Diverging conformations guide dipeptide self-assembly into crystals or hydrogels†

M. Monti, [ID] [a] E. Scarel, [ID] [a] A. Hassanali,[b] M. Stener [ID] *[a] and S. Marchesan [ID] *[a]

The prediction of dipeptide assembly into crystals or gels is challenging. This work reveals the diverging conformational landscape that guides self-organization towards different outcomes. *In silico* and experimental data enabled deciphering of the electronic circular dichroism (ECD) spectra of self-assembling dipeptides to reveal folded or extended conformers as key players.

Self-assembly is a process through which building blocks form a supramolecular structure held together by non-covalent interactions. Several classes of molecules are known to undergo this process. Among them, short peptides have attracted much interest over the past 20 years due to their easy, modular, and low-cost syntheses, and their biocompatibility. Their use is being explored to develop smart, green materials with potential uses that span broadly from medicine to electronics.[1–6]

The most studied self-assembling dipeptide is diphenylalanine (Phe–Phe).[7] This motif was reported to form various nanomorphologies *e.g.*, nanotubes, nanowires, necklaces, and nanovesicles.[7–11] Its success has prompted modifications to derive new functionalities. For instance, the substitution of the N-terminal Phe with an aromatic N-cap,[12,13] or with another aliphatic amino acid (*e.g.*, Leu,[14] or Ile[15,16]) yields self-assembled hydrogels. The chirality of amino acids can have a drastic impact on the dipeptide ability to self-assemble, too.[14,16,17] In the Phe–Val case, only the heterochiral isomers form hydrogels, while the homochiral ones do not. This divergent behaviour was ascribed to the increased hydrophobicity of the heterochiral isomers. However, their ability to give rise to supramolecular water-channels – unlike their L-analogues – could play a role too,[17] as it is a feature shared by other dipeptide gelators, such as heterochiral Phe–Phe[18] and Phe–Ile.[16]

[a] *Chem. Pharm. Sc. Dept., University of Trieste, Via L. Giorgieri 1, Trieste 34127, Italy. E-mail: stener@units.it, smarchesan@units.it*
[b] *The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, Trieste 34151, Italy*

It is clear that several factors affect the ability of peptides to self-assemble. Consequently, the possibility to predict the supramolecular behaviour of an amino-acid sequence using computational approaches has also attracted much attention. Tuttle and co-workers have worked on di- and tri-peptides, screening all the possible combinations for aggregation propensity through coarse-grained molecular dynamics (MD).[19–22] Using this approach, several gelators were discovered. Coarse-grained and all-atomistic MD simulations were also used to model the self-assembly behaviour of classes of peptides and their derivatives.[23–25] Recently, machine-learning techniques were employed as predictive tools for the formation of peptide-based hydrogels.[26,27] To the best of our knowledge, the modelling of the aggregation of heterochiral dipeptides has not been explored thus far.

Over the years, recurrent structural features have been identified for dipeptide hydrogelators, such as hydrophobicity. One of the outstanding unanswered questions pertains to the molecular ingredients that drive self-assembly towards hydrogelation, as opposed to crystallisation. For instance, heterochirality was found to promote hydrogelation of Phe–Val and Phe–Ile.[16,17] However, changing the order of amino acids to Val–Phe and Ile–Phe,[16,17] or substituting Val for its linear isomer norvaline (Nva), led to crystallisation.[28] To shed light on the divergent supramolecular behaviour of such similar structures, here we investigate the conformational landscape of heterochiral (L,D, or D,L) dipeptides (Scheme 1) containing D-Phe and one aliphatic L-amino acid, *i.e.* Val, Nva, or Ile.

For this purpose, we adopted a computational approach[29] that allows to calculate accurately the electronic circular dichroism (ECD) spectrum, exploiting its ability to assess which are the main conformers in solution.[29,30] Extraction of the most significant conformations, through a combined MD-Essential Dynamics (ED) approach, enabled the calculation of a statistically averaged ECD spectrum, subsequently used to rationalise the corresponding experimental data.[16,17,28] Details on the procedure are reported in Section S2 of the ESI.† Remarkably, for each dipeptide, the conformers defining the ECD features in
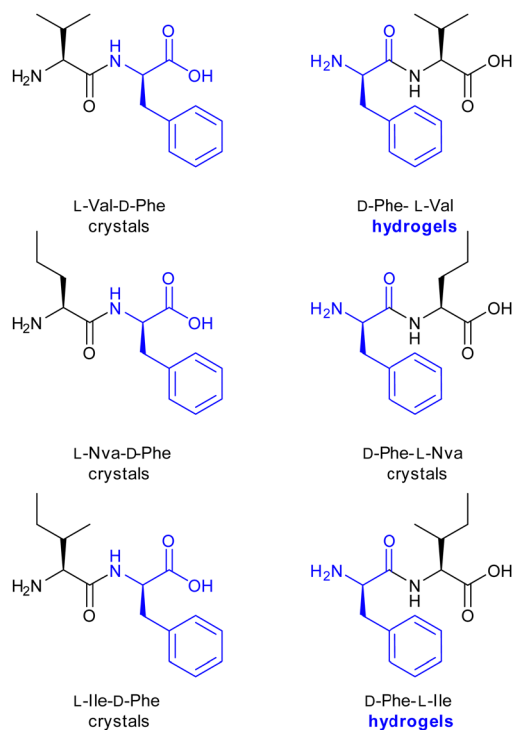
Scheme 1 Heterochiral dipeptide sequences with D-Phe and an L-aliphatic amino acid (*i.e.*, Val, Nva, or Ile) studied in this work to rationalise their supramolecular behaviour towards hydrogels or crystals in phosphate buffer at neutral pH.[16,17,28]
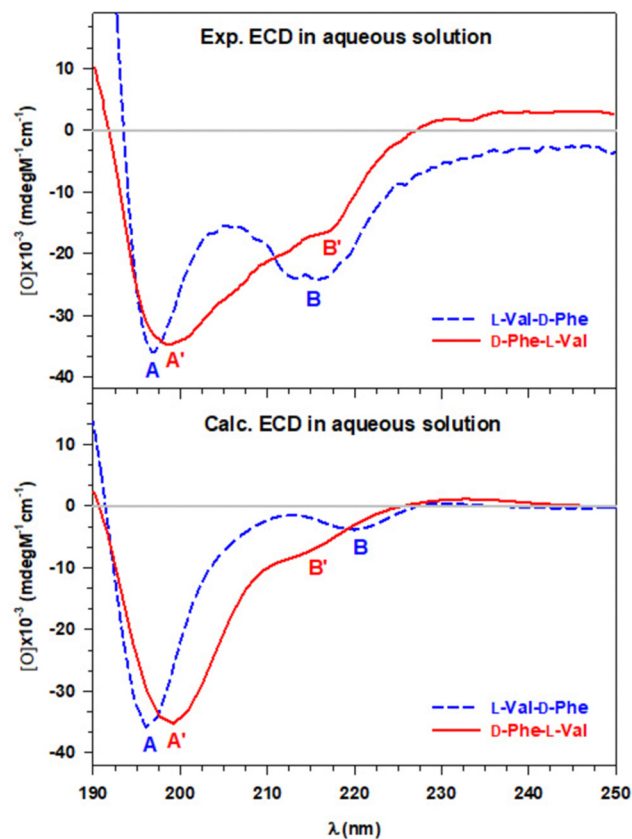

Fig. 1 Comparison between the experimental (Exp., top) and calculated (Calc., bottom) ECD in aqueous solution of L-Val-D-Phe (dashed blue line) and D-Phe-L-Val (solid red line). The Calc. ECD for L-Val-D-Phe and D-Phe-L-Val were shifted by +15 nm, and +0.7 nm, respectively. The intensity scales of the calc. ECD were normalised to match the exp. peak A/A'.

solution can be correlated with those observed in the solid-state (*vide infra*), *i.e.*, in the hydrogels or crystals.[16,17,28]

The first pair of D,L-dipeptides investigated in this work was L-Val-D-Phe and D-Phe-L-Val. The former readily crystallises in phosphate buffer at neutral pH, while the latter gels, then undergoes crystallisation after two days.[17] The comparison between the experimental and calculated ECD response in solution in Fig. 1 shows a satisfactory qualitative agreement, within the limits of the present model and the complexity of the real system. Indeed, the calculations reproduce the two minima (*i.e.*, A/B, and A'/B') observed for both peptides. Fig. 1 also points out that for L-Val-D-Phe, peaks A (197 nm) and B (216 nm) are well-defined and distinguished. By contrast, for D-Phe-L-Val, A' (198 nm) and B' (217 nm) are broader and coalesce together. In addition, peak A is narrower than peak A' in both the experimental and calculated spectra. The maximum at 205 nm is more defined for L-Val-D-Phe, than for D-Phe-L-Val. Hence, switching the amino acid position, we observe variations in the spectra reflecting the ability of ECD to capture conformational changes, even in such small systems.

In Fig. 2 we report the conformational landscapes obtained from the MD-ED analysis together with some of the most representative conformers ($\Delta G < 1.0$ kJ mol^{-1}). The space of L-Val-D-Phe (Fig. 2a) is defined by two shallow low-energy basins, A and B, which contain only folded conformations. Instead, the landscape of D-Phe-L-Val is more complex, with four stable basins, A–D, that are separated by lower energy barriers, relative to the L-Val-D-Phe case, and a fifth independent low-energy region, E.

The complexity of the space is reflected in the observed conformers, ranging from strongly folded in basin D, towards more or less extended in basins A and E. The same conformational analysis was performed for the other dipeptides in Scheme 1, and their free energy landscapes are collected in Fig. S1 and S2 (ESI†). All the crystallising peptides (*i.e.*, L-Nva-D-Phe, L-Ile-D-Phe, and D-Phe-L-Nva) show two main low-energy regions ($\Delta G \leq 1.0$ kJ mol^{-1}) in Fig. S1 (ESI†), resembling the L-Val-D-Phe conformational space. This is particularly true for L-Nva-D-Phe and L-Ile-D-Phe, while for D-Phe-L-Nva three additional small basins appear in the energy range of 1.5–2.5 kJ mol^{-1}. The conformational landscape of gelling D-Phe-L-Ile is defined by five low-energy regions (Fig. S2, ESI†), in agreement with D-Phe-L-Val.

Remarkably, the low-energy ($\Delta G \leq 2.0$ kJ mol^{-1}) conformations extracted for the crystallising dipeptides (Fig. S3 and Table S1, ESI†) are exclusively folded, closely resembling those of the corresponding crystal-unit structures.[16,17,28] For D-Phe-L-Nva, we found a small number of stretched conformations associated with a ΔG of 2.1–2.5 kJ mol^{-1} (basin E, Fig. S4, ESI†). Notably, this dipeptide does transiently gel in MeCN.[28]

We repeated the same procedure of overlapping the MD and crystal structures for the two hydrogelators[16,17] shown in Fig. S5 (ESI†). Here, both adopt folded conformations, in agreement
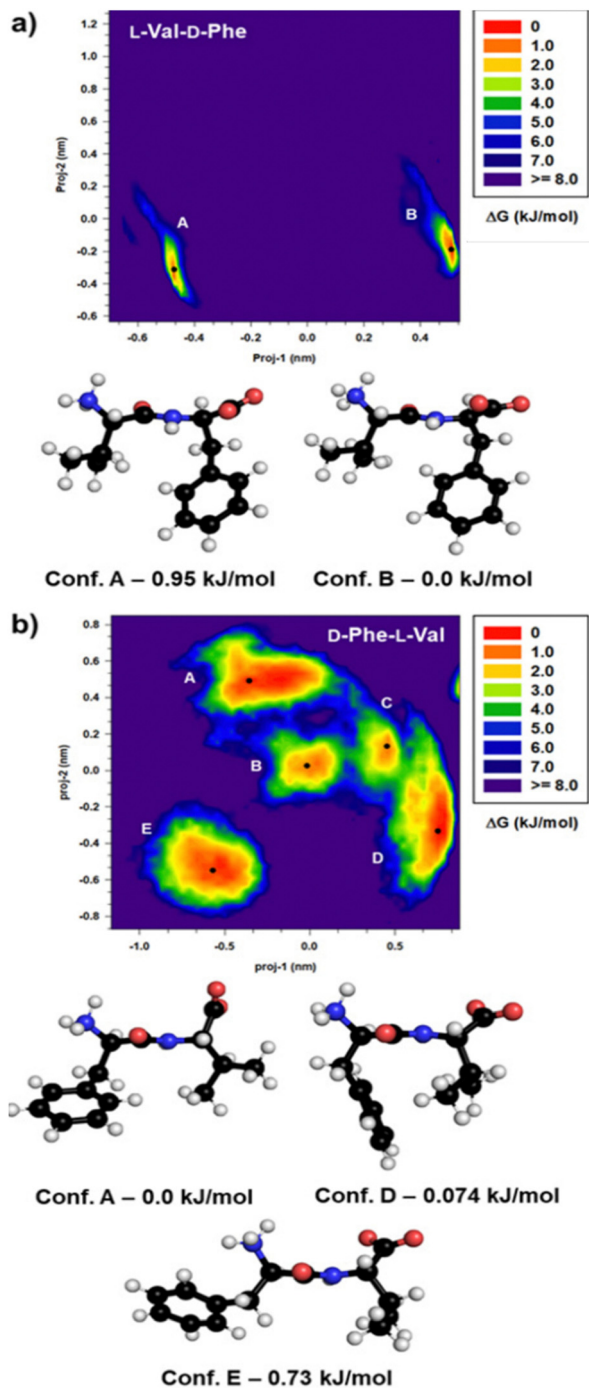
Fig. 2 Relative free energy landscape in the essential plane of L-Val-D-Phe (panel a, top) and D-Phe-L-Val (panel b, bottom) at room temperature (RT). The energy scale, in kJ mol^{-1}, is reported as a vertical colour bar. All of the low-energy regions are labelled with capital letters. The most prevalent conformations with $\Delta G \leq 1.0$ kJ mol^{-1} extracted from these basins are shown in the bottom panel.

with the crystal structure, and extended ones which deviate from them. Interestingly, the conformational equilibrium for the gelators is shifted towards the extended structures.

The clear effect of different conformations on the spectral features is also noteworthy. As shown in Fig. S6 (ESI†) for

D-Phe-L-Val, differences in the energy, intensity, and even sign of the peaks arise because of both the secondary structure, and the solvation shell. ECD confirms to be very sensitive to even small, conformational changes and solvent effects.[29]

This conformational analysis suggests that the systems that crystallise assume folded conformations very quickly in solution similar to those found in the crystal structures. In contrast, hydrogelators are mostly defined by β-strand-like conformations that favour fibrillation.[31] However, the folded conformations, which are still significant, can play a role in proceeding towards the thermodynamically stable phase over days. Therefore, the conformational landscape of a single peptide in solution may indicate its solid-state evolution.

To assess similarities between solution and crystal phases, we also calculated the ECD of models composed of four unit cells for L-Val-D-Phe, D-Phe-L-Val, and L-Nva-D-Phe. All the spectra calculated from the crystals, reported in Fig. S7 and S8 (ESI†), are defined by two minima, thus resembling the spectral features in solution and supporting our analysis above.

Analysis of the conformational distribution is essential to justify and predict why and whether a certain sequence undergoes gelation. Detailed investigations on the conformational preferences of short peptides are available.[32–35] Folded conformers are favoured enthalpically, whereas extended ones are entropically-driven. Thus, electrostatic, steric, hydrophobic, and thermal effects, as well as the stabilisation through intra-, and inter-molecular H-bonds play a role in the folded-extended equilibrium. For the dipeptides in Scheme 1, we suggest that repulsion takes place when Phe is at the C-terminus, contributing to the folding of the aromatic ring, while the $-NH_3^+-\pi$ attractive interaction favours the opening of N-terminal Phe. Cation–π interactions promote peptide gelling.[36] In addition, C-terminal bulky residues with β-branching (e.g., Val, Ile) can destabilise the folded conformer, which is instead favoured with the linear Nva.

To validate these findings, we heated samples to promote a conformational change towards fibrils, as opposed to crystals. We first measured the ECD spectra at 70 °C for L-Nva-D-Phe and L-Val-D-Phe (Fig. S9 and S10, ESI†). Our computational procedure was repeated for the former case, and the calculated ECD is shown in Fig. S9 (ESI†). In both cases, the chiroptical features (peaks B/B′ for L-Nva-D-Phe, and peaks A/A′, and B/B′ for L-Val-D-Phe) are reduced, being indicative of a conformational change, which resulted in a differing solid phase, i.e. fibres (Fig. 3 and Fig. S11, ESI†).

In summary, this work sought a correlation between the conformational behaviour in aqueous solution and the solid-state structures of D,L Phe-based dipeptides. Combining previous[16,17,28] and new experimental data (i.e., ECD spectra, and microscopy data) with a recent computational procedure,[29] we rationalised the chiroptical features that were observed by identifying representative conformers.

The crystallising dipeptides (i.e., L-Val-D-Phe, L-Nva-D-Phe, L-Ile-D-Phe, and D-Phe-L-Nva) are defined by conformational landscapes with two main low-energy regions containing only folded conformers that resemble their crystal units. To assess
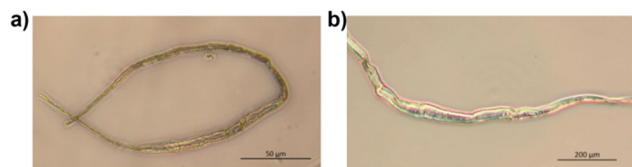
Fig. 3 Optical microscope images of fibres under normal light for L-Nva-D-Phe (panel a, left) and L-Val-D-Phe (panel b, right) after the assembly at 70 °C. Magnification 20× (left) and 10× (right).

the similarities between structures in solution and in the crystal, we calculated ECD spectra on four unit cell models, obtaining an optical response in qualitative agreement with the experimental one in solution, particularly for L-Nva-D-Phe.

In contrast, complex conformational landscapes characterise the peptides that undergo gelation (*i.e.*, D-Phe-L-Val, and D-Phe-L-Ile) followed by crystallisation. In these latter cases, both folded and extended conformers are present, with the latter ones predominating and likely guiding gelation. However, the presence of folded conformers in solution could be responsible for the observed transition from gel to crystal over days, as confirmed by the structural similarity between the calculated folded conformers and those found in the crystal by XRD. Finally, heating drove a conformational change from folded to extended for crystallising dipeptides, as confirmed by ECD and fibrillation (Fig. 3).

The agreement between experimental and *in silico* data validates the utility of gaining key insights about dipeptide assembly in the solid state by investigating the conformational preferences in solution. Extension of this approach to other peptides could offer a tool to predict crystallisation or gelation based simply on their ECD response, potentially enabling more rapid progress in the area of smart biodegradable materials.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 H. Wang, Z. Feng and B. Xu, *Angew. Chem., Int. Ed.*, 2019, **58**, 10423.
2 E. Radvar and H. S. Azevedo, *Macromol. Biosci.*, 2019, **19**, e1800221.
3 G. Uzunalli and M. O. Guler, *Ther. Delivery*, 2020, **11**, 193.
4 M. G. Herrera and V. I. Dodero, *Biophys. Rev.*, 2021, **13**, 1147.
5 D. Giuri, P. Ravarino and C. Tomasini, *Org. Biomol. Chem.*, 2021, **19**, 4622.
6 M. Amit, S. Yuran, E. Gazit, M. Reches and N. Ashkenasy, *Adv. Mater.*, 2018, **30**, 1707083.
7 M. Reches and E. Gazit, *Science*, 2003, **300**, 625.
8 X. Yan, Q. He, K. Wang, L. Duan, Y. Cui and J. Li, *Angew. Chem., Int. Ed.*, 2007, **46**, 2431.
9 J. Ryu and C. B. Park, *Angew. Chem., Int. Ed.*, 2009, **48**, 4820.
10 S. Yuran, Y. Razvag and M. Reches, *ACS Nano*, 2012, **6**, 9559.
11 A. N. Rissanou, E. Georgilis, E. Kasotakis, A. Mitraki and V. Harmandaris, *J. Phys. Chem. B*, 2013, **117**, 3962.
12 A. M. Garcia, R. Lavendomme, S. Kralj, M. Kurbasic, O. Bellotto, M. C. Cringoli, S. Semeraro, A. Bandiera, R. De Zorzi and S. Marchesan, *Chem. – Eur. J.*, 2020, **26**, 1880.
13 S. Roy and A. Banerjee, *Soft Matter*, 2011, **7**, 5300.
14 O. Bellotto, S. Kralj, R. De Zorzi, S. Geremia and S. Marchesan, *Soft Matter*, 2020, **16**, 10151.
15 N. S. de Groot, T. Parella, F. X. Aviles, J. Vendrell and S. Ventura, *Biophys. J.*, 2007, **92**, 1732.
16 O. Bellotto, S. Kralj, M. Melchionna, P. Pengo, M. Kisovec, M. Podobnik, R. De Zorzi and S. Marchesan, *ChemBioChem*, 2022, **23**, e202100518.
17 O. Bellotto, G. Pierri, P. Rozhin, M. Polentarutti, S. Kralj, P. D'Andrea, C. Tedesco and S. Marchesan, *Org. Biomol. Chem.*, 2022, **20**, 6211.
18 S. Kralj, O. Bellotto, E. Parisi, A. M. Garcia, D. Iglesias, S. Semeraro, C. Deganutti, P. D'Andrea, A. V. Vargiu, S. Geremia, R. De Zorzi and S. Marchesan, *ACS Nano*, 2020, **14**, 16951.
19 P. W. J. M. Frederix, R. V. Ulijn, N. T. Hunt and T. Tuttle, *J. Phys. Chem. Lett.*, 2011, **2**, 2380.
20 P. W. J. M. Frederix, G. G. Scott, Y. M. Abul-Haija, D. Kalafatovic, C. G. Pappas, N. Javid, N. T. Hunt, R. V. Ulijn and T. Tuttle, *Nat. Chem.*, 2015, **7**, 30.
21 A. Lampel, R. V. Ulijn and T. Tuttle, *Chem. Soc. Rev.*, 2018, **47**, 3737.
22 A. van Teijlingen, M. C. Smith and T. Tuttle, *Acc. Chem. Res.*, 2023, **56**, 644.
23 O. S. Lee, S. I. Stupp and G. C. Schatz, *J. Am. Chem. Soc.*, 2011, **133**, 3677.
24 O. S. Lee, V. Cho and G. C. Schatz, *Nano Lett.*, 2012, **12**, 4907.
25 P. Divanach, E. Fanouraki, A. Mitraki, V. Harmandaris and A. N. Rissanou, *J. Phys. Chem. B*, 2023, **127**, 4208.
26 F. Li, J. Han, T. Cao and L. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 11259.
27 R. Batra, T. D. Loeffler, H. Chan, S. Srinivasan, H. Cui, I. V. Korendovych, V. Nanda, L. C. Palmer, L. A. Solomon, H. C. Fry and S. K. R. S. Sankaranarayanan, *Nat. Chem.*, 2023, **14**, 1427.
28 E. Scarel, G. Pierri, P. Rozhin, S. Adorinni, M. Polentarutti, C. Tedesco and S. Marchesan, *Chemistry*, 2022, **4**, 1417.
29 M. Monti, M. Stener and M. Aschi, *J. Comput. Chem.*, 2022, **43**, 1997.
30 N. Berova, L. Di Bari and G. Pescitelli, *Chem. Soc. Rev.*, 2007, **36**, 914.
31 K. Sivanesam and N. Andersen, *Arch. Biochem. Biophys.*, 2019, **664**, 51.
32 W. Chin, F. Piuzzi, I. Dimicoli and M. Mons, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1033.
33 Y. Loquais, E. Gloaguen, S. Habka, V. Vaquero-Vara, V. Brenner, B. Tardivel and M. Mons, *J. Phys. Chem. A*, 2015, **119**, 5932.
34 W. Ji, C. Yuan, P. Chakraborty, S. Gilead, X. Yan and E. Gazit, *Commun. Chem.*, 2019, **2**, 65.
35 T. Vermeyen and C. Merten, *Phys. Chem. Chem. Phys.*, 2020, **22**, 15640.
36 C. C. Cole, M. Misiura, S. A. H. Hulgan, C. M. Peterson, J. W. Williams III, A. B. Kolomeisky and J. D. Hartgerink, *Biomacromolecules*, 2022, **23**, 4645.