Supporting Information

Influence of Solvent and Intra-molecular Hydrogen Bonds on the Conformation of *O*-Linked Glycopeptides.

Sairam S. Mallajosyula and Alexander D. MacKerell, Jr.

Department of Pharmaceutical Sciences, 20 Penn Street HSF II, University of Maryland, Baltimore, Maryland 21201.

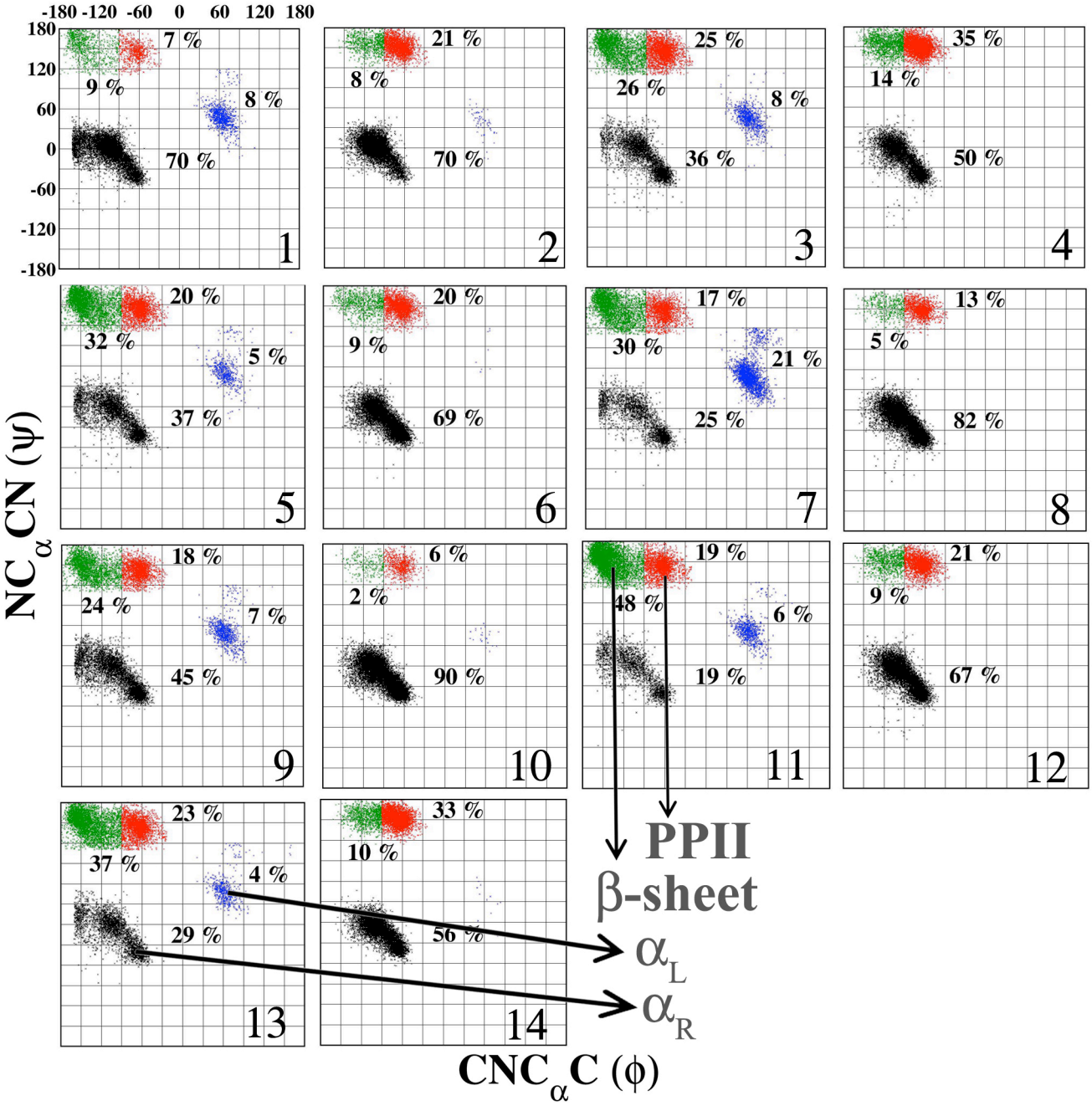

Figure S1: Two-dimensional distributions for ϕ(CNCαC) versus ψ(NCαCN) dihedrals, given in degrees, calculated from HREX simulations for all the model glycopeptide systems. The (ϕ/ψ) space is classified into α (-160° < ϕ < -20° and -120° < ψ < 50°) (black), β (-180° < ϕ < -90° and 110° < ψ < 180°) (green), ppII (-90° < ϕ < -20° and 110° < ψ < 180°) (red) and α_R (20° < ϕ < 160° and -50° < ψ < 120°) (blue) regions. This classification of the (ϕ/ψ) space has been used earlier by Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07– L09. The α_R region is defined as the mirror image of the α region. The total probability corresponding to each region has also been presented.

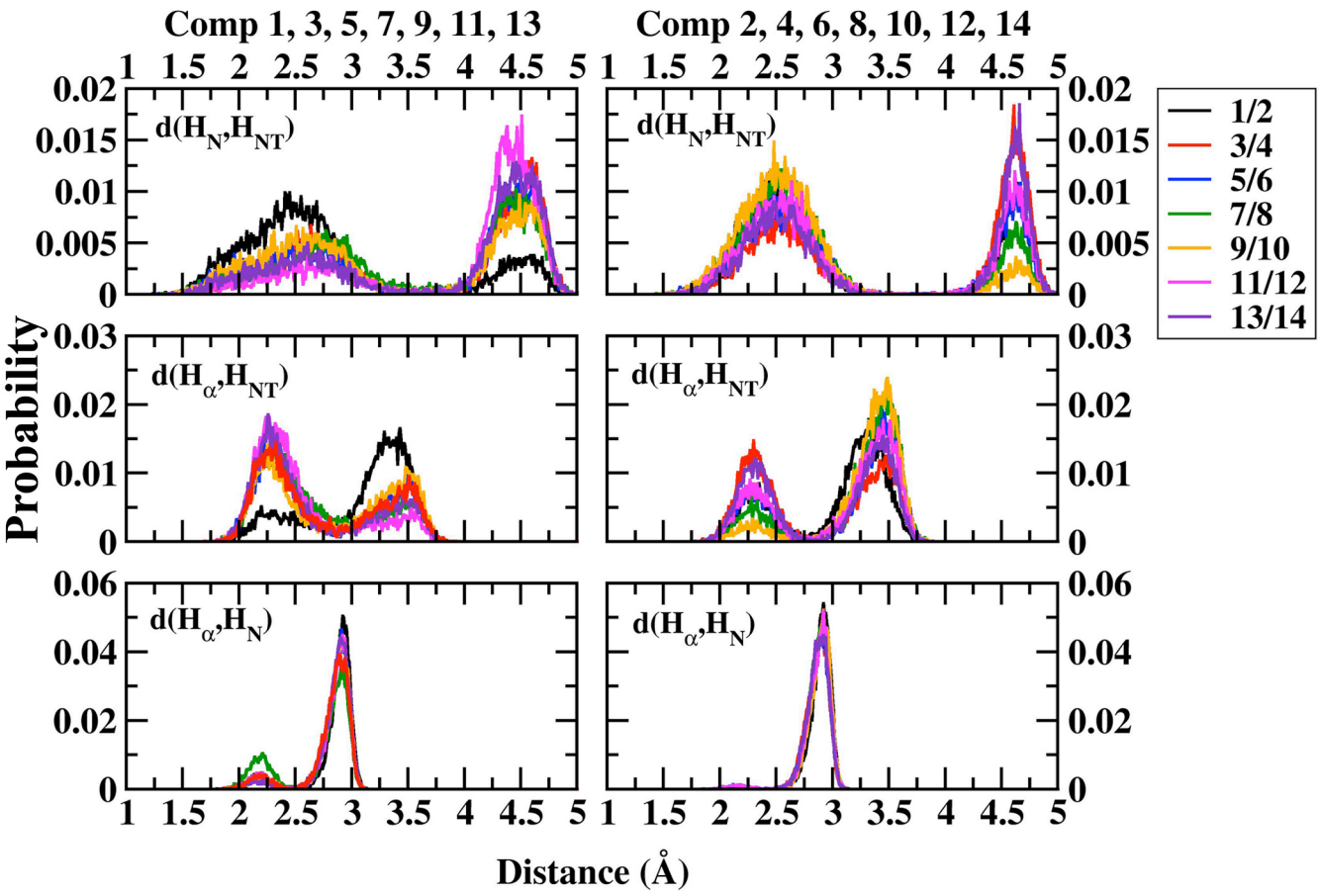

Figure S2: Proton-proton distance distributions of $d(H_N,H_{NT})$, $d(H_\alpha,H_{NT})$ and $d(H_\alpha,H_N)$ from HREX MD simulations for all the model glycopeptide systems. For all the proton-proton distances the serine linkage (odd–numbered systems) distributions are presented in the left panel while the threonine linkage (even–numbered systems) distributions are presented in the right panel. All the distances are given in angstroms and calculated as an $\langle r^{-6}\rangle^{-1/6}$ average.

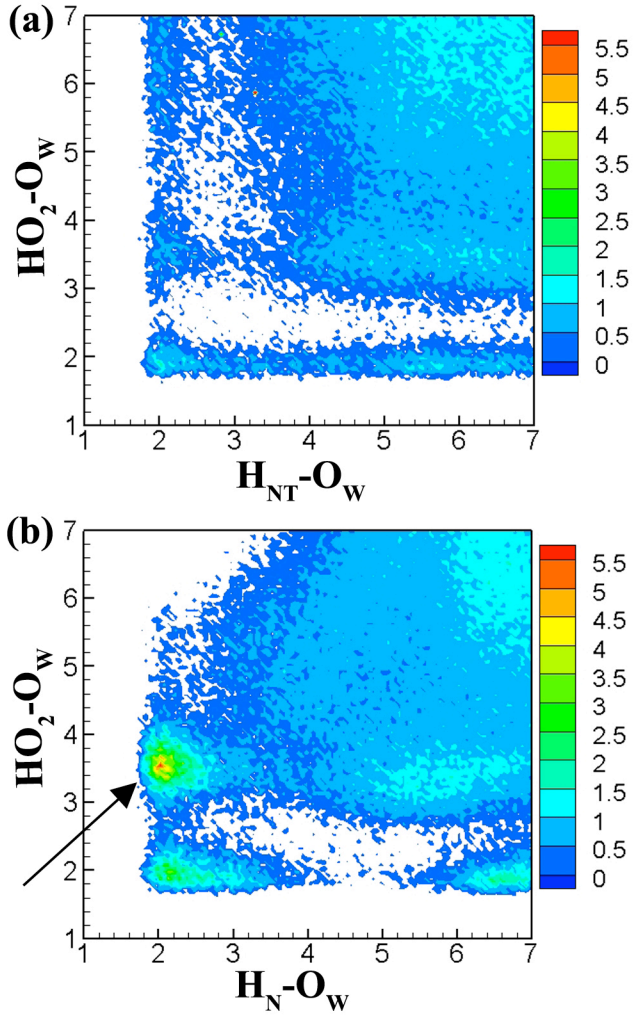

Figure S3: 2D radial distribution plots to identify significant $HO_2…H_2O…H_{NT}$ and $HO_2…H_2O…H_N$ water bridges from HREX simulations for (a) **13** and (b) **14**. Arrow indicates the secondary water density for **14.** All the distances are given in angstroms (Å).

Table T1: NOE derived distances from NMR experiments and mean Proton-Proton distances from MD simulations calculated as a $\langle r^{-6} \rangle^{-1/6}$ average for all model glycopeptide systems (compounds 1 to 14). Sampling of the proton-proton distances exhibiting bimodal distribution from MD simulations.

Comp	Experiment			MD			$d(H_N,H_{NT})$		$d(H_\alpha,H_{NT})$	
	d1[a]	d2[a]	d3[a]	d1[a]	d2[a]	d3[a]	<3.6Å	>3.6Å	<2.9Å	>2.9Å
1	2.9 [40]	2.3 [40]	2.9 [40]	2.31	2.72	2.70	77.8%	22.2%	28.1%	71.9%
2	2.8[41]	2.4[41]	2.9[41]	2.36	3.26	2.91	100.0%	0.0%	1.0%	99.0%
3	2.9[41]	2.2[41]	2.6[41]	2.69	2.41	2.71	40.9%	59.1%	66.2%	33.8%
4	3.0[41]	2.3[41]	2.8[41]	2,57	2.62	2.87	62.9%	37.1%	37.6%	62.4%
5	2.7[43]	2.3[43]	2.6[43]	2.86	2.36	2.68	29.3%	70.7%	79.8%	20.2%
6	2.3[43]	2.5[43]	2.8[43]	2,56	2.63	2.86	65.8%	34.2%	34.5%	65.5%
7	3.0[42]	2.3[42]	2.7[42]	2.82	2.33	2.77	19.9%	80.1%	82.1%	17.9%
8	2.8[42]	--	--	2,46	2.80	2.87	79.2%	20.8%	21.2%	78.8%
9				2.48	2.54	2.69	61.7%	38.3%	48.3%	51.7%
10				2,41	2.87	2.87	84.6%	15.4%	16.1%	83.9%
11				2.78	2.38	2.68	30.2%	69.8%	76.5%	23.5%
12				2.47	2.74	2.87	74.8%	25.2%	25.6%	74.4%
13				2.49	2.52	2.69	58.6%	41.4%	51.4%	48.6%
14				2.37	3.36	2.89	100.0%	0.0%	0.0%	100.0%

d1 - $d(H_N,H_{NT})$, d2 - $d(H_\alpha,H_{NT})$, d3 - $d(H_\alpha,H_N)$, [a]Distances are given in angstroms.

Table T2: J-coupling constants and associated sampling of the H-N-C_2-H_2 dihedrals in the six model glycopeptides with the acetlyamino (-$NHCOCH_3$) side chain.

Comp	Expt	MD	Motional Aver	Anti[a]	Int[a]	Ecl[a]
1	9.2[40]	7.6	7.4	86.2%	13.8%	0.0%
2	9.5[41]	8.6	8.3	92.9%	7.1%	0.0%
3	9.6[41]	8.6	8.3	1.0%	0.0%	99.0%**
4	9.9[41]	10.4	10.0	99.0%	0.5%	0.5%
7	9.4[42]	10.4	10.0	99.5%	0.5%	0.0%
8	9.4[42]	9.8	9.4	65.7%	0.3%	34.0%
	RMS Diff[b]	0.9	0.8			
	Average[b]	0.3	0.6			

[a]Anti: Distributions binned from -120˚ to 120˚, Intermediate (Int): Distributions binned from -120˚ to –60˚ and 60˚ to 120˚, Eclipsed (Ecl): Distributions binned from -60˚ to 60˚.

** For the β-GalNAc Ser linkage (**3**) during the standard MD simulation it was found that the H-N-C_2-H_2 dihedral favored the eclipsed conformation over the anti conformation. Even though an initial anti conformation was chosen for the H-N-C_2-H_2 dihedral, the dihedral flipped to the eclipsed conformation within 500 ps of the production run and did not revert back to the anti conformation. Note that this may be due to the convergence issues associated with short MD simulations for carbohydrates as has been seen in previous studies (Sattelle B. M. et. al., *J. Am. Chem. Soc.*, **2010**, *132*, 13132–13134). On the other hand the HREX simulations starting from the same minimized coordinates as the standard MD simulations sampled the anti and eclipsed conformations for 78% and 21% of the total simulation time, respectively, with close to 400 exchanges between the two conformational bins (Table 3).

From the HREX results when compared to the other systems the population of the β-GalNAc Ser linkage (**3**) in the anti conformation is lowered by close to 10%. Upon searching the PDB database five crystal structures were identified where the H-N-C2-H2 dihedral in β-GalNAc adopted the eclipsed conformation (PDB id: 3GH7, 3DWQ, 1ULG, 2FYD, 2DNJ), indicating that the sampling of the eclipsed conformation is not a simulation artifact. These results again highlight the advantage of HREX conformational sampling and the need for longer sampling time for standard MD simulations.

Table T3: J-coupling constants and the associated sampling of the N-C_α-C_β-O_1 (χ_s) dihedral for the fourteen glycopeptides. The standard deviation of the calculated coupling constant values is presented in parenthesis. J-coupling constants in Hz.

Comp	Expt[d]	MD[d]	Motional Aver[d]	+60[oa]	-60[oa]	±180[oa]
1	5.5,4,5 [40]	7.8(4.2),3.9(3.1)	5.3(1.4),5.4(2.4)	46.6%	10.8%	42.6%
2	2.5 [41]	3.0(0.8)	3.1(0.8)	100.0%	0.0%	0.0%
3	6.8,... [41]	7.1(4.0),8.5(4.7)	5.2(1.4),9.5(3.2)	8.2%	58.4%	33.4%
4	3.5 [41]	3.1(1.0)	3.3(1.0)	100.0%	0.0%	0.0%
5	4.6,... [43]	9.9(3.9),5.1(3.8)	6.0(1.2),7.2(2.6)	11.2%	20.5%	68.3%
6	3.4[43]	3.0(0.9)	3.2(0.9)	100.0%	0.0%	0.0%
7	6.6.5,2[42]	11.2(3.1),4.3(3.1)	6.2(1.0),7.1(1.9)	3.9%	11.2%	84.9%
8	---	3.0(1.0)	3.2(0.9)	100.0%	0.0%	0.0%
9		4.9(2.9),9.2(4.6)	4.4(1.4),9.4(3.9)	23.6%	66.3%	10.1%
10		3.3(1.0)	3.4(0.9)	100.0%	0.0%	0.0%
11		10.0(3.9),5.2(4.0)	5.9(1.3),7.5(2.6)	8.5%	22.2%	69.3%
12		3.0(0.9)	3.2(0.9)	100.0%	0.0%	0.0%
13		9.8(3.9),4.7(3.9)	6.0(1.2),7.1(2.7)	13.3%	19.7%	67.0%
14		3.0(0.9)	3.2(0.8)	100.0%	0.0%	0.0%
Avg[b]	5.9,4.9	8.7(3.7),5.8(3.9)	5.6(1.3),7.6(2.8)			
Avg[c]	3.1	3.1(0.9)	3.2(0.9)			

[a]+60° (g+): Distributions binned from 0° to 120°, -60°(g-): Distributions binned from -120° to –0° , ±180° (anti): Distributions binned from -180° to -120° and 120° to 180°. [b]Ser derivatives: J values averaged over the available experimental data and from the odd numbered compounds for MD simulations. [c]Thr derivatives: J values averaged over the available experimental data and from the even numbered compounds for MD simulations. The standard deviations are also averaged for the MD data. [d]For all the odd numbered systems the first value corresponds to $^3J_{H\alpha,H\beta pro\ R}$ and the second value corresponds to $^3J_{H\alpha,H\beta pro\ S}$.