SUPPLEMENTARY INFORMATION

Probing the fibrillation of lysozyme by nanoscale infrared spectroscopy

Zeyaul Islam[1], Mohamed H. Ali[1], Anton Popelka[2], Raghvendra Mall[3], Ehsan Ullah[3], Janarthanan Ponraj[4], Prasanna R. Kolatkar[1*]

[1]Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University (HBKU), Qatar Foundation, Doha, Qatar.

[2]Center for Advanced Materials (CAM), Qatar University, Doha, Qatar.

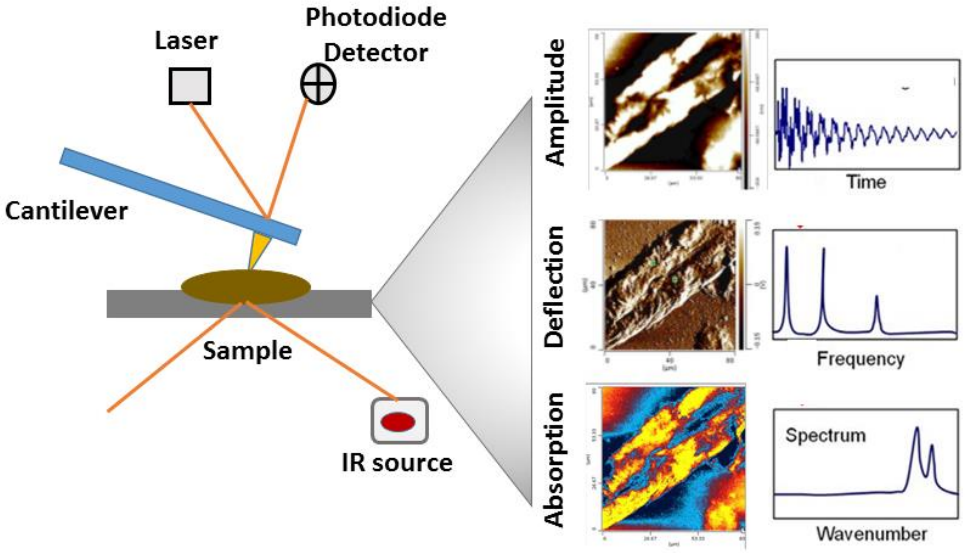
[3]Qatar Computing Research Institute (QCRI), Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.

[4]Qatar Environment and Energy Research Institute (QEERI), Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.

*To whom correspondence should be addressed: Prasanna R. Kolatkar, Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University (HBKU), Qatar Foundation, Doha, Qatar. Tel: +974 445 45889, Fax: +974 445 41770, Email: pkolatkar@hbku.edu.qa

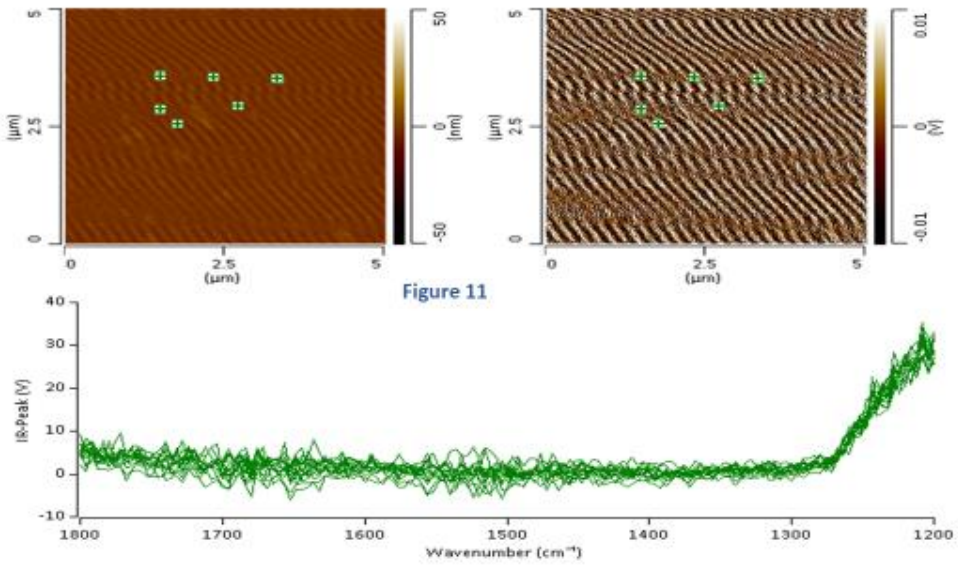
Keywords: Protein aggregation, amyloid fibrils, neurodegenerative diseases, nanoIR, lysozyme

Supplementary Figure 1.



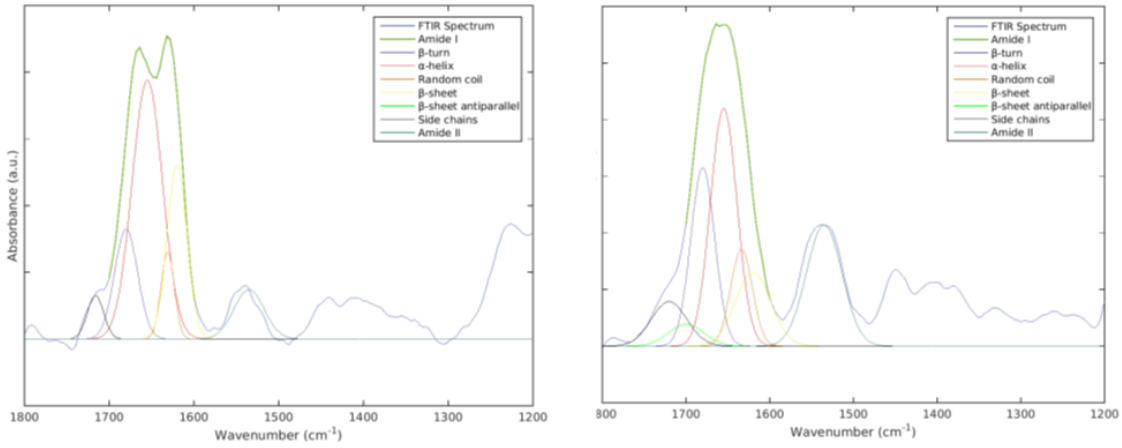
Schematic diagram of a typical nanoIR instrumentation and working pathway. Tunable high speed IR laser light is focussed onto the sample located at the cantilever tip. Thermal expansion occurs when the absorption band matches the wavelength. The cantilever oscillation amplitude as a function of wavelength creates a unique spectral fingerprint. (Adopted from nanoIR webpage https://www.epfl.ch/labs/lpmv/facilities/nanoir/).

Supplementary Figure 2.



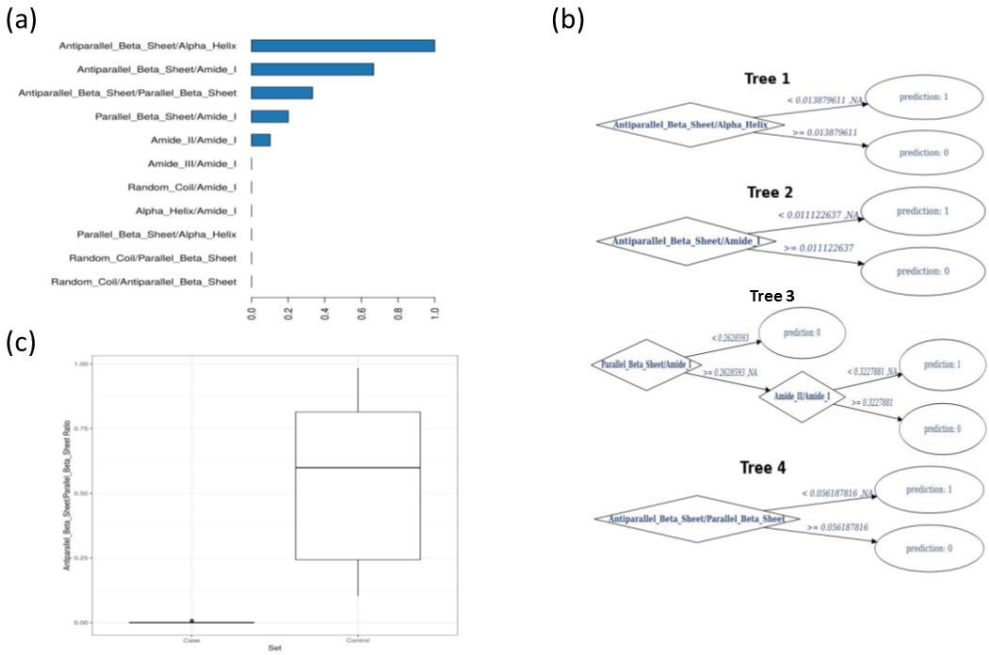
IR spectra of cover slip to use as a background. There is no distinct peak in case of cover slip; all the protein spectra was subtracted from background spectra.

Supplementary Figure 3.



Deconvolution curve showing amide I components assignments acquired from the lysozyme sample incubated for day 21 and 1.

Supplementary Figure 4.



Machine learning Model (Random forest). Here the x-axis represents the scaled feature importance or how important was that ratio in distinguishing the cases from controls. We

can observe that the ratio of antiparallel β- sheet to α-helix. Amide I and parallel β-sheet are the most important ones. (b) The decision trees forming the optimal RF model. (c) Boxplot displaying the difference in the ratio of antiparallel β-sheet to parallel β-sheet in case vs control samples.

Supplementary Table 1. Secondary structure components with their assigned wavelength in IR region.

Protein Components	Wavelength (cm^{-1}) Assignment
Amide I	**1700 – 1600**
Amide II	**1555 - 1535**
Amide III	**1350 - 1200**
Amide I Secondary structure (1700 – 1600)	
α-helix	**1660 - 1650**
Random coil	**1645 - 1630**
Parallel β-sheet	**1610 - 1635**
Beta-shoulder (1665 – 1695)	
β-turn	**1680**
Anti-parallel β-sheet	**1695**

Supplementary Table 2. Secondary structure compositions (%) of lysozyme during fibrillation as elucidated from CD spectral data (deconvulated by CONTIN analysis program on dichroweb).

Days	**α-helices**	**β-sheet**	**Turn**	**Random coil**
0	20.6	21.1	14.3	44.0
1	12.7	23.1	14.1	50.1
3	6.2	23.3	15.4	55.1
6	4.7	28.9	22.3	44.1
10	4.6	27.6	22.3	45.5
21	4.0	34.4	15.7	45.9