**Electronic Supplementary Information for:**

**"Diverging conformations guide dipeptide self-assembly into crystals or hydrogels"**

M. Monti[1], E. Scarel[1], A. Hassanali[2], M. Stener[1], S. Marchesan[1]

[1] Chem. Pharm. Sc. Dept., University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy

[2] The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

# Table of contents

# 1. Experimental part

All the compounds investigated in this work were previously published.[1,2,3] Their purity and molecular characterisation were established with different techniques, including LC-MS, $^{1}$H NMR and $^{13}$C NMR. In particular, LC-MS traces show a single sharp chromatographic peak for each compound and the corresponding mass spectra in positive and negative ion modes are in agreement with the structural elucidation obtained with NMR spectra, corresponding to the desired product.

### 1.1 Circular dichroism:

L-Val-D-Phe and L-Nva-D-Phe were dissolved in milli-Q-water (resistivity > 18 M Ω cm) at the desired concentration and the pH was corrected with a small amount of NaOH 1 M to reach pH 7.0. All samples were loaded into a 0.1 mm quartz cuvette and sealed with a thin layer of grease to prevent the solvent evaporation during the experiment. All spectra were acquired on a Jasco j-815 spectropolarimeter at 70 °C (Peltier) with the following parameters: data pitch 1 nm, scan speed 100 nm/min, D.I.T 1 s. For every sample were collected 20 spectra in order to evaluate the consistency of the signal, check the thermal equilibrium and the possible leaking of solvent due to evaporation. Then all the spectra were merged to increase the signal/noise ratio.

### 1.2 Fibres solutions:

L-Val-D-Phe was dissolved in PBS buffer at 80 mM following the procedure reported in literature.[1] The solution was then placed in an oil bath at 70 °C for 15 minutes and then was left cool down at room temperature.

L-Nva-D-Phe formed fibres after the CD experiment inside the cuvette, therefore the pictures were acquired on that sample.

### 1.3 Optical microscope:

All pictures were acquired on a ZEISS Primovert inverted microscope with Axiocam 208 colour. Few microliters of the fibre solution were placed on a glass slides and then investigated under microscope. Pictures were collected with normal light and with polarized light in order to highlight the chirality of the object. For L-Val-D-Phe some pictures with fibres and crystalline material were acquired. In this case the drop was left dry on the glass slide in order to allow the crystallisation of the remaining peptide in solution.

# 2. Computational procedure

Here we describe the computational set-up employed to study the conformational landscape of the dipeptide series and to simulate the ECD spectra for L-Val-D-Phe, D-Phe-L-Val, and L-Nva-D-Phe. Further details on this procedure have been discussed in previous works.[4,5]

All the molecular Dynamics (MD) simulations were performed with the GROMACS package[6,7] version 2022.3 and the OPLS-AA force field.[8,9] The water solvent was described with the TIP3P model.[10] All the simulations were carried out in the NVT ensemble, keeping the temperature constant with the velocity-rescale thermostat.[11] The LINCS algorithm[12] was used to constrain all the bonds for improving the performance. The long-range electrostatic interactions were treated with the Particle Mesh Ewald method,[13] using a cut-off of 1.0 nm and a 0.12 nm grid spacing. Each dipeptide has been initially inserted in a cubic box and solvated. The volume and number of water molecules were chosen in order to reproduce the concentrations used for the experimental ECD measurements in solution (*i.e.,* 1-5 mM).[1,2] On these initial boxes, energy minimisations were performed with the steepest descent algorithm. Afterwards, each box was adjusted to reproduce the correct density (in infinite dilution in water) at the temperature of interest (*i.e.,* 25, or 70°C), and a pressure of 1.0 bar. In detail, the solute-solvent volume was modified until its average pressure reached that measured on a box of pure water, with the same number of molecules, previously simulated in the NVT ensemble at the experimental density of pure water at the selected temperature: 997.1 g/L (25°C), 977.6 g/L (70°C). For each system, the equilibration was followed by a production MD (hereafter termed full-MD) of 100 ns.

The full-MD simulations have been then analysed with the essential dynamics (ED), whose features are extensively described in the literature,[14,15] and here only summarised. As a first step, we built and diagonalised the covariance matrix of the dipeptide atomic coordinates, thus obtaining a set of eigenvectors and corresponding eigenvalues. Such eigenvectors correspond to the eigendirections along which the system undergoes its internal motion, while the associated eigenvalues represent the actual values of these mean square fluctuations. As a matter of fact, the first eigenvectors, which correspond to the highest eigenvalues, represent the directions essential to account for most of the conformational transitions of the system. Herein, we limited the number of essential eigenvectors to the first pair as a compromise between accuracy and computational cost. For each dipeptide, we projected the Cartesian coordinates of the MD trajectory onto the two eigendirections, obtaining the so-called Principal Components (PCs) which allow us to reduce and simplify the dimension and complexity of the conformational landscape. In practice, the PCs are visualised building a 2D histogram where each square represents a conformational basin $i^{th}$, whose probability P($i$) depends on how frequently that square is spanned (how probable is that conformation). Considering $P_{ref}$ (=1) as the probability of the most probable region, and a negligible difference between the partial molar volumes (MD in the NVT ensemble), the (standard) Gibbs free energy difference ($\Delta G°$) between these basins may be estimated with the Boltzmann statistics:

$$\Delta G^{\circ} = -RT \ln \frac{P(i)}{P_{ref}} \quad (1)$$

Repeating the calculation for each conformational basin (square), we obtain the relative free energy landscapes shown in Fig. S1-S2, and in the main text (Fig. 2). The analysis of these conformational spaces is focused only on the regions associated to a $\Delta G° < k_B T$, where the most significant conformers are found. Therefore, we extracted one or two structures from each low-energy basin as representative of the corresponding region.

We then performed a second set of 15 ns MD simulations (hereafter termed constrained-MD), with the same box and protocol discussed above, keeping frozen the conformations just extracted in order to investigate the conformational role of the solvent. All the constrained-MD trajectories were analysed again with the ED analysis[16] for locating the most plausible peptide-solvent clusters, then used for further analyses and the quantum-chemical calculations. Such clusters, containing 50 water molecules, were built employing a procedure well described in previous works[4,16] and here only briefly summarised. For each constrained-MD, we first constructed the fixed ellipsoid which best describes the frozen dipeptide geometrical shape. Afterwards, we selected the N (=50) water molecules showing the lowest square distances in the so-defined ellipsoidal metrics. This step is repeated at each frame of the constrained-MD, thus obtaining a trajectory for the selected peptide-$(H_2O)_{50}$ cluster. All the trajectories were investigated with ED again, providing us with the 2D conformational landscapes of the peptide-solvent clusters. Once again, we focused our attention on the most probable regions of the space, extracting one or two cluster conformations associated to the higher probability p(j,i) values (lower $\Delta G°$, Eq. 1). Therefore, multiplying P(i) and p(j,i), we obtain a final probability $P_{TOT}(k)$ for each k peptide-$(H_2O)_{50}$ cluster. The $P_{TOT}$ values of the extracted clusters were then normalised. Therefore, considering for each dipeptide case, a certain number N of selected conformations, the total sum ($P_{ALL}$) of the corresponding $P_{TOT}$ values will be:

$$P_{ALL} = \sum_{k=1}^{N} P_{TOT}(k) = 1 \quad (2)$$

The full-, and constrained-MD, together with the ED analysis, were performed for all the heterochiral dipeptides studied in this work (*i.e.,* L-Val-D-Phe, D-Phe-L-Val, L-Nva-D-Phe, D-Phe-L-Nva, L-Ile-D-Phe, and D-Phe-L-Ile).

For L-Val-D-Phe (25°C), D-Phe-L-Val (25°C), and L-Nva-D-Phe (25°C and 70°C) we also carried out the quantum-chemical calculations. More in detail, we first optimised the geometry of the selected N cluster conformations at the Density Functional Theory (DFT)[17] level by using the Amsterdam Density Functional (ADF) engine of the AMS code.[18] A set of 4, 8, 5, and 4 dipeptide-$(H_2O)_{50}$ clusters were considered for L-Val-D-Phe (25°C), D-Phe-L-Val (25°C), L-Nva-D-Phe (25°C),

and L-Nva-D-Phe (70°C), respectively. In order to maintain the features of the extracted conformations, we constrained the internal degrees of freedom corresponding to the semi-classical motions (torsional angles). For all the optimisations, a Slater-type orbital set of triple-zeta quality was used. In order to describe properly the physics of the system, *i.e.,* intra-, inter-molecular H-bonds and charge-transfer excitations, we employed the range-separated exchange-correlation (xc) functional wB97X-D which contains the dispersion and non-local exchange asymptotical corrections.[19]

Finally, we calculated the ECD spectra for the selected conformations at the Time-Dependent DFT (TDDFT) level by using the Casida approach[20] (Davidson algorithm). The same basis set and xc functional mentioned above were used. The lowest 100 excited-states were analysed to cover the experimental energy range.[1,2] Furthermore, the calculated spectra were reported in $[\theta]x10^{-3}$ ($mdegM^{-1}cm^{-1}$) units for an easy comparison with the experimental measurements. Since the calculated spectra are obtained as discrete lines in terms of rotatory strength ($R_{0k}$, in cgs units), we used the following expressions to convert them in Δε units[21] first, and in $[\theta] \times 10^{-3}$ then:

$$\Delta\varepsilon(E) = \frac{1}{2.297x10^{-39}} \frac{1}{\sigma\sqrt{\pi}} \sum_k R_{0k} E_{0k} e^{-\left(\frac{E-E_{0k}}{\sigma}\right)^2} \quad (3)$$

$$[\theta] \times 10^{-3} = \Delta\varepsilon \times 3.298 \quad (4)$$

where σ is related to the Half Width at Half Maximum (HWHM) of the Gaussian functions used to convolute the calculated spectra:

$$\sigma = \frac{HWHM}{\sqrt{\ln 2}} \quad (5)$$

In this work, we used HWHM = 0.15 eV to suitably match the experimental features. Each spectrum was weighted by the corresponding normalised $P_{TOT}(k)$ (k=1,N, see Eq. 2) and summed up to give the final statistically averaged ECD. They were then compared with the experimental references[1,2] for a quality evaluation.

To qualitatively assess the differences between solution and crystal phases, we decided to compare the spectra in solution with those calculated on crystal models (see Fig. S7-S8). Starting from the X-ray crystal structures available for these systems,[1,2] we built four-units models as a compromise between a suitable representation of the crystal molecular orientations and the limitations of the Casida scheme.[22] Although different orientations of the units can be considered, we focused only on the models where all the four units are directly involved in the formation of the water channels that are observed for these crystals. The parameters of the TDDFT calculations are those mentioned above for the calculation in solution.

All the molecular visualisations were realised using the PyMOL software.[23]

## 3. Figures and Tables

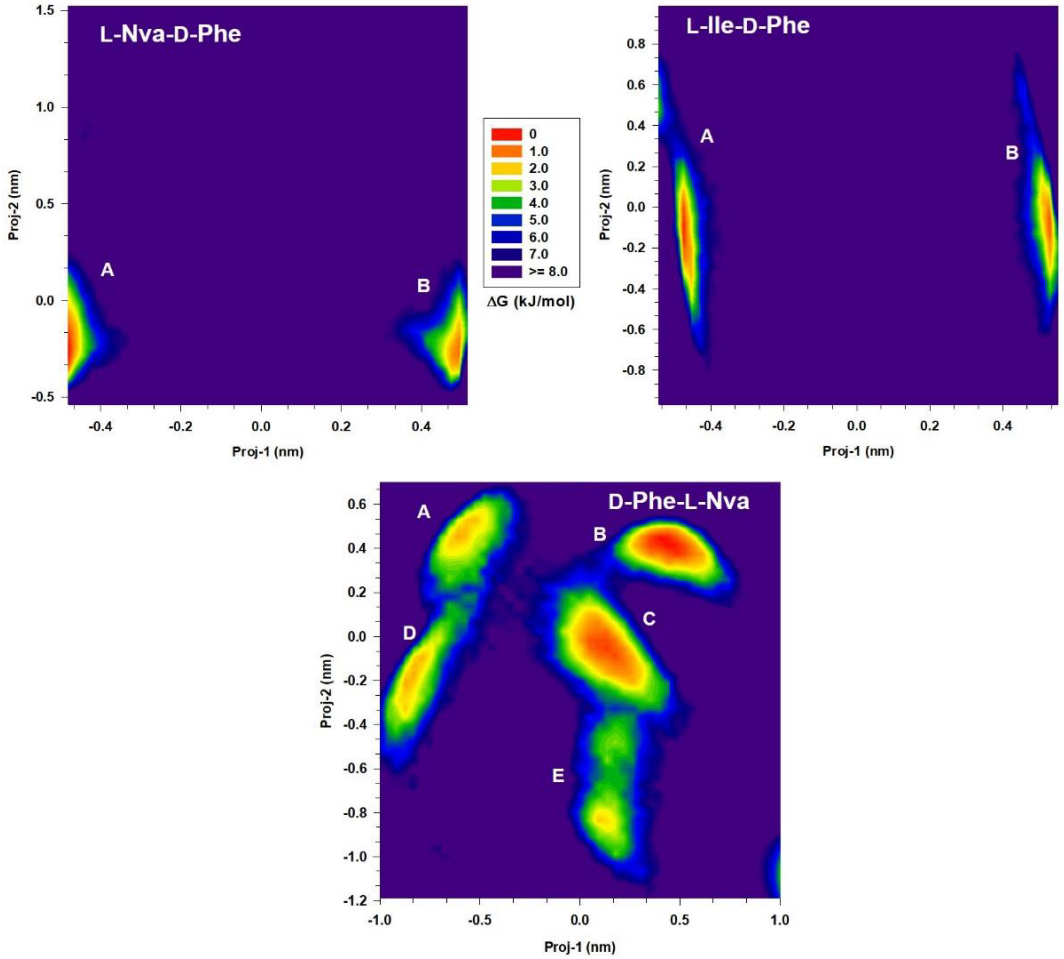


**Figure S1.** Relative free energy landscapes in the essential plane of L-Nva-D-Phe (top left), L-Ile-D-Phe (top right), and D-Phe-L-Nva (bottom) at RT. The energy scale, in kJ/mol, is reported as a vertical-coloured bar on the top left region. All the low-energy regions, within which the most probable conformations lie, have been labelled with capital letters.

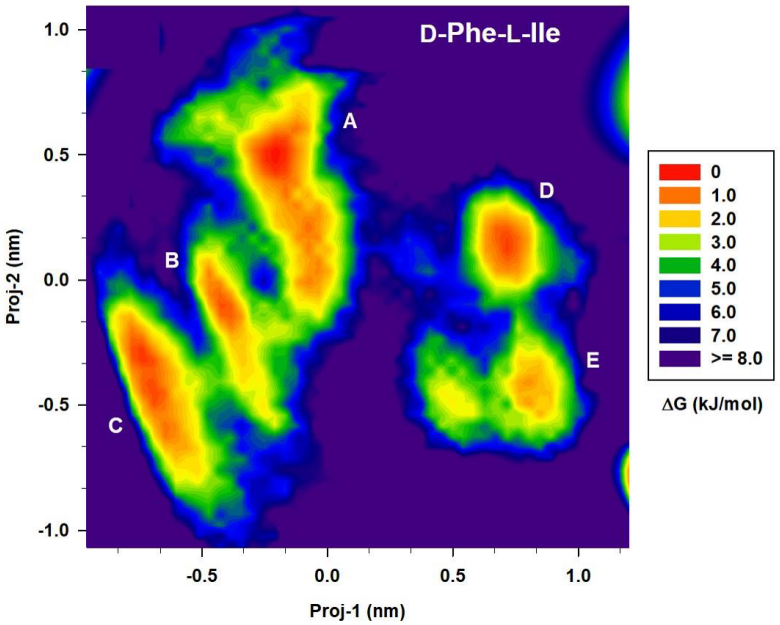


**Figure S2.** Relative free energy landscape in the essential plane of D-Phe-L-Ile at RT. The energy scale, in kJ/mol, is reported as a vertical-coloured bar on the left. All the low-energy regions, within which the most probable conformations lie, have been labelled with capital letters.

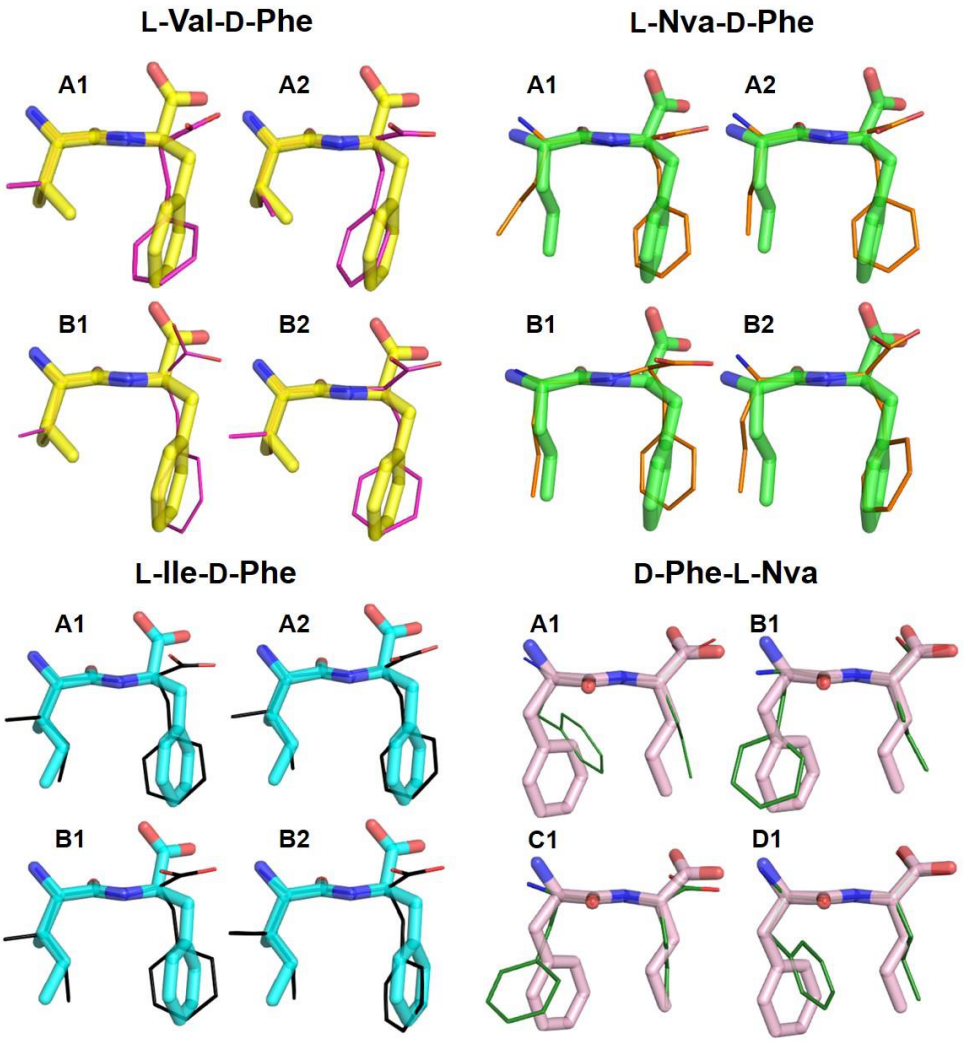


**Figure S3.** Representative significant ($\Delta G \leq 2.0$ kJ/mol) conformers extracted from the corresponding 2D conformational landscape (see Fig. S1) of the L-Val-D-Phe (top left), L-Nva-D-Phe (top right), L-Ile-D-Phe (bottom left), and D-Phe-L-Nva (bottom right) dipeptide. All the conformations (shown as coloured wire lines) have been overlapped with the corresponding experimental crystal units (shown as coloured sticks).[1,2,3] The relative free-energy values associated to these conformations are reported in Table S1.

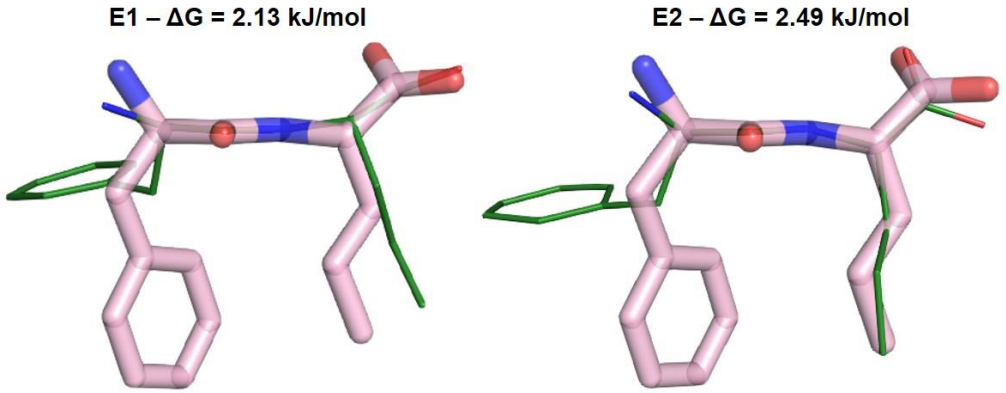


**Figure S4.** Representative conformers with a ΔG around (2.1-2.5) kJ/mol extracted from the 2D conformational landscape (see Fig. S1) of D-Phe-L-Nva. The two conformations (shown as pink wire lines) have been overlapped with the corresponding experimental crystal unit (shown as pink sticks).[2]

**a) L-Val-D-Phe**

| Label | Relative Free Energy (kJ/mol) |
|---|---|
| A1 | 0.95 |
| A2 | 2.04 |
| B1 | 1.90 |
| B2 | 0.0 |

**b) L-Nva-D-Phe**

| Label | Relative Free Energy (kJ/mol) |
|---|---|
| A1 | 1.18 |
| A2 | 0.0 |
| B1 | 1.66 |
| B2 | 0.96 |

**c) L-Ile-D-Phe**

| Label | Relative Free Energy (kJ/mol) |
|---|---|
| A1 | 0.0 |
| A2 | 1.96 |
| B1 | 1.52 |
| B2 | 0.56 |

**d) D-Phe-L-Nva**

| Label | Relative Free Energy (kJ/mol) |
|---|---|
| A1 | 1.68 |
| B1 | 0.0 |
| C1 | 0.64 |
| D1 | 1.71 |

**Table S1.** Relative Free Energy values of the significant conformers (ΔG ≤ 2.0 kJ/mol) extracted for: a) L-Val-D-Phe, b) L-Nva-D-Phe, c) L-Ile-D-Phe, and d) D-Phe-L-Nva. All the conformations have been shown in the previous Fig. S3.

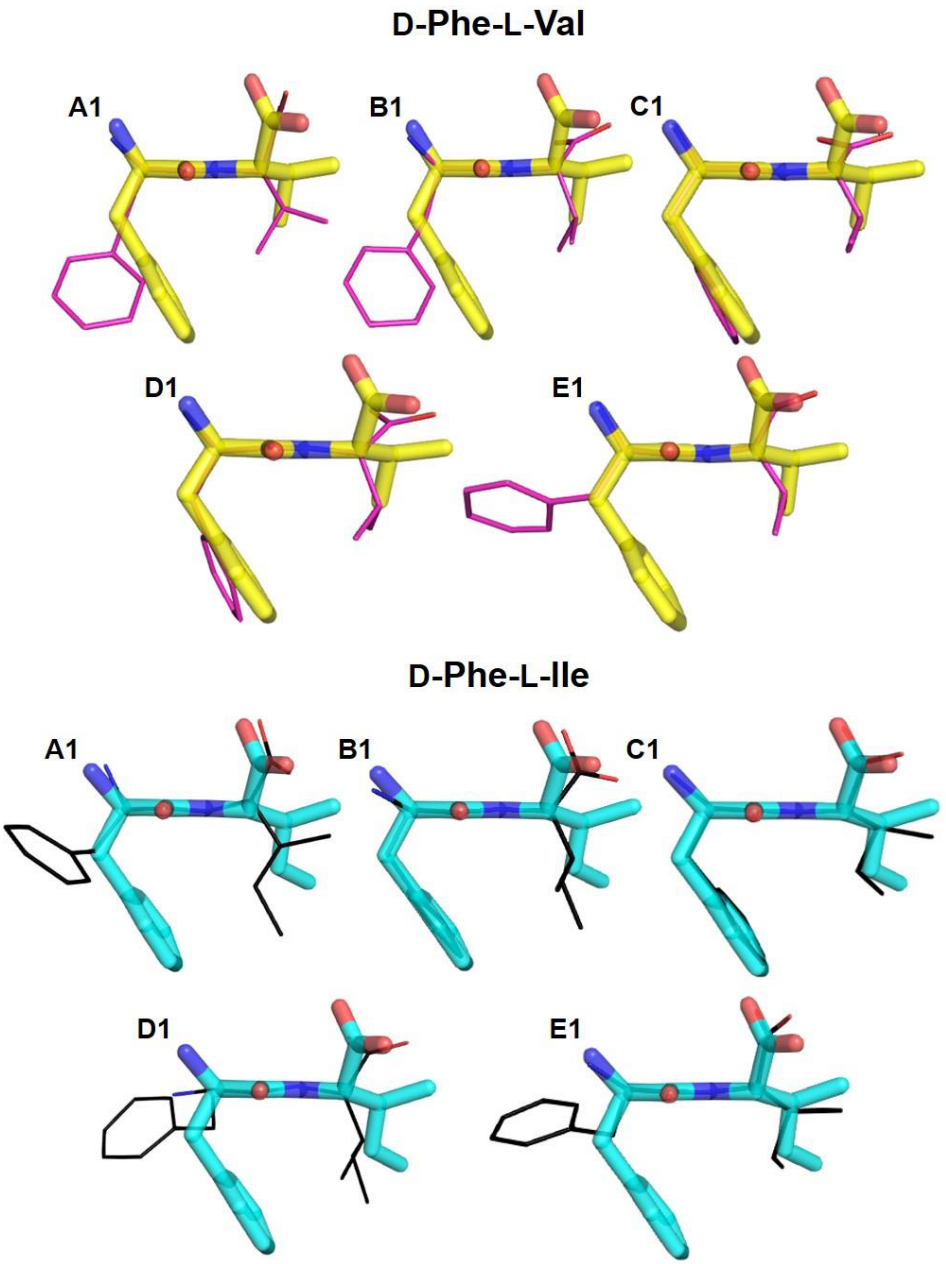


**Figure S5.** Representative significant ($\Delta G \leq 2.0$ kJ/mol) conformers extracted from the corresponding 2D conformational landscape (see Fig. S2) of the D-Phe-L-Val (top), and D-Phe-L-Ile (bottom) dipeptide. All the conformations (shown as coloured wire lines) have been overlapped with the corresponding experimental crystal units (shown as coloured sticks).[1,3] The relative free-energy values associated to these conformations are reported in Table S2.

**a) D-Phe-L-Val**

| Label | Relative Free Energy (kJ/mol) |
|---|---|
| A1 | 0.0 |
| B1 | 1.18 |
| C1 | 1.13 |
| D1 | 0.074 |
| E1 | 0.73 |

**b) D-Phe-L-Ile**

| Label | Relative Free Energy (kJ/mol) |
|---|---|
| A1 | 0.0 |
| B1 | 1.30 |
| C1 | 0.21 |
| D1 | 0.33 |
| E1 | 1.70 |

**Table S2.** Relative Free Energy values of the significant conformers ($\Delta G \leq 2.0$ kJ/mol) extracted for: a) D-Phe-L-Val, and b) D-Phe-L-Ile. All the conformations have been shown in the previous Fig. S5.

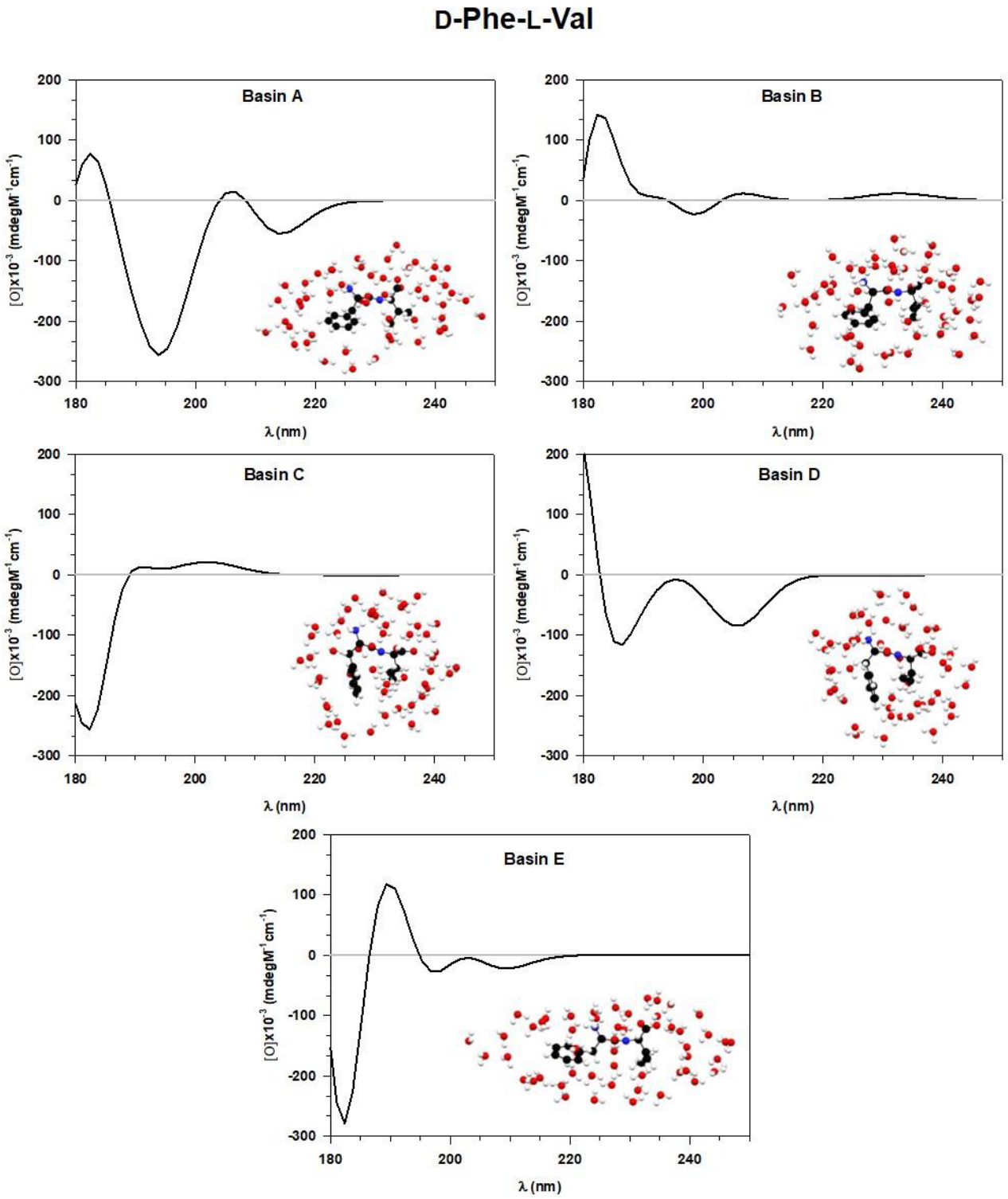


**Figure S6.** Individual calculated ECD spectra of the D-Phe-L-Val-$(H_2O)_{50}$ clusters. Here, we reported only the most probable structures obtained from the statistical analysis of both the dipeptide and the water conformational spaces. The labels of the basin refer to those defined for the conformational landscape of D-Phe-L-Val (see Fig. 1 in the main article). All the spectra are not weighted for their corresponding Boltzmann's factor. Technical details on the calculations have been reported above.

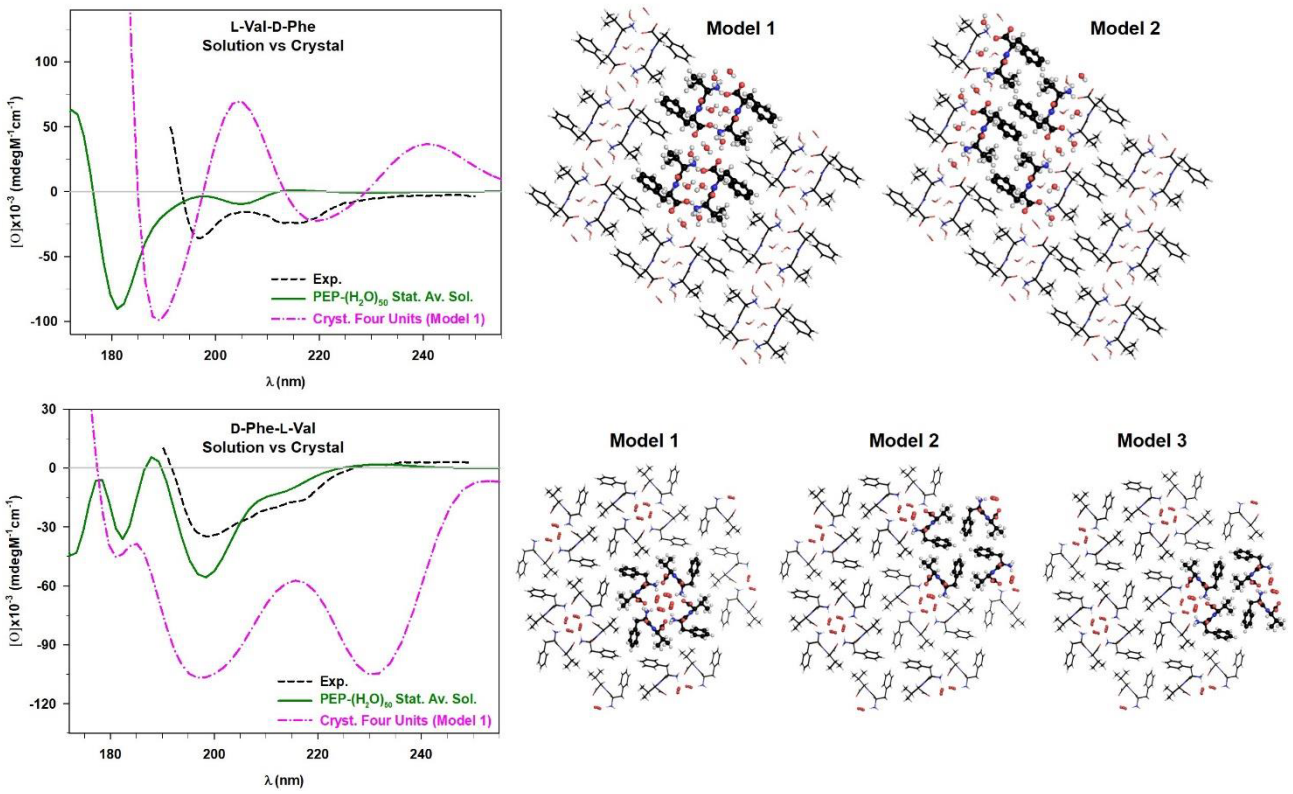


**Figure S7.** Comparison of the experimental (Exp.) ECD spectrum (dashed black line)[1] of L-Val-D-Phe, and D-Phe-L-Val, respectively, with those calculated: i) statistically averaging the ECD of the most probable conformations in solution (PEP + 50 $H_2O$ mols., solid green line), ii) considering a four units crystal model (dash-dotted pink line). Different possible four-unit models are also shown on the right. The intensities of the crystal four units ECD have been divided by a factor 2. Technical details on the calculations have been reported above.

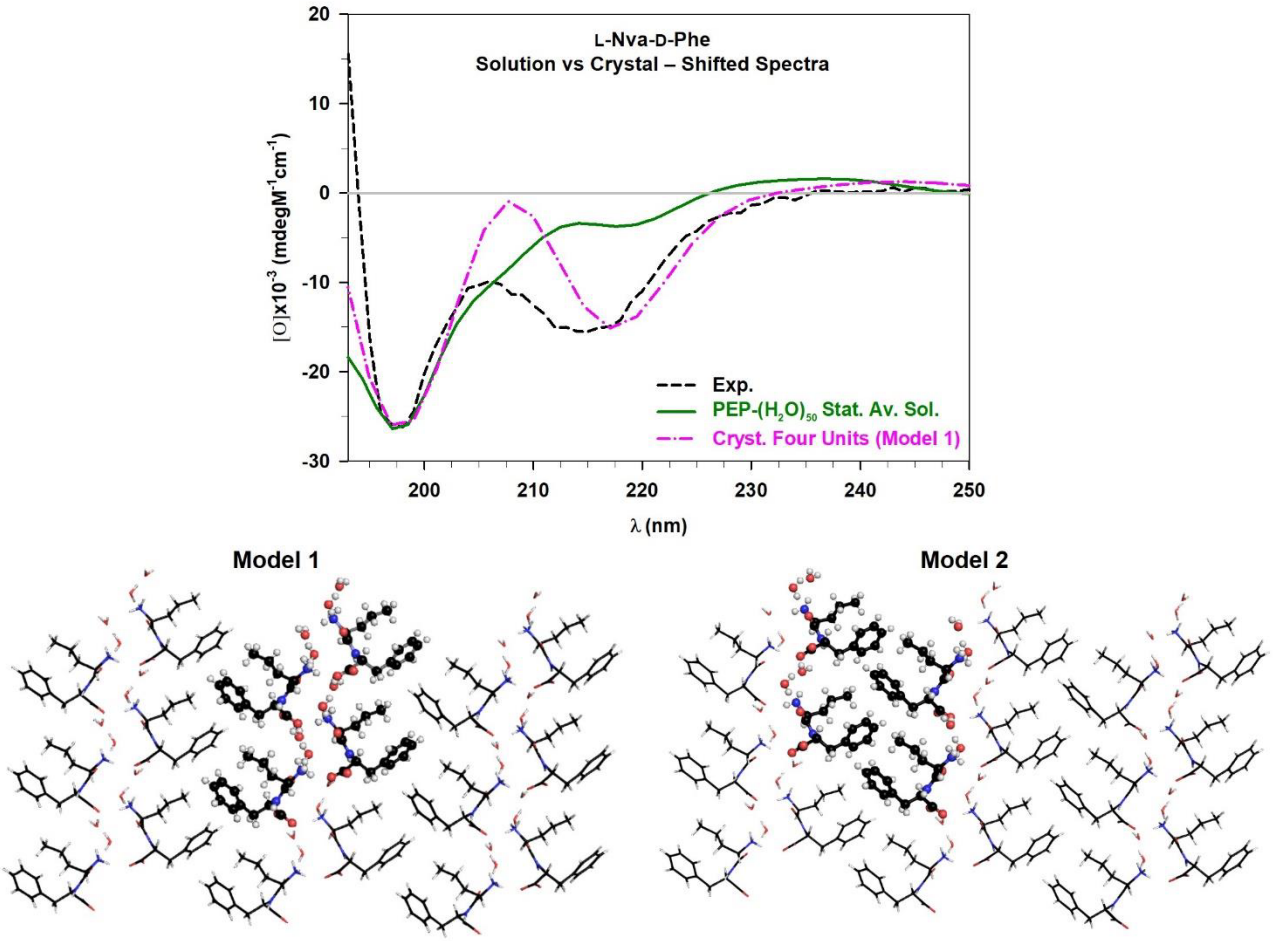


**Figure S8.** Comparison of the experimental (Exp.) ECD spectrum (dashed black line)[2] of L-Nva-D-Phe with those calculated: i) statistically averaging (stat. av.) the ECD of the most probable conformations in solution (PEP + 50 $H_2O$ mols., solid green line), ii) considering a four units crystal model (dash-dotted pink line). Different possible four-unit models are also shown on the bottom. The stat. av. and the four units ECD have been shifted by +9.2 nm, and -28.5 nm, respectively, and their intensities have been normalised in order to match properly the experimental response. Technical details on the calculations have been reported above.

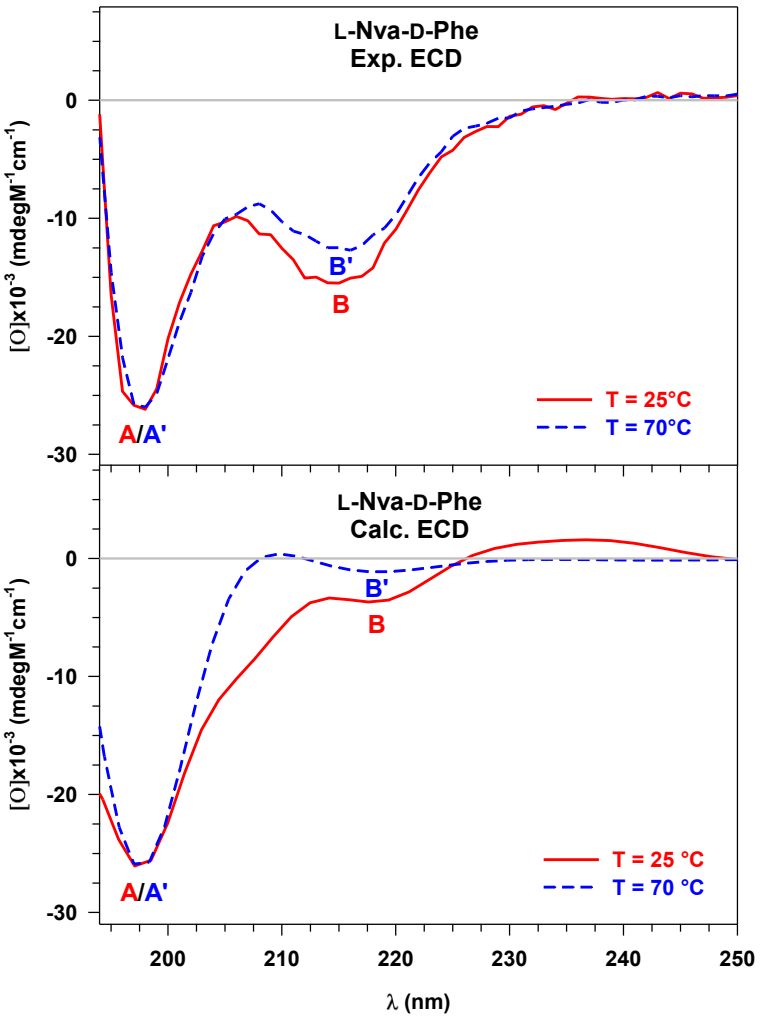


**Figure S9.** Experimental (Exp.) and calculated (Calc.) ECD spectra of L-Nva-D-Phe in aqueous solution at 25°C (solid red line) and 70°C (dashed blue line), respectively. The calculated spectra have been shifted by +9.2 nm (25°C) and +16 nm (70°C), and their intensities have been normalised in order to match properly the experimental response. Technical details on the calculations have been reported above.

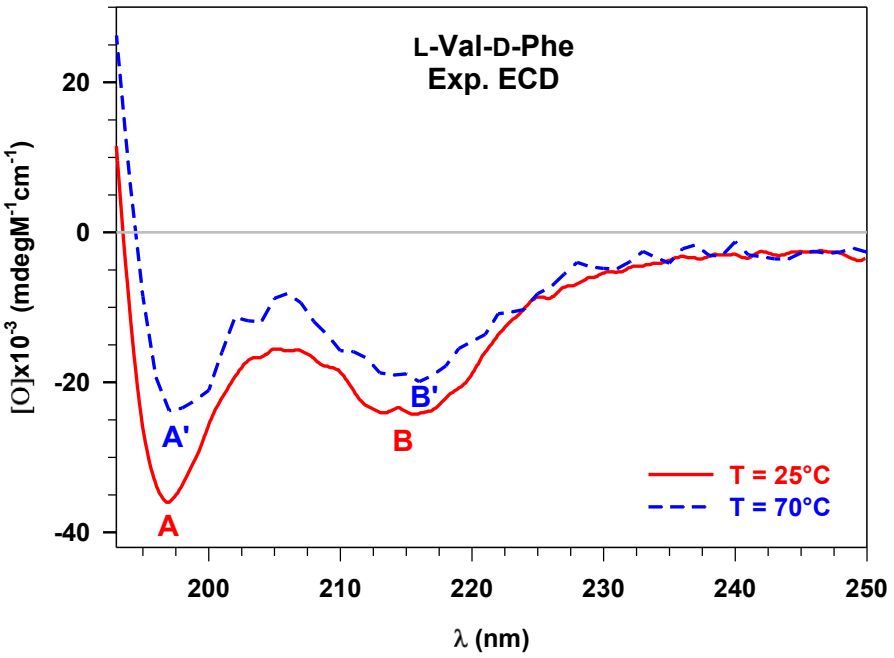


**Figure S10.** Experimental (Exp.) ECD spectra of L-Val-D-Phe in aqueous solution at 25°C (solid red line) and 70°C (dashed blue line), respectively.

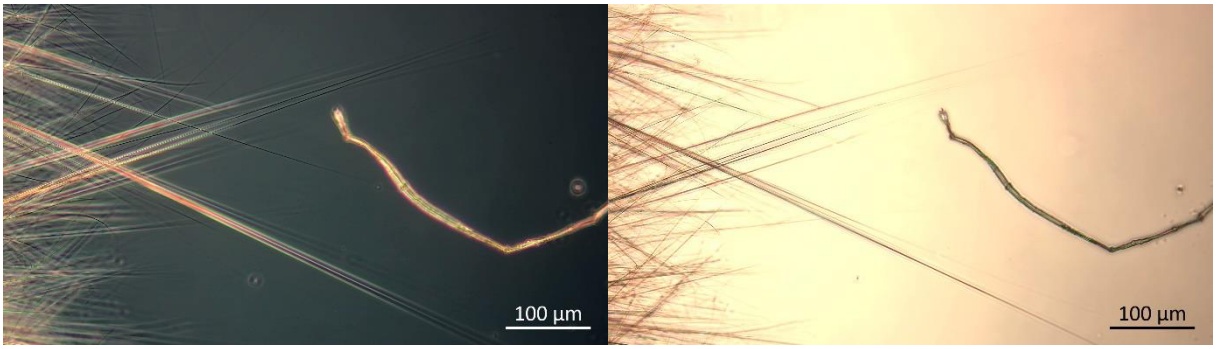


**Figure S11:** L-Val-D-Phe optical microscope images under polarised light (left) and normal light (right). During the collection of the fibre images after the high temperature experiment, the solution was left evaporate on the glass slide. Ordered crystalline needles appeared after several minutes on the left of the image. This highlights the different peptide behaviour at 25 °C.

## References

[1] O. Bellotto, G. Pierri, P. Rozhin, M. Polentarutti, S. Kralj, P. D'Andrea, C. Tedesco, S. Marchesan, *Org. Biomol. Chem.,* 2022, **20**, 6211.

[2] E. Scarel, G. Pierri, P. Rozhin, S. Adorinni, M. Polentarutti, C. Tedesco, S. Marchesan, *Chemistry,* 2022, **4**, 1417.

[3] O. Bellotto, S. Kralj, M. Melchionna, P. Pengo, M. Kisovec, M. Podobnik, R. De Zorzi, S. Marchesan, *ChemBioChem.,* 2022, **23**, e202100518.

[4] M. Monti, M. Stener, M. Aschi, *J. Comput. Chem.,* 2022, **43**, 1997.

[5] M. Monti, G. Brancolini, E. Coccia, D. Toffoli, A. Fortunelli, S. Corni, M. Aschi, M. Stener, *J. Phys. Chem. Lett.,* 2023, **14**, 1941.

[6] H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.,* 1995, **91**, 43.

[7] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhot, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.,* 2005, **26**, 1701.

[8] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.,* 1988*,* **110**, 1657.

[9] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996*,* **118**, 11225.

[10] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.,* 1983, **79**, 926.

[11] G. Bussi, M. Parrinello, *Comput. Phys. Commun.,* 2008, **179**, 26.

[12] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. Fraaije, *J. Comput. Chem.,* 1997, **18**, 1463.

[13] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.,* 1993, **98**, 10089.

[14] A. Amadei, B. M. Linssen, H. J. C. Berendsen, *Proteins,* 1993, **17**, 412.

[15] I. Daidone, A. Amadei, *WIREs Comput. Mol. Sci.,* 2012, **2**, 762.

[16] M. D'Alessando, A. Amadei, M. Stener, M. Aschi, *J. Comput. Chem.,* 2015, **36**, 399.

[17] R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules,* Oxford University Press, New York, Oxford, 1989.

[18] G. Te Velde, M. Bickelhaupt, E. J. Baerendsen, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.,* 2001, **22**, 931.

[19] J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.,* 2008, **10**, 6615.

[20] M. E. Casida, *Recent Advances in Density Functional Methods: (Part I) (World Scientifict),* 1995, 155.

[21] J. A. Schellman, *Chem. Rev.,* 1975, **75**, 323.

[22] O. Baseggio, M. De Vetta, G. Fronzoni, M. Stener, L. Sementa, A. Fortunelli, A. Calzolari, *J. Phys. Chem. C,* 2016*,* **120**, 12773.

[23] The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.