PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE

Proline cis-trans isomerization and its implications for the dimerization of analogues of cyclopeptide stylostatin 1: a combined computational and experimental study

www.Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

C. López-Martínez,[§a] P. Flores-Morales,[§b] M. Cruz,[a] T. González,[c] M. Feliz,[c] A. Diez,[a] and Josep M. Campanera*[d]

Cis and *trans* proline conformers are often associated with dramatic changes in the biological function of peptides. A slow equilibrium between *cis* and *trans* Ile-Pro amide bond conformers occurs in constrained derivatives of the native marine cyclic heptapeptide stylostatin 1 (*cyclo*-(NSLAIPF)); a potential anticancer agent. In this work, four cyclopeptides, *cyclo*-(NSTAIPF), *cyclo*-(KSTAIPF), *cyclo*-(RSTAIPF) and *cyclo*-(DSTAIPF), which are structurally related to stylostatin 1, are experimentally and computationally examined in order to assess the effect of residue mutations on the *cis*–*trans* conformational ratio and the apparent capacity to form dimeric aggregates. Primarily, *cyclo*-(KSTAIPF) and *cyclo*-(RSTAIPF) showed specific trends in circular dichroism, MALDI-TOF and HPLC purification experiments which suggests the occurrence of peptide dimerization. Meanwhile, NMR spectrum of *cyclo*-(KSTAIPF) indicates that this cyclopeptide exists in the two slow-exchange families of conformations mentioned above. Molecular dynamics simulations combined with quantum mechanical calculations have shed light on the factors governing the *cis*/*trans* conformational ratio. In particular, we have found that residue mutations affect the internal hydrogen bond pattern which ultimately tunes the *cis*/*trans* conformational ratio and that only *trans* conformers are capable of aggregating due to the shape complementarity of the two subunits.

1. Introduction

Naturally occurring cyclopeptides represent a rich source of bioactive compounds with unusual and potent biological activities. The size and chemical complexity of cyclic peptides makes them suitable scaffolds to target shallow and extended binding interfaces, including targets with a low-druggability such as protein–protein interactions.[1,2] These compounds fall outside the "drug-likeness" concept generally applied to small organic molecules; but they do define a chemical space that may provide useful scaffolds in drug discovery.[3,4] Compared to linear peptides, cyclic peptides have enhanced therapeutic potential, as they have improved metabolic stability, cell permeability and oral bioavailability.[5,6] Furthermore, cyclization should increase binding affinity by eliminating unproductive conformations that are accessible for linear peptides.

Conformational flexibility is a key factor when it comes to understanding the biological role of cyclopeptides.[7] The choice of cyclization strategy, the presence of non-proteinogenic units, the polarity of the solvent and the physicochemical properties conferred by residues may be exploited to control the conformational behaviour,[8,9,10] which in the case of cyclopeptides containing Pro can be drastically affected by the *cis*-*trans* isomerization.[11] In proteins, approximately 6% of X-Pro peptide bonds are found in the *cis* conformation, mostly lying close to the surface.[12,13] Pro-Pro, aromatic-Pro and Pro-aromatic sequences have the highest propensities to form a *cis*-prolyl amide bond.[14] Due to their tendency to induce turns, the impact of Pro residues on the constrained conformation of cyclic peptides, and hence on their biological activity, is crucial.

For instance, while the two Pro residues in euryjanicin D (*cyclo*-(SIPLFPI) abbreviated to c(SIPLFPI) hereafter) are found in the *cis* conformation,[15] the two Pro residues in its structural isomer stylopeptide 1 (c(PLIFSPI)) are found in *trans* and *cis* conformations.[16] Furthermore, the cyclic heptapeptide ceratospongamide, which contains two Pro residues, is found in both the *cis*,*cis* and *trans*,*trans* conformations, but only the latter leads to potent inhibition of phospholipase A2.[17] Moreover, residue mutations can affect the *cis/trans* conformational ratio, as exemplified by the fact that replacement of the serine residue in the cyclic peptide

[a] *Laboratori de Química Orgànica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, 08028 Barcelona, Spain.*
[b] *CREA-Química, Facultad de Ciencias Químicas, Universidad de Concepción, Concepción, Chile.*
[c] *NMR Facility, Scientific and Technical Centers, University of Barcelona, 08028 Barcelona, Spain.*
[d] *Departament de Fisicoquimica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, 08028 Barcelona, Catalonia, Spain. Email: campanera@ub.edu*
[§] These authors contributed equally to this work.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

sequences c(-dXaa-Ser-Pro-dXaa-Lys-Pro-), by either phosphoserine or tyrosine, increases the *cis* conformation content of both prolines.[18] Likewise, the residue sequence has also been reported to have a decisive influence on the *cis/trans* ratio in contryphans, a 23-membered cyclic disulfide moiety.[19]

In previous studies,[20] we used the cyclic heptapeptide stylostatin 1 (c(NSLAIPF))[21] as a template for the synthesis of pseudostylostatins to be studied as potential anticancer agents. In those compounds, the Ser-Leu pair was replaced by a lactam (piperidone) dipeptide to modulate the molecular conformation.[22] Both NMR and X-ray analysis showed that stylostatin 1 is found in a *cis* Ile-Pro conformation, which is stabilized by three internal hydrogen bonds.[20,21] However, inclusion of the lactam dipeptide yielded pseudostylostatins that adopted both *cis* and *trans* conformations. Although the turn centred on the Ser-Leu dipeptide was successfully mimicked by the piperidone surrogate, it induced the destabilization of the internal hydrogen bonds found in stylostatin 1, and the increased flexibility allowed for the onset of the Ile-Pro *trans* conformation. In contrast, replacement of Ile by D-*allo*-Ile only rendered the *trans* conformation; this probably reflects the steric effect around the α-carbon of D-*allo*-Ile on the D-*allo*-Ile-Pro bond. Overall, these findings revealed the high structural sensitivity of stylostatin 1 to seemingly minor structural alterations.

Here, we now present combined experimental and theoretical work[23] on the factors governing the *cis-trans* isomerization of the stylostatin c(NSLAIPF) and four other pseudostylostatins, namely c(NSTAIPF), c(KSTAIPF), c(RSTAIPF) and c(DSTAIPF). Our principal aim is to establish the structural causes that determine the *cis/trans* Ile-Pro peptide bond conformational ratio and the preferences of these conformers to form aggregates. We are specifically interested in determining how the mutated residues affect the internal hydrogen bonding pattern, which we hypothesize constitutes the main determinant of the *cis-trans* isomerization.[20] Of the five pseudostylostatins, *c*(KSTAIPF) showed a remarkable capacity to form dimeric aggregates and thus was taken as the reference compound for an in-depth study.

The present study uses circular dichroism (CD), MALDI-TOF, HPLC and NMR techniques to experimentally characterize the conformational preferences, and extensive molecular dynamics (MD) simulations (around 6 μs in explicit waters) to elucidate the structure of *cis* and *trans* confomers of the monomer cyclopeptides as well as the structure of the dimer aggregates. Finally, DFT/M06-2X calculations are performed for representative members of the MD conformational ensembles to assess the relative stability of the *cis* and *trans* conformers, and to offer insight into the structural and physicochemical determinants that underlie the *cis*/*trans* conformational ratio and the capacity to form dimeric aggregates.

2. Results and Discussion

2.1. Design and synthesis of the pseudostylostatins

A set of four new cyclic peptides was designed to potentially modify the *cis*/*trans* isomerization ratio of the native stylostatin 1 (c($N^1S^2L^3A^4I^5P^6F^7$) with the numbering of the amino acid residues), which is also included in the present study (Figure 1). The pseudostylostatins were generated by replacing the residues at positions 1 (the position at which one of the internal hydrogen bonds is formed) and 3 by residues that incorporate additional capacity to form hydrogen bonds. So, the mutation of the apolar residue Leu^3 in c(NSLAIPF) by Thr, giving rise to *c*(NSTAIPF). The remaining three compounds are designed by replacing the neutral residue Asn^1 by the charged residues Lys (*c*(KSTAIPF)), Arg (*c*(RSTAIPF)) and Asp (*c*(DSTAIPF)). Apart from exploring the different possibilities for altering the intramolecular hydrogen pattern, these mutations were also motivated by the search for cationic antimicrobial peptides (AMPs), which exhibit cytotoxic activity against cancer cells.[24] Such AMPs require the presence of a positively charged residue (the mutation at position 1 of Asn to Lys and Arg in contrast to Asp) and also a certain amphiphilic character (mutation at position 3 from the hydrophobic residue Leu to the polar residue Thr).

The peptides were synthesized by closure of the Pro^6-Phe^7 bond (Figure S1; see also Experimental Section). We used 2-chlorotrityl (ClTrt) resin, which facilitated the anchoring of Pro as the first amino acid, avoided diketopiperazine formation and enabled cleavage in mild conditions, thus allowing the side chains to remain protected for subsequent cyclization in solution. Since ClTrt resin is sensitive to acid media, for each step we first verified the completion of the coupling (Kaiser test), and then deprotected the amino group with 20% piperidine in dimethylformamide. This procedure was found to improve the synthetic yield compared to our previous work.[20] Commercial ClTrt resin with Pro already anchored as the first amino acid (H-Pro-2-ClTrt, Novabiochem, 0.3-0.9 mmol/g of resin, ref. 04-12-2805) gave the best results. Natural amino acids were coupled using Fmoc-aa-OH/HOBt/DIPCDI (3:3:3). In order to remove the protective groups from the side chains, a final deprotection step was performed using TFA/$H_2$0 (95:5).

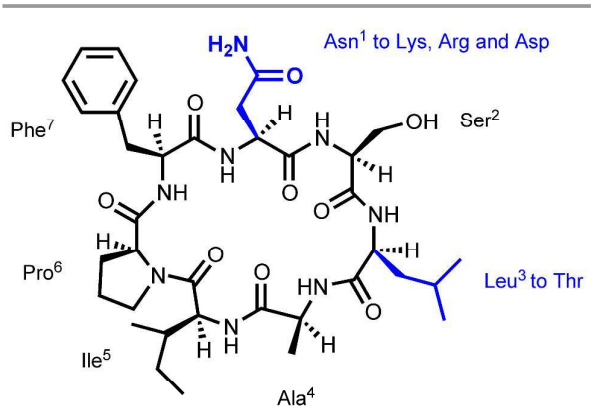
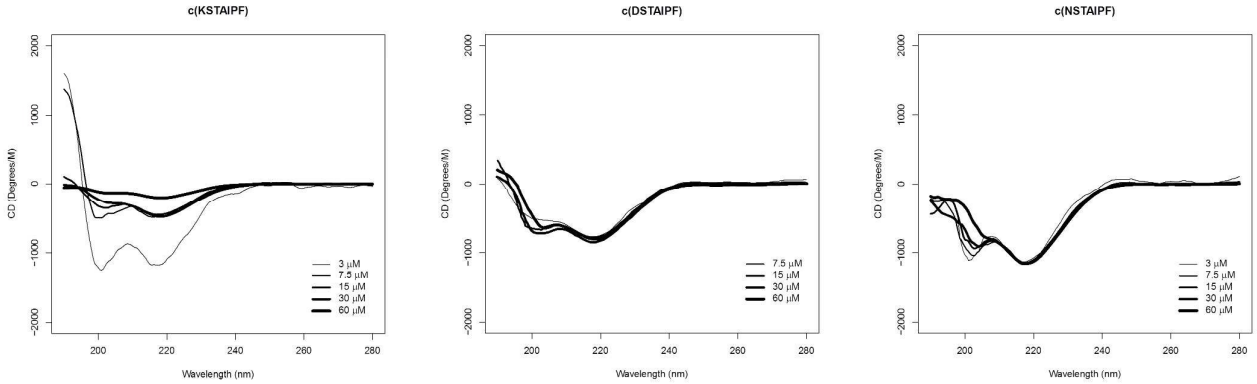


Figure 1. Chemical structure of c(NSLAIPF), also named stylostatin 1. c(NSTAIPF), c(KSTAIPF), c(RSTAIPF) and c(DSTAIPF) derive from mutations of the amino acid residues at position 1 and 3 (highlighted in blue).



(a) CD pattern with clear alteration as the concentration increases

(b) CD pattern with any significant alteration

(c) CD pattern with intermediate behaviour

Figure 2. CD spectra of (a) c(KSTAIPF), (b) c(DSTAIPF) and (c) c(NSLAIPF) recorded at increasing cyclopeptide concentration. The CD spectra of the remaining cyclopeptides can be found in Figure S2.

All the compounds were purified by reverse-phase HPLC. Cyclopeptides were thus obtained in chemical yields of 12%-20% and with HPLC purities of 96%-98%. Finally, because the peptides were highly hygroscopic and difficult to manipulate, they were characterized by standard HPLC, amino acid analysis and MALDI-TOF mass spectrometry.

2.2 Electronic circular dichroism (CD) and MALDI-TOF analysis

Electronic CD spectra were recorded in water solutions at various concentrations for all the compounds. The CD spectra of c(NSLAIPF), c(KSTAIPF) and c(DSTAIPF) (Figure 2), chosen as illustrative examples, show features similar to those found in the CD spectra of other cyclopeptides, such as c(GPSG-(δ)Ava) and c(FpGRGD), where (δ)Ava and p mean 5-amino valeric acid and D-Pro, respectively.[25,26] The spectra show negative bands at around 215-220 and 200-205 nm, which correspond to the n→π* and π→π* excitons, respectively; and a positive band below 200 nm. These features, which resemble those found in α-helix CD spectra but have smaller amplitudes, suggest the presence of a type I β-turn (class C spectra following Woody's nomenclature).[27] This probably reflects the turn geometry of the X^1SXA^4 segment and its analogous NSLA segment in c(NSLAIPF), which possesses such a conformational motif in its structure, as determined by X-ray and NMR analysis.[20,21]

The CD spectra showed three types of patterns when the cyclopeptide concentration was varied in the range 3-60 μM (see Figure 2 for illustrative examples and Figure S2 for the remaining cyclic peptides). On the one hand, the c(DSTAIPF) spectrum did not show any relevant changes as the cyclopeptide concentration increased (Figure 2b); while the c(RSTAIPF) and especially c(KSTAIPF) spectra were significantly altered as the concentration was increased to 60 μM, leading to a flattening of the spectra in the 200-230 nm region (Figure 2a). On the other hand, c(NSLAIPF) and c(NSTAIPF) presented intermediate behaviour between the two abovementioned patterns (Figure 2c), with a slight alteration of the CD spectra due to the increase of the concentration. This finding in c(RSTAIPF) and c(KSTAIPF) suggested that the increase in concentration promotes a structural alteration that could stem from either self-aggregation[28] or the conversion of one conformation of the cyclopeptide to another.[29]

The former interpretation is supported by the analysis of MALDI-TOF spectra, which showed that all the compounds exhibit the expected molecular peaks $[M+H]^+$, $[M+Na]^+$ or $[M+K]^+$, but that c(KSTAIPF), c(RSTAIPF) and c(NSLAIPF) also showed a $[2M+H]^+$ peak, thus indicating the existence of dimeric species for these cyclopeptides (see Figure 3 for c(KSTAIPF) and Figure S3 for the remaining cyclopeptides). The ambiguous results for c(NSLAIPF) might be related to the different sensitivity of the CD and MALDI-TOFF techniques. It could also be related to the slightly different experimental conditions used in the two techniques, which could displace the subtle equilibrium between the monomeric and the dimeric forms in a peptide that clearly shows intermediate

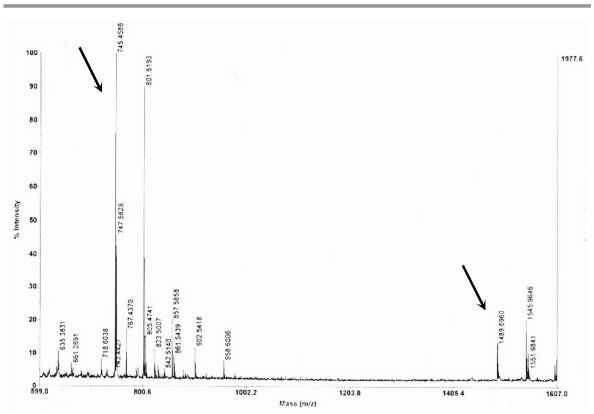


Figure 3. MALDI-TOF spectrum recorded for c(KSTAIPF). The marked peaks are assigned to the monomer of c(KSTAIPF) (745.46 m/z) and to the corresponding dimer (1489.90 m/z). The peak at 801.52 m/z corresponds to c(KSTAIPF) with the addition of the tBu group. The peak at 1977.60 is a reference with the highest intensity. The MALDI-TOF spectra of the remaining cyclopeptides can be found in Figure S3.

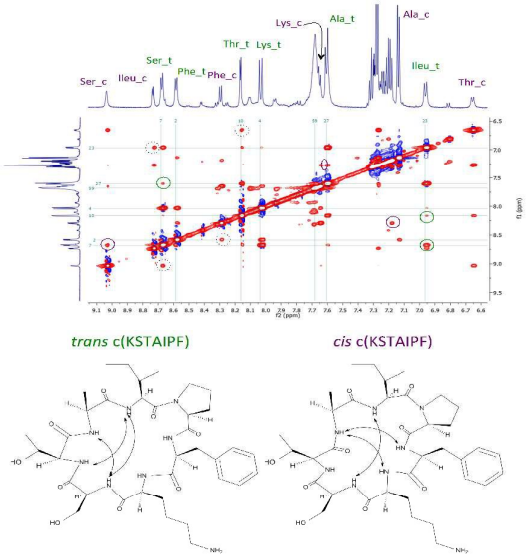


Figure 4. (Top) NOESY spectrum of c(KSTAIPF). (Bottom) Relevant NOE correlations of (left) *trans* Ile-Pro and (right) *cis* Ile-Pro conformers. The two sets of signals are indicated by labels _t (*trans*; green) and _c (*cis*; purple).

behaviour between those that aggregate and those that do not. In line with the previous findings, HPLC purification of c(KSTAIPF) revealed the presence of two major peaks and several smaller peaks that lie close together at room temperature, and also the coalescence of these peaks upon heating to 45ºC (Figure S4). On the basis of the previous findings, it can be speculated that the peaks reflect an equilibrium between distinct conformational states and even aggregated species, which would appear as a single peak in the chromatogram when the temperature is raised. Finally it is important to highlight that the atypical behaviour, especially of c(KSTAIPF) and c(RSTAIPF), cannot be attributed merely to overall physicochemical properties; rather we hypothesize that it could reflect the alteration of the internal hydrogen pattern caused by each specific residue mutation.

2.3. NMR study of c(KSTAIPF)

To gain further insight into the above findings, we undertook ^{1}H and ^{13}C NMR analysis of c(KSTAIPF) in DMSO. Unfortunately the NMR analysis could not be performed in water as the other experiments were since the limited solubility of the cyclopeptides in water impeded to achieve the concentration required for NMR analysis.[30] The spectra showed two main sets of signals, which were identified as the *cis* and *trans* Ile-Pro conformations based on the characteristic ^{13}C chemical shift difference between β- and γ-carbon atoms ($\Delta\delta_{\beta,\gamma}$) of *cis* X-Pro ($\Delta\delta_{\beta,\gamma} \geq 8$ ppm) and *trans* X-Pro ($\Delta\delta_{\beta,\gamma} < 6$ ppm) dipeptides, a method pioneered by Siemion et al.[31,32,33] The higher set of signals showed a $\Delta\delta_{\beta,\gamma}$ of 4.6 ppm, indicating that the major conformation contains a *trans* Ile-Pro; while the

Table 1. Phi and psi backbone torsional angles (degrees) and distances (Å) of β- and γ-turns in *cis* and *trans* c(KSTAIPF).[a]

ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}	d_1[b]	d_2[c]
cis					
Type I β-turn: K^1_i-S^2_{i+1}-T^3_{i+2}-A^4_{i+3}					
-62±11 (-60)	-27±13 (-30)	-101±19 (-90)	-18±19 (0)	5.0±0.2 (<7)	3.2±0.3
Type VIa1 β-turn: A^4_i-I^5_{i+1}-P^6_{i+2}-F^7_{i+3}					
-69±12 (-60)	146±8 (120)	-86±8 (-90)	-10±15 (0)	6.2±0.4 (<7)	3.2±0.3
trans					
Type II β-turn: S^2_i-T^3_{i+1}-A^4_{i+2}-I^5_{i+3}					
-63±13 (-60)	138±17 (120)	59±8 (80)	6±20 (0)	5.9±0.3 (<7)	3.7±0.4
Standard γ-turn: F^7_i-K^1_{i+1}-S^2_{i+2}					
56±10 (80)	-9±34 (-65)	-	-	5.8±0.2 (<7)	2.9±0.2
Inverse γ-turn: I^5_i-P^6_{i+1}-F^7_{i+2}					
-67±11 (-80)	-9±20 (65)	-	-	5.7±0.2 (<7)	3.1±0.2

[a] Ideal values are reported in parenthesis (ref. 35).
[b] Distance between C_α carbons of residues i and i+2 (γ-turn) or i+3 (β-turn).
[c] Distance from carbonyl oxygen of residue i and the nitrogen of the NH peptide unit of residue i+2 (γ-turn) and i+3 (β-turn). See Table S2 for the values of other cyclopeptides.

secondary set of signals ($\Delta\delta_{\beta,\gamma}$ = 9.7 ppm) corresponds to the *cis* Ile-Pro conformation. Notably, Diez et al. measured a value of 9 ppm for $\Delta\delta_{\beta,\gamma}$ of c(NSLAIPF),[20] which indicated a *cis* conformation of the Ile-Pro amide bond, in agreement with the reported X-ray structure.[21] In this previous work, they also reported that an increase of the proportion of DMSO decreases the *cis*/*trans* conformational ratio in constrained derivatives of c(NSLAIPF), as a solvent-induced isomerization effect,[34] which is in line with the present findings that the *trans* conformer dominates over the *cis* conformer in pure DMSO; though in water, the *cis* conformer might be the more populated one, as in the case of c(NSLAIPF).

The structural analysis of the *trans* conformer showed NOE correlations[35] not only between vicinal residues, but also between NH protons of the pairs Ser-Ile, Thr-Ile and Ala-Ser (green circles in Figure 4 and Table S1). This indicates that the cyclopeptide may adopt a conformation that contains a β-turn involving the Ser^2_i-Thr^3_{i+1}-Ala^4_{i+2}-Ile^5_{i+3} fragment and consequently an intramolecular bond between the CO of Ser^2 and the HN of Ile^5. For the *cis* Ile-Pro conformer, NOE correlations between NH protons of the pairs Phe-Ala, Ala-Lys and Ser-Ile were found (purple circles in Figure 4 and Table S1). This could be explained by two extra hydrogen bonds between the NH groups of Ala and Phe and the carbonyl units of Lys and Ala, respectively, which would stabilize the backbone conformation by forming two β-turns. The IR spectrum of c(KSTAIPF) resembles to that of an antiparallel β-sheet, which

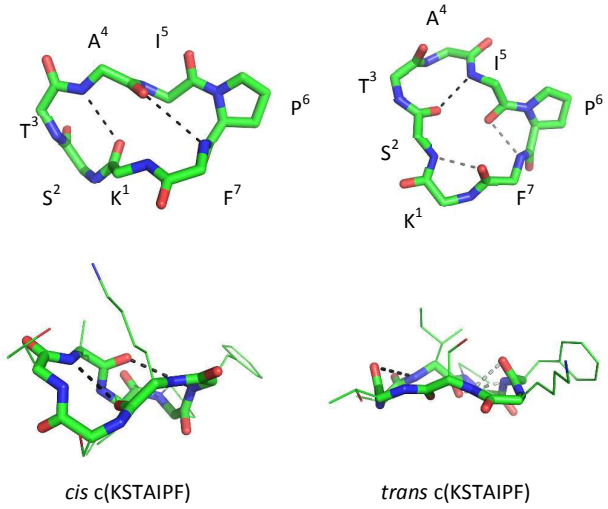


Figure 5. Two views of the molecular representation of a representative structure for (left) *cis* and (right) *trans* c(KSTAIPF) with the residue numbering. Top views do not show side chains except for proline whereas bottom views incorporate side chain structures. Hydrogen bonds of the β-turns are denoted by black lines and those of γ-turns by grey lines. Green, blue and red for carbon, nitrogen and oxygen atoms respectively.

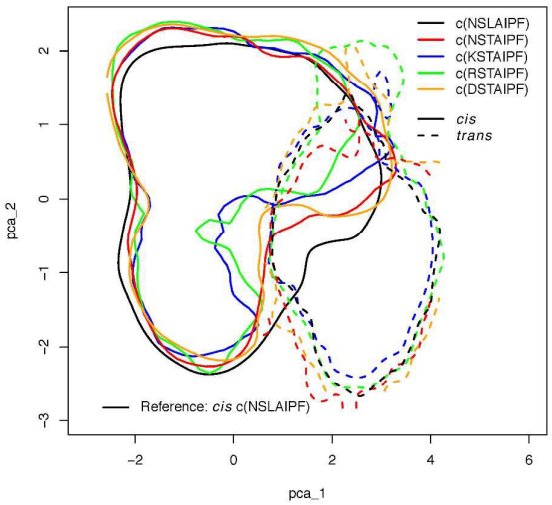


Figure 6. Conformational analysis of the *cis* (solid lines) and *trans* (dotted lines) conformers of all monomeric cyclopeptides. The plot shows the contour map (isodensity lines) of the projection of the backbone Cartesian coordinates of MD simulations of all cyclopeptides over the same conformational space of the *cis* conformer of monomeric c(NSLAIPF) by principal component analysis.

is characterised by two concomitant transitions: a strong transition between 1610 and 1640 cm-1 (peak at 1620 cm^{-1} in our case) and a weak transition at higher wavenumbers (shoulder at 1640 cm^{-1}), see Figure S5. It therefore supports our finding of intramolecular hydrogen bonds.[36,37,38,39] Interestingly, some correlations between the sets of signals for *trans* and *cis* species were observed for Ile, Thr, Phe and Ser (dotted circles in Figure 4). These signals also appeared in the ROESY experiment (Figure S6), but with an inverted sign; thus confirming the slow exchange between conformations. Note that there are no findings regarding the presence of a dimer aggregate in the present NMR study, which has to be related to the effect of the DMSO solvent.

2.4. Molecular dynamics of monomeric *cis* and *trans* conformers of c(KSTAIPF) and the other cyclopeptides

A series of restrained MD simulations were used to obtain structural models of *cis* and *trans* conformers of c(KSTAIPF). The X-ray structure of c(NSLAIPF), where the Ile-Pro peptide bond adopts a *cis* conformation, was used as a template to build the *cis* c(KSTAIPF). Meanwhile, the *trans* conformation was built from its *cis* counterpart by modifying the Ile-Pro bond and applying the geometrical restraints derived from the NOE correlations of the NMR study between non-vicinal residues. All the other conformers of the cyclopeptides were built from the initial models of *cis* c(NSLAIPF) and *trans* c(KSTAIPF) (see Experimental Section and Figure S7).

First of all, it is important to mention that simulated *cis* c(NSLAIPF) reproduces all the structural features of the reported X-ray structure[21] and secondly, that MD simulations yield stable structures of both *cis* and *trans* conformers for all the cyclopeptides. The *cis* c(KSTAIPF) presents a type I β-turn involving the residues Lys^1-Ser^2-Thr^3-Ala^4 and a type VIa β-turn with the residues Ala^4-Ile^5-Pro^6-Phe^7 (Table 1 and Figure 5).[40]

Moreover, the Ile carbonyl group was oriented towards the exterior of the molecule, in agreement with the *cis* Ile-Pro conformation. These turns are accompanied by the corresponding hydrogen bonds: one between Ala(NH) and Lys(CO) in the type I β-turn, and the other between Phe(NH) and Ala(CO) in the type VIa β-turn. Additionally, the NH peptide group of Lys^1 points towards the interior of the cyclopeptide and interacts weakly with the CO of Ala^4. In summary, all these geometrical features of *cis* c(KSTAIPF) resemble those found in the backbone conformation of *cis* c(NSLAIPF), as reported in the work by Pettit et al.[21]

Meanwhile, the *trans* c(KSTAIPF) showed a single type II β-turn involving the Ser^2-Thr^3-Ala^4-Ile^5 fragment and two γ-turns[41] (Phe^7-Lys^1-Ser^2 and Ile^5-Pro^6-Phe^7) with their corresponding distorted hydrogen bonds, see Table 1 and Figure 5 for distances and geometries. It is important to remark that γ-turn involving Phe^7-Lys^1-Ser^2 fragment competes with the transient hydrogen bond between backbone of Phe(CO) and side chain of Ser(OH). Notice also that the orientation of the Ile carbonyl group towards the interior of the molecule is also in agreement with the trans Ile-Pro assignment for the major conformer in the NMR study.

The *cis* and *trans* conformers of the other mutated cyclopeptides keep the same structural features as c(KSTAIPF) with regard to the backbone interactions (see Table S2 for the average distances in β- and γ-turns, and Figure 6 for the conformational space explored by the backbone skeleton of each cyclopeptide). The differences appear in the formation and strength of extra internal hydrogen bonds that involve residue side chains. First, none of the *trans* conformers contain

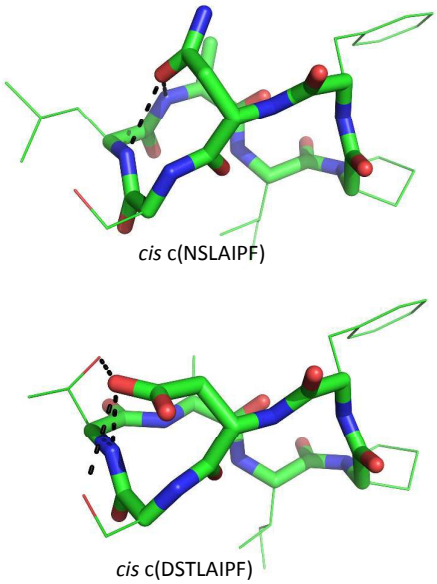


Figure 7. Molecular representation of a representative structure of the *cis* conformer of (top) c(NSLAIPF) and (bottom) c(DSTAIPF). Backbone and side chain of residues at position 1 are pictured as bold sticks and slim sticks represent the rest of side chains. Hydrogen bonds between side chain of residues at position 1 and positions 2, 3 and 4 are denoted by black lines. See also Table 3.

any extra hydrogen bonds other than those involved in the β- and γ-turns described above. This trend reflects the inability of mutated residues of *trans* conformers to form hydrogen bonds involving side-chain residues due to their overall shape, which orients the side chain toward the solvent (see Figure 5). In contrast, *cis* conformers of c(NSLAIPF), c(NSTAIPF) and c(DSTAIPF) contain extra hydrogen bonds between the side-chain residue at position 1 (Asn, Asn and Asp, respectively) and the backbone or side chain at position 3 (Leu, Thr and Thr, respectively). The characterization (bond lengths and intensity of the interactions) of the additional hydrogen bonds for *cis* c(NSLAIPF), c(NSTAIPF) and c(DSTAIPF) is listed in Table 2 and displayed in Figure 7. *Cis* c(NSLAIPF) contains two additional hydrogen bonds: one (Asn^1-CO···HN-Leu^3) was already identified by Petit et al. in their X-ray model;[21] and the other is a weak and transient interaction (Asn^1-CO···HN-Ala^4). *Cis* c(NSTAIPF) also presents an additional weak hydrogen bond (Asn^1-CO···HO-Thr^3), while *cis* c(DSTAIPF) shows one additional strong hydrogen bond (Asp^1-COO-···HN-Ser^2) and two weak interactions. This differences in the internal hydrogen bond pattern between all the cyclopeptides could tune the *cis*/*trans* conformational ratio as already claimed for proline derivatives.[42]

2.5. Estimation of the *cis*/*trans* conformational ratio of monomeric cyclopeptides by semiempirical and quantum mechanical calculations

To further confirm the stabilizing role of the additional hydrogen bonds found in *cis* conformers of c(NSLAIPF), c(NSTAIPF) and c(DSTAIPF) compared to *cis* conformers of c(KSTAIPF) and c(RSTAIPF) and their effect on the *cis*/*trans* conformational ratio, we performed extensive semiempirical, DFT and MP2 calculations. The direct aim of these calculations was to determine the relative stability of both conformers of each cyclopeptide. So, 256 conformations (on average, 26 conformers per isomer of each cyclopeptide, see Table S3 and Figure S8) extracted from the aqueous MD simulations were subjected to geometry optimization and frequency calculation at the semiempirical/PM6-DH+/COSMO[43,44] and DFT/M06-2X/6-31G(d)/SMD[45,46] levels, in order to derive the relative stabilities of each conformer (see Table 3 and Experimental Section for details). It is important to note that these methods explicitly incorporate non-covalent interactions into their parameterizations.[47,48]

Although caution is required when interpreting the absolute values in Table 3, the two methods agree in the relative order of the stability between the *cis* and *trans*

Table 2. Additional hydrogen bonds in *cis* conformers of all studied cyclopeptides.[a]

Cyclopeptide	Bond acceptor	Bond donor	Average distance (Å) along 100 ns	% time below 3.5 Å
c(NSLAIPF)	Side chain Asn^1 (CO)	Main chain Leu^3 (NH)	3.1 ± 0.5	92
		Main chain Ala^4 (NH)	3.5 ± 0.8	61
c(NSTAIPF)	Side chain Asn^1 (CO)	Side Chain Thr^3 (OH)	3.7 ± 1.0	57
c(KSTAIPF)	-	-	-	-
c(RSTAIPF)	-	-	-	-
c(DSTAIPF)	Side chain Asp^1 (COO^-)	Main chain Ser^2 (NH)	3.4 ± 0.5	65
		Main Chain Thr^3 (NH)	3.9 ± 0.9	48
		Side Chain Thr^3 (OH)	4.7 ± 1.0	13

Table 3. Relative energies of *cis* conformers versus *trans* conformers for all studied cyclopeptides, in kcal/mol.[a]

Cyclopeptide	PM6-DH+[b]	M06-2X/6-31G(d)[b]	M06-2X/6-31G(d,p)[c]	MP2/6-31G(d)[d]
c(NSLAIPF)	-7.5	-7.6	-6.7	-4.1
c(NSTAIPF)	-8.2	-6.0	-4.9	-4.7
c(KSTAIPF)	-2.6	-4.8	-3.5	-3.0
c(RSTAIPF)	-3.6	-5.7	-5.5	-4.0
c(DSTAIPF)	-9.4	-13.5	-13.1	-8.4

[a] See methodology section to find out how these values have been estimated. See also Table S5, S6, S7 and S8. A negative sign denotes a preference for the *cis* conformers. Additional results at M06-2X/def2-TZVP can be found in the Table S4.[b] Relative Gibbs free energy using 256 conformations. [c] Relative energy from geometry optimization calculations on the 67 most important conformations from M06-2X/6-31G(d) level of theory. [d] Relative energy from single point calculations on the 67 most important conformations from M06-2X/6-31G(d) level of theory.

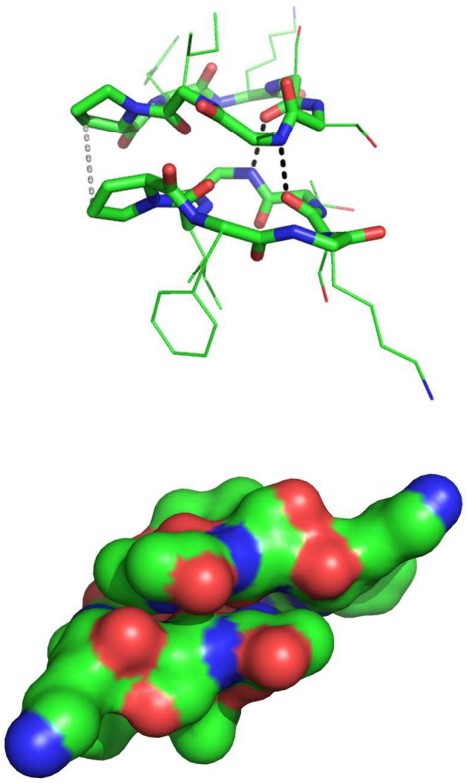


Figure 8. (Top) Molecular representation with bold (backbone) and slim (side chain) sticks and (Bottom) monomeric surfaces of a representative structure of the *trans*,*trans* c(KSTAIPF) dimer. Intermolecular hydrogen bonds are shown as black lines and the hydrophobic Pro-Pro interaction with a grey line.

conformers. The results show that the *cis* conformer is more stable than the *trans* species in all the cyclopeptides regardless of the residue mutation: they range from -2.6 to -9.4 kcal/mol with PM6-DH+, and from -4.8 to -13.5 kcal/mol with M06-2X/6-31G(d). This agrees with the expectation that two β-turns (*cis* conformers) are more stabilizing than one β-turn and two γ-turns (*trans* conformers). So, the two methods give the following order for the relative stability of *cis* and *trans* conformers: c(DSTAIPF) < c(NSTAIPF) ≈ c(NSLAIPF) < c(RSTAIPF) < c(KSTAIPF). Additional calculations with fewer conformations but with larger basis sets (M06-2X/6-31G(d,p) and M06-2X/def2tvzp) and with a wavefunction-based method such as MP2/6-31G(d) yield the same qualitative results, see Table 3 and Table S4. The computed *cis*/*trans* ratio order correlates well with: a) the alteration of the CD spectra as the concentration of the cyclopeptide increases; b) the detection of double mass peaks in the MALDI-TOF spectra; and c) the number of additional hydrogen bonds that result from the mutated side-chain residues. Thus, within the uncertainty of the absolute estimates reported in Table 3, it is important to note that the smallest differences between the relative stabilities of *cis* and *trans* conformers are found for c(KSTAIPF) and c(RSTAIPF); which are the cyclopeptides that exhibit the most significant changes in their CD and MALDI-TOF spectra. In contrast, the differences found for the other three cyclopeptides appear to be large enough to rule out a conformational equilibrium between the *cis* and *trans* conformers, especially for c(DSTAIPF). To sum up, the present QM calculations support the view that there is a *cis*-*trans* conformational equilibrium highly displaced toward the *cis* conformations in water, but with an increasing presence of the *trans* conformations for c(RSTAIPF) and c(KSTAIPF) .

Table 4. Binding free energies of the *trans,trans* dimer cyclopeptides estimated by MM-GBSA, MM-PBSA and MM-RISM methodologies from 100 ns of unconstrained MD simulations, in kcal/mol.[a]

Cyclopeptides	MM-GBSA	MM-PBSA	MM-RISM
c(NSLAIPF)	-24.4	-12.0	-4.3
c(NSTAIPF)	-23.2	-12.7	-4.1
c(KSTAIPF)	-23.2	-12.1	-3.9
c(RSTAIPF)	-22.0	-10.7	-3.3
c(DSTAIPF)	-23.9	-12.2	-4.8

[a] See methodology section to find out how these values have been estimated.

2.6. Dimerization of c(KSTAIPF) and the other cyclopeptides

The concentration-dependent change observed in the CD spectra (Figure 2) and the identification of species with a mass corresponding to the dimer in the MALDI-TOF assay (Figure 3) suggest the capacity of c(KSTAIPF) and c(RSTAIPF) to dimerize. Keeping in mind the presence of both *cis* and *trans* conformers in the NMR study and the similar stability of both species from the previous section, this raises a challenging question regarding the nature of the monomeric species involved in peptide aggregation. From a geometrical point of view, the cyclopeptide ring in *trans* c(KSTAIPF) adopts a planar-like structure, which is in contrast with the folded ring found in the *cis* conformer (see Figure 5). Therefore, whereas in the *cis* conformation the side chains are pointing toward the two faces of the ring, the side chains in *trans* c(KSTAIPF) are oriented toward one face of the ring, leaving the other largely exposed to the bulk solvent. Finally, it is worth noting that *trans* c(KSTAIPF) exposes a highly homogeneous hydrophobic surface to the solvent; a feature that it has been proposed as an essential requirement for promoting aggregation.[41]

The formation of the dimeric species was examined by means of MD simulations. To this end, up to five distinct orientations of dimeric c(KSTAIPF) were generated by changing the relative orientation of the stacked cyclopeptides in order to find the most favourable arrangement (Figure S9). All the attempts made to find stable structures for the *cis*,*cis* and *cis*,*trans* dimeric species were unsuccessful: the MD trajectories led to the dissociation of the aggregate. In contrast, the *trans,trans* conformation led to a stable structure in seven out of eight independent 100 ns MD replicas run for this dimer. Gratifyingly, MM-GBSA, MM-PBSA and MM-RISM calculations confirmed that the binding free energy of the *trans*,*trans* dimer (-23.2, -12.1, -3.9 kcal/mol, respectively) is

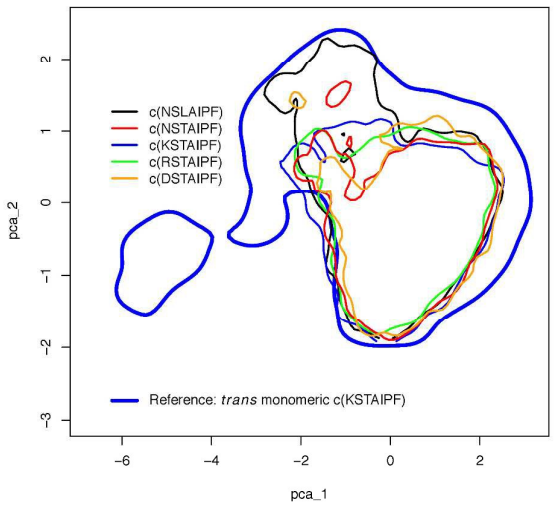


Figure 9. Conformational analysis of one of the monomers of all *trans*,*trans* dimeric cyclopeptides. The plot shows the contour map (isodensity lines) of the projection of the backbone Cartesian coordinates of MD simulations of one monomer of all *trans*,*trans* dimer cyclopeptides over the same conformational space of the *trans* monomeric c(KSTAIPF) taken as a reference by principal component analysis. The equivalent plot for the monomeric counterpart can be found in Figure S9.

more favourable than for the *cis*,*cis* dimer (-13.0, -5.3, -1.3 kcal/mol; in this case, the calculations were performed for the subset of snapshots taken before dissociation of the dimer).

The interaction between *trans* monomers depends upon the way in which the planar faces of the two monomers align (see Figure 8). Whereas the solvent-accessible surface of the monomer amounts to around 900 Å^2, formation of the dimer reduces the surface exposed to solvent by around 20% (a total solvent-accessible area of some 1400 Å^2). This reduction is particularly important for Ala (60%), Thr (30%) and Pro (28%) residues. Three major features characterize the interaction between the monomers. First, the Lys residues in each monomer are well exposed to the solvent, and their side chains are oriented in opposite directions, thus reducing the electrostatic repulsion between the protonated amino groups. Second, the two Pro residues are aligned face to face. Third, two intermolecular hydrogen bonds that involve the backbone groups of the Ala and Ser residues of the interacting monomers are formed in the dimer (average distance of 3.0 Å).

We would expect dimeric species to occur in the other cyclopeptides in a similar way to how it occurs in c(KSTAIPF) due to the structural resemblance of the *cis* and *trans* backbone skeleton in all monomeric cyclopeptides (see Figure 6). Indeed, the simple mutation of the corresponding final *trans*,*trans* c(KSTAIPF) dimer gives rise to stable 200 ns MD simulations with a united dimeric form for all the cyclopeptides, see Table S9. Likewise, the binding free energy estimated using the MM-GBSA, MM-PBSA and MM-RISM methods yielded similar relative results for all the cyclopeptides; although there were great differences in absolute values between methods (Table 5). The higher-level MM-RISM computes binding free energies ranging from -4.8 to -3.3 kcal/mol. Such a small difference between dimers is expected since the mutated residues are not directly involved in the three main features that determine the binding between interacting monomers.

On the other hand, dimerization does not introduce significant changes in the β-turn secondary structure of the separate monomers. The only noteworthy alteration concerns one of the γ-turns (F^7_i-K^1_{i+1}-S^2_{i+2}); not only in terms of average values but also of the reduction of the standard deviation, which may indicate the structure becomes more rigid upon dimerization (compare Tables S10 and 1). This is also reflected in the principal component analysis projection of the backbone space measured by the Cartesian coordinates of the dimers over the corresponding backbone space of *trans* monomer c(KSTAIPF) displayed in Figure 9. Thus, monomers occupy a slightly larger conformational space than the assembled species in the dimer; which only implies a certain reduction in conformational flexibility upon dimerization.

3. Conclusions

We have designed, synthesized and modelled a set of four cyclic heptapeptides based on mutations of the structure of c(NSLAIPF), a potential antibiotic agent, taking into account the need for residues with different capacities to form hydrogen bonds. The final aim was to alter the *cis*-*trans* Ile-Pro amide bond equilibrium and thereby rationalize the structural determinants of the *cis*-*trans* displacement in such derivatives of c(NSLAIPF). Of the synthesized set, two compounds, c(KSTAIPF) and c(RSTAIPF), showed clearly different behaviour that could be interpreted in terms of aggregation: i) a concentration-dependent change in the CD spectra; ii) a $[2M+H]^+$ peak in the MALDI-TOF spectra; and iii) a double HPLC peak at room temperature that coalesces upon heating to 45ºC (in the case of c(KSTAIPF)). Meanwhile, c(NSLAIPF) and c(NSTAIPF) presented intermediate behaviour; whereas c(DSTAIPF) showed no evidence of aggregation whatsoever. Although NMR could not be used to confirm the presence of the dimeric form of c(KSTAIPF) in water, NMR studies of c(KSTAIPF) in DMSO revealed the presence of two major conformations, *cis* and *trans* Ile-Pro conformations, which are primarily characterized by the presence of two β-turns in the former, and one β-turn and two γ-turns in the latter.

MD simulations in explicit water corroborated the feasibility of the monomeric *cis* and *trans* structures, as well as the capacity to form a sufficiently stable dimer only in the case of the *trans* species. Moreover, semiempirical/PM6-DH+/COSMO, DFT/M06-2X/SMD and MP2/SMD calculations involving conformational ensembles of all cyclopeptides corroborate the preference to adopt the *cis* conformation, although only a small difference with the *trans* species is found for c(KSTAIPF) and c(RSTAIPF). The overall greater stability of the *cis* conformers than that of the *trans* conformers has to be attributed to the higher stability of two β-turns (*cis*

conformers) than that of one β-turn and two γ-turns (*trans* conformers). Meanwhile, the relative stability of *cis* conformers of c(DSTAIPF), c(NSTAIPF) and c(NSLAIPF) has to be assigned to the fact that, unlike *cis* conformers of c(KSTAIPF) and c(RSTAIPF), these *cis* conformers contain additional hydrogen bonds between side chains at position 1 and position 3 that confer extra stability to the *cis* conformers compared to the *trans* conformers. From a qualitative point of view, these findings are in agreement with the findings that: i) no experiments indicate that c(DSTAIPF) forms a dimer structure; ii) both CD and MALDI-TOF experiments suggest that dimers of c(KSTAIPF) and c(RSTAIPF) form; and, finally, iii) c(NSLAIPF) and c(NSLAIPF) show intermediate behaviour between the preceding cases, since the CD spectra of both cyclopeptides display only a certain degree of concentration-dependent alteration, and only a tiny peak corresponding to the dimer was found for c(NSLAIPF). To sum up, as only *trans* conformers have the capacity to form aggregates, the displacement of the *cis*-*trans* equilibrium by either destabilizing the *cis* conformers or stabilizing the *trans* conformers will ultimately determine the formation and appearance of *trans*,*trans* dimer aggregates.

Putting all the pieces together, these findings allow us to suggest that this family of stylostatin derivatives exhibits a conformational equilibrium between *trans* and *cis* monomeric conformers, which: i) can be modulated by the capacity of residues to form internal hydrogen bonds; and ii) can be altered by the concentration-dependent formation of the *trans*,*trans* dimer. Thus, even though the dimerization does not lead to drastic structural changes in the monomers, the changes in the CD spectra probably reflect the displacement of the conformational equilibrium between *cis* and *trans* monomeric conformers due to the dominance of the *cis* conformer in the monomeric state and to the exclusive formation of the *trans*,*trans* dimer state. Overall, the results indicate that the capacity to dimerize should not be interpreted merely in terms of specific physicochemical properties of the cyclopeptide, but as indicating the presence of a suitable structural motif that permits interaction between the monomers. These findings could provide a useful guide for exploration of the displacement of the X-Pro *cis-trans* equilibrium in other cyclopeptide families.

4. Experimental Section

4.1. Peptide synthesis

Solid-phase syntheses were carried out at 25°C in polypropylene syringes (5/10 mL) fitted with a polyethylene porous disk. Solvents and soluble reagents were removed by suction. Washing was performed with DMF (5 x 1 min) and DCM (5 x 1 min) using 10 mL of solvent per gram of resin for each wash. Fmoc-protected amino acids were used. The commercial (Novabiochem) side chain-protected amino acids used were: Fmoc-Ser(tBu), Fmoc-Thr(tBu), Fmoc-Asp(Trt), Fmoc-Lys(Boc), and Fmoc-Arg(Boc)2.

Commercial resin H-Pro-2-ClTrt (Novabiochem, 0.3-0.9 mmol/g of resin, ref. 04-12-2805) was swollen with DMF and treated with the appropriate Fmoc protected amino acid, HOBt and DIPCDI in a 3:3:3 ratio. The mixture was manually stirred and left for 1 h, then the solvent was filtered off and the resin was washed with DMF. Piperidine (20% in DMF) was then added to the resin, and the mixture was left for 0.5 h, after which the liquid was filtered off and the resin was washed with DMF and with DCM. The same method was applied to the other five amino acids. Then the protected linear peptides were cleaved from the solid support by treatment of the resin with AcOH/TFE/DCM (2:2:6) for 2 h, and the solution containing the peptides was recovered. Cyclization through Pro^6-Phe^7 was performed by addition of PyBOP/DIEA (1.5:3) in DMF/DCM (97:3) to the linear peptide. The cyclization was monitored by analytical reversed-phase HPLC. Final deprotection of the side chains was performed with TFA/H2O (95:5). The peptides were then purified by HPLC and lyophilized.

4.2. Purification and separation of cyclopeptides

The compounds were purified using a Waters HPLC system with a SymmetryPrep C18 reverse phase preparative column (7.8 x 300 mm) with 7 μm packing material. The crude peptides (50-100 mg) were dissolved in water/acetonitrile 1:2 (1 mL), filtered and injected through a Rheodyne injector with a 1 mL sample loop. The mobile phases used were A: 0.05% aqueous TFA and B: 0.05% TFA in acetonitrile, with a linear gradient from 10%B to 45%B over 40 min (3 mL/min). The fractions were manually collected at 1 min intervals. Elution of the peptide was determined simultaneously from the absorbance at 220 and 254 nm (Waters 2487). The fractions were re-injected to assess purity via analytical reversed-phase HPLC (Waters NovaPak C18, 3.9 x 150 mm, 4 μm). In this case, we used a linear gradient from 0% B to 100% B over 30 min for peptide characterizations. Retention times observed at 220 nm under two different conditions are given. The pure fractions were combined and lyophilized. Each peptide showed $[M+H]^+$, $[M+Na]^+$ and/or $[M+K]^+$ peaks in its mass spectrum, and each was >90% pure. The samples were lyophilized to yield highly hygroscopic white solids. The structure of all the compounds was confirmed by amino acid analysis and MALDI-TOF (using ACH as the matrix). In addition, NMR experiments were performed on c(KSTAIPF). The NMR signal assignments required 2D-TOCSY, COSY (H,H), 2D-HSQC, 2D-NOESY, and 2D-ROESY experiments.

4.3. Spectroscopic analysis

The cyclopeptides were characterized by MALDI-TOF mass spectrometry (Voyager-DE RP MALDI-TOF, PE Biosystems with an N2 laser at 337 nm). 1 μL of the compound solution (0.5-2 mg/mL) was mixed with 1 μL of an α-cyano-4-hydroxycinnamic acid (ACH) matrix then seeded on the MALDI plate and air-dried. The matrix was prepared from a 10 mg/mL solution of ACH in MeCN/H2O 1:1 (v/v) containing 0.1 % TFA.

CD spectra were recorded with a Jasco 810 UV-Vis spectropolarimeter, using a Peltier CDF 426S/426L. The spectra were obtained in a wavelength range of 190 to 260 nm at a spectral bandwidth of 1 nm, with a time response of 4 s, a scan speed of 20 nm/min, and a step resolution of 0.1 nm. Each spectrum was the average of four accumulations. The spectra were measured by dissolving the cyclopeptide in 10 mM phosphate buffer at pH 7.4 and were recorded at 25ºC.[49] The blank was subtracted from each peptide spectrum.

^{1}H and ^{13}C NMR spectra of c(KSTAIPF) were recorded on a Bruker 600-MHz NMR spectrometer. The samples were approximately 3 mM in DMSO-d_6. Chemical shifts were reported in δ (ppm) relative to the undeuterated fraction of DMSO-d_6 and to TMS. Amino acid analysis was performed on a Beckman system Gold 6300.

4.4. Molecular dynamics simulations

All systems were initially solvated with TIP3P[50] waters in an octahedral box spanning 12 Å from the peptide to the edge of the box and neutralized with chloride anions or Na^+ cations. The Amber ff99SB-ILDN[51] force field was used as implemented in the AMBER12 package.[52] The standard protonation state at a physiological pH of 7.4 was assigned to the ionizable residues. Four minimizations (initially with 2000 cycles of steepest descent followed by a maximum of 10000 cycles of conjugate gradient) were run sequentially for hydrogen atoms, ions, water molecules and finally all atoms. Thermalization and first initial MD simulations were run under the canonical isochoric-isothermic (NVT) ensemble, and then the canonical isothermic-isobaric (NPT) ensemble was used for the rest of the MD simulations. Langevin dynamics was employed to control the temperature, whereas Berendsen bath coupling was used for the pressure (1 atm). A time step of 2 ps was used, together with SHAKE,[53] a non-bonded cut-off of 15.0 Å, and the Particle-Mesh-Ewald (PME) method (grid spacing of 1 Å) for electrostatic interactions.

4.5. Elucidation of *cis* and *trans* monomeric c(KSTAIPF) and other monomeric species

The structure of *cis* and *trans* c(KSTAIPF) was resolved by means of restrained MD simulations. The starting structure of *cis* c(KSTAIPF) was derived from the X-ray structure of c(NSLAIPF) and simulated according to the preceding protocol for 200 ns. The other *cis* cyclopeptides were derived in the same way as in the case of *cis* c(KSTAIPF). The *trans* c(KSTAIPF) was generated by rotating the Ile-Pro peptide bond, and the rest of the peptide bonds were manually accommodated using the facilities implemented in Pymol.[54] Then, the NOE correlations observed in the NMR spectra for *cis* and *trans* c(KSTAIPF) were used to derive suitable distance restraints, which were imposed on the subsequent all-atom explicit solvent MD simulations. Thermalization comprised four steps of 200 ps to heat the system from 150 to 300 K under the canonical NVT ensemble, followed by an equilibration step of 1 ns at 300 K under the canonical NPT ensemble, and finally a 20 ns MD simulation in which the harmonic restraints (force constants of 10 kcal mol^{-1} Å^{-2}) were gradually removed over the last 5 ns. To further explore the structural integrity of the candidate structure obtained in the preceding step, two replicas of the final structure were subjected to four series of simulated annealing cycles from 300 K to 500 K and back to 300 K in 5 ns NPT simulations for each cycle. Then, the two resulting structures were subjected to 200 ns of NPT. No significant change was detected in the last 50 ns of the trajectories and, more importantly, the same backbone conformation was achieved in both replicas, see Figure S7. Once the *trans* c(KSTAIPF) was achieved, the other *trans* cyclopeptides were obtained by simply mutating the corresponding residue in the final *trans* c(KSTAIPF) structure and submitting this initial structure to 200 ns of plain MD simulations. Gratifyingly, the conformational space occupied in the unconstrained MD simulations by each cyclopeptide was fairly similar (see Figure 6), which confirms the strategy of the structural elucidation of the *cis* and *trans* isomers for these monomeric cyclopeptides, and also indicates the structural integrity of the backbone of the *cis* and *trans* conformers.

4.6. Relative stability of *cis* and *trans* conformers for all monomeric cyclopeptides

The relative stability of the *cis* and *trans* conformers for each cyclopeptide was calculated by first, extracting the most representative conformations of each conformer; second, calculating the energy of each conformation at several levels of theory (semiempirical/PM6-DH+/COSMO, DFT/M06-2X/6-31G(d)/SMD, DFT/M06-2X/6-31G(d,p)/SMD, DFT/M06-2X/def2tvzp/SMD and MP2/6-31G(d)/SMD); and third, applying a weighting scheme according to the Boltzmann distribution function to derive the relative free energy between the two isomers.

In more detail, the most diverse conformations were selected from the conformational ensemble previously generated by 200 ns (2500 frames) of unconstrained MD simulations for each conformer of all five cyclopeptides. To that end, the k-means clustering technique applied to the Cartesian coordinates of non-hydrogen atoms classified the initial 2500 frames into 75 clusters. The most populated clusters which covered at least 75 % of the conformational space were selected and the closest conformations to the centre of these clusters selected (256 conformations in total). The number of conformations selected for the conformers of each cyclopeptide is shown in Table S3. The Gibbs free energies in implicit aqueous solutions for all the conformations of each cyclopeptide were initially determined by semiempirical calculations using the PM6-DH+/COSMO method, as implemented in MOPAC2012.[55] Then, DFT calculations at the M06-2X/6-31G(d) level, as implemented in Gaussian09 package, were also performed.[56] These semiempirical and DFT calculations involved full geometry optimizations followed by vibrational frequency calculations under the effect of a polarizable continuum model of water (dielectric constant: 78.4), using COSMO[57] for PM6-DH+ and using SMD[58] with radii and non-electrostatic terms from

Truhlar and coworkers for DFT/M06-2X. Vibrational frequencies were computed to verify that the optimized cyclopeptides were at the minimum of their potential energy surface and no imaginary frequencies were found in any of the conformations. The values of the electronic energy were then corrected with the zero-point vibrational energy, as well as the translational, rotational, and vibrational energies at 298.15 K, thus allowing the computation of the Gibbs free energy at the same temperature. In the case of the theoretical levels of theory corresponding to DFT/M06-2X/6-31G(d,p)/SMD, DFT/M06-2X/def2-tvzp/SMD and MP2/6-31G(d)/SMD, only the conformations with a population weight (calculated according to the Boltzmann distribution of the energy without thermal corrections) higher than 5% (67 conformations, see Table S4) at the level of DFT/M06-2X/6-31G(d)/SMD were subjected to geometry optimization (the former) or just single point calculations (the latter two). Finally, the relative free energies and relative energies (depending on the theoretical level of theory) of the *cis* and *trans* isomers of all cyclopeptide were derived from their Boltzmann distributions.

4.7. Elucidation of the dimeric form of *trans* c(KSTAIPF) and the other dimers

A systematic exploration of the relative orientation of two copies of *trans* c(KSTAIPF) was undertaken taking into account that only one of the monomer surfaces appears to be well suited to effective aggregation. This is because the side chains are oriented towards one face of the cyclopeptide, whereas the other face contains peptide bonds that can readily interact in hydrogen bonds. The computational protocol is shown in Figure S9. Briefly, the orientations were built under the premise of maximizing the number of intermonomeric hydrogen bonds and reducing the repulsive electrostatic interaction between the Lys residues in the interacting monomers. This process led to five distinct relative orientations that can be named according to the residues in the two peptides that are placed facing each other in the initial structure. The structural integrity of the five cases was examined via MD simulations. To this end, suitable restraints on the intermolecular hydrogen bonds were introduced in order to avoid artefactual rearrangements during equilibration and in a subsequent 20 ns trajectory, the restraints being gradually removed over the last 5 ns. The systems were then simulated for 100 ns with the only restriction of keeping the centre of mass of the C alpha carbon atoms of both monomers at a distance of less than 10 Å in order to allow the structural relocations of the monomers in the dimer. Analysis of the trajectories revealed a preferential orientation characterized by the intermonomeric backbone hydrogen bond between Ala and Ser, and the apolar interaction between Pro residues. Finally, to further explore the integrity of the dimeric species, a series of eight replicas were simulated for 100 ns. In five cases, the dimer retained the original structure, and in two cases the final structure showed a slight rearrangement to a similar orientation. Only in one replica was the dimer broken, leading to separated monomers at the end of the simulation. Finally, one of the succeeded replicas was extended up to 200 ns and used for the analysis.

4.8 MM-GBSA, MM-PBSA and MM-RISM methods for estimation of the binding energy

The energetics of dimerization was determined using the MM-GBSA (molecular mechanics/generalized Born surface area), MM-PBSA (molecular mechanics/Poisson Boltzmann surface area) and MM-RISM (molecular mechanics/reference interaction site model)[59] methods, as implemented in the python version[60] of AMBER12 for 500 uniformly spaced snapshots taken from the last 100 ns of each dimer simulation. The MM-GB(PB)SA binding free energy accounts for the electrostatic, van der Waals, polar solvation and apolar solvation terms, and is calculated with an ionic strength of 0.150 M, and internal and external dielectric constants of the solute and solvent of 2 and 80, respectively. Meanwhile, MM-RISM is an attractive and a higher-level alternative to continuum solvation models such as MM-GB(PB)SA. It is based on solvent distributions rather than individual molecules, and yields the solvation structure and thermodynamics from first principles of statistical mechanics. The conformational entropic contribution is not included since it can be assumed to have minimal impact on the relative difference in the binding energy between the two systems, due to the isomeric form of the dimers. The R program[61] was used for all statistical analysis and Pymol for molecular graphics.

c(Asn-Ser-Leu-Ala-Ile-Pro-Phe), c(NSLAIPF), Stylostatin 1 (33 mg, 13 % yield, 92 % HPLC purity). HPLC: t_r = 10.43 min (0 → 100% gradient in 15 min); HPLC: t_r = 10.24 min (15 → 70 % gradient in 15 min); HPLC: t_r = 11.61 min (25 → 35 % gradient in 15 min). MALDI-TOF: $[M+H]^+$ 743.45, $[M+Na]^+$ 765.43, $[M+K]^+$ 781.41, $[2M+H]^+$ 1485.63. AAA: Asn 1.00; Ser 0.79; Leu 0.83; Ala 1.05; Ile 1.00; Pro 1.10; Phe 1.01.

c(Asn-Ser-Thr-Ala-Ile-Pro-Phe), c(NSTAIPF), (30 mg, 15% yield, 93% HPLC purity). HPLC: t_r = 9.57min (0 → 100% gradient in 15 min); HPLC: t_r = 8.73min (15 → 70% gradient in 15 min). MALDI-TOF: $[M+Na]^+$ 753.39, $[M+K]^+$ 769.37. AAA: Asn 1.00; Ser 0.7; Thr 0.8; Ala 1.00; Ile 1.00; Pro 1.00; Phe 1.00.

c(Lys-Ser-Thr-Ala-Ile-Pro-Phe), c(KSTAIPF), (15 mg, 5% yield, 95% HPLC purity). HPLC: t_r = 8.5 and 8.7 min, 0 → 100% gradient in 15 min); HPLC: t_r = 8.25 min (15 → 45 % gradient in 15 min, T = 45 ºC). MALDI-TOF: $[M+H]^+$ 745.46, $[M+Na]^+$ 767.44, $[2M+H]^+$ 1489.89. AAA: Lys 1.01; Ser 0.82; Thr 0.90; Ala 1.02; Ile 0.97; Pro 1.08; Phe 1.02. ^{1}H-NMR *trans* (600 MHz, DMSO-d_6, δ in ppm): Lysine: 18.03 (NH), 4.47 (Hα), 1.51 (Hβ), 1.29 and 1.36 (Hγ), 1.47 and 1.56 (Hδ), 2.73 (Hε), 7.67 (NH_2); Serine: 8.67 (NH), 4.56 (Hα), 4.19 and 3.67 (Hβ); Threonine: 8.16 (NH), 3.68 (Hα), 3.98 (Hβ), 1.15 (Hγ); Alanine: 7.60 (NH), 4.31 (Hα), 1.22 (Hβ); Isoleucine: 6.95 (NH), 4.49 (Hα), 1.74 (Hβ), 1.36 and 0.98 (Hγ), 0.85 (Cγ-CH_3), 0.76 (Cδ-CH_3); Proline: 3.96 (Hα), 1.72 and 1.81 (Hβ), 1.89 and 1.98 (Hγ), 3.57 and 3.74 (Hδ); Phenylalanine: 8.57 (NH), 3.71 (Hα), 3.46 and 3.22

(Hβ). ^{13}C-NMR ***trans*** (150 MHz, DMSO-d_6, δ in ppm): Lysine: 53.9 (Cα), 26.0 (Cβ), 21.8 (Cγ), 32.8 (Cδ), 38.4 (Cε); Serine: 53.5 (Cα), 62.1 (Cβ); Threonine: 61.9 (Cα), 64.8 (Cβ), 20.4 (Cγ); Alanine: 48.1 (Cα), 18.7 (Cβ); Isoleucine: 52.9 (Cα), 36.7 (Cβ), 23.6 (Cγ), 13.9 (Cγ-CH_3), 9.9 (Cδ-CH_3); Proline: 60.6 (Cα), 28.7 (Cβ), 24.1 (Cγ), 47.6 (Cδ); Phenylalanine: 57.5 (Cα), 33.7 (Cβ). ^{1}H-NMR ***cis*** (600 MHz, DMSO-d_6, δ in ppm): Lysine: 7.64 (NH), 4.51 (Hα), 1.51 (Hβ), 1.23 (Hγ), 1.70 and 1.95 (Hδ), 2.68 (HCε), 7.59 (NH_2); Serine: 9.03 (NH), 3.87 (Hα), 3.71 (Hβ); Threonine: 6.65 (NH); 3.99 (Hα); 1.04 (Hβ); Alanine: 7.26 (NH), 4.42 (Hα), 1.14 (Hβ); Isoleucine: 8.73 (NH), 3.97 (Hα), 1.73 (Hβ) , 0.81 (Cγ-CH_3 and Cδ-CH_3); Proline: 4.35 (Hα), 1.67 and 1.99 (Hβ), 0.71 and 1.49 (Hγ), 2.80 and 3.23 (Hδ); Phenylalanine: 8.29 (NH), 4.23 (Hα), 3.21 and 3.02 (Hβ). ^{13}C-NMR ***cis*** (150 MHz, DMSO-d_6, δ in ppm) Lysine: 50.9 (Cα), 23.8 (Cβ), 21.2 (Cγ), 32.1 (Cδ), 38.5 (Cε); Serine: 58.9 (Cα), 59.6 (Cβ); Threonine: 57.5 (Cα), 20.3 (Cβ); Alanine: 47.9 (Cα), 16.0 (Cβ); Isoleucine: 60.5 (Cα), 34.9 (Cβ), 14.3 (Cγ-CH_3 and Cδ-CH_3); Proline: 59.8 (Cα), 30.1 (Cβ), 20.4 (Cγ), 45.4 (Cδ); Phenylalanine: 56.6 (Cα), 36.9 (Cβ).

c(Arg-Ser-Thr-Ala-Ile-Pro-Phe), c(RSTAIPF), (16.1 mg, 12% yield, 88% HPLC purity). HPLC: t_r = 8.72 min (0 → 100% gradient in 15 min); HPLC: t_r = 8.63 min (0 → 100% gradient in 15 min, T = 45 ºC); HPLC: t_r = 7,47 min, gradient 15 → 70% in 15 min, T = 45 ºC). MALDI-TOF: $[M+H]^+$ 773.46, $[2M+H]^+$ 1545.89. AAA: Arg 1.06; Ser 0.81; Thr 0.93; Ala 1.01; Ile 0.94; Pro 1.04; Phe 1.02.

c(Asp-Ser-Thr-Ala-Ile-Pro-Phe), c(DSTAIPF), (12 mg, 12% yield, 85% HPLC purity). HPLC: t_r = 9.78 min (0 → 100% gradient in 15 min); HPLC: t_r = 11.69 min (15 → 40 % gradient in 15 min). MALDI-TOF: $[M+H]^+$ 732.45, $[M+Na]^+$ 754.43, $[M+K]^+$ 770.41. AAA: Lys 1.02; Ser 0.76; Thr 0.90; Ala 1.05; Ile 0.91; Pro 1.06; Phe 1.06.

Acknowledgements

We thank Prof. F. Javier Luque from our department for his valuable suggestions and comments. M. Teixidó is acknowledged for technical assistance with MALDI-TOF spectra. This work was supported by grants CTQ2004-01757, CTQ2007-60764 and SAF2014-57094-R (Ministerio de Ciencia e Innovación, Spain), 2009SGR1111 and 2014SGR1189 (Generalitat de Catalunya), and the Centre de Supercomputació de Catalunya for computational facilities. PFM thanks the CONICYT for a postdoctoral fellowship, and CLM thanks the MICINN for his doctoral fellowship.

Notes and references

1. E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. Porco, D. Kozakov, S. Vajda, and A. Whitty, *Nat. Chem. Biol.*, 2014, **10**, 723–31.
2. A. Bhat, L. R. Roberts, and J. J. Dwyer, *Eur. J. Med. Chem.*, 2015, **94**, 471–9.
3. E. M. Driggers, S. P. Hale, J. Lee, and N. K. Terrett, *Nat. Rev. Drug Discov.*, 2008, **7**, 608–24.
4. E. Marsault and M. L. Peterson, *J. Med. Chem.*, 2011, **54**, 1961–2004.
5. A. C. Rand, S. S. F. Leung, H. Eng, C. J. Rotter, R. Sharma, A. S. Kalgutkar, Y. Zhang, M. V Varma, K. A. Farley, B. Khunte, C. Limberakis, D. A. Price, S. Liras, A. M. Mathiowetz, M. P. Jacobson, and R. S. Lokey, *Medchemcomm*, 2012, **3**, 1282–1289.
6. A. T. Bockus, *Curr. Top. Med. Chem.*, 2013, **13**, 821 – 836.
7. J. E. Bock, J. Gavenonis, and J. A. Kritzer, *ACS Chem. Biol.*, 2013, **8**, 488–99.
8. S. R. Gracia, K. Gaus, and N. Sewald, *Future Med. Chem.*, 2009, **1**, 1289–310.
9. M. Góngora-Benítez, J. Tulla-Puche, and F. Albericio, *Chem. Rev.*, 2014, **114**, 901–26.
10. L. D. Walensky and G. H. Bird, *J. Med. Chem.*, 2014, **57**, 6275–88.
11. C. Dugave and L. Demange, *Chem. Rev.*, 2003, **103**, 2475–532.
12. D. Pal and P. Chakrabarti, *J. Mol. Biol.*, 1999, **294**, 271–88.
13. A. Jabs, M. S. Weiss, and R. Hilgenfeld, *J. Mol. Biol.*, 1999, **286**, 291–304.
14. N. J. Zondlo, *Acc. Chem. Res.*, 2013, **46**, 1039–49.
15. B. Vera, J. Vicente, and A. D. Rodríguez, *J. Nat. Prod.*, 2009, **72**, 1555–62.
16. G. R. Pettit, J. K. Srirangam, D. L. Herald, J. Xu, M. R. Boyd, Z. Cichacz, Y. Kamano, J. M. Schmidt, and K. L. Erickson, *J. Org. Chem.*, 1995, **60**, 8257–8261.
17. L. T. Tan, R. T. Williamson, W. H. Gerwick, K. S. Watts, K. McGough, and R. Jacobs, *J. Org. Chem.*, 2000, **65**, 419–425.
18. M. Malešević, M. Schumann, G. Jahreis, G. Fischer, and C. Lücke, *Chembiochem*, 2012, **13**, 2122–7.
19. R. Sonti, K. H. Gowd, K. N. S. Rao, S. Ragothama, A. Rodriguez, J. J. Perez, and P. Balaram, *Chemistry*, 2013, **19**, 15175–89.
20. P. Forns, J. Piró, C. Cuevas, M. García, M. Rubiralta, E. Giralt, and A. Diez, *J. Med. Chem.*, 2003, **46**, 5825–33.
21. G. R. Pettit, J. K. Srirangam, D. L. Herald, K. L. Erickson, D. L. Doubek, J. M. Schmidt, L. P. Tackett, and G. J. Bakus, *J. Org. Chem.*, 1992, **57**, 7217–7220.
22. J. Piró, M. Rubiralta, E. Giralt, and A. Diez, *Tetrahedron Lett.*, 1999, **40**, 4865–4868.
23. M. Izadyar, M. Khavani, and M. R. Housaindokht, *Phys. Chem. Chem. Phys.*, 2015, **17**, 11382–11391.
24. D. W. Hoskin and A. Ramamoorthy, *Biochim. Biophys. Acta*, 2008, **1778**, 357–75.
25. A. Perczel, M. Hollosi, B. M. Foxman, and G. D. Fasman, *J. Am. Chem. Soc.*, 1991, **113**, 9772–9784.
26. P. Bour, J. Kim, J. Kapitan, R. P. Hammer, R. Huang, L. Wu, and T. A. Keiderling, *Chirality*, 2008, **20**, 1104–19.
27. R. W. Woody, in *Peptides, polypeptides and proteins*, ed. B. E. R. . B. F. A. . L. N. . G. M., Wiley, New York, 1974, pp. 338–50.
28. J. Bürck, S. Roth, P. Wadhwani, S. Afonin, N. Kanithasen, E. Strandberg, and A. S. Ulrich, *Biophys. J.*, 2008, **95**, 3872–81.
29. M. H. A. Perczel, in *Circular dichroism and the*

conformational analysis of biomolecules, ed. G. D. Fasman, Plenum Press, New York, 1996, pp. 285–380.
30. This is also the reason why the NMR analysis was not extended to the rest of peptides.
31. I. Z. Siemion, T. Wieland, and K. H. Pook, *Angew. Chem. Int. Ed. Engl.*, 1975, **14**, 702–3.
32. H. Kessler, *Angew. Chemie Int. Ed. English*, 1982, **21**, 512–523.
33. M. S. Donia, B. Wang, D. C. Dunbar, P. V Desai, A. Patny, M. Avery, and M. T. Hamann, *J. Nat. Prod.*, 2008, **71**, 941–5.
34. C. Merten, F. Li, K. Bravo-Rodriguez, E. Sanchez-Garcia, Y. Xu, and W. Sander, *Phys. Chem. Chem. Phys.*, 2014, **16**, 5627–33.
35. K. M. Makwana and R. Mahalakshmi, *J. Phys. Chem. B*, 2015, **119**, 5376–85.
36. N. Demirdöven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester, and A. Tokmakoff, *J. Am. Chem. Soc.*, 2004, **126**, 7981–90.
37. A. Barth and C. Zscherp, *Q. Rev. Biophys.*, 2002, **35**, 369–430.
38. C. Jung, *J. Mol. Recognit.*, 2000, **13**, 325–51.
39. A. Dong, B. Kendrick, L. Kreilgård, J. Matsuura, M. C. Manning, and J. F. Carpenter, *Arch. Biochem. Biophys.*, 1997, **347**, 213–20.
40. E. G. Hutchinson and J. M. Thornton, *Protein Sci.*, 1994, **3**, 2207–16.
41. G. Némethy and M. P. Printz, *Macromolecules*, 1972, **5**, 755–758.
42. M. Kuemin, Y. A. Nagel, S. Schweizer, F. W. Monnard, C. Ochsenfeld, and H. Wennemers, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 6324–7.
43. J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–213.
44. M. Korth, *J. Chem. Theory Comput.*, 2010, **6**, 3808–3816.
45. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
46. Y. Zhao and D. G. Truhlar, *Chem. Phys. Lett.*, 2011, **502**, 1–13.
47. K. E. Riley, M. Pitonák, P. Jurecka, and P. Hobza, *Chem. Rev.*, 2010, **110**, 5023–63.
48. J. H. Lee, H. S. Park, and Y. K. Kang, *New J. Chem.*, 2015, **39**, 4640–4646.
49. N. J. Greenfield, *Nat. Protoc.*, 2006, **1**, 2876–90.
50. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
51. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, *Proteins*, 2010, **78**, 1950–8.
52. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), AMBER 12, University of California, San Francisco.
53. J.-P. Ryckaert, G. Ciccotti, and H. J. . Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
54. The PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC.
55. MOPAC2012, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net, 2012..
56. Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
57. A. Klamt and G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2*, 1993, 799.
58. A. V Marenich, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 4538–4543.
59. A. Kovalenko and F. Hirata, *Chem. Phys. Lett.*, 1998, **290**, 237–244.
60. I. Bill R. Miller, J. T. Dwight McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, 2012.
61. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R.project.org.