Supplemental Data

SOLUTION STRUCTURE OF AN ABC COLLAGEN HETEROTRIMER REVEALS A SINGLE-REGISTER HELIX STABILIZED BY ELECTROSTATIC INTERACTIONS

Jorge A. Fallas, Varun Gauba, Jeffrey D. Hartgerink

Departments of Chemistry and Bioengineering
Rice University, 6100 Main Street. Houston, TX 77005

Peptide Synthesis and Purification of the labeled peptides. The **D***, **O*** and **K*** peptides were synthesized with an Advanced Chemtech Apex 396 solid phase peptide synthesizer using standard Fmoc chemistry and a Rink MBH amide resin on a 0.05 mM scale. The unlabeled amino acids were added in a 4:1 molar ratio to the peptide chain with HBTU/HoBT as activating agents and a coupling time of 45 minutes. The uniformly labeled amino acids were purchased form Cambridge Isotope Laboratories and added in a 1.5:1 molar ratio to the peptide chain. For these steps HATU was used as the activating agent and the coupling reaction was carried out for 4 hours. The peptides were protected at the N-terminus using acetic anhydride and cleaved from the resin with a 38:1:1 mixture of TFA, triisopropylsilane and water yielding an amide C-terminus. Purification was done on a Varian PrepStar220 HPLC using a preparative reverse phase C-18 column with a linear gradient of water and acetonitrile gradient each containing 0.5 % TFA. HPLC fractions were analyzed by MALD/TOF mass spectrometry on a Bruker autoflex II using a prespotted anchor chip with α-cyano-4-hydroxycinnamic acid as the matrix. The data is shown in figures S1.

All experiments were performed on an 800 MHz Varian spectrometer equipped with a cryogenic probe.

NMR characterization of the D and K peptides. 2.4 mg of **D** and 2.4 mg of **K** were each dissolved separately in 560 μL of H_2O and mixed with 70 μL 100 mM phosphate buffer and 70 μL D_2O to afford a 1.2 mM sample. TSP was used as an internal proton standard. TOCSY spectra with a 75 ms spinlock were acquired for each sample on an 800 MHz Varian spectrometer equipped with a cryogenic probe at 25 °C. A total of 1918 complex points were recorded in 16 scans for the directly acquired dimension and 320 increments were recorded in

the sates mode for the indirect dimension. A spectral width of 9600 Hz was used in both dimensions. The data was processed by zero filling to the next power of two and cosine bell apodization functions were applied in both dimensions.

NMR characterization of the K/D/O mixture. 2.3 mg of **O** and 2.4 mg of **K** were dissolved together in 200 μL H_2O and 2.4 mg of **D** was dissolved in 70 μL 100 mM phosphate buffer and 130 μL H_2O affording a 4.2 mM aqueous solution for each peptide. After mixing them, heating to 85 °C for 15 minutes and incubating at room temperature for at least 18 hours, 70 μL D_2O, 30 μL H_2O and a proton standard (TSP) were added. The final concentration for each strand was 1.2 mM and the total peptide concentration in the sample was 3.6 mM. TOCSY spectra with a 75 ms and a 10 ms spinlock duration were recorded at 15 °C and 25 °C**.** A total of 1918 complex points were recorded in 16 scans for the directly acquired dimension for all spectra while 560 increments were used in the indirect dimension for the 25 °C data and 320 increments were used for the 15 °C data. NOESY spectra with a 75 ms mixing time were recorded at 15 °C and 25 °C. A total of 3271 complex points were recorded in 32 scans for the directly acquired dimension and 480 increments in the indirect dimension. A square spectral window of 9600 Hz was used for all spectra. Figure S5 shows the amide region of the 25 °C spectra.

NMR characterization of the K*/D*/O* mixture. A sample was prepared using a similar methodology as described above. $^{1}H,^{13}C$- and $^{1}H,^{15}N$-HSQC, HNHA, HNHB and a 2D ^{13}C, ^{15}N-edited NOESY experiments were performed on an 800 MHz Varian spectrometer at 25 °C. A total of 630 complex points in 16 scans for the direct dimension and 400 increments in the indirect dimension were acquired for the $^{1}H,^{15}N$-HSQC using a spectral window of 8000 Hz in the hydrogen dimension and 4000 Hz in the nitrogen dimension. A total of 961 complex points in 24 scans for the direct dimension and 450 increments in the indirect dimension were acquired for the $^{1}H,^{13}C$- HSQC using a spectral window of 9600 Hz in the hydrogen dimension and 18100 Hz in the carbon dimension. The data was processed by zero filling to the next power of two and cosine bell apodization functions were applied in both dimensions. Forward backwards linear prediction was used to improve the resolution in the heteroatom dimension of both spectra. For the 3D HNHA spectrum a total of 682 complex points in 16 scans for the direct dimension, 120 increments for the first indirect dimension and 18 increments for the second indirect dimensions were acquired using a spectral window of 8000 Hz for direct dimension, 48000 for the hydrogen

indirect dimension and 2432 Hz for the nitrogen indirect dimension. The data was processed by zero filling to the next power of two and cosine bell apodization functions were applied in all dimensions. Forward backwards linear prediction was used to improve the resolution in the indirect hydrogen dimension. For the 3D HNHB 4spectrum a total of 1024 complex points in 4 scans for the direct dimension. 100 increments for the first indirect dimension and 30 increments for the second indirect dimensions were acquired using a spectral window of 12000 Hz for direct dimension, 8000 for the hydrogen indirect dimension and 3243 Hz for the nitrogen indirect dimension. The data was processed by zero filling to the next power of two and cosine bell apodization functions were applied in all dimensions. A total of 1024 complex points in 96 scans for the direct dimension and 200 increments in the indirect dimension were acquired for the 2D ^{13}C, ^{15}N-edited NOESY using a spectral window of 1150 Hz and 8000 Hz.

NMR characterization of the K/D/O* mixture. A sample was prepared using a similar methodology as described above but using a 1.5:1:1 ratio of the peptides (**K:D**:**O***). In this case the D and O concentration was 1.2 mM and the K concentration was 1.8 mM making the total peptide concentration 4.2 mM. A ^{15}N-HSQC experiment was performed on an 800 MHz Varian spectrometer at 25 °C. A total of 1024 complex points in 8 scans for the direct dimension and 180 increments in the indirect dimension were acquired using for a spectral window of 13000 Hz in the hydrogen dimension and 2432 Hz in the nitrogen dimension. The data was processed by zero filling to the next power of two and cosine bell apodization functions were applied in both dimensions. Linear prediction was used to improve the resolution in the indirect dimension of both spectra.

Details about NMR experiments 2D ^{13}C, ^{15}N-edited NOESY. It corresponds to a 2D version of a 4D ^{1}H,^{13}C-HMQC-NOESY-^{1}H,^{15}N-HSQC experiment was recorded at 25 °C(1); to ease further discussion of this experiment it will be referred to as a 2D ^{13}C, ^{15}N-edited NOESY. The initial ^{1}H,^{13}C-heteronuclear multiple quantum coherence (HMQC) pulse train selectively creates magnetization on protons attached to labeled carbon atoms, at the end of which a mixing time is allowed for NOE buildup; the transferred magnetization is then read through an ^{1}H,^{15}N-HSQC

pulse sequence. Our experiment was carried out keeping both heteroatom evolution times (t1 and t3) constant, yielding a 2D NOESY spectrum where cross peaks between hydrogens that are directly to bonded to ^{13}C nuclei and hydrogens that are directly bonded to ^{15}N nuclei are observed. 3D HNHA. For this experiment, the ratio between the intensity of the cross and diagonal peaks can be used to compute the $^{3}J_{HNH\alpha}$ value. A systematic error is incurred in the measurement because the anti-phase magnetization that gives rise to the cross peaks relaxes at a faster rate than the in-phase component which gives rise to the diagonal peak, attenuating the ratio and underestimating the coupling constant. A correction factor of 1.16 has been previously determined for a collagen mimetic peptide to account for this phenomenon(2). Because this factor is a function of the rotational correlation time, τ_c(3), we have applied the same correction to the values reported here. The experimental error in the measurement comes from the integration of the peaks, which was done by fitting the peak profile to a Gaussian shape using a least squares procedure. The error in the integration was estimated by the residual of the fit. The residual was calculated by the sum of the squares of the difference of the value of the peak and the value of the best fit at each point. Using this approach the minimum and maximum possible intensity ratios were computed and used to calculate the error bars on the ^{3}J values. The errors where then propagated when solving the Karplus equation as parameterized in (3) to obtain the φ backbone dihedral angles.

Details about conformational restraints Distance restraints. Intensity categories defined by peak volume: very strong (2.0 - 2.5 Å), strong (2.2 -2.8 Å), medium (2.5–3.5 Å) and weak (2.8-5.0 Å). Dihedral restraints. From the HNHA experiments see table 2 in the text; from the HNHB experiment the χ_1 angle of K (180° ± 40°) and D (-60° ± 40° for D); from proline side ring pucking the the χ_1 and φ values of P (χ_1 = 19° ± 30°, φ = -72° ± 30°) and O (χ_1 = -6° ± 30°, φ = -58° ± 30°).

1. Muhandiram, D. R., Guang, Y. X. and Kay, L. E. (1993) *J. Biomol. NMR* **463-470**, 463-470
2. Li, Y., Brodsky, B. and Baum, J. (2008) *J. Biol. Chem.* **282**, 22699-22706
3. Vuister, G. W. and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 7772-7777

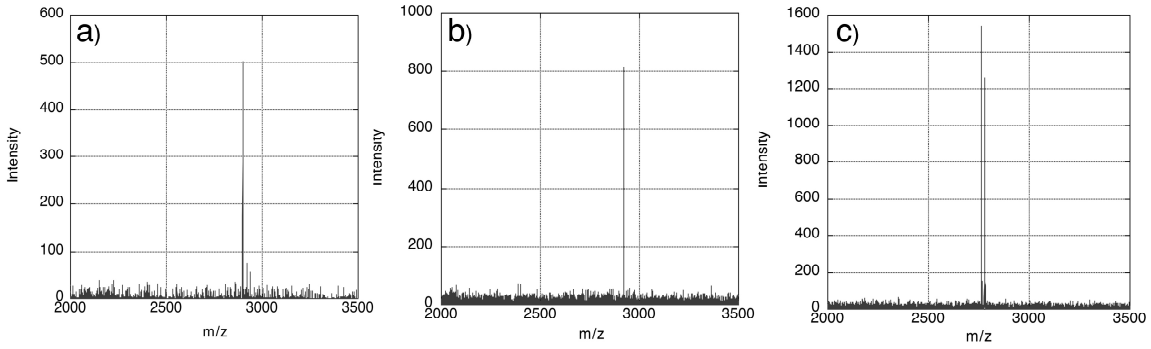

Fig. S1. MALDI-TOF MS data for the peptides synthesized for this work. a) **K*** Expected: 2897.7[H+], Observed: 2898.1[H+]. b) **O*** Expected: 2762.2[Na+], Observed: 2762.4[Na+]. c) **D*** Expected: 2919.6[H+], Observed: 2920.1[H+].

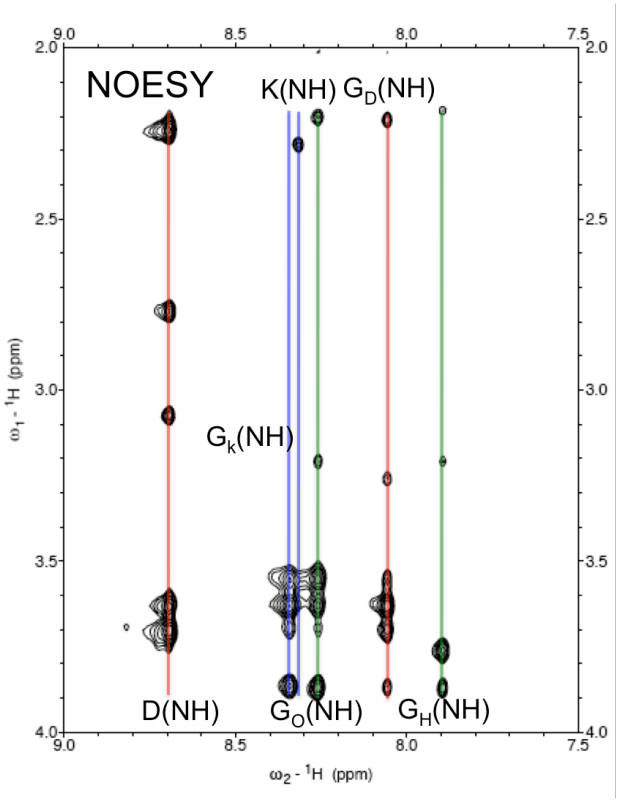

Fig. S2**.** Amide region from the 2D NOESY spectrum of the mixture **K/D/O**. The vertical lines denote the chemical shifts of the different amide resonances using the nomenclature described in the text.

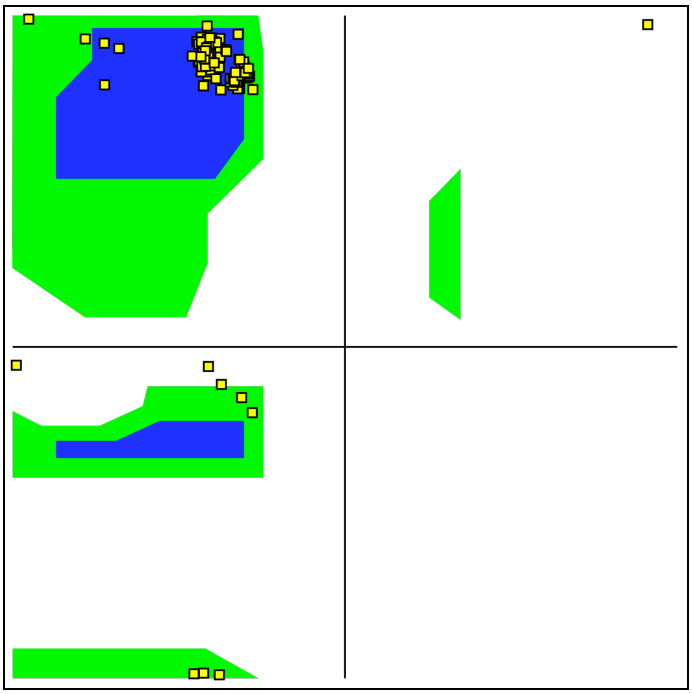

Fig. S3. Representative ramachandran plot of the conformers from our NMR ensemble.

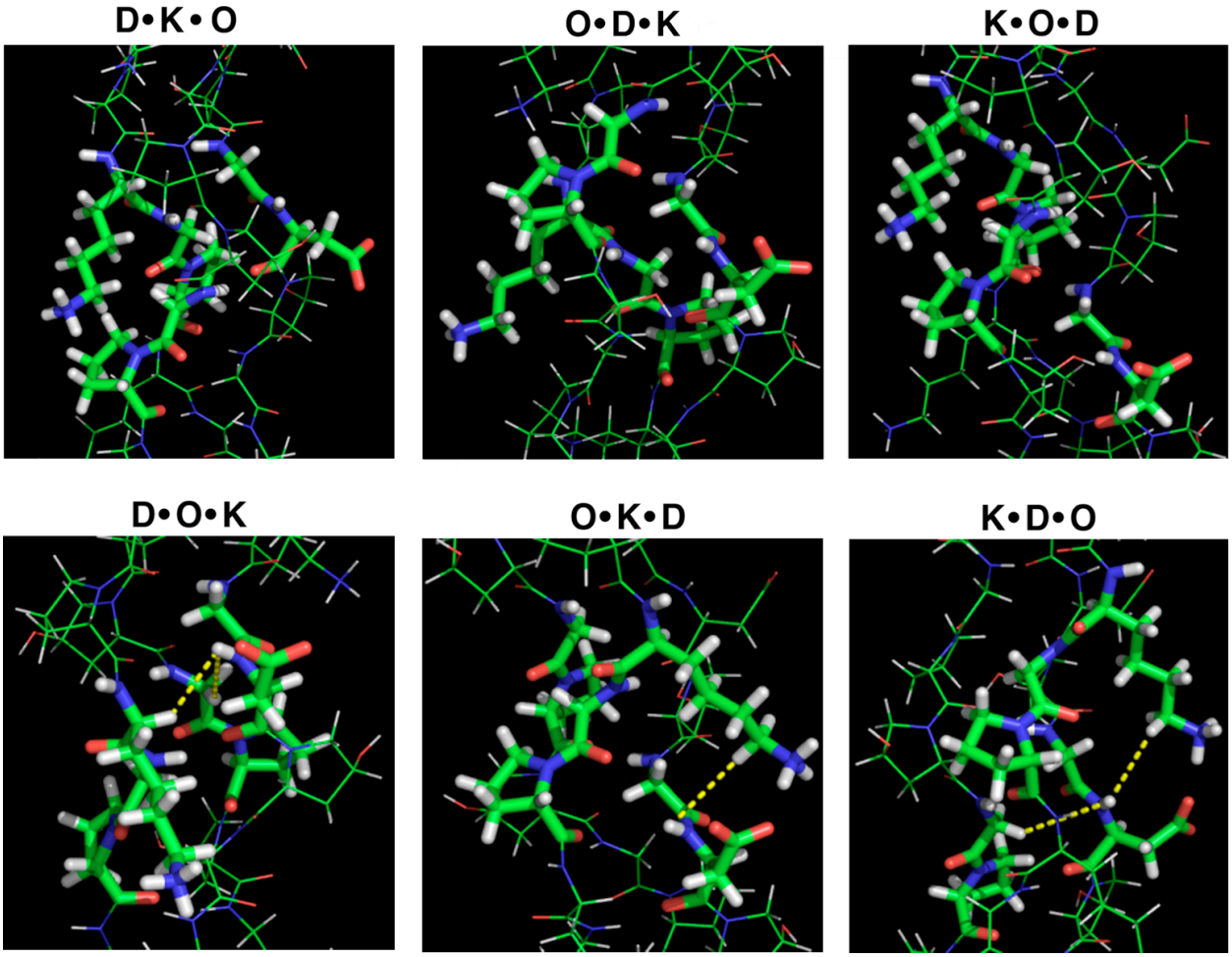

Fig. S4. Homology models for all registers of the heterotrimer. The labeled amino acids (see Table I for sequence) are highlighted and the NOEs discussed in Figure 4B, column E are represented by dashed lines.

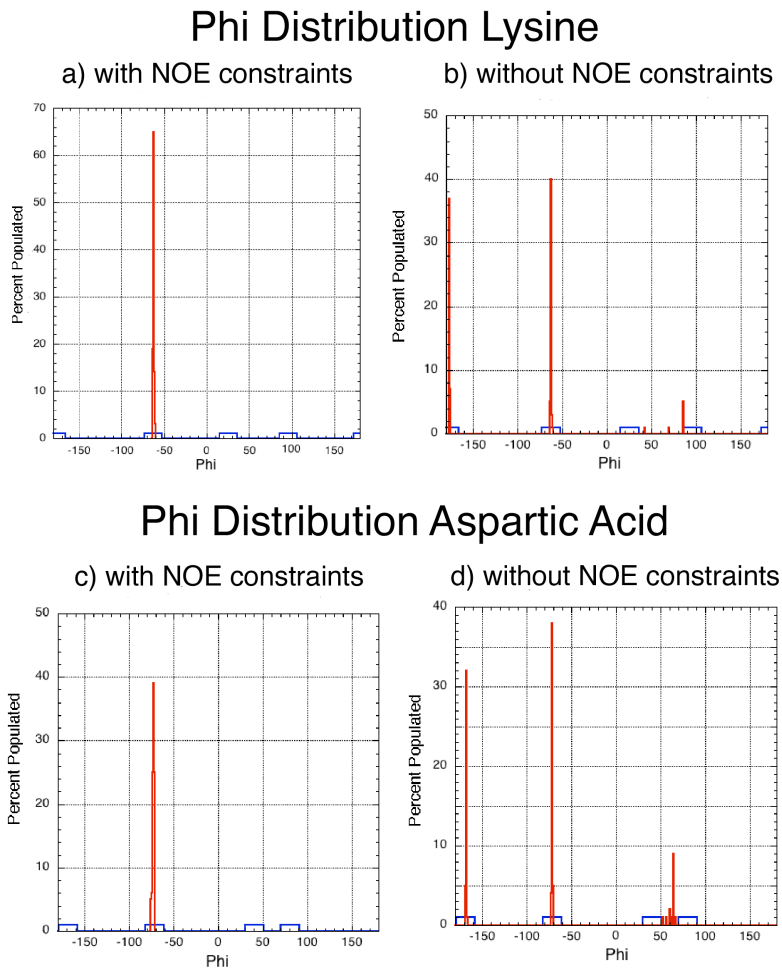

Fig. S5. Distribution of φ dihedrals for K and D amino acids after a preliminary round of simulated annealing with $^{3}J_{HNHa}$ coupling constraints. a) and b) Distribution for K with and without NOE constraints. c) and d) Distribution for D with and without NOE constraints. Highlighted in blue are the four solutions for to the Karplus equation (including experimental uncertainty) and in red the percent of angles found after measuring the 10 lowest energy structures of the run. In both cases supplementing the coupling constraints with NOE data results in the structures populating exclusively one of the possible Karplus solutions.

TABLE S1

Expected interchain NOEs based on a POG homotrimer model and observed interchain NOEs for our heterotrimer.

Expected Homotrimer NOE	Observed Heterotrimer NOE
G(NH)-P($C_\delta H_1$)	$G_{\mathbf{D}}$(NH)-$P_{\mathbf{O}}(C_\delta H_1)$#, $G_{\mathbf{O}}$(NH)-$P_{\mathbf{K}}(C_\delta H_1)$#
O($C_\delta H_2$)-P($C_\delta H_1$)	$O_{\mathbf{O}}(C_\delta H_1)$-$P_{\mathbf{K}}(C_\delta H_1)$, $O_{\mathbf{D}}(C_\delta H_1)$-$P_{\mathbf{O}}(C_\delta H_1)$
O($C_\alpha H$)-P($C_\delta H_1$)	$O_{\mathbf{O}}(C_\alpha H_1)$-$P_{\mathbf{K}}(C_\delta H_1)$, $O_{\mathbf{D}}(C_\alpha H_1)$-$P_{\mathbf{O}}(C_\delta H_1)$
O($C_\alpha H$)-P($C_\delta H_2$)	$O_{\mathbf{O}}(C_\alpha H_1)$-$P_{\mathbf{K}}(C_\delta H_2)$, $O_{\mathbf{D}}(C_\alpha H_1)$- $P_{\mathbf{O}}(C_\delta H_2)$
G($C_\alpha H_1$)- P($C_\alpha H$)	$G_{\mathbf{O}}$ ($C_\alpha H_1$)- $P_{\mathbf{K}}(C_\alpha H)$
P($C_\beta H_1$)-O($C_\beta H_2$)	$P_{\mathbf{K}}(C_\beta H_1)$-$O_{\mathbf{O}}(C_\beta H_2)$*,$P_{\mathbf{O}}(C_\beta H_1)$-$O_{\mathbf{D}}(C_\beta H_2)$*
P($C_\gamma H_1$)-O($C_\alpha H$)	$P_{\mathbf{K}}(C_\gamma H)$-$O_{\mathbf{O}}(C_\alpha H)$
P($C_\gamma H_1$)-O($C_\beta H_1$)	$P_{\mathbf{O}}(C_\gamma H)$-$O_{\mathbf{D}}(C_\beta H_1)^+$
P($C_\gamma H_1$)-O($C_\beta H_2$)	$P_{\mathbf{O}}(C_\gamma H)$-$O_{\mathbf{D}}(C_\beta H_2)^+$
P($C_\gamma H_2$)-O($C_\beta H_2$)	$P_{\mathbf{O}}(C_\gamma H)$-$O_{\mathbf{D}}(C_\gamma H)^+$
P($C_\gamma H_1$)-O($C_\gamma H$)	$P_{\mathbf{O}}(C_\gamma H)$-$O_{\mathbf{D}}(C_\gamma H)^+$
P($C_\delta H_1$)-O($C_\beta H_2$)	$P_{\mathbf{K}}(C_\delta H_1)$-$O_{\mathbf{O}}(C_\beta H_2)$, $P_{\mathbf{O}}(C_\delta H_1)$-$O_{\mathbf{D}}(C_\beta H_2)$
P($C_\delta H_2$)-O($C_\beta H_2$)	$P_{\mathbf{K}}(C_\delta H_2)$-$O_{\mathbf{O}}(C_\beta H_2)$, $P_{\mathbf{O}}(C_\delta H_2)$-$O_{\mathbf{D}}(C_\beta H_2)$

Highlighted in Fig. 2; *Overlapping peaks; $^+P_{\mathbf{O}}(C_\gamma H)$ overlaps with $O_{\mathbf{O}}(C_\beta H_1)$ and $O_{\mathbf{D}}(C_\beta H_1)$