

Pinpointing disulfide connectivity in cysteine-rich proteins

K. Bernardino,^a M. E. F. Pinto,^b V. S. Bolzani,^b A. F. de Moura,^a and J. M. Batista Junior^{*a}

^aDepartment of Chemistry, Federal University of São Carlos – UFSCar, São Carlos, SP 13565-905, Brazil.

^bInstitute of Chemistry, São Paulo State University – UNESP, Araraquara, SP 14800-060, Brazil.

Electronic Supplementary Information

Experimental data

Figure S1. Initial structure of CyO₄

Figure S2. Ramachandran plots for structures A-F

Figure S3. DSSP data for structures A-F

Figure S4. Distributions of C-S-S-C dihedral angles for structures A-E

Figure S5. Calculated UV spectra for structures A-E

Figure S6. Calculated ECD spectra for structures A-E

Figure S7. Free energies for the formation of S-S bonds in the cyclic cyclotide

Figure S8. Free energies for the formation of S-S bonds in the linear cyclotide

Figure S9. Experimental UV spectrum of CyO₄

Table S1. Free energies for the formation of S-S bonds in the cyclic cyclotide

Table S2. Free energies for the formation of S-S bonds in the linear cyclotide

Experimental data

Plant material

The roots of *Pombalia calceolaria* (L.) Oken were collected at Chico Mendes Natural Park, Recreio dos Bandeirantes, Rio de Janeiro in February 2013. The specimen was identified by Dr Marcelo Trovó and a voucher was deposited at Rio de Janeiro Botanical Garden with the collection number MT576 (JBRJ).

Extraction and isolation

Dried and pulverized roots (22.3 g) were extracted with 100 mL of MeOH/H₂O (60:40, v/v) at room temperature during 24 hours (4 times). The extracts were subjected to solvent-solvent partitioning with DCM:MeOH:H₂O (4 times) and the aqueous phases were concentrated under reduced pressure prior to freeze drying. The yield of this procedure is henceforth referred to as aqueous extract. The roots aqueous extracts were dissolved in ACN/H₂O (1:9, v/v) and immediately subjected to solid-phase extraction (SPE). C₁₈ SPE cartridges (Strata-Phenomenex C₁₈ 55 μm, 70 Å, 500 mg) were activated with MeOH and subsequently equilibrated with aqueous 1% formic acid. After the application of the extract, the cartridges were eluted with buffer A (0.1% aqueous TFA) in B (90% acetonitrile, 0.08% TFS), 8:2 (v/v) and 2:8 (v/v), respectively. The fraction eluted in buffer B 80%, named roots_C₁₈_80%, was considered the peptide-rich one. The fraction roots_C₁₈_80% was further purified by semi-preparative HPLC using a linear gradient from 30% to 60% buffer B in A during 60 min, yielding the pure cyclotide CyO₄ (**1**). This purified cyclotide was characterized by *de novo* peptide sequencing using, enzymatic digestion, MALDI-TOF/TOF and homology modeling using CyBase tools.

Cyclotide sequencing

The pure cyclotide **1** had its structure elucidated by MS using a Bruker Daltonics Ultraflex MALDI TOF/TOF Mass Spectrometer and a MALDI-TOF-TOF 4800 Proteomics Analyzer (Applied Biosystems). The reflector mode was adjusted in positive ion mode and 1000-2000 laser shots were acquired per spectrum. The calibration was undertaken using Bruker's Peptide Calibration Standard II (Pepcal II). For acquisition of natural masses the dried compound **1** was dissolved in 0.1% TFA and mixed at a ratio of 1:3 (v/v) with a matrix solution of saturated α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich) in ddH₂O/ACN/TFA, 50/50/0.1% (v/v/v). A 2 μL aliquot of this mixture was directly spotted onto the MALDI target plate and dried. Mass spectra were obtained in the spectral range 2500 to 4500 *m/z*. For sequencing, the reduction and alkylation of isolated peptide was performed according to the protocol previously described. The reduced and alkylated cyclotide was digested using individual endoGlu-C, trypsin, and chymotrypsin enzymes, as described previously. The peptides obtained from enzymatic digestion were desalted using C₁₈ ziptips (Millipore) and reconstituted in 10 μL of 80% (v/v). After being spotted in the plate, the MS/MS spectra were recorded from fragments coming from enzymatic digestion on a Daltonics Ultraflex mass spectrometer. The spectra were obtained with 1000-2000 laser shots and the data were acquired over the mass range *m/z* 1000-4000 Da in positive

ion reflection mode. The cyclotide amino acid sequence was obtained by manual assignment of N-terminal b-ion and C-terminal y-ion series using Bruker FlexAnalysis 3.3 software and Data Explorer® software Version 4.3.

UV and ECD Measurements

UV and ECD spectra of compound **1** were recorded with a JASCO J-815 spectrometer using the following parameters: bandwidth 1 nm; response 1 sec; scanning speed 100 nm/min; 3 accumulations; room temperature; sample in ACN/H₂O 1:1 solution (pH 5.5); 0.1 cm cell; concentration 0.2 mg/mL.

Molecular Dynamics Calculations

a- Thermodynamic integration protocol

The presence of 6 cysteine residues in the cyclotide leads, in principle, to 90 different structures that can be obtained by various combinations in the formation of disulfide bonds, since not only are the pairs involved in the bonds relevant, but also the sequence in which those bonds are formed. In order to determine the relative stability of those structures, a thermodynamic integration protocol is proposed and tested in this work to calculate the free energy differences involved in the formation of the S-S bonds in each possible permutation.

Starting from the structure obtained from www.cybase.org.au (cyclotide ID 52), all S-S bonds were removed and, in order to eliminate the correlation with this initial structure, 500 ps molecular dynamics simulations were performed at 400, 350 and 300 K. The final structure from the 300 K simulation was used as the starting point to the thermodynamic calculations (Figure S1).

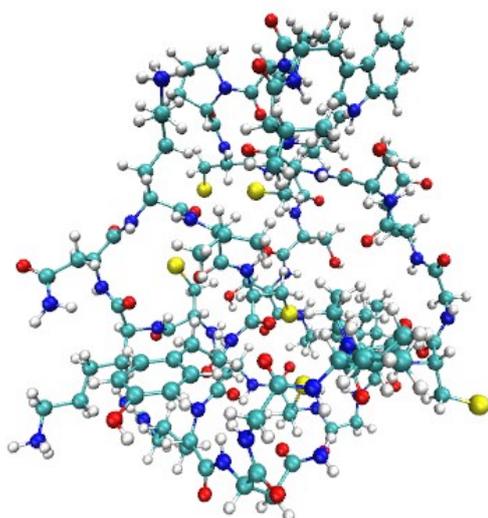


Figure S1 – Initial structure for the thermodynamic integration. Atoms: ● C, ● O, ● H, ● N, ● S.

Some considerations about the initial structure must be made before proceeding. When the disulfide bonds were broken, no hydrogen atoms were added to generate SH groups, and all the interaction parameters were held as in the disulfide ones, including sulphur partial charges. This corresponds to an unrealistic chemical species, since the cysteine residues should present either -SH or -S⁻ terminus after the disulfide bonds were broken. However, since we are interested in the relative stability of the structures with different permutations in S-S bonds, or, in other words, in the free energy difference between the final structures obtained, and those are well-defined molecules, the ill-defined initial structure is not a problem. It happens because we used the same initial structure for all the thermodynamic calculations; thus, the $\Delta\Delta G$ between the final structures is independent of the initial structure. This is shown in Equation S1, where ΔG_A and ΔG_B are the results of the thermodynamic integration for the formation of disulfide bonds in the sequence A and in the sequence B, and those values depend on the initial structure. However, the difference between them, which is the relevant quantity to determine the relative stability of A and B, are independent of the initial structure. If one computes the free energy involved in the conversion of the -SH groups to the ill-defined -S groups in the initial structure, this would only shift all the ΔG values reported by a constant, but would not change their relative stabilities.

$$\Delta\Delta G(A\rightarrow B) = \Delta G_B - \Delta G_A = (G_B - G_{initial}) - (G_A - G_{initial}) = G_B - G_A \text{ (Equation S1)}$$

Since the order in which the disulfide bonds are formed yields different final structures, and hence different free energy variations, all the possible combinations were tested by performing the thermodynamic integration to form one S-S bond at a time. There are $(6 \times 5)/2 = 15$ possibilities for the first bond, and the final structures obtained for the formation of the first bond were used as the starting point for the formation of the second bond. Once the first bonds are formed, there are $(4 \times 3)/2 = 6$ possibilities leading to the second one for each of the first 15 combinations, which results in 90 combinations. Finally, the third bond is defined by the other two since there are only 2 cysteine residues with missing bonds. This lead to a total of 195 thermodynamic integration runs in order to study all the possible sequences of S-S bond formation.

For each run, the S-S bonds are created using the slow-growth method, in which all bonded potentials involved in the formation of the disulfide bond are multiplied by a coupling parameter λ (Equation S2)^a that ranges from 0 (no bond) to 1 (complete disulfide bond). Notice that besides the force constant that defines the S-S bond perturbed by λ , the angles C-S-S and all the dihedrals that involve the S-S bond also need to be perturbed from $\lambda=0$ to $\lambda=1$.

$$V_{bond}(r_{ij}) = \lambda \frac{1}{2} k_b (r_{ij} - r_{eq})^2, V_{angle}(\theta_{ijk}) = \lambda \frac{1}{2} k_a (\theta_{ijk} - \theta_{eq})^2$$

$$V_{dihedral}(\varphi_{ijkl}) = \lambda \sum_{n=0}^5 C_n \cos^n(\varphi_{ijkl} - 180^\circ)$$

(Equation S2)

Since the bonded potential varies with the square of the distance between the atoms (Equation 2), the use of a relatively large λ in the beginning of the thermodynamic integration can lead (and led indeed) to very large forces if the pair of atoms is distant in the initial structure. Also, the greatest energy and structural variations occur before $\lambda = 0.1$ (results not shown). Because of these facts, the best procedure was not the usual linear variation of λ , but instead a finer grid starting with very small λ variations, in the order of 10^{-5} , followed by a coarser grid as λ approached 1. In the present work, the formation of each bond was accomplished using 64 different λ values, as listed below:

0.0, 0.000025, 0.00005, 0.000075, 0.0001, 0.000125, 0.00015, 0.000175, 0.0002, 0.000225, 0.00025, 0.0003, 0.00035, 0.0004, 0.00045, 0.0005, 0.00055, 0.0006, 0.00065, 0.0007, 0.0008, 0.0009, 0.001, 0.0011, 0.00125, 0.0015, 0.00175, 0.002, 0.00225, 0.0025, 0.00275, 0.003, 0.0035, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.0125, 0.015, 0.0165, 0.018, 0.020, 0.0225, 0.025, 0.03, 0.04, 0.05, 0.065, 0.08, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00

For each value of λ , an energy minimization with the L-BGFS algorithm^b was done, followed by a 250 ps simulation. The final structure from the λ_i run was used as the starting point for the λ_{i+1} . The total simulation time for the formation of each bond was hence 16 ns and for the whole ensemble of thermodynamics calculations it reached 3.2 μ s.

The free energy was obtained by the Bennett Acceptance Ratio (BAR).^c The rate of variation of the system Hamiltonian, H , with respect to the λ coordinate was computed and subsequently integrated between the initial and the final states, providing the free energy for the formation of one S-S bond (Equation S3).^{a,c}

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (\text{Equation S3})$$

50 ns equilibrium runs were carried out for the initial structure (without the S-S bonds), for the three lowest free energy structures obtained in the thermodynamic perturbation calculation, and for two of the permutations involving the bonds described in the literature as native.^d

The same protocol was used for the linear cyclotide. Starting from the initial structure from the thermodynamic calculations for the cyclic cyclotide (Figure S1), it was cleaved between the asparagine and glycine residues with glycine becoming the N-terminal and asparagine the C-terminal in the ionized forms. This structure was minimized using steepest descent and L-BGFS algorithms and then simulations of 500 ps at 400, 350 and 300 K were also performed in order to produce the initial structure for the free energy calculations.

b- Molecular dynamics parameters and simulation conditions

All the simulations were performed using GROMACS 5.1^{a,e,f} package, which also includes many analysis tools used in this work. The solvation was described by GBSA implicit model with a relative dielectric constant of 80 to simulate aqueous solution and no counter-ions were added. The stochastic dynamics algorithm was used to simulate the effect of random collisions between the solvent and the cyclotide. OPLS-AA^g parameters were used to describe the cyclotide and, since an implicit solvation model was used, no pressure coupling or periodic boundary conditions were applied and infinite cutoff radius were used for both Coulomb and Lennard-Jones potentials (meaning that all pair interactions were taken into account in all simulations).

The Velocity Rescaling thermostat^h was employed to keep the temperature with $\tau_T=0.1$ ps and $T_{ref}=300$ K in all the simulations except for the initial relaxation described in the previous section, where $T_{ref}=350$ K and $T_{ref}=400$ K were also employed. The integration timestep was of 1 fs in all simulations.

The secondary structures elements were analysed using the DSSP programⁱ interfaced with the GROMACS suite.

UV and ECD Calculations

Single Point DFT calculations at the CAM-B3LYP/6-311G level^l were performed for structures extracted from the 50 ns equilibrium simulations of the three most stable molecules from the thermodynamic calculation for the cyclic cyclotide (13, 58, 65, Table S1) as well as for two of the permutations of the bonds described in the literature as native (14, 44, Table S1), It was performed after energy minimization with GROMACS using the classical force-field, so as to avoid large structural distortions which might result. The simplified time-dependent DFT (sTD-DFT)^{j,k} was employed to calculate the excitation energy, oscillator strength and rotatory strength R in the dipole velocity gauge. The calculated rotatory strengths were used to simulate a convoluted ECD curve using Gaussian broadening with 16 nm FWHM. The predicted wavelength transitions were bathochromically shifted by 20 nm.

The structures were extracted from the classical simulations at every 2.5 ns starting with the structure at 10 ns, which results in 17 DFT calculations for each combination of S-S bonds considered. However, due to numerical problems in the matrix diagonalization in the sTD-DFT step, some of those calculations were not successfully completed even with tighter convergence criteria for the SCF (energy tolerance of 10^{-8} hartree). This problem was more pronounced for the Cys^{III}-Cys^{IV}, Cys^{II}-Cys^V, Cys^I-Cys^{VI} molecule, where only 5 out of 17 calculations were completed (Figures S5-A and S6-A). Due to the poor sampling for this structure, the corresponding spectra should be analyzed with caution. For the Cys^I-Cys^{IV}, Cys^{II}-Cys^V, Cys^{III}-Cys^{VI} 13 calculations were completed and used in the average

calculations while for each one of the other 3 combinations 16 out of the 17 calculations were successful.

- a M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, GROMACS User Manual version 5.1, www.gromacs.org (2015)
- b Zhu, C., Byrd, R. H. and Nocedal, J., *ACM Trans. Math. Softw.* 1997, **23**, 550–560.
- c C. H. E. Bennett, *J. Comp. Phys.* 1976, **22**, 245–268.
- d O. Saether, D. J. Craik, I. D. Campbell, K. Sletten, J. Juul and D. G. Norman, *Biochemistry*, 1995, **34**, 4147.
- e H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comp. Phys. Comm.* 1995, **91**, 43–56.
- f S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics*, 2013, **29**(7):845–854.
- g W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.
- h G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* 2007, **126**, 014101.
- i W. Kabsch, C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
- j Y. T. Risthaus, A. Hansen, S. Grimme, *Phys. Chem. Chem. Phys.* 2014, **16**, 14408.
- k C. Bannwarth and S. Grimme, *Comp. Theor. Chem.* 2014, **1040-1041**, 45.
- l G. Pescitelli and T. Bruhn, *Chirality*, 2016, **28**, 466-474.

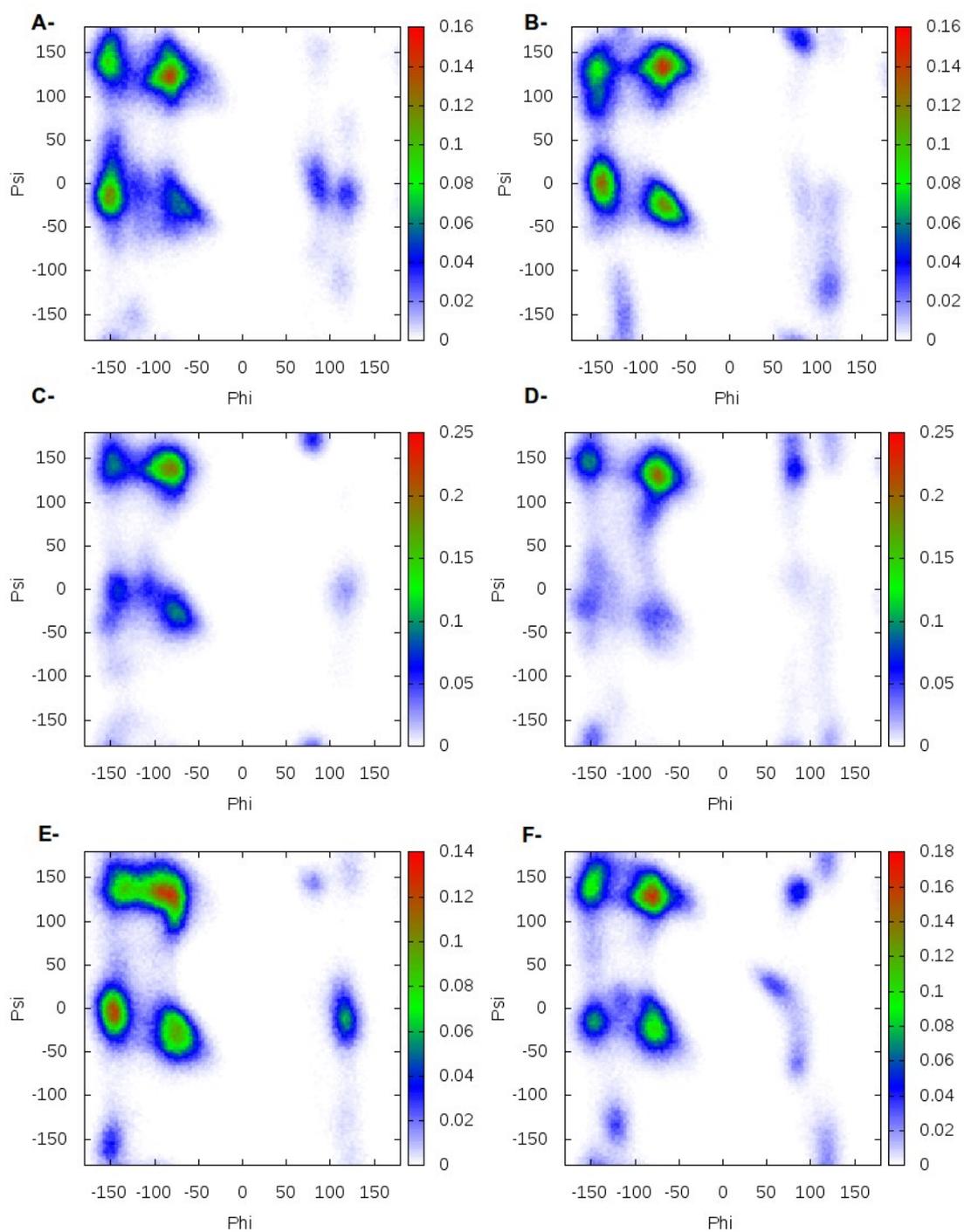


Figure S2 – Ramachandran plots for the three structures with the smallest free energy in the thermodynamics integration: Cys^{III}-Cys^{IV}, Cys^{II}-Cys^V, Cys^I-Cys^{VI} (A); Cys^{III}-Cys^V, Cys^{II}-Cys^{VI}, Cys^I-Cys^{IV} (B), and Cys^I-Cys^{IV}, Cys^{II}-Cys^{III}, Cys^V-Cys^{VI} (C); for two of the permutations of the bonds described in the literature as native: Cys^I-Cys^{IV}, Cys^{II}-Cys^V, Cys^{III}-Cys^{VI} (D); Cys^{II}-Cys^V, Cys^I-Cys^{IV}, Cys^{III}-Cys^{VI} (E), and for the structure without any S-S bond (F).

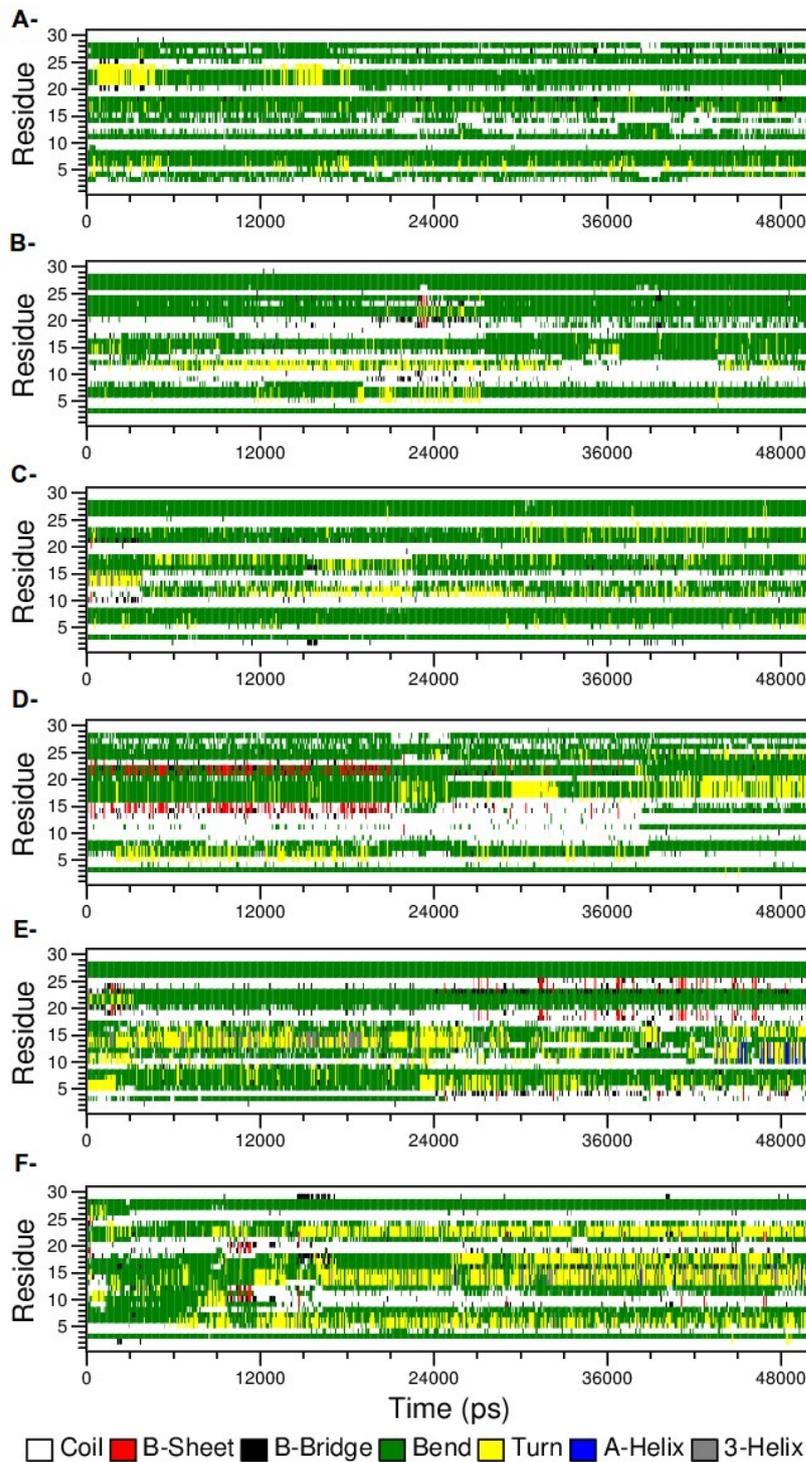


Figure S3 – DSSP data for the three structures with the smallest free energy in the thermodynamics integration: Cys^{III}-Cys^{IV}, Cys^{II}-Cys^V, Cys^I-Cys^{VI} (A); Cys^{III}-Cys^V, Cys^{II}-Cys^{VI}, Cys^I-Cys^{IV} (B), and Cys^I-Cys^{IV}, Cys^{II}-Cys^{III}, Cys^V-Cys^{VI} (C); for two of the permutations of the bonds described in the literature as native: Cys^I-Cys^{IV}, Cys^{II}-Cys^V, Cys^{III}-Cys^{VI} (D); Cys^{II}-Cys^V, Cys^I-Cys^{IV}, Cys^{III}-Cys^{VI} (E), and for the structure without any S-S bond (F).

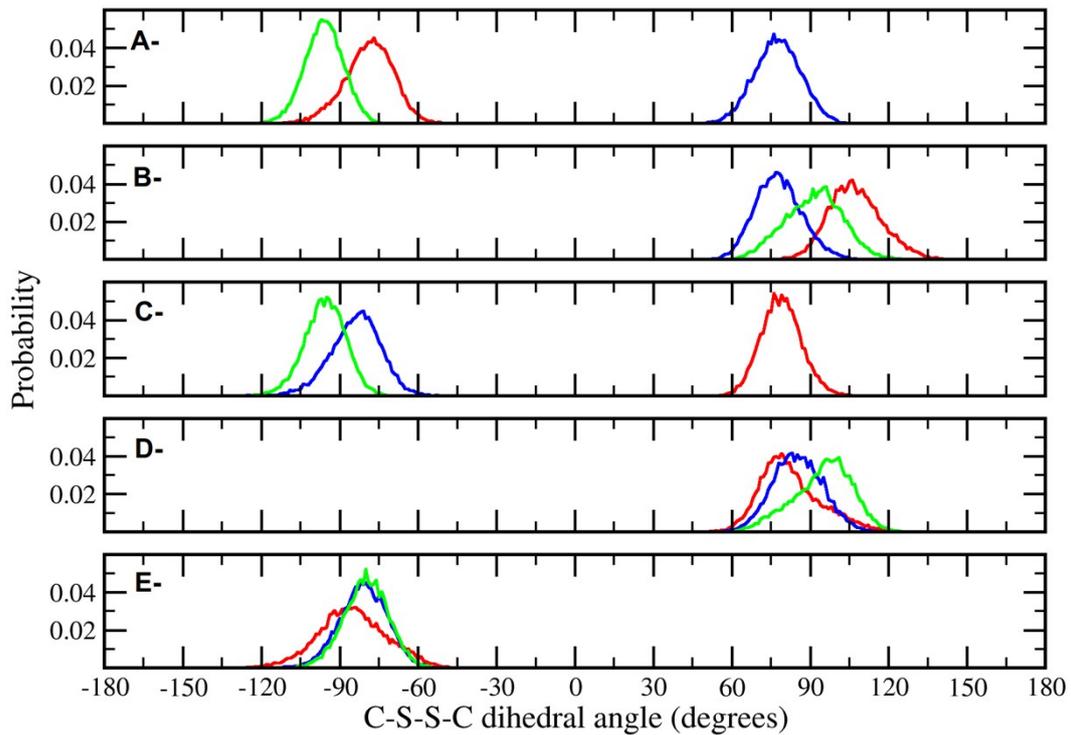


Figure S4 – Distributions of C-S-S-C dihedral angles (χ_3) for the three structures with the smallest free energy in the thermodynamics integration: Cys^{III}-Cys^{IV}, Cys^{II}-Cys^V, Cys^I-Cys^{VI} (**A**); Cys^{III}-Cys^V, Cys^{II}-Cys^{VI}, Cys^I-Cys^{IV} (**B**), and Cys^I-Cys^{IV}, Cys^{II}-Cys^{III}, Cys^V-Cys^{VI} (**C**); and for the two of the permutations of bonds described in the literature as native: Cys^I-Cys^{IV}, Cys^{II}-Cys^V, Cys^{III}-Cys^{VI} (**D**) and Cys^{II}-Cys^V, Cys^I-Cys^{IV}, Cys^{III}-Cys^{VI} (**E**). The different colors stand for the order in which the S-S bonds were created: (—) first S-S bond, (—) second S-S bond and (—) third S-S bond, as indicated in the name of the corresponding structures.

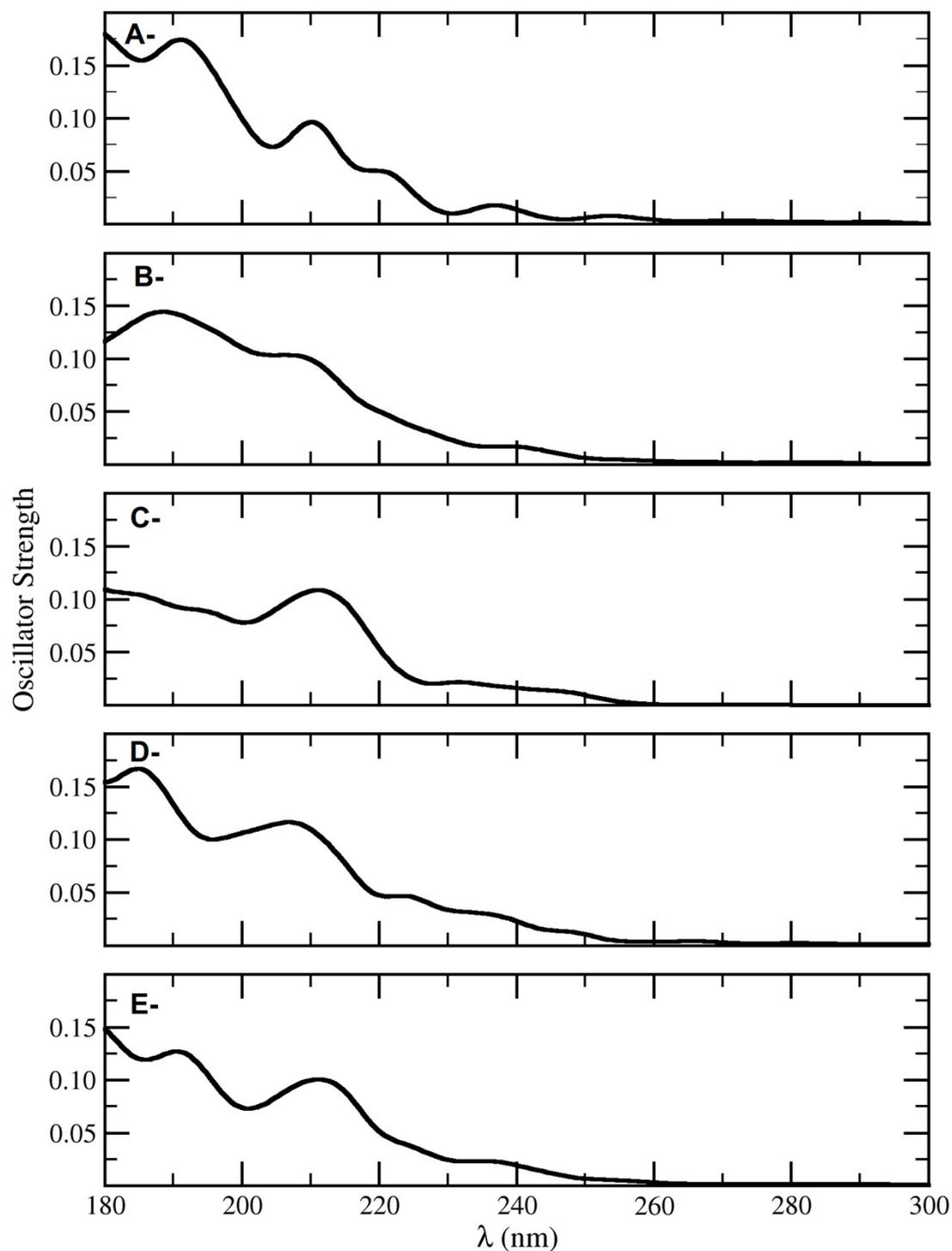


Figure S5 – Calculated (sTD-CAM-B3LYP/6-311G) UV spectra for: Cys^{III}-Cys^{IV}, Cys^{II}-Cys^V, Cys^I-Cys^{VI} (A); Cys^{III}-Cys^V, Cys^{II}-Cys^{VI}, Cys^I-Cys^{IV} (B), and Cys^I-Cys^{IV}, Cys^{II}-Cys^{III}, Cys^V-Cys^{VI} (C); Cys^I-Cys^{IV}, Cys^{II}-Cys^V, Cys^{III}-Cys^{VI} (D) and Cys^{II}-Cys^V, Cys^I-Cys^{IV}, Cys^{III}-Cys^{VI} (E). The predicted wavelength transitions were multiplied by a scaling factor of 0.91 to match experimental UV data.

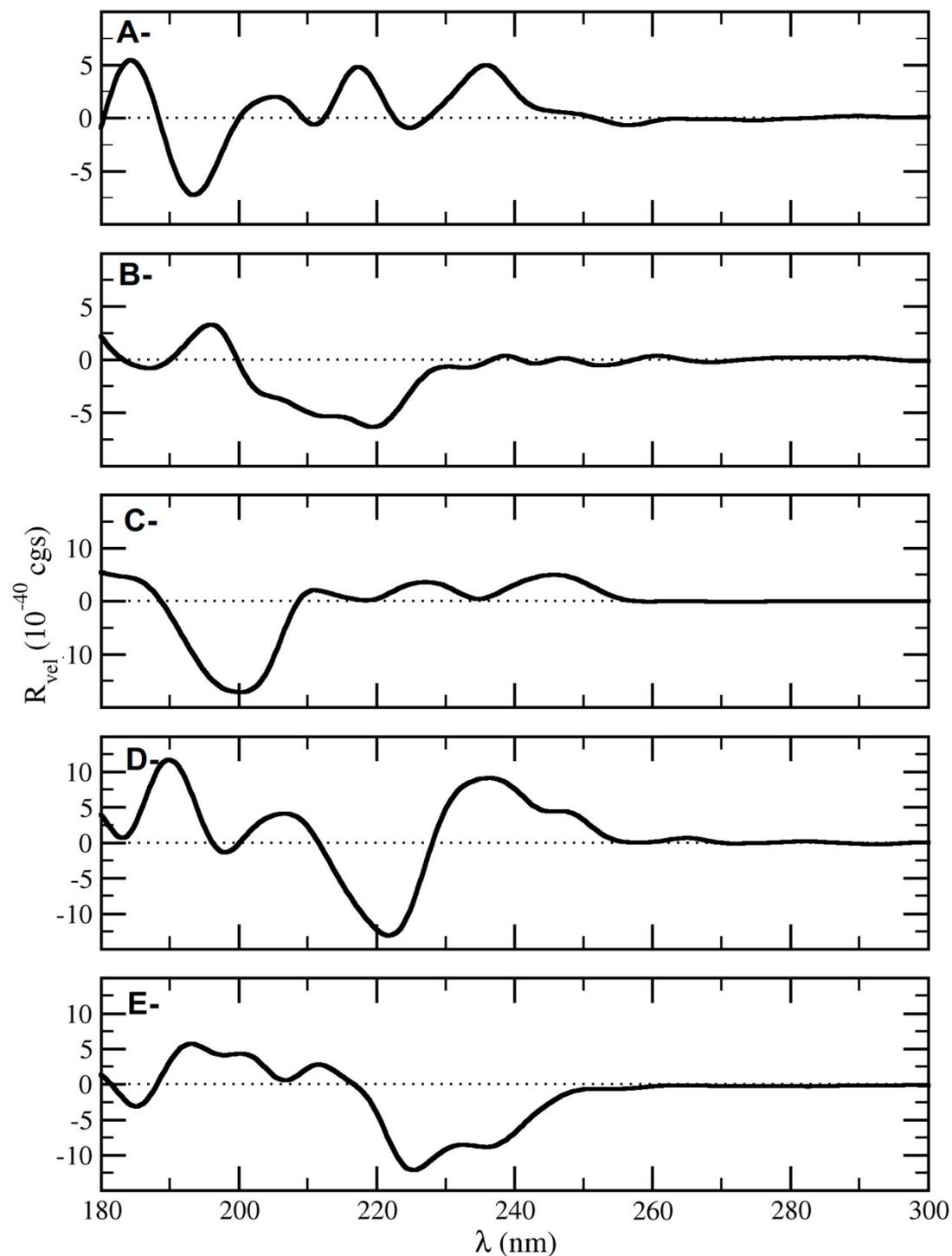


Figure S6 – Calculated (sTD-CAM-B3LYP/6-311G) ECD spectra for: Cys^{III}-Cys^{IV}, Cys^{II}-Cys^V, Cys^I-Cys^{VI} (**A**); Cys^{III}-Cys^V, Cys^{II}-Cys^{VI}, Cys^I-Cys^{IV} (**B**), and Cys^I-Cys^{IV}, Cys^{II}-Cys^{III}, Cys^V-Cys^{VI} (**C**); Cys^I-Cys^{IV}, Cys^{II}-Cys^V, Cys^{III}-Cys^{VI} (**D**) and Cys^{II}-Cys^V, Cys^I-Cys^{IV}, Cys^{III}-Cys^{VI} (**E**). The predicted wavelength transitions were multiplied by a scaling factor of 0.91.

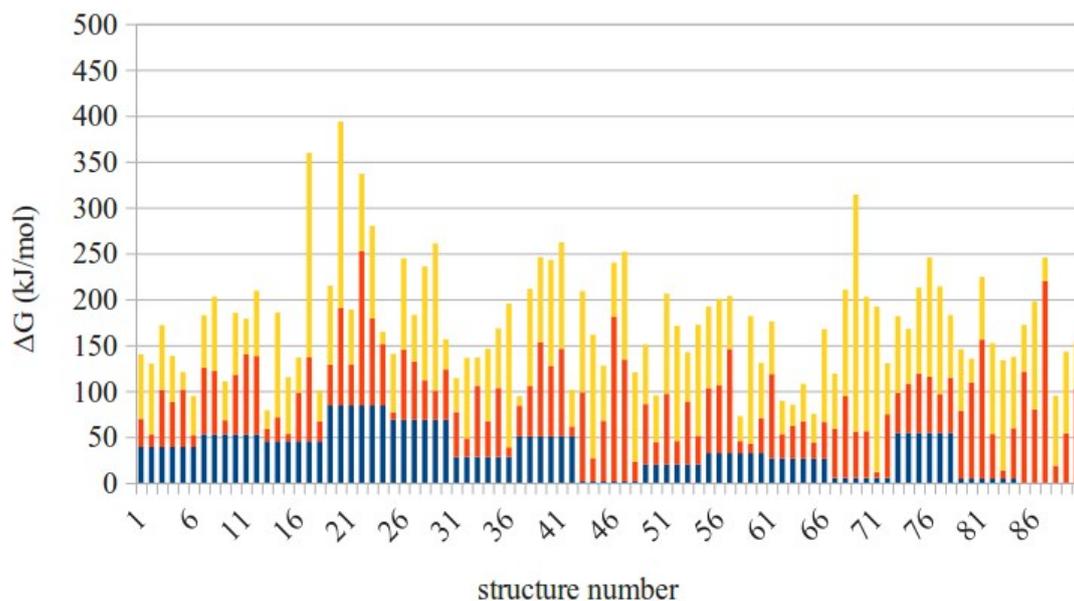


Figure S7 – Free energies for the formation of S-S bonds in each possible combination for the cyclic cyclotide. The blue bars correspond to the variation involved in the first bond, the orange for the second and the yellow ones for the third. The total height corresponds to the total ΔG from the initial to the final structure. The sequence of bond formation for each structure number is given in Table S1.

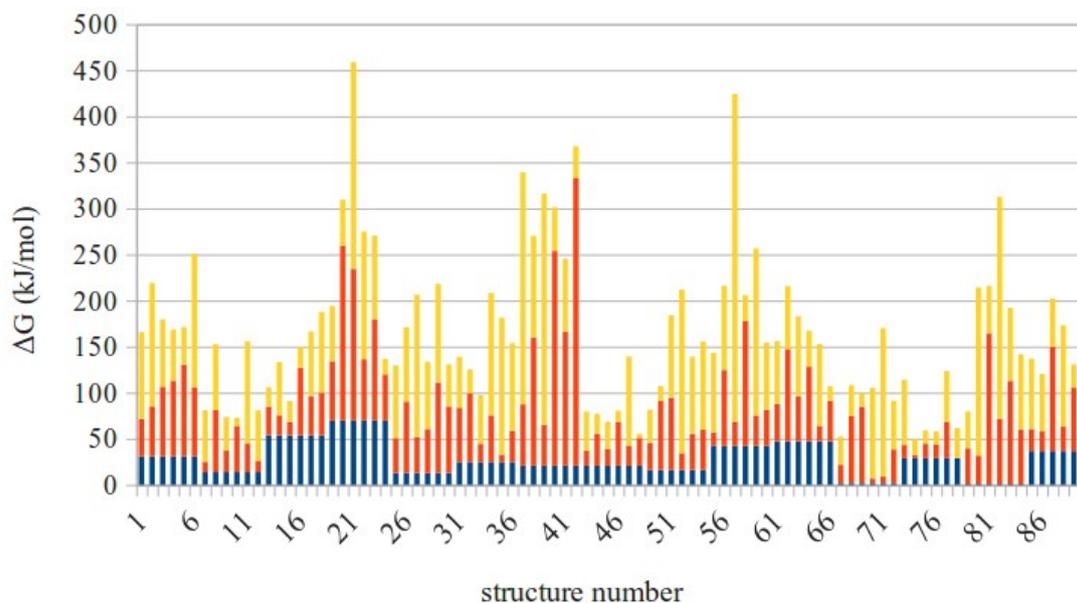


Figure S8 – Free energies for the formation of S-S bonds in each possible combination for the linear cyclotide. The blue bars correspond to the variation involved in the first bond, the orange for the second and the yellow ones for the third. The total height corresponds to the total ΔG from the initial to the final

structure. The sequence of bond formation for each structure number is given in Table S2.

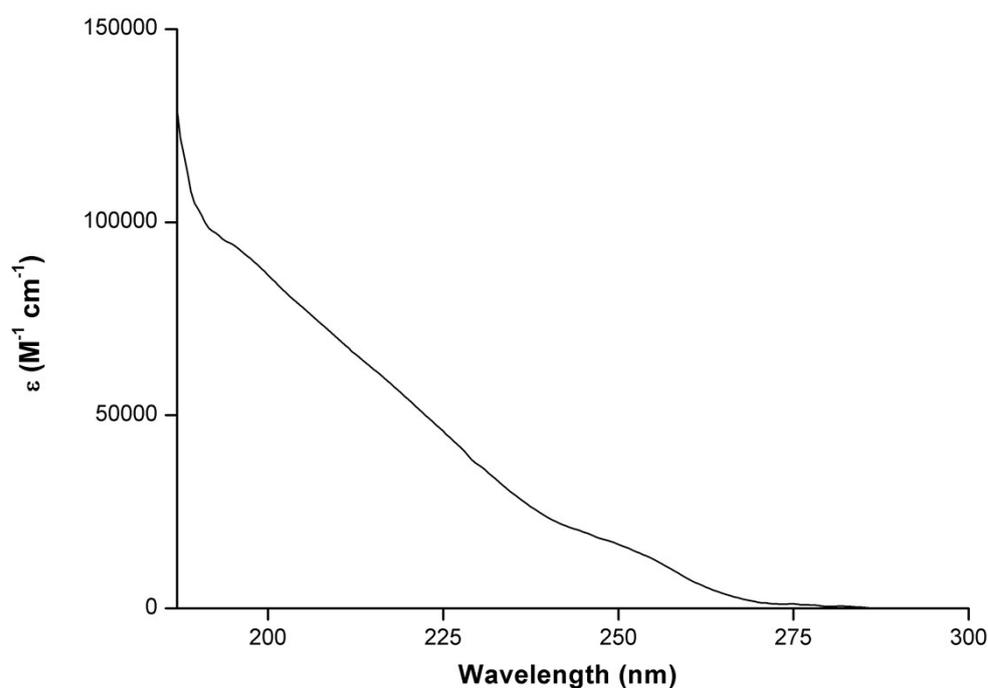


Figure S9 – Experimental UV spectrum of CyO₄ in water/acetonitrile 1:1.

Table S1 – Free energies for the formation of S-S bonds for the cyclic cyclotide and the respective standard deviations. G_1 , G_2 , and G_3 stand for the free energy differences involved in the formation of the first, second and the third S-S bonds. Their sum, G_{tot} , represents the free energy difference involved in the formation of all S-S bonds in the given order, where a_1 and b_1 stands for the cysteine residues involved in the first bond and a_2 and b_2 for the ones of the second bond. The three structures with the lowest-energies chosen for the structural analyses and ECD calculations are highlighted in bold. Free energy values are in kJ/mol.

n	a_1	b_1	G_1	+-	a_2	b_2	G_2	+-	G_3	+-	G_{tot}	+-
1	I	II	39.01	0.22	III	IV	30.55	0.60	70.32	0.89	139.88	1.10
2	I	II	39.01	0.22	III	V	13.91	0.14	76.87	0.53	129.79	0.59
3	I	II	39.01	0.22	III	VI	62.43	0.60	70.34	0.66	171.78	0.92
4	I	II	39.01	0.22	IV	V	48.79	0.43	50.38	0.58	138.18	0.75

5	I	II	39.01	0.22	IV	VI	62.50	0.57	19.15	0.16	120.66	0.63
6	I	II	39.01	0.22	V	VI	12.70	0.31	42.89	0.39	94.60	0.54
7	I	III	52.83	0.72	II	IV	72.54	0.66	57.22	0.25	182.59	1.01
8	I	III	52.83	0.72	II	V	69.02	0.44	80.90	1.54	202.75	1.76
9	I	III	52.83	0.72	II	VI	15.19	0.32	42.38	0.38	110.40	0.87
10	I	III	52.83	0.72	IV	V	65.15	1.69	67.14	0.33	185.12	1.87
11	I	III	52.83	0.72	IV	VI	87.61	1.67	38.22	0.41	178.66	1.86
12	I	III	52.83	0.72	V	VI	85.18	1.27	71.29	0.87	209.30	1.70
13	I	IV	45.04	0.74	II	III	13.68	0.65	20.31	0.51	79.03	1.11
14	I	IV	45.04	0.74	II	V	26.25	0.41	114.16	0.44	185.45	0.95
15	I	IV	45.04	0.74	II	VI	8.29	0.19	61.83	0.96	115.16	1.23
16	I	IV	45.04	0.74	III	V	53.17	0.70	38.21	0.51	136.42	1.14
17	I	IV	45.04	0.74	III	VI	91.73	0.83	222.77	3.76	359.54	3.92
18	I	IV	45.04	0.74	V	VI	21.83	0.34	33.57	0.75	100.44	1.11
19	I	V	84.44	1.09	II	III	44.22	1.33	86.10	0.83	214.76	1.91
20	I	V	84.44	1.09	II	IV	106.45	0.44	202.67	1.19	393.56	1.67
21	I	V	84.44	1.09	II	VI	44.76	0.51	59.59	0.79	188.79	1.44
22	I	V	84.44	1.09	III	IV	168.21	0.86	83.94	1.59	336.59	2.11
23	I	V	84.44	1.09	III	VI	94.50	0.71	101.02	0.91	279.96	1.59
24	I	V	84.44	1.09	IV	VI	66.61	0.74	13.61	0.38	164.66	1.37
25	I	VI	68.53	0.39	II	III	8.49	0.38	63.47	0.45	140.49	0.71
26	I	VI	68.53	0.39	II	IV	76.82	1.65	99.25	0.94	244.60	1.94
27	I	VI	68.53	0.39	II	V	63.87	0.98	50.65	0.78	183.05	1.31
28	I	VI	68.53	0.39	III	IV	43.23	1.17	123.90	0.46	235.66	1.32
29	I	VI	68.53	0.39	III	V	31.80	0.29	160.50	2.25	260.83	2.30
30	I	VI	68.53	0.39	IV	V	54.83	0.43	32.88	0.87	156.24	1.05
31	II	III	28.15	0.48	I	IV	48.96	0.74	36.97	0.60	114.08	1.07
32	II	III	28.15	0.48	I	V	20.19	0.31	87.65	1.42	135.99	1.53
33	II	III	28.15	0.48	I	VI	77.27	0.20	31.09	0.31	136.51	0.61
34	II	III	28.15	0.48	IV	V	38.93	0.43	79.05	0.85	146.13	1.07
35	II	III	28.15	0.48	IV	VI	75.22	0.90	64.92	1.77	168.29	2.04
36	II	III	28.15	0.48	V	VI	10.42	0.17	156.59	1.08	195.16	1.19
37	II	IV	50.49	0.69	I	III	33.57	0.89	10.27	0.22	94.33	1.15
38	II	IV	50.49	0.69	I	V	55.04	0.81	105.87	0.94	211.40	1.42
39	II	IV	50.49	0.69	I	VI	102.48	0.81	92.78	1.37	245.75	1.73

40	II	IV	50.49	0.69	III	V	76.95	0.60	115.27	1.03	242.71	1.38
41	II	IV	50.49	0.69	III	VI	95.57	1.06	116.11	1.53	262.17	1.99
42	II	IV	50.49	0.69	V	VI	10.87	0.49	40.03	0.65	101.39	1.07
43	II	V	1.51	0.22	I	III	97.05	0.82	110.34	1.56	208.90	1.78
44	II	V	1.51	0.22	I	IV	25.44	0.27	134.55	1.17	161.50	1.22
45	II	V	1.51	0.22	I	VI	66.20	0.82	59.91	0.49	127.62	0.98
46	II	V	1.51	0.22	III	IV	179.68	1.77	58.59	1.16	239.78	2.13
47	II	V	1.51	0.22	III	VI	132.95	3.37	117.32	0.97	251.78	3.51
48	II	V	1.51	0.22	IV	VI	21.51	0.34	97.21	1.59	120.23	1.64
49	II	VI	20.04	0.27	I	III	65.77	0.27	64.73	0.27	150.54	0.47
50	II	VI	20.04	0.27	I	IV	24.65	0.38	50.45	0.20	95.14	0.51
51	II	VI	20.04	0.27	I	V	77.05	0.45	109.09	0.87	206.18	1.02
52	II	VI	20.04	0.27	III	IV	25.16	0.42	125.83	1.75	171.03	1.82
53	II	VI	20.04	0.27	III	V	68.47	0.80	53.97	0.45	142.48	0.96
54	II	VI	20.04	0.27	IV	V	31.32	0.36	120.89	3.11	172.25	3.14
55	III	IV	32.34	0.47	I	II	71.02	0.28	88.72	0.48	192.08	0.73
56	III	IV	32.34	0.47	I	V	74.35	0.75	93.75	0.48	200.44	1.01
57	III	IV	32.34	0.47	I	VI	113.49	1.14	57.88	0.15	203.71	1.24
58	III	IV	32.34	0.47	II	V	12.93	0.33	27.24	0.50	72.51	0.76
59	III	IV	32.34	0.47	II	VI	10.12	0.33	139.05	1.94	181.51	2.02
60	III	IV	32.34	0.47	V	VI	38.26	0.43	60.04	0.81	130.64	1.03
61	III	V	26.37	0.34	I	II	92.02	0.82	57.41	1.15	175.80	1.45
62	III	V	26.37	0.34	I	IV	26.68	0.41	36.22	0.61	89.27	0.81
63	III	V	26.37	0.34	I	VI	35.32	0.27	23.34	0.42	85.03	0.60
64	III	V	26.37	0.34	II	IV	40.64	1.00	40.60	0.84	107.61	1.35
65	III	V	26.37	0.34	II	VI	17.32	0.34	31.37	0.90	75.06	1.02
66	III	V	26.37	0.34	IV	VI	39.88	0.77	101.11	0.45	167.36	0.95
67	III	VI	5.42	0.30	I	II	53.63	0.56	59.89	0.66	118.94	0.92
68	III	VI	5.42	0.30	I	IV	88.97	0.50	116.09	1.40	210.48	1.52
69	III	VI	5.42	0.30	I	V	50.25	0.71	258.38	1.24	314.05	1.46
70	III	VI	5.42	0.30	II	IV	50.88	0.96	146.36	2.56	202.66	2.75
71	III	VI	5.42	0.30	II	V	6.27	0.51	180.30	1.09	191.99	1.24
72	III	VI	5.42	0.30	IV	V	69.12	0.47	55.58	0.79	130.12	0.97
73	IV	V	54.63	1.00	I	II	43.10	0.48	83.84	0.58	181.57	1.25
74	IV	V	54.63	1.00	I	III	53.51	0.69	59.64	0.40	167.78	1.28

75	IV	V	54.63	1.00	I	VI	64.29	0.57	93.80	3.29	212.72	3.49
76	IV	V	54.63	1.00	II	III	60.91	0.76	129.94	0.61	245.48	1.40
77	IV	V	54.63	1.00	II	VI	42.22	0.35	117.15	1.30	214.00	1.68
78	IV	V	54.63	1.00	III	VI	59.90	0.67	68.33	0.38	182.86	1.26
79	IV	VI	4.60	0.18	I	II	73.78	2.05	67.23	0.67	145.61	2.16
80	IV	VI	4.60	0.18	I	III	104.89	0.61	25.67	0.59	135.16	0.87
81	IV	VI	4.60	0.18	I	V	151.32	1.20	68.50	1.64	224.42	2.04
82	IV	VI	4.60	0.18	II	III	48.62	0.12	98.93	1.26	152.15	1.28
83	IV	VI	4.60	0.18	II	V	8.94	0.18	119.91	1.08	133.45	1.11
84	IV	VI	4.60	0.18	III	V	54.92	0.25	77.73	0.35	137.25	0.47
85	V	VI	-1.26	0.50	I	II	121.22	0.85	50.86	0.59	170.82	1.15
86	V	VI	-1.26	0.50	I	III	80.10	0.72	117.60	1.93	196.44	2.12
87	V	VI	-1.26	0.50	I	IV	219.81	1.91	25.61	0.62	244.16	2.07
88	V	VI	-1.26	0.50	II	III	18.77	0.27	76.03	1.02	93.54	1.17
89	V	VI	-1.26	0.50	II	IV	53.94	1.03	88.86	0.64	141.54	1.31
90	V	VI	-1.26	0.50	III	IV	103.15	1.07	51.15	0.61	153.04	1.33

Table S2 – Free energies for the formation of S-S bonds for the linear cyclotide and the respective standard deviations. G_1 , G_2 , and G_3 stand for the free energy differences involved in the formation of the first, second and the third S-S bonds. Their sum, G_{tot} , represents the free energy difference involved in the formation of all S-S bonds in the given order, where a_1 and b_1 stands for the cysteine residues involved in the first bond and a_2 and b_2 for the ones of the second bond. Free energy values are in kJ/mol.

n	a_1	b_1	G_1	+-	a_2	b_2	G_2	+-	G_3	+-	G_{tot}	+-
1	I	II	30.69	0.55	III	IV	40.63	0.57	94.30	0.99	165.62	1.27
2	I	II	30.69	0.55	III	V	54.48	0.83	133.89	0.84	219.06	1.30
3	I	II	30.69	0.55	III	VI	75.77	0.86	73.11	1.50	179.57	1.81
4	I	II	30.69	0.55	IV	V	82.23	1.56	55.58	0.56	168.50	1.75
5	I	II	30.69	0.55	IV	VI	99.44	1.15	41.17	0.72	171.30	1.46
6	I	II	30.69	0.55	V	VI	75.03	0.39	144.76	0.89	250.48	1.12
7	I	III	13.77	0.74	II	IV	11.13	0.60	55.92	0.34	80.82	1.01
8	I	III	13.77	0.74	II	V	67.37	0.60	71.35	0.38	152.49	1.03
9	I	III	13.77	0.74	II	VI	23.74	0.57	35.93	0.25	73.44	0.97

10	I	III	13.77	0.74	IV	V	50.12	0.65	8.74	0.33	72.63	1.04
11	I	III	13.77	0.74	IV	VI	30.95	0.67	110.99	0.50	155.71	1.12
12	I	III	13.77	0.74	V	VI	12.38	0.28	54.53	0.99	80.68	1.27
13	I	IV	53.55	0.54	II	III	31.24	0.97	21.21	0.43	106.00	1.19
14	I	IV	53.55	0.54	II	V	21.91	0.40	57.82	1.17	133.28	1.35
15	I	IV	53.55	0.54	II	VI	14.56	0.09	22.72	0.36	90.83	0.66
16	I	IV	53.55	0.54	III	V	73.37	1.64	22.28	0.43	149.20	1.78
17	I	IV	53.55	0.54	III	VI	42.86	0.48	70.01	1.12	166.42	1.33
18	I	IV	53.55	0.54	V	VI	46.41	0.46	87.68	1.33	187.64	1.51
19	I	V	69.66	0.72	II	III	64.52	0.58	59.94	0.26	194.12	0.96
20	I	V	69.66	0.72	II	IV	189.54	1.16	50.19	1.15	309.39	1.79
21	I	V	69.66	0.72	II	VI	164.52	1.55	224.52	3.80	458.70	4.17
22	I	V	69.66	0.72	III	IV	66.58	0.66	138.52	1.46	274.76	1.76
23	I	V	69.66	0.72	III	VI	110.17	1.81	90.52	0.86	270.35	2.13
24	I	V	69.66	0.72	IV	VI	50.06	0.15	17.10	0.18	136.82	0.76
25	I	VI	12.76	0.19	II	III	37.97	0.37	78.85	1.11	129.58	1.19
26	I	VI	12.76	0.19	II	IV	77.47	0.51	80.82	0.81	171.05	0.98
27	I	VI	12.76	0.19	II	V	38.94	0.49	154.57	2.17	206.27	2.23
28	I	VI	12.76	0.19	III	IV	47.94	0.43	72.85	0.45	133.55	0.65
29	I	VI	12.76	0.19	III	V	98.22	1.21	106.82	1.25	217.80	1.75
30	I	VI	12.76	0.19	IV	V	72.44	1.10	45.44	0.89	130.64	1.43
31	II	III	24.64	0.73	I	IV	59.06	0.34	55.13	0.24	138.83	0.84
32	II	III	24.64	0.73	I	V	74.95	0.87	25.43	0.57	125.02	1.27
33	II	III	24.64	0.73	I	VI	19.86	0.27	52.68	0.33	97.18	0.85
34	II	III	24.64	0.73	IV	V	50.62	0.36	132.80	2.02	208.06	2.18
35	II	III	24.64	0.73	IV	VI	8.14	0.36	148.53	2.13	181.31	2.28
36	II	III	24.64	0.73	V	VI	34.07	0.34	95.10	1.73	153.81	1.91
37	II	IV	20.65	0.41	I	III	66.50	1.11	251.91	0.99	339.06	1.54
38	II	IV	20.65	0.41	I	V	139.30	2.16	110.11	1.96	270.06	2.95
39	II	IV	20.65	0.41	I	VI	44.42	0.34	250.87	3.50	315.94	3.54
40	II	IV	20.65	0.41	III	V	233.15	0.84	47.30	1.27	301.10	1.58
41	II	IV	20.65	0.41	III	VI	145.85	0.73	78.93	0.87	245.43	1.21
42	II	IV	20.65	0.41	V	VI	312.77	3.21	33.92	0.24	367.34	3.24
43	II	V	20.45	0.68	I	III	16.67	0.37	42.64	0.41	79.76	0.88
44	II	V	20.45	0.68	I	IV	35.07	0.11	21.47	0.43	76.99	0.81

45	II	V	20.45	0.68	I	VI	18.46	0.31	29.50	0.59	68.41	0.95
46	II	V	20.45	0.68	III	IV	47.72	0.28	12.04	0.59	80.21	0.94
47	II	V	20.45	0.68	III	VI	22.09	0.31	96.57	1.59	139.11	1.76
48	II	V	20.45	0.68	IV	VI	30.48	0.36	4.59	0.14	55.52	0.78
49	II	VI	16.04	0.28	I	III	29.35	0.80	36.03	0.26	81.42	0.89
50	II	VI	16.04	0.28	I	IV	75.33	1.41	15.75	0.32	107.12	1.47
51	II	VI	16.04	0.28	I	V	78.14	1.00	89.60	1.05	183.78	1.48
52	II	VI	16.04	0.28	III	IV	18.13	0.49	177.56	1.06	211.73	1.20
53	II	VI	16.04	0.28	III	V	39.06	0.67	83.88	0.45	138.98	0.85
54	II	VI	16.04	0.28	IV	V	43.69	0.52	95.61	1.17	155.34	1.31
55	III	IV	42.20	0.60	I	II	14.77	0.36	86.33	1.16	143.30	1.35
56	III	IV	42.20	0.60	I	V	82.32	0.94	91.46	0.95	215.98	1.46
57	III	IV	42.20	0.60	I	VI	26.25	0.32	355.54	0.77	423.99	1.03
58	III	IV	42.20	0.60	II	V	135.05	2.73	28.60	0.28	205.85	2.81
59	III	IV	42.20	0.60	II	VI	33.05	0.32	181.33	2.53	256.58	2.62
60	III	IV	42.20	0.60	V	VI	39.02	0.48	73.05	1.12	154.27	1.36
61	III	V	47.24	0.66	I	II	40.51	0.55	68.23	0.49	155.98	0.99
62	III	V	47.24	0.66	I	IV	100.10	1.02	68.37	0.88	215.71	1.50
63	III	V	47.24	0.66	I	VI	48.67	0.48	86.84	1.11	182.75	1.38
64	III	V	47.24	0.66	II	IV	80.80	0.90	39.14	0.69	167.18	1.31
65	III	V	47.24	0.66	II	VI	16.77	0.56	88.79	1.71	152.80	1.92
66	III	V	47.24	0.66	IV	VI	43.64	0.95	15.93	0.38	106.81	1.22
67	III	VI	1.91	0.46	I	II	19.80	0.26	30.71	0.67	52.42	0.85
68	III	VI	1.91	0.46	I	IV	72.79	1.15	33.59	0.27	108.29	1.27
69	III	VI	1.91	0.46	I	V	82.82	0.71	14.84	0.25	99.57	0.88
70	III	VI	1.91	0.46	II	IV	4.98	0.46	98.24	0.67	105.13	0.93
71	III	VI	1.91	0.46	II	V	6.94	0.33	161.12	2.37	169.97	2.44
72	III	VI	1.91	0.46	IV	V	36.22	0.59	53.14	1.49	91.27	1.67
73	IV	V	28.82	0.28	I	II	14.19	0.64	70.85	0.64	113.86	0.95
74	IV	V	28.82	0.28	I	III	3.10	0.50	17.63	0.23	49.55	0.62
75	IV	V	28.82	0.28	I	VI	15.79	0.23	14.29	0.27	58.90	0.45
76	IV	V	28.82	0.28	II	III	14.64	0.27	14.25	0.26	57.71	0.47
77	IV	V	28.82	0.28	II	VI	39.70	0.70	54.86	0.38	123.38	0.84
78	IV	V	28.82	0.28	III	VI	-1.75	0.14	32.62	0.12	59.69	0.34
79	IV	VI	0.82	0.30	I	II	38.86	0.66	40.02	0.36	79.70	0.81

80	IV	VI	0.82	0.30	I	III	30.74	0.44	182.58	2.35	214.14	2.41
81	IV	VI	0.82	0.30	I	V	163.25	0.85	51.75	0.19	215.82	0.92
82	IV	VI	0.82	0.30	II	III	70.82	1.19	240.78	0.77	312.42	1.45
83	IV	VI	0.82	0.30	II	V	112.10	1.63	79.02	0.67	191.94	1.79
84	IV	VI	0.82	0.30	III	V	59.10	0.74	81.47	0.96	141.39	1.25
85	V	VI	36.46	0.51	I	II	24.09	0.43	76.41	0.79	136.96	1.03
86	V	VI	36.46	0.51	I	III	21.43	0.61	62.29	0.54	120.18	0.96
87	V	VI	36.46	0.51	I	IV	113.17	0.56	52.28	1.18	201.91	1.40
88	V	VI	36.46	0.51	II	III	27.03	0.36	109.65	0.91	173.14	1.10
89	V	VI	36.46	0.51	II	IV	68.90	1.94	25.27	0.14	130.63	2.01
90	V	VI	36.46	0.51	III	IV	66.57	1.38	41.80	0.43	144.83	1.53