Detection and Segmentation of Nucleoids Based on Gradient Path Labelling

João Santinha, André D. Mora and José Fonseca
Computational Intelligence Research Group (CA3), UNINOVA – Centre of Technology and Systems, Caparica, Portugal
Email: {jsantinha, atm, jmf}@uninova.pt

Nádia Gonçalves and Andre S. Ribeiro
Laboratory of Biosystem Dynamics – Department of Signal Processing, Tampere University of Technology, Tampere, Finland
Email: {nadia.moreiragoncalves, andre.ribeiro}@tut.fi

***Abstract*—Cellular aging is one of the topics that live cell imaging can assist. With age, there is an increase of aggregates of misfolded proteins, to which age-related diseases have been linked to. In *Escherichia coli*, protein aggregates linked to its aging process exhibit a spatial distribution that appears to be caused by the nucleoid at midcell. To correlate the locations of protein aggregates and the nucleoid, it is necessary to detect and segment the nucleoid from microscopy images. We present an adaptation of methods for Drusens' detection and segmentation to nucleoids in *E. coli*. The size of the nucleoid, extracted using the method here proposed, was compared with an alternative measure (FWHM-based measure) and with the regions of anisotropies in aggregates motions. These comparisons suggest that our new method is of use, providing more accurate minor axis lengths. Also, it provides additional measures, such as the nucleoid's center orientation angle, area, and pixel list.**

***Index Terms*—*Escherichia coli*, cellular aging, nucleoid detection, nucleoid segmentation**

I. Introduction

Ageing is an important topic in current live cell microscopy studies. While it is a fundamental characteristic of any living system, its underlying principles remain mysterious [1]. It has been established that one of its consequences, namely the accumulation of aggregates of misfolded proteins, is a likely cause of age-related diseases (e.g., Huntington's, Alzheimer's, spongiform encephalopathies, Parkinson's, and cataracts) [1], [2].

Recent studies have conducted microscopy observations of unicellular models with the aim of revealing the underlying mechanisms related to cellular ageing processes [1]. The use of *Escherichia coli* as a model organism, has elucidated much facts on the process of protein aggregation in this bacterium and promises significant advances in our understanding of cellular aging [1], [2]. The emergence of "aged" cells in *E. coli* populations has been linked to the accumulation of protein aggregates at the older pole of the aged bacterium [1]. The protein aggregates appear to be deposited at the poles, which, combined with cell division, results in asymmetric damage inheritance [2], [3].

The preference of the protein aggregates to locate at the cell poles has been attributed to macromolecular crowding [1], [4]. The macromolecular crowding is a consequence of the presence of the nucleoid, where the 4.6 kilobases (kb) genome DNA of *Escherichia coli* locates [5]. The DNA forms nucleoprotein complexes with at least 10 major DNA-binding proteins, including the DNA polymerases, the proteins involved in recombination and repair of DNA, and RNA polymerases, along with about 100 species of transcription factors that are associated with the nucleoid at some points in time [6] [7].

To study the effects of macromolecular crowding on the heterogeneous spatial distribution of the protein aggregates, it is advantageous to detect the aggregates and the nucleoids simultaneously, so as to correlate their spatial location. This can be achieved by fluorescent tagging. Even though, with this method, neither the nucleoids nor aggregate 'spots' have clear borders, it is necessary to perform their segmentation in order to establish spatial correlations.

As the nucleoid segmentation is hampered by lack of nucleoid borders, contour-based segmentation methods cannot be used. However, recent studies indicate that the Gaussian function can be implemented as the Point Spread Function (PSF) model, for fitting the position of a fluorescent emitter in a cell [8]. Here, we propose an adaptation of a method previously used in automatic Drusen detection in retinal images [9] to the detection and segmentation of nucleoids and we demonstrate its applicability on images of cells at different temperatures, which causes nucleoid structures to differ, e.g., in size and positioning.

Manuscript received January 28, 2015; revised April 1, 2015.

II. Methods

A. Cell Growth, Nucleoid Staining and Microscopy Imaging

E. coli strain DH5α-PRO cells information is: deoR, endA1, gyrA96, hsdR17(r_K^- m_K^+), recA1, relA1, supE44, thi-1, Δ(*lacZYA-argF*)U169, Φ80δ*lacZ*ΔM15, F-, λ-, P_{N25}/*tetR*, P_{lacI}^q/*lacI*, and SpR.

Bacterial cell cultures were grown in lysogeny broth (LB) media. The chemical components of LB (Tryptone, Yeast extract, and NaCl) were purchased from LabM (Topley House, Bury, Lancashire, UK) and the antibiotics from Sigma-Aldrich (St. Louis, MO). Isopropyl b-D-1-thiogalactopyranoside (IPTG) and anhydrotetracycline (aTc) used for induction of the target genes are from Sigma-Aldrich. These cells contained two plasmids: (i) PROTET-K133 carrying $P_{LtetO-1}$-*MS2d-GFP*, and (ii) pIG-BAC (P_{lac}-*MS2-96bs*) vector, carrying a 96 MS2 binding site array under the control of P_{lac} (a kind gift from I. Golding, Baylor College of Medicine, Houston, TX), allowing the detection of synthetic, fluorescent protein aggregates.

Agarose (Sigma-Aldrich) was used for microscope slide gel preparation. 40, 6-diamidino-2-phenylindole (DAPI) from Sigma-Aldrich was used to stain cell nucleoids, allowing the detection of nucleoids in individual live cells. *E. coli* cells were placed at 37 ℃ from overnight cultures at 250 RPM shaking for 1 hour until reaching an OD600 ≈ 0.3. Cells were fully induced during 1 hour, in order to produce fluorescent protein aggregate complexes, until reaching OD600 ≈ 0.5. Temperatures of 10 ℃, 24 ℃, 37 ℃ and 43 ℃ were set to perform this study. Except for the 37 ℃ case, cells were transferred from optimal temperature 37 ℃ to either 10 ℃, 24 ℃, or 43 ℃, in which they were kept for 1 hour. Cells were then centrifuged at 8000 RPM for 1 minute, suspended in 3.7% phosphate buffered saline (PBS), and incubated for 10 minutes for fixation. DAPI protein stains nucleoids specifically with little or no cytoplasmic labeling [10]. DAPI (0.2 μg/ml) was added to the cells suspended in PBS and incubated in the dark at room temperature for 20 min. DAPI staining "memorizes" the state of cells at the moment cells were fixed. Images were taken for each temperature. Finally, cells were washed twice with PBS, and 5 μl of culture were placed on a 1% agarose gel pad prepared in fresh LB.

Cells were visualized using a Nikon Eclipse (Ti-E, Nikon) inverted microscope with a 100x Apo TIRF (1.49 NA, oil) objective. The software used for image acquisition was NIS-Elements (Nikon). Confocal images were taken by a C2+ (Nikon) confocal laser-scanning system. The pinhole size was set to 1.2 AU. For confocal images, the size of a pixel corresponds to 0.124 μm using scan area resolution of 1024x1024 pixels. To visualize fluorescent complexes, GFP fluorescence was measured using a 488 nm laser (Melles-Griot) and a band-pass 500-530 nm emission filter (HQ515/30m, Nikon). Phase contrast images were captured using a CCD camera (DS-Fi2, Nikon). The size of the images was 2560x1920 pixels, in which a pixel corresponds to 0.049 μm. Epifluorescence images, for the visualization of DAPI-stained nucleoids, were taken using a mercury lamp excitation and a DAPI filter cube (Nikon).

It is noted that the fluorescent aggregates studied in these images, namely, synthetic RNA molecules tagged with multiple MS2-GFP proteins, have been shown to exhibit similar behavior to that of natural aggregates [11].

III. Nucleoid Detection and Segmentation

The method proposed here comprises of two steps: nucleoid detection, using the 'Gradient Path Labelling' (GPL) algorithm, and nucleoid segmentation, using a three-dimensional modified Gaussian function.

A. Nucleoid Detection

Nucleoid detection is performed using the GPL algorithm [9]. This method starts by labelling each pixel, based on its gradient azimuth, and propagating the labels according to its gradient paths. The reduction of labels is obtained by applying equivalences (two labels are tagged as equivalents when both belong to the same maximum). After the reduction of labels, a segmented image is obtained with the number of labels as nucleoids. The nucleoid detection was done for each cell.

B. Modified Gaussian Segmentation

For cells with only one detected nucleoid we performed nucleoid segmentation. The nucleoid segmentation uses a three-dimensional modified Gaussian function, similar to [9] and described by (1).

$$G(x,y) = A \cdot exp(-(a(x - x_0)^2 + 2b(x - x_0)(y - y_0) + c(y - y_0)^2)^{(2/d)}) + z_0, \quad (1)$$

where

$$a = cos^2\theta / 2\sigma_x^2 + sin^2\theta / 2\sigma_x^2, \quad (2)$$

$$b = -sin\,2\theta / 4\sigma_x^2 + sin\,2\theta / 4\sigma_x^2, \quad (3)$$

$$c = sin^2\theta / 2\sigma_x^2 + cos^2\theta / 2\sigma_x^2. \quad (4)$$

This function models the nucleoid. It allows translation in the *xyz* axes (x_0, y_0, z_0), amplitude scaling (A), rotation (θ), width in *x*-plane (σ_x), width in y-plane (σ_y) and the amplitude profile between square shape, bell shape and thin shape (d). The fitting to the nucleoid is done by the Levenberg-Marquardt Least-Squares [12] optimization algorithm, using a predefined $d = 10$ empirically selected to allow the use of the value z_0 as threshold so as to obtain the segmented masks of nucleoids.

From the segmentation mask, we extract the nucleoid's major and minor axes length, the center's position, and the orientation angle, the area and the pixels list of the nucleoid. From the major axis length, we can study changes in the width of the midcell region, defined as the region containing the nucleoid. The measure of the nucleoid's minor axis length is of interest, e.g., in the study of the nucleoids ability to retain complexes at the pole. The additional measures can further assist in

establishing correlations between the spatial distributions of the protein aggregates and nucleoid in each cell.

IV. Results and Discussion

To test the nucleoid detection and segmentation method, for each condition, we selected cells containing only one nucleoid. From these, we extracted the major and the minor axes lengths of the segmented nucleoid, and normalized them by the cell length and width, respectively.

An alternative measure of the nucleoid was obtained for comparison. Namely, for each cell with one single nucleoid detected, we calculated the full width at half maximum (FWHM) of the curve obtained from the sum intensity of the DAPI signal along the minor cell axis, for each position along the major cell axis. This method was also used here to measure the minor cell axis. For that, we summed the intensity of the DAPI signal along the major axis, for each position along the minor cell axis.

In Table I, the mean major axes lengths as measured by the two methods are shown for each temperature condition. A statistical comparison between the measures taken at the same temperature was also performed. From Table I, we find that the two measures are statistically indistinguishable (p-values of the t-test > 0.01), for all temperatures.

Also, both methods show a similar decrease in relative nucleoid size with the increase in temperature.

For the different temperature conditions, we also observe that the mean relative nucleoid major axis length is consistent with the region of anisotropies of the displacement distribution of aggregates along the major cell axis reported in [13].

The measure of the nucleoid minor axis length is also relevant as variations in this quantity are expected to alter the degree of retention of the aggregates at the cell poles. Table II shows the mean relative minor axes lengths measured by the two methods, for each temperature. Although the changes in the mean relative nucleoid minor axis length with temperatures are identical, the two measures are statistically distinguishable (p-values of the t-test < 0.01), for all temperatures.

TABLE I. Mean Relative Nucleoid Major Axis Length Obtained by the Two Methods, in Each Condition.

Temp.	No. of Cells	Relative nucleoid major axis length using GPL detection and modified Gaussian segmentation – mean (std)	p-value of t-test between the two distributions	Relative nucleoid major axis length using fluorescence projection along cell major axis – mean (std)
10 °C	206	0.63 (0.12)	0.59	0.62 (0.14)
24 °C	509	0.56 (0.11)	0.05	0.58 (0.14)
37 °C	367	0.53 (0.11)	0.19	0.52 (0.13)
43 °C	231	0.47 (0.12)	0.03	0.50 (0.14)

A visual comparison between the cells' DAPI stained nucleoid images, the respective segmentation results from GPL-based method (example Fig. 1) and the minor axis lengths measured by the FWHM-based method were used to determine which method obtains a better measure of the minor axis of the nucleoid. From this comparison, we confirmed that the GPL-based method performs an adequate segmentation of the nucleoid, as shown in Fig. 1, while the FWHM-based method underestimates the minor axis length of nucleoids.

TABLE II. Mean Relative Nucleoid Minor Axis Length Obtained by the Proposed Method, in Each Condition.

Temp.	No. of Cells	Relative nucleoid minor axis length using GPL detection and modified Gaussian segmentation – mean (std)	p-value of t-test between the two distributions	Relative nucleoid minor axis length using fluorescence projection along cell major axis – mean (std)
10 °C	206	0.63 (0.09)	< 0.01	0.59 (0.12)
24 °C	509	0.69 (0.09)	< 0.01	0.66 (0.12)
37 °C	367	0.63 (0.09)	< 0.01	0.58 (0.12)
43 °C	231	0.70 (0.12)	< 0.01	0.565 (0.16)

The efficiency of the nucleoid detection method was further verified by inspection, which suggests that the error rate is negligible (see example Fig. 1).

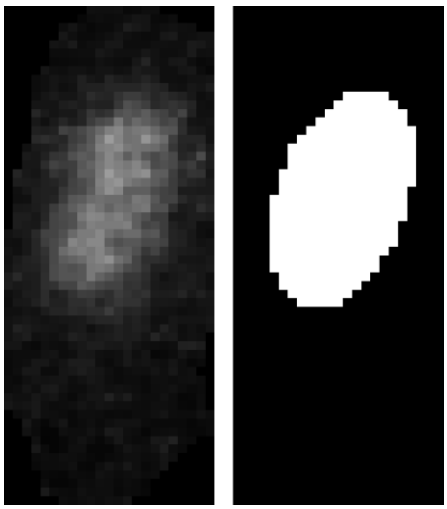

Figure 1. Example of a cell with DAPI stained nucleoid (left) and the nucleoid segmentation in white (right).

In order to evaluate analytically the proposed methodology, the dice coefficient [14] between the automatic segmentation and an atlas-like result could have been calculated. However, to create an atlas we would need the nucleoid segmentation from specialists or a validated software (that to the best of our knowledge is not available). Since the segmentation from specialists typically presents high variability it implies a great number of specialists to reduce the segmentation error and find a valid ground truth that can be used for the

algorithm evaluation. Once that we would need the segmentation of a large number of nucleoid from a considerable number of specialists this evaluation was left for future work.

V. Conclusion

In the present work, the GPL detection algorithm and a modified Gaussian function were adapted to detect and segment nucleoids from microscopy images of cells with tagged nucleoids and synthetic, fluorescent aggregates, shown to behave similarly to natural aggregates [11].

To verify the applicability of the segmentation method we compared its results with an FWHM-based method when applied to images of cells in various temperatures.

Measures of the relative major axis length of the nucleoid obtained from both methods were indistinguishable and showed that both the behavior and the mean relative nucleoid width along the major cell axis are consistent with measured regions of anisotropies of the displacement distribution of the aggregates along the major cell axis reported in [13].

The measures of the minor axis length of the nucleoid obtained by the two methods were statistically distinguishable for all temperatures. To determine which method best measures the nucleoid's minor axis width, we compared their results when applied to images of the cells with DAPI stained nucleoid. We found that the GPL-based method is more adequate as segmentation method as it performs a more realistic description the nucleoid region, while the FWHM-based method tends to underestimate the minor axis length of the nucleoid.

These results demonstrate the usefulness of the new method for nucleoid detection and segmentation. Additionally, it allows extracting additional measures, namely, the center, the orientation angle, area and the pixel list, will assist in the study of the relationship between the protein aggregates' spatial distribution and the nucleoid.

Acknowledgments

Work supported by Fundação para a Ciência e a Tecnologia, Portugal (PTDC/BBB-MET/1084/2012 to JMF and ASR) and the Academy of Finland [126803 to ASR].

References

[1] A. S. Coquel, J. P. Jacob, M. Primet, A. Demarez, M. Dimiccoli, T. Julou, *et al*., "Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid Macromolecular crowding effect," *PLoS Comput. Biol.*, vol. 9, no. 4, p. e1003038, May 2013.

[2] A. B. Lindner, R. Madden, A. Demarez, E. J. Stewart, and F. Taddei, "Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 8, pp. 3076-3081, Feb. 2008.

[3] A. Rokney, M. Shagan, M. Kessel, Y. Smith, I. Rosenshine, and A. B. Oppenheim, "*E. coli* Transports Aggregated Proteins to the Poles by a Specific and Energy-Dependent Process," *J. Mol. Biol.*, vol. 392, no. 3, pp. 589-601, Sep. 2009.

[4] I. Golding and E. C. Cox, "Physical nature of bacterial cytoplasm," *Phys. Rev. Lett.*, vol. 96, no. 9, p. 098102, Mar. 2006.

[5] M. Macvanin and S. Adhya, "Architectural organization in *E. coli* nucleoid," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1819, no. 7, pp. 830-835, Jul. 2012.

[6] J. E. Rebollo, V. François, and J. M. Louarn, "Detection and possible role of two large nondivisible zones on the Escherichia coli chromosome.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, no. 24, pp. 9391-9395, Dec. 1988.

[7] T. A. Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama, "Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid," *J. Bacteriol.*, vol. 181, pp. 6361-6370, Oct. 1999.

[8] S. Stallinga and B. Rieger, "Accuracy of the gaussian point spread function model in 2D localization microscopy.," *Opt. Express*, vol. 18, no. 24, pp. 24461-24476, Nov. 2010.

[9] A. D. Mora, P. M. Vieira, A. Manivannan, and J. M. Fonseca, "Automated drusen detection in retinal images using analytical modelling algorithms," *Biomed. Eng. Online*, vol. 10, no. 59, pp. 1-15, Jan. 2011.

[10] B. Chazotte, "Labeling nuclear DNA using DAPI," *Cold Spring Harb. Protoc.*, vol. 6, pp. 80-82, 2011.

[11] A. Gupta, J. Lloyd-Price, S. M. D. Oliveira, R. N. Venkata, and A. S. Ribeiro, "*In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in Escherichia coli," *Biophys. J.*, vol. 106, pp. 1926-1937, May 2014.

[12] J. J. More, "The Levenberg-marquardt algorithm: implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed. Berlin Heidelberg: Springer-Verlag, 1978, pp. 105-116.

[13] S. M. D. Oliveira, R. N. V. J. A. Santinha, N. Goncalves, M. Barandas, A. Gupta, J. M. Fonseca, *et al*., "Robustness of *Escherichia coli*'s polar retention mechanism of protein complexes to sub-optimal temperatures," unpublished.

[14] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297-302, 1945.

João Santinha was born in Faro, Portugal, in 1988. He graduated in Biomedical Engineering, Faculty of Sciences and Technology of Universidade Nova de Lisboa (FCT/UNL) in 2011.

He worked as an IT Consultant at Novabase – Business Solution, S.A.. Currently, he is a researcher at CA3 – Computational Intelligence Research Group of UNINOVA, where he participates in a research project, along with the Tampere University of Technology, to study the kinetics of aggregates in cell division and its correlation to functional aging from in vivo measurements, one event at a time. His main focus for this research is the development of methods for image segmentation, image registration and tracking of cells in temporal images. He has published 3 papers in international conference proceedings.

Nádia Gonçalves was born in Cape Verde, on 10th of July 1987. She graduated in Biology at the University of Coimbra, Portugal in 2009. She has a Master degree in Biochemistry from the University of Coimbra in 2012, with the master project been conducted at the Institute for Chemical and Biological Technologies, Oeiras, Portugal. The project consisted in the development of a random mutagenesis protocol to create a PpAzoR mutant library and a High-Throughput Screening assay to identify improved mutants.

Currently she is a PhD student at Tampere University of Technology, Tampere, Finland. She studies gene expression dynamics of the bacteriophage promoter T7, in *Escherichia coli* using single-cell, single-molecule *in vivo* measurements. The research is conducted at the Laboratory of Biosystems Dynamics under the supervision of Prof. Andre Ribeiro.

André D. Mora was born in Lisbon, Portugal, in 1975. André Damas Mora has a PhD on Electrical Engineering, specialty on Medical Image Processing, by New University of Lisbon, since 2010 and a MSc in Computer Science, since 2002. He is working has Professor in the Electrical Engineering Department of the New University of Lisbon.
His research activities are in the Centre of Technology and Systems, Uninova, where he collaborated in several National and European Funded Projects. His main research interests are Digital Image Processing, Medical Image Processing, Sensorial Systems and Decision Support Systems. He already published more than 50 papers on Journals and National and International Conferences, is co-author of several book Chapters, is co-author of a National Patent and was co-founder of a spin-off Company.

Andre S. Ribeiro was born in Luanda, Angola, in 1976. Andre S. Ribeiro graduated in Physics, University of Lisbon, Portugal in 1999. He has a PhD from IST, University Tecnica de Lisboa, Portugal (2004), where he studied information propagation in networks. From 2004 to 2007, he was a Postdoc at the University of Calgary, Canada, where he developed dynamic models of genetic circuits. Currently, he is the PI of the Laboratory of Biosystem Dynamics (LBD), and an Associate Professor at the Department of Signal Processing, Tampere University of Technology, Finland. The LBD studies regulatory mechanisms of gene expression dynamics in Escherichia coli using and developing single-cell, single-molecule *in vivo* measurements, along with state-of-the-art Signal Processing techniques and stochastic models and simulators of genetic circuits. Also, the LBD studies the spatio-temporal organization of molecules in cells, related to cellular aging, and synthetic genetic networks. At present, Andre Ribeiro has published 73 papers in international scientific journals, 83 papers in international conference proceedings, and 10 book chapters.
He has been awarded twice by the DSP, TUT for teaching and research activities.

José Fonseca was born in 1962, Lisbon, Portugal. Obtained his PhD on Electrical Engineering specialty Robotics and Integrated Manufacturing by Universidade Nova de Lisboa, 10th May 2001.
He his Professor on the Faculty of Sciences and Technology of the Universidade Nova de Lisboa (FCT/UNL) since 1992, and member of the Electrical Engineering Department since its foundation on 1993. Has been responsible for several courses in the area of sensorial processing and real-time processing.Since its PhD supervised two PhD Thesis, 30 MSc Thesis and 35 Final Diploma Projects. Published more than 60 papers on Journals and International Conferences, is co-author of several book Chapters, co-author of a National Patent and co-founder of a spin-off Company.
Prof. Fonseca has participated on more than ten European Projects as researcher and coordinator of the FCT/UNL participation. Was leader of taskforce eDelivery and leader of the workpackage Integrating Support Activities on the Network of Excellence "BIOPATTERN - Computational Intelligence for BioPattern Analysis to Support eHealth" FP6-2004-IST-2-508803 from 2004 to 2008.