Structural Analysis of Nonapeptides Derived from Elastin

Belén Hernández, Jean-marc Crowet, Joseph Thiery, Sergei Kruglik, N. Belloy, Stéphanie Baud, Manuel Dauchez, Laurent Debelle

Version of Record: https://www.sciencedirect.com/science/article/pii/S0006349520303428
Manuscript_317047d4b7041f09aa05e984d974bc75

Structural analysis of nonapeptides derived from elastin

Hernandez B [a,b], Crowet JM [a,c], Thiery J [a], Kruglik SG [d], Belloy N [a,c], Baud S. [a,c], Dauchez M [a,c], Debelle L* [a,c]

[a] *UMR CNRS 7369 Matrice Extracellulaire et Dynamique Cellulaire (MEDyC), Université de Reims Champagne Ardenne, UFR Sciences Exactes et Naturelles, Moulin de la Housse, 51687 Reims Cedex 2, France*

[b] *Sorbonne Paris Cité, Université Paris 13, Groupe de Biophysique Moléculaire, UFR Santé-Médecine-Biologie Humaine, 74 rue Marcel Cachin, 93017 Bobigny cedex, France*

[c] *Multiscale Molecular Modeling Platform (M3P). Université de Reims Champagne Ardenne, Moulin de la Housse, 51687 Reims Cedex 2, France*

[d] *UMR CNRS 8237, Laboratoire Jean-Perrin, Sorbonne Université, UPMC Paris 06, 75005 Paris, France*

Running title.

Elastin-derived nonapeptides

Corresponding author:

Prof. Laurent DEBELLE

Tel: +33 326 913 205

E-mail: laurent.debelle@univ-reims.fr

Abstract

Elastin-derived peptides are released from the extracellular matrix remodeling by numerous proteases and seem to regulate many biological processes, notably cancer progression. The canonical elastin peptide is VGVAPG which harbors the XGXXPG consensus pattern allowing interaction with the elastin receptor complex located at the surface of cells. Besides these elastokines, another class of peptides has been identified. This group of bioactive elastin peptides presents the XGXPGXGXG consensus sequence but the reason for their bioactivity remains unexplained. In order to better understand their nature and structure-function relationships, herein we searched the current databases for this nonapeptide motif and observed that the XGXPGXGXG elastin peptides define a specific group of tandemly repeated patterns. Further, we focused on four tandemly repeated human elastin nonapeptides, *i.e.* AGIPGLGVG, VGVPGLGVG, AGVPGLGVG and AGVPGFGAG. These peptides were analysed by means of optical spectroscopies and molecular dynamics. UV-circular dichroism and Raman spectra are consistent with a mixture of β-turn, β-strand and random chain secondary elements in aqueous media. Quantitative analysis of their conformations suggested that turns corresponded to half of the total population of structural elements while the remaining half was equally distributed between β-strand and unordered chains. These distributions were confirmed by molecular dynamics simulations. Altogether, our data suggest that these highly dynamic peptides harbor a type II β-turn located in their central part. We hypothesize that this structural element could explain their specific bioactivity.

Statement of Significance

Elastin fragmentation products, the so-called elastin peptides, may exhibit a bioactivity towards normal and tumor cells. This phenomenon depends on the sequence motif they harbor. While XGXXPG sequences bioactivity is explained by the presence of a type VIII β-

turn allowing interaction with the elastin receptor complex, the structural reasons for XGXPGXGXG specific activity remain unexplained. Using data mining, we show that elastin nonapeptides define a specific class of tandemly repeated features. Further, spectroscopic and numerical simulations methods suggest the presence of a type II β-turn in their conformation. This structural element could explain their bioactivity.

Introduction

Elastin is the extracellular matrix protein responsible for the structural integrity and function of tissues undergoing reversible extensibility or deformability (1). This protein is extremely stable and resistant and undergoes virtually no turnover. Nevertheless, during aging, mechanical stress and elastase activities contribute to the fragmentation of this macropolymer into elastin-derived peptides (EDP) (2).

Elastin is characterized by its elasticity and seems devoid of any other biological activity (1). In contrast, EDP have been shown to regulate numerous biological processes and are thought to be involved in several aging-related pathologies such as atherosclerosis (3, 4) and cancer (5–8). Most biologically active EDP, elastokines, possess a GXXPG consensus sequence adopting a type VIII β-turn structure (9) involved in their binding to the elastin receptor complex (10). The interaction of GXXPG-containing elastokines with the elastin receptor complex and the consequences of their interaction has been considerably documented during the last decades. For a comprehensive review on this topic, the reader is referred to (11).

Besides the classical GXXPG-containing sequences, another class of elastokines has also been reported. In 1988, Long and colleagues reported that nonapeptide sequences from elastin were chemoattractant for fibroblasts (12). In 1989, they further demonstrated that this biological activity could be extended to endothelial cells (13). Strikingly, since then, this peculiar class of EDP has been mostly ignored until 2007, when Maeda and colleagues demonstrated that these peptides could promote macrophages migration via a specific, but unknown, receptor (14). More recently, their biological activity has been further linked to lung carcinoma progression (8) and to promotion of invasion in triple-negative cancer cells via MMP-14 and MMP-2 (15).

The relatively low interest encountered by these nonapeptide EDP can be explained by the

fact that considerable advances have been made on both GXXPG-containing EDP and their receptor biology. As a consequence, few groups considered investigating on these peculiar peptides.

The nonapeptide AGVPGLGVG and other EDPs share similar structural features: random coil and β-turns conformations. However, AGVPGLGVG induces angiogenesis, tumor progression, secretion of proteases higher than other EDPs in carcinomas. Moreover, *in vitro*, it behaves like amyloid-like peptides harboring the XGGXG motif by forming cross β-sheets at the supramolecular level (16). Very recently, Brassart and colleagues have shown that the receptor mediating AGVPGLGVG biological effects, and identified as lactose-insensitive, is the ribosomal protein SA (RPSA) (17).

The aim of this work is to analyze the structure and conformational behavior of these peptides. Our structural analysis was conducted using sequence analysis, spectroscopic methods and molecular dynamics simulations.

Our results show that elastin nonapeptides define a peculiar class among peptides harboring the consensus X-G-X-P-G-X-G-X-G sequence. In elastin, these peptides are mostly tandemly repeated. Further our data show that they are engaged in a conformational equilibrium between random coil and turn conformations at the considered concentrations and temperatures. Molecular dynamics simulations suggest that the dominant conformation of these peptides is a type II β-turn. The functional significance of this conformation is discussed.

Material and Methods

Sequence analysis

Sequence retrieval. Sequences harboring the X-G-X-P-G-X-G-X-G consensus motif were searched using Python scripts on UniProtKB (Swiss-Prot and TrEMBL) database Fasta files

where spliced variants were ignored. Only exact matches were retrieved and further cleaned informatically and manually so as to ensure that a sequence belonging to a given species (identified by its unique accession number) was present only once in our results. For each sequence hit, the collected data were its accession number, the originating organism, the length of the protein and the position(s) of the nonapeptide sequence(s).

Pattern analysis. Nonapeptide exact matches were classified and distributed as a function of the taxonomic hierarchical classification lineage of the source organism of the corresponding parent protein. Hits were either individual hits, *i.e.* only one occurrence was found in the sequence, or multiple ones. In the latter case, we distinguished three possibilities. Multiple hits could be found in different regions of the protein and each occurrence was isolated from the others. Another possibility was that the repetitions could overlap, *i.e.* the end of the first occurrence (-X-G) was the beginning of the following one (X-G-). Finally, we observed that nonapeptide sequences could also be tandemly repeated, one after the other. The tree figure has been made with the Python module ETE3 (18) and the Logo sequences with WebLogo (https://github.com/weblogo/weblogo) (19).

Optical spectroscopy

Molecular compounds. Powder samples of amino acids were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). Lyophilized samples of nonapeptides were obtained from Genecust (Luxemburg, Luxemburg) as zwitterionic peptides in acetate salts. Their purity was at least 95% as determined by mass spectrometry.

Solution samples. Lyophilized powder of amino acids and peptides was dissolved in pure water from a Millipore filtration system (Guyancourt, France). The ionic strength of all peptide samples was increased by adding 150 mM NaCl to stock solutions. Upon dissolution, the pH was between 4 and 4.5. Raman spectra were recorded between 10°C and 80°C.

Circular dichroism (CD) spectra were obtained at room temperature between 100 and 200 µM.

CD spectroscopy. Room temperature ultraviolet-circular dichroism (UV-CD) spectra were analyzed on a JASCO J-810 spectrophotometer (Lisses, France) within the 190–300 nm spectral region. A path length of 1 mm and a spectral resolution of 0.2 nm were selected. Each spectrum corresponding to an average of five scans was recorded with a speed of 100 nm per min. The measured ellipticity for each sample, referred to as $[\phi]_{obs}$, was further normalized to obtain the so-called mean residue ellipticity, $[\phi]$, by using the expression $[\phi] = [\phi]_{obs} / ncl$, where n, c, and l are the length of the peptide, its molar concentration, and the optical path length, respectively. The normalized ellipticity was expressed in deg cm^2 $dmol^{-1}$.

Raman spectroscopy. Room temperature Stokes Raman spectra were analyzed in bulk samples at right angle on a Jobin-Yvon T64000 spectrometer (Longjumeau, France) at single spectrograph configuration, 1200 grooves/mm holographic grating and a holographic notch filter. Raman data corresponding to 1200 s acquisition time for each spectrum were collected on a liquid nitrogen cooled CCD detection system (Spectrum One, Jobin-Yvon). The effective slit width was set to 5 cm^{-1}. Solution samples were excited by the 488 nm line of an Ar^+ laser (Spectra Physics, Evry, France), with 200 mW power at the sample.

Post-record spectroscopic data treatment. Buffer subtraction and smoothing of the observed spectra was performed using the GRAMS/AI Z.00 package (Thermo Galactic, Waltham, MA, USA). Final presentation of Raman spectra was done by means of SigmaPlot package 6.10 (SPSS Inc., Chicago, IL, USA).

Secondary structure analysis. Quantification of protein secondary structures using amide I profile decomposition is a method used commonly and its intrinsic limits are readily known (20, 21). For elastin material, its overall accuracy is ± 5% as far as the final structural estimates are concerned (22). The initial step of the reconstruction is to guess how many

underlying bands are present in the complex amide I profile to be analyzed. This is achieved using a second derivative analysis of the smoothed spectrum. For our data, 4 components were suggested. This information was provided to the curve fitting software to reconstruct the experimental spectrum using Gaussian/Lorentzian mixture profiles. The curve fitting procedure was successful when the fit was good (low chi square value), the number of components was unchanged and their width at half height were consistent. The secondary structure estimates were obtained by adding the fractional areas of amide I structural components belonging to the same secondary structure group.

Atomistic molecular dynamics simulations

Ten independent simulations of 500 ns were performed for each of the 4 human elastin nonapeptides using amber99SB-ILDN force field (23) and Gromacs 2016.5 software (24). Peptides were built in an extended conformation with PyMOL (25) and solvated with TIP3P water in a 5 nm cubic box. All studied systems were first minimized by steepest descent for 5000 steps. Then, a 1-ns simulation with the peptides under position restraints was run before the production simulations were performed. Different seeds were used for each simulation during the velocities assignments. Periodic boundary conditions were used with a 2 fs time step. The dynamics were carried out under NPT conditions (310 K and 1 bar). The temperature was maintained using the v-rescale method (26) with τT = 0.1 ps, and an isotropic pressure was maintained using the Parrinello–Rahman barostat (27) with a compressibility of 4.5×10^{-5} bar^{-1} and τP = 2 ps. Short-range non-bonded interactions were treated with a cutoff of 1.0 nm and long-range interactions were calculated with the particle mesh Ewald method (28) with a grid spacing of 0.16 nm. Bond lengths were maintained with the LINCS algorithm (29) and long range dispersion corrections for energy and pressure were applied. The 3D structures were analyzed with both PyMOL and VMD (30) softwares.

The secondary structures were computed with DSSP (31) and the different types of β-turns

have been assigned based on the ΦΨ angles of two residues following each other as defined by Lewis *et al*. in 1973 (32) and Hutchinson *et al*. in 1994 (33). Distances between each *i* and *i+3* Cα have been computed and, if this distance was less than 0.7 nm and the amino acids were not part of an helix conformation, the ΦΨ angles of residues *i+1* and *i+2* were considered. To assign a peculiar type of β-turn, Hutchinson and colleagues (33) considered a range of ±30° or ±40° around the standard angles with a deviation of 50° allowed for one of the angle.

The distribution of ΦΨ angles of residues *i+1* and *i+2* we observed during the simulation permitted to easily identify type I, I', II, II' and VIII β-turn populations as broad plot regions (Fig. S1 in the Supporting Material). As these regions were larger than those proposed by Hutchinson and colleagues (33), we adapted the considered ΦΨ angles ranges to encompass the ranges we observed. The main effect of this change was the reduction of unassigned turn contribution, *i.e.* type IV β-turns. The corresponding figure and ΦΨ angle ranges for type I, I', II, II' and VIII β-turns are provided as supplemental data (Fig. S1, Table S1 in the Supporting Material). As in Hutchinson, other β-turn types are considered with a range of ± 40° around the standard angles and type IV β-turns correspond to otherwise unassigned β-turns. Clustering was performed using the gromos method (34) with a cut-off of 0.1 nm on Cα atoms of residues 2 to 6.

In order to assess the quality of our simulations, CD spectra have been computed from the peptide trajectories by using the SESCA software (35) with the basis set HBSS-3SC1.

Results

Sequence analysis.

The retrieval of sequences possessing the X-G-X-P-G-X-G-X-G yielded 96 293 unique sequences. These sequences (Table S2 in the Supporting Material) were classed as a function

of their taxonomy (Figure 1). Most sequences were from *Bacteria* (71 504 sequences for 75 718 occurrences with 2 129 overlaps and 297 tandem repeats) then *Eukariota* (22 919 sequences for 24 828 occurrences with 811 overlaps and 297 tandem repeats), *Archae* (1 630 sequences for 1 694 occurrences with 16 overlaps and 6 tandem repeats) and *Virus* (240 sequences for 245 occurrences with no overlaps and no tandem repeats). Strikingly, when these results are observed, it comes that elastins define a specific group. They exhibit numerous occurrences per sequences (171 for 38) with no overlap and a very high ratio of tandem repeats. This specificity of elastin nonapeptides is even more evident when the most frequent residues occurring at positions X are considered. While other group mostly favor the presence of G, A, K, P or L at these positions, elastins allow almost none of those with the AGVPGFGVG being the dominating motif. In fact, out of the 140 million sequences searched, this motif was always found in elastin sequences but one bacterial sequence.

The nonapeptide sequences found in elastin entries are reported in Table 1 where they are compared. The sequences are reported as a function of their alphabetical order. In human elastin, nonapeptides can be observed in exon domains 18, 20 or 26. Single occurrences are observed in the two first exons while they appear almost exclusively as tandem repeats in exon 26. Strikingly, the human repeats can also be observed in other species. For instance, the human elastin motif AGVPGFGVG is also present in mouse exon 26. Overall, human exon 26 (or its equivalent in other species) is a sequence domain characterized by tandem repeats of these nonapeptides. As a consequence, those are supposedly important for the biological function of the polymer and we decided to further analyze the structural behavior of the 4 human exon 26 nonapeptides, *i.e.* AGIPGLGVG (hN3), VGVPGLGVG (hN4), AGVPGLGVG (hN5) and AGVPGFGAG (hN6).

CD spectroscopy

In order to describe the structural behavior of the nonapeptides in solution, their CD spectra were recorded in different (polar *vs* low dielectric constant) environments. Routinely, methanol is considered as a primary step in studying the peptides conformational behavior in hydrophobic environments, such as membrane interior or hydrophobic pockets of proteins (36).

The analysis of the CD spectra in water (150 mM NaCl), in a water/methanol mixture (50/50, v/v), as well as in methanol (Fig. 2) reveal a progressive conformational change of the nonapeptides in going from water to methanol solutions.

The CD spectra recorded in water all present a double negative band shape, *i.e.* composed of a deep minimum at ~198 nm followed by a weaker and broader negative band centered at ~225 nm. It is a matter of fact that random chains are generally identified by a deep negative band at ~198 nm (37, 38). However, a negative double band, as that observed in the case of the nonapeptides (Fig. 2), is an indicator of turns. For instance, certain β-turns are known to give rise to a negative double band. It is to be mentioned that β-turns are recognized as the most frequent ones in proteins (39)(40)(41). Particularly, the CD fingerprints of four categories of β-turns, i.e. type-I (37), -II and -II' (42–44) and -VIII (45) β-turns are identified by a double negative band, with unequal ellipticities. It is interesting to note that amongst all the mentioned β-turns, that corresponding to type-VIII is formed by a deep minimum at ~198 nm followed by a broad shoulder at ~220 nm (45), *i.e.* strikingly similar to that observed in the case of the presently analyzed nonapeptides. However, this overwhelming resemblance between CD markers does not permit rejection of possible unordered chains, of which the CD marker can be overlapped with the negative band at ~198 nm. In the same manner, a probable overlap may occur between the CD fingerprint of β-strands, generally characterized by a unique negative band at ~215 nm, and the ~225 nm negative band from β-turns. Nevertheless, any uncertainty about the presence of PPII (polyproline II) type chains can be discarded

because of the recent CD studies highlighting the fact that the PPII fingerprint is composed of a deep negative band at ~198 nm followed by a positive one centered at ~220 nm. The presence of the latter band cannot be suspected in the CD spectra displayed in Figure 2.

In methanol (Fig. 2), the CD signals of the nonapeptides are composed of a broad positive band peaking at ~205 nm followed by a weak and broad negative band at ~225 nm. It is worth noting that the CD spectra observed in water/methanol mixture (1:1, v/v) are somehow an overlap of those observed in water and methanol. To provide a reliable assignment for the CD signal observed in methanol, we can recall that while a type-II β-turn gives rise to a positive band located between 190 and 210 nm (37), the observed negative band at ~225 nm in methanol can be ascribed to β-strands and/or to type-I' β-turn (37). Based on the observed CD spectra, no unordered chain can be expected in the presently studied nonapeptides.

Raman spectroscopy

Raman spectra of the nonapeptides were recorded in the middle wavenumber region (1800-550 cm^{-1}). They presented a low water contribution enabling an accurate analysis of the observed bands (Fig. 3A-B). Whatever the nonapeptide, the spectra were identical for the two tested concentrations, i.e. 2.5 mM and 20 mM (data not shown). Indeed, no substantial change, assignable to either a possible conformational transition or a detectable molecular aggregation, could been observed. Because the spectral features were more pronounced at 20 mM (higher Raman intensities due to higher concentration), we will limit our discussion to the Raman spectra recorded at that concentration.

The tentative assignments of the Raman spectra of the four studied nonapeptides are reported in Table 2. These assignments have been performed in keeping with the features observed in the Raman spectra of their constitutive amino-acids (Fig. S2 in the Supporting Material). The bands observed in the 1500-1320 cm^{-1} range, mainly arising from the bending modes of CH_2 and CH_3 side chain moieties, as well as those located below 1200 cm^{-1} in

amino acids spectra, were taken as references for assigning the observed peaks in the spectra of peptides.

The spectrum of hN6 is characterized by the occurrence of the six Phe characteristic Raman bands, referred to as F1 to F6 (46, 47). Being all of in-plane type and localized in the Phe phenyl ring, the wavenumbers of these six characteristic markers remained very close to those collected from the free amino acid F (Fig. 3B). The Raman spectra of hN3, hN4 and hN5 (Fig. 3A) do not present such narrow and resolved aromatic markers, but their striking spectral shape similarity is to be emphasized (Fig. 3). Two regions corresponding to amide I (1700-1640

cm^{-1}) and amide III (1300-1230 cm^{-1}) vibrations provide valuable structural information through their decomposition into band components (42–44). It is noteworthy that while amide I vibrations results from the backbone C=O bond stretch motion, more or less coupled to its adjacent N-H angular bending, amide III vibrations mainly arise from the backbone N-H angular bending. The Raman spectra obtained for the free amino acids (Fig. S2) are devoid of amide (I and III) vibrations. However, the presence of low intensity bands arising from the side chains of Val, Leu, Ile and Pro, falling within the amide III range should be stressed. These bands can be naturally superimposed to those relative to amide III vibrations in peptides. Nevertheless, because of their weakness, one cannot expect a considerable distortion in the structural analysis on the basis of amide III vibrations. In all nonapeptides, amide I and amide III vibrations are characterized by two strong, broad and incompletely resolved bands peaking at ~1685 and ~1265 cm^{-1}, respectively. To go farther in extracting the structural information from amide vibrations, the decomposition of the amide I region for four nonapeptides is displayed in Figure 4. Through a systematic investigation on the structural analysis of the peptide chains (42–44, 48, 49), in aqueous environment, we have been led to select four band components located at ~1690±5 (assignable to random chain), ~1660±5

(assignable to β-strand), ~1675±5 and ~1650±5 (both assignable to turns) to decompose the amide I region in nonapeptides. The populations corresponding to different secondary structural elements, as estimated by the normalized areas of the band components (expressed in percent) are reported in Table 3. As it can be seen between 50% to 60% of the total population is formed by turn structures, and the rest is equally distributed between β-strands and random chains.

Solution samples containing nonapeptides at 20 mM were heated up to 80°C. Little changes appeared in their Raman spectra, but slight changes could be observed in the amide III region. Figure 5 exemplifies this behavior for hN5 and hN6. The main effect is a wavenumber downshift of the spectral profile consistent with a progressive shift from turn to β-strand upon increasing temperature. Albeit the turns remain the major population of secondary structures in the 10-80° C temperature range (data not shown), they show a little tendency to be transformed into β-strands, presumably because of the weakening/breakdown of the inter-turn stabilizing hydrogen bonds occurring upon thermal annealing.

Atomistic molecular dynamics simulations

To get a more detailed representation of the structures that can be adopted by the four nonapeptides, 10 molecular dynamic simulations of 500 ns have been performed for each peptide. The adopted conformations, analyzed with DSSP (31), are presented on Figure 6A and Figure S3 in the Supporting Material. The figures show that the peptides mainly adopt random coil conformations with some residues favoring turns but no stable structure can be observed. The formation of β-turns is evidenced by distances between Cα atoms of residues *i* and *i+3* being less than 0.7 nm with residues *i+1* and *i+2* being non helical, which is mainly observed in the middle of the peptides between residues 3 and 6 and notably for the hN3 peptide (Fig. S4 in the Supporting Material, panel A). The type of β-turns can be identified by analyzing the ΦΨ angles of two amino acid residues following each other as determined by

Lewis *et al*. (32) and Hutchinson *et al*. (33). Figure 6B shows that type II β-turns are formed during a third to a half of the simulation time (52.9 ± 4.1% for hN3, 37.8 ± 4.7% for hN4, 33.9 ± 2.8% for hN5 and 33.5 ± 3.8% for hN6) for the PG residues being at positions 4 and 5. PG residues are strongly favored at *i+1* and *i+2* positions in experimental structure of type II β-turns and they are known to promote this type of β-turn. The population of type II β-turns is greater for the hN3 peptide than for the hN4, hN5 and hN6 peptides. Figures S4B-E show that type I, I', II, II' and VIII β-turns can also be formed in smaller amounts at other positions of the peptides and that type VIII β-turns are mainly observed at the start of the peptides (up to 3% of the simulation time). A proline at position 4 has been shown to promote the formation of this type of turn (33). Analysis of the simulation underlines that the structures are rapidly changing on the nanosecond time scale from extended to more packed conformations which mainly corresponds to type II β-turns (Fig. S4F). The population of β-turns has also been analyzed through clustering methods and the central structures of the dominant cluster population of each peptide is presented as an insert in Figure 6B. Altogether, our analysis of MD data strongly suggests that the PG motif of the nonapeptides could mainly adopt a type II β-turn conformation.

To test the agreement of these results with the measured CD spectra, CD spectra were computed from our simulated trajectories with the SESCA software (35). The results (Fig. S5 in the Supporting Material) show a very good agreement between the measured and computed spectra despite noticeable differences around ~220 nm. This suggests that the computed profiles might have failed to detect several experimental signals. Nevertheless, they reinforce our proposal that β-turns and notably type II β-turns are the prevailing structures adopted by the nonapeptides.

Discussion

Among elastin peptides, elastin nonapeptides have been identified initially as chemoattractants for fibroblasts (12). Now, there are growing evidences that these elastin peptides could define a peculiar class of matrikines.

Brassart and colleagues have recently shown that hN5 (*i.e.* AGVPGLGVG) plays a key role in tumor progression by promoting tumor cell blebbing and extracellular vesicle shedding following its interaction with the ribosomal protein SA (17). The structure of hN5 has been analyzed previously by CD, FT-IR and NMR spectroscopies (16). This report concluded that the structure of the peptide was a mixture of random coils and β-turns. Our data fully agree with these findings but they also expand our knowledge of the nature and structural behavior of these elastin peptides.

The occurrence of the X-G-X-P-G-X-G-X-G consensus sequence was analyzed in a large sequence data bank. Our results show that among the observed hits, the group of elastin sequences defines a specific group (Fig. 1). Further, our analysis underlines the fact that these peptides mostly occur in tandem repeats in human exon 26 (and corresponding sequences in other species). Finally, as far as human sequences are concerned, we observed that similar sequences were also present in numerous species (Table 1). The most striking observation is that hN6 is present in numerous species and is tandemly repeated five times in mouse sequence. In contrast, hN3 is only found in 3 species. These observations suggest that, besides their matrikines activity, these sequences could play a key, and yet undescribed, role in elastin biology, possibly elastin assembly. Indeed, among elastin sequences, these sequences appear much more conserved that the canonical elastin hexapeptide sequence (say VGVAPG), which occurrence is mostly limited to human exon 24 sequence. Further works are needed to test this hypothesis.

Our data (Fig. 1) show that the X-G-X-P-G-X-G-X-G consensus sequence can be found in most filum and/or species. To the best of our knowledge, sequences harboring this consensus

have caught the attention only in elastin because these elastin peptides are bioactive (8, 12–15, 17). Up to now, neither a biological activity nor a dedicated receptor have been described for the non-elastin sequences.

As evidenced by the consensus profile reported for each of them (Fig. 1), the non-elastin sequences (96 259 sequences) present a sequence pattern that is mostly dominated by the occurrence of G, A or K at X positions. These residues are not favored in elastin sequences where AGVPGFGVG is the prevailing motif. This is remarkably pronounced when non-elastin animal sequences are considered. Their consensus is GGGPGGGGG. These important sequence differences suggest that nonapeptides found in non-elastin sequences would most probably be structured and behave differently than elastin nonapeptides. Nevertheless, a comprehensive structural analysis is required to address this point.

In the present study, we focused our analysis on the nonapeptide sequences found in human exon 26 aiming at understanding their structure and structural behavior. First, this investigation was performed using optical spectroscopies. The results gathered by both CD and Raman spectra are consistent with previous reports underlining a conformational equilibrium between random, strand and turns conformations within elastin sequences (50). The Raman data underline that the conformation of these peptides is dominated by β-turns as deduced from the quantitative analysis of their amide I profiles. These conformations appear very favored as increasing the temperature up to 80°C did not drastically change the spectra and hence their structure, albeit a small increase of the β-strands content could be observed upon heating. Altogether our experimental data suggest that the structures of the nonapeptides are dominated by β-turns and that these conformations could be engaged in a conformational equilibrium with extended and random structures.

In order to have a better understanding of the structural behavior of these peptides, molecular dynamics simulations were undertaken in explicit water. Ten independent 500-ns

simulations were performed for each of the 4 nonapeptides. The results obtained are extremely consistent with our experimental data. Indeed, they show that the most common structures observed during the simulation is a type II β-turn occurring at the X-P-G-X motif. Quantitatively, this structure seems more abundant in hN3 (about 50%), than hN4, hN5 or hN6 (25-35%). This finding is in very good agreement with the conclusions of Lessing and colleagues (51) who observed that the preference for the X-Pro-Gly sequence to form a β-turn increased with the complexity and hydrophobicity of the side chain at position X (Gly < Ala < Val). In contrast to hN4, hN5 and hN6 which harbor a valyl residue before PG, hN3 possesses an isoleucyl residue at that position. Thus, the higher propensity of hN3 to adopt a β-turn when compared to the other three nonapeptides is consistent with the work of Lessing and colleagues.

Since the first report of their bioactivity on bovine fibroblasts (12), nonapeptides sequences have been tested on various cellular systems: bovine endothelial cells (13), rat macrophages (14) and human tumor cell lines (8, 9). Nevertheless, among those publications few used more than one peptide sequence, thereby forbidding the comparison of their respective effects. In the work of Maeda and colleagues, hN5 and hN6 were identified as chemoattractants for rat macrophages (14). In this work, the authors also analyzed the effect of the canonical VGVAPG sequence. Their conclusion was that the two nonapeptides could compete for receptor binding on EBP. Indeed, while VGVAPG appeared as an exclusive ligand for EBP, they underlined that hN5 and hN6 could also bind EBP and a lactose-insensitive receptor, later identified as RPSA. Their work therefore suggest that the EBP-binding motif can be harbored by both hexa- and nonapeptides while RPSA binding motif can be only adopted by nonapeptides. It has been established that VGVAPG bioactivity relies on its ability to adopt mainly a type VIII β-turn conformation (10). Our molecular dynamics simulations show that the major conformation of nonapeptides could be the type II β-turn, but

also that type VIII turns can be present. Our experimental data also suggest that both type VIII and type II β-turn can be present in the conformation of the nonapeptides. Thus, at the difference of VGVAPG, nonapeptides could be favorably structured to bind EBP and RPSA. However, it has to be emphasized that the sequences of the hexa- and nonapeptides are different. That way, binding affinity could also be supported by favorable residue side chain interactions in addition to the locally adopted conformation. Sequence differences could also contribute to the differential bioactivity exhibited by these peptides.

Altogether our data suggest that the structural signature of elastin nonapeptide is a type II β-turn. Consequently, it is reasonable to propose that this conformation could be relevant of the observed bioactivity of these peptides. This proposal is supported by the observations made by Toupance and co-workers where they measured a lesser biological effect of hN6 towards human tumor cells than hN5 (8). In our work, hN6 is predicted to present less β-turn conformations than hN5. Therefore, it seems coherent that the bioactivity of hN5 is greater than that of hN6 supporting the proposal that the turn motif is the conformation sustaining bioactivity of these peptides. If our prediction is correct, then one can anticipate that hN3 would be a peptide more bioactive that others.

Despite decades of efforts, the structure/function relationship of RPSA is still an open question (52). Notably, the interaction site of elastin nonapeptides on RSPA is unknown. In contrast, RPSA laminin binding site was described as a secondary function acquired during evolution (53). This site corresponds to the 161-180 region of the human sequence. However, the crystal structure of the protein (54) locates this binding site mostly inside the fold. Therefore, it is thought that RPSA could possess multiple sites and/or associate with other compounds to exert its functions. Consequently, our current knowledge of RPSA structure cannot explain where and how it binds elastin nonapeptides. Further structural explorations are required to address this question.

Conclusion

Our data throw new light on the nature, structure and conformational behavior of elastin nonapeptides. These peptide sequences exhibit a peculiar composition as demonstrated by our bioinformatic analysis of current databases. Further, experimental (CD and Raman spectroscopies) and theoretical data (molecular dynamics) strongly suggest that the dynamic structure of these peptides is dominated by the type II β-turn conformation. We propose that the observed bioactivity of these peptides towards stromal and tumor cells could be driven by the recognition of this structural feature by their cognate receptor, RPSA. Molecular modeling of their interaction is underway.

Authors contributions

HB and KSG performed the Raman and CD measurements and analyzed the spectra. BS conceived and ran the simulations. CJM, TJ and BN performed the sequence analyses. BS and DM funded the project and edited the manuscript. DL supervised the work and wrote the manuscript.

Acknowledgments

The authors would like to thank the MAgICS chair (Université de Reims Champagne Ardenne) and the Centre National de la Recherche Scientifique (CNRS) for financial support. Molecular dynamics calculations were performed on the ROMEO high performance computing center (https://romeo.univ-reims.fr).

References

1. Mecham, R.P. 2011. The Extracellular Matrix: an Overview. Berlin, Heidelberg: Springer Berlin Heidelberg.

2. Duca, L., S. Blaise, B. Romier, M. Laffargue, S. Gayral, H. El Btaouri, C. Kawecki, A. Guillot, L. Martiny, L. Debelle, and P. Maurice. 2016. Matrix ageing and vascular impacts: focus on elastin fragmentation. Cardiovasc. Res. 110: 298–308.

3. Maurice, P., S. Blaise, S. Gayral, L. Debelle, M. Laffargue, W. Hornebeck, and L. Duca. 2013. Elastin fragmentation and atherosclerosis progression: The elastokine concept. Trends Cardiovasc. Med. 23: 211–221.

4. Gayral, S., R. Garnotel, A. Castaing-Berthou, S. Blaise, A. Fougerat, E. Berge, A. Montheil, N. Malet, M.P. Wymann, P. Maurice, L. Debelle, L. Martiny, L.O. Martinez, A. Pshezhetsky, L. Duca, and M. Laffargue. 2014. Elastin-derived peptides potentiate atherosclerosis through the immune Neu1-PI3Kγ pathway. Cardiovasc. Res. 102: 118-127.

5. Pocza, P., H. Süli-Vargha, Z. Darvas, and A. Falus. 2008. Locally generated VGVAPG and VAPG elastin-derived peptides amplify melanoma invasion via the galectin-3 receptor. Int. J. Cancer. 122: 1972–1980.

6. Coquerel, B., F. Poyer, F. Torossian, V. Dulong, G. Bellon, I. Dubus, A. Reber, and J.-P. Vannier. 2009. Elastin-derived peptides: Matrikines critical for glioblastoma cell aggressiveness in a 3-D system. Glia. 57: 1716–1726.

7. Devy, J., L. Duca, B. Cantarelli, D. Joseph-Pietras, A. Scandolera, A. Rusciani, L. Parent, J. Thevenard, S.B. Pasco, M. Tarpin, L. Martiny, and L. Debelle. 2010. Elastin-derived peptides enhance melanoma growth in vivo by upregulating the activation of Mcol-A (MMP-1) collagenase. Br. J. Cancer. 103: 1562–1570.

8. Toupance, S., B. Brassart, F. Rabenoelina, C. Ghoneim, L. Vallar, M. Polette, L. Debelle, and P. Birembaut. 2012. Elastin-derived peptides increase invasive capacities of lung cancer cells by post-transcriptional regulation of MMP-2 and uPA. Clin. Exp. Metastasis. 29: 511–522.

9. Brassart, B., P. Fuchs, E. Huet, A.J. Alix, J. Wallach, A.M. Tamburro, F. Delacoux, B. Haye, H. Emonard, W. Hornebeck, and L. Debelle. 2001. Conformational dependence of collagenase (matrix metalloproteinase-1) up-regulation by elastin peptides in cultured fibroblasts. J. Biol. Chem. 276: 5222–5227.

10. Blanchevoye, C., N. Floquet, A. Scandolera, S. Baud, P. Maurice, O. Bocquet, S. Blaise, C. Ghoneim, B. Cantarelli, F. Delacoux, M. Dauchez, R.G. Efremov, L. Martiny, L. Duca, and L. Debelle. 2013. Interaction between the Elastin Peptide VGVAPG and Human Elastin Binding Protein. J. Biol. Chem. 288: 1317–1328.

11. Scandolera, A., L. Odoul, S. Salesse, A. Guillot, S. Blaise, C. Kawecki, P. Maurice, H. El Btaouri, B. Romier-Crouzet, L. Martiny, L. Debelle, and L. Duca. 2016. The Elastin Receptor Complex: A Unique Matricellular Receptor with High Anti-tumoral Potential. Front. Pharmacol. 7: 32.

12. Long, M.M., V.J. King, K.U. Prasad, and D.W. Urry. 1988. Chemotaxis of fibroblasts toward nonapeptide of elastin. Biochim. Biophys. Acta. 968: 300–311.

13. Long, M.M., V.J. King, K.U. Prasad, B.A. Freeman, and D.W. Urry. 1989. Elastin repeat peptides as chemoattractants for bovine aortic endothelial cells. J. Cell. Physiol. 140: 512–518.

14. Maeda, I., N. Mizoiri, M.P.P. Briones, and K. Okamoto. 2007. Induction of macrophage migration through lactose-insensitive receptor by elastin-derived nonapeptides and their analog. J. Pept. Sci. 13: 263–268.

15. Salesse, S., L. Odoul, L. Chazée, C. Garbar, L. Duca, L. Martiny, R. Mahmoudi, and L. Debelle. 2018. Elastin molecular aging promotes MDA-MB-231 breast cancer cell invasiveness. FEBS Open Bio. 8: 1395–1404.

16. Da Silva, J., P. Lameiras, A. Beljebbar, A. Berquand, M. Villemin, L. Ramont, S. Dukic, J.-M. Nuzillard, M. Molinari, and M. Gautier. 2018. Structural characterization and in vivo pro-tumor properties of a highly conserved matrikine. Oncotarget. 9: 17839.

17. Brassart, B., J. Da Silva, M. Donet, E. Seurat, F. Hague, C. Terryn, F. Velard, J. Michel, H. Ouadid-Ahidouch, J.-C. Monboisse, A. Hinek, F.-X. Maquart, L. Ramont, and S. Brassart-Pasco. 2019. Tumour cell blebbing and extracellular vesicle shedding: key role of matrikines and ribosomal protein SA. Br. J. Cancer. 120: 453–465.

18. Huerta-Cepas, J., F. Serra, and P. Bork. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. 33: 1635–1638.

19. Crooks, G.E., G. Hon, J.-M. Chandonia, and S.E. Brenner. 2004. WebLogo: A Sequence Logo Generator. Genome Res. 14: 1188–1190.

20. Lefèvre, T., M.-E. Rousseau, and M. Pézolet. 2007. Protein Secondary Structure and Orientation in Silk as Revealed by Raman Spectromicroscopy. Biophys. J. 92: 2885–2895.

21. Dolui, S., A. Mondal, A. Roy, U. Pal, S. Das, A. Saha, and N.C. Maiti. 2020. Order, Disorder, and Reorder State of Lysozyme: Aggregation Mechanism by Raman Spectroscopy. J. Phys. Chem. B. 124: 50–60.

22. Debelle, L., A.J.P. Alix, M.-P. Jacob, J.-P. Huvenne, M. Berjot, B. Sombret, and P. Legrand. 1995. Bovine Elastin and k-Elastin Secondary Structure Determination by Optical Spectroscopies. J. Biol. Chem. 270: 26099–26103.

23. Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, and D.E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78: 1950–1958.

24. Abraham, M.J., T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 1–2: 19–25.

25. Schrodinger, L.L.C. 2010. The PyMOL molecular graphics system. Version. 1: 0.

26. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126: 014101.

27. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52: 7182–7190.

28. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen. 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103: 8577–8593.

29. Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18: 1463–1472.

30. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J. Mol.

Graph. 14: 33–38, 27–28.

31. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 22: 2577–2637.

32. Lewis, P.N., F.A. Momany, and H.A. Scheraga. 1973. Chain reversals in proteins. Biochim. Biophys. Acta BBA - Protein Struct. 303: 211–229.

33. Hutchinson, E.G., and J.M. Thornton. 1994. A revised set of potentials for β-turn formation in proteins. Protein Sci. 3: 2207–2216.

34. Daura, X., K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, and A.E. Mark. 1999. Peptide Folding: When Simulation Meets Experiment. Angew. Chem. Int. Ed. 38: 236–240.

35. Nagy, G., M. Igaev, N.C. Jones, S.V. Hoffmann, and H. Grubmüller. 2019. SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. J. Chem. Theory Comput. 15: 5087–5102.

36. Hernández, B., F.-Z. Boukhalfa-Heniche, O. Seksek, Y.-M. Coïc, and M. Ghomi. 2006. Secondary conformation of short lysine- and leucine-rich peptides assessed by optical spectroscopies: Effect of chain length, concentration, solvent, and time. Biopolymers. 81: 8–19.

37. Perczel, A., and G.D. Fasman. 1992. Quantitative analysis of cyclic β-turn models. Protein Sci. 1: 378–395.

38. Lopes, J.L.S., A.J. Miles, L. Whitmore, and B.A. Wallace. 2014. Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: Applications in secondary structure analyses. Protein Sci. 23: 1765–1772.

39. Sibanda, B.L., and J.M. Thornton. 1985. β -Hairpin families in globular proteins. Nature. 316: 170.

40. Gunasekaran, K., C. Ramakrishnan, and P. Balaram. 1997. Beta-hairpins in proteins revisited: lessons for de novo design. Protein Eng. Des. Sel. 10: 1131–1141.

41. Kaur, H., and G.P.S. Raghava. 2004. A neural network method for prediction of β-turn types in proteins using evolutionary information. Bioinformatics. 20: 2751–2758.

42. Hernández, B., Y.-M. Coïc, S.G. Kruglik, C. Carelli, R. Cohen, and M. Ghomi. 2012. Octreotide Used for Probing the Type-II′ β-Turn CD and Raman Markers. J. Phys. Chem. B. 116: 9337–9345.

43. Hernández, B., Y.-M. Coïc, B. Baron, S.G. Kruglik, F. Pflüger, R. Cohen, C. Carelli, and M. Ghomi. 2014. Low concentration structural dynamics of lanreotide and somatostatin-14. Biopolymers. 101: 1019–1028.

44. Hernández, B., E. López-Tobar, S. Sanchez-Cortes, Y.-M. Coïc, B. Baron, A. Chenal, S.G. Kruglik, F. Pflüger, R. Cohen, and M. Ghomi. 2016. From bulk to plasmonic nanoparticle surfaces: the behavior of two potent therapeutic peptides, octreotide and pasireotide. Phys. Chem. Chem. Phys. 18: 24437–24450.

45. Fuchs, P.F.J., A.M.J.J. Bonvin, B. Bochicchio, A. Pepe, A.J.P. Alix, and A.M. Tamburro. 2006. Kinetics and Thermodynamics of Type VIII β-Turn Formation: A CD, NMR, and Microsecond Explicit Molecular Dynamics Study of the GDNP Tetrapeptide. Biophys. J. 90: 2745–2759.

46. Hernández, B., F. Pflüger, A. Adenier, S.G. Kruglik, and M. Ghomi. 2010. Vibrational Analysis of Amino Acids and Short Peptides in Hydrated Media. VIII. Amino Acids with Aromatic Side

Chains: l-Phenylalanine, l-Tyrosine, and l-Tryptophan. J. Phys. Chem. B. 114: 15319–15330.

47. Hernández, B., F. Pflüger, S.G. Kruglik, and M. Ghomi. 2013. Characteristic Raman lines of phenylalanine analyzed by a multiconformational approach. J. Raman Spectrosc. 44: 827–833.

48. Guiffo-Soh, G., B. Hernández, Y.-M. Coïc, F.-Z. Boukhalfa-Heniche, and M. Ghomi. 2007. Vibrational Analysis of Amino Acids and Short Peptides in Hydrated Media. II. Role of KLLL Repeats To Induce Helical Conformations in Minimalist LK-Peptides. J. Phys. Chem. B. 111: 12563–12572.

49. Guiffo-Soh, G., B. Hernández, Y.-M. Coïc, F.-Z. Boukhalfa-Heniche, G. Fadda, and M. Ghomi. 2008. Vibrational Analysis of Amino Acids and Short Peptides in Hydrated Media. 3. Successive KL Repeats Induce Highly Stable β-Strands Capable of Forming Non-H-Bonded Aggregates. J. Phys. Chem. B. 112: 1282–1289.

50. Debelle, L., and A.M. Tamburro. 1999. Elastin: molecular description and function. Int. J. Biochem. Cell Biol. 31: 261–272.

51. Lessing, J., S. Roy, M. Reppert, M. Baer, D. Marx, T.L.C. Jansen, J. Knoester, and A. Tokmakoff. 2012. Identifying Residual Structure in Intrinsically Disordered Systems: A 2D IR Spectroscopic Study of the GVGXPGVG Peptide. J. Am. Chem. Soc. 134: 5032–5035.

52. DiGiacomo, V., and D. Meruelo. 2016. Looking into laminin receptor: critical discussion regarding the non-integrin 37/67-kDa laminin receptor/RPSA protein: Looking into laminin receptor. Biol. Rev. 91: 288–310.

53. Ardini, E., G. Pesole, E. Tagliabue, A. Magnifico, V. Castronovo, M.E. Sobel, M.I. Colnaghi, and S. Menard. 1998. The 67-kDa laminin receptor originated from a ribosomal protein that acquired a dual function during evolution. Mol. Biol. Evol. 15: 1017–1025.

54. Jamieson, K.V., J. Wu, S.R. Hubbard, and D. Meruelo. 2008. Crystal Structure of the Human Laminin Receptor Precursor. J. Biol. Chem. 283: 3002–3005.

Acknowledgments

The authors would like to thank the MAgICS chair (Université de Reims Champagne Ardenne) and the Centre National de la Recherche Scientifique (CNRS) for financial support. Molecular dynamics calculations were performed on the ROMEO high performance computing center.

Table 1. Nonapeptide stack of elastin Uniprot entries
Species, name of the species. Acc#, accession number of the origin sequence. #, number of the hit within the sequence. Position, position of the hit. Exon, exon number (when available). Human sequences appear in bold faces.

Species	Acc#	#	Sequence	Position	Exon
Oreochromis niloticus	I3JQ46	8	AGGPGSGIG	657 – 665	18
Odobenus rosmarus divergens	A0A2U3ZFC3	5	AGIPGFGAG	541 – 549	
Canis lupus familiaris	J9NW15	5	AGIPGFGVG	571 – 579	23
Homo sapiens	**P15502**	**3**	**AGIPGLGVG**	**576 – 584**	**26**
Rhinopithecus roxellana	A0A2K6NWE2	3	AGIPGLGVG	562 – 570	26
Pan troglodytes	A0A2I3SVF0	3	AGIPGLGVG	580 – 588	27
Macaca mulatta	A0A1D5R663	2	AGIPGVGPG	312 – 320	15
Papio anubis	A0A2I3MW49	3	AGVPGFGAG	472 – 480	16
Cercocebus atys	A0A2K5LP54	3	AGVPGFGAG	512 – 520	18
Felis catus	A0A337S1Z3	4	AGVPGFGAG	404 – 412	18
Ailuropoda melanoleuca	G1MA94	4	AGVPGFGAG	499 – 507	23
Canis lupus familiaris	J9NW15	6	AGVPGFGAG	580 – 588	23
Myotis lucifugus	G1QCG1	2	AGVPGFGAG	483 – 491	24
Myotis lucifugus	G1QCG1	3	AGVPGFGAG	492 – 500	24
Myotis lucifugus	G1QCG1	4	AGVPGFGAG	501 – 509	24
Sus scrofa	A0A287AXU0	5	AGVPGFGAG	599 – 607	25
Homo sapiens	**P15502**	**6**	**AGVPGFGAG**	**603 – 611**	**26**
Rattus norvegicus	Q99372	3	AGVPGFGAG	642 – 650	26
Rattus norvegicus	Q99372	4	AGVPGFGAG	658 – 666	26
Mus musculus	P54320	2	AGVPGFGAG	613 – 621	27
Mus musculus	P54320	3	AGVPGFGAG	622 – 630	27
Mus musculus	P54320	4	AGVPGFGAG	631 – 639	27
Mus musculus	P54320	5	AGVPGFGAG	640 – 648	27
Mus musculus	P54320	6	AGVPGFGAG	649 – 657	27
Mustela putorius furo	M3YXB8	4	AGVPGFGAG	608 – 616	27
Pan troglodytes	A0A2I3SVF0	6	AGVPGFGAG	607 – 615	27
Erinaceus europaeus	A0A1S3W547	3	AGVPGFGAG	584 – 592	
Erinaceus europaeus	A0A1S3W547	4	AGVPGFGAG	593 – 601	
Enhydra lutris kenyoni	A0A2Y9J5B3	4	AGVPGFGAG	615 – 623	
Tursiops truncatus	A0A2U4A8K1	1	AGVPGFGAG	87 – 95	
Ursus maritimus	A0A384CQV9	6	AGVPGFGAG	583 – 591	
Mesocricetus auratus	A0A1U8CCM2	2	AGVPGFGAG	643 – 651	
Mesocricetus auratus	A0A1U8CCM2	3	AGVPGFGAG	652 – 660	
Mesocricetus auratus	A0A1U8CCM2	4	AGVPGFGAG	661 – 669	
Mesocricetus auratus	A0A1U8CCM2	5	AGVPGFGAG	670 – 678	
Dipodomys ordii	A0A1S3GKU4	6	AGVPGFGAG	587 – 595	
Erinaceus europaeus	A0A1S3W547	5	AGVPGFGGG	602 – 610	
Balaenoptera acutorostrata scammoni	A0A384AI97	3	AGVPGFGPG	584 – 592	
Felis catus	A0A337S1Z3	2	AGVPGFGVG	386 – 394	18
Felis catus	A0A337S1Z3	3	AGVPGFGVG	395 – 403	18
Canis lupus familiaris	J9NW15	3	AGVPGFGVG	553 – 561	23
Canis lupus familiaris	J9NW15	4	AGVPGFGVG	562 – 570	23
Saimiri boliviensis boliviensis	A0A2K6UXV4	1	AGVPGFGVG	491 – 499	25
Sus scrofa	A0A287AXU0	3	AGVPGFGVG	581 – 589	25
Sus scrofa	A0A287AXU0	4	AGVPGFGVG	590 – 598	25
Callithrix jacchus	A0A2R8P8B8	3	AGVPGFGVG	523 – 531	26
Mustela putorius furo	M3YXB8	2	AGVPGFGVG	590 – 598	27
Mustela putorius furo	M3YXB8	3	AGVPGFGVG	599 – 607	27
Odobenus rosmarus divergens	A0A2U3ZFC3	4	AGVPGFGVG	532 – 540	
Enhydra lutris kenyoni	A0A2Y9J5B3	2	AGVPGFGVG	597 – 605	
Enhydra lutris kenyoni	A0A2Y9J5B3	3	AGVPGFGVG	606 – 614	
Ursus maritimus	A0A384CQV9	4	AGVPGFGVG	565 – 573	
Ursus maritimus	A0A384CQV9	5	AGVPGFGVG	574 – 582	
Dipodomys ordii	A0A1S3GKU4	4	AGVPGLGAG	569 – 577	
Dipodomys ordii	A0A1S3GKU4	5	AGVPGLGAG	578 – 586	
Danio rerio	F8W4J7	4	AGVPGLGGG	568 – 576	18
Ailuropoda melanoleuca	G1MA94	3	AGVPGLGVG	490 – 498	23
Saimiri boliviensis boliviensis	A0A2K6UXV4	2	AGVPGLGVG	500 – 508	25
Rhinopithecus bieti	A0A2K6MP70	3	AGVPGLGVG	522 – 530	25
Bos taurus	P04985	1	AGVPGLGVG	562 – 570	26
Bos taurus	P04985	2	AGVPGLGVG	571 – 579	26
Bos taurus	P04985	3	AGVPGLGVG	580 – 588	26
Homo sapiens	**P15502**	**5**	**AGVPGLGVG**	**594 – 602**	**26**
Rattus norvegicus	Q99372	2	AGVPGLGVG	633 – 641	26

Species	Acc#	#	Sequence	Position	Exon
Callithrix jacchus	A0A2R8P8B8	4	AGVPGLGVG	532 - 540	26
Callithrix jacchus	A0A2R8P8B8	5	AGVPGLGVG	541 - 549	26
Callithrix jacchus	A0A2R8P8B8	6	AGVPGLGVG	550 - 558	26
Colobus angolensis palliatus	A0A2K5JW24	3	AGVPGLGVG	548 - 556	26
Bos mutus	L8IPH2	1	AGVPGLGVG	564 - 572	26
Macaca nemestrina	A0A2K6BHK1	4	AGVPGLGVG	563 - 571	26
Pan troglodytes	A0A2I3SVF0	5	AGVPGLGVG	598 - 606	27
Ursus maritimus	A0A384CQV9	3	AGVPGLGVG	556 - 564	
Ficedula albicollis	U3KH66	1	AGVPGVGVG	359 - 367	15
Equus caballus	F7BWV3	1	FGVPGYGVG	418 - 426	17
Oryzias latipes	A0A3B3HSK9	1	GGAPGFGPG	383 - 391	14
Tetraodon nigroviridis	H3CXZ1	1	GGAPGSGPG	373 - 381	8
Sus scrofa	A0A287AXU0	2	GGFPGFGVG	429 - 437	19
Homo sapiens	**P15502**	**2**	**GGFPGFGVG**	**401 - 409**	**20**
Pan troglodytes	A0A2I3SVF0	2	GGFPGFGVG	405 - 413	20
Callithrix jacchus	A0A2R8P8B8	2	GGFPGFGVG	408 - 416	21
Balaenoptera acutorostrata scammoni	A0A384AI97	2	GGFPGFGVG	440 - 448	
Erinaceus europaeus	A0A1S3W547	2	GGFPGYGAG	426 - 434	
Canis lupus familiaris	J9NW15	2	GGFPGYGIG	399 - 407	17
Ailuropoda melanoleuca	G1MA94	2	GGFPGYGIG	395 - 403	19
Odobenus rosmarus divergens	A0A2U3ZFC3	2	GGFPGYGIG	396 - 404	
Enhydra lutris kenyoni	A0A2Y9J5B3	1	GGFPGYGIG	456 - 464	
Ursus maritimus	A0A384CQV9	2	GGFPGYGIG	426 - 434	
Felis catus	A0A337S1Z3	1	GGFPGYGVG	289 - 297	14
Papio anubis	A0A2I3MW49	1	GGFPGYGVG	338 - 346	14
Cercocebus atys	A0A2K5LP54	1	GGFPGYGVG	354 - 362	15
Mandrillus leucophaeus	A0A2K5XQM6	2	GGFPGYGVG	284 - 292	15
Rattus norvegicus	Q99372	1	GGFPGYGVG	457 - 465	18
Rhinopithecus bieti	A0A2K6MP70	2	GGFPGYGVG	377 - 385	20
Mus musculus	P54320	1	GGFPGYGVG	454 - 462	21
Colobus angolensis palliatus	A0A2K5JW24	2	GGFPGYGVG	405 - 413	21
Mustela putorius furo	M3YXB8	1	GGFPGYGVG	445 - 453	21
Rhinopithecus roxellana	A0A2K6NWE2	2	GGFPGYGVG	405 - 413	21
Macaca nemestrina	A0A2K6BHK1	2	GGFPGYGVG	405 - 413	21
Mesocricetus auratus	A0A1U8CCM2	1	GGFPGYGVG	466 - 474	
Dipodomys ordii	A0A1S3GKU4	3	GGFPGYGVG	405 - 413	
Canis lupus familiaris	J9NW15	1	GGGPGAGLG	54 - 62	2
Danio rerio	F8W4J7	1	GGGPGAGLG	33 - 41	2
Odobenus rosmarus divergens	A0A2U3ZFC3	1	GGGPGAGLG	71 - 79	
Dipodomys ordii	A0A1S3GKU4	1	GGGPGAGLG	81 - 89	
Oreochromis niloticus	I3JQ46	1	GGGPGFGGG	541 - 549	16
Oreochromis niloticus	I3JQ46	2	GGGPGFGGG	559 - 567	16
Oreochromis niloticus	I3JQ46	3	GGGPGFGGG	577 - 585	16
Oreochromis niloticus	I3JQ46	4	GGGPGFGGG	589 - 597	16
Oreochromis niloticus	I3JQ46	9	GGGPGLGLG	900 - 908	23
Oreochromis niloticus	I3JQ46	7	GGIPGLGYG	632 - 640	16
Danio rerio	F8W4J7	28	GGIPGVGYG	1190 - 1198	30
Danio rerio	F8W4J7	13	GGLPGGGAG	821 - 829	22
Danio rerio	F8W4J7	22	GGLPGGGAG	1033 - 1041	26
Danio rerio	F8W4J7	23	GGLPGGGAG	1041 - 1049	26
Danio rerio	F8W4J7	26	GGLPGGGAG	1117 - 1125	28
Danio rerio	F8W4J7	18	GGLPGGGIG	956 - 964	24
Danio rerio	F8W4J7	16	GGLPGGGLG	895 - 903	24
Danio rerio	F8W4J7	20	GGLPGGGLG	1005 - 1013	26
Danio rerio	F8W4J7	21	GGLPGGGLG	1025 - 1033	26
Danio rerio	F8W4J7	25	GGLPGGGLG	1090 - 1098	28
Oreochromis niloticus	I3JQ46	11	GGLPGGGPG	1844 - 1852	49
Oreochromis niloticus	I3JQ46	10	GGLPGGGTG	932 - 940	23
Danio rerio	F8W4J7	17	GGLPGGGVG	948 - 956	24
Danio rerio	F8W4J7	2	GGLPGIGAG	144 - 152	5
Danio rerio	F8W4J7	19	GGLPGSGIG	964 - 972	24
Danio rerio	F8W4J7	24	GGLPGSGIG	1049 - 1057	26
Danio rerio	F8W4J7	9	GGLPGSGLG	721 - 729	20
Danio rerio	F8W4J7	14	GGLPGSGTG	837 - 845	22
Danio rerio	F8W4J7	5	GGLPGSGVG	661 - 669	20
Danio rerio	F8W4J7	6	GGLPGSGVG	676 - 684	20
Danio rerio	F8W4J7	7	GGLPGSGVG	691 - 699	20
Danio rerio	F8W4J7	8	GGLPGSGVG	706 - 714	20

Species	Acc#	#	Sequence	Position	Exon
Danio rerio	F8W4J7	10	GGLPGSGVG	736 – 744	20
Danio rerio	F8W4J7	27	GGLPGSGVG	1180 – 1188	30
Oreochromis niloticus	I3JQ46	5	GGVPGFGGG	613 – 621	16
Myotis lucifugus	G1QCG1	1	GGVPGLGIG	442 – 450	22
Oreochromis niloticus	I3JQ46	6	GGVPGVGGG	625 – 633	16
Macaca nemestrina	A0A2K6BHK1	3	GGVPGVGVG	421 – 429	21
Sus scrofa	A0A287AXU0	1	PGAPGFGPG	330 – 338	17
Balaenoptera acutorostrata scammoni	A0A384AI97	1	PGAPGFGPG	315 – 323	
Dipodomys ordii	A0A1S3GKU4	2	PGFPGVGAG	303 – 311	
Mandrillus leucophaeus	A0A2K5XQM6	1	PGGPGFGPG	203 – 211	12
Ailuropoda melanoleuca	G1MA94	1	PGGPGFGPG	315 – 323	17
Homo sapiens	**P15502**	**1**	**PGGPGFGPG**	**323 – 331**	**18**
Pan troglodytes	A0A2I3SVF0	1	PGGPGFGPG	327 – 335	18
Rhinopithecus bieti	A0A2K6MP70	1	PGGPGFGPG	294 – 302	18
Callithrix jacchus	A0A2R8P8B8	1	PGGPGFGPG	339 – 347	19
Colobus angolensis palliatus	A0A2K5JW24	1	PGGPGFGPG	322 – 330	19
Rhinopithecus roxellana	A0A2K6NWE2	1	PGGPGFGPG	322 – 330	19
Macaca nemestrina	A0A2K6BHK1	1	PGGPGFGPG	322 – 330	19
Ursus maritimus	A0A384CQV9	1	PGGPGFGPG	333 – 341	
Erinaceus europaeus	A0A1S3W547	1	PGVPGFGRG	343 – 351	
Callorhinchus milii	V9KAK8	1	QGEPGLGGG	243 – 251	
Danio rerio	F8W4J7	3	TGLPGIGPG	463 – 471	12
Danio rerio	F8W4C7	1	TGRPGNGRG	137 – 145	3
Macaca mulatta	A0A1D5R663	1	VGAPGGGLG	302 – 310	13
Danio rerio	F8W4J7	11	VGLPGGGLG	792 – 800	22
Danio rerio	F8W4J7	15	VGLPGGGLG	875 – 883	24
Balaenoptera acutorostrata scammoni	A0A384AI97	5	VGVPGFGAG	602 – 610	
Odobenus rosmarus divergens	A0A2U3ZFC3	3	VGVPGFGAG	523 – 531	
Trichechus manatus latirostris	A0A2Y9RHH3	1	VGVPGFGAG	569 – 577	
Heterocephalus glaber	G5BG87	1	VGVPGFGAG	589 – 597	
Papio anubis	A0A2I3MW49	2	VGVPGLGVG	463 – 471	16
Cercocebus atys	A0A2K5LP54	2	VGVPGLGVG	503 – 511	18
Mandrillus leucophaeus	A0A2K5XQM6	3	VGVPGLGVG	423 – 431	19
Macaca mulatta	A0A1D5R663	3	VGVPGLGVG	535 – 543	23
Macaca mulatta	A0A1D5R663	4	VGVPGLGVG	544 – 552	23
Homo sapiens	**P15502**	**4**	**VGVPGLGVG**	**585 – 593**	**26**
Bos mutus	L8IPH2	2	VGVPGLGVG	573 – 581	26
Bos mutus	L8IPH2	3	VGVPGLGVG	582 – 590	26
Macaca nemestrina	A0A2K6BHK1	5	VGVPGLGVG	572 – 580	26
Pan troglodytes	A0A2I3SVF0	4	VGVPGLGVG	589 – 597	27
Balaenoptera acutorostrata scammoni	A0A384AI97	4	VGVPGLGVG	593 – 601	
Tursiops truncatus	A0A2U4A8K1	2	VGVPGLGVG	96 – 104	
Danio rerio	F8W4J7	12	YGLPGGGLG	813 – 821	22

Table 2. Tentative assignment of the Raman bands observed in the middle wavenumber region.

AGIPGLGVG hN3	VGVPGLGVG hN4	AGVPGLGVG hN5	AGVPGFGAG hN6	Tentative assignment
1700-1620 (br)	1700-1620 (br)	1700-1620 (br)	1700-1620 (br)	Amide I
			1605 (s)	F1
			1586 (m)	F2
1465 (s)	1468 (s)	1466 (s)		P, I, L, V, A
1451 (s)	1452 (s)	1452 (s)	1454 (s)	V, L, I
1421 (s)	1421 (s)	1420 (s)	1421 (s)	G, A, L, I, V, P
1391 (m)	1394 (m)	1395 (m)	1392 (m)	G, A, L, I, V, P
1343 (s)	1345 (s)	1343 (s)	1341 (s)	A, L, I, V, P
1315 (s)	1313 (s)	1317 (s)	1320 (s)	A, L, I
1300-1240 (br)	1300-1240 (br)	1300-1240 (br)	1300-1240 (br)	Amide III
			1207 (s)	F3
			1183 (m)	
1161 (m)	1175 (m)	1161 (m)	1159 (m)	V, L, I, P
1128 (s)	1127 (s)	1126 (s)	1126 (m)	V, L, I
			1101 (m)	P
			1031 (s)	F4
1031 (m)	1034 (m)	1031 (m)		G, I, P
1011 (m)	1014 (m)	1014 (m)		A, V
			1004 (vs)	F5
960 (s)	962 (s)	960 (s)		V, L, I
937 (s)	938 (s)	937 (s)	937 (br)	A, V, L, I
891 (s)	891 (s)	891 (s)		V
857 (m)	860 (m)	856 (sh)	851 (br)	L, P
845 (m)	840 (m)	843 (m)		A, V, L, I, P
			758 (br)	P
736 (br)	735 (br)	737 (br)		V, L, I, P
			621 (m)	F6

(vs) very strong; (s) strong; (m) medium; (br) broad Raman bands.
F1 to F6 refer to the six characteristic Raman bands of phenylalanine. Amide I and amide III vibrations give rise to a broad and strong Raman band peaking at ~1685 and 1255 cm^{-1}, respectively.

Table 3. Underlying secondary structure elements in the Amide I profile of the nonapeptides and fractional areas.

Position (assignment)	hN3	hN4	hN5	hN6
~1690 (random)	20	20	24	25
~1680 (turn)	41	40	37	38
~1660 (β-strand)	26	23	21	25
~1640 (turn)	13	17	18	12
Fractional areas				
Random	20	20	24	25
Turn	54	57	55	50
β-strand	26	23	21	25

Component positions are ± 5 cm^{-1}. During the computation, their widths were kept between 15 and 25 cm^{-1}. The figures correspond to the normalized areas of the components expressed in per cent of the whole Amide I profile. Fractional areas estimate the global secondary structures with an accuracy of ±5% (22).

Figure captions

Figure 1. Nonapeptide occurrences and distribution. The X-G-X-P-G-X-G-X-G consensus was found in 96 293 unique sequences amongst which 22 919 were from *Eukariota*, 71 504 from *Bacteria*, 240 from *Virus* and 1630 from *Archae*. The *Eukariota* branch distribution is dominated by Animals in which Vertebrates are predominant. For each branch of the tree, the total number of hits is reported as well as the number of observed overlapping (O) sequences and of tandem (T) repetitions. The consensus motif is also presented with the most frequent residues found at the four X positions.

Figure 2. Ultraviolet-circular dichroism (UV-CD) spectra of the four nonapeptides observed in water, water-methanol mixture (50/50), and methanol.

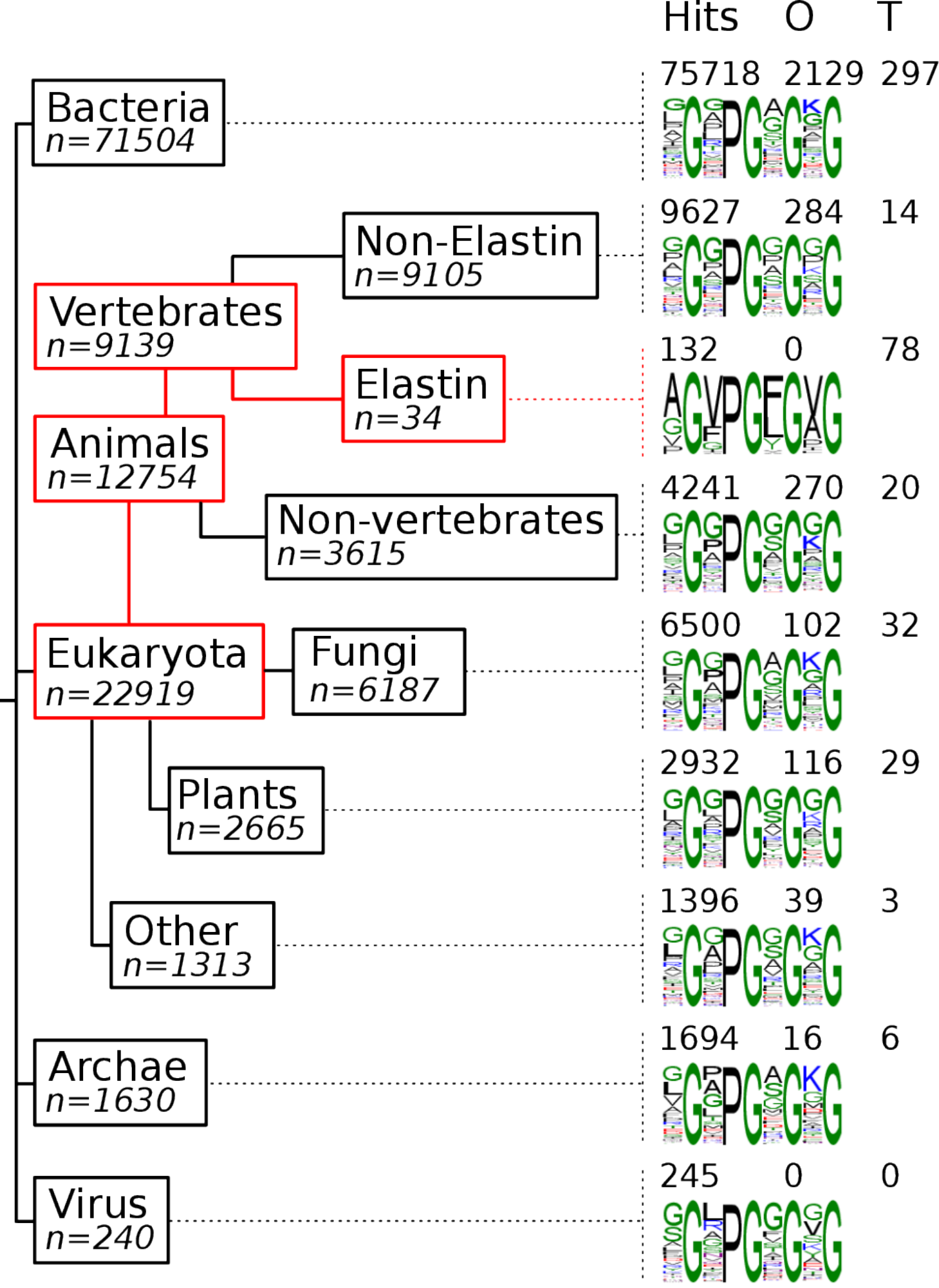
Figure 3. Room temperature Stokes Raman spectra in the middle wavenumber region for (A) hN3, hN4, hN5 and (B) hN6, compared to F. Spectra were recorded at 20 mM. See Table 2 for tentative assignments.

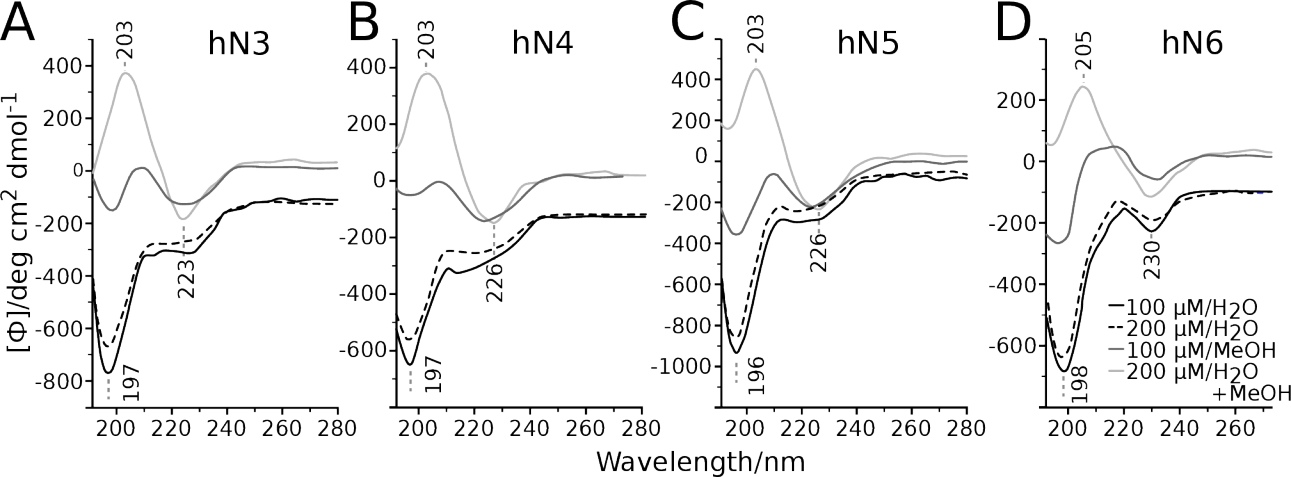
Figure 4. Band decomposition in the amide I region of the nonapeptides observed at 20 mM. Observed spectra are drawn with black lines, component bands with dark gray lines. The light gray trace corresponds to the sum of component bands (fit). Maximum band wavenumber of each used component is reported. In parentheses: (R) random; (T) turn; (S) β-strand. For the estimated populations of different secondary structural elements, see Table 3.

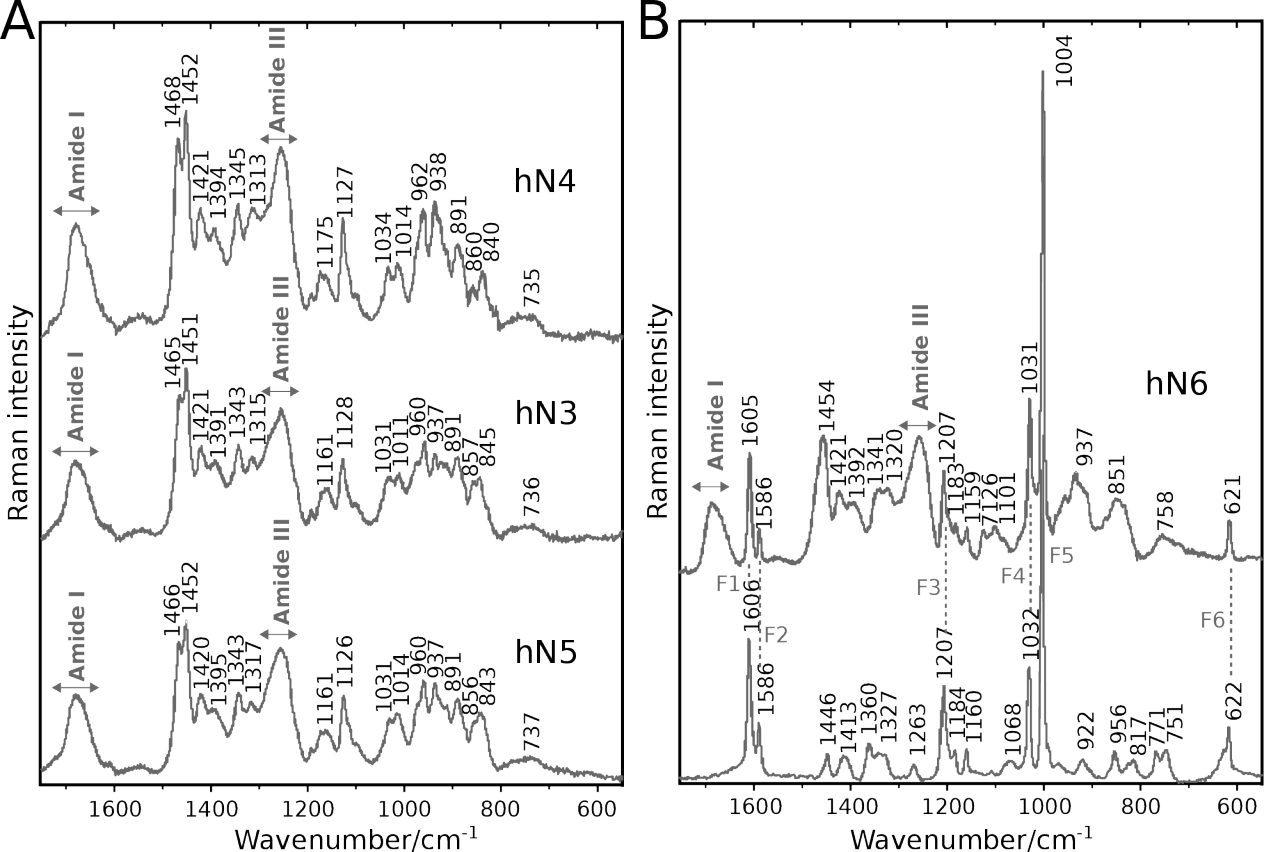
Figure 5. Effect of temperature on the amide III vibrations of nonapeptides. The spectra were recorded for a concentration of 20 mM. Gray scales are used to display the spectra obtained as a function of temperature.

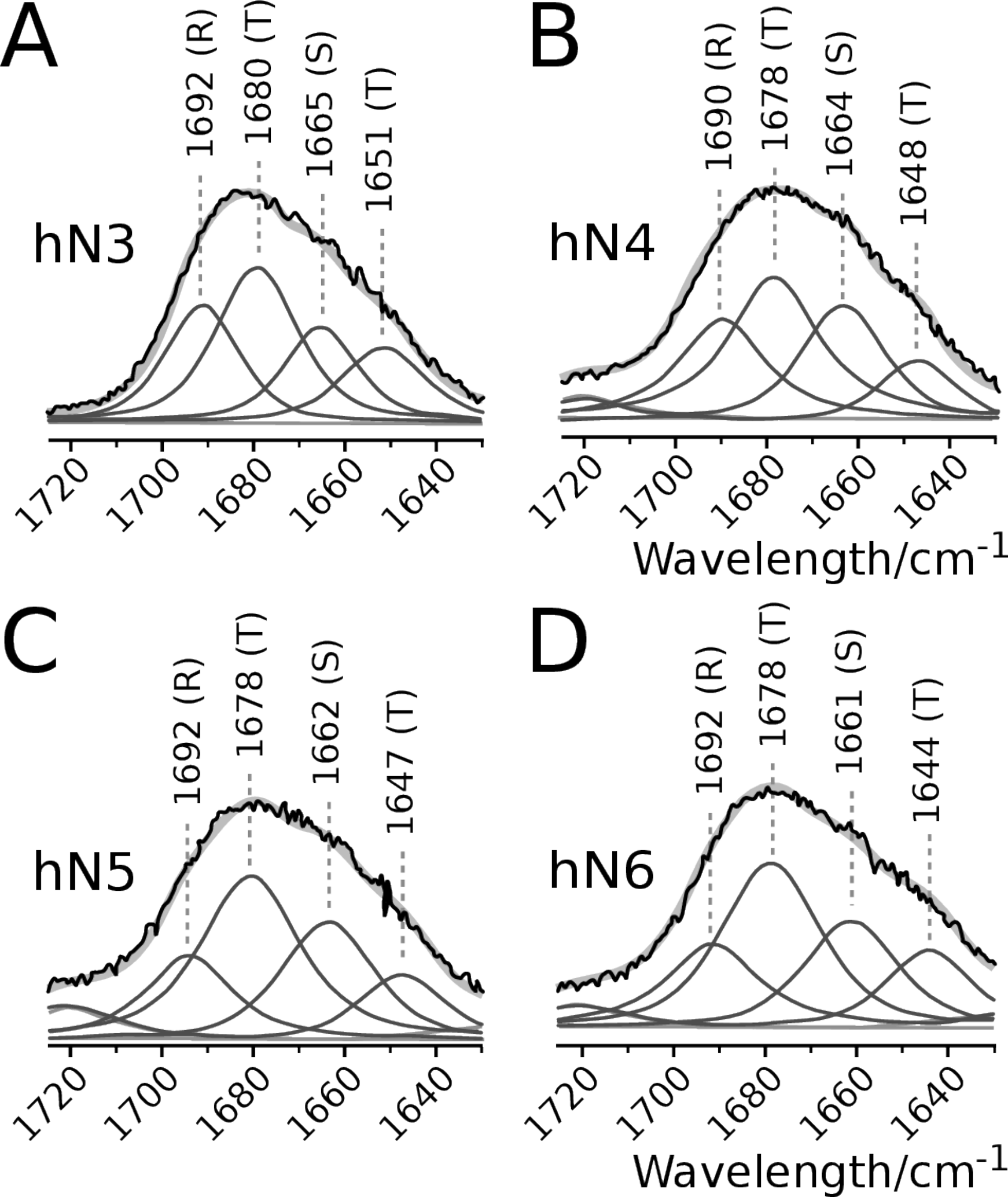
Figure 6. Conformational analysis of the molecular dynamics simulations. A, secondary structures adopted during the simulation calculated with DSSP (31). B, occurrence of β-turns types during each frame of the simulation trajectories. β-turns are computed considering the

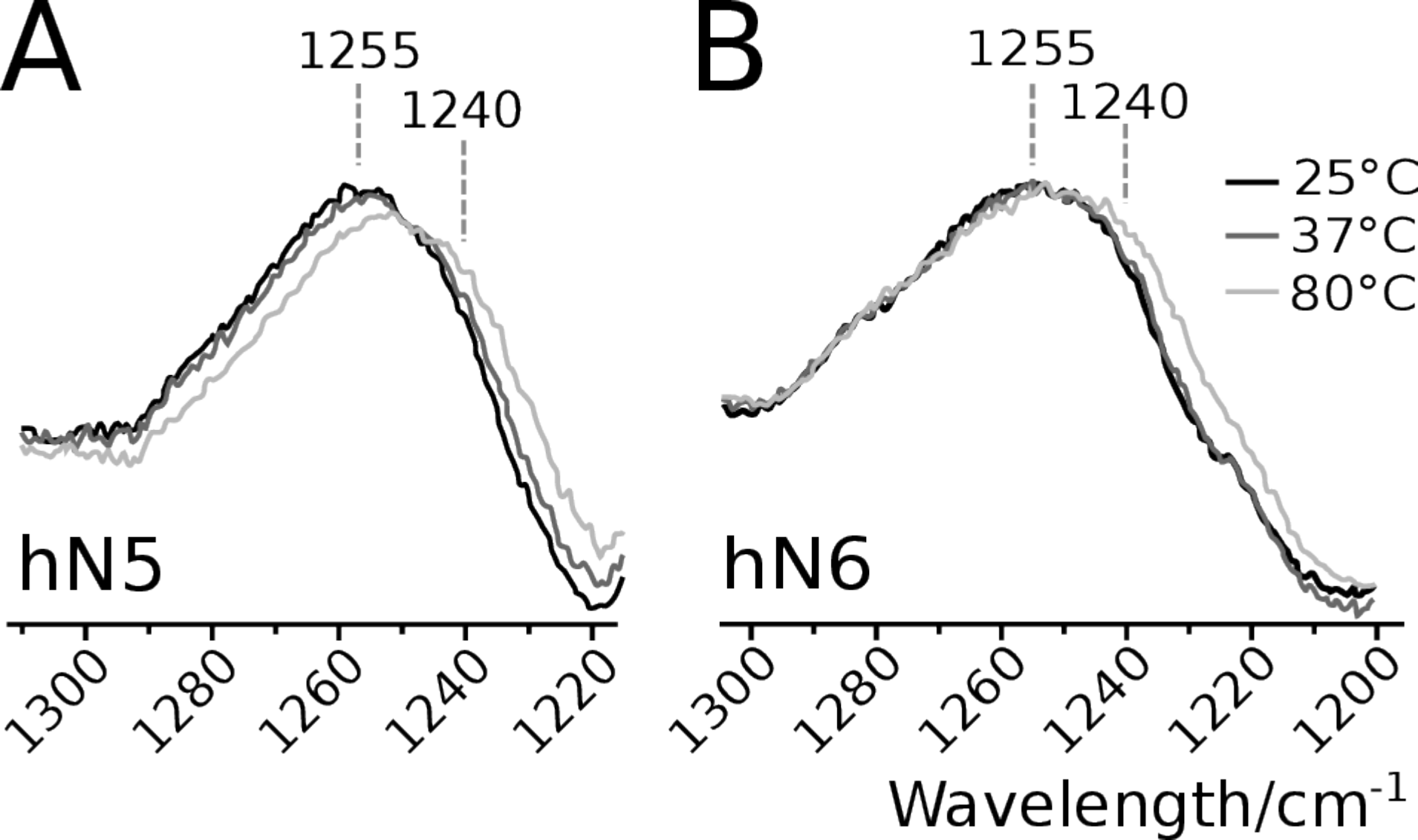
ΦΨ angles of the PG residues at positions 4 and 5. Insert, overlay of the conformation representative of the main cluster computed for each of the four nonapeptides. Residues 3 to 6 are represented with a thicker line. Carbon atoms have the same colors as those defined for each peptide in panel B.

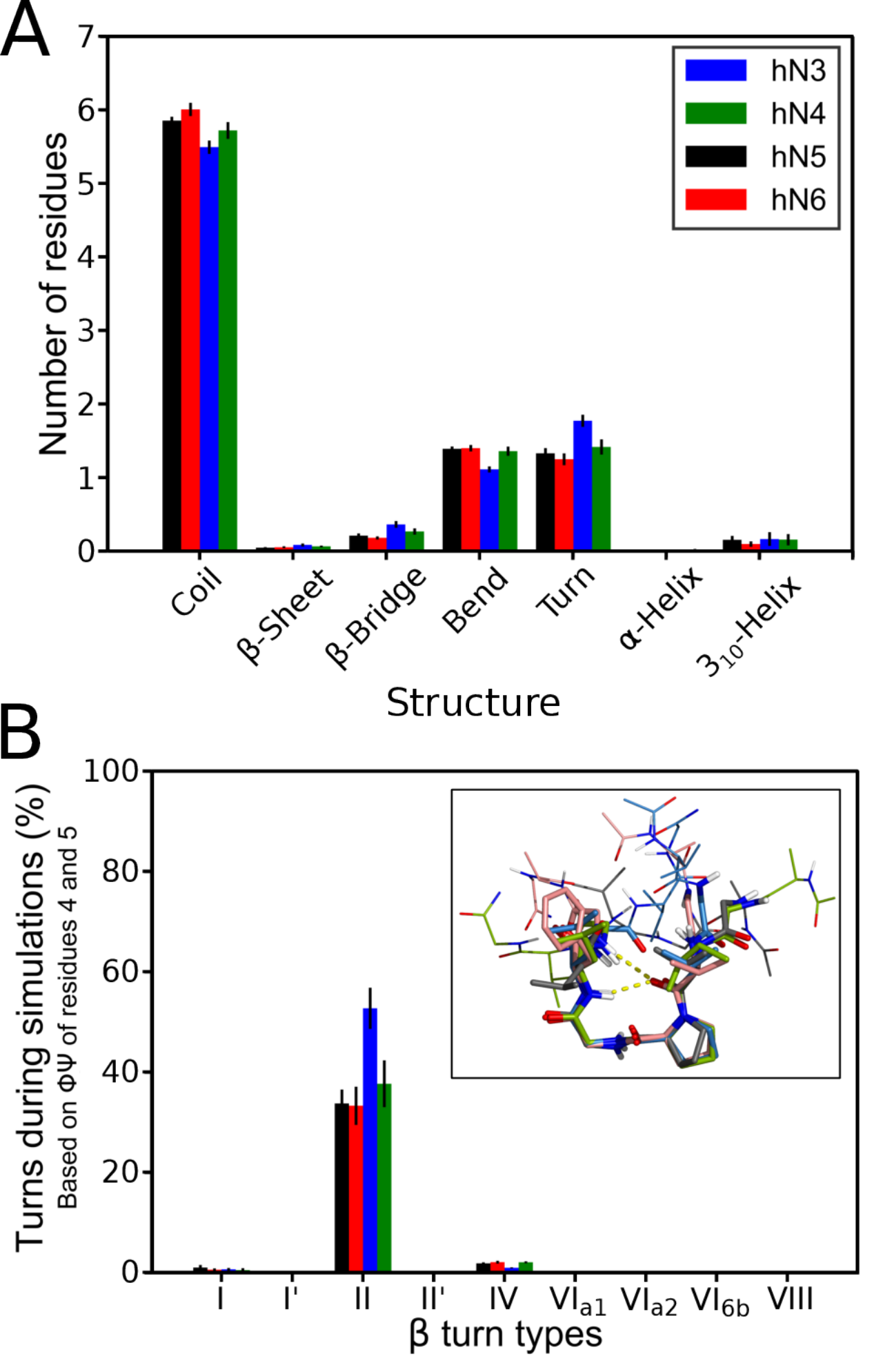

Hits O T
Bacteria
n=71504
75718 2129 297
Non-Elastin
n=9105
Vertebrates
n=9139
9627 284 14
Elastin
n=34
132 0 78
Animals
n=12754
Non-vertebrates
n=3615
4241 270 20
Eukaryota
n=22919
Fungi
n=6187
6500 102 32
Plants
n=2665
2932 116 29
Other
n=1313
1396 39 3
Archae
n=1630
1694 16 6
Virus
n=240
245 0 0


A
hN3
B
hN4
C
hN5
D
hN6
[Φ]/deg cm2 dmol-1
Wavelength/nm
203
223
197
203
226
197
203
226
196
205
230
198
100 μM/H2O
200 μM/H2O
100 μM/MeOH
200 μM/H2O +MeOH


A
B
Raman intensity
Wavenumber/cm-1
1600
1400
1200
1000
800
600
Amide I
Amide III
hN4
1468
1452
1421
1394
1345
1313
1175
1127
1034
1014
962
938
891
860
840
735
hN3
1465
1451
1421
1391
1343
1315
1161
1128
1031
1011
960
937
891
857
845
736
hN5
1466
1452
1420
1395
1343
1317
1161
1126
1031
1014
960
937
891
856
843
737
hN6
1605
1586
1454
1421
1392
1341
1320
1207
1183
1159
1126
1101
1031
1004
937
851
758
621
F1
F2
F3
F4
F5
F6
1606
1586
1207
1032
1446
1413
1360
1327
1263
1184
1160
1068
922
956
817
771
751
622


A
hN3
1692 (R)
1680 (T)
1665 (S)
1651 (T)
1720
1700
1680
1660
1640
B
hN4
1690 (R)
1678 (T)
1664 (S)
1648 (T)
1720
1700
1680
1660
1640
Wavelength/cm-1
C
hN5
1692 (R)
1678 (T)
1662 (S)
1647 (T)
1720
1700
1680
1660
1640
D
hN6
1692 (R)
1678 (T)
1661 (S)
1644 (T)
1720
1700
1680
1660
1640
Wavelength/cm-1


A
1255
1240
hN5
1300
1280
1260
1240
1220
B
1255
1240
25°C
37°C
80°C
hN6
1300
1280
1260
1240
1220
1200
Wavelength/cm-1


A
Number of residues
0
1
2
3
4
5
6
7
hN3
hN4
hN5
hN6
Coil
β-Sheet
β-Bridge
Bend
Turn
α-Helix
3₁₀-Helix
Structure
B
Turns during simulations (%)
Based on ΦΨ of residues 4 and 5
0
20
40
60
80
100
I
I'
II
II'
IV
VIa1
VIa2
VI6b
VIII
β turn types