Supplementary Information on manuscript: Solvent quality and solvent polarity in polypeptides

Cedrix J. Dongmo Foumthuim[a1,†] and Achille Giacometti[c2,§]

[1] *Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia Campus Scientifico, Edificio Alfa, via Torino 155,30172 Venezia Mestre, Italy*

[2] *Dipartimento di Scienze Molecolari e Nanosistemi, Università Ca' Foscari di Venezia Campus Scientifico, Edificio Alfa, via Torino 155,30170 Venezia Mestre, Italy*

(Dated: December 17, 2022)

[a] Corresponding author
[c] Corresponding author
† cedrix.dongmo@unive.it
§ achille.giacometti@unive.it

TABLE SI: Overview of the polymer length considered in this work. The length referred to the equilibrium bond length between CA end-points of the minimized structures.

Oligomers	Length (nm)
GLY3	0.687
ALA3	0.727
ILE3	0.728
ASN3	0.729
LYS3	0.737
ARG3	0.727
GLU3	0.736
GLY5	1.297
ALA5	1.308
ILE5	1.455
ASN5	1.378
LYS5	1.346
ARG5	1.443
GLU5	1.335
GLY7	1.934
ALA7	1.852
ILE7	2.187
ASN7	1.695
LYS7	1.884
ARG7	2.190
GLU7	1.698
GLY9	2.278
ALA9	2.454
ILE9	2.917
ASN9	2.260
LYS9	2.509
ARG9	2.906
GLU9	2.499
GLY11	3.050
ALA11	3.009
ILE11	3.639
ASN11	3.041
LYS11	3.077
ARG11	3.640
GLU11	3.094

TABLE SII: Enthalpic and entropic contributions to the solvation free energies ($kJ\,mol^{-1}$) for different homopeptides studied in water H_2O at 25°C.

	ΔG	ΔH	$-$TΔS
GLY3	-103.96 ± 1.33	-140.03 ± 20.94	36.07 ± 20.36
ALA3	-110.65 ± 3.48	-166.68 ± 10.88	56.02 ± 10.28
ILE3	-147.15 ± 0.85	-222.46 ± 24.43	80.30 ± 24.85
ASN3	-239.20 ± 1.40	-331.29 ± 80.96	92.09 ± 82.23
LYS3	-947.56 ± 3.35	-1071.57 ± 65.25	124.00 ± 66.01
ARG3	-807.52 ± 1.73	-968.39 ± 20.05	160.87 ± 21.48
GLU3	-1483.58 ± 4.41	-1678.94 ± 26.61	195.36 ± 26.72
GLY5	-247.59 ± 1.98	-396.84 ± 24.30	149.25 ± 24.15
ALA5	-188.55 ± 1.34	-304.16 ± 17.68	115.60 ± 17.97
ILE5	-225.04 ± 2.36	-481.68 ± 20.65	256.64 ± 22.29
ASN5	-427.63 ± 4.44	-945.02 ± 48.97	517.39 ± 46.01
LYS5	-2012.55 ± 3.43	-2311.32 ± 64.96	298.77 ± 66.40
ARG5	-1525.69 ± 2.61	-1753.10 ± 43.43	227.41 ± 41.36
GLU5	-2958.77 ± 2.05	-3210.93 ± 34.18	252.16 ± 34.23
GLY7	-248.41 ± 1.11	-418.20 ± 27.76	169.79 ± 26.92
ALA7	-215.44 ± 2.50	-596.16 ± 43.31	380.73 ± 41.40
ILE7	-282.01 ± 3.36	-662.74 ± 15.21	380.74 ± 12.81
ASN7	-499.82 ± 5.52	-924.84 ± 32.24	425.02 ± 31.50
LYS7	-2883.17 ± 3.03	-3160.39 ± 70.84	277.21 ± 71.04
ARG7	-2320.68 ± 4.32	-2565.96 ± 32.71	245.28 ± 31.64
GLU7	-4335.11 ± 3.56	-4675.27 ± 26.63	340.15 ± 26.82
GLY9	-284.09 ± 3.56	-411.45 ± 49.88	127.36 ± 48.68
ALA9	-257.73 ± 8.44	-664.17 ± 25.96	406.44 ± 24.55
ILE9	-335.52 ± 6.64	-689.46 ± 26.19	353.94 ± 26.09
ASN9	-609.45 ± 7.21	-1197.34 ± 123.14	587.89 ± 119.73
LYS9	-3781.51 ± 6.07	-4132.75 ± 87.92	351.24 ± 86.58
ARG9	-3119.82 ± 4.16	-3317.20 ± 33.65	197.38 ± 31.27
GLU9	-5748.86 ± 3.60	-6373.04 ± 90.51	624.18 ± 93.01
GLY11	-319.09 ± 5.48	-590.84 ± 40.74	271.75 ± 42.26
ALA11	-319.88 ± 3.48	-858.95 ± 10.78	539.07 ± 10.21
ILE11	-405.29 ± 7.14	-1037.62 ± 38.83	632.33 ± 35.10
ASN11	-727.72 ± 13.26	-1787.68 ± 101.14	1059.96 ± 88.54
LYS11	-4623.86 ± 9.73	-5076.97 ± 57.91	453.11 ± 58.82
ARG11	-3923.60 ± 7.77	-4144.92 ± 69.78	221.31 ± 64.19
GLU11	-7116.52 ± 14.13	-7713.11 ± 42.26	596.59 ± 49.52

[1] D. Chandler, Nature, 2005, **437**, 640–647.

[2] C. J. Dongmo Foumthuim, M. Carrer, M. Houvet, T. Škrbić, G. Graziano and A. Giacometti, Phys. Chem. Chem. Phys., 2020, **22**, 25848–25858.

TABLE SIII: Enthalpic and entropic contributions to the solvation free energies ($kJ\,mol^{-1}$) for different homopeptides studied in cyclohexane cC_6H_{12} at 25°C.

	ΔG	ΔH	$-T\Delta S$
GLY3	-232.29 ± 3.11	-395.23 ± 27.32	162.94 ± 25.24
ALA3	-248.82 ± 4.17	-386.67 ± 22.74	137.85 ± 23.02
ILE3	-191.46 ± 2.58	-353.30 ± 32.37	161.84 ± 32.60
ASN3	-271.43 ± 7.39	-652.93 ± 44.79	381.50 ± 38.69
LYS3	-1485.29 ± 5.37	-2094.60 ± 19.69	609.31 ± 19.67
ARG3	-1314.45 ± 8.90	-1664.45 ± 37.22	350.00 ± 30.30
GLU3	-1900.58 ± 17.55	-1752.24 ± 159.81	148.33 ± 142.39
GLY5	-281.15 ± 7.83	-642.55 ± 20.29	361.40 ± 13.27
ALA5	-300.77 ± 3.91	-700.65 ± 40.84	399.88 ± 44.53
ILE5	-286.26 ± 5.22	-461.56 ± 21.53	175.30 ± 26.44
ASN5	-425.99 ± 14.92	-860.12 ± 73.30	434.13 ± 87.30
LYS5	-2694.48 ± 17.60	-3167.51 ± 31.33	473.03 ± 41.54
ARG5	-2359.36 ± 17.41	-2784.13 ± 121.95	424.77 ± 104.61
GLU5	-3442.17 ± 19.90	-4048.35 ± 63.45	606.17 ± 55.52
GLY7	-279.38 ± 7.26	-480.52 ± 31.40	201.14 ± 31.06
ALA7	-330.46 ± 12.76	-768.01 ± 33.07	437.55 ± 42.79
ILE7	-362.08 ± 2.95	-657.24 ± 16.85	295.16 ± 17.99
ASN7	-488.48 ± 4.37	-1046.60 ± 67.59	558.12 ± 71.83
LYS7	-3884.66 ± 19.83	-4601.49 ± 63.88	716.83 ± 63.49
ARG7	-3501.44 ± 17.61	-3898.85 ± 57.63	397.41 ± 57.31
GLU7	-5027.21 ± 13.50	-5627.43 ± 79.59	600.22 ± 91.09
GLY9	-367.76 ± 12.69	-872.79 ± 81.04	505.03 ± 79.85
ALA9	-387.98 ± 11.57	-1344.32 ± 49.53	956.34 ± 38.94
ILE9	-432.12 ± 1.94	-892.67 ± 15.64	460.55 ± 14.31
ASN9	-564.66 ± 13.36	-1328.90 ± 69.18	764.24 ± 58.94
LYS9	-5011.07 ± 27.05	-5815.82 ± 45.19	804.75 ± 30.56
ARG9	-4492.76 ± 19.52	-5076.22 ± 118.49	583.47 ± 107.02
GLU9	-6719.31 ± 17.97	-7711.54 ± 94.18	992.23 ± 105.93
GLY11	-522.33 ± 4.72	-1510.80 ± 44.09	988.47 ± 47.32
ALA11	-670.24 ± 13.15	-1779.23 ± 57.24	1108.99 ± 67.95
ILE11	-524.29 ± 5.07	-1144.01 ± 29.47	619.72 ± 26.34
ASN11	-602.91 ± 13.15	-1353.29 ± 57.24	750.38 ± 67.95
LYS11	-6166.10 ± 11.58	-6761.29 ± 143.56	595.19 ± 137.65
ARG11	-5636.40 ± 17.59	-6044.61 ± 49.73	408.21 ± 53.26
GLU11	-8332.91 ± 30.02	-9305.39 ± 116.31	972.48 ± 126.37

TABLE SIV: Enthalpic and entropic contributions to the transfer free energies (kJ mol^{-1}) from water H_2O to cyclohexane cC_6H_{12} for each homopeptide investigated here at 25°C.

	$\Delta\Delta G$	$\Delta\Delta H$	$-T\Delta\Delta S$
GLY3	-128.33 ± 4.44	-255.20 ± 48.26	126.87 ± 45.60
ALA3	-138.17 ± 7.65	-219.99 ± 33.62	81.83 ± 33.30
ILE3	-44.31 ± 3.43	-130.84 ± 56.80	81.54 ± 57.45
ASN3	-32.23 ± 8.79	-321.64 ± 125.75	289.41 ± 121.02
LYS3	-537.73 ± 10.74	-1023.03 ± 84.94	485.31 ± 85.68
ARG3	-506.93 ± 10.63	-696.06 ± 57.27	189.13 ± 51.78
GLU3	-417.00 ± 35.10	-73.30 ± 186.42	-47.03 ± 169.11
GLY5	-33.56 ± 9.81	-245.71 ± 44.59	212.15 ± 37.42
ALA5	-112.22 ± 5.25	-396.49 ± 58.52	284.28 ± 65.50
ILE5	-61.22 ± 7.58	20.12 ± 42.18	-81.34 ± 48.73
ASN5	1.64 ± 19.36	84.90 ± 122.27	-83.26 ± 133.31
LYS5	-681.93 ± 21.03	-856.19 ± 96.29	174.26 ± 107.94
ARG5	-833.67 ± 20.02	-1031.03 ± 165.34	197.36 ± 145.97
GLU5	-483.40 ± 21.95	-837.42 ± 97.63	354.01 ± 87.75
GLY7	-30.97 ± 8.37	-62.32 ± 59.16	31.35 ± 56.98
ALA7	-115.02 ± 15.26	-171.85 ± 76.38	56.82 ± 84.19
ILE7	-80.07 ± 6.31	5.50 ± 32.06	-85.58 ± 30.80
ASN7	11.34 ± 9.89	-121.76 ± 99.83	133.10 ± 103.33
LYS7	-1001.49 ± 22.86	-1441.10 ± 134.72	439.62 ± 134.53
ARG7	-1180.76 ± 21.93	-1332.89 ± 90.34	152.13 ± 88.95
GLU7	-692.10 ± 17.06	-952.16 ± 106.22	260.07 ± 117.91
GLY9	-83.67 ± 16.25	-461.34 ± 130.92	377.67 ± 128.53
ALA9	-130.25 ± 20.01	-680.15 ± 75.49	549.90 ± 63.49
ILE9	-96.60 ± 8.58	-203.21 ± 41.83	106.61 ± 40.40
ASN9	44.79 ± 20.57	-131.56 ± 192.32	176.35 ± 178.67
LYS9	-1229.56 ± 33.12	-1683.07 ± 133.11	453.51 ± 117.14
ARG9	-1372.94 ± 23.68	-1759.02 ± 152.14	386.09 ± 138.29
GLU9	-970.45 ± 21.57	-1338.50 ± 184.69	368.05 ± 198.94
GLY11	-203.24 ± 10.20	-919.96 ± 84.83	716.72 ± 89.58
ALA11	-350.36 ± 26.30	-920.28 ± 68.02	569.92 ± 78.16
ILE11	-119.00 ± 12.21	-106.39 ± 68.30	-12.61 ± 61.44
ASN11	124.81 ± 26.41	434.39 ± 158.38	-309.58 ± 156.49
LYS11	-1542.24 ± 21.31	-1684.32 ± 201.47	142.08 ± 196.47
ARG11	-1712.80 ± 25.36	-1899.69 ± 119.51	186.9 ± 117.45
GLU11	-1216.39 ± 44.15	-1592.28 ± 158.57	375.89 ± 175.89

TABLE SV: Fitting coefficients, a, b and c used to compute the hydration thermodynamics parameters along with their Pearson's correlation coefficients R^2.

	cC_6H_{12}				H_2O			
	a	b	c	R^2	a	b	c	R^2
GLY3	−386.54	0.35	0.03	0.95	−184.55	1.12	−0.14	0.78
ALA3	−927.42	12.61	−1.81	0.99	−706.77	12.32	−1.81	0.97
ILE3	−72.05	−5.77	0.94	0.99	261.82	−10.72	1.64	0.91
ASN3	−4895.90	96.59	−14.23	0.99	452.12	−17.28	2.62	0.91
LYS3	−7837.88	131.05	−19.26	0.99	−125.19	−20.84	3.17	0.89
ARG3	175.01	−40.14	6.16	0.99	−2226.15	28.79	−4.21	0.99
GLU3	−15403.66	306.16	−45.78	0.97	−372.03	−28.70	4.38	0.96
GLY5	−2322.18	38.94	−5.63	0.99	441.93	−18.34	2.81	0.97
ALA5	−2271.59	36.63	−5.26	0.99	409.99	−15.65	2.39	0.97
ILE5	296.05	−16.43	2.54	0.84	341.36	−17.62	2.76	0.99
ASN5	−1917.74	25.21	−3.54	0.98	−4801.74	88.37	−12.93	0.99
LYS5	−6444.36	75.19	−10.99	0.98	−1047.04	−27.39	4.24	0.98
ARG5	−2169.90	−12.37	2.06	0.95	−669.30	−23.58	3.63	0.99
GLU5	−10257.94	141.52	−20.82	0.99	−1064.30	−47.37	7.19	0.99
GLY7	836.72	−28.91	4.41	0.99	−1870.96	33.20	−4.87	0.99
ALA7	−774.92	1.62	−0.02	0.99	−2246.04	38.33	−5.53	0.98
ILE7	−1205.56	13.31	−1.83	0.96	−3141.08	56.95	−8.31	0.99
ASN7	764.39	−38.81	6.07	0.99	−2146.05	28.85	−4.09	0.98
LYS7	−7563.68	68.94	−9.93	0.99	−3260.57	3.18	−0.33	0.98
ARG7	7957.65	−265.01	39.76	0.97	−776.84	−39.36	6.00	0.91
GLU7	−8311.90	62.31	−9.00	0.99	−5.59	2.17	−3.08	0.99
GLY9	−3277.79	55.71	−8.06	0.99	−323.02	1.55	0.29	0.98
ALA9	−15531.18	321.89	−47.58	0.95	−2800.23	49.34	−7.16	0.97
ILE9	2224.87	−68.48	10.45	0.99	−1950.57	29.51	−4.22	0.99
ASN9	−5790.69	102.79	−14.96	0.98	−5357.41	95.42	−13.95	0.98
LYS9	−1263.04	−99.57	15.27	0.90	−6915.64	63.69	−9.33	0.76
ARG9	−12618.43	171.38	−25.29	0.98	−4143.75	19.22	−2.77	0.80
GLU9	−12918.52	120.29	−17.46	0.97	−933.43	−120.10	18.24	0.99
GLY11	6398.38	−174.35	26.52	0.99	−3343.69	62.75	−9.23	0.98
ALA11	3281.17	−109.95	16.97	0.97	−2644.95	41.92	−5.99	0.98
ILE11	616.14	−37.46	5.90	0.99	−1434.39	11.03	−1.33	0.98
ASN11	−6641.13	121.30	−17.73	0.99	−10977.85	210.00	−30.82	0.99
LYS11	−1791.22	−109.65	16.66	0.76	−3979.46	−23.13	3.68	0.99
ARG11	−1213.32	−107.16	16.20	0.86	−944.07	−71.16	10.73	0.83
GLU11	−8367.95	−17.79	3.14	0.97	−11350.91	83.71	−12.20	0.98

TABLE SVI: Summary of the slopes, intercepts to origin, and correlation coefficients for the linear fitting curves $R_g = f(\mathrm{n})$ in water H_2O and cyclohexane cC_6H_{12}. The corresponding plots are visible in figure SII below.

Polymers		GLY_n	ALA_n	ILE_n	ASN_n	LYS_n	ARG_n	GLU_n
H_2O	Slopes	0.02	0.05	0.05	0.05	0.07	0.05	0.06
	Intercepts origin	0.29	0.19	0.25	0.26	0.23	0.40	0.27
	Correlation coefficients	0.88	0.99	0.98	0.99	0.99	0.97	1.00
cC_6H_{12}	Slopes	0.09	0.09	0.05	0.04	0.06	0.04	0.05
	Intercepts origin	-0.14	-0.09	0.22	0.26	0.18	0.41	0.22
	Correlation coefficients	0.84	0.87	0.96	0.71	0.93	0.72	0.89

TABLE SVII: Summary of the slopes, intercepts to origin, and correlation coefficients for the linear fitting curves $-T\Delta S = f(\Delta H)$ a.k.a enthalpy-entropy compensation in water H_2O and cyclohexane cC_6H_{12}.

Polymers		GLY_n	ALA_n	ILE_n	ASN_n	LYS_n	ARG_n	GLU_n
H_2O	Slopes	-0.51	-0.73	-0.67	-0.66	-0.07	-0.01	-0.08
	Intercepts origin	-48.41	-78.80	-73.04	-145.86	69.66	180.86	33.54
	Correlation coefficients	0.97	0.99	0.99	0.99	0.95	0.44	0.95
cC_6H_{12}	Slopes	-0.74	-0.72	-0.61	-0.58	-0.02	-0.03	-0.11
	Intercepts origin	-137.88	-111.98	-82.75	-26.94	509.20	331.30	35.13
	Correlation coefficients	0.99	0.98	0.99	0.98	0.44	0.51	0.95

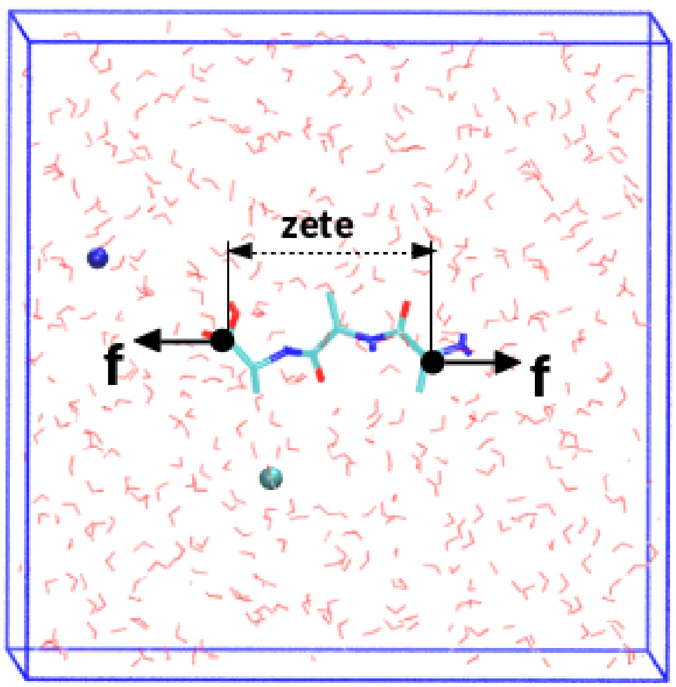
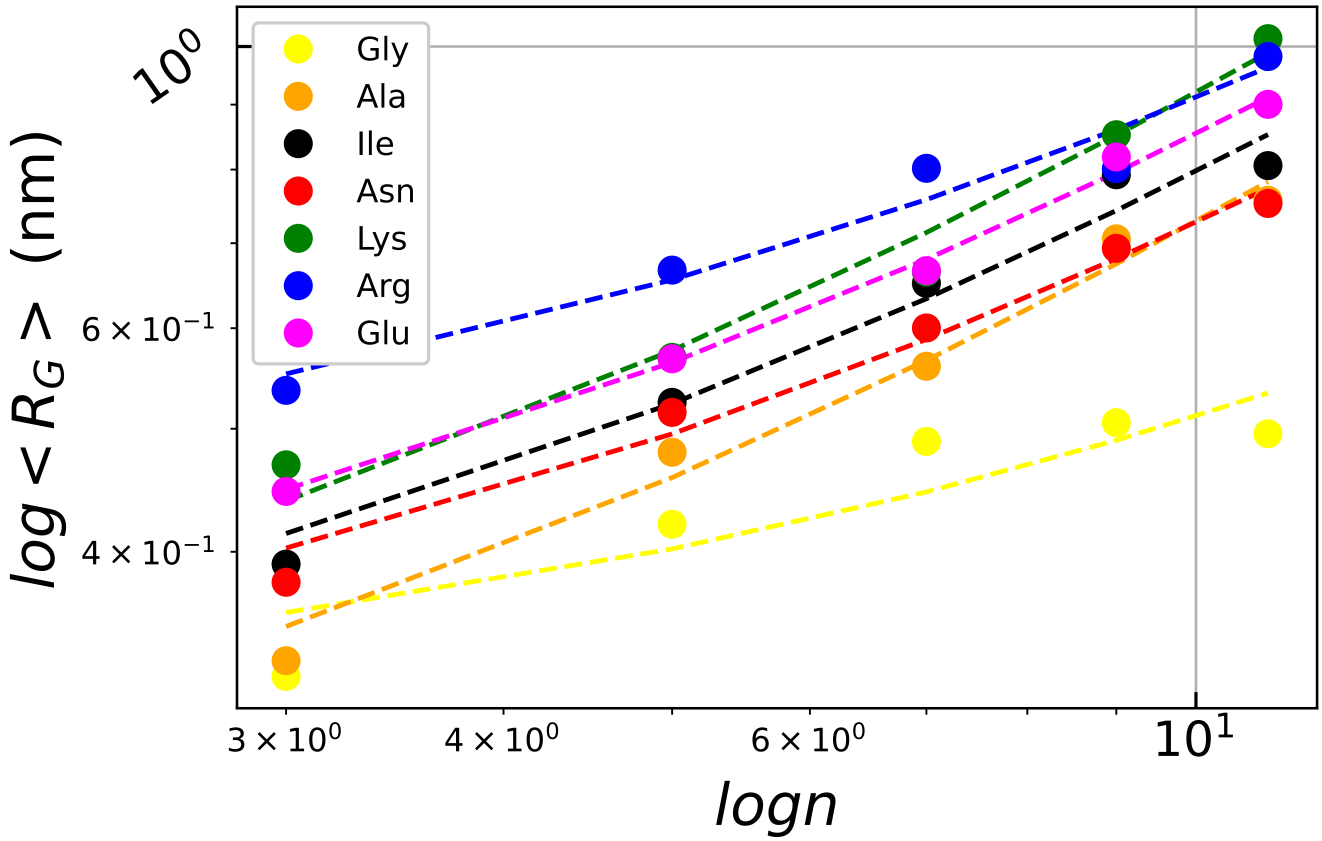


FIG. SI: Illustration of the initial simulated model using trialanine in water. The peptide is aligned to the z-axis and the carbon alpha end-points are used as centers to apply harmonic force restraints that enable maintaining the peptide extended thereby maximizing the number of solute-solvent contacts. Moreover, the polypeptides were simulated in all-atom in order to study the backbone effects on the solvation free energy. The end-to-end distances (z_{ete}) measured by employing CA end-points as references and representing the equilibrium bond lengths of the minimized structures for each polypeptide considered in this work are reported in Table SI above. Note that *all* tripeptides have sizes smaller than 1 nm whereas all other longer peptides have sizes larger that this value that is known to be a critical size in case of water [1].



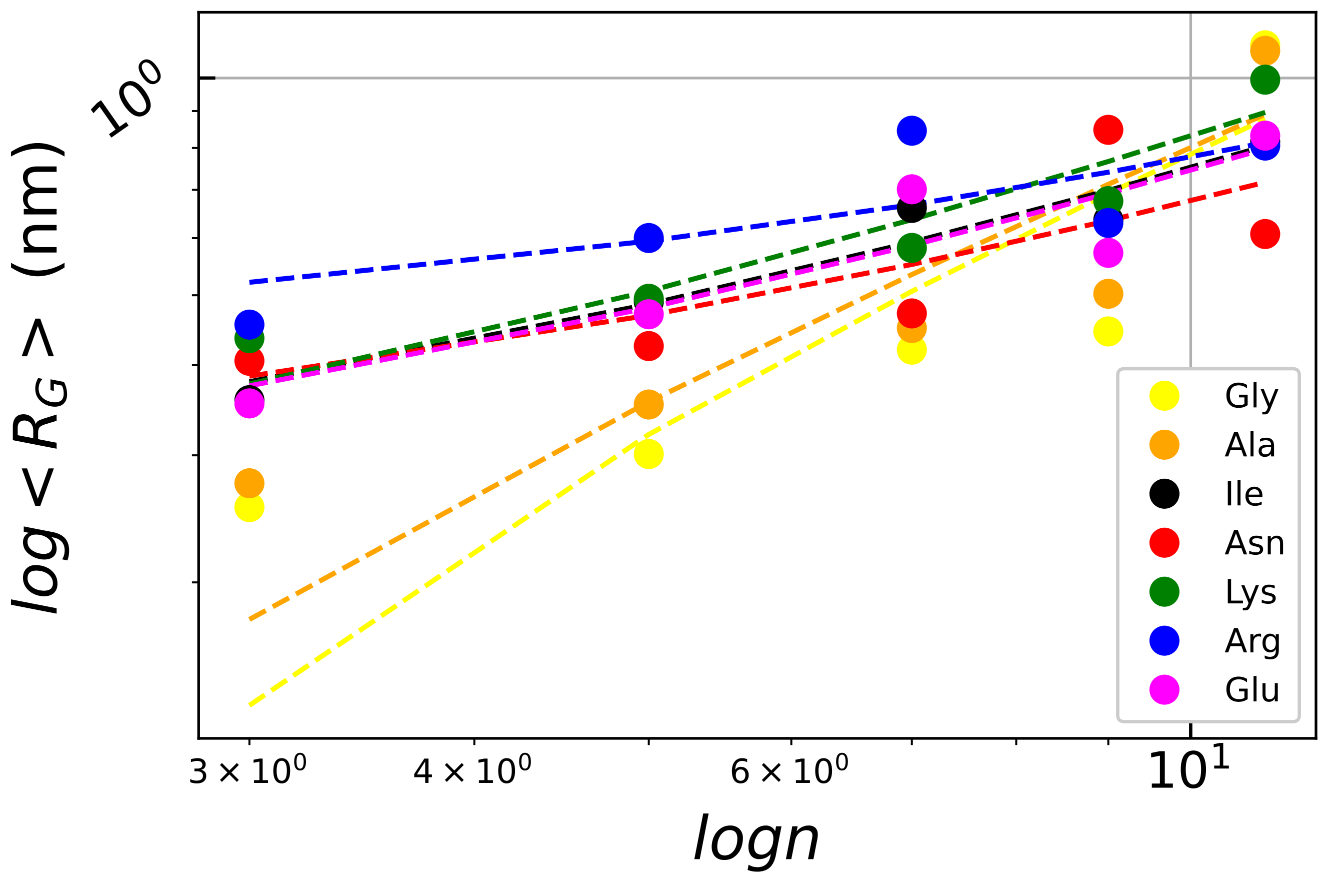


FIG. SII: Dependence of the gyration radius R_g on polypeptide chain length n in water H_2O (top) and cyclohexane cC_6H_{12} (bottom) for each of the oligomers considered in this work. The corresponding slopes, intercept to origin and correlation coefficients are shown in table SVI above.

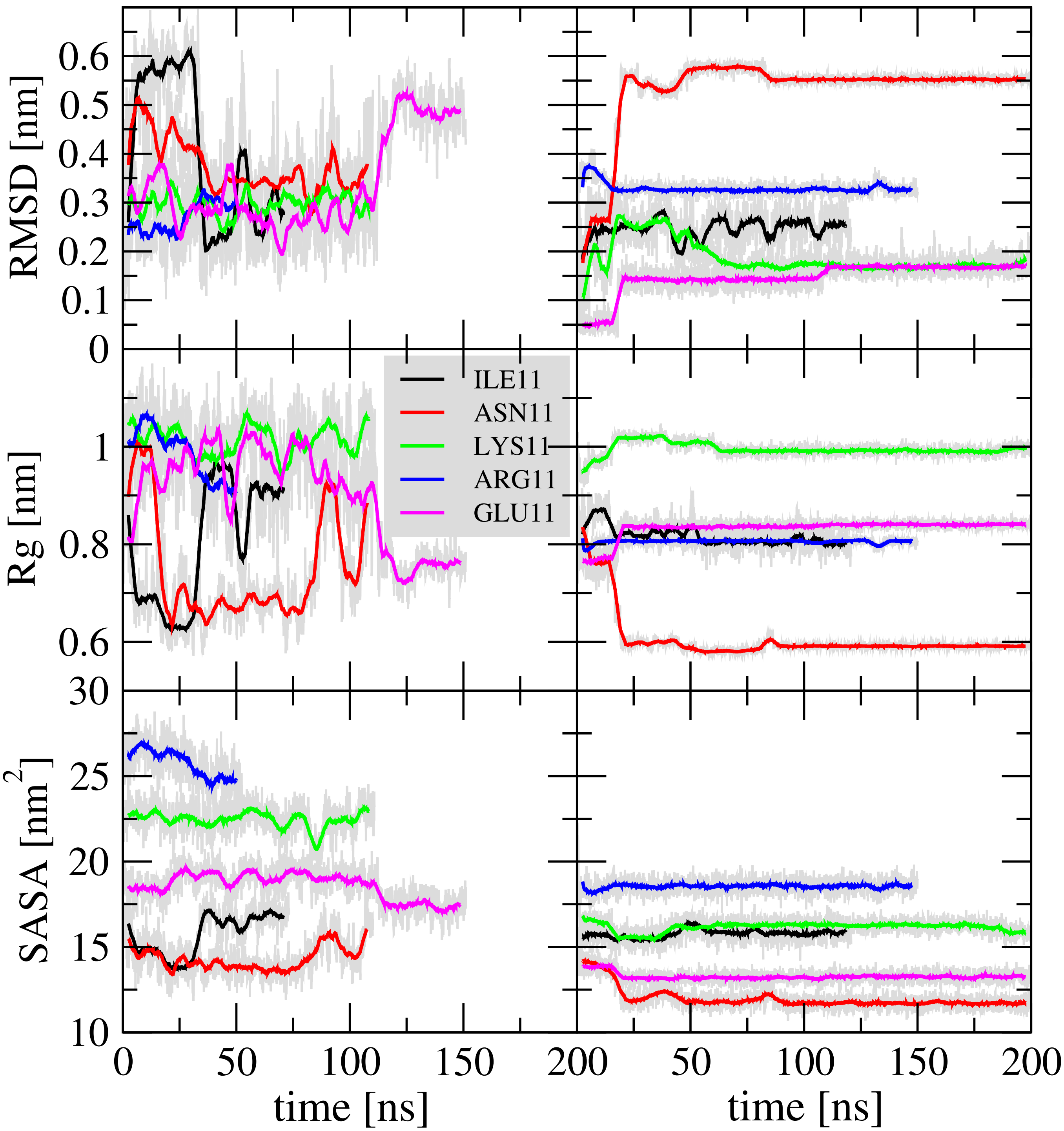


FIG. SIII: Conformational structural dynamics of the undeca-peptides ILE11, ASN11, LYS11, ARG11 and GLU11 both in water H_2O (left panel) and in cyclohexane cC_6H_{12} (right panel). Reported data are the root-mean-square-deviation (RMDS) from the initial coil conformation (top), the radius of gyration R_g (center), and the solvent accessible surface area (SASA, bottom). Full production trajectories are displayed.

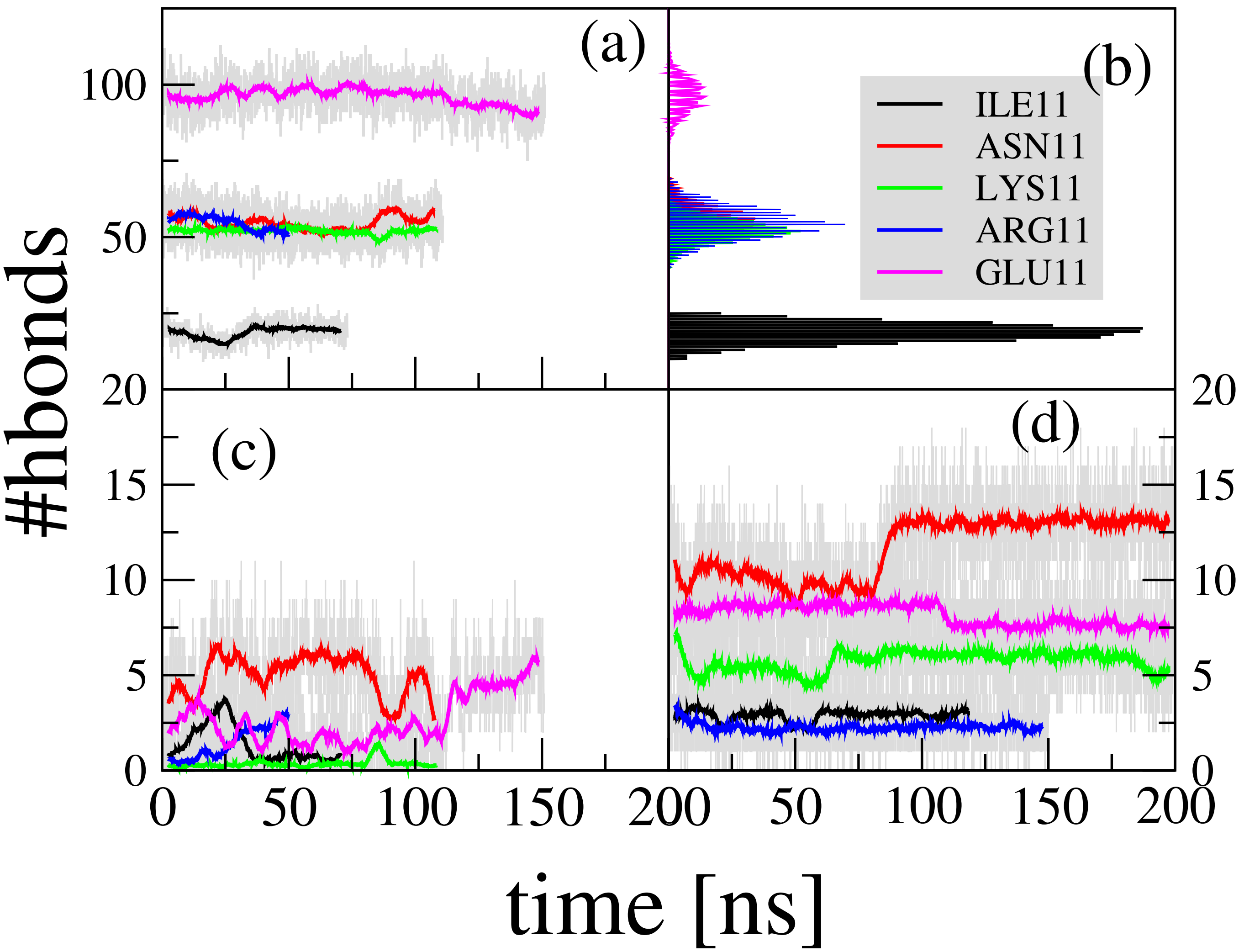


FIG. SIV: Changes in the number of hydrogen bonds along the simulation timescale for ILE11, ASN11, LYS11, ARG11 and GLU11 undeca-peptides considered in this work. The panel (a) reports on the time-based number of solute-solvent hydrogen bonds in water H_2O. On panel (b) the histogram distribution of these latters are shown. In subsets (c) and (d) the time-dependent changes in the intra-molecular (solute-solute) hydrogen bonds is reported in water H_2O and cyclohexane cC_6H_{12}, respectively.

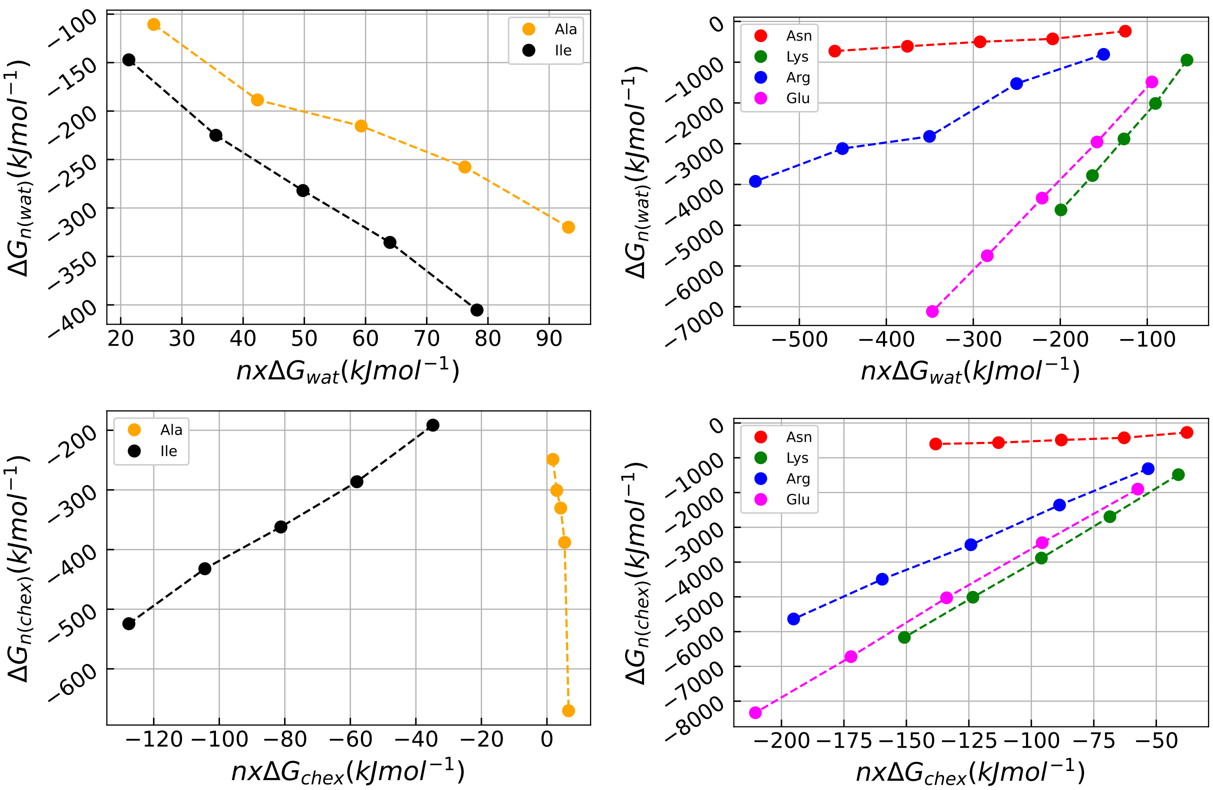


FIG. SV: Solvation free energy of the polypeptides ΔG_n as a function of the monomeric unit counterpart taken from the analog moiety $n \times \Delta G_1$ in water H_2O (top) and cyclohexane cC_6H_{12} (bottom). It is noteworthy that n varies from 3 to 11 and the comparison with GLY sidechain is missing because the latter was not computed in our previous work as GLY lacks of a proper side chain [2].

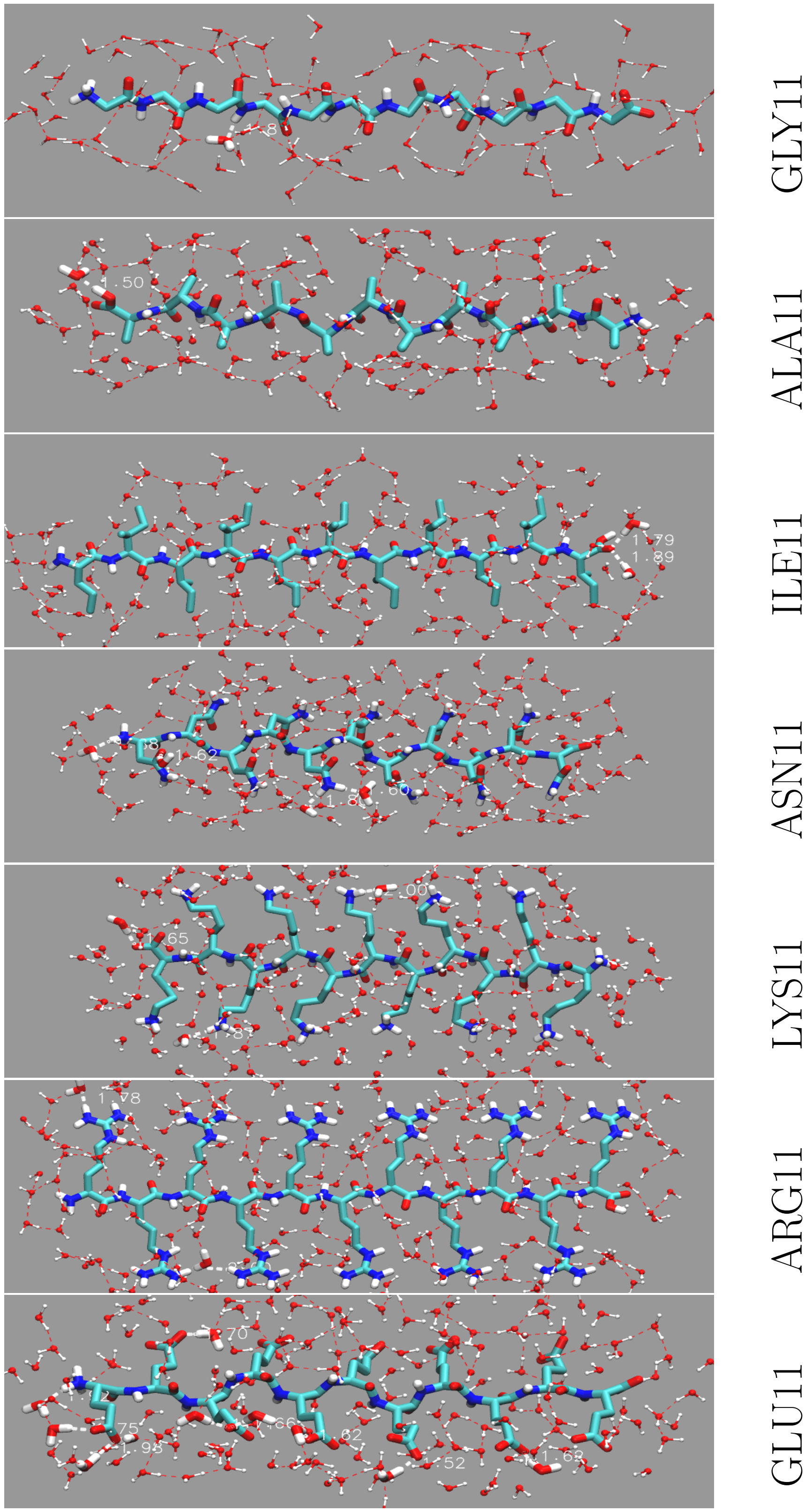


FIG. SVI: Optimized hydrated structures of each of the undecapeptides analysed in this work. From top to bottom the configurations for GLY, ALA, ILE, ASN, LYS, ARG and GLU oligomers are respectively shown. The emphasis is led on the intermolecular solute-solvent hydrogen bonds wherein the solvent molecules involved in the nonbonded interactions are explicitely shown as licorice representation.

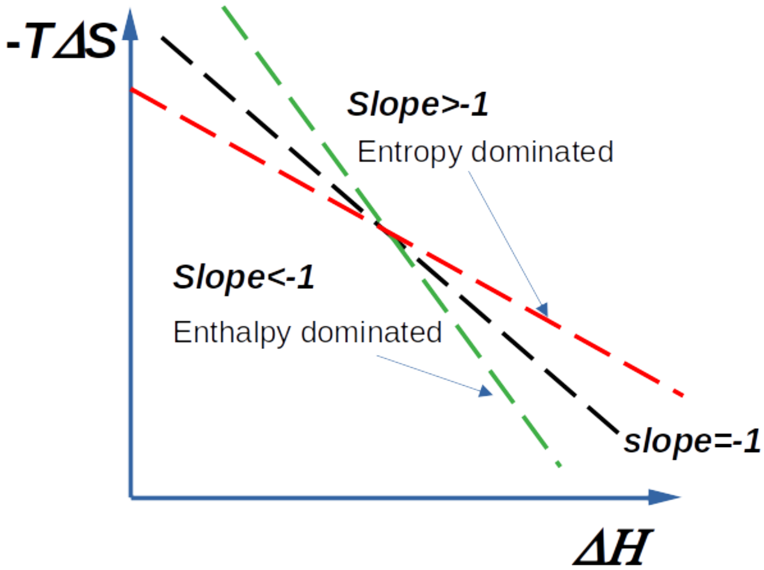


FIG. SVII: Cartoon of the entropy-enthalpy compensation. Black dashed line has slope $= -1$ and corresponds to $\Delta G = 0$ that is an exact entropy-enthalpy compensation. Slope > -1 (green dashed line) corresponds to an entropic dominated process, slope < -1 (red dashed line) corresponds to an enthalpic dominated process.

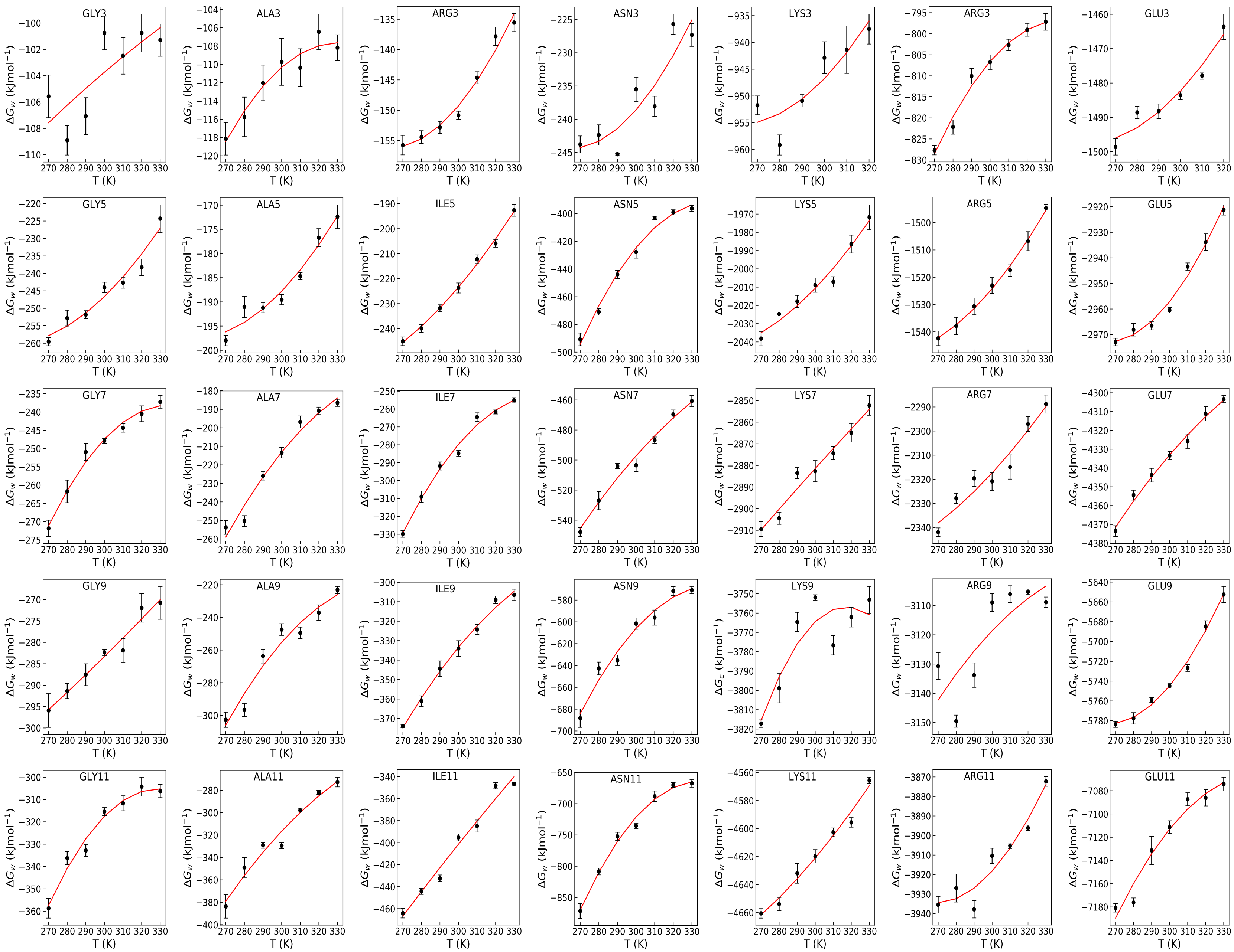


FIG. SVIII: Temperature dependence of the solvation free energy from gas to water H_2O, ΔG_w.

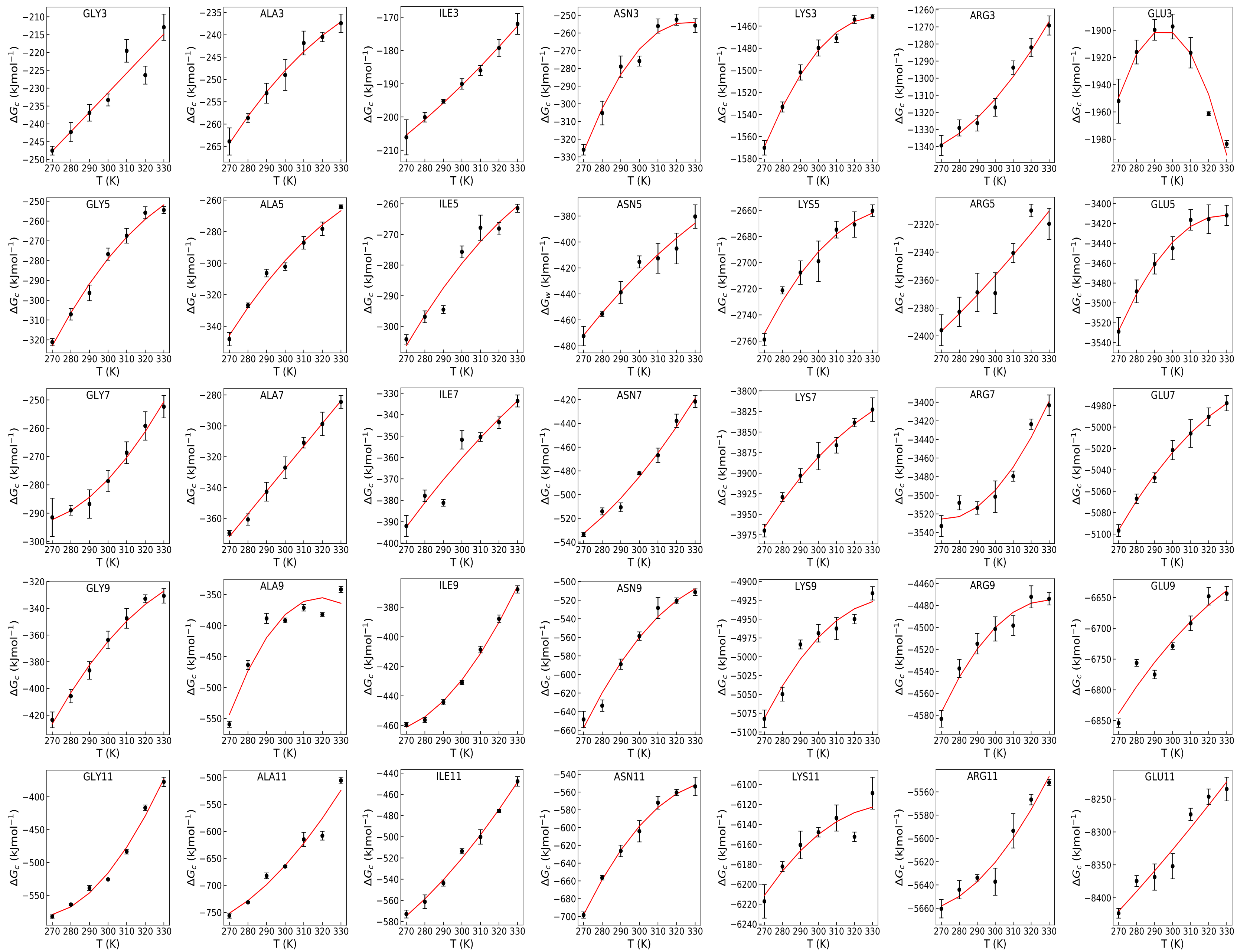


FIG. SIX: Temperature dependence of the solvation free energy from gas to cyclohexane cC_6H_{12}, ΔG_c.

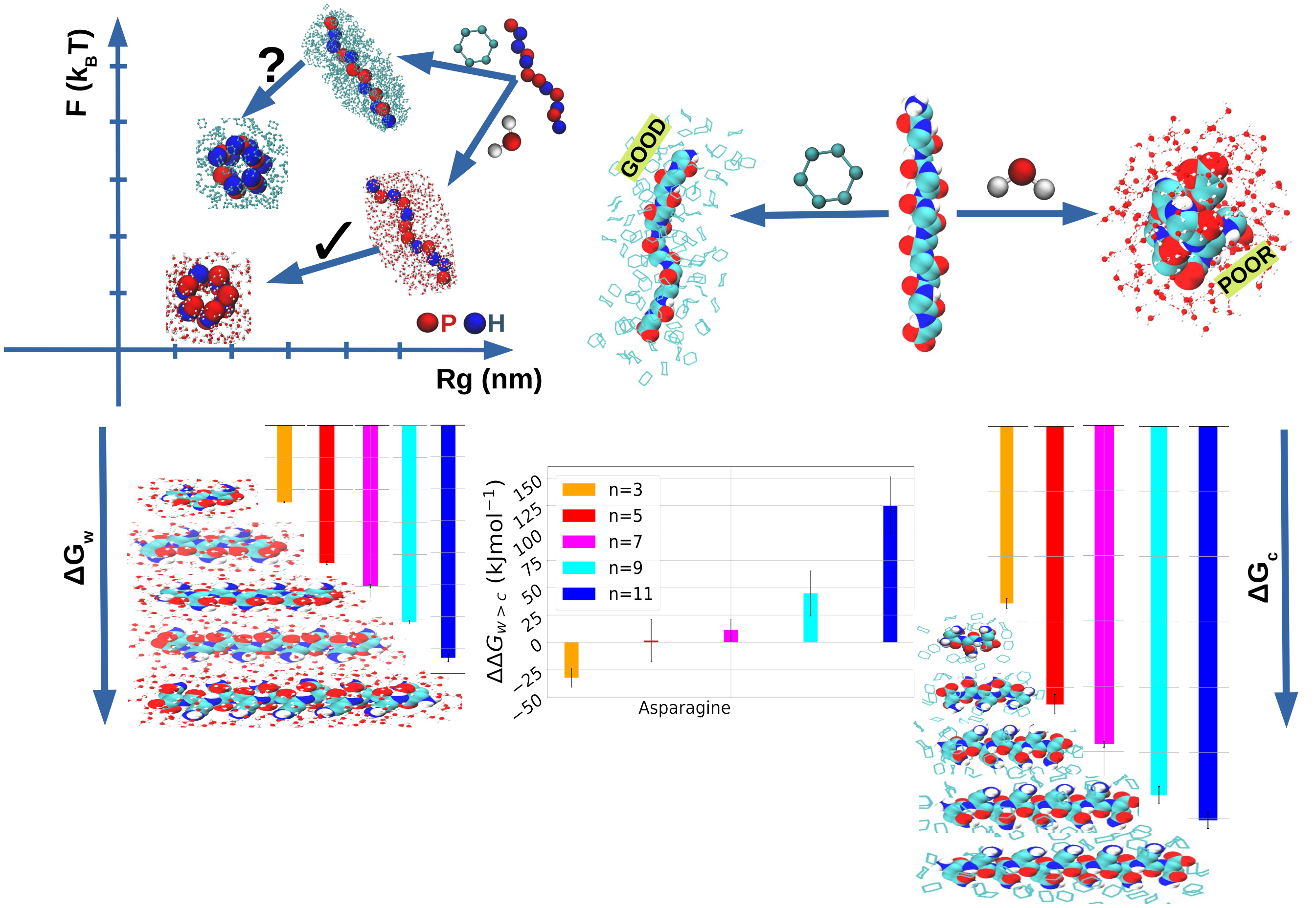


FIG. SX: Synthetic polymers assume a compact/extended conformation in a poor/good solvent. Proteins, formed by both hydrophobic and polar moieties folds in water but not in apolar solvents. This difference can be rationalized and quantified by a calculation of the solvation free energy.