



OPEN

## Evaluation of aromatic amino acids as potential biomarkers in breast cancer by Raman spectroscopy analysis

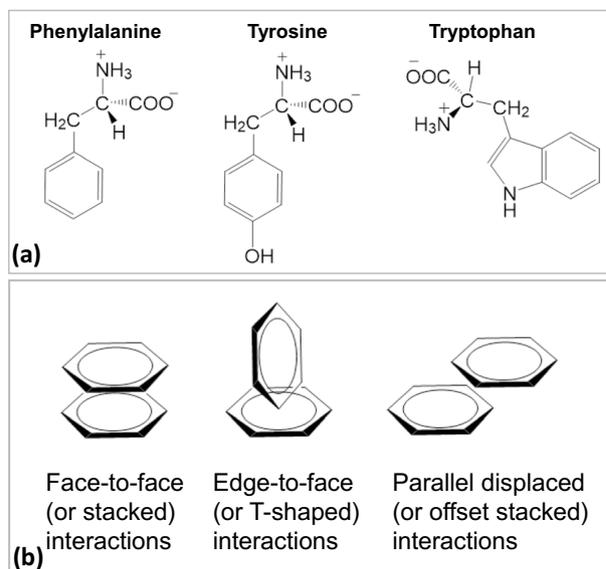
Shaymus Contorno<sup>1</sup>, Richard E. Darienzo<sup>1</sup> & Rina Tannenbaum<sup>1,2</sup>✉

The scope of the work undertaken in this paper was to explore the feasibility and reliability of using the Raman signature of aromatic amino acids as a marker in the detection of the presence of breast cancer and perhaps, even the prediction of cancer development in very early stages of cancer onset. To be able to assess this hypothesis, we collected most recent and relevant literature in which Raman spectroscopy was used as an analytical tool in the evaluation of breast cell lines and breast tissue, re-analyzed all the Raman spectra, and extracted all spectral bands from each spectrum that were indicative of aromatic amino acids. The criteria for the consideration of the various papers for this study, and hence, the inclusion of the data that they contained were two-fold: (1) The papers had to focus on the characterization of breast tissue with Raman spectroscopy, and (2) the spectra provided within these papers included the spectral range of 500–1200  $\text{cm}^{-1}$ , which constitutes the characteristic region for aromatic amino acid vibrational modes. After all the papers that satisfied these criteria were collected, the relevant spectra from each paper were extracted, processed, normalized. All data were then plotted without bias in order to decide whether there is a pattern that can shed light on a possible diagnostic classification. Remarkably, we have been able to demonstrate that cancerous breast tissues and cells decidedly exhibit overexpression of aromatic amino acids and that the difference between the extent of their presence in cancerous cells and healthy cells is overwhelming. On the basis of this analysis, we conclude that it is possible to use the signature Raman bands of aromatic amino acids as a biomarker for the detection, evaluation and diagnosis of breast cancer.

Phenylalanine, tyrosine and tryptophan, shown in Fig. 1a, are the only three amino acids that contain benzyl-based aromatic groups, and hence, are referred to as aromatic amino acids (AAA). These residues have a rigid, planar structure and possess added stability due to the  $\pi$ -electron cloud situated above and below the plane of the aromatic ring<sup>1–3</sup>. The AAA residues can undergo aromatic-aromatic interactions, consisting of hydrogen bonding coupled with attractive, noncovalent, dipole and van der Waals interactions and  $\pi$ -stacking of the benzene rings<sup>4–8</sup>. These interactions result in an optimal geometry where the partially-charged positive hydrogens of the C–C edges in one ring interact favorably with the  $\pi$ -electrons and partially-charged negative carbons of the other ring. The  $\pi$ -system of the aromatic rings gives rise to three types of interactions involving aromatic moieties: (a)  $\pi$ - $\pi$  alignment of the aromatic residues, (b) Dipole- $\pi$  system attraction of cations within 6.0 Å of the face of an aromatic ring, and (c) Long range CH- $\pi$  interactions<sup>6</sup>. These interactions give rise to three different types of geometries involving two aromatic residues, as shown in Fig. 1b: (a) Edge-to-face (or T-shaped), (b) Face-to-face (or stacked)<sup>7</sup>, and (c) Parallel displaced (or offset stacked) interactions. These aromatic-aromatic interactions have been found to play an important role in the stabilization of the overall structure of protein molecules<sup>9–14</sup> and protein-DNA/RNA complexes through free energy lowering by  $-0.6$  and  $-1.3$  kcal/mol, respectively<sup>15–17</sup>. Hence, the interactions between the aromatic residues within a protein and in protein-DNA/RNA complexes constitute a crucial element in the proper functioning of protein molecules, which in turn impacts various cellular processes in which these proteins take part<sup>5,18</sup>.

Besides the structural function in proteins, aromatic amino acids are precursors to many biological compounds that are essential for the normal functioning of live organisms<sup>18,19</sup>. Many of these compounds may be

<sup>1</sup>Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, NY 11794, USA. <sup>2</sup>The Stony Brook Cancer Center, Stony Brook University, Stony Brook, NY 11794, USA. ✉email: irena.tannenbaum@stonybrook.edu



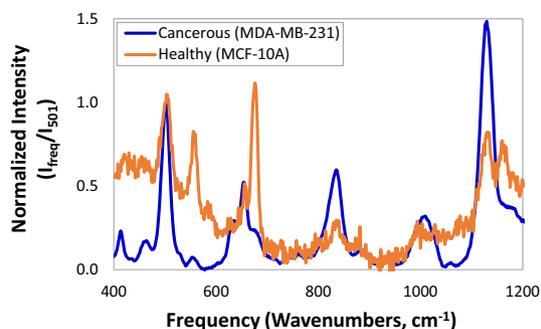
**Figure 1.** (a) The chemical structure of the three phenyl-based aromatic amino acids phenylalanine, tyrosine and tryptophan, and (b) The schematic representation of the three types of interactions between the phenyl rings of the aromatic amino acids.

involved in the pathogenesis of many diseases, such as cancers, psychiatric disorders, depression, cardiovascular diseases, chronic kidney insufficiency or diabetes. In recent years, the impact of the variations in aromatic amino acid content and sequence has been the focus of much effort in understanding the nature of disordered proteins and their involvement in cancer pathology, as well as the impact of proper membrane functioning. It is well established that metabolic abnormalities occur prior to significant morphological changes in malignant tissue<sup>20</sup>. For example, high levels of aromatic amino acids that are present in gastric juice have been deemed indicative of the onset of early gastric cancer (ECG)<sup>21</sup>. Since gastric juice characteristics contain considerable information concerning the metabolic state of the gastric epithelium, high levels of aromatic amino acids could constitute valid biomarkers for the detection of such a cancer<sup>21,22</sup>.

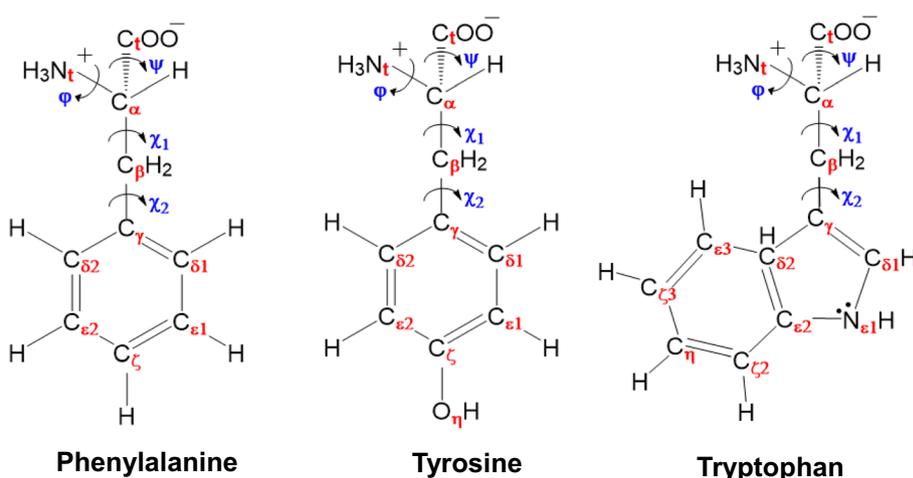
Increased concentrations of AAA have been found in additional various cancer types, such as prostate, lung, breast, oral and esophageal tissue samples. For example, it is well known that some tryptophan metabolites exhibit active involvement in carcinogenesis. Indolic derivatives of tryptophan have been shown to activate the aryl hydrocarbon receptor, which in turn can be transported to the nucleus and bind to a specific DNA sequence, thus acting as a transcription factor. Hence, tryptophan metabolites that are not carcinogens by themselves can promote a neoplastic process when coupled with other enzymatic reactions<sup>23–26</sup>. This has been shown to occur in urinary bladder cancer and lung cancer. The tryptophan metabolites, in addition to acting as chimeric transcription factors, are also involved in the development of tumor-induced immunosuppression through proliferation inhibition and promotion of lymphocyte-T apoptosis, as is the case in some types of breast cancer. In addition to tryptophan, also tyrosine and phenylalanine can constitute the precursors to wide array of substances that may have mutagenic, genotoxic and carcinogenic properties, such as phenolic and indolic compounds<sup>25,27,28</sup>. The overexpression of these molecules in colon tissue may be indicative of the onset of colon cancer.

The enhanced expression of aromatic amino acids in breast cancer became apparent in a multitude of recent studies, in which cancer tissue characteristics was measured by Raman spectroscopy<sup>29–49</sup>. The SK-BR-3 cell lines, characterized by the overexpression of HER-2 receptor, and MDA-MB-231 cell lines, characterized by the lack of estrogen, progesterone and HER-2 receptors, exhibited intense Raman bands that correspond to aromatic amino acids such as tyrosine, phenylalanine, and tryptophan<sup>29</sup>, particularly when using the surface enhance Raman spectroscopy technique. Raman spectroscopy has emerged in recent years as a successful tool for the diagnosis of breast cancer tissue samples by detecting lipid and protein identity, structure and concentrations based on known samples or difference spectra. If the spectral profiles acquired via Raman spectroscopy possess high enough intensity and signal-to-noise ratios, then the subcellular changes present in different cancer phenotypes should be discernible. In view of the considerable involvement and enhanced presence of aromatic amino acids in malignant tissues, they could also be used as an important marker in the diagnosis and prediction of cancer onset.

The scope of the work undertaken in this paper was to explore the feasibility and reliability of using the Raman signature of aromatic amino acids as a marker in the detection of the presence of breast cancer and perhaps, even the prediction of cancer development in very early stages of cancer onset. To be able to assess this hypothesis, we collected most recent and relevant literature in which Raman spectroscopy was used as an analytical tool in the evaluation of breast cell lines and breast tissue, re-analyzed all the Raman spectra, and extracted all spectral bands from each spectrum that were indicative of aromatic amino acids. All data were then plotted without bias in order to observe whether there is a pattern that can shed light on a possible diagnostic classification.



**Figure 2.** Typical Raman spectra of healthy breast cells and cancerous breast cells. The cancerous cells belong to the MDA-MB-231 cell line and consist of cancer cells that lack estrogen receptors (ER), progesterone receptors (PR), and HER2 amplification, and are also known as "triple negative". The healthy cells belong to the MCF-10A cell line and are benign cells with fibrocystic changes. The spectra were normalized with respect to the intensity of the 501  $\text{cm}^{-1}$  peak belonging to the underlying Si substrate onto which the cell samples were placed.



**Figure 3.** The assigned atom identities and main conformational angles of the zwitterionic form of the aromatic amino acids phenylalanine, tyrosine, and tryptophan. Nt and Ct refer to the main atoms involved in the terminal charged groups,  $\phi$  and  $\psi$  are the backbone conformational angles that are represented by (Ct–Ca–Nt–H) and (Nt–Ca–Ct–O), respectively.  $\chi_1$  and  $\chi_2$  are the side chain conformational torsion angles that are represented by (Nt–Ca–C $\beta$ –C $\gamma$ ) and (Ca–C $\beta$ –C $\gamma$ –C $\delta_1$ ), respectively. Note that for phenylalanine, the phenyl ring carbon atoms C $\delta_1$  and C $\delta_2$  are equivalent.

## Results and discussion

Typical Raman spectra of healthy breast cells and cancerous breast cells are shown in Fig. 2. The two human-mammary epithelial cell lines consist of the MCF-10A cells, which are benign cells with fibrocystic changes and MDA-MB-231 cells, which are cancer cells that lack estrogen receptors (ER), progesterone receptors (PR), and HER2 amplification, and are also known as "triple negative". The spectra were normalized with respect to the intensity of the 501  $\text{cm}^{-1}$  peak belonging to the underlying Si substrate onto which the cell samples were placed.

The Raman shift peak frequencies that we followed and analyzed correspond to those characteristic of aromatic amino acids at 657, 816, 837, 845, 965, 1006, 1012 and 1131  $\text{cm}^{-1}$ <sup>129,54–58</sup>, and to S–S bonds at 503  $\text{cm}^{-1}$ <sup>129</sup>. The assignment of the vibrational modes of these peaks is shown in Fig. 3 and summarized in Table 1.

Phenylalanine has two main characteristic bands: one is at 816  $\text{cm}^{-1}$ , corresponding to the in-plane vibrations between the substituted carbon and the immediately-bonded carbon of the side chain (C $\gamma$ –C $\beta$ ) and the double bond between the substituted carbon and its ortho neighbor (C $\gamma$ –C $\delta_1$ ), and the cooperative vibration of the three ring carbons opposite the substituted carbon, (C $\epsilon_1$ –C $\zeta$ –C $\epsilon_3$ ). The second is at 1006  $\text{cm}^{-1}$ , corresponding to the phenyl ring angular bending (i.e. distortion) due to contributions from the changes in the relative position between the substituted carbon and both its ortho neighbors (C $\gamma$ –C $\delta_1$  and C $\gamma$ –C $\delta_2$ ) and contributions from the cooperative vibrations of the conjugated double bond groups (C $\delta_2$ –C $\gamma$ –C $\epsilon_1$ , C $\gamma$ –C $\delta_2$ –C $\epsilon_2$  and C $\zeta$ –C $\epsilon_1$ –C $\delta_1$ ). Tyrosine has a para-substituted phenyl ring, with a hydroxy group at the C $\zeta$  carbon. While the molecule is similar to phenylalanine, it has, however, a lower degree of symmetry due to the para-substitution and hence, exhibits 4 main characteristic bands. The band at 657  $\text{cm}^{-1}$  corresponds to the in-plane ring vibration of the conjugated molecular fragments of the phenyl ring (C $\gamma$ –C $\delta_1$ –C $\epsilon_1$ , C $\gamma$ –C $\delta_2$ –C $\epsilon_2$ , C $\delta_1$ –C $\epsilon_1$ –C $\zeta$  and C $\zeta$ –C $\epsilon_2$ –C $\delta_2$ ). The

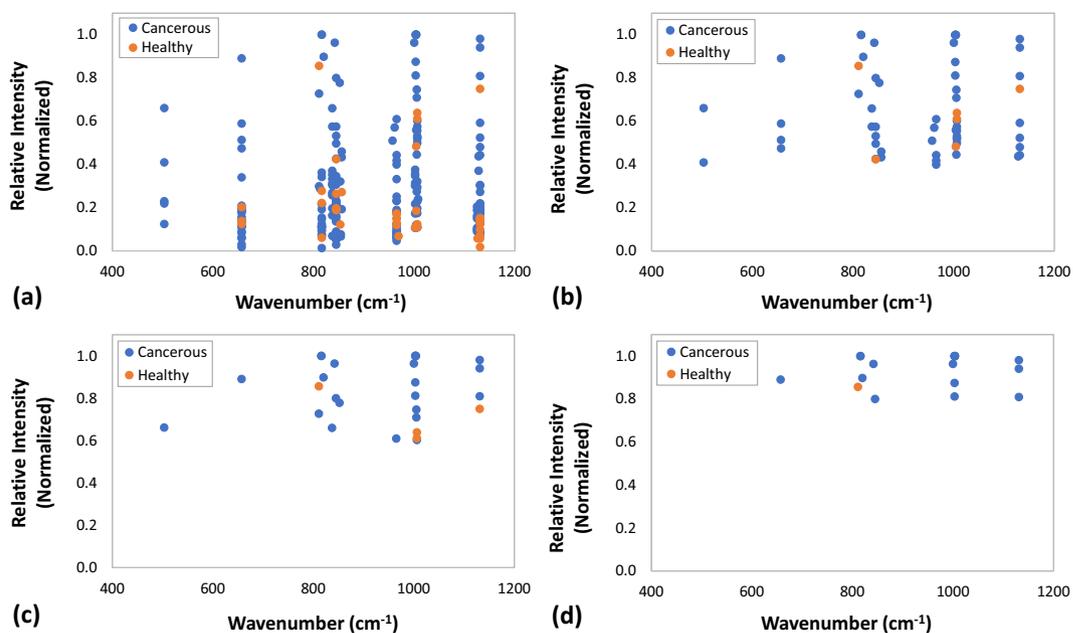
Wavenumber (cm <sup>-1</sup> )	Molecular origin	Vibrational mode assignment	References
503	S-S bonds	Vibration if S-S disulfide bridge	29
657	Tyrosine	In plane ring vibration (6B) C <sub>γ</sub> -C <sub>δ1</sub> -C <sub>ε1</sub> , C <sub>γ</sub> -C <sub>δ2</sub> -C <sub>ε2</sub> , C <sub>δ1</sub> -C <sub>ε1</sub> -C <sub>ζ</sub> , C <sub>ζ</sub> -C <sub>ε2</sub> -C <sub>δ2</sub>	54-58
816	Phenylalanine	C <sub>β</sub> -C <sub>γ</sub> ; C <sub>γ</sub> -C <sub>δ1</sub> , C <sub>ε1</sub> -C <sub>ζ</sub> -C <sub>ε2</sub>	56,58
837	Tyrosine	Out of plane bending (C <sub>ε2</sub> -H) and (C <sub>ζ</sub> -O)	56,58
845	Tyrosine	Out of plane bending (C <sub>ε2</sub> -H)	56,58
965	Tyrosine	Out of plane ring vibration (17a) C <sub>α</sub> -C <sub>β</sub> and Water(H)...O(Ct)	54-58
1006	Phenylalanine	Phenyl ring angular bending C <sub>γ</sub> -C <sub>δ1</sub> , C <sub>γ</sub> -C <sub>δ2</sub> , C <sub>δ2</sub> -C <sub>γ</sub> -C <sub>ε1</sub> , C <sub>γ</sub> -C <sub>δ2</sub> -C <sub>ε2</sub> , C <sub>ζ</sub> -C <sub>ε1</sub> -C <sub>δ1</sub>	56,58
1012	Tryptophan	Symmetric benzene/pyrrole out of phase breathing mode C <sub>ζ3</sub> -C <sub>η</sub> , C <sub>ζ2</sub> -C <sub>η</sub> , C <sub>ε3</sub> -C <sub>ζ3</sub>	56,58
1131	Tryptophan	NtH <sub>3</sub> <sup>+</sup> asymmetric rock and C <sub>η</sub> -C <sub>ζ2</sub> -H	56,58

**Table 1.** Summary of the Raman vibrational peaks and the assignment of the vibrational modes of the three main aromatic amino acids: phenylalanine, tyrosine and tryptophan.

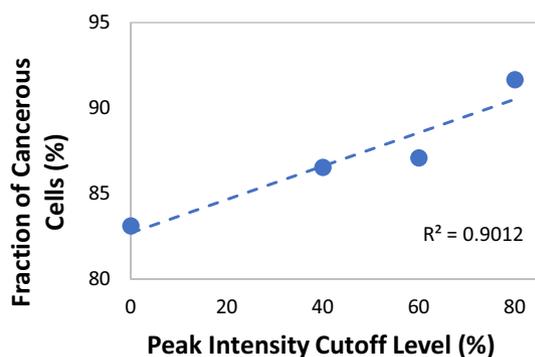
doublet bands at 837 and 845 cm<sup>-1</sup> correspond to the out-of-plane bending along the bond between the unsubstituted phenyl ring carbons and associated hydrogen (e.g. C<sub>ε1</sub>-H and C<sub>ε2</sub>-H) and the out-of-plane bending along the carbon-oxygen bond in the hydroxyl group (C<sub>ζ</sub>-O). The fourth characteristic band at 965 cm<sup>-1</sup> corresponds to the out-of-plane ring vibration of the two main carbons of the substituent functional group on the phenyl carbon (C<sub>α</sub>-C<sub>β</sub>) and the hydrogen bonding associated with the carboxylic group on the terminal carbon and water (HO-H...OOCt-). Finally, tryptophan exhibits 2 characteristic band: one band is at 1012 cm<sup>-1</sup>, corresponding to the symmetric benzene and/or pyrrole out of phase breathing modes (e.g. the cooperative vibrations of C<sub>ζ3</sub>-C<sub>η</sub>, C<sub>ζ2</sub>-C<sub>η</sub> and C<sub>ε3</sub>-C<sub>ζ3</sub>, and the second band is at 1131 cm<sup>-1</sup>, corresponding to the asymmetric rocking vibration of the terminal amine group (NtH<sub>3</sub><sup>+</sup>) and the cooperative vibration modes of unsubstituted ring carbons and associated hydrogens (e.g. C<sub>η</sub>-C<sub>ζ2</sub>-H). It is important to note that while the 1012 cm<sup>-1</sup> band is characteristic of tryptophan, its intensity in all the spectra that were examined is very low and hence, has not been used in the overall peak analysis. Also, the S-S vibration peaks at 503 cm<sup>-1</sup> are very weak and appear in all spectra, they are not specific to the aromatic amino acids, but rather ubiquitous in proteins as part of their tertiary structure, and therefore, were not included in the statistical analysis.

The collection of all the normalized Raman band intensities in the spectral range of 500–1200 cm<sup>-1</sup> for all breast tissues examined<sup>29–49</sup> are shown in Fig. 4a. All the peaks collected from the pertinent Raman spectra in the literature were grouped into two cell categories, cancerous and healthy, the latter category including also cells that have been deemed benign (i.e. non-cancerous) in their respective publications<sup>29–49</sup>. It is evident from this representation of the data, that at higher relative peak intensities, most of the peaks belong to the cancerous cells across all frequencies. The statistical analysis indicates that most of the data points assigned to aromatic amino acids in the analyzed spectra represent Raman bands generated by cancerous cells (or tissues) as they comprise 83% of the data, while healthy cells comprise only over 16%. In order to better highlight this trend, we replotted the normalized peaks having intensities in the upper 60th percentile (intensities ≥ 40%), shown in Fig. 4b, the upper 40th percentile (intensities ≥ 60%), shown in Fig. 4c and in the upper 20th percentile (intensities ≥ 80%), shown in Fig. 4d. As the cutoff for the relative intensity increased, so did the fraction of the data points of aromatic amino acids originating from cancerous cells, as shown in Fig. 5. While the statistical analysis of the contributions of both healthy and cancerous cells becomes less accurate as the total number of data points decrease at higher relative intensity thresholds, it nevertheless, indicates that the characteristic Raman bands of aromatic amino acids in cancerous cells exhibit higher intensities and hence, they are present in higher concentrations<sup>33,43,45</sup>. This conclusion is in accordance with various observations regarding the enhanced prevalence of aromatic amino acids in cancerous tissues. For example, a comparison between the non-malignant MCF-10A with malignant SK-BR-3 breast cell lines indicates an increase in the concentration of tyrosine-rich proteins, as evidenced by the high intensity band at 837 cm<sup>-1</sup>, in accordance with the structure of the HER-2 receptor<sup>29,59–62</sup>. In addition, it has been observed that the Raman spectra corresponding to aromatic amino acids in malignant SK-BR-3 and MDA-MB-231 breast cell lines exhibit intense bands characteristic to tyrosine, phenylalanine, and tryptophan, suggesting the enhancement in the expression of such aromatic amino acid rich proteins in malignant cells<sup>29,63–68</sup>. This phenomenon is not restricted to breast cancer alone, but is prevalent in other malignancies such as pancreatic cancer, colon cancer, gastric cancer, and lung cancer. The overexpression of aromatic amino acids in proteins is considered as a factor that contributes to the destabilization of their tertiary structure and hence, induces the reduction in their 3D structural rigidity and the enhancement of protein disorder, a characteristic that is prevalent in oncoproteins involved in malignant tumor genesis and robustness<sup>66,67</sup>.

To further highlight the specificity of the enhanced presence of aromatic amino acids in cancerous cells, the normalized abundance of the data points corresponding to aromatic amino acids in the analyzed Raman spectra were grouped as a function of presence or absence of cell malignancy and of Raman shift frequency, and shown in Fig. 6. The normalization of data point abundance at a given Raman shift frequency was obtained by dividing the number of data points at a particular frequency for one of the two cell categories by the total number of data points at that frequency and the result was then further multiplied by the relative weight of the specific cell category in the overall data point population, as shown in Eq. (1):



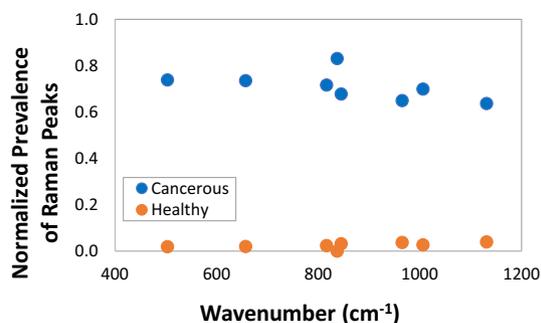
**Figure 4.** The collection of all the Raman bands in the spectral range of 500–1200  $\text{cm}^{-1}$  that were extracted from the pertinent Raman spectra of breast tissues in the literature. **(a)** The collection of all the normalized Raman band intensities, **(b)** The normalized peaks having intensities in the upper 60th percentile (intensities  $\geq 40\%$ ), **(c)** The normalized peaks having intensities in the upper 40th percentile (intensities  $\geq 60\%$ ), and **(d)** The normalized peaks having intensities in the upper 20th percentile (intensities  $\geq 80\%$ ). All the data points were classified into two cell categories, cancerous and healthy.



**Figure 5.** The fraction of the number of data points of aromatic amino acids originating from cancerous cells as a function of the cutoff for the relative intensity of the corresponding Raman bands. As the total number of points decreases as a function of the increase in the normalized intensity threshold, the statistical analysis of the contributions of both healthy and cancerous cells becomes less accurate. However, the trend is indicative of the fact that the characteristic Raman bands of aromatic amino acids in cancerous cells exhibited higher intensities and hence, were present in higher concentrations.

$$A_{\nu_i}^{C_1} = \frac{P_{\nu_i}^{C_1}}{(P_{\nu_i}^{C_1} + P_{\nu_i}^{C_2})} \cdot \frac{\sum_i P_{\nu_i}^{C_1}}{\sum_i (P_{\nu_i}^{C_1} + P_{\nu_i}^{C_2})} \text{ and } A_{\nu_i}^{C_2} = \frac{P_{\nu_i}^{C_2}}{(P_{\nu_i}^{C_1} + P_{\nu_i}^{C_2})} \cdot \frac{\sum_i P_{\nu_i}^{C_2}}{\sum_i (P_{\nu_i}^{C_1} + P_{\nu_i}^{C_2})} \quad (1)$$

where  $A_{\nu_i}^{C_1}$  and  $A_{\nu_i}^{C_2}$  are the normalized abundances of both cell categories,  $C_1$  and  $C_2$ , at a specific Raman shift frequency  $\nu_i$  and  $P_{\nu_i}^{C_1}$  and  $P_{\nu_i}^{C_2}$  are the number of data points for each cell category at that given frequency. The pattern that emerges shows that there is a distinct separation between the data points belonging to aromatic amino acids in healthy cells and those belonging to cancerous cells, with the latter occupying the more abundant regions of the plot. This indicates that the prevalence of aromatic amino acids in cancerous cells is more pronounced and there is a likelihood that this prevalence could be used as a diagnostic tool in identifying breast malignancy.



**Figure 6.** The statistical analysis of the normalized prevalence of the data points corresponding to aromatic amino acids in the analyzed Raman spectra for each cell type category. The normalization of data point abundance at a given Raman shift frequency was obtained by the fraction of data points at a particular frequency for one of the two cell categories multiplied by the weight of the particular cell category in the overall data population.

## Conclusions

The applicability of Raman spectroscopy as a diagnostic tool in the prediction, detection and evaluation of cancer tissues has been gaining considerable momentum in recent years. In particular, new techniques designed to enhance the intensity of Raman peaks, such as resonance Raman or surface enhanced Raman, and the attenuation or even elimination of fluorescence in biological samples by using near infrared incidence lasers, have highlighted the potential of using this type of spectroscopy in a clinical setting. The ability to correlate Raman spectra with specific locations within a sample, afford a cellular and even a subcellular glimpse into the molecular make-up of that given sample. In order to tap into this very promising potential, it would be imperative to determine in each case which molecular details would provide a unique signature that would allow a definitive discernment between healthy and cancerous tissues. In this context, it has been observed in recent years that various types of cancers exhibit an overexpression of aromatic amino acids. Since aromatic amino acids play important roles in a variety of metabolic processes, as well as key functions in the structural integrity of cellular and membrane proteins, anecdotal observations regarding their enhanced presence in cancerous tissues has given way to more dedicated and targeted inquiries. Given some observations that we have also made in our recent exploration of the use of SERS for the identification of breast cancer tissue, we have set out to probe the role of aromatic amino acids in breast cancer and to ascertain the possibility of using their overexpression as a diagnostic tool. The current work is a compilation of all pertinent publications dealing with the characterization of breast cancer cell lines and tissues using various modalities of Raman spectroscopy. All the Raman spectra in these papers, irrespective of the methodology used, have been normalized and have undergone detailed spectroscopic, chemical and statistical analysis. The emphasis of the analysis centered on the spectral region of 500–1200  $\text{cm}^{-1}$  that has been shown to be the most characteristic for aromatic amino acids, specifically tryptophan, phenylalanine and tyrosine. We have been able to demonstrate that cancerous breast tissues and cells markedly exhibit overexpression of aromatic amino acids and that the difference between the extent of their presence in cancerous cells and healthy cells is decisive. The hypothesis for this work does not seek to establish Raman spectroscopy as a de-facto standard for the prediction or diagnosis of cancer, but rather as a tool with the ability to add, as well as complement, to the myriad of tools available to doctors and patients. Based on this analysis of the pertinent literature, we conclude that it may be possible to use the signature Raman bands of aromatic amino acids as a biomarker for the detection, evaluation and diagnosis of breast cancer, paving the way for future work by those in the field to set up in-vitro experiments and, hopefully, in-vivo work as well.

## Methods

The criteria for the consideration of the various papers for this study, and hence, the inclusion of the data that they contained were two-fold: (1) The papers had to focus on the characterization of breast tissue with Raman spectroscopy, and (2) the spectra provided within these papers included the spectral range of 500–1200  $\text{cm}^{-1}$ , which constitutes the characteristic region for aromatic amino acid vibrational modes. After all the papers that satisfied these criteria were collected, the relevant spectra from each paper were extracted. Each spectrum was then uploaded into Photoshop and a baseline was established. Once the baseline was determined, the intensities of the peaks corresponding to aromatic amino acids at 503, 657, 816, 837, 845, 965, 1006, 1012 and 1131  $\text{cm}^{-1}$  were measured together with the highest intensity peak in the specific spectrum. Each peak was then normalized against the corresponding highest intensity band of the specific spectrum. Each such data point was then classified according to the type of breast tissue or breast cell line, i.e. healthy or cancerous, the frequency of the incident laser and the type of Raman technique.

It is important to note that the band processing protocol that we have used, while intending to minimize variance among the multitude of extracted spectra, nevertheless carries an associated possible error in the range of  $\pm 6 \text{ cm}^{-1}$ . This is due to several reasons: (1) The experimental resolution of the source Raman spectra, (2) The resolution of the reproduction of the spectra within the publication, and (3) The resolution of the digitization process and frequency assignment. The implication of this error range is that it could potentially cause an

overlap between the Raman band of phenylalanine at 1006  $\text{cm}^{-1}$  and the weak Raman band of tryptophan peak at 1012  $\text{cm}^{-1}$ , resulting in a more intense tryptophan Raman peak. However, since the Raman band of tryptophan peak at 1012  $\text{cm}^{-1}$  is rather weak, this did neither pose a problem nor did it affect the analysis.

Of the several laser frequencies that have been used as the Raman excitation in the various papers that we have examined, we chose to concentrate on the results obtained with the 785 nm excitation laser. Raman spectra of biological samples tend to exhibit strong background due to tissue autofluorescence, and hence, can often lead to the attenuation of the weak Raman scattering signal. Among the several strategies to reduce tissue autofluorescence and improve the signal-to-noise ratio (SNR), the use of excitation lasers with wavelengths in the near-infrared range<sup>50–53</sup> has proven effective. However, while longer excitation wavelengths effectively reduce tissue autofluorescence background, it also decreases the Raman signal intensity. Hence, the use of the 785 nm excitation laser provides an optimized balance between the reduction of background autofluorescence and band intensity.

## Data availability

All data generated or analyzed during this study are included in this published article.

Received: 20 August 2020; Accepted: 5 January 2021

Published online: 18 January 2021

## References

- Zhang, W. *et al.* Electronic structure of aromatic amino acids studied by soft X-ray spectroscopy. *J. Chem. Phys.* **131**, 035103 (2009).
- Matta, C. F. & Hernández-Trujillo, J. Bonding in polycyclic aromatic hydrocarbons in terms of the electron density and of electron delocalization. *J. Phys. Chem. A* **107**(38), 7496–7504 (2003).
- Scheiner, S., Kar, T. & Pattanayak, J. Comparison of various types of hydrogen bonds involving aromatic amino acids. *J. Am. Chem. Soc.* **124**(44), 13257–13264 (2002).
- Dougherty, D. A. Cation- $\pi$  interactions involving aromatic amino acids. *J. Nutr.* **137**(6), 1504S–1508S (2007).
- Dougherty, D. A. The cation- $\pi$  interaction. *Acc. Chem. Res.* **46**(4), 885–893 (2013).
- Chelli, R., Gervasio, F. L., Procacci, P. & Schettino, V. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J. Am. Chem. Soc.* **124**(21), 6133–6143 (2002).
- McGaughey, G. B., Gagné, M. & Rappé, A. K.  $\pi$ -Stacking interactions alive and well in proteins. *J. Biol. Chem.* **273**(25), 15458–15463 (1998).
- Yajima, T. *et al.*  $\pi$ - $\pi$  stacking assisted binding of aromatic amino acids by copper(II)-aromatic diimine complexes. Effects of ring substituents on ternary complex stability. *Dalton Trans.* **3**, 299–307 (2007).
- Kelkar, D. A. & Chattopadhyay, A. Membrane interfacial localization of aromatic amino acids and membrane protein function. *J. Biosci.* **31**(3), 297–302 (2006).
- Brocchieri, L. & Karlin, S. Geometry of interplanar residue contacts in protein structures. *Proc. Natl. Acad. Sci.* **91**(20), 9297–9301 (1994).
- Burley, S. K. & Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **203**(2), 139–143 (1986).
- Meyer, E. A., Castellano, R. K. & Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Intl. Ed. Engl.* **42**(11), 1210–1250 (2003).
- Espinoza-Fonseca, L. M. Aromatic residues link binding and function of intrinsically disordered proteins. *Mol. Biosyst.* **8**, 237–246 (2012).
- Aravinda, S. *et al.* Aromatic-aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues. *J. Am. Chem. Soc.* **125**(18), 5308–5315 (2003).
- Baker, C. M. & Grant, G. H. Role of aromatic amino acids in protein-nucleic acid recognition. *Biopolymers* **85**(5–6), 456–470 (2007).
- Yarus, M., Widmann, J. J. & Knight, R. RNA-amino acid binding: a stereochemical era for the genetic code. *J. Mol. Evol.* **69**, 406 (2009).
- Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucl. Acids Res.* **29**(13), 2860–2874 (2001).
- Makwana, K. M. & Mahalakshmi, R. Implications of aromatic-aromatic interactions: from protein structures to peptide models. *Protein Sci.* **24**(12), 1920–1933 (2015).
- Krzyściak, W. Activity of selected aromatic amino acids in biological systems. *Acta Biochim. Polon.* **58**(4), 461–466 (2011).
- McDermott, E. W., Barron, E. T., Smyth, P. P. & O'Higgins, N. J. Pre-morphological metabolic changes in human breast carcinogenesis. *Brit. J. Surg.* **77**, 1179–1182 (1990).
- Deng, K. *et al.* High levels of aromatic amino acids in gastric juice during the early stages of gastric cancer progression. *PLoS ONE* **7**(11), e49434 (2012).
- Deng, K. *et al.* Three aromatic amino acids in gastric juice as potential biomarkers for gastric malignancies. *Analyt. Chim. Acta* **694**(1–2), 100–107 (2011).
- Platten, M., Nollen, E. A. A., Röhrig, U. F., Fallarino, F. & Opitz, C. A. Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat. Rev. Drug Discov.* **18**, 379–401 (2019).
- Juhász, C. *et al.* Tryptophan metabolism in breast cancers: molecular imaging and immunohistochemistry studies. *Nucl. Med. Biol.* **39**(7), 926–932 (2012).
- El Refaey, M. *et al.* Oxidation of the aromatic amino acids tryptophan and tyrosine disrupts their anabolic effects on bone marrow mesenchymal stem cells. *Mol. Cell. Endocrinol.* **410**, 87–96 (2015).
- Puccetti, P. *et al.* Accumulation of an endogenous tryptophan-derived metabolite in colorectal and breast cancers. *PLoS ONE* **10**(4), e0122046 (2015).
- Poliaková, M., Aebersold, D. M., Zimmer, Y. & Medová, M. The relevance of tyrosine kinase inhibitors for global metabolic pathways in cancer. *Mol. Cancer* **17**, 27 (2018).
- Sikalidis, A. K. Amino acids and immune response: a role for cysteine, glutamine, phenylalanine, tryptophan and arginine in T-cell function and cancer?. *Pathol. Oncol. Res.* **21**, 9–17 (2015).
- Darienzo, R. E. *et al.* Surface-enhanced Raman spectroscopy characterization of breast cell phenotypes: effect of nanoparticle geometry. *ACS Appl. Nano Mater.* **2**(11), 6960–6970 (2019).
- Bocklitz, T. *et al.* A comprehensive study of classification methods for medical diagnosis. *J. Raman Spectr.* **40**, 1759–1765 (2009).
- Hartmann, K. *et al.* A study of Docetaxel-induced effects in MCF-7 cells by means of Raman microspectroscopy. *Anal. Bioanalyt. Chem.* **403**(3), 745–753 (2012).

32. Jung, G. B. *et al.* Anti-cancer effect of bee venom on human MDA-MB-231 breast cancer cells using Raman spectroscopy. *Biomed. Opt. Express* **9**(11), 57075718 (2018).
33. Gebrekidan, M. T. *et al.* Breast tumor analysis using shifted-excitation Raman difference spectroscopy (SERDS). *Technol. Cancer Res. Treat.* **17**, 1–11 (2018).
34. Wu, Y. *et al.* BRMS1 expression alters the ultrastructural, biomechanical and biochemical properties of MDA-MB-435 human breast carcinoma cells: an AFM and Raman microspectroscopy study. *Cancer Lett.* **293**(1), 82–91 (2010).
35. Keller, M. D. *et al.* Development of a spatially offset Raman spectroscopy probe for breast tumor surgical margin evaluation. *J. Biomed. Opt.* **16**(70), 077006 (2011).
36. Chaturvedi, D. *et al.* Different phases of breast cancer cells: Raman study of immortalized, transformed, and invasive cells. *Biosensors* **6**(4), 57 (2016).
37. Goel, P. N., Singh, S. P., Krishna, C. M. & Gude, R. P. Investigating the effects of pentoxifylline on human breast cancer cells using Raman spectroscopy. *J. Innov. Opt. Health Sci.* **8**(2), 1550008 (2015).
38. Krishna, C. M. *et al.* Micro-Raman spectroscopy of mixed cancer cell populations. *Vibr. Spectr.* **38**(1–2), 95–100 (2005).
39. Zoladek, A., Pascut, F. C., Patel, P. & Notingher, I. Non-invasive time-course imaging of apoptotic cells by confocal Raman microspectroscopy. *J. Raman Spectr.* **42**(3), 251–258 (2011).
40. Aljakouch, K. *et al.* Raman microspectroscopic evidence for the metabolism of a tyrosine kinase inhibitor, neratinib, in cancer cells. *Angew. Chem.* **57**(24), 7250–7254 (2018).
41. Medeiros, P. S. C., de Carvalho, A. L. M. B., Ruano, C., Otero, J. C. & Marques, M. P. M. Raman microspectroscopy for probing the impact of a dietary antioxidant on human breast cancer cells. *Food Funct.* **7**(6), 2800–2810 (2016).
42. McEwen, G. D. *et al.* Subcellular spectroscopic markers, topography and nanomechanics of human lung cancer and breast cancer cells examined by combined confocal Raman microspectroscopy and atomic force microscopy. *Analyst* **138**(3), 287–297 (2013).
43. Hu, C. *et al.* Raman spectra exploring breast tissues: comparison of principal component analysis and support vector machine-recursive feature elimination. *Med. Phys.* **40**(6), 063501 (2013).
44. Zúñiga, W. C. *et al.* Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy. *Sci. Rep.* **9**, 14639 (2019).
45. Krishan, C. M. *et al.* Raman spectroscopy of breast tissues. *Exp. Rev. Mol. Diagn.* **8**(2), 149–166 (2008).
46. Fenn, M. B., Pappu, V., Georgeiv, P. G. & Pardalos, P. M. Raman spectroscopy utilizing fisher-based feature selection combined with support vector machines for the characterization of breast cell lines. *J. Raman Spectr.* **44**(7), 939–948 (2013).
47. Hossain, M. K., Cho, H. Y., Kim, K. J. & Choi, J. W. Silver nanostar patterned substrate for label-free characterization of breast cancer cells based on surface-enhanced Raman spectroscopy. *Sci. Adv. Mater.* **6**(11), 2491–2495 (2014).
48. Yuan, H., Fales, A. M., Khoury, C. G., Liu, J. & Vo-Dinh, T. Spectral characterization and intracellular detection of surface-enhanced Raman scattering (SERS)-encoded plasmonic gold nanostars. *J. Raman Spectr.* **44**(2), 234–239 (2013).
49. Kong, K. *et al.* Towards intra-operative diagnosis of tumors during breast conserving surgery by selective-sampling Raman microspectroscopy. *Phys. Med. Biol.* **59**(20), 6141–6152 (2014).
50. Crow, P. *et al.* The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines. *Brit. J. Cancer* **92**, 2166–2170 (2005).
51. Kaminaka, S., Yamazaki, H., Ito, T., Kohda, E. & Hamaguchi, H. O. Near-infrared Raman spectroscopy of human lung tissues: possibility of molecular-level cancer diagnosis. *J. Raman Spectr.* **32**(2), 139–141 (2001).
52. Kawabata, T. *et al.* Optical diagnosis of gastric cancer using near-infrared multichannel Raman spectroscopy with a 1064-nm excitation wavelength. *J. Gastroenterol.* **43**, 283–290 (2008).
53. Min, Y. K., Yamamoto, T., Kohda, E., Ito, T. & Hamaguchi, H. 1064 nm near-infrared multichannel Raman spectroscopy of fresh human lung tissues. *J. Raman Spectr.* **36**(1), 73–76 (2005).
54. Painter, P. C. & Koenig, J. L. A normal vibrational analysis of benzene. *Spectrochim. Acta A Mol. Spectr.* **33**(11), 1019–1024 (1977).
55. Wilson, E. B. The normal modes and frequencies of vibration of the regular plane hexagon model of the benzene molecule. *Phys. Rev.* **45**, 706–714 (1934).
56. Hernández, B., Pflüger, F., Adenier, A., Kruglik, S. G. & Ghomi, M. 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**(46), 15319–15330 (2010).
57. Grace, L. I., Cohen, R., Dunn, T. M., Lubman, D. M. & de Vries, M. S. The R2PI Spectroscopy of tyrosine: a vibronic analysis. *J. Mol. Spectr.* **215**(2), 204–219 (2002).
58. Wei, F., Zhang, D., Halas, N. J. & Hartgerink, J. D. Aromatic amino acids providing characteristic motifs in the Raman and SERS spectroscopy of peptides. *J. Phys. Chem. B* **112**(30), 9158–9164 (2008).
59. Hartsuiker, L., Zeijen, N. J. L., Terstappen, L. W. M. M. & Otto, C. A comparison of breast cancer tumor cells with varying expression of the Her2/neu receptor by Raman microspectroscopic imaging. *Analyst* **135**(12), 3220–3226 (2010).
60. Aertgeerts, K. *et al.* Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. *J. Biol. Chem.* **286**(21), 18756–18765 (2011).
61. Bi, X., Rexer, B., Arteaga, C. L., Guo, M. & Mahadevan-Jansen, A. Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy. *J. Biomed. Opt.* **19**(2), 025001 (2014).
62. Lyng, F. M. *et al.* Discrimination of breast cancer from benign tumours using Raman spectroscopy. *PLoS ONE* **14**(2), e0212376 (2019).
63. Téllez-Plancarte, A. *et al.* Development of a nanostructured platform for identifying HER2-heterogeneity of breast cancer cell by surface-enhanced Raman scattering. *Nanomaterials* **8**(7), 549 (2018).
64. Lee, H. I., Kim, M. S. & Suh, S. W. Raman spectroscopy of L-phenylalanine, L-tyrosine, and their peptides adsorbed on silver surface. *Bull. Korean Chem. Soc.* **9**(4), 218–223 (1988).
65. Lyng, F. M. *et al.* Discrimination of breast cancer from benign tumors using Raman spectroscopy. *PLoS ONE* **14**(2), e0212376 (2019).
66. Rygula, A. *et al.* Raman spectroscopy of proteins: a review. *J. Raman Spectr.* **44**(8), 1061–1076 (2013).
67. Siddhanta, S. & Narayana, C. Surface enhanced Raman spectroscopy of proteins: implications for drug designing. *Nanomat. Nanotechnol.* **2**(1), 1–13 (2012).
68. Witton, C. J., Reeves, J. R., Going, J. J., Cooke, T. G. & Bartlett, J. M. S. Expression of the HER1-4 family of receptor tyrosine kinases in breast cancer. *J. Pathol.* **200**(3), 290–297 (2003).

### Author contributions

Conceptualization and study design: R.T. Data collection: S.C., R.T., R.E.D. Data analysis and interpretation: S.C., R.T. Manuscript writing and editing: S.C., R.T., R.E.D.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to R.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021