

Supplementary information for:

## Helix-forming carbohydrate amino acids

*Timothy D. W. Claridge,<sup>a</sup> Daniel D. Long,<sup>a</sup> Christopher M. Baker,<sup>c</sup> Barbara Odell,<sup>a</sup> Guy H. Grant,<sup>c</sup>*

*Alison A. Edwards,<sup>a</sup> George E. Tranter,<sup>d</sup> George W. J. Fleet<sup>a\*</sup> and Martin D. Smith.<sup>a, b\*</sup>*

<sup>a</sup> Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK.

<sup>b</sup> Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

<sup>c</sup> Department of Chemistry, Physical & Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK.

<sup>d</sup> Biological Chemistry, Division of Biomedical Sciences, Imperial College, London, SW7 2AZ, UK.

mds44@cam.ac.uk, george.fleet@chem.ox.ac.uk

1. CD spectra of octamer **10** and tetramer **7** in trifluoroethanol, and discussion of the structural information derived from these data. (S1)
2. nOe derived distance constraints used in molecular dynamics simulations. (S3)
3. nOe distance violations for the structure that most satisfies the distance constraints. (S4)
4. Results of molecular dynamics simulations performed without hydrogen bond constraints (S5)
5. Bifurcation study: NH(i) and C=O (i-2) distances over the course of the simulation. (S8)
6. <sup>1</sup>H and <sup>13</sup>C NMR spectra of compounds. (S9)

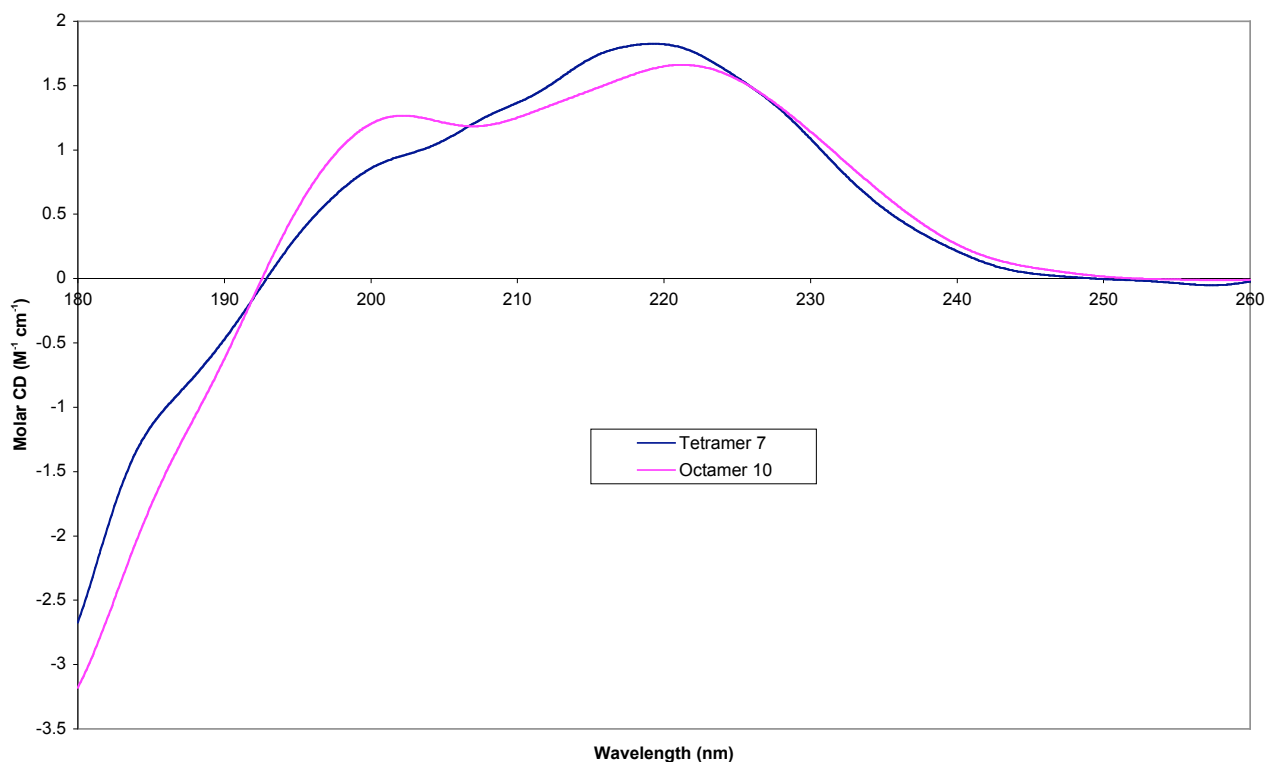
### 1. Circular Dichroism.

Circular dichroism (CD) spectroscopy is a reliable technique for the investigation of amide bond environments of proteins, as typical spectra of common secondary structural elements such as  $\alpha$ -helices,  $\beta$ -turns, and irregular structures are well established. However, the study of CD spectra of carbohydrate amino acid oligomers is in its infancy when compared to the plethora of protein studies. Nonetheless general conclusions may be drawn based on the underlying principles of

chiroptical phenomena. CD spectra of carbohydrate amino acid derivatives may be interpreted in terms of the relative populations of conformers they adopt and hence relationships between conformations of oligomers of different lengths can be investigated. As such, CD can be used to examine the amide environments of related oligomers in order to establish differences in secondary structural preference with chain length.

Unfortunately, it was not possible to record meaningful CD spectra in the solvents utilized for NMR and IR study due to the significant absorption of  $\text{CHCl}_3$  and  $\text{CDCl}_3$  in the wavelength region where amide spectra are observed (260-185 nm). Trifluoro-2,2,2-ethanol (TFE) was therefore employed as the solvent for CD investigation. The CD technique is complementary to the NMR and IR studies reported above, as CD is able to distinguish between regular (both hydrogen bonded and non-hydrogen bonded) and irregular structures. This is highly relevant as the data obtained by NMR and IR for tetramer **7** could equally infer that the conformation adopted was irregular or a non-hydrogen bonded conformation.

From the CD spectra, there is a significant difference in the population of conformers between the tetramer **7** and octamer **10** in TFE (supplementary figure 1). Primarily, the octamer **10** exhibits spectral features at 221 and 201 nm. However, for tetramer **7** the band at 201nm is less pronounced. The octamer **10** exhibits a spectrum comparable to that exhibited by a protein  $\alpha$ -helix formed from D-amino acid residues and is consistent with the hydrogen bonded secondary structure observed by NMR and IR studies in chloroform. The spectrum of the octamer **10** has a negative peak with two separate maxima of similar intensity (for **10**, the maxima occur at 221 nm and 201 nm; the maxima of an  $\alpha$ -helix occur at 222 nm ( $n \rightarrow \pi^*$  transition) and at 208 nm ( $\pi \rightarrow \pi^*$  transition)).



Supplementary figure 1. Circular dichroism spectra of the octamer **10** and tetramer **7** recorded in trifluoroethanol at 293 K, shown with using amide bond concentration (compound concentration 126  $\mu$ M).

Notably, the CD spectrum of the tetramer **7** is not consistent with an irregular conformation but rather with a rigid conformation composed of repeating conformational motifs. This is significant as NMR and solution IR data (in chloroform rather than TFE) indicate that the most populated conformer of the tetramer is not stabilized by hydrogen bonds. If the helical structural preference observed for the octamer was inherent in the THF building blocks then it is not unreasonable to suggest that tetramer may adopt an extended helical conformation (*i.e.* a similar conformation but simply lacking the restraining hydrogen bonds). Nonetheless, the limited possibility of hydrogen-bonding in the tetramer, compared to the octamer, allows effective competition of other conformational motifs.

In summary, the combined spectroscopic evidence suggests that a change in the population of conformers occurs with increased chain length and that the tetramer species **7** may not exhibit an irregular conformation despite indications by NMR and IR.

2. nOe derived distance constraints used in molecular dynamic simulations.

NOEs have been calibrated using H3-H4 intra-residue distances (mean 0.24 nm) determined from energy minimized structures for semi-quantitative analysis to give: Strong: < 0.27 nm, Medium: < 0.35 nm, Weak: < 0.50 nm.

Sequential nOes: NH <sup>(i+1)</sup> to H2 <sup>i</sup>		Strong	Medium	Weak
NH <sup>B</sup>	H2 <sup>A</sup>		✓	
NH <sup>C</sup>	H2 <sup>B</sup>		✓	
NH <sup>D</sup>	H2 <sup>C</sup>		✓	
NH <sup>E</sup>	H2 <sup>D</sup>		✓	
NH <sup>F</sup>	H2 <sup>E</sup>		✓	
NH <sup>G</sup>	H2 <sup>F</sup>		✓	
NH <sup>H</sup>	H2 <sup>G</sup>		✓	

nOes from NH to H3		Strong	Medium	Weak
NH <sup>E</sup>	H3 <sup>B</sup>		✓	
NH <sup>F</sup>	H3 <sup>C</sup>		✓	
NH <sup>G</sup>	H3 <sup>D</sup>		✓	
NH <sup>D</sup>	H3 <sup>A</sup>		✓	
NH <sup>H</sup>	H3 <sup>E</sup>		✓	

nOes from NH to H5		Strong	Medium	Weak
NH <sup>H</sup>	H5 <sup>G</sup>		✓	
NH <sup>B</sup>	H5 <sup>A</sup>		✓	
NH <sup>C</sup>	H5 <sup>B</sup>		✓	
NH <sup>D</sup>	H5 <sup>B</sup>			✓
NH <sup>E</sup>	H5 <sup>C</sup>		✓	
NH <sup>H</sup>	H5 <sup>F</sup>		✓	

THF ring nOes from H2 to H4		Strong	Medium	Weak
H2 <sup>G</sup>	H4 <sup>E</sup>	✓		
H2 <sup>F</sup>	H4 <sup>D</sup>		✓	
H2 <sup>E</sup>	H4 <sup>C</sup>		✓	
H2 <sup>D</sup>	H4 <sup>B</sup>		✓	

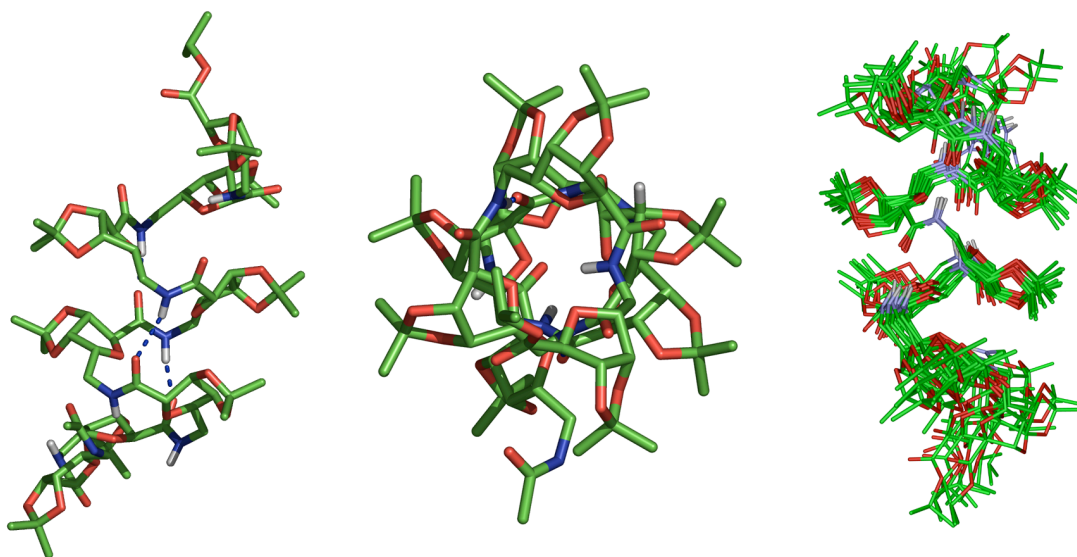
THF ring nOes from H2 to H3		Strong	Medium	Weak
H2 <sup>H</sup>	H3 <sup>F</sup>		✓	
H2 <sup>G</sup>	H3 <sup>E</sup>	✓		
H2 <sup>F</sup>	H3 <sup>D</sup>	✓		
H2 <sup>E</sup>	H3 <sup>C</sup>	✓		
H2 <sup>D</sup>	H3 <sup>B</sup>	✓		

3. Violations of the nOe derived distance constraints for 10 for the structure that most satisfies the distance constraints.

Atom 1	Atom 2	nOe constraint (Å)	Violation(Å)
H2 <sup>H</sup>	H3 <sup>F</sup>	3.5	0.293
H2 <sup>D</sup>	H4 <sup>B</sup>	3.5	0.229
H2 <sup>F</sup>	H4 <sup>D</sup>	3.5	0.179
H2 <sup>G</sup>	H4 <sup>E</sup>	2.7	0.439
NH <sup>B</sup>	H5 <sup>A</sup>	3.5	1.257
NH <sup>H</sup>	H5 <sup>G</sup>	3.5	0.825
NH <sup>D</sup>	H3 <sup>A</sup>	3.5	1.780

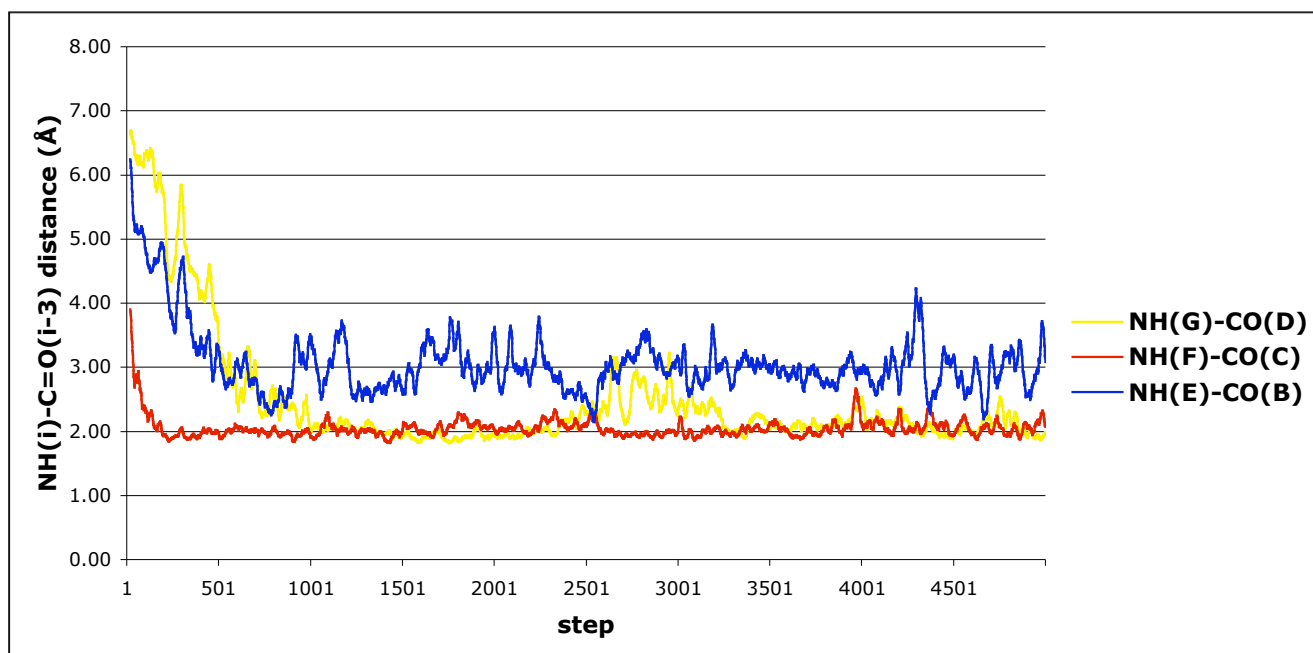
Total violations:	5.002
-------------------	-------

4. Results of molecular dynamic simulations performed without hydrogen bond constraints. Structures produced by dynamics simulations without additional hydrogen bond constraints are pictured below:



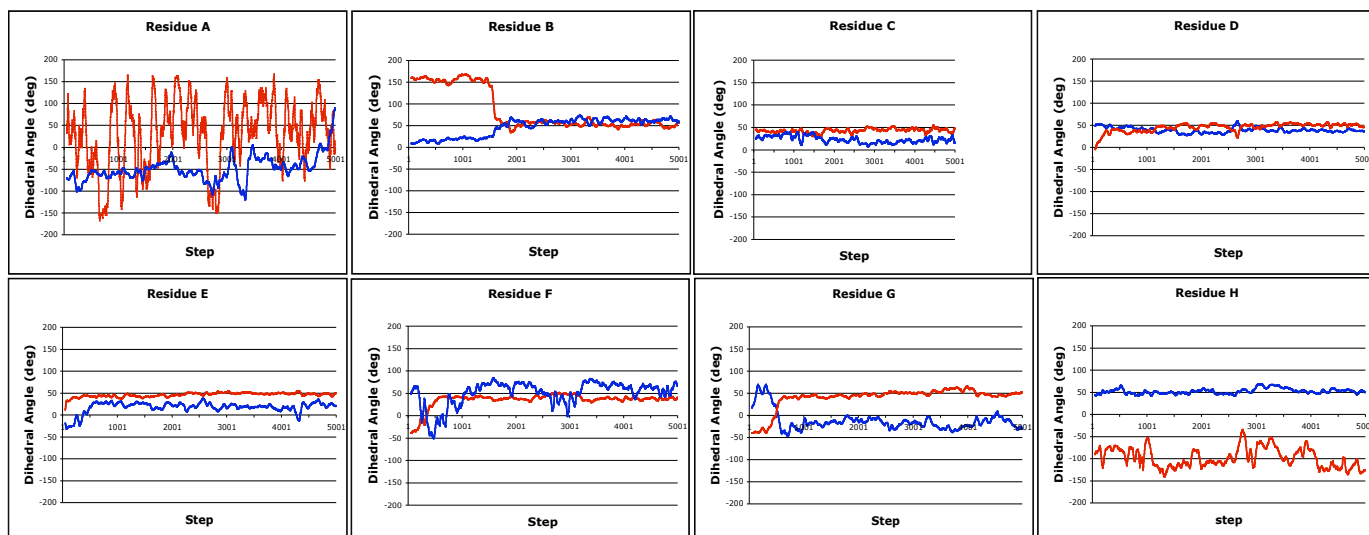
Supplementary figure 2. The final minimized structure produced by a 500 ps nOe constrained simulation on the octamer **10** without hydrogen bonds constraints **B** view looking from *N* to *C* terminus (Hydrogen bonds are indicated by dotted lines) **C**. Superposition of 10 conformations selected every 400 steps throughout the simulation (first structure selected at 1400 steps).

The observed hydrogen bonds form between residues  $\text{NH}^{\text{E}}\text{-C=O}^{\text{B}}$ ,  $\text{NH}^{\text{F}}\text{-C=O}^{\text{C}}$  and  $\text{NH}^{\text{G}}\text{-C=O}^{\text{D}}$ . The absence of any hydrogen bonding involving the two terminal residues may be attributable to the inability of the simulation to model accurately the ends of the molecule or may alternatively indicate that we are seeing true ‘end effects’ and that some ‘critical length’ is necessary before we observe formation of well-defined secondary structure. Monitoring the formation of the three inter-residue hydrogen bonds reveals that all three hydrogen bonds are formed quickly and remain locked in the hydrogen bonding geometry, indicating a fixed conformation for the whole molecule.



Supplementary figure 3. Measurement of inter-residue hydrogen bonding distances over 500 ps simulation with only nOe distance constraints in chloroform.

Analysis of the nOe derived distance constraints reveals that by the end of the simulation, out of the twenty seven restraints imposed, there are seven restraints that are not satisfied, resulting in a total violation of 7.168 Å. The majority of these violations occur in cases where the atomic separation fluctuates around a mean value close to the nOe derived distance, deviating above it only by small amounts. Two of the violations, those corresponding to the  $\text{NH}^{\text{D}}\text{-H3}^{\text{A}}$  and  $\text{H2}^{\text{H}}\text{-H3}^{\text{F}}$  distances, are however, large and consistent. These two major violations account for 57% of the total violations and both occur in situations where we are considering distances involving the terminal residues. As we have previously seen, the terminal residues are not well modelled using these simulations and we can draw the conclusion that, with the exception of the two terminal residues, this simulation gives an accurate representation of the experimentally derived nOe data.

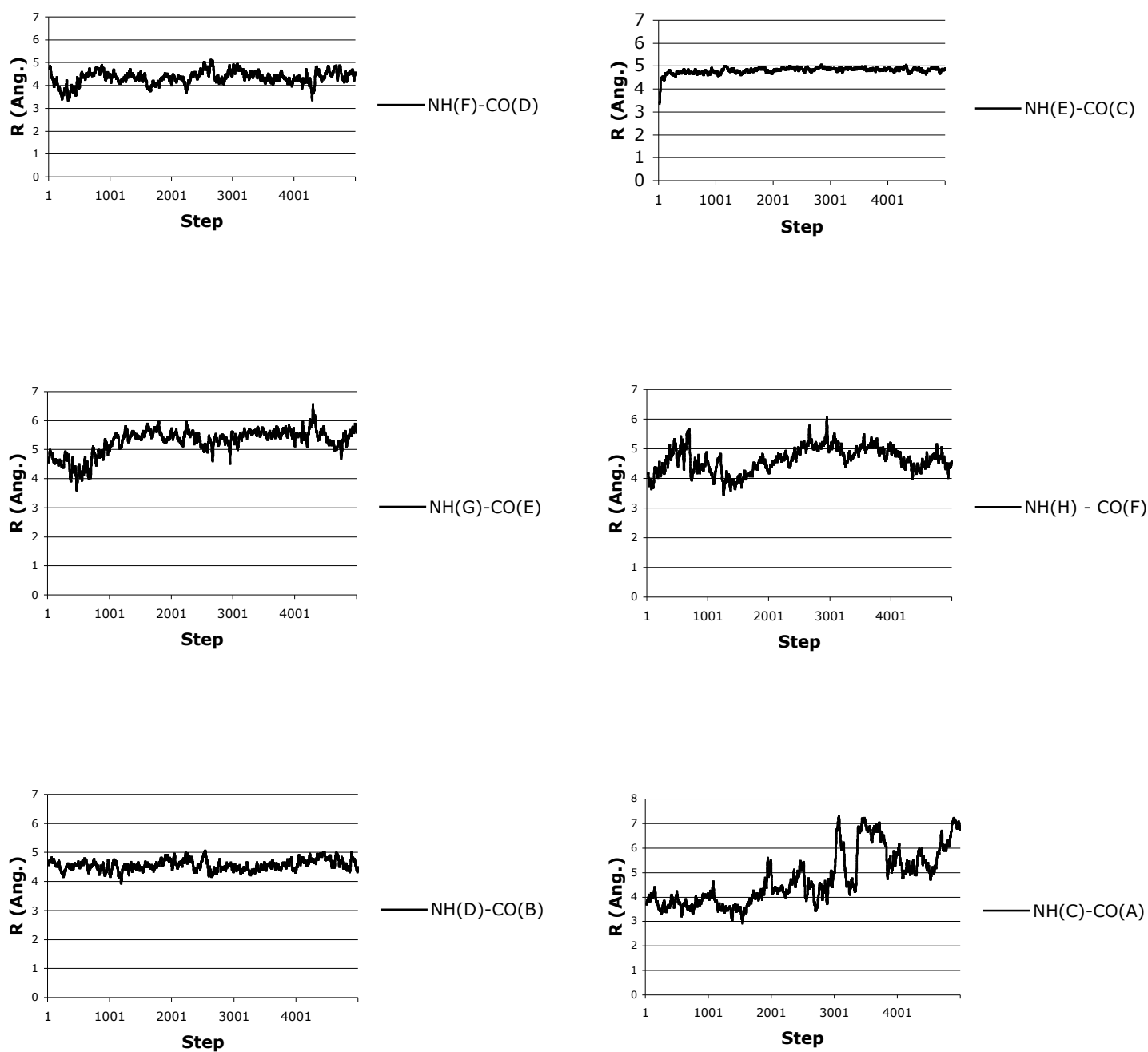


Supplementary figure 4. Dihedral angles  $\Psi_1$  and  $\Psi_2$  throughout the course of the simulation for each residue A-H of the octamer **10**. ( $\Psi_1$ = —,  $\Psi_2$ = —)



## 5. Data for bifurcation study

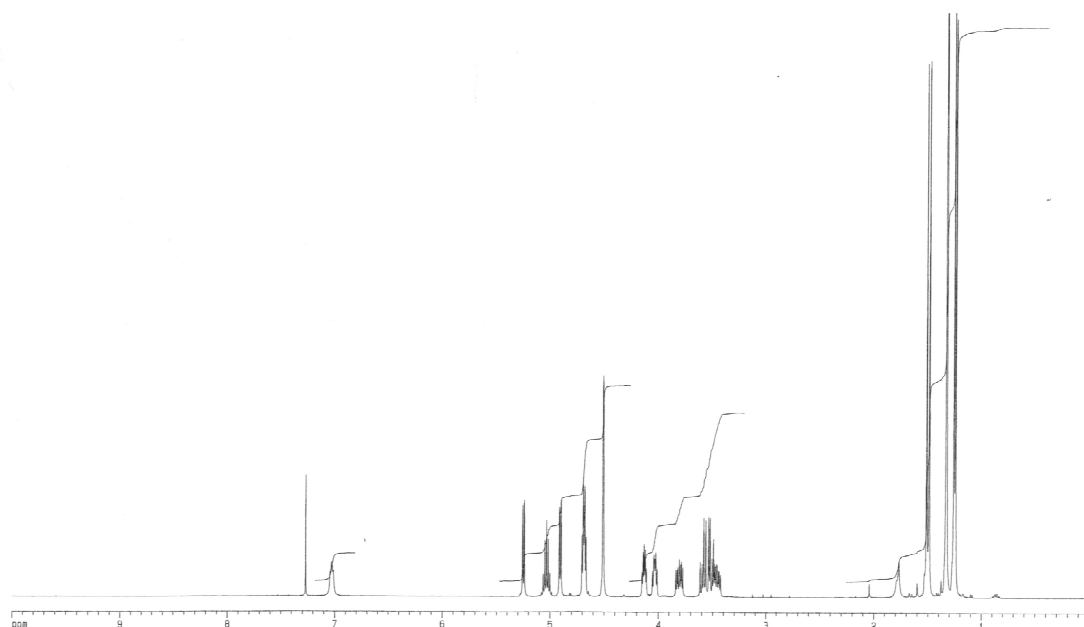
In order to assess whether bifurcation occurs during the simulation, distances between the NH of residue  $i$  and the C=O of residue  $(i-2)$  were monitored. This demonstrated that at all times these distances are well above the range we would expect if hydrogen bonds were being formed, which suggests that no bifurcation is occurring.



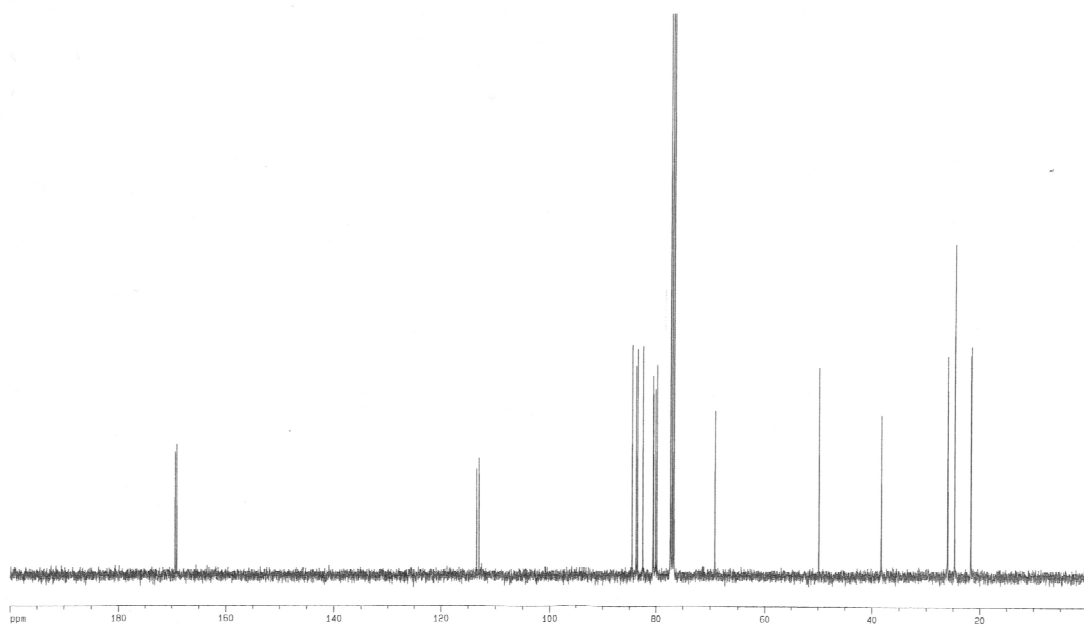
Supplementary figure 5: NH( $i$ ) and C=O ( $i-2$ ) distances over the course of the simulation

6.  $^1\text{H}$  and  $^{13}\text{C}$  spectra of compounds:

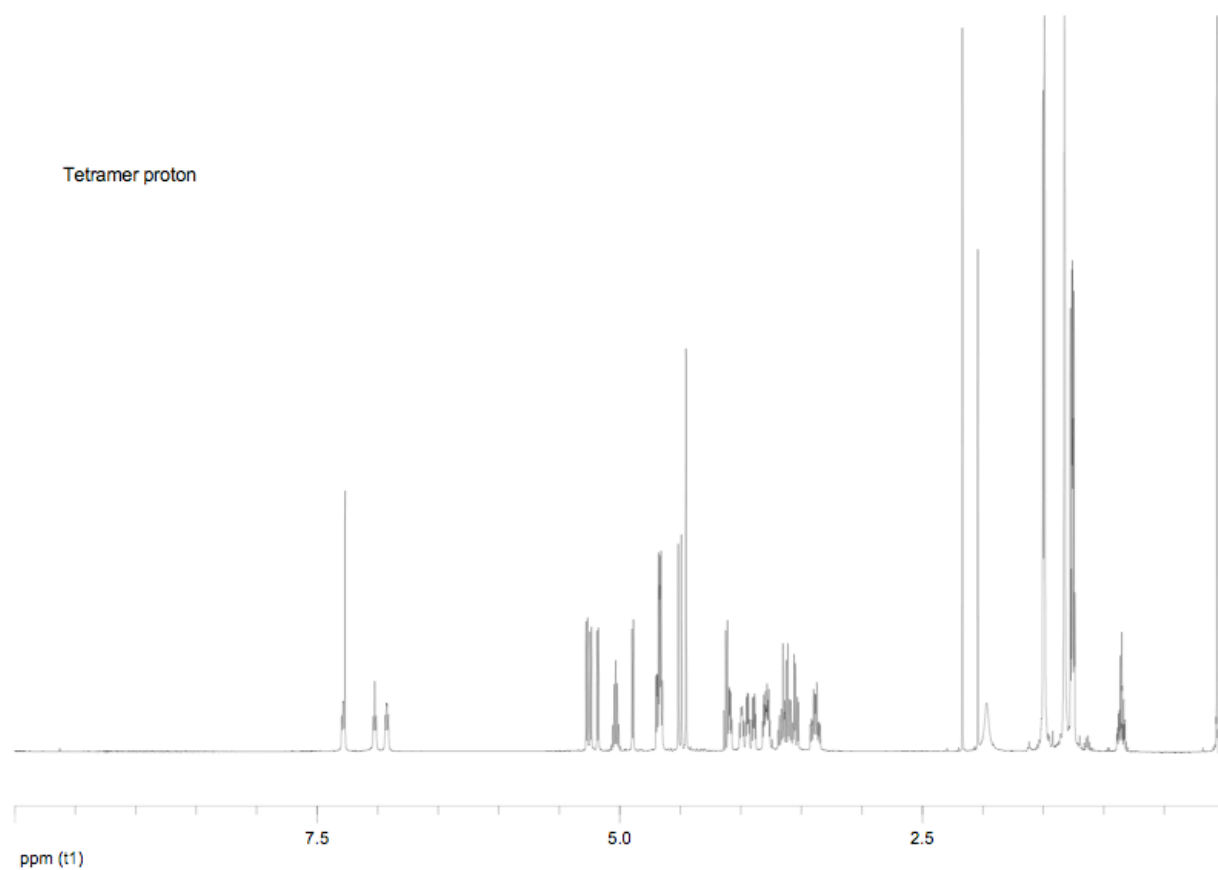
(i) Dimer  $^1\text{H}$ :



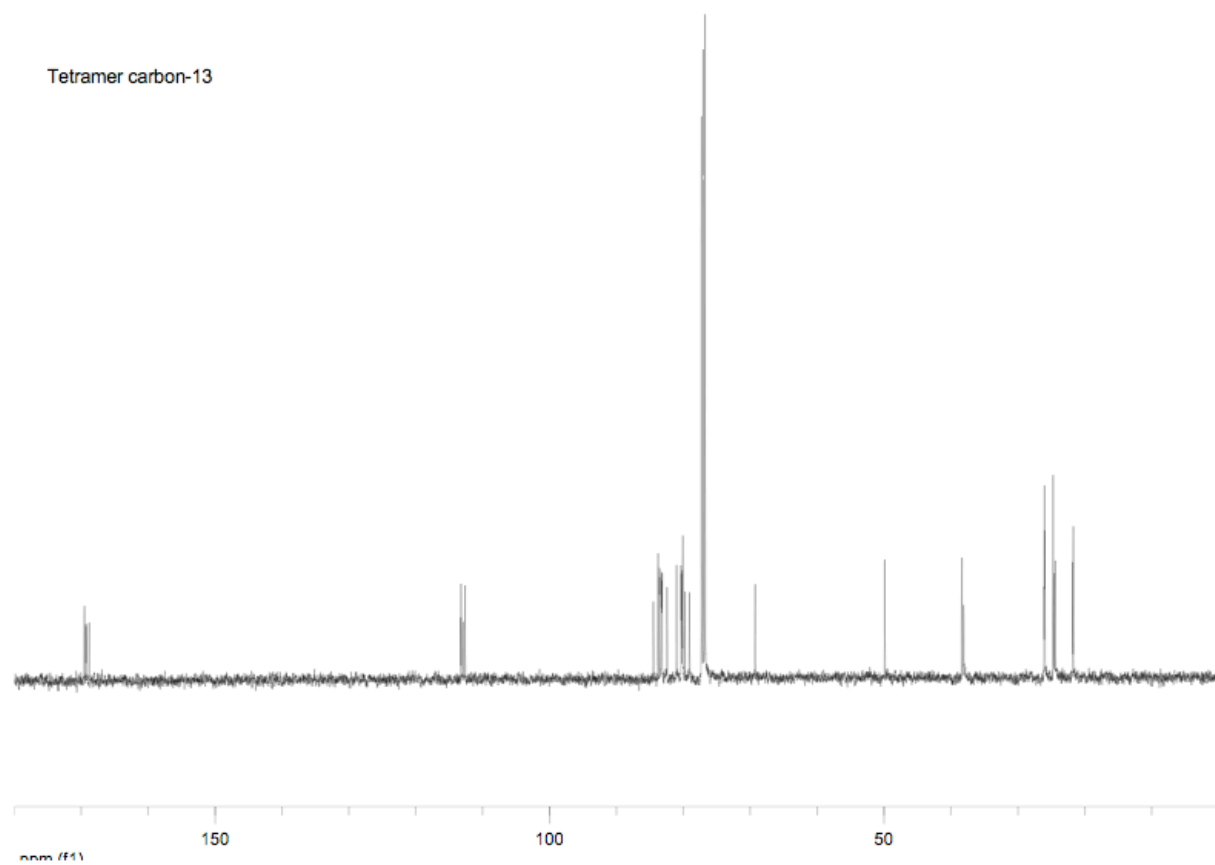
Dimer  $^{13}\text{C}$



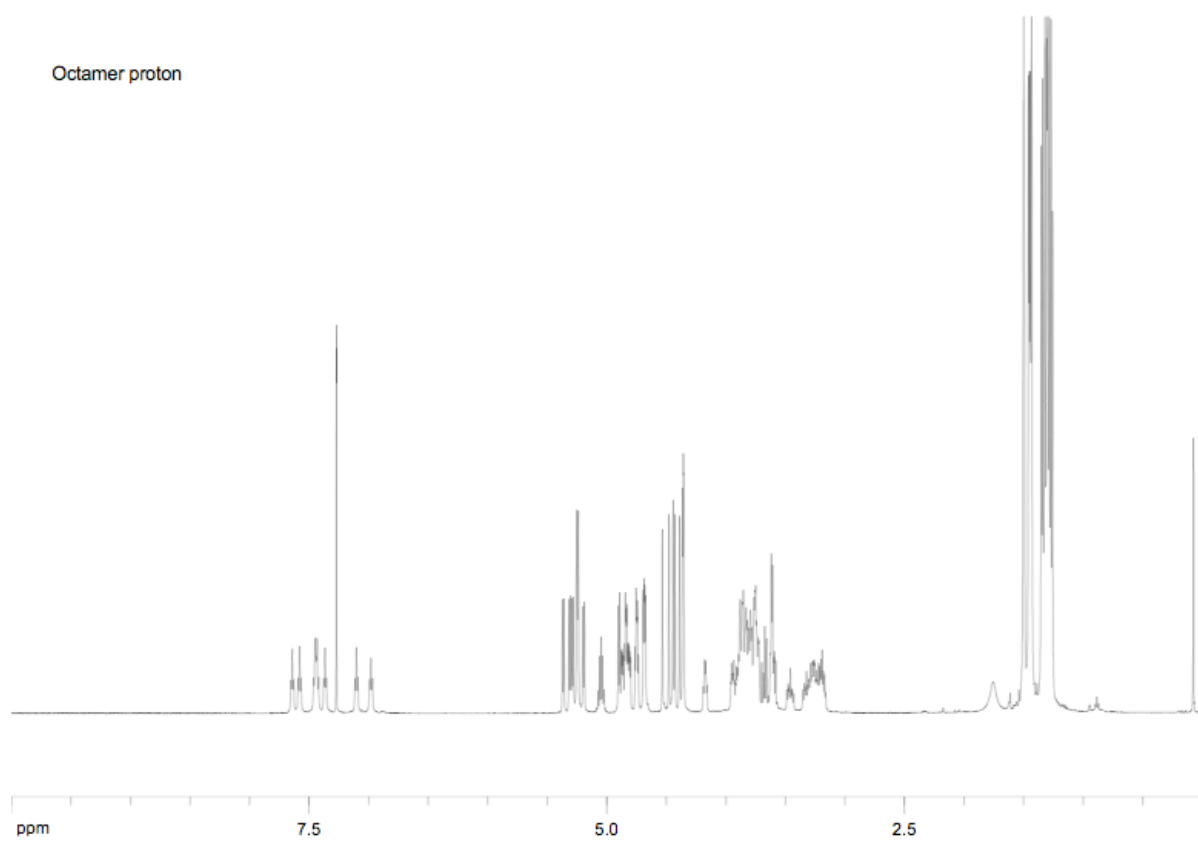
# Tetramer $^1\text{H}$



## tetramer $^{13}\text{C}$ :



Octamer  $^1\text{H}$ :



Octamer  $^{13}\text{C}$

