JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Conformational Determinants of the Activity of Antiproliferative Factor Glycopeptide

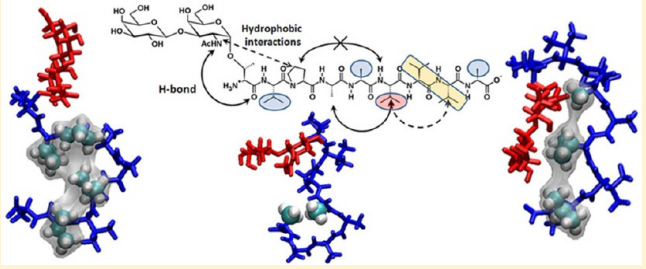
Sairam S. Mallajosyula,[†] Kristie M. Adams,[‡] Joseph J. Barchi,[‡] and Alexander D. MacKerell*,[†]

[†]Department of Pharmaceutical Sciences, University of Maryland, 20 Penn Street HSF II, Baltimore, Maryland 21201, United States

[‡]Chemical Biology Laboratory, Molecular Discovery Program, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland 21702, United States

Supporting Information

ABSTRACT: The antiproliferative factor (APF) involved in interstitial cystitis is a glycosylated nonapeptide (TVPAAVVVA) containing a sialylated core 1 α-O-disaccharide linked to the N-terminal threonine. The chemical structure of APF was deduced using spectroscopic techniques and confirmed using total synthesis. The synthetic APF provided a platform to study amino acid modifications and their effect on APF activity, based on which a structure−activity relationship (SAR) for APF activity was previously proposed. However, this SAR model could not explain the change in activity associated with minor alterations in the peptide sequence. Presented is computational analysis of 14 APF derivatives to identify structural trends from which a more detailed SAR is obtained. The APF activity is found to be dictated by the close interplay between carbohydrate−peptide and peptide−peptide interactions. The former involves hydrogen bond and hydrophobic interactions, and the latter is dominated by hydrophobic interactions. The highly flexible hydrophobic peptide adopts collapsed conformations separated by low energy barriers. APF activity correlates with hydrophobic clustering associated with amino acids 4**A**, 6**V**, and 8**V**. Peptide conformations are highly sensitive to single point mutations, which explain the experimental trends. The presented SAR will act as a guide for lead optimization of more potent APF analogues of potential therapeutic utility.



INTRODUCTION

Interstitial cystitis (IC) is a disease of the urinary bladder which is characterized by the thinning of the bladder epithelium.[1,2] This painful bladder disorder affects approximately 1 million Americans, with evidence suggesting that it occurs eight times more frequently in women than in men.[1,2] While the cause of this disease still remains unknown, urine from IC patients has been shown to contain an antiproliferative factor (APF) that decreases ^{3}H-thymidine incorporation by human bladder epithelial cells.[3] Using a total synthesis approach, APF was identified as a nonapeptide (TVPAAVVVA) containing a 2,3-sialylated core 1 α-O-linked disaccharide (Galβ1−3GalNAc: The Thomsen-Friedenreich antigen) linked to the N-terminal threonine residue (Neu5Acα2−3Galβ1−3GalNAcα-O-TVPAAVVVA).[4] Synthetic APF or its desialylated analogue (asialo APF or as-APF) was shown to inhibit proliferation of normal bladder epithelial cells at nanomolar concentrations,[4] to increase paracellular permeability and decrease the expression of tight junction proteins.[5] These effects were rescued by the D-proline and D-pipecolic acid derivatives of APF, the only two analogues that were shown to inhibit APF action on the normal bladder epithelium.[6] APF was further shown to inhibit proliferation of T24 bladder carcinoma cells,[4] with subsequently proven similar effects on a panel of solid tumor cell lines,[7] making APF a fascinating lead candidate for both anti-IC and anticancer drug design. Some insight into the mechanism of APF activity and the associated targets was recently found by the discovery that cytoskeletal-associated protein 4 (CKAP4) was a functional cellular receptor for APF in bladder epithelial cells.[8]

While as-APF maintains full potency, further truncation of the resultant disaccharide results in a complete loss of activity.[9] In the absence of structural information synthetic analogs of as-APF have been tested for their biological antiproliferative activity to establish structure activity relationships (SAR) in as-APF.[4,9,10] Based on these studies, which targeted extensive modifications in the peptide region of APF, it was found that the smallest synthetic analog of APF that maintained full potency was the synthetic glyco-octapeptide Galβ1−3GalNAcα-O-TVPAAAAA, where the trivaline tail was replaced by three alanine residues.[9,10] The availability of biological activity data for the various synthetic analogs of as-APF with differing amino acid sequences sets the groundwork for an extensive study relating the conformational properties of APF and its analogs to the bioactivity of the glycopeptide.

In this study, we performed Hamiltonian replica exchange (HREX) molecular dynamics (MD) simulations on as-APF and selected analogs to gain insights into the structure activity relationships (SAR) in APF. The as-APF derivatives were

Received: March 9, 2013

   dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

chosen based on three criterion: (i) the availability of experimental ^3J coupling and NOE data for a direct comparison between the experiments and simulations, (ii) differing biological activities (inactive to active), and (iii) differing peptide sequences to develop the SAR model. Greater understanding of the SAR for active APF derivatives will aid in the rational design of APF inhibitors that may be of utility for the treatment of IC.

METHODS

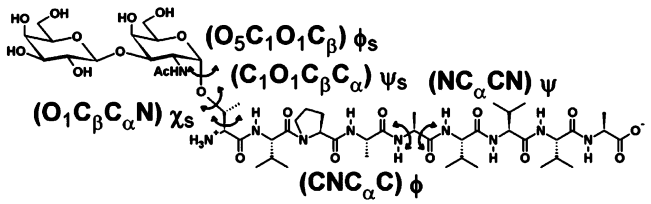
Model Compounds. A total of 12 as-APF derivatives and two unglycosylated amino acid sequences were chosen for the HREX MD simulations (Table 1).[4,9,10] The chosen derivatives

Table 1. APF Analogs Used in the Present Study

compd no.	sequence[a]	% activity	NMR data
1	Galβ1−3GalNAcα-O-TVPAAVVVA	100%	yes
2	Galβ1−3GalNAcα-O-**S**VPAAVVVA	0.01%	yes
3	Galβ1−3GalNAcα-O-**SL**PAAVVVA	1%	yes
4	Galβ1−3GalNAcα-O-T**L**PAAVVVA	0%	yes
5	Galβ1−3GalNAcα-O-TVPA**V**VVVA	0%	no
6	Galβ1−3GalNAcα-O-TVPAA**A**VVA	---	---
7	Galβ1−3GalNAcα-O-TVPAAV**A**VA	---	---
8	Galβ1−3GalNAcα-O-TVPAAVV**A**A	---	---
9	Galβ1−3GalNAcα-O-TVPA**AAAA**A	1%	no
10	Galβ1−3GalNAcα-O-TVPAAVVV	100%	yes
11	Galβ1−3GalNAcα-O-TVPA**AAA**	100%	yes
12	TVPAAVVVA	0%	no
13	TVPAAVVV	0%	yes
14	Galβ1−3GalNAcα-O-TV(**dP**)AAVVVA	0%	no

[a]Modifications in the amino acid sequence are highlighted in bold. The activities reported are normalized to the activity of **1** according to the following equation: % = [IC_{50}(as-APF)/IC_{50} (derivative)] × 100.[4,9,10]

Scheme 1. Structure of as-APF



Galβ1-3GalNAcα-O-TVPAAVVVA (*as*-APF)

including as-APF (compound **1**; Scheme 1) represent a selection of as-APF derivatives with differing biological activities. The details of the synthesis of the glycopeptides have been described in detail in earlier studies.[4,9,10] MD simulations were performed using an equilibration and production strategy as tested earlier for O-linked glycoprotein systems.[11] In brief, the simulations were performed with the CHARMM program.[12] The CHARMM22 protein force field[13] with CMAP (dihedral correction map),[14] the CHARMM carbohydrate force field,[15−19] and the modified TIP3P water model[20] were used to represent the glycopeptides in solution. Initial configurations corresponding to *anti* (180°) and *gauche*+ (60°) conformations were chosen for the H−N−C2−H2 (acetylamino side chain in carbohydrate) and O1−C_β−C_α−N (O-linkage) dihedrals, respectively, in accordance with NMR studies (Tables S1 and S2, Supporting Information (SI)). The rest of the geometries of the model compounds were generated from the topology information present in the force field. These initial geometries were subjected to a 1000-step steepest descent (SD) minimization followed by an adopted basis Newton−Raphson (ABNR) minimization to a force gradient tolerance of 10^{-6} kcal/mol/Å.[21,22] The minimized geometries of the glycopeptides were then immersed in a pre-equilibrated cubic water box of size 45 Å × 45 Å × 45 Å. The size of the water box was selected based on the condition that it extended at least 10 Å beyond the non-hydrogen atoms of the glycopeptides. Water molecules with the oxygen overlapping with the non-hydrogen solute atoms within a distance of 2.8 Å were deleted. For all of the subsequent minimizations and MD simulations, periodic boundary conditions were employed using the CRYSTAL module implemented in the CHARMM program.

System equilibration was initiated with a 50-step SD minimization followed by a 50-step ABNR minimization cycle in which mass-weighted harmonic restraints of 1.0 kcal/mol/Å were applied on the non-hydrogen atoms of the glycopeptides. This was followed by a 100 ps simulation in the NVT ensemble with the same harmonic restraints to equilibrate the solvent molecules around the glycopeptides. A 200 ps NPT simulation at 1 atm and 298 K followed the NVT simulation, wherein all the previous restraints were removed. In the NPT simulation the center of mass of the glycopeptide was restrained near the origin by using the MMFP module[23] in CHARMM using a harmonic restraint of 1.0 kcal/mol/Å applied to the center of mass of the glycopeptides. The electrostatic interactions were treated via the particle mesh Ewald method with a real-space cutoff of 12 Å, a kappa value of 0.34 $Å^{-1}$, and a sixth-order spline.[24] Nonbond interaction lists were updated heuristically out to 16 Å with a force switch smoothing function from 10 to 12 Å used for the Lennard-Jones interactions.[25] The Leapfrog integrator employing an integration time step of 2 fs with the SHAKE algorithm was used to constrain all covalent bonds involving hydrogen atoms.[26] The temperature was maintained at 298 K by a Nose−Hoover heat bath with a thermal piston parameter of 2000 kcal mol^{-1} ps^2.[27] Constant pressure of 1 atm was controlled using the Langevin piston with a mass calculated using the equation P_{mass} = integer(system mass/50.0).[28]

HREX MD production simulations were performed using the REPDSTR module of a modified version of CHARMM c36a2.[29] The HREX simulations were started from the equilibrated coordinates obtained after the 200 ps unbiased NPT simulation at 1 atm and 298 K. The same MMFP restraints used in the NPT runs were utilized in the HREX runs to constrain the glycopeptide at the center of the simulation box. An exchange between neighboring replicas was attempted every 1000 MD steps, and the coordinates were saved every 1 ps. Each replica was simulated for 11.4 ns, thereby amounting to a cumulative simulation time of 91.2 ns (11.4 ns × 8), and the trajectories from the final 10 ns of the first replica (unbiased, ground state replica) were used for subsequent analysis.

A combination of the two-dimensional (2D) dihedral grid-based energy correction map (CMAP) extension of the CHARMM force field and a Saxon-Wood potential were used as the biasing potential across the different replicas.[14] Two CMAP biasing potentials were used, corresponding to the ϕ_s/

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

ψ_s and ϕ/ψ dihedral pairs (Scheme 1) to sample the conformational space of the glycopeptides around the O-glycosidic linkage and the peptide backbone. Additionally, the Saxon-Wood potential, eq 1, was used to enhance the conformational sampling about the χ_s dihedral (Scheme 1) in the Thr/Ser side chain

$$U = h\left[1 + \exp\left\{\frac{p_2 - \|\theta - \theta_{ref}\|}{p_1}\right\}\right]^{-1} \quad (1)$$

where $h = (n \times -0.75)$ kcal/mol, n goes from 0 to 7 for replicas 1 to 8, $p_1 = 0.1$, $p_2 = 0.3$, and $\theta_{ref} = 90°$. The biasing potential CMAPs were obtained using an established protocol for glycopeptide O-linkages which has been reported in detail earlier and successfully applied in studying O-linked glycopeptides.[11]

NMR ^3J coupling constants about the dihedral torsions H−N−C_2−H_2 (for GalNAc), N−C_α−C_β−O_1, and C−N−C_α−C were calculated from the simulations using the following Karplus equations.[30−32]

$$^3J_{HNH2} = 9.60\cos^2\theta - 1.51\cos\theta + 0.99 \quad (2)$$

$(\theta = \text{H−N−C}_2\text{−H}_2)$

$$^3J_{H\alpha H\beta} = 9.5\cos^2\theta - 1.6\cos\theta + 1.8 \quad (3)$$

$(\theta = \text{N−C}_\alpha\text{−C}_\beta\text{−O}_1 \text{ for H}\beta_2 \text{ and } \theta = \text{N−C}_\alpha\text{−C}_\beta\text{−O}_1 - 120° \text{ for H}\beta_3)$

$$^3J_{HNH\alpha} = 6.51\cos^2\theta - 1.76\cos\theta + 1.60 \quad (4)$$

$(\theta = \text{C−N−C}_\alpha\text{−C} - 60°)$

RESULTS

Developing a SAR model for conformationally flexible molecules such as APF requires appropriate conformational sampling. NMR experiments provide indirect insight into the conformation of a molecule via information on dihedral angles and close contacts obtained from NMR ^3J coupling and NOE data, respectively. In this study we use this information to test the performance of the HREX conformational sampling method. As detailed below, the HREX protocol successfully reproduces the ^3J coupling and NOE data for the seven compounds with available NMR data, thus providing confidence in the conformational sampling protocol. The HREX protocol was then extended to the remainder of the model compounds (Table 1), resulting in the data set that allowed for development of a refined SAR model for APF.

^3J Coupling Constants. ^3J coupling constants calculated from HREX simulations by applying the appropriate Karplus equations (eqs 2, 3, 4) are summarized in Tables S1, S2 and S3 (SI) for the H−N−C_2−H_2 (for GalNAc), N−C_α−C_β−O_1, and C−N−C_α−C dihedral torsions, respectively. The ^3J coupling constants obtained by averaging over the simulations are in agreement with the experimentally observed values for the various dihedrals, indicating that the HREX simulations sample the experimentally observed conformational space. The H−N−C_2−H_2 dihedral largely samples the anti conformation, which corresponds to the large ^3J value of ∼10.08 Hz observed experimentally. For all the Thr O-linked carbohydrate systems the N−C_α−C_β−O_1 dihedral predominantly samples the +60°(gauche+) conformation, as has been observed experimentally for model Thr O-linkages.[33−35] This conformational rigidity is lost on substituting **T** with **S** (**2**), which also samples the −60°(gauche-) conformation for close to 20% of the simulation time (Table S2). This loss in rigidity can be related to the removal of the $-CH_3$ side chain on substituting **T** to **S** which results in the release of steric strain. Interestingly on substituting the subsequent amino acid S**V** (**2**) by S**L** (**3**), which amounts to introducing the $-CH_3$ side chain at the second position, the conformational rigidity is regained, with the N−C_α−C_β−O_1 dihedral preferentially sampling the +60°(gauche +) conformation (Table S2). The loss and gain in the rigidity of the N−C_α−C_β−O_1 dihedral with the number of $-CH_3$ side chains matches well with the observed biological activity (Table 1). Substituting **T** (**1**) with **S** (**2**) resulted in a 4 orders of magnitude loss of activity, going from 100% to 0.01%, while resubstituting a $-CH_3$ group at the second position (S**V** to S**L**, **3**) to create an "isosteric" substitution resulted in a 100-fold gain "back" in activity (1%). However adding additional methyls to the system (T**V** to T**L**, **4**), resulted in a complete loss of activity. These results are indicative of carbohydrate−peptide interactions controlling the observed biological activity. Finally, the ^3J coupling constants corresponding to the C−N−C_α−C peptide backbone torsions are summarized in Table S3 of the SI. The ^3J coupling constants are in close agreement with the experimental values; with the average RMS difference with respect to the experimental values (1.3 Hz) being less than the RMS deviation observed in the simulations (2.3 Hz). The only exception to this pattern was the last amino acid wherein the RMS difference of 2.2 Hz was greater than the RMS deviation of 1.5 Hz. The strong agreement with the experimentally observed ^3J coupling in the carbohydrate, carbohydrate−peptide linkage, and peptide regions highlights the appropriate conformational sampling in the HREX simulations.

ROE (Rotating Frame Nuclear Overhauser Effect) gives insight into the noncovalent interactions in the systems based on the proximity of the interacting magnetic nuclei, in our case the protons. Experimental ROEs can be compared to close contact probabilities obtained from HREX simulations. ROE interactions in the glycopeptides can be classified into (a) carbohydrate−peptide and (b) peptide−peptide interactions. All the observed experimental ROE's are tabulated in Table 2.

Carbohydrate−Peptide Interactions. NMR studies show the presence of two medium strength carbohydrate−peptide ROE signals (Table 2) in the glycopeptides systems. The first is between the H5 of GalNAc and the $-CH_3$ side chain of 1**T** in all the **T** substituted glycopeptides (**1**, **4**−**11**, **14**), and the second is between the $-CH_3$ in the N-acetyl group of GalNAc and the $H_{\gamma'}$ hydrogen of 3**P**. It is of interest to note that no ROE signal was observed between H5 of GalNAc and 1**S** in **2** and **3**. An analysis of the simulation of **1** revealed the presence of a strong hydrogen bond (H-bond) between the acetylamino side chain of GalNAc and the amide backbone of 1**T**/1**S**, which locks the sugar conformation with respect to the peptide (Figure 1a). Such H-bonds between the acetylamino side chain of GalNAc and the amide backbone have been observed previously in mucin architecture.[36,37] This H-bond between GalNAc and 1**T** (NH−O) brings the methyl group in the acetylamino ($-NHCOCH_3$) side chain to the proximity of the proline ring, resulting in the GalNAc ($-CH_3$) - 3P (H_δ) ROE signal. In Table 3 the H-bond occupancy corresponding to the GalNAc (HN) − 1**T**/1**S** (O) H-bond is presented. The

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. NMR ROE Signals in the Glycopeptide Systems

compd no.	carbohydrate−peptide	peptide−peptide[a]
1	GalNAc (-H5) - 1T (H_γ) GalNAc ($-CH_3$) - 3P (H_δ)	4A (NH) − 5A (NH) **m** 5A (NH) − 6 V (NH) **w** 6 V (NH) − 7 V (NH) **m** 8 V (NH) − 9A (NH) **w**
2	GalNAc ($-CH_3$) - 3P (H_α) GalNAc ($-CH_3$) - 3P (H_δ)	4A (NH) − 5A (NH) **m** 5A (NH) − 6 V (NH) **m** 6 V (NH) − 7 V (NH) **m** 8 V (NH) − 9A (NH) **m**
3	GalNAc ($-CH_3$) - 3P (H_α) GalNAc ($-CH_3$) - 3P (H_δ)	4A (NH) − 5A (NH) **w** 5A (NH) − 6 V (NH) **m** 6 V (NH) − 7 V (NH) **m** 8 V (NH) − 9A (NH) **m**
4	GalNAc (-H5) - 1T (H_γ) GalNAc ($-CH_3$) - 3P (H_δ)	4A (NH) − 5A (NH) **m** 5A (NH) − 6 V (NH) **m** 6 V (NH) − 7 V (NH) **m** 8 V (NH) − 9A (NH) **m**
10	GalNAc (-H5) - 1T (H_γ) GalNAc ($-CH_3$) - 3P (H_δ)	4A (NH) − 5A (NH) **w** 5A (NH) − 6 V (NH) **w** 6 V (NH) − 7 V (NH) **w**
11	GalNAc (-H5) - 1T (H_γ) GalNAc ($-CH_3$) - 3P (H_δ) GalNAc ($-CH_3$) - 2 V (H_α)	4A (NH) − 5A (NH) **m**
13		4A (NH) − 5A (NH) **w** 5A (NH) − 6 V (NH) **w** 6 V (NH) − 7 V (NH) **w**

[a]The intensity of the peptide−peptide ROE signal is presented in bold.

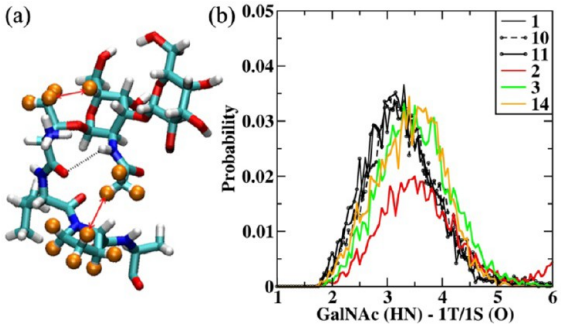


Figure 1. (a) Representative structure of **1** from HREX simulations showing the carbohydrate unit and the first three amino acids (TVP). The H-bond is shown in the black dotted line. The hydrogens involved in the ROE signals are identified using orange spheres, and the interaction is highlighted using red arrows. (b) GalNAc (HN) − 1T/1S (O) H-bond distribution for **1**, **10**, **11**, **2**, **3**, and **14**. Distances are presented in Å.

probability distribution associated with this H-bond for all the systems is presented in Figure S1 of the SI. In Figure 1b we present the H-bond distribution for **2**, **3**, and **14**, which show reduced H-bond occupancy (Table 3), and for compounds **1**, **10**, and **11**, which show 100% activity for direct comparison.

From the probability distribution it is observed that as-APF (**1**) presents a unimodal distribution centered on 2.9 Å. This distribution gets affected on substituting **T** (**1**) with **S** (**2**), with the latter exhibiting a bimodal distribution (Figure S1). This bimodal distribution is consistent with the sampling of the −60°(gauche-) conformation for the N−C_α−C_β−O_1 dihedral in **2** (Table S2). However, this distribution becomes unimodal on substituting $-CH_3$ at the second position (**SV** to **SL**, **3**),

Table 3. Carbohydrate−Peptide and Peptide−Peptide Hydrogen Bond Occupancies[a]

	carbohydrate−peptide H-bond occupancy		peptide−peptide H-bond occupancy		
compd no.	GalNAc (HN) − 1T/1S (O)	GalNAc (O) − 4A (HN)	4A (O)- 7 V/7A (HN)	3P (O)- 6 V/6A (HN)	5A/5 V (O)-8 V/ 8A (HN)
1	0.672	0.092	0.035	0.103	0.050
2	0.292	0.059	0.020	0.106	---
3	0.491	0.122	0.092	0.282	0.102
4	0.653	0.064	---	0.177	---
5	0.638	0.098	---	0.026	---
6	0.706	0.127	0.164	0.070	0.039
7	0.664	0.117	0.011	0.034	0.034
8	0.527	0.059	0.015	0.019	0.081
9[b]	0.629	0.103	0.625	0.623	0.492
10	0.621	0.082	---	---	0.093
11[b]	0.681	0.091	0.481	0.531	0.247
12			---	0.015	0.014
13			0.043	0.096	0.023
14	0.520	0.465	---	---	----

[a]Occupancies calculated using a H-bond cutoff distance of 3.5 Å. [b]Additional H-bonds between 3P(O)-7A(HN) (0.589 and 0.436) and 4A(O)-8A(HN) (0.543 and 0.209) found in **9** and **11**.

albeit the H-bond is elongated in both **2** and **3**, with the distribution centered on 3.1 Å. From the occupancy data presented in Table 3 it can be seen that the H-bond occupancy is affected for **2** (0.292), **3** (0.491), and **14** (0.520). From these data it becomes clear that the H-bond is necessary to lock the conformation of the sugar unit with respect to the peptide and may explain the observed activity pattern for the **TV** (**1**) versus **SV** (**2**) or **SL** (**3**) mutations. However the H-bond analysis fails to provide insight into the activity profile observed in the other **T** variants, starting from **TL** (**4**) which showed 0% activity but has the same H-bond profile as **1** (Figure S1).

These results suggest that secondary interactions involving the peptide unit affect the activity. To gain insight into the peptide−peptide interactions, the peptide region of as-APF (**1**) was closely scrutinized, including comparison with ROE data.

In the standard NMR study of peptides, sequential ROE signals between the amide protons (−NH) on consecutive amino acids are analyzed for indications of bending/kinking in the peptide backbone. For as-APF (**1**) medium strength signals were observed between 4**A**-5**A** and 6**V**-7**V**, while weak signals were observed between 5**A**-6**V** and 8**V**-9**A**. The strengths of these signals are altered to varying degrees for the other as-APF derivatives as tabulated in Table 2. This experimental evidence provides the incentive to analyze these distance distributions to establish a potential structural descriptor of activity. The distance distribution for these amide pairs for **1** is presented in Figure 2a. It is seen that the amide protons pairs 4**A**-5**A** remain in close contact for close to 40% of the simulation time ($d_{4A\text{-}5A}$ < 3.5 Å), while the 5**A**-6**V** pair remains in close contact for close to 57% of the simulation time ($d_{5A\text{-}6\,V}$ < 3.5 Å). The proximity of these particular proton pairs is traced back to the conformations sampled by the underlying peptide backbone. The ϕ/ψ dihedral distributions associated with 4**A**, 5**A**, 6**V**, 7**V**, and 8**V** for **1** are also presented in Figure 2. The peptide backbone samples both the α-helical and extended conformations (β and PPII) around the alanines (4**A** and 5**A**), while it predominantly samples only the extended conformations around the three valines (6**V**, 7**V**, 8**V**). This observation is in

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

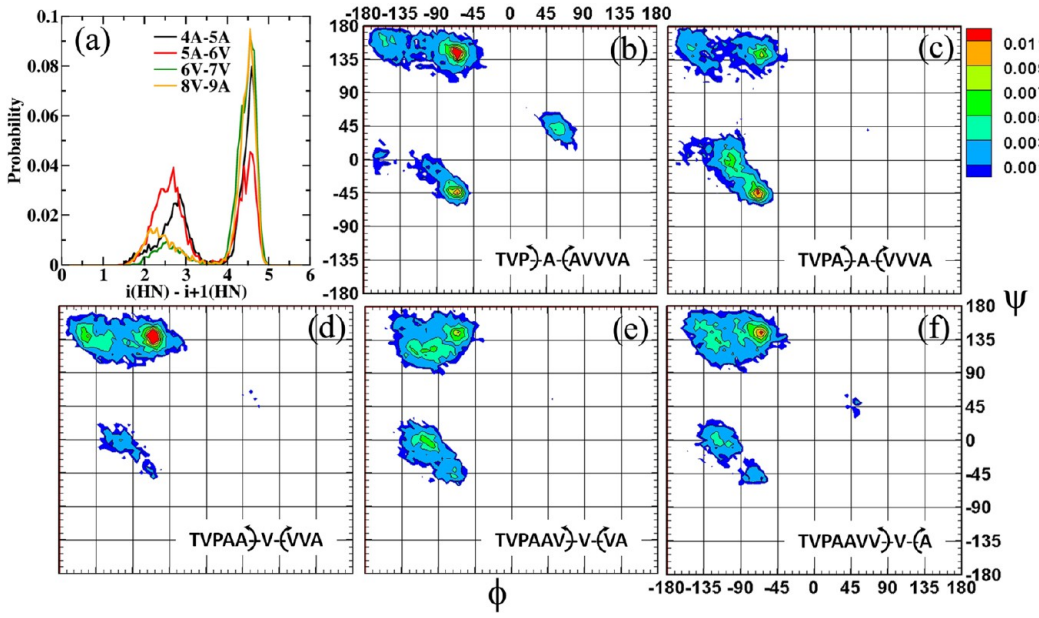


Figure 2. (a) Distance distribution between the amide proton pairs for **1**. ϕ/ψ dihedral distributions for (b) 4**A**, (c) 5**A**, (d) 6**V**, (e) 7**V**, and (f) 8**V** for **1**. Distances are presented in Å, and the dihedral angles are presented in degrees.

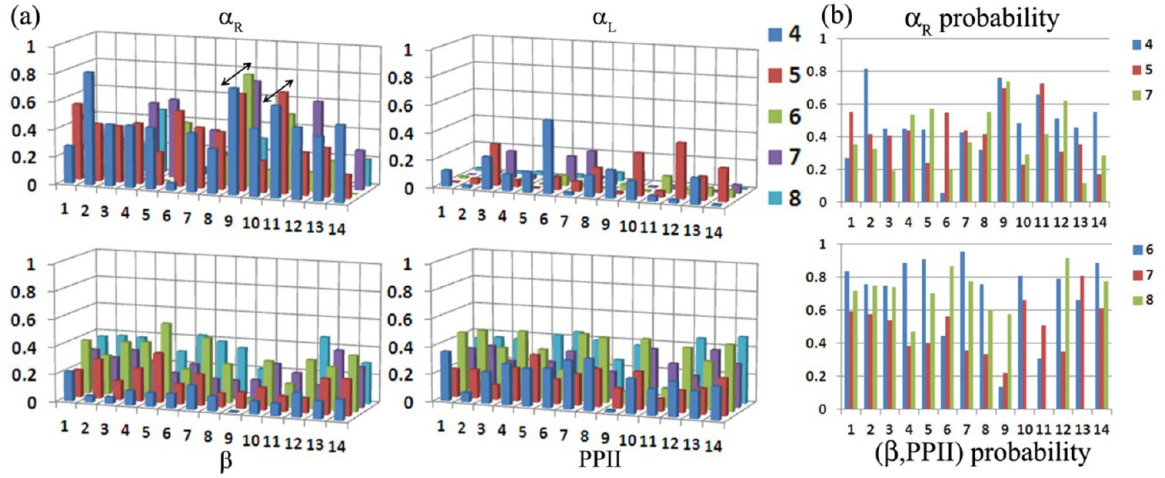


Figure 3. (a) Population distribution of the ϕ/ψ dihedrals in the four commonly observed backbone geometry regions as defined in the Ramachandran surface for amino acids 4 to 8/9. Arrows are used to highlight the formation of an α-helical fold in **9** and **11**. (b) Probability distribution in the α_R region for amino acids 4, 5, and 7 and the summed probability in the extended regions for amino acids 6, 7, and 8. Backbone geometry regions are defined as α_R ($-160° < \phi < -20°$ and $-120° < \psi < 50°$), α_L ($20° < \phi < 160°$ and $-50° < \psi < 120°$), β ($-180° < \phi < -90°$ and $110° < \psi < 180°$), and PPII ($-90° < \phi < -20°$ and $110° < \psi < 180°$).

line with the large $^3J_{HNH\alpha}$ coupling constants (between 7.0 and 8.9 Hz) experimentally observed for the trivaline region.[4,9,10] Thus, while the alanines also sample the folded α-helical conformations the trivaline tail does not. The presence of the commonly found α-helical AXXXA (X = V in as-APF) motif[38] in APF had initially encouraged the hypothesis of a well formed α-helical tail region in APF in the original SAR proposal.[9] However, further structural characterization using circular dichroism (CD) spectroscopy and chemical shift data suggested an extended tail-like region around the trivalines (6**V**, 7**V**, 8**V**) and a folded region around the alanines (4**A** and 5**A**).[10]

Presented in Figure 3a are the population distributions of the ϕ/ψ dihedrals in the four commonly observed backbone geometry regions for the amino acids in the tail region 4 to 8 or 9 depending on the number of amino acids for all the systems.[39] The same for amino acids at positions 2 and 3 (**VP**) is presented in Figure S2 of the SI. The amino acids prefer the extended conformation around **2** and **3** (Figure S2). The only exception to this trend is for 3**P** in **14**, wherein the D-proline also samples 50% α-helical conformation. Other than this the head regions of the glycopeptides (TVP) generally remain only in extended conformations. Variations in the population distribution are observed from 4**A** onwards (Figure 3a). At the outset it is seen that only **9** (Galβ1−3GalNAcα-O-TVPAAAAAA) and **11** (Galβ1−3GalNAcα-O-TVPAAAAA) exhibit a complete α-helical fold with all the amino acids in positions 4 to 8 sampling the α-helical region. This is also established by the large H-bond occupancies between 4**A**-7**A**, 3**P**-6**A**, and 5**A**-8**A** in **9** and **11** (Table 3), which are not seen in the other systems. In all the other cases amino acids at 4, 5, and 7 sampled the α-helical region substantially, while the rest of the amino acids predominantly sample the extended con-

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

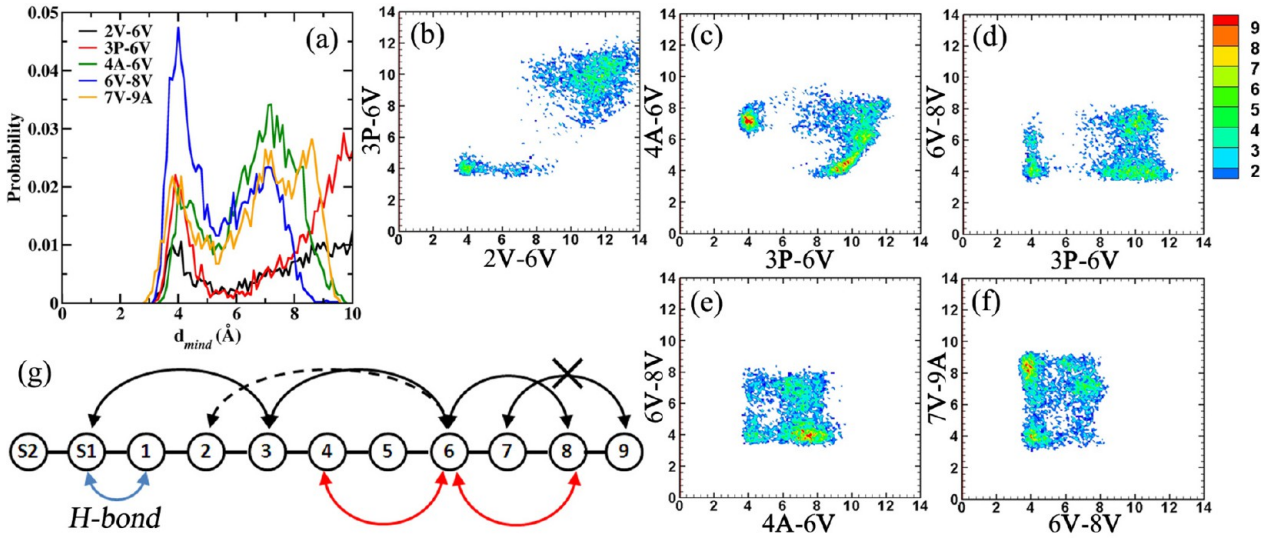


Figure 4. (a) Close contact probabilities observed between the amino acids in as-APF (**1**). Selected 2D distributions of the significant close contact probabilities for (b) 2**V**-6**V**/3**P**-6**V**, (c) 3**P**-6**V**/4**A**-6**V**, (d) 3**P**-6**V**/6**V**-8**V**, (e) 4**A**-6**V**/6**V**-8**V**, and (f) 6**V**-8**V**/7**V**-9**A**. (g) Schematic representation of the various interactiosn in as-APF (**1**). Distances are presented in Å.

formations (β-sheet or polyproline II (PPII)). In Figure 3b we present the probability distributions in the α_R region for amino acids 4, 5, and 7 and the summed probability in the extended regions (β and PPII) for amino acids 6, 7, and 8 for all the systems.

It is observed that in all the systems a valine at position 6 (6**V**) disrupts the formation of the α-helical fold by predominantly favoring the extended geometries (probability >0.6). Substituting 6**V** by 6**A** results in a reduction in the number of extended conformations (**6**: 0.4, **9**: 0.1, and **11**: 0.3). However the remainder of the tail regions (6 to 8/9) do not favor a truly extended β-sheet or PPII geometry, with the peptide backbone sampling both the folded and extended conformations at position 7, irrespective of amino acid (**A** or **V**) type. Thus, it would be incorrect to classify the tail region as being in a truly extended geometry conformation. These results are indicative of stronger long-range interactions in the peptide region. For example, indications of this long-range effect are seen on the population of the folded geometry at 4**A** in **2** and **6**. In **2** the α_R probability increases to 0.8 while the substitution is at position 1 (**S**VPAAVVVA), while in **6** the α_R probability decreases to 0.05 while the substitution is at position 6 (TVPAAAVVA). Analyzing the simulation revealed the formation of various hydrophobic-collapsed structures brought about by the interaction of the side chains. Hence, to identify the structural descriptors of as-APF activity it is important to probe the long-range interactions between the side chains. This is first attempted and described in detail for as-APF (**1**), while the interactions in all the other systems are compared to as-APF.

Close contact distances involving the side chains in as-APF (**1**) were evaluated between all the amino acid side chain pairs separated by one amino acid. The contact probability was evaluated using the minimum distance (*mind*) between the non-hydrogen atoms in the two respective selections. Five significant close contact (d_{mind} < 6.0 Å) probabilities were observed between the amino acid pairs 2**V**-6**V**, 3**P**-6**V**, 4**A**-6**V**, 6**V**-8**V**, and 7**V**-9**A**, with the distance distributions presented in Figure 4a. It is seen that the 6**V**-8**V** pair remains in close contact for a significant portion of the simulation followed by 3**P**-6**V**, 4**A**-6**V**, and 7**V**-9**A**, with the 2**V**-6**V** pair being in the least contact. Following the identification of these close-contact pairs, two-dimensional (2D) distributions were evaluated between all the contact pairs to identify the occurrence of folded geometries. The analysis of the 2D distributions revealed that the 2**V**-6**V** and 3**P**-6**V** contact pairs occur simultaneously (Figure 4b), while the 2**V**-6**V** or 3**P**-6**V** and 4**A**-6**V** pairs are not formed simultaneously (Figure S3a/Figure 4c). Interestingly, both 3**P**-6**V** and 4**A**-6**V** contact pairs can form simultaneously with the 6**V**-8**V** contact pair (Figure 4d and Figure 4e), while they do not form in tandem with the 7**V**-9**A** contact pair (Figure S3b and Figure S3c). However, the tail region from 6**V** to 9**A** favors the simultaneous formation of the 6**V**-8**V** and 7**V**-9**A** contact pairs (Figure 4f). The occurrence of a few of these close-contact pairs revealed the formation of two well-defined folded geometries in as-APF; the first in which the hyrobhobic clustering of 2**V**-6**V**, 3**P**-6**V**, and 6**V**-8**V** side chains lead to the formation of folded structures, and, alternatively, the second in which the hydrophobic clustering of 4**A**-6**V** and 6**V**-8**V** leads to the formation of folded structures. These situations are schematically presented in Figure 4g. It is important to note that in all these collapsed geometries the $-CH_3$ side chain of α-GalNAc interacts with the 3**P** proline ring via hydrophobic interactions and with the 1**T**/1**S** backbone via H-bonding. Working back from the data it was found that the various folded geometries of the peptide backbone could be classified into three distinct regions using the 3**P**-6**V**/4**A**-6**V** 2D distribution.

Presented in Figure 5a is the Boltzmann inverted 3**P**-6**V**/4**A**-6**V** 2D distribution. Inspection reveals a free energy surface that contains two minima separated by a low free energy barrier (<2 kcal/mol). The free energy surface can be classified into three well-defined regions:

Region I: $d_{3P\text{-}6V}$ < 6.0 Å and $d_{4A\text{-}6V}$ > 6.0 Å. This region favors the hydrophobic clustering of 2**V**, 3**P**, 6**V**, and 8**V**.
Region II: $d_{3P\text{-}6V}$ > 6.0 Å and $d_{4A\text{-}6V}$ < 6.0 Å. This region favors the hydrophobic clustering of 4**A**, 6**V**, and 8**V**.
Region III: $d_{3P\text{-}6V}$ > 6.0 Å and $d_{4A\text{-}6V}$ > 6.0 Å. This region favors the extended geometries for the tail region PAAVVVA. Analysis of the conformations belonging to

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

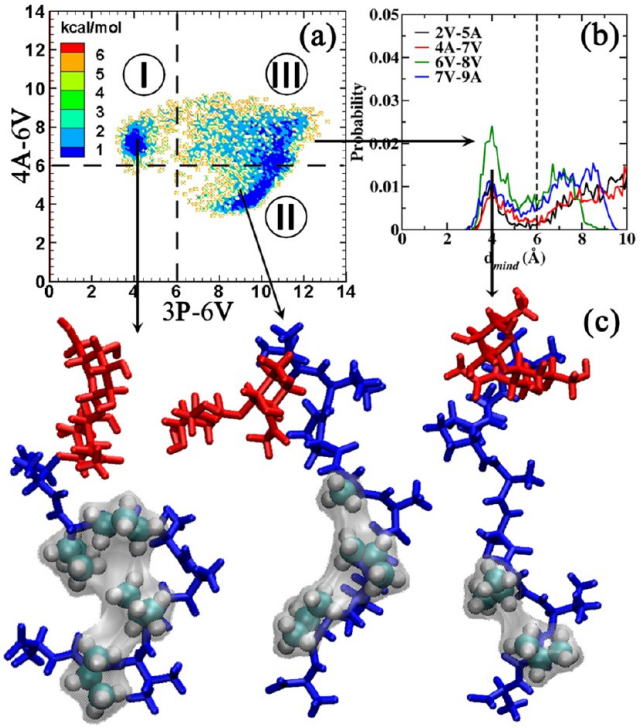


Figure 5. (a) Boltzmann inverted 3**P**-6**V**/4**A**-6**V** 2D distribution for **1**. Regions identified using distance cutoffs are distinguished using dotted lines. (b) Close contact probabilities observed between conformations belonging to region III. (c) Representative structures obtained by clustering the conformations belonging to each region using the backbone ϕ/ψ dihedrals. Color coding: the carbohydrate region is presented in red and the peptide region in blue. Amino acids involved in the hydrophobic clustering are presented in VDW representation. The surface representation is used to highlight the hydrophobic clustering.

this region revealed significant close contacts between 6**V**-8**V** (Figure 5b). This region can be further classified into the 6**V**-8**V** close contact ($d_{6V\text{-}8V}$ < 6.0 Å) and 6**V**-8**V** extended regions ($d_{6V\text{-}8V}$ > 6.0 Å).

The population distribution of the conformations in these three regions for all the systems is presented in Table 4. From

Table 4. Population Distribution of the Conformations in the Regions Identified by the 3P-6V/4A-6V 2D Distribution

			region III		
compd no.	region I	region II	$d_{6V\text{-}8V}$ < 6.0 Å	$d_{6V\text{-}8V}$ > 6.0 Å	outliers
1	19.1%	29.9%	29.6%	21.3%	0.2%
2	29.6%	13.4%	31.5%	23.8%	1.7%
3	43.2%	13.9%	9.3%	31.2%	2.3%
4	25.4%	11.0%	51.8%	11.1%	0.7%
5	15.4%	16.3%	46.5%	21.7%	0.0%
6	10.3%	2.4%	23.7%	63.6%	0.0%
7	12.5%	38.0%	21.3%	27.8%	0.4%
8	1.9%	51.7%	21.6%	24.7%	0.1%
9	63.6%	0.1%	4.5%	31.8%	0.0%
10	5.5%	37.0%	30.6%	26.5%	0.3%
11	58.1%	4.7%	4.5%	32.7%	0.0%
12	6.4%	64.0%	18.4%	10.8%	0.4%
13	17.0%	26.4%	33.4%	22.9%	0.2%
14	2.6%	28.7%	37.6%	30.6%	0.4%

the population distributions it is seen that the population of the folded conformations in region II (plotted for clarity in Figure 6), which corresponds to the hydrophobic clustering of amino

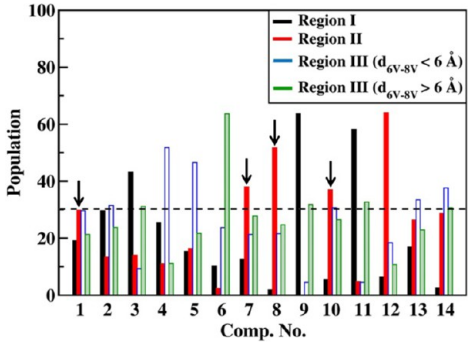


Figure 6. Population distribution corresponding to the various regions of the 3**P**-6**V**/4**A**-6**V** distribution.

acids at positions 4, 6, and 8, shows correlation with the observed activity profile (Table 1). Both **1** and **10**, which exhibit 100% activity, favor the hydrobhobic clustering of amino acids at 4, 6, and 8 (population ∼30%). The derivatives which show 0% activity (**4** and **5**) show an appreciable reduction in population in region II (population ∼15%). However, this descriptor is limited as it fails to describe the 0% activity observed in amino acid only derivatives (**12**, **13**) and the D-proline derivative (**14**), which show appreciable population in region II. Thus, the observed activity in as-APF is a combination of various structural features. This becomes clearer on attempting a case-by-case analysis of the various as-APF derivatives. Accordingly, the as-APF derivatives are divided into various subclasses based on the position of the subsitution.

DISCUSSION

A case-by-case summary is presented of the various as-APF derivatives which lead to an improved understanding of as-APF and a refined SAR model.

Serine Derivatives (2 and 3). The serine derivatives form the first class of as-APF derivatives, wherein the modifications directly affect the carbohydrate−peptide interactions. The observed activity profile of these derivatives can be explained by the effect of the modification on the strength of the intramolecular H-bond (α-GalNAc (HN) − 1**T**/1**S** (O)) and hydrobhobic interactions (α-GalNAc ($-CH_3$) − 3**P** (proline ring)) between the carbohydrate and the peptide regions (Figure 1). It is also observed that the weakening of the carbohydrate−peptide interactions affects the peptide−peptide interactions, as an appreciable reduction is seen in the conformations belonging to region II (Table 4) of the 3**P**-6**V**/4**A**-6**V** 2D distribution (population ∼13%). Thus, it is seen that the interactions in the head region propagate to the tail region of as-APF.

Compound 4. The modification of 2**V** to 2**L** in **4** leads to a complete loss of activity. This effect can be traced to the depletion of the population in region II of the population in region II of the 3**P**-6**V**/4**A**-6**V** 2D distribution (Table 4 and Figure 6) to 11.0%. Analyzing the close contacts in **4** revealed that increasing the length of the hydrophobic side chain at position 2 led to an increase in the number of collapsed structures in region I (population: 25.4%) via stronger hydrophobic interactions between 2**L**, 3**P**, 6**V**, and 8**V**, an

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

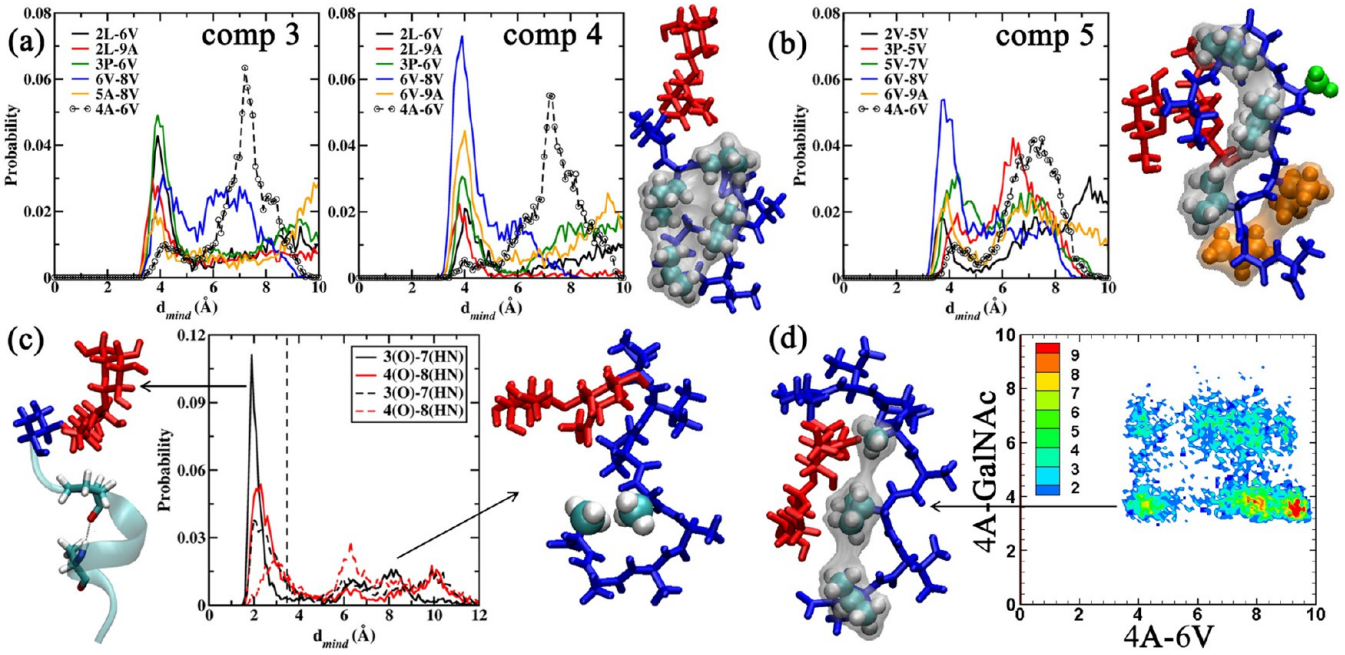


Figure 7. (a) Close contact probabilities observed between the amino acids in **3** and **4**. Representative structure for **4** obtained by clustering the conformations belonging to region I. Amino acids 2**L**, 3**P**, 6**V**, and 8**V** are shown in VWD representation to highlight the hydrophobic clustering. (b) Close contact probabilities in **5**. Representative structure for **5** obtained upon clustering conformations in region III. Amino acids 2**V**, 5**V**, and 7**V** are presented in VWD representation with the corresponding surface to highlight the hydrophobic clustering. The other amino acids 4**A** (green VDW representation), 6**V** and 8**V** (orange VDW representation) highlight other features. (c) H-bond distribution for 3**P**(O)-7**A**(HN) (black lines) and 4**A**(O)-8**A**(HN) (red lines) for **9** (solid lines) and **11** (dashed lines). Representative structures for **9** and **11** depicting the favored α-helical folded and extended geometries. The peptide tail region is presented in new cartoon representation to highlight the α-helical fold for **9**. Amino acids 4**A** and 8**A** are presented in VDW representation to highlight the hydrophobic clustering for **11.** (d) 4**A**-6**V**/4**A**-GalNAc 2D distribution for **14**. The representative structure corresponding to the collapsed region is presented. Amino acids 4**A**, 6**V**, and 8**V** are presented in VDW representation. General color coding: carbohydrate region is presented in red and the peptide region in blue. All distances are presented in Å.

effect that is also seen in **3** (region I population: 43.2%), the other 2**L** variant. The strong hydrophobic clustering induced due to 2**L** in **4** is depicted in the representative structure for region I presented in Figure 7a. The difference in the APF activity between **3** and **4** is thus due to the reduction in the 4**A**-6**V** contacts in **4** when compared to **3** (Figure 7a dotted lines), which results in the lowering of the population in region II.

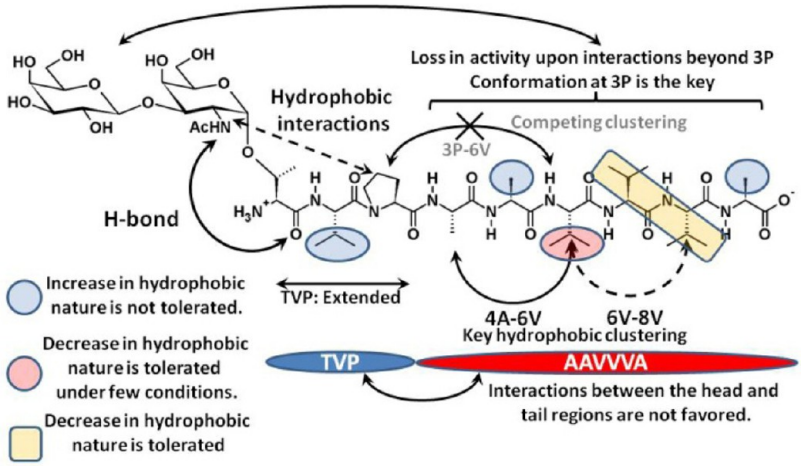
Compound 5. The substitution of 5**A** to 5**V** also led to a complete loss in activity. This substitution leads to the stiffening of the underlying peptide backbone resulting in the loss of folded α_R conformations (Figure 3b), which are necessary to bring about the hydrophobic interactions between 4**A** and 6**V** (population in region II: 16.3%). The stiffening was induced by the increase in the number of hydrophobic contacts with 5**V**, due to an increase in the hydrophobic nature of the amino acid. Figure 7b presents the result of the close contact analysis for **5** along with the representative structure for region III ($d_{6V\text{-}8V}$ < 6.0 Å) (population 46.5%) to illustrate the effect of the substitution.

Trivaline Modifications (Compounds 6, 7, and 8). With 6**V** being involved in the formation of the hydrophobic collapsed structures (region II), it is of interest to study the site-by-site modifications of the trivaline region. From the population distribution presented in Figure 6 it is observed that the substitution of 6**V** to 6**A** (comp **6**) adversely affects the formation of the 4−6−8 collapsed structures with the population in region II going down to 2.4%. In fact the substitution resulted in the formation of extended geometries, with region III ($d_{6V\text{-}8V}$ > 6.0 Å) being populated for close to 63.6%. It is seen that the modification of 7**V** to 7**A** or 8**V** to 8**A** in fact improves the 4**A**-6**V** interactions (population in region II increasing to 38.0% and 51.7% for **7** and **8**, respectively). Thus, reducing the hydrophobic content at positions 7 and 8 improves the hydrophobic collapse involving amino acids at position 4 and 6. Here, it must also be mentioned that an attempt to increase the hydrophobicity at position 9 adversely affected the biological activity. Two attempts at this, TVPAAVVV**L** and TVPAAVVA**V**, resulted in activities of 0.01% and inactive variants.[9] This could be due to this substitution leading to competing hydrophobic collapse involving position 6 which would disrupt the clustering of 4**A** with 6**V**.

All Alanine Modifications (9 and 11). The substitution of all the valines in the tail region leads to a separate class of as-APF derivatives. From the analysis thus far it is observed that both 7**V** and 8**V** modifications to 7**A** and 8**A** are favored by as-APF provided the hydrophobicity of the tail amino acid remained unchanged, i.e., 9**A**. However, changing 6**V** to 6**A** has an adverse effect on the population in region II (population being lowered to 2.4% for **6**). All the above observations were made for the cases wherein a valine at position 6 (6**V**) acted as an α-helix breaker. However, the replacement of 6**V** to 6**A** with the subsequent amino acids also favoring the α-helix geometry opens up a new conformational profile wherein the tail region of as-APF (TVPAAAAAA or TVPAAAAA) may favor an α-helical fold (Figure 3a and Table 3). Interestingly, the shortening of the tail region by one alanine (**9** to **11**) led to an increase in activity of 2 orders of magnitude back to "full" potency (1% to 100%). Analyzing the H-bond profiles for these two systems indicates that the change in one amino acid

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Scheme 2. Schematic Illustration of the Structural Requirements for Biological Activity



resulted in the weakening of the α-helical geometry as is evident by the lower H-bond occupancies between i and i+4 H-bond pairs, 3**P**-7**A** (0.589 to 0.436) and 4**A**-8**A** (0.543 to 0.209) on comparing **9** with **11** (Table 3).

This is also illustrated by the H-bond distribution for these two H-bonds presented in Figure 7c. The shortening of the tail by one amino acid has a drastic effect on the strong 3**P**-7**A** H-bond. In **9** this H-bond was observed in close to 58.9% of the conformations (population calculated using a distance cutoff of 3.5 Å), while in **11** this dropped to 43.6%. It is noticed that the weakening of the H-bond adds flexibility to the peptide, which resulted in hydrophobic interactions being favored over H-bonding. Thus, even in the case of a strong α-helix "inducer," the reduction in the helical content resulted in a potent derivative. Interestingly, the attempt to replace alanine with β-alanine in **11**, a modification which induces helical content,[40] resulted in a 3 orders of magnitude loss in activity (0.1%).[10] Also the replacement of the trivaline tail by triglycine in **11** resulted in a complete loss of activity which is consistent with the observation described above, viz., a triglycine tail would not favor any hydrophobic clustering and would instead favor an α-helical fold, thereby lowering the APF activity of this derivative.[10] These results consistently highlight the fact that enhancement of the helical content in fact adversely affects the activity.

Peptide Only Derivatives (12 and 13). The experimental results clearly indicate that the carbohydrate moiety, at least Galβ1−3GalNAc, is required for the activity. The analysis of the peptide only fragments is insightful as it reveals the influence of the amino acid sequence on the conformations adopted by the peptide region. From the population distribution presented in Table 4 and Figure 6 it is seen that the peptide only fragments significantly populate region II of the 3**P**-6**V**/4**A**-6**V** distribution, with the population being 64.0% and 26.4% for **12** and **13**, respectively. Thus, the hydrophobic clustering of amino acids at positions 4, 6, and 8 is inherent to the peptide. This observation is consistent with the CD results which found the structure of the TVPAAVVV peptide only fragment (comp **13**) to be more organized than the glycosylated fragment (comp **10**).[10] In all the CD results for the APF derivatives, the presence of the carbohydrate induces structural disorder into the systems. These results also highlight the fact that there is interplay of the carbohydrate−peptide and peptide−peptide interactions in AFP systems which governs the activity.

D-Proline (14). The substitution of proline with D-proline led to a complete loss of activity. However, it was found that of all the inactive derivatives of as-APF, the D-proline derivative was one of only two that blocked the antiproliferative activity of as-APF, thereby acting as an antagonist.[6] Structural analysis of **14** reveals that it retains the peptide conformation observed in **1** and **10** (Table 4 and Figure 6), favoring the hydrophobic interactions between 4**A**, 6**V**, and 8**V**. However, the change in the conformation at the proline ring exposes the peptide to the carbohydrate with 4**A** interacting with GalNAc. This is illustrated by the distance distribution between GalNAc and 4**A** for all the systems, which is presented in Figure S4 of the SI. It is observed that 4**A** interacts with GalNAc significantly in **14** and for an appreciable time in **2** (where T is substituted by S). For all other systems GalNAc interacts only with 3**P**. In Figure 7d, the 4**A**-6**V**/4**A**-GalNAc 2D distribution is presented which highlights the fact that **14** can accommodate both close contacts in tandem with each other. This feature highlights the influence of the proline ring on the activity of APF wherein the proline ring acts as a barrier to curb the interactions between the peptide head region (**TV**) and the rest of the peptide (**AAVVVA**).

■ CONCLUSIONS

Presented is an overview of the structural features of as-APF and analogs that impact biological activity. Scheme 2 and the following text summarize the key findings.

a) The carbohydarte interacts with the peptide backbone via strong H-bond (α-GalNAc (HN) − 1**T**/1**S** (O)) and hydrobhobic interactions (α-GalNAc ($-CH_3$) − 3**P** (proline ring)). The presence of the carbohydrate is required for the activity as demonstrated by 0% activity for the peptide only compounds **12** and **13**. Interactions of the carbohydrate with the rest of the peptide can cause a significant loss in activity, as seen for **14** and **2**.

b) The head region of the peptide portion (**TVP**) adopts an extended conformation and generally does not interact with the tail region (**AAVVVA**).

c) The hydrophobic clustering of amino acids at positions 4, 6, and 8 leads to the formation of a pseudoextended geometry that is predicted to be the active as-APF conformation as its presence correlates with the experimental trends. While the

dx.doi.org/10.1021/ci400147s | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

clustering of 4 and 6 is seen to be imporant for the activity, the clustering of 6 and 8 is found to be auxiliary to the activity profile.

d) The hydrophobic clustering between positions 3 and 6 competes with the 4−6 clustering. Thus, subsitutions which favor these interactions adversely affect the activity as observed in **2**, **3**, and **4**.

e) Increasing the hydrophobic nature at amino acids 2, 5, and 9 adversely affects the activity as it disrupts the hydrophobic clustering of 4−6−8. An increase in the hydrophobic nature at position 2 leads to increased 3−6 clustering and interactions between the head (**TVP**) and the tail regions of the peptide (**AAVVVA**). Changes at amino acids 5 and 9 lead to the disruption of the 4−6−8 clustering and favor the formation of alternate peptide conformations in the tail region (**AAVVVA**).

f) The activity of as-APF is highly sensitive to the hydrophobic content of the amino acid at position 6. A valine at this position acts as a α-helix breaker. Substituting valine with alanine is favored only if the subsequent amino acids are also substituted to alanines.

g) Substitution of valine to alanine at positions 7 and 8 is favored as it enhances the 4−6−8 clustering.

h) The all alanine tail derivatives (**9**, **11**) form a separate class of as-APF analogs which cannot be described by the 3**P**-6**A**/4**A**-6**A** 2D distribution profile (Table 4 and Figure 6). On comparing the activities of these two derivatives a 2 orders of magnitude increase in the activity was observed upon a loss in the α-helix character, indicating that conformations which favor hydrophobic clustering over H-bonding are more potent.

In summary, our analysis shows that the hydrophobic clustering of amino acids at position 4−6−8 leads to potent APF glycopeptides derivatives provided the conformation at 3P is maintained. This hydrophobic clustering is found to be very sensitive to single point mutations. However the analysis reveals regions that can tolerate both an increase as well as decrease in the hydrophobic nature of a substitution, indicating that appropriate modifications may lead to more potent as-APF derivatives. With regards to the antagonist activity, it is found that the conformational transition brought about by substituting proline with D-proline (antagonist analog of as-APF) is also partially observed in mutating **T** to **S**.

Comprehensive NMR and CD analysis has been performed on many of these APF analogues,[4,9,10] However, the disordered nature of APF glycopeptides precludes determination of the conformational properties of the peptides with these techniques alone. The computational sampling methods utilized here produce a more complete picture of the glycopeptide conformational properties to be obtained allowing for determination of a SAR model that should greatly assist in the development of glycopeptide-mimetics as viable therapeutic lead compounds for IC and cancer.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

^3J coupling constants calculated from HREX simulations for the H−N−C_2−H_2 (for GalNAc), N−C_α−C_β−O_1, and C−N−C_α−C. GalNAc (HN) − 1**T**/1**S** (O) H-bond distribution. Population distribution of the ϕ/ψ dihedrals for 2**V**/**L** and 3**P**. 2D distributions of the close contact probabilities for 2**V**-6**V**/4**A**-6**V**, 3**P**-6**V**/7**V**-9**A**, and 4**A**-6**V**/7**V**-9**A**. Close contact probabilities between GalNAc and 4**A**. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 410 706-7442. Fax: 410 706-5017. E-mail: alex@outerbanks.umaryland.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support from the NIH (GM070855) is acknowledged. The modified REPDSTR module to perform periodic boundary simulations with CHARMM c36a2 was kindly provided by Prof. Benoit Roux and Dr. Wei Jiang (University of Chicago).

■ ABBREVIATIONS:

APF: antiproliferative factor; SAR: structure−activity relationship; IC: interstitial cystitis; as-APF: asialo synthetic derivative of APF; CKAP4: cytoskeletal-associated protein 4; HREX: Hamiltonian replica exchange; MD: molecular dynamics; ROE: rotating frame nuclear Overhauser effect; Gal: galactose; GalNAc: N-acetylgalactosamine; Neu5Ac: N-acetylneuraminic acid

■ REFERENCES

(1) Clemens, J. Q.; Meenan, R. T.; Rosetti, M. C. O. K.; Gao, S. Y.; Calhoun, E. A. Prevalence and incidence of interstitial cystitis in a managed care population. *J. Urol.* **2005**, *173*, 98−102.

(2) Parsons, J. K.; Parsons, C. L. The historical origins of interstitial cystitis. *J. Urol.* **2004**, *171*, 20−22.

(3) Keay, S.; Zhang, C. O.; Trifillis, A. L.; Hise, M. K.; Hebel, J. R.; Jacobs, S. C.; Warren, J. W. Decreased sup 3 H-thymidine incorporation by human bladder epithelial cells following exposure to urine from interstitial cystitis patients. *J. Urol.* **1996**, *156*, 2073−2078.

(4) Keay, S. K.; Szekely, Z.; Conrads, T. P.; Veenstra, T. D.; Barchi, J. J.; Zhang, C. O.; Koch, K. R.; Michejda, C. J. An antiproliferative factor from interstitial cystitis patients is a frizzled 8 protein-related sialoglycopeptide. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11803−11808.

(5) Zhang, C.-O.; Wang, J.-Y.; Koch, K. R.; Keay, S. Regulation of tight junction proteins and bladder epithelial paracellular permeability by an antiproliferative factor from patients with interstitial cystitis. *J. Urol.* **2005**, *174*, 2382−2387.

(6) Keay, S.; Kaczmarek, P.; Zhang, C. O.; Koch, K.; Szekely, Z.; Barchi, J. J.; Michejda, C. Normalization of proliferation and tight junction formation in bladder epithelial cells from patients with interstitial cystitis/painful bladder syndrome by d-proline and d-pipecolic acid derivatives of antiproliferative factor. *Chem. Biol. Drug. Des.* **2011**, *77*, 421−430.

(7) Koch, K.; Zhang, C.-O.; Kaczmarek, P.; Barchi, J., Jr.; Guo, L.; Shahjee, H.; Keay, S. The effect of a novel frizzled 8-related antiproliferative factor on in vitro carcinoma and melanoma cell proliferation and invasion. *Invest. New Drugs* **2012**, *30*, 1849−1864.

(8) Conrads, T. P.; Tocci, G. M.; Hood, B. L.; Zhang, C.-O.; Guo, L.; Koch, K. R.; Michejda, C. J.; Veenstra, T. D.; Keay, S. K. CKAP4/p63 is a receptor for the frizzled-8 protein-related antiproliferative factor from interstitial cystitis patients. *J. Biol. Chem.* **2006**, *281*, 37836−37843.

(9) Kaczmarek, P.; Keay, S. K.; Tocci, G. M.; Koch, K. R.; Zhang, C. O.; Barchi, J. J.; Grkovic, D.; Guo, L.; Michejda, C. J. Structure-activity relationship studies for the peptide portion of the bladder epithelial cell antiproliferative factor from interstitial cystitis patients. *J. Med. Chem.* **2008**, *51*, 5974−5983.

(10) Kaczmarek, P.; Tocci, G. M.; Keay, S. K.; Adams, K. M.; Zhang, C. O.; Koch, K. R.; Grkovic, D.; Guo, L.; Michejda, C. J.; Barchi, J. J.

Structure-activity studies on antiproliferative factor (APF) glycooctapeptide derivatives. *ACS Med. Chem. Lett.* **2010**, *1*, 390−394.
(11) Mallajosyula, S. S.; MacKerell, A. D., Jr. Influence of solvent and intramolecular hydrogen bonding on the conformational properties of O-linked glycopeptides. *J. Phys. Chem. B* **2011**, *115*, 11215−11229.
(12) Brooks, B. R.; Brooks, C. L., 3rd; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−614.
(13) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(14) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(15) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543−64.
(16) Guvench, O.; Hatcher, E. R.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.
(17) Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr. CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates and inositol. *J. Chem. Theory Comput.* **2009**, *5*, 1315−1327.
(18) Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr. CHARMM additive all-atom force field for aldopentofuranoses, methyl-aldopentofuranosides, and fructofuranose. *J. Phys. Chem. B* **2009**, *113*, 12466−76.
(19) Raman, E. P.; Guvench, O.; MacKerell, A. D., Jr. CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *J. Phys. Chem. B* **2010**, *114*, 12981−94.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(21) *Computational Biochemistry and Biophysics*; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel-Dekker, Inc,: New York, 2001.
(22) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; John Wiley & Sons: New York, 1988.
(23) Beglov, D.; Roux, B. An integral equation to describe the solvation of polar molecules in liquid water. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.
(24) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(25) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **1994**, *15*, 667−683.
(26) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(27) Nose, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(28) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(29) Woodcock, H. L.; Hodošček, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F.; Brooks, B. R. Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* **2007**, *28*, 1485−1502.
(30) Mobli, M.; Almond, A. N-Acetylated amino sugars: the dependence of NMR 3J(HH)-couplings on conformation, dynamics and solvent. *Org. Biomol. Chem.* **2007**, *5*, 2243−2251.
(31) Demarco, A.; Llinás, M.; Wüthrich, K. Analysis of the 1H-NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers* **1978**, *17*, 617−636.
(32) Vuister, G. W.; Bax, A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772−7777.
(33) Corzana, F.; Busto, J. H.; Engelsen, S. B.; Jimenez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Avenoza, A. Effect of beta-O-glucosylation on L-Ser and L-Thr diamides: a bias toward alpha-helical conformations. *Chem.—Eur. J.* **2006**, *12*, 7864−7871.
(34) Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A. New insights into alpha-GalNAc-Ser motif: influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **2006**, *128*, 14640−14648.
(35) Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; de Luis, M. G.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A. Serine versus threonine glycosylation: the methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **2007**, *129*, 9458−9467.
(36) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.; Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. Principles of mucin architecture: structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains. *J. Am. Chem. Soc.* **2002**, *124*, 9833−9844.
(37) Hashimoto, R.; Fujitani, N.; Takegawa, Y.; Kurogochi, M.; Matsushita, T.; Naruchi, K.; Ohyabu, N.; Hinou, H.; Gao, X. D.; Manri, N.; Satake, H.; Kaneko, A.; Sakamoto, T.; Nishimura, S.-I. An efficient approach for the characterization of mucin-type glycopeptides: the effect of O-glycosylation on the conformation of synthetic mucin peptides. *Chem.—Eur. J.* **2011**, *17*, 2393−2404.
(38) Kleiger, G.; Grothe, R.; Mallick, P.; Eisenberg, D. GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* **2002**, *41*, 5990−5997.
(39) Best, R. B.; Buchete, N.-V.; Hummer, G. Are current molecular dynamics force fields too helical? *Biophys. J .* **2008**, *95*, L07−L09.
(40) Seebach, D.; Beck, A. K.; Bierbaum, D. J. The world of β- and γ-peptides comprised of homologated proteinogenic amino acids and other components. *Chem. Biodiversity* **2004**, *1*, 1111−1239.