Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Nuclear Instruments and Methods in Physics Research A ▮ (▮▮▮▮) ▮▮▮–▮▮▮
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

Identifying markers of pathology in SAXS data of malignant tissues of the brain

K.K.W. Siu[a,*], S.M. Butler[b], T. Beveridge[a], J.E. Gillam[a], C.J. Hall[c], A.H. Kaye[d,e], R.A. Lewis[f], K. Mannan[g], G. McLoughlin[g], S. Pearson[h], A.R. Round[i], E. Schültke[g,j,k], G.I. Webb[b], S.J. Wilkinson[i]

[a]*School of Physics and Materials Engineering, PO Box 27, Monash University, Victoria 3800, Australia*
[b]*School of Computer Science and Software Engineering, PO Box 75, Monash University, Victoria 3800, Australia*
[c]*Daresbury Laboratory, Warrington, Cheshire WA4 4AD, United Kingdom*
[d]*Department of Surgery, University of Melbourne, Parkville 3050, Australia*
[e]*Department of Neurosurgery, Royal Melbourne Hospital, Parkville 3050, Australia*
[f]*Monash Centre for Synchrotron Science, PO Box 27, Monash University, Victoria 3800, Australia*
[g]*Division of Neurosurgery, University of Saskatchewan, Saskatoon S7N 5E5, Canada*
[h]*Physics & Electronics, University of New England, Armidale, New South Wales 2351, Australia*
[i]*Department of Materials and Medical Sciences, Cranfield University, Wiltshire SN6 8LA, United Kingdom*
[j]*Department of Anatomy and Cell Biology, University of Saskatchewan, Saskatoon S7N 5E5, Canada*
[k]*Department of Neurological Science, Walton Medical Centre, University of Liverpool, L97 LJ, United Kingdom*

Abstract

Conventional neuropathological analysis for brain malignancies is heavily reliant on the observation of morphological abnormalities, observed in thin, stained sections of tissue. Small Angle X-ray Scattering (SAXS) data provide an alternative means of distinguishing pathology by examining the ultra-structural (nanometer length scales) characteristics of tissue. To evaluate the diagnostic potential of SAXS for brain tumors, data was collected from normal, malignant and benign tissues of the human brain at station 2.1 of the Daresbury Laboratory Synchrotron Radiation Source and subjected to data mining and multivariate statistical analysis. The results suggest SAXS data may be an effective classifier of malignancy.

PACS: 61.10.Eq; 41.60.Ap; 87.15.−v

Keywords: Small angle X-ray scattering; Brain tumors; Data mining; Synchrotron radiation

*Corresponding author. Tel.: +61 (0)3 9905 4922; fax: +61 (0)3 9905 3637.
E-mail address: Karen.Siu@spme.monash.edu.au (K.K.W. Siu).

0168-9002/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2005.03.081

1. Introduction

Brain tumors are the third most frequent cause of cancer related death in adults, being the cause of approximately 2% of all cancer deaths [1]. High grade gliomas (the tumors arising from the supporting tissue of the nervous system, WHO grades III–IV), are the most malignant brain tumors known, having a mean survival time of only 9–12 months [2].

The standard diagnosis of gliomas is reliant on neuropathological analysis under the framework of the World Health Organization (WHO) or Kernohan classification systems [3]. These systems attempt to give some indication of biological behavior, and hence form a basis for treatment protocols. However, the efficacy of the conventional treatments of radiotherapy and chemotherapy is non-uniform and difficult to predict. Treatment of gliomas by local removal is particularly difficult due to their highly diffuse structure. There is therefore a need for improved understanding of malignancy progression at the molecular level, to improve both disease diagnosis and patient treatments [4,5].

This study sought to evaluate Small Angle X-ray Scattering (SAXS) data as a means of distinguishing normal, malignant and benign tissues of the human brain. SAXS is well-suited to the study of the fibrous proteins of the extracellular matrix (ECM) as it probes ultra-structural information on length scales from 10–200 nm [6]. SAXS data has previously demonstrated diagnostic potential for breast cancer [7–9], in which the features distinguishing malignancy were related to changes in the ECM.

The ECM is also strongly implicated in brain tumor migration [10–12]. Invading glioma cells migrate along distinct ECM-containing anatomical structures, particularly blood vessels and tracts of myelinated fibers. This migratory and invasive behavior is regulated by the interactions between the glioma cells and components of the ECM, and necessitates ECM degradation and remodelling.

Myelin itself has also been extensively characterized by SAXS since its crystalline like structure gives rise to strong X-ray scattering [13–15]. Although X-ray diffraction data has long been used to elucidate the structural and compositional characteristics of myelin and its membrane interactions [14,16–18], relatively little work has considered pathologies of the nervous tissues. X-ray diffraction data have been used to examine the amyloid fibers characteristic of Alzheimer's disease [19] and changes in the organization of multiple sclerosis myelin [20]. Lazarev et al. [21] have examined normal and pathological tissues of the human breast, prostate and brain. Although the diffraction patterns from the pathological brain samples (Alzheimer's disease and cerebral hemorrhage) appeared different from normal tissue, samples were restricted to only one of each type.

We have performed SAXS data collection from 41 samples from 32 patients of malignant, benign and normal tissues of the human brain at the Daresbury Laboratory Synchrotron Radiation Source. To our knowledge, this is the first study that attempts to systematically identify pathological markers in SAXS data from malignant tissues of the brain.

2. Materials and methods

2.1. Experimental

Brain tumor tissue samples were sourced, with informed consent, from patients admitted for tumor resection surgery at the Royal Melbourne Hospital, Melbourne, Australia, and the Royal University Hospital, Saskatoon, Canada. Tumor-free control tissue was obtained from patients undergoing surgery for unrelated conditions, e.g. partial temporal lobectomy for medically intractable seizures. Human ethics approval for these studies has been granted at the relevant institutions. SAXS data were collected from total of 41 samples arising from 32 patients (see Table 1). Under the WHO grading system, glioblastomas (GBMs) are classified as grade IV gliomas, whilst the meningiomas and schwanomas are considered benign.

Immediately after surgical removal, tissue samples were snap frozen in liquid nitrogen for storage at $-80\,^{\circ}C$ until transported to the Daresbury Laboratory synchrotron radiation source. SAXS data were collected on the fixed energy (1.54 Å)

Table 1
Tissue sample types and numbers

Tissue type	Number of patients	Number of samples
Normal	1	2
Schwanoma	6	6
Meningioma	14	16
Glioblastoma multiforme	11	17
Total	32	41

small angle X-ray scattering camera at station 2.1 using the 200×200 mm multiwire detector (512×512 pixels) [22]. A sample to detector length of 2.3 m was used, giving a scattering length range of 18–460 Å. Rat tail tendon was used as a calibrant.

Tissue samples, typically 15 mm^2 in area and 1 mm thick, were held between two mica sheets in liquid cells such that the sample remained hydrated and maintained at 20 °C by means of a water bath. The beam size at the sample was 1×1 mm^2 which gave adequate signal to noise in the detector using exposure times of 100 s. Samples were raster scanned at 1 mm horizontal and vertical translations to provide up to 16 images per sample. Some images showed artifacts from X-ray scatter arising from the edges of the sample chamber. These were eliminated from further analysis. The remaining images were then averaged to provide SAXS data representative of the entire sample without loss of resolution in the data.

Immediately following data collection, samples were removed and fixed in 10% buffered formalin. They were then routinely processed for neuropathological examination.

The data were corrected for detector artifacts, incident beam decay and sample attenuation. The two-dimensional data were also radially averaged to provide one-dimensional data as a function of scattering length. We employed two methods of analysis of the SAXS data: Data mining and hierarchical cluster analysis.

2.2. Data mining

Data mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [23]. In a previous study into SAXS breast tumor data classification [9] we obtained useful markers of malignancy using a rule discovery based technique [24]. This exploratory rule discovery approach was selected over a classification rule-based approach so as to enable us to use alternative measures of interest to simple rule accuracy. In this way we hoped gain some further insight into the brain SAXS data by later assessing the rules.

We used version 1.3 of the Magnum Opus exploratory rule discovery software [25,26]. The default objective measure of leverage was used in these experiments, but the system can also find rules by coverage, support, and strength, where coverage is the proportion of all cases for which the rule conditions are true; support is the proportion of all cases for which the rule conditions and the rule outcome co-occur; strength is the proportion of examples for which the rule conditions are met and that also have the rule outcome; and leverage is the proportion of samples that satisfy both the conditions and outcome of the rule in excess of those that would be expected if the rule conditions where independent of the tumor type. Equal frequency discretization with 5 bins was employed for all numeric attributes. That is, all quantitative variables were recoded as qualitative variables by dividing the range of values into 5 sub-ranges such that equal numbers of samples fell into each sub-range.

In this analysis, the features used included properties such as the sum, maximum, and the locations of these maxima for different regions of the images in an effort to capture information stored in the rings at some basic level (see Fig. 1). Additionally, in this analysis we manually added two extra variables representing the presence of the 2nd and 4th order myelin scattering rings.

2.3. Hierarchical clustering analysis

Cluster analysis is a multi-variate statistics technique that groups objects into clusters that are more similar to one another than objects in other clusters [27]. In the present context, it is well-suited to investigate whether SAXS data can form the basis of meaningful classifications, which can

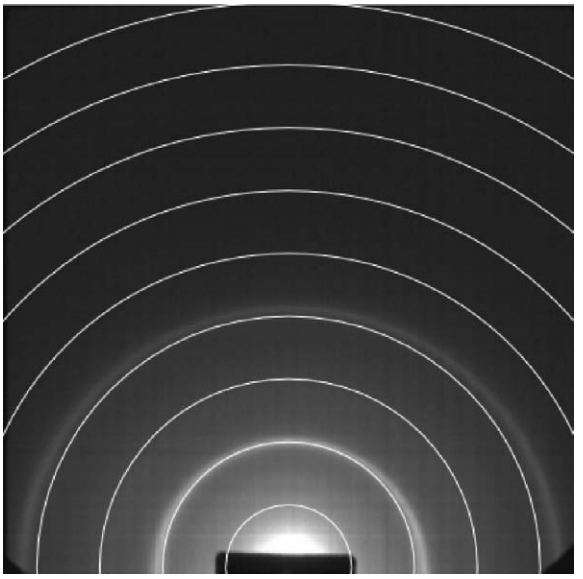

Fig. 1. A typical SAXS image from a glioblastoma sample. The strong 2nd and 4th order myelin scattering rings are clearly seen in this sample. The white lines show the segments used in the data mining analysis.

then be related to pathology. This means of analysis is inherently exploratory, rather than drawing on statistical inferences, as it quantitatively characterizes the structural differences between groups.

Since we do not expect a preferential scattering direction due to the heterogeneity of brain tissue in the gross volumes examined, we performed hierarchical cluster analysis on radially averaged data. In this analysis, the variables used for clustering are the intensities at each scattering length. The software SPSS v.12.0.2 (SPSS Inc., Chicago, USA) was used to perform the analysis.

We chose to restrict our classification investigation to two clusters of meningiomas vs. glioblastomas only at this early stage in our analysis since cluster analysis is highly sensitive to undersampling. Hierarchical cluster analysis using the between-groups method (in which the average of each group is compared to the average of another) was performed on these samples. Two similarity measures were used to group the data, squared Euclidian distance and Pearson correlation. The former measures the similarity of the data in magnitude, whereas the latter compares the patterns.

3. Results

A typical scattering pattern is shown in Fig. 1. The strong scattering rings arising from the 2nd and 4th orders of the myelin membrane pair thicknesses (153–161 Å) [18] are clearly seen in this example. Although predominate in this particular image, myelin scattering rings were not uniformly observed in all the data.

Fig. 2 shows the radially averaged data, by tissue type. Inspection of the standard deviations from the average for each of the meningiomas, glioblastomas and schwanomas (indicated by bars) shows that there is considerable overlap between the schwanomas and the other types. It should be remembered, however, that the number of schwanoma samples available was significantly smaller than meningiomas or glioblastomas.

Running the Magnum Opus system on the data yielded 49 rules that satisfied the constraints. We present the most predictive of these rules in Tables 2 and 3. Two rules, describing the presence of the myelin scattering rings in the radially averaged data, are shown in Table 2. Both rules cover 4 additional cases than would be expected if the rings' presence was independent of the GBM

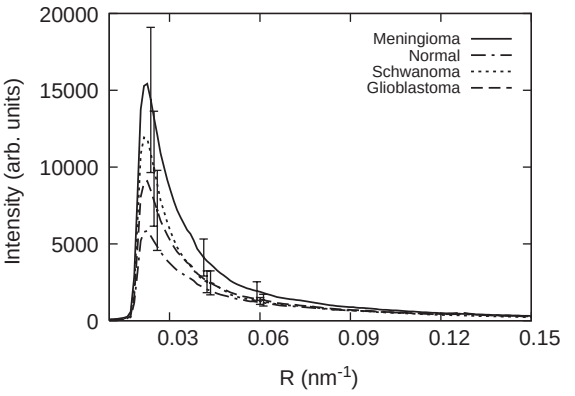

Fig. 2. Radially averaged data for each of the tissue types. The error bars indicate ±1 standard deviation for the meningiomas, schwanomas and glioblastomas only.

Table 2
Association rules relating to the 2nd and 4th order myelin rings. The number in brackets after the coverage, support and leverage indicates the number of samples described by these values respectively

Rule	Tissue type	Coverage	Support	Strength	Leverage	p-value
4th order myelin ring present	GBM	0.439 (18)	0.293 (12)	0.667	0.1106 (4)	0.004
2nd order myelin ring present	GBM	0.220 (9)	0.195 (8)	0.889	0.1041 (4)	0.001

Table 3
Association rules relating to the area near the beam stop. Numbers in brackets as for Table 2

Rule	Tissue type	Coverage	Support	Strength	Leverage
Max intensity in circular region #2 $>$ 2558	Meningioma	0.195 (8)	0.171 (7)	0.875	0.0946 (3)
Max intensity in circular region #1 $>$ 20120	Meningioma	0.195 (8)	0.146 (6)	0.750	0.0702 (2)
Max intensity in circular region #1 $<$ 8633	Normal	0.195 (8)	0.049 (2)	0.250	0.0393 (1)

Table 4
Hierarchical clustering results

Similarity measure	R range (nm^{-1})	Tissue type	Cluster #1	Cluster #2	Outliers	As diagnosed
Squared Euclidean distance	0.019–0.308	Meningioma	6	9	1	16
		Glioblastoma multiforme	15	2	0	17
Pearson correlation	0.170–0.548	Meningioma	14	1	1	16
		Glioblastoma multiforme	6	10	1	17

tumor type. To test whether this could have occurred by chance we applied a chi-square test and both rules passed at $\alpha = 0.05$. This concurs with our knowledge that these malignancies migrate along myelinated fiber tracts. However, this rule should be treated with caution, since the presence of myelin in samples of benign tumors is dependent on the exact location and amount of tissue excised during surgery.

Table 3 presents the rules associated with the maximum intensity of a circular region of the data near the beam stop. These rules show high strength values, suggesting that this is a good indicator of pathology. To verify this result we subjected the peak intensity values in the radially averaged data to a one-tailed t-test for independent samples. GBM, meningioma and normal tissue types all passed the statistical comparison when each tissue type was individually compared to all other types (e.g. GBM to all non-GBM samples) at $\alpha = 0.05$. GBMs passed with $p = 0.004$; meningioma with $p = 0.005$; and normal with $p < 0.001$. Schwanomas were the only tissue type that failed the statistical comparison ($p = 0.340$).

We also examined different scattering vector ranges using hierarchical clustering to determine which features of the scattering patterns might be able to discriminate tissue types. A summary of the most successful two-cluster results is shown in Table 4. Samples that were consistently grouped into clusters with only one member were considered outliers. In the scattering vector range between the beam stop and the 2nd order myelin peak, the squared Euclidean distance measure yielded one cluster composed nearly exclusively of only GBMs, in which 15 from a total of 17 GBM samples were grouped. The majority of the meningiomas were distributed approximately

40:60% between the two clusters. The converse is observed if a Pearson correlation is instead used as the similarity measure. The data range included all the data except that very close to the beam stop. This method groups 14 from 16 meningiomas nearly exclusively in one cluster, with a similar 60:40% distribution of GBMs between the two clusters. These results suggest that the absolute magnitude of the maximum scattering intensity is significant in the classification of GBMs, supporting the data mining results, whilst it is the overall pattern of the scattering data that distinguishes the meningiomas. Analysis of subsets of scattering ranges, such as in the vicinity of each of the myelin scattering rings, did not reveal any apparent consistency in clusters with tissue type.

4. Conclusions and directions for further work

These preliminary analyses imply that SAXS data has the ability to discriminate between tissue types without reliance on subjective interpretation. Both the data mining and hierarchical cluster analysis suggest that the maximum intensity at high *d*-spacings may be an effective marker of pathology. In addition, the results suggest that the presence of myelin scattering and the overall scattering pattern profile warrant further investigation for the identification of glioblastomas. Interpretation of the data from the schwanoma and normal tissue is difficult due to the limited sample numbers.

The features distinguishing glioblastomas are of particular interest as a long term-goal of this work is to define sub-classifications in the malignant tumor group correlated with therapeutic outcomes. Considerable effort will be required to understand the relationship of these features to the biological changes underlying pathology.

Future experimental work is currently planned to increase our database of scattering data, and in particular to balance the numbers of samples in the underrepresented schwanoma and normal groups. In addition to continuing work on exploratory rule discovery approaches, we also plan to subject the results of our cluster analyses to further multivariate statistics techniques to probe which features contribute to the inter-cluster variance. We expect that the symbiosis of these different approaches will be critical to determining if SAXS data has diagnostic potential.

Acknowledgements

We are indebted to Dr Mark Aylim from Bolton Hospital who carried out the histopathology preparations. We are also grateful to Mr S. Stylii of the Royal Melbourne Hospital for his help with sample collection.

This work was supported in part by the Commonwealth of Australia under the Access to Major Research Facilities Program, the Saskatchewan Synchrotron Institute (SSI), the University of Saskatchewan, The Department of Surgery at the Royal University Hospital, Saskatoon, and from the Royal University Hospital Foundation of Saskatchewan.

JEG acknowledges the support provided by an Australian Postgraduate Award. TB acknowledges the support of the J.L. William Scholarship.

References

[1] CBTRUS, Statistical report: Primary brain tumors in the United States, 1995–1999, 2002.
[2] E. Silverberg, et al., CA. Cancer J. Clin. (1990) 409.
[3] M.F. Gonzales, in: A. Kaye, E.R. Laws Jr. (Eds.), Brain Tumors: An Encyclopedic Approach, Churchill Livingstone, New York, 1995.
[4] M. Iida, et al., J. Cancer Res. Clin. Oncol. 123 (1997) 619.
[5] H. Kudo, et al., J. Neurosurg. 72 (1990) 610.
[6] W. Bras, et al., Nucl. Instr. and Meth. A 326 (1993) 587.
[7] M. Fernandez, et al., Phys. Med. Biol. (2002) 577.
[8] R.A. Lewis, et al., J. Synchrotron Rad. 7 (2000) 348.
[9] S.M. Butler, et al., Proceedings of the 16th Australian Joint Conference on Artificial Intelligence, vol. 2903, 2003, p. 677.
[10] E. Kaczarek, et al., Int. J. Dev. Neurosci. 17 (1999) 625.
[11] A. Giese, et al., Neurosurgery 38 (1996) 755.
[12] A. Giese, M. Westphal, Neurosurgery 39 (1996) 235.
[13] A.E. Blaurock, Brain Res. 210 (1981) 383.
[14] D.A. Kirschner, et al., J. Neurochem. 53 (1989) 1599.
[15] A.E. Blaurock, et al., Neurochem. Res. 11 (1986) 1103.
[16] F.O. Schmitt, et al., Science 80 (1934) 567.

[17] C.R. Worthington, A.E. Blaurock, Nature 217 (1968) 87.
[18] R.J. Chandross, et al., J. Comp. Neurol. 177 (1978) 1.
[19] H. Inouye, et al., Biophys. J. 64 (1993) 502.
[20] L.S. Chia, et al., Proc. Natl. Acad. Sci. USA 81 (1984) 1871.
[21] P. Lazarev, et al., Proceedings IEEE 4 (2000) 3230.
[22] R.A. Lewis, J. Synchrotron Rad. 1 (1994) 43.
[23] U. Fayyad, et al., AI Magazine 17 (1996) 37.
[24] R. Agrawal, et al., in: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 1993, p. 207.
[25] G.I. Webb, in: The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, p. 99.
[26] G.I. Webb, Magnum Opus v. 1.3 (2001).
[27] J.F. Hair, et al., Multivariate Data Analysis, Prentice-Hall, Upper Saddle River, NJ, 1998.