.¹⁴

In this study, at variance with a previous work on solvated peptides,¹⁵ we adopted a Hamiltonian REMD approach, based on the unequal scaling of the potential function terms.¹⁶ The results of our REMD simulations are in good agreement with available experimental data.⁹ When the glycosylated and unglycosylated heptapeptides were simulated in the same solvent, we found a significant superposition for the probability distributions of all the structural parameters we monitored. This is due to the fact that the glucose unit, being rich in polar hydroxyl groups, interacts preferentially with the water without interfering too much with the folding of the peptide. This

suggests that the sugar has only a minor effect on the folding of the chain, supporting the idea of a direct interaction between the glucose ring and the binding pocket of the MS autoantibodies.

More importantly, comparison of the equilibrium ensembles of the glycosylated peptide in water and water/HFA revealed that the fraction of hairpin-like conformations is basically the same in the two solvents. The expected structuring effect of HFA, in fact, does not increase the fraction of hairpin conformations but rather determines the appearance of a motif with a curved peptide chain and the sugar pointing toward the interior of the loop. Our data thus suggest that the NMR measurements may have underestimated the amount of hairpin conformations in water, possibly due to the short times of conformational transition. This hypothesis was confirmed by a kinetic clustering of the population using a Markov state model.¹⁷ This approach showed that in the solvent mixture a large amount of the hairpin conformations are sequestered in a metastable state separated by a high energy barrier from another metastable state including the rest of the population. By contrast, in water only a single metastable state exists and hairpin conformations can be turned into any other structural motif at a very high rate. The presence of a high fraction of hairpin conformations in the water simulations may have important pharmaceutical implications, suggesting the use of CSF114(*Glc*) derivatives not just as probes of MS biomarkers in diagnostic assays but as drugs for treatment of the disease.

The paper is organized as follows. In the Methods section we provide details about our computational protocol. In the Results section we comment on the main results of our simulations, starting with a discussion on the conformational ensembles of the bioactive glycopeptide in the two solvents, then performing a kinetic analysis by means of a Markov state model, and finally reviewing the main results concerning the effect of glycosylation of the peptide on structural features. In the Conclusions section we draw the conclusions of our work.

METHODS

In Table 1 we report the sequence of the heptapeptide system simulated in this study and the nomenclature that will be hereafter adopted.

Table 1. Peptide Sequences and Abbreviations Used in the Text

peptide sequence	abbreviation
<i>Ac</i> -ERPNTV-NH ₂	Hepta
<i>Ac</i> -ERP(<i>Glc</i>)HTV-NH ₂	N ⁴ (<i>Glc</i>)Hepta

Molecular Dynamics. Four systems have been set up and simulated, namely, the glycosylated and unglycosylated heptapeptide in water and a 50% in volume mixture of water and HFA. All simulations have been performed with the ORAC suite of programs^{16,18} using the Amber ff99SB force field¹⁹ and the TIP3P water model.²⁰ The parameters for the glucose unit have been taken from the GLYCAM06 force field for carbohydrates.²¹ At variance with previous versions, in GLYCAM06 the 1–4 electrostatic and nonbonded scaling factors were set to unity in order to correct for the unbalance of the O6–O4 and O6–O5 interactions determined by 1–4 scaling. Simulation of a glycopeptide system such as ours requires a mixing of the GLYCAM force field with the Amber amino acid force field. Following the suggestion of the Amber9

manual,²² we retained the scaling factors typical of protein systems (a 5/6 scaling factor for electrostatic interactions and a 0.5 scaling factor for van der Waals interaction). This might cause the loss of accuracy in the populations for the omega angle rotation (O5–C5–C6–O6) and in the corresponding barrier heights, but it affects only marginally the overall carbohydrate structure or stability (see Supporting Information, section Comparison GLYCAM06/Amber for β -D-glucopyranose in water solution). Force field parameters for the HFA were derived using the Antechamber package²² and are detailed in Tables 1–5 of the Supporting Information (section Parameters of HFA).

Peptides were built in the extended conformation using the xLEaP program²² and minimized in vacuo to remove atomic clashes. For the simulations in water the peptide was solvated with 972 water molecules in a simple cubic box with periodic boundary conditions. The system was then equilibrated during a 100 ps simulation in the isothermal–isobaric ensemble at $T = 300$ K and $P = 1$ atm. Constant pressure was obtained using a modification of the Parrinello–Rahman Lagrangian,²³ and temperature control was achieved using the Nosé thermostat.²⁴ Electrostatic interactions were computed using the smooth particle mesh Ewald algorithm with the convergence parameter set to 0.43 \AA^{-1} and a grid spacing of 1.2 \AA .²⁵ The equations of motion were integrated using a multiple time step r-RESPA algorithm²⁶ with a potential subdivision specifically tuned for biomolecular systems.^{18,23}

Before simulating the peptide in the water/HFA environment a preliminary simulation was run to test the solvent mixing. Details of this simulation along with a discussion on microsolvation effects are reported in the Supporting Information (section Microsolvation effects). Once the solvent mixing was ascertained, the glycosylated and unglycosylated peptides were immersed in a simple cubic box where the system containing 1 HFA·H₂O (geminal diol) and 10 water molecules had been replicated 64 times. Since the solvent molecules clashing with the peptide were discarded, the heptapeptide was finally solvated with 59 HFA and 590 water molecules. The system was then equilibrated in a 100 ps NPT run using the same protocol employed for equilibration in the water environment. The last conformation produced by these simulations was then used as the input for the REMD simulation.

REMD simulations were performed using the Hamiltonian REM approach.^{27,28} With Hamiltonian REMD one can compute the acceptance probability of the exchanges based on the energy of a relevant subset of the degrees of freedom of the investigated system. This is achieved by partitioning the energy in a number of additive contributions (in the ORAC implementation, bonded, torsional, and nonbonded terms) weighted by different scaling factors that change along the replica progression. In our case, the bonded term that is expected to play only a minor role in conformational changes was left unscaled, no longer contributing to the exchange probability that therefore depends only on the torsional and nonbonded degrees of freedom. Conversely, the scaling factor of the nonbonded potential was assigned exponentially decreasing values ranging from 1.0 (in the target replica) to 0.5, the latter corresponding to an effective temperature of 600 K. The scaling factor for the torsional term was chosen according to the specific environment where the peptide was simulated. For the water simulations we set the minimal scaling factor to 0.1 corresponding to a maximal effective temperature

of 3000 K, which enabled an effective overcoming of all energy barriers using only 24 replicas. This setting however turned out to be unsuitable for simulation in the water/HFA mixture since it led to an insufficient overlap between the torsional energy distributions of neighboring replicas. This is why in the solvent mixture the torsional scaling factor was increased to 0.3, thus decreasing the maximal effective temperature to 1000 K, and the number of replicas was raised to 32. The new setting determined a more extensive superposition between the torsional energy distributions, improving the acceptance ratio of the exchanges.

Exchanges were attempted every 250 fs, leading to an average acceptance ratio of 28%. The N replicas are sorted in an array based on their vectors of scaling coefficients. One-half of the replicas (with even array indices) are chosen as exchange initiators. These initiators pair with their right and left neighbors alternatively at each call of the REM subroutine, so that $N/2$ exchanges are attempted at each iteration. The total simulation length was 24 ns per replica for simulations in pure water and 92 ns per replica for simulations in water/HFA. The longer time required to reach convergence in the simulations in water/HFA is due to the higher viscosity of the HFA/water mixture with respect to water that results in a slowing down of the exploration of the conformational space. In all four REM simulations the first 25% of the trajectory of the target (unscaled) replica was discarded and analysis was performed on 3600 structures (for the simulations in water) and on 14 000 structures (for the simulation in water/HFA) sampled at regular time intervals. Extensive error analysis and convergence tests for all four REMD simulations have been reported in the Supporting Information (section Error analysis).

Quality Threshold Clustering. Structural insights on the equilibrium population of the target replica can be attained through a quality threshold clustering.²⁹ This algorithm requires a distance matrix for all pairs of structures of the population. The structural metrics that we chose was the maximum difference between corresponding pairs of carbon atoms

$$d_{S_m, S_n} = \max_{i,j} |d_{ij}(S_m) - d_{ij}(S_n)|$$

where d_{ij} is the distance between carbon atoms i and j belonging to structures S_m and S_n . A comparison of this metrics with the more customary RMSD can be found in the Supporting Information (see section Metrics and Clustering). For each structure in the ensemble the algorithm builds a candidate cluster in such a way that the distance between any two structures of the cluster does not exceed the cutoff distance (that we chose to be $d_{\text{cut}} = 6.0 \text{ \AA}$). The program then retains only the largest cluster and removes its structures from the population. The procedure is iterated until all structures of the populations are used.

Markov State Model. The dynamical characterization of a macromolecule in solution requires the determination of the long-lived, metastable states and the transition rates between them. Clustering algorithms group a population in subfamilies based on their geometric features that, however, might not necessarily correspond to kinetic properties. In fact, two conformations belonging to the same cluster may be geometrically close but kinetically distant, being separated by an energy barrier. Markov state models (MSM) perform a kinetic clustering, assigning to the same cluster only structures capable of fast conversion between each other.^{17,30} At first glance MSM

seem not to be applicable to generalized ensemble simulations that perform a random walk in temperature space and thus have no physical kinetics. However, the data of the target replica feature the canonical distribution at the temperature of interest, and thus, they contain information on all energy barriers responsible for separation of time scales. It is thus possible to identify metastable states (also referred to as macrostates) that are characterized by fast intrastate transitions and slow interstate transitions. Pande and co-workers recently developed software to extract dominant metastable states from generalized ensemble simulations (MSMBuilder package, available at <https://simtk.org/home/msmbuilder>). The algorithm first divides the population in a set of small microstates based on their geometric features and then lumps the microstates in larger metastable states based on kinetic features.³¹

The first step of the algorithm is based on the idea that as long as the cutoff radius of a cluster is sufficiently small two structures belonging to the same cluster are so structurally similar that they must quickly interconvert between each other. We performed this step using the quality-threshold clustering with the metrics $d_{\text{sm},\text{sn}}$ applied to carbon atoms only. We chose a cutoff of 6.0 Å, which leads to clusters with an average pairwise RMSD of carbon atoms of 1.5–2.0 Å (See Supporting Information, section Metrics and Clustering).

The second stage of the algorithm, lumping of microstates into macrostates, is based on the idea that if the equilibrium population is partitioned into k metastable states with fast intrastate kinetics and slow interstate kinetics, then the dynamics of the system can be modeled in terms of a nearly uncoupled Markov chain. In such a case it can be proved³² that the spectrum of the transition matrix can be divided into 3 parts: (i) the Perron root $\lambda = 1$; (ii) a cluster of $k - 1$ eigenvalues approaching 1; (iii) the remaining $n - k$ eigenvalues (n is the number of the microstates) with values very different from 1. The eigenvectors of the transition matrix will also be divided in three corresponding groups: (i) The eigenvector corresponding to the Perron root $\mathbf{X}_1 = (1, \dots, 1)$; (ii) the $k - 1$ eigenvectors corresponding to eigenvalues close to $\lambda = 1$. These eigenvectors are of the form $\mathbf{X}_i = \sum_{j=1}^k \alpha_{ij} \chi_{A_j}$, where the α_{ij} are real numbers and χ_{A_j} are the characteristic functions of the k metastable states. This implies that except for small perturbations these eigenvectors feature a piecewise constant structure that allows identification of the metastable states. (iii) Finally, there are $n - k$ remaining eigenvectors associated to the eigenvalues bounded away from 1. For identification of metastable states it is thus crucial to identify the Perron cluster eigenvalues and the corresponding eigenvectors. Fortunately this task can be easily accomplished since the Perron cluster eigenvalues are separated by a spectral gap from the remaining eigenvalues.

The existence of the spectral gap is the basis for identification of the number of metastable states. Basically the algorithm first computes the matrix of transition probabilities between microstates. The entries of this matrix represent the probability of going from microstate i at time t to microstate j at time $t + \tau$. The transition matrix eigenvalues are computed for different values of the lag time τ : as $\tau \rightarrow \infty$ the matrix will converge and its eigenvalues become constant too. As a consequence, if the implied time scales $\tau_k = -\tau[\ln \lambda_k]^{-1}$ are plotted as a function of the lag time, the time scales will level out, showing a major gap. The gap corresponds to the largest separation of time scales: the time scales above the gap correspond to transitions between

macrostates, whereas those below the gap represent transitions within macrostates. This predicts the number of macrostates to be one more than the number of implied time scales above the gap. Once the number of macrostates is determined, a first guess of the assignment of microstates to macrostates can be attained by means of a Perron cluster analysis (PCCA)³³ based on analysis of the piecewise constant structure of the eigenvectors associated to Perron cluster eigenvalues. The assignment of microstates to macrostates is then fine tuned through a Monte Carlo simulated annealing procedure aimed at maximizing the metastability $Q = \sum_{i=1}^k T_{ii}$, where T_{ii} is the self-transition probability for metastable state i . In each step of the Monte Carlo run a randomly chosen microstate is tentatively assigned to a random macrostate. The move is accepted with probability $\min[1, e^{\beta \Delta Q}]$, where β is set equal to the step number, corresponding to a decrease in the effective temperature. Once microstates have been assigned to macrostates it is then possible to compute the transition probabilities between macrostates

$$W(A, B) = \frac{\sum_{a \in I_A, b \in I_B} \pi_a p_{ab}}{\sum_{a \in I_A} \pi_a}$$

where $W(A, B)$ is the probability of transition from macrostate A to macrostate B , π_a is the stationary probability of being in microstate a , p_{ab} is the conditional probability to move to state b provided that the system is in microstate a , and I_A and I_B denote the index sets corresponding to A and B , respectively.

RESULTS

N-Glycosylated Peptide in Water and Water/HFA.

Since the presence of the glucose unit turned out to be essential for recognition of the MS autoantibodies, our discussion is focused on the description of the conformational ensembles of the glycosylated heptapeptide *Ac*-ERP(N(Glc)HTV-NH₂) in water and a 50% in volume mixture of water and HFA. As the biologically active conformation is expected to be, to some extent, hairpin-like, we first analyzed the end-to-end distance of the peptide in the two solvents (Figure 1). The probability distributions of this property in the two solvents are vastly overlapping, with a single high peak at about 18 Å, suggesting that the majority of conformations in the equilibrium ensemble

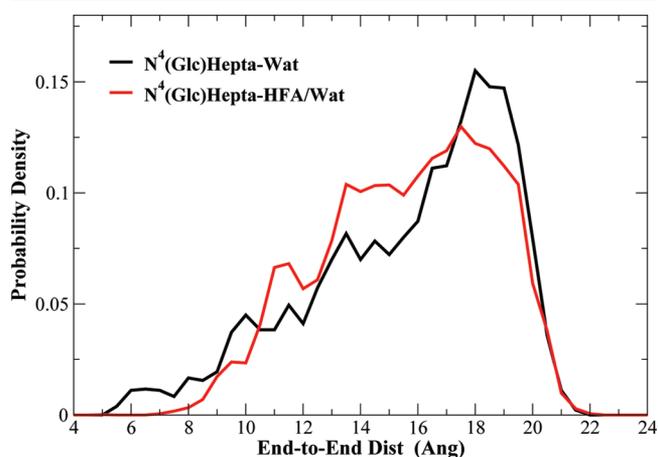


Figure 1. Probability distributions of the end-to-end distance of the glycosylated peptide in water (black curve) and the 50% volume mixture of water and HFA (red curve).

are extended. However, the distribution in water shows slightly higher probability values in correspondence of very large and very small values of the end-to-end distance, while in the water/HFA mixture the intermediate values of this structural parameter are more populated. This result not surprisingly suggests that the extended conformation is favored in water where the backbone and side-chain groups of the peptide can establish hydrogen bonds with the polar solvent.

In order to locate the position of a possible β -turn we monitored the set of all possible $C_{\alpha}(n)-C_{\alpha}(n+3)$ distances in the peptide (a β -turn exists when this distance is below 7 \AA^{34}). A discussion of the merits and limitations of this criterion along with a more detailed classification of the β -turn types of our peptides can be found in the Supporting Information (section β -turn types).

Figure 2 shows the probability distributions of the $C_{\alpha}(n)-C_{\alpha}(n+3)$ distances in the glycosylated peptide in water and the

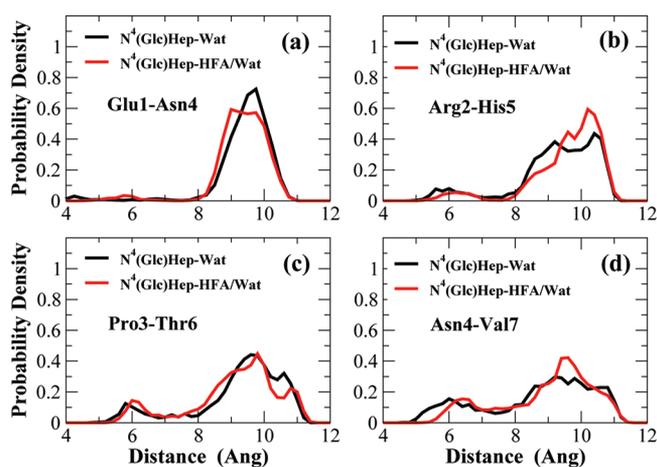


Figure 2. Probability distributions of the $C_{\alpha}(n)-C_{\alpha}(n+3)$ distances between pairs of residues of the glycosylated peptide in water (black line) and the water/HFA mixture (red line). β -Turns are formed when the distance is below 7.0 \AA . Panels a, b, c, and d refer to distances Glu1-Asn4, Arg2-His5, Pro3-Thr6, and Asn4-Val7, respectively.

HFA/water mixture. It can be noticed that such distributions are strikingly similar in the two solvents. Moreover, as one moves from the N-terminus to the C-terminus of the peptide the area under the distribution below the threshold distance of 7 \AA tends to increase, which suggests that the most likely position for the β -hairpin is near the C-terminus. In order to quantify the mean curvature in correspondence of $C_{\alpha}(i)$ we considered the triangle having vertexes on the three subsequent α carbons: $C_{\alpha}(i-1)$, $C_{\alpha}(i)$, $C_{\alpha}(i+1)$. The curvature can be simply quantified as the ratio r_c between the base (the distance $C_{\alpha}(i-1)-C_{\alpha}(i+1)$) and the height of the triangle. If the peptide is locally extended, the curvature ratio tends to infinity, while small values of the ratio are the signature of curved regions. A more detailed discussion on this indicator can be found in the Supporting Information (section Curvature ratio). Figure 3 reveals the existence of a gradient of curvature that tends to increase from the N-terminus to the C-terminus of the peptide. The highest curvature can be detected on His5, where the distribution shows a single sharp and very high peak at $r_c = 2$ and the probability density is larger than 0.8. Pronounced but less populated peaks at $r_c = 2$ can also be observed at Asn4 and Thr6, confirming the preferential location of a loop near the C-terminus. Once again, we must stress here the similarities of the

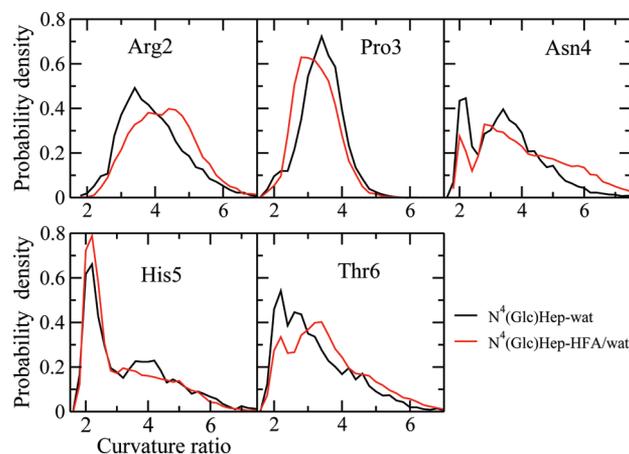


Figure 3. Probability distributions of the curvature ratio of residues 2–6 of the glycosylated peptide in water (black curve) and the water/HFA mixture (red curve). Panels a, b, c, d, and e refer to Arg2, Pro3, Asn4, His5, and Thr6, respectively.

curvature distributions of the glycosylated peptide in the two solvents, pointing to a common β -turn conformation with vertex on His5.

An important structural indicator is the relative position of the saccharide and peptide components of the glycopeptide. This can be measured by the distance d_{GP} between the center of mass of the glucose ring and the center of mass of the peptide: large values of d_{GP} refer to β -turn or bent structures with the glucose unit pointing away from the loop; small values are indicative of conformations where the glucose unit is buried in a partially folded peptide, while intermediate values are compatible with extended/disordered structures. As we shall see, this simple indicator reveals important conformational differences of the peptide in the two solvents. Figure 4 refers to the glycosylated peptide in the HFA/water mixture and shows that, in this solvent, the probability distribution of this quantity features three peaks relative to the indicative ranges 0–5, 5–10, and 10–12 \AA , roughly corresponding to 25%, 64%, and 11%, respectively, of the whole equilibrium ensemble. It can therefore be expected that in the HFA/water mixture three structural motifs exist with different positions of the glucose ring with respect to the peptide. This prediction that relies on the simple d_{GP} indicator can be confirmed by a partitioning of the conformational space based on a more accurate metrics, performed through QT clustering.²⁹ The amounts provided hereafter and in Table 2 derive from inspection of the 20 most populated clusters (full details of clusters population are reported in the Supporting Information, section Metrics and Clustering).

Below we report the main results of the clustering analysis for the glycopeptide in HFA/water mixture. The conformations with the smallest distance between the center of mass of the sugar and the center of mass of the peptide account for about 40% of the total conformational space and are exemplified by the representatives of Cluster 1 and Cluster 3 depicted in Figure 5. In this motif the peptide follows a trapezoidal profile, that is, it is extended in the central region, from Arg2 to His5, where it shows curvature points. The glucose ring is projected toward the interior of the loop in order to establish hydrogen bonds with the backbone and side-chain groups of the peptide. This motif is presumably due to the existence of a transient hydrophobic environment around the solute that induces the

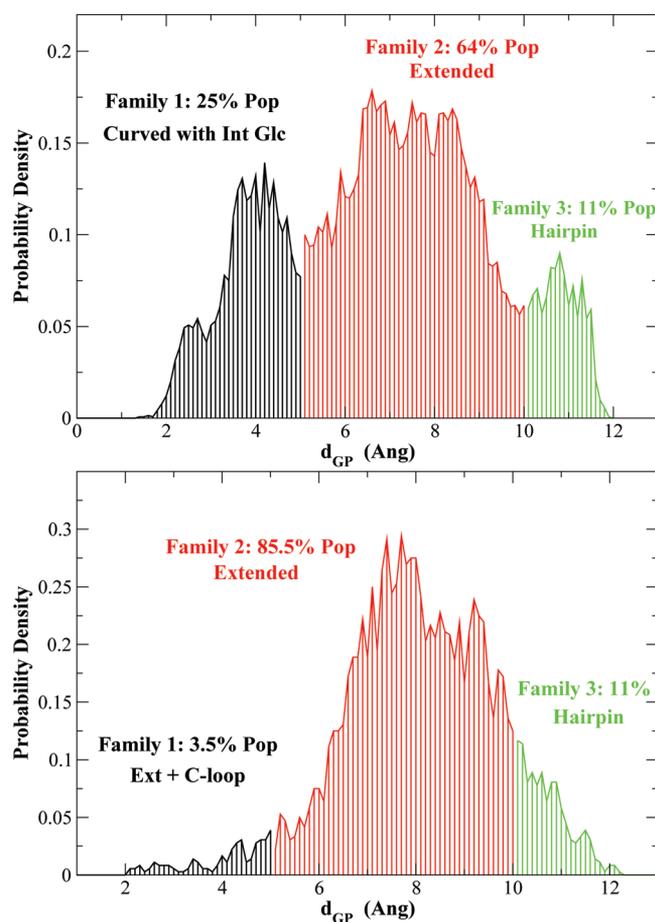


Figure 4. (Top) Probability distribution of the distance between the centers of mass of the glucose ring and the peptide in the HFA/water mixture. Areas shaded in black, red, and green roughly represent the three structural motifs, i.e., the curved motif with the glucose pointing toward the interior of the loop, the extended conformations, and the hairpin-like conformations. (Bottom) Probability distribution of the distance between the centers of mass of the glucose ring and the peptide in the pure water. Areas shaded in black, red, and green roughly represent the three structural motifs, i.e., the extended motif with a C-terminal loop, the extended conformations, and the hairpin-like conformations.

Table 2. Structural Composition of the Equilibrium Ensemble of the Glycosylated ($N^4(\text{Glc})\text{Hepta}$) and Unglycosylated Heptapeptide (Hepta) When Simulated in Water (W) and the HFA/Water (HFA/W) Mixture

peptide	solvent	extended, %	hairpin, %	Curv + GlcInt, %	extended + C-loop, %
$N^4(\text{Glc})$ Hepta	HFA/water	42	17	41	
$N^4(\text{Glc})$ Hepta	water	70	17		13
Hepta	HFA/water	65	16		19
Hepta	water	57	16		27

sugar to interact with the peptide (see also discussion in the Supporting Information, section Microsolvation effects). The hydrogen bonds established by the glucose then determine the curvature of the backbone.

Figure 5 also shows the representative of Cluster 2 that exemplifies the structures (about 40% of the total) with

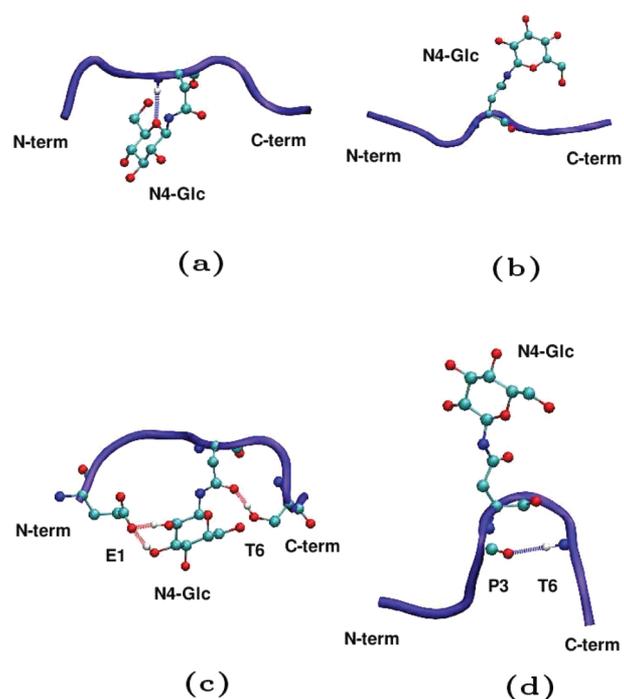


Figure 5. Representatives of the four most populated clusters of the glycosylated heptapeptide in the water/HFA mixture. Panels a, b, c, and d correspond to the representatives of clusters 1, 2, 3, and 4, respectively.

intermediate values of the distance between the centers of mass of the sugar and the peptide. In this motif the peptide is almost perfectly extended and the glucose ring projects away from the axis of the backbone. In this conformation the sugar preferentially interacts with the solvent even if the tilting toward the C-terminus of the side chain of the Asn-Glc unit does not exclude the possibility of interactions between the sugar and the peptide.

The third group of clusters, accounting for 17% of the conformational ensemble and characterized by a high value of the distance between the centers of mass of the sugar and the peptide, is illustrated by the representative of Cluster 4. This structure shows a typical hairpin-like conformation with a β -turn between Pro3 and Thr6 (the distance between the C_{α} s of these atoms is 5.7 Å) and the glucose unit pointing outside the loop.

When the glycosylated heptapeptide is simulated in water, the distribution of the d_{GP} distance undergoes significant changes with respect to that obtained in the simulation in the HFA/water mixture, featuring a single peak (see Figure 4(bottom)). In the 2–5 Å interval the distribution only shows a flat tail corresponding to 3.5% of the population, suggesting almost complete disappearance of the curved motif with the glucose oriented toward the interior of the loop. Expectedly, a large share of conformations (85.5%) exhibit in water d_{GP} values in the intermediate range 5–10 Å, corresponding to extended/disordered structures. Interestingly, the area under the distribution in the 10–12 Å range still corresponds to 11% of the population, indicating that the amount of hairpin-like conformations does not appear to change significantly on going from HFA/water to water, a result that is indeed in agreement with the data reported in Figures 2 and 3, pointing to a common β -turn conformation in the two solvents.

Cluster analysis reveals the existence of three main structural motifs in water. The representative of Cluster 1, depicted in Figure 6, is a typical example of the structures with the d_{CP}

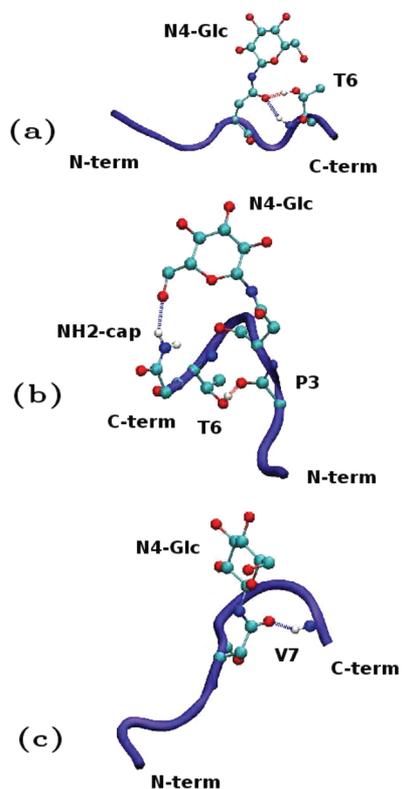


Figure 6. Typical structural motifs of the glycosylated peptide in water. Panels a, b, and c show the representatives of clusters 1, 5, and 6, respectively.

distribution in the 5–10 Å range. The peptide is fully extended, and the side chain of the Asn-Glc unit projects away from the axis of the peptide, tilting toward the C-terminus so that O_{δ} can occasionally establish H bonds with the side chain and NH backbone group of Thr6. These structures account for 70% of the equilibrium population of the glycopeptide in water, while they only amounted to 42% of the population in the water/HFA mixture. This pattern is the expected result of the larger polarity of water with respect to the water/HFA mixture: the possibility to establish a large network of hydrogen bonds with water discourages establishment of intramolecular interactions in the solute. The increase in the fraction of extended conformations in water with respect to HFA/water mixture is compensated by the almost complete disappearance of the curved structures with the sugar pointing toward the interior of the loop. These structures are replaced in water by a motif, accounting for 13% of the structures, where the peptide is almost extended but features a loop near the C-terminus. This motif is exemplified by the representative of Cluster 6, appearing in Figure 6. In this structure the C-terminal curvature is very pronounced and the distance between the $C_{\alpha S}$ of Asn4 and Val7 (7.10 Å) is just above the threshold of 7.0 Å defining the β -turn. The third structural motif, corresponding to the d_{CP} distributions in the 10–12 Å range, is the hairpin-like conformation exemplified by the representative of Cluster 5 and depicted in Figure 6, revealing structural similarities with Cluster 4 in the HFA/water mixture (see Figure 5). As in Cluster 4 for the HFA/water mixture, the glucose ring points

outside the β -turn but in water an H bond between hydroxyl 6 of the sugar and the C-terminal amide ending group is detected with a high incidence. This bond, together with that of the carbonyl of Asn4 with the same amide, induces a curvature of the C-terminal strand of the hairpin, so that the two strands are not parallel. Interestingly, the fraction of hairpin-like conformations based on clustering analysis is 17%, the same as that observed in the water/HFA mixture. This finding does not contradict the notion of the structuring effect of fluoro-alcohols. However, it must be noted that the structuring effect of HFA results in stabilization of the curved structures with the sugar pointing inside rather than that of the biologically active hairpin structures.

Conformational Kinetics by MSM. Conformational analysis of the glycosylated peptide in water and the HFA/water mixture reveals that the proportion of bioactive hairpin-like conformation is basically the same in both solvents. This result is seemingly in disagreement with NMR measurements that indicate that in a water environment the peptide is in a disordered, random coil conformation.⁹ A possible explanation for this inconsistency relies on the high flexibility of the peptide that in water could quickly interconvert between any pair of conformations. If these conformational changes are faster than the time required for two spins to interact, the signal could be averaged over all conformations populated during the NMR time scale thus causing the amount of hairpin conformations to be underestimated.³⁵ By contrast, if in the water/HFA mixture the conformational transitions of the hairpin conformations were somehow slowed down, then these structures could be stable enough to be detected by NMR.

In order to test our working hypothesis, we used the MSMBuilder package^{17,30} to build a Markov state model of the glycosylated peptide in the two solvents.

MSM in Water. The microstates of the glycopeptide in water correspond to the clusters identified through the quality threshold algorithm²⁹ with $d_{cut} = 6.0$ Å (see Supporting Information, section Metrics and Clustering), which led to identification of 206 clusters. In order to determine the number of metastable states we computed the first 80 eigenvalues of the 206×206 matrix of transition probabilities between microstates. The eigenvalues and thus the corresponding implied time scales were computed for 180 different lag times (from a minimum lag time of 5.0 ps to a maximum lag time of 900.0 ps at intervals of 5.0 ps). The plot of implied time scales as a function of lag time (Figure 7) reveals the existence of a single gap above which only one individual time scale can be detected, that is the time scale of transitions between two metastable states. The other low-lying time scales conversely represent the typical transition times between sub-basins inside the same metastable state. Once it is ascertained that our system only includes two macrostates, it is possible to lump together kinetically related microstates using the Perron cluster analysis³³ that exploits the peculiar structure of the eigenvectors of the matrix of transition probabilities between microstates. We used the simplex version of the PCCA algorithm (PCCA+)³⁶ that spared us the necessity to perform a refinement stage through simulated annealing. Instead, we used the built-in optimization routine of the PCCA+ algorithm. The PCCA calculation assigned to the first macrostate only a single microstate containing 4 hairpin-like conformations, while the rest of the population was assigned to the second metastable state. This results implies that in water practically all the

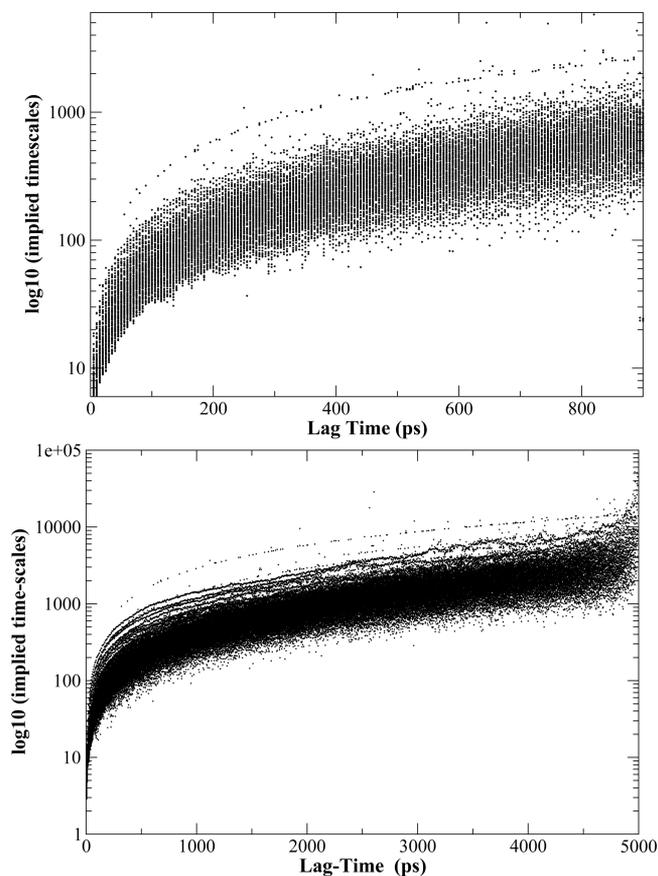


Figure 7. Time scales as a function of lag time for the Markov state model of the glycosylated peptide in water (top) and the water/HFA mixture (bottom).

population belongs to a single metastable state and any structure can quickly interconvert into any other conformation.

MSM in HFA/Water. A similar analysis was performed on the equilibrium population generated by simulation of the glycopeptide in the water/HFA mixture. The quality threshold clustering with $d_{\text{cut}} = 6.0 \text{ \AA}$ split the population in 193 clusters so that a 193×193 matrix of transition probabilities between microstates could be built. Figure 7(bottom) reports the first 80 implied time scales as a function of the lag time that was varied from a minimum of 5 ps to a maximum of 5 ns in increments of 5 ps. From the plot it is not clear whether above the main gap there are one or two main time scales. We therefore applied the PCCA+ algorithm in the hypothesis of 2 and 3 metastable states, choosing the scenario that guarantees the highest metastability. The metastability Q is a heuristics for maximizing the separation of time scales. The two-macrostate scenario yields a metastability $Q = 1.64$, while in the case of three metastable states we found $Q = 2.08$. Considering that the maximum possible value of the metastability corresponds to the number of macrostates, it is clear that the metastability is closer to its maximum in the two-macrostate situation. The microstate-to-macrostate mapping performed by the PCCA+ algorithm reveals that the first metastable state includes about 10% of the equilibrium population mostly comprising hairpin-like conformations (88% of the population of the metastable state), so that 52% of the hairpin-like structures of the whole equilibrium ensemble belong to this macrostate. Conversely, the second metastable state accounts for 90% of the equilibrium population, and it includes all possible structural motifs.

Calculation of the transition probabilities between macrostates reveals that the self-transition probabilities of the two macrostates are 68% and 96%, respectively. Conversely, the probability of transition from the hairpin-populated macrostate 1 to the more promiscuous macrostate 2 is 32%, and the probability of the reverse transition is only 4%. In other words, this scenario implies that a significant fraction of hairpin conformations is sequestered in macrostate 1 and can rearrange into structures of macrostate 2 only very slowly. This stabilization of hairpin conformations is presumably the reason why NMR spectroscopy can detect this motif in the water/HFA mixture but not in pure water. This result is confirmed by further data reported in the Supporting Information (see section Comparison with NOE data), where we show that the probability distribution of the distance HA(Pro)-HN(His) (corresponding to an intense NOE interaction) has a marked bimodal distribution in the water/HFA mixture with barrier heights on the order of few kcal/mol along the distance coordinate. Such bimodal behavior of the HA(Pro)-HN(His) probability distribution is still present in water, although with an attenuated free energy barrier.

Effect of the Glucose Unit on the Peptide Structure.

The fact that CSF114(Glc) specifically recognizes MS autoantibodies while the unglycosylated counterpart cannot clearly shows the importance of the glucose moiety for the immunochemical reactivity of the peptide. The role of the sugar could be due to either direct interaction with the autoantibodies circulating in blood or stabilization of a different native state corresponding to a structure sterically complementary to the binding pocket of the antibodies. This is why it is important to compare the conformational ensembles of the N-glycosylated heptapeptide ($N^{\text{G}}(\text{Glc})\text{Hepta}$) and unglycosylated heptapeptide (Hepta). Such comparison is fully detailed in the Supporting Information (section The effect of the glucose unit), and we give here the salient results. The glucose moiety in the HFA/water mixture has a moderate effect on the conformational landscape. Its main effect with respect to the unglycosylated variant is that of decreasing the incidence of extended conformations (see Table 2) with basically no impact on the population of hairpin structures. In pure water, the effect of the glucose unit is practically undetectable with all analyzed conformational properties of the glycosylated and unglycosylated form yielding the same behavior within statistical error (see Supporting Information Figures 15–17). These results rule out the hypothesis that the sugar moiety can alter significantly the conformational state of the peptide, lending further support to the notion of a direct interaction between Asn(Glc) and the autoantibodies in multiple sclerosis patients' sera, as recently speculated.^{6,9}

CONCLUSIONS

In this work we performed a computational study of a small glycosylated heptapeptide derived from CSF114(Glc) and engineered to favor β -turn structure, achieving an increase by 1 order of magnitude of the affinity for MS autoantibodies.⁹ According to NMR measurements,^{6,9} CSF114(Glc) and its derivatives adopt in water a random coil conformation that can be turned into the bioactive β -hairpin structure when a 50% in volume water/HFA mixture is used as solvent. The dominance of a random coil reported in water solvent, however, may be simply due to the averaging of the signal over all fast-interchanging conformations populated during the NMR time scale. In order to test this issue we performed Hamiltonian

REMD simulations of the glycosylated and unglycosylated heptapeptide Ac-ERPN(Glc)HTV-NH₂ in water and in a 50% in volume mixture of water and HFA. The unglycosylated peptide was also simulated in the two solvents as a reference. The structural compositions of the equilibrium ensembles in the four systems obtained by means of the QT-clustering analysis are summarized in Table 2.

Table 2 reveals that the fraction of β -hairpin conformations (17%) is the same when the glycopeptide is simulated in HFA/water and water. The greater hydrophobicity of the HFA/water solvent determines a decrease in the fraction of extended conformations (70% in water vs 42% in HFA/water) and the appearance of a new structural motif with a curved peptide chain and the glucose ring oriented toward the interior of the loop. These structures, representing 41% of the equilibrium ensemble of N⁴(Glc)Hepta in HFA/water may be compared to a family of structures with a basically extended peptide chain and a C-terminal loop that correspond to 13% of the population of N⁴(Glc)Hepta in water. In both situations the curvature is induced by a network of hydrogen bonds that Asn4-Glc establishes with the C-terminal residues, especially Thr6 and Val7.

The failure of NMR measurements to detect the hairpin structures in water may be due to the intrinsic limitations of this technique. In fact, NMR measures space couplings of nuclear spins in NOE experiments, and the slow relaxation in NMR spectroscopy sets a lower limit for spin interaction time scale. If the molecular conformation changes faster than the time required for two spins to interact, the signal is averaged over all conformations populated during the NMR time scale. Thus, the conformation of small, flexible peptides may not be revealed by NMR. This hypothesis was confirmed through the use of a Markov state model. This technique showed that in water the whole equilibrium population can be considered as a single metastable state, suggesting that all conformations can rapidly interconvert between each other. By contrast, in the water/HFA mixture the conformational space is partitioned in two metastable states, one of which contains about 50% of all hairpin-like conformations. The existence of high energy barriers between the two macrostates (testified by the low interstate transition rates) determines longer residence times of the peptide in the hairpin conformation that make it detectable by NMR. This effect is presumably also enhanced by the increase in viscosity induced by HFA and testified by the longer convergence times in the REM simulations.

The abundance of hairpin-like conformations in the aqueous environment suggests that peptide Ac-ERPN(Glc)HTV-NH₂ could be used not just as a probe for MS biomarkers in a diagnostic assay but also as a drug (possibly in dendrimeric form³⁷) for treatment of the disease. Our study also suggests the opportunity to further optimize the peptide sequence so as to increase the fraction of hairpin-like conformations that turned out to be below 20% in both solvents.

■ ASSOCIATED CONTENT

● Supporting Information

Force field parameters for HFA; comparison of GLYCAM06/Amber for β -D-glycopyranose in water solution; microsolvation effects in HFA; simulation data and analysis for the unglycosylated peptide in water and HFA/water mixture; β -turn classification of the hairpin-like structures in glycosylated and unglycosylated peptides in water and HFA/water mixture; error analysis and convergence tests for MD simulations;

comparison of simulation data with experimental NOE-derived upper limit constraints; metrics and clustering; curvature ratio. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: piero.procacci@unifi.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was performed as a part of the project *In silico assessment of peptides with potential pharmaceutical properties for Multiple Sclerosis* supported by the Regione Toscana (Regional Health Research Program 2009). All calculations were performed on the high-performance architectures of the CRESCO (Computational Research Center for Complex Systems) Project. We thank Samuele Pierattini of ENEA for his constant and competent technical support on the CRESCO facility.

■ REFERENCES

- (1) Sospedra, M.; Martin, R. *Ann. Rev. Immun.* **2005**, *23*, 683–747.
- (2) Holmes, B. A.; Madgwick, T.; Bates, D. *Br. J. Med. Econ.* **1995**, *8*, 181–193.
- (3) Pittcock, S. J. *Lancet* **2007**, *370*, 1417–1422.
- (4) Barclay, L. *Neurology* **2003**, *61*, 602–611.
- (5) Leslie, D.; Lipsky, P.; Notkins, A. L. *J. Clin. Invest.* **2001**, *108*, 1417–1422.
- (6) Lolli, F.; Mulinacci, B.; Carotenuto, A.; Bonetti, B.; Sabatino, G.; Mazzanti, B.; D'Ursi, A.; Novellino, E.; Pazzagli, M.; Lovato, L.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10273–10278.
- (7) Mazzucco, S.; Matà, S.; Vergelli, M.; Fioresi, R.; Nardi, E.; Mazzanti, B.; Chelli, M.; Lolli, F.; Ginanneschi, M.; Pinto, F.; et al. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 167–172.
- (8) Carotenuto, A.; D'Ursi, A. M.; Nardi, E.; Papini, A. M.; Rovero, P. *J. Med. Chem.* **2001**, *44*, 2378–2381.
- (9) Carotenuto, A.; Alcaro, M. C.; Saviello, M. R.; Peroni, E.; Nuti, F.; Papini, A. M.; Novellino, E.; Rovero, P. *J. Med. Chem.* **2008**, *51*, 5304–5309.
- (10) Rajan, R.; Awasthi, S. K.; Bhattacharjya, S.; Balaram, P. *Biopolymers* **1997**, *42*, 125–128.
- (11) Chatterjee, C.; Hovagimyan, G.; Gerig, J. T. *Current fluoroorganic chemistry*; ACS Symposium Series; Oxford University Press: Cary, NC, 2007; Vol. 949; pp 379–392.
- (12) Carotenuto, A.; D'Ursi, A. M.; Mulinacci, B.; Paolini, I.; Lolli, F.; Papini, A. M.; Novellino, E.; Rovero, P. *J. Med. Chem.* **2006**, *49*, 5072–5079.
- (13) Williamson, M. P. In *Spectroscopic Methods and Analyses*; Jones, C., Mulloy, B., Thomas, A. H., Eds.; Humana Press, 1993; Chapter 2.
- (14) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (15) Guardiani, C.; Marsili, S.; Marchetti, S.; Gambi, C.; Procacci, P.; Livi, R. *Biopolymers (Pept. Sci.)* **2010**, *96*, 245–251.
- (16) Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. *J. Comput. Chem.* **2010**, *31*, 1106–1116.
- (17) Bowman, G.; Huang, X.; Pande, V. *Methods* **2009**, *49*, 197–201.
- (18) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848–1862.
- (19) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712–725.
- (20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (21) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

- (22) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M. et al. *AMBER 9*; University of California: San Francisco, 2006.
- (23) Marchi, M.; Procacci, P. *J. Chem. Phys.* **1998**, *109*, 5194–5202.
- (24) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
- (25) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (26) Tuckerman, M.; Berne, B. J. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
- (27) Sugita, A., Y.; Kitao; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (28) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.
- (29) Heyer, L. J.; Kruglyak, S.; Yooseph, S. *Genome Res.* **1999**, *9*, 1106–1115.
- (30) Bowman, G.; Beauchamp, K.; Boxer, G.; Pande, V. *J. Chem. Phys.* **2009**, *131*, 124101.
- (31) Chodera, J.; singhal, N.; Pande, V.; Dill, K.; Swope, W. *J. Chem. Phys.* **2007**, *126*, 155101.
- (32) Stewart, G. In *Mathematical computer performance and reliability*; Iazeolla, G., Courtois, P., Eds.; Elsevier: New York, 1984; pp 287–302.
- (33) Deuffhard, P.; Huisinga, W.; Fischer, A.; Schütte, C. *Lin. Alg. Appl.* **2000**, *315*, 39–59.
- (34) Chou, K. *Anal. Biochem.* **2000**, *286*, 1–16.
- (35) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519–538.
- (36) Deuffhard, P.; Weber, M. In *Linear Algebra and Its Applications: Special Issue on Matrices and Mathematical Biology*; Dellnitz, M., Kirkland, S., Neumann, M., Schütte, C., Eds.; Elsevier: New York, 2005; pp 161–184.
- (37) Sadler, K.; Tam, J. P. *Rev. Mol. Biotechnol.* **2002**, *90*, 195–229.