

Electronic Supplementary Information for

**Identifying the Protonation Site and the Scope of Non-Proline
Cis-Peptide Bond Conformations: A First-Principles Study on
Protonated Oligopeptides**

Rui Xiong, Li Xu, Yong Tang, Mengge Cao, Hongbao Li *

Institutes of Physical Science and Information Technology, Key Laboratory of Structure and
Functional Regulation of Hybrid Materials, Ministry of Education. Anhui Graphene Engineering
Laboratory. Anhui University, Hefei, Anhui, 230601, China.

*Correspondence to: lihb@mail.ustc.edu.cn

Table S1 The relative electronic energies, relative zero-point vibrational energy correction ($\Delta ZPVE$), relative Gibbs free energy correction (ΔG) and the corrected electronic and Gibbs free energies (in kcal mol⁻¹) of the most stable conformers of different GGGGH isomers at 298 K, with optimization at the B3LYP-D3(BJ)/6-311++G (d, p) level and single-point energy calculations by the MP2/6-311++G (2df, 2p) and advanced DSD-PBEP86-D3(BJ)/aug-cc-pVTZ methods.

Conf.	Relative electronic energy		$\Delta ZPVE$	ΔG	Corrected relative electronic energy		Corrected relative Gibbs free energy	
	E_{mp2}	E_{dsd}			E_{mp2}	E_{dsd}	G_{mp2}	G_{dsd}
	GGGGH _{a1}	0.00			0.00	0.00	0.00	0.00
GGGGH _{a2}	-0.16	0.38	0.32	0.22	0.16	0.71	0.06	0.60
GGGGH _{a3} *	1.46	1.22	-0.23	-1.26	1.23	0.99	0.21	-0.04
GGGGH _{a4}	0.82	1.67	-0.33	-0.07	0.49	1.34	0.75	1.60
GGGGH _{a5}	0.51	1.60	0.18	0.37	0.69	1.78	0.88	1.97
GGGGH _{a6}	1.48	2.00	-0.20	0.50	1.28	1.80	1.98	2.50
GGGGH _{a7} *	2.70	2.81	-0.34	0.14	2.36	2.47	2.85	2.95
GGGGH _{a8} *	2.75	3.11	-0.46	0.14	2.30	2.65	2.90	3.25
GGGGH _{a9}	2.12	2.58	0.09	0.25	2.21	2.67	2.37	2.83
GGGGH _{a10} *	3.45	3.20	-0.10	-0.01	3.35	3.11	3.43	3.19
GGGGH _{b1} *	10.62	9.31	-2.12	-3.92	8.51	7.19	6.70	5.38
GGGGH _{b2}	13.22	10.86	-2.88	-6.19	10.34	7.99	7.03	4.68
GGGGH _{b3}	13.30	10.93	-2.90	-6.23	10.40	8.04	7.07	4.70
GGGGH _{b4}	12.36	10.82	-2.14	-4.26	10.22	8.69	8.09	6.56
GGGGH _{b5} *	12.82	11.22	-1.80	-3.91	11.01	9.42	8.91	7.31
GGGGH _{c1} *	8.2	7.32	-1.07	0.02	7.13	6.24	8.23	7.34
GGGGH _{c2}	13.5	11.12	-3.8	-7.34	9.71	7.33	6.16	3.78
GGGGH _{c3}	13.29	11.31	-2.43	-5.03	10.87	8.89	8.26	6.29
GGGGH _{c4}	13.57	11.45	-2.52	-5.17	11.05	8.93	8.39	6.27
GGGGH _{c5}	13.63	11.67	-2.36	-5.09	11.27	9.31	8.54	6.58
GGGGH _{d1}	11.63	10.15	-2.3	-4.05	9.32	7.84	7.58	6.1
GGGGH _{d2}	12.32	10.57	-2.6	-4.94	9.72	7.97	7.38	5.63
GGGGH _{d3}	13.27	12.03	-2.31	-4.31	10.95	9.71	8.96	7.72
GGGGH _{d4}	13.96	12.48	-2.58	-5.16	11.38	9.9	8.8	7.32
GGGGH _{d5} *	15.45	13.53	-2.74	-5.05	12.72	10.79	10.4	8.47

Table S2 The relative electronic energies, relative zero-point vibrational energy correction ($\Delta ZPVE$), relative Gibbs free energy correction (ΔG) and the corrected electronic and Gibbs free energies (in kcal mol⁻¹) of the most stable conformers of different GGGGGH isomers at 298 K, with optimization at the B3LYP-D3(BJ)/6-311++G (d, p) level and single-point energy calculations by the MP2/6-311++G (2df, 2p) and advanced DSD-PBEP86-D3(BJ)/aug-cc-pVTZ methods.

Conf.	Relative electronic energy		$\Delta ZPVE$	ΔG	Corrected relative electronic energy		Corrected relative Gibbs free energy	
	Emp2	E ^d _{sd}			Emp2	E ^d _{sd}	G ^{mp2}	G ^d _{sd}
	GGGGGH _{a1}	0.00			0.00	0.00	0.00	0.00
GGGGGH _{a2}	1.56	0.87	-0.64	-0.82	0.92	0.23	0.74	0.05
GGGGGH _{a3}	0.32	0.31	0.18	0.75	0.50	0.49	1.07	1.06
GGGGGH _{a4}	1.10	0.76	0.22	0.77	1.32	0.99	1.86	1.53
GGGGGH _{a5}	0.71	0.81	0.30	0.20	1.02	1.12	0.91	1.01
GGGGGH _{a6}	2.41	2.33	-0.53	-1.40	1.88	1.80	1.01	0.93
GGGGGH _{a7} *	3.90	3.19	-1.01	-3.00	2.89	2.18	0.89	0.19
GGGGGH _{a8} *	4.37	3.46	-0.93	-3.05	3.44	2.54	1.32	0.41
GGGGGH _{a9}	3.67	3.18	-0.27	-0.86	3.40	2.92	2.80	2.32
GGGGGH _{a10}	3.95	4.06	-1.12	-3.00	2.83	2.94	0.95	1.06
GGGGGH _{b1}	11.36	10.33	-1.56	-3.03	9.80	8.78	8.34	7.31
GGGGGH _{b2}	11.80	10.90	-1.59	-1.96	10.21	9.32	9.85	8.95
GGGGGH _{b3}	12.36	11.28	-1.55	-3.36	10.81	9.73	9.01	7.92
GGGGGH _{b4}	13.01	11.43	-1.49	-3.06	11.51	9.93	9.95	8.37
GGGGGH _{b5}	12.80	12.11	-1.78	-3.89	11.02	10.33	8.91	8.21
GGGGGH _{c1} *	13.15	10.97	-2.71	-3.48	10.44	8.26	9.67	7.49
GGGGGH _{c2} *	14.84	12.4	-2.47	-3.91	12.36	9.93	10.92	8.49
GGGGGH _{c3} *	15.25	12.84	-2.44	-3.94	12.81	10.41	11.31	8.90
GGGGGH _{c4} *	16.20	13.88	-2.55	-3.32	13.65	11.33	12.88	10.56
GGGGGH _{c5}	15.89	14.16	-1.5	-2.02	14.39	12.66	13.87	12.13

Table S3 The relative electronic energies, relative zero-point vibrational energy correction ($\Delta ZPVE$), relative Gibbs free energy correction (ΔG) and the corrected electronic and Gibbs free energies (in kcal mol⁻¹) of the most stable conformers of different GAGH isomers at 298 K, with optimization at the B3LYP-D3(BJ)/6-311++G (d, p) level and single-point energy calculations by the MP2/6-311++G (2df, 2p) and advanced DSD-PBEP86-D3(BJ)/aug-cc-pVTZ methods.

Conf.	Relative electronic energy		$\Delta ZPVE$	ΔG	Corrected relative electronic energy		Corrected relative Gibbs free energy	
	Emp2	E ^d _{sd}			Emp2	E ^d _{sd}	G ^{mp2}	G ^d _{sd}
	GAGH _{a1} *	0.00			0.00	0.00	0.00	0.00
GAGH _{a2}	0.75	0.71	-0.17	-0.43	0.59	0.54	0.32	0.28
GAGH _{a3} *	1.83	1.80	-0.27	-0.83	1.56	1.53	1.00	0.97
GAGH _{a4} *	1.84	1.89	0.13	0.37	1.97	2.02	2.22	2.27
GAGH _{a5} *	1.84	1.90	0.13	0.36	1.97	2.03	2.21	2.26
GAGH _{a6}	2.94	2.84	-0.22	-0.43	2.72	2.62	2.52	2.41
GAGH _{a7}	2.70	2.89	-0.05	0.04	2.64	2.83	2.74	2.93
GAGH _{a8}	2.74	2.86	0.23	1.11	2.97	3.09	3.85	3.97
GAGH _{a9} *	3.64	3.27	-0.06	-0.87	3.58	3.21	2.78	2.40
GAGH _{a10} *	3.89	3.75	-0.31	-0.57	3.59	3.44	3.32	3.18
GAGH _{b1}	3.80	1.98	-1.95	-2.82	1.85	0.03	0.99	-0.84
GAGH _{b2}	4.41	2.38	-2.07	-2.85	2.34	0.30	1.56	-0.47
GAGH _{b3}	4.62	3.05	-1.76	-2.84	2.86	1.29	1.78	0.21
GAGH _{b4}	4.72	3.30	-1.72	-2.79	3.00	1.58	1.93	0.51
GAGH _{b5}	5.26	3.64	-1.78	-2.75	3.49	1.86	2.51	0.89
GAGH _{c1}	9.84	7.94	-1.62	-2.07	8.22	6.32	7.77	5.87

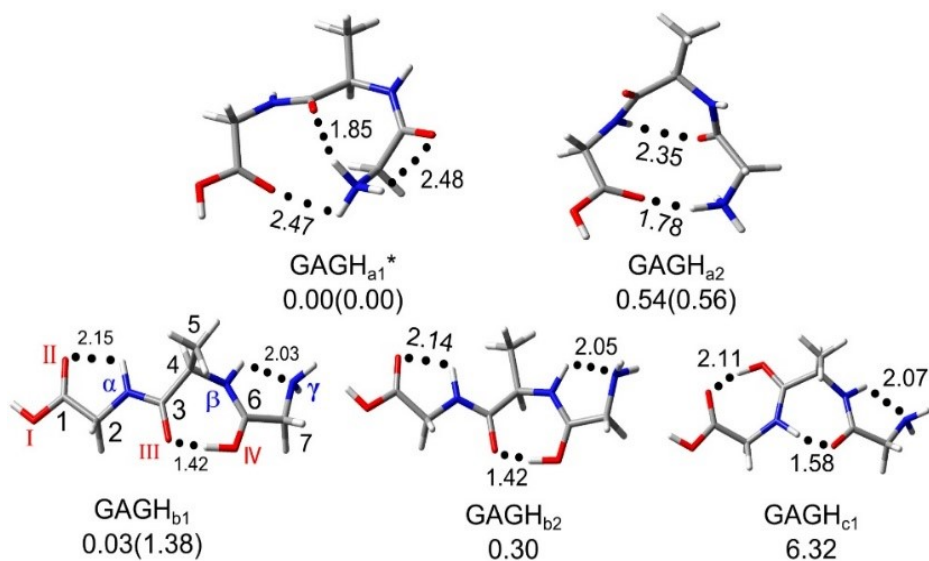


Fig. S1 Geometric structures of the most stable conformers of different isomers of GAGH and their relative energies based on temperature independent electronic energies (in kcal mol⁻¹) calculated at the DSD-PBEP86-D3(BJ)/aug-cc-pVTZ level with zero-point correction from B3LYP-D3(BJ)/6-311++G (d, p). The relative energies in parentheses were determined by further CBS-QB3 calculations.

Table S4 Theoretical values of PA and GB (in kcal mol⁻¹) for different oligopeptides GG, AAA, GAG, GGGG and GGGGG based on different protonated isomers and the corresponding neutral molecules calculated at the DSD-PBEP86-D3(BJ)/aug-cc-pVTZ level with enthalpy and Gibbs free energy corrections at the B3LYP-D3(BJ)/6-311++G (d, p) level, together with previous experimental results ^{a-e} for comparison.

		GG							
		Experiment			Theory				
		MIKE ^a	FT-MS ^b	FT-ICR ^c	GGH _a	GGH _a [*]	GGH _b	Ave	
PA		219.1±0.8	224.5	223.6	221.66	219.96	220.79	221.58	
GB		-	213.5±2	215.3±2	213.38	210.96	212.04	213.25	
		AAA							
		Experiment			Theory				
		Ref. ^d			AAAH _a [*]	AAAH _a	AAAH _b	Ave	
PA		-	-	-	233.16	230.83	231.49	232.94	
GB		221.08	-	-	224.32	221.19	223.00	224.14	
		GAG							
		Experiment			Theory				
				FT-ICR ^c	GAGH _a [*]	GAGH _a	GAGH _b	Ave	
PA				226.8	228.37	228.01	228.28	228.35	
GB				218.5±2.4	219.32	219.18	220.11	220.00	
		GGGG							
		Experiment			Theory				
		MIKE ^a	FT-MS ^b	FT-ICR ^c	GGGGH _a	GGGGH _a [*]	GGGGH _b	GGGGH _b [*]	GGGGH _c
PA		227.2±0.4	226	233.3	232.68	231.36	224.89	224.16	224.68
GB		-	215±2	225±2.3	221.69	221.84	216.38	217.23	218.10
		GGGGG							
		Experiment			Theory				
		MIKE ^a	FT-MS ^b	FT-ICR ^c	GGGGGH _a	GGGGGH _a [*]	GGGGGH _b	GGGGGH _c	Ave
PA		231.8±0.7	228.1	233.6	235.39	232.84	226.14	226.90	234.55
GB		-	217.1±2	225.3±4	226.48	226.48	219.07	219.03	226.64

^a MIKE (mass analyzed ion kinetic energy) experiment from ref. 52.

^b FT-MS (Fourier transform mass spectrometry) experiment from ref. 53.

^c FT-ICR (Fourier transform ion cyclotron resonance mass spectrometer) experiment from ref. 54.

^d from ref. 55.

^e FT-ICR experiment from ref. 56.

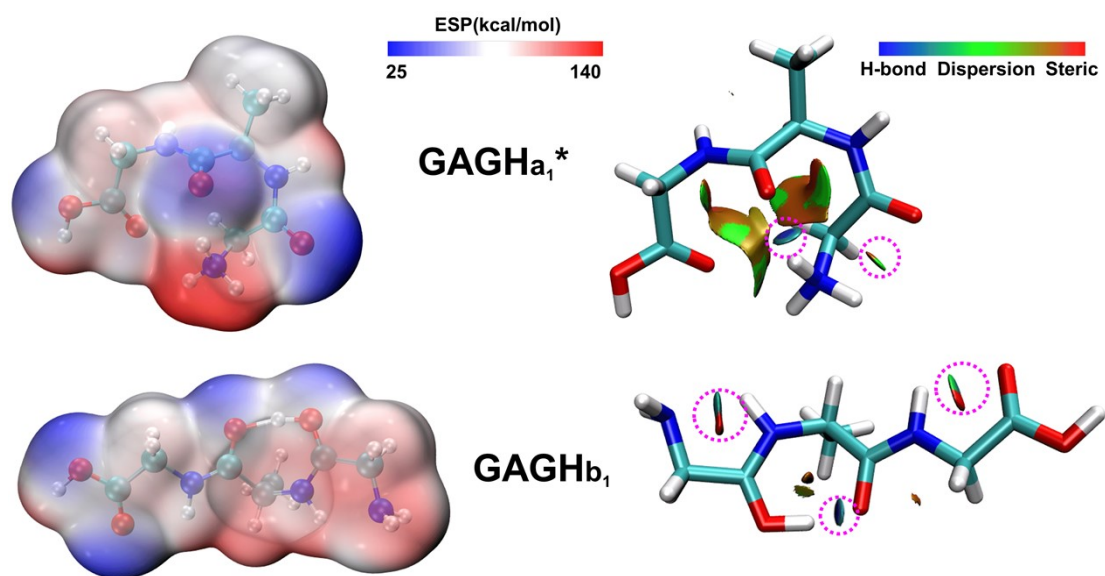


Fig. S2 The ESP distributions and the intra-molecular interactions (including the dispersion interaction) for the most stable conformers of the cis- and trans- tripeptide $GAGH_{a_1}^*$ and $GAGH_{b_1}$.

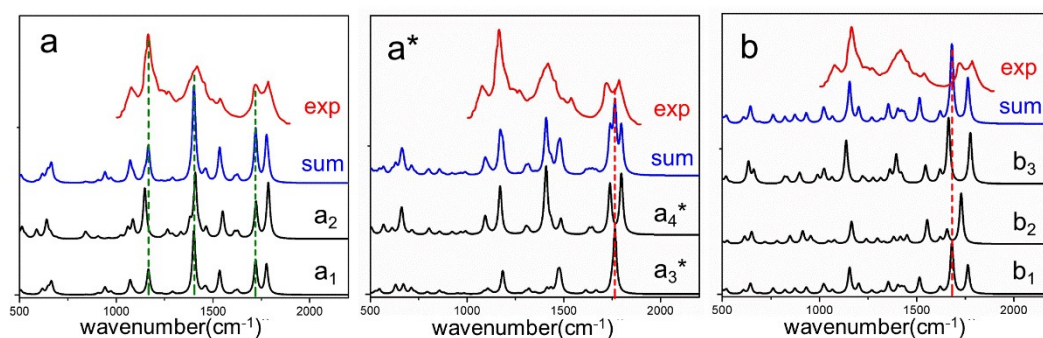
Table S5 Summary of surface analysis for the protonated oligopeptides

Conf.	Overall surface area (Å ²)	V _{max} (kcal/mol)	V _{min} (kcal/mol)	MPI (Kcal/mol)
GGH _{a3} *	158.51	149.66	36.01	98.03
GGH _{a1}	168.03	151.84	32.22	93.57
AAAH _{a1} *	263.14	135.60	25.54	79.27
AAAH _{a3}	256.75	140.16	24.03	78.05
GAGH _{a1} *	223.34	132.17	25.31	84.44
GAGH _{a2}	227.04	144.50	21.22	82.73
GGGGH _{a3} *	253.32	119.07	20.65	79.27
GGGGH _{a1}	252.96	134.57	27.75	79.82
GGGGGH _{a7} *	298.92	116.03	18.73	73.43
GGGGGH _{a1}	275.28	127.35	33.92	75.99

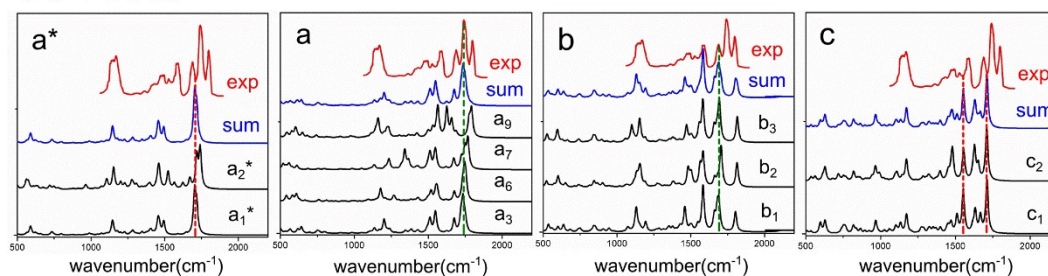
Table S6 Representative isomers of protonated oligopeptides, together with their respective percent abundances at 298K, which were calculated by the DSD-PBEP86-D3(BJ)/aug-cc-pVTZ with Gibbs free energy corrections at the B3LYP-D3(BJ)/6-311++G (d, p) level.

GGH _a	Percent (%)	GGH _a *	Percent (%)	GGH _b	Percent (%)		
GGH _{a1}	84.78	GGH _{a3} *	43.49	GGH _{b1}	92.60		
GGH _{a2}	14.88	GGH _{a4} *	53.27	GGH _{b2}	1.58		
SUM	99.66	SUM	96.76	GGH _{b3}	4.16		
				SUM	98.34		
AAAHA*	Percent (%)	AAAHA	Percent (%)	AAAHB	Percent (%)	AAAHC	Percent (%)
AAAHA ₁ *	93.74	AAAHA ₃	65.61	AAAHB ₁	60.88	AAAHC ₁	63.40
AAAHA ₂ *	4.38	AAAHA ₆	19.86	AAAHB ₂	26.88	AAAHC ₂	36.6
SUM	98.12	AAAHA ₇	4.98	AAAHB ₃	10.21	SUM	100
		AAAHA ₉	4.68	SUM	97.97		
		SUM	95.13				
GAGH _a *	Percent (%)	GAGH _a	Percent (%)	GAGH _b	Percent (%)		
GAGH _{a1} *	79.40	GAGH _{a2}	96.09	GAGH _{b1}	52.47		
GAGH _{a3} *	15.39	GAGH _{a6}	2.62	GAGH _{b2}	28.32		
SUM	94.79	SUM	98.71	GAGH _{b3}	8.97		
				GAGH _{b4}	5.37		
				SUM	95.13		
GGGGH _a	Percent (%)	GGGGH _a *	Percent (%)	GGGGH _b *	Percent (%)		
GGGGH _{a1}	65.84	GGGGH _{a3} *	98.57	GGGGH _{b1} *	89.87		
GGGGH _{a2}	23.81	GGGGH _{a7} *	0.63	GGGGH _{b5} *	3.46		
GGGGH _{a4}	4.43	SUM	99.20	GGGGH _{b7} *	4.41		
SUM	94.08			SUM	97.74		
GGGGH _b	Percent (%)	GGGGH _c	Percent (%)	GGGGH _d	Percent (%)		
GGGGH _{b2}	49.24	GGGGH _{c2}	95.20	GGGGH _{d1}	28.99		
GGGGH _{b3}	46.85	GGGGH _{c3}	1.39	GGGGH _{d2}	64.90		
GGGGH _{b4}	2.05	GGGGH _{c4}	1.42	GGGGH _{d4}	3.69		
SUM	98.14	SUM	98.01	SUM	97.58		
GGGGGH _a	Percent (%)	GGGGGH _a *	Percent (%)	GGGGGH _b	Percent (%)		
GGGGGH _{a1}	33.42	GGGGGH _{a7} *	59.24	GGGGGH _{b1}	42.82		
GGGGGH _{a2}	30.53	GGGGGH _{a8} *	40.76	GGGGGH _{b3}	15.18		
GGGGGH _{a5}	6.07	SUM	100	GGGGGH _{b4}	7.14		
GGGGGH _{a6}	6.93			GGGGGH _{b5}	9.30		
SUM	76.95			GGGGGH _{b10}	6.00		
				SUM	80.44		
GGGGGH _b *	Percent (%)	GGGGH _c	Percent (%)				
GGGGGH _{b15} *	100	GGGGH _{c1}	77.98				
SUM	100	GGGGH _{c2}	14.40				
		GGGGH _{c3}	7.14				
		SUM	99.52				

(1) GGH



(2) AAAH



(3) GAGH

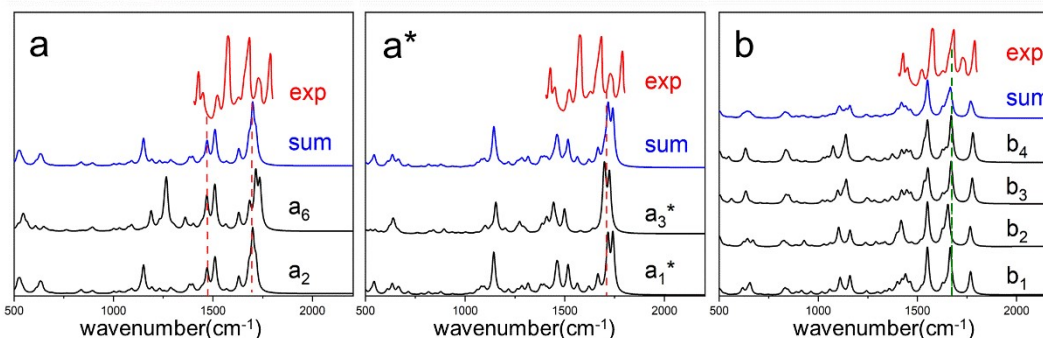
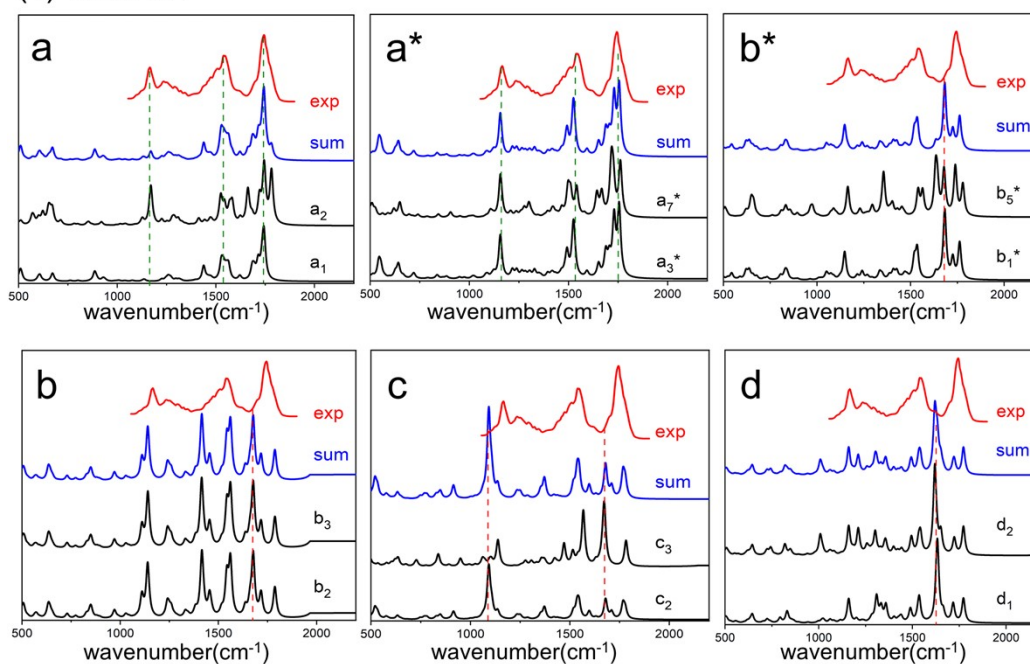


Fig. S3 Simulated IR spectra of the dominant conformers for different isomers of (1) GGH, (2) AAAH and (3) GAGH at 298 K, calculated at the B3LYP-D3(BJ)/6-311++G (d, p) level of theory, as well as their summation (“sum” in blue) calculated using the percentage abundances listed in Table S6, together with the experimental IRMPD results (“exp” in red from ref. 8, 9, 58) for comparison. A Lorentzian profile with a full width at half maximum (FWHM) of 20 cm⁻¹ is used to convolute the calculated spectra.

(1) GGGGH



(2) GGGGGH

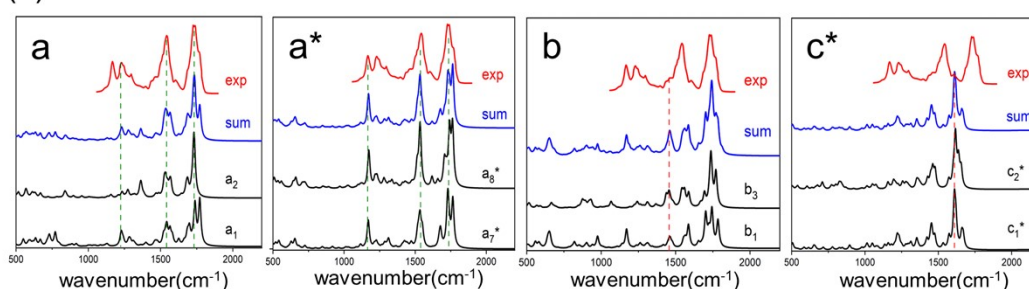


Fig. S4 Simulated IR spectra of the dominant conformers for different isomers of (1) GGGGH and (2) GGGGGH at 298K calculated at the B3LYP-D3(BJ)/6-311++G (d, p) level of theory, as well as their summation (“sum” in blue) calculated using the percentage abundances listed in Table S6, together with the experimental IRMPD results (“exp” in red from ref. 9) for comparison. A Lorentzian profile with a full width at half maximum (FWHM) of 20 cm⁻¹ is used to convolute the calculated spectra.

Table S7 Scaling factors obtained according to the IR spectral comparisons between the anharmonic and harmonic frequencies calculations on the most stable conformer of each isomer at the B3LYP-D3(BJ)/6-311++G (d, p) level.

GGH	GGH _a 0.98	GGH _a * 0.992	GGH _b 0.970		
AAAH	AAAH _a * 0.975	AAAH _a 0.975	AAAH _b 1	AAAH _c 0.970	
GAGH	GAGH _a * 0.972	GAGH _a 0.970	GAGH _b 0.975	GAGH _c 0.975	
GGGH	GGGGH _a 0.995	GGGGH _a * 0.980	GGGGH _b * 0.980	GGGGH _c 0.975	GGGGH _d 0.975
GGGGH	GGGGH _a 0.987	GGGGH _a * 0.995	GGGGH _b 0.990	GGGGH _c 0.940	

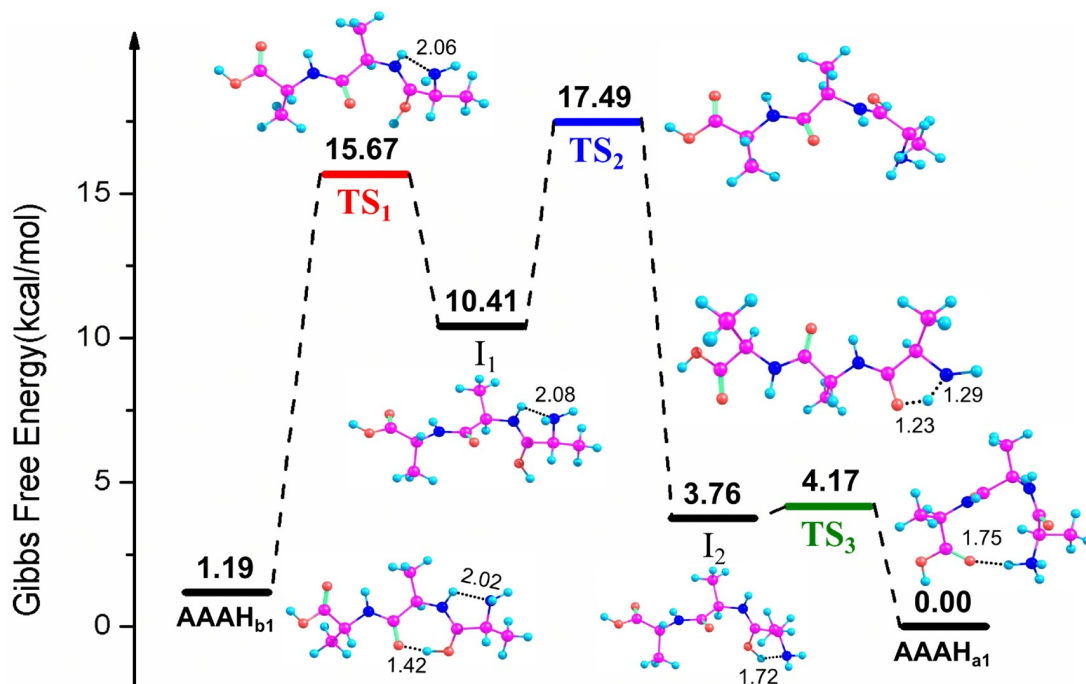


Fig. S5 Transition states (TS) for the interconversion between the most abundant conformers of AAAH_{a1} and AAAH_{b1} at 298K. The relative Gibbs free energy values were determined at the DSD-PBEP86-D3(BJ)/aug-cc-pVTZ level of theory with the Gibbs free energy corrections at the B3LYP-D3(BJ)/6-311++G (d, p) level. Here I₁ and I₂ are two intermediate structures. TS₁: AAAH_{b1} to I₁; TS₂: I₁ to I₂; TS₃: I₂ to AAAH_{a1}.

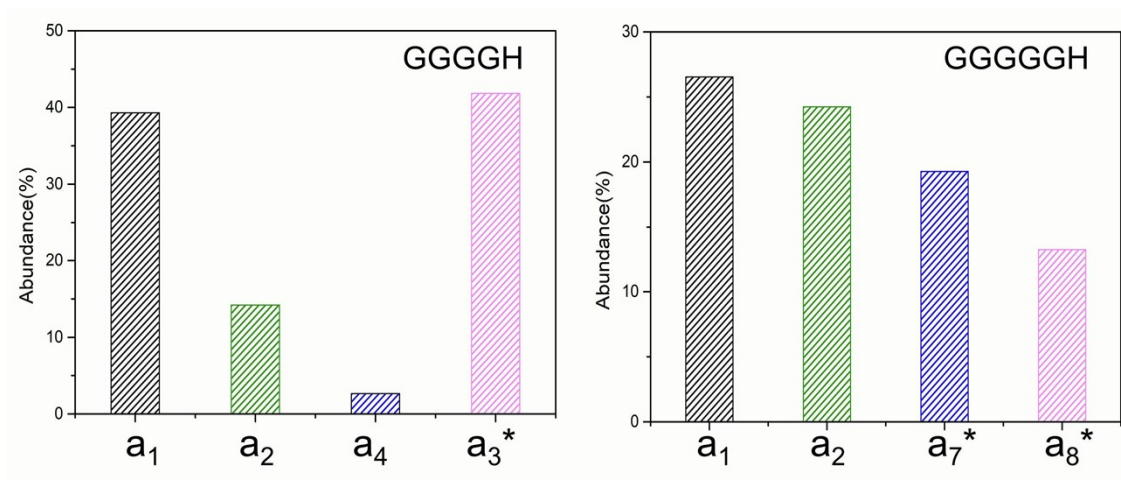


Fig. S6 Percent abundances for different isomers of protonated oligopeptide GGGGH and GGGGGH at 298 K. The abundances were calculated by DSD-PBEP86-D3(BJ)/aug-cc-pVTZ with Gibbs free energy corrections at the B3LYP-D3(BJ)/6-311++G (d, p) level of theory.

Table S8 Representative isomers of protonated oligopeptides GGH, AAAH, GAGH, GGGGH and GGGGGH, together with their respective percent abundances at 498K. The abundances were calculated by DSD-PBEP86-D3(BJ)/aug-cc-pvtz with Gibbs free energy corrections at the B3LYP-D3(BJ)/6-311++G (d, p) level of theory.

GGH	Percent (%)	AAAH	Percent (%)	GAGH	Percent (%)	GGGGH	Percent (%)	GGGGGH	Percent (%)
GGH _{a1}	56.79	AAAH _{a1} *	48.37	GAGH _{a1} *	7.28	GGGGH _{a1}	14.51	GGGGGH _{a1}	6.47
GGH _{a2}	22.66	AAAH _{a2} *	4.93	GAGH _{a2}	6.75	GGGGH _{a2}	9.71	GGGGGH _{a2}	7.00
GGH _{a4} *	3.89	AAAH _{a3}	1.29	GAGH _{b1}	31.87	GGGGH _{a3} *	45.31	GGGGGH _{a10}	11.46
GGH _{b1}	11.03	AAAH _{b1}	15.18	GAGH _{b2}	20.21	GGGGH _{a7} *	16.60	GGGGGH _{a11}	11.50
SUM	94.37	AAAH _{b2}	14.55	GAGH _{b3}	13.07	GGGGH _{b1} *	0.37	GGGGGH _{b7} *	32.60
		AAAH _{b3}	7.09	GAGH _{b4}	9.58	GGGGH _{b2}	2.29	GGGGGH _{b8} *	30.93
		AAAH _{c1}	0.05	GAGH _{c1}	0.00	GGGGH _{c2}	7.47	GGGGGH _{b1}	0.01
		SUM	91.46	SUM	88.76	GGGGH _{d2}	0.46	GGGGGH _{c1}	0.00
						SUM	96.72	SUM	99.97

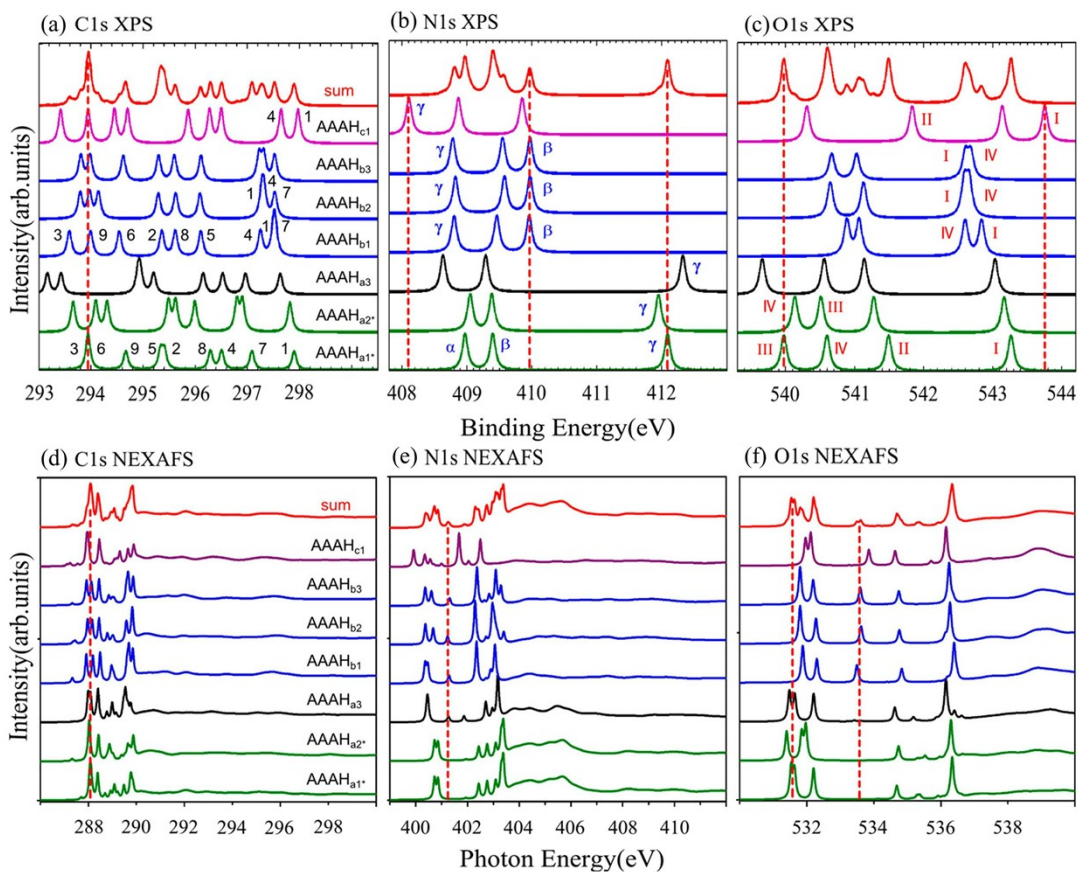


Fig. S7 Calculated (a) C1s, (b) N1s, and (c) O1s XPS spectra and (d) C1s, (e) N1s, and (f) O1s NEXAFS spectra of the low-energy conformers of AAAH as well as the averaged spectra (“sum” plotted in red) at 498 K, computed according to their equilibrium distributions. Spectra of each conformer are plotted in a colour connoting the isomer to which the conformer belongs.

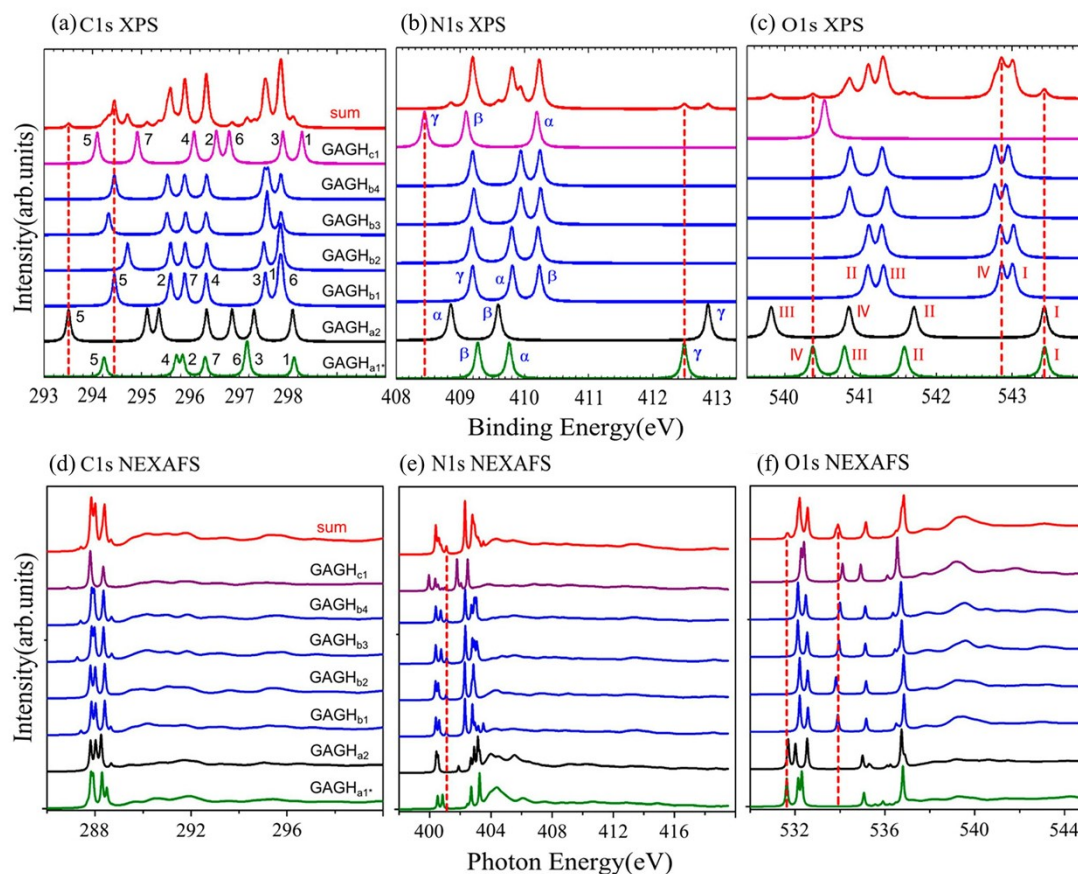


Fig. S8 Calculated (a) C1s, (b) N1s, and (c) O1s XPS spectra and (d) C1s, (e) N1s, and (f) O1s NEXAFS spectra of the lowest-energy conformers of GAGH as well as the averaged spectra (“sum” plotted in red) at 498 K, computed according to their equilibrium distributions. Spectra of each conformer are plotted in a colour connoting the isomer to which the conformer belongs.

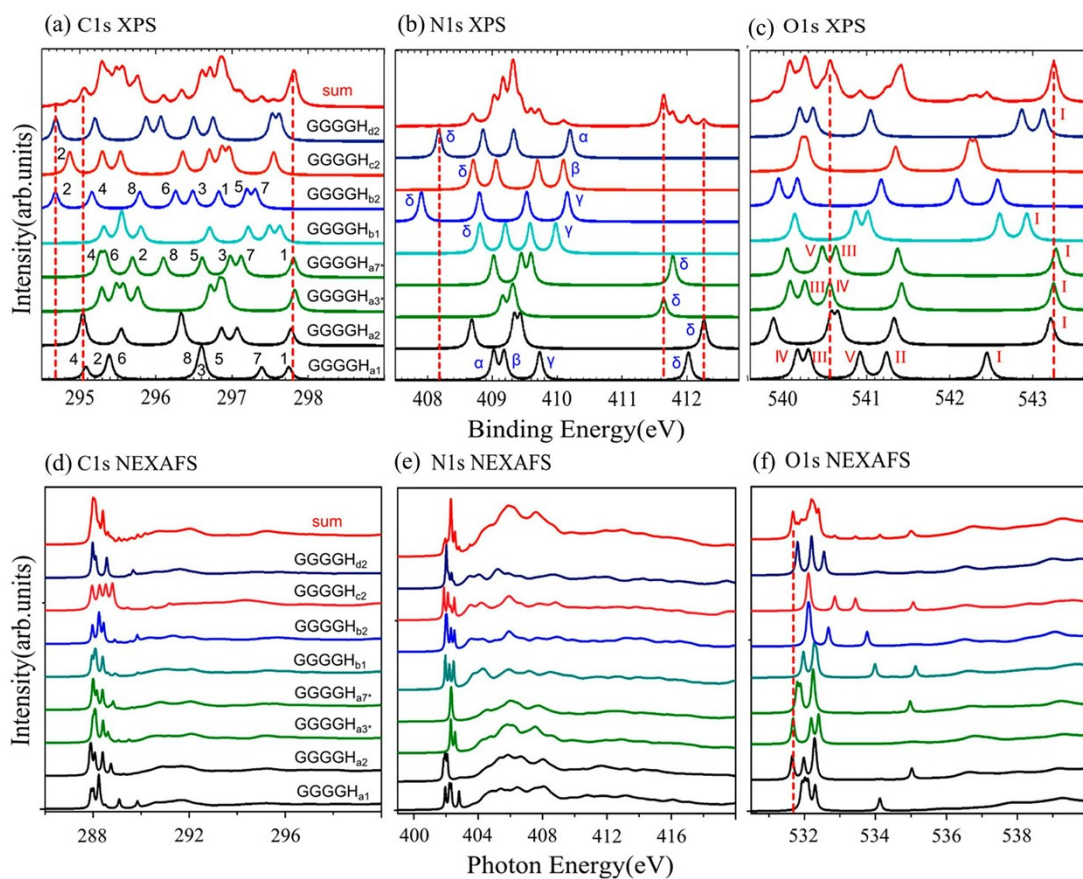


Fig. S9 Calculated (a) C1s, (b) N1s, and (c) O1s XPS spectra and (d) C1s, (e) N1s, and (f) O1s NEXAFS spectra of the lowest-energy conformers of GGGGH as well as the averaged spectra (“sum” plotted in red) at 498 K, computed according to their equilibrium distributions. Spectra of each conformer are plotted in a colour connoting the isomer to which the conformer belongs.

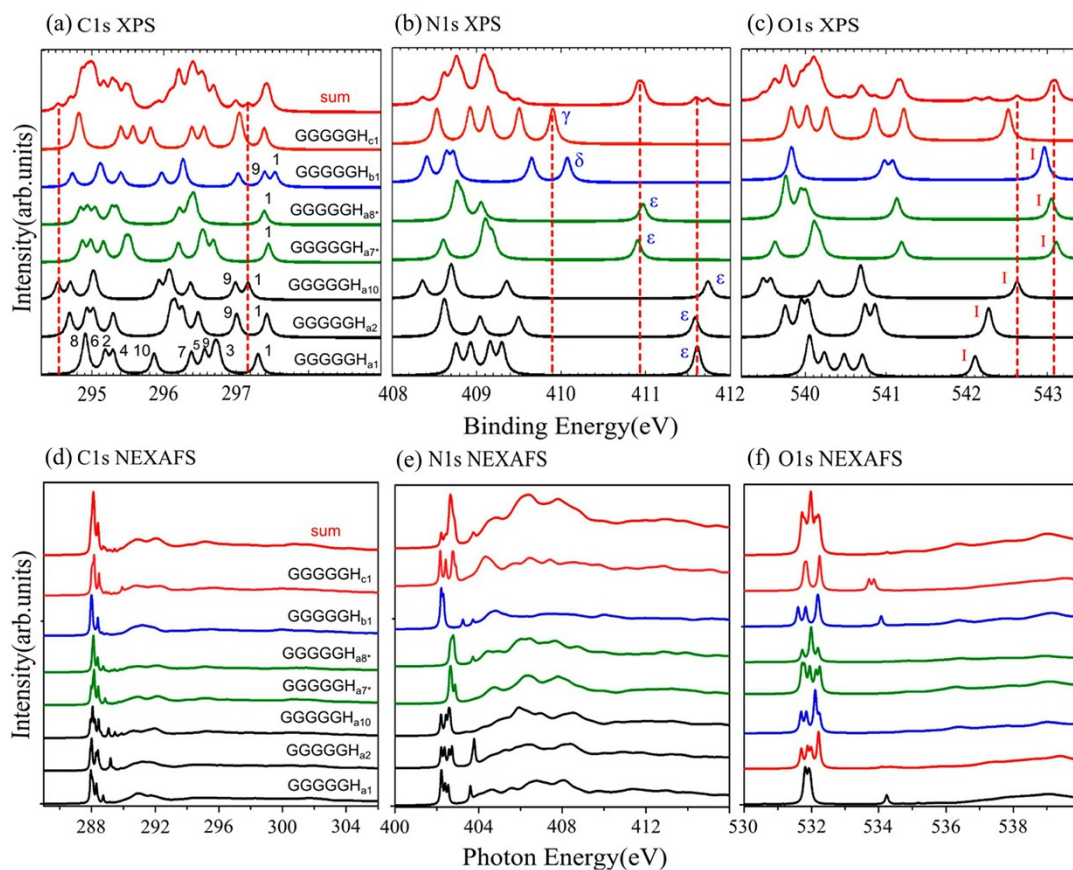


Fig. S10 Calculated (a) C1s, (b) N1s, and (c) O1s XPS spectra and (d) C1s, (d) N1s, and (f) O1s NEXAFS spectra of the lowest-energy conformers of GGGGGH as well as the averaged spectra (“sum” plotted in red) at 498 K, computed according to their equilibrium distributions. Spectra of each conformer are plotted in a colour connoting the isomer to which the conformer belongs.