Image processing analysis of stem cell antigens

Marc-Olivier M. P. Baradez

Kingston University

2005

PAGES NOT SCANNED AT
THE REQUEST OF THE
UNIVERSITY

SEE ORIGINAL COPY OF
THE THESIS FOR THIS
MATERIAL

Image Processing analysis of stem cell antigens.

Abstract

This thesis aims to investigate the automation of an image processing driven analysis of antigen distributions in the membrane of early human Haematopoietic Stem/Progenitor Cells (HSPCs) imaged by Laser Scanning Confocal Microscopy (LSCM). LSCM experiments generated a vast amount of images of both single and dual labelled HSPCs. Special focus was given to the analysis of colocalised antigen distributions, as colocalisation may involve functional relationships. However, quantitative methods are also investigated to characterise both single and dual labelled antigen distributions.

Firstly, novel segmentation algorithms are developed and assessed for their performances in automatically achieving fast fluorescence signal identification. Special attention is given to global histogram-based thresholding methods due to their potential use in real time applications. A new approach to fluorescence quantification is proposed and tested. Secondly, visualisation techniques are developed in order to further assist the analysis of the antigen distributions in cell membranes. They include 3D reconstruction of the fluorescence, newly proposed 2D Antigen Density Maps (ADMs) and new 3D graphs of the spatial distributions (sphere models). Thirdly, original methods to quantitatively characterise the fluorescence distributions are developed. They are applied to both single and dual/colocalised distributions. For the latest, specific approaches are investigated and applied to colocalised CD34/CD164 distributions and to colocalised $CD34^{class\ I}CD34^{class\ II}$ and $CD34^{class\ I}CD34^{class\ III}$ epitopes distributions (two combinations of

the three known different isoforms of the CD34 molecule, a major clinical marker for HSPCs).

The visualisation tools revealed that HSPC membrane antigens are often clustered within membrane domains. Three main types of clusters were identified: small clusters, large patch-like clusters and newly identified meridian-shaped crest-like (MSCL) clusters. Quantitative analysis of antigen distributions showed heterogeneous distributions of the various measured features (such as polarity or colocalisation patterns) within the HSPC populations analysed.

Finally, the proposed methodology to characterise membrane antigen distributions is discussed, and its potential application to other biomedical studies is commented. The potential extensions of the innovative linear diffusion-based MultiScale Analysis (MSA) algorithm to other applications are outlined. Visual and quantitative analyses of antigen membrane distributions are eventually used to generate hypotheses on the potential, yet unknown roles of these early antigens and are discussed in the context of haematopoietic theories.
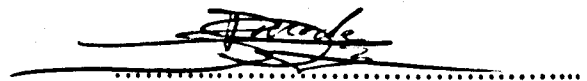
Declaration

I hereby declare that whilst registered for a research degree at Kingston University, I have not been a registered candidate or enrolled student at any other academic institution.

I declare that the materials contained in this thesis have not been used in any other submission for an academic award. All the sources of investigation have been duly acknowledged.

This thesis has been composed by myself and is the result of my own investigation.

Marc-Olivier Baradez, Bsc (Hons)

December 2004

List of publications, presentations and awards

Reprints of published papers are in Appendix A.

Full research papers:

- **Baradez M-O**, Forraz N, Hoppe A, Pettengell R, McGuckin CP. Subcellular analysis of the sialomucins CD34 and CD164 in stem-progenitor cell membranes. 2004. **(submitted).**

- **Baradez M-O**, McGuckin CP, Forraz N, Pettengell R, Hoppe A. Robust and automated unimodal histogram thresholding and potential applications. Pattern Recognition 2004; 37(1):1131-1148.

- McGuckin CP, Forraz N, **Baradez M-O**, Lojo-Rial C, Wertheim D, Whiting K, Watt SM, Pettengell R. Colocalization analysis of sialomuicins CD34 and CD164. Stem Cells 2003;21: 162-170.

Conference paper:

- **Baradez M-O**, Forraz N, Hoppe A, Pettengell R, McGuckin CP. Thresholding based on linear diffusion for feature segmentation. In Proceedings of the British Machine Vision Conference 2003; 9th to 11th September, Norwich, UK.

Conference abstracts and presentations:

- **Baradez M-O**, Forraz N, Hoppe A, Pettengell R, McGuckin CP. Thresholding based on linear diffusion for feature segmentation. In Proceedings of the British Machine Vision Conference 2003; 9th to 11th September, Norwich, UK.

- **Baradez M-O**, Pettengell R, McGuckin C, Hoppe A, Wertheim D. Semi-automatic recognition of cell boundaries. Photon02 conference, Cardiff, September 2002.

- **Baradez M-O**, Pettengell R, McGuckin C, Hoppe A, Wertheim D. Computer system for semi-automatic recognition of cell boundaries. House of Commons Presentation by Britain's Younger Scientists, Engineers and Technologists, London, March 2002.

- **Baradez M-O**, Wertheim D, Pearce D, Forraz N, Whiting K, Pettengell R, McGuckin CP. Quantification of antigen distributions around human blood progenitor cells. 5th International Conference on Cellular Engineering, Aachen, Germany, July 2001.

- Wertheim D, **Baradez M-O**, Whiting K, Forraz N, Pearce D, Pettengell R, McGuckin CP. Computer analysis of confocal microscope images. 5th International Conference on Cellular Engineering, Aachen, Germany, July 2001.

- Forraz N, **Baradez M-O**, Watt S, Whiting K, Pettengell R, McGuckin CP. Confocal laser scanning microscope 3D representation supports co-localisation of sialomucin molecules CD164/CD34 on bone marrow and

umbilical cord blood primary cells during adhesion. British Society of Haematology meeting, Harrogate, March 2001.

- McGuckin CP, **Baradez M-O**, Wertheim D, Pearce D, Forraz N, Pettengell R. A new computer system for assessment of progenitor antigenic label distribution from confocal microscope images. 2000. **Blood 96[11], 122b.**

Presentations:

- **Baradez M-O**, McGuckin CP, Pearce D, Forraz N, Pettengell R, Wertheim D. A semi-automatic computer system for recognition of cell boundaries. Oral presentation at Kingston University School of Computing and Information Systems Research Seminar, September 2001.

- **Baradez M-O**, McGuckin CP, Pettengell R, Wertheim D. Quantification of antigen distributions in stem cell membranes. Oral presentation at Kingston University School of Computing and Information Systems Research Seminar, October 2000.

Awards:

- The "Travel Scholarship Award" awarded by **the British Journal of Haematology Research Trust** to attend the 45th Conference of the American Society of Hematology, San Diego, CA, awarded in November 2003.

- The "**T. Kamalagharan Rose Bowl Prize**" for the best poster entered by a postgraduate student at St George's Hospital Medical School, Cellular Molecular Sciences Group 9th Scientific meeting and symposium awarded in March 2001.

Other conference abstracts and presentations:

(Not directly related to the present thesis)

- C. Dessapt, **M.O. Baradez**, A. Hayward, GC. Viberti, L. Gnudi. TGF-β1 and mechanical stretch reduce murine podocyte adhesion to extracellular matrix substrate and modulate β1 integrin expression / maturation *in vitro*. European Association for the Study of Diabetes, 41st Annual Meeting 2005, Athens, Greece, September 2005.

- C. Dessapt, **M.O. Baradez**, A. Hayward, GC. Viberti, L. Gnudi. TGF-β1 and Mechanical Stretch reduce murine podocyte adhesion to extracellular matrix substrate and modulate β1 integrin expression / maturation *in vitro*. 65th Scientific Sessions at the American Diabetes Association, in San Diego, USA, June 2005. Diabetes, June 2005, Vol 54, Supplement 1, p A200.

- C. Dessapt, **M.O. Baradez**, A. Hayward, GC. Viberti, L. Gnudi. Effects of mechanical stretch and TGF-β1 on β1 integrin expression in murine podocytes *in vitro*. European Diabetic Nephropathy Study Group, an Official Group of the European Association for the Study of Diabetes, Arnhem, The Netherlands, May 2005.

- Dessapt C, **Baradez M-O**, Hayward A, Viberti GC, Gnudi L. Effects of TGFβ_1 and mechanical stretch on β_1 integrin expression in podocytes in vitro. Accepted for *poster* presentation at the National Kidney Research Fund (NKRF) Fellows Day, Belfast, Republic of Ireland, April 2005.

- Dessapt C, **Baradez M-O**, Hayward A, Viberti GC, Gnudi L. Effects of TGFβ_1 and mechanical stretch on β_1 integrin expression in podocytes in vitro. Accepted for *oral* presentation at the Renal Association / British Transplantation Society Congress, Belfast, Republic of Ireland, April 2005.

- Dessapt C, Dei Cas A, **Baradez M-O**, Hayward A, Thomas S, Viberti GC, Gnudi L. High glucose and mechanical stretch downregulate $\alpha3\beta1$ integrin expression and reduce cell adhesion in podocytes in vitro. 40th Annual Meeting of European Association for the Study of Diabetes. Published in Diabetologia, Vol. 47, Supplement 1, p A98. September 5-9, 2004, Munich, Germany.

- Dessapt C, Dei Cas A, **Baradez M-O**, Hayward A, Hargraeves R, Viberti G, Gnudi L. High glucose and mechanical stretch regulate $\alpha3\beta1$ integrin expression and cell adhesion in murine podocytes in vitro. 64th Annual Scientific Sessions of the American Diabetes Association, June 2004, Olrando, published in *"Diabetes"*.

- McGuckin CP, Forraz N, **Baradez M-O**, Myriknas S, Ouvrard R, Lowe R, Swaby D. Extracellular matrix components and umbilical cord blood stem cells support ex vivo regeneration of the cornea. 26th Meeting of the

European Study Group for Cell Proliferation, Cell proliferation Volume 37(2), pp. 130, May 2004, Prague, Czech Republic.

- Dessapt C, Dei Cas A, **Baradez M-O**, Hayward A, Hargraeves R, Viberti G, Gnudi L. Modulation of $\alpha3\beta1$ integrin expression and cell adhesion in murine glomerular visceral epithelial cells by high glucose and mechanical stretch in vitro. Meeting of the European Diabetic Nephropathy Study Group – an official group of the European Association for the study of Diabetes – Bergamo, Italy, May 2004.

- Dessapt C, Dei Cas A, **Baradez M-O**, Hargraeves R, Viberti G, Gnudi L. High Glucose and mechanical stretch regulate $\alpha3\beta1$ integrin expression and cell adhesion in murine glomerular visceral epithelial cells in vitro. National Kidney Research Fund Fellows Day, March 2004, Aberdeen.

- Dessapt C, Dei Cas A, **Baradez M-O**, Hayward A, Hargraeves R, Viberti G, Gnudi L. High glucose and mechanical stretch regulate $\alpha3\beta1$ integrin expression and cell adhesion in murine podocytes in vitro. Presentation at Britain's Younger Scientists, Engineers and Technologists, House of Commons, London, March 2004.

- Forraz N, Myriknas S, **Baradez M-O**, Ouvrard R, Lowe R, Swaby D, Morgan R, Rostron CK, McGuckin CP. Tissue regeneration for cornea using cord blood stem cells. American Society of Cell Biology Annual Meeting, San Francisco, USA, December 2003, # L341.

List of figures and tables

Figures:

Chapter 1

Figure 1.1: **The haematopoietic process.**

Figure 1.2: **Models of haematopoiesis.**

Figure 1.3: **The haematopoietic microenvironment,** adapted from Hoffbrand *et al*, 2001.

Figure 1.4: **Some antigen expressed in HSPC membrane and their potential role in regulating signalling pathways,** taken from Simmons *et al*, 1997.

Figure 1.5: **Example of membrane molecules used to characterise and investigate HSPCs,** taken from Simmons *et al*, 1999.

Figure 1.6: **Molecular structure of the CD34 antigen,** taken from Krause *et al*, 1996.

Figure 1.7: **Molecular structure of the CD164 antigen,** taken from Zannetino *et al*, 1996.

Figure 1.8: **Molecular structure of the CD133 antigen,** taken from Miraglia *et al*, 1997.

Figure 1.9: **Principles of Laser Scanning Confocal Microscopy (LSCM).**

Figure 1.10: **LSCM Z-series.**

Figure 1.11: **Images obtained by LSCM.**

Figure 1.12: **Effects of filtering on background regions histograms.**

Chapter 4

Tables:

Chapter 6

Acknowledgments

I would like to thank my Director of Studies Dr Andreas Hoppe for his valuable help and guidance throughout these years. I am particularly grateful for his open-minded views on the methodologies developed during the course of this PhD. I also thank you for the enjoyable extra late hours of work and discussions which were very appreciated.

I am extremely grateful to my supervisor Dr Colin McGuckin for his constant support over the years since 2000, when I joined the group as a project student. I wish to acknowledge the valuable professional experience he taught me and all the opportunities he kindly offered to me since. I also wish to thank him for standing by me all along in difficult times. A huge thank you for this!

I also thank my supervisor Dr Ruth Pettengell (St. George's Hospital Medical School) and her PhD student Mrs. Karen Whiting for providing the microscope images used in the present work. Thanks to Dr Daniel Pearce for providing the first confocal datasets.

I am very grateful to Dr Nicolas Forraz for providing microscope images. Moreover, during these years, he was a source of inspiration and a great person to discuss the work. It was a real pleasure working with him.

Thanks are extended to Pr Graeme Wilkinson and Dr Graeme Jones for their help with the course of this PhD. I am sincerely thankful to Dr Jones for his guidance and advices in difficult times.

Thanks to Dr Luidgi Gnudi, King's College London, for kindly providing electrophoresis gel images on which I could test quantification approaches and investigate future directions.

A big thank you to the many friends who were part of my english landscape : Colin and Nico, both the best friends I could ever have found when I left France, Paul and Rachael, so honestly crazy!, Pierrot, with whom I had long discussions over night on how to change the world (and get better marks), Carlos our future family doctor, Marie-Christine and her cooking, Estelle and her wonderful chicken you must try, Maryline, Mohson, Suzy and Sean. Many thanks to Pierre, Lilette, Lydia and Magalie Dessapt for their constant encouragements and affection, it helped tremendously!

Cecile Dessapt deserves the warmest thanks for her love, support and faith in me over these years. Her presence by my side was a great and constant motivating force. Thank you also for supporting me when it was hard enough for you. Merci du fond du coeur.

Finally, I wish to thank my family, my Great-parents for their financial support and their wise remarks, my parents Yves and Anne for the financial support during my studies, and more importantly for their encouragements and love. Thank you so much for all these years. I would like to humbly dedicate this work to them and to the loving memory of Mr Marian Borkowski and Mr Paul Baradez.

Marc-Olivier Baradez

December 2004

Aims and objectives

This thesis aims to develop new algorithms and methodologies in order to characterise human blood stem/progenitor cell membrane antigens. Specific membrane antigens are visualised using laser scanning confocal microscopy. Such images are heavily contaminated by a noisy background which is predominant over rarer fluorescence signal pixels. This results in a strong unimodal histogram of the image grey intensities. This unimodality complicates the decision of a threshold that would appropriately separate the relevant brighter signal regions from the background, prior to signal processing, representation and analysis. Furthermore, confocal microscopy generates large datasets. As a consequence, this work aims first to the development and application of automatic segmentation algorithms in order to reliably extract signal regions.

This work also aims to the analysis of the segmented antigen distributions. This analysis is both qualitative (i.e. visualisation of the 3D distributions in cell membranes, 3D reconstruction of the fluorescence) and quantitative. Different visualisation approaches are investigated to display the fluorescence distributions. The quantitative analysis aims to characterise the distributions according to sets of measurements performed on the segmented and processed signal regions. Different measurements reflecting different characteristics of the distributions are therefore developed and applied. The collection of these measurements can be subjected to statistical analysis in order compare different stem/progenitor cell populations. Quantitative methods to characterise stem cell antigens are developed. These methods allow the characterisation of single labelled (only one target molecule of interest is fluorescently labelled in cell membranes) or dual

labelled cells (two different antigens are labelled on the same cell membrane). For dual labelled cells, special methods to quantify the colocalisation properties of the distributions are also developed.

Finally, this thesis aims to compare different cell populations. A statistical analysis of the measurements is performed and commented in the context of labelling and imaging procedures. Together with the qualitative analysis of the membrane antigens distributions, these analyses are used in the context of haematopoiesis to generate hypothesis on the potential roles and functions (unknown to date) of the investigated antigens.

One of the segmentation algorithms is based on linear diffusion and analyses intensity histograms of the images observed at different levels of observations. This novel approach has segmentation potential for other types of images, not necessarily with unimodal histograms. Such potential is therefore briefly investigated at the end of the thesis.

The global methodology adopted to characterise stem cell membrane antigen distributions by image processing approaches can be further extended to other biomedical investigations.

Table of content

Chapter 5. Fluorescence processing, representation and analysis

Chapter 7. Discussion   250

7.1. LSCM image processing methodology   250

7.2. Potential applications of the MSA algorithm 259

Reference list and bibliography

Appendices

List of abbreviations

ADM – antigen density map

ACD-A – acid citrate dextrose formula-A acid

APC – antigen presenting cell

BCVM – Between-Class Variance Maximisation

BIHC – Background/Image Histogram Comparison

BM – bone marrow

CAM– cell adhesion molecule

CB – umbilical cord blood

CD – cluster of differentiation

CFC – colony forming cell

CFU – colony forming unit

2D – two dimensional

3D – three dimensional

Da – Dalton

DIC – differential interference contrast

DNA – deoxyribonucleic acid

ECM – extra-cellular matrix

FACS – fluorescence activated cell sorting

FISH – fluorescence *in situ* hybridation

FITC – fluorescein isothiocyanate

HIM – haemopoietic inductive microenvironment

HSC – haemopoietic stem cell

HSPC – haemopoietic stem / progenitor cell

ICAM – intracellular adhesion molecule

Ig - immunoglobulin

IL – interleukin

LFA – lymphocyte function-associated antigen

LinNeg – lineage negative

LSCM – laser scanning confocal microscopy

LTC – long term culture

LTC-IC – long term culture-initiating cell

LSM – laser scanning microscope

MACS – magnetic activated cell sorting

MHC – major histocompatibility complex

MNC – mononucleated cell

MoAbs – monoclonal antibodies

MSA – multiscale approach

Neg – negative

PBS – phosphate buffered saline

PDF – probability density function

PE – phycoerythrin

PERCP – perinidin chlorophyll protein

PSF – point spread function

ROI – region of interest

SD – standard deviation

TCR – T-cell receptor

TRITC – tetramethyl rhodamine isothiocyanate

UCB – umbilical cord blood

Chapter 1. General introduction

The work presented in this thesis was initiated by the collaboration of Stem Cells scientists with Bioinformatics scientists. Amongst a large body of work, this stem cell research involved an extensive use of Laser Scanning Confocal Microscopy (LSCM) in order to get deeper insights on the potential role of stem cell membrane molecules, or antigens, whose functions are unclear or unknown to date. These antigens are thought to play important roles in the haematopoietic (blood formation) process by regulating and modulating stem cell responses to microenvironmental stimuli, cytokines and other molecules, involved in cell-to-cell contact, adhesion, migration and maturation.

This thesis aims to quantitatively characterise stem cell membrane antigen distributions in order to allow statistical comparison of different antigens and cell populations. Stem cell membrane antigens were investigated by laser scanning confocal microscopy (LSCM), an imaging technique allowing the three-dimensional visualisation of specific molecules in cells and tissues. A new methodology and novel algorithms were developed and applied to process LSCM images in order to extract quantitative information. Such quantitative analysis is used in turn to suggest hypothesis about the potential role(s) of the target antigens/cell populations. While quantitative analysis of confocal fluorescence images is widely used in biomedical research, first extracting the relevant information, such as image fluorescent regions, is still mostly performed subjectively or by following heuristics approaches. The independence given to the operator at this stage may prevent reliable reproduction of the quantitative results. This is likely to introduce some bias in the resultant analysis. The possible

existence of such a bias in the analysis of rare samples, such as actual stem cell populations or certain cancer biopsies, justifies the attempt for automating the analysis process from its early stages (sample preparation, imaging) to its final stages (analyses of results and interpretation).

The thesis is organised as follows. First, the background relevant to this study is presented in chapter 1. It encompasses the haematological process (i.e. mechanisms of blood formation), the microscopy techniques for cell imaging and a review of common segmentation algorithms used to separate relevant image regions from background. It follows a description of the biology protocols involved in the study, from stem cell isolation and labelling with specific fluorescent markers to antigen distribution imaging in chapter 2. In chapter 3, new algorithms for fluorescence signal segmentation of LSCM images are developed and described. Several approaches were investigated. Emphasis was given to global thresholding approaches as they are fast methods, and three new global thresholding algorithms are presented. A new region-based algorithm is also presented to demonstrate the feasibility of spatial segmentation approaches. The global thresholding algorithms are tested and validated in chapter 4, along with a proposed fluorescence quantification approach. In chapter 5, an original methodology is designed to process, represent and characterise stem cell membrane antigen distributions in order to extract relevant and quantifiable information. The quantitative results are presented and further detailed and commented in chapter 6. They are finally discussed and used to generate hypothesises specific to the algorithms, the protocols/methodologies, and the interpretation of the results and potential applications of this research in chapter 7.

1.1. Background

Blood is a special, liquid tissue as vital to the body as the brain or the heart. It is distributed throughout the whole body via a complex ramified network of arteries and veins, and reaches all organs in order, for example, to oxygenate or protect them against infections. The blood is constituted of a nutritive and protective liquid fraction, the plasma, which serves as a mean of transportation for the blood cells. The plasma also transports a very wide range of diverse molecules such as coagulation factors, glucose, antibodies, essential for the immune response, or high and low density lipoproteins, essential for transportation of essential molecules (e.g. cholesterol), and the metabolic products to their processing site.

Blood cells can be divided in three categories:

- the red blood cells, which function is to carry the oxygen from the lung to all organs and tissues, and to collect the carbon dioxide from the organs and bring it to the lungs where it is expulsed out of the body,

- the white blood cells, or leukocytes, are members of an extremely diverse family which encompasses three main cellular types, the polynucleated cells, the lymphocytes and the monocytes:

  o the polynucleated cells, named after the pluri-lobed aspect of their nucleus, are further sub-divided into three groups, namely the neutrophils, eosinophils and basophils.

o   the lymphocytes are responsible for immunity as only them have the properties of diversity, specificity, memory and self/non-self recognition, all essential to coordinate the immune response.

o   the monocytes, from which are derived the macrophages, are an essential element of the immunity defence mechanisms as they have the ability to track, capture and digest bacteria and foreign bodies. They are highly motile, chemo-sensitive cells able to respond to chemotactic gradients triggered by infected or injured tissues.

- The platelets represent the smallest cellular elements of the blood. These non-nucleated cells are produced by a specialised cell type, the thrombocytogenic megacariocytes. Platelets are the result of the fragmentation of pseudopod-like protrusions emerging from the thrombocytogenic megacariocyte cell membrane. Their function is to act as the primary agent of homeostasis, a process which prevents blood leakage from altered blood vessels by forming the "blood clot". This phenomenon is efficient when small blood vessels are involved, but the clotting process of bigger vessels requires further complex processes.

This wide diversity of blood cells, and the complex intricate network of chemical and cellular responses it can manage, is entirely generated from a unique, undifferentiated pool of cells known as the Haemopoietic Stem Cells, or HSC.

1.1.1. Haematopoiesis

Haematopoiesis, or haemopoiesis, is a highly dynamic, finely tuned process giving rise to the production of all blood cell lineages (fig. 1.1). In this process, a small subset of extremely immature cells, the Haemopoietic Stem Cells (HSCs), is of particular importance. HSCs are characterised by two properties: their aptitude to sustain self-renewal of the HSC pool, and their potential to exponentially proliferate to produce the whole variety of blood cell lineages [Krause *et al*, 2001].

Indeed, these cells have the capability to greatly proliferate and differentiate into fully mature blood cells such as lymphocytes or red blood cells [Akashi *et al*, 2000], in order to sustain normal blood function. It is experimentally estimated that a single HSC can generate about 10^{15} mature cells after only 50 cell divisions [McNiece and Bridell, 2001, Brummendorf *et al*, 1998]. However, in order not to exhaust the stem cell pool, HSCs also have to self-renew themselves throughout the lifespan of the individual [Bjørnsson *et al*, 2003]. The mechanisms controlling HSC dynamics are carefully balanced via cell-to-cell contacts between HSC and its microenvironment through adhesion molecules [Krämer *et al*, 1999], in conjunction with signalling molecules such as growth factors and cytokines. Abnormalities in the haematopoietic process may trigger important complications such as anaemia, lymphoma or leukaemia.

As the HSC differentiate, they become progenitor cells, then precursors, and eventually become mature, fully differentiated and functional cells. The experimental discrimination between stem and progenitor cells is a difficult and uncertain task, and such cells are commonly referred to as haematopoietic stem/progenitor cells (HSPC). As stem cells are recruited in the process of
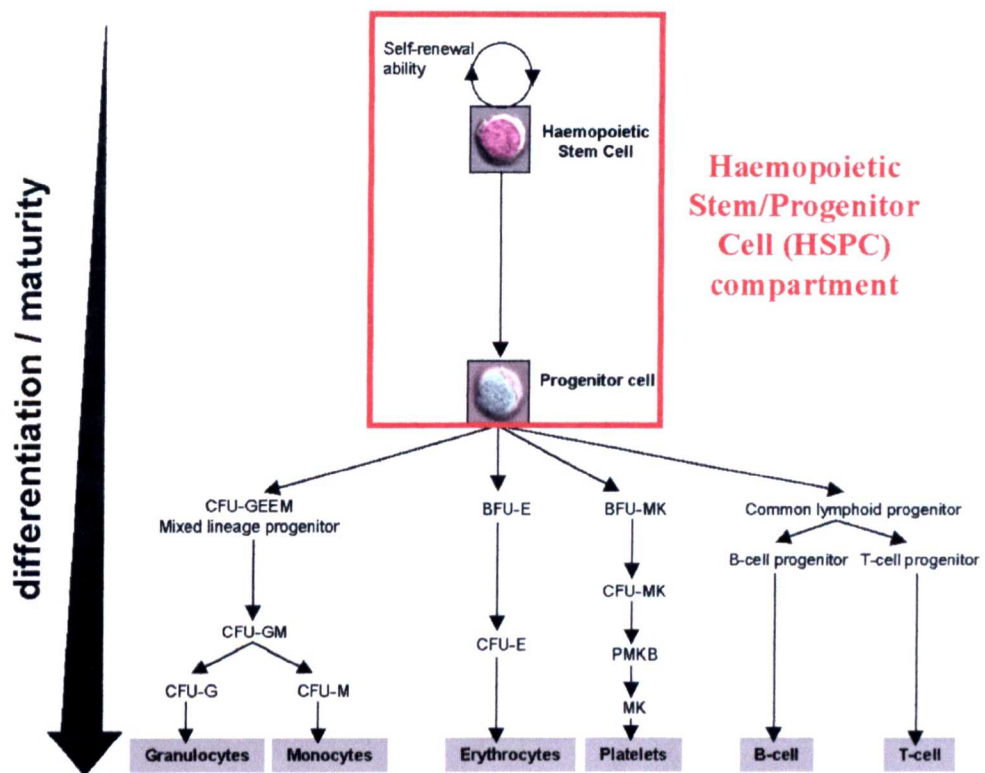
Figure 1.1: **The haematopoietic process.**

Haemopoiesis is a pyramidal process. Haemopoietic Stem/Progenitor cells are found at the top of the pyramid. As they are involved in the proliferation and differentiation process, they lose their self-renewal potential and become progenitors. In turn, as differentiation increases, the committed progenitors give rise to increasingly specialised and more mature cells. At the bottom of the pyramid are fully differentiated, functional blood cell types.

proliferation and differentiation, they progressively lose their self-renewal ability, while they show a higher cell cycling activity and proliferation [Aglietta *et al*, 1998].

Two predominant models were proposed to explain the haematopoietic process, namely the stochastic and the determinist (or inductive) models (fig. 1.2).

In the stochastic model, it is supposed that HSC generate fully committed progenitors, i.e. as HSC become progenitors they gradually express random combinations of various receptors to cytokines and other key growth factors as the consequence of complex intrinsic molecular events that encode for such receptors. Blood lineages are determined by suitable combinations of signalling molecules secreted in the progenitor environment. The variety of possible signalling molecule combinations is responsible for the lineage, while the absence of suitable combination leads to apoptosis (orchestrated cell death), a critical and essential mechanism necessary to control cell proliferation [Ogawa et al, 1983; Koury, 1992].

In the determinist model, HSC are assumed to express a range of lineage specific combinations of growth factor receptors. The role of the environment is therefore to secrete the appropriate cocktail of cytokines in order to activate functionally clustered genes that are eventually responsible for committed differentiation and maturation. Similarly to the previous model, if appropriate cytokine cocktails are not encountered by the progenitors, they undergo apoptosis in order to prevent abnormal haematopoiesis.

To date, the actual nature of the haematopoietic process has not yet been discovered. Recent studies demonstrated the reversibility of the HSC phenotype [Sato et al, 1999; Knaan-Shanzer, 2000; Dao et al, 2003], which allows new views on the nature of this haematopoietic process. It seems likely that a combination of both models takes place interactively in the body and is actually responsible for

the flexible commitment of cell lineages given physiologically possible combinations of cytokine cocktails.

Model 1   Instructional hypothesis

Microenvironmental signals

HSPC

HSPC   Lineage specific progenitors

Microenvironmental signal determined differentiation or self-renwal

Model 2   Stochastic hypothesis

Random decision regarding self-renewal or lineage -specific differentiation commitment

HSPC

HSPC   Lineage specific progenitors

Microenvironmental survival signals
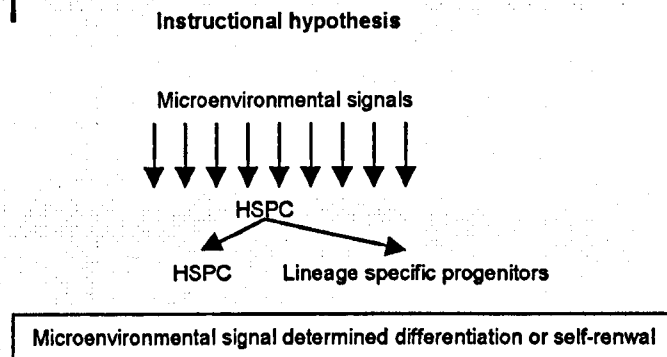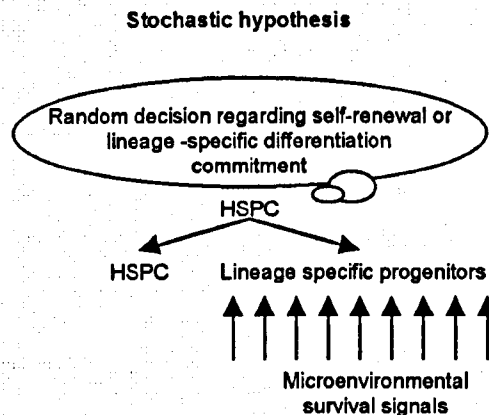
Figure 1.2: **Models of haematopoiesis.**

In the first model – the instructional model – microenvironmental signals are responsible for the fate of the HSPCs. Under the appropriate stimulations (via cytokines and adhesion molecules), HSPCs are kept immature and quiescent, or engage themselves in the differentiation process. For the second model – the stochastic model – HSPCs have a random potential for either remaining in an immature state or entering the differentiation process. Their fate is decided depending on the microenvironmental signals received and their potential to be stimulated by such signals.

1.1.1.1. Haematopoietic microenvironment

During embryonic and foetal development, the haematopoiesis takes successively place in different organs [Kumaravelu *et al*, 2002], and can be divided into three phases.

- Mesoblastic phase: at week 3 of gestation, some mesenchymal cells evolve into primitive endothelial cells, or angioblasts, and haematocytoblasts
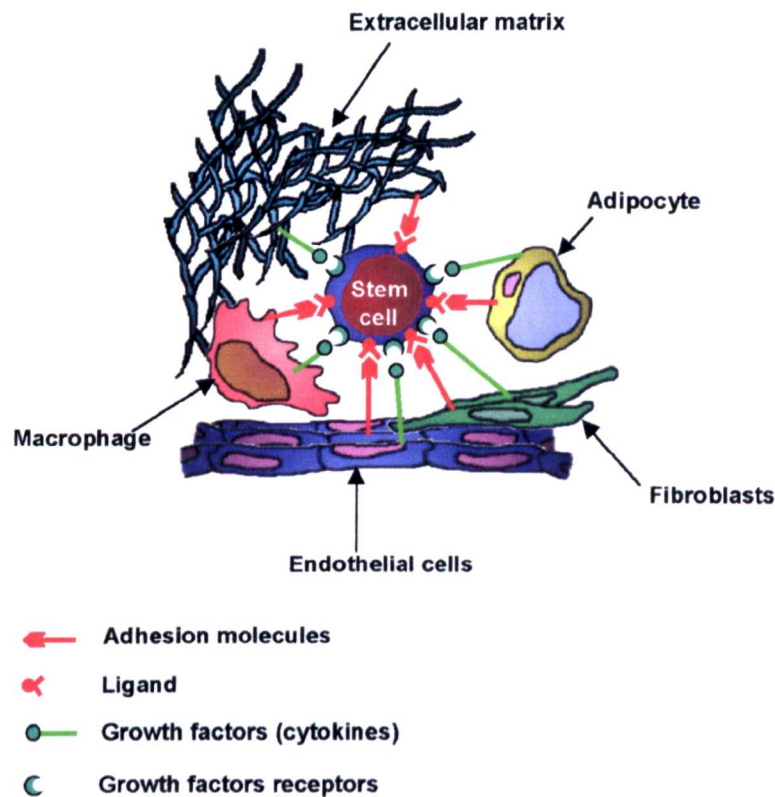
Figure 1.3: **The haematopoietic microenvironment.**

The fate of HSPCs is regulated by finely tuned molecular pathways. Interaction of growth factors and adhesion molecules with their cognate ligands and growth factor receptors triggers signalling cascades which modulate these pathways. These molecules are secreted or expressed by a wide variety of other cells (macrophages, adipocytes, fibroblasts, endothelial cells) and some are components of the extracellular matrix, all present in the haemopoietic microenvironment.

(primitive erythroid-like, granulocytic-like, and non-platelet producing megakaryocytic-like haematopoietic cells).

- Hepatic phase: at week 6, the liver becomes the main haemopoietic organ, with increased erythropoiesis and granulopoiesis. In the same time, the spleen also supports the haemopoiesis, to a smaller extent, and contributes to granulopoiesis, while lymphopoiesis begins in the foetal thymus and lymphoid tissue until birth.

- The myeloid phase: at week 22, the bone marrow starts to increasingly support the haematopoietic process and becomes the major site for life-long haematopoesis.

1.1.1.2. The bone marrow as an haemopoietic organ

The bone marrow is typically located in the medullary cavities in spongy parts of long bones and diploe of flat bones. This particular organ is structured as a complex three-dimensional network formed by the extra-cellular matrix (ECM), stromal cells and accessory cells, which altogether is referred as the haemopoietic inductive microenvironment (HIM) [Mayani *et al*, 1992] (fig. 1.3). HSC self-renewal, differentiation and maturation predominantly take place in the bone marrow. It constitutes an ideal environment where the finely tuned balance of intrinsic retro-active molecular pathways can effectively control the fate of HSC [Taichman and Emerson, 1998]. These regulation pathways are constituted by positive and negative signals triggered by the release of soluble cytokines and growth factors by the surrounding cells in conjunction with the pathways activated

by cell-adhesion molecules (CAM) at the adhesion sites of HSC with the haemopoietic inductive microenvironment.

1.1.1.3. Control of haematopoiesis

The haemopoiesis is controlled by successions of complex and complementary intrinsic and extrinsic stimuli [Alexander *et al*, 1996; Phillips *et al*, 2000]. Ex-vivo expansion studies stressed the importance of external stimuli through particular interactions taking place at the interface of the HSC with the environment, i.e. the cell membrane [Christensen *et al*, 2004; Batard *et al*, 2000] (fig. 1.4). Indeed it is
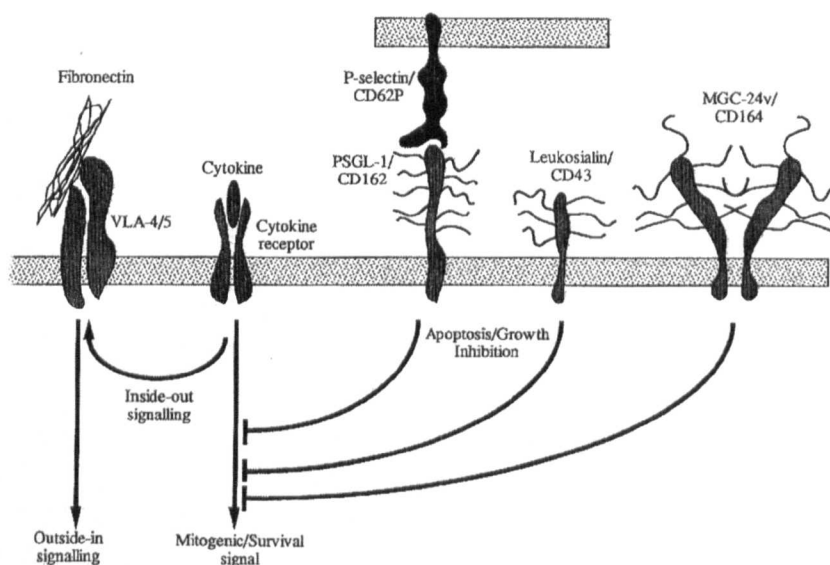
Figure 1.4: **Some antigen expressed in HSPC membrane and their potential role in regulating signalling pathways.**

These antigens are part of larger families of membrane antigens potentially involved in regulating signalling pathways. CD43 has a portion of its associated signalling pathway shared with a HSPC marker, the sialomucin CD34. CD164 is another sialomucin recently identified in early HSPC populations.

known that extrinsic stimuli are mediated inside the cell via a diverse range of membrane molecules such as CAM, which allow cells to exchange information at the site of cell-to-cell contact or at the adhesion site of the cells with the ECM [Bhatia *et al*, 2002]. Also critical to mediate external stimuli are the receptors for cytokines and other growth factors. The activation of the membrane receptors generate intra-cellular cascade of molecular events leading to the generation and modulation of the intrinsic regulatory stimuli with production of appropriate transcription factors [Tenen *et al,* 1997; Shivadsani and Orkin 1996; Behringer *et al*, 1997]. Some membrane antigens have found important clinical application [Elghetany, 1998; Rafii *et al*, 2002; Michova *et al*, 2003; Brugger *et al*, 1999; Piacibello *et al*, 1997 ; Thomas, 1999].

1.1.2. Investigating Haemopoietic Stem Cells

After the early experiments demonstrating the existence of HSC [Mitchinson, 1956; Nowell *et al*, 1956], the description of the first clonal haemopoietic stem cell assay used to characterise the colony-forming-unit-spleen (CFU-S) by [Till and McCulloch, 1961] was opening the way of a rigorous experimental framework to investigate quantitatively the dynamics of blood cell lineages formation. Further studies confirmed the status of progeny cells and demonstrated that they originated from a single clonogenic progenitor [reviewed in Weissman *et al*, 2001; Nilsson and Quesenberry, 1998]. The therapeutic interest of HSC was demonstrated by [Goodman and Hodgson, 1962] who showed the definite
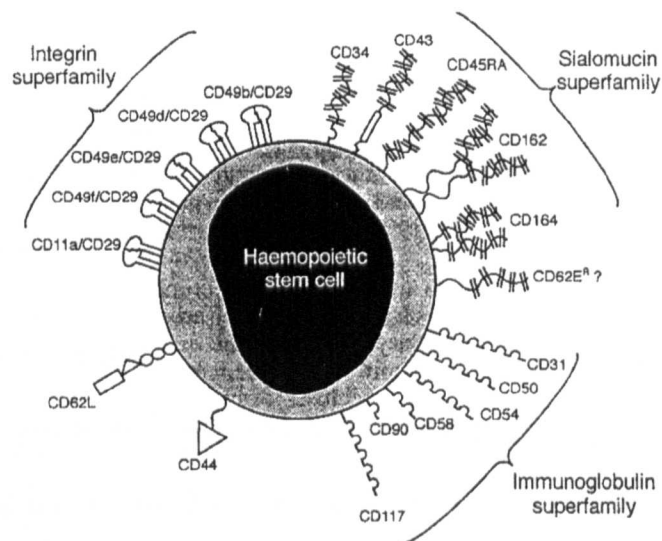
Figure 1.5: **Example of membrane molecules used to characterise and investigate HSPCs.**

This example shows that a wide diversity of molecules are present in HSPC membranes, many of these potentially interacting in order to sustain an appropriate regulation of survival signals. The membrane molecules are classified into families and superfamilies depending on their molecular structures and properties.

evidence that circulating stem cells were able to restore irradiation-myeloablated haemopoiesis in mice.

With such a therapeutic potential, the need for better characterising HSC led to the development of

- *in vitro* clonal progenitor and stromal-based assays [Pluznik and Sachs, 1965; Bradley and Metcalf, 1966; Dexter *et al*, 1977], which allow to investigate the formation of monoclonal colonies derived from progenitor cells,

- specific monoclonal antibodies with affinity for haemopoietic cell surface antigens, which found a critical application in high speed HSPC isolation [Muller-Sieburg *et al*, 1986],

- human recombined cytokines enabling studies of external pathway stimulating specific lineage differentiation and/or stem cell maintenance [Metcalf, 1993],

- *in vivo* animal models where competitive engraftment assays, in myeloablated and immuno-compromised animal hosts, between one HSPC population competing with another, which enables the cinetics and kinematics analysis of early or long-term bone marrow repopulation [Harrisson *et al*, 1988].

Recently, the attention of researchers has focused on some specific membrane antigens [Greenberg *et al*, 2000; Gazitt, 2000]. Indeed, some of these antigens are only expressed (by blood cells) at the very early stages of haematopoiesis and are

therefore used as markers to identify immature cells. HSPC mostly express CAM which belong to the immunoglobulin, the integrin and the sialomucin superfamilies [Simmons *et al*, 1997; Verfaillie, 1998; Chan and Watt, 2001; Jiang *et al*, 2000] (fig. 1.5), with the exception of L-selectin (involved in HSPC homing) and CD44 (which mediates high affinity cell adhesion to bone marrow stroma and ECM components such as hyalunoric acid and fibronectin) [Denning-Kendall *et al*, 2003]. Such antigens are the CD34, CD133, CD164 molecules. The functions of these antigens is not well known or not known at all, and trying to understand their functions provides with a challenge for biomedical scientists.

1.1.3. The sialomucin family

Sialomucins (or mucine-like molecules) are a family of glycoproteins which share common structural features. Typically, the extracellular domain of these proteins has numerous serine and threonine rich residues which confer them a variety of 0-glycosilation sites. As a consequence, the O-linked carbohydrate array forms a dense globular thread-like structure. Such structure provides a convenient platform for displaying multiple sugar combinations [Chan and Watt, 2001; Simmons *et al*, 1999; Baumheter *et al*, 1993]. The function of the sialomucins is not yet well established. Increasingly, recent investigations suggest a role in mediating outside-in signalling, which appear to have powerful negative regulator ability on the haematopoiesis. This negatively regulates the expansion of HSPC stimulated by cytokines. It is also hypothesised that sialomucins can trigger positive regulatory stimuli once binding to their cognate, yet unknown ligands [Park *et al*, 1991; Simmons *et al*, 1999; Fackler *et al*, 1995; Zannettino *et al*,

1998]. Therefore investigating sialomucins can provide information on their functions, and understanding how the work could optimise ex-vivo expansion strategies, so critical to generate large pools of HSCP prior to transplantation.

1.1.3.1. The CD34 antigen

CD34 (fig. 1.6) is a type I, 353 amino acids long (105-120kDa) transmembrane glycoprotein of the sialomucin family. The first part of the extracellular domain is estimated to be 130 amino acid long and typically exhibits multiple sites for O-linked and N-linked glycosylations, which makes this domain to be heavily glycosylated. Further away from this region is another extra cellular section
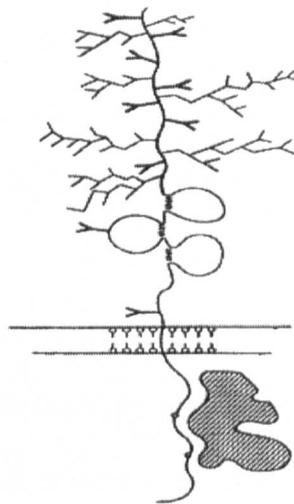
Figure 1.6: **Molecular structure of the CD34 antigen.**

The intracellular domain present sites for protein kinase C (dashed intracellular body) and tyrosine phosphorylation, the proximal extracellular present immunoglobulin-like domains while the distal part is heavily glycosylated due to the presence of multiple sites for O-linked and N-linked glycosylations along the peptide chain, conferring to the CD34 antigen a global thread-like configuration.

encompassing an estimated 100 amino acid long segment which displays several immunoglobulin-like structures. The cytoplasmic domain is much shorter than the extracellular counterpart, and it presents two sites for protein kinase C phosphorylation and one for tyrosine phosphorylation [Fackler *et al*, 1990; Simmons, *et al* 1992].

The precise function of the CD34 antigen and counter receptor/cognate ligand still remains unknown after twenty years of investigation. Several studies have shown a possible role in cytoadhesion to BM stroma and/ or endothelium [Madjic *et* al, 1994; Healy *et al*, 1995] and binding of CD34 with specific antibodies have been shown *in vitro* to have an effect on HSPC migration across various substrates, such as L-selectins [Baumheter *et al*, 1993; de Boers *et al*, 2002].

CD34 was found to be expressed on a small, discrete cell subset of the bone marrow which encompassed HSCP [Civin *et al*, 1984; Sutherland *et al*, 1989]. Indeed, CD34+ cells accounts for about 3% of nucleated cells in the bone marrow, 0.01-0.05% in steady state normal peripheral blood, and around 1% in umbilical cord blood [Barnett *et al*, 1999; de Wynter *et al*, 1998; McGuckin *et al*, 2003]. Fluorescence activated cell sorting (FACS) studies showed that the CD34+ cells are a heterogeneous population of primitive haemopoietic cells able of initiating early and long term cell culture (LTC-IC). They were further characterised as committed progenitors, with the critical ability of *in vivo* reconstitution of the myelo-lymphopoietic system in myeloablated hosts [Krause, *et al*, 1996]. The lineage-committed CD34 expression is gradually down-regulated as HSPC progress in the differentiation process [Sutherland *et al*, 1992; Krause *et al*, 1994; McGuckin *et al*, 2003]. However other tissues in the body can express this

sialomucin, as detected in fine vessels endothelium and some fibroblasts of the

bone marrow [Steen *et al*, 1996]. Moreover, it is thought that CD34 may have a

function in cell trans-differentiation [Jankowski *et al*, 2002], as observed with its

consequences on the regeneration capacity of myogenic progenitor cells. It is also

known that some CD34^{Neg} cell subsets are precursors for blood progenitors [Huss,

1998]. CD34 is a very important and widely used clinical marker of progenitor

cells [Vogel *et al*, 2000] and a large body of work on HSPC deals with the

manipulation of CD34$^+$ cells (e.g. [Wu *et al*, 2001]).

1.1.3.2 The CD164 antigen

The CD164 antigen (fig. 1.7) is a heavily glycosylated type I transmembrane

sialomucin protein estimated to be a 368 amino acid long (160-180 kDa)
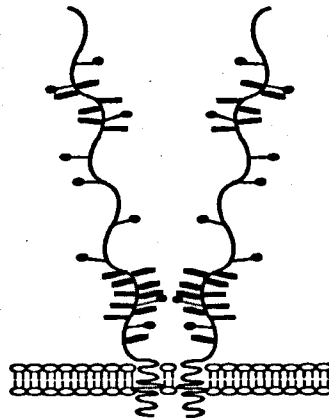
Figure 1.7: Molecular structure of the CD164 antigen.

This molecule has been observed in monomeric and dimeric forms, possibly tetrameric as well. Protein dimerisation is observed for some membrane receptors, suggesting such a function for CD164. The internal portion of the molecule is very short while the longer external portion present heavily glycosylated domains separated by more poorly glycosylated domains.

homodimer (so each protein core is 184 amino acids long). It shares thread-like structural similarities with CD34. Its extracellular domain also bears multiple O- and N- glycosylation sites. However CD164 does not possess any globular structure on its inner extracellular segment [Zannettino *et* al, 1998; Watt *et al,* 1998]. The expression of this antigen is not restricted to HSCP, but is also found in many tissues throughout the body. It is found in the HSPC membrane and also on bone marrow stromal cells, developing erythroid cells, monocytes and pre-B cells [Watt *et al*, 2000; Watt and Chan, 2000]. It is estimated that, on average, 63-82% of bone marrow and 55-93% of umbilical cord blood CD34+ cells are positive for CD164 [Watt *et al*, 1998]. The ligand to CD164 remains unknown [Chan *et al*, 2001]. However antibody ligation studies revealed an interactive relationship between CD34 and CD164 responsible for regulating HSPC proliferation and adhesion to the bone marrow stroma [Zannettino *et al,* 1998].
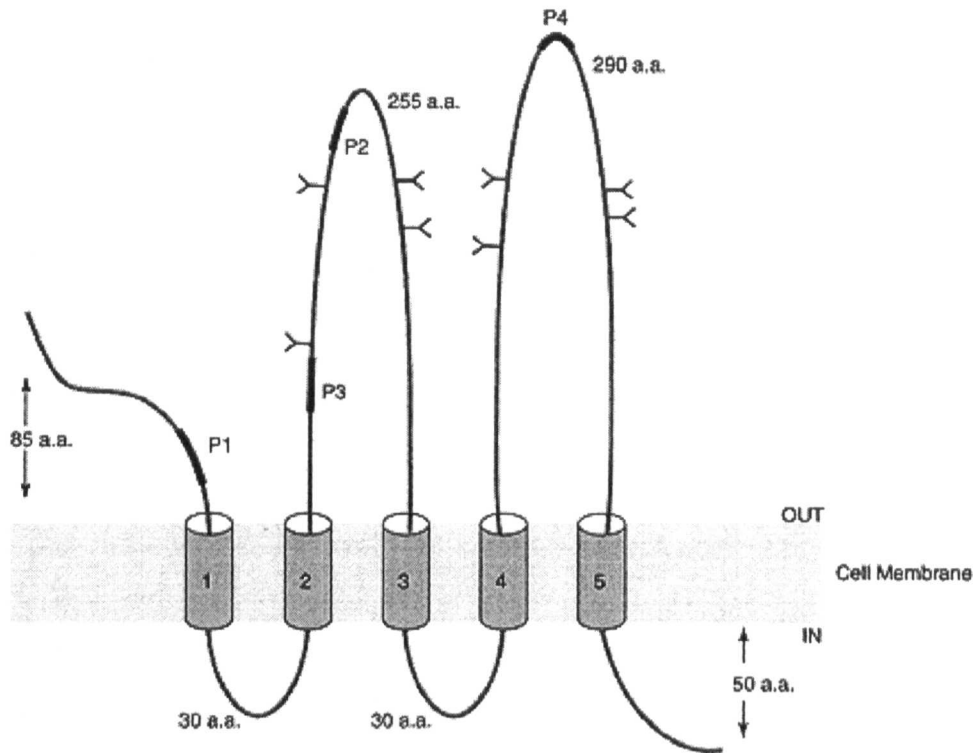
Figure 1.8: **Molecular structure of the CD133 antigen.**

The CD133 molecule (early known as AC133) is unusually predicted to have 5 transmembrane domains. Its actual 3D structure and function are unknown to date.

1.1.4 The CD133 antigen

CD133 (early referred to as AC133, see fig. 1.8) is a highly hydrophobic 865 amino acid (120kDa) glycoprotein. It is predicted to have five transmembrane domains and possesses 8 N-linked glycosylation sites. CD133 is the human equivalent, or ortholog, of the murine molecule prominin 1 [Fargeas *et al*, 2003; Yin *et al*, 1997; Miraglia *et al*, 1997; Corbeil *et al*, 1998].

The CD133 expression is predominantly restricted to a small HSPC subset of CD34+ cells [Gallacher *et al*, 2000; Corbeil *et al*, 2000]. It is also found on

primitive endothelial cells and several non haemopoietic tissues including the foetal brain, pancreas and kidney [Corbeil *et al*, 1999; Miraglia *et al*, 1997; Uchida *et al*, 2000]. In the very early stages of the haematopoiesis, recent investigations showed that CD133 expression on CD34+ HSPC precedes that of CD34 [McGuckin *et al,* 2003b; Bhatia, 2001].

A very small fraction of blood cells are CD133+. About 0.52% of the bone marrow nucleated cells are CD133+, and 0.16% of the umbilical cord blood cells [de Wynter *et al,* 1998]. The function of this antigen remains unknown to date. Its structure and composition, with the presence of a leucine-zipper motif on the second extracellular loop, tend to suggest a possible dimerisation capability. Also considering its selective membrane location onto plasma pseudopod-like protrusions, CD133 could be involved in mediating signalling used for cell-to-cell communication [Corbeil *et al*, 2000; McGuckin *et al*, 1999].

CD133+ cells were found to have a higher proliferation potential than CD34+ cells and also to contain long term culture initiating cells (LTC-IC) at a higher frequency than CD34+ cells [de Wynter *et al*, 1998, Pasino *et al*, 2000; Matsumoto *et al*, 2000]. CD133+ cells were recently shown to potentially undergo trans-differentiation [Quirici *et al*, 2001]. This antigen is also an important clinical marker for the identification of certain cancers [Horn *et al*, 1999].

1.2. Microscopy techniques for cell imaging

Imaging techniques can be used to quantify features on the investigated object, in order to correlate these features with the object intrinsic properties. Cells are visualised using microscopy techniques [Stephens and Allan, 2003]. These microscopy techniques exploit particular properties of light.

Wide field microscopy is the most basic form of microscopy. The sample is placed underneath the objective lens and is illuminated from below. The light goes through the sample and is collected by the objective. The magnifying system projects the magnified image through the ocular piece, where it is visually observed by the experimenter. Digital cameras are increasingly used to capture the image, which allows to store it digitally in a computer drive, making it available for further use and/or analysis. The optical properties of the sample, such as the absorption coefficients of its constituents, affect the amount of light allowed to travel through it. Therefore, optically denser regions produce darker areas in the output image, which generally allows discriminating between the cells and their environment. The environment is often a more uniform background, and when it is not it can usually be characterised by its texture or other properties. This technique has major limitations when applied on biological samples. Indeed, the main abundant constituent of cells is water, which transmits light very well and does not produce a sufficient contrast to visualise the cells. For this reason, histological techniques use dyes that specifically bind to classes of molecules (proteins, DNA, lipids, etc.) which have a much higher absorption coefficient, therefore providing more contrast on the magnified image.

Several techniques based on the interference properties of light have been used to further improve the contrast provided by wide field microscopy. Condensers and filters can be used to selectively allow light rays of specific wavelength and phase to pass through the sample. The optical characteristics of the sample deviate or slow down the photons which are consequently shifted in phase and/or trajectory. This results in producing interferences on the final image. These interferences emphasise the contrast of the optical properties of the different regions of the sample and allow to visualise smaller details of the subcellular structures. Such approaches are commonly implemented in differential interference microscopy (DIC) or phase contrast microscopy [Parthasarathy, 2001].

Scientists may only be interested in visualising the distribution of a small set of molecules. As mentioned above, some dyes can be used to specifically visualise intracellular components such as lipids or proteins. However, for accurate identification of unique molecules, such dies reveal themselves not to be specific enough, and other markers have to be used. Monoclonal antibodies have increasingly become the markers of choice to bind specifically to the target of interest. Indeed, these molecules produced in animal hosts (mice, rabbits) have a very specific affinity to bind to molecular domains of the target molecule. Such affinity property allows to locate particular molecules, or even different isoforms and epitopes of given molecules. Monoclonal antibodies (moAbs) need to be conjugated with a marker, either a dye with high optical density (for wide field microscopy investigation) or a fluorochrome.

Fluorescence microscopy makes an extensive use of fluorochromes. The atoms of a fluorochrome reach an excited state when photons of appropriate wavelength

collide with them. Some of the energy carried by the photons is emitted back as a photons of lower energy/different wavelength. This constitutes the fluorescence emission, which is captured by a digital camera and therefore used to locate the molecule of interest.

1.2.1. Laser-scanning confocal microscopy

Laser scanning confocal microscopy (LSCM) currently represents one of the

A TYPICAL LSCM SYSTEM

Lance Ladic
ladic@cs.ubc.ca
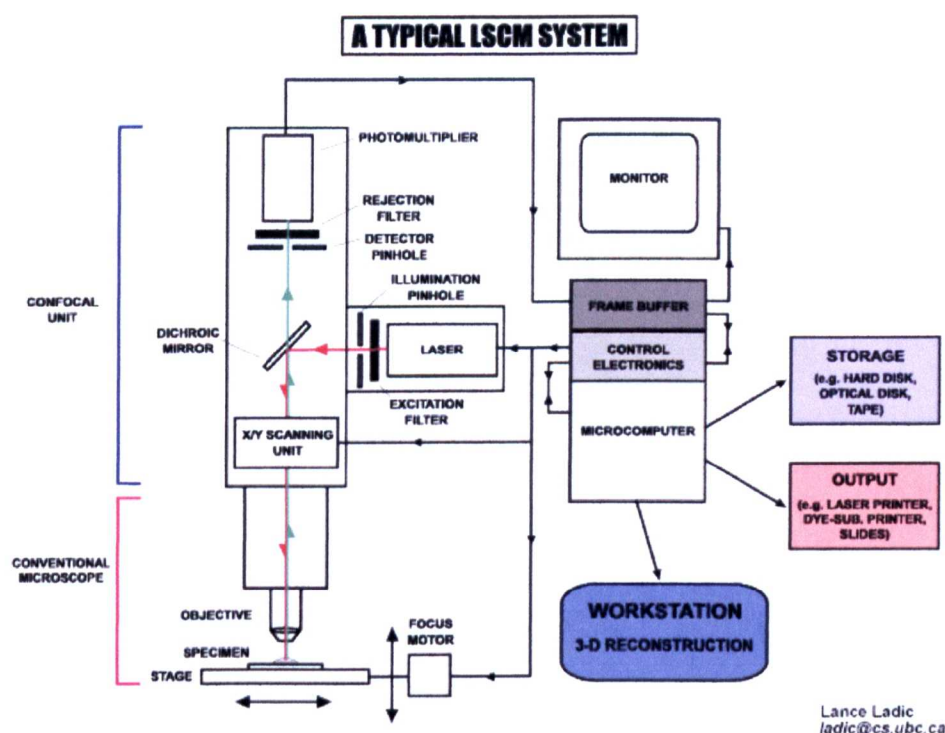
Figure 1.9: **Principles of Laser Scanning Confocal Microscopy (LSCM).**

A laser beam at specific wavelength is focused on a labelled specimen. The focal point of the focused light is scanned at constant Z elevation (the focal plane). The light emitted in the focal plane as a result of laser excitation travels backward towards a photomultiplier. Filters are used to separate different wavelength, and the pinhole aperture ensures that only the light coming from the focal point is detected. The LSCM images are stored in a computer and available for processing and analysis on a worksation.

state-of-the-art light microscopy techniques [Stelzer, 1998] (fig. 1.9). Samples are incubated with fluorochrome-conjugated moAbs which bind to the target molecules. The so-labelled samples are placed under the microscope objective and are illuminated by a laser of appropriate frequency in order to excite the fluorochrome. A thin circular aperture is placed between the sample and the objective, thus only the light coming from a small localised region in space (in the three-dimensional referential of the sample) is captured, amplified through photomultiplier tubes and digitally recorded (fig. 1.9). The focal point is scanned over the surface in order to capture an optical section of the specimen. The focal plane is then moved along the Z axis in order to image the complete 3D fluorescent structure under investigation (fig. 1.10). The capability of LSCM to physically remove out-of-focus light and to scan the specimen in 3D with high definition make this microscopy technique being increasingly used in biomedical studies to investigate particular structures and molecules of interest [Boutet de Monvel *et al*, 2001; Guo et al, 2001; Liu *et al*, 1997; Swedlow et al, 2002].
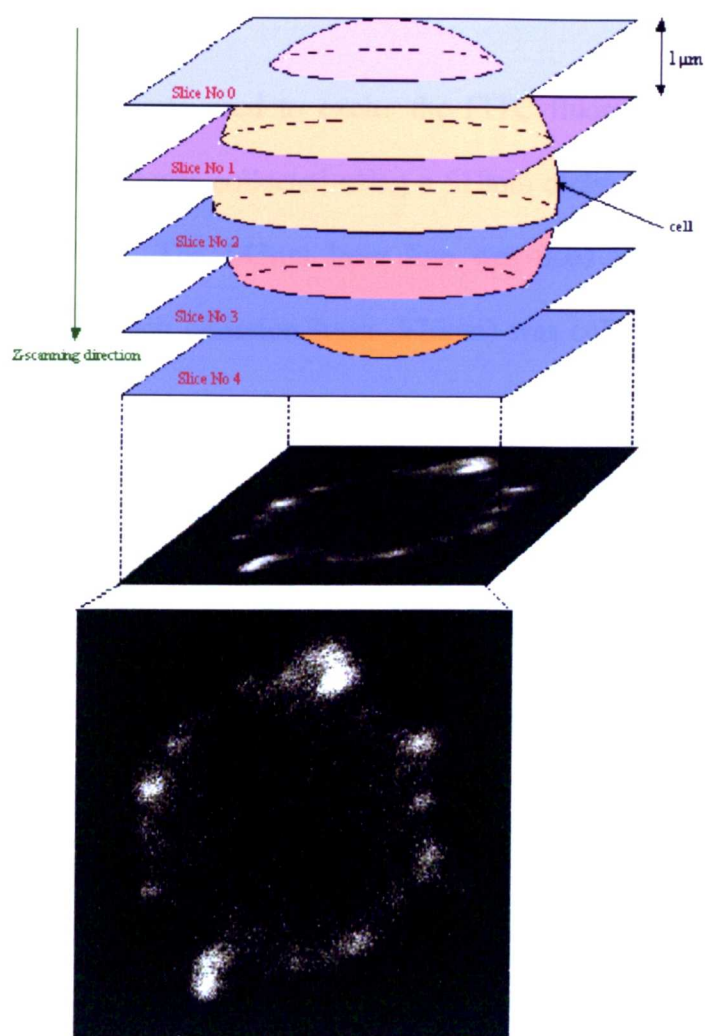
Figure 1.10: **LSCM Z-series.**

Confocal Z-series are produced by imaging optical Z-sections at equally spaced intervals along the vertical axis, or Z-axis. In this example, a Z-section through a cell labelled for CD34 is shown (bottom). The fluorescence emitted by the fluorochrome-conjugated specific monoclonal antibodies bound to the CD34 antigens in the cell membrane. In this example, CD34 is distributed randomly around the cell as dense or weak clusters and a diffuse halo. The image is affected by a strong noise component.

In this study, images were acquired with a Zeiss LSM.440 LSCM equipped with

Argon Ion laser providing to excitation laser lines of 488 and 543nm respectively.

Fluorescence and DIC images were acquired via photomultiplier tubes employing

a 63 x oil objective lens. Attached x10 condensation lens and computerised "false" magnification (up to x8) provided further means for image enhancement.

The 488nm laser line was used to excite the FITC fluorescence signal which emission (max .520nm) was collected using a 510nm dichroic mirror and a band pass 515-540nm filter. The 543nm laser line was used to excite the TRITC fluorescence signal which emission (max .576nm) was collected using a 560nm

A)

Differential Interference Contrast (DIC) image taken through the middle of the cell body (adherent stem/progenitor cell).

B)

LSCM optical sections (z-scans), taken 2 micrometers apart along the z-axis
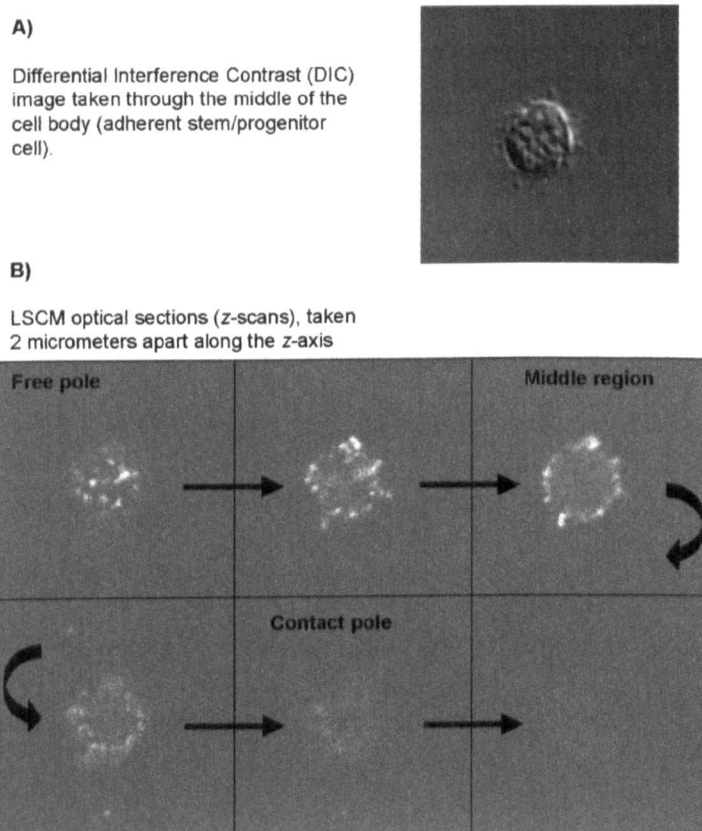
Figure 1.11: **Images obtained by LSCM.**

Top – DIC image taken in the middle of the cell body. DIC images are useful to assess the morphology of the cells. This cell is very spherical but displays membrane protrusions, a sign of membrane activity. *Bottom* – Z-series of the cell displayed on the top. The Z-sections still reflect the spherical shape of the cell along the Z axis. In the middle region, corresponding to the Z level where the DIC is taken, the positions of fluorescent antigen clusters are spatially correlated with membrane protrusions.

dichroic mirror and a long pass 610nm filter.

The specimens were manually scanned at low power in order to target cells of interest. Enhanced fluorescence and DIC images (fig. 1.11-A) of the cells middle focal plane were acquired and stored on the computer system. For each fluorescent signal (i.e. each labelled antigen) 1μm equidistant series of z-section were acquired from the bottom to the top of the cell (fig. 1.11-B).

The digitation of LSCM images further allows the development of computer image processing algorithms which are widely used nowadays to extract information from such images (and now generally provided in user-friendly packages with confocal microscopes). For most digital image analysis applications, the first step of the process is to identify and/or extract the object of interest in the image (e.g. cells, organelles, tissues), a process called "segmentation". For high magnification LSCM images, objects to be segmented are the image regions were fluorescence signal is present [Xavier et al, 2001, Wählby et al, 2002]. Various approaches have been described to perform image segmentation, depending on the nature of the objects to be segmented. In section 1.3, some common approaches are reviewed and commented.

1.2.2. LSCM image noise filtering

LSCM images are usually heavily contaminated by speckle noise. This occurs in the image acquisition process, during which a small amount of light collected at each pixel position and amplified through photomultiplier tubes. It is assumed that this noise is signal dependent (i.e. it is not additive). The distribution of such noise

is generally modelled with Poisson statistics, which are particularly well adapted to the description of rare discrete events (such as photon counts per pixel during LSCM image acquisition) [Rooms et al, 2004]. Also, as the number of event increases, Poisson distributions approach Gaussian distributions, which are commonly used as approximation of actual noise distributions prior to data processing.
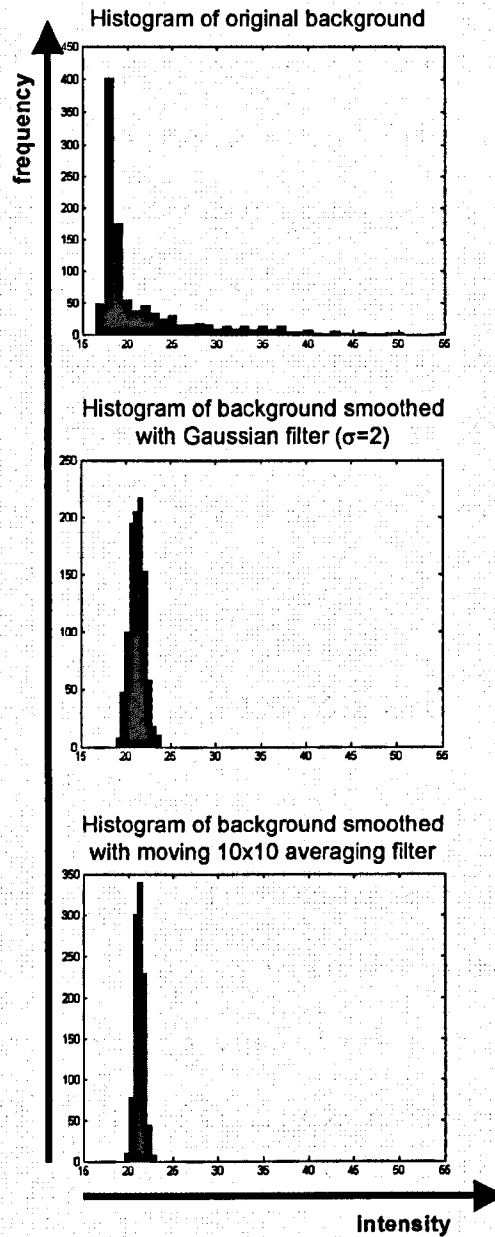
Figure 1.12: Effects of filtering on background regions histograms.

Top – Histogram of original noisy background. The shape approaches a Poisson distribution. There is only one main peak close to the mean image intensity. The long tail spans over a wide range of intensities, a large proportion of which also belong to signal pixels in the whole image. *Middle* – Histogram of the same background region after filtering with a 10x10 Gaussian filter. The shape now approaches a Gaussian distribution. *Bottom* – Histogram of the same region filtered with a 10x10 moving averaging filter. The distribution is still Gaussian, but its standard deviation is smaller than after Gaussian filtering.

Common ways of reducing the noise consist in filtering the images with low pass

filters, mostly Gaussian and averaging kernels, or empirically using rank filters

such as median filters. Figure 1.12 shows an actual measured noise distribution which has characteristics of Poisson noise. The filtering of the LSCM image used to obtain this figure with a gaussian kernel (10x10 pixels, $\sigma=2$) reduces this noise to a more Gaussian-like distribution spanning over a smaller range of intensities. However, the simple use of a 10x10 averaging filter also generates a Gaussian-like distribution spanning over a smaller range of intensities. For this reason and others described later in the text (section 3.1), 10x10 averaging filters were applied consistently prior to applying segmentation algorithms.

1.2.3. Analysis of fluorescence in LSCM images

Fluorescence images are used to locate the distribution of fluorochrome-conjugated molecular markers in biological samples [Spear *et al*, 1999]. Such spatial distribution can be used to get insights on the molecular patterns of the molecule under investigation, at various levels of observations. Some applications may consider the analysis of tissues (i.e. masses comprising large number of spatially organised cells). Other applications may focus on the analysis of some individual cells in cell populations (such as HSPC in the blood), which is done at the "cellular level of observation", while other studies may concentrate on the analysis of structures at the "subcellular level of observation" (i.e. structures inside cells themselves, such as particular organelles or clusters of receptor molecules in cell membranes).

Z cross sections	Binary masks of fluorescence	Mean fluorescence intensity
Free pole		1.149
Equator		1.015
Contact pole		0.647
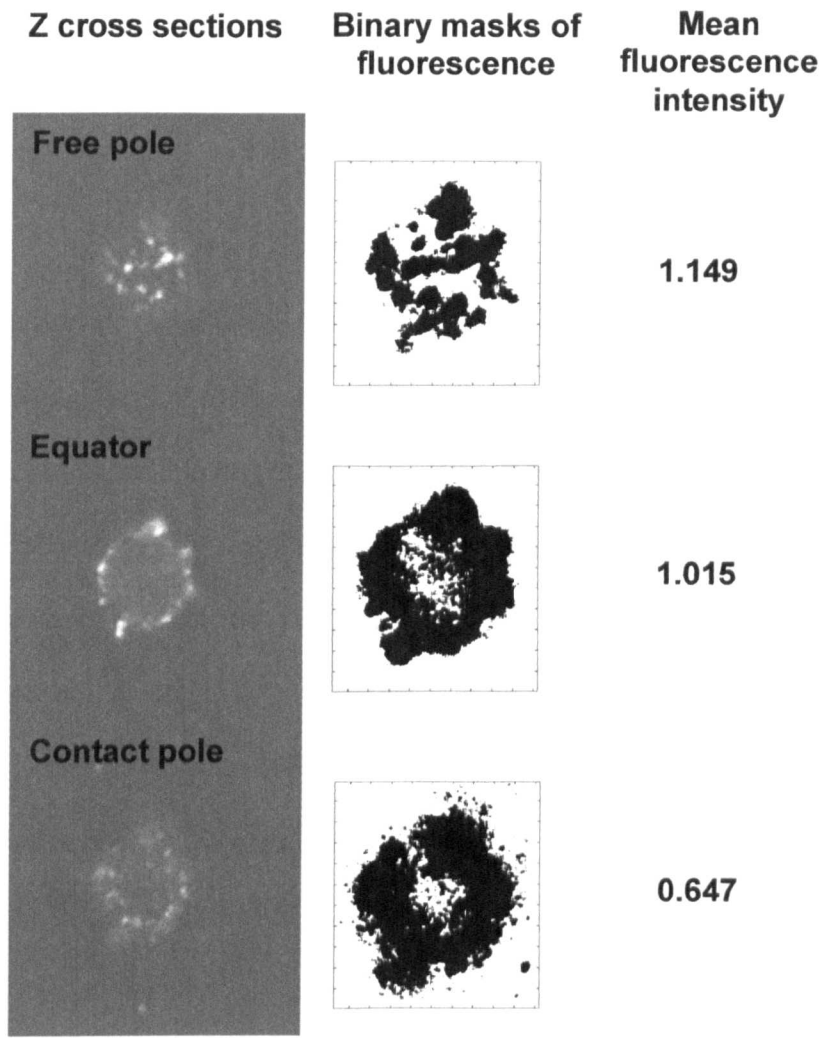
Figure 1.13: **Antigen quantification using total fluorescence area or mean fluorescence intensity.**

Left – Three LSCM images taken at decreasing Z positions. *Middle* – Binary masks of signal regions. As the Z position decreases, the surface covered by the signal increases, so quantify the size of signal regions would give increasing values. These regions were obtained by manual thresholding. *Right* – Mean fluorescence intensity in signal regions (arbitrary units). As the Z position decreases, the mean signal intensity decreases. Antigen quantification using these two common approaches provides with contradictory results.

In biomedical research, it is often relevant to perform measurements on the fluorescence signal [Andrews *et al*, 2002; Danckaert *et al*, 2002]. Such measurements are generally used to quantify the degree of expression of the molecule of interest [Matsuoka *et al*, 2001; Nishio *et al*, 1996]. This can be relevant since the expression level of some markers is directly related to

biological features [Ito *et al*, 2000; McCloskey *et al*, 2001]. For instance, the expression level of the antigen Ki-67 in the membrane of cancerous lymphoma cells directly reflects the position, or grade, of the tumour in the classification scale used to characterise the likely outcome of the disease [Matkowskyj *et al*, 2004].

Stem cell research employs membrane antigens to assess the degree of differentiation of the HSPCs. One way to perform quantitative measurements is by calculating the size of the area where the fluorescence is expressed (fig. 1.13, middle column). In this process, an appropriate threshold has to be selected either automatically or by the user. This latter option may introduce a subjective, human bias which could affect the quantitative analysis. It will be shown in this thesis that measuring the area of interest in fluorescence images is a very unstable criterion. Other groups express the fluorescence degree by measuring the intensity of the fluorescence in the area of interest (fig. 1.13, right column). Here again, an appropriate threshold has to be selected either automatically or by the user. The intensity of the signal is calculated by summing or averaging the fluorescence value from each pixel in the segmented region.

Fluorescence microscopy is also used to investigate the distribution of several markers at the same time, on the same sample. A particular technique looks at marker colocalisation. Indeed, colocalisation studies are employed to investigate the molecular and potentially functional relationships of colocalised antigens.

1.2.3.1. Colocalisation analysis methods

Colocalisation analysis and quantification is mostly performed through two approaches, the first takes into account the sizes of the areas shared by both markers, relative to the area occupied by one or the other marker (the AND/OR ratio performed by the free NIH software, http://rsb.info.nih.gov/nih-image/) [Leitinger and Hogg, 2002], while the second takes into account the relative contributions of the intensity of each fluorescent signal to the total intensity in colocalised regions [Matkowskuj *et al*, 2004].

Prior to colocalisation analysis, dual LSCM images must be perfectly aligned. Indeed, because the different fluorochromes have different excitation and emission wavelengths, fluorescence light beams respond differently to their passage through the medium used in the experimental set up and through the glass of the coverslip and the optical parts of the microscope. It results in shifted images, which cannot be analysed under the assumption of pixel correspondence as this introduces a bias in the measurements. Therefore, methods to correct the shift have to be employed. We will present a method based on correlation measurements which performs this task robustly.

1.2.3.1.1. Quantitative colocalisation analysis: the area method

This method involves the correct identification of signal (fluorescence) areas. Such areas are thresholded either manually (which happens in most cases, as biomedical scientists prefer to have control on this step in order to prevent under or over segmentation) or using various algorithms. The simplest segmentation

approach is global thresholding which consists in thresholding the image based on the properties of its grey histogram (reviewed in section 1.3.1). Other methods usually used in biomedical image analysis designed for specific applications may make use of active contour models (e.g. snakes) to delimitate the boundaries of compact fluorescent regions (e.g. nucleus in Fluorescent *In Situ* Hybridation – FISH – studies). Such spatial approaches are also reviewed and commented in section 1.3.3.

Fluorescence area segmentation using global image thresholding was preferred in this work due to its computational efficiency. Also, conversely to many region-based segmentation techniques, the number of parameters used to achieve appropriate segmentation could be reduced to a minimum and prevent the use of empirically set up parameters (which may affect the segmentation of images with various intensity distributions).

The manual selection of an appropriate threshold can reveal to be a very subjective task, and, as demonstrated in the present work, can lead to non-robust, unstable quantification of fluorescence areas. After signal segmentation, fluorescence quantification is performed by measuring the surface area of the total segmented region [Leitinger and Hogg, 2002; Zaho *et al*, 2002]. Colocalisation is then achieved by comparing the sizes of the areas shared by both markers relatively to the area occupied by one or the other marker (AND/OR ratio).

1.2.3.1.2. Quantitative colocalisation analysis: relative fluorescence method

This technique aims to express the contribution of each marker fluorescence in a colocalisation region to the total fluorescence in this region [Ortiz de Solorzano *et al*, 2002; Zaho *et al*, 2002; Ameen *et al*, 1999; Silver and Stryker, 2000]. Such an approach is more accurate than the area method, as it considers the various intensity levels present in a region instead of simply calculating the size of the region, but it requires an accurate calibration of each marker fluorescence. A drawback of the method is that intensity calibration is not necessarily achieved by biomedical scientists. Instead, the observed intensity in commonly adjusted mostly using brightness and contrast adjustments whose functions are commonly provided with imaging software. Generally, the parameters subjectively used for such intensity mappings are not known.

1.2.3.1.3. Quantitative colocalisation analysis: 2D intensity histogram and Pearson's correlation coefficient

Another method makes use of the computation of an alternative to the cooccurence matrix used to analyse textures in individual images. Two LSCM images taken at the same position are analysed. For each pixel position in the pair of images, the element of a matrix whose rows correspond to the intensity value for the first image while the columns correspond to the intensity values are incremented by one each time such intensity pair is encountered in the dual images. The resultant matrix is a 2D histogram of the dual images. Strictly colocalised fluorescence distributions would produce a 2D histogram with values

of higher magnitude closely distributed along the main diagonal of the matrix, while poorly or non-colocalised distribution tend to display 2D histograms with high magnitude values away from or scattered around the main diagonal [Enderle *et al*, 1997; Guild *et al*, 2001; Malide *et al*, 2001]. In addition to the analysis of the 2D histogram matrix, the Pearson's correlation coefficient R is calculated between the images of the pair and can be used to help quantifying the degree of colocalisation of the fluorescence distributions [Heilbronn *et al*, 2003; Kreft *et al*, 2004; Phee *et al*, 2001].

These methods to characterise colocalisation properties of dual labelled antigen distributions are generally performed on 2D optical sections taken at the middle of cell (equatorial Z section) where the best contrast is usually achieved. A careful visual analysis of highly colocalised CD34/CD164 distributions further showed that two types of colocalisation patterns are observed (figure 1.14). Indeed, in some membrane regions, the expression of one molecule is linearly related to the expression of the other antigen. Such colocalisation patterns are referred to as *linear* colocalisation. In other regions, despite both antigens also being expressed at identical locations, the linear relation is not observed and one antigen may be expressed in an apparently unrelated manner in comparison to the other one. Such a situation is consequently referred to as *differential* colocalisation. Various degrees of linear and differential colocalisation can be observed in cell membranes. Approaches for quantitatively characterising such patterns are investigated in this thesis.
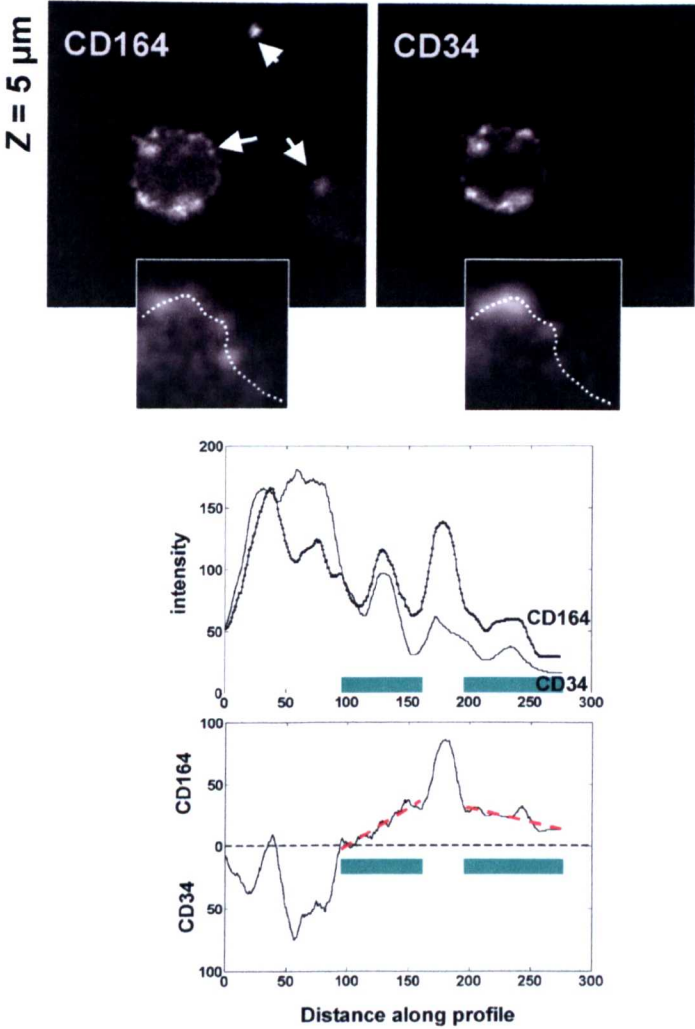
Figure 1.14: **Antigen colocalisation.**

Top – LSCM micrographs taken at the equatorial focal plane, for CD164 and CD34 antigens. They are very similar (demonstrating antigen colocalisation). The arrows point at fluorescent structures found only in the CD164 image proving that the observed colocalisation is not an experimental artefact. The insets represent magnified portions of the cell membrane. The dashed line indicates the position of a path along which the fluorescence for both antigens was measured. Both fluorescence curves are plotted on the graph below. *Top graph* – It can be observed that some parts of the curves follow similar behaviours, as if the fluorescence of one channel was linearly related to the fluorescence from the other channel. In some other parts of the curves, their evolution seems unrelated to each other, despite true fluorescence expression. *Bottom graph* – Intensity difference between both paths. Perfect colocalisation would be reflected by small fluctuations around zero. However it can be observed that some parts of the curve strongly fluctuate far from zero, a consequence of the predominance of one antigen over the other one (despite both being expressed at the same location). Such case is referred to as *differential* colocalisation in the thesis. In some other regions (marked with green segments and red dashed lines) the expression of one antigen is more linearly related to the expression of the other one, and the term *linear* colocalisation is used.

38

1.2.3.2. Visualisation: 3D representation of fluorescence distribution

A usual way of getting insights on the antigen distribution in its natural three-dimensional space is by computing the projection of the fluorescence intensity along a particular direction [Constans, 2004; Sarti *et al*, 2000; Daly *et al*, 2002]. Such projection is two-dimensional by definition. Therefore it does not represent an actual 3D reconstruction of the distribution. For this purpose, techniques borrowed from the field of computer graphics may be used.

1.2.3.2.1. Isosurfaces

Isosurfaces (fig. 1.15, middle and right) are surfaces which link regions of a defined constant intensity through the whole z-stack of LSCM images. This process uses intensity interpolation to estimate the intensities values between the optical sections [Sarti *et al*, 2000; Daly *et al*, 2002, Constans, 2004]. The selection of only one intensity level to locate the position of the isosurface acts as a thresholding procedure [Gerlich *et al*, 2001]. As a consequence, this approach does not allow the visualisation of the 3D structures with intensity values above the threshold, which are enclosed by the isosurface. Furthermore, the lower intensities are not visualised either, and low, weaker signals are not represented in the artificial 3D space. Several layers of isosurfaces set up at various intensity levels and with various degrees of transparency may be used to increase the level of observable details. However, this solution requires a large number of calculations, which slows the display down. The presence of noise also affects the 3D rendering as it makes the surface look rough and highly irregular. Smoothing

procedures are likely to improve the final rendering of the isosurface, but it may be at the cost of losing some of the original resolution.

1.2.3.2.2. Isocontours

Isocontours methods consist in drawing contours at N constant, but different, intensity values, usually separated by a constant incremental value. The contours are calculated in the 2D plane of the LSCM image, and are displayed at their respective z positions along the z-axis of the artificial 3D space (fig. 1.15, left).
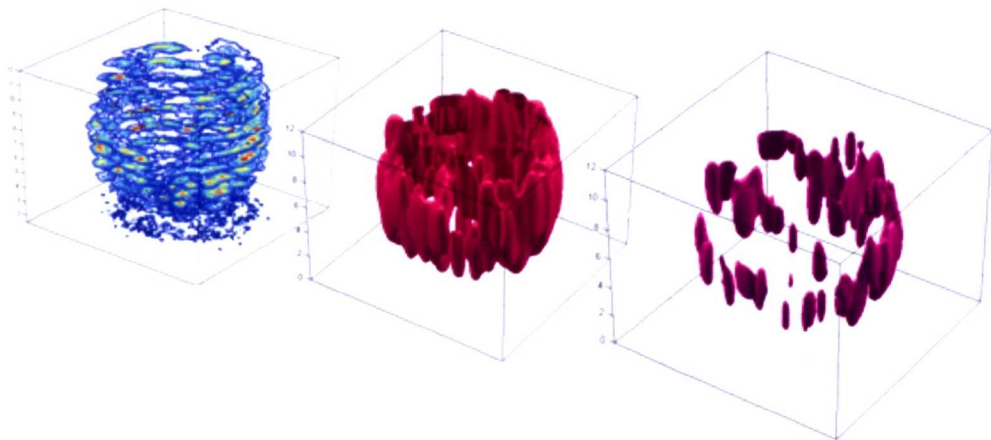
Figure 1.15: **Common 3D representations of antigen distribution from LSCM data.**

Left – Isocontours method. Lines of iso-intensities are represented in a 3D referential. This methods allows to see various antigen concentrations in the same figure. *Middle and left* – Isosurface method. Isosurfaces envelop 3D regions containing pixels with intensities equal or lower than a fixed threshold. Due to the unimodal nature of the intensity distribution in LSCM images, a small variation of the threshold value generates large variations of the volume enclosed by the isosurface. The three reconstructions are from the same LSCM dataset and the view angle is identical. The middle and right distributions are obtained by using close thresholds to illustrate the qualitatively different reconstructions, consequence of histogram unimodality. The distribution on the right shows the clusters with high antigen content while the middle one also include lower intensity halo regions.

Consequently, the computation of the isocontours at a particular z level does not affect the computation of the isocontours in the adjacent z sections. The continuity of 3D fluorescent structures is visualised by selecting manually the best view points, and the user makes a subjective interpretation of the resultant figure. Because there are several isocontours used in each LSCM image, it is possible to visualise structures of various sizes and intensities at the same z position, therefore compensating for the thresholding effect of the isosurface method. Isocontours are faster to implement than isosurfaces, as no interpolation between z-scans is used [McGuckin *et al*, 2003].

1.2.3.2.3. Clouds of points

Point clouds are an interesting alternative to isosurfaces and isocontours for 3D rendering of spatial antigen distributions. The principle is to display more points in the spatial regions with higher intensity values (i.e. regions of higher densities), while lesser points are used in low density regions, and no points at all in background regions [Constans *et al*, 2004]. The number of points to be displayed at every location can be calculated using some mapping functions, not necessarily linear. Interpolation between z-scans can also be used to display points between the optical sections, but can affect the rendering by surcharging the artificial 3D space. Colours of the points can be chosen according to the local densities in small neighbourhoods. Because the whole spatial antigen distribution is represented by a cloud of points, if the densities of points are locally well chosen, this approach allows to visualise at the same time all the structures in the cell membrane, for all intensity levels. Therefore, both weak and high signal regions are represented,

which also compensates for the thresholding effect of the isosurface method. Possibly, for these three 3D visualisation methods, transparency could be included in order to improve the clarity of the rendering.

1.3. Image segmentation methods

Fluorescence LSCM image analysis involves the prior segmentation of the fluorescent signal. Image segmentation describes a process in which regions or objects sharing similar characteristics are identified. It is often used to distinguish foreground objects from the surrounding background [Davies, 1990; Gonzales and Woods, 2002]. In these sections, some common approaches to segment grey scale images are described. Various approaches may be employed, some being global (a threshold is applied to the whole image simultaneously) and some others being region-based (thresholds are chosen locally, in sub-regions of the original image).

1.3.1. Global thresholding

In order to segment signal from background values, a single threshold may be applied simultaneously to all pixels in an image. This is global thresholding. The threshold is commonly selected after analysis of the image intensity histogram.

[Tsai, 1995] describes a method for segmenting either plurimodal or unimodal histograms into N classes, based on their shape, under the assumption that classes (homogeneous regions in the spatial domain) are represented by peaks in the histogram. The histogram peaks have to be separated by local minima. In the case of unimodal histogram, it is assumed that two peaks (corresponding to two classes with poor between-class contrast) are too close to each other, but their boundary can be found by locating a between-class curve discontinuity, also assumed to be the maximum curvature, in the histogram.

The number of classes N is chosen by the user. The small fluctuations in the original histogram are iteratively smoothed using convolution with a Gaussian kernel. Similar histogram smoothing is also used by [Draghici, 1997]. After each iteration, the number of maxima and minima is determined. The process ends when the desired number of peaks is obtained. The minima locations between the peaks indicate the thresholds. When the number of classes found is lower than the expected number N, is it hypothesised that two neighbouring peaks were merged during the blurring process. In this instance, the smoothed histogram may be discontinuous at the intersection point of the two overlapping distributions. This point is detected using a local maximum curvature method which does not explore points in the neighbourhood of the peak. This non-parametric approach is shown to be computationally more efficient than the well known between-class variance ([Otsu, 1979] described below) or the entropy methods, which do not work well if the object classes are too small, or the contrast between objects too low.

[Otsu, 1979] proposed a method based on discriminant analysis which maximises a measure of class separability. The criterion function to maximise is given by the following formula (*h(z)* represents the PDF measured from the image for any intensity *z*, and *T* the threshold which is moved along the histogram)

$$J_{OT}(T) = \frac{P_1(T)P_2(T)[m_1(T) - m_2(T)]^2}{P_1(T)\sigma_1^{\ 2}(T) + P_2(T)\sigma_2^{\ 2}(T)}$$

where

$$P_1(T) = \Pr\{C_1\} = \sum_{z=0}^{T} h(z)$$

$$P_2(T) = \Pr\{C_2\} = \sum_{z=T+1}^{L-1} h(z) = 1 - P_1(T)$$

$$m_1(T) = \sum_{z=0}^{T} z\Pr\{z|C_1\} = \frac{1}{P_1(T)} \sum_{z=0}^{T} zh(z)$$

$$m_2(T) = \sum_{z=T+1}^{L-1} z\Pr\{z|C_2\} = \frac{1}{P_2(T)} \sum_{z=T+1}^{L-1} zh(z)$$

$$\sigma_1(T) = \sum_{z=0}^{T} [z - m_1(T)]^2 \Pr\{z|C_1\} = \frac{1}{P_1(T)} \sum_{z=0}^{T} [z - m_1(T)]^2 h(z)$$

$$\sigma_2(T) = \sum_{z=T+1}^{L-1} [z - m_2(T)]^2 \Pr\{z|C_2\} = \frac{1}{P_2(T)} \sum_{z=T+1}^{L-1} [z - m_2(T)]^2 h(z).$$

Otsu's method has become one of the most popular optimal, global thresholding technique from grey intensity histograms. [Cheriet *et al*, 1998] also extended this algorithm through a recursive technique for multimodal histogram segmentation. These authors initially pre-process their images using a 3x3 averaging filter, in order to get better, smoother histogram for further analysis. The algorithm has been reported to work well when both classes are well separated, which involves that the class means are as well separated as possible, and their variances as small as possible. However, the stronger the histogram unimodality, the poorer the segmentation using Otsu's method, as two distinguishable, Normally distributed classes are not easily detectable in the histogram [Petrou, 1999].

Similar discriminant analysis has been investigated by other authors. [Kittler *et al*, 1986] proposed a minimum error thresholding method. Their criterion function is defined as

$$J_{KI} = \sum_{z=0}^{L-1} h(z)c_{KI}(z,T) = \sum_{z=0}^{T} h(z)c_{KI}^{(1)}(z,T) + \sum_{z=T+1}^{L-1} h(z)c_{KI}^{(2)}(z,T)$$

where $c_{KI}^{(1)}$ and $c_{KI}^{(2)}$ are considered to be two parts of a cost function c_{KI}. This cost function is derived based on Bayes' rule. Under the assumption that both classes to be segmented have Gaussian distributions, the criterion function becomes

$$J_{KI}(T) = 1 + 2[P_1(T)\ln\sigma_1(T) + P_2(T)\ln\sigma_2(T)] - 2[P_1(T)\ln P_1(T) + P_2(T)\ln P_2(T)].$$

The optimal threshold is given by maximising this function, i.e.

$$T_{KI}^{\bullet} = \arg\min_{0 \le T \le L\text{-}1} J_{KI}(T)\,.$$

[Huang *et al*, 1995] proposed to select a threshold based on fuzzy membership functions and entropy measurements. The image is seen as an array of fuzzy singletons (the pixels), each having a membership value associated with a certain feature of the pixel. An image I is represented as

$$I = \{(f(x,y), \mu_1(f(x,y)))\}$$

where $\mu_1(f(x,y))$ is the membership function which defines the grade of pixels belonging either to the background or foreground class. The criterion function is

$$J_{HW}(T) = E(I) = -\frac{1}{MN\ln 2}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} S_e(\mu_I(f(x,y))$$

where

$$S_e(\mu) = -\mu \ln \mu - (1 - \mu) \ln(1 - \mu) \qquad 0 \le \mu \le 1$$

used as a cost function, represents Shannon's entropy. The optimal threshold is given by the value T that minimises $J_{HW}(T)$. This approach classifies the pixels in two classes, as close as possible to their class mean intensity. Like Otsu's algorithm, the class means need to be well separated and their variances as small as possible.

[Yan, 1996] showed that [Otsu, 1979], [Huang et al., 1995] and [Kittler et al., 1986] share similar formalism despite their different approaches.

[Liao *et al*, 2001] proposed a faster, recursive extension of Otsu's method to segment plurimodal histograms. A global optimisation method which seeks to maximise the between-N-classes variance is described. Otsu's between-class variance criterion is modified so its computation is faster. Its actual measurement is computed from zero- and first-order moments of classes defined over known ranges of grey intensities. These quantities are calculated from the histogram as the thresholds are moved along the grey scale, and stored in two look-up tables. The modified between-class variance criterion is computed from both look-up table and is a function of the N-1 thresholds. The optimal combination of thresholds is the one that maximises the between-class variance criterion.

[Rosin, 2001] specifically addresses the problem of unimodal histogram thresholding *via* an empirical approach, similar to [Tsai, 1995]. It is hypothesised that the unimodality of the histogram is due to a dominant class, with lower intensities, which prevents to discern a second, smaller peak at higher intensities (the signal). The point of maximum curvature is identified using the standard

recursive subdivision technique for determining the polynomial approximation of a curve, the point of maximum histogram curvature between the bin with highest value and the first empty bin following the last filled bin. This method works on the basis of histogram shape assumption, and therefore the approach works robustly to detect the maximum curvature point of various curves. However there is no theoretical justification of its validity, since no physical or mathematical explanation is provided regarding the relevance of such a point having fundamental properties of discriminating the object from the background, and therefore can only work on certain type of images. Similar approach is adopted in [Snidaro and Foresti, 2003; Hervás *et al*, 2003].

The author notes that current thresholding methods from grey level histograms require critical parameters to specify the amount of preliminary smoothing applied to the histogram in order to remove spurious, insignificant extrema and the size of the region of support over which the histogram curvature is calculated. This implies that such parameters have to be found empirically, which does not make the algorithms fully automated.

[Antoine *et al*, 2001] also describe a method based on the shape analysis of weakly bimodal or unimodal histogram curves, where critical inflexion points are found using the shape properties of the histogram and its first derivative. This approach requires preliminary smoothing of the curve using a 4th order Butterworth filter in order to remove high frequency components. Landmark points such as beginning, end, and highest points of the curve are detected. The histogram is differentiated and smoothed using a running averaging operator with supporting region of size 21 bins. Using this derivative, several conditions are

then tested to determine if either a local minimum is present between two histogram peaks, or an inflexion point if both peaks are not well separated. It is noted that when the histogram is very much unimodal in shape, the effect of the initial smoothing procedure is too strong and the shoulder should be looked for on the non-smoothed curve, which may introduce some error on the threshold decision due to noise considerations. Strong unimodal histograms are detected and the algorithm cannot threshold these.

[Li and Tam, 1998] propose an iterative algorithm to speed up the threshold evaluation from the minimum cross-entropy method calculated from a histogram h. The optimal threshold is found at the value t that minimises the criterion function

$$\eta(t) = -m_{1a}(t)\log(\mu_a(t)) - m_{1b}(t)\log(\mu_b(t)), \quad t \in [1, L]$$

where

$$\mu_a(t) = \frac{m_{1a}(t)}{m_{0a}(t)} = \frac{\sum_{i=1}^{t-1} h(i)}{\sum_{i=1}^{t-1} ih(i)} \quad \text{and} \quad \mu_b(t) = \frac{m_{1b}(t)}{m_{0b}(t)} = \frac{\sum_{i=t}^{L} h(i)}{\sum_{i=t}^{L} ih(i)}.$$

i representing the intensity value and h the PDF. By definition, the thresholding algorithm looks for t that gives minimum value to $\eta(t)$. The derivative of $\eta(t)$ is calculated and simplified when equalling zeros (which indicates the minimum of $\eta(t)$). Looking for non trivial solutions and expressing in terms of t, we get

$$t = \frac{\mu_b(t) - \mu_a(t)}{\log(\mu_b(t)) - \log(\mu_a(t))}.$$

Applying the one point iteration method to this formula leads to

$$t_{n+1} = \text{round}\left\{\frac{\mu_b(t_n) - \mu_a(t_n)}{\log(\mu_b(t_n)) - \log(\mu_a(t_n))}\right\}.$$

The convergence is achieved when $t_{n+1} = t_n$. It is shown to take approximately five to nine iterations to converge to the desired threshold, instead of having to calculate all the values of $\eta(t)$. Therefore the iterative method is faster than the conventional exhaustive search, but the author notes that it does not always lead to the correct threshold. Tests and performances are measured on artificial histograms, and only one microscope image of stained nuclei is used as a potential application for the proposed method.

1.3.2. Comments on global, histogram-based thresholding methods

Global thresholding is the fastest approach to segment fluorescence in LSCM images. However, most histogram-based algorithms found in the literature address bi- or multimodal histograms thresholding. As described above, high magnification LSCM micrographs of HSPC have a strongly unimodal histogram, due to the strength of the speckle noise in the predominant background areas. Therefore, most of the reviewed thresholding algorithm cannot be applied successfully to extract the fluorescence in the datasets analysed in the present study. The approach described in [Rosin, 2001] aims to threshold unimodal histograms, but the method is heuristic in nature and does not provide a theoretical background. Consequently, there is no way of ensuring that this algorithm is

reliable and that it can be blindly applied to segment our datasets. Thus, new and theoretically sound algorithms need to be developed in order to perform such task.

1.3.3. Region-based segmentation methods (spatial methods)

Region-based segmentation techniques rely upon feature similarities of image areas, i.e. image areas sharing same properties (e.g. texture, intensity) are assimilated to a common region. Several criteria can be used to further improve the definition of a region, such as connectivity or fuzzy membership [Gavrielides *et al*, 2001]. Below are described some common approaches. They are further commented.

1.3.3.1. Watershed segmentation

In this thesis, watershed segmentation is used in two sections. The first appearance of the algorithm is in section 3.5, where watersheds are used to describe and demonstrate the possibility of segmenting LSCM images using a region-based approach. However, this is done only for demonstration purpose and since the method is computationally more expensive than global thresholding approaches, it is not used to process the whole LSCM datasets. The second appearance of the watershed algorithm is in section 5.2.3.2, where it is used to segment fluorescent antigen clusters in dual labelled equatorial Z section prior to colocalisation analysis. This colocalisation analysis is then compared with an

alternative analysis performed on Antigen Density Maps (sections 5.2.3.3 and 6.4.6).

The concept of watershed segmentation involves interpreting the (grey level) image as a three-dimensional landscape, where two dimensions, the "geographical" coordinates, describe the spatial coordinates of the pixels, and the third dimension, the "altitude", is given by the grey value. This method was initially proposed by [Digabel and Lantuéjoul, 1978] and was further extended to a more general framework by [Lantuéjoul and Beucher, 1981]. In its common version, the three-dimensional image-landscape is fictionally submerged in "water". The water continuously fills the landscape, successively flooding it at the local minima positions. As the water level increases, the frontiers of these catchment basins, or watersheds, eventually reach a state where they merge. This frontier is labelled as a boundary between the watersheds. Generally, each time a watershed is about to merge with another, the location of the boundary between them is stored in memory until no further merging is possible (when the rising water has reached the highest – maximal – grey intensity value in the landscape). The final set of boundaries corresponds to the watershed lines, which delimitate the different objects in the image. Pixels belonging to a same catchment basin are attributed a same label value, so each segmented object, consequently delimitated by a continuous boundary, has a distinct and unique label [Haris et al, 1998; Murphy et al, 2001]. This approach has the advantage, under certain conditions, to segment clustered objects, and have been extensively used in cell image segmentation. However, pre-processing of the original images is often required, as noise or local irregularities rapidly lead to over-segmentation if not filtered or removed [Wählby et al, 2002].

This version of the watershed algorithm works well when the objects can be described as mounds and mountains within the 3D framework established above. However this case is not necessarily met, and a useful transformation of the image may be used to properly segment the regions of interest. Indeed, the watershed lines are located at local image maxima. When the objects cannot be described as mounds, mountains or peaks, it can be useful to apply this technique to the gradient image, where high gradient values are displayed with high intensities. In such a case, the high gradient regions are likely to represent the image object boundaries. Therefore, applying the watershed algorithm to gradient, or edge images can help to approximate the actual boundaries of clustered objects. Practically, the watershed algorithm is commonly implemented by using fast mathematical morphology operators.

In a pre-processing stage, over segmentation, a classical practical problem, can also be overcome by using sets of connected components, or markers (somewhat similar to the seeds used in the region growing algorithm). Indeed, using additional knowledge *via* markers may help to discriminate between the regions of potential interest versus the background [Haris *et al*, 1998]. Markers locating the objects are referred to as internal markers while those locating the background are referred to as external markers. Furthermore, the number of internal markers sets the number of watershed boundaries to a limited amount, which consequently helps preventing, or at least reducing, over-segmentation, while the external markers make the segmentation more robust.

1.3.3.2. Edge-based methods

This thesis does not make use of edge-based segmentation methods. However, they are briefly described in order to be commented, and also to describe the gradient of an image, its mathematical formulation and approximation using operators. Indeed, if the gradient is not used for segmentation purposes, it is used (section 5.2.4.2) to characterise the distribution of antigen clusters in the cell membrane, what will be referred to as "distribution energy".

The region-based methods described above are using some criteria of homogeneity to identify and segment image regions. Edge-based methods rely upon the detection of the boundaries between homogeneous areas to identify regions [Xavier *et al*, 2001]. A well-known algorithm making use of these techniques is the Canny edge detector [Canny, 1986].

Edges in images are regions where the intensity changes quickly, conversely to homogeneous regions where it is assumed that no significant change of intensity occurs quickly. Seeking the image as a three-dimensional landscape, edges are the areas with high slope value, the steepest the slope, the stronger (or sharper) the edge. Consequently, it is natural to use the gradient of an image to detect edges. Gradients are computed from the image function behaviour in a given neighbourhood, and are characterised by their magnitude and their direction. Edges are perpendicular to the direction of the gradient. Given a (presumably continuous) image $I(x,y)$, its gradient magnitude $|\text{grad } I(x,y)|$ and gradient direction ψ are calculated as

$$\left| \text{grad } I(x,y) \right| = \sqrt{\left(\frac{\partial I}{\partial x} \right)^2 + \left(\frac{\partial I}{\partial y} \right)^2}$$

$$\psi = \arg\left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)$$

where arg(x,y) is the angle, in radians, from the x axis to the point (x,y).

A variety of applications need only to consider the edge magnitude without regard to their orientation. A very popular linear differential operator, the Laplacian, can be used as it has same properties in all directions (it is rotation invariant), and is defined as

$$\nabla^2 I(x,y) = \frac{\partial^2 I(x,y)}{\partial x^2} + \frac{\partial^2 I(x,y)}{\partial x^2} .$$

Because digital images are discrete in nature, gradients can be approximated by differences along given directions. The interval over which is computed the difference is usually small (e.g. one pixel only). Operators estimating gradients in small neighbourhoods can be expressed as convolution masks. One of the oldest operators of this type is the Roberts operator [Roberts, 1965] which works on a 2x2 neighbourhood:

$$h_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad h_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

and therefore the edge magnitude is given by

$$\left| I(i,j) - I(i+1,j+1) \right| + \left| I(i,j+1) - I(i+1,j) \right| .$$

The main drawback of this operator, due to the small size of the supporting region over which it is computed, is its high sensitivity to noise.

Another popular operator for approximating the second derivative (i.e. gradient magnitude only) is the Laplacian, which can take several forms:

$$h = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad h = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

and sometimes the significance of the central pixel may be emphasised:

$$h = \begin{bmatrix} 2 & -1 & 2 \\ -1 & -4 & -1 \\ 2 & -1 & 2 \end{bmatrix} \qquad h = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix}.$$

Such kernels can have the disadvantage to respond doubly to some edges. Depending on the application and the strength of the expected response to the application of such operators, users may adjust the size of the kernels or the size of the images in order to get practically useful results.

1.3.3.3. Multiscale analysis and scale-spaces

Scale-spaces and image pyramids [Rezaee *et al*, 2000] are related forms of multiscale analysis. The mathematical formulation of 1D scale space analysis was first described by [Witkin, 1983] who showed how to analyse a one dimensional signal at different resolution by continuously smoothing the original input with a Gaussian kernel. This was further investigated in 2D spaces by [Koenderink,

1984]. In this approach, any image is shown to be embedded in a one parameter family of images derived from the original image, with resolution as the parameter. In its original formulation, the transformation of an original image *I(x)* into its associated scale image *I(x, σ)* can be described as the convolution of the image with a Gaussian kernel *G* of standard deviation *σ*

$$I(x,\sigma) = I(x) \otimes G(x,\sigma) \qquad\qquad 1.1$$

where $\otimes$ is the convolution operator and *G* is

$$G(x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} . \qquad\qquad 1.2$$

The higher the value of *σ*, the highest the degree of blurring. The parameter *σ* represents the scale at which the original image is observed [Kuijper and Florack, 2001]. The scale space is therefore constructed by smoothing the original image with Gaussian kernels of increasing standard deviation, which blurs *I(x)* increasingly, reducing its resolution.

This transformation respects the constraint that no spurious artefact is generated when decreasing the resolution, a condition known as the causality principle. This causality principle is a very important requirement for scale space theories. It states that every feature obtained at a coarser scale must have a cause at a finer scale. As an extreme example, a local maximum at coarse scale cannot occur if the initial image is totally flat. New image structures should not be introduced as scale is increased.

The dynamic behaviour of the intensity values in this scale space, for an infinite domain, has been described as a diffusion process [Koenderink, 1984]. The general form of the diffusion equation is given as

$$\frac{\partial I}{\partial t} = D\Delta I(x,t)$$ 1.3

where the parameter D controls the blurring of the image at each scale t. When D is constant, the process is called linear diffusion, but diffusion can also be anisotropic [Venegas-Martinez *et al*, 2002; Black *et al*, 1998]. Gaussian scale spaces are a solution of the diffusion equation 1.3. The diffusion process can be interpreted as a flow of intensity values from pixels towards the mean image intensity in scale space. Eventually, the diffusion process reaches a state of equilibrium ($t\rightarrow\infty$) where all intensities have converged to the mean intensity of the image.

Scale space image generation using equation 1.1 is computationally expensive for large sets of data, even when performed in the Fourier domain. In practice, generating scale space images can be approximated by reducing the size of the original image using intensity interpolation. This idea is the principle behind image pyramid formation, where the pyramids usually represent a discrete scale space image. Constructing and using image pyramids consists of successively down sampling the original image, possibly using (e.g. mean, bilinear) intensity interpolation, prior to further sampling it up and comparing the transformed image to the image at the same resolution. An homogeneity criterion can be used to identify regions having similar properties between these images. Therefore image pyramids are often used to identify and segment regions which bear homogeneous

properties over decreasing resolution levels [Rezaee *et al*, 2000; Venegas-Martinez *et al*, 2002]. Such an approach was recently adapted to segment uniform areas in images [Rezaee *et al*, 2000; Kuijper and Florack, 2001; Tabbone and Wendling, 2003; Rooms and Philips, 2002].

It was mentioned that for scale space generation using equation 1.1, when the original image is iterated an infinity of times, the theoretical output of this transformation is an uniform image having the mean intensity of the original image at any location. Similar dynamics also occur with discrete image pyramids when the image is successively resized to a constant fraction of its size from scale to scale using intensity interpolation. If mean intensity interpolation is used to calculate the pixel intensities over the scales, in discrete space, the process eventually leads to the creation of a single pixel at the top of the pyramid. Its intensity is the average intensity of the original image. This is an important remark which will be used in section 3.2 to produce a computationally efficient method for approximating a rescaling process used for segmentation purposes.

Algorithms using scale spaces for image segmentation rely on the analysis of the original image observed at different scales. Usually, features are located at coarser scales, and they are back-tracked at the finest scales or possibly in the original image [Rezaee *et al*, 2000; Venegas-Martinez *et al*, 2002; Kuijper and Florack, 2001; Tabbone and Wendling, 2003; Rooms and Philips, 2002].

There are other ways of defining scale spaces, briefly described here. The morphological scale space is defined using mathematical morphology. In such spaces, erosion and dilation are the primary operation used to construct opening and closing operators. The scale is defined by the size of the structural element

used in these operations (often a disk). Succession of opening and closing steps act as sieves which extract regions of particular size (which are scale-related) [Jalba et al., 2004]. Reaction-diffusion scale spaces combine both linear and morphological scale spaces. On grey scale images, morphological operators extract regions at different scales and smoothing is region size dependent (also consequently scale-dependent), being therefore inhomogeneous [Salembier and Serra, 1995].

1.3.3.4. Region growing

In its simplest form, region growing is initiated at a starting point of the image, or seed. The set of features describing the region to be segmented must be established prior to running the algorithm. Pixels or areas around the seed are then analysed, and if they share identical features with the description of the expected region they are clustered to the seed as belonging to the expected region. The process is iterated at the growing region border until no further pixels can be clustered and attributed to this region according to their feature description. This method has the advantage of segmenting homogeneous, consistent regions, due to the connectivity property of the growing process. This cannot necessarily be achieved through global thresholding. Limitations can be found when regions have complex descriptors or geometry. Considerations about the stopping rule, which is often based on local information, and therefore does not take into account the history of the region growth or its shape, may also affect the segmentation [Jin et al, 2003; Choi et al, 2001]. This problem is overcome by specifying and embedding these *a priori* criteria into the growing process

[Precioso *et al*, 2003; Yang and Jiang, 2001]. Another consideration is about the seed selection. This procedure obviously depends on the nature of the image and regions to be segmented. Such an approach was found to be valuable for image with multimodal histograms, usually under the assumption that the modes in the histogram are generated by homogenous grey regions in the image [Jin *et al*, 2003; Wirth and Stapinski, 2003].

1.3.3.5. Splitting and merging

As stated for the region growing method, the features – or predicate – describing the targeted region have to be known prior to running the split and merge algorithm. Here again, the predicate can be a vector for which elements represent quantities attributed to the target region, e.g. texture, intensity, or combinations of these. However, conversely to starting from seed points, the whole image is divided into several arbitrary, usually large disjointed regions [Bonton *et al*, 2002]. If one or more regions have properties close enough to the predicate, these regions are no further divided. Otherwise, they are subsequently divided into smaller regions in the next iteration and the process is repeated until a fixed resolution, or region size, is achieved. In order to avoid having separate adjacent regions with similar features, merging is simultaneously introduced [Haris *et al*, 1998; Choi *et al*, 2001], which allows to allocate and fuse such touching regions into a bigger, still uniform (according to the predicate) region.

1.3.4. General comments on segmentation approaches

All image segmentation algorithms reviewed above have specific drawbacks which prevent their application in automatically segmenting the LSCM images dealt with in this thesis.

1.3.4.1. Spatial methods

The spatial methods are computationally expensive. This is a main limitation to their application for segmenting large numbers of 1024x1024 LSCM images. Furthermore, they include several parameters which need to be estimated prior processing the image, or which may be set up subjectively. Despite the computational efficiency of histogram-based approaches, published algorithms are also subjected to drawbacks, described below.

Spatial methods (region-based methods) are likely to perform the fluorescence segmentation task successfully. However, their automation would require the careful set up of various parameters contained in the algorithm. For example, snakes require parameters to stop the growing process after the energy of the snake reaches stability. These parameters are evaluated subjectively [Santos *et al*, 2001], with the inherent risk that bias from human decision may be introduced in the process. Furthermore, spatial methods are computationally much more expensive than histogram-based methods. Taking these drawbacks into account, it seems preferable to investigate and develop automated histogram-based algorithm to perform the task of segmenting fluorescence in the LSCM micrograph datasets.

1.3.4.2. Texture measurements

Texture measurements are useful to discriminate between regions with different statistics or patterns [Malpica *et al*, 2003], and they are often used in various region-based segmentation algorithms (e.g. region growing, split-and-merge) [Basset *et al*, 2000]. However, the unimodal nature of the LSCM images prevents their use for segmentation purpose. Indeed, performing texture measurements on such images generates measures which follow themselves a unimodal distribution. There is a direct bijection between the intensities of smaller sub-regions and texture measures performed in these regions. The existence of such a bijective correspondence between intensity and texture measures makes the application of texture measures useless for segmenting LSCM images. Working directly on the intensities remains therefore the most likely computationally efficient approach for this task.

1.3.4.3. Scale spaces

Scale spaces (in the spatial domain) are attractive techniques as they analyse the images at various resolution levels, taking into account details easily observable at specific scales and describe the structures of an image in a tree-like hierarchy naturally well adapted for segmentation purposes. These algorithms have particularly low numbers of parameters, an important criterion for designing popular methods (Ockam's razor). However, the processing of large datasets is computationally expensive. Furthermore, their exploitation for segmentation

purposes requires the definition of some criteria, such as region homogeneity or texture, which are also subjectively defined, with the risk that human error or bias may be introduced in the process and add to the computational load.

1.3.4.4. Histogram-based methods

These techniques are computationally efficient and consequently are further investigated in the present work. The main drawback is that current algorithms do not perform well on images with unimodal histograms or have an empirical basis [Rosin, 2001; Wiltschi et al, 2000]. Defining a good threshold from histogram is not a straightforward task in most cases [Bonnet *et al*, 2002].Therefore new algorithms are developed and tested to address this problem.

Chapter 2

HSPC isolation, labelling and imaging

Chapter 2. HSPC collection, labelling and imaging

The process from HSPC isolation to laser scanning confocal imaging is described in this chapter, followed by a preliminary analysis of LSCM images.

2.1. HSPC isolation and labelling for confocal image acquisition

2.1.1. Haemopoietic stem/progenitor cell isolation

Most investigators rely on the antigen profile of haemopoietic stem/progenitor cells (HSPC) to either positively or negatively select HSPC from heterogeneous blood samples. Positive selection depends on monoclonal antibodies (conjugated to various compounds) raised specifically against markers expressed on primitive HSPC. These markers include the CD34 antigen and more recently the CD133 antigen, which are usually targeted for the selection of HSPC away from developing and differentiated cells.

2.1.2. Umbilical cord blood and bone marrow collection and mononucleated cell isolation

Umbilical cord blood (UCB) specimens were collected from full-term third stage of labour deliveries after elective caesarean sections. Bone marrow (BM) samples were obtained from informed haematologically normal volunteers. CB and BM specimens were collected following St George's Hospital Medical School ethical regulations.

Blood samples were mixed with acid citrate dextrose formula-A acid anti-coagulant (ACD-A, 0.6% final concentration, Baxter) immediately after collection

in a ratio 1:5. Four volumes of (blood + buffer) solution was carefully overlaid onto one volume of research grade Ficoll-Paque solution (d: $1.077g/cm^3$, Pharmacia Biotech, Uppsala, Sweden) prior to centrifugation (400g, 30 minutes, 22°C). The mononucleated cell (MNC) layer, at the interface of the ficoll and the lower density (plasma + buffer) solution, was collected with a sterile Pasteur pipette, washed twice in ACD-A buffer, and pelleted (400g, 10 minutes) before resuspension in ACD-A buffer. Cell aliquots taken for cell viability / enumeration using trypan blue (0.25% in PBS, Sigma-Aldrich).

2.1.3. CD133+ cell immunomagnetic positive selection

$CD133^{+/neg}$ cells were obtained from MNC after immunomagnetic separation using the CD133 mini-MACS selection kit (Miltenyi Biotec, Bergish Gladbach, Germany) following the manufacturer's instructions. The labelling volume was $500\mu l/10^8$ MNC in ACD-A buffer supplemented with Fc receptor-blocking reagent (100µl, 5min incubation, 4°C) before adding colloidal super-paramagnetic MACS MicroBeads conjugated to monoclonal mouse anti-human AC133/1 antibody (100µl IgG1 isotype, 25min incubation, 4°C). Cells were then washed (5ml ACD-A buffer, 400g, 10min, 4°C) and resuspended in 500µl ACD-A buffer. Labelled MNC fraction was applied to a chilled MACS positive selection column (MS+/RS+) attached to a cold magnet. The so-applied high gradient magnetic field permitted the $CD133^+$ cells to be magnetically retained onto the column (fig. 2.1). The column was simultaneously rinsed with cool ACD-A buffer (4x500µl) eluting the CD133-negative ($CD133^{Neg}$) cell population. After magnet removal $CD133^+$ cells were eluted with a plunger in 1ml of cold ACD-A buffer. The
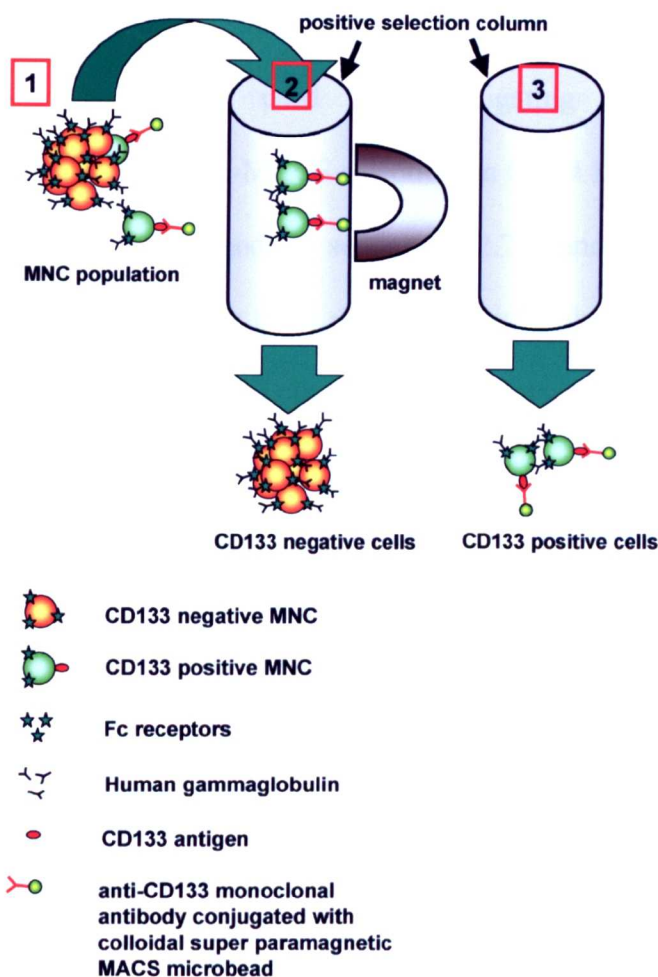
Figure 2.1: **CD133+ cell immunomagnetic positive selection using the mini-MACS kit.**

1 – The MNC fraction is labelled with human gammaglobulins solution to block non specific cell surface Fc receptors prior to incubation with magnetic bead-conjugated monoclonal mouse anti-human AC133/1 antibody, which directly labels CD133+ cells. 2 – Labelled cells are passed through a positive selection column submitted to a magnetic field. This magnetic field attracts the magnetic bead-conjugated antibodies and thus retained CD133+ cells onto the column whilst eluting CD133Neg MNC. 3 – Upon magnet removal and buffer washes, CD133+ cells are recovered. To enhance CD133+ purity, steps 2 and 3 were repeated using a new column.

CD133$^+$ cell fraction was reapplied to a fresh positive selection column and eluted as described above, prior to cell enumeration and viability assays.

In early experiments, CD34$^+$ cells were mini-MACS selected directly on the CD34 antigen using anti-CD34 MoAbs. This procedure forced the CD34

molecules to be artificially and predominantly concentrated on one side of the cell membrane, a case of artificial polarisation. Since the corresponding LSCM datasets could not be reliably used for investigating normal HSPCs in physiological conditions, the mini-MACS selection on CD34 LSCM datasets were used to measure polarity (defined in sections 5.2.2.1 and 5.2.2.2) on these artificially polarised cells in order to be compared with polarity measured under more physiological conditions for all other LSCM datasets.

2.1.4. Immunofluorescent labelling of CD34 and CD164 antigen on cord blood and bone marrow HSPC

In the studies reported in this thesis LSCM imaging was used for analysis of antigen distribution on haemopoietic stem and progenitor cell (HSPC) membrane. CB or BM MNC were adhered at room temperature on gold positive slides (BDH, UK) before incubation at 4°C with human gamma globulins (2%, Sigma-Aldrich) to block Fc receptors.

Adhered cells were then indirectly labelled, firstly using primary mouse anti-human CD164 (IgG3, 103B2/9E10 labelling class II epitope reported to be ubiquitously distributed on most primitive HSPC subset [Watt *et al*, 2000]) and/or CD34 (IgG1, HPCA-II labeling class III epitope, reported to be expressed on more primitive HSPC subset than class I/II counterparts [Krause *et al*, 1996; Steen *et al*, 1996] antibodies (BD-Pharmingen). Cells were also labelled against isotype-matched monoclonal antibody controls: mouse IgG1 pure & mouse IgG3 pure respectively (BD-Pharmingen).

After five washes in excess volume of staining buffer, cells were secondarily labelled (30min, 4°C) with FITC-conjugated goat anti-mouse IgG3 specific, F(ab')$_2$ fragments (Southern Biotechnology, USA) or tetramethyl rhodamine isothiocyanate (TRITC)-conjugated rabbit anti-mouse IgG F(ab')$_2$ fragments (DAKO, Sweden) against IgG1 primary antibodies. After five washes in staining buffer, labelled cells were fixed in a 3.5% paraformaldehyde solution (BDH) and carefully mounted in 25µl Vector-Shield anti-bleaching solution.

CD34 class I, II and III epitopes were labelled following similar protocol. Adhered cells were indirectly labelled for CD34, Epitope III, using mouse anti-CD34 (IgG$_1$, HPCA-II, PerCP) (Becton Dickinson Immunocytometry Systems. California. USA) Epitope II (IgG$_1$, QBEND-10), or Epitope I, (IgG$_1$, My10). Irrelevant matched isotype antibody controls were also undertaken (IgG$_1$). The secondary fluorescent conjugated antibody used was Alexa fluor 488 (Molecular Probes). Antibody incubations were carried out at 4°C for 30 minutes with slides washed in excess PBS, sodium azide 0.1% between incubations. Dual immunolabelling was carried out sequentially for expression of CD34 epitopes; Epitope III and 1 or Epitope II and I. The epitope proximal to the cell membrane was labelled first to minimise blocking of distal epitopes to the subsequent antibodies.

Coverslips were sealed onto the slide with acrylic varnish Mounted slides were then stored at 4°C prior to images acquisition with a Zeiss LSM.440 LSCM (previously described in section 1.2.1).

2.2. Preliminary analysis of LSCM images

The LSCM images analysed in the studies presented in this thesis are 1024x1024

pixels, which is the highest size and resolution available on the Zeiss LSM.440

confocal microscope used. Single or clustered Haemopoietic Stem/Progenitor

Cells (HSPCs) were imaged. They were either single or dual labelled for diverse

A)

LSCM micrograph taken in
the middle region of the cell
(negative image).

B)

Histogram of A.

C)

Mean intensity of representative
background areas in function of
the focal plane position along the
Z-axis.

Different symbols indicate different
cells
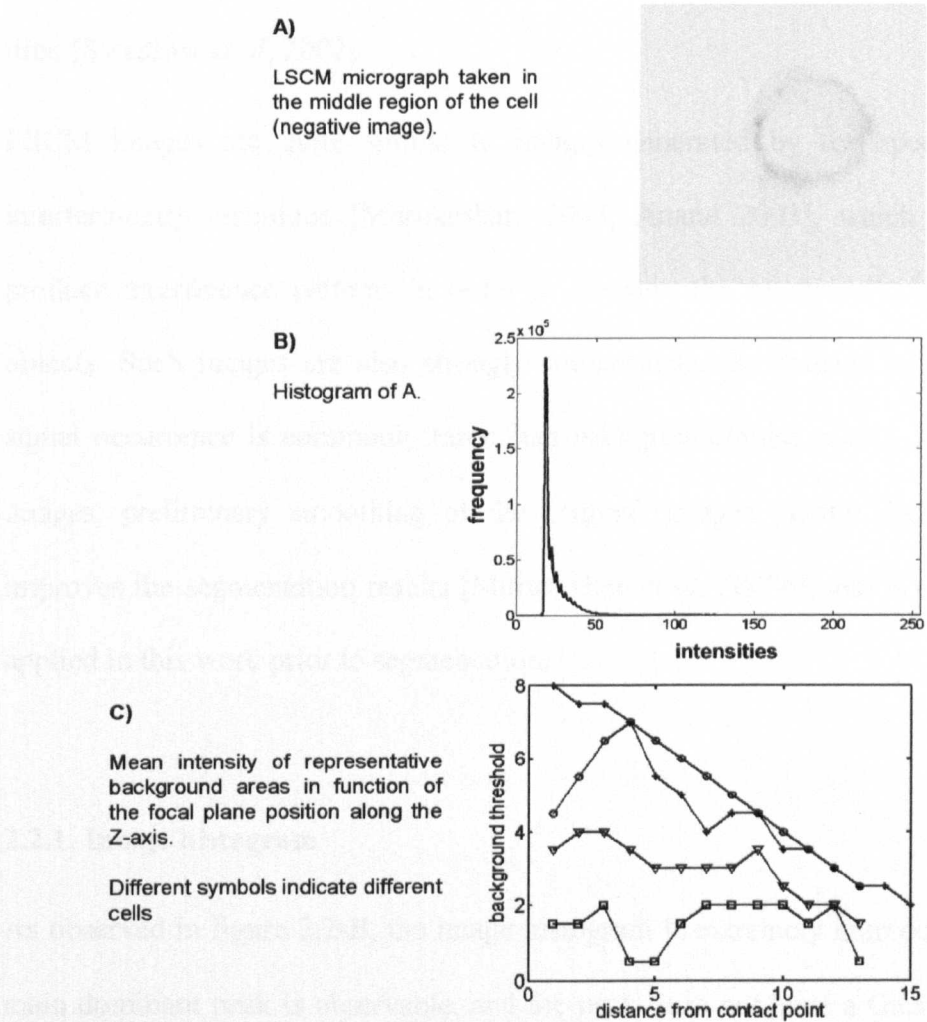
Figure 2.2: **Preliminary analysis of LSCM images.**

A – A typical LSCM image. B – Histogram of A. The strong unimodality of the
curve is obvious. C – Mean intensity of representative background areas vs. Z
position, for 5 different cells. This illustrates that the mean background value
fluctuates unpredictably with the z position. Consequently, a constant threshold to
segment signal regions from background cannot be applied systematically to the
whole Z-series, but must be evaluated separately for each LSCM image.

membrane antigens, or different forms (or epitopes) of the same molecule. The images are commonly contaminated by a strong noise component (fig. 2.2-A). Due to the nature of the distributions imaged, signal pixels are rarer than noise/background pixels. Furthermore, it was consistently observed that the average background values fluctuate unpredictably with the Z position (fig. 2.2-C), making the use of a constant threshold to segment the signal from the background impossible. Such fluctuations are mostly due to laser instability over time [Swedlow *et al*, 2002].

LSCM images are quite similar to images generated by the speckle fringe interferometry technique [Murukeshan, 2003; Anand 2003], which is used to produce interference patterns in order to measure the relative displacement of objects. Such images are also strongly contaminated by speckle noise and the signal occurrence is commonly rarer than background/noise pixels. As for such images, preliminary smoothing of the original images greatly facilitates and improves the segmentation results [Murukeshan *et al*, 2003b], and is consistently applied in this work prior to segmentation.

2.2.1. Image histogram

As observed in figure 2.2-B, the image histogram is extremely unimodal, i.e. one main dominant peak is observable, and the peak does not have a Gaussian shape but looks rather like a skewed Poisson distribution. This may be explained by the Poisson process involved during signal generation and brightness/contrast adjustments made during the set-up of the confocal microscope. This peak is located on the left of the histogram, at the lowest intensities, since it is due to the predominance of background noisy pixels in comparison to the number of

fluorescent pixels [Antoine *et al*, 2001]. It appears that this strong unimodal distribution prevents the use of popular segmentation algorithms based on histogram shape [Ammouche *et al*, 2001]. Indeed, these approaches make use of assumptions which are not suitable in the unimodal situation. The algorithm by [Rosin, 2001] addresses and proposes a solution to this histogram-based segmentation problem, but unfortunately the algorithm is heuristic in nature and consequently does not have any theoretical justifications.

2.2.2. Background and signal statistics

A preliminary analysis of the original LSCM images indicated that such images
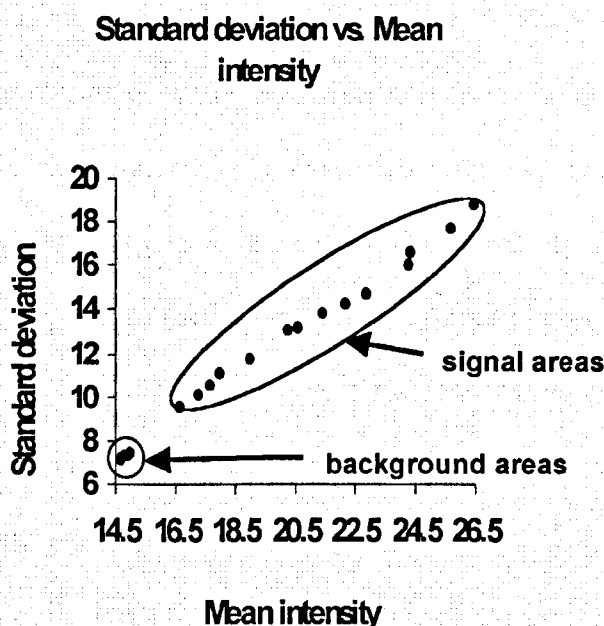
Figure 2.3: **Background and signal intensity preliminary statistics.**

The mean intensities measured in background and signal regions are plotted versus the respective standard deviation (SD) of the intensities in these regions, for all Z-scans of a LSCM Z-series. The mean intensities and SD are higher for signal regions than background. They also vary over a wider range of intensity than for background areas.

have a strong unimodal histogram. Figure 2.3 shows a dot plot of measurements taken from a confocal Z-series. Each image was manually thresholded to separate the background from the signal. However, because of the speckle noise affecting the images, the original micrographs were initially smoothed with a 10x10 averaging filter in order to reduce the speckle noise and make signal regions more homogeneous. The smoothed images were then manually thresholded using a global threshold. The threshold was stet up by the operator at a value subjectively thought to separate correctly the signal from the background. The so obtained binary masks were used to extract the actual background and signal areas in the original, non smoothed micrographs. The x-axis of figure 2.3 represents the mean intensity of both the segmented background and signal areas. The y-axis represents the standard deviation of the intensities in these areas. The points are globally aligned along a trend line. However, it can be observed that the points of lower values are clustered in the lower left corner of the graph. This cluster is constituted by the measurements obtained from the background areas. The set of points on the right of this cluster are the measurements obtained from the signal areas. Although they are aligned with the background cluster, they extend over a much wider range of values. This is due to the fact that signal intensities spread themselves over a broad range of intensities. As a consequence, the standard deviations in signal areas are also higher than in background areas. This observation will be the basis for an assumption used to develop the Between Class Variance Maximisation algorithm presented in section 3.4.

2.2.3. Comments on actual cell shapes
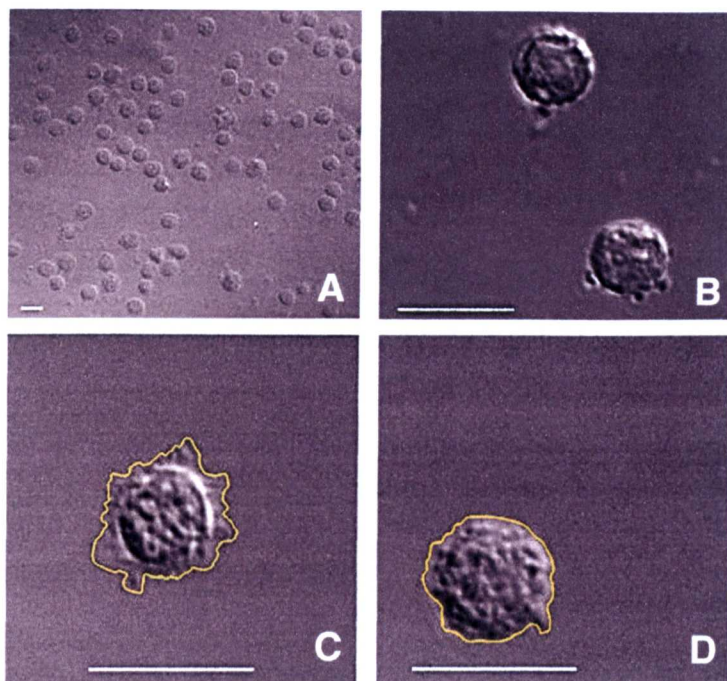
Figure 2.4: **HSPC morphology assessed by DIC images.**

A – At low magnification (x630), cord blood MNCs appear to have a spherical shape for the whole population. *B* – Identified cord blood HSPCs have similar spherical morphology, as confirmed at higher magnification (x630x6). However, some membrane protrusions are also clearly visible. *C,D* – Bone marrow progenitor cells also have a blast morphology, but they show increased membrane activity, as outlined by the yellow contours, in comparison to cord blood HSPCs. This may reflect their physiological interacting status in the bone marrow when compared to normally circulating cord blood progenitor cells. Bars are 10 µm long.

The actual cell shape will influence the way the antigen distributions will be analysed. However, none of the LSCM datasets used fluorescent dyes for staining the cell membrane (e.g. PKH26), and consequently the actual cell shape is not known. DIC images taken in the middle region of the cell (in the middle of the z-series, where it provides the neater image) helps to estimate the morphology and average diameter of the cell. On one hand, the visual examination of such DIC images (fig. 2.4) reveals that HSC have a typical blast morphology, i.e. small (approximately 10 µm in diameter) rounded, quasi-spherical cells when observed

in cross-section. On the other hand, DIC images do not provide information on the three-dimensional structure of the cells. 3D sectioning using DIC images has been attempted, but reconstructing the 3D shape of the cell using them was revealed impossible so far due to the strong blurring effects occurring in the images taken further away from the middle z-sections. From the LSCM image datasets used in this work, the fluorescence signal from the cell membranes was used to get some preliminary information on the actual cell shape (algorithm in Appendix C). Basically, the fluorescent clusters and halos are used to estimate the mean radius of the cell at each z-level. It appears that HSPC are either rather spherical or more cylinder-like shaped (figure 2.5). A flat basis is often observed at the contact point of the cell with the slide, probably a consequence of gravity and an increased adhesion area triggered by the cell. As a result of this observation, a general geometry for describing the HSPCs has been adopted and used all the way through the present thesis. This geometrical system is presented in figure 2.6 along with the terminology used throughout the thesis.

Estimated cell diameter vs. z-level

Estimated cell diameter vs. z-level
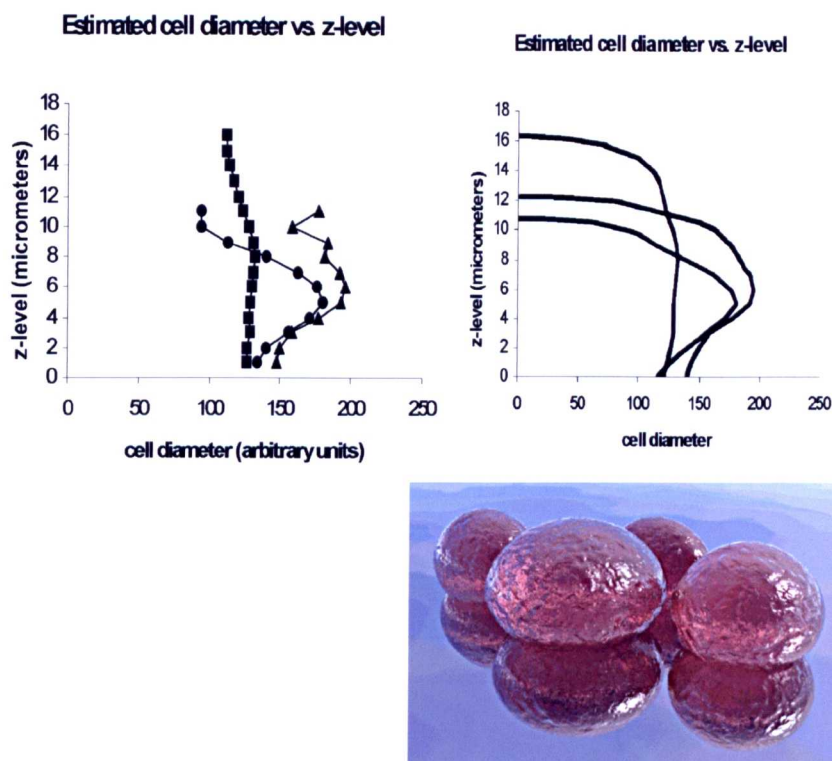
Figure 2.5: **HSPC morphology assessed by LSCM images.**

Top left – Cell diameter (versus Z position) of three cells estimated using the fluorescence of the membrane antigen. In all cases, the cell diameters are reasonably circular. The diameter is automatically estimated as the median distance of the fluorescent pixels from the cell centroid (program given in appendix). The cells used were chosen to have a rather uniform membrane antigen distribution. The profiles of the cells can be visualised and clearly indicate that the cells are not truly spherical. *Top right* – Subjective interpretation of the actual cell profiles by interpolation of the previous measurements. The interpolation is particularly important at the top of the cells, whose position is estimated as the Z-position where no fluorescence is recorded. *Bottom* – Artist view of cell shapes. The cells are pictured as flattened sphere, which is more in agreement with the actual measurements than with a perfect theoretical sphere.
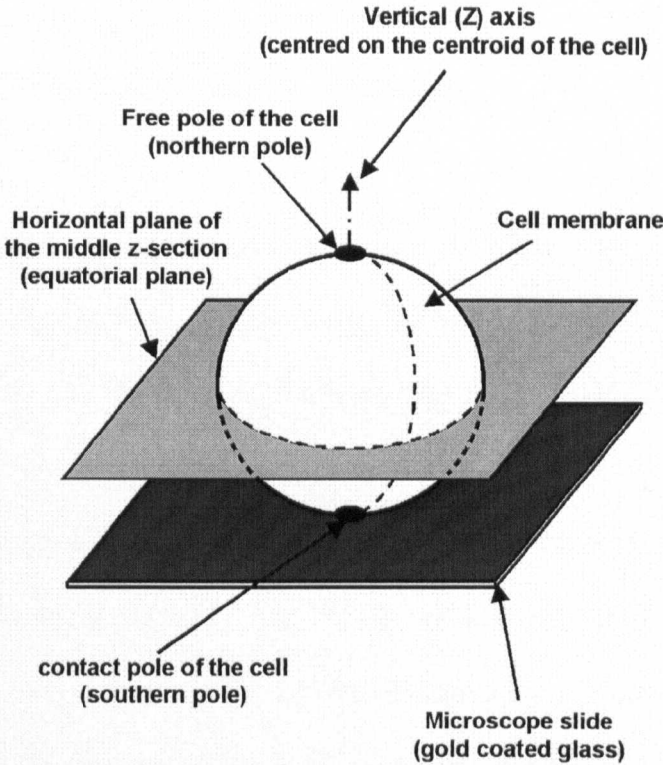
Figure 2.6: **Theoretical geometry of the system and terminology.**

Idealised representation of a theoretical cell and some of the terminology used throughout this thesis. The cell is assumed to be a perfect 3D sphere adhering to the microscope slide at its contact point. It is divided in two hemispheres, northern and southern, by the equatorial plane. The cell is centred to the vertical axis, so it is possible to express any location in the membrane using either Cartesian or polar coordinates.

Chapter 3

Algorithms for automated fluorescence signal segmentation in LSCM images

Chapter 3. Algorithms for automated fluorescence signal segmentation in LSCM images

In the following sections, three new automated global thresholding algorithms are presented. The first one is based on a multiscale analysis of scale image histograms, the second is based on histogram comparison of a rescaled image with a background area, and the third one is based on the maximisation of a between-class criterion. A fourth region-based segmentation algorithm is also presented in order to demonstrate the feasibility of such an approach. Due to its innovative nature and the potentially wider range of applications of the first global thresholding algorithm, the MultiScale Approach (MSA), the phenomena investigated and used for its design are described and commented in detail first, prior to the description of the MSA algorithm itself, followed by the shorter descriptions of the other algorithms. All algorithms presented in this thesis were implemented using Matlab 6.1 (the MathWorks Inc., Natick, Massachusetts, USA) and were not compiled.

3.1. Preliminary qualitative analysis of scale image histograms

Generating scale images in order to perform segmentation is computationally expensive. For instance, in a Gaussian scale-space, the scale is represented by the standard deviation σ of the Gaussian filter used to blur the original image. There is a continuum of values for σ, from extremely small to extremely high. In computational applications, such continuum of values can be approximated with discrete values. Once the (linear or discrete) scale image is constructed, segmentation approaches make use of the analysis of the paths of critical points (such as saddle points) through the scales, or regions bearing certain properties

(e.g. high gradient for edges, homogeneity for smooth surfaces), and then back tracking (through scale space) such features up to the original scale where segmentation is performed. These steps greatly add to the computational load of the segmentation process. These make scale spaces, despite being attractive and theoretically powerful methods, expensive to implement in practice, especially for large images. They are rarely used in real applications.

How could segmentation be performed on LSCM images using classical scale spaces (i.e. region-based approaches)? Since the features to be segmented are fluorescent clusters, or "blobs", it could be imagined to back track the blobs from coarse to fine scales by identifying local maxima in the scale image. Similar back tracking procedure is common for segmenting image regions using scale spaces. Theoretically, the paths of such maxima would have a tree-like nature, the stem being located at the coarser scale while the number of branches would increase at finer scales. At the finest scales, a problem would emerge: there would be as many branches as the number of local maxima, those from fluorescent blobs and those from local background fluctuations. Due to the unimodal nature of the image histogram, deciding which intensity threshold separates fluorescent from background blobs is not straightforward (this is actually a problem similar to defining a threshold from a unimodal intensity histogram, the problem addressed in this thesis). Consequently using scale image for LSCM image segmentation does not appear to be a computationally efficient approach which can be easily and robustly automated.

It is herein investigated the intensity histograms of scale images instead, for reasons partially described in this section and further detailed with the description of the actual algorithm in section 3.2[1].

As a first introductory graphical example, a 1024x1024 artificial test image was produced which consists of two disks of different radii and intensities contaminated with 80% random noise smoothed with a 5x5 Gaussian filter (figure 3.1). A portion of the scale image is constructed by convolving this original image with an 11x11 Gaussian kernel with $\sigma=2$ applied iteratively over 800 iterations. It can be noticed that the disks still appear distinctly despite being blurred after 800 iterations. At each iteration, the intensity histogram is computed and stored. The histograms along scales (iterations) are presented colour-coded in figure 3.1. Three main lines (i.e. the histogram peaks corresponding to both disks and the background) are observed. The middle line corresponds to the smaller disk, while the left line is from the background and the right line from the largest disk.

The positions of both the left and right lines do not vary much along the scales. However, the middle one slightly (but clearly) deviates progressively towards the left, lower intensities. This is a consequence of the diffusion process of the small disk's intensities towards background intensities as the degree of blurring is increased. It can also be observed that the spaces between the three lines contain no or very small values (sky blue colour) at low scales, but become progressively more filled with (in order of both appearance and increasing magnitude) cyan, green and yellow colours. This phenomenon is particularly

---

[1] The reason for splitting the explanations is that the qualitative analysis carried out in the present section is used to justify the design of the MSA algorithm described later, but some of these preliminary explanations make use of objects computed with the MSA algorithm (i.e. the matrix R from equation 3.10).

distinct between the left and middle lines. For both these lines, the interpretation is that, as blurring increases, more pixels from the smallest disk and the background are "mixed" together and their intensities weighted according to the local convolution product with the iteratively applied Gaussian kernel. Therefore, as scales increase (i.e. more iterations are performed), more pixels are involved in this process at the boundaries of the disk and its surrounding background, leading to the diffusion process observed on the histogram along the scales.
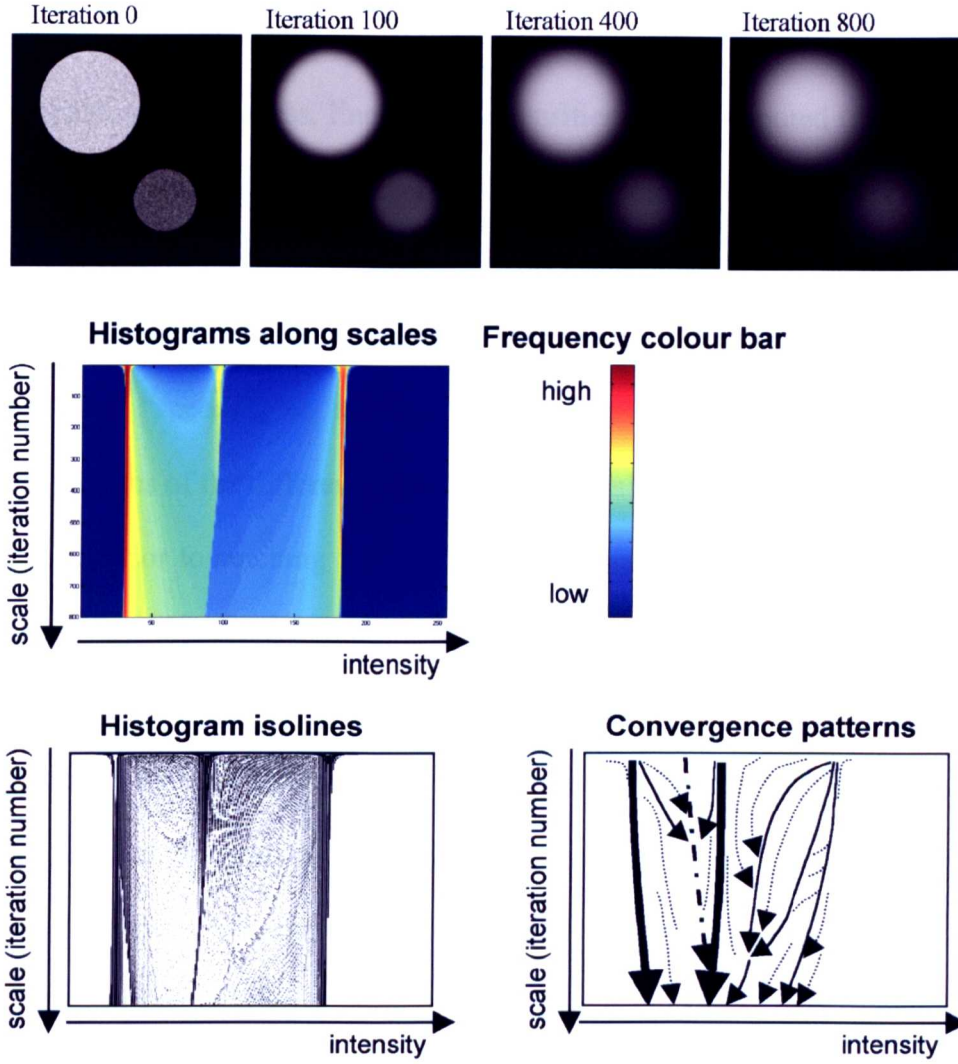
Figure 3.1: **Intensity diffusion in scale space. Qualitative analysis of histograms.**

Top – Original test image used to produce this figure and its smoothed versions after iterative convolution with an 11x11 gaussian filter (σ=2). *Middle* – Histograms along scales. Three main lines/peaks are observed which correspond to both disks and the background in the original and scale images. *Bottom* – Histogram isolines (*left*) and graphical interpretation of the convergence patterns observed with the isolines (*right*). As diffusion increases with the scale, the relative frequencies of the intensities in the histograms are modified and convergence patterns are observed. It can be noticed that some intensities converge faster (i.e. over less scales) than others. Convergence rates are affected by the initial relative frequencies and the spatial organisation of the pixels.

Such diffusion manifests itself as an increased frequency of pixels with intensities close to those of the disks. As blurring increases, diffusion increases:

there is a transfer of intensities between main histogram peaks. This transfer works both ways, from one peak to the other and *vice-versa*. In order to further emphasise this process, lines of iso-intensities were computed (figure 3.1) from the histogram along scales. This figure illustrates the complex patterns of intensity transfer/diffusion along the scales. Using these isolines, a schematic of the dynamical patterns was drawn. It is observed that isolines converge with different "speeds", or rates, at various scales, depending on the relative magnitudes of the different peaks and the spatial relationships of the pixels in the image, prior to reaching a slower, more uniform progress along the scales. The frequency of particular intensities at and close to the main peaks is also depicted (figure 3.2, A, B, C, D). Generally, it is observed that the frequency increases markedly over the first scales prior to decreasing at various rates. Eventually the frequencies increase again steadily with much lower rates. The graph E shows the frequencies measured half-way between both peaks, which globally increase over scales with different rates. Such increase is not steady over small local scale ranges (it fluctuates), a consequence of the complex dynamics occurring during the diffusion process. These figures illustrate the effects on the image histogram of the diffusion process as scales are decreased.

It is worth noticing that generating such figures from an initial 1024x1024 image is very computationally expensive. In definitive, after 800 iterations, little seems exploitable from these histograms.
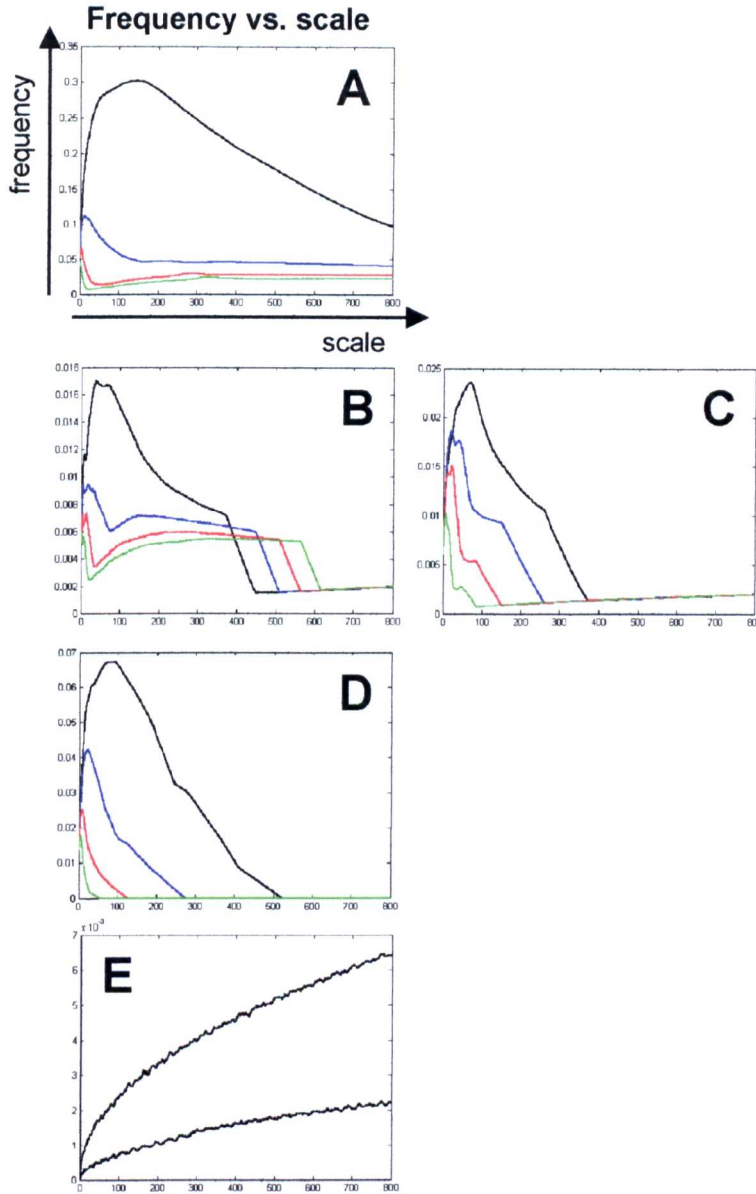
Figure 3.2: **Frequency vs. scales for particular intensities of the scale image histograms.**

In all cases, the **black** lines indicate values measured at the original peak locations, the **blue** lines at one intensity unit from the peaks, **red** lines at two intensity units from peaks, and **green** lines at three intensity units from peaks. A – Frequencies measured at and on the right of the first histogram peak. B – Frequencies measured at and on the left of the second histogram peak. C – Frequencies measured on the right of the second histogram peak. D – Frequencies measured at and on the left of the third histogram peak. These graphs $(A - D)$ reflect different evolutions of the frequencies around the histogram peaks which are due to the diffusion process taking place down the scales of the scale image. E – Frequencies measured half way between the first and second histogram peak (top curve) and between second and third peak (bottom curve). See text for comments.

Figure 3.3 shows another example of the effect of diffusion on histograms. Instead of using a 2D image, an arbitrary 1D signal was used for speeding up the diffusion process. Both histograms B and C were produced as described above, using iterative convolution with a Gaussian filter of fixed standard deviation. Again, convergence of the intensities through diffusion is reflected in the histograms by having convergence and merging of peaks. Smaller peaks converge faster while it takes relatively much longer for the biggest ones. Global convergence of all peaks happens within less scales if the standard deviation is increased, but at the cost of losing the resolution of convergence process at finest scales. Alternatively, increasing σ at each iteration allows to visualise the convergence details at all scales and to speed up the process at high scales. The choice of how σ should be modified through the iterations is rather subjective, depending on the performance of the processor used and the level of details needed.
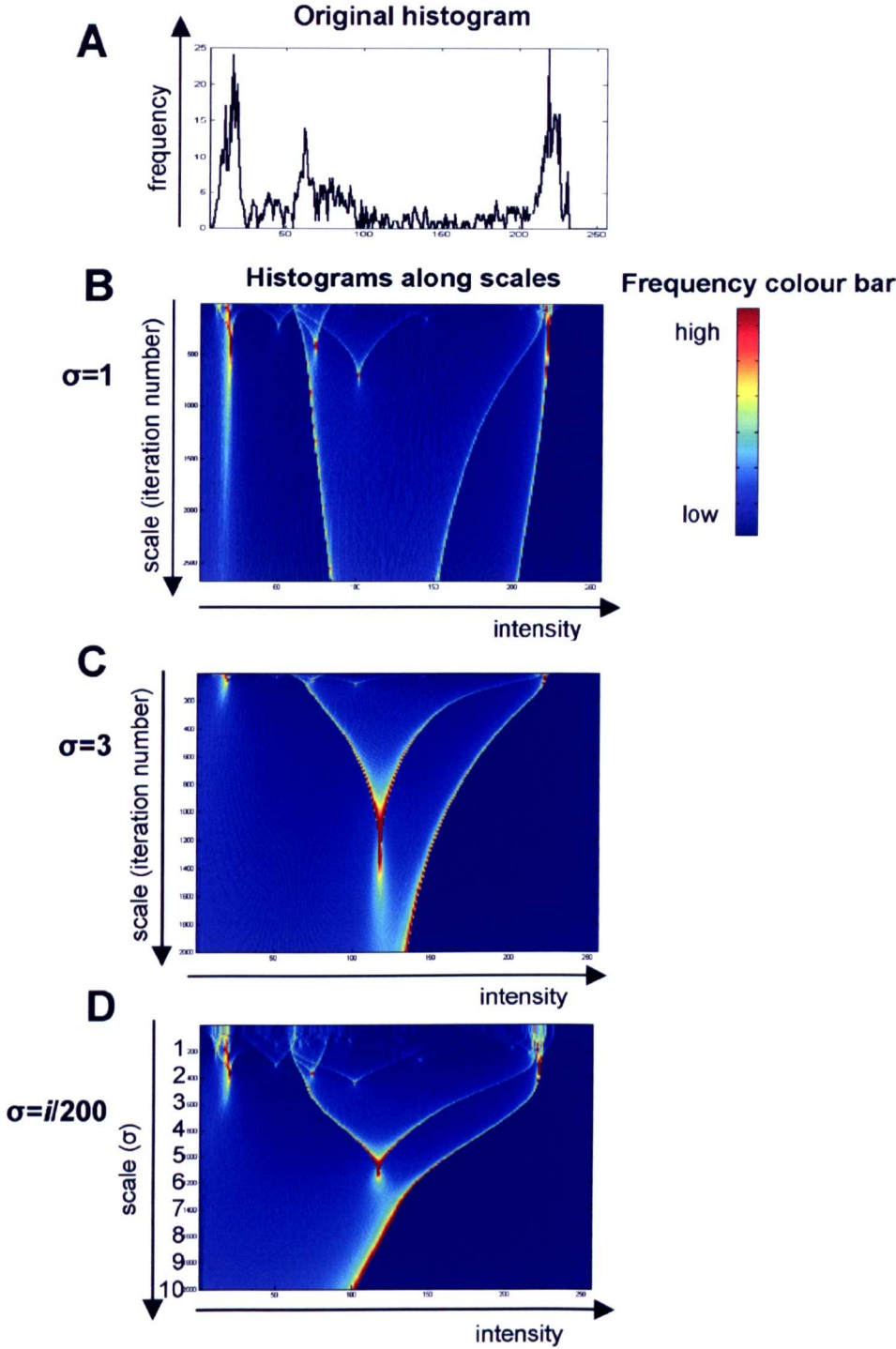
Figure 3.3: Frequency vs. scales using different scale parameters.

A – Original histogram of the 1D signal used to compute these histograms in scale spaces. *B, C* – Histogram of scale signal generated by iterative convolution with Gaussian filter of fixed SD (indicated on the left of the figures). A higher SD speeds the intensity diffusion, as reflected by faster convergence of the histogram peaks, but at the cost of losing details observed at finer scales. *D* – Histograms obtained by varying the SD as a function of the iteration number *i*. Details are observable at both fine and coarse scales.

3.1.1. Convergence rate of histogram frequencies along scales

It was previously mentioned that the rate of convergence of peaks along the scales depends on the relative sizes of these peaks and spatial pixel distributions, and they carry some information about the intensity distribution of the image pixels. In the MultiScale Analysis (MSA algorithm) described below, a measure of the diffusion rates of the pixel intensities is performed (detailed in section 3.2) from the histograms. Such measure has high values when, for a given intensity, the frequency changes significantly between two consecutive scales, and low in the opposite case. When there is no frequency change between scales (the frequency is either zero or a positive constant), the diffusion rate is zero.

Figure 3.5 shows both histograms and corresponding diffusion rates for different images with strongly unimodal histograms (original LSCM image – fig. 3.4-*top*, smoothed LSCM image – not shown – , pure smoothed fluorescent signal – fig. 3.4-*bottom*, pure original background – fig. 3.4-*middle* – and smoothed background – not shown). Histograms and diffusion rates were computed from approximated scale image until complete diffusion was achieved (which theoretically takes infinity steps with Gaussian scale spaces). The schematics on the right column symbolise the features of interest observed on the diffusion rate figures. Lines are similar to the arrows used in figure 3.1, indicating the dominant diffusion patterns, or "diffusion lines", applied this time to the diffusion rates instead of the histograms. The red points indicate the locations where main diffusion rates disappear. The green areas show the regions were the positive histogram frequencies vary little or not over scales, and the corresponding diffusion rates consequently equal zero. The vertical dashed lines indicate the mean image intensity.

Original LSCM image

Pure noisy background

Pure smoothed signal
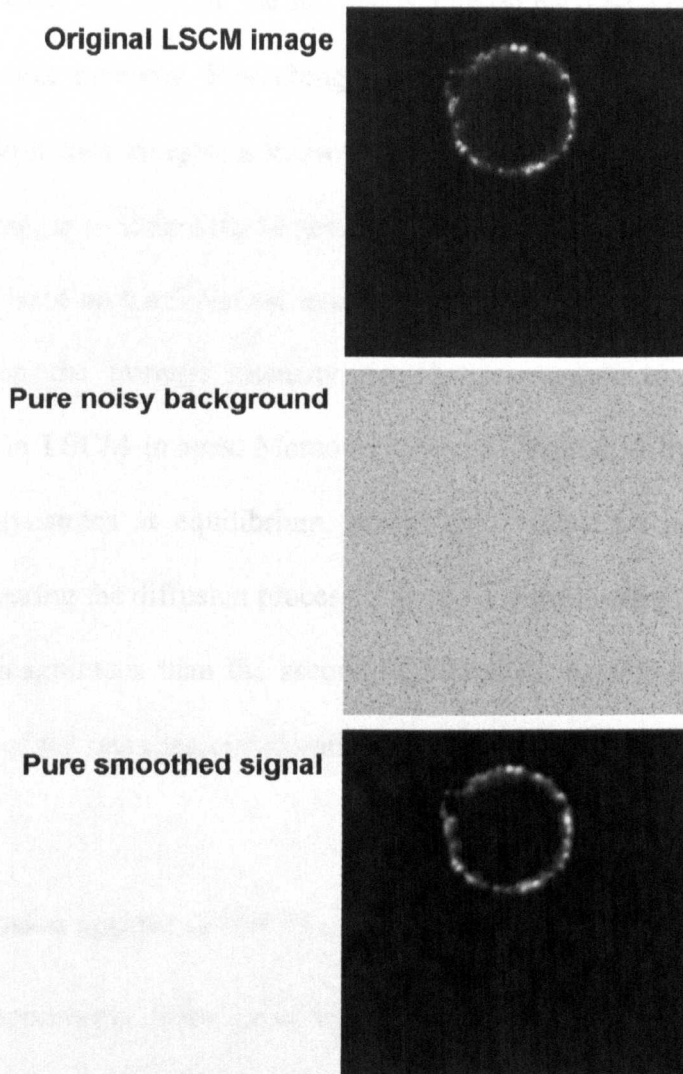
Figure 3.4: **Images used for qualitative multiscale analysis of histograms.**

Top – Original LSCM image (noisy background + noisy signal). *Middle* – Pure (negative) noisy background with same PDF as original noise background in LSCM images. *Bottom* – Pure fluorescence signal smoothed with a 10x10 moving averaging filter.

3.1.2. Diffusion applied to LSCM image noise background

For an image containing only background noise (distributed as in an original LSCM image) (fig. 3.4-*middle*), half way along the scales, the diffusion rates disappear as diffusion has reached an equilibrium (fig. 3.5-second row). Two

"tails" (black lines) indicate the rate paths, one on each side of, and converging to, the average intensity. Smoothing LSCM images, or similar type of images (e.g. interferometry images) is known to improve segmentation. The effect of an initial averaging filtering (10x10 kernel) is depicted in figure 3.5-third row. Two tails are present on the diffusion rates, but (importantly) they are located on the left side on the average intensity, which corresponds to background-only intensities in LSCM images. Moreover, they are separated by a green area of diffusion dynamics at equilibrium, scales over which no important changes happened during the diffusion process. The tail at finest scales (top tail) has rates of higher magnitudes than the second tail. Overall, smoothing did reduce the magnitude of the rates associated with background intensity diffusion.

3.1.3. Diffusion applied to LSCM image signal

Similar experiments were repeated on pure fluorescence signal (similarly smoothed) segmented manually to remove background areas (so the noise does not interfere with the process) (fig. 3.4-*bottom*). A more complex pattern of rate paths emerged (fig. 3.5-top row). Again, an equilibrium phase is achieved along some scales which extends over approximately 5 scales. Leaving this equilibrium phase, diffusion rates took positive values again, with two main tails situated on both sides of the average intensity. The longest tail is situated on the right side of the average intensity, which corresponds to the region of signal pixels in LSCM images. Most tails (black lines) are actually located to the right of the average image intensity. The rate's magnitudes are higher at coarser scales (i.e. after the equilibrium phase) than finest ones, the opposite of what was observed with the pure smoothed background image.
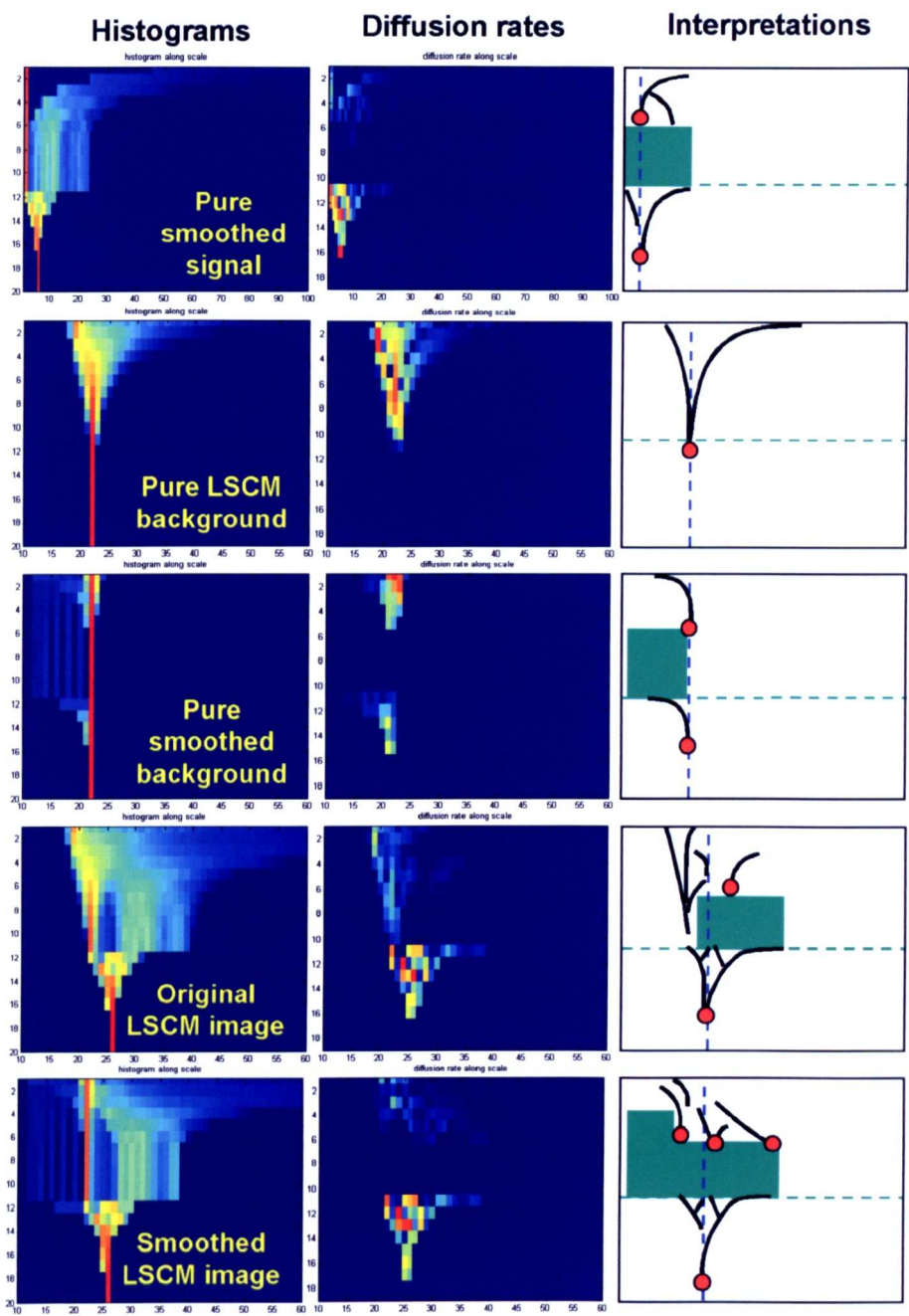
Figure 3.5: **Qualitative analysis of histograms and diffusion rates along scales.**

Scale space versions of the LSCM image components displayed in figure 3.4 were generated and their histograms computed along scales (left column, similar to figure 3.1). These histograms were further processed to measure the diffusion rates between consecutive scales (middle column), and a qualitative interpretation of both histogram and rate matrices are given (right column). Refer to text for details.

Therefore it appears that diffusion rates of signal and background images have qualitatively different behaviours over scales. Tails for background images are located on the left of the average image intensity while those from signals are on the right. Higher rate magnitudes are located at finest scales for background images while they are observed at the coarsest scales for signal images.

3.1.4. Diffusion applied to LSCM images (background noise + fluorescent signal)

Similar experiments were performed on original LSCM images and smoothed images. In both cases (figure 3.5, fourth and fifth rows), complex patterns of rate paths are observed. The tails of highest magnitudes are still those generated by the presence of the signal component, in coarser scale regions and have quite similar patterns for both smoothed and unsmoothed images.

However, for an unsmoothed image (fig. 3.4-fourth row), the equilibrium phase takes place mostly on the right of the average image intensity, were signal pixel intensities are expected. For a smoothed image (fig. 3.4-fifth row), it also takes place on the left of the average intensity, corresponding to background intensities. In comparison to tails on the right of the average intensity, the left tails extend over less scales. The consequence is that smoothing reduces the magnitude of diffusion rates along scales for background intensities and the number of scales over which diffusion is greater than zero. Furthermore, highest rate magnitudes corresponding to noise diffusion are observed closer to the average intensity in smoothed images than in unsmoothed images. Also, while

these rates have lower magnitudes in background regions of the figure, they are more present on the right side of the average intensity. Smoothing increased the effect of the "interactions" between signal and noise pixels through diffusion by shifting the diffusion rates more to the right of the average intensity (the tails).

From these observations, it appears that analysing diffusion rates instead of the histograms provides with more information about the dynamics of these histograms along scales. They reveal patterns which are not obvious from the histogram figures. Rates also identify regions of the histograms where background and signal can be characterised for developing a segmentation approach. Such an approach is provided below, where details on how the histogram and rate matrices were constructed are described. In addition, this preliminary analysis showed the importance of smoothing the LSCM images prior to applying the diffusion process and its influence on the behaviour of the rates along scales.

3.2. MultiScale Approach (MSA)

Unimodal histograms observed on LSCM images are the consequence of the noisy background which predominates over small but significant signal areas with lower occurrence of signal pixels. Furthermore, the pixels in original signal areas share a wide range of intensities with background pixels, which complicates the selection of an appropriate global threshold which would well separate noise and signal. This new algorithm makes use of a property of scale space images or image pyramids, i.e. the convergence to the mean image intensity as scale increases. However, an important and novel aspect is that it

works from the information contained in respective "scale histograms", thus it remains a histogram based approach.

It is hypothesised that the intensity of signal pixels would converge to the mean image intensity in a different way than background pixels, due to the unbalanced relative frequencies of both these populations. It is expected that, as scale increases, the intensity of the highest value pixels (signal) would diffuse quicker towards the average image intensity than background pixels, due to the relative rarity of such signal pixels. As a consequence, it is expected that the frequency stored in the bins corresponding to signal pixels in the scale histograms would decrease quickly and converge to zeros while the signal disappears. Thus, the analysis of the diffusion rate of each histogram bin along the scales should behave differently (i.e. have different dynamics) along the scale, somehow revealing the intensity values respectively corresponding to background and signal regions, and the decision of an appropriate threshold would be derived from this analysis.[2]

Such a histogram-based approach can be computationally very efficient, as the important convergence rates are solely calculated from first order statistics, if the image rescaling process is computed efficiently. In order to speed up the convergence process to reach the average image intensity, the re-scaling process of the original image was implemented so that the total number of pixels was reduced by half at each scale. The original 1024x1024 LSCM image was first

---

[2] The term "diffusion rate" is preferred to retain the idea of the diffusion process despite the fact that the term "convergence" is used to describe the process by which the frequency per histogram bin, over scales, approaches – converge to – the value zero. In addition during this process, the original histogram peaks convergent hierarchically and merge as the scale decreases, eventually converging to the mean image histogram (figure 3.5).

smoothed with a 10x10 averaging filter (the benefit of this operation were qualitatively outlined in section 3.1), and vectorised as a vector C_1 so

$$C_1 = \{C(1), C(2), ..., C(n)\} \qquad n = 1024^2.$$

3.1

The first 1024 elements of this vector are constituted by the first column of the image, the next 1024 elements are from the second column, and so on. A normalised histogram of this vector C_1 was computed as

$$H_C^1 \qquad \text{with} \qquad \sum_{i=0}^{255} H_C^1(i) = 1.$$

3.2

which corresponds to the histogram of the original image after the initial smoothing step. The re-scaling of vector C was implemented by re-organising the 1×1024^2 elements of C into a $2 \times [(1024^2)/2]$ matrix C', so

$$C_1' = \begin{bmatrix} C(1) & C(3) & ... & C(n-1) \\ C(2) & C(4) & ... & C(n) \end{bmatrix}$$

3.3

and averaging down the columns to produce a new vector C_2 containing $1 \times (n/2)$ elements. By definition, this new vector contains half the number of original pixels, and intensity interpolation (averaging) is used to calculate the values of the new elements.

As for the original vector image, the normalised histogram of C_2 was computed as

$$H_C^2 \qquad \text{with} \qquad \sum_{i=0}^{255} H_C^2(i) = 1.$$

3.4

This process is iterated, so C_2 is reorganised according to equation 3.3, then averaged along the $n/4$ columns, H_C^3 is computed, and so on.

This averaging process is related to mean intensity interpolation used in some pyramid images. It can be seen that, despite vectorisation of the initial image, the values in C_2 are mean intensity interpolations of spatially adjacent pixels grouped in pairs. The spatial relationship of pixels holds during the rescaling process, as it does for scale space images or pyramids, except that it is not based on a similar local basis (i.e. the averaging is not performed in $n \times n$ neighbourhoods). Indeed, the elements of C_2 are the average values of the pixel pairs

$$C_2 = \{(1,2) \quad (3,4) \quad ... \quad (n-1,n)\} \qquad 3.5$$

where the brackets contain the indices of the elements from C_1 used for the averaging. Iterating the process leads to

$$C_3 = \{(1,2,3,4) \quad (5,6,7,8) \quad ... \quad (n-3,n-2,n-1,n)\}, \qquad 3.6$$

$$C_4 = \{(1,2,...,8) \quad (9,10,...,16) \quad ... \quad (n-7,n-6,...,n)\}, \qquad 3.7$$

$$C_5 = \{(1,2,...,16) \quad (17,18,...,32) \quad ... \quad (n-15,n-14,...,n)\} \qquad 3.8$$

and so on. After $\log(1024)/\log(2)=10$ steps, C_{10} contains the 1024 average intensities calculated down each of the 1024 columns of the initial smoothed square LSCM image. Following the same construction, C_{11} contains the average intensities of $1024/2=512$ pairs of columns from the square LSCM image grouped such as the first element is the average of the first two columns of the image, the second element the average of the next two columns, and so on.

This dynamics explains the structures observed in figure 3.5 depicting the histograms H_C^i computed along the scales i. The whole matrix containing these histograms is referred to as H. At the finest scales (first iterations of the rescaling

process), the intensities of the rarer signal pixels are diffused rapidly into the background. This is because average intensities stored in the C_i are computed from non-overlapping image segments (aligned in the 2D image from top to bottom and left to right) of exponentially increasing sizes. It explains the initial fast convergence in the direction of the average intensity value observed in H, on the right of the average intensity. As scales decrease, exponentially more aligned adjacent pixels are involved in the averaging locally computed in the columns of the square image as shown by the equations 3.6, 3.7 and 3.8 above, so the impact of signal pixels whose intensities have already diffused significantly have an exponentially lower impact on the averages. A similar effect would have been observed if the averaging had been performed directly on the 2D image in spatial sub-regions of exponentially increasing sizes. This explains why an equilibrium state is achieved in H after a few scales (typically 5). Indeed, during these 10 first scales, convergence to a mean intensity per segments of the 2D image is achieved quickly (5 steps), so the histograms computed from these mean intensities are rapidly found to vary little over scales. This is the equilibrium. This also explains why the equilibrium lasts until the 10th or 11th scale, since the 10th scale corresponds to the scale where each single column of the 2D image has been fully averaged. At the 11th and above scales, the columns in the 2D images start to be averaged by pairs, generating new histogram dynamics as new averages of now stable averages are computed. Again, in this process, the column averaging involves an exponentially increasing number of columns. This leads to the forced convergence of mean intensities to the global image average value at the last scale.

Through the whole process, only adjacent pixels are involved, i.e. no pixels from separate image regions are simultaneously averaged. This is important as it means that the features measured here and computed at one scale, the averages, are directly related to those computed at the previous and finer scales, while spatial pixel relationships are maintained. This property meets the requirement stated by the causality principle of scale space theories, which makes the proposed method a scale space approach (section 1.3.3.3).

The re-scaling of the $1 \times M^2$ vector initially obtained from the $M \times M$ smoothed original image takes

$$k = \frac{2 \log(M)}{\log(2)} \hspace{3cm} 3.9$$

steps. This approach is computationally very efficient as it works in $O(2\log(M)/\log(2))$. Applied to 1024x1024 LSCM images, the whole re-scaling process takes $k = 20$ steps, each one being twice faster than the previous one, after which the image is reduced to a single pixel whose intensity equals the average original image intensity (similarly to the single pixel at the top of an image pyramid constructed using mean intensity interpolation). Therefore the requirement of designing a fast scale space histogram-based method is met.

At the end of this procedure, H is a 20x256 elements matrix. H contains the normalised histogram information (horizontally) of the original image observed at decreasing scales (vertically), eventually converging to the mean intensity (fig. 3.6-A). The first row contains the values of the original normalised image histogram, and the last row contains only zero-valued elements, except for one element which represents the intensity value at equilibrium of the diffusion process.

As qualitatively investigated in section 3.1, the rate of diffusion is influenced by the relative occurrence of pixel intensities and by the local density in their neighbourhood. For instance, if there is initially a very low proportion (a small population) of pixels with intensity j in the image, more pixels with different intensities are present in their neighbourhoods. Consequently, through averaging along scales, the intensity of this population quickly diffuses into the intensity of the surrounding pixels, which makes $H_{i,j}$ quickly converge to 0 as i increases (i and j respectively being the index of the row – scale – and the column – intensity – in matrix H). Conversely, if $H_{i,j}$ has an initial high frequency, the diffusion to 0 is slower.

The diffusion rates were calculated from the histograms in H by computing a rate function R such as

$$R^{i-1} = \left(\frac{1+H_c^i}{1+H_c^{i-1}} - 1 \right)^2$$

3.10

$$i \in \{2, \quad 3, \quad ..., \quad k\}$$

where H_c^i and H_c^{i-1} represent the normalised histograms at scales i and $i-1$ respectively. Consequently, if $H_{i,j} = 0$ and $H_{i+1,j} = 0$, then $R_{i,j} = 0$ accordingly to equation 3.10. A diffusion rate of zero indicates no further diffusion at a particular intensity value j, at a particular scale i. As uniform areas generate increasingly predominant peaks in the vectors ordered along the y-axis of R, the apparition of only one major peak (close to the mean grey intensity level in the smoothed image) is expected as a consequence of the histogram unimodality (fig. 3.6-B). This peak is generated due to the dominance of the uniform noisy

background in LSCM images and its position in R is affected by the presence of signal pixels (section 3.1).

The next step consists of using the information stored in R to define the appropriate threshold used to binarise the smoothed original image. It is convenient to reduce the 2-dimensionality of R to a one dimensional curve from which the threshold would be selected. As the one dimension to be retained in the process is the intensity axis, the x-axis of R, the scales, the reduction has to be performed down the columns of R.

In section 3.1, the properties of diffusion rates were described. In summary, most of the matrix R contains zero-valued elements, except in regions of dynamical activity along scales. Regions of R corresponding to background intensities (on the left of the mean intensity value) display tail-like structures which have lower magnitudes and lesser extent over the finest scales in comparison to those observed in signal regions. For these background intensity regions, rates are situated closer to the mean image intensity. Most rates are found in regions around the average intensity and on its right, which predominantly corresponds to signal regions. Therefore most activity is observed at the intensities separating background and signal, where an appropriate threshold should be set up. It is proposed to use the information in R by calculating the median values along the columns to produce the 1x256 vector

$$R_{med} = \text{median}(R) \qquad\qquad 3.11$$

which allows the identification of the predominant diffusion rate for each particular grey value of the histogram in the scale space (down the columns). However, other measurements may also be used. This rank filter is commonly

used to identify the predominant value of a series, which is what is aimed to be done here down the columns of R as more diffusion rates are observed for the intensities separating background and signal pixels.

The curve R_{med} has specific properties due to the unimodal nature of the original image histogram (fig. 3.6-C). The relative rarity of bright significant signal pixels causes these pixels to initially diffuse quickly to lower intensities. Therefore, after a relatively short number of re-scaling steps, the rows of R corresponding to most signal intensities contain predominantly zero-valued elements. Thus, $R_{med} = 0$ for these intensities. After these first re-scaling steps, the scale image contains values closer to the mean intensity and the histograms reach an equilibrium dynamics, for which $R_{med} = 0$. The initial smoothing process makes the lower intensities from the background to quickly diffuse towards the mean image intensity and the equilibrium state is reached earlier in these regions than for the signal, and it lasts as long as for the signal. Moreover, at these first scales, the presence of signal pixels "forces" the diffusion rates to be shifted towards signal regions of the matrix R, for intensities higher than the mean image value.
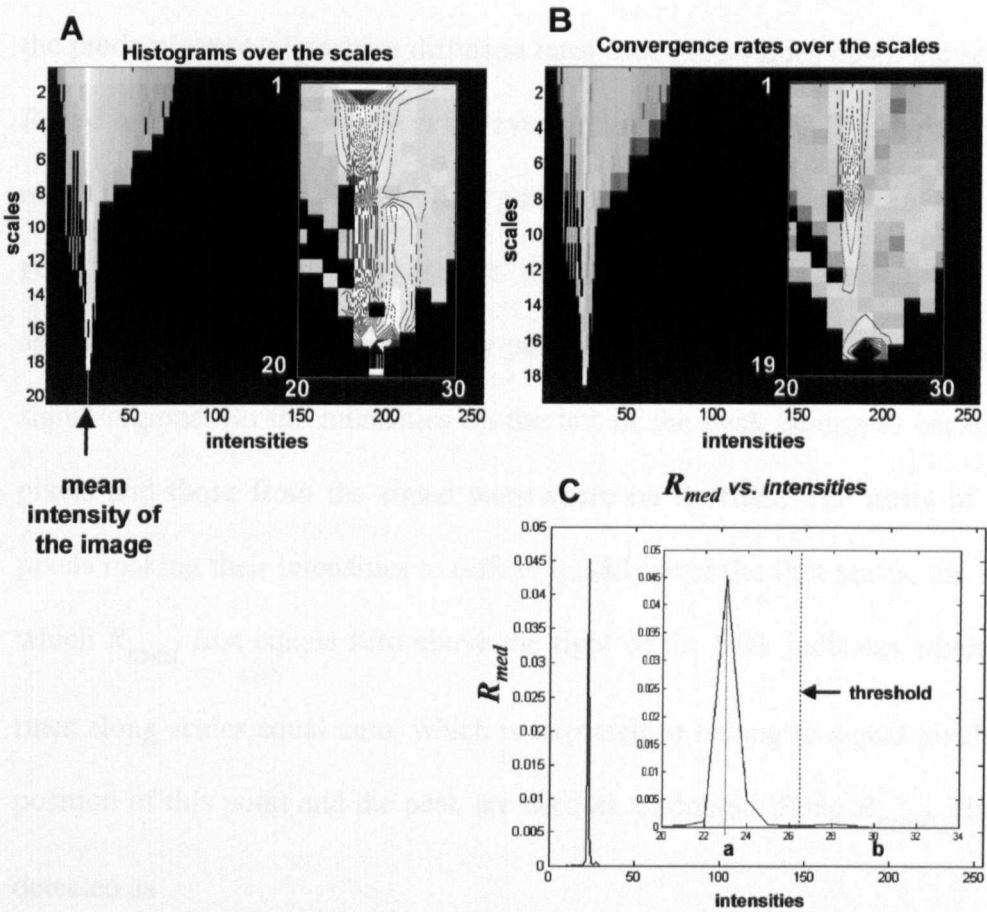
Figure 3.6: **Threshold calculation using the MSA algorithm.**

A – Typical histogram matrix obtained after the 20 rescaling steps of a smoothed LSCM image as involved in the MSA algorithm. *B* – Corresponding diffusion rate matrix. *C* – Median rates calculated down the columns of *B*. The inlets represent magnified portions of the figures. The landmark points **a** and **b** are indicated as well as the threshold set up half way between these points.

The curve properties are used to identify landmarks along R_{med} from which the threshold value is set up. As a result of these properties, R_{med} presents a distinct peak located at, or very close to, the mean image intensities. This peak is due to the predominance of positive diffusion rates over zero-values down the scales in R, and has highest magnitude at the average image intensity. Thresholding at the mean intensity (as is heuristically performed sometimes) does not provide a good segmentation for images with unimodal histograms. Indeed, the image appears over-segmented and a many patchy background regions are included in signal regions. So the intensities on the left of the peak belong to background pixels and those from the signal somewhere on its right. The rarity of signal pixels making their intensities to diffuse quickly over the first scales, the bin for which R_{med} first equals zero above the right of the peak indicates where most rates along scales equal zero, which is expected to belong to signal pixels. The position of this point and the peak are used as landmarks along R_{med}. They are detected as

$$a = \arg \max(R_{med}). \qquad\qquad 3.12$$

as well as the coordinate that first realises

$$b = \arg \left[R_{med} \, (i) = 0 \right] \qquad\qquad 3.13$$

$$i \in [a; 255].$$

The thresholds T that separates background and signal areas belongs to the interval $]a; b[$. Since we are dealing with strong unimodal distributions, the

interval $[a; b]$ is generally small (typically between 7-8 up to 15-20 intensity units, from the range 0-255), as indicated in figure 3.6 (magnified window). All values of R_{med} in $[a; b[$ are positive, and all values in $[b; 255]$ equal 0. Due to the relatively small distance separating a and b, the intensity T used to threshold the original smoothed image is calculated as the average of a and b:

$$T = \frac{a+b}{2}$$
3.14

but more elaborated procedures to calculate T could also be implemented according to the shape information of R_{med} in the interval $[a; b]$. This threshold was found to provide segmentation in very close agreement with human operators (see section 4.2). The output of this algorithm is a binary mask obtained from the smoothed image thresholded at T, used to separate the signal from the background in the original LSCM image. Both the signal and the background are then available for further analysis.

3.3. Background/Image Histogram Comparison (BIHC)

The principle behind this algorithm is to compare the histogram of the original image resized to a much lower scale (obtained using mean intensity interpolation) with the histogram of a typical background area, as illustrated on figure 3.7. This allows the identification of an intensity range distinguishing background from foreground values where a potential threshold could be set (as the landmark points on R_{med} which were used for the same purpose).
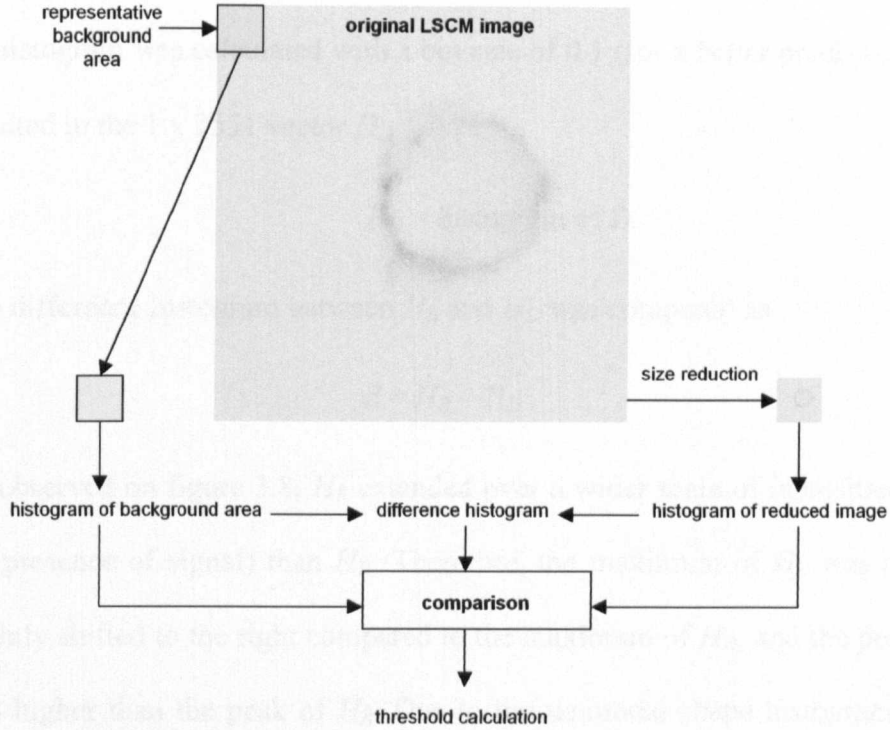
Figure 3.7: **Principle of the BIHC algorithm.**

Steps involved in the BIHC algorithm. Figure 3.8 shows the threshold calculation using histogram comparison.

The original 1024 x 1024 LSCM image was reduced to a 100 x 100 image B using mean interpolation. Since the reduced image contained mainly non-integer elements, we computed the histogram of B in the range [0; 255] with a bin size of 0.1 (for a better precision), which results in the 1 x 2551 vector H_B

$$H_B = \text{histogram of B.} \qquad 3.15$$

In the next step, a histogram representing a background distribution was calculated. To represent a background region, an area C (110 x 110) was selected from the upper left corner of the original LSCM image. This was a reasonable assumption as single cells were generally positioned in the middle of the images. The area C was smoothed using a moving averaging filter (10x10), which generated a 100x100 matrix D while excluding the border values.

As for B, D contained mainly non-integer elements and had the same size as B. Its histogram was calculated with a bin size of 0.1 (for a better precision), which resulted in the 1 x 2551 vector H_D

$$H_D = \text{histogram of D.} \qquad\qquad 3.16$$

The difference histogram between H_B and H_D was computed as

$$d = H_B - H_D . \qquad\qquad 3.17$$

As observed on figure 3.8, H_B extended over a wider scale of intensities (due to the presence of signal) than H_D. Therefore, the maximum of H_B was generally slightly shifted to the right compared to the maximum of H_D, and the peak of H_D was higher than the peak of H_B. Due to the unimodal shape histogram H_B, the mean image intensity is located on the right of the peak. As discussed in section 3.1, the intensities on the left of the peak belong to background pixels while those from the signal are somewhere on its right. In order to find automatically an appropriate threshold, a potential range of values had to be identified. The maximum J of d was first detected such as (see figure 3.8)

$$H_B(J) = \max(H_B). \qquad\qquad 3.18$$

The upper limit T_{max} of the range of potential threshold values was identified as the first intensity that satisfied the condition

$$H_D(i) = 0, \quad \text{with } i \in [J,255]. \qquad\qquad 3.19$$
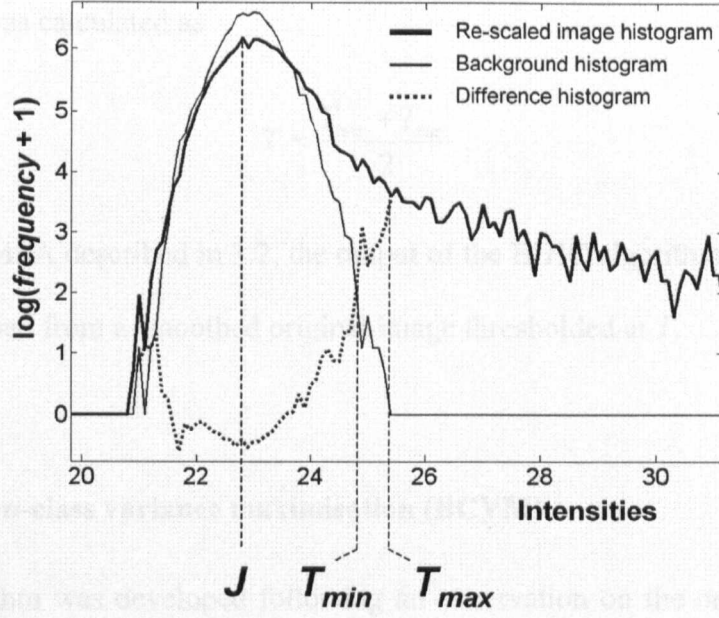
Figure 3.8: **Threshold calculation using the BIHC algorithm.**

Background and rescaled image histograms are compared by calculating the difference histogram. The three curves are used to identify landmark points (J, T_{min}, T_{max}) and the threshold is set up half way between T_{min} and T_{max}.

This indicated that there was no more contribution of background pixels in the histogram of the reduced original image. By comparing the difference histogram d with the background histogram H_D, it was possible to define the range of intensities where most pixels would statistically belong to the background. The lower limit T_{min} of the range of potential threshold values was identified as the first intensity that satisfied the condition

$$d(i) > H_D(i), \quad \text{with } i \in [J; 255]. \qquad 3.20$$

Statistically, all intensities above T_{min} were predominantly found in signal pixels while all intensities below T_{min} were predominantly found in background pixels. As a consequence of the histogram unimodality, the interval $[T_{min}; T_{max}]$ was

generally very small, and a threshold that separated well the background from the signal was calculated as

$$T = \frac{T_{min} + T_{max}}{2}.$$

3.21

As for the MSA described in 3.2, the output of the BIHC algorithm was a binary mask obtained from a smoothed original image thresholded at T.

3.4. Between-class variance maximisation (BCVM)

This algorithm was developed following an observation on the original LSCM images that the variance of a background area was significantly lower than the variance of an area containing signal pixels (figure 2.3). Signal areas comprise pixels with background intensities and also pixels with brighter signal intensities. The range of intensities covered in signal areas is consequently wider than in the background, explaining the higher variance in such areas (see section 2.2). Due to the unimodality of the histogram and the nature of the signal, foreground signal pixels are represented somewhere on the negative slope of the histogram. The proposed algorithm works on the hypothesis that the difference in variance between background and signal intensities is also reflected in the histogram. Thus, a threshold could be found by maximising the between-class variance from the histogram itself, which can be seen as a simplification of Otsu's approach.

The original LSCM image was smoothed with a 10 x 10 averaging kernel and its normalised histogram computed as

$$H, \quad \text{so} \sum_{i=0}^{255} H(i) = 1 \qquad\qquad 3.22$$

A threshold T was moved along H that separates H in two classes $H_{0 \to T}$ and

$H_{T+1 \to 255}$. The variances of both classes were computed such as

$$a(T) = \text{var}(H_{0 \to T}) \qquad\qquad 3.23$$

$$b(T) = \text{var}(H_{T+1 \to 255}) \qquad\qquad 3.24$$

$$\text{with } T \in [0,255]$$

and their absolute difference was calculated as

$$D(T) = |a(T) - b(T)| \qquad\qquad 3.25$$
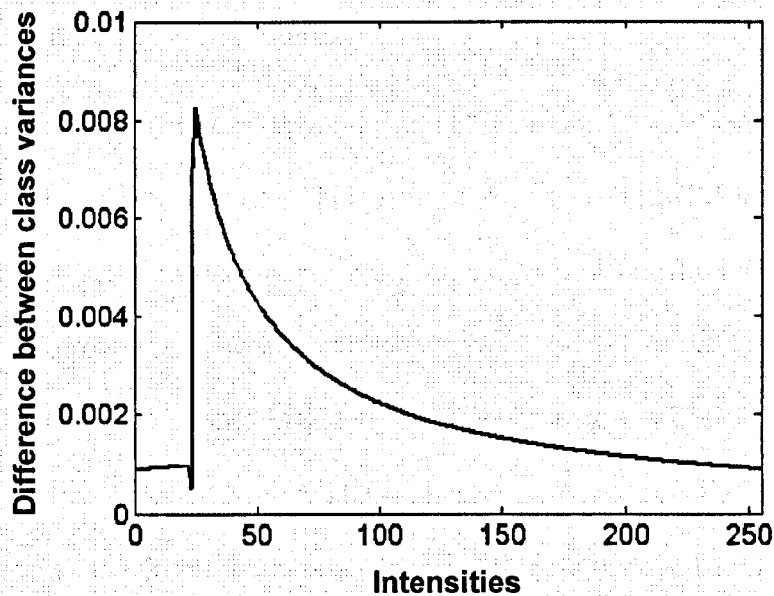
Figure 3.9: **Typical response curve of the BCVM algorithm.**

Absolute difference of classes' standard deviations as the threshold is moved along the X axis. The algorithm sets the binarisation threshold at the position of the peak, indicating the intensity for which this between class variance criterion is maximised.

and was shown in figure 3.9. The appropriate threshold T was set up as

$$T = \arg\max(D) \qquad\qquad 3.26$$

and was used to generate the binary mask as previously described.

3.5. Spatial method: a watershed based algorithm

In this section, a spatial segmentation method is described using the watershed algorithm (described in section 1.3.3.1), as a demonstration of the feasibility of the approach. However this region-based approach does not match the computational efficiency of the global segmentation algorithms previously described and was not subsequently used. In LSCM images, the signal is naturally brighter than the background and consists of several well-defined or loose clusters embedded in a weaker halo. It should be possible to segment such images using a region-based approach.

Watershed methods work best when the signal is distinctly observable as regions superimposed to the background. This is not the case with LSCM images, where the boundaries between the fluorescent clusters, the halo in the cell membrane, and the background are hard to define, either visually or using algorithms. However, the signal being brighter than the background, the first region segmented using watershed segmentation necessarily contains the signal region which contains the highest intensity values. This region is labelled 1 in the corresponding label matrix. The second watershed is also likely to encompass a signal region, it is labelled 2 in the label matrix, and so on for the first watersheds. The hierarchical nature of watershed segmentation is responsible for segmenting signal regions first, and the background later, due to the properties

of LSCM images. The rationale behind the new proposed algorithm is to describe a method which would allow to estimate which of the region label separates well signal and background watersheds, simply taking into account the labels and size of these watershed regions.

The watershed algorithm is well known to be sensitive to noise. Hence, in this study, a 10x10 median filter was first applied to attenuate the effect of the speckle noise. Then, the watershed algorithm was applied to the filtered LSCM image. A label matrix L was computed which contained the segmented regions whose elements are labelled with the order number in which they were segmented, separated by boundaries consisting of zero-valued pixels. This label matrix L is further used to discriminate between signal and background. The histogram H_L of L is then computed in the range [1; L_{max}], where L_{max} represents the label value of the last region segmented by the watershed algorithm. It is hypothesized that the sizes of the watersheds in signal regions, and the order in which they are detected, will present different and distinct behaviours/distributions along the histogram of L. Such a histogram is displayed in figure 3.10-A. It can be noticed that approximately the first third of the histogram seem to contain lower values in comparison to the remaining two thirds of the curve. This tendency is made even more obvious after two convolutions of the histogram with an averaging filter of length 5 (figure 3.10-B). After this smoothing operation, the baseline of the first third of L is noticeably lower than the baseline of the remaining two thirds of the curve. This observation can be further demonstrated by computing the histogram of the histogram of H_L, between the minimum and maximum values of H_L. This histogram is displayed in figure 3.10-C. Two peaks, each corresponding to the

different behaviours of H_L are visible. They are separated around the value 50. These observations tend to confirm that watershed segmentation can indeed be used to segment the fluorescence signal in LSCM images.
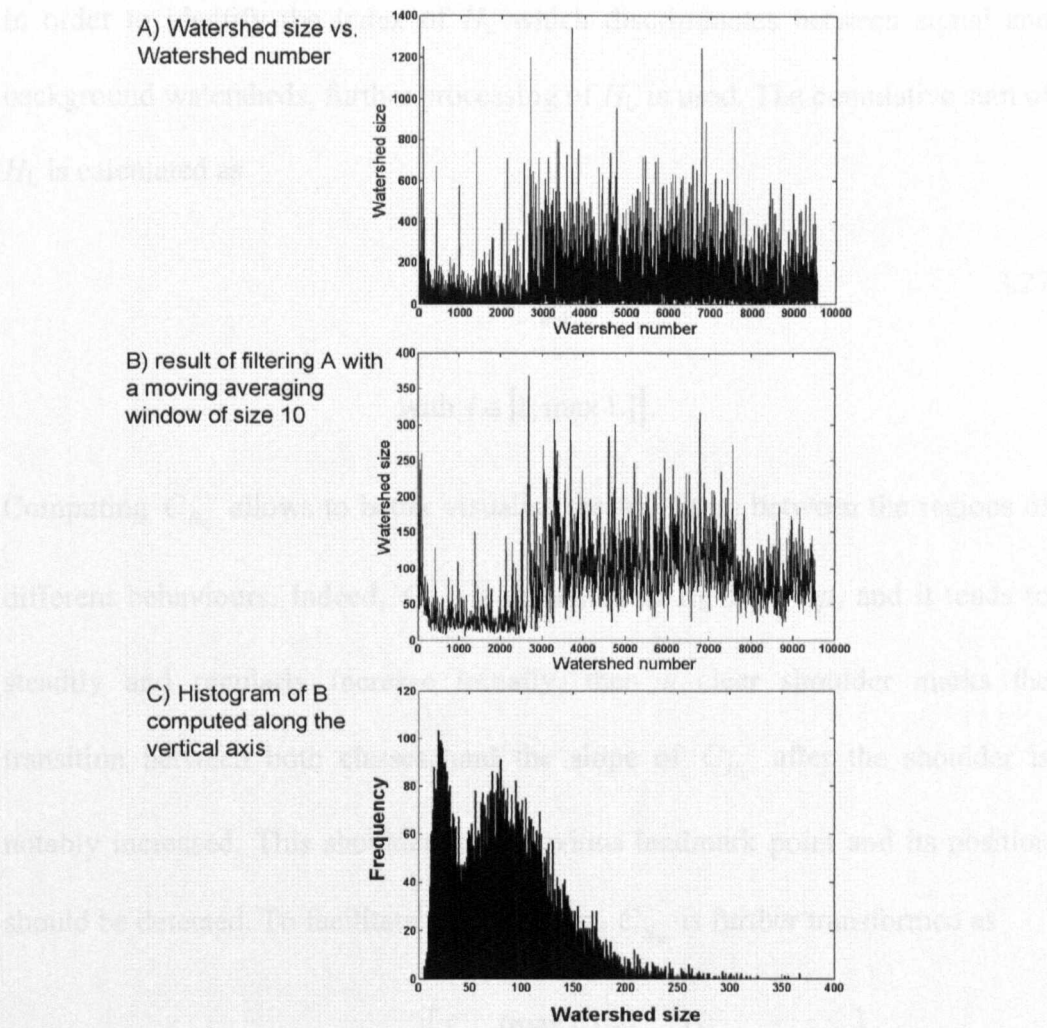
Figure 3.10: **Using watershed segmentation to produce bimodal data from unimodal LSCM images.**

A – Watershed size plotted against the watershed index (i.e. the order in which each watershed was detected). The first watershed basins detected belong to the signal regions, in principle, while the last ones belong to background regions. It seems that above the index 2500, the watersheds are significantly bigger than below this index, indicating a potential threshold for discriminating signal from background. This is more obvious after smoothing – B – the original curve. After this operation, the baseline of the curve is much lower below index 2500 than above. C – The histogram of curve B computed along the vertical axis eventually demonstrates that this data is bimodal in nature, the first peak being representative of the signal regions while the second peak is mainly due to background regions.

In order to identify the index of H_L which discriminates between signal and background watersheds, further processing of H_L is used. The cumulative sum of H_L is calculated as

$$C_{H_L}(i) = \sum_{n=1}^{i} H_L(n) \qquad\qquad 3.27$$

with $i \in [1; \max L]]$.

Computing C_{H_L} allows to better visualise the transition between the regions of different behaviours. Indeed, C_{H_L} is presented in fig 3.11-*top*, and it tends to steadily and regularly increase initially, then a clear shoulder marks the transition between both classes, and the slope of C_{H_L} after the shoulder is notably increased. This shoulder is an obvious landmark point and its position should be detected. To facilitate this detection, C_{H_L} is further transformed as

$$C'_{H_L}(i) = C_{H_L} - \left[\frac{C_{H_L}(\max L) - C_{H_L}(1)}{\max L} i + C_{H_L}(i) \right] \qquad 3.28$$

where the terms in the square brackets represent the equation of the straight line which joins the first and last point of C_{H_L}. Therefore, C'_{H_L} is the difference between C_{H_L} and the straight line which joins the first and last point of C_{H_L}. Such transformation makes the detection of the shoulder in C_{H_L} much easier. Indeed, the maximum distance point is located at the position S of the extrema of C'_{H_L} without considering the sign of this extrema. This point indicates the transition from one type of regime (or behaviour) to another for the sizes of the segmented watershed basins. The curve C'_{H_L} is displayed in figure 3.11 and it
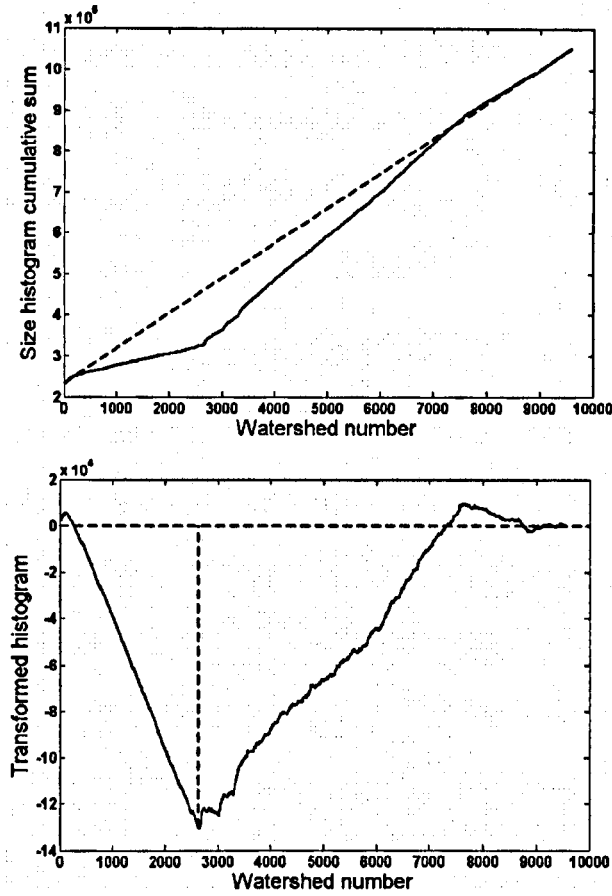
Figure 3.11: **Locating a threshold from watershed data.**

Top – Cumulative sum of the first curve presented in figure 3.10-A. In this case, three main linear segments are visually identified, each with a different slope. The dashed line joins the first and last points of this curve. *Bottom* – Result of subtracting the dashed line to the cumulative curve. This procedure emphasises the transition between the linear segments. The location of the minimum point is located as a landmark point: watersheds with indexes higher than this point belong to background. For simplicity, the indexes of the watersheds belonging to signal regions are those between 1 and half the value of the landmark curve minimum position.

can be observed that the extrema corresponds to the minimum of C'_{H_L}, which is located around the watershed number S=2600. Using the transformation described in equation 3.28, the steady regime observed of the left of the shoulder is also made clear and appears as a rather straight line with negative slope (with a small peak located at the first watershed regions). The more erratic nature of

116

the watershed segmentation process in background areas is reflected by the rougher evolution of C'_{H_L} after the position of the shoulder.

From experimental observations, the threshold T is set up half way between 1 and the position of the shoulder, i.e.

$$T = \frac{S}{2}.$$

3.29

In order to obtain the binary mask (fig. 3.12) of the signal in the LSCM image, the first T watersheds and their boundaries are labelled 1 (for signal regions), while the rest of them are labelled 0 (for background regions). Therefore, this thresholding procedure is not a global thresholding approach. It is region-based and has the effect of local thresholding (i.e. different thresholds are chosen locally as a consequence of the watershed procedure).

Negative original LSCM image

Result of filtering the LSCM micrograph with a 10x10 averaging filter

Watershed segmentation result. The darkest areas indicate the earliest segmented watershed

Segmentation of the watershed output
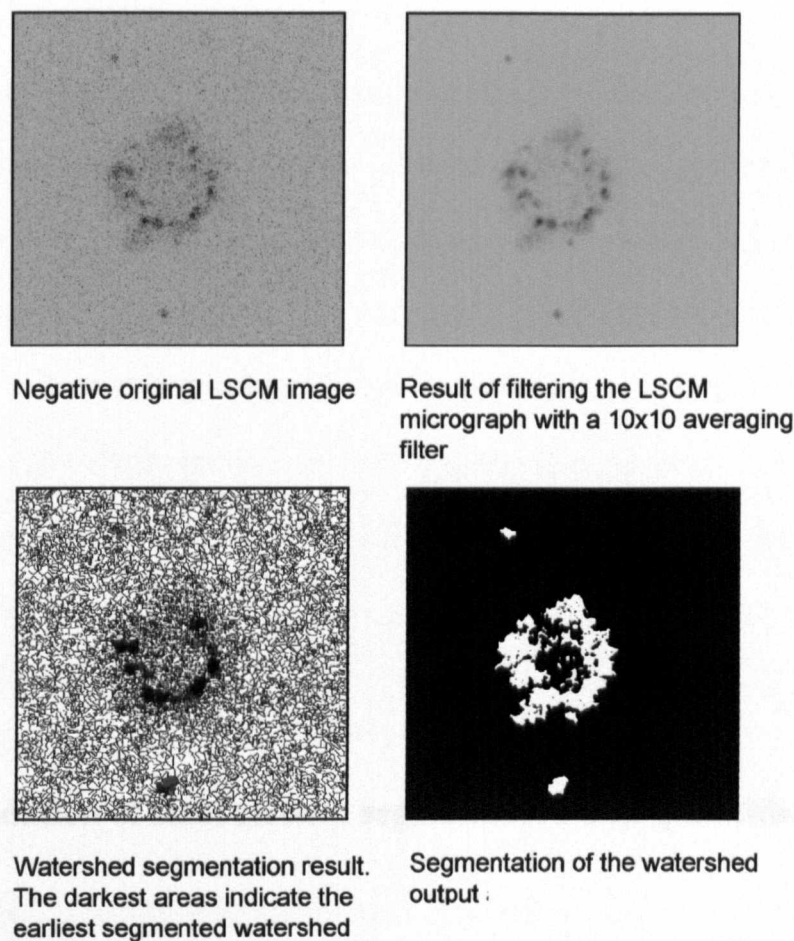
Figure 3.12: Example of image segmented using the watershed algorithm proposed in this thesis.

The top left image is the original LSCM image, and the bottom right image is the binary mask obtained from the proposed watershed algorithm. This watershed algorithm provides good segmentation result in images with a reasonable contrast, but was observed to loose its consistency with poorer contrast images. Furthermore, it is much more time consuming than the global thresholding approaches developed in this work.

Chapter 4

Validation of fluorescence segmentation and quantification

Chapter 4. Validation of fluorescence segmentation and quantification

Segmentation performances of these algorithms were assessed by comparing their outputs. The MSA algorithm was potentially more interesting to use due to theoretical considerations and it was subsequently used to segment all datasets prior to their analysis. Since this analysis relies on the accurate identification and quantification of the signal in LSCM images, artificial LSCM images were generated to test the reliability of the segmentation performed by the MSA algorithm. The first section of this chapter (section 4.1) describes the construction of such artificial images. In the second section (section 4.2), the segmentation algorithms are compared and fluorescence quantification using the MSA algorithm is validated.

4.1. LSCM image synthesis for segmentation validation

4.1.1. Laser scanning confocal image synthesis

Theoretically, several ways are possible to generate artificial LSCM images. LSCM is commonly used to obtain 2D images through an optically transparent sample, or in 3D in order to capture the three-dimensional structure of this sample. Thus, for some test applications, the artificial/phantom data can be generated from a two dimensional model which aims to describe the signal generation in particular optical planes. Some other applications may consider the influence of the 3D structure of the sample on the generation of artificial images, and therefore a 3D model may be constructed to generate artificial z-sections series (not covered in this thesis).

For both approaches, preliminary measurements can reveal useful information to estimate the parameters of the models. Such parameters include the probability density function (PDF) of the noise, the size and shape of the cells, possibly the point spread function of the system, etc. In order to make artificial data as similar as possible to real data, these preliminary measurements may be performed on small sets of original LSCM images.

In the first step of the process, the noisy background of the image is generated from a measured PDF from real images. The second step simulates the generation of the fluorescence signal according to different assumptions on the nature of its spatial distribution.

4.1.2. Background noise generation

Various noise models are commonly used in image synthesis. Well-known models include uniform, Gaussian, Rayleigh, gamma, exponential and impulse noise. The PDF of some of these distributions are given below, and can be used to generate spatially non-correlated background noise in artificial LSCM images:

Gaussian PDF:
$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

Rayleigh PDF:
$$p(z) = \begin{cases} \frac{2}{b}(z-a)e^{-\frac{(z-a)^2}{b}} & \text{for } z \geq a \\ 0 & \text{for } < a \end{cases}$$

Gamma PDF:
$$p(z) = \begin{cases} \dfrac{a^b z^{b-1}}{(b-1)!} e^{-az} & \text{for } z \geq 0 \\ 0 & \text{for } z < 0 \end{cases}$$

Exponential PDF:
$$p(z) = \begin{cases} a e^{-az} & \text{for } z \geq 0 \\ 0 & \text{for } z < 0 \end{cases}$$

where z is a random variable. The quantities a and b are positive constants which are used to define the mean and variance of these PDF.

The noise in microscope images is theoretically assumed to have a Poisson distribution, or a mixture of Poisson distributions. In some real datasets, the experimenter may have adjusted the brightness and contrast of the image so that the noise histogram is stretched and shifted to the left of the graph. This results in an observed exponential PDF. Therefore, an estimate of the noise actual PDF is valuable to generate realistic artificial images, and the noise in the proposed synthetic images was modelled from PDF measured on real datasets. Indeed, simplistic assumptions may reveal not to be accurate enough or may generate wrong and biased results when the algorithms are tested on them. The PDF $p_{measured}(i)$ of a manually selected background region (supporting 8-bits encoded intensity values) is obtained by dividing each value of its histogram $h(i)$, where i is the intensity index, by the total number of pixels used to compute it, so

$$\sum \text{PDF} = \sum_{i=0}^{255} p_{measured}(i) = 1. \qquad\qquad 4.1$$

The generation of the artificial noisy background is achieved by directly calculating the number of pixels of intensity i multiplied by the total number of pixels in the final $N \times N$ image, i.e

number of pixels with intensity $i = NP(i) = N^2 \times p_{measured}(i)$.   4.2

Random variability is introduced in $NP(i)$ to avoid the synthesis of background with exactly similar histograms, with a condition ensuring that the final number of pixels produced equals N^2. For any intensity i, $NP(i)$ non-overlapping pixels are randomly selected in the $N \times N$ image which are consequently attributed the intensity value i. The histogram of such an artificial background synthesised from a real background PDF is shown in figure 4.1. It does compare well with the actual noise histogram. This method allows to generate artificial, uncorrelated noisy backgrounds given any arbitrary PDF. Such process can therefore model any type of noise PDF from actual data.

Figure 4.1: **Background noise generation for artificial LSCM synthesis.**

A – Negative optical section (HSPC labelled for CD164). The dashed square encloses a fluorescence-free area, used to estimate the probability density function of the noisy background. B – *Left*. Histogram of the whole image A (the frequency is displayed on a logarithmic scale). *Right*. Histogram of the background area in A (log-scale). Such histogram does not look like any common noise distribution, but more like a combination of several distributions, possibly of exponential and/or Poisson nature. The probability density function of the background noise is estimated by dividing this histogram by the number of pixels contained in the selected background area. C – Histogram of an artificial background generated using the probability density function estimated from the background histogram.

4.1.3. Fluorescence signal modelling

As stated earlier, both 2D and 3D models of the fluorescence signal produced by fluorochrome-labelled antigens in the stem cell membrane can be used to generate synthetic confocal images. In LSCM, the PSF in the horizontal (*x-y*) plane is commonly assumed to have a Gaussian form [Boutet de Monvel *et al*, 2001]. In this study, we considered that the PSF is normally distributed both in a 2D or a 3D referential. For simplicity, we show in this section how a 2D optical section is synthesised.

4.1.4. Two-dimensional model

For simplification, the model assumes that the PSF function of the system is Gaussian. Therefore the observed fluorescence distribution is built by the circular superposition of 2D Gaussian functions, each centred on the locations of the fluorochrome/antigen complexes in the circular cell membrane, plus the noise component. These are reasonable assumptions when compared with real LSCM images. However the existence of a weaker fluorescent signal was also noticed inside the cell as a much more diffuse internal halo, (probably due to the internalisation of the fluorochrome during the time-consuming labelling steps, heavily blurred by the transfer function of the system). We integrated this observation in the model by adding extra smaller rings of Gaussians with variances increasing proportionally to an inverse function of the radius.

The parameters of the model are the number of rings, or layers, the shape parameters of the rings, the number of Gaussians per ring, their variances, their

distribution along the rings, and the PDF of the noise. The radius r of the external ring is expressed in polar coordinates α as a sum of K cosine functions with decreasing amplitudes and increasing frequencies

$$r(\alpha) = \sum_{n=1}^{K} \tau_n \cos(A_n \alpha + B_n)$$ 4.3

with $\tau_1 > \tau_2 > ... > \tau_K$ and $A_1 < A_2 < ... < A_K$

where τ_n, A_n, and B_n respectively represent the amplitude, the frequency and the phase of the nth ring. Constraints are imposed on the range of values which can be taken by these variables, in order to ensure roughly circular shapes of the cell membrane as observed in real LSCM images. The internal rings are scaled versions of the external one. In this study, we used different uniform random distributions to select the locations of the fluorescent clusters on the different rings.

The artificial fluorescence is generated by averaging the different rings, which represent different layers of fluorescence clusters blurred at various degrees. The noise is added to the image according to the procedure described in section 4.1.2. Figure 4.2 shows an artificial LSCM image generated according to this method and using only three layers. Arbitrary shapes can be obtained by carefully choosing the values of the parameters of $r(\alpha)$.
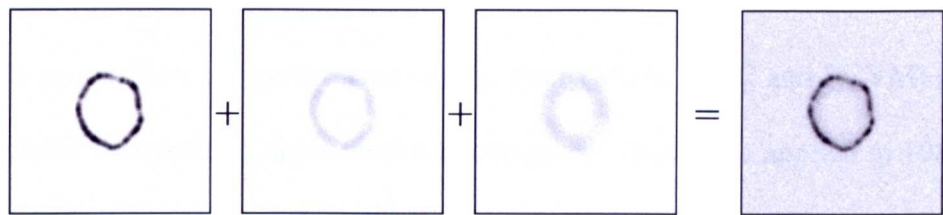
Figure 4.2: **Artificial LSCM synthesis.**

Artificial LSCM image generated by averaging three layers. The first layer represents the actual fluorescence signal. The second and third layers are blurred versions of the signal. Each layer is generated by successively decreasing the radius of the cell and by increasing the variance of the Gaussian functions used to model the fluorescence signal. The same noise background was added to all layers, so averaging the layers does not affect the actual background noise distribution.

4.2. Performance analysis of the proposed segmentation algorithms

In this section, the performance of the three (MSA, BIHC and BCVM) proposed global thresholding algorithms was compared. They were applied to 198 distinct LSCM images of umbilical cord blood HSPCs from 10 individual datasets. These cells were labelled for the CD34 and CD164 membrane antigens. The comparison was firstly performed by visually comparing the binary masks obtained from the 198 LSCM images, secondly by comparing the thresholds calculated to obtain these binary masks, thirdly by comparing the mean intensity of the signal within the masks, and finally by comparing the binary masks' signal areas. The last two comparisons were performed to assess the feasibility of automated signal quantification. Indeed, signal quantification in fluorescence imaging is often performed by either comparing the signal intensities or the areas expressing the signal after thresholding.

Only images thresholded using the three global thresholding algorithms (MSA, BIHC and BCVM) were compared to allow consistency: the watershed-based algorithm, despite performing the desired task, does not provide a unique global threshold to be compared with the others. Furthermore, this approach was implemented to demonstrate the principle of thresholding bimodal histogram obtained through processing of unimodal data, and is shown to be much slower than the global approaches.

The following analysis demonstrated that antigen quantification is far more reliable if fluorescence intensities are used rather than only the extent of the mask covering signal regions (sections 4.2.3 and 4.2.4). The effect of the new quantification formula proposed in section 5.1.1 is also assessed using a slightly

modified version of the graphical Bland-Altman approach used to assess the agreement between different (here quantification) methods.

4.2.1. Visual assessment of binary masks

Figure 4.3 displays the binary masks obtained with the MSA, BIHC and BCVM algorithms for three representative LSCM images taken at different z-positions. The first one was taken at the free pole of the cell where, in this case, almost no signal is distinguishable, even on the negative image which emphasises low
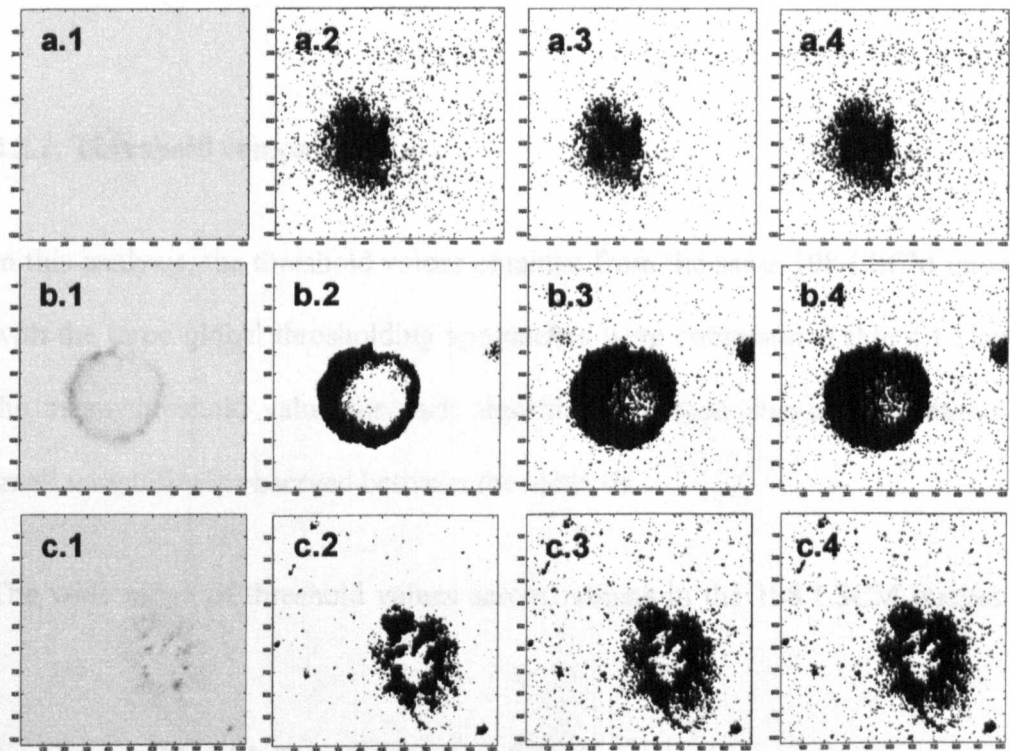
Figure 4.3: **Visual assessment of fluorescence masks.**

First column – Inverted original LSCM images respectively taken at the free pole of the cell, the equator and the contact point with the microscope slide. *Second column* – Binary signal masks obtained with the MSA algorithm. *Third column* – Binary masks obtained with the BIHC algorithm. *Fourth column* – Binary masks obtained with the BCVM algorithm.

intensities. The second image was taken at the middle of the cell, where generally the best contrast in the z-stack is obtained. The last image was taken close to the contact point of the cell with the microscope slide where stained debris are visible.

From this initial, subjective assessment, the MSA algorithm showed good results, and appeared to be less sensitive to very weak signal areas in images with good signal contrast (fig.4.3-b2, 4.3-c2). However, in poorly contrasted images, very weak signals were identified with a high sensitivity (fig.4.3-a2). The BCVM and the BIHC algorithm showed good performances at detecting extremely low intensity signals, regardless of the image contrast.

4.2.2. Threshold comparison

In this analysis, the threshold values obtained from the same 198 LSCM images with the three global thresholding approaches were compared. Table 4.1 shows the mean threshold value for each algorithm and their standard deviation. A small variability is observed between the methods.

The wide range of threshold values across z-scans in the 198 LSCM images is

Table 4.1. **Threshold comparison.**

	Algorithms		
	MSA	BIHC	BCVM
Mean threshold	21.98	21.76	20.76
Standard deviation	6.68	6.01	6.29

responsible for the high standard deviations. However, the mean threshold values and their standard deviations are in good agreement for all three methods.

For a more detailed comparison, each of the algorithms was compared with the others using regression analysis. Figure 4.4 shows the 198 thresholds obtained by each algorithm plotted versus those obtained from the other algorithms. Homogeneous data sets (i.e. z-scans performed on the same microscope slide) are responsible for the dot clusters on the graphs. Indeed, several z-scans were performed on different slides, generating different background values, which results in the clusters observed on the graphs.

The first graph in figure 4.4 presents the thresholds from BIHC vs. MSA. The thresholds vary between 10 to 40, reflecting a wide range of intensities. It is observed that a good linear relationship ($R \approx 0.92$) exists between these thresholds. The slope of the best fitted line is equal to 1.018, indicating that the MSA method calculates, on average, thresholds slightly higher than the BIHC method. This was also confirmed by the visual examination of the binary masks. The high correlation coefficient and the slope of the best fitting line indicate that both methods, while different in principle, produced similar results.

The second graph in figure 4.4 presents the thresholds from the BIHC vs. BCVM. Once again, the linear relationship is very well verified, with a correlation coefficient $R \approx 0.96$. The slope of the best fitting line is 1.008, indicating that both the BIHC and the BCVM algorithms are in good agreement.

The last graph presents the thresholds from the MSA vs. BCVM. Again, a good linear relationship ($R \approx 0.96$) can be observed. The slope of the best fitting line

is equal to 0.903, indicating that most thresholds obtained with MSA are slightly

higher than those obtained with BCVM.

This analysis allows the following general classification

$$\overline{T}_{MSA} > \overline{T}_{BIHC} > \overline{T}_{BCVM}$$

where $\overline{T}$ indicates the mean threshold given by each indexed algorithm.
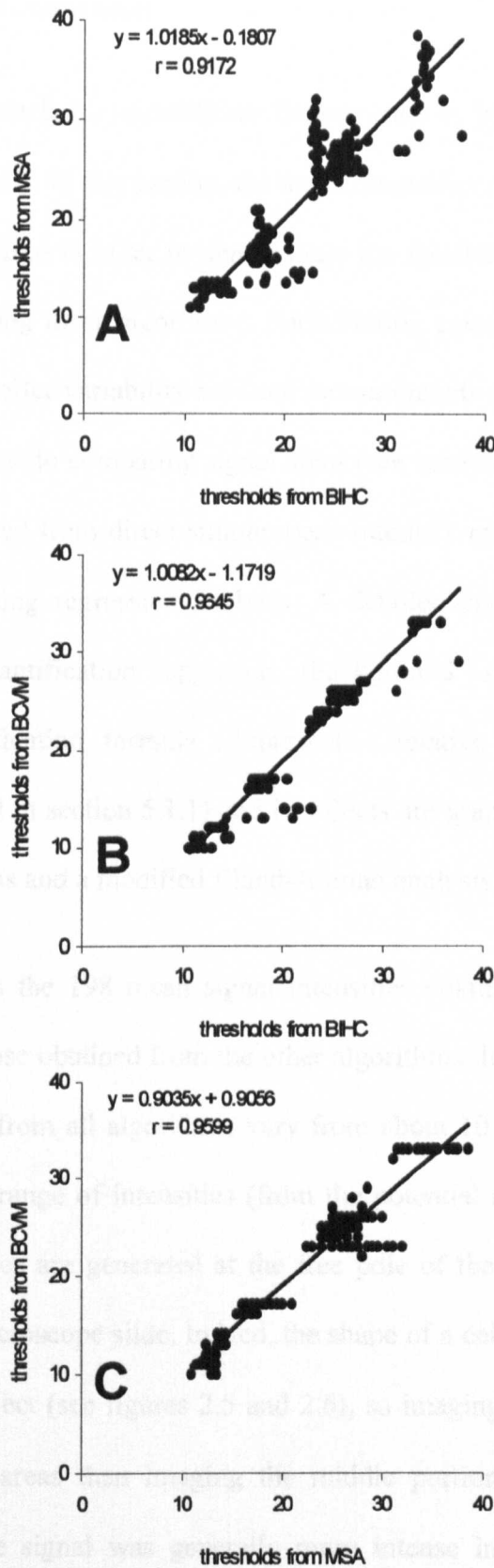
Figure 4.4: **Inter-algorithm threshold comparison.**

A – MSA vs. BIHC. *B* – BCVM vs. BIHC. *C* – BCVM vs. MSA. The equations of the best fitted lines, obtained by linear regression, and the Pearson's correlation coefficient *r* are also indicated for each plot.

4.2.3. Mean signal comparison

A common approach for quantifying fluorescence is by measuring the mean intensity in the ROI. In this section, the mean intensities of the segmented signal areas were compared in order to demonstrate the feasibility of automatic signal quantification using these algorithms. Furthermore, comparing mean intensities provides with smaller variability between measurements obtained from different methods in contrast to comparing signal areas (see section 4.2.4). In this section, the results obtained from direct simple mean intensity measurement in the ROI are compared using regression analysis. A detailed comparison of the usual fluorescence quantification approach (background subtraction) with the proposed quantification formula (fluorescence relative to mean background intensity, detailed in section 5.1.1) and its effects are graphically analysed using regression analysis and a modified Bland-Altman analysis in section 4.2.7.

Figure 4.5 shows the 198 mean signal intensities obtained by each algorithm plotted versus those obtained from the other algorithms. It was observed that the mean intensities from all algorithms vary from about 10 to about 70, therefore covering a wide range of intensities (from the potential range [0; 255]). Weak mean signal values are generated at the free pole of the cell, or at its contact point with the microscope slide. Indeed, the shape of a cell can be approximated as a spherical object (see figures 2.5 and 2.6), so imaging both poles takes into account smaller areas than imaging the middle portions of the cell. It was observed that the signal was generally more intense in the middle sections, where the best resolution was obtained, explaining the high mean values observed on the graph.

Regression analysis showed a good linear relationship between the results from the three algorithms (R ≈ 0.91, 0.94 and 0.94). The slope for the best fitted line is 1.245 for the mean intensities provided by MSA vs. BIHC, 0.983 for BIHC vs. BCVM and 0.716 for BCVM vs. MSA. Overall, the mean intensities calculated from MSA are, on average, higher than those obtained from BIHC and BCVM. This is due to the fact that the MSA thresholds are, on average, slightly higher than the thresholds from the other algorithms. As a consequence, the weakest signal intensities are lost in the MSA algorithm and thus the mean signal intensity is higher. The best agreement was obtained for BIHC vs. BCVM, with R ≈ 0.94, and the slope of the best fitted line equals to 0.983, indicating the close similarity between the mean signal values obtained from both algorithms.
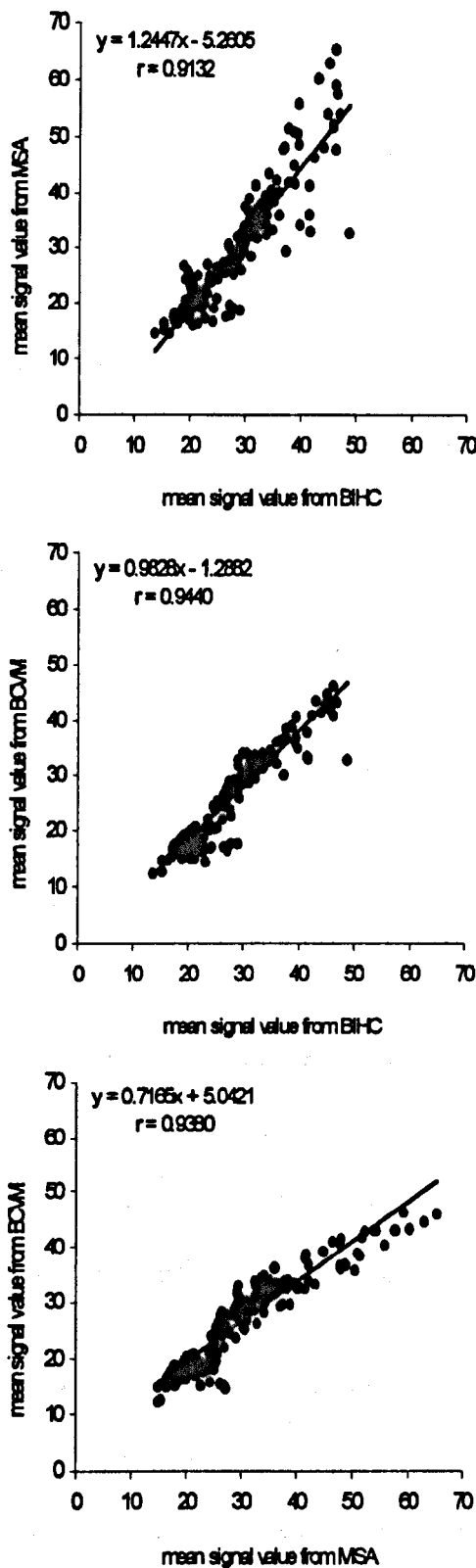
Figure 4.5: Mean signal intensity comparison.

Top – MSA vs. BIHC. *Middle* – BCVM vs. BIHC. *Bottom* – BCVM vs. MSA. The equations of the best fitted lines, obtained by linear regression, and the Pearson's correlation coefficient *r* are also indicated for each plot.

4.2.4. Binary mask area comparison

As mentioned earlier, a common approach to quantify fluorescence consists in measuring the area of a ROI where the fluorescence above a certain threshold is present. The threshold is often chosen manually by an experimented operator. This decision is highly subjective, and different results are likely to be obtained by different experimenters. In this section, the binary masks obtained using the MSA, BIHC and BCVM algorithms were compared using regression analysis, assuming that the area where the fluorescence is expressed may be a relevant criterion to quantify fluorescence.

The areas from one algorithm were plotted versus the areas from another algorithm (i.e. three graphs are needed to compare the three possible combinations, figure 4.6). It can be observed that the good linear relationships previously obtained when comparing thresholds and mean signal intensities is lost when comparing the areas of the binary masks. This is the consequence of the histogram unimodality, where small variations of the threshold values generate large variations of the segmented areas as thresholds are set up where the histogram slopes have high negative values. This is an important observation with consequences for signal quantification as detailed in the discussion.

Table 4.2 shows the relative comparison between signal areas obtained from all three algorithms. The highest areas were obtained with BCVM (due to the high sensitivity of this algorithm in detecting very low signals). MSA generated the smallest areas. Interestingly, it was found that 11.61% of the areas obtained with MSA were exactly identical as those obtained through BCVM. Also 3.03% of the areas obtained through BIHC were exactly identical as those obtained

through BCVM, and 2.02% of the areas obtained through MSA were exactly identical as those obtained through BIHC. On one occasion, all three algorithms provided exactly the same binary mask.

In our investigation, all masks did not exceed 30% the size of the image. The BCVM and the BIHC masks covered a wider range of areas than the MSA masks. Most of the areas segmented by MSA and BIHC were found to be around 10% the size of the image, while BCVM predominantly generated areas around 15% the size of the original image.
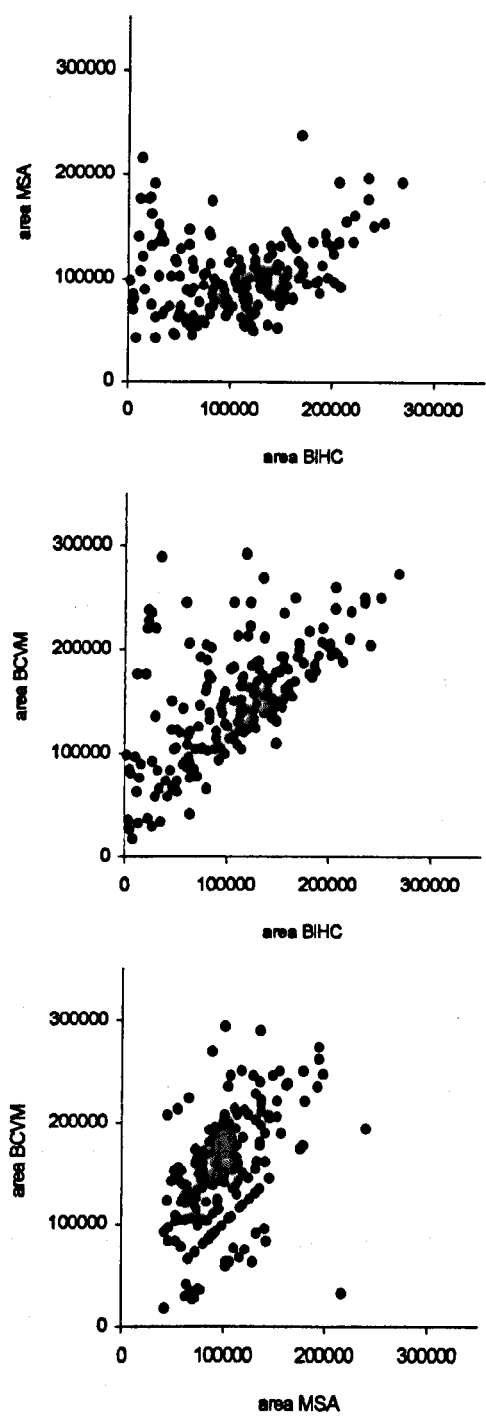
Figure 4.6: Area comparison.

Top – MSA vs. BIHC. *Middle* – BCVM vs. BIHC. *Bottom* – BCVM vs. MSA.

Table 4.2. **Binary masks area comparison.**

Area comparison	
BCVM > MSA	89.40%
BCVM > BIHC	87.38%
MSA > BIHC	34.85%
BCVM = MSA	11.61%
BCVM = BIHC	3.03%
MSA = BIHC	2.02%

4.2.5. Validation of the MSA LSCM image segmentation using artificially generated LSCM test images

It was proposed in section 4.1 to generate realistic artificial LSCM images in order to assess the segmentation performance of the MSA algorithm, the most interesting of the proposed segmentation approaches. In that purpose, over 100 artificial test images were automatically generated and segmented using this algorithm.

The position of the true signal was known. This position was the area encompassed by the mixture of Gaussian kernels used to produce the first image layer in figure 4.2. (without taking into account the noise component which is added separately for realistic rendering). The mean signal intensity was therefore also known.

In order to assess the performance of the MSA segmentation, two quantities were measured from the actual binary mask A_{true} of the true signal and from the binary mask A_{obs} provided by the MSA method. The first quantity is an area ratio

$$Ratio_{\text{area}} = \frac{A_{true} \cap A_{obs}}{A_{true} \cup A_{obs}} = \text{AND/OR ratio}. \qquad 4.4$$

This quantity takes into account true colocalised/overlapping areas (the nominator) and the total areas covered by either masks (the denominator). For perfect segmentation, the AND/OR area ratio would be equal to 1, or 100%. In any case of non perfect segmentation, this ratio would drop. It was subjectively assumed that a reasonable segmentation should produce a ratio equal or higher than 70%, so colocalised binary masks should cover at least 70% of the true signal location. It is noticed that most of the true signal location contains pixels with intensities very close to background values, due to the Gaussian nature of the signal, and therefore only a small proportion of the true signal area contains significantly bright signal pixels. Consequently, by segmenting at least 70% of the true signal area, it is very likely to also correctly segment the most relevant signal pixels.

The second quantity used to assess MSA segmentation is an intensity ratio

$$Ratio_{\text{intensity}} = \frac{\overline{I_{A_{true} \cup A_{obs}}}}{\overline{I_{A_{true} \cap A_{obs}}}} = \text{OR/AND ratio}, \qquad 4.5$$

where $\overline{I_{A_{true} \cup A_{obs}}}$ and $\overline{I_{A_{true} \cap A_{obs}}}$ are the mean intensities in the OR and AND areas defined above. Because the mean intensity of the AND area is higher than the mean intensity of the OR area (which encompasses larger areas of weaker intensities), the ratio OR/AND was used to provide values below 1, or 100%. In case of exact segmentation, this ratio would be equal to 100%. In case of over-segmentation, the ratio would drop dramatically as a consequence of the image

histogram unimodality, and for under-segmentation it would be higher than

100%. Preliminary testing showed that an OR/AND ratio between 80% and

100% would reflect acceptable segmentation performances.

The results of area and intensity ratio measurements are presented in figure 4.7.

The mean area ratio is 80.36% ± 5.10, and the mean intensity ratio is 90.84% ±

Validation of segmentation algorithm
MSA

Area ratio   = 80.36% ± 5.10
Intensity ratio   = 90.84% ± 1.11
Number of cells = 102
Cell radius   = 123.94 ± 26.76 pixels
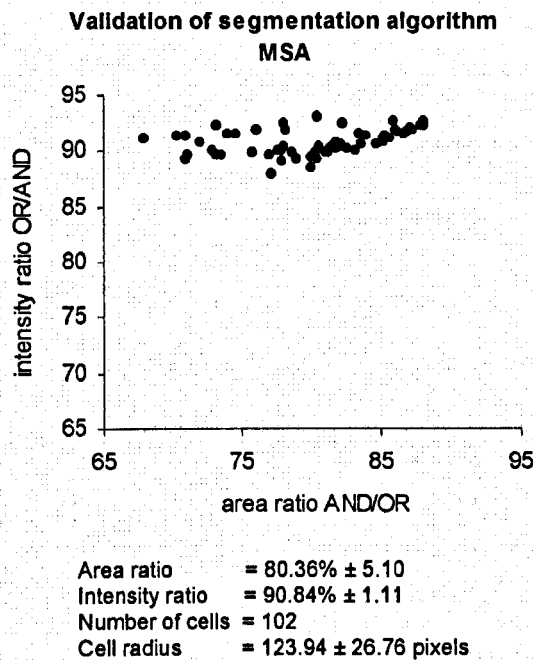
Figure 4.7: **Validation of the MSA segmentation performances.**

Ratios of areas and intensities calculated over more than 100 artificial LSCM images of cell cross-sections. The images were generated using the method described in section 4.1. The area ratio is calculated from the areas of the binary masks obtained from the MSA algorithm (applied to the artificial LSCM images) and from the (known by definition) area containing the artificial fluorescence signal. In average, 80% of the signal area segmented using the MSA covers the true position of the signal. This confirms that the MSA algorithm segments real LSCM images properly.

1.11. These values are well inside the intervals considered to indicate good segmentation. They also confirm that MSA does not under- or over-segment LSCM images non-reasonably. Thus segmenting real LSCM images using MSA is reliable and can be used for quantification purposes.

4.2.6. Bland-Altman analysis of quantification approaches

Measurements provided by two different methods are commonly compared in the literature by using Linear Regression Analysis and calculation of correlation coefficients, classically the Pearson's correlation coefficient R being used. Such statistical methods were initially designed to assess the degree of linear relationship between measures. They do not necessarily reflect the agreement, or the lack of it, between measurements obtained from different methods. This situation has been well described and explored by [Bland and Altman, 1986]. Indeed, if the methods aim to measure the same quantity, it is likely that they will give similar results in the first place. As a consequence, these measures are likely to be well correlated. The authors report that correlation coefficients around 0.9 (which indicates strong correlation) are commonly encountered in the literature. In such cases, measuring correlation is not enough to assess the degree of agreement between two methods. They propose an alternative, yet simple graphical approach to assess this degree of similarity between measurements, obtained from two different apparatus in their study. By extension, the method is directly applicable to measures obtained from different algorithms. Furthermore, the method was initially described for small samples, which are often encountered in biomedical research.

The first step of the analysis consists in plotting the measurements from the first method versus those obtained from the second method. This is done to confirm the expected linear relationship between both methods. Then, the differences between pairs of measurements are plotted against the average value of the pair. This allows to assess the relation between the difference and the mean. Indeed, it is expected that most of the differences would lie between $\overline{d} - 2s$ and $\overline{d} + 2s$, or more precisely between $\overline{d} - 1.96s$ and $\overline{d} + 1.96s$, which is the 95% confidence interval, if the differences are Normally distributed (where $\overline{d}$ and s represent the mean difference and the standard deviation of the differences). This interval $\overline{d} \pm 2s$ is referred to as the limit of agreement. A large interval, relatively to the actual measures, reflects a lack of agreement between methods, despite a high correlation coefficient.

This method was applied to paired mean intensity measurements from the MSA, BIHC and BCVM algorithms, which consist in three different combinations (MSA/BIHC, MSA/BCVM, and BIHC/BCVM). The outputs of such graphical analysis were compared for the results from the usual quantification approach (the background subtraction method) and from the new formula (equation 5.2) proposed in this thesis (quantification relative to mean background value), in order to get an insight on the modifications introduced by this procedure.

Prior to the Altman-Bland analysis method, for the first step of the fluorescence signal quantification analysis, signal intensities quantified from each of the three pairs are plotted versus each others. This results in three graphs, whose linear relationship between pairs are assessed by calculating the correlation coefficients. This is done to confirm the expected linear relationship between the

results provided by each pair of algorithm, as they are subjectively designed to provide close signal quantification results.

The Altman-Bland procedure does assume that the differences between paired measurements are Normally distributed around their mean. In their paper, they stress that such a situation may not occur when dealing with real datasets. Thus, the mean difference (bias) may not be independent from the mean intensity and may vary as the mean intensity varies.

4.2.7. Modified Bland-Altman analysis of fluorescence quantification approaches

In order to analyse the behaviour of the bias (the mean difference in a given interval of mean intensities) as a function of the mean intensity, first the usual Bland-Altman plot was computed, with the mean difference and the 95% confidence agreement boundaries (as described in section 4.2.6). It is therefore expected that 95% of the points will be inside this interval. The proportion of points outside this interval was calculated for both the results obtained from the usual quantification method (mean background intensity subtracted to mean intensity in ROI) and the proposed quantification method (mean signal intensity normalised in respect to the mean background intensity). Then, the whole range of mean intensities was divided into smaller intervals, each containing 15 points (i.e. the intervals do not necessarily have same width). This is done to reduce locally the effect of the bias by assuming that a smaller set of points is be more likely to be Normally distributed around their mean, on a local basis. Such an

approach allows to graphically visualise the evolution of the bias over the range of mean intensities, and to assess its behaviour in respect to the estimated 95% agreement interval calculated over the whole population of measurements.

A close visual inspection of the regression graphs obtained using the usual quantification method (background subtraction, figure 4.8, left column) tends to indicate that, despite good general linear relationship between measures, the plot does not follow an exact straight line. This is particularly observable in the region of low mean intensities. This consequently suggests that the bias is not independent from the mean intensity.

The same analysis is applied to a new fluorescence quantification approach further described in section 5.1.1. It is presented here in order to be compared with the usual background subtraction method and in order to validate this new approach. Similar analysis of the regression graphs obtained with the proposed quantification method (relative to mean background value, figure 4.8, right column) indicates generally improved correlation coefficients between quantitative measures from segmentation performed by pairs of independent methods.

Figure 4.9 shows the Bland-Altman plot for comparing the fluorescence measures obtained from three global thresholding algorithms, using both the usual and proposed quantification methods. On the graphs, the actual percentage of points falling outside the 95% confidence interval are indicated (one expects 5% or less points to be outside this interval). On the left hand side are the plots obtained from the classical quantification method. It is noted that more than 5% of the points fall outside the 95% confidence interval, for all three combinations

of methods. On the right hand side are the plots obtained from the proposed quantification method. Except for the BIHC/MSA comparison, where 8.08% of the points fall outside the 95% confidence interval, the other graphs indicate that less than 5% of the points now fall inside the confidence interval. This reflects an improved agreement of the fluorescence measurements provided by the BIHC/BCVM and MSA/BCVM.

The second step of the analysis is to look at the behaviour of the bias in respect to the mean intensity. The presence of a bias *independent* from the mean intensity would be indicated by the points being mostly distributed horizontally either above or below the intensity difference zero. The presence of a bias *dependent* on the mean intensity would be indicated by the points not being distributed in parallel to the horizontal zero axis (i.e. a trend would appear).

For the usual quantification method, all three plots indicate the presence of biases. This is not surprising as the existence of biases was already noted from the regression analysis presented in figure 4.5. Nevertheless, it appears here that the biases are not independent from the mean intensity. Indeed, for low mean intensities, the points are globally tightly clustered around the value zero for each combination of method. For the BIHC/BCVM Bland-Altman plot, the means of the sub-intervals defined along the horizontal axis keep close to zeros, with slightly higher deviations to a positive bias in the mean intensity range of [7, 12]. However, for the BIHC/MSA and MSA/BCVM Bland-Altman plots, the points at higher x values (10 and above) clearly show a tendency to have higher differences (they "move" towards the upper limit of the 95% CI). For these two combinations of algorithms, the estimated 95% CI of the sub-intervals

increase at higher mean intensities, while this phenomenon occurs around a mean intensity range of [5, 10] for the BIHC/BCVM combination.

For the proposed quantification method, the dependence of the biases to the mean intensity has been reduced. Overall, the points are more clustered around the zero value, and the general aspect of the plots is quite similar. Still remains the tendency of the points to be more dispersed for higher mean intensities, but this tendency now has a similar aspect for all three algorithm combinations. Indeed, the 95% CI of the sub-intervals at lower mean intensities are clearly smaller than those at higher mean intensities (above 0.4). It is worth noting that most of the fluorescence measurements belong in the lower mean intensity range (approximately the range [0, 7] for the usual quantification method, and [0, 0.35] for the proposed method). Therefore the proposed quantification formula seems to transform the fluorescence measurements in such a way that their distributions are more similar to each other. Also taking into account that most points found in the regions of lower mean intensities have smaller dispersion around zero, this explains why the correlation coefficients are improved using the proposed method.
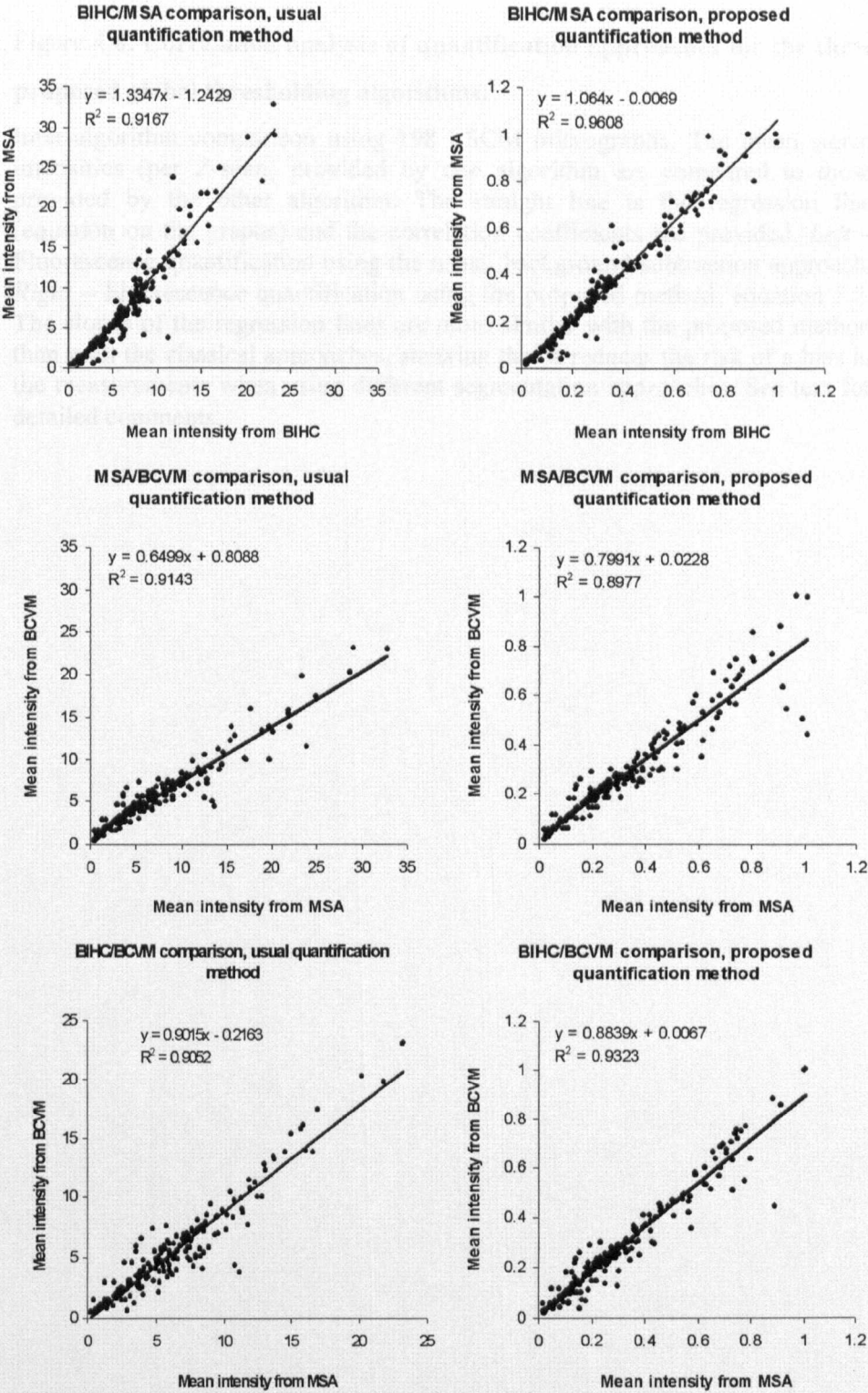
Figure 4.8: **Correlation analysis of quantification approaches for the three proposed global thresholding algorithms.** *(See next page for legend)*.

Figure 4.8: **Correlation analysis of quantification approaches for the three proposed global thresholding algorithms.**

Inter-algorithm comparison using 198 LSCM micrographs. The mean signal intensities (per Z-scan) provided by one algorithm are compared to those provided by the other algorithm. The straight line is the regression line (equation on the graphs) and the correlation coefficients are provided. *Left* – Fluorescence quantification using the usual, background subtraction approach. *Right* – Fluorescence quantification using the proposed method, equation 5.2. The slopes of the regression lines are more similar with the proposed method than with the classical approaches, showing that it reduces the risk of a bias in the measurements when using different segmentation approaches. See text for detailed comments.
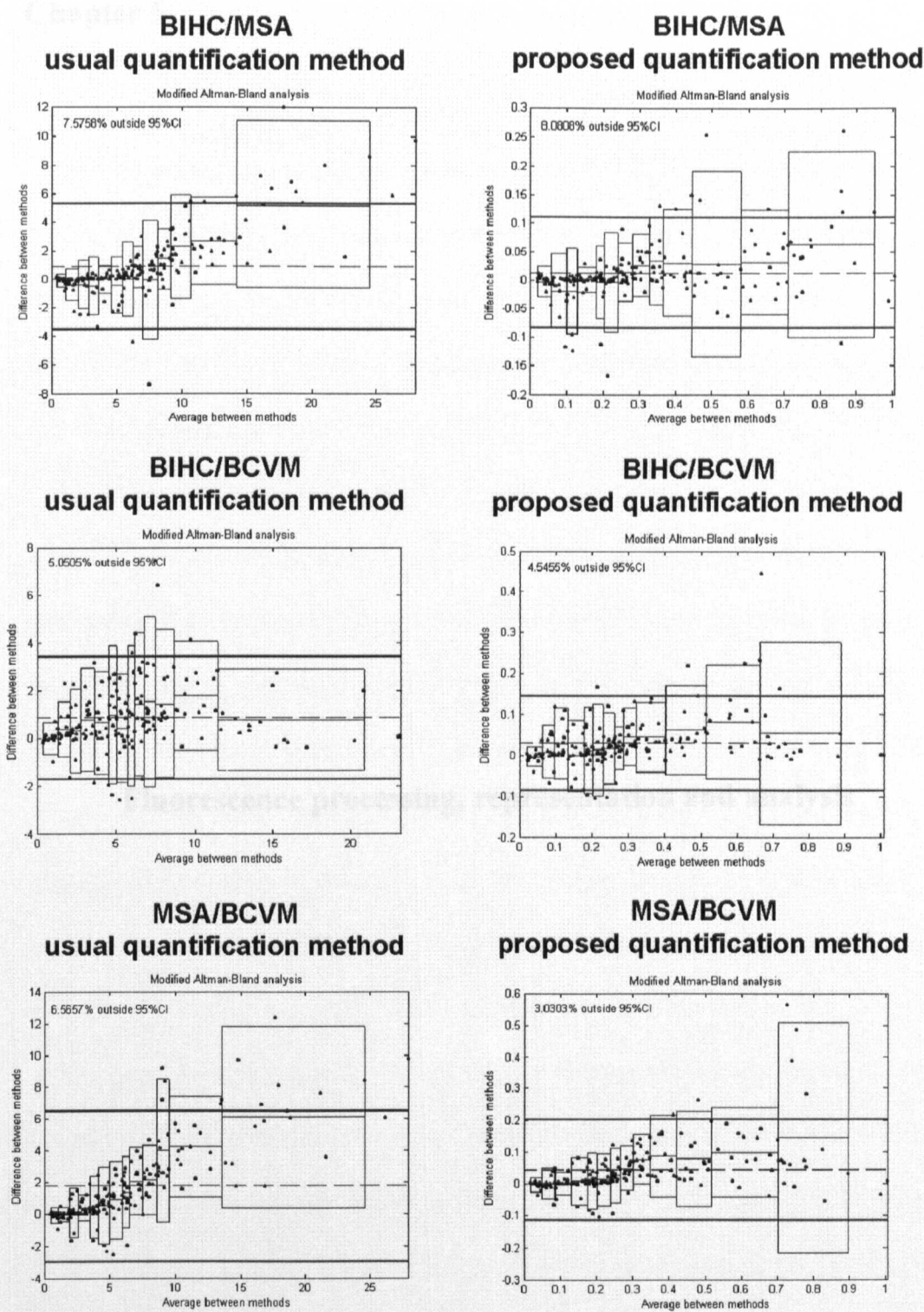
Figure 4.9: **Bland–Altman analysis of plots displayed in fig. 4.8.**

The dashed line indicates the position of the mean difference (bias). The solid lines indicate the limit of agreement, i.e. mean ± 1.96 SD. Refer to text for detailed comments.

Chapter 5

Fluorescence processing, representation and analysis

Chapter 5. Fluorescence processing, representation and analysis

5.1. Fluorescence signal processing and representation

Once the fluorescence signal is segmented from the LSCM image background, the measured intensities are used to characterise the antigen distribution. Such characterisation is both qualitative and quantitative. The quantitative analysis of the CD164 and CD34 epitopes distributions is described in chapter 6 and makes use of the Antigen Density Maps, or ADMs, developed in this chapter. Prior to the computation of the ADM, the measured signal has to be processed to compensate from the presence of the background which affects the quantification. Background subtraction is a common procedure used to remove the contribution of background intensities for a measured signal, which is also widely applied for LSCM measurements and is herein described and commented in the context of LSCM images. This approach assumes the noise to have an additive origin. It cannot be guaranteed that such an assumption is appropriate for LSCM images since the process that generates them is thought to follow signal-dependent Poisson statistics [Rooms et al, 2004]. Thus this approach is described in the following section, and another quantification formula is described which does not consider the noise to be additive but multiplicative. The consistency of the results obtained using both formulas on the measured signal segmented using the MSA, BIHC and BCVM algorithms and processed using either the usual background subtraction method or the new proposed method was assessed in section 4.2.7 using linear regression and Bland-Altman analysis. The background-corrected signal intensities obtained using the proposed quantification formula (equation 5.2, further detailed in next section) is used for quantitative analysis using the ADMs (chapter 6), and also for visualisation purposes by employing, along with the

ADMs, 3D reconstructions and 3D spherical graph models also described in this chapter.

5.1.1. Fluorescence signal quantification

Despite fluorescence measurements being now commonly published in quantitative or semi-quantitative studies, achieving a correct quantification is still a huge challenge. Indeed, there are many factors which can drastically impair such measurements. Actually, over 30 parameters to be controlled were reported, all of which potentially having "biasing" effects. These encompass the chemical properties of the medium used in the experiment, which may be responsible for different spectral responses of the same fluorochrome subjected to various pH and molecular environment. Fluorescence bleaching after repeated application of the laser at the same location also introduces a well known bias in fluorescence measurements, and corrective image processing methods are available to correct this effect [Boutet de Monvel *et al*, 2001; Umesh Adiga and Chaudhuri, 2001].

This problem was overcome by imaging each cell or group of cells only once for a particular wavelength, so the bleaching of the fluorochrome was minimal and similar for each cell imaged. The thickness of the sample through which the light has to travel is also a common problem, as light finds harder to penetrate deep in the sample. In the experiments analysed in this thesis, the depth imaged was only 12 to 18 μm maximum, which is very thin in confocal microscopy in comparison to the commonly achieved 100 μm or above. Furthermore, the cells were not imaged in dense cellular clusters or when cells were overlapping. Physical considerations can significantly affect fluorescence quantification. Indeed, LSCM

experiments can easily span over several hours, and some users have reported important fluctuations of the laser power over the course of the experiments [Swedlow *et al*, 2002]. As the measured fluorescence intensity is directly related to the laser power, laser instability is a real, hard to control problem. Other parameters related to the preparation of the sample, the imaging parameters such as brightness and contrast adjustments altogether contribute to make signal quantification a difficult task.

Having control over all these parameters is impossible in practice. Therefore, scientists use calibration curves to relate the measured fluorescence to known concentrations of the fluorochrome [Gaigalas *et al*, 2001]. Usually, fluorochrome-coated micro-beads are used to construct such curves. They are supplied by manufacturers with various sizes and surface fluorochrome concentrations. It is assumed that fluorochromes give the same spectral responses with coated micro-beads or real samples, which may not be true. However, since such calibration curves are not always available (as in this work), the mean fluorescence of a control sample may be used as a baseline for quantification, and the levels of fluorescence in investigated samples would be either expressed as a fold increase (or decrease) of the control fluorescence, or given in arbitrary units as the (absolute) excess of fluorescence above the control baseline. This latest method is referred to as "background subtraction" method and is predominantly used in published fluorescence measurements. However, it works best when large samples containing many cells are imaged, as the inter-cell variability is compensated by the number of cells analysed. This was not possible in our cellular imaging experiments where unique or small number of HSPCs were imaged at a time. Furthermore, negative controls were performed on non immuno-reactive cells (i.e.

negative control for CD34 labelling performed on CD34^{Neg} cells), which expectedly produced no fluorescence at all.

When control data is not available, the simplest and widely used quantification approach consists in measuring the mean background fluorescence, and subtracting it from the mean signal fluorescence. This is equivalent to baseline correction. As no calibration is available, the fluorescence measurements are given in arbitrary units. In this process, it must be assumed that the noise has an additive nature, while it may be multiplicative, or combinations of different natures. The noise generated by the Poisson process taking place during LSCM image acquisition is signal-dependent [Rooms et al, 2004], therefore not additive, so background subtraction may not be the most appropriate approach.

In this thesis, a new method is proposed which could improve the robustness of fluorescence quantification, when the calibration curve is not available. Instead of using the traditional background subtraction method described by

$$I_{FS} = \overline{I_S} - \overline{I_B} \qquad\qquad 5.1$$

where I_{FS} represents the quantified fluorescence of signal intensities, $\overline{I_S}$ and $\overline{I_B}$ respectively being the mean signal intensity in the segmented signal regions and the mean background intensity. Instead, the simple following ratio was introduced

$$I_N = \frac{\overline{I_S}}{\overline{I_B}} - 1 \qquad\qquad 5.2$$

which is equivalent to

$$I_N = \frac{\overline{I_{FS}}}{\overline{I_B}}. \qquad\qquad 5.3$$

I_N represents the mean fluorescence signal after subtraction of mean background value divided by the mean background value. It expresses the excess intensity above the background level arising from the fluorescence as a multiple of the mean background intensity (minus 1, so mean signal intensities close to background levels are close to 0 according to equation 5.2). As will be demonstrated, this simple modification of the usual approach tends to improve the correlation between fluorescence signals quantified using independent methods, better compensating for the unpredictable fluctuations of the background along the z-axis. It is noticed that, because I_N is a ratio of intensities, it is expressed without units by definition.

5.1.2. Computation of the Antigen Density Maps (ADMs)

An important aspect of this thesis was to address the readability of LSCM datasets. As mentioned, LSCM images are used both for quantification purposes and visualisation of 3D antigen distributions. 3D visualisation techniques can use maximum intensity projection along a specified direction, or they can involve approaches borrowed from the field of Computer Graphics, such as isosurfaces, isocontours and point clouds. These later techniques aim to represent the distributions in an artificial 3D space projected on the 2D surface of the computer screen.

The research project of this thesis was to study the distributions of antigens in the membrane of single cells. HSC are naturally very spherical (the so named "blast" morphology) under physiological conditions (and also here, under carefully adjusted experimental conditions), and their membrane antigen distributions were

imaged at high resolution, which allowed to observe details of the distribution at the subcellular level. As a consequence, it was decided to describe these distributions by computing 2D maps of the antigen densities around the cells according to a spherical model of the cell shape. This new approach can be seen as cutting (from northern to southern pole) and unfolding the 3D cell membrane on a 2D surface, such as a map of the world flattened on a school wall. Since no calibration of the fluorescence is available, the expression "antigen density" is used in throughout this thesis instead of "antigen concentration". Therefore the proposed maps were named Antigen Density Maps, or ADMs.

The ADMs are computed from the z-series LSCM images for which the signal was quantified according to equation 5.2 after segmentation using the MSA algorithm. First, a projection of the fluorescence along the z-axis of the z-series is
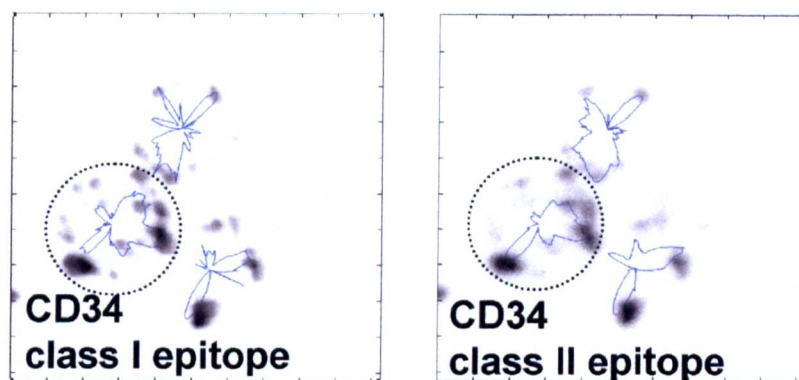
Figure 5.1: **Z projection of LSCM Z-series.**

Intensity projections are used to select the region containing the fluorescence of the cell of interest (dashed circle in this example). Here, three HSPCs are clustered together. The blue curves centred on the cell centroids show the antigen density in polar coordinates. The signal from the whole Z series was used for computing these projections. However, the antigen density curves (in blue) were calculated from the middle equatorial Z sections, and therefore do not directly map to the most intense regions of the images.

calculated and displayed (as shown in figure 5.1). The projected image is used to allow the users to manually select the image regions in which they wish to perform the analysis, i.e. a region containing a cell image. This region is referred to as the Region Of Interest, or ROI. This step is the only manual step involved in the whole analysis. However, it was deliberately kept manual in order to provide some freedom to the users about which cell they wish to analyse. Furthermore, under certain conditions, ROI detection may easily be automated in the case of single cell experiments. In order to assist the manual selection of the ROI, the DIC image taken at the middle section of the cell can also be used to correctly identify the position of the cell membrane. For this purpose, landmark points are manually placed around the cell boundary to exclude possible surrounding fluorescent debris and reduce quantification bias. The centroid of the ROI is then calculated.

The computation of the ADM is based on the equal circular segmentation of the ROI by small angular sections, all centred on the ROI centroid. An angular resolution of approximately 2.24° is selected since it allows the detection of smallest clusters correctly at the magnification used in this study. Therefore, the ROI is divided into 161 angular sections. The first section corresponds to pixels with angular coordinate in the range [0°, 2.24°[, and the 161th one to pixels in the range [357.76°, 360°[.

For the next step, the MultiScale Algorithm (MSA) was used for signal segmentation due to its speed and good segmentation performances in close agreement with human operator results. In each angular section, the pixel intensities were processed in respect to the mean background intensity, according to equation 5.2, and they are summed in each section. The same process is repeated over all z-sections. These mean angular intensities are computed at each

z-position and stored in a matrix $M_{z,\theta}$, where z and θ are respectively the z level of

the optical section and the angular index in the corresponding optical section. This

matrix $M_{z,\theta}$ constitutes the ADM (figure 5.2).
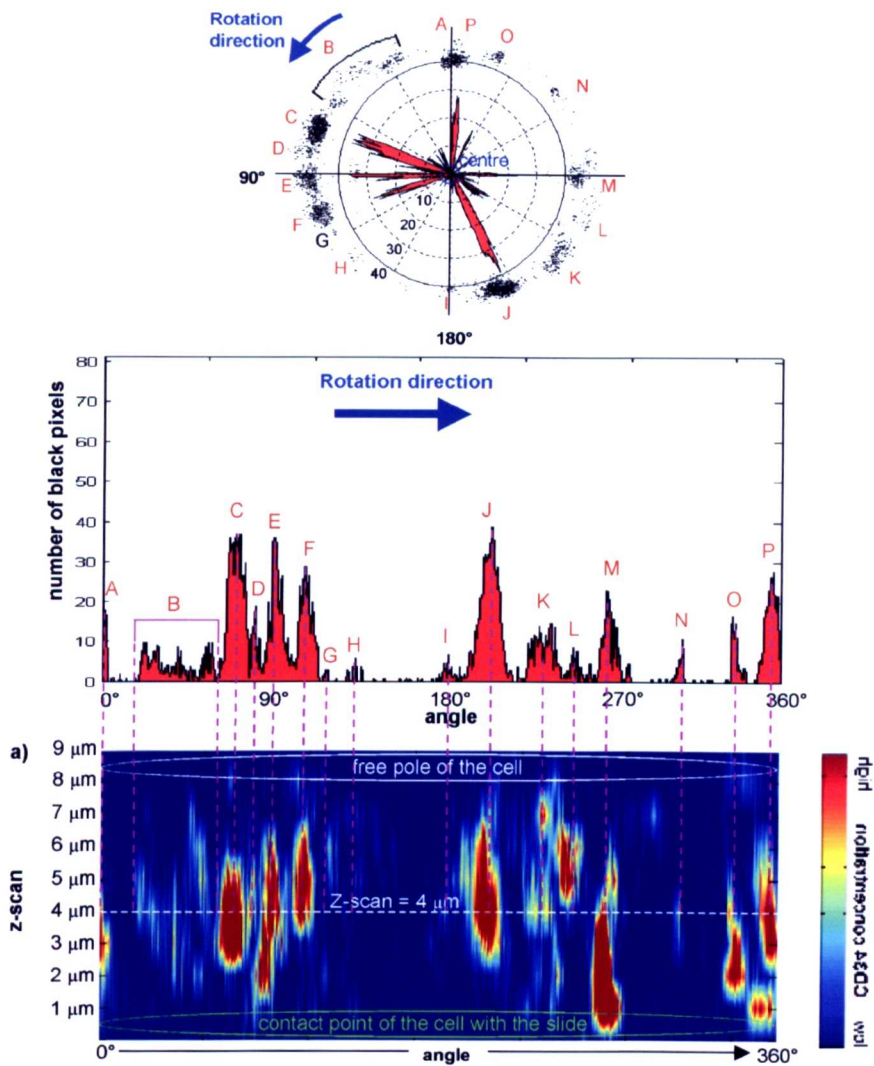
Figure 5.2: **Principles for computation of the Antigen Density Map (ADM).**

Top – For each Z-scan of a series, an orthonormal set of axis is centred to the cell's centroid in the 2D plane. The segmented fluorescence is summed in small angular sections, anti-clockwise. *Middle and bottom* – These measurements are projected in a matrix whose rows indicate the Z-level (4 μm from contact point in this example), and columns indicate the angular section. Repeated to all Z-scans, this constructs the ADM which is stored in computer and on CD prior to further analysis. The ADM presented here has undergone bilinear interpolation between Z levels for visualisation purpose.

Since the resolution in the x-y plane is much better than the resolution along the z-axis, bilinear interpolation is used for some applications of the ADM (ADM visualisation, sphere models described below), as it allows reasonable extrapolation of the signal intensities between each z-slice (as illustrated in figure 5.3).

Unprocessed ADM

Bilinear interpolation of ADM
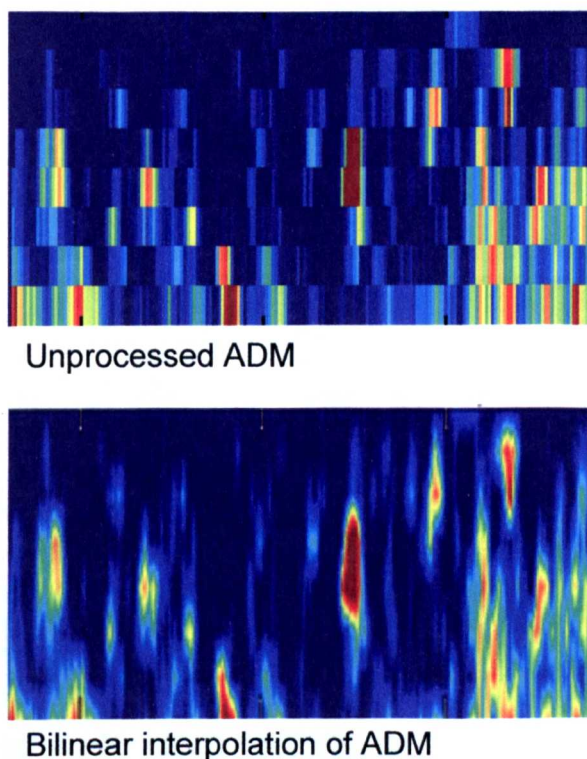
Figure 5.3: **Bilinear interpolation of ADM.**

Top – Original ADM. The readability of such a matrix is not very clear. Therefore bilinear interpolation is used – *bottom* – between the Z levels. It allows to link corresponding antigen structures between Z scans. Interpolated ADM are used for ADM visualisation and sphere models.
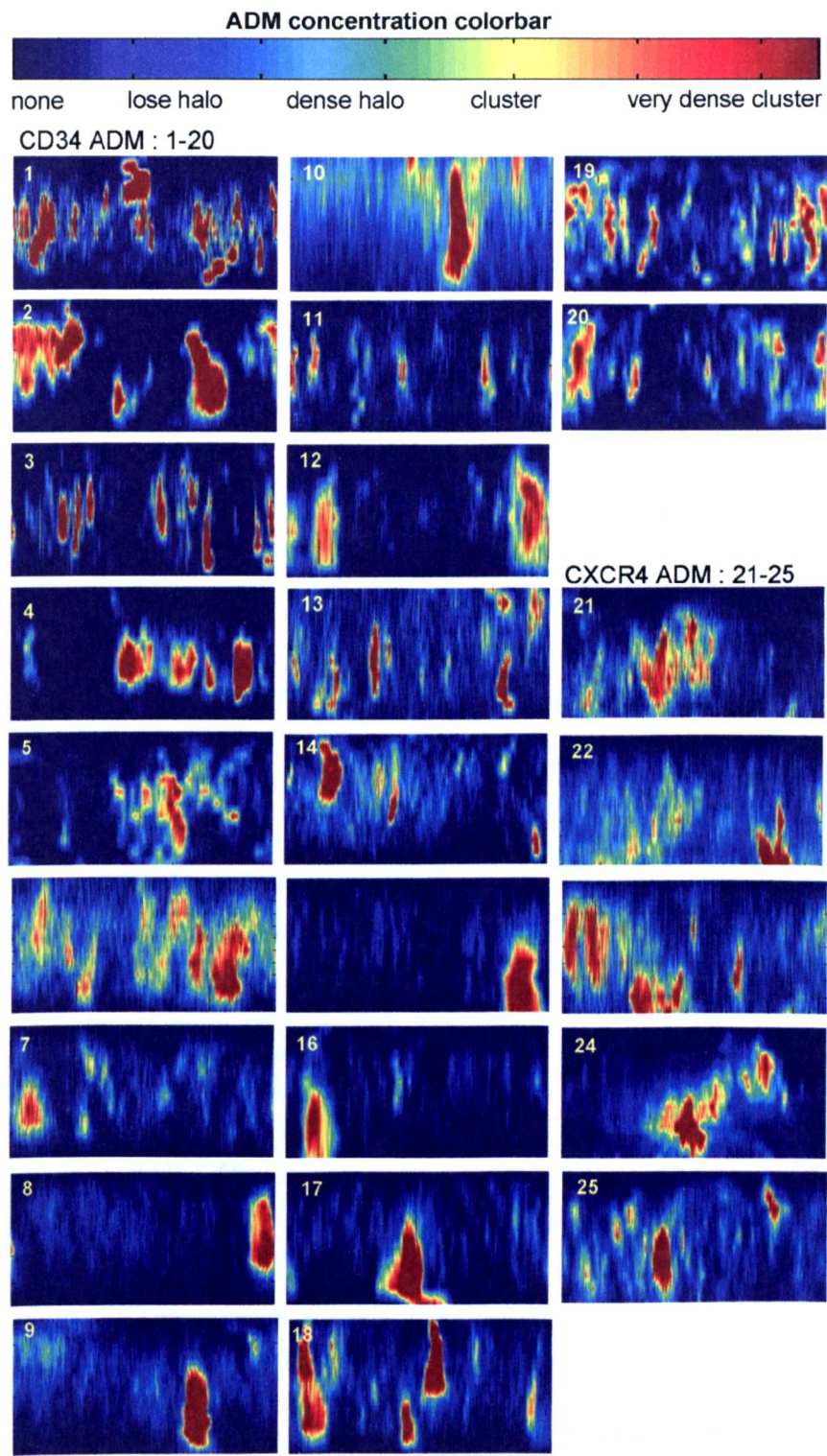
Figure 5.4: **A sample collection of CD34 and CXCR4 ADMs.**

This figure illustrates the application of ADM for comparing membrane antigen distributions. Since the ADMs are 2D projections of 3D distributions, they make easier to assess distribution similarities or differences. It can be observed that these cells can express a few big antigen clusters (e.g. 8 and 12), many smaller clusters in a halo (e.g. 1, 19), or combinations of each to various extents (e.g. 2, 6, 25).
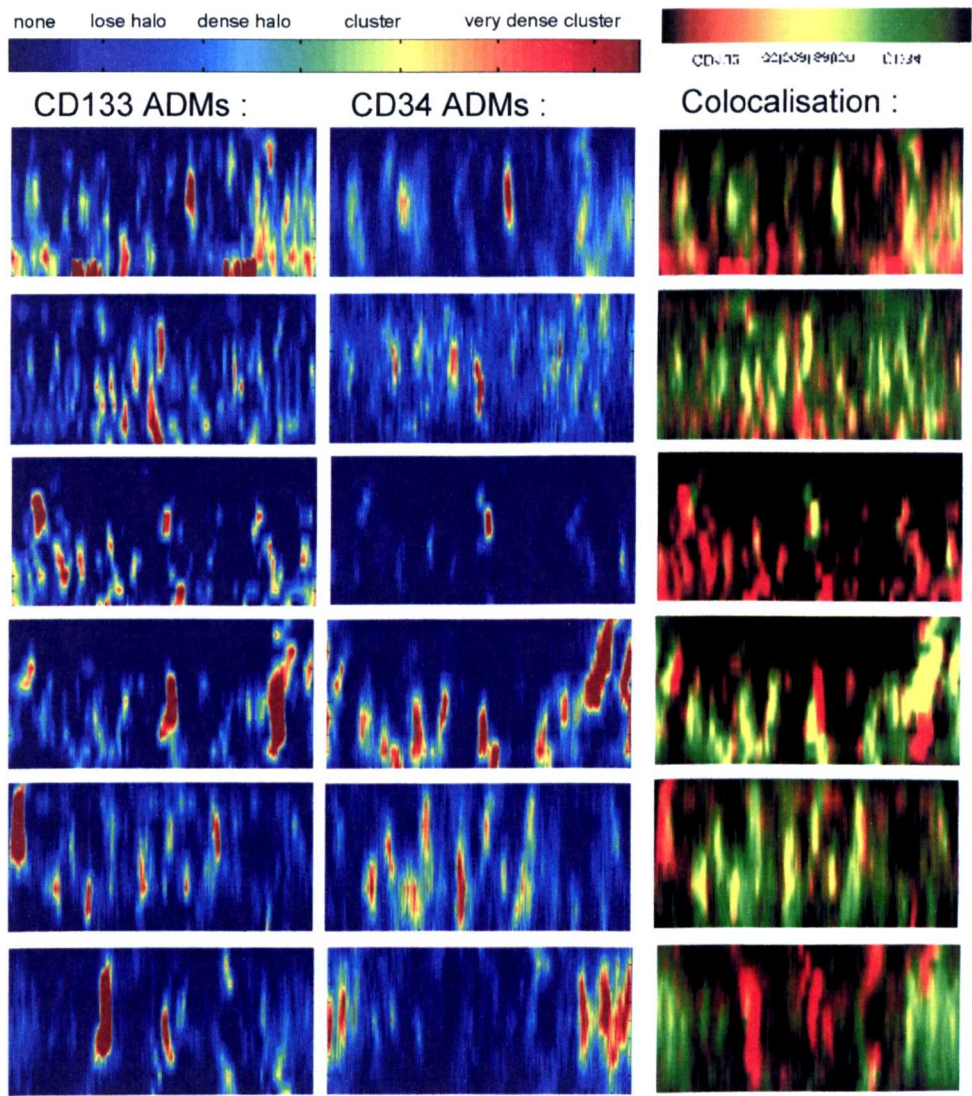
Figure 5.5: **Using ADMs for visual analysis of colocalisation.**

Similarly to merging 2D confocal images to show colocalised regions, ADMs can be merged for the same purpose. This is applied here to CD34 and CD133 antigens in the membrane of CD34$^+$CD133$^+$ cells. The fourth set of ADM is particularly interesting. Differential colocalisation of CD34 and CD133 is observed. The elongated CD133 cluster on the far right of the CD133 ADM is found to colocalise very well with a slightly shorter CD34 cluster in the north hemisphere of the cell, but not in the south. The other clusters of this ADM colocalise well with each other, but in a differential way.

5.1.3. Visualisation: fast 3D reconstruction of fluorescence distributions

A fast 3D reconstruction was developed to allow the user to have a quick view of the actual 3D structure of the antigen distribution. It works by using the information contained in the segmented background areas in LSCM images, but other approaches can be used to process the segmented signal in order to obtain such 3D reconstructions.

The histogram of the original, non-smoothed background regions (segmented using the fluorescence mask provided by the segmentation algorithm) is computed. From this curve, the cumulative histogram is calculated and normalised so its maximum value (in the last bin of the histogram) equals 1. Then a threshold is arbitrarily chosen along the y-axis (the axis of normalised frequencies). The intensity (on the x-axis) for which the cumulative histogram first goes above 0.9 is chosen to be a binarisation threshold to be applied to the segmented fluorescence regions, since it was observed to provide appropriate results.

This approach takes advantage of the strong noisy nature of the LSCM digital micrographs. Binarising the original signal regions by applying a threshold proportional to the cumulated background (noise) histogram generates 2D fluorescence distributions similar to point clouds. The more intense the fluorescent cluster, the denser its point cloud representation in binary form (i.e. the more points contained in the binary 2D cloud representing this particular 2D cluster). The threshold (between 0 and 1) has to be selected subjectively according to the visual aspect of the resultant binarised signal. The same process is applied to all LSCM images in the Z-series, and the collection of the resultant binary pixels are plotted in a 3D virtual space, each z-plane at the corresponding z-level. This 3D plot represents the reconstructed antigen distribution on the cell

membrane which can be arbitrarily rotated at any angle. In order to make such reconstructions more visually understandable, a colour map was applied to the clouds of points when appropriate. In this case, the colour of each point reflects the local mean intensity, or density, in its small neighbourhood (figure 5.11).
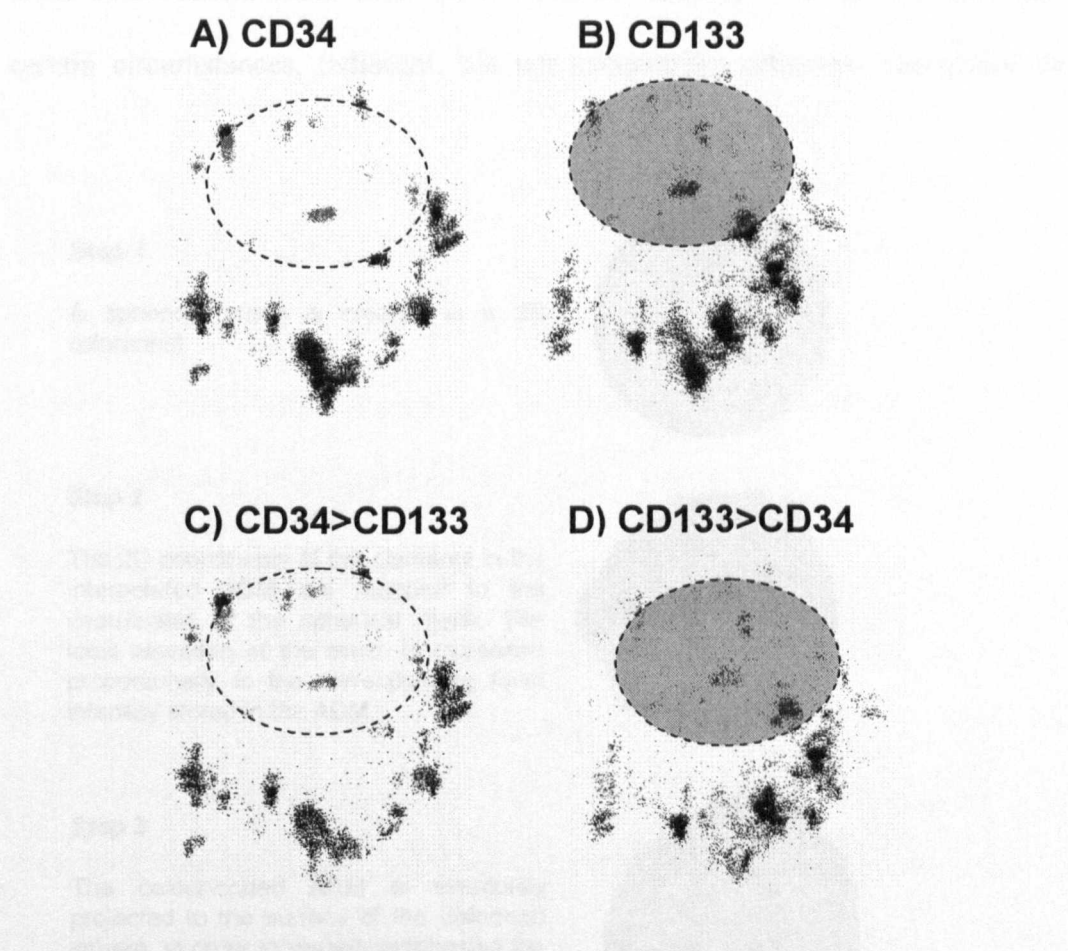
Figure 5.6: Example of fast 3D reconstruction of LSCM datasets.

Top – 3D reconstruction of the spatial CD34 and CD133 distributions in the cell membrane. In practice, the reconstruction can be rotated and viewed under any angle, facilitating its interpretation. The dashed circles indicate the contact area of the cell with the slide (the cells are inverted for visualisation purposes). Some similarities (due to colocalisation) and dissimilarities are noticeable. *Bottom* – Only the pixels following the relationship expressed by (**C**) or (**D**) were reconstructed in order to better show the differences between the top reconstructions.

5.1.4. Visualisation: Three-dimensional graphs – sphere models

The Antigen Density Maps (ADM) are a convenient and meaningful way of presenting the antigen distribution in the cell membrane of isolated cells. However, it was visually difficult to assess the three-dimensional relationship of these distributions when cells were clustered. Indeed, it is known that under certain circumstances, (adjacent, but not necessarily) cells may reorganise the

Step 1

A spherical mesh is created in a 3D referential.
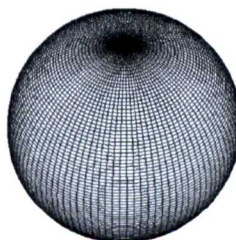
Step 2

The 2D coordinates of the elements in the interpolated ADM are mapped to the coordinates of the spherical mesh. The local elevation of the mesh is increased proportionally to the corresponding local intensity stored in the ADM.

Step 3

The colour-coded ADM is eventually projected to the surface of the deformed sphere, in order to visually emphasise the position of the fluorescent clusters. The colour map can be changed.

Figure 5.7: **Creating sphere models.**

Generation of a sphere model, which is a projection of the antigen distribution onto a sphere assumed to simplify the actual cell shape. First a sphere is created, then is surface is locally displaced proportionally to the antigen density stored in the ADM at the corresponding location. Finally, the ADM is mapped onto this spherical graph. Such sphere models are convenient to assess the spatial distribution of membrane antigens in adjacent and clustered cells.

spatial distribution of their membrane antigen in order to perform biological functions. For instance, such a phenomenon occurs during cell polarisation events, e.g. formation of the immunological synapse at the contact point between T-cells and Antigen Presenting Cells (APCs), where some particular antigens (MHC-peptide, ICAM-1, TCR, LFA-1) are recruited to allow the exchange of information between cells. Cell migration also requires antigen polarisation. Indeed, f-actin and chemosensory receptors are recruited at the leading edge of the cell while myosin, CD43, CD44, PSGL1, ICAM-1, ICAM-2, ICAM-3 are all recruited at the opposite side of the cell, in the uropod (or trailing edge) [Christensen *et al*, 2004; de Boer *et al*, 2002; Dravid and Rao, 2002; Ferrero *et al*, 2003]. Since CD34 has been reported to have some influence on HSPC migration capabilities, we designed a useful visualisation method of the 3D membrane antigen distributions of clustered cells.

None of the LSCM images acquired in this thesis made use of a dye to locate the position of the membrane, and therefore the true location of the cell membrane is unknown. However, in order to get an estimate of the position of the membrane, 3D spheres were generated. The relative radii R of the clustered cells are estimated manually from the DIC image. The location of the centroid of each cell is estimated from the mask drawn manually around the cells during the ADM computation step. Finally, the radial coordinates $R(x,y,z)$ of the sphere are modified according to the antigen density values stored in the ADM. Therefore, higher values generate a high local increase of the sphere radius. Consequently, the sphere is deformed and peaks (representing antigen clusters) appear where a fluorescence cluster is present in the cell membrane (see figure 5.7). In this respect, the sphere model is a 3D graph of the ADM. The local antigen densities

are represented by the local increases of the sphere radius. Furthermore, the modified spheres are coloured according to the height of the peaks. Various colour maps may be employed by the user. The proposed sphere model was found to be a reasonable, simplistic assumption of HSC cell shape, as their blast morphology is described as compact and spherical.
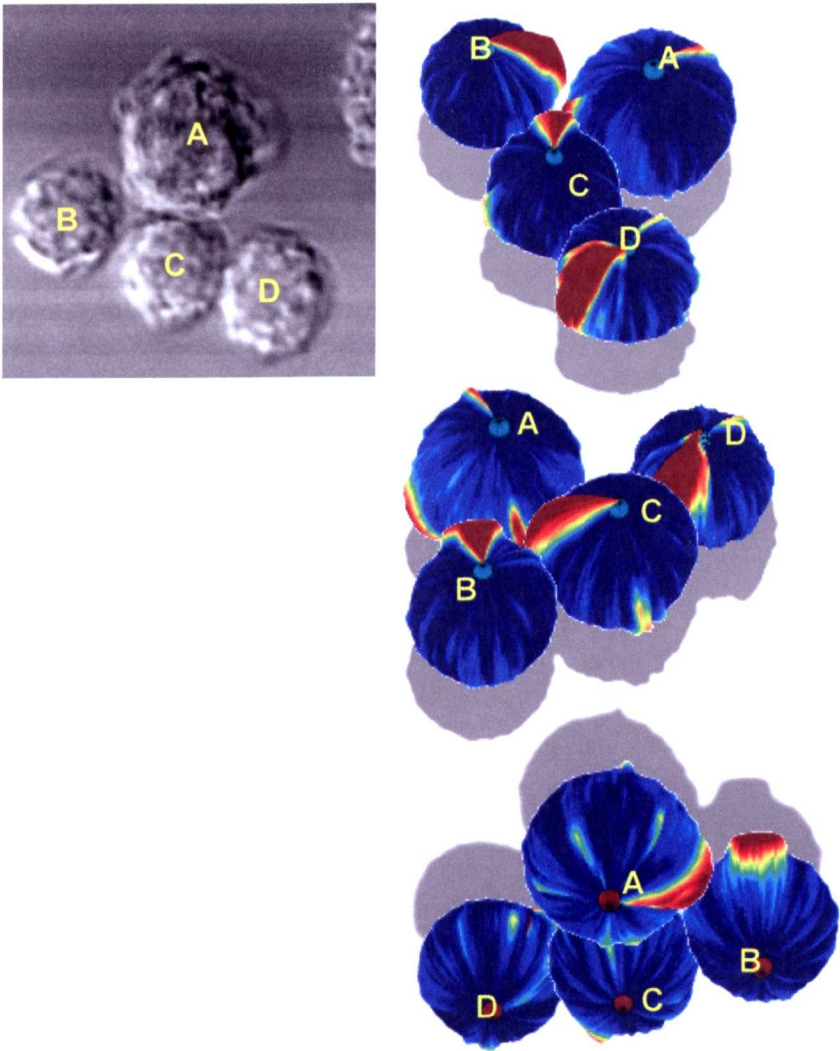
Figure 5.8: **Application of the sphere models.**

The spherical models where applied to the visualisation of CD34 distribution in clustered cells (DIC image on the left). The red and blue spheres at the poles respectively indicate the contact point with the slide and the free pole of the cells. Each cell contains one major CD34 cluster. In this example, many clusters seem directed towards adjacent cells.

Figure 5.9: **Application of the sphere models, another example.**

In this example, a different colour map of the figure was applied to sphere models. Both cells were imaged for the CD34 antigen. Each cell contains a collection of cluster with various sizes. However, there is a clear concentration of CD34 in the portions of cell membrane facing each other. This suggests that, at a distance, there may be some early signalling activity taking place between the cells which allows CD34 to be recruited in such areas. This is consolidated by the fact that this observation has been made consistently over the datasets and is shown here as a representative example.

5.2. Analysis of the segmented antigen distributions

In this chapter, new methods to quantitatively characterise membrane antigen distributions are described. These methods were applied either to LSCM images (section 5.2.3.2) or (for all other applications listed below) to the newly developed Antigen Density Maps (ADMs). Quantitative characterisation of distributions was performed using various specifically developed measurements, which were

- fluorescence quantification at various levels of observation

    o cellular level

    o Z-scan level

    o subcellular level

- polarity analysis of the antigen distributions in cell membranes

    o polarity in the XY plane

    o polarity along the Z axis

- antigen colocalisation analysis of dual labelled cells

    o colocalisation using equatorial LSCM Z section

    o colocalisation using the ADMs

- analysis of antigen clusters

    o clusters detection and proportion in southern hemisphere

o   cluster characterisation by distribution energy

Each of these measurements and their relevance are described in the following sections and their results are summarised in section 6.3 and further detailed in section 6.4.

5.2.1. Fluorescence quantification using ADMs and correlation of the measurements

Antigen quantification using equation 5.2 was explored at three different levels using the ADMs:

- the **cellular level** refers to the average fluorescence measured per whole ADM. The cellular level quantification was measured to compare the average fluorescence levels of either single or dual labelled cells. In chapter 6, the quantity measured at this observation level is referred to as "mean ADM intensity".

- the **Z-scan level** refers to the total amount of fluorescence measured on each different Z-section within the Z-series. The Z-scan level quantification was performed on dual labelled cells, in order to compare the fluorescence measurements from each Z-scan of paired Z-series using correlation analysis.

- the **subcellular level** refers to each single element/position of any ADM matrix. Quantitative analysis at the subcellular level was performed using

correlation analysis applied element by element on ADMs obtained from dual distributions.

Fluorescence quantification at the cellular level is similar to FACS analysis, where antigen amount per cell is quantified by measuring the average fluorescence for each cell of a population. Comparing quantification results at both the Z-scan and subcellular levels provides with allows to characterise the degree of colocalisation of two antigens in the cell membrane. Indeed (section 1.2.3.1.3) the Pearson's correlation coefficient is sometimes used to quantify the degree of colocalisation between distributions. If two antigenic distributions are spatially colocalised, the correlation coefficient is close to 1. Commonly, correlation is performed between paired equatorial LSCM cross sections. Since the ADMs contain quantitative fluorescence measurements of whole Z-series, more information is available and it is interesting to plot the correlation coefficients obtained at the Z-level versus those obtained at the subcellular level in order to compare different cell populations (see figure 7.3). However, other colocalisation analysis approaches are investigated in section 5.2.3.

The Pearson's correlation coefficient R between two vectors X and Y is defined as

$$R = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\sum_i (X_i - \bar{X})^2\right)\left(\sum_i (Y_i - \bar{Y})^2\right)}} \qquad 5.4$$

where here X is the vector containing the total measured fluorescence for the first antigen per Z cross-section i, and Y is the vector containing the total measured fluorescence per Z cross-section i for the second antigen. $\bar{X}$ and $\bar{Y}$ are the

arithmetic means of vectors X and Y. Similarly, the correlation coefficient between two ADMs A and B is defined as

$$R = \frac{\sum_{m}\sum_{n}(A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{\left(\sum_{m}\sum_{n}(A_{mn} - \overline{A})^2\right)\left(\sum_{m}\sum_{n}(B_{mn} - \overline{B})^2\right)}} \qquad 5.5$$

where m and n are the indexes of the elements in A and B, and $\overline{A}$ and $\overline{B}$ the means of the elements in A and B.

5.2.2. Polarisation analysis

Antigen polarisation in the cell membrane occurs when the antigen are predominantly found concentrated on one side of the cell [Madruga *et al*, 1997; van Buul et al, 2003]. As described earlier, this phenomenon is typically observed on migrating cells, where particular molecules involved in the process are expressed at the leading front of the cell membrane (e.g. f-actin), and other molecules are exclusively expressed at the opposite part of the membrane (e.g. myosin).
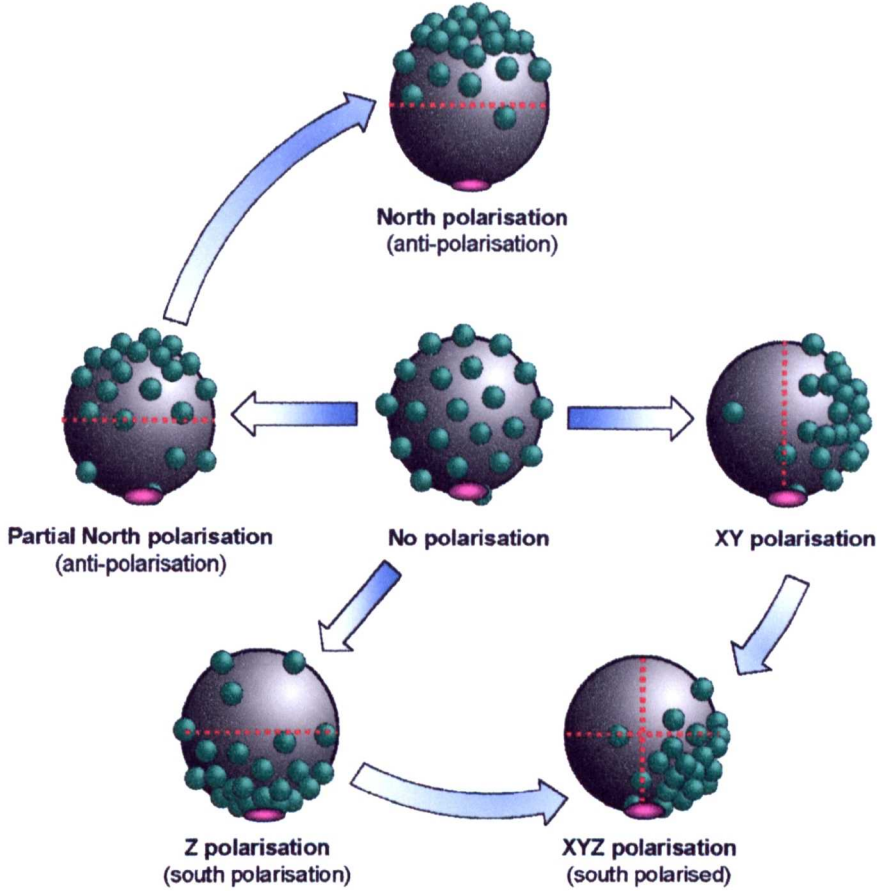
Figure 5.10: **Antigen polarisation in cell membrane.**

Different instances of antigen polarisation encountered in HSPCs. It is assumed that membrane antigens are initially randomly distributed around the cell. Under appropriate stimuli, these antigens concentrate on particular sides of the membrane. For instance, XY polarisation is observed on migrating cells or during cell-to-cell contact events. Z polarisation can be triggered by cell adhesion to an appropriate substrate. Combination of Z and XY polarisation generates XYZ polarisation. The purple ellipse indicates the contact point of the cell with the microscope slide. Z and XY polarisation in HSPC membrane are quantified in this thesis.

The functions of the CD34 and CD164 antigens are not known. Some experimental evidence suggests that CD34 has some involvement in cell migration. Our group previously hypothesised that polarity could also be triggered by the adhesion of a cell to an appropriate substrate. Indeed, it was observed that a proportion of HSPCs display instances of antigen polarisation along the z-axis when adhering to the gold coated surface of the microscope slide. Therefore, we

proposed to investigate antigen polarisation both in the XY-plane, where it could be triggered by cell migration or cell-to-cell contact events, and along the Z-axis, where it could be triggered by active recruitment towards the contact area of the cell with a substrate.
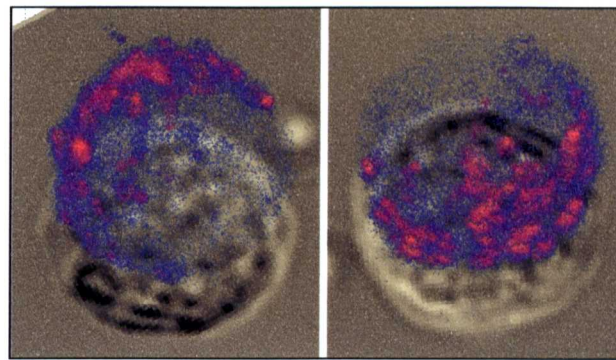
XY-polarised CD34 distribution
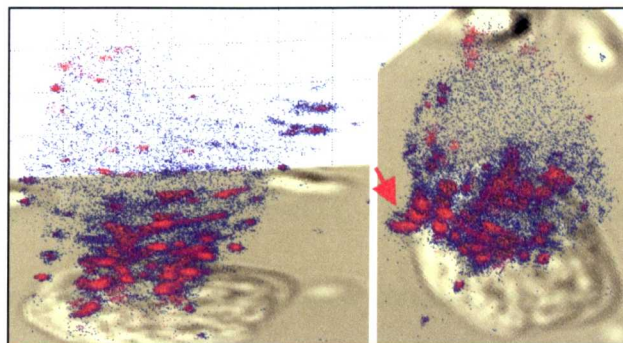
Z-polarised CD34 distribution

Figure 5.11: **Z and XY Polarised CD34 distributions.**

These 3D reconstructions use a range of colours to better represent the fluorescence intensity. Shades of red indicate higher antigen densities. DIC images of the cells are also shown below the reconstructions. *Top* – Example of XY polarised cell viewed at two opposite angles. CD34 is clearly concentrated on one side of the cell. *Bottom* – Example of Z polarised CD34 distribution. Clearly, most CD34 clusters are located towards the adhesion site. The arrow indicates a small CD34-containing pseudopod extending over 3 μm towards the slide, as would be expected if CD34 was actively involved in cell adhesion. In both cells, a CD34 halo can be observed all around the membrane. These datasets were collected on HSPC adhering to fibronectin, an adhesion molecule, for 30 minutes.

5.2.2.1. Polarity in the XY-plane

The sum of antigen density $S(\alpha)$ was calculated down the columns of the ADMs, which is a projection of the total amount of antigen in XY-plane, between the

polar angles $\alpha \in [0^{\circ}\text{-}360^{\circ}]$ (or $\alpha \in [0\text{-}2\pi]$ radians). For any angle α a ratio $R(\alpha)$ was

calculated such as

$$R(\alpha) = \frac{\sum\limits_{i=\alpha-\pi/2}^{\alpha+\pi/2} S(i)}{\sum\limits_{j=\alpha}^{\alpha+2\pi} S(j)} \qquad\qquad 5.6$$

where

$$\sum\limits_{j=\alpha}^{\alpha+2\pi} S(j) = \sum_{i}\sum_{j} M_{i,j} \, . \qquad\qquad 5.7$$

$M_{i,j}$ represents the ADM matrix. The denominator (equation 5.7) is constant and

represents the sum of the antigen densities stored in $M_{i,j}$. By definition, R can

only take values in the range [0; 1]. The polarity of the antigen distribution in the

x-y plane was defined as

$$\text{Polarity}_{x\text{-}y} = \frac{\sum R(\alpha) \in R | R(\alpha) \ge 0.5}{\sum R(\alpha) \in R | R(\alpha) < 0.5} \, . \qquad\qquad 5.8$$

This ratio is higher than 1 for XY polarised distributions. In such cases, $R(\alpha)$ has

a shape close to one period of the cosine function, i.e. it has one maximum and

one minimum and varies monotonically between them. In this situation, the

numerator of equation 5.8 represents a compact set of $R(\alpha)$ (all α for $R(\alpha) \ge 0.5$

are separated by one increment unit), and similar remark applies to the

denominator. When the antigen distribution is not polarised, $R(\alpha)$ fluctuates

several times around the value 0.5 and is not a compact set anymore. Measuring

polarity using equation 5.8 is therefore meaningless in this case, and instead the

Polarity XY = 1.1532

Polarity XY = 1.3489

Polarity XY = 1.5809
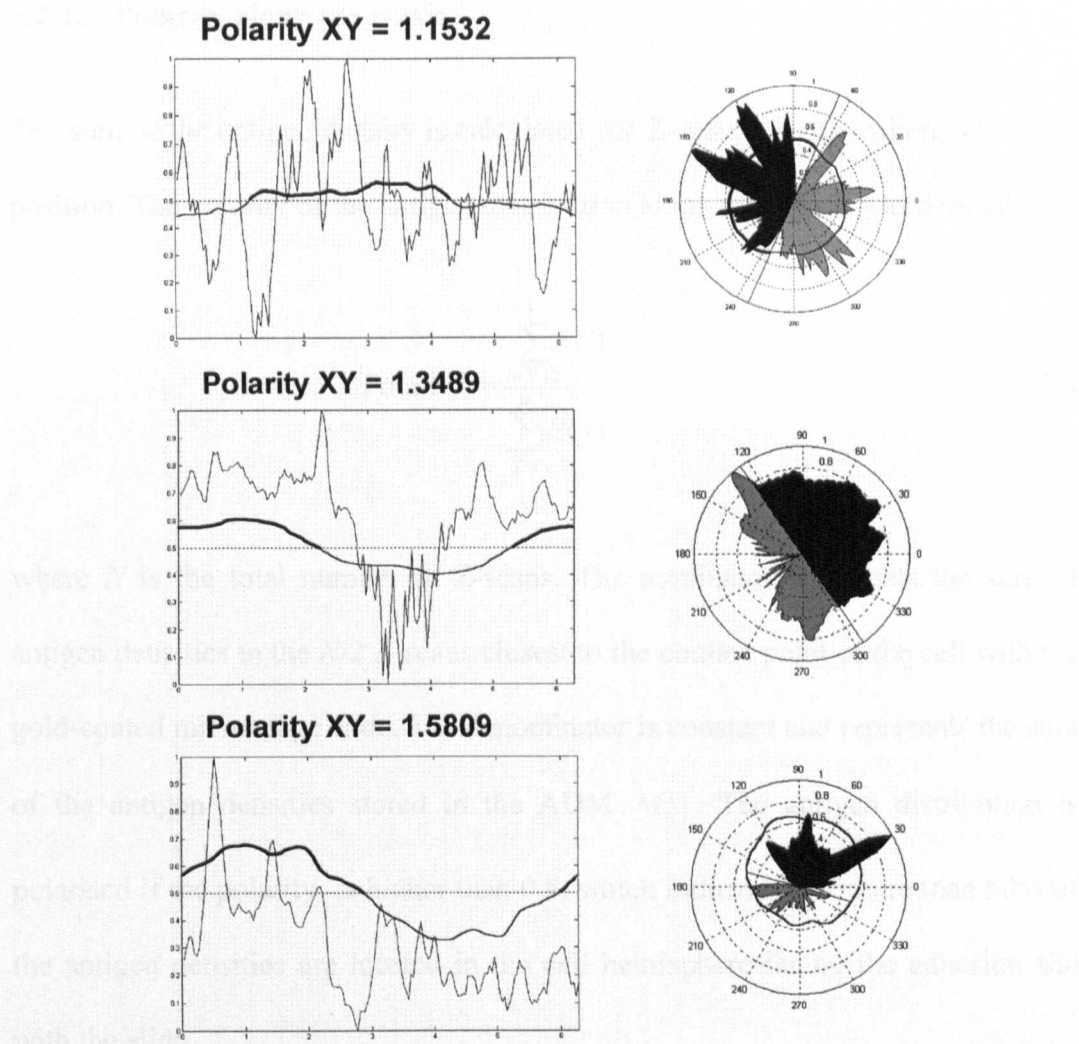
Figure 5.12: **Polarity graphs and quantification in the XY plane.**

Left – Normalised fluorescence intensity vs. angular coordinate. The curve fluctuating closely around the dashed line is $R(\alpha)$ in the text (equation 5.6), where α is the angle. The bold line indicates where $R(\alpha)$ is higher than 0.5 (dashed horizontal line). This curve is used to calculate the polarity XY according to equation 5.8. The value of the polarity is indicated. *Right* – Polarity graphs in the XY plane. Such graphs are similar to the top graph in figure 5.2.

distribution was attributed a XY-polarity score of 1 (non polarised). It was subjectively found that antigen distributions can be considered polarised in the XY-plane (i.e. the antigens are mostly expressed on one side of the cell) when the polarity is equal or higher than 1.5.

178

5.2.2.2. Polarity along the z-axis

The sum of the antigen density is calculated per Z-scan, i.e. $S(n)$ where n is the Z position. The polarity of the antigen distribution along the Z-axis was defined as

$$\text{Polarity}_z = \frac{\sum_{i=N/2}^{N} S(i)}{\sum_{j=1}^{N} S(j)} \qquad 5.9$$

where N is the total number of Z-scans. The nominator represents the sum of antigen densities in the $N/2$ Z-scans closest to the contact point of the cell with the gold-coated microscope slide. The denominator is constant and represents the sum of the antigen densities stored in the ADM $M_{i,j}$. The antigen distribution is polarised if the polarity$_z$ is higher than 0.5, which indicates that more than 50% of the antigen densities are located in the cell hemisphere facing the adhesion site with the slide.

5.2.3. Colocalisation analysis

The most common approaches to investigate colocalisation were described in section 1.2.3.1. A common approach consists, using paired LSCM images, in plotting a scatter plot for which each axis represents the intensity of each fluorochrome. The corresponding dual images are scanned on a pixel basis. Each point of the scatter plot represents a pair of dual pixels at a particular location in the dual LSCM images. The proportion of points belonging to the region of colocalisation of the graph is used to quantitatively characterise the distribution

colocalisation of the two antigens. The Pearson's correlation coefficient can also be used to quantify the extent of the linear relationship between both distributions. It must be stressed that such a correlation coefficient only measures the linear agreement between two variables. Therefore, two antigens may be spatially colocalised while having no linear relationship between their levels of expression. Furthermore, it is noticed that high correlation coefficients can be obtained if the mask used to identify the relevant fluorescent regions is set up to low and encompass lower intensity areas. Due to the unimodal nature of the image, a lower threshold used to separate the background to the signal generates a mask which encompasses proportionally wide insignificant areas for which low intensity pixel pairs participate to increase significantly the correlation coefficient. It is consequently important to make sure that the segmentation threshold is set up properly, as previously demonstrated in the chapter 4.

In the following sections, two different approaches for performing colocalisation analysis are described, one involving LSCM images and one involving the ADMs.

5.2.3.1. Image pair re-alignment

In most experimental datasets used in the present work, corresponding dual LSCM images were not strictly aligned or displayed a strong shift. Accordingly to observations from the literature, most datasets were affected by shifts in the XY-plane while few were affected by shifts in along the Z-axis. This latest case was observed for some datasets of the epitope experiments, due to some motion of

cells which did not stick strongly to the coverslip. Such sets were not used for colocalisation analysis.

A correlation method was implemented to correct the XY shift. Correlation was considered as a good approach since the dual signals to be realigned were often significantly colocalised. Also, the fluorescence signal was only present in the membrane of the cells for which the shape is similar whatever the distribution of the antigen under investigation. As observed, this shape similarity made the use of a correlation re-alignment method robust and reliable.

The method requires the manual selection of a square region encompassing the signal of one fluorochrome from the cell under investigation. The whole cell area is selected each time. This area is then artificially shifted at known position along the X- and Y-axis. For each position, the Pearson's correlation coefficient is calculated between the shifted and the corresponding non-shifted image of the pair. The shifts along the X- and Y- direction, respectively sx and sy, were chosen to belong to the ranges

$$\{sx, sy\} \in [-20;20] \text{ pixels.} \qquad 5.10$$

This interval was found wide enough to correct most encountered shifts. Higher shifts were considered as resulting from an experimental problem (i.e. cells unstuck from the coverslip), and the data sets were not used in colocalisation analysis. The correlation coefficients were stored in a 41x41 matrix M_r. The position of the maximum element in M_r was located as

$$(sx_{opt}, sy_{opt}) = \arg\max_{i,j}(M_r(i,j)). \qquad 5.11$$

Correspondingly, the appropriate image of the pair was shifted by the amount sx_{opt} and sy_{opt} prior to colocalisation analysis. It was observed that using the whole signal areas from the cell of interest made this approach very robust in comparison to selecting only predominant clusters which may not necessarily have an equivalent counterpart in the other LSCM image of the pair.
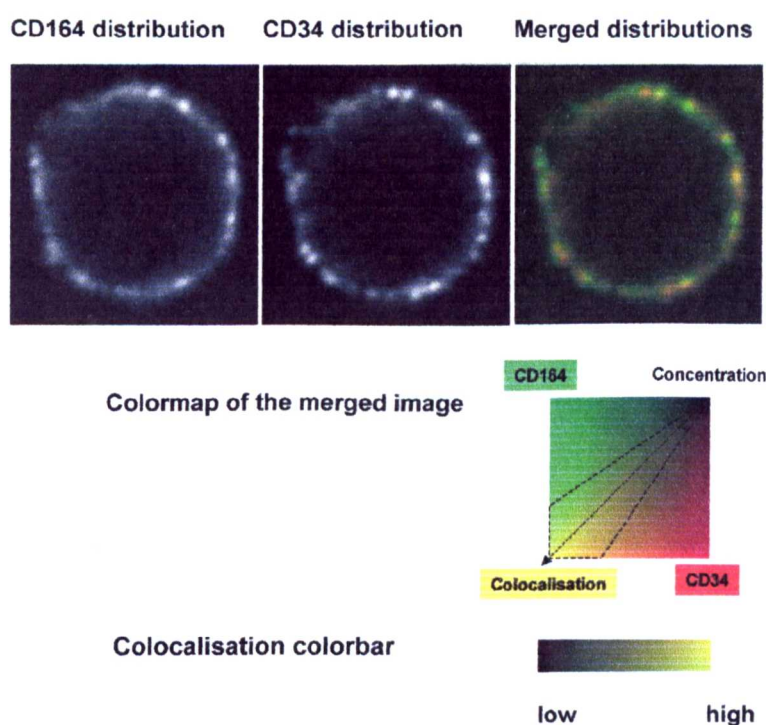
Figure 5.13: Illustration of merging the dual fluorescence after automatic re-alignment.

Two Z-sections, corresponding to CD164 and CD34 in this example, are re-aligned using the correlation method described in the text. Once the shift corrected, they are merged. Colour map of the merged image and colocalisation colour bar. Colocalisation is assumed to take place when the colour is close to the dashed diagonal (in colour map image), in the area enclosed by the dashed lines.

5.2.3.2. Colocalisation analysis using correlation method after watershed segmentation

Visual inspection of the merged fluorescence distributions after correlation-driven re-alignment indicated that, in many cases, the antigen clusters or the halos did not share similar degree of colocalisation despite them sharing identical locations (see figure 5.13). As a consequence of this observation, a way of segmenting the different fluorescent clusters and analysing their antigen content using correlation analysis and linear regression was investigated and described in this section.

Many approaches [Phee *et al*, 2001] use the middle image in the confocal *z*-stack for colocalisation analysis, since it usually provides the best contrast of the fluorescent signal. This is also true for the datasets used in this work, and the equatorial Z-section was used for this approach of colocalisation analysis.

As previously mentioned, watershed segmentation is particularly relevant for segmenting good quality LSCM fluorescence images (i.e. when the contrast is strong enough to discriminate between signal and background) [Lin *et al*, 2003]. However it is also very sensitive to noise, and original LSCM micrographs are naturally strongly affected by it. Consequently, one or several noise removal filtering steps were required prior to applying watershed segmentation to the fluorescent clusters. It was observed that various smoothing procedures using different filter parameters produced different segmentation results. It was decided to avoid heavy under- or over-segmentation of the fluorescence distributions. After experimenting with several approaches, the following filtering steps were consistently used as pre-processing steps on the investigated image pairs.

Firstly, the original LSCM images were filtered with a 11x11 median filter. This rank filter removed the speckle noise components and made the lower intensity background very uniform. Secondly, in order to smooth and attenuate the boundaries between the fluorescent clusters, weaker filters (described below) were applied iteratively. The number of iterations was set up to 3. At each iterative loop i, a filter F_i was designed as a 11x11 matrix whose elements were initially all set to 1, except the middle element which was empirically defined at each iteration i as

$$F_{6,6} = 5 + i, \qquad i \in \{1,2,3\}. \tag{5.12}$$

The filter F was then normalised as F^N by dividing each of its elements by their total sum, i.e.

$$F_i^N = \frac{Fi}{121 + 5 + i} = \frac{Fi}{126 + i}, \tag{5.13}$$

so

$$\sum_{x,y} F_i^N(x,y) = 1. \tag{5.14}$$

This smoothing procedure further diffused the remaining higher frequency components in the image after the preliminary application of the median filter.

Watershed segmentation was then applied to the smoothed images. A label matrix the same size as the original LSCM image was created, in which same numbers indicated pixels belonging to the same watershed. The watershed boundaries were set up to 0. Since watershed segmentation was also applied to background areas in

the image, the application of a mask to isolate the fluorescence signal was required. This signal mask was obtained using the Multi-Scale algorithm (MSA).

The fragmentation of the fluorescence signal according to watershed segmentation produced patches of signal areas. The careful set up of the parameters used in the smoothing process was chosen so that the biggest patches encompassed one or several antigen clusters. Therefore these segmented patches were available for further analysis. For this purpose, the focus was given to some quantities to characterise the colocalisation of two antigens under the preliminary hypothesis that it exists a linear relationship between their levels of expression, i.e.

$$[Ag_1] = a[Ag_2] + b, \qquad\qquad 5.15$$

where $[Ag_1]$ and $[Ag2]$ represent the antigen "concentrations" (as reflected by the fluorescence) of antigens 1 and 2, and a and b are constants (a is the slope of the straight line while b is the value of this line when $[Ag2]=0$). If so, the Pearson's correlation coefficient is a relevant quantity to compute, as a value close to 1 would indicate clear linear relationship between two antigens while a value around 0 would indicate the lack of linear correlation (a third extreme case would be a correlation coefficient close to −1, which would indicate that the antigen levels are anti-correlated). In addition, the determination of the value of a can help getting further insights on relative proportions between the expression levels of the two antigens. Indeed, a value of a close to 1 would indicate equality between the levels of each antigen, while lower or higher values (in association with a high correlation coefficient) provide information on the proportionality coefficient between both expression levels, or the value of the bias if fluorescence measures

are not calibrated. These quantities are able to differentiate between various degrees of colocalisation, but do not provide information on the actual fluorescence intensity. Therefore, for each watershed basin, the mean intensity of both antigens is also measured and stored. In this study, fluorescence calibration was not available so no use was made of the mean fluorescence per watershed, but this would be important information in other LSCM experiments where calibration is known.
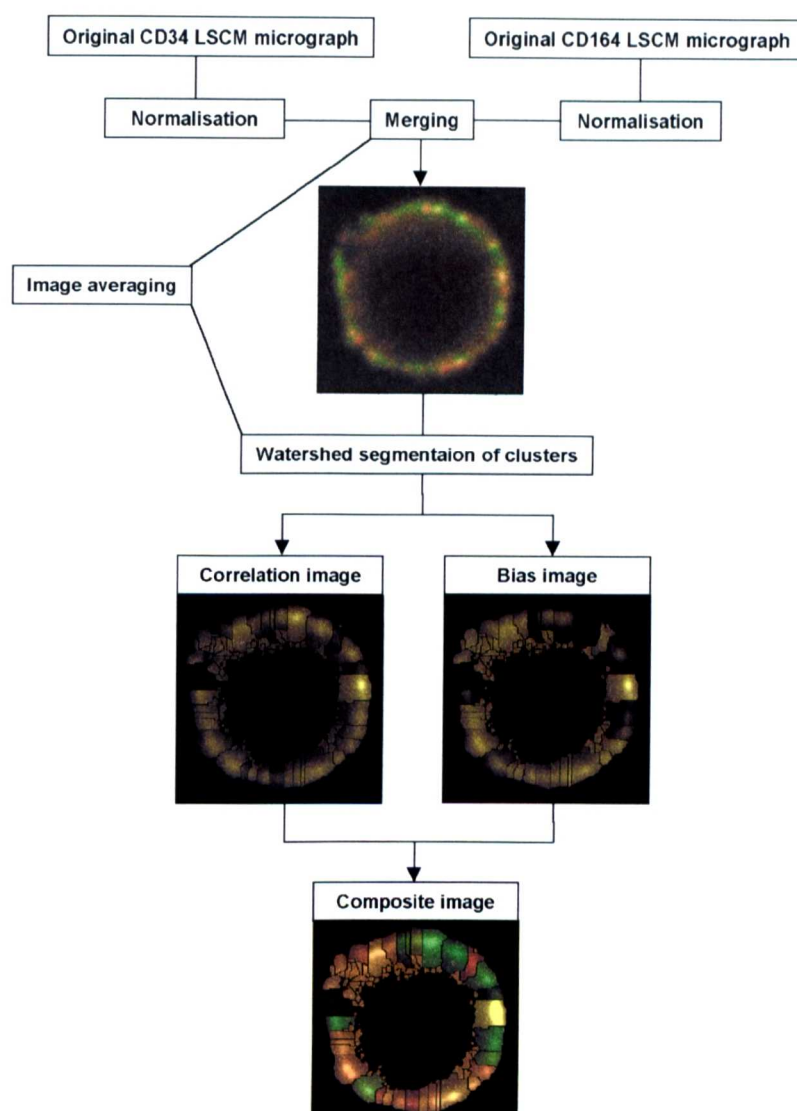
Figure 5.14: **Example of watershed segmentation for colocalisation analysis.**

The (middle) left image shows the colour-coded correlation coefficient computed in each single watershed basin (weighted by the mean intensity in each basin). The right image shows the colour-coded bias calculated in each basin. It can be noticed that in some areas, correlation and bias are very similar while in some other areas they are very different (e.g. top of the cell). This illustrates visually the differential expression of CD34 and CD164 around the cell membrane. The composite image at the bottom is a combination of both correlation and bias images. It is shown here to illustrate how the proposed watershed colocalisation analysis could be used to visually emphasise the colocalisation properties of the merged distributions. For instance, clusters with high correlation coefficients and low biases are found predominantly on the top-right part of the distribution, while clusters with high correlation coefficients and higher biases are found predominantly on the bottom half of the distribution. This is by no mean obvious by looking at the merged distribution.

5.2.3.3. Colocalisation analysis using the ADM

For comparison, colocalisation analysis was investigated using the newly developed ADMs. The proposed analysis also works on the assumption that a linear relationship exists between the expression levels of the dual labelled antigens, and aims to measure the deviation from this linear model. A high deviation would show that, despite antigens being co-localised at the same locations, they are differentially expressed and thus do not strictly follow a linear model. If the antigen distributions of colocalised antigens are following a linear model, as it would likely to be expected when labelling different epitopes of the same molecule, on the same cell, the corresponding ADMs should be extremely similar. However, in some situations, the contrast and brightness parameters of the image acquisition system may have affected the range of intensities available in the ADMs. Under the condition that such experimental adjustments of the image quality are applied reasonably to the original LSCM images by the operator (as it is recommended in good confocal microscopy practice, e.g. refer to the guidelines of the free NIH image processing and analysis software), their effect should result in a modification of the ADM histogram shape. Theoretically, the histogram of one ADM which could look like a Gaussian curve would also have similar shape when computed from the other ADM. However, the position of the peak may be different (due to brightness adjustment) and the standard deviation may not be the same either (a narrower peak would be expected if the contrast of the LSMC images used to generate this ADM was stronger than for the other micrographs used to generate the first ADM). The following procedure was applied to compensate this problem and perform colocalisation analysis on paired ADM.

CD34 ADM   **CD164 ADM**

ADM thresholded in three classes: halo, low and dense clusters

halo

low density clusters

high density clusters

Classes	Intensity range
Halo	mean ADM intensity - SD to mean ADM intensity
Low density clusters	mean ADM intensity to mean ADM intensity + SD
High density clusters	mean ADM intensity + SD and above
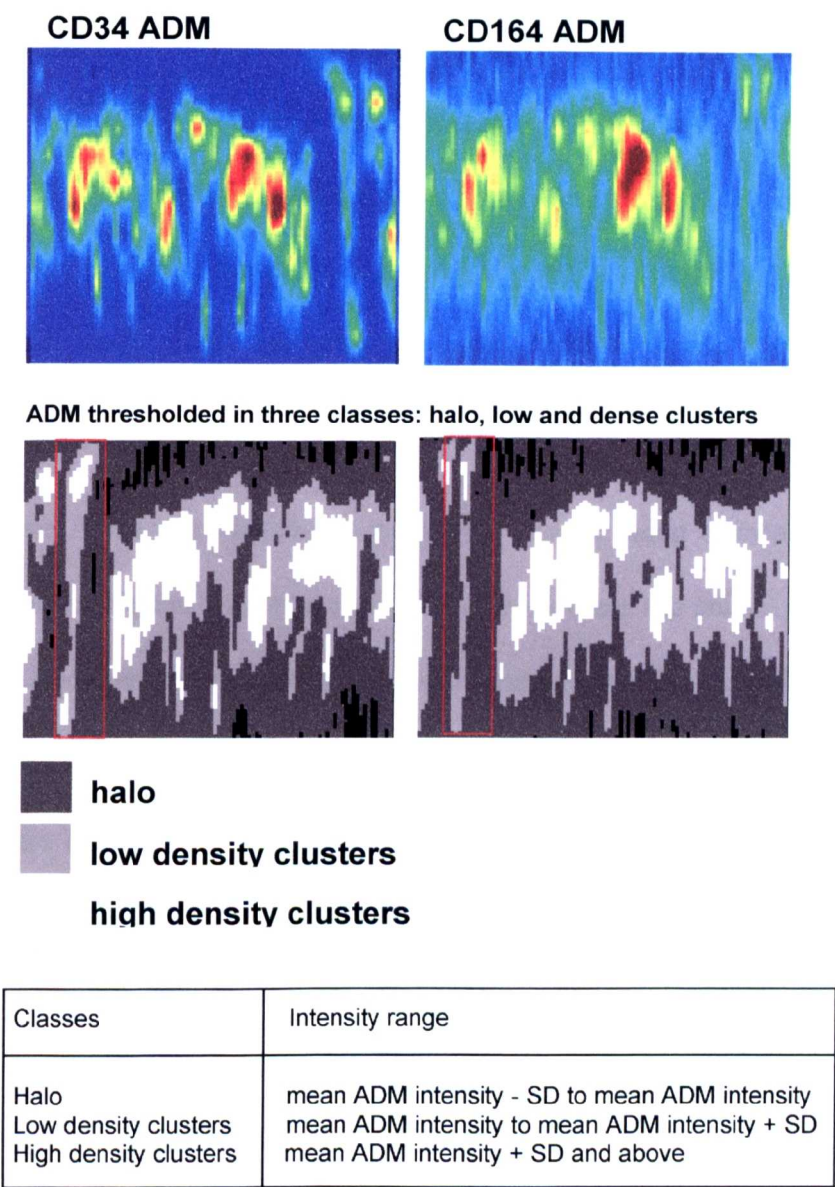
Figure 5.15: **Colocalisation analysis using the ADM.**

Top – Two CD34 and CD164 ADM to be analysed for colocalisation. *Middle* – Thresholded ADMs in order to locate three different classes, namely the halo, low density clusters and high density clusters. Classes are identified below. Despite the original ADMs having different intensity distribution, thresholding according to the procedure described in the text leads to very similar classes well colocalised. The red rectangles show a meridian-shape crest-like (MSCL) structure extending from one pole of the cell to the other. It can be observed that antigen density varies along the MSCL structure, but both CD34 and CD164 are found colocalised in this structure. *Bottom* – Table providing the intensity ranges used to define each class.

In this procedure, dual ADMs were thresholded using N equally spaced thresholds

(N=80 in this work) chosen to cover the range of intensities corresponding to halo,

low and high density antigen clusters. For each threshold, dual binary ADMs were generated in which positive values indicated regions with intensities equal or higher than the threshold. Colocalisation analysis was performed by measuring the degree of overlap between binary ADMs and the proportion of overlapping regions for each of the antigens, for every threshold. The method is further detailed below.

Firstly, the ADM were resized to a 100x161 matrix using bilinear interpolation. Bilinear interpolation was used in this case to increase the number of grey levels in the ADM. Secondly, each ADM was successively thresholded in the range $\left[\overline{m} - \sigma; \overline{m} + \sigma\right]$, where $\overline{m}$ and σ are the mean and standard deviation of the ADM (which are not changed after bilinear re-sampling). From this interval (different for each ADM), 80 equally spaced thresholds were selected to binarise the ADM. (This procedure has exactly the same effect as standardising the ADM by subtracting the average ADM value to each elements and dividing them by the ADM standard deviation, then thresholding the ADM in the range [-1; 1] using 80 equally spaced thresholds). The range $\left[\overline{m} - \sigma; \overline{m} + \sigma\right]$ was chosen to prevent the analysis of the extreme (low and high) fluorescence intensities, which may have been affected by a non-linear response due to possible, unknown brightness-contrast adjustment from the operator. The lower intensity values in this range are found in the ADM regions of the halo, while high intensities are found in the antigen cluster regions.

Let A_{CD34}^{T} and A_{CD164}^{T} be the binary images of the CD34 and CD164 ADMs thresholded at T, $T \in \{\hat{m} \pm \sigma\}$. For each T, the ADM regions where both CD34

and CD164 were co-expressed, i.e. where both binary ADMs shared white-positive areas, was defined as

$$A_{\cap}^{T} = A_{CD34}^{T} \cap A_{CD164}^{T} \, .$$ 5.16

The "$\cap$" symbol is equivalent to the logical "AND" operation. The respective proportions P_{CD34}^{T} and P_{CD164}^{T} of each antigen found colocalised with the other antigen were calculated as

$$P_{CD34}^{T} = \frac{A_{\cap}^{T}}{A_{CD34}^{T}} \quad \text{and} \quad P_{CD164}^{T} = \frac{A_{\cap}^{T}}{A_{CD164}^{T}} \, .$$ 5.17

By definition, both P_{CD34}^{T} and P_{CD164}^{T} have values in the range [0; 1], where a value 0 indicates that the antigen in question, e.g. CD34, is not colocalised with the other antigen, e.g. CD164, while a value 1 indicates that CD34 is entirely colocalised with CD164 (but a proportion of CD164 may not be colocalised with CD34).

It is stressed that this is measured for antigen intensities, or densities, above the threshold T. Moving the threshold may produce other values for P_{CD34}^{T} and P_{CD164}^{T}. Nevertheless, if the hypothetical existence of linear relationship expressed by equation 5.15 is true, P_{CD34}^{T} and P_{CD164}^{T} should be both equal to 1 for any T. Therefore it arises that these quantities are appropriate to quantitatively characterise the degree of colocalisation of each antigen, in respect to the other. However, in this state, they do not provide an overall quantitative measurement which could describe the degree of colocalisation of the merged antigen distribution. Consequently, another quantity is defined as

$$Col^T_{CD34/CD164} = \frac{A^T_\cap}{A^T_{CD34} \cup A^T_{CD164}} , \qquad 5.18$$

where $A^T_{CD34} \cup A^T_{CD164}$ is the union of the sets A^T_{CD34} and A^T_{CD164}, i.e. the collection of any elements belonging to both A^T_{CD34} and A^T_{CD164}. Again, if the linear relationship expressed by equation 5.15 is true, $Col^T_{CD34/CD164}$ should be equal to 1 for any value of T. Therefore, $Col^T_{CD34/CD164}$ is also an appropriate measure to quantitatively characterise the degree of colocalisation between two antigens, here applied on the ADMs, and particularly to observe (see figure 6.2) the likely deviations from the linear hypothesis of equation 5.15.

5.2.4. Investigating the potential role of CD34 by analysing CD34 antigen clusters distribution using the ADM

Our group and others (refer to chapters 1 and 7) previously made assumptions on the potential biological functions of the antigens investigated in the present thesis. It has been reported that CD34 may have a role in cell migration and HSC homing under physiological conditions. Such conditions imply that circulating CD34[+] HSPC migrate from blood vessels, through the vessel endothelium, to HSPC niches (particularly the bone marrow in adults) by sensing the appropriate chemotactic gradient in the local micro-environment. HSPCs may consequently react to a chemotactic gradient by migrating along that gradient (theoretically by moving towards the source of the highest chemoattractant concentrations). Nevertheless, such hypothesis arose from migration experiments [Phee et al, 2001], mostly using the well known and used trans-well designs (despite

criticisms about their reliability and repeatability properties). In such experiments, a well is divided into two chambers using a horizontal membrane (that membrane has pores of known sizes). The lowest chamber contains the chemoattractant while the upper chamber contains the HSC incubated in the culture medium. A chemotactic gradient is theoretically established between the lower and upper chamber, which potentially triggers CD34$^+$ HSPC migration through the membrane. The migration activity is quantified by measuring the ratio of cells migrated to the lower chamber in respect to the cell number in the upper chamber (commented in [Jeon *et al*, 2002]).

Under this experimental set up, non-direct evidence that CD34 is involved in trans-well migration is provided by blocking experiments, where anti-CD34 monoclonal antibodies (moAbs) are added to the medium in order to bind specifically to the membrane CD34 antigens. This presumably inhibits the CD34 function. It was observed that lower migration rates of migrating CD34$^+$ cells were obtained under these conditions, but it could be argued that other molecular phenomena are taking place (e.g. anti CD34 moAbs may artificially activate CD34 which in turn lowers the migrating potential of the cells, or the moAbs make the cells to stick more strongly to the membrane, preventing cell migration).

CD34 may also be involved in participating to cell-to-cell contact. Indeed, LSCM image analysis and 3D reconstructions of CD34 spatial distributions tend to suggest that, despite strong colocalisation of the whole CD34 membrane pool, antigen clusters may often been found at the contact areas between adjacent cells. CD34 clusters may also been found preferentially distributed towards the contact pole of the cell with the microscope slide. Why clusters form and are sometimes

found in a polarised manner is not understood. Under the light of some time-lapse experiments [Francis *et al*, 1998] and electron microscopy data (data not shown), human HSPC seem not to deform their shape in order to communicate to each other. Instead, they produce thin filaments (tenupodia, pseudopodia) of various lengths to reach each other, get closer, or maybe exchange biochemical signals. In this perspective, CD34 clusters could be a mean of concentrating CD34 molecules in the membrane region where a filament is to be produced, consequently coating the filament, during its growth, with CD34, and bringing the molecule in contact to other cells in order to perform the desired function.

Antigen polarisation may involve the whole antigen pool, or mostly the halo, or the clusters, or a combination of both. In order to further investigate the potential function of CD34, the spatial distribution of the CD34 antigen clusters in the HSC membrane was analysed. Two ADM-based approaches are further detailed below, one consisting in the direct detection of clusters and the other taking account of the gradient of the ADM with the assumption that clusters generate higher gradient than halo regions. For both methods, a ratio similar to the Z-polarity is calculated to express the polarisation of the clusters in the southern cell hemisphere.

5.2.4.1. Antigen cluster detection

The first approach aimed to detect antigen clusters from the ADMs. The definition of clusters is not as straightforward as it may seem. There are small dense clusters, small loose clusters, big clusters of various densities, and regions containing

mixtures of different types of overlapping clusters. This explains why clusters are difficult to define and locate using the intuitively simple natural terms of morphology. Regarding the ADMs, clusters were defined as local maxima (points where surrounding values in a defined neighbourhood are lower than the middle one). In order to retain the most significant clusters, only the maxima whose intensities were higher than the mean ADM intensity were considered.

The ADMs were first filtered to eliminate the random intensity fluctuations caused by noise. The smoothing procedure made use of morphological operators, erosion and dilation, applied to each row of the ADMs. The morphological operation of the dilation, noted $\oplus$, of the set A by the set B is mathematically defined [Gonzales and Woods, 2002] as

$$A \oplus B = \left\{ z \,\middle|\, \left(\hat{B} \right)_z \cap A \neq \varnothing \right\} \qquad\qquad 5.19$$

where z is a point of Cartesian coordinates (z_1, z_2) and $\left(\hat{B} \right)_z$ is the *translation* to the point z of the *reflection* of B. The translation is defined as

$$(B)_z = \left\{ c \,\middle|\, c = a + z, \quad \text{for} \quad a \in A \right\} \qquad\qquad 5.20$$

and the reflection is defined as

$$\hat{B} = \left\{ w \,\middle|\, w = -b, \quad \text{for} \quad b \in B \right\}. \qquad\qquad 5.21$$

Other definitions of the dilation may be found in literature. The erosion operation is defined as

$$A \ominus B = \left\{ z \,\middle|\, (B)_z \subseteq A \right\}. \qquad\qquad 5.22$$

In grey scale images, the dilation operator assigns the maximum intensity value found in the 1x5 region covered by the filter to the middle element of the visited region. Then the filter was displaced by one bin in the horizontal direction and the process was repeated, until the filter reached the end of the row. The erosion operator works conversely, assigning the minimum intensity value found in the 1x5 region covered by the filter to the middle element of this region.

Firstly each row of the ADM was twice dilated using a 1x5 dilating operator, and then twice eroded using a 1x5 erosion filter. These operations were performed on rows, not columns, since the antigen clusters were located on a z-scan basis (one row = one z-scan) and were defined as local maxima (see below). The size of these morphological filters was manually selected to remove noisy areas while globally preserving the predominant, most significant clusters.

Secondly, the local maxima in the ADM were identified as indicating clusters. Two maxima were merged if they were too close to each other (typically closer than five distance units), which indicated that they belonged to the same cluster while the smoothing step was not strong enough to merge them despite the presence of the noise. This minimum allowed distance was chosen subjectively according to the size of the smallest but clearly identifiable clusters.

Once identified, the proportion of clusters located in the "southern" hemisphere was calculated for each ADM and stored for further analysis. A proportion higher than 0.5 (more than 50% of the clusters located towards the contact area of the cell with the slide) suggested that the antigen clusters on this cell may have been involved in adhesion function. It was noted that the low vertical resolution of the ADM may affect the calculation of the proportion of "southbound" clusters

(values of 0.49 – clusters polarised to contact pole – or 0.51 – polarity towards free pole – may actually not reflect any particular biological tendency, and may have been affected by increasing the vertical resolution). Furthermore, the clusters are detected on a local maxima basis, which does not provide information on the actual size of the clusters or their antigen content. Following these reflections, it appeared that another means of characterising the involvement of cluster to distribution polarity was needed (next section).

5.2.4.2. Antigen cluster characterisation using "distribution energy"

In section 5.2.2.2, the polarity along the z-axis was defined from the ADM as the ratio of the amount of fluorescence in the southern hemisphere to the total amount of fluorescence. This definition did not distinguish between the contribution of the halo, loose or dense clusters to the polarity. Trying to identify them individually in order to assess their contribution to Z polarisation, as described in the previous section, is not a straightforward task. A similar problem is encountered in the literature presenting algorithms to segment spots on 2D electrophoresis gel. Many approaches use modified watershed segmentation or 2D spot modelling [Bettens et al, 1996]. Such techniques do not work when applied to the ADM due to the low resolution along the z-axis. In this section, a method reflecting both the contribution of the clusters and their antigen content to Z polarisation is presented.

First, only the absolute magnitude of the gradient $\nabla f(x, y)$ of the interpolated ADM $f(x,y)$ is computed. Since the gradient has values proportional to local intensity changes in the ADM., wider and more uniform halo areas have a very

low absolute gradient in comparison to cluster regions. Secondly, the intensity of the actual signal, which reflects the amount of antigen [Kara *et al*, 2001], especially the signal contained in the clusters, is taken into account. This is realised by element-by-element multiplication of the re-sized ADM by its absolute gradient map, and the product represents the distribution energy map *E*, i.e.

$$E(x,y) = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2} \times f(x,y)$$

$$= \left|\nabla f(x,y)\right| \times f(x,y)$$

5.23

The "×" indicates element-by-element multiplication. The distribution energy contained in the southern hemisphere of the cell membrane is evaluated and referred as the energy ratio in the southern hemisphere defined, for a *MxN* matrix *E* as

$$Ratio_E = \frac{\displaystyle\sum_{x=M/2}^{M}\sum_{y=1}^{N} E(x,y)}{\displaystyle\sum_{x}\sum_{y} E(x,y)}.$$

5.24

The denominator is constant and represents the sum of all elements in *E*. As for the polarity Z (but regarding the clusters), an energy ratio higher than 50% indicates a preferential tendency of the cluster to be polarised towards the contact point of the cell with the slide. This measure is more convenient to assess the contribution of cluster to Z polarity than the simple relative count of clusters since it does not rely on the accurate detection of clusters and also takes into account the amount of fluorescence per cluster.

A – Negative original LSCM image filtered with a 10x10 averaging window

Image normalised in the range 0–255
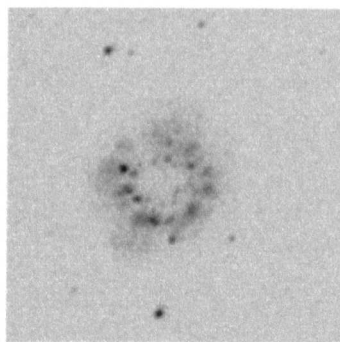
B – Negative of the gradient image

Image normalised in the range 0–255
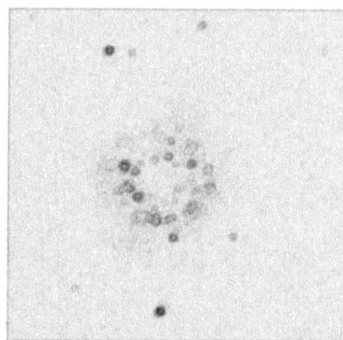
C – Product of **A** and **B**

Image normalised in the range 0–255
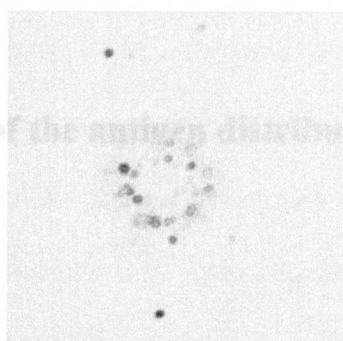
Figure 5.16: **Effect of calculating energy distribution rather than antigen density.**

Energy distribution is proposed to quantify the effect of the clusters in the polarisation of the antigen distribution along the Z axis. The computation was applied on ADM, but its effect is shown here on a single LSCM image in order to visually benefit from the better resolution of the picture. A – The original LSCM image is first smoothed using a 10x10 averaging filter and B – its absolute gradient is calculated. C – The distribution energy is obtained by multiplying A by B. Therefore the energy image is the fluorescence image (or "density", or "mass" image) weighted by its absolute gradient. The effect of the procedure is to emphasize the clusters while reducing the effect of the halo. The three images are normalised in the intensity range 0-255. The strong background and the important loose fluorescent halo have been noticeably attenuated in the energy image while the clusters have been retained.

Chapter 6

Quantitative analysis of the antigen distributions

Chapter 6. Quantitative analysis of the antigen distributions

In chapter 5, several quantities measured on the ADM were defined in order to characterise the antigen distributions. In this section, a short review of these quantities is given and the results of the analysis from different cell samples are summarised.

1 – The first quantity is the mean fluorescence intensity per ADM, or *mean ADM intensity*. It reflects the mean antigen concentration per imaged cell. Comparing mean ADM intensities may potentially help discriminating between different HSPC, or cells at various stage of activation. Thus, cell populations or sub-populations could be identified. Following the same principle, the current techniques of choice for cell population analysis and comparison are Fluorescence Activated Cell Sorting (figure 6.1) [Grabarek and Darzynkiewicz, 2002 ; Ramarao and Meyer, 2001] and flow cytometry (similar to FACS, but cells are not separated [Lore *et al*, 2001; Macey *et al*, 1997]), which records the mean fluorescence intensity of each individual cell (or event) in a preparation [Pasino *et al*, 2000;] (see figure 6.2). Therefore, comparing the distributions of mean ADM intensities is similar to performing FACS analysis. Moreover it is known that image processing techniques can provide more information about cell populations [Lore et al, 2001] as more features can be measured [Swedlow *et al*, 2003]. Consequently,

such measurements obtained on various small HSC samples are also presented and described below.

2 – The *mean number of fluorescent clusters* in a cell population can be measured from the ADM according to the method described in section 5.2.4.1, and was used to compare the cell samples. The *mean fluorescence intensity in clusters* can also be measured. Because such clusters are detected in the ADM, their spatial location was recorded as well, which allowed to calculate the proportion of clusters located in the south hemisphere of the cell body, i.e. half the cell portion adhering to the microscope slide. If these antigens are involved in cell adhesion and belong to cluster structures, then one should expect that – predominantly – more than 50% of the clusters are located in the south hemisphere.

3 – The antigen distribution polarity along the z-axis – *polarity Z* – or in the XY-plane – *polarity XY* – can also be quantified using the ADM (section 5.2.2) to assess the potential involvement of the investigated antigens in cell adhesion, cell migration or cell-to-cell communication.

4 – As the *percentage of clusters located in the southern hemisphere* was observed not to be a very robust criterion when measured in ADM computed from low intensity / poorly contrasted LSCM images, the *distribution energy* was introduced and defined (section 5.2.4.2). It was used to quantify the proportion of high density structures in the cell membrane (clusters) in the cell hemisphere.

In summary, the quantities measured from the ADM and for any antigen distribution (from either single or dual labelled cells) are

- mean ADM intensity (or mean antigen density)

- mean number of clusters per ADM

- mean intensity in clusters

- % of clusters in southern hemisphere

- polarity z

- polarity xy

- distribution energy.
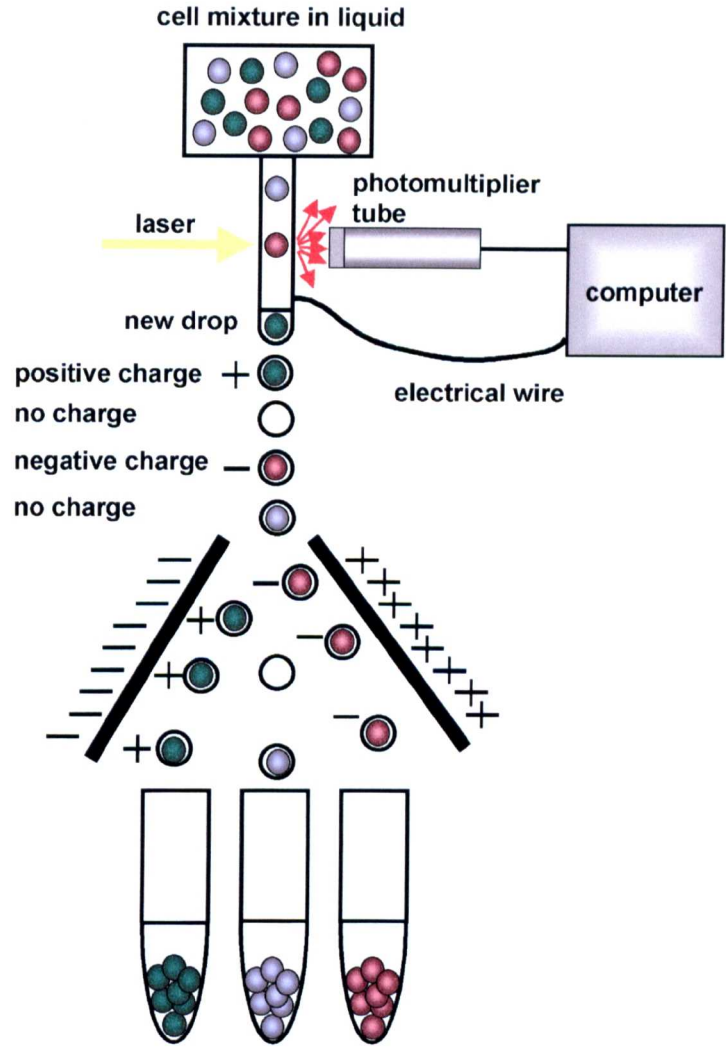
Figure 6.1: **Principles of Fluorescence Activated Cell Sorting (FACS).**

A mixture of cell is labelled with (generally) two fluorescent dies (bound to specific monoclonal antibodies targeting particular antigens) are passed through a very narrow tube in which individual cells are excited by a laser. Depending on the excited fluorochrome, cells acquire a positive or negative charge, or remain neutral (if the cell carries none of the target antigens). The emitted fluorescence is recorded in a photomultiplier tube and stored in a computer for further analysis. Each detection is referred to as one event. Individual droplets (either containing one cell or none) are deflected between two electrically charged plates according to their polarity. The different, sorted cell fractions are then collected. FACS analysis quantifies the average fluorescence measured per cell/drop.
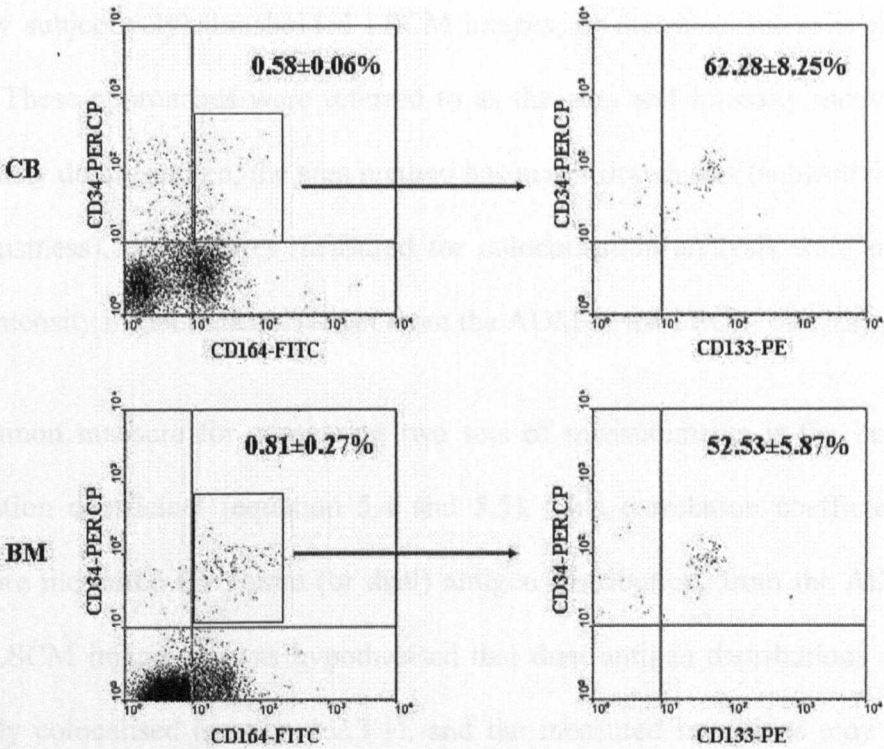
Figure 6.2: **Flow cytometric distribution analysis of CD34 and CD164 on CB and BM.**

Left panels are representative examples of the CD34 and CD164 antigenic profile on CB and BM MNC. The smaller rectangles (left panels) highlight CD34+CD164+ discrete cell subset in each source. Each point represents a single event, the average fluorescence count for an individual cell. Right panels represent CD133 co distribution on CD34+CD164+ cells. Results are expressed as mean ± SEM of percentage expression (CB, n=11; BM, n=4). FACS analysis identifies a very immature HSPC population but does not provide the possibility to further characterise the different antigen distributions in cell membranes and the spatial relationships between co-expressed antigens.

6.1. Features measured for colocalisation analysis of dual labelled cells

As mentioned in section 5.2.3, usual colocalisation analysis methods compare area ratios where two fluorochromes are colocalised, i.e. area comparison of (mostly subjectively) thresholded LSCM images, or the mean intensity ratios in ROIs. These approaches were referred to as the area and intensity methods. As previously demonstrated, the area method has major drawbacks (subjectivity, lack of robustness), the features measured for colocalisation analysis were obtained from intensity measurements (either from the ADM or the LSCM images).

A common measure for comparing two sets of measurements is the Pearson's correlation coefficient (equation 5.4 and 5.5). This correlation coefficient was therefore measured for paired (or dual) antigen distribution, from the ADM and from LSCM images. It was hypothesised that dual antigen distributions may be spatially colocalised (section 1.2.3.1), and the measured intensities may have a linear relationship expressed by equation 5.15. If this assumption is true, the correlation coefficient would be high. However, the correlation coefficient does not inform on the bias ("a" in equation 5.15). Therefore, further to measuring correlation, the bias itself was also measured and stored for comparison between different sample. Both the correlation coefficient and bias were measured at two different levels on the ADM (section 5.2.1). Furthermore, they were measured for any of the measurements summarized above.

6.2. Statistical analysis

For paired measurements, a two-tailed Student's test was performed on the measurements, and P values were stored (see table 6.3). For independent measure comparisons, a Levene's test for variance equality was applied since this test is less sensitive to deviation from Normality (indeed, the populations of measurements analysed were very small, about 9-13 elements each time, so they were likely not to be Normally distributed), and an independent t-test was used to test mean equality. Examples of such tests are provided in Appendix E.

Using the proposed features (chapter 5) obtained from the CD34 epitope datasets, the measurements were further processed to compensate for the bias ("a" in equation 5.15) using the equation of the regression line that best fits the data. A t-test was then re-applied to this transformed data, and the P values were stored (table 6.3).

6.3. Detailed analysis and comparison of CD164$^+$CD34$^{\text{class III}+}$, CD34$^{\text{class I}+}$, CD34$^{\text{class II}+}$ and CD34$^{\text{class III}+}$ HSPC populations

Most of the ADM measurements were performed on all the available and documented LSCM datasets, which were:

CD34 class III epitope / CD164 class II epitope

CD34 class I epitope / CD34 class II epitope, sample 1

CD34 class I epitope / CD34 class II epitope, sample 2

CD34 class I epitope / CD34 class III epitope, sample 3.

These measurements were:

— mean ADM intensity (or mean antigen density)

— mean number of clusters per ADM

— mean intensity in clusters

— percentage of clusters in southern hemisphere

— polarity Z

— polarity XY

— percentage of distribution energy in southern hemisphere

Colocalisation analysis involved the measurements of other features measured either on the LSCM images themselves of the ADMs.

All series of measurements are presented in different tables, depending on the type of analysis carried out, and are described and commented in the following sections. Results are expressed as mean ± standard deviation (SD) for each distribution. For each type of analysis, the statistical test used is indicated and, when relevant, the P values are provided. P values lower than 0.05 were considered significant.

These sections are organised as follow:

- first, most of the above features are presented and compared between the various HSPC populations investigated.

- Then special emphasis is given to the analysis of the CD34 epitopes datasets since they were the most abundant data available for the present study.

- It is followed by a description of the correlation analysis performed on the CD34 class III epitope / CD164 class II epitope cell population (i.e. $CD164^{+}CD34^{+}$ cells).

- A whole section is then devoted to the polarity analysis.

- The remaining sections are dedicated to colocalisation analysis of dual labelled populations,

  o firstly using correlation analysis between measurements obtained at the Z and subcellular levels of observation,

o secondly using results from watershed segmentation of equatorial Z-section, and thirdly using the results obtained from colocalisation analysis developed from the ADMs.

6.3.1. Comparison of HSPC populations

For each of the antigens/epitopes combinations, most of the above features are presented in table 6.1. For comparison, the percentage of clusters in south hemisphere and the polarity z are shown on the right side of the table. The bold numbers in these two last columns indicate the mean values above 0.5, which

Table 6.1: **Measurements from ADMs. Comparison of HSPC populations.**

antigen combinations	antibodies	measurements from ADM				
		mean ADM intensity	mean number of clusters	mean intensity in clusters	% clusters in south hemisphere	% polarity Z
CD34 - CD164	CD34 IgG1 + HPCA-II class III epitope	1.43 ± 0.21	21.20 ± 5.89	2.55 ± 0.56	46 ± 27	50.31 ± 3.88
	CD164 IgG3 + 103B2/9E10 class II epitope	1.30 ± 0.16	19.80 ± 6.06	2.04 ± 0.35	39 ± 22	49.85 ± 1.96
CD34 epitopes I - II sample 1	My10 class I epitope	3.07 ± 0.59	19.62 ± 3.97	7.07 ± 4.02	55 ± 22	50.95 ± 3.98
	QBEND10 class II epitope	3.85 ± 0.95	18.92 ± 2.87	7.87 ± 2.95	47 ± 23	53.00 ± 8.10
CD34 epitopes I - II sample 2	My10 class I epitope	2.56 ± 1.41	9.92 ± 3.17	7.52 ± 3.20	49 ± 20	49.75 ± 13.77
	QBEND10 class II epitope	2.55 ± 1.08	10.62 ± 2.79	8.36 ± 3.74	42 ± 22	47.00 ± 10.50
CD34 epitopes I - III sample 3	My10 class I epitope	2.15 ± 1.19	18.45 ± 5.61	7.12 ± 1.46	40 ± 24	47.389 ± 7.65
	HPCA-II class III epitope	2.68 ± 1.13	18.27 ± 5.61	7.04 ± 2.88	39 ± 17	51.28 ± 9.03

indicate (by definition) some degree of polarisation. Furthermore, the antibodies used in these dual labelling experiments are also presented in the table.

6.3.1.1. Mean ADM intensity

6.3.1.1.1. CD34 – CD164 comparison

For the CD34/CD164 dual labelling experiments, the mean ADM intensities appear similar for both antigens, but a two-tailed paired t-test provides a value of $P = 0.010$, which indicates that these mean antigen densities are significantly different. From the measurements, it appears that less CD164 fluorescence is consistently measured in this cell sample. It is also noted that these mean values are significantly lower than any other mean antigen density from the epitope analysis, so it appears from this investigation that CD164 is expressed at the lowest antigen density levels in comparison to CD34 epitopes.

6.3.1.1.2. CD34 class III epitope comparison

In comparison to the only other data on CD34 class III epitope (sample 3), the measured fluorescence of CD34 from $CD34^+CD164^+$ cells is significantly lower than the measured fluorescence from $CD34^{class\ I\ +}CD34^{class\ III\ +}$ ($P = 0.030$ from independent t-test for mean equality, after variances were shown not to be significantly different – $P=0.143$ from Levene's test for variance equality).

6.3.1.1.3. CD34 class II epitope comparison

CD34 class II epitope positive cells were imaged from two different CB samples and delivery dates (sample 1 and sample 2). A Levene's test and a t-test reveals

that both variances are similar (P=0.830) and the means are significantly different (P=0.003), sample 2 being lower than sample 1 by the antigen density 1.304 (background normalised units). This means that CD34 class II epitope is expressed in an excess of over 50% in sample 1 in comparison to the mean fluorescence measured in sample 2. However, because no fluorescence calibration was available, the possibility that this difference may be due to experimental considerations (despite the fact that the same protocol was applied throughout the whole investigation) has to be considered.

6.3.1.1.4. CD34 class I epitope comparison

The three samples (sample 1, 2, and 3 in the first column of the table) were all analysed for CD34 class I epitope. Levene's test and independent t-test indicated that samples 1 and 2 have different variances (P=0.001) while their means are similar (P=0.248). Samples 2 and 3 are distributed similarly, having similar variances and means (P=0.231 and P=0.456 respectively). Samples 1 and 3 have significantly different variances and means (P=0.085 and P=0.035 respectively), so an excess of 0.92 fluorescence (in background normalised units) was measured for CD34 class I epitope in sample 1 in comparison to sample 3.

Therefore CD34 class I epitope was measured to be expressed in lower amount between one of the two CD34 $^{class\ I}$ $^{+}$CD34 $^{class\ II}$ $^{+}$ samples (sample 1) in comparison to the CD34 $^{class\ I}$ $^{+}$CD34 $^{class\ III}$ $^{+}$ sample (sample 3). The mean fluorescence intensities measured for CD34 $^{class\ I}$ $^{+}$CD34 $^{class\ II}$ $^{+}$ (sample 2) and CD34 $^{class\ I}$ $^{+}$CD34 $^{class\ III}$ $^{+}$ (sample 3) are not statistically different. It is

consequently difficult to assess whether or not the observed differences reflect some fundamental biological facts are an experimental artefact.

6.3.1.2. Mean number of clusters

6.3.1.2.1. CD34 – CD164 comparison

The mean number of clusters for CD34 and CD164 on dual labelled CD34+CD164+ cells were not statistically different.

6.3.1.2.2. CD34 class III epitope comparison

Again, the CD34 class III epitope distributions from $CD34^+CD164^+$ cells were compared with $CD34^{\text{ class I }+}CD34^{\text{ class III+}}$ cells (sample 3). They were not found to be different (P = 0.356 from independent t-test for mean equality, after variances were also shown to be similar – P=0.939 from Levene's test for variance equality).

6.3.1.2.3. CD34 class II epitope comparison

The CD34 class II epitope distributions from samples 1 and 2 were compared, and the average number of clusters in each sample were found to be significantly different (P=0.796 from Levene's test for variance equality, so the variances are similar, and P<0.001 from t-test for mean equality, so the means are different). The t-test indicates that the CD34 class II epitope from sample 1 contains, in

average, 8.3 more clusters than CD34 class II epitope from sample 2. This almost double (80% increase) the mean number of clusters measured in sample 2.

6.3.1.2.4. CD34 class I epitope comparison

The comparison of the three samples of CD34 class I epitope further revealed that two out of the three possible sample combinations have significantly different average number of antigen clusters per cell. Indeed, twice more CD34 class I epitope clusters were measured in sample 1, in comparison with sample 2 (P=0.0.780 from Levene's test, so the variances are similar, and P<0.001 from t-test, so the means are different). Similarly, almost twice more CD34 class I epitope clusters were measured in sample 3, in comparison with sample 2 (P=0.0.083 from Levene's test, so the variances are different, and P<0.001 from t-test, so the means are also different). However, the sample 1 / sample 3 combination was statistically identical.

6.3.1.3. Mean intensity in clusters

6.3.1.3.1. CD34 – CD164 comparison

The mean fluorescence intensities in the clusters of both dual labelled distributions were not statistically different.

6.3.1.3.2. CD34 class III epitope comparison

No significant difference was measured for the mean intensity in the clusters of the $CD34^+CD164^+$ cells when compared with $CD34$ class I + $CD34$ class III+ cells (sample 3). The average number of clusters for these samples was reported to be statistically similar.

6.3.1.3.3. CD34 class II epitope comparison

The mean fluorescence intensity in the clusters of the two samples compared (1 and 2) is identical amongst the samples. Despite the fact that a much lower number of clusters was measured on sample 2, the mean fluorescence measured in the clusters of sample 1 is statistically comparable to sample 2.

6.3.1.3.4. CD34 class I epitope comparison

Again, all of the three possible combinations for comparing the CD34 class I epitope samples showed that the mean fluorescence intensity in the clusters is identical amongst the samples. Therefore, despite the fact that some samples may display a lower number of antigen clusters, the mean fluorescence measured in these clusters is statistically comparable to the other samples, richer in clusters.

Interestingly, it results from this analysis that the mean intensity in the clusters is comparable for any of the imaged antigen distributions.

6.3.1.4. Percentage of clusters in southern hemisphere

6.3.1.4.1. CD34 – CD164 comparison

No significant difference is observed, and both percentage of clusters polarised in the southern hemisphere of the cells were below 0.5, indicating that the antigen clusters were not preferentially polarised towards the contact points of the cells with the microscope slide.

6.3.1.4.2. CD34 class III epitope comparison

Once more, no difference is observed, and again both percentage of clusters polarised in the south hemisphere of the cells were below 0.5, so the CD34 class III antigen clusters were not preferentially polarised towards the contact points of the cells with the microscope slide for these samples.

6.3.1.4.3. CD34 class II epitope comparison

Yet again, no difference was measured between the cluster polarisation on CD34 class II epitope distribution, and still both percentage of clusters polarised in the south hemisphere of the cells were below 0.5.

6.3.1.4.4. CD34 class I epitope comparison

Even though the measures seem to be different, the differences were not significant.

6.3.1.5. Polarity Z

The polarity Z values were provided in this table 6.1 to allow the comparison with the percentage of cluster located in the cell's southern hemisphere. Similarly to the results obtained from the analysis of the percentage of cluster located in the cell southern hemisphere, no significant differences were measured when comparing the sample combinations. For some combinations, the difference of the mean seems important, but the corresponding standard deviations are high, which makes both distributions significantly overlap. Hence the similarities between measurements when compared at the population level.

It is important to notice that within the cell population themselves, some cells are found highly polarised. This may indicate that some HSPCs can still actively polarise their antigen distributions when specifically stimulated at their contact point with the microscope slide, the actual nature of the stimulus having to be elucidated. The gold of the slide coating is unlikely to be a strong promoter of polarisation. Biologically sound molecules classically used in adhesion experiments, such as fibronectin or collagen, could be used to repeat the LSCM investigation. Then, a significant difference in cell polarisation may be measured. Preliminary experiments carried out using fibronectin-coated slides revealed cells with a very strong antigen polarisation, which seemed to have been triggered by the cell adhesion to fibronectin (figure 5.10). If the gold substrate used in this thesis was not a polarisation stimulant, then the higher polarity observed in some

samples (bold numbers in far right column, indicating polarity Z > 0.5, or 50%) may reflect a non-specific tendency of these cells to polarise, suggesting indeed that a proper substrate would activate an actual polarisation of the antigen towards the cell contact point with the slide.

6.3.1.6. Comparison of the percentage of southbound clusters and polarity Z

In order to assess if the antigen clusters are particularly involved in the polarisation of the antigen distribution, the percentage of southbound clusters and the polarity Z were compared to each other, for each sample of the antigen combinations.

Despite the fact that all standard deviations of the compared measurements were significantly different, most differences of their means were not significant (P>0.1). The only polarised distribution of clusters in the southern hemisphere was observed for the class I epitope of sample 1 (percentage of southbound clusters = 0.55 ± 0.22). Similarly, the corresponding polarity Z is found slightly polarised as well (percentage of polarity Z = 50.95 ± 3.98). Therefore, in this case, the clusters are involved in the antigen distribution polarisation. However, whenever the polarity Z is over 50% for other samples (CD34 in $CD34^+CD164^+$ cells, CD34 class II epitope in sample 1, and CD34 class III epitope in sample 3), the percentage of southbound clusters is lower than 50%, indicating that the clusters are predominantly located in the northern hemisphere of the cell, and do not contribute to the polarisation of the antigen distribution. Therefore, in such instances, the polarisation is due to the predominance of the halo or very low density clusters in the southern pole. This is particularly observed for the CD34

class III epitope distribution (sample 3), for which the clusters are clearly polarised towards the free (north) pole of the cells (percentage of southbound clusters = 0.39 ± 0.17) while the whole antigen distribution is slightly polarised towards the contact point (percentage polarity Z = 51.28 ± 9.03).

This analysis therefore reveals that the antigen clusters do not participate actively to the polarisation of the antigen distribution in the membrane of the HSC. This also further strengthens the observations made in the previous section (6.3.1.5), that gold coating does not trigger specific polarisation.

6.3.2. General descriptive statistics for the CD34 class I, II and III epitope distributions

In this section, the results obtained for each of the epitope (i.e. class I, II and III) are summarised and commented.

Table 6.2 presents, for each epitope class, the mean ± standard deviation for each type of measurement. For all classes, the differences observed were not significant (P>0.05). As expected, the cluster antigen densities are (over twice) higher than the total antigen density (which actually represents the mean ADM intensity). The high standard deviations are due to the high variances observed in the various cell populations used in the study. These standard deviations are comparable to the detailed results (for each epitope class, for each sample) presented in table 6.1.

FACS analyses reported in the literature [Lanza *et al*, 1999; Maynadié *et al*, 2002] show little, if any, differences between the mean fluorescence intensities measured for each epitope class, in normal and cancer cell populations. Generally, a slightly weaker fluorescence is measured by FACS analysis for the class II CD34 epitope, in comparison to the other classes. In our data, the opposite tendency is observed, i.e. class II CD34 epitope displayed an increased

Table 6.2. General descriptive statistics for the CD34 class I, II and III epitope distributions.

Classes I, II and III CD34 epitope expression and membrane distribution features measured on populations of CD133 selected cells.

Features	CD34 MoAb			Comments
	class I My10	class II QBEND10	class III HPCA-II	
Total Antigen Density	2.62 ± 1.14	3.20 ± 1.20	2.68 ± 1.13	classes I, II, III similar results
Cluster Antigen Density	7.24 ± 3.07	8.11 ± 3.11	7.04 ± 2.88	classes I, II, III similar results
Number of clusters	15.86 ± 6.10	14.77 ± 5.06	18.27 ± 5.60	classes I, II, similar results, class III significantly higher
% southbound clusters	0.48 ± 0.22	0.44 ± 0.22	0.39 ± 0.16	classes I, II, III similar results, class III less variable
Z polarity	0.52 ± 0.15	0.49 ± 0.09	0.53 ± 0.10	classes I, II, III similar results
XY polarity	1.91 ± 1.09	1.90 ± 1.40	1.82 ± 0.77	classes I, II, III similar results, class III less variable

Data expressed as mean ± standard deviation

fluorescence, both reflected by the mean antigen densities and the cluster antigen densities. As for FACS results, this difference is not significant.

The number of antigen clusters is also similar, but CD34 class III epitope distributions showed a higher number, despite no statistical difference with the other classes.

The percentage of southbound clusters is consistent for each class. However, they are all below 0.5, which indicates no polarisation of the clusters towards the

contact point of the cells with the slide. This is comparable to the detailed results from table 6.1.

The axial polarity Z is also again found consistently distributed around 0.5 for all epitope classes, indicating no particular tendency of the CD34 antigen to polarise towards the adhesion sites.

The polarity XY measurements are all higher 1.5, which indicates that these cells all had a tendency to concentrate the CD34 antigen preferential on one side of the cell. This type of polarisation is found on migrating cells and in cell-to-cell contact events. Therefore, these high XY polarity values may suggest that CD34 has a role to play in either migration and/or cell-to-cell contact events. This further strengthens the reports made in the literature about the potential role of CD34.

6.3.2.1. Bias correction of paired measurements

Table 6.3 shows the Pearson's correlation coefficients obtained when comparing the dual measurements for the class I and II, and class I and III CD34 epitopes. As previously mentioned, this indicates the degree of linear agreement between two sets of measurements, but it does not inform on the potential bias which may affect the datasets. Therefore, the similarity of the dual datasets was assessed using a paired t-test (P-values given in table 6.3, second and fifth columns), directly measured from the raw measurements. It was expected to obtain very similar results for the features measures on the ADM for the CD34 epitope data, a high correlation should be observed, and (theoretically) the paired measurements

should not be biased. However, due to experimental considerations (problems with experiment repeatability), measurement similarity may not be observed. Therefore, we further made the hypothesis that a bias was present in our data, and that such bias could be corrected. The bias correction between two sets of measures x_i and y_i involved computing the slope of the best fitting straight line using linear regression method. This slope a is given by the equation

$$a = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}.$$ (55)

A new set of paired measures was then generated for which the slope of the best fitting line was equal to 1. This new set represents the (theoretically) unbiased measures. For any y, the corresponding unbiased measure y' is calculated as

$$y' = \frac{y - b}{a},$$ (56)

where

$$b = \bar{y} - a\bar{x}.$$ (57)

The P-value of this unbiased set was calculated in order to check that unbiased paired measurements were distributed similarly. P-values close to 1 were therefore expected.

Table 6.3. Bias correction of paired measurements.

Classes I-II and I-III CD34 epitope dual expression and membrane distribution features measured on dual labelled CD133 selected cells.

Features	CD34 MoAbs					
	Class I, II (My10, QBEND10) n = 27			Class I, III (My10, HPCA-II) n = 11		
	R value	P value [a]	P value [b]	R value	P value [a]	P value [b]
Total Antigen Density	0.72	0.034	0.999	0.27	0.243	1
Cluster Antigen Density	0.01	0.400	1	0.16	0.932	1
Number of Clusters	0.98	1	1	0.58	0.909	1
% southbound clusters	0.65	0.050	0.896	0.06	0.936	0.999
Z polarity	0.83	0.165	0.998	-0.07	0.845	0
XY polarity	0.95	0.999	0.999	0.20	0.599	1

R = Pearson's correlation coefficient
P from paired t-test
[a] before bias correction
[b] after bias correction

6.3.2.2. Comments on R values

The correlation for class I and II CD34 epitopes is much higher for the features assessed than for class I and III, for which the features seem totally uncorrelated. The fact that class I and III CD34 epitopes are uncorrelated was suspected at the visual examination of the LSCM images.

For class I and II CD34 epitopes, the cluster antigen densities are found to be uncorrelated. This is due to the fact that the clusters express the epitopes differentially, particularly in the high antigen density clusters (see sections 6.3.6 and 6.3.7), while the expected linear relationship described by equation 5.15 is more verified in lower intensity membrane domains, such as weaker clusters and halos. The proportion of southbound clusters also have a low correlation (R= 0.65). This reflects the fact that CD34 epitope clusters are distributed differently for different epitopes, thus showing that the CD34 clusters are not similar to each other in the cell membrane. It was indeed visually observed from the LSCM

images that some CD34 clusters express only one class of epitope instead of expressing the expected three classes. This suggests that different forms of the CD34 molecule may have different biological roles or functions.

The uncorrelated features measured for dually labelled CD34 class I and III epitopes further underlines the difference in antigenic distributions for different epitopes. Class III epitopes are part of the CD34 molecule the closest to the cell membrane, while class I is located the further away from the membrane. The lack of correlation seems to indicate that there is no functional relationship between class I and III epitopes.

6.3.2.3. Comments on P values

The P values of the original (potentially biased) measures do not indicate significant differences between the measurements (P>0.05) except for the percentage of southbound clusters of class I and II CD34 epitopes. A great variability of these values is observed, particularly for the comparison of class I and II CD34 epitopes. As stated above, because different forms of the same molecules were targeted, similar measurements should be obtained. The unbiased correlation analysis indicated that this is not the case. However, if it was the case, correcting the bias as described earlier should make paired measurements to be similarly distributed, which is confirmed by the P values after bias correction. Indeed, they all are close to 1 (min/max range = 0.896 - 1) except for the Z polarity of class I and III CD34 epitopes, for which the resultant p value is 0 (indicating no distribution similarity, even after bias correction).

The bias correction procedure applied to the epitope data shows that dual labelling experiments of antigens known to colocalise could be analysed using the methods presented in this thesis. Indeed, even the presence of an experimental bias (e.g. one channel recording twice more fluorescence than the other channel due to a faulty set up), could be compensated. Consequently, the distribution of features measured on the ADM could still be compared at the cell population level.

6.3.3. Correlation analysis of the CD34/CD164 distributions

Table 6.4 shows the correlation coefficients obtained from comparing the features measured from the ADM. The number of clusters and the polarity in the XY-plane are not correlated between CD34 and CD164. Regarding the uncorrelated number of clusters, the low R value is due to the nature of the CD34/CD164 LSCM original dataset. Indeed, the fluorescent distributions imaged expressed the presence of a strong halo in the membrane of all HSC imaged which was due to the presence of non-clustered CD34 and CD164 antigens. This was rarely the case for the CD34 epitopes data and can be verified by expressing the mean ADM intensity from table 6.1 as a percentage of the mean intensity of the clusters. For the epitope data, this percentage is found to be approximately 37% (min/max range = 30 - 49), while it is equal to 60% for the CD34/CD164 data (respectively 56% for CD34 and 64% for CD164). The higher value obtained for CD34/CD164 is the direct consequence of the presence of a stronger antigenic halo in HSC

Table 6.4. Correlation analysis of CD34/CD164 distributions.

Features measured from ADM	Pearson's correlation coefficient between dual CD34/CD164 membrane distributions
mean ADM intensity	0.98
number of clusters	0.43
mean intensity in clusters	0.83
% southbound clusters	0.74
polarity Z	0.98
polarity XY	0.35

membrane than for the CD34 class I, II or III distributions. This strong halo rose the thresholds used to define relevant antigen clusters, as described in section 6.4.2.1. Consequently, the number of clusters detected was subject to more variability, explaining the lack of correlation observed for this feature.

The polarity XY is also uncorrelated. This is no artefact from the analysis, and it really indicates a biological difference in the way the investigated cells spatially organise their CD34 and CD164 (see table 6.5 also) surface antigen content.

Both the mean ADM intensity and the mean intensity in the clusters are well correlated (R = 0.98 and 0.83 respectively), but a two-tailed paired t-test revealed that there was significant differences between the distributions of these measures (P = 0.010 and 0.026 respectively). This is an illustration of the potential presence of a bias between the paired measurements, as previously described (section 6.3.2.1): despite good correlation, the distributions differ. A way of possibly answering this question, calibration of the fluorescence should be performed using the LSCM microscope and calibrated microbeads. This procedure in itself can be difficult and may lead to pitfalls. This is beyond the scope of this thesis. However, if the measurements were biased, the bias correction procedure previously mentioned could be applied and the data further analysed reliably.

The percentage of southbound clusters is also found to be correlated (R = 0.74), but the lower value of R is due to the higher variability in cluster detection for CD34/CD164 data than for the CD34 epitope data. The measurements are distributed similarly in the population (P > 0.05).

The polarity Z is also found to be well correlated (R = 0.98) and the measurements are distributed similarly in the population (P > 0.05).

6.3.4. Polarity analysis

Table 6.5 displays the statistics (mean ± SD) for all cell populations analysed. The first row contains the measurements performed on a population of CD34$^+$ cells selected using anti-CD34 moAbs (section 2.3.1). Due to the strong magnetic field of the selection column, the CD34 antigen was predominantly moved on one side of the cells, generating an artificial XY polarisation. This dataset was used to compare the polarity measures with the other sets. The third row of the table

Table 6.5. Polarity analysis.

sample	antigen	polarity	
		Z polarity	XY polarity
miniMACS CD34-selected CD34+ cells	CD34	53.92 ± 14.90	1.70 ± 0.47
CD34+ CD164+	CD34	50.31 ± 3.88	1.21 ± 0.16
	CD164	49.85 ± 1.96	1.36 ± 0.09
CD34Neg CD164+	CD164	53.52 ± 1.33	1.51 ± 0.20
1	CD34 ep. I	50.95 ± 3.98	2.03 ± 0.63
	CD34 ep. II	53.00 ± 8.10	1.73 ± 0.49
2	CD34 ep. I	49.75 ± 13.77	1.99 ± 1.67
	CD34 ep. II	47.00 ± 10.50	2.07 ± 1.95
3	CD34 ep. I	47.389 ± 7.65	1.68 ± 0.62
	CD34 ep. III	51.28 ± 9.03	1.82 ± 0.77

displays the results obtained on a sample of $CD34^{neg}CD164^{+}$ cells. The rest of the table contains the results for the cell populations previously described and analysed. Bold numbers indicate polarity ratios considered to represent polarisation (i.e. polarity $Z > 50\%$ and polarity $XY > 1.5$).

6.3.4.1. MiniMACS CD34-selected CD34^{+} cells

The miniMACS CD34-selected CD34^{+} cells have both polarity Z and XY higher than the thresholds used to indicate preferential polarisation. Amongst all polarity Z values contained in the table, this cell population has the highest Z polarity ratio (and also the highest corresponding SD). Furthermore, the polarity XY is higher than 1.5, which indicates that most cells are predominantly polarised in the XY plane. These two values numerically confirm the visual examination of the LSCM images from which the artificial polarity was initially suspected. They also provide, at the cell population level, an estimate of the magnitude of the ratios in polarised populations (bearing in mind that some cells would have much higher or lower ratios than others).

6.3.4.2. CD34^{+}CD164^{+} cells

The CD34^{+}CD164^{+} cell population has the CD34 distributions slightly polarised (polarity Z = 50.31) while CD164 is slightly less polarised (polarity Z = 49.85). So the Z polarity is not very significant for these cells. Their XY polarity is low (both antigen distributions have a polarity XY<1.5). Both these polarity values are

the lowest measured on all cell samples. They also are the only one of the collection not to indicate significant polarisation. Therefore, it appears from this sample that cells with the $CD34^+CD164^+$ phenotype do not significantly polarise, either in the XY plane or along the Z axis, under these experimental conditions.

6.3.4.3. $CD34^{Neg}CD164^+$ cell

In contrast, the CD164 distributions from $CD34^{Neg}CD164^+$ cells (from the same sample, same experimental preparation, same microscope slide) showed a significant polarisation along the Z axis (polarity Z = 53.52 ± 1.33, with variance much lower than for the miniMACS CD34-selected $CD34^+$ cells. Significant polarisation in the XY plane was also measured (polarity XY = 1.51). Therefore it appears that completely different polarisation patterns are observed between $CD34^+CD164^+$ and $CD34^{Neg}CD164^+$ cells. The difference of polarity Z measured for the CD164 antigen from both $CD34^+CD164^+$ and $CD34^{Neg}CD164^+$ cells was statistically significant (P<0.001), while the polarity XY measured was not (P>0.05).

6.3.4.4. Class I, II and III epitopes from $CD34^+$ cells

The three samples of the epitope datasets display various polarisation patterns. All XY polarity values are of the same magnitude or higher than the value measured on miniMACS CD34-selected $CD34^+$ cells (i.e. around 1.7 and above), which

indicates clear, strong tendency of these CD34 epitopes to polarise preferentially in one direction of the plane.

Sample 1 was measured to be polarised both in the XY plane and along the Z axis for both class I and II epitope distributions, class II having a stronger Z polarity while class I having a higher XY polarity. Sample 2 is not particularly polarised along the Z axis. Sample three shows that class III epitopes are polarised towards the contact point with the slide while class I is not (so in that instance, class I behaved similarly to class I from sample 2).

Finally, the XY polarity was compared with t-tests as applied in section 6.3.1 for the comparison of the features measured from the ADM. No significant differences were observed, so the XY polarity values were distributed similarly amongst the cell populations.

6.3.4.5. Comparison of Z polarity with percentage energy ratio

In order to assess the involvement of the antigen clusters in the polarity towards the contact point of the cell with the slide, the polarity Z ratios (or "mass" ratios) were compared to the percentage of the distribution energy polarised towards the contact point (detailed graphs can be found in Appendix D). The distribution energy was defined in section 5.2.4.2. An energy ratio higher than 50% indicates that the antigen clusters are more concentrated in the southern hemisphere than in the northern, therefore possibly having a biological involvement in cell adhesion. In order to test this hypothesis, the class I, II and III epitope data were used to compare the polarity Z and the energy ratio. The results are presented in table 6.6

Table 6.6. Comparison of Z polarity with percentage energy ratio.

samples	antigens	% mass ratio (in south hemisphere)	% energy ratio (in south hemisphere)
sample 1	epitope I	50.95 ± 3.98	51.52 ± 14.18
	epitope II	53.00 ± 8.10	52.42 ± 9.71
sample 2	epitope I	49.75 ± 13.77	52.42 ± 15.10
	epitope II	47.00 ± 10.50	48.86 ± 13.51
sample 3	epitope I	47.389 ± 7.65	39.28 ± 12.48
	epitope III	51.28 ± 9. 03	42.59 ± 12.72

(mean ± standard deviation). The polarity Z results were already compared in section 6.3.4.

The comparison of polarity Z measurements with energy ratios revealed that there is no difference between the distributions of class I and II CD34 epitopes on $CD34^{class\ I+}CD34^{class\ II+}$ cells (sample 1 and 2). However, the Z polarity was higher than the energy ratio for class I and III CD34 epitopes from $CD34^{class\ I+}CD34^{class\ III+}$ cells in sample 3 (P=0.003 and P=0.005 for class I and III respectively). Furthermore, the energy ratios clearly indicate no polarisation of the cluster in the cell south hemisphere. Therefore, the slight polarisation observed for the class III epitope is mostly due to the halo than the clusters. For the class I epitope, the Z polarity does nor reflect polarisation along the Z axis, and the energy ratio further confirms it, indicating that the antigen clusters are located mainly in the northern pole.

In the next section, the energy ratios are compared for class I and II between the different samples, as described in section 6.3.1.

6.3.4.5.1. Class I CD34 epitope

Class I CD34 epitope was not found to have significant energy ratio difference between sample 1 and 2. However, a significant difference was measured for this epitope class in sample 3 in comparison to samples 1 and 2 (P=0.037 and P=0.035 respectively). For samples 1 and 2 (which are both $CD34^{class\ I+}CD34^{class\ II+}$) the energy ratio are around 50%, therefore not clearly indicating polarisation towards the contact point. For sample 3 (for which cells are both $CD34^{class\ I+}CD34^{class\ III+}$) the energy ratio are around 40%, indicating that the class I CD34 clusters are predominantly located in the northern hemisphere. For this sample, the clusters do not contribute highly to cell adhesion.

6.3.4.5.2. Class II CD34 epitope

No significant difference was observed between samples 1 and 2. The energy ratios indicate a contribution of the clusters to the polarisation for sample 1 (52.42 ± 9.71) while they indicate the opposite for sample 2 (48.86 ± 13.51). These results tend to confirm that CD34 polarisation is not specific in these experiments.

6.3.5. Colocalisation analysis using correlation of quantitative measurements performed at the z- and subcellular level

This section presents the results obtained from the correlation analysis described in section 5.2.1 based on the ADMs. Firstly, the total fluorescence per z-scan level was calculated for one antigen distribution of the pair, and is compared to the similar measures obtained from the other ADM. The correlation coefficient was calculated and stored for each ADM pair analysed. Secondly, the slope of the

Table 6.7. Correlation analysis using correlation of quantitative measurements performed at the Z- and sub-cellular levels.

Antigen combinations	Z-scan level		ADM level (subcellular level)	
	Pearson's correlation coefficient R	Bias coefficient a	Pearson's correlation coefficient R	Bias coefficient a
CD34 - CD164	0.94 ± 0.06	0.61 ± 0.07	0.84 ± 0.07	0.54 ± 0.11
CD34 ep. I-II (sample 1)	0.32 ± 0.47	0.35 ± 0.63	0.55 ± 0.22	0.57 ± 0.22
CD34 ep. I-II (sample 2)	0.47 ± 0.48	0.71 ± 0.88	0.68 ± 0.14	0.65 ± 0.14
CD34 ep. I-III (sample 3)	0.52 ± 1.17	0.54 ± 0.19	0.53 ± 0.14	0.55 ± 0.18

regression line was also calculated as it may indicate the presence of a bias between both ADMs. Results are presented in table 6.7. The slope of the regression line is named "bias coefficient a". Similar procedure was applied for the intensity correlation at the subcellular level.

6.3.5.1. CD34⁺CD164⁺ antigen distributions

The correlation measured between CD34 and CD164 from the CD34$^+$CD164$^+$ cell

sample at the z-scan level is the highest in the whole table (0.94). At this level of

observation, high values indicate that both antigen distributions are colocalised

and follow the hypothetical linear relationship described by equation 5.15, or, that

if the linear relationship is not observed, the antigen clusters are small in size and

high in number, so their spatial distribution is random. In this case, measuring the

total intensity per z-level would be correlated between both ADM (because these

values would vary with changes of cell shape mostly) but it would not reflect any

colocalisation. Consequently, it is relevant to further perform correlation

measures at the subcellular level. At the subcellular level, the correlation is still

very high (0.84), which confirms that true CD34/CD164 colocalisation is taking

place in the HSC membrane. This second value is lower than the one obtained at

the z-level, due the differential expression of the antigen densities, particularly in

high intensity regions (see section 6.3.7).

The bias coefficients indicate that CD164 fluorescence is between 55-60% of the

intensity of the CD34 fluorescence. Without calibration of the fluorescence, it is

not possible to assess if this observation has a biological significance or is due to

experimental bias.

6.3.5.2. Class I, II and III CD34 epitope distributions

In all three samples, the correlation coefficients are much lower than for the

CD34$^+$CD164$^+$ cell. They vary in the range 0.32-0.52 for the Z-level. This

indicates a lack of strong colocalisation. In contrast to CD34$^+$CD164$^+$ cells, they vary on the range 0.53-0.68 at the subcellular level of observation, therefore being higher than for the z-level. This small improvement is due to local colocalisation in some parts of the cell membrane, were the hypothetical linear relationship (equation 5.15) is approximately followed. The lack of higher values can be due to the fact that the epitopes are weakly colocalised (which is surprising since they are meant to represent different parts of the same CD34 molecule, and should therefore be strongly colocalised), or that the antigen are colocalised but the hypothetical linear relationship does not take place (and therefore this is a case of differential expression, see figure 1.14). Further investigation is required to decide which hypothesis is true (sections 6.3.6 and 6.3.7).

The biases are all below 1, which indicates that the second epitope of the combination is expressed as a percentage of the fluorescence of the first epitope.

6.3.5.3. General comment on the biases

Because it has been noticed that the bias values were always below 1 (so less fluorescence was recorded each time when imaging the second antigen distribution), it may be possible that the binding ability of the antibodies used in the labelling of the second antigen may be affected by the presence of already labelled cell membrane. In such a case, the biases are the consequence of the experimental labelling protocol.

Figure 6.3 shows graphs obtained using some the previously described measurements in order to potentially investigate CD34 function (details on possible interpretation in legend).
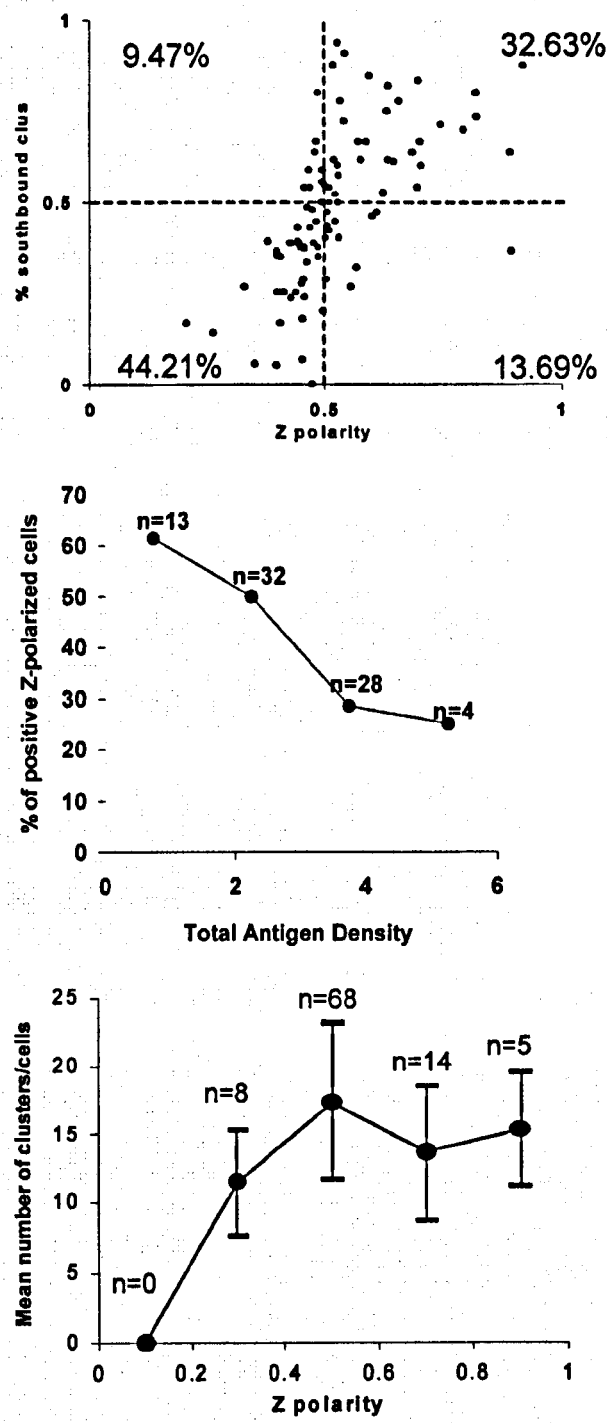
Figure 6.3: **Analysis of the CD34 antigen clusters.** *(legend next page)*

Figure 6.3: **Analysis of the CD34 antigen clusters.**

Top – Percentage of southbound clusters vs. Z polarity. This graph has been plotted using all CD34 distributions from the CD34 epitope datasets. The graph is divided in four quadrants. The top-left quadrant represents CD34 distributions whose antigen clusters were predominantly polarised towards the contact point with the slide (% southbound clusters > 0.5, or 50%) while the whole antigen distribution (clusters + halo) was not (Z polarity < 0.5). Such distributions only represent less than 10% of all. The top-right quadrant is important: it represents CD34 distributions whose both antigen clusters and whole distribution were predominantly polarised towards the contact point. About a third of the distributions fall in this quadrant, indicating that overall 33% of the cells were polarised towards the contact point, and that CD34 clusters were involved in this polarisation. The bottom-right quadrant represents distributions whose antigen clusters were not recruited towards the contact point while the whole antigen distribution was. This does not have a great occurrence, less than 14% of total distributions. Finally, the bottom-left quadrant represents distributions for which both the antigen clusters and whole antigen distribution were not polarised at all, indicating that over 44% of the cells where more polarised to their north hemisphere than south. From this, it is calculated that 42.1% of the cells have their CD34 clusters towards the contact point, 46.32% have their whole antigen distribution polarised in the south hemisphere. Furthermore, for 70.45% of Z polarised cells, clusters are involved in the polarisation. From these figures it is concluded that CD34 distribution do not polarise significantly when the cells are adhered on gold-coated slides.

Middle – Proportion of Z polarised cells vs. antigen density per cell. It is observed that as the antigen density per cell increases, the fraction of cell polarised to the contact point decreases from around 60% to less than 30%, so it undergoes over a 50% reduction. This indicates that cells with lower CD34 content use it predominantly in polarisation activity, and as the density increases, less CD34 is used for polarity purpose and is also distributed elsewhere around the cells. This suggests that CD34 have a predominant role in cell polarisation, but also have potentially other roles since it is not used for polarisation at higher density. Extrapolating this to the XY polarisation, instances of this observation are illustrated in figure 7.13, where the right cell (low CD34 density) is clearly polarised (but towards the adjacent cell, not the contact point with the slide) while the left cell (high CD34 density) uses some of its CD34 to the site of contact with the other cell, and recruits CD34 in other structure such as the crescent and the protrusion.

Bottom – Mean number of clusters per cell vs. Z polarity. As the Z polarity increases, the average number of CD34 clusters per cell increases and reaches a plateau at polarity Z = 0.5. However, such tendency is not very strong due to the overlap of the measurements.

6.3.6. Colocalisation analysis using the watershed segmentation of equatorial z-sections

The results obtained from the watershed segmentation (section 5.2.3.2) are presented in this section and commented. The analysis has been restricted to the CD34/CD164 antigens from $CD34^+CD164^+$ cells, and to class I and II CD34 epitopes from $CD34^{class\ I+}CD34^{class\ II+}$ cells. $CD34^{class\ I+}CD34^{class\ III+}$ cells were not analysed as the watershed segmentation process did not work properly due to the poor contrast and image quality of the LSCM images from this set.

Table 6.8 presents the results of the analysis. In each watershed basin, three measurements were performed: the mean fluorescence intensity (after each image was normalised in the range 0-255), the Pearson's correlation coefficient R and the bias coefficient. For each equatorial z-section, the standard deviation of these

Table 6.8. Colocalisation analysis using the watershed segmentation of equatorial Z sections.

Measurements from watershed basins		Antigen combinations		
		CD34 - CD164	CD34 ep. I-II (sample 1)	CD34 ep. I-II (sample 2)
Fluorescence Intensity	mean intensity	31.75	21.80	8.84
	mean intensity standard deviation	16.10	14.49	9.30
	standard deviation of mean intensity	10.51	11.47	5.56
Pearson's correlation coefficient R	mean R	0.86	0.61	0.69
	mean R standard deviation	0.03	0.09	0.13
	standard deviation of R	0.05	0.24	0.21
Bias coefficient a	mean a	1.28	1.04	0.56
	mean a standard deviation	0.11	0.26	0.13
	standard deviation of a	0.46	0.63	0.26

three features measured amongst the watershed basins is measured and averaged at the population level (mean intensity standard deviation, mean *R* standard deviation, mean *a* standard deviation). The standard deviation of the collection of features measured at the population level is also measured and referred in the table as "standard deviation of (feature)". All these measurements may be of value for comparing bigger cell populations analysed under different experimental set ups (e.g. different concentrations of a drug), where they could help spotting the difference in measures and quantitatively discriminate between different antigenic distributions. For example, in this table, the fluorescence measurements are provided to demonstrate that such measures can be performed and analysed, but are of no value in this colocalisation analysis.

6.3.6.1. Pearson's correlation coefficient

The correlation values for the $CD34^+CD164^+$ cells and $CD34^{class\ I+}CD34^{class\ II+}$ cells (samples 1 and 2) respectively are 0.86 ± 0.05, 0.61 ± 0.24 and 0.69 ± 0.21. They compare remarkably well with the values obtained at the subcellular level of observation using the ADM (section 6.3.5), where they were found to be equal (in the same order) to 0.84 ± 0.07, 0.55 ± 0.22 and 0.68 ± 0.14. This good match suggests that performing correlation analysis directly on the ADM is a valid approach. Furthermore, it is probably better to do so since the watershed analysis takes only account of the equatorial z-section (which represents around 10% of the data in most case). Moreover, the correlation performed on watershed basins takes almost as long as the full processing of the LSCM Z-series (including computation of the ADM).

Both sample 1 and 2 of the epitope data have a higher mean R standard deviation than CD34/CD164 distributions. This reflects a higher variability of the R values for the epitope, per analysed Z-scan, variability which is between 3 and 4 times the one for CD34/CD164. Such higher variability also supports the idea that CD34 epitopes are weakly colocalised and differentially expressed, as suspected by the results in section 6.3.5.2.

6.3.6.2. Bias coefficients

The bias coefficients for the $CD34^+CD164^+$ cells and $CD34^{class\ I+}CD34^{class\ II+}$ cells (samples 1 and 2) respectively are 1.28 ± 0.46, 1.04 ± 0.63 and 0.56 ± 0.26. They do not compare with the values obtained at the subcellular level of observation using the ADM, where they were found to be equal (in the same order) to 0.54 ± 0.11, 0.55 ± 0.22 and 0.65 ± 0.14. This is due to the normalisation step which stretches the measured range of intensities in the range 0-255. However, if class I and II CD34 epitopes were expected to have similar membrane distributions, the bias should be roughly similar between the samples. The fact that the bias measured on sample 1 is twice the one measured on sample 2 further emphasises the differences between these samples.

6.3.7. Colocalisation analysis using the ADM

Figure 6.4 presents the results of the colocalisation investigation using the approach described in section 5.2.3.3. This approach aimed to measure the

deviation from the linear hypothesis (equation 5.15) of the antigen distributions on dual labelled cells. The top graphs represent the relative proportion of dual antigens (combination indicated above the graphs) colocalised with each other. They show the curve $P^T_{Antigen\,1} - P^T_{Antigen\,2}$ from equation 5.17 as a function of the threshold T. The points indicate the mean value in the cell population and the solid lines above and below are means ± SD. In the case of perfect linear colocalisation of both antigens, the curve should be close to zero for any T. Deviation from 0 indicates that the hypothetical linear relationship from equation 5.15 is not respected (differential colocalisation or lack of colocalisation). Great difference can be observed between the first two graphs, for which the same epitope combination was used. This reflects differences in colocalisation patterns between both cell populations. Interestingly, $CD34^{class\ III+}CD164^+$ distributions display a similar pattern to the first epitope graph, but it also shows much less variability for low density clusters.

The middle graphs show the colocalisation ratio obtained using equation 5.18 (means ± SD). Perfect linear colocalisation would involve that the curves stay around value 1. The dashed line indicates the level where 50% of the amount of antigen 1 and 2 are linearly colocalised with each other, while the remaining 50% for each antigen is expressed on its own or colocalised differentially. For both class I and II CD34 epitope samples, the colocalisation ratio falls below this limit in the region of low density clusters. This reveals that most CD34 clusters contain predominantly diverse proportions of each epitope (differential or no colocalisation). For $CD34^{class\ III+}CD164^+$ distributions, the colocalisation ratio falls below the 50% threshold in the region of high density clusters, indicating that in such clusters, still a bit less than 50% of each antigen linearly colocalised

with the other. Overall, the graph for this antigen combination reveals that $CD34^{class\ III+}CD164^{+}$ have a much lower differential colocalisation than CD34 epitope combinations. This is further emphasised in the bottom graph, which shows the mean colocalisation behaviour between low and high density clusters for the three cell populations. A polynomial fit has been added for each curve and the equations are shown to the right of the graph. The last term in the equations represent the antigen colocalisation ratio in low density clusters while this ratio in high density clusters in indicated on the far right. In denser clusters, the $CD34^{class\ III+}CD164^{+}$ population still had 44% of both CD34 and CD164 colocalised. Comparison with the corresponding top-left curve revealed that CD34 (class III) was predominantly involved in the colocalisation process. $CD34^{class\ I+}CD34^{class\ II+}$ populations had 35% and 26% of class I and II colocalised in such clusters (i.e. 80% and 60% of the $CD34^{class\ III+}CD164^{+}$ colocalisation respectively). For one of the two $CD34^{class\ I+}CD34^{class\ II+}$ populations, class I was predominantly involved in colocalisation while both epitopes were, on average, equally involved in this process. In summary, this analysis revealed that $CD34^{class\ I+}CD34^{class\ II+}$ cells express a higher degree of differential colocalisation pattern than $CD34^{class\ III+}CD164^{+}$ cells.
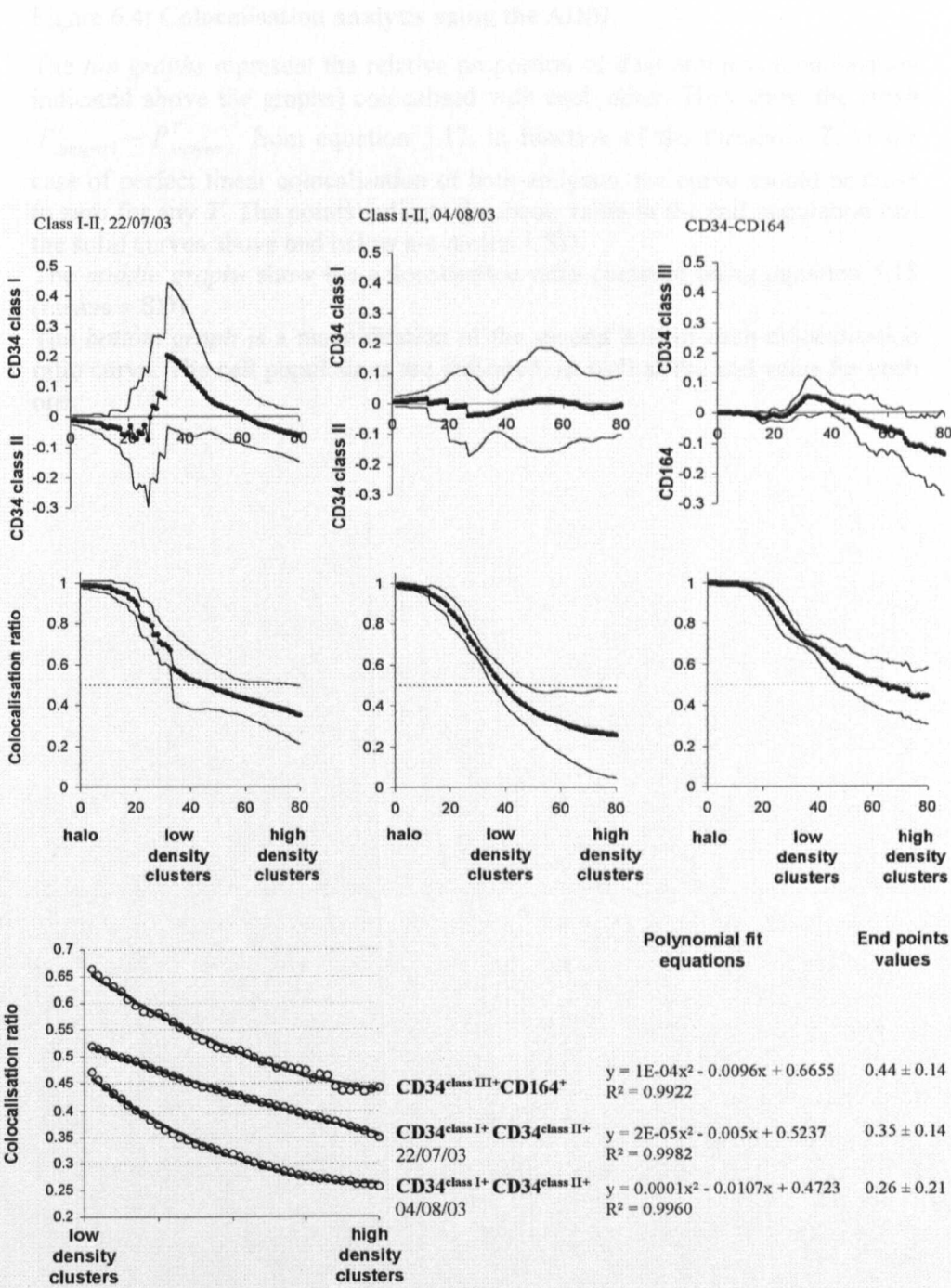
Figure 6.4: **Colocalisation analysis using the ADM.** (*See next page for legend*)

Figure 6.4: **Colocalisation analysis using the ADM.**

The *top graphs* represent the relative proportion of dual antigens (combination indicated above the graphs) colocalised with each other. They show the curve $P^T_{Antigen\,1} - P^T_{Antigen\,2}$ from equation 5.17, in function of the threshold T. In the case of perfect linear colocalisation of both antigens, the curve should be close to zero for any T. The points indicate the mean value in the cell population and the solid curves above and below are means ± SD.

The *middle graphs* show the colocalisation ratio obtained using equation 5.18 (means ± SD).

The *bottom graph* is a magnification of the second half of each colocalisation ratio curve. The cell populations are indicated, as well as the end value for each one.

6.4. Summary of the quantitative analysis of antigen distribution in HSPC membrane

The quantitative analysis from the ADMs and watershed segmented LSCM images showed very different patterns for the expression of the different antigens in HSPC membranes. These differences seem to be depending on the cell phenotype. The cell populations analysed had the following phenotypes:

- $CD34^{class\ III+}CD164^{+}$

- $CD133^{+}CD34^{Neg}CD164^{+}$

- $CD133^{+}CD34^{class\ I+}CD34^{class\ II+}$

- $CD133^{+}CD34^{class\ I+}CD34^{class\ III+}$.

The expression patterns of CD133 could not be assessed as this antigen was used during cell selection and isolation process, in order to ensure the immaturity of the selected population. However, some LSCM datasets for dual labelled CD133+CD34+ cells were still available from an early experiment (documentation not available) and could be used solely for the visualisation applications (ADM and 3D reconstruction) with interesting results. The most quantitatively significant results are summarised below.

6.4.1. Mean cell fluorescence

The mean intensity of CD164 is lower than for the other antigens (p<0.01). Class III CD34 epitope for this cell population ($CD34^{class\ III+}CD164^{+}$) is also less

expressed than in the CD34$^{class\ I+}$CD34$^{class\ III+}$ population (p<0.01). For the class II epitope, one of the two CD34$^{class\ I+}$CD34$^{class\ II+}$ population showed a significant lower expression (p<0.01). However, when comparing each class to each other, no difference in the distribution patterns were significant, as reported in literature [Lanza *et al*, 1999; Maynadié *et al*, 2002]. This pattern was

fluorescence class I<class III<class II

(2.62<2.68<3.20).

6.4.2. Analysis of clusters

6.4.2.1. Number of clusters

On the CD34$^{class\ III+}$CD164$^+$ population, each antigen displayed on average the same number of clusters per cells. Similarly, on CD34$^{class\ I+}$CD34$^{class\ III+}$ cells, the number of clusters was not different. Nevertheless, there were significant differences for other cell populations. Between the two CD34$^{class\ I+}$CD34$^{class\ II+}$ populations, one sample had 80% more class II clusters than the other, i.e. an average of 8.3 clusters (p<0.001). Out of the three CD34class I+ populations, two samples had over twice more class I clusters than the third one (p<0.001). One of these two sample was CD34$^{class\ I+}$CD34$^{class\ II+}$ and the other was CD34^{class} $^{I+}$CD34$^{class\ III+}$. No differences were observed when comparing each class to each other (no discrimination between samples). The average pattern for this collection was

number of clusters class III>class I and class III>class II.

6.4.2.2. Cluster fluorescence

Within all HSPC samples, the mean fluorescence from the antigen clusters was not significant. Therefore the mean cell fluorescence differences observed for some samples and antigens are the consequence of weaker intensity clusters and/or low intensity halos.

6.4.2.3. Clusters in cell south hemisphere

The proportion of clusters in the southern hemisphere of the cells (facing adhesion site) are found to be dispersed around the 50% limit of neutrality, with a tendency for some samples to indicates that most clusters are, to a small extend, more concentrated in the north pole of the cell (% southbound clusters<50%), which possibly suggests no involvement of the clusters. However, the differences are not significant. Furthermore, when comparing the CD34 epitope classes to each other indicated a tendency of the clusters to concentrate more on free pole of the cell than in the south hemisphere.

6.4.3. Polarity analysis

6.4.3.1. Z polarity

A great variability in Z polarity was observed amongst the samples. Dual labelled distributions did not show significant differences between the corresponding antigens. The polarised (and non polarised) populations also expressed high SD, so overall the cells had Z polarity distributed around the 50% limit of neutrality. The high SD reflected the heterogeneity in the polarisation ability of HSPCs, possibly related to the fact that these populations were not homogeneous themselves. Nevertheless the CD164 antigen on CD34$^{class~III+}$CD164^{+} cells was Z polarised while it was not on CD34$^{class~IIINeg}$CD164^{+} cells ($p < 0.001$). The patterns observed in this study were

CD34^{+} MACS selected cells : polarised

CD34$^{class~IIINeg}$CD164^{+} : polarised

CD34$^{class~III+}$CD164^{+} : both antigens non polarised

CD34$^{class~I+}$CD34$^{class~II+}$: class I and II polarised for one of the two samples (class II>class I)

CD34$^{class~I+}$CD34$^{class~III+}$: class III polarised, class I non polarised.

6.4.3.2. Z polarity and proportion of southbound clusters

Great variability amongst the samples were observed, but overall no significant differences were found (the proportion of southbound clusters is similar to the Z polarity). For one sample of the epitope data, both the percentage of southbound

clusters and the Z polarity indicated a weak tendency for both to polarise towards the contact point.

6.4.3.3. XY polarity

$CD34^{class\ III+}CD164^{+}$ cells were not XY polarised despite the fact that $CD34^{class\ IIINeg}CD164^{+}$ cells were polarised. All other populations were strongly polarised despite high variability intra- and inter-samples. Artificially polarised MACS selected cells had an average polarity value of 1.70, which gave an indication of the polarity magnitude to be expected in polarised populations. All polarised samples had polarity distributed around this value. Amongst the dual labelled samples, only one (of the two) $CD34^{class\ I+}CD34^{class\ II+}$ population was both X and XY polarised for both antigens, but different patterns were observed for each antigen: class II had higher Z polarity than class I, while class I had higher XY polarity than class I.

6.4.3.4. Comparison of Z polarity with energy ratio

Only the epitope datasets were analysed for this comparison. No differences were observed for $CD34^{class\ I+}CD34^{class\ II+}$ cells, where both polarity and energy ratios were distributed around the 50% limit of neutrality. For $CD34^{class\ I+}CD34^{class\ III+}$ cells, for each epitope, the polarity ratios were higher than the energy ratios (p=0.003 and p=0.005 for class I and III respectively). The energy ratio for class I epitope in this population was significantly lower (ratio = 40%) than class I in

CD34class I+CD34class II+ cells (p<0.05), so most clusters were found in the north pole of the cells.

6.4.4. Colocalisation analysis

6.4.4.1. Colocalisation analysis using fluorescence correlation at Z and subcellular levels

Overall, much higher correlation values were obtained for $CD34^{class\ III+}CD164^{+}$ cells than for $CD34^{class\ I+}CD34^{class\ II+}$ and $CD34^{class\ I+}CD34^{class\ III+}$ cells. Linear colocalisation was observed for $CD34^{class\ III+}CD164^{+}$ cells as high correlation values were observed at the Z level, and slightly lower values at the subcellular level. For this population, analysis of the bias coefficients showed that the CD164 fluorescence is about 55-60% the fluorescence for CD34. Differential colocalisation was observed for $CD34^{class\ I+}CD34^{class\ II+}$ and $CD34^{class\ I+}CD34^{class\ III+}$ cells where low values were measured at the Z level, while they slightly increased at the subcellular level, barely reaching a significant value (R=0.68) for one of the $CD34^{class\ I+}CD34^{class\ II+}$ population.

6.4.4.2. Colocalisation using watershed segmentation

The correlation values measured in the watershed basins segmenting clusters in equatorial Z sections compared very well with those obtained from the ADM analysis at the subcellular level, for both $CD34^{class\ III+}CD164^{+}$ and $CD34^{class}$

$^{I+}$CD34$^{class\ II+}$ populations. And similarly, a much higher variability was observed for the values measured on the CD34 epitope LSCM images.

6.4.4.3. Colocalisation using the ADM

The CD34$^{class\ III+}$CD164^{+} and both CD34$^{class\ I+}$CD34$^{class\ II+}$ populations all expressed different relative colocalisation patterns. The percentage of linearly colocalised antigens is close to 100% in low antigen density regions while it decreases steadily and to different extents for the three populations, indicating different degrees of differential colocalisation. In high density clusters, the CD34$^{class\ III+}$CD164^{+} population still had 44% of both CD34 and CD164 colocalised, CD34 (class III) being mostly involved in the colocalisation process, while CD34$^{class\ I+}$CD34$^{class\ II+}$ populations had 35% and 26% of class I and II colocalised in such clusters (i.e. 80% and 60% of the CD34$^{class\ III+}$CD164^{+} colocalisation respectively). For one of the two CD34$^{class\ I+}$CD34$^{class\ II+}$ populations, class I was predominantly involved in colocalisation while both epitopes were, in average, equally involved in this process. These figures revealed that CD34$^{class\ I+}$CD34$^{class\ II+}$ cells express a higher degree of differential expression than CD34$^{class\ III+}$CD164^{+} cells.

Chapter 7

Discussion

Chapter 7. Discussion

In this section, the methodology developed for the automated analysis of LSCM images is first commented. Some figures are also presented in order to illustrate how the results from the application of this methodology could be used for biomedical investigations. Great use was made of the new MSA algorithm, based on linear diffusion, after it was shown to segment fluorescence accurately. Further applications of this algorithm are also investigated and briefly outlined in the second section of this chapter. The measurements and observations from the processed HSPC LSCM datasets and their potential implications in the context of haemopoietic theories are also discussed.

7.1 LSCM image processing methodology

New approaches were introduced to process and analyse LSCM datasets. The application of these methods could easily be extended to the analysis of other datasets.

LSCM images contain background and signal values overlapping over a wide range of intensities. Other automatic techniques and global thresholding algorithms do not provide satisfactory segmentation results on images with such a strong unimodal histogram. The first stage therefore consisted in automating accurate fluorescent signal segmentation in LSCM images. This step prevents manual or heuristic subjective thresholding which otherwise may introduce early bias, possibly altering the subsequent analysis of the fluorescence signal. Three global thresholding algorithms were developed. Consistent results were found

when comparing the outputs of these algorithms. The MSA algorithm based on linear diffusion relies on a novel approach, i.e. the analysis of histogram behaviour as the image is processed through the scales, and it provides results in close agreement with appropriate manual segmentation. Therefore the MSA algorithm was subsequently used for the segmentation of all datasets used in this work. It also operates much faster than the region-based watershed segmentation algorithm.

Once the raw images were segmented and signal area correctly identified, fluorescence quantification as performed. Two common approaches, area or intensity measurements, were compared. The area method (used in many studies) was demonstrated not to be robust, while the intensity method provided much more consistent results. In laser scanning confocal microscopy, the intensity measured per pixel reflects an average photon count per local voxel. This average count itself reflects different concentrations of the fluorochrome, and consequently different concentrations of the labelled antigen of interest. Quantifying antigen concentration or density using intensity measurements was therefore a more sound approach. Intensity measurements are subject to potential processing in order to improve the quantification. Commonly, an improvement of the measure is performed through baseline correction (mean background intensity subtracted to mean fluorescence intensity). We confirmed that this method improves the correlation coefficient between fluorescence measures from independent methods (or users). Moreover, a new quantification formula was proposed and compared to the usual background subtraction method. The correlation of independent measurements was globally further improved, and the behaviour of the inter-method bias became more comparable. This suggests that

this new formula could further help reducing the bias between different methods, contributing to improve the quality of the comparison of experimental results.

Segmented and quantified fluorescence distributions were stored in a compact, yet meaningful form, the Antigen Density Maps (ADMs). The ADM were particularly appropriate for storing and displaying the fluorescence patterns on HSPC membranes as HSPC were naturally quite spherical. They facilitated the visual comparison of large LSCM datasets of antigen distribution patterns in the membrane of individual cells. Additional visualisation tools were also implemented in order to assess the spatial relationship of antigen distributions in HSPC clusters (3D spherical models, fast 3D point cloud reconstruction of the actual distributions). They revealed to be particularly useful for demonstrating the presence of different types of CD34 clusters at contact point of adjacent cells, such as single clusters, patch-like clusters and meridian-shaped clusters.

The ADM were further used to quantitatively characterise the antigen distributions through various measurements, such as polarity ratios, correlation

Sparse small clusters   Patch-like clusters   Meridian-shaped crescent
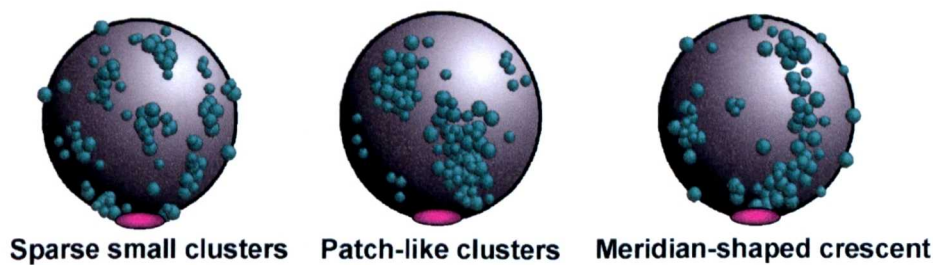
Figure 7.1: **Types of antigen clustering patterns observed on HSPCs.**

Such antigen clusters are frequently observed in HSPC membranes. They express both linear and differential colocalisation patterns. Sparse small cluster are often found associated with membrane protrusions. Patch-like clusters are predominantly observed at cell-to-cell junctions. Meridian-shaped crest-like (MSCL) structures are found on both individual and clustered cells, sometimes at cell-to-cell junctions (but not necessarily).

coefficients, colocalisation ratios. These measurements could be used for graphical display of the data characteristics (e.g. polarity graphs), which allowed a quick visual inspection of feature distributions for different cell populations, and for statistical analysis of the different cell populations (figures 7.2, 7.3, 7.4 are examples of such applications).
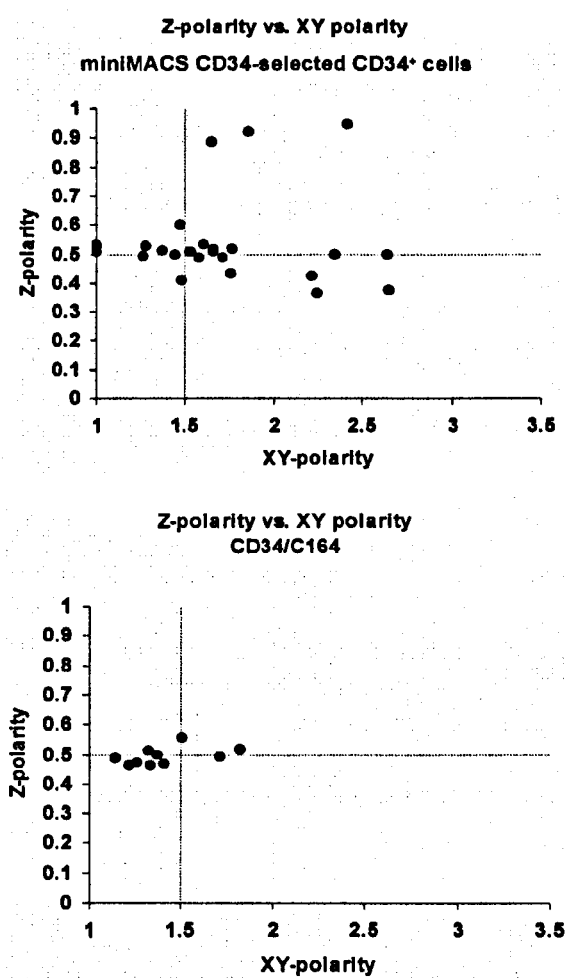
Figure 7.2: Comparison of polarity patterns between miniMACS CD34-selected CD34$^+$ cells and non-selected CD34$^+$CD164$^+$ cells obtained from a CB MNC fraction.

This example demonstrates the great difference in polarity patterns between two samples. Dashed lines indicate polarity thresholds. For the bottom graph, no discrimination was made between CD34 and CD164 distributions, and the points are used to show that most distributions are not polarised in the XY plane, in great contrast with the top plot, where distributions are both generally polarised along the Z axis and in the XY plane.

Altogether, the features measured with this methodology allow a thorough characterisation of antigenic distributions, particularly useful and relevant for the analysis of small and rare samples, as demonstrated in this thesis with HSPCs. Indeed, the extensive analysis of such samples is a complicated task using the current biomedical methodologies (FACS, cytology examination by experts, protein assays), and often only a few measurements are obtained from the analysis of large cell populations (several thousands cells or more). The approach described in this work proposed to increase the number of quantifiable features per cell and to apply such measurements to smaller, but still statistically significant, cell populations. The characterisation of biological samples may therefore be optimised. As demonstrated, in addition to the mean cell fluorescence measured by one of the most popular tool available to characterise cell populations, the Fluorescence Activated Cell Sorter (FACS), six other quantifiable and meaningful features were automatically extracted from the LSCM image analysis (mean number of clusters per cell, mean intensity in clusters, percentage of clusters in southern hemisphere, Z polarity, XY polarity, distribution energy). Further to these measurements, quantitative colocalisation analysis methods were developed, which have the capability to characterise linear and differential colocalisation patterns.

The application of these algorithms and methods could easily be extended to the study of other cell populations, especially cancer cells. Indeed, cancer cells are abnormal cells which would normally be removed by the immune system. However, there are complex, yet not understood molecular phenomena taking place at the interface of these cells with their environment which prevent immune responses and allow cell survival. Thus quantifying the relationships of specific

molecules in cancer cell membranes with other molecules and cell types could help getting a better understanding of some of the mechanisms involved in cancer cell proliferation. Another important field in cell biology is the study of cell migration. This happens normally for many blood cells in adults, for cells during embryogenesis, and abnormally for metastatic cancer cells. Many molecules (e.g. Cell Adhesion Molecules) are cooperatively involved to promote cell migration

Pearson's correlation coefficient r, Z vs. subcellular level
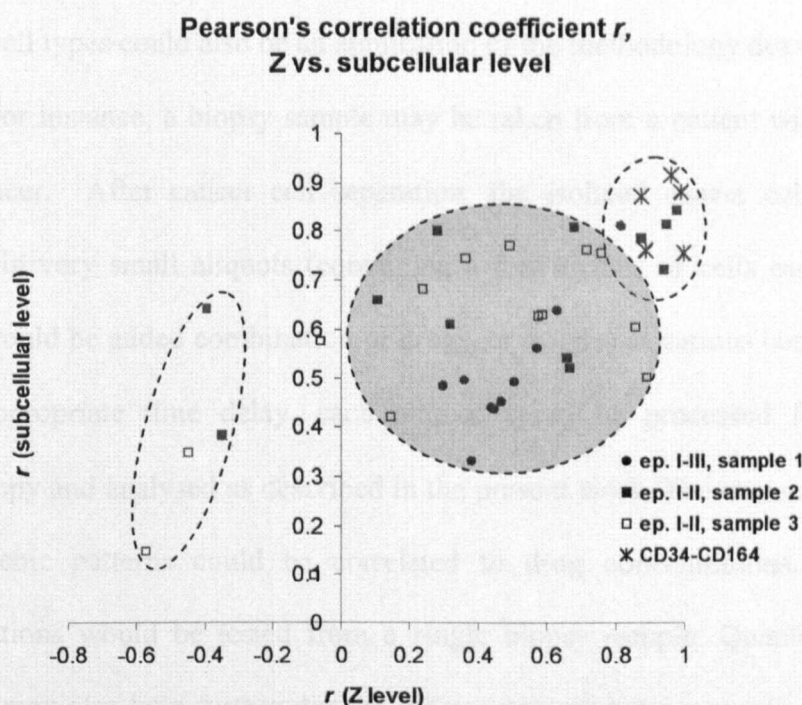
Figure 7.3: **Correlation coefficients for colocalisation analysis. Z-level vs. subcellular level.**

Correlation coefficients calculated using the ADMs (see section 5.2.1). Linearly colocalised distributions (those following equation 5.15) have their correlation coefficients clustered in the top right corner of the figure. More extended scattering of the points indicates diverse degrees of differential colocalisation. CD34/CD164 from $CD34^+CD164^+$ cells are clustered in the top right corner, reflecting the high linear colocalisation of these distributions. Conversely, if some of the CD34 epitope distributions are also linearly colocalised, most of them show various degrees of differential colocalisation or even poor correlation (negative coefficients).

on/through appropriate biological substrates and tissues. Disorders and unpaired antigenic expression of such molecules, to various degrees, are responsible for medical complications. Polarity measures of the membrane antigen distributions could further help discriminating between cell types, or could be correlated to impaired or excessive biological responses (lack of, or stimulated migration) [Peled *et al*, 2002]. Such measures could be compared to those obtained from migration experiments currently used (Dunn chamber systems, trans-well systems, etc., commented in [Jeon et al, 2002]). Quantifying the response to drugs for certain cell types could also be an application of the methodology described in this thesis. For instance, a biopsy sample may be taken from a patient with a (rare or not) cancer. After cancer cell separation, the isolated cancer cells could be divided in very small aliquots (containing a few dozens of cells each), each of which would be added combination of drugs, or drug(s) at various concentrations. After appropriate time delay, each aliquot could be processed for confocal microscopy and analysed as described in the present work. The measured changes in antigenic patterns could be correlated to drug concentrations, and many combinations would be tested from a single biopsy sample. Quantifying these features may also help further discriminating between heterogeneous populations which are currently considered to be homogeneous (such as the known example of $CD34^+$ cells, which actually encompass many sub-populations – some discriminated on the basis of the expression of other membrane antigens – with different proliferation and differentiation potentials). The discrimination would not be performed on the basis on the antigenic profile but rather on the features measured from antigen distributions, e.g. different polarity patterns may identify different sub-populations with identical antigenic expressions.

In summary, a new automated methodology was developed to process and analyse LSCM images. This approach does not depend on human users, so no subjective biases are introduced in the analytical process. Consequently, these methods facilitate the comparison of experimental data from different users, providing that the biological protocols were comparable. Furthermore they extend the quantification possibilities currently performed in real biomedical situations by increasing the number of features measured per cells, providing deeper insights on cell biology. In the light of more powerful laser scanning confocal microscopes available today [Gerlich *et al*, 2001; May, 2004], with programmable automated scanning paths on the slide and image acquisition, a full automation of the procedure – from image acquisition to statistical analysis – is potentially a realistic future possibility.

Figure 7.4: **Degree of homology between class I, II and III CD34 epitopes distributions calculated from the features measured on the ADM.**

The features used for these homology measurements are presented in table 6.2. Overall, six features were available. The left image shows a schematic of the molecular structure of CD34 with the location of the different epitopes indicated. The degree of homology between class I and II, and class I and III was calculated using all six features computed from the ADM. It indicates that class I and II distributions were 69% similar while class I and III distributions were only 22% similar. Therefore there are much greater difference patterns between class I and III than between class I and II. This is further emphasized when calculating the degree of homology using only three of the features, i.e. the antigen density, Z and XY polarity. Class I and II epitopes are shown to have even more similar expression patterns (83% homology) while class I and III patterns are more dissimilar (18% homology). If the same structural form of the CD34 molecule was to be found in the membrane of any HSPC, the degree of homology between epitope classes should be very high. Therefore these results indicate that the CD34 molecule is expressed under various isoforms, with structural differences, in the HSPC membrane. It could suggest that the presence (in the membrane of single HSPC) of different proportion of each CD34 isoform allows the modulation of the function of CD34 and its effect on single cells.

7.2 Potential applications of the MSA algorithm

7.2.1 Segmentation of other types of unimodal images

The MSA algorithm presented in this work makes use of the analysis of the diffusion rates of grey levels in scale spaces. Its implementation is computationally efficient compared to more conventional scale space approaches. Thus it a good candidate for real time application in the machine vision arena. Furthermore, histograms of scale space images are analysed at different scales. The approach adopted in this work uses the detection of an intensity range were high activity of histogram convergence rates are observed by measuring the median rate values down the scales. This was found to correctly segment images with strongly unimodal histograms [Baradez *et al*, 2004, Baradez *et al*, 2003]. Consequently, it was also applied to other types of images expressing such histograms encountered in research and industrial applications, for instance bright field microscopy, microarrays, electrophoresis gels, scanned text documents, speckle fringe patterns or edge images. For these images, the unimodality is due to the predominance of the background over the relevant signal with the addition of noise. Examples of such images are presented in figure 7.5, with the corresponding binary masks extracted using the MSA algorithm. Despite these images having different signal areas, different spatial organisation of these areas and different intensity ranges, all masks obtained seem reasonably accurate. Therefore the MSA algorithm is thus potentially interesting for segmenting a wide range of unimodal images and could be used to get an initial "guess" on the position of ROIs. Furthermore, the information contained in the rate matrix R

(equation 3.10) could be exploited in different ways, as briefly described at the end of this section.
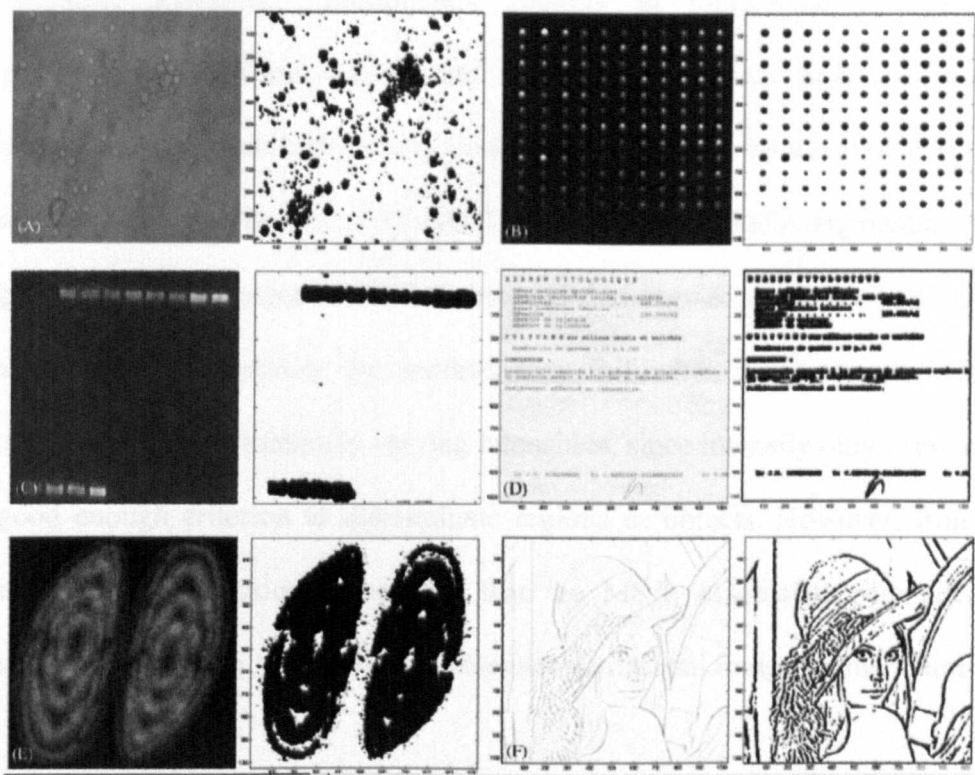
Figure 7.5: **Binary masks of different types of unimodal images obtained with the MSA algorithm.**

A – Wide field microscope image (the high frequency image of this picture was used to generate the mask), B – simulated microarray, C – electrophoresis gel, D – scanned text document, E – simulated interferometry fringe pattern, F – edge image of Lenna.

7.2.2 Iterative segmentation of natural plurimodal images

Preliminary experiments were performed to investigate the potential application of the MSA to images with plurimodal histograms. Such an attempt iteratively applied the algorithm to plurimodal images. After one initial application step, a threshold was calculated and used to separate two subsets of image regions (those above and below the threshold). For each subsets, pixels in the black regions

(indicating which subset was removed) were attributed an intensity equal to the value of the threshold. The MSA was then re-applied to both subsets. The number of iterations is chosen beforehand but it was noticed that a few iterations provide meaningful segments. Consequently regions in plurimodal images were segmented into different classes, and it was observed that such classes often corresponded to semantically sound (in the visual sense of the term) objects. Such results are presented in figures 7.6 and 7.7 where iterative MSA segmentation was applied to natural images. It can be observed that segmentation is meaningful in many instances. However, the method may fail when images contain fuzzy, blurred regions of monotically varying intensities, since intensity only is no longer a good enough criterion to discriminate regions or objects. However, from this preliminary investigation, it appears that the MSA, associated with adequate image pre-processing, could perform relevant plurimodal image segmentation.
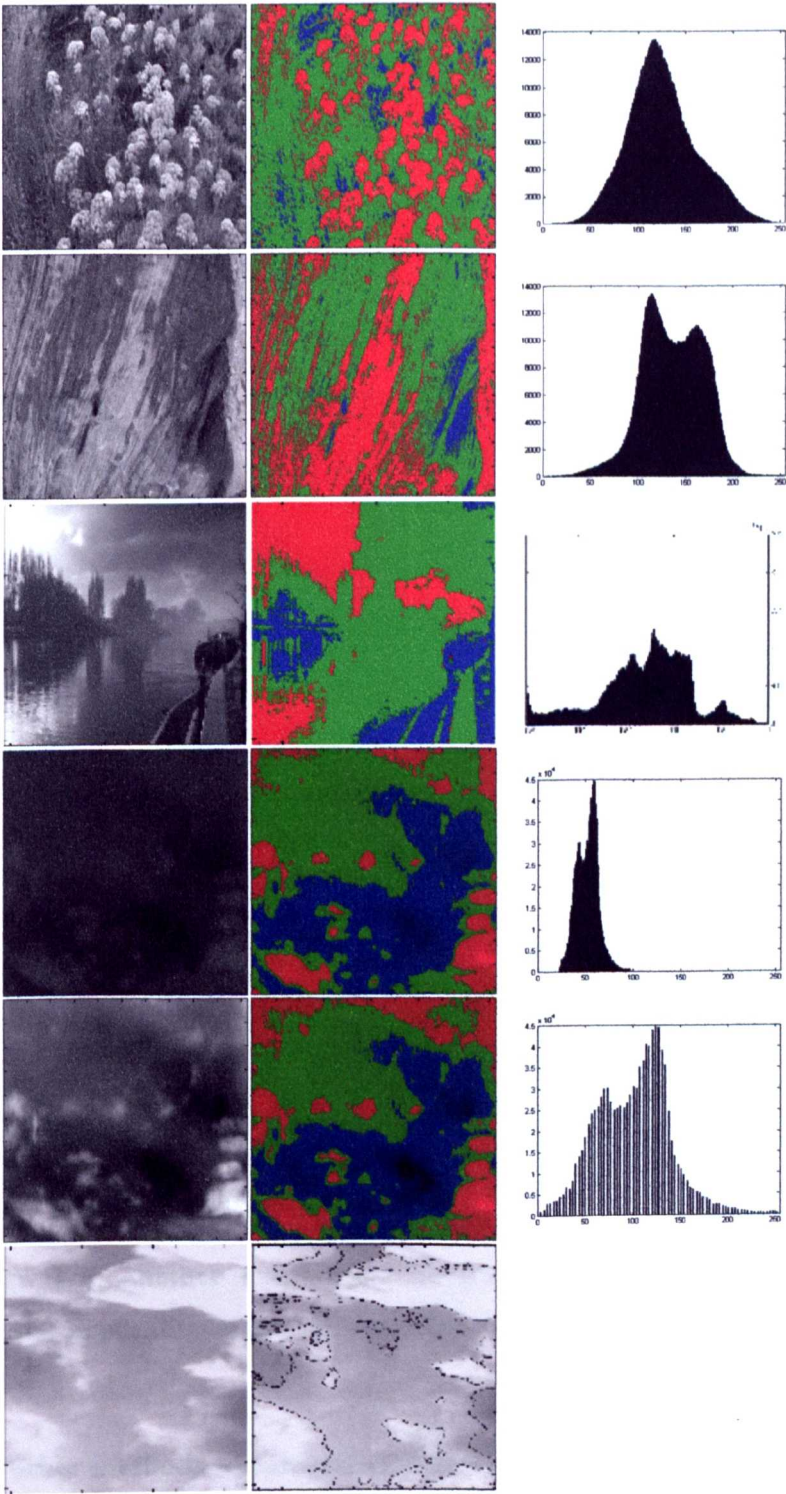
Figure 7.6 : **Example of natural image segmentation by iterative application of the MSA algorithm.** *(legend next page)*

Figure 7.6: **Example of natural image segmentation by iterative application of the MSA algorithm.**

Left – Original images. *Middle* – Images segmented into three classes. *Right* – Histograms of the original images. The third row is an example of inappropriate segmentation as grey levels do not correspond to homogeneous and distinct objects. Fourth to sixth rows are cloud images. The fifth image is the brightness/contrast adjusted image of the one above. Despite very different dynamic ranges, the segmentation is very similar.

Figure 7.7: **Another example of natural image segmentation by iterative application of the MSA algorithm.**

7.2.3 Two other methods to calculate a threshold from the rate matrix

Finally, since other methods could be used to exploit the information from the rate matrix, another approach (Approach 1) was briefly investigated to calculate the

binarisation threshold of plurimodal images. First, the Shannon entropy associated

with each rate curve along the scales in R (equation 3.10) was measured and

summed for each scale to produce the curve E. It appeared that some scales

contained very little information/entropy, particularly in the corresponding

histogram regions of equilibrium dynamics. These scales typically have an

associated entropy lower than the value 0.0002. Therefore this value was used to

threshold the matrix R and to retain only elements above this threshold. Such

elements are predominantly located at the lower scales where convergence of the

histograms eventually collapses to the average image intensity, but some may be

found at higher scales, depending on the structure of the original image. The

information contained in the thresholded rate matrix was used by summing the

elements down the scales (i.e. the columns). The curve C obtained presents

characteristics similar to the one obtained by taking the median values down the

columns of the original R, notably one predominant peak located at or close to the

mean intensity. Due to the thresholding effect, this curve contains mostly zero-

value elements. Experiments showed that locating the position where two

consecutive zeros were first located along the curve C either before or after the

peak provided a visually good threshold. For natural images, it was noticed that

segmented regions were very stable under Gamma coreection within wide range

of values, in comparison to Otsu's thresholding (figure 7.8).
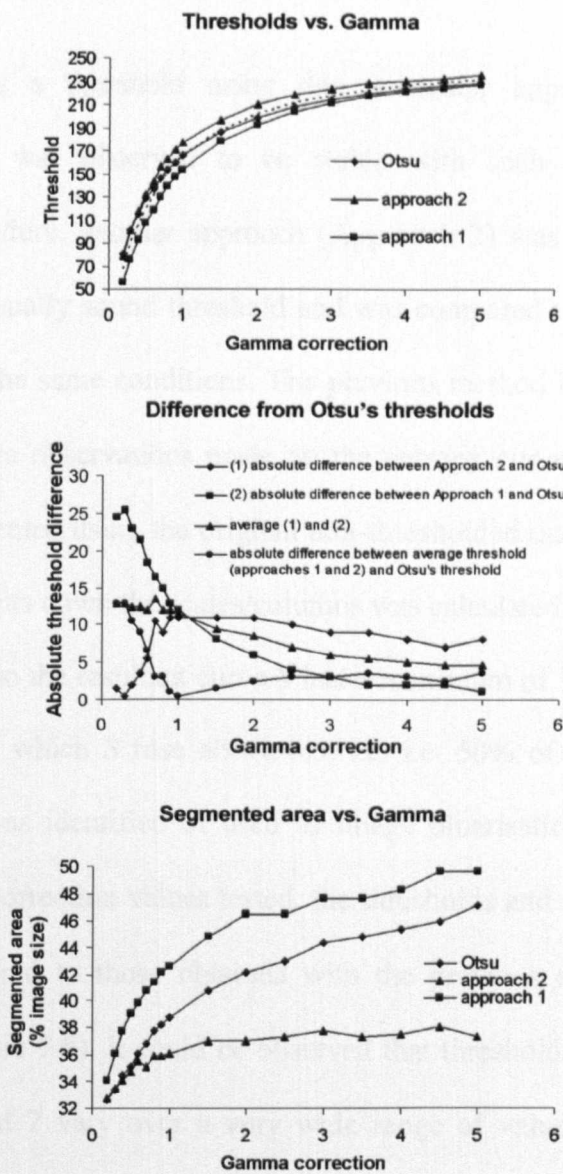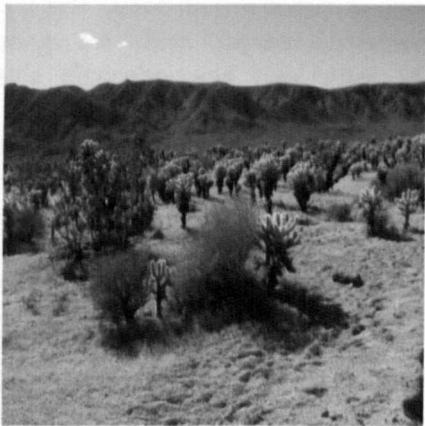
Figure 7.8 : **Potential application of the MSA algorithm to plurimodal image segmentation.** *(see next page for legend)*

Figure 7.8: **Potential application of the MSA algorithm to plurimodal image segmentation.**
Top – Original natural image for which Gamma correction was applied prior to segmentation using Otsu's algorithm and two other MSA-based approaches (described in text). *First graph* – Thresholds obtained by the three approaches vs. Gamma correction values. As a consequence of Gamma correction, the thresholds vary over a wide range of values, but Otsu's thresholds are bounded by those obtained from both MSA-based approaches. The dashed line indicates the average values of both MSA-based thresholds. *Second graph* – Absolute difference of MSA-based thresholds with Otsu. *Third graph* – Segmented area size vs. Gamma parameter. It is noticed that the second approache provides stable segmented areas (i.e. the same areas are segmented, whatever the magnitude of Gamma correction), which makes this approach more robust than the popular Otsu's method.

Since computing a threshold using this particular approach and particular landmark on C was observed to be stable with such a common intensity adjustment procedure, another approach (Approach 2) was also implemented to locate another visually sound threshold and was compared to this one and Otsu's threshold under the same conditions. The previous method involved thresholding R using subjective observations made on the entropy curve E, so an alternative could be implemented using the original non-thresholded rates R. The cumulative sum of the elements down the scales/columns was calculated and normalised to its maximum value so the resultant curve S had a maximum of 1. The position of the first intensity for which S rose above 0.5, i.e. i.e. 50% of the total sum of the elements in R, was identified of used as image binarisation threshold. For the various Gamma correction values tested, the thresholds and segmented area sizes were also compared to those obtained with the previous approach and Otsu's thresholding (figure 7.8). It could be observed that thresholds obtained from both Approaches 1 and 2 vary over a very wide range of values (from 60 to 230), follow a similar pattern than Otsu's thresholds over the Gamma values tested (from 0.1 to 5) and keep very close to Otsu's threshold. It was also observed that

thresholds from Approaches 1 and 2 envelop Otsu's threshold, and the average threshold from these approaches is close to Otsu's thresholds. A more detailed analysis of these measurements was carried out by calculating the absolute distance between them and Otsu's results (figure 7.8). Approach 2 is very close to Otsu's thresholds in the Gamma range [0.1, 1] where absolute distances gradually increase in the range [0, 11]. Above Gamma = 1, Approach 1 is closer than Approach 2, as the absolute distances decrease in the range [11, 1]. The stability of these distances is more apparent for Gamma values above 1, and despite Approach 1 being closer to Otsu than Approach 2, Approach 2 keeps a more stable distance (varying in the range [11-8] in comparison to [11, 1] for Approach 1). It is noticed that both Approaches 1 and 2 have absolute distances varying quickly over a higher range of values for Gamma in the range [0.1, 1] while they evolve more monotically for Gamma>1. However, the average absolute distance of both approaches to Otsu varies much more monotically over all Gamma values and remains in the range [13, 5]. Finally, the size of the segmented image regions is plotted in percentage of the total image size in figure 7.8. It can be seen that regions segmented using Approach 2 remain very similar despite large variations of the Gamma correction parameter (i.e. the same regions are segmented, which represent roughly 36-37% of the image area) despite being smaller than Otsu's segmented regions.

These potential further developments on image segmentation using the MSA algorithm may indicate that more consideration could be given to this algorithm since it seems to have the potential to be used to segment plurimodal images along with unimodal images.

7.3. Biological implications of the results

7.3.1. Antigen expression levels and colocalisation

The study of HSPC membrane antigens using the algorithms and methodologies developed in this thesis indicated statistically significant differences between different HSPC populations. While the number of antigen clusters and their antigen content were similar between these populations, colocalisation analysis showed different patterns between $CD164^+CD34^+$, $CD34^{class\ I+}CD34^{class\ II+}$ and $CD34^{class\ I+}CD34^{class\ III+}$ populations. Indeed, it was expected that CD34 epitopes would present higher degrees of colocalisation in comparison with $CD164^+CD34^+$ cells, since epitopes are simply different parts of the same molecule or variants (obtained through RNA splicing or enzymatic processing of a bigger template) of this molecule. Despite it was not *in situ* image analysis, it was reported using FACS studies that CD34 epitope expression levels were very similar on HSPC populations [Lanza *et al.*, 1999; Steen *et al.*, 1996], further strengthening the hypothesis that high colocalisation should be expected. However, the present LSCM image analysis demonstrated that CD164 and CD34, despite being different molecules, had a higher degree of colocalisation than CD34 epitopes. While in all cases the colocalisation is high in halo and low antigen density membrane regions (i.e. weak fluorescent clusters), it decreases as more antigens are concentrated in the clusters. Such a decrease reflects a differential colocalisation of the antigen, meaning that two antigens both located in the same region of the cell membrane have their relative expression levels which varies from clusters to clusters (in case of linear colocalisation, the relative proportion of both antigens would remain rather constant for different clusters, whatever the

expression level). Differential colocalisation was observed for all HSPC populations analysed and are discussed here in the context of haemopoietic theories.
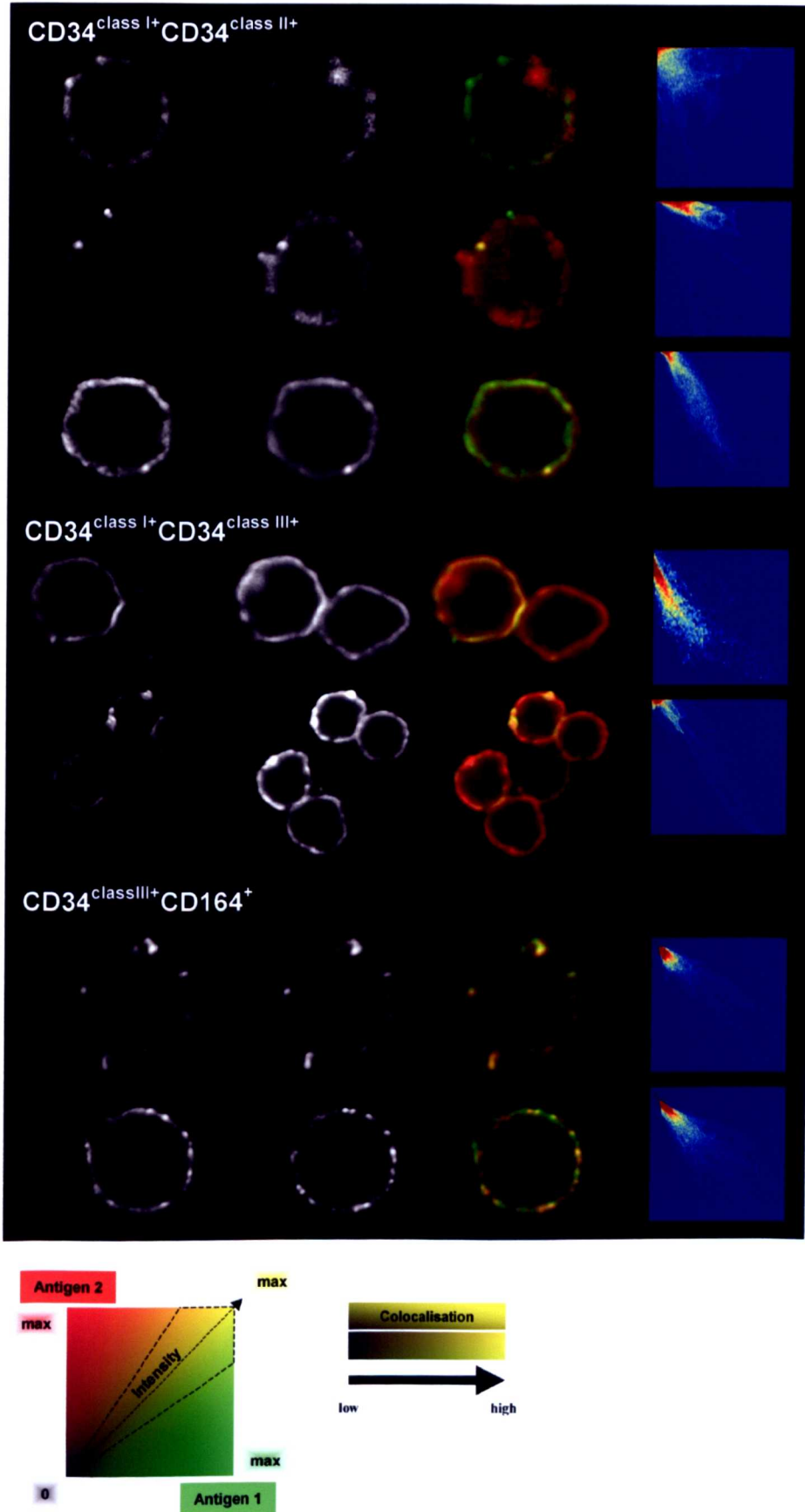
Figure 7.9: **Example of automatically re-aligned optical sections for CD34^{class} $^{I+}$CD34$^{class\ II+}$, CD34$^{class\ I+}$CD34$^{class\ III+}$ and CD34^{+}CD164^{+} cells.** *(legend next page).*

Figure 7.9: **Examples of automatically re-aligned optical sections for** $CD34^{class\ I+}CD34^{das\ II+}$ **cells,** $CD34^{das\ I+}CD34^{class\ III+}$ **cells and** $CD34^{+}CD164^{+}$ **cells.**

On the right of the figure are represented the 2D histograms of the merged images. The origin is situated in the top left corner. The rows indicate the first antigen of the combination (e.g. $CD34^{class\ I+}$ for $CD34^{class\ I+}CD34^{class\ II+}$ cells) while the columns indicate the second antigen. Well colocalised distributions should follow the diagonal joining the top left corner to the bottom right corner (which is the case for both $CD34^{+}CD164^{+}$ cells presented here). The merged images for the CD34 epitope distributions show various degrees of colocalisation, which are reflected in the 2D histograms. The second $CD34^{class\ I+}CD34^{class\ II+}$ cell illustrates that CD34 epitopes do not necessarily colocalise consistently. Indeed, most of the CD34 distribution around the cell is due to class II epitope while class I is poorly represented, except in two bright clusters (top and left of the $CD34^{class\ I+}$ distribution). The left cluster is very well colocalised with a class II epitope cluster, but the top cluster contains only class I epitope. The first $CD34^{class\ I+}CD34^{class\ II+}$ cell seems to have almost non-colocalised distributions, while the third cell, despite expressing more class I epitope, colocalises well with class II epitope.

7.3.2. Antigen clusters

The presence of antigen clusters is thought to reflect some biological functions taking place at the interface of the cells with their environment. Cluster formation is not well understood [Kansas, 1996] and several processes were found to be involved. Small clusters can be formed by aggregation of halo antigens *via* diffusion in the cell membrane. In this scenario, freely diffusing antigens are clustered where some activity is taking place in particular membrane regions, or microdomains (domains generally ≤ 0.1 µm width), and nowhere else. Membrane and cytoskeletal molecules attached to the inner side of the membrane are involved in the process. Some clustering processes were reported to involve fyn kinase, glycosylphosphatidyl inositol-anchored proteins Thy-1 and F3 that selectively co-patch with Reggie molecules after antibody cross-linking activation [Stuermer et al, 2001]. This triggers substantial colocalisation of Reggie–1 and –2 with Thy-1, GM1, T-cell receptor complex and fyn. In this example, such clusters are thought to participate to the formation of signal transduction centres. The merging of adjacent microdomains may be responsible for the creation of bigger clusters. However, as the cluster size increases, other membrane molecules also participate to cluster formation.
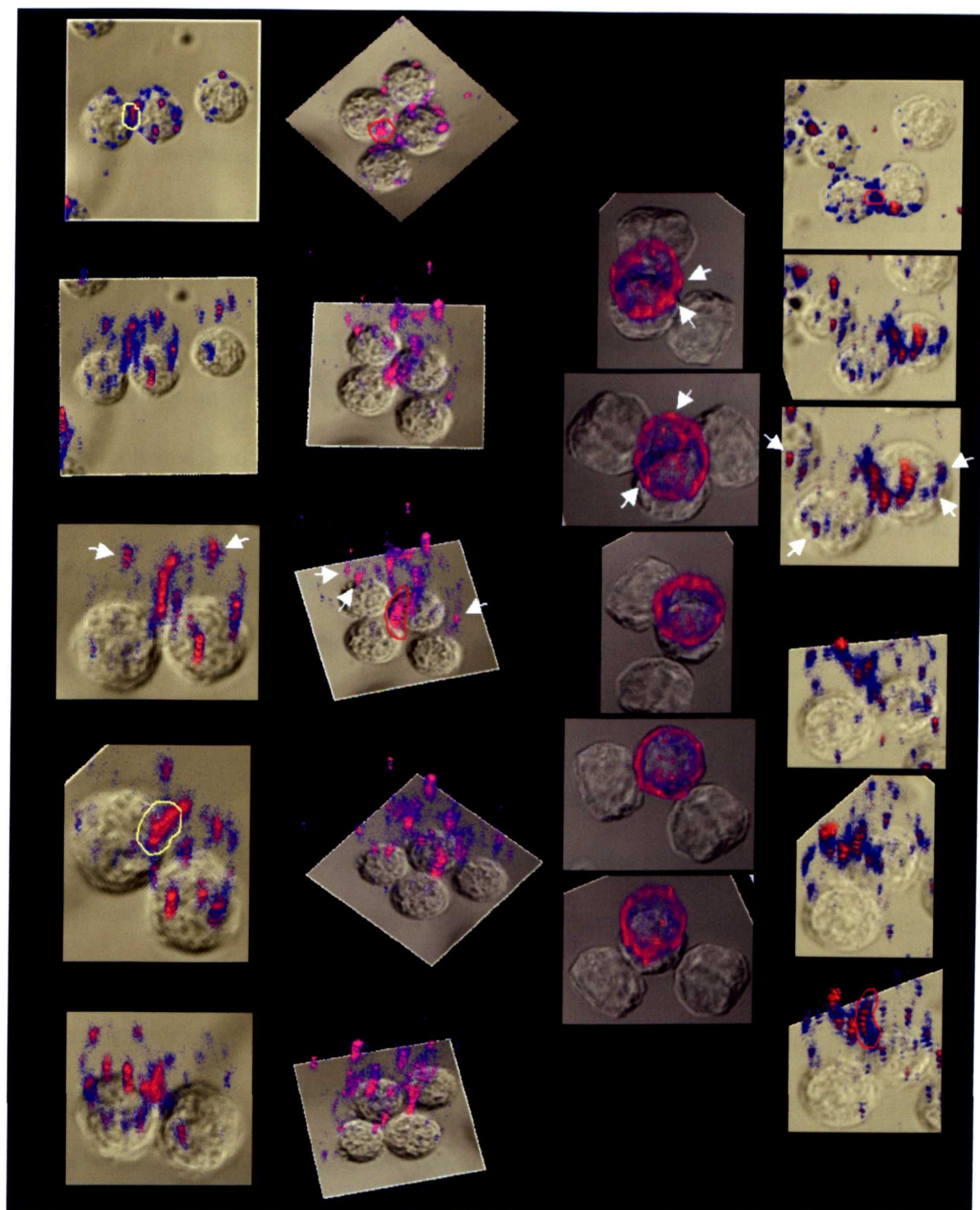
Figure 7.10: **Example of point cloud 3D reconstructions of CD34 distributions on clustered CD133$^+$ cells.**

White arrows point to small clusters, yellow curves enclose patch-like clusters and red curves enclose meridian-shaped crest-like clusters. These two latest types of clusters are located at cell-to-cell junctions.

7.3.3. Lipid rafts

Lipid rafts are a typical example of molecules involved in big cluster formation. These plasma membrane lipid microdomains sequester signalling proteins and are located to the outer leaflet of the membrane lipid bilayer. They are coupled to microdomains in the inner leaflet of the membrane, which are activated and initiate signalling cascades when lipid raft proteins bind to their ligand. The classification of lipid rafts according to their chemical composition is not well established, and the composition of some of them is still unknown (particularly in the inner leaflet). However, it is known that other molecules such as transmembrane adaptor proteins are also involved with lipid rafts and that the actin cytoskeleton is involved in raft migration and coalescence, despite the actual mechanisms being unknown.

These rafts have been found to have important functions, as particularly well observed for cells of the immune system, and noticeably as polarised distributions [Fassett *et al*, 2001]. These cells do perform many tasks that stem cells perform, such as rolling, tethering, transendothelial migration, adhesion, or cell-to-cell communication. Amongst these functions, lipid rafts were suggested to be important for the modulation of signal transduction [Gupta and DeFranco, 2003; Leitinger and Hogg, 2002] and cell adhesion [Yanagisawa et al, 2004; Wu et al, 1997], both critical functions for the HSPCs investigated in the present study. Under the light of these comments, lipid rafts may well play a key factor in antigen cluster formation and function in HSPC membranes.

7.3.4. Theoretical mechanisms for antigen cluster formation in HSPC membranes

Our group previously suggested that antigen clusters may reflect membrane activity and be responsible for performing some unknown functions. Here is proposed a theoretical formation mechanism which could explain the antigen expression patterns observed in the HSPC membrane, backed up by the quantitative measurements developed and reported in this work.

7.3.4.1. Intracellular compartment

DIC visual image analysis showed that HSPCs have a very large nucleus that occupies most of the space inside cells. Only a tiny fraction of cytoplasm can be visually identified, mostly found on one predominant side of the nucleus. Therefore it could be hypothesised that protein production, which requires the presence of rough and smooth endoplasmic reticulum in the cytoplasm, would take place mainly on these intracellular regions where enough cytoplasm is present. Indeed, early intra-cellular labelling of CD34 (figure 7.11) showed that internal CD34 molecules were located in such cytoplasmic pockets, within one compact sub-volume of internal cell compartments. The antigen produced in this small volume must then be migrated to the cell surface where their external structures are used for ligand recognition while their internal structures may be involved in signal pathways.
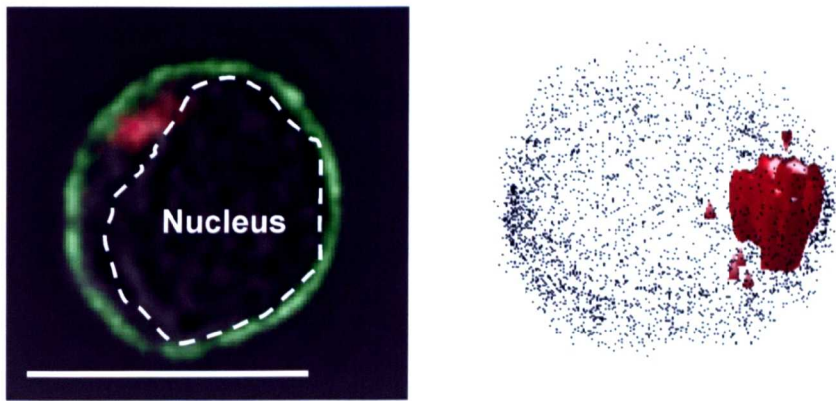
Figure 7.11: **Internal/external CD34 distributions.**

Left – HSPC DIC image overlaid with internal (red) and external (green) CD34 fluorescence (intensities adjusted for visualisation). This HSPC is at a very early stage as more mature cells do not express internal C34. *Right* – 3D reconstruction of the corresponding Z series. The cloud point technique applied to membrane CD34 shows quite a faint uniform halo in the cell membrane. The isosurface encloses the volume where internal CD34 is expressed. It is located in a single small pocket between the nucleus and the membrane. The pocket is 2-3 μm width and 5 μm height. Bar = 10 μm.

7.3.4.2 Antigen halo and low density clusters

The antigen halos observed in the cell membranes were not experimental artefacts due to membrane autofluorescent molecules since they were not observed on negative cells. In some rare instances, the halo was totally missing from some membrane regions while expressed on some other regions. It was also noticed that the bigger the clusters, the weaker the halo, and some cells labelled for CD34 epitopes displayed highly dense clusters while no halo was present. These regions (i.e. halo, loose weak clusters) contain antigens (CD34, CD164) which are not yet recruited to clusters.

In the process of cluster formation, the first step for HSPCs is to express the antigens in the cell membrane. Since the antigens are generally concentrated internally on one side of the cell, their transport to cell surface is also likely to be polarised. However, cell surface polarisation of antigens is only reported for very specific functions (adhesion, migration, cell-to-cell contact events), so initial (polarised) externalisation of HSPC antigens should lead to uniform surface expression, potentially achieved via passive diffusion of the molecules within the membrane since this mechanism does not involve energy consumption and eventually leads to uniform distribution. This would be consistent with uniform halos observed on weakly clustered cells. Moreover, this is also consistent with the measurements which indicated that low antigen density regions of cell membranes display linear colocalisation. Indeed, following the externalisation process just described, all membrane regions should express the same relative proportions of the different antigens, a signature of linear colocalisation as observed on figure 6.4. At the first step of antigen recruitment to microdomains, small amounts of locally present antigens would simply follow an aggregation process, during which linear colocalisation is still to be expected (which is observed indeed) to a certain degree.

7.3.4.3 High density and big clusters

High density clusters contain much more antigens than surrounding regions. Some of them also have a patch-like structure covering an extensive membrane surface.

For the smaller of these dense clusters, an aggregation process of the primary microdomains may still be active. As aggregation increases, local differences of antigen densities may be amplified, which would explain why such clusters start displaying a deviation of linear colocalisation towards more differential colocalisation patterns. The production of such dense clusters, particularly the patch-like clusters, is likely to involve other membrane and cytoplasmic molecules, as observed with lipid raft structures. These unknown mechanisms would be responsible for the aggregation of smaller dense clusters, further increasing differential colocalisation. Figure 5.14 shows a composite image of an LSCM cross-section of a HSPC dual labelled for CD34 and CD164. Watershed segmented clusters were analysed for colocalisation and this composite image obtained by combining the correlation coefficients between antigen expression levels per cluster and the biases (proportionality coefficients) between these levels. Colour-coded, it can be observed that extended, often adjacent, domains of the membrane share similar colours, therefore sharing similar formulations of their antigen contents. This further supports the theoretical mechanisms for cluster formation proposed in this discussion. This is also consistent with the measures indicating an increased of differential colocalisation with the cluster densities.

7.3.4.4 Energy use by HSPC

The cluster formation using the suggested mechanisms described above requires energy consumption from the cell. Once the clusters are formed, no more energy is needed in this process, except if these structures have to be moved or further clustered within the membrane. Experiments investigating antigen formation and

cellular responses (in terms of functions, such as adhesion or shape remodelling) showed that $CD34^+$ progenitor cells can sustain these responses for about 30 minutes to one hour, but no functional responses are observed over this time limit. It may be concluded that stem cells use little internal energy in order to survive in a quiescent state and that they optimise their energy resources to perform short-time responses before and after which long delay are available to regain this energy.

Considering these remarks, since clusters are big enough to suggest energy consumption during their creation, they well may have a functional role as the cell energy should not be wasted, role unknown to date. The large sizes of the clusters also suggest involvement of other types of molecules, possibly lipid rafts, kinases and signalling molecules after comparison with structures and functions observed of immune cells.

Our group was first to identify unusual long, thin membrane extensions (pseudopodia) extending from some of the earliest immunologically well characterised lineage negative HSPC populations ever identified in healthy human adults [Forraz et al, 2004]. Such pseudopodia can extend over to 80 μm. Some morphologically identical blood immune and cancer cells, respectively B-cells [Gupta and DeFranco, 2003] and KG1a cells, were found very recently to produce similar membrane protrusions [Francis *et al*, 1998]. Electron microscopy studies (data not shown) showed that KG1a cell pseudopodia are predominantly coated with CD34. CD34 clusters are also observed on much smaller but larger membrane protrusions, while loose distributions of single CD34 molecules are sparsely found in smooth domains of the cell membrane. KG1a cells are primitive

blood stem cells turned into leukaemic cancer cells at an early developmental stage, so they retain many features associated with normal HSPCs. Their electron microscopy study suggests that CD34 clusters are associated with membrane protrusion. This was also supported by a time-lapse confocal study of live KG1a cells labelled for membrane components [Francis *et al*, 1998]. Long, thin pseudopodia were produced and used to sense their environment and distant KG1a cells. They were produced and retracted within seconds, which involves energy consumption and fine tuning of the molecular pathways participating to the process [Yang *et al*, 2001]. Studies on $CD34^+CXCR4^+$ cell migration showed that such cells deform their bodies during migration, which involves membrane protrusions [Peled *et al.*, 2002]. After one hour, migration rates decrease, possibly due to exhaustion of the energy resources available in the cell cytoplasm.

7.3.5 Potential function of the CD34, CD164 and CD133 cluster antigens and synthesis with haemopoetic theories

How and why HSPCs leave the bone marrow is not well understood, neither is why only a fraction of these cells participate to the process at a time, while evidence suggest that HSPCs have to circulate in the blood stream in order to enter the haemopoietic process. However all circulating cells do not undergo proliferation [Pierelli *et al*, 2000] followed by differentiation and remain quiescent after homing back to the bone marrow. It is likely that molecular signals and the balance/unbalance of chemical composition in the haemopoietic microenvironment activate cell migration [Plett *et al*, 2002] and extravasation to blood vessels [Kronenwett *et al*, 2000]. It is known that adhesion is an important

process to regulate HSPC proliferation [Kröger *et al*, 1998]. Since the haemopoietic microenvironment expresses such adhesion molecules [Suehiro *et al*, 1999] and cytokines [Ueda *et al*, 2000], and since a pool of HSPCs do not proliferate and differentiate throughout life (so blood renewal is sustained), it may be hypothesised that, during the phase where HSPCs undergo transmigration to blood stream prior to bone marrow homing, the cells are not exposed anymore to the appropriate environmental adhesion molecules and therefore are more likely to either enter cell cycling activity or undergo apoptosis [Voermans *et al*, 2000]. There are some recent experimental observations that transendothelial migration of CD34$^+$ cells prevents starvation induced apoptosis [Ferrero *et al*, 2003]. Consequently, it would be important for circulating HSPCs to find quickly the appropriate adhesion molecules and microenvironment in order to remain in an immature and quiescent state [Fruehauf *et al*, 1998] or reverse to such a state if the quiescence property was lost during circulation [Huygen *et al*, 2002]. In turn, the proportion of circulating HSPCs which are not regulated quickly enough *via* adhesion enters proliferation and differentiation process [Puznel *et al*, 2003], sustained as long as the right environment and cytokine cocktails are reached (also in the bone marrow). In this model, HSPC circulation in the blood acts as a trigger point for engaging a fraction of HSPCs into the haemopoietic process.
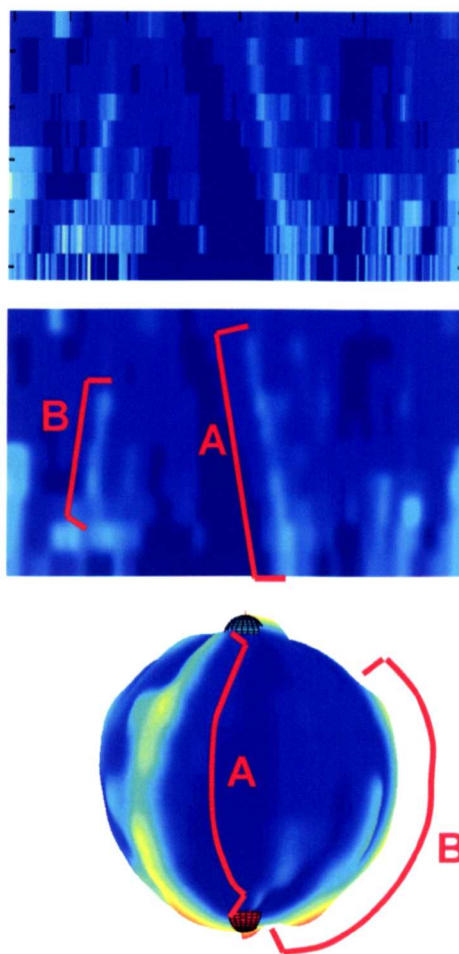
Figure 7.12: **Representative meridian-shaped crest-like structures.**

Top – Original CD34 ADM. *Middle* – Interpolated ADM. Several elongated high CD34 concentration structures are observable. The longest one (in the middle of the ADM) extends to 9 μm length, which is approximately the diameter of the cell. Such meridian-shaped crest-like (MSCL) structures are not experimental artefacts due to the PSF of the system (in which case MSCL structures would be observed strictly vertically, or would all follow similar orientation). *Bottom* – Sphere model of the CD34 membrane distribution. The MSCL structure A extends from one pole of the cell to the other, hence its name. A smaller MSCL structure B is also observed, which extends over 5 μm.

As demonstrated in this work, CD34 and CD164 are found to colocalise and to concentrate in clusters and meridian-shaped crest-like structures (figure 7.12). CD34 epitopes also colocalise to a lesser extent and express a higher degree of differential colocalisation. CD34 is thought to be involved in migration process

287

and cytoadhesion [Madjic *et al*, 1994] but its short intracellular tail does not allow proliferation signalling [Hu and Chien, 1998], and some evidence suggests that CD164 plays a role in negative control of cell proliferation. Furthermore, preliminary work showed that CD34 and CD133 are also found to colocalise in clusters, despite that colocalisation is specific to some particular rare clusters while most clusters contain either one or the other antigen (figure 5.5). No function is known for CD133, but its leucine-zipper motif and very unusual 5-transmembrane domains suggest it may serve as a signalling receptor, allowing cell-to-cell communication or environment sensing.

Given the fact that the cells analysed in this work where predominantly circulating cells, a potential function for the observed antigen clusters and their composition may be hypothesised. Indeed, further to adhesion, cell-to-cell contact/communication between HSPCs takes place and is probably important for HSPC survival. Indeed, while HSPCs are more abundant in the bone marrow, they are extremely rare in the blood stream. This may well explain why HSPCs produce within seconds unusually long membrane extensions in order to reach each others. This could allow the transmission of survival message between cells, while those who are isolated may not receive them (and consequently may undergo apoptosis). Shorter and more abundant pseudopodia are also observed which are involved in tight junctions between adjacent HSPCs and to sense the environment. They are likely to participate to cell migration, as observed with other motile blood cells (macrophages, B-cells). Electron microscopy showed that CD34 is predominantly found on membrane extension and protrusions in KG1a cells (data not shown), while it was consistently observed in the present analysis of the LSCM images coupled with the visual analysis of DIC images that antigen

clusters were also associated with short membrane protrusions, despite the fact that all clusters are not systematically found with visually observable membrane protrusions (still these are protrusions, as observed on KG1a cells [Oh *et al*, 1999; Francis *et al*, 1998]).

Antigen clusters may act as individual, functional membrane domains, containing various relative proportions of the different CD34, CD164, CD133, and other proteins which differ from cluster to clusters. Other molecules are likely to be associated to these clusters (lipids, kinases [Zhang *et al*, 2001], cadherins [Puch *et al*, 2001; van Buul *et al*, 2002], selectins, and other adhesion molecules [Leppanen *et al*, 1999; Lévesque and Simmons, 1999 ; Nakamura *et al*, 2000 ; Peichev *et al*, 2000]; Sackstein and Dimitroff, 2000]), here referred to as "cluster associated molecules", in order to perform signal transduction and adhesion functions [Dravid and Rao, 2002]. Different relative proportions of the various antigens composing the clusters, i.e. differential colocalisation, may confer to the clusters the aptitude to react specifically to particular combinations of cytokines. Therefore, if the right combination of cytokine is encountered by a cluster with the right proportions of antigens, signalling cascades may take place [Adam *et al*, 2003], be amplified by the cluster associated molecules and lead to actin polymerisation (see figure 7.14). Consequently, only the clusters with the right antigen formulation lead to efficient signalling cascades and fast actin polymerisation [Lévesque and Simmons, 1999]. Actin fibres therefore elongate and protrude in the cell membrane, forming quickly the short and long pseudopodia observed. Since most clusters do not have the appropriate antigen formulation, most of them would generate no membrane protrusion, small ones or slightly more elongated ones. This may explain the relative abundance of short

protrusion in comparison to longer one. This may also explain why protrusion are found coated with CD34 (and likely with other cluster molecules), as actin polymerisation would occur behind and in the vicinity of the clusters with appropriate antigen formulation.

The production of pseudopodia coated with CD34, CD164 and CD133 may be used in turn to mediate survival signals to the cell body. Indeed, along such long pseudopodia, CD34 and CD164 are more likely to encounter spatial regions coated with their specific (yet unknown) ligands. CD34 function would be to perform adhesion (to substrate or other cells). Indeed, it has been reported that CD34 shares structural similarities with another sialomucin, CD43, a protein implicated in cell-cell adhesion, tyrosine kinase activation and cytoskeleton interactions [Felschow et *al.*, 2001; Anzai *et al*, 1999; Park *et al*, 1991]. As for CD34, CD43 is also reported to display differential colocalisation patterns in cell membranes by confocal analysis [Brown *et al*, 1996]. In addition, CD34 and CD43 share a common signalling pathway via Syk and Lyn tyrosine kinases [Tada *et al*, 1999]. Furthermore, CD34 is capable of inducing adhesion (possibly to L-selectins, as for CD43) specifically related to class II epitope expression, and signal transduction. Its predominant form is the full length protein (class I, II and III altogether). CD34 class I and II epitopes (the most distal) but not class III monoclonal antibodies induce cell-cell adhesion. CD34 does not have intrinsic kinase activity, so signalling cascades can only be triggered through intracellular CD34 tail association with other molecules. Indeed, an unknown 45 kd protein was detected associating with the intracellular tail of the full length CD34 molecule immediately after CD34 engagement with antibodies [Felschow *et al*, 2001]. CD34 is found to *differentially* bind to the most abundantly expressed

adapter protein in HSPCs, the 39 kd CrkL. Such adapter can enable membrane receptors (such as integrins) to directly or indirectly interact with tyrosine or serine/threonine kinases in order to transmit signals. CrkL overexpression leads to increased adhesion to fibronectin through activation of C3G. The CrkL protein family has also been reported to be involved in a variety of signalling pathways following T-cell stimulation, haemopoietic cytokine activation and integrin cross-linking. Signals transmitted after CD34 engagement lead to actin polymerisation. Moreover, there is in vitro evidence that CrkL forms complexes with other proteins (such as C3G, c-Abl, Sos and Dock 180), strongly suggesting that other molecules are likely to be associated with the HSPC membrane antigen clusters (so called cluster-associated proteins). Intracellular kinases (e.g. protein kinase C) are also known to play a role in CD34 function through phosphorylation [Fackler *et al*, 1990] and the control of its activity [Lanza *et al*, 2001].

CD164, located in the same regions as CD34, would also be activated by its ligand and subsequently generate signalling cascades. The signalling cascades triggered by CD164 would send negative regulatory messages to the cell body up to the nuclear material which would prevent cell proliferation, impairing cell engagement into the haemopoietic process [Doyonnas *et al*, 2000; Lee *et al*, 2001]. Therefore, in this proposed mechanism, cell adhesion (*via* CD34 or other molecules [Goodell, 1999; Guo *et al*, 2003; Kuci *et al*, 2003; Verfaillie, 1998]) prevents cell division (*via* CD164), which is consistent with haemopoietic theories (i.e. adhesion prevents cell cycling). It also provides a potential explanation for the observed membrane protrusions and pseudopodia, together with their observed antigen coating (observed with electron microscopy), but it is stressed that antigen clusters may not necessarily trigger podia formation (as observed with LSCM, e.g.

fig. 7.13). In any case, adhesion using CD34 prevents cell proliferation using CD164. The selective presence of CD133 within a few CD34 clusters and none other may also follow this hypothesis. CD133 containing clusters may have the right balance of antigens to perform their biological function which is to create (short or long) membrane extension in order to reach other cells [Corbeil *et al*, 1999]. Therefore CD133 is directly brought to contact points between touching or neighbouring cells, and since this antigen is potentially a signalling receptor, its presence at cell-to-cell interface would possibly allow cell-to-cell communication.

As the haemopoietic microenvironment secretes cocktails of cytokines, a fraction of the HSPCs activated after circulating in the blood stream engages in the maturation process, CD34, CD164 and CD133 expression is down-regulated until no more expression takes place, and in turn an increasing diversity of other antigens are expressed in order to participate to the tuning of differentiation events [Aglietta *et al*, 1998; Ogawa *et al*, 1983; Koury, 1992].

In this interpretation, clusters are small, specialised and spatially segregated pieces of machinery with random and different antigen formulations. The relative proportions of antigens are responsible for the specific activation and signal amplification in particular clusters in response to appropriate cytokine cocktails, and random antigen formulation confers a wide sensitivity range to cytokine/adhesion molecule stimulation (fig. 7.14). Thus HSPCs are potentially well equipped to develop fast responses to appropriate stimuli in localised and specific microenvironments. This theory weights in favour of the non-deterministic model of haematopoiesis: indeed, random combinations of membrane receptors are expressed in HSPCs, particular and lineage specific

environmental factors stimulate the appropriate antigen clusters, leading to signalling cascade and cell engagement in the differentiation process. Non appropriately stimulated HSPCs remain quiescent or undergo apoptosis.
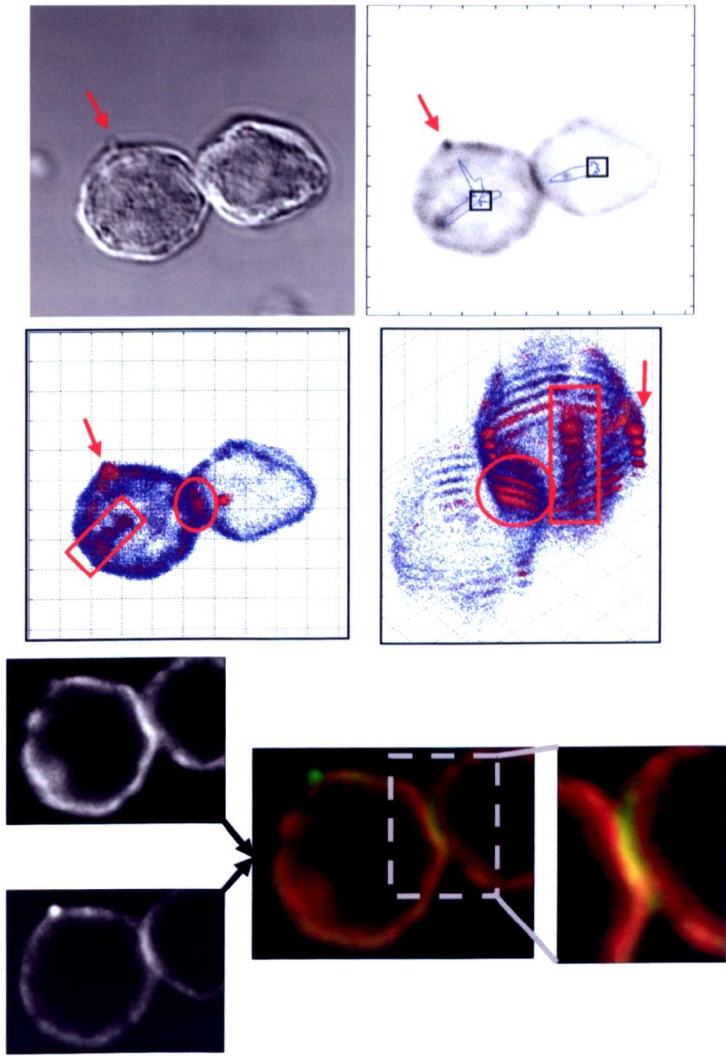
Figure 7.13 : **Representative example of the three types of clusters observed on HSPCs.**

These are CD34 distributions. The circle indicates a patch-like cluster located at the cell-to-cell contact area. The rectangle indicates a MSCL structure and the arrow points to a small cluster. This small dense cluster is clearly associated with a membrane protrusion (visible on the DIC image) and predominantly contains class III CD34 epitope (green). At cell-to-cell contact site, the left cell expresses mostly class I CD34 epitope while the right cell expresses mostly class III CD34 epitope. Both classes are clearly colocalised in the patch like area, as indicated by the yellow colour in the bottom right image. Polarity graphs (top right) show that theCD34 from the right cell is predominantly polarised towards the contact area.
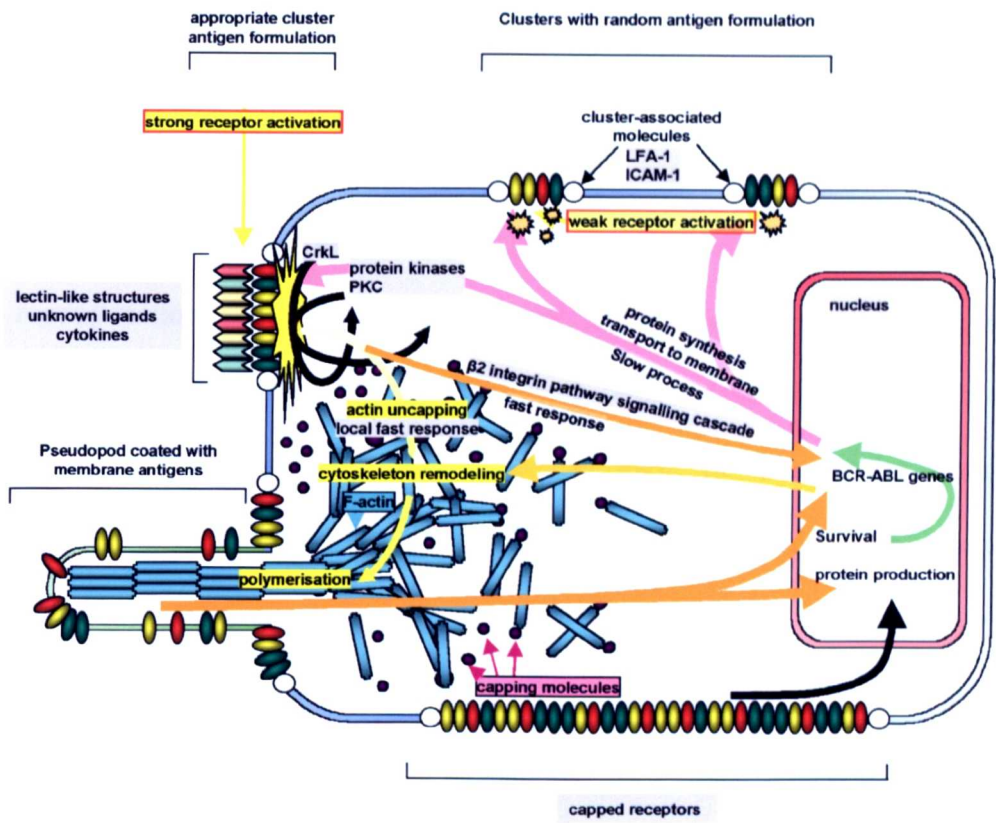
Figure 7.14: **Potential function of antigen clusters in HSPC membrane.**
Under either external (cytokines, adhesion molecules) or internal (genetic) stimuli, early membrane antigens are synthesised in the cytoplasm and transported to the membrane. As HSPC circulate the blood stream, these antigens aggregate within cluster structures, probably with participation of cytoskeleton molecules and lipid rafts. The resultant clusters have different relative proportions of various antigens. Some of these molecule interact directly or indirectly with each others, modulating the signalling pathways taking place at these sites. The relative proportions of cluster antigens is responsible for the differential tuning of signalling activity, after appropriate activation through ligation of corresponding ligands. It is hypothesised that functional cluster signalling activity relies on the amplification of initial phosphorylation cascades leading to actin polymerisation and gene activation. Such an amplification happens only if clusters with appropriate antigen formulation encounter the corresponding ligand/cytokine combinations from the surrounding environment. Actin polymerisation may generate membrane protrusions and pseudopods coated with cluster antigens.

7.4 Conclusion

A novel methodology has been developed to qualitatively and quantitatively characterise haematopoietic membrane antigens. This method could be extended to the study of other cell types. More accurate membrane analysis could be performed by adding a membrane dye to the target cells, which would allow better 3D reconstruction [Ortiz De Solorzano *et al*, 2001] and antigen mapping, and possibly allow the analysis of clusters and their relations with membrane protrusions. These methods could also be used to characterise antigen distributions in cells subjected to various experimental conditions. For example, they could be used to quantitatively assess the effect of drugs at various concentrations. They could also be used to characterise time lapse series. Importantly, several new visualisation tools and measurements were introduced:

- the Antigen Density Maps (ADMs) provide a concise way for visualising and comparing 3D antigen distributions,

- the 3D sphere models are particularly helpful for assessing and comparing fluorescence distribution of clustered cells, while actual 3D reconstruction may still be subject to difficult interpretation,

- polarity measurements allow to quantitatively characterise the various antigen distribution polarisation states encountered in cell membranes, making this information available for statistical comparison of various cell populations under different experimental conditions, and to represent this information in the practical form of graphs

- the various quantitative colocalisation methods herein developed greatly extend the possibility of characterising accurately colocalised antigen distributions, describing numerically and graphically various degrees of colocalisation, in a much finer way than the conventional use of Pearson's correlation measurements applied between images.

Altogether, these innovative tools allow to multiply the measurements performed on LSCM images, therefore increasing the amount of information available for further analysis and interpretation of the data.

The novel fluorescence quantification method adopted here has been shown to improve the measurements obtained using different segmentation algorithms. It could possibly be used to improve the comparison of measurements performed by different users while reducing the subjective biases. Biases are introduced at several points in the experiments (e.g. problems with fluorescence and microscope calibration). They are also commonly introduced by user performing manual segmentation. In this study, the development and application of automatic segmentation algorithms was investigated and was shown to be robust [Baradez et al., 2004], therefore minimising the risk of initial subjective bias.

Moreover, the MSA algorithm has many potential segmentation applications. Its high computational efficiency suggests that it could easily be implemented for real time applications for research and industry.

Reference list and Bibliography

Adams GB, Chabner KT, Foxall RB, Wibrecht KW, Rodrigues NP, Dombkowski D, Fallon R, Poznansky MC, Scadden DT. Heterologous cells cooperate to augment stem cell migration, homing and engraftment. Blood 2003; 101(1)45-51.

Aglietta M, Bertolini F, Carlo-Stella C, De Vincentiis A, Lanata L, Lemoli RM et al. Ex vivo expansion of hematopoietic cells and their clinical use. Haematologica 1998; 83(9):824-848.

Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 2000; 404(6774):193-7.

Alexander WS, Roberts AW, Nicola NA, Li R, Metcalf D. Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-Mpl. Blood 1996; 87(6):2162-2170.

Ameen NA, Martensson B, Bourguinon L, Marino C, Isenberg J, McLaughlin GE. CFTR channel insertion to the apical surface in rat duodenal villus epithelial cells is upregulated in vivo. Journal of Cell Science 1999; 112:887-894.

Ammouche A, Riss J, Breysse D, Marchand J. Image analysis for the automated study of microcracks in concrete. Cement and Concrete Composites 2001; 23:267-278.

Anand A. Tracing of interference fringes using average gray value and simultaneous row and column scan. Optics & Laser Technology 2003; 35(2):73-79.

Andrews PD, Harper IS, Swedlow JR. To 5D and beyond: quantitative fluorescence microscopy in the postgenomic era. Traffic 2002; 3:29-36.

Antoine C, Lloyd MD, Antoine J. A robust thresholding algorithm for halftone dots. Journal of Pulp and Paper Science 2001; 27(8):268-272.

Anzai N, Gotoh A, Shibayama H, Broxmeyer HE. Modulation of integrin function in hematopoietic progenitor cells by CD43 engagement: possible involvement of protein tyrosine kinase and phospholipase C-gamma. Blood 1999; 93(10):3317-26.

Baradez M-O, McGuckin CP, Forraz N, Pettengell R, Hoppe A. Robust and automated unimodal histogram thresholding and potential applications. Pattern Recognition 2004; 37(1):1131-1148.

Baradez M-O, McGuckin CP, Forraz N, Pettengell R, Hoppe A. Thresholding based on linear diffusion for feature segmentation. In Proceedings of the British Machine Vision Conference 2003; 9th to 11th September, Norwich, UK.

Barnett D, Janossy G, Lubenko A, Matutes E, Newland A, Reilly JT. Guideline for the flow cytometric enumeration of CD34+ haematopoietic stem cells. Prepared by the CD34+ haematopoietic stem cell working party. General Haematology Task Force of the British Committee for Standards in Haematology. Clinical Laboratory Haematology 1999; 21(5):301-308.

Basset O, Buquet B, Abouelkaram S, Delachartre P, Culioli J. Application of texture image analysis for the classification of bovine meat. Food Chemistry 2000; 69:437-445.

Batard P, Monier MN, Fortunel N, Ducos K, Sansilvestry-Morel P, Phan TH, Hatzfeld A, Hatfzeld JA. TGF-β1 maintains hematopoietic immaturity by a reversible negative control of cell cycle and induces CD34 antigen up-modulation. Journal of Cell Science 2000; 113:383-290.

Baumheter S, Singer MS, Henzel W, Hemmerich S, Renz M, Rosen SD et al. Binding of L-selectin to the vascular sialomucin CD34. Science 1993; 262(5132):436-438.

Behringer D, Kresin V, Henszchler R, Mertelsmann R, Lindemann A. Cytokine and chemokine production by CD34$^+$ haemopoietic progenitor cells: detection in single cells. British Journal of Haematology 1997; 97:9-14.

Bettens E, Scheunders P, Sijbers J, Van Dyck D, Moens L. Automatic segmentation and modelling of two-dimensional electrophoresis gels. In International Conference on Image Processing 1996; 2 :665-668

Bhatia M. AC133 expression in human stem cells. Leukemia. 2001; 15(11):1685-8.

Bhatia R, Williams AD, Munthe HA. Contact with fibronectin enhances preservation of normal but not chronic myelogenous leukemia primitive hematopoietic progenitors. Experimental Hematology 2002; 30:324-332.

Bjørnsson JM, Larsson N, Brun ACM, Magnusson M, Andersson E, Lundström P, Larsson J, Repetowska E, Ehinger M, Humphries RK, Karlsson S. Reduced proliferative capacity of hematopoietic stem cells deficient in *Hoxb3* and *Hoxb4*. Molecular and Cellular Biology 2003; 23(11):3872-3883.

Black MJ, Sapiro G, Marimont DH, Heeger D. Robust anisotropic diffusion. IEEE Transactions on Image Processing 1998; 7(3):421-432.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 ; 1(8476):307-10.

Bonnet N, Cutrona J, Herbin M. A 'no-threshold' histogram-based image segmentation method. Pattern Recognition 2002; 35(10):2319-2322.

Bonton P, Boucher A, Thonnat M, Tomczak R, Hidalgo PJ, Belmonte J, Galan C. Colour image in 2D and 3D microscopy for the automation of pollen rate measurement. Image Analysis and Stereology 2002; 21:25-30.

Boutet de Monvel J, le Calvez S, Ulfendhal M. Image restoration for confocal microscopy: improving the limits of deconvolution, with application to the visualisation of the mammalian hearing organ. Biophysical Journal 2001; 80:2455-2470.

Bradley TR, Metcalf D. The growth of mouse bone marrow cells in vitro. Australian Journal of Experimental Biology and Medical Sciences 1966; 44:287-300.

Brown TJ, Shuford WW, Wang WC, Nadler SG, Bailey TS, Marquardt H, Mittler RS. Characterization of a CD43/leukosialin-mediated pathway for inducing apoptosis in human T-lymphoblastoid cells. Journal of Biological Chemistry 1996; 271(44):27686-27695.

Brugger W, Bühring HJ, Grünebach F, Vogel W, Kaul S, Müller R, Brümmendorf TH, Ziegler BL, Rappold I, Brossart P, Scheding S, Kanz L. Expression of MUC-1 epitopes on normal bone marrow: implications for the

detection of micrometastatic tumor cells. Journal of Clinical Oncology 1999; 17(5):1535-1544.

Brummendorf TH, Dragowska W, Zijlmans JMJM, Thonbury G, Lansdorp PM. Asymmetric cell divisions sustain long-term hematopoiesis from single-sorted human fetal liver cells. Journal of Experimental Medicine 1998; 188(6):1117-1124.

Canny J. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1986; 8(6):679-698.

Chan JYH, Lee-Prudhoe JE, Jorgensen B, Ihrke G, Doyonnas R, Zannettino ACW, Buckle VJ, Ward CJ, Simmons PJ, Watt SM. Relationship between novel isoforms, functionally important domains, and subcellular distribution of CD164/Endolyn. The Journal of Biological Chemistry 2001; 276(3):2139-2152.

Chan JY, Watt SM. Adhesion receptors on haematopoietic progenitor cells. British Journal of Haematology 2001; 112(3):541-57.

Cheriet M, Said JN, Suen CY. A recursive thresholding technique for image segmentation. IEEE Transactions on Image Processing 1998; 7(6):918-921.

Choi WP, Lam KM, Siu WC. An adaptive active contour model for highly irregular boundaries. Pattern Recognition 2001; 34:323-331.

Christensen JL, Wright DE, Wagers AJ, Weissman IL. Circulation and chemotaxis of fetal hematopoietic stem cells. Public Library of Science (PLOS) Biology 2004; 2(3):368-377.

Civin CI, Strauss LC, Brovall C, Fackler MJ, Schwartz JF, Shaper JH. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. Journal of Immunology 1984; 133(1):157-165.

Constans A. Rendering images in 3D. The Scientist 2004; April 26:40-42.

Corbeil D, Roper K, Hannah MJ, Hellwig A, Huttner WB. Selective localization of the polytopic membrane protein prominin in microvilli of epithelial cells - a combination of apical sorting and retention in plasma membrane protrusions. Journal of Cell Science 1999; 112(Pt 7):1023-33.

Corbeil D, Roper K, Hellwig A, Tavian M, Miraglia S, Watt SM et al. The human AC133 hematopoietic stem cell antigen is also expressed in epithelial cells and targeted to plasma membrane protrusions. Journal of Biological Chemistry 2000; 275(8):5512-5520.

Corbeil D, Roper K, Weigmann A, Huttner WB. AC133 hematopoietic stem cell antigen: human homologue of mouse kidney prominin or distinct member of a novel protein family? [letter]. Blood 1998; 91(7):2625-2626.

Daly C, McGee A, Vila E, Briones A, Giraldo J, Arribas S, González C, Somoza B, Pagakis SN, Adler J, Provost JC, Merle A, Maddison J, Pederson J, McGrath JC. Analysing the 3D structure of blood vessels using confocal microscopy. Microscopy and Analysis 2002; Nov. 5-8.

Danckaert A, Gonzalez-Couto E, Bollondi L, Thompson N, Hayes B. Automated recognition of intracellular organelles in confocal microscope images. Traffic 2002; 3: 66-73.

Dao MA, Arevalo J, Nolta JA. Reversibility of CD34 expression on human hematopoietic stem cells that retain the capacity for secondary reconstitution. Blood 2003; 101(1):112-118.

Davies ER. Machine Vision: Theory, Algorithms and Practicalities. Academic Press, 1990, Chap 4 (Thresholding techniques).

de Boer F, Kessler FL, Netelenbos T, Zweegman S, Huijgens PC, van der Wall E, van der Linden JAM, Pinedo HM, Jan Schuurhuis G, Dräger AM. Homing and clonogenic outgrowth of CD34[+] peripheral blood stem cells: a role for L-selectin? Experimental Hematology 2002; 30:590-597.

Denning-Kendall P, Singha S, Bradley B, Hows J. Cytokine expansion culture of cord blood CD34[+] cells induces marked and sustained changes in adhesion receptor and CXCR4 expressions. Stem Cells 2003; 21:61-70.

de Wynter EA, Buck D, Hart C, Heywood R, Coutinho LH, Clayton A et al. CD34+AC133+ cells isolated from cord blood are highly enriched in long-term culture-initiating cells, NOD/SCID-repopulating cells and dendritic cell progenitors. Stem Cells 1998; 16(6):387-396.

Dexter TM, Allen TD, Lajtha LG. Conditions controlling the proliferation of haemopoietic stem cells in vitro. Journal of Cellular Physiology 1977; 91(3):335-344.

Digabel H, Lantuejoul C. Iterative Algorithms. In Actes du Second Symposium Europeen d'Analyse Quantitative des Microstructures en Sciences des Materiaux, Biologie et Medecine, Caen, J-L Chermant, Ed. Riederer Verlag, Stuttgarts, pp 85-99, 4-7 October 1978.

Doyonnas R, Chan JYH, Butler LH, Rappold I, Lee-Prudhoe JE, Zannettino ACW, Simmons PJ, Bühring HJ, Levesque JP, Watt SM. CD164 monoclonal antibodies that block hemopoietic progenitor cell adhesion and proliferation interact with the first mucin domain of the CD164 receptor. Journal of Immunology 2000; 165: 840-851.

Draghici S. A neural network based artificial vision system for licence plate recognition. International Journal of Neural Systems 1997; 8(1):113-126.

Dravid G, Rao SGA. Ex Vivo expansion of stem cells from umbilical cord blood: expression of cell adhesion molecules. Stem Cells 2002; 20:183-189.

Elghetany MT. Surface marker abnormalities in myelodysplastic syndromes. Haematologica 1998; 83: 1104-1115.

Enderle T, Ha T, Ogletree DF, Chemla DS, Magowan C, Weiss S. Membrane specific mapping and colocalisation of material and host skeletal proteins in the *Plasmodium falciparum* infected erythrocyte by dual-color near-field scanning optical microscopy. Proceedings of the National Academy of Sciences of the USA 1997; 94:520-525.

Fackler MJ, Civin CI, Sutherland DR, Baker MA, May WS. Activated protein kinase C directly phosphorylates the CD34 antigen on hematopoietic cells. Journal of Biological Chemistry 1990; 265(19):11056-11061.

Fackler MJ, Krause DS, Smith OM, Civin CI, May WS. Full-length but not truncated CD34 inhibits hematopoietic cell differentiation of M1 cells. Blood 1995; 85(11)3040-3047.

Fargeas CA, Florek M, Huttner WB, Corbeil D. Characterization of prominin-2, a new member of the prominin family of pentaspan membrane glycoproteins. Journal of Biological Chemistry 2003; 278(10):8586-96.

Fassett MS, Davis DM, Valter MM, Cohen GB, Strominger JL. Signalling at the inhibitory natural killer cell immune synapse regulates lipid raft polarization but not class I MHC clustering. Proceedings of the National Academy of Sciences of the USA 2001; 98(25):14547-14552.

Ferrero E, Belloni D, Foglieni P, Ferrero ME, Fabbri M, Poggi A, Zocchi MR. Transendothelial migration leads to protection from starvation-induced apoptosis in CD34$^+$CD14$^+$ circulating progenitors: evidence for PECAM-1 involvement through Akt/PKB activation. Blood 2003; 101(1):186-93.

Forraz N, Pettengell R, McGuckin CP. Characterization of a lineage-negative stem-progenitor cell population optimized for ex vivo expansion and enriched for LTC-IC. Stem Cells 2004; 22:100-108.

Francis K, Ramakrishna R, Holloway W, Palsson BO. Two New Pseudopod Morphologies Displayed by the Human Hematopoietic KG1a Progenitor Cell Line and by Primary Human CD34$^+$ Cells. Blood 1998; 92(10)3616-3623

Fruehauf S, Veldwijk MR, Krämer A, Haas R, Zeller WJ. Delineation of cell cycle state and correlation to adhesion molecule expression of human CD34$^+$ cells from steady-state bone marrow and peripheral blood mobilized following G-CSF-supported chemotherapy. Stem Cells 1998; 16:271-279.

Gaigalas AK, Li L, Henderson O, Vogt R, Barr J, Marti G, Weaver J, Schwartz A. The development of fluorescence intensity standards. Journal of

Research of the National Institute of Standards and Technology 2001; 106:381-389.

Gallacher L, Murdoch B, Wu DM, Karanu FN, Keeney M, Bhatia M. Isolation and characterization of human CD34-Lin-and CD34+Lin- hematopoietic stem cells using cell surface markers AC133 and CD7. Blood 2000; 95(9):2813-2820.

Gavrielides MA, Lo JY, Floyd Jr CE. Parameter optimisation of a computer-aided diagnosis scheme for the segmentation of microcalcification clusters in mammograms. Medical Physics 2001; 29(4):475-483.

Gazitt Y. Immunologic profiles of effector cells and peripheral blood stem cells mobilized with different hematopoietic growth factors. Stem Cells 2000; 18:390-398.

Gerlich D, Beaudouin J, Gebhard M, Ellenberg J, Eils R. Four-dimensional imaging and quantitative reconstruction to analyse complex spatiotemporal processes in live cells. Nature Cell Biology 2001; (9):852-5.

Gonzalez R, Woods R. Digital Image Processing. Addison-Wesley Publishing Company, 2002, Chap 10 (Image segmentation).

Goodell MA. CD34[+] or CD34[-]: Does it Really Matter? Blood 1999; 94(8):2545-2547.

Goodman JW, Hodgson GS. Evidence for Stem cells in the peripheral blood of mice. Blood 1962; 19:702-14.

Grabarek J, Darzynkiewicz Z. Versatility of analytical capabilities of laser scanning Cytometry (LSC). Clinical and Applied Immunology Reviews 2002; 2:75-92.

Greenberg AW, Kerr WG, Hammer DA. Relationship between selectin-mediated rolling of hematopoietic stem and progenitor cells and progression in hematopoietic development. Blood 2000; 95(2):478-486.

Guild SB, Murray AT, Wilson ML, Wiegand UK, Apps DK, Jin Y, Rindler M, Roder J, Jeromin A. Over-expression of NCS-1 in AtT-20 cells affects ACTH secretion and storage. Molecular and Cellular Endocrinology 2001; 184:51-63.

Guo L, Burke P, Lo SH, Gandour-Edwards R, Lau D. Quantitative analysis of angiogenesis using confocal laser scanning microscopy. Angiogenesis 2001; 4:187-191.

Guo Y, Lübbert M, Engelhardt M. CD34⁻ haematopoietic stem cells: current concepts and controversies. Stem Cells 2003; 21:15-20.

Gupta N, DeFranco AL. Visualising lipid raft dynamics and early signalling events during antigen receptor-mediated B-lymphocyte activation. Molecular Biology of the Cell 2003; 14:432-444.

Haris K, Efstratiadis SN, Maglaveras N, Katsaggelos AK. Hybrid image segmentation using watersheds and fast region merging. IEEE Transactions on Image Processing 1998; 7(12):1684-1699.

Harrisson DE, Astle CM, Lerner C. Number and continous proliferative pattern of transplanted primitive immunohematopoietic stem cells. Proceedings of the National Academy of Sciences of the USA 1988; 85:822-6.

Healy L, May G, Gale K, Grosveld F, Greaves M, Enver T. The stem cell antigen CD34 functions as a regulator of hemopoietic cell adhesion. Proceedings of the National Academy of Sciences of the USA 1995; 92: 12240-12244.

Heilbronn R, Engstler M, Weger S, Krahn A, Schetter C, Boshart M. ssDNA-dependent colocalization of adeno-associated virus Rep and herpes simplex virus ICP8 in nuclear replication domains. Nucleic Acids Research 2003; 32(21)6206:6213.

Hervás J, Barredo JI, Rosin PL, Pasuto A, Mantovani F, Silvano S. Monitoring landslides from optical remotely sensed imagery: the case history of Tessina landslide, Italy. Geomorphology 2003; 1346:1-13.

Horn PA, Tesch H, Staib P, Kube D, Diehl V, Voliotis D. Expression of AC133, a novel hematopoietic precursor antigen, on acute myeloid leukemia cells. Blood 1999; 93(4):1435-7.

Hu MCT, Chien SL. The Cytoplasmic Domain of Stem Cell Antigen CD34 Is Essential for Cytoadhesion Signalling But Not Sufficient for Proliferation Signalling. Blood 1998; 91(4): 1152-1162.

Huss R. CD34-negative stem cells as the earliest precursors of hematopoietic progeny. Experimental Haematology 1998; 26:1022-1023.

Huygen S, Giet O, Artisien V, Di Stefano I, Beguin Y, Gothot A. Adhesion of synchronized human hematopoietic progenitor cells to fibronectin and vascular cell adhesion molecule-1 fluctuates reversibly during cell cycle transit in *ex vivo* culture. Blood 2002; 100(8):2744-2752.

Ito T, Tajima F, Ogawa M. Developmental changes of CD34 expression by murine hematopoietic stem cells. Experimental Haematology 2000; 28:1269-1273.

Jalba AC, Wilkinson MHF, Roerdink JBT. Morphological hat-transform scale spaces and their use in pattern classification. Pattern Recognition 2004; 37(5):901-915.

Jankowski RJ, Deasy BM, Cao B, Gates C, Huard J. The role of CD34 expression and cellular fusion in the regeneration capacity of myogenic progenitor cells. Journal of Cell Science 2002; 115:4361-4374.

Jeon NL, Baskaran H, Dertinger SKW, Whitesides GM, Van de Water L, Toner M. Neutrophil chemotaxis in linear and complex gradients of interleukin-8 formed in microfabricated device. Nature Biotechnology 2002; 20:826-830.

Jiang Y, Prosper F, Verfaillie CM. Opposing effects of engagement of integrins and stimulation of cytokine receptors on cell cycle progression of normal human hematopoietic progenitors. Blood 2000; 95(3):846-854.

Kansas GS. Selectins and their ligands: current concepts and controversies. Blood 1996; 88(9):3259-87.

Knaan-Shanzer S. No hierarchy within the primitive hemopoietic stem cell compartment. Abstract presented at: The New Stem Cells; September 18-19, 2000; Leiden, The Netherlands.

Koenderink JJ. The structure of images. Biological Cybernetics 1984; 50:363-370.

Koury MJ. Programmed cell death (apoptosis) in hematopoiesis. Experimental Haematology 1992; 20:391-394.

Krämer A, Hörner S, Willer A, Fruehauf S, Hochhaus A, Hallek M, Hehlmann R. Adhesion to fibronectin stimulates proliferation of wild-type and *bcr / abl* - transfected murine hematopoietic cells. Proceedings of the National Academy of Sciences of the USA 1999; 96:1087-1092.

Krause DS, Fackler MJ, Civin CI, May WS. CD34: structure, biology, and clinical utility. Blood 1996; 87(1):1-13.

Krause DS, Theise ND, Collector MI, Henegariu O, Hwang S, Gardner R, Neutzel S, Sharkis SJ. Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. Cell 2001; 105(3):369-77.

Kreft M, Milisav I, Ptokar M, Zorec R. Automated high through-put colocalization analysis of multichannel confocal images. Computer Methods and Programs in Biomedecine 2004; 74:63-67.

Kröger N, Zeller W, Hassan HT, Dierlamm J, Zander AR. Difference between expression of adhesion molecules on CD34$^+$ cells from bone marrow and G-CSF-stimulated peripheral blood. Stem Cells 1998; 16:49-53

Kronenwett R, Martin S, Haas R. The role of cytokines and adhesion molecules for mobilization of peripheral blood stem cells. Stem Cells 2000; 18:320-330.

Kuci S, Wessels JT, Buehring HJ, Schilbach K, Schumm M, Seitz G, Loeffler J, Bader P, Schlegel PG, Niethammer D, Handgretinger R. Identification of a novel class of human adherent CD34⁻ stem cells that give rise to SCID-repopulating cells. Blood 2003; 101(3):869-876.

Kuijper A, Florack LMJ. Hierachical pre-segmentation without prior knowledge. Proceedings of the 8th International Conference on Computer Vision 2001, Vancouver, Canada, July 9-12, 2001, Vol II, 487-493

Kumaravelu P, Hook L, Morrison AM, Ure J, Zhao S, Zuyev S, Ansell J, Medvinsky A. Quantitative developmental anatomy of the definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. Development 2002; 129:4891-4899.

Lanza F, Healy L, Sutherland DR. Structural and functional features of the CD34 antigen: an update. Journal of Biological Regulation of Homeostatic Agents [Review] 2001; 15:1-13.

Lanza F, Moretti S, Castagnari B, Montanelli F, Latorraca A, Ferrari L, Bardi A, Dominici M, Campioni D, Dabusti M, Piva N, Lodi G, Reverberi R, Castoldi G. Assessment of distribution of CD34 epitope classes in fresh and cryopreserved peripheral blood progenitor cells and acute myeloid leukemic blasts. Haematologica 1999; 84(11):969-977.

Lee YN, Kang JS, Krauss RS. Identification of a role for the sialomucin CD164 in Myogenic differentiation by signal sequence trapping in yeast. Molecular and Cellular Biology 2001; 2(22):7696-7706.

Leitinger B, Hogg N. The involvement of lipid rafts in the regulation of integrin function. Journal of Cell Science 2002; 115:963-972.

Leppanen A, Mehta P, Ouyang YB, Ju T, Helin J, Moore KL, van Die I, Canfield WM, McEver RP, Cummings RD. A novel glycosulfopeptide binds to P-selectin and inhibits leukocyte adhesion to P-selectin. Journal of Biological Chemistry 1999; 274(35):24838-48.

Lévesque JP, Simmons PJ. Cytoskeleton and integrin-mediated adhesion signalling in human CD34+ hemopoietic progenitor cells. Experimental Hematology 1999; 27:579-586.

Li CH, Tam PKS. An iterative algorithm for minimum cross entropy thresholding. Pattern Recognition Letters 1998; 19:771-776.

Liao PS, Chen TS, Chung PC. A fast algorithm for multilevel thresholding. Journal of Information Science and Engineering 2001; 17:713-727.

Lin G, Adiga U, Olson K, Guzowski JF, Barnes CA, Roysam B. A hybrid 3D watershed algorithm incorporating gradient cues and object models for

automatic segmentation of nuclei in confocal image stacks. Cytometry 2003; 56A:23-36.

Liu S, Weaver DL, Taatjes DJ. Three-dimensional reconstruction by confocal laser scanning microscopy in routine pathologic specimens of benign and malignant lesions of the human breast. Histochemistry and Cell Biology 1997; 107(4):267-78.

Lore K, Spetz AL, Fehniger TE, Sonnerborg A, Landay AL, Andersson J. Quantitative single cell methods that identify cytokine and chemokine expression in dendritic cells. Journal of Immunological Methods 2001; 249:207-22.

Macey MG, McCarthy DA, van Agthoven A, Newland AC. How should CD34+ cells be analysed? A study of three classes of antibody and five leucocyte preparation procedures. Journal of Immunological Methods 1997; 204:175-188.

Madruga J, Koritschoner NP, Diebold SS, Kurz SM, Zenke M. Polarised expression pattern of focal contact proteins in highly motile antigen presenting dentritic cells. Journal of Cell Science 1999; 112:1685-1696.

Majdic O, Stockl J, Pickl WF, Bohuslav J, Strobl H, Scheinecker C et al. Signaling and induction of enhanced cytoadhesiveness via the hematopoietic progenitor cell surface molecule CD34. Blood 1994; 83(5):1226-1234.

Malide D, Yewdell JW, Bennink JR, Cushman SW. The export of major histocompatibility complex class I molecules from the endoplasmic reticulum of rat brown adipose cells is acutely stimulated by insulin. Molecular Biology of the Cell 2001; 12:101-114.

Malpica N, Santos A, Tejedor A, Torres A, Castilla M, García-Barreno P, Desco M. Automatic quantification of viability in epithelial cell cultures by texture analysis. Journal of Microscopy 2003; 209(Pt 1):34-40.

Matkowskyj K, Glover S, Benya R. Quantitative immunohistochemistry: a new algorithm measuring cumulative signal strength and receptor number. Microscopy and Analysis 2004; 18(3):17-18.

Matsumoto K, Yasui K, Yamashita N, Horie Y, Yamada T, Tani Y, Shibata H, Nakano T. In vitro proliferation potential of AC133 positive cells in peripheral blood. Stem Cells 2000; 18:196-203.

Matsuoka S, Ebihara Y, Xu MJ, Ishii T, Sagiyama D, Yoshino H, Ueda T, Manabe A, Tanaka R, Ikeda Y, Nakahata T, Tsuji K. CD34 expression on long-term repopulating hematopoietic stem cells changes during developmental stages. Blood 2001; 97(2)419-425.

May M. Advances in cellular image processing. The Scientist 2004, March 15, pp. 40-43

Mayani H, Guilbert LJ, Janowska-Wieczorek A. Biology of the hemopoietic microenvironment. European Journal of Haematology 1992; 49(5):225-33.

Maynadié M, Gerland L, Aho S, Girodon F, Bernier M, Brunet C, Campos L, Daliphard S, Deneys V, Falkenrodt A, Jacob MC, Kühlein E, LeCalvez G, Moskovtchenko P, Philip P, Carli PM, Faure GC, Béné MC. Clinical value of the quantitative expression of the three epitopes of CD34 in 300 cases of acute myeloid leukaemia. Haematologica 2002; 87(8):795-803.

McCloskey KE, Zborowski M, Chalmers JJ. Measurement of CD2 expression levels of IFN-α-treated fibrosarcomas using cell tracking velocimetry. Cytometry 2001; 44:137-147.

McGuckin CP, Forraz N, Baradez MO, Lojo-Rial C, Wertheim D, Whiting K, Watt SM, Pettengell R. Colocalization analysis of sialomuicins CD34 and CD164. Stem Cells 2003;21: 162-170.

McGuckin CP, Pearce D, Forraz N, Watt SM, Pettengell R. Multiparametric analysis of immature cell populations in umbilical cord blood and bone marrow. European Journal of Haematology 2003b; 71(5):341-50.

McGuckin CP, Pettengell R, Martin F, Jones K, Gordon-Smith EC, Pearce D. Is AC133 involved in direct cell to cell communication? Experimental Hematology 1999; 28[7, S1]160.

McNiece I, Briddell R. Ex-vivo expansion of haemopoietic progenitor cells and mature cells. Experimental Hematology 2001; 29:3-11.

Metcalf D. The molecular control of proliferation and differentiation in hemopoietic cells. Comptes-Rendus de l'Académie des Sciences III 1993; 316(9):860-870.

Michova A, Guenova M, Nikolova M, Taskov H. CD13 epitope 7H5 exhibits a particular expression pattern in haematological malignancies. Haematology 2003; 6(3): 344-351.

Miraglia S, Godfrey W, Yin AH, Atkins K, Warnke R, Holden JT et al. A novel five-transmembrane hematopoietic stem cell antigen: isolation, characterization, and molecular cloning. Blood 1997; 90(12):5013-5021.

Mitchinson NA. The colonisation of irradiated tissue by transplanted spleen cells. British Journal of Experimental Pathology 1956; 37:239-47.

Muller-Sieburg C, Whitlock CA, Weissman IL. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1lo hematopoietic stem cell. Cell 1986; 44:653-62.

Murphy RF, Velliste M, Yao J, Porreca G. Searching online journals for fluorescence microscope images depicting protein subcellular location patterns. Proceedings of the 2nd IEEE International Symposium in Bio-Informatics and Biomedical Engineering (BIBE 2001) pp. 119-128

Murukeshan VM. Trends in digital speckle pattern interferometry. Optics and Lasers in Engineering 2003; 39:409-410.

Murukeshan VM, Fei LY, Krishnakumar V, Ong LS, Asundi A. Development of Matlab filtering techniques in digital speckle pattern interferometry. Optics and Lasers in Engineering 2003b; 39(4):441-448.

Nakamura Y, Ando K, Muguruma Y, Sumita M, Terashima S, Yoshida M, Yamashita Y, Hotta T, Kato S. Enrichment of lineage-CD34- cells using a newly developed filter system. British Journal of Haematology 2000;108(4):801-4.

Nilsson SK, Quesenberry PJ. Stem cell systems: basic principles and methodologies. In: Quesenberry PJ, Stein GS, Forget BG, Weissman SM, editors. Stem cell biology and gene therapy. New York: Wiley-Liss., 1998: 566.

Nishio N, Hisha H, Ogata H, Inaba M, Yamamoto Y, Amoh Y, Yasumizu R, Hanada KI, Hamada H, Ikehara S. Changes in markers, receptors and adhesion

molecules expressed on murine hemopoietic stem cells after single injection of 5-fluorouracil. Stem Cells 1996; 14:584-591

Nowell PC, Cole LJ, Habermeyer JG, Roan PL. Growth and continued function of rat marrow cells in x-radiated mice. Cancer Research 1956 16(3); 258-61.

Nyongesa HO, Otieno AW, Rosin PL. Neural fuzzy analysis of delaminated composites from shearography imaging. Composite Structures 2001; 54:313-318.

Ogawa M, Porter PN, Nakahata T. Renewal and commitment to differentiation of hemopoietic stem cells (an interpretive review). Blood 1983; 61(5):823-829.

Oh DJ, Lee GM, Francis K, Palsson BO. Phototoxicity of the fluorescent membrane Dyes PKH2 and PKH26 an the human hematopoietic KG1a progenitor cell line. Cytometry 1999; 36:312-318.

Ortiz de Solórzano C, Costes S, Callahan DE, Parvin B, Barcellos-Hoff MH. Application of quantitative digital image analysis to breast cancer research. Microscopy Research and Techniques 2002; 59:119-127.

Ortiz de Solórzano C, Malladi R, Lelièvre A, Lockett S. Segmentation of nuclei and cells using membrane related protein markers. Journal of Microscopy 2001; 201(3)404-415.

Park JK, Rosenstein YJ, Remold-O'Donnell E, Bierer BE, Rosen FS, Burakoff SJ. Enhancement of T-cell activation by the CD43 molecule whose expression is defective in Wiskott-Aldrich syndrome. Nature 1991; 350(6320):706-709.

Parthasarathy M. Extracting structural information about living cells from DIC images. Microscopy and Analysis, September 2001, pp. 13-15

Pasino M, Lanza T, Marotta F, Scarso L, De Biasio P, Amato S, Corcione A, Pistoia V, Mori PG. Flow cytometric and functional characterization of AC133+ cells from human umbilical cord blood. British Journal of Haematology 2000;108(4):793-800.

Peichev M, Naiyer AJ, Pereira D, Zhu Z, Lane WJ, Williams M, Oz MC, Hicklin DJ, Witte L, Moore MAS, Rafii S. Expression of VEGFR-2 and AC133 by circulating human CD34+ cells identifies a population of functional endothelial precursors. Blood 2000; 95(3):952-958.

Peled A, Hardan I, Trackhtenbrot L, Gur E, Magid M, Darash-Yahana M, Cohen N, Grabovsky V, Franitza S, Kollet O, Lider O, Alon R, Rechavi G, Lapidot T. Immature leukemic CD34+CXCR4+ cells from CML patients have

lower integrin-dependent migration and adhesion in response to the chemokine SDF-1. Stem Cells 2002; 20: 259-266.

Petrou M, Bosdogianni P. Image processing, the fundamentals. 1999, John Wiley & Sons Ltd, p. 282

Phee H, Rodgers W, Coggeshall KM. Visualization of negative signalling in B cells by quantitative confocal microscopy. Molecular and Cellular Biology 2001; 21(24):8615-8625.

Phillips R, Ernst R, Brunk B, Ivanova N, Mahan M, Deanehan J et al. The genetic program of hematopoietic stem cells. Science 2000; 288:1635-40.

Piacibello W, Sanavio F, Garetto L, Severino A, Bergandi D, Ferrario J, Fagioli F, Berger M, Aglietta M. Extensive amplification and self-renewal of human primitive hematopoietic stem cells from cord blood. Blood 1997; 89(8):2644-53.

Pierelli L, Scambia G, Bonanno G, Rutella S, Puggioni P, Battaglia A, Mozetti S, Marone M, Menichella G, Rumi C, Mancuso S, Leone G. CD34+/CD105+ cell are enriched in primitive circulating progenitors residing in the G0 phase of the cell cycle and contain all bone marrow and cord blood CD34+/CD38low/- precursors. British Journal of Haematology 2000; 108:610-620.

Plett PA, Frankovitz SM, Wolber FM, Abonour R, Orshell-Traycoff CM. Treatment of circulating CD34+ cells with SDF-1α or anti-CXCR4 antibody enhances migration and NOD/SCID repopulating potential. Experimental Hematology 2002; 30:1061-1069.

Pluznik DH, Sachs L. The cloning of normal "mast" cells in tissue culture. Journal of Cellular Physiology 1965; 66:319.

Precioso F, Barlaud M, Blu T, Unser M. Smoothing B-Spline Active Contour for Fast and Robust Image and Video Segmentation. Proceedings of the 2003 IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain, September 14-17, 2003, pp. I.137-I.140.

Puch S, Armeanu S, Kibler C, Johnson KR, Muller CA, Wheelock MJ, Klein G. N-cadherin is developmentally regulated and functionally involved in early hematopoietic cell differentiation. Journal of Cell Science 2001; 114(8):1567-77.

Punzel M, Liu D, Zhang T, Eckstein V, Miesala K, Ho AD. The symmetry of the initial divisions of the human hematopoietic progenitors is altered only by the cellular microenvironment. Experimental Hematology 2003; 31:339-347.

Quesenberry PJ, Colvin GA, Lambert JF. The chiaroscuro stem cell: a unified stem cell theory. Blood 2002; 100(13):4266-4271.

Quirici N, Soligo D, Caneva L, Servida F, Bossolasco P, Deliliers GL. Differentiation and expansion of endothelial cells from human bone marrow CD133(+) cells. British Journal of Haematology 2001;115(1):186-94.

Rafii S, Lyden D, Benezra R, Hattori K, Heissig B. Vascular and haematopoietic stem cells: novel targets for anti-angiogenesis therapy? Nature 2002; 2:826-835.

Ramarao N, Meyer TF. Helicobacter pylori resists phagocytosis by macrophages: quantitative assessment by confocal microscopy and fluorescence-activated cell sorting. Infection and Immunity 2001; 69(4):2604-2611.

Rezaee MR, van der Zwet PMJ, Lelieveldt BPF, van der Geest RJ, Reiber JHC. A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. IEEE Transactions on Image Processing 2000; 9(7):1238-1248.

Roberts LG. Machine Perception of Three-Dimensional Solids. Optical and Electro-Optical Information Processing 1965; MIT Press, Cambridge, Massachusetts, pp. 159-197.

Rooms F, Philips W. Multiresolution techniques in microscopical image processing. G.I.T. Imaging & Microscopy April 2002 pp 60-61

Rooms F, Philips W, Van Oostveldt P. Estimation of anisotropic blur for the restoration of confocal images. Wavelet Applications in Industrial Processing. Edited by Truchetet Frédéric. Proceedings of the SPIE 2004; 5266:9-17

Rosin PL. Unimodal thresholding. Pattern Recognition 2001; 34:2083-2096.

Rosin PL, Nyongesa HO, Otieno AW. Classification of delaminated composites using neuro-fuzzy image analysis. 6th UK Workshop on Fuzzy Systems 267-273, Uxbridge, 1999

Sackstein R, Dimitroff CJ. A hematopoietic cell L-selectin ligand that is distinct from PSGL-1 and displays N-glycan-dependent binding activity. Blood 2000; 96(8)2765-2774.

Salembier P, Serra J. Flat zones filtering, connected operators and filters by reconstruction. IEEE Transactions on Image Processing 1995; 8(4):1153-1160.

Santos A, Ramiro C, Desco M, Malpica N, Tejedor A, Torres A, Ledesma-Carbayo MJ, Castilla M, García-Barreno P. Automatic detection of cellular necrosis in epithelial cell cultures. Medical Imaging 2001: Image Processing Proceedings of SPIE Vol. 4322 (2001) pp. 1836-1844

Sarti A, Ortiz de Solórzano C, Lockett S, Malladi R. A geometric model for 3-D confocal image analysis. IEEE Transactions on Biomedical Engineering 2000; 47(12):1600:1609.

Sato T, Laver JH, Ogawa M. Reversible expression of CD34 by murine hematopoietic stem cells. Blood 1999; 94(8):2548-2554.

Silver MA, Stryker MP. A method for measuring colocalization of presynaptic markers with anatomically labelled axons using double label immunofluorescence and confocal microscopy. Journal of Neuroscience Methods 2000; 94;205-215.

Simmons PJ, Haylock DN, Levesque J-P. Influence of cytokines and adhesion molecules on hematopoietic stem cell development. In: Schindhelm K, Nordon R, editors. Ex vivo cell therapy. San Diego: Academic Press, 1999.

Simmons PJ, Levesque JP, Zannettino AC. Adhesion molecules in haemopoiesis. Baillieres Clinical Haematology 1997; 10(3):485-505.

Simmons DL, Satterthwaite AB, Tenen DG, Seed B. Molecular cloning of a cDNA encoding CD34, a sialomucin of human hematopoietic stem cells. Journal of Immunology 1992; 148(1):267-271.

Snidaro L, Foresti GL. Real-time thresholding with Euler numbers. Pattern Recognition Letters 2003; 24:1533-1544.

Spear RN, Cullen D, Andrews JH. Fluorescent labels, confocal microscopy, and quantitative image analysis in the study of fungal biology. Methods in Enzymology 1999; 307:607-623.

Steen R, TjØnnfjord GE, GrØnseth LAG, Egeland T. Characteristics of haemopoietic progenitor cells related to CD34 epitope class expression. European Journal of Haematology 2000; 64: 245-251.

Steen R, TjØnnfjord G, Gaudernack G, Brinch L, Egeland T. Differences in the distribution of CD34 epitopes on normal haemopoietic progenitor cells and leukaemic blast cells. British Journal of Haematology 1996; 94:597-605.

Stelzer EHK. Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy. Journal of Microscopy, Volume 189 Issue 1 Page 15-24 - January 1998

Stephens DJ, Allan VJ. Light Microscopy Techniques for Live Cell Imaging. Science 2003 300: 82-86.

Suehiro Y, Muta K, Umemura T, Abe Y, Nishimura J, Nawata H. Macrophage inflammatory protein 1α enhances in a different manner adhesion of

hematopoietic progenitor cells from bone marrow, cord blood, and mobilized peripheral blood. Experimental Hematology 1999; 27:1637-1645.

Sutherland HJ, Eaves CJ, Eaves AC, Dragowska W, Lansdorp PM. Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. Blood 1989; 74(5):1563-1570.

Sutherland DR, Stewart AK, Keating A. CD34 antigen: molecular features and potential clinical applications. Stem Cells 1993; Oct;11 Suppl 3:50-7. Review.

Swedlow JR, Goldberg I, Brauner E, Sorger PK. Informatics and quantitative analysis in biological imaging. Science 2003; 300:100-102.

Swedlow JR, Hu K, Andreaws PD, Roos DS, Murray JM. Measuring tubulin content on Toxoplasma gondii: a comparison of laser-scanning confocal and wide-field fluorescence microscopy. Proceedings of the National Academy of Sciences of the USA 2002; 99(4):2014-2019.

Tabbone S, Wendling L. Multi-scale binarization of images. Pattern Recognition Letters 2003; 24(1-3):403-411.

Tada JI, Omine M, Yamaguchi N. A common signalling pathway via Syk and Lyn tyrosine kinases generated from capping of the sialomucins CD34 and CD43 in immature hematopoietic cells. Blood 1999; 93(11):3723-3735.

Taichman RS, Emerson SG. The role of osteoblasts in the hematopoietic microenvironment. Stem Cells 1998; 16:7-15.

Tenen DG, Hromas R, Licht JD, Zhang D-E. Transcription factors, normal myeloid development and leukaemia. Blood 1997; 90:489-519.

Thomas ED. Bone marrow transplantation : a review. Seminar Hematology 1999 Oct;36(4 Suppl 7):95-103.

Till JE, McCulloch EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiation Research 1961; 14:213-22.

Tsai DM. A fast thresholding selection procedure for multimodal and unimodal histograms. Pattern Recognition Letters 1995; 16:653-666.

Tsai RYL, McKay RDG. Cell contact regulates fate choice by cortical stem cells. Journal of Neuroscience 2000 ; 20(10) :3725-3735.

Uchida N, Buck DW, He D, Reitsma MJ, Masek M, Phan TV et al. Direct isolation of human central nervous system stem cells. Proceedings of the National Academy of Sciences of the USA 2000; 97(26):14720-5.

Ueda T, Tsuji K, Yoshino H, Ebihara Y, Yagasaki H, Hisakawa H, Mitsui T, Manabe A, Tanaka R, Kobayashi K, Ito M, Yasukawa K, Nakahata T.

Expansion of human NOD/SCID-repopulating cells by stem cell factor, Flk2/Flt3 ligand, thrombopoietin, IL-6, and soluble IL-6 receptor. Journal of Clinical Investigation 2000 ; 105(7):1013-21.

Umesh Adiga PS, Chaudhuri BB. Some efficient methods to correct confocal images for easy interpretation. Micron. 2001; 32(4):363-70.

van Buul JD, Voermans C, van der Berg V, Anthony EC, Mul FPJ, van Wetering S, van der Schoot CE, Hordijk PL. Migration of human hematopoietic progenitor cells across bone marrow endothelium is regulated by vascular endothelial cadherin. Journal of Immunology 2002; 168:588-596.

van Buul JD, Voermans C, van Gelderen J, Anthony EC, van der Shoot CE, Hordijk PL. Leukocyte-endothelium interaction promotes SDF-1-dependent polarisation of CXCR4. Journal of Biological Chemistry 2003; 278(32):30302-30310.

Venegas-Martinez S, Rendon JM, Stamon G. Anisotropic diffusion by a recursive linear convolving method: application to space-time segmentation and pattern recognition. 15th International Conference on Vision Interface, May 27-29, 2002, Calgary, Canada

Verfaillie CM. Adhesion receptors as regulators of the hematopoietic process. Blood 1998; 92(8):2609-2612.

Voermans C, Rood PML, Hordijk PL, Gerritsen WR, van der Schoot CE. Adhesion molecules involved in transendothelial migration of human hematopoietic progenitor cells. Stem Cells 2000, 18: 435-443.

Vogel W, Scheding S, Kanz L, Brugger W. Clinical application of CD34+ peripheral blood progenitor cells (PBPC). Stem Cells 2000; 18:87-92.

Wählby C, Linblad J, Vondrus M, Bengtsson E, Björkesten L. Algorithms for cytoplasm segmentation of fluorescence labelled cells. Analytical Cellular Pathology 2002; 24:101-111.

Watt SM, Chan JY. CD164--a novel sialomucin on CD34+ cells. Leukemia & Lymphoma 2000; 37(1-2):1-25.

Watt SM, Bühring HJ, Rappold I, Chan JYH, Lee-Prudhoe J, Jones T, Zannettino ACW, Simmons PJ, Doyonnas R, Sheer D, Butler LH. CD164, a novel sialomucin on CD34+ and erythroid subsets, is located on human chromosome 6q21. Blood 1998; 92(3):849-866.

Watt SM, Butler LH, Tavian M, Bühring HJ, Rappold I, Simmons PJ, Zannettino ACW, Buck D, Fuchs A, Doyonnas R, Chan JYH, Levesque JP, Peault B, Roxanis I. Functionally defined CD164 epitopes are expressed on CD34+ cells throughout ontogeny but display distinct distribution patterns in adult hematopoietic and nonhematopoitic tissues. Blood 2000; 95(10):3113-3124.

Weissman IL, Anderson DJ, Gage F. Stem and progenitor cells: origins, phenotypes, lineage commitments and transdifferentiations. Annual Review of Cellular and Developmental Biology 2001; 17:387-403.

Wiltschi K, Pinz A, Lindeberg T. An automatic assessment scheme for steel quality inspection. Machine Vision and Application 2000; 12:113-128.

Wirth MA, Stapinski A. Segmentation of the breast region in mammograms using active contours. Proceedings of SPIE - Volume 5150, Visual Communications and Image Processing, June 2003, pp. 1995-2006.

Witkin AP. Scale-space filtering. Proceedings of the Eighth International Joint Conference On Artificial Intelligence (IJCAI), Karlsruhe, Aug. 8-12 1983; 2 :1019-1022.

Wu C, Butz S, Ying Y, Anderson RG. Tyrosine kinase receptors concentrated in caveolae-like domains from neuronal plasma membrane. Journal of Biological Chemistry 1997; 272 :3554-3559.

Wu MH, Smith SL, Dolan ME. High efficiency electroporation of human umbilical cord blood CD34+ hematopoietic precursor cells. Stem Cells 2001; 19:492-499.

Xavier JB, Schnell A, Wuertz S, Palmer R, White DC, Almeida JS. Objective threshold selection procedure (OTS) for segmentation of scanning laser confocal microscope images. Journal of Microbiological Methods 2001; 47:169-180.

Yan H. Unified formulation of a class of image thresholding techniques. Pattern Recognition 1996; 29(12):2025-2032.

Yanagisawa M, Nakamura K, Taga T. Roles of lipid rafts in integrin-dependent adhesion and gp130 signalling pathway in mouse embryonic neuronal precursor cells. Genes to Cells 2004; 9:801-809.

Yang FC, Atkinson SJ, Gu Y, Borneo JB, Roberts AW, Zheng Y, Pennington J, Williams DA. Rac and Cdc42 GTPases control hematopoietic stem cell shape, adhesion, migration, and mobilization. Proceedings of the National Academy of Sciences of the USA 2001;98(10):5614-8

Yang F, Jiang T. Cell Image Segmentation with Kernel-Based Dynamic Clustering and an Ellipsoidal Cell Shape Model. Journal of Biomedical Informatics 2001; 34:67-73.

Yin AH, Miraglia S, Zanjani ED, Almeida-Porada G, Ogawa M, Leary AG, Olweus J, Kearney J, Buck DW. AC133, a novel marker for human hematopoietic stem and progenitor cells. Blood 1997; 90(12):5002-12.

Zhao X, Lasell TKR, Melançon P. Localization of large ADP-ribosylation factor-guanine nucleotide exchange factors to different Golgi compartments: evidence for distinct functions in protein traffic. Molecular Biology of the Cell 2002; 13:119-133.

Zannettino ACW, Bürhing HJ, Niutta S, Watt SM, Benton MA, Simmons PJ. The sialomucin CD164 (MGC-24v) is an adhesive glycoprotein expressed by human hematopoietic progenitors and bone marrow stromal cells that serves as a potent negative regulator of hematopoiesis. Blood 1998; 92(8):2613-2628.

Zannettino ACW, Roubelakis M, Welldon KJ, Jackson DE, Simmons PJ, Bendall LJ, Henniker A, Harrisson KL, Niutta S, Bradstock KF. Novel mesenchymal and haematopoietic cell isoforms of the SHP-2 docking receptor, PZR: identification, molecular cloning and effects on cell migration. Biochemistry Journal 2003; 370:537-549.

Zhang XF, Wang JF, Matczak E, Proper J, Groopman JE. Janus kinase 2 is involved in stromal cell-derived factor-1α-induced tyrosine phosphorylation of focal adhesion proteins and migration of hematopoietic progenitor cells. Blood 2001; 97(11):3342:8.

Appendix A

Publications

This appendix contains reprints of the published paper produced from this thesis.

- **Baradez M-O**, McGuckin CP, Forraz N, Pettengell R, Hoppe A. <u>Robust and automated unimodal histogram thresholding and potential applications</u>. Patterm Recognition 2004; 37(6):1131-1148.

- **Baradez M-O**, McGuckin CP, Forraz N, Pettengell R, Hoppe A. <u>Thresholding based on linear diffusion for feature segmentation</u>. British Machine Vision Conference, 8th September 2003, Norwich, UK

- McGuckin CP, Forraz N, **Baradez M-O**, Lojo-Rial C, Wertheim D, Whiting K, Watt SM, Pettengell R. <u>Colocalisation analysis of sialomucins CD34 and CD164</u>. **Stem Cells 2003; 21:162-170.**

Appendix B

Buffers, solutions, cell count and viability

This appendix contains descriptions of buffers and solutions used for HSPC preparations, and the description of cell count and viability calculation.

Buffers, solutions, cell count and cell viability

Please note: Town and country of a given supplier are quoted in brackets only upon first appearance in the text but not subsequently.

Buffers and solutions:

Phosphate buffer saline (PBS): 5 PBS tablets (Sigma-Aldrich, Poole, UK) were dissolved in 1 litre of double distilled water following the manufacturer's instructions in order to reach the following concentrations: 137 mmol Sodium Chloride, 2.7 mmol Potassium Chloride, 10 mmol phosphate, pH 7.2-7.4

ACD-A buffer: PBS (Sigma-Aldrich) supplemented with 0.6% ACD-A (Baxter, Maurepas, France) and bovine serum albumin (0.5% fraction V, Sigma-Aldrich, UK) equilibrated at pH 7.4. with 1N NaOH.

Staining buffer: (for fluorescent immunophenotyping) PBS (Sigma-Aldrich) supplemented with 0.1% sodium azide (Sigma-Aldrich) and 2 % foetal calf serum (PAA laboratories, Yeoville, UK) at pH 7.4.

Cell count and viability:

Cells were enumerated and assessed for viability by diluting cell samples with a trypan blue solution (0.25% in PBS). Cells were then counted in an Improved Neubauer standard haemocytometer chamber (Fischer Scientific, Loughborough, UK).
The trypan blue solution permitted discrimination between dead cells (staining blue due to impaired membrane integrity) and viable cells (remaining translucent).For enumeration at least three squares content were scored.

Cell number = (cell count/number of squares)x 10^4 x cell volume x dilution factor

% Viability = 100 x (number of viable cells / total number of cells)

Appendix C

Algorithm for estimation of cell radius using fluorescence images

This appendix contains the Matlab code used to estimate cell radii from fluorescence LSCM images used in section 2.2.3.

Algorithm for estimation of cell radius using fluorescence images

% Estimation of cell diameter using LSCM fluroescence images

% Load and display LSCM image:
[fname1,pname1]=uigetfile(['c:*.tif']),Fn=[pname1 fname1];[H, map]=imread(Fn);H=double(H);
figure,colormap(gray(256)),image(H),axis image

% Manual positioning of cell centroid (mx,my):
r=1;n=0;
hold on
while r==1
    n=n+1;
    [x(n),y(n),r]=ginput(1);
    plot(x(n),y(n),'r*'),hold on,plot(x(n),y(n),'ro')
end
hold off
mx=y(n);
my=x(n);

% Compute fluorescence mask D using the MSA algorithm:
[D,d]=f_MSA(H);

% H is thresholded at mean+4SD to produce D2:
D2=im2bw(H/255,(mean(H(:))+4*std(H(:)))/(255));

% Background binary pixels are removed using the mask D:
D3=D2.*D/255;

% Find (x,y) coordinates of signal pixels:
[x,y]=find(D3>0);

% Convert cartesian to polar coordinates, after translation of cell
% centroid to (0,0):
% t is the angle; r is the radius;
[t,r]=cart2pol(x-mx,y-my);

% Calculate a first estimate of the mean radius:
mr=mean(r);
rr=mr*ones(size(t));
[df,dff]=pol2cart(t,rr);

```
% Remove pixels out of the range 0.5*(mean radius) to 1.5*(mean
radius)
% This step prevents distant fluorescent debris to influence the cell
radius
% estimation:
BB=roipoly(H,1.5*df+my,1.5*dff+mx);
BB2=roipoly(H,0.5*df+my,0.5*dff+mx);
BX=xor(BB,BB2);
D4=D3.*BX;

% Final cell radius estimation as median radius "mr" of remaining
pixels,
% graphical outpout:
[x,y]=find(D4>0);
figure,plot(x-mean(x),y-mean(y),'.k','markersize',0.2)
h=gcf;
[t,r]=cart2pol(x-mx,y-my);
mr=median(r);
rr=mr*ones(size(t));
[df,dff]=pol2cart(t,rr);

% Superimpose circle to fluorescent pixels and display radius value:
figure(h),hold on,plot(df,dff,'.r'),axis equal
hold on,text(-40,0,num2str(mr))
```
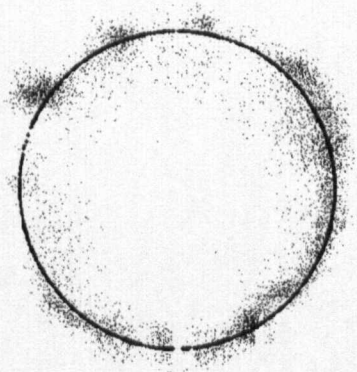

Estimated diameter = 193.6759 pixels
11 pixels = 1 μm

Exemple of automated cell radius estimation using fluorescence

Appendix D

Example of Polarity Z and distribution energy comparison. Comparison of class I, II and III CD34 epitopes.

This appendix presents graphs of Z polarity vs. distribution energy ratios obtained from three HSPC population labelled for CD34 epitopes. These graphs illustrate the difference between both measurements, between the different epitopes, and show how such measurements could be used for graphical displays.

Epitope analysis of 3 cell populations
Distribution energy vs. Z polarity

Epitopes I-III, Sample 1

% distribution energy (contact pole)

0%   22.7273%

59.0909%   18.1818%

% antigen density (contact pole)

Epitopes I-II, Sample 2

% distribution energy (contact pole)

12.5000%   33.3333%

50%   4.1667%

% antigen density (contact pole)

Epitopes I-II, Sample 3

% distribution energy (contact pole)

14.2857%   50%

25%   10.7143%

% antigen density (contact pole)
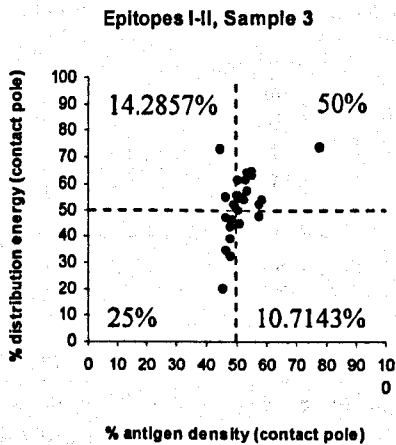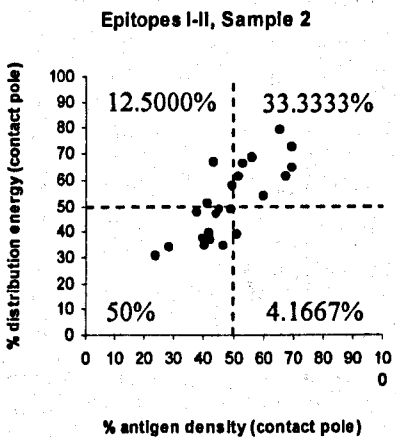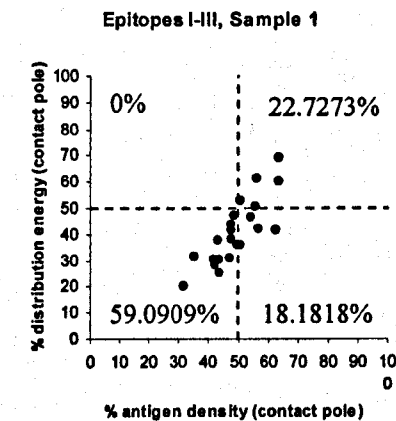
Detailed epitope analysis of 3 cell populations
Distribution energy vs. Z polarity

Epitope I, energy vs. density ratios, sample 1
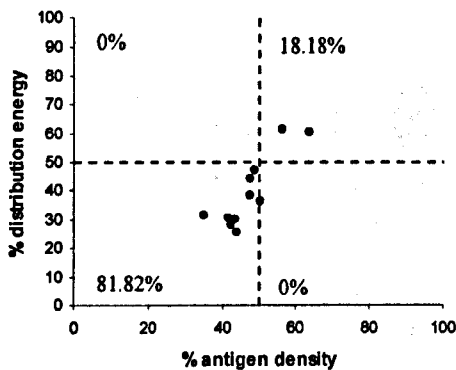
Epitope III, energy vs. density ratios, sample 1
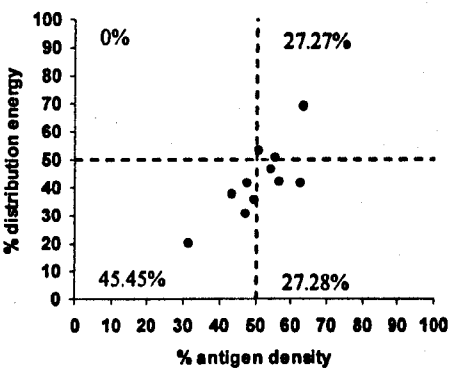
Epitope I, energy vs. density ratios, sample 2
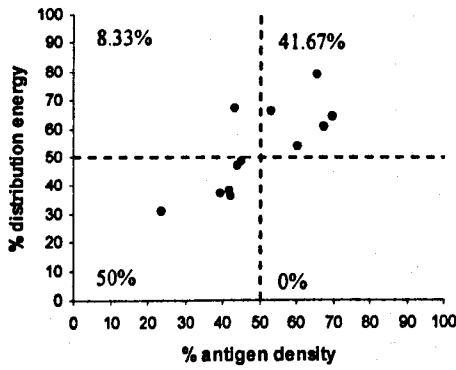
Epitope II, energy vs. density ratios, sample 2
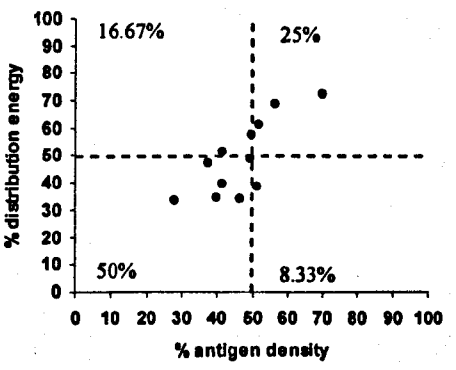
Epitope I, energy vs. density ratios, sample 3
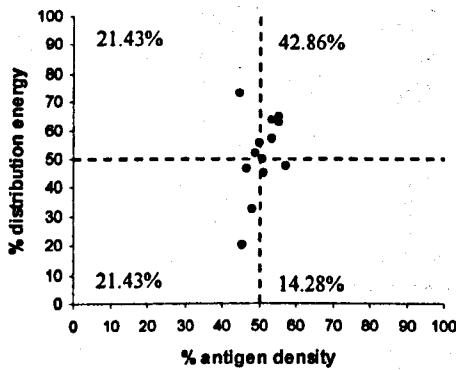
Epitope II, energy vs. density ratios, sample 3
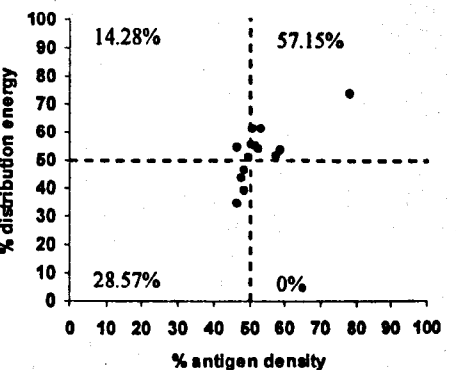
Epitope analysis of 3 cell populations
Distribution energy compared to Z polarity
Class X epitope vs. Class Y epitope

Epitope I vs. epitope III ratios, sample 1
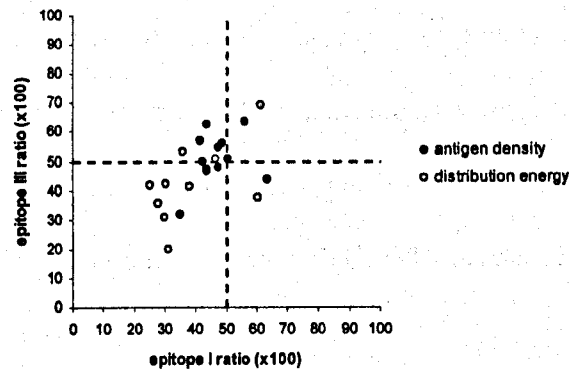
Epitope I vs. epitope II ratios, sample 2
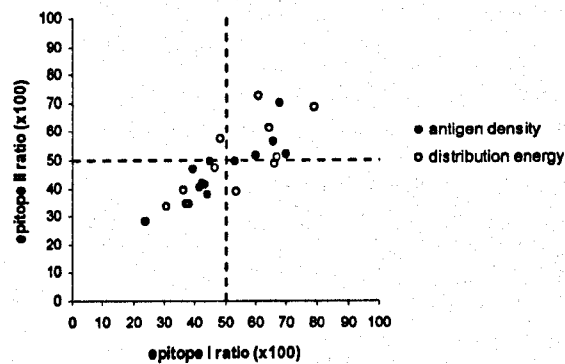
Epitope I vs. epitope II ratios, sample 3

Appendix E

E-I
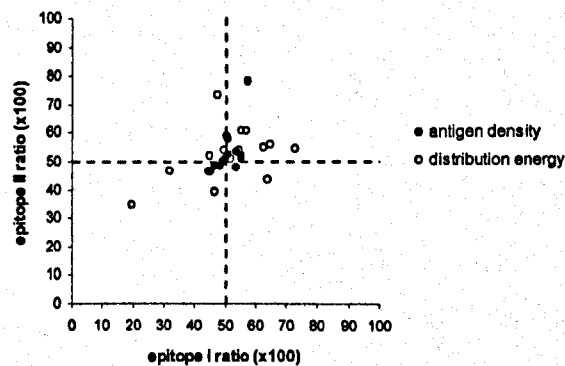
Example of statistical analysis performed for comparison of series of ADM measurements.

This appendix presents paired T-tests used to assess the similarities/differences between features measured from CD34 epitope combinations. Such statistical tests are used to compare all series of measurements within the HSPC samples. Between HSPC feature comparison was performed by unpaired T-test. An example of descriptive statistics from dual CD34/CD164 distributions is also presented along with a typical correlation analysis.

Inter-epitope T-test analysis

Antigen density

Pair 1 = ep. I-II, CB sample 04/08/03
Pair 2 = ep. I-II, CB sample 22/07/03
Pair 3 = ep. I-III, CB sample 25/06/03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00005	3.0725	13	.58678	.16274
	VAR00006	3.8521	13	.95424	.26466
Pair 2	VAR00003	2.5640	13	1.41243	.39174
	VAR00004	2.5481	13	1.07985	.29950
Pair 3	VAR00024	2.1545	11	1.19209	.35943
	VAR00025	2.6787	11	1.13026	.34079

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00005 & VAR00006	13	.685	.010
Pair 2	VAR00003 & VAR00004	13	.795	.001
Pair 3	VAR00024 & VAR00025	11	.274	.415

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	VAR00005 - VAR00006	-.7796	.69857	.19375	-1.2018	-.3575	-4.024	12	.002
Pair 2	VAR00003 - VAR00004	.0159	.85796	.23795	-.5025	.5344	.067	12	.948
Pair 3	VAR00024 - VAR00025	-.5242	1.40023	.42218	-1.4649	.4165	-1.242	10	.243

Inter-epitope T-test analysis

polarity XY

Pair 1 = ep. I-II, CB sample 22/07/03
Pair 2 = ep. I-II, CB sample 04/08/03
Pair 3 = ep. I-III, CB sampl 25/06/03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00003	1.9941	13	1.67133	.46354
	VAR00004	2.0724	13	1.95466	.54212
Pair 2	VAR00005	2.0270	13	.63401	.17584
	VAR00006	1.7350	13	.49006	.13592
Pair 3	VAR00024	1.6762	11	.62093	.18722
	VAR00025	1.8217	11	.77402	.23338

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00003 & VAR00004	13	.988	.000
Pair 2	VAR00005 & VAR00006	13	.328	.274
Pair 3	VAR00024 & VAR00025	11	.204	.546

Paired Samples Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	VAR00003 - VAR00004	-.0783	.40102	.11122	-.3206	.1640	-.704	12	.495
Pair 2	VAR00005 - VAR00006	.2920	.66220	.18366	-.1081	.6922	1.590	12	.138
Pair 3	VAR00024 - VAR00025	-.1455	.88776	.26767	-.7419	.4509	-.544	10	.599

Inter-epitope T-test analysis

polarity Z

Pair 1 = ep. I-II, CB sample 22/07/03
Pair 2 = ep. I-II, CB sample 04/08/03
Pair 3 = ep. I-III, CB sample 25/06/03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00003	.5365	13	.19608	.05438
	VAR00004	.4856	13	.13062	.03623
Pair 2	VAR00005	.4955	13	.03363	.00933
	VAR00006	.5011	13	.03558	.00987
Pair 3	VAR00024	.5432	11	.19127	.05767
	VAR00025	.5298	11	.09858	.02972

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00003 & VAR00004	13	.880	.000
Pair 2	VAR00005 & VAR00006	13	.410	.164
Pair 3	VAR00024 & VAR00025	11	-.070	.838

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	VAR00003 - VAR00004	.0509	.10205	.02830	-.0108	.1126	1.798	12	.097
Pair 2	VAR00005 - VAR00006	-.0056	.03763	.01044	-.0283	.0172	-.535	12	.602
Pair 3	VAR00024 - VAR00025	.0133	.22123	.06670	-.1353	.1620	.200	10	.845

E-IV

Inter-epitope T-test analysis

Number of clusters

Pair 1 = ep. I-II, CB sample 22/07/03
Pair 2 = ep. I-II, CB sample 04/08/03
Pair 3 = ep. I-III, CB sample 25/06/03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00003	9.9231	13	3.17442	.88042
	VAR00004	10.6154	13	2.78503	.77243
Pair 2	VAR00005	19.6154	13	3.96943	1.10092
	VAR00006	18.9231	13	2.87117	.79632
Pair 3	VAR00024	18.4545	11	5.61006	1.69150
	VAR00025	18.2727	11	5.60519	1.69003

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00003 & VAR00004	13	.977	.000
Pair 2	VAR00005 & VAR00006	13	.933	.000
Pair 3	VAR00024 & VAR00025	11	.578	.063

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	VAR00003 - VAR00004	-.6923	.75107	.20831	-1.1462	-.2384	-3.323	12	.006
Pair 2	VAR00005 - VAR00006	.6923	1.65250	.45832	-.3063	1.6909	1.511	12	.157
Pair 3	VAR00024 - VAR00025	.1818	5.15399	1.55399	-3.2807	3.6443	.117	10	.909

Inter-epitope T-test analysis

Antigen density in clusters

Pair 1 = ep. I-II, CB sample 22/07/03
Pair 2 = ep. I-II, CB sample 04/08/03
Pair 3 = ep. I-III, CB sample 25/06/03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00003	7.5224	13	3.20233	.88817
	VAR00004	8.3586	13	3.74155	1.03772
Pair 2	VAR00005	7.0748	13	4.02398	1.11605
	VAR00006	7.8664	13	2.94748	.81748
Pair 3	VAR00024	7.1233	11	1.45989	.44017
	VAR00025	7.0442	11	2.87995	.86834

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00003 & VAR00004	13	-.222	.466
Pair 2	VAR00005 & VAR00006	13	.234	.442
Pair 3	VAR00024 & VAR00025	11	.166	.626

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	VAR00003 - VAR00004	-.8362	5.43874	1.50844	-4.1228	2.4504	-.554	12	.590
Pair 2	VAR00005 - VAR00006	-.7916	4.39720	1.21956	-3.4488	1.8656	-.649	12	.528
Pair 3	VAR00024 - VAR00025	.0791	3.00488	.90601	-1.9396	2.0978	.087	10	.932

Inter-epitope T-test analysis

% southbound clusters

Pair 1 = ep. I-II, CB sample 22/07/03
Pair 2 = ep. I-II, CB sample 04/08/03
Pair 3 = ep. I-III, CB sample 25/06/03

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00003	.4855	13	.20196	.05601
	VAR00004	8.3586	13	3.74155	1.03772
Pair 2	VAR00005	.5514	13	.21563	.05980
	VAR00006	.4179	13	.21953	.06089
Pair 3	VAR00024	.3971	11	.23931	.07216
	VAR00025	.3900	11	.16620	.05011

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00003 & VAR00004	13	.401	.175
Pair 2	VAR00005 & VAR00006	13	-.124	.687
Pair 3	VAR00024 & VAR00025	11	.057	.867

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	VAR00003 - VAR00004	-7.8732	3.66528	1.01657	-10.0881	-5.6583	-7.745	12	.000
Pair 2	VAR00005 - VAR00006	.1335	.32619	.09047	-.0636	.3306	1.476	12	.166
Pair 3	VAR00024 - VAR00025	.0071	.28341	.08545	-.1833	.1975	.083	10	.936

CD164/CD34 distributions
Descriptive statistics

Features measured from ADM on CD164/CD34 dual distributions

Pair 1 = Antigen density
Pair 2 = Polarity XY
Pair 3 = Polarity Z
Pair 4 = Number of clusters
Pair 5 = Antigen density in clusters
Pair 6 = % southbound clusters

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00003	1.2993	5	.16215	.07252
	VAR00029	1.4292	5	.21307	.09529
Pair 2	VAR00004	1.5156	5	.23985	.10726
	VAR00030	1.3062	5	.13748	.06148
Pair 3	VAR00005	.4901	5	.01953	.00873
	VAR00031	.4973	5	.04007	.01792
Pair 4	VAR00006	19.8000	5	6.05805	2.70924
	VAR00032	21.2000	5	5.89067	2.63439
Pair 5	VAR00024	2.0406	5	.34527	.15441
	VAR00033	2.5528	5	.56003	.25045
Pair 6	VAR00026	.3934	5	.21750	.09727
	VAR00035	.4598	5	.27108	.12123

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	VAR00003 & VAR00029	5	.979	.004
Pair 2	VAR00004 & VAR00030	5	.347	.568
Pair 3	VAR00005 & VAR00031	5	.979	.004
Pair 4	VAR00006 & VAR00032	5	.429	.471
Pair 5	VAR00024 & VAR00033	5	.834	.079
Pair 6	VAR00026 & VAR00035	5	.745	.149

CD164/CD34 distributions
T-Test analysis

Features measured from ADM on CD164/CD34 dual distributions

Pair 1 = Antigen density
Pair 2 = Polarity XY
Pair 3 = Polarity Z
Pair 4 = Number of clusters
Pair 5 = Antigen density in clusters
Pair 6 = % southbound clusters

Paired Samples Test

| | | Paired Differences | | | | | | |
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper	t	df	Sig. (2-tailed)
Pair 1 VAR00003 - VAR00029	-.1300	.06352	.02841	-.2088	-.0511	-4.575	4	.010
Pair 2 VAR00004 - VAR00030	.2094	.23145	.10351	-.0779	.4968	2.023	4	.113
Pair 3 VAR00005 - VAR00031	-.0072	.02134	.00954	-.0337	.0193	-.757	4	.491
Pair 4 VAR00006 - VAR00032	-1.4000	6.38749	2.85657	-9.3311	6.5311	-.490	4	.650
Pair 5 VAR00024 - VAR00033	-.5121	.33207	.14851	-.9245	-.0998	-3.449	4	.026
Pair 6 VAR00026 - VAR00035	-.0664	.18163	.08123	-.2919	.1591	-.817	4	.460