Submitted to Anal Chem

ATR-FTIR spectroscopy supported by multivariate analysis for the characterization of adipose tissue aspirates from patients affected by systemic amyloidosis.

Diletta Ami[‡$], Paolo Mereghetti[‡$], Andrea Foli[§], Masayoshi Tasaki[§#], Paolo Milani[§], Mario Nuvolone[§], Giovanni Palladini[§], Giampaolo Merlini[§], Francesca Lavatelli[§*], Antonino Natalello[$*]

$Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy.

§Amyloidosis Research and Treatment Center, Fondazione IRCCS Policlinico San Matteo and Department of Molecular Medicine, University of Pavia, Viale Golgi 19, 27100, Pavia, Italy.

Department of Morphological and Physiological Sciences, Graduate School of Health Sciences, Kumamoto University, 4-24-1 Kuhonji, Kumamoto 862-0976, Japan, and Department of Neurology, Graduate School of Medical Sciences, Kumamoto University, 1-1-1 Honjo, Kumamoto 860-0811, Japan.

ABSTRACT: Deposition of misfolded proteins as extracellular amyloid aggregates is the pathological hallmark of systemic amyloidoses. Subcutaneous fat acquired by fine needle aspiration is the preferred screening tissue in suspected patients. In this study we employed Fourier transform infrared (FTIR) spectroscopy in attenuated total reflection (ATR) to investigate human abdominal fat aspirates with the aim to detect disease-related changes in the molecular structure and composition of the tissue and to exploit the potentiality of the method to discriminate between amyloid-positive and negative samples. The absorption and second derivative spectra of Congo Red (CR) positive and CR-negative specimens were analyzed by three multivariate methods in four spectral regions. The proposed ATR-FTIR method is label-free, rapid, relatively inexpensive and requires minimal sample preparation. We found that the ATR-FTIR approach can differentiate fat aspirates containing amyloid deposits from control specimens with high sensitivity and specificity, both of 100 [89-100]%. Noteworthy, the wavenumbers most important for discrimination indicate that changes both in the protein conformation and in resident lipids are intrinsic features of affected subcutaneous fat in comparison with the CR-negative controls. In this proof of concept study, we show that this approach could be useful for assessing tissue amyloid aggregates and for acquiring novel knowledge on the molecular bases of the disease.

Amyloidoses are conformational diseases characterized by deposition of misfolded proteins as extracellular amyloid fibrils in target tissues[1,2]. Aggregation is related to loss of protein native conformation and acquisition of an alternative cross β-sheet-rich structure. This phenomenon subtends several human diseases with enormous overall social and economic impact. At least 36 distinct proteins are known causative agents of amyloid diseases[3]; these range from central nervous system forms such as neurodegenerative diseases, to systemic forms, in which amyloid deposits are widespread and can affect multiple sites. Major examples of these latter include light chain (AL), transthyretin (ATTR) and reactive (AA) amyloidosis. Fibril deposition leads to subversion of the tissue architecture and dysfunction of the affected organs. Demonstrating the presence of tissue amyloid deposits is required for the diagnosis of systemic amyloidoses. Amyloid-specific dyes, such as Congo red (CR), which display specific tinctorial properties when interacting with the ordered structure of the fibrils, are the diagnostic mainstay for this task. As a less invasive procedure compared to target organ biopsy, subcutaneous fat acquired by fine needle aspiration is the preferred screening site in suspected patients[4]. This tissue, in fact, is frequently affected and its acquisition is minimally invasive. In these progressive diseases, timely diagnosis is crucial, since an array of effective therapies are now available to halt organ function deterioration in several of these forms[5].

Fourier transform infrared (FTIR) spectroscopy is an emerging tool of potential clinical usefulness in several pathological contexts. This label free approach can be applied to intact cells, tissues and biofluids, providing a "spectroscopic fingerprint", which represents a snapshot of biomolecular composition and structure of the investigated sample[6–9]. The clinical potential of FTIR in the field of protein misfolding diseases is significant. The information provided by FTIR, in fact, is unique and complementary with respect to all other tools for assessing fibrillar aggregates in tissues, including microscopy, amyloid-specific dyes, and *in vivo* imaging (e.g. scintigraphy). The distinctive value of FTIR stands in its ability to provide label-free information on the structural properties of proteins *in situ*, including the presence of intermolecular β-sheets. After the pioneering work of Choo et al.[10] reporting the *in situ* structural characterization of β-amyloids in a section of Alzheimer's disease (AD) brain by infrared microspectroscopy, a number of FTIR investigations have provided promising evidence on the potentiality of this approach for the diagnosis of dementia using accessible

specimens such as biological fluids (recently reviewed in[7]). We have recently characterized *in situ* by FTIR microspectroscopy the infrared (IR) response of human biopsy tissues affected by AL amyloidosis[11]. In particular, the intermolecular β-sheet marker band was detected in cardiac and adipose tissue sections. Moreover, in cardiac tissues the protein deposits were found to be enriched with cholesterol, glycosaminoglycan and collagen[11]. However, FTIR microspectroscopy was not suitable to study adipose tissue aspirates, which cannot be cut into sections of proper thickness required for transmission measurements.

In this study we employed FTIR spectroscopy in attenuated total reflection (ATR), supported by multivariate analyses, to investigate human abdominal fat aspirates with the aim to discriminate samples containing amyloid fibrils from amyloid-negative ones. The identification of the wavenumbers responsible for the segregation between affected and unaffected individuals allowed us to characterize *in situ* the main structural changes associated with the presence of amyloidosis.

EXPERIMENTAL SECTION

Patients and samples. Subcutaneous abdominal fat samples were obtained by fine needle aspiration during the routine diagnostic procedures at Centro per lo Studio e la Cura delle Amiloidosi Sistemiche of Fondazione IRCCS Policlinico San Matteo in Pavia, Italy. Consecutive individuals evaluated between February and April 2015 were included in the study. All individuals gave written informed consent for their biological samples to be stored and used for research purposes and the study was approved by the Ethical Committee of Policlinico San Matteo. The presence of amyloid fibrils was evaluated by CR staining (Figure S-1a) and the amyloid load was graded semiquantitatively with a CR score from 0+ (negative) to 4+[12]. Amyloid typing was confirmed on adipose tissue in CR positive cases by immunoelectron microscopy (IEM)[4] (Figure S-1b). Clinical characterization was performed as previously described[4,13]. Samples indicated herein as controls are CR-negative and derive from individuals who had no biochemical, genetic, instrumental and clinical signs consistent with the presence of systemic amyloidosis. None of the controls developed signs consistent with systemic amyloidosis over a median follow up of 30 months. All specimens were stored unfixed and frozen at -80 °C until use. Sample processing prior to FTIR analysis was minimal; fat aspirates were thawed on ice and washed 3 times with 1 ml of ice-cold isotonic saline (0.9% NaCl) to completely remove blood. Each washing was followed by centrifugation (1 min, 13200 rpm, 4 °C). Only samples without visible blood after washing were retained for analysis.

FTIR measurements. The subcutaneous fat aspirates were placed onto the crystal surface of the single reflection ATR device (Quest, Specac) for FTIR measurements. The sample was forced into close contact with the diamond ATR crystal using the clamp arm assembly of the device (Figure 1). ATR-FTIR spectra were collected by the Varian 670-IR spectrometer (Varian Australia Pty Ltd.) under the following conditions: 2 cm^{-1} resolution, scan speed of 25 kHz, 512 scan coadditions, triangular apodization, and a nitrogen-cooled Mercury Cadmium Telluride detector. The spectrometer was continuously purged with dry air[14,15]. To take into account the sample macro-heterogeneity, up to 8 spectra were collected from different areas of the same abdominal fat sample. Spectra with very high lipid absorption were not considered for the subsequent analyses. In particular, the proteins to lipids ratio was evaluated from the area of the protein Amide I band (1700-1600 cm^{-1}, mainly due to the C=O stretching vibrations of the peptide bond) and from that of the lipid C=O band (area between 1770-1715 cm^{-1}). Only spectra with Amide I/C=O ratios above 2.7 were considered. The overall features of acceptable absorption spectra are reported in Figure 1 and Figure S-2. Measured absorption spectra were offset to zero at 1800 cm^{-1} and normalized at the same Amide I band area to compensate for differences in the protein content. Absorption spectra were smoothed using the Savitsky-Golay method (25 points) before the second derivative calculation that was obtained without additional smoothing points. Spectral analyses were performed with the Resolutions-Pro software (Varian Australia Pty Ltd.) We should note that the measured absorption spectra (Figure S-2) were characterized by a flat baseline without evidences of scattering artifacts (such as the typical slopping baseline of the scattering-affected spectra[16]). This result is in agreement with the common observation that ATR spectra are not influenced by resonant Mie scattering to the same extent compared to transmission spectra[16].

Multivariate analysis. Multivariate analysis has been performed using R version 3.4.3. Raw spectra have been checked for outliers using an algorithm based on principal component analysis which is especially suited for high-dimensional datasets where the number of variables is much larger than the number of observations, as described in[17] and implemented in the R package mvoutlier version 2.0.5. No outlier were found, so all spectra were retained for further analysis.

Spectra have been split into four spectral regions and on each region three different multivariate data analysis methods have been tested, namely extreme gradient boosting (xgbTree), multivariate adaptive regression splines (MARS) and partial least square discriminant analysis (PLS-DA). The xgbTree belongs to the family of the boosting methods, i.e. classification methods that produce classification models from an ensemble of weak classifiers, classification trees in this case. The extreme adjective refers to the algorithmic tricks implemented in order to increase the computational performances[18,19]. The MARS method is an extension of linear regression that models nonlinearities and interactions among variables[20]. PLS-DA is a widely used multidimensional regression method, which is a variant of the classical partial least square method when the dependent variable is categorical[21].

In order to assess the predictive discrimination and avoid over-fitting, for each method a 10-time repeated 5-fold cross-validation was applied. So for each method 50 models were trained. Since each individual has multiple spectra, folds have been created at the individual level, ensuring that all spectra for a given individual are either in the training or in the test set. More specifically, having N individual each with m_N spectra, on every round of cross-validation, the individuals have been partitioned into 5 folds. Four folds (containing N*4/5 individuals) have been used to train the model and the remaining fold (containing N*1/5 individuals) was used to test the model. Folds are complementary (i.e. no repeated individuals in different folds) and the individuals are randomly chosen. The training of the model is repeated five times, each time varying the test partition. The 5-fold cross-validation is then repeated 10-times in order to lower the risk of partition-dependent artifacts. The

best model has been selected using the "one standard error rule". In this case, the model with the best performance value is identified and, using resampling, we can estimate the standard error of performance. The final model used was the simplest model within one standard error of the (empirically) best model[22]. As performance measure the area under the curve (AUC) was used (Figure S-3). The AUC is computed from the receiver operating characteristic curve (ROC curve) that is created by plotting the true positive rate (i.e. sensitivity) against the false positive rate (i.e. specificity) at various threshold (value that discriminate between positive and negative outcome) settings. When using normalized units, the AUC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative')[23]. For all trained models the sensitivity and specificity have been also computed. The sensitivity is computed as true positive / (true positive + false negative), while the specificity as true negative / (true negative + false positive). Variable importance has been obtained differently for each method. For xgbTree method the variable importance expresses the gain contribution of each feature to the model. Each gain of each feature of each tree is taken into account, then average gain per feature obtained[22]. For MARS method the variable importance tracks the reduction in the generalized cross-validation statistic as terms are added[24]. For the PLS-DA method the variable importance measure here is based on weighted sums of the absolute regression coefficients[25].

RESULTS AND DISCUSSION

FTIR characterization of abdominal fat aspirate. Samples from 10 patients affected by systemic light chain (AL) amyloidosis (8 AL λ, 2 AL κ) and from 1 patient affected by reactive (AA) amyloidosis were included in the study. Patients' samples presented different degrees of amyloid infiltration. The control population included 13 unaffected individuals (Figure 1, Figure S-1). The main clinical features of patients' and controls' samples are indicated in Table S1. Sample weight ranged between 10 and 30 mg.

The ATR-FTIR spectrum of the abdominal fat aspirate from a CR negative sample is reported in Figure 1 (right spectrum). The spectrum provides information on the content and structures of the interrogated biological specimen[26]. In particular, the main peaks assigned to the sample biomolecules are indicated in Figure 1. The following main proteins bands are observed: Amide A (NH stretching vibration around 3295 cm^{-1}), Amide I (mainly due to the C=O stretching vibration between 1700-1600 cm^{-1}) and Amide II (mainly due to the NH bending and the CN stretching vibrations between 1580-1516 cm^{-1})[27]. The absorption of the lipid hydrocarbon chains (CHx) occurs in the 3050-2800 cm^{-1} spectral region, while the absorption of the lipid C=O around 1744 cm^{-1}. In the 1500-1200 cm^{-1} region the absorption of the lipid heads and hydrocarbon chains[28], as well as of proteins and amino acid side chains takes place[27,29]. Moreover, IR peaks due to nucleic acid phosphate groups and to sugar moieties are expected in the 1400-1000 cm^{-1} spectral regions[26].

With the aim to detect marker bands of amyloids and to exploit the potentiality of ATR-FTIR spectroscopy to discriminate between amyloid-positive and negative individuals, the spectra from CR-positive and CR-negative samples were measured (Figure 1). Noteworthy, this spectroscopic method does not require any sample preparation or labeling. Indeed, the fat aspirates were only washed with isotonic saline to remove residual blood, which was found to interfere with the subsequent spectral analyses (data not shown). In this work, the ATR-FTIR approach was selected over other sampling methods because it is practically infeasible to cut adipose tissue aspirates into sections of proper thickness required for transmission measurements by IR microspectroscopy. On the other hand, the ATR-FTIR spectroscopy does not allow to study sample micro-heterogeneity as instead previously reported for cardiac and adipose tissue sections of AL positive patients[11].

As shown in Figure 1, the ATR-FTIR absorption spectra from the amyloid-positive patients appeared to overlap to a great extent with those collected from controls. Therefore, to disclose and validate possible spectral differences, complex analyses are typically required, such as the second derivative calculation and the multivariate analyses[6,7,26,30–32], which will be reported in the following paragraphs.

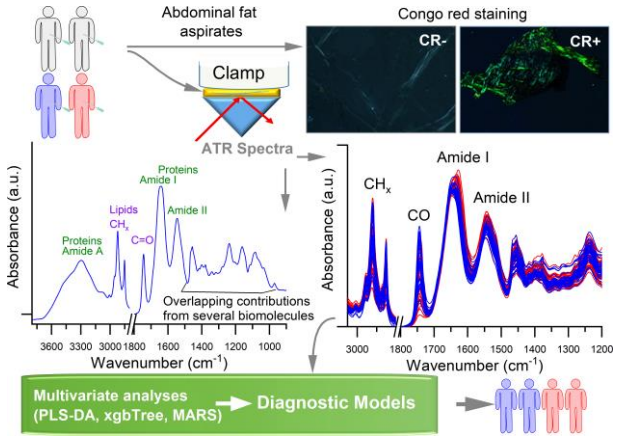


Figure 1. Overview of the experimental workflow. Abdominal fat aspirates from 24 individuals (11 CR-positive amyloid patients and 13 CR-negative controls) were deposed onto the ATR diamond crystal for FTIR measurements. Representative images of CR positive and negative samples, viewed using polarized light microscopy, are shown in the upper right panels. A representative ATR-FTIR absorption spectrum of the fat aspirate from an amyloid-negative control and the 78 absorption spectra from the 24 individuals (amyloid-positive as red spectra and CR-negative controls as blue spectra) are reported in the left and right panels, respectively. The assignment of selected bands to the main biomolecules is shown[26]. The absorption or second derivative spectra from well characterized CR-positive (red) and CR-negative individuals (blue) were subjected to multivariate analyses to evaluate the discrimination potential of the method. The overall performance, sensitivity, and specificity of the multivariate models were estimated using 10-time repeated 5-fold cross-validation as described in the experimental section.

Multivariate analyses of the ATR-FTIR absorption spectra. To explore the ability of ATR-FTIR spectroscopy to provide molecular information that could allow discriminating between amyloid-positive patients and negative individuals, a multivariate analysis of the spectral data was performed. In particular, we evaluated the performance (AUC value), the sensitivity, and the specificity of three different approaches on four spectral regions, namely between 3050-2800 cm^{-1} (mainly due to lipid hydrocarbon chains), 1700-1500 cm^{-1} (protein Amide I

and Amide II bands), 1500-1200 cm^{-1} (overlapping contributions from lipids, proteins, nucleic acids and other biomolecules, such as glycosaminoglycans)[11,26,33] and 1700-1200 cm^{-1} spectral ranges. The predictive performance was estimated by applying a 10-time repeated 5-fold cross-validation. During each repetition the training and the test set have been created splitting the spectra at the individual level, i.e. ensuring that all spectra for a given patient are either in the training or in the test set (see the experimental section for details).

Three different classification methods have been tested. PLS-DA, extreme gradient boosting (xgbTree) and MARS. The PLS-DA was chosen because it is a well-known and robust method. The xbgTree is a recent improvement of the boosting method which turns out to be a really powerful method in many different scenarios[34]. The latter, MARS, was selected because it is a flexible, robust, and fast training method which generally gives good classification performances in large datasets[20]. Another reason we chosen these methods is because they are drastically different among each others. Hence, obtaining similar results will be an indication of the confidence of the outcome of the analysis.

The overall performance of the results obtained by the three multivariate approaches is summarized in Figure 2. The best classification ratios were obtained by PLS-DA that gives an average AUC (expressed as median [first, third quartile] over the repeated k-fold cross-validation) of 93 [91-95]%, specificity of 81 [73-86]% and sensitivity of 97 [90-100]% when performed in the 1700-1500 cm^{-1} spectral region, and AUC of 94 [92-96]%, specificity of 88 [85-93]% and sensitivity of 94 [92-100]% for the 1700-1200 cm^{-1} spectral regions, respectively. Worse performances were obtained in the other tested spectral regions and using different chemometrics approaches (Figure 2 and Table S2).

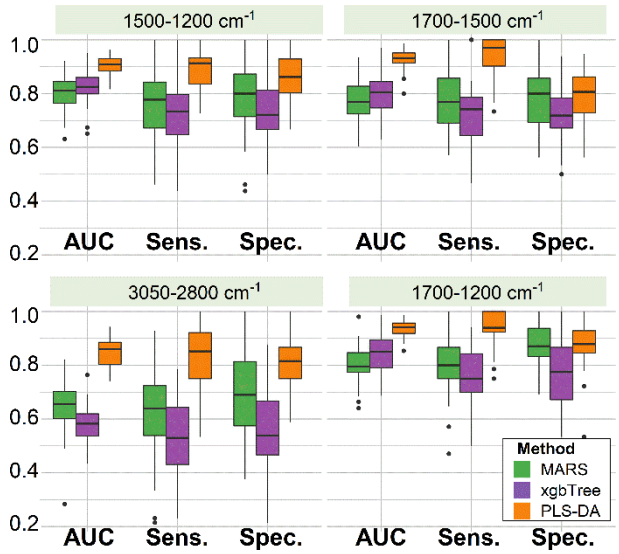


Figure 2. Overall performances of the three multivariate methods performed on the absorption spectra. Four spectral regions were tested: 3050-2800 cm^{-1}, 1700-1500 cm^{-1}, 1500-1200 cm^{-1}, and 1700-1200 cm^{-1} ranges. For each condition the resampled area under the curve (AUC), sensitivity (Sens.) and specificity (Spec.) are reported as box-plots. Black horizontal line within the box is the median, box ends show the first (Q1) and third quartile (Q3), lower whisker computed as the maximum value between the absolute minimum and Q1-1.5*IQR, and upper whisker as the minimum between the absolute maximum and Q3+1.5*IQR. Here, IQR is the interquartile range computed as Q3-Q1. Values beyond whiskers (outliers) are shown as black dots.

One of the main advantages of the above multivariate methods is the possibility to extract the wavenumbers that are responsible for the discrimination in a multivariate way (i.e. removal of one component will affect the contributions of the remaining ones). These wavenumbers can be therefore taken as marker bands whose intensity variations reflect the relevant and statistically significant molecular changes associated with the presence of the disease. In particular, the PLS-DA analysis detects as the most relevant the following peaks (Figure 3a): ~1650 cm^{-1} (assigned to protein α-helical and random coil structures[15,27]), ~1624 cm^{-1} (assigned to protein β-sheets[15,27]), ~1612 cm^{-1} (assigned to protein intermolecular β-sheets and/or amino acid side chains[14,15,27,35,36]), and ~1471 cm^{-1} (assigned mainly to the hydrocarbon chain CH_2 groups[28,37]). This band is sensitive to hydrocarbon chain packing and conformation[28]. In agreement with the PLS-DA outcomes, intensity differences of the above wavenumbers were observed in the absorbance (Figure 3b) and second derivative spectra (see later) of CR-positive patients and unaffected controls. In particular, the α-helical/random coil structures were found to decrease in the CR-positive samples, compared to controls, while the β structures increase. Moreover, the 1471 cm^{-1} band, mainly due to hydrocarbon chain CH_2 groups, was found to decrease in CR-positive samples, suggesting a variation of the lipid properties. Similar results were obtained by the MARS and xgbTree analyses (Figure S-4), which in addition suggested relevant spectral changes around 1540 cm^{-1} (Amide II[27]), ~1392 cm^{-1} (assigned to acidic side chains, as well as to lipid moieties[28,29,37–39]) and ~1294 cm^{-1} (tentatively assigned to the protein Amide III band[27]). To further investigate the lipid contribution in discriminating between CR-positive and CR-negative specimens, the ratio of absorption at 2925 cm^{-1} and at 2960 cm^{-1} (due to the CH_2 and CH_3 stretching, respectively) was considered. This ratio has also been reported to be important to evaluate the contribution of polypeptide absorption to other spectral regions in addition to the Amide I band when proteins accumulate in particular sample areas[40]. In the present study, there was no statistically significant difference in the 2925 cm^{-1} and 2960 cm^{-1} absorption ratio between the CR-positive and CR-negative specimens according to a two-tailed t-test: t=-1.94, p-value=0.054 (data not shown). This result is in agreement with the lowest discrimination accuracy obtained by the multivariate analyses performed in the 3050-2800 cm^{-1} spectral region (Figure 2 and Figure S-5). Noteworthy, abdominal fat aspirates are *per-se* particularly enriched of lipids with very high absorption in CH_2 and CH_3 stretching spectral region, which can eventually hide the spectral differences between the two groups.

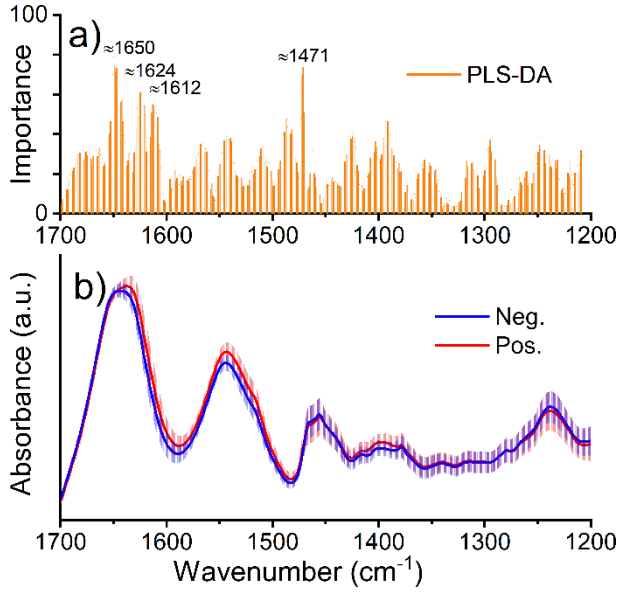


Figure 3. ATR-FTIR absorption spectra. (a) Wavenumber importance (domain 0-100) for PLS-DA discrimination performed in the 1700-1200 cm^{-1} spectral region. (b) Average absorption spectra of amyloid-positive (Pos.) and negative (Neg.) individuals. Error bars represent the standard deviations.

Multivariate analyses of the ATR-FTIR second derivative spectra. To further explore the potential of ATR-FTIR spectroscopy to characterize the fat aspirate, the multivariate analyses were also performed on the second derivatives of the absorption spectra. The overall performance in the discrimination between amyloid-positive patients and unaffected controls is reported in Figure 4 and Table S-2. Comparing these results with the one obtained using the absorption spectra, we can see a higher overall discrimination performance in general. The use of the second derivative spectra increased the overall discrimination performance (expressed as median [first, third quartile] over the repeated k-fold cross-validation) of the methods up to 97 [95-98]% AUC, 100 [89-100]% sensitivity, and 100 [89-100]% specificity in the case of PLS-DA performed in the 1700-1200 cm^{-1} (Figure 4 and Table S-2). Such a high performance may imply some degrees of over-fitting of the PLS-DA model. For this reason, it is important to take into account and compare the PLS-DA results with those of the other two methods (MARS and xgbTree) in order to perform a significant evaluation.

The wavenumber importance profiles obtained from the three methods are consistent among each other when performed on the second derivative spectra (Figure S-6), as well as on the absorption data (Figure S-4). Moreover, the wavenumbers that are more relevant for the discrimination in the multivariate analyses (Figure 5 and S-6) occurred in the same region where spectral differences between the two groups were observed in the second derivatives, whose minima correspond to absorption maxima (Figure 5). As already described in the "Multivariate analyses of the ATR-FTIR absorption spectra" paragraph, these spectral changes involve the protein and lipid IR responses (see Figure 5 and S-6 for peak assignment). In particular, they are indicative of an overall decrease of the α-helical/random coil structures (component around 1654 cm^{-1}) and an increase of β-sheets in the CR-positive samples (components in the 1630-1616 cm^{-1} range), compared to controls.

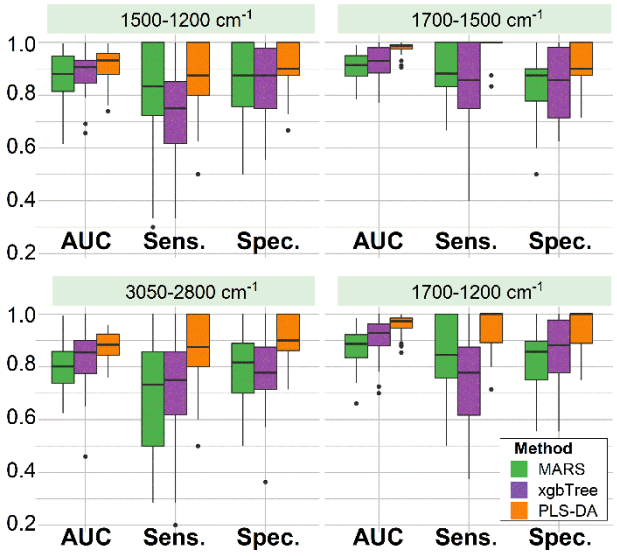


Figure 4. Overall performances of the MARS, xgbTree and PLS-DA methods performed on the second derivative spectra. Four spectral regions were tested and the resampled area under the curve (AUC), sensitivity (Sens.) and specificity (Spec.) are reported as in Figure 2.

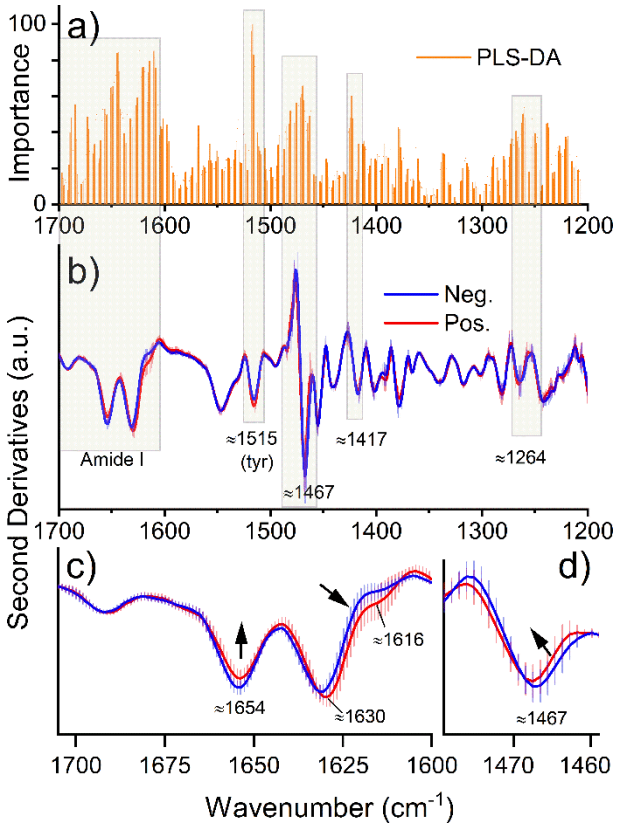


Figure 5. Second derivatives of the ATR-FTIR absorption spectra. (a) Wavenumber importance (domain 0-100) for PLS-DA discrimination performed in the 1700-1200 cm^{-1} spectral region. (b) Average second derivative spectra of amyloid-positive (Pos.) and unaffected (Neg.) individuals. The Amide I spectral region (c) and the components around 1467 cm^{-1} (d) were showed in expanded scales. Arrows point to the spectral changes occurring from control to CR-positive individuals. Error bars represent the standard deviations.

CONCLUSIONS

We showed that ATR-FTIR spectroscopy, coupled with multivariate analysis, can differentiate adipose tissue aspirates containing amyloid deposits from unaffected samples with high sensitivity and specificity, both of 100 [89-100]% (expressed as median [first, third quartile] over the repeated k-fold cross-validation) in the case of PLS-DA performed in the 1700-1200 cm^{-1} spectral region. This approach is rapid and requires no sample processing besides blood removal. The inherently structural information provided by FTIR is additive to that provided by CR. This dye, in fact, delivers indirect clues on the fibril conformation, thanks to its ability to interact with the repetitive β-sheet structure[41]. On the contrary, FTIR provides label-free knowledge on inter and intra-molecular β-sheets of proteins in the samples, and on other molecular species. Importantly, discrimination between amyloid-affected and negative samples was obtained on the basis of the whole spectrum, and not only by considering the spectral regions sensitive to the protein conformation. This indicates that changes in the molecular composition of tissues, in particular resident lipids, are intrinsic features of affected subcutaneous fat. Our report has notable implications of potential clinical relevance. Of note, the classification algorithm also correctly discriminated spectra from patients with small amounts of CR-positive areas. This may be due to the ability of FTIR spectroscopy, coupled to multivariate analysis, to reveal *in situ* not only protein deposits but also changes in the overall biomolecule composition and structural features. Moreover, the active area of the ATR crystal (about 1.8 mm of diameter in our device) allows obtaining the average infrared response of a relatively large part of the sample possibly enabling the detection of very small and scattered amyloid deposits.

Overall, this report opens the road for further applications of ATR-FTIR to analyze subcutaneous fat and other tissues in systemic amyloidosis; this could be most useful for diagnosis and for acquiring novel knowledge on the molecular bases of the disease.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Two supporting tables and 6 supporting figures (PDF).

AUTHOR INFORMATION

Corresponding Authors

* Antonino Natalello, Email: antonino.natalello@unimib.it

* Francesca Lavatelli, Email: francesca.lavatelli@unipv.it

Author Contributions

D.A., F.L., G.M., A.N. conceived the study; AF, PM, MN, GP acquired patients' samples and performed clinical and pathological characterization; FL and MT processed samples prior to ATR-FTIR; D.A. and A.N. performed the ATR-FTIR analyses; P.M. performed the multivariate analyses; D.A., P.M., F.L., G.M., A.N. wrote the manuscript. All authors have read and approved the final version of the manuscript.

‡These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by the University of Milano-Bicocca (grant "Fondo di Ateneo per la Ricerca" n. 2017-ATE-0339 to A.N.); Fondazione Cariplo (grant n. 2013–0964 to D.A., G.M., and A.N; grant n. 2015-0591 to G.P.; grant n. 2016-0489 to F.L.); Associazione Italiana per la Ricerca sul Cancro special program "5 per mille" (number 9965 to G.M.); the Italian Ministry of Health (grants RF-2013-02355259 to G.M. and RF-2016-02361756 to G.P.); Italian Medicines Agency (grant AIFA-2016-02364602 to G.P.), the European Union (E-Rare 2016 grant to G.P) and the Amyloidosis Foundation (to M.N.). We are grateful to Laura Verga and Gianluca Capello for electron microscopy imaging.

REFERENCES

(1) Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (6), 384–396.

(2) Chiti, F.; Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86* (1), 27–68.

(3) Sipe, J. D.; Benson, M. D.; Buxbaum, J. N.; Ikeda, S.; Merlini, G.; Saraiva, M. J. M.; Westermark, P. Amyloid Fibril Proteins and Amyloidosis: Chemical Identification and Clinical Classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid* **2016**, *23* (4), 209–213.

(4) Fernandez de Larrea, C.; Verga, L.; Morbini, P.; Klersy, C.; Lavatelli, F.; Foli, A.; Obici, L.; Milani, P.; Capello, G. L.; Paulli, M.; et al. A Practical Approach to the Diagnosis of Systemic Amyloidoses. *Blood* **2015**, *125* (14), 2239–2244.

(5) Nuvolone, M.; Merlini, G. Emerging Therapeutic Targets Currently under Investigation for the Treatment of Systemic Amyloidosis. *Expert Opin. Ther. Targets* **2017**, *21* (12), 1095–1110.

(6) Baker, M. J.; Hussain, S. R.; Lovergne, L.; Untereiner, V.; Hughes, C.; Lukaszewski, R. A.; Thiéfin, G.; Sockalingum, G. D. Developing and Understanding Biofluid Vibrational Spectroscopy: A Critical Review. *Chem. Soc. Rev.* **2016**, *45* (7), 1803–1818.

(7) Paraskevaidi, M.; Martin-Hirsch, P. L.; Martin, F. L. Progress and Challenges in the Diagnosis of Dementia: A Critical Review. *ACS Chem. Neurosci.* **2018**, *9* (3), 446–461.

(8) Ami, D.; Natalello, A.; Doglia, S. M. Fourier Transform Infrared Microspectroscopy of Complex Biological Systems: From Intact Cells to Whole Organisms. *Methods Mol. Biol.* **2012**, *895*, 85–100.

(9) Ami, D.; Mereghetti, P.; Leri, M.; Giorgetti, S.; Natalello, A.; Doglia, S. M.; Stefani, M.; Bucciantini, M. A FTIR Microspectroscopy Study of the Structural and Biochemical Perturbations Induced by Natively Folded and Aggregated Transthyretin in HL-1 Cardiomyocytes. *Sci. Rep.* **2018**, *8* (1), 12508.

(10) Choo, L. P.; Wetzel, D. L.; Halliday, W. C.; Jackson, M.; LeVine, S. M.; Mantsch, H. H. In Situ Characterization of Beta-Amyloid in Alzheimer's Diseased Tissue by Synchrotron Fourier Transform Infrared Microspectroscopy. *Biophys. J.* **1996**, *71* (4), 1672–1679.

(11) Ami, D.; Lavatelli, F.; Rognoni, P.; Palladini, G.; Raimondi, S.; Giorgetti, S.; Monti, L.; Doglia, S. M.; Natalello, A.; Merlini, G. In Situ Characterization of Protein Aggregates in Human Tissues Affected by Light Chain Amyloidosis: A FTIR Microspectroscopy Study. *Sci. Rep.* **2016**, *6* (1), 29096.

(12) van Gameren, I. I.; Hazenberg, B. P. C.; Bijzet, J.; Haagsma, E. B.; Vellenga, E.; Posthumus, M. D.; Jager, P. L.; van Rijswijk, M. H. Amyloid Load in Fat Tissue Reflects Disease Severity and Predicts Survival in Amyloidosis. *Arthritis Care Res. (Hoboken).* **2010**, *62* (3), 296–301.

(13) Gertz, M. A.; Comenzo, R.; Falk, R. H.; Fermand, J. P.; Hazenberg, B. P.; Hawkins, P. N.; Merlini, G.; Moreau, P.; Ronco, P.; Sanchorawala, V.; et al. Definition of Organ Involvement and Treatment Response in Immunoglobulin Light Chain Amyloidosis (AL): A Consensus Opinion from the 10th International Symposium on Amyloid and Amyloidosis. *Am. J. Hematol.* **2005**, *79* (4), 319–328.

(14) Natalello, A.; Mangione, P. P.; Giorgetti, S.; Porcari, R.; Marchese, L.; Zorzoli, I.; Relini, A.; Ami, D.; Faravelli, G.; Valli, M.; et al. Co-Fibrillogenesis of Wild-Type and D76N β 2 -Microglobulin. *J. Biol. Chem.* **2016**, *291* (18), 9678–9689.

(15) Natalello, A.; Doglia, S. M. S. M. Insoluble Protein Assemblies Characterized by Fourier Transform Infrared Spectroscopy. *Methods Mol. Biol.* **2015**, *1258*, 347–369.

(16) Bassan, P.; Sachdeva, A.; Kohler, A.; Hughes, C.; Henderson, A.; Boyle, J.; Shanks, J. H.; Brown, M.; Clarke, N. W.; Gardner, P. FTIR Microscopy of Biological Cells and Tissue: Data Analysis Using Resonant Mie Scattering (RMieS) EMSC Algorithm. *Analyst* **2012**, *137* (6), 1370–1377.

(17) Filzmoser, P.; Maronna, R.; Werner, M. Outlier Identification in High Dimensions. *Comput. Stat. Data Anal.* **2008**, *52* (3), 1694–1711.

(18) Friedman, J.; Hastie, T.; Tibshirani, R. Additive Logistic Regression: A Statistical View of Boosting (With Discussion and a Rejoinder by the Authors). *Ann. Stat.* **2000**, *28* (2), 337–407.

(19) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29* (5), 1189–1232.

(20) Friedman, J. H. Multivariate Adaptive Regression Splines. *Ann. Stat.* **1991**, *19* (1), 1–67.

(21) Pérez-Enciso, M.; Tenenhaus, M. Prediction of Clinical Outcome with Microarray Data: A Partial Least Squares Discriminant Analysis (PLS-DA) Approach. *Hum. Genet.* **2003**, *112* (5), 581–592.

(22) Breiman, L.; Friedman, J.; Stone, C. J.; Olshen, R. A. Classification and Regression Trees; The Wadsworth and Brooks-Cole statistics-probability series; Taylor & Francis, 1984.

(23) Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27* (8), 861–874.

(24) Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28* (5).

(25) Martens, H.; Næs, T. Multivariate Calibration; John Wiley and Sons Inc.: New York, 1989.

(26) Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A.; et al. Using Fourier Transform IR Spectroscopy to Analyze Biological Materials. *Nat. Protoc.* **2014**, *9* (8), 1771–1791.

(27) Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta - Bioenerg.* **2007**, *1767* (9), 1073–1101.

(28) Casal, H. L.; Mantsch, H. H. Polymorphic Phase Behaviour of Phospholipid Membranes Studied by Infrared Spectroscopy. *BBA - Rev. Biomembr.* **1984**, *779* (4), 381–401.

(29) Barth, A. The Infrared Absorption of Amino Acid Side Chains. *Prog. Biophys. Mol. Biol.* **2000**, *74* (3–5), 141–173.

(30) Paraskevaidi, M.; Morais, C. L. M.; Lima, K. M. G.; Snowden, J. S.; Saxon, J. A.; Richardson, A. M. T.; Jones, M.; Mann, D. M. A.; Allsop, D.; Martin-Hirsch, P. L.; et al. Differential Diagnosis of Alzheimer's Disease Using Spectrochemical Analysis of Blood. *Proc. Natl. Acad. Sci.* **2017**, *114* (38), E7929–E7938.

(31) Ami, D.; Mereghetti, P.; Natalello, A.; Doglia, S. M.; Zanoni, M.; Redi, C. A.; Monti, M. FTIR Spectral Signatures of Mouse Antral Oocytes: Molecular Markers of Oocyte Maturation and Developmental Competence. *Biochim. Biophys. Acta - Mol. Cell Res.* **2011**, *1813* (6), 1220–1229.

(32) Ami, D.; Doglia, S. M.; Mereghetti, P. Multivariate Analysis for Fourier Transform Infrared Spectra of Complex Biological Systems and Processes; IntechOpen, 2013.

(33) Brézillon, S.; Untereiner, V.; Lovergne, L.; Tadeo, I.; Noguera, R.; Maquart, F. X.; Wegrowski, Y.; Sockalingum, G. D. Glycosaminoglycan Profiling in Different Cell Types Using Infrared Spectroscopy and Imaging. *Anal. Bioanal. Chem.* **2014**, *406* (24), 5795–5803.

(34) Adam-Bourdarios, C.; Cowan, G.; Germain-Renaud, C.; Guyon, I.; Kégl, B.; Rousseau, D. The Higgs Machine Learning Challenge. *J. Phys. Conf. Ser.* **2015**, *664* (7), 072015.

(35) Lomont, J. P.; Rich, K. L.; Maj, M.; Ho, J.-J.; Ostrander, J. S.; Zanni, M. T. Spectroscopic Signature for Stable β-Amyloid Fibrils versus β-Sheet-Rich Oligomers. *J. Phys. Chem. B* **2018**, *122* (1), 144–153.

(36) Natalello, A.; Relini, A.; Penco, A.; Halabelian, L.; Bolognesi, M.; Doglia, S. M.; Ricagno, S. Wild Type Beta-2 Microglobulin and DE Loop Mutants Display a Common Fibrillar Architecture. *PLoS One* **2015**, *10* (3), e0122449.

(37) Arrondo, J. L. R.; Goñi, F. M. Infrared Studies of Protein-Induced Perturbation of Lipids in Lipoproteins and Membranes. *Chem. Phys. Lipids* **1998**, *96* (1–2), 53–68.

(38) Levchuk, Y. N. Infrared Spectra of Steroids. *Russ. Chem. Rev.* **1968**, *37* (2), 155–171.

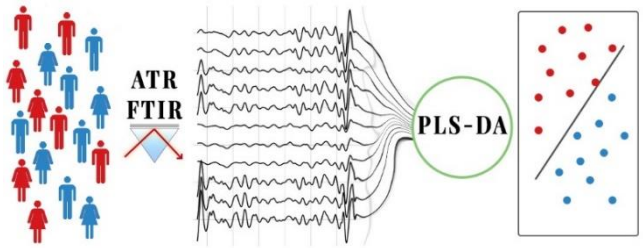
(39) Lewis, R. N. A. H.; McElhaney, R. N. Fourier Transform Infrared Spectroscopy in the Study of Lipid Phase Transitions in Model and Biological Membranes. Practical Considerations.; Dopico, A. M., Ed.; Humana Press: Totowa, NJ, 2007; pp 207–226.

(40) Benseny-Cases, N.; Álvarez-Marimon, E.; Castillo-Michel, H.; Cotte, M.; Falcon, C.; Cladera, J. Synchrotron-Based Fourier Transform Infrared Microspectroscopy (MFTIR) Study on the Effect of Alzheimer's Aβ Amorphous and Fibrillar Aggregates on PC12 Cells. *Anal. Chem.* **2018**, *90* (4), 2772–2779.

(41) Groenning, M. Binding Mode of Thioflavin T and Other Molecular Probes in the Context of Amyloid Fibrils—current Status. *J. Chem. Biol.* **2010**, *3* (1), 1–18.

For toc only

Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. Consult the journal's Instructions for Authors for TOC graphic specifications.



Insert Table of Contents artwork here