Calculating CD spectra of flexible peptides: An assessment of TD-DFT functionals

Zlatko Brkljača,[†] Momir Mališ,[†,‡] David M. Smith,[‡,§,*] Ana-Sunčana Smith[†,*]

[†]Institute for Theoretical Physics, Friedrich Alexander University Erlangen-Nürnberg, Staudtstrasse 7, Erlangen, 91058, Germany
[‡]Ruđer Bošković Institute, Bijenička 54, 10000, Zagreb, Croatia
[§]Center for Computational Chemistry, Friedrich Alexander University of Erlangen-Nürnberg, Nägelsbachstraße 25, Erlangen, 91052, Germany

Supporting Information

9 pages

Conformational Phase Space of Met_N and Met_Z in TFE

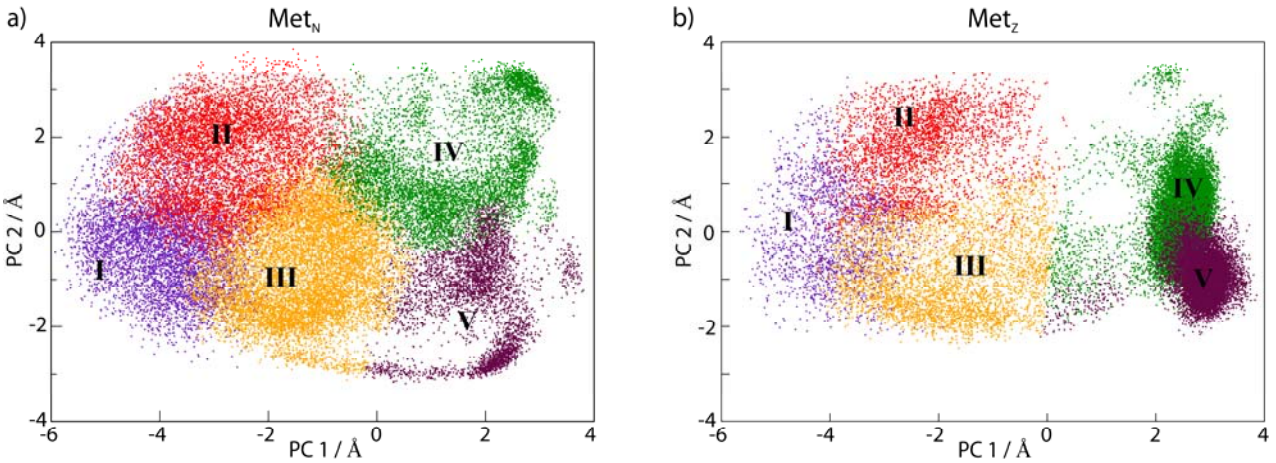


Figure S1. a) The conformational phase space of Met_N at T = 299 K obtained using REMD. The principle component analysis was performed to project 34000 conformations of Met_N onto two first principle components (PC1, PC2). The clustering analysis (K-means algorithm), on the basis of C_α positions, produced five distinct clusters of Met_N conformations, numbered I to V. Each cluster was further sub-clustered, using the same clustering algorithm, to produce 50 distinct all-atom-based sub-clusters (see ref. 8 for the details). The precursor Met_N I-V structures (Figure 2 in the article) correspond to the representative structures of the largest all-atom-based sub-clusters of each of the initial clusters I-V, respectively. b) The conformational phase space of Met_Z at T = 299 K, also obtained using REMD. Identical procedure to the one described in a) was used to analyze the conformational phase space and to obtain the precursor Met_Z structures from which the zwitterionic model systems ($CMet_Z$, Figure 3) were obtained.

Number of the Excited States Lying above 185 nm

Table S1. Number of excited states (N_λ) found above 185 nm for all 15 model systems.

Model Structure	RICC2	CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE0
	N_λ					
$CMet_N$ I	8	10	8	9	26	20
$CMet_N$ II	6	7	7	7	25	20
$CMet_N$ III	7	7	7	7	26	20
$CMet_N$ IV	7	8	7	8	25	19
$CMet_N$ V	7	7	7	8	26	22
$NMet_N$ I	4	4	4	4	13	9
$NMet_N$ II	4	4	4	4	15	10
$NMet_N$ III	4	4	4	4	17	13
$NMet_N$ IV	4	4	4	4	16	12
$NMet_N$ V	4	4	4	4	14	9
$CMet_Z$ I	7	7	7	8	35	24
$CMet_Z$ II	8	8	7	9	30	22
$CMet_Z$ III	8	7	7	7	33	24
$CMet_Z$ IV	7	8	7	7	42	32
$CMet_Z$ V	9	8	7	9	36	27

Solvent effect – brute-force approach

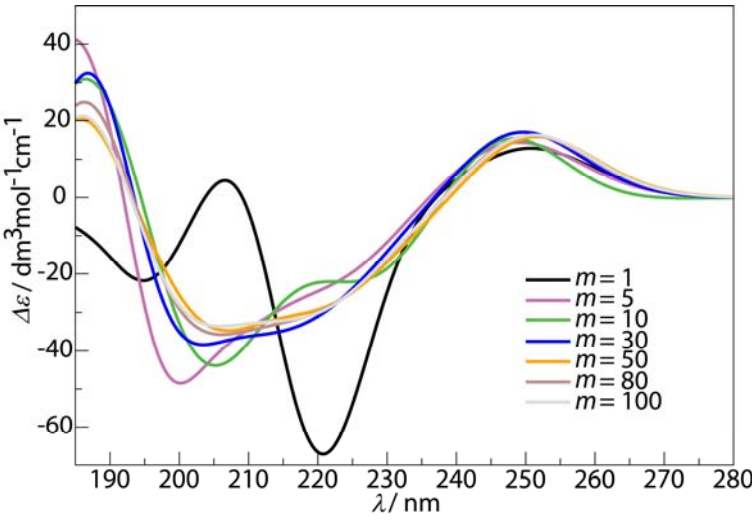


Figure S2. The CD spectrum of a single peptide conformation depends greatly on the position of the surrounding solvent molecules (ref. 8, Figure 2: http://pubs.acs.org/action/showImage?doi=10.1021%2Fct200868y&iName=master.img-001.jpg&type=master). It is thus neccessary to average over a large number of solvent configurations in order to obtain a reliable and reproducible CD spectrum of a single peptide conformation. In the Figure, we show the results from a single selected Met_N conformation, on which we performed constant temperature molecular dynamics (NVT) at 300 K with strong positional restraints on the peptide. This produces a set of distinct and uncorrelated solvent configurations (solvent was extracted every 10 ps after a 200 ps of equilibration period. For further details consult ref. 8). The CD spectrum of the selected Met_N conformation was then evaluated (TD-B3LYP/STO-3G, σ = 0.15 eV) under the influence of each of the obtained solvent configurations, in a manner entirely equivalent to a QM/MM calculation with electrostatic embedding (e.g. ONIOM[TD-B3LYP/STO-3G:AMBER]). The averaging over m of the resulting single solvent configuration CD spectra was then performed, yielding the results shown in the Figure. Satisfactory converence of the CD spectrum is already obtained with m = 30 and additional configurations have only a minor effect, implying that reliable results can certainly be obtained with m = 50. See also Fig S3 from the SI of ref. 8: http://pubs.acs.org/doi/suppl/10.1021/ct200868y/suppl_file/ct200868y_si_001.pdf

Solvent effect – A modified Average Solvent Electrostatic Configuration approach

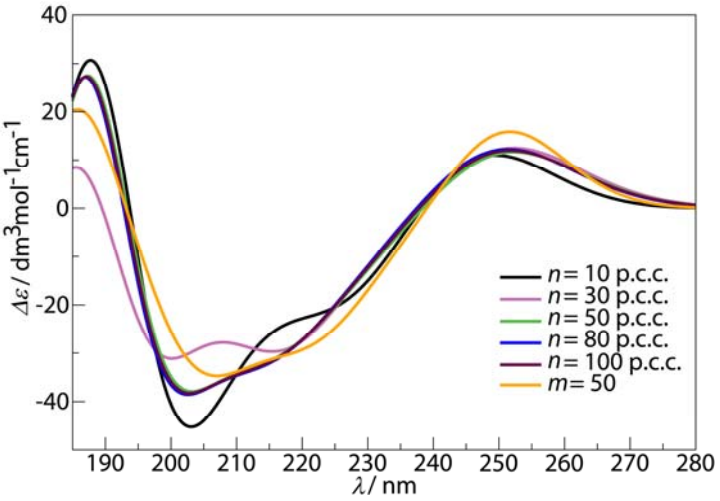


Figure S3. The same Met_N conformation used in Figure S2 is used to test our modified implementation of Average Solvent Electrostatic Configuration (ASEC) approach. Therein, we superimpose n distinct solvent configurations (obtained as explained in the caption of the Figure S2), where each partial charge in each configuration was multiplied by $1/n$ and, in this way, generate a single averaged solvent configuration per peptide conformation. The CD spectra calculated under the influence of such an averaged solvent configuration (B3LYP/STO-3G level of theory, σ = 0.19 eV) are shown as a function of n in the adjacent Figure. The data not only show that the modified ASEC approach exhibits satisfactory convergence but also that it corresponds extremely well with the brute force approach discussed above (orange curve, m = 50). See also Figure 3 from ref. 8: http://pubs.acs.org/action/showImage?doi=10.1021%2Fct200868y&iName=master.img-002.jpg&type=master

Basis set dependence - Met_N

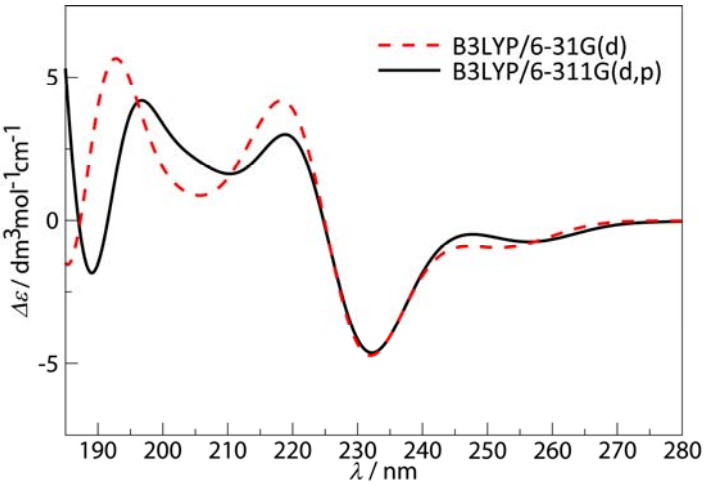


Figure S4. Average CD spectra of Met_N calculated with B3LYP/6-31G(d) (see Figure 9a, red dashed line) and B3LYP/6-311G(d,p) quantum mechanical level of theory. Number of calculated excited states was set to 100 for the 6-31G(d), while 120 states were taken into account for larger 6-311G(d,p) basis set (in both cases σ = 0.135 eV). See also Figure S4 of the SI for ref. 8: http://pubs.acs.org/doi/suppl/10.1021/ct200868y/suppl_file/ct200868y_si_001.pdf.

CD Spectra of Neutral Species vs. Composite Spectra

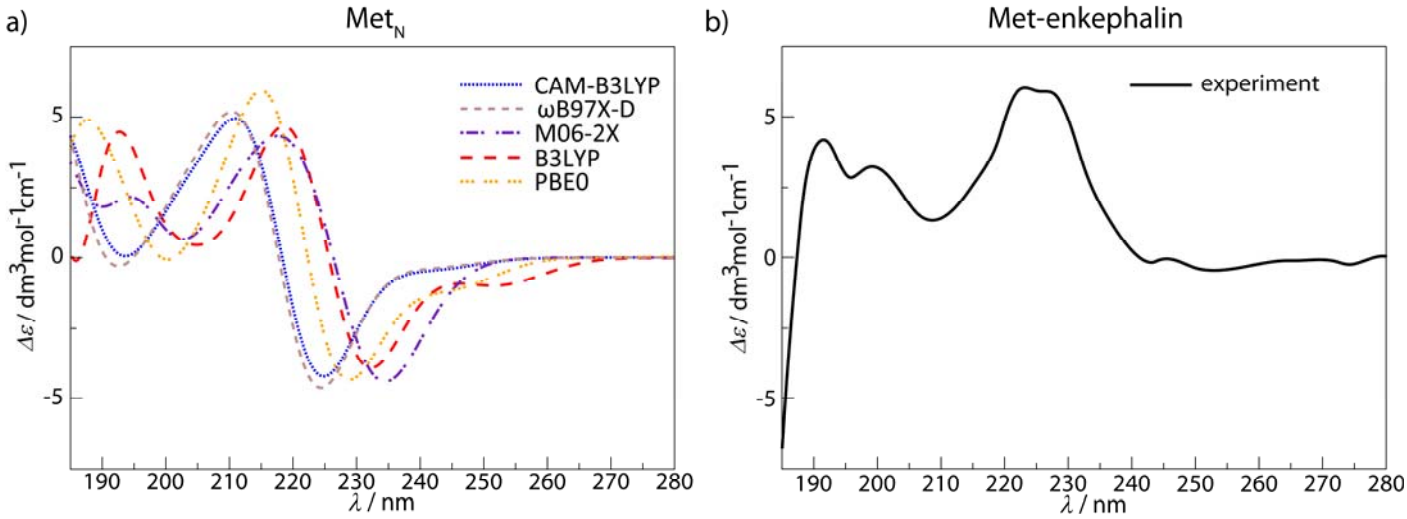


Figure S5. a) Composite spectra of Met-enkephalin obtained by taking into account 9:1 ratio of neutral and zwitterionic forms of peptides in TFE. b) Experimentally obtained CD spectrum of Met-enkephalin in TFE.

CD Spectra of Met_Z

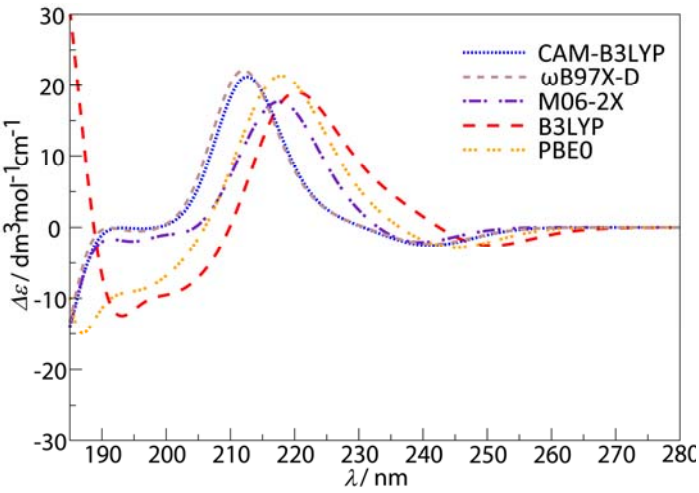


Figure S6. Theoretically calculated CD spectra of Met_Z. No strong minimum at around 230 nm is observed, in contrast to all the average CD spectra of the neutral species (Figure 4a, Figure S5a). However, even though Met_Z spectra correlate better with the experiment in this respect, they also lack any discernible resemblence to the experimental spectrum below 220 nm (Figure S5b).

Convergence of the CD spectrum over the conformational phase space

When comparing the calculated spectra with those obtained experimentally, the quality and convergence features of the theoretical sampling procedure need to be established. We have addressed this issue in greater detail in the main text[8] and in the supporting information of our previous manuscript. Therein, we found that:

1. The use of REMD with about 34000 structures produced satisfactory convergence of the phase space of a peptide of a length like Met_N (Ref.8, Fig. S1 - http://pubs.acs.org/doi/suppl/10.1021/ct200868y/suppl_file/ct200868y_si_001.pdf).

2. The representation of, on average, 700 structures by a single representative structure (amounting to 50 representative structures for the entire phase space) was found as the best compromise in the context of local cluster convergence (Ref. 8, Fig. 10: http://pubs.acs.org/action/showImage?doi=10.1021%2Fct200868y&iName=master.img-010.jpg&type=master). Decreasing the average number of structures in a sub-cluster resulted in numerous clusters with very few members and a limited number of clusters containing many structures.

3. This level of statistical sampling was sufficient to reproduce the signal obtained from two clusters occupying the same phase space but belonging to different although related peptides (Ref. 8, Fig. 11: http://pubs.acs.org/action/showImage?doi=10.1021%2Fct200868y&iName=master.img-011.jpg&type=master).

4. This level of sampling was also sufficient to obtain mirror imaged spectra from two completely independent simulations of enantiomeric peptides. (Ref. 8, Fig. 9: http://pubs.acs.org/action/showImage?doi=10.1021%2Fct200868y&iName=master.img-009.jpg&type=master). Following this calculation, we were able to estimate the convergence uncertainty to about 20%, which is in the same range (or lower) than the uncertainties arising from the calculation of the electronic transitions with TD-DFT.

Figure S7 below examines alternative sampling protocols.

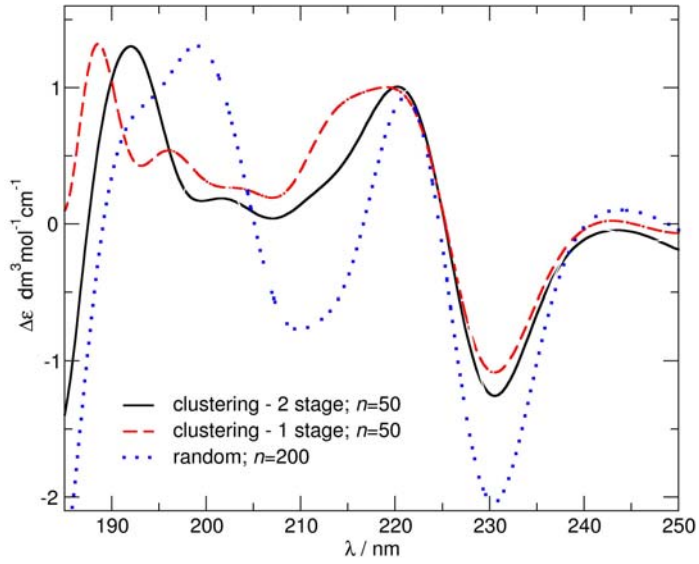
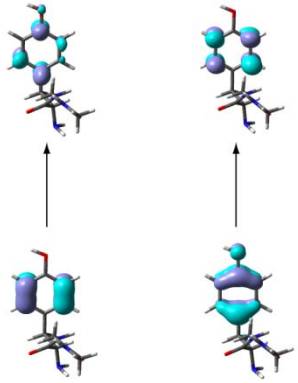
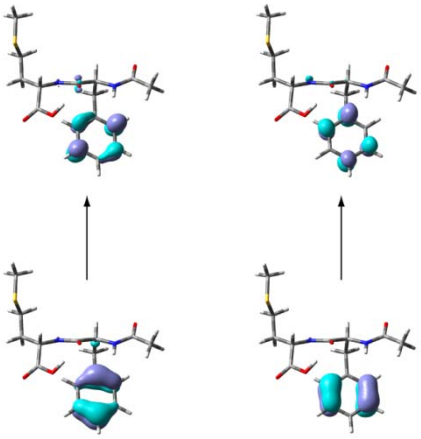
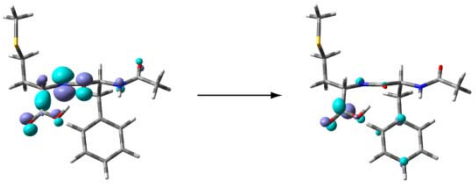
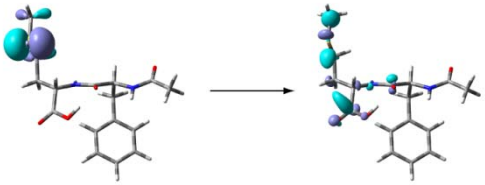
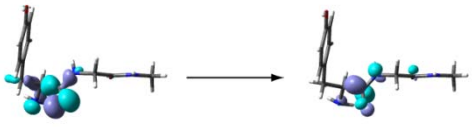
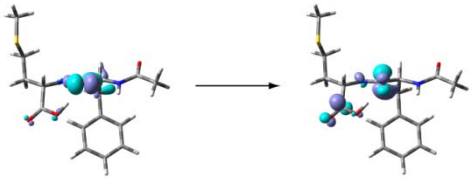
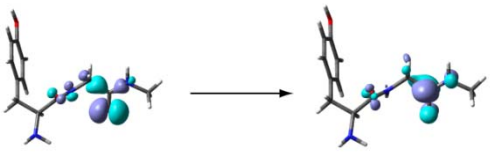
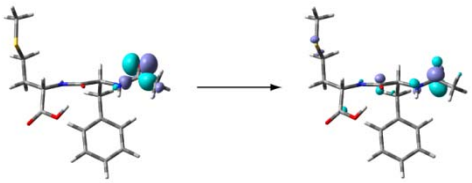
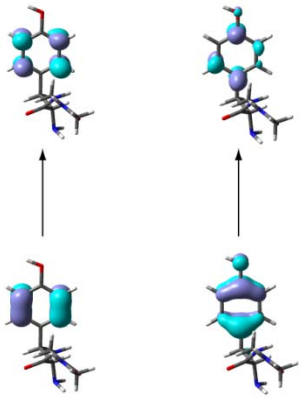
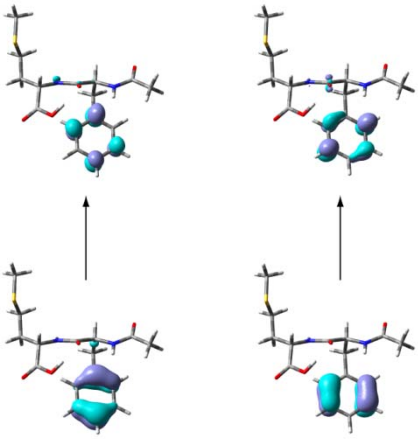
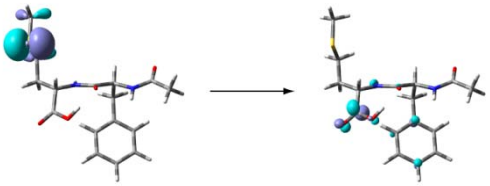


Figure S7. To further examine the convergence issues in the context of the current work, we present a comparison of Met_N spectra, obtained using TD-B3LYP/6-31G(d) (with σ=0.085) with different approaches for sampling the structural phase space. Specifically, the black curve was obtained the two-stage clustering procedure outlined in ref. 8, where the structures were initially clustered, on the basis of the backbone atoms, into five primary clusters. Further clustering was subsequently performed, on the basis of all peptide atoms, so as to produce a total of 50 clusters (and hence 50 representative structures). The red-dashed curve, on the other hand, was produced by a single-stage clustering procedure, based on all peptide atoms, also to produce 50 representative structures. The spectra arising from the application of the appropriate weighting factors from these two procedures are in close agreement with one another and with the main features of the experiment. In contrast, a sampling procedure based on random extraction of structures from the total pool of 34000 is slowly converging and, even with the inclusion of 200 structures (blue dotted line), does not offer the same degree of agreement with experiment as that exhibited by the clustering approaches. The intensities of the three spectra were normalized to the amide transition at ~220 nm.

Orbital presentations of excited states

Table S2. Orbital presentations of the investigated excitations (consult Table 1 in the article).

Excited state	Orbital presentation
Tyrosine ^1L_b	
Phenylalanine ^1L_b	
$n_O(COOH) + \pi(F\text{-}M) \rightarrow \pi^*(COOH)$	
$n_S(M) \rightarrow n_S(M)^*$	
$n_O(Y) \rightarrow \pi^*(Y\text{-}G)$	

$n_O(F) \rightarrow \pi^*(F\text{-}M)$	
$n_O(G) \rightarrow \pi^*(G\text{-}Nme)$	
$n_O(Ace) \rightarrow \pi^*(Ace\text{-}F)$	
Tyrosine ^1L_a	
Phenylalanine ^1L_a	
$n_S(M) \rightarrow \pi^*(COOH)$	

Root mean square errors (RMSE) tables

Table S3. RMSE of TD-DFT transition wavelengths from RICC2 values are calculated as $\sqrt{\frac{1}{n}\sum_i(\boldsymbol{TDDFT_i} - \boldsymbol{RICC2_i})^2}$. $\boldsymbol{TDDFT_i}$ and $\boldsymbol{RICC2_i}$ correspond to the values of the wavelength of the particular excitation in the *i*-th peptide model found using the TD-DFT method of interest and RICC2, respectively.

Excitation type	Excited state	Wavelength / nm				
		CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE0
Aromatic excitations	Tyrosine ^1L_b	17.0	16.8	20.3	9.1	13.3
	Phenylalanine ^1L_b	15.7	15.6	17.9	9.9	14.0
	Tyrosine ^1L_a	3.4	2.8	1.4	8.8	9.3
	Phenylalanine ^1L_a	7.9	7.4	2.5	15.4	11.3
CT	n_S(M) -> π*(COOH)	6.6	4.6	7.7	75.3	59.4
Amide excitations	n_O(Y) -> π*(Y-G)	2.2	1.8	9.2	8.1	6.0
	n_O(F) -> π*(F-M)	0.5	0.5	6.8	9.8	4.8
	n_O(G) -> π*(G-Nme)	3.5	3.0	10.5	8.1	4.9
	n_O(Ace) -> π*(Ace-F)	2.9	2.3	9.4	7.4	4.3

Table S4. RMSE of TD-DFT rotatory strengths from RICC2 values are calculated as $\sqrt{\frac{1}{n}\sum_i(\boldsymbol{TDDFT_i} - \boldsymbol{RICC2_i})^2}$. $\boldsymbol{TDDFT_i}$ and $\boldsymbol{RICC2_i}$ correspond to the values of the rotatory strength of the particular excitation in the *i*-th peptide model found using the TD-DFT method of interest and RICC2, respectively.

Excitation type	Excited state	Rotatory strength / 10^{-40} erg esu cm Gauss^{-1}				
		CAM-B3LYP	ωB97X-D	M06-2X	B3LYP	PBE0
Aromatic excitations	Tyrosine ^1L_b	9.7	8.4	10.5	10.0	9.7
	Phenylalanine ^1L_b	0.9	0.8	2.1	1.2	1.9
	Tyrosine ^1L_a	15.6	13.1	13.2	22.8	19.9
	Phenylalanine ^1L_a	6.6	6.7	4.2	11.0	14.7
CT	n_S(M) -> π*(COOH)	17.6	13.0	21.0	23.0	20.8
Amide excitations	n_O(Y) -> π*(Y-G)	9.4	8.4	11.2	15.1	9.4
	n_O(F) -> π*(F-M)	4.5	7.8	3.5	5.5	14.0
	n_O(G) -> π*(G-Nme)	7.0	4.4	3.7	2.4	6.6
	n_O(Ace) -> π*(Ace-F)	2.9	2.4	2.5	5.3	5.4