Small Angle X-ray Scattering as a Diagnostic Tool for Breast Cancer

Sabeena Sidhu
BSc, MSc

School of Physics, Monash University

February 12, 2009

ERRATA

p 20 para 3, last sentence. “it gives some indication of the amount of collagen present in the sample, and how well the fibrils are distributed and ordered within the sample related to the dimensions of the peak (Table 1.1).” for “it describes the degree of long range order and distribution of the fibrils in the sample, as well as gives some indication of the amount of collagen in the sample.”

p 22 para 2, first sentence. “The intensity of the amorphous scatter from malignant tissues has been shown to be higher than the intensity of the amorphous scatter from normal tissues [20, 44, 49, 51, 68, 77, 101, 102].” for “The amorphous scatter from malignant tissues has been shown to be higher than the amorphous scatter from normal tissues [16, 41, 43, 57, 68, 71, 87, 90, 91, 105].”

p 31 para 2, after 5th sentence. Additional sentence: “Mammoplasty procedures were mainly performed as breast reductions for patients presenting with back pain or other symptoms that affect their quality of life due to a large breast mass.”

p 34 para 2, end of paragraph. Additional sentence: “Tissue samples were kept in storage for up to 2 years at -80° C, before being defrosted and imaged during one of the experimental sessions.”

p 72 para 2, after 1st sentence. Additional sentence: “All of the post-parameterized data, such as the parameters listed in Table 1.1, were utilised for PCA.”

p 88 para 1, 11th line. “Figure 3.17” for “Figure 1.10”.

p 108 2nd paragraph. “Paper II: Medical Physics [112] Paper III: Cancer Research [113]” for “Paper II: Cancer Research [113] Paper III: Medical Physics [112]”.

<u>Correction of References:</u>

p 2 line 6-7. [20, 41, 44, 49, 50, 68, 77, 101, 102, 128] for [34, 38, 41, 43, 68, 75, 103, 105, 124].

p 2 line 19. [25] for [24].

p 2 line 21. [77, 102] for [75, 105].

p 2 line 29-30. [15, 44, 49, 68, 77, 91, 103] for [14, 41, 43, 68, 75, 90, 103, 105].

p 3 line 2. [85, 97, 98] for [85, 96, 98].

p 5 end of caption. [12, 28, 107] for [130].

p 7 end of 1st paragraph in section 1.2.2. [34, 39, 40] for [30, 38, 39].

p 8 end of caption in Figure 1.4. Add [102].

p 9 line 19. [35, 63] for [10].

p 10 line 6. [36, 67, 87] for [56, 58, 88].

p 10 line 8. [30, 47, 67, 94, 110] for [29, 46, 63, 67, 95, 128].

p 11 line 3. [96, 97] for [97, 98].

p 11 line 4. [98] for [99].

p 14 1st line in Section 1.3. [44, 50, 51, 77, 101, 103, 127-129] for [44, 51, 55, 77, 101, 103, 127-129]

p 14 4th line in Section 1.3. "…tendon [4, 36, 55], bone [50, 106, 129], and corneal tissue [30, 57, 82]." for "…tendon [11, 48, 49, 123], bone [50, 104, 129], and corneal tissue [52, 53, 98, 99]."

p 16 1st column and 2nd row of Table 1.2. [49, 51] for [41, 43].

p 18 line 15 in reference []. 68 for 67.

p 21 1st sentence in 2nd paragraph of Section 1.3.2. [49, 51] for [41, 43].

p 22 1st sentence in 2nd paragraph of Section 1.3.3. [20, 44, 49, 51, 44, 68, 77, 101, 102] for [16, 41, 43, 57, 68, 71, 87, 90, 91, 105].

p 22 end of 3rd sentence in section 1.3.4. [121] for [131].

p 29 2nd sentence of 1st paragraph. [18, 20, 44, 49-51, 77, 91, 102, 119, 128] for 16, 34, 38, 41-43, 75, 86, 103, 105, 124].

p 61 line 3. [35, 36] for [28, 29].

p 72 line 7. [27, 29, 43, 74, 81] for [21, 26, 37, 66].

p 75 1st sentence of 2nd paragraph. [20, 44, 49-51, 77, 102] for [16, 38, 41-43, 75, 106].

p 123 last sentence of 1st paragraph. [8, 39, 78, 92, 93, 110, 116] for [6, 39, 84, 88, 89, 110, 116].

p 123 2nd sentence of 2nd paragraph. [36, 62, 68, 106] for [42, 62, 68, 113].

p 123 2nd sentence of 3rd paragraph. [28, 93] for [22, 82].

p 159 1st sentence. [20, 44, 49, 51, 77, 102] for [16, 38, 41, 43, 75, 105].

p 165 line 9. [20, 44, 49, 51, 77] for [16, 38, 41, 43, 75].

p 165 line 24. [49-51] for [41-43].

p 172 line 20. [20, 51] for [16, 43].

p 173 last line on page. [1, 32, 93] for [1, 44, 102].

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other institution. To the best of my knowledge, the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Sabeena Sidhu:_____ Date: _____

I dedicate this research to all of the women who participated in this study.
Despite the difficulties of their illness, they still made the effort to help others.

Abstract

Breast cancer is the most common cause of cancer death in Australian women. Current pathological analysis examines a small section of tissue for cellular and plasma abnormalities using a light microscope. However, this method of diagnosis, despite being the current gold standard, has its limitations, where human error and professional experience can influence a patient's diagnosis. A potential alternative or adjunct to conventional histopathology for classifying tissue disease status is offered by Small Angle X-ray Scattering (SAXS).

At the time of commencement of this work, there had been several small scale studies which examined the potential of SAXS to classify the disease status of breast tissue. These tended to focus on the supramolecular structure of collagen fibrils found in the breast, where it is known that the degradation of these fibres is related to the spread of disease. Most previous studies also used a synchrotron as an X-ray source, due to the intense and highly collimated flux available. This study used a synchrotron source, but also evaluated the use of a laboratory X-ray source, as a more convenient and relatively inexpensive alternative that could one day find application in the clinic. The work presented in this thesis analyses the largest cohort of patients and breast tissue samples studied to date using SAXS: 130 patients with 543 tissue samples. Tissues were sourced from surgical waste and classified into four groups: invasive carcinoma, benign, normal, and mammoplasty. Mammoplasty tissue samples were harvested from patients undergoing breast reduction and/or reconstruction, where no history or presence of disease was indicated. Normal tissue was sampled from patients with known disease, but pathological analysis of the tissue core diagnosed it as normal. A comprehensive analysis of the scattering patterns was carried out, analysing features arising from the collagen structure and orientation, the total scattered intensity, and adipose tissue in the breast. Features related to the axial D-spacing of the collagen fibrils within the breast tissue as well as the integrated scattering intensity (called amorphous scatter) demonstrated the highest ability to discriminate tissue types, in SAXS images acquired from both the synchrotron source and the laboratory X-ray source.

The amorphous scatter intensities obtained using a synchrotron source showed highly significant differences ($p < 0.01$) for almost all of the tissue pair comparisons: invasive carcinoma vs. benign, invasive carcinoma vs. normal, invasive carcinoma vs. mammoplasty, benign vs. mammoplasty, and normal vs.

mammoplasty. However, no significant difference was seen in the amorphous scatter between benign versus normal tissues ($p = 0.30$). The amorphous scatter values increased with severity of disease, i.e. it was the highest for invaded tissues and decreased progressively from benign to normal to mammoplasty. There was a significant difference between normal and mammoplasty tissue types using the amorphous scatter as a discriminator ($p = 0.0025$). Pathological assessment cannot differentiate between these two tissue types, which suggests that there may be changes occurring in these tissue structures at the supramolecular level that can be characterised using SAXS.

The ability of SAXS to reveal structural differences between normal and mammoplasty tissue types is highly significant, for both disease diagnosis and treatment, as well as for understanding disease progression. For example, these differences might aid in determining surgical margin clearance of excised breast lesions as well as potentially provide a means of pre-screening or perhaps improve false-negative rates of diagnosis. The potential of SAXS to reveal macroscopic extent and directional spread of disease was explored using two-dimensional mapping of the amorphous scatter. These maps showed broad agreement with histopathological diagnosis, but further investigation regarding their reliability and interpretation for clinical utility is still needed. Changes in both the amorphous scatter and the axial D-spacing were seen in tissue samples up to 6 cm away from the primary site of disease. In particular, a significant decrease in both parameters was seen between the centre of the tumour (at 0 cm) and 2 cm away, suggesting that closer examination of the tissue structures over the disease/healthy tissue border may provide information regarding the mechanisms of metastasis and growth of cancerous tumours.

The combination of the amorphous scattering results from the two X-ray sources indicates that the size of the scatterers may be the key in classifying tissue types. The synchrotron source was able to access a lower q-range ($q = 0.1\text{-}0.6\ nm^{-1}$) and the laboratory source covered a larger q-range ($q = 0.25\text{-}2.3\ nm^{-1}$). Mammoplasty tissues appear to be characterised by large scattering components ($d > 25.13$ nm), whereas normal tissues are characterised by slightly smaller scattering components ($10.47\ nm < d < 25.13\ nm$) and benign tissues by even smaller scattering components ($4.83 < d < 10.47$ nm). It appears that the size of the scatterers contributing to the total scattering intensity decreases with severity of disease, which was seen independently with both X-ray sources. Further investigation is warranted to determine the biological origin of these differences. These results also suggest that the optimum SAXS instrument may need to cover a scattering vector range of $q < 0.25\ nm^{-1}$ to identify differences in healthy tissue types, and $q > 2.3\ nm^{-1}$ to possibly

investigate invasive carcinoma tissue types. A SAXS apparatus that can examine a large q-range may provide all of the necessary information from the amorphous scatter to differentiate between tissue groups.

The periodic structure of collagen fibrils along their longitudinal axis can be characterised by the axial D-spacing, where this spacing was found to change with the presence of disease. The axial D-spacing for healthy breast tissues was found to be significantly lower in normal and mammoplasty tissues compared to invaded tissues ($p = 0.0050$ and $p = 0.0093$, respectively). However, no significant differences between the other tissue group pairs were seen ($p > 0.05$). These differences were evident in classification modelling of the four tissue groups, where the amorphous scatter and the amplitude of a collagen axial peak were used to build a probability model for disease status. The model showed high sensitivities (> 70%) and widely variable specificities (ranged from 18-97%) for the data examined with the synchrotron source. This means that the model was a good indicator of disease, but poor at indentifying healthy tissue types.

The work presented in this thesis shows that SAXS is capable of distinguishing breast tissue types with high sensitivity and has the potential to become a significant tool for the investigation of cancer progression or even diagnosis. Further investigation into the amorphous scatter and axial D-spacing in particular may provide insight into the biological mechanisms related to tissue degradation associated with invasive disease.

Acknowledgements

This thesis could not have been completed without the generous support and encouragement from a multitude of people from several disciplinary fields. I would like to extend my thanks to my main supervisor, Prof. Rob Lewis, for taking me on as a doctoral candidate and giving me the opportunity to further my research career. I owe a particular debt of gratitude to my associate supervisor, Dr. Karen Siu, whose patience and encouragement never wavered during my candidature. Without Dr. Siu's experience, intellect, and perseverance, the investigation would have floundered.

I am most grateful and appreciative to Dr. Greg Falzon who assisted with the statistical analysis presented in this thesis and to Simon Nazaretian who provided all of the histopathological diagnoses of the 543 tissue samples examined in this study. These two individuals ensured that the work presented in this thesis was based upon a strong statistical and clinical foundation.

I would like to express my thanks to the staff at Southern Health Medical, especially the nurses, surgeons, and pathologists who shared my passion and commitment for obtaining the largest cohort of patients possible. In particular, the support and advice provided by Mr. Stewart Hart and Ms. Jane Fox will not be forgotten.

Special thanks must be made to Drs. Chris Hall, David Paganin, and Adam Round for providing a great deal of assistance and guidance during the investigation. Also to Drs. Steven Mudie and Andrew Stevenson who provided so much assistance with the laboratory source apparatus at CSIRO. Drs. Sarah Pearson, Gunter Grossmann, and Tina Geraki provided assistance with the synchrotron experiments at the SRS Daresbury, and their help was much appreciated. Without the "drop everything" attitude of the Electronics and Mechanical Workshop with the School of Physics, Monash University, the beam stop and motorized stage used in the experiments would have been in ruin. I would also like to thank Dr. John McDougall from the Monash Centre for Synchrotron Science for keeping me in line and telling me to "get over it".

I would like to thank the numerous fellow students within the School of Physics at Monash University with whom I shared post-graduate life. In particular, my office mates Dr. Marcus Kitchen, Dr. Ivan Williams, and Jeff Crosbie who kept things in perspective. Coffee and lunchtime conversations were never dull with the company of Dr. John Gillam. A special thanks to Associate Professor Michael

Morgan, Jean Pettigrew, Julia Barnes, Tracey Lawrence, James Gibbons, and Antonio Benci for their administrative and technical assistance as well as their compassionate ears. For the intellectual break afforded by dealing with first-year physics students I thank Dr. Susan Feteris, Jenny McCabe, and Dr. David Mills.

I would like to thank the Monash Centre for Synchrotron Science for scholarship support for both tuition and fees and for providing me with a decent way of life. I am very appreciative of the funding for travel to the SRS Daresbury provided by the Australian Government's Access to Major Research Facilities Programme.

To my family and friends who kept me sane during the past three and a half years. To my Australian clone, Bree Polkinghorne, and her family who made my life so much fun and opened my eyes to the beauty of Melbourne. To the "doctor girls": Thanu Shekar, Fiona Gutteridge, Tamara Jones, Vivian Lee, Cheryl Choong, Mihiri Weerasooriya, and Tracy Shang who are a group of friends that people search a lifetime for, and who also made me so glad that I stayed in physics. I would like to give a special and heart-felt thanks to Toby Beveridge who was there for me for all of the good times and bad times and showed me how wonderful life could be. Trudy and David Beveridge, as well as Sam and Ben Beveridge also deserve special thanks for being my surrogate family in Melbourne. My brother and his wife, Jason and Jessica Sidhu, as well as to my sister Natasha Sidhu who always knew I could do it. And most importantly, my most grateful thanks to my father and mother, Baldev and Chindow Sidhu, whose love and support during my entire educational career will never be forgotten.

Contents

ABSTRACT ..... **I**

ACKNOWLEDGEMENTS ..... **V**

CONTENTS ..... **VII**

LIST OF FIGURES ..... **XI**

LIST OF TABLES ..... **XV**

LIST OF TABLES ..... **XV**

GLOSSARY ..... **XIX**

ACRONYMS ..... **XXIII**

CHAPTER 1 ..... **1**

INTRODUCTION 1
1.1 Project Aims *1*
1.2 Background *3*
1.2.1 Features of Breast Cancer 3
1.2.2 Collagen 7
1.2.3 Collagen and Breast Cancer 9
1.2.4 Small Angle X-ray Scattering (SAXS) 11
1.3 SAXS and Breast Tissue *14*
1.3.1 Axial (Meridional) Peak Features 17
1.3.2 Equatorial Peak Features 21
1.3.3 Amorphous Scatter 22
1.3.4 Fat (Adipose) Tissue 22
1.3.5 Collagen Fibril Orientation 24
1.4 Motivation and Contribution to Knowledge *25*
1.5 Outline of the Thesis *27*

CHAPTER 2 ..... **29**

BREAST TISSUE SAMPLES 29
2.1 Patient Recruitment and Sample Collection *29*
2.1.1 Ethics Approval and Patient Consent 33
2.2 Tissue Harvesting and Storage *34*
2.3 Collection of SAXS Tissue Samples *36*
2.4 Tissue Handling During an Experiment *39*
2.5 Histopathological Reviews *39*
2.6 Conclusion *44*

CHAPTER 3 ..... **47**

SMALL ANGLE X-RAY SCATTERING: EXPERIMENTAL METHODOLOGY 47
3.1 Experimental Set-up: Synchrotron Source *48*
3.2 Experimental Set-Up: Laboratory SAXS Camera *52*
3.3 Image Acquisition and Pre-Processing *59*
3.3.1 Removal of wire structure 60
3.3.2 Calibration of q-spacing 60
3.3.3 Normalisation of the Intensity 65
3.3.4 Radial Integration 67
3.4 Methods of Analysis *70*

CHAPTER 4 ..... **75**

ANALYSIS OF PRIMARY SITE TISSUE SAMPLES USING A SYNCHROTRON SOURCE 75
4.1 Introduction *75*
4.2 Determining Useful SAXS Parameters *77*
4.2.1 Axial (Longitudinal) Variables 77
4.2.2 Equatorial (Lateral) Variables 83

4.2.3 Total Scattering Intensity Variables 86
4.2.4 Polar Angle Distribution Data 88
4.3 Classification of Tissue Groups 89
4.4 Discussion 96
4.4.1 Identifying Optimal Parameters for Classification 96
4.4.2 Differences between Experiment #3 Analysis and a Combined Experimental Analysis101
4.5 Conclusions 103

CHAPTER 5 107

ANALYSIS OF SERIES SITE TISSUE SAMPLES USING A SYNCHROTRON SOURCE 107
5.1 Introduction 107
5.2 Methods and Materials 108
5.2.1 Tissue samples 108
5.2.2 SAXS Data Analysis 109
5.3 Systematic Changes with Distance 111
5.4 Mapping Parameter Changes in Individual Patients 115
5.5 Discussion 122
5.6 Conclusions 127

CHAPTER 6 129

LABORATORY SOURCE DATA ANALYSIS 129
6.1 Introduction 129
6.2 Methods and Materials 131
6.2.1 Tissue Samples 131
6.2.2 Methods 133
6.3 Feature Analysis 134
6.3.1 Amorphous Scatter 134
6.3.2 Fifth Order Axial Collagen Peak 135
6.3.3 Fat (Adipose) Tissue 136
6.4 Classification of Tissue Samples 138
6.4.1 Probability Models and ROC curves 138
6.4.2 Discriminate Variables 142
6.5 Comparing Amorphous Scatter from Laboratory Source Data with Amorphous Scatter from Synchrotron Source Data 145
6.5.1 Summary of Amorphous Scatter: Laboratory Source versus Synchrotron Source 147
6.6 Series Tissue Analysis: 4 mastectomy patients 148
6.7 Two-Dimensional Mapping of Four Patients Using the Laboratory Source 151
6.8 Discussion 154

CHAPTER 7 159

SUMMARY AND CONCLUSIONS 159
7.1 Summary of Findings 159
7.2 Discussion of the Main Findings 161
7.3 Improvements and Future Work 170
7.4 Conclusions 174

APPENDIX A 177

A.1 Participant Information Form 177
A.2 Informed Consent Form 181
A.3 Ethics Approval Extensions 183

APPENDIX B 187

HISTOPATHOLOGY 187
B.1 Histopathological Features for Diagnosing Breast Disease 187
B.2 Histopathologic Grading 192

APPENDIX C 195

FUJIFILM IMAGING PLATE (IP) DETECTORS 195
C.1 Characteristics 196

APPENDIX D 201

DESIGN OF SEMI-TRANSPARENT BEAM-STOP FOR LABORATORY SAXS CAMERA 201
D.1: Determination of the Diameter of the Beam *201*
D.2: Calculating the Appropriate Thickness of the Semi-Transparent Beam-Stop *203*

APPENDIX E 207

STATISTICS 207
E.1 Student t-test and p-values *207*
E.2 Analysis of Variance (ANOVA) *208*
E.3 Bartlett's Test *211*
E.4 Flinger (Killeen) Median Test *212*
E.5 Two-Sample Kolmogorov-Smirnov Test *213*
E.6 Shapiro-Wilk Test for Normality *214*
E.7 Principal Component Analysis (PCA) *216*
E.8 Receiver Operating Characteristic (ROC) *218*
E.9 Quadratic Discriminant Analysis (QDA) *219*
E.10 Bivariate Gaussian Kernel Estimation *220*

REFERENCES 225

SUPPORTING DOCUMENTS 237

PAPER I: X-RAY SCATTERING FOR CLASSIFYING TISSUE TYPES ASSOCIATED WITH BREAST DISEASE 238
PAPER II: MAPPING OF STRUCTURAL CHANGES IN BREAST TISSUE DISEASE USING X-RAY SCATTERING 249
PAPER III: MAPPING OF STRUCTURAL CHANGES IN BREAST TISSUE DISEASE USING X-RAY SCATTERING 261

List of Figures

FIGURE 1.1: CROSS-SECTION OF A BREAST SHOWING THE BASIC ANATOMY. 4
FIGURE 1.2: NORMAL AND CANCEROUS BREAST MORPHOLOGY 5
FIGURE 1.3: THE LEVELS OF ARRANGEMENT IN FIBRILLAR COLLAGEN.....8
FIGURE 1.4: THE STAGGERED ARRANGEMENT OF THE COLLAGEN MOLECULES. ..8
FIGURE 1.5: ELECTRON MICROGRAPHS OF COLLAGEN TISSUE IN NORMAL TISSUE AND INVASIVE CARCINOMA TISSUE 10
FIGURE 1.6: SMALL ANGLE X-RAY SCATTERING GEOMETRY.....13
FIGURE 1.7: CONSTRUCTION OF THE SCATTERING VECTOR $\vec{q}$ 13
FIGURE 1.8: MERIDIONAL (AXIAL) AND EQUATORIAL SCATTERING FROM WELL-ORIENTED FIBRILS 17
FIGURE 1.9: PLOT OF Q VERSUS THE INTENSITY*Q^2 OF THE RADIALLY AVERAGED DATA.....19
FIGURE 1.10: SAXS IMAGE TAKEN WITH THE SAXS OSMIC CAMERA.....24

FIGURE 2.1: HISTOGRAM PLOT IN 5 YEAR BINS OF THE PATIENT AGES.....32
FIGURE 2.2: SAMPLING METHOD FROM A MASTECTOMY BREAST MASS (TYPICAL PATIENTS.....35
FIGURE 2.3: SAMPLING METHOD FROM A MASTECTOMY BREAST MASS (NON-TYPICAL).....35
FIGURE 2.4: TISSUE SAMPLE SHOWING GLASS ARTEFACTS 42
FIGURE 2.5: TISSUE SAMPLE ILLUSTRATING FOLDING OF THE TISSUE SAMPLE .42
FIGURE 2.6: EXAMPLES TISSUE SAMPLES AFTER H&E STAINING 43

FIGURE 3.1: SCHEMATIC OF THE SRS DARESBURY EXPERIMENTAL SET-UP AT STATION 2.1.....49
FIGURE 3.2: THE SAMPLE HOLDER USED TO ALIGN THE TISSUE SAMPLES IN GLASS CAPILLARIES 51
FIGURE 3.3: THE RECTANGULAR BEAM STOP FOR THE SYNCHROTRON SOURCE 51
FIGURE 3.4: THE LABORATORY SAXS CAMERA SET-UP.....53
FIGURE 3.5: CONFOCAL MAX-FLUX OPTIC 2-D REFLECTION MONOCHROMATOR SYSTEM, ...54
FIGURE 3.6: THE GRADED D-SPACING MULTILAYER PRINCIPAL OF THE CONFOCAL MAX-FLUX OPTIC 54
FIGURE 3.7: THE LABORATORY SAXS CAMERA SAMPLE CHAMBER.....57
FIGURE 3.8: THE SEMI-TRANSPARENT BEAM STOP USED WITH THE LABORATORY SAXS CAMERA 58
FIGURE 3.9 A&B: THE STEPS TAKEN TO REMOVE THE WIRE STRUCTURE FROM THE SRS DARESBURY SAXS IMAGES 62
FIGURE 3.10 C&D: THE STEPS TAKEN TO REMOVE THE WIRE STRUCTURE FROM THE SRS DARESBURY SAXS IMAGES. 63
FIGURE 3.11 E: THE STEPS TAKEN TO REMOVE THE WIRE STRUCTURE FROM THE SRS DARESBURY SAXS IMAGES. 64
FIGURE 3.12: SAXS IMAGES OF CALIBRANTS USED RAT TAIL TENDON AND SILVER BEHENATE. 65

FIGURE 3.13: RAT TAIL TENDON IMAGE WITH REGION OF INTEREST OUTLINED.66
FIGURE 3.14: REGION OF INTEREST FOR THE SRS DARESBURY IMAGES AND FOR THE LABORATORY SAXS CAMERA IMAGES.....68
FIGURE 3.15: SECTOR REGIONS OF INTEREST USED TO SPECIFICALLY LOOK AT MERIDIONAL FEATURES FROM THE SAXS IMAGE.....68
FIGURE 3.16: INTENSITY*Q^4 VERSUS Q FOR RAW DATA.....69
FIGURE 3.17: 2D SAXS IMAGE SHOWING THE REGION OF INTEREST AROUND THE THIRD ORDER AXIAL COLLAGEN PEAK CONVERTED TO RTHETA 71
FIGURE 3.18: 1D PLOT OF THE INTEGRATED SECTION BETWEEN Q = 0.265-0.310 NM^{-1} 71

FIGURE 4.1: AVERAGE OF THE AMPLITUDES OF THE THIRD AXIAL PEAK FOR THE FOUR TISSUE GROUPS. 78
FIGURE 4.2: AVERAGE OF THE CENTRES OF THE THIRD AXIAL PEAK FOR THE FOUR TISSUE GROUPS 79
FIGURE 4.3: HOW DIFFRACTION IMAGES WERE TAKEN AT 1 MM INTERVALS THROUGH THE LENGTH OF THE TISSUE CORE 81
FIGURE 4.4: THE SQUARE ROOT OF THE EQUATORIAL PEAK AREA PLOTTED WITH THE EQUATORIAL PEAK AMPLITUDE 85
FIGURE 4.5: AVERAGES OF THE TOTAL AMORPHOUS SCATTER FOR EACH OF THE FOUR TISSUE GROUPS OVER THE OBSERVED Q-RANGE 0.1-0.6 NM^{-1}.....87
FIGURE 4.6: THE AMPLITUDE OF THE THIRD AXIAL PEAK IN RELATION TO THE AMORPHOUS SCATTER FOR THE FOUR TISSUE TYPES.....91
FIGURE 4.7: THE D-SPACING OF THE THIRD ORDER AXIAL PEAK IN RELATION TO THE AMORPHOUS SCATTER FOR THE FOUR TISSUE TYPES.....91
FIGURE 4.8 (A & B): QDA PROBABILITY PLOTS OF THE DATA SHOWN IN FIGURE 4.6.....93
FIGURE 4.9 (C & D): QDA PROBABILITY PLOTS OF THE DATA SHOWN IN FIGURE 4.6.....94
FIGURE 4.10 (E & F): QDA PROBABILITY PLOTS OF THE DATA SHOWN IN FIGURE 4.6.....95
FIGURE 4.11: RECEIVER OPERATOR CHARACTERISTIC (ROC) CURVES 97

FIGURE 5.1: AVERAGE VALUES FOR THE AMORPHOUS SCATTER AND THE THIRD AXIAL PEAK D-SPACING FOR THE CENTRE TISSUE POINTS AND AT THE 2 CM INTERVALS AWAY FROM THE CENTRE 113
FIGURE 5.2: AVERAGE VALUES FOR THE AMORPHOUS SCATTER AT EACH DISTANCE INTERVAL ALONG THE N-AXIS AND THE P-AXIS.....114
FIGURE 5.3: AVERAGE VALUES FOR THE AXIAL D-SPACING AT EACH DISTANCE INTERVAL ALONG THE N-AXIS AND THE P-AXIS 114
FIGURE 5.4: 2D MODELS FOR PATIENT 133, MAPPING THE AMORPHOUS SCATTER AND THE AXIAL D-SPACING.....117
FIGURE 5.5: 2D MODELS FOR PATIENT 35, MAPPING THE AMORPHOUS SCATTER AND THE AXIAL D-SPACING.....118
FIGURE 5.6: 2D MODELS FOR PATIENT 53, MAPPING THE AMORPHOUS SCATTER AND THE AXIAL D-SPACING.....119
FIGURE 5.7: 2D MODELS FOR PATIENT 85 (WLE), MAPPING THE AMORPHOUS SCATTER AND THE AXIAL D-SPACING.....120

FIGURE 6.1: TYPICAL 1D PROFILE EXTRACTED FROM A LABORATORY SOURCE IMAGE 134
FIGURE 6.2: AVERAGE AMORPHOUS SCATTER VALUES CALCULATED IN RANGE 1 FOR THE 66 PATIENTS IMAGED WITH THE LABORATORY SAXS CAMERA. 135
FIGURE 6.3: AVERAGES OF THE FWHM OF THE FAT PEAK CALCULATED USING THE LABORATORY SOURCE FOR THE FOUR TISSUE GROUPS. 138
FIGURE 6.4: THE AMORPHOUS SCATTER FOR RANGE 1 WITH THE FWHM OF THE FAT PEAK FOR THE PRIMARY SITE TISSUE SAMPLES IMAGED WITH THE LABORATORY SOURCE 140
FIGURE 6.5: QDA PROBABILITY PLOTS USING THE AMORPHOUS SCATTER FOR RANGE 1 AND THE FWHM OF THE FAT PEAK.. 141
FIGURE 6.6: DISCRIMINATE VARIABLES 1 AND 2 OF THE 20 TISSUE SAMPLES THAT HAD ALL 16 POSSIBLE PARAMETERS PRESENT IN THEIR SAXS IMAGES 143
FIGURE 6.7: COMPARISON OF EQUIVALENT AMORPHOUS SCATTER RANGES (Q = 0.25-0.6 NM^{-1}) FOR THE 21 MATCHED PATIENTS BETWEEN THE SRS DAREBURY DATA AND THE LABORATORY SAXS CAMERA DATA 144
FIGURE 6.8: SUMMARY OF THE AMORPHOUS SCATTER RANGES THAT SHOWED SIGNIFICANT DIFFERENCES BETWEEN THE TISSUE GROUP PAIRS FOR THE LABORATORY SOURCE AND THE SYNCHROTRON SOURCE. 148
FIGURE 6.9: AVERAGE VALUES OF THE FWHM OF THE FAT PEAK FOR THE FOUR PATIENTS STUDIED WITH DISTANCE AWAY FROM THE CENTRE OF THE PALPABLE LESION 150
FIGURE 6.10: AVERAGE VALUES OF THE AMORPHOUS SCATTER CALCULATED FOR RANGE 1 (Q = 0.25–0.6 NM^{-1}) FOR THE FOUR PATIENTS STUDIED, AGAINST DISTANCE FROM THE CENTRE OF THE PALPABLE LESION. 150
FIGURE 6.11: 2D GAUSSIAN MAPS OF THE AMORPHOUS SCATTER FOR RANGE 1 OF PATIENT 1. 152
FIGURE 6 12: 2D GAUSSIAN MAPS OF THE AMORPHOUS SCATTER FOR RANGE 1 OF PATIENT 2. 152
FIGURE 6.13: 2D GAUSSIAN MAP OF THE AMORPHOUS SCATTER FOR RANGE 1 OF PATIENT 3.. 153
FIGURE 6.14: 2D GAUSSIAN MAP OF THE AMORPHOUS SCATTER FOR RANGE 1 OF PATIENT 4 153

FIGURE B.1: FIBROCYSTIC CHANGE 189
FIGURE B.2: ADENOSIS 189
FIGURE B.3: FIBROADENOMA 190
FIGURE B.4: DUCTAL (OR EPITHELIAL) HYPERPLASIA 190
FIGURE B.5: DUCTAL CARCINOMA *IN SITU* (DCIS) 191
FIGURE B.6: INVASIVE LOBULAR CARCINOMA 191

FIGURE C.1: COMPOSITE STRUCTURE OF THE IMAGING PLATE 195
FIGURE C.2: THE DYNAMIC RANGE OF THE PSL OF THE IMAGING PLATE 196
FIGURE C.3: FADING OF THE IP SIGNALS AS A FUNCTION OF TIME WITH TWO DIFFERENT X-RAY ENERGIES (5.9 AND 59.9 KEV) 199

FIGURE D.1: RADIAL PROFILE OF SPOT IMAGED WITH THE CCD CAMERA.....202
FIGURE D.2: DESIGN SCHEMATIC OF THE SEMI-TRANSPARENT BEAM STOP USED WITH THE LABORATORY SAXS CAMERA. 204
FIGURE D.3: THE AVERAGE PSL WITH VARYING THICKNESS OF ALUMINIUM. 205

FIGURE E.1: 2D MODELS USING THE AMORPHOUS SCATTER FOR PATIENT 18 AND PATIENT 24 222
FIGURE E.2: PADDED 2D MODELS USING THE AMORPHOUS SCATTER FOR PATIENT 18 AND PATIENT 24 223

List of Tables

TABLE 1.1: THE SAXS FEATURES ANALYSED IN THIS STUDY AND THEIR RELATION TO THE STRUCTURAL COMPONENTS IN BREAST TISSUE. 15
TABLE 1.2: SUMMARY OF THE FEATURES STUDIED BY SEVERAL SAXS GROUPS THAT MAY BE USEFUL IN CLASSIFYING DISEASE IN BREAST TISSUES.....16

TABLE 2.1: SAXS BREAST TISSUE STUDIES LISTING THE NUMBER OF PATIENTS AND THE NUMBER OF TISSUE SAMPLES IN EACH STUDY. 30
TABLE 2.2: STATISTICS FOR THE 114 PATIENTS WHO HAD THEIR AGES AVAILABLE FOR RECORDING 33
TABLE 2.3: THE NUMBER OF PATIENTS IMAGED DURING EACH SYNCHROTRON EXPERIMENTAL SESSION AND WITH THE LABORATORY SOURCE 37
TABLE 2.4: THE NUMBER OF SERIES SITE PATIENTS IMAGED DURING EACH SYNCHROTRON EXPERIMENTAL SESSION AND WITH THE LABORATORY SAXS CAMERA. 37

TABLE 3.1: PROPERTIES OF STATION 2.1 AT THE SRS DARESBURY 49
TABLE 3.2: PROPERTIES OF THE SAXS LABORATORY CAMERA BASED ON A 3-PINHOLE DESIGN IN HIGH-FLUX MODE. 53

TABLE 4.1: COMPARISON OF THE AXIAL D-SPACINGS BETWEEN EACH TISSUE GROUP.....78
TABLE 4.2: THE THIRD AXIAL PEAK D-SPACING FOR EACH OF THE POINT IMAGES 81
TABLE 4.3: P-VALUES CALCULATED TO COMPARE THE THIRD ORDER AXIAL D-SPACING OF THE POINT SCAN IMAGES WITHIN EACH PATIENT. 83
TABLE 4.4: THE LINEAR FITS CALCULATED FOR THE DATA SHOWN IN FIGURE 4.3 FOR EACH OF THE TISSUE GROUPS 85
TABLE 4.5: P-VALUES BETWEEN TISSUE GROUPS FOR THE AMORPHOUS SCATTER. 87
TABLE 4.6: ROC CURVE CHARACTERISTICS FOR EACH PAIRING OF TISSUE GROUPS. 92
TABLE 4.7: COMPARISON OF AXIAL D-SPACING BETWEEN THE DATA FROM ALL THREE EXPERIMENTAL SESSIONS AND ONLY EXPERIMENT #3. 103
TABLE 4.8: COMPARISON OF THE SLOPES CALCULATED USING THE BESSEL AMPLITUDE VS √BESSEL AREA BETWEEN THE TWO DATA SETS. 103

TABLE 5.1: BREAKDOWN OF THE NUMBER OF TISSUE SAMPLES AND THEIR DIAGNOSIS AT EACH DISTANCE POINT FOR THE 56 PATIENTS WHO HAD SERIES TISSUE DATA TAKEN 109
TABLE 5.2: P-VALUES FOR THE AMORPHOUS SCATTER AND AXIAL D-SPACING VALUES BETWEEN THE CENTRES OF THE PRIMARY LESIONS AND THE TISSUE SAMPLES TAKEN AT 2 CM INTERVALS OVER BOTH AXES. 111

TABLE 5.3: THE NUMBER OF TISSUE SAMPLES WITH AMORPHOUS SCATTER VALUES DESIGNATED AS INVASIVE CARCINOMA AND BENIGN/NORMAL USING MAPPING DATA. 122
TABLE 5.4: THE NUMBER OF TISSUE SAMPLES WITH AMORPHOUS SCATTER VALUES DESIGNATED AS INVASIVE CARCINOMA AND BENIGN/NORMAL USING MAPPING DATA 122

TABLE 6.1: SUMMARY OF THE TISSUE SAMPLES IMAGED WITH THE LABORATORY SAXS CAMERA FROM 66 PATIENTS. 131
TABLE 6.2: THE NUMBER OF TISSUE SAMPLES AND THEIR DIAGNOSIS FOR THE 66 PATIENTS WHO HAD TISSUE IMAGED WITH BOTH THE SYNCHROTRON SOURCE AND THE LABORATORY SOURCE. 132
TABLE 6.3: COMPARISONS OF P-VALUES FOR THE AMORPHOUS SCATTER IN EACH Q-RANGE FOR THE 66 PRIMARY SITE TISSUE SAMPLES IMAGED WITH THE LABORATORY SOURCE. 136
TABLE 6.4: ROC CURVE CHARACTERISTICS FOR EACH PAIRING OF TISSUE GROUPS WITH THE LABORATORY SAXS CAMERA. 142
TABLE 6.5: COMPARISON OF THE P-VALUES FOR THE AMORPHOUS SCATTER RANGES BETWEEN EACH TISSUE GROUP FOR THE SRS DARESBURY AND THE LABORATORY SAXS CAMERA. 144

TABLE 7.1: THE MEAN AXIAL D-SPACING VALUES WITH STANDARD DEVIATIONS (±SD) FOR 130 PRIMARY SITE TISSUE SAMPLES IMAGED WITH THE SRS DARESBURY 166

TABLE B.1: THE BLOOM AND RICHARDSON GRADING SYSTEM (BRE) FOR BREAST CANCERS. 193
TABLE B.2: THE 45 PATIENTS WHO HAD THEIR CARCINOMA GRADE AVAILABLE FROM THEIR MEDICAL RECORDS 193

TABLE C.1: READOUT SETTINGS USED FOR THE BAS-5000 WHEN READING IPS 198

Glossary

Words in *italic* font have individual entries within the glossary.

A **Adipose Tissue** Medical term referring to fatty tissue. The connective tissue that contains stored fat.
Axilla The armpit.
Axillary Nodal Dissection A surgical procedure in which the *lymph nodes* in the armpit (axillary nodes) are removed and examined to determined if breast disease has spread to those nodes and also to remove any cancerous *lymph nodes*.

B **Basement Membrane (BM)** A thin layer of connective tissue underlying the epithelium of the breast. The extracellular supporting layer of the proteins underlying the epithelia.
Benign Tissue that is neither cancerous nor malignant, but cannot be classified as normal tissue either. A growth that does not spread to other parts of the body. This type of *lesion* usually has defined margins and is confined to one area of the breast.
Biopsy A procedure in which tissue samples are removed from the body for examination of their appearance under a microscope to find out whether *cancer* or other abnormal cells are present. A biopsy is done with a needle or by surgery.

C **Calcification** The process of forming hardened deposits of calcium salts within the breast. The exact composition and structure of calcifications within the breast is unknown, but is based upon the prototype calcium hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$] with extensive ionic substitution, e.g. carbonate ions replacing the hydroxy or phosphate ions. The presence of calcifications with mammography or biopsy usually suggests the presence of *DCIS*.
Cancer The uncontrolled growth of cells in the human body and the ability of these cells to migrate from their primary origin to other parts of the body. The spread of these cells can cause failure of vital organs which can lead to death.
Carcinoma Cancer that begins in the lining of the lobules or ducts, resulting in *lobular carcinoma* or *ductal carcinoma*.
Cyst Commonly occurring fluid filled sacs which develop in the breast. These are sometimes part of fibrocystic breast disease.

D **DCIS (Ductal Carcinoma In-Situ)** The name given to cells that are contained within the milk ducts of the breast that become abnormal in shape and size, and multiply in an uncontrolled way.

E **Epithelial *Hyperplasia*** Increased cellularity and altered architecture within interlobular ducts or within the lobules.

F **Fat *Necrosis*** A benign breast condition that may occur when fatty breast tissue becomes swollen or tender spontaneously, or as the result of injury or trauma.

H **Hyperplasia** Increased cell production, an increase of normal tissue.

I **In Situ** Confined to the site of origin without invasion of neighbouring tissues. An in situ *carcinoma*, means that *malignant* cells are present but have not *metastasised* beyond the original site where the *tumour* was discovered.

Invasive Carcinoma A neoplasm in which collections of epithelial cells infiltrate or destroy the surrounding tissue.

L **Lesion** A definite abnormality in the tissue. May or may not be *palpable*.

Lobular Carcinoma *Cancer* that begins in the lobules (i.e. the glands that make milk)

Lymph Node Small bean-shaped structures of immune system tissue located along the *lymphatic* vessels. Their function is to remove waste and fluids from lymph and help fight infections.

Lymphatic System The tissues and organs (including bone marrow, spleen, thymus, and *lymph nodes*) that produce and store lymphocytes (cells that fight infection) and the channels that carry the lymph fluid. Invasive cancers sometimes penetrate the lymphatic vessels and *metastasise* to lymph nodes.

M **Malignant** Refers to a tumour, or cells, that may invade surrounding tissues or *metastasise* to distance areas of the body. This is the precise medical term for "*cancer*".

Mammoplasty Plastic surgery to reconstruct the breast or to change the shape, size, or position of the breast. Reduction mammoplasty reduces the size of the breast. In the context of this thesis, the term mammoplasty was used to describe the breast tissue samples excised from patients who underwent a mammoplasty procedure. These patients had no presence or history of breast disease.

Mammogram or Mammography An X-ray of the breast, used to screen or investigate breast abnormalities and breast *cancer*, particularly those which are too small to be felt by physical examination.

Mastectomy Surgery to remove all or part of the breast and sometimes other tissue, which can include the *axilla* to remove infected *lymph nodes*. This type of treatment is usually offered to patients with confirmed *invasive carcinoma*.

Mesenchyme The meshwork of embryonic connective tissue from which connective tissues of the body and the blood and lymphatic vessels are formed.

Metastasis The spread of cancer cells to distant areas of the body by way of the *lymphatic system* or bloodstream. Usually spreads to the bone first.

N **Necrosis** A term used to describe the death of cellular tissue. Necrosis within a cancerous tumour may indicate that the tumour is growing so rapidly that blood vessels are not able to multiply fast enough to nourish some of the cancer cells. Necrosis usually indicates that the tumour is very aggressive and can spread quickly.

Neoplasm An abnormal growth that starts from a single altered cell. A neoplasm may be *benign* or *malignant*.

Non-neoplastic Not *benign* or *malignant* disease, but cannot be diagnosed as "normal" tissue i.e. somewhere in between normal and diseased tissue. Non-neoplastic tissue usually gives rise to a *palpable* mass or a *mammographic* abnormality.

Normal Tissue Healthy tissue with no abnormalities or evidence of disease.

P **Palpable** Able to be felt. With breast lumps, the lump is considered palpable if it can be felt by manual examination of the breasts.
Parenchyma The glandular tissue in the breast, as opposed to the fatty or connective tissues.
Primary Site The centre of the palpable mass.

Q **Quality of Life (QOL)** Refers to the patient's ability to enjoy normal life activities. Some medical treatments can impair quality of life without providing appreciable benefit, while others greatly enhance quality of life without consequential side-effects.

S **Sensitivity** The proportion of tissue samples with disease that are correctly identified by defined criteria.
Specificity The proportion of tissue samples without disease that are correctly identified by a test. The specificity is reflective of the number of true negative results divided by the sum of the numbers of true negative plus false positive results.
Surgical Margin Usually an extra 2-3 cm from the edge of the diseased *lesion*. The surgical margin is expected to contain healthy tissue.
Standard Deviation (SD) The average distance of all of the values in a data array from the mean of that data array. Note: ± 1,2 and 3 SD includes 68%, 95%, and 99% of a Normal Distribution.

T **Tumour** A growth (lump or mass) which has formed due to excessive accumulation of abnormal cells. The term “tumour” is not a precise medical term and its use has been avoided in this thesis. Tumours can be *benign* or *malignant*.

W **Wide Local Excision (WLE)** (a.k.a. Breast Conserving Therapy, Lumpectomy) The surgical removal of a breast lump and a small amount of non-cancerous tissue around the lump (called the *surgical margin*), without removing any other part of the breast. WLE is usually followed by radiation therapy if the lump contained *invasive carcinoma*.

Acronyms

Acronym	Abbreviated term
ADH	Atypical Ductal Hyperplasia
ALH	Atypical Lobular Hyperplasia
ANOVA	Analysis of Variance
BM	Basement Membrane
BRE	Bloom Richarson and Elston & Ellis tissue grading system
CI	Confidence Interval
DCIS	Ductal Carcinoma In Situ
ECM	Extracellular Matrix
FWHM	Full Width at Half Maximum
IP	Imaging Plate
LCIS	Lobular Carcinoma In Situ
PC	Principal Component
SAXS	Small Angle X-ray Scattering
SD	Standard Deviation
SE	Standard Error
WLE	Wide Local Excision

Chapter 1

Introduction

1.1 Project Aims

Small Angle X-ray Scattering (SAXS) can be used to characterise structures at the tens to hundreds of nanometre length scales appropriate to evaluating the structures of fibrillar collagens. Several studies have shown that abnormalities in the collagen structure, in particular the amount of degradation, are associated with invasive carcinoma in the breast [28, 98]. Current histopathological analysis (used mainly in the clinic) uses information at the micrometre scale to determine if disease is present within the breast, where observation is limited to cellular structure and overall plasma formation. SAXS, on the other hand, can characterise structural components of breast tissue at the molecular and supramolecular level, and has the potential to provide different information compared to standard histopathology that can be used in the diagnosis and treatment of breast cancer.

This project had three main aims:

1. To determine what SAXS parameters can be used to classify and distinguish between healthy and diseased breast tissue types.
2. To determine if systematic changes in breast tissue structures occur several centimetres away from the centre of the diseased breast lesion.

3. To determine if a laboratory based SAXS camera can be used for the classification and distinction of diseased and healthy breast tissues (for the purpose of applying SAXS in a clinical setting).

Several groups have investigated a range of SAXS parameters relating to the tissue structure and collagen fibre components in breast tissue related to disease [34, 38, 41, 43, 68, 75, 103, 105, 124]. Most of these studies examined patient cohorts that could not be used for classification or diagnostic comparisons, since the number of patients lacked adequate statistical power, or the diversity of tissue types imaged was not sufficient. Nonetheless, these previous studies have established that SAXS has potential to be used as a diagnostic indicator of disease. The work presented in this thesis further investigated the SAXS parameters that showed evidence of distinguishing between diseased and non-diseased breast tissue types with a view to developing the use of SAXS for potential diagnostic or disease aetiology research.

The research presented in this thesis examined breast tissue from the largest patient cohort studied for SAXS investigation. A total of 542 tissue samples excised from 130 patients were analysed using synchrotron radiation (at the Synchrotron Radiation Source (SRS) Daresbury) and a laboratory based X-ray source (Osmic SAXS Camera, Rigaku Corporation, Texas, USA, 2008 [24]). Tissue samples taken throughout the breast mass (not just at the centre of the palpable lesion) have been studied by two other groups [75, 105], where evidence of systematic changes with distances >2 cm away from the tumour were seen. We examined 346 tissue samples taken at varying distances away from the centre of the tumour for 56 patients in order to determine if systematic changes with distance away from the tumour were statistically significant.

The data obtained using a synchrotron radiation source was used as a basis for establishing SAXS feature and classification parameters. Previous SAXS investigations that have attempted to differentiate between diseased and healthy breast tissues have been performed using a synchrotron radiation source [14, 41, 43, 68, 75, 90, 103, 105]. Although synchrotron radiation provides higher quality diffraction patterns than laboratory systems, due to a small collimated beam of high flux, the use of a synchrotron facility for examining breast tissue is not a practical means of providing day-to-day clinical information. In order to determine if SAXS could be used in a clinical setting, a laboratory SAXS camera was used to analyse breast tissue samples. A laboratory based source that can be used directly within the clinic would provide a means of implementing SAXS as a viable part of diagnostic procedures.

1.2 Background

Previous studies have identified that the structural properties of collagen in the breast may be linked to the progression of invasive carcinoma [85, 96, 98]. Specifically, the degradation of the collagen present in the Extracellular Matrix (ECM) has been correlated to the invasion and growth of cancer in surrounding tissues. It is the link between collagen and tissue degradation and breast disease that was investigated in the research presented in this thesis.

A discussion of the anatomy of the breast, the formation and mechanisms of breast cancer, and the structure of collagen is presented in this section. Collagen fibril components, which vary in scale between tens and hundreds of nanometres, can be characterised using SAXS and a brief description of SAXS theory and how SAXS can be used to study collagen in the breast are also reviewed in this section. Finally, comparisons between previous SAXS breast tissue studies, as well as the potential for applying SAXS as a classification tool for breast tissue types, are discussed in Section 1.3.

1.2.1 Features of Breast Cancer

Cancer is the generic term for any malignant tumour, where a tumour is a growth which has formed due the excessive accumulation of abnormal cells. Tumours can be benign or malignant, where benign tumours usually do not cause death, but malignant tumours invade vital organs within the body and impede their normal functions which can result in organ failure and death. There are many types of breast cancers and breast diseases, which are discussed in more detail in Appendix B.

A cross-section of a breast is shown in Figure 1.1. A normal breast is composed of a fatty support structure connected by a series of ducts, each with an orifice at the nipple. The individual ducts, 2-4.5 mm in diameter, have convoluted walls enabling expansion for milk delivery [7]. Further into the breast the ducts divide into smaller ductal units. The deeper ducts are composed of mesenchymal cells, which make up part of the connective tissues within the breast. Ductal units

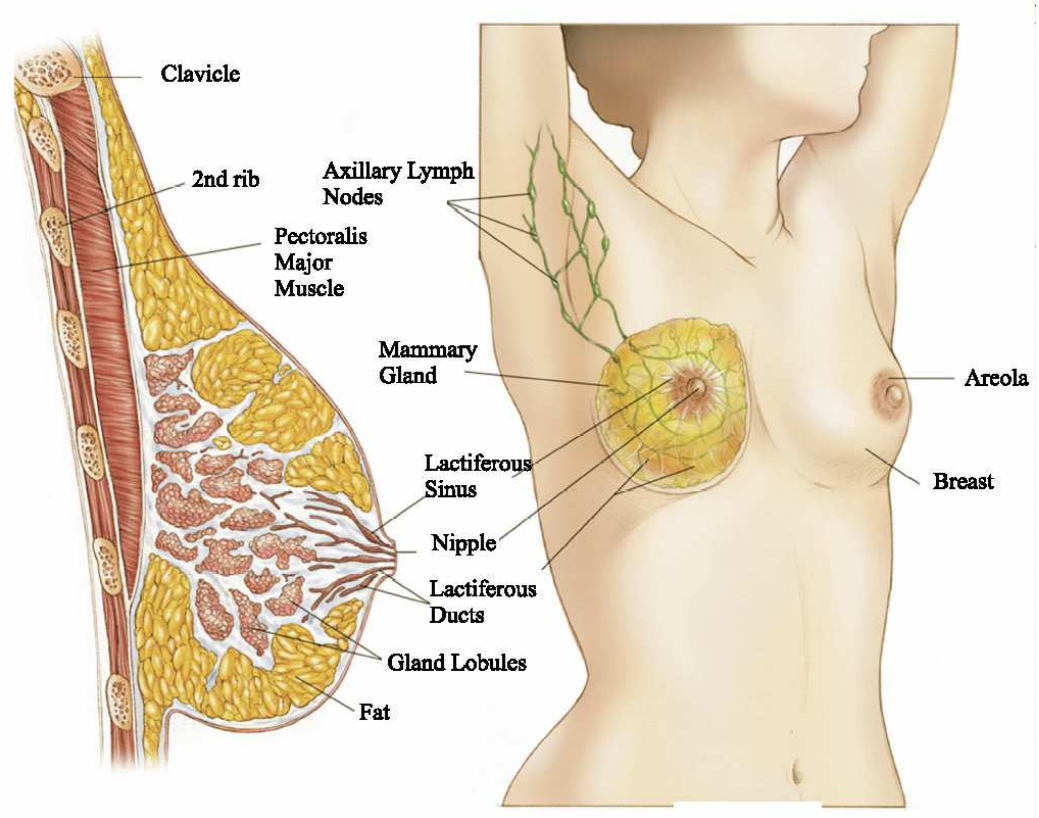


Figure 1.1: Cross-section of a breast showing the basic anatomy described in the text. The image on the right shows the location of the breast and axillary lymph nodes relative to the rest of the torso. (Modified and used with permission from the American Society of Clinical Oncology & The ASCO foundation [54]).

connect to hormonally sensitive lactation terminal ductal lobular units which are 0.3-0.6 mm in diameter and are connected by many extra lobular terminal ductules. Each unit (ductal and/or lobular) is situated in a mycoxoid (mucus-like) loose connective tissue and is supported by (but separate from) denser collagenous cells and fat [32]. Lymphatic vessels in the breast drain into a network of lymph nodes located around the breast's edges, in the axilla (or underarm), and near the clavicle. These lymph nodes are embedded within the fat and are often the first site of metastatic breast cancer.

Figure 1.2 shows the progression of a common type of breast carcinoma: Ductal Carcinoma *In Situ* (DCIS) (see Appendix B). Cancer of the breast usually originates from the epithelial cells that line the terminal duct lobular unit. In the

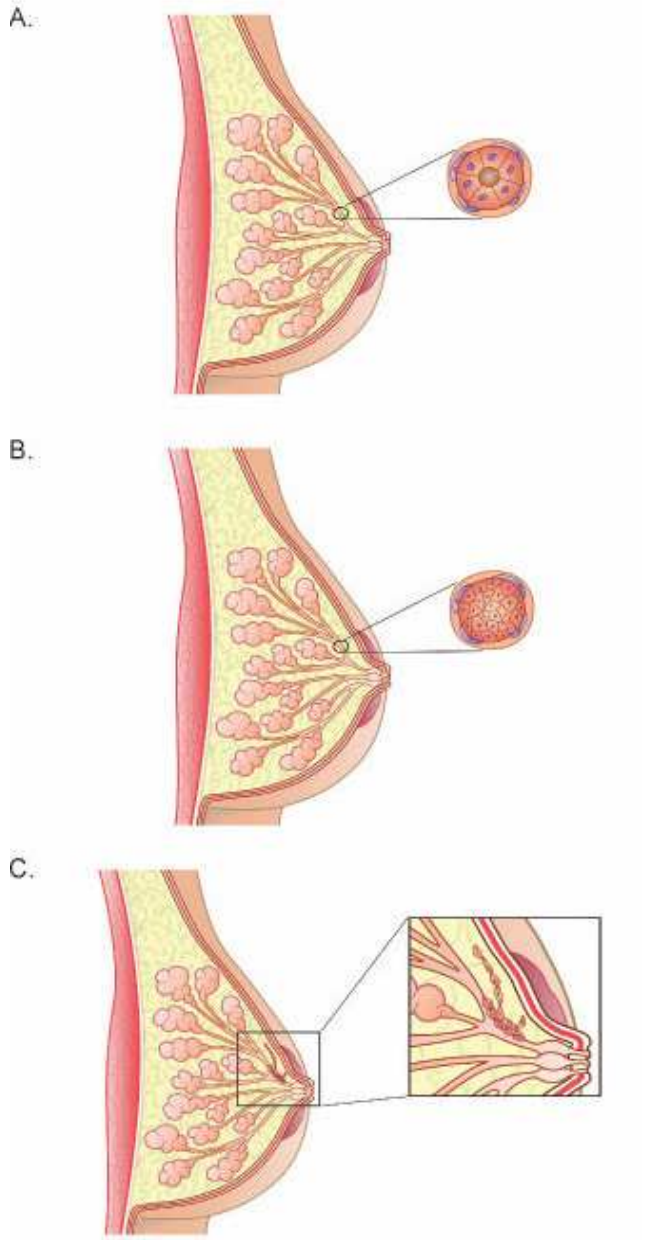


Figure 1.2: Normal and cancerous breast morphology. A) Diagram showing the normal organization of ducts and acini (small saclike dilations) in the human breast. The cross-section demonstrates luminal epithelial cells aligned in a polar manner so their apical side faces and surrounds the lumen. The luminal epithelial cells are surrounded by a non-continuous layer of myoepithelial cells. Surrounding these cells is the basement membrane. Fibroblasts align the basement membrane and this entire structure is surrounded by the stroma, which is predominantly, but not exclusively, composed of type I collagen. B) During ductal carcinoma in situ (DCIS), the normal polar organization of the luminal epithelial cells is lost, as these cells de-differentiate and proliferate. The cross-section shows the epithelial cells completely filling the lumen. In less severe cases of DCIS, the luminal epithelial cells do not completely block the lumen. In DCIS, the transformed epithelial cells do not cross the basement membrane, but remain within the duct. C) DCIS sometimes leads to invasive, or infiltrating, carcinoma, in which the epithelial cells migrate and invade through the basement membrane and into the surrounding stroma [130]. (Printed with permission from Biological Procedures Online)

early proliferative stages of most types of cancer, there is a continuous basement membrane (BM) that separates the hyperplastic epithelial cells from the surrounding stroma. Pathologically, the transition from *in situ*[*] to invasive carcinoma is usually accompanied by the disorganisation and discontinuity of the BM. It is generally believed that penetration of the BM and the subsequent invasion of the ECM is a crucial step in the processes that lead to tumour cell invasion and metastasis [31, 53, 107] (discussed further in Section 1.2.2).

There are many diagnostic imaging techniques that are used to detect breast cancer, but mammography is the most common screening procedure. If a breast abnormality is detected with a mammogram, additional breast imaging with diagnostic mammography, ultrasound, or other type of imaging (for example, Computed Tomography (CT) or Magnetic Resonance Imaging (MRI)) is usually performed. Depending on the results of these imaging tests, a breast biopsy may be necessary. A small piece of tissue is removed from the suspicious lesion by means of needle, or surgical excision which is then prepared for histopathological assessment (see Appendix B for details regarding tissue processing and tissue diagnosis). Biopsy is currently the gold-standard for determining if cancer is present. The ability to obtain an accurate histopathological diagnosis of breast lesions is crucial to any diagnostic test in terms of appropriate treatment planning and patient counselling.

The diagnosis of breast biopsy tissue is performed by a pathologist using a light microscope. Diagnosis is based on the appearance of cytoplasm, nucleic features, and connective tissue seen within a small slice of the tissue specimen (~2 μm thick) (see Appendix B). Healthy tissue has evenly dispersed cells, or well-ordered and aligned cells, with no extraneous proliferation or foreign materials (such as calcifications) present. The first stages of disease show an increase in the number of cells (hyperplasia) which change into atypical cells (atypical hyperplasia), followed by carcinoma *in situ* (non invasive cancer) and finally invasive carcinoma. Not all instances of the disease progress in this manner: cell proliferation can cease at any stage and not become invasive, or can rapidly spread throughout the breast and metastasize. Comparisons of tissue types are discussed further in Chapter 2, which discusses the tissue used in this project, and Appendix B which details histopathology.

Histopathological examination of biopsy tissue, by a pathologist, is currently the standard diagnostic protocol used in the clinic for identifying breast disease [6]. As with any diagnostic technique, there are errors associated in the reporting and

[*] *In situ* carcinoma is defined by the proliferation of tumour cells within the duct and/or lobule.

assessment of the tissue, which can lead to false diagnosis and unnecessary treatment procedures for the patient. Inter-observer variability in histopathological diagnosis represents the uncertainty of this 'gold-standard' methodology of diagnosing breast disease. In a recent study by Verkooijen *et al*. [125], diagnosis was confirmed for 90% of open excision biopsies[†] between an expert breast pathologist and a general pathologist. Tissue samples taken from open excision biopsies are excised from the centre of the known lesion, where the probability of the tissue sample containing disease is high to begin with. Of more concern is the diagnosis of tissue at the margins of surgical masses where any presence of disease left in the patient can lead to further spread or a higher risk of recurrence. When it came to borderline lesions (tissues sampled at the tumour/healthy edge), Verkooijen *et al*. [125] found that only 43% of the samples were diagnosed similarly by both pathologists. Not all breast biopsies can be evaluated by an expert pathologist (this is not feasible considering the number of pathological requests in any clinic or medical centre) and dual reporting is not sufficiently cost-effective to justify implementation [22]. A diagnostic method that can accurately identify disease at borderline lesions, or improve central lesion diagnosis rates, can reduce the risks of misdiagnosis and provide patients and clinicians with more information before considering treatments and surgical procedures.

1.2.2 Collagen

The interaction of invasive carcinoma cells with the extracellular matrix (ECM), the main structural component of which is collagen, have been shown to induce increased production of proteins that degrade the ECM, enabling invasive carcinoma cells to invade the surrounding tissues and progressively grow [30, 38, 39].

Collagen types I and III are the most common structural proteins comprising the ECM, and form fibrils of similar structure within the breast [39]. Type IV collagen is also present in the breast, but is not as abundant as types I and III. Collagen fibrils form a matrix that provides structural support in multiple directions in connective tissue. The fibrils are composed of groups of micro-fibrillar bundles

[†] These tissue samples are typically large (where lots of disease is present and known) and specimens used for diagnosis are easily obtained (i.e. less chance of sampling error)

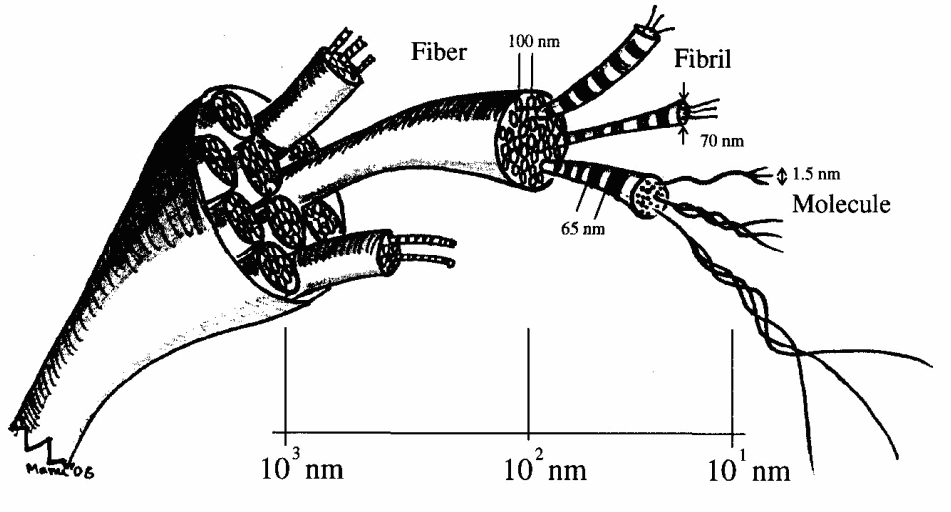
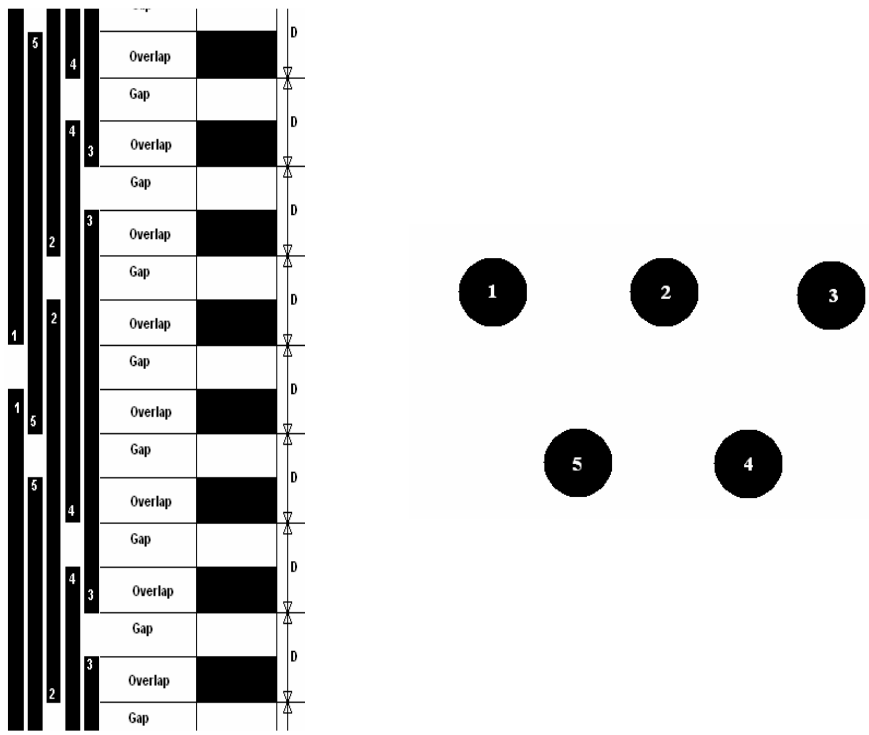


Figure 1.3: The levels of arrangement in fibrillar collagen in breast tissue. Typical dimensions taken from Fernandez *et al*. [51] (Printed with permission from Dr. M. Fernandez, 2009)



(a) (b)

Figure 1.4: The staggered arrangement of the collagen molecules showing the gap-overlap sections created (a) and the hexagonal close-packed arrangement of the molecules in the lateral direction (b). (Printed with permission from Dr. A. Round).

(see Figure 1.3), which are packed laterally in an approximately hexagonal close-packed formation (see Figure 1.4b). The collagen molecules hydrogen bond to each other in a staggered arrangement longitudinally (see Figure 1.4a) and this structure is repeated periodically along the fibril axis [64, 127].

Collagen is composed of three chains, wound together in a tight triple helix. Each chain is over 1400 amino acids long where a repeated amino acid tri-sequence forms a stable structure [39, 40]. Every third amino acid is glycine and the remaining positions in the chain are filled with proline and hydroxyproline. The inordinate number of glycine residues allows tight coiling of each of the alpha chain subunits of tropocollagen, where there is a rise per turn of 0.3 nm as opposed to the 0.36 nm of a regular alpha helical coil [40]. Type I collagen consists of two α1(I) chains and one α2 chain, and type III collagen has three identical α1(III) chains. In normal conditions, type I collagen shows little affinity for cells of epithelial origin, which in healthy tissues are prevented from entering into the collagenous stroma [94].

The axial arrangement of collagen molecules in their fibrils has an overlapping pattern, in which each molecule, approximately 290 nm in length, is staggered a distance D from its neighbours, where D is the axial periodicity of the collagen (see Figure 1.4). Each period comprises a 0.6D long ‘gap’ zone and a 0.4D long ‘overlap’ zone [10]. In healthy breast tissue, collagen fibrils have a radius of about 90 nm, have an inter-fibril distance of approximately 100 nm, and have an axial periodicity, D = 65 nm [49].

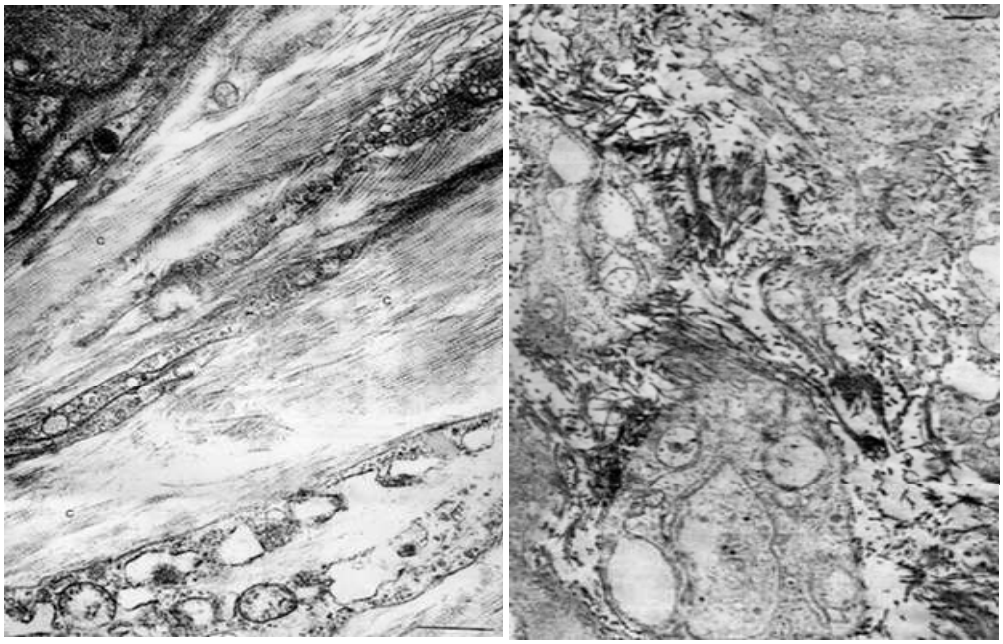
1.2.3 Collagen and Breast Cancer

In order for dissemination of tumour cells to occur, the first step in the life a tumour cell which has just become motile is to break all stable physical contacts with neighbouring cells. Cells move through the surrounding BM and towards and through the blood vessels surrounding the BM. Tumour cell-secreted proteases, the expression of which is influenced by ECM components, are causative in the degradation of basement membranes around epithelial cells. They are also involved in the penetration of other physical barriers, such as dense mesenchyme. Tumour cell locomotion is also influenced by migratory signals. Type IV collagen and type I-trimer collagen, but not normal type I collagen, has been suggested to promote tumour cell migration in the ECM [79]. Interactions of tumour cells with the BM and the ECM are important events in the metastasis of cancer [78]. The implication that cancer cells become invasive by adhesion to, and degradation of, the BM and

the ECM, is central to the activity of tumour cells entering and exiting blood vessels and lymphatics in order to proliferate at distant sites [93].

The components of the ECM (amino acids and sugars) encode information which is interpreted by the cell through interaction with specific plasma membrane receptors [39]. Intracellular pathways that convey signals from the ECM are known to be disrupted by oncogenes and tumour suppressor genes [56, 58, 88]. Also, embryonic traits, such as oncofetal fibronectin, type I-triple collagen, OF/LB collagen and tenascin [29, 46, 63, 67, 95, 128], which are characteristic of the stromal fibrotic response to tumour cells, further constitutes the tumour-promoting role of the ECM [94].

Collagen degradation and dispersal of the collagen bundles in invasive carcinogenic tissues have been observed with an electron microscope; an abundance of thin fibrils chaotically running in all directions throughout the stroma was seen [96] (see Figure 1.5). The fragmented appearance of fibrils may indicate that the



(a) (b)

Figure 1.5: Electron micrographs of collagen tissue in (a) normal tissue and (b) invasive carcinoma tissue. In the normal tissue, the in-tact collagen fibres are aligned and well structured, where the collagen strands in the invasive carcinoma are broken and unstructured. (Printed with permission from Oxford University Press 2009, [96])

new collagen is unable to aggregate properly to reconstitute the typical bundles found in healthy tissues. The degradation of collagen fibrils in tissue has been linked to the presence of the acidic collagen component (αa) [97, 98], where biochemical results have suggested that (αa) causes collagen chain modifications [99], which has been shown in a majority of tumour samples studied. In the study by Pucci-Minafra *et al*. [96], the (αa) component was absent in the adjacent normal mucosa and in two cases of *in situ* carcinoma. They concluded that the accumulation of altered collagen molecules during cancer progression impairs the correct supramolecular organisation of the collagenous stroma and may provide a permissive pathway for the invasive growth of neoplastic cells.

Collagen types I and III are abundant in breast tissues and have structural properties that make it ideal for SAXS analysis. Before discussing how SAXS can be used to characterise collagen in the breast, an understanding of the basic theory and methods of SAXS are explained in Section 1.2.4.

1.2.4 Small Angle X-ray Scattering (SAXS)

The organisation of matter within an object can be analysed using X-ray scattering and diffraction. Discrete peaks of X-ray intensity result from the regular packing of the material being analysed. The ordered structure of the sample gives rise to coherent scattering of the material, which can be observed at specific angles diffracted from the sample. The position of the diffraction peaks, their intensity, and their angular distribution can provide information related to the structures within the sample. Diffraction patterns arise from the inhomogeneities in the distribution of electrons in the material and these differences can be between inclusions and the surrounding matrix or between molecular assemblies. The shape and radial distribution of the observed scattering pattern can also be used to infer the particle size, density, and surface area of the sample constituents [127].

SAXS interrogates length scales at the tens to hundreds of nanometre scale, so SAXS patterns give information on structural ordering arising from the macromolecular features of the sample. In order to obtain large signals, the sample must have systematic ordering over long distances.

Coherent scattering follows Bragg's Law [13]:

$$n\lambda = 2d\sin\theta \qquad (1.1)$$

where n is the order of the diffraction maxima (and is an integer), λ is the wavelength of the incident X-rays, d is the periodicity of the order within the

scattering medium, and θ is the angle between the incident and scattered X-ray beams. Figure 1.6 illustrates the scattering geometry for SAXS, where the elastic scattering of the incident photons are conserved and interference in the wavefronts of the scattered photons can be seen on the detector. The scattering pattern that arises gives information related to the structure of the object based on Bragg's Law. Scattering data can be presented by relating the observed intensity to the modulus of the scattering vector $\vec{q}$ (see Figure 1.7):

$$\vec{q} = \vec{k}_f - \vec{k}_i \tag{1.2}$$

and

$$\left|\vec{k}_i\right| = \frac{2\pi}{\lambda} \tag{1.3}$$

for elastic scattering [26], where the incoming photon, $\vec{k}_i$, and the scattered photon, $\vec{k}_f$ define $\vec{q}$. From the geometry defined in Figure 1.6 and 1.7, the scattering vector can be described as:

$$q = \frac{4\pi \sin\theta}{\lambda}. \tag{1.4}$$

Using Equation 1.1, a real space distance *d* and the length of the scattering vector *q* are related by [51]:

$$d = \frac{n\lambda}{2\sin\theta} = \frac{2\pi n}{q} \tag{1.5}$$

In this thesis, the data is given in terms of *q* (nm^{-1}) and *d* (nm).

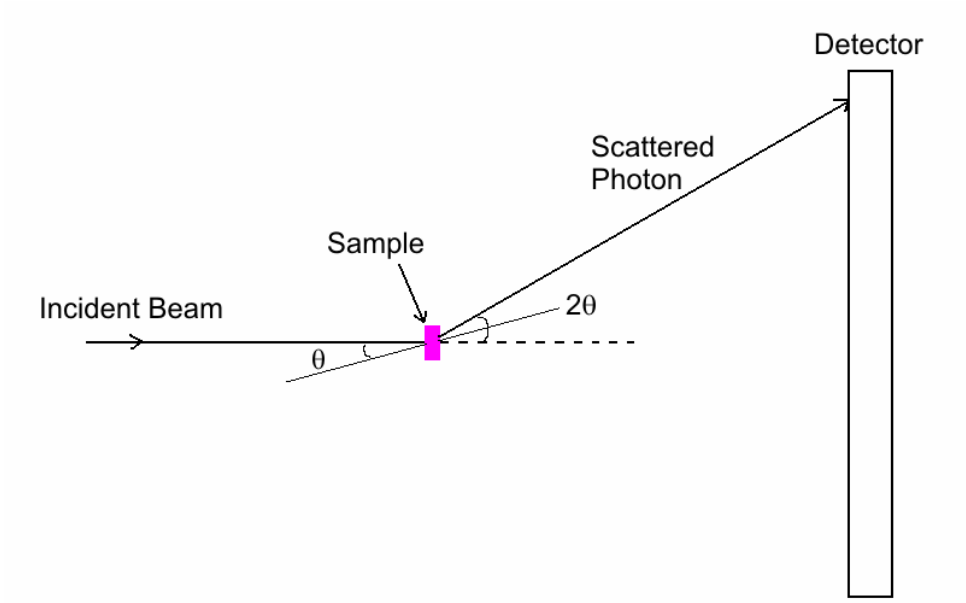


Figure 1.6: Small angle X-ray scattering geometry where the incident beam of photons interacts with the sample and is scattered at an angle of 2θ.

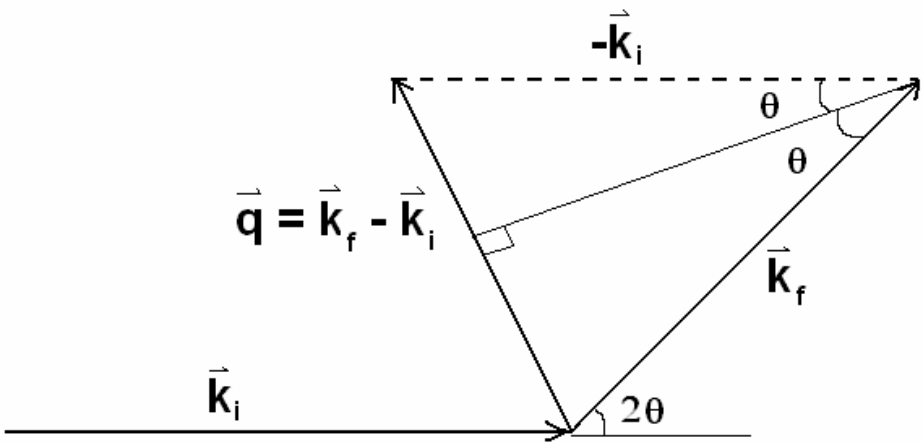


Figure 1.7: Construction of the scattering vector $\vec{q}$ with the incoming photon, $\vec{k}_i$, and the scattered photon, $\vec{k}_f$ where $\left|\vec{k}_f\right| = \vec{k}_i$.

There are two techniques used to obtain the information described by the Bragg equation (Equation 1.1) in the small angle regime: angular-dispersive X-ray diffraction (ADXRD), and energy dispersive X-ray diffraction (EDXRD). With EDXRD, a polychromatic X-ray source and an energy-resolving detector are used at fixed θ in order to determine the characteristic period d of the scatterer. The use of an energy dispersive detector permits rapid measurement over a large momentum

transfer range (energy values can be transformed into momentum transfer values by using $q = \sin(\theta/2)/\lambda$), which allows EDXRD to be performed on compact, conventional laboratory sources [132]. However, the relatively large acceptance angle necessary at the detector in EDXRD means that it is less powerful for the retrieval of structural information compared to ADXRD. The intensity peaks arising from the scattering of ordered structures are revealed with poor resolution using EDXRD techniques, so accurate information related to the specific packing and ordering of the components is difficult to achieve. ADXRD, on the other hand, uses a monochromatic X-ray source that is incident on a sample and the intensity of the scattered radiation is measured as a function of angle using a position sensitive detector (i.e. λ is fixed). Using a 2-dimensional detector for acquiring SAXS information allows for observation of SAXS features to be related to the position on the detector. To our knowledge, only two laboratory ADXRD type SAXS systems have been used to study breast tissue: the SAXSess system used by Round *et al*. [102] and the SAXS Osmic system used for the work presented in this thesis.

Another X-ray scattering technique that utilises the concepts of Bragg scattering to image structures that are in the nanometre range, is wide-angle X-ray scattering (WAXS). WAXS uses the sample principles as Small-Angle X-ray Scattering (SAXS), but the distance from the sample to the detector is shorter. At this shorter sample-to-detector distance, the diffraction maxima are observed at larger angles ($2\theta > 5^{o}$) compared to that of SAXS ($2\theta < 5^{o}$). Patterns from WAXS are ideal for data concerning correlations on an intra-molecular or inter-atomic level [82, 116], whereas patterns from SAXS concern correlations on an inter-molecular level [35, 46, 58, 82]. Thus, SAXS is more suited for analysing samples where there is macromolecular or aggregate order, such as in collagen fibres.

1.3 SAXS and Breast Tissue

SAXS has been used to examine breast [44, 51, 55, 77, 101, 103, 127-129], and prostate cancers [75] all of which contain collagen structures affected by invasive disease. Collagen structural properties have also been studied using X-ray scattering in relation to tendon [11, 48, 49, 123], bone [50, 104, 129], and corneal tissue [52, 53, 98, 99]. With these tissues, changes in the structure and distribution of the collagens were related to diseases such as myopia for corneal tissues and sclerosis for bone.

Table 1.1: The SAXS features analysed in this study and their relation to the structural components in breast tissue.

SAXS Feature	Physically Represents
Collagen Axial D-spacing (Position of the axial peak)	Spacing between the gap-overlap sections of the collagen molecules.
Collagen Axial Peak Amplitude	Degree of long-range order in the collagen fibrils.
Collagen Axial Peak Width (the Full Width at Half Maximum – FWHM)	Wide peaks correspond to a large variability in the d-spacings and suggest that the fibrillar collagen overlap spacings are variable. Narrow peaks correspond to low variability in the d-spacings, suggesting highly ordered fibrillar collagen overlap spacings.
Area Under Axial Peak	A combination of the factors attributed to the peak width and peak amplitude. Describes the degree of long-range order and distribution of the collagen fibrils in the tissue sample (longitudinally). Gives some indication of the amount of collagen present in the tissue sample
Amorphous Scatter (Background, equivalent to the area subtracted before peak fitting)	Proportional to the total area per unit mass. Tells us about variations in the electron density of the entire heterogenous sample.
Amorphous Scatter between the third and fifth order peak	Integrated intensity between $q = 0.3\text{-}0.5\ nm^{-1}$. Used to describe the specific surface area of the scatterers.
Centre of Equatorial Peak	The distance between the hexagonal packing of the collagen fibrils.
Equatorial Peak Width (FWHM)	Diameter of the collagen fibrils.
Equatorial Peak Amplitude	Degree of long-range order of the packing of the collagen fibrils.
Equatorial Peak Area	A combination of the equatorial peak amplitude and equatorial peak width. Describes the degree of long-range order and distribution of the collagen fibrils in the tissue sample (laterally). Gives some indication of the amount of collagen material present in the sample.
Polar Angle Data (Collagen fibril orientation)	A measure of the degree of alignment of the collagen fibrils. If the peaks form along a distinct axis, then this suggests that the fibrils are highly orientated. If the peaks form rings on the SAXS pattern, this suggests that the fibrils have no preferred orientation.

Table 1.2: Summary of the features studied by several SAXS groups that may be useful in classifying disease in breast tissues.

Group	Features Observed	Conclusions
Lewis *et al.* [77]	Intensity in Axial and Bessel peaks Spacing of 3rd axial peak: D-spacing	Different for tumour, benign, and normal tissues. Changed for normal tissues up to 8 cm away from tumour site. Larger for benign lesions compared to malignant and normal tissue.
Fernandez *et al.* [41, 43]	Axial period of collagen structure Average intensity of scattering Area of the 3rd and 5th collagen peaks Adipose Tissue: average intensity	Distinguish between normal and diseased tissues. May be able to distinguish between different types of disease. Third axial peak spacing larger for tumours. Differences between normal and cancerous tissue. Different between normal and cancerous tissue. Detect inhomogeneous structures due to invasion of cancer in adipose tissue.
Butler *et al.* [15]	Segmentation, peak locations for 3rd and 5th axial, ratio of peak intensities (5th/3rd), patient age	These features may be useful in classification.
Round *et al.* [102, 103]	3rd order axial peak area Equatorial Peak Area	Detected differences in normal samples up to 6 cm away from tumour site Differences between normal and benign tissues, but no differences seen with cancerous tissues.
Falzon *et al.* [44]	Wavelet coefficients to describe axial d-spacing of collagen and orientation	Can differentiate between healthy and diseased tissues.
Castro *et al.* [17, 18]	Fat Peak: Height and position of peak	Differences between healthy and cancerous breast tissues.

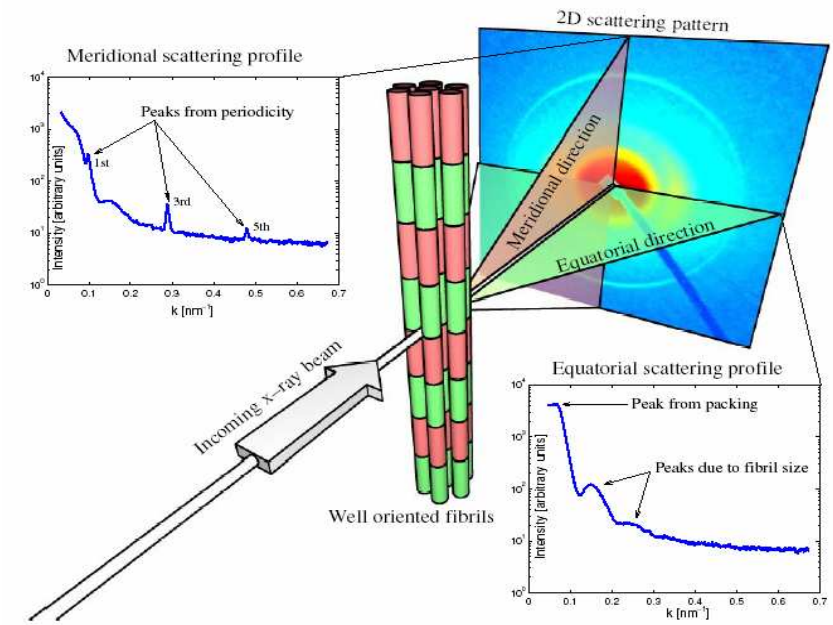


Figure 1.8: Meridional (axial) and equatorial scattering from well-oriented fibrils. The equatorial axis was designated 0º, where the meridional axis was then at a polar angle of 90º. (printed with permission from the Institute of Physics Publishing, 2009 [119]).

In this section, the SAXS features observed in breast tissue studies prior to this thesis are discussed. The diagnostic implementation as realised by other groups are used as a starting point for the work performed for this thesis. All of the SAXS breast tissue studies to date have not examined all of the possible features together with a large patient cohort. Table 1.1 lists the features studied for this project and their relation to the structural parameters in breast tissue. Table 1.2 summarises the results of previous SAXS breast tissue studies and how the features differed between diseased and healthy breast tissue types examined by each group.

1.3.1 Axial (Meridional) Peak Features

As a long periodic structure, collagen produces maxima in scattered intensity on SAXS images (as illustrated in Figure 1.8 and in Figure 1.9), where the sharpness and width of these peaks reflect the regularity in electron density. For well-oriented systems such as collagen, the period length (the gap overlap spacing, illustrated in Figure 1.8) can be easily read from the Axial Bragg peaks in the meridional direction.

The SAXS patterns from oriented collagen fibrils have distinct features in the meridional direction due to this axial periodicity [119]. If the peaks appear as arcs in the scattering pattern, this is evidence of a variation in the alignment of the fibres (discussed in Section 1.3.5).

Axial peaks from the axially periodic collagen structures are wider for malignant than for healthy tissue, which has been interpreted as a decrease of ordering with malignancy [119]. Figure 1.9 shows an example of the q-spacing versus the intensity for one of the breast tissue samples imaged using SAXS at the Daresbury Synchrotron Radiation Source (SRS). The third and fifth order axial peaks can be seen at q = 0.282 nm^{-1} and q = 0.470 nm^{-1}, respectively[‡] [35, 63, 84]. Other peaks within this q-range (such as the 4th peak) are not visible because their reflections are forbidden [35, 56, 64, 119]. The odd ordered Bragg peaks indicate a near-symmetric square-well density profile for the alternating gap/overlap arrangement along the fibre axis. For this reason, most SAXS studies dealing with biological tissues, such as the breast [15, 20, 44, 49-51, 67, 77, 101, 102, 119, 128, 129], pay specific attention to the odd ordered Bragg peaks to obtain a sense of periodicity within the collagen present in the sample. Specifically, in the current study, special attention was made to the 3rd and 5th order Bragg peaks.

Two independent groups, both using similar experimental methods and synchrotron radiation, have used SAXS to examine the state of collagen fibrils in breast tissues in order to determine if differences can be attributed to breast disease: Lewis *et al*. [77] and Fernandez *et al*. [49, 51]. Lewis *et al*. [77] had 43 patients in their study, comprised of 18 normal, 18 invasive carcinoma, and 7 benign fresh tissue samples. Fernandez *et al*. [51] had 7 patients in their study, comprised of 7 invasive carcinoma and 1 benign tissue sample (i.e. 8 tissue samples from 7 patients), where histologically healthy tissue samples were additionally taken from the benign patient and one invasive carcinoma patient. Fernandez *et al*. conducted two studies: (1) using fresh tissue samples [51] from 7 patients and (2) using formalin fixed tissue samples [49, 50] from 3 patients. The manner in which the SAXS data were acquired between the two studies were somewhat different: Lewis *et al*. [77] obtained SAXS images averaged over the whole tissue sample (i.e. one 2D scattering image for one patient), whereas Fernandez *et al*. [49, 51] obtained SAXS images at many points in the tissue sample (1 mm to 0.25 mm spacings) and treated each resulting image as a separate observation corresponding to the pathology of the sample at that point (i.e. several images for one patient). Fernandez *et al*. [49, 50] found that the axial D-spacing

[‡]Where the peaks are at measured positions based on the orders derived from the collagen periodic D-spacing of 66.8 nm [35] for healthy breast tissue.

between formalin fixed tissue and fresh tissue were systematically different by 0.1 nm. This suggests that the differences in the axial D-spacing between tissue types are preserved, but the absolute values of the axial D-spacing are different compared to fresh tissue. Lewis *et al*. [77] imaged the tissue sample before fixing the piece in formalin and pathology reporting was done after the experiment. With the work presented in this thesis, only fresh breast tissue samples were used for analysis, and a description of how the tissue samples were handled before, during, and after SAXS imaging is given in Chapter 2.

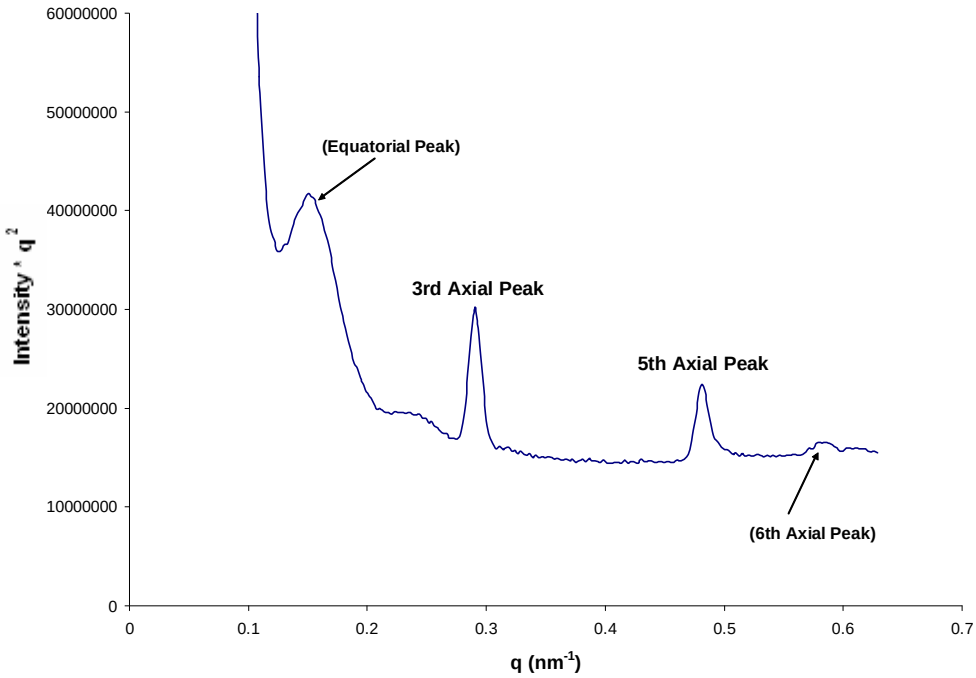


Figure 1.9: Plot of q versus the Intensity*q^2 of the radially averaged data for a mammoplasty tissue sample. The third and fifth order axial peak reflections are prominent in the profile. The 6th order axial peak and the 2nd order equatorial peak are also indicated.

Fernandez *et al*. found the axial d-spacing of collagen embedded in invaded adipose tissue (fresh tissue) to be 65.3 ± 0.2 nm and 65.0 ± 0.1 nm in benign lesions and healthy connective tissue [51]. Lewis *et al*. found the axial D-spacing of collagen for invasive carcinoma and normal tissues to be 64.90 ± 0.05 nm and 64.87 ± 0.02 nm, respectively. Lewis *et al*. also found that benign tissues had a smaller axial D-spacing of 64.52 ± 0.04 nm compared to invaded and normal breast tissues [77]. Both studies suggest that the axial D-spacing is larger in invaded tissues than in healthy or benign tissues, but the small differences in the absolute axial D-spacing

values between the two studies can perhaps be attributed to differences in experimental conditions or calibration. Furthermore, the number of patients used for both studies may not have been large enough to show significant differences in disease types with certainty: Fernandez *et al.* [49, 51] analysed 7 patients, and Lewis *et al*. [77] analysed 43 patients. Both studies found that the axial d-spacing alone was not sufficient for differentiating breast disease with high sensitivity. Instead, other features (discussed in the following sections), perhaps in combination with axial D-spacing, may be more reliable for use in distinguishing differences between healthy and diseased tissues.

Another axial feature that has shown differences between diseased and healthy breast tissues is the area under the collagen peaks [15, 17, 44, 49, 51, 77, 102]. The area encompasses both the peak height (amplitude) and the peak full-width-at-half-maximum (FWHM). The peak amplitude indicates the degree of long range order in the collagen fibrils which arises from the fact that X-ray scattering is a process of constructive interference. Round *et al*. [102] found that the amplitude of the 3rd order axial peak was higher for benign tissues compared to malignant tissues. The reduction in the peak amplitude for malignant tissues can be attributed to the action of metalloproteinases (MMPs) on the malignant phenotype [8, 34, 45, 87, 92, 110], however, without accurate knowledge of the amount of collagen in the tissue sample, the peak amplitude cannot be used as a marker for disease.

The FWHM of the peak describes the distribution of the particle sizes within the sample. In the case of collagen, the FWHM relates to the distribution of axial d-spacings in the fibrillar collagen. For example, a narrow peak relates to a low variability in the axial d-spacings which indicates that all of the fibrils are very similar in terms of their d-spacing and structure (and vice versa for wide peaks). Therefore, the area of the peak is a combination of the features relating to the peak amplitude and peak FWHM: it describes the degree of long range order and distribution of the fibrils in the sample, as well as gives some indication of the amount of collagen in the sample (Table 1.1).

Fernandez *et al.* [49] observed large differences in the area of the collagen peaks between different types of tissues, but no quantitative analysis was performed for statistical comparison. They did, however, suggest that this feature can be useful for tissue mapping and may provide more information on the changes of tissue structures at the molecular level during cancer growth. Round *et al.* [102] analysed tissues at various distances away from the centre of the primary lesion in breast patients and detected differences in the 3rd order axial peak area in pathologically normal tissue and invaded tissues ($p = 0.013$), however, there were no significant

differences with benign tissues ($p = 0.098$). A full statistical analysis performed on a larger sampling population of the axial peak features may show differences between tissue types that can be useful for classifying and diagnosis.

1.3.2 Equatorial Peak Features

In fibrils, collagen molecules pack equatorially and hydrogen bond to each other in a quasi-hexagonal arrangement (as shown in Figure 1.4b). The lateral characteristics are seen in SAXS images as peaks in the equatorial direction of the pattern (see Figure 1.8). These discrete diffraction maxima can be used to determine the diameter of the cylindrical collagen fibrils and the spacing between fibrils. These structures can be modelled as Bessel function oscillations [36], since they arise from a combination of the cylinder diameter and the packing of the cylinders.

Fernandez *et al*. [41, 43] found clear differences in the lateral ordering and packing of collagen between healthy and invaded tissues in SAXS images. The equatorial peak maxima were relatively broad for invaded tissues, suggesting a range of fibril diameters; however the distances between the fibrils, indicated by the separation between the maxima, seemed to remain the same irrespective of tissue type. Round *et al.* [102] used the SAXS data obtained from the same 43 patients examined by Lewis *et al*. [77], plus SAXS data from an additional 39 patients, and imaged these patient samples using similar experimental methods to Lewis *et al*. [77]. Round *et al*. found a significant difference in the total area under the equatorial peaks between normal tissues and benign neoplasms ($p = 0.0055$), but there was no distinction found between invaded tissues and normal or benign tissues ($p > 0.05$)[§]. Their data indicates that there is a high degree of inter-patient variability, which is characteristic with such small patient cohorts (43 patients with Lewis *et al*. [77] and 82 patients with Round *et al*. [102]). Otherwise, the different conclusions made by each group suggest that either the calculation methods were not meticulously examined, or that the equatorial peak areas are randomly distributed between patients and cannot be used to differentiate between tissue types.

[§] How the values in these peak areas differed was not discussed by Round *et al*. [102] , only that the peak areas varied significantly between the tissue groups.

1.3.3 Amorphous Scatter

Another parameter that has been examined with SAXS [49, 51, 91] is the smoothly varying scattering intensity between the identifiable diffraction features, often called amorphous scatter. The amorphous scatter describes the variations in the electron density of the entire heterogeneous sample and excludes any peaks or patterns associated with characteristic lengths or dimensions of the collagen fibrils. For this thesis, a quadratic function fitted to the data was integrated over a q-range to obtain a single measure of the amorphous scattering intensity (between q = 0.1-0.6 nm^{-1} was examined with the synchrotron images, and between q = 0.25-2.3 nm^{-1} with the laboratory images, discussed further in Chapter 3).

The amorphous scatter from malignant tissues has been shown to be higher than the amorphous scatter from normal tissues [16, 41, 43, 57, 68, 71, 87, 90, 91, 105]. Rogers *et al*. [101] suggested that the amorphous scatter between q = 0.11 nm^{-1} and q = 0.23 nm^{-1} differs in the presence of invasive carcinoma tissue. Fernandez *et al*. [49] looked at the range between q = 0.50 nm^{-1} and q = 0.56 nm^{-1} and found that lower intensities corresponded to adipose, medium intensities to healthy, slightly higher intensities to invaded, and the highest intensities to necrotic tissues. Lewis *et al*. [77] considered the fraction of scatter in the diffraction peaks (arising from the axial and lateral fibril features as discussed above) as a fraction of the total scatter (i.e. peak scatter plus amorphous scatter). They found that the fraction of total scattered intensity contained in the peaks was much less in invasive carcinoma tissue than in either mammoplasty tissue or benign lesions. This finding, coupled with the axial d-spacing, was statistically significant in distinguishing between tissue types (mammoplasty, invasive carcinoma, benign).

1.3.4 Fat (Adipose) Tissue

Fat is composed of hydrocarbon chains that are packed into a hexagonal lattice, which form lipid molecules. The molecules come together to create a lipid bilayer which forms a lamellar structure. The spacing between each of these lamellae is about d = 4.2 nm (q = 1.49 nm^{-1}) [131]. A characteristic ‘fat ring’ is seen on SAXS images at this q-spacing (see Figure 1.10), and has been observed with breast tissue at d = 4.26 nm (q = 1.47 nm^{-1}) by Suortti *et al*. [120]. Most diffraction studies that have characterised the spacings of the lamellae in fat have utilised EDXRD [17, 20, 68, 72, 121], where the momentum transfer of this feature was measured. Measuring

the scattering features of the fat peak with EDXRD techniques is difficult due to the poor resolution of peak features obtained with EDXRD.

SAXS has sufficient q resolution to allow for diffraction of the fat peak's accurate position, amplitude, and width. The centre of the fat peak, which measures the spacing between the lamellae, varies with the amount of water and the percentage of adipose content within the sample [18, 72]. Tartari *et al*. [121] used filtered and fresh excised pig fat for analysis (powder X-ray diffractometer) and found that the scattering features varied up to 6.3 nm, clearly showing that the inhomogeneity factors associated with fat tissue can greatly affect the position of the peak. Tartari *et al*. suggested that it was the inclusion of other tissue components within the sample (such as water, proteins, and residual lean tissue) that interferes with the scattered signature of fat tissue [121]. Castro *et al*. [18] used EDXRD and found differences in the ratios of the peak heights to the background intensity of the fat feature between healthy (5.320 ± 0.160) and cancerous (3.846 ± 0.115) breast tissues. This suggests that healthy fat has a highly ordered lamellar structure, whereas the lamellar components in invaded fat have begun to break down. The height of the fat peak may show differences between healthy and diseased breast tissues, which can be used as a classification feature.

Fernandez *et al*. [51] used SAXS to analyse breast tissue between q = 0.02-0.7 nm^{-1}. Although the fat peak is not visible within this q-range, the group analysed the average intensities between healthy and invaded fatty breast tissues. They found a significant change in the SAXS pattern between invaded fat and healthy fat. The average intensity between q = 0.02-0.7 nm^{-1} increased by a factor of 5 in the invaded regions of tissue compared to normal regions. This could indicate that the invasion of cancer in adipose tissue introduces inhomogeneous structures at the supramolecular level. Newly formed collagen fibrils (type I trimer collagen: an embryonic foetal collagen phenotype) are known to form in developing invasive carcinoma which provides a pathway throughout the fatty tissue to spread disease to other areas [93, 94, 96]. This may account for the increase in the average intensity between this q-range, where the newly formed collagen fibrils introduce more inhomogeneous structures of this size in diseased fatty tissues.

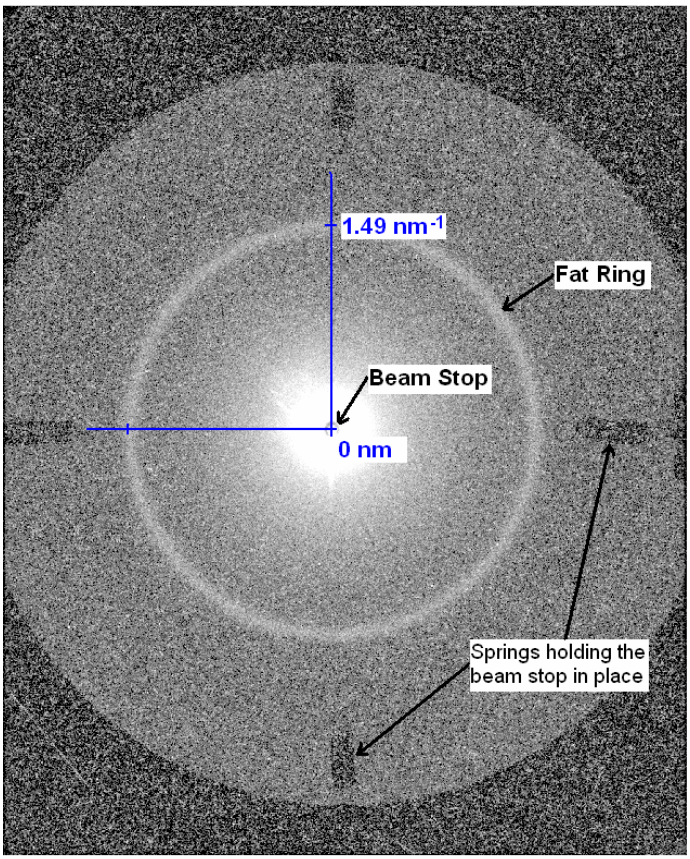


Figure 1.10: SAXS image taken with the SAXS Osmic Camera. The 'fat ring' is clearly seen on the image, located at q = 1.49 nm^{-1}. The scale shows the q-spacing axis, which is radial from the centre of the beam stop. The beam stop, which stops the transmitted beam from reaching the detector, is seen in the middle of the image, which is held in place with thin wire attached to springs.

1.3.5 Collagen Fibril Orientation

Assuming preferentially oriented fibrils, inter-fibrillar scattering by these fibrils can be approximated by a linear superposition of the scattering from the individual fibrils [128]. Hence, the distribution of scattering intensities as a function of polar angle (the equatorial axis was designated 0°) in the SAXS image provides information on how the collagen fibrils are orientated in the tissue sample. Wilkinson *et al*. [128] and Falzon *et al*. [44] have suggested that healthy breast tissues have a preferred orientation or show evidence of multiple alignments in one direction on SAXS patterns compared to invaded tissues. Electron microscopy

images (see Figure 1.5) have showed that healthy collagen fibrils align within the tissue structures naturally, but have chaotically distributed segmented collagen fibrils in diseased tissues [50, 62, 69]. This suggests that the degree of orientation of the collagen fibrils may indicate some differences between tissue types.

1.4 Motivation and Contribution to Knowledge

Although many studies have looked at one or more SAXS features that characterise breast tissue (listed in Table 1.1), none have analysed all of these features exhaustively (Table 1.2) using robust statistical analysis for classification of breast tissue. Examining all of the studies to date, the conclusions of different authors are sometimes ambiguous or inconsistent.

The aim of the work presented in this thesis was to perform a comprehensive analysis of all of the parameters listed in Table 1.1 for 543 breast tissue samples excised from 130 patients. This is the largest breast patient cohort ever studied using SAXS, which allows for better statistical power for analysing features between tissue groups and is more representative of a patient population. A multitude of patient ages, disease grades and disease types, as well as tumour sizes are represented within the 130 patient cohort analysed in this work. A description of the patient cohort and tissue sample collection procedures are described in Chapter 2.

Lewis *et al*. [77] were the only group who studied pathologically normal tissue sampled from patients with known invasive disease and normal tissue sampled from patients with no history or presence of disease. Differences were apparent between these two tissue types, but no conclusive results were presented due to the small cohort of patients used in the study (43 patients). A more detailed study of the SAXS parameters between normal tissue from patients with known disease and normal tissue from patients with no history or presence of disease is described in Chapters 4-6. A thorough examination of the SAXS features between these two types of normal tissue may provide significant evidence that supramolecular changes occur in the entire breast when disease is present, but are not detected by conventional histopathology,

Systematic differences in SAXS features with distances >2 cm away from the centre of the tumour have not been extensively analysed. Round *et al*. [102] and Lewis *et al*. [77] were the only groups that have analysed breast tissues at varying distances (up to 6 cm) away from the centre of the tumour. A model that maps the

changes in tissue structure over the breast area is presented in this thesis (Chapters 5 and 6). An accurate model would be a valuable tool in diagnosis and compliment conventional histopathology to show areas where disease may be present at a supramolecular level. Mapping of structural changes over the breast may reveal directional spread of disease which may contribute to a better understanding to the mechanisms behind metastasis. Furthermore, changes to tissue structure at the supramolecular level can be seen using SAXS that standard histopathology diagnosis cannot observe; this information may be useful for examining surgical margins for ensuring all of the disease has been removed from the patient.

Breast tissue has been analysed with laboratory based EDXRD systems and with synchrotron (typically ADXRD) systems. As discussed in Section 1.2.3, laboratory EDXRD does not provide the high quality information needed to characterise the peak features in SAXS images that can distinguish between breast tissue types. The use of synchrotron radiation provides a high intensity source of monochromatic radiation, allowing for accurate q-space resolution of features. However, using a synchrotron facility is not a feasible experimental option when continual use is needed, as would be the case if this technique was to be developed for possible diagnostic use in the clinic. The SAXS Osmic system used to acquire SAXS images for this thesis is only the second laboratory based ADXRD SAXS system to study breast tissue (see Chapter 3). SAXS features obtained using a laboratory X-ray source that can distinguish between breast tissue types has the potential for application of SAXS in the clinic, where the diagnostic information related to the structural components of breast tissue could be readily available for medical practitioners.

A comprehensive examination of all of the SAXS features listed in Table 1.1 that also adequately represents a patient population, is the first step in providing insight into the differences between the supramolecular components of healthy and disease breast tissue at the tens to hundreds of nanometre scale. Information that is different from what current histopathological techniques use for the diagnosis of breast disease may aid in deceasing the rates of misdiagnosis and assist surgeons in confirming diagnosis at surgical margins. Knowing that the SAXS characteristics associated with collagen fibril structures in breast tissues can differentiate between tissue types can lead to further research regarding the implications these characteristics have on diagnostic and metastatic control. Furthermore, in order to make SAXS clinically feasible, tissue type discriminator sensitivity, similar to that obtained from a synchrotron must be available with a laboratory X-ray source. Having a reliable, accurate, and efficient alternative to conventional histopathology

readily available within a clinic would be advantageous to both health care professionals and patients in terms of prompt treatment and reduced psychological trauma caused by prolonged diagnosis.

1.5 Outline of the Thesis

The work presented in this thesis explores aspects of SAXS for breast tissue characterization. In the next two chapters, methods of tissue collection and histopathologic assessments used in the project are explained (Chapter 2) and SAXS data methodologies for both synchrotron acquired data and the laboratory SAXS camera are discussed (Chapter 3). The analyses of tissue samples excised from the centre of the lesion that were imaged using synchrotron radiation is presented in Chapter 4. All of the features listed in Table 1.1 are thoroughly examined in Chapter 4, and their features that show the best distinction between tissue types are compared (Aim 1). Chapter 5 presents an analysis of tissue samples taken at varying distances away from the centre of the tumour to see if systematic changes in SAXS features with distance away from the tumour are present (Aim 2). These changes are further mapped over the entire area of the excised breast mass to show areas of disease progression.

In Chapter 6, the information obtained with the laboratory SAXS camera with a subset of the 130 patients imaged with a synchrotron radiation source (66 patients) is discussed. Significance of the parameters visible with the laboratory camera is reviewed and compared for effectiveness and features not seen with the synchrotron data (in particular, the fat peak features) are characterised between tissue groups. The feasibility of using the laboratory SAXS camera for aid in clinical diagnosis and screening was also addressed (Aim 3). The final chapter (Chapter 7) discusses the overall project results, future work, and conclusions.

Chapter 2

Breast Tissue Samples

The patients recruited for the project, tissue harvesting and storage, and the histopathological assessment of each tissue sample are discussed in this chapter. Previous SAXS breast tissue studies [16, 34, 38, 41-43, 75, 86, 103, 105, 124] used multiple samples from one patient and treated these samples as independent tissues. Table 2.1 lists the number of patients and the tissue samples used for SAXS breast tissue studies. The work presented in this thesis is the largest study of breast tissue ever examined using SAXS and a full description of the tissue samples is described in this chapter.

2.1 Patient Recruitment and Sample Collection

Patient recruitment was carried out over the period of August 2005 to February 2007, at Southern Health Medical Centres[**]: Clayton, Moorabin, and Berwick. The last SAXS synchrotron experimental session allocated for imaging the breast tissue samples collected for this thesis was February 2007, and that is when patient recruitment ceased. All patients who presented with a palpable mass, before surgical

[**] Southern Health is the state of Victoria's largest metropolitan health service located across south eastern Melbourne. They serve over 866 000 people (21% of the city's population) and have over 12 000 staff.

Table 2.1: SAXS breast tissue studies listing the number of patients and the number of tissue samples in each study. In some cases, the number of patients or the number of tissue samples were not stated, and were assumed.

SAXS Study	# Patients	# Tissue Samples	Tissue Types
Kidane *et al*. [68]	100	Assumed: 100	Inv. Carcinoma Benign Normal Adipose
Round [102]	82	225	Inv. Carcinoma Benign Normal Mammoplasty
Lewis *et al*. [77]	43	99	Inv. Carcinoma Benign Normal Mammoplasty
Fernandez *et al*. [51] [49] [50]	9 7 28	Multiple Scattering Images (3100)	Inv. Carcinoma Benign Normal
Rogers *et al*. [101]	6	12	Inv. Carcinoma Normal
Changizi *et al*. [20]	Assumed: 99	99	Inv. Carcinoma Benign Normal Adipose
Falzon *et al*. [44]	49	Assumed: 49	Inv. Carcinoma Benign Normal Mammoplasty

Inv. Carcinoma = Invasive Carcinoma
Normal = Tissue that is histopathologically normal but excised from patients with known disease present.
Mammoplasty = Tissue that is histopathologically normal and excised from patients with no history or presence of disease (harvested from patients undergoing breast reduction/mammoplasty procedures).

removal, or confirmed local mass by a radiologist (from X-ray, mammogram, or ultrasound) were asked to participate in the study. For safety reasons, patients with known infectious disease status (such as Hepatitis or HIV) were not asked to participate, as well as anyone under the age of 18. Other than these criteria, patient selection was random and was largely determined by the availability of collection staff. Collecting breast tissue samples in this manner provides a patient cohort that is representative of a clinical population with a diverse range of disease types and patient backgrounds (related to age, menopausal status, and medical history).

A total of 185 patients agreed to participate in the study. There was only one patient who did not participate because their native language was not English and it was felt that this would disadvantage their understanding of the implications of the study. Out of the 185 patients, only 130 patients had tissue samples that could be analysed at the centre of the palpable lesion (termed as 'primary site'). The other 55 patients had tissue samples that were unsuitable for accurate histopathological diagnosis. These 55 primary site tissue samples either disintegrated during the freezing and storage process (see Section 2.2), or lacked enough material needed for adequate staining and visualisation of the tissue for diagnosis by a pathologist (see Section 2.5).

Four tissue groups were analysed: invasive carcinoma, benign, normal, and mammoplasty. Tissue samples that had any presence of invasive carcinoma (for example, 5% of the tissue core contained invasive carcinoma and 95% of the tissue core was normal tissue) were diagnosed as invasive carcinoma (and similar with benign). Normal tissue was breast tissue sampled from patients with known disease, but was histopathologically diagnosed as normal parenchyma. These normal tissues were expected to contain invasive carcinoma, since the tissue biopsy was taken at the centre of the palpable mass which was confirmed by surgical pathology to be diseased. Mammoplasty tissue was normal tissue excised from patients with no known presence or history of breast disease. The tissue groupings and histological diagnosis of the samples is described in Appendix B.

Clinical information and patient history regarding age, diagnosis, menopausal status, and pathological diagnosis of the lesion (based on the surgical pathology report) were recorded from the patient's medical chart. Not all of the patient's charts were available for review due to location of the charts[††], patient's switching hospitals after their surgery, or the chart was needed by the patient's health care practitioner for an extended period of time. Also, some information was not available because it was not recorded.

Out of 130 patients, only 114 patients had their medical charts available. The ages of these 114 patients were recorded and available for analysis. Table 2.2 lists the average ages of the 114 patients with the respective diagnosis of their primary site tissue samples. T-tests showed that the ages between each of the tissue groups were significantly different ($p < 0.05$), except between benign/mammoplasty ($p = 0.94$) and normal/mammoplasty patients ($p = 0.14$). The distribution of ages for

[††] Patient's medical records were not always at one of the three hospitals with Southern Health. Sometimes they were kept with the patient's General Practitioner where they could not be accessed.

the 114 patients over the four tissue groups is shown in Figure 2.1. The age distributions of the pateints should be considered when analysing the tissue samples for structural difference because breast tissue changes with age. Younger women have dense breasts with a high amount of fibrous tissue whereas older women have less dense and fibrous breast tissue and much more fat and adipose tissue. It is known that menopause causes changes in the tissue components within the breast. Out of the 114 patients who had their medical charts available, only 67 patients had their menopausal status recorded: 43 patients were post-menopause and 24 patients were pre-menopause. As expected, the ages between the post-menopausal patients (average = 62 ± 10 years) and the pre-menopausal patients (average = 33 ± 9 years) was significantly different ($p = 7 \times 10^{-18}$). The differences in the ages can be attributed to Australia's national breast cancer screening program, which provides free screening mammograms at two-yearly intervals for women aged 50-69 [6], resulting in more patients diagnosed with disease in this age group compared to other ages. Mammoplasty tissue samples were obtained from patients undergoing breast reduction or breast augmentation surgery, who are typically younger than

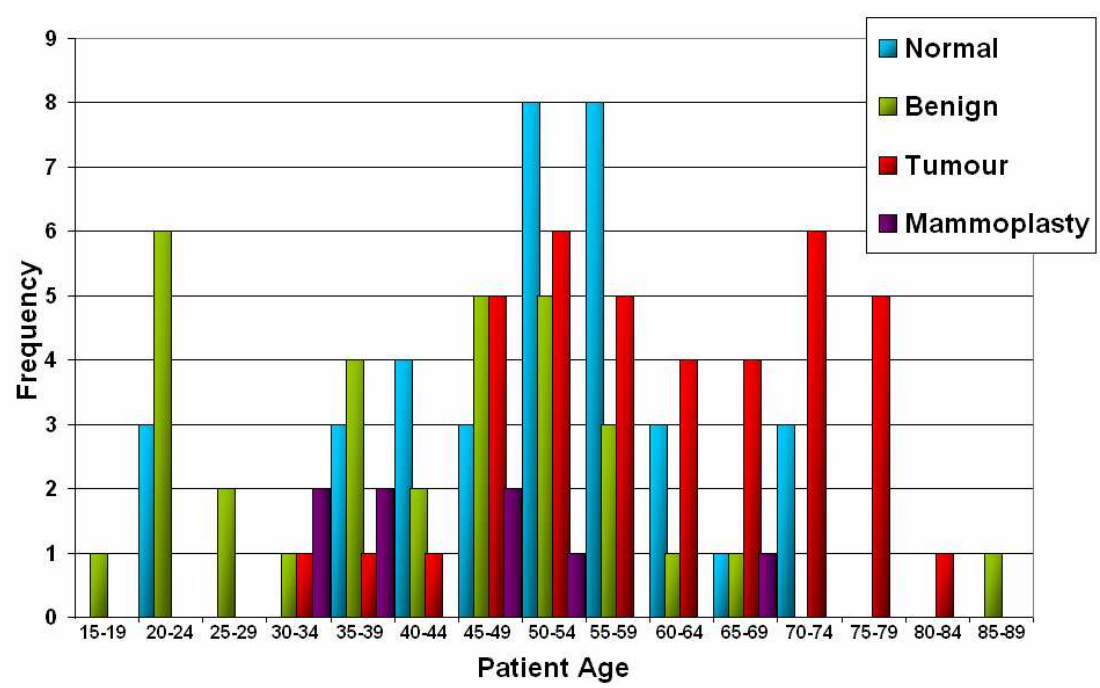


Figure 2.1: Histogram plot in 5 year bins of the patient ages for each of the four tissue groups. Only 114 patients (out of 130) had their age recorded.

Table 2.2: Statistics for the 114 patients who had their ages available for recording. Values in brackets represent the number of patients in each group that had their age recorded over the total number of patients in that tissue group.

Tissue Group	Average Age (years)	Maximum (years)	Minimum (years)	Standard Deviation (years)
Invasive Carcinoma (39/42)	61	84	30	13
Benign (32/36)	42	84	18	16
Normal (35/35)	51	71	21	13
Mammoplasty (8/17)	44	66	32	12

women who are diagnosed with breast cancer. The differences in the patients' ages, as well as their menopausal status must be considered when interpreting the results of tissue data, especially since structural information and tissue components were analysed.

2.1.1 Ethics Approval and Patient Consent

Ethics approval regarding collection and experimentation on human tissue was granted prior to commencing this study by both Monash University (Project 2003/021) and Southern Health Human Research Ethics Committees (No. 03094B). An extension of this approval was granted to 30 June 2010 (See Appendix A.3). Explanation of the study both verbal and written (see Appendix A.1: Participant Information Form), as well as any implications and contact information was addressed personally with each patient and any family or friends present with the patient. If the patient decided to take part in the study, the Informed Consent Form (Appendix A.2) was signed by the patient, family/friend present, and the consulting surgeon.

After the patient consented to participate in the study, the patient's hospital identification number (known as the Unique Reference (UR) number in the Australian public health system) was de-identified to the university researchers. Patient anonymity was maintained by only the consulting surgeon being aware of the both the identification codes used by the university researchers and the UR numbers.

2.2 Tissue Harvesting and Storage

Location of the primary lesion, as well as approximate size of the palpable lump, in relation to the breast was recorded on a sample sheet (red circle on Figure 2.2). The surgical procedure was observed in the operating theatre to ensure correct location of the palpable mass. When the surgical mass was removed from the patient by the operating surgeon, it was placed in a tissue bucket and labelled with the patient's hospital UR number. It was immediately taken to the pathology laboratory (located in the same building) where the consulting pathologist supervised core tissue removal. A biopsy gun (MEAD 16Gauge Needle Biopsy Gun) was used to remove the required samples from the mass. The core was a piece of tissue approximately 2 inches in length and 1 mm in diameter. Extracting the tissue samples with a biopsy gun kept the excised mass intact without mutilating the surgical margin, and so did not disrupt routine pathological analysis. The thickness of the tissue sample had to be small because as the tissue thickness increases, multiple scattering effects caused by photons re-scattered by succeeding collagen fibre layers also increases. This causes potential loss of structural information in the diffraction pattern [106].

Tissue was harvested "fresh" with no chemical contact, such as formalin or staining and excised within 20 minutes of vascular removal from the patient. Chemicals such as formalin affect the collagen structure in the tissue and cause degradation and dehydration, which would affect the SAXS structural analysis. Each tissue core was placed on a separate piece of aluminium foil which was folded to contain each sample, then flash-frozen in liquid nitrogen for storage in a -80ºC freezer until the time of imaging.

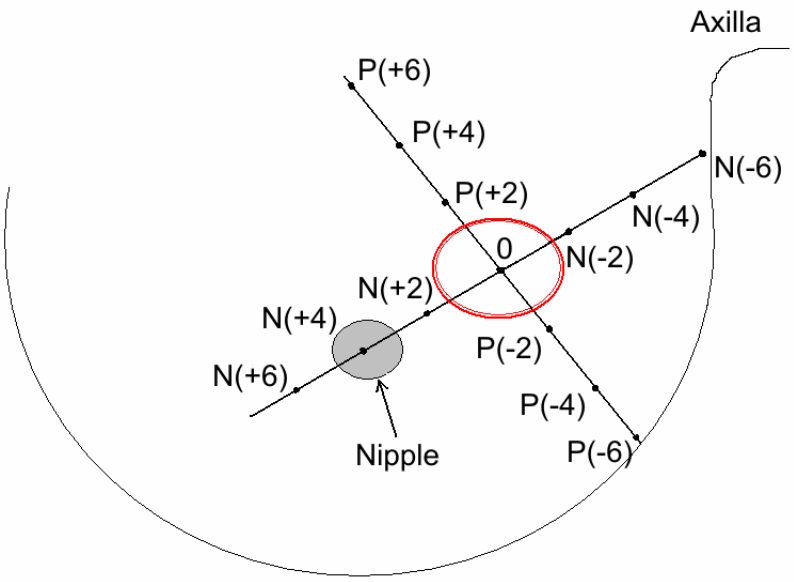


Figure 2.2: Sampling method from a mastectomy breast mass (typical: 122/130 patients). The lump or main tumour mass (red circle) was found and considered to be the origin. Samples towards the nipple (N+) and samples away from the nipple (N-) were taken at 2-3 cm intervals from the origin. Sampling was also done on the perpendicular axis from N, where P+ samples were taken superior and P- samples were taken inferior from the origin.

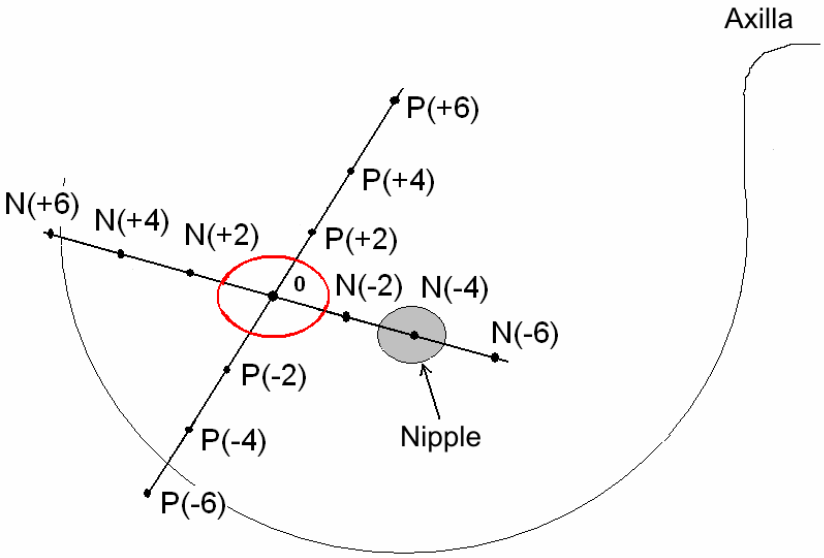


Figure 2.3: Sampling method from a mastectomy breast mass (non-typical: 8/130 patients). The lump or main tumour mass (red circle) was found and considered to be the origin. Samples towards the nipple (N-) and samples away from the nipple (N+) were taken at 2-3 cm intervals from the origin. Sampling was also done on the perpendicular axis from N, where P+ samples were taken superior and P- samples were taken inferior from the origin.

2.3 Collection of SAXS Tissue Samples

In total, 543 tissue samples from 130 patients were imaged using the synchrotron source and the laboratory source. 'Primary' site tissue samples were those taken at the centre of the lesion, and 'series' site tissue samples were taken at ~2 cm intervals away from the primary site. The total number of primary sites analysed was 196 and the total number of series tissues analysed was 347. This was the largest SAXS breast study undertaken at the time of writing.

Pairs of tissue samples (approximately 2 mm apart) were taken at the primary site of the palpable mass and, where possible, up to 6 cm away at 2 cm increments, radially outward along two perpendicular axes. The axes were determined based on the location of the tumour in relation to the breast nipple. The majority of patients (122/130) had tumours located laterally from the nipple (Figure 2.2). The N-axis was determined as the line between the nipple and axilla, and the P-axis was taken perpendicular to the N-axis, where the two axes intersect at the centre of the primary site. The (+2), (+4), (+6) notation represented samples taken towards the nipple (N-axis) or towards the superior part of the breast (P-axis) at each of the 2 cm intervals. The (-2), (-4), (-6) notation represented samples taken towards the axilla (N-axis) or towards to the inferior part of the breast (P-axis) at each of the 2 cm intervals. For example, a point labelled N(-4) represents a tissue sampled 4 cm from the centre of the primary lesion and towards the axilla direction.

There were only 8 patients (out of the 130 patients) who had palpable masses located medially from the nipple (see Figure 2.3). For these 8 patients, series tissue samples were taken towards the nipple and axilla, but labelled negative points (i.e. (-2), (-4), and (-6)). Tissue samples taken away from then nipple (i.e. medially) were labelled positive points (i.e. (+2), (+4), and (+6)). The P-axis points were taken similarly as described above, perpendicular to the N-axis tissue points.

The maximum distance at which a tissue sample was taken was ~ 6 cm away from the centre of the primary lesion; thus, a maximum of 13 tissue samples were collected from any one patient. Southern Health Pathology requested that tissue samples be excised no more than 6 cm from the centre of the lesion. The largest breast masses used for tissue sampling were from patients undergoing mastectomy, where most of the masses were approximately 12 cm in diameter. Patients who underwent wide-local-excision (WLE) had breast masses less than 8 cm in diameter.

Table 2.3: The number of patients imaged during each synchrotron experimental session and with the laboratory source. The diagnosis of the tissue samples taken at the centre of the lesion (primary sites) as diagnosed by histopathology is listed in relation to each experimental session. Note: primary site tissue samples represent one tissue sample from one patient (i.e. # patients = # tissue samples) taken at the centre of the patient's palpable mass.

	Synchrotron Source			**Laboratory Source**		
	Primary Site	**Diagnosis**		**Primary Site**	**Diagnosis**	
Session 1: Nov 2005	21	Inv. Carcinoma	8	66	Inv. Carcinoma	21
		Benign	6		Benign	16
		Normal	7		Normal	18
		Mammoplasty	0		Mammoplasty	11
Session 2: Feb 2006	29	Inv. Carcinoma	11			
		Benign	3			
		Normal	7			
		Mammoplasty	8			
Session 3: Feb 2007	80	Inv. Carcinoma	23			
		Benign	27			
		Normal	21			
		Mammoplasty	9			
Total	**130**	**Inv. Carcinoma**	**42**			
		Benign	**36**			
		Normal	**35**			
		Mammoplasty	**17**			

Table 2.4: The number of series site patients imaged during each synchrotron experimental session and with the laboratory SAXS camera (# Centre Points). The number of tissue samples taken at the centre of the lesion (primary sites) diagnosed by histopathology, with the number of tissue samples taken at each distance interval away from the centre point.

	# Centre Points (0 cm)	**Diagnosis**		**# Distance Sites**	**Distance from Centre**	
Synchrotron Source	56	Inv. Carcinoma	31	301	2 cm	158
		Benign	6		4 cm	87
		Normal	19		6 cm	56
Laboratory Source	4	Inv. Carcinoma	3	46	2 cm	15
		Benign	1		4 cm	16
		Normal	0		6 cm	15

Since tissue samples were taken after surgical removal of the mass, the excised tissue mass was difficult to handle and position, and it was consequently challenging to make accurate distance measurements from the primary site. Therefore, the measurements of 2 cm, 4 cm, and 6 cm were only approximate intervals (± 1 cm).

Table 2.3 lists the patients and tissue samples imaged for each of the three synchrotron experimental sessions and imaged using the laboratory source. A total of 130 patients were imaged with the synchrotron source, and 66 patients (who were a subset of the 130 patients imaged with the synchrotron source) were imaged with the laboratory SAXS camera. A smaller number of laboratory source tissues was imaged due to the time constraints to complete the project. Table 2.4 lists the number of patients who had series data imaged with the synchrotron source and the laboratory source. Only 56 patients had sufficiently large masses (11 mastectomy, 45 Wide Local Excision (WLE)) to obtain series tissue samples as described above. The number of samples in these 'series' patients varied according to the size of the mass, with WLE masses generally being much smaller than mastectomy masses (i.e. only 2-4 cm from the primary site). These 56 'series' patients had a total of 357 tissue cores sampled from their excised mass: 56 samples at 0 cm, 158 samples at 2 cm, 87 samples at 4 cm, and 56 samples at 6 cm.

The four patients who had series tissues analysed with the laboratory source were patients who had full mastectomy performed, so a full set of 13 tissue samples were available for these four patients to analyse[♯♯]. In order to compare the synchrotron source data with the laboratory source data, only the patients whose pairs of tissue samples were diagnosed the same (i.e. one of each pair for each X-ray source) were used. Only 28 patients (out of the 66) had the same histopathology diagnosis at the primary site with both the laboratory source and the synchrotron source: 6 invasive carcinoma, 6 benign, 5 normal, and 11 mammoplasty. Only these 28 patients were analysed together when directly comparing the synchrotron source data with the laboratory source data (Section 6.4). The heterogenous nature of breast tissue and the sampling error from biopsy contributed to the differences in tissue diagnosis between the tissue samples.

[♯♯] Two tissue samples were very fatty and could not be preserved for histopathological analysis.

2.4 Tissue Handling During an Experiment

Just before imaging, the tissue core was defrosted[§§] and placed into a 1 mm diameter glass capillary tube (1.0 outside diameter Special Glass capillaries, Wolfgang Muller Glas Technik, Germany) with saline solution to maintain hydration. Small angle scattering was performed on water-filled capillaries as well as empty (air) capillaries to verify that scattering from the glass capillaries was negligible compared to the scattering from the tissue sample.

In order to insert the tissue sample into the capillary tube, the bottom end of the capillary was cut and the tissue sample was placed into the top of the capillary tube. The tissue sample was inserted carefully into the capillary tube with the aid of a thin wire then hydrated by adding buffered saline with the aid of a small gauge syringe and needle into the tube. Both ends of the capillary tube were sealed with wax to avoid dehydration and to maintain the position of the tissue sample within the tube. Biopsy cores were cylindrical in shape and typically 1 mm in diameter and 2 cm long, which made it easy to place the tissue cores into the 1 mm diameter capillary tubes.

After the sample was loaded into the capillary tube, they were placed into the sample holder and aligned with the radiation beam and imaged using SAXS to obtain a diffraction pattern. See Section 3.4 for more detail on SAXS image acquisition and the sample holder. Once image acquisition was completed, the tissue was removed from the capillary tube by breaking both ends of the tube and pushing the tissue out gently with a wire. The tissue sample was fixed in formalin for preservation until it could be embedded into paraffin blocks for further histopathological analysis (see Sections 2.5 and Appendix B).

2.5 Histopathological Reviews

The current gold standard for breast diagnosis is having a pathologist view and grade/type a tissue sample of the lesion using standard laboratory histopathological techniques. For this reason, the SAXS data was compared against this as the primary diagnostic comparison.

The standard procedure for histopathology is to view the tissue specimens using a light microscope after processing, sectioning (sliced to a 2 μm thickness), and mounting onto a slide. All 543 tissue samples imaged with both synchrotron

[§§] The tissue was allowed to defrost naturally – at room temperature, this took ~30 seconds.

radiation and the laboratory SAXS camera were prepared by Histology Laboratory Services at Monash University. Preparing the tissue samples for pathological diagnosis included embedding the tissue cores into paraffin blocks, sectioning the tissue samples and mounting them onto slides, and staining the specimens in order to make the cells visible for histopathological diagnosis (see Appendix B). The basic nature of histology slide staining is to stain the slide with two or more contrasting agents that highlight specific areas with one colour, and leave a counterstaining background colour. The routine stain for breast lesions and cell abnormalities is the hematoxylin and eosin (H&E) stain. This stain uses two separate dyes, one staining the nucleus and the other staining the cytoplasm and connective tissue. Hematoxylin is a dark purplish dye that will stain the chromatin within the nucleus, leaving it a deep purplish-blue colour. Eosin is an orange-red dye that stains the cytoplasmic material, the connective tissue and collagen, and leaves an orange-pink counterstain. This counterstain acts as a sharp contrast to the purplish-blue nuclear stain of the nucleus, and helps identify other entities in the tissues such as cell membrane, red blood cells, and fluid. Appendix B shows examples of tissue samples and the types of characteristics that pathologists look for when diagnosing specimens. In order to maintain consistency, a single pathologist (Simon Nazaretian, Department of Pathology, Monash Medical Centre, Southern Health, Clayton) reviewed the tissue samples.

Overall diagnosis pertaining to the tissue core was recorded along with the percentage of each tissue type present within each core observed. Histopathologic review of these samples was difficult because the tissue material presented with damage (most likely due to the freezing and unfreezing process; radiation damage may also have occurred but was not directly observed during SAXS imaging) as well as artefacts caused during the process of removal from capillaries, blocking and staining. Figure 2.4 shows glass artefacts in a tissue slide, and Figure 2.5 shows the folding of the tissue slice which made it difficult to observe proper percentages of tissue typing within the sample. However, reviewing of the tissue samples was consistent and care was taken to diagnose according to standard pathological reviewing (see Appendix B) [19]. The pathologist was blinded to any information pertaining to the tissue samples or patient history or medical treatments during review.

Tissues were diagnosed into four groups: invasive carcinoma, benign, normal, and mammoplasty. Appendix B shows examples of the tissue types examined in this study and describes some histopathological features from which the pathologist makes a diagnosis. A tissue core was labelled as invasive carcinoma if

there were any carcinoma cells present in the sample (typically >5% invaded). Benign tissues were those diagnosed by the pathologist with distinct benign features: fibroadenoma, cellular fibrosis, fibrocystic changes. Tissues diagnosed as being 100% normal tissue architecture by the reviewing pathologist were deemed normal, however, this tissue grouping was later separated into two groups: Normal and Mammoplasty. Normal tissue was tissue deemed 100% normal by the pathologist, but the sample was taken from a patient with known presence of disease. Some primary sites, where the biopsy was believed to have been taken from the palpable lesion, contained normal tissue, and this can be ascribed to either heterogeneity of tissue within the palpable mass or sampling error. However, these normal tissue samples were taken within 1 cm of the primary lesion site (well within the surgical margins of excision where macroscopic disease was believed to be present). Mammoplasty tissue samples were those diagnosed as normal, but from those patients who underwent breast reduction or breast reconstructive surgery (i.e. no presence or history of disease). Mammoplasty tissue samples were taken within 5 cm of the nipple and excised by a plastic surgeon. Mammoplasty tissue samples were confirmed as normal by the pathologist, as well as with the surgical pathology reports and patient history. Figure 2.6 shows examples of the four tissue types.

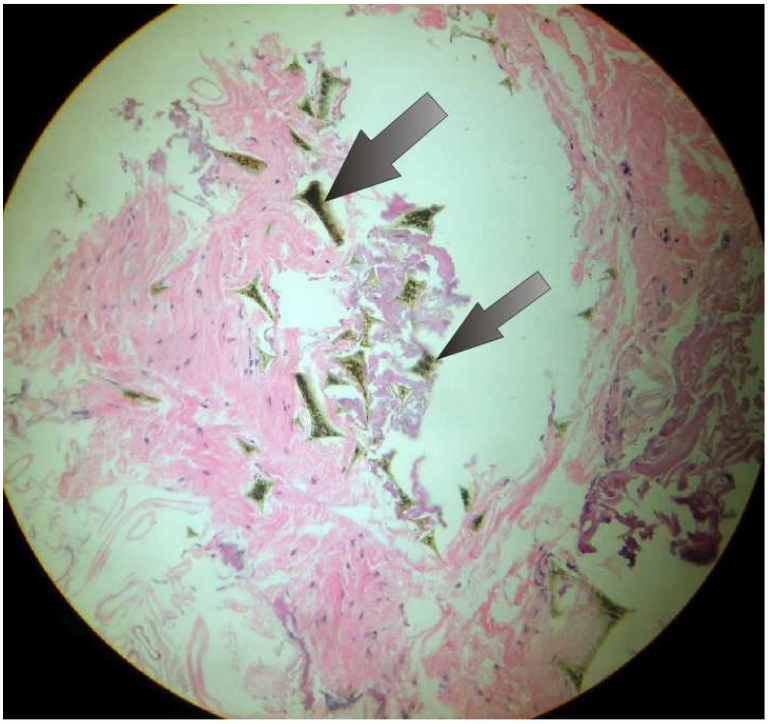


Figure 2.4: Tissue sample showing glass artefacts from shards of glass from the capillary tube that was mixed in with the tissue. The yellowish-brown pieces of glass are indicated by arrows. The glass artefacts affect proper cutting and staining of the tissue sample and deter from the pathologist's review of the sample.

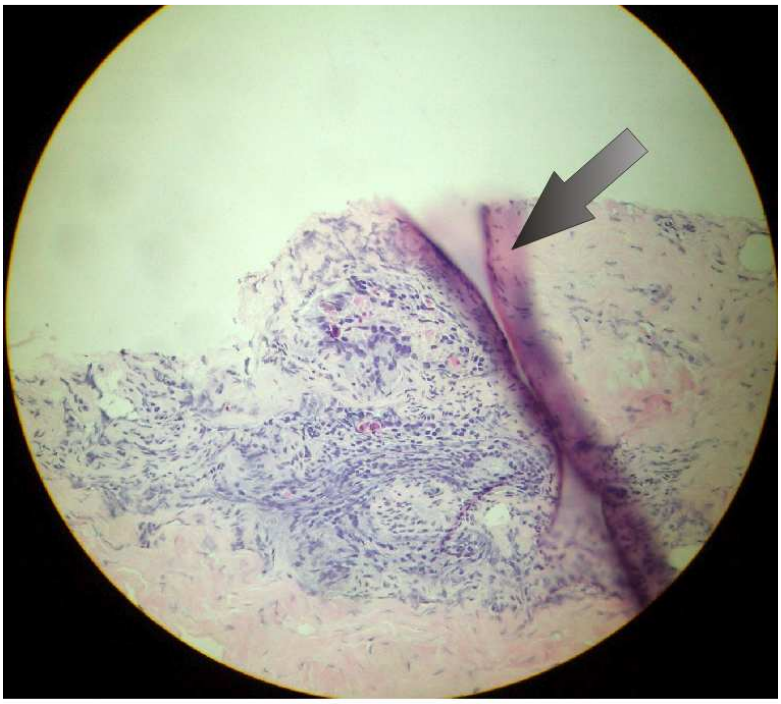


Figure 2.5: Tissue sample illustrating folding of the tissue sample (arrow). Staining and accurate analysis of the folded area becomes difficult.

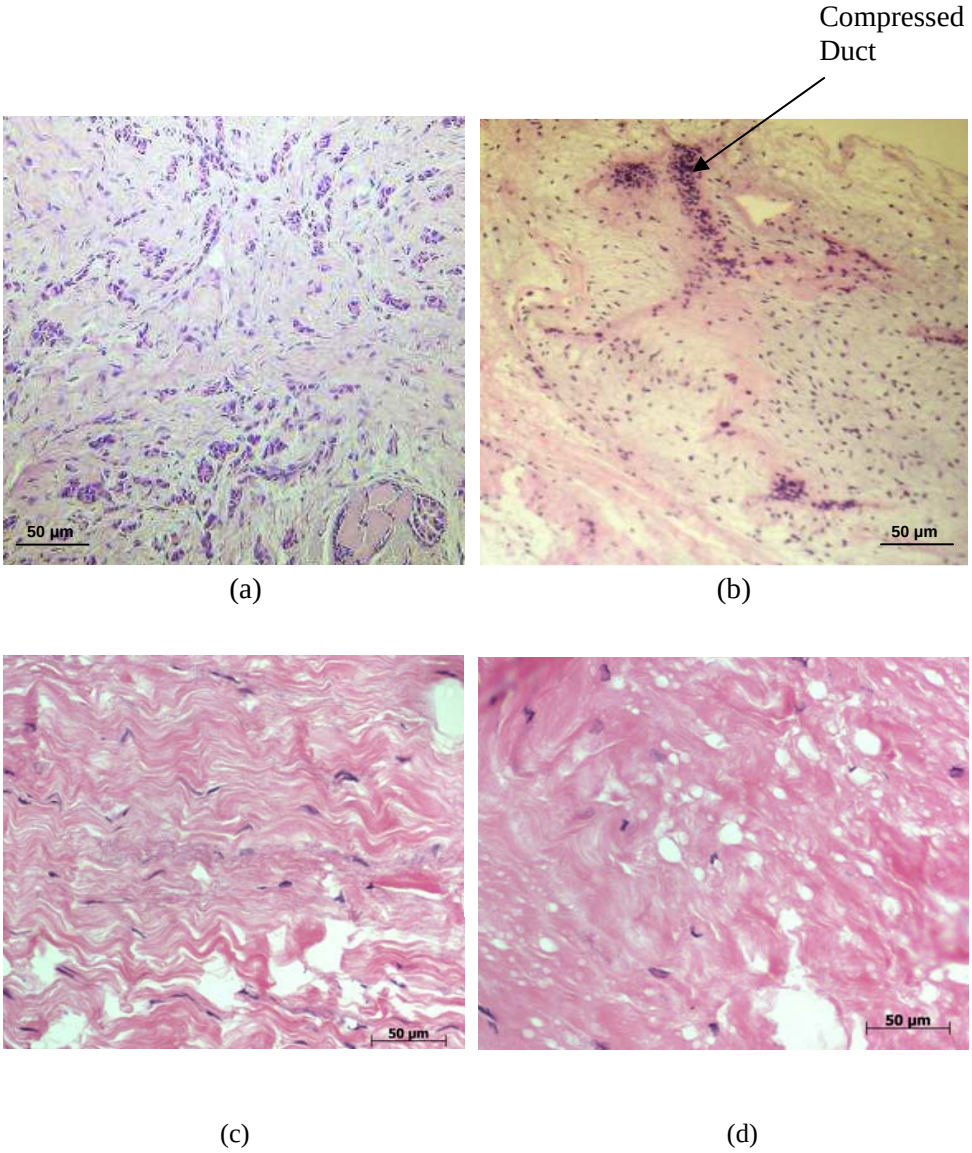


Figure 2.6: Examples tissue samples after H&E staining: (a) invasive carcinoma (ductal type) with atypical cells infiltrating myxoid fibrocellular stroma as small groups, tubules, single-files and occasional single cells (b) fibroadenoma (benign); there is a lobule with a compressed duct surrounded by a myxoid fibrous stroma, evidence of dense fibrosis (c) normal (fibrous stroma); (d) mammoplasty (fibrous stroma).

2.6 Conclusion

Tissue samples were collected from 130 patients over 30 months, with a total of 543 tissue samples imaged with a synchrotron source and the laboratory source. This was the largest SAXS breast tissue study undertaken at the time of writing this thesis, with a diverse number of diseased groups and tissue samples taken over a large area of the breast mass. The patient ages showed significant differences between the tissue groupings as did the ages between patients who were pre-menopausal and post-menopausal. Therefore, the age of patients must be considered when analysing the results between tissue groups (Chapters 4-6). The experimental design and equipment used for obtaining the SAXS images with both the synchrotron source and the laboratory source are described in the next chapter (Chapter 3). Image processing, feature extraction, and statistical methods are also explained.

Chapter 3

Small Angle X-Ray Scattering: Experimental Methodology

This chapter describes the two X-ray sources used for the collection of SAXS data: a synchrotron source (Section 3.1) and a laboratory source (Section 3.2). An ideal X-ray beam for SAXS would have a well-collimated X-ray beam with a small cross-sectional area (i.e. point focus), since the diffraction patterns in all directions are equally important. A well-focussed X-ray beam gives well-focussed images, where a large spot size can dramatically increase the peak widths (i.e. smearing effects) [100]. The main advantages of using a synchrotron source is its intrinsic high brightness and narrow angular divergence, which results in high quality SAXS data. Section 3.1 describes the experimental set-up and characteristics of the SAXS beamline used at Station 2.1 at the Synchrotron Radiation Source (SRS) Daresbury. Access to a synchrotron facility is limited and is not a feasible approach to implementing a SAXS technique clinically, so a laboratory source was also used for examining breast tissue samples with SAXS and is described in Section 3.2. Section 3.3 describes the preliminary processing and normalisation of the data to make it suitable for further analysis and comparison, not only between experimental sessions, but also between each sample. The data obtained using the synchrotron source will be presented in Chapters 4 and 5, and the data obtained using the laboratory source will be presented in Chapter 6.

3.1 Experimental Set-up: Synchrotron Source

The initial SAXS breast tissue experiments were performed in 1999 at the Daresbury Synchrotron Radiation Source (SRS)[***] in the United Kingdom by Lewis *et al*. [77], where the experimental protocols and tissue preparation methods were first developed. The SRS Daresbury was also used for imaging the breast tissue samples used in this thesis and the same experimental protocols and tissue preparation methods used by Lewis *et al*. [77] were also implemented for the work presented in this thesis.

In a synchrotron, the electrons circulate with a speed close to that of light and are kept in a closed orbit by a series of bending magnets. The electrons experience acceleration when their direction is changed and they emit electromagnetic radiation. Due to relativistic effects as the electrons pass through a bending magnet, the radiation is emitted as a fan of very well-collimated radiation, tangential to their pseudo-circular orbit.

The main properties of the X-ray scattering beamline at the SRS Daresbury Station 2.1 are listed in Table 3.1. A schematic of the beamline is shown in Figure 3.1. Station 2.1 was a fixed wavelength station (λ = 1.54 Å, Energy = 8.050 keV) using a reversed mirror-monochromator arrangement which resulted in a reduced heat loading of the mirror, thus producing a highly stable focal spot [122]. The beamline optics were able to produce a focal spot, collimated with the use slits, of 0.5 mm x 0.5 mm at the sample position. The camera length was variable between 1.00 m and 8.25 m that provided a d-spacing range from 0.8 nm to 150 nm. The detectors that were used to collect the X-ray scattering data for the experiments on station 2.1 were 2-dimensional gas filled, wire, delay line detectors [76]. This detector, with an active area of approximately 200 x 200 mm, permitted collection of 2-dimensional data. For these experiments, a camera length of ~6.3 m was used to produce data within a q-range of 0.1-0.6 nm^{-1}. The data for this project was collected over three experimental sessions: November 2005, February 2006, and February 2007. Travel funding to the facility to perform the experiments was provided by The Access to Major Facilities Program[†††]. Table 2.3 in Section 2.2 lists the tissue samples imaged during each of the experimental sessions.

[***] The Synchrotron Radiation Source at the Science and Technology Facilities Council's Daresbury Laboratory in Warrington, UK, closed on August 4th, 2008.
[†††] This program is a component of the International Science Linkages Programme established under the Australian Government's innovative statement, *Backing Australia's Ability*.

Table 3.1: Properties of Station 2.1 at the SRS Daresbury [122]

Generation	2nd
Energy	2 GeV
Ring Current (max)	300 mA
Station 2.1	SAXS/WAXS
Energy	8.050 keV
Wavelength	1.54 Å
Horizontal Focussing	Triangular monochromator
Vertical Focussing	Plane mirror
Sample-to-Detector Distance	1-8.25 metres
Focal Spot	0.3 mm x 2.5 mm
d-spacing range	0.8 - 150 nm
Flux	1.4 x 10^{13} photons/sec
Intensity	9 x 10^{13} photons/sec/mm^2

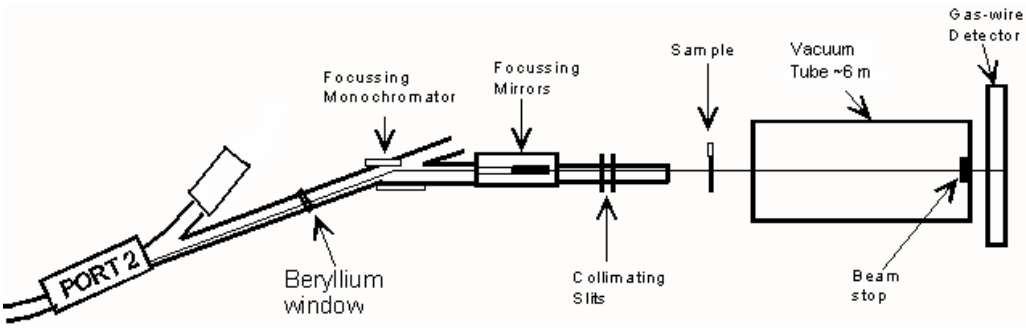


Figure 3.1: Schematic of the SRS Daresbury experimental set-up at Station 2.1. The synchrotron beam was utilised from Port 2 within the storage ring and then focussed using a monochromator (Crystal Type Ge III) and mirror before being collimated further by slits when reaching the tissue sample. A semi-transparent beam-stop was placed at the end of the tube in order to block the direct beam at the centre of the multi-wire gas detector (see Figure 3.3) (sample-to-detector distance ≈ 6.3 m).

Tissue samples (placed in glass capillaries) were placed in a sample holder ~6.3 m from the detector. The sample holder was able to hold up to 5 glass capillaries (with tissue samples, vertically) 0.5 cm apart (see Figure 3.2). By having 5 samples in the holder, the hutch door at Station 2.1 only had to be opened once every sample set, which expedited data acquisition compared to loading one sample at a time. Stepper motors were used to align each sample horizontally with the beam and position and oscillate the sample vertically through the beam during image acquisition. Small angle scattering from the glass capillaries was verified to be negligible in comparison with scattering from the tissue sample provided the beam

intersected the central vertical axis of the capillary and scattering from the capillary walls was avoided. Samples were either oscillated (experimental sessions November 2005 and February 2006), or raster scanned by integrating a series of images at equally spaced increments over the length of the sample (experimental session February 2007). The tissue samples were kept in air because they were easier to handle and to maintain hydration of the samples during the experiment, rather than in a vacuum environment. The path of scattered radiation from the sample is predominately in vacuum in order to avoid scattering from air interfering with the data. This was accomplished by means of a vacuum tube between the sample and detector (shown in Figure 3.1). There is a small air path around the capillary of less than 5 cm. In order to minimize parasitic scattering, beryllium windows with an X-ray transmission at 8 keV = 95% separated the high vacuum of the storage ring and subsequent optical elements [80].

A beam stop is typically inserted between the sample and the detector in order to protect the detector from the direct X-ray beam. The beam stop used at the SRS Daresbury was rectangular in shape and suspended with wire near the detector end of the vacuum chamber (Figure 3.3). The same beam stop was used for all three experimental sessions at the SRS Daresbury. The beam stop consisted of a hollow rectangular lead casing 10 mm x 20 mm with a wall thickness of ~ 1 mm (shown in Figure 3.3). The lead shields the detector from any parasitic scatter in the beam stop. The beam passed through the centre part of the rectangle with the largest face orthogonal to the incident beam. To ensure that the transmitted intensity was within an acceptable dynamic range of the gas-wire detector (discussed below), the side of the beam stop that was incident to the beam was covered with nickel, 0.5 mm thick, which attenuated the X-ray beam by 80.2%. The beam stop was suspended with nylon thread at the end of the scattering vacuum tube and a schematic is shown in Figure 3.3. The beam stop was positioned at the centre of the detector in order to obtain 2-dimensional scattering data in both the meridional and equatorial directions (as described in Section 1.3).

The beam stop was semi-transparent to allow for measurement of the transmitted intensity of the beam (i.e. the incident beam intensity less the absorbed intensity due to the sample). Knowing the transmitted intensity of the beam allows for comparisons between each tissue since the variations in absorption due to the thickness of the tissue sample, variations due to electron beam decay, variations in the downstream optical elements, and glass capillary absorption can be corrected for.

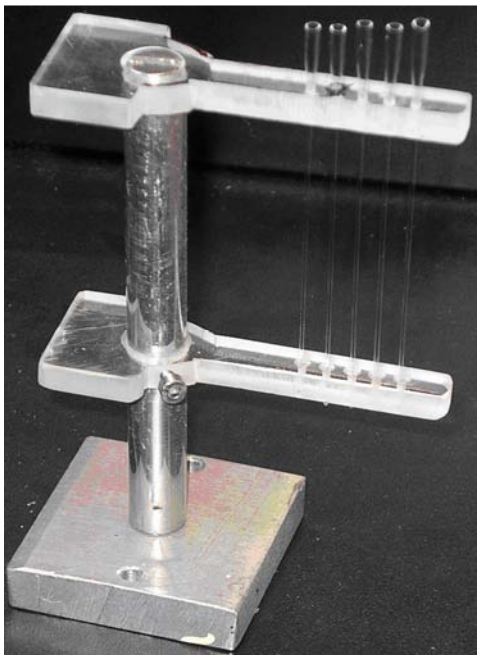


Figure 3.2: The sample holder used to align the tissue samples in glass capillaries (shown empty). The sample holder could hold up to 5 glass capillaries, 0.5 cm apart, and were positioned in the beam for individual data collection using the sample stage motors. (Printed with permission from Dr. Adam Round, 2009 [102])

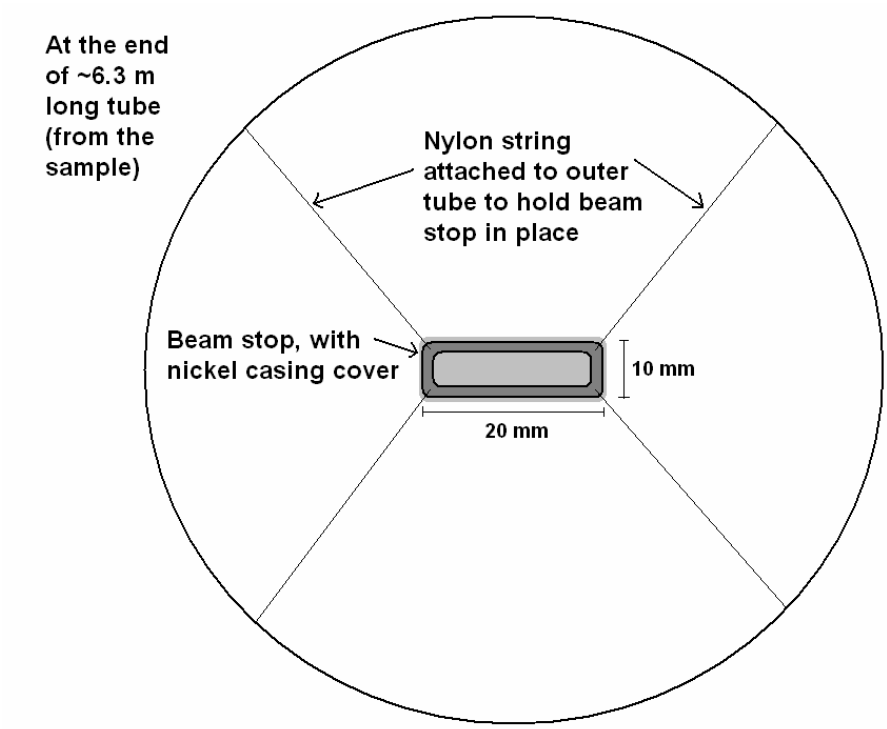


Figure 3.3: The rectangular beam stop, with the nickel cover placed overtop, was suspended at the end of the ~6 m long scattering vacuum tube (sample-to-detector distance ≈ 6.3 m) with nylon string at the four corners of the beam stop. The beam goes into the page, through the centre of the beam stop.

A multi-wire gas proportional counter was used at the SRS Daresbury. A detailed description of the specifications of this detector have been described in [76], but is summarized here. Gas-wire detectors are useful for high flux sources such as synchrotrons, as they have high count rates, high dynamic ranges, and excellent signal-to-noise ratios due to their photon counting nature. A major advantage of using a detector that is based on a photon counting system is that there is no noise associated with the image. In order to detect photons across its input face, the detector uses a gas volume for detecting ionisation products. The absorption of an X-ray by the gas atom initialised the detection event of a photon. A photo or Auger electron results from this event, from which a cloud of ions and electrons drift in opposite directions under the influence of an applied electric field. The detector used at the SRS Daresbury was filled with Argon gas. Each ionisation pair yielded 250 electrons per 8 keV photon [76]. The detector was 200 x 200 mm with pixel sizes of 0.38 x 0.38 mm.

3.2 Experimental Set-Up: Laboratory SAXS Camera

In this section, the properties of the laboratory SAXS set-up are described. The laboratory SAXS camera was purchased by Monash University from Osmic Inc. (now part of the Rigaku Americas Corporation, Texas, USA 2008) and is a compact apparatus compared to a synchrotron facility. Disadvantages with the laboratory camera system over a synchrotron source are the larger divergence of the beam and the lower flux. These differences are due to both lower intrinsic brightness of the fixed anode source as well as the difficulties associated with focussing a tube source and the flux reduction that inevitably occurs when collimation is utilised.

Figure 3.4 shows a photo of the laboratory SAXS camera as installed at the Rigaku factory before delivery at Monash University. The X-ray source consisted of a microfocused electron beam on a copper target (Bede, Jordon Valley Semiconductors Inc., 2003), with a wavelength λ = 1.54056 Å and Energy = 8.050 keV. The properties of the laboratory SAXS system are listed in Table 3.2.

The laboratory SAXS camera uses a microfocusing copper X-ray source coupled with a Confocal Max-Flux^{TM} (CMF) (Rigaku Corporation, Texas, USA, 2008) optics and provides a monochromatic focused hard x-ray beam of λ = 0.154

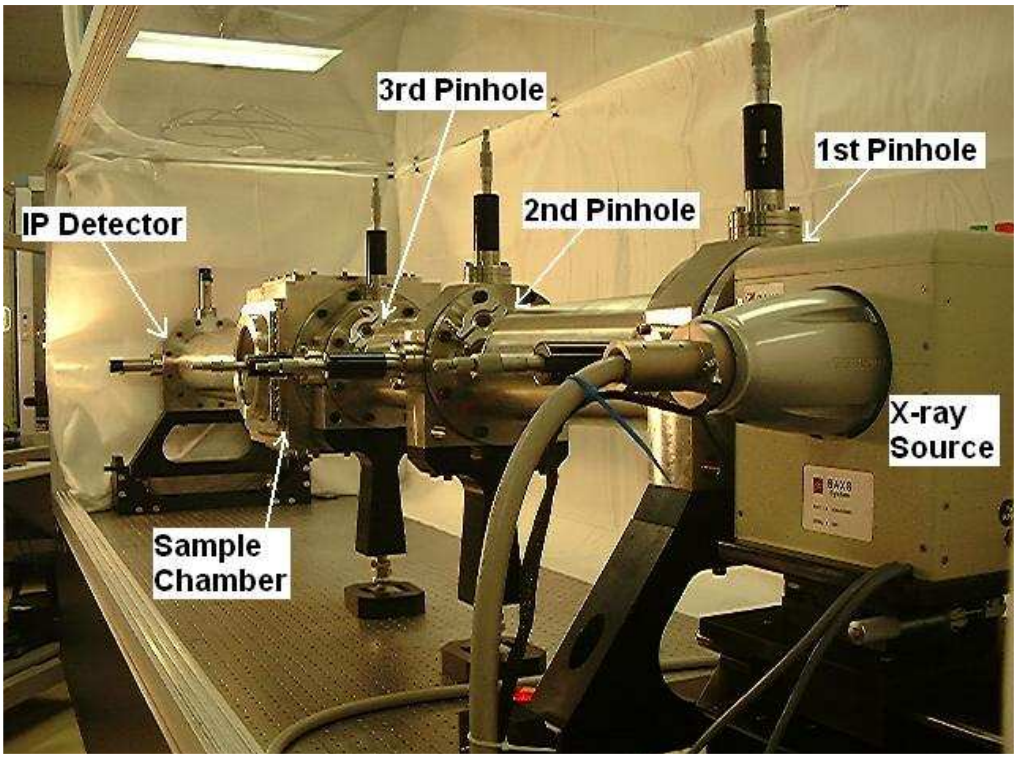


Figure 3.4: The laboratory SAXS camera set-up, with the X-ray source, three pinholes, sample chamber, and detector.

Table 3.2: Properties of the SAXS Laboratory Camera based on a 3-pinhole design in high-flux mode.

Energy	8.050 keV
Current	0.66 mA
Operating Voltage	45 kV
Wavelength	1.54056 Å
Horizontal Focussing	Confocal Max-Flux Optic
Vertical Focussing	Confocal Max-Flux Optic
Sample-to-Detector Distance	~1.6 metres
Focal Spot (sample)	0.33 mm FWHM
d-spacing range	< 89.5 nm
Flux	2.8×10^7 photons/sec

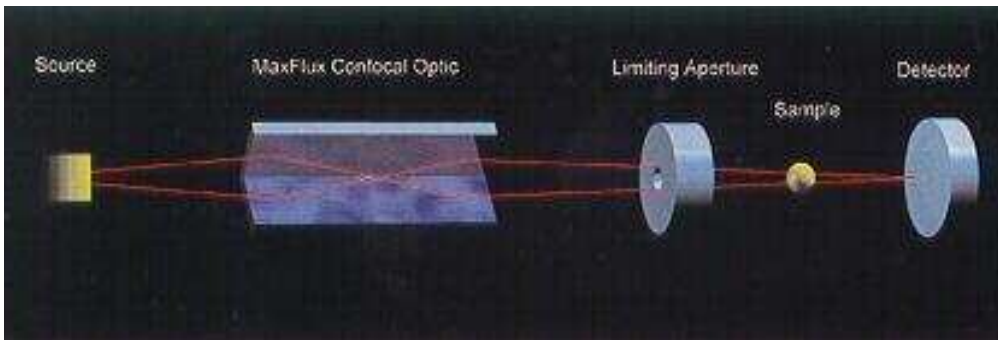


Figure 3.5: Confocal Max-Flux Optic 2-D reflection monochromator system, shown for 2-pinhole configuration. (Printed with permission from Rigaku Corporation, Texas, USA, 2008: **http://www.osmic.com/MSC/MSC1/MSC.htm**).

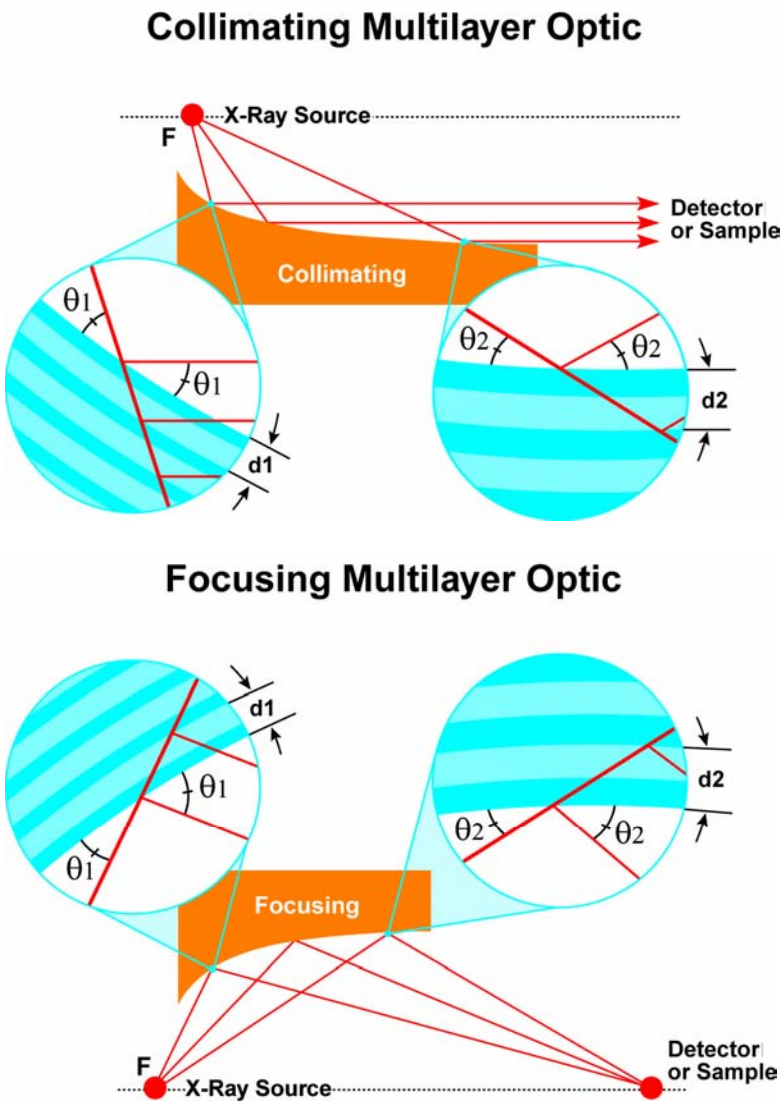


Figure 3.6: The graded d-spacing multilayer principal of the Confocal Max-Flux optic. The collimating and focussing optics are realised by curving the surfaces, which satisfies Bragg's Law at every point. (Printed with permission from the Rigaku Corporation, Texas, USA, 2008 [25])

nm and Energy = 8.050 keV by combining a side-by-side approach as in a Kirkpatrick-Baez optic with high performance multilayers [66] (see Figure 3.5). This multilayer optic effectively focuses the beam at the sample positionwhilst simultaneously making the beam monochromatic. The d-spacing of the multilayer system is varied non-linearly along the optic from 2.84 – 3.12 nm, with a value of 3.0 nm in the middle, which allows for Bragg's law to be satisfied at every point along the optic (see Figure 3.6) [25]. This results in a mean angle of incidence of 1.4^{o} [24, 25, 66]. The advantage of using a side-by-side system is that both optics are the same distance from the X-ray source, and thus capture a larger view of the source resulting in a higher flux beam.

The camera was configured into a 3-pinhole design: the first pinhole defines the beam, the second pinhole defines the beam and also blocks the parasitic scattering from the multilayer optic, and the third pinhole blocks the parasitic scattering from the second pinhole, resulting in a very clean background. For the work presented in this thesis, the laboratory SAXS system was set-up to utilise the high-flux mode, rather than the low-q mode option. In the high-flux configuration, the first, second, and third pinholes were larger in diameter (0.6 mm, 0.3 mm, and 0.95 mm, respectively) compared to the low-q configuration (0.4 mm, 0.2 mm, and 0.65 mm, respectively). The larger diameter pinholes in the high flux configuration allowed for more rapid data collection, which was preferred, compared to the low-q configuration. The 3-pinhole configuration is well suited for imaging samples with low scattering power and high d-spacing.

The sample chamber is shown in Figure 3.7, which housed the sample holder and motorised positioning stage. The sample holder was a similar apparatus used at the SRS Daresbury where up to five tissue samples (in glass capillaries) could be placed. The motorised positioning stage moved in two-directions (vertical and horizontal), so that the tissue samples could be aligned with the beam for imaging. Sample alignment perpendicular to the beam direction was verified with a CCD camera by observing the attenuation of the direct beam with respect to the beam stop (the beam stop was centred with the direct beam, so when the beam was masked entirely by the beam stop, the sample was centred with the beam). During exposure, the tissue samples were oscillated vertically through the length of the tissue sample, as was done similarly with the synchrotron experiments. The scattered radiation passed from the sample to the detector through an evacuated tube. At the end closest to the sample, the tube forms a cone that protrudes into the sample chamber to permit samples being maintained in an air environment, whilst

minimizing the distance of this air path. The tissue samples were kept in air in order to make changing the tissue samples easier between imaging. Also, because of the delicate nature of wet samples in glass capillary tubes, the pressure associated with a vacuum chamber would damage the glass capillaries that held the tissue samples in place during the experiment. The resulting sample-to-detector distance with the laboratory SAXS system was ~1.6 m.

As with the SRS Daresbury experimental set-up, a beam stop was also used with the laboratory SAXS camera to block the direct beam from reaching the detector. The beam stop was semi-transparent to allow for measurement of the transmitted intensity of the beam. Figure 3.8 shows the suspension of the beam stop with the laboratory source, which was positioned near the end of the vacuum tube (see Figure 3.7), close to the detector. The beam stop was designed as a semi-transparent lead cylinder, where several different thicknesses of aluminium could be used for attenuating the X-ray beam. For all of the data collected with the laboratory SAXS camera, the semi-transparent beam stop thickness was kept constant at 0.80 mm, which allowed for calculation of the transmitted intensity of the beam for an exposure of 60 minutes. A thickness of 0.8 mm of aluminium (attenuates the beam by 89.7%) provided a low enough intensity so the Imaging Plate (IP) detector would not saturate, but a high enough intensity to provide good statistics (signal-to-noise ratio) for normalisation purposes. Appendix D explains in more detail the design of the semi-transparent beam stop used with the laboratory source.

Images were acquired using 20 cm x 25 cm Image Plate (IP) detectors (Fujifilm BAS-MS2025, Fuji Photo Film Co., Ltd. Tokyo, Japan). An in-depth description of the IPs used for this project is presented in Appendix C. The IP is a flexible plastic plate which is coated with photo-stimulable phosphor ($BaF(Br,I):Eu^{2+}$). The photo-stimulable phosphor stores a fraction of the absorbed X-ray energy it is exposed to by trapping electrons in Br^- and F^- vacancies. The IPs were read using a Fuji Film BAS-5000 IP Reader, which uses a laser to stimulate the trapped electrons, which release the stored energy as visible light. The emitted photo-stimulated luminescence (PSL) is proportional to the absorbed X-ray intensity [117]. The PSL is collected by a photomultiplier tube (PMT) by scanning the plate point-by-point in order to make the final image. Any residual image on the IP was then erased by irradiation with visible light, which allowed repeated use of the IP.

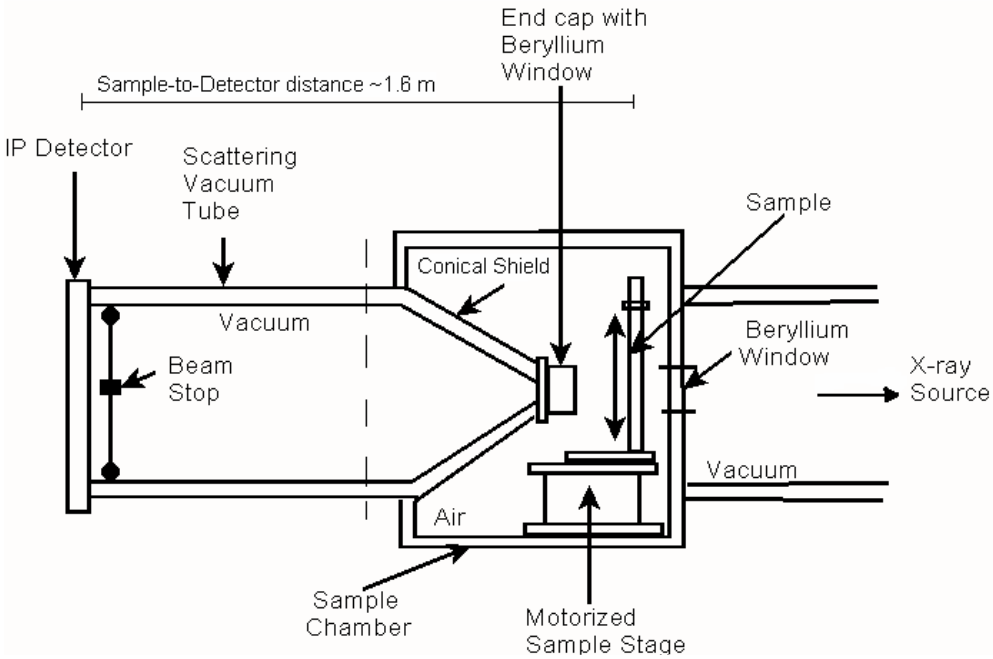
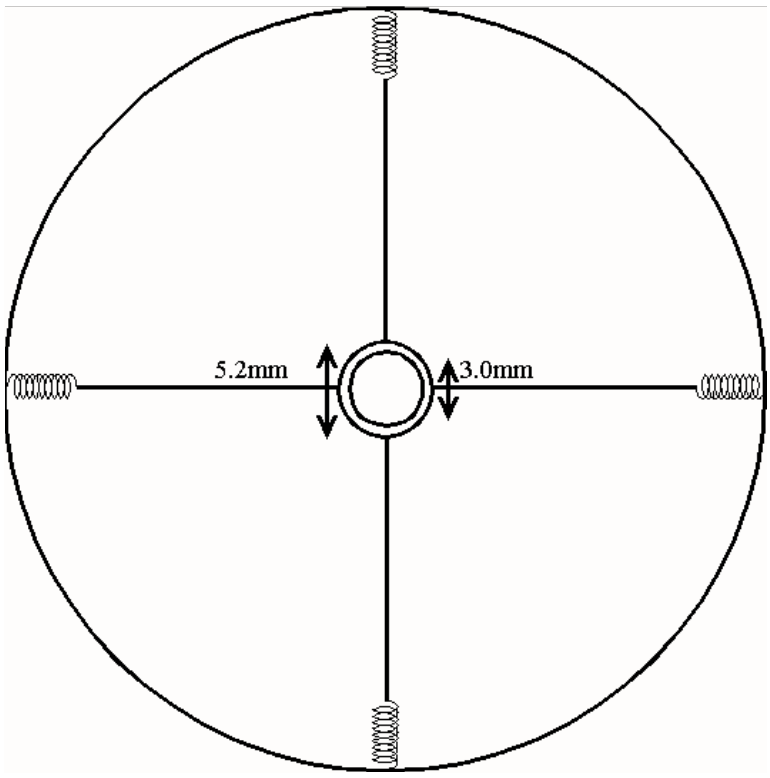
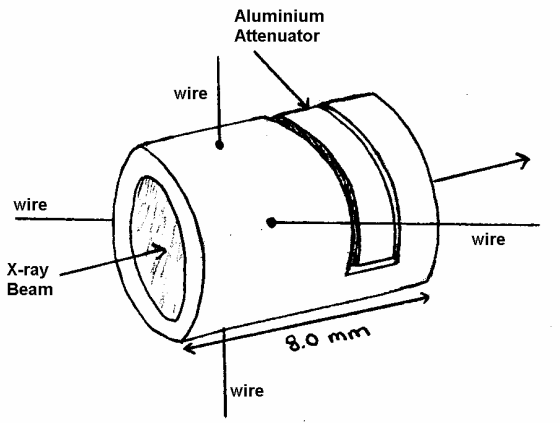


Figure 3.7: The laboratory SAXS camera sample chamber (not to scale). The sample holder could hold up to five glass capillaries and the motorized stage could move both vertically and horizontally for positioning the sample in the beam and for oscillating (vertically) the tissue sample during imaging. A canonical shield is placed on the left side of the sample chamber and attached at the beginning of the scattering vacuum tube. The beam stop was attached to the tube with thin wire and springs on the inside of the scattering vacuum tube (see Figure 3.8).



(a)



(b)

Figure 3.8: The semi-transparent beam stop used with the laboratory SAXS camera. (a) View of the beam stop looking directly at the source. The beam stop was suspended with thin wires attached to springs on the inside of the scattering vacuum tube. (b) Detailed view of the beam stop. The wires were glued into small holes on the surface of the beam stop, and the aluminium attenuator (0.8 mm thick) was positioned into the slot to block the direct X-ray beam.

An advantage of using an IP detector is that it can be reused many times over, making IPs very cost efficient. Images are obtained in a digital format, which makes it easy to manipulate and process the data. However, a disadvantage of using IPs is the long exposure times needed to acquire an image (60 minutes per sample in this study). There is also baseline noise associated with IPs since they are integrating detectors. Signals with low intensity disappear in the noise of the image plate, whereas the gas-wire detector can record low intensities without any difficulty [76]. A pixel size of 0.2 x 0.2 mm was used with the IP detectors for the work presented in this thesis.

3.3 Image Acquisition and Pre-Processing

The tissue samples were handled similarly in both the synchrotron source experiments and the laboratory source experiments. As discussed in Chapter 2, tissue samples were placed into glass capillaries and imaged with an X-ray source. After each sample was imaged, the tissue was removed from the capillary and placed in 4% formalin solution for preservation so that H&E processing of the tissue sample could be performed later (see Section 2.5).

Each image acquired at the synchrotron source was saved as two files: (1) the header file containing the time the image was taken, the image name, and other unique identifications, and (2) the 2D binary image data file. These files were then converted into one `.det` file using PCDetPak32 (Version 6.0, R. Lewis, Monash University, 2006) which contained all of the data from both files. The files acquired using the laboratory source were output by the IP reader as `.img` files with a header file containing the unique information for each image. These `.img` files were then converted to `.det` files using PCDetPak32. The `.det` file images were then ready for pre-processing. This is described in more detail in Appendix C.

Images acquired with both the SRS Daresbury and the laboratory SAXS camera required a series of pre-processing steps to be performed before parameters for further processing could be extracted. Image processing was performed using PCDetPak32. There were four processing steps that were performed in order to extract information from the SAXS images:

(1) Removal of wire structure (SRS Daresbury images only) to correct for detector artefacts.

(2) Calibration of the images for q-spacing.

(3) Normalisation of the intensity. This was necessary for comparison of images between experimental sessions and x-ray sources.

(4) Radial integration over the 2D images to obtain 1D data of the intensity as a function of the scattering vector q, from which feature parameters were extracted.

These steps are discussed here in more detail.

3.3.1 Removal of wire structure

Removal of the wire structure corrects for the inhomogenieties in the detector response introduced by the wires that form the active area of the gas filled detector used at the SRS Daresbury. Figure 3.11a shows a raw image prior to wire structure removal and Figure 3.11e shows the same image after wire structure removal.

The high-frequency content of the raw image was used to correct the intensity variations. First, the intensities of the raw image were integrated row by row in the vertical direction, and then repeated in the horizontal direction of the image, in order to obtain 1D profiles of the intensity. Figure 3.11b shows a 1D horizontal profile through the raw image data. This projection shows both the high frequency and low frequency components of the image. The projection was smoothed by averaging; the result is shown on a 1D profile in Figure 3.11c. The smoothing threshold was set to compensate for the wire structure: the size and shift of the position of the wires. The integrated profile was then divided by the smoothed profiles (Figure 3.9c) to obtain the high frequency components of the image, seen in Figure 3.9d. Lastly, this high frequency correction was used to correct the apparent wire structure in the original image by dividing each row of the raw image. A similar process was carried out in the vertical structure, resulting in the final image (shown in Figure 3.11e).

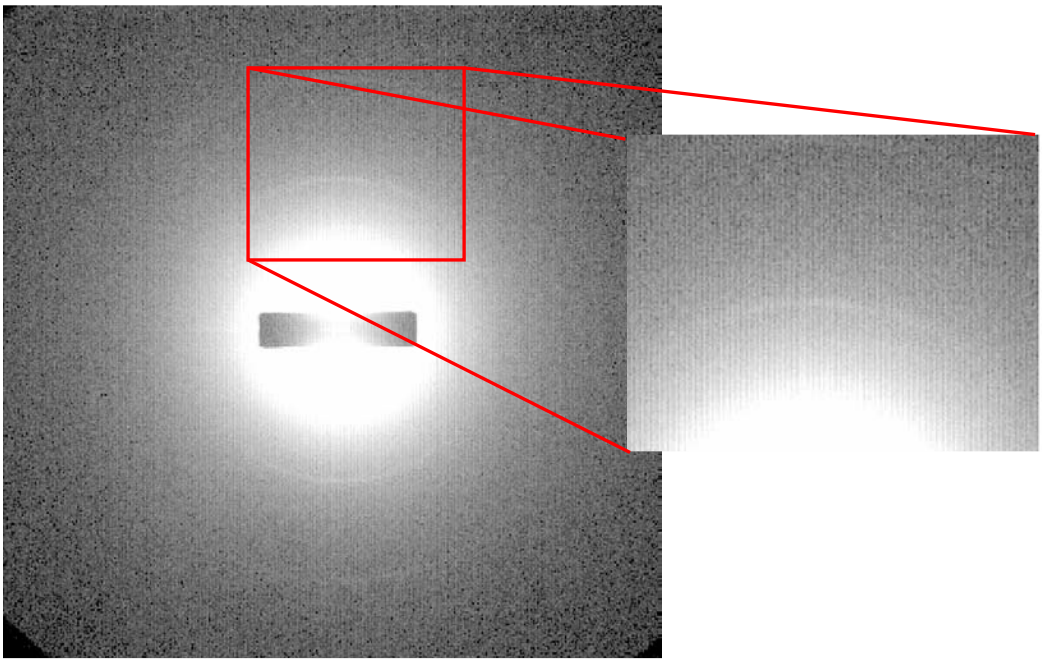
3.3.2 Calibration of q-spacing

For the synchrotron experiments, the sample-to-detector distance was ~6.3 m, which resulted in smaller q-range seen on the SAXS images (q = 0.1-0.6 nm^{-1}) compared to the larger q-range seen with the laboratory source images (q = 0.25-2.3 nm^{-1}) (sample-to-detector distance ≈ 1.6 m). At smaller q-spacings, the common biophysical standard used for calibration is type I collagen from rat tail tendon. The images from the synchrotron source were calibrated for q-spacing using rat-tail

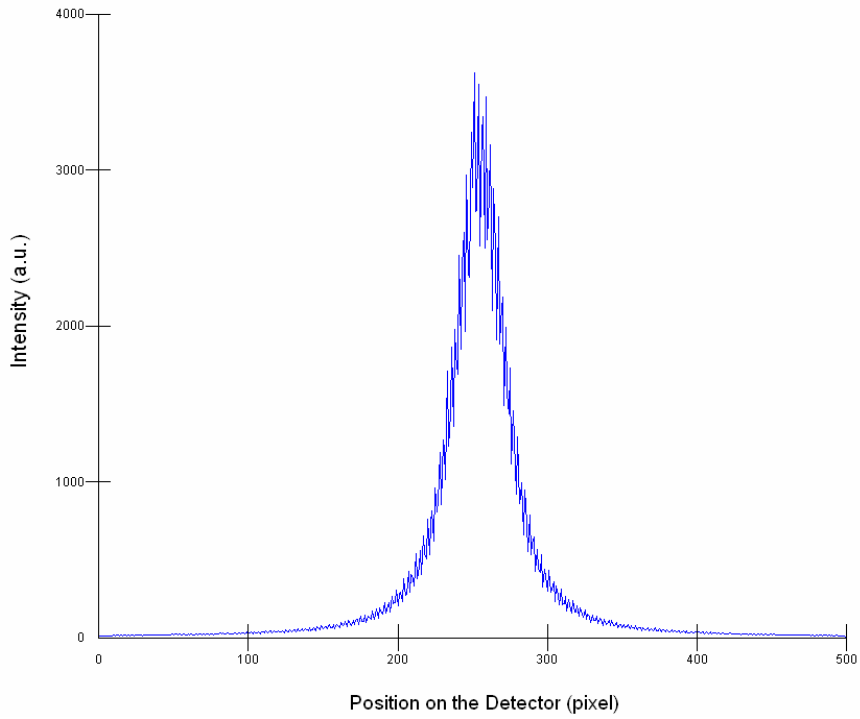
tendon. This also determined the exact sample-to-detector distance. The lattice constant d for rat tail collagen used for calibrating the synchrotron source images was 66.8 nm[‡‡‡] [28, 29]. Rat tail tendon was inserted into glass capillaries with buffered saline and imaged as was similarly done with the breast tissue samples. The exposure time for acquiring rat tail collagen images was 5 minutes, the same exposure time used for imaging the breast tissue samples with the synchrotron source. Rat tail images were taken at the beginning of each beam injection at the SRS Daresbury (once a day at 8:00am), or whenever the sample position was changed, to account for changes in the beam position or effective camera length. An example of a rat tail tendon image taken with the SRS Daresbury is shown in Figure 3.12a. Care was taken to ensure rat tail tendon was freshly excised and fully immersed in saline solution for each experiment, as the d-spacing is known to vary with hydration levels [57, 64].

With the three synchrotron experimental sessions, the rat tail tendon images were calibrated between each experimental session to ensure that the d-spacings were exactly the same. This was accomplished by overlaying all of the rat-tail SAXS images together and aligning their patterns before setting the known d-spacings. This allowed for comparison of the d-spacing for all of the SAXS breast tissue images for all three synchrotron experimental sessions.

[‡‡‡] See Equation 1.6, where $q = 2\pi n/d$

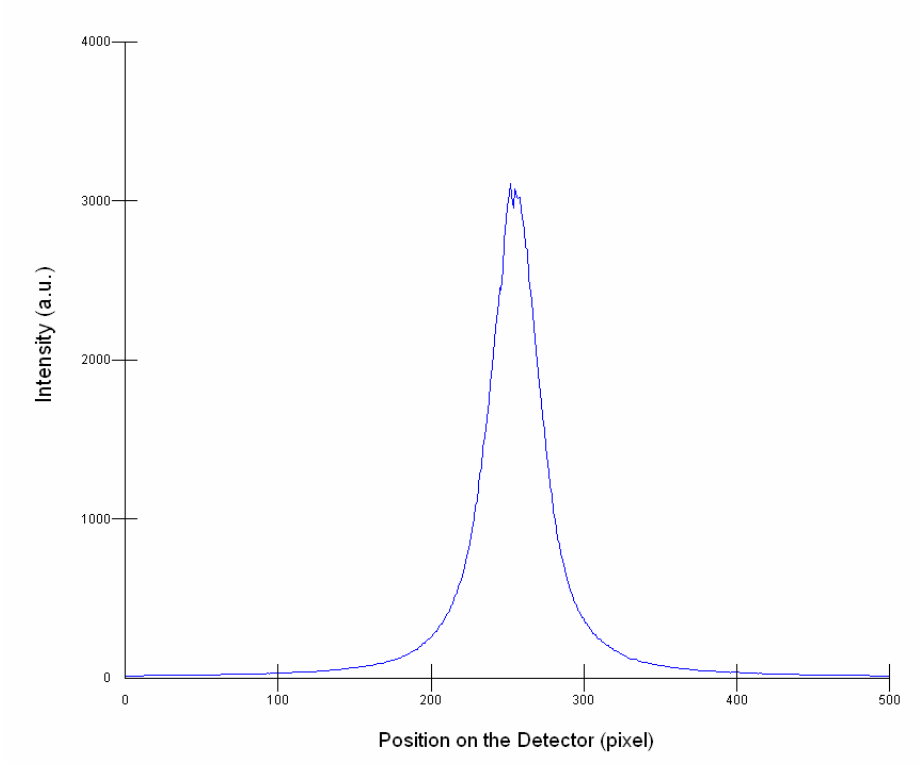


(a)

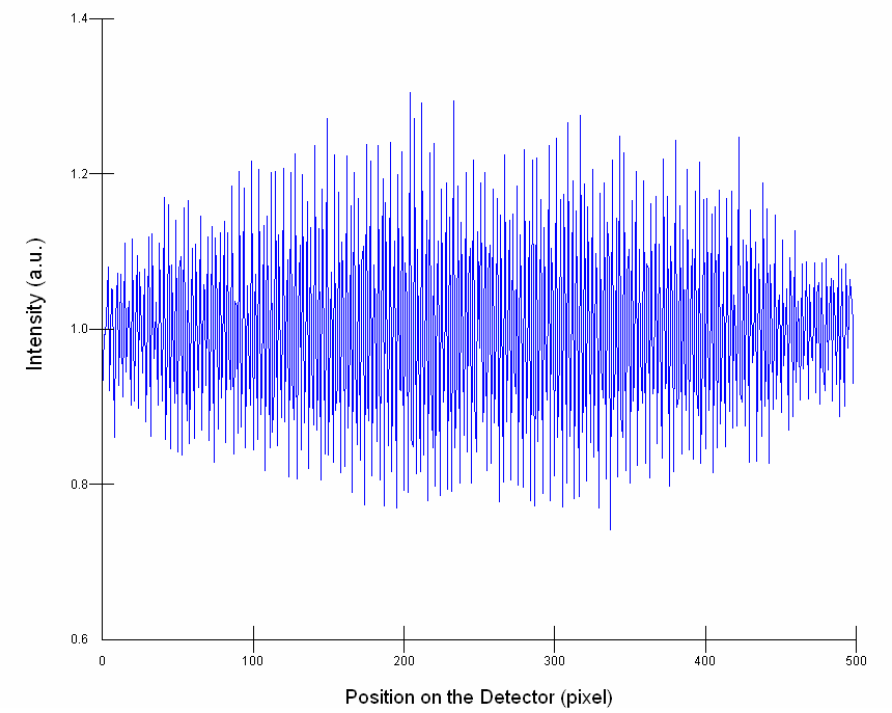


(b)

Figure 3.9 a&b: The steps taken to remove the wire structure from the SRS Daresbury SAXS images. (a) Raw SAXS image with no corrections or normalisations, showing a close-up of the wire-structure effects. (b) Projection profile (one row) of a raw image where both the low and high frequency components present.

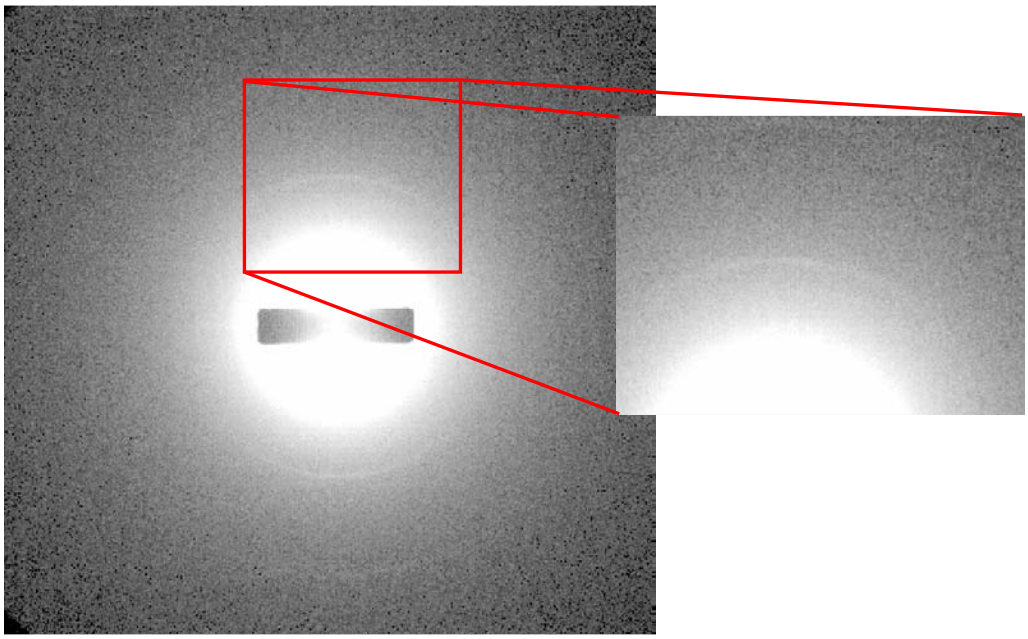


(c)



(d)

Figure 3.10 c&d: The steps taken to remove the wire structure from the SRS Daresbury SAXS images. (c) Smoothed profile obtained from low-pass filtering of the profile shown in (b). (d) The wire structure profile (one row) – high frequency profile from a raw image.



(e)

Figure 3.11 e: The steps taken to remove the wire structure from the SRS Daresbury SAXS images. (d) SAXS image after wire structure removal has been performed (from b-d) where the close-up of the wire-structure removal is shown.

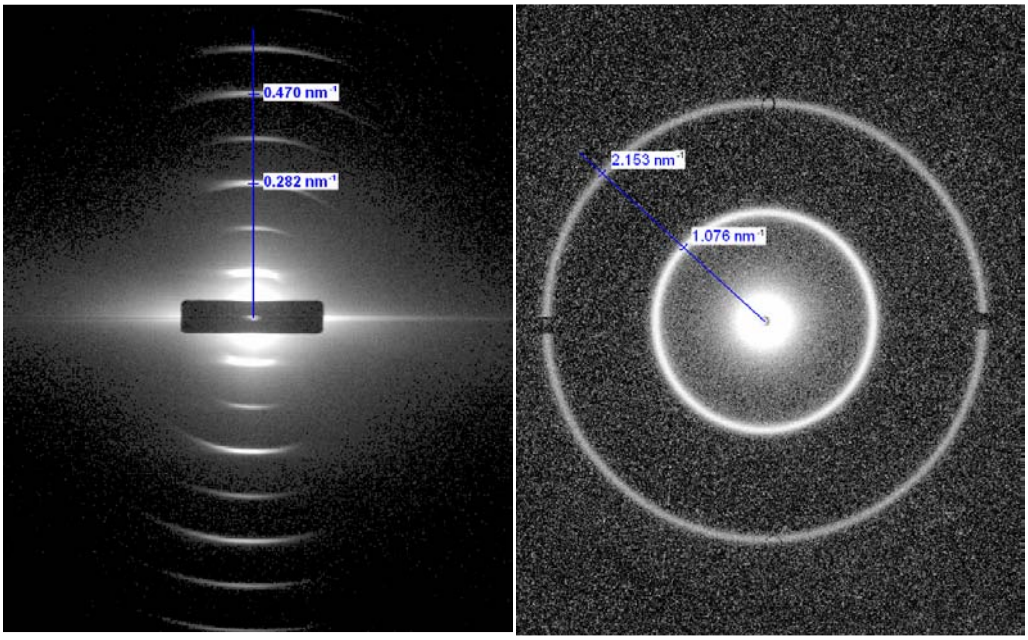
The laboratory source images were taken at a shorter sample-to-detector distance (~1.6 m) than the images taken with a synchrotron source (~6.3 m). Silver behenate was used to calibrate the d-spacing for the images taken with the laboratory source (Figure 3.12b). Silver behenate is a periodic calibrate used for low-range d-spacing calibration with a constant d = 5.8376 nm [11, 61]. Silver behenate, in a dry powder form, was placed into a glass capillary for imaging with the laboratory source. Because silver behenate is a good scattering medium (i.e. high scattering intensity), an exposure time of 10 minutes was all that was needed to obtain an image that could be used for calibration (compared to 60 minutes which was needed for the breast tissue samples). Silver behenate images were taken every time the sample holder or any part of the apparatus was moved during the experiment. To protect the silver behenate from photo-degradation, which may alter the d-spacing in the sample, the silver behenate was stored in a covered container that did not allow any light to penetrate the sample.

The observed intensity peaks from the sample of known d-spacing (rat-tail tendon or silver behenate) were fitted with Gaussian functions in order to determine the centre of each peak. Regions of interest highlighting the observed peaks of the sample of known d-spacing (see Figure 3.13) restrict the search area to find the position each peak. The known orders of the visible peaks permitted the centre of

diffraction, and the camera length to be calculated, and these parameters were inputted into all of the diffraction image header information so that pixel co-ordinates could be converted to q (which is done automatically in PCDetpak).

3.3.3 Normalisation of the Intensity

Normalisation of the intensity in the SAXS images was needed to correct for the variation in the incident beam intensity and sample attenuation. Data was normalised to compensate for the variation in the intensity of the incident beam flux and to take into account the exposure time and the decay of the synchrotron beam current or the laboratory X-ray tube current. Normalisation of the intensity was crucial for accurately comparing the data between the three experimental sessions at the SRS Daresbury and all of the data obtained with the laboratory source, as well as for comparisons of individual samples. Two types of normalisation were performed for the intensity: (1) glassy carbon normalisation for incident beam intensity for comparing between experimental sessions and the two X-ray sources, and (2) the transmitted intensity division for sample attenuation and incident beam decay.



(a) (b)

Figure 3.12: SAXS images of calibrants used. (a) Rat tail tendon at a camera length of ~6.3 metres (synchrotron source), showing the q-spacing for the 3rd and 5th order Bragg reflection, and (b) silver behenate at a camera length of ~1.6 metres (laboratory source) showing the q-spacing for the 1st and 2nd order Bragg reflection. The strong horizontal streak in (a) is a result of diffraction from the wall of the glass capillary.

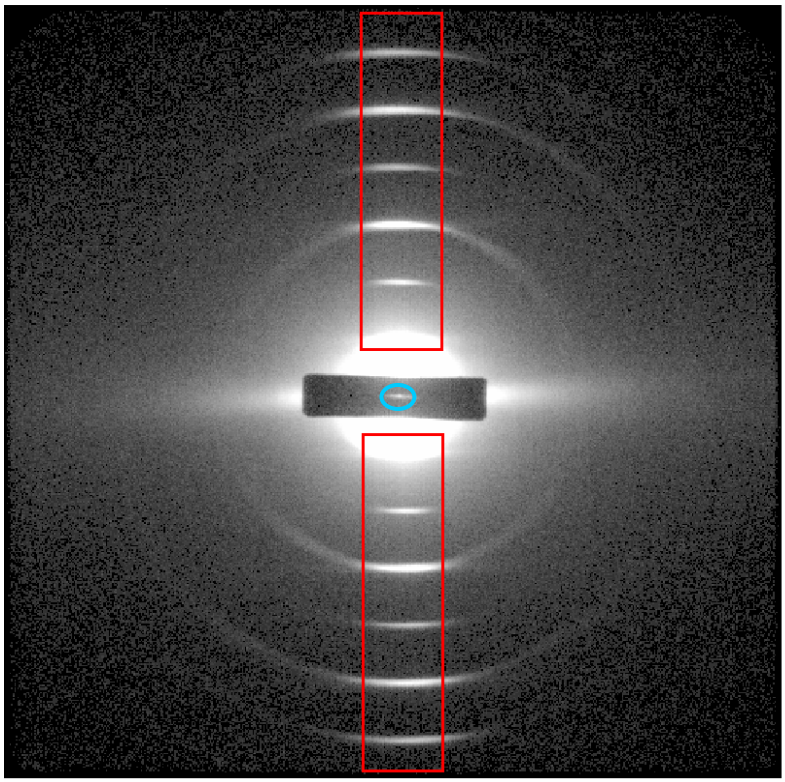


Figure 3.13: Rat tail tendon image with region of interest outlined (in red) to determine the positions of the peaks used for calibrating q. The central oval (in blue) shows the area on the semi-transparent beam stop which was used for calculating the transmitted intensity of the beam.

The intensity of the images were normalised to correct for incident intensity differences using glassy carbon (Glassy Carbon, Sigradur® (K), HTW Hochtemperatur-Werkstoffe GmbH, Germany), where the same piece of glassy carbon (K) was used for all three synchrotron experimental sessions and all of the laboratory source images. The piece of glassy carbon was solid (density = 1.54 g/cm^3) and had the dimensions 10 mm x 10 mm x 3 mm. Glassy carbon is best suited for inter-calibration because the scattering arises from voids in the carbon matrix which gives rise to an isotropic strong signal [105]. The piece of glassy carbon was scanned over its area and the scattering was measured to ensure that the piece being used was uniform. Exposure times for the glassy carbon were 5 minutes with the synchrotron source and 5 minutes with the laboratory source. However, with the synchrotron images of glassy carbon, the scattering was so strong that a sheet of aluminium attenuator was placed in front of the detector so that saturation would not occur. Since the same piece of glassy carbon was used for all of the experiments for both X-ray sources, the absorption from the aluminium attenuators were accounted for, so the scattering intensities were constant. The glassy carbon

images were similarly used to calculate a normalisation factor for each image, relative to a reference glassy carbon image, and this factor was used to correct the subsequent diffraction images.

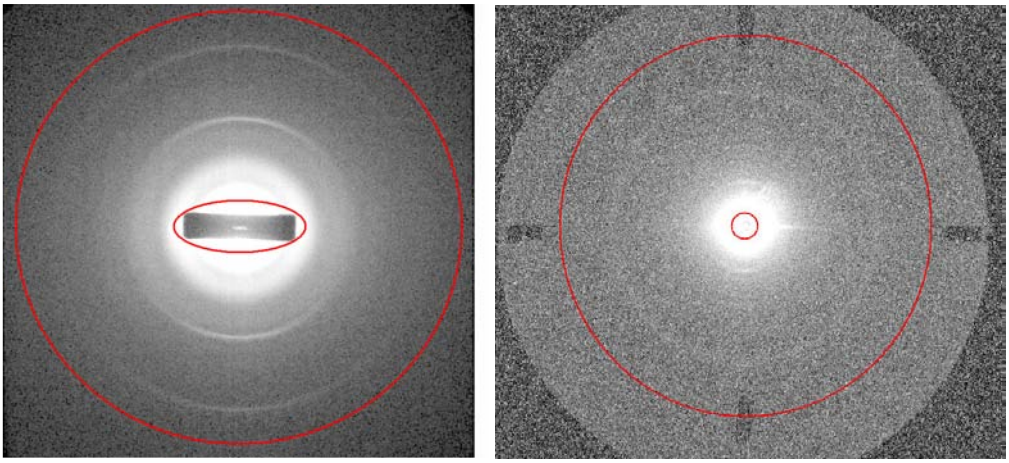
In order to normalise for sample variability due to differences in X-ray absorption of each sample as well as for the variations in the incident beam intensity within one experimental run, the intensity of the direct X-ray beam after it passed through the sample was recorded. This was accomplished by calculating the integrated intensity of the beam in the image observed through a semi-transparent beam stop (see Section 3.2) for each image. Figure 3.11 shows the area within the semi-transparent beam stop that was used to calculate the normalisation factor for each image (shown in blue). Each image was then multiplied by this normalisation factor to obtain an image corrected by sample absorption.

3.3.4 Radial Integration

As seen on the SAXS images in Figure 3.12, intensity maxima due to scattering for well-ordered systems are observed as 'rings' about the centre of the image. By radially integrating the image, a 1-dimensional (1D) profile of the intensity over the entire image can be obtained. Fitting procedures (such as Gaussian fitting of the intensity peaks) are easier to perform on 1D data, rather than with a 2D image.

Before the 2D image data was radially integrated, the beam stop was masked out. The data in the masked area was not used during radial integration. The sizes of these masks, for both the synchrotron source and the laboratory source, are shown in Figure 3.14.

A circular region of interest, centred on the diffraction image, was defined and applied to the images to radially average the image (see Figure 3.14). The intensities were averaged as a function of q (i.e. as a function of distance from the diffraction centre), starting from the edge of the masked beam-stop and outward to the defined edge of the circle region of interest. The resulting 1D data was exported to an Excel file where the data was recorded as q as a function of $Intensity*q^4$. Multiplying the intensity by q^4 allowed for easier fitting of the quadratic background function to the linear data.



(a) (b)

Figure 3.14: Region of interest (shown as outer red circles) for the (a) SRS Daresbury images and (b) for the laboratory SAXS camera images. The area integrated over each image was maintained as constant (in q) for all samples, irrespective of sample-to-detector distance. The beam-stop was also marked where it was masked so as to be excluded from integration (inner red ellipse for (a) and inner red circle for (b)).

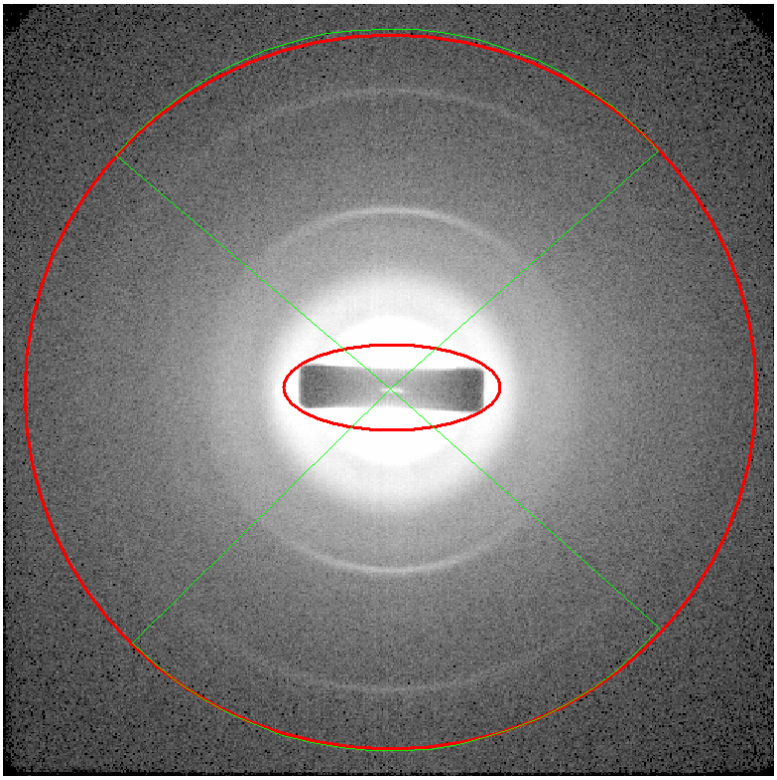


Figure 3.15: Sector regions of interest used to specifically look at meridional features from the SAXS image (marked in green). The beam stop was also masked from integration.

Sector integration was also performed in order to obtain a more detailed analysis of the SAXS features in the meridional direction, with less information contributing to the profiles from the equatorial features, and vice versa. An example of the sector areas is shown in Figure 3.15. The four sector pieces equated to the same area of the original circle region of interest.

The intensity profiles for *q* versus *Intensity*q*^{4} were graphed (as shown in Figure 3.16) and the peaks were fitted with a sum of Gaussian functions [42, 73] and a quadratic function using a script within Igor Pro (Igor Pro 5.0, WaveMetrics Inc.) that automated the peak and background fitting for batch processing of the images.

A quadratic function[§§§] was fitted to the 1-D profile SAXS data for each patient after the peak features were subtracted from the profiles:

$$q^4 I(q) = B(q-A)^2 + C \qquad (3.1)$$

where $q = 4\pi\sin\theta/\lambda$. These coefficients (A, B, and C), as well as the total area under the fitted quadratic function (amorphous scatter) were calculated for each tissue sample. The peak amplitudes, position of the peaks in q-space, areas under the peaks, and the other features listed in Table 1.1 were also extracted from the 1-dimensional plots.

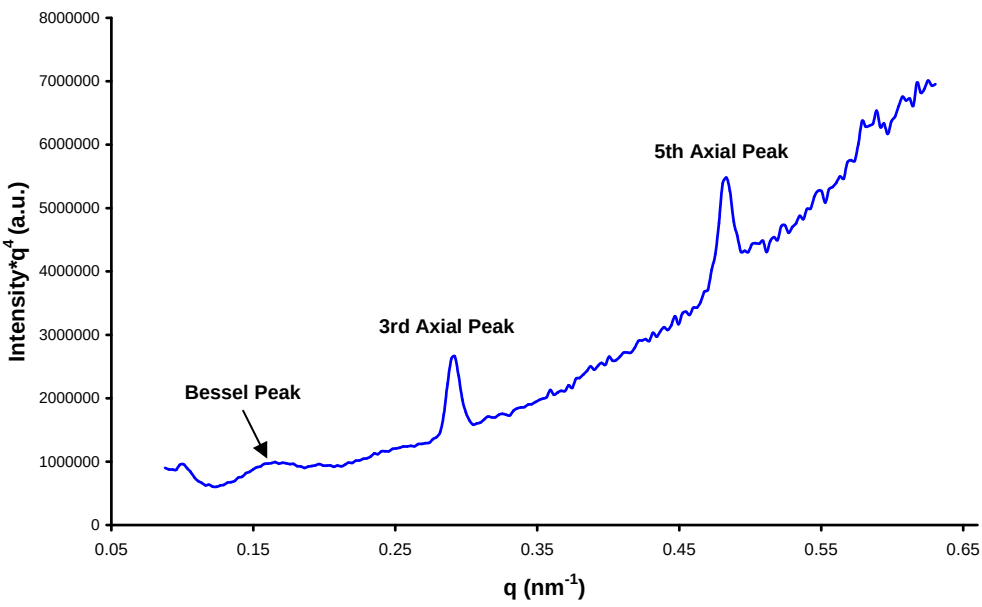


Figure 3.16: Intensity*q^4 versus q for raw data of a benign lesion. The 3rd and 5th order axial peaks are prominent as well as the second order Bessel peak.

[§§§] Derived from the quadratic function: $I(q^4) = ax^2 + bx + c$, where $A = -b/a^2$, $B = a$, and $C = c - (b^2/4a)$.

3.3.4a Orientation Distribution data (Polar Angle Data)

The collagen axial peaks seen on SAXS images appear as extended 'arcs', which is indicative of a low degree of preferential orientation of the collagen fibres (i.e. a large polar angle range). The 'arcs' were typical of all of the tissue samples, irrespective of tissue type. However, there is some evidence of fibres aligning in the longitudinal direction (see Section 4.2.4). Despite their breadth, the arcs arising from the lateral direction are centred about a polar angle of 90º. It is likely that this effect is partially as a result of inserting the tissue samples into capillaries [90].

Polar angle data from the third-order axial d-spacing peak maxima was obtained for the 80 patients imaged during Experiment #3 at the SRS Daresbury. Each patient's 2D image (Figure 3.17a) was converted into 'RTheta' images (Figure 3.17b) where the co-ordinate system is remapped as $I(\theta)$ versus q, where θ is the polar angle defined in Figure 3.15. The RTheta images were then folded along the 0º axis and the data between 180º to 360º was averaged with the data between 0º and 180º. Because the third order axial peak was the most intense peak, the area between q = 0.265-0.310 nm^{-1} was examined where 1D profiles of $I(\theta)$ were calculated for this area by integrating the RTheta image over all q (see Figure 3.18). From these 1D profiles, Gaussian curves were fitted and information regarding the FWHM, amplitude, centre, and area were extracted.

3.4 Methods of Analysis

Once all of the features were extracted from the SAXS images, further analysis was done similarly with both the synchrotron data and the laboratory data. The aim of the analysis was to test for classification utility and to determine what features related to tissue structure can distinguish between tissue types (as discussed in Chapter 1). All of the statistical tests used for analysis of the data are described in detail in Appendix E. The data was tested for homogeneity of variance (Bartlett's Test) and agreement with Gaussian distribution (Shapiro-Wilk Test for Normality) in order to validate the use of ANalysis Of VAriance (ANOVA) and t-testing. A Kolmogorov-Smirnov (KS) test was also used to determine if a sample can be adequately described by a particular statistical distribution and was mainly used to assess the assumption of normality used in the ANOVA models. All of the statistical tests were performed

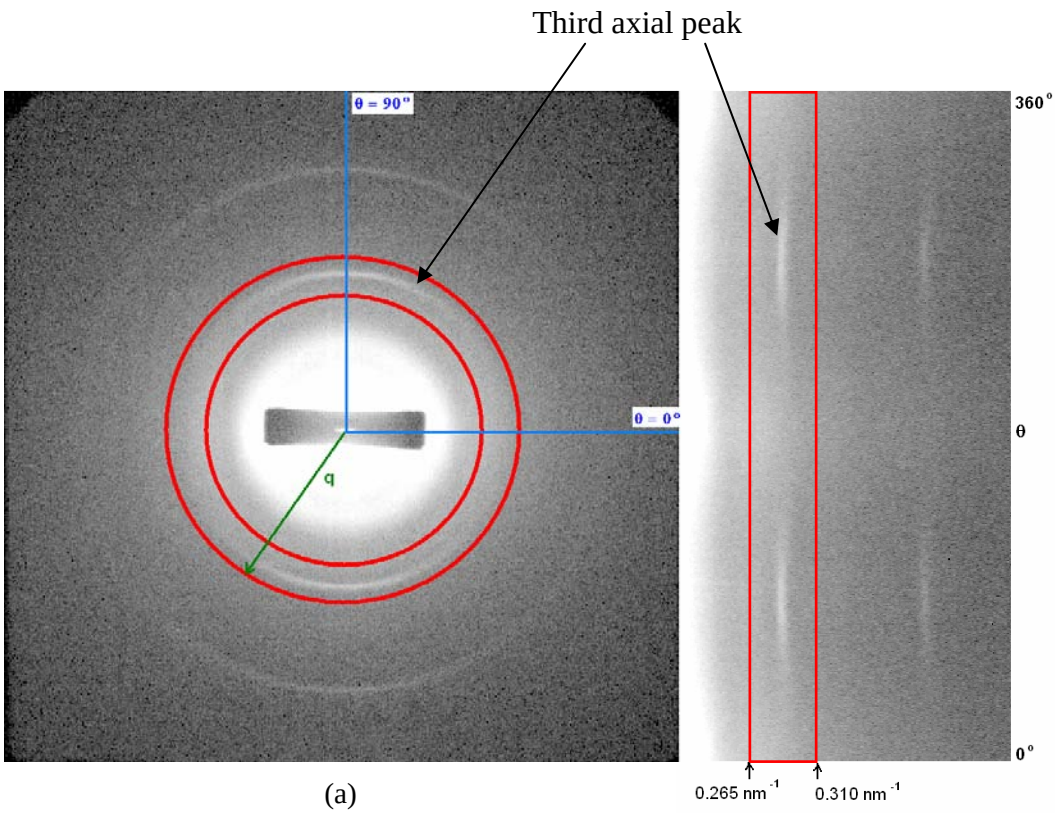


Figure 3.17: (a) 2D SAXS image showing the region of interest around the third order axial collagen peak. The image was converted to (b) RTheta, where each pixel (x,y) is remapped to (θ,q) for the radius from the diffraction centre at polar angle θ for the equatorial axis.

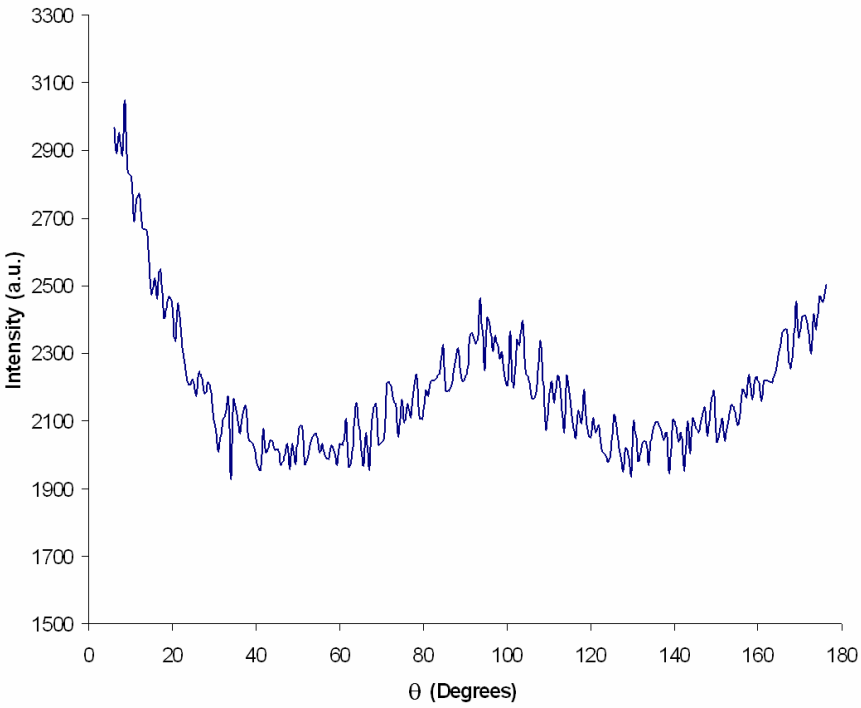


Figure 3.18: 1D plot of the integrated section between q = 0.265-0.310 nm^{-1} shown in Figure 3.17b, where the data from 180º-360º has been averaged with the data from 0º-180º. The peak is representative of the third order axial collagen peak found within this q-range, and was fitted with a Gaussian function in order to extract the peak features for analysis (amplitude, height, centre, and FWHM).

using R (Version 2.6.1, The R Foundation for Statistical Computing, 2007), MatLab (Version 2.6.1 R2007a with Statistical Toolbox, MathWorks Inc. 2007), and Microsoft Office Excel (Microsoft Office Professional Edition 2003, Microsoft Corporation 2003).

Principal Component Analysis (PCA) was used to reduce the multidimensionality of the data set to lower dimensions in order to perform a more discrete analysis [21, 26, 37, 66]. PCA involves the calculation of the eigenvalue decomposition of the data matrix after it has been mean centred for each variable. PCA can reveal the internal structure of the data set in such a way that best explains the variance in the data. Once these significant variables were determined, their diagnostic performance was evaluated for separation between tissue groups using Receiver Operating Characteristic (ROC) curves. ROC curves plot the sensitivity and (1-specificity) for a binary classifier (between two tissue groups) and provides a tool to check classification models. All of the ROC curves were calculated using a web-based calculator (John Eng, John Hopkins University, Baltimore, Maryland USA 2007 [38]). The particulars of how the ROC curves were calculated are further discussed in Appendix E.

Further modelling of the extracted SAXS features was studied using quadratic discriminate analysis (QDA) (discussed in Appendix E). QDA is used in statistical classification to separate measurements of two or more classes of objects by a quadratic surface, which was performed once the best variables for classification were determined using PCA. The probabilities between the separations of the tissue groups using these variables can be assessed for the accuracy of classification. Chapters 4 (synchrotron source) and 6 (laboratory source) analyse the primary site tissue samples using QDA for classification of the tissue groups.

The data associated with the series data (the data taken from samples at increasing distances away from the primary lesion) were more challenging to analyse because of the distance parameter associated with each tissue sample. In order to properly represent the data over the area of the excised tissue mass, Gaussian mixture modelling and model-based clustering techniques were used to create 2-dimensional maps of the features. Feature values were mapped over the 2D area of the tissue mass where probability distributions (probability density functions) were calculated to visualize the variables. This allowed visualisation of the feature changes with distance, or feature gradients.

In the next chapter (Chapter 4), 130 tissue samples harvested from the centre of the tumour (i.e. primary site) and imaged using a synchrotron source are analysed. Detailed statistical analysis of the features listed in Table 1.1 are presented and

discussed. QDA modelling and ROC curves of the data are discussed in relation to accuracy of distinguishing between the tissue groups and the potential for classification.

Chapter 4

Analysis of Primary Site Tissue Samples using a Synchrotron Source

4.1 Introduction

Results from the primary site tissue samples taken at the centre of the lesion for 130 patients and imaged using a synchrotron source (Station 2.1 at the SRS Daresbury, UK) are presented in this chapter. SAXS parameters relating to the intensity, axial features, and equatorial features, were extracted and used to determine the most significant features that could be used for classification of tissue samples. There were 19 patient samples imaged in November 2005 (Experiment #1), 31 patient samples imaged in February 2006 (Experiment #2), and 80 patient samples imaged in February 2007 (Experiment #3).

There have been a several studies that have examined the use of SAXS for tissue identification [16, 38, 41-43, 75, 106] and whilst many have suggested features that show some ability to distinguish between disease types, none have looked at large patient cohorts that can be used for classification, modelling, and for assessing possible suitability for a new diagnostic technique. Chapter 1 discussed the structural features of collagen that are observable in SAXS images (see Table 1.1). It was not the purpose of this study to perform a structural analysis of collagen in the breast. Instead, we examine a broader range of features, many of which are directly related to collagen structure, but some of which derive from other tissue scattering

characteristics, in order to determine which features may be useful for tissue classification.

Section 4.2 discusses the determination of useful classification parameters from the 130 patients analysed in the three experimental sessions, with a particular emphasis on the axial D-spacing experiments by Fernandez *et al*. [49, 51] and Lewis *et al*. [77]. The parameters identified in Section 4.2 were further used in a model that can potentially classify the four tissue groups: invasive carcinoma, benign, normal, and mammoplasty. Linear analysis and quadratic discriminate analysis were used to create models that have the best potential for classification. Receiver Operator Characteristic (ROC) curves were calculated to verify the reliability of each of the models and a discussion of the sensitivity and specificity of using SAXS for classification analysis is presented.

A subset of the data analysed in this chapter has been published in Medical Physics (Sidhu *et al*. "X-ray scattering for classifying tissue types associated with breast disease," Med. Phys. 35, 4660-4670 (2008)) and is referred to as Paper I [114], which is included at the end of this thesis. Only the data from Experiment #3 (80 patients) was presented in Paper I in order to remove any inter-experimental effects. Experiment #3 had the largest number of tissue samples examined with the most diverse range of tissue types. Experiment #3 acquired SAXS images by a point-by-point basis then integrated all of the point images to obtain one summed SAXS image for the tissue sample. This allowed for a comparison analysis between individual point-scan SAXS images and the average (or integrated) SAXS image of the entire tissue sample. This addresses the experimental differences between Fernandez *et al*.[41, 43] and Lewis *et al.* [77], which may have contributed to the differences in results by these two groups (Section 4.2.1a). The data acquired during Experiments #1 and #2 was recorded as a single image whilst the entire tissue core was scanned. Therefore, in order to combine the data from all three experiments, the averaged SAXS images from Experiment #3 were used in conjunction with the SAXS images acquired from Experiments #1 and #2.

In this chapter, the same analysis that was performed in Paper I is extended to the data from all three experimental sessions combined. Calibration and normalisations were performed (as discussed in Section 3.5) to ensure that the data from all three experiments could be combined for analysis. A comparison between the data of Experiment #3 (80 patients) with all of the data from the three experiments combined (130 patients) is discussed in Section 4.4.1.

4.2 Determining Useful SAXS Parameters

4.2.1 Axial (Longitudinal) Variables

The longitudinal arrangement of collagen fibrils (the axial d-spacing peaks) give rise to the axial variables obtained from the SAXS images (as discussed in Section 1.3.1 and Table 1.1). The following characteristics were extracted from the fitted peaks for both the third and fifth order axial d-spacing peaks for the 130 patients: peak amplitude, peak full-width-half-maximum (FWHM), peak centre (axial D-spacing), and peak area.

The use of Principal Component Analysis (PCA) (discussed in Section 3.4 and described further in Appendix E) revealed that the amplitudes of the axial peaks were the most significant parameter of all the axial variables considered. The first principal component (PC1) accounted for 99.5% of the data and was given by the following linear combination of features:

$$PC1 = -0.583*(3^{rd}\ \text{Amplitude}) - 0.812*(5^{th}\ \text{Amplitude}) \qquad (4.1)$$

The amplitude of the 3rd axial peak was used in the discriminate classification modelling in Section 4.3 because this peak feature was visible for all of the patients and was shown to represent most of the data (99.5%) according to PCA. However, when the average values for the 3rd axial peak amplitude (with 95%CI) were calculated for each tissue group (shown in Figure 4.1), no separation of tissue groups was observed ($p > 0.05$). Therefore, the absolute values of the 3rd axial peak amplitude showed no significant differences between tissue groups, but the variances between each of the tissue groups showed distinctions between the tissue types.

Previous studies have shown that the axial D-spacing of collagen can be used to distinguish between breast disease types [20, 49, 51, 77, 102]. Figure 4.2 shows the mean third order axial peak d-spacing for each tissue group with calculated 95% confidence intervals. The normal tissue group had the lowest axial D-spacing of 64.88 ± 0.04 nm. Both mammoplasty and benign tissue groups had similar axial D-spacings of d = 64.92 ± 0.04 nm and d = 64.91 ± 0.02 nm, respectively. Invasive carcinoma tissue had the highest axial D-spacing of d = 65.00 ± 0.05 nm. The p-values for each of the tissue group comparisons are shown in

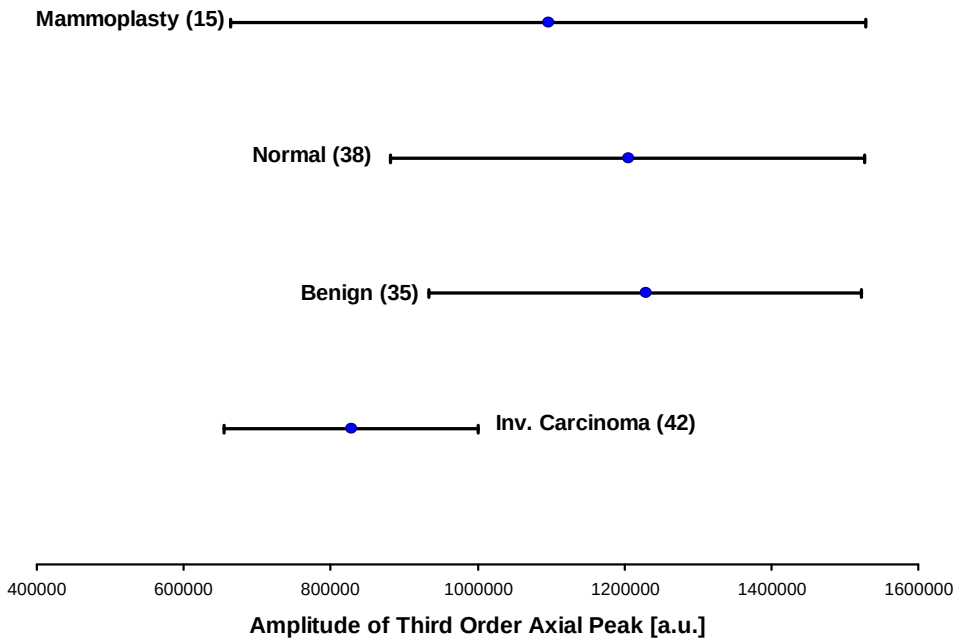


Figure 4.1: Average of the amplitudes of the third axial peak for the four tissue groups. Error bars represent 95% confidence intervals. Numbers in brackets represent the number of tissue samples in each tissue group.

Table 4.1: Comparison of the axial D-spacings between each tissue group, using p-values calculated from a two-sample t-test.

Tissue Group Comparison	p-value
Inv. Carcinoma & Benign	0.061
Inv. Carcinoma & Normal	0.0050**
Inv. Carcinoma & Mammoplasty	0.0093**
Benign & Normal	0.26
Benign & Mammoplasty	0.11
Normal & Mammoplasty	0.62

** Significantly different $p < 0.01$

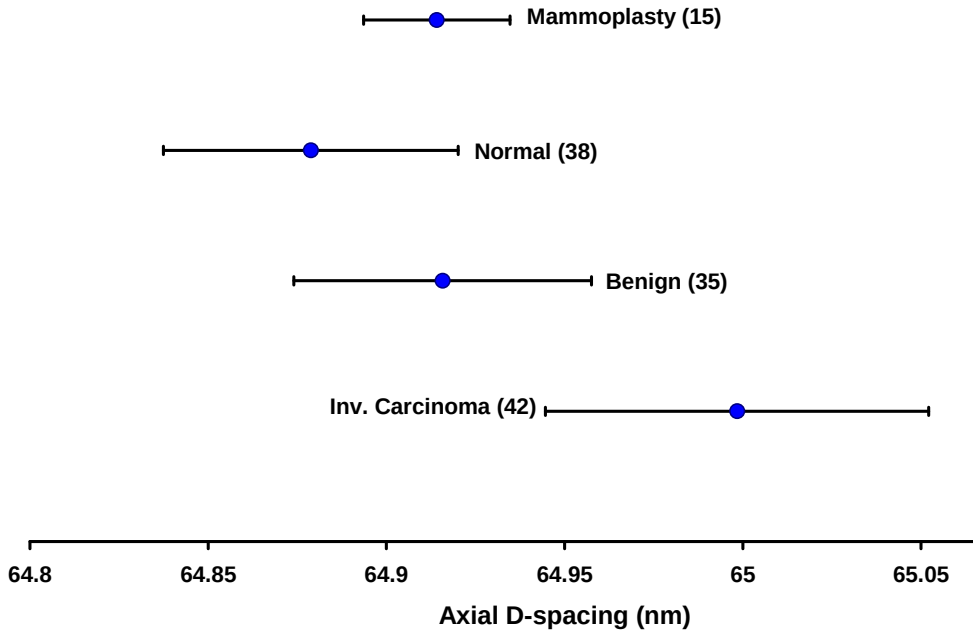


Figure 4.2: Average of the centres of the third axial peak for the four tissue groups (axial d-spacing). Error bars represent 95% confidence intervals. Numbers in brackets represent the number of tissue samples in each tissue group.

Table 4.1. There were statistically significant differences ($p < 0.05$) in the axial D-spacing between invasive carcinoma and normal tissues and between invasive carcinoma and mammoplasty tissues. However, there was no significant difference in the axial D-spacing between any of the other tissue comparisons, including between normal tissue and mammoplasty tissue.

4.2.1a Point Scans: Comparison of Two Experimental Techniques to Determine Collagen Axial D-spacing

SAXS is able to probe a relatively large volume of a tissue sample, since scattering from a volume equal to the incident beam area multiplied by the sample thickness is imaged. In the study by Lewis *et al*. [77], the volume of tissue imaged was on the order of 0.25 mm^3 (0.5 x 0.5 x 1 mm), whereas the study by Fernandez *et al*. [49, 51] imaged a volume on the order of 0.04 mm^3 (0.2 x 0.2 x 1 mm)[****]. In both studies, the volume imaged with SAXS was larger than the small area probed in histology for diagnosis, where a thin section is examined that is typically ~0.002 mm^3 (0.002

[****] Lewis *et al*. used a beam size of 0.5 x 0.5 mm and Fernandez *et al*. used a beam size of 0.2 x 0.2 mm. Both studies used tissue a tissue thickness of 1 mm.

mm thick and ~1 mm^2 area). Both Lewis *et al*. [77] and Fernandez *et al*. [49, 51] found that the axial D-spacing for breast fibrillar collagen was larger in invaded tissues compared to benign tissues; however, the specific values of these d-spacings differed by up to 0.2 nm.

In order to determine if the differing experimental methodology was the underlying reason for the different axial D-spacings from each study, the methods of Lewis *et al*. [77] and Fernandez *et al*. [49, 51] were compared. Lewis *et al*. [77] acquired SAXS image data over the entire sample volume by oscillating the sample over it's length during image acquisition to produce one image from the entire tissue volume. Fernandez *et al.* [49, 51] acquired multiple individual images acquired by incrementing point scans through the tissue sample. Each point scan image was analysed as a separate tissue sample. In order to compare the experimental methodologies used by each group, 15 patients were directly compared by two methods in this study:

1) Point Images (Image # in Figure 4.3): individual SAXS images were taken at 1 mm increments in the tissue sample for each patient (as with Fernandez *et al*. [49, 51]). Each image was analysed as an individual SAXS image. The number of images taken over one tissue sample ranged from 4-19 per patient (mean = 9) with 141 images in total, over the 15 samples.
2) Average Image (Image T in Figure 4.3): Obtained by averaging all of the point images together from each patient to get one SAXS image representative of that tissue sample as a whole (similar to Lewis *et al*. [77]), resulting in 15 images for 15 tissue samples.

The 15 tissue samples imaged consisted of 6 invasive carcinoma, 4 benign, and 5 mammoplasty. The purpose of this analysis was to examine the inter and intra variability between patient tissue samples, rather than tissue type differences. Therefore, normal tissue (tissue that was histologically normal but sampled from patients with known disease) was not studied. Table 4.2 lists the axial d-spacing values for each of the point scans imaged for the 15 tissue samples, and the axial d-spacing from the integrated image for that tissue sample.

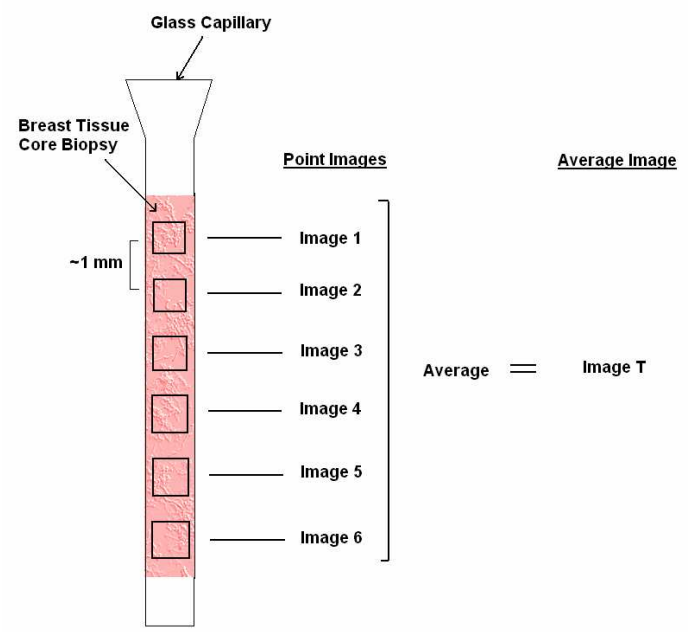


Figure 4.3: Diffraction images were taken at 1 mm intervals through the length of the tissue core. Each of these images were analysed separately (Point Images) and then each of these images were averaged to obtain one average diffraction image representing the entire tissue core (Average Image).

Table 4.2: The third axial peak d-spacing for each of the point images with the average D-spacing ± (standard deviation) for the point images. The D-spacing calculated from the integrated image for each of the 15 patients is in the last column.

			D-spacing (nm)	
Patient	**Tissue Type (as diagnosed by histology)**	**# Images**	**Average of the Point Images (±SD)**	**Integrated Image**
115	Invasive Carcinoma	8	64.98 ± 0.21	64.85
127	Invasive Carcinoma	4	64.82 ± 0.14	64.97
133	Invasive Carcinoma	5	64.99 ± 0.37	64.88
141	Invasive Carcinoma	15	64.82 ± 0.09	64.97
104	Invasive Carcinoma	19	65.03 ± 0.19	64.95
109	Invasive Carcinoma	10	65.11 ± 0.56	65.00
120	Benign	9	64.87 ± 0.21	64.98
124	Benign	5	64.73 ± 0.12	64.90
112	Benign	8	65.10 ± 0.37	64.90
137	Benign	8	64.90 ± 0.20	64.63
144	Benign	10	65.07 ± 0.23	64.87
101	Mammoplasty	10	64.74 ± 0.38	65.00
108	Mammoplasty	10	64.75 ± 0.07	65.00
110	Mammoplasty	10	64.80 ± 0.10	65.00
111	Mammoplasty	10	64.85 ± 0.23	64.78

The distribution of d-spacing values within the point scans of each individual patient was calculated and then compared with the distribution of d-spacing values for all 15 patients, irrespective of tissue type (Table 4.3), using the Bartlett's Test for significance (see Appendix E for details of statistical tests). This suggested that there was a distinct difference in the range (i.e. the distribution) of d-spacing values between different patients, $p = 5.7 \times 10^{-10}$. Closer examination, now including tissue type as a discriminating variable, revealed that a significant difference in the variation of the d-spacing values was present in both the mammoplasty ($p = 1.1 \times 10^{-5}$) and the invasive carcinoma ($p = 2.3 \times 10^{-7}$) tissue groups but not in the benign group ($p = 0.195$). The benign group appeared to be unique in that it had a very homogenous range of d-spacing values irrespective of the patient (i.e. within each benign tissue sample there is no significant difference). This indicates that the axial D-spacing for benign tissues does not exhibit intra-patient variability, but there is inter-patient variability with benign tissues. The large variation in the d-spacing within the invasive carcinoma and mammoplasty tissue samples can probably be attributed to the heterogenous nature of tissue.

Inter-patient variability, in addition to disease status, was also examined (Table 4.3). The variation of the d-spacing values of the point-scan data, between patients but within each tissue group (invasive carcinoma, benign, and mammoplasty) was examined using Analysis of Variance (ANOVA), which showed a p-value of 0.0045. This indicated that within at least one of the three tissue types, the magnitude of the d-spacing varied significantly between patients. Further investigation revealed that it was the benign group that had the most amount of variation in d-spacings ($p = 0.051$) between patients, compared to invasive carcinoma and mammoplasty patients.

The use of ANOVA also revealed that the d-spacing differed between tissue groups ($p = 0.002$) when using the point scan method, which was also concluded in the Fernandez *et al*. [49, 51] studies, and therefore may be used to distinguish between tissue types. When using the averaging method for each patient t-tests were able to distinguish between invasive carcinoma and mammoplasty tissue ($p = 0.02$) but not between invasive carcinoma and benign tissue ($p = 0.78$). Thus, the d-spacing calculated from the entire tissue sample (the integrated image which represents all of the tissue core) could be used to separate between certain tissue groups (i.e. invasive carcinoma and mammoplasty), but was not as sensitive to the differences revealed by a point scan analysis.

Table 4.3: Using the Bartlett's Test for significance, p-values were calculated to compare the third order axial d-spacing of the point scan images within each patient, as well as between each of the 15 patients.

	Irrespective of Tissue Type	**Within each Tissue Type**	
Intra-Patient Variability (within each patient)	5.7×10^{-10}**	Invasive Carcinoma	2.3×10^{-7}**
		Benign	0.195
		Mammoplasty	1.1×10^{-5}**
Inter-Patient Variability (between each patient)	0.0045**	Invasive Carcinoma	0.158
		Benign	0.051
		Mammoplasty	0.667

*Significantly different $p < 0.05$
** Significantly different $p < 0.01$

In summary, each patient had their own characteristic collagen d-spacing value and variation about that average value. The mean value of the d-spacing was found to differ between patients of the benign tissue group. The range of d-spacing values was different for patients in both the mammoplasty and invasive carcinoma tissue groups, but not in the benign tissue group. Therefore, there are distinct differences in how the d-spacing value varies between patients in the invasive carcinoma and mammoplasty tissue groups. These characteristics are expected to be influenced by biological factors, hormone therapy, chemotherapy and/or radiotherapy, genetics, and age. Analysing the d-spacing at each point in a tissue sample is more sensitive for distinguishing between tissue types than using the d-spacing over the whole tissue sample. This is to be expected, since tissue samples are typically very heterogeneous (containing a mixture of normal and diseased tissues) over several millimetres.

4.2.2 Equatorial (Lateral) Variables

The second harmonic equatorial peak was visible for 90% (117/130) of the patients (34 invasive carcinoma, 34 benign, 35 normal, and 14 mammoplasty patients). The amplitude, centre, FWHM, and area of the fitted peak were determined and analysed using PCA. The results showed that 100% of the variance in the data projected onto PC1 and it is dependent on the area of the equatorial peak:

$$PC1 = 1.0*(\text{Equatorial Peak Area}) \qquad (4.2)$$

In order to determine if PC1 is correlated with tissue diagnosis, a Kolmogorov-Smirnov (KS) test (see Appendix E for statistical details) was performed on the data. The p-values showed no evidence that PC1 was significantly different between tissue groups ($p > 0.05$). A Flinger (median) Test (see Appendix E for statistical details) on the data resulted in a p-value = 0.131, showing that the variance is homogeneous between tissue groups. Therefore, the variance in the equatorial variables is due completely to the equatorial peak area, but the peak area cannot discern between tissue groups (i.e. is not a good classifier) because the variance is similar for all of the tissue types.

Linear, quadratic, exponential, and logarithmic comparisons were made with the amplitude, centre, FWHM, and area of the fitted equatorial peak to determine if any other relationships existed in the data. A linear relationship was observed between the square root of the equatorial peak area and the amplitude of the fitted equatorial peak (Figure 4.4) for each of the tissue groups. None of the other comparisons showed any correlation. Table 4.4 lists the slope and R^2 values for a linear fit to each of the tissue groups. The slopes of the fitted lines showed a decrease from healthy tissue to progressively diseased tissues, particularly with the invasive carcinoma tissues. The area encompasses both the amplitude and width (FWHM) of the peak, where the width represents the distribution of equatorial packing of the fibrils: a larger width means that there is more variability in the distance between the fibrils compared to a peak with a smaller width. The amplitude is representative of the degree of long range order in the packing structure (distances between the tropocollagen molecules, see Section 1.3.2): the larger the amplitude, the higher degree of long range order in the packing structure of the collagen present in the sample. Therefore, in Table 4.4, a larger slope means that the tissue group has more ordered collagen present in the tissue samples with less variability in the equatorial packing distance compared to tissue groups with a lower slope. The slopes represented by mammoplasty, normal, and benign tissues were very close to one another (Table 4.4), indicating that the packing distance of the fibrils in these tissue groups were relatively equivalent (within SE). In comparison, the slope for invasive carcinoma tissues was less than the other tissue groups, which means that the equatorial packing distances between the fibrils for invasive carcinoma tissues are more variable and perhaps more disrupted as was seen in Figure 1.3.

Table 4.4: The linear fits calculated for the data shown in Figure 4.3 for each of the tissue groups. The Standard Error (SE) of the slope is shown in brackets.

Tissue Group	Slope (SE)	R^2
Mammoplasty	5814 (342)	0.96
Normal	5601 (371)	0.87
Benign	5531 (374)	0.87
Invasive Carcinoma	3237 (231)	0.86

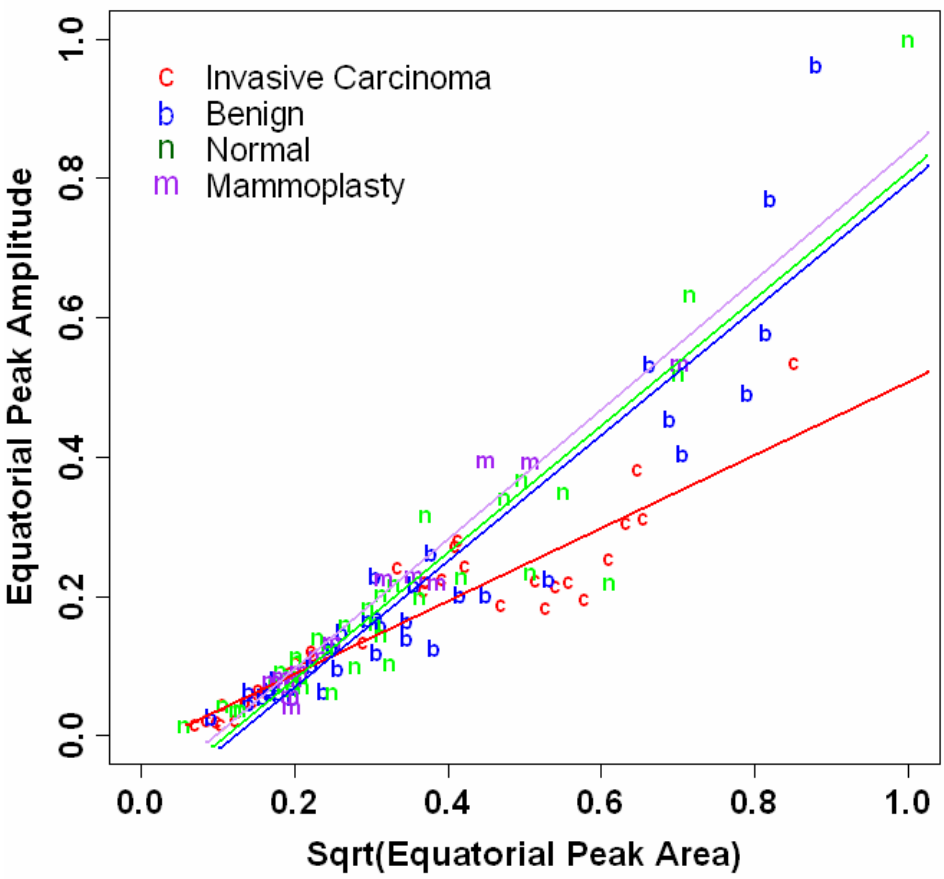


Figure 4.4: The square root of the equatorial peak area plotted with the equatorial peak amplitude for each of the four tissue groups. Axis values have been normalised with the data. The linear fits are also shown for each tissue group.

4.2.3 Total Scattering Intensity Variables

Scattering intensity was previously found to be of interest in distinguishing between invaded breast tissue and healthy breast tissue by Fernandez *et al*. [49] The group examined the power-law exponent between q = 0.50 nm^{-1} and q = 0.56 nm^{-1} (i.e. between the fifth and sixth order collagen axial peaks) and found that the specific surface area of scatterers that have dimensions of a few tens of nanometres was very much larger in invaded tissue than in healthy tissue [51]. Lewis *et al*. [77] examined the fraction of the total scattered intensity contained in the axial and equatorial peaks, and found that it was much less in tumour tissue than in either normal tissue or benign lesions. In this section, the total scattering over the entire q-range of the SAXS data (from 0.1-0.6 nm^{-1}) as well as the fraction of scattering intensity in the collagen axial peaks was examined for classification potential.

As described in Chapter 3, a quadratic function (Equation 3.1) was fitted to the 1-D profile SAXS data for each patient after the peak features were subtracted from the profiles. The Lorentz correction factor of q^2 was applied as appropriate for isotropic samples [58]. These coefficients (*A*, *B*, and *C*), as well as the total area of the fitted quadratic function (amorphous scatter between q = 0.1-0.6 nm^{-1}), the total area of the fitted peaks of the collagen features, and the peak fraction of total diffraction[††††] were analysed using PCA to examine the scattering intensity features. Additionally, the power law exponent between q = 0.3 nm^{-1} and q = 0.5 nm^{-1} (i.e. between the third and fifth order collagen axial peaks), was also examined.

PCA showed that the variable *C* encompassed all of the data and, thus, can be used to describe the entire data set:

$$PC1 = 1.0*(C) \qquad (4.3)$$

The coefficient *C* is the vertical offset to the quadratic function and essentially measures the magnitude of ($I \cdot q^2$). Therefore, *C* is a good estimator of the total scattering intensity (or amorphous scatter). A Bartlett's Test between the tissue groups with the variable *C* gave a *p*-value of 0.000198, suggesting a highly significant difference in the variances between the four tissue groups using this variable.

The differences in the amorphous scatter between tissue groups can be further seen in Figure 4.5, where the amorphous scatter (i.e. the area of the fitted quadratic, subsequent to peak subtraction) averaged over all samples for each tissue

[††††] The peak fraction of total diffraction = the total area in the fitted peaks of the collagen features divided by (the total area of the fitted quadratic function + total area in the fitted peaks of the collagen features).

group is plotted with 95% confidence intervals (CI). Table 4.5 is a list of the p-values for the amorphous scatter between each of the possible tissue group pairs, and further shows how the tissue group pairs can be separated by their amorphous scatter. All of the tissue group pairs, with the single exception of the normal vs. benign group pair (p = 0.30) were significantly different from one another. These results are discussed further in Section 4.5.

Table 4.5: P-values between tissue groups for the amorphous scatter.

	p-value
Inv. Carcinoma & Benign	0.0072**
Inv. Carcinoma & Normal	0.0023**
Inv. Carcinoma & Mammoplasty	4.1×10^{-7}**
Benign & Normal	0.30
Benign & Mammoplasty	0.0012**
Normal & Mammoplasty	0.0025**

**Significantly different $p < 0.01$

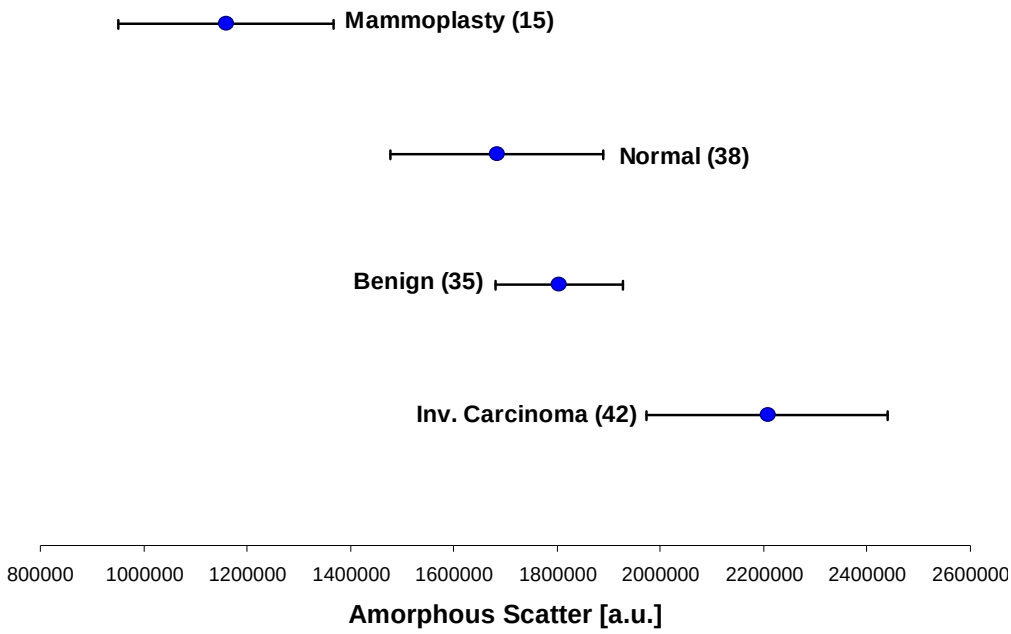


Figure 4.5: Averages of the total amorphous scatter (integrated under the quadratic function) for each of the four tissue groups over the observed q-range 0.1-0.6 nm^{-1}. Numbers in brackets indicate the number of patients in each tissue group. Error bars represent 95% confidence intervals.

4.2.4 Polar Angle Distribution Data

The SAXS images obtained using the SRS Daresbury demonstrate the extended 'arcs' corresponding to the axial d-spacing peaks indicative of a low degree of preferential orientation of the collagen fibres (see Figure 3.17, Section 3.3.3a). This relatively low degree of preferred orientation was typical of all samples, irrespective of tissue type. There is some evidence of fibres aligning in the longitudinal direction, which may be due to how the tissue sample is placed into the capillary tubes during sample preparation (as discussed in Section 2.3). The tissue biopsies are approximately the same diameter as the capillary tube (1 mm diameter), which allows the tissue core to be slid, or pushed, through the capillary tube. The tissue core is thus aligned along the vertical axis of the capillary tube. The resulting arcs are centred about a polar angle of 90º (as defined in Figure 1.10). The partial orientation of the fibres is most likely due the vertical placement of the tissue sample within the capillary tubes, which can also be attributed to the fibrous nature of the tissue sample [90].

Polar angular data output from the third-order axial d-spacing peak maxima position (taken between q = 0.265-0.310 nm^{-1}) was obtained for a sample set of 80 patients (all patients imaged in Experiment #3 as described in Section 3.3.3a and shown in Figure 3.17). From these 1D profiles, Gaussian curves were fitted and information regarding the FWHM, amplitude, centre, and area were extracted. For patients showing a constant intensity over all theta in this q-range (i.e. no particular orientation), a width of 180º was inserted into the data set. PCA was performed on these variables and it was found that the area of the fitted Gaussian represented 100% of the data. The area of a Gaussian is dependent upon both its amplitude and width. However, the amplitudes calculated for this polar angle data have essentially been analysed in Section 4.2.1: peak amplitudes associated with the axial data. It is the width (FWHM) of the fitted distribution that provides information regarding the orientation of the collagen fibrils within the tissue sample. Tissue samples that exhibit narrow peaks have collagen fibrils that have less variability in their direction of orientation within the sample. On the other hand, tissue samples that exhibit wide peaks have collagen fibrils that are distributed in many directions: meaning that the fibrils are not well oriented within the tissue sample. ANOVA showed no significant difference in the variances ($p = 0.74$) between the tissue groups. Thus, the degree of orientation does not show a correlation with tissue types, suggesting that orientation of the fibrils may not be correlated with disease presence.

4.3 Classification of Tissue Groups

The previous section (Section 4.2) examined a total of 26 parameters related to collagen axial and equatorial features, amorphous scatter, and polar distribution. The key parameters determined by PCA as potentially useful classifiers were (listed in order of significance)

(1) the total scattering intensity (i.e. predominately the amorphous scatter),

(2) the third axial peak amplitude,

(3) the third order axial D-spacing, and

(4) the second order equatorial peak area and equatorial peak amplitude

Other parameters, notably the polar distribution of the 3rd axial peak intensity, did not show any difference between tissue groupings, thus discounting these parameters as potential differentiating factors.

Of the variables (1) – (4), the axial peak amplitude and the amorphous scatter showed the most separation between tissue groups than compared with the third order axial D-spacing and equatorial peak features. Consequently, this section focuses on the evaluation of the axial peak amplitude and the amorphous scatter as potential classifiers to determine if they can be used to discriminate between tissue types. Scatter plots, Quadratic Discriminant Analysis (QDA), and Receiver Operating Characteristic (ROC) curves were used to evaluate the sensitivity and specificity of using the amorphous scatter and axial peak amplitude to separate tissue types. Details of these techniques are described in Appendix E. Scatter plots show the absolute values of the parameters in relation to each other, and were used to visually examine the data to determine if clustering or grouping was present in the data. Scatter plots that show wide separation between groups are a good indication of the effectiveness of applying an accurate model that can distinguish between the tissue groups. QDA calculates the probability contours using known data point from two groups. This model can then be used to classify unknown samples from each group with a probability of each sample being either in group 1 or group 2. Thus, the model can be considered a binary classification technique. Based on the QDA model and the probability associated with each tissue sample, ROC curves were calculated. ROC curves were used for testing the reliability of the model (in this case the QDA probability model) where sensitivity and specificity values can be used as a criterion for diagnostic performance.

A scatter plot of the 130 patients analysed over all three experiments comparing the amorphous scatter and the amplitude of the third axial peak (normalised) is shown in Figure 4.6 with tissue types indicated by symbol and

colour. The third axial peak d-spacing versus the amorphous scatter (normalised to the data) is shown in Figure 4.7. There is no clear clustering or separation of the four tissue groups in either plot (Figure 4.6 and Figure 4.7), although some separation between the tissue groups is visible with the amorphous scatter. The data in Figure 4.6, which compares the amorphous scatter and amplitude of the third axial peak, shows more spread in the data over the amplitude of the third axial peak compared to the data in Figure 4.7 with the D-spacing of the third axial peak. For this reason, the data represented in Figure 4.6, using the amplitude of the third axial peak, was used for classification modelling because it had more potential in showing separation between the tissue groups.

Inspection of both Figure 4.6 and Figure 4.7 indicate that the amplitude of the third order axial peak shows more separation in tissue groups than the d-spacing. The data points in Figure 4.7 (with the axial D-spacing) are more clustered together, whereas the points in Figure 4.6 (with the amplitude of the third axial peak) are distributed more evenly over the x-axis. The distribution of the points allow for easier modelling and classification, thus, the amorphous scatter and amplitude of the third axial peak were used for classification modelling.

A probability model was created using QDA using the amplitude of the third order axial peak and the amorphous scatter as classifiers. Figure 4.10 shows the models for the combinations of the four tissue groups analysed for 130 patients. In Figure 4.10c (for example), a probability of 0.8 means that the tissue has an 80% probability of being an invasive carcinoma, and 20% probability of being a mammoplasty tissue.

The comparison between invasive carcinoma and benign (Figure 4.10a) reflects a linear model (i.e. the probability contours are linear), whereas the benign and normal comparison (Figure 4.10d) reflects the similarities in the supramolecular components in these two tissue types in that the probability lines have no preferred orientation, so no distinction between benign and normal tissues can be attributed. The similarities between benign and normal tissues, at least in terms of their amorphous scatter characteristics, are further seen in Figure 4.10a and Figure 4.10b, which show similar probability lines when each is compared with invasive

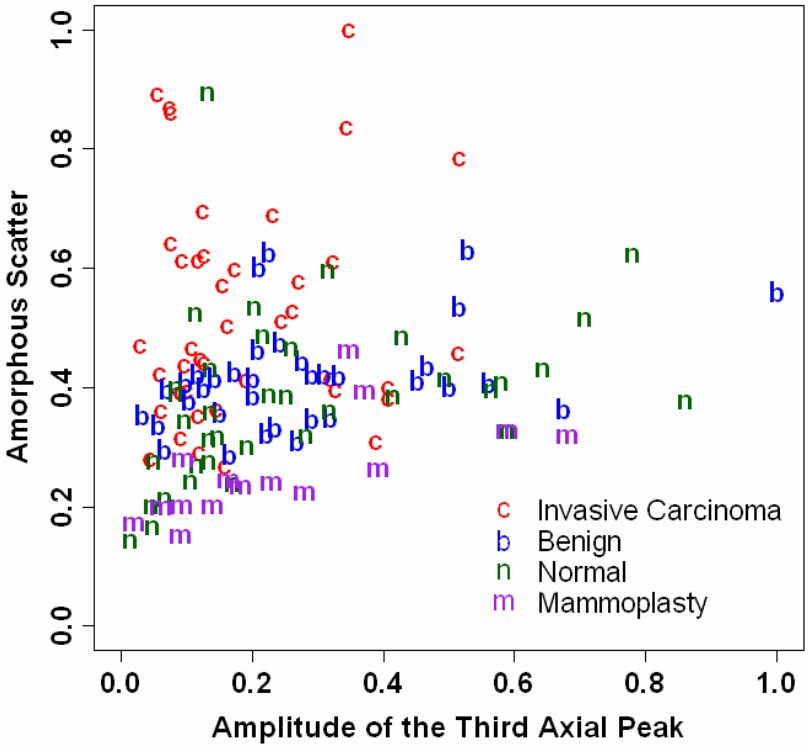


Figure 4.6: The amplitude of the third axial peak in relation to the amorphous scatter for the four tissue types. Axis values have been normalised to the range of the data.

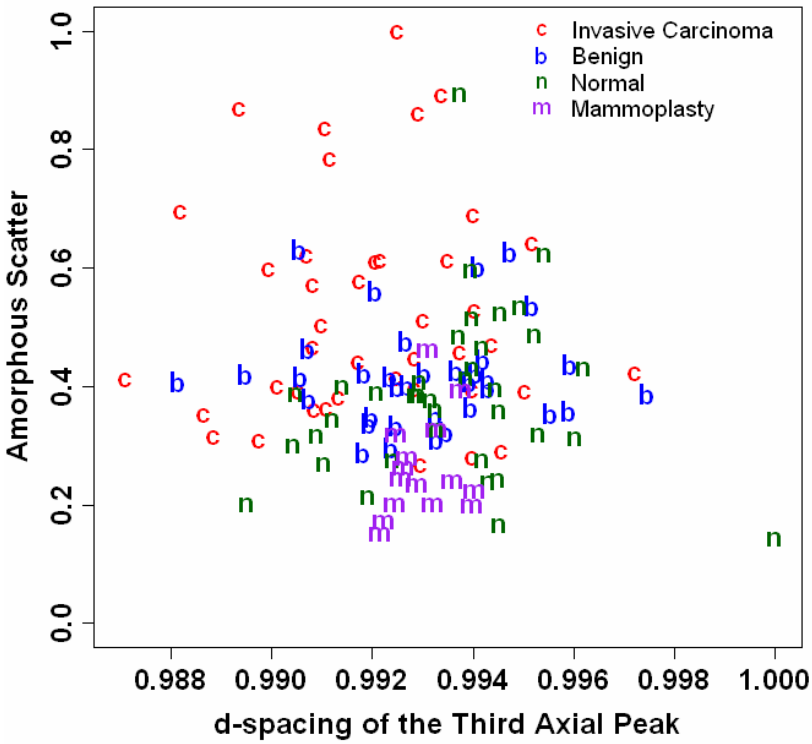


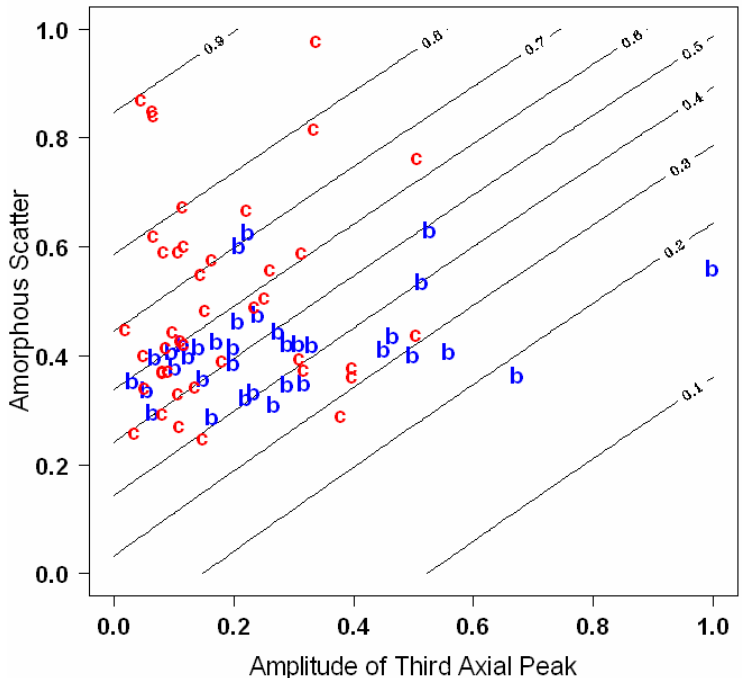
Figure 4.7: The d-spacing of the third order axial peak in relation to the amorphous scatter for the four tissue types. Axis values have been normalised to the range of the data.

Table 4.6: ROC curve characteristics for each pairing of tissue groups. The first listed tissue group was considered "positive" (e.g., invasive carcinoma) and the second tissue group considered "negative" (e.g., benign).

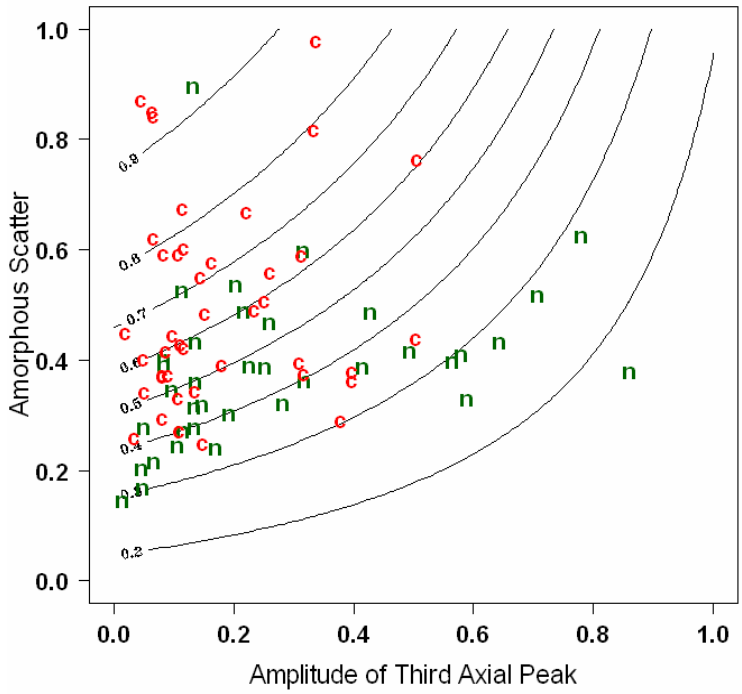
Tissue Comparisons	Sensitivity	Specificity	Accuracy	Area Under the Curve (AUC)
Inv. Carcinoma & Benign	75.0	54.3	65.3	0.74
Inv. Carcinoma & Normal	72.5	76.3	74.4	0.82
Inv. Carcinoma & Mammoplasty	97.6	66.7	89.5	0.94
Normal & Benign	72.2	40.6	57.4	0.63
Mammoplasty & Benign	73.3	97.1	90.0	0.92
Normal & Mammoplasty	94.1	17.6	68.6	0.81

carcinoma tissues. The model which shows the most separation between tissue groups is shown for invasive carcinoma compared to mammoplasty (Figure 4.10c). This can be considered the best model for tissue discrimination. This was as expected, as the amorphous scatter analysis in Section 4.2.3 and shown in Figure 4.5, also demonstrated that amorphous scatter values for invasive carcinoma and mammoplasty tissues were widely separated.

The Receiver Operator Characteristic (ROC) [38] (see Appendix E) curves for the data presented in Figure 4.10a-c are illustrated in Figure 4.11a and the ROC curves for the data presented in Figure 4.10d-f are illustrated in Figure 4.11b. ROC curves were calculated using a threshold probability of 0.6 or higher (by convention [59, 83, 133]) as being positive. A test with perfect discrimination has an ROC curve that passes through the upper left corner of a plot of sensitivity versus (1 – specificity) (100% *sensitivity*, 100% *specificity*). Therefore, the closer the ROC curve is to reaching the upper left corner, the higher the accuracy of the QDA model [133]. However, if the ROC curve is close to, or follows, a 45 degree diagonal in ROC space, then the test is considered to be completely random in determining a diagnosis. It follows then, that the area under the ROC curve represents the accuracy of the QDA model: an area of 1 represents a perfect model; an area of 0.5 represents

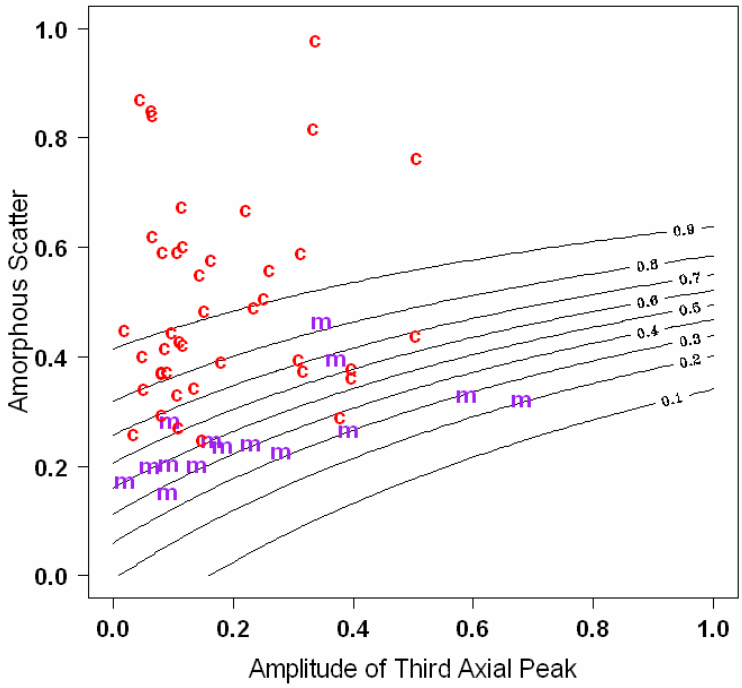


(a)

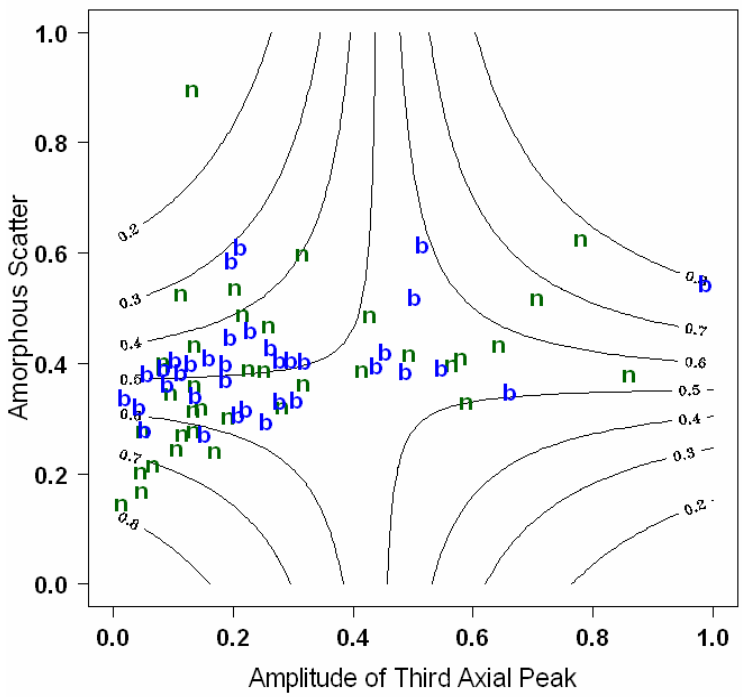


(b)

Figure 4.8 (a & b): QDA probability plots of the data shown in Figure 4.6. The symbols c, b, n, and m represent invasive carcinoma, benign, normal, and mammoplasty tissue types, respectively. (a) Probability of Invasive Carcinoma vs. Benign (b) Probability of Invasive Carcinoma vs. Normal.

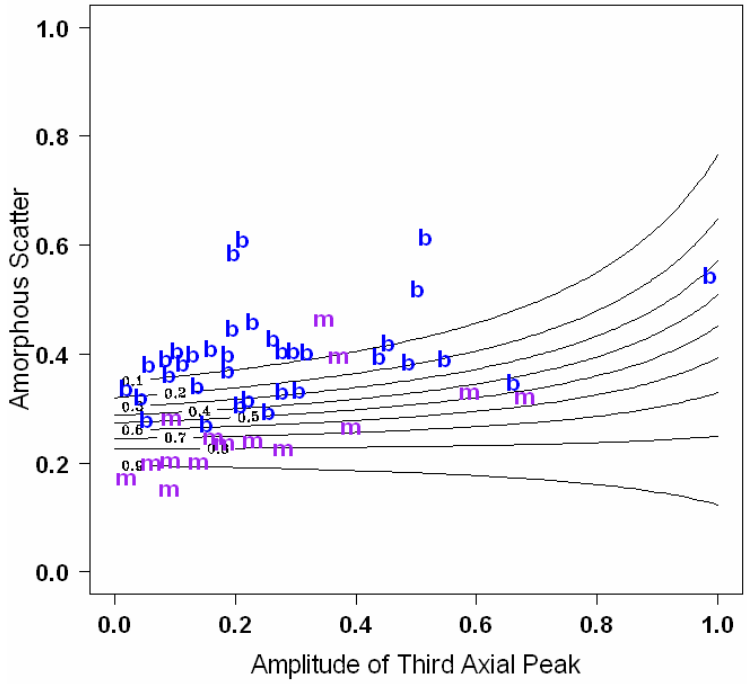


(c)

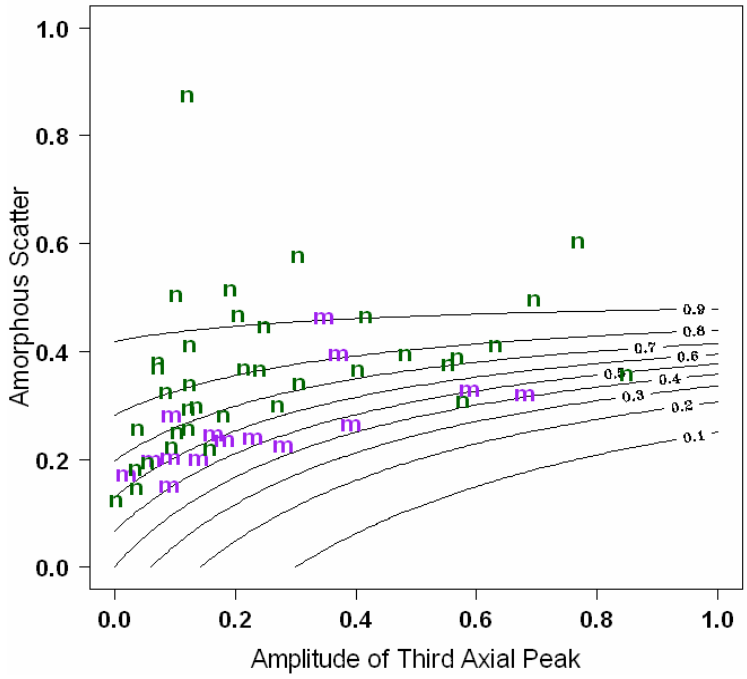


(d)

Figure 4.9 (c & d): QDA probability plots of the data shown in Figure 4.6. The symbols c, b, n, and m represent invasive carcinoma, benign, normal, and mammoplasty tissue types, respectively. (c) Probability of Invasive Carcinoma vs. Mammoplasty (d) Probability of Normal vs. Benign.



(e)



(f)

Figure 4.10 (e & f): QDA probability plots of the data shown in Figure 4.6. The symbols c, b, n, and m represent invasive carcinoma, benign, normal, and mammoplasty tissue types, respectively. (e) Probability of Mammoplasty vs. Benign (f) Probability of Normal vs. Mammoplasty.

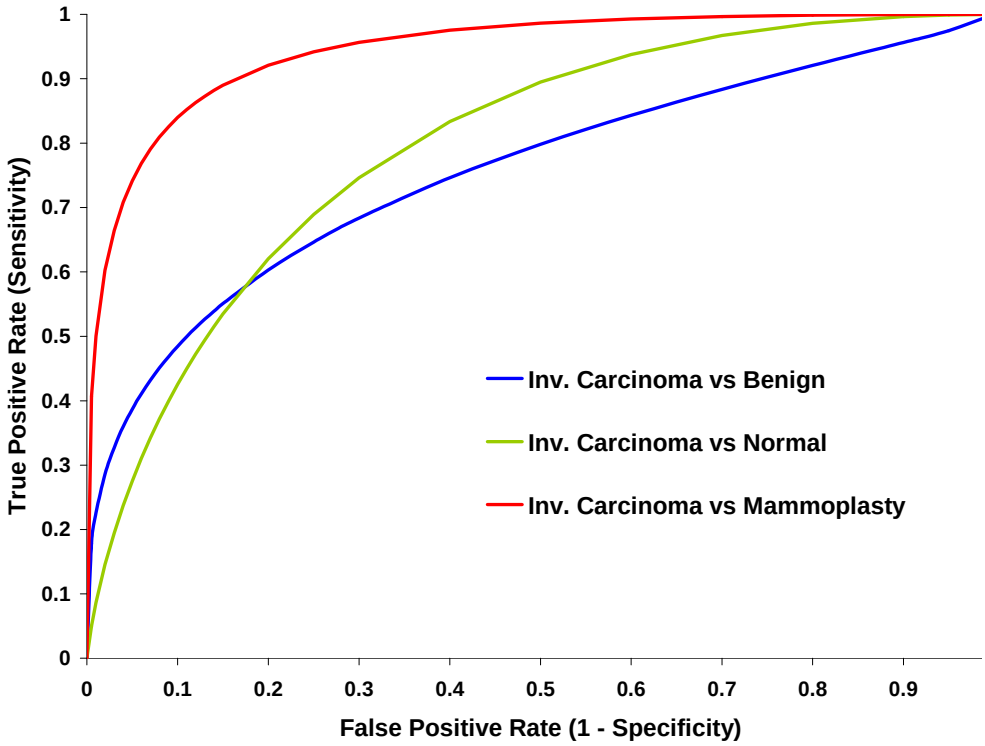
a worthless model. Typically, an area between 0.90-1 is excellent, 0.80-0.90 is good, 0.70-0.80 is fair, 0.60-0.70 is poor, and 0.50-0.60 is a fail. Table 4.6 lists the ROC curve characteristics for each tissue group pairing. The sensitivity and specificity values shown in Table 4.6 reflect the models in Figure 4.10. Based on this criteria, the data presented in Table 4.6 shows that the QDA model shows excellent classification between invasive carcinoma versus mammoplasty and mammoplasty versus benign (high sensitivity and high specificity, respectively). The other comparisons show fair to good accuracy. However, the specificity with the normal versus mammoplasty and normal versus benign failed in being able to predict mammoplasty and normal type tissues, respectively. The sensitivity was excellent between normal and mammoplasty tissue types, meaning that the QDA model was good at classifying the normal tissues, but not the mammoplasty tissues.

The same analysis (probability QDA models and ROC curve analysis) was performed using the equatorial peak area versus the amorphous scatter as well as the equatorial peak amplitude versus the amorphous scatter. The results were similar to those shown in Figure 4.10, verifying that it is in fact the amorphous scatter that is the dominant classifying variable. The equatorial or axial peak feature respectively permits the construction of the probability plot and in this sense, acts as an independent variable to permit discrimination between tissue types based on the amorphous scatter variable.

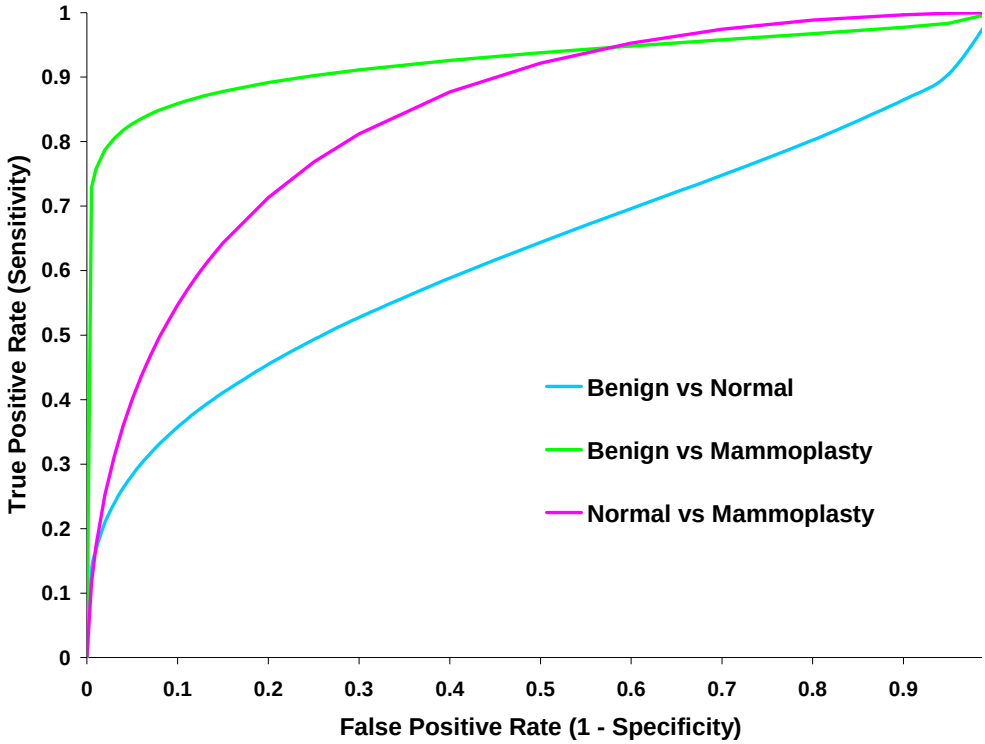
4.4 Discussion

4.4.1 Identifying Optimal Parameters for Classification

Based on our analysis, there is a clear distinction between invasive carcinoma and mammoplasty tissue in terms of the axial d-spacing, the amplitude of the third order axial peak, and the amorphous scatter. Interestingly, tissue from disease-free patients (mammoplasty) and histopathological normal tissue, sampled from patients with known disease, seems to show differences in the amorphous scatter. However, features from benign and normal tissues are not as clearly separated, suggesting that the supra-molecular structures in benign and normal tissues from cancer patients are different to those from mammoplasty tissues with no history or presence of disease. From the pathologist’s assessments, there is no obvious histological difference



(a)



(b)

Figure 4.11: Receiver Operator Characteristic (ROC) curves for (a) the Invasive Carcinoma tissues diagnosed against Benign, Normal, and Mammoplasty tissues and (b) the Benign and Normal tissues diagnosed against each other and Mammoplasty tissues (b). Sensitivities and Specificities are listed in Table 4.6.

between the normal and mammoplasty tissue samples used in this study. However, our results suggest that there may be changes in the fibrillar collagen at the molecular level between these two tissue types that can be characterised using SAXS.

The differences between histopathological normal tissue and mammoplasty tissue are most evident in the average amorphous scatter comparisons of the tissue groups shown in Figure 4.5. The average amorphous scatter of normal tissue is in between the invasive carcinoma and mammoplasty profile intensities (whose 95%CI are clearly separated), suggesting that these normal tissues, as diagnosed by a pathologist, may not be completely devoid of disease. The ROC curves comparing the mammoplasty against invasive carcinoma and mammoplasty against normal tissue types showed high sensitivity (>90%) (Table 4.6), i.e. good true-positive rates for disease. For the invasive carcinoma versus mammoplasty tissues, the ROC curve (Figure 4.11a) showed high sensitivity and high accuracy: large differences between these two tissue groups means that the QDA model subsequently showed good performance. However, this tissue pairing had low specificity (66.7%), meaning that the false-positive rate would be unacceptably high in a diagnostic context. Mammoplasty tissues versus Normal tissues had the lowest specificity in all of the tissue comparisons, which indicated the QDA model is not good at identifying the mammoplasty tissues. However, the high sensitivity means that the QDA model for the mammoplasty and normal tissue comparison was very good at identifying the normal tissues.

High sensitivity is desirable if the consequences of the disease are relatively serious (i.e. fatal or permanently affecting the patient’s quality of life) and the cure, or treatment, is relatively inexpensive and easily available. There is a trade-off between specificity and sensitivity, where a high specificity usually has a low sensitivity associated with it, and vice versa. Models that show high sensitivity and low specificity occur especially when the threshold is very high for inclusion of disease, where patients with a false-positive diagnosis are given unnecessary treatment. However, if inclusion for treatment is conservative, where the threshold is very low (low sensitivity, high specificity), patients who require treatment will be missed. Invasive carcinoma is a serious disease, where the treatments are expensive and have negative side-effects (poor quality of life). Normal tissue is considered as not serious because there are no consequences of normal tissue to the patient’s health. However, a high specificity is needed to identify normal and tissues, since the treatment (if diagnosed incorrectly as invasive carcinoma) is relatively

expensive. In terms of invasive carcinoma, a higher sensitivity would be more acceptable based on a diagnostic and treatment aspect.

Looking at ROC Areas Under the Curves (AUCs), only the QDA models between the mammoplasty tissues and disease (invasive carcinoma and benign) showed high values (AUC > 0.90). It is important to note that the sensitivities and specificities calculated using this QDA model, were comparable with current screening and diagnostic techniques. Mammography has sensitivities between 63-95% and specificities between 14-90% [9, 65, 71]. The large range of specificities with mammography is due to early impalpable lesions detected during screening, and subsequent needle biopsy must be imaged guided to ensure proper sampling. Core biopsy diagnoses are considerably more accurate, with sensitivities between 88-98% and specificities between 95-100%[9, 37, 65]. It is more appropriate to compare our results with core biopsy sensitivities and specificities, since core tissue biopsies were examined in this work. With our data, only the comparison with mammoplasty and benign tissues showed high specificity in the QDA model (97.1%), even when the sensitivity was relatively low (73.3%). This demonstrates that the QDA model also performed adequately to identify true negatives, and was able to separate the benign from mammoplasty tissues. Although the accuracy between normal and mammoplasty tissues was poor, as was also the specificity, the sensitivity was very high. To our knowledge, there have been no studies comparing the sensitivities and specificities between normal and mammoplasty tissue types, and the analysis presented in this thesis is the first to do so.

The importance of the amorphous scatter in distinguishing tissue types was also demonstrated by Fernandez *et al*. [49, 51] who found that the amorphous scatter (in adipose tissue) increases by a factor of 5 in invasive carcinoma tissues in comparison to normal tissues. Our results indicated an increase in scatter by a factor of 1.7 between mammoplasty and invasive carcinoma tissues, and an increase by a factor of 1.4 between benign/normal and invasive carcinoma tissues in the q-range between 0.31-0.46 nm^{-1}. Interestingly, benign tissues have an amorphous scatter in between mammoplasty and invasive carcinoma tissues, suggesting that benign structures have some evidence of increased scattering potential.

The most significant changes in the fibrillar collagen features in the equatorial and longitudinal peak features appear to be between mammoplasty tissue and invasive carcinomas, which may imply that underlying mechanisms involved with the spread of disease is affecting the supra-molecular nature of the collagen. Our data found that the d-spacing in tissues with invasive carcinoma (d = 65.00 ± 0.05 nm) was larger than histopathologically normal (d = 64.88 ± 0.04 nm) and

benign (d = 64.92 ± 0.04 nm) tissues, which agrees with the d-spacing analysis done by Lewis *et al*. [77] However, our results showed a significant distinction in the d-spacing between invasive carcinoma and mammoplasty (d = 64.91 ± 0.02 nm) tissues, whereas Lewis *et al*. [77] found that the d-spacings in mammoplasty tissues were not distinguishable from invasive carcinoma tissue. Lewis *et al*. [77] only had 16 invasive carcinoma patients and 18 mammoplasty patients, whereas this work investigated the axial D-spacings for 42 invasive carcinoma patients and 15 mammoplasty patients. Even with the larger sample size of invasive carcinoma patients, which would be expected to have more variations in disease types, age, and carcinoma grade, this work showed a significant difference in the axial D-spacing between invasive carcinoma and mammoplasty tissues. Fernandez *et al*. [49, 51] did not examine mammoplasty tissues, but did find that the d-spacing in tissue with invasive carcinoma was larger than in histopathologically normal tissues, which we have also confirmed with a much larger patient cohort. However, our point-scan analysis also showed that the d-spacings in benign tissues had significant variance, and cannot be distinguished from the d-spacings in invasive carcinoma tissues. A difference in d-spacing further suggests that a closer examination of the processes involved in changing the molecular components of the collagen may reveal information regarding how neoplasms form

As discussed in the Chapter 1 (Figure 1.5), electron microscopy images of diseased tissues show that diseased tissue exhibit fibrils that are chaotically distributed throughout the tissue and healthy tissue exhibits well-ordered and highly aligned fibrils. The polar angle distribution analysis (Section 4.2.4 Polar Angle Distribution Data) showed that the variation in the orientation of the collagen fibrils was not sufficiently different to distinguish between tissue types. However, examination of the fibril distribution in smaller tissue pieces may show differences in tissue types, as shown in electron microscopy images.

Although disease, whether benign or invasive, appears to have an effect on the molecular structure of the collagen, other factors may contribute to the differences between the tissue groups. Based on t-tests, the ages of the patients in the invasive carcinoma group (60 ± 13 years (standard deviation)) were significantly different from the other tissue groups ($p < 0.01$). Age was not a factor between the histopathologically normal (48 ± 15 years), benign (41 ± 16 years), and mammoplasty (49 ± 16 years) patients ($p > 0.08$). Even though the patients in the invasive carcinoma group were older, it is important to note that the patients in the normal group were taken from patients presenting with benign and invasive

carcinoma lesions, and that these normal patients, who had known presence of disease but not in the tissue sample examined using SAXS, had comparable ages to the mammoplasty patients. Other clinical differences that may affect the collagen structure in patients include: previous chemotherapy or radiotherapy, hormone replacement therapy, menopausal status, child bearing and breast feeding history, and trauma to the breast. The clinical information available for the 130 patients was not sufficiently powerful in order to perform a more detailed analysis to assess the effects of all of these factors.

An advantage of SAXS compared to histological core biopsy analysis is that the entire volume of the biopsy core, which in this study was up to 20 mm^3, can be analysed, where the tissue is fresh and unaltered from its natural state. Pathological analysis of breast tissue biopsies is usually based on chemically fixed and stained tissue slices taken from a core, typically 2 μm thick. Most tissue core biopsy samples are heterogenous, where an array of normal, benign, and invasive carcinoma tissue can be present within one tissue core. Some biopsy cores may only contain 5-10% invasive carcinoma tissue, in which case it is diagnosed as invasive carcinoma as is obviously required in a clinical context. Since histopathologically normal samples from patients with a history of disease displayed SAXS parameters lying in-between those of invasive carcinoma samples and samples from disease-free patients, it may be that SAXS is more tolerant to sampling error (e.g. not obtaining tissue directly from the lesion). When a patient presents with a palpable mass and the biopsy is diagnosed as normal, the diagnosis is suspicious in a clinical context in which it is likely to be a false-negative diagnosis. The analysis presented in this work suggests that SAXS may have the potential at excluding false negative diagnoses, in particular with the discriminations between normal and mammoplasty tissue types. Whilst taking the average SAXS image over the entire tissue sample may not discern between the different tissue types within a biopsy core, it may still be capable of showing underlying changes in the molecular structures within the tissue which warrant further investigation for diagnosis.

4.4.2 Differences between Experiment #3 Analysis and a Combined Experimental Analysis

The results presented and discussed in this chapter have been for a cohort of 130 patients over three experimental sessions. As mentioned in Section 4.1, a subset of 80 patients were analysed for an earlier publication [114]. In this section, the

differences in the data analysed for this subset of patients versus all of the data combined from the three experiments are compared.

Most of the results using the data from all three experimental sessions were similar to the results presented in Paper I [114], where only the data from Experiment #3 was analysed. This was, of course, expected since the analysis of 61.5% (80/130) of the total number of patients was presented in Paper I. The PCA outcomes were equivalent, showing the same SAXS features that could be used to distinguish between tissue types with similar statistical weighting. The amorphous scatter also presented with the same statistical results, where the p-values < 0.05 between each of the tissue types showed significant differences, except for benign and normal tissues ($p = 0.30$ for all data, $p = 0.25$ for Experiment #3 only), further suggesting that the amorphous scatter feature has the potential to be used as a classification variable. The subset of patients in Experiment #3 and the 130 patients from all of the experimental sessions had identical d-spacing results for each tissue groups (within experimental uncertainties). Table 4.7 shows the axial d-spacings for each of the tissue groups for the two data sets. Consequently, the QDA models were also consistent between the two data sets and the ROC curves produced similar comparisons in sensitivity and specificity. This shows that the results are consistent between experiments and that inter-experimental errors are minimal.

One difference between the two patient sets was the equatorial feature analysis regarding the linear comparisons in the amplitude of the Bessel peak with the square root of the Bessel area (Section 4.2.2 Equatorial (Lateral) Variables). Table 4.8 gives the slopes calculated for each of the tissue groups between the two data sets. Although the decreasing trend from healthy to diseased tissue is still apparent, the slopes for the mammoplasty, normal, and benign tissues using all three experiments are equal within standard error (SE). However, invasive carcinoma tissues are significantly different from the other three tissue groups as was also seen with Experiment #3 only. The differences between tissue types with their slopes was more prominent with Experiment #3, suggesting that the variance is larger for bigger patient populations, which may also produce larger standard errors in the data.

Table 4.7: Comparison of axial D-spacing between the data from all three experimental sessions and only Experiment #3 for each of the tissue types. The number of patients in each tissue group (All three experiments/ Experiment #3) is shown in the brackets in the first column.

Tissue Group	130 patients – All three experiments (±SD)	80 patients – Experiment #3 only (±SD)
Invasive Carcinoma (42/23)	65.00 ± 0.05 nm	64.97 ± 0.05 nm
Benign (36/27)	64.92 ± 0.04 nm	64.93 ± 0.04 nm
Normal (35/21)	64.88 ± 0.04 nm	64.88 ± 0.04 nm
Mammoplasty (17/9)	64.91 ± 0.02 nm	64.89 ± 0.03 nm

Table 4.8: Comparison of the slopes calculated using the Bessel Amplitude vs √Bessel Area between the two data sets.

Tissue Group	Slope (SE) All Experiments	Slope (SE) Experiment #3 Only
Mammoplasty	5841 (342)	6385 (1075)
Normal	5601 (371)	5748 (492)
Benign	5531 (374)	4211 (227)
Invasive Carcinoma	3237 (231)	2702 (297)

4.5 Conclusions

We have found that the collagen fibril d-spacing, the axial peak amplitude, and the area of the lateral peaks (equatorial peak features) were distinguishing parameters between tissue types. This finding agrees with previous studies [41, 44, 49, 51, 77, 102, 128]. The amorphous scatter showed the best separation between tissue types. Invasive carcinoma tissues and mammoplasty tissues showed the most separation, and consequently when these tissue types were modelled using the amorphous scatter as the discriminating parameter, they also had high sensitivity and specificity values. Benign and normal (presence of disease) tissue parameters tend to have values between those of invasive carcinoma and mammoplasty tissues, suggesting that the amorphous scatter and axial D-spacing can indicate changes in the molecular structures of the tissues in the absence of conventional indications seen in histopathology assessment. These properties need to be examined more closely to

understand how disease contributes to the changing molecular structures in the breast.

The ability of SAXS to distinguish normal tissue from disease-free tissue gives a novel means of providing insight into disease presence and progression. It may also be useful in conjunction with current diagnostic methods, to assist in early detection or more information on the extent of disease with defining surgical margins. In particular, significant differences between invasive carcinoma tissues and normal tissues were evident with the amorphous scatter and axial D-spacing, which is consistent with preliminary results from Lewis *et al*.[77] who examined tissue samples from 5 patients taken at 2 cm intervals outward from the centre of the primary tumour (centre = invasive carcinoma, interval tissues = normal). They found an increase in the intensity of the diffraction peaks with distance away from the centre lesion. In other words, the amplitude of the peaks was lower for the centre points (that were invasive carcinoma tissues) compared to the surrounding tissue samples (that were typically normal parenchyma).

In the next chapter (Chapter 5), we present data that has mapped the changes in collagen fibril parameters as a function of distance from the centre of the invasive carcinoma lesion in mastectomy patients, in order to determine if systematic differences can be seen over macroscopic distances (up to 10 cm from the primary lesion). We examine a much larger cohort of patients than has previously been studied, and focus particularly on the parameters discussed in this chapter that have shown to have the best discrimination potential between tissue groups.

Chapter 5

Analysis of Series Site Tissue Samples using a Synchrotron Source

5.1 Introduction

In the previous chapter (Chapter 4), the most promising SAXS features that can be used for classification of the breast tissue groups were identified, which addressed the first aim of this thesis. These features were:

(1) the total scattering intensity (i.e. predominantly the amorphous scatter),

(2) the axial D-spacing of the third order axial peak,

(3) the axial peak amplitude, and

(4) the equatorial peak area and the equatorial peak amplitude.

This chapter further investigates these features in relation to tissue samples taken at 2 cm increments away from the centre of the palpable lesion (referred to as the primary site), imaged with a synchrotron source. These samples are referred to as series tissue samples (see Section 2.1.2). The second aim of this thesis is addressed in this chapter, where systematic changes in tissue structure were mapped in relation to distance away from the centre of the primary lesion.

The series tissue analysis presented in this chapter focuses on the amorphous scatter and third order axial peak features. Because there was evidence that SAXS showed structural changes between diseased and non-diseased breast tissue types (Chapter 4), particularly between normal and mammoplasty tissues, it was possible

to see systematic changes in these features throughout the mass of the breast. If proven to correlate with disease, these changes can be mapped over the breast to potentially show areas of high risk disease or to identify surgical margin extraction. Lewis *et al*. [77] and Round *et al*. [102, 103] are the only studies that we know of that have examined breast tissue samples at distances away from the centre of the primary mass. However, Lewis *et al*. [77] examined 5 patients for changes with distance, where the fraction of total scattered intensity in the diffraction peaks increased with distance away from the primary site. Round *et al*. [102] examined 4 patients with 17 tissue samples taken at varying distances from the tumour. Round *et al*. [102, 103] found that axial peak features (peak area and axial D-spacing) showed clear distinctions between invasive carcinoma and normal tissues, but there was no evidence of trends in these features with distance away from the primary site of the tumour. With a larger sample population, changes in SAXS features with distance from the tumour may be possible to identify. The results presented in this chapter represent the first study with systematic sampling of breast tissue of a large cohort of patients where SAXS images were taken from the centre of the primary lesion and at regular intervals away from this centre.

The results presented in this chapter have been submitted for publication in two journals:

Paper II: Cancer Research [113]

Paper III: Medical Physics [112].

The article submitted to Cancer Research summarises the clinical relevance of the results and the article submitted to Medical Physics discusses the physical origins of the SAXS features analysed, and also presents a more comprehensive statistical analysis of the results. The submitted manuscripts of these two papers are included at the end of this thesis as supplementary material.

5.2 Methods and Materials

5.2.1 Tissue samples

Table 5.1 lists the tissue samples and their diagnosis by the pathologist at each distance point for the 56 patients who had series tissue excised from their breast mass. A total of 357 tissue cores sampled from 14 mastectomy patients and 42

Table 5.1: Breakdown of the number of tissue samples and their diagnosis at each distance point for the 56 patients who had series tissue data taken. (Also see Table 2.4)

Distance Point	Total # of Tissue Samples	Invasive Carcinoma	Benign	Normal
0 cm	56	31	6	19
N(+2)	52	15	1	36
N(+4)	30	6	0	24
N(+6)	22	3	0	19
N(-2)	44	11	1	32
N(-4)	21	4	2	15
N(-6)	15	2	0	13
P(+2)	30	15	0	15
P(+4)	21	5	1	15
P(+6)	12	2	1	9
P(-2)	32	11	1	20
P(-4)	15	5	0	10
P(-6)	7	1	0	6

Wide-Local-Excision (WLE) patients were analysed: 56 samples at 0 cm (primary site), 158 samples at 2 cm, 87 samples at 4 cm, and 56 samples at 6 cm. The amorphous scatter and the third order axial peak parameters (peak amplitude, Full Width at Half Maximum (FWHM), area, and axial D-spacing) were analysed for changes with distance away from the primary site of the lesion. Of these 357 tissue samples, 111 samples were histopathologically invasive carcinoma, 13 were benign, and 233 were diagnosed as normal parenchyma. All of the normal tissue analysed in this chapter was histopathologically diagnosed as normal parenchyma from patients with known invasive disease within the breast. Mammoplasty tissue was not analysed in this chapter since series tissue was not harvested from disease-free patients.

5.2.2 SAXS Data Analysis

Section 5.3 analyses the third order axial peak features (axial D-spacing, amplitude, FWHM, and area) and the amorphous scatter using the series tissues for 56 patients. Systematic changes in these parameters were examined at each distance increment (~2 cm) from the centre of the primary lesion, where general trends were evident over the distance increments. Directional trends of these parameters along the P-axis (superior-inferior of the breast) and N-axis (nipple-axilla line of the breast) (see Figure 2.2 and Figure 2.3) were also examined to determine if changes in tissue structure favoured certain areas of the breast. Invasion and metastasis is known to

occur primarily along the nipple-axilla line (N-axis) where lymphatic ducts and vascular tissues are typically more concentrated and facilitate disease spread [28, 32]. Thus, structural changes in the tissue along the N-axis would be expected to be more prominent than along the P-axis.

As discussed in Chapter 4, the amorphous scatter and axial D-spacing were the two parameters that showed the most significant differences between tissue types. These two parameters are explored in Section 5.4 in the form of 2-dimensional (2D) maps. Because there were a limited number of known data points over the breast mass (up to 13 tissues were sampled from each patient), an interpolation of the data in between the known point values was estimated based on a discrete convolution structure which was computed quickly using the fast Fourier transform (FFT). This interpolation procedure is known as Gaussian Kernel Estimation, and is discussed in detail in Appendix E. The estimation uses a weighted probability to determine the values in between the known data points and was calculated for a 32 x 32 point-grid for each patient.

Wrap around effects were seen in the 2D maps when less than 6 tissue samples were available for calculation. These arose from the FFT calculation and smoothing criterion used to interpolate the values between the data points. Calculating the map using a padded grid of zero values could be used to show the areas of low and high amorphous scatter or axial D-spacing values without the wrap around effects. This issue is discussed in Appendix E**;** however, the padded maps cannot be taken as accurate around the edges, where the extended distance was set as zero, especially when considering the gradient of the values around the extended area.

Mapping of the structural changes over a 2D area allows for prediction of trends in areas between sampling sites or beyond the known edges of the excised mass to hypothesize possible directions or areas of disease extension. The 2D maps cannot be used as a predictive model, but rather provides information relating to directional trends, and areas, of structural changes in the breast. The implications of the results and the mapping techniques are discussed and compared with the histological diagnosis of the tissue samples for confirmation.

Table 5.2: P-values for the amorphous scatter and axial D-spacing values between the centres of the primary lesions and the tissue samples taken at 2 cm intervals over both axes. The number of tissue samples at each distance point and the diagnosis is listed in Table 5.1.

	Amorphous Scatter	Third Axial Peak D-Spacing	Peak Amplitude	Peak FWHM	Peak Area
Centre (0) & All (2,4,6 cm)	2.36×10^{-15}**	0.010*	0.14	0.61	0.065
Centre (0) & 2 cm	2.07×10^{-11}**	0.020*	0.024*	0.72	0.017*
Centre (0) & 4 cm	4.29×10^{-10}**	0.014*	0.22	0.37	0.15
Centre (0) & 6 cm	7.99×10^{-9}**	0.081	0.83	0.018*	0.60
2 cm & 4 cm	0.61	0.83	0.44	0.17	0.57
2 cm & 6 cm	0.25	0.54	0.022*	0.0094**	0.13
4 cm & 6 cm	0.49	0.44	0.19	0.077	0.43

*Significantly different $p < 0.05$
**Significantly different $p < 0.01$

5.3 Systematic Changes with Distance

Data was analysed by distance from the primary lesion, which is defined as being at 0 cm or the centre position, and at radial increments of ~2 cm for all of the patients. For example, values for samples taken at N(+2) N(-2) P(+2) and P(-2) for all of the patients were averaged to give a value at a radial distance of 2 cm from the centre of the primary lesion. The data was also analysed according to the two axes examined: the superior-inferior direction of the breast (P-axis), and the nipple-axilla direction of the breast (N-axis). The axis analysis was performed in order to determine if structural changes in each of the quadrants of the breast showed significant differences, which would have affected the grouped radial analysis.

First, the data was averaged at each of the radial intervals and grouped together irrespective of patient or tissue type in order to determine if there was a general trend in parameters with distance away from the centre of the primary lesion. In order to determine if the difference between the centre point and the distance interval points were significant, p-values were calculated using t-tests and are shown in Table 5.2 for the amorphous scatter, third axial peak D-spacing, and the other peak features. The amorphous scatter values showed the most significant difference

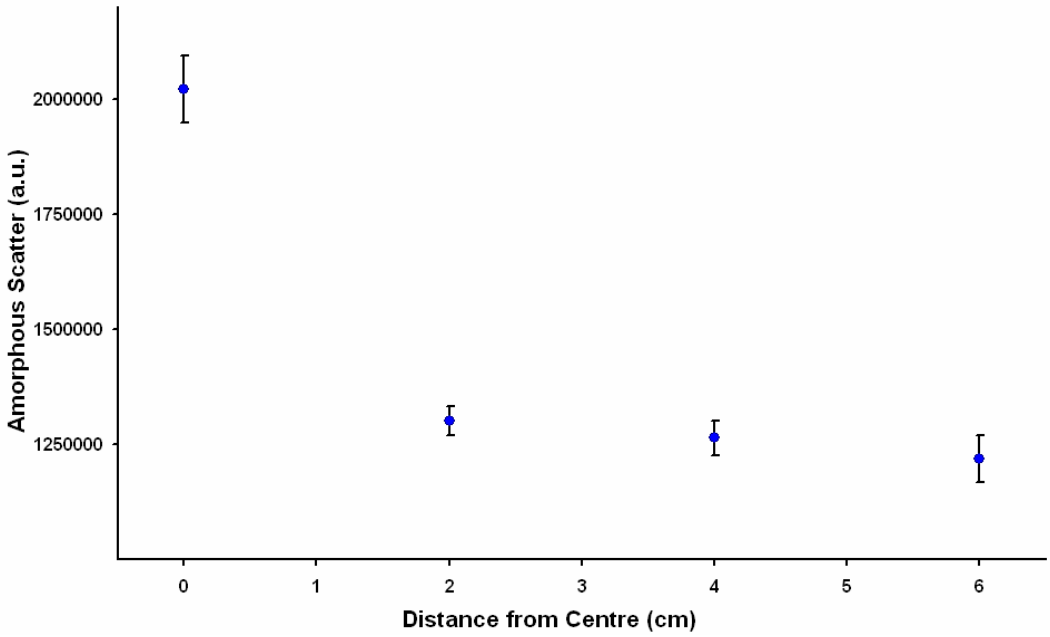
($p = 2.36 \times 10^{-15}$) between the centre and all distance points. The third order axial peak D-spacing also showed a consistent significant difference ($p = 0.010$) between the centre points and the distance points. Most of the other comparisons (third order axial peak amplitude, width, and area), including between the distance points (for example, between 2 cm and 4 cm), had $p > 0.05$, and therefore showed no significant differences. The comparisons with the 6 cm tissue points, which showed worse p-values in comparison with the other distant points, can possibly be explained by the fewer tissue samples in this distance group[15], which would diminish the statistical power of the calculations. Also, the differences between the distance groups are not consistent with both the peak amplitude and peak FWHM parameters. In particular, the p-values between the centre and 2 cm points, and the centre and 6 cm points are not as consistent as the amorphous scatter and axial D-spacing parameters.

Since the amorphous scatter and the axial D-spacing showed consistent evidence of changes with distance, these two parameters were investigated further. The mean values with respective 95% confidence intervals (CI) for the amorphous scatter and axial D-spacing parameters for each of the 2 cm distance interval points are shown in Figure 5.1a and Figure 5.1b, respectively. Both parameters show a consistent decrease in value from the centre of the lesion to 2 cm away, then a plateau between 2 cm and 6 cm.

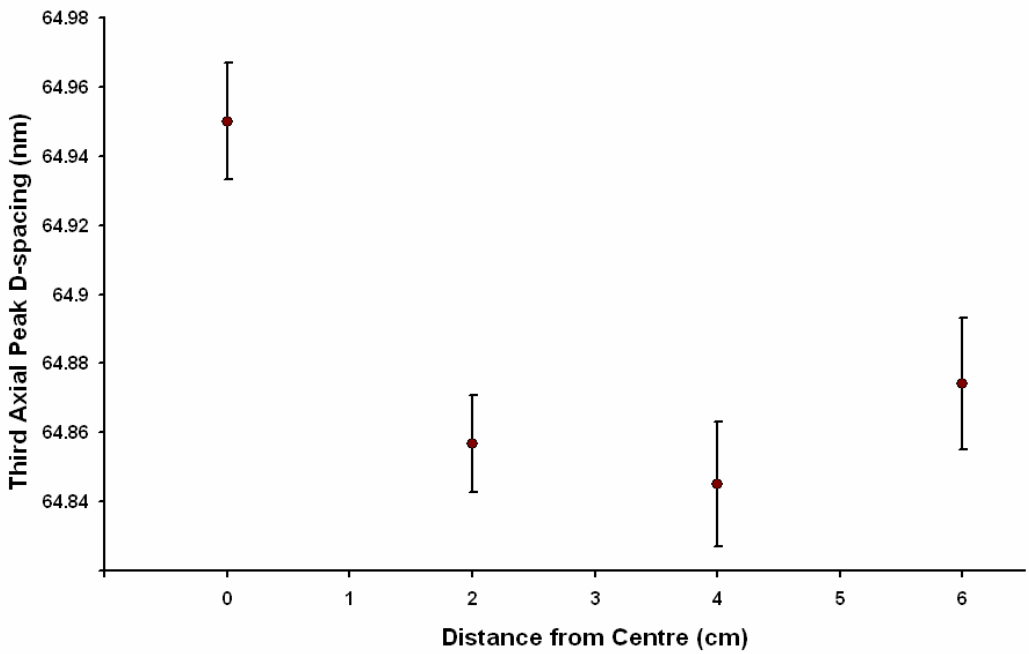
These two parameters were also analysed along each axis (P-axis and N-axis) in order to determine if directional trends were evident. Figure 5.2 shows the average values and their 95%CI along the N-axis and the P-axis for the amorphous scatter variable. There was a significant difference between the amorphous scatter values between the centre and distant tissues ($p < 0.01$). However, there were no significant differences between the tissues at distances other than 0 cm ($p > 0.01$). This suggests that the direction of sampling does not affect the general trend in the amorphous scatter, i.e. there is no directional preference in the scattering potential of tissue in the breast.

The third order axial D-spacing is shown in Figure 5.3 for the data along the two axes. There was no evident trend in the axial D-spacing for the N-axis, but there was some evidence of a decreasing trend from the centre to the P(+2), P(+4), and P(-4) points, as indicated in Figure 5.2b. The p-values comparing the axial D-spacing at the centre and the distance tissues (0 cm & All (2, 4, 6 cm)) were significantly

[15] Not all of the breast masses from mastectomy patients were large enough to take tissue samples as far away as 6 cm.



(a)



(b)

Figure 5.1: Average values for the (a) amorphous scatter and the (b) third axial peak D-spacing for the centre tissue points and at the 2 cm intervals away from the centre. Error bars represent the 95% confidence intervals. The number of samples and the diagnoses at each distance increment are listed in Table 5.1.

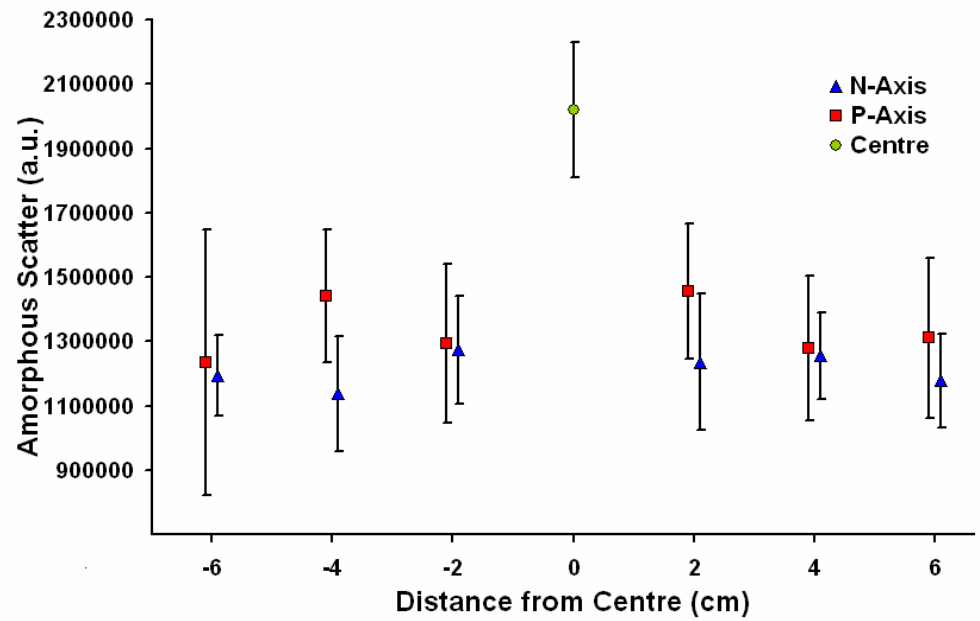


Figure 5.2: Average values for the amorphous scatter at each distance interval along the N-axis and the P-axis. Negative distances represent tissues sampled toward the axilla (N-) or inferior from the primary lesion (P-). Positive distances represent tissues sampled toward the nipple (N+) or superior from the primary lesion (P+). Error bars represent 95%CI. The number of tissue samples and diagnostic classifications of the tissues at each distance interval are listed in Table 5.1.

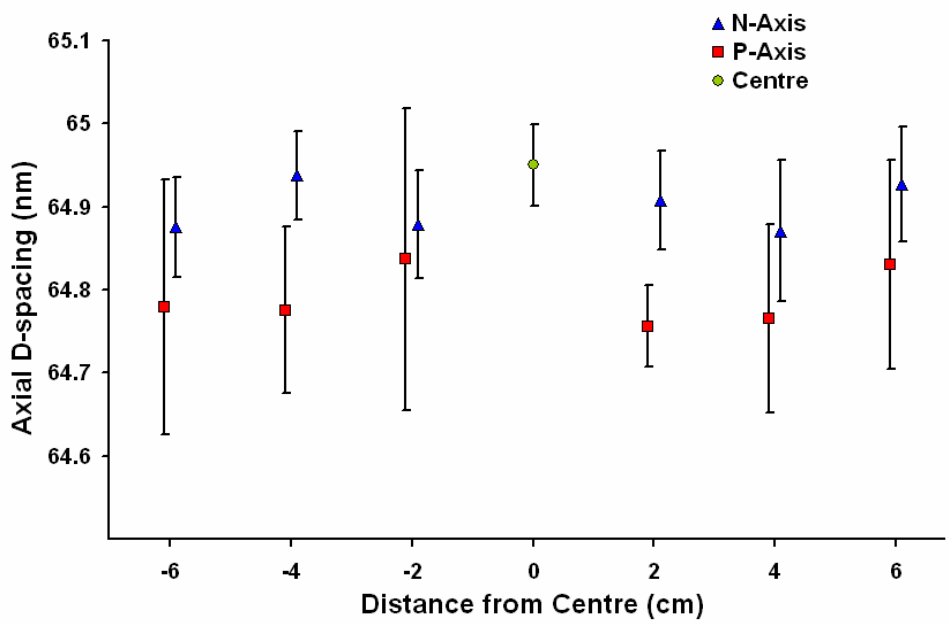


Figure 5.3: Average values for the axial D-spacing at each distance interval along the N-axis and the P-axis. Negative distances represent tissues sampled toward the axilla (N-) or inferior from the primary lesion (P-). Positive distances represent tissues sampled toward the nipple (N+) or superior from the primary lesion (P+). Error bars represent 95%CI. The number of tissue samples and diagnostic classifications of the tissues at each distance interval are listed in Table 5.1.

different with the P-axis ($p = 0.00049$), but showed no significant difference in the N-axis ($p = 0.12$). Thus, the decreasing trend in the axial D-spacing seen in Figure 5.1b is primarily due to the tissue samples from the P-axis. Although the amorphous scatter showed a decrease from 0 cm to the distance points for both axes (i.e. Figure 5.1a is correlated with Figure 5.2), the decreasing trend in the axial D-spacing arises from the P-axis only. This was expected, since the collagen components along the nipple-axilla line were expected to be similar to the diseased lesion if spread of disease is most likely to occur along the N-axis before spreading inferiorly and superiorly in the fatty connective tissues along the P-axes [28, 32].

5.4 Mapping Parameter Changes in Individual Patients

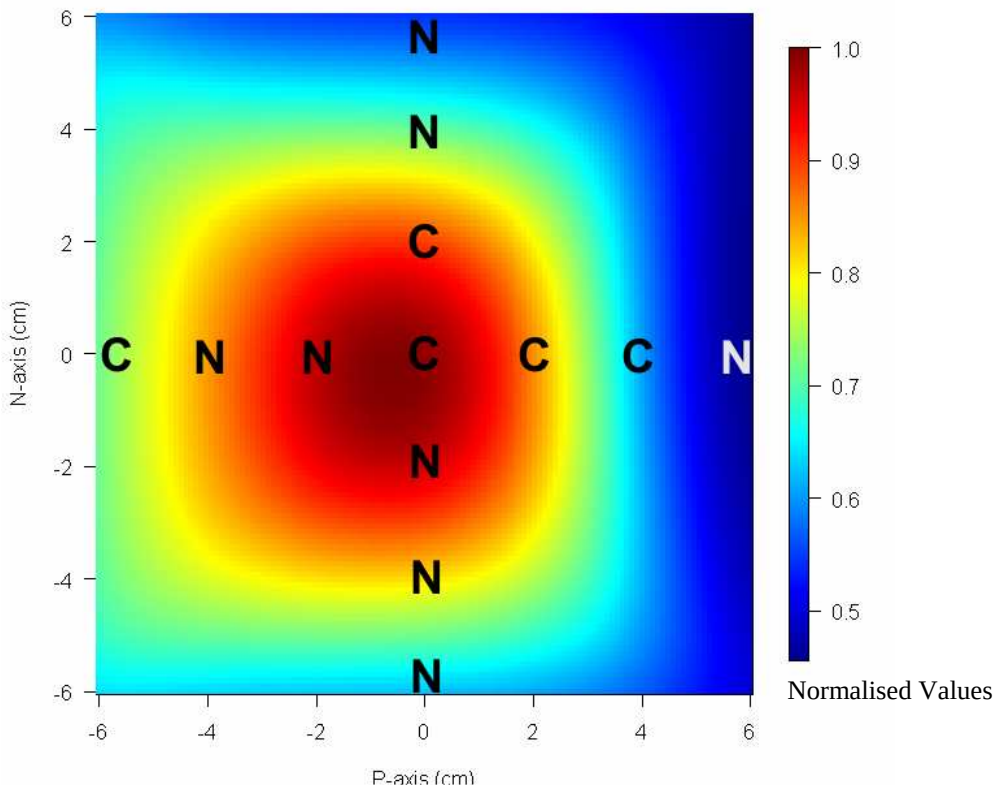
A spatial model, using Gaussian Kernel Estimation (discussed in Section 5.2.2 and further detailed in Appendix E) was used to visualise the values of the amorphous scatter and third order axial D-spacing parameters for each patient in 2-dimensions (2D). Each parameter was mapped onto a 2D grid of 32 x 32 pixels. Each patient's amorphous scatter data was normalised by the maximum value for each patient to allow comparisons between each patient, where the trends can be observed, rather than the differences in absolute values. In order to preserve an absolute scale for the axial D-spacings, these were not normalised for each patient.

The 2D models for four patients who had close to the maximum possible 13 tissue sample set that could be excised from one patient are shown in Figure 5.4-Figure 5.7 using the amorphous scatter (part (a) of each figure) and the third order axial D-spacings (part (b) of each figure). Figure 5.4 and Figure 5.5 show mastectomy patients with nearly full data sets (13 tissue samples) and Figure 5.6 and Figure 5.7 show wide-local-excision (WLE) patients where the mass removed from the breast was smaller than the entire breast (for breast conserving surgery). The histopathological diagnosis for each of the tissue samples is overlaid for comparison. Using this mapping model, red (high values) indicates areas of probable invasive carcinoma, and blue (lower values) indicates areas of probable normal tissue, based on the results from Chapter 4. On first examination, the distribution of invasive carcinoma vs. normal tissue suggested by the mapping agrees well with the histopathological diagnosis. This correlation between map intensities (based on color: red = high, blue = low) and histopathology was also consistent with all 56 patients studied.

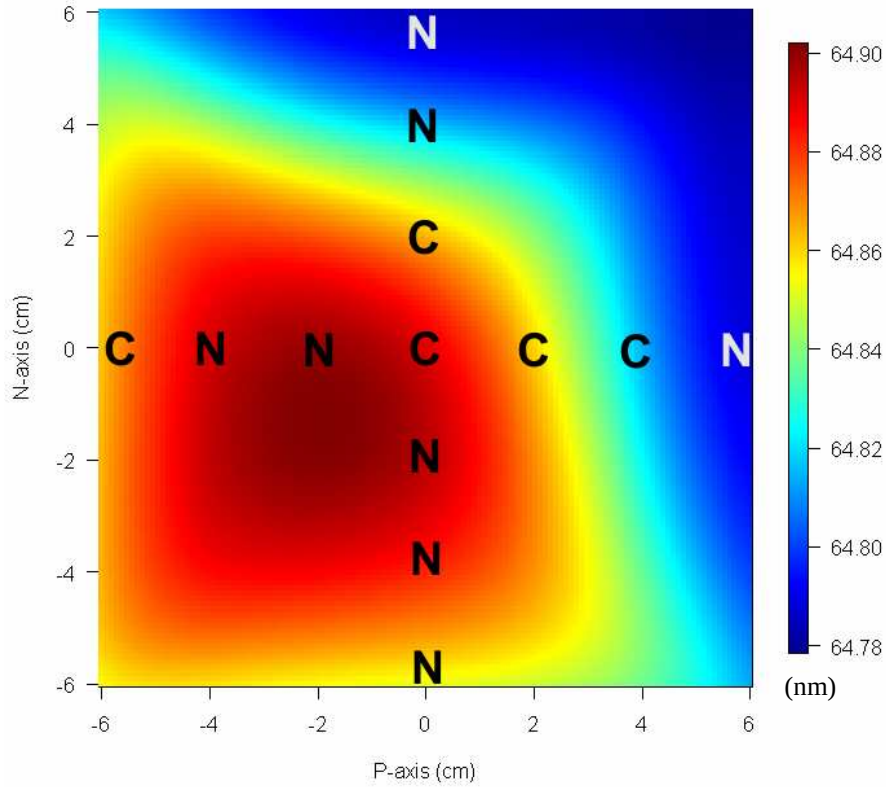
Patient 133 (Figure 5.4) was typical of those patients with confirmed invasive disease at the centre point. Large values of both the amorphous scatter and axial D-spacing were seen at the centre (0 cm) and decreased towards the margins. With patient 133, in particular, the red area is positioned slightly towards the negative side of the P-axis, which is representative of the inferior part of the breast area. This is reflective of the P(-6) tissue point containing invasive carcinoma at this margin (i.e. the map shows evidence of disease in that direction). Patient 35 (Figure 5.5) shows some agreement with histopathology diagnosis with the amorphous scatter mapping, where a red area appears in the N(-6) region. The third order axial D-spacing for this patient (Figure 5.5b) suggests that disease is mainly in the N(-) region, showing that the origin of disease may be around the N(-4) area, rather than at the centre. However, the amorphous scatter model in Figure 5.5a did not show the histopathologically confirmed disease at the centre and P(+2) areas as effectively as the axial D-spacing model shown in Figure 5.5b. Patient 35 had axillary dissection with the removal of lymph nodes, which was deemed necessary by the surgeon because invasive disease was feared to have reached the axilla[16]. The maps for this patient suggest that disease spread is toward the axilla direction (the N(-) direction).

The amorphous scatter and axial D-spacing maps of Patient 53 and Patient 85 are shown in Figure 5.6 and Figure 5.7, respectively. These two patients did not have the full 13 tissue samples excised from their breast mass, but were wide-local-excision (WLE) patients who had breast masses smaller than 12 cm in diameter. The amorphous scatter maps for Patient 53 (Figure 5.6a) showed good agreement with histopathological analysis, where the red areas of the map correspond to the invasive carcinoma tissues and the normal tissue diagnosed at N(-2) and P(+4) are blue on the map. The axial D-spacing map for Patient 53 (Figure 5.6b), however, suggests that there is invaded tissue at the N(-2) point and normal tissue at the P(-4) point, which does not agree with histopathological diagnosis. The other WLE patient, Patient 85 (Figure 5.7), shows good agreement in the colour mapping with histopathological diagnosis, except with the tissue point at P(-4). Both the amorphous scatter and axial D-spacing maps correlate the tissue point at P(-4) to be normal tissue (map is blue in this area), whereas histopathology found this tissue sample to contain invasive carcinoma.

[16] It is unknown if this patient had positive lymph nodes removed from the axilla because it was not recorded in the patient's medical chart. However, the surgeon who performed the surgery confirmed that Patient 35 did show evidence of disease spread towards the axilla.

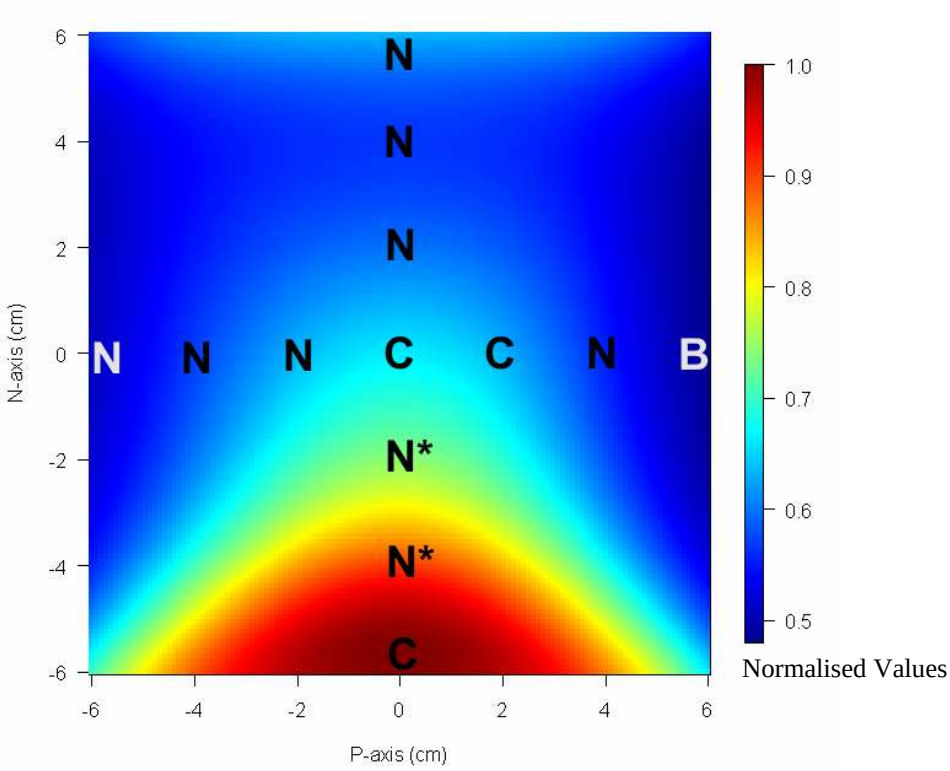


(a)

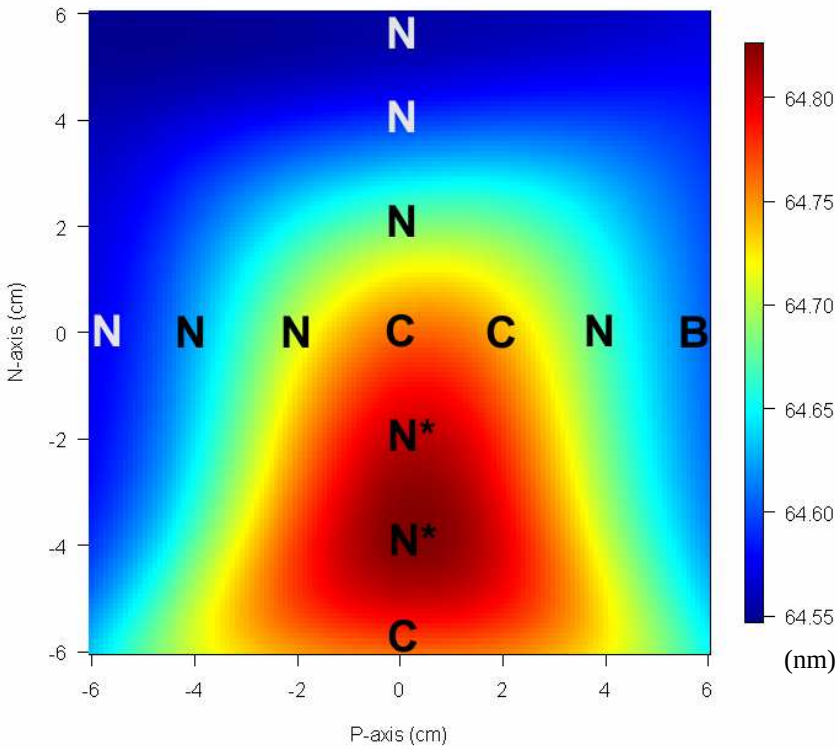


(b)

Figure 5.4: 2D models for Patient 133, mapping (a) the amorphous scatter and (b) the axial D-spacing. Histopathological diagnosis is overlaid at each tissue sample point: C represents tissue samples diagnosed as invasive carcinoma and N represents tissue samples diagnosed as normal parenchyma.

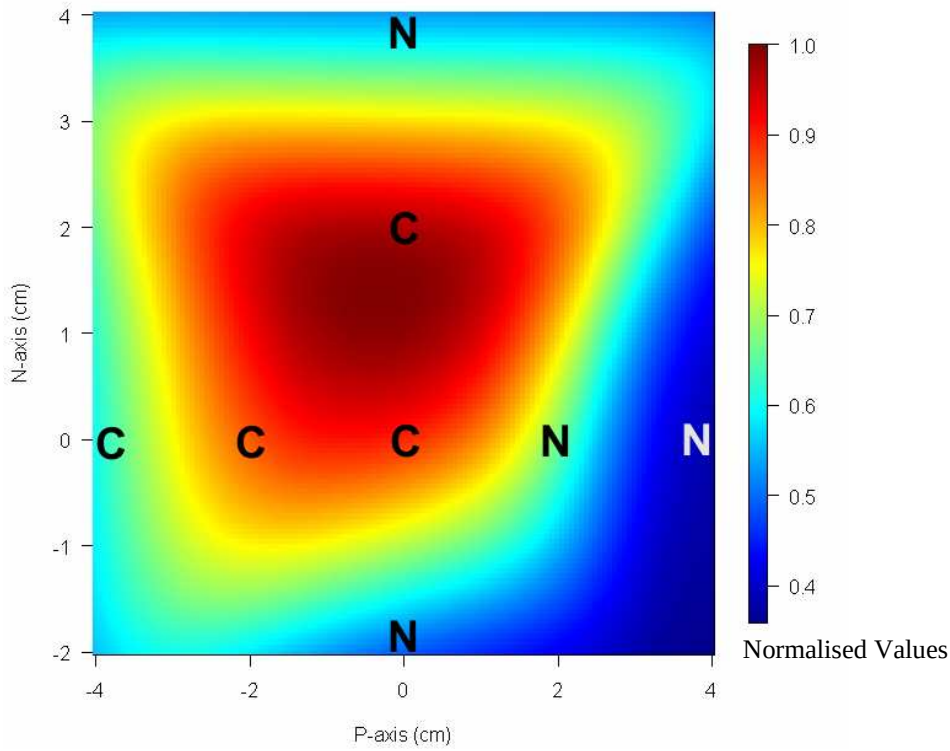


(a)

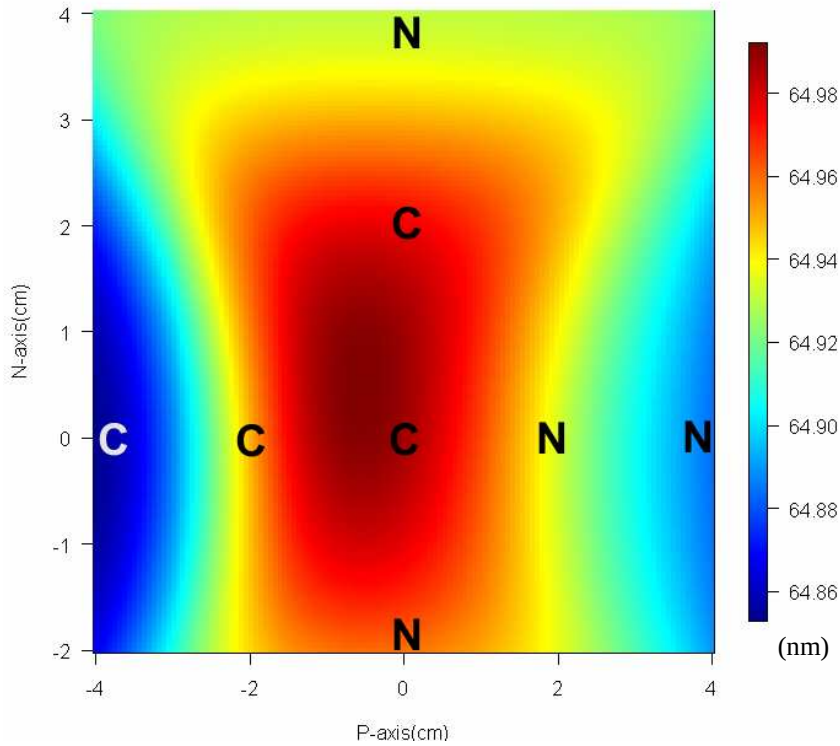


(b)

Figure 5.5: 2D models for Patient 35, mapping (a) the amorphous scatter and (b) the axial D-spacing. Histopathological diagnosis is overlaid at each tissue sample point: C represents tissue samples diagnosed as invasive carcinoma and N represents tissue samples diagnosed as normal parenchyma. B represents benign diagnosed tissue. N* is tissue that could not be diagnosed due to artefacts and degradation in the stained slide, but was diagnosed from a matched tissue sample used for laboratory camera examination.

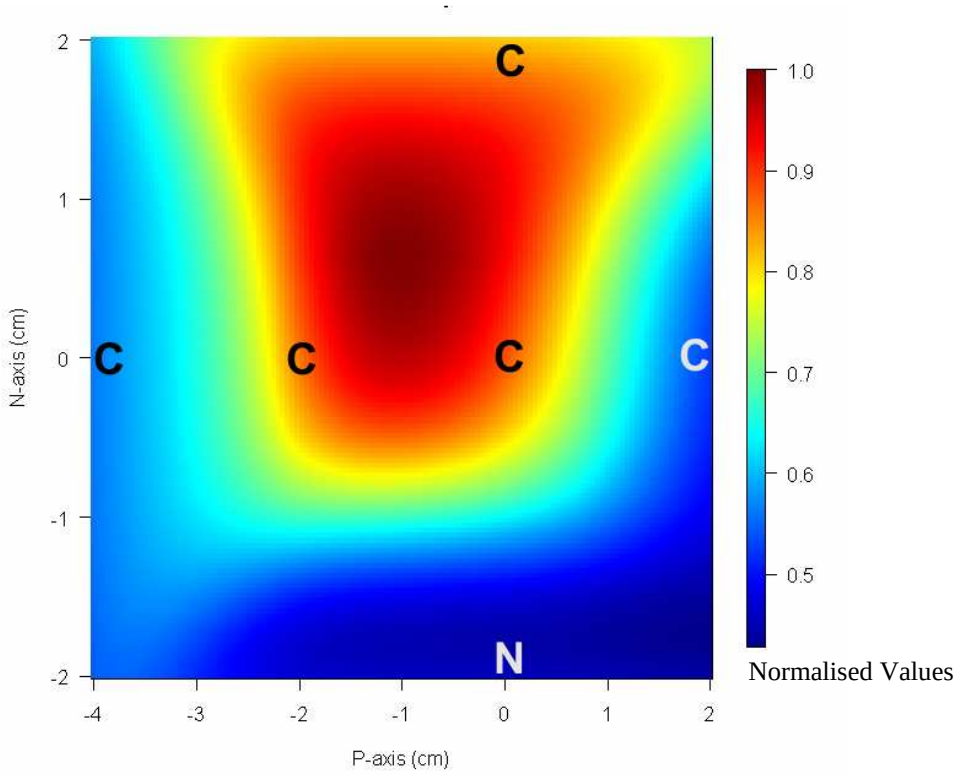


(a)

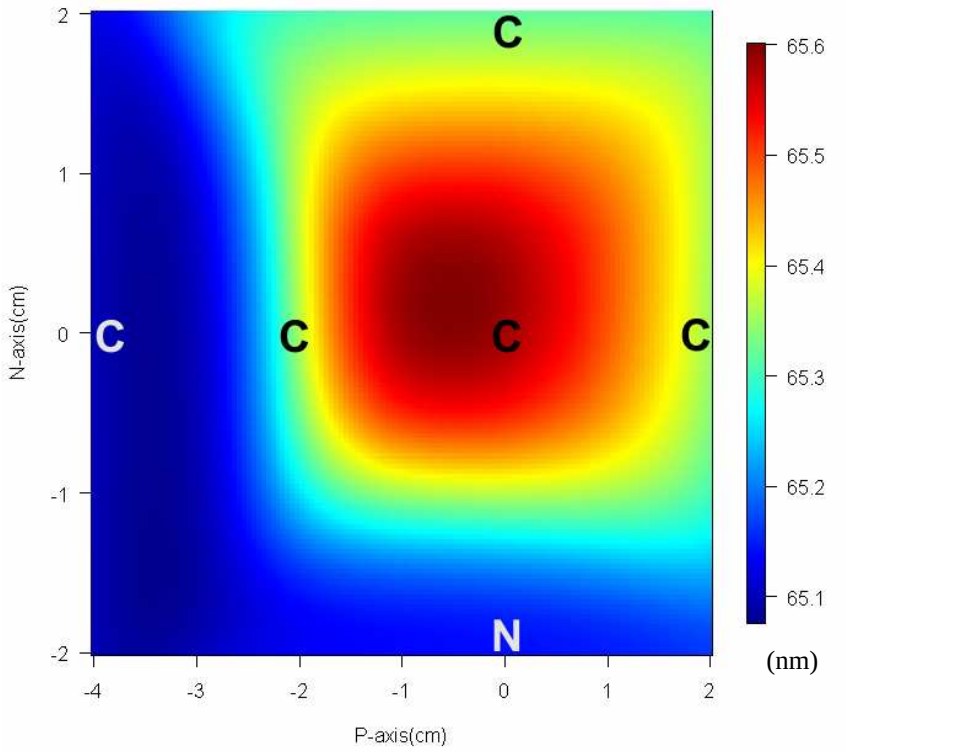


(b)

Figure 5.6: 2D models for Patient 53 (WLE), mapping (a) the amorphous scatter and (b) the axial D-spacing. Histopathological diagnosis is overlaid at each tissue sample point: C represents tissue samples diagnosed as invasive carcinoma and N represents tissue samples diagnosed as normal parenchyma.



(a)



(b)

Figure 5.7: 2D models for Patient 85 (WLE), mapping (a) the amorphous scatter and (b) the axial D-spacing. Histopathological diagnosis is overlaid at each tissue sample point: C represents tissue samples diagnosed as invasive carcinoma and N represents tissue samples diagnosed as normal parenchyma.

A quantitative comparison of the series samples to the histological data was also performed based on the results from Chapter 4, which showed that the amorphous scatter values for invasive carcinoma tissues were significantly higher compared to the benign and normal tissue. A threshold value was used to assign a tissue type to the mapping data: samples above the threshold were designated invasive carcinoma, and samples below the threshold were designated normal/benign. The threshold value of 1920000 was chosen to lie between the lower 95%CI for invasive carcinoma tissues and the upper 95%CI for benign/normal tissues[17] (as seen in Figure 4.5, Section 4.2.3). The results are listed in Table 5.3 for all 56 patients who had series tissue sampled from their excised breast mass. Using the amorphous scatter as a discriminator, only 21/111 of invasive carcinoma tissues and 212/233 of normal tissues were correctly identified, which is equivalent to a sensitivity of 19% and specificity of 91%, respectively. Thus, the amorphous scatter was good at predicting normal tissue, but was not reliable at identifying invasive carcinoma tissue.

In order to determine if the diagnosis of the centre point (primary site at 0 cm) tissue sample affected the sensitivity and specificity, the amorphous scatter for the 22 patients (out of 56 patients who had series data) who had histopathologically confirmed invasive disease at their centre point are listed in Table 5.4. The subsequent classification of the 121 series tissues for these 22 patients with reference to the threshold value used, yielding an improved sensitivity of 55% for the primary sites. The sensitivity in distinguishing invasive carcinoma tissue was better, but the specificity (80% of normal tissue was correctly identified) was slightly compromised. Both sensitivities for the benign and normal tissues for these 22 patients increased slightly compared to the data for all 56 patients analysed in Table 5.3, but the specificities also decreased.

[17] Figure 4.5 (Section 4.2.3), the lower 95%CI for invasive carcinoma was ~1950000 and the upper 95%CI for benign and normal tissues was ~1880000. An intermediate values of 1920000 was used as the threshold value.

Table 5.3: The number of tissue samples with amorphous scatter values designated as invasive carcinoma and benign/normal using mapping data. Samples above the threshold of 1920000 are considered invasive carcinoma and samples below the threshold of 1920000 are considered normal/benign (absolute values).

Tissue Type	Total # of Tissue Samples	# of Tissue Samples > Threshold (Inv. Carcinoma)	# of Tissue Samples < Threshold (Normal/Benign)
Invasive Carcinoma	111	21 (19%)	90 (81%)
Benign	13	4 (31%)	9 (69%)
Normal	233	21 (9%)	212 (91%)

Table 5.4: The number of tissue samples with amorphous scatter values designated as invasive carcinoma and benign/normal using mapping data. Samples above the threshold of 1920000 are considered invasive carcinoma and samples below the threshold of 1920000 are considered normal/benign (absolute values). The tissue samples are from the 22 patients who had histopathologically confirmed invasive carcinoma at their primary site. The normal and benign tissue samples are series samples from the 22 patients who had confirmed disease at their primary site.

Tissue Type	Total # of Tissue Samples	# of Tissue Samples > Threshold (Inv. Carcinoma)	# of Tissue Samples < Threshold (Normal/Benign)
Invasive Carcinoma (Primary Site)	22	12 (55%)	10 (45%)
Invasive Carcinoma (Series Sites)	37	6 (16%)	31 (84%)
Benign (Series Sites)	5	2 (40%)	3 (60%)
Normal (Series Sites)	79	16 (20%)	63 (80%)

5.5 Discussion

Specific trends with distance from the primary lesion in both the amorphous scatter and the axial D-spacing were evident when inspecting the values grouped together for the entire population (Section 5.3), as well as with individual patient analysis (Section 5.4). There were obvious changes in these two parameters along the two axes studied (Figure 5.1-Figure 5.3), where a significant decrease in both the amorphous scatter and axial D-spacing was evident between the primary site at the 0 cm and tissue points ≥ 2 cm away. The decreasing trends with distance away from the primary site of both of these parameters agree with the results found in Chapter 4, where invasive carcinoma tissues had higher amorphous scatter and axial D-spacing values compared to normal and benign tissue.

The amorphous scatter showed consistent trends with distance away from the primary site of the lesion as well as with the 2D mapping of the amorphous scatter for individual patients. The absolute values of the amorphous scatter showed a general decreasing trend from the centre of the primary lesion to the outward margins of the dissected mass. The decrease in the amorphous scatter with distance suggests that it may be an important component in distinguishing the spread of disease within the breast. A decrease in the amorphous scatter value from 0 cm to 2 cm, seen in Figure 5.1a and Figure 5.2, suggests that invaded tissue has a higher scattering potential compared to tissue at 2 cm and beyond, which may be caused by a disruption of the molecular and supra-molecular structures in cancerous regions. As discussed in Chapter 1, such an idea is consistent with the expression of matrix metalloproteinases (MMPs) in malignant tumours, which facilitate tumour cell invasion and metastasis by removing the physical barriers to invasion through degradation of the extracellular matrix (ECM) components [6, 39, 84, 88, 89, 110, 116] (see Section 1.2.3).

In Chapter 4, the primary site tissue samples from the centre of the lesion had larger axial D-spacing for malignant tissue than for normal tissue ($p = 0.0050$, Table 4.3). This has been interpreted as an increase in the axial periodicity with malignancy [42, 62, 68, 113]. It was also observed, as seen in Figure 5.1b, that the periodicity of the axial stacking of collagen fibrils in breast tissue decreased with distance away from the primary site of the diseased lesion. This confirmed the earlier results in Chapter 4 that the axial D-spacing is indeed sensitive to changes in tissue disease status.

When the data was examined along each axis with the axial D-spacing (Figure 5.3), the P-axis appeared to show a weak trend of decreasing axial D-spacing from the centre (0 cm) toward 6 cm, but the N-axis showed no significant differences between any of the distance points – including the centre point ($p > 0.05$). Invasion and metastasis is known to occur primarily along the nipple-axilla line (N-axis) where lymphatic ducts and vascular tissues are typically more concentrated and facilitate disease spread [22, 82]. Thus, changes in the collagen structure components similar to the primary site where the disease most likely originated, were expected to be seen in the N-axis because disease is more likely to be present than in the P-axis. The patients who had breast masses larger than 12 cm in diameter, where tissue samples at 6 cm from the primary site could be excised, were mastectomy patients who most likely had disease spread to the axillary lymph nodes and lymphatic ducts. Collagen axial D-spacing in the superior and inferior parts of the breast show evidence of collagen structural changes more than tissue

sampled along the nipple-axilla direction. This suggests that spread of disease in the superior and inferior parts of the breast is secondary compared to the nipple-axilla line.

The results in relation to directional dependency of the axial D-spacing must be considered with caution because the axes were not fixed. The P and N-axes were defined by the location of the lesion with respect to the nipple, rather than the same location on every patient. For example, the N(-) axis was not directly toward the axilla for 8 patients (as diagrammed in Figure 2.3) and for the remaining 122 patients, the nipple-axilla line varied by ~45º. The P-axis varied depending on the position of the N-axis. A more precise method of determining if a directional dependency exists with axial D-spacing is needed to accurately assess the trends over the mass of the breast. Patients with lesions at similar positions need to be analysed, where the positions of both the N-axis and P-axis can be determined on an absolute scale, rather than relative to the nipple (for example, the use of a bony landmark). Further investigation regarding the tissue structures and components in the superior and inferior sections of the breast may provide more information about the mechanisms related to collagen fibril degradation.

The size of the tumour will also affect the trends with distance (from 0 cm toward 6 cm away from the primary site) shown in Figure 5.1-Figure 5.3, since some patient's tumours were more than 4 or 6 cm in diameter. For these larger lesions, it cannot be assumed that the tissue samples at 2 cm, 4 cm, or 6 cm will be normal, although a trend away from the 0 cm might be expected [77]. With larger tumours, necrotic tissue is sometimes present at the centre of the lesion which would affect the SAXS pattern. Fernandez *et al*. [51] found that the intensity in the power-law regime was much higher in necrotic tissues than in healthy collagen-rich tissues and there was no evidence of organised supramolecular structures in the necrotic tissue SAXS patterns. They suggest that necrotic tissues contain molecular coils (due to the power-law exponent = -2) and the specific surface area of the coils is more than tens times larger than that of collagen fibrils, which explains the higher intensities. Necrotic tissue was not specifically examined in the work presented in this thesis. Further work is needed in order to identify the increasing or decreasing gradients of the parameters occurring within the volume of the palpable lesion as well as at the margins of known invasive disease with normal parenchyma and necrotic tissue. This could be achieved by taking more tissue samples at smaller distance intervals throughout the tumour and surrounding mass and analysing the trends in the amorphous scatter and axial D-spacing with distance more precisely.

Instead of sampling along two axes, tissue harvesting over a denser grid may show multiple disease foci and better discrimination of marginal disease. A more thorough investigation of how the parameters change throughout the breast might be achieved by taking a slice of tissue (or a cross section) and obtain SAXS parameters over the entire slice (similar to Fernandez *et al*. [49, 51] where they acquired images at 0.25 mm increments, but on a much larger scale), and then repeating this for a series of contiguous slices through an entire breast. A 3-dimensional map of how the SAXS parameters change over the breast could be created in this way. This issue is further discussed in Chapter 7 in relation to future work.

The 2D maps presented in Section 5.4 interpolated values of amorphous scatter and axial D-spacing from a moderately sparse amount of data. These were compared with histopathological diagnosis of each tissue sample. Using a threshold value to classify tissue samples into groups and comparing the normalised individual patient amorphous scatter data with histopathological diagnosis, a classification of 19% was achieved with invasive carcinoma tissues. Examining patients whose primary site tissues had confirmed invasive carcinoma disease and comparing their series tissue samples revealed a sensitivity and specificity of 55% and 80%, respectively, for the 2D maps. The specificity of 80% was close to the specificity calculated for the QDA classification models in Chapter 4, which was found to be 76.3% for the primary site comparison of invasive carcinoma and normal tissue (Table 4.6, Section 4.3). The sensitivity of 55%, however, was lower than the sensitivity calculated in Chapter 4, which was 72.5%. This shows that the 2D maps are good at identifying normal tissue (i.e. true negatives) but lack the high sensitivities required to identify invaded tissues. High specificity rates can still be useful for identifying normal tissues, particularly around surgical margins where the 2D maps can confirm that normal tissue is present. The differences seen between the sensitivity and specificity values of the amorphous scatter could also reflect the heterogeneity of the tissue samples. Many of the tissue samples diagnosed as invasive carcinoma only contained 5-10% invaded tissues, where the remaining 90-95% was normal tissue. Taking an average SAXS pattern over the entire tissue sample is a downfall when attempting to use the SAXS data for classification of tissue types. Further investigation into the amorphous scatter in relation to tissue heterogeneity effects may improve these classification rates.

Using a threshold for comparing histological diagnosis to the absolute values of the amorphous scatter for each patient has sensitivities too low for using this mapping model for diagnosis of invaded tissue. Although the amorphous scatter showed highly significant differences between invasive carcinoma and normal

tissues (as described in Chapter 4), the threshold value may need to be refined to include a ‘grey-area’ for more accurate diagnosis. If the amorphous scatter of a tissue sample falls within this ‘grey-area’, the diagnosis can be deemed inconclusive and further investigation of the tissue would be warranted. In particular, tissue samples taken in or around the tumour-healthy tissue border may have amorphous scatter values that fall within this ‘grey-area’, which would suggest changes at the supramolecular level are occurring in these regions. The 2D maps can be re-calculated using an absolute scale of the amorphous scatter using threshold ranges to reflect areas that are inconclusive. Tissue with amorphous scatter values within the ‘grey-area’ can also be used to define surgical margins, where these tissues show evidence of supramolecular changes occurring in the tissues. Comparing tissues that fall within the ‘grey-area’ with histopathological diagnosis may provide insight into the changes occurring in and around the tumour, as well as define surgical margin removal.

A limitation of the Gaussian mapping was that wrap around effects were evident when less than 6 tissue samples were available for mapping due to the periodicity assumptions of the Fourier Transform. Obviously, these wrap-around effects would not be present if tissue samples were taken over a larger grid area, rather than along two axes. Zero-padding the margins of the 2D maps increases the number of points used for calculating the map, which essentially increases the grid area so that proper calculation of the Fourier Transform is permitted. Patients who had > 6 tissue points available for their computation, or had evenly distributed tissue points around the centre point (for example, N(+2), N(-2), P(+2), and P(-2)) were needed for the model to be calculated with no computational artefacts (such as wrap-around effects). Zero-padding of the boundaries for patients with < 6 tissue samples largely eliminated the wrap around effects; however, interpreting the maps needs to be done with caution. Determination of diseased margins and parameter gradients over the 2D grid are discussed further in Appendix E.

It is important to recognise that SAXS observes smaller structures (in the tens to hundreds of nanometre range) of the tissue core compared to standard histopathology (in the micrometre range). Attempting to compare SAXS tissue indications with histopathology diagnostics may not be appropriate since different information is being compared. However, SAXS may provide complimentary information in addition to histopathology analysis, particularly for showing directional changes in tissue structures.

5.6 Conclusions

In this chapter, a significant decrease in the amorphous scatter and axial D-spacing was found at distances away from the centre of the primary lesion. The p-values comparing these two parameters between each of the centre and distant points indicated no significant differences in the distance points (2 cm, 4 cm, and 6 cm), but did show a significant difference between the centre tissue point (0 cm) and the 2 cm tissue point.

The 2D maps of the amorphous scatter and the axial D-spacing values were well correlated to the histopathology analysis of the tissues, provided that the grid is sufficiently sampled. However, even with smaller sampling numbers (for example, from WLE), the maps show changes in the amorphous scatter and axial D-spacing over the breast mass, but should be treated with caution as the smoothing process of Gaussian Kernel Estimation tends to mean they are unreliable. Using a threshold value for comparing tissue types with amorphous scatter showed a low sensitivity in defining invasive carcinoma tissues, but the threshold analysis was useful in identifying normal tissue types (high specificity). A better defined threshold, or even using threshold ranges may be more suitable for classifying tissue types based on the amorphous scatter. Evaluating tissue samples using threshold ranges may be of clinical benefit to define surgical margins, or even aid pathologists in delineating areas between the tumour/healthy tissue border.

The data taken using a synchrotron source has the ability to characterise the directional spread of supramolecular changes within the breast using SAXS, which has some ability for tissue classification. However, using a synchrotron facility is not likely to be feasible clinically. In the next chapter, analysis of tissue samples using a laboratory source is examined using the same techniques presented in Chapters 4 and 5 in order to determine if a conventional laboratory SAXS instrument can yield similar classification information.

Chapter 6

Laboratory Source Data Analysis

6.1 Introduction

In the previous chapters it was discussed how SAXS, using a synchrotron radiation source (Station 2.1, SRS Daresbury), could be used to classify breast tissue samples. Feature analysis revealed that the amorphous scatter and the axial D-spacing were the best classifiers of the four tissue groups, particularly showing differences between normal tissue from patients with breast cancer and mammoplasty tissue. If SAXS were ever implemented in a clinical setting to provide information for diagnostic or screening analysis, a laboratory SAXS apparatus capable of producing the required quality of data would be a more feasible apparatus to use rather than a synchrotron facility.

To our knowledge, only Round *et al*. [102, 103] has analysed tissue samples using SAXS from both a synchrotron radiation source and a laboratory based system (SAXSess (Panalytical) system, University of Graz, Department of Chemistry, 2005). Round *et al*. [102, 103] used a camera length of 0.267 m (laboratory based system), so no collagen axial peaks were observed on the images. Lower ordered collagen axial peaks are difficult to analyse because they are located near the beam stop (low q), which either blocks the peaks in the SAXS image or the high scattering close to the beam stop interferes with obtaining accurate peak data. Higher order peaks are less likely to be visible in tissue since very long range order is necessary

for their intensities to be visible. Since peak analysis was not possible, Round *et al.* [102, 103] found that the amorphous scatter showed differences between 18 invasive carcinoma tissue samples and 5 normal tissues (from the contralateral breast) he imaged using a laboratory X-ray source. This thesis is the first to analyse both intensity features as well as peak features from matched tissue sample data (two tissue samples taken from the same patient, discussed further in Section 6.2.1) and compare these features for classification between a laboratory source and a synchrotron source. A comparison between the laboratory source data and the synchrotron source data is presented in Section 6.5.

As described in Chapter 3, the laboratory source used for imaging the tissue samples in this thesis was the SAXS Osmic Camera (Rigaku Corporation, Texas, USA). The laboratory source has a more divergent beam compared to the synchrotron source, which decreases the resolution of the scattering. The sample-to-detector distance of the laboratory source is limited by this lower divergence and hence is shorter (~1.6 m) than that used with the synchrotron source (~6.3 m). The q-ranges seen with a sample-to-detector distances of ~1.6 m and ~6.3 m are 0.25-2.3 nm^{-1} and 0.1-0.6 nm^{-1}, respectively. Thus, the features that could be seen with the laboratory source were different to those visible with the synchrotron source. The synchrotron images contained up to the 6th order axial collagen peak (q = 0.564), whereas the laboratory images could reveal up to the 12th order axial collagen peak (q = 1.129 nm^{-1}). In particular, the higher maximum *q* permits visualisation of the diffraction components related to adipose tissue (which is referred to as “fat” in this thesis) in the samples. Examination of the fat peak, which forms a complete ring on the laboratory source images, is presented in Section 6.3.3, where the peak features relating to the amplitude, Full Width at Half Maximum (FWHM), centre (or axial D-spacing), and peak area were analysed. The third axial peak was visible with the laboratory source images (as well as with the synchrotron images), but scattering near the beam stop interfered with accurate extraction of this peak feature with the laboratory source. The fifth order axial collagen peak features were studied (Section 6.3.2) because this peak was visible on both the synchrotron images and the laboratory images. In order to compare the amorphous scatter data between the laboratory source and the synchrotron source, the amorphous scatter was examined over several q-ranges (Section 6.3.1 and Section 6.4) with the laboratory source data. The classification potential of these features was examined in a similar manner as described in Chapter 4.

Table 6.1: Summary of the tissue samples imaged with the laboratory SAXS camera from 66 patients. Also see Table 2.3.

	Invasive Carcinoma	**Benign**	**Normal**	**Mammoplasty**	**Total # Tissue Samples**
Total # of Patients	21	16	18	11	66
Amorphous scatter present	21	16	18	11	66
5th axial peak present	12	10	10	3	35
Fat peak present	14	6	18	7	45
Amorphous, 5th peak, and fat peak present	5	4	9	2	20
Tissue samples histopathologically matched diagnosis with synchrotron and lab source	6	6	5	4	21

6.2 Methods and Materials

6.2.1 Tissue Samples

A total of 66 patients were examined with the laboratory source: 21 invasive carcinoma, 16 benign, 18 normal, and 11 mammoplasty. The SAXS features examined from these tissue samples are described in Section 6.3, where the features relating to the fat peak, fifth order axial collagen peak, and amorphous scatter were examined. However, the fat peak and fifth order axial collagen peak were not visible in all of the tissue images because some tissue samples (presumably) did not contain sufficient fat and/or collagen to give rise to these features. Table 6.1 lists the number of tissue samples and the features that were visible using the laboratory source for the 66 patients studied (primary sites only).

The 66 patients who had tissue samples imaged with the laboratory SAXS camera were a subset of the 130 patients who had tissue samples imaged with the synchrotron source. Due to time restrictions, the remaining 64 patients could not be imaged with the laboratory source. For the 66 patients that were imaged with the laboratory source, two tissue samples were harvested from each sample site, where the second tissue sample from the same patient was imaged with the synchrotron source. Each tissue sample was taken using a biopsy gun (see Section 2.2), where the pair of tissues were harvested within 2 mm of each other. We expected that each pair of tissue samples would be diagnosed as the same tissue type (i.e. the

synchrotron sample diagnosed the same as the laboratory sample), since both tissues were excised close together. Although these biopsies were in close proximity, the tissue around the centre of the lesion, where the biopsies were taken, is typically heterogenous, so not all of the 66 pairs of tissue samples were diagnosed equivalently. Table 6.2 lists the tissue samples of the 66 patients imaged with both X-ray sources. The 66 patients who had their (paired) tissue sample imaged with the synchrotron consisted of 19 invasive carcinoma, 13 benign, 25 normal, and 9 mammoplasty. Out of the 66 patients imaged with both X-ray sources, 21 patients had the same histopathology diagnosis for their pair of tissues: 6 invasive carcinoma, 6 benign, 5 normal, and 4 mammoplasty. There were 7 patients who underwent mammoplasty surgery, but their tissue samples were diagnosed by the pathologist as normal[†††††] or benign. Analysis and classification was also performed with these 21 patient's samples in order to compare the feature information that can be obtained from both X-ray sources.

Of the 66 patients who had primary site tissue samples imaged with the laboratory source, 4 patients had their series tissue samples analysed. The four patients had at total of 46 tissue samples taken radially outward from their primary site imaged with the laboratory source: 4 invasive carcinoma, 4 benign, and 38 normal. The primary site tissue points of the 4 patients who had series tissue samples examined consisted of 3 primary sites diagnosed as invasive carcinoma, and 1 primary site diagnosed as benign. The features found in Section 6.3 that showed the most separation between tissue types were examined in Section 6.6 for the series tissue samples.

Table 6.2: The number of tissue samples and their diagnosis for the 66 patients who had tissue imaged with both the synchrotron source and the laboratory source.

Tissue Type	**# Tissues imaged with Laboratory Source**	**# Tissues imaged with Synchrotron Source**	**# Tissue pairs that had the same histopathological diagnosis from the same patient**
Invasive Carcinoma	21	20	6
Benign	16	13	6
Normal	18	29	5
Mammoplasty	11	4	4

[†††††] The pathologist diagnosed 7 of the synchrotron mammoplasty tissues as 'normal' because there was evidence of fibrocystic change in these tissue samples, which could not be deemed as mammoplasty type tissues.

6.2.2 Methods

The experimental set-up and laboratory SAXS camera components were described in Chapter 3. A camera length of ~1.6 m was used, which produced SAXS images between the q-range of 0.25-2.4 nm^{-1} using IP detectors. The IPs were scanned using a Fuji Film BAS-5000 IP Reader, which produced a digitized SAXS image that was further processed using PCDetPak32 (Version 6.0, R. Lewis, Monash University, 2006). Calibration of the images for q-spacing (using silver behenate) and intensity (using Glassy Carbon) was performed, as well as radial integration of the 2D images to obtain 1D data over q-space (as previously described in Section 3.5). From these 1D profiles, the amorphous scatter, 5th order axial collagen d-spacing, and fat peak features were extracted for analysis. The peak amplitudes, peak full-widths-at-half-maximum (FWHM), peak centres (d-spacing), and peak areas were extracted from those patients who had the fat peak and the 5th order axial collagen peak visible on their SAXS image.

As discussed in Chapter 4, the amorphous scatter showed the most separation between tissue groups for the synchrotron data and was the best SAXS feature for use in classification. For this reason, the amorphous scatter for the laboratory source images was divided into three q-ranges in order to fully examine the classification potential of this SAXS feature. Figure 6.1 indicates the three ranges of the amorphous scatter that were examined with the laboratory images, showing where the fat peak was typically located, The range used with the synchrotron source (q = 0.1-0.6 nm^{-1}) is also shown in Figure 6.1 for comparison. Range 1 was defined as q = 0.25-0.6 nm^{-1}, Range 2 between q = 0.6-1.3 nm^{-1}, and Range 3 between q = 1.7-2.3 nm^{-1}. The range between q = 1.3-1.7 nm^{-1} contained the fat peak so the amorphous scatter in this range was not calculated. The total amorphous scatter (between the q-range 0.25-2.3 nm^{-1}) was also included in the analysis. Range 1 (q = 0.25-0.6 nm^{-1}) was chosen because this was the similar q-range examined with the synchrotron data (q = 0.1-0.6 nm^{-1}). The q-range between q = 0.1-0.25 nm^{-1} could not be reliably analysed with the laboratory source due to parasitic scattering around the beam stop. The amorphous scatter in Range 1 of the laboratory source data and the amorphous scatter calculated with the synchrotron data are directly compared in Section 6.5.

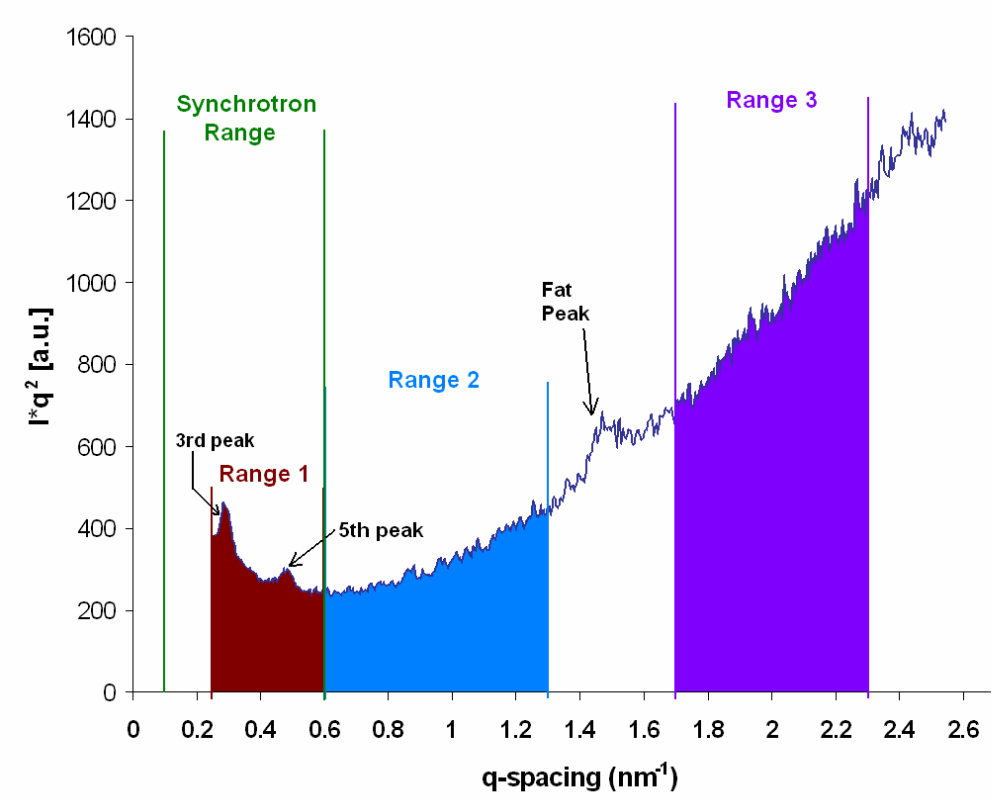


Figure 6.1: Typical 1D profile extracted from a laboratory source image (normal tissue sample, with not much fat). The three q-ranges that were used to examine the amorphous scatter are indicated, with the fat peak shown and 5th order axial collagen peak location. Range 1 is between q = 0.25-0.6 nm^{-1}, Range 2 is between q = 0.6-1.3 nm^{-1}, and Range 3 is between q = 1.7-2.3 nm^{-1}. The synchrotron range is indicated between q = 0.1-0.6 nm^{-1}.

6.3 Feature Analysis

6.3.1 Amorphous Scatter

As described in Chapter 3, a quadratic function (Equation 3.1) was fitted to the 1-D profile SAXS data for each patient after the peak features were subtracted from the profiles to calculate the total amorphous scatter. The coefficients of Equation 3.1 (*A*, *B*, and *C*), as well as the total area of the fitted quadratic function (amorphous scatter) between q = 0.25-2.3 nm^{-1} were calculated for each tissue sample. The peak fraction of total diffraction (total area in the fitted peaks of the collagen features divided by the total area of all the scattered intensity) and the amorphous scatter within three q-ranges (as described in Section 6.2.2 and shown in Figure 6.1) were also calculated and compared.

The results of Principal Component Analysis (PCA) (see Appendix E for details of statistical analysis) showed that the variable *C* accounted for all of the variation in the data and, can thus be used to describe variability in the entire data set. This was consistent with the results in Chapter 4 with the synchrotron data, which also showed that the variable *C* accounted for all of the variation in the data. However, Kolmogorov-Smirnov (KS) tests (see Appendix E) performed on the intensity features showed that PC1 did not distinguish between tissue groups ($p > 0.05$) and could not be used as a classifier. There was, however, evidence that the variance between tissue groups showed a significant difference when a Bartlett's Test (see Appendix E) was performed: $p = 2.15 \times 10^{-8}$. This means that the absolute values of the parameter could not be used to classify the tissue groups, but the variance in the parameter suggests that significant differences do exist between the tissue types.

6.3.2 Fifth Order Axial Collagen Peak

The fifth order intensity maxima for the collagen axial D-spacing diffraction peak was visible in 35/66 patients (see Table 6.1) and the peak features (amplitude, FWHM, centre, area) were extracted from their laboratory SAXS images and analysed using PCA.

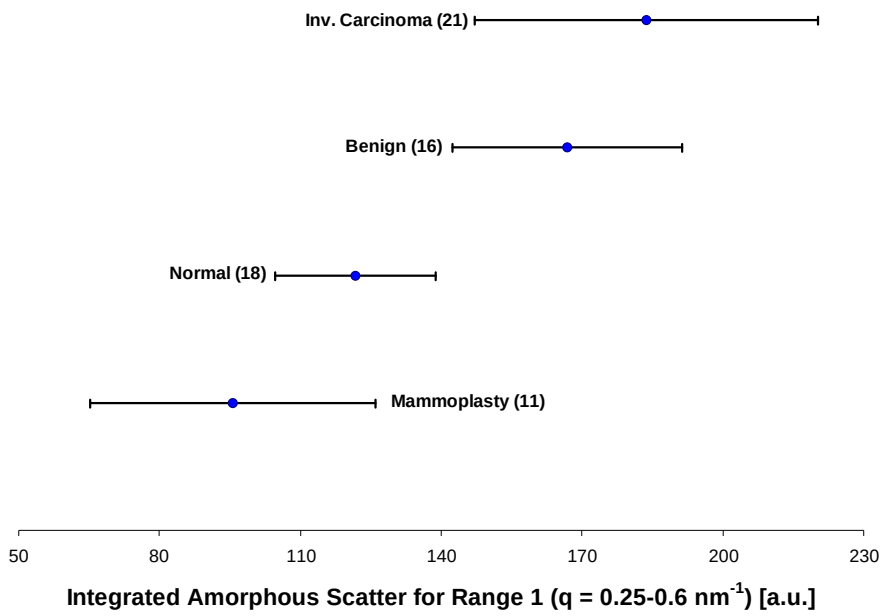


Figure 6.2: Average amorphous scatter values calculated in Range 1 for the 66 patients imaged with the laboratory SAXS camera. The number of tissue samples imaged in each tissue group is shown in brackets. Error bars represent 95% confidence intervals.

Table 6.3: Comparisons of p-values for the amorphous scatter in each q-range for the 66 primary site tissue samples imaged with the laboratory source.

Tissue Groups	Total (0.25-2.3 nm^{-1})	Range 1 (0.25-0.6 nm^{-1})	Range 2 (0.6-1.3 nm^{-1})	Range 3 (1.7-2.3 nm^{-1})
Inv. Carcinoma & Mammoplasty	0.031*	0.0042**	0.0023**	0.040*
Benign & Mammoplasty	0.083*	0.0025**	0.0021**	0.11
Benign & Normal	0.14	0.016*	0.0084**	0.087
Normal & Mammoplasty	0.21	0.086	0.14	0.55
Inv. Carcinoma & Normal	0.031*	0.0089**	0.0040**	0.019*
Inv. Carcinoma & Benign	0.36	0.32	0.40	0.45

* Significantly significant $p < 0.05$
**Significantly significant $p < 0.01$

The use of PCA revealed that the amplitude of the fifth order axial peak was the most significant variable out of all of the peak variables, which accounted for 99% of the variability in the data. In order to determine if PC1 was correlated with tissue diagnosis, a KS test was performed and the p-values showed no evidence that the tissue groups were significantly different ($p > 0.05$). Thus, the amplitude of the fifth order axial collagen peak was not a good classifier of disease. The variances were also calculated and a Bartlett's Test confirmed that there were no significant differences in the variances between the tissue groups ($p = 0.40$).

6.3.3 Fat (Adipose) Tissue

Intensity maxima due to the scatter from fat within the tissue sample were visible in 45/66 patients (see Table 6.1). The fat peak amplitude, peak full-width-half-maximum (FWHM), peak centre (d-spacing), and peak area were calculated and analysed for the four tissue groups.

The fat peak features were analysed using PCA which showed that 99% of the data related to the fat peak could be explained by the first Principal Component (PC1), equivalent to the amplitude of the fat peak. Although PC1 accounted for 99% of the fat peak data, a KS test revealed that this component (the amplitude of the fat peak) did not vary significantly between the tissue groups. Therefore, the amplitude of the fat peak was not a useful classification parameter by itself, but rather, showed

that the variation in the values of the amplitude was significantly different between the tissue groups, which was further confirmed with a Bartlett's test which showed $p = 6.5 \times 10^{-12}$ for the variance between the tissue groups.

However, the peak FWHM showed a general trend between groups as shown in Figure 6.3. A significant difference in the FWHM of invasive carcinoma compared to mammoplasty tissues was observed ($p = 0.017$), but none of the other tissue groupings showed distinctions ($p > 0.05$) between each other. It appears from Figure 6.3 that there may be a significant difference between benign and mammoplasty tissue types, however, due to the small number of samples in each of these tissue groups, no statistical difference was found ($p = 0.092$). Although PCA did not show that the FWHM of the fat peaks was able to separate the tissue groups (except between invasive carcinoma and mammoplasty), inspection of Figure 6.3 suggests that the FWHM of the fat peak decreases with severity of disease status (tends to show a decrease).

In order to compare our results with Fernandez *et al*. [51], the amorphous scatter for Range 1 was compared for only the 45/66 tissue samples that had the fat peak present on their SAXS images. Only the amorphous scatter in Range 1 showed significant differences between invasive carcinoma and normal tissues, and invasive carcinoma and mammoplasty tissues ($p = 0.012$ and $p = 0.020$, respectively). All of the other tissue comparisons showed no significant differences with the amorphous scatter in Range 1 ($p > 0.05$). Our results agreed with Fernandez *et al*. [51], where our data clearly showed that invaded tissues that contain fat have a higher amorphous scatter than healthy tissues that contain fat.

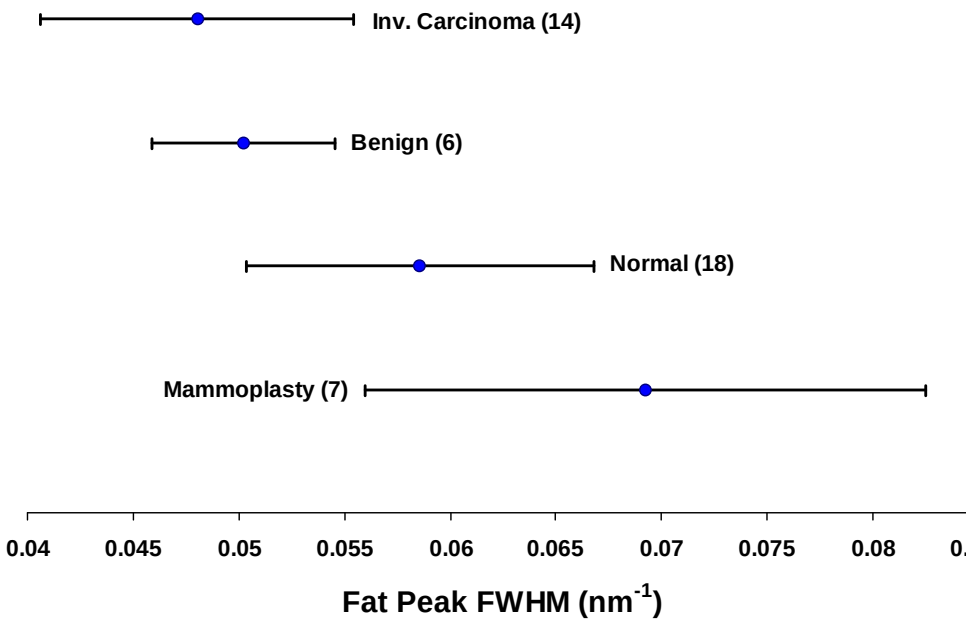


Figure 6.3: Averages of the FWHM of the fat peak calculated using the laboratory source for the four tissue groups. The number of patients that had the fat peak visible on their SAXS images in each tissue group is shown in brackets (total number of patients with fat peak visible = 45). Error bars represent the 95% confidence intervals.

6.4 Classification of Tissue Samples

6.4.1 Probability Models and ROC curves

From the analysis presented in Section 6.3, the amorphous scatter in Range 1 and the FWHM of the fat peak showed the most significant difference between the tissue groups. A plot (normalised to the range of the data) of these two parameters is shown in Figure 6.4. On first inspection, there appears to be no separation of tissue groups when using these two variables. The invasive carcinoma tissues appear to cluster into two groups. The grades and ages were compared as a function of these two invasive carcinoma clusters and there was no correlation. Hormone therapy, radiotherapy, and prescription drugs were also compared and no correlation was found.

Figure 6.5 shows the tissue pair comparisons from Quadratic Discriminate Analysis (QDA) probability models (see Appendix E) using the two most differentiating variables, as also used in Figure 6.4. The QDA calculates the

probability over the more severe disease states in relation to the less severe disease states. For example, in Figure 6.5a, a probability of 0.8 means that there was an 80% probability of a tissue sample being invasive carcinoma (more severe state) when compared to benign tissue (less severe state).

The ability of the QDA model using the amorphous scatter in Range 1 and the FWHM of the fat peak to classify tissue samples was quantitatively assessed using Receiver Operator Characteristic (ROC) curves (see Appendix E). Table 6.4 shows the values obtained from the ROC curves calculated for the tissue type pair comparisons shown in Figure 6.5. The tissue comparisons between invasive carcinoma versus mammoplasty, benign versus mammoplasty, and normal versus mammoplasty (first three rows of Table 6.4, which had valid calculations) showed good sensitivities (above 80%) and, on average, good specificities, indicating that the laboratory camera has potential in providing good classification between these tissue groups. Also, the QDA model was able to discriminate between normal and mammoplasty tissues (sensitivity = 81.3%, specificity = 66.7%).

However, the small number of tissue samples in each group, as well as the points being very closely clustered together, made calculating the values for the tissue pair comparisons in the last three rows of Table 6.4 invalid for classification (invasive carcinoma vs. benign, invasive carcinoma vs. normal, and benign vs. normal). This means that there were no differences, using the amorphous scatter and FWHM of the fat peak for the model, between these tissue pair groupings that could be used for classification. Failure of the model to provide any useful distinction between certain tissue types limits the model's ability to provide useful diagnostic information, which is important in determining treatment modalities for patients. With higher sampling numbers, the three tissue comparisons that did not have meaningful ROC curve information (the last three rows in Table 6.4) may show valid results. Further work is needed to complete the ROC curve data in Table 6.4 by analysing a larger number of tissue samples (discussed further in Chapter 7), which may prove to be difficult, since not all tissue samples have the fat ring present on their SAXS image.

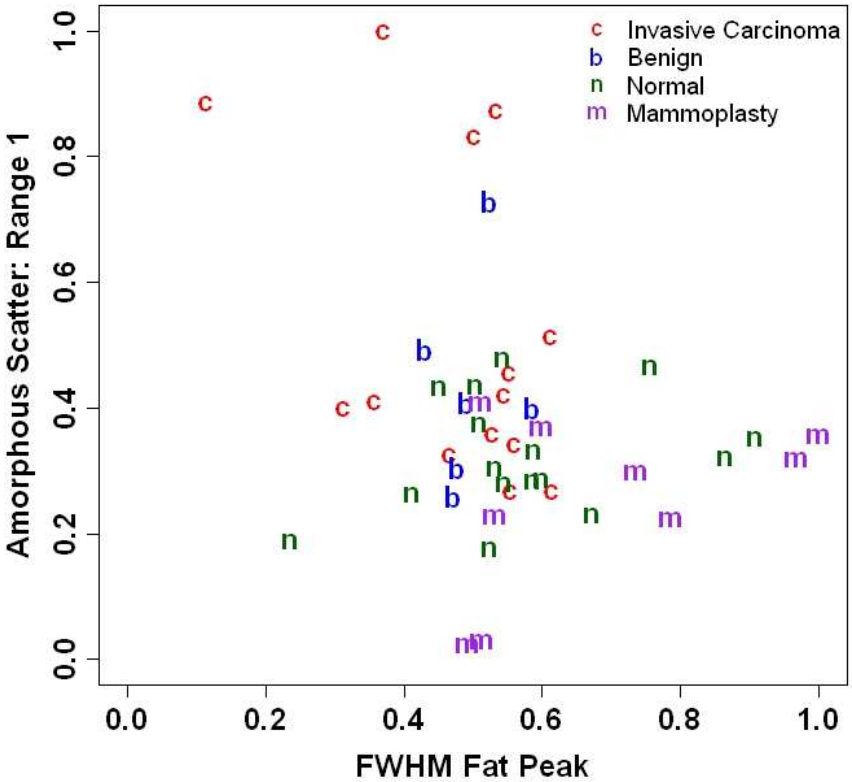


Figure 6.4: The amorphous scatter for Range 1 with the FWHM of the fat peak for the primary site tissue samples imaged with the laboratory source. Values have been normalised.

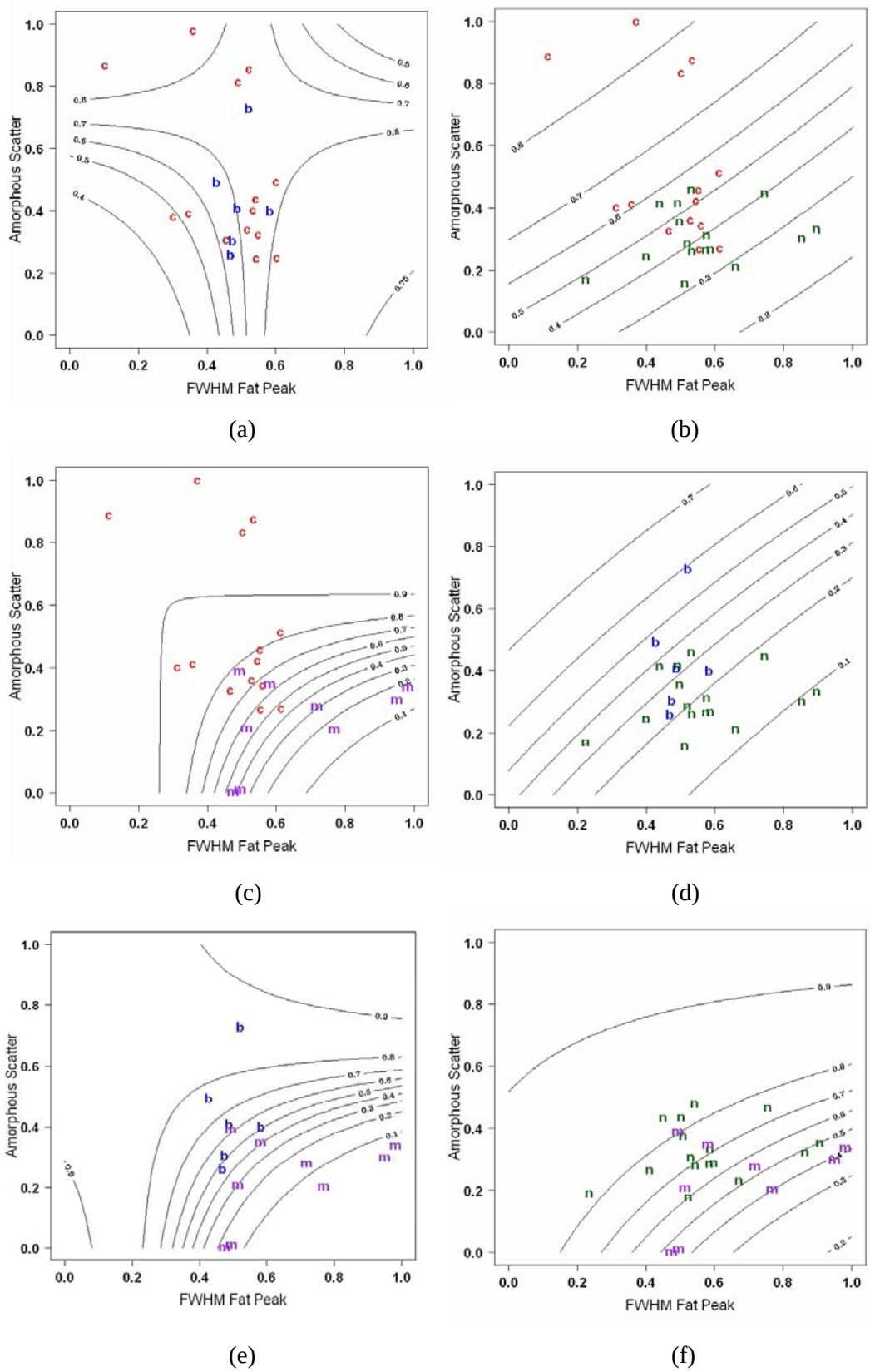


Figure 6.5: QDA probability plots of the data shown in Figure 6.4 using the amorphous scatter for Range 1 and the FWHM of the fat peak. The tissue groups are identified as c = invasive carcinoma, b = benign, n = normal, and m = mammoplasty. (a) Probability of Invasive Carcinoma with Benign (b) Probability of Inv. Carcinoma with Normal (c) Probability of Invasive Carcinoma with Mammoplasty (d) Probability of Normal with Benign (e) Probability of Mammoplasty with Benign (f) Probability of Normal with Mammoplasty.

Table 6.4: ROC curve characteristics for each pairing of tissue groups with the laboratory SAXS camera. The first listed tissue group was considered "positive" (e.g., invasive carcinoma) and the second tissue group considered "negative" (e.g., versus benign). Only the mammoplasty comparisons in the first three rows provided valid information, whereas the other tissue comparisons (last three rows) did not.

Tissue Comparisons	Sensitivity	Specificity	Accuracy	AUC (empiric)
Inv. Carcinoma & Mammoplasty	85.7	77.8	82.6	0.93
Benign & Mammoplasty	83.3	88.9	86.7	0.93
Normal & Mammoplasty	81.3	66.7	76.0	0.77
Inv. Carcinoma & Benign	84.6	0.0	57.9	0.65
Inv. Carcinoma & Normal	42.9	100	73.3	0.81
Benign & Normal	0.0	100	68.2	0.79

6.4.2 Discriminate Variables

In order to determine if there were any other linear combinations of parameters that could be used to classify the tissue groups, a discriminate analysis was performed on those samples that had all three features present (20/66) in their SAXS image (see Table 6.1): amorphous scatter, fat peak, and fifth order axial peak. A total of 16 parameters were used to describe the three features:

A) Amorphous Scatter (as described in Section 6.3.1):

Total , Range 1, Range 2, Range 3, the quadratic coefficients (A,B,C from Equation 3.1), peak fraction of total diffraction

B) Fifth order axial collagen peak (as discussed in Section 6.3.2):

Peak amplitude, peak FWHM, peak centre, and peak area

C) Fat peak (as discussed in Section 6.3.3):

Peak amplitude, peak FWHM, peak centre, and peak area

These 16 parameters were used to calculate the first and second order discriminate variables which are plotted in Figure 6.6. Removing one or more parameters to calculate the discriminate variables (i.e. with less than 16 parameters) proved to be ineffective at providing any apparent tissue group clustering. Thus, only the 20 tissue samples that had all 16 parameters available on their SAXS images could be used for this analysis. This suggested that all 16 parameters from all three features were needed to classify the four tissue groups. Nonetheless, it appears that the 1st

discriminant variable (d1) is sufficient to provide discrimination between tissue groups, whereas the 2nd discrminant variable (d2) only serves to further separate mammoplasty tissues from the rest of the tissue types.

Using all 16 parameters is not a practical approach to classifying tissue for diagnostic purposes because only 30% of the patients had sufficient information available for the model. A patient eligibility of only 30% for use of this model for classification is clearly not useful as a diagnostic tool.

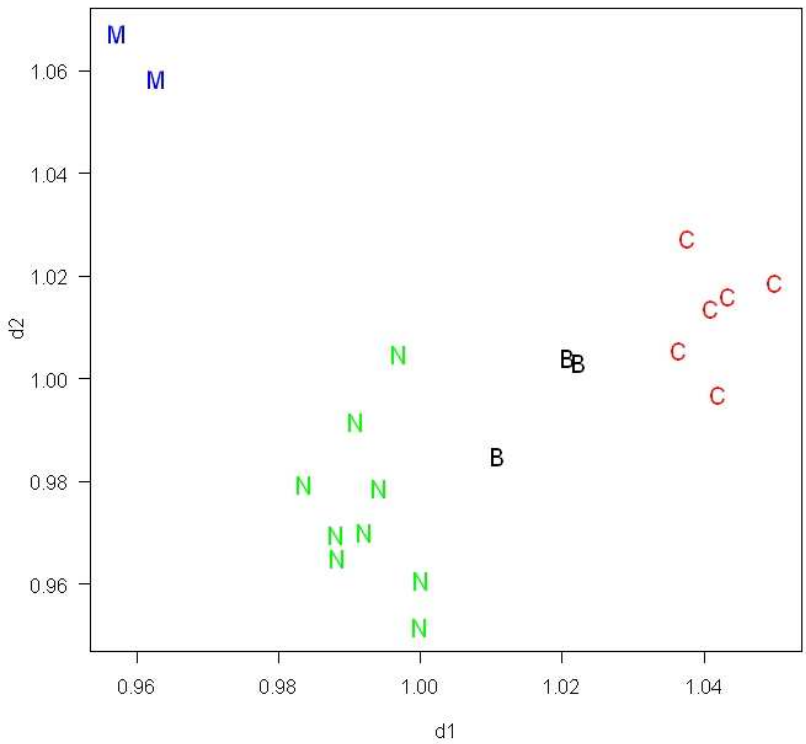


Figure 6.6: Discriminate variables 1 and 2 of the 20 tissue samples that had all 16 possible parameters present in their SAXS images. M represents mammoplasty tissues, N normal tissues, B benign tissues, and C represents invasive carcinoma tissues.

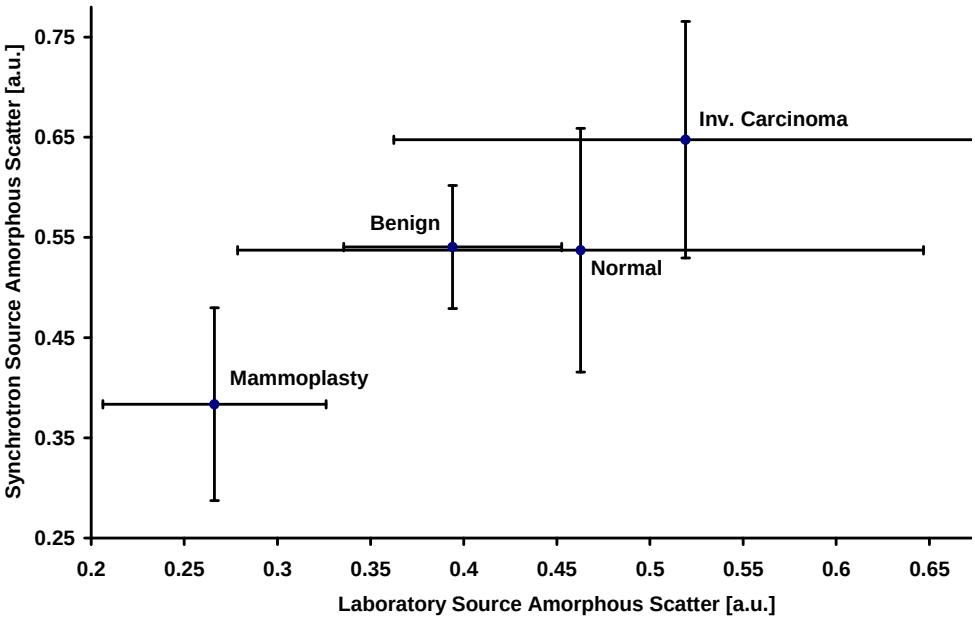


Figure 6.7: Comparison of equivalent amorphous scatter ranges (q = 0.25-0.6 nm^{-1}) for the 21 matched patients between the SRS Darebury data and the laboratory SAXS camera data. The data was normalised for each X-ray source, then plotted. Error bars represent 95% confidence intervals.

Table 6.5: Comparison of the p-values for the amorphous scatter ranges between each tissue group for the SRS Daresbury (130 patients, and 66 matched patients) and the laboratory SAXS camera (66 patients).

	Laboratory 66 patients q = 0.25-0.6 nm^{-1} (Range 1)	**Synchrotron 130 patients q = 0.1-0.25 nm^{-1}**	**Synchrotron 66 patients q = 0.25-0.6 nm^{-1} (Range 1)**	**Synchrotron 130 patients q = 0.1-0.6 nm^{-1}**
Inv. Carcinoma & Mammoplasty	0.0042**	0.030*	0.0039**	4.1×10^{-7}**
Benign & Mammoplasty	0.0025**	0.0019**	0.0080**	0.0012**
Benign & Normal	0.016*	0.57	0.54	0.30
Normal & Mammoplasty	0.086	0.011*	0.040*	0.0025**
Inv. Carcinoma & Normal	0.0089**	0.098	0.088	0.0023**
Inv. Carcinoma & Benign	0.32	0.17	0.13	0.0072**

* Statistically significant $p < 0.05$
** Statistically significant $p < 0.01$

6.5 Comparing Amorphous Scatter from Laboratory Source Data with Amorphous Scatter from Synchrotron Source Data

The images obtained from the two X-ray sources for each sample pair were compared when patients had matching histopathological diagnosis for both sources: 6 invasive carcinoma, 6 benign, 5 normal, and 4 mammoplasty tissue pairs from 21 patients. The amorphous scatter for the 21 patients imaged with both X-ray sources were calculated for Range 1: q = 0.25 – 0.6 nm^{-1} and are directly compared in Figure 6.7. In Figure 6.7, it appears that the mammoplasty tissues are clearly separated from the other three tissue groups. However, closer inspection of the 95%CI of the mammoplasty and normal tissues reveals that the intervals for both tissue groups overlap, for both the synchrotron and laboratory source data. Thus, the graph in Figure 6.7 must be considered carefully. Nonetheless, the amorphous scatter in Range 1 for the laboratory source, showed the same increasing trend with severity of disease as the data for the synchrotron source (i.e. the amorphous scatter increased from mammoplasty → normal → benign → invasive carcinoma for both X-ray sources). This suggests that even with a small sample size (21 tissue samples) the amorphous scatter can show evidence of characterising differences between tissue groups.

The p-values for each of the tissue type pairs, including those that did not have matching histopathology, are presented in Table 6.5, where the amorphous scatter in several q-ranges was calculated for the 66 laboratory source tissue samples, the 66 paired tissue samples imaged with the synchrotron, and the 130 tissue samples imaged with the synchrotron. The lower end of the q-range could not be calculated for the laboratory source (from q = 0.1-0.25 nm^{-1}) because of the presence of the beam stop, but was calculated for the 130 tissue samples imaged with the synchrotron source in order to determine if the amorphous scatter within this low q-range may affect tissue group separation. The amorphous scatter calculated for q = 0.1-0.6 nm^{-1} for the 130 tissue samples imaged with the synchrotron are also shown (the same results were presented previously in Table 4.5).

For all of the q-ranges examined for both X-ray sources, the amorphous scatter clearly differentiates between invasive carcinoma and mammoplasty tissues, and between benign and mammoplasty tissues ($p < 0.05$). The amorphous scatter differences between each of these tissue group comparisons are strong (i.e. almost all of the p-values are less than 0.01), and both X-ray sources can reveal these

differences. In particular, the benign and mammoplasty tissue comparisons in Table 6.5 and Table 6.3, show significant differences in the amorphous scatter in the low q-range data ($q < 1.3$ nm^{-1}) , but not at the higher q-values in Range 3 ($p = 0.11$). This suggests that differences in these two tissue groups lie in the amount of large sized scatterers in these tissues (i.e. $q < 1.3$ nm^{-1}, $d > 4.83$ nm).

An interesting comparison of the two X-ray sources is revealed by the amorphous scatter differences for the benign versus normal, and the normal versus mammoplasty tissue types. According to the p-values in Table 6.5, the amorphous scatter calculated from the synchrotron data cannot differentiate between normal and benign tissue types. However, the laboratory source can see a significant difference between normal and benign tissues with the amorphous scatter in Range 1 ($p = 0.016$), which is a subsection of the synchrotron range. From Table 6.3, the amorphous scatter in Range 2 also showed a significant difference between normal and benign tissue types ($p = 0.0084$), and to a lesser degree in Range 1 ($p = 0.016$). Furthermore, the amorphous scatter in Range 1 for benign versus normal tissue types between the two X-ray sources (1st and 3rd column of values in Table 6.5) are contradictory: the laboratory source can statistically differentiate the two tissue types, but the synchrotron source could not with the 66 patients and same q-range. This discrepancy is most likely due to patient variability as well as tissue diagnosis, i.e. the tissue samples imaged with both X-ray sources were matched with patient, but the diagnosis of the tissue pairs was not consistent. This suggests that the amorphous scatter in Range 2 ($q = 0.6$-1.3 nm^{-1}) can characterise the differences between normal and benign tissue types, which corresponds to scatterers of size $d = 4.83$-10.47 nm.

The opposite was true for the normal and mammoplasty tissue comparisons, where the synchrotron data showed significant separation in the amorphous scatter between these two tissue groups, but the laboratory source cannot differentiate normal and mammoplasty tissue types. The highest significance in the amorphous scatter between normal and mammoplasty tissues was for the 130 patients imaged with the synchrotron data for the range $q = 0.1$-0.6 nm^{-1}. However, the amorphous scatter for the range $q = 0.1$-0.25 nm^{-1} had a higher significant difference compared to the amorphous scatter in the range $q = 0.25$-0.6 nm^{-1} ($p = 0.011$ versus $p = 0.040$, respectively). This suggests that the amorphous scatter at this low q-range ($q = 0.1$-0.25 nm^{-1}, $d = 25.13$-62.83 nm) may be important in distinguishing between these two tissue groups.

The inability of the amorphous scatter values from the laboratory source to distinguish normal and mammoplasty tissues could be due to sample size variations,

where the number of normal and mammoplasty tissues for the 66 patients imaged with both X-ray sources was markedly different (only 9 pairs of tissue samples had the same diagnosis of normal and mammoplasty groups, see Table 6.2). The 7 patients whose mammoplasty tissues imaged with the synchrotron contained fibrocystic changes (thus were diagnosed as 'normal' for the synchrotron sample population) that were not present in the paired and tissue samples imaged with the laboratory source. These differences may have contributed to the discrepancy in the p-values seen between the normal and mammoplasty tissues in Range 1 (see Table 6.5). This suggests that fibrocystic changes occurring in the breast contribute to the ability of the amorphous scatter to classify the differences between normal and mammoplasty tissue types.

The differences in amorphous scattering between invasive carcinoma versus normal tissues as well between invasive carcinoma versus benign tissues are inconsistent for the two X-ray sources. With invasive carcinoma versus normal tissues, significant differences were seen with the laboratory source for all q-ranges examined, but the synchrotron source showed a difference with only the 130 tissue samples for the q-range q = 0.1-0.6 nm^{-1}. As shown in Table 6.3, the laboratory source was able to differentiate between invasive carcinoma tissue and normal tissues over all of the q-ranges examined, including the total from q = 0.25-2.3 nm^{-1}. The laboratory source data has $p < 0.01$ for the amorphous scatter in Range 1, but the same range for the synchrotron source data has $p > 0.05$. In addition, the synchrotron source data has $p < 0.01$ for the range q = 0.1-0.6 nm^{-1} (d = 10.47-62.83 nm), which includes Range 1 (q = 0.25-0.6 nm^{-1}, d = 10.47-25.13 nm), but $p > 0.05$ for both of the two sub-ranges: q = 0.1-0.25 nm^{-1} and q = 0.25-0.6 nm^{-1}. These results are obviously contradictory and not easily explained except for the possibility that the invasive carcinoma versus normal tissue groups are likely to form a continuum rather than two clearly separated groups (i.e. individual samples can lie anywhere on a spectrum of supramolecular change between the extremes of normal and invasive carcinoma).

6.5.1 Summary of Amorphous Scatter: Laboratory Source versus Synchrotron Source

The amorphous scattering results for the two X-ray sources can be combined and examined for specific differences between tissue groups. Figure 6.8 shows the tissue pair comparisons and their significant q-ranges for both X-ray sources.

1) Invasive carcinoma can be distinguished from mammoplasty tissues with high significance using the entire scattering range q = 0.1-2.3 nm^{-1} (d = 2.73-62.83 nm).
2) Benign tissues can be distinguished from mammoplasty tissues with high significance using the range q = 0.1-1.3 nm^{-1} arising from scatterers of the size d = 4.83-62.83 nm.
3) The distinguishing feature for benign versus normal tissue groups is the scattering in the range q = 0.6-1.3 nm^{-1} (d = 4.83-10.47 nm).
4) The distinguishing feature for normal versus mammoplasty tissue groups is the amorphous scattering in the range q = 0.1-0.25 nm^{-1}, arising from large scatterers of size d = 25.13-62.83 nm.
5) The results for invasive carcinoma versus normal and invasive carcinoma versus benign tissue are contradictory between the two X-ray sources and require further investigation.

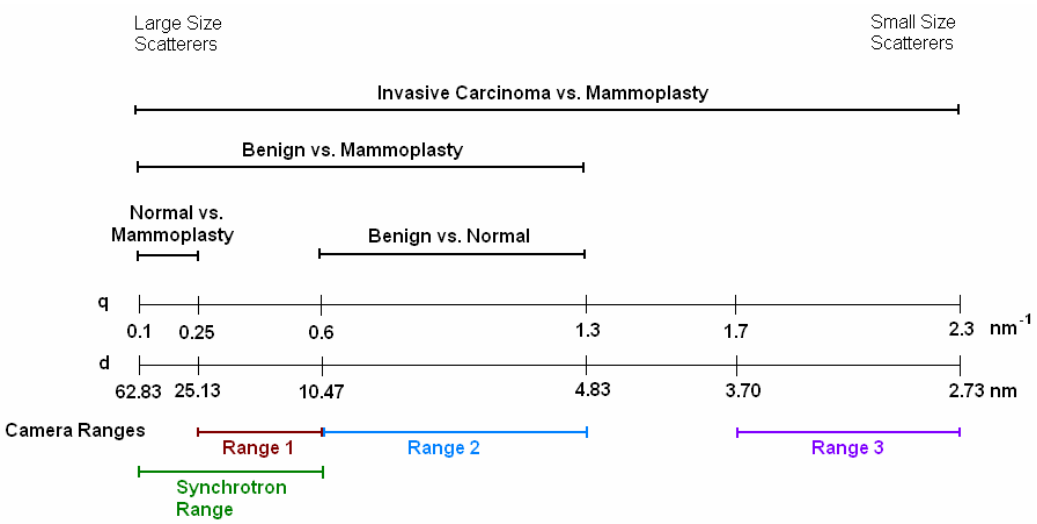


Figure 6.8: Summary of the amorphous scatter ranges that showed significant differences between the tissue group pairs for the laboratory source and the synchrotron source. Invasive carcinoma vs. normal and invasive carcinoma vs. benign were inconsistent.

6.6 Series Tissue Analysis: 4 mastectomy patients

Four mastectomy patients had full series tissues extracted and imaged with the laboratory source. For each patient, primary site tissue samples and up to 12 tissue

samples taken radially outward along two perpendicular axes (in the same manner to that described in Section 2.2) were analysed. Patient 1 had insufficient tissue at their N(-2) point to yield a good quality SAXS image for analysis; nor did Patient 4 at their N(-6) point. Therefore, only 46 (out of a possible 48) series tissue samples were imaged, along with 4 tissue points extracted from the primary site (total = 50).

The fat peak was visible for all 4 primary site tissue samples for the four patients studied. A t-test (see Appendix E) showed a significant difference in the FWHM of the fat peak between the 4 primary site tissue samples and all of the distance tissues grouped together ($p = 0.041$) (i.e. between the 0 cm points and all of the tissues at 2 cm, 4 cm, and 6 cm). This trend was similar to those presented in Chapter 5 with the analysis of the amorphous scatter and axial D-spacing of the series tissue samples imaged with the synchrotron source, where a significant difference was seen between the 0 cm point and the rest of the distance tissues, but not between the distance tissues. The fat peak was only visible for 34 (out of 48) laboratory series tissue samples. Figure 6.9 shows the FWHM of the fat peak as a function of distance from the primary site for the 4 patients studied. A t-test showed no significant difference between the distance tissues for the FWHM of the fat peak of these four patients ($p > 0.05$). This suggests that the distribution of the fat tissue at the centre of the palpable mass was more ordered than the fat tissue more than 2 cm away from the centre of the palpable mass.

Figure 6.10 shows the average values and 95%CI for Range 1 as a function of distance from the primary site for the four patients. The total amorphous scatter ($q = 0.25\text{-}2.3\ nm^{-1}$) and Range 2 ($q = 0.6\text{-}1.3\ nm^{-1}$) showed similar results. There was no directional dependency along either axis or a general trend in the amorphous scatter values with distance away from the centre of the palpable mass, which was most likely due to the small number of tissue samples available for statistical analysis.

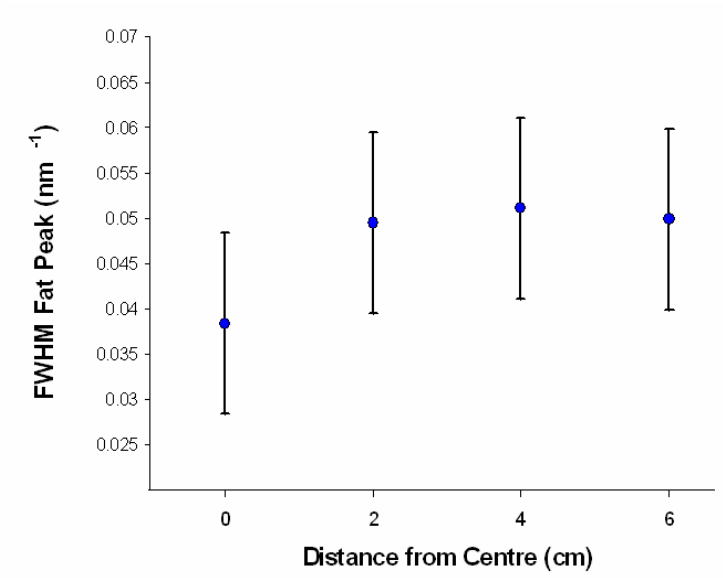


Figure 6.9: Average values of the FWHM of the fat peak for the four patients studied with distance away from the centre of the palpable lesion. Error bars represent 95% confidence intervals.

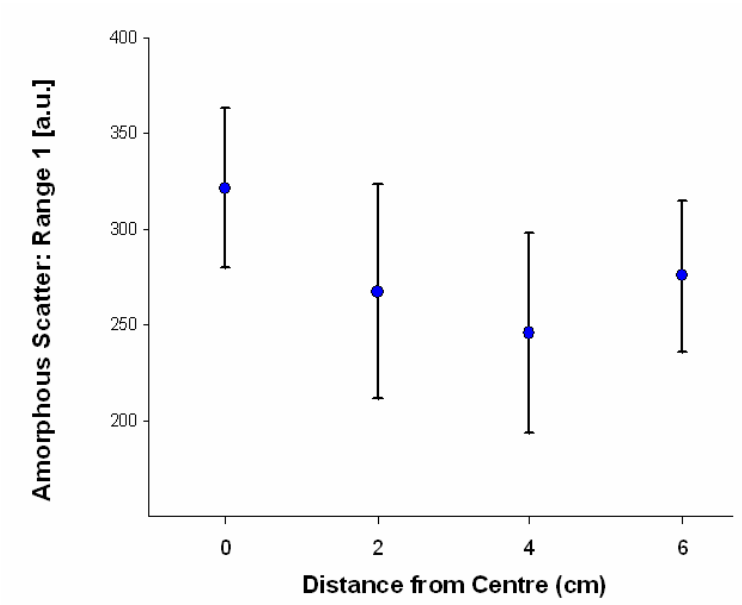


Figure 6.10: Average values of the amorphous scatter calculated for Range 1 ($q = 0.25\text{–}0.6\,nm^{-1}$) for the four patients studied, against distance from the centre of the palpable lesion. Error bars represent 95% confidence intervals.

6.7 Two-Dimensional Mapping of Four Patients Using the Laboratory Source

Although no statistically significant differences with distance from the primary site could be seen with the series tissue samples from the four patients imaged with the laboratory source (Section 6.5), based on the results from Section 6.3, the amorphous scatter for Range 1 showed the most classification potential between the tissue groups (see Table 6.3). For this reason, the amorphous scatter for Range 1 was used to create 2D maps for each of the 4 patients. Gaussian Kernel Smoothing was performed on the four series mastectomy patients as described in Section 5.2.2. Figures 6.11-6.14 show the 2D maps for the four patients with the histopathological diagnosis for each tissue point overlaid.

Patient 1 (Figure 6.11) showed the best agreement with histological diagnosis, except for the P(+4) point, which was diagnosed as invasive carcinoma, but the 2D model suggested that the area contained normal tissue. The histological diagnosis for Patient's 2 and 4 (Figure 6.12 and Figure 6.14, respectively) suggest that the benign tissues had a high degree of disorder in the tissue structure, which was consistent with the high amorphous scatter data of the benign primary site tissues imaged with the laboratory source (as described in Section 6.3.1). Patient 2 had their centre point diagnosed as invasive carcinoma, but the 2D model suggested normal tissue at the centre (blue represents low amorphous scatter, which correlates to normal tissue).

Patient 3 (Figure 6.13) had a full mastectomy procedure where invasive carcinoma was confirmed by surgical pathology examination therefore, we expected the centre tissue point for this patient to be invasive carcinoma. The pathologist was blind to any patient or other information related to the tissue samples during histopathological diagnosis of the tissue samples used in this study. Although the centre tissue point for Patient 3 was most likely invasive carcinoma, the pathologist's diagnosis was that it was benign. Patient 3 did not show the centre point (diagnosed as benign) as the focal point of increased amorphous scatter for Range 1, but rather at the P(-6) point which was diagnosed as being normal parenchyma. Patient 3 also showed some evidence of wrap-around effects on the P(+6) edge of the model.

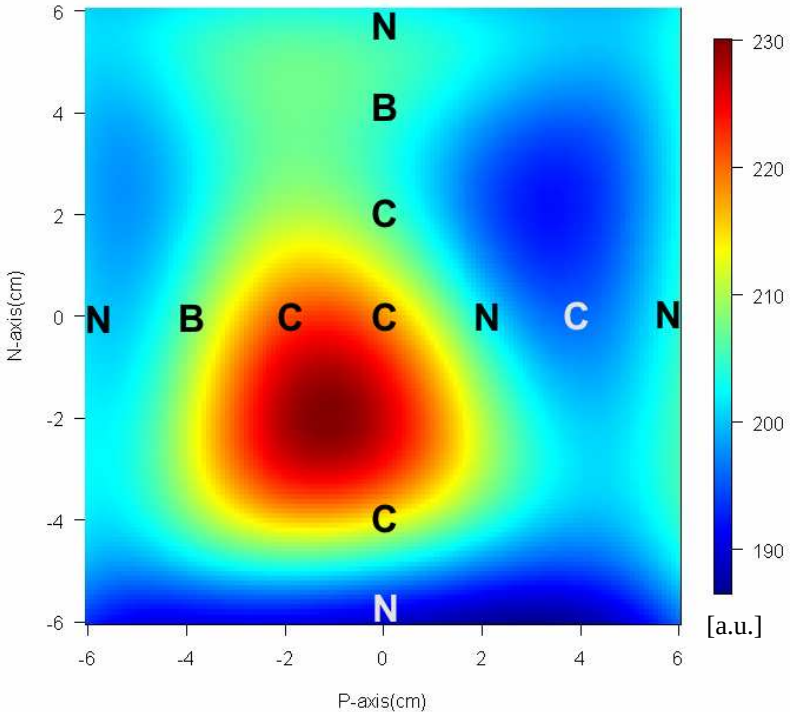


Figure 6.11: 2D Gaussian maps of the amorphous scatter for Range 1 of Patient 1. Values were not normalized, so the scale shown is absolute for the amorphous scatter for Patient 1. C represents tissue diagnosed as invasive carcinoma, B diagnosed as benign, and N diagnosed as normal parenchyma.

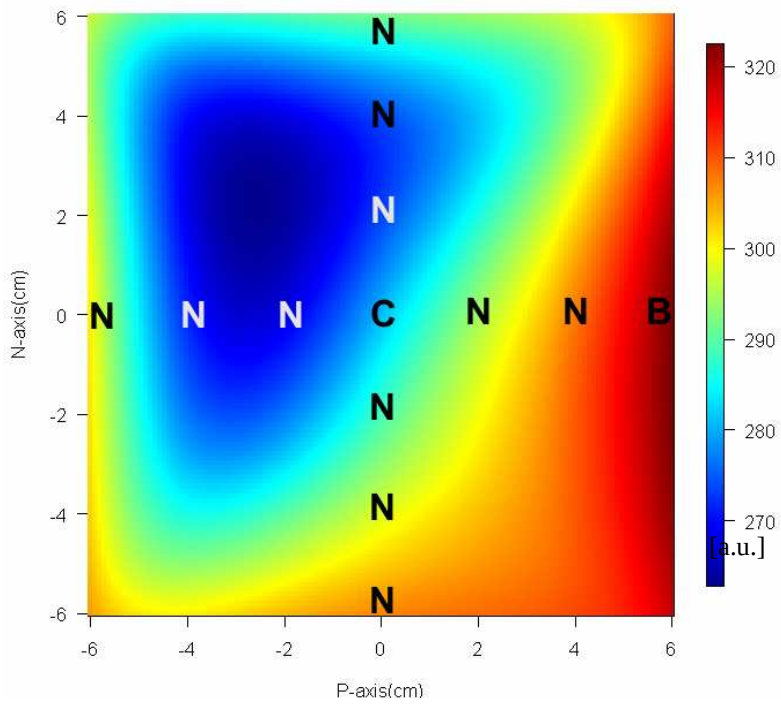


Figure 6 12: 2D Gaussian maps of the amorphous scatter for Range 1 of Patient 2. Values were not normalized, so the scale shown is absolute for the amorphous scatter for Patient 2. C represents tissue diagnosed as invasive carcinoma, B diagnosed as benign, and N diagnosed as normal parenchyma.

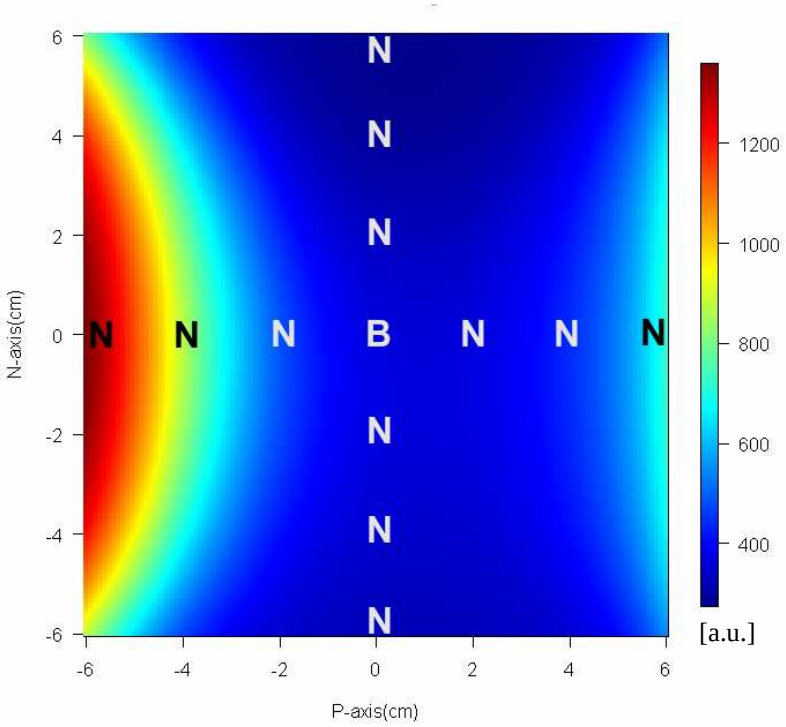


Figure 6.13: 2D Gaussian map of the amorphous scatter for Range 1 of Patient 3. Values were not normalized, so the scale shown is absolute for the amorphous scatter for Patient 3. B represents tissue diagnosed as benign and N represents tissue diagnosed as normal parenchyma.

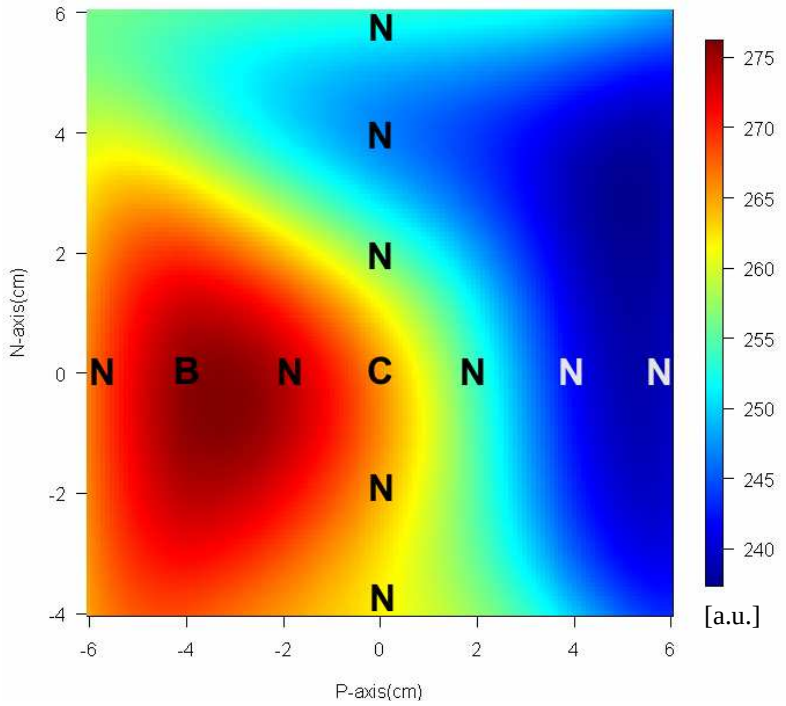


Figure 6.14: 2D Gaussian map of the amorphous scatter for Range 1 of Patient 4. Values were not normalized, so the scale shown is absolute for the amorphous scatter for Patient 4. C represents tissue diagnosed as invasive carcinoma, B diagnosed as benign, and N diagnosed as normal parenchyma.

In general, the 2D maps were inconsistent with histopathological diagnosis of these four patients and cannot be used reliably to show changes in tissue structures that can be correlated with disease. This is perhaps not surprising given the contradiction in the amorphous scatter in invasive carcinoma versus normal and invasive carcinoma versus benign tissues between the two X-ray sources discussed in Section 6.5.

6.8 Discussion

Analysis of the synchrotron data in Chapter 4 found that the most useful SAXS features that can be used for classification of the four breast tissue groups (invasive carcinoma, benign, normal, and mammoplasty) were the amorphous scatter and the third order axial collagen D-spacing. For the laboratory source data, the amorphous scatter also showed to be the most useful classification feature.

The results presented in Section 6.5 suggest that the size of the scatterers is characteristic of the breast tissue type. In particular, the differences between normal and mammoplasty tissue types are attributed to large size scatterers ($d > 25.13$ nm), so in order to differentiate between these two tissue types, SAXS data at this low q-range must be visible ($q < 0.25$ nm^{-1}). The laboratory source was used in high flux mode with a customised beam stop which meant that the minimum q was 0.25 nm^{-1} (as discussed in Section 3.2), but can be operated in low-q mode by using the smaller pinhole diameters which gives a minimum q of ~0.05 nm^{-1}, according to the manufacturer [24, 25]. This would also require a smaller beam stop. Although the flux available would be compromised in the low-q configuration, the laboratory source may show the distinction between normal and mammoplasty tissue types at the lower q-ranges. This warrants further investigation because normal and mammoplasty tissues cannot be differentiated with histopathological techniques (as discussed in Chapter 2). If changes in tissue structure are indeed occurring within the normal tissue of the breast, further investigation into the biological mechanisms causing these changes may provide insight into how disease forms and progresses through tissue.

The amorphous scatter between benign and normal tissues was not distinguishable at the q-range examined with the synchrotron source ($q = 0.1$-0.6 nm^{-1}). However, the laboratory source identified that the differences between these two tissue groups are due to scatterers of size $d = 4.83$-10.47 nm ($q = 0.6$-1.3 nm^{-1}). The size of the scatterers that characterise the differences between these two tissue types

are smaller than the scatterers associated with differentiating normal versus mammoplasty tissues. This could suggest that mammoplasty tissues are characterised by large scattering components of sizes d > 25.13 nm, and benign tissues are characterised by scattering components of sizes 4.83 nm < d < 10.47 nm. It would follow that normal tissues can be identified by scatterers of sizes 10.47nm < d < 25.13 nm. This hypothesis is further supported by the results indicating that the differences between benign and mammoplasty tissues are characterized by scatterers over this entire range: d > 4.83 nm, q < 1.3 nm^{-1}, which includes both the benign tissue range (d = 4.83-10.47 nm) and the mammoplasty tissue range (d = 25.13-62.83 nm). The size of the scatterers within breast tissue therefore decreases from mammoplasty → normal → benign tissue types. This follows from biochemical analysis and electron micrographs (see Figure 1.3) which show that diseased tissues are more degraded than healthier tissues.

The amorphous scatter between invasive tissue and mammoplasty tissues is highly significant over the entire q-range studied. Clearly, there are highly significant differences between diseased and healthy tissues that can be identified using SAXS with both a synchrotron source and a laboratory source. However, the other amorphous scatter comparisons with invasive carcinoma tissues (invasive carcinoma vs. benign and invasive carcinoma vs. normal) were inconclusive and showed conflicting results. This suggests that the amorphous scatter for invasive carcinoma tissues is not restricted to the size of the scatterers, but is rather a disease that affects breast tissue in a highly complex and variable manner. Tissue sample heterogeneity and sampling numbers may be contributing to the conflicting comparisons seen in Table 6.5 with invasive carcinoma versus benign and invasive carcinoma versus normal tissue types. The differences in the comparisons of the 66 patients with both the synchrotron and lab camera of these two comparisons can be attributed to the different sample sizes in each tissue group (see Table 6.2).

For the purposes of classification and diagnosis, it is very important to be able to distinguish between invasive disease, benign disease, and normal tissue types so that proper treatment can be administered (or unnecessary treatment can be avoided). Further investigation into the size of the scatterers and their classification potential with regards to invasive carcinoma tissue is warranted. Differences in grade, previous treatment, menopausal or hormonal status, and prescription drug use may affect the amorphous scatter values and future work in relating these clinical factors with the size of the scatterers may provide insight into disease components.

The analysis of the three SAXS features (amorphous scatter, fifth axial D-spacing, and FWHM of the fat peak) obtained using the laboratory source showed that the parameters associated with the amorphous scatter can be used to differentiate between the four tissue groups studied. The FWHM of the fat peak also showed statistical significance in differentiating between the tissue groups. The discriminate function analysis shown in Figure 6.6 showed excellent grouping of tissue types. However, all three SAXS features (relating to the amorphous scatter, fifth order axial collagen peak, and the fat peak) must be present in the SAXS image in order for this model to work. Only 30% of the patients (20/66) had all 16 variables present in their SAXS image, suggesting that each of the features plays an important role in the classification of breast tissue. However, an exclusion of 70% is obviously not sufficient in order to be effectively used as a classification model (i.e. only 30% of patients could be evaluated) if this sampling is typical of a larger population.

One feature that could be seen on the laboratory source images that could not be seen on the synchrotron source images was the peak from the fat within the tissue sample. For the 45 patients who had the fat peak present on their laboratory SAXS camera images, the position of the fat peak was found to be at $q = 1.48 \pm 0.03$ nm^{-1}. This is equal to the q position of the fat peak quoted in the literature: $q = 1.49$ nm^{-1} [120].

The only parameter associated with the fat peak that showed a trend with disease was the FWHM. The FWHM of the diffraction peak decreased with disease and was significantly different between invasive carcinoma and mammoplasty tissues. The width of the peak indicates the amount of variability in the distribution of the particle sizes in the sample. Therefore, a wide peak suggests that the distribution of the particles is highly variable, whereas a narrow peak indicates that the particles in the fat are have a more ordered structure. The FWHM of the fat peak (shown in Figure 6.3) showed a general trend (although not statistically significant) of increasing FWHM with severity of disease (when considering d-space). Although the laboratory source was not able to significantly distinguish this feature between all of the tissue groups, it may be worthwhile to explore the features associated with fat using a larger sample population with tissues containing high amounts of fat to ensure that the fat peak is present on the SAXS image for analysis.

To our knowledge, only two other groups have examined the scattering features of fat tissue in breast patients: Castro *et al.*[17, 18] and Fernandez *et al*. [49-51, 120]. Castro *et al*. [17, 18] found that the amplitude of the fat peak was higher

for healthy tissues compared to cancerous tissues (5.320 ± 0.160 and 3.846 ± 0.115, relative to the background intensity, respectively). Our data did not find any correlations in the amplitude of the fat peak with tissue type. Fernandez *et al*. [51] examined the average intensity between healthy and invaded fat between the q-range of 0.30-0.50 nm^{-1} and found that the intensity increased by a factor of 5 with invaded tissues. They suggested this could indicate that the invasion of cancer in adipose tissue introduces inhomogeneous structures at the supramolecular level. Newly formed collagen fibrils are known to form in developing invasive carcinoma which provides a pathway throughout the fatty tissue to spread to other areas [94, 95, 97]. So it would be expected that diseased fat tissue would have a higher scattering intensity compared to healthy fat tissue because of the presence of other contributing scatters.

In the next and final chapter of this thesis, the major findings of the studies using both the synchrotron source and the laboratory source are summarised and discussed, and directions for future work are proposed.

Chapter 7

Summary and Conclusions

7.1 Summary of Findings

At the commencement of this work, several studies had suggested that SAXS data could be used to discriminate between healthy and diseased tissue of the breast [16, 38, 41, 43, 75, 105]. This work addressed several aspects of SAXS for tissue diagnosis, including the evaluation of a large patient cohort with a diverse set of tissue types, representative of clinical populations that have not been examined in previous studies.

The purpose of this work was to further investigate the relationship between SAXS parameters and diseased and healthy breast tissue types with the view of developing SAXS for potential tissue classification and also to provide information that could be related to disease aetiology research. The project's three aims (Chapter 1) have been addressed throughout this thesis via the analysis of SAXS images of tissue samples excised from the centre of the primary lesion and at ~2 cm distances away from the centre of the lesion. This was accomplished by acquiring SAXS images from 543 breast tissue samples. A total of 431 tissue samples were imaged with a synchrotron source, and 112 tissue samples were imaged using a laboratory source.

The first aim of this project was to identify which SAXS features could be used for classification of tissue types. Features relating to the supramolecular structures of breast tissue, including collagen and fat components, were analysed to determine if any features could be used to classify four tissue groups: invasive carcinoma, benign, normal parenchyma, and mammoplasty tissue. Normal tissue was excised from patients with known disease in the breast, where these tissue samples were histopathologically normal. Mammoplasty tissue was excised from patients who had no history or presence of disease in the breast. Comparisons between normal tissue and mammoplasty tissue had not been performed with a large cohort of patients prior to this study, nor had the SAXS features between these two tissue types been extensively studied.

The second aim of this thesis was to use the features that showed the best separation between the tissue groups to investigate the differences in breast tissue samples at varying distances away from the primary lesion. This was accomplished by examining tissue samples taken from the centre of the palpable mass (primary site) (Chapters 4 and 6) and tissue samples taken at 2 cm intervals radially outward along two axes (series sites) (Chapters 5 and 6).

The third aim of this thesis was to determine if a laboratory SAXS camera can show differences between tissue types that would further justify the clinical applications of SAXS for diagnosis and classification of breast disease. Tissue samples from 66 patients were analysed, and feature analysis as a function of distance away from the centre of the tumour was conducted (Chapter 6).

The major findings of this thesis were:

(1) There are significant differences in the tissue structure between healthy tissue from patients with no history or presence of disease (mammoplasty) and histopathologically normal tissue from patients with known disease present in the breast. This suggests that SAXS can identify differences between tissues that standard histopathological analysis cannot.

(2) The axial D-spacing of collagen in both normal and mammoplasty tissue is significantly lower than the axial D-spacing of collagen in invaded breast tissue. Higher axial D-spacings indicate that modifications in the collagen fibril structure occur in invaded tissues.

(3) The amorphous scatter is the best classifier of the four tissue types: invaded carcinoma, benign, normal, and mammoplasty.

(4) The amorphous scatter in different q-ranges may provide better classification between different breast tissue groups. This suggests that the size of the scatterers is the differentiating factor between tissue types.

(5) There is a significant decrease in both the amorphous scatter and axial D-spacing from the centre of the primary lesion and ≥ 2 cm away from the centre. Changes occurring at the tens to hundreds of nanometre scale can be seen between the lesion and the normal tissue surrounding the lesion, which may have implications relating to surgical margin disease clearance and provides a better understanding of malignant aetiology between the diseased/normal tissue border.

(6) Mapping of the amorphous scatter and axial D-spacing over the 2D area of the excised breast mass shows areas of disease and non-disease when using 6 or more data points. The maps derived from synchrotron data agree with histopathology diagnosis and may be useful for identifying directional spread of invasive carcinoma, as well as validate clearance of disease from margins of surgically removed disease.

(7) The laboratory SAXS camera can be used to classify the four tissue groups using 16 features related to the amorphous scatter, the fat peak, and the 5th order axial collagen peak.

In this final chapter of the thesis, the implications of the main findings as well as further suggestions for future work are discussed.

7.2 Discussion of the Main Findings

A comprehensive investigation of SAXS features related to breast tissue disease was performed in this work. All of the features listed in Table 1.1 were examined using a synchrotron source, and it was found that the amorphous scatter and the axial D-spacing were the most significant features that were able to distinguish between the four breast tissue groups (Chapter 4). The axial D-spacing, which describes the distance between the gap-overlap spacing of the collagen molecules, increased with presence of disease. The amorphous scatter, which represents the area under the fitted quadratic function of the SAXS 1D profile, also increased with presence of disease and showed the most significance between tissue groups. The slopes of the plotted √(equatorial peak area) versus equatorial peak amplitude showed that the equatorial packing distances between the fibrils for invasive carcinoma tissues are

more variable and perhaps more disrupted compared to the other three tissue groups studied. The polar distribution data showed no correlations in separating the tissue groups which suggests that orientation of the fibrils does not differ between tissue types.

The amorphous scatter showed a significant difference ($p = 0.0025$) between mammoplasty and normal tissue types, suggesting that there are supramolecular changes in histopathologically similar tissues. A higher scattering potential also corresponds to an increase in the specific surface area of the scatterers, which is related to a higher degree of degradation of tissue structures. This was also evident in the analysis of the amorphous scattering over the q-ranges examined with both the laboratory source and the synchrotron source. The differences between normal and mammoplasty tissues are characterised by large scatterers of size $d > 25.13$ nm ($q < 0.25$ nm^{-1}). Benign and normal tissue types can be differentiated by scatterers of size 4.83 nm $< d < 10.47$ nm (0.6 nm^{-1} $< q < 1.3$ nm^{-1}), and benign and mammoplasty tissue types of scatterers $d > 4.83$ nm ($q < 1.3$ nm^{-1}). Thus, it was hypothesized that the size of the scatterers decreases with severity of disease types from mammoplasty → normal → benign. This follows from biochemical analysis and electron micrographs (see Figure 1.3) which show that diseased tissues are more degraded than healthier tissues, thus, the scatterers within degraded tissues are smaller than those in healthy tissues.

The amorphous scatter comparisons with invasive carcinoma tissues were somewhat inconsistent between the two X-ray sources. Although the amorphous scatter over the entire q-range examined ($q = 0.1$-2.3 nm^{-1}) showed significant differences between invasive carcinoma and mammoplasty tissue types for both X-ray sources, the tissue pair comparisons between invasive carcinoma versus normal and invasive carcinoma versus benign were contradictory between the laboratory source data and the synchrotron source data. This may be due to the heterogeneity of the tissue cores, or perhaps shows that invaded tissue has a complex tissue structure that is varied. Further investigation of the size of the scatterers in relation to tissue types and disease may provide insight into the structural changes occurring within the breast for each tissue group.

It is known that laminin is primarily involved in the process of invasion and metastasis of malignant tumours, an idea which is supported by biochemical analysis and electron micrograph images that show that the laminin-binding type I trimer collagen fibrils are disintegrated in diseased tissues [94]. This may account for the higher scattering intensities we observed in normal tissues sampled from patients

with known disease present in the breast compared to mammoplasty tissues that have no history or presence of disease. The amorphous scatter from the synchrotron source could not distinguish, however, between benign and normal tissue types (Table 4.5 and Figure 4.5), but the laboratory source data suggests that the amorphous scatter at higher q-ranges may be useful for distinguishing between these two tissue groups. A synchrotron source is capable of obtaining SAXS images over many q-ranges since the camera length and beam energy can easily be changed to accommodate data acquisition. A comprehensive analysis of the amorphous scatter over a multitude of q-ranges (even less than $q = 0.1$ nm^{-1} and greater than $q = 2.3$ nm^{-1}) may show more information in discriminating components of tissue types that can be used for classification. In particular, higher q-ranges may prove to show discriminating features related to invasive carcinoma, since invaded tissues are highly degraded and would therefore have increased scattering from very small scattering components. The laboratory source is limited in experimental configuration, where the pinhole diameters can be smaller to accommodate a smaller q-range ($q < 0.25$ nm^{-1}), but the camera length cannot be changed due to beam divergence. However, an X-ray tube that produces a higher energy beam could be used to obtain lower q-values.

It is important to note that current histopathological diagnosis procedures (such as H&E staining) cannot discern between normal and mammoplasty tissues (see Figure 2.6c and d), whereas SAXS can detect variations. Differences between normal and mammoplasty tissues reveal that structural changes are occurring in the breast before standard histopathological diagnosis can characterise the presence of these changes. This has implications for diagnostics, therapeutic methods, and understanding disease progression.

The difficulties associated with obtaining a tissue sample with all of the information needed to provide an accurate histopathological analysis, with respect to biopsy, is a major limitation in the diagnosis of breast cancer [5, 9, 37, 65, 71]. Obtaining tissue from a palpable mass is considerably easier than from a lesion that can only be seen with the aid of mammography or ultrasound, but sampling error still occurs. In fact, normal tissue can be biopsied from a known palpable breast mass, where the histopathological diagnosis of the normal tissue becomes suspicious. These types of sampling errors occurred with the tissue samples examined for this study. Out of the 116 tissue samples excised by biopsy from the centre of a lesion which was confirmed invasive carcinoma, only 54 (47%) tissue samples had histopathologically confirmed invasive disease present in the tissue

core. The high variability in tissue sampling was also seen with the series tissue sample analysis. With the series tissues, it was expected that all of the centre points contained invasive carcinoma, which was not the case for 31/56 patients. This type of biopsy error occurs in the clinic [5, 71] and is considered a diagnostic limitation. Sampling errors of this magnitude make it even more important to identify the presence of disease within histopathologically normal tissue and the differences associated with mammoplasty tissues.

Radiation damage of biological samples causes breaking of bonds and disruption of the molecular and supramolecular structures within the sample. Fernandez *et al*.[51] studied the effects of radiation damage on several samples of tissue (imaged at Station ID02 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France) and found clear changes in scattered intensity and collagen ordering after approximately 10 seconds of exposure, but no damage occurred to tissues with accumulated doses of $< 10^5$ Gy. The dose received by the tissue samples at the SRS Daresbury and at the laboratory source received $< 10^5$ Gy, therefore radiation damage did not affect the tissue samples in this study. If radiation damage of the tissue cores becomes an issue, the use of a gas wire detector can reduce exposure without compromising image information. Gas wire detectors are photon counting which means that several images can be taken of a sample and added together in order to increase the signal-to-noise ratio. If radiation damage to the tissue samples after a certain exposure time becomes an issue, the photon counts taken after the damage point can be discarded from the final image. Integrating detectors, such as IPs, are not capable of retaining threshold measurements.

The results presented in this thesis suggest that synchrotron acquired SAXS images of breast tissue can be used to distinguish between normal tissue sampled from a patient with known disease and normal tissue sampled from a healthy patient (mammoplasty). Therefore, a particular advantage of SAXS is that disease can be identified in a patient, even though there was sampling error in the biopsy, which histopathology would not have been able to identify. Our work suggests that SAXS may provide a highly sensitive diagnostic aid that overcomes issues dealing with tissue heterogeneity and sampling error from both histopathological processing and biopsy.

Perhaps the most immediately obvious potential of differentiating between normal and mammoplasty tissues is that surgical margins can be tested for evidence of tissue changes to ensure that all of the infected tissue has been cleared from the patient. The risks associated with recurrence are dependent upon the

histopathological diagnosis in and around the surgical margins of removed tissue to ensure no microscopic evidence of disease is present. SAXS may provide a more sensitive tool for ensuring that disease at the tens to hundreds of nanometre scale has been sufficiently removed from the patient. This can have serious repercussions relating to risk of recurrence and survival rates of breast cancer.

The axial D-spacing of collagen has been suggested by several groups to have potential in differentiating between diseased and healthy breast tissue types [16, 38, 41, 43, 75]. This feature was extensively examined in Chapter 4, where the axial D-spacing of normal and mammoplasty tissue types (d = 64.88 ± 0.04 nm and d = 64.91 ± 0.02 nm, respectively) were significantly lower than the axial D-spacing of invasive carcinoma tissues (d = 65.00 ± 0.05 nm), which are listed in Table 7.1. An increase in the axial D-spacing of the collagen suggests that invasive carcinoma tissues have a larger periodical spacing of the collagen fibril structures (see Section 1.3.1). The change in the axial D-spacing may be due to the replacement of normal type I collagen molecules by the trimer, where the three chains are identical [94]. Ageing may also affect the collagen period [30], since younger women have dense breasts with a high amount of fibrous tissue whereas older women have less dense and fibrous breast tissue and much more fat and adipose tissue. It would follow then, that differences between disease types in younger women may be more sensitive to collagen feature analysis, whereas differences attributed to fat may be more suitable for classification of tissue types in older women.

The collagen axial D-spacing between breast tissue types has been extensively studied by Fernandez *et al*. [41-43] and Lewis *et al*. [77]. Table 7.1 shows the results presented in Chapter 4 for the axial D-spacing of the four tissue groups and compares the D-spacings found by both Fernandez *et al*. [49, 51] (fresh tissue) and Lewis *et al*. [77]. Most of the D-spacing values for the tissue types are in close agreement between the three studies (within ± 2 standard deviations (SD)), except with the benign tissue groups in this thesis, which had a much larger axial D-spacing (d_{benign} = 64.92 ± 0.04 nm) compared to those of Lewis *et al*. [77] (d_{benign} = 64.52 ± 0.04 nm). These absolute differences can be attributed to experimental differences (discussed in Section 4.2.1a). It should be noted that the current study had a substantially reduced SD for axial D-spacing measurements compared to Fernandez *et al*. [51], reflecting the much larger number of samples examined (Thesis = 130 patients, Fernandez *et al*. [51] = 7 patients). Given the differing SDs

Table 7.1: The mean axial D-spacing values with Standard Deviations (±SD) for 130 primary site tissue samples imaged with the SRS Daresbury in this thesis and the results obtained from the 7 patients by Fernandez *et al*. [51]. Lewis *et al*. [77] stated the axial D-spacing ranges found in their 44 patients in their publication, but these values were obtained through personal correspondence.

Tissue Type	**Sidhu Thesis**	**Fernandez *et al*.[51]**	**Lewis *et al*.[77]**
Invasive Carcinoma	65.00 ± 0.05 nm	65.3 ± 0.2 nm	64.90 ± 0.05 nm
Benign	64.92 ± 0.04 nm	65.0 ± 0.1 nm	64.52 ± 0.04 nm
Normal	64.88 ± 0.04 nm	65.0 ± 0.1 nm	64.87 ± 0.02 nm
Mammoplasty	64.91 ± 0.02 nm	N/A	64.85 ± 0.03 nm

of each study, it is more useful to instead compare the general trends between the tissue groups. The decrease in the axial D-spacing between invaded tissue and benign/healthy tissue in this thesis agrees with the results found by Fernandez *et al*. [51]. However, Lewis *et al*. [77] found that the axial D-spacing of invaded tissues and mammoplasty tissues were similar, which was not the case with the 130 patients analysed in Chapter 4.

In order to determine if differences in parameters as a function of distance away from the centre of the palpable lesion are due to local or systematic changes in the tissue structures (Chapter 5), 165 series tissues from 56 patients imaged with the SRS Daresbury were analysed. Both the amorphous scatter and axial D-spacing showed a significant decrease between the centre tissue point and the other tissue samples taken up to 6 cm away ($p = 2.36 \times 10^{-15}$ and $p = 0.010$, respectively). Values in the D-spacing along the nipple-axilla line (N-axis) were similar to the D-spacing of the centre point, but the tissue samples taken superior and inferior within the breast (P-axis) showed a decrease compared to the primary site point. This was expected because there are more lymphatic vessel structures and ductal components in the breast between the nipple and axilla (N-axis), which are known to facilitate the spread of disease [28, 32]. Thus, if a cancerous lesion originates along the nipple-axilla line, these structures have a higher probability of becoming infected before the tissues superior and inferior of the nipple (P-axis), which is mostly composed of connective and fatty tissues. The results presented in Chapter 5 suggest that it is in these lymphatic and vascular breast components that disease spread is facilitated. If SAXS was to be used to detect supramolecular changes in breast tissue as a screening protocol, tissue samples along the nipple-axilla line would be the most appropriate location to investigate, however, further research is needed to verify this assumption since the axes the tissue samples were harvested from were not consistently in the same direction. The direction of the N-axis varied by ~45º, which of course affected the placement of the P-axis. Determining a more constant

sampling area would provide more specific information of the changes in tissue structures in certain areas of the breast which could then be correlated to the ductal or vascular structures of the breast. Axes or grid locations of where tissue samples are to be harvested in reference to a bony landmark may be more appropriate.

Directional spread of disease was expected to be seen with the 2D mapping of the amorphous scatter and axial D-spacing variables for patients who had series tissues samples excised over their breast mass. The 2D maps calculated with the synchrotron source data showed some agreement with histopathological diagnosis of the tissue samples. Assuming that low amorphous scatter and axial D-spacing values are correlated with normal tissue and high values are correlated with invasive disease, the 2D maps can potentially be used to see directional spread of disease, in particular, towards the axilla and lymph nodes (as mentioned previously), and also to determine if surgical margins need to be extended if structural changes are present around the edges of the removed tissue mass. The 2D maps may be more informative if more tissue samples were examined at smaller distance intervals throughout the breast mass, which may provide more information related to the disease/normal boundary. This is discussed further as possible future work for the research in Section 7.3. The 2D maps of the amorphous scatter calculated for the four patients examined with the laboratory source did not have good correlation with histopathological diagnosis. The poor maps seen with the laboratory source data may be due to the relatively low sensitivity of the laboratory source parameters (compared to the synchrotron derived parameters) to distinguish invasive carcinoma with normal tissues (discussed below).

A similar comprehensive review of SAXS features was performed using a laboratory source in order to determine if a SAXS system can be used in a clinical setting (Chapter 6). Due to the different camera length of the laboratory source (~1.6 m) compared to the synchrotron source (~6.3 m), further investigation of the SAXS features related to the fat peak and the amorphous scatter in several q-ranges were possible. There were 112 tissue samples (66 primary sites, 46 series sites from 4 mastectomy patients) imaged with the laboratory source. The amorphous scatter feature showed the most potential for classification of the tissue groups with both the synchrotron source and the laboratory source.

Clear classification of the four tissue groups, with excellent separation between each group, was possible only if all 16 parameters related to the amorphous scatter, fat peak, and 5th axial collagen peak were obtainable from the laboratory source SAXS images (Figure 6.6). This requirement suggests that a multitude of

features contribute to the supramolecular changes occurring in the breast when disease was present. All 16 parameters were visible for 20/66 of the patients examined using the laboratory source. Although the model shown in Figure 6.6 shows perfect specificity and sensitivity, a patient eligibility of 30% is not clinically advantageous as a diagnostic tool. In order to have both the fat peak and the collagen peak visible on the SAXS images (the amorphous scatter is always available), the tissue samples must have both fatty tissue and fibrous tissue present. To achieve this, two tissue samples could be harvested from the patient, one with high fat content and one with high fibrous content, and the 16 parameters could be obtained from two SAXS images for use in the classification model. Identifying these two tissue types, as well as ensuring that the samples are taken from the diseased lesion, would limit the accuracy of the model. However, based on the results presented in this thesis, SAXS can distinguish between normal and mammoplasty tissue types using the amorphous scatter. The amorphous scatter can be used to identify possible sampling errors associated with biopsy, as well as changes at varying distances away from the primary lesion. This suggests that SAXS information does not have to be dependent on the information from one tissue sample. Instead, the information culminating from several tissue samples may provide the information necessary to classify tissue samples using the 16 parameter discriminant variable model. Because the 16 parameter discriminant model has perfect classification of the four tissue groups, further investigation of these parameters with more patients is warranted.

In order to increase the patient eligibility criteria, a classification model based only on the amorphous scatter and the Full-Width-at-Half-Maximum (FWHM) of the fat peak was used to determine the sensitivities and specificities. Good sensitivities (> 80%) were found between invasive carcinoma versus mammoplasty, benign versus mammoplasty, and normal versus mammoplasty (see Table 6.4). Thus, the model used with the laboratory source was capable of distinguishing diseased tissues from healthy tissues excised from patients with no history or presence of disease. The sensitivities and specificities may be improved if further investigation into the amorphous scatter at different q-ranges can be explored. The q-range for the laboratory source extends to much larger q than that probed by the synchrotron source, where differences in the size of the scatterers were able to classify benign from normal tissues, which the synchrotron range could not (q = 0.1-0.6 nm^{-1}). Differences between tissue types based on the amorphous scatter within specific q-ranges suggests that changes in supramolecular structures also occur over a much wider range of scatterer sizes than has previously been explored.

The sensitivities and specificities from the Receiver Operator Characteristic (ROC) curves calculated with the laboratory source between invasive carcinoma versus benign, invasive carcinoma versus normal, and benign versus normal tissues were invalid due to the tight clustering of these tissue groups to one another and the small number of tissue samples used to generate the model. Only 66 primary site tissue samples could be imaged with the laboratory camera, rather than all 130 primary site tissue samples, due to time constraints. In order to perform an accurate ROC curve analysis, additional invasive carcinoma, benign, and normal tissue samples are needed to obtain the sensitivities and specificities associated in identifying these tissue types. These three tissue groups are important to differentiate in clinical diagnosis because the avenues of treatment for each of these tissue groups are completely different. In order to implement SAXS within the clinic, the laboratory source must be able to classify invasive carcinoma, benign, and normal tissue types with high accuracy. Further work into the amorphous scatter ranges as well as increasing the number of tissue samples analysed for classification, is important for determining whether the laboratory source can be used for clinical diagnosis.

The comparison between invasive carcinoma and normal tissues is important in classification modelling in order to quantify the sensitivity and specificity of a diagnostic model. With all of the q-ranges examined with the laboratory source, the amorphous scatter showed significant differences between invasive carcinoma and normal tissue types. Thus, we expected that the 2D mapping of this feature with the laboratory source data would show good correlation with histopathology and would also show accurate gradients of the changes in the amorphous scatter with distance from the primary site. These relationships were not observed, and the 2D maps of the amorphous scatter in Range 1 of the four series patients did not show good correlation with histopathological results. The variations between histopathology and the 2D maps may have been due to tissue heterogeneity effects and patient variability, which was seen between the amorphous scatter comparisons of the 66 patients between the two X-ray sources: the laboratory source significantly differentiated between invasive carcinoma and normal tissue, but the synchrotron data for the 66 patients (in Range 1) showed no significant difference between these two tissue types. Correlation between the 2D maps and histopathology using the synchrotron source was far better than the 2D maps calculated with the laboratory source (Chapter 5). The higher significant difference between invasive carcinoma and normal tissues with the synchrotron source (p =

0.0023) may have been the limiting factor for producing accurate 2D maps, compared to the laboratory source (p = 0.0089). It therefore can be assumed that the laboratory source is not sensitive enough to accurately differentiate between invaded and normal tissue types. In fact, a threshold p-value may need to be determined that can justify the use of 2D mapping parameters with laboratory source data. Analysing the amorphous scatter within specific q-ranges, as mentioned previously, may provide better mapping criteria for series tissue analysis with the laboratory source.

7.3 Improvements and Future Work

Using a synchrotron radiation source for clinical use of SAXS for diagnosis and classification of tissue samples is not a feasible alternative to current diagnostic methods. The laboratory source does show potential in classifying breast tissue, which can aid pathologists and surgeons for diagnostic purposes. Before the laboratory source can be implemented for use in the clinic, further investigation is needed to prove the efficacy of tissue classification with the SAXS Osmic instrument. As discussed previously (in Section 7.2 and Chapter 6), the amorphous scatter at the lower q-ranges may provide the necessary information to classify the tissue groups. By changing the configuration of the laboratory apparatus to low-q mode, where the diameter of the pinholes are decreased compared to the diameter of the pinholes used in high-flux mode for obtaining the SAXS data in this thesis (as discussed in Chapter 3), $q < 0.25$ nm^{-1} can be achieved. The lowest q-value seen in the high-flux mode is theoretically $q = 0.070$ nm^{-1}; however, due to interference from the beam stop, $q = 0.25$ nm^{-1} was the lowest value accessible for accurate analysis. In the low-q configuration of the SAXS Osmic system, q-values down to $q = 0.048$ nm^{-1} are achievable, but the flux decreases from 2.8×10^7 photons/sec to 8.0×10^6 photons/sec.

Of course, even though data at low q-ranges are technically achievable with the SAXS Osmic system, the size of the beam stop and parasitic scattering increases the limits of visible q-space. A smaller beam stop would also have to be used in order to see information within lower ranges $q < 0.25$ nm^{-1}. Using a smaller beam stop is not mechanically difficult, but the measurement of the transmitted beam may become challenging if the beam stop is too small to allow for a semi-transparent beam stop. Instead, the use of an ion chamber would be needed to obtain the transmitted intensity corrections, which may prove to be difficult. More concerning is that a decrease in the flux results in an increase in the exposure time required,

along with a decrease in the signal-to-noise ratio, which is not a favourable compromise when using IP detectors. Using a gas-wire proportional counter increases the signal-to-noise ratio (noise becomes negligible) and alleviates these challenges[‡‡‡‡‡]. Further examination of the amorphous scatter at lower q-ranges is possible by using the low-q configuration of the laboratory system as well as a gas-wire detector. The ability of the laboratory source to differentiate between disease types would further support the use of SAXS for clinical diagnosis of breast tissue, and warrants further study.

An experiment that compares matched fresh tissue with formalin fixed tissue would be worthwhile, especially if clinical applications of SAXS are warranted. Formalin fixed tissue samples are more accessible than fresh tissue samples, because most tissue repositories and tissue banks (hospitals, cancer centres) store formalin fixed tissues. Tissue banks are also useful for obtaining a representative population of patients with matched ages, previous radiation/hormone therapy, and menopausal status. If SAXS is implemented as a clinical aid in diagnostic procedures, formalin fixed tissue analysis would be much easier to handle and obtain for analysis, rather than dealing with fresh tissue samples. Current standard procedure places any excised tissue mass immediately in formalin so the tissue remains preserved for later histopathological examination. Obtaining fresh tissue samples involves excision of the tissue and either immediately imaging the tissue sample or freezing the tissue for later analysis. In terms of clinical efficiency and cost-effectiveness regarding staffing and storage, imaging formalin fixed tissue is a more feasible alternative than using fresh tissue, if the information is not compromised.

Another benefit of examining formalin fixed tissues is that freezing the tissue samples for storage would not be necessary. Freezing and defrosting tissue samples causes some structural damage to the tissue which then makes histopathological analysis difficult. The cellular components of the tissue used for histopathological diagnosis are damaged from freezing and defrosting, which results in incomplete staining of the features necessary for tissue identification. Fixing the tissue in formalin would make it much easier for histopathological diagnosis of the tissue sample after SAXS imaging is performed on the sample.

A disadvantage of fixing the tissue samples in formalin is that the process removes the water content from the samples which may affect the values of the SAXS features analysed in this work. For example, formalin fixing the tissue

[‡‡‡‡‡] A gas-wire detector has been commissioned for use on the laboratory SAXS system to be operational in 2009.

samples would be of particular concern when imaging fat components in the sample, since fat cells contain mostly water. Also, the axial periodicity of collagen fibrils in biological samples is dependent on the amount of hydration in the sample. This has been seen with rat tail tendon, which is commonly used as a calibrant for the axial D-spacing of biological samples in solution [10, 55, 57, 64]. Fernandez *et al*. [49] found that the D-spacing between formalin fixed tissue and fresh tissue were systematically different by 0.1 nm. This difference suggests that the variation in the axial D-spacing between tissue types is preserved between the tissue groups, but the absolute values of the axial D-spacing are different compared to fresh tissue. A comprehensive study that examines formalin fixed tissue and compares the axial D-spacing differences between each tissue group is needed to show that formalin fixed tissue can be used as a classifier of breast tissue types.

There are a number of sample handling improvements that could be made to reduce the error associated with histopathological diagnosis of the tissue samples. The pathologist found it very difficult to diagnose the post-SAXS experiment tissue samples. Artefacts were clearly evident in most of the tissue samples, which also made H&E staining and processing difficult. The glass artefacts (seen in Figure 2.4) seen with many of the tissue samples could be eliminated by using a different sample holder apparatus rather than glass capillaries. Sandwiching the tissue samples between mica slides has been used by several other SAXS groups [16, 43] because mica has a high transparency to X-rays and does not generate background scatter in the small angle regime. In order to perform a more accurate diagnosis of the tissue samples, a frozen section of the tissue core could be taken before SAXS imaging. A downfall to this option, however, is that only a thin slice of the tissue would be histopathologically analysed and that slice would not be part of the SAXS image. As mentioned, using formalin fixed tissue would alleviate some of the challenges associated with histopathological analysis of fresh tissues. The cellular features that are stained for histopathological identification of disease would be more in-tact, and the amount of artefacts seen in the tissues would be reduced.

Larger pieces of breast tissue, rather than 1 mm diameter tissue cores, could be imaged and analysed. In particular, SAXS information throughout a large section of a breast could be mapped in 2D, similar to the 2D maps presented in Chapters 5 and 6. The synchrotron acquired data showed that systematic differences existed in the tissues sampled at varying distances away from the primary site. More tissue samples taken at smaller distance intervals, particular between 0 cm and 2 cm which showed the highest differences, may show a functional trend in the amount of tissue degradation with distance away from the lesion. The 2D mapping models would also

be improved if multiple tissue samples taken over the entire grid were analysed, rather than across two axes. If pathological edges of healthy/invaded tissues could be determined with histopathology, differences in the supramolecular structures across these edges may shed light on the structural differences between disease types and mechanisms of disease spread. A 3D reconstruction of the SAXS parameters over an entire mastectomy volume would also be possible. Scanning SAXS over multiple slices of a mastectomy would be needed to create an accurate 3D SAXS model. However, due to the long exposure times currently needed to obtain SAXS images from the laboratory source, using a synchrotron source would be more efficient for obtaining this kind of data. Thus, large scale SAXS for individual patient analysis is not diagnostically feasible, however, investigating the changes in structural components throughout the entire breast may prove useful in understanding disease progression.

In order to exclude inter-patient variability as well as other clinical effects (such as hormone status, age, and previous drug and cancer treatments which can affect tissue structure), an animal model has the advantage of producing controlled lesions, where specific properties of the tissue can be examined without the inclusion of extraneous variables. Tissue interfaces between tumour and normal areas using SAXS can also be investigated for different types of aggressive carcinomas to provide information related to invasion and disease components. In particular, the differences between histopathologically normal and mammoplasty tissue types can be further examined using controlled tissue, which may provide insight into the specific supramolecular differences without the interferences of age, hormonal effects, or inter-patient variability. Fibrocystic changes that occur in healthy tissues need to be investigated further in order to understand how these structural changes occur within the breast. Understanding the structural constituents of disease progression can provide more research avenues towards better treatments and screening techniques.

Another potential avenue of interest is using SAXS to examine lymph node tissue for presence of disease. Surgical removal of one or more lymph nodes in the axilla is sometimes performed on mastectomy or Wide-Local-Excision (WLE) patients to determine if metastasis could be a factor for the patient. There have been several studies that have found the size of the sentinel lymph node (SLN) and extracapsular extension (ECE) around the SLN to be independent predictors of non-SLN involvement [1, 44, 102]. The cell composition within the ECE is typically

fibrous and very dense. However, if the lymph node is inflamed, which occurs when disease has infiltrated the node, fibrosis may exist and fatty degeneration of the connective tissue may be present [1]. SAXS could be used to see the amount of degradation in the fatty and connective tissues in diseased lymph nodes, where an increase in the amorphous scatter would indicate a higher scattering potential in the tissue. Healthy lymph nodes would have a lower amorphous scatter because the tissue is highly dense and fibrous, where more ordered collagen fibril structures are present.

With the series data analysis, the axial D-spacing of the collagen fibrils along the nipple-axilla line showed similar values as the tissue sample taken at the primary site of the lesion. This correlation suggests that the tissue structures along the nipple-axilla line are similar to the tissue structures present within the diseased lesion, meaning that disease has spread toward the axillary lymph nodes. By determining if tissue samples taken from within the breast are correlated to SAXS information obtained from diseased lymph nodes, information related to the spread of disease through the lymphatic system may be obtained. Lymph node tissue samples were harvested from 28 patients, where a total of 42 lymph node samples were taken: 18 diseased, 24 non-diseased. These lymph node tissues were imaged using SAXS at the Advanced Photon Source (APS) in Chicago, USA. However, there was not enough time to analyse this SAXS data for inclusion in this thesis.

7.4 Conclusions

This study was the largest SAXS study of breast cancer tissues that we know of undertaken to date, with tissue samples from a total of 130 patients imaged with the synchrotron source and 66 patients imaged with the laboratory source. The key features that can significantly distinguish between tissue groups were the amorphous scatter and the axial D-spacing of collagen fibrils in breast tissue. The fat peak also showed potential in separating between tissue types.

Invaded tissues presented with higher amorphous scatter values, suggesting that the electron densities in diseased tissues was higher than in healthy tissues. Thus, invaded tissues have a larger scattering potential due to the higher degradation in the tissue components. In particular, differences in SAXS feature components of normal and mammoplasty tissues suggest that structural changes occurring at the supramolecular level of tissue is occurring in diseased patients, even when

histopathological diagnosis indicates that the tissue is healthy. A comparison of the differing q-ranges probed by the synchrotron source and the laboratory source revealed that the amorphous scatter within certain q-ranges may provide more information specific to each tissue group that can distinguish between certain tissue types. The size of the scatterers appeared to decrease from mammoplasty to normal to benign, suggesting that the degradation of the tissue structures with severity of disease can be seen using SAXS. Invasive carcinoma tissues did not have a particular size of scatterer that could be used to identify invasive tissues. The lack of a single characteristic scatterer size suggests that invasive carcinoma tissues have a complex and varied structure. Also, there were conflicting results in the tissue group comparisons in the amorphous scatter q-ranges between the synchrotron source and the laboratory camera source. Further investigation of the amorphous scatter at different q-ranges is warranted to determine the full potential of using this SAXS feature as a diagnostic indicator of disease. Analysis of the laboratory source data suggests that a combination of the key features of the amorphous scatter, axial D-spacing, and fat peak must be used together for adequate discrimination of breast tissue types. This information can provide insight into the structural components that are responsible for characterising disease, and how they differ between tissue groups.

Tissue samples taken at 2 cm intervals away from the primary site of the lesion were also examined using SAXS in this study. There was a significant decrease in both the amorphous scatter and the axial D-spacing from the centre of the primary site and 2 cm away from the centre of the lesion with the synchrotron source data. Two-dimensional mapping of the changes in the amorphous scatter and axial D-spacing over the area of the breast demonstrated that these parameters can effectively infer the extent of tissue degradation over significant macroscopic distances many centimetres from the primary lesion with relatively few samples. The 2D mapping of the amorphous scatter using the laboratory source data did not show consistent correlation with histopathological results. Further investigation into the accuracy of the laboratory source in differentiating between tissue groups is needed to verify that using this apparatus is suitable for clinical diagnosis of breast cancers. Nonetheless, the 2D maps calculated using synchrotron source data have the potential of showing surgical margins that may have presence of increased supramolecular changes that may be related to microscopic disease. Further investigation is needed to establish the necessary confidence in these features in order to form a clinically viable diagnostic technique.

Appendix A

A.1 Participant Information Form

MONASH University

School of Physics

Participant Information Form

25 August 2005

Project Title:

Mapping of changes in the extra-cellular matrix during breast cancer progression.

Investigators:

Professor Robert Lewis, Dr Karen Siu, Ms Sabeena Sidhu,
School of Physics, Monash University
Telephone: 9905 3652

Dr Don McNaughton, Dr Bayden Wood
School of Chemistry, Monash University
Telephone: 9905 4525

Assoc. Professor Merrill Rowley, Assoc. Professor Tracey Brown, Dr Gary Brownlee
Department of Biochemistry and Molecular Biology, Monash University
Telephone: 9905 3760

Ms Jane Fox, Mr Stewart Hart, Dr. Beatrice Susil
Department of Medicine, Monash University
and Southern Health, Monash Medical Centre, Clayton
Telephone: 9594 2336

Dr Sarah Pearson, Mr Greg Falzon
School of Biological, Biomedical and Molecular Sciences,
University of New England, NSW 2351
Telephone: (02) 6773 2061

1. Your Consent

You are invited to take part in this research project.

This Participant Information contains detailed information about the research project. Its purpose is to explain to you as openly and clearly as possible all the procedures involved in this project before you decide whether or not to take part in it.

Please read this Participant Information carefully. Feel free to ask questions about any information in the document. You may also wish to discuss the project with a relative or friend or your local health worker. Feel free to do this.

Once you understand what the project is about and if you agree to take part in it, you will be asked to sign the Consent Form. By signing the Consent Form, you indicate that you understand the information and that you give your consent to participate in the research project.

School of Physics
PO Box 27
Monash University
Victoria 3800
Australia
Telephone +61 3 9905 3652 Facsimile +61 3 9905 3637
www.monash.edu.au
ABN 12 377 614 012 CRICOS provider number 00008C

You will be given a copy of the Participant Information and Consent Form to keep as a record.

2. Purpose and Background

One of the difficulties with diagnosing and treating cancer is that diseased tissue often looks very similar to normal tissue. The normal way to tell the difference is by taking a small tissue sample, a biopsy, and looking at it under a microscope. The process is quite complicated, can take several days, and requires a lot of expertise and care.

A team of researchers at Monash University, led by Professor Robert Lewis, are working together with your doctors to see if two new types of examination of tissue samples can tell the difference between cancerous and normal tissue. One examination uses X-rays and the other infrared light.

We hope it may be possible to simply shine a beam of X-rays through a sample of tissue and find out if it is cancer or not in a few minutes. The test may also help us find the exact edges of a tumour and understand how cancer spreads. Previous experience has shown that both the X-ray and infrared techniques are capable of distinguishing between healthy and cancerous tissues. To quantitatively assess whether these techniques have potential as diagnostic methods, we need to apply the techniques to much larger numbers of tissue samples. .

The special kinds of test will be done either at Monash University or for more detailed investigation at a facility called a synchrotron. At the moment Australia does not have a synchrotron but one is being built at Monash University and will be completed in 2007. Until then we will use synchrotrons overseas in Japan, the USA and Europe.

A total of up to 400 people are expected to participate in this project.

The results of this research may be used to help Ms Sabeena Sidhu and Mr Greg Falzon to obtain PhDs.

3. Procedures

To help us in this research, all we need is your permission to try the new test on some of the tissue that the surgeon will remove from you during your operation. This will not affect your treatment at all, since the tissue would have been removed anyway, but it will help us find out if the new test is useful or not. There will be no additional discomfort to that experienced during surgery. This research will not require you to commit further time to aid the study. To evaluate the test effectiveness we also need to have access to limited clinical information to compare with our X-ray and infrared results. This information will consist only of the histopathology test results alongside your patient code number (see point 7).

4. Collection of Tissue Samples for Research Purposes

By consenting to take part in this study, you also consent to the collection, storage and use of tissue samples. Only tissues already being removed by the surgeon will be used in this research. Thus only tissue being removed in the treatment of cancer, or that being removed as part of a cosmetic procedure, is requested, depending upon your situation. The latter will be used as control tissue. That is, it will be used as a source of normal tissue to compare with cancerous tissue. After collection of the X-ray and infrared data, the tissues will be returned to the hospital pathology department. They will be subjected to standard pathology tests and stored in this form according to normal hospital procedures.

5. Possible Benefits

The research does not offer you any direct benefits, but an increased understanding of how cancer spreads may eventually aid in the treatment of other patients. However, if an unexpected cancer is found by the extra histology, the results will be referred back to the treating surgeon by the pathologist, and you will be informed.

School of Physics
PO Box 27
Monash University
Victoria 3800
Australia
Telephone +61 3 9905 3652 Facsimile +61 3 9905 3637
www.monash.edu.au
ABN 12 377 614 012 CRICOS provider number 00008C

6. Possible Risks

There are no risks to you as the research cannot affect you directly, nor will it affect your medical treatment in any way.

7. Privacy, Confidentiality and Disclosure of Information

If you choose to participate in the project by signing the consent form, you will be identified only by a code number which your pathologist sends to the research team at the University so you will remain anonymous. Only the research team will have access to the coded data which will be securely stored for at least seven years as prescribed by the university regulations. The original information will be kept by your doctor in the same way as the rest of your medical records. All data will be confidentially destroyed at the end of the seven year period. We plan to publish the results of the project in peer reviewed scientific journals. Only the combined results of all patients will be published, such that no individual will be able to identified.

8. New Information Arising During the Project

During the research project, new information about the risks and benefits of the project may become known to the researchers. If this occurs, you will be told about this new information. This new information may mean that you can no longer participate in this research. If this occurs, the person(s) supervising the research will stop your participation. In all cases, you will be offered all available care to suit your needs and medical condition.

9. Results of Project

A successful outcome of the research would be the demonstration that the X-ray and/or infra red tests are able to distinguish between healthy and cancerous tissue with greater reliability and/or sensitivity than current techniques. The scientific outcomes of the project will be published in the scientific literature and will be made available to you upon request. There will be no financial reward for the investigators through the life of the research project.

10. Further Information or Any Problems

If you require further information or if you have any problems concerning this project, you can contact the principal researcher, Professor Rob Lewis, or any of the researchers listed at the beginning of this form.

11. Other Issues

If you have any complaints about any aspect of the project, the way it is being conducted or any questions about your rights as a research participant, you can phone 9905 2052. You can then ask to speak to the secretary of the Human Ethics Committee and tell him or her that the number of the project is 2003/021. You could also write to the secretary. That person's address is:

The Secretary
The Standing Committee on Ethics in Research Involving Humans
PO Box No 3A
Monash University, Victoria 3800
Telephone (03) 9905 2052 Fax (03) 9905 1420
Email: SCERH@adm.monash.edu.au

You can also contact:

Malar Thiagarajan
Executive Officer of the Southern Health Human Research Ethics Committee
Monash Medical Centre
Telephone (03) 9594 3025

You will need to tell Ms. Thiagarajan the name of one of the research team.

School of Physics
PO Box 27
Monash University
Victoria 3800
Australia
Telephone +61 3 9905 3652 Facsimile +61 3 9905 3637
www.monash.edu.au
ABN 12 377 614 012 CRICOS provider number 00008C

12. Participation is Voluntary

Participation in any research project is voluntary. If you do not wish to take part you are not obliged to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage.

Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect your routine treatment, your relationship with those treating you or your relationship with Monash Medical Centre.

If at any time you desire to withdraw tissue samples or data from the research project you have the right to do so.

Before you make your decision, a member of the research team will be available so that you can ask any questions you have about the research project. You can ask for any information you want. Please sign the Consent Form only after you have had a chance to ask your questions and have received satisfactory answers.

If you decide to withdraw from this project, please notify your breast care nurse before you withdraw. This notice will allow that person or the research supervisor to inform you if there are any health risks or special requirements linked to withdrawing.

13. Reimbursement for your costs

You will not be paid for your participation in this trial. However, you will be reimbursed for any costs that you incur as a result of participating in this trial. We anticipate that no such costs will arise.

14. Ethical Guidelines

This project will be carried out according to the *National Statement on Ethical Conduct in Research Involving Humans* (June 1999) produced by the National Health and Medical Research Council of Australia. This statement has been developed to protect the interests of people who agree to participate in human research studies. The Southern Health Human Research Ethics Committee of Monash Medical Centre and the Standing Committee on Ethics in Research Involving Humans of Monash University have granted approval for this research .

15. Injury

In the event that you suffer an injury as a result of participating in this research project, hospital care and treatment will be provided by the public health service at no extra cost to you.

16. Termination of the Study

This research project may be stopped for a variety of reasons. The research team reserves the right to terminate the study at any given time.

Thank you.

Prof. Robert Lewis
Monash University
Tel: 9905 3652

School of Physics
PO Box 27
Monash University
Victoria 3800
Australia
Telephone +61 3 9905 3652 Facsimile +61 3 9905 3637
www.monash.edu.au
ABN 12 377 614 012 CRICOS provider number 00008C

A.2 Informed Consent Form

Informed Consent Form

Project Title: Mapping of changes in the extra-cellular matrix during breast cancer progression
Monash University (Clayton), and
Southern Health (Monash Medical Centre, Clayton)

This Informed Consent Form is endorsed by the Department of Human Services, Victoria, as well as VMIA for the purpose of insurance coverage under the Department of Human Services Malpractice Policy.

I agree to take part in the above Monash University research project. I have had the project explained to me, and I have read the Participant Information Form, which I will keep for my records. I understand that agreeing to take part means that I am willing to:

- Allow samples to be taken from the tissue that will be removed from me during my surgery.
- Allow these samples to be investigated by specialized X-ray and infrared analysis.
- Allow these samples to be investigated by standard pathology techniques for comparison.
- Allow the researchers to have access to the results of these tests.

The tissue that I provide will be identified only by a code number and my identity will not be made known to the researchers.

I understand that any information I provide is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party.

I have chosen to take part in this research on the understanding that the research study will be carried out in a manner conforming to the principles set out by the National Statement on Ethical Conduct in Research Involving Humans, and further that:

1. I understand the general purposes, methods, demands and benefits and possible risks, inconveniences and discomforts of the study as outlined in the 'Explanatory Statement' that has been given to me.
2. Although I understand that the purpose of this research project is to improve the quality of care, it has also been explained that my involvement may not be of any personal benefit to me.
3. My participation in the research study is voluntary, and I am free to withdraw at any time, and to continue receiving appropriate treatment for my condition, as will be the case if I do not volunteer to enter the study.
4. I have been given the opportunity to have a member of my family or a friend present while the project was explained.
5. I have been given the opportunity to ask questions in relation to the research study, and I have received all the information and explanations I have requested.
6. If at any time during the study I need to obtain further information I am free to telephone any of the investigators on the numbers supplied on the explanatory form.
7. I understand that the confidentiality of my medical history will be safeguarded save where required by law. Although information regarding my medical history may be divulged and results of research may be published, my identity will not be revealed in such publication.
8. This research project has been approved by the Monash University Standing Committee on Ethics in Research Involving Humans and the Southern Health Human Research Ethics Committee.

Version 1.0
25 August 2005
Page 1/2

Signature …….. Date……

Name: ..... (please print)

Signature: ..... Date: .....

Independent witness to participant's voluntary and informed consent

I witness that the signature above is that of the participant, and I believe that she/he understands the above project and gives her/his consent voluntarily.

Name: .....(please print)

Signature: ..... Date: .....

Address:.....

Investigator*

I have fully explained the aims, risks and procedures of the above-named study to the person named herein.

Name: .....(please print)

Signature: ..... Date: .....

* A senior member of the research team must provide the explanation and provision of information concerning the research project.

Note: All parties signing the Consent Form must date their own signature.

Version 1.0
25 August 2005
Page 2/2

A.3 Ethics Approval Extensions

Southern Health

246 Clayton Road
Clayton 3168
Victoria
Australia

Postal address:
Locked Bag 29
Clayton Victoria 3168
Australia

Human Resear
Secretariat offic
tel (03) 9594 3
fax (03) 9594 E

22 December 2005

Prof R Lewis
School of Physics and Materials Engineering
PO Box 27,
Monash University
VICTORIA 3800

Dear Researcher,

Research Project Application No. 03094B : Mapping of extra-cellular matrix changes during breast cancer progression

At its meeting on 08 December 2005 2005, the Human Research Ethics Committee A received and approved your Annual Report in relation to the above project.

Additionally, as requested and in order to address the expected completion date of the project, an extension was granted to 30 September 2008.

Yours sincerely,

MALAR THIAGARAJAN
Executive Officer
Human Research Ethics Committee A

Southern Health

Dandenong Hospital
Kingston Centre
Cranbourne Integrated Care Centre

Monash Medical Centre
Casey Hospital

Community Health Services across the South East

KS

Southern Health

246 Clayton Road
Clayton, Victoria 3168
Australia

Postal address:
Locked Bag 29
Clayton South, Victoria 3169
Australia

tel 03 9594 6666
fax 03 9594 6111

RECEIVED
- 3 NOV 2008

Research Directorate
Level 4, Main Block
246 Clayton Rd, Clayton, 3168
Tel (03) 9594 4611
Fax (03) 9594 6306

22 October 2008

Prof R Lewis
School of Physics and Materials Engineering
Monash University
PO Box 27
CLAYTON
VIC 3800

Dear Researcher,

Research Project Application 03094B : Mapping of extra-cellular matrix changes during breast cancer progression

At its meeting on 16 October 2008, the Human Research Ethics Committee B received and approved your Annual Report in relation to the above project.

Additionally, as requested and in order to address the expected completion date of the project, an extension was granted to 30 June 2010.

Yours sincerely,

D. Dell

DEBORAH DELL
Administrative Officer
Research Directorate

Please quote Project No. and Title for all correspondence

Appendix B

Histopathology

This appendix was written with the help of Simon Nazaretian, who performed all of the histopathological diagnoses for the work presented in this thesis.

B.1 Histopathological Features for Diagnosing Breast Disease

As described in Chapter 2, conventional histopathology analysis was performed on all tissue samples following SAXS imaging. This appendix provides background information on the histopathological reviews conducted by Anatomical Pathologist, Simon Nazaretian (Monash Medical Centre, Southern Health), for the purposes of diagnosing the tissue samples used in this experiment. Although patient diagnostic information was also sourced from patient histories (surgical pathology reports, investigative biopsies, patient medical records), the variability of tissue samples taken in and around suspicious lesions warranted individual analysis of the SAXS tissue samples. The analysis described here is typical of that carried out in standard Australian clinical practice [19].

In this study, both neoplastic[§§§§§] and non-neoplastic breast disease were analysed. Non-neoplastic breast disease includes a number of disorders that may give rise to a palpable mass or a mammographic abnormality. Because non-neoplastic disease is difficult to diagnose from mammography alone, surgical removal or biopsy of the mass is usually performed. Fibrocystic change involves cystic dilation, as well as stromal fibrosis, and there is usually associated apocrine metaplasia[******], adenosis of ducts and acini[††††††] and epithelial hyperplasia present (see Figure B.1). The cause of fibrocystic change in the breast is not well understood, however it has been associated with hormonal activity and genetics [21]. The menstrual cycle can cause variations in the breast tissue, due to hormonal changes occurring during this time [48]. Adenosis refers to a benign proliferation of glands that may be combined with fibrosis of the stroma to give rise to sclerosing adenosis (Figure B.2) which is difficult to distinguish from invasive carcinoma. The

[§§§§§] Neoplastic: an abnormal growth of tissue that can be benign or malignant.
[******] Apocrine Metaplasia: a type of fibrocystic disease of the breast involving transformation of acinar epithelium of breast tissue.
[††††††] Acini: a small saclike dilation within the breast.

most common benign neoplasm of the breast is the fibroadenoma (see Figure B.3). This is a localized proliferation of breast ducts and stroma and occurs most frequently in women aged 25-35 [118]. Another type of benign lesion is ductal hyperplasia which involves the epithelial layer of the ducts without proliferation of the glands (see Figure B.4). Atypical ductal hyperplasia (ADH) is an intermediate condition between hyperplasia and carcinoma *in situ* and has a small but significantly increased risk of developing into carcinoma. Although considered benign disease, there are relative risks of developing into subsequent breast cancer. Benign lesions can be categorised as non-proliferative, proliferative without atypia[‡‡‡‡‡] and proliferative with atypia. Various studies have shown non-proliferative disorders like cysts, apocrine change and mild epithelial hyperplasia do not show increased risk of subsequent carcinoma development. The other categories that include moderate and florid epithelial hyperplasia, intraductal papilloma, sclerosing adenosis, fibroadenoma, and atypical ductal hyperplasia (ADH) or atypical lobular hyperplasia (ALH), do show increased relative risks [115]. With the exception of ADH and sclerosing adenosis, these non-neoplastic conditions have little or no increased risk of malignancy. Benign breast disorders may be difficult to differentiate clinically from carcinoma and may occur concurrently with, but independently of, invasive carcinomas [118].

Neoplasia is considered to be a process in which normal cells change to abnormal or atypical cells which then multiply out of control, which is called dysplasia. Severe dysplasia is termed as carcinoma in situ, when the cells cytologically appear malignant. Examples of dysplasia are ductal carcinoma *in situ* (DCIS) (Figure B.5) and lobular carcinoma *in situ* (LCIS) (lobular carcinoma is seen in Figure B.6), which have not invaded through the basement membrane (i.e. not invasive disease or metastasised). Most are adenocarcinomas arising from the epithelium of either the mammary lobules (lobular carcinoma) or the mammary ducts (ductal carcinoma). The development of invasive breast cancer apears to be preceded by carcinoma *in situ* in which the malignant cells proliferate within the mammary ducts or lobules but do not breach the basement membrane (ductal or lobular carcinoma *in situ*) [47, 118]. DCIS usually has calcifications present in the tissue, which can also be seen on mammography.

[‡‡‡‡‡] Atypia: clinical term for atypical cell.

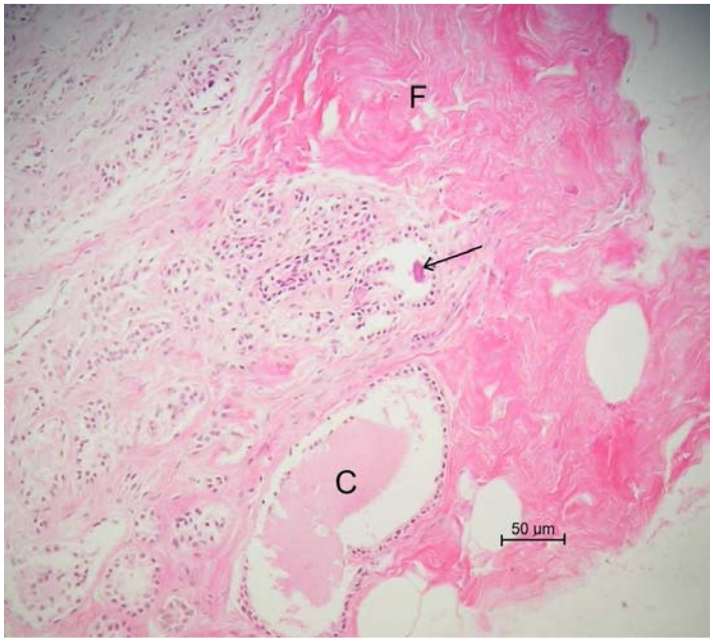


Figure B.1: Fibrocystic change with partly dense stromal fibrosis (F) and microcyst formation (C) as well as focal calcification (arrow). Fibrocystic change also may include large cyst formation and proliferating ductal epithelium. The epithelium usually shows a variable pattern of apocrine metaplasia and occasional papillary growth pattern. There is no cellular atypia [115].

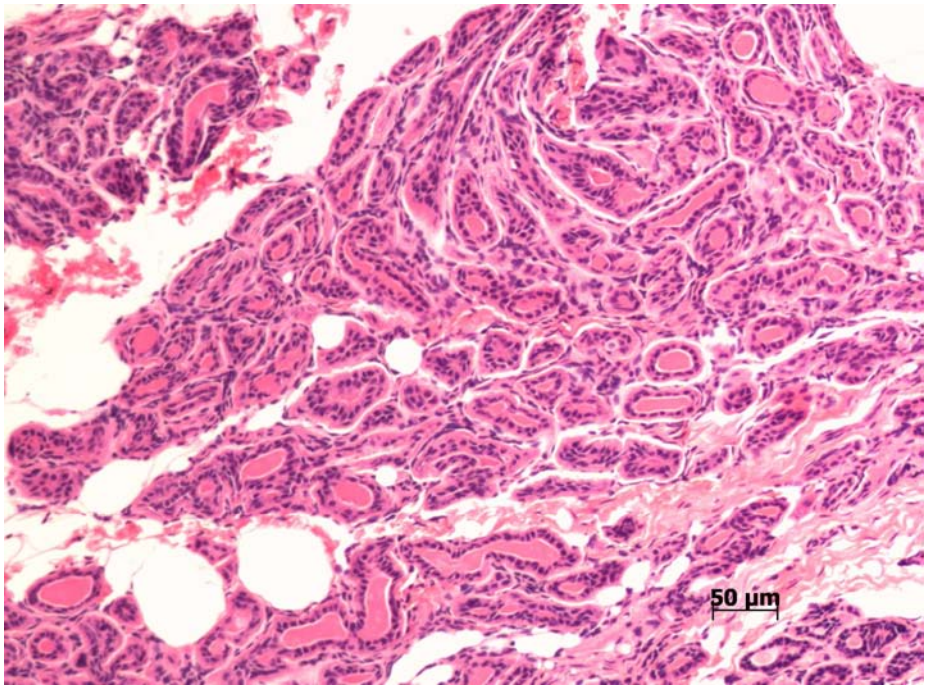


Figure B.2: Adenosis with compact proliferation of small regular glands showing both luminal ductal epithelium and outer myoepithelial cell layers. On low power examination, retained lobular architecture can be appreciated. [115]

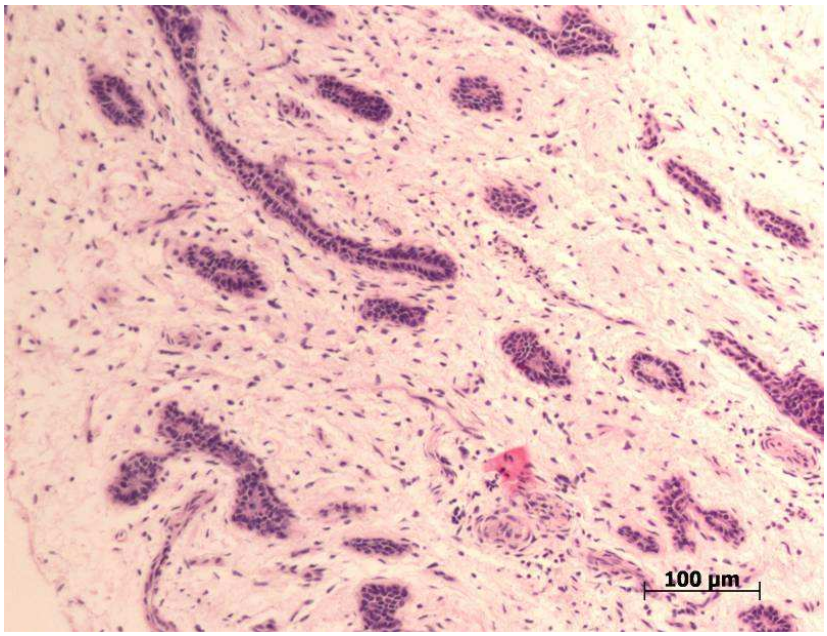


Figure B.3: Fibroadenoma with proliferating stroma which is loose and myxoid and consists of spindled cells compressing slit-like ducts. There is no proliferation or cytological atypia seen in this example.[115]

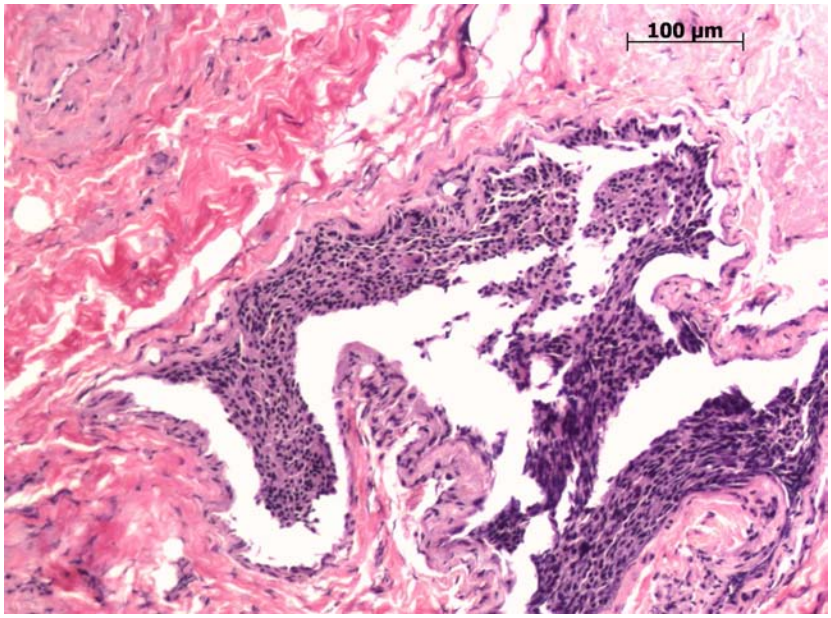


Figure B.4: Ductal (or Epithelial) Hyperplasia of usual type shows ductal epithelial cells proliferating and distending ducts. Cells are oval, show overlapping and occasional slit-like spaces. There is no cytological atypia. Epithelial hyperplasia is a proliferative intraductal lesion and moderate or florid types are associated with increased risk for subsequent development of invasive carcinoma. In this example, artefact complicates assessment but there is some suggestion of cellular monotony and architectural rigidity and this raises the suspicion of low grade ductal carcinoma in situ. [115]

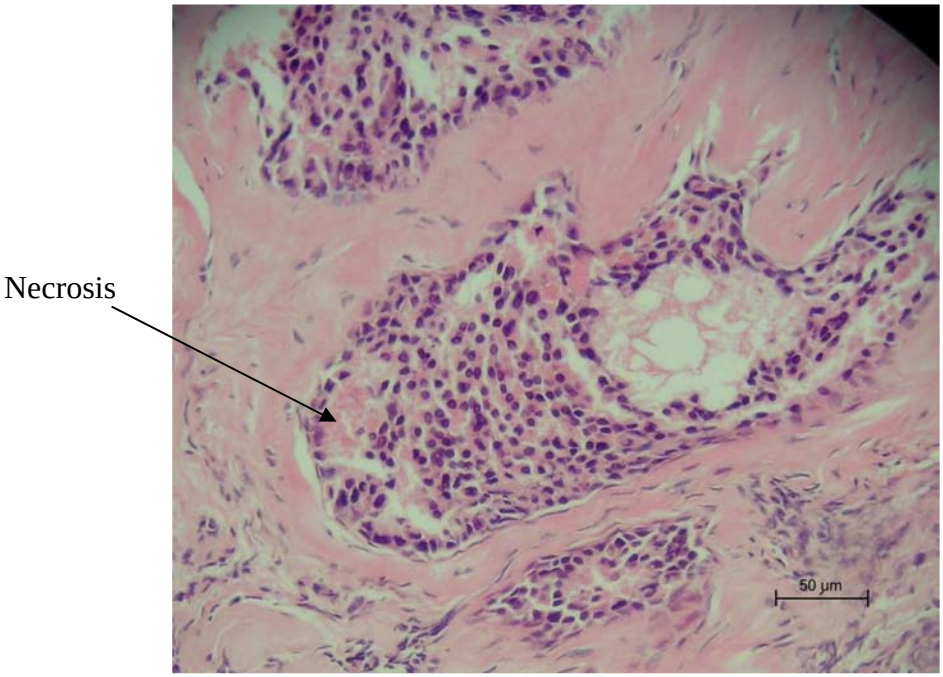


Figure B.5: Ductal Carcinoma *In Situ* (DCIS) with a cribriform pattern of solid spaces within the duct lined by mildly atypical epithelium. A few spaces contain necrotic debris (arrow). DCIS is a heterogeneous group of proliferations showing atypical cytological and architectural features. Growth patterns are characterised as solid, cribriform, comedo, micropapillary or papillary. Different histological patterns often coexist within the same lesion.[115]

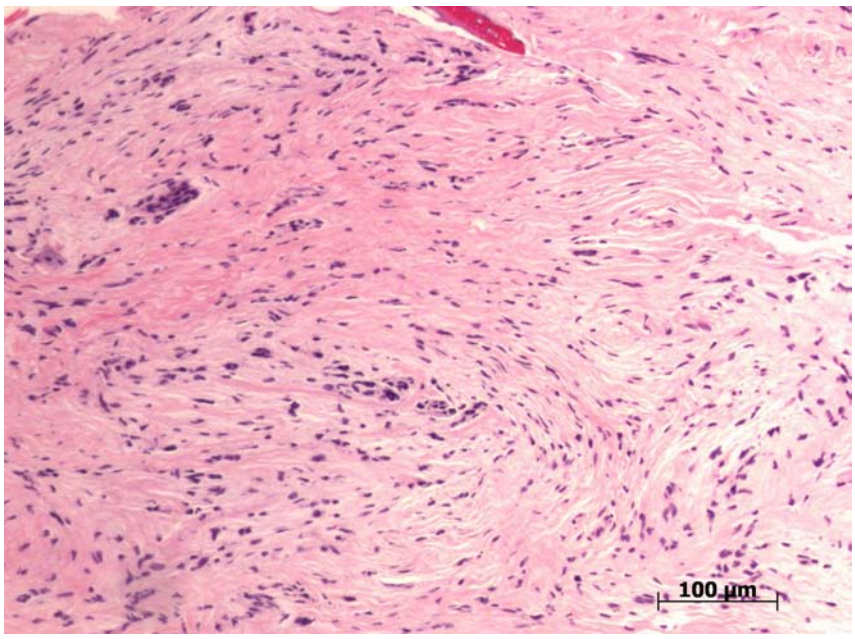


Figure B.6: Invasive lobular carcinoma is composed of non-cohesive cells individually dispersed or arranged in single-file linear patterns in fibrous stroma. The lobular carcinoma cells often surround non-malignant ducts forming targetoid lesions. Lobular carcinoma in situ may accompany lobular carcinoma [115].

B.2 Histopathologic Grading

Pathologists often assign a histological grade to a patient's cancerous breast tumour to identify the type of tumour present and to help determine their patient's prognosis [19]. The Elston and Ellis modification of the Bloom and Richardson grading system (BRE), is the grading system used by the pathologists at the Monash Medical Centre. Using the BRE system of classification [12, 109], three features are used in determining the grade of a cancer: (1) the mitotic frequency score, (2) tubular formation score (the percentage of cancer composed of tubular structures), and (3) nuclear score (the change in cell size, coarseness of chromatin and nucleolar prominence). Each of these characteristics is assigned a score ranging from 1 to 3. The scores of each of the cells features are then added together for a final sum that ranges between 3 and 9. Table B.1 summarizes how the BRE system of grading is used to classify tumour types. Pathologists also look for necrosis when determining tumour grade. Typically, cancers presenting with a high BRE grade and necrosis close to the surrounding margin of a breast tissue sample, or samples with a large areas of DCIS (ductal carcinoma *in situ*), are more likely to recur after treatment compared to other breast cancers [2, 16]. Thus, it is critical that accurate grading of the cells be determined so that proper treatment and patient counselling can be administered accordingly [19].

Out of the 543 tissue samples imaged, 178 tissue samples were diagnosed with invasive carcinoma. Due to the poor state of the tissue at the time of histopathological analysis (see Section 2.5), it was very difficult to assign a pathological grade to each tissue sample. Freezing and defrosting the tissue samples attributed to some cellular damage which made staining of the tissues difficult to process for identifying the diagnostic features used for histopathological analysis. Pathological grade could be assumed equivalent to that reported from the surgical pathology analysis. However, only 45 patients (out of the 130 patients in total) had their invasive carcinoma grade available for recording from their medical records and only 37 patients had tissue samples with confirmed disease (8 patients had their primary site cores diagnosed as benign or normal). Table B.2 lists the patients and their reported pathological grade of their carcinoma. Due to the difficulty in grading the tissue samples after SAXS imaging and the limited population size of those patients with data available, analysis regarding tissue grade did not have enough discriminatory power to pursue. However, all of the tissue samples diagnosed with invasive carcinoma were grouped together for analysis, irrespective of carcinoma grade.

Table B.1: The Bloom and Richardson grading system (BRE) for breast cancers (Page 26 of [19]).

Grade	Description	Score
Grade 1 (lowest)	Well-differentiated breast cells; cells generally appear normal and are not growing rapidly; cancer arranged in small tubules	3,4,5
Grade 2	Moderately-differentiated breast cells; have characteristics between Grade 1 and Grade 3 tumours	6,7
Grade 3 (highest)	Poorly-differentiated breast cells; cells do not appear normal and tend to grow and spread more aggressively	8,9

Table B.2: The 45 patients who had their carcinoma grade available from their medical records. Patient numbers represent the de-identified values assigned for the patients. The diagnosis was that determined by Simon Nazaretian from the SAXS tissue samples. C = invasive carcinoma, B = benign, N = normal tissue. (The Grade was not determined by Simon Nazaretian, but from the patient medical records.)

Patient	Diagnosis	Grade
3	C	1
7	C	1
24	C	1
39	C	1
74	C	1
126	C	1
152	B	1
171	N	1
187	N	1
1	C	2
31	C	2
35	C	2
42	C	2
47	C	2
50	C	2
51	C	2
70	C	2
71	C	2
72	C	2
78	C	2
82	C	2
104	C	2
113	T	2

Patient	Diagnosis	Grade
122	C	2
148	C	2
149	C	2
160	C	2
163	C	2
174	C	2
181	C	2
127	B	2
139	N	2
180	N	2
53	C	3
83	C	3
87	C	3
143	C	3
147	C	3
151	C	3
153	C	3
154	C	3
168	C	3
185	C	3
155	N	3
170	B	3

Appendix C

FujiFilm Imaging Plate (IP) Detectors

Imaging plates (IP) have been shown experimentally to satisfy the optical requirements for X-ray detectors for synchrotron experiments [86]. The IPs used for acquiring SAXS images with the laboratory source were Fujifilm BAS-MS2025 IPs (Fuji Photo Film Co., Ltd. Tokyo, Japan). The IP was an excellent detector to use for obtaining diffraction images from the Osmic SAXS camera since a large dynamic range and high spatial resolution was required to obtain images with the laboratory source.

The IP is a flexible plastic plate which was coated with bunches of very small crystals (grain size ~5 µm) of photo-stimulable phosphor ($BaF(Br,I):Eu^{2+}$) by using an organic binder (see Figure C.1 for the composite structure). When the IP is exposed to X-rays, some of the electrons in the valence band are excited to the conduction band of the phosphor crystals (i.e. ionization of Eu^{+2} to Eu^{+3}). The electrons are trapped in the doped Br^- and F^- vacancies, forming temporary color centres. Exposure to visible light again excites the trapped electrons so that they generate energy for luminescence, while returning to the valence band of the crystal. The release of the stored energy within the phosphor by stimulation of visible light, is the photostimulated luminescence (PSL). Because the response time of the PSL is very short (in the µs range), it is possible to read an X-ray image with a speed of 5-10 µs per pixel with high efficiency. The wavelength of the PSL ($\lambda \approx 390$ nm) is reasonably separated from that of the stimulating light ($\lambda = 632.8$ nm), allowing it to be collected by the photo-multiplier tube (PMT). The residual image on the IP is erased by irradiation with visible light, to allow for repeated use.

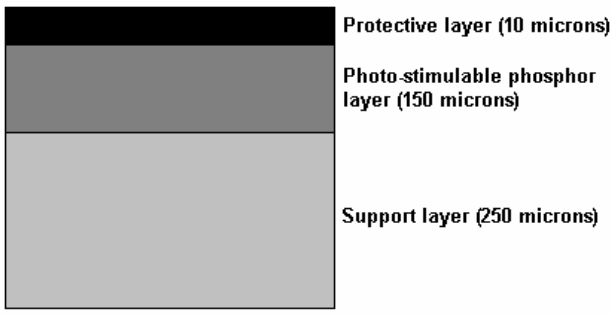


Figure C.1: Composite structure of the Imaging Plate [3]

C.1 Characteristics

The BAS-MS2025 white IPs of dimension 20 cm x 25 cm were used with the BAS5000 scanning system (FUJI Photo Film Co. Ltd. Science Systems, Tokyo, Japan 2004).

The background noise level of the IP usually corresponds to a signal level of less than 3 X-ray photons/(100 μm)^2 [3]. The saturation of the relative uncertainty of the IP results from 'system fluctuation noise'. The origins of the system fluctuation noise are non-uniformity of absorption, non-uniformity of the color-centre density, fluctuation of the laser intensity, non-uniformity of PSL collection, and fluctuation of the high-voltage supply to the photo-multiplier tube (PMT) [88]. The ultimate precision in intensity measurements with the IP was of the order of 1%.

The dynamic range of the IP is of the order of 10^5 [4]. The response of the PSL is linear over a dynamic range of < 10^3/100 μm^2, with an error rate of less than 5% [3]. Figure C.2 shows the dynamic range of the PSL imaging plate in comparison with the dynamic range for X-ray film as a comparison [4].

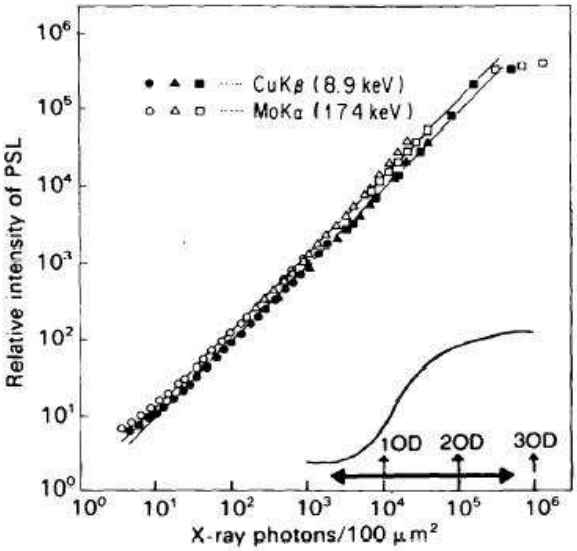


Figure C.2: The dynamic range of the PSL of the Imaging Plate. The dynamic range of typical high-sensitivity X-ray film is also shown. (Printed with permission from the IUCr, 2009 [3]).

The spatial resolution of the IP is 170 µm FWHM of the line-spread function (LSF) (when using a Resolution (R) = 100 µm). The spatial resolution was determined by the laser-light scattering in the phosphor during the readout. The laser-light scattering originates from a mismatching of the refractive indices at the boundaries of phosphor crystal grains [4].

The BAS system generates image data as a logarithmic compression and the software used to process the images (PCDetPak32 6.0, R. Lewis, Monash University, 2007) made the necessary log to linear conversion on the BAS-5000 system to obtain the PSL according to the equation:

$$PSL = \left(\frac{\mathrm{Re}\,s}{100}\right)^2 \times \frac{4000}{S} \times 10^{L\times\left(\frac{QL}{G}-\frac{1}{2}\right)} \qquad \text{(C.1)}$$

where the PSL is in a linear scale, *Res* is the resolution in µm, *S* is the sensitivity, *L* is the latitude[§§§§§§], *G* is the gradation (either 8 bit or 16 bit), and *QL* is the quantum level which represents the pixel value in the digital image file. A header file, which contains the values used for scanning in equation C.1 as well as the date and time the image was scanned, and a binary digital image file, which contains the pixel data and image information, are created when the IP is scanned. However, if *QL = 0*, then the PSL is defined as being zero. In this experiment, a 16 bit gradation was used for all image outputs, thus, the range of *QL* is from 0 to 65535. The latitude was selectable during readout (*L* = 4 or 5), which dictated the dynamic range of the IP during scanning (controls the slope of the amplifier response): in this experiment, an *L* = 5 was used so a higher order of magnitude could be utilized. There were three sensitivity options available when using *L* = 5: 4000, 10000, and 30000. According to Equation C.1, increasing the sensitivity decreases the PSL. The sensitivity controls the pixel read amplifier gain on the Fuji system pixel readout amplification. If S is too small, the PSL values will not be over a wide dynamic range, so separation of similar intensities cannot be achieved. For the purposes of this experiment, an S = 30000 was used for all IP readouts. This value was adequate for obtaining the required diffraction data from each image and required a reasonable exposure time, providing efficiency in the experiment. Table C.1 summarises the IP readout settings used in this experiment.

[§§§§§§] It is important to note that the PSL value was independent from changing the sensitivity or latitude parameters. Since the radiation dose to the IP was the same, the PSL value was always the same, even if the sensitivity or latitude was changed.

Table C.1: Readout settings used for the BAS-5000 when reading IPs

Characteristic	Setting
Resolution	200 μm
Sensitivity	30000
Latitude	5
Gradation	65535 (16 bit)

One disadvantage of using IPs was the effect of fading (due to thermal stimulation), where the image stored in the IP decreases with time after exposure to X-rays. The fading rate depends on temperature: higher temperatures increase the rate of fading. The IPs were kept in an opaque bag to protect them from temperature variations in the environment. However, fading does not depend on the exposure level or the X-ray photon energy of the image [4]. Figure C.3 shows the fading of an IP as a function of time for two different energies. This decay of the latent image on the IP with the time between exposure and reading needs to be considered in the calibration curves for any delay times which were used in the experiment. A delay of 1 hour between exposure and readout time introduced a difference of 1% within a 5 minute interval [52]. However, at 5 minutes after exposure, a time difference of 1.25 minutes introduced a difference of 1%. If less delay is required, the uncertainty due to time differences was greater. Floyd *et al*. [52] suggested that plate fogging was not a problem when less than 4 hours were elapsed between erasure and readout. It would seem then, that the plates should be read between 5 minutes and 4 hours. For this experiment, the time delay between erasure and readout was less than 2 hours for each IP.

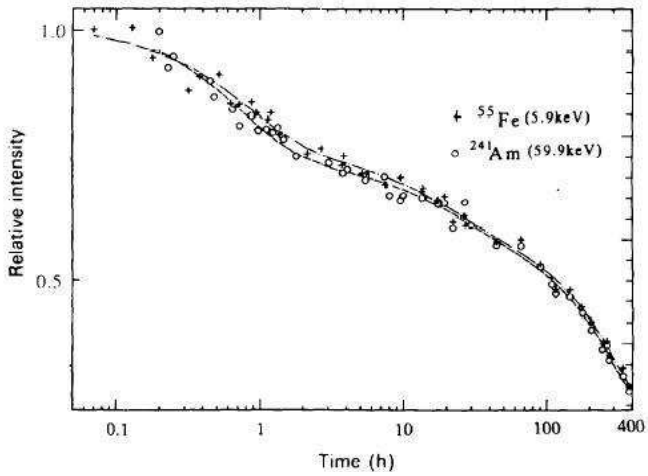


Figure C.3: Fading of the IP signals as a function of time with two different X-ray energies (5.9 and 59.9 keV). Temperature: 293 K. (Printed with permission from the IUCr. [3].

Appendix D

Design of Semi-Transparent Beam-Stop for Laboratory SAXS Camera

A beam stop is typically inserted between the sample and the detector in order to protect the detector from the direct X-ray beam. A semi-transparent beam stop allows for measurement of the transmitted intensity of the beam (incident beam intensity – absorbed intensity due to the sample). Knowing the transmitted intensity of the beam allows for comparisons between each tissue since the variations in absorption due to the thickness of the tissue sample, variations due to electron beam decay, variations in the downstream optical elements, and glass capillary absorption can be corrected. The laboratory SAXS camera did not have a semi-transparent beam stop that could be used to measure the transmitted intensity, so one had to be designed and tested for use. In order to determine the appropriate dimensions needed for optimum performance and data acquisition with Imaging Plates (IP) (See Appendix C), two parameters had to be established: (1) the diameter of the photon beam to determine the dimensions of the beam stop, and (2) the thickness of the semi-transparent beam-stop so the transmitted intensity of the beam can be calculated. The former is explained in D.1: Determination of the Diameter of the Beam, and the latter is explained in D.2: Calculating the Appropriate Thickness of the Semi-Transparent Beam Stop.

D.1: Determination of the Diameter of the Beam

An image of the beam was taken using a Photonic Science CCD Camera (XDI, Photonic Science Ltd., Millham, UK) (http://www.photonic-science.co.uk/PDF/xdi.pdf). The intensity through the centre of the bright spot, both horizontally and vertically were averaged and fitted with a Gaussian function (shown in Figure D.1), where the Full Width at Tenth Maximum (FWTM) was calculated to be 1.87 mm.

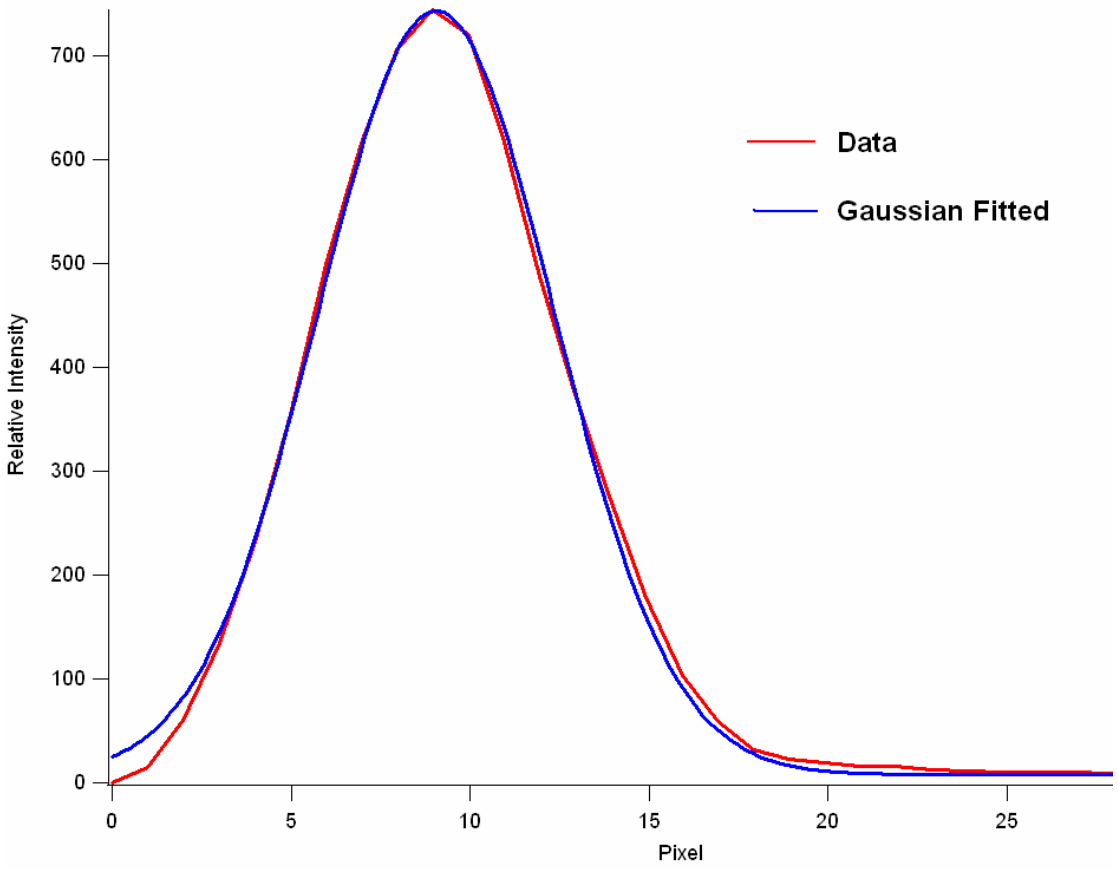


Figure D.1: Radial profile of spot imaged with the CCD camera with the raw data (in red) fitted with a Gaussian function (in blue).

According to the SAXS Osmic Camera Manual, the FWHM (Full Width at Half Maximum) and the FWTM at the detector are stated to be 1.38 mm and 2.14 mm, respectively. The measurement made from the above fitting was smaller than this (FWTM = 1.87) which could be due to the position of the CCD camera used for our calculation, which was closer to the sample (by ~5 cm) than where the detector position is located (this is where the manufacture value was measured). The position of the beam stop is ~5 cm from the detector, so the FWTM we calculated is more appropriate to use as an estimation of the spot size. A minimum window diameter of 3.0 mm (an extra 0.58 mm on either side) was used to compensate for possible positioning errors so that the entire beam goes through the semi-transparent part.

D.2: Calculating the Appropriate Thickness of the Semi-Transparent Beam-Stop

Figure D.2 shows the design of the semi-transparent beam stop used with the laboratory source. In order to calculate the transmitted intensity of the beam using a semi-transparent beam stop, a low enough intensity so the Imaging Plate (IP) detector would not saturate, but a high enough intensity to provide good statistics (signal-to-noise ratio) for normalisation is ideal. Varying thicknesses of aluminium attenuator (between 0.1-1.0 mm) can be placed in a notch made in the beam stop cylinder. Figure D.3 shows a plot of the Photo-stimulated Luminescence (PSL) with varying thicknesses of aluminium. All exposure times were 5 minutes in duration. Four thicknesses of aluminium were examined in order to find the optimum thickness for calculation of the transmitted intensity of the beam: 480 μm, 560 μm, 640 μm, and 720 μm. The fitted line was established with using Excel's trendline function to obtain the relationship between the PSL and the thickness of aluminium.

$$PSL = (166875)e^{-0.015x} \qquad \text{(D.1)}$$

For aluminium, the coefficient in the exponent is 0.01864 (m^{-1}) for 8.05 keV [70].

Judging from the plot shown in Figure D.3, a semi-transparent beam-stop of aluminium >700 μm thick was sufficient in absorption of the beam so saturation of the IP did not occur (saturation occurs when the PSL = 126). If a Quantum Level (QL) = 32767 (half saturation) is taken, this corresponds to a PSL = 0.4, in which case, a thickness of > 980 μm is needed. Therefore, variations for the thickness of the aluminium semi-transparent beam stop were made between 700 and 1000 μm in the design of the beam-stop. For the purposes of this experiment, an aluminium thickness of 800 μm was used for calculating the transmitted intensity of the laboratory source. This thickness did not allow saturation to occur on the IP during the 60 minute exposures, but did allow for calculation of the transmitted intensity at a high enough PSL for meaningful measurement.

$$PSL = \left(\frac{\mathrm{Re}\,s}{100}\right)^2 \times \frac{4000}{S} \times 10^{L\left(\frac{QL}{G}-\frac{1}{2}\right)} \qquad \text{(D.2)}$$

Res = 100 (μm) S = 10000 L = 5 G = 65535

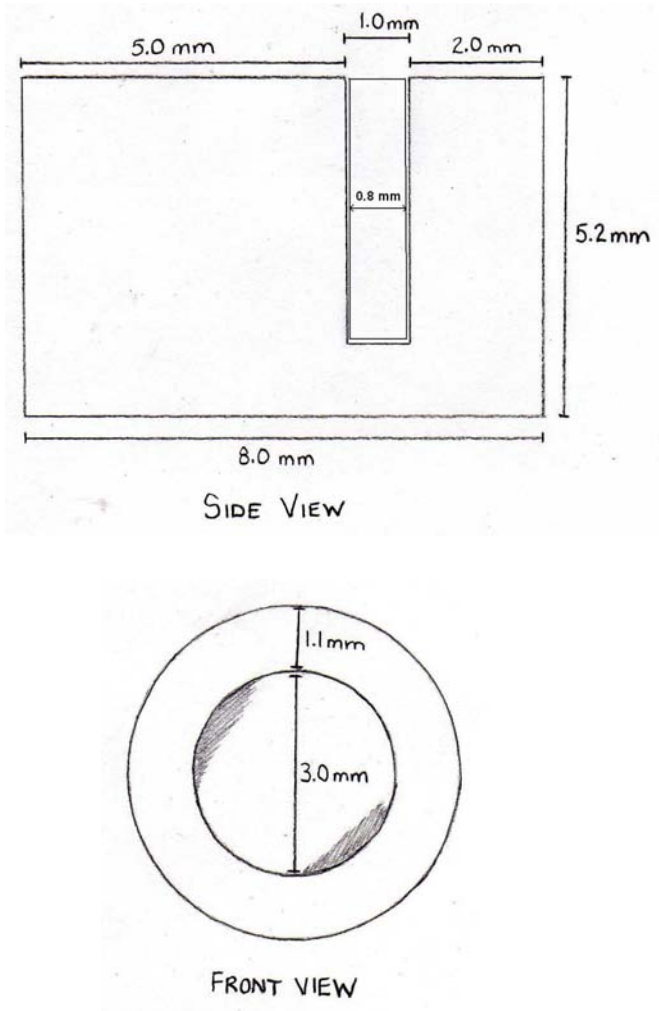


Figure D.2: Design schematic of the semi-transparent beam stop used with the Laboratory SAXS Camera (manufactured by the Physics Machine Shop, Faculty of Science, Monash University). The beam stop was cylindrical in shape and composed of lead, where the centre 3.0 mm was hollowed out (front view). A 1.0 mm slit was cut out 5.0 mm from the incident beam side entry to allow for an aluminium sheet, 0.8 mm thick, to be placed in order to block the incident beam and calculate the absorbed intensity.

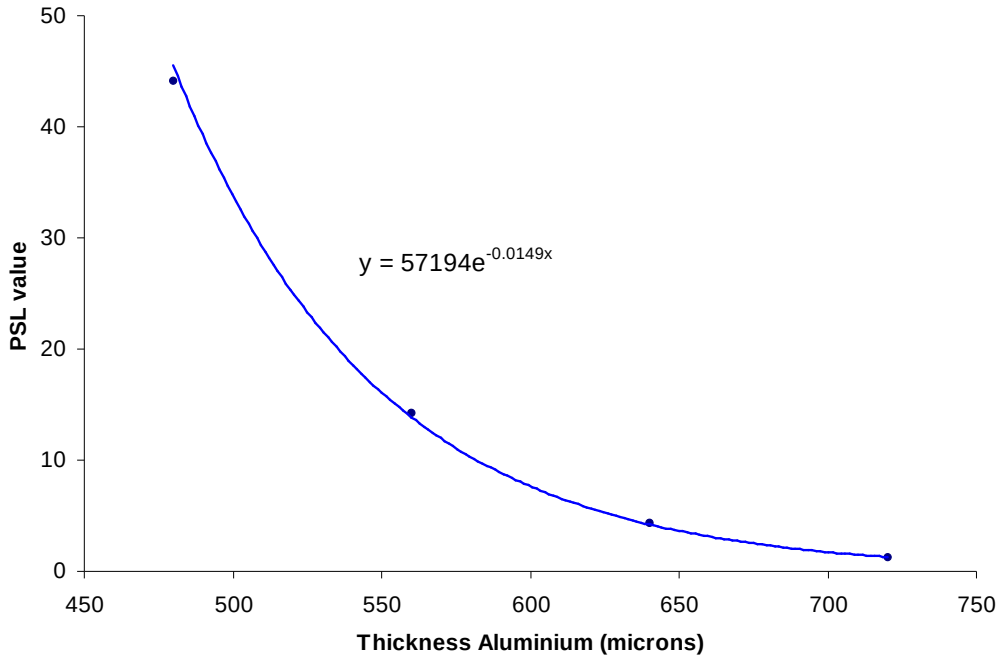


Figure D.3: The average PSL with varying thickness of Aluminium. All data was taken with an exposure time = 5 minutes, Res = 100 μm, S = 1000, and L = 5. Saturation occurs when PSL = 126 (QL=65535).

Appendix E

Statistics

This appendix was written with the help of Dr. Greg Falzon, who also assisted with the analysis presented in Chapters 4-6 of this thesis.

E.1 Student t-test and p-values

The t-test was used to infer differences in the means of two populations. The null hypothesis (H_o) means that there is no difference in population means, compared to the alternative hypothesis (H_A), which describes a difference in population means. The hypotheses are specified mathematically as:

The null hypothesis H_o: $\mu_1 = \mu_2$

The alternative hypothesis H_A: $\mu_1 \neq \mu_2$

where the sample means ($\bar{x}_1$, $\bar{x}_2$) of each data set were compared in order to test whether or not the groups had equal population means (μ_1, μ_2). Differences or similarities between the data sets were determined by calculating a test statistic (t_s) and a corresponding p-value. The test (t-statistic) was calculated from the data using:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{(E.1)}$$

where s represents the standard deviation and n represents the number of samples. This statistic describes a general type of two-sample t-test for which the population variances (σ_1^2, σ_2^2) are not required to be equal. The test weights the differences in the sample means and makes these weights proportional to both of the sample variances and degrees of freedom. If the null hypothesis is true, then the t-statistic obeys a Student t-distribution that has η degrees of freedom [42]. The Student t-distribution is specified as:

$$f(t) = \frac{\Gamma\left\{\frac{1}{2}(\upsilon+1)\right\}}{(\upsilon\pi)^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\upsilon\right)}\left(1+\frac{t^2}{\upsilon}\right)^{\frac{1}{2}(\upsilon+1)} \qquad \text{(E.2)}$$

and the degrees of freedom calculated by [42]:

$$\upsilon = \left[\frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \right]^{-1} \tag{E.3}$$

$$c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}} \tag{E.4}$$

A hypothesis test can then be performed by calculating a p-value. A p-value is the probability of obtaining the data (or the data with a more extreme departure from the null-hypothesis) under the assumption that the null hypothesis is true [42]. For a t-statistic obeying a Student t-distribution, it is calculated as [73]:

For x > 0,

$$P(t_\upsilon \le x) = \frac{1}{2}\left[P\left(Y \le \frac{x^2}{n + x^2} \right) + 1 \right] \tag{E.5}$$

where Y is a $\beta\left(\frac{1}{2}, \frac{\upsilon}{2}\right)$ distributed random variable and $P(t_\upsilon \le x) = 1 - P(t_\upsilon \le -x)$. The p-value measured the compatibility between the data and the null hypothesis, H_o: a large p-value (close to 1) indicated a value of t_n near the centre of the t-distribution (compatible with H_o), whereas a small p-value (close to 0) indicated a value of t_n in the far tails of the t-distribution (incompatible with H_o).

Significance levels (or threshold values) were chosen to represent the cut-off values of the p-value to which H_o was accepted or rejected. The value chosen for the analysis of this project was p-value = 0.05. It is important to note that the p-value and t-test were only valid if the data was normally distributed about the mean. The test also assumes the data was sampled independently and randomly from the populations under study, which can be assured by careful experimental design. Tests for normality were performed on the data and are described in Sections E.3, E.4, and E.5.

E.2 Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) was used for analysing the effect of categorical factors for the data sets examined in this project. This model partitions the total variance of a data set into identifiable components attributable to particular factors (for instance, partitioning the data according to the tissue pathology). In this project,

a single factor design was used, where *N* subjects were attributed to *k* different levels of a factor [43]. In this case, the subjects were assumed to associate with the factor levels randomly. The model was specified in factor-effects form[89]:

$$X_{i,j} = \mu + \tau_i + \varepsilon_{i,j} \qquad \text{(E.6)}$$

for parameters $i = 1, \ldots, k$ (level indicator) and $j = 1, \ldots, n_i$ (per level sample size) and responses $X_{i,j}, \mu = \sum_{i=1}^{k} \mu_i / k$ is the un-weighted treatment mean for the *k* levels, $\tau_i = \mu_i - \mu$ describes the treatment effect[*******], $\varepsilon_{i,j}$ are independent modelling errors that are each assumed to obey a normal distribution $\mathrm{N}(0, \sigma^2)$.

The model is based upon the following assumptions [89]:

i. The set of samples in each level of the factor obeys a normal probability distribution.
ii. All of these probability distributions have the same variance.
iii. The data was sampled at random.
iv. The sample response was independent of all other levels for each factor level.

The validity of the ANOVA model was verified prior to analysis by checking the model assumptions. Tests regarding the normality (Kolmorogrov-Smirnov (Section E.4) and Shapiro-Wilk tests (Section E.5)) and homogeneity of variances (Bartlett's Test (Section E.3)) of the data were used to assess the first two assumptions above. The assumptions of random sampling and independence were justified from a careful experimental design in each application. The ANOVA model was estimated using the maximum-likelihood technique, which is equivalent to least-squares minimisation in the analysis of this thesis, and the model estimates are given by [89]:

$$\hat{\tau}_i = \overline{X}_i \qquad \text{(E.7)}$$

that is the average of all of the data corresponding to the *i*th level. The residuals[89]:

$$\varepsilon_{i,j} = X_{i,j} - \hat{X}_{i,j} = X_{i,j} - \overline{X}_i \qquad \text{(E.8)}$$

were then assessed using diagnostic plots (standardised residual time series, normal quantile plots) after the model was fit to ensure that they obeyed the model assumptions of independence and normality.

A statistical test for the equality of factor level means could then be performed [89]:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0$$

$$H_A : \text{ not all of the } \tau_i \text{ equal zero} \qquad \text{(E.9)}$$

[*******] Treatment effect: the difference of the mean of the *i*th factor level and the overall mean.

The F-test statistic defined as [89]:

$$F^{*} = \frac{MSTR}{MSE} \tag{E.10}$$

where the 'mean-squares of the treatments' is

$$MSTR = \frac{SSTR}{k-1} = \frac{\sum_{j=1}^{n_i}\left(\overline{X}_i - \mu\right)^2}{k-1} \tag{E.11}$$

and the 'mean squares of the errors' is

$$MSE = \frac{SSE}{n_T - k} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{i,j} - \mu\right)^2 \tag{E.12}$$

for a total sample size of n_T. This statistic ranges in magnitude from 0 to ± ∞, values of the F statistic close to unity support the null hypothesis (H_o) because the mean squares of the treatments is then of similar size to the mean squares of the errors. In contrast, large magnitudes of the F statistic support the alternative hypothesis (H_A). When the null hypothesis holds, the F statistic is described by the central F-distribution [73, 89]:

$$F\left(x \mid k-1, n_T - k\right) = \left\{\frac{\Gamma\left(\frac{(k-1)+(n_T-k)}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{n_T-k}{2}\right)}\right\}\left\{\frac{\left(\frac{k-1}{2}\right)^{\frac{(k-1)}{2}} x^{\frac{(k-1)}{2}-1}}{\left(\frac{(n_T-k)}{2}\right)^{\frac{(k-1)}{2}}\left[1+\frac{(k-1)x}{(n_T-k)}\right]^{\left(\frac{(k-1)}{2}+\frac{(n_T-k)}{2}\right)}}\right\}$$

subject to *(k-1) > 0* $(n_T - k)$ > and *x > 0*. (E.13)

Using the F statistic and the F-distribution, the p-value of the hypothesis test was calculated [89]:

$$P\left(F_{(k-1),(n_T-k)} \le x\right) = P\left(Y \le \frac{(k-1)x}{(n_T-k)+(k-1)x}\right) \tag{E.14}$$

where Y is a $\beta\left(\frac{(k-1)}{2}, \frac{(n_T-k)}{2}\right)$ distributed random variable.

If the evidence suggested a difference in treatment means, then further investigations were conducted to determine precisely which treatments were significant.

E.3 Bartlett's Test

The calculation of p-values in the ANOVA models assumed that the variances were equal across the data groups analysed. The Bartlett's test for homogeneity of variance was used to justify this assumption. The null hypothesis of equal population variances was tested against the alternative hypothesis [70]:

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$$

$$H_A \text{ : the variances are not all equal} \qquad \text{(E.15)}$$

for k independent samples, n_i (i = 1, 2, …, k) observations per sample and $n_T = \sum_{i=1}^{k} n_i$ samples in total. Test statistics for exact and approximate methods exist, the approximate method is more convenient in practice as exact critical values are available only for specific cases [81]. The test statistic of interest is [70]

$$B_{k,n_T} = \frac{\left(\sum_{i=1}^{k} \gamma_i\right) \ln \bar{s}^2 - \sum_{i=1}^{k} \gamma_i \ln s_i^2}{1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k} \frac{1}{\gamma_i} - \left(\frac{1}{\sum_{i=1}^{k} \gamma_i}\right)\right]} \qquad \text{(E.16)}$$

where $\gamma_i = n_i - 1$ and $\bar{s}^2 = \frac{\sum_{i=1}^{k} \gamma s_i^2}{\sum_{i=1}^{k} \gamma_i}$ for the sample variance s_i^2. When the null hypothesis is true, this statistic has a distribution that approximates a chi-square distribution with *k-1* degrees of freedom. The chi-square probability density function, with k-1 degrees of freedom is [73]

$$f(x \mid (k-1)) = \frac{1}{2^{\frac{(k-1)}{2}} \Gamma\left(\frac{(k-1)}{2}\right)} \exp^{-\frac{x}{2}} x^{\frac{(k-1)}{2}-1} \qquad \text{(E.17)}$$

where $x > 0$ and $(k-1) > 0$. The variances were judged to be unique if [70]

$$B_{k,n_T} > \chi^2_{\alpha,(k-1)} \qquad \text{(E.18)}$$

Which was the upper critical value of the chi-square distribution with k-1 degrees of freedom and a significance level of α (which we chose to be 0.05). The p-values associated with this test can be computed by [73]:

$$P\left(\chi^2_{(k-1)} \le x \mid (k-1)\right) = P\left(Y \le x \mid \frac{(k-1)}{2}, 2\right) \qquad \text{(E.19)}$$

where Y is a $\Gamma\left(\frac{(k-1)}{2},2\right)$ distributed random variable. Since the rejection region corresponds only to values greater than the test statistic, the p-value for the test is [73]

$$1-P\left(\chi^2_{(k-1)} \le x \mid (k-1)\right) \qquad \text{(E.20)}$$

Caution must be used when using this test as it is highly sensitive to departures of the sample distribution from normality [43]. Such departures may incorrectly indicate a difference in population variances when in fact there is not one. To guard against this possibility graphical procedures such as histograms, density estimates, and quantile plots of the data were examined and statistical tests of normality (such as the Shapiro-Wilk test) were performed.

E.4 Flinger (Killeen) *Median Test*

The Flinger-Killeen median test is a nonparametric statistical test that assesses the null hypothesis that the medians of the populations from which two samples are drawn are identical [108]. The hypothesis tested is:

$$H_0 : \sigma_1 = \sigma_2$$

$$H_A : \sigma_1 \neq \sigma_2 \qquad \text{(E.21)}$$

for group 1of sample size n_1 and group 2 of sample size n_2, that otherwise have a common statistical distribution.

Define, $z_{jk} = \left| y_{jk} - M_j \right|$ as the centred samples for y_{jk} the kth observation k = (1, …, n_j) from the jth group j = (1,2) and M_j the median value of the samples belonging to the jth group. Order the values of z_i according to increasing magnitude for all of the samples in the entire data set, i = (1, …, n) where $n = n_1 + n_2$ [123]. R_i is the position for each element z_i. Elements of equal magnitude are treated as ties and are assigned an average value. Similarly, define $R_{j,i}$ as the position of the ith element in the ordered set with elements $z_{j,i}$ consisting of those observations only from the jth group. Define *a(i)* for the ith element of the ordered set z as,

$$a(i) = \Phi^{-1}\left(\frac{i}{2(n+1)} + \frac{1}{2}\right) \qquad \text{(E.22)}$$

where Φ denotes the standard normal cumulative distribution function [123]. The test statistic has a χ^2 distribution when the null-hypothesis holds and inference can be performed using a chi-square test.

E.5 Two-Sample Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test is a non-parametric statistical technique, used to determine if a sample can be adequately described by a particular statistical distribution. It is a goodness of fit statistical test that was used to assess the assumption of normality used in the ANOVA models. The test compares the cumulative distribution function (cdf) of a known probability distribution with the empirical cumulative distribution function (ecdf) that is obtained from the data. For instance, the cdf of the univariate normal distribution is [73]:

$$F(x \mid \mu, \sigma) = \int_{-\infty}^{x} f(t \mid \mu, \sigma) dt \tag{E.23}$$

and the ecdf of the data is obtained via[70]:

$$F_n(x) = \frac{K}{n} x_{(i)} \leq x < x_{(i+1)} \tag{E.24}$$

for K observations less than or equal to x and the sequence of observations arranged in increasing order $\{x_{(i)}\}, (i = 1, \ldots, (n-1))$. The following hypothesis was then tested [70]:

$$\begin{aligned} H_o &: F(x \mid \mu, \sigma) = F_n(x) \\ H_A &: F(x \mid \mu, \sigma) \neq F_n(x) \end{aligned} \tag{E.25}$$

If the cdf and ecdf match well then the null hypothesis holds and the proposed distribution is correct. The D test statistic is used to perform inference which in practice is defined as [70]:

$$D = \max \left| F_n(x) - F(x) \right| \tag{E.26}$$

It is the absolute maximum difference between the cdf and the ecdf and ranges between 0 and 1. Tables of critical values can be consulted [70] to perform statistical comparisons at a fixed significance level, α, or p-values can be calculated for a cdf and ecdf with equal sample sizes ($m = n$) using [74]:

$$P(D^- \mid \leq x \leq D^+) = 2 \frac{\binom{2n}{\lfloor n(x+1) \rfloor}}{\binom{2n}{n}} \tag{E.27}$$

where $\lfloor n(x+1) \rfloor$ denotes the largest magnitude integer less than or equal to $n(x+1)$ and

$$\begin{aligned} D^{+} &= \sup_x \left(F\left(x \mid \mu,\sigma\right)\right) \\ D^{-} &= \sup_x \left(F_n\left(x\right) - F\left(x \mid \mu,\sigma\right)\right) \end{aligned} . \qquad \text{(E.28)}$$

Alternatively, when the samples sizes are unequal ($m \neq n$) an approximation can be used based upon a modified statistic [74]:

$$P\left(\sqrt{\frac{mn}{m+n}}D \leq x\right) \approx 1 - 2\sum_{k=1}^{+\infty} \exp^{-2k^2x^2} \qquad \text{(E.29)}$$

where m,n > 30 and of similar magnitude.

An advantage of using the KS test was that under the null hypothesis the distribution of the KS test statistic did not depend on the underlying cumulative distribution function being tested. Limitations of the KS test include that it was devised for samples from continuous distributions and that it relies on sampled data so that sparse observations in the tails of the distributions can contribute less weight to the inferences. It is a general nonparametric test and other tests (such as the Shapiro-Wilks test) designed to test more specific assumptions often out perform it in practice. The test distribution had to be specified in advance, this problem was overcome in the assessment of the ANOVA models by standardising the sample values to z-scores allowing the cdf to be based upon the standard normal distribution [73]:

$$\Phi\left(z\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{\frac{(x-\mu)}{\sigma}} \exp\left(\frac{-l^2}{2}\right) dl \qquad \text{(E.30)}$$

The proposal and estimation of other candidate distributions can prove more troublesome.

E.6 Shapiro-Wilk Test for Normality

The Shapiro-Wilk Test was used to assess the assumption of normality of the data. That is, the following hypotheses were compared:

$$\begin{aligned} H_0 &: f\left(x\right) \sim N\left(\mu,\sigma^2\right) \\ H_A &: f\left(x\right) \bar{\sim} N\left(\mu,\sigma^2\right) \end{aligned} \qquad \text{(E.31)}$$

where ~ denotes ‘distributed as’ and $\bar{\sim}$ denotes ‘not distributed as’, so that $f\left(x\right) \sim N\left(\mu,\sigma^2\right)$ means that the data obeys a normal distribution with mean μ and variance σ^2. The Shapiro-Wilk W-test is preferable than the KS test to assess normality because it has a greater level of statistical power in general [27].

For a random sample of data $X^T = (x_1, x_2,\ldots,x_n)$ for which the assumption of normality is to be assessed, the data can be ordered from smallest to largest to obtain the order statistics, $(x_{(1)}, x_{(2)},\ldots,x_{(n)})$ of the sample. Denote $\boldsymbol{m}^T = (m_1, m_2,\ldots, m_n)$ the vector of expected values of a sequence $\boldsymbol{z}^T = (z_1, z_2,\ldots, z_n)$ of ordered random numbers obeying a standard normal distribution $N(0,1)$. Specify V to be the corresponding covariance matrix that has elements denoted by $\upsilon_{i,j}$, $(i, j = 1,\ldots, n)$. That is, $E(z_i) = m_i (i = 1,.., n)$ and $cov(z_i,z_j) = \upsilon_{i,j}$. The Shapiro-Wilk W-test statistic is then given by [27]:

$$W = \frac{\left(\vec{a}^T \vec{x}\right)^2}{(n-1)s^2} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{E.32}$$

where s^2 is the sample variance and the coefficient vector

$$\left\{\vec{a}^T\right\} = \vec{m}^T \vec{V}^{-1}\left[\left(\vec{m}^T \vec{V}^{-1}\right)\left(\vec{V}^{-1}\vec{m}\right)\right]^{-\frac{1}{2}} \tag{E.33}$$

are normalised 'best linear unbiased coefficients' [104]. Explicit use of the covariance matrix $\vec{V}$ can be avoided in the calculation of $\left\{\vec{a}^T\right\}$ (which is beneficial for increasing sample size, n and avoids excessive matrix operations) by an approximation. Modify the coefficients to give $\vec{a}^* = \vec{m}^T \vec{V}^{-1}$, the approximation $\left(\hat{a}^*\right)$ to these modified coefficients are given by:

$$\hat{a}_i^* = \begin{cases} 2m_i, i = 2,3,\ldots, n-1 \\ \left(\dfrac{\hat{a}_1^2}{1-2\hat{a}_1^2}\sum_{i=2}^{n-1} a_i^{*2}\right)^{\frac{1}{2}}, i = 1, i = n, \end{cases} \tag{E.34}$$

where

$$\hat{a}_1^2 = \hat{a}_n^2 = \begin{cases} g(n-1), n \le 20 \\ g(n), n > 20 \end{cases} \tag{E.35}$$

and

$$g(n) = \frac{\Gamma\left(\frac{1}{2}[n+1]\right)}{\sqrt{2}\Gamma\left(\frac{1}{2}n+1\right)} \tag{E.36}$$

In practice, a close approximation to $g(n)$ is utilised [104]. The approximation is in error by less than 1% and gets more accurate as the sample size, n, increases [111]. To perform statistical inference, the following transformation is applied:

$$y = (1-W)^{\lambda} \text{ and } z = \frac{(y-\mu_y)}{\sigma} \tag{E.37}$$

where λ is a constant, z is the standard normal statistic, μ_y and σ_y are the mean and standard deviation of the transformed variable y, respectively [104]. The appropriate value of the power λ is determined by:

$$\lambda = \mathrm{P}(n) = \sum_{i=0}^{K} c_i (\log n - d)^i \qquad \text{(E.38)}$$

where K is the number of coefficients in the polynomial, which vary with sample size: $\{(7 \le n \le 20) \rightarrow K = 2, (21 \le n \le 2000) \rightarrow K = 5\}$ and d is a parameter that also varies with sample size: $\{d = 3$ for $n \le 20$, $d = 5$ for $21 \le n \le 2000\}$. The coefficients, $\vec{c}_j^{\,T}\,(j = 1,2)$, of the polynomials are [104]:

$$\vec{c}_1^{\,T} = [0.118898, 0.133414, 0.327907] \text{ for } (7 \le n \le 20) \text{ and,}$$

$$\vec{c}_2^{\,T} = [0.480385, 0.318828, 0, -0.0241665, 0.00879701, 0.002989646]$$

$$\text{for } (21 \le n \le 2000) \qquad \text{(E.39)}$$

After applying this transformation, the W-statistic has an approximate standard normal distribution and statistical inference performed using this distribution. If the sample size, n, is less than 7, then the transform is inadequate, but tabulated critical values of the W-statistic are known for many cases and an exact probability density function specified for the special case of n = 3 [104, 111]. The Shapiro-Wilk test is a useful statistical tool to assess normality of data and can be readily used up to sample sizes of n = 2000. Nonetheless, caution must be used when implementing the test as it is known to be sensitive to outlying data and it is recommended that the data must undergo other assessments using box-plots, quantile plots, as well as the KS test to guard against model misspecifications.

E.7 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) transforms data described by a vector of correlated random variables $\vec{x}$ into data described by linear combination of a vector of orthogonal random variables $\vec{z}$. That is, for

$$\vec{x} = (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) \qquad \text{(E.40)}$$

the representation is

$$\vec{z}_k = \vec{\alpha}_k^T \vec{x} \qquad \text{(E.41)}$$

where $\vec{\alpha}_k$ is the kth eigenvector ($k = 1,\ldots,N$) of the singular value decomposition (SVD) of the covariance matrix Σ of the data. In practice, the covariance matrix Σ is estimated using the sample covariance matrix $\vec{S}$, so that the SVD gives $\vec{S} = \vec{\alpha}^T \vec{\Lambda} \vec{\alpha}$

where $\vec{\alpha}$ is the matrix of N eigenvectors and $\vec{\Lambda}$ a diagonal matrix of eigenvalues of rank N. The eigenvalues λ_k of the SVD are ordered in decreasing magnitude along the main diagonal of the matrix $\vec{\Lambda}$ [124].

The solution is constrained, in our applications, so that $\vec{\alpha}_k^T \vec{\alpha}_k = 1$, which is convenient as then the vector $\alpha_1^T x$, which describes the maximum variation in the data set $\vec{X}$ corresponds to $\mathrm{var}(\vec{z}_1) = \vec{\alpha}_1^T \vec{S} \vec{\alpha}_1 = \lambda_1$. The vectors $\vec{\alpha}_k^T \vec{x}$ are referred to as the principal components and the eigenvectors $\vec{\alpha}_k$ as the loadings. The first principal component corresponds to the eigenvector $\vec{\alpha}_1$ with the largest magnitude eigenvalue, λ_1. In general, the kth principal component corresponds to the kth eigenvector, $\vec{\alpha}_k$. The random variables $\vec{x}$ were described for observations of multiple samples as a data matrix $\vec{X}$ of n rows (corresponding to the number of samples) and N columns (corresponding to the number of variables). Prior to the PCA transformation, the data was scaled to ease interpretation. The scaling process is defined as:

$$\vec{x}_i^* = \frac{\left(\vec{x}_i - \bar{x}_i\right)}{\sqrt{\sum_{j=1}^{n} \vec{x}_j^2 \Big/ (n-1)}} \tag{E.42}$$

For each of the (i = 1,…, N) variables. These x_i^* variables were the data to which the PCA transform was actually applied. Without prior scaling, there is a risk that PCA would give misleading results. For instance, two random variables $\vec{x}_1$ and $\vec{x}_2$ might have values that differ by several orders of magnitude. The random variable $\vec{x}_1$ has a large mean but little variation, whilst the other random variable $\vec{x}_2$ has a very small mean value but a relatively large variation about that value. Application of PCA without adjusting these two variables to be on the same scale will tend to favour the selection of principal components that describe variation in the first variable as the magnitude of the variations of the second variable will be insignificant in comparison.

The objective of PCA is to find a few principal components $\vec{\alpha}_k^T \vec{x}$ that effectively describes the variation in the data set. The number of variables describing the data set can be then reduced and allow for a graphical and mathematical interpretation in terms of orthogonal vectors. In this work, the number of principal components selected was determined by the requirement that at lease 95% of the

variation in the data set was explained. The contribution of the jth principal component was assessed using the ratio,

$$\gamma_j = \frac{\lambda_j}{\sum_{i=1}^{N} \lambda_i^2} \tag{E.43}$$

and the cumulative contribution by

$$f(j) = \sum i = 1^{j \leq N} \gamma_i \tag{E.44}$$

E.8 Receiver Operating Characteristic (ROC)

Receiver Operating Characteristic (ROC) curve analysis was used to evaluate the diagnostic performance of a test or the accuracy of a model to discriminate diseased cases from normal cases [59, 83, 133]. Cut-off points or decision criterion that was used to discriminate between the two tissue groups compared in a model, was used to correctly classify diseased samples as positive (TP = true positive fraction). However, some cases with disease were, of course, classified as negative (FN = false negative fraction). Some cases without disease were correctly classified as negative (TN = true negative fraction) and some cases without disease were also classified as positive (FP = false positive fraction). From this, the sensitivity and specificity of the diagnostic test was determined:

$$Sensitivity = \frac{TP}{TP + FN} \tag{E.45}$$

$$Specificity = \frac{TN}{TN + FP} \tag{E.46}$$

The sensitivity represented the probability that the test result was positive when disease was present (i.e. the true positive rate) and the specificity represented the probability that the test result was negative when the disease was not present (i.e. the true negative rate).

ROC curves were plotted using the true positive rate (*Sensitivity*) as a function of the false positive rate (*1 - Specificity*) for different cut-off points. In this experiment, a cut-off point of 60% was used to calculate all ROC curves. A test with perfect discrimination has an ROC curve that passes through the upper left corner (100% *sensitivity*, 100% *specificity*). Therefore, the closer the ROC curve was to the upper left corner, the higher the accuracy of the test or model being analysed [133]. However, if the ROC curve was close to, or followed, a 45 degree diagonal of the ROC space, then the test was considered to be completely random in determining a

diagnosis. It follows then, that the area under the ROC curve (AUC) represented the accuracy of the test: an area of 1 represents a perfect test; an area of 0.5 represents a test with no useful diagnostic information and an area less than 0.5 corresponds to a test that has useful diagnostic information but that has been applied incorrectly. The area measured discrimination, or the ability of the test to correctly classify those with and without disease. Typically, an area between 0.90-1 is excellent, 0.80-0.90 is good, 0.70-0.80 is fair, 0.60-0.70 is poor, and 0.50-0.60 is a fail.

A limitation of ROC curves is that it does not readily account for the frequency (prior probability) of each disease state. Another major issue is the potential for two very different shaped ROC curves to have the same AUC statistics. A ROC curve with a higher AUC may also perform poorly in a certain region as compared to a curve with a lower AUC value. To counter these problems prior probabilities were used (where possible) in the diagnostic model and the area statistic was assessed in light of the inspection of the ROC curve.

E.9 Quadratic Discriminant Analysis (QDA)

Quadratic discriminant analysis (QDA) is a model for assigning classification to data from different groups that is based upon the normal distribution. For a training sample [126]:

$$\vec{x} = \left(\vec{x}_1, \vec{x}_2, ..., \vec{x}_K\right) \tag{E.47}$$

estimates of the mean vectors $\vec{\mu}$ and covariance matrices $\vec{\sigma}$ are obtained conditional on group membership (that is, the parameters are extracted from data belonging from group1, group 2, …, group G). These estimates are denoted [126]:

$$\hat{\mu}_i \text{ and } \hat{\Sigma}_i = \vec{S}_i \tag{E.48}$$

and are found using the method of maximum likelihood to be [126]:

$$\hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \vec{x}_i \tag{E.49}$$

$$\hat{\Sigma}_i = \vec{S}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(\vec{x}_{i,j} - \hat{\mu}_i\right)\left(\vec{x}_{i,j} - \hat{\mu}_i\right)^T \tag{E.50}$$

for groups $g_i(i = 1,\ldots, G)$ and n_i observations for the ith group. Once found, they provide 'plug-in' parameters to the normal distribution of each group [126]:

$$\hat{f}\left(x \mid g_i\right) = \frac{1}{(2\pi)^{\frac{d}{2}}\left|\vec{S}_i\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\vec{x} - \hat{\mu}_i\right)^T S_i^{-1}\left(\vec{x} - \hat{\mu}_i\right)\right\} \tag{E.51}$$

for groups g_i(i = 1,…, G), and d dimensions, to produce the quadratic discrimination rule [124]:

$$\hat{Q}_i(\vec{x}) = (\vec{x}_i - \hat{\mu}_i)S_i^{-1}(\vec{x}_i - \hat{\mu}_i)^T + \log\left|\vec{S}_i\right| - 2\log(\pi_i) \qquad \text{(E.52)}$$

Where $\pi_i = n_i/N$ is the prior probability for the ith group (as calculated by sampling) for a total number of N observations in the data set irrespective of group label. The posterior probability of membership of each observation to each group was then calculated by determining the probability that the observation belonged to the normal distribution describing that group, and then by the result weighting by the corresponding prior probability for that group. The group label with the largest posterior probability was then assigned as the 'classification' for each observation.

E.10 Bivariate Gaussian Kernel Estimation

Visualisation of the series data for both the synchrotron and camera analysis was best shown using smoothed estimates of the point-values obtained from each tissue sample analysed for the patient. An in depth discussion of this technique can be reviewed in the references [23, 29, 33], but a general description is presented here for a basic, simplified explanation.

In order to map the data over a 2D area, a kernel density estimate over a mesh of grid points was created so that its graph could be plotted. Because there were a limited number of known data points over the grid, an approximation of the data in between the known point values was estimated based on a discrete convolution structure which was computed quickly using the fast Fourier transform (FFT).

The first step in the computation of the 2D connecting function was to obtain grid counts over the 2D area: in this project, a 32 x 32 grid was created for each patient, regardless of the number of tissue points. The binned kernel density estimator relied on assigning certain weights to the grid points, based on neighbouring observations. The kernel weights are designed to provide a discrete representation of the bivariate normal distribution:

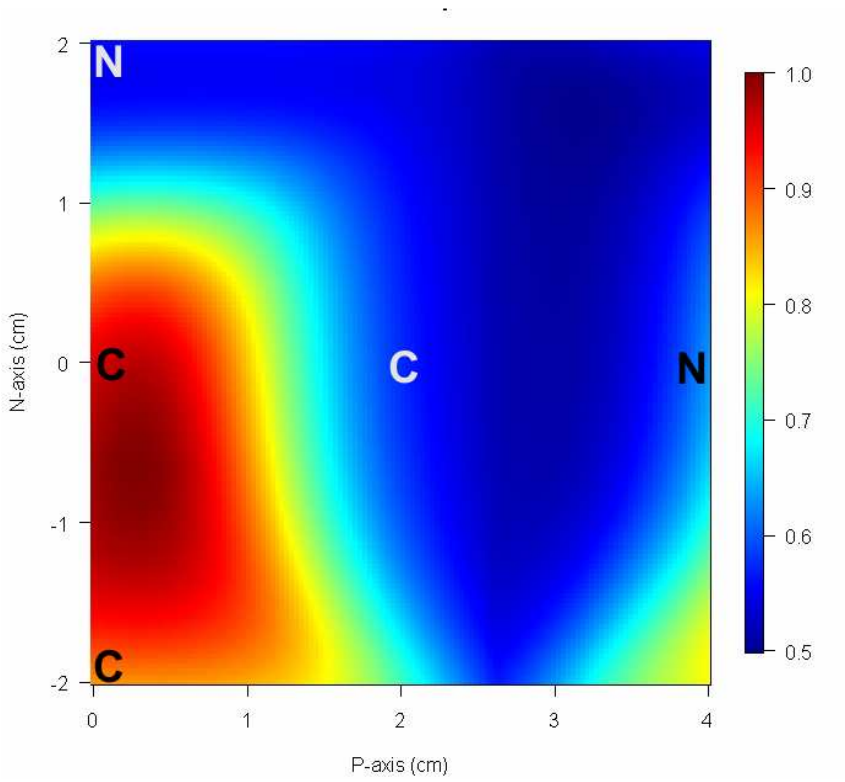
$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp-\left[(x-\mu_x)^2/2\sigma_x^2 + (y-\mu_y)^2/2\sigma_y^2\right] \qquad \text{(E.53)}$$

Once the kernel weights were computed, the grid points were weighted by convolving the grid counts with the kernel weights.

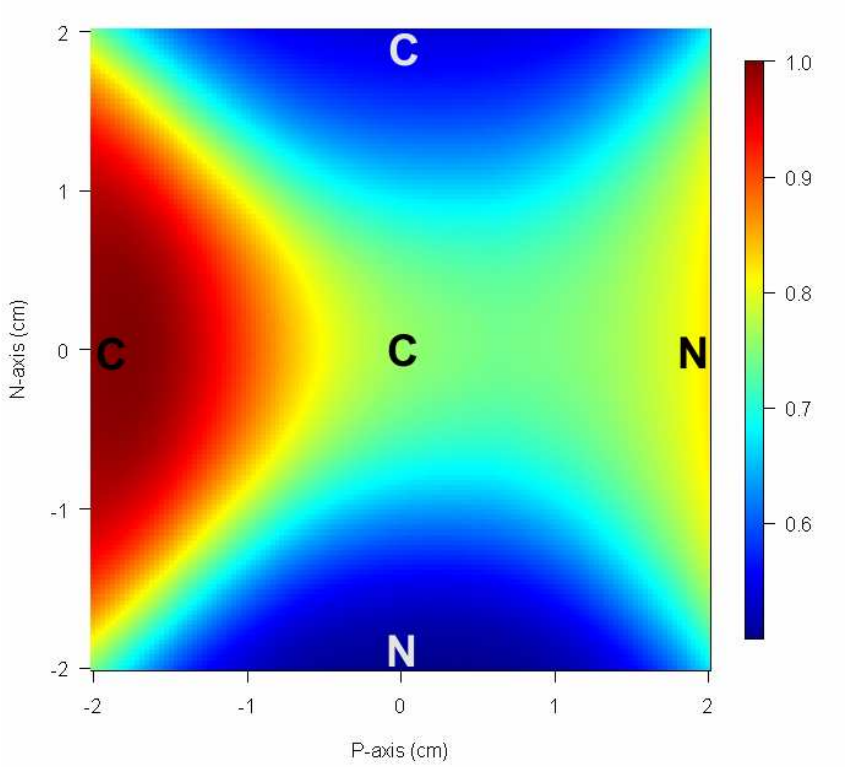
The discrete convolution of these two vectors (grid counts and kernel weights) was computed using the FFT by using the discrete convolution theorem [14, 60]. This was done by simply multiplying the Fourier transforms of the two vectors element-by-element and then inverting the result to obtain the convolution vector. One downfall was that certain periodicity assumptions had to be met in order to avoid wrap-around effects. This was seen with those patients who had a small number of data points available for their computation. Patient who underwent Wide Local Excision (WLE) had smaller excised masses than mastectomy patients, where less than 6 tissue samples were available for mapping, which resulted in wrap around effects being seen with their maps. In these cases, it may be more appropriate to re-calculate the 2D grid with zero-padding the boundaries.

Figure E.1 shows an example of two WLE patients who had less than 6 tissue points sampled from their excised mass. The patient's map in Figure E.1a showed wrap-around effects at the lower right corner (or P(+4):N(-2) sector), which suggested tissue degradation was present in this area of the excised mass. This was misleading, since there were no tissue samples taken specifically at this point to suggest higher values of amorphous scatter in this sector. The patient shown in Figure E.1b exhibited wrap around effects along the P(+2) edge, where the tissue sample taken at P(+2) was diagnosed as normal parenchyma.

Calculating the maps using a padded grid of zero values could be used to show the areas of low and high amorphous scatter or D-spacing values without the wrap-around effects. Figure E.2a shows the padded 2D model for the patient in Figure E.1a and Figure E.2b shows the padded 2D model for the patient in Figure E.1b. An extra 2 cm interval along each axis was padded with zero values to re-calculate the map. The padded map did not show wrap around effects with any of the patients analysed in this study. However, the padded model cannot be taken as accurate around the edges, where the extended distance was set as zero, especially when considering the gradient of the values around the extended area.

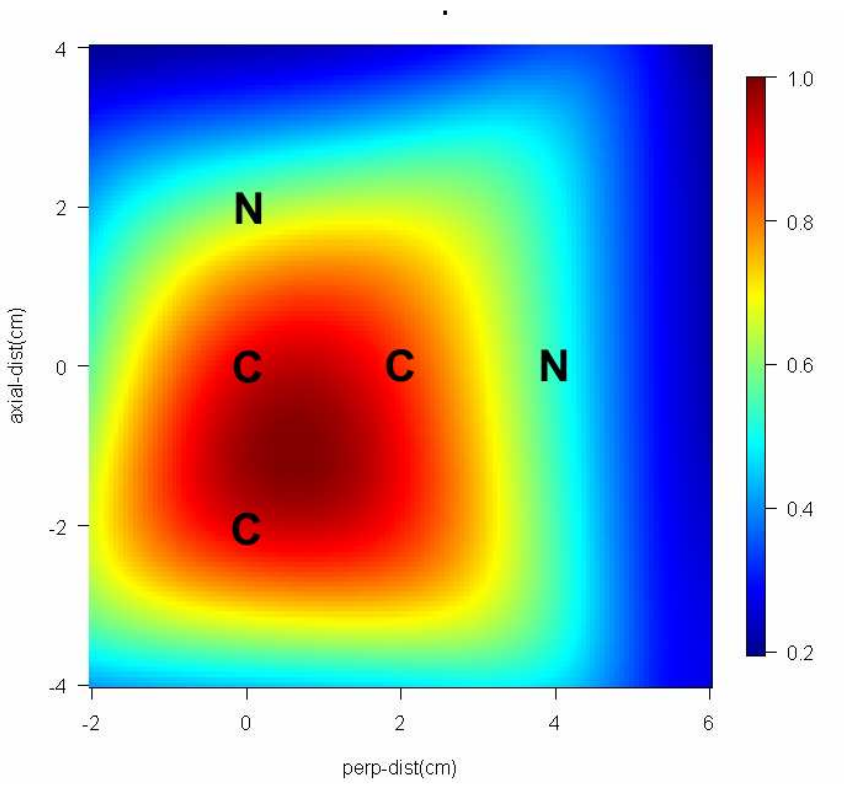


(a)

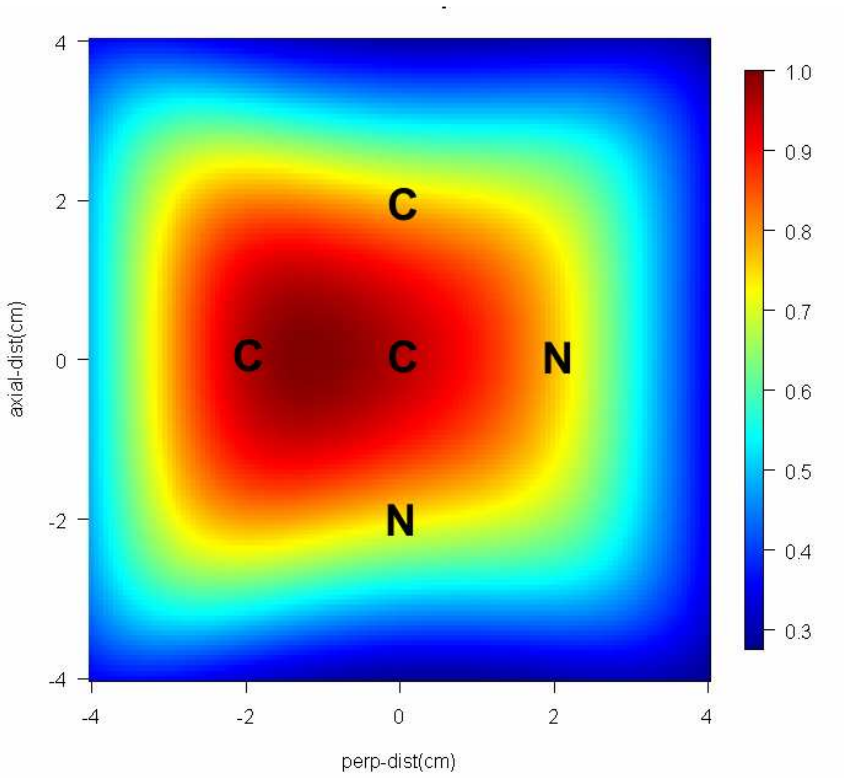


(b)

Figure E.1: 2D models using the amorphous scatter for (a) Patient 18 and (b) Patient 24 who both underwent WLE where less than 6 tissue samples were taken from the excised mass. The model shows wrap-around effects, where the Kernel does not accurately calculate the model because there is not enough data for proper computation. C represents tissue samples diagnosed as invasive carcinoma and N represents tissue samples diagnosed as normal parenchyma.



(a)



(b)

Figure E.2: Padded 2D models using the amorphous scatter for (a) Patient 18 and (b) Patient 24 who both underwent WLE where less than 6 tissue samples were taken from the excised mass. An extra 2 cm interval was added on the edge of each axis and padded with zero values in order to calculate the model. C represents tissue samples diagnosed as invasive carcinoma and N represents tissue samples diagnosed as normal parenchyma.

References

[1] Alitalo K, Tammela T and Pterova T V 2005 Lymphangiogenesis in development and human disease *Nature* 438 946-953

[2] Amat S, Penault-Llorca F, Cure H, Le Bouedec G, Achard J, van Praagh I, Feillel V, Mouret-Reynier M, Dauplat J and Chollet P 2002 Scarff-bloom-richardson (sbr) grading: A pleiotropic marker of chemosensitivity in invasive ductal breast carcinomas treated by neoadjuvant chemotherapy *International Journal of Oncology* 20 791-796

[3] Amemiya Y 1995 Imaging plates for use with synchrotron radiation *Journal of Synchrotron Radiation* 2 13-21

[4] Amemiya Y, Wakabayashi K, Tanaka H, Ueno Y and Miyahara J 1987 Laser-stimulated luminescence used to measure X-ray diffraction of a contracting striated muscle *Science* 237 164-168

[5] Antley C, Mooney E and Layfield L 1998 A comparison of accuracy rates between open biopsy, cutting-needle biopsy, and fine-needle aspiration biopsy of the breast: A 3-year experience *Breast Journal* 4 3-8

[6] Australia, BreastScreen 2004 National accreditation standards *BreastScreen Australia*, Canberra

[7] Azar F S 2001 *A deformable finite element model of the breast for predicting mechanical deformations under external perturbations*. Thesis: Department of Bioengineering, University of Pennsylvania, Philadelphia

[8] Aznavoorian S, Moore B A, Alexander-Lister L D, Hallit S L, Windsor J and Engler J A 2001 Membrane type I-matrix metalloproteinase-mediated degradation of type I collagen by oral squamous cell carcinoma cells *Cancer Research* 61 6264-6275

[9] Berg W A, Gutierrez L, NessAiver M S, Carter W B, Bhargavan M, Lewis R S and Ioffe O B 2004 Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer *Radiology* 233 830-849

[10] Bigi A and Roveri N 1991 *Handbook on synchrotron radiation* Editors: E Ebashi, *et al*., Elsevier, Amsterdam pp 199-239

[11] Blanton T N, Huang T C, Toraya H, Hubbard C R, Robie S B, Louer D, Gobel H E, Will G, Gilles R and Raferty T 1995 JCPDS - international centre for diffraction data round robin study of silver behenate. A possible low-angle X-ray diffraction calibration standard *Powder Diffraction* 10 91-95

[12] Bloom H and Richardson W 1957 Histological grading and prognosis in breast cancer *British Journal of Cancer* 11 359-377

[13] Bragg W L 1913 *The diffraction of short electromagnetic waves by a crystal* Proceedings of the Cambridge Philosphical Society 17 43-57

[14] Brigham E O 1988 *The Fast Fourier Transform and its applications* Prentice Hall, Englewood Cliffs, New Jersey

[15] Butler S M, Webb G I and Lewis R A 2003 *A case study in feature invention for breast cancer diagnosis using X-ray scatter images* Proceedings of the 16th Australian Joint Conference on Artificial Intelligence 2903 677-685

[16] Carriaga M and Henson D 1995 The histologic grading of cancer *Cancer* 75 406-421

[17] Castro C R F, Barroso R C, Anjos M J, Lopes R T and Braz D 2004 Coherent scattering characteristics of normal and pathological human breast tissues *Radiation Physics and Chemistry* 71 649-651

[18] Castro C R F, Barroso R C and Lopes R T 2005 Scattering signatures for some human tissues using synchrotron radiation *X-ray Spectrometry* 34 477-480

[19] Centre, National Breast and Ovarian Cancer 2008 *The pathology reporting of breast cancer: A guide for pathologists, surgeons, radiologists and oncologists* www.nbocc.org.au: Australian Cancer Network, Cancer Council 1-52

[20] Changizi V, Oghabian M A, Speller R, Sarkar S and Kheradmand A A 2005 Application of small angle x-ray scattering (SAXS) for differentiation between normal and cancerous breast tissue *International Journal of Medical Sciences* 2 118-121

[21] Chen C, Sakoda L C, Doherty J A, Loomis M M, Fish S, Ray R M, Lin M G, Fan W, Zhao L P, Gao D L, Stalsberg H, Feng Z and Thomas D B 2008 Genetic variation in CYP19a1 and risk of breast cancer and fibrocystic breast conditions among women in Shanghai, China *Cancer Epidemiology Biomarkers and Prevention* 17 3457-3466

[22] Coberneck R C 1980 Breast biopsy: A study of cost-effectiveness *Annals of Surgery* 192 152-156

[23] Collaboration D, Abazov V M, Abbott B and Abdesselam A 2001 Search for new physics using QUAERO: A general interface to D0 event data *Physical Review Letters* 87 231801

[24] Corporation, Rigaku 2008 *SAXS Osmic camera user manual* Rigaku Corporation, Auburn Hills, Texas

[25] Corporation, Rigaku 2008 *User manual: Copper microfocused x-ray beam generator for SAXS* Rigaku Corporation, Auburn Hills, Texas

[26] Cullity B and Stock S 2001 *Elements of x-ray diffraction* Prentice Hall, Pearson Education International

[27] D'Agostino R B and Stephens M A 1986 *Goodness of fit techniques* CRC Press, Boca Raton, Florida

[28] D'Arrigo C C and Fentiman I S 2004 Pathology of breast carcinoma *International Journal of Clinical Practice* 58 29-34

[29] Davis D T and Hwang J-N 1998 *Expanding Gaussian kernels for multivariate conditional density estimation* IEEE Transactions on Signal Processing 46 269-275

[30] Daxer A, Misof K, Grabner B, Ettl A and Fratzl P 1998 Collagen fibrils in the human corneal stroma: Structure and aging Investigative *Ophthalmology & Visual Science* 39 644-648

[31] De Vita V, Hellmann S and Rosenberg S 2001 *Cancer: Principles & practice of oncology* Lippincott-Raven

[32] Donegan W and Spratt J 2002 *Microscopic anatomy of the breast* Saunders, Philadelphia

[33] Duda R and Hart P 1973 *Pattern classification and scene analysis* John Wiley & Sons

[34] Duffy M J, Maguire T M, Hill A, McDermott E and O'Higgins N 2000 Metalloproteinases: Role in breast carcinogenesis, invasion and metastasis *Breast Cancer Research* 2 252 - 257

[35] Eikenberry E F and Brodsky B 1980 X-ray diffraction of reconstituted collagen fibres *Journal of Molecular Biology* 144 397-404

[36] Eikenberry E F, Brodsky B and Parry D A D 1982 Collagen fibril morphology in developing chick metatarsal tendons 1. X-ray diffraction studies *International Journal of Biological Macromolecules* 4 322-328

[37] Elmore J G, Armstrong K, Lehman C D and Fletcher S W 2005 Screening for breast cancer *JAMA* 293 1245-1256

[38] Eng J 2007 Receiver operator characteristic curve calculator www.jrocfit.org.

[39] Engel J 1991 Common structural motifs in proteins of the extracellular matrix *Current Opinions in Cell Biology* 3 779-785

[40] Engel J and Bachinger H P 2005 *Structure, stability and folding of the collagen triple helix* Springer Berlin/Heidelberg

[41] Evans S, Bradley D, Dance D, Bateman J and Jones C 1991 Measurement of small-angle photon scattering for some breast tissues and tissue substitute materials *Physics in Medicine and Biology* 36 7-18

[42] Everitt B S 2006 *The Cambridge dictionary of statistics* University Press, New York

[43] Everitt B S and Hothorn T 2006 *A handbook of statistical analysis using R* Taylor and Francis, Boca Roton, Florida

[44] Falzon G, Pearson S, Murison R, Hall C J, Siu K K W, Evans A, Rogers K D and Lewis R A 2006 Wavelet-based feature extraction applied to small angle x-ray scattering patterns from breast tissue: A tool for differentiating between tissue types *Physics in Medicine and Biology* 51 2465-2477

[45] Federman S, Miller L M and Sagi I 2002 Following matrix metalloproteinases activity near the cell boundary by infrared micro-spectroscopy *Matrix Biology* 21 567-577

[46] Feigin L A and Svergun D I 1987 *Structure analysis by small-angle X-ray and neutron scattering* Plenum Press, New York

[47] Fentiman I S I S 2004 Pathogenesis of breast carcinoma *International Journal of Clinical Practice* 58 35-40

[48] Ferguson J E, Schor A M, Howell A and Ferguson M W 1992 Changes in the extracellular matrix of the normal human breast during the menstrual cycle *Cell & Tissue Research* 268 167-177

[49] Fernandez M, Keyrilainen J, Karjalainen-Lindsberg M-L, Leidenius M, Von Smitten K, Fiedler S and Suortti P 2004 Human breast tissue characterisation with small-angle X-ray scattering *Spectroscopy* (Eugene Or.) 18 167-176

[50] Fernandez M, Keyrilainen J, Serimaa R, Torkkeli M, Karjalainen-Lindsberg M-L, Leidenius M, Smitten K v, Tenhunen M, Fiedler S, Bravin A, Weiss T M and Suortti P 2005 Human breast cancer in vitro: Matching histo-pathology with small-angle x-ray scattering and diffraction enhanced X-ray imaging *Physics in Medicine and Biology* 50 2991-3006

[51] Fernandez M, Keyrilainen J, Serimaa R, Torkkeli M, Karjalainen-Lindsberg M-L, Tenhunen M, Thomlinson W, Urban V and Suortti P 2002 Small-angle x-ray scattering studies of human breast tissue samples *Physics in Medicine and Biology* 47 577-592

[52] Floyd C E, Chotas H G, Dobbins J T I and Ravin C E 1990 Quantitative radiographic imaging using a photostimulable phosphor system *Medical Physics* 17 454-459

[53] Flug M and Kopf-Maier P 1995 The basement membrane, its involvement in carcinoma cell invasion *Acta Anatomica* 152 69-84

[54] Foundation, American Society of Clinical Oncology & the ASCO 2007 *Report: People living with cancer*

[55] Fratzl P, Fratzl-Zelman N and Klaushofer K 1993 Collagen packing and mineralization. An x-ray scattering investigation of turkey leg tendon *Biophysical Journal* 64 260-266

[56] Fratzl P, Misof K, Zizak I, Rapp G, Amenitsch H and Bernstorff S 1998 Fribrillar structure and mechanical properties of collagen *Journal of Structural Biology* 122 119-122

[57] Fullwood N J and Meek K M 1994 An ultrastructural, time-resolved study of freezing in the corneal stroma *Journal of Molecular Biology* 236 749-758

[58] Glatter O and Kratky O (eds.) 1982 *Small angle X-ray scattering* Academic Press, London

[59] Griner P F, Mayewski R J, Mushlin A I and Greenland P 1981 Selection and interpretation of diagnostic tests and procedures *Annals of Internal Medicine* 94 555-600

[60] Grunbaum F A 1982 The eigenvectors of the discrete Fourier Transform: A version of the Hermite functions Journal of mathematical *Analysis and Applications* 88 355-363

[61] Huang T C, Toraya H, Blanton T N and Wu Y 1993 X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard *Journal of Applied Crystallography* 26 180-184

[62] Hulmes D J, Jesior J C, Miller A, Berthet-Colominas C and Wolff C 1981 *Electron microscopy shows periodic structure in collagen fibril cross sections* Proceedings of the National Academy of Sciences 78 3567-3571

[63] Hulmes D J and Miller A 1979 Quasi-hexagonal molecular packing in collagen fibrils *Nature* 282 878-880

[64] Hulmes D J, Wess T J, Prockop D J and Fratzl P 1995 Radial packing, order, and disorder in collagen fibrils *Biophysical Journal* 68 1661-1670

[65] Irwig L, Houssami N and van Vliet C 2004 New technologies in screening for breast cancer: A systematic review of their accuracy *British Journal of Cancer* 90 2118-2122

[66] Jiang L, Verman B and Dan Joensen K 2000 A general approach in multilayer optical system design for SAXS *Journal of Applied Crystallography* 33 801-803

[67] Kauppila S, Stenback F, Risteli J, Jukkola A and Risteli L 1998 Aberrant type I and type III collagen gene expression in human breast cancer in vivo *Journal of Pathology* 186 262-268

[68] Kidane G, Speller R D, Royle G J and Hanby A M 1999 X-ray scatter signatures for normal and neoplastic breast tissues *Physics in Medicine and Biology* 44 1791-1802

[69] Kobayashi K, Niwa J, Hoshino T and Nagatani T 1992 Electron microscopic visualization of collagen aggregates without chemical staining *Journal of Electron Microscopy* 41 235-241

[70] Kokoska S and Zwillinger D 2000 *CRC standard probability and statistics tables and formulae* CRC Press, Boca Raton, Florida

[71] Kolb T M, Lichy J and Newhouse J H 2002 Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: An analysis of 27,825 patient evaluations *Radiology* 225 165-175

[72] Kosanetzky J, Knoerr B, Harding G and Neitzel U 1987 X-ray diffraction measurements of some plastic materials and body tissues *Medical Physics* 14 526-532

[73] Krishnamoorthy K 2008 *Handbook of statistical distributions with applications* Chapman and Hall/CRC Press, Boca Raton, Florida

[74] Kvam P H and Vidakovic B 2007 *Nonparametric statistics with applications to science and engineering* Wiley-Interscience, New Jersey

[75] Lazarev P, Paukshto M, Pelc N and Sakharova A 2000 *Human tissue x-ray diffraction: Breast, brain, and prostate* Proceedings of the 22nd Annual EMBS International Conference, Chicago, IL 3230-3233

[76] Lewis R A 1994 Multiwire gas proportional counters: Decrepit antiques or classic performers? *Journal of Synchrotron Radiation* 1 43-53

[77] Lewis R A, Rogers K D, Hall C J, Towns-Andrews E, Slawson S, Evans A, Pinder S E, Ellis I O, Boggis C R M, Hufton A P and Dance D R 2000 Breast cancer diagnosis using scattered X-rays *Journal of Synchrotron Radiation* 7 348-352

[78] Lochter A and Bissell M J 1995 Involvement of extracellular matrix constituents in breast cancer *Seminars in Cancer Biology* 6 165-173

[79] Luparello C, Sheterline P, Pucci-Minafra I and Minafra S 1991 A comparison of spreading and motility behaviours of 8701-bc breast carcinoma cells on type I, II-trimer, and type V collagen substrata *Journal of Cell Science* 100 179-185

[80] Lurio L, Mulders N, Paetkau M, Jemian P R, Narayanan S and Sandy A 2007 Windows for small-angle X-ray scattering cryostats *Journal of Synchrotron Radiation* 14 527-531

[81] Manoukian E B 1986 *Mathematical nonparametric statistics* Gordon and Breach, New York

[82] Meek K M and Quantock A J 2001 The use of x-ray scattering techniques to determine corneal ultrastructure *Progress in Retinal and Eye Research* 20 95-137

[83] Metz C E 1978 *Basic principles of ROC analysis* Seminars in Nuclear Medicine 8 283-298

[84] Miller A and Wray J S 1971 Molecular packing in collagen *Nature London* 230 437-439

[85] Minafra S, Luparello C, Rallo F and Pucci-Minafra I 1988 Collagen biosynthesis by a breast carcinoma cell strain and biopsy fragments of the primary tumour *Cell Biology International Reports* 12 895-905

[86] Miyahara J, Takahashi K, Amemiya Y, Kamiya N and Satow Y 1986 A new type of X-ray area detector utilizing laser stimulated luminescence *Nuclear Instruments & Methods in Physics Research* A246 572-578

[87] Nakopoulou L, Tsirmpa I, Alexandrou P, Louvrou A, Ampela C, Markaki S and Davaris P S 2003 Mmp-2 protein in invasive breast cancer and the impact of MMP-2/timp-2 phenotype on overall survival *Breast Cancer Research & Treatment* 77 145-155

[88] Ne F, Gazeau D, Lambard J, Lesieur P, Zemb T and Gabriel A 1993 Characterization of an image-plate detector used for quantitative small-angle scattering studies *Journal of Applied Crystallography* 26 763-773

[89] Neter J, Kutner M H, Nachtsheim C J and Wasserman W 1996 *Applied linear statistical models* WCB/McGraw Hill, New York

[90] Ooi G J, Fox J, Siu K K W, Lewis R A, Bambery K R, McNaughton D and Wood B R 2008 Fourier transform infrared imaging and small angle X-ray scattering as a combined biomolecular approach to diagnosis of breast cancer *Medical Physics* 35 2151-2161

[91] Pearson S J, Siu K K W, Hall C E, Reid C and Falzon G 2006 *Small angle x-ray scattering and second harmonic generation imaging studies of collagen in invasive carcinoma* Australian Institute of Physics 17th National Congress, Brisbane, Australia 4

[92] Polette M and Birembaut P 1995 Membrane-type metalloproteinases in tumour invasion *International Journal of Biochemistry & Cell Biology* 30 1195-1202

[93] Poste G and Fidler I 1990 The pathogenesis of cancer metastasis *Nature* 283 139-146

[94] Pucci-Minafra I, Luparello C, Andriolo M, Basirico L, Aquino A and Minafra S 1993 A new form of tumor and fetal collagen that binds laminin *Biochemistry* 32 7421-7427

[95] Pucci-Minafra I, Luparello C, Schillaci R and Sciarrino S 1987 Ultrastructural evidence of collagenolytic activity in ductal infiltrating carcinoma of the human breast *International Journal of Cancer* 39 599-603

[96] Pucci Minafra I, Andriolo M, Basirico L, Alessandro R, Luparello C, Buccellato C, Garbelli R and Minafra S 1998 Absence of regular 2(I) collagen chains in colon carcinoma biopsy fragments *Carcinogenesis* 19 575-584

[97] Pucci Minafra I, Luparello C, Sciarrino S, Tomasino R M and Minafra S 1985 Quantitative determination of collagen types present in the ductal infiltrating carcinoma of human mammary gland Cell Biology International Reports 9 291-296

[98] Pucci Minafra I, Minafra S, Tomasino R M, Sciarrino S and Tinervia R 1986 Collagen changes in the ductal infiltrating (scirrhous) carcinoma of the human breast. A possible role played by type I trimer collagen on the invasive growth *Journal of Submicroscopic Cytology* 18 795-805

[99] Rajkumar L, Balasubramanian K, Arunakaran J, Govindarajulu P and Srinivasan N 2005 Influence of Estradiol on mammary tumour collagen solubility in DMBA-induced rat mammary tumours *Cell Biology International* 30 164-168

[100] Ramakrishnan V 1985 A treatment of instrumental smearing effects in circularly symmetric small-angle scattering *Journal of Applied Crystallography* 18 42-46

[101] Rogers K D, Lewis R A, Hall C J, Towns-Andrews E, Slawson S, Evans A, Pinder S E, Ellis I O, Boggis C R M and Hufton A 1999 Preliminary observations of breast tumour collagen *Synchrotron Radiation News* 12 15-20

[102] Round A R 2006 *Ultra-structural analysis of breast tissue* Thesis: Materials and Medical Sciences, Cranfield University, Cranfield, United Kingdom

[103] Round A R, Wilkinson S J, Hall C J, Rogers K D, Glatter O, Wess T and Ellis I O 2005 A preliminary study of breast cancer diagnosis using laboratory based small angle X-ray scattering *Physics in Medicine and Biology* 50 4159-4168

[104] Royston J P 1982 An extension of Shapiro and Wilk's W test for normality to large samples *Applied Statistics* 31 115-124

[105] Russell T P, Lin J S, Spooner S and Wignall G D 1988 Intercalibration of small angle X-ray and neutron scattering data *Journal of Applied Crystallography* 21 629-638

[106] Sacks M S, Smith D B and Hiester E D 1997 A small angle light scattering device for planar connective tissue microstructural analysis *Annals of Biomedical Engineering* 25 678-689

[107] Sainsbury J R, Anderson T J and Morgan D A 2000 ABC of breast diseases: *Breast cancer British Medical Journal* 321 745-750

[108] Samuels M L 1989 *Statistics for the life sciences* Collier MacMillan Publishers, London

[109] Scarff R and Torloni H 1968 *Histological typing of breast tumours* International Histological Classification of Tumours, (Geneva: World Health Organisation) 13-20

[110] Seiki M 2003 Membrane-type 1 matrix metalloproteinase: A key enzyme for tumour invasion *Cancer Letters* 194 1-11

[111] Shapiro S S and Wilk M B 1965 An analysis of variance test for normality (complete samples) *Biometrika* 52 591-611

[112] Sidhu S, Siu K K, Falzon G, Hart S, Fox J and Lewis R A 2009 Systematic changes in SAXS features with distance away from the primary breast lesion *Medical Physics* (submitted October 2008)

[113] Sidhu S, Siu K K W, Falzon G, Hart S A, Fox J, Nazaretian S and Lewis R A 2009 Mapping of structural changes in breast tissue disease using X-ray scattering *Cancer Research* (submitted October 2008)

[114] Sidhu S, Siu K K W, Falzon G, Nazaretian S, Hart S, Fox J, Susil B and Lewis R A 2008 X-ray scattering for classifying tissue types associated with breast disease *Medical Physics* 35 4660-4670

[115] Silverberg S G, DeLellis R A, Frable W J, LiVolsi V A and Wick M R 2006 *Silverberg's principles and practice of surgical pathology and cytopathology* Elsevier B.V.

[116] Sionkowska A, Wisniewski M, Skopinska J, Kennedy C J and Wess T J 2003 Molecular interactions in collagen and chitosan blends *Biomaterials* 25 795-801

[117] Sonoda M, Takano M, Miyahara J and Kato H 1983 Computed radiography utilizing scanning laser stimulated luminescence *Radiology* 148 833

[118] Stevens A, Lowe J and Young B 2002 *Wheater's basic histopathology* Elsevier Science Limited, Churchill, Livingstone

[119] Suhonen H, Fernandez M, Serimaa R and Suortti P 2005 Simulation of small-angle X-ray scattering from collagen fibrils and comparison with experimental patterns *Physics in Medicine and Biology* 50 5401-5416

[120] Suortti P, Fernandez M and Urban V 2003 Comments on synchrotron fibre diffraction identifies and locates foetal collagenous breast tissue associated with breast carcinoma by V.J. James (2002) Journal of Synchrotron Radiation 9 71-76 *Journal of Synchrotron Radiation* 10 198

[121] Tartari A, Casnati E, Bonifazzi C and Baraldi C 1997 Molecular differential cross sections for x-ray coherent scattering in fat and polymethly methacrylate *Physics in Medicine and Biology* 42 2551-2560

[122] Towns-Andrews E, Berry A, Bordas J, Mant G R, Murray P K, Roberts K, Sumner I, Worgan J S and Lewis R A 1989 Time-resolved X-ray diffraction station: X-ray optics, detectors and data acquisition *Review of Scientific Instruments* 60 2346-2349

[123] Upton G and Cook I 2008 *A dictionary of statistics* Oxford University Press, Oxford

[124] Venables W N and Ripley B D 1998 *Modern applied statistics with S-plus* Springer-Verlag, New York

[125] Verkooijen H M, Peterse J L, Schipper M E I, Buskens E, Hendriks J H C L, Pijnappel R M, Peeters P H M, Borel Rinkes I H M, Mali W P T M and Holland R 2003 Interobserver variability between general and expert pathologists during the histopathological assessment of large-core needle and open biopsies of non-palpable breast lesions *European Journal of Cancer* 39 2187-2191

[126] Webb A R 2002 *Statistical pattern recognition* John Wiley and Sons, New Jersey

[127] Wess T, Hammersley A, Wess L and Miller A 1998 A consensus model for molecular packing of type I collagen *Journal of Structural Biology* 122 92-101

[128] Wilkinson S J and Hukins D W L 1999 Determination of collagen fibril structure and orientation in connective tissues by X-ray diffraction *Radiation Physics and Chemistry* 56 197-204

[129] Wilkinson S J, Rogers K D and Hall C J 2006 Model fitting in two dimensions to small angle diffraction patterns from soft tissue *Physics in Medicine and Biology* 51 1819-1830

[130] Wozniak M and Keely P J 2005 Use of three-dimensional collagen gels to study mechanotransduction in t47d breast epithelial cells *Biological Procedures Online* 7 144-161

[131] Yorke M A and Dickson D H 1985 A cytochemical study of myeloid bodies in the retinal pigment epithelium of the newt notophthalmus viridescens *Cell & Tissue Research* 240 641-648

[132] Zschenderlein U, Kampfe B, Schultrich B and Fritsche G 2007 Application of energy dispersive X-ray diffraction for the efficient investigation of internal stresses in thin films *Solid State Phenomena* 130 39-42

[133] Zweig M H and Campbell G 1993 Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine *Clinical Chemistry* 39 561-577

Supporting Documents

This section contains supporting documents which were published or submitted during the author's candidature. In chronological order:

Paper I: Sidhu S, Siu K K W, Falzon G, Nazaretian S, Hart S, Fox J, Susil B and Lewis R A 2008 X-ray scattering for classifying tissue types associated with breast disease *Medical Physics* **35** 4660-4670.

Paper II: Sidhu S, Siu K K, Falzon G, Hart S, Fox J and Lewis R A 2009 Systematic changes in SAXS features with distance away from the primary breast lesion *Medical Physics* (*submitted October 2008*)

Paper III: Sidhu S, Siu K K W, Falzon G, Hart S A, Fox J, Nazaretian S and Lewis R A 2009 Mapping of structural changes in breast tissue disease using x-ray scattering *Cancer Research (submitted October 2008)*

Paper I: X-ray Scattering for Classifying Tissue Types Associated with Breast Disease

X-ray scattering for classifying tissue types associated with breast disease

Sabeena Sidhu[a] and Karen K. W. Siu[b]
Monash Centre for Synchrotron Science, and School of Physics, Monash University, Clayton, Victoria 3800, Australia

Gregory Falzon[c]
School of Science and Technology, University of New England, Armidale, New South Wales 2351, Australia

Simon Nazaretian[d]
Department of Anatomical Pathology, The Royal Women's Hospital, Carlton, Victoria 3053, Australia, Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria 3800, Australia, and Department of Pathology, Monash Medical Centre, Clayton, Victoria 3168, Australia

Stewart A. Hart[e]
Department of Surgery, Monash Medical Centre, Clayton, Victoria 3168, Australia and Monash Breast Unit, Monash Medical Centre, Moorabin, Victoria 3165, Australia

Jane G. Fox[f]
Department of Surgery, Monash Medical Centre, Clayton, Victoria 3168, Australia and Monash Breast Unit, Monash Medical Centre, Moorabin, Victoria 3165, Australia

Beatrice J. Susil[g]
Department of Pathology, Monash Medical Centre, Clayton, Victoria 3168, Australia and Monash Breast Unit, Monash Medical Centre, Moorabin, Victoria 3165, Australia

Robert A. Lewis[h]
Monash Centre for Synchrotron Science, and School of Physics, Monash University, Clayton, Victoria 3800, Australia

(Received 2 March 2008; revised 5 August 2008; accepted for publication 7 August 2008; published 23 September 2008)

Collagen types I and III can be characterized at the molecular level (at the tens to hundreds of nanometers scale) using small angle x-ray scattering (SAXS). Although collagen fibril structural parameters at this length scale have shown differences between diseased and nondiseased breast tissues, a comprehensive analysis involving a multitude of features with a large (>50) patient cohort has not previously been investigated. Breast tissue samples were excised from 80 patients presenting with either a breast lump or reduction mammoplasty. From these, invasive carcinoma, benign tissue, and normal parenchyma were analyzed. Parameters related to collagen structure, including longitudinal (axial) and lateral (equatorial) features, polar angle features, total scattering intensity, and tissue heterogeneity effects, were extracted from the SAXS patterns and examined. The amplitude of the third-order axial peak and the total scattering intensity (amorphous scatter) showed the most separation between tissue groups and a classification model using these two parameters demonstrated an accuracy of over 95% between invasive carcinoma and mammoplasty patients. Normal tissue taken from disease-free patients (mammoplasty) and normal tissue taken from patients with presence of disease showed significant differences, suggesting that SAXS may provide different diagnostic information from that of conventional histopathology. © *2008 American Association of Physicists in Medicine.* [DOI: 10.1118/1.2977667]

Key words: SAXS, collagen, breast disease, synchrotron

I. INTRODUCTION

The diagnosis of invasive (or infiltrating) carcinoma is of management and prognostic importance since this type of neoplasm has invaded the basement membrane and the surrounding breast tissue. There are many indications that collagen is the medium in which invasion and growth of certain types of carcinomas thrive in the body.[1–4] Several studies have shown that abnormalities in the collagen structure, in particular the amount of degradation, are associated with invasive carcinoma in certain types of breast disease.[5,6] In particular, interaction of invasive carcinoma cells with the extracellular matrix (ECM) has been shown to induce increased production of proteins that degrade the ECM, enabling the invasive carcinoma cells to invade the surrounding tissues and progressively grow.[6–8] It is in the ECM, in which fibrillar collagen types I and III are dominant,[9] where the molecular changes in the fibrillar collagen can be examined for evidence of invasive carcinoma.

Small angle x-ray scattering (SAXS) can be used to characterize structures at the tens to hundreds of nanometer length scales appropriate to evaluate the structure of fibrillar collagens. It has been used to examine breast tissues[10–18] and can be described by the Bragg equation:

$$2D \sin \theta = n\lambda, \quad (1)$$

where D is the distance between the periodic scatterers, θ is the Bragg angle (half the total scattering angle) of the incident x-rays of wavelength λ, and n is the order of the scattering maxima.

Although there have been a few studies that have examined the use of SAXS for tissue identification,[10–13,17–21] none have looked at large patient cohorts that can be used for classification, modeling, and perhaps form the first steps toward a new diagnostic technique. It is not the purpose of this study to perform a structural analysis of collagen in the breast, but to determine any differences in the SAXS diffraction images between different types of breast tissue, using known parameters as a guide for comparison.

The collagen in breast tissue biopsy cores is not perfectly oriented, but is randomly distributed within the tissue core. The scattering pattern from randomly oriented cylinders is similar to the azimuthally averaged scattering pattern from systems with some degree of orientation.[22–24] This makes a full structural analysis complex and not practical for clinical diagnosis. However, collagen structural parameters appear evident on SAXS images[11,12,19] and properties of these components are well known.[23,25,26]

Because collagen fibrils are randomly oriented in breast tissues, a feature not commonly studied with SAXS is the degree of orientation in the fibrils.[17,18] Assuming preferentially oriented fibrils, intrafibrillar scattering of these fibrils can be approximated by a linear superposition of the scattering from individual fibrils.[17] Hence, the distribution of scattering intensities as a function of polar angle in the SAXS image provides information on how the collagen fibrils are orientated in the tissue sample, and if fibrils tend to align with a preferred orientation within the tissue. We have focused on the intensities within a small range (D spacing between 21 and 22 nm, or the third order axial d spacing peak) as a function of polar angle, so the distribution of fibril orientation can be studied.

Fibrillar collagen found in the breast has an axial d spacing of ~65 nm, which can change with the pathological state of the tissue.[11–13,19,20] Changes in the axial d spacing between healthy and malignant breast tissues has been attributed to degradation of the collagen fibrils, where the presence of cancerous cells between the collagen fibers seems to be a necessary condition for modification of the collagen structure.[6,20,27] In addition to longitudinal parameters, the diameter of collagen fibrils (~70 nm), the lateral distance between each fibril, and the degree of order in these structures are features that can also be examined for changes using SAXS.[11]

Two independent groups, both using similar experimental methods and synchrotron radiation, have used SAXS to examine the state of collagen fibrils in breast tissues in order to determine if differences can be attributed to breast disease: Lewis *et al.*[12] and Fernandez *et al.*[11,19,20] Lewis *et al.*[12] had 43 patients in their study, comprised of 18 normal, 18 invasive carcinoma, and 7 benign tissue samples. Fernandez *et al.*[11,19] had 8 patients in their study, comprised of 7 invasive carcinoma and 1 benign tissue sample, where histologically healthy tissue samples were additionally taken from the benign patient and one invasive carcinoma patient. The manner in which the SAXS data were acquired between the two studies were somewhat different: Lewis *et al.*[12] obtained SAXS images averaged over the whole tissue sample (i.e., one 2D scattering image for one patient), whereas Fernandez *et al.*[11,19] obtained SAXS images at many points in the tissue sample (1–0.25 mm spacings) and treated each resulting image as a separate observation corresponding to the pathology of the sample at that point (i.e., several images for one patient). Their tissue samples were fixed in formalin and pathological analysis done prior to SAXS imaging. Lewis *et al.*[12] imaged the tissue sample before fixing the piece in formalin and pathology reporting was done after the experiment. Furthermore, the beam size used by Lewis *et al.*[12] (diameter =0.5 mm) was larger than that used by Fernandez *et al.*[11,19] (diameter=0.2 mm). Breast tissue is intrinsically heterogeneous, consisting of a mixture of adipose tissue, fibrous tissue, and normal parenchyma. In order to evaluate if the differences observed by these two groups were due to their slightly differing methodologies, this study directly compared the two methods: (1) SAXS images taken point by point (at 1 mm intervals) through a tissue volume, and (2) an average SAXS image of the whole tissue volume.

Both studies suggest that the axial d spacing is larger in invaded tissues than in healthy or benign tissues and the small differences in the specific axial d spacing values between the two studies can perhaps be attributed to differences in experimental conditions or calibration. Fernandez *et al.*[11,19] found the axial d spacing of collagen embedded in invaded adipose tissue to be 65.3 ± 0.2 and 65.0 ± 0.1 nm in benign lesions and healthy connective tissue. Lewis *et al.*[12] found the axial d spacing of collagen for invasive carcinoma and normal tissues to be within the range of 64.7–65.22 nm, but benign tissues to have a smaller axial d spacing of between 64.4 and 64.7 nm. Both studies found that axial d spacing alone was not sufficient for differentiating breast disease with high sensitivity. Instead, other features, perhaps in combination with axial d spacing, may be more reliable for use in distinguishing differences between healthy and diseased tissues.

Fernandez *et al.*[11,19] found clear differences in the ordering and packing of collagen between healthy and invaded tissues in SAXS images. Scattering of collagen fibrils results in diffraction maxima related to the fibril diameter, interfibril distance, and ordering,[28,29] referred to herein as "equatorial features" or "lateral features." These maxima can be modeled by Bessel functions, which oscillate in a manner similar to sine waves damped with decreasing electron periodic density.[22] The "Bessel" maxima were relatively broad for invaded tissues, suggesting a range of fibril diameters; however the distances between the fibrils, indicated by the separation between the maxima, seemed to remain the same.

Another parameter examined by previous groups was the smooth scattering intensity between the identifiable diffraction features, often called amorphous scatter. The amorphous

scatter from malignant tissues has been shown to be higher than the amorphous scatter from normal tissues.[13,19] Rogers *et al.*[15] suggested that the amorphous scatter between $q = 0.11\ \mathrm{nm}^{-1}$ and $q = 0.23\ \mathrm{nm}^{-1}$ (where $q = 4\pi \sin\theta/\lambda = 2\pi/D$, using Bragg's law) differs in the presence of invasive carcinoma tissue. Fernandez *et al.*[19] looked at the range between $q = 0.50\ \mathrm{nm}^{-1}$ and $q = 0.56\ \mathrm{nm}^{-1}$ and found that the lower intensities corresponded to adipose, the medium intensities to healthy, slightly higher intensities to invaded, and the highest intensities to necrotic tissues. Lewis *et al.*[12] considered the fraction of scatter in the diffraction peaks (arising from the axial and lateral fibril features as discussed earlier) as a fraction of the total scatter (i.e., peak scatter plus amorphous scatter). They found that the fraction of total scattered intensity contained in the peaks was much less in invasive carcinoma tissue than in either mammoplasty tissue or benign lesions. This finding, coupled with the axial d spacing, was statistically significant in distinguishing among tissue types (mammoplasty, invasive carcinoma, benign).

The present study aimed to thoroughly examine a wide range of parameters using SAXS and to compare and evaluate these variables for their diagnostic potential. Moreover, we present this comprehensive study with a larger cohort of benign tissue samples as well as two types of normal tissue, not previously examined by other SAXS breast tissue studies. We aim to specifically determine which parameters, if any, can be used to differentiate between tissue types.

II. METHODS AND MATERIALS

II.A. Tissue samples

Core biopsy samples (MEAD 16 Gauge Needle Biopsy Gun) were excised from 80 patients and separated into four tissue groups based on histopathology analysis (by S.N.): 23 invasive carcinoma, 27 benign, 21 normal, and 9 mammoplasty. Tissues were taken from patients who had mammographically, ultrasonographically, or biopsy confirmed lesions that were subsequently treated by mastectomy or wide-local excision. Normal tissue samples were those excised from patients presenting with either a benign or invasive carcinoma lesion, but the tissue sample was subsequently diagnosed as "normal" by a pathologist. These tissue samples were close (within 2 mm) to the primary lesion and are histopathologically normal due to sampling error during tissue collection. Mammoplasty tissue samples were taken from patients undergoing breast reduction surgery; histopathology was used to confirm no presence of disease and patient records also confirmed no history of disease in these patients. Tissue core samples were typically cylindrical in shape with a diameter of 1 mm and lengths from 4 to 25 mm. All tissue samples were harvested within 20 min of surgery and snap frozen in liquid nitrogen for preservation and storage at −80 °C. Tissues were obtained with full informed consent from patients undergoing surgical treatment, under ethics approvals for this study [Southern Health: Human Research Ethics Committee B (Project No. 03094B) and Monash University: SCERH Human Ethics (Project No. 2003/021)].

Just before imaging, the tissue was defrosted and placed in a 1-mm-diam glass capillary tube (1.0 o.d. Special Glass capillaries, Wolfgang Muller Glas Technik, Germany) with phosphate-buffered saline solution to maintain hydration. These capillary tubes were then held vertically in a sample holder and aligned with the radiation beam and imaged using SAXS to obtain a diffraction pattern. Once imaged, the tissue was removed from the capillary tube and fixed in formalin for preservation until they could be embedded into paraffin blocks and stained with hematoxylin and eosin. Classification of tissues was based on the pathologist's diagnosis (by S.N.) postexperiment, and correlating patient information collected from each patient's surgical pathology report.

II.B. Image acquisition and processing

All diffraction images were recorded at Station 2.1 at the SRS Daresbury, UK, using a beam size of ~0.5 mm × 0.5 mm at the sample and a wavelength of 1.54 Å. Data acquisition and subsequent image processing were performed in a manner similar to Lewis *et al.*,[12] but is summarized here.

The sample-to-detector distance (L) was 6.3 m and data recorded using a 200 mm × 200 mm imaging multiwire proportional counter operated at 512 × 512 pixels.[30] Rat-tail tendon was used as a calibrant for D spacing, whilst glassy carbon [Sigradur® (K), HTW Hochtemperatur-Werkstoffe GmbH, Germany] was used as a standard to normalize for incident intensity variations. SAXS images were acquired at ~1 mm intervals over the length of each tissue sample and the series of images was summed to obtain an average diffraction pattern for the whole tissue, maintaining a constant total exposure time of 300 s. Two examples of diffraction images are shown in Fig. 1: (a) the average diffraction pattern for a mammoplasty tissue and (b) the average diffraction pattern for an invasive carcinoma tissue.

The diffraction data were processed to remove detector artifacts and corrected for the incident beam decay and sample attenuation (PCDETPAK32 6.0, R. Lewis, Monash University, 2007). The two-dimensional images were radially integrated to obtain the subsequent one-dimensional intensity data as a function of scattering vector $q = (4\pi \sin\theta/\lambda)$. The resulting 1D $I(q)$ data were multiplied by q^2 to facilitate peak fitting. A quadratic background was fitted to the intensity profiles and was removed before fitting Gaussian peaks using IGOR PRO (WaveMetrics Inc., 2005). Orientation distributions were obtained by expressing the data averaged over the q range between 0.265 and 0.310 nm^{-1} as a function of polar angle, where 0° was taken as the dominant diffraction axis. Images tended to demonstrate symmetry about the longitudinal axis due to the tendency of fibrils to align in the capillary, which was held vertically in the x-ray beam.

All statistical analysis and calculations were performed using R (Version 2.6.1, The R Foundation for Statistical Computing, 2007), MATLAB (Version 2.6.1 R2007a with Statistical Toolbox, MathWorks Inc., 2007), and Microsoft Office EXCEL (Microsoft Office Professional Edition 2003, Microsoft Corporation, 2003). Receiver operator characteristic

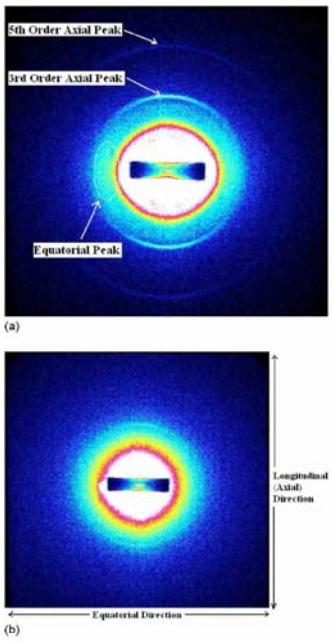


FIG. 1. SAXS diffraction images (2D) from (a) mammoplasty and (b) invasive carcinoma breast tissue biopsy core samples. In these images, the beam position is indicated by the maximum intensity in the semitransparent beamstop (dark rectangle at the center of the image), and the polar angle, φ, is defined as 0° in the longitudinal direction. Lateral features in the collagen fibrils are evident along the equatorial direction. The mammoplasty tissue has distinct third- and fifth-order axial reflections visible (the third-order axial reflection is visible with the invasive carcinoma, but faint). Both equatorial and longitudinal features are spread over a large polar angle range.

(ROC) curves were calculated using a freely available calculator.[31]

III. RESULTS

III.A. Point scans

Both Lewis *et al.*[12] and Fernandez *et al.*[11,19] found that the axial d spacing for breast fibrillar collagen was larger in invaded tissues compared to benign tissues; however, the specific values of these d spacings differed by up to 0.2 nm. This may be due to the methodology each group used to produce SAXS images of their tissue samples. In order to determine if the axial d spacing can be used as a diagnostic

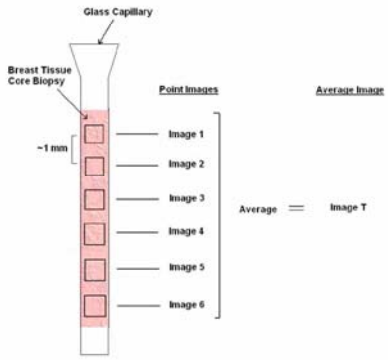


FIG. 2. Diffraction images were taken at 1 mm intervals through the length of the tissue core. Each of these images were analyzed separately (Point Images) and then each of these images were summed and averaged to obtain one average diffraction image representing the entire tissue core (average image).

indicator with some accuracy, SAXS data were analyzed for a subset of patient samples in two ways (illustrated in Fig. 2):

(1) Point images (Image #): several individual SAXS images taken at 1 mm increments in the tissue sample for one patient (as with Fernandez *et al.*[11,19]), and
(2) Average image (Image T): one SAXS image, integrated from all of the 1 mm increments, representing an average pattern over one patient sample (similar to Lewis *et al.*[12]).

This part of the study only examined the third-order axial d spacing peak of the collagen fibrils, as was done in these previous studies since it is the most prominent of the axial peaks in this q range.

Point scans of SAXS images were analyzed for 15 patients (6 invasive carcinoma, 4 benign, and 5 mammoplasty) to determine the effects of tissue heterogeneity with d spacing. Normal tissue (with history of disease) samples were excluded from this subset. Diffraction images were taken at 1 mm intervals over the length of each tissue sample. The number of images ranged from 4 to 19 per patient (mean =9) with 141 images in total. The third-order axial d spacing peaks were fitted and the axial d spacings calculated for all 141 images for the 15 patient samples.

The distribution of d spacing values within the point scans of each individual patient was investigated and then compared with those of all the patients, irrespective of tissue type (Table I). This suggested that there was a distinct difference in the range (i.e., the distribution) of d spacing values between different patients. Bartlett's test supported this finding which gave a p value of 5.7×10^{-10}. Closer examination,

TABLE I. p value comparison of the third-order axial d spacing of the point scan images.

	Irrespective of tissue type	Within each tissue type	
Intrapatient variability (within each patient)	5.7×10^{-10}	Invasive carcinoma	2.3×10^{-7}
		Benign	0.195
		Mammoplasty	1.1×10^{-5}
Interpatient variability (between each patient)	0.0045	Invasive carcinoma	0.158
		Benign	0.051
		Mammoplasty	0.667

now including tissue type as a discriminating variable, revealed that this significant variation was present in both the mammoplasty ($p=1.1\times10^{-5}$) and the invasive carcinoma ($p=2.3\times10^{-7}$) tissue groups but not in the benign group ($p=0.195$). The benign group appears to be unique in that it has a very homogeneous range of d spacing values irrespective of the patients. This suggests that benign tissue may have a more complex collagen structure compared to mammoplasty and invasive carcinoma tissue. There appears to be two different mechanisms generating the differences in d spacing properties for the different tissue types.

Interpatient variability, in addition to disease status, was also examined (Table I). The variation of the d spacing values of the point-scan data: Analysis of variance (ANOVA) of the d spacings between patients within each tissue group (mammoplasty, invasive carcinoma, benign) showed a p value of 0.0045. This result indicated that within at least one of the three tissue types, the magnitude of the d spacing varied significantly between patients. Further investigation revealed that it was only the benign group that had d spacings that varied significantly ($p=0.051$) between patients. The large variation in the d spacing within a sample can probably be attributed to the heterogeneous nature of tissue samples.

The use of ANOVA also revealed that the d spacing differed between tissue groups ($p=0.002$), and therefore may be used to distinguish between tissue types, when using point scan values. Therefore, d spacings at specific points in a tissue sample (as in Fernandez *et al.*[11,19]) are a possible means of identifying tissue types. When using the average d spacings for each patient (as used in Lewis *et al.*[12]), t tests were able to distinguish between invasive carcinoma and mammoplasty tissue ($p=0.02$) but not between invasive carcinoma and benign tissue ($p=0.78$). Thus, the d spacing averaged over an entire tissue sample can be used to separate between certain tissue groups (i.e., invasive carcinoma and mammoplasty), but is not as sensitive to the differences when compared to a point analysis.

In summary, each patient had their own characteristic collagen d spacing value and variation about that average value. The mean value of the d spacing was found to differ between patients of the benign tissue group. The range of d spacing values was different for patients in both the mammoplasty and invasive carcinoma tissue groups, but not in the benign tissue group. Therefore, there are distinct differences in how the d spacing value varies between patients in these tissue groups. These characteristics are expected to be influenced by biological factors, hormone therapy, chemotherapy and/or radiotherapy, genetics, and age. Analyzing the d spacing at each point in a tissue sample is more sensitive in distinguishing between tissue types than using the average d spacing over the tissue sample. However, the average can still be used as a diagnostic tool, but with caution.

III.B. Longitudinal (axial) variables

The third- and fifth-order intensity maxima for axial d spacing collagen diffraction peaks were easily visible in the data (seen in Fig. 1). The following characteristics were extracted from the fitted peaks for both the third- and fifth-order axial d spacing peaks: peak amplitude, peak full width at half-maximum (FWHM), peak center (D spacing), and peak area.

Figure 3 shows the average third-order axial peak d spacing for each tissue group with calculated 95% confidence intervals. The mammoplasty tissues are clearly separated from the invasive carcinoma tissues, whilst the peak centers for benign tissues are in between the normal/mammoplasty and invasive carcinoma tissue peak centers. The axial d spacing is smaller for mammoplasty tissue, $d=64.89\pm0.03$ nm, and larger compared to invasive carcinoma tissue, $d=64.97\pm0.05$ nm, where benign tissue is in between, $d=64.93\pm0.04$ nm (normal tissue, $d=64.88\pm0.04$ nm).

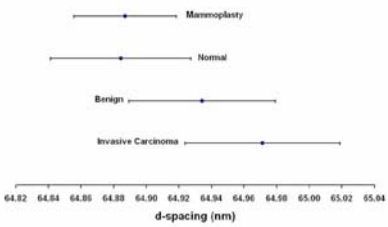


FIG. 3. Average of the centers of the third axial peak for the four tissue groups (axial d spacing). Error bars represent 95% confidence intervals. Normal tissues were diagnosed as such by a pathologist, but taken from patients with known presence of disease. Mammoplasty tissues were taken from patients with no presence or history of disease.

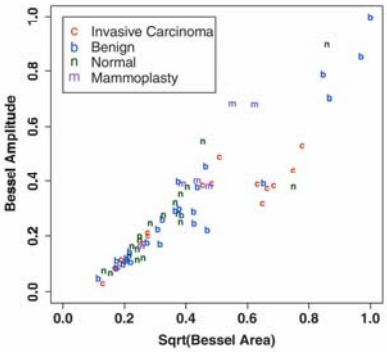


FIG. 4. The square root of the equatorial peak area plotted with the equatorial peak amplitude for each of the four tissue groups. Axis values have been normalized with the data.

Principal component analysis (PCA) revealed that the amplitudes of the axial peaks were the most significant variable of the axial parameters considered. The first principal component (PC1) accounted for 99.5% of the data and is given by the following linear combination of features:

$$\mathrm{PC1} = -0.582(\text{third amplitude}) - 0.813(\text{fifth amplitude}). \tag{2}$$

III.C. Lateral variables

The second harmonic Bessel (equatorial peak) was visible for 18 invasive carcinoma, 26 benign, 21 normal, and 6 mammoplasty patients. The amplitude, center, FWHM, and area of the fitted peak were determined and analyzed using PCA. The results showed that 99.8% of the variance in the data projected onto PC1 and it is given by the linear combination of features:

$$\mathrm{PC1} = 0.101(\text{equatorial peak area}) + 0.995(\text{equatorial peak amplitude}). \tag{3}$$

In order to determine if PC1 is affected by tissue diagnosis, a Kolmogorov—Smirnov test was performed on the data. The p values showed no evidence that tissue groups were significantly different with PC1 ($p>0.05$), with the exception of the mammoplasty and normal tissue samples ($p=0.030$). A Flinger (median) test on the data resulted in a p value=0.988, showing that the variance is homogeneous between tissue groups.

A linear relationship was observed between the square root of the equatorial peak area and the amplitude of the fitted equatorial peak (Fig. 4) and Table II lists the slope and R^2 values for a linear fit to each of the tissue groups. The slopes of the fitted lines showed a decrease from healthy tissue to progressively diseased tissues. The amplitude is representative of the relative amount of ordered fibrillar collagen in the tissue sample: the larger the amplitude, the more collagen present in the sample. The area encompasses both the height and width of the peak, where the width represents the distribution of how the fibrils are packed in the equatorial direction: A larger width means that there is more variability in the fibril packing structure than compared to a peak with a smaller width. Therefore, in Fig. 4, a larger slope means that there was more collagen present in the tissue sample that had less variability in the equatorial packing distance, i.e., the collagen fibrils were tightly packed together and it was abundant.

TABLE II. The linear fits calculated for the data shown in Fig. 4 for each of the tissue groups. The standard error (s.e.) of the slope is shown in parentheses.

Tissue group	Slope (s.e.)	R^2
Mammoplasty	6385 (1075)	0.87
Normal	5748 (492)	0.88
Benign	4211 (227)	0.93
Invasive carcinoma	2702 (297)	0.83

III.D. Total scattering intensity variables

A quadratic function was fitted to the modified 1D profile SAXS data for each patient after the peak features were subtracted from the profiles:

$$q^2 I(q) = B(q-A)^2 + C, \tag{4}$$

where $q=4\pi \sin\theta/\lambda$. The Lorentz correction factor of q^2 was applied as appropriate for isotropic samples. These coefficients (A, B, and C), as well as the total area of the fitted quadratic function (amorphous scatter), the total area of the fitted peaks of the collagen features, and the peak fraction of total diffraction (total area in the fitted peaks of the collagen features divided by the total area of the fitted quadratic function) were used for PCA of the scattering intensity features. Additionally, as was done in Fernandez *et al.*[19] (who examined the power-law exponent between $q=0.50\ \mathrm{nm}^{-1}$ and $q=0.56\ \mathrm{nm}^{-1}$, i.e., between the fifth- and sixth-order collagen axial peaks), the power-law exponent between $q=0.3\ \mathrm{nm}^{-1}$ and $q=0.5\ \mathrm{nm}^{-1}$ (i.e., between the third- and fifth-order collagen axial peaks), was also examined in this study.

PCA showed that the variable $-C$ encompassed all of the data and, thus, can be used to describe the entire data set. A Bartlett test between the tissue groups with the variable $-C$ gives a p value of 0.00023, suggesting a highly significant difference in the variances between the four tissue groups on this variable.

The differences between tissue groups and scattering intensities are further seen in Fig. 5, where the average intensity profiles are plotted from the SAXS diffraction images for the four tissue groups. The 95% confidence intervals (CI) indicate that the profiles of the invasive carcinoma and mammoplasty groups are separated. The average intensity profiles

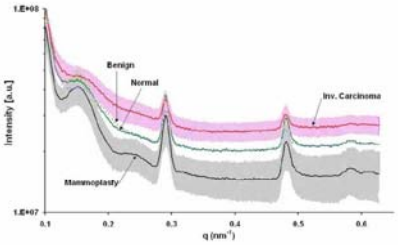


FIG. 5. Average intensity profiles for the four tissue groups over the observed q range. Confidence intervals of 95% are shown for the mammoplasty and invasive carcinoma profiles (shaded).

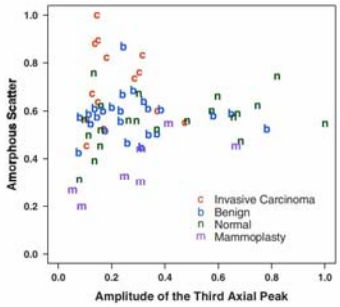


FIG. 6. The amplitude of the third axial peak in relation to the amorphous scatter for the four tissue types. Axis values have been normalized with the data.

for the benign and normal groups are approximately equal, and their 95%CI (not shown) coincide approximately with the lower and upper limits of the invasive carcinoma and mammoplasty profiles, respectively.

III.E. Polar angle distribution data

The SAXS images in Fig. 1 demonstrate the extended "arcs" corresponding to the axial d spacing peaks indicative of a low degree of preferential orientation of the collagen fibers and similarly demonstrate a large polar angle range. This relatively low degree of preferred orientation was typical of all samples, irrespective of tissue type. There is some evidence of fibers aligning in the longitudinal direction (despite their breadth, the arcs arising from the lateral direction are centred about a polar angle of 0°, or longitudinal in these images). It is likely that this effect is partially as a result of inserting the tissue samples into capillaries.[32]

Polar angular data output from the third-order axial d spacing peak maxima position (taken between $q = 0.265-0.310\ \mathrm{nm}^{-1}$) was obtained for all 80 patients from 0° to 360° and folded and averaged about 180° to 360°. From these 1D profiles, Gaussian curves were fitted and information regarding the FWHM, amplitude, center, and area were extracted. For patients showing a constant intensity over all theta in this q range (i.e., no particular orientation), a width of 180° was inserted into the data set. PCA was performed on these variables and it was found that the area of the fitted Gaussian represented 100% of the data. The area of a Gaussian is dependent upon both its amplitude and width. However, the amplitudes calculated for this polar angle data have essentially been analyzed in Sec. III B in the analysis of peak amplitudes associated with the longitudinal data. It is the width (FWHM) of the fitted distribution that provides information regarding the orientation of the collagen fibrils within the tissue sample. ANOVA showed no significant difference in the variances ($p=0.74$) between the tissue groups. Thus, the degree of orientation does not show a correlation with tissue types, suggesting that orientation of the fibrils may not be indicative of disease.

III.F. Classification

The key variables from each group of parameters determined by PCA as potentially useful classifiers were (1) the d spacing of the third-order axial peak, (2) the axial peak amplitude, (3) equatorial peak area and the equatorial peak amplitude, and (4) the total scattering intensity (i.e., predominantly the amorphous scatter).

Other parameters, notably the third axial peak intensity as a function of polar angle, did not show any difference between tissue groupings, thus removing these parameters as potential diagnostic factors. Of the variables (1)–(4), the axial peak amplitude and the amorphous scatter showed the most separation between tissue groups (Fig. 6).

A probability model was created using quadratic discriminant analysis (QDA) using the amplitude of the third-order axial d spacing peak and the amorphous scatter as classifiers. An example of the model is shown in Fig. 7, comparing the two extreme tissue groups: invasive carcinoma and mammoplasty. Here, the probabilities represent the tissue being an invasive carcinoma (e.g., probability of 0.8 means that the tissue has an 80% probability of being an invasive carcinoma, and 20% probability of being a mammoplasty tissue).

Figure 8 shows the probability model of invasive carcinomas against benign lesions. The ROC curves for the data in Figs. 7 and 8 are illustrated in Fig. 9. ROC curves were calculated using a probability of 0.6 or higher as being positive. Table III lists the ROC curve characteristics for each tissue group pairing.

The same analysis was performed using the equatorial peak area versus the amorphous scatter as well as the equatorial peak amplitude versus the amorphous scatter. The results were similar as seen in Figs. 6–8, verifying that it is in fact the amorphous scatter that is the dominant classifying variable in this model. The equatorial or axial peak feature respectively permits the construction of the probability plot

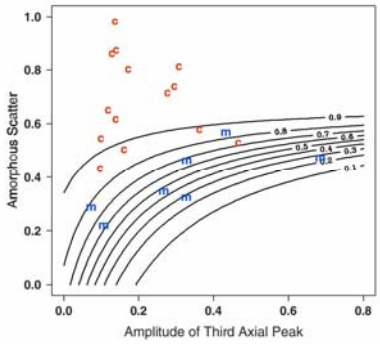


FIG. 7. Quadratic discriminant analysis classification of invasive carcinoma tissues (positive) with mammoplasty tissues (negative) (tissue from disease-free patients) using the amorphous scatter and the amplitude of the third axial collagen peak as classifiers. Axis values have been normalized with the data. Invasive carcinomas are labeled as "c" and mammoplasties are labeled as "m" and are overlayed onto the model.

and, in this sense, acts as an independent variable to permit discrimination between tissue types based on the amorphous scatter variable.

IV. DISCUSSION

Based on our data, there is a clear distinction between invasive carcinoma and mammoplasty tissue in terms of the axial *d* spacing, the amplitude of the third-order axial peak, and the amorphous scatter. However, features from benign and normal tissues are not as clearly separated, suggesting that the supramolecular structures in benign and normal tissues are similar. Interestingly, tissue from disease-free patients (mammoplasty) and normal tissue, sampled from patients with known presence of disease, seems to show differences. To date, there have been no SAXS studies that have specifically investigated the collagen fibril parameter differences between normal and mammoplasty tissues. From the pathologist's assessments, there is no obvious histological difference between the normal and mammoplasty tissue samples used in this study. However, our results suggest that there may be changes in the fibrillar collagen at the molecular level that may show evidence of disease, not usually seen with standard histopathological techniques, even though only normal parenchyma is seen in the biopsy core.

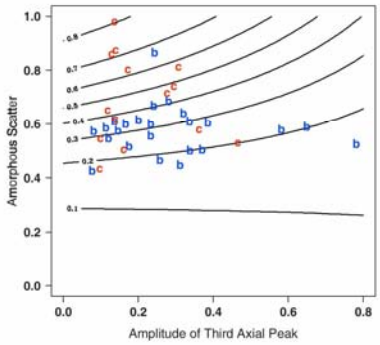


FIG. 8. Quadratic discriminate analysis classification of invasive carcinoma tissues (positive) with benign tissues (negative) using the amorphous scatter and the amplitude of the third axial collagen peak as classifiers. Axis values have been normalized with the data. Invasive carcinomas are labeled as "c" and benigns are labeled as "b" and are overlayed onto the model.

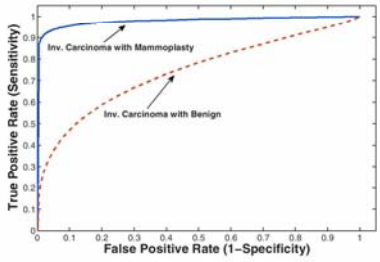


FIG. 9. ROC curves for (1) invasive carcinoma (positive) with mammoplasty (negative)—solid line—based on Fig. 7 model, and (2) invasive carcinoma (positive) with benign (negative)—dashed line—based on the model in Fig. 8. A threshold value of 0.6 was used (i.e., probability>0.6 is positive).

TABLE III. ROC curve characteristics for each pairing of tissue groups. The first listed tissue group was considered "positive" (e.g., invasive carcinoma) and the second tissue group considered "negative" (e.g., with mammoplasty).

Tissue comparisons	Sensitivity	Specificity	Accuracy	AUC (empiric)
Invasive carcinoma with mammoplasty	92.9	100	95.2	0.98
Benign with mammoplasty	95.8	71.4	90.3	0.95
Normal with mammoplasty	95.2	57.1	85.7	0.92
Invasive carcinoma with normal	64.3	85.7	77.1	0.81
Invasive carcinoma with benign	50.0	87.5	73.7	0.72
Benign with normal	83.3	47.6	66.7	0.72

This is perhaps most evident in the amorphous scattering intensity profiles (shown in Fig. 5), where the normal profile intensity is in between the invasive carcinoma and mammoplasty profile intensities, suggesting that these pathologically normal tissues may not be completely devoid of disease. All of the ROC curves that compared the mammoplasty patients against other tissue types showed high sensitivity (Table III), i.e., good true-positive rates for disease. For the invasive carcinoma versus mammoplasty tissues, however, the ROC curve (Fig. 9) showed both high sensitivity and high specificity: Large differences between these two tissue groups means that the QDA model subsequently showed good performance. Looking at the ROC area under the curves (AUCs), only the QDA models using the mammoplasty tissues showed high values (AUC$>$0.90). It is important to note that using this QDA model, SAXS was comparable with current screening and diagnostic techniques. Mammography, a widely used screening technique in developed countries, has sensitivities between 63% and 95% and specificities between 14% and 90%.[33–35] The lower specificities in mammography are due to early impalpable lesions detected during screening, and subsequent needle biopsy must be imaged guided to ensure proper sampling. Core biopsies were considerably better, with sensitivities between 88% and 98% and specificities between 95% and 100%.[33,34,36] Comparisons with invasive carcinoma tissues showed high specificity in the QDA model ($>$85%), even when the sensitivity was relatively low, demonstrating that the QDA model also performed adequately to identify true negatives, and was able to separate the noninvasive carcinoma tissues.

The importance of the amorphous scatter in distinguishing tissue types was also demonstrated by Fernandez *et al.*,[11,19] who found that the amorphous scatter (in adipose tissue) increases by a factor of 5 in invasive carcinoma tissues in comparison to normal tissues. Our results indicated an increase by a factor of 1.7 between mammoplasty and invasive carcinoma tissues, and an increase by a factor of 1.4 between benign/normal and invasive carcinoma tissues in the q range between 0.31 and 0.46 nm^{-1} (see Fig. 5). It is possible that invasive carcinoma samples have a more disordered tissue structure due to collagen degradation compared to mammoplasty tissues, since there is more amorphous scattering present in these samples. Interestingly, benign tissues have an amorphous scatter in between mammoplasty and invasive carcinoma tissues, suggesting that benign structures have some evidence of increased scattering potential.

The most significant changes in the fibrillar collagen features in the lateral and longitudinal directions seem to be between mammoplasty tissue and invasive carcinomas, which may imply that underlying mechanisms involved with the spread of disease is affecting the supramolecular nature of the collagen. This study found that the d spacing in tissues with invasive carcinoma (d=64.97 $\pm$ 0.05 nm) was larger than histopathologically normal (d=64.88 $\pm$ 0.04 nm) and benign (d=64.93 $\pm$ 0.04 nm) tissues, which agrees with the d spacing analysis done by Lewis *et al.*[12] However, our results showed a significant distinction in the d spacing between invasive carcinoma and mammoplasty (d=64.89 $\pm$ 0.03 nm) tissues, whereas Lewis *et al.*[12] found that the d spacings in mammoplasty tissues were not distinguishable from invasive carcinoma tissue. Fernandez *et al.*[11,19] did not examine mammoplasty tissues, but did find that the d spacing in tissue with invasive carcinoma was larger than in histopathologically normal tissues, which we have also confirmed with a much larger patient cohort. However, our point-scan analysis also showed that the d spacings in benign tissues had significant variance, and cannot be distinguished from the d spacings in invasive carcinoma tissues. A difference in d spacing further suggests that a closer examination of the processes involved in changing the molecular components of the collagen may reveal information regarding how neoplasms form and become invasive.

Although disease, whether benign or invasive, appears to have an effect on the molecular structure of the collagen, other factors may contribute to the differences we have seen between the tissue groups. Based on t tests, only the ages of the patients in the invasive carcinoma group [61 $\pm$ 12 years (s.d.)] were significantly different from the other groups (p $<$0.05). Age was not a factor between the normal (50 $\pm$ 14 years), benign (43 $\pm$ 15 years), and mammoplasty (46 $\pm$ 15 years) patients ($p$$>$0.12). Even though the patients in the invasive carcinoma group were older, it is important to note that the patients in the normal group were taken from patients presenting with benign and invasive carcinoma lesions, and that these normal patients had comparable ages to the mammoplasty patients. Thus, the techniques used in this study still have diagnostic potential despite the slight differences seen in the ages between the tissue groups. Other clinical differences that may affect the collagen structure in patients include: previous chemotherapy or radiotherapy, hormone replacement therapy, menopausal status, child bearing and breast feeding history, and trauma to the breast. This information was not sufficiently powerful with the sample size of 80 patients in order to perform a more detailed analysis.

An advantage of SAXS compared to histological core biopsy analysis is that the entire volume of the biopsy core can be analyzed, which in this study was up to 20 mm^3, where the tissue is fresh and unaltered from its natural state. Pathological analysis of breast tissue biopsies is usually based on chemically fixed and stained tissue slices taken from a core, typically 3–4 μm thick. Most tissue samples are heterogeneous, containing an array of normal, benign, and invasive carcinoma tissue. Some biopsy cores may only contain 5%–10% invasive carcinoma tissue, in which case it is diagnosed as invasive carcinoma when 90% of the core is normal parenchyma, as is obviously required in a clinical context.

Analyzing the entire biopsy sample using SAXS is not only a more efficient use of the tissue being examined, it also decreases the impact of any sampling error in the biopsy or sectioning processes. Since histopathologically normal samples from patients with a history of disease demonstrated SAXS parameters lying in-between those of invasive carcinoma samples and samples from disease-free patients, the

use of SAXS data may be much more tolerant to sampling error (e.g., not obtaining tissue directly from the lesion). Thus, whilst taking the average SAXS image over the entire tissue sample may not discern between the different tissue types within a biopsy core, it may still be capable of showing underlying changes in the molecular structures within the tissue which warrants further investigation for diagnosis.

SAXS information may be useful in conjunction with current diagnostic methods to assist in early detection or provide insight on the extent of disease with defining surgical margins. Our study is consistent with very preliminary results from Lewis *et al.*,[12] who examined tissue samples taken at 2 cm intervals outward from the invasive carcinoma site, and found an increase in the intensity of the diffraction peaks with distance away from the primary carcinoma site. We are currently mapping the changes in collagen fibril parameters as a function of distance from the center of the invasive carcinoma lesion in mastectomy patients, in order to determine if systematic differences can be seen over macroscopic distances (up to 10 cm from the primary lesion). This will provide further insight into the efficacy of using SAXS as a diagnostic tool for breast disease.

In summary, we have confirmed the diagnostic potential of SAXS for breast disease, using a broader sample of disease types and larger patient numbers that have previously been studied. We have shown that SAXS has the potential of being an important tool in diagnosing breast disease, and may show differences in tissue types not typically seen with current pathological techniques. In particular, we have shown that there are differences in the collagen fibrils present in mammoplasty (disease-free) tissues compared to pathologically normal tissues in patients with presence of disease.

As with other studies, we have found that the collagen fibril *d* spacing, the axial peak amplitude, and the area of the lateral peaks (equatorial peak features) were distinguishing parameters. The most significant factor was the amorphous scatter, which showed the best separation between tissue types. Invasive carcinoma tissues and mammoplasty tissues showed the most separation, as well as the best modeling parameters resulting in high sensitivity and specificity values between these two tissue types. Benign and normal (presence of disease) tissue parameters tend to lie between the invasive carcinoma and mammoplasty tissue parameters, suggesting that these properties need to be examined more closely to understand how disease contributes to the changing molecular structure of collagen fibrils in the breast.

ACKNOWLEDGMENTS

We acknowledge financial support from the *Access to Major Research Facilities Programme*, which is a component of the *International Science Linkages Programme* established under the Australian Government's innovation statement, *Backing Australia's Ability*. S.S. is grateful for scholarship support from the Monash Centre for Synchrotron Science. K.K.W.S. acknowledges the support of a Strategic Grant from the Faculty of Medicine, Nursing and Health Sciences, Monash University. The authors would like to thank the nurses and staff at Southern Health for their support and patience, in particular, Timothy Tsui and Peter Wells in the Department of Pathology, Moorabbin Centre. The authors acknowledge helpful discussions with Dr. Chris Hall, Dr. David Paganin, and Dr. Adam Round.

[a]Electronic mail: sabeena.sidhu@sci.monash.edu.au
[b]Electronic mail: karen.siu@sync.monash.edu.au
[c]Electronic mail: gfalzon@une.edu.au
[d]Electronic mail: simon.nazaretian@med.monash.edu.au
[e]Electronic mail: centreroad@ozemail.com.au
[f]Electronic mail: jane.fox@med.monash.edu.au
[g]Electronic mail: beatrice.susil@southernhealth.org.au
[h]Electronic mail: rob.lewis@sync.monash.edu.au

[1]H. Morgan and P. A. Hill, "Human breast cancer cell-mediated bone collagen degradation requires plasminogen activation and matrix metalloproteinase activity," Cancer Cell Int. **5**, 1–10 (2005).
[2]H. Tanjore and R. Kalluri, "The role of type iv collagen and basement membranes in cancer progression and metastasis," Am. J. Path. **168**, 715–717 (2006).
[3]Z.-S. Zeng, A. M. Cohen, and J. G. Guillem, "Loss of basement membrane type iv collagen is associated with increased expression of metalloproteinases 2 and 9 (mmp-2 and mmp-9) during human colorectal tumorigenesis," Carcinogenesis **20**, 749–755 (1999).
[4]X. H. Zhang and I. Takenaka, "Morphological changes in the basement membrane during progression from non-invasive to invasive rat bladder cancer induced by *n*-butyl-*n*-(4-hydoxybutyl) nitrosamine," BJU Int. **79**, 378–382 (1997).
[5]C. C. D'Arrigo and I. S. Fentiman, "Pathology of breast carcinoma," Int. J. Clin. Pract. **58**, 29–34 (2004).
[6]I. Pucci Minafra, S. Minafra, R. M. Tomasino, S. Sciarrino, and R. Tinervia, "Collagen changes in the ductal infiltrating (scirrhous) carcinoma of the human breast. A possible role played by type i trimer collagen on the invasive growth," J. Submicrosc. Cytol. **18**, 795–805 (1986).
[7]S. Minafra, C. Giambelluca, M. Andriolo, and I. Pucci-Minafra, "Cell-cell and cell-collagen interactions influence gelatinase production by human breast-carcinoma cell line 8701-bc," 62, 777–783 (1995).
[8]I. Pucci Minafra, C. Luparello, S. Sciarrino, R. M. Tomasino, and S. Minafra, "Quantitative determination of collagen types present in the ductal infiltrating carcinoma of human mammary gland," Cathet Cardiovasc. Diagn. **9**, 291–296 (1985).
[9]S. Kauppila, F. Stenback, J. Risteli, A. Jukkola, and L. Risteli, "Aberrant type i and type iii collagen gene expression in human breast cancer in vivo," J. Pathol. **186**, 262–268 (1998).
[10]G. Falzon, S. Pearson, R. Murison, C. J. Hall, K. K. W. Siu, A. Evans, K. D. Rogers, and R. A. Lewis, "Wavelet-based feature extraction applied to small angle x-ray scattering patterns from breast tissue: A tool for differentiating between tissue types," Phys. Med. Biol. **51**, 2465–2477 (2006).
[11]M. Fernandez, J. Keyrilainen, R. Serimaa, M. Torkkeli, M.-L. Karjalainen-Lindsberg, M. Tenhunen, W. Thomlinson, V. Urban, and P. Suortti, "Small-angle x-ray scattering studies of human breast tissue samples," Phys. Med. Biol. **47**, 577–592 (2002).
[12]R. A. Lewis, K. D. Rogers, C. J. Hall, E. Towns-Andrews, S. Slawson, A. Evans, S. E. Pinder, I. O. Ellis, C. R. M. Boggis, A. P. Hufton, and D. R. Dance, "Breast cancer diagnosis using scattered x-rays," J. Synchrotron Radiat. **7**, 348–352 (2000).
[13]A. R. Round, S. J. Wilkinson, C. J. Hall, K. D. Rogers, O. Glatter, T. Wess, and I. O. Ellis, "A preliminary study of breast cancer diagnosis using laboratory based small angle x-ray scattering," Phys. Med. Biol. **50**, 4159–4168 (2005).
[14]P. Fratzl, N. Fratzl-Zelman, and K. Klaushofer, "Collagen packing and mineralization. An x-ray scattering investigation of turkey leg tendon," Biophys. J. **64**, 260–266 (1993).
[15]K. D. Rogers, R. A. Lewis, C. J. Hall, E. Towns-Andrews, S. Slawson, A. Evans, S. E. Pinder, I. O. Ellis, C. R. M. Boggis, and A. Hufton, "Preliminary observations of breast tumour collagen," Synchrotron Radiat. News **12**, 15–20 (1999).
[16]T. Wess, A. Hammersley, L. Wess, and A. Miller, "A consensus model for molecular packing of type i collagen," J. Struct. Biol. **122**, 92–101 (1998).
[17]S. J. Wilkinson and D. W. L. Hukins, "Determination of collagen fibril

structure and orientation in connective tissues by x-ray diffraction," Radiat. Phys. Chem. **56**, 197–204 (1999).
[18]S. J. Wilkinson, K. D. Rogers, and C. J. Hall, "Model fitting in two dimensions to small angle diffraction patterns from soft tissue," Phys. Med. Biol. **51**, 1819–1830 (2006).
[19]M. Fernandez, J. Keyrilainen, M.-L. Karjalainen-Lindsberg, M. Leidenius, K. Von Smitten, S. Fiedler, and P. Suortti, "Human breast tissue characterisation with small-angle x-ray scattering," Spectroscopy (Eugene, Or.) **18**, 167–176 (2004).
[20]M. Fernandez, J. Keyrilainen, R. Serimaa, M. Torkkeli, M.-L. Karjalainen-Lindsberg, M. Leidenius, K. v. Smitten, M. Tenhunen, S. Fiedler, A. Bravin, T. M. Weiss, and P. Suortti, "Human breast cancer in vitro: Matching histo-pathology with small-angle x-ray scattering and diffraction enhanced x-ray imaging," Phys. Med. Biol. **50**, 2991–3006 (2005).
[21]V. Changizi, M. A. Oghabian, R. Speller, S. Sarkar, and A. A. Kheradmand, "Application of small angle x-ray scattering (saxs) for differentiation between normal and cancerous breast tissue," Int. J. Med. Sci. **2**, 118–121 (2005).
[22]E. F. Eikenberry, B. Brodsky, and D. A. D. Parry, "Collagen fibril morphology in developing chick metatarsal tendons 1. X-ray diffraction studies," Int. J. Biol. Macromol. **4**(6), 322–328 (1982).
[23]L. A. Feigin and D. I. Svergun, *Structure Analysis by Small-Angle X-Ray and Neutron Scattering* (Plenum, New York, 1987).
[24]G. Oster and D. P. Riley, "Scattering from cylindrically symmetric systems," Acta Crystallogr. **5**, 272–276 (1952).
[25]E. F. Eikenberry and B. Brodsky, "X-ray diffraction of reconstituted collagen fibers," J. Mol. Biol. **144**(3), 397–404 (1980).
[26]R. D. B. Fraser, T. P. MacRae, A. Miller, and E. Suzuki, "Molecular conformation and packing in collagen fibrils," J. Mol. Biol. **167**, 497–521 (1983).
[27]M. J. Duffy, T. M. Maguire, A. Hill, E. McDermott, and N. O'Higgins, "Metalloproteinases: Role in breast carcinogenesis, invasion and metastasis," Breast Cancer Res. Treat **2**, 252–257 (2000).
[28]B. Brodsky, E. F. Eikenberry, K. C. Belbruno, and K. Sterling, "Variations in collagen fibril structure in tendons," Biopolymers **21**(5), 935–952 (1982).
[29]D. J. Hulmes and A. Miller, "Quasi-hexagonal molecular packing in collagen fibrils," Nature (London) **282**, 878–880 (1979).
[30]R. A. Lewis, K. D. Rogers, C. J. Hall, E. Towns-Andrews, S. Slawson, A. Evans, S. Pinder, I. O. Ellis, C. R. M. Boggis, A. Hufton, and D. R. Dance, "Preliminary observations of breast tumour collagen," Proc. SPIE **3770**, 32–37 (1999).
[31]J. Eng, www.jrocfit.org, John Hopkins University, Baltimore, MD, 2007.
[32]G. J. Ooi, J. Fox, K. K. W. Siu, R. A. Lewis, K. R. Bambery, D. McNaughton, and B. R. Wood, "Fourier transform infrared imaging and small angle x-ray scattering as a combined biomolecular approach to diagnosis of breast cancer," Med. Phys. **35**, 2151–2161 (2008).
[33]W. A. Berg, L. Gutierrez, M. S. NessAiver, W. B. Carter, M. Bhargavan, R. S. Lewis, and O. B. Ioffe, "Diagnostic accuracy of mammography, clinical examination, us, and mr imaging in preoperative assessment of breast cancer," Radiology **233**, 830–849 (2004).
[34]L. Irwig, N. Houssami, and C. van Vliet, "New technologies in screening for breast cancer: A systematic review of their accuracy," Br. J. Cancer **90**, 2118–2122 (2004).
[35]T. M. Kolb, J. Lichy, and J. H. Newhouse, "Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: An analysis of 27,825 patient evaluations," Radiology **225**, 165–175 (2002).
[36]J. G. Elmore, K. Armstrong, C. D. Lehman, and S. W. Fletcher, "Screening for breast cancer," JAMA, J. Am. Med. Assoc. **293**, 1245–1256 (2005).

Paper II: Mapping of Structural Changes in Breast Tissue Disease Using X-ray Scattering

Authors (in order):

Sabeena Sidhu
Monash Centre for Synchrotron Science, Monash University, Clayton, VIC Australia, 3800.
School of Physics, Monash University, Clayton Campus, VIC Australia 3800
Sabeena.Sidhu@sci.monash.edu.au
+613 9905 8063

Karen KW Siu
Monash Centre for Synchrotron Science, Monash University, Clayton, VIC Australia, 3800.
School of Physics, Monash University, Clayton Campus, VIC Australia 3800
Karen.Siu@sync.monash.edu.au
+613 9905 4922
+613 9905 3637

Gregory Falzon
School of Science and Technology, University of New South Wales, Armidale, NSW Australia 2351
gfalzon@une.edu.au
+61 0409 749 154

Stewart A Hart
Department of Surgery, Monash University, Clayton, VIC Australia 3800.
Monash Breast Unit, Monash Medical Centre, Moorabin, VIC Australia 3165
centreroad@ozemail.com.au
+613 9579 6122

Jane G Fox
Department of Surgery, Monash University, Clayton, VIC Australia 3800.
Monash Breast Unit, Monash Medical Centre, Moorabin, VIC Australia 3165
Jane.Fox@med.monash.edu.au
+613 9579 6122

Robert A Lewis
Monash Centre for Synchrotron Science, Monash University, Clayton, VIC Australia, 3800.
School of Physics, Monash University, Clayton Campus, VIC Australia 3800
Rob.Lewis@sync.monash.edu.au
+613 9905 3622
+613 9905 3637

Abstract

Spread of invasive carcinoma throughout breast tissue is believed to occur at supramolecular levels, beyond the range of standard histopathology identification. Small Angle X-ray Scattering (SAXS) is a novel technique that is capable of characterising the structural properties of collagen found in the breast. 31 patients who were treated with wide local excision or mastectomy had tissue core samples analysed at 2 cm intervals along two perpendicular axes over their excised mass. Parameters were extracted from SAXS images and analysed to determine if there were systematic differences in supramolecular structures with distance away from the primary site of the lesion. The amorphous scatter and the third order collagen axial d-spacing showed significant differences between the centre (0 cm) and distant tissues ($p < 0.01$). Mapping of these two variables over a 2-dimensional grid showed good matching with independent histopathology diagnosis, and may be capable of identifying areas of invasion or directional spread of disease.

Keywords: SAXS, collagen, breast cancer

1. Introduction

Small Angle X-ray Scattering (SAXS) is capable of observing changes in breast tissue at the nanometre scale, several magnitudes beyond standard histopathology analysis, making SAXS a potentially more sensitive diagnostic technique[1-8]. The degradation of the collagen fibrils in breast tissue is known to occur with infiltrating disease[30, 97, 113, 116]. Systematic variations in the structural properties of breast tissue can be determined over the entire breast which can lead to a better understanding of the spread of invasive disease.

Earlier studies have observed some indications of systematic differences in the SAXS patterns with distance away from the primary site[6, 7]. Lewis *et al.* [88] collected tissue samples at 2 cm intervals away from the primary lesion and analysed these with SAXS. These tissue samples showed variation in the intensity of the diffraction peaks with distance, but nothing could be definitively concluded with their sample size. Round[123] also found evidence of tissue differences up to 6 cm away from the primary site in both the peak areas of the collagen axial features and in the intensity ratio. However, no conclusive evidence that these features showed systematic changes with distance and tissue types could be confirmed with their sample size.

Several groups have found distinctions between invasive carcinoma, benign, and normal breast tissue types using SAXS[3, 4, 6, 7, 13], where these differences were statistically significant. In our recent study[134], we showed that SAXS can determine differences between pathologically normal tissue (breast tissue diagnosed as normal, but from a patient with known disease) and disease-free normal tissue (breast tissue taken from reduction/mammoplasty procedures, with no history or presence of disease). This raises the possibility that SAXS may be sensitive enough to detect malignancies even when conventional histopathology returns a negative result. This would be of great importance to diagnostic and surgical procedures, especially where sampling error is a concern[6, 14, 148].

This is the first study with systematic sampling of breast tissue with a large cohort of patients where SAXS images were taken from the centre of the primary lesion and at regular intervals away from this centre. Analysis of the tissue samples from the primary lesions has been presented in Sidhu *et al.*[134]. The variable that showed the most correlation with disease type was the amorphous scatter[134]. An increase in the

amorphous scatter over a wide range suggests that tissue components have been modified over a broad range of length scales[49-51, 88, 134]. In these studies, we believe this is most likely a result of tissue degradation in the presence of disease. Four other variables were also analysed to see if changes were evident with distance: third order axial peak width, amplitude, peak area, and d-spacing. These third order axial peak width, peak amplitude and peak area have been shown in previous studies[18, 45, 49, 51, 88, 123, 134] to be correlated with disease, however, based on our group's previous analysis of the primary site tissues[134], we have focussed on the amorphous scatter and axial D-spacing in this study. Using these variables, structural differences in the tissue fibres of the breast may be characterised with distance away from the centre of a palpable lesion.

2. Methods

2.1 Tissue Samples

Tissues samples were taken from patients who had mammographic, ultrasonographic, or biopsy confirmed lesions that were subsequently treated by mastectomy or wide-local excision. Core biopsy samples (MEAD 16 Gauge Needle Biopsy Gun) were excised from the centre of the primary lesion, then at two centimeter intervals radially outwards along two perpendicular axes (see Figure 1). The N-axis was determined as the line between the nipple and axilla, and the P-axis was taken perpendicular to the N-axis. (+2), (+4), (+6) notation represented samples taken towards or close to the nipple (N-axis) or towards the superior part of the breast (P-axis). (-2), (-4), (-6) notation represented samples taken towards the axilla (N-axis) or towards to the inferior part of the breast (P-axis). The maximum distance of a tissue sample was ~ 6 cm away from the centre of the primary lesion; thus, a maximum of 13 tissue samples could be collected from any one patient. However, 24 patients were treated with wide-local excisions (7 patients were treated with mastectomy), where sometimes the maximum distance at which samples could be harvested were only 2 cm or 4 cm away from the primary site. Tissue samples were taken after surgical removal of the mass. This made it challenging to make accurate distance measurements from the primary site, because the excised tissue mass is difficult to handle and position. Therefore, the measurements of 2 cm, 4 cm, and 6 cm are only approximate intervals (± 1 cm).

Data from 31 patients (median age 57 ± 11 years) with a total of 165 tissue samples were analyzed in this study: 31 samples at 0 cm, 81 samples at 2 cm, 36 samples at 4 cm, and 17 samples at 6 cm. Tissue core samples were typically cylindrical in shape with a diameter of 1 mm and lengths from 4 mm to 25 mm. All tissue samples were harvested within 20 minutes of surgery and snap frozen in liquid nitrogen for preservation and storage at -80°C. Tissues were obtained with full informed consent from patients undergoing surgical treatment, under ethics approvals for this study (Southern Health: Human Research Ethics Committee B (Project No. 03094B) and Monash University: SCERH Human Ethics (Project No. 2003/021)).

Just before SAXS imaging, the tissue was defrosted and placed in a 1 mm diameter glass capillary tube (1.0 O.D. Special Glass capillaries, Wolfgang Muller Glas Technik, Germany) with phosphate-buffered saline solution to maintain hydration. These capillary tubes were then held vertically in a sample holder and aligned with the radiation beam and imaged to obtain a SAXS pattern. Once imaged, the tissue was removed from the capillary tube and fixed in formalin for preservation until they could be embedded into paraffin blocks and stained with hematoxylin and eosin (H&E). Classification of tissues was based on the pathologist's diagnosis post-experiment, and correlating patient information collected from each patient's surgical pathology report. Most of the patients (27/31) presented in this study had pathologically confirmed invasive carcinoma present in the tissue sample from the palpable, primary site lesion (i.e. tissue core taken at 0 cm).

Subsequent tissue samples (tissue samples taken radially outward from the primary site) were typically diagnosed as normal (103), however, there were samples diagnosed with benign disease (5) and invasive carcinoma (total 57, non-primary site = 30).

2.2 Image Acquisition and Processing

All diffraction images were recorded at Station 2.1 at the SRS Daresbury, UK, over three experimental sessions. A beam size of ~0.5 mm x 0.5 mm was used at the sample with a wavelength of 1.54 Å. Data acquisition and subsequent image processing were performed as discussed in Sidhu *et al.*[134] and Lewis *et al.*[88], but is summarised here.

The sample-to-detector distance (L) was 6.3 m and data recorded using a 200 mm x 200 mm imaging multi-wire proportional counter operated at 512 x 512 pixels. Rat-tail tendon was used as a calibrant for d-spacing whilst glassy carbon (Sigradur® (K), HTW Hochtemperatur-Werkstoffe GmbH, Germany) was used as a standard to normalize for intensity variations. SAXS images were acquired over the length of each tissue sample by oscillating the tissue sample vertically during the 300 seconds of exposure.

The diffraction data was processed to remove detector artifacts and corrected for the incident beam decay and sample attenuation (PCDetpak32 6.0, Rob Lewis, Monash University, 2007). The 2-dimensional images were radially integrated to obtain the subsequent 1-dimensional intensity data as a function of scattering vector $q = (4\pi sin\theta/\lambda)$. The resulting 1-dimensional $I(q)$ data was multiplied by q^2 to facilitate peak fitting. A quadratic background was fitted to the intensity profiles and was removed before fitting Gaussian peaks using IGOR Pro (WaveMetrics Inc. 2005). Images tended to demonstrate symmetry about the longitudinal axis due to the tendency of fibrils to align in the capillary, which was held vertically in the X-ray beam.

All statistical analysis and calculations were performed using R (Version 2.6.1, The R Foundation for Statistical Computing, 2007), MatLab (Version 2.6.1 R2007a with Statistical Toolbox, MathWorks Inc. 2007), and Microsoft Office Excel (Microsoft Office Professional Edition 2003, Microsoft Corporation 2003).

3. Results

3.1 Data Analysis of All Samples

All data was analysed together, irrespective of patient or tissue type, in order to determine if there was a general trend in parameters with distance. Data was grouped by distance, with 0 as the centre of the primary lesion and then up to four tissue samples from each patient at 2 cm intervals (for example, N(+2) N(-2) P(+2) and P(-2)). The mean values with respective 95% confidence intervals (CI) for the amorphous scatter and axial D-spacing are shown in Figure 2 for the distance intervals. Both variables show a marked decrease in value from the centre of the lesion to 2 cm away, then a plateau between 2 cm and 6 cm. In order to determine if the difference between the centre point and the interval points were significant, p-values were calculated and are shown in Table 1 for the amorphous scatter and axial D-spacing variables. The amorphous scatter values show the most significant difference ($p = 2.36 \times 10^{-15}$) between the centre and each of the distant points. The axial D-spacing also shows a consistent significant differences, except for the centre with the 6 cm points ($p = 0.081$). All other comparisons (third order axial peak amplitude, peak width, and peak area), including between the distant points (for example, between 2 cm and 4 cm), had $p > 0.05$, and therefore showed no significant differences.

3.2 Directional Analysis

(nipple-axilla) and the P-axis (perpendicular to the nipple-axilla line) for the amorphous scatter variable. The axial D-spacing showed similar results compared to the grouped analysis of all of the samples in Figure 2 (a downward trend from the centre to 2 cm, then approximately flat from 2 cm to 6 cm). There was still a significant difference between amorphous scatter values between the centre and distant tissues ($p < 0.0001$), however, there were no significant differences between each of the distant tissues ($p > 0.01$) (for instance, between 2 cm and 4 cm), as was similar to the analysis of all of the tissues in Section 3.1.

3.3 Individual Patient Data Analysis

A two-dimensional spatial model was used to visualise the values of the amorphous scatter and axial D-spacing parameters for each patient. The parameters were determined for each tissue sample and mapped onto a 2-dimensional grid of 32 x 32 pixels. Each patient's amorphous scatter data was normalized for each model (i.e. a value assigned to lie between 0 and 1). The axial D-spacing values were maintained as absolute values (nm^{-1}). Gaussian Kernel estimation was used to calculate the missing values within the grid, between the known data points, implemented using a Fast-Fourier Transform (FFT)[149].

The 2D models for two sample patients who had full mastectomy data (i.e. these patients had close to the maximum 13 samples possible) are shown in Figure 4 using the amorphous scatter (Figures 4a and 4b) and the axial peak D-spacing (Figures 4c and 4d). The histopathological diagnosis for each of the tissue samples is overlaid for independent comparison. Using this model, light shades (white, large values) indicate areas of invasive carcinoma, and darker shades (red, lower values) indicate areas of normal tissue. On first examination, the model agrees well with the histopathological diagnosis. This was also consistent with the other 29 patients studied.

Patient 1 (Figures 4a and 4c) is typical of those patients with centrally located (on or near the nipple) lesions with axillary dissection. Large values of both the amorphous scatter and D-spacing are seen at the centre (0 cm) and decrease towards the margins. Patient 2 (Figures 4b and 4d) shows good agreement with histopathology diagnosis and the model with the amorphous scatter, where disease appears in the N(+) region and on the P(+) region. The axial D-spacing for this patient (Figure 4d) suggests that disease is mainly in the P(+) region, showing that the origin of disease may be at the P(+6) area, rather than at the centre. However, the model in Figure 4d does not show the confirmed disease in the N(+) and P(-) regions as well as the amorphous scatter model shown in Figure 4b. This agrees with the p-values calculated for the amorphous scatter and the axial D-spacing in Table 1, which suggest that amorphous scatter is a more powerful discriminator of disease status.

Discussion

Specific trends in the parameters are evident when inspecting the values grouped together for the entire population, as well as with individual patient analysis. The SAXS amorphous scatter and the axial D-spacing demonstrate systematic differences between tissues on the molecular and supra-molecular level with distance from the centre of the lesion. There is an obvious change in these two components along the two axes studied. The amorphous scatter and axial D-spacing values compared with histopathology analysis agree with our group's previous results[134] as well as with other studies[18, 45, 49, 51, 88, 123, 134]. In particular, the amorphous scatter variable has proven to be a useful variable in classification models[3, 4, 13, 18] and has shown the same results in this analysis. Furthermore, the amorphous scatter is consistent in both the grouped data analysis and individual patient analysis, suggesting that this may be an important component in distinguishing spread of disease within the breast.

Changes in breast tissue structure seem to be occurring significantly from the site of the palpable lesion to the normal tissue occupying the margins around the lesion. A decrease in the amorphous scatter value from 0 cm to 2 cm, seen in Figure 2a, suggests that invaded tissue has a higher specific surface area of scatterers compared to tissue 2 cm and beyond. This may be caused by a disruption of the molecular and supra-molecular structures in cancerous regions, indicative of inhomogeneous structures due to the invasion of cancer within the tissue.

Collagen molecules pack laterally to each other with hydrogen bonds and bind to each other in a staggered arrangement longitudinally. This structure type is repeated periodically along the fibril axis. The diffraction peaks seen on SAXS images give information on the periodic axial stacking of the microfibril arrangement of the collagen. The positions of these peaks, or the axial D-spacings, are larger for malignant tissue than for healthy tissue, which is interpreted as an increase in the axial periodicity with malignancy[49, 51, 88, 134]. Figure 2b shows that the periodic axial stacking for the third ordered peak (D-spacing) is larger at the centre of the primary site (0 cm) where these tissue samples have confirmed invaded tissue present. It appears that the periodicity of the axial stacking of collagen fibrils in breast tissue decreases with distance away from the primary site of a diseased lesion.

The size of the tumour may affect the trends shown in Figure 2, where some patient's tumours may be more than 4 or 6 cm in diameter. There were 17 patients who had a lesion size reported in their pathology report. In 6/17 cases, the size of the lesion was < 2 cm and in 5/17 cases the size of the lesion was between 2 cm and 4 cm. The other six patients who had data available presented with lesions > 4 cm (largest, 7.5 cm). Although general trends seem to be evident, further work is needed in order to identify the specific gradients occurring within the volume of the palpable lesion as well as at the margins of known invasive disease with normal parenchyma. For this reason, individual patient analysis using these 2D models specific to each patient may provide more reliable information on the differences in parameters with distance from the centre of the primary lesion.

The simple model presented here has been shown to correlate with histopathology analysis and may even provide further information on disease site and extent within the breast, even though only a relatively sparse sampling grid was used. A more accurate model can be developed if more tissue samples were extracted over the area of the surgical mass. Instead of sampling along two axes, tissue harvesting at smaller intervals (every 1 cm) over a grid may show multiple disease foci and better discrimination of marginal disease. A disadvantage of the model is that multiple samples must be taken from the patient so that certain periodicity assumptions can be met, as necessitated by use of kernel estimation on a finite grid using the Fourier Transform, in order to avoid wrap-around effects. This was seen with patients who had a small number of data points available for computation of their 2D maps. It was observed empirically that more than 4 tissue samples were needed for the model to generate map free of artefacts. These can be avoided, however, if the 2D grid is re-calculated with zero-padding on the boundaries. However, interpretation of diseased margins and parameter gradients over the 2D grid would need to be considered more carefully in these zero-padded data sets.

We have found that there are distinct changes occurring in breast tissue between the centre of the palpable diseased lesion and up to 6 cm away. With a sufficiently large sample number for each patient, our model is capable of mapping changes in the amorphous scatter and axial D-spacing values which can be compared to histopathology analysis of these tissues. Visualisation of these changes over the entire excised mass may show presence of disease where standard histology cannot, and moreover appears to identify pathology accurately in regions between sampling sites. The use of this modelling

technique along the margins of excised tissue would, for example, greatly aid in the determination of surgical margin removal and may lead to the evaluation of a patient's risk of metastatic recurrence. Breast tissue analysis using SAXS in conjunction with current diagnostic methods may prove advantageous in determining areas and directional spread of disease as well as provide a more cautious approach towards treatment of patients normally given clear surgical margins.

Acknowledgements
We acknowledge financial support from the *Access to Major Research Facilities Programme* which is a component of the *International Science Linkages Programme* established under the Australian Government's innovation statement, *Backing Australia's Ability*. SS is grateful for scholarship support from the Monash Centre for Synchrotron Science. KKWS acknowledges the support of a Strategic Grant from the Faculty of Medicine, Nursing and Health Sciences, Monash University. The authors would like to thank the nurses and staff at Southern Health for their support and patience, and in particular, Dr. Beatrice Susil in the Department of Pathology. Also, Simon Nazaretian for his assistance with histopathology. The authors acknowledge helpful discussions with Dr. Chris Hall, Dr. David Paganin, and Dr. Adam Round.

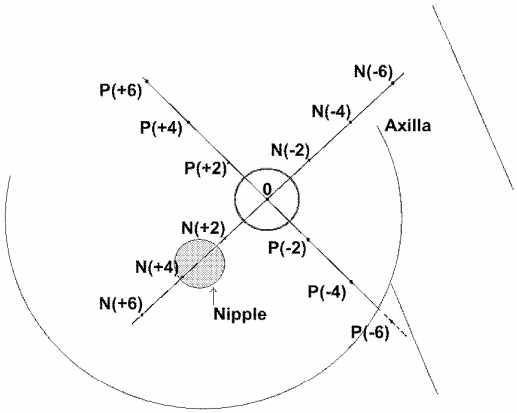
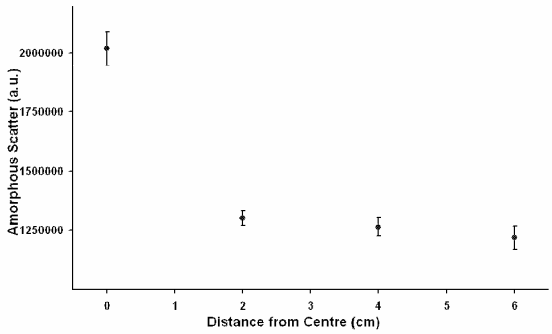
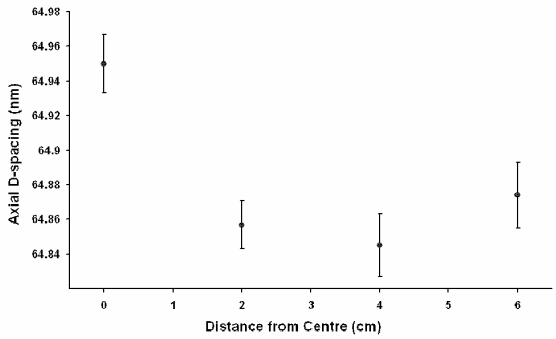


Figure 1: Locations of biopsy samples taken over an excised breast mass. The nipple is shown as a shaded circle, and the palpable mass is outlined (red circle). Typically, the centre (0 cm) was defined as the middle of the largest palpable lesion. Up to 6 tissue samples were taken at ~2 cm intervals from the centre along the nipple-axilla direction (N-axis), then similarly perpendicular to this axis (P-axis).



(a)



(b)

Figure 2: Average values for the amorphous scatter (a) and the axial D-spacing (b) for the centre tissue points and at the 2 cm intervals away from the centre. Error bars represent the 95% confidence intervals.

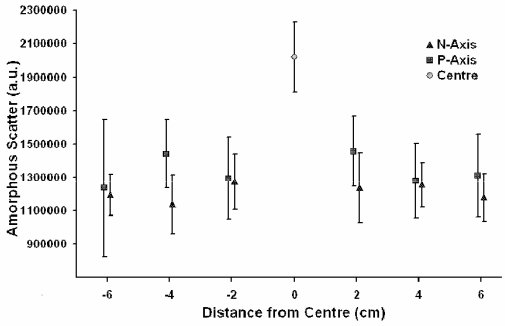


Figure 3: Average values of the amorphous scatter for the centre (0 cm, circle) and across the N-axis (triangles) and P-axis (squares). Negative values represent N(-2), N(-4), N(-6), P(-2), P(-4), and P(-6) tissue samples. Positive values represent N(+2), N(+4), N(+6), P(+2), P(+4), and P(-6) tissue samples. Error bars represent 95% confidence intervals.

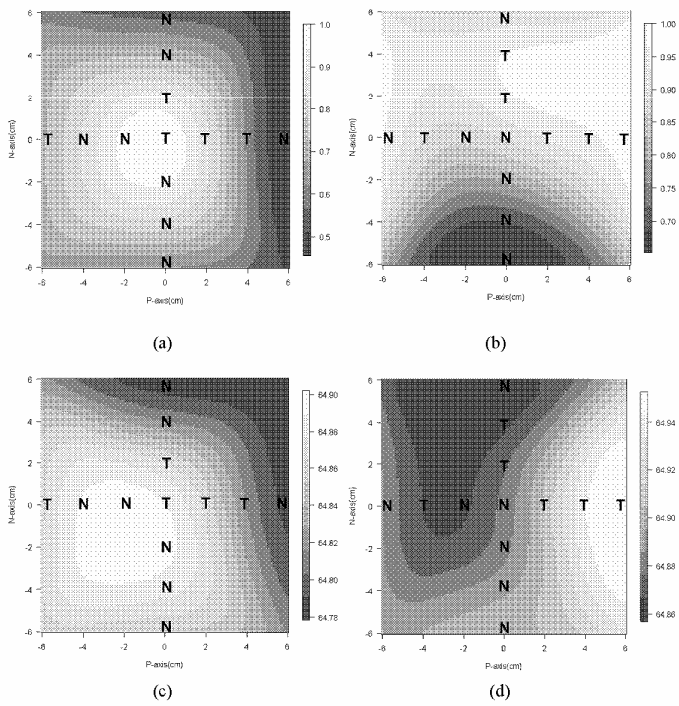


Figure 4: 2-dimenstional spatial models using the amorphous scatter (a & b) and the axial D-spacing (c & d). The amorphous scatter is normalized to each patient and the axial D-spacing is kept absolute. Patient 1 is shown in (a) and (c) and Patient 2 is shown in (b) and (d). Histopathology for each tissue sample is overlaid for each point measured: N represents normal parenchyma and T represents invasive carcinoma. (Colour reproductions shown in online version)

Table 1: P-values for the amorphous scatter and axial D-spacing values between the centers of the primary lesions and the tissue samples taken at 2 cm intervals over both axes.

	Amorphous Scatter	**Axial D-spacing**
Centre (0) & All (2,4,6 cm)	2.36×10^{-15}	0.010
Centre (0) & 2 cm	2.07×10^{-11}	0.019
Centre (0) & 4 cm	4.29×10^{-10}	0.014
Centre (0) & 6 cm	7.99×10^{-9}	0.081

Paper III: Mapping of Structural Changes in Breast Tissue Disease Using X-ray Scattering

Sabeena Sidhu
Monash Centre for Synchrotron Science, Monash University, Clayton, VIC Australia, 3800.
School of Physics, Monash University, Clayton Campus, VIC Australia 3800
Sabeena.Sidhu@sci.monash.edu.au
+613 9905 8063

Karen KW Siu
Monash Centre for Synchrotron Science, Monash University, Clayton, VIC Australia, 3800.
School of Physics, Monash University, Clayton Campus, VIC Australia 3800
Karen.Siu@sync.monash.edu.au
+613 9905 4922

Gregory Falzon
School of Science and Technology, University of New South Wales, Armidale, NSW Australia 2351
gfalzon@une.edu.au
+61 0409 749 154

Stewart A Hart
Department of Surgery, Monash University, Clayton, VIC Australia 3800.
Monash Breast Unit, Monash Medical Centre, Moorabin, VIC Australia 3165
centreroad@ozemail.com.au
+613 9579 6122

Jane G Fox
Department of Surgery, Monash University, Clayton, VIC Australia 3800.
Monash Breast Unit, Monash Medical Centre, Moorabin, VIC Australia 3165
Jane.Fox@med.monash.edu.au
+613 9579 6122

Simon P Nazaretian
Department of Anatomical Pathology, The Royal Women's Hospital, Parkville, VIC Australia3053.
Department of Anatomy and Developmental Biology, Monash University, Clayton, VIC Australia 3800.
Simon.nazaretian@med.monash.edu.au
+613 9344 2292

Robert A Lewis
Monash Centre for Synchrotron Science, Monash University, Clayton, VIC Australia, 3800.
School of Physics, Monash University, Clayton Campus, VIC Australia 3800
Rob.Lewis@sync.monash.edu.au
+613 9905 3622

Running Title: Mapping changes in breast tissue
Key Words: Breast cancer, tissue morphology, molecular structure

Abstract
Current standard histopathological analysis of breast tissue is limited to the cellular and tissue morphology and immunoperoxidase stains visible at the micrometre scale. Small Angle X-ray Scattering can characterise fibrillar structures of breast tissue at the tens to hundreds of nanometre range. This technique can probe supramolecular changes in breast tissue components before cellular abnormalities are visible. Changes at the supramolecular scale can further be mapped over the entire breast, showing areas of significant tissue degradation which can be related to potential diseased sites. Our group has previously shown that the amount of disorder in the fibril structures of collagen components in breast tissue is correlated with disease. The scattering intensity and the d-spacing of the third order axial collagen peak are two components that can be used to classify tissue types. Breast tissue samples were harvested from 31 mastectomies and wide local excisions over two axes at 2 cm intervals. The amorphous scatter and third order axial d-spacing were mapped over a 2-dimensional grid and a model was applied to show changes in the tissue structure. Standard histopathological reporting was used for a comparison. The model correlates with areas of disease and with areas of normal tissue, as diagnosed by conventional histopathology. Further examination indicates that SAXS can show evidence of disease, based on structural degradation of the tissue, where pathological analysis does not. The ability of SAXS to reveal structural changes in breast tissue specimens suggests that it may form a complimentary tool in diagnosis with conventional histopathology.

Introduction
Current histopathological analysis of breast tissue, obtained from core needle biopsy or surgical excision, uses information at the micrometre scale, where observation is limited to cellular structure and overall tissue architecture. Biological information at the nanometre scale, where molecular formation and fibre structures can be examined, may be useful in understanding the origins of cellular mutation and the proliferation of diseased cells. These changes can be mapped throughout the breast to show areas of high risk disease or to identify surgical margin extraction.

It has been shown from morphological data using zymography, immunoblotting, in situ hybridization, and immunohistochemical analysis of some invasive carcinomas that the disruption of the basal lamina present in these tissues is accompanied by an alteration of the underlying extracellular matrix (ECM) [8, 74, 143, 160]. Pathologically, the transition from *in situ* to invasive carcinoma is usually accompanied by the disorganisation and discontinuity of the basement membrane (BM) [54]. The collagen degradation and dispersal of the collagen bundles by invasive carcinomas have been observed using electron microscopy to show an abundance of thin fibrils chaotically running in all directions of the stroma [114]. The fragmented appearance of these fibrils is believed to result from the formation of new collagen that is unable to properly aggregate to reconstitute the bundles found in typical healthy tissues. Furthermore, collagen chain modifications, including acidic collagen component (αa) is present in most tumours, and has been shown to be absent in the adjacent normal mucosa [101]. On the other hand, loss of αa expression in invasive carcinomas may be attributed to the remodelling of the BM [74, 144]. The accumulation of altered collagen molecules during carcinoma progression impairs the correct supramolecular organisation of the collagenous stroma and may provide a permissive pathway for the invasive growth of neoplastic cells.

Visualisation of these molecular changes would provide an alternate probe of tissue characterization to compliment standard histology. X-ray diffraction is a well

known method of revealing the structure of biological molecules which can be crystallized. The most famous example was the determination of the molecular configuration of DNA by Watson and Crick [150] who used the information from X-ray diffraction photographs [154] to determine the regular spacing of nucleotides along the chain. For biological molecules that are not crystalline in nature, Small Angle X-ray Scatting (SAXS) is able to probe the supramolecular parameters of highly ordered structures and examine their motifs. No staining, fixing, or preparation of the tissue is necessary and samples can be studied in air or wet environments.

Several studies have investigated aspects of collagen structure in the breast as an indicator of disease [49, 50, 134]. Our previous study [134] showed two features that correlated well with disease: (1) amorphous scatter, and (2) third order axial d-spacing. The amorphous scatter is a good quantifier of the overall order (or disorder) in the sample: a higher scattering intensity represents more degradation and disordered structural components. Axial fibril d-spacing has also been correlated to disease [23, 45, 49, 51, 88], where a larger spacing has corresponded to evidence of invasive disease. In our recent study [134], we found distinctions in these two SAXS parameters between invasive carcinoma, benign breast disease, and normal breast tissues. Importantly, we also found differences between morphologically normal tissue (normal breast tissue in patients with definite known disease elsewhere in the breast) and completely disease-free normal tissue (breast tissue taken from reduction/mammoplasty procedures, with no past history or presence of disease) [134].

In this study, structural changes known to be correlated with breast disease were mapped over two perpendicular axes of mastectomy or wide local excision. The tissue samples were interpreted histologically and this was used as our gold standard. Mapping of the structural changes over a 2D area allows for spatial comparisons over the entire breast mass, which also allows for prediction of trends in areas between sampling sites or beyond the known edges of the excised mass to hypothesize possible directions or areas of disease extension.

Materials and Methods

Tissues samples were harvested from the removed tissue mass of patients who had biopsy confirmed lesions that were subsequently treated by mastectomy or wide-local excision. Core biopsy samples (MEAD 16 Gauge Needle Biopsy Gun) were taken from the centre of the primary lesion and also at approximately two centimetre intervals radially outwards along two perpendicular axes (see Figure 1). Figure 1a (Patient 1) was typical of most patients, where the 0 cm was defined at the centre of the palpable mass and then tissue samples taken at 2 cm intervals away from the centre. Figure 1b (Patient 2) was unusual, where more than one palpable mass was present. In this case, the centre (0 cm) was taken at the nipple and the N-axis was along the direction connecting the two palpable masses.

Accurate distance measurements from the primary site were challenging because the excised tissue mass was difficult to handle and position. Therefore, the measurements of 2 cm, 4 cm, and 6 cm are only approximate intervals (± 1 cm). The N-axis was determined as the line toward the nipple, and the P-axis was taken perpendicular to the N-axis. N(+2), N(+4), N(+6) notation represented samples taken towards or close to the nipple, and N(-2), N(-4), N(-6) notation represented samples taken towards the axilla. P(+2), P(+4), P(+6) represented samples taken towards the superior part of the breast, and P(-2), P(-4), P(-6) represented samples taken towards the inferior part of the breast. The maximum distance of a tissue sample was approximately 6 cm away from the centre of the primary lesion; thus, a maximum of 13 tissue samples was collected from any one patient. Most patients were, however, treated with wide local excision, and so sometimes the maximum distance from the primary lesion that samples could be harvested were only 2 cm or 4 cm away from the primary site.

Data from 31 patients with a total of 165 tissue samples were analysed in this study: 31 samples at 0 cm, 81 samples at 2 cm, 36 samples at 4 cm, and 17 samples at 6 cm. Tissue core samples were typically cylindrical in shape with a diameter of 1 mm and lengths from 4 mm to 25 mm. Tissues were obtained with full informed consent from patients undergoing surgical treatment, under ethics approvals for this study (Southern Health: Human Research Ethics Committee B (Project No. 03094B) and Monash University: SCERH Human Ethics (Project No. 2003/021)).

Classification of tissue samples was based on the pathologist's assessment of the tissue cores after the sample was imaged. Patient information, such as age and disease diagnosis, was also collected from each patient's final surgical report. All patients presented in this study had pathologically confirmed invasive carcinoma present in the primary tissue sample (tissue core taken at 0 cm, 31 patients). Subsequent tissue samples (tissue samples taken away from the primary site) were typically diagnosed as normal (99), however, there were a few samples diagnosed as invasive carcinoma (30) and benign disease (5).

All SAXS images were recorded at Station 2.1 at the SRS Daresbury, UK, over three experimental sessions. A beam size of ~0.5 mm x 0.5 mm was used at the sample with a wavelength of 1.54 Å. A full description of the experimental procedures, data acquisition, and subsequent image processing of the diffraction patterns is discussed in Sidhu *et al.* [134] and Lewis *et al.* [88]

All statistical analysis and calculations were performed using R (Version 2.6.1, The R Foundation for Statistical Computing, 2007), MatLab (Version 2.6.1 R2007a with Statistical Toolbox, MathWorks Inc. 2007), and Microsoft Office Excel (Microsoft Office Professional Edition 2003, Microsoft Corporation 2003). The total amorphous scatter and third order axial d-spacing values were determined for each tissue sample and mapped onto a 2-dimensional grid of 32 x 32 pixels. The size of the grid remained constant regardless of the number of tissue samples examined for each patient. Each sample's amorphous scatter data was normalised to the highest parameter value determined for that patient. This allowed for examination of each individual patient and between patient models based on the changes of the amorphous scatter, rather than absolute values. Third axial d-spacing values were given as absolute values, since the specific value of the third axial d-spacing is related to disease [49, 51, 88, 134]. Gaussian Kernel estimation was used to predict the missing values within the grid, between the known data points. This discrete convolution theorem was implemented using the Fast-Fourier Transform (FFT) [149]. Using this modelling technique allowed us to produce a 2D map over a large area from a limited number of tissue samples.

Results

The 2D models are shown by colour in Figure 2 for two sample patients who had mastectomy, in which the entire tissue mass excised exceeded 12 cm in both directions (i.e. 13 tissue points per patient). The histopathologic diagnosis for each of the tissue samples is overlaid for comparison. For both parameters, areas of red correspond to large values, which correspond to disease, and areas of blue correspond to low values, which correspond to normal tissue. The model agrees well with the histopathologic diagnosis and this was consistent with the other 29 patients studied.

With Patient 1, both the amorphous scatter (Figure 2a) and the third order axial d-spacing (Figure 2c) showed changes relating to tissue degradation towards the P(-6) edge (inferior part of the breast) of the tissue specimen. Patient 2 (Figure 2b and 2d) shows structural degradation present in the N(-6) area, rather than at the N(+6) area that was reported (surgical pathology report) to be the main site of disease. For Patient 2, the model shows no evidence of increased tissue degradation in that region (N(+6)) compared to the site at N(-6), suggesting that the origin of disease may be located at N(-6). Figure 2b and 2d also shows that molecular changes are extended along the margin of the N(-6). The tissue sample at the P(+6) point was histologically

diagnosed as benign, but shows similar structural degradation properties as the normal tissues for that patient.

Patient 1 (Figures 2a and 2c) shows that SAXS is consistent in showing areas of disease whereas pathological analysis shows a normal tissue diagnosis: Centre and P(-6) were diagnosed with disease, but the tissue samples between these points (P(-2) and P(-4)) are histopathologically normal. The SAXS data shows that these two tissue samples (P(-2) and P(-4)) have high levels of degradation, consistent with invasive carcinoma, but this was not evident on histology. Furthermore, histopathology shows that disease extends towards the nipple (N(+2) was diagnosed as invasive carcinoma); however, the SAXS model suggests that molecular changes extend toward the axilla (higher values towards N(-2) to N(-6)), particularly with the third order axial d-spacing parameter (Figure 2c).

Discussion

The SAXS data shows consistent changes, through specific gradients, of the molecular structure along the two axes examined. Most patients showed a "hot spot" on or near the primary site tissue sample with a decrease in parameter value toward the margins. If tissue sampling was taken at smaller intervals and more frequently over the excised mass, the model may even more accurately assess the lesion, which, if previously assessed to be a single large lesion, may turn out to be multifocal. For now, visualisation of these molecular changes would greatly aid in the determination of surgical margins and may even show presence of disease and even direct where samples can be taken for histology.

Multiple tissue samples must be collected over the breast in order to calculate a useful and predictive 2D mapping of structural changes. A minimum of 3 tissue samples on each axis were needed in order for the model to be calculated effectively. Otherwise, the truncated array tends to introduce numerical artefacts due to the use of the discrete FFT. In these cases, it may be appropriate to re-calculate the 2D grid with zero-padding on the boundaries to avoid aliasing (wrap-around) effects. Tissue sample collection over the entire grid, rather than only across two axes may show better mapping of structural changes occurring over the entire breast. However, an advantage of using SAXS as a diagnostic indicator of disease is that issues relating to tissue sampling error are decreased. With SAXS, the entire tissue core is used to obtain an overall quantified parameter to describe the state of the tissue. Standard histopathology information relies on a thin section of the excised tissue core, which may not be representative of the lesion.

The amorphous scatter is an excellent parameter to consider for mapping tissue structure changes, since it encompasses information from the entire tissue sample. Several studies by independent groups have found that the amorphous scatter is directly correlated with disease [49, 51, 134] and can easily be determined using SAXS. The third order axial collagen d-spacing has also been correlated with disease [49, 51, 88, 123, 134], however, this parameter is very sensitive to inter-patient variability [134], where patient age, previous radiotherapy/chemotherapy, and menopausal status may affect this value. However, with the amorphous scatter variable, normalising each patient, as was performed in this study, avoids sensitivities related to inter-patient variability. This also makes the gradients seen over the area of the mass more informative to assess the consistency of data from a diverse patient population. A disadvantage of normalisation, however, is that more than one tissue sample is needed for adequate comparison and mapping of the parameters over the 2D spatial plane.

Our group has previously shown that SAXS can distinguish between histologically diagnosed normal tissues from patients with known invasive disease with normal tissue from reduction/mammoplasty patients with no history or presence of disease [134]. Sensitivity and specificity, relative to histopathology, between these two normal tissue groups were 95.2 and 57.1, respectively. This suggests that SAXS

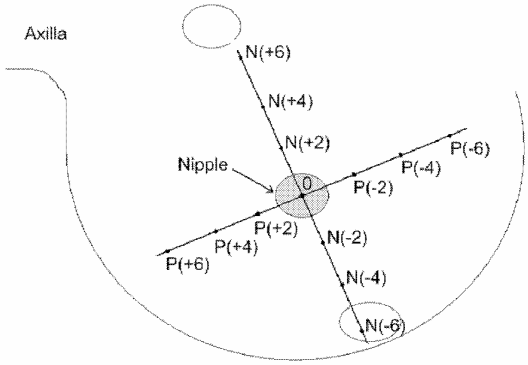
can distinguish structural tissue changes occurring in the breast caused by invasive disease elsewhere in the breast before detection by standard histopathology techniques. In particular, further clearance of tissue towards the axilla may show molecular degradation that could be correlated with axillary lymph node invasion. Future work with regards to increasing the sampling within the model and longer term follow up of the patients may show information in regards to directional spread of invasive disease and predict patients with higher metastatic risk.

Although synchrotron radiation is preferred in obtaining the structural parameter information presented, with the advent of better optics and detector systems, laboratory based SAXS is definitely a possibility for clinical analysis. Several groups have already presented promising results where differences between tissue types can be distinguished without synchrotron radiation [23, 62, 108, 109, 124]. Our group is currently analysing SAXS information from tissue samples imaged using a laboratory SAXS camera. Using SAXS in conjunction with current diagnostic methods may prove advantageous in determining areas of disease and possible direction of spread of invasive carcinoma, thus, these patients may be more closely monitored.

Acknowledgements

We acknowledge financial support from the *Access to Major Research Facilities Programme* which is a component of the *International Science Linkages Programme* established under the Australian Government's innovation statement, *Backing Australia's Ability*. SS is grateful for scholarship support from the Monash Centre for Synchrotron Science. KKWS acknowledges the support of a Strategic Grant from the Faculty of Medicine, Nursing and Health Sciences, Monash University. The authors would like to thank the nurses and staff at Southern Health for their support and patience, in particular, Dr. Beatrice Susil for histopathological support. The authors acknowledge helpful discussions with Dr. Chris Hall, Dr. David Paganin, and Dr. Adam Round.

(a)



(b)

Figure 1: Locations of excised biopsy samples taken from Patient 1 (a) left breast, and Patient 2 (b) right breast. Shaded circle represents nipple location, and clear ovals (red) indicate where palpable masses were located as reported by surgical pathology reports.

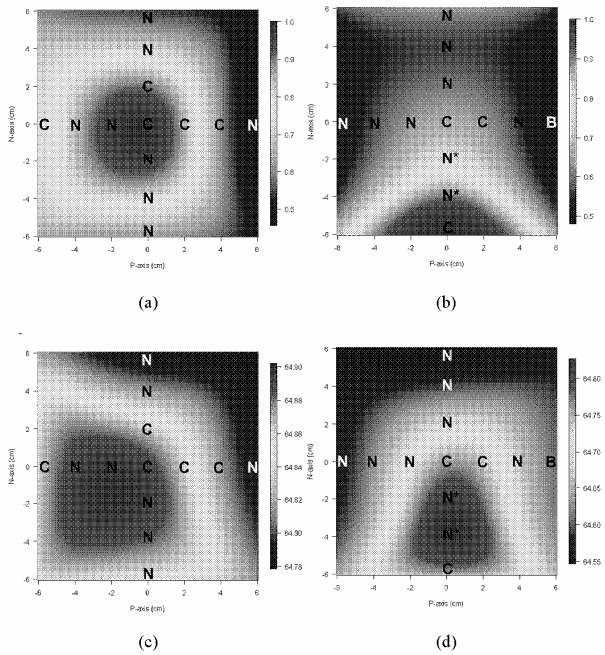


Figure 2: 2D Maps for Patient 1 (a & c) and Patient 2 (b & d), using the amorphous scatter (a & b) and using the third axial d-spacing (c & d). The amorphous scatter maps are normalised for each patient, where red (or a value of 1) corresponds to highly degraded tissue with a high probability of containing invasive carcinoma, and dark blue (or a value of 0) corresponds to highly ordered tissue with a high probability of being mostly normal parenchyma. The d-spacing maps (c & d) are not normalized, but show absolute third axial d-spacing differences. Large d-spacing corresponds to invaded tissue, and low d-spacing corresponds to normal parenchyma. Patient 1 is typical, where the centre (0 cm) was sampled at the centre of the palpable mass. Patient 2 is unusual, where the centre (0 cm) was taken at the nipple and N(+6) and N(-6) were areas of two separate palpable masses. Diagnosis of each tissue sample is overlaid as N for normal parenchyma, C for invasive carcinoma, and B for benign disease. (Note: the N* tissue diagnosis is based on a matched tissue sample of the patient, since the tissue at these points was not available for diagnosis)