Supporting Information

Chiral Inversion of Amino Acids in Anti-parallel β-sheets at Interfaces Probed by Vibrational Sum Frequency Generation Spectroscopy

Ethan A. Perets, a,† Pablo E. Videla, a,b,† Elsa C. Y. Yan, a,* Victor S. Batista, a,b,*

a Department of Chemistry, Yale University, 225 Prospect St., New Haven, CT 06520
b Energy Sciences Institute, Yale University, 810 West Campus Drive, West Haven, CT 06516

*To whom correspondence may be addressed.

E-mail: victor.batista@yale.edu, elsa.yan@yale.edu

† E.A.P and P.E.V. made equal contributions.
1. ATR-FTIR of (L-), (D-), and (L-Leu)/(D-Lys) LK7β

Methods. The LK7β samples were prepared on glass in the same manner as for VSFG experiments (see Methods in main text). ATR-FTIR data were measured with a Nicolet FTIR with diamond ATR SmartOrbit accessory (ThermoElectron Corporation). The acquisition time was 10 minutes using a resolution of 1 cm⁻¹.

Analysis. Figure S1 shows the ATR-FTIR spectra for anti-parallel β-sheet peptide LK7β composed of (L-) or (D-) amino acids only, and LK7β composed of alternating (L-Leu) and (D-Lys) amino acids. The three LK7β species show very similar ATR-FTIR response in the C-H region (left). At least three peaks can be observed in this spectral region: 2875 cm⁻¹ (CH₃), 2935 cm⁻¹ (CH₂), and a major peak that appears at 2955 cm⁻¹ for (L-) and (D-) LK7β, but appears blue-shifted by 5 cm⁻¹ in (L-Leu)/(D-Lys) LK7β. This peak has contributions from CH₃ asymmetric stretch and could be related to structural differences of the (L-Leu)/(D-Lys) LK7β on glass compared to the (L-) and (D-) species. The spectra are normalized by the peak at 2955 cm⁻¹.

Figure S1. C-H ATR-FTIR response of (L-), (D-), and (L-Leu)/(D-Lys) LK7β prepared in H₂O and adsorbed on glass.
2. Contribution of Fermi resonances to the vibrational C-H spectral region

In order to analyze the contribution of Fermi resonance to the C-H stretch spectral region, we simulated the IR spectra of single leucine and lysine residues by including couplings between fundamental modes and combination and overtone bands of lower energy in an effective way. Fermi resonance splittings for each fundamental mode are computed by diagonalizing the coupling matrix

\[ H = \begin{pmatrix} \omega_0 & \gamma_{01} & \gamma_{02} & \cdots \\ \gamma_{01} & \omega_1 & 0 & \cdots \\ \gamma_{02} & 0 & \omega_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \]

where \( \omega_0 \) is the frequency of the fundamental mode, \( \omega_n (n = 1,2,\ldots,N) \) are the frequencies of the \( N \) combination and overtone bands within an energy threshold of the fundamental mode, and \( \gamma_{0n} \) are the couplings associated with each resonance. The diagonalization of \( H \) in Eq. (1) leads to \((1+N)\) eigenvalues \( \tilde{\omega}_i \) that represent the frequencies of the new spectral bands and to \((1+N)\) eigenfunctions of the form

\[ |\psi_i\rangle = c_i^0 |\tilde{0}\rangle + \sum_{n=1}^{N} c_i^n |\tilde{n}\rangle, \]

where \( |\tilde{0}\rangle \equiv |000 \ldots 0\rangle \) represents the wavefunction of the fundamental mode in the harmonic product basis set \( |\nu_0\nu_1 \ldots\nu_N\rangle = |\nu_0\rangle|\nu_1\rangle \ldots|\nu_N\rangle \), and \( |\tilde{n}\rangle \) represents the wavefunction of combinations and overtone modes (i.e., \(|002 \ldots 0\rangle\)). Assuming that only the fundamental stretching carries the IR intensity, the band intensity of the Fermi resonance is borrowed from the fundamental band with relative intensity according to \( I_i/I_k = |c_i^0|^2 / |c_k^0|^2 \).

Electronic structure calculations of single gas-phase leucine and lysine residues were performed with the Gaussian 2009 software package, at the DFT level of theory. The B3LYP\(^5\) hybrid functional and the 6-31G(d) basis set\(^6\) were employed for geometry optimizations and
frequency calculations. Coupling parameters $\gamma_{on}$ were obtained from anharmonic frequency calculations based on second-order perturbation theory (VPT2), whereas frequencies of the fundamental, combination and overtone vibrational modes were computed at the harmonic level. An “ultrafine” integration grid (99 radial shells and 590 angular points per shell) was used for the frequency calculations. Only combination and overtone bands within an energy threshold of 4 cm$^{-1}$ of the fundamental mode were included in the coupling matrix $H$ (Eq. 1). A Lorentzian line shape with a width of 7.5 cm$^{-1}$ were apply to generated the IR spectra.

In Figure S2a we present the harmonic IR spectrum (black line) of a single leucine residue in the C-H stretch spectral region. The spectrum is characterized at low frequency by a broad band comprising the symmetric CH$_3$ stretching of the leucine’s isopropyl groups with a small shoulder at $\sim$2855 cm$^{-1}$ due to the C$_\gamma$-H stretching mode. The band at high frequency correspond to asymmetric stretching modes of the CH$_2$/CH$_3$ groups. After accounting for Fermi resonances, the band corresponding to the symmetric CH$_3$ stretching modes (black arrows) splits into a series of new peaked features in the spectra (red line, red arrows), some of which appears in the spectral region in between symmetric and asymmetric CH$_3$ stretching modes. An analysis of the composition of these new features (namely, $c_1^0$ of Eq. 2) shows that the new low frequency peaks can be mainly attributed to the fundamental transition whereas the higher frequency peaks correspond to Fermi resonances (overtone and combination bands). Hence, these results support the presence of symmetric CH$_3$ Fermi resonances for Leucine residues in the middle C-H stretching spectral region.

Similar results are obtained for the lysine residue (Figure S2b). The harmonic IR spectrum for this amino acid (black line) comprises a low frequency band that can be assigned to the symmetric CH$_2$ stretching and a principal peak around 2900 cm$^{-1}$ characterized by asymmetric CH$_2$
stretching vibrational modes. A small peak at ~2950 cm\(^{-1}\) due to the CH\(_2\) motif of the lysine’s sidechain next to the NH\(_3^+\) terminal is also visible. The inclusion of Fermi resonance in the spectra (red line) produces predominantly a splitting of the symmetric CH\(_2\) stretching band and gives rise to new peaks buried under the asymmetric CH\(_2\) stretching band. Note that the splitting of the CH\(_2\) band in Lysine is smaller than the splitting of the CH\(_3\) band in leucine, suggesting that the Fermi resonance CH\(_3\) mode should appear at higher frequencies than the corresponding Fermi resonance CH\(_2\) mode. The overall conclusion of this analysis supports the presence of Fermi resonances in the ~2885-2930 cm\(^{-1}\) C-H stretch spectral region that arise from the splitting of the symmetric CH\(_3\) and CH\(_2\) stretching of leucine and lysine sidechains respectively.
Figure S2. IR spectra of (a) leucine and (b) lysine residues in the C-H stretching region with (red line) and without (black line) accounting for Fermi resonances. Black arrows indicate the (harmonic) transitions that split due to the Fermi resonance; red arrows show the new position of the newly coupled transitions.
3. IR and normal modes analysis of LK$_3$β model

In Figure S3 we present the DFT-based IR spectrum of the model peptide LK$_3$β composed of two (L-) LK$_3$ peptides in an antiparallel β-sheet conformation. The spectrum is characterized by a complex lineshape and comprises four broad peaks involving symmetric and asymmetric C-H stretching. A normal mode analysis revealed that most of the modes are delocalized, making it difficult to identify and assign each peak to local vibrational moieties (representative normal mode displacements are presented in Figure S4). Nevertheless, the spectrum can be divided in four regions according to the type of vibrations involved. The shoulder at low frequency (shaded purple region) can be assigned to the symmetric CH$_2$ stretching of the lysine sidechain as well as stretching of the C$_\gamma$-H moiety of leucine. The strong peak observed at ~2870 cm$^{-1}$ (shaded red region) is due to the symmetric CH$_3$ stretching of the leucine isopropyl groups. The region around 2900 cm$^{-1}$ (shaded green region) is characterized by asymmetric CH$_2$ stretching of the lysine and leucine sidechains. The higher frequencies (shaded cyan region) correspond to asymmetric stretch of the CH$_3$ groups in leucine. We also found that the modes involving the displacement of the C$_\alpha$-H motif of the lysine residue (whose frequencies appear in the region marked with the red circle in Figure S3) are highly coupled to the asymmetric CH$_2$ stretching (Figure S4). The C$_\alpha$-H of the leucine residues (blue circle) can be characterized as local modes. Furthermore, the N-terminal acetyl capping groups may introduce additional resonances corresponding to CH$_3$ symmetric (~2893 cm$^{-1}$, black circle) and asymmetric stretch modes (~2960 cm$^{-1}$ and ~2970 cm$^{-1}$, green circles).
Figure S3. DFT-based IR spectra for (L)-LK₃β model. Shaded areas correspond to symmetric CH₂ (purple), symmetric CH₃ (red), asymmetric CH₂ (green) and asymmetric CH₃ (cyan) stretching modes.
Figure S4. Representative normal mode displacements of the LK$_3$β model system.
4. Additional characterization of backbone conformation

In Figures S5 and S6 we present Ramachandran-like free energy profiles of backbone dihedral angles $\phi$ and $\psi$ for individual residues along the LK$_7$β peptide, obtained from molecular dynamics simulations of (L-), (D-), and (L-Leu)/(D-Lys) LK$_7$β at the air-water interface. In the case of multiple strands, the results correspond to averages over identical pair of residues.

The results for the (L-) LK$_7$β composed of two strands are presented in the left panels of Figure S5. The residues along the LK$_7$β peptide adopt dihedral angles consistent predominantly with $\beta$-sheet conformations (-160°$<\phi<$-80°, 100°$<\psi<$150°). Note that the residues corresponding to the N- and C-terminal of the peptide are somehow more flexible and expand a greater free-energy landscape than the residues in the middle of the $\beta$-sheet.

Right panels of Figure S5 show results for two strands of (D-) LK$_7$β. The free energy landscape sampled by the D-amino acids mirrors those of the L-amino acids, albeit a 180° rotation of the distributions around the origin due to the inverted chirality at the C$_\alpha$ atom.

On the left panels of Figure S6 we present the free energy profile per residue for the (L-Leu)/(D-Lys) LK$_7$β system. The backbone conformations sampled by this system is much richer and complex than the ones presented above, which can be ascribed to the lack of secondary structure. Note also that the conformations sampled by each individual residue comprises several minima in the free energy landscape, suggesting a dynamical nature of the polypeptide. To disentangle the effect of chiral inversion from the absence of secondary structure in the conformational structure of the (L-Leu)/(D-Lys) LK$_7$β, we performed additional simulations of a single strand of (L-) LK$_7$β at the water-air interface. The results are presented in the right panels
of Figure S6. Several similarities and difference are worth mentioning between these two systems. Quite remarkably, the conformational distributions of the leucine residues seem to be comparable for both (L-Leu)/(D-Lys) and (L-) LK7β, although some difference persist (i.e. residue Leu3 in Figure S6). For the case of the lysine residues, however, the free energy landscape for the (L-) LK7β cannot be reproduce by a mirror image of the (L-Leu)/(D-Lys) LK7β results. Of particular significance is the $-90^\circ<\phi<-30^\circ$ and $-80^\circ<\psi<0^\circ$ prominent region spanned by (D-) lysine residues that is not sampled at all by the (L-) lysine residues. These results demonstrate that the inversion in chirality of the amino acids can not only disrupt the secondary structure of the peptide, but also modify the conformation landscape sampled.¹⁴
**Figure S5.** Free energy (units of k\(_B\)T) as a function of backbone dihedral angles \(\phi\) and \(\psi\) per residue along the LK-\(\beta\) peptide. Left: two strands of (L-) LK-\(\beta\). Right: two strands of (D-) LK-\(\beta\).
Figure S6. Free energy (k_B T units) as a function of the backbone dihedral angles φ and ψ per residue. Left: one strand of (L-Leu)/(D-Lys) LK_7β. Right: one strand of (L-) LK_7β.
5. Inversion dynamics of (L-) LK\(\beta\) on the air-water interface

In order to test the strong preference of the hydrophobic/hydrophilic interactions at the air-water interface, we performed a simulation test in which the two-strands (L-) LK\(\beta\) system, which originally has the hydrophobic leucine residues pointing into the air and the hydrophilic lysine residues into the water, is placed upside-down into the water interface. From this initial configuration the system is allowed to relax using conjugate gradient algorithm for 50 steps to eliminate any close contact interaction between atoms, and then a 450 ps thermalization run followed by a 4 ns production simulation was performed.

In Figure S7 we present the time evolution of the orientation of the sidechains of each residue in the (L-) LK\(\beta\) system. The sidechain’s orientation is characterized by the projection of the C\(\alpha\)-C\(\beta\) bond along the surface normal. At time zero, all leucine residues are pointing down into the water phase \((\cos\theta = -1)\) whereas all lysine residues are pointing up in to the air \((\cos\theta = 1)\). However, during the course of the simulation, the system naturally evolves to invert the orientation of the sidechains of each residue, resulting in the stable conformation with all the hydrophilic lysine residues solvated by the water phase \((\cos\theta = -1)\) and all the hydrophobic leucine residues pointing in to the air \((\cos\theta = 1)\). These results demonstrate the importance of hydrophobic-hydrophilic interactions at the air-water interface for modulating the dynamics and stability of secondary structures in proteins.
Figure S7. Time evolution of the orientation of the sidechains of two strands of (L-) LK₁β, characterized by the Cα-Cβ projection along the surface normal, after inversion of the β-sheet at the air-water interface (time origin).
6. Dynamics of two strands of (L-Leu)/(D-Lys) LK7β on the air-water interface

Several attempts were made to simulate two strands of (L-Leu)/(D-Lys) LK7β in an anti-parallel β-sheet conformation at the water-air interface. However, we found that the system is not stable in this conformation and the strands break apart over the course of the simulation. In Figure S8 we present two examples of the dynamical evolution of two strands of (L-Leu)/(D-Lys) LK7β by using the distance between the N-terminal of one strand and the C-terminal of the other (d_{NC}) as an order parameter. Note that small values of this parameter represent close contact between the strands whereas increase d_{NC} values indicate dissociation of the β-sheet.

Top panel of Figure S8 shows an example in which the antiparallel β-sheet breaks at one extreme first within 200 ps of starting the simulation, whereas the other extreme remains bonded (with d_{NC} ≈ 4 Å). However, after 3 ns, the second extreme also breaks and the two peptide strands separate. On the example presented on the bottom panel of Figure S8 both extremes break at the same time (around 500 ps) and the β-sheet dissociates apart. These results suggest that the (L-Leu)/(D-Lys) LK7β do not form secondary structure at the water-air interface, and are consistent with the VSFG silent chiral amide I/amide II spectrum.
Figure S8. Time evolution of the distance between the N-terminal of one strand and the C-terminal of the other in the antiparallel (L-Leu)/(D-Lys) LK7β. Dashed line marks the end of the thermalization steps and the beginning of production dynamics. As a reference, the average distance for the (L-) LK7β system is ~5. Insets correspond to snapshots of the peptide conformation at the time marked with arrows.

7. Additional characterization of angular orientation

In the top panels of Figures S9 and S10 we present the angular distribution of tilt angles (θ) for selected bonds with respect to the surface normal for leucine and lysine residues, respectively, for a system composed of two strands of (L-) LK7β (panel a) and two strands of (D-) LK7β (panel b). Results for (L-) LK7β were presented in Figures 7 and 8 and discussed in the main text, but are reproduced here for completeness. The angular distributions for (D-) LK7β are essentially the
same as the ones for (L-) LK-β for both leucine and lysine residues, and are characterized by an intense preference of the Ca-Cβ bonds of leucine (lysine) sidechains to point towards the air (water), as well as a strong ordering at the interface of the isopropyl group of leucine residues.\textsuperscript{12, 15-20} Note also that the CH\textsubscript{2} groups of the lysine sidechain are ordered. Since the achiral VSFG spectra is sensible to the orientation of sidechains with respect to the surface normal, the results presented here supports the indistinguishability between (L-) LK-β and (D-) LK-β, consistent with the experimental results presented in Figures 1b and 2b.

In the bottom panels of Figures S9 and S10 we present results for a system composed of a single strand of (L-Leu)/(D-Lys) LK-β (panel c) and (L-) LK-β (panel d). Interestingly, the angular distributions of the leucine residues for both single-strand’s systems (Figure S9c,d) look remarkably similar, suggesting that the ordering of the (L-) leucine residue is not affected by the inversion of chirality in the (L-Leu)/(D-Lys) LK-β system. In comparison with the leucine two-strands systems (Figure S9a), the one-strand leucine’s distributions become slightly flatter and wider, consistent with the more flexible nature of the polypeptide strand in the absence of antiparallel β-sheet secondary structure (Figure S6). Note that the distribution of the Ca-Cβ bonds (blue lines) is still peaked at θ = 0°, demonstrating the strong preference of leucine residues to point into the air. The overall conclusion that emerge from this analysis is that the angular orientation of leucine sidechains remains roughly intact over the different systems studied, albeit with wider distributions for the single-strand peptides.
The picture drastically changes when analyzing the lysine residues. In Figure S10c we present the angular distributions for the lysine residue of (L-Leu)/(D-Lys) LK7β. With the exception of the Cα-Cβ bonds (blue line), which still present a strong preference in pointing towards the water, the bond distributions of the CH2 groups in the lysine sidechain become nearly isotropic. Remarkably, the absence of orientationally preference of the (D-) lysine sidechain is not just a consequence of the lack of secondary structure but also of the intrinsic complex dynamics of the (L-Leu)/(D-Lys) LK7β system. In fact, Figure S10d shows the angular distributions for a system comprising one strand of (L-) LK7β. Note that for this system, the distributions of CβH2 and CγH2 groups, although wider, retain some ordering similar to the one found in Figure S10a, suggesting that the orientation of (L-) lysine sidechains is roughly independent of the formation of secondary structure. Contrary, the results presented here suggest that the orientation of (D-) lysine sidechains significantly differ from the ones obtained for (L-) lysine amino acids in LK7β polypeptides.
Figure S9. Angular distributions of selected bond vectors of leucine residue with respect to the surface normal for (a) two-strands of (L-) LKβ, (b) two-strands of (D-) LKβ, (c) one-strand of (L-Leu)/(D-Lys) LKβ and (d) one-strand of (L-) LKβ. Dashed lines correspond to an isotropic distribution of the angles. Shaded areas correspond to one standard deviation. Inset: Schematic representation of leucine residue with definition of bond vectors used to describe the orientation with respect to the surface normal.
Figure S10. Angular distributions of selected bond vectors of lysine residue with respect to the surface normal for (a) two-strands of (L-) LKβ, (b) two-strands of (D-) LKβ, (c) one-strand of (L-Leu)/(D-Lys) LKβ and (d) one-strand of (L-) LKβ. Dashed lines correspond to an isotropic distribution of the angles. Shaded areas correspond to one standard deviation. Inset: Schematic representation of lysine residue with definition of bond vectors used to describe the orientation with respect to the surface normal.
Figure S11. Top: Schematic representation of leucine and lysine residues with definition of bond vectors used to describe the orientation with respect to the surface normal. Bottom: Angular distributions of selected bond vectors of leucine and lysine residues with respect to the surface normal for (a) two-strands of (L-) LK7β, (b) two-strands of (D-) LK7β, (c) one-strand of (L-Leu)/(D-Lys) LK7β and (d) one-strand of (L-) LK7β. Dashed lines correspond to an isotropic distribution of the angles.
**Table S1.** Average values and standard deviation of angular orientation of leucine and lysine residues in LKβ.¹

<table>
<thead>
<tr>
<th>residue</th>
<th>angle</th>
<th>all-L two-strands</th>
<th>all-D two-strands</th>
<th>all-L one-strand</th>
<th>mix-L/D one-strand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\bar{\theta}$</td>
<td>$\sigma_\theta$</td>
<td>$\bar{\theta}$</td>
<td>$\sigma_\theta$</td>
</tr>
<tr>
<td>LEU</td>
<td>$C_\gamma$-H</td>
<td>114</td>
<td>22</td>
<td>114</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>$C_\alpha$-Cβ</td>
<td>40</td>
<td>19</td>
<td>39</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>$C_\alpha$-H</td>
<td>88</td>
<td>23</td>
<td>87</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>$C_\beta$-H₂</td>
<td>76</td>
<td>20</td>
<td>78</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>C=O</td>
<td>101</td>
<td>25</td>
<td>98</td>
<td>23</td>
</tr>
<tr>
<td>LYS</td>
<td>$C_\beta$-H₂</td>
<td>111</td>
<td>20</td>
<td>113</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>$C_\gamma$-H₂</td>
<td>94</td>
<td>29</td>
<td>98</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>$C_\alpha$-H₂</td>
<td>88</td>
<td>30</td>
<td>84</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>$C_\beta$-H₂</td>
<td>99</td>
<td>33</td>
<td>100</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>$C_\alpha$-Cβ</td>
<td>147</td>
<td>23</td>
<td>151</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>$C_\alpha$-H</td>
<td>92</td>
<td>26</td>
<td>91</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>C=O</td>
<td>88</td>
<td>22</td>
<td>88</td>
<td>19</td>
</tr>
</tbody>
</table>

¹ All values are in degrees. For a definition of the angles, see Figures 8 and 9 in the main text and Figure S9. As a reference, for an isotropic distribution $\bar{\theta} = 90^\circ$ and $\sigma_\theta \approx 39^\circ$. 
Table S2. Relevant Fresnel factors for the ssp and psp polarization.\(^{(a)}\)

<table>
<thead>
<tr>
<th>Fresnel factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L(_{yyz})</td>
<td>0.25</td>
</tr>
<tr>
<td>L(_{xxz})</td>
<td>0.20</td>
</tr>
<tr>
<td>L(_{zxy})</td>
<td>0.17</td>
</tr>
</tbody>
</table>

\(^{(a)}\) The refractive index of the interfacial layer is assumed to be 1.20 and the refractive index of air and water was taken to be 1 and 1.333 respectively.\(^{21-22}\)